JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

"ANALYSIS AND DESIGN OF R.C.C OVERHEAD TANK"

A project report submitted to the

Jawaharlal Nehru Technological University – Kakinada

In partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY – CIVIL ENGINEERING

BY

BHANUKIRAN.B - (17H71A0106)

KAVYA SRI.T - (17H71A0155)

AMITH.H - (17H71A0146)

GOPINADH.K - (17H71A0114)

GANESH RAJA.P - (17H71A0112)

Under the esteemed guidance of

Mr. V. VENKATESH, B.Tech

Assistant Professor

DEPARTMENT OF CIVIL ENGINEERING

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
Kanchikacherla - 521 180. Phone : 08678-273535
(Approved by AICTE & Affiliated to JNTU Kakinada)

2017-2021

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
Kanchikacherla - 521 180. Phone : 08678-273535
(Approved by AICTE & Affiliated to JNTU Kakinada)

CERTIFICATE

This is to certify that the project report entitled "ANALYSIS AND DESIGN R.C.C OVERHEAD TANK" Bachelor of Technology in Civil Engineering from Jawaharlal Nehru Technological University ,Kakinada is bonafied work done by **BHANUKIRAN.B(17H71A0106),KAVYA SRI.T (17H71A0155) ,AMITH.H (17H71A0146), GOPINADH.K (17H71A0114),GANESH RAJA.P (17H71A0112)** under my guidance during the academic year 2020-2021 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. V.VENKATESH, B.Tech	**Mr. N.V SUBBA RAO, M.E (Ph.D.)**
Assistant Professor	Associate professor
Project Guide	Head of the Department

Dr. K.SRINIVAS, Ph.D

Professor& Principal **External Examiner**

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

ANALYSIS AND DESIGN OF RETAINING WALLS WITH AND WITHOUT SHELVES

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

In partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY

In

CIVIL ENGINEERING

By

B. Chetana (17H71A0108)

Sd. Zaheer Abbas (17H71A0189)

K. Mahesh (17H71A0122)

G. Yaswanth (17H71A0188)

V. Rakesh Kumar (17H71A0131)

Under the Guidance of

Mr. A.V.V. SAIRAM, MTech

Assistant professor

Department of Civil Engineering

Devineni Venkata Ramana & Dr Hima Sekhar

MIC College of Technology

Accredited by NBA: EEE/MECH/ECE & NAAC with 'A' Grade

Kanchikacherla - 521180, Krishna Dist. A.P.

2020-21

CERTIFICATE

This is to certify that the project report entitled "**ANALYSIS AND DESIGNOF RETAINING WALLS WITH AND WITHOUT SHELVES**" being submitted for the partial fulfillment of the requirements for the award of the degree, **Bachelor of Technology in Civil Engineering from JAWAHARLALNEHRU TECHNOLOGICAL UNIVERSITY, KAKINADA,** is a bonafide work done by **Ms. BCHETANA(17H71A0108), Mr. SYED ZAHEER ABBAS (17H71A0189), Mr. K MAHESH (17H71A0122), Mr. G YASWANTH(17H71A0188), Mr. V RAKESH KUMAR (17H71A0131)** under my guidance during the academic year 2020-21 and it has been suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

External Examiner

Mr. A.V.V. SAIRAM, MTech
Assistant professor
Project Guide

Mr. N.V. SUBBA RAO, M.E, (PhD)
Assistant Professor
Head of the Department

Dr. K. SRINIVAS, PhD
Principal

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY
KAKINADA

DESIGN AND DETAILLING OF FORM WORK FOR A MULTI STOREYED BUILDING

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

In partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

CIVIL ENGINEERING
By

M.DHARANI (17H71A0109)
G.DHANUNJAY (18H75A0125)
I.GOPALA KRISHNA (18H75A0107)
M.ARUN KUMAR (18H75A0102)
P.AJAY KUMAR (18H75A0101)

Under the esteemed guidance of

Mr. N. V. SUBBARAO, (M.E Ph.D.)

HEAD OF THE DEPARTMENT

DEPARTMENT OF CIVIL ENGINEERING

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
Accredited by NBA: EEE|ME|CSE|ECE & NAAC with 'A' Grade
Kanchikacherla - 521180, Krishna Dist. A.P.

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
Accredited by NBA: EEE|ME|CSE|ECE & NAAC with 'A' Grade
Kanchikacherla - 521180, Krishna Dist. A.P.

CERTIFICATE

This is to certify that the project report entitled **"DESIGN AND DETAILING OF FORMWORK FOR A MULTISTOREYED BUILDING " Bachelor of Technology in Civil Engineering** from **Jawaharlal Nehru Technological University, Kakinada** is a bonafide work done by **M.DHARANI(17H71A0109),G.DHANUNJAY(18H75A0125), I.GOPALAKRISHNA(18H75A0107) M.ARUN KUMAR CHOWDARY (18H75A0102), P.AJAY KUMAR (18H75A0101)** under my guidance during the academic year 2020-2021 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. N.V. SubbaRao, (M.E Ph.D.)
Head of the department
Project Guide

Mr. N.V. SubbaRao,(M.E Ph.D.
Associate Professor
Head of the Department

Dr. K. Srinivas Rao, Ph.D.
Professor & Principal

External Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY
KAKINADA

A CASE STUDY ON SEWAGE TREATMENT PLANT

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

In partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
IN
CIVIL ENGINEERING
BY

CH.HARISH (17H71A0115)
G.MANIKANTA(17H71A0158)
B.HANUMANTHA RAO(17H71A0152)
K.VENKATA BABU RAO(17H71A0184)

Under the esteemed guidance of
MS.SK.AYEESHA
Assistant professor

Department of civil engineering

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
An Autonomous Institution

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
An Autonomous Institution

CERTIFICATE

This is to certify that the project report entitled **"A CASE STUDY ON SEWAGE TREATMENT PLANT" Bachelor of Technology in Civil Engineering** from **Jawaharlal Nehru Technological University, Kakinada** is a bonafide work done **CH.HARISH(17H71A0115),G.MANIKANTA(17H71A0158),B.HANUMANTHA RAO (17H71A0152),K.VENKATA BABU RAO(17H71A0184).**under my guidance during the academic year 2020-2021 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

MS.SK.AYEESHA
Assistant professor
Project Guide

Mr. N.V. Subba Rao,(M.E Ph.D.)
Associate Professor
Head of the Department

Dr.K Srinivas,Ph.D
Professor and principal

External Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

EXPERIMENTAL INVESTIGATION ON SOIL STABILIZATION USING TERRAZYME

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

In partial fulfilment of the award of the degree of

BACHELOR OF TECHNOLOY

In

CIVIL ENGINEERING

By

K. JOSHITHA (17H71A0116)

CH. RAJESH (17H71A0130)

V.SAI PAWAN (18H75A0119)

T. SANDEEP (17H71A0135)

T. LALU SAI CHANDRA (17H71A0121)

Under the esteemed guidance of

Ms. P. L. K. SOWJANYA, M. Tech

Assistant Professor

Department of Civil Engineering

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
An Autonomous Institution

NAAC 'A'

CERTIFICATE

This is to certify that the project report entitled " **STABILIZATION OF SOIL USING TERRAZYME FOR ROAD CONSTRUCTION " BACHELOR OF TECHNOLOGY IN CIVIL ENGINEERING** from **JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, KAKINADA,** is a bonafide work done by **K. JOSHITHA (17H71A0116), CH. RAJESH (17H71A0130), V. SAI PAWAN (18H75A0119), T. SANDEEP (17H71A0135), T. LALU SAI CHANDRA (17H71A0121).** Under my guidance during the academic year 2020-2021 and it has been for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Ms. P. L. K. SOWJANYA, M. Tech
Assistant Professor & Project Guide

Mr. M. V. SUBBA RAO M.E (Ph.D.)
Head of the Department

Dr. K. Srinivas, Ph.D.
Professor & Principle

External Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY
KAKINADA

DESIGN OF CONCRETE JACKETING SYSTEM FOR REPAIRING OF RCC COLUMN

A project report submitted to the
Jawaharlal Nehru Technological University Kakinada
In partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
CIVIL ENGINEERING

By

K.KEERTHIKA	**(17H71A0118)**
M.KONDAL RAO	**(18H75A0110)**
P.RAM GOPAL	**(18H75A0117)**
N.AVINASH	**(18H75A0103)**

Under the Guidance of
Mr. S.V. TEJAS KUMAR RAJU, M.Tech
Assistant professor
Department of Civil Engineering

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
Accredited by NBA: EEE,MECS,ECE,CE & NAAC with 'A' Grade
Kanchikacherla - 521180, Krishna Dist. A.P.

2020-2021

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
Accredited by NBA: EEE,MECS,ECE,CE & NAAC with 'A' Grade
Kanchikacherla - 521180, Krishna Dist. A.P.

CERTIFICATE

This is to certify that the project report entitled **"DESIGN OF CONCRETE JACKETING FOR REPAIRING OF RCC COLUMN" BACHELOR OF TECHNOLOGY IN CIVIL ENGINEERING** from **JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, KAKINADA,** is a bonafied work done by **Mrs. K.KEERTHIKA (17H71A0118), Mr. M.KONDAL RAO(18H75A0110), P.RAM GOPAL (18H75A0117), Mr. N. AVINASH (18H75A0103)** under my guidance during the academic year 2020-2021 and it has been suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. S.V.TEJAS KUMAR RAJU, M.Tech
Assistant professor
Project Guide

Mr. N.V.SUBBA RAO, (Ph.D)
Associate professor
Head of the Department

Dr. K.SRINIVAS, (Ph.D)
Principal

External Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

STOREY RESPONSE OF COMMERCIAL STRUCTURE IN DIFFERENT SEISMIC ZONES BY USING ETABS

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

In partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

CIVIL ENGINEERING

By

G. MANI SREE (17H71A0123) **R. PRASANTH (18H75A0115)**

A.SRAVAN (18H75A0122) **K. JITENDHRA (17H71A0153)**

SK. BADEMIYA (17H71A0148)

Under the esteemed guidance of

Mr. K. PRASAD, M.Tech.

Assistant Professor

DEPARTMENT OF CIVIL ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
An Autonomous Institution

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
An Autonomous Institution

CERTIFICATE

This is to certify that the project report entitled **"STOREY RESPONSE OF COMMERCIAL STRUCTURE IN DIFFERENT SEISMIC ZONES BY USING ETABS " Bachelor of Technology in Civil Engineering** from **Jawaharlal Nehru Technological University, Kakinada** is a bonafide work done by **G. MANI SREE(17H71A0123), R.PRASANTH(18H75A0115), A. SRAVAN(18H75A0122), K. JITENDHRA(17H71A0153), SK. BADEMIYA(17H71A0148).** under my guidance during the academic year 2020-2021 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. K. PRASAD, M.Tech
Assistant professor
Project Guide

Mr. N.V. Subba Rao, M.E (Ph.D.)
Associate Professor
Head of the Department

Dr. K. Srinivas, Ph.D.
Professor & Principal

External Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

A PREVIEW OF COVID-19 IMPACT ON INFRASTRUCTURE OF INDIA

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

In partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

CIVIL ENGINEERING

By

MYTHILI SRAVANI GUNJA (17H71A0125)

RAGHUNANDAN (18H75A0116)

SAI KUMAR (18H75A0120)

KOWSHIK (17H71A0182)

SATYA PRASAD (17H71A0174)

Under the esteemed guidance of

Mr. AJAY KUMAR VERMA, M.Tech.

Assistant Professor

DEPARTMENT OF CIVIL ENGINEERING

Devineni Venkata Ramana & Dr Hima Sekhar

MIC College of Technology

Accredited by NBA, HYDERABAD & NAAC with 'A' Grade

Kanchikacherla - 521180, Krishna Dist., A.P.

Devineni Venkata Ramana & Dr Hima Sekhar

MIC College of Technology

Accredited by NBA, HYDERABAD & NAAC with 'A' Grade

Kanchikacherla - 521180, Krishna Dist., A.P.

CERTIFICATE

This is to certify that the project report entitled "**A PREVIEW OF COVID-19 IMPACT ON INFRASTRUCTURE OF INDIA**" **Bachelor of Technology in Civil Engineering** from **Jawaharlal Nehru Technological University, Kakinada** is a bonafied work done by **MYTHILI SRAVANI GUNJA (17H71A0125), RAGHUNANDAN (18H75A0116), SAI KUMAR (18H75A0120) , KOWSHIK (17H71A0182), SATYA PRASAD (17H71A0174)** under my guidance during the academic year 2020-2021 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. AJAY KUMAR VERMA, M.Tech

Assistant professor

Project Guide

Mr. N.V. Subba Rao, (M.E Ph.D.)

Associate Professor

Head of the Department

Dr Kadivendi Srinivas, (PhD)

Principal

External Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

EFFECT OF WIND LOAD ON MULTI STOREY BUILDING IN DIFFERENT TERRAIN CATEGORIES BY USING STAAD PRO

A project report submitted to the

Jawaharlal Nehru Technological University – Kakinada

In partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY – CIVIL ENGINEERING

BY

SAI PRADEEP.K - (17H71A0133)

LOKESH.Y - (17H71A0157)

NAVEEN TEJA.N - (17H71A0161)

GOVINDARAJULU.S - (18H75A0108)

MANOJ.M - (17H71A0124)

Under the esteemed guidance of

Mr. AAMEER SOHIEL

DEPARTMENT OF CIVIL ENGINEERING

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
Kanchikacherla - 521 180. Phone : 08678-273535
(Approved by AICTE & Affiliated to JNTU Kakinada)

2017-2021

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
Kanchikacherla - 521 180. Phone : 08678-273535
(Approved by AICTE & Affiliated to JNTU Kakinada)

CERTIFICATE

This is to certify that the project report entitled "EFFECT OF WIND LOAD ON MULTI STOREY BUILDING IN DIFFERENT TERRAIN CATEGORIES BY USING STAAD PRO" Bachelor of Technology in Civil Engineering from Jawaharlal Nehru Technological University ,Kakinada is a bonified work done by SAI PRADEEP.K(17H71A0133),LOKESH.Y (17H71A0157),NAVEEN TEJA.N(17H71A0161),GOVINDARAJULU.S(18H75A0108),MANOJ.M(17H71A0124) under my guidance during the academic year 2020-2021 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. AAMEER SOHIEL, M.tech	Mr. N.V SUBBA RAO, M.E (Ph.D.)
Assistant Professor	Associate Professor
Project Guide	Head of the Department

Dr. K.SRINIVAS

Professor & Principal

External Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY
KAKINADA

NON-LINEAR DYNAMIC ANALYSIS OF T-SHAPED IRREGULAR RC BUILDING

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

In partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

CIVIL ENGINEERING

Submitted by

A. SAI TRIVENI	**17H71A0134**
MD. SHARUK	**17H71A0175**
G. SAI VINAY	**17H71A0173**
B.TEJA HARINI	**17H71A0179**
P.PRASAD	**17H71A0164**

Under the esteemed guidance of

Mr. AJAY KUMAR VERMA, M.Tech

Assistance Professor

DEPARTMENT OF CIVIL ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
An Autonomous Institution

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
An Autonomous Institution

CERTIFICATE

This is to certify that the project report entitled **"NON-LINEAR DYNAMIC ANALYSIS OF T-SHAPED IRREGULAR RC BUILDING " Bachelor of Technology in Civil Engineering** from **Jawaharlal Nehru Technological University, Kakinada** is a bonafide work carried out by **A. SAI TRIVENI (17H71A0134), MD.SHAUK (17H71A0175), G.SAI VINAY(17H71A0173), B. TEJA HARINI (17H71A0179), P.PRASAD(17H71A0164).** Under my guidance during the academic year 2020-2021 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. AJAY KUMAR VERMA, M.Tech
Assistant professor
Project Guide

Mr. N.V. Subba Rao, M.E (Ph.D.)
Associate Professor
Head of the Department

Dr. K. Srinivas, Ph.D.
Professor& Principal

External Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY
KAKINADA

DESIGN AND ANALYSIS OF NONUNIFORM SHOPPING COMPLEX

A project report submitted to the
Jawaharlal Nehru Technological University Kakinada
In partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY

In
CIVIL ENGINEERING

By

CH Venu (17H71A0142)
D Sai Suresh (17H71A0120)
M Vardhan (18H75A0124)
M Suresh (17H71A0139)
K Sriram (17H71A0177)

Under the Guidance of
Mr. S.V. TEJAS KUMAR RAJU, MTech
Assistant professor
Department of Civil Engineering

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
Accredited by NBA: EEE/ME/CSE/ECE & NAAC with 'A' Grade
Kanchikacherla - 521180, Krishna Dist. A.P.

2020-21

CERTIFICATE

This is to certify that the project report entitled "**DESIGN AND ANALYSIS OF NONUNIFORM SHOPPING COMPLEX**" being submitted for the partial fulfilment of the requirements for the award of the degree, **Bachelor of Technology in Civil Engineering from JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, KAKINADA,** is a bonafide work done by **Mr. CH VENU (17H71A01), Mr. D SAI SURESH (17H71A0120), Mr. M VARDHAN (18H75A0124), Mr. M SURESH (17H71A0139), Mr. K SRIRAM (17H71A0177)** under my guidance during the academic year 2020-21 and it has been suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. S.V. TEJAS KUMAR RAJU, MTech	**Mr. N.V. SUBBA RAO, M.E, (PhD)**
Assistant professor	**Assistant professor**
Project Guide	**Head of the Department**
Dr. K. SRINIVAS, PhD	**External Examiner Principal**

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

REMEDIATION OF EXPANSIVE SOILS BY USING BAGASSE ASH

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

In partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

CIVIL ENGINEERING

By

G. JYOTHI SREE (17H71A0154) **N.S. SANTJOSH KUMAR(17H71A0138)**

S. BABY SUSHMA (17H71A0147) **P. VINAY KIRAN (17H71A0144)**

B. VIVEK (17H71A0145)

Under the esteemed guidance of

Mr. K. PRASAD, M.Tech.

Assistant Professor

DEPARTMENT OF CIVIL ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
An Autonomous Institution

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
An Autonomous Institution

CERTIFICATE

This is to certify that the project report entitled **"REMEDIATION OF EXPANSIVE SOILS USING BAGASSE ASH" Bachelor of Technology in Civil Engineering** from **Jawaharlal Nehru Technological University, Kakinada** is a bonafidework done by**JYOTHISREE.G(17H71A0154), N.S.SANTHOSHKUMAR(17H71A0138),BABYSUSHMA.S (17H71A0147), VINAYKIRAN.P(17H71A0144), VIVEK.B(17H71A0145)**under myguidance during the academic year 2020-2021 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. K. PRASAD, M.Tech
Assistant professor
Project Guide

Mr. N.V. Subba Rao, M.E (Ph.D.)
Associate Professor
Head of the Department

Dr. K. Srinivas, Ph.D.
Professor & Principal

External Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

ANALYSIS AND DESIGN OF INDUSTRIAL SHED

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

In partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

CIVIL ENGINEERING

By

P.NAVEEN(17H71A0160) **M.SRINIVASA RAO (18H75A0123)**

K.SAI KUMAR (17H71A0169) **SRI HARSHA GUPTA CH.N.V (17H71A0137)**

M.BHARGAV (18H75A0104)

Under the esteemed guidance of

Mr. A.V.V. SAI RAM

Assistant Professor

DEPARTMENT OF CIVIL ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
An Autonomous Institution

CERTIFICATE

This is to certify that the project report entitled **"ANALYSIS AND DESIGN OF INDUSTRIAL SHED" Bachelor of Technology in Civil Engineering** from **Jawaharlal Nehru Technological University, Kakinada** is a bonafide work done by **P.NAVEEN (17H71A0160**) **M.SRINIVASA RAO (18H75A0123)**

K.SAI KUMAR(17H71A0169)SRI HARSHA GUPTA CH.N.V (17H71A0137)M.BHARGAV(18H75A0104).under my guidance during the academic year 2020-2021 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

External Examiner

Mr. A.V.V.SAI RAM,M.Tech
Assistant professor
Project Guide

Mr. N.V. Subba Rao, M.E (Ph.D.)
Associate Professor
Head of the Department

Dr. K. Srinivas, Ph.D.
Professor & Principal

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY
KAKINADA

DESIGN OF MULTI STOREY GREEN BUILDING BY USING SVA GRIHA RECOMENDATIONS

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

In partial fulfillment for the award of the Degree of

BACHELOR OF TECHNOLOGY

IN

CIVIL ENGINEERING

By

PRAGNA.T	**(17H71A0163)**	**NIKITHA.K**	**(17H71A0127)**
SIVA.K	**(18H75A0121)**	**PUJITH KUMAR.D**	**(17H71A0129)**

Under the esteemed guidance of

Mr. AAMEER SOHIEL, M.Tech

Assistant Professor

Department of Civil Engineering

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
Accredited by NBA: EEE/MECS/ECE & NAAC with 'A' Grade
Kanchikacherla - 521180, Krishna Dist. A.P.

2020-2021

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
Accredited by NBA: EEE/MECS/ECE & NAAC with 'A' Grade
Kanchikacherla - 521180, Krishna Dist. A.P.

CERTIFICATE

This is to certify that the project report entitled " **STUDY AND DESIGN OF MULTI STOREY GREEN BUILDING BY USING SVA GRIHA RECOMMENDATIONS"** being submitted for the partial fulfillment of the requirements for the award of the degree in **Bachelor of Technology in Civil Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a Bonafide work done by **PRAGNA.T , NIKITHA.K , SIVA.K , & PUJITHKUMAR.D** with **Regd Nos: 17H71A0163, 17H71A0127, 18H75A0121, 17H71A0129** under my guidance during the academic year **2020-2021** and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. AAMEER SOHIEL
Assistant Professor
&
Project Guide

Mr. N. V. SUBBA RAO, M.E (PhD)
Associate Professor
&
Head of the Department

Dr. K. SRINIVAS, Ph.D.
Professor & Principal

EXTERNAL EXAMINER

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

ANALYSIS AND DESIGN FOR OF MULTI STOREY FRAMED STRUCTURE AT VARIOUS SESMIC ZONES USING STAAD PRO

A Project report submitted to the

Jawaharlal Nehru Technological University Kakinada

In partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

In

CIVIL ENGINEERING

By

B.TARAKA RAMA RAO	**(17H71A0178)**	**E.VAMSI**	**(17H71A0180)**
L.LOHITH	**(17H71A0156)**	**A.LAKSHMI NARAYANA**	**(17H71A0119)**

Under the Esteemed Guidance of
Ms P.L.K.SOWJANYA, M.Tech
Assistant Professor
Department of Civil Engineering

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Autonomous)
Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada
Accredited with NAAC 'A' Grade, NBA Accredited B.Tech(CSE, ECE, EEE, ME), An ISO 9001:2008 Certified Institute
Kanchikacherla - 521180, Krishna Dist., A.P, India. Phone No.: 08678-273533, 273623

2020-2021

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Autonomous)
Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada
Accredited with NAAC 'A' Grade, NBA Accredited B.Tech(CSE, ECE, EEE, ME), An ISO 9001:2008 Certified Institute
Kanchikacherla - 521180, Krishna Dist., A.P, India. Phone No.: 08678-273533, 273623

CERTIFICATE

This is to certify that the project report entitled "**ANALYSIS AND DESIGN FOR OF MULTI STOREY FRAMED STRUCTURE AT VARIOUS SESMIC ZONES USING STAAD PRO** " being submitted for the partial fulfillment of the requirements for the award of the degree in **Bachelor of Technology in Civil Engineering** from **JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, KAKINADA**,is a Bonafide work done by **B.TARAKA RAMA RAO (17H71A0178) , E.VAMSI (17H71A0180) L.LOHITH (17H71A0156) ,A.LAKSHMI NARAYANA (17H71A0119)** under my guidance during the academic year 2020-21 and it has been suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Ms. P.L.K.SOWJANYA M.Tech	**Mr. N.V.SUBBARAO** ME,(Ph.D.)
Assistant Professor	Associate Professor
Project Guide	**Head Of The Department**

Dr K. SRINIVAS Ph.D. — **External Examiner**

Principal

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

A CUMULATIVE STUDY ON TRAFFIC VOLUME & ROAD ACCIDENTS BETWEEN PARITALA & KANCHIKACHERLA ON NH-65

Submitted to

Jawaharlal Nehru Technological University, Kakinada

In Partial Fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

CIVIL ENGINEERING

BY

SRI GANESH MANNE (17H71A0190) **AJATHA REDDY TENALI (17H71A0101)**

ABDUL REHAMAN SAIYED (17H75A0101) **RACHANA KAMSANI (17H71A0166)**

RAJA SEKHAR CHINNAM (17H71A0167)

Under the Esteemed Guidance of

Sri S Ashok Kumar

M. Tech,(Ph D), MIE,MISTE,MASCE

Associate Professor

DEPARTMENT OF CIVIL ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
An Autonomous Institution

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
An Autonomous Institution

CERTIFICATE

This is to certify that the project report entitled **"A CUMULATIVE STUDY ON TRAFFIC VOLUME AND ROAD ACCIDENTS BETWEEN PARITALA & KANCHIKACHERLA ON NH-65"** submitted for the degree of **Bachelor of Technology in Civil Engineering** from **Jawaharlal Nehru Technological University, Kakinada** is a bonafide work done by **SRI GANESH MANNE (17H71A0190), AJATHA REDDY TENALI (17H71A0101), ABDUL REHAMAN SAIYED (17H75A0101), RACHANA KAMSANI (17H71A0166) and RAJA SEKHAR CHINNAM (17H71A0167)** under my guidance during the academic year 2020-2021 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other Institution or University for the award of any degree.

Sri S. Ashok Kumar	Mr.N.V. SubbaRao
M. Tech,(Ph D), MIE,MISTE,MASCE	M.E (Ph.D)
Project Guide	**Head Of Department**

Dr. K. Srinivas, Ph.D

Principal **External Examiner**

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

PLANNING, DESIGNING AND ESTIMATION OF SEWERAGE SYSTEM FOR LAYOUT OF FERRY VILLAGE

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

In partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

CIVIL ENGINEERING

Submitted by

CHANDRIKA.K	**18H75A0105**
SRAVANTHI.N	**17H71A0151**
NETHAJI GANESH.B	**18H75A0114**
SAGAR.G	**18H75A0118**
AJAY KUMAR.V	**17H71A0102**

Under the esteemed guidance of

Mr. N. V. SUBBARAO, M.E Ph.D.

Head of the Department

DEPARTMENT OF CIVIL ENGINEERING

CERTIFICATE

This is to certify that the project report entitled **"PLANNING, DESIGNING AND ESTIMATION OF SEWERAGE SYSTEM FOR LAYOUT OF FERRY VILLAGE"** is a bonafide work carried out by **Chandrika K(18H75A0105), Sravanthi N(17H71A0151), Nethaji ganesh B(18H75A0114), Sagar G(18H75A0118),Ajay kumar V(17H71A0102)** in partial fulfillement for the award of degree of Bachelor of Technology in **Civil Engineering** from **Jawaharlal Nehru Technological University Kakinada** during the academic year 2020-2021 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. N.V. Subba Rao, M.E Ph.D.
Associate professor
Project Guide

Mr. N.V. Subba Rao, M.E Ph.D.
Associate Professor
Head of the Department

Dr. K. Srinivas, Ph.D.
Professor & Principal

External Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

ANALYSIS AND DESIGN OF G+7 RESIDENTIAL BUILDING BY USING STAAD PRO

A Project report submitted to the

Jawaharlal Nehru Technological University Kakinada

In partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

CIVIL ENGINEERING

By

P. LOHITHA (18H75A0111)

J. NAGA SAI (17H71A0126)

R. BHANU PRASAD (17H71A0149)

E. PRAVEEN KUMAR (17H71A0165)

P.SAI PAVAN (17H71A0171)

Under The Esteemed Guidance Of

Sri. S. ASHOK KUMAR
M.TECH, (Ph D), MIE, MISTE, MASCE
Associate Professor
Department of Civil Engineering

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
Accredited by NBA, EEE,MECS,ECE & NAAC with 'A' Grade
Kanchikacherla - 521180, Krishna Dist. A.P.

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
Accredited by NBA, EEE,MECS,ECE & NAAC with 'A' Grade
Kanchikacherla - 521180, Krishna Dist. A.P.

CERTIFICATE

This is to certify that the project report entitled "**ANALYSIS AND DESIGN OF A G+7 RESIDENTAL BUILDING BY USING STAAD PRO"** being submitted for the partial fulfillment of the requirements for the degree in Bachelor of Technology in Civil Engineering from **JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, KAKINADA,** is a Bonafede work done **by P. LOHITHA, J. NAGA SAI, R. BHANU PRASAD, E. PRAVEEN KUMAR, P.SAI PAVAN** with Registered Nos: 18H75A0111, 17H71A0126, 17H71A0149, 17H71A0165, 17H71A0171 under my guidance during the academic year 2020-2021 and it has been found suitable for acceptance according to the Requirement of the university. This work is not submitted to any other Institution or university for the award of any degree.

SRI S. ASHOK KUMAR
M. TECH (Ph D), MIE, MISTE, MASCE
Associate Professor
Project Guide

Mr. N. V. SUBBA RAO, M.E (Ph. D)
Associate Professor

Head of the department

(Dr. K. SRINIVAS)
Principal

EXTERNAL EXAMINER

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY
KAKINADA

ANALYSIS AND DESIGN OF RESIDENTIAL BUILDING BY USING ETABS

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

In partial fulfillment for the award of the Degree of

BACHELOR OF TECHNOLOGY

IN

CIVIL ENGINEERING

By

B. MOSHE	**(18H75A0113)**
N. MOHAN KRISHNA	**(18H75A0112)**
T. JAYA SAINAADH	**(18H75A0109)**
R.SAI KUMAR	**(17H71A0170)**
D.GANESH	**(17H71A0111)**

Under the esteemed guidance of

SK.AYESHA, M.Tech

Assistant Professor

Department Of Civil Engineering

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
An Autonomous Institution

2020-2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
An Autonomous Institution

CERTIFICATE

This is to certify that the project report entitled "**ANALYSIS AND DESIGN OF RESIDENTIAL BUILDING BY USING ETABS**" being submitted for the partial fulfillment of the requirements for the award of the degree in **Bachelor of Technology in Civil Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a Bonafide work done by **B.MOSHE, N.MOHAN KRISHNA, T.JAYA SAINAADH & R.SAI KUMAR, D.GANESH** with Regd Nos: 18H75A0113, 18H75A0112, 18H75A0109, 17H71A0170, 17H71A0111 under my guidance during the academic year 2020-2021 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

SK.AYESHA M.Tech	**Mr. N. V. SUBBA RAO,** M.E (PhD)
Assistant Professor	Associate Professor
Project Guide	**Head of the Department**

(Dr. K. Srinivas)

Principal

EXTERNAL EXAMINER

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited B.Tech (CSE|ECE|ME)
Accredited by NAAC With Grade A

DEPARTMENT:EEE **F-031**

ACADEMIC YEAR: 2020-21 **CLASS/ SEMESTER: IV/II**

Batch No	Roll No	CandidateName	Project Title	Internal Guide
1	17H71A0239	A.Ramya Sai	Distribution TransformerMonitoring System UsingInternet of Things (IoT)	Dr.P.Pradeep
	17H71A0266	Ch. Vinod SaiKumar		
	17H71A0251	R. Sri Gayathri		
	18H75A0246	K. Vijay Balaji		
2	17H71A0249	Sravya G	Iot Based Robotic Arm	Mr. A.V. RaviKumar
	17H71A0217	HrudaiVikas G		
	17H71A0216	Hemanth Ganesh Ch		
	17H71A0219	Jayavanth P		
3	17H71A0248	SaiVenkataDurgaPhaniMadhav M	A Wide Output Voltage Resonant EV Charger WithPower Factor Correction	Mr. M. Nageswara Rao
	18H75A0254	Sandeep Kumar P		
	18H75A0238	Sowmya B		
	18H75A0216	Guru SaiCh		
	18H75A0208	DundiSai Kishore Ch		
4	18H75A0214	GouseMohiddeenSk	Developing an islandingarrangement for grid on sensing variation of voltage and frequency	Mr. N. RajeshBabu
	18H75A0203	Ajay M		
	17H71A0231	NayeemAhammadSk		
	17H75A0219	PremchandCh		
5	17H71A0253	Srinavya K	Minimization Of CollisionAnd Alcohol Detection In Automobiles	Mr. M. SunilKumar
	17H71A0243	Sai Kumar K		
	17H71A0233	Prasanth Reddy K		
	15H71A0277	Zain Ahmed		
6	18H75A0245	Venkatapadmavathi K	Power Quality Improvement Using Dynamic Voltage RestorerIn Electrical Distribution System	Mr. K.N. Swamy
	17H71A0225	Lakshmi Siva Naga Gayathri I		
	17H71A0245	Sai Kumar Y		
	17H71A0230	Na Srinadh K		

7	17H71A0252	Sri Ramani M	Optimal Power Flow ByUsing Bald Eagle SearchOptimization Algorithm	Mr. M. RamaMohana Rao
	18H75A0211	Ganesh K		
	17H71A0228	Madhu Sri Harsha R		
	18H75A0231	RajkamalNaik D		
8	17H71A0211	Chinmai P	Power Quality Improvement by active shunt filter with hysteresiscurrent controller	Mr. N. VenuBabu
	17H71A0260	Vanditha V		
	17H71A0208	Blessy M		
	18H75A0242	Vamsi G		
9	17H71A0234	Pravallika B	Simulation and implementation of sevenlevel Inverter based Bldcmotor (60)	Mrs. P. DurgaBhavani
	18H75A0241	SyamSundarM		
	17H71A0201	Adithya T		
	17H71A0236	Ram Chaitanya P		
10	17H71A0255	Susmitha M	Design And Operation OfSmart Energy Meter For Effective Energy Utilization In Smart Cities	Mr.S.Ravikanth
	18H75A0215	GowthamAshok O		
	18H75A0228	Navyasri V		
	18H75A0220	Komal NagaTeja A		
11	17H71A0246	Sai Sri R	Dual solar trackingsystem using microcontroller	Mr. M. Tharun
	18H75A0205	Ashok KumarN		
	18H75A0212	Gopi L		
	18H75A0223	Lakshmi DasM		
12	18H75A0226	Monica Yadav D	Minimizing Penalty In Industrial Power Consumption By EngagingApfc Unit	Mr. T. Jayaraju
	17H71A0222	KhajaMoienSk		
	18H75A0249	Chalapathi Rao G		
	17H71A0224	Lahari P		
13	17H71A0221	KarimullaSk	Analysis of Three-Switch-Based Integrated Dual-DCBoost Converter Topology	Mr. P. NagaGangadhar
	17H71A0254	SuneethaCh		
	18H75A0244	Venkata Sunil Kumar G		
	18H75A0206	BalaAbhishekK		
14	17H71A0247	SaiTejaswini M	Modelling and voltagecontrol of solar wind hybrid microgrid withoptmised statcom.	Mr. A. VamsiKrishna
	17H71A0227	Madhan Mohan G		
	17H71A0209	Brahma Reddy P		

	17H71A0261	VasaviCh		
15	18H75A0210	Ganesh K	STATCOM for Improved Dynamic Performance of Wind Farms in Power Grid	Dr.P.Pradeep
	17H71A0214	Gopinadh D		
	17H71A0256	Tejaswini P		
	18H75A0255	LeelaKrishna. U		
16	18H75A0248	Aravind N	Wireless Power TransferSystem For Moving Electric Vehicles Using Magnetic Resonance	Mr. A.V. RaviKumar
	17H71A0264	Venkateswara Rao M		
	18H75A0243	VenkataAdises hu Kumar S		
	18H75A0207	Divya A		
17	18H75A0237	Sindhu K	A Canonical Switching Cell(Csc) Converter BasedPower Factor Corrected Battery Charger For E Rickshaw	Mr. M. Nageswara Rao
	18H75A0240	Swathi Y		
	17H71A0229	N V UdayaSriCh		
	17H71A0265	Vinod KumarB		
18	18H75A0225	Mohan D	Hybrid power generation	Mr.N.RajeshBabu
	18H75A0217	Indrani K		
	17H71A0267	Vishnu Vardhan P		
	17H71A0213	Gopi KrishnaA		
19	17H71A0241	ReshmaSk	Wireless power transmission using solar energy Wireless power transmission using solarenergy	Mr. M. SunilKumar
	18H75A0204	Aruna K		
	17H71A0203	Ankit Raj		
	17H71A0223	L SeshasaiDurgaMani Sujith N		
20	17H71A0215	Haswanth J	Autonomous robot for GPS controlled environment monitoring and military purpose Basedon iot and arm (40)	Mr.K.N.Swamy
	18H75A0234	SaiPavankuma r Y		
	17H71A0250	Sreeja L		
	18H75A0219	Kalyan A		
21	18H75A0209	Durga Naga Pavan Kishore G	Optimal Reactive PowerDispatch Using Lévy Flight Distribution Algorithm	Mr. M. RamaMohana Rao
	18H75A0251	Prabhakara Rao M		
	17H71A0210	Charan Rajesh P		
	18H75A0221	Koteswara Rao S		
22	17H71A0232	NazeerHussain Sk	Protection Of InductionMotor From Over	Mr. N.

	18H75A0247	Vinod Kumar L	Voltage And Under Voltage. (50)	VenuBabu
	17H71A0240	Rathna B		
	17H71A0237	Ramya A		
23	17H71A0238	Ramya K	Monitoring of PV performance using aurdino	Mrs. P. DurgaBhavani
	18H75A0222	Krishnaveni K		
	18H75A0230	Prasad J		
	18H75A0232	RevanthSai Y		
24	18H75A0201	Abdul Raheem Sk	Automatic Phase Selectorwith IOT Notifier.	Mr.S.Ravikanth
	18H75A0218	Jaswanth J		
	18H75A0224	Madhu G		
	18H75A0213	Gopinadh P		
25	17H71A0262	Venkata Naga Rakesh N	A UPFC for Voltage Regulation in LV Distribution Feeders with aDC-Link Ripple Voltage Suppression Technique. (40)	Mr. M. Tharun
	18H75A0236	SethuSai D		
	17H71A0258	V S L N SaiAbhishek G		
	18H75A0235	SankarSatish G		
26	17H71A0220	Jessymole K	Bidirectional Rotation Of	Mr. T. Jayaraju
	18H75A0250	DurgaMalleswari M	An Induction Motor WithA Remote Control Device	
	17H71A0202	Anil G		
	17H71A0205	Ashok A		
27	18H75A0227	Mounika G	Design and development ofmultilevel inverter with reduced switches for electric vehicles	Dr.B.K.Karunakar Rao
	17H71A0207	BalaSaiKiran T		
	18H75A0229	Phirod KumarP		
	17H71A0204	Arun Kumar K		
28	17H71A0257	V RohithKumar A	Design and control standalone microgrid	Mr .K. Naresh
	18H75A0252	Rama Krishna A		
	17H71A0226	Lavanya T		
	17H71A0235	Rajesh B		

Project-Coordinator **Head of the Department**

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

IoT BASED DISTRIBUTION TRANSFORMER MONITORING SYSTEM

A Project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
ELECTRICAL AND ELECTRONICS ENGINEERING
By

Akula Ramya Sai	**17H71A0239**
Kantu Vijay Balaji	**18H75A0246**
Chilaka Vinod Sai Kumar	**17H71A0266**
Rachuri sri Gayathri	**17H71A0251**

Under the esteemed guidance of
Dr. P. Pradeep, M.Tech,. PhD
Associate Professor
Head of the Department
Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India

2020-2021

Devineni Venkata Ramana & Dr. Himasekhar

MIC College of Technology

An Autonomous Institution

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada

Kanchikacherla - 521180, Krishna Dist, A.P, India.

CERTIFICATE

This is to certify that the project report entitled **"IoT based Distribution Transformer Monitoring System"** being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a bonafide work done by **A.Ramya Sai, K.Vijay Balaji, Ch.Vinod Sai Kumar, R.Sri Gayathri,** under my guidance during the academic year 2020-2021 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Dr.P. Pradeep, M.Tech.,Ph.D
Associate Professor
Project Guide

Dr.P. Pradeep, M.Tech.,Ph.D
Associate Professor
Head of the Department

Dr.K.Srinivas, Ph.D, JNTUH
Principal

Examinar

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

IOT BASED ROBOTIC ARM

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

by

G. SRAVYA	(17H71A0249)
G. HRUDAI VIKAS	(17H71A0217)
CH. HEMANTH GANESH	(17H71A0216)
P. JAYAVANTH	(17H71A0219)

Under the esteemed guidance of

Mr.A.V. Ravi Kumar M.Tech (Ph.D)

Assistant Professor

Department of EEE

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)

Kanchikacherla – 521180, Krishna Dist., AP, India

Phone: 08678-273535, Fax: 08678-273569

e-mail: dvrhmic@mictech.ac.in, website: www.mictech.ac.in

NBA-Accredited

NAAC A With Grade

2020-21

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist., A.P. India.

CERTIFICATE

This is to certify that the seminar report entitled **"IOT BASED ROBOTIC ARM"** being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a bonafide work done by **G. Sravya, G. Hrudai Vikas, CH. Hemanth Ganesh, P. Jayavanth** under my guidance during the academic year 2020-21 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. A. Vamsi Krishna, M.Tech .
Assistant Professor
Project Guide

Dr.P.Pradeep, Ph.D
Associate Professor
Head of the Department

Dr. K.Srinivas, Ph.D,
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

A WIDE OUTPUT VOLTAGE RESONANT EV CHARGER WITH POWER FACTOR CORRECTION

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

In

ELECTRICAL AND ELECTRONICS ENGINEERING

By

M.S.V.D. Phani Madhav	17H71A0248
P. Sandeep Kumar	18H75A0254
B. Sowmya	18H75A0238
Ch. Guru Sai	18H75A0216
Ch. Dundi Sai Kishore	18H75A0208

Under the esteemed guidance of

Mr. M. Nageswara Rao, M. Tech., (Ph. D)

Assistant Professor

Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar

MIC College of Technology

An Autonomous Institution

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada

Kanchikacherla - 521180, Krishna Dist, A.P, India

2020-21

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist., A.P. India

CERTIFICATE

This is to certify that the project report entitled **"A WIDE OUTPUT VOLTAGE RESONANT EV CHARGER WITH POWER FACTOR CORRECTION"** being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada**, is a bonafide work done by **M S V D Phani Madhav, P Sandeep Kumar, B Sowmya, Ch Guru Sai, Ch Dundi Sai Kishore** under my guidance during the academic year 2020-21 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. M. Nageswara Rao, MTech., (Ph. D)
Assistant Professor
Project Guide

Dr. P. Pradeep, MTech., Ph. D
Professor
Head of the Department

Dr. K. Srinivas, MTech., Ph. D
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

DEVELOPING AN ISLANDING ARRANGEMENT FOR GRID ON SENSING VARIATION OF VOLTAGE & VARIATION OF FREQUENCY

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRICAL AND ELECTRONICS ENGINEERING

by

Sk. Gouse Mohiddeen (18H75A0214)

Sk. Nayeem Ahammad (17H71A0231)

Ch. Premchand (18H75A0219)

M. Ajay (18H75A0203)

Under the guidance of

Mr. N. Rajesh Babu M. Tech., MBA., (Ph. D)

Assistant Professor

Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar

MIC College of Technology

An Autonomous Institution

Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada

Kanchikacherla - 521180, Krishna Dist., A.P, India.

2020-21

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist., A.P, India.

CERTIFICATE

This is to certify that the project report entitled **"Developing An Islanding Arrangement For Grid On Sensing Variation Voltage And Variation Frequency"** being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a bona fide work done by **Sk. Gouse Mohiddeen, Sk. Nayeem Ahammad, Ch. Premchand, M. Ajay** under my guidance during the academic year 2020-21 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. N. Rajesh Babu M. Tech., MBA., (Ph. D)
Assistant Professor
Project Guide

Dr. P. Pradeep, M. Tech., Ph. D
Associate Professor
Head of the Department

Dr K. Srinivas M. Tech., Ph. D
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

MINIMISATION OF COLLISION BY ELECTRO MAGNETIC REPULSION BRAKING AND ALCOHOL DETECTION IN AUTOMOBILES

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRICAL AND ELECTRONICS ENGINEERING
by

SRI NAVYA KILARU	(17H71A0253)
SAI KUMAR K	(17H71A0243)
PRASANTH REDDY K	(17H71A0233)
ZAIN AHMED	(15H71A0277)

Under the esteemed guidance of
Mr. M. SUNIL KUMAR, M.TECH (PhD)
Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

CERTIFICATE

This is to certify that the project report entitled **"MINIMISATION OF COLLISION BY ELECTRO MAGNETIC REPULSION AND ALCOHOL DETECTION IN AUTOMOBILES"** being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a bonafide work done by **Sri Navya K, Sai Kumar K, Prasanth Reddy K and Zain Ahmed** under my guidance during the academic year 2020-21 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. M. Sunil Kumar, M.Tech.
Assistant Professor
Project Guide

Dr. P. Pradeep, Ph.D
Associate Professor
Head of the Department

Dr. K.Srinivas, Ph.D,
Principal

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

POWER QUALITY IMPROVEMENT USING DYNAMIC VOLTAGE RESTORER IN ELECTRICAL DISTRIBUTION SYSTEM

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

by

K.VENKTAPADMAVATHI(18H75A0245)

I.LASHMI SIVA NAGA GAYATHRI(17H71A0225)

Y.SAI KUMAR(17H71A0245)

K.SRINADH(17H71A0230)

Under the esteemed guidance of

Mr.K.N.Swamy, M.Tech,(Ph.D)

Assistant Professor

Department of EEE

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)

Kanchikacherla – 521180, Krishna Dist., AP, India

Phone: 08678-273535, Fax: 08678-273569

e-mail: dvrhsmic@mictech.ac.in, website: www.mictech.ac.in

NBA-Accredited

NAAC A

2020-21

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

CERTIFICATE

This is to certify that the seminar report entitled **"Design and Operation of Smat Energy Meter for Effective Energy Utilization in Smart Cities"** being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a bonafide work done by **K.Venkta Padmavathi, I.Lashmi Siva Naga Gayathri, Y. Sai Kumar, K.Srinadh** under my guidance during the academic year 2020-21 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. K.N.Swamy, M.Tech,(Ph.D).
Assistant Professor
Project Guide

Dr.P.Pradeep, Ph.D
Associate Professor
Head of the Department

Dr. K.Srinivas, Ph.D,
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

OPTIMAL POWER FLOW USING BALD EAGLE SEARCH OPTIMIZATION ALGORITHM

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRICAL AND ELECTRONICS ENGINEERING

by

SRI RAMANI M (17H71A0252)
GANESH K (18H75A0211)
MADHU SRI HARSHA (17H71A0228)
RAJ KAMAL NAIK D (18H75A0231)

Under the esteemed guidance of

Mr. M.RAMA MOHANA RAO, M.Tech
Assistant Professor
Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

2020-2021

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

CERTIFICATE

This is to certify that the project report entitled **"Optimal Power Dispatch using Bald Eagle Search Optimization Algorithm"** being submitted for the partial fulfilment of the requirements for the award of **the degree in Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a bonafide work done by **Sri Ramani M, Ganesh K, Madhu Sri Harsha R and Raj Kamal Naik D** under my guidance during the academic **year 2020-21** and it has been found **suitable for acceptance** according to the **requirement of the university**. This work is not submitted to any other institution or university for the award of any degree.

Mr. M.Rama Mohana Rao, M.Tech.
Assistant Professor
Project Guide

Dr.P.Pradeep, Ph.D
Professor
Head of the Department

Dr. K.Srinivas, Ph.D.
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

"POWER QUALITY IMPROVEMENT BY ACTIVE SHUNT FILTER WITH HYSTERESIS CURRENT CONTROLLER"

A Project report submitted to the

Jawaharlal Nehru Technological University, Kakinada

In partial fulfillment for the award of the degree

BACHELOR OF TECHNOLOGY

In

ELECTRICAL AND ELECTRONICS ENGINEERING

by

CHINMAI P (17H71A0211)

VANDITHA V (17H71A0260)

BLESSY M (17H71A0208)

VAMSI G (18H75A0242)

Under the esteemed guidance of,

Mr. N. VENU BABU, M. Tech, Assistant professor, Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar

MIC College of Technology

An Autonomous Institution

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada

Kanchikacherla - 521180, Krishna Dist, A.P, India.

2020-2021

Devineni Venkata Ramana & Dr. Himasekhar

MIC College of Technology

An Autonomous Institution

Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada

Kanchikacherla - 521180, Krishna Dist., A.P, India.

CERTIFICATE

This is to certify that the project report entitled **"POWER QUALITY IMPROVEMENT BY ACTIVE SHUNT FILTER WITH HYSTERESIS CURRENT CONTROLLER"** being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada** is a bona fide work done by **CHINMAI P, VANDITHA V, BLESSY M, VAMSI G,** during the academic year **2020-2021** and it has been found suitable for acceptance according to the requirement of the university.

Mr. N. VENU BABU, M. Tech
Assistant Professor
Project Guide

DR. P. PRADEEP, M. Tech., Ph. D
Associate Professor
Head of the Department

DR. K. SRINIVAS, PH. D
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

SIMULATION AND IMPLEMENTATION OF SEVEN-LEVEL INVERTER BASED BLDC MOTOR

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

In partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
In

ELECTRICAL & ELECTRONICS ENGINEERING

By

B.PRAVALLIKA (17H71A0234)

T.ADITYA (17H71A0201)

P.RAMCHAITANYA (17H71A0236)

M.SHYAM SUNDHAR(18H75A0241)

Under the esteemed guidance of

Mrs.P.DURGABHAVANI,M.Tech.,

Assistant Professor

Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar

MIC College of Technology

An Autonomous Institution

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada

Kanchikacherla - 521180, Krishna Dist, A.P, India.

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist., A.P, India.

2020-2021

CERTIFICATE

This is to certify that the project report entitled **"Simulation & Implementation of seven level inverter based BLDC motor"** being submitted for the partial fulfillment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University,Kakinada**,is a bonafide work done by **B.Pravallika, T.Adithya, P.Ram chaitanya , M.Shyam sundhar** under my guidance during the academic year 2020-21 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

P. Durga Bhavani

Mrs.P.Durga Bhavani,M.tech.,
Assistant Professor
Project Guide

Dr.P.Pradeep,Ph.D
Professor
Head of the Department

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

Design and Operation of Smart Energy Meter for Effective Energy Utilization in Smart Cities

A Project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRICAL AND ELECTRONICS ENGINEERING

by

Susmitha M	17H71A0255
Gowtham Ashok O	18H75A0215
Navyasri V	18H75A0228
Komal Naga Teja A	18H75A0220

Under the esteemed guidance of
Mr. Mr.S.Ravikanth, M. Tech.,
Assistant Professor
Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

2020-21

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

CERTIFICATE

This is to certify that the seminar report entitled **"Design and Operation of Smat Energy Meter for Effective Energy Utilization in Smart Cities"** being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a bonafide work done by **Susmitha M, Gowtham Ashok O, Navyasri V, Komal Naga Teja A** under my guidance during the academic year 2020-21 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. S.Ravikanth, M.Tech .
Assistant Professor
Project Guide

Dr.P.Pradeep, Ph.D
Assistant Professor
Head of the Department

Dr. K.Srinivas, Ph.D.
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

DUAL AXIS SOLAR TRACKER USING MICROCONTROLLER

A Project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

ELECTRICAL AND ELECTRONICS ENGINEERING

by

SAI SRI R	**(17H71A0246)**
ASHOK KUMAR N	**(18H75A0205)**
GOPI L	**(18H75A0212)**
LAKSHMI DAS M	**(18H75A0223)**

Under the esteemed guidance of
Mr. M.THARUN, M.Tech,
Assistant Professor
Department of EEE

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently affiliated to:JNTUK,Kakinada
kanchikacherla – 521180,Krishna Dist,A.P,India

2020-2021

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently affiliated to:JNTUK,Kakinada
kanchikacherla – 521180,Krishna Dist,A.P,India

NBA-Accredited
B.Tech (CSE | ECE | EEE | ME)
Accredited by NAAC A With Grade

CERTIFICATE

This is to certify that the project report entitled **"Dual Axis Solar Tracker Using Microcontroller"** being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a bonafide work done by **Sai Sri R, Ashok Kumar N, Gopi L and Lakshmi Das M** under my guidance during the academic year **2020-21** and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr.M.Tharun, M.Tech.
Assistant Professor
Project Guide

DR.P.Pradeep, Ph.D
Associate Professor
Head Of The Department

DR.K.Srinivas, Ph.D
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

MINIMIZING PENALTY IN INDUSTRIAL POWER CONSUMPTION BY ENGAGING APFC UNIT

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRICAL AND ELECTRONICS ENGINEERING
by

D. MONICA YADAV (18H75A0226)
SK. KHAJA MOIEN (17H71A0222)
G. CHALAPATHI RAO (18H75A0249)
P. LAHARI (17H71A0224)

Under the esteemed guidance of

Mr. T. JAYA RAJU, MTech

Assistant Professor
Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

2020-21

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

2020-2021

CERTIFICATE

This is to certify that the project report entitled **"Minimizing Penalty in Industrial Power Consumption by engaging APFC Unit"** being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a bonafide work done by **D. Monica Yadav, Sk. Khaja Moien, G. Chalapathi Rao, P.Lahari** under my guidance during the academic year 2020-21 and it has been found suitable for acceptance according to the requirement of the university.

T. JAYA RAJU

Mr. T. Jaya Raju, MTech.
Assistant Professor
Project Guide

Dr. P. Pradeep, PhD
Associate Professor
Head of the Department

Dr. K. Srinivas, PhD.
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

Analysis of Three-Switch Based Integrated Dual -DC Boost Converter Topology

A project report submitted to

Jawaharlal Nehru Technological University, Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

ELECTRICAL AND ELECTRONICS ENGINEERING

by

SHAIK KARIMULLA	(17H71A0221)
CHANDRA SUNEETHA	(17H71A0254)
GANGUMALLA VENKATA SUNIL KUMAR	(18H75A0244)
KOMIRISETTY BALA ABHISHEK	(18H75A0206)

Under the esteemed guidance of

Mr. PALURU NAGA GANGADHAR, M. Tech

Assistant Professor

Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar

MIC College of Technology

An Autonomous Institution

Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada

Kanchikacherla - 521180, Krishna Dist, A.P, India.

2020-2021

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

CERTIFICATE

This is to certify that the project report entitled **"Analysis of Three-Switch Based Integrated Dual - DC Boost Converter Topology"** being submitted for the partial fulfillment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada**, is a bonafide work done by **KARIMULLA SHAIK, SUNEETHA CHANDRA ,VENKATA SUNIL KUMAR GANGUMALLA,BALA ABHISHEK KOMMIRISETTY** under my guidance during the academic year 2020-2021 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. P. Naga Gangadhar, M.Tech.
Assistant Professor
Project Guide

Dr.P. Pradeep, M.Tech., Ph.D.
Associate Professor
Head of the Department

Dr.K. Srinivas Rao M.Tech., Ph.D.
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

Modelling and voltage control of solar wind hybrid microgrid with optmised statcom.

A Project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

ELECTRICAL AND ELECTRONICS ENGINEERING

By

SaiTejaswini.M	17H71A0247
Madhan Mohan G	17H71A0227
Brahma Reddy P	17H71A0209
Vasavi Ch	17H71A0261

Under the esteemed guidance of

Mr. A. VamsiKrishna, M. Tech.,

Assistant Professor

Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar

MIC College of Technology

An Autonomous Institution

Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada

Kanchikacherla - 521180, Krishna Dist, A.P, India.

2020-21

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

CERTIFICATE

This is to certify that the seminar report entitled **"Modelling and voltage control of solar wind hybrid microgrid with optmised statcom"** being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a bonafide work done by **SaiTejaswini.M, Madhan Mohan G, Brahma Reddy P, Vasavi Ch** under my guidance during the academic year 2020-21 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. A. Vamsi Krishna, M.Tech .
Assistant Professor
Project Guide

Dr.P.Pradeep, Ph.D
Assistant Professor
Head of the Department

Dr. K.Srinivas, Ph.D.
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

STATCOM for Improved Dynamic Performance of Wind Farms in Power Grid

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
ELECTRICAL AND ELECTRONICS ENGINEERING
by

GANESH K (18H75A0210)
GOPINADH D (17H71A0214)
TEJASWINI P (17H71A0256)
LEELA KRISHNA U (18H75A0255)

Under the esteemed guidance of

Dr. P. PRADEEP M.Tech., Ph. D

Associate Professor
Department of EEE

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla – 521180, Krishna Dist., AP, India
Phone: 08678-273535, Fax: 08678-273569
e-mail: dvrhsmic@mictech.ac.in, website: www.mictech.ac.in

NBA-Accredited
NAAC A With Grade

2020-21

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla – 521180, Krishna Dist., AP, India
Phone: 08678-273535, Fax: 08678-273569
e-mail: dvrhsmicet@mictech.ac.in, website: www.mictech.ac.in

NBA-Accredited
B.Tech: ECE | ECE | EEE | MECH
NAAC A
Grade

2020-21

CERTIFICATE

This is to certify that the project report entitled **"STATCOM for Improved Dynamic Performance of Wind Farms in Power Grid"** being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a bonafide work done by **Ganesh k, Gopinadh d, Tejasvini p and Leela krishna U** under my guidance during the academic year 2020-21 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Dr. P. Pradeep, M.Tech, Ph.D
Associate Professor
Project Guide

Dr. P.Pradeep, M.Tech, Ph.D
Associate Professor
Head of the Department

Dr. K.Srinivas, M.Tech, Ph.D.
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

Wireless Power Transfer System for Moving Electric Vehicles Using Magnetic Resonance

A Project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRICAL AND ELECTRONICS ENGINEERING

By

Aravind N	18H75A0248
Venkateswara Rao M	17H71A0264
VenkataAdiseshu Kumar S	18H75A0243
Divya A	18H75A0207

Under the esteemed guidance of
Mr. A.V. RaviKumar, M. Tech.,
Assistant Professor
Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P. India.

2020-21

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

CERTIFICATE

This is to certify that the seminar report entitled **"Wireless Power Transfer System for Moving Electric Vehicles Using Magnetic Resonance"** being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a bonafide work done by **Aravind N, Venkateswara Rao M, VenkataAdiseshu Kumar S, Divya A** under my guidance during the academic year 2020-21 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. A.V. RaviKumar, M.Tech
Assistant Professor
Project Guide

Dr.P.Pradeep, Ph.D
Assistant Professor
Head of the Department

Dr. K.Srinivas, Ph.D.
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

A CANONICAL SWITCHING CELL(CSC) CONVERTER BASED POWER FACTOR CORRECTED BATTERY CHARGER FOR E-RICKSHAW

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment the award of the degree of

BACHELOR OF TECHNOLOGY

In

ELECTRICAL AND ELECTRONICS ENGINEERING

by

K.SINDHU	(18H75A0237)
Y.SWATHI	(18H75A0240)
CH.N.V.UDAYA SRI	(17H71A0229)
B.VINOD KUMAR	(17H71A0265)

Under the esteemed guidance of

Mr.M.Nageswara Rao, M.Tech., (Ph.D)

Assistant Professor

Department of EEEE

Devineni Venkata Ramana & Dr. Himasekhar

MIC College of Technology

An Autonomous Institution

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada

Kanchikacherla - 521180, Krishna Dist, A.P, India.

CERTIFICATE

This is to certify that the project report entitled **"A Canonical switching cell (CSC) converter based power factor corrected battery charger for E-rickshaw"** being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a bonafide work done by **K.Sindhu, Y.Swathi, CH.N.V.Udaya Sri, R.Vinod kumar** under my guidance during the academic year 2020-21 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr.M.Nageswara Rao, M.tech.,(Ph.D)
Assistance Professor
Project Guide

Dr. P.Pradeep, Ph.D
Associate Professor
Head of the Department

DrK.Srinivas., M.Tech,Ph.D

Principal

Examinar

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

Hybrid Power Generation

A Project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRICAL AND ELECTRONICS ENGINEERING

By

Mohan D	18H75A0225
Indrani K	18H75A0217
Vishnu Vardhan P	17H71A0267
Gopi Krishna A	17H71A0213

Under the esteemed guidance of
Mr. N.Rajesh Babu, M. Tech.,
Assistant Professor
Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

2020-21

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK., Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P. India.

CERTIFICATE

This is to certify that the seminar report entitled **"Wireless Power Transfer System for Moving Electric Vehicles Using Magnetic Resonance"** being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a bonafide work done by **Aravind N, Venkateswara Rao M, VenkataAdiseshu Kumar S, Divya A** under my guidance during the academic year 2020-21 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. A.V. RaviKumar, M.Tech
Assistant Professor
Project Guide

Dr.P.Pradeep, Ph.D
Assistant Professor
Head of the Department

Dr. K.Srinivas, Ph.D.
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

WIRELESS POWER TRANSMISSION USING SOLAR ENERGY

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRICAL AND ELECTRONICS ENGINEERING

by

SHAIK RESHMA (17H71A0241)
NANDIWADA L.S.D.M SUJITH (17H71A0223)
KOTA ARUNA (18H75A0204)
ANKIT RAJ(17H71A0203)

Under the esteemed guidance of

Mr. M. SUNIL KUMAR, M.Tech , Ph.D

Assistant Professor

Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

2020-2021

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P. India.

CERTIFICATE

This is to certify that the project report entitled **"Wireless power transmission using Solar Energy"** being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a bonafide work done by **Shaik Reshma, N.L.S.D.M Sujith , K Aruna and Ankit Raj** under my guidance during the academic **year 2020-21 and** it has been found suitable for **acceptance** according to the requirement **of the university.** This work is not submitted to any other institution or university for the award of any degree.

Mr. M. Sunil kumar, M.Tech., Ph.D
Assistant Professor
Project Guide

Dr.P.Pradeep, Ph.D
Professor
Head of the Department

Dr. K.Srinivas, Ph.D.
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

USER CONTROLLED ROBOT FOR ENVIRONMENT MONITORING BASED ON IOT

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

ELECTRICAL AND ELECTRONICS ENGINEERING

by

J.HASWANTH	**[17H71A0215]**
Y.SAI PAVAN KUMAR	**[18H75A0234]**
L.SREEJA	**[17H71A0250]**
A.KALYAN	**[18H75A0219]**

Under the esteemed guidance of

Mr. K.N.SWAMY, M. Tech (Ph.D)

Assistant Professor

Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar

MIC College of Technology

An Autonomous Institution

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada

Kanchikacherla - 521180, Krishna Dist, A.P, India.

CERTIFICATE

This is to certify that the project report entitled " **USER CONTROLLED ROBOT FOR ENVIRONMENT MONITORING BASED ON IOT** " being submitted for the partial fulfillment of the requirements for the award of the degree in**Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a bonafide work done by **J.Haswanth, Y.Sai Pavan Kumar, L.Sreeja** and **A.Kalyan,** under my guidance during the academic year 2020-21 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

K.N.Swamy

Mr. K.N.SWAMY, M. Tech (Ph.D)
Assistant Professor
Project Guide

Dr.P. Pradeep, MTech, Ph.D.
Associate Professor
Head of the department

Dr.K. Srinivas Rao MTech, Ph.D.
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

OPTIMAL REACTIVE POWER DISPATCH USING LÉVY FLIGHT DISTRIBUTION ALGORITHM

A project report submitted to the

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITYKAKINADA

BACHELOR OF TECHNOLOGY

in

ELECTRICAL AND ELECTRONICS ENGINEERING

by

DURGA NAGA PAVAN KISHORE G (18H75A0209)

PRABHAKARA RAO M (18H75A0251)

CHARAN RAJESH P (17H71A0210)

KOTESWARA RAO S (18H75A0221)

Under the Guidance of

Mr. M. RAMAMOHANA RAO M. Tech

Assistant Professor

Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar

MIC College of Technology

An Autonomous Institution

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada

Kanchikacherla - 521180, Krishna Dist, A.P, India

2020-2021

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

CERTIFICATE

This is to certify that the project report entitled **"Optimal Reactive Power Dispatch using Levy Flight Distribution Algorithm"** being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a bonafide work done by **Durga Naga Pavan Kishore G, Prabhakar Rao M, Charan Rajesh P and Koteswaru Rao S** under my guidance during the **academic year 2020-21** and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. M. Rama Mohana Rao, MTech.
Assistant Professor
Project Guide

Dr.P. Pradeep, Ph. D
Professor
Head of the Department

Dr. K. Srinivas, Ph. D
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

Protection of induction motor from over voltage and under voltage

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

ELECTRICAL AND ELECTRONICS ENGINEERING

By

Sk.Nazeer Hussain	(17H1A0232)
Lam.Vinod Kumar	(18H75A0247)
B . Rathna	(17H71A0240)
A.Ramya	(17H71A0237)

Under the Guidance of
Mr.N. Venu Babu, M.Tech
Assistant Professor

Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

2020-2021

CERTIFICATE

This is to certify that the project report entitled **"PROTECTION OF INDUCTION MOTOR FROM OVER VOLATAGE AND UNDER VOLTAGE** - being submitted for the partial fulfillment of the requirements for the award of the degree in **BACHELOR OF TECHNOLOGY IN ELECTRICAL AND ELECTRONICS ENGINEERING** FROM **JAWAHARLAL NEHRU TECHNOLOGICAL** University, Kakinada, is a bonafide work done by **Nazeer Hussain Shaik , L. Vinod Kumar , B. Rathna** and **A.Ramya** under my guidance during the academic year 2010-2021 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. N .Venu babu , M . Tech
Assistant Professor
Project Guide

Dr. P. Pradeep, Ph. D
Associate Professor
Head of the Department

Dr. K. Srinivas, Ph. D

Principal

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

Monitoring of PV performance using aurdino

A Project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRICAL AND ELECTRONICS ENGINEERING

By

Ramya K	17H71A0238
Krishnaveni K	18H75A0222
Prasad J	18H75A0230
RevanthSai Y	18H75A0232

Under the esteemed guidance of
Mrs. P. DurgaBhavani, M. Tech.,
Assistant Professor
Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

2020-21

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P. India.

CERTIFICATE

This is to certify that the seminar report entitled **"Monitoring of PV performance using aurdino"** being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a bonafide work done by **Ramya K, Krishnaveni K, Prasad J, RevanthSai Y** under my guidance during the academic year 2020-21 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

P. Durga Bhavani

Mrs. P. DurgaBhavani, M.Tech
Assistant Professor
Project Guide

Dr. P. Pradeep, Ph.D
Assistant Professor
Head of the Department

Dr. K.Srinivas, Ph.D.
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

AUTOMATIC PHASE SELECTOR WITH POWER THEFT DETECTION

A project report submitted to the

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

in partial fulfillment for the award of

Bachelor of Technology

in

ELECTRICAL AND ELECTRONICS ENGINEERING

by

Shaik. Abdul Raheem	**(18H75A0201)**
J Jaswanth	**(18H75A0218)**
G Madhu	**(18H75A0224)**
P Gopinadh	**(18H75A0213)**

Under the esteemed guidance of

Mr S Ravikanth, M.Tech

Assistant Professor

Department of EEE

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)

Kanchikacherla – 521180, Krishna Dist., AP, India

Phone: 08678-273535, Fax: 08678-273569

e-mail: dvrhumic@mictech.ac.in, website: www.mictech.ac.in

2020-21

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla – 521180, Krishna Dist., AP, India
Phone: 08678-273535, Fax: 08678-273569
e-mail: dvrhsmic@mictech.ac.in, website: www.mictech.ac.in

NBA-Accredited
NAAC With grade A

2020-21

CERTIFICATE

This is to certify that the project report entitled **"AUTOMATIC PHASE SELECTOR POWER THEFT DETECTION"** being submitted for the partial fulfilment of the requirements for the award of **Bachelor of Technology in Electrical and Electronics Engineering** from **JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA** is a bonafide work done by **Sk Abdul Raheem (18H75A0201), J Jaswanth (18H75A0218), G Madhu (18H75A0224), P Gopinadh (18H75A0213),** under my guidance during the academic year 2020-21 and it has been found suitable for acceptance according to the requirement of the **JNTUK.** This work is not submitted to any other institution or university for the award of any degree.

Mr S.Ravikanth, M.Tech
Assistant Professor
Project Guide

Dr P.Pradeep, M.Tech, Ph.D.
Associate Professor
Head of the Department

Dr. K.Srinivas, Ph.D.
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

A UPFC FOR VOLTAGE REGULATION IN LV DISTRIBUTION FEEDER WITH A DC-LINK RIPPLE VOLTAGE SUPPRESSION TECHNIQUE

A Project report submitted to the

Jawaharlal Nehru Technological University, Kakinada

In partial fulfillment for the award of the degree

BACHELOR OF TECHNOLOGY

in

ELECTRICAL AND ELECTRONICS ENGINEERING

by

VENKATA NAGA RAKESH NALLURU	**(17H71A0262)**
SEATHU SAI D	**(18H75A0236)**
V S L N SAI ABHISHEK G	**(17H71A0258)**
SANKAR SATISH G	**(18H75A0235)**

Under the Esteemed Guidance of

Mr. M. Tharun, M. Tech

Assistant Professor

Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar

MIC College of Technology

An Autonomous Institution

Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada

Kanchikacherla - 521180, Krishna Dist, A.P. India

2020-2021

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

CERTIFICATE

This is to certify that the seminar report entitled "**A UPFC for Voltage Regulation in LV Distribution Feeders with a DC-Link Ripple Voltage Suppression Technique**" being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada** is a bonafide work done by **Venkata Naga Rakesh Nalluru, Seathu Sai D, V S L N Sai Abhishek G, Sankar Satish G** during the academic year **2020-2021** and it has been found suitable for acceptance according to the requirement of the university.

Mr. M. Tharun, M. Tech
Assistant Professor
Project Guide

Dr. P. Pradeep, M. Tech., Ph. D
Professor
Head of the Department

Dr. K. Srinivas, Ph. D
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

GSM AND ARDUINO BASED POWER THEFT DETECTION AND PROTECTION

A Project report submitted to the

Jawaharlal Nehru Technological University,Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OFTECHNOLOGY

In

ELECTRICAL AND ELECTRONICSENGINEERING

By

K.JESSYMOLE	(17H71A0220)
M.DURGAMALLESWARI	(18H75A0250)
G.ANIL	(17H71A0202)
A.ASHOK	(17H71A0205)

Under the esteemed guidance

Mr.T.JAYARAJU, M.Tech,

Assistant professor

Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar

MIC College of Technology

An Autonomous Institution

Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada

Kanchikacherla - 521180, Krishna Dist, A.P, India.

2020-2021

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

CERTIFICATE

This is to certify that the project report entitled "**GSM AND ARDUINO BASED POWER THEFT DETECTION AND PROTECTION**" being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a bonafide work done by **K.Jessy Mole, M.DurgaMalleswari, G.Anil, A.Ashok** under my guidance during the academic year 2020-2021 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree

T.JAYARAJU

M.JAYARAJU,M.TECH
Assistant professor
Project Guide

P.PRADEEP,M.TECH,PH.D
Associate Professor
Head of the Department

DR. K. SRINIVAS, PH. D
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

Design and development of multilevel inverter with reduced switches for electric vehicle applications

A Project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRICAL AND ELECTRONICS ENGINEERING

By

Mounika G	18H75A0227
BalaSaiKiran.T	17H71A0207
Phirod Kumar P	18H75A0229
Arun Kumar K	17H71A0204

Under the esteemed guidance of
Dr.B.K.Karunakar Rao, M. Tech.PhD,
Associate Professor

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to: JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

Department of EEE

2020-21

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.

CERTIFICATE

This is to certify that the seminar report entitled **"Design and development of multilevel inverter with reduced switches for electric vehicle applications"** being submitted for the partial fulfilment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering** from **Jawaharlal Nehru Technological University, Kakinada,** is a bonafide work done by **Mounika G, BalaSaiKiran.T, Phirod Kumar P , Arun Kumar K** under my guidance during the academic year 2020-21 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Dr.B.K.Karunakar Rao,PhD
Associate Professor
Project Guide

Dr. P. Pradeep, Ph.D
Associate Professor
Head of the Department

Dr. K.Srinivas, Ph.D.
Principal

Examiner

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

KAKINADA

DESIGN AND CONTROL OF A STANDALONE MICROGRID SYSTEM

A project report submitted to the

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRICAL AND ELECTRONICS ENGINEERING

By

A.V. ROHITKUMAR	(17H71A0257)
A. RAMAKRISHNA	(18H75A0252)
T. LAVANYA	(17H71A0226)
B. RAJESH	(17H71A0235)

Under the esteemed guidance of

Mr. K. NARESH, M. Tech

Assistant Professor

Department of EEE

Devineni Venkata Ramana & Dr. Himasekhar

MIC College of Technology

An Autonomous Institution

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada

Kanchikacherla - 521180, Krishna Dist., A.P. India.

2020-2021

Devineni Venkata Ramana & Dr. Himasekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada
Kanchikacherla - 521180, Krishna Dist., A.P, India

CERTIFICATE

This is to certify that the project report entitled **"DESIGN AND CONTROL OF A STAND ALONE MICROGRID SYSTEM"** being submitted for the partial fulfillment of the requirements for the award of the degree in **Bachelor of Technology in Electrical and Electronics Engineering from Jawaharlal Nehru Technological University, Kakinada**, is a bonafide work done by **A. V. ROHITHKUMAR, A. RAMAKRISHNA, T. LAVANYA** and **B. RAJESH** under my guidance during the academic year 2020-21 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr. K. Naresh, M.Tech
Assistant Professor
Project Guide

Dr. P. Pradeep, M.Tech, Ph.D
Professor
Head of the Department

Dr. K. Srinivas, M.Tech, Ph.D
Principal

Examiner

Department of Electrons & Communication Engineering

A.Y:2021-22

List of Major Projects

S.No	REG No	NAME OF THE STUDENT	PROJECT SUPERVISOR	Title of the Project
A1	18H71A0430	ANNAPUREDDY NIKHITHA	Dr. CH. PULLARAO	Network Slicing and Non-Orthogonal Multiple Access to Transmit data in a Mobile Hospital System using 5G
	18H71A0436	RAYAPUDI SATYA LAKSHMI CHARITHA		
	18H71A0410	GANDLAPATHI CHENNAREDDY		
A2	18H71A0439	SARANYA NAGUBANDI	Ms. T. SRIDEVI	Number Plate Detection
	18H71A0417	PEMMASANI HEMANTHA RAJU		
	18H71A0407	SHAIK BONI NEHA		
	18H71A0425	VEERLA NAGARJUNA		
A3	18H71A0424	KOYA NAGA MOUNIKA	Ms. S. T. MRUDULA	Under water image quality asssessment based on human visual system
	18H71A0421	NAGOTHI LOHITHA		
	18H71A0408	PUNYAPU CHAITANYA		
	18H71A0446	KANUMURI SRI CHANDRA KUMAR		
A4	18H71A0412	KANTHETI HARIKA	Mr. T. VIJAY KANTH	Photovoltaic thermoelectric generator monitring system using DAQ & atmega328p
	18H71A0444	KASULA SOMA SEKHAR		
	18H71A0434	NALLANI RASI		
	18H71A0401	NIMMAGADDA AKASH		
A5	18H71A0429	PASUPULETI NAVYA SREE	Mr. B. R. K. SINGH	Multi-Band Guided Decomposition forsingle image enhancement
	18H71A0406	GUMMA BHAVANA		
	19H75A0404	KATARAPU KISHORE BABU		
	18H71A0454	PEDDINENI UDAY KIRAN		
A6	18H71A0402	MANDADAPU ANITHA	Mr. Y. V. N. M. SARMA	

	18H71A0403	VADLAMUDI ANUDEEPTHI		Monitoring and managing
	18H71A0414	BHUKYA HARISHNAIK		
	18H71A0411	NAKKA DURGA MALLESWARA RAO		

A7	18H71A0420	AVULURI LAVANYA LAHARI	Dr. P. RAJANI	Reversible data embeding for tamper proof watermarking
	18H71A0416	DARA HARSHITHA		
	18H71A0443	SUVVADA SIVA MANIKANTA		
	18H71A0445	VINNAKOTA SRAVAN VISWANATH		

A8	18H71A0437	MULE SAI LAKSHMI	Mr. K VEENANAND	Low Power Coding Approach Using of lagger Algorithm
	18H71A0431	MANGALA PAUL VINCENT		
	18H71A0451	BALUSUPATI SWARNA		
	19H75A0402	BADUGULA GOPINATH		

A9	18H71A0455	C V VISHNUPRIYA	Mr. G. KIRAN KUMAR	Low cost VLSI Architecture for image scaling
	18H71A0449	TALLURI SRUJANA		
	18H71A0404	MACHHA BALA NAGA NARASIMHA REDDY		
	18H71A0435	ABDUL RAWOOF		

A10	18H71A0413	BOMMAREDDY HARINI REDDY	Ms. K. MANASA LAKSHMI	Design of Raspberry pi web based energy monitoring system for residential electricity consumption
	18H71A0440	KANNETI SITA RAMA SAI		
	18H71A0405	ALLADA BHANU MURTHY		
	18H71A0432	CHIRUMAMILLA PRANEETHA		

A11	18H71A0418	PATHAN KHADHAR HAMEED KHAN	Mr. T. NARENDRA	Real-time wireless Embedded Electronics for Soldier Security
	18H71A0427	KOTAGIRI NARMADA		
	18H71A0458	VENNABOYINA VENKAT RAJ		
	18H71A0456	NALLAMOTHU VANAJA		

A12	18H71A0422	KAVARTAPU MAHESH BABU	Ms. B RAJYALAKSHMI	

	18H71A0441	ADIRAJU SESHA SAI PADMA MANOJNA		Brain tumor Detection and segmentation from MRI Images
	18H71A0433	SEEMAKURTHI RAJASHREE		
	18H71A0419	BEZAWADA KARTHIK		

A13	18H71A0453	ANUMALASETTY TARAKANADH	Dr. B PRAGATHI	Smart Saline Monitoring System
	18H71A0409	KOTHA CHAITANYA SRI		
	18H71A0415	HARSHINI REDDY SURA		
	19H75A0401	BOLLI BHANU SHANKAR		

A14	18H71A0459	JEEGURU VIJAYA KUMARI	Mr. D. RAHUL	Detection Lung Cancer using CT scan
	18H71A0447	TARIGOPULA SRIKANTH		
	19H75A0403	TEEGALA JYOTHI		
	18H71A0426	KAKARLA NAGENDRA BABU		

A15	18H71A0423	YANAGANDALA MOUNIKA	Mr. L. TIRU GANESH	Detection and Classification of COVID-19 cases using CNN and X-ray Figures in Matlab
	18H71A0442	RACHAGOLLA SIREESHAA		
	18H71A0438	BOJEDLA SAI SUDHEER		

A16	18H71A0457	HARI VANDITHA	Mr. M. NAGESWARA Rao	Automatic control system for electric vehicles
	18H71A0428	KANDRU NATHAN ISRAEL		
	18H71A0452	LAKKISETTI SYAM KUMAR		

B1	18H71A0497	GURRALA RISHITHA	Dr. B PRAGATHI	Video image deblurring algorithm based on denoising engine
	18H71A04B0	UJJWAL VADEGHARU		
	18H71A0493	CHUNDU PRANATHI		
	18H71A0474	MUNAGALA GOPI		

B2	18H71A04A2	CHARUGUNDLA SHALINI	Mr. D. RAHUL	A novel image processing approach to detect retina
	18H71A0480	KORIKANI JEEVITESH		
	19H75A0407	GULIPALLI POOJITHA		
	18H71A0468	JETTI CHARAN SAI		

				lesins in fundus images

B3	18H71A0473	REDDY GEETHIKA	Dr. CH. PULLARAO	Analytical review of cybersecurity for embedded systems
	18H71A04A5	GANNAMANENI SUMANJALI		
	18H71A0477	MANDADI HARISH		
	18H71A0482	MEENUGU KARTHEEK		

B4	18H71A0469	CHOPPAVARAPU CHARAN TEJA	Mr. CH. LAKSHAMANA	Virtual Doctor Robot
	18H71A0496	ADIMADHYAM RAMANI PRIYADARSINI WAGDEVI		
	18H71A04B2	MARAMREDDY VENKATA HARITHA		
	19H75A0409	UPPULURI RAHUL ROY		

B5	18H71A0485	KEERTHI S	Ms. G. NAGA LAKSHMI	Solar Tracking system using two axis solar panel
	18H71A04C0	DANDABATTULA YEGNITHA		
	18H71A0475	JONNALAGADDA GOPI CHAND		
	18H71A0491	KASTHALA NAVEEN		

B6	18H71A04B4	CHELLU VENKATA SIVA SAI	Mr. G.V.P. CHANDRA SEKHAR	Visualization of sunsurface defects using barker coded thermal wave imaging
	18H71A0476	AKULA GUNA SEKHAR		
	18H71A04A4	PARSA LAKSHMI SRUTHA KEERTHI		
	18H71A0494	CHERUKU PRIYANKA		

B7	18H71A04A3	GONE SRILATHA	Mr. G.V.P. CHANDRA SEKHAR	Detection of osteoprosis using non-stationary thermal wave imaging
	19H75A0406	SYED MEER MALIK		
	18H71A0484	DASARI KEERTHI REDDY		
	18H71A0464	CHINTALA CHAKRAPANI		

B8	18H71A04A7	VEERANKI SUPRIYA	Mr. S. RAMA KOTESWARA RAO	Segmentation of the blood vessels and optic disk in retinal images
	18H71A0488	TIRUNAGARI LAVANYA		
	18H71A04B6	CHITTARU VIJAY CHANDU		
	18H71A04B5	KOMMINENI VENKATESH		

B9	18H71A04B7	INALA VIJAY DEEPAK	Ms. S. T. MRUDULA	An adaptive over-segmentation
	18H71A04B3	TANGELLAMUDI VENKATA RAKESH		
	18H71A0487	LAVANYA THUMATI		

	18H71A0465	AMBATI CHANDANA		and feature point matching approach for detection of image forestry

B10	18H71A0478	MUTYALA HEMALATHA	Mr. T. VIJAY KANTH	An improved frequency resolution in a mircrocontroller using two phase lock-in amplifier
	18H71A0492	GUNDABATHINA PAVANI		
	18H71A0483	UPPALAPU KARTHIK NAGA SAI		
	18H71A0466	BALASANI CHANDRA VAMSI		

B11	18H71A0470	THOTA DHARANI DEVI	Ms. G. ANANTHA LAKSHMI	Credit card fraud detection using machine learning
	18H71A04B9	DAMARLA YASASWINI		
	18H71A0489	SANGANA MUKESH REDDY		
	19H75A0405	MANDA MAHESH BABU		

B12	18H71A04A0	MOHAMMAD SAMREEN	Mr. Y. V. N. M. SARMA	An inversible message embedding for JPEG images VIA
	18H71A0472	VUNGARALA DURGA LAKSHMI VENKATA DEEPAK		
	18H71A0462	MANDURI ANUSHA		
	18H71A04B8	ALLU VINAY KUMAR REDDY		

B13	18H71A0463	BATHULA BHANU REKHA	Mr. B.R.K. SINGH	Identification of Parkinson's disease using voice signal
	18H71A0471	MITTAPALLI DIVYA TEJA		
	18H71A04A1	MOHAMMED SAQLAIN ALI		
	18H71A0498	GRANDHI SAI GANESH		

B14	19H75A0408	GUDIMETLA PUSHPA LEKHA	Dr. P. RAJANI	An image enhancement algorithm of x-ray security inspection images
	18H71A0481	THOTAKURA KALYAN KUMAR		
	18H71A04A8	YANGALA SURYA PRAKASH		
	18H71A0486	ADDAGARLA LAKSHMI PRASANNA		

B15	18H71A0490	KANNETI NAGA SAI SRI	Mr. M. NAGESWARA RAO	Bee foraging algorithm based Multi-
	18H71A0495	VEMA PRUDHVI SAI		
	18H71A0467	POTTIPATI CHANDRIKA		

	18H71A0479	KANDULA JAYANTH NAGA KARTHIKEYA SWAMY		level thresholding for image segmentation

B16	18H71A04A6	JORIGE SUNANDA	Ms. G. ANANTHA LAKSHMI	Plant diesease recognition: A large-scale benchmark Datasheet and a visual region and loss reweighting approach
	18H71A0461	SHAIK ALMAS		
	18H71A04A9	YARRABOLU TEJA SRI		
C1	18H71A04H8	JONNALAGADDA VYSHNAVI	Mr. B.R.K. SINGH	Multi-Lingual Home Automation Using Node MCU
	18H71A04G2	YARRAMSETTI SINDHU		
	18H71A04G8	KURAPATI SRI SAI TARUN		
	18H71A04H0	BOINA SRIKANTH		

C2	18H71A04E5	YERRAGONDA NAGA BABU	Mr. K VEENANAND	Towars A Sustainable Development
	18H71A04H2	LANKALAPALLI SUPRAJA		
	18H71A04E6	MARIDU NAGA VYSHNAVI		
	19H75A0412	KANKIPATI SAIDEEP		

C3	19H75A0416	AYANACHU VANI	Mr. G. KIRAN KUMAR	Sun Tracking Solar Panel
	18H71A04F6	MAREMALLA SAI GUPTA VINAY		
	18H71A04G4	VASANTHA SIVA PARVATHI		
	18H71A04E3	GADIDESI MANOJ		

C4	18H71A04E7	KAKANI NAVYA	Ms. K. MANASA LAKSHMI	Reversible data hiding technique with high message embedding capacity in images
	18H71A04F9	KALISETI SASHANK BABU		
	18H71A04F7	SUNKARA SAI KAMAL		

C5	18H71A04H4	ALLU TEJASRI	Mr. T. NARENDRA	Water Supply controlling system with water theft identification using node MCU
	18H71A04C1	CHALLAPUREDDI ANUSHA		
	18H71A04C6	MARRAPU BHARATH KISHORE		
	18H71A04F3	GOBERU RAMA RAO		

C6	18H71A04H6	MERAKANAPALLI VAMSI	Ms. B RAJYALAKSHMI	Underwater Wireless Communication System
	18H71A04E8	PILLI NITHIN		
	18H71A04E9	GURRAM PAVAN KIRAN		
	18H71A04C9	CHIRUMAMILLA CHINMAI SAI GOPI CHAND		

C7	18H71A04G7	NAMBURI SRAVYA	Dr. B PRAGATHI	Design and Implementation
	19H75A0417	AILAPURAPU VENKATA VISWANADH		
	18H71A04F1	MALAJI RAJITHA		
	18H71A04C4	SHAIK BASHID		

C8	18H71A04D8	KOTHARU LALITHA	Mr. S. RAMA KOTESWARA RAO	Prediction of heart Disease
	18H71A04H7	VEGINATI VENKATESWARLU		
	19H75A0414	SUDHULA SRAVYA		
	18H71A04G0	DAMARACHARLA SASI KUMAR		

C9	18H71A04C3	VANKAYALAPATI ARCHANA	Mr. L. TIRU GANESH	IOT Covid Patient Health Monitor in Quarentine
	19H75A0411	YADAVALLI RUSHITHA		
	18H71A04D5	CHILAKALA KASI VISWANATH REDDY		
	18H71A04F4	UPPELA ROHITH REDDY		

C10	18H71A04G5	CHERUKU SIVA SRI	Mr. CH. LAKSHAMANA	Smart Shoe for Visually Impaired
	18H71A04C8	CHIGURUPATI CHANDANA		
	19H75A0413	CHERUKU SIVA NAGA MALLESWARA RAO		
	18H71A04D0	DEVARAPALLI DEEVAN KUMAR		

C11	19H75A0410	PALAGANI RUCHITHA	Ms. G. NAGA LAKSHMI	Digital watermarking technology in information security
	18H71A04D9	GUNTUPALLI LAVANYA		
	18H71A04F5	CHILAKALA SAI GREESHMANTH		
	18H71A04C5	REVALLA BHANU PRAKASH		

C12	18H71A04G1	DEVALLA SHIRISHA	Ms. G. ANANTHA LAKSHMI	Mask Classification and head temperature detection combined with Deep Learning networks
	18H71A04C2	BATTINENI ANVITHA		
	18H71A04I0	PODILA LAKSHMI TEJA		
	18H71A04D1	BATTHULA DILIP		

C13	18H71A04C7	AMBATI BHAVANI	Mr. CH. PULLARAO	Neural Voice Cloing
	18H71A04E4	BORRA MONIKA		

	18H71A04F0	BAVISETTI PAVAN KUMAR		
	19H75A0418	GARIKAMUKKU VIKAS		

C14	18H71A04D4	VALLABHANENI JASWANTH SAI	Ms. T. SRIDEVI	Ultra-Wideband Imulse radar through-Wall Detection of Vital Signs
	18H71A04H9	GUDIMETLA YASHASWINI		
	18H71A04G9	SEELAM SRIVANI REDDY		
	18H71A04E1	VADDADI LOKESH		

C15	18H71A04D6	MUPPASANI KEERTHANA	Dr. P. RAJANI	Design of Agriculture Irrigation system Based on Wireless Communications
	18H71A04G6	CHEKURI SRAVANI		
	18H71A04E2	TATINENI MAN VENKATESH		
	18H71A04H3	MATTE SURESH BABU		

Signature of Coordinator
(G.V.P.Chandra Sekhar)

Signature of HoD
(Dr.Ch.Pulla Rao)

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"AUTOMATIC COVID – 19 FACE MASK AND TEMPERATURE DETECTION SYSTEM"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

K. KEERTHI (17H71A0416) **P. ALEKHYA (17H71A0463)**

P. LEELA KUMARI (17H71A0483)

Under the Esteemed Guidance of
Mr. B. R. K. Singh, M. Tech, MIE
ASSOCIATE PROFESSOR

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist., A.P. India
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

NBA-Accredited

2020 - 2021

CERTIFICATE

This is to certify that the Main Project entitled **"AUTOMATIC COVID – 19 FACE MASK AND TEMPERATURE DETECTION SYSTEM"** is a bonafide work carried out by, **K. Keerthi (17H71A0416), P. Alekhya (17H71A0463), P. Leela Kumari (17H71A0483)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year **2020-2021**.

(Mr. R. K. Singh)
Project supervisor

(Mr. CH. PULLA RAO)
Head of the Department

(Dr. K. SRINIVAS RAO)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"SMART CROP MONITORING AND AUTOMATIC IRRIGATION SYSTEM USING RASPBERRY PI"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

V.Sunoj Kumar (17H71A04A7)

A.Sai Chakravarthi (17H71A0494)

S.H.V.N. Sai Venkat(17H71A0410)

Sd.Reshma(17H71A0446)

Under the Esteemed Guidance of
Mr. L. Tiru Ganesh, M.Tech.,
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist., A.P, India
Phone: 08678 - 273535, 273621, Fax: 08678 - 271569
E mail: dvrhimic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist. A.P. India
Phone: 08678 - 273535, 273623, Fax 08678 - 273569
E-mail: [illegible], Website: www.mictech.ac.in
2020 - 2021

NBA Accredited

Accredited by NAAC A

CERTIFICATE

This is to certify that the Main Project entitled **"SMART CROP MONITORING AND AUTOMATIC IRRIGATION SYSTEM USING RASPBERRY PI"** is a bonafide work carried out by **V.Sunnj Kumar (17H71A04A7), A. Sai Chakravarthi (17H71A0494), S.H.V.N.Sai Venkat (17H71A0410), Sd.Reshma (17H71A0446)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-2021.

(Mr. L. Tiru Ganesh)
Project supervisor

(Mr.CH. Pulla Rao)
Head of the Department

(Dr. K.Srinivas)
Principal

S.T. Mrudula

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"SMART EYE TRACKING SYSTEM"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

M.Sri Naga Sai Durga Tejaswi(17H71A04G2)

M.Gopinadh(17H71A0408)

K.Bhargav (17H71A0466)

A.Sai Dinesh(17H71A0495)

Under the Esteemed Guidance of
Mr.G.SIVARAMAKRISHNA, M.TECH.
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist. A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@gmail.com, Website: www.micet.ac.in

2020-2021

NBA-Accredited
Accredited by
NAAC

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 272535, 272622, Fax: 08678 - 271565
E mail: dvrlamic@mictech.ac.in, Website: www.mictech.ac.in
2020-21

NBA-Accredited
Accredited by

CERTIFICATE

This is to certify that the Main Project entitled **"SMART EYE TRACKING SYSTEM"** is a bonafide work carried out by **M.Sri Naga Sai Durga Tejaswi(17H71A04G2) , M.Gopinadh(17H71A0408) , K.Bhargav(17H71A0466) , A.Sai Dinesh(17H71A0495)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mr. G. SIVARAMAKRISHNA.)
Project Supervisor

(Mr. CH. Pulla Rao)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"DETECTION OF RENAL CALCULI USING K-MEANS CLUSTERING ALGORITHM"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

D.SATYA VENKATA TEJASWI (17H71A04F9) **J.MANIKANTA(17H71A0485)**

G.RUPESH(17H71A0493) **L.SUSHMITHA(17H71A04A8)**

Under the Esteemed Guidance of
Mr. Ch Lakshmana, M. Tech,
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone : 08678 - 273535, 273623, Fax 08678 - 273[illegible]
E mail: [illegible], Website: [illegible]

2020-2021

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E-mail: dvrhsmictech@mictech.ac.in, Website: www.mictech.ac.in

2020-21

NBA-Accredited

Accredited by NAAC With Grade A

CERTIFICATE

This is to certify that the Main Project entitled **"DETECTION OF RENAL CALCULI USING K-MEANS CLUSTERING ALGORITHM"** is a bonafide work carried out by **D.S.V.TEJASWI(17H71A04F9), J .MANIKANTA(17H71A0485), G.RUPESH(17H71A0493), L.SUSHMITHA(17H71A04A8)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mr. Ch Lakshmana)
Project Supervisor

(Mr. CH. PullaRao)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"Multi-purpose Solar Grass Cutter"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

P. Sailaja (17H71A04F5)
T.Ashok kumar (17H71A04C4)
V R C K Bharadwaj T (18H75A0419)
D. Mohana Lavanya (17H71A0486)

Under the Esteemed Guidance of

Mr.Y V N M SARMA, M.Tech
ASSISTANT PROFESSOR

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E- mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in
2020-2021

NBA-Accredited
Accredited by NAAC

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@micctech.ac.in, Website: www.micctech.ac.in

NBA-Accredited
Accredited by NAAC A

2020-21

CERTIFICATE

This is to certify that the seminar entitled **"Multi-purpose Solar Grass Cutter"** is a bona-fide work carried out by **P.Sailaja (17H71A04F5), T.Ashok kumar (17H71A04C4), V R C K Bharadwaj T (18H75A0419), D. Mohana Lavanya (17H71A0486)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering of Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mr.V V N M SARMA)
Project Supervisor

(Mr. CH. Pulla Rao)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

KAKINADA

"DETECTION OF LEAF DISEASE USING SVM CLASSIFIER"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

V. VEERA VENKATA SATISH KUMAR (17H71A04H1)
D. VINEELA (17H71A0460)
CH. SUNEETHA (18H75A0415)
P. MAHESH (17H71A0420)

Under the Esteemed Guidance of
Mr. Y. VIKAS, M. Tech.,
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhamic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 275535, 275623, Fax : 08678 - 273569
Email: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"DETECTION OF LEAF DISEASE USING SVM CLASSIFIER"** is a bonafide work carried out by **V. V. V. SATISH KUMAR (17H71A04H1), D. VINEELA (17H71A0460), CH. SUNEETHA (18H75A0415), P. MAHESH (17H71A0420)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mr. Y. Vikas)
Project Supervisor

(Mr. CH. Pulla Rao)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLALNEHRUTECHNOLOGICALUNIVERSITYKAKINADA

KAKINADA

"LECARM: LOW-LIGHT IMAGE ENHANCEMENT USING CAMERA RESPONSE MODEL"

Submitted to
Jawaharlal Nehru Technological University Kakinada
In partial fulfillment for the award of the degree of

BACHELOROFTECHNOLOGY
in
ELECTRONICS&COMMUNICATIONENGINEERING

Submitted by

B. Tejaswini (17H71A0455)

A. Sucharita (17H71A04G5)

SD. Irfan (17H71A0411)

A. Pavan Kumar (17H71A0438)

Under the Esteemed Guidance of
Mr. L. TIRU GANESH, M. Tech.,
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

DevineniVenkataRamana& Dr.HimaSekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)Kanchikacherla -521180,KrishnaDist., A.P,India.

Phone:08678 -273535, 273623, Fax:08678 -273569

Email:devhanic@mictech.ac.in,Website:www.mictech.ac.in

2020-2021

NBA-Accredited

Accredited by

DevineniVenkataRamana& Dr.HimaSekhar
MIC College of Technology
(ApprovedbyAICTE&PermanentlyAffiliatedtoJNTUK,Kakinada)Kan
chikacherla - 521180, KrishnaDist,A.P, India.
Phone:08678-273535,273623,Fax:08678-273569
Email:dvrhsmic@mictech.ac.in,Website:www.mictech.ac.in

NBA-Accredited
Accredited by NAAC A With Grade

2020 -2021

CERTIFICATE

This is to certify that the Main Project entitled **"LECARM: LOW-LIGHT IMAGE ENHANCEMENT USING CAMERA RESPONSE MODEL"** is a bonafide work carried out by, **B. Tejaswini (17H71A0455), A. Sucharita (17H71A04G5), SD. Irfan (17H71A0411), A. Pavan Kumar (17H71A0438)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year **2020-2021.**

(MR. L. TIRU GANESH)
Project supervisor

(Mr. CH. PULLA RAO)
Head of the Department

(Dr. K. SRINIVAS RAO)
Principal

Examiner1

Examiner2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

KAKINADA

"A Novel Authentic Watermarking for QR Code Image using Wavelet Transform"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

U. Gayathri (18H75A0403)

KVSSN. Raj Kumar (17H71A04B6)

M. Druga Akhila (18H75A0401)

R. Hari Prasad (17H71A04D2)

Under the Esteemed Guidance of
Mr. D. RAHUL, M.S.,
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist., A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

NBA-Accredited
B.Tech (CSE | ECE | EEE | MEC)

Accredited by
NAAC A
With Grade

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated
to JNTUK, Kakinada)Kanchikacherla -
521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhamic@mictech.ac.in, Website: www.mictech.ac.in

2020 - 2021

CERTIFICATE

This is to certify that the Main Project entitled **"A NOVEL AUTHENTIC WATERMARKING FOR QR CODE IMAGE USING WAVELET TRANSFORM"** is a bonafide work carried out by **U. Gayathri (18H75A0403), KVSSN. Raj Kumar (17H71A04B6), M. Druga Akhila (18H75A0401), R. Hari Prasad (17H71A04D2)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mr. D. Rahul)
Project Supervisor

(Mr. CH. Pulla Rao)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

KAKINADA

"A NOISE-AWARE ENHACEMENT METHOD FOR UNDEREXPOSED IMAGE"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

D.RAJESH (17H71A0490)

M.NARENDRA REDDY (17H71A04E5)

CH.NIHARIKA (17H71A04E6)

K.SRIVANI (17H71A04G4)

Under the Esteemed Guidance of
Mr.Y.VIKAS, M.Tech
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist., A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

NBA-Accredited

Accredited by

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

NBA-Accredited
Accredited by NAAC A

2020 - 2021

CERTIFICATE

This is to certify that the Main Project entitled **"A NOISE-AWARE ENHANCEMENT METHOD FOR UNDEREXPOSED IMAGE"** is a bonafide work carried out by **D.Rajesh (17H71A0490), M.Narendra Reddy (17H71A04E5), Ch.Niharika (17H71A04E6), K.Srivani (17H71A04G4)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-2021.

(Mr. Y. VIKAS)
Project supervisor

(Mr. CH. Pulla Rao)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

KAKINADA

"UNDERWATER IMAGE ENHANCEMENT"

Submitted to

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

S.N.D.V.M Ashwith (17H71A04E3)

M. Hari Chandana (17H71A0474)

SK. Rasoolbi (17H71A04F0)

G. Sai Ganesh (17H71A0447)

Under the Esteemed Guidance of

Mr. D. Rahul, M.S.

Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)

Kanchikacherla - 521180, Krishna Dist, A.P, India.

Phone : 08678 - 273535, 273623, Fax: 08678 - 273569

E mail: dvhsmic@mictech.ac.in, Website: www.mictech.ac.in

NBA-Accredited

Accredited by NAAC With Grade A

2020-2021

Devineni Venkata Ramana & Dr Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)

Kanchikacherla - 521180, Krishna Dist, A.P, India.

Phone : 08678 - 273535, 273623, Fax: 08678 - 273569

E mail: dvchumic@mictech.ac.in, Website: www.mictech.ac.in

NBA-Accredited B.Tech (CSE, ECE, EEE, ME)

Accredited by NAAC A With Grade

An ISO 9001:2008 Certified Institute

2020-2021

CERTIFICATE

This is to certify that the Main Project entitled **"UNDERWATER IMAGE ENHANCEMENT"** is a Bonafede work carried out by **S.N.D.V.M Ashwith (17H71A04E3), SK. Rasoolbi (17H71A04F0), M. Hari Chandana (17H71A0474), G. Sai Ganesh (17H71A0447)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2018-19.

(Mr. D. Rahul)
Project Supervisor

(Mr. Ch. Pullarao)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

KAKINADA

"COVID-19 INDOOR SAFETY MONITORING SYSTEM"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

SD. NURJAHAN (17H71A0436)	S. CHANDANA TANVI (17H71A0468)
M.L.V.V. SUMANTH (17H71A0484)	J. VARUN (17H71A04B4)

Under the Esteemed Guidance of
Mr. G. SIVA RAMA KRISHNA, M.TECH
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist., A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

NBA-Accredited

Accredited by NAAC A

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

NBA-Accredited (B.Tech - CSE, ECE, EEE, ME)
Accredited by NAAC A With Grade

2020 - 2021

CERTIFICATE

This is to certify that the Main Project entitled **"COVID-19 INDOOR SAFETY MONITORING SYSTEM"** is a bonafide work carried out by, **Sd. Nurjahan (17H71A0436), S. Chandana Tanvi (17H71A0468), M.L.V.V.Sumanth (17H71A0484), J. Varun (17H71A04B4)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-2021.

(Mr. G. Siva Rama Krishna)
Project supervisor

(Mr. CH. Pulla Rao)
Head of the Department

(Dr. K . Srinivas)
Principal

PRINCIPAL
Devineni Venkata Ramana & Dr Himasekhar
MIC College of Technology
Kanchikacherla, Krishna District

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"SOLAR BASED HOME AUTOMATION"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS AND COMMUNICATION ENGINEERING

Submitted by

B. RAJITHA(17H71A04E9)

A. NAGA SIVA KYATHI(17H71A0428)

V. RAVI KUMAR (18H75A0411)

MD.AYUB(16H71A04B4)

Under the Esteemed Guidance of
Mr .CH.LAKSHAMANA, M.Tech.,
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@gmictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana &Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in
2020-2021

CERTIFICATE

This is to certify that the Main Project entitled **"SOLAR BASED HOME AUTOMATION"** is a bonafide work carried out by **B.RAJITHA (17H71A04E9), A.NAGA SIVA KYATHI (17H71A0428), V.RAVI KUMAR (18H75A0411) and MD.AYUB (16H71A04B4)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21

(Mr.CH.Lakshamana)
Project Supervisor

(Mr.CH.Pulla Rao)
Head of the Department

(Dr. K. Srinivasa Rao)
Principal

Examiner1

Examiner2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"TOXIC GASES DETECTION USING IoT AND GSM MODULE"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

B. Niteesha (17H71A0435) **R. Lakshmi Prasanna(17H71A0418)**

G. Ramya (17H71A0442) **CH. Annapurna (17H71A0403)**

Under the Esteemed Guidance of

Ms. G. Naga Lakshmi, M. Tech.,
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna dist., A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: [illegible]@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

NBA-Accredited
Accredited by NAAC A

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@micctech.ac.in, Website: www.micctech.ac.in
2020-2021

NBA-Accredited
Accredited by NAAC With Grade A

CERTIFICATE

This is to certify that the Main Project entitled **"TOXIC GASES DETECTION USING IoT AND GSM MODULE"** is a bonafide work carried out by **B. Niteesha (17H71A0435), R. Lakshmi Prasanna (17H71A0418), G. Ramya (17H71A0442), CH. Annapurna (17H71A0403)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Ms. G. Naga Lakshmi)
Project Supervisor

(Mr. CH. PullaRao)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"CORPSE DISPENSING AND SANITIZATION USING ARDUINO"

Submitted to
Jawaharlal Nehru Technological University Kakinada
In partial fulfillment for the award if the degree of

BACHELOR OF TECHNOLOGY
In
ELECTRONICS AND COMMUNICATION ENGINEERING

Submitted by

V.ANUSHA(17H71A0464) **A.BAVYA (17H71A0406)**

K.PRIYANKA (17H71A0439) **K.SUDHARANI (17H71A0453)**

Under the Esteemed Guidance of

Mr.Y V N M SARMA , M.Tech

ASSISTANT PROFESSOR

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone : 08678 - 273535, 273623, Fax: 08678 - 271569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone : 08678 - 273535, 273621, Fax 08678 - 273569
E mail: devineni@mictech.ac.in, Website: www.mictech.ac.in

NBA-Accredited
Accredited By NAAC With Grade A

2020-21

CERTIFICATE

This is to certify that the project entitled **"CORPSE DISPENSING AND SANITIZATION USING ARDUINO"** is a bonafide work carried out by **V. ANUSHA (17H71A0464), A. BAVYA (17H71A0406), K. PRIYANKA (17H71A0439), K. SUDHA RANI (17H71A0453)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-2021.

(Mr. Y V N M SARMA)
Project Supervisor

(Mr. CH. Pulla Rao)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"MEDICAL IMAGE CONTRAST ENHANCEMENT USING RETINEX ALGORITHM"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

S. Keerthana (17H71A0479) **A. S. Safa (17H71A0461)**

N. Jala Samba Siva Rao (17H71A04D6) **P. Sri Vinay (17H71A0452)**

Under the Esteemed Guidance of

Ms. G. Naga Lakshmi, M. Tech.,
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone : 08678 - 273535, 273623, Fax 08678 - 273569
E-mail: drdvhmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone: 08678 - 271535, 271621, Fax: 08678 - 271569
E-mail: info@mictech.ac.in, Website: www.mictech.ac.in

NBA Accredited
Accredited by

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"MEDICAL IMAGE CONTRAST ENHANCEMENT USING RETINEX ALGORITHM"** is a bonafide work carried out by **S. Keerthana (17H71A0479), A. S. Safa (17H71A0461), N. Jala Sambu Siva Rao (17H71A04D6), P. Sri Vinay (17H71A0452)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Ms. G. Naga Lakshmi)
Project Supervisor

(Mr. CH. Pulla Rao)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

KAKINADA

"DETECTION OF CANCER IN PANCREATIC CELLS USING MINIMUM DISTANCE CLASSIFIER ALGORITHM"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

K.YESWANTH (17H71A04H7)

N.HAJARATH REDDY (17H71A0409)

A.P S S DHARANI (17H71A0469)

P.SAMUEL ANANDA RAO (17H71A0449)

Under the Esteemed Guidance of
Mr.G.SIVA RAMA KRISHNA , M.Tech.,
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E-mail: deftonic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

NBA-Accredited

Accredited by

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone : 08678 - 273535, 273623, Fax: 08678 - 273560
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-21

NBA-Accredited

Accredited by

CERTIFICATE

This is to certify that the Main Project entitled **"DETECTION OF CANCER IN PANCREATIC CELLS USING MINIMUM DISTANCE CLASSIFIER ALGORITHM"** is a bonafide work carried out by **K.YESWANTH (17H71A04H7) N.HAJARATH REDDY (17H71A0409),A.PSSDHARNI(17H71A0489)P.SAMUEL ANANDA RAO (17H71A0449)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mr.G. Siva Rama Krishna)
Project Supervisor

(Mr. CH. Pulla Rao)
Head of the Department

(Dr. K. Srinivas)
Principal

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"HIGH RESOLUTION SATELLITE IMAGES USING WAVELET TRANSFORMS "

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

Suchitha Lakshmi K(17H71A04A3)

Lavanya N(17H71A04E0)

Hussain Sk(17H71A04D4)

Vamsi Krishna A(17H71A04G8)

Under the Esteemed Guidance of
Mr. D. Rahul, M.S.
Assisstant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kandrika Kuduru - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273568
E mail: deebamic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E-mail: [illegible], Website: www.mictech.ac.in

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"HIGH RESOLUTION SATELLITE IMAGES USING WAVELET TRNSFORMS"** is a bonafide work carried out by **K. Snehitha Lakshmi (17H71A04A3), Sk. Hussain (17H71A04D4)**,**N. Lavanya (17H71A04E0), A. Vamsi Krishna (17H71A04G8)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mr. D. Rahul)
Project Supervisor

(Mr. CH. Pulla Rao)
Head of the Department

(Dr. K. Srinivas)
Principal

g. T. Mrudula
Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

KAKINADA

"PRECISION AGRICULTURE and YIELD PREDICTION BASED ON IN-FIELD ELECTRONIC SYSTEM AND METHODOLOGY"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

ELECTRONICS&COMMUNICATION ENGINEERING

Submitted by

T. DIVYA SRI LAKSHMI (17H71A0471)I. KALPANA (17H71A0414)

N. SIVARAM (17H71A04A2) **D. PAVAN CHANDRA(17H71A0437)**

Under the Esteemed Guidance of

Mr. T.VIJAYA KANTH, M.Tech, MISTE

Assistant professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)

Kanchikacherla - 521180, Krishna Dist,A.P, India

Phone : 08678 - 273535, 273623, Fax: 08678 – 273569

E mail: dvrhsmic@micetech.ac.in, Website: www.micetech.ac.in

NBA-Accredited

Accredited by

NAAC

2020-2021

Devineni Venkata Ramana & Dr.Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist,A.P, India
Phone : 08678 - 273535, 273623, Fax 08678 - [illegible]
E-mail: [illegible], Website: www.mictech.ac.in

2020-2021

NBA-Accredited
(B.Tech: CSE, ECE, EEE, ME)
Accredited by
NAAC
With Grade A
An ISO 9001:2008
Certified Institute

CERTIFICATE

This is to certify that the Main Projectentitled "**PRECISION AGRICULTURE AND YIELD PREDICTION BASED ON IN FIELD ELECTRONIC SYSTEM AND METHODOLOGY**"is a bonafide work carried out by**T. DIVYA SRI LAKSHMI (17H71A0471), I. KALPANA (17H71A0414),N. SIVARAM (17H71A04A2), D. PAVAN CHANDRA(17H71A0437)**in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mr. T. VIJAYA KANTH)
Project Supervisor

(CH.PULLARAO)
Head of the Department

(Dr. K. Srinivas Rao)
Principal

S. T. Mrudula
Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"Development of Congestion Level Based Dynamic Traffic Management System Using IoT"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in Partial fulfilment for the award of the degree of

"BACHELOR OF TECHNOLOGY

in

ELECTRONICS & COMMUNICATION ENGINEERING"

Submitted by

Sk.Sadhika (17H71A04F2)	**Sk.Hameedha Begum (17H71A04D1)**
D.Karthik (17H71A04D8)	**K.Harish (17H71A04D3)**

Under the Esteemed Guidance of

Mr. T. VIJAYA KANTH (M.Tech, MISTE)
Assistant Professor
DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist., A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhimic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)

Kanchikacherla - 521180, Krishna Dist., A.P, India

Phone: 08678 - 273535, 273623, Fax: 08678 - 273569

E mail: dvrhmic@mictech.ac.in, Website: www.mictech.ac.in

NBA-Accredited

2020-2021

CERTIFICATE

This is to certify that the Main Project entitled -Development of Congestion Level Based Dynamic Traffic Management System Using IoT- is a bonafide work carried out by **Sk.Sadhika (17H71A04F2), Sk.Hameedha Begum (17H71A04D1), D.Karthik (17H71A04D8), K.Harish (17H71A04D3)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2017-2021.

(T. VIJAYA KANTH)
Project Supervisor

(Dr. CH.PULLARAO)
Head of the Department

(Dr. Srinivas)
Principal

Examiner1

Examiner2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"AN EFFICIENT IoT BASED REAL REMOTE TIME CARDIAC ACTIVITY"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

CH.S.V.PRIYANKA (17H71A0497) **G.JAYA SREE (17H71A0412)**

MD SHAKEEL (17H71A04A0) **M.VENKATA SAIRAM ROHITH(17H71A04H3)**

Under the Esteemed Guidance of
Mr. TANNERU NARENDRA KUMAR, M.Tech ,MISTE.
Assistant professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dirtanic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

NBA-Accredited

Accredited by NAAC

Devineni Venkata Ramana & Dr Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P. India
Phone : 08678 - 271535, 271621, Fax: 08678 - 271569
E mail: dvrhsmictech@gmail.com, Website: www.mictech.ac.in

2020-2021

NBA-Accredited

CERTIFICATE

This is to certify that the Main Project entitled **"AN EFFICIENT IoT BASED PLATFORM FOR REAL REMOTE TIME CARDIAC ACTIVITY"** is a bonafide work carried out by **CH.S.V.PRIYANKA (17H71A0497), G.JAYA SREE (17H71A0412),MD SHAKEEL (17H71A04A0), M.VENKATA SAIRAM ROHITH (17H71A04H3)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mr. TANNERU NARENDRA KUMAR)
Project Supervisor

(CH.PULLARAO)
Head of the Department

(Dr. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"A HYBRID FEATURE EXTRACTION METHOD WITH REGULARIZED EXTREME LEARNING MACHINE FOR BRAIN TUMOR CLASSIFICATION"

Submitted to

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

T. S. S. L. Bhavya Sri (17H71A04F8)

SK. Saida (17H71A04F4)

SK. Tasneem (17H71A04B0)

P. Naveen Kumar (17H71A0432)

Under the Esteemed Guidance of

Mr. Ch. Pullarao, M.Tech,

Associate Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr. Hima Nekkar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)

Kanchikacherla - 521180, Krishna Dist, A.P, India

Phone : 08678 - 273533, 273623, Fax: 08678 - 273569

E-mail: dvrhamic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)

Kanchikacherla - 521180, Krishna Dist, A.P, India

Phone: 08678 - 273535, 273623, Fax: 08678 - 273569

E-mail: dvrhimacet@mictech.ac.in, Website: www.mictech.ac.in

NBA-Accredited

Accredited by NAAC

2020-2021

CERTIFICATE

This is to certify that the Main Project entitled **"A Hybrid Feature Extraction Method with Regularized Extreme Learning Machine for Brain Tumor Classification"** is a bonafide work carried out by **T. S. S. L. Bhavya Sri (17H71A04F8), SK. Saida (17H71A04F4), SK. Tasneem (17H71A04B0), P. Naveen Kumar (17H71A0432)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mr. Ch. Pullarao)
Project Supervisor

(Mr. Ch. Pullarao)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner1

Examiner2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

KAKINADA

"SURVEILLANCE ROBOT WITH FACE RECOGNITION USING RASPBERRY PI"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

K.THARUN (17H71A04A9)

SK.MUNNY (17H71A0425)

K.ROHITH (17H71A0492)

V.VINUSHA (17H71A04H4)

Under the Esteemed Guidance of
MR.B.RADHA KRISHNA SINGH ,MIE,M.TECH
Associate Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist., A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhamic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

NBA-Accredited

Accredited by

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhamic@mictech.ac.in, Website: www.mictech.ac.in

NBA-Accredited

2020 - 2021

CERTIFICATE

This is to certify that the Main Project entitled **"SURVEILLANCE ROBOT WITH FACE RECOGNITION USING RASPBERRY PI"** is a bonafide work carried out by, **K.Tharun(17H71A04A9),SK.Munny(17H71A0425), K.Rohith (17H71A0492), V.Vinusha (17H71A04H4)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-2021.

(Mr.B.Radha Krishna Singh)
Project supervisor

(Mr.CH.Pulla Rao)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"DEPARTMENT INFORMATION SYSTEM USING SCRIPTING LANGUAGES"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS AND COMMUNICATION ENGINEERING

Submitted by

K.MEGHANA	**(17H71A04E1)**	**K.VENKATA GOWTHAM**	**(17H71A04B7)**
K. RAVI KIRAN	**(17H71A0445)**	**S.NANDINI**	**(17H71A0429)**

Under the Esteemed Guidance of
Mr. CH. PULLARAO, M.Tech.,
Associate Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone : 08678 - 273535, 273623, Fax : 08678 - 273569
E mail: drrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

NBA Accredited

Accredited by NAAC

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone : 08678 - [illegible], Fax 08678 - [illegible]
E mail: [illegible], Website: www.mictech.ac.in
2020-2021

NBA-Accredited

Accredited by NAAC

CERTIFICATE

This is to certify that the Main Project entitled **"DEPARTMENT INFORMATION SYSTEM USING SCRIPTING LANGUAGES"** is a bonafide work carried out by **K.MEGHANA (17H71A04E1), K.VENKATA GOWTHAM (17H71A04B7), K. RAVI KIRAN (17H71A0445) and S.NANDINI(17H71A0429)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mr. CH.PULLARAO)
Project Supervisor

(Mr. CH.PULLARAO)
Head of the Department

(Dr. K.SRINIVAS)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"RECOGNIZATION OF LEAF DISEASE USING IMAGE PROCESSING"

Submitted to

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

G.Keerthi(17H71A0480) **S.Chandrika(17H71A04C9)**

A.Venkata Manohar(17H71A04B8) **D.Yugandhar Kumar(17H71A04H8)**

Under the Esteemed Guidance of

Dr.B.PRAGATHI ,M.Tech,Ph.D

ASSOCIATE PROFESSOR

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

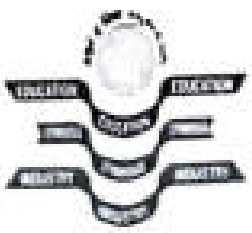
Devineni Venkata Ramana &Dr.Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada) Kanchikacherla - 521180, Krishna Dist., A.P.

India.Phone: 08678 - 273535, 273623, Fax: 08678 – 273569

E mail: dvrhamics@mictech.ac.in, Website: www.mictech.ac.in



2020-2021

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 271535, 271621, Fax: 08678 - 271569
E mail: dvrhimsr@mictech.ac.in, Website: www.mictech.ac.in

2020 - 2021

NBA-Accredited
Accredited by

CERTIFICATE

This is to certify that the Main Project entitled **"RECOGNIZATION OF LEAF DISEASE USING IMAGE PROCESSING"** is a bonafide work carried out by,**G.Keerthi (17H71A0480), S.Chandrika(17H71A04C9), A.Venkata Manohar(17H71A04B8), D.Yugandhar Kumar(17H71A04H8)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year2019-2020.

(Mrs. Dr. B. Pragathi)
Projectsupervisor

(Mr. CH. Pulla Rao)
Head of theDepartment

(Dr. K. Srinivas)
Principal

Examiner1

Examiner2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"WALK TO POWER AND SECURE"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

N.HEMA SRI (17H71A0477)

N. BALAJI(17H71A04C7)

G.SRAVYA((17H71A04G1)

G.SATHWIK BINNY JOSEPH(17H71A0450)

Under the Esteemed Guidance of

Mrs.Dr.B.PRAGATHI,M.Tech,Ph.D

ASSOCIATE PROFESSOR

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist., A.P. India
Phone: 08678 - 271935, 271621, Fax: 08678 - 271389
E-mail: deptece@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr.Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P. India
Phone : 08678 - 273535, 273623, Fax 08678 - 273569
E-mail: [illegible] Website: [illegible]

NBA-Accredited

Accredited by NAAC with A Grade

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"WALK TO POWER AND SECURE"** is a bonafide work carried out by **N.Hema Sri(17H71A0477)**, **N.Balaji(17H71A04C7)**, **G.Sravya(17H71A04G1)**, **G.Sathwik Binny Joseph (17H71A0450)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mrs.Dr.B.PRAGATHI)
Project Supervisor

(Mr. CH. PULLA RAO)
Head of the Department

(Dr. K. Srinivas)
Principal

S.T. Mrudula
Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"GSM BASED GARBAGE SORTING AND MONITORING SYSTEM"

Submitted to

Jawaharlal Nehru Technological University

Kakinada in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

T.Kalyan Babu(17H71A0415)

M.Hema Sai Annapurna(18H75A0405)

Y.Manikanta(17H71A0421)

Md.Abdul Raqib(17H71A0401)

Under the Esteemed Guidance of

Mr. CH. PULLARAO ,M.Tech

Associate Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata RamanaK. Dr.HimaSekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)

Kanchikacherla - 521180, Krishna Dist, A.P, India.

Phone : 08678 - 273535, 273623, Fax: 08678 - 273569

E mail: dehonic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK,Kakinada
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273823, Fax: 08678 - 272569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

NBA Accredited
Accredited by NAAC

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"GSM BASED GARBAGE SORTING AND MONITORING SYSTEM"** is a bonafide work carried out by **T.Kalyan Babu(17H71A0415),Y.Manikanta(17H71A0421),M.HemaSaiAnnapurna(18H75A0405), Md.AbdulRaqib(17H71A0401)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mr.CH. PullaRao)
ProjectSupervisor

(Mr. CH. PullaRao)
Head of theDepartment

(Dr.K.Srinivas)
PRINCIPAL
Principal
Devineni Venkata Ramana & Dr Himasekhar
MIC College of Technology
Kanchikacherla, Krishna District

Examiner1

Examiner2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"AES BASED IMAGE ENCRYPTION AND DECRYPTION "

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

N.Vardhan (17H71A04G9)

K.Jyothsna Sai (17H71A0413)

P.Nathan (17H71A0431)

M.Thanuja (17H71A04B2)

Under the Esteemed Guidance of

Mr.CH.PULLA RAO

Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr Hanu Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

NBA-Accredited

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"AES BASED IMAGE ENCRYPTION AND DECRYPTION"** is a bonafide work carried out by **N.Vardhan (17H71A04G9), P.Nathan (17H71A0431), K.Jyothsna Sai (17H71A0413), M.Thanuja(17H71A04B2)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mr. CH. Pulla Rao)
Project Supervisor

(Mr. CH. Pulla Rao)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"IMAGE BASED PASSWORD AUTHENTICATION FOR ILLITERATES USING TOUCH SCREEN"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS&COMMUNICATION ENGINEERING

Submitted by

R.DURGAMBA	(17H71A0472)	P.DURGA BHAVANI	(18H75A0402)
K.SRI DIVYA	(17H71A04A5)	N.SUDHEER	(17H71A04A6)

Under the Esteemed Guidance of
Mr.B.Radha Krishna Singh,M.Tech, MIE, DCHE..,

Associate Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone : 08678 - 273535, 273622, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

NBA-Accredited

Accredited by NAAC

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: drvhsmiccet@mictech.ac.in, Website: www.mictech.ac.in

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"IMAGE BASED PASSWORD AUTHENTICATION FOR ILLITERATES USING TOUCH SCREEN"** is a bonafide work carried out by **R. DURGAMBA (17H71A0472),P.DURGA BHAVANI.(18H75A0402), K. SRI DIVYA (17H71A04A5) and N. SUDHEER (17H71A04A6)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mr.B.Radha Krishna Singh)
Project Supervisor

(Mr.CH. PullaRao)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"Advanced Military Spying & Bomb Disposal Robot"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

DIVYA GARMIDI (17H71A0470) TEJASRI PERA (17H71A04B1)

NIKHILESHWAR K (17H71A0434) ANVESH VEMPATI (17H71A0404)

Under the Esteemed Guidance of
Mr. TANNERU NARENDRA KUMAR, [illegible]
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
[illegible]
[illegible]
[illegible]

2020-2021

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone : 08678 - 273535, 273023, Fax 08678 - 273560
E-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in
2020-2021

CERTIFICATE

This is to certify that the Main Project entitled **"ADVANCED MILITARY SPYING AND BOMB DISPOSAL ROBOT"** is a bonafide work carried out by **DIVYA GARMIDI (17H71A0470), TEJASRI PERA (17H71A04B1), NIKHILESHWAR KALYANAM (17H71A0434) and ANVESH VEMPATI (17H71A0404)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mr. T NARENDRA KUMAR)
Project Supervisor

(Mr.CH.PULLA RAO)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

KAKINADA

"LoRa BASED INTELLIGENT HOME AUTOMATION SYSTEM"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS AND COMMUNICATION ENGINEERING

Submitted by

Y. ANAND KUMAR	(17H71A04C1)	P. TEJESWI DEVI PRIYA	(17H71A0456)
D.PRADHIEP	(17H71A04E7)	M.SRINIVAS REDDY	(17H71A04G3)

Under the Esteemed Guidance of

Mr. K. VEENANAND, M.Tech.,

Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 – 273569
Email: dotheic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

An ISO 9001:2015 Certified Institute

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone : 08678 - 273535, 273623, Fax 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

An ISO 9001:2015 Certified Institute

CERTIFICATE

This is to certify that the Main Project entitled **"LoRa BASED INTELLIGENT HOME AUTOMATION SYSTEM"** is a bonafide work carried out by **Y.ANAND KUMAR (17H71A04C1), P.TEJESWI DEVI PRIYA (17H71A0456), D.PRADHIEP (17H71A04E7)** and **M.SRINIVAS REDDY (17H71A04G3)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mr. K. VEENENAND)
Project Supervisor

(CH. PULLA RAO)
Head of the Department

(Dr. K. SRINIVAS)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

KAKINADA

"AN IMPROVED VERSION OF ATTENDANCE REGISTRATION SYSTEM BY USING RFID"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS&COMMUNICATION ENGINEERING

Submitted by

K.USHA RANI(17H71A04G7)

T.HARSHAVARDHINI(17H71A0475)

G.HARSHITHA(17H71A0476)

S.N.V.HARIKA PRIYA(18H75A0410)

Under the Esteemed Guidance of
Mr. K. VEENANAND, M.Tech.,
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

DevineniVenkataRamana&Dr.HimaSekhar
MIC College of Technology
proved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist,A.P, India.
Phone : 08678 - 273535, 273621, Fax: 08678 - 273569
E mail: dvrhumic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

DevineniVenkataRamana&Dr.HimaSekhar
MIC College of Technology
(approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E.mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

NBA-Accredited
Accredited by NAAC A With Grade

2020-2021

CERTIFICATE

This is to certify that theMain Projectentitled **"AN IMPROVED VERSION OF ATTENDANCE REGISTRATION SYSTEM BY USING RFID"** is a bonafide work carried out by **K.USHA RANI(17H71A04G7) , T.HARSHAVARDHINI (17H71A0475) , G.HARSHITHA(17H71A0476) and S.N.V.HARIKA PRIYA(18H75A0410)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering**of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mr. K. Veeraanand)
Project Supervisor

(Mr.CH.Pullarao)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

KAKINADA

"IDENTIFICATION OF COVID-19 IN LUNGS BY DEEP LEARNING"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS&COMMUNICATION ENGINEERING

Submitted by

F. ASWANI (17H71A04C5)

B. MUNNI KRISHNA (17H75A0487)

B. PAVANI GANGA (17H71A0488)

M. MOUNICA (18H75A0409)

Under the Esteemed Guidance of
Mr. K. VEENANAND, M.Tech.,
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

DevineniVenkataRamana&Dr.HimaSekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

NBA-Accredited

Accredited by NAAC

DevineniVenkataRamana&Dr.HimaSekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E-mail: [illegible], Website: www.mictech.ac.in

2020-2021

NBA-Accredited

Accredited by NAAC With Grade A

CERTIFICATE

This is to certify that the Main Project entitled **"IDENTIFICATION OF COVID-19 IN LUNGS BY DEEP LEARNING"** is a bonafide work carried out by **P. ASWANI (17H71A04C5), B. MUNNI KRISHNA (17H71A0487), B. PAVANI GANGA (17H71A0488)** and **M. MOUNICA (18H75A0409)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-2021.

(Mr. K. Veenanand)
Project Supervisor

(CH. Pullarao)
Head of the Department

(Dr.K.Srinivas)
Principal

S.T. Mrudula
Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

KAKINADA

"RFID BASED FUEL DISPENING SYSTEM"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

P.JAGADEESH (17H71A04D5)

N. BALAKIRAN(17H71A04C6)

M. NIHARIKA (17H71A0433)

T. ANUSHA (17H71A04C3)

Under the Esteemed Guidance of
Mr. G. KIRAN KUMAR, M.Tech
Assistant professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

NBA-Accredited

Accredited by NAAC

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone: [illegible], Fax: [illegible]
E-mail: [illegible], Website: www.mictech.ac.in

NBA Accredited

Accredited by NAAC A

2020-2021

CERTIFICATE

This is to certify that the Main Project entitled **"RFID BASED FUEL DISPENSING SYTEM"** is a bonafide work carried out by **P. Jagadeesh (17H71A04D5), M. Niharika (17H71A0433), N. Balakiran (17H71A04C6), T. Anusha (17H71A04C3)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-2021

(Mr. G. Kiran kumar)
Project Supervisor

(Mr. CH. PullaRao)
Head of the Department

(Dr. K. Srinivas)
Principal

S. T. Mrudula

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

KAKINADA

"CoCMA: Energy-Efficient Coverage Control in Cluster-Based Wireless Sensor Networks Using a Memetic Algorithm"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

G. Bhavana (17H71A0405)

G. Bhargavi (17H71A04C8)

MD. Shabaaz (17H71A0499)

V.V.V. Koteswararao (17H71A04H0)

Under the Esteemed Guidance of

Mr. G. KIRAN KUMAR, M.TECH.,

ASSISTANT PROFESSOR

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273589
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

NBA-Accredited
B.Tech (CSE | ECE | EEE | ME)

Accredited by
N.A.A.C
With Grade A

An ISO 9001:2015
Certified Institute

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone : 08678 - 273333, 273621, Fax: 08678 - 273369
E mail: dvrhsmic@gmail.com, Website: www.micrec.edu.in
2020-21

NBA-Accredited
Accredited by NAAC A
An ISO 9001:2015 Certified Institute

CERTIFICATE

This is to certify that the Main Project entitled **"CoCMA: Energy efficient coverage control in a cluster based wireless sensor networks using Memetic Algorithm"** is a bonafide work carried out by **G.Bhavana(17H71A0405), G.Bhargavi(17H71A04C8),MD.Shabaaz(17H71A0499),V.V.V.Koteswararao(17H71A4H0)**in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mr.G. Kiran Kumar)
Project Supervisor

(Mr. CH. Pulla Rao)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

KAKINADA

"TEXTURE SYNTHESIS OF IMAGE BY QUILTING METHOD"

Submitted to
Jawaharlal Nehru Technological University
Kakinadain partial fulfillment for the award of the
degree of

BACHELOR OF TECHNOLOGY

In

ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

V.VENKATA REDDY (17H71A0459) **Y.YASHWINI (17H71A04H6)**

R.JAGADEESWAR (18H75A0407) **G.V.L.KAVYA (18H75A0420)**

Under the Esteemed Guidance of
Mr.G.KIRAN KUMAR,MTech.,

Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK,
Kakinada)Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK,
Kakinada)Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dirhamic@mictech.ac.in, Website: www.mictech.ac.in

NBA-Accredited
B.Tech (CSE, ECE, EEE, ME)
Accredited by
NAAC A
With Grade
An ISO 9001:2008
Certified Institute

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"TEXTURE SYNTHESIS OF IMAGE BY QUILTING METHOD"** is a bonafide work carried out by **V.VENKATA REDDY (17H71A0459) , Y.YASHWINI (17H71A04H6), R. JAGADEESWAR (18H75A0407), G.V.L.KAVYA (18H75A0420),** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mr. G. KIRAN KUMAR)
Project Supervisor

(Mr. CH. Pulla Rao)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

KAKINADA

"SMART AND SAFE CHILD RESCUE SYSTEM USING ARDUINO"

Submitted to
Jawaharlal Nehru Technological University Kakinada
In partial fulfillment for the award if the degree of

BACHELOR OF TECHNOLOGY
In
ELECTRONICS AND COMMUNICATION ENGINEERING

Submitted by

R. SAI RAMYA (17H71A0448)

K. BHAVANI (17H71A0467)

G. AJAY KIRITI (17H71A0462)

G. VENKATA YOGANAND (17H71A04H2)

Under the Esteemed Guidance of

Mrs. T. SRIDEVI, M.Tech
Asst. Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERINGS

Devineni Venkata Ramana & Dr.Hanu Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)Kanchikacherla
521180, Krishna Dist, A.P., India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhanu@mictech.ac.in, Website: www.mictech.ac.in

NBA-Accredited

2020-2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada|Kanchikacherla
521180, Krishna Dist., A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273560
E mail: dvrhemlc@mictech.ac.in, Website: www.mictech.ac.in

NBA-Accredited
B.Tech: CSE | ECE | EEE | MEC

Accredited by NAAC With Grade A

2020-2021

CERTIFICATE

This is to certify that the Main project entitled **"SMART AND SAFE CHILD RESCUE SYSTEM USING ARDUINO"** is a bonafide work carried out by **R. SAI RAMYA (17H71A0448), K. BHAVANI (17H71A0467), G. AJAY KIRITI (17H71A0462), G. VENKATA YOGANAND (17H71A04H2)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-2021.

T. Sri Devi

(Mrs. T. SRI DEVI)

Project Guide

Ch. Pulla Rao

(Mr. CH. PULLA RAO)

Head of the Department

K

(Dr. K. Srinivas)

Principal

S. T. Hruduka

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

KAKINADA

"MEDICAL IMAGE AMALGAMATION USING SYNCHRONIZED ANISOTROPIC DIFFUSION MODEL"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

P. NAGA SIRISHA (17H71A04E4)

S. SRI RAMYA KRISHNA (18H75A0414)

T. TARUN DAS (18H75A0416)

D. SATYA BRAHMESH (17H71A0498)

Under the Esteemed Guidance of
Mrs. T. SRIDEVI, M.Tech
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr. Uma Seshu

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist., A.P. India
Phone: 08678 - 273515, 273621, Fax: 08678 - 273569
E mail: dvrhumc@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Accredited

Accredited by

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E-mail: drvhsmictech@gmail.com, Website: www.mictech.ac.in

NBA-Accredited

Accredited by

2020 - 2021

CERTIFICATE

This is to certify that the Main Project entitled **"MEDICAL IMAGE AMALGAMATION USING SYNCHRONIZED ANISOTROPIC DIFFUSION MODEL"** is a bonafide work carried out by, **P Naga Sirisha (17H71A04E4), S Sri Ramya Krishna (18H175A0414), T Tarun Das (18H175A0416), D Satya Brahmesh (17H71A0498)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-2021.

(Ms. T. Sridevi)

Project supervisor

(Mr. CH. Pulla Rao)

Head of the Department

(Dr. K. Srinivas)

Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"Wireless 2 Way E-menu Food Ordering System In Hotels & Restaurants"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

Sai Meghana .H (17H71A0496)	Venkata Sai Mallikarjun .V (17H71A04C0)
Durga Prasad.V (17H71A04D0)	Rashmika .P (17H71A0444)

Under the Esteemed Guidance of
Mrs. T. Sridevi, M.Tech.,

Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & I.N.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist.A.P. India
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist. A.P. India
Phone : 08678 - 273535, 271623, Fax 08678 - 271567
E-mail: devineniventkatramana@mictech.ac.in, Website: www.mictech.ac.in
2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"Wireless 2-way E-menu Food Ordering System In Hotels and Resturants"** is a bonafide work carried out by **Sai Meghana.H(17H71A0496), Venkata Sai Mallikarjun.V(17H71A04C0), Durga Prasad.V (17H71A04D0), Rashmika.P(17H71A0444)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineeringof Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mrs. T. Sridevi)
Project Supervisor

(Mr. CH. Pulla Rao)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"DATA HIDING THROUGH STEGANOGRAPHY USING MODIFIED DWT AND LSB"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

K. KRISHNA PRIYA (17H71A04D9)

V. RAJESH (17H71A0441)

K. SIVA NAGA RANI (17H71A04A1)

G. RAJA (17H71A0440)

Under the Esteemed Guidance of

Mrs. B. RAJYA LAKSHMI, M.Tech

Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 271535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@gmail.com, Website: www.mictech.ac.in

NBA-Accredited

Accredited by NAAC With Grade A

2020-2021

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kandrika Kothuru - 521180, Krishna Dist., A.P., India
Phone: 08678 - 273535, 273633, Fax: 08678 - 275569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

CERTIFICATE

This is to certify that the Main Project entitled **"DATA HIDING THROUGH STEGANOGRAPHY USING MODIFIED DWT AND LSB"** is a bonafide work carried out by **K. KRISHNA PRIYA (17H71A04D9) ,V. RAJESH (17H71A0441), K. SIVA NAGA RANI (17H71A04A1), G. RAJA (17H71A0440)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year **2020-2021**.

(Mrs. B. Rajya Lakshmi)
Project Supervisor

(Mr. CH. Pulla Rao)
Head of the Department

(Dr .K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"ACCIDENT PREVENTION SYSTEM"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

V. SANDHYA RANI (17H71A04F6)
T. AGNI MITHRA (17H71A0402)
CH. HEMANTH KUMAR (18H75A0406)
N. TEJA SRI MALLIK (18H75A0417)

Under the Esteemed Guidance of
Mrs. B. Rajya Lakshmi, M. Tech.,
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P., India
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E-mail: drvhsmic@mictech.ac.in, Website: www.mictech.ac.in

NBA-Accredited

2020-2021

Devineni Venkata Ramana & Dr Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P. India
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

NBA-Accredited
Accredited by NAAC With Grade A

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"Accident Prevention System"** is a bonafide work carried out by **V. Sandhya Rani (17H71A04F6), T. Agni Mithra(17H71A0402), CH. Hemanth Kumar (18H75A0406), N. Teja Sri Mallik (18H75A0417)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-2021

(Mrs. B. Rajya Lakshmi)
Project Supervisor

(Mr. CH. Pulla Rao)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"ANTI-POACHING ALERT SYSTEM FOR VALUABLE TREES SAFETY"

Submitted to
Jawaharlal Nehru Technological University Kakinada in
partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

V. SAI BHAVANI (17H71A04F3)

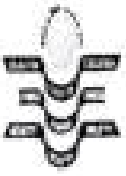
K. ANUSHA (17H71A04C2)

K.MAHENDAR (18H75A0405)

V. RAVI (17H71A0491)

Under the Esteemed Guidance of
Mrs. K. MANASA LAKSHMI, M.Tech
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING



Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College ofTechnology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone : 08678 - 273535, 273623, Fax: 08678 – 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada) Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273621, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

NBA-Accredited
Accredited by
NAAC A

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"ANTI-POACHING ALERT SYSTEM FOR VALUABLE TREES SAFETY"** is a bonafide work carried out by **V.SAI BHAVANI (17H71A04F3),K.ANUSHA (17H71A04C2),K.MAHENDAR(18H75A0408), V.RAVI (17H71A0491)** in partial fulfillment for the award of degree of Bachelor of Technology in **ELECTRONICS AND COMMUNICATION ENGINEERING** OF **JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA** during the year **2020-21.**

(Mrs. K. Manasa Lakshmi)
Project Supervisor

(Mr. CH. Pulla Rao)
Head of the Department

(Dr. K. Srinivas)

Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"IMPLEMENTATION OF IMAGE STEGANOGRAPHY WITH MODIFIED RSA ALGORITHM"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

A. Manjusha (17H71A0422) **L. Sirisha (17H71A04G0)**

M. Shyam Phani Kumar (18H75A0413) **D. Chandra Kiran (17H71A0469)**

Under the Esteemed Guidance of

Mrs. M. MANASA LAKSHMI, M. Tech.,

Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P. India
[illegible]

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"IMPLEMENTATION OF IMAGE STEGANOGRAPHY WITH MODIFIED RSA ALGORITHM"** is a bonafide work carried out by **A. Manjusha (17H71A0422), L. Sirisha (17H71A04G0), M. Shyam Phani Kumar (18H75A0413), D. Chandra Kiran (17H71A0469)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mrs. M. Manasa Lakshmi)
Project Supervisor

(Mr. CH. Pulla Rao)
Head of Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"FINGERPRINT FORGERY DETECTION USING ADAPTIVE OVERSEGMENTATION"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degreeof

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

N. SANKEERTHANA (17H71A04F7) S. TRILOK KUMAR (17H71A0457)

CH. V.N.KALYAN KUMAR (17H71A04B9) CH. NAGA CHAITRA (17H71A0426)

Under the Esteemed Guidance of

Mrs. M. MANASA LAKSHMI, M. Tech.,

Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

MIC College of Technology
(Approved by AICTE & Permanently Affiliated
to JNTUK, Kakinada)Kanchikacherla -521180
Krishna Dist. A.P, India

2020-2021

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P. India
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
Email: [illegible], Website: www.mictech.ac.in

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"FINGERPRINT FORGERY DETECTION USING ADAPTIVE OVERSEGMENTATION"** is a bonafide work carriedout by **N.SANKEERTHANA (17H71A04F7), S.TRILOK KUMAR (17H71A0457), CH.V.N.KALYAN KUMAR (17H71A04B9), CH. NAGA CHAITRA (17H71A0426)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mrs. K. Manasa Lakshmi)
Project Supervisor

(Mr. CH. Pulla Rao)
Head of Department

(Dr. K. Srinivas)
Principal

S.T. Mridula
Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"FAST AND ROBUST FRAME EXTRACTION METHOD FROM VIDEOS"

Submitted to

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

K.YASHWANTH (17H71A04H5)

T.SRI SAI VEENA TARANGINI (17H71A04A4)

S. TRIVENI (18H75A0418)

N. ASHOK KUMAR (17H71A0465)

Under the Esteemed Guidance of

Mrs. S. T. MRUDULA,

M.Tech, (PH.D)

ASSISTANT PROFESSOR

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada) Kanchikacherla - 521180, Krishna Dist., A.P, India.

Phone: 08678 - 273535, 273623, Fax: 08678 - 273569

E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada) Kanchikacherla - 521180, Krishna Dist, A.P,India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in
2020 - 2021

CERTIFICATE

This is to certify that the Main Project entitled with **"FAST AND ROBUST FRAME EXTRACTION METHOD FROM VIDEOS "** is a bonafide work carried out by **K.Yashwanth (17H71A04H5), T.Sri Sai Veena Tarangini (17H71A04A4), S.Triveni (18H75A0418), N.Ashok Kumar (17H71A0465)** in partial fulfillment for the award of degree of Bachelor of Technology in Electronics and Communication Engineering of Jawaharlal Nehru Technological University Kakinada during the year 2020-21.

S. T. Mrudula
(Mrs. S. T. Mrudula)
Project Supervisor

(CH. Pulla rao)
Head of the Department

(Dr. K. Srinivas)
Principal

S. T. Mrudula
Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"ELECTRONIC WATER SUPPLY AND ELECTRICITY MANAGEMENT SYSTEM FOR EDUCATIONAL INSTITUTIONS"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

K.KRISHNASRI(17H71A0481)

P.JYOTHSNA(17H71A04D7)

M.SAIDURGA (18H75A0412)

K.RESHMA(17H71A04F1)

Under the Esteemed Guidance of
Ms. S.T.MRUDULA MTECH(Ph.d)
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Devineni Venkata Ramana & Dr Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone : 08678 - 273535, 273623, Fax: 08678 - 273369
E-mail: dvrhimic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone : 08678 - 273535, 273623, Fax 08678 - 273569
E-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-21

NBA-Accredited

Accredited by NAAC with Grade A

CERTIFICATE

This is to certify that the Main Project entitled **"ELECTRONIC WATER SUPPLY AND ELECTRICITY MANAGEMENT SYSTEM FOR EDUCATIONAL INSTITUTIONS"** is a bonafide work carried out by **K.KRISHNASRI(17H71A0481), P.JYOTHSNA(17H71A04D7), M.SAIDURGA(18H75A0412), K.RESHMA(17H71A04F1)** in partial fulfillment for the award of degree of Bachelor of Technology in **Electronics and Communication Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

S.T. Mrudula

(Ms.S.T.Mrudula)
Project Supervisor

(Mr. CH. Pulla Rao)
Head of the Department

(Dr. K. Srinivas)
Principal

S.T. Mrudula

Examiner 1

Examiner 2

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

Kakinada

FACE MASK DETECTION

A Project Report submitted to

Jawaharlal Nehru Technological University Kakinada in

partial fulfilment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

By

P. Venkateswari (17H71A0556)

S. Bhanu Prasad (17H71A0502)

P. Karuna Jyothi (17H71A0520)

P.Sai Teja (17H71A0538)

A. Bindu Kumari (17H71A0504)

Under the Esteemed guidance of

Mrs. Ch. Sabitha

Associate Professor

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar

MIC College of Technology

an ISO 9001:2015 Certified Institute

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)

Kanchikacherla - 521180, Krishna Dist, A.P, India.

Phone: 08678 - 273535, 273623, Fax: 08678 - 273569

E mail: dvrhmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS

NBA-Accredited B.Tech (CSE|ECE|ME)

Accredited by NAAC With Grade A

2020-21

I

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC A
With Grade

CERTIFICATE

This is to Certify that main project entitled **"Face Mask Detection"** is a bonafied work carried out by P. Venkateswari (17H71A0556), P. Karuna Jyothi (17H71A0520), S. Bhanu Prasad (17H71A0502), P. Sai Teja (17H71A0538), A. Bindu Kumari (17H71A0504) in partial fulfilment for the award of degree of Bachelor of Technology in **Computer Science and Engineering** of **Jawaharlal Nehru Technological University, Kakinada** during the year 2020-21.

Ch. Sabitha — **D. Prasad**

Project Guide — **Head of the Department**

Dr K Srinivas

Examiner — **PRINCIPAL**

ACKNOWLEDGEMENT

We would like to take this opportunity to express my deepest appreciation to the following people for their valuable contributions and assistance with this Project.

Initially, we would like to thank my project supervisor, **Ch. Sabitha, Associate professor, Department of Computer Science and Engineering** for the guidance and support, especially for the valuable ideas and knowledge shared to me throughout the Project.

We have the immense pleasure in expressing my thanks and deep sense of gratitude to**, Mr. D. Prasad, Head of the Department, Computer Science and Engineering** for extending necessary facilities and support for the completion of the Project.

We whole sincerely acknowledge **Dr. K. Srinivas, Principal and Prof D. PANDURANGA RAO, CEO** for giving opportunity to take up the Project work. We also extend my thanks to all faculty members of **Computer Science and Engineering**, for their valuable guidance and encouragement.

We would like to extend my warm appreciation to all my friends for sharing us their knowledge, valuable contributions and help with this Project.

Finally, my special thanks go to my family for their continuous support and help throughout my academic years and for their continual support and encouragement for the completion of the project.

DECLARATION

We **P. Venkateswari, P. Karuna Jyothi, S. Bhanu Prasad, P. Sai Teja, A.Bindu Kumari** of the Main-Project "**FACE MASK DETECTION**", hereby declare that the matter embodied in this Project is the genuine work done by us and has not been submitted either to this university or to any other university/institute for the fulfilment of the requirement of any course of study.

.

P. Venkateswari (17H71A0556)

P. Karuna Jyothi (17H71A0520)

S. Bhanu Prasad (17H71A0502)

P. Sai Teja (17H71A0538)

A. Bindu Kumari (17H71A0504)

ABSTRACT

COVID-19 has affected the world badly. Studies have proved that wearing a face mask is one of the precautions to reduce the risk of viral transmission. And many public places as well as public service providers require customers to use the service and place only if they wear mask correctly. So, it is not possible to manually track the customer, whether they have the mask or not. That's why this technology holds the key here. In this paper, we propose face mask detection using image processing which is one of the high-accuracy and efficient face mask detector. This proposed system mainly consists of three stages i.e. 1. Image preprocessing 2. Face detection and crop 3. Face mask classifier. Our system is capable of unmasked faces and send them fine alert message to their mobile number. The proposed system also identifies the social distance among the persons and alert them to maintain social distance. This system will help to tack safety violations, promote the use of face masks and it ensure a safe working environment.

CONTENTS

1 INTRODUCTION

1.1 OVERVIEW OF THE PROJECT

The trend of porting face mask publically is rising because of the Covid-19 epidemic everywhere in the world. Because Covid-19 people wont to wear mask to shield their health from air pollution. Whereas other are self-conscious concerning their looks, they hide their emotions from the general public by activity their faces. Somebody treated the wearing face masks works on hindering Covid-19 transmission. Covid-19 is that the last epidemic virus that hit the human health within the last century. In 2020, the fast spreading of Covid-19 has forced the WHO to declare Covid-19 as international pandemic. Quite 5 million cases were infected by Covid-19 in not up to half dozen month across 188 countries. The virus spreads through shut contact and in packed and overcrowded areas. The corona virus epidemic has given rise to a unprecedented degree of worldwide scientific cooperation. Computer science supported machine learning and deep learning will facilitate to fight Covid-19 in several ways. Machine learning evaluate huge quantities of knowledge to forecast the distribution of Covid-19, to function early warning mechanism for potential pandemics, and classify vulnerable population. Folks are forced by laws to wear face masks publically many countries. These rules and law we have a tendency yore developed as associate degree action to the exponential growth in cases an deaths in several areas. However, the method observation massive teams of individuals is changing into a lot of difficult. The monitoring process involves the finding of anyone who isn't sporting a face mask. Here we introduce a mask face detection model that's supported machine learning and image process techniques. The planned model may be detect the mask with image and real time detection people wearing mask or not wearing a mask. The model is integration between deep learning and classical machine learning techniques with OpenCV, TensorFlow and Keras. We have a tendency to introduce a comparison between them to seek out the foremost appropriate algorithm program that achieved the very best accuracy and consumed the smallest amount time within the method of coaching and detection.

1.2 FEASIBILITY STUDY

In this we study various feasibilities where existing and software equipment were sufficient for completing the project. The economic Feasibility determines whether doing the project is economically beneficial. The outcome of first phase was that the request and various studies are

approved, and it was decided that the project taken up will serve the end user. On developing and implementation this software saves a lot of amounts and sharing of valuable time

- Economical Feasibility
- Technical Feasibility
- Social Feasibility

1.2.1 Economical Feasibility

The study is carried out to check the economic impact that the system will have on the organization. The amount of fund pour in research and development of system is limited. The expenditure is justified.

1.2.2 Technical Feasibility

This is carried out to check the technical feasibility that is, the technical requirements of the system. Any system developed must not have a high demand on available technical recourses. The developed System must have modest requirements and are required for implementing this system. Our project has modest technical requirements.

1.2.3 Social Feasibility

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not be threatened by the system. His/her level of confidence must be increased so that he/she is able to make some constructive criticism which is welcomed.

1.3 SCOPE

Detecting whether people are wearing face masks or not. The detection can be done in images, videos, and webcam feeds. Depending on the results, the system alerts the officials by sounding an alarm so that the police staff or other officials can take the necessary actions against the people. The Scope of the Project The system is a generic system which is not affected by any background disturbance. Hence it has a very wide scope. The system can be used in CCTV cameras in public places like shopping malls, gardens, market area and so on

2 LITERATURE SURVEY

Generally, most of the projects specialize in face construction identity recognition when wearing mask. During this projects, the focus is on recognizing the people that wearing mask, or not help in decreasing the transmission and spreading of covid-19. The scientist has proven that wearing a mask help in minimizing the spreading rate of Covid-19.

The authors developed a face mask wearing condition identification method. They were ready to classify three categories of face mask-wearing. The categories are face mask- wearing, incorrect face mask-wearing and no face mask-wearing. Saber et al have applied the principal component analysis on masked and unmasked face recognition to acknowledge the person. Also, PCA was utilized. The author proposed a way that's used for removing glasses from human frontal faces. The authors used the YOLOv3 algorithm for face detection. YOLOv3 uses Darknet-53 because the backbone. Nizam et al proposed a completely unique GAN-based network, which will automatically remove mask covering the face area and regenerate the image by building the missing hole. The authors presented a system for detecting the presence or absence of a compulsory medical mask within the OR. The general is to attenuate the false positive face detection as possible without missing mask detection so as to trigger alarms just for medical staff who don't wear a surgical mask. Shaik et al used deep learning real-time face emotion classification and recognition. They used VGG-16 to classify seven countenance. Under the present Covid-19 lock-in time, this technique is effective in preventing spread in may use cases. Here are some use cases which will benefit from system.

Airports: the proposed system could also be vital find travelers at airports. theres no mask. The traveler's data are often captured as a video within the system at the doorway. Any passenger who finds no mask will alert the airport authorities send in order that they can act quickly.

Hospital: the proposed system are often integrated with CCTV cameras, and therefore the data are often manage to ascertain if its employees are wearing masks. If you discover some doctors. If the aren't wearing a mask, they're going to receive a reminder to wear a mask.

Office: The proposed system can help to take care of safety standards to stop. The spread of covid-19 or any such airborne disease. If some employees aren't wearing masks, they're going to receive reminders to wear mask. The choice of the system must be supported the simplest performance. So, I'm using the simplest system performance indicators in order that you'll large –scale implementation. The system has been used with the MobileNetV2 classifier:

MobileNetV2: MobileNetV2 is that the latest technology of mobile visual recognition, including classification, object detection and semantic segmentation. The classifier uses deep intelligent

separable convolution, its purpose is to significantly reduce the complexity cost and model size of the network, so it's suitable for mobile devices, or devices with low computing power. In MobileNetV2, another best module introduced is that the reverse residual structure. The nonlinearity within the narrow layer is removed. Maintain because the backbone of feature extraction, MobileNetV2 achieves the simplest performance in object detection and semantic segmentation.

For MobileNetV2 classifier, ADAM optimizer has been applied to see performance:

ADAM: Adam, a stochastic optimization algorithm supported step the target function is predicated on an adaptive estimation of low-order moments. This manner it's computationally efficient and may be executed almost without memory. It's the diagonal of the gradient is rescaled unchanged, which is extremely suitable for the subsequent problems large in terms of knowledge and/or parameters. Hyper parameters are intuitive explain that they typically don't require much adjustment. The empirical results show that Adam it works well in practice and may be compared with other stochastic optimization method.

2.1 EXISTING SYSTEM AND DRAWBACKS

1. One of the issues we encountered while doing the literature survey was that all the systems mentioned are either hardware-based solutions or software solutions which require different equipment based on their respective applications. Our system is a generalized system which can be used at multiple places with the same results and the software used in completely open sourced, hence no expense is involved.
2. The mentioned systems do not give any alert if any rules are broken. Our system gives an alert by sounding an alarm if anyone is found not wearing a mask.

2.2 PROPOSED SYSTEM AND ADVANTAGES

- The proposed system helps to ensure the safety of the people at public places by automatically monitoring them whether or not and individual wears face mask.
- The proposed system is a deep learning solution that uses OpenCV and TensorFlow, to train the model.
- We combine the deep learning MobileNetV2 modal with the SSD framework for a fast and efficient deep learning solution for human detection in video streams.
    - It uses a triangular similarity technique to measure distance between persons detected by camera.

3 SYSTEM ANALYSIS

3.1 OVERVIEW OF SYSTEM ANALYSIS

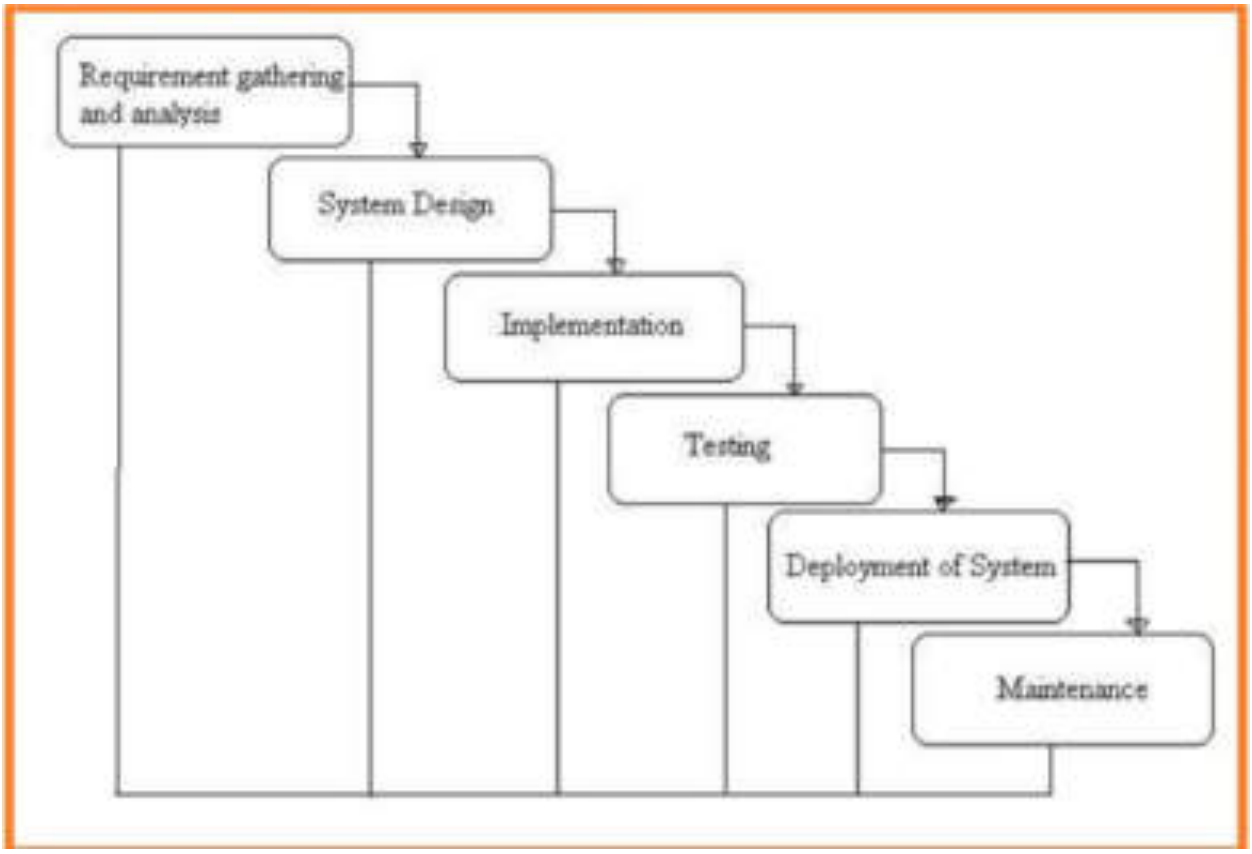


Fig 1: Project SDLC

- Project Requisites Accumulating and Analysis
- Application System Design
- Practical Implementation
- Manual Testing of My Application
- Application Deployment of System
- Maintenance of the Project

3.1.1 Requisites Accumulating and Analysis

It's the first and foremost stage of the any project as our is a an academic leave for requisites amassing we followed of IEEE Journals and Amassed so many IEEE Related papers and final culled a Paper designated by setting and substance importance input and for analysis stage we took referees

from the paper and did literature survey of some papers and amassed all the Requisites of the project in this stage.

3.1.2 System Design

In System Design has divided into three types like GUI Designing, UML Designing with avails in development of project in facile way with different actor and its utilizer case by utilizer case diagram, flow of the project utilizing sequence, Class diagram gives information about different class in the project with methods that have to be utilized in the project if comes to our project our UML Will utilizable in this way The third and post import for the project in system design is Data base design where we endeavor to design data base predicated on the number of modules in our project

3.1.3 Implementation

The Implementation is Phase where we endeavor to give the practical output of the work done in designing stage and most of Coding in Business logic lay coms into action in this stage its main and crucial part of the project

3.1.4 Testing

Unit Testing

It is done by the developer itself in every stage of the project and fine-tuning the bug and module predicated additionally done by the developer only here we are going to solve all the runtime errors

Manual Testing

As our Project is academic Leave we can do any automatic testing so we follow manual testing by endeavor and error methods

3.1.5 Deployment of System

Once the project is total yare we will come to deployment of client system in genuinely world as its academic leave we did deployment i our college lab only with all need Software's with having Windows OS

3.1.6 Maintenance

The Maintenance of our Project is one time process only

3.2 SOFTWARE USED IN THE PROJECT

What things you need to install the software and how to install them:

1. Python 3.6
   - This setup requires that your machine has python 3.6 installed on it. you can refer to this url https://www.python.org/downloads/ to download python. Once you have python downloaded and installed, you will need to setup PATH variables (if you want to run python program directly, detail instructions are below in *how to run software section*). To do that check this: https://www.pythoncentral.io/addpython-to-path-python-is-not-recognized-as-an-internal-or-external-command/.
   - Setting up PATH variable is optional as you can also run program without it and more instructon are given below on this topic.
2. Second and easier option is to download jupyter notebook on command prompt using the command pip install jupyter notebook.
3. You will also need to download and install below packages after you install python and jupyter notebook from the steps above
   - tensoflow
   - keras
   - openCV -seaborn

➢ Use below commands in command prompt to install these packages

pip install tensorflow

pip install numpy

pip install keras

pip install seaborn

Computer Vision

It is a field that include processing analyzinng and understanding image in general high dimensional data from the real world in order to produce numerical and symbolic information or it is a technology of science and machine that see it obtain information from images.

Deep Learning

Deep learning is a powerful set of techniques for learning using neural network. Neural network is beautiful biologically inspired program paradigm which enables a computer to learn from data. These are learning algorithms.

Neural Network

Neural networks reflect the behavior of the human brain, allowing computer programs to recognize patterns and solve common problems in the fields of AI, machine learning, and deep learning. Neural networks, also known as artificial neural networks (ANNs) or simulated neural networks (SNNs), are a subset of machine learning and are at the heart of deep learning algorithms. Their name and structure are inspired by the human brain, mimicking the way that biological neurons signal to one another.

The three important types of neural networks that form the basis for most pre-trained models in deep learning:

1. Artificial Neural Networks (ANN)
2. Convolution Neural Networks (CNN)
3. Recurrent Neural Networks (RNN)

Convolutional Neural Network

Pioneer of Convolutional Neural Network is Yann Lecun.He is the director of Facebook AI Research Group. He Built the first convolutional Neural Network called LeNet in 1988.LeNet was used for character recognition tasks like reading zip codes and digits.
CNN also known as a ConvNet. A convolutional neural network is a feed-forward neural network that is generally used to analyze visual images by processing data with grid-like topology. A convolutional neural network is used to detect and classify objects in an image.In CNN, every image is represented in the form of an array of pixel values.

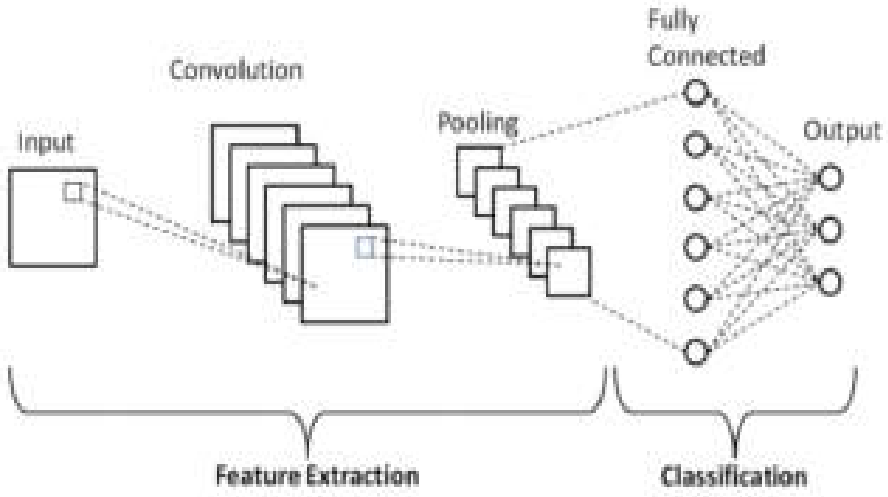
Layers in a Convolutional Neural Network

A convolution neural network has multiple hidden layers that help in extracting information from an image. The four important layers in CNN are:

1.Convolution Layer
2.ReLU Layer

3.Pooling Layer

4.Fully Connected Layer



Input layer:

It Accepts the pixels of the Image as input in the form of arrays.

Hidden Layers:

Carry out features extraction by performing certain calculations & manipulations. These are multiple hidden layers like convolution layer, ReLU layer, Pooling layer etc. that performs features extraction.

Convolution Layer:

This is the first step in the process of extracting valuable features from an image. This layer uses a matrix filter and performs convolution operation to detect patterns in the image. A convolution layer has several filters that perform the convolution operation. Every image is considered as a matrix of pixel values. Consider the following 5x5 image whose pixel values are either 0 or 1.

There's also a filter matrix with a dimension of 3x3.

ReLU Layer:

ReLU stands for the rectified linear unit. Once the feature maps are extracted, the next step is to move them to a ReLU layer. ReLU performs an elementwise operation and sets all the negative pixels to 0. It introduces non-linearity to the network, and the generated output is a rectified feature map. Its activation function is applied to the convolution layer to get a rectified feature map of the image. Below is the graph of a ReLU function:

Pooling Layer:

Pooling is a down-sampling operation that reduces the dimensionality of the feature map. The rectified feature map now goes through a pooling layer to generate a pooled feature map.Also uses multiple filters to detect edges, corners, eyes, features, etc.Different kind of pooling operations can be applied out of which some are explained below.

Fully Connected Layer (FC):

Fully Connected layers in a neural networks are those layers where all the inputs from one layer are connected to every activation unit of the next layer. The output from the convolutional layers represents high-level features in the data. While that output could be flattened and connected to the output layer, adding a fully-connected layer is a (usually) cheap way of learning non-linear combinations of these features. The fully connected layer operates on a flattened input where each input is connected to all neurons.

OpenCV:

OpenCV is a cross-platform library using which we can develop real-time computer vision applications. It mainly focuses on image processing, video capture and analysis including features like face detection and object detection. All the OpenCV array structures are converted to and from Numpy arrays. This also makes it easier to integrate with other libraries that use Numpy such as SciPy and Matplotlib. In this, all OpenCV data types are preserved as such. Later, OpenCV came with both cv and cv2 . Now, there in the latest releases, there is only the cv2 module, and cv is a subclass inside cv2 . You need to call import cv2.cv as cv to access it.

TensorFlow :

TensorFlow is a mathematical computation library for training and building you machine learning and deep learning model with simple to use high level APIs. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML-powered applications. TensorFlow was originally developed by researchers and engineers working on the Google Brain team within Google's Machine Intelligence Research organization to conduct machine learning and deep neural networks research. The system is general enough to be applicable in a wide variety of other domains, as well.

Keras :

Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation. Being able to go from idea to result with the least possible delay is key to doing good research. Features of keras are:

- Allows for easy and fast prototyping (through user friendliness, modularity, and extensibility).
- Supports both convolutional networks and recurrent networks, as well as combinations of the two.
- Runs seamlessly on CPU and GPU.
- Keras is compatible with: Python 2.7-3.8.

3.3 MODULES

Train.py

Predict_webcam.py

3.4 SYSTEM REQUIREMENTS

3.4.1 SOFTWARE REQUIREMENTS

- Software:Python IDLE
- Version:3.6
- Numpy, tensorflow, keras
- operating system :windows

3.4.2 HARDWARE REQUIREMENTS

- Processor: Intel core I5
- Hard Disk :500GB or More
- A simple enhanced windows keyboard
- GSM Module

4. SYSTEM DESIGN

4.1 OVERVIEW OF SYSTEM DESIGN

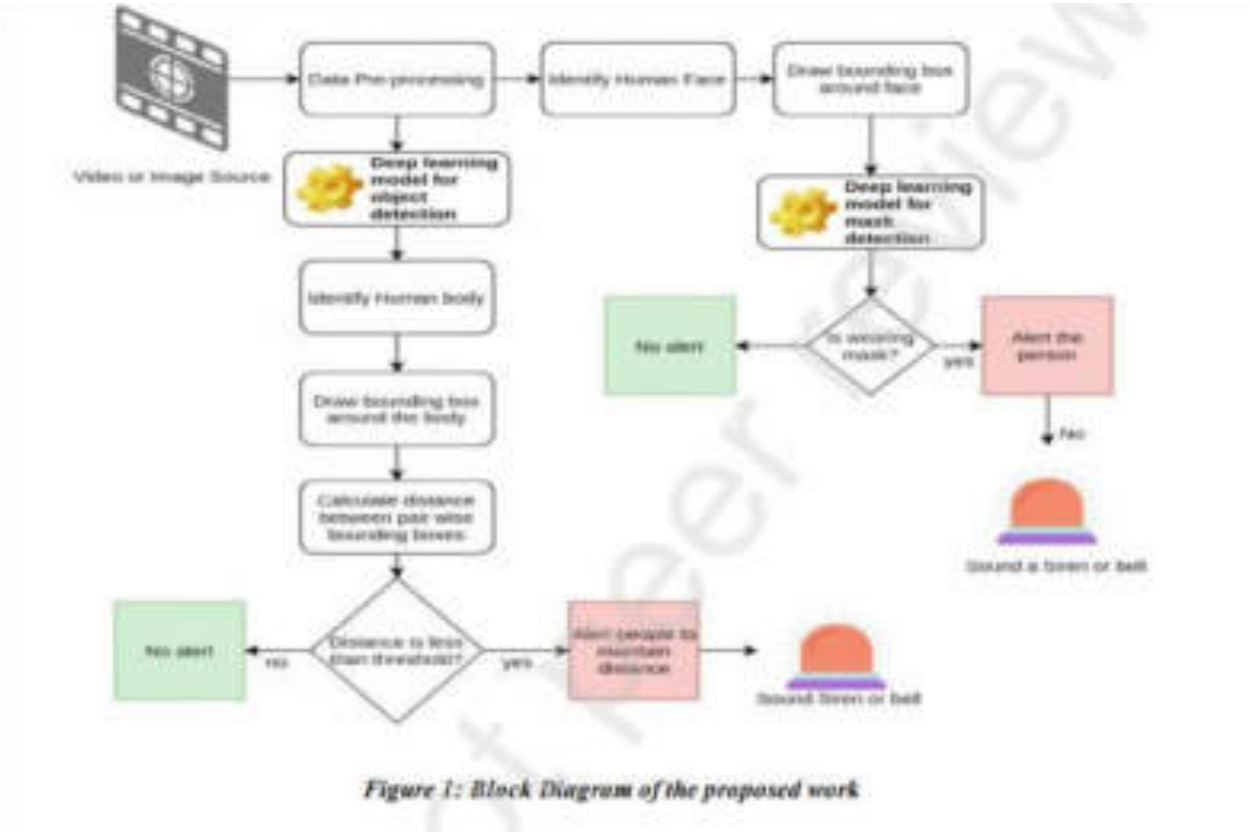


Figure 1: Block Diagram of the proposed work

Fig 2: Architecture

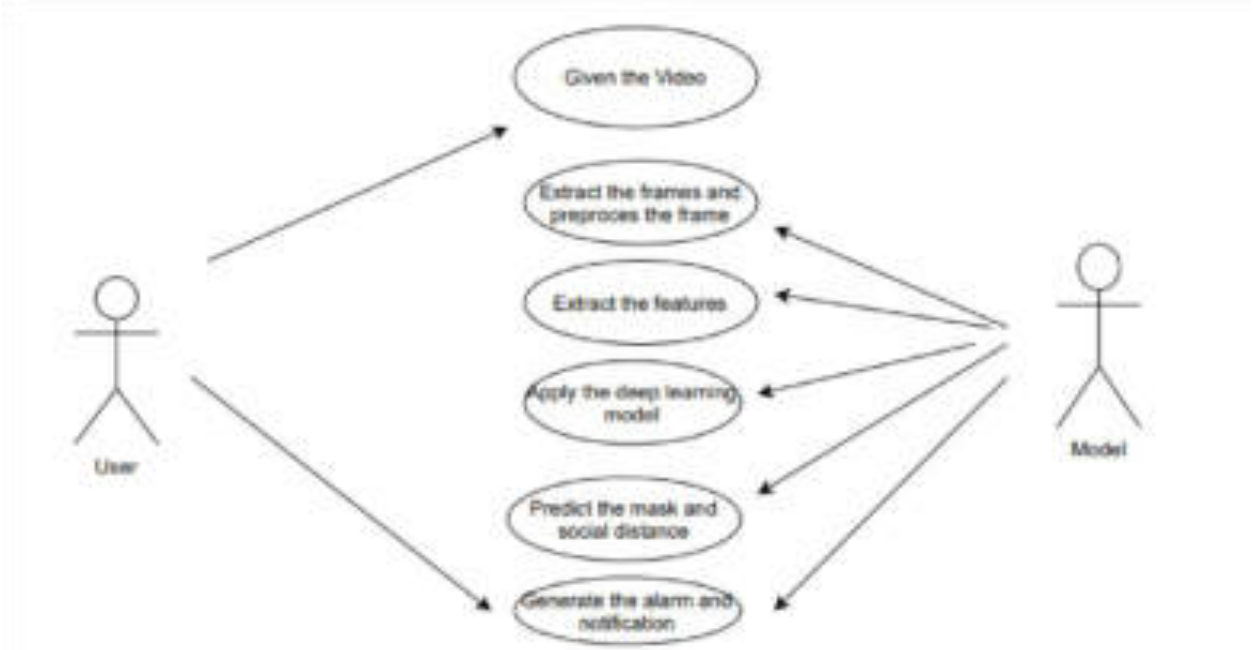
4.2 UML DIAGRAMS

A UML diagram is a diagram based on the UML (Unified Modeling Language) with the purpose of visually representing a system along with its main actors, roles, actions, artifacts or classes, in order to better understand, alter, maintain, or document information about the system. As every diagram need not to be included in our project, we tested out what are the best suited diagrams for our project. Some of them are,

1. Use case Diagram
2. Class Diagram
3. Sequence Diagram
4. Activity Diagram
5. State Diagram

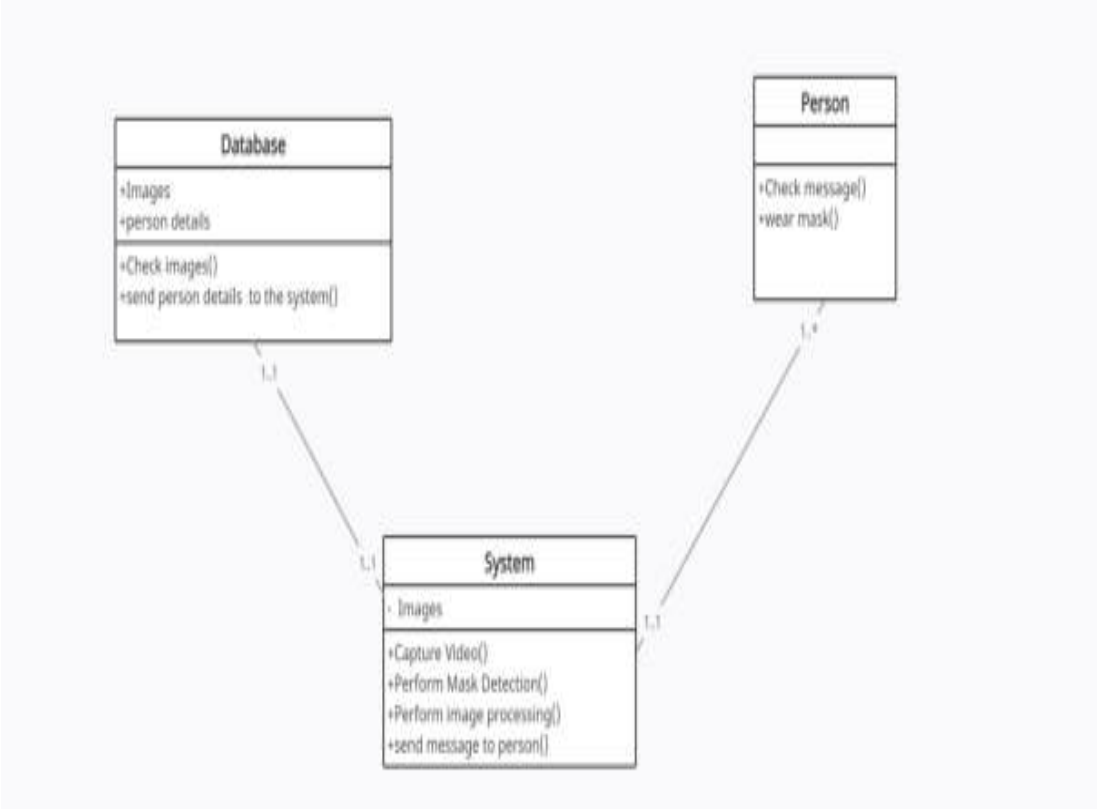
Use Case Diagram:

A usage case outline inside the Unified Modeling Language we used in our Project Development is Star (UML) could be a sort of behavioral chart portrayed out by and produced using a Usecase examination. Its inspiration is to gift a graphical layout of the presence of mind gave by a system to the extent performing specialists, their targets (addressed as use cases), and any conditions between those use cases. The most explanation behind a use case diagram is to show what structure limits are played out that on-screen character. Parts of the entertainers inside the system will be diagram.


Given the Video
Extract the frames and preproces the frame
Extract the features
Apply the deep learning model
Predict the mask and social distance
Generate the alarm and notification
User
Model

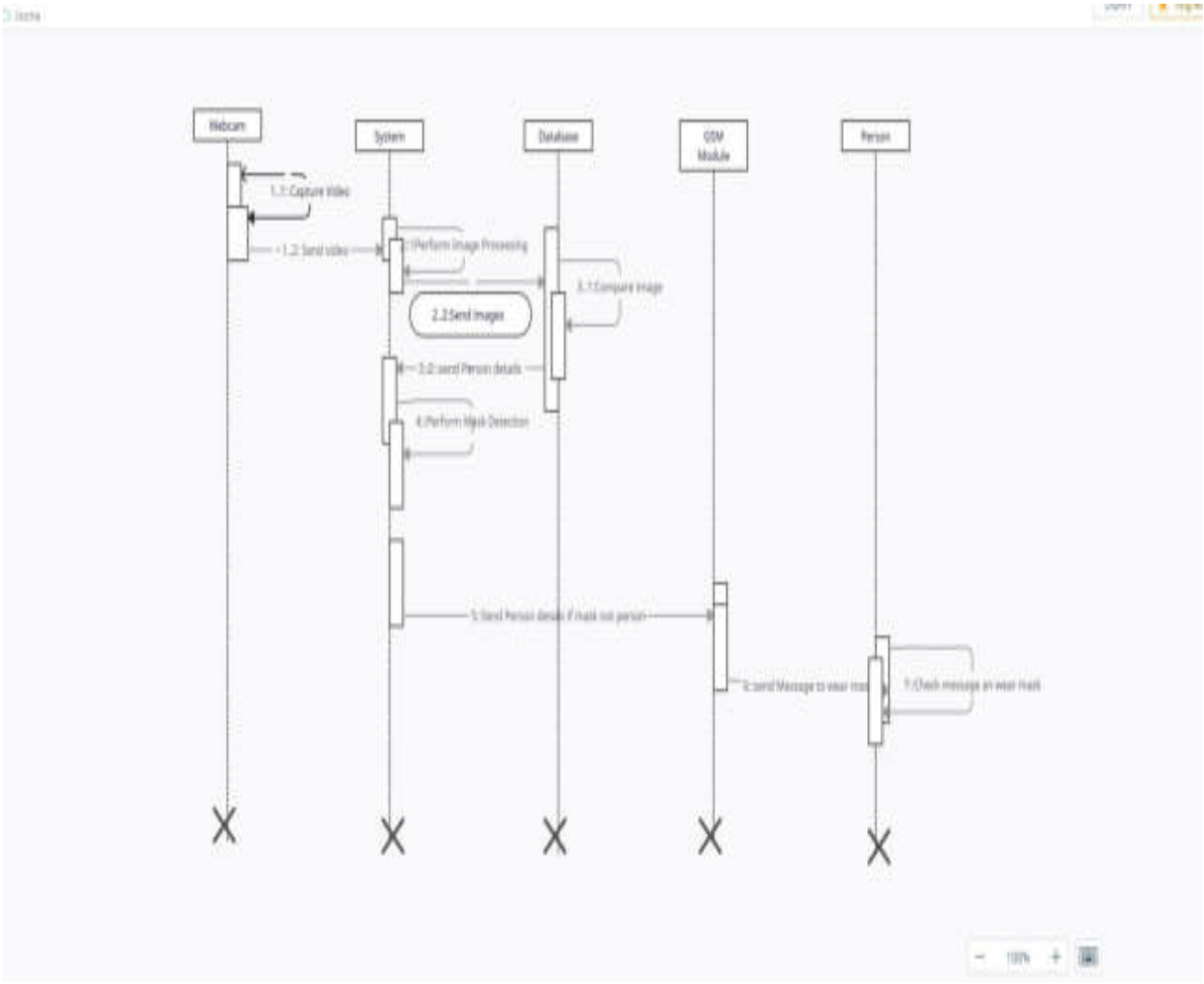
Class Diagram:

In PC code planning, a class plot inside the Unified Modeling Language we used in or Project Development in star (UML) could be a sort of static structure outline the delineates the structure of a system by showing the systems' characterizations, their qualities, operations (or methodologies), and moreover the associations among the groupings. It elucidates that class contains data.



Sequence Diagram:

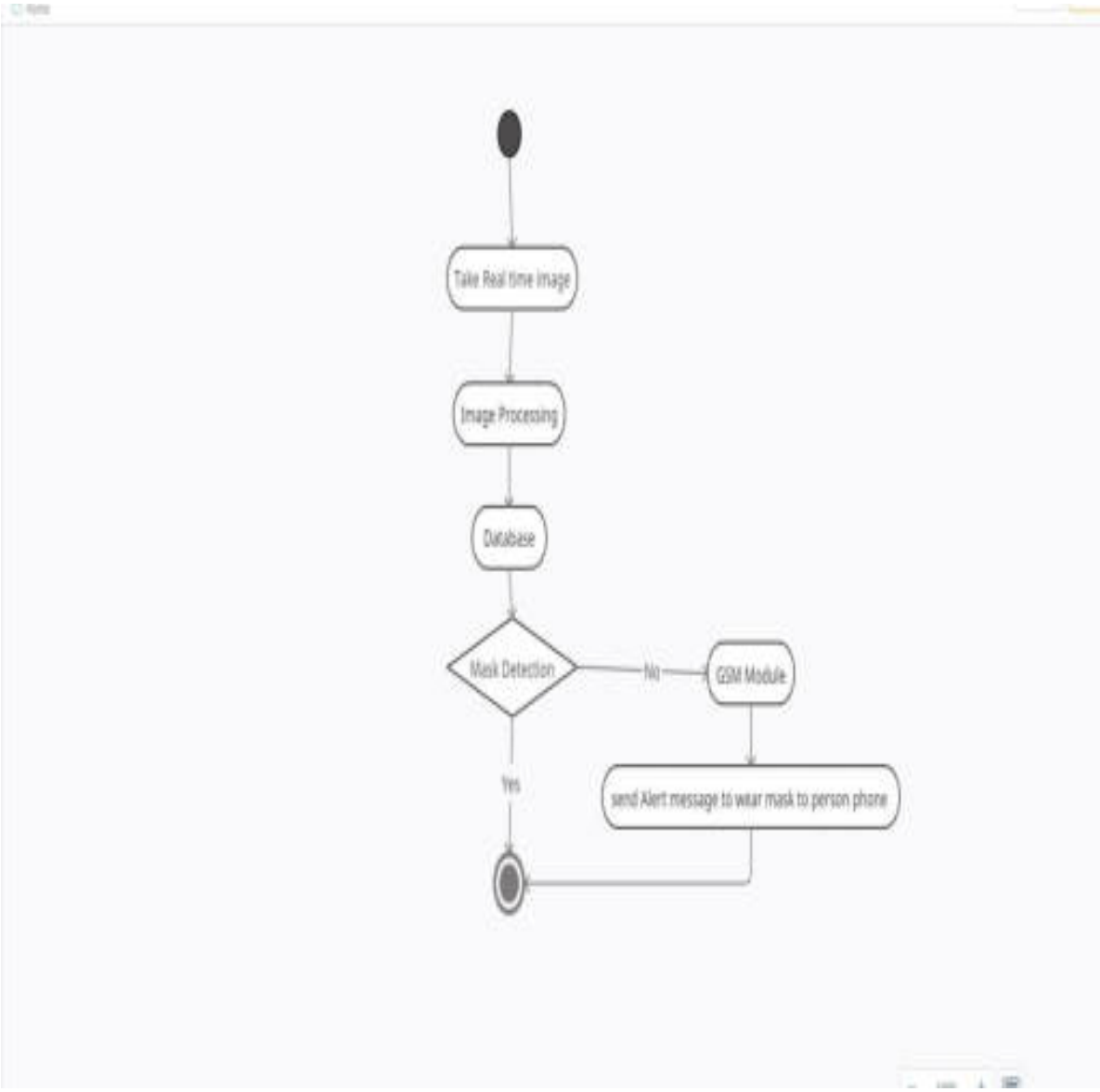
A sequence diagram simply depicts interaction between objects in a sequential order i.e. the order in which these interactions take place. We can also use the terms event diagrams or event scenarios to refer to a sequence diagram. Sequence diagrams describe how and in what order the objects in a system function. These diagrams are widely used by businessmen and software developers to document and understand requirements for new and existing systems.



Activity Diagram:

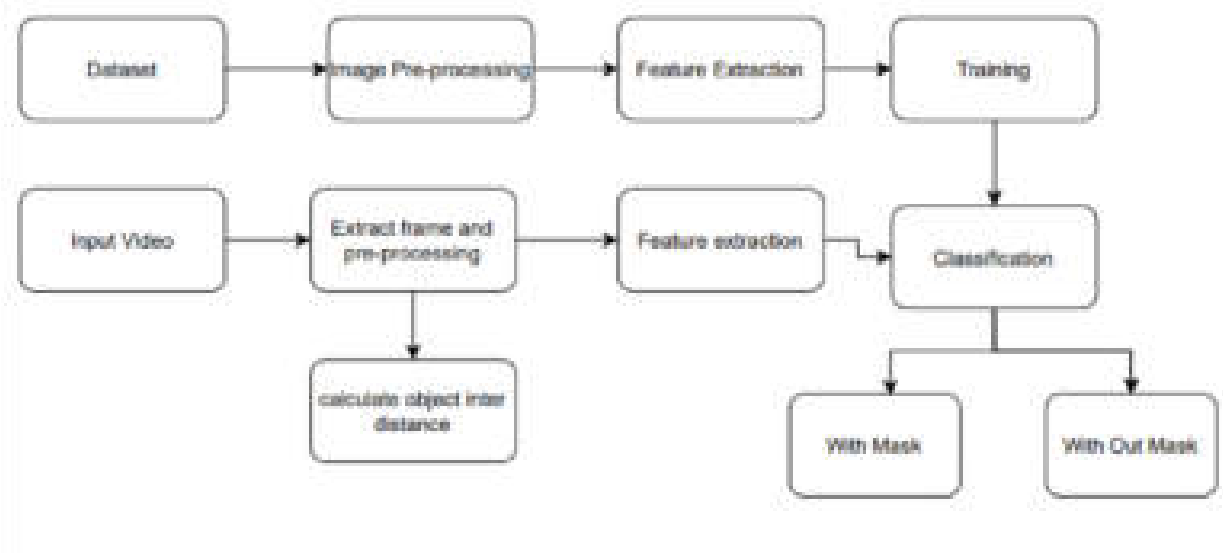
Activity diagram is basically a flowchart to represent the flow from one activity to another activity. The activity can be described as an operation of the system.

The control flow is drawn from one operation to another. This flow can be sequential, branched, or concurrent. Activity diagrams deal with all type of flow control by using different elements such as fork, join, etc.



State Diagram:

A state diagram is used to represent the condition of the system or part of the system at finite instances of time. It's a behavioral diagram and it represents the behavior using finite state transitions. State diagrams are also referred to as State machines and State-chart Diagrams. These terms are often used interchangeably. So simply, a state diagram is used to model the dynamic behavior of a class in response to time and changing external stimuli. We can say that each and every class has a state but we don't model every class using State diagrams.



5 CODING & IMPLEMENTATION

CODE

```
# USAGE check the output mask led need to maintain proper disytance from camera

# python detect_mask_image.py --image examples/example_01.png

# 10 min its complete/ wait for a min// call me once if it is ok for you

# import the necessary packages remove the mask, move out of camera

# ok we will c remaining part by tomorrow morning

from tensorflow.keras.applications.mobilenet_v2 import preprocess_input

from tensorflow.keras.preprocessing.image import img_to_array

from tensorflow.keras.models import load_model

import numpy as np

import argparse

import cv2

import os

from imutils.video import VideoStream

import time

from playsound import playsound

# construct the argument parser and parse the arguments

ap = argparse.ArgumentParser()

#ap.add_argument("-i", "--image", required=True,

#       help="path to input image")
```

```
ap.add_argument("-f", "--face", type=str,
	default="face_detector",
	help="path to face detector model directory")
ap.add_argument("-m", "--model", type=str,
	default="mask_detector.model",
	help="path to trained face mask detector model")
ap.add_argument("-c", "--confidence", type=float, default=0.5,
	help="minimum probability to filter weak detections")
args = vars(ap.parse_args())

# load our serialized face detector model from disk
print("[INFO] loading face detector model...")
prototxtPath = os.path.sep.join([args["face"], "deploy.prototxt"])
weightsPath = os.path.sep.join([args["face"],
	"res10_300x300_ssd_iter_140000.caffemodel"])
net = cv2.dnn.readNet(prototxtPath, weightsPath)

# load the face mask detector model from disk
print("[INFO] loading face mask detector model...")
model = load_model(args["model"])

# load the input image from disk, clone it, and grab the image spatial
# dimensions  side move avvandi once
def SendDataCommand(cmd):
```

```
    message = str(cmd)

    if serialPort.IsOpen():

        serialPort.Send(message)
video_capture = cv2.VideoCapture(0)

time.sleep(2.0)

iii=0


while(1):


        _,image = video_capture.read()

        orig = image.copy()

        (h, w) = image.shape[:2]


        # construct a blob from the image

        blob = cv2.dnn.blobFromImage(image, 1.0, (300, 300),

                (104.0, 177.0, 123.0))


        # pass the blob through the network and obtain the face detections

        print("[INFO] computing face detections...")

        net.setInput(blob)

        detections = net.forward()

        print(len(detections))
        # loop over the detections

        for i in range(0, detections.shape[2]):

                # extract the confidence (i.e., probability) associated with

                # the detection
```

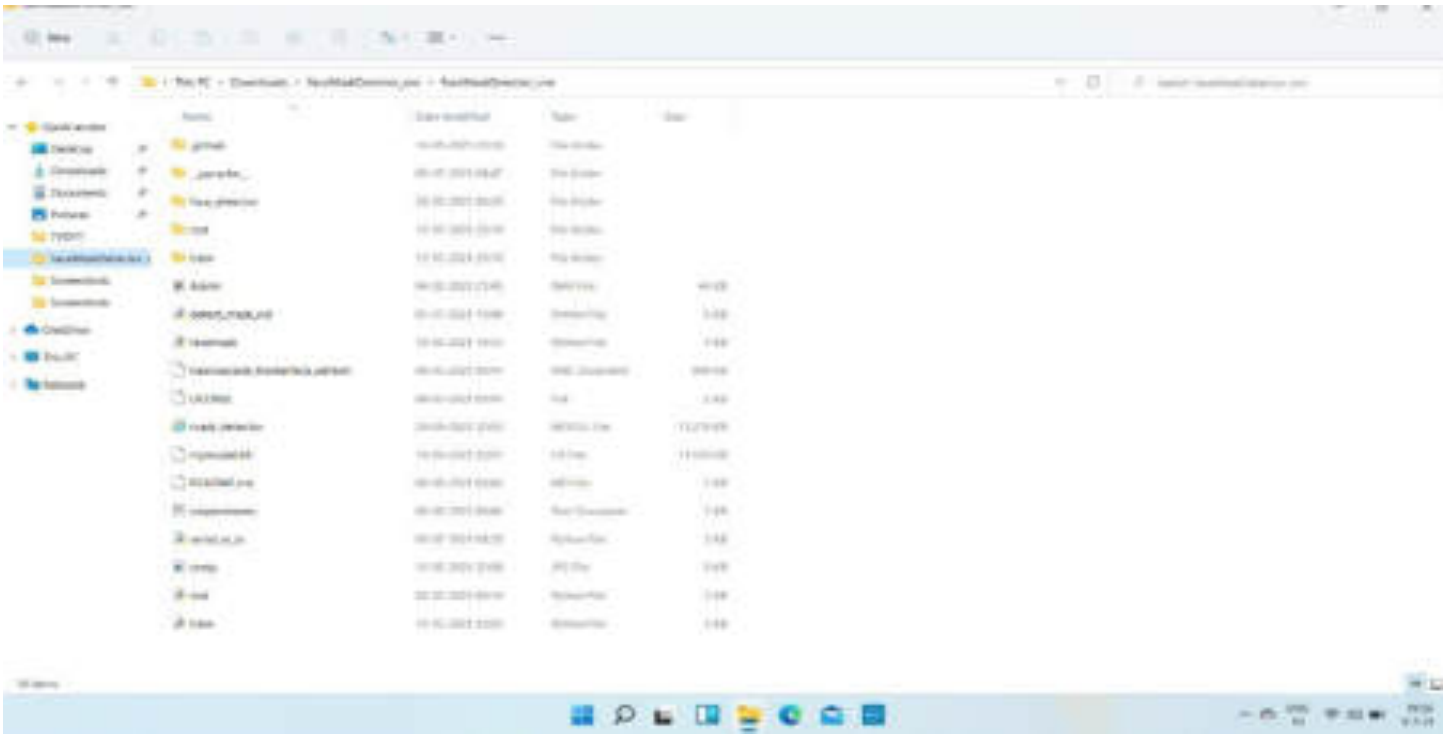
```
confidence = detections[0, 0, i, 2]

# filter out weak detections by ensuring the confidence is
# greater than the minimum confidence
if confidence > args["confidence"]:
    # compute the (x, y)-coordinates of the bounding box for
    # the object
    box = detections[0, 0, i, 3:7] * np.array([w, h, w, h])
    (startX, startY, endX, endY) = box.astype("int")

    # ensure the bounding boxes fall within the dimensions of
    # the frame  side vvu
    (startX, startY) = (max(0, startX), max(0, startY))
    (endX, endY) = (min(w - 1, endX), min(h - 1, endY))

    # extract the face ROI, convert it from BGR to RGB channel
    # ordering, resize it to 224x224, and preprocess it
    face = image[startY:endY, startX:endX]
    face = cv2.cvtColor(face, cv2.COLOR_BGR2RGB)
    face = cv2.resize(face, (224, 224))
    face = img_to_array(face)
    face = preprocess_input(face)
    face = np.expand_dims(face, axis=0)
    # pass the face through the model to determine if the face
```

```
            # has a mask or not

            (mask, withoutmask) = model.predict(face)[0]
            # determine the class label and color we'll use to draw

            # the bounding box and text

            label = "mask" if mask > withoutmask else "No mask"

            color = (0, 255, 0) if label == "mask" else (0, 0, 255)

            # include the probability in the label

            label = "{}: {:.2f}%".format(label, max(mask, withoutmask) * 100)


            # display the label and bounding box rectangle on the output

            # frame

            cv2.putText(image, label, (startX, startY - 10),

                  cv2.FONT_HERSHEY_SIMPLEX, 0.45, color, 2)

            cv2.rectangle(image, (startX, startY), (endX, endY), color, 2)
            if (mask < withoutmask) :

                  iii=iii+1

                  if(iii>5):

                        path = os.path.abspath("Alarm.wav")

                        #playsound(path)
            else:

                  iii=0
# show the output image

cv2.imshow("Output", image)

if cv2.waitKey(1) & 0xFF == ord('q'):

      break
```

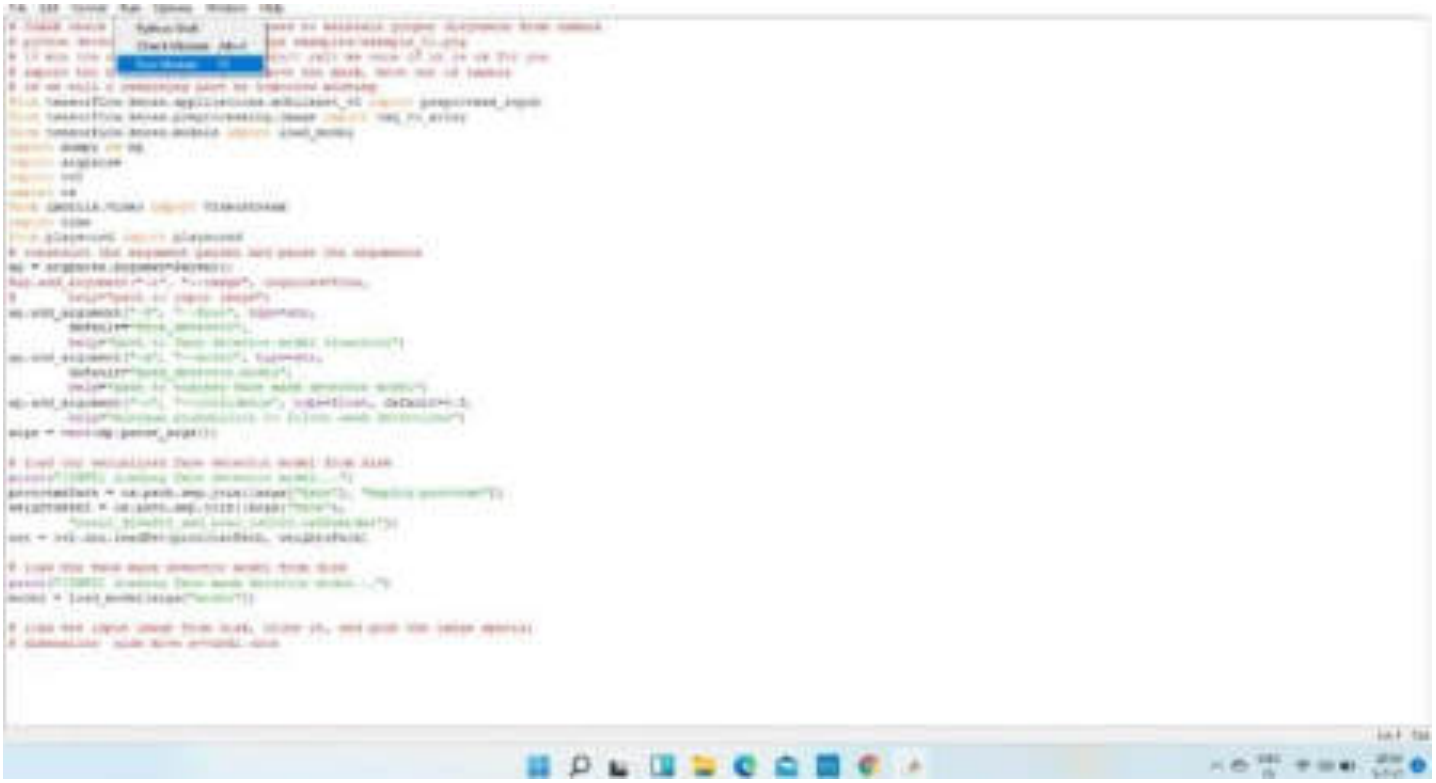
IMPLEMENTATION

This is an important domain that needs special attention during a pandemic. To safeguard the health of doctors and nurses, this system can be implemented to detect whether a patient is wearing a mask. A CCTV camera can be placed to detect mask and social distancing violations which can in turn help reduce crowding while patients are waiting in queues for their appointments. we come up with an efficient way for detecting and notifying officials when a person does not follow the COVID 19 safety protocols in a workplace, business establishments etc. In this work, we have trained a model for face mask detection using TensorFlow and Keras and used mobilenetV2 Object detection for detecting social distancing.

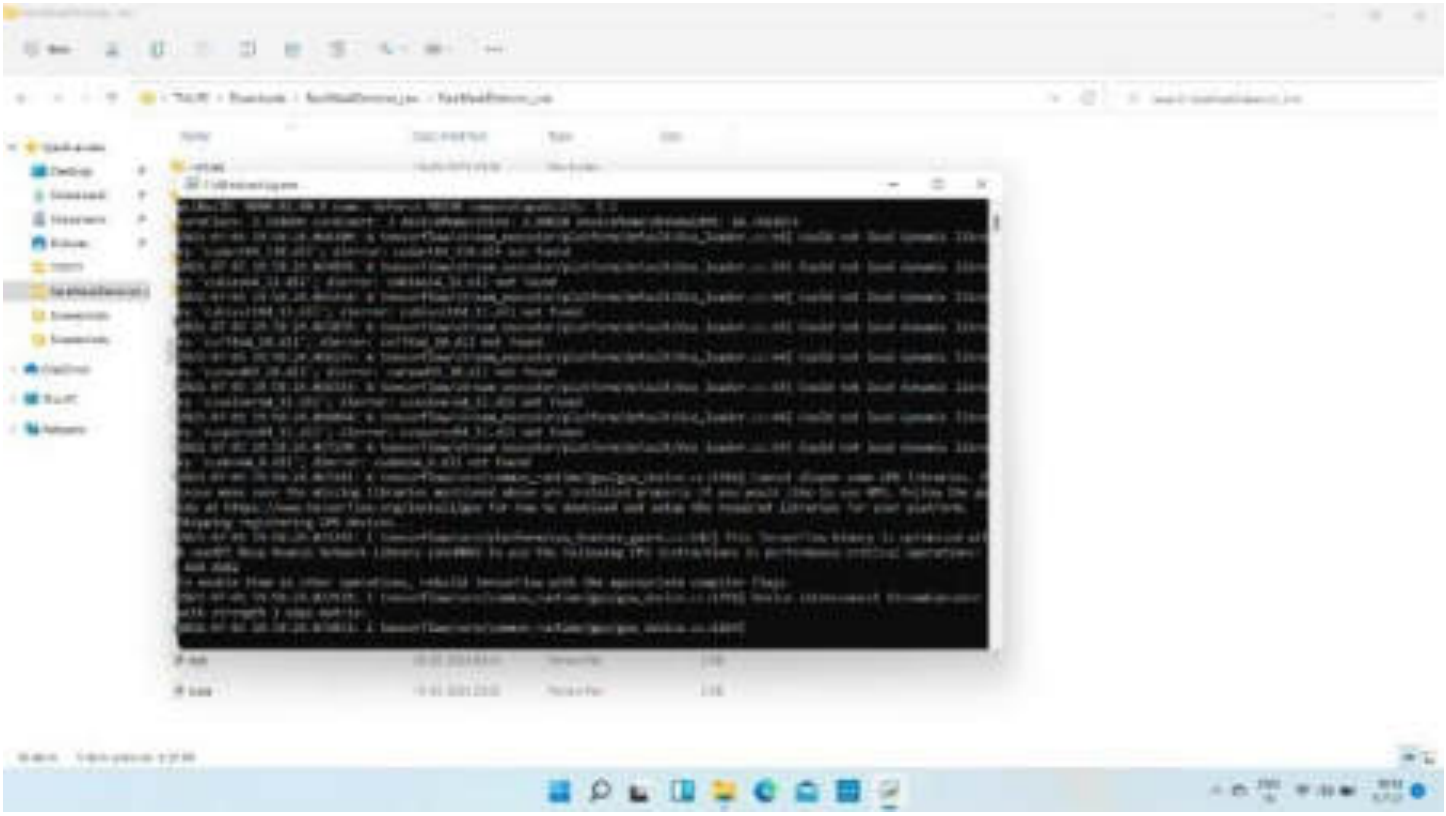
Files:



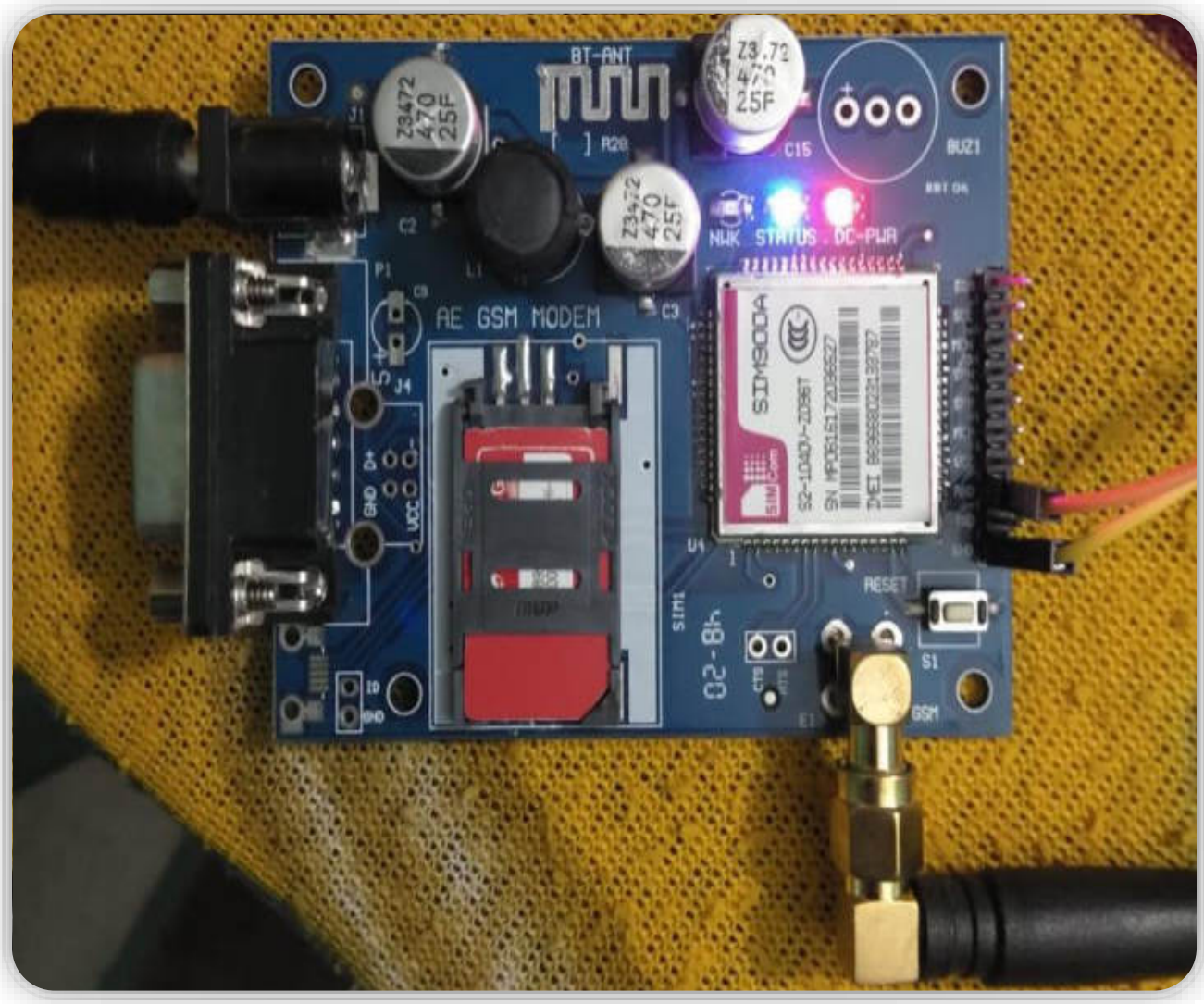
Source code:



Running code:

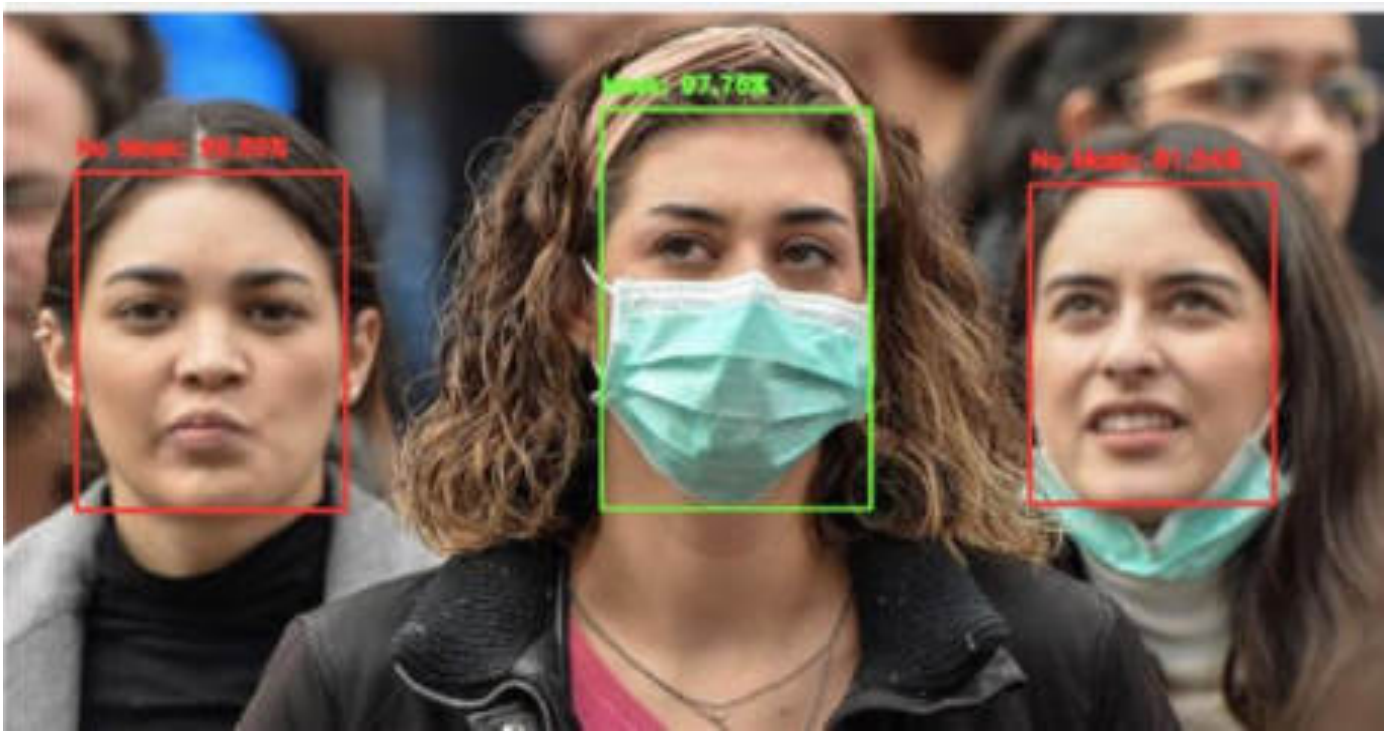


GSM Module:

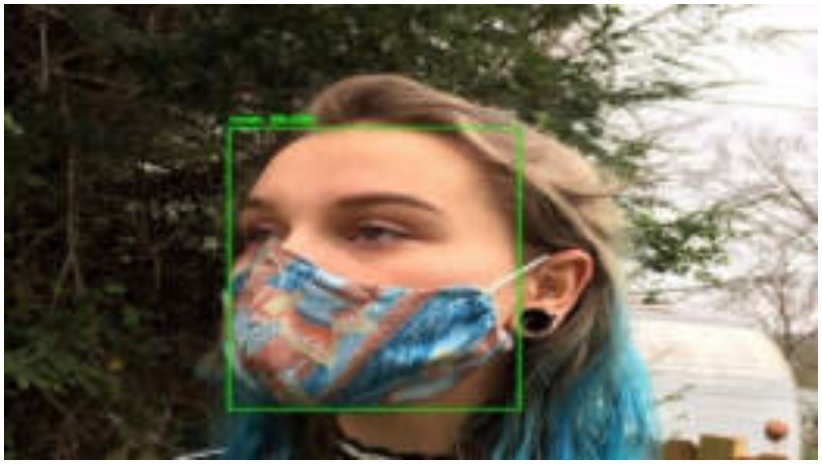


6 OUTPUT

Here is the screenshot of the face mask prediction From Multiple Image:



Here is the screenshot of the face mask prediction From Single Image for Mask with Different Accuracy:



Here is the screenshot of the face mask prediction From Single Image for no Mask with Different Accuracy:

7.SYSTEM TESTING

7.1 OVERVIEW OF TESTING

Software Testing is evaluation of the software against requirements gathered from users and system specifications. Testing is conducted at the phase level in software development life cycle or at module level in program code. Software testing comprises of Validation and Verification.

7.2 TYPES OF TEST

7.2.1 UNIT TESTING

UNIT TESTING is a level of software testing where individual units/ components of a software are tested. The purpose is to validate that each unit of the software performs as designed. A unit is the smallest testable part of any software. It usually has one or a few inputs and usually a single output.

7.2.2 INTEGRATION TESTING

INTEGRATION TESTING is a level of software testing where individual units are combined and tested as a group. The purpose of this level of testing is to expose faults in the interaction between integrated units. Test drivers and test stubs are used to assist in Integration Testing.

8.CONCLUSION

Thus, we have presented an effective and advanced solution during these tough periods we all are facing. This project will be of great help for the govt. officials for keeping an eye on the common public to see whether they are following the rules or not. In turn, it will be beneficial for the common public because they can be safe from the rampant pandemic.

Future Work For the first stage of the face mask detection i.e., face detection, if the input image or video frame is too obscured, the model will not be able to identify human faces. Hence, the mask detector will not work. To overcome this limitation an end to end object detector can be developed

9.REFERENCES

[1] Narinder Singh Punn, Sanjay Kumar Sonbhadra, Sonali Agarwal, "Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques", arXiv:2005.01385v2 [cs.CV], 6 May 2020.

[2] Mahdi Rezaei, Mohsen Azarmi," DeepSOCIAL: Social Distancing Monitoring and Infection Risk Assessment in COVID-19 Pandemic" - MDPI article, published in October 2020.

[3] Shashi Yadav, "Deep Learning based Safe Social Distancing and Face Mask Detection in Public Areas for COVID19 Safety Guidelines Adherence", International Journal for Research in Applied Science Engineering Technology (IJRASET), Volume 8, Issue VII, July 2020.

[4] Dongfang Yang, Ekim Yurtsever, Vishnu Renganathan Keith A. Redmill, Umit Ozguner, "A Vision-based Social Distancing and Critical Density Detection System for COVID-19", arXiv:2007.03578, July 2020.

[5] Kerim Ku¨r¸sat C¸evik, "Computer Vision Based Distance Measurement System using Stereo Camera View", 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), October 2019.

[6] Rinkal Keniya, Ninad Mehendale, "Real-time social distancing detector using SocialdistancingNet19 deep learning network", SSRN preprint, August 2020.

[7] Vincent Chi-Chung Cheng, Shuk-Ching Wong, Vivien Wai-Man Chuang, Simon Yung-Chun So, Jonathan Hon-Kwan Chen, Siddharth Sridhar, Kelvin Kai-Wang To, Jasper Fuk-Woo Chan, Ivan Fan-Ngai Hung, PakLeung Ho, KwokYung Yuen, "The role of community-wide wearing of face mask for control of coronavirus disease 2019 (COVID-19) epidemic due to SARS-CoV2", Journal of Infection, Volume 81, Issue 1, Pages 107-114 July 2020.

[8] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi - "You Only Look Once: Unified, Real-Time Object Detection", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pg. 779-788, 2016.

[9] Huaizu Jiang and Erik Learned-Miller, "Face Detection with the Faster R-CNN", arXiv:1606.03473v1 [cs.CV], 10 Jun 2016.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

Kakinada

DEFENSIVE MODELLING OF FAKE NEWS THROUGH ONLINE SOCIAL NETWORKS

A Project Report submitted to

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING.

By

V.Praneetha(17H71A0533)

K.Trinadh(17H71A0550)

Ch.Sushma(17H71A0547)

P.Vishnu(17H71A0557)

Under the Guidance of

Ms.CH.SABITHA
Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada | ISO 9001:2015 Certified
NBA Accredited (B.Tech-CSE/ECE/MECH) | NAAC Accredited with 'A' Grade
Kanchikacherla-521180, Krishna Dist., A.P., India. Phone: 08678-273535, 273623, Fax: 08678-273569

2019-20

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
An Autonomous Institution
Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada | ISO 9001:2015 Certified
NBA Accredited (B.Tech-CSE/ECE/MECH) | NAAC Accredited with 'A' Grade
Kanchikacherla-521180, Krishna Dist., A.P., India. Phone: 08678-273535, 273623, Fax: 08678-273569

CERTIFICATE

This is to certify that the project work entitled "**Defensive Modelling Of Fake News Detection Through Online Social Network**" is a bonafied work carried out by Ms. V.PRANEETHA (17H71A0533), Mr. K.TRINADH (17H71AO550), Mr. P.VISHNU (17H71A0557), Mr. CH.SUSHMA (17H71A0547) inpartial fulfillment for the award of the degree of Bachelor of Technology in Computer Science and Engineering of Jawaharlal Nehru Technological University, Kakinada during the year 2020-2021. It is certified that all corrections/ suggestions indicated for assessment have been incorporated in the report. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the abovedegree.

ProjectGuide **Head of theDepartment**

Examiner **Principal**

DECLARATION

Hereby we, who carried out the project on "DEFENSIVE MODELLING OF FAKE NEWS DETECTION THROUGH ONLINE SOCIAL NETWORKS", declare that the matter included in this project report is a genuine work done by us and has not, been submitted to this university or any other university/institute for the fulfilment of the requirement of the degree.

V.PRANEETHA(17H71A0533)

K.TRINADH(17H71A0550)

P.VISHNU(17H71A0557)

CH.SUSHMA(17H71A0547)

ACKNOWLEDGEMENT

We are grateful to **Ms.CH.SABITHA, Asst.Professor, Department of Computer Science and Engineering,** whose guidance has been immense all throughout the project work.

Our sincere thanks to **Mr.D.PRASAD, Head of the Department, Computer Science and Engineering,** for her valuable guidance and timely advices in the completion of ourproject.

We express our heartfelt gratitude to **Dr. K.SRINIVAS, Principal** for permitting us to take up our project work and to complete the project successfully.

We would like to express our sincere and heartful thanks to all the lecturers of the department for their continuous cooperation, which has given us the cogency to buildup adamant aspiration over the completion of our project.

Finally we thank one and all who directly and indirectly helped us to complete our project successfully.

ABSTRACT

Online social networks (OSNs) have become an integral mode of communication among people and even nonhuman scenarios can also be integrated into OSNs. The ever growing rise in the popularity of OSNs can be attributed to the rapid growth of Internet technology. OSN becomes the easiest way to broadcast media (news/content) over the Internet. In the wake of emerging technologies, there is dire need to develop methodologies, which can minimize the spread of fake messages or rumors that can harm society in any manner. In this article, a model is proposed to investigate the propagation of such messages currently coined as fake news. The proposed model describes how misinformation gets disseminated among groups with the influence of different misinformation refuting measures. With the onset of the novel coronavirus-19 pandemic, dubbed COVID-19, the propagation of fake news related to the pandemic is higher than ever. In this article, we aim to develop a model that will be able to detect and eliminate fake news from OSNs and help ease some OSN users stress regarding the pandemic. A system of differential equations is used to formulate the model. Its stability and equilibrium are also thoroughly analyzed. The basic reproduction number (R0) is obtained which is a significant parameter for the analysis of message spreading in the OSNs. If the value of R0 is less than one ($R0 < 1$), then fake message spreading in the online network will not be prominent, otherwise if $R0 > 1$ the rumor will persist in the OSN. Realworld trends of misinformation spreading in OSNs are discussed. In addition, the model discusses the controlling mechanism for untrusted message propagation. The proposed model has also been validated through extensive simulation and experimentation.

CONTENTS

LIST OF FIGURES

1. INTRODUCTION

1.1 BACKGROUND

The advent of the World Wide Web and the rapid adoption of social media platforms (such as Facebook and Twitter) paved the way for information dissemination that has never been witnessed in the human history before. Besides other use cases, news outlets benefitted from the widespread use of social media platforms by providing updated newsin near real time to its subscribers. The news media evolved from newspapers, tabloids, and magazines to a digital form such as online news platforms, blogs, social media feeds, and other digital media formats .

It became easier for consumers to acquire the latest news at their fingertips. Facebook referrals account for 70% of traffic to news websites . These social media platforms in their current state are extremely powerful and useful for their ability to allow users to discuss and share ideas and debate over issues such as democracy, education, and health. However, such platforms are also used with a negative perspective by certain entities commonly for monetary gain and in other cases for creating biased opinions, manipulating mindsets, and spreading satire or absurdity. The phenomenon is commonly known as fake news.

1.2 EXISTING SYSTEM

Effective transfer of media from one location to another. With the development of Internet technology, social networks such as Facebook, WhatsApp, Twitter, Instagram, and Google plus have become a vital platform for information exchange. Nowadays, people are connected through online social networks (OSNs) and exchange information in a cost-efficient mannerthrough data transfer. However, information exchanged on OSN platforms may comprise rumors that may affect the social lives of people . Take COVID-19 as an example, where the proliferation of fake news related to the virus has left many people skeptical of any information they read information related to the virus on social media .

Some recent fake news related to a cure for COVID-19 has spread through Facebook [4]. Due to this type of misinformation, people from different corners of the world died. The impact of fake news on people related to a well-known Zika virus case study was presented by Sommariva et al. [5]. The authors found that the speed of fake news spread on OSNs is tremendous and tends to cover large audiences. One major challenge that is associated with OSNs is verification of messages exchanged as well as the authenticity of users. Some messages that are spread through these social networks may create horrible situations regarding peace and harmony in society. Such messages, currently coined as fake news, can also be life-threatening.

The usage of these OSN portals by the criminal group is also increasing rapidly. These users aim to spread false information, thereby creating harmful and damaging situations in the world. Due to such messages, people get affected and panicked. The high penetration rate of social networks into the daily lives of the people has led to another problem of concern. The spread of messages on the social network is very quick and it becomes a challenge to block and remove the untrusted type of messages. Hence, to protect the OSN from this type of activity, there is a need to develop models which can control the rumors and avoid the unforeseen situation in the world.

1.3PURPOSE OF THE PROJECT

There has been a rapid increase in the spread of fake news in the last decade, most prominently observed in the 2016 US elections [5]. Such proliferation of sharing articles online that do not conform to facts has led to many problems not just limited to politics but covering various other domains such as sports, health, and also science [3]. One such area affected by fake news is the financial markets [6], where a rumor can have disastrous consequences and may bring the market to a halt.

Our ability to take a decision relies mostly on the type of information we consume; our world view is shaped on the basis of information we digest. There is increasing evidence that consumers have reacted absurdly to news that later proved to be fake. One recent case is the spread of novel corona virus, where fake reports spread over the Internet about the origin, nature, and behavior of the virus. The situation worsened as more people read about the fake contents online. Identifying such news online is a daunting task.

Fortunately, there are a number of computational techniques that can be used to mark certain articles as fake on the basis of their textual content . Majority of these techniques use fact checking websites such as “PolitiFact” and “Snopes.” There are a number of repositories maintained by researchers that contain lists of websites that are identified as ambiguous and fake .However, the problem with these resources is that human expertise is required to identify articles/websites as fake. More importantly, the fact checking websites contain articles from particular domains such as politics and are not generalized to identify fake news articles from multiple domains such as entertainment, sports, and technology.

The World Wide Web contains data in diverse formats such as documents, videos, and audios. News published online in an unstructured format (such as news, articles, videos, and audios) is relatively difficult to detect and classify as this strictly requires human expertise. However, computational techniques such as natural language processing (NLP) can be used to detect anomalies that separate a text article that is deceptive in nature from articles that are based on facts . Other techniques involve the analysis of propagation of fake news in contrast with real news . More specifically, the approach analyzes how a fake news article propagates differently on a network relative to a true article. The response that an article gets can be differentiated at a theoretical level to classify the article as real or fake. A more hybrid approach can also be used to analyze the social response of an article along with exploring the textual features to examine whether an article is deceptive in nature or not.

1.4CLASSIFICATION OF FAKE NEWS

A number of studies have primarily focused on detection and classification of fake news on social media platforms such as Facebook and Twitter . At conceptual level, fake news has been classified into different types; the knowledge is then expanded to generalizemachine learning (ML) models for multiple domains . The included extracting linguistic features such as n-grams from textual articles and training multiple ML models including K-nearest neighbor (KNN), support vector machine (SVM), logistic regression (LR), linear support vector machine (LSVM), decision tree (DT), and stochastic gradient descent (SGD), achieving the highest accuracy (92%) with SVM and logistic regression.

According to the research, as the number of increased in -grams calculated for a particular article, the overall accuracy decreased. The phenomenon has been observed for learning models that are used for classification. achieved better accuracies with different models by combining textual features with auxiliary information such as user social engagements on social media. The authors also discussed the social and psychological theories and how they can be used to detect false information online. Further, the authors discussed different data mining algorithms for model constructions and techniques shared for features extraction. These models are based on knowledge such as writing style, and social context such as stance and propagation.

1.5DIFFERENT APPROCHES

A different approach is followed The author used textual features and metadata for training various ML models. The author focused mainly on using convolutional neural network (CNN). A convolutional layer is used to capture the dependency between the metadata vectors, followed by a bidirectional LSTM layer. The max-pooled text representations were concatenated with the metadata representation from the bidirectional LSTM, which was fed to fully connected layer with a softmax activation function to generate the final prediction. The research is conducted on a dataset from political domain which contains statements from two different parties. Along with that, some metadata such as subject, speaker, job, state, party, context, and history are also included as a feature set. Accuracy of 27.7% was achieved with combination of features such as text and speaker, whereas 27.4% accuracy was achieved by combining all the different metadata elements with text. A competitive solution is provided, which is a stance detection system that assigns one of four labels to an article, “agree,” “disagree,” “discuss,” or “unrelated,” depending on the conformity of article headline with article text.

1.6OUR CONTRIBUTION

In the current fake news corpus, there have been multiple instances where both supervised and unsupervised learning algorithms are used to classify text. However, most of the literature focuses on specific datasets or domains, most prominently the politics domain .Therefore, the algorithm trained works best on a particular type of article's domain and does not achieve optimal results when exposed to articles from other domains. Since articles from different domains have a unique textual structure, it is difficult to train a generic algorithm that works best on all particular news domains. In this paper, we propose a solution to the fake news detection problem using the machine learning ensemble approach.

Our study explores different textual properties that could be used to distinguish fake contents from real. By using those properties, we train a combination of different machine learning algorithms using various ensemble methods that are not thoroughly explored in the current literature. The ensemble learners have proven to be useful in a wide variety of applications, as the learning models have the tendency to reduce error rate by using techniques such as bagging and boosting . These techniques facilitate the training of different machine learning algorithms in an effective and efficient manner. We also conducted extensive experiments on 4 real world publicly available datasets.

1.7PROBLEMSTATEMENT

The main goal of this project is to detect the FAKE NEWS through online social network in a effective way.

1.8SCOPE OF THE PROJECT

The following steps can help you to effectively define the scope of a project.

Step 1: Identify the project needs:

When you are clearly able to identify the needs of a project, you are more likely set a

Sound bench mark for the beginning.

Understanding the ‘what and why’ of a project will enable you to set specific goals and objectives. It also sets the groundwork for what tasks are to follow and how they are to be performed.

Step 2: Confirm the objectives and goals of the Project

The basis of the project scope should entail your goals and objectives to be one that follows a SMART guideline. That is, to be Specific, Measurable and Achievable. It should also be Realistic and completed within a specificTimeframe.

Specific–This involves stating accurately what the project wants to achieve. That is, what, why and how these will be done. Clarity will reduce the chances of ambiguities and misunderstandings.

Measurable –Are your goals and objectives able to provide feedback and be accountable for?

Achievable –Can your project’s goals and objectives be achieved, given the resources on hand?

Realistic –Are the goals and objectives easy to deliver, especially if you face problems or complications. Will these reduce the overall quality of the project’s outcome and cause running over budget and not meeting the setdeadlines.

Time Frame –Can your project goals and objectives be met within the allocated time frame? Is it a key criterion to meet these deadlines?

Step 3: Project Scope description

You as a leader, need to be clear about the features and functioning required for your product or service.

For example, you are building a website. You need a list that provides how you will build your website, the type of branding required and so on. In other words, what certain qualities will increase achieving your project’ssuccess.

Step 4: Expectations and acceptance

Successful projects are ones that take into account the satisfaction of the end-user. Whether they meet the end-users expectations and accept the product, service or process. The end-users could be your customers or your internal team.

For customers, this includes pricing, value, and quality of products/services as well as availability, delivery and return policies. For employees, this includes the effectiveness and efficiency of new operational processes. Ultimately, your project scope is one that should be attuned to giving better outcomes to whoever your end users may be.

Step 5: Identify constraints

There are always roadblocks to achieving what you were set out to do. When being aware of possible limitations along the way, it can help you minimize problems that may delay or constrain your ability to achieve your project's outcome.

These can be caused by dynamic environmental conditions (internal and external), technological glitches and/or lack of resources. Communicating such problems with your team early on and taking steps to overcome these hurdles will reduce delays in project completion and keep spending within budget. Whether these are based on assumptions or uncertainty, analyzing their impact throughout the projects timeline further reduces the risk of failure.

Step 6: Identify necessary changes

It is always best to avoid reworking the scope of your project, as it means investing in more time, money and resources.

However, at times these changes are inevitable and necessary .Limit changes by taking on the perspectives of customers, stakeholders, and employees involved in the project. This minimizes disagreements later on.

1.9 SURVEY

The creditability of information was defined by many words such as trustworthiness, believability, reliability, accuracy, fairness, objectivity, and other with the same concepts and definitions . There are several researches that use the machine learning approach to calculate the creditability of message. Fake news is the contents that claim people to believe with the falsification, sometimes it is the sensitive messages.

When the messages were received, they will rapidly disperse it to other. The dissemination of fake news in today's digital world has affected beyond a specific group. Mixing both believable and unbelievable information on social media has made the confusion of truth. That is the truth will be hardly classified. However, the appearance of fake news causes great threat on the safety of people's lives and property. There is misinformation (the distributer believes there are true) or disinformation (the distributer knows it is not fact but he intentionally hoax) in fake news proliferation. In this paper, we develop computational resources and models for the task of fake news detection.

We present the construction of a novel dataset covering two different domains. The dataset is collected using a combination of manual and crowdsourced annotation efforts. Using this dataset, we conduct several exploratory analyses to identify linguistic properties that are predominantly present in fake content, and we build fake news detectors relying on linguistic features that achieve accuracies of up to 78%. To place our results in perspective, we also compare the accuracy of our fake news detection models with an empirical human baseline accuracy. Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.

Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

There exists a large body of research on the topic of machine learning methods for deception detection, most of it has been focusing on classifying online reviews and publicly available social media posts. Particularly since late 2016 during the American Presidential election, the question of determining 'fake news' has also been the subject of particular attention within the literature. Conroy, Rubin, and Chen outline several approaches that seem promising towards the aim of perfectly classify the misleading articles . They note that simple content-related n-grams and shallow parts-of-speech (POS) tagging have proven insufficient for the classification task, often failing to account for important context information. Rather, these methods have been shown useful only in tandem with more complex methods of analysis. Deep Syntax analysis using Probabilistic Context Free Grammars (PCFG) have been shown to be particularly valuable in combination with n-gram methods. Feng, Banerjee, and Choi are able to achieve 85%-91% accuracy in deception related classification tasks using online review corpora .

Classification of Alert Here machine learning model will be used that will train and update the data based on feedback and delayed samples. Classifier will be trained separately using feedback and delayed samples and their probabilities will be aggregated to identify alerts. Transaction that will be having high probability will be alerted. Hence only limited number of alerted transaction is reported to investigators.

2.AN HYBRID NURAL NETWORK FOR THE FAKE NEWS DETECTION

The fast spreading of fake news stories on social media can cause inestimable social harm. Developing effective methods to detect them early is of paramount importance. A major challenge of fake news early detection is fully utilizing the limited data observed at the early stage of news propagation and then learning useful patterns from it for identifying fake news. In this article, we propose a novel deep neural network to detect fake news early. It has three novel components:

1) identify the user before joining the network;

2) identify when the user after the spreading rumor in the network.

3) a status-sensitive crowd response feature extractor that extracts both text features and user features from combinations of users' text response and their corresponding user profiles,

4) a position-aware attention mechanism that highlights important user responses at specific ranking positions, and

5) a multi-region mean-pooling mechanism to perform feature aggregation based on multiple window sizes.

2.1 Need of the Hybrid Neural Network

Newspapers are the primary source of news for people worldwide. However, off late, due to the significant growth and updates in technologies, there has been a stupendous rise in the popularity of social media. The number of people who use social media has increased remarkably. As a consequence, social networks such as social media, websites, blogs, etc. have emerged as relevant platforms to gather all kinds of news. People rely more on social networks than newspapers these days. With the availability of the internet, these networks can be accessed easily. This can lead to easy manipulation of the existing news, thereby causing fake news. Fake news can be used as a vital tool to project people in a wrong way. It can spread hate among people which can further harm the society. Hence, it is very necessary to prevent the spread of fake news.

2.2 Algorithm used

Quick technological advancement have authorized newspapers and journalism to be distributed over the web and the rise of Twitter, Youtube, Instagram, Facebook and some other social networking sites.

Networking Sites have become a noteworthy method to speak for people with each other and offer schemes and thoughts. Critical components of a person these networking sites is quick sharing of information. Specifically in this situation, exactness of the news or information distributed is critical. Fake news spreading on different networking sites has become the most concerning issue. Fake news has majorly influenced everyday lives and the social requests of many individuals & caused some negative impacts. Here, the most thorough electronic databases have been broken down to take a greater look at articles about identification of news that is fake on networking sites using an efficient practice of literature review. The fundamental point to study this is revealing the advantages that AI uses for the knowledge about fake news & its victory in one application or the other. Accordingly, assumptions were made that the victory of computerized reasoning gadgets is more than 90%. This is accepted to be a manual for anyone related to this field(researchers and individuals).

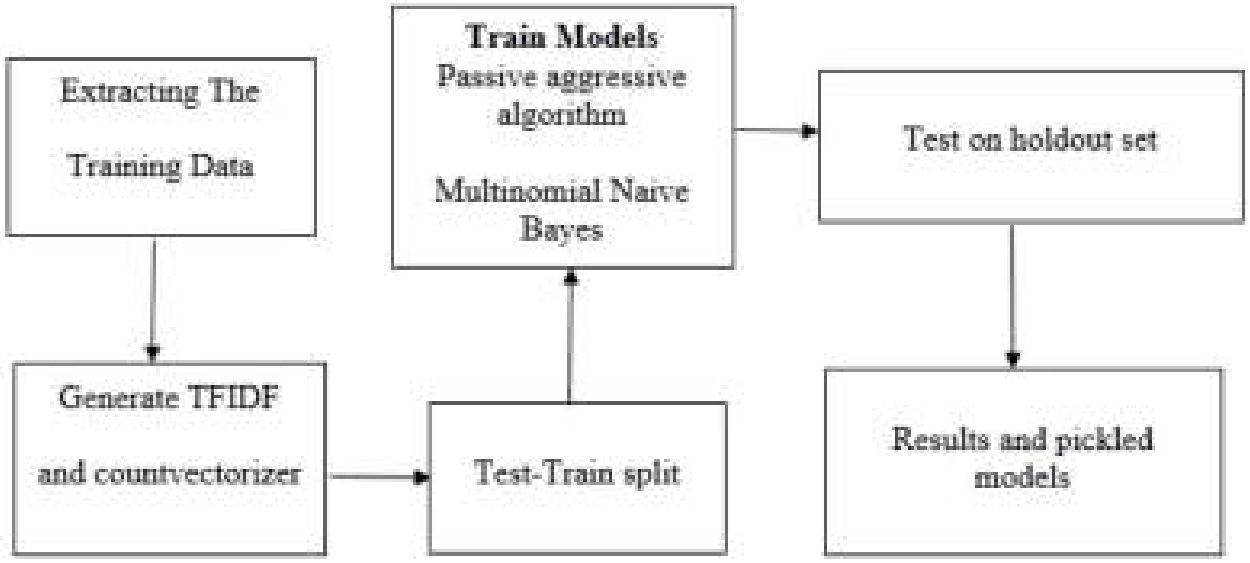


FIG: data flow model

2.3 WHY EXTENSION

A model is build based on the count vectorizer or a tfidf matrix (i.e) word tallies relatives to how often they are used in other articles in your dataset) can help . Since this problem is a kind of text classification, Implementing a HNN will be best as this is standard for text-based processing. The actual goal is in developing a model which was the text transformation (count vectorizer vs tfidf vectorizer) and choosing which type of text to use (headlines vs full text). Now the next step is to

extract the most optimal features for countvectorizer or tfIdf-vectorizer, this is done by using a n-number of the most used words, and/or phrases, lower casing or not, mainly removing the stop words which are common words such as “the”, “when”, and “there” and only using those words that appear at least a given number of times in a given text dataset.

2.4 DRAWBACKS OF EXISTING

- Implemented a semantic analysis looking at 'object:descriptor' pairs for contradictions with the text on top of Feng's initial deep syntax model for additional improvement.
- Rubin, Lukoianova and Tatiana analyze rhetorical structure using a vector space model with similar success.
- Employ language pattern similarity networks requiring a pre-existing knowledge base.
- Hardware dependence.
- Unexplained behavior of the network.
- Determination of proper network structure.

2.5ADVANTAGES OF PROPOSED SYSTEM

- Very High accuracy in image recognition problems.
- Automatically detects the important features without any human supervision.
- Weight sharing.

2.6SOFTWARE OR PLATFORM REQUIRED

- Python
- numpy
- pandas

- itertools
- matplotlib
- sklearn

2.7SYSTEM REQUIREMENTS

H/W SystemConfiguration:-

- Processor - Pentium–IV
- RAM - 4 GB(min)
- Hard Disk - 20 GB
- KeyBoard - Standard Windows Keyboard
- Mouse - Two or ThreeButton Mouse
- Monitor - SVGA

Software Requirements:

- Operating System - Windows XP
- Coding Language - PYTHON

3.DESIGN

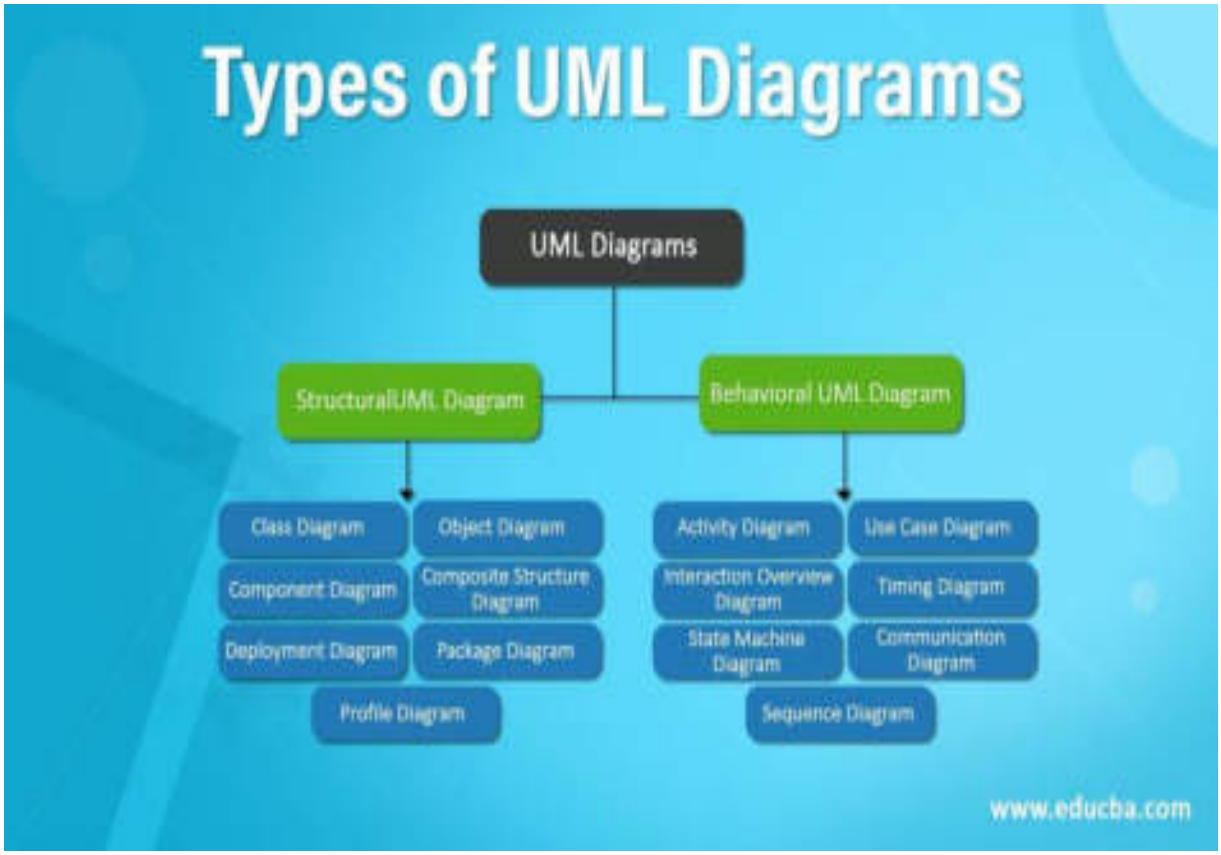
3.1 UML DIAGRAMS

The UML is a standard language that is used to effectively describe a model, using both visual elements and text, just as in your map.

Note that UML is not a programming language. You can rather call it as a pictorial language consisting of symbols, diagrams, text, pseudo-code or anything that describes a software system. Every stakeholder speaks UML in a software engineering process, particularly during the requirements design phase.

A UML model is a representation of a software in terms of it's structure, behavior and interactions, before the actual coding process begins. This model is usually used to describe an object oriented programming approach. Considering the complexity of a software product, a UML model is made up of different types of diagrams. Just as the English language consists of 26 alphabets, the UML model diagrams also consist of some basic **elements** as follows:

Fig: OUTLINE OF UML DIAGRAMS



In this article, we discuss some of the UML diagrams that depicts the structural and behavioural aspects of our proposed system. The following are the required diagrams.

1.Class Diagram
2.Usecase Diagram
3.Sequence Diagram
4.State Diagram
5.Collaboration Diagram

3.1.1CLASS DIAGRAM

In software engineering, a **class diagram** in the Unified Modelling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among objects.

The class diagram is the main building block of object-oriented modelling. It is used for general conceptual modelling of the structure of the application, and for detailed modelling, translating the models into programming code. Class diagrams can also be used for data modelling. The classes in a class diagram represent both the main elements, interactions in the application, and the classes to be programmed.

In the diagram, classes are represented with boxes that contain three compartments:

- The top compartment contains the name of the class. It is printed in bold and centered, and the first letter is capitalized.
- The middle compartment contains the attributes of the class. They are left-aligned and the first letter is lowercase.
- The bottom compartment contains the operations the class can execute. They are also left-aligned and the first letter is lowercase.

A class with three compartments.

In the design of a system, a number of classes are identified and grouped together in a class diagram that helps to determine the static relations between them. In detailed modelling, the classes of the conceptual design are often split into subclasses.

In order to further describe the behaviour of systems, these class diagrams can be complemented by a state diagram or UML machine.

UML provides mechanisms to represent class members, such as attributes and methods, and additional information about them like constructors.

Visibility

To specify the visibility of a class member (i.e. any attribute or method), these notations must be placed before the member's name:

+	Public
−	Private
#	Protected
~	Package

A **derived property** is a property whose value (or values) is produced or computed from other information, for example, by using values of other properties.

A derived property is shown with its name preceded by a forward slash '/'

Scope

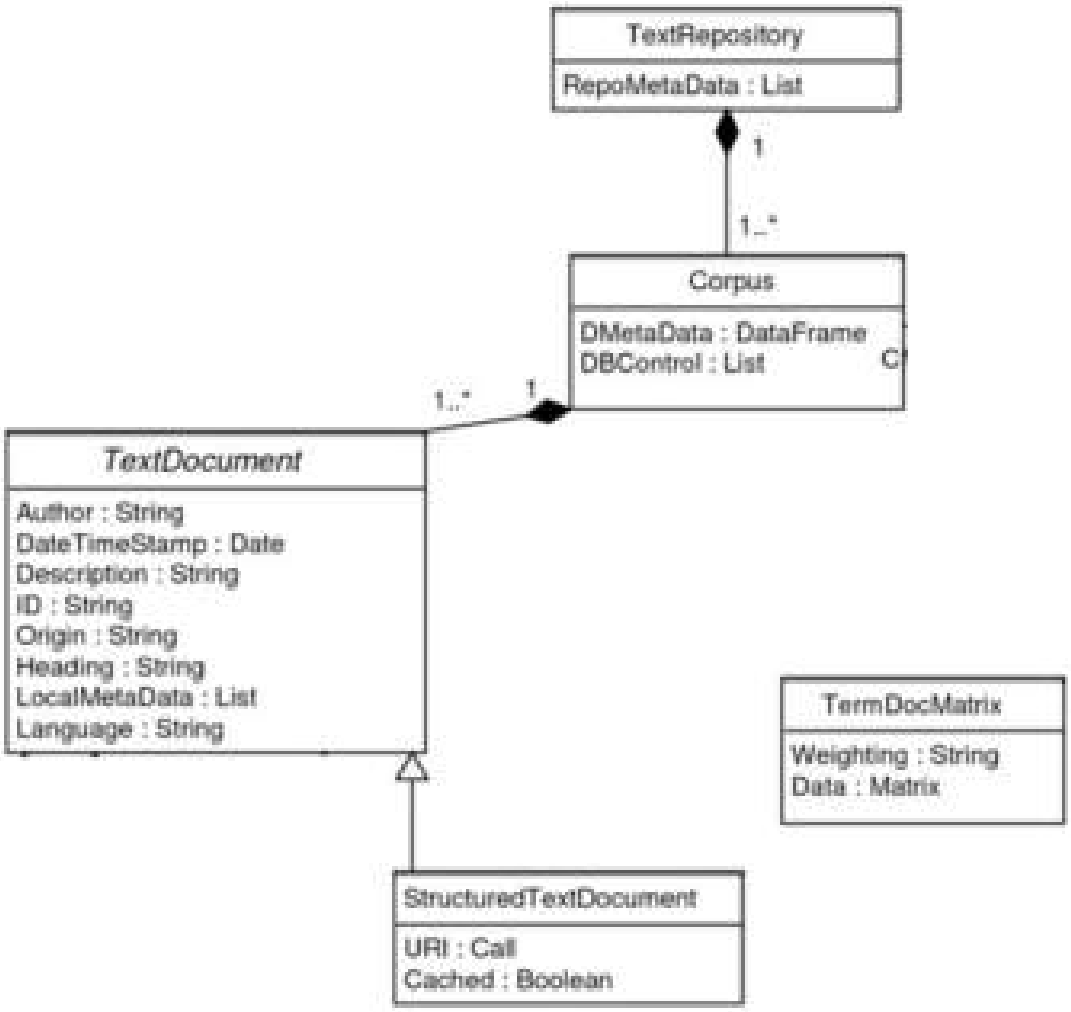
The UML specifies two types of scope for members: *instance* and *classifier*, and the latter is represented by underlined names.

- **Classifier members** are commonly recognized as “static” in many programming languages. The scope is the class itself.
    - Attribute values are equal for all instances

- Method invocation does not affect the classifier's state

- **Instance members** are scoped to a specific instance.
  - Attribute values may vary between instances
  - Method invocation may affect the instance's state (i.e. change instance's attributes)

To indicate a classifier scope for a member, its name must be underlined. Otherwise, instance scope is assumed by default.



3.1.2.USECASE DIAGRAM

A **use case diagram** is a graphical depiction of a user's possible interactions with a system. A use case diagram shows various use cases and different types of users the system has and will often be accompanied by other types of diagrams as well. The use cases are represented by either circles or ellipses. The actors are often shown as stick figures.
The purpose of a use case diagram is to capture the dynamic aspect of a system. They provide a simplified graphical representation of what the system should do in a use case. Further diagrams and documentation are needed for a complete functional and technical outlook on the system.

While a use case itself might drill into a lot of detail about every possibility, a use-case diagram can help provide a higher-level view of the system. It has been said before that "Use case diagrams are the blueprints for your system".

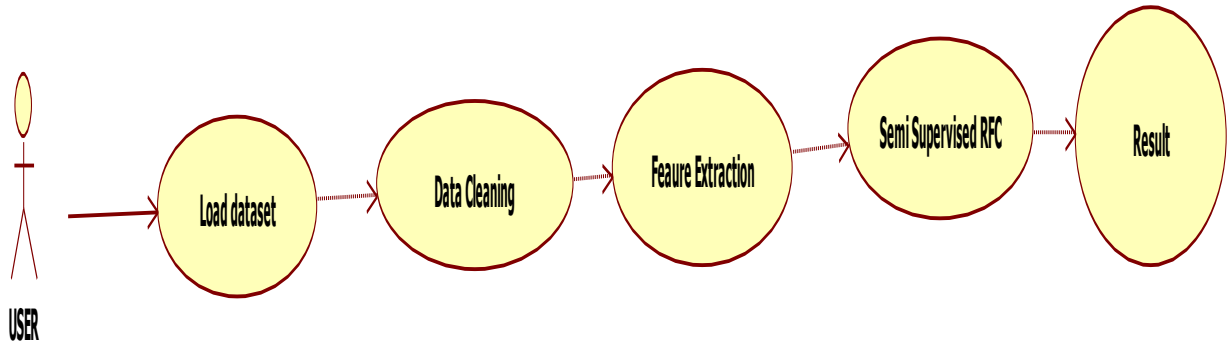
Due to their simplistic nature, use case diagrams can be a good communication tool for stakeholders. The drawings attempt to mimic the real world and provide a view for the stakeholder to understand how the system is going to be designed. Siau and Lee conducted research to determine if there was a valid situation for use case diagrams at all or if they were unnecessary. What was found was that the use case diagrams conveyed the intent of the system in a more simplified manner to stakeholders and that they were "interpreted more completely than class diagrams".

Use case diagrams are usually referred to as **behaviour diagrams** used to describe a set of actions (use case) that some system or systems (subject) should or can perform in collaboration with one or more **external users** of the system (actors). Each use case should provide some observable and valuable result to the actors or other stakeholders of the system.

The following diagram depicts the components of the use case diagram. They are

1.Actor
2.Action

In this article, we discuss the behavioural properties of the fake news detection system by using the use case structure. From the below diagram, **User** is the actor and the actions performed by him are load dataset, data cleaning, feature extraction, Semi supervised RFC and finally result is obtained.



3.1.3.SEQUENCE DIAGRAM

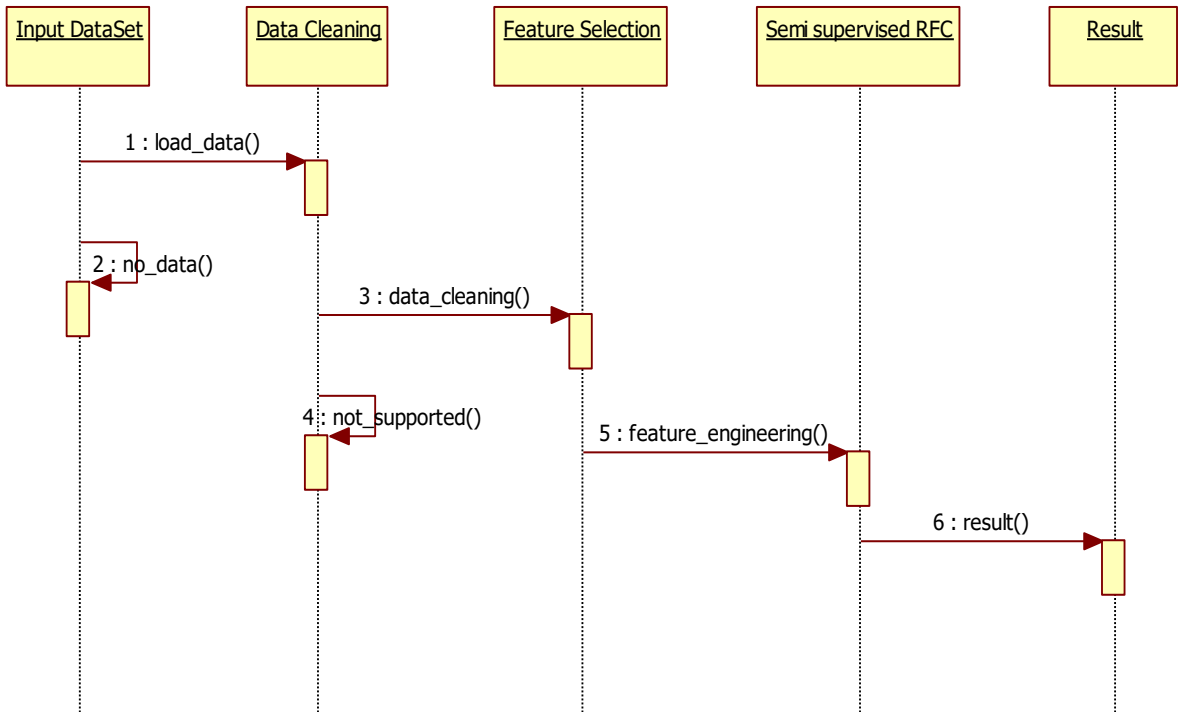
A **sequence diagram** shows object interactions arranged in time sequence. It depicts the objects involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario. Sequence diagrams are typically associated with use case realizations in the logical view of the system under development. Sequence diagrams are sometimes called **event diagrams** or **event scenarios**.

A sequence diagram shows, as parallel vertical lines (*lifelines*), different processes or objects that live simultaneously, and, as horizontal arrows, the messages exchanged between them, in the order in which they occur. This allows the specification of simple runtime scenarios in a graphical manner.

If the lifeline is that of an object, it demonstrates a role. Leaving the instance name blank can represent anonymous and unnamed instances.

Messages, written with horizontal arrows with the message name written above them, display interaction. Solid arrow heads represent synchronous calls, open arrow heads represent asynchronous messages, and dashed lines represent reply messages. If a caller sends a synchronous message, it must wait until the message is done, such as invoking a subroutine.

If a caller sends an asynchronous message, it can continue processing and doesn't have to wait for a response. Asynchronous calls are present in multithreaded applications, event-driven applications and in message-oriented middleware. Activation boxes, or method-call boxes, are opaque rectangles drawn on top of lifelines to represent that processes are being performed in response to the message (Execution Specifications in **uml**).



A SEQUENCE DIAGRAM FOR FAKE NEWS DETECTION SYSTEM

3.1.4.STATE DIAGRAM

A **state diagram** is a type of diagram used in computer science and related fields to describe the behaviour of systems. State diagrams require that the system described is composed of a finite number of states; sometimes, this is indeed the case, while at other times this is a reasonable abstraction. Many forms of state diagrams exist, which differ slightly and have different semantics.

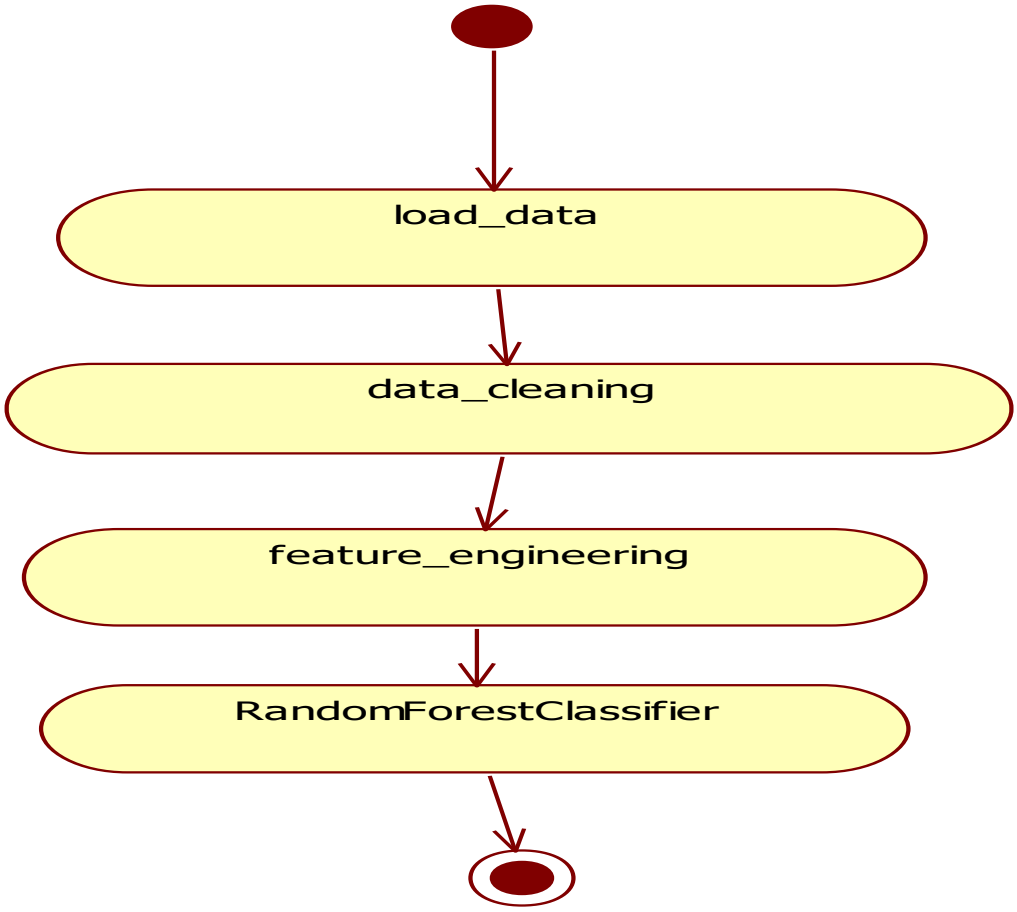
State diagrams are used to give an abstract description of the behaviour of a system. This behaviour is analyzed and represented by a series of events that can occur in one or more possible states. Hereby "each diagram usually represents objects of a single class and track the different states of its objects through the system".

State diagrams can be used to graphically represent finite-state machines (also called finite automata). This was introduced by Claude Shannon and Warren Weaver in their 1949 book ***The Mathematical Theory of Communication***. Another source is Taylor Booth in his 1967 book *Sequential Machines and Automata Theory*.

In this article ,the state diagram in the UML depicts the behaviour of the fake news detection system. The state box describes the state of the system and we have all about four different states such as load data, data cleaning, feature engineering, randomforest classifier.

- It captures the behaviour of an entity present in the system.
- It is used to represent the specific implementation of an element.
- The behaviour of a system can be modelled using behavioural state machine diagram in OOAD

The below figure depicts the flow of state chart diagram.



A STATE DIAGRAM FOR FAKE NEWS DETECTION SYSTEM

3.1.5.COLLABORATION DIAGRAM

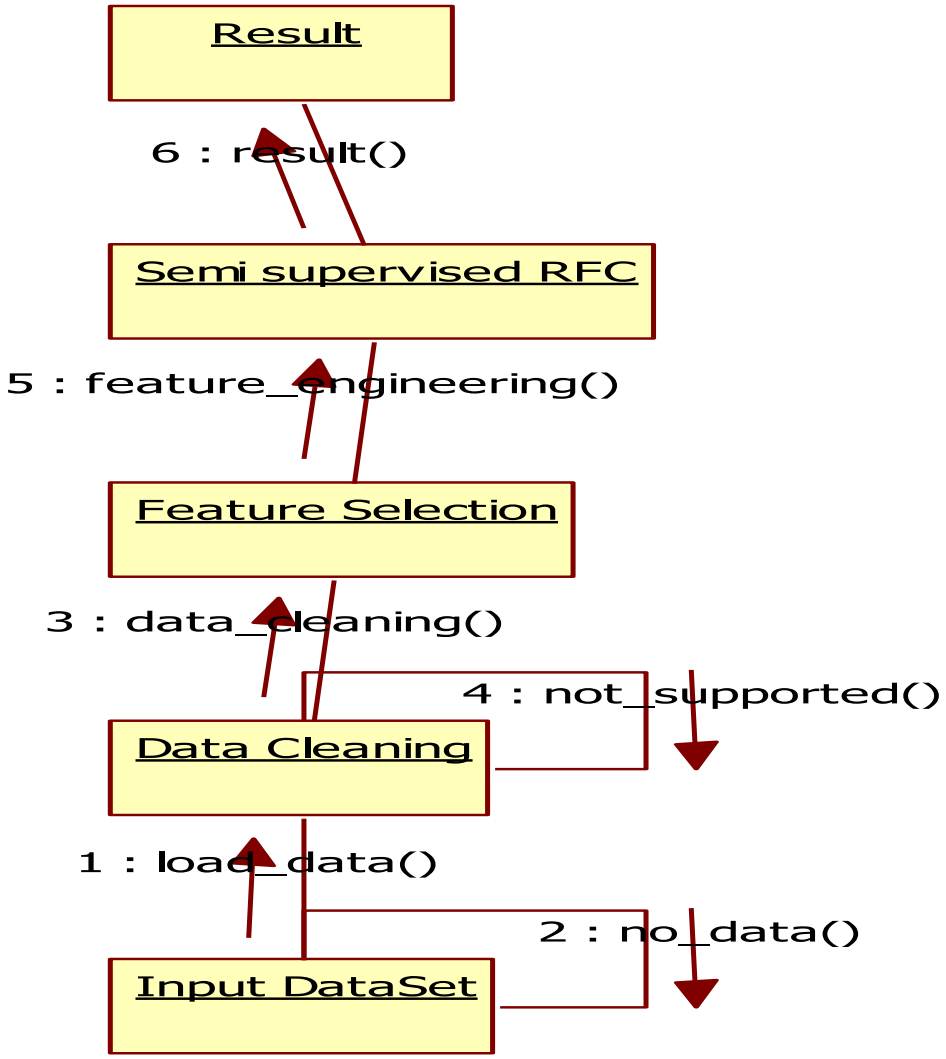
A collaboration diagram, also known as a communication diagram, is an illustration of the relationships and interactions among software objects in the Unified Modelling Language (UML). These diagrams can be used to portray the dynamic behaviour of a particular use case and define the role of each object.

Collaboration diagrams are created by first identifying the structural elements required to carry out the functionality of an interaction. A model is then built using the relationships between those elements. Several vendors offer software for creating and editing collaboration diagrams.

- It reinforces the structural aspects of an interaction system which is how the lifeline is connected.
- Messages transmitted over sequencing are shown by hierarchical numeration of each message.
- It enables to focus on the structural elements and not on the flow of message as stated in sequence diagrams.

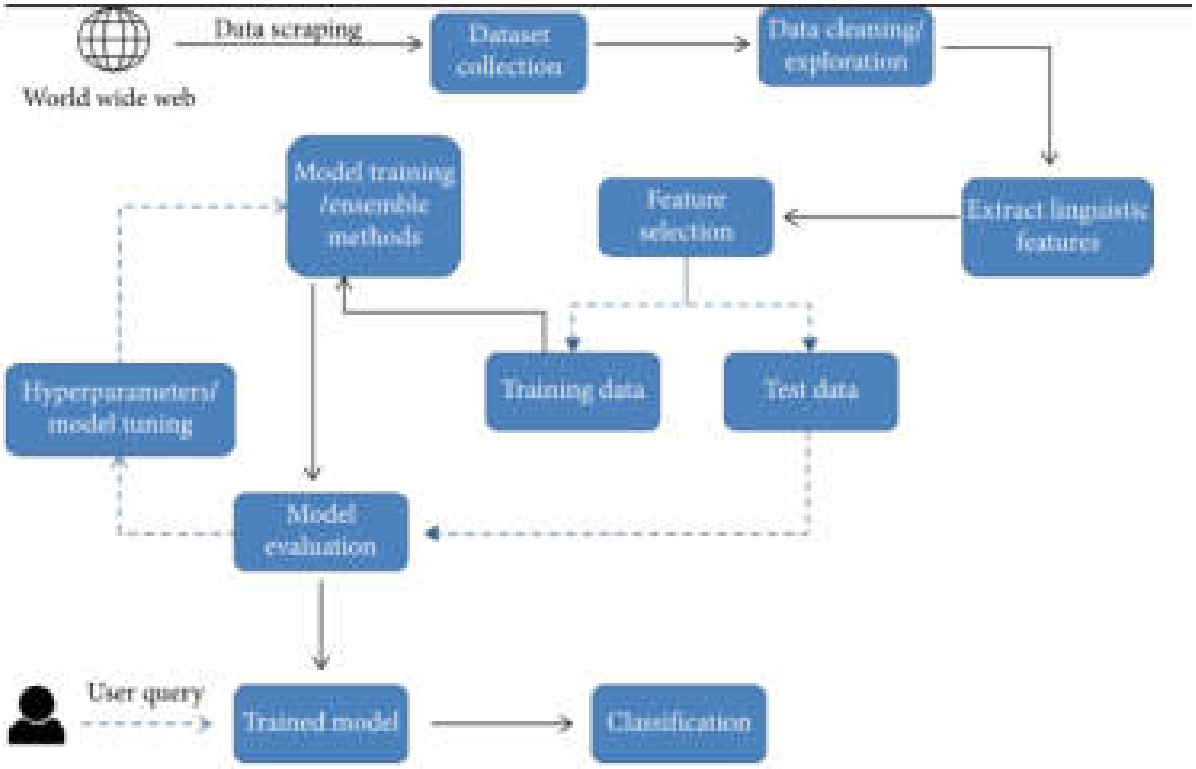
Collaboration diagrams have been used to envision the structural organization and interactions between objects. At the other hand, sequence diagrams concentrate on the order of messages moving in between objects. Nonetheless, a single figure is not enough in most situations to explain a system 's behaviour, so both diagrams are necessary.

Hence, In our article the fake news detection system has been described in the below collaboration diagrams with a clear structure.



A COLLABORATION DIAGRAM FOR FAKE NEWS DETECTION SYSTEM.

3.2 ARCHITECTURE OVERVIEW



Data collection: William Yang Wang's "A New Benchmark Dataset for Fake News Detection" was used to introduce and validate the proposed framework. This repository's dataset is divided into three sets: training, validation, and test. There are 12,836 short statements in the dataset that have been labelled for truthfulness, topic, context/venue, speaker, state, group, and prior history. The dataset initially classified the news into six fine-grained labels for the truthfulness ratings: pantsfire, false, barelyreal, half-true, mostlytrue, and true.

Data preprocessing: The six-label classification problem was translated into a binary classification problem with True and False labels for the proposed scheme. The labels were converted using the following mapping: • pantsfire --> False • false --> False • barely-true --> False • halftrue --> True • mostly-true --> True • true --> True In addition, only the news headline was used as an input for classification. Thus, in the preprocessing stage, the labels were first mapped using the above-mentioned mapping, after which only the labels and news statement columns were extracted from the dataset and saved in csv format for future use.

Following the preprocessing, we were able to obtain the following three cleaned files: • train.csv • valid.csv • test.csv.

Feature extraction: The following feature extraction method is used to help machine learning models gain insights from news headlines:

Count Vectorizer: First, the English stopwords were stripped from all of the news headlines using scikitlearn'sCountVectorizer, and then they were tokenized using spaces and punctuation marks as the delimiter. After all of the headlines had been tokenized, a sparse matrix of all of the news headlines as rows and the tokens as columns was restored. In addition to their morphological use, a number of n-grams were returned to make the tokens reflect the sense in which they were used.

Modelling: the models used for training. The features extracted from the CountVectorizer are used to train the models. After that, using GridSearchCV and a 5-hold out cross validation set, all of the models were hyperparameter tuned for all of the different possible parameters. The aim of this hyperparameter tuning was to boost the models' f1-score. After the models were fine-tuned, they were evaluated on a test range, and evaluation metrics for the models were determined.

4.IMPLEMENTATION

4.1INTRODUCTION TO PYTHON

Python is an interpreted high-level general-purpose programming language. Python's design philosophy emphasizes code readability with its notable use of significant indentation. Its language constructs as well as its object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

Python is dynamically-typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library.

- ✓ Python is an open source, object-oriented, high-level powerful programming language.
- ✓ Developed by Guido van Rossum in the early 1990s. Named after Monty Python
- ✓ Python runs on many Unix variants, on the Mac, and on Windows 2000 and later.
- ✓ Available for download from http://www.python.org.

4.1.1 HISTORY OF PYTHON

The name Python was selected from "Monty Python's Flying Circus" which was a British sketch comedy series created by the comedy group Monty Python and broadcast by the BBC from 1969 to 1974.

Python was created in the early 1990s by Guido van Rossum at the National Research Institute for Mathematics and Computer Science in Netherlands.

Python was created as a successor of a language called ABC (All Basic Code) and released publicly in1991. Guido remains Python's principal author, although it includes many contributions from active user community.

Between 1991 and 2001 there are several versions released, current stable release is 3.2. In 2001 the Python Software Foundation (PSF) was formed, a non-profit organization created specifically to own Python-related Intellectual Property. Zope Corporation is a sponsoring member of the PSF.

Python 2.0 was released on 16 October 2000, with many major new features, including a cycle-detecting garbage collector and support for Unicode.

Python 3.0 was released on 3 December 2008. It was a major revision of the language that is not completely backward-compatible. Many of its major features were backported to Python 2.6.x and 2.7.x version series. Releases of Python 3 include the 2 to 3 utility, which automates (at least partially) the translation of Python 2 code to Python 3.

Python 2.7's end-of-life date was initially set at 2015 then postponed to 2020 out of concern that a large body of existing code could not easily be forward-ported to Python 3. No more security patches or other improvements will be released for it. With Python 2's end-of-life, only Python 3.6.x and later are supported.

Python 3.9.2 and 3.8.8 were expedited as all versions of Python (including 2.7[1]) had security issues, leading to possible remote code execution and web cache poisoning.

4.1.2PYTHON EASY TO USE

Yes. It is still common to start students with a procedural and statically typed language such as Pascal, C, or a subset of C++ or Java. Students may be better served by learning Python as their first language. Python has a very simple and consistent syntax and a large standard library and, most importantly, using Python in a beginning programming course lets students concentrate on important programming skills such as problem decomposition and data type design. With Python, students can be quickly introduced to basic concepts such as loops and

procedures. They can probably even work with user-defined objects in their very first course.

Many other aspects of Python make it a good first language. Like Java, Python has a large standard library so that students can be assigned programming projects very early in the course that do something. Assignments aren't restricted to the standard four-function calculator and check balancing programs. By using the standard library, students can gain the satisfaction of working on realistic applications as they learn the fundamentals of programming. Using the standard library also teaches students about code reuse. Third-party modules such as PyGame are also helpful in extending the students' reach.

For a student who has never programmed before, using a statically typed language seems unnatural. It presents additional complexity that the student must master and slows the pace of the course. The students are trying to learn to think like a computer, decompose problems, design consistent interfaces, and encapsulate data. While learning to use a statically typed language is important in the long term, it is not necessarily the best topic to address in the students' first programming course.

4.1.3PYTHON ENVIRONMENT

AIX	AROS	AS/400 (OS/400)	BeOS
MorphOS	MS-DOS	OS/2	OS/390 and z/OS
Palm OS	PlayStation and PSP	Psion	QNX
RISC OS	Series 60	Solaris	VMS

Windows CE or Pocket PC	HP-UX	Linux	

A virtual **environment** is a **Python environment** such that the **Python** interpreter, libraries and scripts installed into it are isolated from those installed in other virtual **environments**, and (by default) any libraries installed in a “system” **Python**, i.e., one which is installed as part of your operating system.

4.1.4 PYTHON FEATURES

Open source: Python is publicly available open source software,any one can use source code that doesn't cost anything.

Easy-to-learn: Popular (scripting/extension) language, clear and easy syntax, no type declarations, automatic memory management, high-level data types and operations, design to read (more English like syntax) and write (shorter code compared to C, C++, and Java) fast.

High-levelLanguage:
High-level language (closer to human) refers to the higher level of concept from machine language (for example assembly languages). Python is an example of a high-level language like C, C++, Perl, and Java with low-level optimization.

Portable:
High level languages are portable, which means they are able to run across all major hardware and software platforms with few or no change in source code. Python is portable and can be used on Linux, Windows, Macintosh, Solaris, FreeBSD, OS/2, Amiga, AROS, AS/400 and many more.

Object-Oriented: Python is a full-featured object-oriented programming language, with features such as classes, inheritance, objects, and overloading.

Python is Interactive:

Python has an interactive console where you get a Python prompt (command line) and interact with the interpreter directly to write and test your programs. This is useful for mathematical programming.

Interpreted : Python programs are interpreted, takes source code as input, and then compiles (to portable byte-code) each statement and executes it immediately. No need to compiling or linking

Extendable : Python is often referred to as a "glue" language, meaning that it is capable to work in mixed-language environment. The Python interpreter is easily extended and can add a new built-in function or modules written in C/C++/Java code.

Libraries : Databases, web services, networking, numerical packages, graphical user interfaces, 3D graphics, others.

Supports : Support from online Python community

4.1.5MAJOR USES OF PYTHON

- System utilities (system admin tools, command line programs).
- Web Development.
- Graphical User Interfaces (Tkinter, gtk, Qt).
- Internet scripting.
- Embedded scripting.
- Database access and programming.
- Game programming.

- Rapid prototyping and development.
- Distributed programming

4.1.6 0RGANIZATIONS USING PYTHON

Web Development : Yahoo Maps, Yahoo Groups, Google, Zope Corporation, Ultraseek, Linux Weekly News, ElasticHosts Cloud Servers, Mojam.com, hunch, Shopzilla, Movieplayer.it, Multiplayer.it.

Games: Battlefield 2, Crystal Space, Star Trek Bridge Commander, The Temple of Elemental Evil, Vampire: The Masquerade: Bloodlines, Civilization 4, QuArK (Quake Army Knife)

Graphics : Industrial Light & Magic, Walt Disney Feature Animation, HKS, Inc. (ABAQUS/CAE), RoboFog, Caligari Corporation, Blender 3D, Jasc Software, Paint Shop Pro.

Financial : Altis Investment Management, ABN AMRO Bank, Treasury Systems, Bellco Credit Union, Journyx Timesheet and Resource Management Software.

Science : National Weather Service, Radar Remote Sensing Group, Applied Maths, Biosoft, The National Research Council of Canada, Los Alamos National Laboratory (LANL) Theoretical Physics Division, AlphaGene, Inc., LLNL, NASA, Swedish Meteorological and Hydrological Institute (SMHI), Environmental Systems Research Institute (ESRI), Objexx Engineering, Nmag Computational Micromagnetics

Electronic Design Automation: Ciranova, Productivity Design Tools, Object Domain, Pardus, Red Hat, SGI, Inc., MCI Worldcom, Nokia,

Education : University of California, Irvine, Smeal College of Business, The Pennsylvania State University, New Zealand Digital Library, IT Certification Exam preparation, SchoolTool,

Business Software : Raven Bear Systems Corporation, Thawte Consulting, Advanced Management Solutions Inc., IBM, Arakn<E9>, RealNetworks, dSPACE, Escom, The Tiny Company, Nexedi, Piensa Technologies - Bufete Consultor de Mexico, Nektra, WuBook.

4.2 SOURCE CODE

4.2.1 MAIN

```
CONSUMER_KEY = "ZZCQIC9BpywndHAM0AiefpgLf"

#CONSUMER_SECRET = "zGy5D2Z2AN2xZUblym7SjYRW37fKlEV2euYbT86eIUqOVZJ96W"

#OAUTH_TOKEN = "845406513338429442-VMJDy4xMuEmEL8p4Ung1aTWCa4WyGF3"

#OAUTH_TOKEN_SECRET = "gll1RwCMthjM5sLVLLdtaFYxR2wSyAnc1LJ71FyzXSRWg"

#pi install tweepy

import tweepy

from tweepy import OAuthHandler

import pandas as pd

import matplotlib.pyplot as plt

import nltk

import numpy as np



pd.options.display.max_columns = 50

pd.options.display.max_rows= 50
```

```
pd.options.display.width= 120


consumer_key = 'ZZCQIC9BpywndHAM0AiefpgLf'
consumer_secret = 'zGy5D2Z2AN2xZUblym7SjYRW37fKIEV2euYbT86elUqOVZJ96W'
access_token = '845406513338429442-VMJDy4xMuEmEL8p4Ung1aTWCa4WyGF3'
access_secret = 'gll1RwCMthjM5sLVLLdtaFYxR2wSyAnc1LJ71FyzXSRWg'


auth = OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_secret)


api = tweepy.API(auth)


results = []
i=0
for tweet in tweepy.Cursor(api.search, q='COVID19', lang="en").items():
    if (not tweet.retweeted) and ('RT @' not in tweet.text):
results.append(tweet)
        id = tweet.id
i=i+1
        if i==50:
            break
```

```
print (len(results))

def process_results(results):

id_list = [tweet.id for tweet in results]

data_set = pd.DataFrame(id_list, columns=["id"])

# Processing Tweet Data

data_set["text"] = [tweet.text for tweet in results] #text of tweet

data_set["created_at"] = [tweet.created_at for tweet in results] #when the tweet was created

data_set["retweet_count"] = [tweet.retweet_count for tweet in results] #number of retweets

data_set["favorite_count"] = [tweet.favorite_count for tweet in results] #number of favourites

data_set["source"] = [tweet.source for tweet in results] #source of the tweet

data_set["length"] = [len(tweet.text) for tweet in results] #number of characters in tweet

# Processing User Data

data_set["user_id"] = [tweet.author.id for tweet in results] #id of the author

data_set["user_screen_name"] = [tweet.author.screen_name for tweet in results]

data_set["user_name"] = [tweet.author.name for tweet in results]

data_set["user_created_at"] = [tweet.author.created_at for tweet in results] #age of user account

data_set["user_description"] = [tweet.author.description for tweet in results]
```

```
    data_set["user_followers_count"] = [tweet.author.followers_count for tweet in results] #number of followers

    data_set["user_friends_count"] = [tweet.author.friends_count for tweet in results] #number of friends

    data_set["user_location"] = [tweet.author.location for tweet in results] #user has a location in profile?

    data_set["user_statuses_count"] = [tweet.author.statuses_count for tweet in results] #number of statuses

    data_set["user_verified"] = [tweet.author.verified for tweet in results] #user is verified?

    data_set["user_url"] = [tweet.author.url for tweet in results] #user has a URL?

        return data_set

    data_set = process_results(results)

    data_set.to_csv("Tweets.csv", index=False, encoding='utf-8')

    df=pd.read_csv("Tweets.csv")

    print(df)

    df["user_has_url?"] = ""

    df["user_has_url?"] = np.where(df["user_url"].isnull(), 'No', 'Yes')


import  test main.
```

4.2.2 HYBRID NURAL NETWORK MODEL BUILDING

```
from keras.models import Sequential,load_model
from keras.layers import Dense,Dropout
from keras.layers import LSTM,Conv1D
from keras.layers import MaxPooling1D
from keras.layers import Flatten
from keras.layers.embeddings import Embedding
from keras.preprocessing import sequence
from keras.preprocessing.text import Tokenizer
from keras import optimizers
from keras.layers import TimeDistributed
import pandas as pd
from sklearn.model_selection import train_test_split
import numpy as np
import pickle

from sklearn.preprocessing import LabelEncoder

vocabulary_size = 400000
#*****
time_step=300
embedding_size=100

dataVal_Fake_Real=pd.read_csv('Tweets.csv')

texts=[]
texts=dataVal_Fake_Real['text']#######################################
label=dataVal_Fake_Real['user_verified']
#print(label)
#X=texts.astype(str).values.tolist()
```

```
#X=np.reshape(X,(-1,1))
from Clean_Texts import clean_text
X=texts.map(lambda x: clean_text(x))
#print(X)
#label=label.astype(int).values
labelEncoder=LabelEncoder()
encoded_label=labelEncoder.fit_transform(label)
y=np.reshape(encoded_label,(-1,1))



#X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2)

training_size=int(0.8*X.shape[0])
print(X.shape[0],training_size)
X_train=X[:training_size]
y_train=y[:training_size]
X_test=X[training_size:]
y_test=y[training_size:]

#Tokenizing texts
tokenizer = Tokenizer(num_words= vocabulary_size)
tokenizer.fit_on_texts(X_train)
sequences_train= tokenizer.texts_to_sequences(X_train)
X_train = sequence.pad_sequences(sequences_train,
maxlen=time_step,padding='post')

print(len(tokenizer.word_index))



vocab_size=len(tokenizer.word_index)+1
```

```
#Reading Glove
f = open('glove.6B.100d.txt',encoding='utf-8')
embeddings={}
for line in f:
    values = line.split()
    word = values[0]
coefs = np.asarray(values[1:], dtype='float32')
    embeddings[word] = coefs
f.close()

print('Total %s word vectors.' % len(embeddings))

# create a weight matrix for words in training docs
embedding_matrix = np.zeros((vocab_size, embedding_size))
for word, i in tokenizer.word_index.items():
    embedding_vector = embeddings.get(word)
    if embedding_vector is not None:
            embedding_matrix[i] = embedding_vector

print(embedding_matrix.shape[0],embedding_matrix.shape[1])



sequences_test= tokenizer.texts_to_sequences(X_test)
X_test = sequence.pad_sequences(sequences_test, maxlen=time_step,padding='post')

#print(len(X_test),len(y_test))
#print(label)



# Embedding
```

```
#maxlen = 100
#embedding_size = 32
## create model
model = Sequential()
model.add(Embedding(np.array(embedding_matrix).shape[0],
embedding_size, weights=[embedding_matrix], trainable=False))

model.add(LSTM(300))

model.add(Dense(1, activation='sigmoid'))

model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])
model.summary()
## Fit train data
history=model.fit(X_train, y_train, epochs = 10,batch_size=64,shuffle=True)



print("Saving Model...")
model_name =
'Models/Model_lstm_FR_2.h5'##########################################3
model.save(model_name)##################################################
#
score=model.evaluate(X_test,y_test,verbose=1)
print('acc: '+str(score[1]))
from sklearn.metrics import precision_recall_fscore_support,classification_report
y_pred=model.predict_classes(X_test)
print('Classification report:\n',classification_report(y_test,y_pred))
from sklearn.metrics import precision_recall_fscore_support, classification_report
y_pred=model.predict_classes(X_test)
print('Classification Report: '+classification_report(y_test, y_pred))
##########################################
```

4.3 TEST DATA

CONSUMER_KEY = "ZZCQIC9BpywndHAM0AiefpgLf"

CONSUMER_SECRET = "zGy5D2Z2AN2xZUblym7SjYRW37fKlEV2euYbT86eIUqOVZJ96W"

OAUTH_TOKEN = "845406513338429442-VMJDy4xMuEmEL8p4Ung1aTWCa4WyGF3"

OAUTH_TOKEN_SECRET = "gll1RwCMthjM5sLVLLdtaFYxR2wSyAnc1LJ71FyzXSRWg"

5.TESTING

5.1 WHAT IS TESTING ?

Testing is the process of executing a program with the aim of finding errors. To make our software perform well it should be error-free. If testing is done successfully it will remove all the errors from the software.

5.2 PRINCIPLES OF TESTING

- All the test should meet the customer requirements
  To make our software testing should be performed by a third party
  Exhaustive testing is not possible.
- As we need the optimal amount of testing based on the risk assessment of the application.
- All the test to be conducted should be planned before implementing it
  It follows the Pareto rule(80/20 rule) which states that 80% of errors come from 20% of program components.
  Start testing with small parts and extend it to large parts.

5.3 TESTING METHODOLOGIES

The following are the Testing Methodologies:

- **Unit Testing.**
- **Integration Testing.**
- **User Acceptance Testing.**
- **Output Testing.**
- **Validation Testing.**

Unit Testing

Unit testing focuses verification effort on the smallest unit of Software design that is the module. Unit testing exercises specific paths in a module's control structure to ensure complete coverage and maximum error detection. This test focuses on each module individually, ensuring that it functions properly as a unit. Hence, the naming is Unit Testing.

During this testing, each module is tested individually and the module interfaces are verified for the consistency with design specification. All important processing path are tested for the expected results. All error handling paths are also tested.

Integration Testing

Integration testing addresses the issues associated with the dual problems of verification and program construction. After the software has been integrated a set of high order tests are conducted. The main objective in this testing process is to take unit tested modules and builds a program structure that has been dictated by design.

The following are the types of Integration Testing:

1.Top Down Integration

This method is an incremental approach to the construction of program structure. Modules are integrated by moving downward through the control hierarchy, beginning with the main program module. The module subordinates to the main program module are incorporated into the structure in either a depth first or breadth first manner.

In this method, the software is tested from main module and individual stubs are replaced when the test proceeds downwards.

2. Bottom-up Integration

This method begins the construction and testing with the modules at the lowest level in the program structure. Since the modules are integrated from the bottom up, processing required for modules subordinate to a given level is always available and the need for stubs is eliminated. The bottom up integration strategy may be implemented with the following steps:

- The low-level modules are combined into clusters into clusters that perform a specific Software sub-function.
- A driver (i.e.) the control program for testing is written to coordinate test case input and output.
- The cluster is tested.
- Drivers are removed and clusters are combined moving upward in the program structure

The bottom up approaches tests each module individually and then each module is module is integrated with a main module and tested for functionality.

User Acceptance Testing

User Acceptance of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changes wherever required. The system developed provides a friendly user interface that can easily be understood even by a person who is new to the system.

Output Testing

After performing the validation testing, the next step is output testing of the proposed system, since no system could be useful if it does not produce the required output in the specified format. Asking the users about the format required by them tests the outputs generated or displayed by the system under consideration. Hence the output format is considered in 2 ways – one is on screen and another in printed format.

Validation Checking

Validation checks are performed on the following fields.

Text Field:

The text field can contain only the number of characters lesser than or equal to its size. The text fields are alphanumeric in some tables and alphabetic in other tables. Incorrect entry always flashes and error message.

Numeric Field:

The numeric field can contain only numbers from 0 to 9. An entry of any character flashes an error messages. The individual modules are checked for accuracy and what it has to perform. Each module is subjected to test run along with sample data. The individually tested modules are integrated into a single system. Testing involves executing the real data information is used in the program the existence of any program defect is inferred from the output. The testing should be planned so that all the requirements are individually tested.

A successful test is one that gives out the defects for the inappropriate data and produces and output revealing the errors in the system.

5.4Preparation of Test Data

Taking various kinds of test data does the above testing. Preparation of test data plays a vital role in the system testing. After preparing the test data the system under study is tested using that test data. While testing the system by using test data errors are again uncovered and corrected by using above testing steps and corrections are also noted for future use.

Using Live Test Data:

Live test data are those that are actually extracted from organization files. After a system is partially constructed, programmers or analysts often ask users to key in a set of data from their normal activities. Then, the systems person uses this data as a way to partially test the system. In other instances, programmers or analysts extract a set of live data from the files and have them entered themselves.

It is difficult to obtain live data in sufficient amounts to conduct extensive testing. And, although it is realistic data that will show how the system will perform for the typical processing requirement, assuming that the live data entered are in fact typical, such data generally will not test all combinations or formats that can enter the system. This bias toward typical values then does not provide a true systems test and in fact ignores the cases most likely to cause system failure.

Using Artificial Test Data:

Artificial test data are created solely for test purposes, since they can be generated to test all combinations of formats and values. In other words, the artificial data, which can quickly be prepared by a data generating utility program in the information systems department, make possible the testing of all login and control paths through the program.

The most effective test programs use artificial test data generated by persons other than those who wrote the programs. Often, an independent team of testers formulates a testing plan, using the systems specifications.

The package “Virtual Private Network” has satisfied all the requirements specified as per software requirement specification and was accepted.

5.5 USER TRAINING

Whenever a new system is developed, user training is required to educate them about the working of the system so that it can be put to efficient use by those for whom the system has been primarily designed. For this purpose the normal working of the project was demonstrated to the prospective users. Its working is easily understandable and since the expected users are people who have good knowledge of computers, the use of this system is very easy.

5.6 TYPES OF TESTS

Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

1.*White Box Testing*

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

2. Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as

specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

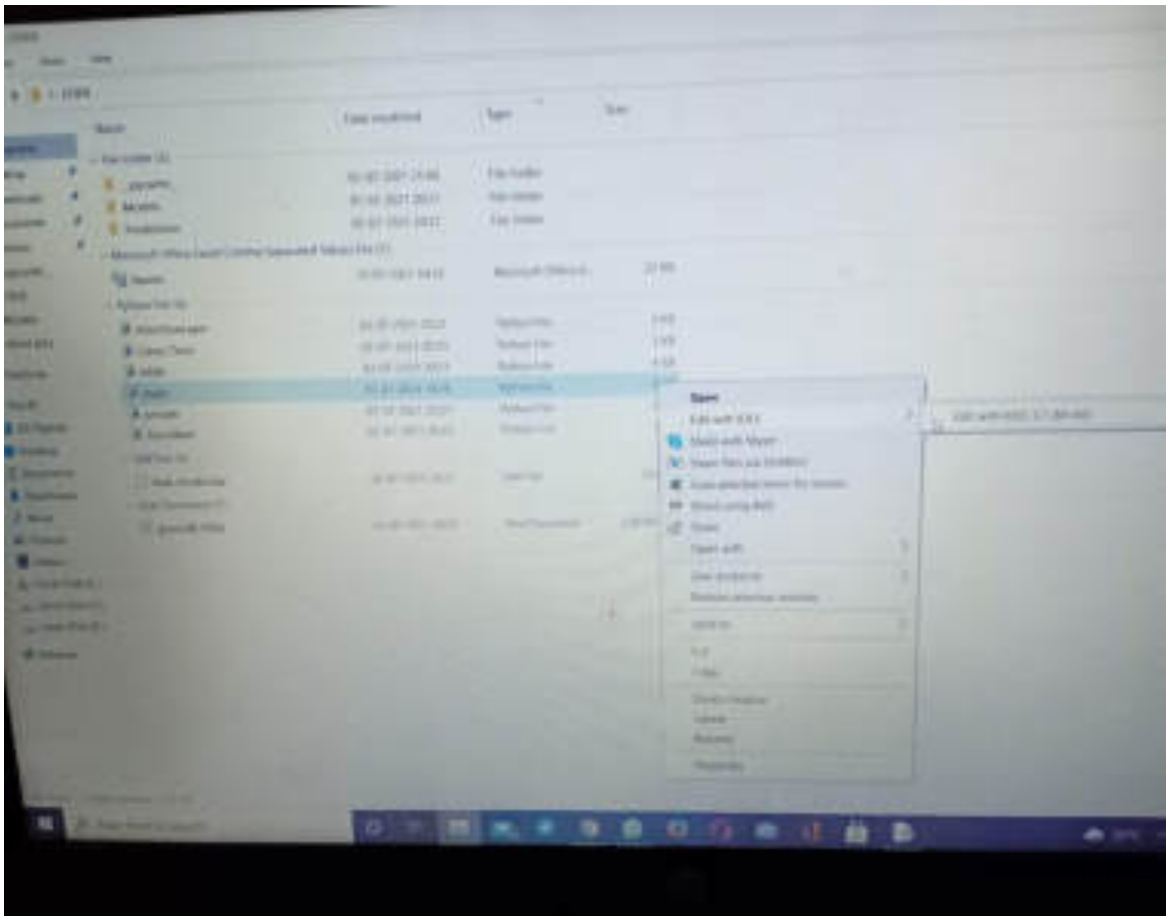
5.7MAINTAINENCE

This covers a wide range of activities including correcting code and design errors. To reduce the need for maintenance in the long run, we have more accurately defined the user’s requirements during the process of system development. Depending on the requirements, this system has been developed to satisfy the needs to the largest possible extent. With development in technology, it may be possible to add many more features based on the requirements in future. The coding and designing is simple and easy to understand which will make maintenance easier.

6.RESULTS

6.1EXECUTION

Path for execution



As our system contains the PyCharm software, opening our source file in the IDLE(integrated development and learning environment) will get executed automatically.

OUTPUT 1

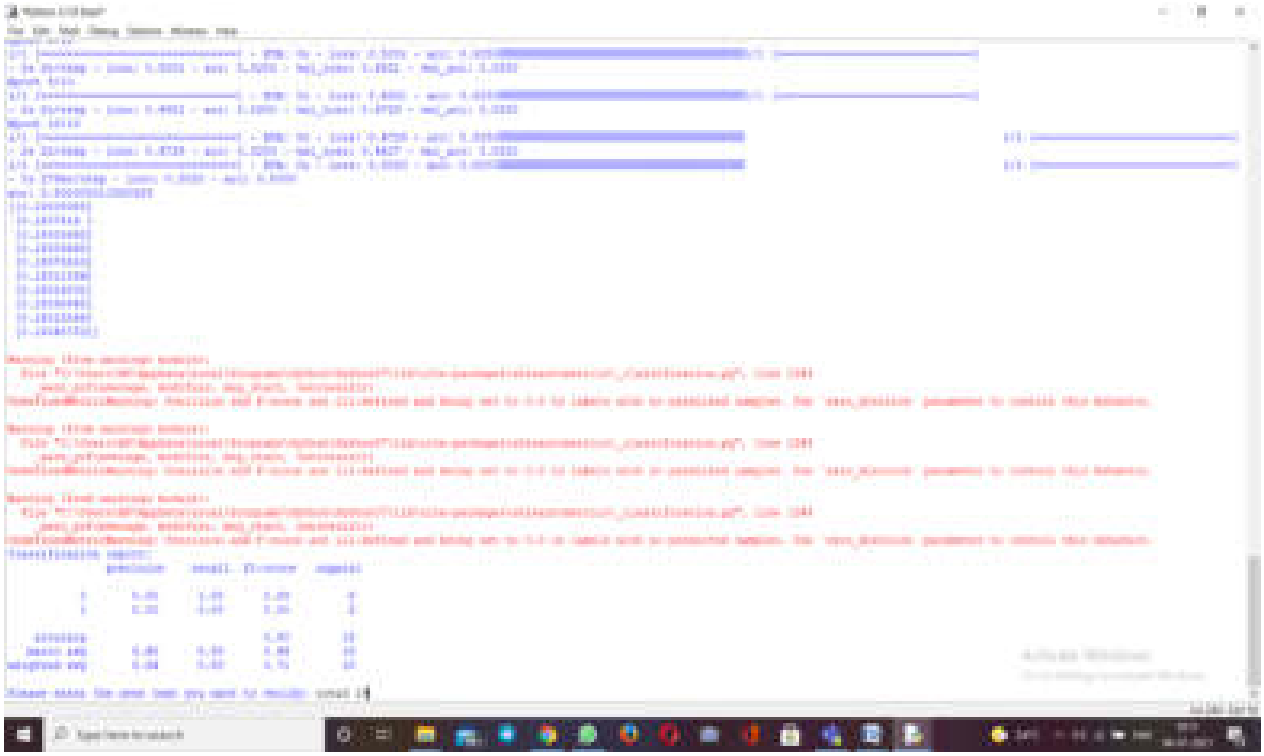


Fig: Building a Hybrid Neural Network Model

OUTPUT2

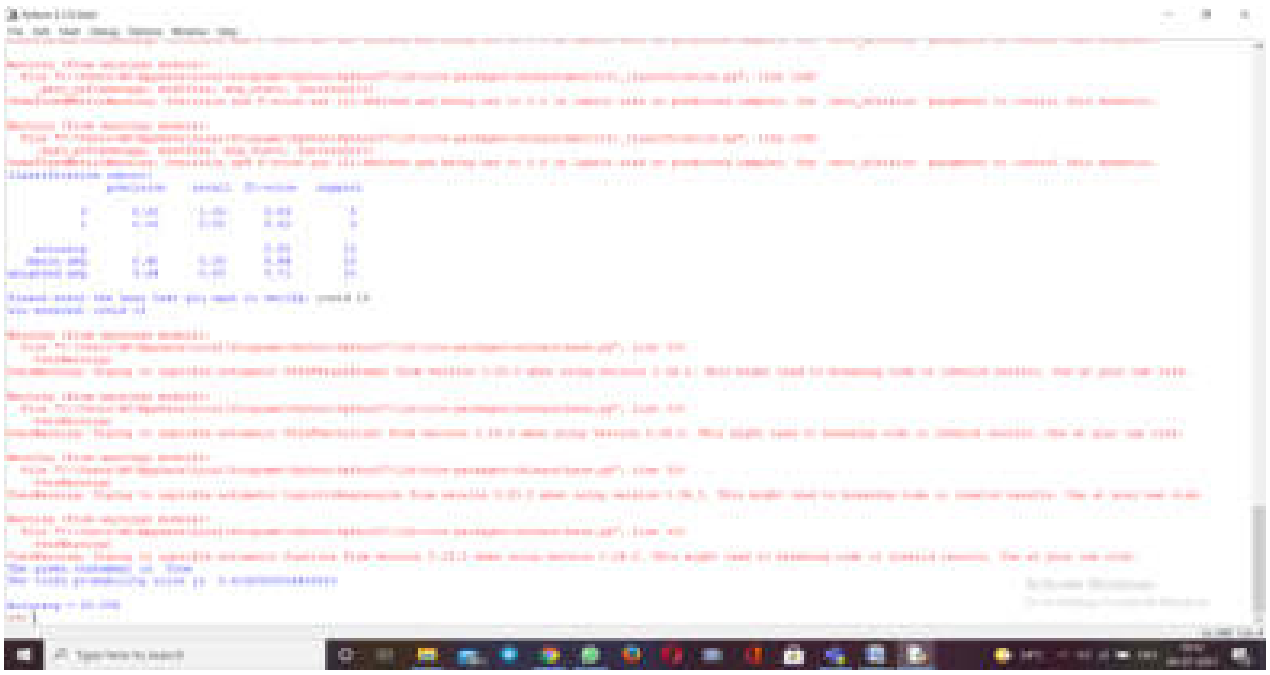


Fig: Accuracy of News in social media.

CONCLUSION AND FUTURE WORK

we addressed the task of automatic identification of fake news. We introduced two new fake news datasets, one obtained through crowdsourcing and another one obtained from the web covering celebrities. We developed classification models that rely on a combination of lexical, syntactic, and semantic information, as well features representing text readability properties. Our best per-forming models achieved accuracies that are com-parable to human ability to spot fake content. This paper is based on the crowd sourcing dataset and the web covering dataset. These are the static datasets. Through these we can only test the data which is present in the predefined training data sets. The paper gives the appropriate result for the test data that is present in the training datasets. Thus, the future scope of the paper is connecting this methodology to the internet news which gives results even for the test data that is not present in the training data sets. We can even change to some other better classifier to classify the data other than naïve bayes and logistic regression.

REFERENCES

[1] Majed Alrubaian, Muhammad Al-Qurishi, A Credibility Analysis System for Assessing Information on Twitter, IEEE Transactions on Dependable and Secure Computing, 1-14. DOI : http://dx.doi.org/ 10.1109/TDSC.2016.2602338 .

[2] Manish Gupta, Peixiang Zhao, Jiawei Han, 2012. Evaluating Event Credibility on Twitter, Proceedings of the 2012 SIAM International Conference on Data Mining, 153-164, DOI: http://epubs.siam.org/doi/abs/10.1137/ 1.9781611972825.14 .

[3] Krzysztof Lorek, Jacek Suehiro-Wici´nski, Micha l Jankowski-Lorek, Amit Gupta, Automated credibility assessment on twitter, Computer Science, 2015, Vol.16(2), 157-168, DOI: http://dx.doi.org/10.7494/csci. 2015.16.2.157.

[4] Ballouli, Rim El, Wassim El-Hajj, Ahmad Ghandour, Shady Elbassuoni, Hazem M. Hajj and Khaled Bashir Shaban. 2017. "CAT: Credibility Analysis of Arabic Content on Twitter." WANLP@EACL (2017).

[5] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," Proceedings of the Association for Information Science and Technology, vol. 52, no. 1, pp. 1–4, 2015.

[6] S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for deception detection," in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2, Association for Computational Linguistics, 2012, pp. 171–175.

[7] Granik, M., &Mesyura, V. 2017. Fake news detection using naive Bayes classifier. 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), 900-903.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"RATING PREDICTION BASED ON SOCIAL SENTIMENT FROM TEXTUAL REVIEWS"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
COMPUTER SCIENCE AND ENGINEERING

Submitted by

M.Srivani (17H71A0545)

K.Roshitha (17H71A0535)

P.Srivalli (17H71A0549)

Under the Esteemed Guidance of

Mrs.V.Lakshmi Chetana
Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr.Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"RATING PREDICTION BASED ON SOCIAL SENTIMENT FROM TEXTUAL REVIEWS"** is a bonafide work carried out by **all M.Srivani(17H71A0545),K.Roshitha(17H71A0535) and P.Srivalli(17H71A0549)** in partial fulfillment for the award of degree of Bachelor of Technology in **computer science and Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

Mrs.V.Lakshmi Chetana
Project Supervisor

Mr.D.Prasad
Head of the Department

Dr. K. Srinivas
Principal

Examiner 1

Examiner 2

ACKNOWLEDGEMENT

We would like to take this opportunity to express our deepest appreciation to the following people for their valuable contributions and assistance with this Project.

Initially, we would like to thank my project supervisor, **Mrs.V.Lakshmi Chetana , Assistant Professor, Department of Computer Science and Engineering** for the guidance and support, especially for the valuable ideas and knowledge shared to me throughout the Project.

We have the immense pleasure in expressing my thanks and deep sense of gratitude to**, Mr. D. Prasad, Head of the Department, Computer Science and Engineering** for extending necessary facilities and support for the completion of the Project.

We whole sincerely acknowledge **Dr. K. Srinivas, Principal and Prof D. PANDURANGA RAO, CEO** for giving opportunity to take up the Project work. We also extend our thanks to all faculty members of **Computer Science and Engineering**, for their valuable guidance and encouragement.

We would like to extend our warm appreciation to all our friends for sharing us their knowledge, valuable contributions and help to this Project.

Finally, our special thanks go to my family for their continuous support and help throughout my academic years and for their continual support and encouragement for the completion of the project.

M.Srivani (17H71A0545)

K.Roshitha (17H71A0535)

P.Srivalli (17H71A0549)

DECLARATION

We M.Srivani,K.Roshitha,P.Srivalli of the main-Project "**RATING PREDICTION BASED ON SOCIAL SENTIMENT FROM TEXTUAL REVIEWS**",hereby declare that the matter embodied in this Project is the genuine work done by us and has not been submitted either to this university or to any other university/institute for the fulfilment of the requirement of any course of study.

M.Srivani (17H71A0545)
K.Roshitha (17H71A0535)
P.Srivalli (17H71A0549)

ABSTRACT

In recent years, we have witnessed a flourish of review websites. It presents a great opportunity to share our viewpoints for various products we purchase. However, we face the information overloading problem. How to mine valuable information from reviews to understand a user's preferences and make an accurate recommendation is crucial. Traditional recommender systems (RS) consider some factors, such as user's purchase records, product category, and geographic location.

In this work, we propose a sentiment-based rating prediction method (RPS) to improve prediction accuracy in recommender systems. Firstly, we propose a social user sentimental measurement approach and calculate each user's sentiment on items/products. Secondly, we not only consider a user's own sentimental attributes but also take interpersonal sentimental influence into consideration.

Then, we consider product reputation, which can be inferred by the sentimental distributions of a user set that reflect customers' comprehensive evaluation. At last, we fuse three factors-user sentiment similarity, interpersonal sentimental influence, and item's reputation similarity into our recommender system to make an accurate rating prediction. We conduct a performance evaluation of the three sentimental factors on a real-world dataset collected from Yelp. Our experimental results show the sentiment can well characterize user preferences, which help to improve the recommendation performance.

CONTENTS

INTRODUCTION

Sentiment analysis is the most fundamental and important work in extracting user's interest preferences. In general, sentiment is used to describe user's own attitude on items. We observe that in many practical cases, it is more important to provide numerical scores rather than binary decisions. Generally, reviews are divided into two groups, positive and negative. However, it is difficult for customers to make a choice when all candidate products reflect positive sentiment or negative sentiment. To make a purchase decision, customers not only need to know whether the product is good, but also need to know how good the product is. It's also agreed that different people may have different sentimental expression preferences. For example, some users prefer to use "good" to describe an "excellent" product, while others may prefer to use "good" to describe a "just so so" product.

In our daily life, customers are most likely to buy those products with highly-praised reviews. That is, customers are more concerned about item's reputation, which reflects consumers' comprehensive evaluation based on the intrinsic value of a specific product. To obtain the reputation of a product, sentiment in reviews is necessary. Normally, if item's reviews reflect positive sentiment, the item may be with good reputation to a great extent. Oppositely, if item's reviews are full of negative sentiment, then the item is to be with bad reputation. To a given product, if we know user sentiment, we can infer the reputation and even the comprehensive ratings. When we search the net for purchasing, both positive reviews and negative reviews are valuable to be as reference. For positive reviews, we can know the advantages of a product. For negative reviews, we can obtain the shortcomings in case of being cheated. So it's worth to explore those reviewers who have obvious and objective attitude on items. We observe that reviewers' sentiment will influence others: if a reviewer has clear like and dislike sentiment, other users will pay much attention to him/her. However, user's sentiment is hard to predict and the unpredictability of interpersonal sentimental influence makes a great difficulty in exploring social users.

In addition to extracting user preferences, there is much work paying attention to the interpersonal interaction. Many approaches about the interpersonal influence in social networks have proved good performance in recommendation, which can effectively solve the "cold start"

problems. However, the existing approaches mainly leverage product category information or tag information to study the interpersonal influence. These methods are all restricted on the structured data, which is not always available on some websites. However, user reviews can provide us ideas in mining interpersonal inference and user preferences.

To address these problems, we propose a sentiment-based rating prediction method in the framework of matrix factorization. In our work, we make use of social users' sentiment to infer ratings. Fig. 1 is an example that illustrates our motivation. First, we extract product features from user reviews. Then, we find out the sentiment words, which are used to describe the product features. Besides, we leverage sentiment dictionaries to calculate sentiment of a specific user on an item/product. What is more, we combine social friend circle with sentiment to recommend. In Fig.1, the last user is interested in those product features, so based on the user reviews and the sentiment dictionaries, the last item will be recommended. Compared with previous work [2-5], [8], [9], the main difference is that: we use unstructured information to recommend instead of other structured social factors. Compared with [6] the main difference is that: their work mainly focuses on classifying users into binary sentiment (i.e. positive or negative), and they do not go further in mining user's sentiment. In our paper, we not only mine social user's sentiment, but also explore interpersonal sentimental influence and item's reputation. Finally, we take all of them into the recommender system.

The main contributions of our approach are as follows: 1) we propose a user sentimental measurement approach, which is based on the mined sentiment words and sentiment degree words from user reviews. Besides, some scalable applications are proposed. For example, we explore how the mined sentiment spread among users' friends. What is more, we leverage social users' sentiment to infer item's reputation, which showed great improvement in accuracy of rating prediction. 2) We make use of sentiment for rating prediction. User sentiment similarity focuses on the user interest preferences. User sentiment influence reflects how the sentiment spreads among the trusted users. Item reputation similarity shows the potential relevance of items. 3) We fuse the three factors: user sentiment similarity, interpersonal sentimental influence, and item reputation similarity into a probabilistic matrix factorization framework to carry out an accurate recommendation. The experimental results and discussions show that user's social

sentiment that we mined is a key factor in improving rating prediction performances.

1.1 PROBLEM STATEMENT:

- In recent years, we have witnessed a flourish of review websites. It presents a great opportunity to share our viewpoints for various products we purchase. However, we face the information overloading problem. How to mine valuable information from reviews to understand a user's preferences and make an accurate recommendation is crucial. Traditional recommender systems (RS) consider some factors, such as user's purchase records, product category, and geographic location.

1.2 PURPOSE:

- Firstly, we propose a social user sentimental measurement approach and calculate each user's sentiment on items/products. Secondly, we not only consider a user's own sentimental attributes but also take interpersonal sentimental influence into consideration.

1.3 MOTIVATION:

- Traditional recommender systems (RS) consider some factors, such as user's purchase records, product category, and geographic location. In this work, we propose a sentiment-based rating prediction method (RPS) to improve prediction accuracy in recommender systems

1.4 SCOPE OF THE PROJECT:

- At last, we fuse three factors-user sentiment similarity, interpersonal sentimental influence, and item's reputation similarity into our recommender system to make an accurate rating prediction

1. LITERATURE SURVEY

In this section, we survey recent work related to our approach. Firstly, we review some approaches based on collaborative filtering (CF). Then, we review the often utilized rating prediction/recommendation methods based on matrix factorization. Also, the review based approaches as well as the sentiment mining and applications are provided in detail.

A. Collaborative Filtering:

The task of CF is to predict user preferences for the unrated items, after which a list of most preferred items can be recommended to users. To improve recommendation performance, many CF algorithms have been proposed. One of the most well known CF algorithms is the user-based CF algorithm proposed. The basic idea is that people expressed similar preferences in the past will prefer to buy similar items in the future. Tso-Sutter et al. propose a generic method that allows tags to be incorporated to standard CF algorithms and to fuse the 3-dimensional correlations between users, items and tags. Moreover, item-based CF algorithm produces the rating from a user to an item based on the average ratings of similar or correlated items by the same user. It obtains better performance in computing the similarity between items. Gao et al. propose a review expert collaborative recommendation algorithm based on the assumption that those projects/experts with similar topics have similar feature vectors. Fletcher et al. propose a CF-based service This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2016.2575738, IEEE Transactions on Multimedia recommendation method that considers users' personalized preferences on nonfunctional attributes.

B. Matrix Factorization based Approaches:

1) Basic Matrix Factorization:

Matrix factorization is one of the most popular approaches for low-dimensional matrix decomposition. Here, we review the Basic MF. The rating matrix $R \in \boldsymbol{R}m \times n$ (m is the number of users and n is the number of items) can be predicted according to Eq. (1), where $Uu \in \boldsymbol{U}m \times k$ denotes the user Potential Eigen vectors matrix and $Pi \in \boldsymbol{P}n \times k$ denotes item Potential Eigen vectors matrix, and k is the dimension of the vectors. $R\hat{}\, u$, denotes the predicted objective star level of item i, $\bar{R}$ denotes the average value of all ratings. $R\hat{}\, u,i = \bar{R} + UuPiT$ (1) We learn Potential Eigen vectors of users and items on the observed rating data by minimizing the objective function. The objective function Ψ is defined as follows: $\Psi(\boldsymbol{R}, \boldsymbol{U}, \boldsymbol{P}) = 1\ 2\sum (Ru,i - R\hat{}\, u,i)\ 2\ u,i + \lambda\ 2\ (\|\boldsymbol{U}\|F\ 2 + \|\boldsymbol{P}\|F\ 2\ )$ (2) where $\|\boldsymbol{X}\|F$ is the Frobenius norm of matrix X, which is utilized to avoid over-fitting. The optimization of the objective function can be solved by gradient descent method [8].

2) Social Recommendation:

Some matrix factorization based social recommendations are proposed to solve the "cold start" problems. Jamali et al.[4] explore a matrix factorization based approach for recommendation in social networks. They incorporate the mechanism of trust propagation into the recommendation model. Trust propagation has been shown to be a crucial factor in social network analysis and in trust-based recommendation. Yang et al.[2] propose the concept of "Trust Circles" in social networks. Their model outperforms the Basic MF[1] and Social MF[4]. The trusted value between users is represented by a matrix S, and directed and weighted social relationship of user u with user v is represented by a positive value $Su,c* \in$ [0,1]. The basic idea is that the user latent feature should be similar to the average of his/her friends' latent features with weight of $Su,vc*$ in category c. Except for the factor of interpersonal influence in , Jiang et al.[3] propose another important factor, the individual preference. They conduct experiments on Renren dataset and TencentWeibo dataset in China, and the results demonstrate the significance of social contextual factors (individual preference and interpersonal influence) in their model. Qian et al.propose a personalized recommender model (PRM) combing with user interpersonal interest similarity, interpersonal influence and personal interest factor. They make use of

categories of products, and user personal interest is the main contributions. Wang et al. propose to use social propagation simulation and content similarity analysis to update the user-content matrix. They also construct a joint social-content space to measure the relevance between users and videos, which provides a high accuracy for video importing and re-sharing recommendation. However, some websites do not always offer structured information, and all of these methods do not leverage users' unstructured information, i.e. reviews. In addition, there also remain a few questions: some users may have no social relation with each other or even worse, explicit social networks information is not always available and it is difficult to provide a good prediction for each user. In this paper, we elaborate the sentiment factor to improve social recommendation.

C. Reviews based Applications

There are also many reviews based work for the task of recommendation. Qu et al. propose a bag-of-opinions model to predict a user's numeric rating in a product review. And they develop a constrained ridge regression method for learning scores of opinions. Wang et al. propose a review rating prediction method by incorporating the social relations of a reviewer. In addition, they classify the social relations of reviewers into strong social relation and ordinary social relation. Zhang et al. incorporate various product review factors including content related to product quality, time of the review, product durability and historically older positive customer reviews. They present a product ranking model that applies weights to product review factors to calculate the ranking score. Ling et al. propose a unified model that combines content-based collaborative filtering, and harnessing the information of both ratings and reviews. Luo et al. define and solve a new problem: aspect identification and rating, together with overall rating prediction in unrated reviews. They propose a LDA-style topic model which generates ratable aspects over sentiment and associates modifiers with ratings.

D. Sentiment based Applications

Sentiment analysis can be conducted on three different levels: review-level, sentence-level, and phrase-level. Review-level analysis and sentence- level analysis attempt to classify the sentiment of a whole review to one of the predefined sentiment polarities, including positive, negative and sometimes neutral. While phrase-level analysis attempt to extract the sentiment polarity of each feature that a user expresses his/her attitude to the specific feature of a specific product. The main task of phrase-level sentiment analysis is the construction of sentiment

lexicon. Pang et al. propose a context insensitive evaluative lexical method. However, they can not deal with the mismatch between the base valence of the term and the author's usage. Polanyi et al. describe how the base attitudinal valence of a lexical item is modified by lexical and discourse context and propose a simple implementation for some contextual shifters. They calculate user sentiment based on a finer grained method on all levels. Taboada et al. present a semantic orientation calculator which uses dictionaries of words annotated with their semantic orientation (polarity and strength), and incorporates intensification and negation. Lu et al. propose an optimization framework that provides a unified and principled way to combine different sources of information for learning a context- dependent sentiment lexicon. The proposed framework is quite general and applicable for opinionated text collection in any domain. Wang et al. analyze user opinions about an entity in a review at the level of topical aspects. They discover each individual reviewer's latent opinion on each aspect when forming the overall judgment of the entity.

2. SYSTEM ANALYSIS

EXISTING SYSTEM:

- Sentiment analysis can be conducted on three different levels: review-level, sentence-level, and phrase-level.
- Review-level analysis and sentence-level analysis attempt to classify the sentiment of a whole review to one of the predefined sentiment polarities, including positive, negative and sometimes neutral.
- While phrase-level analysis attempt to extract the sentiment polarity of each feature that a user expresses his/her attitude to the specific feature of a specific product.
- Zhang *et al.* propose a self-supervised and lexicon-based sentiment classification approach to determine sentiment polarity of a review that contains both textual words and emoticons. And they use sentiment for recommendation.
- Lee *et al.* propose a recommender system using the concept of Experts to find both novel and relevant recommendations. By analyzing the user ratings, they can recommend special experts to a target user based on the user population.

DISADVANTAGES OF EXISTING SYSTEM:

- The existing work mainly focuses on classifying users into binary sentiment (i.e. positive or negative), and they do not go further in mining user's sentiment.
- The existing approaches mainly leverage product category information or tag information to study the interpersonal influence.
- These methods are all restricted on the structured data, which is not always available on some websites. However, user reviews can provide us ideas in mining interpersonal inference and user preferences.

PROPOSED SYSTEM:

- We propose a sentiment-based rating prediction method in the framework of matrix factorization. In our work, we make use of social users' sentiment to infer ratings.
- First, we extract product features from user reviews. Then, we find out the sentiment words, which are used to describe the product features. Besides, we leverage sentiment dictionaries to calculate sentiment of a specific user on an item/product.
- The main contributions of our approach are as follows:
- We propose a user sentimental measurement approach, which is based on the mined sentiment words and sentiment degree words from user reviews.
- We make use of sentiment for rating prediction. User sentiment similarity focuses on the user interest preferences. User sentiment influence reflects how the sentiment spreads among the trusted users. Item reputation similarity shows the potential relevance of items.
- We fuse the three factors: user sentiment similarity, interpersonal sentimental influence, and item reputation similarity into a probabilistic matrix factorization framework to carry out an accurate recommendation. The experimental results and discussions show that user's social sentiment that we mined is a key factor in improving rating prediction performances.

ADVANTAGES OF PROPOSED SYSTEM:

- In our paper, we not only mine social user's sentiment, but also explore interpersonal sentimental influence and item's reputation. Finally, we take all of them into the recommender system.
- The purpose of our approach is to find effective clues from reviews and predict social users' ratings.
- We fuse user sentiment similarity, inter personal sentiment influence, and item reputation similarity into a unified matrix factorization frame work to achieve the rating prediction task.

SYSTEM SPECIFICATION:

HARDWARE REQUIREMENTS:

- System : Pentium IV 2.4 GHz.
- Hard Disk : 40 GB.
- Floppy Drive : 1.44 Mb.
- Monitor : 15 VGA Colour.
- Mouse : Logitech.
- Ram : 512 Mb.

SOFTWARE REQUIREMENTS:

- Operating system : Windows XP/7 and above.
- Coding Language : JAVA/J2EE
- Data Base : MYSQL

MODULES:

- Data preprocessing for LDA
- Extracting product features
- User Sentimental Measurement
- Sentiment Evaluation

MODULES DESCSRIPTION:

Data preprocessing for LDA

- In the first module we develop the data preprocessing for LDA. We have collected rating data set from http://www.yelp.com. We give this dataset as the input to our system. The data set are product items dataset, user ratings dataset and user feedback dataset. We have to separate dataset feedback and ratings based. The purpose of our approach is to find effective clues from reviews and predict social users' ratings. In this module, we firstly extract product features from user review corpus, and then we introduce the method of identifying social users' sentiment.
- The dataset are categories into three factors. 1. Item's reputation 2.interpersonal sentimental influence 1.item's reputation 3.user sentiment similarity.

Extracting product features

- In this module, we extract product features from textual reviews using LDA. We mainly want to get the product features including some named entities and some product/item/service attributes. LDA is a Bayesian model, which is utilized to model the relationship of reviews, topics and words
- To construct the vocabulary, we firstly regard each user's review as a collection of words without considering the order. Then we filter out "Stop Words", "Noise Words" and sentiment words, sentiment degree words, and negation words.
- A stop word can be identified as a word that has the same likelihood of occurring in those documents not relevant to a query as in those documents relevant to the query. For example, the "Stop Words" could be some prepositions, articles, and pronouns etc.. After words filtering, the input text is clear and without much interference for generating topics. All the unique words are constructed in the vocabulary V, each word has a label.

- From each topic, we have some frequent words. However, we need to filter the noisy features from the candidate set based on their co-occurrence with adjective words and their frequencies in background corpus.

User Sentimental Measurement

- We extend HowNet Sentiment Dictionary3 to calculate social user's sentiment on items. In this module, we merge the positive sentiment words list and positive evaluation words list of HowNet Sentiment Dictionary into one list, and named it as POS-Words; also, we merge the negative sentiment words list and negative evaluation words list of HowNet Sentiment Dictionary into one list, and named it as NEG-Words.
- In this module we develop five different levels in sentiment degree dictionary (**SDD**), which has 128 words in total. There are 52 words in the **Level-1,** which means the highest degree of sentiment, such as the words "*most*", and "*best*". And 48 words in the **Level-2**, which means higher degree of sentiment, such as the words "*better*", and "*very*". There are 12 words in the **Level-3**, such as the words "*more*", and "*such*". There are 9 words in the **Level-4,** such as the words "*a little*", "*a bit*", and "*more or less*". And there are 7 words in the **Level-5,** such as the words "*less*", "*bit*", and "*not very*". Also, we built the negation dictionary (**ND**) by collecting frequently-used negative prefix words, such as "*no*", "*hardly*", "*never*", etc. These words are used to reverse the polarity of sentiment words.

Sentiment Evaluation

- We firstly divide the original review into several clauses by the punctuation mark. Then for each clause, we firstly look up the dictionary **SD** to find the sentiment words before the product features. A positive word is initially assigned with the score +1.0, while a negative word is assigned with the score −1.0. Secondly, we find out the sentiment degree words based on the dictionary **SDD** and take the sentiment degree words into consideration to strengthen sentiment for the found sentiment words. Finally, we check the negative prefix words based on the dictionary **ND** and add a negation check coefficient that has a default value of +1.0. If the sentiment word is preceded by an odd number of negative prefix words within the specified zone, we reverse the sentiment polarity, and the coefficient is set to −1.0.

- Each sentiment factor is described as follows:
    - User Sentiment Similarity, 2) Interpersonal Sentiment Influence, 3) Item Reputation Similarity
- We compare the performance of our method with the existing models on Yelp dataset. In the objective function of RPS, *k* is the dimension of user and item latent feature vectors.
- The experimental results show the high accuracy of RPS. Meanwhile, we demonstrate the importance of social friend factors (i.e. CircleCon2b, PRM) and explicit features (i.e. EFM) in a recommender system.

SOFTWARE MODEL:

The Waterfall Model was the first Process Model to be introduced. It is also referred to as a **linear-sequential life cycle model**. It is very simple to understand and use. In a waterfall model, each phase must be completed before the next phase can begin and there is no overlapping in the phases.

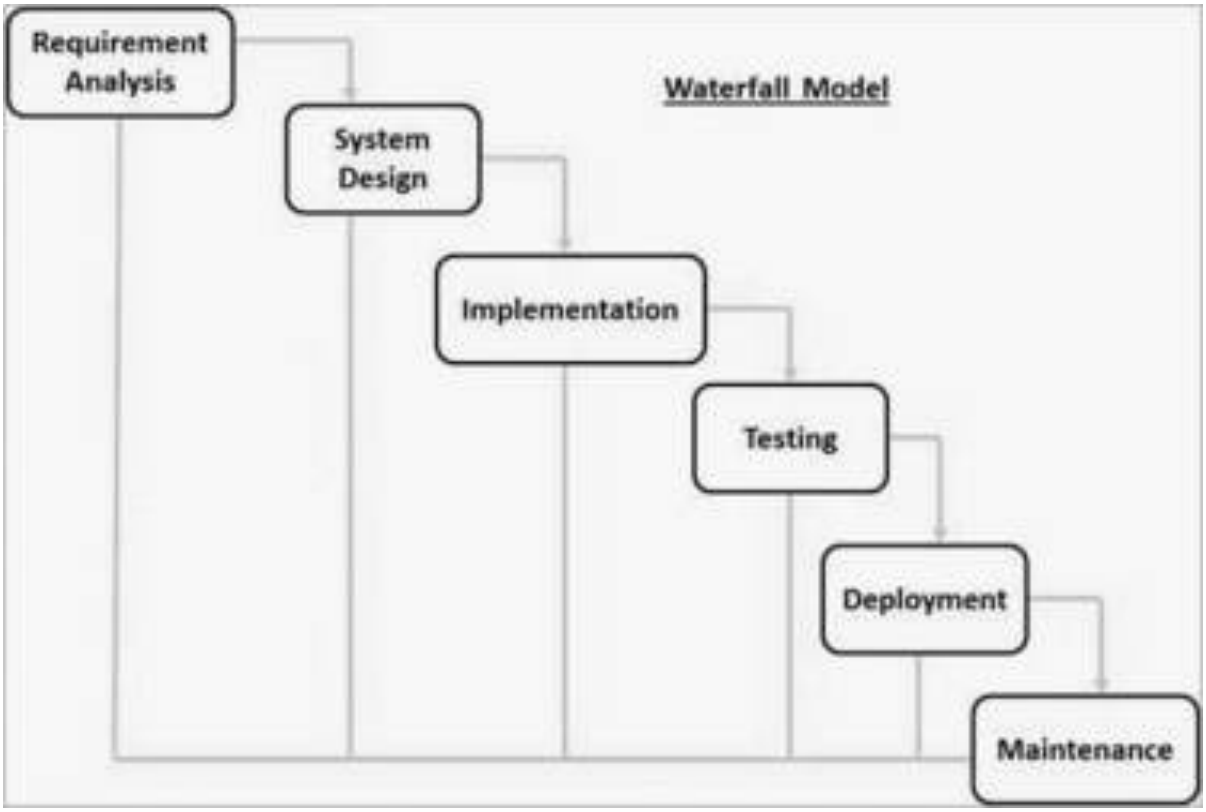
The Waterfall model is the earliest SDLC approach that was used for software development.

The waterfall Model illustrates the software development process in a linear sequential flow. This means that any phase in the development process begins only if the previous phase is complete. In this waterfall model, the phases do not overlap.

Waterfall Model - Design

Waterfall approach was first SDLC Model to be used widely in Software Engineering to ensure success of the project. In "The Waterfall" approach, the whole process of software development is divided into separate phases. In this Waterfall model, typically, the outcome of one phase acts as the input for the next phase sequentially.

The following illustration is a representation of the different phases of the Waterfall Model.



The sequential phases in Waterfall model are −

- **Requirement Gathering and analysis** − All possible requirements of the system to be developed are captured in this phase and documented in a requirement specification document.
- **System Design** − The requirement specifications from first phase are studied in this phase and the system design is prepared. This system design helps in specifying hardware and system requirements and helps in defining the overall system architecture.
- **Implementation** − With inputs from the system design, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality, which is referred to as Unit Testing.
- **Integration and Testing** − All the units developed in the implementation phase are integrated into a system after testing of each unit. Post integration the entire system is tested for any faults and failures.

- **Deployment of system** − Once the functional and non-functional testing is done; the product is deployed in the customer environment or released into the market.
- **Maintenance** − There are some issues which come up in the client environment. To fix those issues, patches are released. Also to enhance the product some better versions are released. Maintenance is done to deliver these changes in the customer environment.

FEASIBILITY STUDY:

The first and foremost strategy for development of a project starts from the thought of designing a mail enabled platform for a small firm in which it is easy and convenient of sending and receiving messages, there is a search engine,address book and also including some entertaining games. When it is approved by the organization and our project guide the first activity, i.e., preliminary investigation begins. The activity has three parts:

- **Request Clarification**
- **Feasibility Study**
- **Request Approval**

REQUEST CLARIFICATION:

After the approval of the request to the organization and project guide, with an investigation being considered, the project request must be examined to determine precisely what the system requires.

Here our project is basically meant for users within the company whose systems can be interconnected by the Local Area Network (LAN). In today's busy schedule man need everything should be provided in a readymade manner. So taking into consideration of the vastly use of the net in day to day life, the corresponding development of the portal came into existence.

FEASIBILITY STUDY:

An important outcome of preliminary investigation is the determination that the system request is feasible. This is possible only if it is feasible within limited resource and time. The different

feasibilities that have to be analyzed are

- **Operational Feasibility**
- **Economic Feasibility**
- **Technical Feasibility**

Operational Feasibility:

Operational Feasibility deals with the study of prospects of the system to be developed. This system operationally eliminates all the tensions of the Admin and helps him in effectively tracking the project progress. This kind of automation will surely reduce the time and energy, which previously consumed in manual work. Based on the study, the system is proved to be operationally feasible.

Economic Feasibility:

Economic Feasibility or Cost-benefit is an assessment of the economic justification for a computer based project. As hardware was installed from the beginning & for lots of purposes thus the cost on project of hardware is low. Since the system is a network based, any number of employees connected to the LAN within that organization can use this tool from at anytime. The Virtual Private Network is to be developed using the existing resources of the organization. So the project is economically feasible.

Technical Feasibility:

According to Roger S. Pressman, Technical Feasibility is the assessment of the technical resources of the organization. The organization needs IBM compatible machines with a graphical web browser connected to the Internet and Intranet. The system is developed for platform Independent environment. Java Server Pages, JavaScript, HTML, SQL server and WebLogic Server are used to develop the system. The technical feasibility has been carried out. The system is technically feasible for development and can be developed with the existing facility

REQUEST APPROVAL

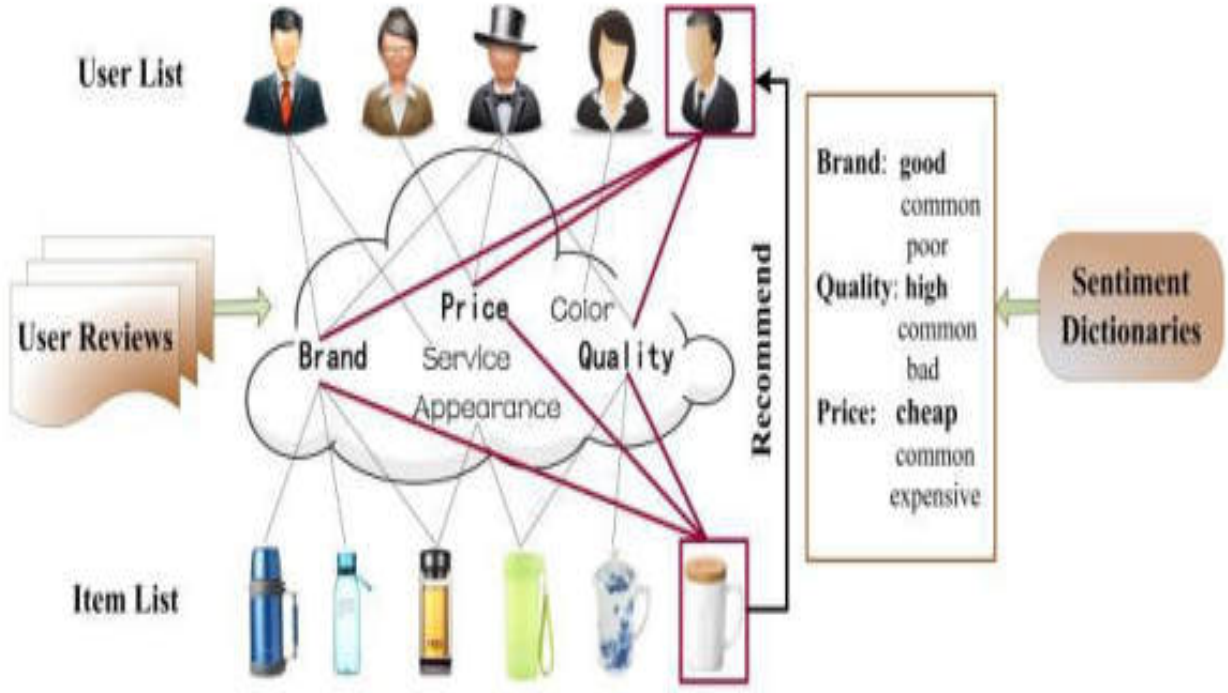
Not all request projects are desirable or feasible. Some organization receives so

many project requests from client users that only few of them are pursued. However, those projects that are both feasible and desirable should be put into schedule. After a project request is [i]approved, it cost, priority, completion time and personnel requirement is estimated and used to determine where to add it to any project list. Truly speaking, the approval of those above factors, development works can be launched.

3. SYSTEM DESIGN

Systems design is the process of defining elements of a system like modules, architecture, components and their interfaces and data for a system based on the specified requirements. It is the process of defining, developing and designing systems which satisfies the specific needs and requirements of a business or organization.

4.1 Architecture Diagram:



4.2 UML DIAGRAMS:

UML (Unified Modeling Language) is a general-purpose, graphical modeling language in the field of Software Engineering. UML is used to specify, visualize, construct, and document the artifacts (major elements) of the software system. It was initially developed by Grady Booch, Ivar Jacobson, and James Rumbaugh in 1994-95 at Rational software, and its further development was carried out through 1996. In 1997, it got adopted as a standard by the Object Management Group.

UML (Unified Modeling Language) is a general-purpose, graphical modeling language in the field of Software Engineering. UML is used to specify, visualize, construct, and document the artifacts (major elements) of the software system. It was initially developed by Grady Booch, Ivar Jacobson, and James Rumbaugh in 1994-95 at Rational software, and its further development was carried out through 1996. In 1997, it got adopted as a standard by the Object Management Group.

What is UML

The UML stands for Unified modeling language, is a standardized general-purpose visual modeling language in the field of Software Engineering. It is used for specifying, visualizing, constructing, and documenting the primary artifacts of the software system. It helps in designing and characterizing, especially those software systems that incorporate the concept of Object orientation. It describes the working of both the software and hardware systems.

The UML was developed in 1994-95 by Grady Booch, Ivar Jacobson, and James Rumbaugh at the Rational Software. In 1997, it got adopted as a standard by the Object Management Group (OMG).

The Object Management Group (OMG) is an association of several companies that controls the open standard UML. The OMG was established to build an open standard that mainly supports the interoperability of object-oriented systems. It is not restricted within the boundaries, but it can also be utilized for modeling the non-software systems. The OMG is best recognized for the Common Object Request Broker Architecture (CORBA) standards.

Goals of UML

- Since it is a general-purpose modeling language, it can be utilized by all the modelers.
- UML came into existence after the introduction of object-oriented concepts to systemize and consolidate the object-oriented development, due to the absence of standard methods at that time.
- The UML diagrams are made for business users, developers, ordinary people, or anyone who is looking forward to understand the system, such that the system can be software or non-software.
- Thus it can be concluded that the UML is a simple modeling approach that is used to model all the practical systems.

Characteristics of UML

The UML has the following features:

- It is a generalized modeling language.
- It is distinct from other programming languages like C++, Python, etc.
- It is interrelated to object-oriented analysis and design.
- It is used to visualize the workflow of the system.
- It is a pictorial language, used to generate powerful modeling artifacts.

Conceptual Modeling

Before moving ahead with the concept of UML, we should first understand the basics of the conceptual model.

A conceptual model is composed of several interrelated concepts. It makes it easy to understand the objects and how they interact with each other. This is the first step before drawing UML diagrams.

Following are some object-oriented concepts that are needed to begin with UML:

- **Object:** An object is a real world entity. There are many objects present within a single system. It is a fundamental building block of UML.
- **Class:** A class is a software blueprint for objects, which means that it defines the variables and methods common to all the objects of a particular type.
- **Abstraction:** Abstraction is the process of portraying the essential characteristics of an object to the users while hiding the irrelevant information. Basically, it is used to envision the functioning of an object.
- **Inheritance:** Inheritance is the process of deriving a new class from the existing ones.
- **Polymorphism:** It is a mechanism of representing objects having multiple forms used for different purposes.
- **Encapsulation:** It binds the data and the object together as a single unit, enabling tight coupling between them.

You can also create your own set of diagrams to meet your requirements. Diagrams are generally made in an incremental and iterative way.

There are two broad categories of diagrams and they are again divided into subcategories −

- Structural Diagrams
- Behavioral Diagrams

Structural Diagrams

The structural diagrams represent the static aspect of the system. These static aspects represent those parts of a diagram, which forms the main structure and are therefore stable.

These static parts are represented by classes, interfaces, objects, components, and nodes. The four structural diagrams are −

- Class diagram

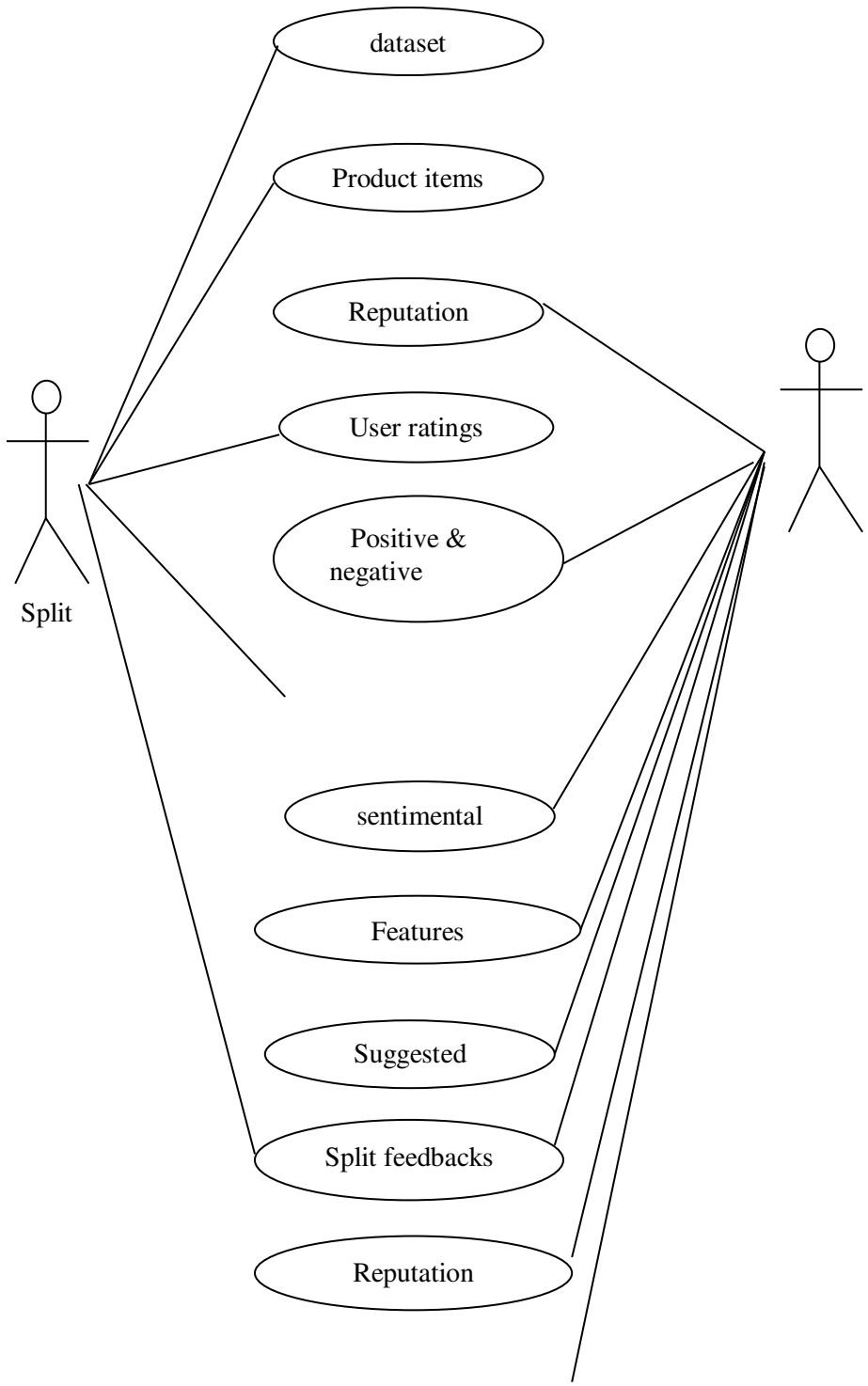
- Object diagram
- Component diagram
- Deployment diagram

BEHAVIOURAL DIAGRAMS:

- Behavioral diagrams basically capture the dynamic aspect of a system. Dynamic aspect can be further described as the changing/moving parts of a system.
- UML has the following five types of behavioral diagrams −

- Use case diagram
- Sequence diagram
- Collaboration diagram
- Statechart diagram
- Activity diagram

USECASE DIAGRAM:



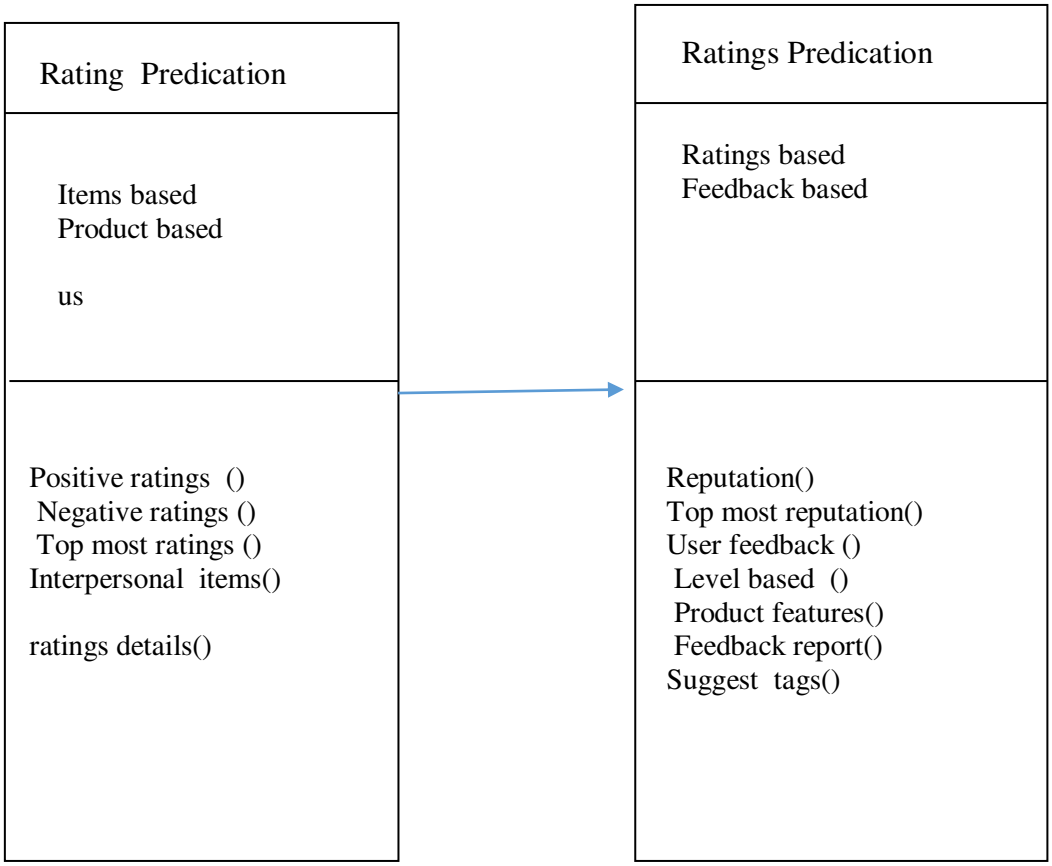
CLASS DIAGRAM:

Model, objects are entities that combine state (i.e., data), behavior (i.e., procedures, or methods) and identity (unique existence among all other objects). The structure and behavior of an object are defined by a class, which is a definition, or blueprint, of all objects of a specific type. An object must be explicitly created based on a class and an object thus created is considered to be an instance of that class. An object is similar to a structure, with the addition of method pointers, member access control, and an implicit data member which locates instances of the class (i.e. actual objects of that class) in the class hierarchy (essential for runtime inheritance features)

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, and the relationships between the classes.

The class diagram is the main building block in the object oriented modeling. It is used both for general conceptual modeling of the semantics of the application, and for detailed modeling translating the models into programming code. The classes in a class diagram represent both the main objects and or interactions in the application and the objects to be programmed. In the class diagram these classes are represented with boxes which contain the two parts:

- The upper part holds the name of the class.
- The middle part contains the attributes of the class.
- The lower part contains the operations of the class.


Rating Predication
Items based
Product based
us
Positive ratings ()
Negative ratings ()
Top most ratings ()
Interpersonal items()
ratings details()
Ratings Predication
Ratings based
Feedback based
Reputation()
Top most reputation()
User feedback ()
Level based ()
Product features()
Feedback report()
Suggest tags()

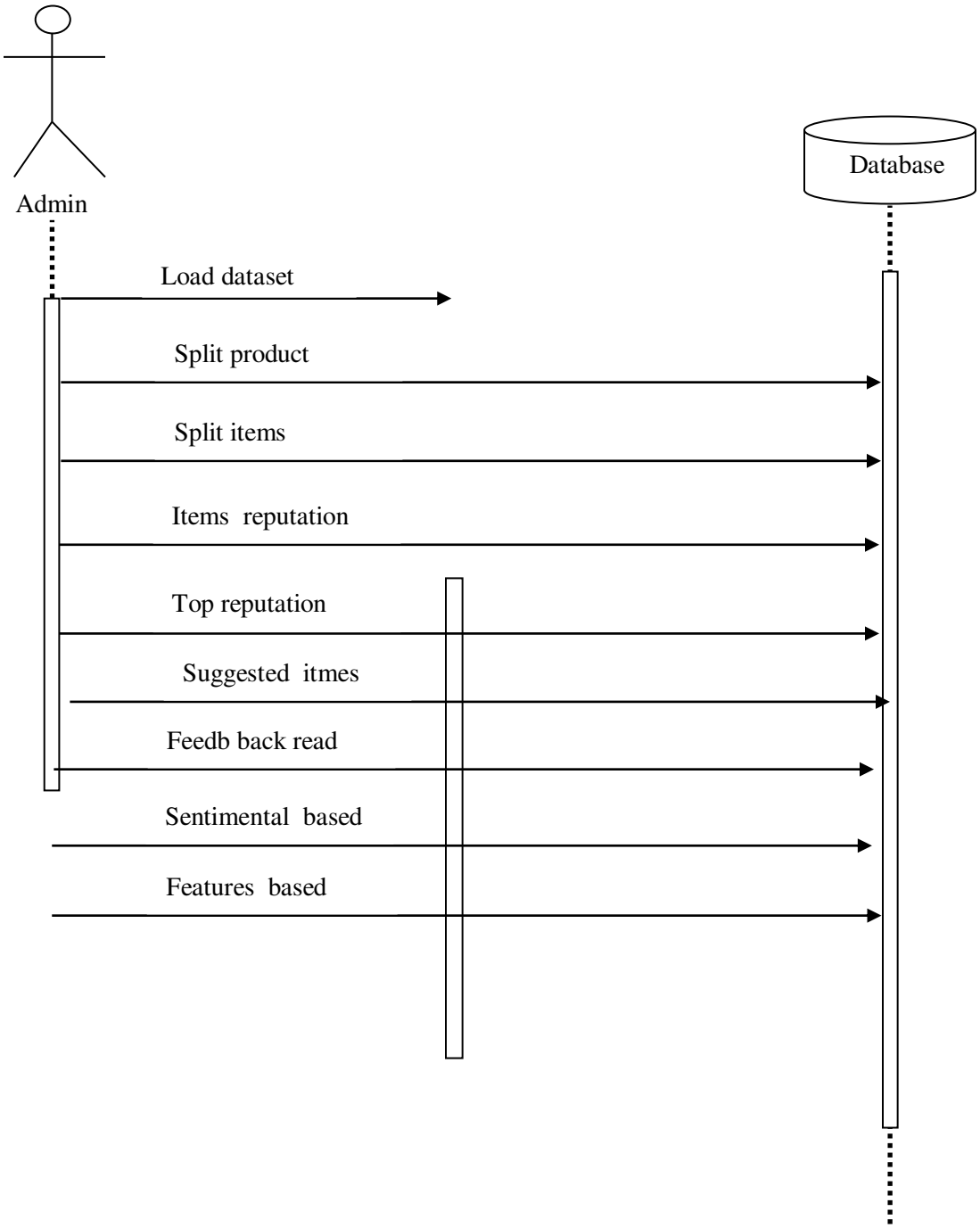
Sequence Diagram

The sequence diagram represents the flow of messages in the system and is also termed as an event diagram. It helps in envisioning several dynamic scenarios. It portrays the communication between any two lifelines as a time-ordered sequence of events, such that these lifelines took part at the run time. In UML, the lifeline is represented by a vertical bar, whereas the message flow is represented by a vertical dotted line that extends across the bottom of the page. It incorporates the iterations as well as branching.

Purpose of a Sequence Diagram

1. To model high-level interaction among active objects within a system.
2. To model interaction among objects inside a collaboration realizing a use case.
3. It either models generic interactions or some certain instances of interaction.

SEQUENCE DIAGRAM



UML Activity Diagram

In UML, the activity diagram is used to demonstrate the flow of control within the system rather than the implementation. It models the concurrent and sequential activities.

The activity diagram helps in envisioning the workflow from one activity to another. It put emphasis on the condition of flow and the order in which it occurs. The flow can be sequential, branched, or concurrent, and to deal with such kinds of flows, the activity diagram has come up with a fork, join, etc.

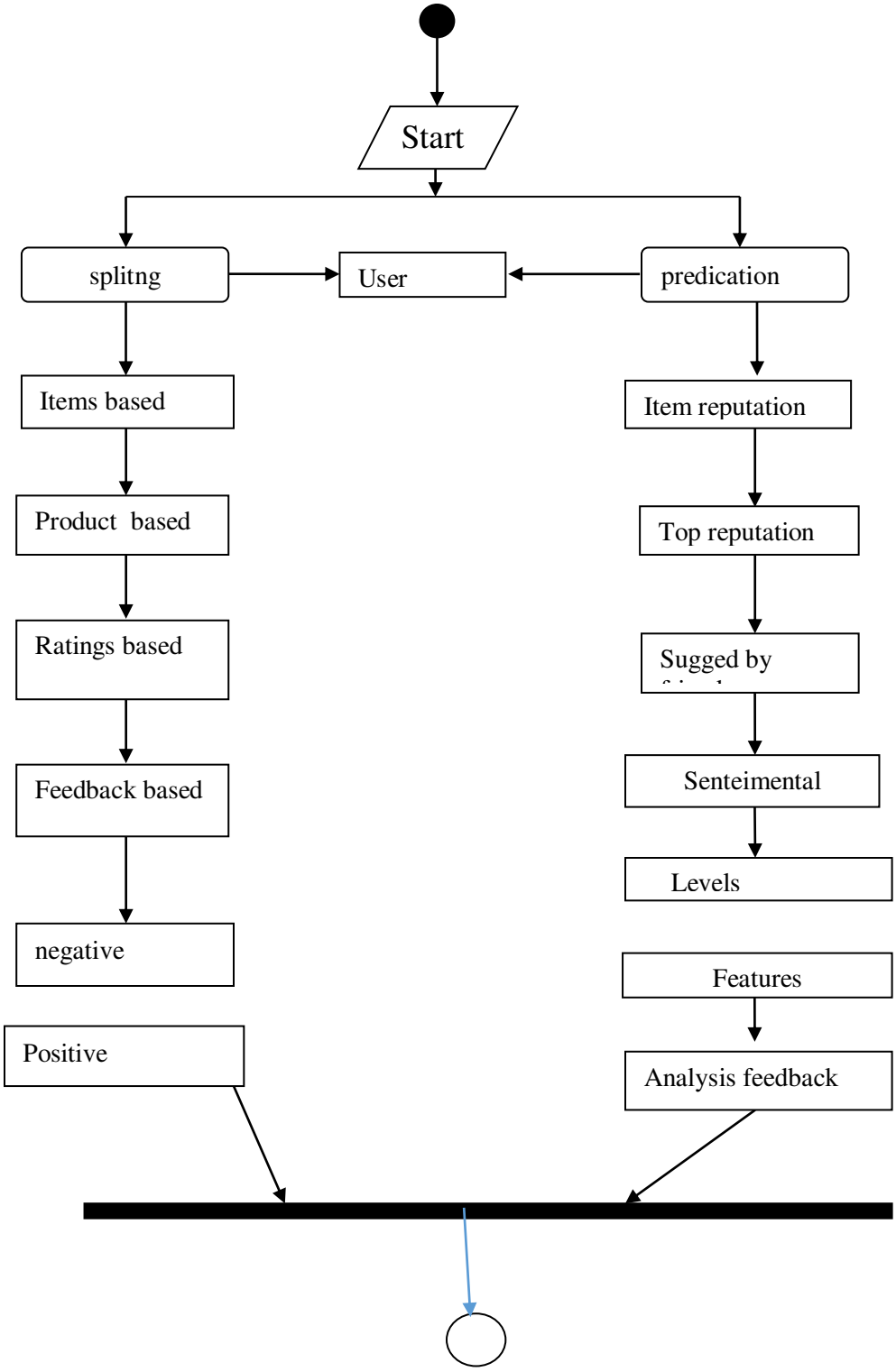
It is also termed as an object-oriented flowchart. It encompasses activities composed of a set of actions or operations that are applied to model the behavioral diagram.

Why use Activity Diagram?

An event is created as an activity diagram encompassing a group of nodes associated with edges. To model the behavior of activities, they can be attached to any modeling element. It can model use cases, classes, interfaces, components, and collaborations.

It mainly models processes and workflows. It envisions the dynamic behavior of the system as well as constructs a runnable system that incorporates forward and reverse engineering. It does not include the message part, which means message flow is not represented in an activity diagram.

It is the same as that of a flowchart but not exactly a flowchart itself. It is used to depict the flow between several activities.


Start
splitng
User
predication
Items based
Product based
Ratings based
Feedback based
negative
Positive
Item reputation
Top reputation
Sugged by
Senteimental
Levels
Features
Analysis feedback

5. IMPLEMENTATIONS

5.1 TECHNOLOGIES:

Java Technology

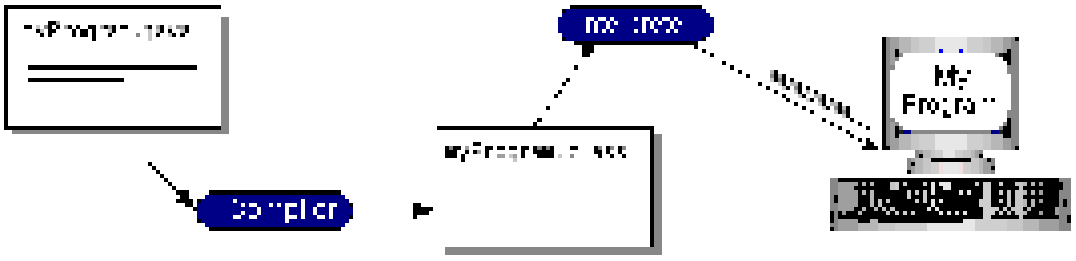
Java technology is both a programming language and a platform.

The Java Programming Language

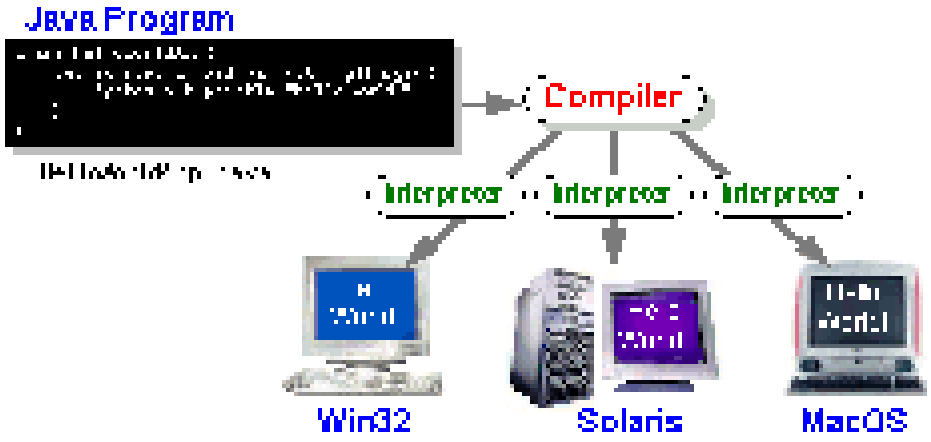
The Java programming language is a high-level language that can be characterized by all of the following buzzwords:

- Simple
- Architecture neutral
- Object oriented
- Portable
- Distributed
- High performance
- Interpreted
- Multithreaded
- Robust
- Dynamic
- Secure

With most programming languages, you either compile or interpret a program so that you can run it on your computer. The Java programming language is unusual in that a program is both compiled and interpreted. With the compiler, first you translate a program into an intermediate language called *Java byte codes* —the platform-independent codes interpreted by the interpreter on the Java platform. The interpreter parses and runs each Java byte code instruction on the computer. Compilation happens just once; interpretation occurs each time the program is executed. The following figure illustrates how this works.



You can think of Java byte codes as the machine code instructions for the *Java Virtual Machine* (Java VM). Every Java interpreter, whether it's a development tool or a Web browser that can run applets, is an implementation of the Java VM. Java byte codes help make "write once, run anywhere" possible. You can compile your program into byte codes on any platform that has a Java compiler. The byte codes can then be run on any implementation of the Java VM. That means that as long as a computer has a Java VM, the same program written in the Java programming language can run on Windows 2000, a Solaris workstation, or on an iMac.



The Java Platform

A *platform* is the hardware or software environment in which a program runs. We've already mentioned some of the most popular platforms like Windows 2000, Linux, Solaris, and MacOS. Most platforms can be described as a combination of the operating system and hardware. The Java platform differs from most other platforms in that it's a software-only platform that runs on top of other hardware-based platforms.

The Java platform has two components:

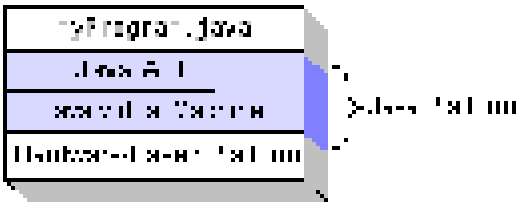
- The *Java Virtual Machine* (Java VM)

- The *Java Application Programming Interface* (Java API)

You've already been introduced to the Java VM. It's the base for the Java platform and is ported onto various hardware-based platforms.

The Java API is a large collection of ready-made software components that provide many useful capabilities, such as graphical user interface (GUI) widgets. The Java API is grouped into libraries of related classes and interfaces; these libraries are known as *packages*. The next section, What Can Java Technology Do? Highlights what functionality some of the packages in the Java API provide.

The following figure depicts a program that's running on the Java platform. As the figure shows, the Java API and the virtual machine insulate the program from the hardware.



Native code is code that after you compile it, the compiled code runs on a specific hardware platform. As a platform-independent environment, the Java platform can be a bit slower than native code. However, smart compilers, well-tuned interpreters, and just-in-time byte code compilers can bring performance close to that of native code without threatening portability.

What Can Java Technology Do?

The most common types of programs written in the Java programming language are *applets* and *applications*. If you've surfed the Web, you're probably already familiar with applets. An applet is a program that adheres to certain conventions that allow it to run within a Java-enabled browser.

However, the Java programming language is not just for writing cute, entertaining applets for the Web. The general-purpose, high-level Java programming language is also a powerful software platform. Using the generous API, you can write many types of programs.

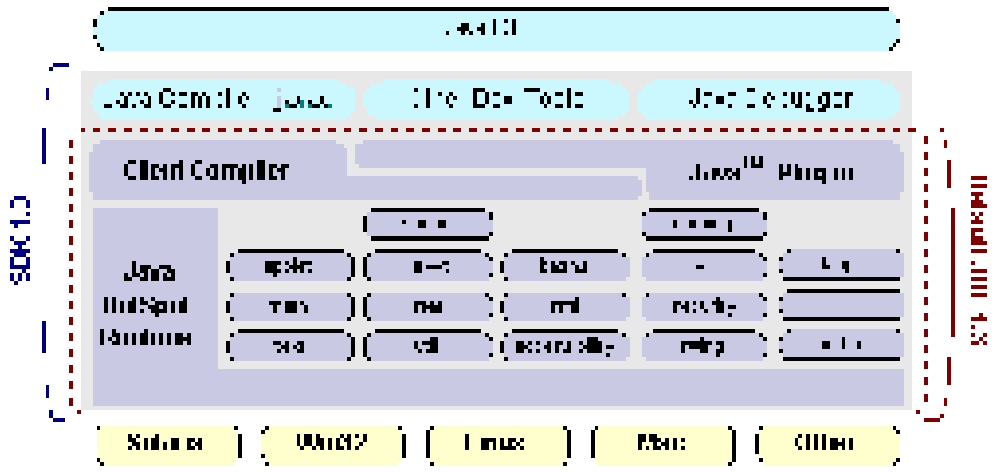
An application is a standalone program that runs directly on the Java platform. A special kind of application known as a *server* serves and supports clients on a network. Examples

of servers are Web servers, proxy servers, mail servers, and print servers. Another specialized program is a *servlet*. A servlet can almost be thought of as an applet that runs on the server side. Java Servlets are a popular choice for building interactive web applications, replacing the use of CGI scripts. Servlets are similar to applets in that they are runtime extensions of applications. Instead of working in browsers, though, servlets run within Java Web servers, configuring or tailoring the server.

How does the API support all these kinds of programs? It does so with packages of software components that provides a wide range of functionality. Every full implementation of the Java platform gives you the following features:

- **The essentials**: Objects, strings, threads, numbers, input and output, data structures, system properties, date and time, and so on.
- **Applets**: The set of conventions used by applets.
- **Networking**: URLs, TCP (Transmission Control Protocol), UDP (User Data gram Protocol) sockets, and IP (Internet Protocol) addresses.
- **Internationalization**: Help for writing programs that can be localized for users worldwide. Programs can automatically adapt to specific locales and be displayed in the appropriate language.
- **Security**: Both low level and high level, including electronic signatures, public and private key management, access control, and certificates.
- **Software components**: Known as JavaBeans™, can plug into existing component architectures.
- **Object serialization**: Allows lightweight persistence and communication via Remote Method Invocation (RMI).
- **Java Database Connectivity (JDBC™)**: Provides uniform access to a wide range of relational databases.

The Java platform also has APIs for 2D and 3D graphics, accessibility, servers, collaboration, telephony, speech, animation, and more. The following figure depicts what is included in the Java 2 SDK.



How Will Java Technology Change My Life?

We can't promise you fame, fortune, or even a job if you learn the Java programming language. Still, it is likely to make your programs better and requires less effort than other languages. We believe that Java technology will help you do the following:

- **Get started quickly**: Although the Java programming language is a powerful object-oriented language, it's easy to learn, especially for programmers already familiar with C or C++.
- **Write less code**: Comparisons of program metrics (class counts, method counts, and so on) suggest that a program written in the Java programming language can be four times smaller than the same program in C++.
- **Write better code**: The Java programming language encourages good coding practices, and its garbage collection helps you avoid memory leaks. Its object orientation, its JavaBeans component architecture, and its wide-ranging, easily extendible API let you reuse other people's tested code and introduce fewer bugs.
- **Develop programs more quickly**: Your development time may be as much as twice as fast versus writing the same program in C++. Why? You write fewer lines of code and it is a simpler programming language than C++.
- **Avoid platform dependencies with 100% Pure Java**: You can keep your program portable by avoiding the use of libraries written in other languages. The 100% Pure Java™Product Certification Program has a repository of historical process manuals, white papers, brochures, and similar materials online.

- **Write once, run anywhere**: Because 100% Pure Java programs are compiled into machine-independent byte codes, they run consistently on any Java platform.
- **Distribute software more easily**: You can upgrade applets easily from a central server. Applets take advantage of the feature of allowing new classes to be loaded "on the fly," without recompiling the entire program.

ODBC

Microsoft Open Database Connectivity (ODBC) is a standard programming interface for application developers and database systems providers. Before ODBC became a *de facto* standard for Windows programs to interface with database systems, programmers had to use proprietary languages for each database they wanted to connect to. Now, ODBC has made the choice of the database system almost irrelevant from a coding perspective, which is as it should be. Application developers have much more important things to worry about than the syntax that is needed to port their program from one database to another when business needs suddenly change.

Through the ODBC Administrator in Control Panel, you can specify the particular database that is associated with a data source that an ODBC application program is written to use. Think of an ODBC data source as a door with a name on it. Each door will lead you to a particular database. For example, the data source named Sales Figures might be a SQL Server database, whereas the Accounts Payable data source could refer to an Access database. The physical database referred to by a data source can reside anywhere on the LAN.

The ODBC system files are not installed on your system by Windows 95. Rather, they are installed when you setup a separate database application, such as SQL Server Client or Visual Basic 4.0. When the ODBC icon is installed in Control Panel, it uses a file called ODBCINST.DLL. It is also possible to administer your ODBC data sources through a stand-alone program called ODBCADM.EXE. There is a 16-bit and a 32-bit version of this program and each maintains a separate list of ODBC data sources.

From a programming perspective, the beauty of ODBC is that the application can be written to use the same set of function calls to interface with any data source, regardless of the database vendor. The source code of the application doesn't change whether it talks to Oracle or SQL Server. We only mention these two as an example. There are ODBC drivers available for

several dozen popular database systems. Even Excel spreadsheets and plain text files can be turned into data sources. The operating system uses the Registry information written by ODBC Administrator to determine which low-level ODBC drivers are needed to talk to the data source (such as the interface to Oracle or SQL Server). The loading of the ODBC drivers is transparent to the ODBC application program. In a client/server environment, the ODBC API even handles many of the network issues for the application programmer.

The advantages of this scheme are so numerous that you are probably thinking there must be some catch. The only disadvantage of ODBC is that it isn't as efficient as talking directly to the native database interface. ODBC has had many detractors make the charge that it is too slow. Microsoft has always claimed that the critical factor in performance is the quality of the driver software that is used. In our humble opinion, this is true. The availability of good ODBC drivers has improved a great deal recently. And anyway, the criticism about performance is somewhat analogous to those who said that compilers would never match the speed of pure assembly language. Maybe not, but the compiler (or ODBC) gives you the opportunity to write cleaner programs, which means you finish sooner. Meanwhile, computers get faster every year.

JDBC

In an effort to set an independent database standard API for Java; Sun Microsystems developed Java Database Connectivity, or JDBC. JDBC offers a generic SQL database access mechanism that provides a consistent interface to a variety of RDBMSs. This consistent interface is achieved through the use of "plug-in" database connectivity modules, or *drivers*. If a database vendor wishes to have JDBC support, he or she must provide the driver for each platform that the database and Java run on.

To gain a wider acceptance of JDBC, Sun based JDBC's framework on ODBC. As you discovered earlier in this chapter, ODBC has widespread support on a variety of platforms. Basing JDBC on ODBC will allow vendors to bring JDBC drivers to market much faster than developing a completely new connectivity solution.

JDBC was announced in March of 1996. It was released for a 90 day public review that ended June 8, 1996. Because of user input, the final JDBC v1.0 specification was released soon after.

The remainder of this section will cover enough information about JDBC for you to know what it is about and how to use it effectively. This is by no means a complete overview of JDBC. That would fill an entire book.

JDBC Goals

Few software packages are designed without goals in mind. JDBC is one that, because of its many goals, drove the development of the API. These goals, in conjunction with early reviewer feedback, have finalized the JDBC class library into a solid framework for building database applications in Java.

The goals that were set for JDBC are important. They will give you some insight as to why certain classes and functionalities behave the way they do. The eight design goals for JDBC are as follows:

1. ***SQL Level API***

   The designers felt that their main goal was to define a SQL interface for Java. Although not the lowest database interface level possible, it is at a low enough level for higher-level tools and APIs to be created. Conversely, it is at a high enough level for application programmers to use it confidently. Attaining this goal allows for future tool vendors to "generate" JDBC code and to hide many of JDBC's complexities from the end user.

2. ***SQL Conformance***

   SQL syntax varies as you move from database vendor to database vendor. In an effort to support a wide variety of vendors, JDBC will allow any query statement to be passed through it to the underlying database driver. This allows the connectivity module to handle non-standard functionality in a manner that is suitable for its users.

3. ***JDBC must be implemental on top of common database interfaces***

   The JDBC SQL API must "sit" on top of other common SQL level APIs. This goal allows JDBC to use existing ODBC level drivers by the use of a software interface. This interface would translate JDBC calls to ODBC and vice versa.

4. ***Provide a Java interface that is consistent with the rest of the Java system***

   Because of Java's acceptance in the user community thus far, the designers feel that they should not stray from the current design of the core Java system.

5. ***Keep it simple***

This goal probably appears in all software design goal listings. JDBC is no exception. Sun felt that the design of JDBC should be very simple, allowing for only one method of completing a task per mechanism. Allowing duplicate functionality only serves to confuse the users of the API.

6. ***Use strong, static typing wherever possible***

Strong typing allows for more error checking to be done at compile time; also, less error appear at runtime.

7. ***Keep the common cases simple***

Because more often than not, the usual SQL calls used by the programmer are simple SELECT's, INSERT's, DELETE's and UPDATE's, these queries should be simple to perform with JDBC. However, more complex SQL statements should also be possible.

Finally we decided to proceed the implementation using Java Networking.

And for dynamically updating the cache table we go for MS Access database.

Java ha two things: a programming language and a platform.

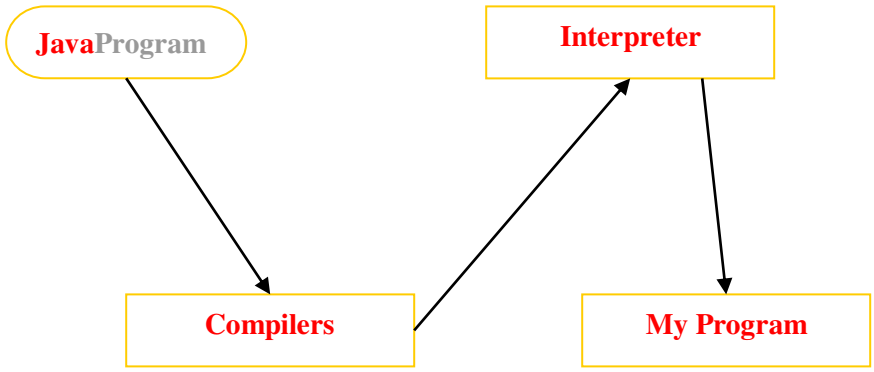
Java is a high-level programming language that is all of the following

Simple	Architecture-neutral
Object-oriented	Portable
Distributed	High-performance
Interpreted	multithreaded
Robust	Dynamic

Secure

Java is also unusual in that each Java program is both compiled and interpreted. With a compile you translate a Java program into an intermediate language called Java byte codes the platform-independent code instruction is passed and run on the computer.

Compilation happens just once; interpretation occurs each time the program is executed. The figure illustrates how this works.



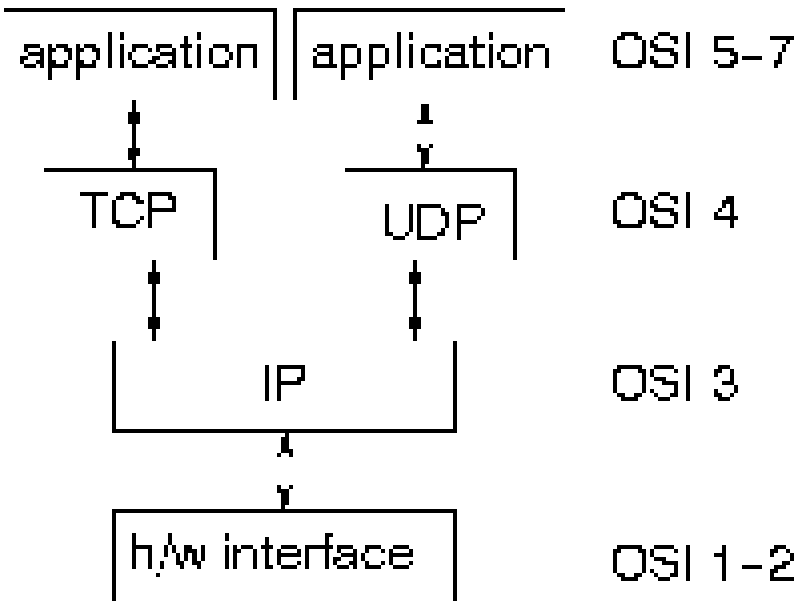
You can think of Java byte codes as the machine code instructions for the Java Virtual Machine (Java VM). Every Java interpreter, whether it's a Java development tool or a Web browser that can run Java applets, is an implementation of the Java VM. The Java VM can also be implemented in hardware.

Java byte codes help make “write once, run anywhere” possible. You can compile your Java program into byte codes on my platform that has a Java compiler. The byte codes can then be run any implementation of the Java VM. For example, the same Java program can run Windows NT, Solaris, and Macintosh.

Networking

TCP/IP stack

The TCP/IP stack is shorter than the OSI one:



TCP is a connection-oriented protocol; UDP (User Datagram Protocol) is a connectionless protocol.

IP datagram’s

The IP layer provides a connectionless and unreliable delivery system. It considers each datagram independently of the others. Any association between datagram must be

supplied by the higher layers. The IP layer supplies a checksum that includes its own header. The header includes the source and destination addresses. The IP layer handles routing through an Internet. It is also responsible for breaking up large datagram into smaller ones for transmission and reassembling them at the other end.

UDP

UDP is also connectionless and unreliable. What it adds to IP is a checksum for the contents of the datagram and port numbers. These are used to give a client/server model - see later.

TCP

TCP supplies logic to give a reliable connection-oriented protocol above IP. It provides a virtual circuit that two processes can use to communicate.

Internet addresses

In order to use a service, you must be able to find it. The Internet uses an address scheme for machines so that they can be located. The address is a 32 bit integer which gives the IP address. This encodes a network ID and more addressing. The network ID falls into various classes according to the size of the network address.

Network address

Class A uses 8 bits for the network address with 24 bits left over for other addressing. Class B uses 16 bit network addressing. Class C uses 24 bit network addressing and class D uses all 32.

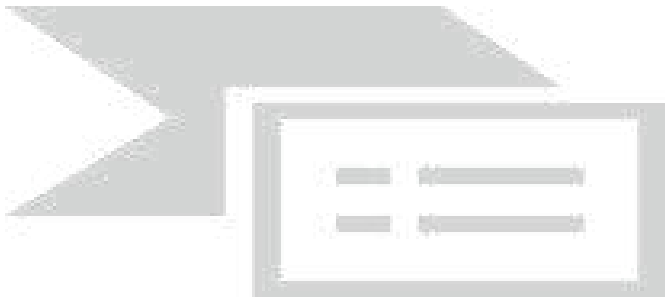
Subnet address

Internally, the UNIX network is divided into sub networks. Building 11 is currently on one sub network and uses 10-bit addressing, allowing 1024 different hosts.

Host address

8 bits are finally used for host addresses within our subnet. This places a limit of 256 machines that can be on the subnet.

Total address



The 32 bit address is usually written as 4 integers separated by dots.

Port addresses

A service exists on a host, and is identified by its port. This is a 16 bit number. To send a message to a server, you send it to the port for that service of the host that it is running on. This is not location transparency! Certain of these ports are "well known".

Sockets

A socket is a data structure maintained by the system to handle network connections. A socket is created using the call socket. It returns an integer that is like a file descriptor. In fact, under Windows, this handle can be used with Read File and Write File functions.

```
#include <sys/types.h>
#include <sys/socket.h>
int socket(int family, int type, int protocol);
```

Here "family" will be AF_INET for IP communications, protocol will be zero, and type will depend on whether TCP or UDP is used. Two processes wishing to communicate over a network create a socket each. These are similar to two ends of a pipe - but the actual pipe does not yet exist.

JFree Chart

JFreeChart is a free 100% Java chart library that makes it easy for developers to display professional quality charts in their applications. JFreeChart's extensive feature set includes:

A consistent and well-documented API, supporting a wide range of chart types;

A flexible design that is easy to extend, and targets both server-side and client-side applications;

Support for many output types, including Swing components, image files (including PNG and JPEG), and vector graphics file formats (including PDF, EPS and SVG);

JFreeChart is "open source" or, more specifically, free software. It is distributed under the terms of the GNU Lesser General Public Licence (LGPL), which permits use in proprietary applications.

1. Map Visualizations

Charts showing values that relate to geographical areas. Some examples include: (a) population density in each state of the United States, (b) income per capita for each

country in Europe, (c) life expectancy in each country of the world. The tasks in this project include:

Sourcing freely redistributable vector outlines for the countries of the world, states/provinces in particular countries (USA in particular, but also other areas);

Creating an appropriate dataset interface (plus default implementation), a rendered, and integrating this with the existing XYPlot class in JFreeChart;

Testing, documenting, testing some more, documenting some more.

2. Time Series Chart Interactivity

Implement a new (to JFreeChart) feature for interactive time series charts --- to display a separate control that shows a small version of ALL the time series data, with a sliding "view" rectangle that allows you to select the subset of the time series data to display in the main chart.

3. Dashboards

There is currently a lot of interest in dashboard displays. Create a flexible dashboard mechanism that supports a subset of JFreeChart chart types (dials, pies, thermometers, bars, and lines/time series) that can be delivered easily via both Java Web Start and an applet.

4. Property Editors

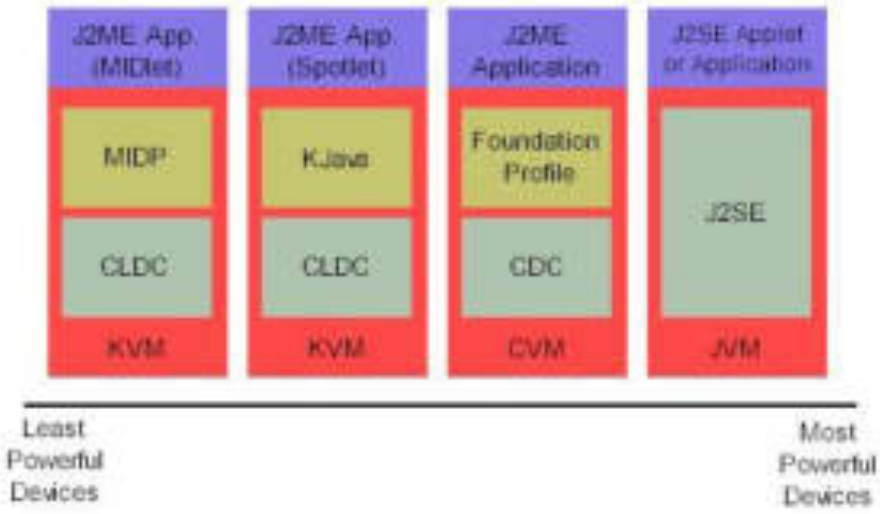
The property editor mechanism in JFreeChart only handles a small subset of the properties that can be set for charts. Extend (or reimplement) this mechanism to provide greater end-user control over the appearance of the charts.

J2ME (Java 2 Micro edition):-

Sun Microsystems defines J2ME as "a highly optimized Java run-time environment targeting a wide range of consumer products, including pagers, cellular phones, screen-phones, digital set-top boxes and car navigation systems." Announced in June 1999 at the JavaOne Developer

Conference, J2ME brings the cross-platform functionality of the Java language to smaller devices, allowing mobile wireless devices to share applications. With J2ME, Sun has adapted the Java platform for consumer products that incorporate or are based on small computing devices.

1. General J2ME architecture



J2ME uses configurations and profiles to customize the Java Runtime Environment (JRE). As a complete JRE, J2ME is comprised of a configuration, which determines the JVM used, and a profile, which defines the application by adding domain-specific classes. The configuration defines the basic run-time environment as a set of core classes and a specific JVM that run on specific types of devices. We'll discuss configurations in detail in the Theprofile defines the application; specifically, it adds domain-specific classes to the J2ME configuration to define certain uses for devices. We'll cover profiles in depth in the The following graphic depicts the relationship between the different virtual machines, configurations, and profiles. It also draws a parallel with the J2SE API and its Java virtual machine. While the J2SE virtual machine is generally referred to as a JVM, the J2ME virtual machines, KVM and CVM, are subsets of JVM. Both KVM and CVM can be thought of as a kind of Java virtual machine -- it's just that they are shrunken versions of the J2SE JVM and are specific to J2ME.

2. Developing J2ME applications

Introduction In this section, we will go over some considerations you need to keep in mind when developing applications for smaller devices. We'll take a look at the way the compiler is invoked when using J2SE to compile J2ME applications. Finally, we'll explore packaging and deployment and the role preverification plays in this process.

3. Design considerations for small devices

Developing applications for small devices requires you to keep certain strategies in mind during the design phase. It is best to strategically design an application for a small device before you begin coding. Correcting the code because you failed to consider all of the "gotchas" before developing the application can be a painful process. Here are some design strategies to consider:

* Keep it simple. Remove unnecessary features, possibly making those features a separate, secondary application.
* Smaller is better. This consideration should be a "no brainer" for all developers. Smaller applications use less memory on the device and require shorter installation times. Consider packaging your Java applications as compressed Java Archive (jar) files.
* Minimize run-time memory use. To minimize the amount of memory used at run time, use scalar types in place of object types. Also, do not depend on the garbage collector. You should manage the memory efficiently yourself by setting object references to null when you are finished with them. Another way to reduce run-time memory is to use lazy instantiation, only allocating objects on an as-needed basis. Other ways of reducing overall and peak memory use on small devices are to release resources quickly, reuse objects, and avoid exceptions.

4. Configurations overview

The configuration defines the basic run-time environment as a set of core classes and a specific JVM that run on specific types of devices. Currently, two configurations exist for J2ME, though others may be defined in the future:

* **Connected Limited Device Configuration (CLDC)**is used specifically with the KVM for 16-bit or 32-bit devices with limited amounts of memory. This is the configuration (and the virtual

machine) used for developing small J2ME applications. Its size limitations make CLDC more interesting and challenging (from a development point of view) than CDC. CLDC is also the configuration that we will use for developing our drawing tool application. An example of a small wireless device running small applications is a Palm hand-held computer.

* **Connected Device Configuration (CDC)**is used with the C virtual machine (CVM) and is used for 32-bit architectures requiring more than 2 MB of memory. An example of such a device is a Net TV box.

5. J2ME profiles

What is a J2ME profile?

As we mentioned earlier in this tutorial, a profile defines the type of device supported. The Mobile Information Device Profile (MIDP), for example, defines classes for cellular phones. It adds domain-specific classes to the J2ME configuration to define uses for similar devices. Two profiles have been defined for J2ME and are built upon CLDC: KJava and MIDP. Both KJava and MIDP are associated with CLDC and smaller devices. Profiles are built on top of configurations. Because profiles are specific to the size of the device (amount of memory) on which an application runs, certain profiles are associated with certain configurations.

A skeleton profile upon which you can create your own profile, the Foundation Profile, is available for CDC.

Profile 1: KJava

KJava is Sun's proprietary profile and contains the KJava API. The KJava profile is built on top of the CLDC configuration. The KJava virtual machine, KVM, accepts the same byte codes and class file format as the classic J2SE virtual machine. KJava contains a Sun-specific API that runs on the Palm OS. The KJava API has a great deal in common with the J2SE Abstract Windowing Toolkit (AWT). However, because it is not a standard J2ME package, its main package is com.sun.kjava. We'll learn more about the KJava API later in this tutorial when we develop some sample applications.

Profile 2: MIDP

MIDP is geared toward mobile devices such as cellular phones and pagers. The MIDP, like KJava, is built upon CLDC and provides a standard run-time environment that allows new

applications and services to be deployed dynamically on end user devices. MIDP is a common, industry-standard profile for mobile devices that is not dependent on a specific vendor. It is a complete and supported foundation for mobile application

development. MIDP contains the following packages, the first three of which are core CLDC packages, plus three MIDP-specific packages.

* java.lang
* java.io
* java.util
* javax.microedition.io
* javax.microedition.lcdui
* javax.microedition.midlet
* javax.microedition.rms

5.2. SAMPLE CODE

Home.java

```
packagerating.prediction.based;

public class home extends javax.swing.JFrame {

public home() {

initComponents();

    }


   /**

     * This method is called from within the constructor to initialize the form.

     * WARNING: Do NOT modify this code. The content of this method is always

     * regenerated by the Form Editor.

     */

@SuppressWarnings("unchecked")

   // <editor-fold defaultstate="collapsed" desc="Generated Code">//GEN-
BEGIN:initComponents

private void initComponents() {


        jLabel4 = new javax.swing.JLabel();

        jLabel5 = new javax.swing.JLabel();

        jButton1 = new javax.swing.JButton();

        jLabel1 = new javax.swing.JLabel();
```

```
setDefaultCloseOperation(javax.swing.WindowConstants.EXIT_ON_CLOSE);

getContentPane().setLayout(new org.netbeans.lib.awtextra.AbsoluteLayout());


jLabel4.setFont(new java.awt.Font("Bodoni MT Black", 1, 18)); // NOI18N

jLabel4.setText("SENTIMENT  FROM TEXTUAL REVIEWS");

getContentPane().add(jLabel4, new org.netbeans.lib.awtextra.AbsoluteConstraints(70, 210, 490,
40));


jLabel5.setFont(new java.awt.Font("Bodoni MT Black", 1, 18)); // NOI18N

jLabel5.setText("RATING PREDICTION BASED ON SOCIAL ");

getContentPane().add(jLabel5, new org.netbeans.lib.awtextra.AbsoluteConstraints(70, 130, 470,
40));


jButton1.setBackground(new java.awt.Color(153, 153, 0));

jButton1.setFont(new java.awt.Font("Times New Roman", 1, 14)); // NOI18N

jButton1.setText("NEXT");

jButton1.addActionListener(new java.awt.event.ActionListener() {

public void actionPerformed(java.awt.event.ActionEventevt) {

jButton1ActionPerformed(evt);

            }

        });

getContentPane().add(jButton1, new org.netbeans.lib.awtextra.AbsoluteConstraints(460, 313,
80, 30));
```

```
jLabel1.setIcon(new
javax.swing.ImageIcon(getClass().getResource("/rating/prediction/based/bgg.jpg"))); // NOI18N

jLabel1.setText("jLabel1");

getContentPane().add(jLabel1, new org.netbeans.lib.awtextra.AbsoluteConstraints(0, 0, 660,
470));


java.awt.DimensionscreenSize = java.awt.Toolkit.getDefaultToolkit().getScreenSize();

setBounds((screenSize.width-638)/2, (screenSize.height-500)/2, 638, 500);

    }// </editor-fold>//GEN-END:initComponents


private void jButton1ActionPerformed(java.awt.event.ActionEventevt) {//GEN-
FIRST:event_jButton1ActionPerformed

        // TODO add your handling code here:

newallproductitems().setVisible(true);

    }//GEN-LAST:event_jButton1ActionPerformed


    /**

     * @paramargs the command line arguments

     */

public static void main(String args[]) {

        /* Set the Nimbus look and feel */

        //<editor-fold defaultstate="collapsed" desc=" Look and feel setting code (optional) ">
```

```
        /* If Nimbus (introduced in Java SE 6) is not available, stay with the default look and feel.
         * For details see http://download.oracle.com/javase/tutorial/uiswing/lookandfeel/plaf.html
         */
try {
for (javax.swing.UIManager.LookAndFeelInfo info :
javax.swing.UIManager.getInstalledLookAndFeels()) {
if ("Nimbus".equals(info.getName())) {
javax.swing.UIManager.setLookAndFeel(info.getClassName());
break;
                }
            }
        } catch (ClassNotFoundException ex) {
java.util.logging.Logger.getLogger(home.class.getName()).log(java.util.logging.Level.SEVERE,
null, ex);
        } catch (InstantiationException ex) {
java.util.logging.Logger.getLogger(home.class.getName()).log(java.util.logging.Level.SEVERE,
null, ex);
        } catch (IllegalAccessException ex) {
java.util.logging.Logger.getLogger(home.class.getName()).log(java.util.logging.Level.SEVERE,
null, ex);
        } catch (javax.swing.UnsupportedLookAndFeelException ex) {
java.util.logging.Logger.getLogger(home.class.getName()).log(java.util.logging.Level.SEVERE,
null, ex);
        }
```

```
        //</editor-fold>
        /* Create and display the form */
java.awt.EventQueue.invokeLater(new Runnable() {
public void run() {
new home().setVisible(true);
            }
        });
    }
    // Variables declaration - do not modify//GEN-BEGIN:variables
privatejavax.swing.JButton jButton1;
privatejavax.swing.JLabel jLabel1;
privatejavax.swing.JLabel jLabel4;
privatejavax.swing.JLabel jLabel5;
    // End of variables declaration//GEN-END:variables
}
```

Allproductitems.java

```
/*
 * To change this template, choose Tools | Templates
 * and open the template in the editor.
 */
packagerating.prediction.based;


importjava.sql.Connection;
```

```
importjava.sql.ResultSet;
importjava.sql.ResultSetMetaData;
importjava.sql.SQLException;
importjava.sql.Statement;
importjava.util.Vector;
importjava.util.logging.Level;
importjava.util.logging.Logger;
importjavax.swing.JOptionPane;
importjavax.swing.table.DefaultTableModel;

/**
 *
 * @author java1
 */
public class allproductitems extends javax.swing.JFrame {

    Connection con;
static Statement st;
staticResultSet rs,rs1,rs2,rs3,rs4,rs5;
static String path;
staticint row;
static String s,name;
String filename;
```

```
privateinti;



    /**
     * Creates new form allproductitems
     */
publicallproductitems() {
try {
initComponents();
con =  dobc.getConne();
st = (Statement) con.createStatement();
        } catch (SQLException ex) {
Logger.getLogger(allproductitems.class.getName()).log(Level.SEVERE, null, ex);
        }



    }


    /**
     * This method is called from within the constructor to initialize the form.
     * WARNING: Do NOT modify this code. The content of this method is always
     * regenerated by the Form Editor.
     */
```

```
@SuppressWarnings("unchecked")
    // <editor-fold defaultstate="collapsed" desc="Generated Code">//GEN-
BEGIN:initComponents
private void initComponents() {

        jScrollPane2 = new javax.swing.JScrollPane();
        jTable1 = new javax.swing.JTable();
        jButton1 = new javax.swing.JButton();
        jTextField1 = new javax.swing.JTextField();
        jLabel1 = new javax.swing.JLabel();
        jButton2 = new javax.swing.JButton();
        jLabel2 = new javax.swing.JLabel();
        jLabel3 = new javax.swing.JLabel();

setDefaultCloseOperation(javax.swing.WindowConstants.EXIT_ON_CLOSE);
getContentPane().setLayout(new org.netbeans.lib.awtextra.AbsoluteLayout());

jTable1.setFont(new java.awt.Font("Baskerville Old Face", 1, 14)); // NOI18N
jTable1.setModel(new javax.swing.table.DefaultTableModel(
new Object [][] {

            },
new String [] {
```

```
            }
        ));
jScrollPane2.setViewportView(jTable1);

getContentPane().add(jScrollPane2, new org.netbeans.lib.awtextra.AbsoluteConstraints(50, 70, 340, 230));

jButton1.setBackground(new java.awt.Color(153, 153, 0));
jButton1.setFont(new java.awt.Font("Times New Roman", 1, 11)); // NOI18N
jButton1.setText("VIEW_ALL_PRODUCT");
jButton1.addActionListener(new java.awt.event.ActionListener() {
public void actionPerformed(java.awt.event.ActionEventevt) {
jButton1ActionPerformed(evt);
            }
        });
getContentPane().add(jButton1, new org.netbeans.lib.awtextra.AbsoluteConstraints(410, 150, 160, 30));

jTextField1.setText("jTextField1");
getContentPane().add(jTextField1, new org.netbeans.lib.awtextra.AbsoluteConstraints(290, 330, 160, 30));
```

```
jLabel1.setFont(new java.awt.Font("Times New Roman", 1, 14)); // NOI18N

jLabel1.setText("OVER_ALL_ITEMS_COUNT");

getContentPane().add(jLabel1, new org.netbeans.lib.awtextra.AbsoluteConstraints(70, 330, 180, 30));


jButton2.setBackground(new java.awt.Color(153, 153, 0));

jButton2.setFont(new java.awt.Font("Times New Roman", 1, 11)); // NOI18N

jButton2.setText("NEXT");

jButton2.addActionListener(new java.awt.event.ActionListener() {

public void actionPerformed(java.awt.event.ActionEventevt) {

jButton2ActionPerformed(evt);

            }

        });

getContentPane().add(jButton2, new org.netbeans.lib.awtextra.AbsoluteConstraints(440, 210, 90, 30));


jLabel2.setFont(new java.awt.Font("Bodoni MT Black", 1, 18)); // NOI18N

jLabel2.setText("ALL PRODUCT ITEMS");

getContentPane().add(jLabel2, new org.netbeans.lib.awtextra.AbsoluteConstraints(90, 30, 260, -1));


jLabel3.setIcon(new javax.swing.ImageIcon(getClass().getResource("/rating/prediction/based/bgg.jpg"))); // NOI18N

jLabel3.setText("jLabel3");
```

```
getContentPane().add(jLabel3, new org.netbeans.lib.awtextra.AbsoluteConstraints(0, 0, 610,
390));


java.awt.DimensionscreenSize = java.awt.Toolkit.getDefaultToolkit().getScreenSize();

setBounds((screenSize.width-628)/2, (screenSize.height-433)/2, 628, 433);

    }// </editor-fold>//GEN-END:initComponents


private void jButton1ActionPerformed(java.awt.event.ActionEventevt) {//GEN-
FIRST:event_jButton1ActionPerformed

try {

          // TODO add your handling code here:

           Vector col = new Vector();

           Vector data = new Vector();

rs = (ResultSet) st.executeQuery("SELECT *  FROM  c_items_product  ");

ResultSetMetaData md = (ResultSetMetaData) rs.getMetaData();

int columns = md.getColumnCount();

int count=0;

for (inti = 1; i<= columns; i++) {

col.addElement(md.getColumnName(i));


                }


while (rs.next()) {
```

```
count++;
                    Vector row = new Vector(columns);
for (inti = 1; i<= columns; i++) {
row.addElement(rs.getObject(i));

                    }

data.addElement(row);
                }
jTextField1.setText(Integer.toString(count));
con.close();

DefaultTableModel model = new DefaultTableModel(data, col);
jTable1.setModel(model);
JOptionPane.showMessageDialog(null, "over all product items");
        } catch (SQLException ex) {
Logger.getLogger(allproductitems.class.getName()).log(Level.SEVERE, null, ex);
        }
    }//GEN-LAST:event_jButton1ActionPerformed

private void jButton2ActionPerformed(java.awt.event.ActionEventevt) {//GEN-
FIRST:event_jButton2ActionPerformed
        // TODO add your handling code here:
```

```
newalluserratings().setVisible(true);
    }//GEN-LAST:event_jButton2ActionPerformed


    /**
     * @paramargs the command line arguments
     */
public static void main(String args[]) {
        /* Set the Nimbus look and feel */
        //<editor-fold defaultstate="collapsed" desc=" Look and feel setting code (optional) ">
        /* If Nimbus (introduced in Java SE 6) is not available, stay with the default look and feel.
         * For details see http://download.oracle.com/javase/tutorial/uiswing/lookandfeel/plaf.html
         */
try {
for (javax.swing.UIManager.LookAndFeelInfo info :
javax.swing.UIManager.getInstalledLookAndFeels()) {
if ("Nimbus".equals(info.getName())) {
javax.swing.UIManager.setLookAndFeel(info.getClassName());
break;
                }
            }
        } catch (ClassNotFoundException ex) {
java.util.logging.Logger.getLogger(allproductitems.class.getName()).log(java.util.logging.Level.
SEVERE, null, ex);
```

```
        } catch (InstantiationException ex) {
java.util.logging.Logger.getLogger(allproductitems.class.getName()).log(java.util.logging.Level.
SEVERE, null, ex);
        } catch (IllegalAccessException ex) {
java.util.logging.Logger.getLogger(allproductitems.class.getName()).log(java.util.logging.Level.
SEVERE, null, ex);
        } catch (javax.swing.UnsupportedLookAndFeelException ex) {
java.util.logging.Logger.getLogger(allproductitems.class.getName()).log(java.util.logging.Level.
SEVERE, null, ex);
        }
        //</editor-fold>

        /* Create and display the form */
java.awt.EventQueue.invokeLater(new Runnable() {
public void run() {
newallproductitems().setVisible(true);
            }
        });
    }
    // Variables declaration - do not modify//GEN-BEGIN:variables
privatejavax.swing.JButton jButton1;
privatejavax.swing.JButton jButton2;
privatejavax.swing.JLabel jLabel1;
```

```
privatejavax.swing.JLabel jLabel2;
privatejavax.swing.JLabel jLabel3;
privatejavax.swing.JScrollPane jScrollPane2;
privatejavax.swing.JTable jTable1;
privatejavax.swing.JTextField jTextField1;
    // End of variables declaration//GEN-END:variables
}
```

6.SYSTEM TESTING

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

Introduction:

In general, software engineers distinguish software faults from software failures. In case of a failure, the software does not do what the user expects. A fault is a programming error that may or may not actually manifest as a failure. A fault can also be described as an error in the correctness of the semantic of a computer program. A fault will become a failure if the exact computation conditions are met, one of them being that the faulty portion of computer software executes on the CPU.

A fault can also turn into a failure when the software is ported to a different hardware

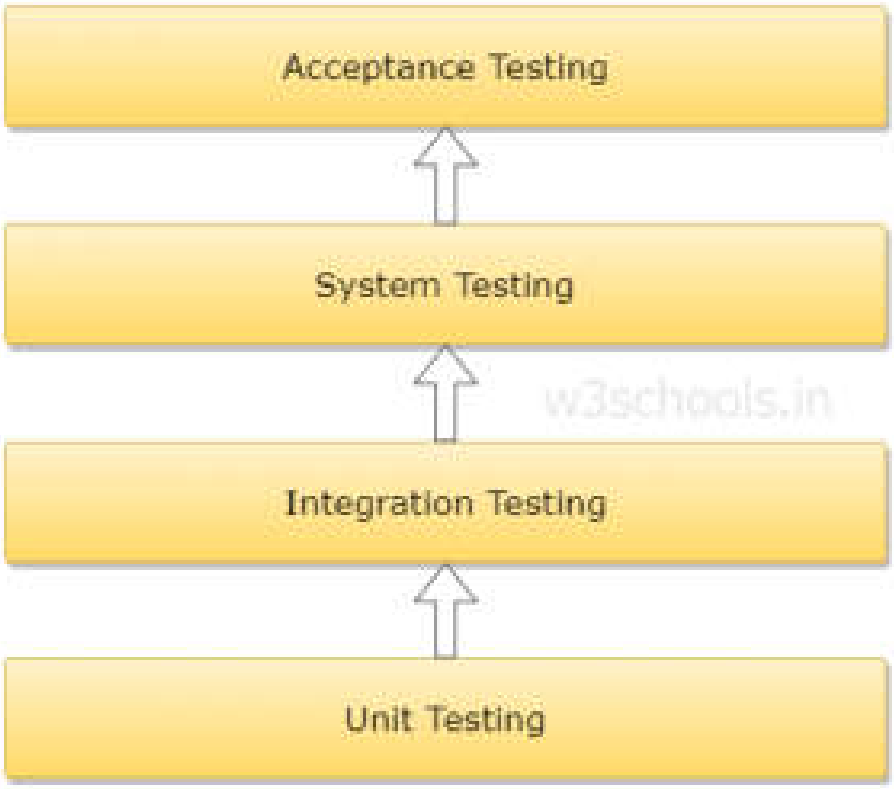
Software testing may be viewed as a sub-field of Software Quality Assurance but typically exists independently (and there may be no SQA areas in some companies). In SQA, software process specialists and auditors take a broader view on software and its development. They examine and change the software engineering process itself to reduce the amount of faults that end up in the code or deliverfaster.

Regardless of the methods used or level of formality involved the desired result of testing is a level of confidence in the software so that the organization is confident that the software has an acceptable defect rate. What constitutes an acceptable defect rate depends on the nature of the software. An arcade video game designed to simulate flying an airplane would presumably have a much higher tolerance for defects than software used to control an actual airliner.

A problem with software testing is that the number of defects in a software product can be very large, and the number of configurations of the product larger still. Bugs that occur infrequently are difficult to find in testing. A rule of thumb is that a system that is expected to function without faults for a certain length of time must have already been tested for at least that length of time. This has severe consequences for projects to write long-lived reliablesoftware.

6.1 Test Cases:

Different levels of testing:



Levels of Testing

TYPES OF TESTS

Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs

accurately to the documented specifications and contains clearly defined inputs and expected results.

Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems/Procedures : interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

Unit Testing

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

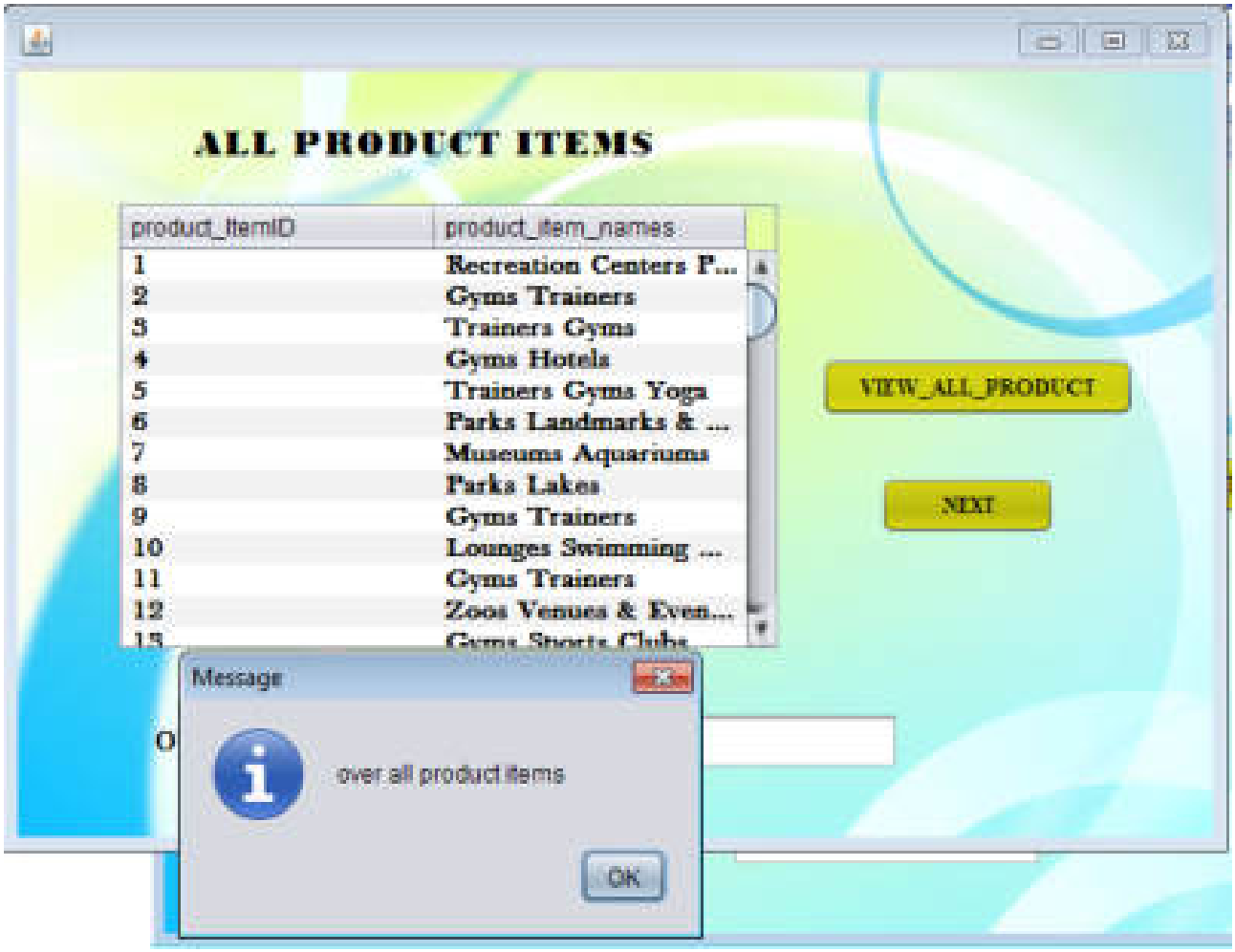
Acceptance Testing

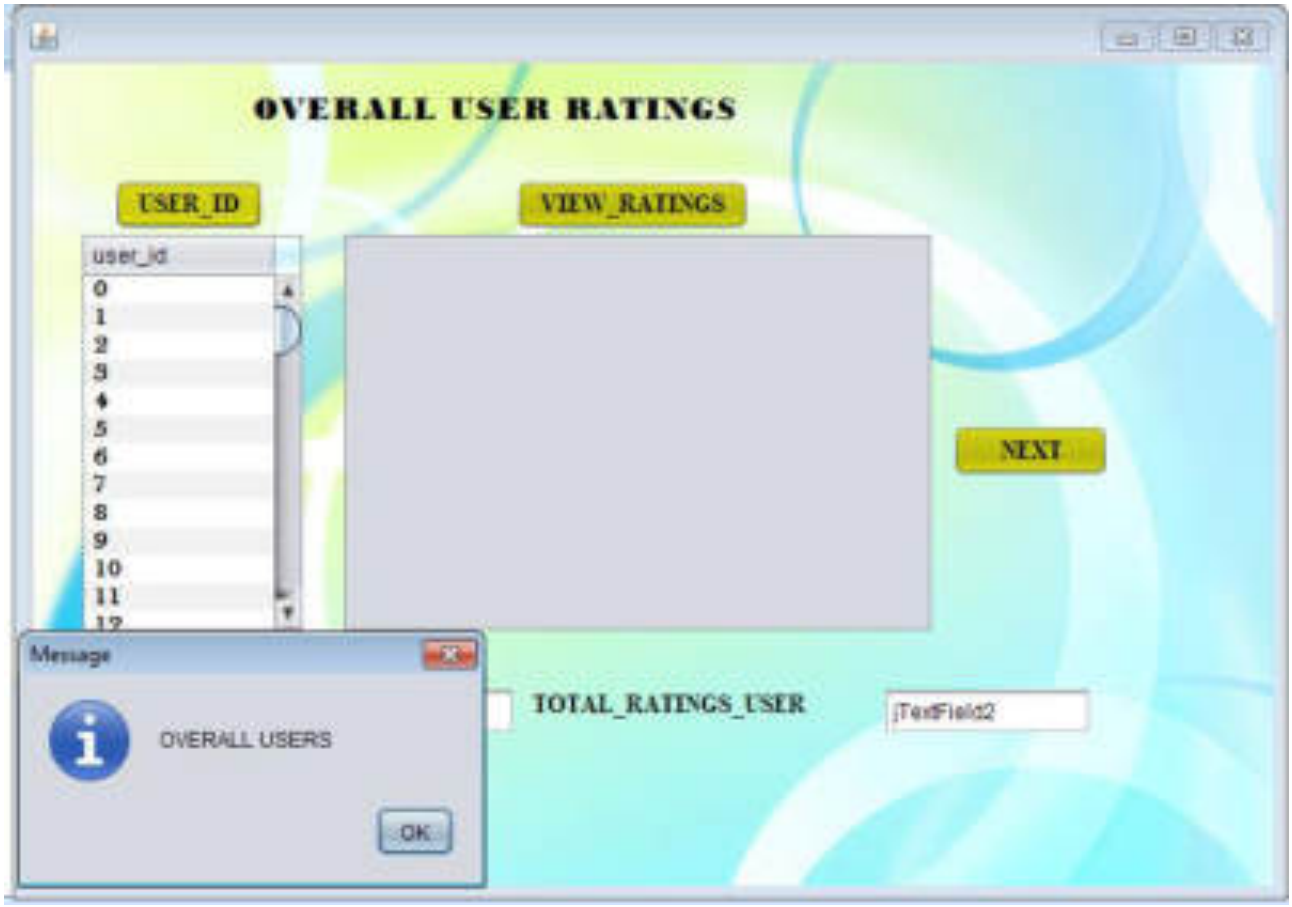
User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

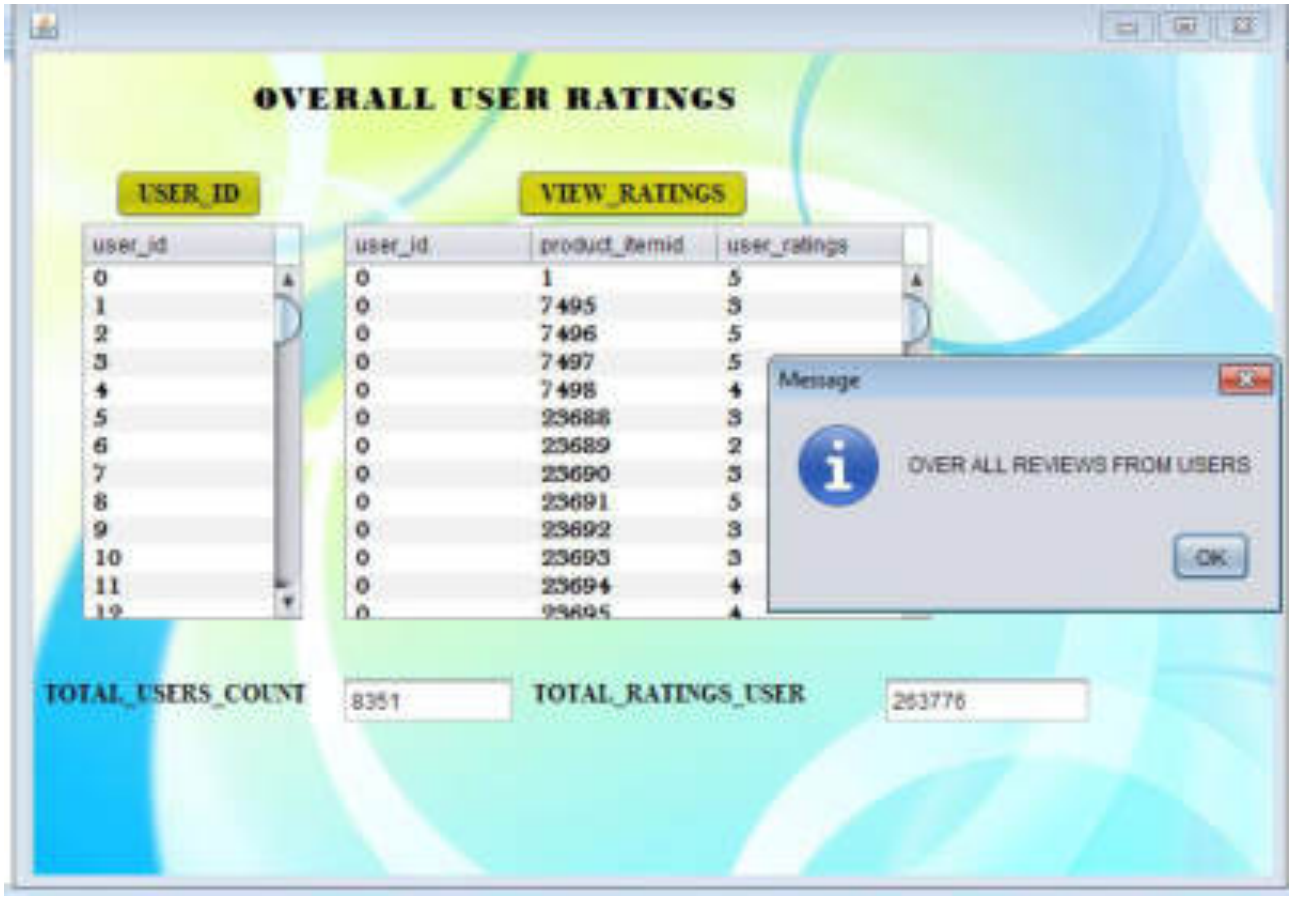
Test Results: All the test cases mentioned above passed successfully. No defects encountered.

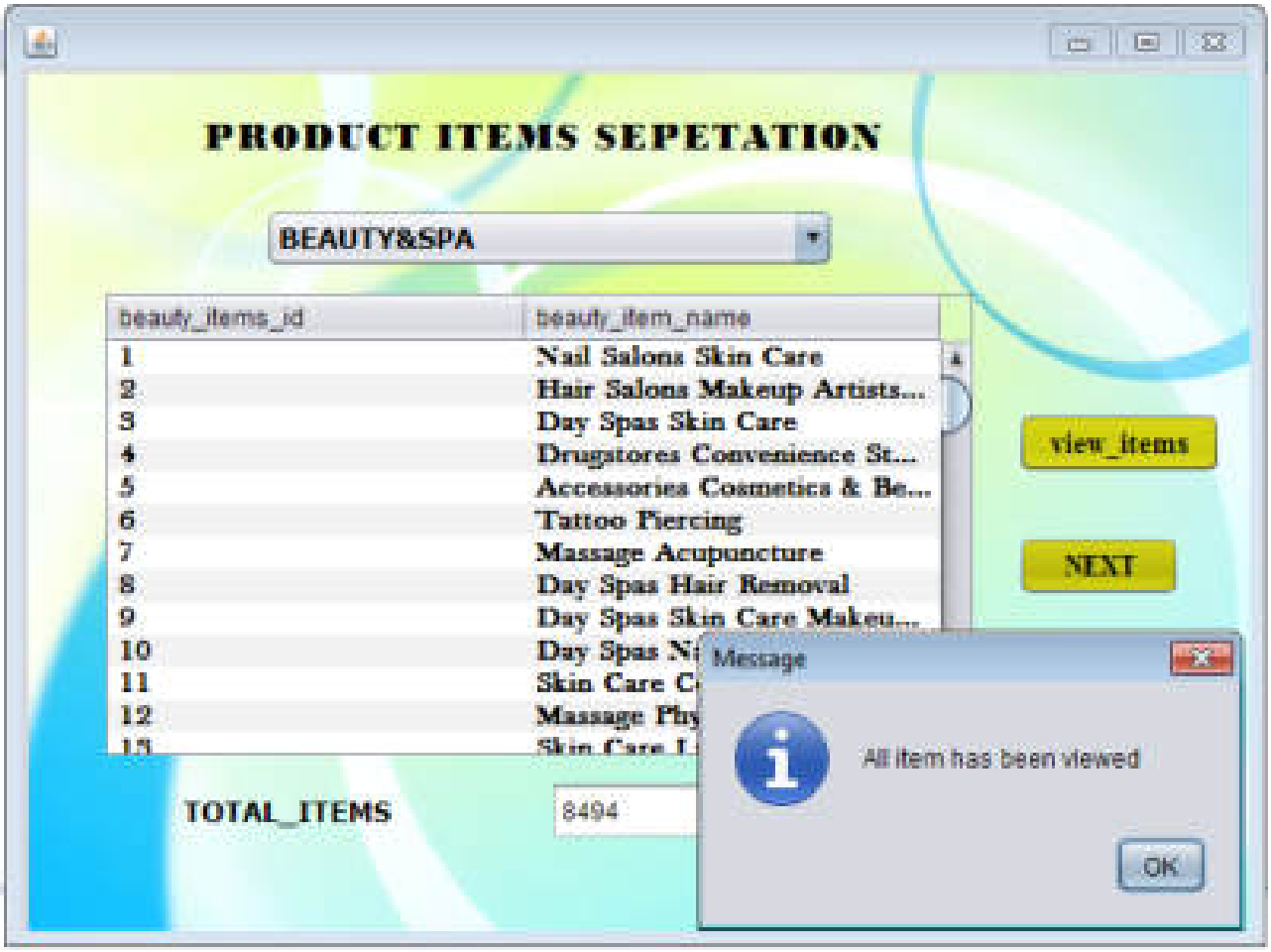
7. RESULTS

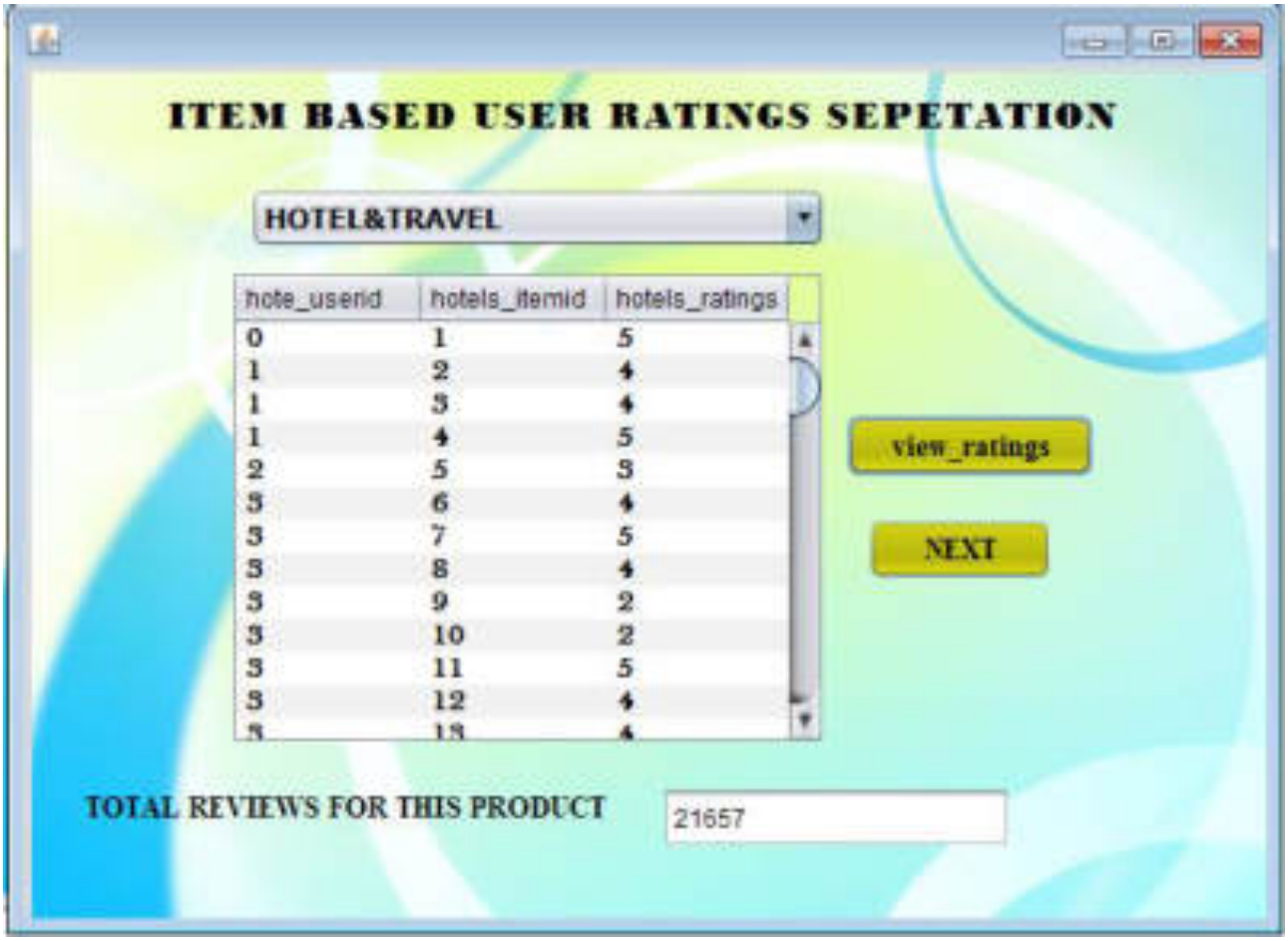
SCREEN SHOTS:

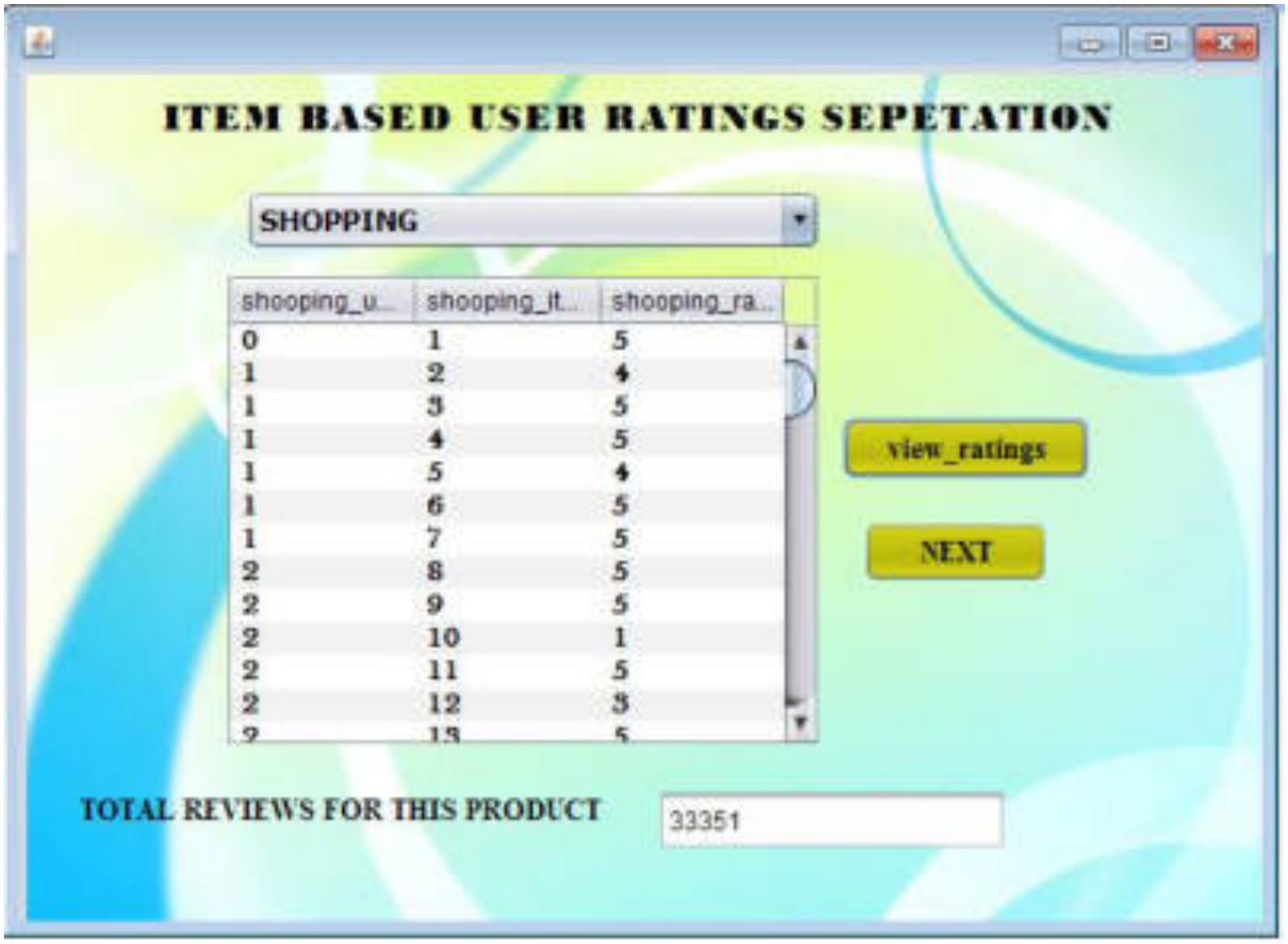

ALL PRODUCT ITEMS
product_ItemID
product_item_names
1
Recreation Centers P...
2
Gyms Trainers
3
Trainers Gyms
4
Gyms Hotels
5
Trainers Gyms Yoga
6
Parks Landmarks & ...
7
Museums Aquariums
8
Parks Lakes
9
Gyms Trainers
10
Lounges Swimming ...
11
Gyms Trainers
12
Zoos Venues & Even...
VIEW_ALL_PRODUCT
NEXT
Message
over all product items
OK


OVERALL USER RATINGS
USER_ID
VIEW_RATINGS
user_id
0
1
2
3
4
5
6
7
8
9
10
11
NEXT
Message
OVERALL USERS
OK
TOTAL_RATINGS_USER
jTextField2


OVERALL USER RATINGS
USER_ID
VIEW_RATINGS
user_id
0
1
2
3
4
5
6
7
8
9
10
11
user_id
product_itemid
user_ratings
0 1 5
0 7495 3
0 7496 5
0 7497 5
0 7498 4
0 23688 3
0 23689 2
0 23690 3
0 23691 5
0 23692 3
0 23693 3
0 23694 4
Message
OVER ALL REVIEWS FROM USERS
OK
TOTAL_USERS_COUNT
8351
TOTAL_RATINGS_USER
263776

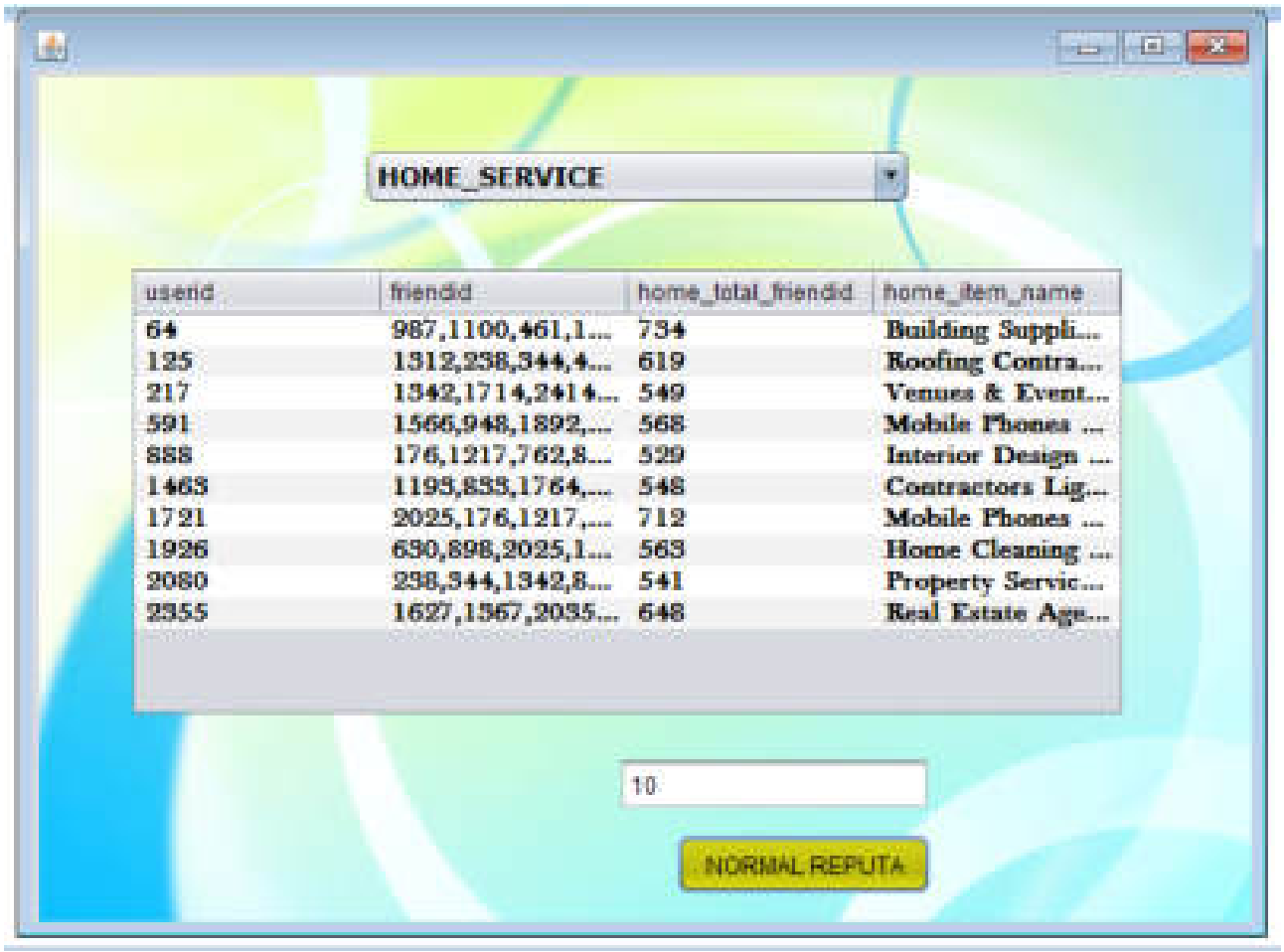

PRODUCT ITEMS SEPETATION
BEAUTY&SPA
beauty_items_id
beauty_item_name
1
Nail Salons Skin Care
2
Hair Salons Makeup Artists...
3
Day Spas Skin Care
4
Drugstores Convenience St...
5
Accessories Cosmetics & Be...
6
Tattoo Piercing
7
Massage Acupuncture
8
Day Spas Hair Removal
9
Day Spas Skin Care Makeu...
10
11
12
view_items
NEXT
Message
All item has been viewed
OK
TOTAL_ITEMS
8494

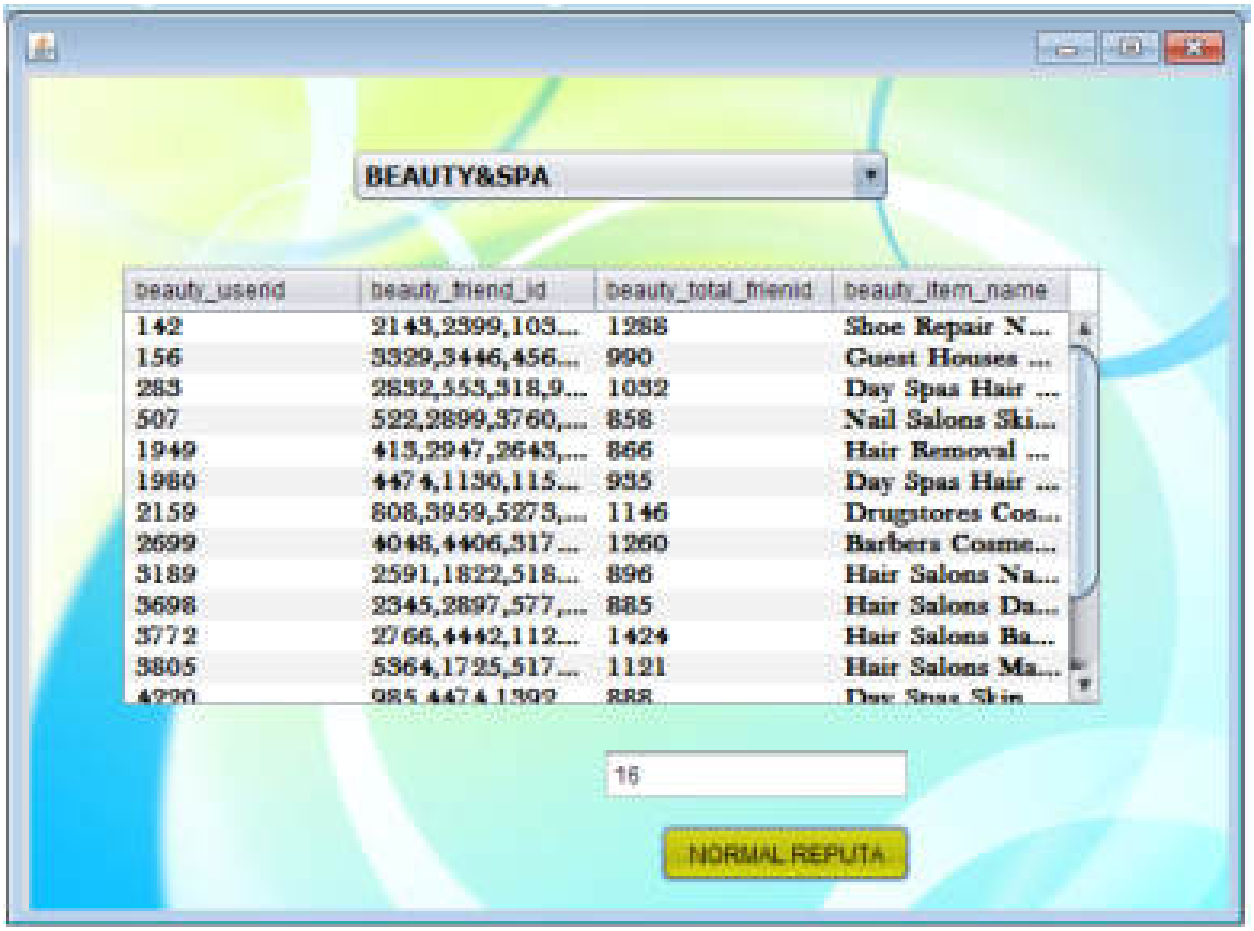

ITEM BASED USER RATINGS SEPETATION
HOTEL&TRAVEL
hote_userid
hotels_itemid
hotels_ratings
0 1 5
1 2 4
1 3 4
1 4 5
2 5 3
3 6 4
3 7 5
3 8 4
3 9 2
3 10 2
3 11 5
3 12 4
view_ratings
NEXT
TOTAL REVIEWS FOR THIS PRODUCT
21657

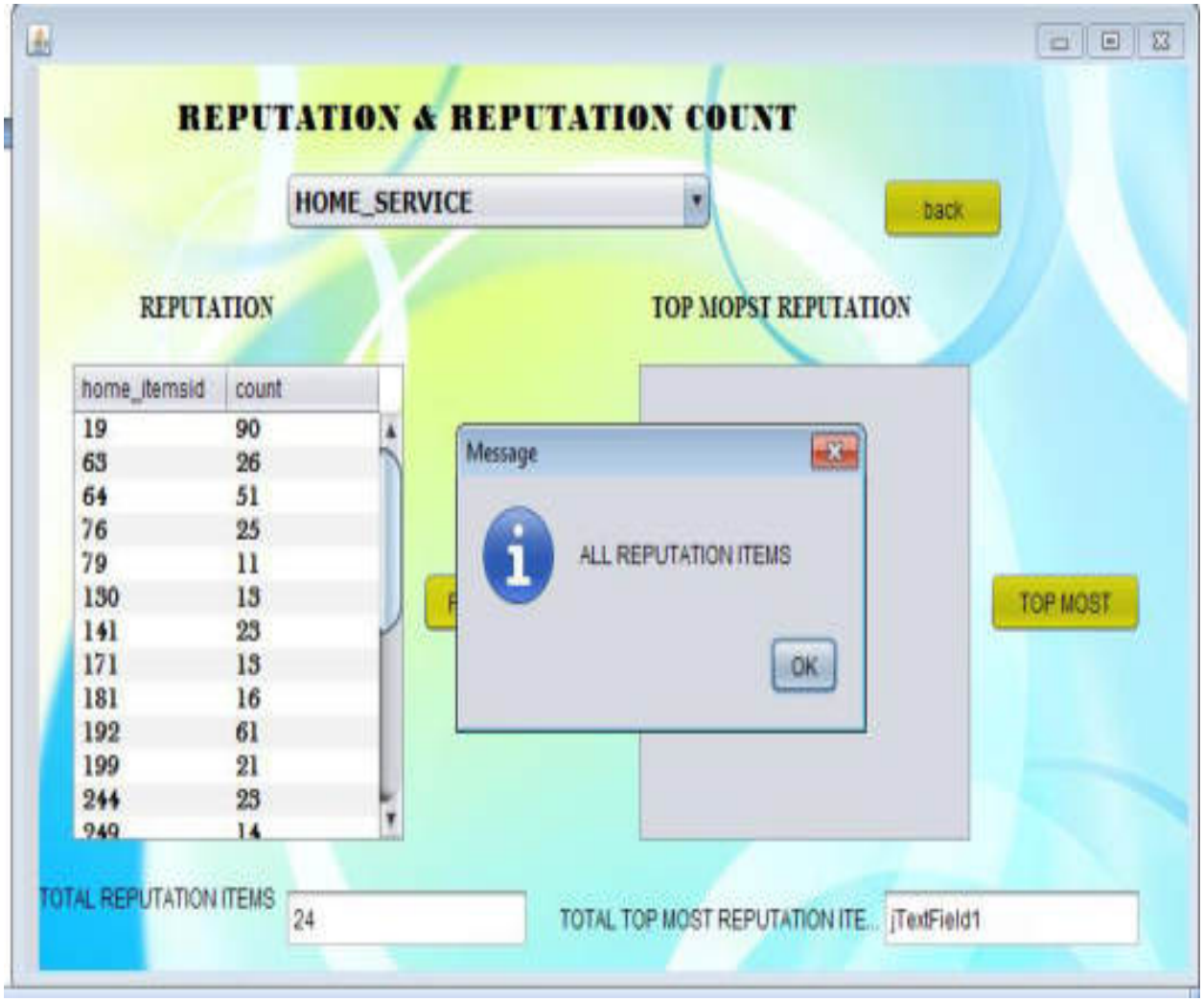

ITEM BASED USER RATINGS SEPETATION
SHOPPING
shooping_u...
shooping_it...
shooping_ra...
0 1 5
1 2 4
1 3 5
1 4 5
1 5 4
1 6 5
1 7 5
2 8 5
2 9 5
2 10 1
2 11 5
2 12 3
view_ratings
NEXT
TOTAL REVIEWS FOR THIS PRODUCT
33351

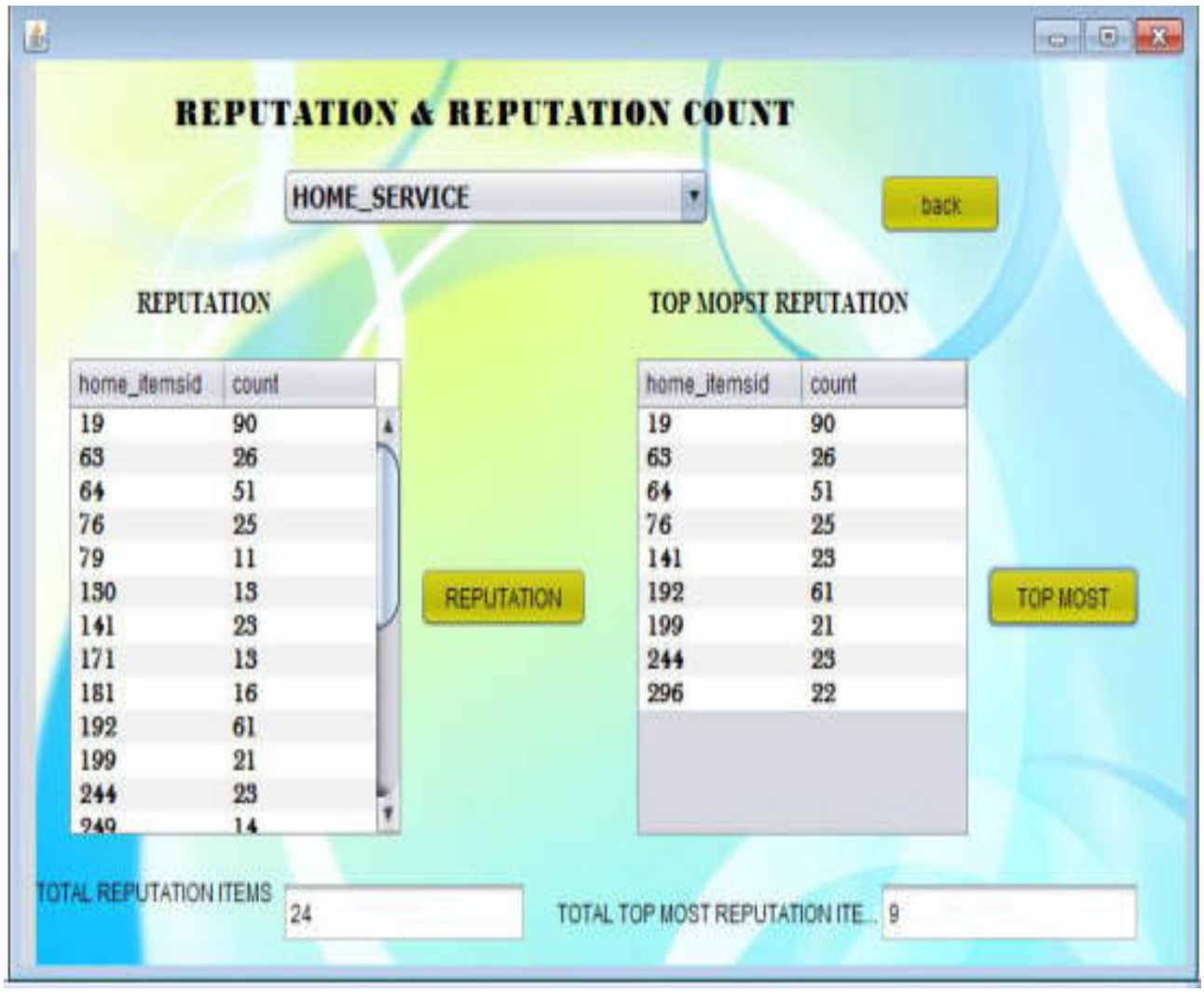

THREE FACTORS
USER'S TRUSTED FRIEND
ITEM'S REPUTATION
USER SENTIMENT SIMILARITY

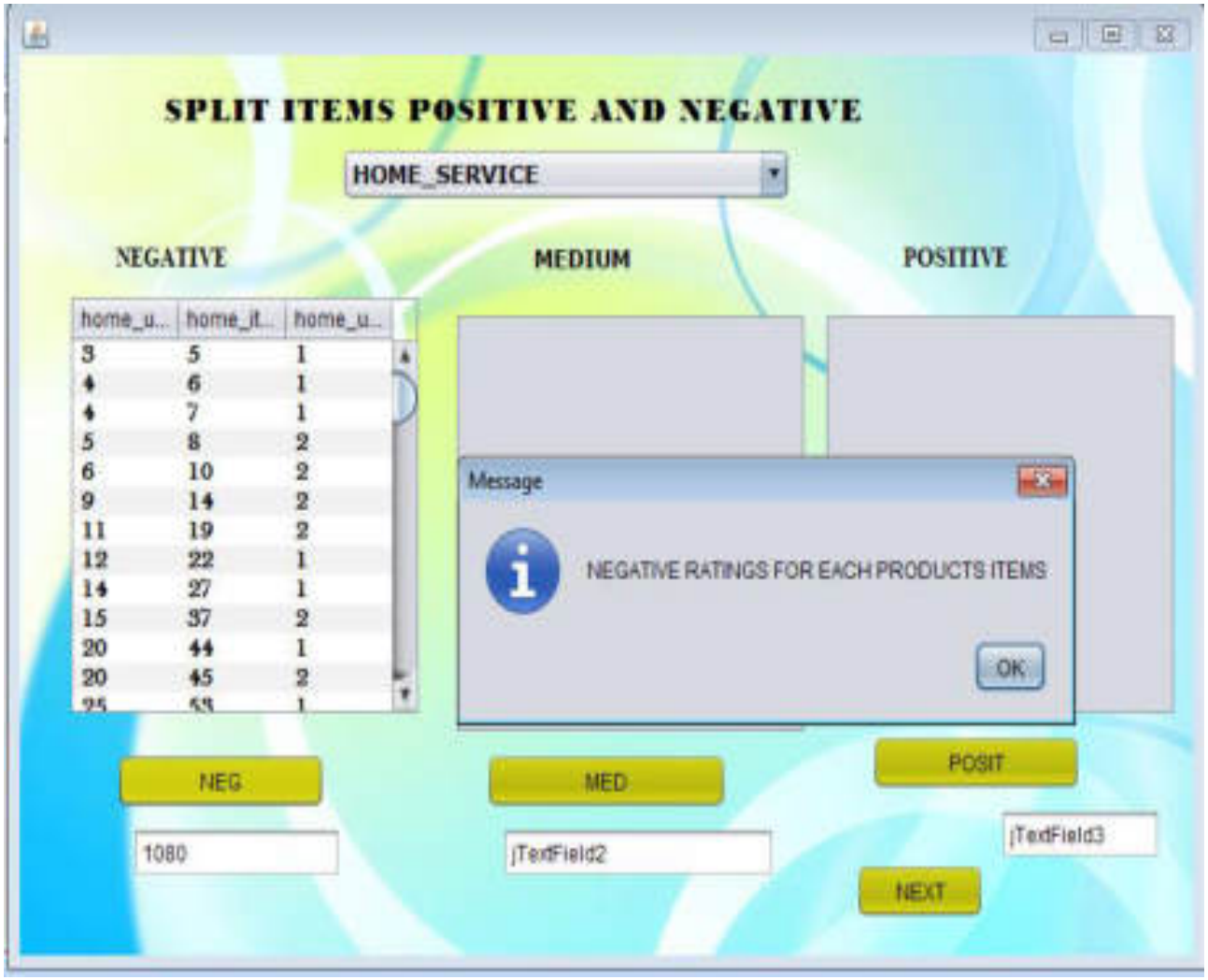

INTERPERSONAL SENTIMENTAL USER ITEMS
BEAUTY&SPA
beauty_userid
beauty_friend_id
beauty_total_friendid
beauty_item_name
1 2931,2515,142,... 8 Nail Salons Ski...
2 2783,5253,216... 16 Hair Salons Ma...
3 2202,2143,479... 57 Day Spas Skin ...
4 5178,657,5188,... 36 Drugstores Co...
5 2143,4470,289... 79 Accessories Co...
6 4724,1915,206,... 95 Tattoo Piercing
7 2143,1710,452,... 454 Massage Acupu...
8 677,1435, 2 Day Spas Hair ...
9 142, 1 Day Spas Skin ...
10 5303,4013,269... 8 Day Spas Nail ...
11 4567,2354,540... 19 Skin Care Cos...
12 1915,1517,97,1... 273 Massage Physic...
VIEW_USER_FRIENDS
VIEW USER FRIEND
5333
NEXT

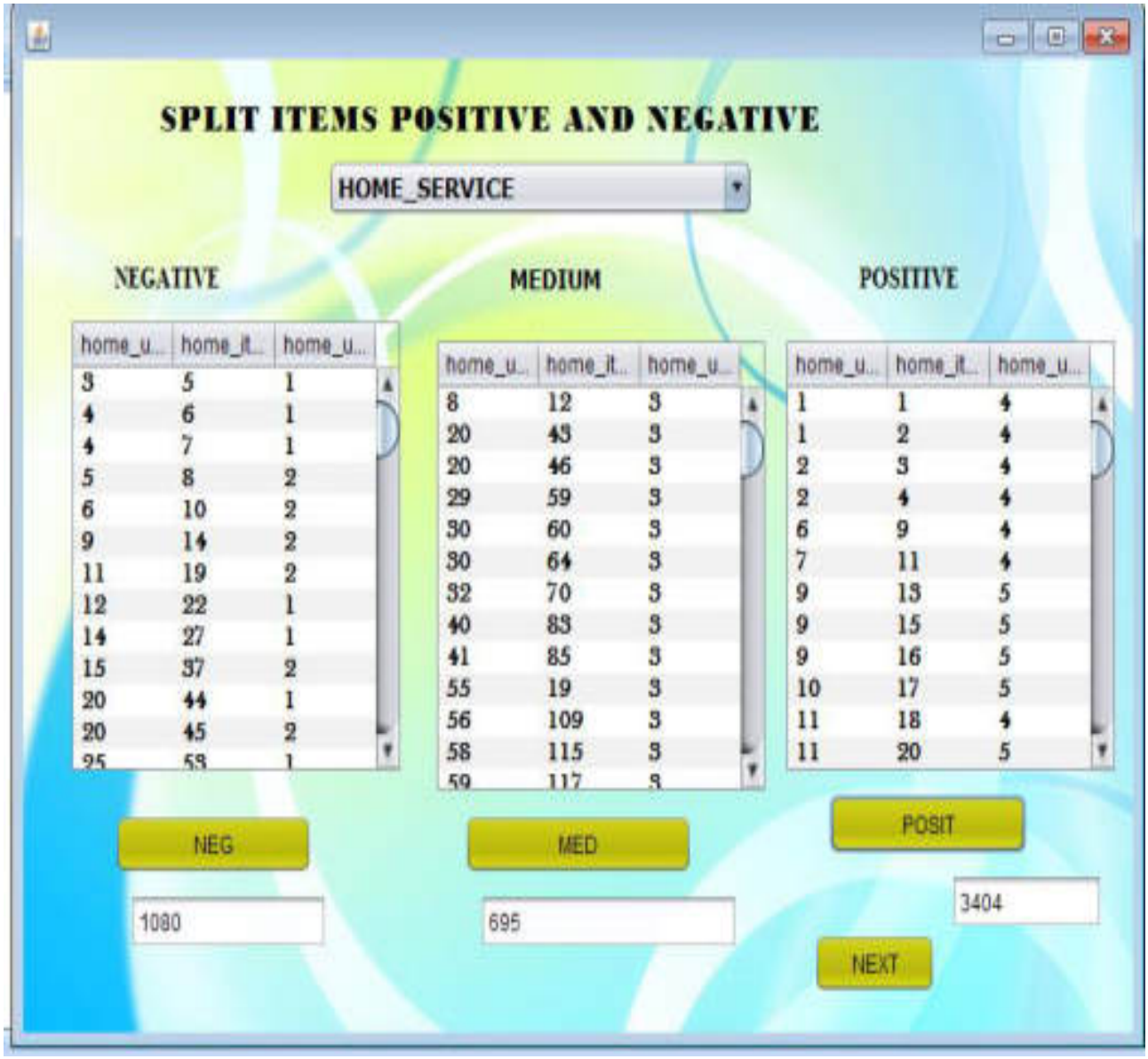

HOME_SERVICE
usend friendid home_total_friendid home_item_name
64 987,1100,461,1... 734 Building Suppli...
125 1312,238,344,4... 619 Roofing Contra...
217 1342,1714,2414... 549 Venues & Event...
591 1566,948,1892,... 568 Mobile Phones ...
888 176,1217,762,8... 529 Interior Design ...
1463 1195,833,1764,... 548 Contractors Lig...
1721 2025,176,1217,... 712 Mobile Phones ...
1926 630,898,2025,1... 563 Home Cleaning ...
2080 238,344,1342,8... 541 Property Servic...
2355 1627,1367,2035... 648 Real Estate Age...
10
NORMAL REPUTA


BEAUTY&SPA
beauty_usend
beauty_friend_id
beauty_total_frienid
beauty_item_name
142 2143,2399,103... 1288 Shoe Repair N...
156 3329,3446,456... 990 Guest Houses ...
283 2832,553,318,9... 1032 Day Spas Hair ...
507 522,2899,3760,... 858 Nail Salons Ski...
1949 413,2947,2643,... 866 Hair Removal ...
1980 4474,1130,115... 935 Day Spas Hair ...
2159 808,3959,5273,... 1146 Drugstores Cos...
2699 4048,4406,317... 1260 Barbers Cosme...
3189 2591,1822,518... 896 Hair Salons Na...
3698 2345,2897,577,... 885 Hair Salons Da...
3772 2766,4442,112... 1424 Hair Salons Ba...
3805 5364,1725,517... 1121 Hair Salons Ma...
4290 985,4474,1302 888 Day Spas Skin
16
NORMAL REPUTA


REPUTATION & REPUTATION COUNT
HOME_SERVICE
back
REPUTATION
TOP MOPST REPUTATION
home_itemsid
count
19 90
63 26
64 51
76 25
79 11
130 13
141 23
171 13
181 16
192 61
199 21
244 23
249 14
Message
ALL REPUTATION ITEMS
OK
TOP MOST
TOTAL REPUTATION ITEMS
24
TOTAL TOP MOST REPUTATION ITE...
jTextField1


REPUTATION & REPUTATION COUNT
HOME_SERVICE
back
REPUTATION
home_itemsid count
19 90
63 26
64 51
76 25
79 11
130 13
141 23
171 13
181 16
192 61
199 21
244 23
249 14
REPUTATION
TOP MOPST REPUTATION
home_itemsid count
19 90
63 26
64 51
76 25
141 23
192 61
199 21
244 23
296 22
TOP MOST
TOTAL REPUTATION ITEMS
24
TOTAL TOP MOST REPUTATION ITE...
9


SPLIT ITEMS POSITIVE AND NEGATIVE
HOME_SERVICE
NEGATIVE
MEDIUM
POSITIVE
home_u...
home_it...
home_u...
3 5 1
4 6 1
4 7 1
5 8 2
6 10 2
9 14 2
11 19 2
12 22 1
14 27 1
15 37 2
20 44 1
20 45 2
Message
NEGATIVE RATINGS FOR EACH PRODUCTS ITEMS
OK
NEG
MED
POSIT
1080
jTextField2
jTextField3
NEXT


SPLIT ITEMS POSITIVE AND NEGATIVE
HOME_SERVICE
NEGATIVE
MEDIUM
POSITIVE
home_u... home_it... home_u...
3 5 1
4 6 1
4 7 1
5 8 2
6 10 2
9 14 2
11 19 2
12 22 1
14 27 1
15 37 2
20 44 1
20 45 2
home_u... home_it... home_u...
8 12 3
20 43 3
20 46 3
29 59 3
30 60 3
30 64 3
32 70 3
40 83 3
41 85 3
55 19 3
56 109 3
58 115 3
home_u... home_it... home_u...
1 1 4
1 2 4
2 3 4
2 4 4
6 9 4
7 11 4
9 13 5
9 15 5
9 16 5
10 17 5
11 18 4
11 20 5
NEG
MED
POSIT
1080
695
3404
NEXT

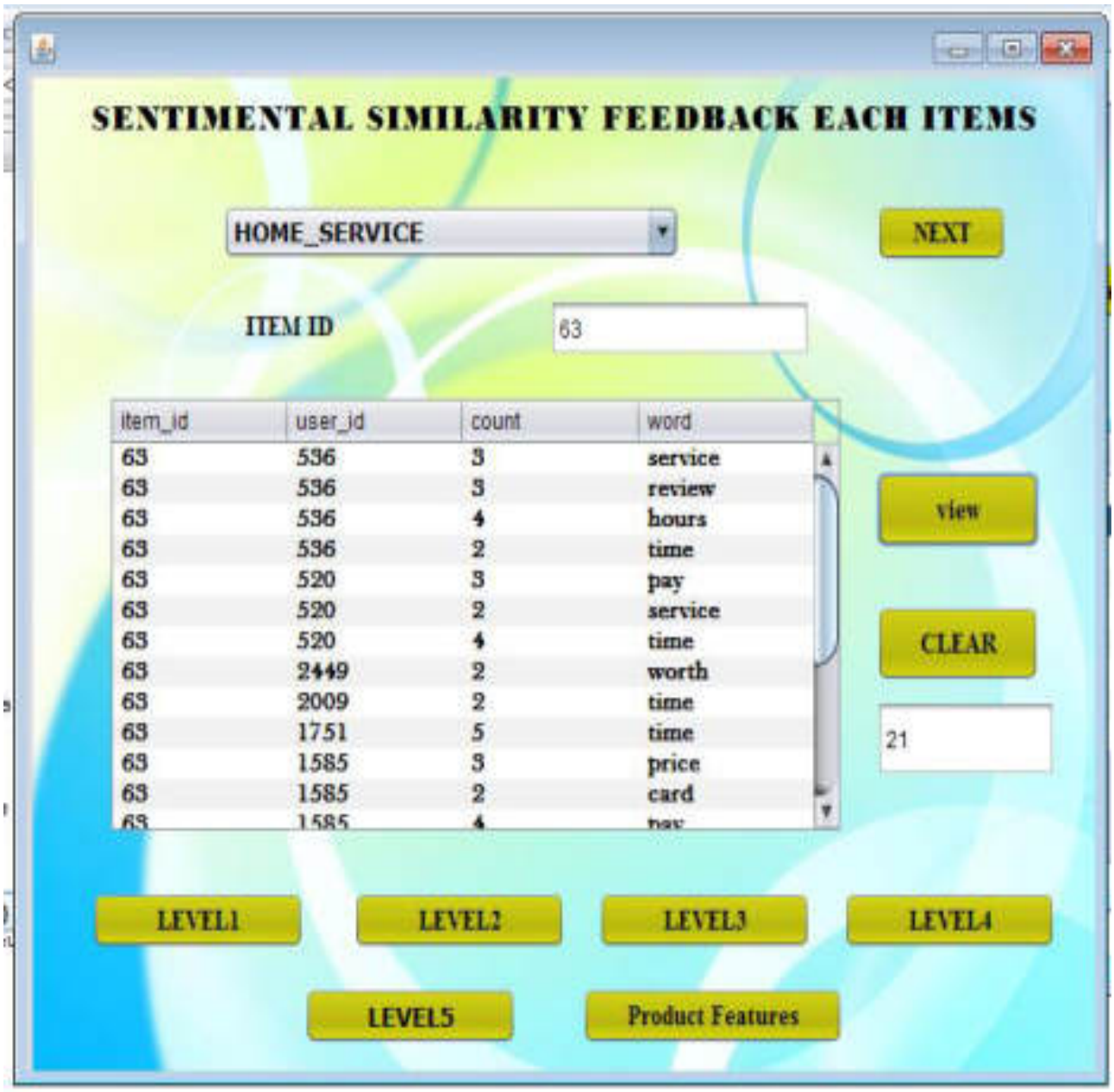

ALL USER FEEDBACK
BEAUTY&SPA
beauty_user_id
beauty_item_id
beauty_feedback
0 0 i absolutely love th...
0 2 I go here for threa...
0 3 UPDATE on the l...
0 1 This isn't my usual...
1 4 WARNING, CHE...
1 5 Awe I love Marcy ...
1 6 This shop has a ba...
10 39 The massages are t...
100 319 Typical Walgreens...
10 40 Painless, Careful, a...
100 321 I like Rick's (obvio...
100 322 Great Walgreens. ...
1000 9679 best salon in vanco
NORMAL REPUTA
NEXT
TOTAL USER FEDBACK
21345

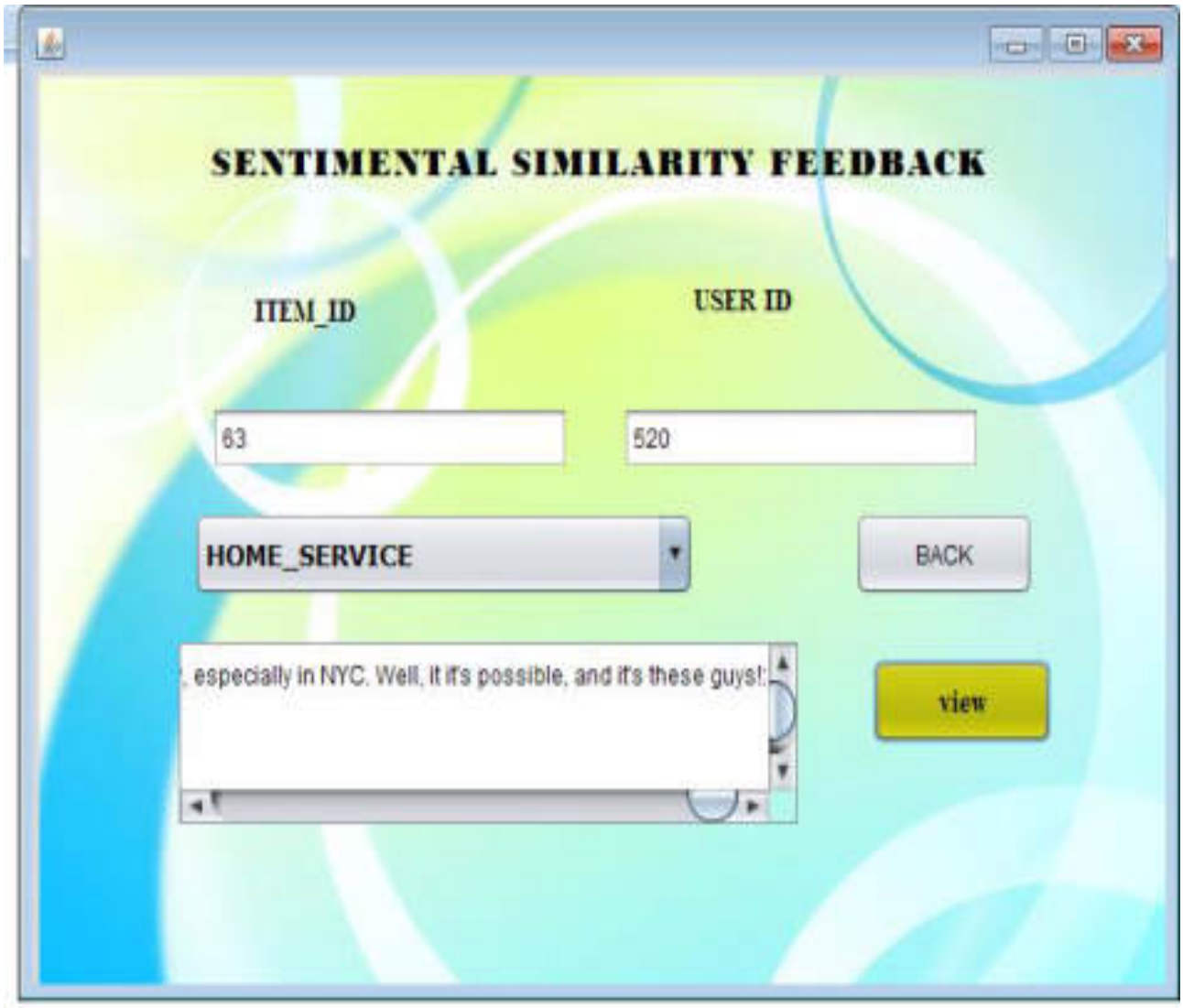

ALL USER FEEDBACK
HOTEL&TRAVEL
hotels_userid
hotel_itemid
hotels_feedback
1991 99 Duck. Go with the...
1992 869 If a romantic spa ...
1993 23 I recently stayed h...
1993 325 Everyone who lives...
1993 545 For my boyfriend's...
1994 3462 if there was the op...
1995 410 I love Southwest A...
1996 100 Ahhh, memories fl...
1997 1562 Review is on SPO...
1998 208 Loved my time her...
1998 2422 I was VERY happy...
1998 3463 My husband surpri...
1998 400 You get what you
NORMAL REPUTA
NEXT
TOTAL USER FEDBACK
16446


SENTIMENTAL SIMILARITY FEEDBACK EACH ITEMS
HOME_SERVICE
NEXT
ITEM ID
63
item_id user_id count word
63 536 5 all
63 536 2 highest
63 520 13 all
63 520 2 every
63 433 3 all
63 336 3 all
63 30 3 all
63 268 2 every
63 2451 2 all
63 2449 2 every
63 2446 2 all
63 2290 2 all
63 1959 4 all
view
CLEAR
30
LEVEL1
LEVEL2
LEVEL3
LEVEL4
LEVEL5
Product Features


SENTIMENTAL SIMILARITY FEEDBACK EACH ITEMS
HOME_SERVICE
NEXT
ITEM ID
63
item_id user_id count word
63 536 3 better
63 536 3 better
63 536 2 too
63 536 4 very
63 520 2 a lot
63 520 3 especially
63 520 3 how
63 520 2 lot
63 520 4 over
63 520 2 too
63 520 3 very
63 30 2 over
view
CLEAR
48
LEVEL1
LEVEL2
LEVEL3
LEVEL4
LEVEL5
Product Features


SENTIMENTAL SIMILARITY FEEDBACK EACH ITEMS
HOME_SERVICE
NEXT
ITEM ID
63
item_id user_id count word
63 536 3 more
63 536 3 so
63 536 2 great
63 520 2 even
63 520 7 so
63 336 2 great
63 30 2 so
63 30 3 great
63 2449 2 so
63 2290 3 so
63 2290 2 great
63 2009 2 even
63 1980 2 even
view
CLEAR
33
LEVEL1
LEVEL2
LEVEL3
LEVEL4
LEVEL5
Product Features


SENTIMENTAL SIMILARITY FEEDBACK EACH ITEMS
HOME_SERVICE
NEXT
ITEM ID
63
item_id user_id count word
63 536 3 service
63 536 3 review
63 536 4 hours
63 536 2 time
63 520 3 pay
63 520 2 service
63 520 4 time
63 2449 2 worth
63 2009 2 time
63 1751 5 time
63 1585 3 price
63 1585 2 card
view
CLEAR
21
LEVEL1
LEVEL2
LEVEL3
LEVEL4
LEVEL5
Product Features


SENTIMENTAL SIMILARITY FEEDBACK
ITEM_ID
USER ID
63
520
HOME_SERVICE
BACK
especially in NYC. Well, it it's possible, and it's these guys!:
view


Line Chart Example with JFreechart
ALL PRODUCT CATEGORIES
RMSE
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
0
5
10
20
30
40
50
60
100
200
500
1000
1500
2000
LEVEL
Beauty
Hotel&Travel
Pets
Shopping
Home Service

8. CONCLUSION

In this paper, a recommendation modelis proposed by mining sentiment information from social users'reviews. We fuseusersentimentsimilarity, interpersonalsentimentinfluence, and itemreputation similarity into a unified matrix factorization frameworkto achieve the rating prediction task. In particular, we use social users' sentiment to denote user preferences. Besides, we build a new relationship named interpersonal sentiment influence between the user and friends, which reflectshow users' friends influence usersin a sentimental angle. Whatismore, as long as we obtainuser's textual reviews, we can quantitively measure user's sentiment,and we leverage items' sentiment distribution among users to infer item's reputation. The experiment results demonstrate that the three sentimental factors make great contributions to the rating prediction. Also, it shows significant improvements over existing approaches on a real-world dataset.In our future work, we can considermore linguistic rules when analyzing the context, and we can enrich the sentiment dictionaries to apply fine-grained sentiment analysis. Besides, wecanadapt or develop other hybrid factorization models such as tensor factorization or deep learning technique to integrate phrase-level sentiment analysis.

9. REFERENCES

[1] R. Salakhutdinov, and A. Mnih, "Probabilistic matrix factorization," in *NIPS*, 2008.

[2] X. Yang, H. Steck, and Y. Liu, "Circle-based recommendation in online social networks, " in *Proc. 18th ACM SIGKDD Int. Conf. KDD*, New York, NY, USA, Aug. 2012, pp. 1267–1275.

[3] M. Jiang, P. Cui, R. Liu, Q. Yang, F. Wang, W. Zhu, and S. Yang, "Social contextual recommendation," in proc. *21st ACM Int. CIKM*, 2012, pp. 45-54.

[4] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in *Proc. ACM conf. RecSys*, Barcelona, Spain. 2010, pp. 135-142.

[5] Z. Fu, X. Sun, Q. Liu, et al., "Achieving Efficient Cloud Search Services: Multi-Keyword Ranked Search over Encrypted Cloud Data Supporting Parallel Computing," *IEICE Transactions on Communications*, 2015, 98(1):190-200.

[6] G. Ganu, N. Elhadad, A Marian, "Beyond the stars: Improving rating predictions using Review text content," in *12th International Workshop on the Web and Databases (WebDB 2009)*. pp. 1-6.

[7] J. Xu, X. Zheng, W. Ding, "Personalized recommendation based on reviews and ratings alleviating the sparsity problem of collaborative filtering," *IEEE International Conference on e-business Engineering*. 2012, pp. 9-16.

[8] X. Qian, H. Feng, G. Zhao, and T. Mei, "Personalized recommendation combining user interest and social circle," *IEEE Trans. Knowledge and data engineering*. 2014, pp. 1763-1777.

[9] H. Feng, and X. Qian, "Recommendation via user's personality and social contextual," in

Proc. 22nd ACM international conference on information & knowledge management. 2013, pp. 1521-1524.

[10] Z. Fu, K. Ren, J. Shu, et al., "Enabling Personalized Search over Encrypted Outsourced Data with Efficiency Improvement," *IEEE Transactions on Parallel & Distributed Systems*, 2015:1-1.

[11] D.M. Blei, A.Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of machine learning research* 3. 2003, pp. 993-1022.

[12] W. Zhang, G. Ding, L. Chen, C. Li , and C. Zhang, " Generating virtual ratings from Chinese reviews to augment online recommendations," *ACM TIST*, vol.4, no.1. 2013, pp. 1-17.

[13] Z. Xia, X. Wang, X. Sun, and Q. Wang, "A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 2, 2015, pp. 340-352.

[14] J. Weston, R. J. Weiss, H. Yee, "Nonlinear latent factorization by embedding multiple user interests," *7th ACM, RecSys*, 2013, pp. 65-68.

[15] J. Huang, X. Cheng, J. Guo, H. Shen, and K. Yang, " Social recommendation with interpersonal influence," in *Proc. ECAI*, 2010, pp. 601-606.

[16] Y. Lu, M. Castellanos, U. Dayal, C. Zhai, "Automatic construction of a context-aware sentiment lexicon: an optimization approach," *World Wide Web Conference Series*. 2011, pp. 347-356.

[17] T. Kawashima, T. Ogawa, M. Haseyama, "A rating prediction method for e-commerce application using ordinal regression based on LDA with multi-modal features," *IEEE 2nd Global Conference on Consumer Electronics (GCCE)*. 2013,
pp. 260-261.

[18] K.H. L. Tso-Sutter, L. B. Marinho, L. Schmidt-Thieme, “Tag-aware recommender systems by fusion of collaborative filtering algorithms,” in *Proceedings of the 2008 ACM symposium on Applied computing*, 2008, pp. 1995-1999.

[19] B. Wang, Y. Min, Y. Huang, X. Li, F. Wu, “ Review rating prediction based on the content and weighting strong social relation of reviewers,” in *Proceedings of the 2013 international workshop of Mining unstructured big data using natural language processing, ACM*. 2013, pp. 23-30.

[20] F. Li, N. Liu, H. Jin, K. Zhao, Q. Yang, X. Zhu, “Incorporating reviewer and product information for review rating prediction,” in *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence*, 2011, pp. 1820-1825.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"A Hybrid Trust based Graph Cluster Algorithm for Online Social Network (OSN) Data."

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
COMPUTER SCIENCE AND ENGINEERING

Submitted by

G. Sairam (17H71A0539)

G. Deepika (17H71A0509)

B. Sai Deepika (17H71A0537)

K. Charithasri (17H71A0507)

Under the Esteemed Guidance of

Mrs. V. Lakshmi Chetana

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist., A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist., A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in
2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"A Hybrid Trust based Graph Cluster Algorithm for Online Social Network (OSN) Data."** is a bonafide work carried out by **G. Sairam (17H71A0539), G. Deepika (17H71A0509), B. Sai Deepika (17H71A0537), K. Charithasri (17H71A0507)** in partial fulfilment for the award of degree of Bachelor of Technology in **Computer Science and Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

(Mrs. V. Lakshmi Chetana)
Project Supervisor

(Mr. D. Prasad)
Head of the Department

(Dr. K. Srinivas)
Principal

Examiner 1

Examiner 2

ACKNOWLEDGEMENT

We would like to take this opportunity to express my deepest appreciation to the following people for their valuable contributions and assistance with this Project.

Initially, we would like to thank my project supervisor, **Mrs. V. Lakshmi Chetana, Asst. Professor** for her guidance and support, especially for her valuable ideas and knowledge provided throughout the Project. Her expertise and experience in **guide specialization** makes valuable comments and suggestions have been very useful in solving problems encountered during the Project.

We have the immense pleasure in expressing my thanks and deep sense of gratitude to**, Mr. D. Prasad, Head of the Department, Computer Science and Engineering** for extending necessary facilities for the completion of the Project.

We whole heartedly acknowledge **Dr. K. Srinivas, Principal and Prof D. PANDURANGA RAO, CEO** for giving opportunity to execute this Project. We also extend my thanks to all faculty members of **Computer Science and Engineering**, for their valuable guidance and encouragement in this Project.

We would like to extend my warm appreciation to all my friends for sharing us their knowledge, valuable contributions and help with this Project.

Finally, my special thanks go to my family for their continuous support and help throughout my academic years and for their continual support and encouragement for this Project.

Batch Members Names and Roll Numbers

G. Sairam (17H71A0539)

G. Deepika (17H71A0509)

B. Sai Deepika (17H71A0537)

K. Charithasri (17H71A0507)

DECLARATION

We **G. Sairam, G. Deepika, B. Sai Deepika, K. Charithasri** of the main-Project "**A Hybrid Trust based Graph Cluster Algorithm for Online Social Network (OSN) Data**", hereby declare that the matter embodied in this Project is the genuine work done by us and has not been submitted either to this university or to any other university/institute for the fulfilment of the requirement of any course of study.

G. Sairam (17H71A0539)

G. Deepika (17H71A0509)

B. Sai Deepika (17H71A0537)

K. Charithasri (17H71A0507)

ABSTRACT

Online social networking is a way to establish a well-balanced connection between people. A larger number of online social network (OSN) sites such as Facebook, WhatsApp, and LinkedIn are used to find the contextual relations among the users and their friend list. In the current system, if any user can generate a friend request then send a request even if he doesn't have any connection with random person. The user can also accept unknown requests which can be a threat or insecurity. The current services of social networking suggest friends based on the network of the respective individual. This may not be the perfect way to recommend friends to the respective user due to privacy issues. In this proposed work, a novel social network friend recommended system is developed using the integrity of each user and their friend's list. In this work, a graph clustering algorithm is implemented using the MD5 integrity technique as trust-based friend recommended system. Experimental results proved that the present model is better than the traditional OSN friend recommended systems in terms of accuracy is concerned.

LIST OF FIGURES

CONTENTS

PAGE NO

1) INTRODUCTION

1.1) BACKGROUND

Social Networking Sites (SNS) is a fast-growing world in the present scenario. This is one of the most powerful media for communicating various users in one platform. OSN is an online service or site to facilitate social interaction to help individuals find others of a common interest. Security and privacy are two important aspects of SNS.

It is very important to secure the user's data and provide privacy for the user's information. Online social networking is a way to establish a well-balanced connection between people. A larger number of online social network (OSN) sites such as Facebook, WhatsApp, and LinkedIn are used to find the contextual relations among the users and their friend list.

Various ponders have indicated that members unmistakably speak to the feeble connection for security in OSN and that they are powerless against a few kinds of social building assaults. This is in part brought about by an absence of attention to the results of straightforward and probably private activities, such as tolerating contact demands, or labelling pictures, just as correspondence tasks like remarking on proles or posting on dividers. The low level of ease of use of protection controls by the SNS, and above all intrinsic presumptions about different members and trust in different proles, which are really an ideal trademark, unquestionably add to the issue.

1.2) TRUST-BASED RECOMMENDATION SYSTEM

The recommendations may show you people you might want to add as a friend who you haven't seen on SNS yet or who may be new to SNS. We identify some open challenges that all trust models are facing. Interpersonal organization Services (SNS) are definitely changing the manner in which individuals communicate. We use trust based recommendation to overcome these challenges.

To generate recommendations we use graph clustering algorithm. These recommendations are given accurately using the concept called collaborative filtering. In numerous person to person communication, users utilize their genuine name to speak to their records and get suggestions. So, two types of recommendations are used. One is based on the graph clustering

i.e. a graph formed from the disjoint union of complete graphs. The other type is the collaborative filtering and content-based which can be combined and called as hybrid recommendation system. Hybrid recommender systems combine two or more recommendation strategies in different ways to benefit from their complementary advantages.

1.3) MOTIVATION OF THE PROJECT

In particular, we use the friend recommendation application as it has gained special importance from both the social network administrators and the users. We use machine learning algorithms as machine learning models learn, identify patterns, and make decisions with minimal intervention from humans.

Ideally, machines increase accuracy and efficiency and remove (or greatly reduce) the possibility of human error. The graph clustering and MD5 algorithms are used because recommendations are genuine and better compared to the existing algorithm. We also use MD5 algorithm to get trust based recommendations. From the users' side, friend recommendations help them grow their social contacts and explore for new friends based on their own interests.

1.4) PROBLEM STATEMENT

The accuracy of suggestions in the current existing system is very poor. The current existing system does not work properly with the community and wrong suggestions are given. There can be fake accounts which may lead to fraud like cybercrimes. The user can accidentally accept fraud friend requests and can become a victim in cybercrimes. As we use the graph clustering, a certain group forms a cluster and based on this cluster we get the friend suggestions.

1.5) SCOPE OF THE PROJECT

We use graph clustering which means we set a number of nodes into a group called cluster based on the similarities of these nodes. The importance of graph clustering is that it is desired to partition the graphs in such a way so as to minimize the weights of the edges across the partitions. In general way, we would like to partition the graphs into k groups of nodes. One is to make the clustering algorithm applied to the discovery issue of the community and the standard should be carefully chosen.

The second is to minimize the complexities of time and space for large data and to ensure the algorithm's accuracy. The brief idea of the algorithm should be suggested before introducing the algorithm process. Choose the relationship measure between nodes is proposed for the community discovery issue. The measure of relationship is a high-dimensional relationship, so the next step is to introduce characteristics of high-dimensional relationship. The previous idea is introduced separately in large data structures.

2) LITERATURE SURVEY

Nissenbaum [5] clarified a definitive estimation of the information be ensured so as to protect the relevant trustworthiness of the online shared information. Data assembled from internet based life for examination reasons for existing is commonly unintended and frequently immaterial. Be that as it may, it might be identified with the private exercises of an individual, for instance, religion or political affiliations [6].

Client produced content via web-based networking media may incorporate clients' encounters, feelings, and information. Furthermore, it might likewise incorporate private information, for instance, name, sexual orientation, area, and private photographs [7]. Online-shared data is electronically put away and is along these lines lasting, replicable, and re shareable [8]. OSN clients for the most part face the difficulties of dealing with their social character while trading off their social security.

Considering this worldwide number of clients, protection is one of the conspicuous and basic issues with respect to OSNs. Different protection issues are encouraged in view of OSNs, for example, observation, in which the social circle of OSNs changes to a business circle and OSN specialist organizations regulate client activities for showcase power get to control. Standard OSNs share clients' close to home information with outsiders for ad purposes that might be misused [9].

Great dangers are online dangers that make OSN clients defenceless, yet in addition other online clients who don't utilize any OSN. The second kind of dangers is present day dangers, which are identified with OSN clients simply because of the OSN framework that can bargain client protection and security [10].

Great dangers have been an issue as far back as the advancement of the Internet. These dangers are spam [11], malware [12], and phishing [13], or cross-site scripting (XSS) assaults [14]. In spite of the fact that specialists and businesses have tended to these dangers in the past with the creation of OSNs, they can spread in another way and more rapidly than any other time in recent memory. Great dangers are utilized to remove the individual data of clients, which are shared through an OSN, not exclusively to assault the objective clients yet in addition their friends by modifying the risk to associate to clients' private qualities.

Malware represents vindictive programming. The most noticeably terrible malware case is to get to clients' certifications and mimic them to send messages to their friends. For instance, the Koobface malware was spread through OSNs, for example, MySpace, Facebook, and Twitter. It was utilized to gather login certifications and make the objective contaminated PC a piece of a botnet [15].

Baltazar et al. [15] is focused to fortify the clients' notoriety in online commercial centers like eBay. The chance to make accounts unreservedly drives Sybil clients to make different records and causes the exercise in futility and critical financial misfortunes for cheated clients. Here Baltazar distinguished the imperfections in sybil's procedure. He focused on chance to make accounts unreservedly leads sybil clients to make different assaults.

3) SYSTEM ANALYSIS

3.1) EXIXTING SYSTEM

- By analysing the previous methodologies, in complex networks, community detection research is divided into two categories.
- The first category includes disjoint community research dealing with nodes belonging to a single community.
- The second category involves finding the communities that overlap in the network where nodes belong to more than one community.
- The need to overlap community detection algorithms arises when networks have structures where a clear assignment of a node to a community is not possible or not desirable.
- Individuals are usually part of several natural groups in social networks.

DRAWBACKS:

- The accuracy of suggestions is very poor
- It does not work properly with the community and wrong suggestions are given.

3.2) PROPOSED SYSTEM

- We use graph clustering which means we set a number of nodes into a group called cluster based on the similarities of these nodes.
- The importance of graph clustering is that it is desired to partition the graphs in such a way so as to minimize the weights of the edges across the partitions. In general way, we would like to partition the graphs into k groups of nodes.
- One is to make the clustering algorithm applied to the discovery issue of the community and the standard should be carefully chosen.

- The second is to minimize the complexities of time and space for large data and to ensure the algorithm's accuracy.
- The brief idea of the algorithm should be suggested before introducing the algorithm process.
- Choose the relationship measure between nodes is proposed for the community discovery issue.
- The measure of relationship is a high-dimensional relationship, so the next step is to introduce characteristics of high-dimensional relationship.
- The previous idea is introduced separately in large data structures.
- Uses MD5 algorithm for evaluating trust based systems.
- In the existing system the recommendation may not be genuine i.e. can be fake.

ADVANTAGES:

- It gives better results.
- It works more accurately compared to previous work.
- We are using k-mean machine learning algorithm.

3.3) SYSTEM SPECIFICATIONS

HARDWARE REQUIREMENTS SPECIFICATION:

- System : Pentium IV 2.4 GHz.
- Hard Disk : 40 GB.
- Monitor : 15 VGA Colour.
- Mouse : Logitech.
- Ram : 512 Mb.
- Windows : v 10

SOFTWARE REQUIREMENTS SPECIFICATION:

- Operating system : Windows XP/7.
- Coding Language : JAVA 8.0
- IDE : ECLIPSE oxygen 4.7
- Database : MYSQL
- Database platform : XAMPP v3.3.0

3.4) SYSTEM ARCHITECTURE

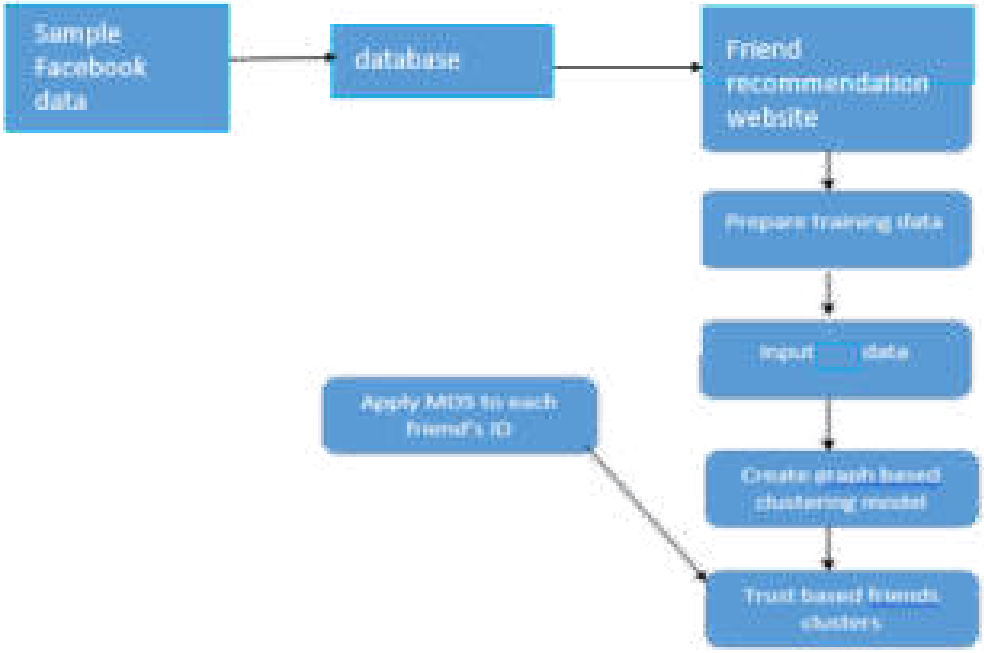


Fig 3.1: System Architecture

1) We take sample Facebook data and store it into our database.

2) This data is used for the friend recommendation and a website is created for displaying the data.

3) The training data is prepared and stored into the database.

4) The test and training data are given as an input.

5) Using the graph based clustering algorithm, we create a clustering model that is to be applied to the friend recommendations.

6) Then as a result we get trust based friend clusters.

7) On these clusters the MD5 integrity is applied to each users ID.

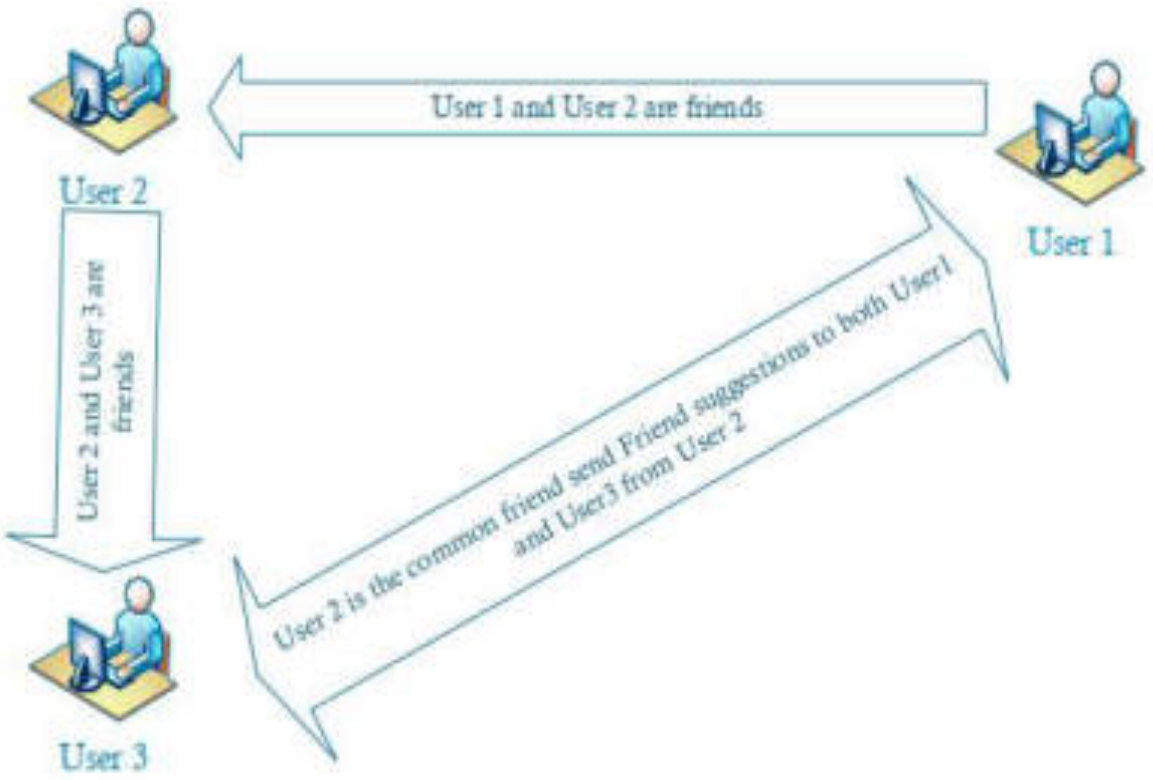


Fig 3.2: Collaborative Filtering

Collaborative filtering filters information by using the interactions and data collected by the system from other users. It's based on the idea that people who agreed in their evaluation of certain items are likely to agree again in the future.

The concept is simple: when we want to find a new movie to watch we'll often ask our friends for recommendations. Naturally, we have greater trust in the recommendations from friends who share tastes similar to our own.

Most collaborative filtering systems apply the so-called similarity index-based technique. In the neighborhood-based approach, a number of users are selected based on their similarity to the active user. Inference for the active user is made by calculating a weighted average of the ratings of the selected users.

Collaborative-filtering systems focus on the relationship between users and items. The similarity of items is determined by the similarity of the ratings of those items by the users who have rated both items.

There are two classes of Collaborative Filtering:

- User-based, which measures the similarity between target users and other users.
- Item-based, which measures the similarity between the items that target users rate or interact with and other items.

4) DESIGN

Systems design is the process of defining elements of a system like modules, architecture, components and their interfaces and data for a system based on the specified requirements. It is the process of defining, developing and designing systems which satisfies the specific needs and requirements of a business or organization.

4.1) UML DIAGRAMS:

UML (Unified Modeling Language) is a general-purpose, graphical modeling language in the field of Software Engineering. UML is used to specify, visualize, construct, and document the artifacts (major elements) of the software system. It was initially developed by Grady Booch, Ivar Jacobson, and James Rum Baugh in 1994-95 at rational software, and its further development was carried out through 1996. In 1997, it got adopted as a standard by the Object Management Group. UML (Unified Modeling Language) is a general-purpose, graphical modeling language in the field of Software Engineering. UML is used to specify, visualize, construct, and document the artifacts (major elements) of the software system. It was initially developed by Grady Booch, Ivar Jacobson, and James Rum Baugh in 1994-95 at rational software, and its further development was carried out through 1996. In 1997, it got adopted as a standard by the Object Management Group.

What is UML?

The UML stands for Unified modeling language, is a standardized general-purpose visual modeling language in the field of Software Engineering. It is used for specifying, visualizing, constructing, and documenting the primary artifacts of the software system. It helps in designing and characterizing, especially those software systems that incorporate the concept of Object orientation. It describes the working of both the software and hardware systems. The UML was developed in 1994-95 by Grady Booch, Ivar Jacobson, and James Rumbaugh at the Rational Software.

The Object Management Group (OMG) is an association of several companies that controls the open standard UML. The OMG was established to build an open standard that mainly supports the interoperability of object-oriented systems. It is not restricted within the boundaries, but it can also be utilized for modeling the non-software systems. The OMG is best recognized for the Common Object Request Broker Architecture (CORBA) standards.

Goals of UML

- Since it is a general-purpose modeling language, it can be utilized by all the modelers.
- UML came into existence after the introduction of object-oriented concepts to systemize and consolidate the object-oriented development, due to the absence of standard methods at that time.
- The UML diagrams are made for business users, developers, ordinary people, or anyone who is looking forward to understand the system, such that the system can be software or non-software.
- Thus it can be concluded that the UML is a simple modeling approach that is used to model all the practical systems.

Characteristics of UML

The UML has the following features:

- It is a generalized modeling language.
- It is distinct from other programming languages like C++, Python, etc.
- It is interrelated to object-oriented analysis and design.
- It is used to visualize the workflow of the system.
- It is a pictorial language, used to generate powerful modeling artifacts.

Conceptual Modeling

Before moving ahead with the concept of UML, we should first understand the basics of the conceptual model.

A conceptual model is composed of several interrelated concepts. It makes it easy to understand the objects and how they interact with each other. This is the first step before drawing UML diagrams.

Following are some object-oriented concepts that are needed to begin with UML:

- **Object:** An object is a real world entity. There are many objects present within a single system. It is a fundamental building block of UML.

- **Class:** A class is a software blueprint for objects, which means that it defines the variables and methods common to all the objects of a particular type.
- **Abstraction:** Abstraction is the process of portraying the essential characteristics of an object to the users while hiding the irrelevant information. Basically, it is used to envision the functioning of an object.
- **Inheritance:** Inheritance is the process of deriving a new class from the existing ones.
- **Polymorphism:** It is a mechanism of representing objects having multiple forms used for different purposes.
- **Encapsulation:** It binds the data and the object together as a single unit, enabling tight coupling between them.

You can also create your own set of diagrams to meet your requirements. Diagrams are generally made in an incremental and iterative way.

There are two broad categories of diagrams and they are again divided into subcategories −

- Structural Diagrams
- Behavioral Diagrams

Structural Diagrams

The structural diagrams represent the static aspect of the system. These static aspects represent those parts of a diagram, which forms the main structure and are therefore stable.

These static parts are represented by classes, interfaces, objects, components, and nodes. The four structural diagrams are −

- Class diagram
- Object diagram
- Component diagram
- Deployment diagram

BEHAVIOURAL DIAGRAMS:

• Behavioral diagrams basically capture the dynamic aspect of a system. Dynamic aspect can be further described as the changing/moving parts of a system.

UML has the following five types of behavioral diagrams –

- Use case diagram
- Sequence diagram
- Collaboration diagram
- State chart diagram
- Activity diagram

USECASE DIAGRAM:

Fig 4.1: Use Case Diagram

SEQUENCE DIAGRAM

The sequence diagram represents the flow of messages in the system and is also termed as an event diagram. It helps in envisioning several dynamic scenarios. It portrays the communication between any two lifelines as a time-ordered sequence of events, such that these lifelines took part at the run time. In UML, the lifeline is represented by a vertical bar, whereas the message flow is represented by a vertical dotted line that extends across the bottom of the page. It incorporates the iterations as well as branching.

Purpose of a Sequence Diagram

1. To model high-level interaction among active objects within a system.

2. To model interaction among objects inside a collaboration realizing a use case.

3. It either models generic interactions or some certain instances of interaction.

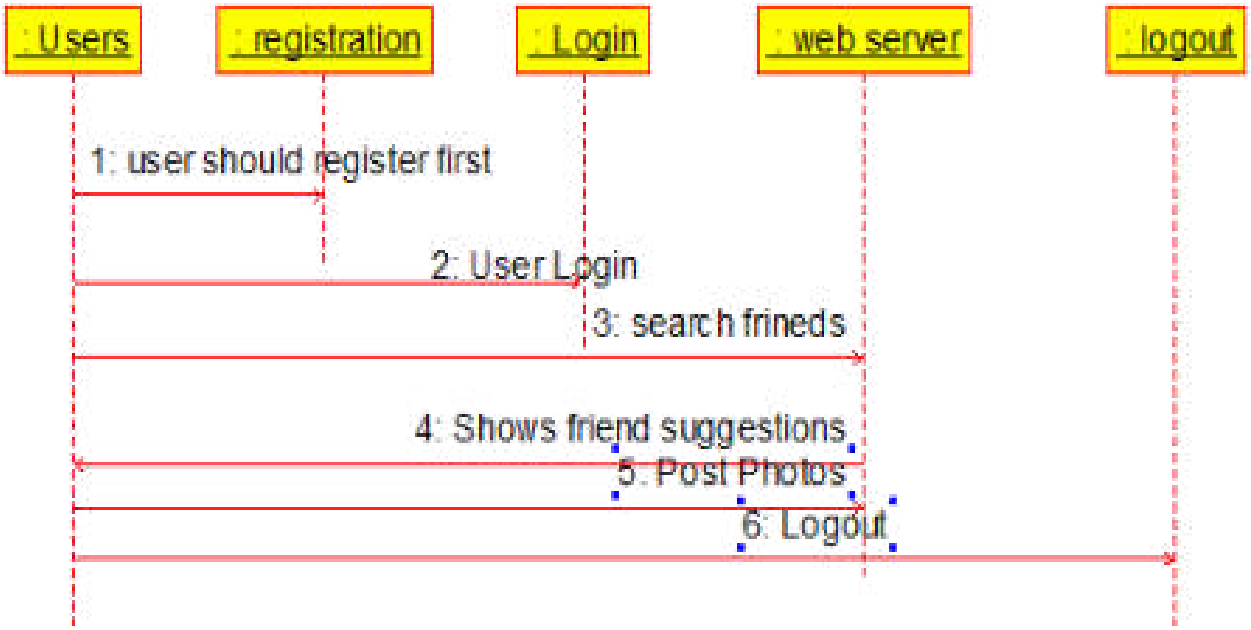


Fig 4.2: Sequence Diagram

CLASS DIAGRAM:

Model, objects are entities that combine state (i.e., data), behavior (i.e., procedures, or methods) and identity (unique existence among all other objects). The structure and behavior of an object are defined by a class, which is a definition, or blueprint, of all objects of a specific type. An object must be explicitly created based on a class and an object thus created is considered to be an instance of that class. An object is similar to a structure, with the addition of method pointers, member access control, and an implicit data member which locates instances of the class (i.e. actual objects of that class) in the class hierarchy (essential for runtime inheritance features)

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, and the relationships between the classes.

The class diagram is the main building block in the object oriented modeling. It is used both for general conceptual modeling of the semantics of the application, and for detailed modeling translating the models into programming code. The classes in a class diagram represent both the main objects and or interactions in the application and the objects to be programmed. In the class diagram these classes are represented with boxes which contain the two parts:

- The upper part holds the name of the class.

- The middle part contains the attributes of the class.

- The lower part contains the operations of the class.

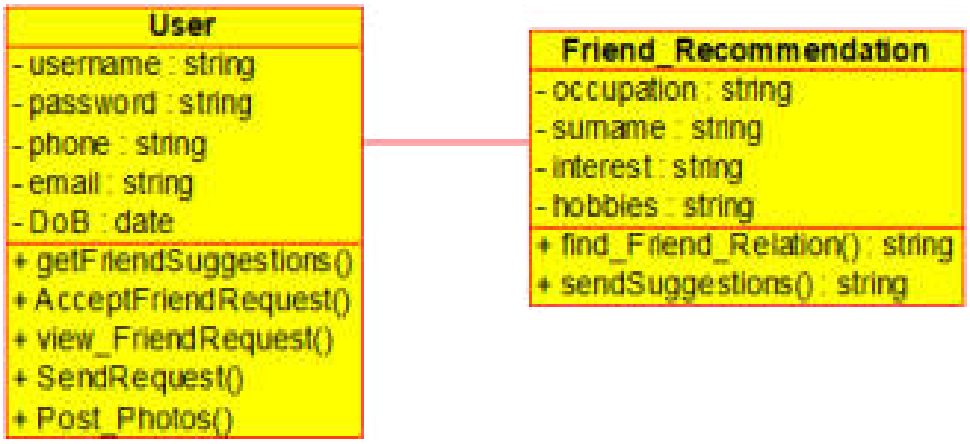


Fig 4.3: Class Diagram

COMMUNICATION DIAGRAM:

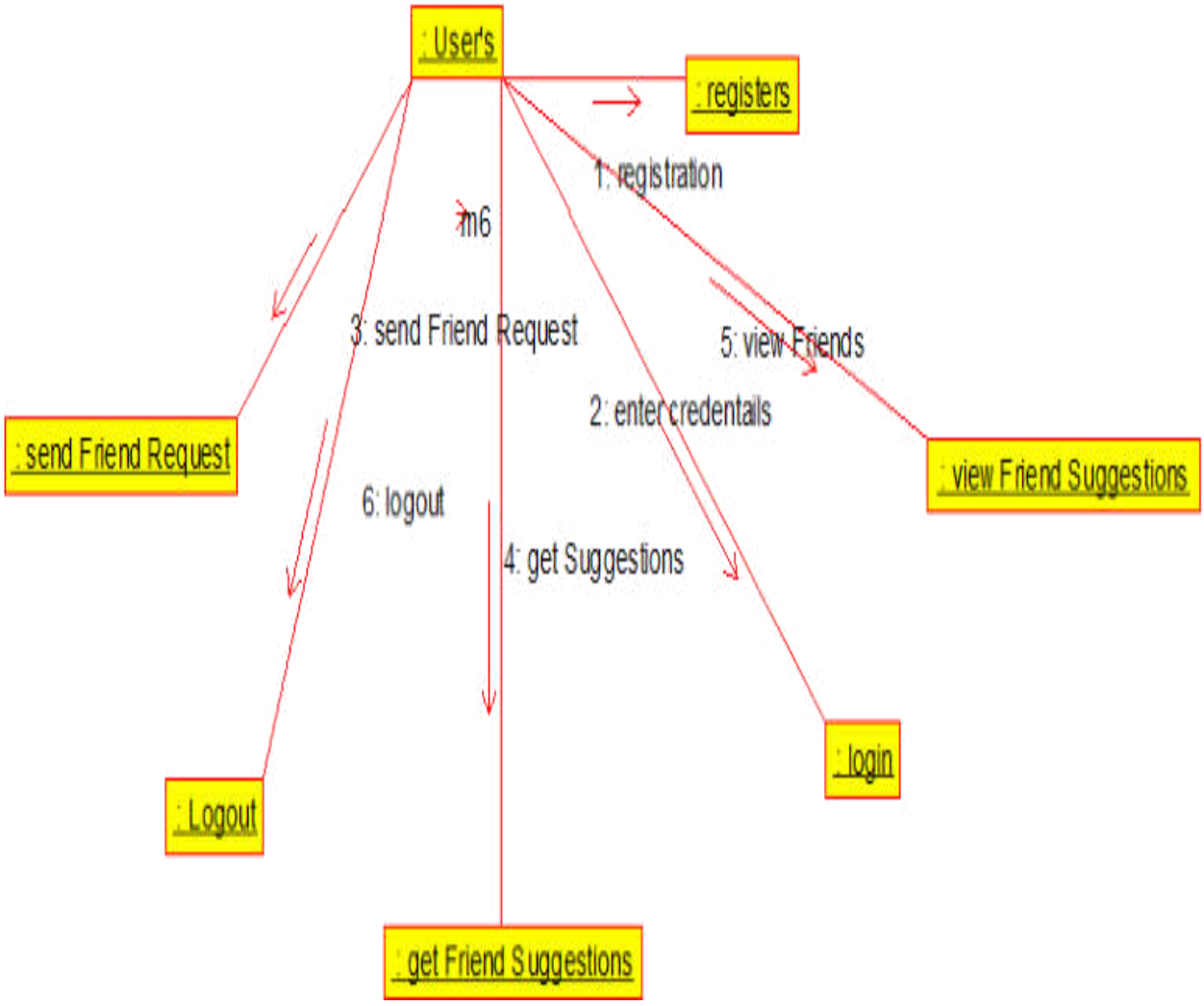


Fig 4.4: Communication Diagram

5) IMPLEMENTATION

5.1) SOFTWARE ENVIRONMENT:

5.1.1) Java Technology:

Java technology is both a programming language and a platform.

The Java Programming Language

The Java programming language is a high-level language that can be characterized by all of the following buzzwords:

- Simple
- Architecture neutral
- Object oriented
- Portable
- Distributed
- High performance
- Interpreted
- Multithreaded
- Robust
- Dynamic
- Secure

With most programming languages, you either compile or interpret a program so that you can run it on your computer. The Java programming language is unusual in that a program is both compiled and interpreted. With the compiler, first you translate a program into an intermediate language called *Java byte codes* —the platform-independent codes interpreted by the interpreter on the Java platform. The interpreter parses and runs each Java byte code instruction on the computer. Compilation happens just once; interpretation occurs each time the program is executed. The following figure illustrates how this works.

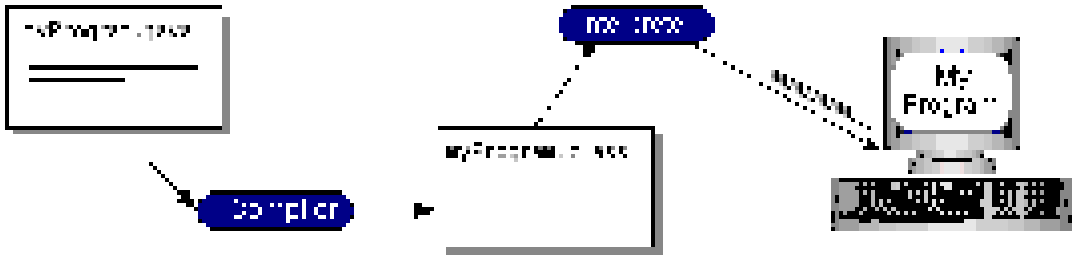


Fig 5.1: JAVA Compilation Process

You can think of Java byte codes as the machine code instructions for the *Java Virtual Machine* (Java VM). Every Java interpreter, whether it's a development tool or a Web browser that can run applets, is an implementation of the Java VM. Java byte codes help make "write once, run anywhere" possible. You can compile your program into byte codes on any platform that has a Java compiler. The byte codes can then be run on any implementation of the Java VM. That means that as long as a computer has a Java VM, the same program written in the Java programming language can run on Windows 2000, a Solaris workstation, or on an iMac.

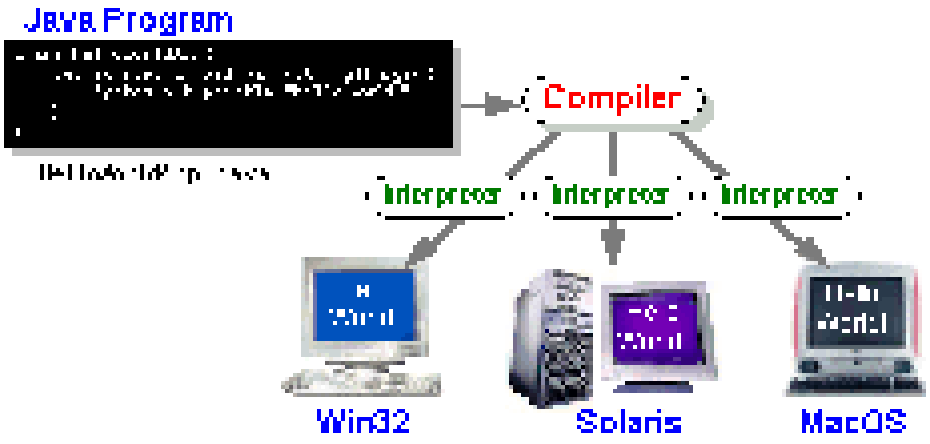


Fig 5.2: JAVA-Platform Independent

The Java Platform

A *platform* is the hardware or software environment in which a program runs. We've already mentioned some of the most popular platforms like Windows 2000, Linux, Solaris, and

MacOS. Most platforms can be described as a combination of the operating system and hardware. The Java platform differs from most other platforms in that it's a software-only platform that runs on top of other hardware-based platforms.

The Java platform has two components:

- The *Java Virtual Machine* (Java VM)
- The *Java Application Programming Interface* (Java API)

You've already been introduced to the Java VM. It's the base for the Java platform and is ported onto various hardware-based platforms.

The Java API is a large collection of ready-made software components that provide many useful capabilities, such as graphical user interface (GUI) widgets. The Java API is grouped into libraries of related classes and interfaces; these libraries are known as *packages*. The next section, What Can Java Technology Do? Highlights what functionality some of the packages in the Java API provide.

The following figure depicts a program that's running on the Java platform. As the figure shows, the Java API and the virtual machine insulate the program from the hardware.

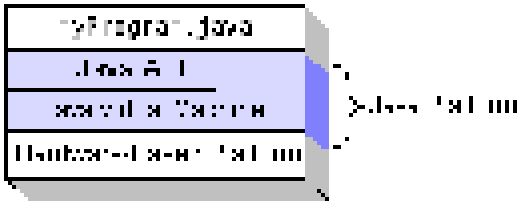


Fig 5.3: JAVA Platform

Native code is code that after you compile it, the compiled code runs on a specific hardware platform. As a platform-independent environment, the Java platform can be a bit slower than native code. However, smart compilers, well-tuned interpreters, and just-in-time byte code compilers can bring performance close to that of native code without threatening portability.

What Can Java Technology Do?

The most common types of programs written in the Java programming language are *applets* and *applications*. If you've surfed the Web, you're probably already familiar with applets. An applet is a program that adheres to certain conventions that allow it to run within a Java-enabled browser.

However, the Java programming language is not just for writing cute, entertaining applets for the Web. The general-purpose, high-level Java programming language is also a powerful software platform. Using the generous API, you can write many types of programs.

An application is a standalone program that runs directly on the Java platform. A special kind of application known as a *server* serves and supports clients on a network. Examples of servers are Web servers, proxy servers, mail servers, and print servers. Another specialized program is a *servlet*. A servlet can almost be thought of as an applet that runs on the server side. Java Servlets are a popular choice for building interactive web applications, replacing the use of CGI scripts. Servlets are similar to applets in that they are runtime extensions of applications. Instead of working in browsers, though, servlets run within Java Web servers, configuring or tailoring the server.

How does the API support all these kinds of programs? It does so with packages of software components that provides a wide range of functionality. Every full implementation of the Java platform gives you the following features:

- **The essentials**: Objects, strings, threads, numbers, input and output, data structures, system properties, date and time, and so on.
- **Applets**: The set of conventions used by applets.
- **Networking**: URLs, TCP (Transmission Control Protocol), UDP (User Data gram Protocol) sockets, and IP (Internet Protocol) addresses.
- **Internationalization**: Help for writing programs that can be localized for users worldwide. Programs can automatically adapt to specific locales and be displayed in the appropriate language.
- **Security**: Both low level and high level, including electronic signatures, public and private key management, access control, and certificates.
- **Software components**: Known as JavaBeans™, can plug into existing component architectures.
- **Object serialization**: Allows lightweight persistence and communication via Remote Method Invocation (RMI).
- **Java Database Connectivity (JDBC™)**: Provides uniform access to a wide range of relational databases.

The Java platform also has APIs for 2D and 3D graphics, accessibility, servers, collaboration, telephony, speech, animation, and more. The following figure depicts what is included in the Java 2 SDK.

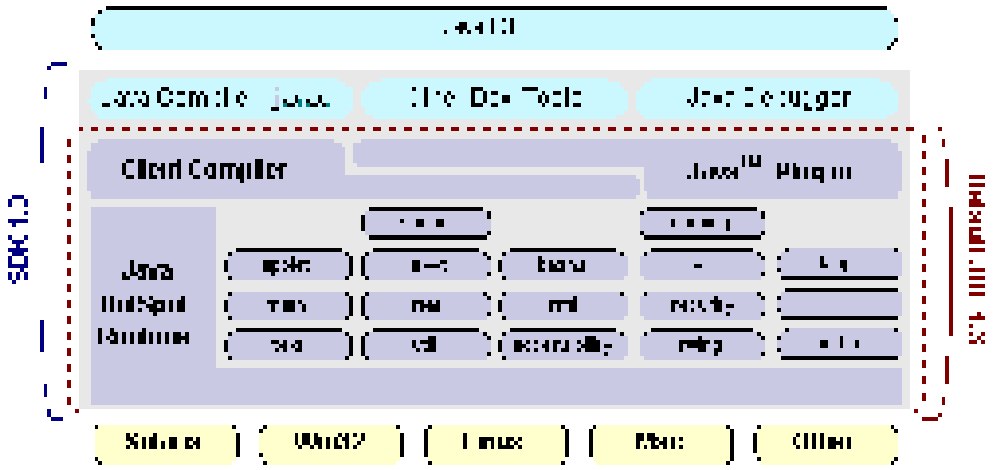


Fig 5.4: JAVA Software Development Kit (SDK)

How Will Java Technology Change My Life?

We can't promise you fame, fortune, or even a job if you learn the Java programming language. Still, it is likely to make your programs better and requires less effort than other languages. We believe that Java technology will help you do the following:

- **Get started quickly**: Although the Java programming language is a powerful object-oriented language, it's easy to learn, especially for programmers already familiar with C or C++.
- **Write less code**: Comparisons of program metrics (class counts, method counts, and so on) suggest that a program written in the Java programming language can be four times smaller than the same program in C++.
- **Write better code**: The Java programming language encourages good coding practices, and its garbage collection helps you avoid memory leaks. Its object orientation, its JavaBeans component architecture, and its wide-ranging, easily extendible API let you reuse other people's tested code and introduce fewer bugs.

- **Develop programs more quickly**: Your development time may be as much as twice as fast versus writing the same program in C++. Why? You write fewer lines of code and it is a simpler programming language than C++.
- **Avoid platform dependencies with 100% Pure Java**: You can keep your program portable by avoiding the use of libraries written in other languages. The 100% Pure Java™ Product Certification Program has a repository of historical process manuals, white papers, brochures, and similar materials online.
- **Write once, run anywhere**: Because 100% Pure Java programs are compiled into machine-independent byte codes, they run consistently on any Java platform.
- **Distribute software more easily**: You can upgrade applets easily from a central server. Applets take advantage of the feature of allowing new classes to be loaded "on the fly," without recompiling the entire program.

ODBC

Microsoft Open Database Connectivity (ODBC) is a standard programming interface for application developers and database systems providers. Before ODBC became a *de facto* standard for Windows programs to interface with database systems, programmers had to use proprietary languages for each database they wanted to connect to. Now, ODBC has made the choice of the database system almost irrelevant from a coding perspective, which is as it should be. Application developers have much more important things to worry about than the syntax that is needed to port their program from one database to another when business needs suddenly change.

Through the ODBC Administrator in Control Panel, you can specify the particular database that is associated with a data source that an ODBC application program is written to use. Think of an ODBC data source as a door with a name on it. Each door will lead you to a particular database. For example, the data source named Sales Figures might be a SQL Server database, whereas the Accounts Payable data source could refer to an Access database. The physical database referred to by a data source can reside anywhere on the LAN.

The ODBC system files are not installed on your system by Windows 95. Rather, they are installed when you setup a separate database application, such as SQL Server Client or Visual Basic 4.0. When the ODBC icon is installed in Control Panel, it uses a file called ODBCINST.DLL. It is also possible to administer your ODBC data sources through a stand-alone

program called ODBCADM.EXE. There is a 16-bit and a 32-bit version of this program and each maintains a separate list of ODBC data sources.

From a programming perspective, the beauty of ODBC is that the application can be written to use the same set of function calls to interface with any data source, regardless of the database vendor. The source code of the application doesn't change whether it talks to Oracle or SQL Server. We only mention these two as an example. There are ODBC drivers available for several dozen popular database systems. Even Excel spreadsheets and plain text files can be turned into data sources. The operating system uses the Registry information written by ODBC Administrator to determine which low-level ODBC drivers are needed to talk to the data source (such as the interface to Oracle or SQL Server). The loading of the ODBC drivers is transparent to the ODBC application program. In a client/server environment, the ODBC API even handles many of the network issues for the application programmer.

The advantages of this scheme are so numerous that you are probably thinking there must be some catch. The only disadvantage of ODBC is that it isn't as efficient as talking directly to the native database interface. ODBC has had many detractors make the charge that it is too slow. Microsoft has always claimed that the critical factor in performance is the quality of the driver software that is used. In our humble opinion, this is true. The availability of good ODBC drivers has improved a great deal recently. And anyway, the criticism about performance is somewhat analogous to those who said that compilers would never match the speed of pure assembly language. Maybe not, but the compiler (or ODBC) gives you the opportunity to write cleaner programs, which means you finish sooner. Meanwhile, computers get faster every year.

JDBC

In an effort to set an independent database standard API for Java; Sun Microsystems developed Java Database Connectivity, or JDBC. JDBC offers a generic SQL database access mechanism that provides a consistent interface to a variety of RDBMSs. This consistent interface is achieved through the use of "plug-in" database connectivity modules, or *drivers*. If a database vendor wishes to have JDBC support, he or she must provide the driver for each platform that the database and Java run on.

To gain a wider acceptance of JDBC, Sun based JDBC's framework on ODBC. As you discovered earlier in this chapter, ODBC has widespread support on a variety of platforms. Basing JDBC on ODBC will allow vendors to bring JDBC drivers to market much faster than developing a completely new connectivity solution.

JDBC was announced in March of 1996. It was released for a 90 day public review that ended June 8, 1996. Because of user input, the final JDBC v1.0 specification was released soon after.

The remainder of this section will cover enough information about JDBC for you to know what it is about and how to use it effectively. This is by no means a complete overview of JDBC. That would fill an entire book.

JDBC Goals

Few software packages are designed without goals in mind. JDBC is one that, because of its many goals, drove the development of the API. These goals, in conjunction with early reviewer feedback, have finalized the JDBC class library into a solid framework for building database applications in Java.

The goals that were set for JDBC are important. They will give you some insight as to why certain classes and functionalities behave the way they do. The eight design goals for JDBC are as follows:

1. ***SQL Level API***

   The designers felt that their main goal was to define a SQL interface for Java. Although not the lowest database interface level possible, it is at a low enough level for higher-level tools and APIs to be created. Conversely, it is at a high enough level for application programmers to use it confidently. Attaining this goal allows for future tool vendors to "generate" JDBC code and to hide many of JDBC's complexities from the end user.

2. ***SQL Conformance***

   SQL syntax varies as you move from database vendor to database vendor. In an effort to support a wide variety of vendors, JDBC will allow any query statement to be passed through it to the underlying database driver. This allows the connectivity module to handle non-standard functionality in a manner that is suitable for its users.

3. ***JDBC must be implemental on top of common database interfaces***
   The JDBC SQL API must "sit" on top of other common SQL level APIs. This goal allows JDBC to use existing ODBC level drivers by the use of a software interface. This interface would translate JDBC calls to ODBC and vice versa.

4. ***Provide a Java interface that is consistent with the rest of the Java system***
   Because of Java's acceptance in the user community thus far, the designers feel that they should not stray from the current design of the core Java system.

5. ***Keep it simple***
   This goal probably appears in all software design goal listings. JDBC is no exception. Sun felt that the design of JDBC should be very simple, allowing for only one method of completing a task per mechanism. Allowing duplicate functionality only serves to confuse the users of the API.

6. ***Use strong, static typing wherever possible***
   Strong typing allows for more error checking to be done at compile time; also, less error appear at runtime.

7. ***Keep the common cases simple***
   Because more often than not, the usual SQL calls used by the programmer are simple SELECT's, INSERT's, DELETE's and UPDATE's, these queries should be simple to perform with JDBC. However, more complex SQL statements should also be possible.

Finally we decided to proceed the implementation using Java Networking. And for dynamically updating the cache table we go for MS Access database.

Java has two things: a programming language and a platform. Java is a high-level programming language that is all of the following:

- Simple
- Architecture-neutral
- Object-oriented
- Portable

- Distributed
- High-performance
- Interpreted
- Multithreaded
- Robust
- Dynamic
- Secure

Java is also unusual in that each Java program is both compiled and interpreted. With a compile you translate a Java program into an intermediate language called Java byte codes the platform-independent code instruction is passed and run on the computer.

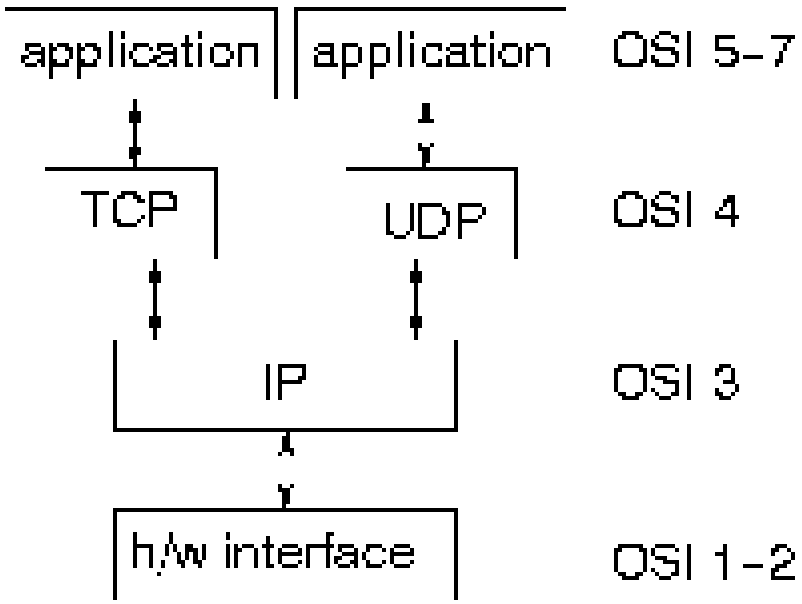
Compilation happens just once; interpretation occurs each time the program is executed. The figure illustrates how this works.

You can think of Java byte codes as the machine code instructions for the Java Virtual Machine (Java VM). Every Java interpreter, whether it's a Java development tool or a Web browser that can run Java applets, is an implementation of the Java VM. The Java VM can also be implemented in hardware.

Java byte codes help make "write once, run anywhere" possible. You can compile your Java program into byte codes on my platform that has a Java compiler. The byte codes can then be run any implementation of the Java VM. For example, the same Java program can run Windows NT, Solaris, and Macintosh.

Networking

TCP/IP stack: The TCP/IP stack is shorter than the OSI one:



TCP is a connection-oriented protocol; UDP (User Datagram Protocol) is a connectionless protocol.

IP datagram's:

The IP layer provides a connectionless and unreliable delivery system. It considers each datagram independently of the others. Any association between datagram must be supplied by the higher layers. The IP layer supplies a checksum that includes its own header. The header includes the source and destination addresses. The IP layer handles routing through an Internet. It is also responsible for breaking up large datagram into smaller ones for transmission and reassembling them at the other end.

UDP:

UDP is also connectionless and unreliable. What it adds to IP is a checksum for the contents of the datagram and port numbers. These are used to give a client/server model - see later.

TCP:

TCP supplies logic to give a reliable connection-oriented protocol above IP. It provides a virtual circuit that two processes can use to communicate.

Internet addresses:

In order to use a service, you must be able to find it. The Internet uses an address scheme for machines so that they can be located. The address is a 32 bit integer which gives the IP address. This encodes a network ID and more addressing. The network ID falls into various classes according to the size of the network address.

Network address:

Class A uses 8 bits for the network address with 24 bits left over for other addressing. Class B uses 16 bit network addressing. Class C uses 24 bit network addressing and class D uses all 32.

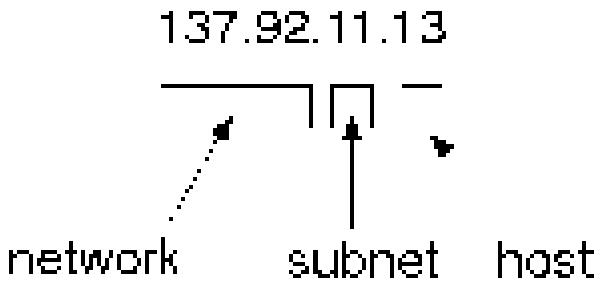
Subnet address:

Internally, the UNIX network is divided into sub networks. Building 11 is currently on one sub network and uses 10-bit addressing, allowing 1024 different hosts.

Host address:

8 bits are finally used for host addresses within our subnet. This places a limit of 256 machines that can be on the subnet.

Total address:



The 32 bit address is usually written as 4 integers separated by dots.

Port addresses:

A service exists on a host, and is identified by its port. This is a 16 bit number. To send a message to a server, you send it to the port for that service of the host that it is running on. This is not location transparency! Certain of these ports are "well known".

Sockets:

A socket is a data structure maintained by the system to handle network connections. A socket is created using the call socket. It returns an integer that is like a file descriptor. In fact, under Windows, this handle can be used with Read File and Write File functions.

```
#include <sys/types.h>
#include <sys/socket.h>
int socket(int family, int type, int protocol);
```

Here "family" will be AF_INET for IP communications, protocol will be zero, and type will depend on whether TCP or UDP is used. Two processes wishing to communicate over a network create a socket each. These are similar to two ends of a pipe - but the actual pipe does not yet exist.

JFREE CHART

JFreeChart is a free 100% Java chart library that makes it easy for developers to display professional quality charts in their applications. JFreeChart's extensive feature set includes:

- A consistent and well-documented API, supporting a wide range of chart types;

- A flexible design that is easy to extend, and targets both server-side and client-side applications;

Support for many output types, including Swing components, image files (including PNG and JPEG), and vector graphics file formats (including PDF, EPS and SVG);

JFreeChart is "open source" or, more specifically, free software. It is distributed under the terms of the GNU Lesser General Public Licence (LGPL), which permits use in proprietary applications.

1. Map Visualizations

Charts showing values that relate to geographical areas. Some examples include: (a) population density in each state of the United States, (b) income per capita for each country in Europe, (c) life expectancy in each country of the world. The tasks in this project include:

Sourcing freely redistributable vector outlines for the countries of the world, states/provinces in particular countries (USA in particular, but also other areas);

Creating an appropriate dataset interface (plus default implementation), a rendered, and integrating this with the existing XYPlot class in JFreeChart;

Testing, documenting, testing some more, documenting some more.

2. Time Series Chart Interactivity

Implement a new (to JFreeChart) feature for interactive time series charts --- to display a separate control that shows a small version of ALL the time series data, with a sliding "view" rectangle that allows you to select the subset of the time series data to display in the main chart.

3. Dashboards

There is currently a lot of interest in dashboard displays. Create a flexible dashboard mechanism that supports a subset of JFreeChart chart types (dials, pies, thermometers, bars, and lines/time series) that can be delivered easily via both Java Web Start and an applet.

4. Property Editors

The property editor mechanism in JFreeChart only handles a small subset of the properties that can be set for charts. Extend (or reimplement) this mechanism to provide greater end-user control over the appearance of the charts.

J2ME (Java 2 Micro edition):-

Sun Microsystems defines J2ME as "a highly optimized Java run-time environment targeting a wide range of consumer products, including pagers, cellular phones, screen-phones, digital set-top boxes and car navigation systems." Announced in June 1999 at the JavaOne Developer Conference, J2ME brings the cross-platform functionality of the Java language to smaller devices, allowing mobile wireless devices to share applications. With J2ME, Sun has adapted the Java platform for consumer products that incorporate or are based on small computing devices.

1. General J2ME architecture

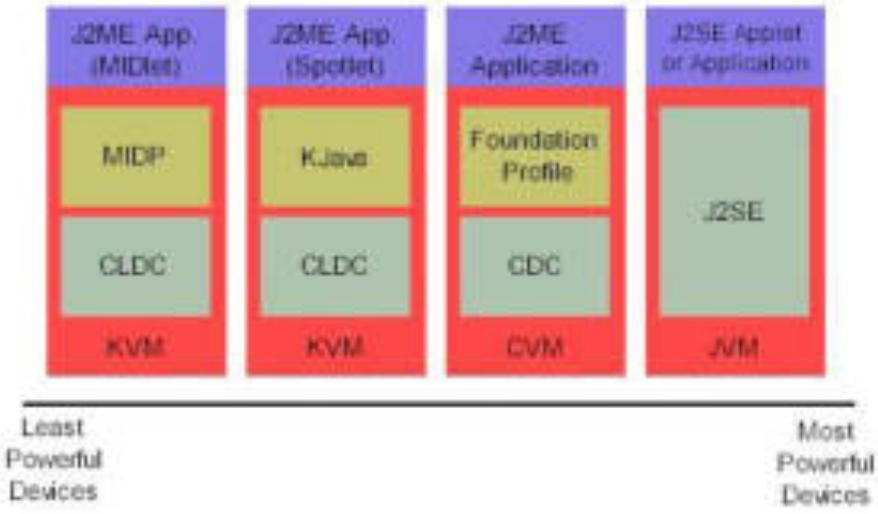


Fig 5.5: General J2ME architecture

J2ME uses configurations and profiles to customize the Java Runtime Environment (JRE). As a complete JRE, J2ME is comprised of a configuration, which determines the JVM used, and a profile, which defines the application by adding domain-specific classes. The configuration defines the basic run-time environment as a set of core classes and a specific JVM that run on specific types of devices. We'll discuss configurations in detail in the The profile defines the application; specifically, it adds domain-specific classes to the J2ME configuration to define certain uses for devices. We'll cover profiles in depth in the The following graphic depicts the relationship between the different virtual machines, configurations, and profiles. It also draws a parallel with the J2SE API and its Java virtual machine. While the J2SE virtual machine is generally referred to

as a JVM, the J2ME virtual machines, KVM and CVM, are subsets of JVM. Both KVM and CVM can be thought of as a kind of Java virtual machine -- it's just that they are shrunken versions of the J2SE JVM and are specific to J2ME.

2. Developing J2ME applications

Introduction In this section, we will go over some considerations you need to keep in mind when developing applications for smaller devices. We'll take a look at the way the compiler is invoked when using J2SE to compile J2ME applications. Finally, we'll explore packaging and deployment and the role preverification plays in this process.

3. Design considerations for small devices

Developing applications for small devices requires you to keep certain strategies in mind during the design phase. It is best to strategically design an application for a small device before you begin coding. Correcting the code because you failed to consider all of the "gotchas" before developing the application can be a painful process. Here are some design strategies to consider:

* Keep it simple. Remove unnecessary features, possibly making those features a separate, secondary application.

* Smaller is better. This consideration should be a "no brainer" for all developers. Smaller applications use less memory on the device and require shorter installation times. Consider packaging your Java applications as compressed Java Archive (jar) files.

* Minimize run-time memory use. To minimize the amount of memory used at run time, use scalar types in place of object types. Also, do not depend on the garbage collector. You should manage the memory efficiently yourself by setting object references to null when you are finished with them. Another way to reduce run-time memory is to use lazy instantiation, only allocating objects on an as-needed basis. Other ways of reducing overall and peak memory use on small devices are to release resources quickly, reuse objects, and avoid exceptions.

4. Configurations overview

The configuration defines the basic run-time environment as a set of core classes and a specific JVM that run on specific types of devices. Currently, two configurations exist for J2ME, though others may be defined in the future:

* **Connected Limited Device Configuration (CLDC)** is used specifically with the KVM for 16-bit or 32-bit devices with limited amounts of memory. This is the configuration (and the virtual machine) used for developing small J2ME applications. Its size limitations make CLDC more interesting and challenging (from a development point of view) than CDC. CLDC is also the configuration that we will use for developing our drawing tool application. An example of a small wireless device running small applications is a Palm hand-held computer.

* **Connected Device Configuration (CDC)** is used with the C virtual machine (CVM) and is used for 32-bit architectures requiring more than 2 MB of memory. An example of such a device is a Net TV box.

5. J2ME profiles

What is a J2ME profile?

As we mentioned earlier in this tutorial, a profile defines the type of device supported. The Mobile Information Device Profile (MIDP), for example, defines classes for cellular phones. It adds domain-specific classes to the J2ME configuration to define uses for similar devices. Two profiles have been defined for J2ME and are built upon CLDC: KJava and MIDP. Both KJava and MIDP are associated with CLDC and smaller devices. Profiles are built on top of configurations. Because profiles are specific to the size of the device (amount of memory) on which an application runs, certain profiles are associated with certain configurations.

A skeleton profile upon which you can create your own profile, the Foundation Profile, is available for CDC.

Profile 1: KJava

KJava is Sun's proprietary profile and contains the KJava API. The KJava profile is built on top of the CLDC configuration. The KJava virtual machine, KVM, accepts the same byte codes and class file format as the classic J2SE virtual machine. KJava contains a Sun-specific API that runs on the Palm OS. The KJava API has a great deal in common with the J2SE Abstract Windowing Toolkit (AWT). However, because it is not a standard J2ME package, its main package is com.sun.kjava. We'll learn more about the KJava API later in this tutorial when we develop some sample applications.

Profile 2: MIDP

MIDP is geared toward mobile devices such as cellular phones and pagers. The MIDP, like KJava, is built upon CLDC and provides a standard run-time environment that allows new applications and services to be deployed dynamically on end user devices. MIDP is a common, industry-standard profile for mobile devices that is not dependent on a specific vendor. It is a complete and supported foundation for mobile application development. MIDP contains the following packages, the first three of which are core CLDC packages, plus three MIDP-specific packages.

* java.lang

* java.io

* java.util

* javax.microedition.io

* javax.microedition.lcdui

* javax.microedition.midlet

* javax.microedition.rms

5.2) SOURCE CODE

DATABASE CONNECTION:

```
package com.mysql;
import java.sql.*;

public class DatabaseCon
{
	static Connection co;
	public static Connection getConnection()
	{


		try
		{
			Class.forName("com.mysql.jdbc.Driver");
			co =
DriverManager.getConnection("jdbc:mysql://localhost:3306/TBPA","root","");
		}
		catch(Exception e)
		{
			System.out.println("Database Error"+e);
		}
		return co;
	}

}
```

USER SIGNUP:

```
<%
String title="";
%>

<%@ include file="header.jsp"%>

<%String message=request.getParameter("id");

if(message!=null )
{
out.println("<font color='Green'><h2>Registration Successful</h2></font>");
}
%>
<body class="login-bg">
```

```
		<div class="login-body">
			<div class="login-heading">
				<h1>Sign up</h1>
			</div>
			<div class="login-info">
			<form method="post" action="signup2.jsp">
					<input type="text" class="user" name="name"
placeholder="Name" required="">
					<input type="text" class="user" name="email"
placeholder="Email" required="">
					<input type="text" class="user" name="ph"
placeholder="Phone Number" required="">
					<input type="password" name="pwd" class="lock"
placeholder="Password">

					<input type="submit" name="Sign In" value="Sign up">
					<div class="signup-text">
						<a href="user.jsp">Already have an account? Login
here.</a>
					</div>
				</form>
			</div>
		</div>
		<div class="go-back login-go-back">
				<a href="index.jsp">Go To Home</a>
			</div>


		<%@ include file="footer.jsp"%>
```

USER SIGN-IN:

```
<%
String title="";
%>

<%@ include file="header.jsp"%>
                                                <%String
message=request.getParameter("id");
                                                if(message!=null )
                                                {
                                                        out.println("<font
color='Green'><h2>Registration Successful</h2></font>");
                                                }
                                        %>
<body class="login-bg">
```

```
		<div class="login-body">
			<div class="login-heading">
				<h1>Login</h1>
			</div>			<div class="login-info">
			<form method="post" action="ulogin.jsp">
					<input type="text" class="user" name="email"
placeholder="Email" required="">
					<input type="password" name="pwd" class="lock"
placeholder="Password">

					<input type="submit" name="Login" value="Login">
					<div class="signup-text">
						<a href="signup.jsp">You have 'nt account?
Register here.</a>
					</div>
				</form>
			</div>
		</div>
		<div class="go-back login-go-back">
				<a href="index.jsp">Go To Home</a>
			</div>


		<%@ include file="footer.jsp"%>
```

VIEW FRIEND SUGGESTIONS:

```
<%String title="View Friend Suggestions";%>

<%@ include file="uheader.jsp"%>
<%@ page  import="java.sql.*" import="com.mysql.*" import="CT.*" %>

<div>

<% //String mail=request.getParameter("email");
%>
<%
try{
Connection con = DatabaseCon.getConnection();
Statement st=con.createStatement();
ArrayList<String>al=Details.getSuggestFriendsList(session.getAttribute("email").toString());

String sql="select * from users where  email!='"+session.getAttribute("email")+"' ";

System.out.println(sql);
```

```
ResultSet rs=st.executeQuery(sql);


%>
<form method="post" action="search2.jsp">
<table align="center" >

<tr><td>
<tr><td> <h2>Results are</h2>
<%


while(rs.next())
{

        String em=rs.getString("email");
        int count=0;
         for(String name:al)
         {
                if(em.equals(name))
                {
                        count++;
                }

        }

        if(count==0)
        {


%>

<tr><td colspan=2> <img src="view1.jsp?id=<%=rs.getString("email")%>" alt=""
style="width:258px;height:251px;" >
<tr><td colspan=2> <br><h3>Email:           <%=rs.getString("email")%>
<input type="hidden" name="mail" value="<%=rs.getString("email")%>"/>
<tr><td colspan=2><input  class="form-control1"  type="submit" value="Send Friend Reqest"/>

</form><td>

<%
        }
}
}
catch(Exception e){
}
```

```
%>
</table>


</div>

<%@ include file="ufooter.jsp"%>
```

6) SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

6.1) TYPES OF TESTS

Unit testing:

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Integration testing:

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Functional testing:

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centred on the following items:

- Valid Input : identified classes of valid input must be accepted.
- Invalid Input : identified classes of invalid input must be rejected.
- Functions : identified functions must be exercised.
- Output : identified classes of application outputs must be exercised.
- Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

System Testing:

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

White Box Testing:

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

Black Box Testing:

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document,

such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

6.1.1) Unit Testing:

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

6.1.2) Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

6.1.3) Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

7) RESULTS

SCREENSHOTS:

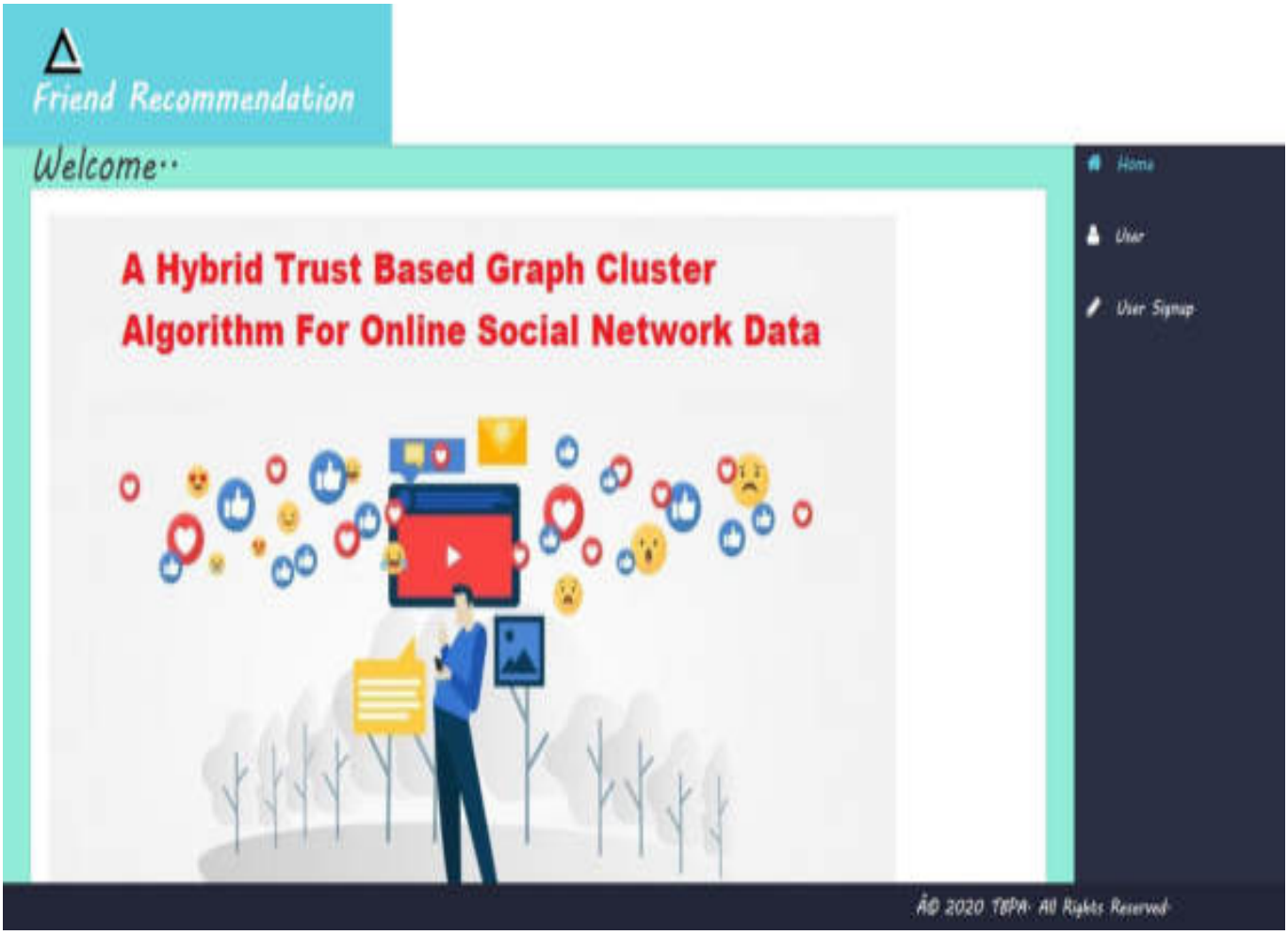


Fig 7.1: Welcome page

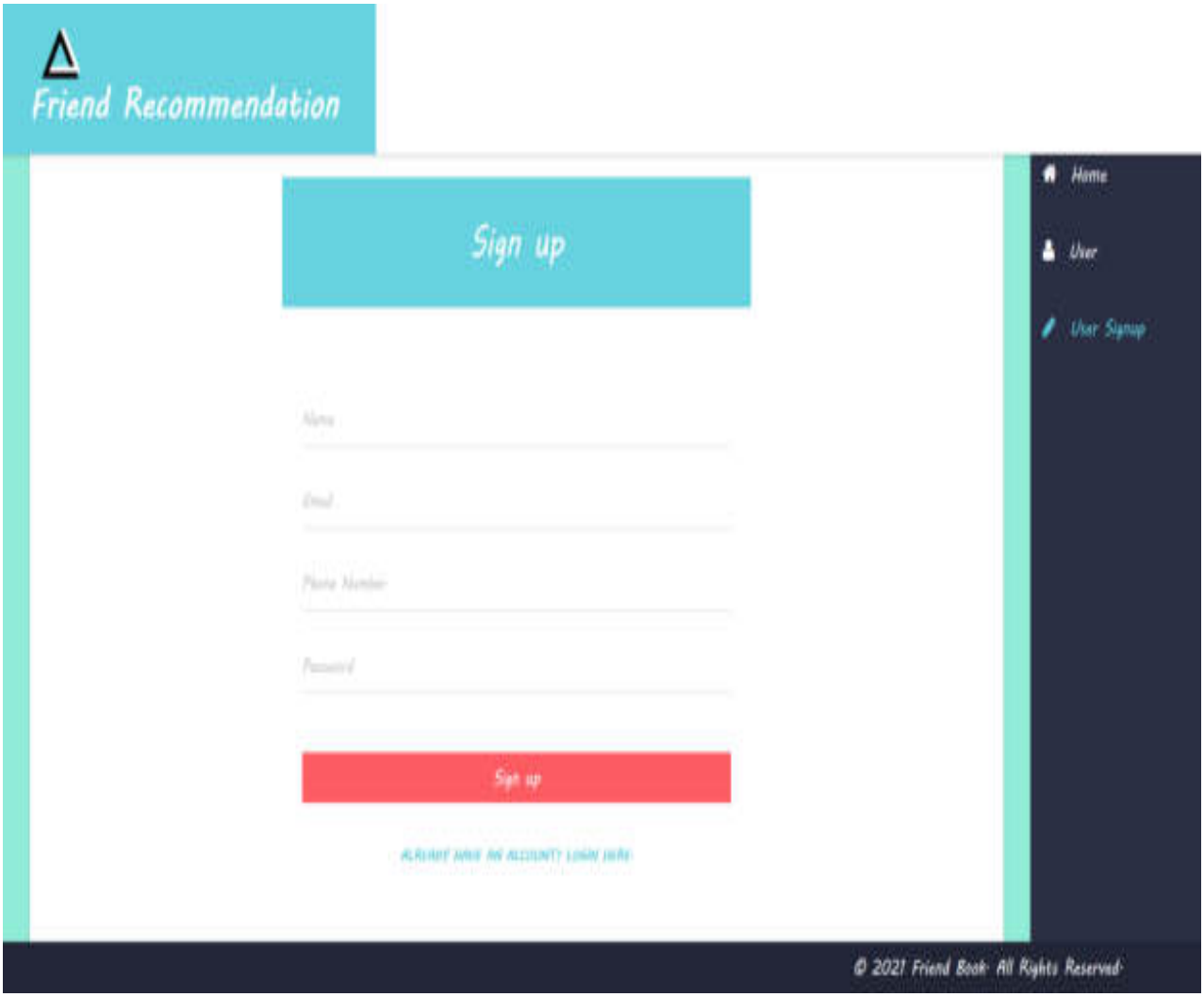


Fig 7.2: Sign-up page

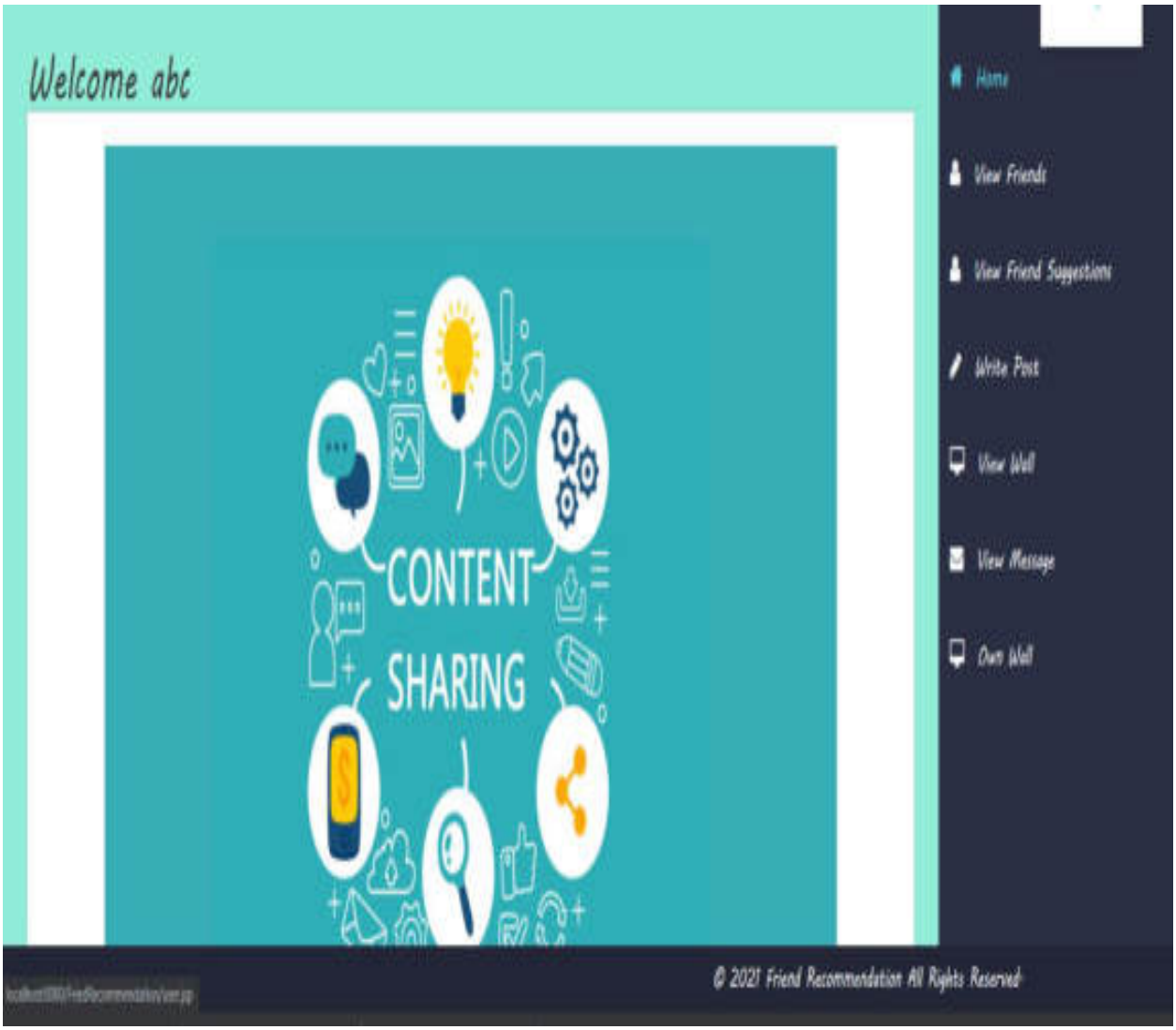


Fig 7.3: Home Page

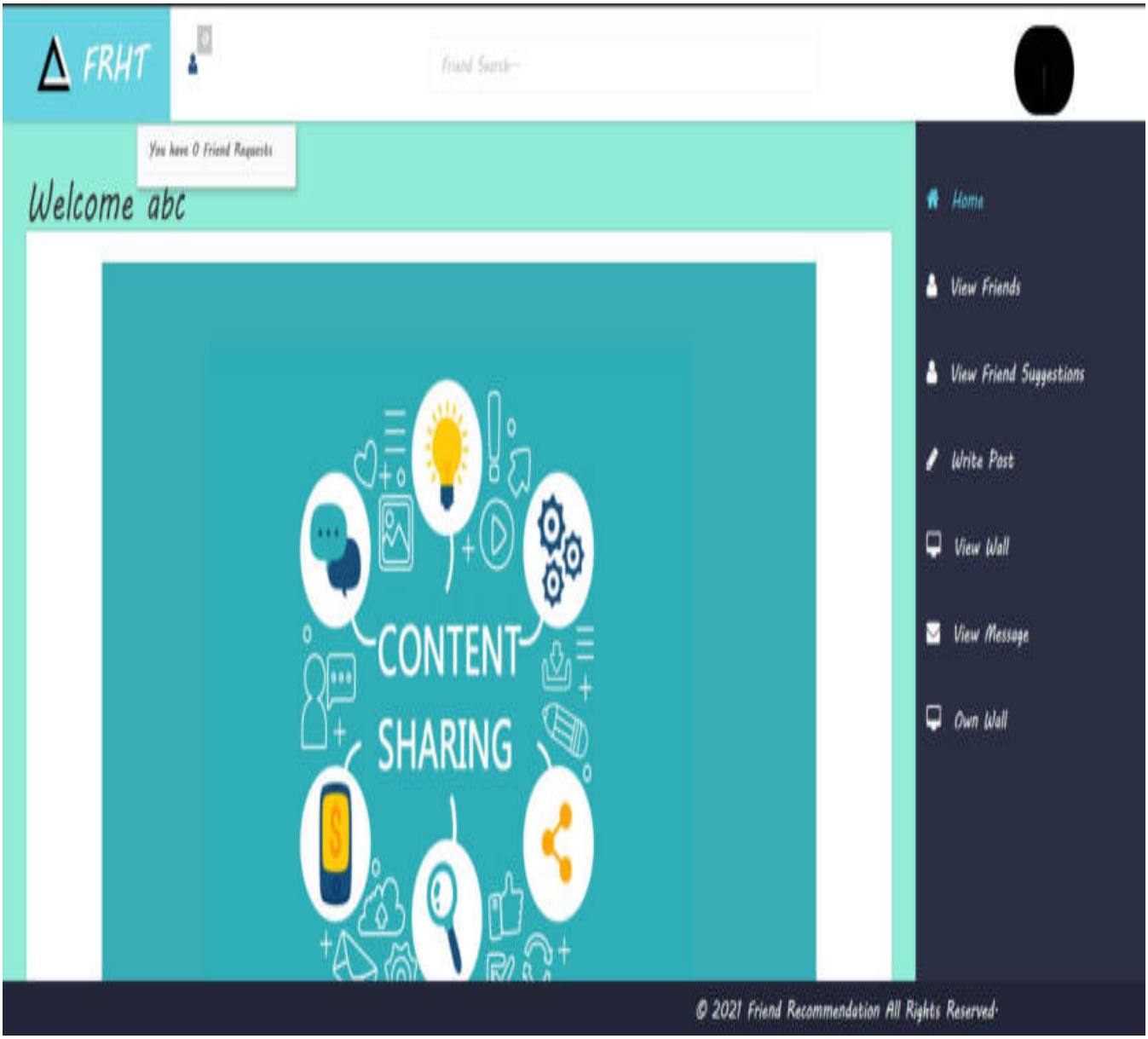


Fig 7.4: Friend Request List

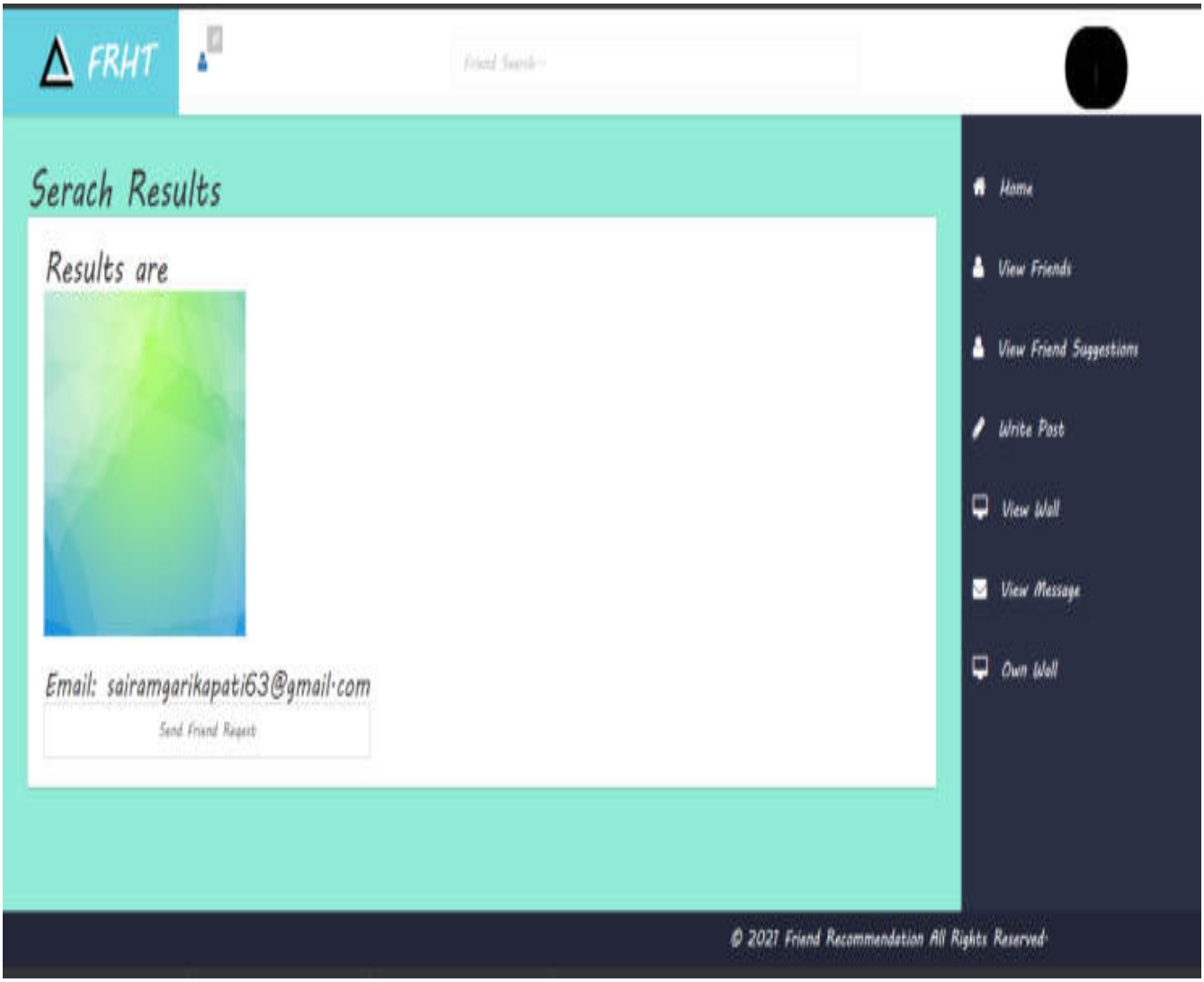


Fig 7.5: Search Results

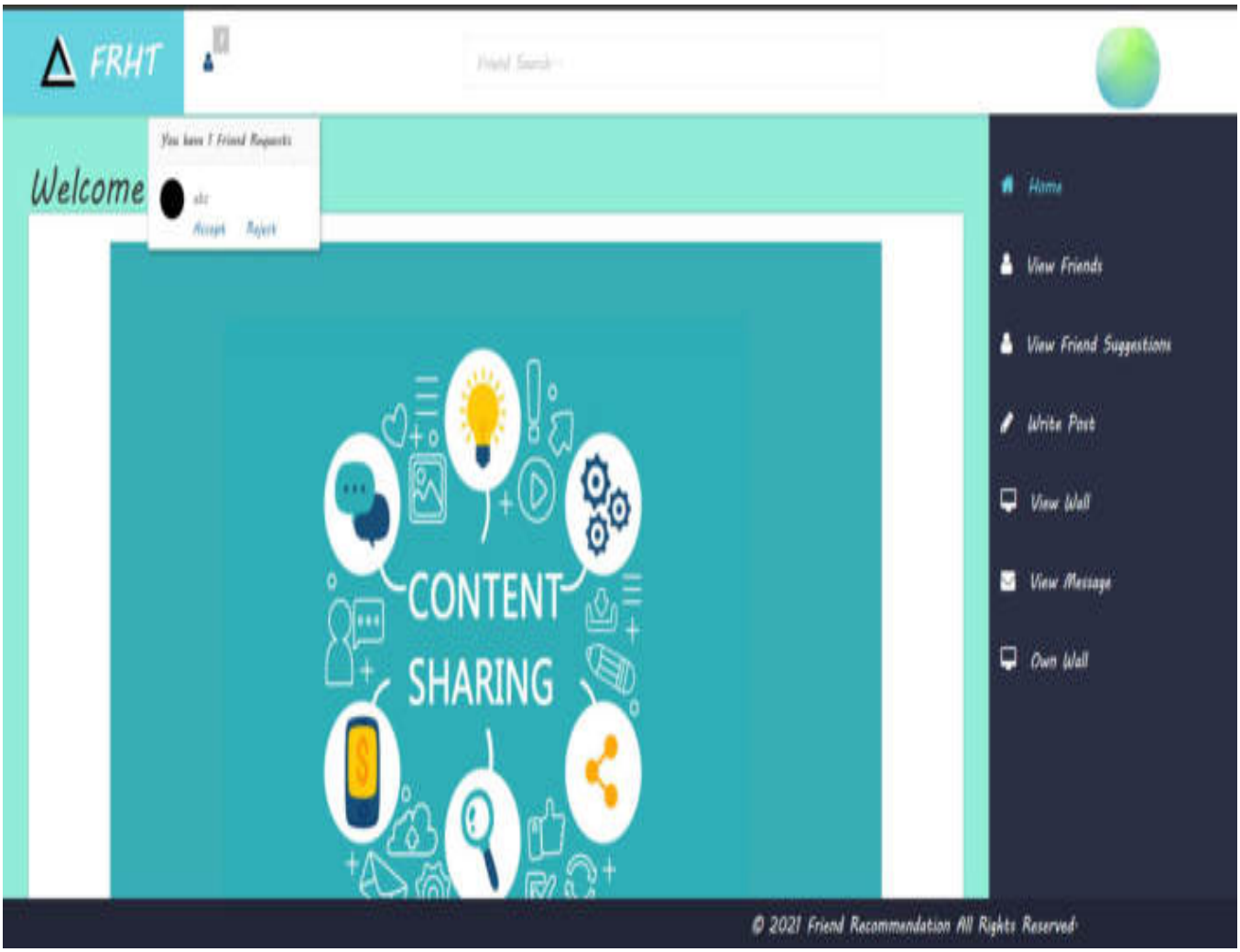


Fig 7.6: Request Status

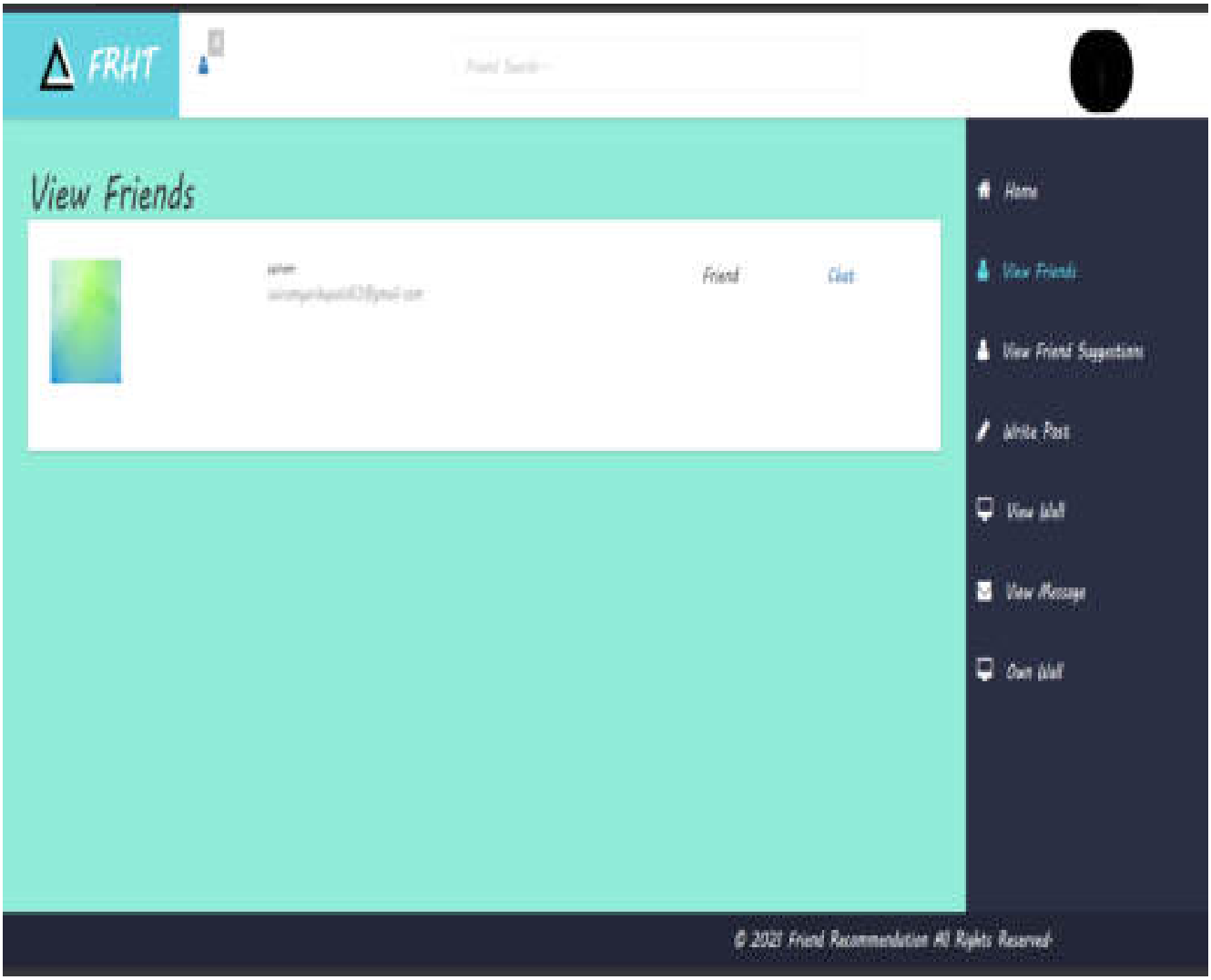


Fig 7.7: View Friends List

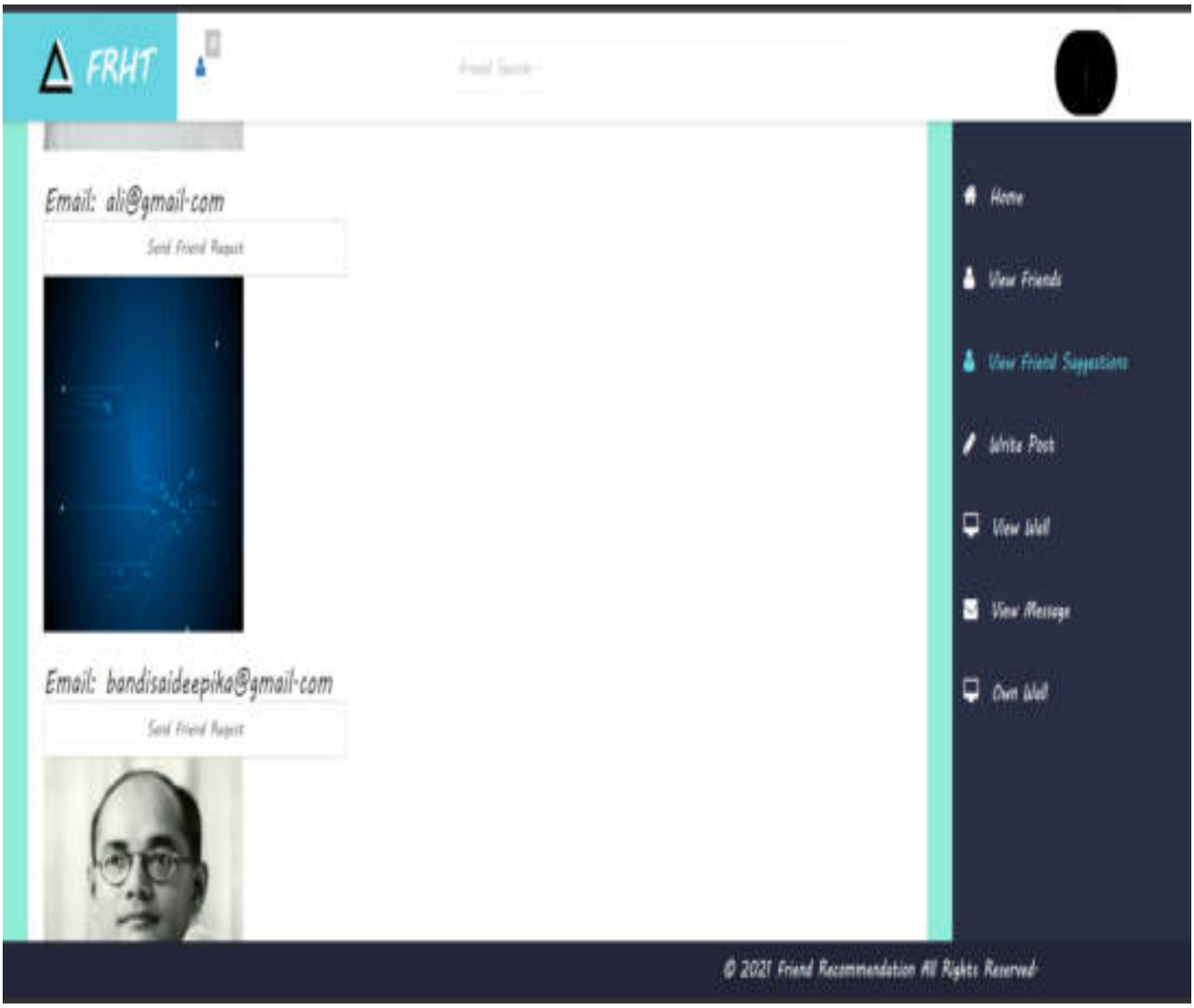


Fig 7.8: Friend Suggestions List

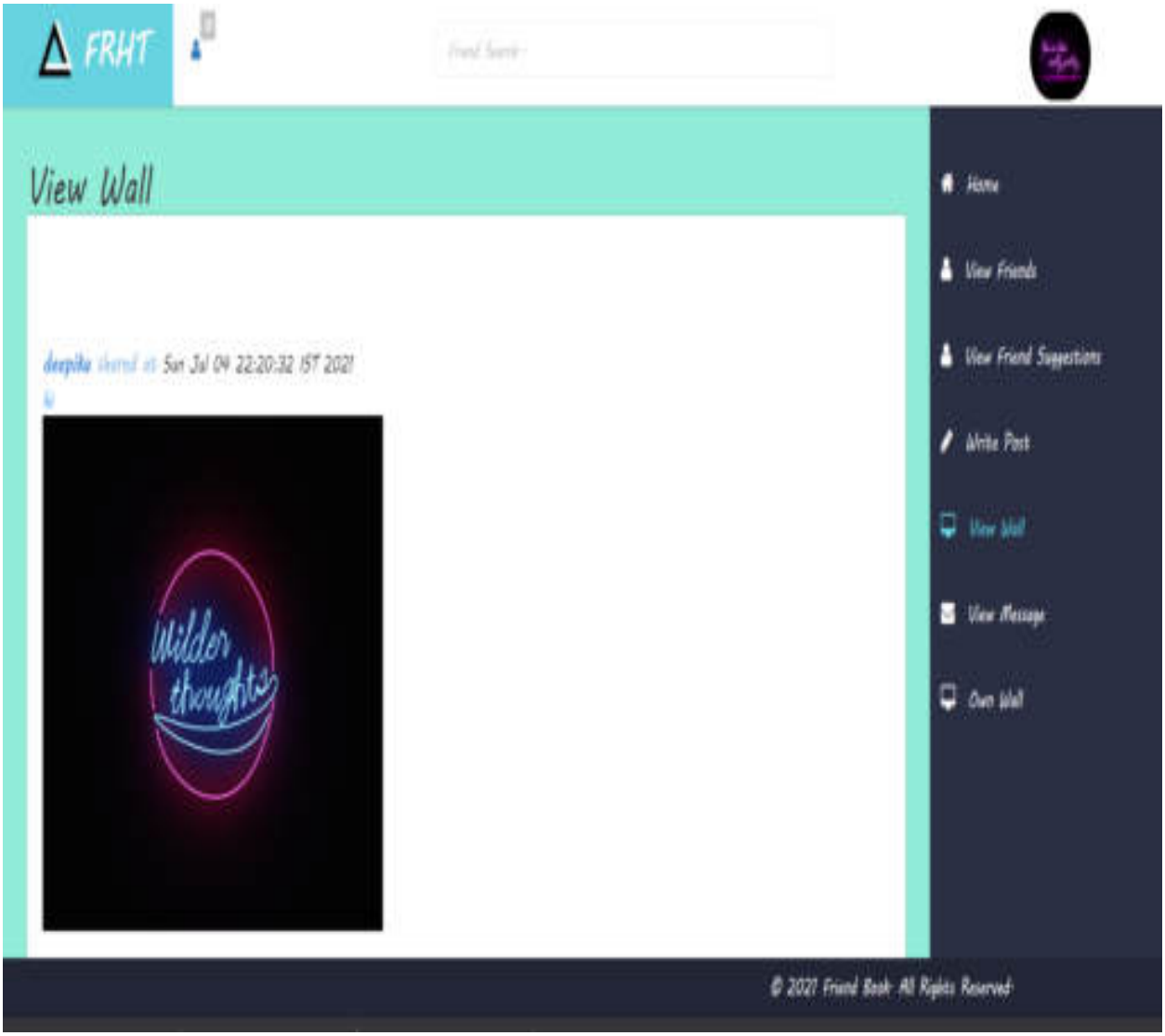


Fig 7.9: Friend's Posts Wall

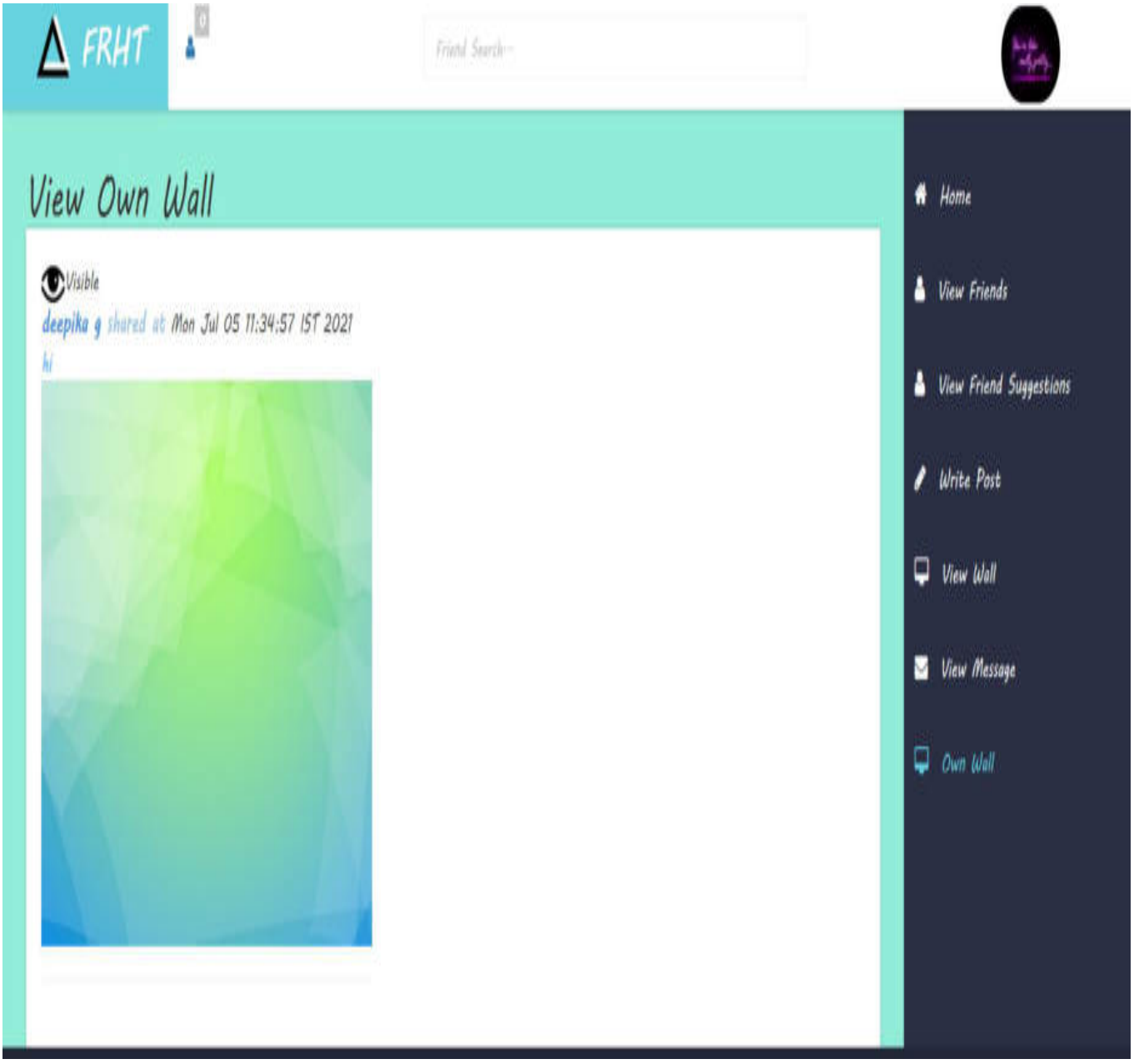


Fig 7.10: Own Post Wall

Fig 7.11: Logout

8) CONCLUSION

OSN (Online Social Networking) is an online platform which people use to build social networks or social relationships with other people who share similar personal or career interests, activities, backgrounds or real-life connections. We are going to get recommendations based on the concept of graph clustering algorithm. Finally, we summarize the features of OSNs and the properties of recommendation systems. The working of the proposed system is very good. The proposed system has good accuracy. It gives better recommendations. On experimenting, it is observed that it gives better results.

9) REFERENCES

[1] Mohamed Shehab et al., Recommendation Models for Open Authorization, IEEE Transactions on dependable and secure computing, Vol. 9, No. 4, July/August 2012.

[2] Robert Koch et al., Data Control in Social Networks, 978-1-4577-0460-4/11/$26.00 ©2011 IEEE.

[3] Gorrell P. Cheek et al., Human Effects of Enhanced Privacy Management Models, IEEE Transactions on dependable and secure computing, Vol. 11, No. 2, March/April 2014.

[4] Nidhi Grover et al., Digital Image Authentication Model Based on Edge Adaptive Steganography, 2013 Second International Conference on Advanced Computing, Networking and Security.

[5] Nissenbaum, H. Privacy as Contextual Integrity. Wash. L. Rev. 2004, 79, 101–139.

[6] Davison, H.K.; Maraist, C.C.; Hamilton, R.; Bing, M.N. To Screen or Not to Screen? Using the Internet for Selection Decisions. Empl. Responsib. Rights J. 2012, 24, 1–21. [CrossRef]

[7] Taddicken, M. The 'Privacy Paradox' in the Social Web: The Impact of Privacy Concerns, Individual Characteristics, and the Perceived Social Relevance on Different Forms of Self-Disclosure. J. Comput.-Mediat. Commun. 2014, 19, 248–273.

[8] Marwick, A.E.; Boyd, D. Networked privacy: How teenagers negotiate context in social media. New Media Soci. 2014, 16, 1051–1067. [CrossRef]

[9] Ashtari, S. I Know Who You Are and I Saw What You Did: Social Networks and the Death of Privacy. J. Inf. Priv. Secur. 2013, 9, 80–82. [CrossRef]

[10] Fire, M.; Goldschmidt, R.; Elovici, Y. Online social networks: Threats and solutions. IEEE Commun. Surv. Tutor. 2014, 16, 2019–2036. [CrossRef]

[11] Heymann, P.; Koutrika, G.; Garcia-Molina, H. Fighting spam on social web sites: A survey of approaches and future challenges. IEEE Internet Comput. 2007, 11, 36–45. [CrossRef]

[12] Everett, C. Social media: Opportunity or risk? Comput. Fraud Secur. 2010, 2010, 8–10. [CrossRef]

[13] Alarm, S.; El-Khatib, K. Phishing Susceptibility Detection through Social Media Analytics. In Proceedings of the 9th International Conference on Security of Information and Networks, Newark, NJ, USA, 20–22 July 2016; pp. 61–64.

[14] Nithya, V.; Pandian, S.L.; Malarvizhi, C. A survey on detection and prevention of cross-site scripting attack. Int. J. Secur. Appl. 2015, 9, 139–152. [CrossRef]

[15] Baltazar, J.; Costoya, J.; Flores, R. The Real Face of Koobface: The Largest Web 2.0 Botnet Explained. Trend Micro Threat Research. 2009. Available online: https://www.trendmicro.de/cloud-content/us/pdfs/ security-intelligence/white-papers/wp_the-real-face-of-koobface.pdf (accessed on 21 October 2018).

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"A Data Driven Approach For Predicting Patient Readmission Using Ensembled Model "

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
COMPUTER SCIENCE AND ENGINEERING

Submitted by

B.Hima Bindu(17H71A0516)

N.Bhavana Yadav(17H71A0503)

K.Venkata Sai Rahul Teja(17H71A0555)

Sk.Jailauddin (17H71A0519)

Under the Esteemed Guidance of

Mrs.V.Lakshmi Chetana

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by NAAC With Grade A

2020-21

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC A
With Grade

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"A Data Driven Approach For Predicting Patient Readmission Using Ensembled Model"** is a bonafide work carried out by **B.Hima Bindu(17H71A0516),N.Bhavana Yadav(17H71A0503), Venkata Sai Rahul teja(17H71A0555), Sk.Jailauddin(17H71A0519)** in partial fulfilment for the award of degree of Bachelor of Technology in **Computer Science and Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

V.Lakshmi Chetana
Project Guide

D. Prasad
Head of the Department

Examiner

Dr K Srinivas
PRINCIPAL

ACKNOWLEDGEMENT

We would like to take this opportunity to express my deepest appreciation to the following people for their valuable contributions and assistance with this Project.

Initially, we would like to thank my project supervisor, **Mrs, V. Lakshmi Chetana, Assistant Professor, Department of Computer Science and Engineering** for the guidance and support, especially for the valuable ideas and knowledge shared to me throughout the Project.

We have the immense pleasure in expressing my thanks and deep sense of gratitude to**, Mr. D. Prasad, Head of the Department, Computer Science and Engineering** for extending necessary facilities and support for the completion of the Project.

We whole sincerely acknowledge **Dr. K. Srinivas, Principal and Prof D. PANDURANGA RAO, CEO** for giving opportunity to take up the Project work. We also extend my thanks to all faculty members of **Computer Science and Engineering**, for their valuable guidance and encouragement.

We would like to extend my warm appreciation to all my friends for sharing us their knowledge, valuable contributions and help with this Project.

Finally, my special thanks go to my family for their continuous support and help throughout my academic years and for their continual support and encouragement for the completion of the project.

DECLARATION

We **B. Hima Bindu, N. Bhavana Yadav, K. Venkata Sai Rahul teja, Sk. Jailauddin** of the Main-Project "**A Data Driven Approach For Predicting Patient Readmission Using Ensembled Model**", hereby declare that the matter embodied in this Project is the genuine work done by us and has not been submitted either to this university or to any other university/institute for the fulfilment of the requirement of any course of study.

Batch Members Names and Roll Numbers

B. Hima Bindu (17H71A0516)

N. Bhavana Yadav (17H71A0503)

K. Venkata Sai Rahul Teja (17H71A0555)

Sk. Jailauddin (17H71A0519)

ABSTRACT

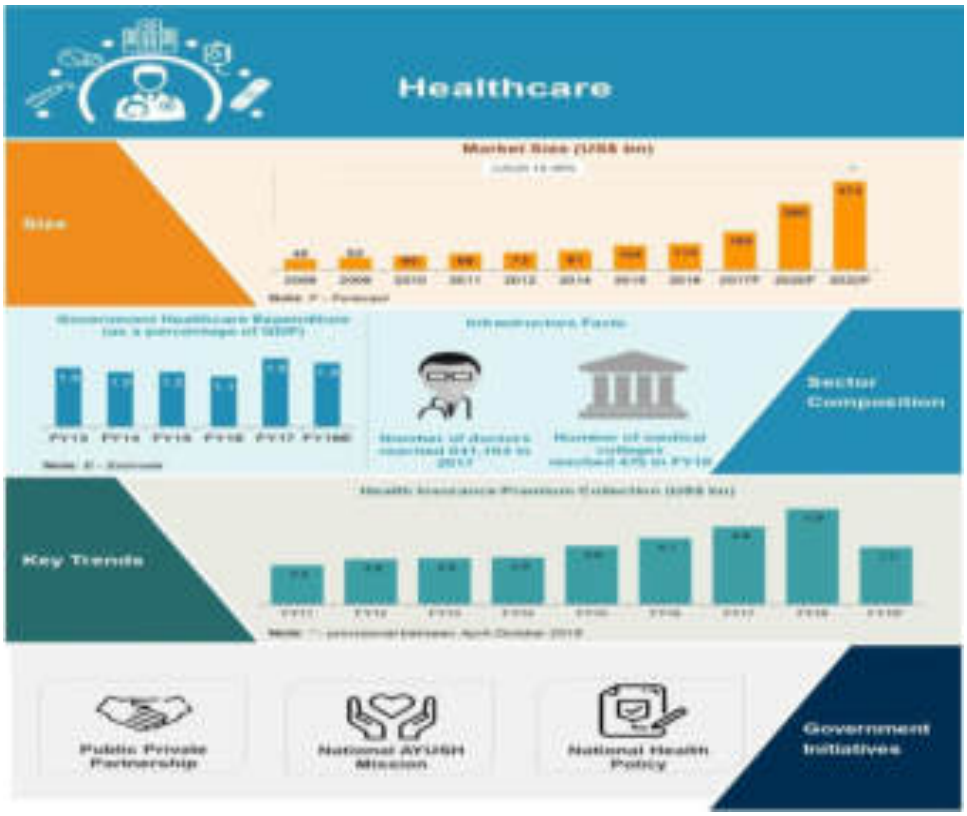
A major concern for modern hospitals is counting how many patients are readmitted to their Hospital within 30 days after an original admission. As the healthcare system moves toward value-based care, many programs to improve the quality of care of patients were developed.

One of these programs is called the Hospital Readmission Reduction Program (HRRP), which reduces reimbursement to hospitals with above average readmissions. Machine Learning can help the Hospital system identify patients with chronic illness that are undiagnosed or misdiagnosed, predict the possibility and present patient-specific preventive interventions. Recent advancement in ML in developing and executing it on patient's health record dataset has depicted prominent outcomes. The ML models are completely dependent on labelled data for better predictions of the results. The state of art of ML models has given a bunch of solutions to provide better results for clinical tasks which seem to be impossible with traditional practices. To develop a strong and more accurate machine learning model, we can use data collected from studies carried out, medical health records, and other sources.

CONTENTS

1. INTRODUCTION

A patient who had been discharged from a hospital is admitted again within as specified time interval. Readmission rates have increasingly been used as an outcome measure in health services research and as a quality benchmark for health systems. Generally, higher readmission rate indicates in effectiveness of treatment during past hospitalizations. Hospital readmission rates were formally included in reimbursement decisions for the Centers for Medicare and Medicaid Services (CMS) as part of the Patient Protection and Affordable (ACA) of 2020.which penalizes health systems with higher than expected readmission rates through the Hospital Readmission Reduction Program. Since the inception of this penalty, there have been other programs that have been introduced, with the aim to decrease hospital readmission. The Community Based Care Transition Program, Independence At Home Demonstration Program,and Bundled Payments for Care Improvement Initiative are all examples of these programs. While many time frames have been used historical the most common time frame is within 30 days of discharge, and this is what CMS uses.



1.1 INTRODUCTION IN HEALTHCARE INDUSTRY WITH ML

Introduction: The emerging use cases of Artificial Intelligence (AI) in the healthcare sectorcan be seen as a collection of technologies enabling machines to sense, comprehend, act and learn so they can perform administrative and clinical healthcare functions, as well as be used in research and for training purposes.8 Unlike legacy technologies that only complemented human skills, health AI today can significantly expand the scope of human activity. These technologies9 include, among others, natural language processing, intelligent agents, computervision, machine learning, expert systems, chat bots and voice recognition.10 These technologies can also potentially be used to compensate for a physician's cognitive biases (such as "recency bias," where one is more likely to allow the last case one treated to inform the course of treatment for the next patient.)11 This use and adoption of AI can be seen at varying levels across the healthcare ecosystem. Machine learning can be used to address the issue of reporting in siloed Electronic Health Records (EHRs) and instead redirect these reports toward analysis and predictive modelling.12 This technology can also be applied to preventative health programs. Machine learning can be used to merge an individual's –omic (genome, proteome, metabolome, microbiome) data with other data sources such EHRs to predict the likelihood ofdeveloping a disease, which can then be addressed through timely interventions such as preventative therapy

The healthcare industry in India is made up of a number of segments.40 Through a review of companies developing AI solutions for health, health practitioners using AI, and researchers looking into the potential of AI and health, it was found that AI is employed in a variety of ways across the different segments including:

1. Hospitals: These include government hospitals, including healthcare centres, district hospitals and general hospitals; and private hospitals, which include nursing homes and mid- tier and top-tier private hospitals.

From a review of solutions adopted it appears that hospitals in India are employing descriptive and predictive AI. For instance, the Manipal Group of Hospitals has tied up withIBM's Watson for Oncology to aid doctors in the diagnosis and treatment of 7 types of cancer.

Here, AI is used to analyse data and research evidence and improve the quality of the report, in turn increasing patient trust. Importantly, patients are fully aware of the process and providetheir express consent. Due care is also taken to preserve patient anonymity. However, at the global level, Watson has recently come under fire from physicians across the world for allegedly posing as "a 'mechanical turk' - a human-driven engine masquerading as an artificialintelligence." It was reported that instead of using AI, it actually works by convening a small panel of cancer experts, who recommendations for specific patient profiles - "These recommendations represent the best guesses of these experts, supported by medical literature and personal experience. IBM has never allowed an independent study of Watson for Oncology. No follow up is done to evaluate whether its recommendations help patients."42 Foreign physicians have also complained that the population that Watson is trained on does notaccurately reflect the diversity of cancer patients across the world, and as a result, it is heavilybiased towards American patients and standards of care.43 Aravind Eye Care Systems is presently working with Google Brain, after previously helping Google develop its retinal screening system by contributing images to train its image parsing algorithms. After successfulclinical trials to detect signs of diabetes-related eye disease, it is now attempting to put it to routine use with patients.

2. Pharmaceuticals: These include manufacturing, extraction, processing, purification and packaging of chemical materials for use as medications for humans or animals. From a reviewof solutions adopted it appears that pharmaceuticals in India are employing descriptive and predictive AI with prototypes for prescriptive AI being developed and tested. The most common use of AI in pharmaceuticals is in drug discovery, where AI is mobilised to scan through all available literature on a particular molecule for a drug (eg. targeted molecule discovery), which would otherwise be impossible for even a group of people to manually carry out. In addition to streamlining the process of drug discovery, the application of AI in pharmacy offers additional advantages such as identification of both tangible and intangible enhanced value proposition, enhanced competitor differentiation, optimal resource allocation for maximum market share gain, revenue and profit, ability to maximize growth,

customizing sales and marketing messaging for greater customer engagement, andautomation of sales and marketing messages and channels.

3. Diagnostics: These comprise businesses and laboratories that offer analytical or diagnostic

services. In addition to bigger companies such as Google and IBM, India is also host to

start-up companies that specialise in harnessing AI to diagnose disease. From a review of solutions adopted it appears that diagnostics in India are employing descriptive and predictive AI.

According to the WHO, India is home to over five crore Indians suffering from depression, and is a major contributor to global suicides.56 However, seeking help for mental health issues is still stigmatised. Firms are addressing this issue by using technology to help deal with mental health issues, usually in the form of chat bots that offer counselling while maintaining privacy. In India, AI is being employed through chat bots such as Wysa that provide mental health support. A person can chat anonymously with an AI-enabled system, and the Chabot is intended to provide empathetic support and suggest practitioners to consult. However, these Chabot.

4. Medical Equipment and Supplies: s have not been designed to provide diagnosis or deal with more serious issues (these are transferred to doctors). The advantage of these chat bots is the potential anonymity and privacy they could provide if designed with a privacy enhancing approach – there is no need to sign up, and no personal information needs to be collected. Chat bots have also been presented as an interface which is non-judgmental and consequently, more empathetic towards the concerns of patients. This could encourage people to open up without hesitation.

This includes establishment's primarily manufacturing medical equipment and supplies, e.g. surgical, dental, orthopaedic, ophthalmologic, laboratory instruments, etc. From a review of solutions adopted it appears that companies developing medical equipment and supplies in India are employing descriptive and predictive AI.

5. Medical Insurance: This includes health insurance and medical reimbursement facilities, covering an individual's hospitalisation expenses incurred due to sickness. From a review of solutions adopted it appears that companies offering medical insurance in India are employing descriptive and predictive AI. Machine learning is able to automate claims management by analysing vast amounts of data in less time, which reduces processing time

and handling costs and improves customer experience.

Identifying suspicious patterns in data can also help identify fraudulent claims, which could speed up settlement of genuine claims. By combining big data with AI, insurers can identify the lifestyle habits of customers to provide them customised offerings.

Big data can also be harnessed to enable insurers to identify early-stage illnesses and reduce the risk of treatment-related complications.63 At present, insurers in India are limited to managing operations.

1.2 MOTIVATION:

If patients at risk of readmission could be identified and offered early interventions then their lives and long-term health may be improved by reducing the chances of readmission, and hopefully their cost of care reduced.

1.3 PROBLEM STATEMENT:

The existing system is time consuming process. And it is very difficult to detect it in early stages. To the best of our knowledge, traditional machine learning prediction algorithms may not give better accuracyresults so, we are using ensembled model to improve the accuracy.
Main aim is to predict if a patient with diabetes will be readmitted to the hospital within 30 days using Random Forest Algorithm.

1.4 OBJECTIVES:

Currently, the scenario is if the patient is suffering from any symptoms, then he/she must visit to the doctor or to the hospital to diagnose the disease. But our main objective is to reduce such efforts taken by patients only to diagnose the disease. Many patients are losing their life only because of the late diagnosis of their disease. So our main aim is to reduce such deaths.

- To identify the patients with increased risk of readmission.
- Build a predictive model that focuses on predicting hospital readmission for patients with diabetes.

2. REVIEW OF LITERATURE

Diabetes is one of the chronic non-communicable diseases that are on the rise with massive urbanization and a drastic change of lifestyle in many countries. It is expected to turn into the seventh most prevalent mortality factor by 2030 and millions of deaths could be prevented each year through better analytics. Therefore, diabetes is on the health agenda of most developed and developing countries. Healthcare industry collects and process diabetes patient medical data in huge volume, diverse structure, and real-time flow of data. With the rise of technology,both from the diagnosis and monitoring, storage and analysis, novel solutions are now available to better address challenges like non-invasive screening, tailor made treatment, and hospital readmissions.

When assessing the quality of care delivered by a health canter, readmission is the metric of choice. It measures the number of patients that need to come back to the hospital after their initial discharge. The readmission can be classified into three broad categories such as unavoidable, planned, and unplanned. The unavoidable readmission that is highly predictable mostly due to the nature of the pathology or patient's condition (i.e. cancer phase IV,metastasis). Secondly in the planned readmission which is directly prescribed by the healthcareprofessional to the patient (i.e. check-up, transfusion). Lastly, the unplanned is defined as readmission that shouldn't have happened given the practitioner's diagnosis and could have been avoided if proper care was given to the patient post-discharge.

Unavoidable and planned readmissions already are highly anticipated. However, predicting unplanned readmission is of prime interest due to its inherent uncertainty. Unplanned readmission is the most useful type when evaluating the quality of care of a hospitalas it highlights a practitioner's diagnosis or treatment error. Beyond being a core indicator of the quality of care, unplanned readmissions also constitute a financial problem for nations. Therefore, with a predictive model to assess unplanned readmission risk could optimize the quality of hospital services.

Moreover, several researchers have proposed a predictive model for readmission in healthcare for all types of diseases and only limited work are dedicated to diabetes. As different pathologies have different conditions and behaviors, prediction on specific pathology subset would highly benefit the prediction model's performance.

Readmissions occurring after 30 days have less correlation with the quality of care from the health center and might be an impact due to external factors such as complications or patient's lifestyle. Numerous researches highlighted clear interest in 30-day unplanned readmission prediction models based on diabetes complications.

The purpose of this study is to propose a prediction model for 30-day unplanned readmission among diabetes patients in US hospitals. The analysis will be based on risk factors such as a patient's demographics, admission details, diagnosis, and medical data. In a broader sense, the goal of the study is to allow health centres to better anticipate and address unplanned readmissions while improving their quality of care and cost efficiency.

3. SYSTEM ANALYSIS

EXISTING SYSTEM:

- The existing work will be done on unstructured or textual data. For the prediction of diseases, the existing willbe done on linear, KNN etc.
- The existing models are characterized by limited prediction power, generalizability and pre- processing. so we are moving to ensemble method which gives better accuracy.

Disadvantages:

- In the existing system, practical use of various collected data is time consuming.
- Detection is not possible at an earlier stage.

PROPOSED SYSTEM:

- It takes less training time as compared to other system.
- It predicts output with high accuracy even for the large data set it runs efficiently.
- It also utilizes bootstrap aggregating, also known as bagging, to reduce over fitting and improve generalization accuracy.

Advantage:

- High performance and accuracy rate.

Software Requirements:

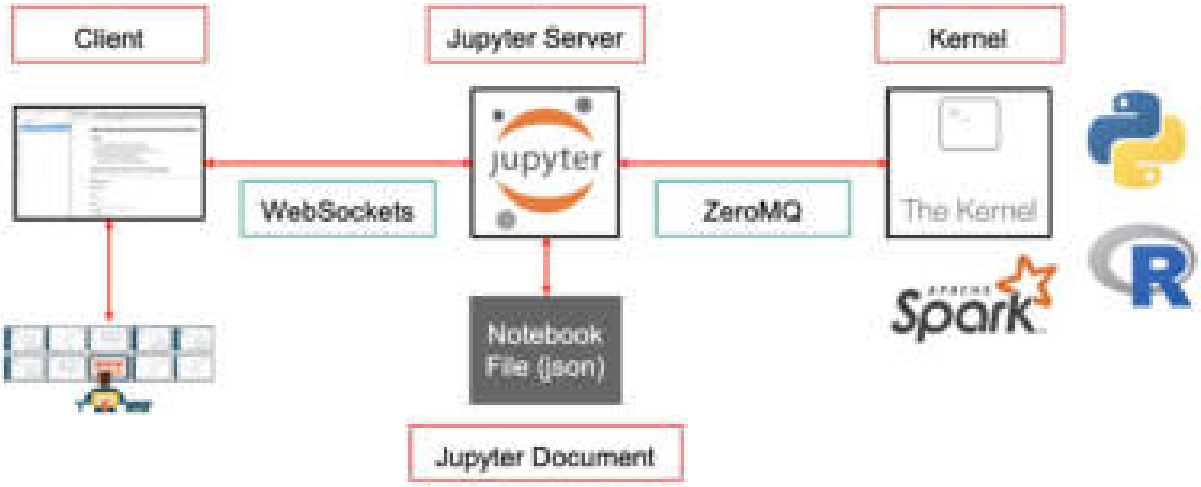
1. Python IDLE 3.7 version (or)
2. Anaconda 3.7 (or)
3. Jupiter (or)
4. Google colab

Hardware Requirements:

1. Operating system: Windows, Linux.
2. Processor: minimum Intel i3
3. Ram: minimum 4 GB
4. Hard disk: minimum 250gb

4. SYSTEM DESIGN

Jupyter Notebooks work with what is called a two-process model based on a kernel-client infrastructure. This model applies a similar concept to the Read-Evaluate-Print Loop (REPL) programming environment that takes a single user's inputs, evaluates them, and returns the result to the user. Based on the two-process model concept, we can explain the main components of Jupyter in the following way



Jupyter Client:

- It allows a user to send code to the kernel in a form of a Qt Console or a browser via notebook documents.
- From a REPL perspective, the client does the read and print operations.
- Notebooks are hosted by a Jupyter web server which uses Tornado to serve HTTP requests

Jupyter Kernel:

- It receives the code sent by the client, executes it, and returns the results back to the client for display. A kernel process can have multiple clients communicating with it which is why this model is also referred as the decoupled two-process model.
- From a REPL perspective, the kernel does the evaluate operation.
- kernel and clients communicate via an interactive computing protocol based on an asynchronous messaging library named ZeroMQ (low-level transport layer) and Web Sockets (TCP-based)
- Makes Jupyter a language agnostic application (Julia, Python, R, etc.)

What is Anaconda Python?

Together with a list of Python packages, tools like editors, Python distributions include the Python interpreter.

Anaconda is one of several Python distributions. Anaconda is a new distribution of the Python and R data science package. It was formerly known as Continuum Analytics. Anaconda has more than 100 new packages.

This work environment, Anaconda is used for scientific computing, data science, statistical analysis, and machine learning. The latest version of Anaconda 5.0.1 is released in October 2017.

The released version 5.0.1 addresses some minor bugs and adds useful features, such as updated R language support. All of these features weren't available in the original 5.0.0 release.

This package manager is also an environment manager, a Python distribution, and a collection of open-source packages and contains more than 1000 R and Python Data Science Packages.

Modules in project:

1. Data Collection
2. Data Preparation
3. Choosing Learning Algorithm
4. Training Model
5. Evaluating Model
6. Predictions

1. Data Collection-

In this stage,

- Data is collected from different sources.
- The type of data collected depends upon the type of desired project.
- Data may be collected from various sources such as files, databases etc.
- The quality and quantity of gathered data directly affects the accuracy of the desired system.

2. Data Preparation-

In this stage,

- Data preparation is done to clean the raw data.
- Data collected from the real world is transformed to a clean dataset.
- Raw data may contain missing values, inconsistent values, duplicate instances etc.
- So, raw data cannot be directly used for building a model

Different methods of cleaning the dataset are-

- Ignoring the missing values
- Removing instances having missing values from the dataset.
- Estimating the missing values of instances using mean, median or mode.

- Removing duplicate instances from the dataset.
- Normalizing the data in the dataset.

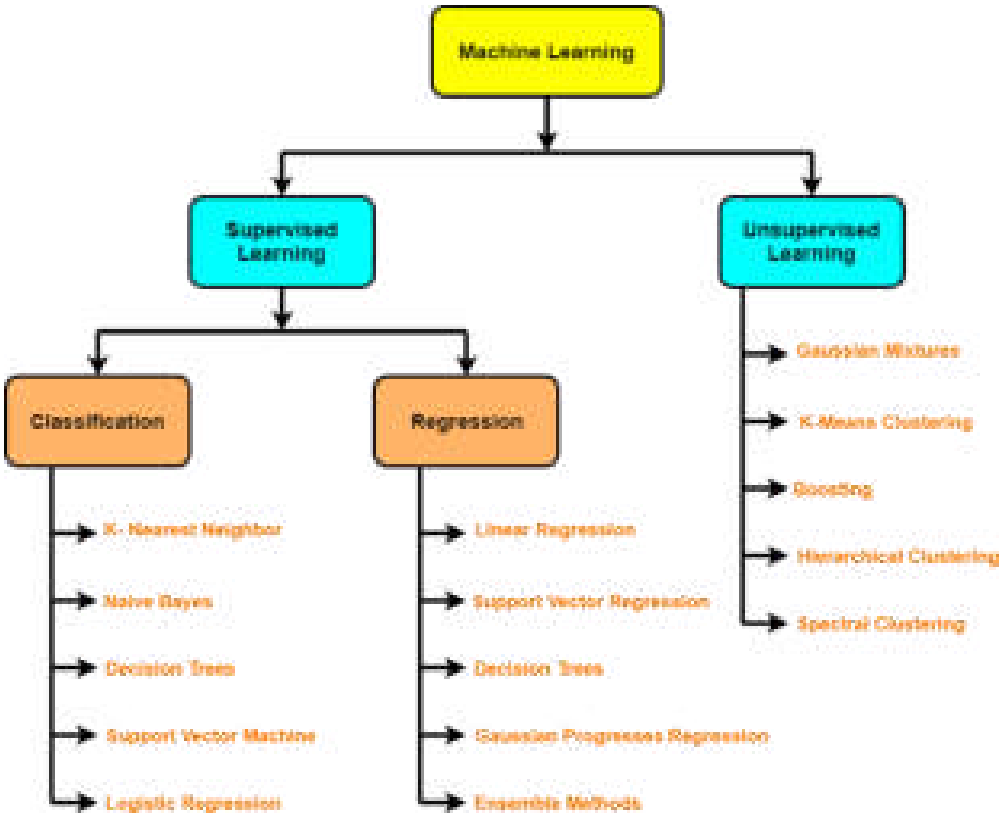
This is the most time-consuming stage in machine learning workflow

3. Choosing Learning Algorithm-

In this stage,

- The best performing learning algorithm is researched.
- It depends upon the type of problem that needs to solved and the type of data we have.
- If the problem is to classify and the data is labeled, classification algorithms are used.
- If the problem is to perform a regression task and the data is labeled, regression algorithms are used.
- If the problem is to create clusters and the data is unlabeled, clustering algorithms are used.

The following chart provides the overview of learning algorithms-



4. Training Model-

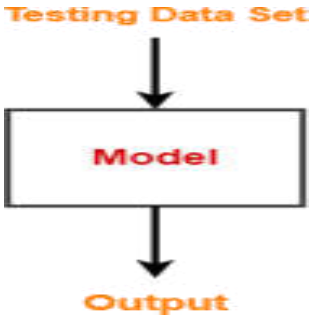
In this stage,

- The model is trained to improve its ability.
- The dataset is divided into training dataset and testing dataset.
- The training and testing split is order of 80/20 or 70/30.
- It also depends upon the size of the dataset.
- Training dataset is used for training purpose.
- Testing dataset is used for the testing purpose.
- Training dataset is fed to the learning algorithm
- The learning algorithm finds a mapping between the input and the output and generates the model.

5. Evaluating Model-

In this stage,

- The model is evaluated to test if the model is any good.
- The model is evaluated using the kept-aside testing dataset.
- It allows to test the model against data that has never been used before for training.
- Metrics such as accuracy, precision, recall etc are used to test the performance.
- If the model does not perform well, the model is re-built using different hyperparameters.
- The accuracy may be further improved by tuning the hyper parameters.



6. Predictions-

In this stage,

- The built system is finally used to do something useful in the real world.
- Here, the true value of machine learning is realized.

Data Set Information:

The dataset represents 10 years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks. It includes over 50 features representing patient and hospital outcomes. Information was extracted from the database for encounters that satisfied the following criteria:

- It is an inpatient encounter (a hospital admission).
- It is a diabetic encounter, that is, one during which any kind of diabetes was entered to the system as a diagnosis.
- The length of stay was at least 1 day and at most 14 days.
- Laboratory tests were performed during the encounter.
- Medications were administered during the encounter.

The data contains such attributes as patient number, race, gender, age, admission type, time in hospital, medical specialty of admitting physician, number of lab test performed, HbA1c test result, diagnosis, number of medication, diabetic medications, number of outpatient, inpatient, and emergency visits in the year before the hospitalization, etc.

5. DATA ANALYSIS & INTERPRETATION

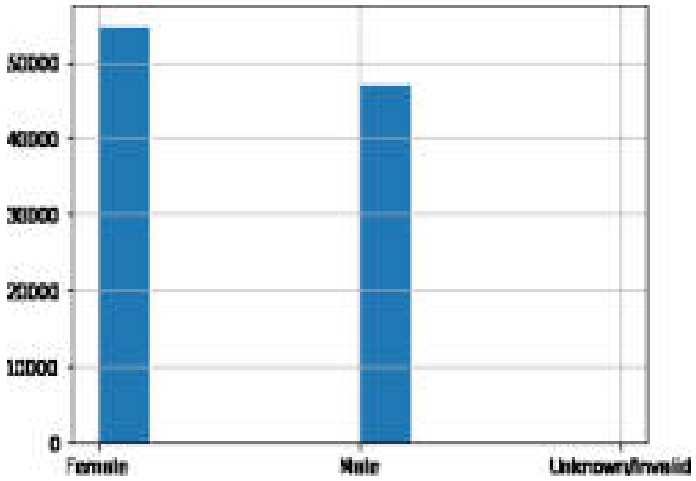
5.1 EXPLORATORY DATA ANANLYSIS (EDA) BEFORE CLEANING:

In the given data set you need to import the data using NumPy, Pandas, matplotlib and scikit learn.

After that using exploratory commends to explore the data.

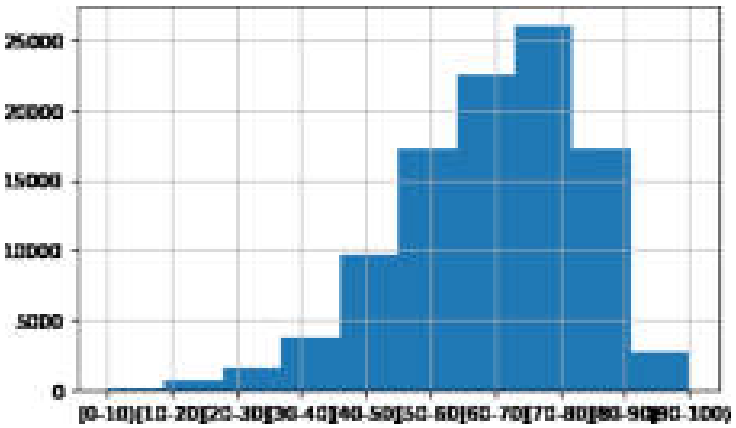
- Using describe command what are the minimum values, maximum values, 25% of the given data and 50% of the given data and 75% of the given data. mean of each column, count value of each column and standardization of the data.
- In the data you check top rows of the given data using head function.
- In the data you check bottom rows of the given data using tail function.
- In the data set you check the data set column names and data type using info function. in this command shows the column names and data type like int, float, string, double and objects
- Using covariance function to find out the covariance of each column if the covariance is positive or negative covariance using covariance function.
- Using Correlation function check whether the correlation between the each column is the correlation is positive correlation or negative or week or strong
- correlation check using this command.
- Using shape command you find out the total number rows and total number ofcolumns in the given data set.
- Using pivot table to find out the pivot values of each column.

1) Histogram for Gender



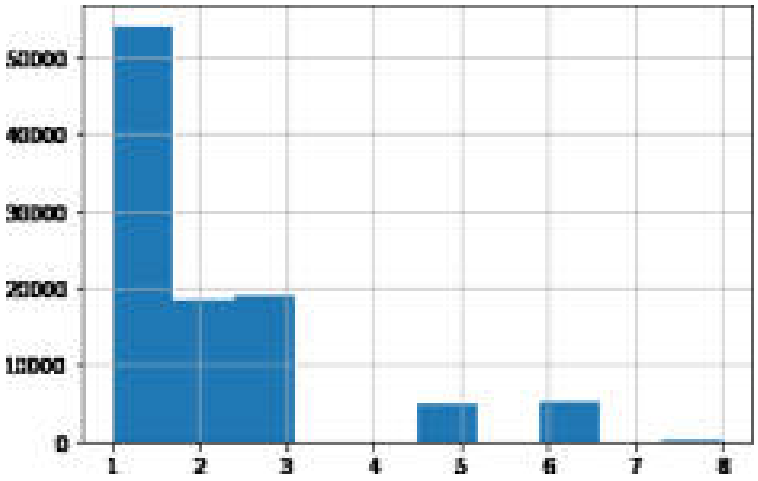
In this gender histogram female patients are high when compared to male &unknown/invalid.

2) Histogram for AGE



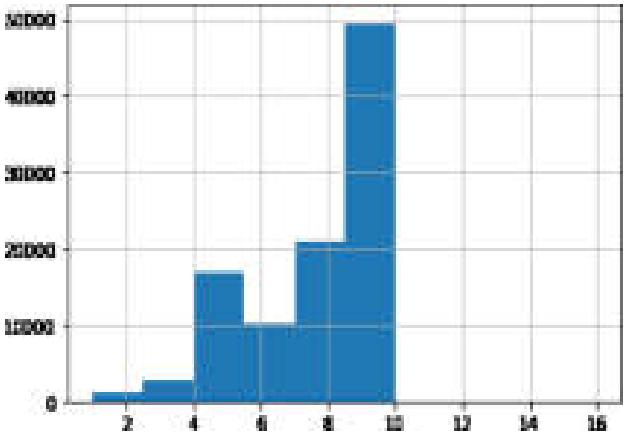
In above histogram Age between (70-80] are above 25000 patients when compared to other ages.

3. Histogram for Admission _ type _ id



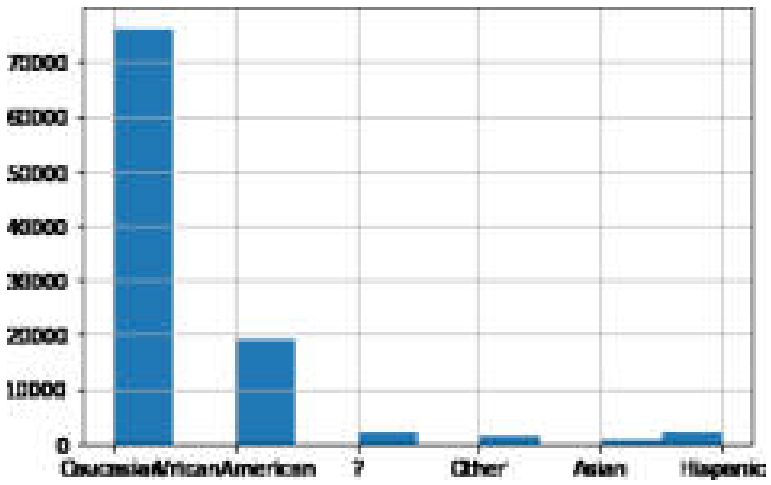
In the above histogram admission type id 1 is above 50000 patients when compared to other numbers

4. Histogram for number _ diagnosis



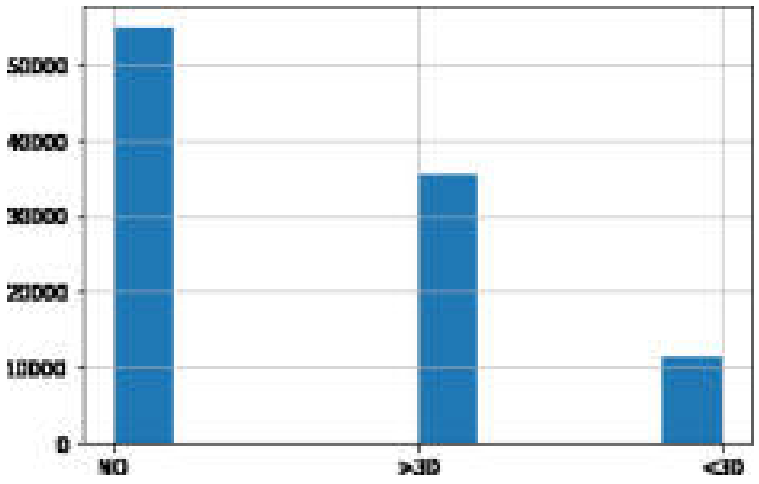
In the above histogram10 maximum number of diagnoses patients.

5. Histogram for Race



In the above histogram in the Caucasian city is the maximum number of patients shows in thediagram.

6. Histogram for readmitted

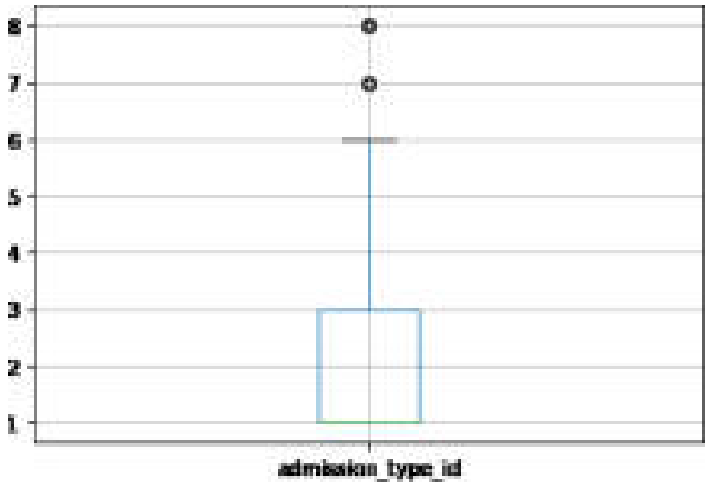


In the above histogram shows maximum number of patients are not readmitted.

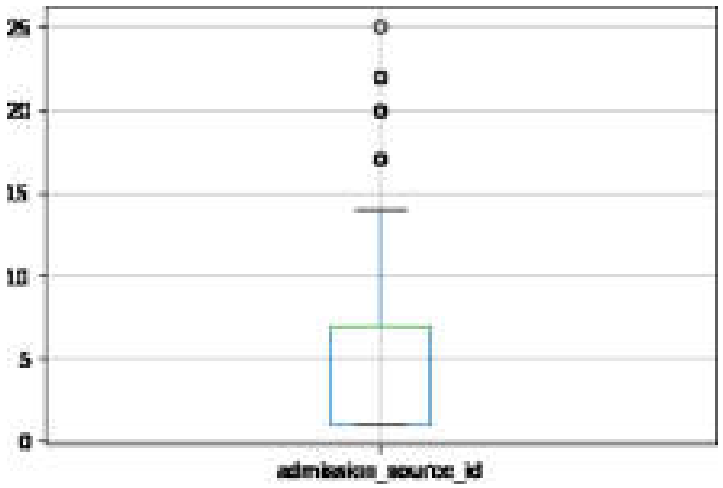
Just 35000 patients are readmitted above 30 days.11000 patients are readmitted less than 30 days.

BOX PLOTS:

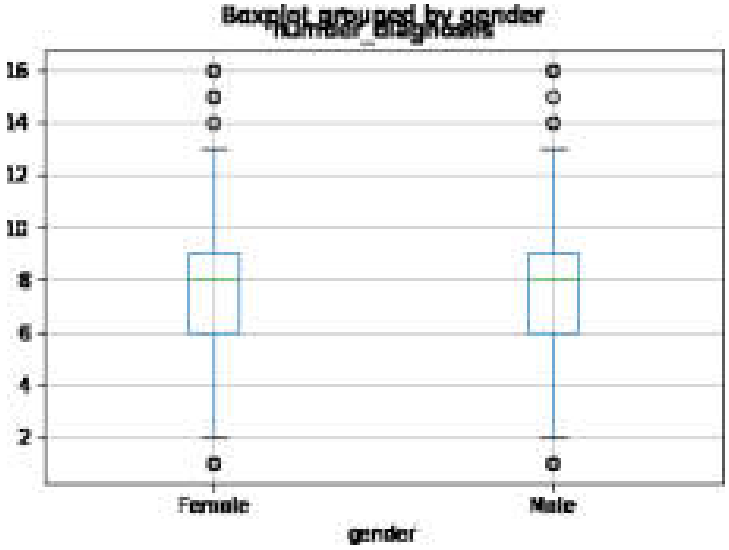
1. Box plot for Admission _ type _ id



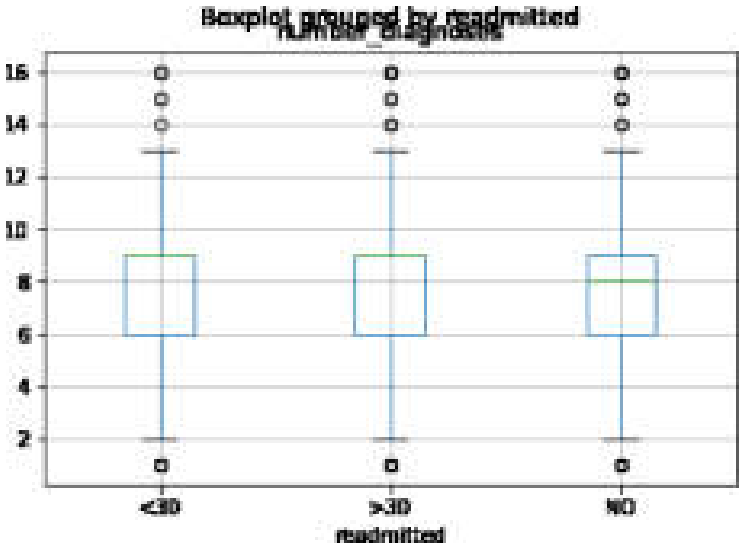
2) Box plot for admission _ source _ id



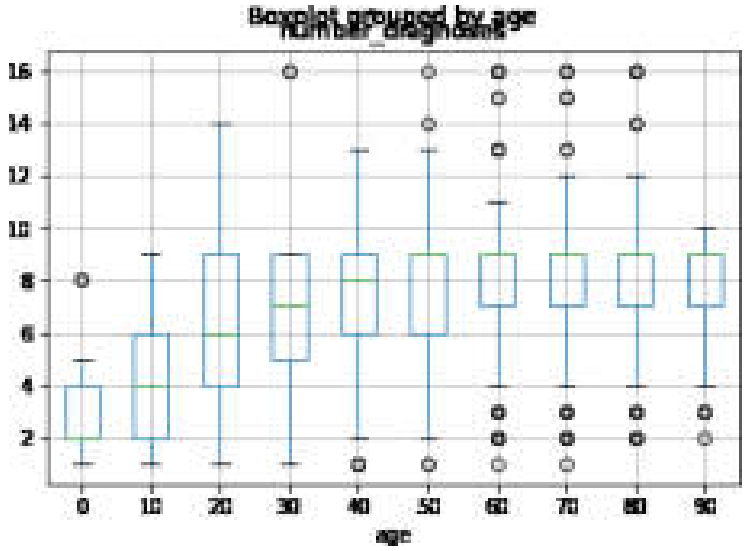
3) Box plot for Number_ diagnosis & Admission _ type _ id



4) Box plot for number _ diagnoses & readmitted



5. Box plot for number _ diagnoses & age



.

5.2 DATA CLEANING:

In the given data set if the data is missing the data set you need to clean the data usingfollowing steps.

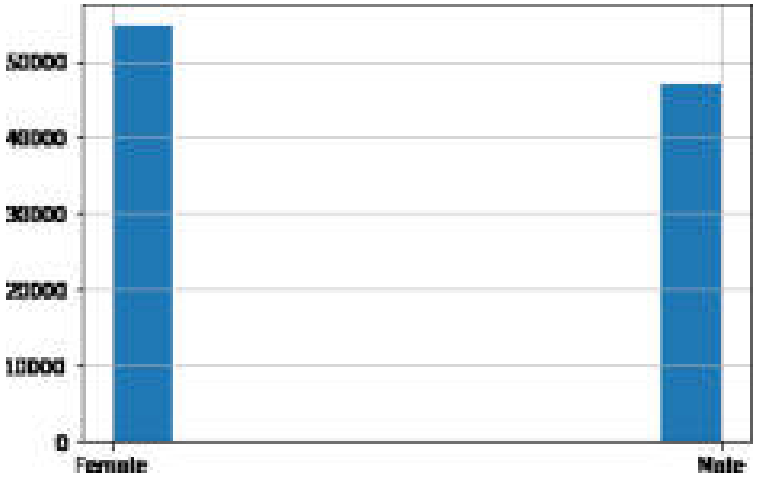
- First in the given data set if the column is “blank” and “?” to clean the data. if the data is alphabets to convert alphabets to numeric values and empty data and? to filled with some values.
- In the diabetic data sets lot of missing values like black column, alphabets and “?”
- First you need clear the missing values using following steps.
- First values count using count function. if the data missing in the rows filled with maximum count values. If the count value is numeric filled with mean value.
- In the data lot of “? “Is there you filled the “? “Maximum count function.
- In the given data set lot of between values like this (75-100]. when the data this type one time you filled with minimum values and one time filled with maximum values.
- in the given data sets some alphabets values like V14, V15. You need to convert alphabets to numeric values.
- Final step of the data cleaning is converting all data into numeric values. this is thedata cleaning process.

5.3 EXPLORATORY DATA ANANLYSIS (EDA) AFTER CLEANING:

- Using describe command what are the minimum values, maximum values, 25% of the given data and 50% of the given data and 75% of the given data. Mean of each column, count value of each columns and standardization of the data.
- In the data you check top rows of the given data using head function.
- In the data you check bottom rows of the given data using tail function.
- In the data set you check the data set column names and data type using info function. in this command shows the column names and data type like int,float,string,double and objects.
- Using covariance function to find out the covariance of each column if the covariance is positive or negative covariance using covariance function.
- Using Correlation function check whether the correlation between the each column is the correlation is positive correlation or negative or week or strong
- correlation check using this command.
- Using shape command you find out the total number rows and total number ofcolumns in the given data set.
- Using pivot table to find out the pivot values of each column.

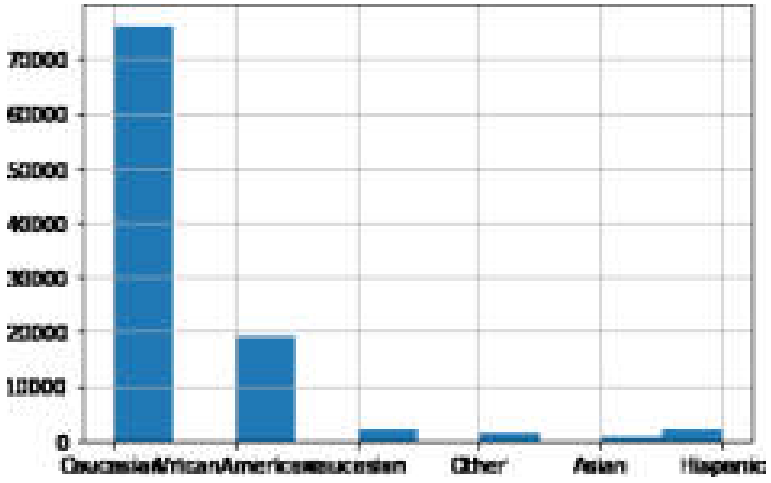
HISTOGRAMS:

1) Histogram for gender



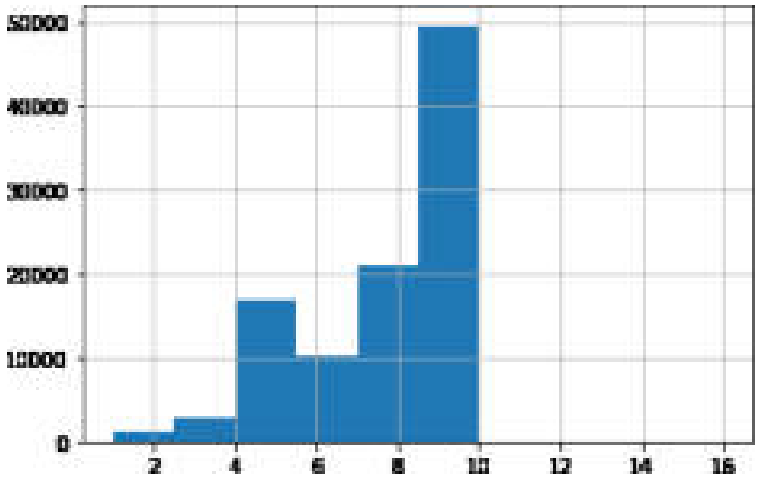
In this gender histogram female patients are high when compared to male &unknown/invalid

2) Histogram for race



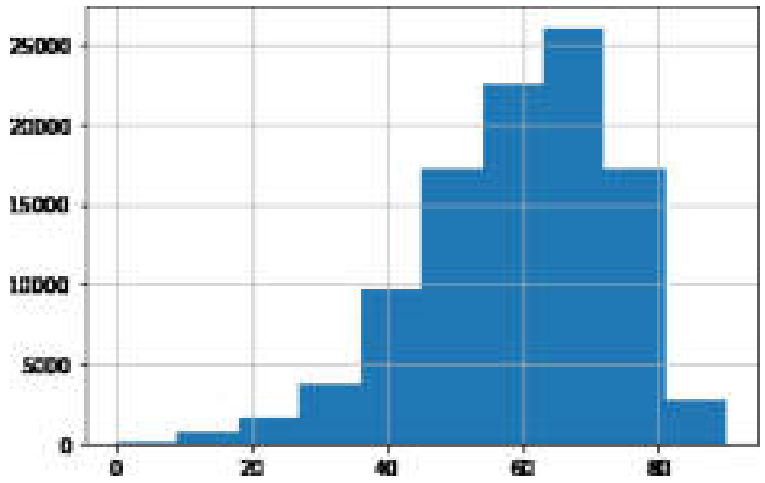
In above histogram Age between (70-80] are above 25000 patients when compared to otherages.

3) Histogram for number _ diagnoses



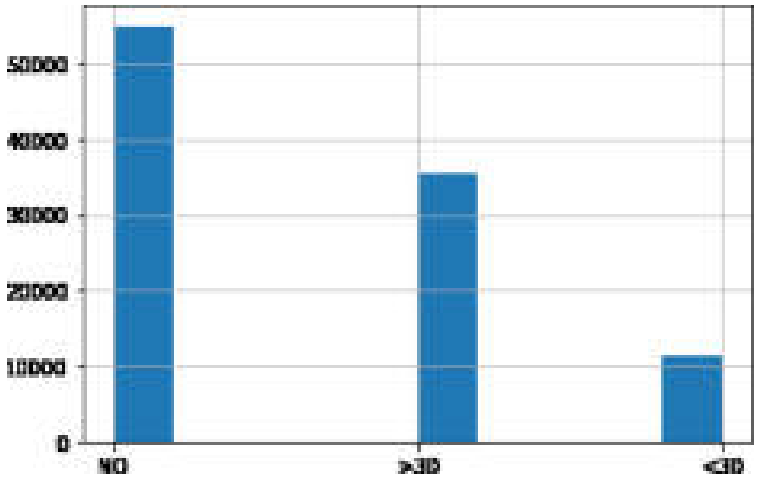
In the above histogram10 maximum number of diagnoses patients.

4) Histogram for age



In above histogram Age between (70-80] are above 25000 patients when compared to otherages.

5) Histogram for readmitted

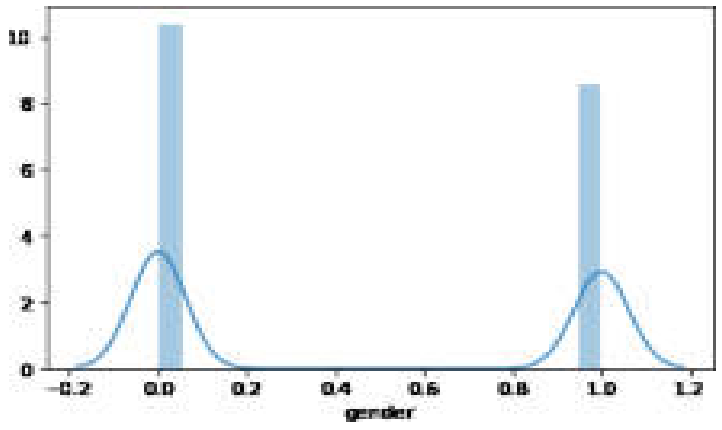


In the above histogram shows maximum numbers of patients are not readmitted.Just 35000 patients are readmitted above 30 days.

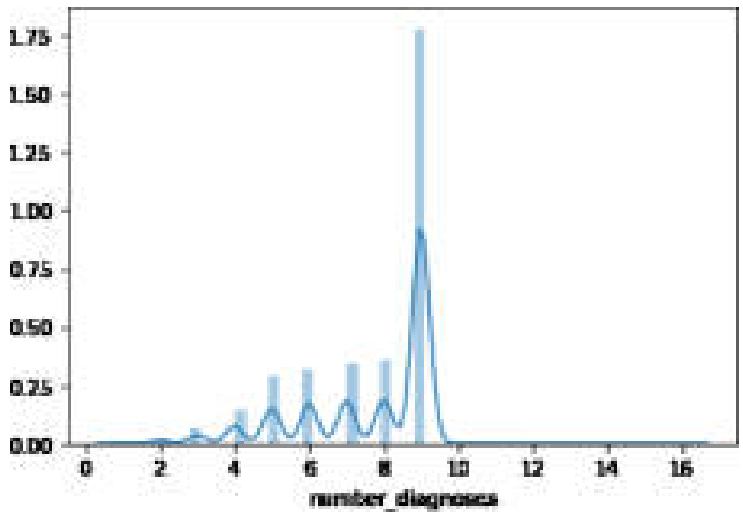
11000 patients are readmitted less than 30 days

SNS PLOTS:

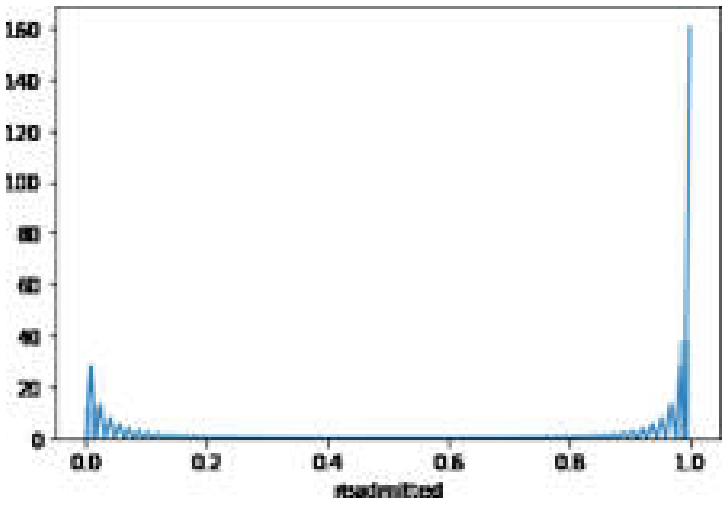
1) This is SNS plot for gender



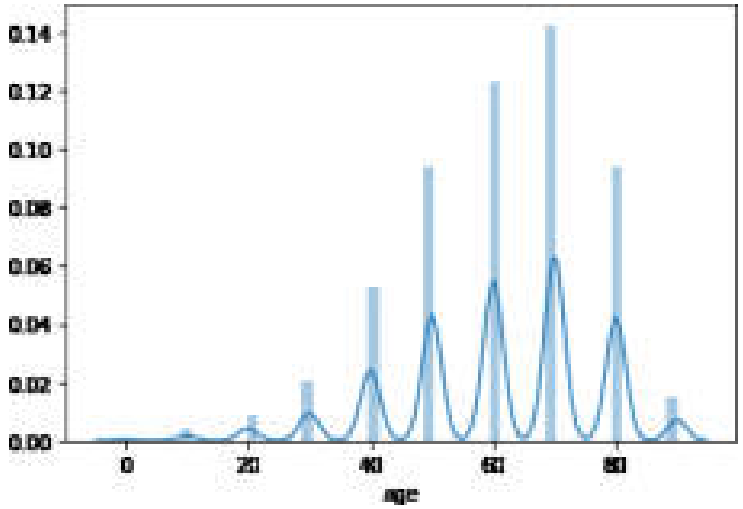
2) This is SNS plot for number _ diagnoses



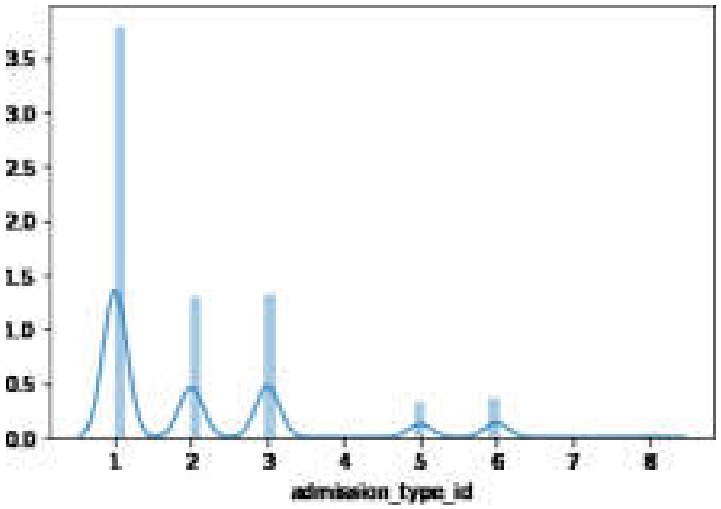
3) This is SNS plot for readmitted



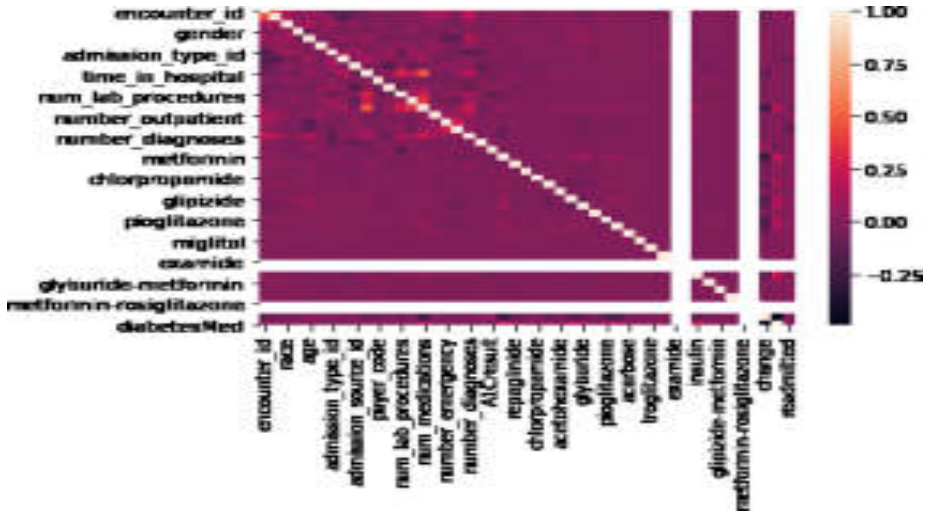
4) This is SNS plot for AGE



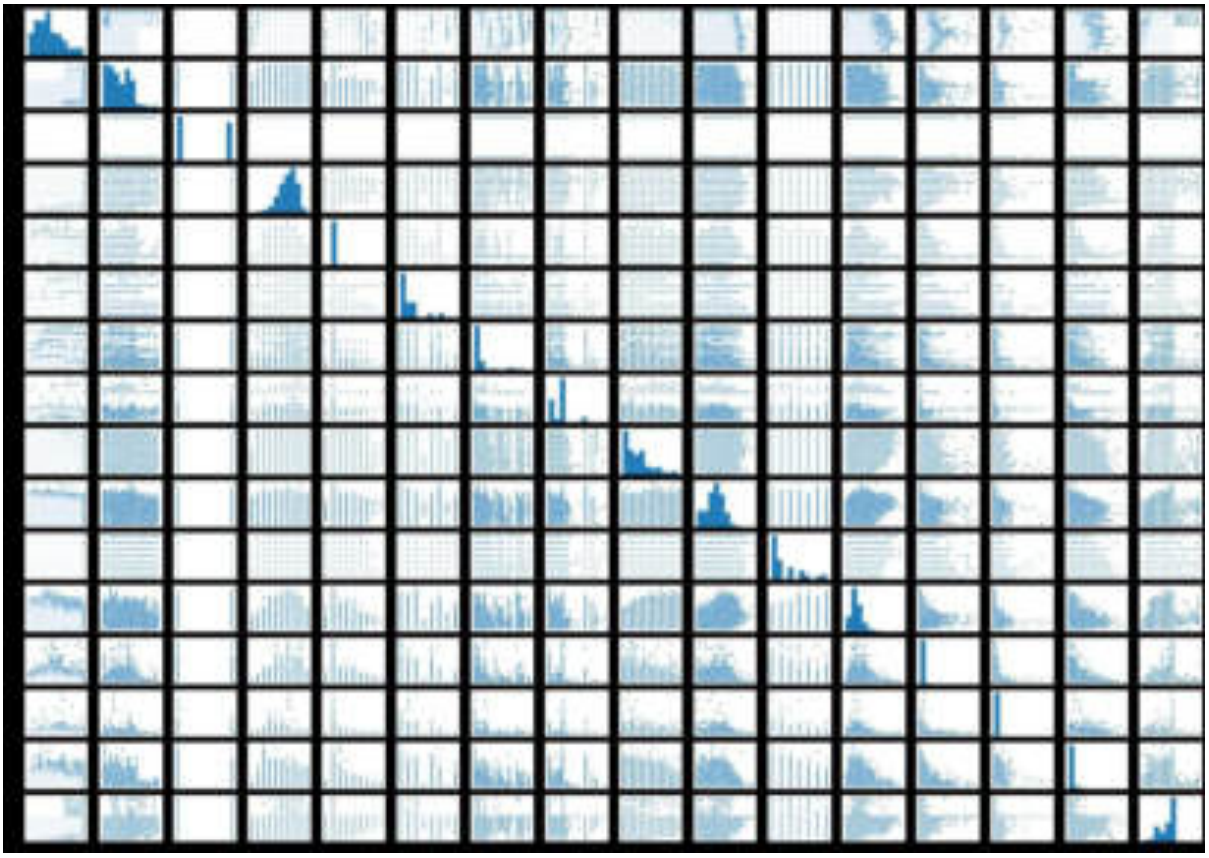
5) This is SNS plot for admission _type _id



This is heat map for the correlation



7) This is SNS pair plot for



5.4 FEATURE ENGINEERING:

Feature engineering is the process of using domain knowledge to extract features from raw data via data mining techniques. These features can be used to improve the performance of machine learning algorithms. Feature engineering can be considered as applied machine learning itself. What is a feature and why we need the engineering of it? Basically, all machine learning algorithms use some input data to create outputs. This input data comprises features, which are usually in the form of structured columns. Algorithms require features with some specific characteristic to work properly. Here, the need for **feature engineering** arises. I think feature engineering efforts mainly have two goals:

- Preparing the proper input dataset, compatible with the machine learning algorithm requirements.
- Improving the performance of machine learning models.

5.5 PREDICTIVE MODELLING:

Predictive modelling is the general concept of building a model that is capable of making predictions. Typically, such a model includes a machine learning algorithm that learns certain properties from a training dataset in order to make those predictions.

5.5.1 KNEAREST NEIGHBORS:

- Deals with classifying data based on similar characteristics or “nearest” neighbours.
- Nearest neighbour classifiers are defined by their characteristic of classifying unlabeled examples by assigning them the class of the most similar labelled examples.
- Well-suited for classification tasks where relationships among the features and the target classes are numerous, complicated, or otherwise extremely difficult to understand.
- The nearest neighbour’s approach to classification is utilized by the kNN algorithm.

Pros:

- Simple and effective
- Makes no assumptions about the underlying data distribution
- Fast training phase

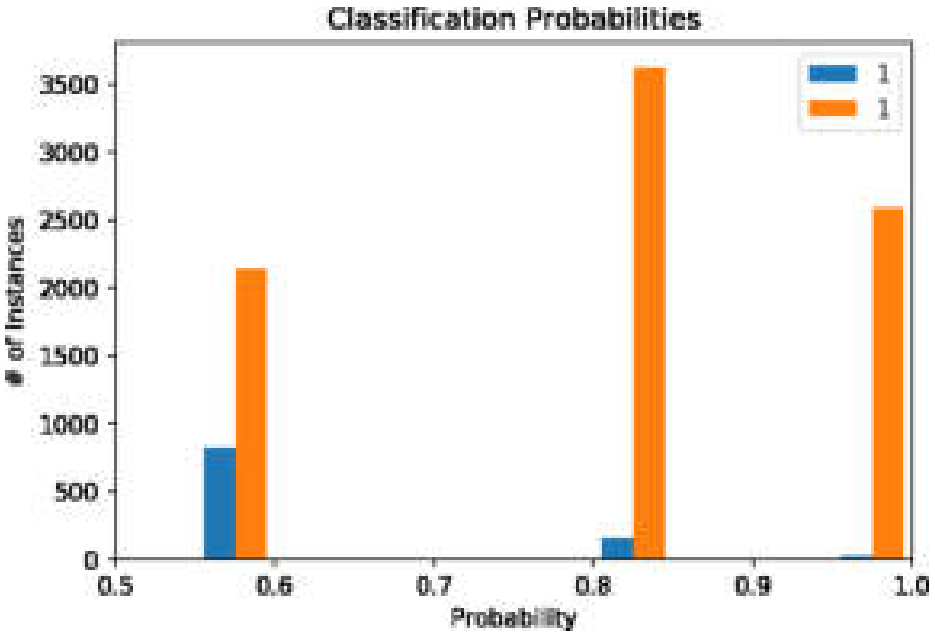
Cons:

- Slow classification phase.
- Requires a large amount of memory.
- Does not produce a model, which limits the ability to find novel insights in relationships among features.

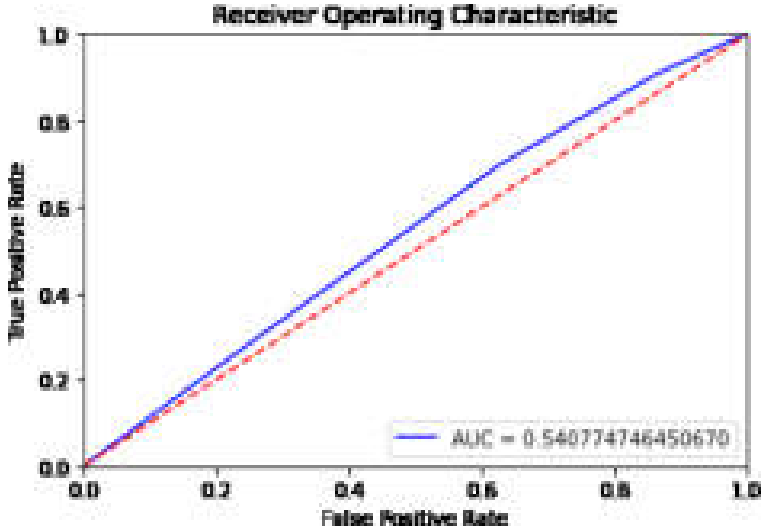
WORKING PRINCIPLE:

Begins with a training dataset made up of examples that are classified into several categories,as labelled by a nominal variable.

- For each record in the test dataset, KNN identifies k records n the training data thatare the "nearest" in similarity.
- The unlabeled test instance is assigned the class of the majority of the k nearest neighbours.
- An instance-based learner in class of non-parametric learning methods.
- No parameters are learned about the data, instead patterns are found rather than fitting data into a preconceived form.
- In the given data set the data splitting 80% of the data is training and 20% of the datais testing.
- This is the confusion matrix for the test data.
- Array ([[288, 1922],[722, 6449]], dtype = int64)
- This the accuracy of the test data is 71.81
- Visualization graph of the data



❖ ROC curve of the KNN is



KNN Algorithm example:

Sample example for KNN

X1 = Acid Durability (Seconds)	X2 = Strength (Kg/Square meter)	Y = Classification)
7	7	Bad
7	4	Bad
3	4	Good
1	4	Good
3	7	?

1. Determine parameter k = number of nearest neighbour , Suppose k = 3
2. Calculate the distance between the query-instance and all the training samples. Coordinateof query instance is (3,7), instead of calculating the distance which is faster to calculate (without square root)

X1 = Acid Durability (Seconds)	X2 = Strength (Kg/Square meter)	Square Distance to query instance(3,7)
7	7	(7-3)^2 + (7-7)^2 = 16
7	4	(7-3)^2 + (4-7)^2 = 25
3	4	(3-3)^2 + (4-7)^2 = 9
1	4	(1-3)^2 + (4-7)^2 = 13

3. Sort the distance and determine nearest neighbours based on the K-th minimum distance.

X1 = Acid Durability (Seconds)	X2 = Strength (Kg/Square meter)	Square Distance Rank to Query min instance distance (3,7)	Is it includedin 3-Nearest neighbours ?
7	7	(7-3) ^2 + (7-7) ^2 = 16(3)	Yes
7	4	(7-3) ^2 + (4-7) ^2 = 25(4)	No
3	4	(3-3) ^2 + (4-7) ^2 = 9(1)	Yes
1	4	(1-3) ^2 + (4-7) ^2 = 13(2)	Yes

4. Gather the category Y of the nearest neighbours.

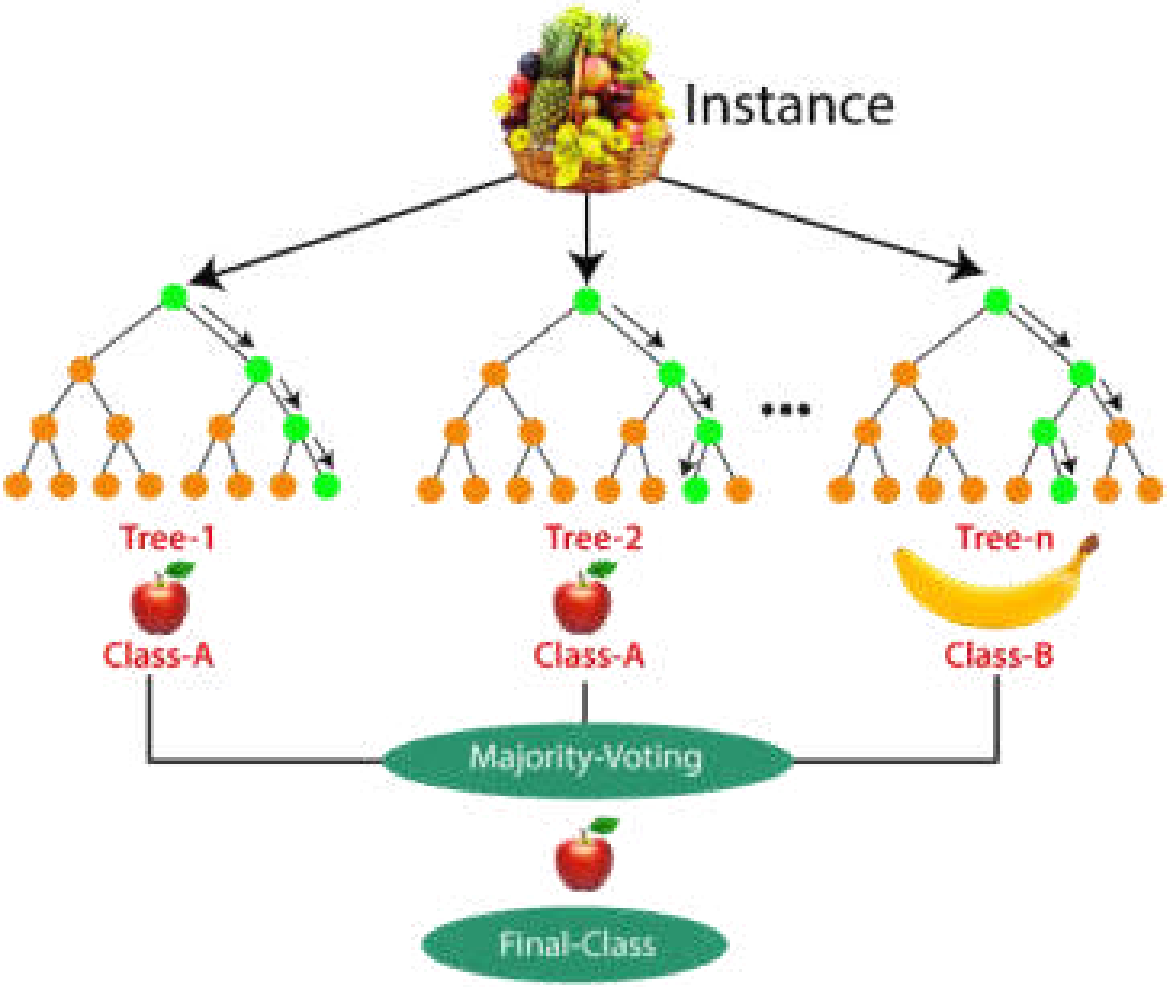
Notice in the second-row last column that the category of nearest neighbour (Y) isnot included Because the rank of this data is more than 3(=k).

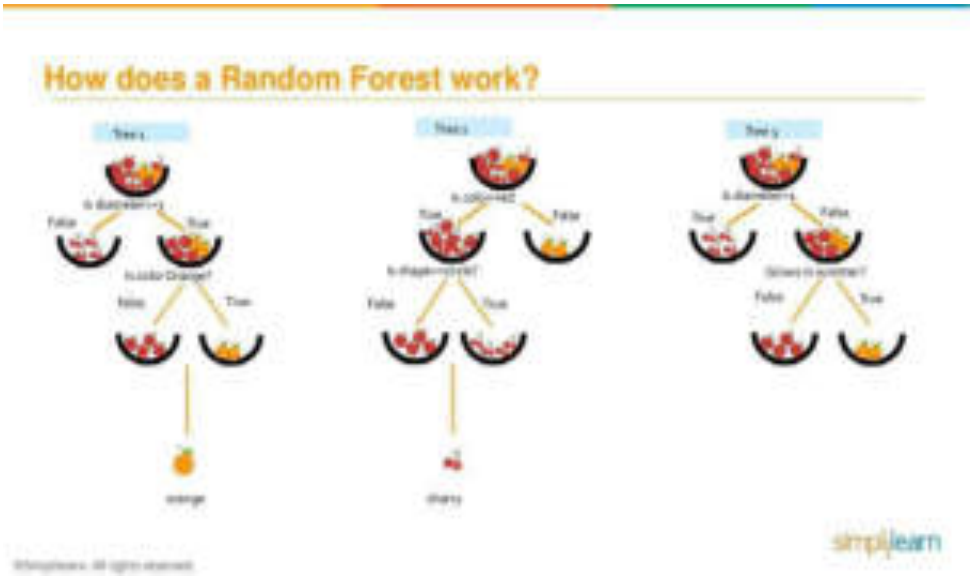
X1 = Acid Durability	X2 = Strength	Square Distance Rank to Query min instance distance (3,7)	Is it included in 3-Nearest neighbors?	Y=categoryof nearest neighbor
7	7	(7-3) ^2 + (7- 7) ^2= 16(3)	Yes	Bad
7	4	(7-3) ^2 + (4-7) ^2 = 25(4)	No	-
3	4	(3-3) ^2 + (4-7) ^2 = 9(1)	Yes	Good
1	4	(1-3) ^2 + (4-7) ^2 = 13(2)	Yes	Good

We have 2 good and 1 bad, then we conclude that pass test with X1 = 3 and X2 = 7 is includedin **good** category.

5.5.2 RANDOM FOREST:

- Random Forest is a popular machine learning Algorithm that belongs to the supervised learning technique.
- As we know that a forest is made up of trees and more trees means more robust forest. Similarly, random forest algorithm creates decision trees on data samples and then getsthe prediction from each of them and finally selects the best solution by means of voting.
- It can be used for both Classification and Regression problems in ML. It is based on the concept of **ensemble learning,** which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model





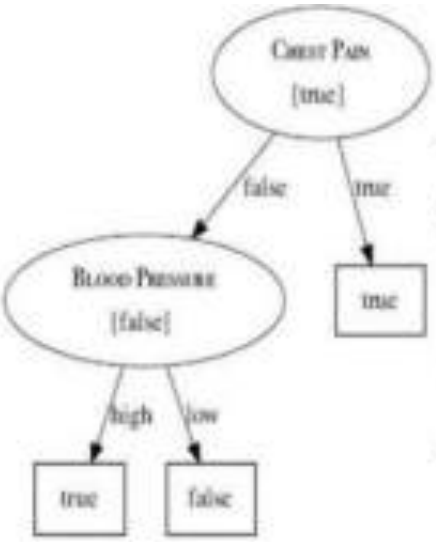
Random Forest algorithm:

Random Forest works in two-phase first is to create the random forest by combining N decision tree, and second is to make predictions for each tree created in the first phase.

The Working process can be explained in the below steps and diagram:

- **Step-1:** Select random K data points from the training set.
- **Step-2:** Build the decision trees associated with the selected data points (Subsets).
- **Step-3:** Choose the number N for decision trees that you want to build.
- **Step-4:** Repeat Step 1 & 2.
- **Step-5:** For new data points, find the predictions of each decision tree, and assign the new datapoints to the category that wins the majority votes.

Example For Random Forest:



ID	Chest Pain	Blood Pressure	Heart Disease
1	false	high	false
2	true	low	true
3	false	low	false
4	true	high	true
5	false	high	false

Random forest example of an ensemble method, meaning that it relies on aggregating the results of an ensemble of simpler estimators.

The somewhat surprising result with such ensemble methods is that the sum can be greater thanthe parts: that is, a majority vote among a number of estimators can end up being better than any of the individual estimators doing the voting! We will see examples of this in the followingsections. We begin with the standard imports: first import the random forest classifier randomstate is 0.

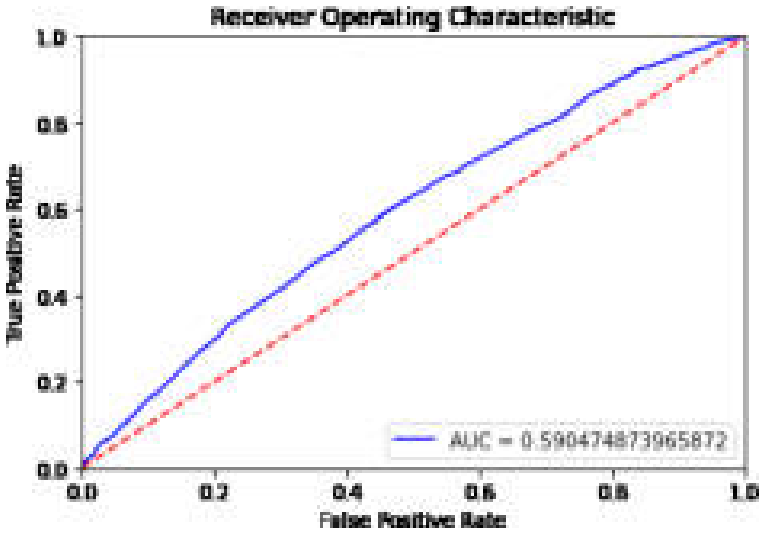
This section contained a brief introduction to the concept of ensemble estimators, and in particular the random forest – an ensemble of randomized decision trees. Random forests are a powerful method with several advantages:

- Both training and prediction are very fast, because of the simplicity of the underlying decision trees. In addition, both tasks can be straight forwardly parallelized, because the individual trees are entirely independent entities.

- The multiple trees allow for a probabilistic classification: a majority vote among estimators gives an estimate of the probability (accessed in Scikit – Learn with the predict_ (probability method).
- The nonparametric model is extremely flexible, and can thus perform well on tasks thatare under-fit by other estimators.

A primary disadvantage of random forests is that the results are not easily interpretable: that is,if you would like to draw conclusions about the meaning of the classification model, random.

- Build confusion matrix on test data is
- Array ([[0, 2210], [0, 7171]], dtype =int64)
- Build confusion matrix on train data is
- Array ([[0, 9147], [0, 28374]], dtype =int64)
- Accuracy score on test data is 76.4%
- Accuracy score on trained data is 75.6%
- Precision score of the test data is 58.4%
- Precision score of the train data is 57.1%
- Recall score of the test data is 76.4%
- Recall score of the train data is 75.6%
- Ada Boost classifier is 75.6%
- ROC curve for the random forest



6. IMPLEMENTATION

Technologies and Packages:

- Python idle 3.7 version
- Anaconda 3.7
- Jupyter Notebook
- Google colab

Python

Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library.

Python is Interactive − you can actually sit at a Python prompt and interact with the interpreter directly to write your programs. Python also acknowledges that speed of development is important. Readable and terse code is part of this, and so is access to powerful constructs that avoid tedious repetition of code. Maintainability also ties into this may be an all but useless metric, but it does say something about how much code you have to scan, read and/or understand to troubleshoot problems or tweak behaviours. This speed of development, the easewith which a programmer of other languages can pick up basic Python skills and the huge standard library is key to another area where Python excels. All its tools have been quick to implement, saved a lot of time, and several of them have later been patched and updated by people with no Python background - without breaking.

Modules Used in Project: -

NumPy

NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays.

It is the fundamental package for scientific computing with Python. It contains various features including these important ones:

- A powerful N-dimensional array object
- Sophisticated (broadcasting) functions
- Tools for integrating C/C++ and Fortran code
- Useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined using NumPy which allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

Pandas

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data load, prepare, manipulate, model, and analyze. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

Matplotlib

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter Notebook, web application servers, and four graphical user interface toolkits. Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, error charts; scatter plots, etc., with just a few lines of code. For examples, see the sample plots and thumbnail gallery.

For simple plotting the pyplot module provides a MATLAB-like interface, particularly when combined with IPython. For the power user, you have full control of line styles, font properties, axes properties, etc, via an object oriented interface or via a set of functions familiar to MATLAB users.

Scikit – learn

Scikit-learn provide a range of supervised and unsupervised learning algorithms via a consistent interface in Python. It is licensed under a permissive simplified BSD license and is distributed under many Linux distributions, encouraging academic and commercial use.

DATA SET

The dataset represents 10 years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks. It includes over 50 features representing patient and hospital outcomes. (https://github.com/pmacinec/diabetes-patients-readmissions-prediction.git)

	encounter_id	patient_nbr	race	gender	age	weight	admission_type_id	discharge_disposition_id	admission_source_id	time_in_hospital	...
0	2278392	8222157	Caucasian	Female	[0-10)	?	6	25	1	1	...
1	149190	55629189	Caucasian	Female	[10-20)	?	1	1	7	3	...
2	64410	86047875	AfricanAmerican	Female	[20-30)	?	1	1	7	2	...
3	500364	82442376	Caucasian	Male	[30-40)	?	1	1	7	2	...
4	16680	42519267	Caucasian	Male	[40-50)	?	1	1	7	1	...
5	35754	82637451	Caucasian	Male	[50-60)	?	2	1	2	3	...
6	55842	84259809	Caucasian	Male	[60-70)	?	3	1	2	4	...
7	63768	114882984	Caucasian	Male	[70-80)	?	1	1	7	5	...
8	12522	48330783	Caucasian	Female	[80-90)	?	2	1	4	13	...
9	15738	63555939	Caucasian	Female	[90-100)	?	3	3	4	12	...
10	28236	89869032	AfricanAmerican	Female	[40-50)	?	1	1	7	9	...
11	36900	77391171	AfricanAmerican	Male	[60-70)	?	2	1	4	7	...
12	40926	85504905	Caucasian	Female	[40-50)	?	1	3	7	7	...
13	42570	77586282	Caucasian	Male	[80-90)	?	1	6	7	10	...
14	62256	49726791	AfricanAmerican	Female	[60-70)	?	3	1	2	1	...

15 rows × 50 columns

Sample Code:

```
In [1]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import time
```

```
In [2]: df = pd.read_csv('diabetic_data.csv')
```

```
In [3]: print('Number of samples:',len(df))
```

```
Number of samples: 101766
```

```
In [4]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 101766 entries, 0 to 101765
Data columns (total 50 columns):
 #   Column                    Non-Null Count   Dtype
---  ------                    --------------   -----
 0   encounter_id              101766 non-null  int64
 1   patient_nbr               101766 non-null  int64
 2   race                      101766 non-null  object
 3   gender                    101766 non-null  object
 4   age                       101766 non-null  object
 5   weight                    101766 non-null  object
 6   admission_type_id         101766 non-null  int64
 7   discharge_disposition_id  101766 non-null  int64
 8   admission_source_id       101766 non-null  int64
 9   time_in_hospital          101766 non-null  int64
 10  payer_code                101766 non-null  object
 11  medical_specialty         101766 non-null  object
 12  num_lab_procedures        101766 non-null  int64
 13  num_procedures            101766 non-null  int64
 14  num_medications           101766 non-null  int64
 15  number_outpatient         101766 non-null  int64
 16  number_emergency          101766 non-null  int64
 17  number_inpatient          101766 non-null  int64
 18  diag_1                    101766 non-null  object
 19  diag_2                    101766 non-null  object
 20  diag_3                    101766 non-null  object
 21  number_diagnoses          101766 non-null  int64
 22  max_glu_serum             101766 non-null  object
 23  A1Cresult                 101766 non-null  object
 24  metformin                 101766 non-null  object
 25  repaglinide               101766 non-null  object
 26  nateglinide               101766 non-null  object
 27  chlorpropamide            101766 non-null  object
 28  glimepiride               101766 non-null  object
 29  acetohexamide             101766 non-null  object
 30  glipizide                 101766 non-null  object
 31  glyburide                 101766 non-null  object
 32  tolbutamide               101766 non-null  object
 33  pioglitazone              101766 non-null  object
 34  rosiglitazone             101766 non-null  object
 35  acarbose                  101766 non-null  object
 36  miglitol                  101766 non-null  object
 37  troglitazone              101766 non-null  object
 38  tolazamide                101766 non-null  object
 39  examide                   101766 non-null  object
 40  citoglipton               101766 non-null  object
 41  insulin                   101766 non-null  object
 42  glyburide-metformin       101766 non-null  object
 43  glipizide-metformin       101766 non-null  object
 44  glimepiride-pioglitazone  101766 non-null  object
 45  metformin-rosiglitazone   101766 non-null  object
 46  metformin-pioglitazone    101766 non-null  object
 47  change                    101766 non-null  object
 48  diabetesMed               101766 non-null  object
 49  readmitted                101766 non-null  object
dtypes: int64(13), object(37)
memory usage: 38.9+ MB
```

In [5]:
```
df.head()
```

Out[5]:

	encounter_id	patient_nbr	race	gender	age	weight	admission_type_id	discharge_disposition_id	admission_source_id	time_in_hospital	...	co
0	2278392	8222157	Caucasian	Female	[0-10)	?	6	25	1	1	...	
1	149190	55629189	Caucasian	Female	[10-20)	?	1	1	7	3	...	
2	64410	86047875	AfricanAmerican	Female	[20-30)	?	1	1	7	2	...	
3	500364	82442376	Caucasian	Male	[30-40)	?	1	1	7	2	...	
4	16680	42519267	Caucasian	Male	[40-50)	?	1	1	7	1	...	

5 rows × 50 columns

In [14]:
```
# For each column
for c in list(df.columns):

    # get a list of unique values
    n = df[c].unique()

    # if number of unique values is less than 30, print the values. Otherwise print the number of unique values
    if len(n)<30:
        print(c)
        print(n)
    else:
        print(c + ': ' +str(len(n)) + ' unique values')
```

```
encounter_id: 99343 unique values
patient_nbr: 69990 unique values
race
['Caucasian' 'AfricanAmerican' '?' 'Other' 'Asian' 'Hispanic']
gender
['Female' 'Male' 'Unknown/Invalid']
age
['[0-10)' '[10-20)' '[20-30)' '[30-40)' '[40-50)' '[50-60)' '[60-70)'
 '[70-80)' '[80-90)' '[90-100)']
weight
['?' '[75-100)' '[50-75)' '[0-25)' '[100-125)' '[25-50)' '[125-150)'
 '[175-200)' '[150-175)' '>200']
admission_type_id
[6 1 2 5 4 3 8 7]
discharge_disposition_id
[25  1  3  6  2  5  7 10  4 18  8 12 16 17 22 19  9 13 24 28 27]
admission_source_id
[ 1  7  2  4  5 20  6  8 17  9  3 14 10 22 11 25 13]
time_in_hospital
[ 1  3  2  4  5 13 12  9  7 10  8 11  6 14]
payer_code
['?' 'MC' 'MD' 'HM' 'UN' 'BC' 'SP' 'CP' 'SI' 'DM' 'CM' 'CH' 'PO' 'WC' 'OT'
 'OG' 'MP' 'FR']
medical_specialty: 73 unique values
num_lab_procedures: 118 unique values
num_procedures
[0 5 1 3 2 4 6]
num_medications: 75 unique values
number_outpatient: 39 unique values
number_emergency: 33 unique values
number_inpatient
[ 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 21 19 20]
diag_1: 716 unique values
diag_2: 749 unique values
diag_3: 787 unique values
number_diagnoses
[ 1  9  8  7  6  5  4  3 16 13 12 15 10 11 14]
max_glu_serum
['None' '>200' 'Norm' '>300']
A1Cresult
['None' '>7' '>8' 'Norm']
metformin
['No' 'Steady' 'Up' 'Down']
repaglinide
['No' 'Up' 'Steady' 'Down']
nateglinide
['No' 'Steady' 'Down' 'Up']
```

In [34]:
```
print('Total number of features:', len(cols_num + cols_all_cat + cols_extra))
print('Numerical Features:',len(cols_num))
print('Categorical Features:',len(cols_all_cat))
print('Extra features:',len(cols_extra))
```

```
Total number of features: 143
Numerical Features: 8
Categorical Features: 133
Extra features: 2
```

In [37]:
```
# shuffle the samples
df_data = df_data.sample(n = len(df_data), random_state = 42)
df_data = df_data.reset_index(drop = True)
```

In [38]:
```
# Save 30% of the data as validation and test data
df_valid_test=df_data.sample(frac=0.30,random_state=42)
print('Split size: %.3f'%(len(df_valid_test)/len(df_data)))
```

```
Split size: 0.300
```

In [39]:
```
df_test = df_valid_test.sample(frac = 0.5, random_state = 42)
df_valid = df_valid_test.drop(df_test.index)
```

In [40]:
```
# use the rest of the data as training data
df_train_all=df_data.drop(df_valid_test.index)
```

In [41]:
```
# split the training data into positive and negative
rows_pos = df_train_all.OUTPUT_LABEL == 1
df_train_pos = df_train_all.loc[rows_pos]
df_train_neg = df_train_all.loc[~rows_pos]

# merge the balanced data
df_train = pd.concat([df_train_pos, df_train_neg.sample(n = len(df_train_pos), random_state = 42)],axis = 0)

# shuffle the order of training samples
df_train = df_train.sample(n = len(df_train), random_state = 42).reset_index(drop = True)

print('Train balanced prevalence(n = %d):%.3f'%(len(df_train), calc_prevalence(df_train.OUTPUT_LABEL.values)))
```

```
Train balanced prevalence(n = 13766):0.500
```

In [42]:
```
df_train_all.to_csv('df_train_all.csv',index=False)
df_train.to_csv('df_train.csv',index=False)
df_valid.to_csv('df_valid.csv',index=False)
df_test.to_csv('df_test.csv',index=False)
```

In [47]:
```
from sklearn.metrics import roc_auc_score, accuracy_score, precision_score, recall_score
def calc_specificity(y_actual, y_pred, thresh):
    # calculates specificity
    return sum((y_pred < thresh) & (y_actual == 0)) /sum(y_actual ==0)

def print_report(y_actual, y_pred, thresh):


    accuracy = accuracy_score(y_actual, (y_pred > thresh))
    recall = recall_score(y_actual, (y_pred > thresh))
    precision = precision_score(y_actual, (y_pred > thresh))


    print('accuracy:%.3f'%accuracy)
    print('recall:%.3f'%recall)
    print('precision:%.3f'%precision)


    print(' ')
    return accuracy, recall, precision
```

In [50]:
```
# k-nearest neighbors
from sklearn.neighbors import KNeighborsClassifier
knn=KNeighborsClassifier(n_neighbors = 100)
knn.fit(X_train_tf, y_train)
```

Out[50]:
```
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                     metric_params=None, n_jobs=None, n_neighbors=100, p=2,
                     weights='uniform')
```

In [51]:
```
y_train_preds = knn.predict_proba(X_train_tf)[:,1]
y_test_preds = knn.predict_proba(X_test_tf)[:,1]

print('KNN')
print('Training:')
knn_train_accuracy, knn_train_recall, \
    knn_train_precision = print_report(y_train,y_train_preds, thresh)
print('Test:')
knn_test_accuracy, knn_test_recall, \
    knn_test_precision = print_report(y_test,y_test_preds, thresh)
```

```
KNN
Training:
accuracy:0.603
recall:0.491
precision:0.633

Test:
accuracy:0.661
recall:0.469
precision:0.166
```

In [54]:
```
#Random Forest Classifier
from sklearn.ensemble import RandomForestClassifier
rf=RandomForestClassifier(max_depth = 6, random_state = 42)
rf.fit(X_train_tf, y_train)
```

Out[54]:
```
RandomForestClassifier(bootstrap=True, ccp_alpha=0.0, class_weight=None,
                       criterion='gini', max_depth=6, max_features='auto',
                       max_leaf_nodes=None, max_samples=None,
                       min_impurity_decrease=0.0, min_impurity_split=None,
                       min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, n_estimators=100,
                       n_jobs=None, oob_score=False, random_state=42, verbose=0,
                       warm_start=False)
```

In [55]:
```
y_train_preds = rf.predict_proba(X_train_tf)[:,1]
y_test_preds = rf.predict_proba(X_test_tf)[:,1]

print('Random Forest')
print('Training:')
rf_train_accuracy, rf_train_recall, rf_train_precision = print_report(y_train,y_train_preds, thresh)
print('Test:')
rf_test_accuracy, rf_test_recall, rf_test_precision = print_report(y_test,y_test_preds, thresh)
```

```
Random Forest
Training:
accuracy:0.643
recall:0.612
precision:0.652

Test:
accuracy:0.626
recall:0.604
precision:0.177
```

7. EVALUATION ANALYSIS AND RESULTS

7.1 Evaluation Metrics:

Accuracy: The accuracy of a machine learning classification algorithm is one way to assess how often model classifies a data point correctly. The numerator is total number of predictions that were correct. The denominator is the total number of predictions.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

Precision: It attempts to answers the question: What proportion of positive identifications was actually correct?

$$\text{Recall} = \frac{TP}{TP + FN}$$

Recall: It attempts to answer the question: What proportion of actual positives was identified correctly?

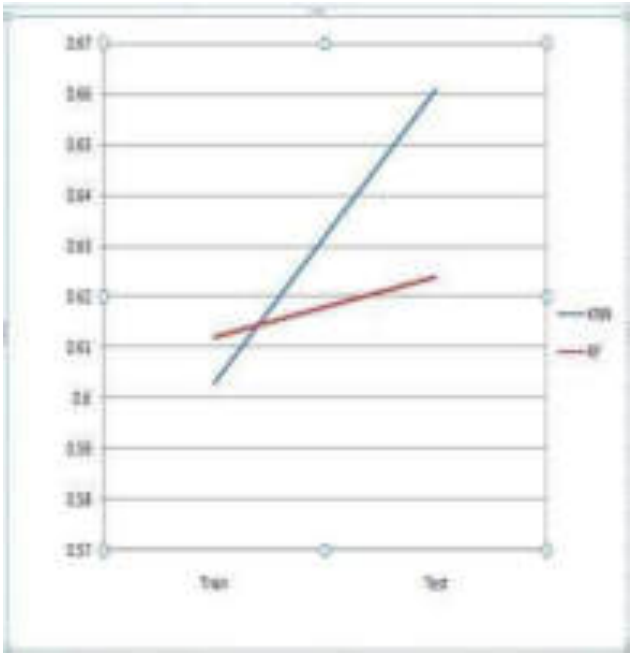
$$\text{Recall} = \frac{TP}{TP + FN}$$

Train data = 70

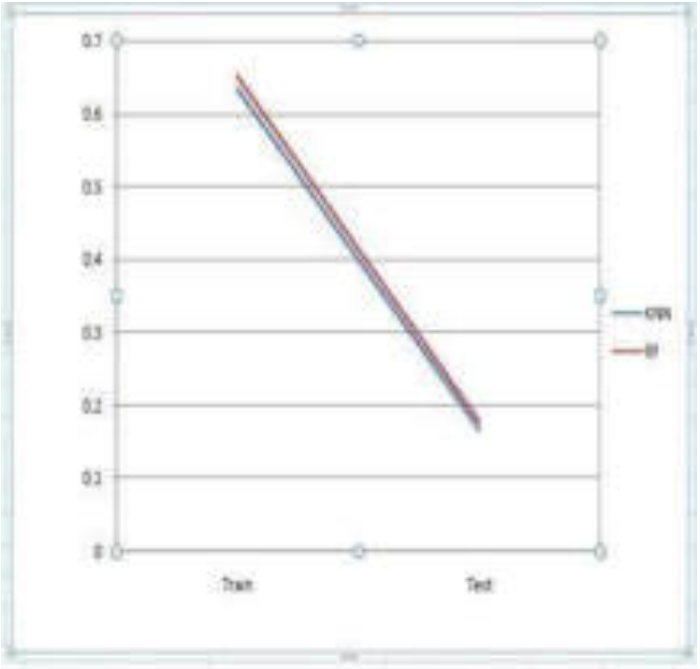
Test data = 30

Algorithm	Accuracy		Recall		Precision	
	Train	Test	Train	Test	Train	Test
KNN	0.603	0.661	0.491	0.469	0.633	0.166
RandomForest	0.643	0.624	0.612	0.604	0.652	0.177

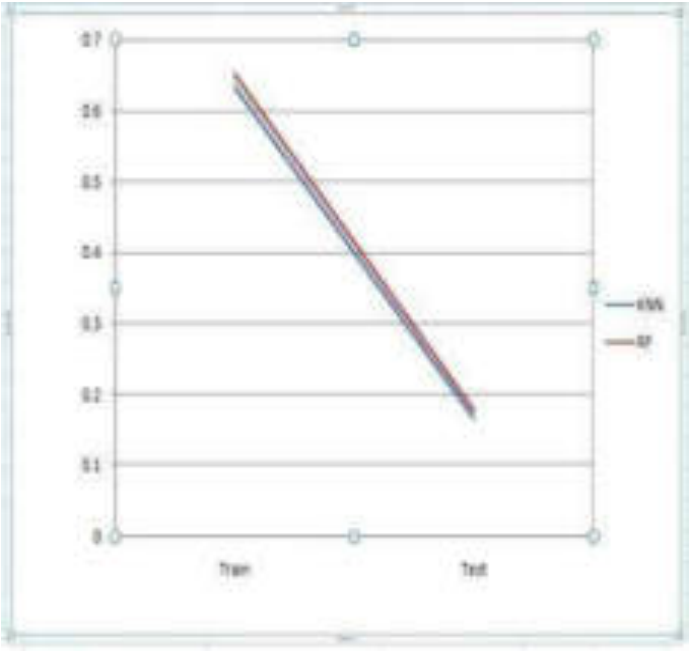
Accuracy:



Recall:



Precision:



RESULTS:

```
In [34]: #Random Forest Classifier
         from sklearn.ensemble import RandomForestClassifier
         rf=RandomForestClassifier(max_depth = 6, random_state = 42)
         rf.fit(X_train_tf, y_train)
```

```
Out[34]: RandomForestClassifier(bootstrap=True, ccp_alpha=0.0, class_weight=None,
                                criterion='gini', max_depth=6, max_features='auto',
                                max_leaf_nodes=None, max_samples=None,
                                min_impurity_decrease=0.0, min_impurity_split=None,
                                min_samples_leaf=1, min_samples_split=2,
                                min_weight_fraction_leaf=0.0, n_estimators=100,
                                n_jobs=None, oob_score=False, random_state=42, verbose=0,
                                warm_start=False)
```

```
In [35]: y_train_preds = rf.predict_proba(X_train_tf)[:,1]
         y_test_preds = rf.predict_proba(X_test_tf)[:,1]

         print('Random Forest')
         print('Training:')
         rf_train_accuracy, rf_train_recall, rf_train_precision = print_report(y_train,y_train_preds, thresh)
         print('Test:')
         rf_test_accuracy, rf_test_recall, rf_test_precision = print_report(y_test,y_test_preds, thresh)
```

```
Random Forest
Training:
accuracy:0.643
recall:0.612
precision:0.652

Test:
accuracy:0.624
recall:0.604
precision:0.177
```

```
In [36]: # K-nearest neighbors
         from sklearn.neighbors import KNeighborsClassifier
         knn=KNeighborsClassifier(n_neighbors = 100)
         knn.fit(X_train_tf, y_train)
```

```
Out[36]: KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                              metric_params=None, n_jobs=None, n_neighbors=100, p=2,
                              weights='uniform')
```

```
In [37]: y_train_preds = knn.predict_proba(X_train_tf)[:,1]
         y_test_preds = knn.predict_proba(X_test_tf)[:,1]

         print('KNN')
         print('Training:')
         knn_train_accuracy, knn_train_recall, \
             knn_train_precision = print_report(y_train,y_train_preds, thresh)
         print('Test:')
         knn_test_accuracy, knn_test_recall, \
             knn_test_precision = print_report(y_test,y_test_preds, thresh)
```

```
KNN
Training:
accuracy:0.603
recall:0.491
precision:0.633

Test:
accuracy:0.661
recall:0.469
precision:0.166
```

8. CONCLUSION

30-day hospital readmission of diabetes patients is of prime importance for health centers and is found very stressful due to the current models limit in term of performance and generalizability. To cope with this challenge, this study implemented a comprehensive pre-processing framework in order to improve the initial data quality, hence empowering the model's efficiency. The suggested pre-processing framework included comprehensive data cleaning, data reduction and transformation aiming at better optimizing and selecting prominent features for 30-day unplanned readmission among diabetes patients. More specifically, the performance of the designed model was found robust, scalable on trained data and with particularly balanced across different metrics and achieved a high score on all evaluated metrics, optimal performance.

Finally, just predicting if someone will get a disease within a certain period of time might not be good enough. Predicting when the disease is likely to occur can facilitate more targeted and effective preventive care and early treatment. The task of predicting when an event will occur is generally known as survival analysis. Developing survival analysis models using the same type of medical code sequence data considered in this work would be a valuable next step.

9. REFERENCES

1. Axon RN, Williams MV. Hospital readmission as an accountability measure. JAMA. 2011;305(5):504–5.

2. Stone J, Hoffman, G. Medicare Hospital Readmissions: Issues, Policy Options and PPACA. Vol 7–5700, R40972: Congressional Research Service, Penny Hill Press, 2010.

3. Epstein AM. Revisiting readmissions—changing the incentives for shared accountability. N Engl J Med. 2009;360(14):1457–9.

4. Cook CB, Seifert KM, Hull BP, Hovan MJ, Charles JC, Miller-Cage V, et al. Inpatient to outpatient transfer of diabetes care: planning for an effective hospital discharge. Endocr Pract. 2009;15(3):263–9.

5. Draznin B, Gilden J, Golden SH, Inzucchi SE, Baldwin D, Bode BW, et al. Pathways to quality inpatient management of hyperglycemia and diabetes: a call to action. Diabetes Care. 2013;36(7):1807–14.

6. Dungan KM. The effect of diabetes on hospital readmissions. J Diab Sci Technol. 2012;6(5):1045–52.

7. Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care: advantages and limitations. Arch Intern Med. 2000;160(8):1074–81.

8. Umpierrez GE, Isaacs SD, Bazargan N, You X, Thaler LM, Kitabchi AE. Hyperglycemia: an independent marker of in-hospital mortality in patients with undiagnosed diabetes. J Clin Endocrinol Metab. 2002;87(3):978–82.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"Driver Drowsiness Detection System with OpenCV and Keras"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
COMPUTER SCIENCE AND ENGINEERING

Submitted by

G.Divya (17H71A0511)

CH.Sowmya (17H71A0544)

T.Divya Sree (17H71A0512)

K.Pradeep (17H71A0532)

Under the Esteemed Guidance of
Mrs.A.Aunradha

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
Kanchikacherla - 521 180. Phone : 08678-273535
(Approved by AICTE & Affiliated to JNTU-Kakinada)

2020-2021

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
Kanchikacherla - 521 180. Phone : 08678-273535
(Approved by AICTE & Affiliated to JNTU-Kakinada)

2020-2021

CERTIFICATE

This is to certify that the Main Project entitled **"Driver Drowsiness Detection System with OpenCV and Keras"** is a bonafide work carried out by **G.Divya (17H71A0511), CH.Sowmya (17H71A0544), T.Divya Sree (17H71A0512), K.Pradeep (17H71A0532)** in partial fulfilment for the award of degree of Bachelor of Technology in **Computer Science and Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

Mrs. A. Anuradha
Project Guide

Mr. D. Prasad
Head of the Department

Examiner

Dr. K. Srinivas
PRINCIPAL

ACKNOWLEDGEMENT

We would like to take this opportunity to express my deepest appreciation to the following people for their valuable contributions and assistance with this Project.

Initially, we would like to thank my project supervisor, **Mrs. A. Anuradha, Assistant Professor, Department of Computer Science and Engineering** for the guidance and support, especially for the valuable ideas and knowledge shared to me throughout the Project.

We have the immense pleasure in expressing my thanks and deep sense of gratitude to**, Mr. D. Prasad, Head of the Department, Computer Science and Engineering** for extending necessary facilities and support for the completion of the Project.

We whole sincerely acknowledge **Dr. K. Srinivas, Principal and Prof D. PANDURANGA RAO, CEO** for giving opportunity to take up the Project work. We also extend my thanks to all faculty members of **Computer Science and Engineering**, for their valuable guidance and encouragement.

We would like to extend my warm appreciation to all my friends for sharing us their knowledge, valuable contributions and help with this Project.

Finally, my special thanks go to my family for their continuous support and help throughout my academic years and for their continual support and encouragement for the completion of the project.

DECLARATION

We **G.Divya, CH. Sowmya, T.Divya Seer, K.Pradeep** of the Main-Project "**Driver Drowsiness Detection System with OpenCV and Keras**", hereby declare that the matter embodied in this Project is the genuine work done by us and has not been submitted either to this university or to any other university/institute for the fulfilment of the requirement of any course of study.

G.DIVYA (17H71A0511)
CH.SOWMYA (17H71A0544)
T.DIVYA SREE (17H71A0512)
K.PRADEEP (17H71A0532)

ABSTRACT

This document is a review report on the research conducted and the project made in the field of computer engineering to develop a system for driver drowsiness detection to prevent accidents from happening because of driver fatigue and sleepiness. The report proposed the results and solutions on the limited implementation of the various techniques that are introduced in the project. Whereas the implementation of the project give the real world idea of how the system works and what changes can be done in order to improve the utility of the overall system.

Furthermore, the paper states the overview of the observations made by the authors in order to help further optimization in the mentioned field to achieve the utility at a better efficiency for a safer road.

LIST OF FIGURES

1. INTRODUCTION

PURPOSE

HUMAN PSYCHOLOGY WITH CURRENT TECHNOLOGY

Humans have always invented machines and devised techniques to ease and protect their lives, for mundane activities like traveling to work, or for more interesting purposes like aircraft travel. With the advancement in technology, modes of transportation kept on advancing and our dependency on it started increasing exponentially. It has greatly affected our lives as we know it. Now, we can travel to places at a pace that even our grandparents wouldn't have thought possible. In modern times, almost everyone in this world uses some sort of transportation every day. Some people are rich enough to have their own vehicles while others use public transportation. However, there are some rules and codes of conduct for those who drive irrespective of their social status. One of them is staying alert and active while driving.

Neglecting our duties towards safer travel has enabled hundreds of thousands of tragedies to get associated with this wonderful invention every year. It may seem like a trivial thing to most folks but following rules and regulations on the road is of utmost importance. While on road, an automobile wields the most power and in irresponsible hands, it can be destructive and sometimes, that carelessness can harm lives even of the people on the road. One kind of carelessness is not admitting when we are too tired to drive. In order to monitor and prevent a destructive outcome from such negligence, many researchers have written research papers on driver drowsiness detection systems. But at times, some of the points and observations made by the system are not accurate enough. Hence, to provide data and another perspective on the problem at hand, in order to improve their implementations and to further optimize the solution, this project has been done.

FACTS & STATISTICS

Our current statistics reveal that just in 2015 in India alone, 148,707 people died due to car related accidents. Of these, at least 21 percent were caused due to fatigue causing drivers to make mistakes. This can be a relatively smaller number still, as among the multiple causes that can lead to an accident, the involvement of fatigue as a cause is generally grossly underestimated. Fatigue combined with bad infrastructure in developing countries like India is a recipe for disaster. Fatigue, in general, is very difficult to measure or observe unlike alcohol and drugs, which have clear key indicators and tests that are available easily. Probably, the best solutions to this problem are awareness about fatigue-related accidents and promoting drivers to admit fatigue when needed. The former is hard and much more expensive to achieve, and the latter is not possible without the former as driving for long hours is very lucrative. When there is an increased need for a job, the wages associated with it increases leading to more and more people adopting it. Such is the case for driving transport vehicles at night. Money motivates drivers to make unwise decisions like driving all night even with fatigue. This is mainly because the drivers are not themselves aware of the huge risk associated with driving when fatigued. Some countries have imposed restrictions on the number of hours a driver can drive at a stretch, but it is still not enough to solve this problem as its implementation is very difficult and costly.

INTENDED AUDIENCE

The intended audience for this document is the development team, the project evaluation jury, and other tech-savvy enthusiasts who wish to further work on the project.

PRODUCT SCOPE

There are many products out there that provide the measure of fatigue level in the drivers which are implemented in many vehicles. The driver drowsiness detection system provides the similar functionality but with better results and additional benefits. Also, it alerts the user on reaching a certain saturation point of the drowsiness measure.

PROBLEM DEFINITION

Fatigue is a safety problem that has not yet been deeply tackled by any country in the world mainly because of its nature. Fatigue, in general, is very difficult to measure or observe unlike alcohol and drugs, which have clear key indicators and tests that are available easily. Probably, the best solutions to this problem are awareness about fatigue-related accidents and promoting drivers to admit fatigue when needed. The former is hard and much more expensive to achieve, and the latter is not possible without the former as driving for long hours is very lucrative.

2. LITERATURE SURVEY

Driver Fatigue is among the most common reason for fatal road accidents around the world. This shows that in the transportation industry especially, where a driver of a heavy vehicle is often exposed to hours of monotonous driving which causes fatigue without frequent rest period. Due to the frequent occurrence of driver fatigue this has become an area of great socio economic concern. Detecting driver's drowsiness has been a research topic for many years, with many approaches being studied thus so far. The work presented in takes advantage of some mouth geometrical features to detect yawning. The work in proposes the detection of the face region using the difference image between two images. Driver's yawn is then detected based on the distance between the midpoint of nostrils and the chin Uses Gravity-center template to detect the face. It then uses grey projection and Gabor wavelets to detect the mouth corners. Finally LDA is applied to classify feature vectors to detect yawning It presents a system where the face is located through Viola-Jones face detection method in a video frame. Then, a mouth window is extracted from the face region, in which lips are searched through spatial fuzzy means(s- FCM) clustering .In there is an advantage of two cameras: a low resolution camera for the face and a high resolution one for the mouth. It then uses haar-like features to detect driver's mouth and yawning is detected by the ratio of mouth height and width. In a method is adopted for yawning detection based on the changes in mouth geometric features. The work in [10] driver's drowsiness is determined using vehicle based measures, behavioral measures and psychological measures which makes this a hybrid drowsiness detection system. [11] Shows detection of drowsiness based on head movement and geometrical features of mouth is proposed. Experiment was conducted on sample size of 50 video clips and observed that head movement contributes about 8% and yawning contributes about49%.

E. Makinen et al. Presented a system which classifies the detected and aligned face images based on the gender. The paper concluded that manual alignment method provides better classification rates as compared to that of automatic alignment method. One of the findings was that different input image sizes did not affect the classification accuracy rates. S. U. Rehman et al. [2] presented a new architecture for face image classification named unsupervised CNN.CNN is required where a single CNN handles multitask (i.e. Facial detection and emotional classification) by merging CNN with other modules/algorithms.Hence,CNN is used as an algorithm for the project.

3. SYSTEM ANALYSIS

EXISTING SYSTEM

By using a non-intrusive machine vision based concepts, drowsiness of the driver detected system is developed. Many Existing System require a camera which is installed infront of driver. It points straight towards the face of the driver and monitors the driver's eye, mouth in order to identify the drowsiness. For large vehicle such as heavy trucks and buses this arrangement is not pertinent. Bus has large front glass window to have a broad view for safe driving. If we place camera on the window of front glass, the camera blocks the frontal view of driver so it is not practical

Disadvantages of Existing System

- Existing System has two cameras in the system one for monitoring the head movement and the other one for facial expressions.
- The other disadvantage is aging of sensors and all these sensors are attached to the driver's body which may affect the driver

PROPOSED SYSTEM

For detection of drowsiness the per closure value of eye, yawning status and head tilt position are considered. So when the closure of eyes, mouth open or head tilt exceeds a certain predefined thresholds then the driver is identified to be sleepy, and alert the driver and control the vehicle speed along with indicators. The proposed system also send the driver status to authorized persons through IoT.

Advantages of proposed System

- It is an efficient system to identify user attentiveness based on fatigue detection.

OBJECTIVES

Driver drowsiness detection is a car safety technology which spares the life of the driver by avoiding mishaps when the driver is getting languid.

- The primary goal is to initially plan a framework to distinguish driver's sluggishness by persistently checking retina of the eye.
- The framework works disregarding driver wearing displays and in different lighting conditions
- To caution the driver on the identification of laziness by utilizing ringer or alert.
- Speed of the vehicle can be reduced.
- Traffic management can be maintained by reducing the accidents.

4. REQUIREMENT ANALYSIS

PRELIMINARYINVESTIGATION

The first and foremost strategy for development of a project starts from the thought of designing a mail enabled platform for a small firm in which it is easy and convenient of sending and receiving messages, there is a search engine, address book and also including some entertaining games. When it is approved by the organization and our project guide the first activity, i.e. preliminary investigation begins. The activity has three parts:

- Request Clarification
- Feasibility Study
- Request Approval

REQUEST CLARIFICATION

After the approval of the request to the organization and project guide, with an investigation being considered, the project request must be examined to determine precisely what the system requires.

Here our project is basically meant for users within the company whose systems can be interconnected by the Local Area Network(LAN). In today's busy schedule man need everything should be provided in a readymade manner. So taking into consideration of the vastly use of the net in day to day life, the corresponding development of the portal came into existence

FEASIBILITY ANALYSIS

An important outcome of preliminary investigation is the determination that the system request is feasible. This is possible only if it is feasible within limited resource and time. The different feasibilities that have to be analyzed are

- Operational Feasibility
- Economic Feasibility
- Technical Feasibility

Operational Feasibility

Operational Feasibility deals with the study of prospects of the system to be developed. This system operationally eliminates all the tensions of the admin and helps him in effectively tracking the project progress. This kind of automation will surely reduce the time and energy, which previously consumed in manual work. Based on the study, the system is proved to be operationally feasible

Economic Feasibility

Economic Feasibility or Cost-benefit is an assessment of the economic justification for a computer based project. As hardware was installed from the beginning & for lots of purposes thus the cost on project of hardware is low. Since the system is a network based, any number of employees connected to the LAN within that organization can use this tool from at any time. The Virtual Private Network is to be developed using the existing resources of the organization. So the project is economically feasible

Technical Feasibility

According to Roger S. Pressman, Technical Feasibility is the assessment of the technical resources of the organization. The organization needs IBM compatible machines with a graphical web browser connected to the Internet and Intranet. The system is developed for platform independent environment. Java Server Pages, JavaScript, HTML, SQL server and Web Logic Server are used to develop the system. The technical feasibility has been carried out. The system is technically feasible for development and can be developed with the existing facility.

FUNCTIONALREQUIREMENTS

In software engineering, a functional requirement defines a function of a software system or its component. A function is described as a set of inputs, the behavior, and outputs (see also software). Functional requirements may be calculations, technical details, data manipulation and processing and other specific functionality that define what a system is supposed to accomplish. Behavioral requirements describing all the cases where the system uses the functional requirements are captured in use cases. Generally, functional requirements

are expressed in the form **system shall do <requirement>.** The plan for implementing functional requirements is detailed in the system design. In requirements engineering, functional requirements specify particular results of a system. Functional requirements drive the application architecture of a system. A requirements analyst generates use cases after gathering and validating a set of functional requirements. The hierarchy of functional requirements is: **user/stakeholder request -> feature -> use case -> business rule.**

Functional requirements drive the application architecture of a system. A requirements analyst generates use cases after gathering and validating a set of functional requirements. Functional requirements may be technical details, data manipulation and other specific functionality of the project is to provide the information to the user.

NON FUNCTIONAL REQUIREMENTS

In systems engineering and requirements engineering, a non-functional requirement is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific behaviors

The project non-functional requirements include the following.

- Updating Work status
- .Problem resolution.
- Error occurrence in the system.
- Customer requests.

Availability: A system's "availability" or "uptime" is the amount of time that is operational and available for use. It's related to is the server providing the service to the users in displaying images. As our system will be used by thousands of users at any time our system must be available always. If there are any cases of updations they must be performed in a short interval of time without interrupting the normal services made available to the users.

Efficiency: Specifies how well the software utilizes scarce resources: CPU cycles, disk space, memory, bandwidth etc. All of the above mentioned resources can be effectively used by performing most of the validations at client side and reducing the workload on server by using

JSP instead of CGI which is being implemented now.

Flexibility: If the organization intends to increase or extend the functionality of the software after it is deployed, that should be planned from the beginning; it influences choices made 13 during the design, development, testing and deployment of the system. New modules can be easily integrated to our system without disturbing the existing modules or modifying the logical database schema of the existing applications.

Portability: Portability specifies the ease with which the software can be installed on all necessary platforms, and the platforms on which it is expected to run. By using appropriate server versions released for different platforms our project can be easily operated on any operating system, hence can be said highly portable.

Scalability: Software that is scalable has the ability to handle a wide variety of system configuration sizes. The nonfunctional requirements should specify the ways in which the system may be expected to scale up (by increasing hardware capacity, adding machines etc.). Our system can be easily expandable. Any additional requirements such as hardware or software which increase the performance of the system can be easily added. An additional server would be useful to speed up the application

Integrity: Integrity requirements define the security attributes of the system, restricting access to features or data to certain users and protecting the privacy of data entered into the software. Certain features access must be disabled to normal users such as adding the details of files, searching etc which is the sole responsibility of the server. Access can be disabled by providing appropriate logins to the users for only access.

Usability: Ease-of-use requirements address the factors that constitute the capacity of the software to be understood, learned, and used by its intended users. Hyperlinks will be provided for each and every service the system provides through which navigation will be easier. A system that has high usability coefficient makes the work of the user easier.

Performance: The performance constraints specify the timing characteristics of the software.

3.6 .1 SOFTWARE REQUIREMENTS

What things you need to install the software and how to install them:

1. Python3.6

   - This setup requires that your machine has python 3.6 installed on it. you can refertothisurlhttps://www.python.org/downloads/ to download python. Once you have python downloaded and installed, you will need to setup PATH variables (if you want to run python program directly, detail instructions are below in *how to run software section*). To do that check this: https://www.pythoncentral.io/add-python-to-path-python-is-not-recognized-as-an-internal-or-external-command/

   - Setting up PATH variable is optional as you can also run program without it and more instruction are given below on this topic.

2. Second and easier option is to download jupyter notebook on command prompt using the command pip install jupyter notebook

3. You will also need to download and install below packages after you install python and jupyter notebook from the steps above

Packages used are:

- Tensor flow
- Keras
- Opencv

.

- Use below commands in command prompt to install packages.

- Pip install tensor flow

  pip install numpy

  pip install seaborn

  pip install keras

Computer Vision

It is a field that includes processing analyzing and understanding image in general high dimensional data from the real world in order to produce numerical and symbolic information or it is a technology of science and machine that see it obtain information from images.

Deep Learning

Deep learning is a powerful set of techniques for learning using neural network. Neural network is beautiful biologically inspired program paradigm which enables a computer to learn from data. These are learning algorithms.

Neural Network

Neural networks reflect the behavior of the human brain, allowing computer programs to recognize patterns and solve common problems in the fields of AI, machine learning, and deep learning. Neural networks, also known as artificial neural networks (ANNs) or simulated neural networks(SNNs),are a subset of machine learning and are at the heart of deep learning algorithms. Their name and structure are inspired by the human brain, mimicking the way that biological neurons signal to one another.

The three important types of neural networks that form the basis for most pre-trained models in deep learning:

1. Artificial Neural Networks(ANN)
2. Convolution Neural Networks(CNN)
3. Recurrent Neural Networks(RNN)

Convolutional Neural Network

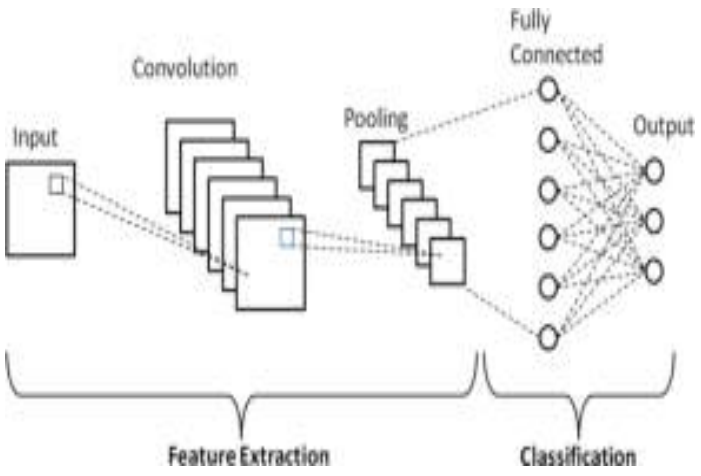
Pioneer of Convolutional Neural Network is Yann Lecun.He is the director of Facebook AI Research Group. He built the first convolutional Neural Network called LeNet in1988.LeNetwasusedfor character recognition tasks like reading zip codes and digits.

CNN also known as a ConvNet. A convolutional neural network is a feed-forward neural network that is generally used to analyze visual images by processing data with grid- like topology. A convolutional neural network is used to detect and classify objects in an image. In CNN, every image is represented in the form of an array of pixel values.

Layers in a Convolutional Neural Network

A convolutional neural network has multiple hidden layers that help in extracting information from an image. The four important layers in CNN are:

1. Convolution Layer
2. ReLU Layer
3. Pooling Layer
4. Fully Connected Layer



Input layer:

It accepts the pixels of the Image as input in the form of arrays.

Hidden Layers:

Carry out features extraction by performing certain calculations & manipulations. These are multiple hidden layers like convolution layer, ReLU layer, Pooling layer etc. that performs features extraction.

Convolution Layer:

This is the first step in the process of extracting valuable features from an image. This layer uses a matrix filter and performs convolution operation to detect patterns in the image. A convolution layer has several filters that perform the convolution operation. Every image is considered as a matrix of pixel values. Consider the following 5x5 image whose pixel values are either 0 or 1.There'salsoa filter matrix with a dimensionof3x3.

ReLU Layer:

ReLU stands for the rectified linear unit. Once the feature maps are extracted, the next step is to move them to a ReLU layer. ReLU performs an element wise operation and sets all the negativepixelsto0.Itintroducesnon-linearitytothenetwork,and the generated output is a rectified feature map. Its activation function is applied to the convolution layer to get a rectified feature map of the image.

Pooling Layer:

Pooling is a down-sampling operation that reduces the dimensionality of the feature map. The rectified feature map now goes through a pooling layer to generate a pooled feature map. Also uses multiple filters to detect edges, corners, eyes features, etc. Different kind of pooling operations can be applied out of which some are explained below.

Fully Connected Layer (FC):

Fully Connected layers in neural networks are those layers where all the inputs from one layer are connected to every activation unit of the next layer. The output from the convolutional layers represents high-level features in the data. While that output could be flattened and connected to the output layer, adding a fully-connected layer is a (usually)

Cheap way of learning non-linear combinations of these features. The fully connected layer operates on a flattened input where each input is connected to all neurons.

OpenCV

OpenCV is a cross-platform library using which we can develop real-time computer vision applications. It mainly focuses on image processing; video capture and analysis including features like face detection and object detection. All the OpenCV array structures are converted to and from Numpy arrays. This also makes it easier to integrate with other libraries that use Numpy such as SciPy and Matplotlib. In this, all OpenCV data types are preserved as such. Later, OpenCV came with both cv and cv2. Now, there in the late streleases,the, the is only the cv2 module, and cv is a sub class inside cv2 .You need to call import cv2.cv as cv to access it.

Tensor Flow

Tensor Flow is a mathematical computation library for training and building you machine learning and deep learning model with simple to use high level APIs. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML- powered applications. Tensor Flow was originally developed by researchers and engineers working on the Google Brain team within Google's Machine Intelligence Research organization to conduct machine learning and deep neural networks research. The system is general enough to be applicable in a wide variety of other domains, as well.

6. SYSTEMREQUIREMENTS

SOFTWAREREQUIREMENTS

- Software: Python IDLE
- Version:3.6
- Numpy, tensor flow, keras
- Operating system: windows

HARDWAREREQUIREMENTS

- Processor: IntelcoreI3
- Ram: 4GB
- Hard Disk:256GBorMore
- Monitor: 15 VGA Colour
- Web Camera
- Mouse: Logitech

7. MODULES DESCRIPTION

MODULES OF PROJECT

- **Web Capturing**
- **Face and eye Detection**
- **Drowsiness Detection(with Alert message and alarm sound)**

Web Capturing

Initializing the camera and capturing the frames containing 64 landmarks**.**

Face and eye Detection

This is the initialization stage of the system. Every time the system is started it needs to be setup and optimized for current user and conditions. The key step in this stage is success full head detection. If the driver's head is correctly located we can proceed to extract the features necessary for setting up the system. Setup steps include: (i) extracting driver's skin color and using that information to create custom skin color model and (ii) collecting a set of open/closed eyes samples, along with driver's normal head position.

To help achieve these goals, user interaction might be required. The driver might be asked to sit comfortably in its normal driving position so that system can determine upper and lower thresholds needed for detecting potential nodding. The driver might also be asked to hold their eyes closed and then open for a matter of few seconds each time. This is enough to get the system started. Over time, the system will expand the dataset of obtained images and will become more error resistant and overall more robust.

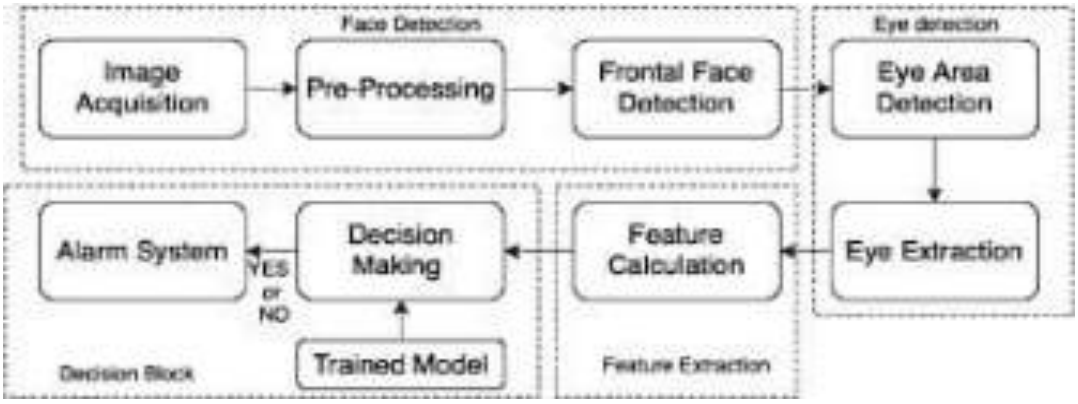
Drowsiness Detection

If the driver keeps his eyes closed for prolonged period of time or starts to nod, alertness has to be raised. The key step within this stage is close monitoring of drivers eyes. The system must determine whether the eyes are still closed, and what is the eyes 'position relative to previously established thresholds. We cannot afford to skip frames in this stage. In practice, tracking of eyes is performed much in the same way as in the tracking stage with the addition of the following processes: calculation of velocity and trajectory of the yes and threshold monitoring. These additional computations are required to improve the system's ability to determine het her the driver is drowsy or not.

Once it has been determined that the driver appears to be in a abnormal driving state, the system has to be proactive and alert the driver of potential dangers that an arise. Combination of audio/visual alerts are used to attract the driver's attention and raise their alertness level. Alerting has to be implemented in such a way as not to cause the opposite effect of intended and start let he driver into causing an accident

8. SYSTEM DESIGN

OVERVIEW OF SYSTEM DESIGN:



BLOCK DIAGRAM:

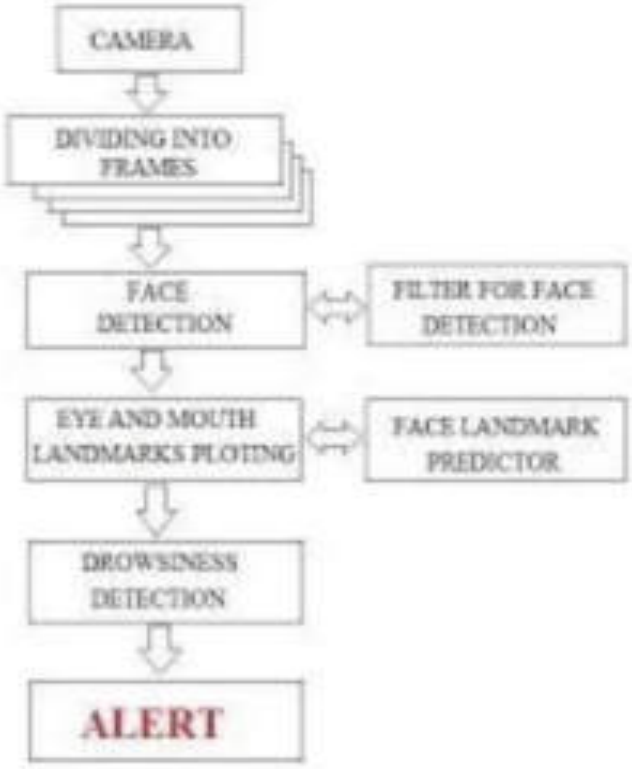


Fig: 8.1 Block Diagram

The driver's face is continuously recorded using a video camera that is installed under the front mirror. In order to detect the yawn, the first step is to detect and track the face using the series of frame shots taken by the camera. Then the location of the eyes is detected and the mouth in the detected face. The closed eye gesture is detected along with closed eyes for yawning detection. This makes segmentation procedure more robust to false detections. The mouth and eye geometrical features are then used to detect the yawn. The system will alert the driver of his fatigue by use of beep or buzzer and the improper driving situation in case of yawning detection. This is to be done in various phases given following:

- Phase 1: Face Detection
- Phase 2: Eyes Detection
- Phase 3: Mouth Detection
- Phase 4: Yawn Detection
- Phase5: Alert System

This is the initialization stage of the system. Every time the system is started it needs to be setup and optimized for current user and conditions. The key step in this stage is success full head detection. If the driver's head is correctly located we can proceed to extract the features necessary for setting up the system. Setup steps include: (i) extracting driver's skin color and using that information to create custom skin color model and (ii) collecting a set of open/closed eyes samples, along with driver's normal head position.

To help achieve these goals, user interaction might be required. The driver might be asked to sit comfortably in its normal driving position so that system can determine upper and lower thresholds needed for detecting potential nodding. The driver might also be asked to hold their eyes closed and then open for a matter of few seconds each time. This is enough to get the system started. Over time, the system will expand the dataset of obtained images and will become more error resistant and overall more robust

.

If the driver keeps his eyes closed for prolonged period of time or starts to nod, alertness has to be raised. The key step within this stage is close monitoring of drivers eyes. The system must determine whether the eyes are still closed, and what is the eyes 'position relative to previously established thresholds. We cannot afford to skip frames in this stage. In practice, tracking of eyes is performed much in the same way as in the tracking stage with the addition of the following processes: calculation of velocity and trajectory of the yes and threshold monitoring. These additional computations are required to improve the system's ability to determine het her the driver is drowsy or not.

Once it has been determined that the driver appears to be in a abnormal driving state, the system has to be proactive and alert the driver of potential dangers that an arise. Combination of audio/visual alerts are used to attract the driver's attention and raise their alertness level. Alerting has to be implemented in such a way as not to cause the opposite effect of intended and start let he driver into causing an accident

9. IMPLEMENTATION

In our program we used Dlib, a pre-trained program trained on the HELEN dataset to detect human faces using the pre-defined 68 landmarks.

Fig: 9.1 Landmarked Image of a person by Dlib

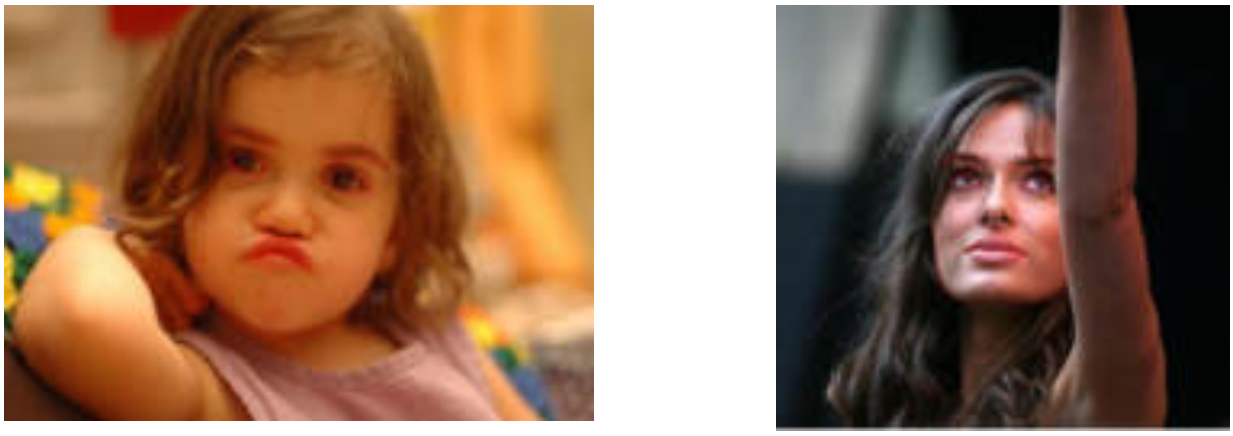


Fig: 9.2 HELENDATASET SAMPLES

- After passing our video feed to the dlib frame by frame, we are able to detect left eye and right eye features of the face.
- Now, we drew contours around it using OpenCV.

- Using Scipy's Euclidean function, we calculated sum of both eyes 'aspect ratio which is the sum of 2 distinct vertical distances between the eyelids divided by its horizontal distance.

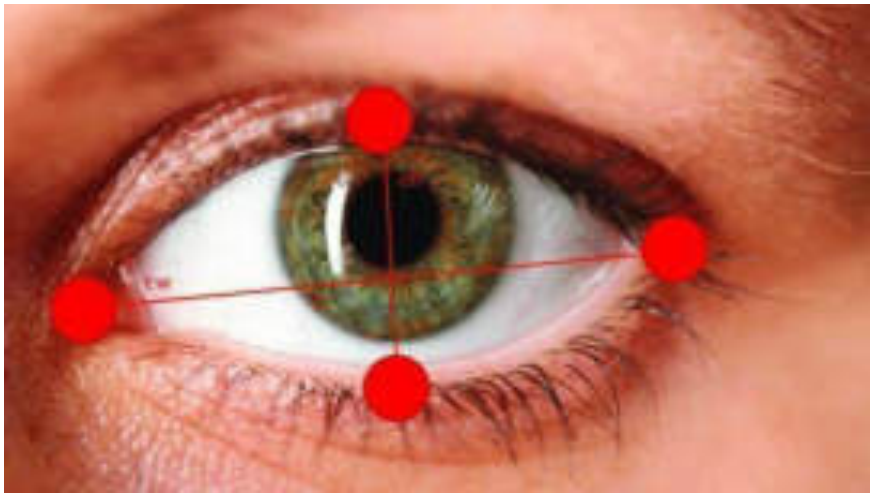


Fig:9.3 Eyes with horizontal and vertical distance marked for Eye Aspect Ratio calculation.

- Now we check if the aspect ratio value is less than 0.25 (0.25 was chosen as a base case after some tests). If it is less an alarm is sounded and user is warned.

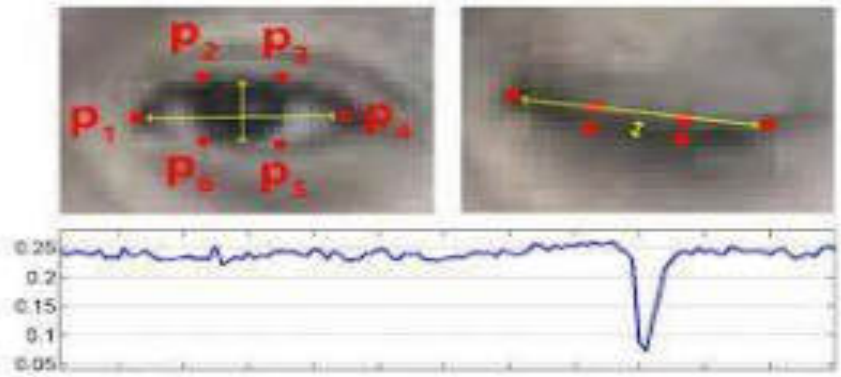
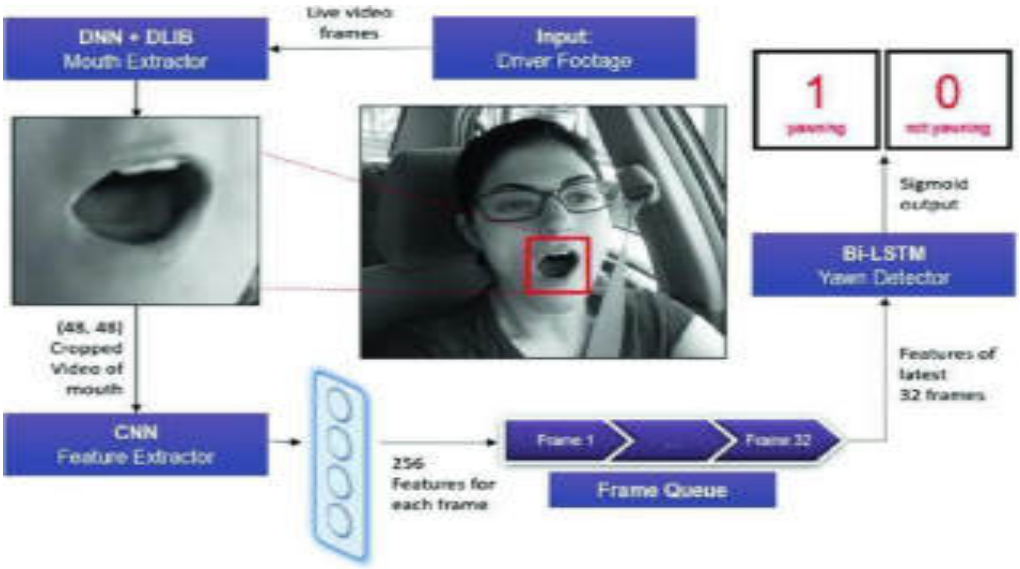
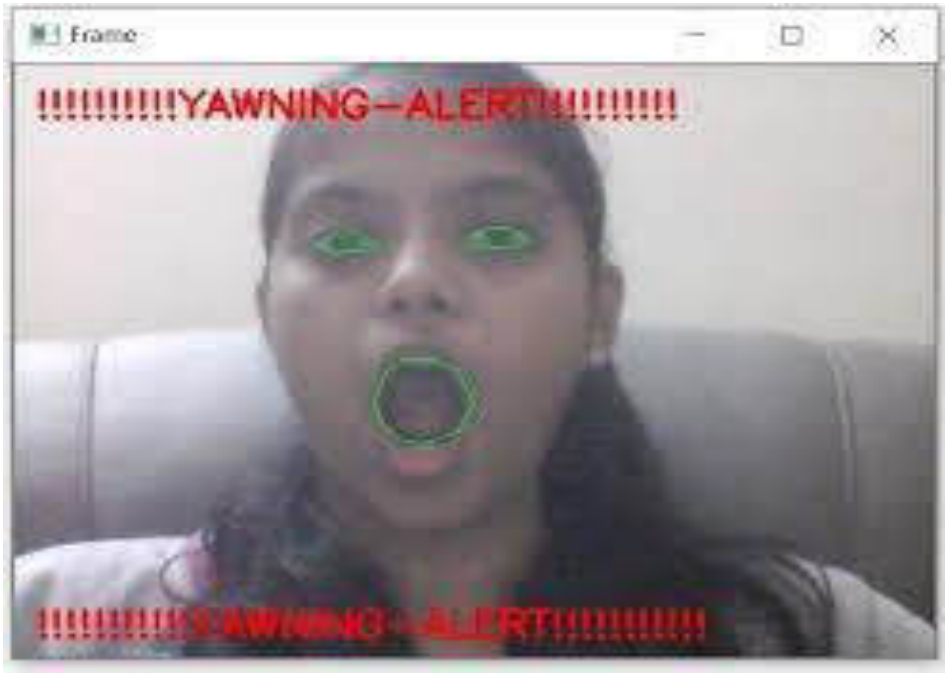


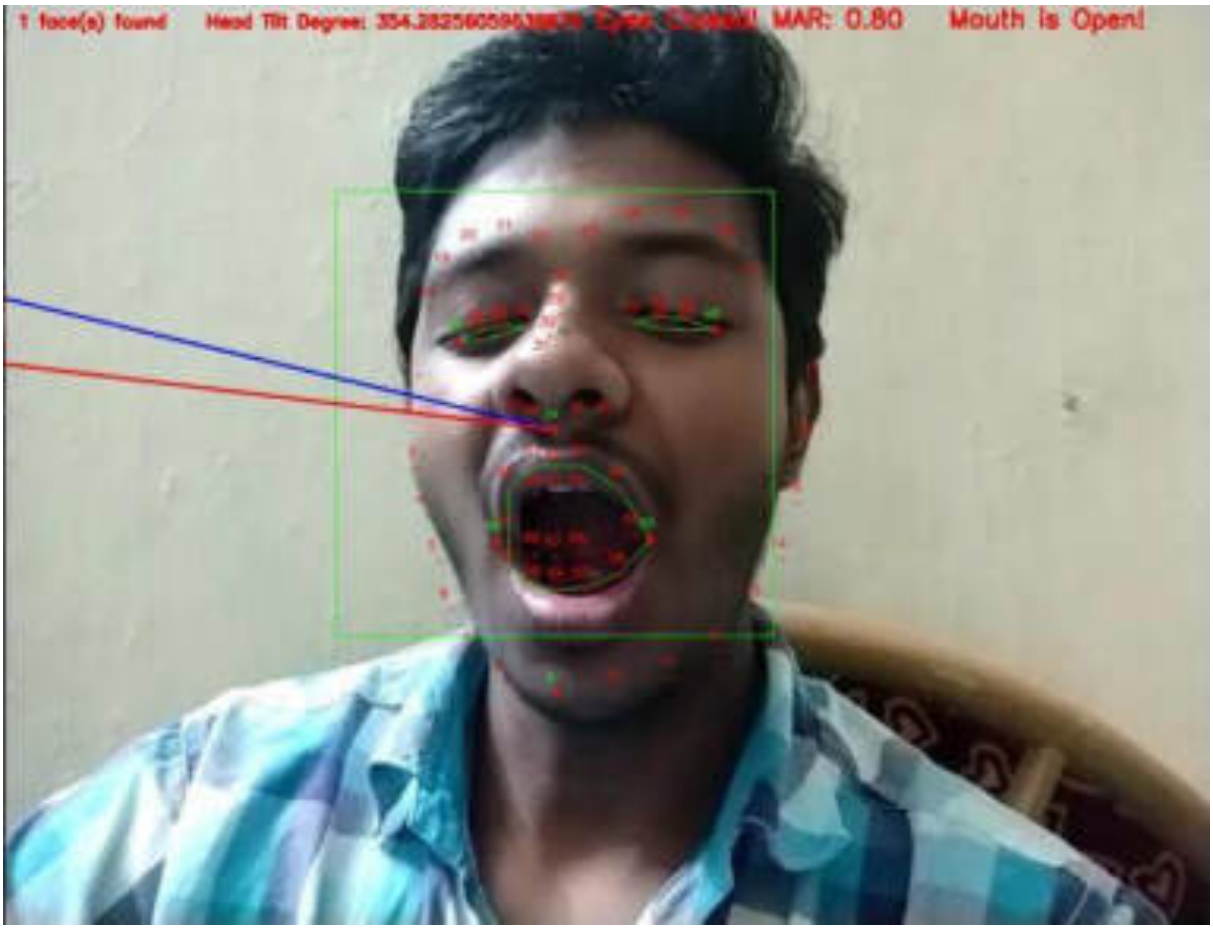
Fig:9.4 Eye Status

- In the Similar way we will determine the yawning status.

Fig: 9.5 Yawning Alert







- Along with closing of eye, yawning status we can also consider another parameter that is head tilt position in detecting in the driver fatigue. In head tilt position, consider the 360 degree (both left side and left side).

10. UML DIAGRAMS

A UML diagram is a diagram based on the UML (Unified Modeling Language) with the purpose of visually representing a system along with its main actors, roles, actions, artifacts or classes, in order to better understand, alter, maintain, or document information about the system. As every diagram need not to be included in our project, we tested out what are the best suited diagrams for our project. Some of them are

1. Use case Diagram
2. Class Diagram
3. State Diagram
4. Sequence Diagram
5. Collaboration Diagram

Use Case Diagram:

A use case describes a sequence of actions that provide something of measurable value to an actor and is drawn as a horizontal ellipse.

Actors: An actor is a person, organization, or external system that plays a role in one or more interactions with the system.

System boundary boxes (optional):

A rectangle is drawn around the use cases, called the system boundary box, to indicate the scope of system. Anything within the box represents functionality that is in scope and anything outside the box is not.

Four relationships among use cases are used often in practice.

a. Include:

In one form of interaction, a given use case may include another. "Include is a Directed Relationship between two use cases, implying that the behavior of the included use case is inserted into the behavior of the including use case.

The first use case often depends on the outcome of the included use case. This is useful for extracting truly common behaviors from multiple use cases into a single description. The notation is a dashed arrow from the including to the included use case, with the label

"«include»". There are no parameters or return values. To specify the location in a flow of events in which the base use case includes the behavior of another, you simply write include followed by the name of use case you want to include, as in the following flow for track order.

b. Extend:

In another form of interaction, a given use case (the extension) may extend another. This relationship indicates that the behavior of the extension use case may be inserted in the extended use case under some conditions. The notation is a dashed arrow from the extension to the extended use case, with the label "«extend»". Modelers use the «extend» relationship to indicate use cases that are "optional" to the base use case.

c. Generalization:

In the third form of relationship among use cases, a generalization/specialization relationship exists. A given use case may have common behaviors, requirements, constraints, and assumptions with a more general use case. In this case, describe them once, and deal with it in the same way, describing any differences in the specialized cases. The notation is a solid line ending in a hollow triangle drawn from the specialized to the more general use case (following the standard generalization notation.

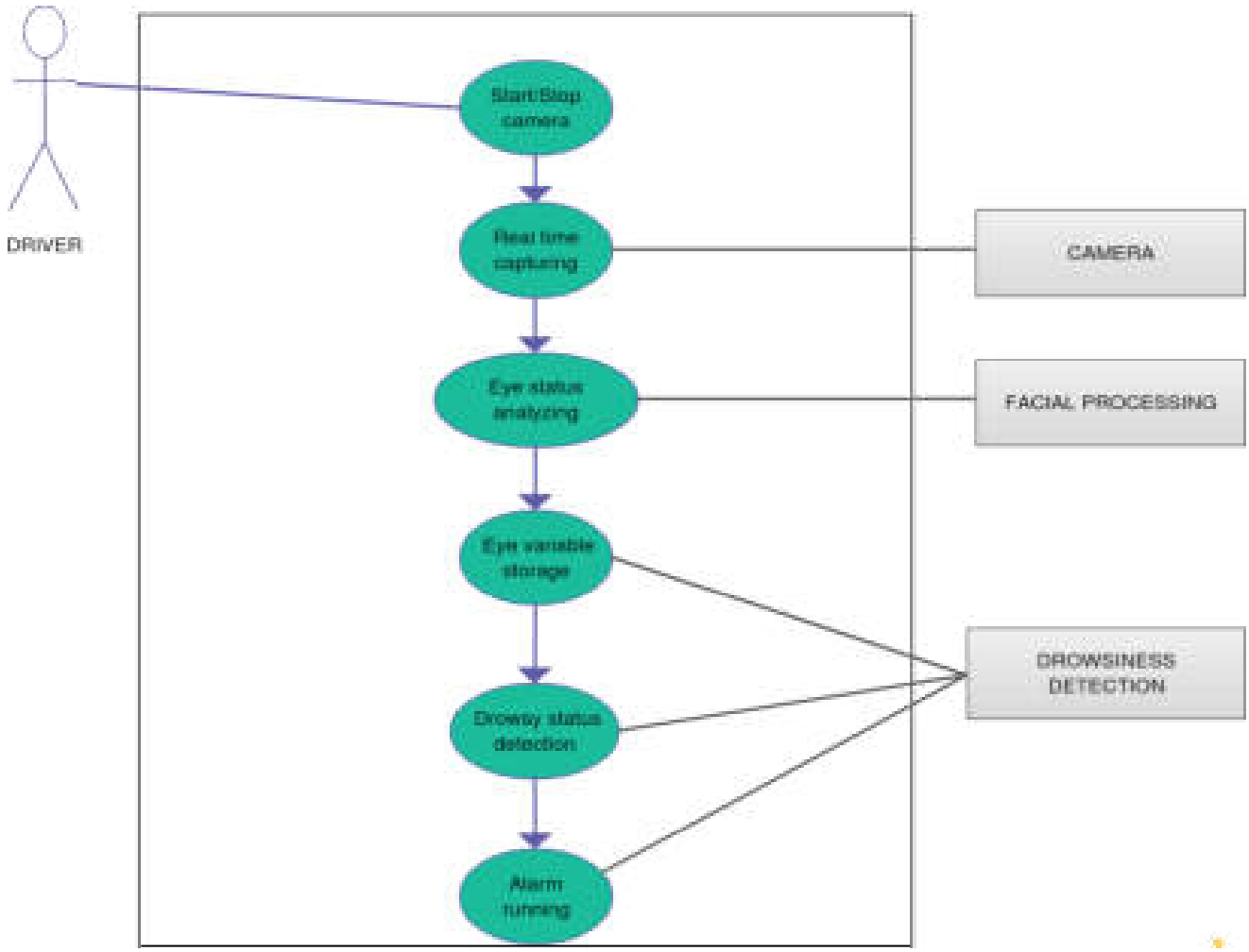
d. Associations:

Associations between actors and use cases are indicated in use case diagrams by solid lines. An association exists whenever an actor is involved with an interaction described by a use case. Associations are modeled as lines connecting use cases and actors to one another, with an optional arrowhead on one end of the line. The arrowhead is often used to indicating the direction of the initial invocation of the relationship or to indicate the primary actor within the use case.

Identified Use Cases

The "user model view" encompasses a problem and solution from the preservative of those individuals whose problem the solution addresses. The view presents the goals and objectives of the problem owners and their requirements of the solution. This view is composed of "use case diagrams". These diagrams describe the functionality provided by a system to external integrators. These diagrams contain actors, use cases, and their relationships.

Fig: 10.1 Use case diagram



Class Diagram

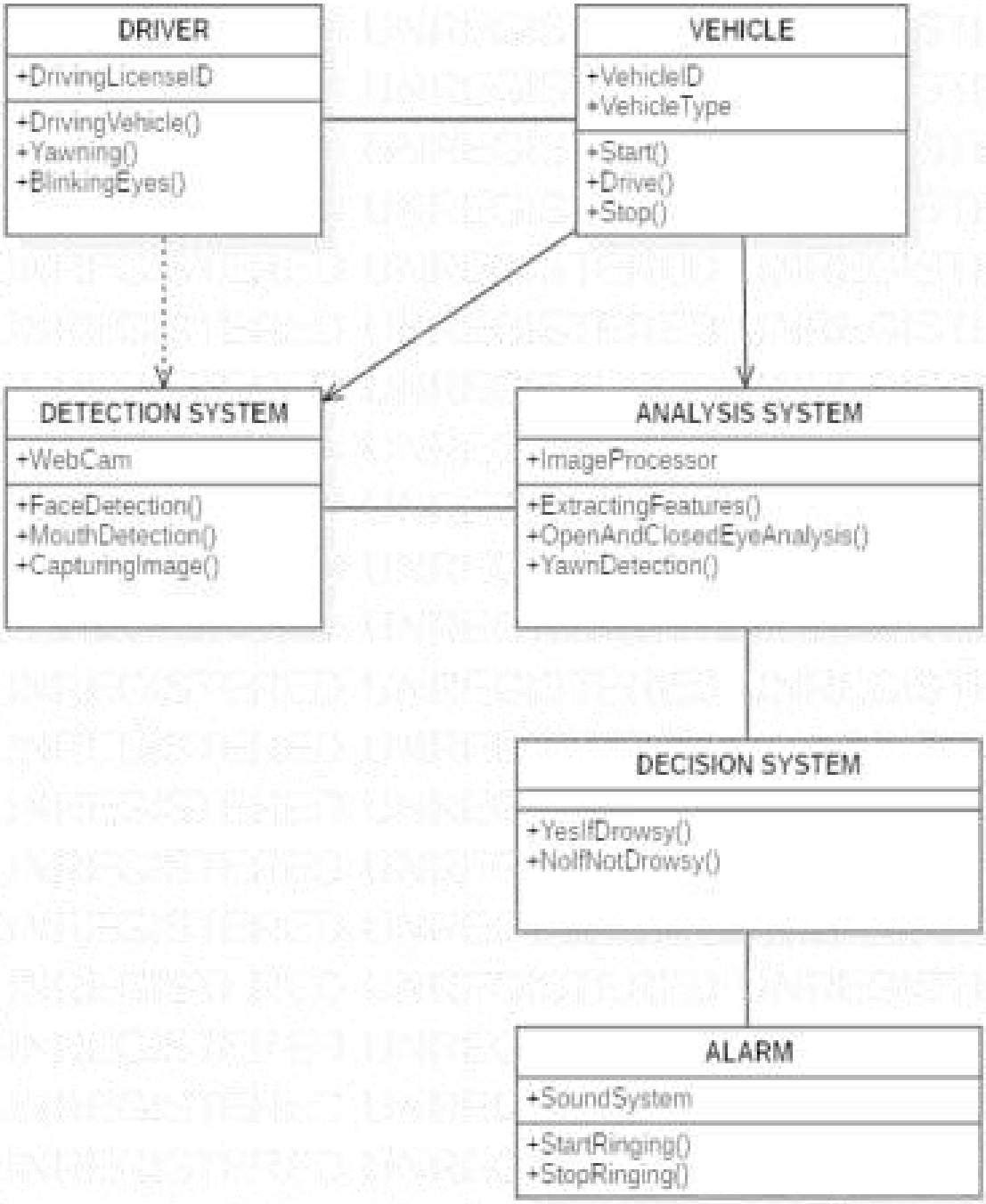
Model, objects are entities that combine state (i.e., data), behavior (i.e., procedures, or methods) and identity (unique existence among all other objects). The structure and behavior of an object are defined by a class, which is a definition, or blueprint, of all objects of a specific type. An object must be explicitly created based on a class and an object thus created is considered to be an instance of that class. An object is similar to a structure, with the addition of method pointers, member access control, and an implicit data member which locates instances of the class (i.e. actual objects of that class) in the class hierarchy (essential for runtime inheritance features)

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, and the relationships between the classes.

The class diagram is the main building block in object oriented modeling. It is used both for general conceptual modeling of the semantics of the application, and for detailed modeling translating the models into programming code. The classes in a class diagram represent both the main objects and or interactions in the application and the objects to be programmed. In the class diagram these classes are represented with boxes which contain the two parts:

- The upper part holds the name of the class.
- The middle part contains the attributes of the class.
- The lower part contains the operations of the class.

Fig: 10.2 Class diagram



Activity Diagram:

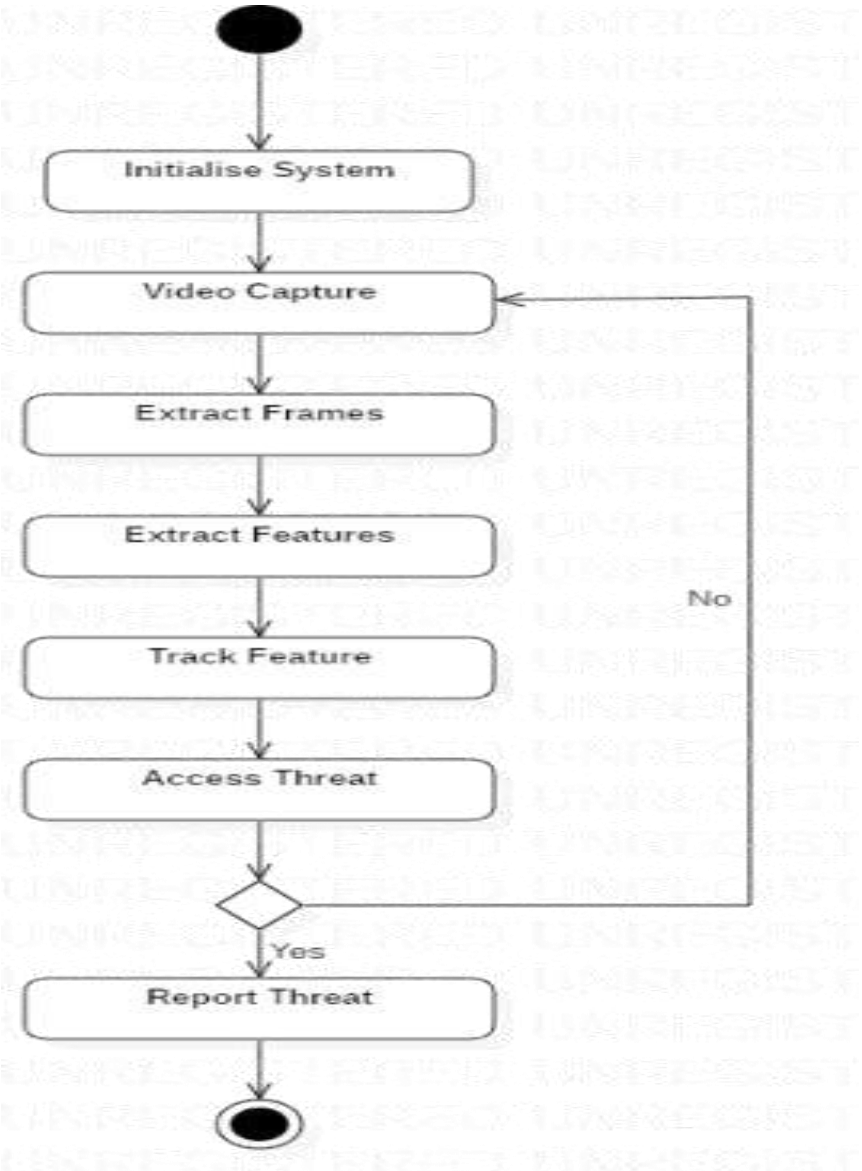
Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control. Activity diagrams are constructed from a limited repertoire of shapes, connected with arrows.

The most important shape types:

- rounded rectangles represent activities;
- bars represent diamonds represent decisions;
- the start (split) or end (join) of concurrent activities;
- a black circle represents the start (initial state) of the work flow;
  ;
- An encircled black circle represents the end (final state).

Arrows run from the start towards the end and represent the order in which activities happen. However, the join and split symbols in activity diagrams only resolve this for simple cases; the meaning of the model is not clear when they are arbitrarily combined with the decisions or loops.

Fig: 10.3 Activity diagram

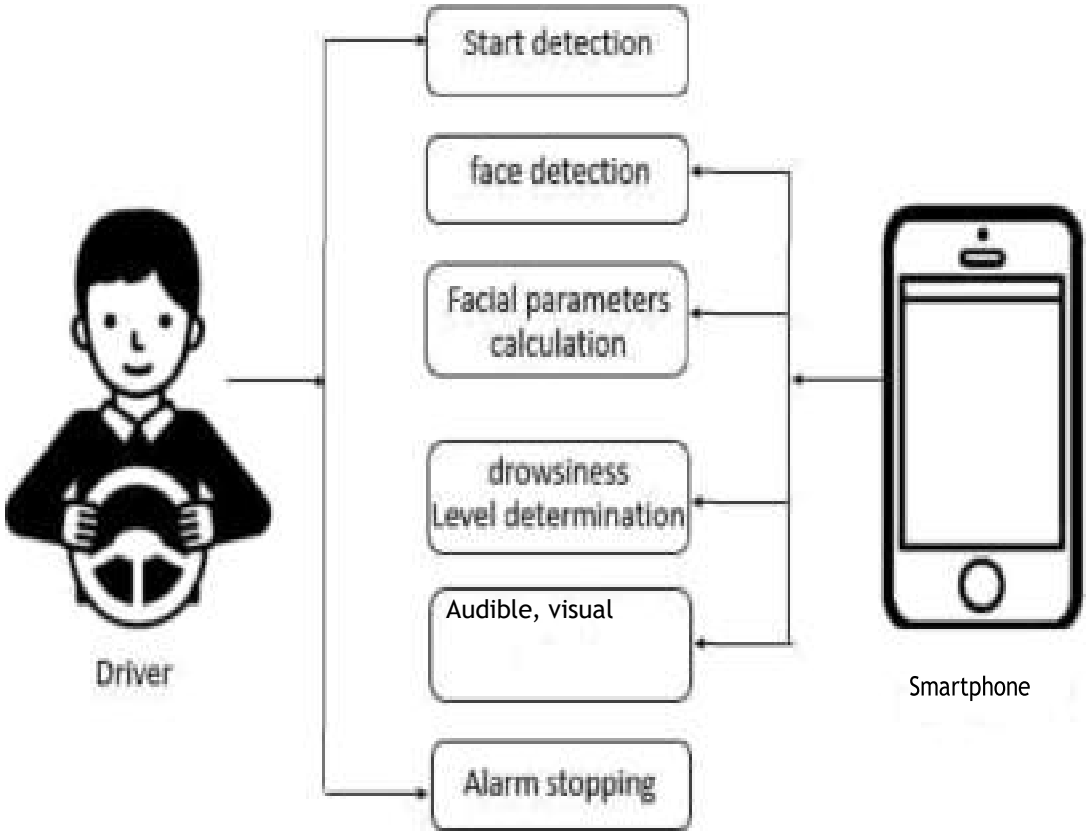


State Chart Diagram:

Objects have behaviors and states. The state of an object depends on its current activity or condition. A state chart diagram shows the possible states of the object and the transitions that cause a change in state. A state diagram, also called a state machine diagram or state chart diagram, is an illustration of the states an object can attain as well as the transitions between those states in the Unified Modeling Language. A state diagram resembles a flowchart in which the initial state is represented by a large black dot and subsequent states are portrayed as boxes with rounded corners. There may be one or two horizontal lines through a box, dividing it into stacked sections. In that case, the upper section contains the name of the state, the middle section (if any) contains the state variables and the lower section contains the actions performed in that state. If there are no horizontal lines through a box, only the name of the state is written inside it. External straight lines, each with an arrow at one end, connect various pairs of boxes. These lines define the transitions between states. The final state is portrayed as a large black dot with a circle around it. Historical states are denoted as circles with the letter H letter

Each state diagram typically begins with dark circle that indicates the initial state and ends with a bordered circle that denotes the final state. However despite having clear state and end points, state diagrams are not necessarily the best tool for capturing an overall progression of events. It mainly depict state and transitions.

Fig: **10.4** State chart diagram



11. SOURCE CODE

```
#!/usr/bin/env python
from scipy.spatial import distance as dist
from imutils.video import VideoStream
from imutils import face_utils
import argparse
import imutils
import time
import dlib
import math
from cv2 import cv2
import numpy as np
from EAR import eye_aspect_ratio
from MAR import mouth_aspect_ratio
from HeadPose import getHeadTiltAndCoords

# Camera is initializing
# initializedlib's face detector (HOG-based) and then create the
# Facial landmark predictor
print("[INFO] loading facial landmark predictor...")
detector = dlib.get_frontal_face_detector()
predictor = dlib.shape_predictor(
    './dlib_shape_predictor/shape_predictor_68_face_landmarks.dat')

# initialize the video stream and sleep for a bit, allowing the
# Camera sensor to warm up
print("[INFO] initializing camera...")

vs = VideoStream(src=0).start()
# vs = VideoStream(usePiCamera=True).start() # Raspberry Pi
time.sleep(2.0)

# 400x225 to 1024x576
frame_width = 1024
frame_height = 576

# loop over the frames from the video stream
# 2D image points. If you change the image, you need to change vector
image_points = np.array([
    (359, 391),     # Nose tip 34
    (399, 561),     # Chin 9
    (337, 297),     # Left eye left corner 37
    (513, 301),     # Right eye right corner46
    (345, 465),     # Left Mouth corner 49
    (453, 469)      # Right mouth corner 55
], dtype="double")

(lStart, lEnd) = face_utils.FACIAL_LANDMARKS_IDXS["left_eye"]
```

```
(rStart, rEnd) = face_utils.FACIAL_LANDMARKS_IDXS["right_eye"]

EYE_AR_THRESH = 0.25
MOUTH_AR_THRESH = 0.79
EYE_AR_CONSEC_FRAMES = 3
COUNTER = 0

# grab the indexes of the facial landmarks for the mouth
(mStart, mEnd) = (49, 68)

while True:
   # grab the frame from the threaded video stream, resize it to
   # have a maximum width of 400 pixels, and convert it to
   # grayscale
   frame = vs.read()
   frame = imutils.resize(frame, width=1024, height=576)
   gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)
   size = gray.shape

   # detect faces in the grayscale frame
rects = detector(gray, 0)

# Check to see if a face was detected, and if so, draws the total
   # number of faces on the frame
   if len(rects) > 0:
      text = "{} face(s) found".format(len(rects))
      cv2.putText(frame, text, (10, 20),
               cv2.FONT_HERSHEY_SIMPLEX, 0.5, (0, 0, 255), 2)

   # loop over the face detections
   for rect in rects:
      # compute the bounding box of the face and draw it on the
      # frame
      (bX, bY, bW, bH) = face_utils.rect_to_bb(rect)
      cv2.rectangle(frame, (bX, bY), (bX + bW, bY + bH), (0, 255, 0), 1)
      # determine the facial landmarks for the face region, then
      # convert the facial landmark (x, y)-coordinates to a NumPy
      # array
      shape = predictor(gray, rect)
      shape = face_utils.shape_to_np(shape)

      # Extract the left and right eye coordinates, then use the
      # coordinates to compute the eye aspect ratio for both eyes
leftEye = shape[lStart:lEnd]
rightEye = shape[rStart:rEnd]
leftEAR = eye_aspect_ratio(leftEye)
rightEAR = eye_aspect_ratio(rightEye)
      # average the eye aspect ratio together for both eyes
      ear = (leftEAR + rightEAR) / 2.0

      # compute the convex hull for the left and right eye, then
      # visualize each of the eyes
```

```
leftEyeHull = cv2.convexHull(leftEye)
rightEyeHull = cv2.convexHull (rightEye)
        cv2.drawContours(frame, [leftEyeHull], -1, (0, 255, 0), 1)
        cv2.drawContours(frame, [rightEyeHull], -1, (0, 255, 0), 1)

        # check to see if the eye aspect ratio is below the blink
        # threshold, and if so, increment the blink frame counter
        if ear < EYE_AR_THRESH:
            COUNTER += 1
            # if the eyes were closed for a sufficient number of times
            # then show the warning
            if COUNTER >= EYE_AR_CONSEC_FRAMES:
                cv2.putText(frame, "Eyes Closed!", (500, 20),
                        cv2.FONT_HERSHEY_SIMPLEX, 0.7, (0, 0, 255), 2)
            # otherwise, the eye aspect ratio is not below the blink
            # threshold, so reset the counter and alarm
        else:
            COUNTER = 0

        mouth = shape[mStart:mEnd]

mouthMAR = mouth_aspect_ratio(mouth)
        mar = mouthMAR
        # compute the convex hull for the mouth, then
        # visualize the mouth
mouthHull = cv2.convexHull(mouth)

        cv2.drawContours(frame, [mouthHull], -1, (0, 255, 0), 1)
        cv2.putText(frame, "MAR: {:.2f}".format(mar), (650, 20),
                cv2.FONT_HERSHEY_SIMPLEX, 0.7, (0, 0, 255), 2)

        # Draw text if mouth is open
        if mar > MOUTH_AR_THRESH:
            cv2.putText(frame, "Yawning!", (800, 20),
                    cv2.FONT_HERSHEY_SIMPLEX, 0.7, (0, 0, 255), 2)


        # loop over the (x, y)-coordinates for the facial landmarks
        # and draw each of them
        for (i, (x, y)) in enumerate(shape):
            if i == 33:
                # something to our key landmarks
                # save to our new key point list
                # i.e. keypoints = [(i,(x,y))]
image_points[0] = np.array([x, y], dtype='double')
                # write on frame in Green
                cv2.circle(frame, (x, y), 1, (0, 255, 0), -1)
                cv2.putText(frame, str(i + 1), (x - 10, y - 10),
                        cv2.FONT_HERSHEY_SIMPLEX, 0.35, (0, 255, 0), 1)
elifi == 8:
                # something to our key landmarks
                # save to our new key point list
```

```
            # i.e. keypoints = [(i,(x,y))]
image_points[1] = np.array([x, y], dtype='double')
            # write on frame in Green
            cv2.circle(frame, (x, y), 1, (0, 255, 0), -1)
            cv2.putText(frame, str(i + 1), (x - 10, y - 10),
                        cv2.FONT_HERSHEY_SIMPLEX, 0.35, (0, 255, 0), 1)
elifi == 36:
            # Something to our key landmarks
            # save to our new key point list
            # i.e. keypoints = [(i,(x,y))]
image_points[2] = np.array([x, y], dtype='double')
            # write on frame in Green
            cv2.circle(frame, (x, y), 1, (0, 255, 0), -1)
            cv2.putText(frame, str(i + 1), (x - 10, y - 10),
                        cv2.FONT_HERSHEY_SIMPLEX, 0.35, (0, 255, 0), 1)
elifi == 45:
            # something to our key landmarks
            # save to our new key point list
            # i.e. keypoints = [(i,(x,y))]
image_points[3] = np.array([x, y], dtype='double')
            # write on frame in Green
cv2.circle (frame, (x, y), 1, (0, 255, 0), -1)
            cv2.putText(frame, str(i + 1), (x - 10, y - 10),
                        cv2.FONT_HERSHEY_SIMPLEX, 0.35, (0, 255, 0), 1)
elifi == 48:
            # something to our key landmarks
            # save to our new key point list
            # i.e. keypoints = [(i,(x,y))]
image_points[4] = np.array([x, y], dtype='double')
            # write on frame in Green
            cv2.circle(frame, (x, y), 1, (0, 255, 0), -1)
            cv2.putText(frame, str(i + 1), (x - 10, y - 10),
                        cv2.FONT_HERSHEY_SIMPLEX, 0.35, (0, 255, 0), 1)
elifi == 54:
            # something to our key landmarks
            # save to our new key point list
            # i.e. keypoints = [(i,(x,y))]
image_points[5] = np.array([x, y], dtype='double')
            # write on frame in Green
            cv2.circle(frame, (x, y), 1, (0, 255, 0), -1)
            cv2.putText(frame, str(i + 1), (x - 10, y - 10),
                        cv2.FONT_HERSHEY_SIMPLEX, 0.35, (0, 255, 0), 1)
        else:
            # everything to all other landmarks
            # write on frame in Red
            cv2.circle(frame, (x, y), 1, (0, 0, 255), -1)
            cv2.putText(frame, str(i + 1), (x - 10, y - 10),
                        cv2.FONT_HERSHEY_SIMPLEX, 0.35, (0, 0, 255), 1)

    #Draw the determinant image points onto the person's face
    for p in image_points:
        cv2.circle(frame, (int(p[0]), int(p[1])), 3, (0, 0, 255), -1)
```

```
        (head_tilt_degree, start_point, end_point,
end_point_alt) = getHeadTiltAndCoords(size, image_points, frame_height)

        cv2.line(frame, start_point, end_point, (255, 0, 0), 2)
        cv2.line(frame, start_point, end_point_alt, (0, 0, 255), 2)

 if head_tilt_degree:
            cv2.putText(frame, 'Head Tilt Degree: ' + str(head_tilt_degree[0]), (170, 20),
                        cv2.FONT_HERSHEY_SIMPLEX, 0.5, (0, 0, 255), 2)

 # extract the mouth coordinates, then use the
 # coordinates to compute the mouth aspect ratio
    # show the frameq
    cv2.imshow("Frame", frame)
    key = cv2.waitKey(1) & 0xFF

    # if the `q` key was pressed, break from the loop
    if key == ord("q"):
        break

# print(image_points)

# do a bit of cleanup
cv2.destroyAllWindows()
vs.stop()
```

OUTPUT SCREEN SHOTS:

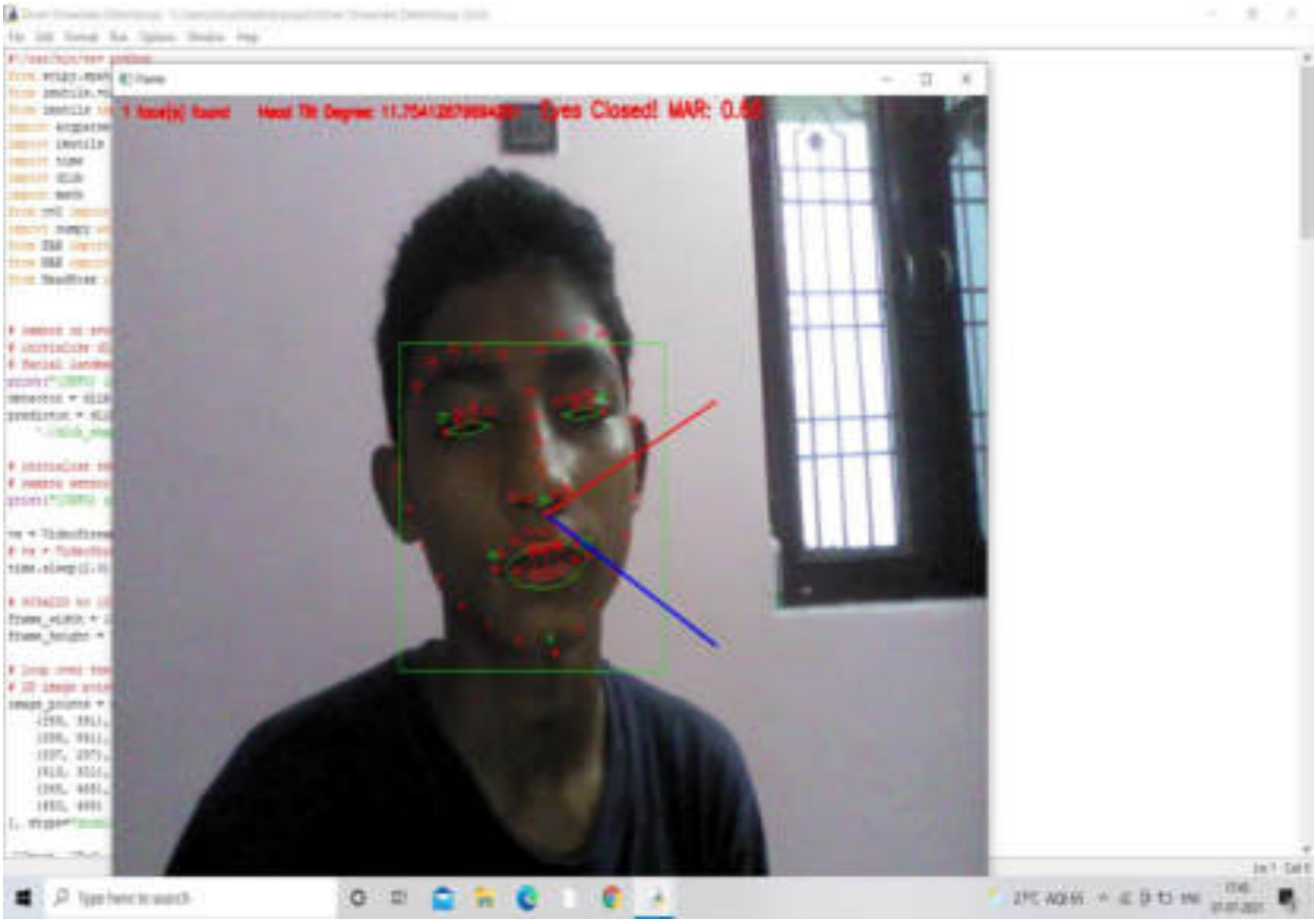


Fig: 11.1 output Screenshot for closing of eye

Fig: 11.2 output Screenshot for Yawning Status

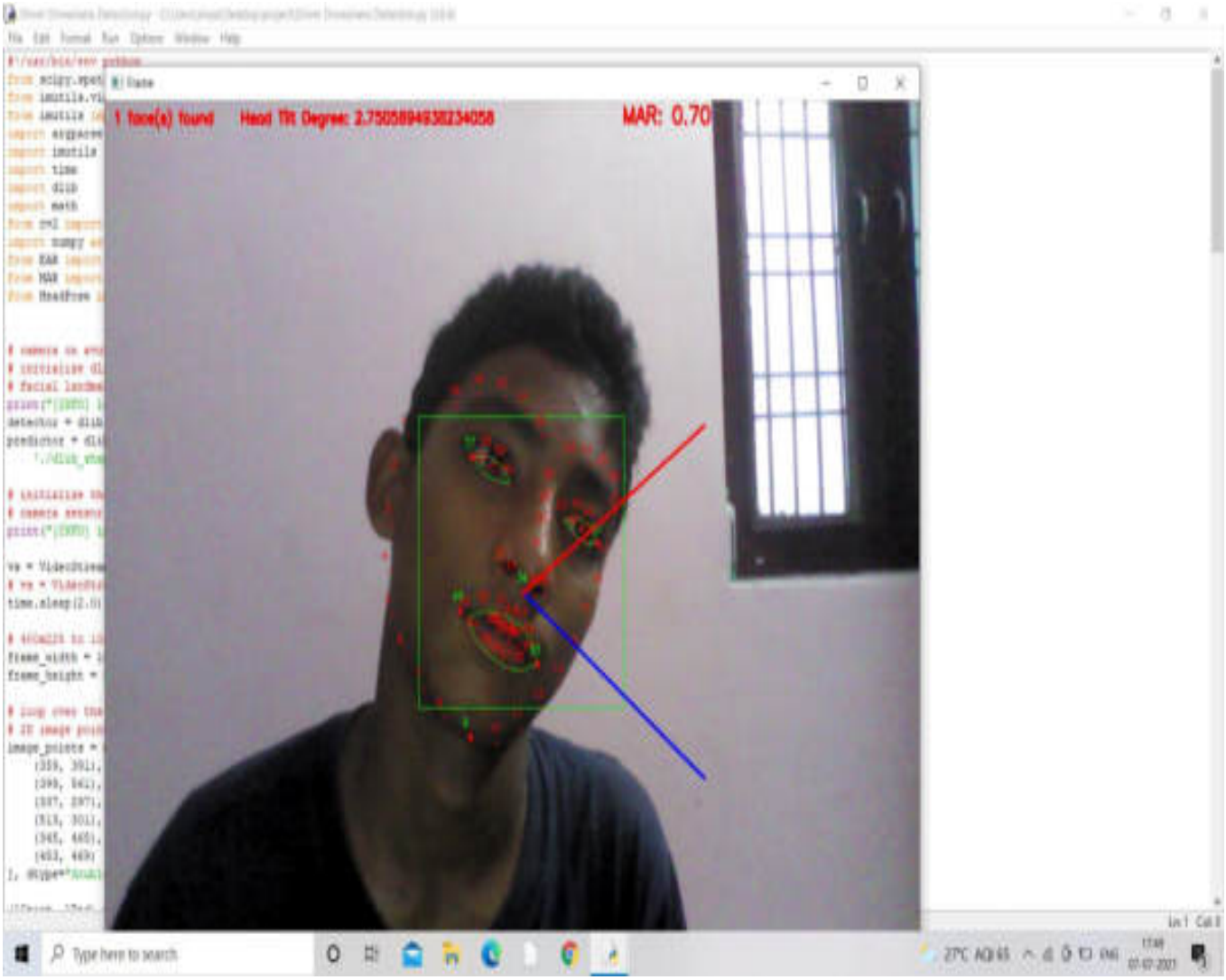


Fig:11.3 output Screenshot for head tilt position

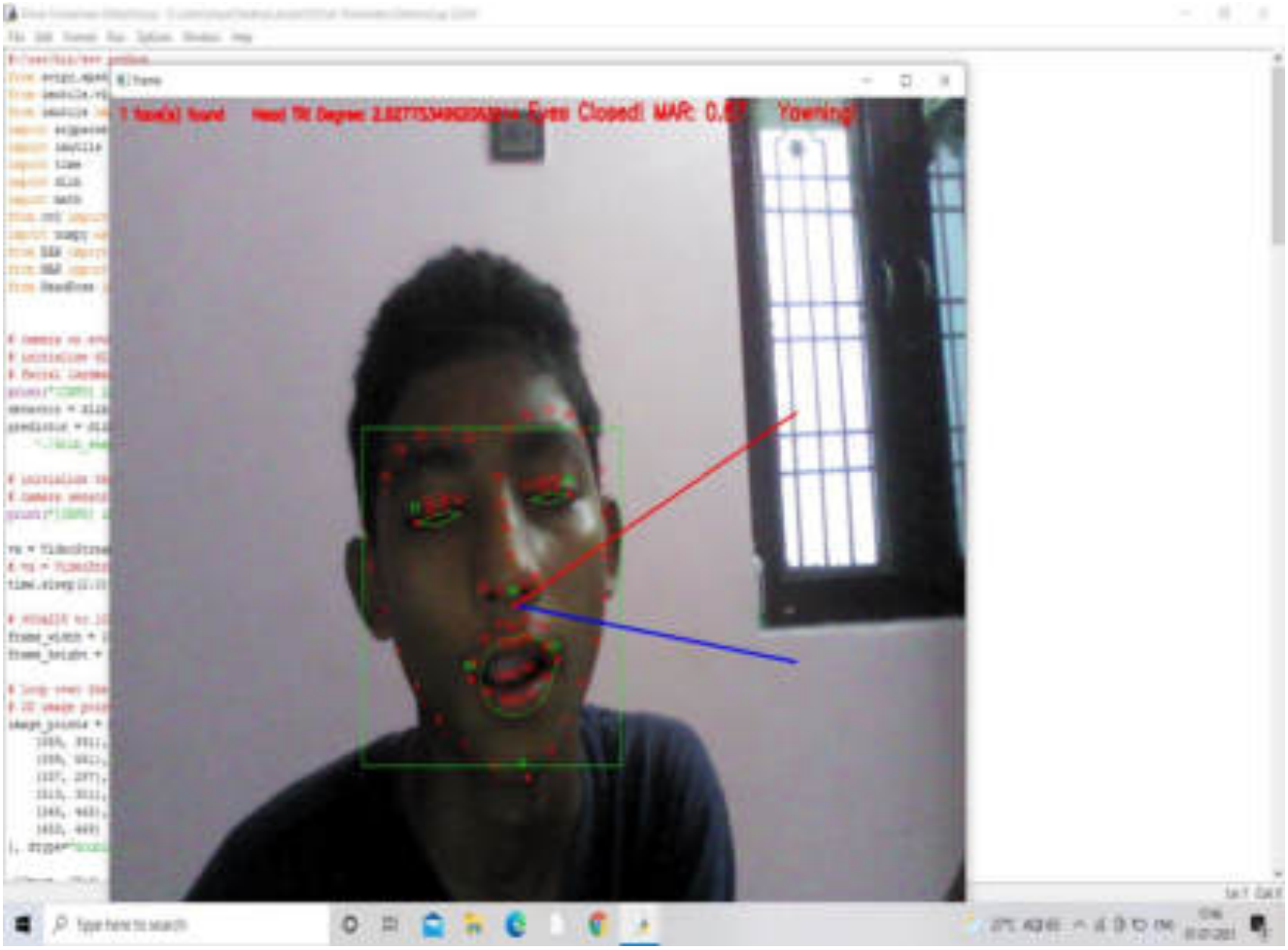


Fig: 11.1 output Screenshot for all the parameters

12. SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, subassemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

TYPES OF TESTS

Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user

manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output :identified classes of application outputs must be exercised

Systems/ Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box. you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

Unit Testing

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link
- The entry screen, messages and responses must not be delayed.

Features to be tested:

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects. The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

Test Results

All the test cases mentioned above passed successfully. No defects encountered.

TESTINGMETHODOLOGIES

The following are the Testing Methodologies:

- **Unit Testing.**
- **Integration Testing**
- **User Acceptance Testing.**
- **Output Testing.**
- **Validation Testing**

Unit Testing

Unit testing focuses verification effort on the smallest unit of Software design that is the module. Unit testing exercises specific paths in a module's control structure to ensure complete coverage and maximum error detection. This test focuses on each module individually, ensuring that it functions properly as a unit. Hence, the naming is Unit Testing.

During this testing, each module is tested individually and the module interfaces are verified for the consistency with design specification. All the important processing path are tested for the expected results. All error handling paths are also tested.

Integration Testing

Integration testing addresses the issues associated with the dual problems of verification and program construction. After the software has been integrated a set of high order tests are conducted. The main objective in this testing process is to take unit tested modules and builds a program structure that has been dictated by design.

The following are the types of Integration Testing:

Top Down Integration

This method is an incremental approach to the construction of program structure. Modules are integrated by moving downward through the control hierarchy, beginning with the main program module. The module subordinates to the main program module are incorporated into the structure in either a depth first or breadth first manner.

In this method, the software is tested from main module and individual stubs are replaced when the test proceeds downwards.

Bottom-up Integration

This method begins the construction and testing with the modules at the lowest level in the

program structure. Since the modules are integrated from the bottom up, processing required for modules subordinate to a given level is always available and the need for stubs is eliminated. The bottom up integration strategy may be implemented with the following steps:

- The low-level modules are combined into clusters into clusters that perform a specific Software sub-function.
- A driver (the control program) for testing is written to coordinate test case input and output.
- The cluster is tested.
- Drivers are removed and clusters are combined moving upward in the program structure

The bottom up approaches tests each module individually and then each module is module is integrated with a main module and tested for functionality.

User Acceptance Testing.

User Acceptance of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changes wherever required. The system developed provides a friendly user interface that can easily be understood even by a person who is new to the system.

Output Testing:

After performing the validation testing, the next step is output testing of the proposed system, since no system could be useful if it does not produce the required output in the specified format. Asking the users about the format required by them tests the outputs generated or displayed by the system under consideration. Hence the output format is considered in 2 ways– one is on screen and another in printed format.

13. CONCLUSION

Convolutional Neural Network, a supervised machine learning algorithm gives accurate and better results as compared to other algorithms. For Drowsiness Detection, the model is trained on the pre-processed data and hence is able to determine the driver fatigue of the face image. The parameters considered for drowsiness detection are: eyes, mouth and head tilt. In order to reduce road accidents we need to detect the causes such as drowsiness, fatigue and to alert the driver. This project enables to choose an efficient method to reduce read accidents due to driver fatigue.

14. FUTURE SCOPE

- Rather using alarm we can use Automatic Braking System which will reduce the speed of the vehicle.
- The vehicle can be automatically parked by first using Automatic braking system, which will reduce the speed and concurrently will turn on the parking lights of the vehicle
- Using Pressure sensor on the steering alarm or Automatic braking System can be set in case of drowsiness. By using wireless Technology if the driver gets drowsy an alert message can be sent to a selected person's mobile by using GSM module along with the alarm in vehicle.

15. REFERENCES

- Computationally efficient face detection; B. Schlkopf-A. Blake,S.Romdhani, and P. Torr
- Use of the Hough transformation to detect lines and curves in picture; R.Duda and P.E.Hart.
- Open/Closed eye analysis for drowsiness detection; .R. Tabrizi and R.A. Zoroofi
- http://ncrb.gov.in/StatPublications/ADSI/ADSI2015/chapter1A%20traffic%20accidents.pdf
- http://www.jotr.in/text.asp?2013/6/1/1/118718
- http://dlib.net/face_landmark_detection_ex.cpp.html
- Ameratunga.S,Bailey.J,Connor.J,Civil.I,Dunn.R,Jackson.R,Norton.R,and,Robinson.E,Driversleepiness and risk of serious injury to car occupants: Populationcontrolstudy.‖*BritishMedical Journal*, vol. 324, 2002, pp. 1125–1129.
- Distante.A,D'Orazio.T, Guaragnella.C,andLeo.M, ―A visualapproach for Bronte.S, Bergasa.L, Delgado.B, Garc´ıa.I, Hern´andez.N,and Sevillano.M,Vision-based drowsiness detector for a realistic driving simulator,‖in *IEEE Intelligent TransportationsSystemsConference(ITSC)*, 2010.
- Bronte.S, Bergasa.L, Delgado.B, Garc´ıa.I, Hern´andez.N,and Sevillano.M,―Vision-based drowsiness detector for a realistic driving simulator,‖in *IEEE Intelligent TransportationsSystemsConference(ITSC)*, 2010.
- Tippayanate.N,Theimpactofsleepdeprivationondrivingperformance.*KhonKaenHospitalMedical Journal,* 30(3): 320-325, 2006.
- Dingus.T.A, Klauer.S.G, Neale.V.L, and Sudweeks.J.D,‖ The impact of driver inattention on near-crash/crash risk: An analysis using the 100 car naturalistic driving study data‖, National Highway Traffic Safety Adminstration, DC, DOT HS, 2006, vol.810.
- Faber.J, ―Detection of different levels of vigilance by eeg pseudo spectra.‖ Neural Network World, vol. 14, no. 3-4, pp. 285–290, 2004.

- Igarashi.K ,Itou.K, Itakura.F, Miyajima.C, Ozawa.T, Takeda.K and Wakita.T, ―Driver identification using driving behavior signals,‖ IEICE - Trans. Inf. Syst., vol. E89- D, no. 3, 2006, pp. 1188–1194.

- Nakano.T, Suzuki.M, Yamamoto.N, Yamamoto.O, and Yamamoto.S, ―Measurement of driver's consciousness by image processing a method for presuming driver's drowsiness by eyeblinks coping with individual differences.‖ Systems, Man and Cybernetics, vol. 4, 2006, p. 2891– 2896.

- World Health Organization, Global Status Report on Road Safety 2013: Supporting a Decade of Action : Summary. World Health Organization, 2013. [Online]. Available: http://books.google.com/books?id=qzK2nQEACAAJ

- T. Akertedt, P. Fredlung, M. Gillberg, and B. Jansson, "A prospective study of fatal occupational accidents relationship to sleeping difficulties and occupational factors," Journal of Sleep Research, vol. 11, no. 1, pp. 69–71, 2002. [Online]. Available: http://dx.doi.org/10.1046/j.1365- 2869.2002.00287.x

- E. Vural, "Video based detection of driver fatigue," Ph.D. dissertation, Sabanci University, 2009.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

Home Security with Real Time Image Processing Using OpenCV

A Project Report submitted to

Jawaharlal Nehru Technological University, Kakinada

In partial fulfilment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING.

By

T Deeksha (17H71A0508)
J Homeshwar (17H71A0517)
J Nikhil Teja (17H71A0529)
M Yogeshwar (17H71A0559)

Under the Guidance of

Ms. A Anuradha

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

2020-2021

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Autonomous)
Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada
Accredited with NAAC 'A' Grade, NBA Accredited B.Tech(CSE, ECE, EEE, MEC), An ISO 9001:2008 Certified Institute
Kanchikacherla - 521180, Krishna Dist., A.P, India. Phone No.: 08678-277030, 277033

CERTIFICATE

This is to certify that the project work entitled **"Home Security with Real Time Image Processing Using OpenCV"** is a bonafied work carried out by Ms. T DEEKSHA (17H71A0508), Mr. J HOMESHWAR (17H71A0517), Mr. J NIKHIL TEJA (17H71A0529), Mr. M YOGESHWAR (17H71A0559) in partial fulfillment for the award of the degree of Bachelor of Technology in Computer Science and Engineering of Jawaharlal Nehru Technological University, Kakinada during the year 2020-2021. It is certified that all corrections/suggestions indicated for assessment have been incorporated in the report. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the above degree.

Project Guide
(Ms. A Anuradha)

Head of the Department
(Mr. D Prasad)

Principal
(Dr K Srinivas)

Examiner 1

Examiner 2

DECLARATION

Hereby we, who carried out the project on "**Home Security with Real-Time Image Processing Using OpenCV**", declare that the matter included in this project report is a genuine work done by us and has not, been submitted to this university or any other university/institute for the fulfilment of the requirement of the degree.

T Deeksha (17H71A0508)
J Homeshwar (17H71A0517)
J Nikhil Teja (17H71A0529)
M Yogeshwar (17H71A0559)

ACKNOWLEDGEMENT

We are grateful to **Ms. A Anuradha, Assistant Professor, Department of Computer Science and Engineering**, whose guidance has been immense all throughout the project work.

Our sincere thanks to **Dr. D Prasad, Head of the Department, Computer Science and Engineering**, for his valuable guidance and timely advices in the completion of our project.

We express our heartfelt gratitude to **Dr. K. Srinivasa Rao, Principal** for permitting us to take up our project work and to complete the project successfully.

We would like to express our sincere and heartful thanks to all the lecturers of the department for their continuous cooperation, which has given us the cogency to buildup adamant aspiration over the completion of our project.

Finally, we thank one and all who directly and indirectly helped us to complete our project successfully.

ABSTRACT

Home security system has established its importance and benefits numerous times by providing immediate monitoring of the house. This is because of the increasing home theft and burglary incidents that create an awareness among most of the house owners. CCTV-based security systems are not real-time because the alert comes to the owner after the incident occurred unless they are at home during the incident. To overcome this problem, many researchers are developing cost-effective custom-based security systems, which are affordable for everyone. Most of these systems use a Passive Infrared (PIR) motion sensor for motion detection. Although affordable, such a system still has many limitations. For example, false alarms triggered due to an abnormal condition such as rapid heating from sunlight exposure. In this work, a vision-based home security system using OpenCV on Raspberry Pi 3 model B was developed to improve the effectiveness of motion detection. This system applied the Haar-Cascade algorithm coupled with background subtraction as well as considered the Histogram of Oriented Gradients (HOG) during the development stage. The developed prototype was tested under a few conditions to determine the accuracy of motion detection and compare the results with a system that uses a PIR motion sensor for motion detection. From the results obtained, the developed vision-based home security system using OpenCV has 100% of detection rate compared to the PIR motion sensor-based security system with 76% of the detection rate.

CONTENTS

1. INTRODUCTION

Nowadays, the evolution of technology-based systems has drastically increased over the past few years. As the technology grows, it is no surprise that most of the work that was done by human will be taken over by machines. Although many people believe that this will make everyone to be lazy, it is an undeniable fact that this is for the betterment of humankind. Consequently, they have to confront this technology every day, which undoubtedly affects their lifestyle from the way they live until the way they work or relax. The convenience that technology provides them is the most common reason for their willingness to get it to affect their daily lifestyle to such extent. Building management is one particular area where the technology is slowly taking over the responsibilities, which belong to a human. Many corporate buildings have an integrated automation system, which controls their power management, ventilation, security and other operations, often with less or no involvement of human. This automation system is also integrated into many homes now, which is known as smart homes. Similar to their corporate counterparts, the advantages of smart homes compared to a regular one includes more efficient power management, higher surveillance and other household operations. House owners or tenants are able to monitor their household operations in detail and they can be controlled remotely using active internet connections, which helps to improve comfort. In general, most of the smart homes or home automation systems cover the security aspect in the house passively or actively contributing to the increase of the home security.

Surveillance plays an important role in the security aspect from smart homes to corporate-based automation. It refers to close observation using electronic equipment such as CCTV cameras. In a corporate building, security refers to the monitoring of people's activity or behavior to protect valuable information or things compared to smart homes or regular homes, where the security aspects usually focus on burglary and theft of valuable things from the house as these incidents have always been a predicament. However, implementing CCTV-based security system is considered not affordable for an average resident because of the expensive price and some of CCTV systems do not provide a real-time notification to the user when a theft or burglary incident takes place. Security has always been a major issue everywhere around the globe and the importance of security cannot be denied in today's society because of the increasing crime rate. For instance, in Malaysia, the high crime rate can make it a less safe place to stay. Home theft rate in Malaysia is the second highest crime [1] and this creates

awareness in the society. Home security systems development using IoT infrastructure has become ubiquitous because of the high home theft rate. The most common features of the home security system are motion detection, live monitoring, and alert notification. Systems relying only on a Passive Infrared (PIR) sensor to accommodate for motion detection have unreliable detection rate because it could trigger a false alarm due to abnormal conditions such as pet intrusion or rapid heating [2] e.g., from sunlight exposure. False alarms can have significant impacts such as in security systems that trigger calls to the police [3] or other emergency agencies. In order to overcome the limitations of motion detection using only PIR sensors, this paper presents an implementation of vision-based motion detection system using OpenCV and prototyped on a Raspberry Pi 3 Model B. The main objective of this work is to evaluate the accuracy of such a system under a few conditions including pet or non-human intrusion and rapid heating, in relative to a PIR-only system aimed for home security. The rest of this paper is organized as follows. Previous Work reviews a few relevant work, Software & Hardware Implementation describes the software and hardware components used in our prototype, System implementation explains how the system is implemented, Performance Evaluation presents the evaluation results, and finally, Conclusion draws a conclusion and offers a few suggestions as future work.

2. REVIEW OF LITERATURE

Vigneswari et al. presented an automated security system with surveillance. A PIR sensor and a camera were installed respectively to detect the presence of an intruder and capture his/her picture. The owner will be alerted through Short Message Service (SMS) using the GSM technology. At the heart of the system was an Atmega644p microcontroller, which receives and processes signals from the PIR sensor and decides whether it is necessary to send a notification message with the captured image over SMS. Suresh et al. proposed a home monitoring and security system where a PIR sensor and a temperature and humidity sensor are connected to an Arduino Uno microcontroller. The system intends to apply changes in both motion and temperature in a monitored room to improve the accuracy of the intrusion detection by reducing false detections based on line of sight that can be cut by any entity and not necessarily an intruder. If the temperature is above a set threshold and a change in motion is detected, an SMS message will then be sent to the owner's mobile phone via GSM.

An Arduino-based, low-cost, and multi-level home security system was proposed by Zaman et al. [6] for developing countries. Their system consists of two parts, namely internal and external parts to make the system more effective. The internal part is controlled by an Arduino Mega microcontroller with GSM shield to detect intruder(s) and notify the owner through SMS or phone call. Each room was installed with PIR sensors and LED lights attached to the ceiling. When a PIR sensor is triggered, a signal is sent to the microcontroller, LED lights blink, respective lights of the affected room get turned on, an installed alarm starts to buzz for three seconds with intervals to alert the surrounding, and the GSM module sends an SMS notification to the owner. On the other hand, the external part is controlled by an Arduino Uno microcontroller that includes a fingerprint-based door lock and a webcam. Only users with authorized fingerprints saved in the sensor memory will be granted access. Sruthy and George [7] presented a Wi-Fi enabled home security and surveillance system using Raspberry Pi and NodeMCU IoT/Wi-Fi module integrating sensor alerts with video surveillance. Intrusion and fire detections are the prime features of this system. As the system is purely based on Wi-Fi connectivity, data can be fetched from almost anywhere or moved to a cloud storage for future monitoring. A PIR sensor node and a fire sensor node are connected to the NodeMCU module to detect human presence and fire, respectively. If a detection takes place, a signal is then sent to the Raspberry Pi and an installed webcam will be activated to record the event. This feature allows the user to watch a live video of the scene remotely by connecting to the IP address of

the Raspberry Pi. The system has two ways of sending an alert notification, which are through SMS with the help of a GSM module as well as by email. The GSM module will also notify the police for an intrusion alert and the fire force for fire alert.

Embedded home surveillance system with pyroelectric infrared sensor using GSM was introduced by Ragrade [8]. The system is designed with the use of PIR sensor and ultrasonic sensor to detect intruders in a home. If there is an intruder present, a buzzer will be triggered, and SMS will be sent and camera is used to capture image. The microcontroller unit used is the Atmega 328. Two sets of PIR and ultrasonic sensors are used to detect for intruders, one for outdoors and one for indoors.

Smart Surveillance Monitoring System Using Raspberry PI and PIR Sensor was introduced by Prasad et. al. [9]. The system makes use of a 3G dongle to transmit information to a smartphone on web application. Live streaming is also carried out by the camera and is stored for future playback. When motion is detected by the PIR sensor, the cameras automatically begin the recording and owner is alerted via his smartphone regarding the possible intrusion. The two main components of the Raspberry Pi interacting are the web application on the smartphone browser, and the server-scripts that run on a cloud.

Kaur et al. [10] introduced An Illustration of Making a Home Automation System Using Raspberry Pi and PIR Sensor. When PIR sensor detects motion, it displays string outputs stating there is an intruder on the Raspbian terminal, and vice versa. No external user interface was set up, and only made use of the Raspbian terminal to show whether or not there is an intruder present in the house.

These systems make use of PIR sensors, which may be unreliable at times, hence could cause false detections and alarms when implemented in surveillance systems. One example is how PIR sensor might detect motion when receiving rapid heating from being exposed to the sun. On the other hand, there are in fact other options such as computer vision techniques which can be used to replace the said sensor. Hence, this work aims to improve the effectiveness of motion detection by using OpenCV by carrying out a comparison between the use of PIR sensor and OpenCV techniques in motion detection.

3. SYSTEM ANALYSIS

EXISTING SYSTEMS:

- CCTV Security System is a traditional system used so far and considered as the first option in most situations.
- Systems with Passive Infrared sensors are engineered for real-time monitoring.
- Systems relying on PIR sensor can integrate motion detection.

Disadvantages:

- CCTV-based security systems are not real-time and are expensive for a common man's home security.
- PIR sensor-based solutions have unreliable detection rate as it could trigger a false alarm due to abnormal conditions such as pet intrusion.
- False alarms can have significant impacts such as in security systems that trigger calls to the police or other emergency agencies.

PROPOSED SYSTEM:

In order to overcome the limitations of motion detection using only PIR sensors, we present an implementation of vision-based motion detection system using real time Image Processing with OpenCV, prototyped on a Raspberry Pi SBC.

Advantages:

- Consistent and reliable detection rate.
- Completely software-based detection.
- Real-time monitoring and reporting system.
- Image based recognition allows authentication for selected people.
- Immune to false alarms.

Software Requirements:

- OS: Raspberry PI OS (Linux Based)
- Language: Python 3.7
- Libraries:
    - OpenCV
    - NumPy
    - SMTP
    - Email
- Development Environment: Microsoft Visual Studio Code

Hardware Requirements:

- Raspberry Pi 3 Model B plus, (Single Board Computer)
    - Broadcom BCM2837B0, ARMv8 64-bit SoC @ 1.4GHz
    - 1GB LPDDR2 RAM
    - Fast Ethernet
    - 2.4/5GHz wireless LAN
    - Bluetooth 4.2
- Pi Camera v1.3
- Micro SD (8GB)

4. TECHNOLOGY INSIGHT

4.1 OpenCV

OpenCV (Open-Source Computer Vision Library) is an open-source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. Being a BSD-licensed product, OpenCV makes it easy for businesses to utilize and modify the code.

The library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms. These algorithms can be used to detect and recognize faces, identify objects, classify human actions in videos, track camera movements, track moving objects, extract 3D models of objects, produce 3D point clouds from stereo cameras, stitch images together to produce a high resolution image of an entire scene, find similar images from an image database, remove red eyes from images taken using flash, follow eye movements, recognize scenery and establish markers to overlay it with augmented reality, etc. OpenCV has more than 47 thousand people of user community and estimated number of downloads exceeding 18 million. The library is used extensively in companies, research groups and by governmental bodies.

Along with well-established companies like Google, Yahoo, Microsoft, Intel, IBM, Sony, Honda, Toyota that employ the library, there are many start-ups such as Applied Minds, VideoSurf, and Zeitera, that make extensive use of OpenCV. OpenCV's deployed uses span the range from stitching street view images together, detecting intrusions in surveillance video in Israel, monitoring mine equipment in China, helping robots navigate and pick up objects at Willow Garage, detection of swimming pool drowning accidents in Europe, running interactive art in Spain and New York, checking runways for debris in Turkey, inspecting labels on products in factories around the world on to rapid face detection in Japan.

It has C++, Python, Java and MATLAB interfaces and supports Windows, Linux, Android and Mac OS. OpenCV leans mostly towards real-time vision applications and takes advantage of MMX and SSE instructions when available. A full-featured CUDA and OpenCL interfaces are being actively developed right now. There are over 500 algorithms and about 10 times as many functions that compose or support those algorithms. OpenCV is written natively in C++ and has a templated interface that works seamlessly with STL containers.

4.1.1 Applications

OpenCV's application areas include:

- 2D and 3D feature toolkits
- Egomotion estimation
- Facial recognition system
- Gesture recognition
- Human–computer interaction (HCI)
- Mobile robotics
- Motion understanding
- Object detection
- Segmentation and recognition
- Stereopsis stereo vision: depth perception from 2 cameras
- Structure from motion (SFM)
- Motion tracking
- Augmented reality

To support some of the above areas, OpenCV includes a statistical machine learning library that contains:

- Boosting
- Decision tree learning
- Gradient boosting trees
- Expectation-maximization algorithm
- k-nearest neighbours' algorithm
- Naive Bayes classifier
- Artificial neural networks
- Random forest
- Support vector machine (SVM)
- Deep neural networks (DNN)

4.1.2 Face Detection

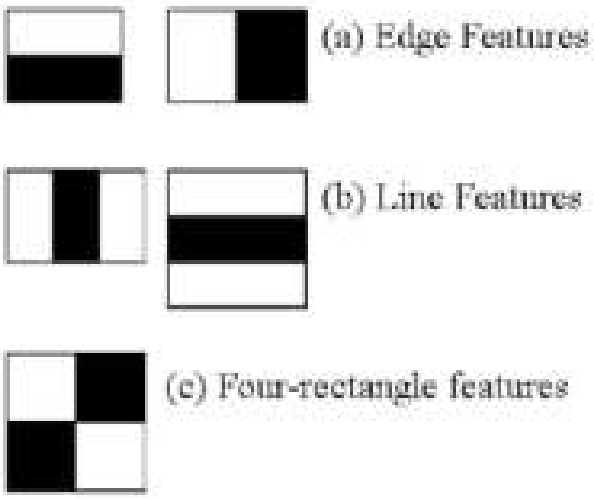
In computer vision, one essential problem we are trying to figure out is to automatically detect objects in an image without human intervention. Face detection can be thought of as such a problem where we detect human faces in an image. There may be slight differences in the faces of humans but overall, it is safe to say that there are certain features that are associated with all the human faces. There are various face detection algorithms but Viola-Jones Algorithm is one of the oldest methods that is also used today and we will use the same later in the article. You can go through the Viola-Jones Algorithm after completing this article as I'll link it at the end of this article.

Face detection is usually the first step towards many face-related technologies, such as face recognition or verification. However, face detection can have very useful applications. The most successful application of face detection would probably be photo taking. When you take a photo of your friends, the face detection algorithm built into your digital camera detects where the faces are and adjusts the focus accordingly.

4.1.3 Object Detection using Haar Cascades

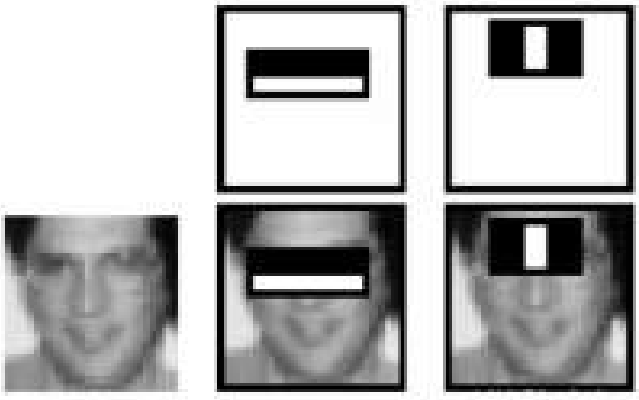
Object Detection using Haar feature-based cascade classifiers is an effective method proposed by Paul Viola and Michael Jones in the 2001 paper, "Rapid Object Detection using a Boosted Cascade of Simple Features". It is a machine learning based approach in which a cascade function is trained from a lot of positive and negative images. It is then used to detect objects in other images.

Here we will work with face detection. Initially, the algorithm needs a lot of positive images (images of faces) and negative images (images without faces) to train the classifier. Then we need to extract features from it. For this, Haar features shown in below image are used. They are just like our convolutional kernel. Each feature is a single value obtained by subtracting the sum of pixels under the white rectangle from the sum of pixels under the black rectangle.



Now all possible sizes and locations of each kernel are used to calculate plenty of features. For each feature calculation, we need to find the sum of the pixels under the white and black rectangles. To solve this, they introduced the integral images. It simplifies calculation of the sum of the pixels, how large may be the number of pixels, to an operation involving just four pixels.

But among all these features we calculated, most of them are irrelevant. For example, consider the image below. Top row shows two good features. The first feature selected seems to focus on the property that the region of the eyes is often darker than the region of the nose and cheeks. The second feature selected relies on the property that the eyes are darker than the bridge of the nose. But the same windows applying on cheeks or any other place is irrelevant. So how do we select the best features out of 160000+ features? It is achieved by Adaboost.



For this, we apply each and every feature on all the training images. For each feature, it finds the best threshold which will classify the faces to positive and negative. But obviously, there will be errors or misclassifications. We select the features with minimum error rate, which

means they are the features that best classifies the face and non-face images. (The process is not as simple as this. Each image is given an equal weight in the beginning. After each classification, weights of misclassified images are increased. Then again same process is done. New error rates are calculated. Also new weights. The process is continued until required accuracy or error rate is achieved or required number of features are found).

Final classifier is a weighted sum of these weak classifiers. It is called weak because it alone can't classify the image, but together with others forms a strong classifier. The paper says even 200 features provide detection with 95% accuracy. Their final setup had around 6000 features. (Imagine a reduction from 160000+ features to 6000 features. That is a big gain).

So now you take an image. Take each 24x24 window. Apply 6000 features to it. Check if it is face or not. Wow... Wow... Isn't it a little inefficient and time consuming? Yes, it is. Authors have a good solution for that.

In an image, most of the image region is non-face region. So, it is a better idea to have a simple method to check if a window is not a face region. If it is not, discard it in a single shot. Don't process it again. Instead focus on region where there can be a face. This way, we can find more time to check a possible face region.

For this they introduced the concept of **Cascade of Classifiers**. Instead of applying all the 6000 features on a window, group the features into different stages of classifiers and apply one-by-one. (Normally first few stages will contain very a smaller number of features). If a window fails the first stage, discard it. We don't consider remaining features on it. If it passes, apply the second stage of features and continue the process. The window which passes all stages is a face region. How is the plan!!!

Authors' detector had 6000+ features with 38 stages with 1, 10, 25, 25 and 50 features in first five stages. (Two features in the above image is actually obtained as the best two features from Adaboost). According to authors, on an average, 10 features out of 6000+ are evaluated per sub-window.

So, this is a simple intuitive explanation of how Viola-Jones face detection works. Read paper for more details.

4.1.4 Face Recognition

Thanks to OpenCV, coding facial recognition is now easier than ever. There are three easy steps to computer coding facial recognition, which are similar to the steps that our brains use for recognizing faces.

These steps are:

1. Data Gathering: Gather face data (face images in this case) of the persons you want to identify.
2. Train the Recognizer: Feed that face data and respective names of each face to the recognizer so that it can learn.
3. Recognition: Feed new faces of that people and see if the face recognizer you just trained recognizes them.

OpenCV has three built-in face recognizers and thanks to its clean coding, you can use any of them just by changing a single line of code. Here are the names of those face recognizers and their OpenCV calls:

- EigenFaces – cv2.face.createEigenFaceRecognizer()
- FisherFaces – cv2.face.createFisherFaceRecognizer()
- Local Binary Patterns Histograms (LBPH) – cv2.face.createLBPHFaceRecognizer()

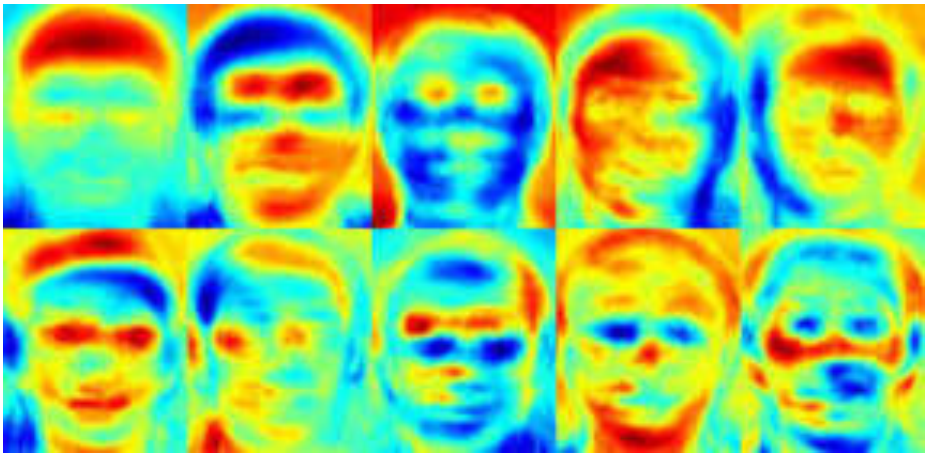
Eigenfaces face recognizer –

This algorithm considers the fact that **not all parts of a face are equally important or useful for face recognition**. Indeed, when you look at someone, you recognize that person by his distinct features, like the eyes, nose, cheeks or forehead; and how they vary respect to each other.

In that sense, you are focusing on the areas of maximum change. For example, from the eyes to the nose there is a significant change, and same applies from the nose to the mouth. When you look at multiple faces, you compare them by looking at these areas, because by catching the maximum variation among faces, they help you differentiate one face from the other.

In this way, is how EigenFaces recognizer works. It looks at all the training images of all the people as a whole and tries to extract the components which are relevant and useful and discards the rest. These important features are called **principal components**.

Below is an image showing the variance extracted from a list of faces.



So, EigenFaces recognizer trains itself by extracting principal components, but it also keeps a record of which ones belong to which person. Thus, whenever you introduce a new image to the algorithm, it repeats the same process as follows:

1. Extract the principal components from the new picture.
2. Compare those features with the list of elements stored during training.
3. Find the ones with the best match.
4. Return the 'person' label associated with that best match component.

FisherFaces face recognizer –

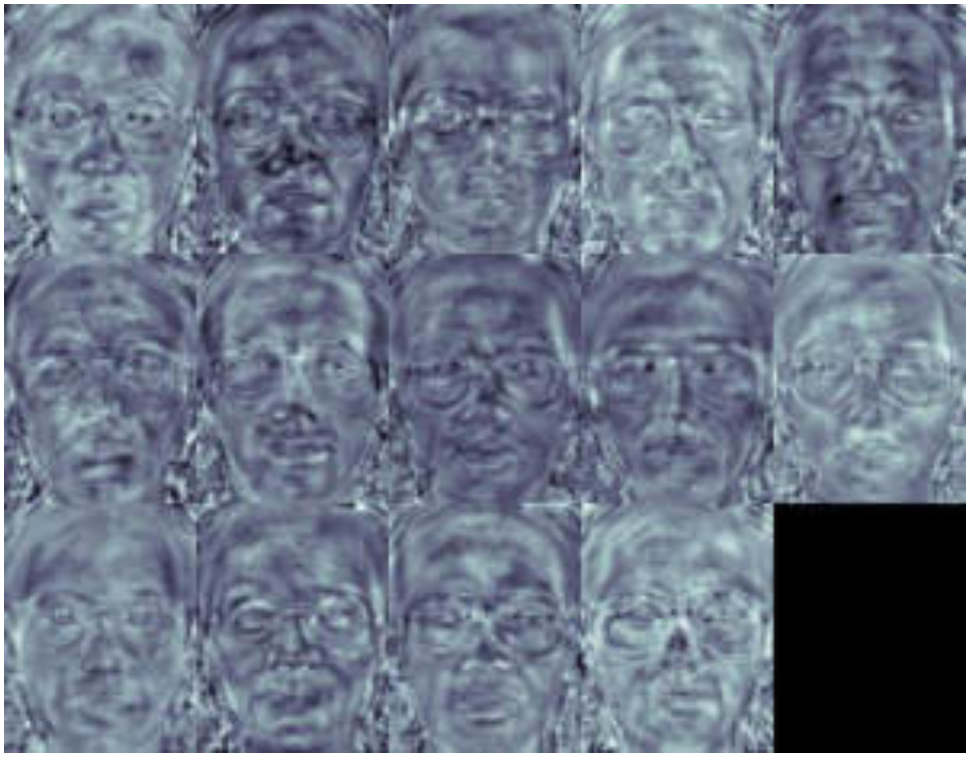
This algorithm is an improved version of the last one. As we just saw, EigenFaces looks at all the training faces of all the people at once and finds principal components from all of them combined. By doing that, it doesn't focus on the features that discriminate one individual from another. Instead, it concentrates on the ones that represent all the faces of all the people in the training data, as a whole.

Consider the lighting changes in following images:

Since EigenFaces also finds illumination as a useful component, it will find this variation very relevant for face recognition and may discard the features of the other people's faces, considering them less useful. In the end, the variance that EigenFaces has extracted represents just one individual's facial features.

Precisely, **FisherFaces** face recognizer algorithm extracts principal components that differentiate one person from the others. In that sense, an individual's components do not dominate (become more useful) over the others.

Below is an image of principal components using FisherFaces algorithm.

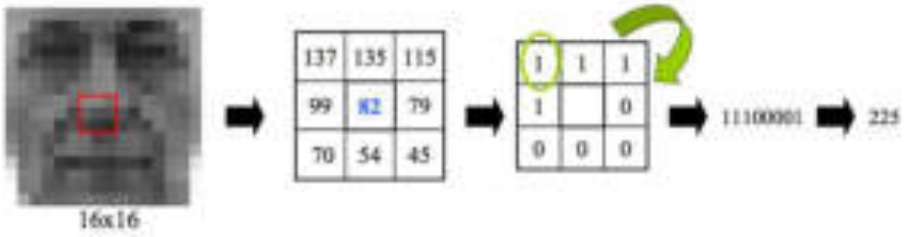


One thing to note here is that FisherFaces only prevents features of one person from becoming dominant, but it still considers illumination changes as a useful feature. We know that light variation is not a useful feature to extract as it is not part of the actual face.

LBPH face recognizer –

Take a 3×3 window and move it across one image. At each move (each local part of the picture), compare the pixel at the center, with its surrounding pixels. Denote the neighbors with intensity value less than or equal to the center pixel by 1 and the rest by 0.

After you read these 0/1 values under the 3×3 window in a clockwise order, you will have a binary pattern like 11100011 that is local to a particular area of the picture. When you finish doing this on the whole image, you will have a list of **local binary patterns**.



The LBP histograms extracted from each sub-region are used for calculation and combined into a single, histogram with spatial advanced features defined as:

$$H_{i,j} = \sum_{x,y} I\left\{f_l(x,y) = i\right\} I\left\{(x,y) \in R_j\right\}, i = 0,\ldots,n-1, j = 0,\ldots,m-1.$$

Where:

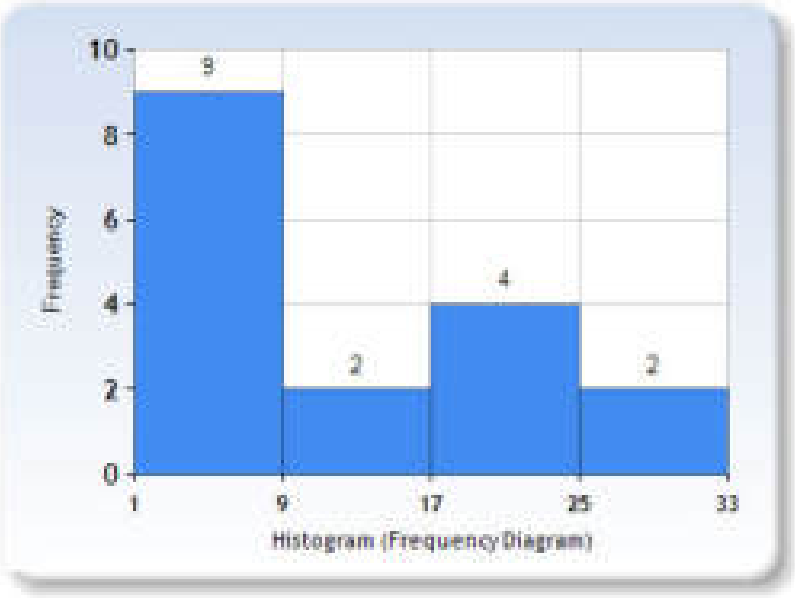
n: is the number of different labels produced by the LBP operator.

m: is the number of sub-regions. and I {…} is defined as follow:

$$I\{A\} = \begin{cases} 1, & \text{A is true} \\ 0, & \text{A is false.} \end{cases}$$

LBP conversion to binary. Source: López & Ruiz; Local Binary Patterns applied to Face Detection and Recognition.

Now, after you get a list of local binary patterns, you convert each one into a decimal number using binary to decimal conversion (as shown in above image) and then you make a histogram of all of those decimal values. A sample histogram looks like this:

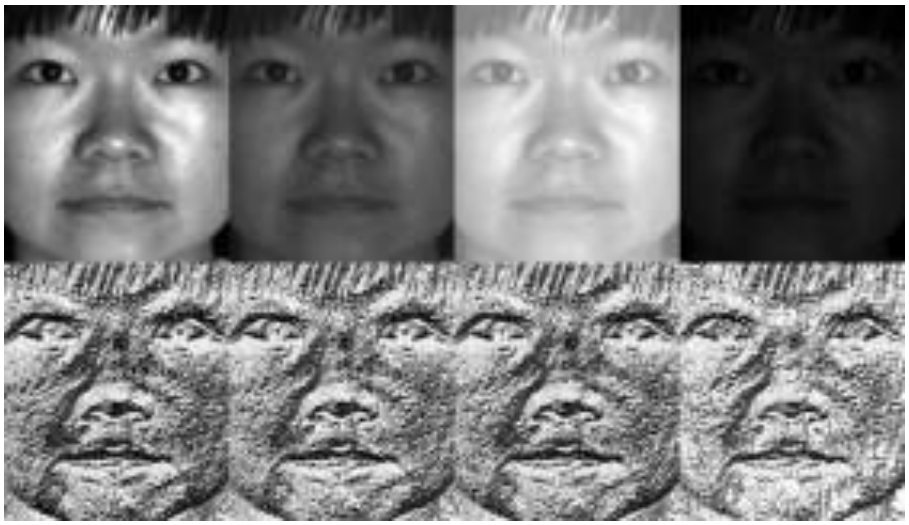


In the end, you will have one histogram for each face in the training data set. That means that if there were 100 images in the training data set then LBPH will extract 100 histograms after training and store them for later recognition. Remember, the algorithm also keeps track of which histogram belongs to which person.

Later during recognition, the process is as follows:

- Feed a new image to the recognizer for face recognition.
- The recognizer generates a histogram for that new picture.
- It then compares that histogram with the histograms it already has.
- Finally, it finds the best match and returns the person label associated with it.

Below is a group of faces and their respective local binary patterns images. You can see that the **LBP faces are not affected by changes in light conditions**:



4.2 SMTP

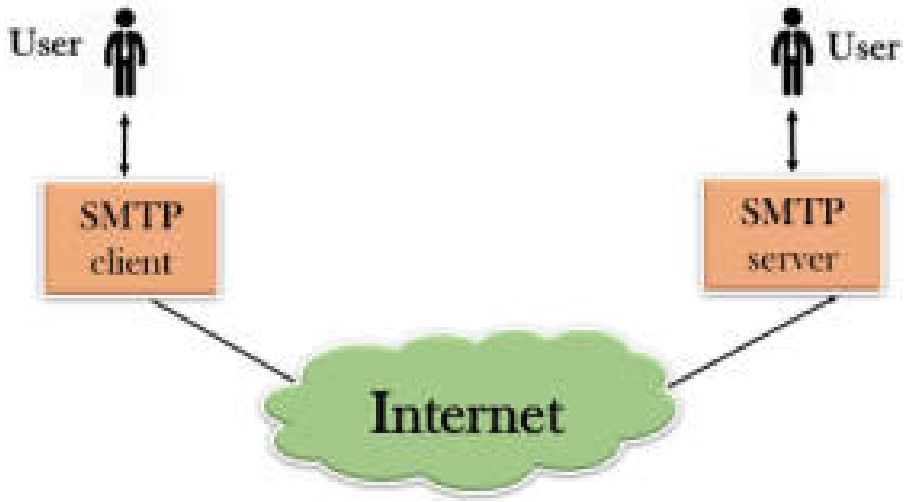
SMTP stands for Simple Mail Transfer Protocol. SMTP is a set of communication guidelines that allow software to transmit an electronic mail over the internet is called Simple Mail Transfer Protocol. It is a program used for sending messages to other computer users based on e-mail addresses. It provides a mail exchange between users on the same or different computers, and it also supports:

It can send a single message to one or more recipients.

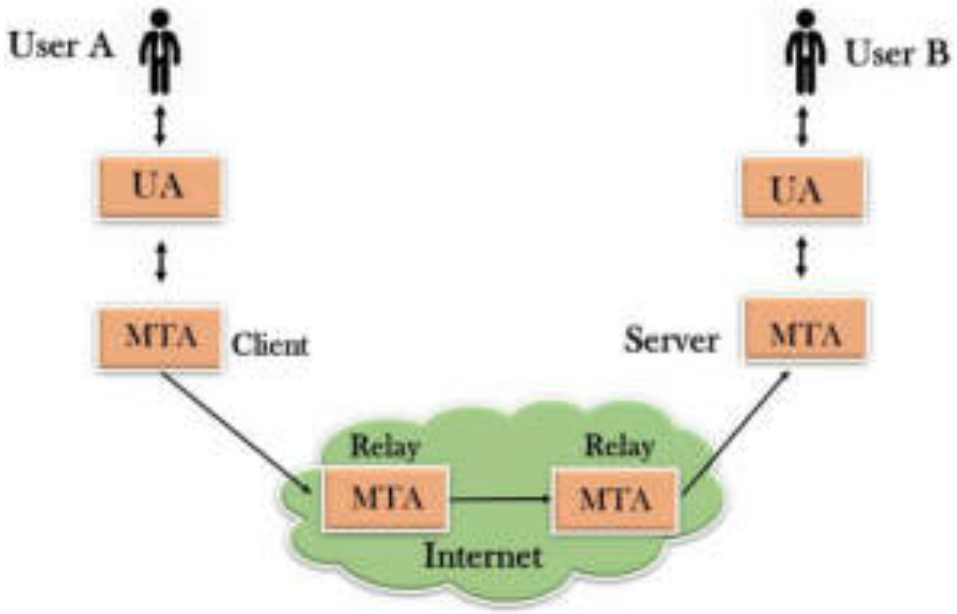
Sending message can include text, voice, video or graphics. It can also send the messages on networks outside the internet.

The main purpose of SMTP is used to set up communication rules between servers. The servers have a way of identifying themselves and announcing what kind of communication they are trying to perform. They also have a way of handling the errors such as incorrect email address. For example, if the recipient address is wrong, then receiving server reply with an error message of some kind.

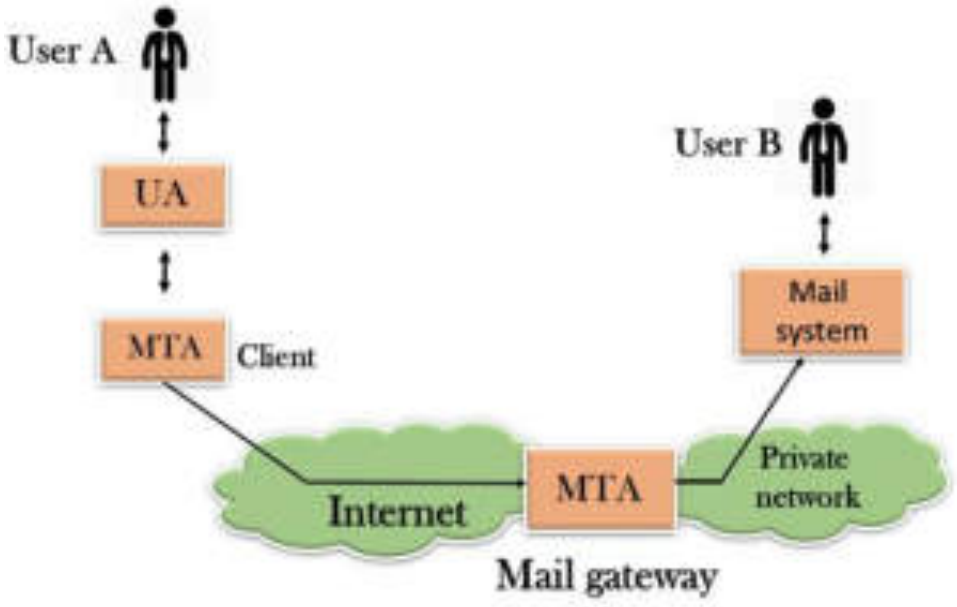
4.2.1 Components of SMTP



- First, we will break the SMTP client and SMTP server into two components such as user agent (UA) and mail transfer agent (MTA). The user agent (UA) prepares the message, creates the envelope and then puts the message in the envelope. The mail transfer agent (MTA) transfers this mail across the internet.



- The relaying system without TCP/IP protocol can also be used to send the emails to users, and this is achieved by the use of the mail gateway. The mail gateway is a relay MTA that can be used to receive an email.



4.2.2 Working of SMTP

1. **Composition of Mail**: A user sends an e-mail by composing an electronic mail message using a Mail User Agent (MUA). Mail User Agent is a program which is used to send and receive mail. The message contains two parts: body and header. The body is the main part of the message while the header includes information such as the sender and recipient address. The header also includes descriptive information such as the subject of the message. In this case, the message body is like a letter and header is like an envelope that contains the recipient's address.
2. **Submission of Mail**: After composing an email, the mail client then submits the completed e-mail to the SMTP server by using SMTP on TCP port 25.
3. **Delivery of Mail**: E-mail addresses contain two parts: username of the recipient and domain name. For example, vivek@gmail.com, where "vivek" is the username of the recipient and "gmail.com" is the domain name. If the domain name of the recipient's email address is different from the sender's domain name, then MSA will send the mail to the Mail Transfer Agent (MTA). To relay the email, the MTA will find the target domain. It checks the MX record from Domain Name System to obtain the target domain. The MX record contains the domain name and IP address of the recipient's domain. Once the record is located, MTA connects to the exchange server to relay the message.
4. **Receipt and Processing of Mail**: Once the incoming message is received, the exchange server delivers it to the incoming server (Mail Delivery Agent) which stores the e-mail where it waits for the user to retrieve it.
5. **Access and Retrieval of Mail**: The stored email in MDA can be retrieved by using MUA (Mail User Agent). MUA can be accessed by using login and password.

4.3 EMAIL

The email package is a library for managing email messages. It is specifically not designed to do any sending of email messages to SMTP (RFC 2821), NNTP, or other servers; those are functions of modules such as smtplib and nntplib. The email package attempts to be as RFC-compliant as possible, supporting RFC 5322 and RFC 6532, as well as such MIME-related RFCs as RFC 2045, RFC 2046, RFC 2047, RFC 2183, and RFC 2231.

The overall structure of the email package can be divided into three major components, plus a fourth component that controls the behavior of the other components.

The central component of the package is an “object model” that represents email messages. An application interacts with the package primarily through the object model interface defined in the message sub-module. The application can use this API to ask questions about an existing email, to construct a new email, or to add or remove email subcomponents that themselves use the same object model interface. That is, following the nature of email messages and their MIME subcomponents, the email object model is a tree structure of objects that all provide the EmailMessage API.

The other two major components of the package are the parser and the generator. The parser takes the serialized version of an email message (a stream of bytes) and converts it into a tree of EmailMessage objects. The generator takes an EmailMessage and turns it back into a serialized byte stream. (The parser and generator also handle streams of text characters, but this usage is discouraged as it is too easy to end up with messages that are not valid in one way or another.)

The control component is the policy module. Every EmailMessage, every generator, and every parser has an associated policy object that controls its behavior. Usually, an application only needs to specify the policy when an EmailMessage is created, either by directly instantiating an EmailMessage to create a new email, or by parsing an input stream using a parser. But the policy can be changed when the message is serialized using a generator. This allows, for example, a generic email message to be parsed from disk, but to serialize it using standard SMTP settings when sending it to an email server.

The email package does its best to hide the details of the various governing RFCs from the application. Conceptually the application should be able to treat the email message as a structured tree of unicode text and binary attachments, without having to worry about how

these are represented when serialized. In practice, however, it is often necessary to be aware of at least some of the rules governing MIME messages and their structure, specifically the names and nature of the MIME “content types” and how they identify multipart documents. For the most part this knowledge should only be required for more complex applications, and even then, it should only be the high-level structure in question, and not the details of how those structures are represented. Since MIME content types are used widely in modern internet software (not just email), this will be a familiar concept to many programmers.

The following sections describe the functionality of the email package. We start with the message object model, which is the primary interface an application will use, and follow that with the parser and generator components. Then we cover the policy controls, which completes the treatment of the main components of the library.

The next three sections cover the exceptions the package may raise and the defects (non-compliance with the RFCs) that the parser may detect. Then we cover the header registry and the content manager sub-components, which provide tools for doing more detailed manipulation of headers and payloads, respectively. Both of these components contain features relevant to consuming and producing non-trivial messages, but also document their extensibility APIs, which will be of interest to advanced applications.

Following those is a set of examples of using the fundamental parts of the APIs covered in the preceding sections.

The foregoing represents the modern (Unicode friendly) API of the email package. The remaining sections, starting with the Message class, cover the legacy compat32 API that deals much more directly with the details of how email messages are represented. The compat32 API does not hide the details of the RFCs from the application, but for applications that need to operate at that level, they can be useful tools. This documentation is also relevant for applications that are still using the compat32 API for backward compatibility reasons.

4.4 PYTHON

Python is a general purpose, dynamic, high level and interpreted programming language. It supports Object Oriented programming approach to develop applications. It is simple and easy to learn and provides lots of high-level data structures. Python is easy to learn yet powerful and versatile scripting language which makes it attractive for Application Development. Python's syntax and dynamic typing with its interpreted nature, makes it an ideal language for scripting and rapid application development. Python supports multiple programming pattern, including object oriented, imperative and functional or procedural programming styles. Python is not intended to work on special area such as web programming. That is why it is known as multipurpose because it can be used with web, enterprise, 3D CAD etc. Python makes the development and debugging fast because there is no compilation step included in python development and edit-test-debug cycle is very fast.

4.4.1 Features

- Python provides lots of features that are listed below.
- **Easy to Learn and Use**: Python is easy to learn and use. It is developer-friendly and high-level programming language.
- **Expressive Language**: Python language is more expressive means that it is more understandable and readable.
- **Interpreted Language**: Python is an interpreted language i.e., interpreter executes the code line by line at a time. This makes debugging easy and thus suitable for beginners.
- **Cross-platform Language**: Python can run equally on different platforms such as Windows, Linux, Unix and Macintosh etc. So, we can say that Python is a portable language.
- **Free and Open Source**: Python language is freely available at official web address. The source-code is also available. Therefore, it is opensource.
- **Object-Oriented Language**: Python supports object-oriented language and concepts of classes and objects come into existence.
- **Extensible**: It implies that other languages such as C/C++ can be used to compile the code and thus it can be used further in our python code.
- **Large Standard Library**: Python has a large and broad library and provides rich set of module and functions for rapid application development.
- **GUI Programming Support**: GUI applications can be developed using Python.
- **Integrated**: It can be easily integrated with languages like C, C++, JAVA etc.

4.4.2 Applications

Python is known for its general-purpose nature that makes it applicable in almost each domain of software development. Python as a whole can be used in any sphere of development. Here, we are specifying applications areas where python can be applied.

- **Web Applications –** We can use Python to develop web applications. It provides libraries to handle internet protocols such as HTML and XML, JSON, Email processing, request, beautifulSoup, Feed parser etc. It also provides Frameworks such as Django, Pyramid, Flask etc. to design and develop web-based applications. Some important developments are: Python WikiEngines, Pocoo, Python BlogSoftware etc.
- **Desktop GUI Applications –** Python provides Tk GUI library to develop user interface in python-based application. Some other useful toolkits wxWidgets, Kivy, pyqt that are useable on several platforms. The Kivy is popular for writing multi touch applications.
- **Software Development –** Python is helpful for software development process. It works as a support language and can be used for build control and management, testing etc.
- **Scientific and Numeric –** Python is popular and widely used in scientific and numeric computing. Some useful library and package are SciPy, Pandas, IPython etc. SciPy is group of packages of engineering, science and mathematics.
- **Business Applications –** Python is used to build business applications like ERP and e-commerce systems. Tryton is a high-level application platform.
- **Console Based Application –** We can use Python to develop console-based applications. For example: IPython.
- **Audio or Video based Applications –** Python is awesome to perform multiple tasks and can be used to develop multimedia applications. Some of real applications are: TimPlayer, cplay etc.
- **3D CAD Applications –** To create CAD application Fandango is a real application which provides full features of CAD.
- **Enterprise Applications –** Python can be used to create applications which can be used within an Enterprise or an Organization. Some real time applications are: OpenErp, Tryton, Picalo etc.
- **Applications for Images –** Using Python several applications can be developed for image. Applications developed are: VPython, Gogh, img Seek etc.

4.5 Visual Studio Code

Visual Studio Code is a source-code editor made by Microsoft for Windows, Linux and macOS. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git. Users can change the theme, keyboard shortcuts, preferences, and install extensions that add additional functionality.

Microsoft has released most of Visual Studio Code's source code on the Microsoft/vscode repository of GitHub using the "Code – OSS" name, under the permissive MIT License, while the releases by Microsoft are proprietary freeware.

In the Stack Overflow 2019 Developer Survey, Visual Studio Code was ranked the most popular developer environment tool, with 50.7% of 87,317 respondents reporting that they use it.

Visual Studio Code was first announced on April 29, 2015, by Microsoft at the 2015 Build conference. A Preview build was released shortly thereafter.

On November 18, 2015, Visual Studio Code was released under the MIT License, having its source code available on GitHub. Extension support was also announced. On April 14, 2016, Visual Studio Code graduated from the public preview stage and was released to the Web.

4.5.1 Features

Visual Studio Code is a source-code editor that can be used with a variety of programming languages, including Java, JavaScript, Go, Nodejs, Python and C. It is based on the Electron framework, which is used to develop Node.js Web applications that run on the Blink layout engine. Visual Studio Code employs the same editor component (codenamed "Monaco") used in Azure DevOps (formerly called Visual Studio Online and Visual Studio Team Services).

Instead of a project system, it allows users to open one or more directories, which can then be saved in workspaces for future reuse. This allows it to operate as a language-agnostic code editor for any language. It supports a number of programming languages and a set of features that differs per language. Unwanted files and folders can be excluded from the project tree via the settings. Many Visual Studio Code features are not exposed through menus or the user interface but can be accessed via the command palette.

Visual Studio Code can be extended via extensions, available through a central repository. This includes additions to the editor and language support. A notable feature is the ability to create

extensions that add support for new languages, themes, and debuggers, perform static code analysis, and add code linters using the Language Server Protocol.

Visual Studio Code includes multiple extensions for FTP, allowing the software to be used as a free alternative for web development. Code can be synced between the editor and the server, without downloading any extra software.

Visual Studio Code allows users to set the code page in which the active document is saved, the newline character, and the programming language of the active document. This allows it to be used on any platform, in any locale, and for any given programming language.

4.5.2 Language support

Out-of-the-box, Visual Studio Code includes basic support for most common programming languages. This basic support includes syntax highlighting, bracket matching, code folding, and configurable snippets. Visual Studio Code also ships with IntelliSense for JavaScript, TypeScript, JSON, CSS, and HTML, as well as debugging support for Node.js. Support for additional languages can be provided by freely available extensions on the VS Code Marketplace.

4.5.3 Data collection

Visual Studio Code collects usage data and sends it to Microsoft, although this can be disabled. In addition, because of the open-source nature of the application, the telemetry code is accessible to the public, who can see exactly what is collected. According to Microsoft, the data is shared with Microsoft-controlled affiliates and subsidiaries, although law enforcement may request it as part of a legal process.

4.5.4 Version control

Source control is a built-in feature of Visual Studio Code. It has a dedicated tab inside of the menu bar where you can access version control settings and view changes made to the current project. To use the feature, you must link Visual Studio Code to any supported version control system (Git, Apache Subversion, Perforce, etc.). This allows you to create repositories as well as make push and pull requests directly from the Visual Studio Code program.

4.6 Raspberry Pi

Raspberry Pi is a low cost, credit card-sized single-board computer developed by the Raspberry Pi Foundation in the United Kingdom. It is known as a single-board computer as it contains Central Processing Unit (CPU), Random Access Memory, Graphics Processing Unit (GPU) and other peripherals included in a single circuit. It was initially designed to improve programming skills in pre-university level students. It has a 64-bit ARM processor and uses Raspbian, a Linux distributed Debian based 32-bit operating system. Raspberry Pi can also be plugged into a TV or monitor together with a keyboard, mouse, speakers, and camera can be used like a computer. Although it is not as powerful as a normal desktop computer, for the size of a credit card, it is considered powerful enough for embedded systems. There are several generations of Raspberry Pi, namely Raspberry Pi 1, Raspberry Pi 2, Raspberry Pi 3, and Raspberry Pi Zero and the latest high-end generation of the Pi was released in a single model called Raspberry Pi 3 Model B with an additional on-board wireless LAN 802.11n and Bluetooth connectivity.

The Raspberry Pi 3 Model B+ is the latest product in the Raspberry Pi 3 range, boasting a 64-bit quad core processor running at 1.4GHz, dual-band 2.4GHz and 5GHz wireless LAN, Bluetooth 4.2/BLE, faster Ethernet, and PoE capability via a separate PoE HAT the dual-band wireless LAN comes with modular compliance certification, allowing the board to be designed into end products with significantly reduced wireless LAN compliance testing, improving both cost and time to market. The Raspberry Pi 3 Model B+ maintains the same mechanical footprint as both the Raspberry Pi 2 Model B and the Raspberry Pi 3 Model B.

4.7 Pi Camera

This Raspberry Pi Camera Module is a custom designed add-on for Raspberry Pi. It attaches to Raspberry Pi by way of one of the two small sockets on the board upper surface. This interface uses the dedicated CSI interface, which was designed especially for interfacing with cameras. The CSI bus is capable of extremely high data rates, and it exclusively carries pixel data.

The camera connects to the BCM2835 processor on the Pi via the CSI bus, a higher bandwidth link that carries pixel data from the camera back to the processor. This bus travels along the ribbon cable that attaches the camera board to the Pi.

5. SYSTEM DESIGN

5.1 Unified Modeling Language

UML (Unified Modeling Language) is a general-purpose, graphical modeling language in the field of Software Engineering. UML is used to specify, visualize, construct, and document the artifacts (major elements) of the software system. It was initially developed by Grady Booch, Ivar Jacobson, and James Rumbaugh in 1994-95 at Rational software, and its further development was carried out through 1996. In 1997, it got adopted as a standard by the Object Management Group.

UML (Unified Modeling Language) is a general-purpose, graphical modeling language in the field of Software Engineering. UML is used to specify, visualize, construct, and document the artifacts (major elements) of the software system. It was initially developed by Grady Booch, Ivar Jacobson, and James Rumbaugh in 1994-95 at Rational software, and its further development was carried out through 1996. In 1997, it got adopted as a standard by the Object Management Group.

5.1.1 Goals of UML

- Since it is a general-purpose modeling language, it can be utilized by all the modelers.
- UML came into existence after the introduction of object-oriented concepts to systemize and consolidate the object-oriented development, due to the absence of standard methods at that time.
- The UML diagrams are made for business users, developers, ordinary people, or anyone who is looking forward to understand the system, such that the system can be software or non-software.
- Thus, it can be concluded that the UML is a simple modeling approach that is used to model all the practical systems.

5.1.2 Characteristics of UML

- It is a generalized modeling language.
- It is distinct from other programming languages like C++, Python, etc.
- It is interrelated to object-oriented analysis and design.
- It is used to visualize the workflow of the system.
- It is a pictorial language, used to generate powerful modeling artifacts.

5.1.3 Conceptual Modeling

Before moving ahead with the concept of UML, we should first understand the basics of the conceptual model.

A conceptual model is composed of several interrelated concepts. It makes it easy to understand the objects and how they interact with each other. This is the first step before drawing UML diagrams.

Following are some object-oriented concepts that are needed to begin with UML:

- **Object:** An object is a real-world entity. There are many objects present within a single system. It is a fundamental building block of UML.
- **Class:** A class is a software blueprint for objects, which means that it defines the variables and methods common to all the objects of a particular type.
- **Abstraction:** Abstraction is the process of portraying the essential characteristics of an object to the users while hiding the irrelevant information. Basically, it is used to envision the functioning of an object.
- **Inheritance:** Inheritance is the process of deriving a new class from the existing ones.
- **Polymorphism:** It is a mechanism of representing objects having multiple forms used for different purposes.
- **Encapsulation:** It binds the data and the object together as a single unit, enabling tight coupling between them.

5.1.4 Object-Oriented Analysis and Design

Object oriented analysis is an analysis of objects, and design means combining those identified objects. So, the main purpose of Object-oriented analysis is identifying the objects for designing a system. The analysis can also be done for an existing system. The analysis can be more efficient if we can identify the objects. Once we have identified the objects, their relationships are then identified, and the design is also produced.

The purpose of Object-oriented is given below:

- To identify the objects of a system.
- To identify their relationships.
- To make a design that is executable when the concepts of Object-oriented are employed.

Following are the steps where Object-oriented concepts are applied and implemented:

Step 1: Object-oriented Analysis

The main purpose of Object-oriented analysis is identifying the objects and describing them correctly. After the objects are identified, the designing step is easily carried out. It is a must to identify the objects with responsibilities. Here the responsibility refers to the functions performed by the objects. Each individual object has its own functions to perform. The purpose of the system is fulfilled by collaborating these responsibilities.

Step 2: Object-oriented Design

This phase mainly emphasizes on meeting the requirements. In this phase, the objects are joined together as per the intended associations. After the association is completed, the designing phase also gets complete.

Step 3: Object-oriented Implementation

This is the last phase that comes after the designing is done. It implements the design using any Object-oriented languages like C++, Java, etc.

5.1.5 Role of UML in Object-oriented design

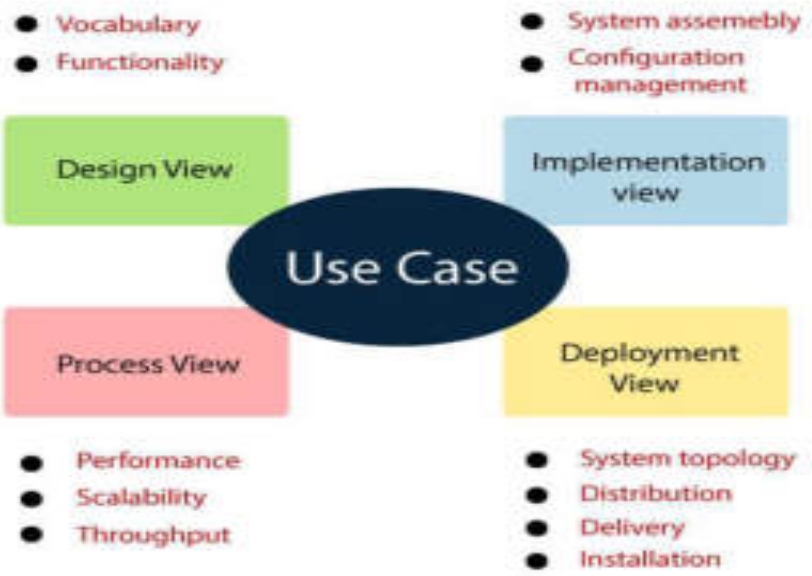
As the UML is a modeling language used to model software as well as non-software systems, but here it focuses on modeling object-oriented software applications. It is essential to understand the relation between the object-oriented design and UML. The object-oriented design can be converted into the UML as and when required. The object-oriented languages influence the programming world as they model real world objects.

The UML itself is an amalgamation of object-oriented notations like Object-Oriented Design (OOD), Object Modeling Technique (OMT), and Object-Oriented Software Engineering (OOSE). The strength of these three approaches is utilized by the UML to represent more consistency.

5.2 UML- Architecture

Any real-world system is used by different users. The users can be developers, testers, business people, analysts, and many more. Hence, before designing a system, the architecture is made with different perspectives in mind. The most important part is to visualize the system from the perspective of different viewers. The better we understand the better we can build the system.

UML plays an important role in defining different perspectives of a system. These perspectives are −



The centre is the **Use Case** view which connects all these four. A **Use Case** represents the functionality of the system. Hence, other perspectives are connected with use case.

Design of a system consists of classes, interfaces, and collaboration. UML provides class diagram, object diagram to support this.

Implementation defines the components assembled together to make a complete physical system. UML component diagram is used to support the implementation perspective.

Process defines the flow of the system. Hence, the same elements as used in Design are also used to support this perspective.

Deployment represents the physical nodes of the system that forms the hardware. UML deployment diagram is used to support this perspective.

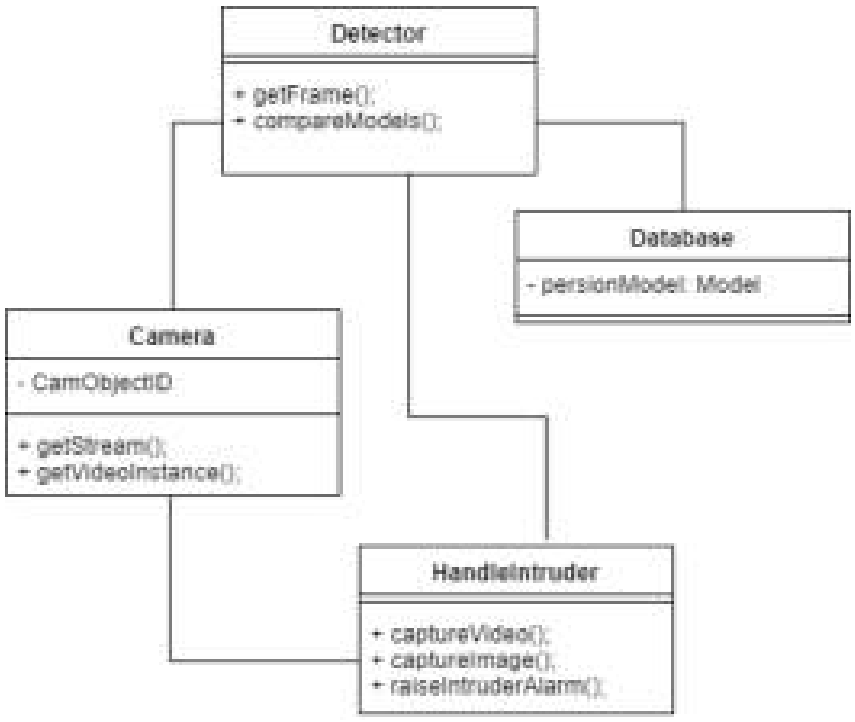
5.3 UML-Diagrams

5.3.1 UML Class Diagram

The class diagram depicts a static view of an application. It represents the types of objects residing in the system and the relationships between them. A class consists of its objects, and also it may inherit from other classes. A class diagram is used to visualize, describe, document various different aspects of the system, and also construct executable software code.

It shows the attributes, classes, functions, and relationships to give an overview of the software system. It constitutes class names, attributes, and functions in a separate compartment that helps in software development. Since it is a collection of classes, interfaces, associations, collaborations, and constraints, it is termed as a structural diagram.

Our class Diagram –



5.3.2 UML Use Case Diagram

A use case diagram is used to represent the dynamic behavior of a system. It encapsulates the system's functionality by incorporating use cases, actors, and their relationships. It models the tasks, services, and functions required by a system/subsystem of an application. It depicts the high-level functionality of a system and also tells how the user handles a system.

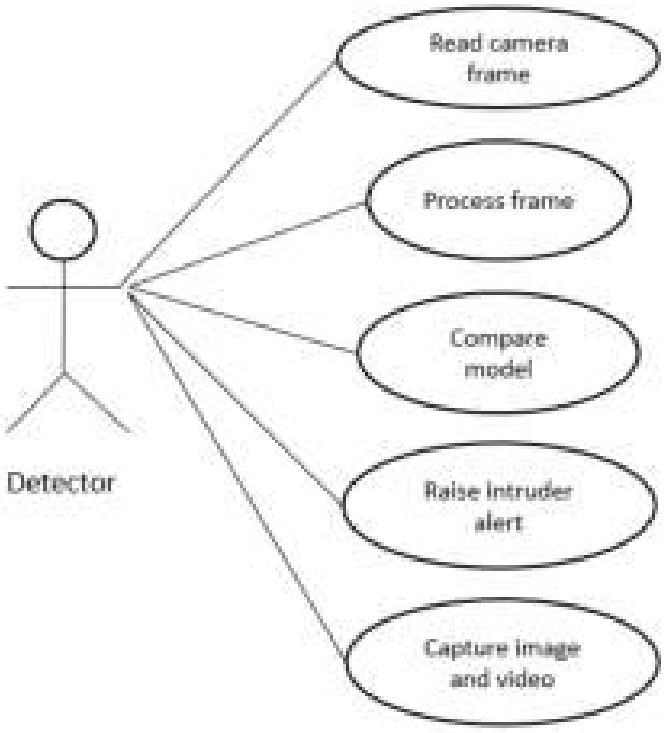
Purpose of Use Case Diagrams

The main purpose of a use case diagram is to portray the dynamic aspect of a system. It accumulates the system's requirement, which includes both internal as well as external influences. It invokes persons, use cases, and several things that invoke the actors and elements accountable for the implementation of use case diagrams. It represents how an entity from the external environment can interact with a part of the system.

Following are the purposes of a use case diagram given below:

1. It gathers the system's needs.
2. It depicts the external view of the system.
3. It recognizes the internal as well as external factors that influence the system.
4. It represents the interaction between the actors.

Our Use Case Diagram –



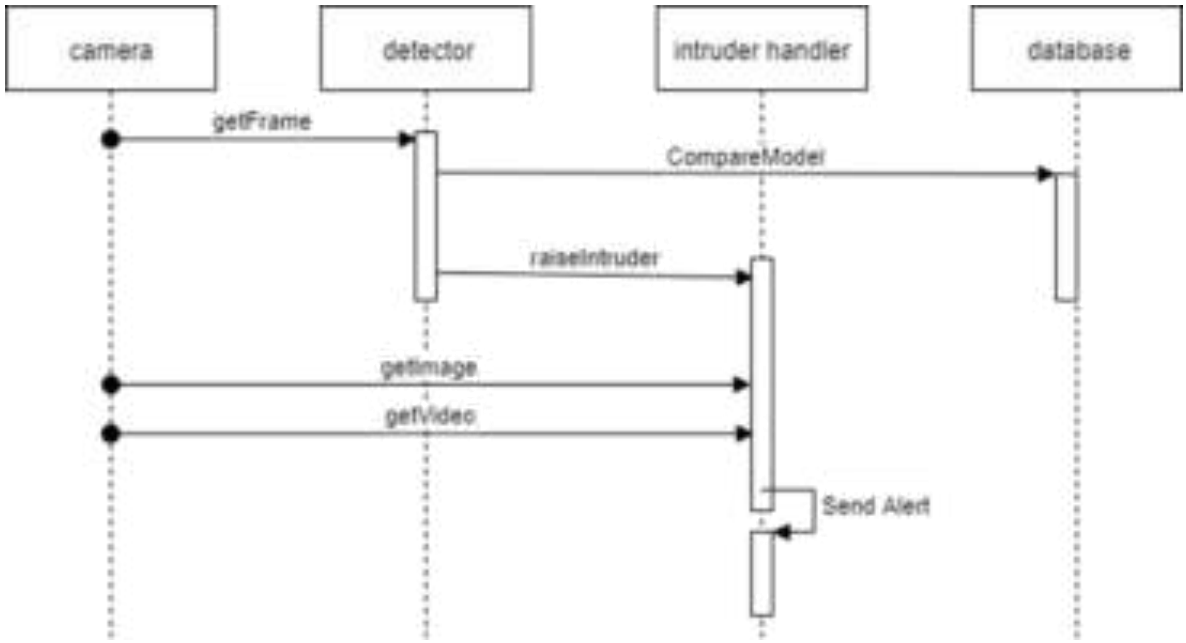
5.3.3 Sequence Diagram

The sequence diagram represents the flow of messages in the system and is also termed as an event diagram. It helps in envisioning several dynamic scenarios. It portrays the communication between any two lifelines as a time-ordered sequence of events, such that these lifelines took part at the run time. In UML, the lifeline is represented by a vertical bar, whereas the message flow is represented by a vertical dotted line that extends across the bottom of the page. It incorporates the iterations as well as branching.

Purpose of a Sequence Diagram

1. To model high-level interaction among active objects within a system.
2. To model interaction among objects inside a collaboration realizing a use case.
3. It either models' generic interactions or some certain instances of interaction.

Our Sequence Diagram –



5.3.4 UML Activity Diagram

In UML, the activity diagram is used to demonstrate the flow of control within the system rather than the implementation. It models the concurrent and sequential activities. The activity diagram helps in envisioning the workflow from one activity to another. It put emphasis on the condition of flow and the order in which it occurs. The flow can be sequential, branched, or concurrent, and to deal with such kinds of flows, the activity diagram has come up with a fork, join, etc. It is also termed as an object-oriented flowchart. It encompasses activities composed of a set of actions or operations that are applied to model the behavioral diagram.

Activity diagram constitutes following notations:

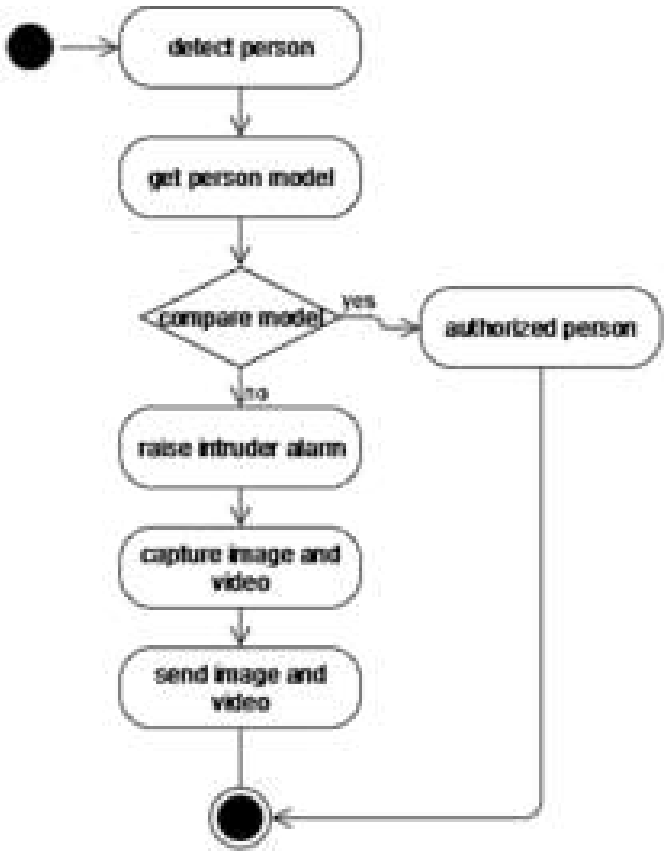
Initial State: It depicts the initial stage or beginning of the set of actions.

Final State: It is the stage where all the control flows and object flows end.

Decision Box: It makes sure that the control flow or object flow will follow only one path.

Action Box: It represents the set of actions that are to be performed.

Our Activity Diagram –



6. IMPLEMENTATION

6.1 Dataset

We are using a simple live generated dataset of 30 image samples per authorized person. The *dataset.py* module contains the method to generate the dataset sample for the person we want to add to our authorized people model.

This dataset, after creation/modification, is immediately used to train the recognizer model through *train.py* module.

6.2 Project Structure

Project/
- Cascade/
  - haarcascade_frontalface_default.xml
- Dataset/
  - Samples/
    - <User Samples.jpg>
  - names.txt
- Model/
  - model.yml
- Buffer/
- dataset.py
- intrude.py
- main.py
- recognize.py
- trainer.py

6.3 Source code

dataset.py

```
import cv2
import os
import sys
def makeDataset(cam, face_detector, face_id):
    count = 0
    while(True):
        if face_detector.empty():
            print("[ERR] Cascade load failed")
            sys.exit(1)
        ret, img = cam.read()
        gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
        faces = face_detector.detectMultiScale(gray, 1.3, 5)
        for (x, y, w, h) in faces:
            cv2.rectangle(img, (x, y), (x+w, y+h), (255, 0, 0), 2)
            count += 1
            print("[INFO] Captured {} image(s)".format(count))
            # Save the captured image into the datasets folder
            cv2.imwrite("dataset/samples/User." + str(face_id) +
                    '.' + str(count) + ".jpg", gray[y:y+h, x:x+w])
        if count >= 30:
            break
    print("[INFO] Exiting Program and cleanup stuff")
```

train.py

```
import cv2
import numpy as np
from PIL import Image
import os
def train():
    path = 'dataset/samples'
    recognizer = cv2.face.createLBPHFaceRecognizer()
    print ("[INFO] Training faces. It will take a few seconds. Wait ...")
    imagePaths = [os.path.join(path,f) for f in os.listdir(path)]
    faceSamples=[]
    ids = []
    for imagePath in imagePaths:
        PIL_img = Image.open(imagePath).convert('L') # grayscale
        img_numpy = np.array(PIL_img,'uint8')
        id = int(os.path.split(imagePath)[-1].split(".")[1])
        faceSamples.append(img_numpy)
        ids.append(id)
    recognizer.train(faceSamples, np.array(ids))
    recognizer.save('model/model.yml')
    print("[INFO] {0} faces trained. Exiting Program"
                                    .format(len(np.unique(ids))))
```

recognize.py

```
import cv2
import numpy as np
import os
import pickle

import intrude
import time

def validate(img, faces, cam, classifier):
    recognizer = cv2.face.createLBPHFaceRecognizer()
    recognizer.load('model/model.yml')
    font = cv2.FONT_HERSHEY_SIMPLEX
    gray = cv2.cvtColor(img,cv2.COLOR_BGR2GRAY)
    authorized = False
    names = dict()
    with open("dataset/names.txt") as f:
        names = f.readlines()
        #print(names)
    print("[INFO] Analysing Detected Faces. . .")
    for (x,y,w,h) in faces:
        id, confidence = recognizer.predict(gray[y:y+h,x:x+w])
        #print(id, confidence)

        # If confidence is less them 100 ==> "0" : perfect match
        if (confidence < 100 and id <= int(names[0].strip())):
            print("[INFO] Detected '{}'. . .".format(names[id].strip()))
            cv2.rectangle(img, (x, y), (x+w, y+h), (0, 255, 0), 2)
            authorized = True
        else:
            print("[INFO] Intruder Detected!")
            cv2.rectangle(img, (x, y), (x+w, y+h), (0, 0, 255), 2)
    if authorized:
        print("[INFO] Authorized person detected, no intrusion.")
        intrude.raiseIntrude(cam, img, sendVideo = False)
        print("[INFO] Program Halt for next 30sec")
        time.sleep(int(60.0 * 0.5))
        print("[INFO] Resuming program in 5Sec")
        time.sleep(int(60.0 * 1 / 12))
        print("[INFO] Resuming Program Now.")
    else:
        print("[INFO] Initializing Alert System")
        intrude.raiseIntrude(cam, img, sendVideo = True)
        print("[INFO] Program Halt for next 15sec")
        time.sleep(int(60.0 * 0.25))
        print("[INFO] Resuming program in 5Sec")
        time.sleep(int(60.0 * 1 / 12))
        print("[INFO] Resuming Program Now.")
```

intrude.py

```
import cv2
import os
import time

import smtplib
from email.mime.multipart import MIMEMultipart
from email.mime.text import MIMEText
from email.mime.base import MIMEBase
from email import encoders

def raiseIntrude(cam, img, sendVideo = True):
    fromaddr = "btech.project.mail@gmail.com"
    toaddr = "yogeshwarm10145@gmail.com"

    print("[INFO] Processing subject image")
    img_bytes = cv2.imencode('.jpg', img)[1].tobytes()

    print("[INFO] Preparing Mail for Image")
    msg_img = MIMEMultipart()
    msg_img['From'] = fromaddr
    msg_img['To'] = toaddr
    msg_img['Subject'] = "Home Security Alert!"

    if sendVideo:
        body = "There is a breach at your secure location. Take a look...\n" +
"video in its way."
    else:
        body = "A person has entered your secure location."

    msg_img.attach(MIMEText(body, 'plain'))

    filename = "Sample.jpg"
    #attachment = open("test/test (1).png", "rb")
    p = MIMEBase('application', 'octet-stream')
    p.set_payload(img_bytes)
    encoders.encode_base64(p)
    p.add_header('Content-Disposition', "attachment; filename= %s" % filename)
    msg_img.attach(p)

    print("[INFO] Sending image mail")
    s = smtplib.SMTP('smtp.gmail.com', 587)
    s.starttls()
    s.login(fromaddr, "iotproject")
    text = msg_img.as_string()
    s.sendmail(fromaddr, toaddr, text)
    print("[INFO] Image mail Sent")
    s.quit()
```

```
    if not sendVideo:
        return

    print("[INFO] Collecting Video")
    filename = "Sample-vid.mp4"
    fourcc = cv2.VideoWriter_fourcc(*'X264')
    vid = cv2.VideoWriter("buffer/" + filename, fourcc, 16, (640, 480))

    for _ in range(1000):
        _, frame = cam.read()
        vid.write(frame)
        time.sleep(1 / 16)
    vid.release()

    print("[INFO] Preparing Mail for Video")
    msg_vid = MIMEMultipart()
    msg_vid['From'] = fromaddr
    msg_vid['To'] = toaddr
    msg_vid['Subject'] = "Home Security Alert! - Video"

    body = "There is a breach at your secure location. Take a look..."
    msg_vid.attach(MIMEText(body, 'plain'))

    with open("buffer/" + filename, "rb") as video:
        vid_bytes = video.read()

    p = MIMEBase('application', 'octet-stream')
    p.set_payload(vid_bytes)
    encoders.encode_base64(p)
    p.add_header('Content-Disposition', "attachment; filename= %s" % filename)
    msg_vid.attach(p)

    print("[INFO] Sending video mail")
    s = smtplib.SMTP('smtp.gmail.com', 587)
    s.starttls()
    s.login(fromaddr, "iotproject")
    text = msg_vid.as_string()
    s.sendmail(fromaddr, toaddr, text)
    print("[INFO] Video mail Sent")
    s.quit()
    os.remove("buffer/" + filename)
```

main.py

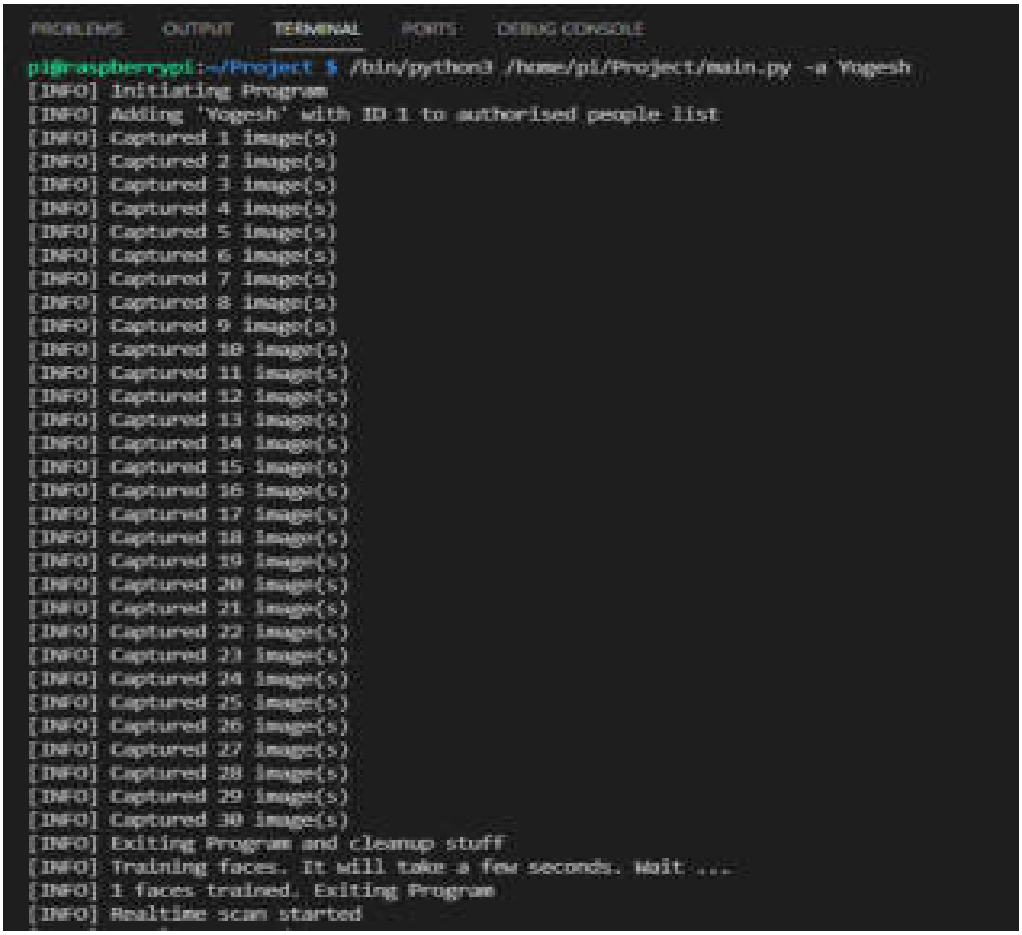
```
import cv2
import os
import argparse
import sys
import pickle

import dataset
import trainer
import recognize

os.chdir("/home/pi/Project/")

argParser = argparse.ArgumentParser()
argParser.add_argument("-a", "--addPerson", metavar="NAME", type=ascii,
        help="Add new person to the authorised people list")
argParser.add_argument("-t", "--train", action='store_true',
        help="Train the dataset before starting application")
args = argParser.parse_args()

print("[INFO] Initiating Program")

cam = cv2.VideoCapture(0)
cam.set(3, 640)
cam.set(4, 480)

classifier = 'cascade/haarcascade_frontalface_default.xml'
face_detector = cv2.CascadeClassifier(classifier)
if face_detector.empty():
    print("[ERR] Cascade load failed")
    cam.release()
    sys.exit(1)

if args.addPerson is not None:
    name = args.addPerson
    with open("dataset/names.txt") as f:
        count = int(f.readline().strip())
        names = f.readlines()
        id = count + 1
    print("[INFO] Adding {} with ID {} to authorised people list"
                    .format(name, id))
    dataset.makeDataset(cam, face_detector, id)
    names.append(name)
    with open("dataset/names.txt", "w") as f:
        f.write(str(count + 1) + '\n')
        for i in names:
            f.write(i.strip() + '\n')
    trainer.train()
```

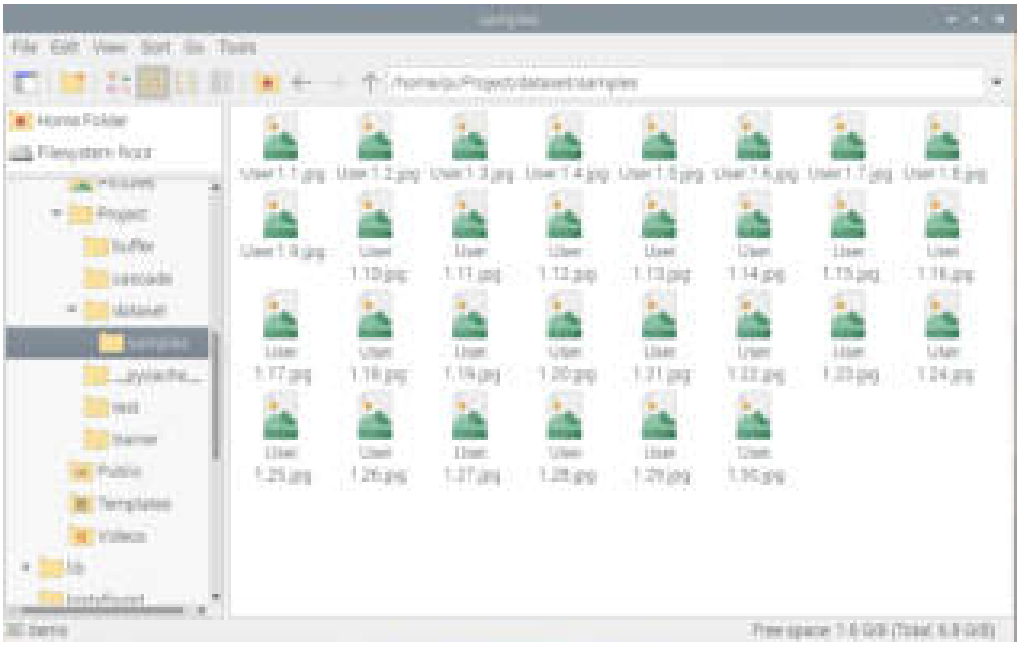
```
if args.train:
    trainer.train()

minW = 0.1*cam.get(3)
minH = 0.1*cam.get(4)
count = 0
print("[INFO] Realtime scan started")
while True:
    ret, img =cam.read()
    if not ret:
        print("[ERR] Camera read failed!")
        cam.release()
        sys.exit(1)
    gray = cv2.cvtColor(img,cv2.COLOR_BGR2GRAY)
    faces = face_detector.detectMultiScale(gray, scaleFactor = 1.2, minNeighbo
rs = 5, minSize = (int(minW), int(minH)),)
    if len(faces) != 0:
        count += 1
        #print("Count  incresed to -", count)
        if count == 5:
            print("[INFO] People Detected. . .")
            recognize.validate(img, faces, cam, face_detector)
            count = 0
    elif count != 0:
        count -= 1
        #print("Count reduced to -", count)
    #cv2.imshow("image", img)
    k = cv2.waitKey(10) & 0xff # Press 'q' for exiting video
    if k == ord("q"):
        break

cam.release()
cv2.destroyAllWindows()
```

7. PROJECT OUTCOME

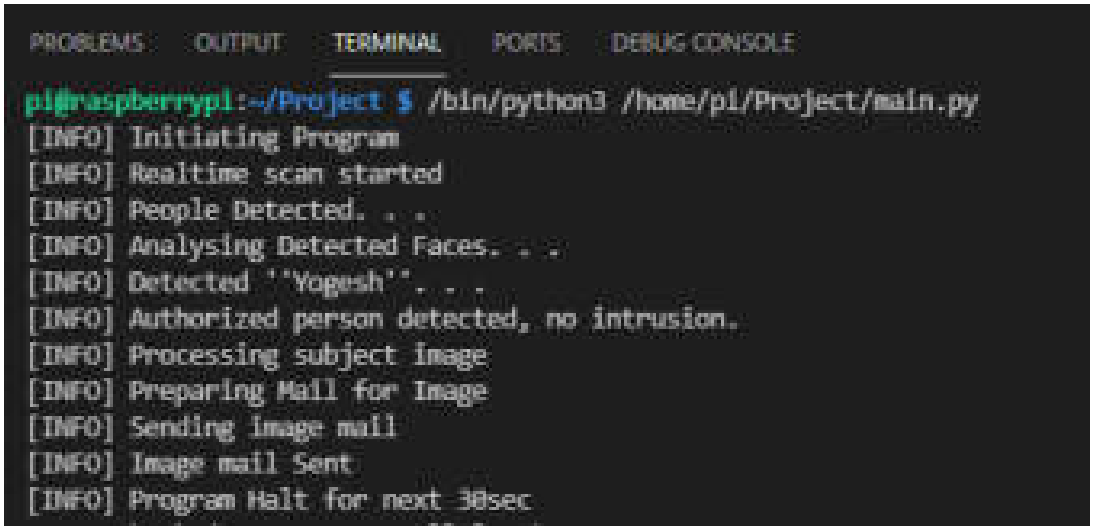
7.1 Dataset Making

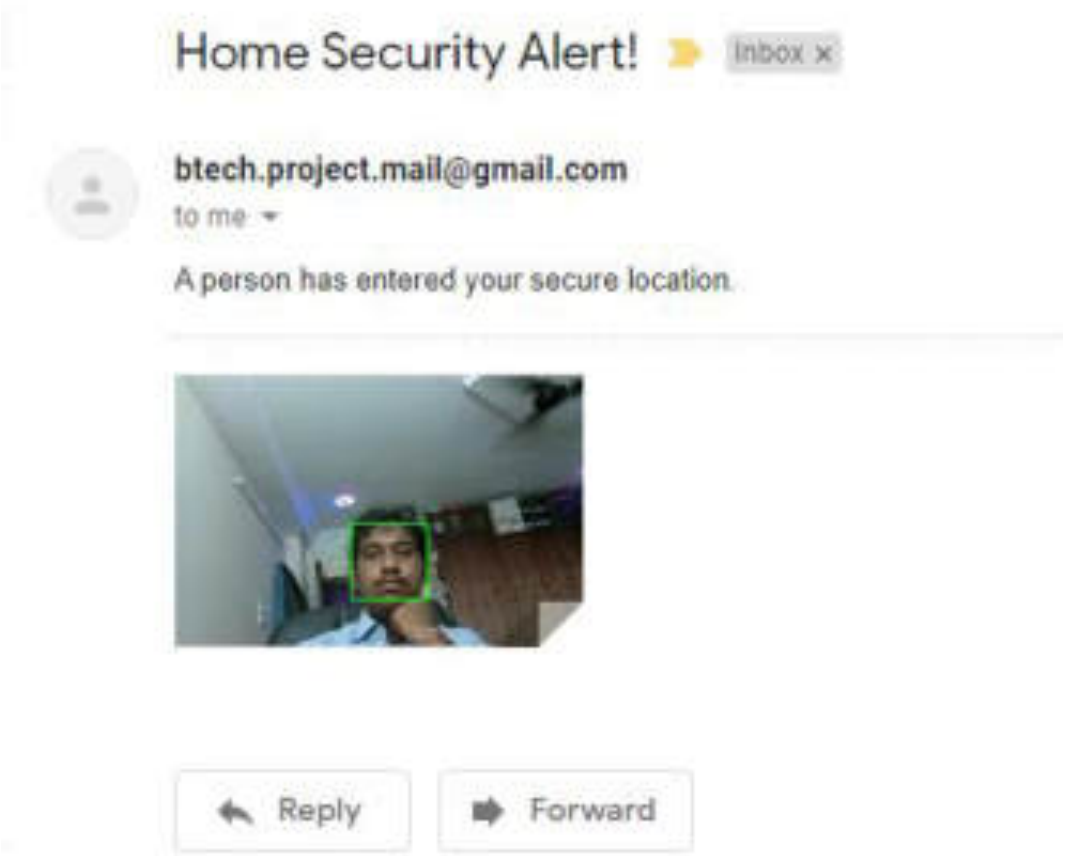
$ python3 main.py --addPerson NAME



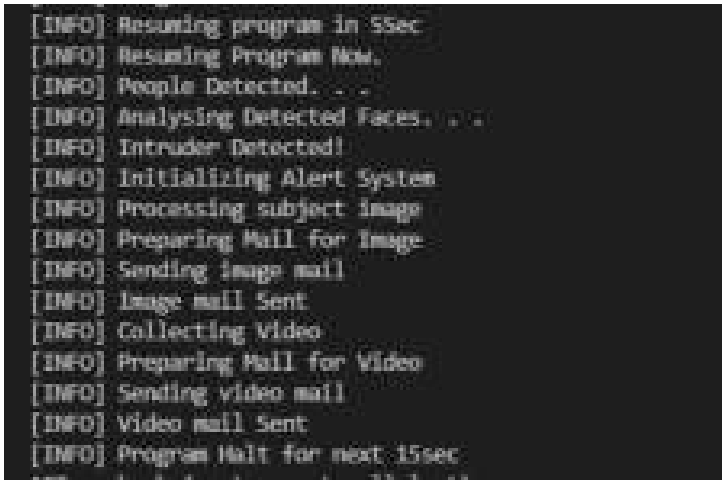


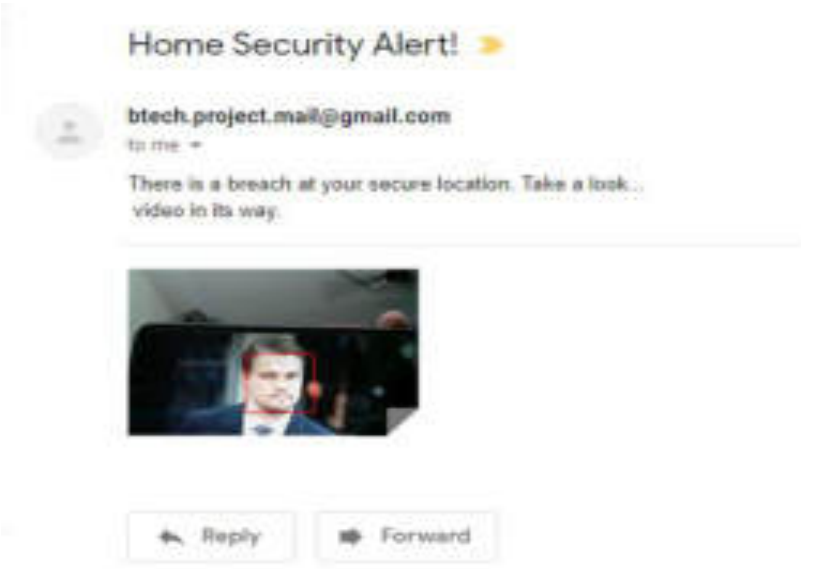
7.2 Authorized Person Detected

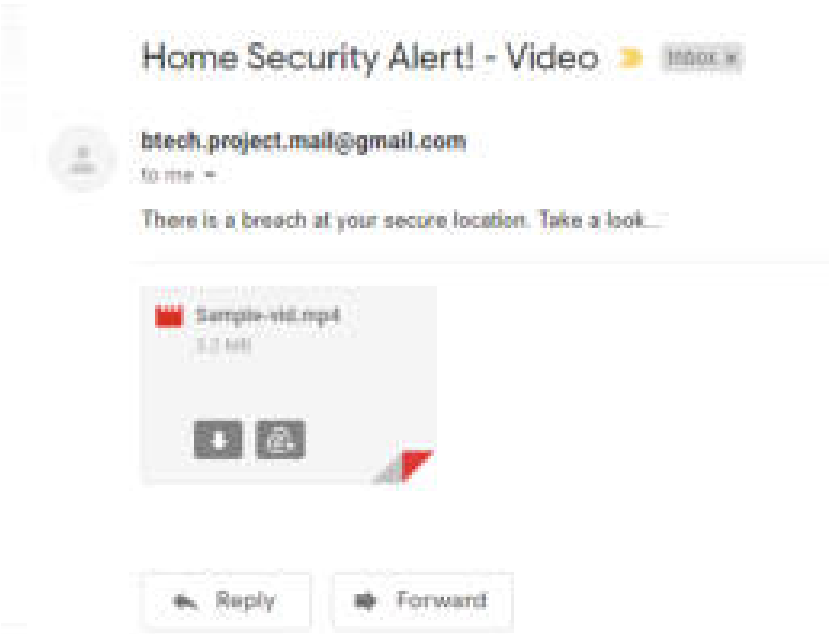




7.3 Intruder Detected







8. CONCLUSION & FUTURE WORK

In this work, a home security system using a camera with OpenCV implementation has been successfully developed on Raspberry Pi 3 Model B. The system composed of both hardware and software implementations, where both parts collaborate to form an effective motion detection mechanism. The Haar-Cascade algorithm coupled with a background subtraction method was applied in the OpenCV implementation. The results obtained from the three conducted experiments suggest the high-accuracy of a vision-based motion detection system, eliminating false alarms. This OpenCV motion detection system was shown to be more effective than the developed PIR motion detection system.

A deeper performance analysis can be made for instance by carrying out more experiments or use cases with more than 10 trials per experiment, plotting a confusion matrix by conducting separate positive and negative tests, etc.

Moreover, a variety of enhancements could be made to this system to achieve greater accuracy in sensing, detection, and alert notification, such as:

i. Integrate with Global System for Mobile communications (GSM) technology for Short Message Service (SMS) based alert. Notification through email is will definitely receive a slow response/feedback from user or respective person especially during a dangerous situation. Using the GSM-based technology, user or respective person can send/receive response even faster which can be very effective to handle dangerous situations.
ii. Implement Facial Recognition algorithm to the system. Using the facial images that the owner provides, the system will be able to recognize those faces and it can differentiate between a familiar faces and unfamiliar faces. During a burglary, the system easily recognizes the burglar because of his unfamiliar face and automatically triggers the alarm.
iii. Integrate Deep Learning algorithm to the system. The facial characteristics of a person will differ from time to time. Thus, it will be troublesome for the user to keep updating the latest facial images of the familiar faces. Using Deep Learning algorithm, a collection of data from the owner's smartphone such as captured images and videos is used to train the system automatically to recognize familiar faces. This will further improve the rate of detection and accuracy of the system to make it more robust.

9. REFERENCES

[1] M. Menon, "Innovating to zero crime rate," in Digital News Asia, ed, (2015).

[2] M. Pugh, J. Brewer, and J. Kvam, "Sensor fusion for intrusion detection under false alarm constraints," in 2015 IEEE Sensors Applications Symposium (SAS), (2015), 1-6.

[3] R. Sampson, "False burglar alarms," in Problem-Oriented Guides for Police: Problem-Specific Guides Series No. 5, U. S. D. o. J. Office of Community Oriented Policing Services, Ed., 2 ed, (2007).

[4] P. Vigneswari, V. Indhu, R. Narmatha, A. Sathinisha, and J. Subashini, "Automated security system using surveillance," International journal of current engineering and technology, vol. 5, no. 2, 882-884, (2015).

[5] S. Suresh, J. Bhavya, S. Sakshi, K. Varun, and G. Debarshi, "Home monitoring and security system," in 2016 International Conference on ICT in Business Industry & Government (ICTBIG, 2016), 1-5.

[6] H. U. Zaman, T. E. Tabassum, T. Islam, and N. Mohammad, "Low cost multi-level home security system for developing countries," in 2017 International Conference on Intelligent Computing and Control Systems (ICICCS, 2017), 549-554.

[7] S. Sruthy and S. N. George, "WiFi enabled home security surveillance system using Raspberry Pi and IoT module," in 2017 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES, 2017), 1-6.

[8] R. R. Ragade, "Embedded home surveillance system with pyroelectric infrared sensor using GSM," in 2017 1st International Conference on Intelligent Systems and Information Management (ICISIM, 2017), 321-324,

[9] S. Prasad, P. Mahalakshmi, A. J. C. Sunder, and R. Swathi, "Smart surveillance monitoring system using Raspberry Pi and PIR sensor," in International Journal of Computer Science and Information Technologies, vol. 5, no. 6, 7107-7109, (2014).

[10] B. Kaur, P. K. Pateriya, and M. K. Rai, "An illustration of making a home automation system using raspberry Pi and PIR sensor," in 2018 International Conference on Intelligent Circuits and Systems (ICICS), 2018, pp. 439-444.

[11] G. Bradski and A. Kaehler, Learning OpenCV: Computer vision with the OpenCV library. O'Reilly Media, Inc.", (2008).

[12] T. Agarwal. (2017, 5 May 2018). PIR Sensor Circuit and Module Working. Available: https://www.elprocus.com/pir-sensor-circuit-with-working/ 020013-10

[13] C. Zamshed Iqbal, I. Masudul Haider, A. Muhammad Moinul, S. Mst. Rumana Aktar, and N. Nafisa Shahera, "Design and implementation of Pyroelectric Infrared sensor based security system using microcontroller," in IEEE Technology Students' Symposium (2011), 1-5.

[14] A. Raza, A. A. Ikram, A. Amin, and A. J. Ikram, "A review of low cost and power efficient development boards for IoT applications," in 2016 Future Technologies Conference (FTC) (2016), 786-790.

[15] R. Pi. (2015, 10 May 2018). Raspberry Pi Model B. Available: https://www.raspberrypi.org/products /raspberry -pi-3-model-b/

[16] R. Pi. (2018, 10 May 2018). Run Android on Raspberry Pi. Available: https://www.raspberrypi.org/magpiissues/MagPi71.pd

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"PREVENTION OF SQL INJECTION ATTACKS USING AWS WAF"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
COMPUTER SCIENCE AND ENGINEERING

Submitted by

G Rishi Bhavani Prasad (17H71A0534)

S Yuva Adithya (17H71A0560)

A Nandini (17H71A0526)

G Chaitanya (17H71A0505)

Under the Esteemed Guidance of

Ms. A Anuradha
Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E-mail: dvhimic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech: ECE|EEE|CSE|ME
Accredited by NAAC A With Grade

2020-21

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273560
E mail: dvhimic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
(B.Tech: ECE|EEE|CSE|ME)
Accredited by NAAC With Grade A

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"Prevention of SQL Injection Attacks using AWS WAF"** is a bonafide work carried out by **Mr. G Rishi Bhavani Prasad (17H71A0534), Mr. S Yuva Adithya (17H71A0560), Ms. A Nandini (17H71A0526), Mr. G Chaitanya (17H71A0505)** in partial fulfillment for the award of the degree of Bachelor of Technology in **Computer Science and Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

A Anuradha
Project Guide

D. Prasad
Head of the Department

Examiner

Dr K Srinivas
PRINCIPAL

ACKNOWLEDGEMENT

We would like to take this opportunity to express my deepest appreciation to the following people for their valuable contributions and assistance with this Project.

Initially, we would like to thank my project supervisor, **Ms. A. Anuradha, Assistant Professor, Department of Computer Science and Engineering** for the guidance and support, especially for the valuable ideas and knowledge shared to me throughout the Project.

We have the immense pleasure in expressing my thanks and deep sense of gratitude to**, Mr. D. Prasad, Head of the Department, Computer Science and Engineering** for extending necessary facilities and support for the completion of the Project.

We whole sincerely acknowledge **Dr. K. Srinivas, Principal and Prof D. PANDURANGA RAO, CEO** for giving opportunity to take up the Project work. We also extend my thanks to all faculty members of **Computer Science and Engineering**, for their valuable guidance and encouragement.

We would like to extend my warm appreciation to all my friends for sharing us their knowledge, valuable contributions and help with this Project.

Finally, my special thanks go to my family for their continuous support and help throughout my academic years and for their continual support and encouragement for the completion of the project.

DECLARATION

We, **G Rishi Bhavani Prasad, S Yuva Adithya, A Nandini, G Chaitanya** of the Main-Project "Prevention of SQL Injection Attacks using AWS WAF", hereby declare that the matter embodied in this Project is the genuine work done by us and has not been submitted either to this university or to any other university/institute for the fulfillment of the requirement of any course of study.

G Rishi Bhavani Prasad (17H71A0534)

S Yuva Adithya (17H71A0560)

A Nandini (17H71A0526)

G Chaitanya (17H71A0505)

ABSTRACT

SQL Injection (SQLi) is a type of injection attack that makes it possible to execute malicious SQL statements. This Injection attack is used to retrieve sensitive information from the database of the web application. So the data that we are giving to the companies in the form of login details are not safe. This Vulnerability is due to the lack of input validation, which is the most critical part of software security that is often not properly covered in the software development life cycle design phase. The negligence in the development phase of the software may need to face severe damage in the future. This project presents the countermeasures for detecting and preventing SQL injection attacks. The proposed solution in the project is to use a database firewall between the client (user) side and the database server through AWS to avoid the malicious codes injected by the attackers. So AWS WAF prevents the malicious user from injecting the malicious queries, it prevents the execution of malicious queries on the client-side rather than executing the database server.

LIST OF FIGURES

CONTENTS

CHAPTER 1: INTRODUCTION

1.1 INTRODUCTION

In today's world where almost every task is performed through web applications such as banking, online shopping, and bill payments we entrust our personal information to these web applications and their underlying databases because of the trust on the confidentiality and integrity of the security of their data. As the usage of these services is increasing day by day on a large scale we are also facing a devastating increase in the number of attacks which can potentially give an attacker complete access to an individual's database such as one containing credit card information underlying the secured database.

SQL injection attacks (SQLIAs) are the most effective and malicious system attacks which can be used to gain or manipulate the data in data-driven systems. The risk of SQLIAs is that when they are performed by the victim back-end system, they will be running with the same privileges that the system has in the database, that means if the system has been assigned a role as a power user or administrator which has the read and write permissions then the injection code could be executed with disastrous effect on the victim machine.

A SQL Injection attack (SQLIA) is one in which a malicious minded person injects their own crafted query as an input and replaces the default query. The backend server executes the injected query statement and sends the result to the attackers. Therefore, most of the attackers use SQL for accessing the database and for the detection and prevention of these attacks various tools have been developed. There are multiple types of SQLIA's and each one of them has a different approach and effect for attacks on the target website.

To counter these attacks, we will be extensively discussing some of the modern SQL Injection attacks and the ways to protect and defend against these types of attacks. The negligence at the initial stage of development can lead to monetary losses at later stages.

1.2 PROBLEM STATEMENT

Unfortunately, many applications do not frequently validate user input. Attackers utilize these flaws to hack the backend database. The data to be stolen includes sensitive information of employees, customers. A successful attack can destroy information or execute a DOS (Denial of Service) attack.

1.3 WEB APPLICATION ENVIRONMENT

Before we initiate any discussion on the approaches for detection and prevention of SQL injection attacks, let's first explore the Web application environment itself. In a Web Application environment, the web application information is presented to the Web server by the user's client, in the form of URLs, cookies and form inputs (POSTs and GETs). These inputs drive both the logic of the applications as well as the queries which help the attacker to gain access to these applications for creating and sending a query to the database to extract relevant data.

Unfortunately, many applications do not frequently validate user input and so are more susceptible to SQL injection. Attackers capitalize on these flaws to attempt to hack the backend database to do something different than what the application or the search is intended for. This can include extracting sensitive information of employees, customers, destroying information or executing a DOS (Denial of Service) attack that limits the usage of the application.

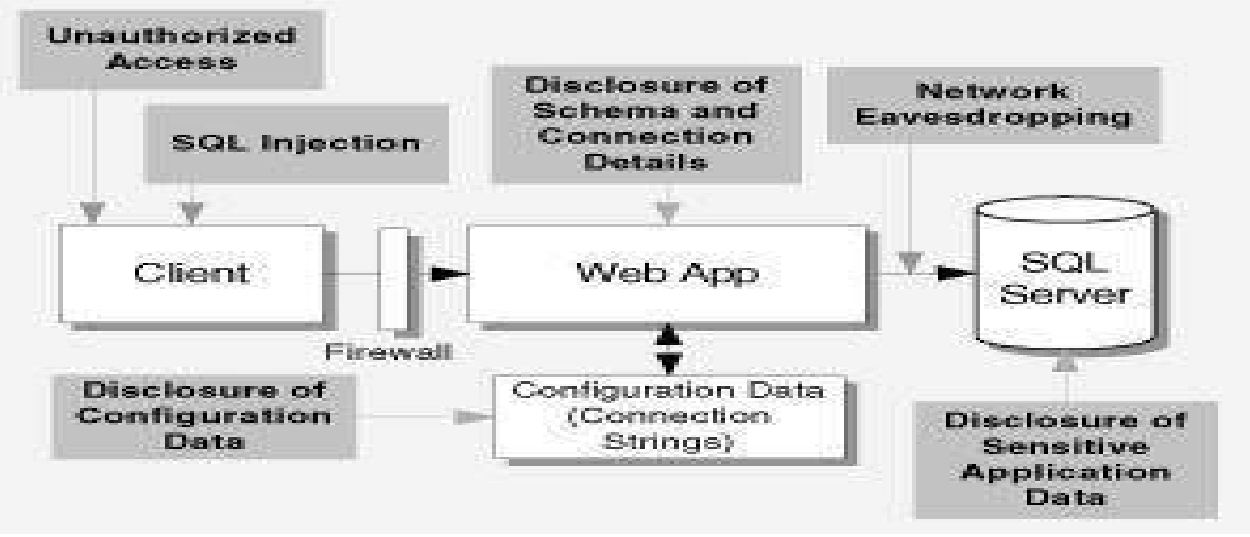


Figure 1.1 Web Tier Environment. SQL injection is still alive: A study on SQL injection signature evasion techniques (p.256), by Sadeghian, A., Zamani, M., & Ibrahim, S., 2013, International Conference on Informatics and Creative Multimedia.

1.4 OBJECTIVE OF THE PROJECT

- The main objective of this project is to provide multiple layers of security to protect databases from SQL Injection from a method which has highly durable storage and high-performance databases.
- To provide virtual clouds for organizations which are easy to access, have low maintenance and a capital prerequisite which can be taken care of even by small private companies and startup firms.

1.5 SQL INJECTION ATTACK OVERVIEW

SQL injection attacks are initiated by the manipulation of the data input on a Web form such that the traces of the SQL instructions are passed to the Web applications and these Web applications then combine with the rogue SQL fragments with the proper SQL dynamically generated by the application and create valid SQL requests. These new, unanticipated requests cause the database to perform the tasks intended by the attacker.

To have a clear understanding let us consider an example: If we have an application whose web page contains a simple form of the query with the input fields for username and password. With these credentials, the user can get a list of all the credit card accounts the various customers hold with a bank. Further, if the bank's application was built without taking into consideration the potential of SQL injection attacks.

In this situation, it is reasonable to assume that the application merely takes an input the user types and places it directly into the SQL query constructed to retrieve that user's information. In PHP, the query string would be like this:

$query = "select accountName, accountNumber from creditCardAccounts where username='".$_POST["username"]."' and password='".$_POST["password"]."'" Normally this would work properly as a user entered their credentials, say johnSmith and my Password, and forms the query:

$query = “select accountName, accountNumber from creditCardAccounts where username='johnSmith' and password='myPassword'

This query will come up with the total number of accounts Mr. John Smith is holding. Now consider someone with a fraudulent intent. If the person attempts viewing the account information of one or more of the bank’s customers, he enters the following credential into the form:

' or 1=1 -- and anyThingsAtAll

When this SQL fragment is inserted into the SQL query by the application it becomes: $query = “select accountName, accountNumber from creditCardAccounts where username='' or 1=1 -- and password= anyThingsAtAll

The injection of the term, ' or 1=1 --, accomplishes two things. Firstly, it causes the first term to be true for all the rows of the query in the SQL statement; Secondly, it causes the rest of the statement to be treated as a comment and is ignored during runtime. Thus, as a result, the attacker has all the valuable information customers were seeking all the credit card information up to the limit the Web page will list.

It should be noted that this simple example is just one of an infinite number of variations that can be used to accomplish the same attack. Further, there are many other ways to exploit a vulnerable application.

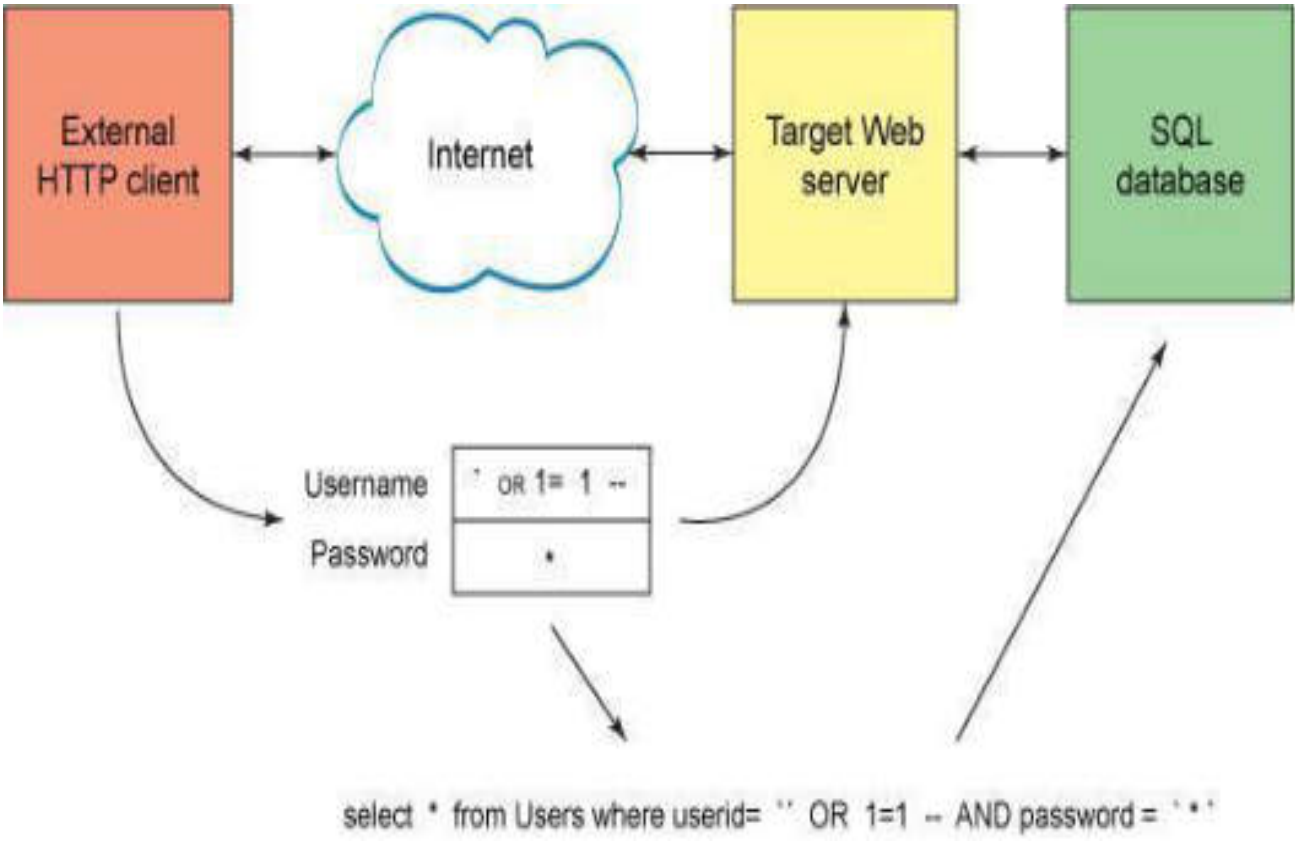


Figure 1.2 SQL Injection Attack Overview. Runtime monitors for tautology-based SQL injection attacks (p. 26), by Dharam, R.; Shiva, S.G., 2012, Cyber Security, International conference on Cyber Warfare and Digital Forensic (CyberSec).

1.6 APPLICATIONS VULNERABLE TO SQL INJECTION

Due to several factors, writing these applications securely has become very rare. Many applications were written at the time when Web security was not a major threat. While due to the recent discussions on SQL injection at security conferences and other settings, an awareness was spread that the attack frequency of SQL injection only five or so years ago was so low that most developers were simply not aware. In addition, the applications were exposed to the web with a lower security threshold and subsequently exposed to the web without even considering the security threats that it might have in the future because of SQL injections. Even applications which are written and deployed today often inadequately addresses security concerns. IBM's X-Force project recently found that 47% of all vulnerabilities that result in unauthorized disclosures are Web application vulnerabilities by Kar, Panigrahi, Sundarajan (2016) For packaged applications from commercial software vendors Cross-Site Scripting & SQL injection vulnerabilities continue to dominate as the attack vector of choice. Vulnerabilities in custom applications were not reported. Since this software is generally not as carefully treated for security robustness, it is reasonable to assume that the problem is much greater because 97% of data breaches worldwide are still due to SQL injection somewhere along the line by Kar, Panigrahi, Sundarajan (2016).

Interestingly, modern environments and development approaches create a subtle vulnerability. By the advent of Web 2.0, there has been a massive shift in how developers treat user input. In these applications, the input transmits the information to the web server directly in a simpler form for processing. Most frequently the JavaScript portion of the application performs input validation so the feedback to the user is handled more smoothly. This often creates the sense that the application is protected because of this very specific input validation; resulting in the negligence of the server side on a large scale. Unfortunately, attackers will not inject their input into an application using another application rather they leverage intermediate applications to capture the client-side input and allow them to manipulate it.

1.7 SUMMARY

The introduction gives a brief overview of the different types of SQL Injection attacks and how the web application firewalls are used to obstruct the unwanted queries in malfunctioning the codes of any web application. A brief overview of how a SQL injection attack is performed by an attacker is explained with the causes initiating the SQL injection attacks.

CHAPTER 2. BACKGROUND AND LITERATURE

2.1 INTRODUCTION

Over the past few years, SQL Injection attacks have been slipping seamlessly through the network firewalls over port 80 (HTTP) or 443 .Moreover, the count of SQL injection attacks against organizations has increased over the years causing devastating effects on their databases and security. At that point, the attacker can exploit the soft internal network and vulnerable databases because SQL injection has become the most dangerous threat that is being tackled by many organizations.

Detection of SQL fragments injected into a Web application has proven extremely challenging. There are several precautions and security measures that enterprises can adopt. When implementing prevention and remediation efforts, the enterprise strives to develop secure code and/or encrypt confidential data stored in the database. However, these are not always available options. For example, in some cases, the application source code may have been developed by a third party and not be available for modification. Additionally, patching deployed code requires significant resources and time because rewriting an existing operational application would need to be prioritized ahead of projects driving new business. Similarly, efforts to encrypt the confidential data stored in the database can take even longer time and require more resources. Given today's compressed development cycles, and a limited number of developers with security domain experience, even getting the code rewrite project off the ground could prove difficult.

2.2 LITERATURE RELATED TO THE PROBLEM

A novel technique was proposed by Wei Ke, Muthuprasanna, and Kothari,
(2006) to defend the SQL Injection attacks targeted at stored procedures. This technique was the combination of static application code analysis with runtime validation which can eliminate the occurrence of such attacks. The technique, in which a stored procedure parser was designed for any SQL statement which depends on user inputs to compare the original SQL statement structure to the user inputs was used.

An anomaly-based approach was described by Kiani, Clark & Mohay, (2008) which utilizes the character distribution of certain sections of HTTP requests to detect previously unseen SQL injection attacks. This approach does not require user interaction, and no modification of, or access to, either the backend database or the source code of the web application itself.

The hybrid approach based on the Adaptive Intelligent Intrusion 725 Detector Agent (AIIDA-SQL) proposed by Pinzon, Paz, Bajo & Herrero, (2010) was used for the detection of various SQL Injection attacks. "The AIIDA-SQL agent incorporates a Case-Based Reasoning (CBR) engine which is equipped with learning and adaptation capabilities for the classification of SQL queries and detection of malicious user requests" Pinzon et.al (2010). To carry out the tasks of attack classification and detection, the agent incorporates advanced algorithms in the reasoning cycle stages.

Basically, an innovative classification model based on a mixture of an Artificial Neural Network together with a Support Vector Machine is applied in the reuse stage of the CBR cycle. This strategy enables to classify the received SQL queries in a reliable way. Finally, a projection neural technique is incorporated, which notably eases the revision stage carried out by human experts in the case of suspicious queries by Pinzon et.al (2010).

The Database driven web application is subsequently threatened by SQL Injection Attacks (SQLIAs) because this type of attack can compromise confidentiality and integrity of information in databases and to stop these type of attacks various approaches had been proposed but because of their respective limitations they are not enough to block these attacks Tajpour & Jor, (2010).

To test the tools in a realistic scenario, Vulnerability and Attack Injection is applied in a setup based on three web applications of different sizes and complexities designed by Elia, Fonseca & Vieira, (2010). Results show that the assessed tools have a very low efficiency and only perform well under specific circumstances, which highlight the limitations of current intrusion detection tools in detecting SQL Injection attacks.

Based on the class of injection flaw in which specially crafted input strings lead to illegal queries to databases, an effective solution TransSQL was developed by Zhang, Lin, Chen, Hwang, Huang & Hsu (2011). TransSQL automatically translates a SQL request to an LDAP-equivalent request. After queries are executed on a SQL database and an LDAP one, TransSQL checks the difference in responses between a SQL database and an LDAP one to detect and block SQL injection attacks.

A framework which can be used to handle tautology-based SQL Injection Attacks using a post-deployment monitoring technique was proposed by Dharam & Shiva, (2012). Their framework uses two pre-deployment testing techniques i.e. basis path and data flow testing techniques to identify legal execution paths of the software. Runtime monitors are then developed and integrated to observe the behavior of the software for identified execution paths such that their violation will help to detect and prevent tautology-based SQL Injection Attacks.

Wu & Chan (2012), proposed a very effective method named k-centers (KC) to detect SQL injection attacks (SQLIAs). The number and the centers of the clusters in KC are adjusted according to unseen SQL statements in the practical environment, and in which the types of attacks are changed after a period to adapt to different kinds of attacks.

One of the most common solutions for defending against SQL Injection Attacks is the use of web application firewalls. Usually, these firewalls use signature-based techniques as the main core for the detection in which the firewall checks each packet against a long list of predefined SQL injection attacks known as signatures. “The problem with this technique is that an attacker with a good knowledge of SQL language can change the look of the SQL queries in a way that a firewall cannot detect them but still they lead to the same malicious results” Sadeghian, Zamani & Abdullah (2013).

2.3 LITERATURE RELATED TO THE METHODOLOGY

Amazon Web Services (AWS) provides a variety of infrastructure services, such as computing power, storage options, networking, and databases. These databases will be available in seconds and are delivered as a utility. “This allows enterprises, start-ups, small and medium-sized businesses, and customers in the public sector to access the building blocks they need to respond quickly required to change the business requirements” Mathew, (2006).

In 2006, Amazon Web Services (AWS) began offering IT infrastructure services to businesses in the form of web services—now commonly known as cloud computing Amazon Web Services, including the Web Application Firewall (AWS WAF) by Mathew (2006).

One of the key benefits of the AWS WAF is the opportunity to replace up-front capital infrastructure expenses with low variable costs that scale with the business. With the help of AWS WAF, businesses no longer need to plan for and procure servers and other IT infrastructure weeks or months in advance. Instead, they can instantly spin up hundreds or thousands of servers in minutes and deliver results faster.

2.4 TYPES OF SQL INJECTION ATTACKS

There are different types of attacks depending upon the goal of an attacker which are performed together or sequentially.

TAUTOLOGIES

The tautology-based attack is basically injecting the code into one or more conditional statements, so the statements always evaluate to true. The results of this attack depend on how the queries are used within the application proposed by Anley, (2002). The most common usages are to bypass authentication pages and extract data. In this type of injection, an attacker exploits an injectable field that is used in a query's WHERE condition statement.

According to McDonald (2002), the database table gets targeted by the returned query by transforming the conditional query. For a tautology-based attack to work, an attacker must not only consider injecting the vulnerable parameters, but also the coding which evaluates the query results. An attack is successful when the code either displays all the returned records or performs some action if at least one record is returned. Example: In this example attack, an attacker submits "' or 1=1 - - " for the login input field (the input submitted for the other fields is irrelevant). The resulting query is:

SELECT accounts FROM users WHERE login='' or 1=1 -- AND pass='' AND pin=
The code injected in the conditional (OR 1=1) transforms the entire WHERE clause into a tautology" by Halfond, Vieagas & Orso (2006).

The above condition is used as the base for evaluating each row and deciding which ones should return to the application. Because the above condition is a tautology, the query validates to be true for each row in the table and returns all the values related to the query. In the above example, the returned set evaluates to a nonnull value, which causes the application to conclude that the user authentication was successful by Howard, LeBlanc, (2003). Therefore, all the application would invoke method displayAccounts() and show all of the accounts in the set returned by the database.

ILLEGAL OR LOGICALLY INCORRECT QUERIES

The attacker gathers important information about the type and structure of the organization's back-end database of a Web application. This attack is considered a preliminary, information-gathering step for other attacks. Because of the vulnerability caused by this attack, the default error page is returned by the application servers and often are very detailed. In fact, according to Anley, (2002), injectable parameters can be generated by an attacker from the simple error messages that are displayed using any web application. While, additional error information, is used for debugging the applications by the programmers, will adversely help the attackers to gain information about the functioning queries of the back-end database.

As proposed by Litchfield (2002), while performing this attack, the statements that cause a syntax, type conversion, or logical error are manipulated by the attackers into the database. Injectable parameters can be identified by syntax errors. Type errors are used to deduce the data types of certain columns or to extract data. Logical errors can reveal the names of the tables and columns from the database that causes an error.

Example: In this example, the attacker's goal is to cause a type conversion error that can reveal relevant data. To do this, the attacker injects the following text into input field pin: "convert (int,(select top 1 name from sysobjects where xtype='u'))". The resulting query is:

SELECT accounts FROM users WHERE login='' AND pass='' AND pin= convert (int,(select top 1 name from sysobjects where xtype='u'))

In the above example, the select query injected into the attack string attempts to extract the first user table (xtype='u') from the database's metadata table (assume the application is using Microsoft SQL Server, for which the metadata table is called sysobjects). The query then tries to convert the specified table name into an integer and as this type of conversion is not legal in Microsoft SQL Server, the database throws a error stating Microsoft OLE DB Provider (0x80040E07) Error converting varchar value 'CreditCards' to a column of data type int. Halfond, Vieagas & Orso (2006).

There are two useful pieces of information which help an attacker according to Halfond, Vieagas & Orso (2006).

- "First the attacker can see that the database is a SQL Server database, as the error message explicitly states this fact.
- Second, the error message reveals the value of the string that caused the type of conversion to occur." Halfond, Vieagas & Orso (2006).

In the above scenario, the table that is been attacked first is a user-defined table in the database called "CreditCards". Each column in the database can be extracted by using the similar strategy. More threats can be created to the database by an attacker using the same information about the schema of the database, which targets specific pieces of information in the database.

UNION QUERY

For a given Query the attacker exploits a vulnerable parameter and changes the dataset returned in this type of attack. According to Anley, (2002) the application can be tricked into returning data from a different table that was not intended by the developer to be returned for the respective query. The most commonly injected statement used by the attackers is of the form: UNION SELECT. The information of the table can be retrieved by the attackers as they have complete control over the second/injected query which aids in accessing the permission rights to the database. Because of this attack, the final database will be a combination of the original query which was created by the developer and the modified second query injected by an attacker. Example: Referring to the running example, an attacker could inject the text

“’ UNION SELECT cardNo from CreditCards where acctNo=10032 - -” into the login field, which produces the following query:

SELECT accounts FROM users WHERE login=’’ UNION SELECT cardNo from CreditCards where acctNo=10032 -- AND pass=’’ AND pin=

The first query (Original) will return with a result of null set considering there is no login equal to “”, whereas the second query (Injected) returns data from the “CreditCards” table. The column “cardNo” for account “10032” will be returned by the database. Because of these two queries, the database will return the union of them to the application. Because of the union of these two queries, the cardNo would show up with the account information in the application by Halfond, Vieagas & Orso (2006).

PIGGY-BACKED QUERIES

Original query is injected with the additional queries in this attack type. This is distinguished from others as here the attacker modifies the original query instead of a new one. This includes the “piggy-back” queries on the original query. Numerous SQL queries are returned from the database because of this query. First the intended query is executed then the subsequent queries that are entered are the injected ones, and they are in addition to the previous one. According to Anley, (2002) this type of attacks are very vulnerable and if it is successful, any SQL command can be injected by the attackers virtually. The original query is injected and executed along with the stored. procedure as an example into the additional queries. This type of attacks usually happens to a database where the configuration allows multiple statements to be contained in a single string, they are very vulnerable to the structured database. McDonald, (2002).

Example: If the attacker inputs “’; drop table users - -” into the pass field, the application generates the query:

SELECT accounts FROM users WHERE login=’doe’ AND pass=’’; drop table users -- ’ AND pin=123 by Halfond, Vieagas & Orso (2006).

After the completion of the first query, the database will inject the second query after recognizing the query parameter (“;”) and the injected second query will be executed. Valuable information will be destroyed from the database if the injected second query is executed and the tables are dropped. Other types of queries could insert new users into the database or execute stored procedures Howard & LeBlanc, (2003). Simply scanning for a

query separator will not be a good idea to detect the injected queries as the databases do not require special characters to separate and identify distinct queries.

STORED PROCEDURES

According to Halfond, Vieagas & Orso (2006), stored procedures are routines stored in the database and run by the database engine. These procedures can either be user-defined procedures or procedures provided by the database by default. SQLIAs of this type try to execute stored procedures present in the database. The database interaction with the operating system is limited now a day to an extent with the help of stored procedures, as they set a standard functionality and most of the vendors provide that set by default while delivering the database. The SQLIAs can be used to execute the stored procedures in that database, once the attacker knows which type of database is used in the backend. Stored procedures also interact with the operating system. Using the stored procedures while coding the Web applications renders them invulnerable to SQLIAs. The stored procedures are not much dependent from the developer side as these procedures are most vulnerable to the attacks on the applications Howard & LeBlanc, (2003).

The attackers get the access to run the arbitrary codes on the server or to escalate the privileges as the stored procedures are often written in special scripting languages and additionally they can contain other types of vulnerabilities, such as buffer overflows. Labs, (2002).

CREATE PROCEDURE DBO.isAuthenticated @userName varchar2, @pass varchar2, @pin int AS EXEC("SELECT accounts FROM users WHERE login='" +@userName+ "' and pass='" +@password+ "' and pin=" +@pin); GO

Example: The SQLIA can be used to exploit the parameterized stored procedure in the above example. In the example, a stored procedure has been placed as an alternative for the constructed query string. To rightly authenticate the user credentials, the stored procedure returns a true/false statement. The attacker simply injects “ ’ ; SHUTDOWN; - -” into either the userName or password fields to inject the SQLIA attack. Due to this the injection the following query is generated through the stored procedure:

SELECT accounts FROM users WHERE login='doe' AND pass=' '; SHUTDOWN; -- AND pin= Halfond, Vieagas & Orso (2006)

This attack is called a "piggy-bank" type attack. The injected or the malicious query is injected second into the database after the execution of the first normal query, due to which the database shuts down. In the above example, it illustrates that the stored procedures are as vulnerable to the same range of attacks as the traditional application code.

INFERENCE

In this attack, the query is modified in such a way that any action executed will depend on the true or false answer values for the data which is altered in the database. In this type of injection, attackers generally attack a site that has enough security so that, whenever there is a successful injection, there should not be any usable feedback through database error messages. As the database error messages are unavailable or not sufficient for the attacker as no feedback is provided an alternate method should be used by the attackers for obtaining a response from the database by Anley, (2002).

According to Spett, (2003) by using an alternate method malicious commands will be injected by the attacker into the website and is studied for any functional changes on the website. After completely studying the effects caused by the injected commands like what changes the commands are making to the website interface and functioning the attacker can deduce the accurate commands to see what parameters are vulnerable to the change in the behavior of the. Most commonly there are two important attack techniques based on an inference which allows an attacker to extract data from a database and detect vulnerable parameters.

BLIND INJECTION

According to Halfond, Vieagas & Orso (2006) the developers hide the error details during programming which ends up showing a generic page instead of an error message because of which the attacker gets the information of the tables related to the database structure by asking the true/false type of questions through SQL statements.

SELECT accounts FROM users WHERE login= 'doe' and 1 =0 -- AND pass = AND pin=O

SELECT accounts FROM users WHERE login= 'doe' and 1 = 1 -- AND pass = AND pin=O

If there is no input validation the query will execute.

TIMING ATTACKS

According to Halfond, Vieagas & Orso (2006), this attack particularly depends on the time lapses or delays. This time delays aid an attacker in gaining information of the database. The timing attack is pretty much like the blind injection except it uses a different inference method. For performing a timing attack, if/then statements are used as an injected query by the attacker which relates to the content of the database. The WAITFOR keyword which is used to delay the time response for a specified time uses the SQL Queries to construct the amount of time to execute each branch among all the other branches. A specific branch is picked by the attacker which either increases or decreases in response to the time of the database which gives the solution of the injected question to the attacker.

Example: A specific code is used in two different ways in which the attacks are explained by using the inference-based techniques. The parameters are identified using the blind injection technique in the first form while filling up two possible injections in the login field.

The first being “legalUser’ and 1=0 - -” and the second, “legalUser’ and 1=1 - -”. These injections result in the following two queries:

SELECT accounts FROM users WHERE login=’legalUser’ and 1=0 -- ’ AND pass=’’ AND pin=0

SELECT accounts FROM users WHERE login=’legalUser’ and 1=1 -- ’ AND pass=’’ AND pin=0

Considering two scenarios in which assuming the first scenario as a secure application which has a validated login input. As the SQL queries injected by the attacker will return with login error messages because of the incorrect login parameters making the query not vulnerable. In the second scenario, there will be two attempts by the attacker for the injection one with always a true statement and one with always false statement as we have an insecure application and the login parameter is vulnerable to injection. The first statement which will be false is injected by the attacker and as an expected result the application will return with a login error message.

There might be two reasons for an error message during login, one being the attack attempt validated correctly by the application and second, the injected attack itself caused the login error. Now the second statement which is always true is injected.
by the attacker and there won't be any login error message which concludes to the attacker that the login parameter is vulnerable to the injection.

Data extraction can be carried out using the inference-based techniques by injecting a timing-based inference attack and extracting the table name from the database.

In this attack, the following query is injected into the login parameter: ''legalUser' and ASCII(SUBSTRING((select top 1 name from sysobjects),1,1)) > X WAITFOR 5 --''.
This produces the following query:
SELECT accounts FROM users WHERE login='legalUser' and ASCII(SUBSTRING((select top 1 name from sysobjects),1,1)) > X WAITFOR 5 -- ' AND pass='' AND pin=0

In this attack, the attacker asks a series of questions about the first character of the first table's name (SUBSTRING) using a binary search strategy and if the value of X is greater-than or less-than-or-equal-to the value of ASCII value there is an additional 5 second delay in the response of the database, by which the attacker knows that the value injected is greater and then the value of the first character. Therefore, the value of X is adjusted by the attacker accordingly.

ALTERNATE ENCODINGS

This attack is used in combination with other attacks by injecting a modified query altered by defensive coding practices to avoid detection of the automated prevention techniques. In other words, as explained by Anley, (2002) alternate encodings are used as an aid by the attacker for evading the detection and prevention techniques which might be exploitable and can carry vulnerabilities in the application. These evasion techniques are useful in scanning certain "bad characters," such as single quotes and comment operators commonly used in the coding practices.

The common techniques are not enough capable of determining and scanning the specially encoded strings which use hexadecimal, ASCII, and Unicode characters which allows the SQL injection attacks go undetected.

The alternate Encoding technique provides different layers in an application to evaluate all the specially encoded strings by scanning for certain escape characters that represent alternate encodings in its language domain and may even use different methods of encoding by Howard & LeBlanc, (2003).

A perfect code-based defense is practically very much difficult to build and implement in work environment as it requires the developers to consider all the possible scenarios which could affect a query string in different layers of an application through SQL injection.

For example, "a database could use the expression char(120) to represent an alternately-encoded character "x", but char(120) has no special meaning in the application language's context Halfond, Vieagas & Orso (2006)." Therefore, the attackers are very much successful in injecting a coded query in the application code string.

Example: An alternately encoded attack is provided in the example in which the following text is injected into the login field: "legalUser'; exec(0x73687574646f776e) -- ".

The resulting query generated by the application is:

SELECT accounts FROM users WHERE login='legalUser';

exec(char(0x73687574646f776e)) -- AND pass='' AND pin=

In this example char() function is used with the ASCII hexadecimal encoding. The char() function returns the instance of that character and is considered as a parameter an integer or hexadecimal encoding of the particular character. The second line in the example is the ASCII hexadecimal encoding of the string "SHUTDOWN." Therefore, a SHUTDOWN command is executed whenever a code or string is interpreted by the database.

2.5 MAIN CAUSES OF SQL INJECTION

In this section, various causes of SQL injection are presented:

Invalidated input. Any SQL query consists of some parameters such as INSERT, UPDATE, ALTER and some SQL control characters such as a semicolon and quotation mark. If there is no checking for these, web applications can potentially be abused in a SQL injection attack.

Generous privileges. Privileges are some rules for accessing some database for an object. SELECT, INSERT, and DELETE are actions of executing SQL queries that include

typical privileges. Typically, a web application is used for accessing any specific information from the database.

Uncontrollable variable size: If any variable is used for the storage of a large amount of data there might be a chance of SQL injection of faked input values from the attacker

Error message. An error message is generated when the wrong input values are inserted in web applications. Attackers may get the script structure or information about the database so that the attacker may create its own attack.

Client-side only controls. If input validation is implemented in client side-scripts only, then by using cross-site scripting security functions of a script at the client side it can be overridden, and an attacker can invalidate input for accessing the database.

Stored procedure. Stored Procedures are a small program with some functions which are called multiple times in execution. When these functions become calls so that stored procedures become calls in place of that function. These stored procedures become stored in the database. The problem with stored procedures is that an attacker can execute and damage the database.

Into out file support. A text file containing SQL query results may be gotten by manipulating a SQL query. This can be possible by using the condition of INTO OUTFILE clause that is beneficial for some relational databases.

Sub-select. When a SQL query is inserted in the WHERE clause of another SQL query this shows one of the weaknesses for a database. This weakness also makes the web application more vulnerable.

The challenge with detection. The goal of any security technology is to provide a robust threat detection for the database which is very easy to set up or which doesn't require any setup or configuration. Further, if that technology relies on learning or training to improve its ability to detect threats, those learning periods must be short and well-defined. The longer the time period, for learning the higher are the chances that attacks may occur so there is a need to expedite the installation and minimize the risk of attacks.

2.6 DETECTION AND PREVENTION TECHNIQUES

Researchers have proposed a range of techniques to assist developers and compensate for the shortcomings in the application of defensive coding.

Black box testing. A black-box technique called WAVES, was designed by Huang, Lin & Tsai (2003) for testing Web applications for SQL injection vulnerabilities. The technique uses a Web crawler to identify all points in a Web application that can be used to inject SQLIAs. It then builds attacks that target such points based on a specified list of patterns and attack techniques. The time response of the attacks over the application improves in the WAVES technique as it uses the machine learning approaches to guide the testing. This technique is safer compared to the other testing's but still cannot guarantee concerning complete security.

Static code checkers. JDBC-Checker technique is also known as Static code checker technique which is used to prevent the type of SQL injection attacks that occur due to the mismatch of the practically generated query string proposed by Gould, Su & Devanbu, (2004). This technique detects SQLIA code vulnerabilities, typo's in the code input. As this technique was not developed for detection and prevention of the SQL injection attacks is still used for the same purpose of finding the root vulnerabilities in the dynamically generated query string. Even after the combination of the static analysis with the automated reasoning, it was unable to detect different types of SQL injection attacks other than Tautologies.

Combined static and dynamic analysis. AMNESIA is a model-based technique designed by Halfond & Osro, (2005) that combines static analysis and runtime monitoring. There are two phases in this type of analysis static phase and dynamic phase. Static analysis is used to generate legal queries for an application at each point of access to the database by building models of different types of queries through a process called AMNESIA.

Whereas Dynamic analysis validates all the unwanted queries before they are sent to the database for the statically built models through the same process. Queries which does not pass through the validation of AMNESIA are considered as SQLIAs which will be terminated from executing into the database. The primary limitation of this technique is the accuracy of the static analysis which is used for building the query models.

There are two more approaches related to the combined static and dynamic analysis. In the first approach runtime for the queries is verified to confirm the model for the expected queries should pass only the accepted queries. Whereas the SQLGuard model deduces the runtime by adding an additional user input known as SQLCheck -by the developer. Both the approaches share a secret key which is used to insert user input during parsing by the runtime checker.

The developer must rewrite the use of special characters or markers in the code to develop a dynamically generated query so as the to avoid the attackers in finding out the secret key proposed by SQLGuard by Buehrer, Weide & Sivilotti, (2005) and SQLCheck by Wasserman & Su, (2004).

Taint-based approaches. The Taint Based approach uses a method called WebSSARI which is used to check the taint flows for sensitive functions which detect the precondition points in which the filters and sanitization functions can automatically be added to satisfy the precondition parameters. It uses the predefined set of filters to sanitize the input. The primary drawback of this technique is that the sensitive functions in an injected code can be accurately expressed using the typing system through a certain type of filters which are not tainted stated by Huang, Yu, Hang, Lee & Kuo, (2004).

Livshits and Lam, (2005) proposed that using information flow techniques for detecting the tainted input using static analysis vulnerabilities in software can also be detected. A SQL query can be constructed with this technique to avoid the flagged as SQLIA vulnerabilities. Another approach made by Pietraszek and Berghe, (2005) used a context-sensitive analysis which used a PHP interpreter to track precise per-character taint information. The SQL injections would be validated depending on the false positive statements which intercept any untrusted query or code injected by an attacker. Only known patterns of SQLIAs can be detected by these two approaches which cause the common drawback for both the methods as they require modifications to the runtime environment, which affects portability.

Another technique is by using SecuriFly which validates the query strings generated by the tainted inputs, unlike the above two approaches which use a context-sensitive analysis and track the taint information depending on the per-string basis stated by Haldar, Chandra & Franz, (2005) and Martin, Livshits & Lam, (2005).

But as there is no taint-based approach related to this method it does not give enough sanitization to regulate the injection in the numeric fields of the code. The main drawback of this technique is identifying all the sources of tainted user input in web applications and accurately validating them.

New query development paradigms. A combination of two approaches, SQL DOM by McClure & Krugre, (2005) and Safe Query Objects proposed by Cook & Rai, (2005) offers an effective technique by changing the query building process using encapsulation of database queries in combination with the API string concatenation. This approach provides a safe and

reliable way to access the databases and avoids the unwanted SQL injections. This technique needs a new development environment as it is a combination of the latest and the legacy approaches which creates a paradigm in which the SQL queries are developed.

As it is a new environment the only drawback is the developers must learn a new programming language and there won't be any protection for the existing legacy systems.

Intrusion detection systems. IDS system builds models based on a machine learning technique which consists of typical queries and monitors the runtime of the application in real time that is being trained using a set of typical application queries. As the training set is required to monitor the application, a poor training set will generate many false positives and false negatives which is the only limitation of IDS stated by Valeur, Mutz and, Vigna, (2005).

Proxy filters. These filters have security gateways which provide the developer with a Security Policy Descriptor Language (SPDL), which has specified constraints and helps in filtering the unwanted injected codes coming from untrusted proxies to the web application. SPDL provides defensive programming which requires the developers to know which data needs to be filtered and which proxies should be blocked and considered as untrusted and what patterns be applied to the existing database to suspend unwanted SQL injection attacks Scott & Sharp, (2002).

Instruction set randomization. SQLrand is based on the framework which helps the developers in creating the queries based on instruction-set randomization. Instruction-set randomization uses a proxy filter which intercepts the normal SQL keywords and pushes the randomized queries to the database. As the code injected by the attacker might not be constructed using the randomized instruction set the injected SQL query will fail in attacking the application. Like other techniques, SQLrand has a drawback that the code uses a secret key to modify the instructions which result in integration of a proxy with the tables present in the database of a system.

a. INJECTION DETECTION AT THE WEB TIER

There is a large variation in the pattern of SQL attacks, which makes it even more challenging for the detection of the initial point from where the attack is initiating in the Web server. Furthermore, the SQL requests sent to the database has special characters which may not be expected in a typical form sent by the attacker. There are URL's, cookies, and form inputs (POSTs and GETs) to inspect and retrieve and inspecting each set of input values, makes it more

difficult for a WAF. The SQL injection attacks are caused by coding the application using simple coding techniques and words such as “like” and “or” to catch every possible attack which practically is not possible.

Alternatively, as mentioned earlier, much more complex patterns that are clearly indicative of an attack can be used. Unfortunately, as discussed, the different types of SQL injection attack the number and variation of possible attacks are so large that it is impossible to effectively cover all possible attack patterns. Creating the initial pattern set, being updated about the evolving attacks, and verifying that they are sufficiently unique so as not to show up in some fields is an almost impossible task. And now, considering that the applications are also changing and evolving over time, it requires more time so as more learning and hands-on skills for proper security of the databases without any breaches.

The Database Firewall is much more secure and effective than the previously used Web Application Firewall as it follows the structured analysis to build the SQL statements instead of the rudimentary input pattern validation used in WAF. It is more effective and secure because it monitors the networks between the application servers and databases with a much smaller set of SQL build statements. This database firewall is not that easy to build and maintain so we opt for different services such as Oracle but the latest most efficient and economical service to store and to secure the integrity of the data is provided through Amazon Web Services AWS.

2.7 SUMMARY

The Background and Literature review helps in completely understanding about the SQL Injection attacks. Different types of SQL injection attacks are explained with the main causes and some of the detection and prevention techniques. The most efficient method of detecting and preventing the web applications from the SQL Injection attacks (Injection Detection at Web-Tier) is also explained.

CHAPTER 3. DESIGN

3.1 INTRODUCTION

Amazon Web Services Web Application Firewall (AWS WAF) helps to protect web applications from common web exploits like SQL injection attacks that could affect application availability, compromise security, or consume excessive resources.

AWS WAF gives control over the traffic which allows or blocks the web applications by defining customizable web security rules. To create custom rules that block common attack patterns, such as SQL injection or cross-site scripting and to respond quickly for the change of patterns in the traffic, new rules can be deployed within minutes through AWS WAF. Also, AWS WAF includes a full-featured API that can be used to automate the creation, deployment, and maintenance of web security rules.

The strategy of configuring a web application firewall can be challenging and burdensome to large and small organizations alike, especially for those who do not have dedicated security teams. To simplify this process, AWS offers a solution that uses AWS Cloud Formation to automatically deploy a set of AWS WAF rules designed to filter common web-based SQL injection attacks. With AWS WAF we pay only for what we use. AWS WAF pricing is based on how many rules are being deployed and how many web requests the web application receives. These rules can be deployed by AWS WAF on either Amazon Cloud Front as part of the CDN solution or the Application Load Balancer (ALB) that fronts the web servers or origin servers running on EC2.

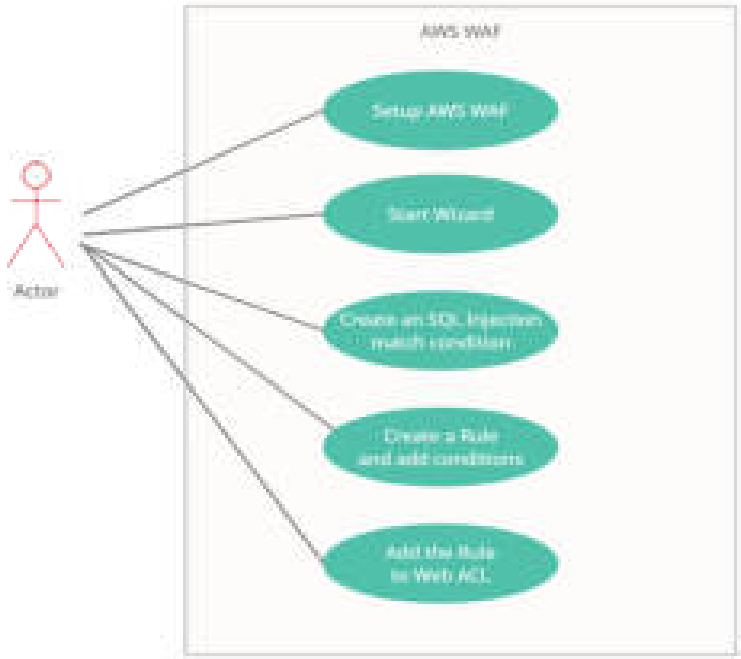
3.2 UML DIAGRAMS

Unified Modelling language (UML) is a standardized modeling language enabling developers to specify, visualize, construct and document artifacts of a software system. Thus, UML makes these artifacts scalable, secure and robust in execution. UML is an important aspect involved in object-oriented software development. It uses graphic notation to create visual models of software system. UML diagrams represent static and dynamic views of a system model. The static view includes class diagrams and composite structure diagrams, which

emphasize static structure of systems using objects, attributes, operations and relations. The dynamic view represents collaboration among objects and changes to internal states of objects through sequence, activity and state machine diagrams. A wide variety of UML modeling tools are available to simplify the modeling process, including IBM Rational Rose, Rational Rhapsody, represents Magic Draw UML, Star UML, Argo UML, BOUML, Power Designer and Dia. The UML a collection of best engineering practices that proven successfully in the modeling of large and complex systems. The UML is a very important part of developing objects oriented software and software development process. The UML uses mostly graphical to expresss the design of software projects.

3.2.1. USE CASE DIAGRAM

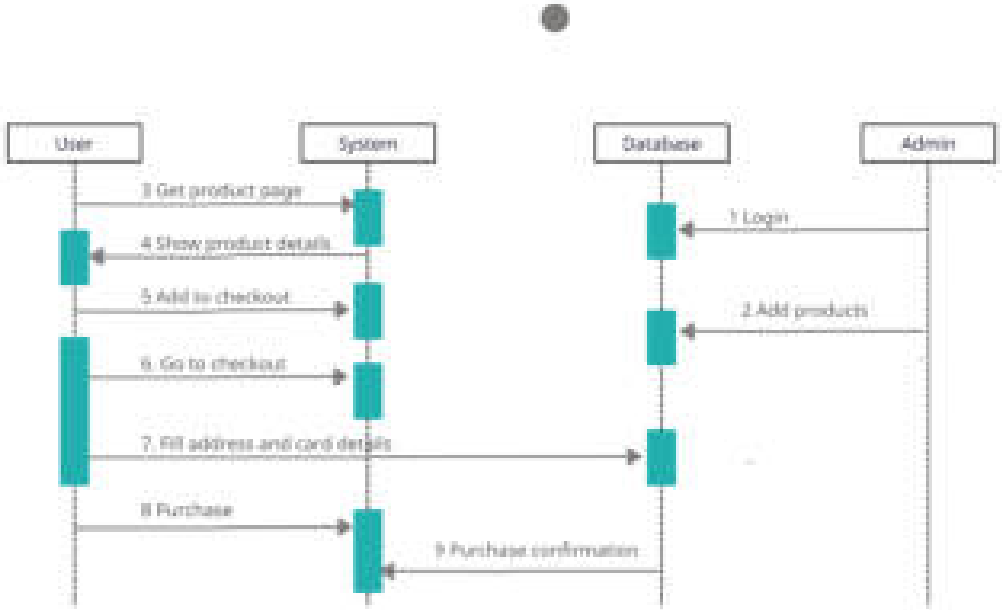
A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



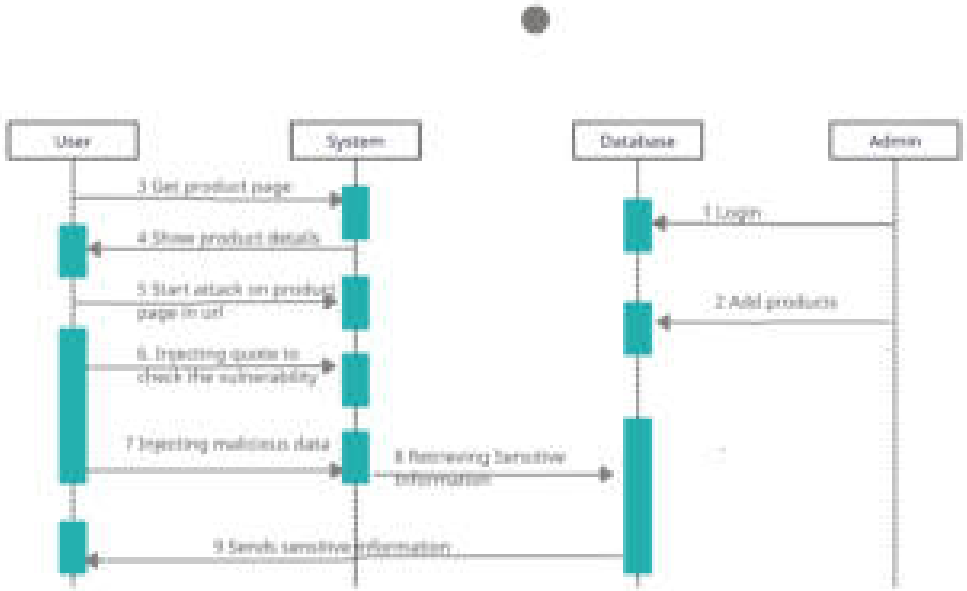
3.2.2. SEQUENCE DIAGRAM

A sequence diagram is a Unified Modeling Language (UML) is a sort of cooperation chartthat shows how forms work with each other and in what order. It is a develop of a MessageSequence Chart. Arrangement outliners are at times called occasion graphs, occasion situations, event scenarios and timing diagrams.

Normal User:

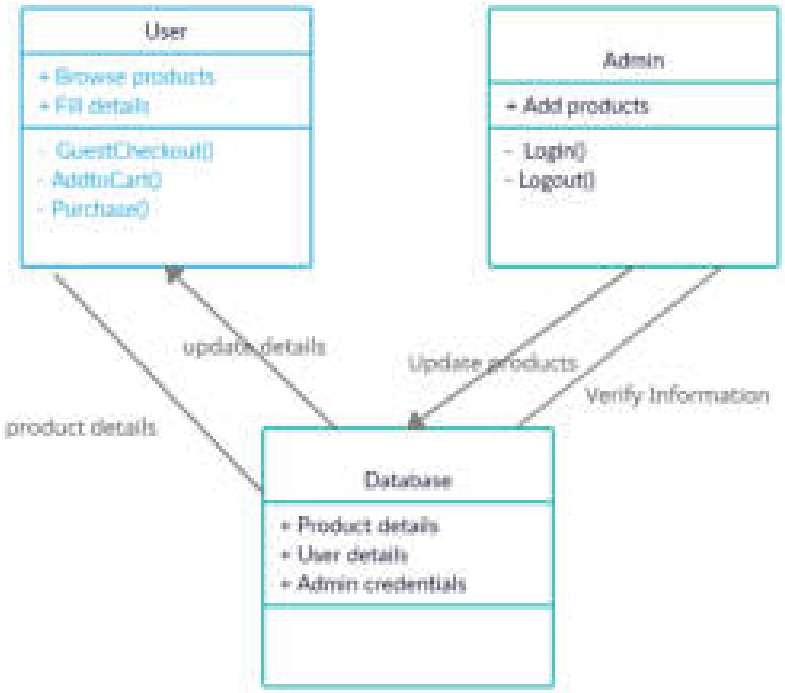


Malicious user:



3.2.3. CLASS DIAGRAM

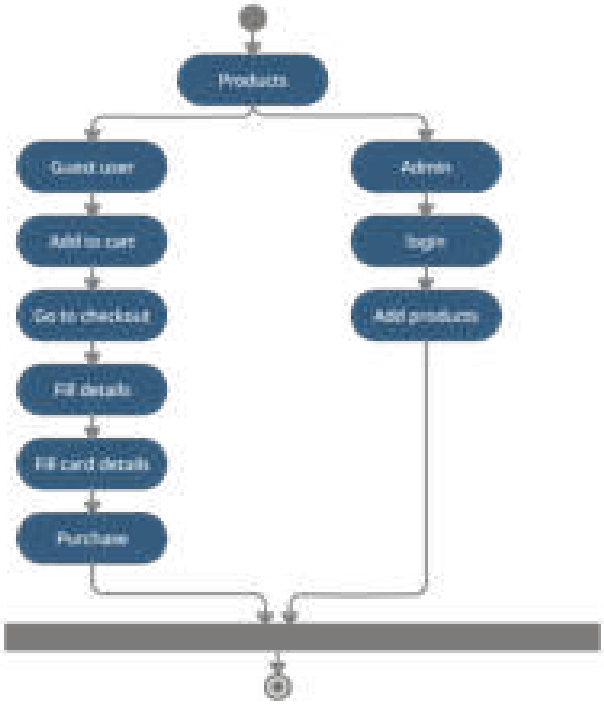
The class diagram is the main building block of object-oriented modeling. It is used for general conceptual modeling of the systematic of the application, and for detailed modeling translating the models into programming code. Class diagrams can also be used for data modeling. The classes in a class diagram represent both the main elements, interactions in the application, and the classes to be programmed.



3.2.4 ACTIVITY DIAGRAM

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. Anactivity diagram shows the overall flow of control.Activity diagrams are constructed from a limited repertoire of shapes, connectedwith arrows. The most important shape types: • rounded rectangles represent activities; • diamonds represent decisions; • bars represent the start (split) or end (join) of concurrent activities; • a black circle represents the start (initial state) of the workflow; • An encircled black circle represents the end (final state). Arrows run from the start towards the end and represent the order in which activities happen. However,

the join and split symbols in activity diagrams only resolve this for simple cases; the meaning of the model is not clear when they are arbitrarily combined with the decisions or loops.



3.2 DESIGN OF THE STUDY

Thus, far we have discussed different types of SQL injection attacks, the main causes of SQL injection and the method of detecting SQL injection attacks at the Web tier interface by a simple WAF system. A more effective and efficient method proposed in this paper to defend against SQL injection attacks is by using AWS WAF. This web application firewall allows us to monitor the HTTP and HTTPS requests which are forwarded to Amazon Cloud Front or an Application Load Balancer and allows us to control and access the content.

Based on conditions specified by the user, such as the IP addresses that the requests originate from or by the query string values, the Cloud Front or an Application Load Balancer responds to requests either with the requested content or with an HTTP 403 status code (Forbidden). The Cloud Front or an Application load balancer can also be configured in such a way that it returns with a custom error page when a request is blocked to analyze the actual SQL generated by the application as presented to the database by a firewall.

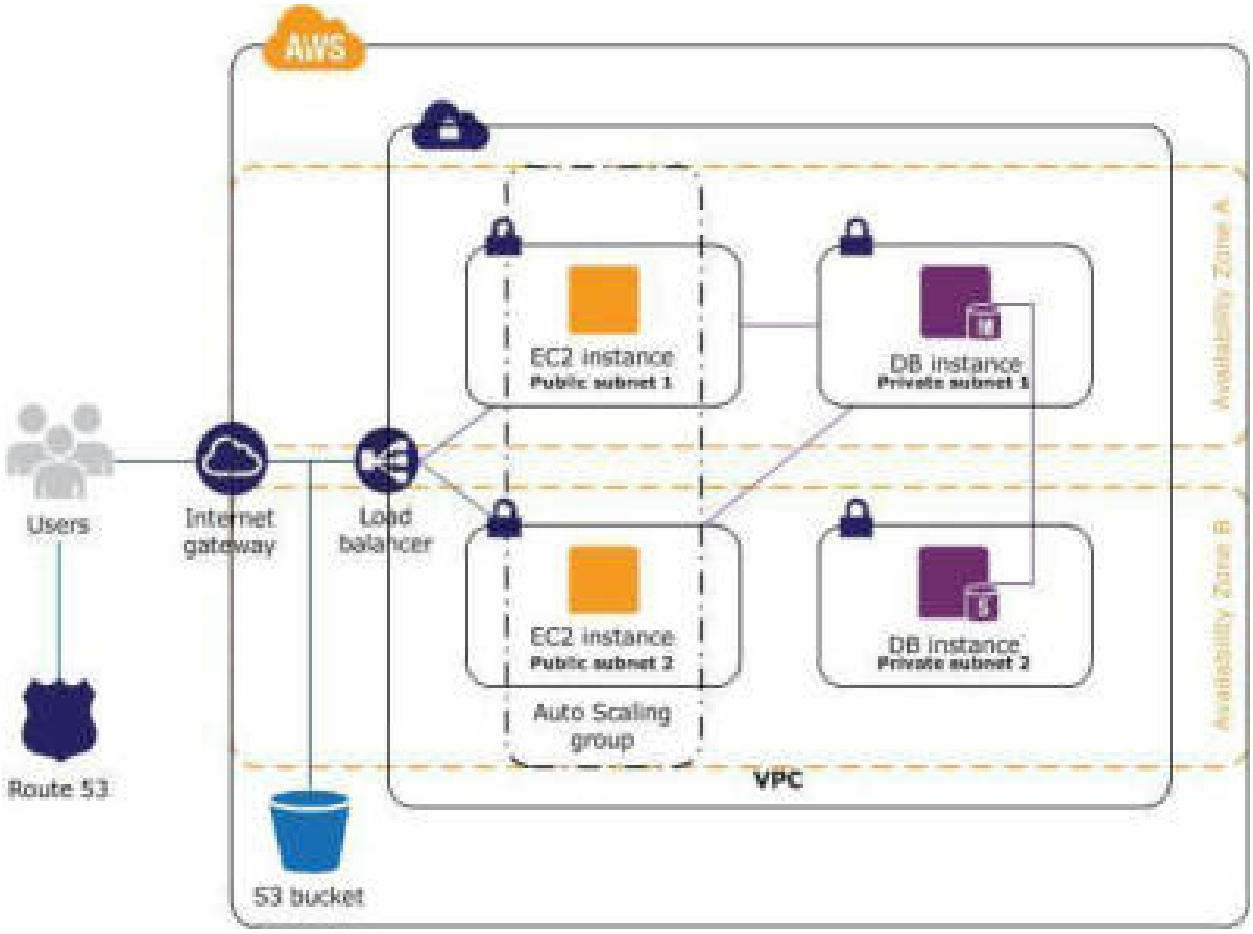


Figure 3.1 AWS WAF Architecture. Amazon Web Services Architecture, by Mathew, M., 2006, White Paper, Amazon Web Services, Inc., 2006. https://aws.amazon.com/security/documents/WhitepaperAWSWAF.pdf.

The AWS WAF allows us to choose only the requests specified and block all the other unwanted requests such as SQL injections. It gives several other potential benefits such as providing rules which can be reused for multiple web applications, automated administration using AWS WAF API, real-time metrics and sampled web requests. A qualitative approach will be best suited for the proposed plan as it does not require any numerical data analysis. The following are the steps used for building an AWS WAF which will be discussed in detail in the next part of the paper.

- Step 1: Set Up for AWS WAF
- Step 2: Start the Wizard
- Step 3: Create an IP Match Condition
- Step 4: Create a String Match Condition
- Step 5: Create a SQL Injection Match Condition
- Step 6: Create Additional Conditions
- Step 7: Create a Rule and Add Conditions
- Step 8: Add the Rule to a Web ACL
- Step 9: Clean Up Your Resources

3.3 DATA ANALYSIS

Hardware and software requirements.

Hardware requirements

A Ec2 t2.micro instance:

Memory :	1GB
Storage:	10GB
vCPUS:	1

Software requirements

Stack	:	LAMP (Linux, Apache, MySQL, PHP)
Frontend	:	HTML, CSS, Javascript
Backend	:	PHP, MySQL
Aws Services	:	EC2, Application load balancer, AWS WAF

3.4 SUMMARY

An effective and efficient approach towards the protection of web applications is explained in the methodology. The AWS technique provides better security, no infrastructure, less capital, on-demand upgrade of processing speed, storage services and many computing clouds and subnets. The design and steps of building an Amazon Web Services Web Application Firewall are explained briefly.

CHAPTER 4: IMPLEMENTATION

4.1 INTRODUCTION

Amazon Web Services was born out of the idea to provide multiple layers of security to avoid SQL injection attacks and to transfer the data from small scale to large scale. Amazon web services are available at any capacity on a moment's notice and without necessarily forecasting demand. Amazon meets this expectation in both of its key AWS products. Amazon's Elastic Cloud Computing (EC2) platform allows applications to run on an instantly scalable number of processors on demand, while Amazon's Simple Storage System (S3) allows access to a practically infinite allocation of disk space on demand. The Amazon EC2 platform allows applications to use as much processing power as they need at any given time, scaling up and down parallel to the demand. Similarly, S3 allows applications to scale storage needs exactly in parallel with demand.

Amazon began AWS by charging directly in proportion to usage (Amazon EC2 charges anywhere from $0.10 to $0.80 per processor hour while S3 charges up to $0.14 per GB per month of storage, with bandwidth costs of $0.10 to $0.15 per GB of bandwidth downloaded or uploaded. This inexpensive, pay-as-you-go price scheme eliminates the risk associated with investing in technologies never tested, encouraging system administrators and curious programmers to play with the service at extremely low costs.

4.2 DATA PRESENTATION

Amazon web services cloud platform. AWS consists of many cloud services and to access these services the AWS Management Console, and the AWS Command Line Interface is used.

AWS management console. Access and manage Amazon Web Services through the AWS Management Console, a simple and intuitive user interface. AWS Command Line Interface

The AWS Command Line Interface (CLI) is a unified tool to manage the AWS services. With just one tool to download and configure, multiple AWS services can be controlled from the command line and automate them through scripts.

4.3 COMPUTE

Amazon EC2. Amazon Elastic Compute Cloud (Amazon EC2) is a web service which provides secure, resizable compute capacity in the cloud. It is designed to make web-scale computing easier for developers and to reduce the time required to obtain and boot new server instances (called Amazon EC2 instances) to minutes, allowing to quickly scale capacity, both up and down, as the computing requirements change time to time.

Benefits.

Elastic web-scale computing. Amazon EC2 enables to increase or decrease the capacity within minutes. Hundreds of thousands of server's instances can be controlled simultaneously. Because the instances are controlled by web service APIs, the application can automatically scale itself up and down depending on its needs.

Completely controlled. There is a complete control of the Amazon EC2 instances having root access to each instance. While retaining the data on the boot partition, the Amazon EC2 instances can be stopped and then can be restarted subsequently using web service APIs. Instances can be rebooted remotely using web service APIs.

Flexible cloud hosting services. There are multiple options for the instance types, operating systems, and software packages to choose from. Amazon EC2 allows the users to select the memory configuration, CPU, instance storage, and boot partition size.

Integrated

Amazon EC2 is integrated with most AWS services, such as Amazon Simple Storage Service (Amazon S3), Amazon Relational Database Service (Amazon RDS), and Amazon Virtual Private Cloud (Amazon VPC) to provide a complete, secure solution for computing, query processing, and cloud storage across a wide range of applications.

Reliable. Amazon EC2 offers a highly reliable environment where replacement instances can be rapidly and predictably commissioned.

Secure. Amazon EC2 works in conjunction with Amazon VPC to provide security and robust networking functionality. The compute instances are in a VPC with an IP address range specified by the user which are exposed to the internet either to remain private or public. Security groups and network access control lists (ACLs) allows the user to control inbound and outbound network access to and from the instances.

The users can connect their existing IT infrastructure to resources in the VPC using industry-standard encrypted IPsec virtual private network (VPN) connections.

Inexpensive. Amazon EC2 instances can be used at a very low rate for the compute capacity consumed by the users. On-Demand Instances

With On-Demand instances, the users pay for computing capacity by the hour with no long-term commitments. The users can increase or decrease the compute capacity depending on the demands of the application and only pay the specified hourly rate for the instances used. The use of On-Demand instances frees the users from the costs and complexities of planning, purchasing, and maintaining hardware and transforms the large fixed costs into much smaller variable costs.

4.4 STORAGE

Amazon Elastic Block Storage. Amazon Elastic Block Store (EBS) is an easy to use, high-performance, block-storage service designed for use with Amazon Elastic Compute Cloud (EC2) for both throughput and transaction intensive workloads at any scale. A broad range of workloads, such as relational and non-relational databases, enterprise applications, containerized applications, big data analytics engines, file systems, and media workflows are widely deployed on Amazon EBS.

You can choose from six different volume types to balance optimal price and performance. You can achieve single-digit-millisecond latency for high-performance database workloads such as SAP HANA or gigabyte per second throughput for large, sequential workloads such as Hadoop. You can change volume types, tune performance, or increase

volume size without disrupting your critical applications, so you have cost-effective storage when you need it.

Designed for mission-critical systems, EBS volumes are replicated within an Availability Zone (AZ) and can easily scale to petabytes of data. Also, you can use EBS Snapshots with automated lifecycle policies to back up your volumes in Amazon S3, while ensuring geographic protection of your data and business continuity.

Amazon ELB features.

Performance for any workload:

EBS volumes are performant for your most demanding workloads, including mission-critical applications such as SAP, Oracle, and Microsoft products. SSD-backed options include a volume designed for high performance applications and a general-purpose volume that offers strong price/performance for most workloads.

Highly available and durable:

Amazon EBS architecture offers reliability for mission-critical applications. EBS volumes are designed to protect against failures by replicating within the Availability Zone (AZ), offering 99.999% availability. Manager (DLM) policies to automate snapshot management.

Cost-effective:

EBS offers six different volumes at various price points and performance benchmarks, enabling you to optimize costs and invest in a precise level of storage for your application needs.

Easy to Use:

Amazon EBS volumes are easy to create, use, encrypt, and protect. Elastic Volumes capability allows you to increase storage, tune performance up and down, and change volume types without any disruption to your workloads.

Virtually unlimited scale:

Amazon EBS enables you to increase storage without any disruption to your critical workloads, build applications that require as little as a single GB of storage, or scale up to petabytes of data — all in just a few clicks.

Secure:

EBS is built to be secure for data compliance. Newly-created EBS volumes can be encrypted by default with a single setting in your account.

4.5 SECURITY

AWS security. Cloud security at AWS is the highest priority because there are no physical servers or datacenters needed for processing and providing security to the database. All the migration, security and processing of the database is provided through software tools which cost far more less time, money and infrastructure for maintenance compared to the physical servers and storage devices.

An advantage of the AWS Cloud is that it allows the user to scale and innovate while maintaining a secure environment and paying only for the services they use. This means that they can have the security at a lower cost than in an on-premises environment.

Benefits of AWS security.

Keep data safe. The AWS infrastructure puts strong safeguards in place to help protect the user privacy. All data is stored in highly secure AWS data centers.

Meet Compliance Requirements: AWS manages dozens of compliance programs in its infrastructure. This means that segments of the compliance have already been completed.

Save money. Cut costs by using AWS data centers. Maintain the highest standard of security without having to manage your own facility.

Scale Quickly: Security scales with the AWS Cloud usage. No matter the size of the business, the AWS infrastructure is designed to keep the user's data safe.

4.6 NETWORKING

Amazon VPC. Amazon Virtual Private Cloud (Amazon VPC) allows the user to create a logically isolated section of the AWS Cloud where they can launch AWS resources in a virtual network as defined. The user has complete control over the virtual networking environment, including the selection of their own IP address range, the creation of subnets, and configuration

of routing tables and network gateways. Both IPv4 and IPv6 in the VPC can be used for secure and easy access to resources and applications.

The network configuration for the VPC can easily be customized. There are basically two subnets for the web servers to access the database. The private subnet comprises of all the sensitive database and backend system which does not have access to internet whereas the public subnets have the web servers which have complete access to the internet.

4.7 DATA ANALYSIS

a. Create and launch EC2 lamp Instance.

To launch the Lamp EC2 instance

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.

2. In the marketplace choose **Lamp.**

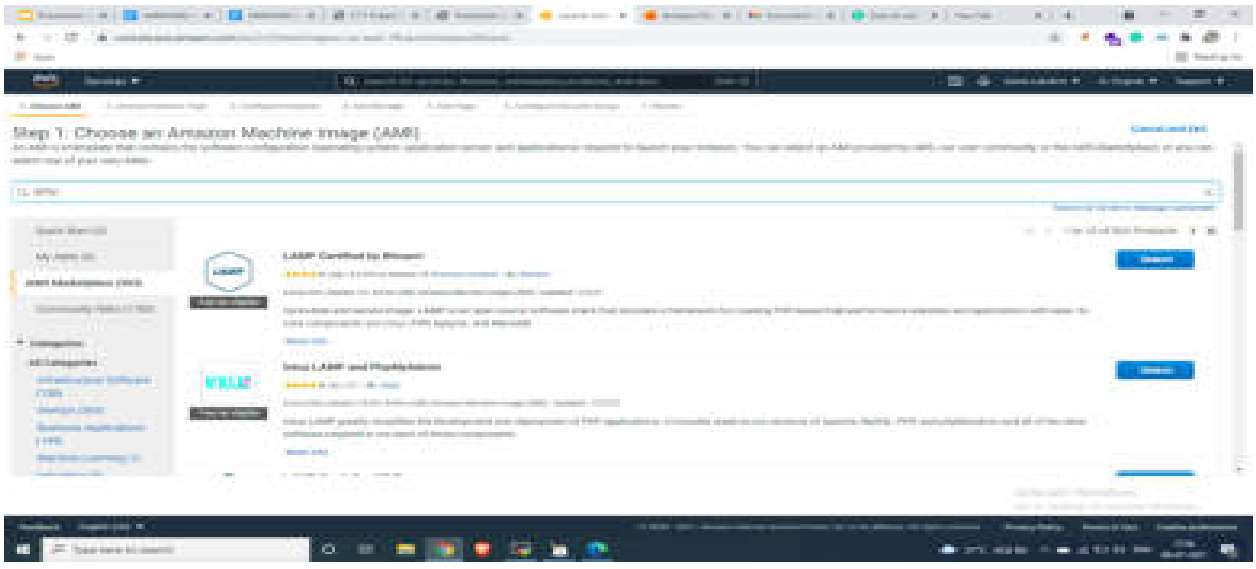


Fig: 4.1

3. Choose Instance type

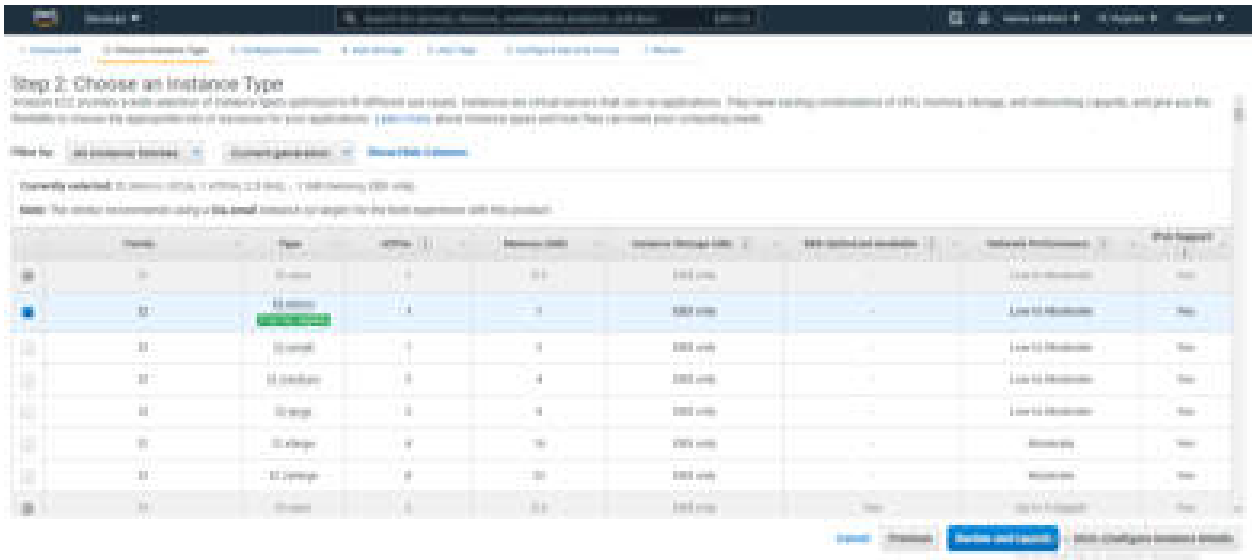


Fig: 4.2

4. Configure Instance

Fig: 4.3

5. Add Storage

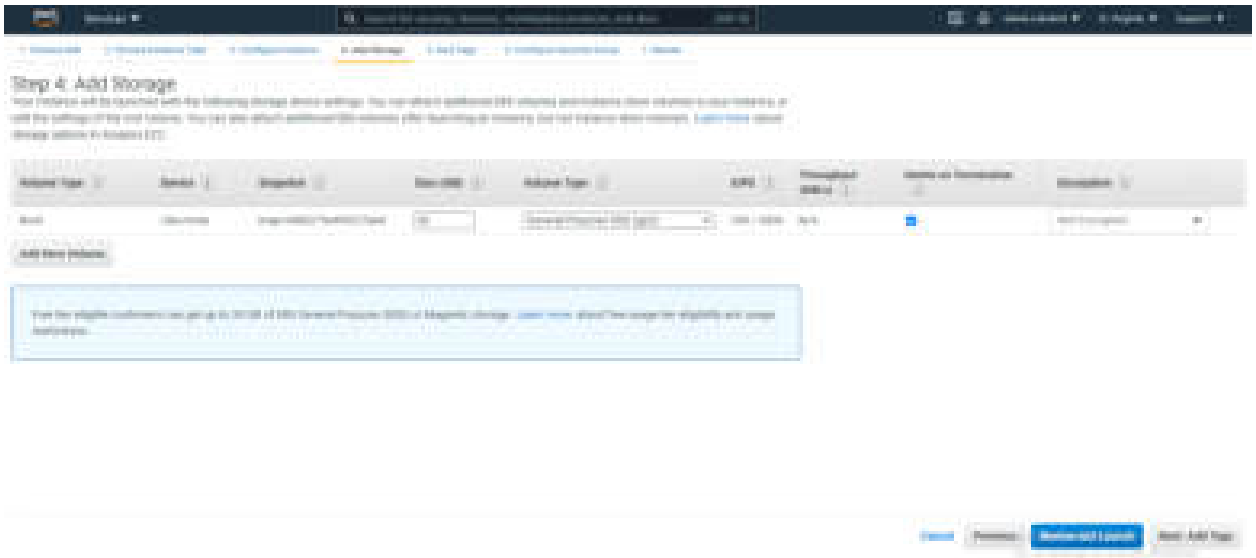


Fig: 4.4

6. Add Tags

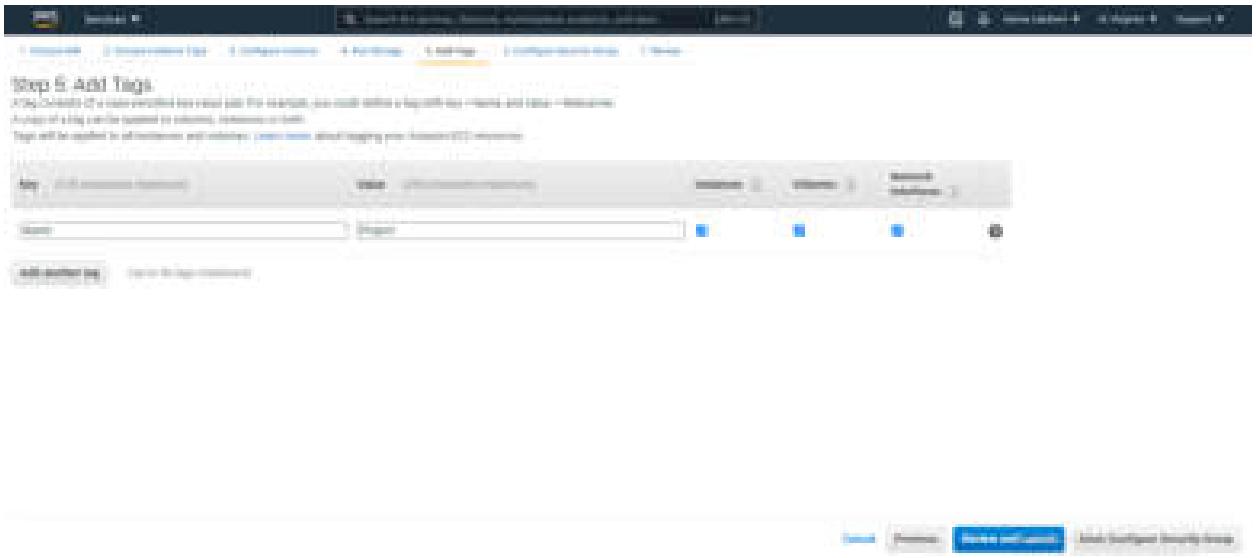


Fig: 4.5

8. Configure Security Group:

Fig: 4.6

9. Review:

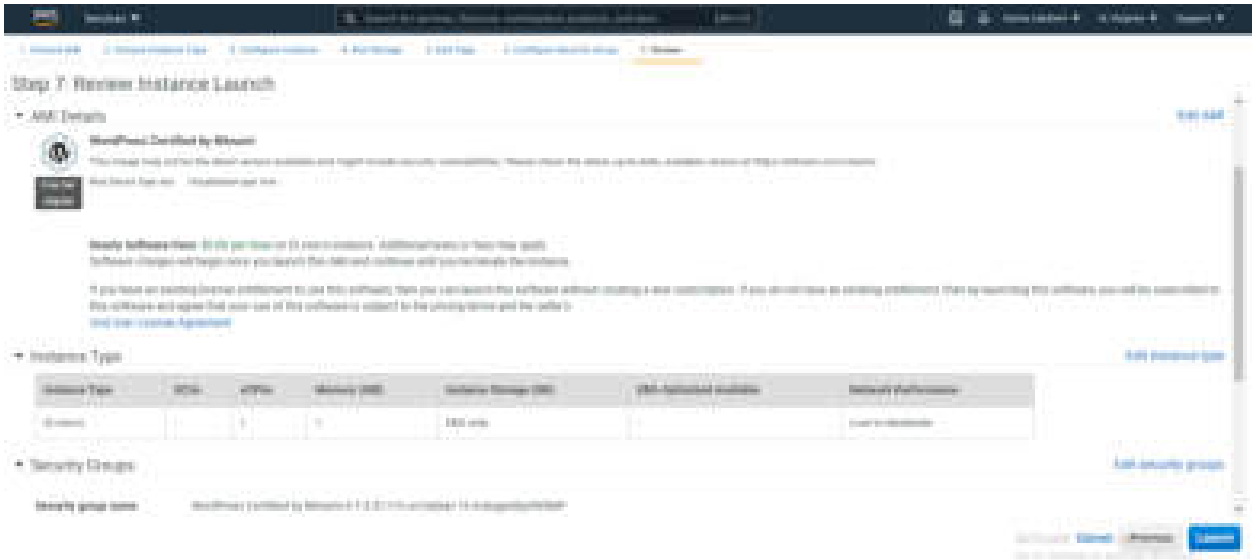


Fig: 4.7

b. Creating Load Balancer:

1. **Configure Load Balancer:**

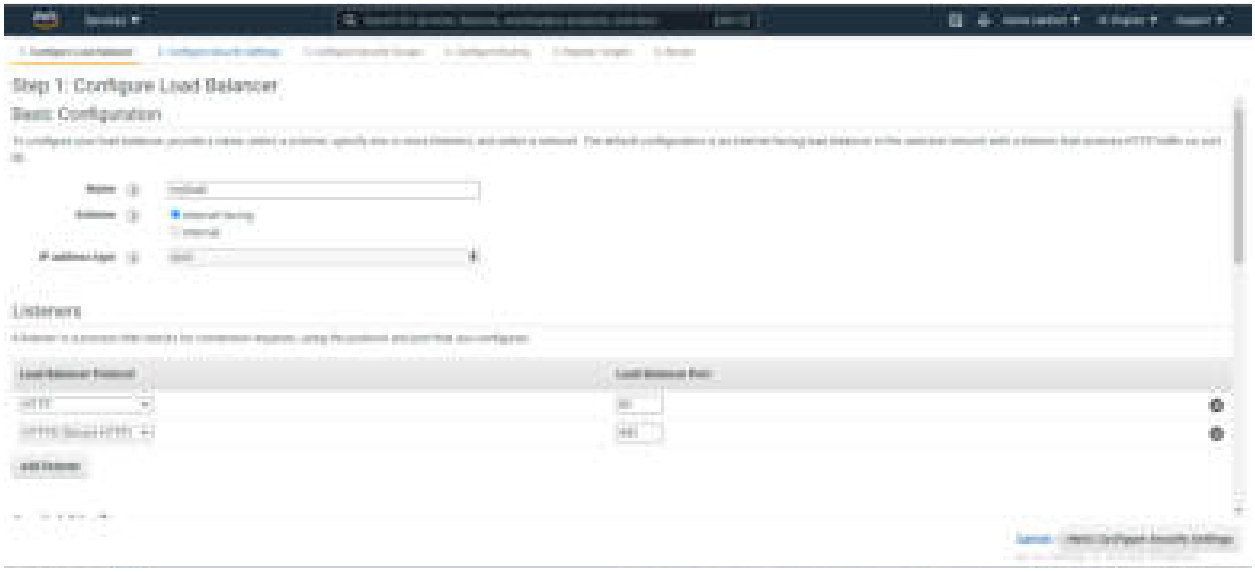


Fig: 4.8

2. **Configure Security Settings:**

Fig: 4.9

3. **Configure Security Groups:**

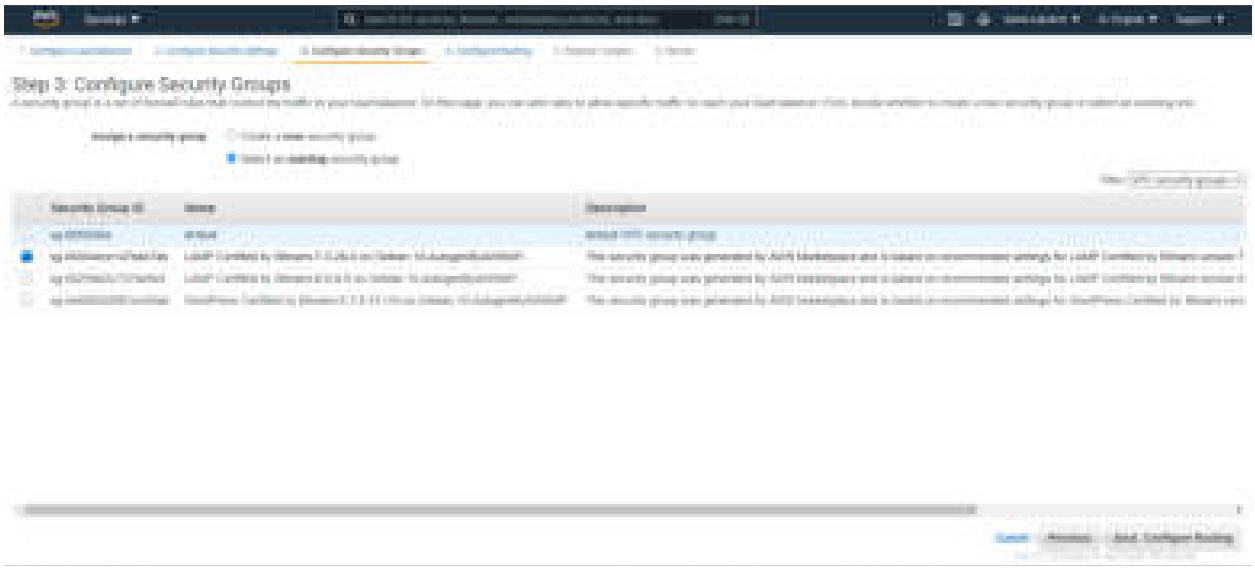


Fig: 4.10

4. Configure Routing:

Fig: 4.11

5. Register targets:

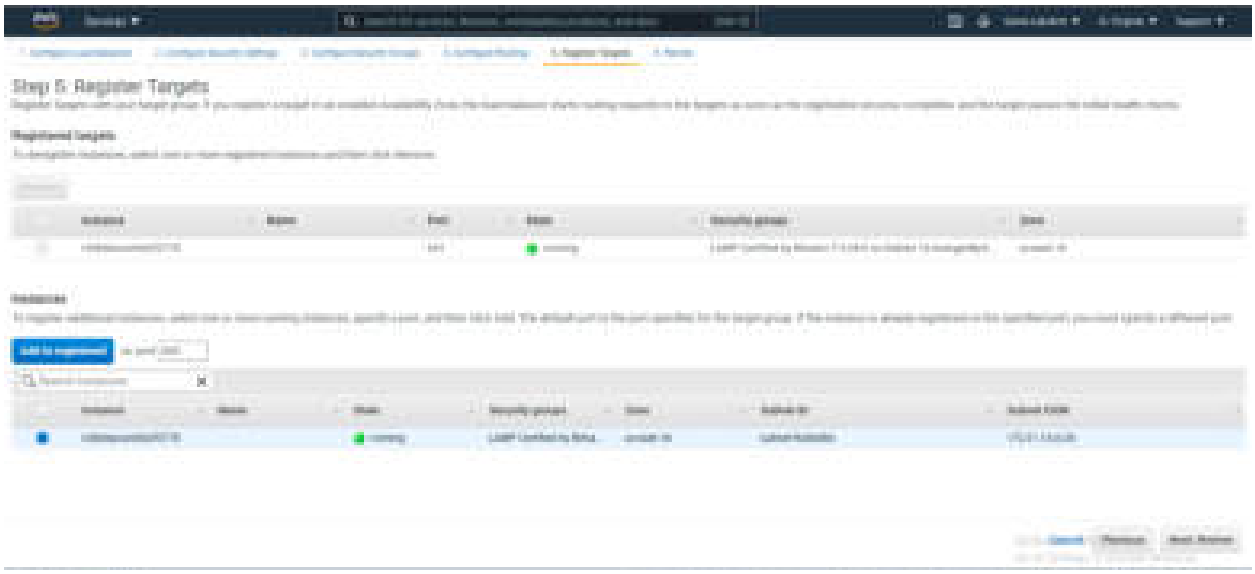


Fig: 4.12

6. Review and Create

Fig: 4.13

Connecting to the Instance:

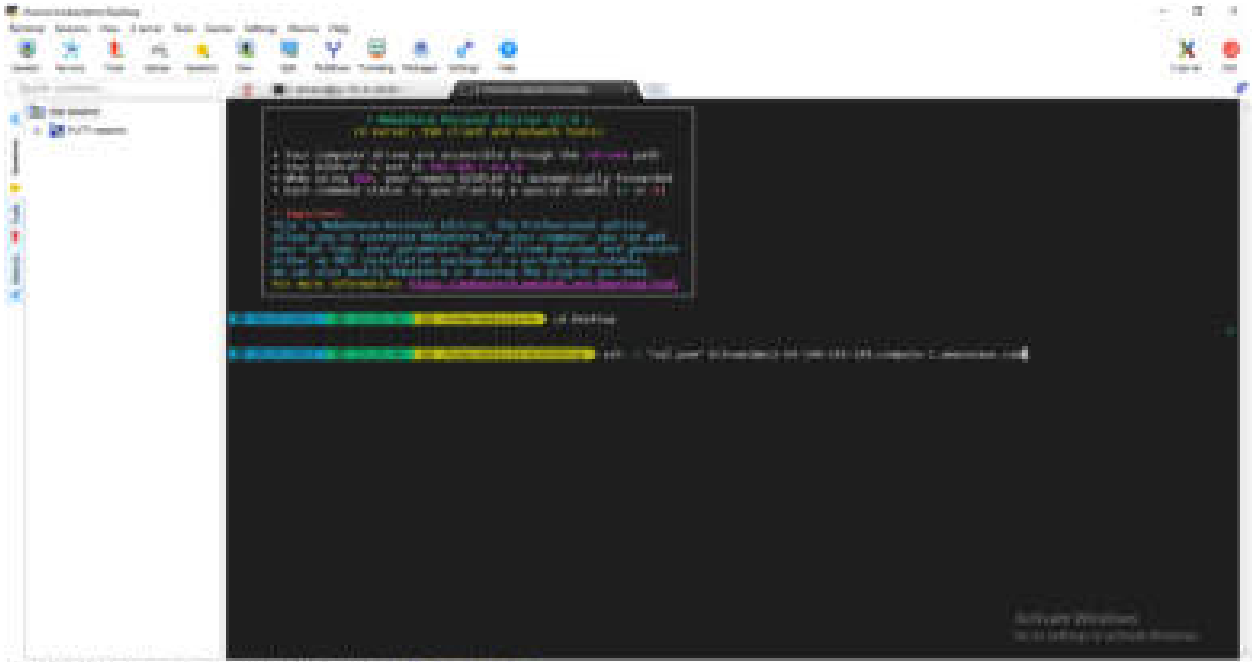


Fig: 4.14

Uploading webpages to the server:

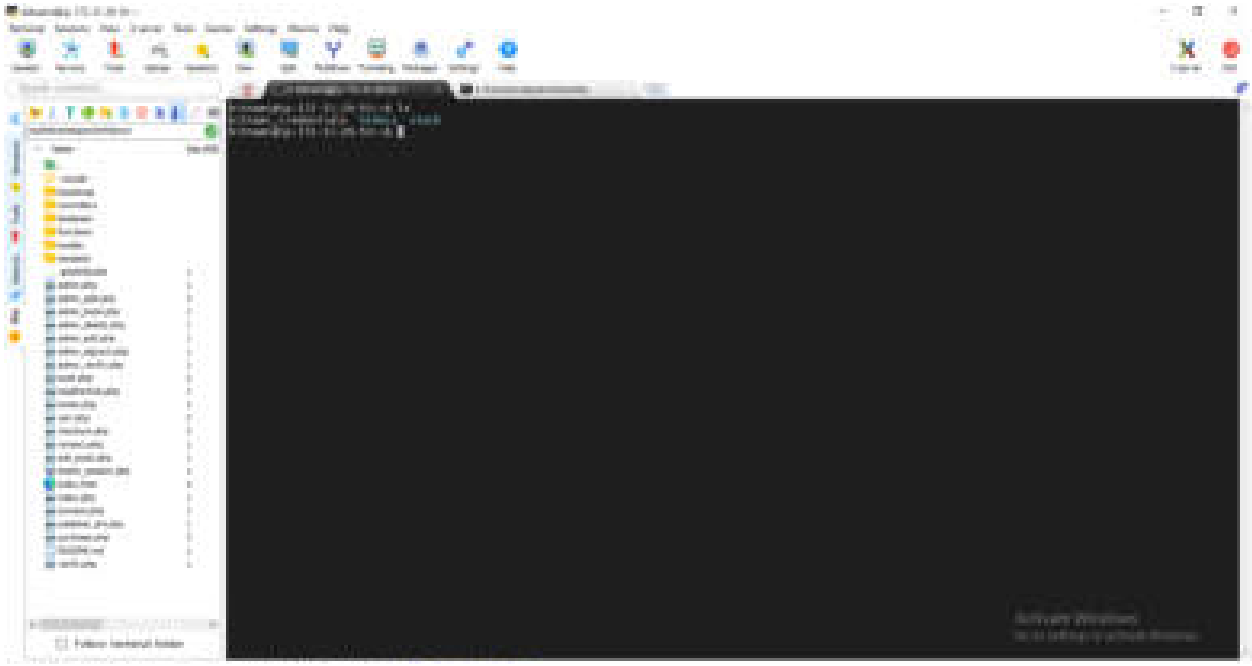


Fig: 4.15

4.8 SOURCE CODE FOR WEB APPLICATION:

Index.php:

<?php

```
  session_start();
  $count = 0;
  // connecto database

  $title = "Index";
  require_once "./template/header.php";
  require_once "./functions/database_functions.php";
  $conn = db_connect();
  $row = select4LatestBook($conn);
?>
     <!-- Example row of columns -->
     <p class="lead text-center text-muted">Latest books</p>
     <div class="row">
       <?php foreach($row as $book) { ?>
      <div class="col-md-3">
          <a href="book.php?bookisbn=<?php echo $book['book_isbn']; ?>">
         <img class="img-responsive img-thumbnail" src="./bootstrap/img/<?php echo
$book['book_image']; ?>">
        </a>
      </div>
       <?php } ?>
     </div>
<?php
  if(isset($conn)) {mysqli_close($conn);}
  require_once "./template/footer.php";
?>
```

ADMIN.PHP

```
<?php
        $title = "Administration section";
        require_once "./template/header.php";
```

```
?>

<div class="row">
<div class="col-md-3"></div>
<div class="col-md-6 text-center">
        <form class="form-horizontal" method="post" action="admin_verify.php">
        <fieldset>
        <legend>Admin</legend>
                <div class="form-group">
                        <label for="name" class="control-label col-md-4">Username</label>
                        <div class="col-md-4">
                                <input type="text" name="name" class="form-control">
                        </div>
                </div>
                <div class="form-group">
                        <label for="pass" class="control-label col-md-4">Password</label>
                        <div class="col-md-4">
                                <input type="password" name="pass" class="form-control">
                        </div>
                </div>
                <input type="submit" name="submit" class="btn btn-primary">
                </fieldset>
        </form>
</div>
</div>
<?php
        require_once "./template/footer.php";
?>
```

ADMIN_BOOK.PHP

```
<?php
```

```
        session_start();
        require_once "./functions/admin.php";
        $title = "List book";
        require_once "./template/header.php";
        require_once "./functions/database_functions.php";
        $conn = db_connect();
        $result = getAll($conn);
?>
        <p class="lead"><a href="admin_add.php">Add new book</a></p>
        <a href="admin_signout.php" class="btn btn-primary">Sign out!</a>
        <table class="table" style="margin-top: 20px">
                <tr>
                        <th>ISBN</th>
                        <th>Title</th>
                        <th>Author</th>
                        <th>Image</th>
                        <th>Description</th>
                        <th>Price</th>
                        <th>Publisher</th>
                        <th> </th>
                        <th> </th>
                </tr>
                <?php while($row = mysqli_fetch_assoc($result)){ ?>
                <tr>
                        <td><?php echo $row['book_isbn']; ?></td>
                        <td><?php echo $row['book_title']; ?></td>
                        <td><?php echo $row['book_author']; ?></td>
                        <td><?php echo $row['book_image']; ?></td>
                        <td><?php echo $row['book_descr']; ?></td>
                        <td><?php echo $row['book_price']; ?></td>
                        <td><?php echo getPubName($conn, $row['publisherid']); ?></td>
```

```
                        <td><a href="admin_edit.php?bookisbn=<?php echo $row['book_isbn'];
?>">Edit</a></td>
                        <td><a href="admin_delete.php?bookisbn=<?php echo
$row['book_isbn']; ?>">Delete</a></td>
                    </tr>
                    <?php } ?>
            </table>

<?php
        if(isset($conn)) {mysqli_close($conn);}
        require_once "./template/footer.php";
?>
```

BOOK.PHP

```
<?php
  session_start();
  $book_isbn = $_GET['bookisbn'];
  // connecto database
  require_once "./functions/database_functions.php";
  $conn = db_connect();

  $query = "SELECT * FROM books WHERE book_isbn = '$book_isbn'";
  $result = mysqli_query($conn, $query);
  if(!$result){
    echo "Can't retrieve data " . mysqli_error($conn);
    exit;
  }

  $row = mysqli_fetch_assoc($result);
  if(!$row){
```

```
    echo "Empty book";
    exit;
  }

  $title = $row['book_title'];
  require "./template/header.php";
?>
      <!-- Example row of columns -->
      <p class="lead" style="margin: 25px 0"><a href="books.php">Books</a> > <?php echo
$row['book_title']; ?></p>
      <div class="row">
        <div class="col-md-3 text-center">
          <img class="img-responsive img-thumbnail" src="./bootstrap/img/<?php echo
$row['book_image']; ?>">
        </div>
        <div class="col-md-6">
          <h4>Book Description</h4>
          <p><?php echo $row['book_descr']; ?></p>
          <h4>Book Details</h4>
          <table class="table">
            <?php foreach($row as $key => $value){
              if($key == "book_descr" || $key == "book_image" || $key == "publisherid" || $key ==
"book_title"){
                continue;
              }
              switch($key){
                case "book_isbn":
                  $key = "ISBN";
                  break;
                case "book_title":
                  $key = "Title";
```

```
            break;
          case "book_author":
            $key = "Author";
            break;
          case "book_price":
            $key = "Price";
            break;
        }
      ?>
      <tr>
        <td><?php echo $key; ?></td>
        <td><?php echo $value; ?></td>
      </tr>
      <?php
        }
        if(isset($conn)) {mysqli_close($conn); }
      ?>
    </table>
    <form method="post" action="cart.php">
      <input type="hidden" name="bookisbn" value="<?php echo $book_isbn;?>">
      <input type="submit" value="Purchase / Add to cart" name="cart" class="btn
btn-primary">
    </form>
      </div>
  </div>
<?php
  require "./template/footer.php";
?>
```

TEMPLATES:

ADMIN.PHP

```
<?php
	if(!isset($_SESSION['admin']) && $_SESSION['admin'] != true){
		header("Location: index.php");
	}
?>
```

CART_FUNCTIONS.PHP

```
<?php
	/*
		loop through array of $_SESSION['cart'][book_isbn] => number
		get isbn => take from database => take book price
		price * number (quantity)
		return sum of price
	*/
	function total_price($cart){
		$price = 0.0;
		if(is_array($cart)){
			foreach($cart as $isbn => $qty){
				$bookprice = getbookprice($isbn);
				if($bookprice){
					$price += $bookprice * $qty;
				}
			}
		}
		return $price;
	}

	/*
		loop through array of $_SESSION['cart'][book_isbn] => number
		$_SESSION['cart'] is associative array which is [book_isbn] => number of books
for each book_isbn
```

```
            calculate sum of books
        */
        function total_items($cart){
            $items = 0;
            if(is_array($cart)){
                foreach($cart as $isbn => $qty){
                    $items += $qty;
                }
            }
            return $items;
        }
?>
```

DATABASE_FUNCTIONS.PHP

```
<?php
        function db_connect(){
            $conn = mysqli_connect("localhost", "root", "", "www_project");
            if(!$conn){
                echo "Can't connect database " . mysqli_connect_error($conn);
                exit;
            }
            return $conn;
        }

        function select4LatestBook($conn){
            $row = array();
            $query = "SELECT book_isbn, book_image FROM books ORDER BY
book_isbn DESC";
            $result = mysqli_query($conn, $query);
            if(!$result){
               echo "Can't retrieve data " . mysqli_error($conn);
```

```
            exit;
        }
        for($i = 0; $i < 4; $i++){
            array_push($row, mysqli_fetch_assoc($result));
        }
        return $row;
    }

    function getBookByIsbn($conn, $isbn){
        $query = "SELECT book_title, book_author, book_price FROM books WHERE
book_isbn = '$isbn'";
        $result = mysqli_query($conn, $query);
        if(!$result){
            echo "Can't retrieve data " . mysqli_error($conn);
            exit;
        }
        return $result;
    }

    function getOrderId($conn, $customerid){
        $query = "SELECT orderid FROM orders WHERE customerid = '$customerid'";
        $result = mysqli_query($conn, $query);
        if(!$result){
            echo "retrieve data failed!" . mysqli_error($conn);
            exit;
        }
        $row = mysqli_fetch_assoc($result);
        return $row['orderid'];
    }
```

```
	function insertIntoOrder($conn, $customerid, $total_price, $date, $ship_name,
$ship_address, $ship_city, $ship_zip_code, $ship_country){
		$query = "INSERT INTO orders VALUES
		('" . $customerid . "', '" . $total_price . "', '" . $date . "', '" . $ship_name . "', '" .
$ship_address . "', '" . $ship_city . "', '" . $ship_zip_code . "', '" . $ship_country . "')";
		$result = mysqli_query($conn, $query);
		if(!$result){
			echo "Insert orders failed " . mysqli_error($conn);
			exit;
		}
	}

	function getbookprice($isbn){
		$conn = db_connect();
		$query = "SELECT book_price FROM books WHERE book_isbn = '$isbn'";
		$result = mysqli_query($conn, $query);
		if(!$result){
			echo "get book price failed! " . mysqli_error($conn);
			exit;
		}
		$row = mysqli_fetch_assoc($result);
		return $row['book_price'];
	}

	function getCustomerId($name, $address, $city, $zip_code, $country){
		$conn = db_connect();
		$query = "SELECT customerid from customers WHERE
		name = '$name' AND
		address= '$address' AND
		city = '$city' AND
		zip_code = '$zip_code' AND
```

```
            country = '$country'";
            $result = mysqli_query($conn, $query);
            // if there is customer in db, take it out
            if($result){
                $row = mysqli_fetch_assoc($result);
                return $row['customerid'];
            } else {
                return null;
            }
        }

        function setCustomerId($name, $address, $city, $zip_code, $country){
            $conn = db_connect();
            $query = "INSERT INTO customers VALUES
                ('', '" . $name . "', '" . $address . "', '" . $city . "', '" . $zip_code . "', '" .
$country . "')";

            $result = mysqli_query($conn, $query);
            if(!$result){
                echo "insert false !" . mysqli_error($conn);
                exit;
            }
            $customerid = mysqli_insert_id($conn);
            return $customerid;
        }

        function getPubName($conn, $pubid){
            $query = "SELECT publisher_name FROM publisher WHERE publisherid =
'$pubid'";
            $result = mysqli_query($conn, $query);
            if(!$result){
```

```
                echo "Can't retrieve data " . mysqli_error($conn);
                exit;
            }
            if(mysqli_num_rows($result) == 0){
                echo "Empty books ! Something wrong! check again";
                exit;
            }

            $row = mysqli_fetch_assoc($result);
            return $row['publisher_name'];
        }

        function getAll($conn){
            $query = "SELECT * from books ORDER BY book_isbn DESC";
            $result = mysqli_query($conn, $query);
            if(!$result){
                echo "Can't retrieve data " . mysqli_error($conn);
                exit;
            }
            return $result;
        }
?>
```

HEADER.PHP

```
<?php
        //session_start();
        error_reporting(E_NOTICE);
?>
<!DOCTYPE html>
<html lang="en">
  <head>
```

```
    <meta charset="utf-8">
    <meta http-equiv="X-UA-Compatible" content="IE=edge">
    <meta name="viewport" content="width=device-width, initial-scale=1">

    <title><?php echo $title; ?></title>

    <link href="./bootstrap/css/bootstrap.min.css" rel="stylesheet">
    <link href="./bootstrap/css/bootstrap-theme.min.css" rel="stylesheet">
    <link href="./bootstrap/css/jumbotron.css" rel="stylesheet">
  </head>

  <body>

    <nav class="navbar navbar-inverse navbar-fixed-top">
      <div class="container">
        <div class="navbar-header">
          <button type="button" class="navbar-toggle collapsed" data-toggle="collapse"
data-target="#navbar" aria-expanded="false" aria-controls="navbar">
            <span class="sr-only">Toggle navigation</span>
            <span class="icon-bar"></span>
            <span class="icon-bar"></span>
            <span class="icon-bar"></span>
          </button>
          <a class="navbar-brand" href="index.php">CSE Bookstore</a>
        </div>

        <!--/.navbar-collapse -->
        <div id="navbar" class="navbar-collapse collapse">
```

```
        <ul class="nav navbar-nav navbar-right">
          <!-- link to publiser_list.php -->
          <li><a href="publisher_list.php"><span class="glyphicon
glyphicon-paperclip"></span>  Publisher</a></li>
          <!-- link to books.php -->
          <li><a href="books.php"><span class="glyphicon glyphicon-book"></span> 
Books</a></li>
          <!-- link to contacts.php -->
          <li><a href="contact.php"><span class="glyphicon
glyphicon-phone-alt"></span>  Contact</a></li>
          <!-- link to shopping cart -->
          <li><a href="cart.php"><span class="glyphicon
glyphicon-shopping-cart"></span>  My Cart</a></li>
        </ul>

      </div>
    </div>
  </nav>
  <?php
    if(isset($title) && $title == "Index") {
  ?>
  <!-- Main jumbotron for a primary marketing message or call to action -->
  <div class="jumbotron">
    <div class="container">
      <h1>Welcome to online CSE bookstore</h1>
      <p class="lead">This site has been made using PHP with MYSQL (procedure
functions)!</p>
      <p>The layout use Bootstrap to make it more responsive. It's just a simple web!</p>
    </div>
  </div>
```

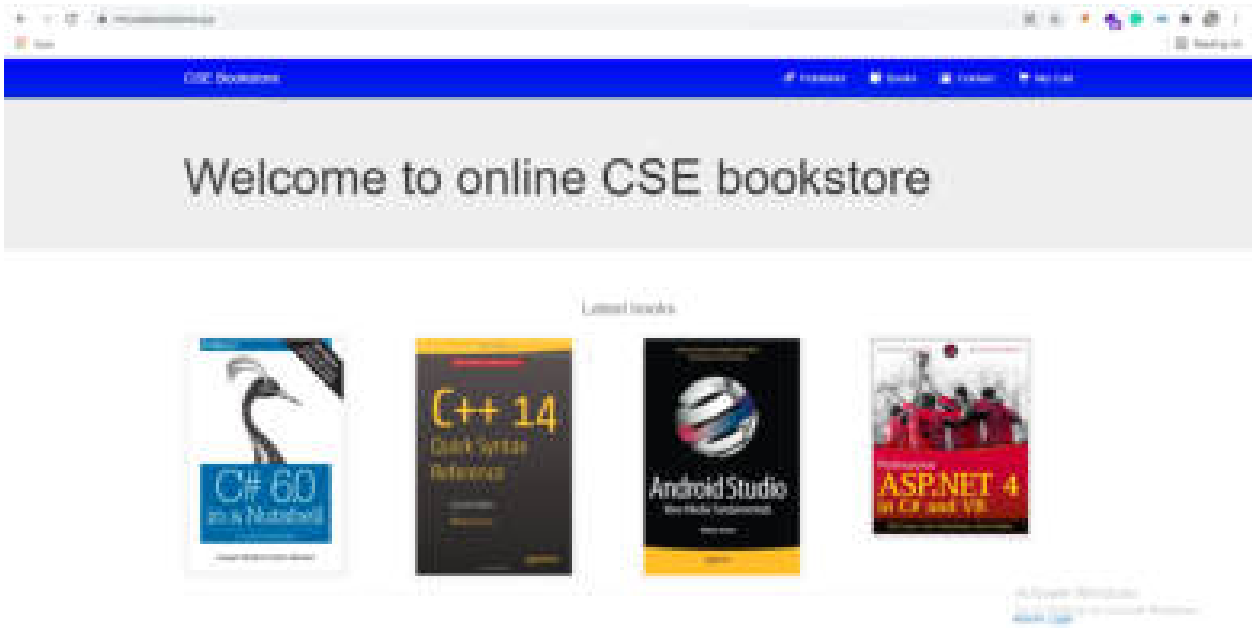
```
<?php } ?>

<div class="container" id="main">
```

FOOTER.PHP:

```
        <hr>

        <footer>

        <div class="text-muted pull-right">
                <a href="admin.php">Admin Login</a>
          <a href="user.php">User Login</a>
        </div>
        </footer>
</div> <!-- /container -->


<!-- Bootstrap core JavaScript
================================================== -->
<!-- Placed at the end of the document so the pages load faster -->
<script type="text/javascript" src="./bootstrap/js/jquery-2.1.4.min.js"></script>
<script type="text/javascript" src="./bootstrap/js/bootstrap.min.js"></script>
</body>
</html>
```

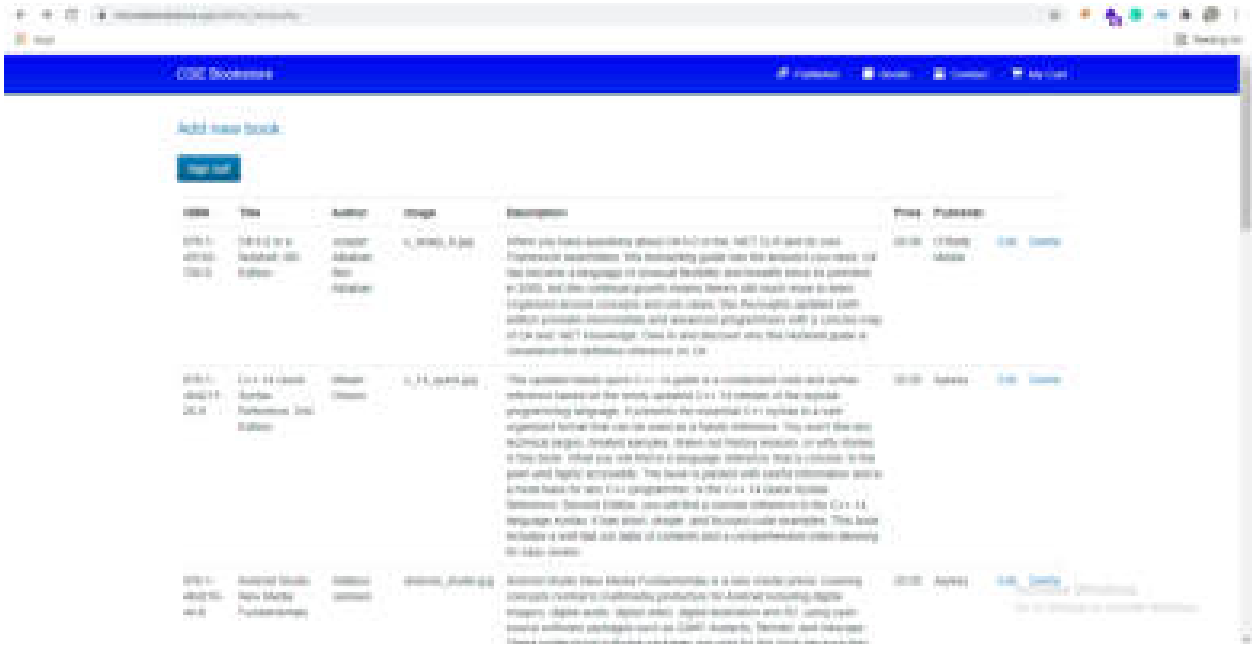
CHAPTER 5: RESULTS

HOMEPAGE:



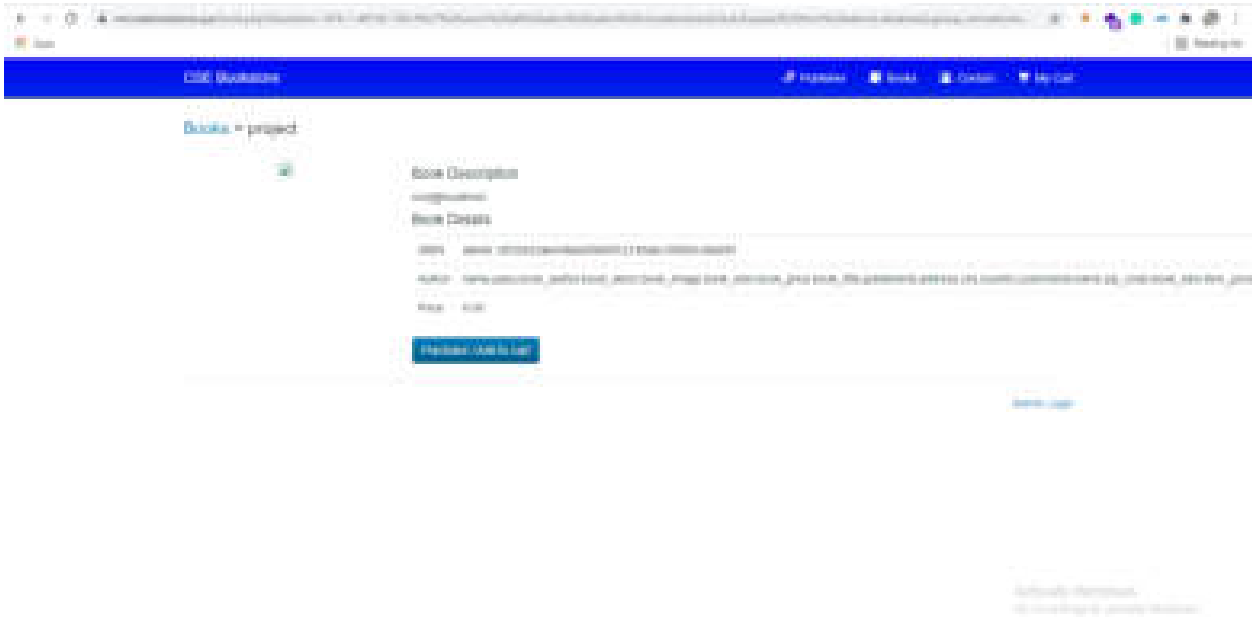
ADMIN LOGIN:

ADMIN DASHBOARD:



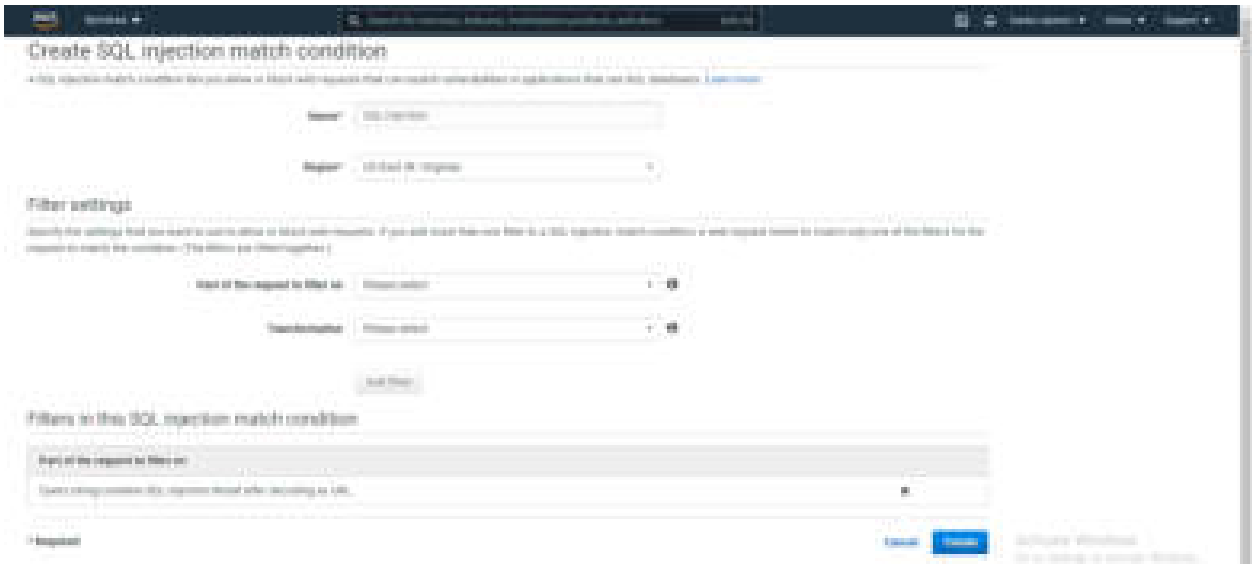
PRODUCT PAGE:

ATTACK:



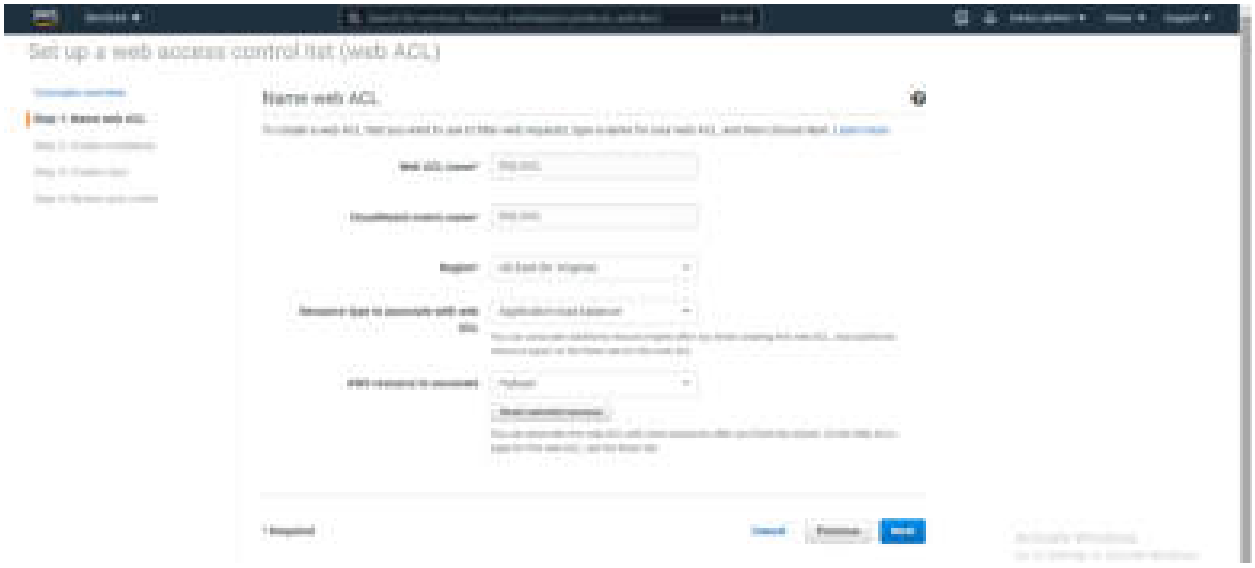
USING AWS WAF:

1. Create sql injection match condition



2. Create sql injection rules

3. Adding rules to webacl



4. Output

SUMMARY:

AWS consists of some major key components like EC2 instances, Storage Services, VPN, public and private subnets which are explained in analyzing the results. Steps in creating the Ec2 instances, coding the IPv4 and IPv6 instances and creating the public and private subnets with steps in creating the amazon s3 buckets both in IPv4 and IPv6 have been explained. Steps for creating the WebserverSG security group is also been discussed.

CHAPTER 5: CONCLUSIONS AND FUTURE WORK

Amazon Web Services Web Application Firewall (AWS WAF) helps to protect web applications from common web exploits like SQL injection attacks that could affect application availability, compromise security, or consume excessive resources. Amazon's Elastic Cloud Computing (EC2) platform allows applications to run on an instantly scalable number of processors on demand, while Amazon Elastic Block Store (EBS) is an easy to use, high-performance, block-storage service designed for use with Amazon Elastic Compute Cloud (EC2) for both throughput and transaction intensive workloads at any scale. A broad range of workloads, such as relational and non-relational databases, enterprise applications, containerized applications, big data analytics engines, file systems, and media workflows are widely deployed on Amazon EBS.

1) Does the AWS firewall provide better security than the WAF?

A) Yes, it does provide multiple layers of security at every stage of process to avoid SQL injection attacks as it provides complete control to the user over the virtual networking environment, including selection of the own IP address range, creation of subnets, and configuration of route tables and network gateways as well as creating public and private subnets.

2) Does AWS provide high performance databases?

A) Amazon EBS provides the most durable, cost effective and highly secured databases.

3) Is AWS feasible for any organization?

A) Yes because of the cost-effective system and the users have to pay for what they use it is very convenient for any scale of organization.

CONCLUSION

As the WAF was not completely capable of defending the SQL injection attacks, AWS is being used because of the multiple layers of security and access it provides for the databases either by providing a private VPN with gateway authorities or by creating multiple subnets or by providing access to ports for the databases.

AWS provides a required number of processors on demand as well as a scalable number of databases on demand with multiple layers of security in the form of VPN's, gateways, portals,

public and private subnets avoiding the hardware requirements for the organizations. As the subnets can be created public and private the data and the permissions can be secured and authenticated at different role levels which protect the integrity and security of the data as the critical information will not be available to all the users. Future Work In-depth study of the AWS management console should be carried out like studying about the glacier, snowball which is an advanced level of data storage services in AWS. Similarly, creating multiple subnets in IPv4, IPv6 and route 53 privacy.

CHAPTER 7: REFERENCES

[1] Anley, C., “Advanced SQL Injection in SQL Server Applications,” White paper, Next Generation Security Software Ltd, 2002.

[2] Boyd, S. W., Keromytis, A. D., “SQLrand: Preventing SQL Injection Attacks,” In Proceedings of the 2nd Applied Cryptography and Network Security (ACNS) Conference, v. 3089, pp. 292–302, June 2004.

[3] Buehrer, G., Weide, B. W., & Sivilotti, P. A. G. “Using parse tree validation to prevent SQL injection attacks.” Proceedings of the 5th International Workshop on Software Engineering and Middleware, Lisbon, Portugal, pp. 106-113. doi:10.1145/1108473.1108496, 2005.

[4] Cook, W. R., & Rai, S. “Safe query objects: Statically typed objects as remotely executable queries.” Proceedings of the 27th International Conference on Software Engineering, St. Louis, MO, USA, pp. 97-106. doi:10.1145/1062455.1062488, 2005.

[5] Dharam, R.; Shiva, S.G., "Runtime monitors for tautology-based SQL injection attacks," Cyber Security, International conference on Cyber Warfare and Digital Forensic (CyberSec), pp. 26-28, June 2012.

[6] Elia, I. A., Fonseca, J., & Vieira, M., “Comparing SQL injection detection tools using attack injection” An experimental study Software Reliability Engineering (ISSRE), IEEE 21st International Symposium on Software Reliability Engineering, pp. 289- 298. doi:10.1109/ISSRE.2010.32, 2010. 84

[7] Gould, C., Su, Z., & Devanbu, P., “JDBC checker: A static analysis tool for SQL/JDBC applications,” In Proceedings of the 26th International Conference on Software Engineering, pp. 697-698. doi:10.1109/ICSE.2004.1317494, 2004.

[8] Gould, C., Su, Z., & Devanbu, P., “Static Checking of Dynamically Generated Queries in Database Applications,” In Proceedings of the 26th International Conference on Software Engineering, pp. 645–654, doi:10.1109/ICSE.2004.1317494, 2004.

[9] Haldar, V., Chandra, D., & Franz, M., “Dynamic taint propagation for java,” In Proceedings of the 21st Annual Computer Security Applications Conference (ACSAC'05), pp. 309-311. doi:10.1109/CSAC.2005.21, Dec 2005.

[10] Halfond, W. G., Viegas, J., & Orso, A. “AMNESIA: Analysis and monitoring for Neutralizing SQL-injection attacks,” Proceedings of the 20th IEEE/ACM International

Conference on Automated Software Engineering, Long Beach, CA, USA, pp. 174-183, doi:10.1145/1101908.1101935, Nov 2005.

[11] Halfond, W. G., Viegas, J., & Orso, A. "Combining Static Analysis and Runtime Monitoring to Counter SQL-Injection Attacks," In Proceedings of the Third International ICSE Workshop on Dynamic Analysis (WODA 2005), St. Louis, MO, USA, pp. 22–28, doi:10.1145/1101908.1101935, May 2005.

[12] Halfond, W. G., Viegas, J., & Orso, A. (2006, March). A classification of SQL-injection attacks and countermeasures. In Proceedings of the IEEE International Symposium on Secure Software Engineering (Vol. 1, pp. 13-15). IEEE.

[13] Howard, M., & Leblanc, D. E., "Writing secure code," (2nd ed.). Redmond, WA, USA: Microsoft Press, 2003.

[14] Huang, Y., Huang, S., Lin, T., & Tsai, C., "Web application security assessment by fault injection and behavior monitoring," Proceedings of the 12th International Conference on World Wide Web, Budapest, Hungary, pp. 148-159. doi:10.1145/775152.775174, 2003.

[15] Huang, Y., Yu, F., Hang, C., Tsai, C., Lee, D., & Kuo, S., "Securing web application code by static analysis and runtime protection," Proceedings of the 13th International Conference on World Wide Web, New York, NY, USA, pp. 40-52. doi:10.1145/988672.988679, 2004.

[16] Kar, D., Panigrahi, S., & Sundararajan, S. (2016). SQLiGoT: Detecting SQL injection attacks using graph of tokens and SVM. Computers & Security, 60, 206-225.

[17] Kiani, M., Clark, A., & Mohay, G., "Evaluation of anomaly-based character distribution models in the detection of SQL injection attacks." Third International Conference on Availability, Reliability and Security, pp. 47-55. doi:10.1109/ARES.2008.123, 2008.

[18] Litchfield, D., "Web Application Disassembly with ODBC Error Messages," Technical document, @Stake, Inc., 2002.

[19] Livshits, V. B., & Lam, M. S., "Finding security vulnerabilities in java applications with static analysis." Proceedings of the 14th Conference on USENIX Security Symposium, Baltimore, MD. Volume 14, pp. 271–286, Aug. 2005.

[20] Labs, S., "SQL Injection," White paper, SPI Dynamics, Inc., 2002. http://www.spidynamics.com/assets/documents/ WhitepaperSQLInjection.pdf. 86

[21] Mathew, M., "Amazon Web Services Architecture," White Paper, Amazon Web Services, Inc., 2006. https://aws.amazon.com/security/documents/WhitepaperAWSWAF.pdf.

[22] Martin, M., Livshits, B., & Lam, M. S., “Finding application errors and security flaws using PQL: A program query language.” In Proceedings of the 20th annual ACM SIGPLAN conference on Object oriented programming systems languages and applications (OOPSLA 2005), v. 40(10), pp. 365-383. doi:10.1145/1103845.1094840, 2005.

[23] McClure, R. A., & Kruger, I. H., “SQL DOM: Compile time checking of dynamic SQL statements.” Proceedings of 27th International Conference on Software Engineering. ICSE 2005. pp. 88-96, doi:10.1109/ICSE.2005.1553551, 2005.

[24] McDonald, S., “SQL Injection: Modes of attack, defense, and why it matters,” White paper, GovernmentSecurity.org, April 2002.

[25] Pietraszek, T., & Berghe, C. V., “Defending against injection attacks through contextsensitive string evaluation.” Proceedings of the 8th International Conference on Recent Advances in Intrusion Detection, Seattle, WA. pp. 124-145. doi:10.1007/11663812_7, 2006.

[26] Pinzón, C., Paz, J. F. D., Bajo, J., Herrero, Á., & Corchado, E., “AIIDA-SQL: An adaptive intelligent intrusion detector agent for detecting SQL injection attacks.” 2010 10th International Conference on Hybrid Intelligent Systems, pp.doi:10.1109/HIS.2010.5600026, Aug 2010.

[26] Sadeghian, A., Zamani, M., & Ibrahim, S., “SQL injection is still alive: A study on SQL injection signature evasion techniques.” 2013 International Conference on Informatics and Creative Multimedia, pp. 265-268. doi:10.1109/ICICM.2013.52, 2013.

[27] Sadeghian, A., Zamani, M., & Manaf, A. A., “A taxonomy of SQL injection detection and prevention techniques.” 2013 International Conference on Informatics and Creative Multimedia, pp. 53-56. doi:10.1109/ICICM.2013.18, Sept 2013.

[28] Sadeghian, A., Zamani, M., & Manaf, A. A., “A taxonomy of SQL injection detection and prevention techniques.” 2013 International Conference on Informatics and Creative Multimedia, pp. 53-56. doi:10.1109/ICICM.2013.18 Sept 2013.

[29] Spett, K., “Blind SQL injection,” White paper, SPI Dynamics, Inc., 2003. http://www.spidynamics.com/whitepapers/ Blind SQLInjection.pdf. Scott, D., & Sharp, R., “Abstracting application-level web security.” Proceedings of the 11th International Conference on World Wide Web, Honolulu, Hawaii, USA, pp. 396-407. doi:10.1145/511446.511498, 2002.

[30] Su, Z., & Wassermann, G., “The essence of command injection attacks in web applications.” In The 33rd Annual Symposium on Principles of Programming Languages (POPL 2006) SIGPLAN Not., 41(1), pp. 372-382. doi:10.1145/1111320.1111070, Jan 2006.

[31] Tajpour, A., & Shooshtari, M. J. z., “Evaluation of SQL injection detection and prevention techniques.” 2010 2nd International Conference on Computational 88 Intelligence, Communication Systems and Networks, pp. 216-221. doi:10.1109/CICSyN.2010.55, July 2010.

[32] Valeur, F., Mutz, D., & Vigna, G., “A learning-based approach to the detection of SQL attacks.” Proceedings of the Second International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment, Vienna, Austria. pp. 123- 140. doi:10.1007/11506881_8, July 2005.

[33] Wan, M., & Liu, K., “An improved eliminating SQL injection attacks based regular expressions matching.” 2012 International Conference on Control Engineering and Communication Technology, pp. 210-212. doi:10.1109/ICCECT.2012.235, Dec 2012.

[34] Wassermann, G., & Su, Z., “An Analysis Framework for Security in Web Applications,” In Proceedings of the FSE Workshop on Specification and Verification of Component-Based Systems (SAVCBS 2004), pp. 70–78, 2004.

[35] Wei, K., Muthuprasanna, M., & Kothari, S., “Preventing SQL injection attacks in stored procedures.” Australian Software Engineering Conference (ASWEC'06), v.8, pp. 18-21. doi:10.1109/ASWEC.2006.40, April 2006.

[36] Wu, X., & Chan, P. P. K., “SQL injection attacks detection in adversarial environments by k-centers.” 2012 International Conference on Machine Learning and Cybernetics, pp. 406-410. doi:10.1109/ICMLC.2012.6358948, July 2012.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

Kakinada

DETECTION OF SUICIDE RELATED POSTS IN TWITTER DATA STREAM

A Project Report submitted to

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING.

Submitted by

B.Harshitha Reddy(17H71A0515)

P.Sri Alekhya(17H71A0546)

K.Gopi(17H71A0513)

Y.Charishma(17H71A0506)

Under the Guidance of

Ms. K. Vinayasree

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar

MIC College of Technology

an ISO 9001:2015 Certified Institute

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)

Kanchikacherla - 521180, Krishna Dist, A.P, India.

Phone: 08678 - 273535, 273623, Fax: 08678 - 273569

E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS

NBA-Accredited B.Tech (CSE|ECE|EEE|ME)

Accredited by NAAC With Grade A

2020-2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC A
With Grade

2020-2021

CERTIFICATE

This is to certify that the Main project work entitled "**Detection of suicide related posts in twitter data stream**" is a bonafide work carried out by Ms. B.Harshitha Reddy(17H71A0515), Ms. P.Sri Alekhya (17H71A0546), Mr. K.Gopi (17H71A0513), Ms.Y.Charishma(17H71A0506) in partial fulfillment for the award of the degree of Bachelor of Technology in Computer Science & Engineering of Jawaharlal Nehru Technological University, Kakinada during the year 2020-2021.

K . Vinayasree
Project Guide

D. Prasad
Head of Department

Examiner

Dr. K.Srinivas
Principal

ACKNOWLEDGEMENT

We would like to take this opportunity to express my deepest appreciation to the following people to their valuable contribution and assistance with this Project.

Initially, we would like to thank my project supervisor, **Ms. K.Vinayasree, Assistant Professor, Department of Computer Science and Engineering** for the guidance and support, especially for the valuable ideas and knowledge shared to me throughout the Project.

We have the immense pleasure in expressing my thanks and deep sense of gratitude, **Mr.D.Prasad, Head of the Department, Computer Science and Engineering** for extending necessary facilities and support for the completion of the Project.

We whole sincerely acknowledge **Dr. K. Srinivas, Principal and Prof D. PANDURANGA RAO, CEO** for giving opportunity to take up the Project work. We also extended my thanks to all faculty members of **Computer Science and Engineering,** for their valuable Guidance and Encouragement.

We would like to extend my warm appreciations to all my friends for sharing us their knowledge, valuable contributions and help with this Project.

Finally, my special thanks go to my family for their continuous support and help throughout my academic years and for their continual support and encouragement for the completion of the project.

B .Harshitha Reddy (17H71A0515)

P. Sri Alekhya (17H71A0546)

K. Gopi (17H71A0513)

Y .Charishma (17H71A0506)

DECLARATION

We **B.Harshitha Reddy, P.Sri Alekhya, K.Gopi, Y.Charishma** of the Main-Project "**Detection of suicide related posts in twitter data stream**", hereby declare that the matter embodied in this Project is the genuine work done by us and has not been submitted either to this University or to any other university/institute for the fulfilment of the requirement of any course of study.

B .Harshitha Reddy (17H71A0515)

P. Sri Alekhya (17H71A0546)

K. Gopi (17H71A0513)

Y .Charishma (17H71A0506)

ABSTRACT

Suicidal ideation detection in online social networks is an emerging research area with major challenges. Recent research has shown that the publicly available information, spread across social media platforms, holds valuable indicators for effectively detecting individuals with suicidal intentions. The key challenge of suicide prevention is understanding and detecting the complex risk factors and warning signs that may precipitate the event. In this paper, we present a new approach that uses the social media platform Twitter to quantify suicide warning signs for individuals and to detect posts containing suicide-related content. The main originality of this approach is the automatic identification of sudden changes in a user's online behavior. To detect such changes, we combine natural language processing techniques to aggregate behavioral and textual features and pass these features through a martingale framework, which is widely used for change detection in data streams. Experiments show that our text-scoring approach effectively captures warning signs in text compared to traditional machine learning classifies. Additionally, the application of the martingale framework highlights changes in online behavior and shows promise for detecting behavioral changes in at-risk individuals.

CONTENTS

1. INTRODUCTION

1.1 Motivation

According to the World Health Organization (WHO), it is estimated that 800,000 people worldwide die by suicide each year with at least as many suicide attempts. The grief felt in the aftermath of such an event is compounded by the fact that a suicide may be prevented. This reality of suicide has motivated WHO member states to commit themselves to reducing the rate of suicide by a significant percent by 2020.

To educate the public, the American Foundation for Suicide Prevention (AFSP) has identified characteristics or conditions that may increase an individual's risk. The three major risk factors are: health factors, environmental factors and historical factors. Identifying these risk factors is the first step in suicide prevention.

Here we address the challenge of real-time analysis of Twitter posts and the detection of suicide-related behavior. To process the stream of an individual's online content, we implement a martingale framework, which is widely used for the detection of changes in data stream .The main contributions of this project are twofold. First, using research from the field of psychology, we design and develop behavioral features to quantify the level second, we monitor the stream of an individual Twitter user and his behavioral features of risk for an individual according to his online behavior on Twitter. In particular, we create a feature for text analysis called the Suicide Prevention Assistant (SPA) text score. using an innovative application of a martingale framework to detect sudden behavioral changes.

1.2 Problem Definition

Sudden behavioral change is one of the most important suicide warning signs. As reported by the AFSP, a person's suicide risk is greater if a behavior is new or has increased, especially if it is related to a painful event, loss, or change. Considering this in conjunction with social media, where users constantly publish messages and deliberately express their feelings, we address suicide warning sign detection as a real-time data stream mining problem. Given a series of observations over time (tweets, messages, or blog posts), the task is to detect an abrupt change in a user behavior that may be considered as suicide warning sign. In the field of data stream mining, this can be specifically seen as change point detection problem.

As highlighted by warning signs signify increased imminent risk for suicide (i.e., within minutes, hours, or days). According to the APA suicide warning signs may include talking about dying, significant recent loss (death, divorce, separation, broken relationship), change in personality, fear of losing control, suicide plan, suicidal thoughts, or no hope for the future. As discussed in the following, recent research has shown the emergence of such signs on social networking sites.

A general framework for detecting suicide-related posts in social networks In this section, we present the proposed framework for the analysis and real-time detection of suicide-related posts on Twitter. First, we introduce the real-time detection problem. Then we define our online proxy measurements (behavior features) for suicide warning signs. Finally, we describe the approach we implement for detecting behavioral change points.

Most of the research at the intersection of behavioral health disorders and social media has focused on depression detection in online communities, specifically Major Depressive Episodes (MDE). However, the risk factors for suicide defined by the APA go far beyond depression alone. It is important to remember that depression does not necessarily imply suicidal ideation. Rather, suicide should be thought of as a potential end symptom of depression.

While mental health issues such as depression, suicidal ideation, and self-mutilation are defined medically as separate illnesses with overlapping symptoms, the approaches proposed to detect them online can be quite similar. Where the approaches vary is in the data they are treating, i.e. Facebook posts, Twitter tweets, Reddit forum threads, etc. and the specific event they are attempting to predict. In Moreno et al. first demonstrated that social networking sites could be a potential avenue for identifying students suffering from depression. The prevalence rates found for depression disclosed on Facebook corresponded to previous works in which such information was self-reported. On a larger scale, Jashinsky showed correlation between Twitter-derived and actual United States per-state suicide data. Together, these works established the presence of depression disclosure in online communities and opened up a new avenue for mental health research.DeChoudhury etal.explored the potential to use social media to detect and predict major depressive episodes in Twitter users. Using crowd-sourcing techniques, the authors built a cohort of Twitter users scoring high for depression on the CES-D (Center for Epidemiologic Studies Depression Scale) scale and other users scoring low. Studying these two

classes, they found that what is known from traditional literature on depressive behavior also translates to social media.

For example, users with a high CES-D score posted more frequently late at night, interacted less with their online friends, and had a higher use of first-person pronouns. Additionally, online linguistic patterns match previous findings regarding language use of depressed individuals. More recently, De Choudhury et al. have shown that linguistic features are important predictors in identifying individuals transitioning from mental discourse on social media to suicidal ideation. The authors showed a number of markers characterizing these shifts including social engagement, manifestation of hopelessness, anxiety and impulsiveness based on a small subset of Reddit posts.Coppersmith et al.examined the data published by Twitter users prior to a suicide attempt and provided an empirical analysis of the language and emotions expressed around their attempt.

One of the interesting results found in this study is the increase in the percentage of tweets expressing sadness in the weeks prior to a suicide attempt, which is then followed by a noticeable increase in anger and sadness emotions the week following a suicide attempt. In the same line of research, O'Dea et al. confirmed that Twitter is used by individuals to express suicidality and demonstrated that it is possible to distinguish the level of concern among suicide-related tweets, using both human coders and an automatic machine classifier. These insights have also been investigated by Braithwaite et al. who demonstrated that machine learning algorithms are efficient in differentiating people who are at a suicidal risk from those who are not. For a more detailed review of the use of social media platforms as a tool for suicide prevention, the reader may refer to the recent systematic survey by Robinson et al. These works have shown that individuals disclose their depression and other struggles to online communities, which indicates that social media networks can be used as a new arena for studying mental health. Despite the solid foundation, the current literature is missing potential key factors in the effort to detect depression and predict suicide. Currently, few works analyze the evolution of an individual's online behavior. Rather, the analysis is static and may take into consideration one post or tweet at a time while ignoring the whole context. Additionally, an individual's online speech is often compared to other individuals and not to their own linguistic style. This is a disadvantage because two individuals suffering the same severity of depression may express themselves very differently online.

Each tweet was reviewed by at least two annotators, with a subset of 55 tweets being validated by the psychologist. The annotated data had a Cohen's kappa statistic of 69.1%, which is considered as a substantial agreement among the annotators. The weighted kappa statistic, which takes into account different levels of disagreement, was around 71.5%. Finally, we used the Fleiss kappa to measure agreement for the 55 tweets with three annotators, which gives 78.3%. Overall, the three statistics show strong agreement between the annotators.

In addition to the cross-sectional dataset, we gathered Twitter history from a collection of users. This dataset, which may be considered as a longitudinal data, was collected from 10 unique real users that demonstrated a serious change in speech or online behavior in their Twitter accounts. Two initial validation cases, which ended with the individual committing suicide, were previously identified. We were able to identify a third fatal case and an additional seven other cases where the individuals demonstrated an abrupt change in behavior. These additional cases were manually chosen and the abrupt change in behavioral was principally judged by a change in speech.

However, unlike retrospective detection settings, which focus on batch processing, here we are interested in the setting where the data arrives as a stream in real time. To address this challenge, we chose an approach employing a martingale framework for change point detection. This algorithm has been successfully applied to detecting changes in unlabeled data streams, video-shot change detection, and more recently, in the detection of news events in social networks. To the best of our knowledge, this is the first attempt to apply the martingale framework on a multi-dimensional data stream generated by Twitter users.

1.3 Limitations of Project

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out.

Limitations are given by the accuracy of the methods applied to infer some of these aspects based on the vocabulary and behavioral patterns of the users. For instance, the weekend tweets count ratio is because of the difference in the posting time according to the user’s time zone and the sleeping time tweets ratio, for which we assume that the sleeping time is the period in a day when the user has

less activity. Finally, there are also limitations given by the nature of the users who post on Twitter, as they might differ from users at risk who do not choose to make their profiles public or even from users at risk who do not have a Twitter account. In addition, it is not guaranteed that the users annotated as users at risk are actually at risk because the annotation was performed just from reading a few tweets.

This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

- **ECONOMICAL FEASIBILITY**

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

- **TECHNICAL FEASIBILITY**

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

- **SOCIAL FEASIBILITY**

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

2. LITERATURE SURVEY

2.1. Introduction

A literature survey or a literature review in a project report is that section which shows the various analyses and research made in the field of your interest and the results already published, taking into account the various parameters of the project and the extent of the project.

In an effort to educate the public, the American Foundation for Suicide Prevention (AFSP) has identified characteristics or conditions that may increase an individual's risk. The three major risk factors are: 1) health factors (e.g. mental health, chronic pain),

2) environmental factors (e.g. harassment, stressful life events), and

3) historical factors (e.g. previous suicide attempts, family history).

Additionally, the time period preceding a suicide can hold clues to an individual's struggle. The AFSP categorizes these warning signs as follows: 1) talk (e.g. mentioning being a burden or having no reason to live), 2) behavior (e.g. withdrawing from activities, sleeping too much or too little), and 3) mood (e.g. depression, rage).

A social networking service (also social networking site, or SNS or social media) is an online platform which people use to build social networks with other people who share similar personal or career interests, activities, backgrounds or real-life connections.

Most of the research at the intersection of behavioral health disorders and social media has focused on depression detection in online communities, specifically Major Depressive Episodes (MDE). However, the risk factors for suicide defined by the APA go far beyond depression alone. It is important to remember that depression does not necessarily imply suicidal ideation. Rather, suicide should be thought of as a potential end symptom of depression.

While mental health issues such as depression, suicidal ideation, and self-mutilation are defined medically as separate illnesses with overlapping symptoms, the approaches proposed to detect them online can be quite similar. Where the approaches vary is in the data they are treating, i.e. Facebook posts, Twitter tweets, Reddit forum threads, etc. and the specific event they are attempting to predict. In

Moreno et al. first demonstrated that social networking sites could be a potential avenue for identifying students suffering from depression. The prevalence rates found for depression disclosed on Facebook corresponded to previous works in which such information was self-reported. On a larger scale, Jashinsky et al. showed correlation between Twitter-derived and actual United States per-state suicide data. Together, these works established the presence of depression disclosure in online communities and opened up a new avenue for mental health research.

[1] How stigma interferes with mental health care by P. Corrigan:

Many people who would benefit from mental health services opt not to pursue them or fail to fully participate once they have begun. One of the reasons for this disconnect is stigma; namely, to avoid the label of mental illness and the harm it brings, people decide not to seek or fully participate in care. Stigma yields 2 kinds of harm that may impede treatment participation: It diminishes self-esteem and robs people of social opportunities. Given the existing literature in this area, recommendations are reviewed for ongoing research that will more comprehensively expand understanding of the stigma-care seeking link. Implications for the development of anti stigma programs that might promote care seeking and participation are also reviewed.

Identifying these risk factors is the first step in suicide prevention. However, the social stigma surrounding mental illnesses means that at-risk individuals may avoid professional assistance. In fact, they may be more willing to turn to less formal resources for support. Recently, online social media networks have become one such informal resource. Research has shown that at-risk individuals are turning to contemporary technologies (forums, micro-blogs) to express their deepest struggles without having to face someone directly. As a result, suicide risk factors and warning signs have been seen in a new arena. There are even instances of suicide victims writing their final thoughts on Twitter, Facebook, and other online communities.

[2]Young people's help seeking for mental health problems by D. Rickwood, F. P. Deane, C. J. Wilson, et al:

This summarizes an ambitious research agenda aiming to uncover the factors that affect help-seeking among young people for mental health problems. The research set out to consider why young people, and particularly young males, do not seek help when they are in psychological distress or

suicidal; how professional services be made more accessible and attractive to young people; the factors that inhibit and facilitate help-seeking; and how community gatekeepers can support young people to access services to help with personal and emotional problems. A range of studies was undertaken in New South Wales, Queensland and the ACT, using both qualitative and quantitative approaches. Data from a total of 2721 young people aged 14–24 years were gathered, as well as information from some of the community gatekeepers to young people's mental health care.

Help-seeking was measured in all the studies using the General Help Seeking Questionnaire (Wilson, Deane, Ciarrochi & Rickwood, 2005), which measures future help-seeking intentions and, through supplementary questions, can also assess prior help-seeking experience. Many of the studies also measured recent help-seeking behavior using the Actual Help Seeking Questionnaire. The types of mental health problems examined varied across the studies and included depressive symptoms, personal-emotional problems, and suicidal thoughts.

The results are reported in terms of: patterns of help-seeking across adolescence and young adulthood; the relationship of help-seeking intentions to behavior; barriers to seeking help—lack of emotional competence, the help-negation effect related to suicidal thoughts, negative attitudes and beliefs about help-seeking and fear of stigma; and facilitators of seeking help—emotional competence, positive past experience, mental health literacy, and supportive social influences. The paper considers the implications of the findings for the development of interventions to encourage young people to seek help for their mental health problems, and concludes by identifying gaps in the help-seeking research and literature and suggesting future directions.

[3]Feeling bad on Facebook: depression disclosures by college students on a social networking site by M. Moreno, L. Jelenchick, K. Egan, et al:

Depression is common and frequently undiagnosed among college students. Social networking sites are popular among college students and can include displayed depression references. The purpose of this study was to evaluate college students' Facebook disclosures that met DSM criteria for a depression symptom or a major depressive episode (MDE).

We selected public Facebook profiles from sophomore and junior undergraduates and evaluated personally written text: "status updates." We applied DSM criteria to one year of status updates from

each profile to determine prevalence of displayed depression symptoms and MDE criteria. Negative binomial regression analysis was used to model the association between depression disclosures and demographics or Facebook use characteristics.

The results are a total of 200 profiles were evaluated, profile owners were 43.5% female with a mean age of 20 years. Overall, 25% of profiles displayed depressive symptoms and 2.5% met criteria for MDE. Profile owners were more likely to reference depression if they averaged at least one online response from their friends to a status update disclosing depressive symptoms (exp (B)=2.1, $p<0.001$), or if they used Facebook more frequently ($p<0.001$).

[6]Predicting depression via social media by Predicting depression via social media:

Major depression constitutes a serious challenge in personal and public health. Tens of millions of people each year suffer from depression and only a fraction receives adequate treatment. We explore the potential to use social media to detect and diagnose major depressive disorder in individuals. We first employ crowd sourcing to compile a set of Twitter users who report being diagnosed with clinical depression, based on a standard psychometric instrument. Through their social media postings over a year preceding the onset of depression, we measure behavioral attributes relating to social engagement, emotion, language and linguistic styles, ego network, and mentions of antidepressant medications. We leverage these behavioral cues, to build a statistical classifier that provides estimates of the risk of depression, before the reported onset. We find that social media contains useful signals for characterizing the onset of depression in individuals, as measured through decrease in social activity, raised negative effect, highly clustered ego networks, heightened relational and medicinal concerns, and greater expression of religious involvement. We believe our findings and methods may be useful in developing tools for identifying the onset of major depression, for use by healthcare agencies; or on behalf of individuals, enabling those suffering from depression to be more proactive about their mental health.

[20] Social media and suicide prevention: A systematic review by .J. Robinson, G. Cox, E. Bailey, et al:

Social media platforms are commonly used for the expression of suicidal thoughts and feelings, particularly by young people. Despite this, little is known about the ways in which social media can be used for suicide prevention. The aim of this study was to conduct a systematic review to identify current

evidence pertaining to the ways in which social media are currently used as a tool for suicide prevention. Medline, PsycInfo, Embase, CINHAL and the Cochrane Library were searched for articles published between 1991 and April 2014. English language articles with a focus on suicide-related behavior and social media were included. No exclusion was placed on study design. Thirty studies were included; 4 described the development of social media sites designed for suicide prevention, 6 examined the potential of social media in terms of its ability to reach or identify people at risk of suicide, 15 examined the ways in which people used social media for suicide prevention-related purposes, and 5 examined the experiences of people who had used social media sites for suicide prevention purposes. No intervention studies were identified. Social media platforms can reach large numbers of otherwise hard-to-engage individuals, may allow others to intervene following an expression of suicidal ideation online, and provide an anonymous, accessible and non-judgmental forum for sharing experiences. Challenges include difficulties controlling user behavior and accurately assessing risk, issues relating to privacy and confidentiality and the possibility of contagion. Social media appears to hold significant potential for suicide prevention; however, additional research into its safety and efficacy is required.

Suicide warning signs in online behavior :

To identify online behaviors that may reflect the mental state of a Twitter user, we established two groups of behavioral features: user-centric and post-centric features. User-centric features characterize the behavior of the user in the Twitter community while post-centric features are characteristics that are extracted from the properties of a tweet. These features have been shown to successfully aid in determining the mental health of a user. Table 1 shows a detailed description of the features we selected. The American Association of Suicidology (AAS) identifies withdrawing from friends, family, or society as one of the warning signs of suicide. With the user-centric behavioral features, we aim to capture changes in a Twitter user's engagement with other users. The friends and followers features can quantify an individual's interaction with their online community, such as a sudden decrease in communication. Contrarily, they can also reflect an expansion of an individual's online community. This is relevant, as at-risk individuals have also been shown to increase their time online developing personal relationships. It is important to note that we have chosen the terms friends and followers to represent the unidirectional relationships that are inherent on Twitter. We acknowledge that this term may not apply for certain user accounts such as celebrities and news outlets.

Additional features include volume, replies, re tweets, and links, which were all identified by as markers for mental health. These measures can help to quantify the number of interactions a user has with their friends and followers for it could be the case that an individual's social network remains stable while their interactions increase or decrease. The final user-centric feature, questions, may also indicate a user's attempt to engage with others online.

Post-centric behavioral features are characteristics originating from the post itself. One important piece of information is the hour at which the tweet is published (time feature). Late night activity can be an indication of unusual rhythms in sleep (insomnia and hypersomnia) and can predict future episodes of depression. In addition to the time feature, we address the text of the post (text score), which holds the most vital information pertaining to an individual's current mood and mental health.

To classify the text of the post, we propose two different approaches. The first approach is a natural language processing (NLP) method that combines features generated from the text based on an ensemble of lexicons. These lexicons are composed of linguistic themes commonly exhibited by at-risk individuals. The second approach, called the distress classifier, is based on machine learning. Although machine learning is commonly used to classify text, the supervised algorithms require annotated datasets, which may be costly in terms of time and potential annotator error. Additionally, traditional machine learning methods are difficult to apply in this context because of the nature of depression and distress in general. Two individuals suffering from depression may not express their symptoms in the same way, which translates to texts of the same severity having vastly different content. This means it is difficult for the algorithm to find the concept mapping between the textual features and the level of depression/distress.

We believe that this large amount of data on people’s feelings and behaviors can be used successfully for early detection of behavioral changes in at-risk individuals and may even help prevent deaths. Social computing research has focused on this topic in recent years. However, few initiatives have been concerned with the real time detection of suicidal ideation on Twitter. Previously proposed detection methods rely heavily on manually annotated speech, which can limit their effectiveness due in part to the varying forms of suicide warning signs in at-risk individuals. Many of these methods also focus on

the messages published by individuals at a specific time independently of the whole context, which may be represented by the sequence of publications over time.

Social Networks & Psychological Features:

These consist of a group of features based on generic lexicons, statistics measured from the users' writings, information of interest for clinicians regarding the behavior of users in time, the users' social network (relational features), and n-grams lexicons, which include terms referring to suicidal ideation or suicide risk factors (we referred to these features as suicide-related lexicon features). Each of these types of features is described in the next subsections.

Behavioral Features:

These features are based on the information extracted from the metadata of tweets. Here, we measured the behavior of users based on their activity within certain periods, which are defined at different granularity levels.

Tweets statistics:

This group refers to 2 types of features that correspond to statistical measures calculated from the tweets of users. We considered elements such as the number of tweets created and their length and the number of tweets that were retained for each user at the SPV in relation to the total number of tweets posted.

Relational Features:

These are informative features regarding the relationships and interactions between users. Elements such as the count of retweets and favorites received and given by the users can provide insight on the social support they receive, along with information regarding the number of followers and users followed, as previously considered for depression screening.

Image-Based Feature:

The methodology proposed in the study by Rodriguez et al, where a method for inferring the personality under the OCEAN model was presented. In this sense, we created a classifier trained on images extracted from Instagram using a subset of the phrases and keywords used in the data collection process for Twitter. These images were considered suicide related, whereas a set of unrelated images was

considered as our control cases. Afterward, this first model was applied to each of the images extracted from the users' tweets of our dataset. To create this model, we used 90,000 images for training and 60,000 images for validation. To obtain a single score per user (images user score), the average of the individual scores of the images of each user was considered as the user's aggregated score.

2.2 Existing System

- In the existing system, the system implemented to understand the connectivity and communication characteristics of Twitter users who post content subsequently classified by human annotators as containing possible suicidal intent or thinking, commonly referred to as suicidal ideation.
- The system achieves this understanding by analyzing the characteristics of their social networks. Starting from a set of human annotated Tweets we retrieved the authors' followers and friends lists, and identified users who re-tweeted the suicidal content. We subsequently built the social network graphs.
- An existing system results show a high degree of reciprocal connectivity between the authors of suicidal content when compared to other studies of Twitter users, suggesting a tightly-coupled virtual community. In addition, an analysis of the re-tweet graph has identified bridge nodes and hub nodes connecting users posting suicidal ideation with users who were not, thus suggesting a potential for information cascade and risk of a possible contagion effect. This is particularly emphasized by considering the combined graph merging friendship and re-tweeting links.
- In this project we address the challenge of real-time analysis of Twitter posts and the detection of suicide-related behavior. To process the stream of an individual's online content we implement a martingale framework, which is widely used for the detection of changes in data stream settings. The input into this framework is a series of behavioral features computed from each individual Twitter post (tweet). These features are compared to previously seen behavior in order to detect a sudden change in emotion that may indicate an elevated risk of suicide.

2.3 Disadvantages of Existing System.

- It is not based on a natural language processing (NLP) based approach.
- There is no technique for emotion change detection.

2.4 Proposed System

- In the proposed system, the system addresses the challenge of real-time analysis of Twitter posts and the detection of suicide-related behavior. To process the stream of an individual's online content, we implement a martingale framework, which is widely used for the detection of changes in data stream settings.
- The input into this framework is a series of behavioral features computed from each individual Twitter post (tweet). These features are compared to previously seen behavior, in order to detect a sudden change in emotion that may indicate an elevated risk of suicide.
- The AAS identifies withdrawing from friends, family, or society as one of the warning signs of suicide. With the user-centric behavioral features, we aim to capture changes in a Twitter user's engagement with other users. The friends and followers features can quantify an individual's interaction with his or her online community, such as a sudden decrease in communication. On the other hand, they can also reflect an expansion of an individual's online community. This is relevant, as at-risk individuals have also been shown to increase their time online developing personal relationships. It is important to note that we have chosen the terms friends and followers to represent the unidirectional relationships that are inherent on Twitter. We acknowledge that this term may not apply for certain user accounts such as celebrities and news outlets.

Advantages

- Very fast Suicide warning signs in online behavior.
- The system is effective due to a general framework for detecting suicide-related posts in social networks.

3. ANALYSIS

3.1 Introduction:

Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, and is used in different business, science, and social science domains. In today's business world, data analysis plays a role in making decisions more scientific and helping businesses operate more effectively.

Mental disorders are a serious health issue worldwide. According to the mortality data presented by the World Health Organization, the number of deaths because of suicide is equivalent to a person dying every 40 seconds. Considering that the signs and symptoms of these disorders have been proven to be traceable on social media, scientists have started to work on the development of automated methods to detect signs and symptoms of these conditions by addressing the importance of early detection.

The methods used for analyzing user-generated data related to suicide focus on tracking social networks at a post level, that is, a tweet or at a user level, that is, a sample of their tweets or posts. The latter is more related to risk assessment, as more data from a single user can be explored. In terms of the type of information extracted and explored, state-of-the-art approaches perform an exhaustive analysis of textual information contained in posts. This has been proven to be relevant for screening and risk assessment tasks. The methods applied often consider bag of words (BoW) models, topic models, lexicons, sentiment analysis tools, and readability and syntactical analysis features. The most recent work with deep learning approaches consists of exploring sequence models encoding vector representations of terms known as word embeddings.

Regarding the use of visual information from posts shared in social networks to address mental disorders, the closest approaches to our goal are methods for personality prediction, and a few recent approaches address self-harm, depression, and anxiety. We believe that our study is the first image-based approach for suicide risk assessment on social media at the user level.

Cleaning:

Cleaning refers to the process of removing invalid data points from a dataset.

Transforming:

Data transforming is the process of converting data or information from one format to another, usually from the format of a source system into the required format of a new destination system.

Modeling:

Data modeling is a process used to define and analyze data requirements needed to support the business processes within the scope of corresponding information systems in organizations.

3.2 Software Requirement Specifications:

A **software requirements specification** (SRS) is a detailed description of a software system to be developed with its functional and non-functional requirements. The SRS is developed based the agreement between customer and contractors. It may include the use cases of how user is going to interact with software system. The software requirement specification document consistent of all necessary requirements required for project development.

A good SRS defines the how Software System will interact with all internal modules, hardware, communication with other programs and human user interactions with wide range of real life scenarios. Using the Software requirements specification (SRS) document on QA lead, managers creates test plan. It is very important that testers must be cleared with every detail specified in this document in order to avoid faults in test cases and its expected results.

3.2.1. Software Requirements

- Operating System - Windows XP or Later Versions
- Coding Language - Java/J2EE (JSP,Servlet)
- Front End - J2EE
- Back End - MySQL

3.2.2. Hardware Requirements

- Processor - Pentium–IV or later version
- RAM - 4 GB(min)
- Hard Disk - 40 GB
- Keyboard - Standard Windows Keyboard

- Mouse - Two or Three Button Mouse
- Monitor - SVGA

3.3. Content Diagram of Project

The Content diagram is an extension of UML notation. The purpose of the Content diagram is to generate or represent a project structure (diagrams) and relations between them. The Content table on works as a table of contents for a project. The Content Shape creates a table of contents of all diagrams of the project.

A context diagram is a visual representation of the relationship between data and business processes. This diagram has 3 main components which include **external entities, system processes, and data flows**. It provides the factors and events you need to consider when developing a system.

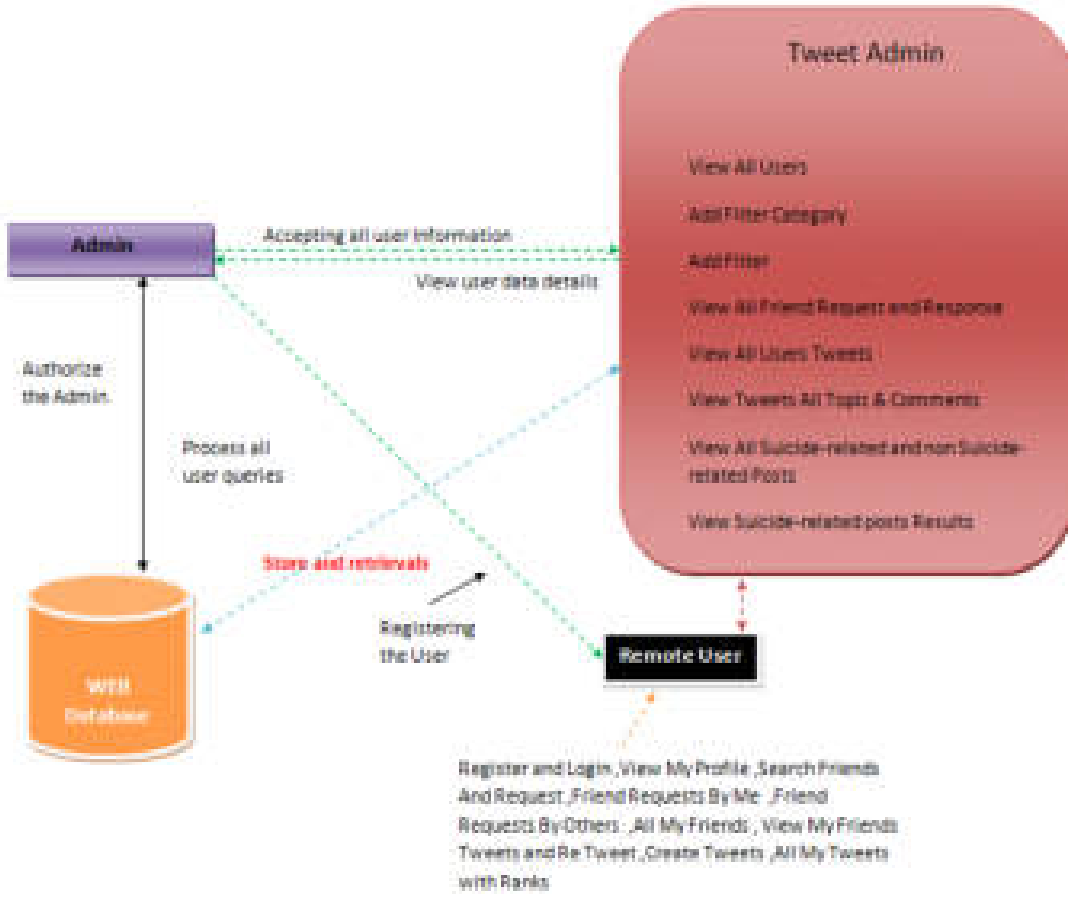


Fig.no:3.3 Content diagram of project

3.4. Algorithm

- Point wise mutual information $PMI(w, c) = \log (P(w, c) / P(w)P(c))$
- Suicide Prevention Assistant (SPA)= $f_{\text{symptoms}} + f_{\text{swear}} + f_{\text{intensifiers}} + f_{\text{first\_pronouns}}$.
- Unified strangeness measure (USM) is defined as:

$$USM_i(X, x_n) = \sum_{k=1}^{m} \left| x_{ik} - \mu\left(\{x_{1k}, \ldots, x_{(n-1)k}\} \cup \{x_{nk}\}\right) \right|$$

- In the second step of the martingale framework

We rank the USM of the new point with respect to the USM of the previously observed points using a statistic. This statistic is denoted as the $\hat{p}$- value for each instance xi. Formally, the $\hat{p}$-value of xi for i: 1…n can be calculated as follow:

$$\hat{p}_i([x_1, \ldots, x_n], \theta_i) = \frac{\#\{j : s_j > s_i\} + \theta_i \#\{j : s_j = s_i\}}{i}$$

- Randomized power martingale

$$M_n^{(\varepsilon)} = \prod_{i=1}^{n} (\varepsilon \hat{p}_i^{\varepsilon - 1})$$

4. DESIGN

4.1. Introduction

A model is a simplification of reality. A model provides the blue prints of a system. A model may be structural, emphasizing the organization of a system or it may be behavioral, emphasizing the dynamics of the system. Our system design follows the architecture of many modern data analytic.

4.2. UML Diagrams

Unified Modeling language (UML) is a standardized modeling language enabling developers to specify, visualize, construct and document artifacts of a software system. Thus, UML makes these artifacts scalable, secure and robust in execution. UML is an important aspect involved in object-oriented software development. It uses graphic notation to create visual models of software system.

UML diagrams represent static and dynamic views of a system model. The static view includes class diagrams and composite structure diagrams, which emphasize static structure of systems using objects, attributes, operations and relations. The dynamic view represents collaboration among objects and changes to internal states of objects through sequence, activity and state machine diagrams. A wide variety of UML modeling tools are available to simplify the modeling process, including IBM Rational Rose, Rational Rhapsody, Magic Draw UML, Star UML, Argo UML, BOUML, Power Designer and Dia.

The UML represents a collection of best engineering practices that proven successfully in the modeling of large and complex systems.

The UML is a very important part of developing objects oriented software and software development process. The UML uses mostly graphical to express the design of software projects.

4.2.1. Use case Diagram

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

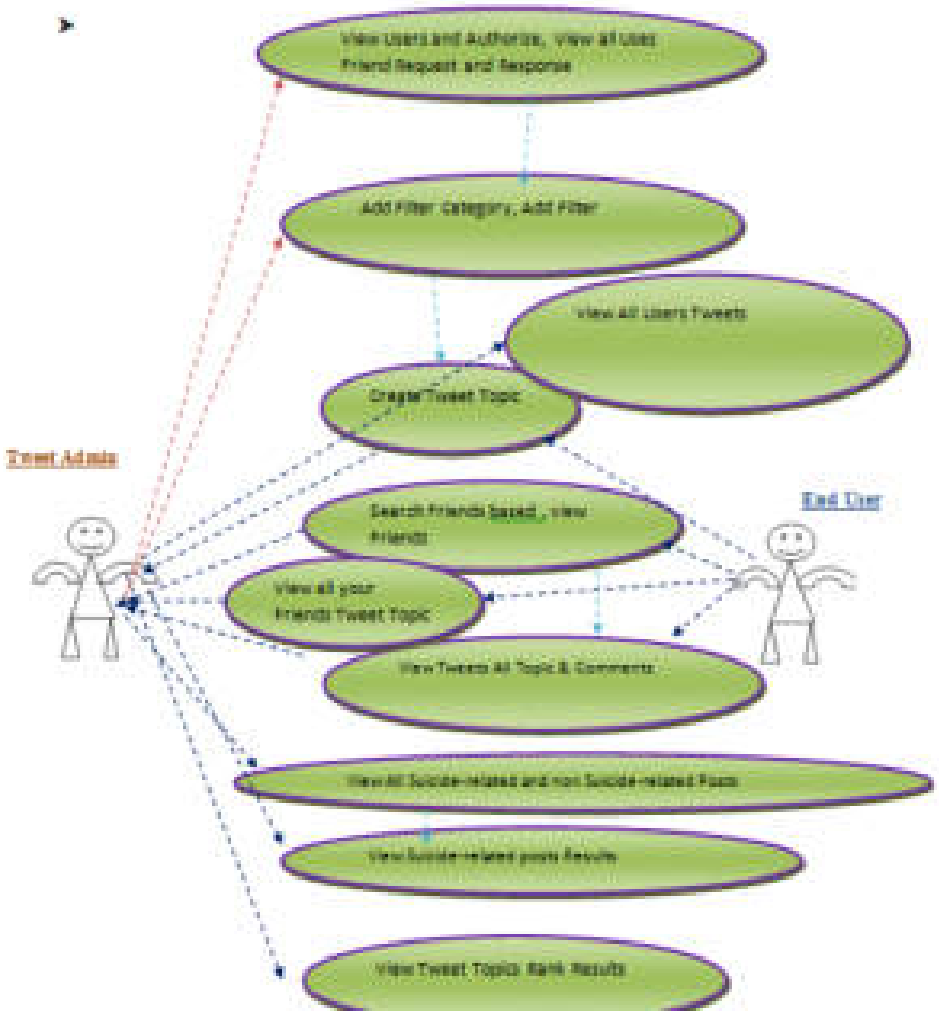


Fig.no:4.2.1: Use case diagram

4.2.2. Sequence Diagram

A sequence diagram is a Unified Modeling Language (UML) is a sort of cooperation chart that shows how forms work with each other and in what order. It is a develop of a Message Sequence Chart. Arrangement outliners are at times called occasion graphs, occasion situations, event scenarios and timing diagrams

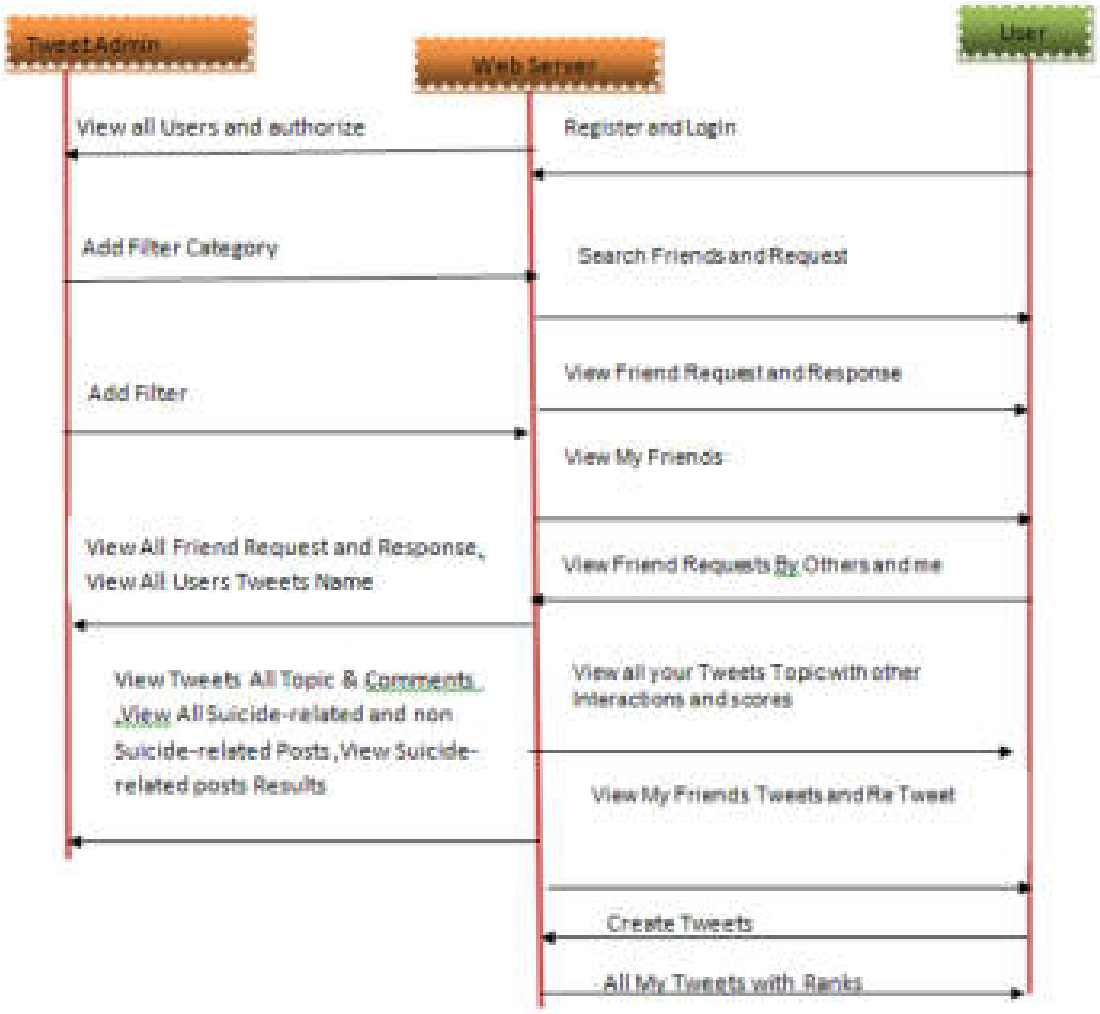


Fig.no:4.2.2. Sequence Diagram

4.2.3. Class Diagram

The class diagram is the main building block of object-oriented modeling. It is used for general conceptual modeling of the systematic of the application, and for detailed modeling translating the models into programming code. Class diagrams can also be used for data modeling. The classes in a class diagram represent both the main elements, interactions in the application, and the classes to be programmed.

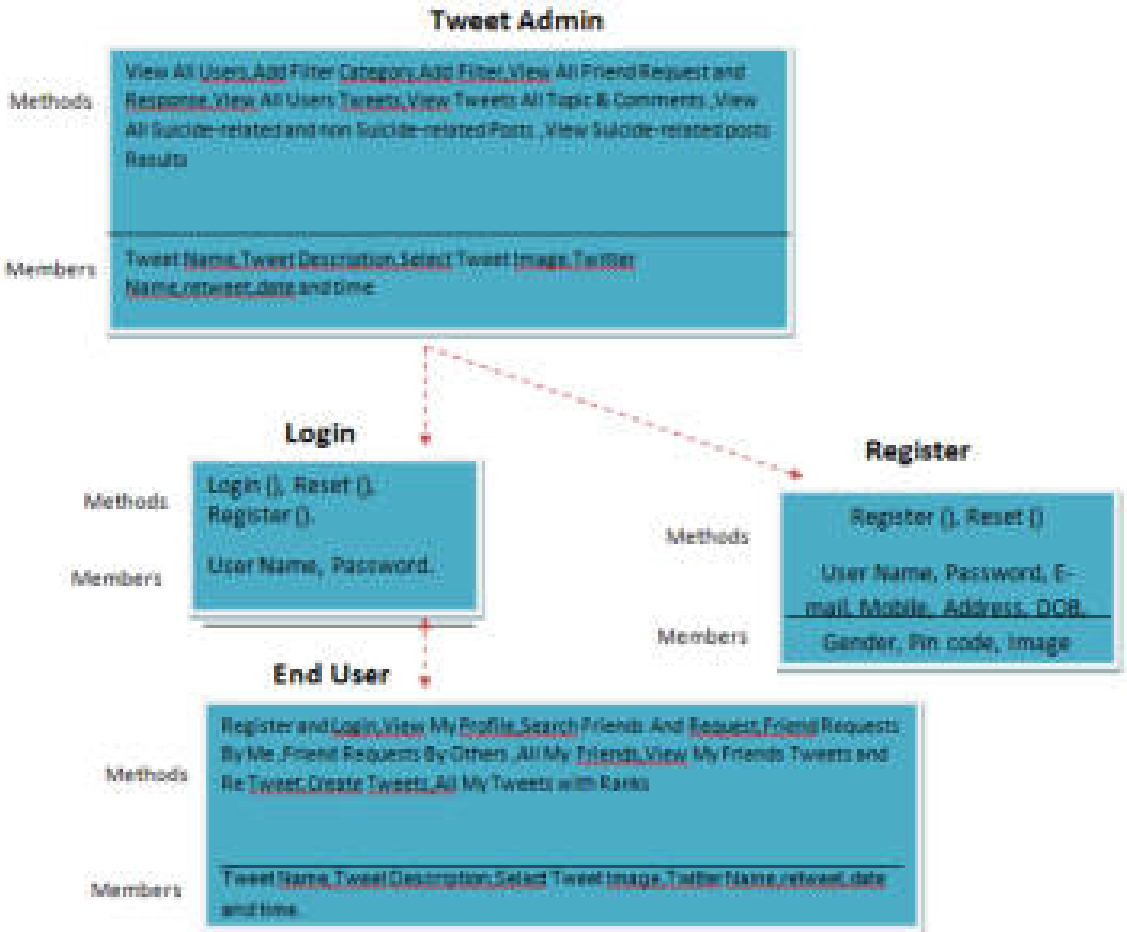


Fig no:4.2.3: Class Diagram

4.2.4. Dataflow Diagram

A data flow diagram (DFD) maps out the flow of information for any process or system. It uses defined symbols like rectangles, circles and arrows, plus short text labels, to show data inputs, outputs, storage points and the routes between each destination. Data flowcharts can range from simple, even hand-drawn process overviews, to in-depth, multi-level DFDs that dig progressively deeper into how the data is handled. They can be used to analyze an existing system or model a new one.

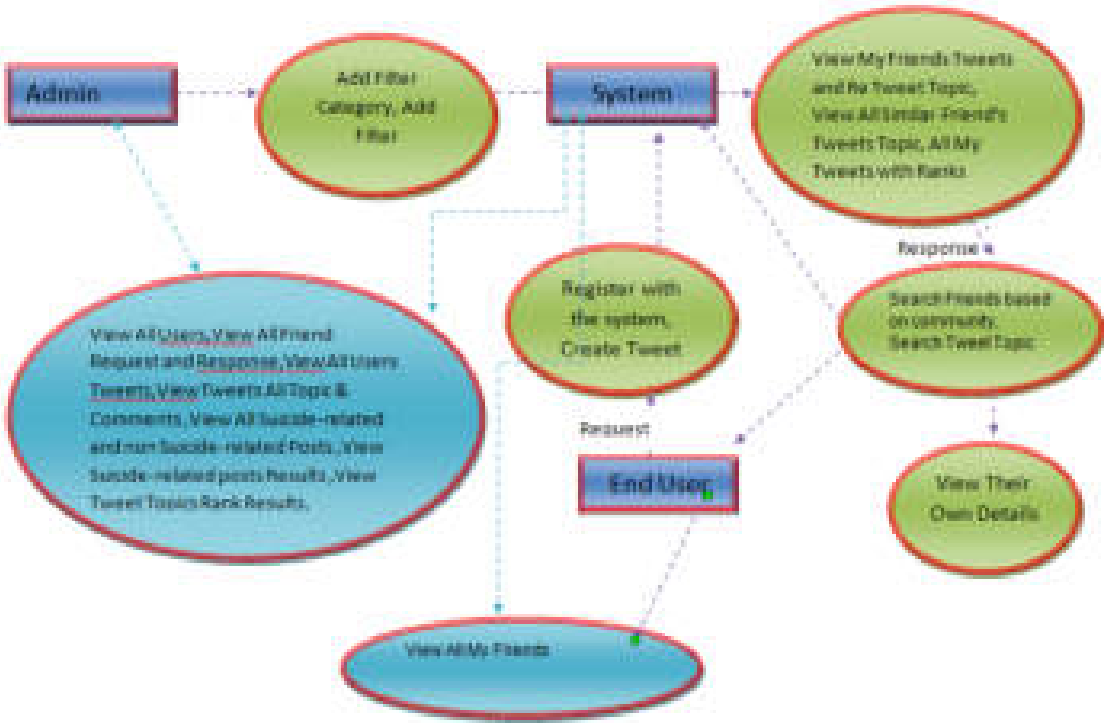


Fig no: 4.2.4:Data flow Diagram

4.2.5. Flowchart Diagram

A flowchart is a diagram that depicts a process, system or computer algorithm. They are widely used in multiple fields to document, study, plan, improve and communicate often complex processes in clear, easy-to-understand diagrams. Flowcharts, sometimes spelled as flow charts, use rectangles, ovals, diamonds and potentially numerous other shapes to define the type of step, along with connecting arrows to define flow and sequence.

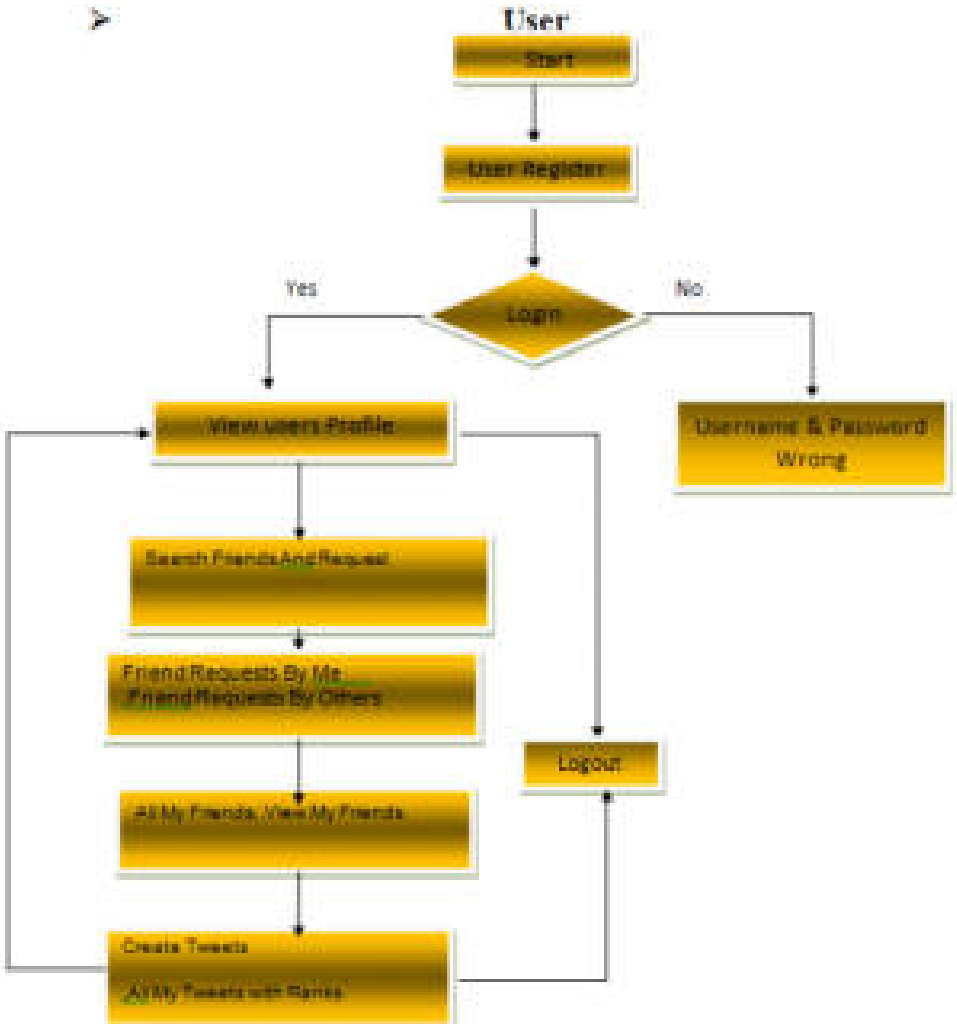


Fig no: 4.2.5: Flow chart Diagram

4.3. Module Design and Organization

Modules:

- Admin
- user

- **Tweet Admin**

In this module, the Admin has to login by using valid user name and password. After login successful he can perform some operations such as View All Users, Add Filter Category, Add Filter, View All Friend Request and Response, View All Users Tweets, View Tweets All Topic & Comments, View All Suicide-related and non Suicide-related Posts, View Suicide-related posts Results, View Tweet Topics Rank Results

Friend Request & Response

In this module, the admin can view all the friend requests and responses. Here all the requests and responses will be displayed with their tags such as Id, requested user photo, requested user name, user name request to, status and time & date. If the user accepts the request then the status will be changed to accepted or else the status will remains as waiting.

- **User**

In this module, there are n numbers of users are present. User should register before performing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user can perform some operations like View My Profile, Search Friends And Request, Friend Requests By Me, Friend Requests By Others, All My Friends, View My Friends Tweets and Re Tweet, Create Tweets, All My Tweets with Ranks.

Searching Users to make friends

In this module, the user searches for users in Same Network and in the Networks and sends friend requests to them. The user can search for users in other Networks to make friends only if they have permission.

5. IMPLEMENTATION & RESULTS

5.1. Introduction

Modules:

- **Admin**

Admin module allows system administrator to set up back-end of the system and perform basic system configuration, mainly definition of predefined drop-down fields, definition of classes time schedule, etc.

- **User**

The user module allows users to register, log in, and log out. Each user is assigned one or more roles. By default there are three roles: anonymous (a user who has not logged in) and authenticated (a user who is registered), and administrator (a signed in user who will be assigned site administrator permissions)

5.2. Explanation of Key Functions

A function key is a key on a computer or terminal keyboard which can be programmed so as to cause an operating system command interpreter or application program to perform certain actions, a form of soft key. On some keyboards/computers, function keys may have default actions, accessible on power-on.

5.2.1 Java Technology:

Java is both a programming language and computing platform. There are lots of applications and websites that will not work unless you have Java installed, and more are created every day. Java is fast, secure, and reliable. Today, Java is needed to run various applications such as games, social media applications, audio and video applications, etc. It is a class based and an object-oriented language similar to C++, but with advanced and simplified features. Java is Concurrent where you can execute many statements instead of sequentially executing it. This language is free to access and can run on all

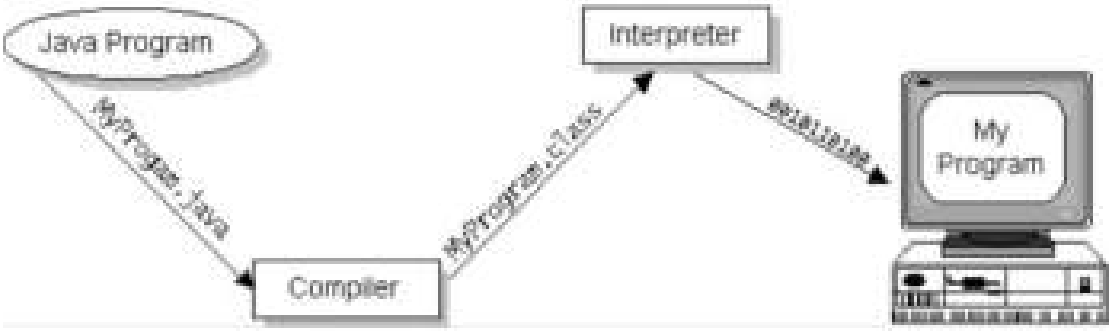
platforms. From laptops to datacenters, game consoles to scientific supercomputers, cell phones to the Internet, Java is everywhere!

5.2.2. The Java Programming Language:

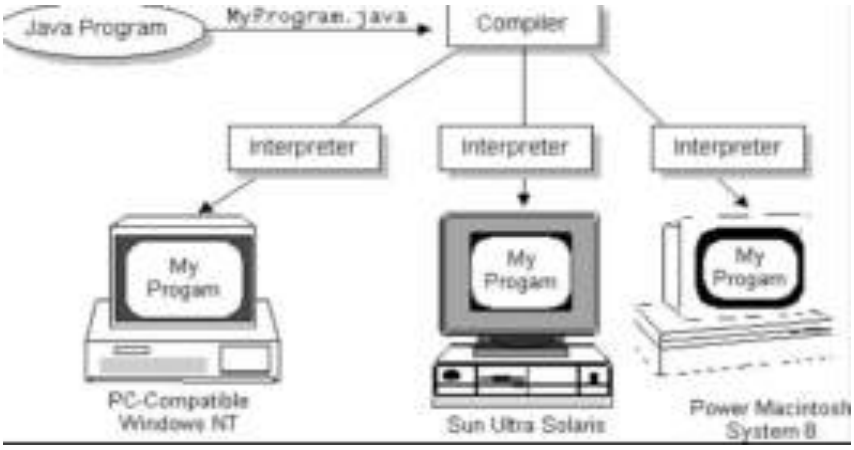
The java programming language is a high level language that can be characterized by all of the following buzz words.

- Simple
- Architectural neutral
- Object oriented
- Portable
- Distributed
- High performance
- Interpreted
- Multithreaded
- Dynamic
- Secured

With the most programming languages you either combine or interpret the program so that run it on your computer. The java programming language is unusual in that a program is both compiled and interpreted. With the compiler, first you translate a program into a intermediate language called java byte codes—the platform independent codes interpreted by the interpreter on the java platform. The interpreter parses and runs each java byte code instruction on the compiler. Compilation happens just once; interpretation occurs each time the program is executed. The following figure illustrates how this works



You can think of java byte codes as the machine code instructions for the java virtual machine (java VM). Every java interpreter, whether it's a development tool or a web browser that can run applets, is an implementation of the java VM. Java byte codes help make "bright ones run anywhere" possible. You can compile your program into byte codes that has a java compiler. The byte codes can then be run on any implementation of the java VM. That means that as long as a computer has a java VM , the same program written in the java programming language can run on windows 2000, a Solaris workstation, or on an i Mac.



5.2.3 Javascript:

JavaScript is a script-based programming language that was developed by Netscape Communication Corporation. JavaScript was originally called Live Script and renamed as JavaScript to indicate its relationship with Java. JavaScript supports the development of both client and server

components of Web-based applications. On the client side, it can be used to write programs that are executed by a Web browser within the context of a Web page. On the server side, it can be used to write Web server programs that can process information submitted by a Web browser and then updates the browser's display accordingly

Even though JavaScript supports both client and server Web programming, we prefer JavaScript at Client side programming since most of the browsers supports it. JavaScript is almost as easy to learn as HTML, and JavaScript statements can be included in HTML documents by enclosing the statements between a pair of scripting tags

<SCRIPTS>..</SCRIPT>.

<SCRIPT LANGUAGE = "JavaScript">

JavaScript statements

</SCRIPT>

Here are a few things we can do with Java Script:

- Validate the contents of a form and make calculations.
- Add scrolling or changing messages to the Browser's status line.
- Animate images or rotate images that change when we move the mouse over them.
- Detect the browser in use and display different content for different browsers.
- Detect installed plug-ins and notify the user if a plug-in is required.

We can do much more with JavaScript, including creating entire application.

5.2.4 Hyper Text Markup Language:

Hypertext Markup Language (HTML), the languages of the World Wide Web (WWW), allows users to produces Web pages that include text, graphics and pointer to other Web pages (Hyperlinks).

HTML is not a programming language but it is an application of ISO Standard 8879, SGML (Standard Generalized Markup Language), but specialized to hypertext and adapted to the Web. The idea behind Hypertext is that instead of reading text in rigid linear structure, we can easily jump from

one point to another point. We can navigate through the information based on our interest and preference. A markup language is simply a series of elements, each delimited with special characters that define how text or other items enclosed within the elements should be displayed. Hyperlinks are underlined or emphasized works that load to other documents or some portions of the same

HTML can be used to display any type of document on the host computer, which can be geographically at a different location. It is a versatile language and can be used on any platform or desktop.

HTML provides tags (special codes) to make the document look attractive. HTML tags are not case-sensitive. Using graphics, fonts, different sizes, color, etc., can enhance the presentation of the document. Anything that is not a tag is part of the document itself.

Basic HTML Tags:

<!-- -->	Specifies comments
<A>…….</A>	Creates hypertext links
<B>…….</B>	Formats text as bold
<BIG>…….</BIG>	Formats text in large font.
<BODY>…</BODY>	Contains all tags and text in the HTML document
<CENTER>...</CENTER>	Creates text
<DD>…</DD>	Definition of a term
<DL>...</DL>	Creates definition list
<FONT>…</FONT>	Formats text with a particular font
<FORM>...</FORM>	Encloses a fill-out form
<FRAME>...</FRAME>	Defines a particular frame in a set of frames
<H#>…</H#>	Creates headings of different levels

<HEAD>...</HEAD>	Contains tags that specify information about a document
<HR>...</HR>	Creates a horizontal rule
<HTML>…</HTML>	Contains all other HTML tags
<META>...</META>	Provides meta-information about a document
<SCRIPT>…</SCRIPT>	Contains client-side or server-side script
<TABLE>…</TABLE>	Creates a table
<TD>…</TD>	Indicates table data in a table
<TR>…</TR>	Designates a table row
<TH>…</TH>	Creates a heading in a table

5.2.5 Java Server Pages (JSP):

Java server Pages is a simple, yet powerful technology for creating and maintaining dynamic-content web pages. Based on the Java programming language, Java Server Pages offers proven portability, open standards, and a mature re-usable component model .The Java Server Pages architecture enables the separation of content generation from content presentation. This separation not eases maintenance headaches; it also allows web team members to focus on their areas of expertise. Now, web page designer can concentrate on layout, and web application designers on programming, with minimal concern about impacting each other's work.

Steps in the execution of a JSP Application:

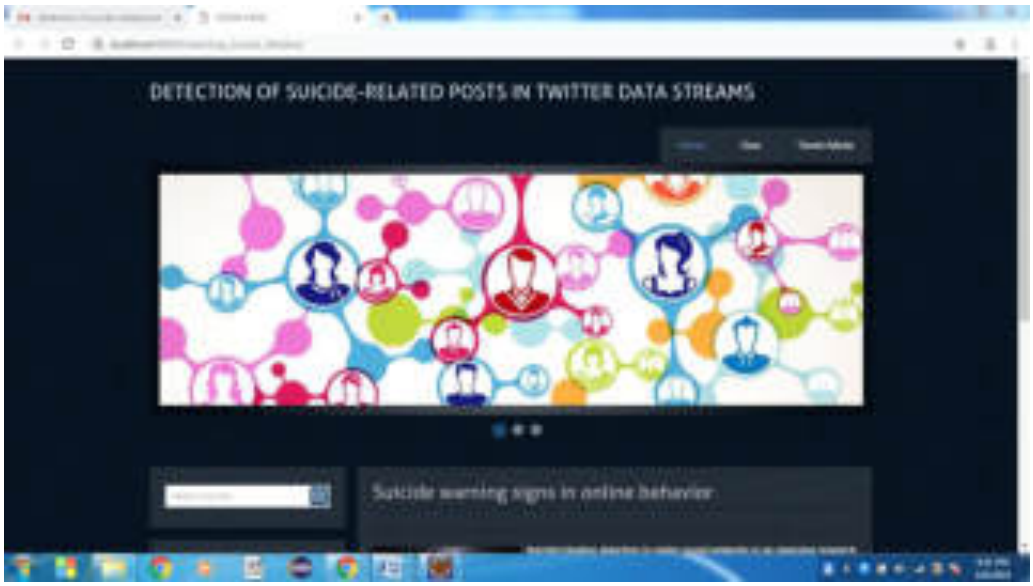
- The client sends a request to the web server for a JSP file by giving the name of the JSP file within the form tag of a HTML page.
- Ths request is transferred to the Java WebServer. At the server side Java WebServer receives the request and if it is a request for a jsp file server gives this request to the JSP engine.

- JSP engine is program which can understands the tags of the jsp and then it converts those tags into a Servlet program and it is stored at the server side. This Servlet is loaded in the memory and then it is executed and the result is given back to the Java WebServer and then it is transferred back to the result is given back to the Java WebServer and then it is transferred back to the client.

5.3. Methods of Implementation

Output screens

5.3.1 Project home page



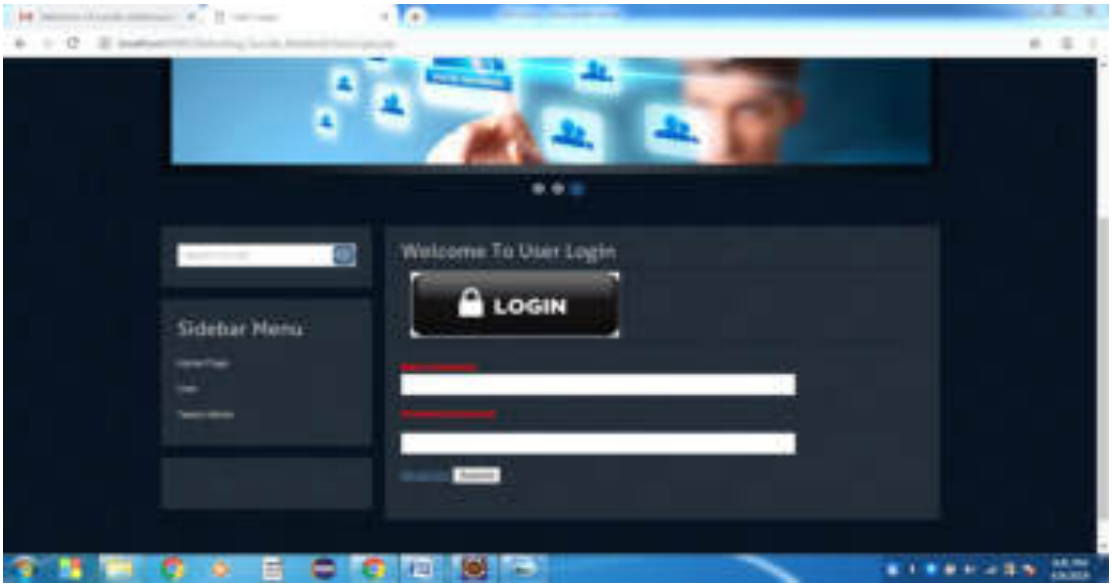
This is the Project Home Page, the essential Tool with important components of effective web design, properly styled with various color schemes and layout, this is the main interface of our project.

5.3.2.User registration form

It's important to establish user's brand and business presence on Twitter. First impressions count, and user's profile is the chance to make a lasting, positive one. Each element of the profile should accurately reflect the brand's purpose and values, and persuade a potential customer to follow.

The User Registration interface inculcates the Very Basic Details Of the desired user like Username, Password, Email Address, Mobile Number, Address, Date of birth, Gender & Profile Picture. Making all these formalities then the profile gets generated.

5.3.3Userloginpage



A standard **user account** is for those **users** who need to run multiple actions on the application, but they need limited or restricted access to **administrative** access to the application.

User needs to enter his Credentials which inculcate User Name & Password, both are mandatory fields. If once the credentials are properly given, then only it will drive the user to access the application.

If incase the specific User try access the application but he didn't have the account which is required, then the page automatically drives the user to create a new User profile. This is nothing but User Registration Form.

5.3.4 Server login page

The administrator account is something like for the user who wants to acquire full control over the interface and attain complete access.

User needs to enter his Credentials which inculcate User Name & Password, both are mandatory fields. If once the credentials are properly given, then only it will drive the user to access the application.

5.3.5 Administrator menu

When the user tries logging into the application, In the User Screen the interface includes The login widget at the center and a navigator Sidebar Menu consists of Home Page, User and a Tweet admin tabs. Home page is what we have seen in the very first screen. User Tab allows the user to get logged with the credentials. And Tweet Admin Tab is the main interface of the twitter application.

If once the User tried accessing the Tweet Admin Menu, then a series of tabs will be visible to the user, each tab has its own way of functioning.

5.3.6 Click on view all users

By clicking on View All Users, the admin can see all the list of users who are all using the twitter application. Which accommodates the list of all users which accommodates all their details including names, id's, address etc..

5.3.7 Shows the list of all the users

It Shows the list of all the users who are accessing and using the application with all the details like User Name, Mobile number, E-mail, Address and the Status of the Account.

5.3.8 The user is waiting to get authorized.

If the Profile of the User is recently created, it will take certain time for the profile gets Authorized. Till then the status of the account will be waiting.

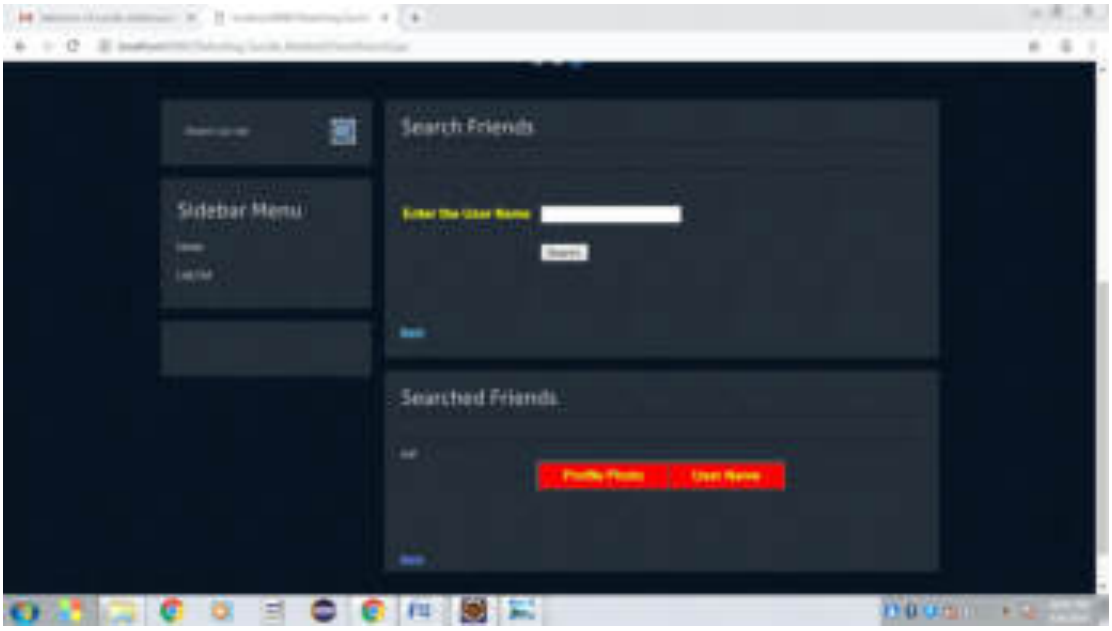
5.3.9 Admin authorizes the user

Only the Admin User have the Access to change the status of the user. The whole rights were issued only to the Admin User.

5.3.10 User successfully logged in

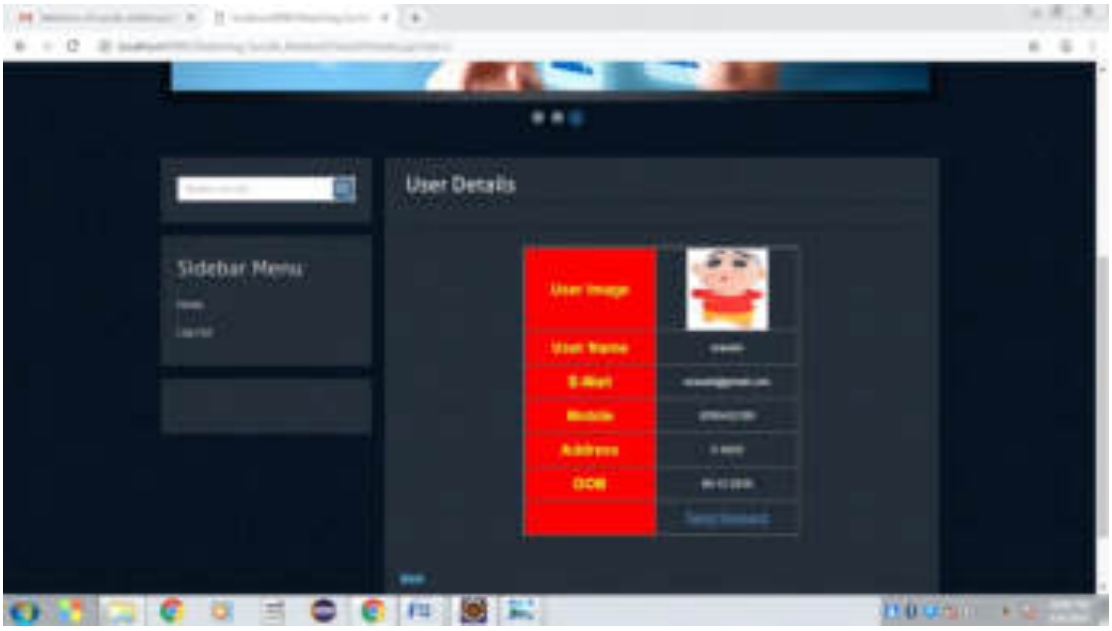
Only after entering the Credentials which inculcates User Name & Password, then only the user will be successfully logged into the application.

5.3.11 User can search friends



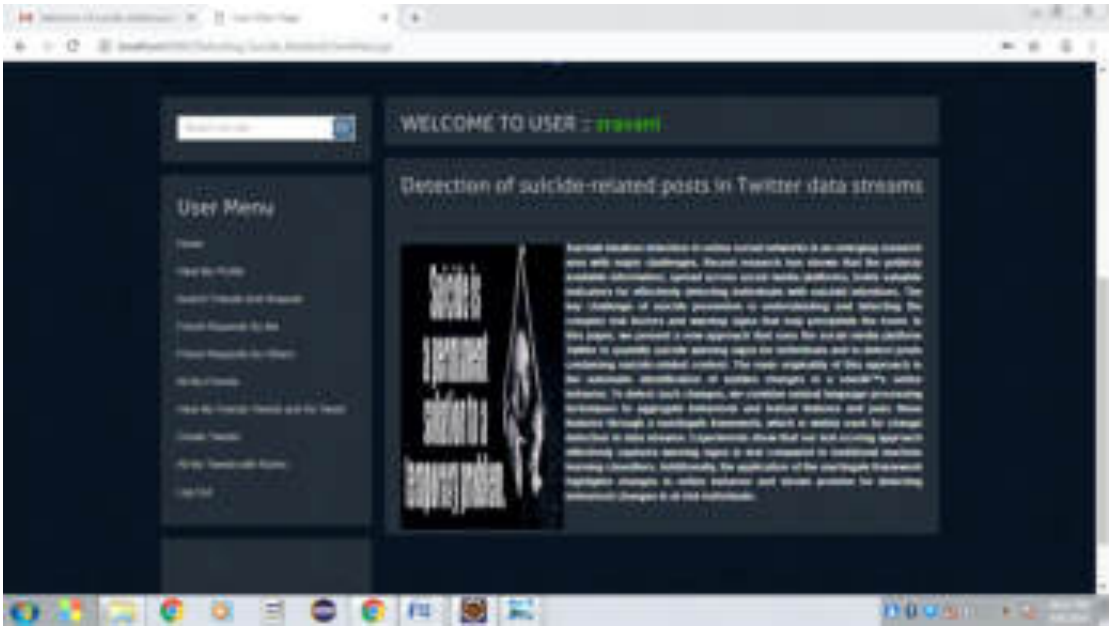
Here the user can search for friends by giving their name or id.

5.3.12 User sending request to friends



If the user finds his friends and wanted to add them to this circle, he can send request to them and can monitor who all are accepting their requests & their responses.

5.3.13 Friend clicks on friend requests by others



Here the user can find the friend requests which are sent to him by others.

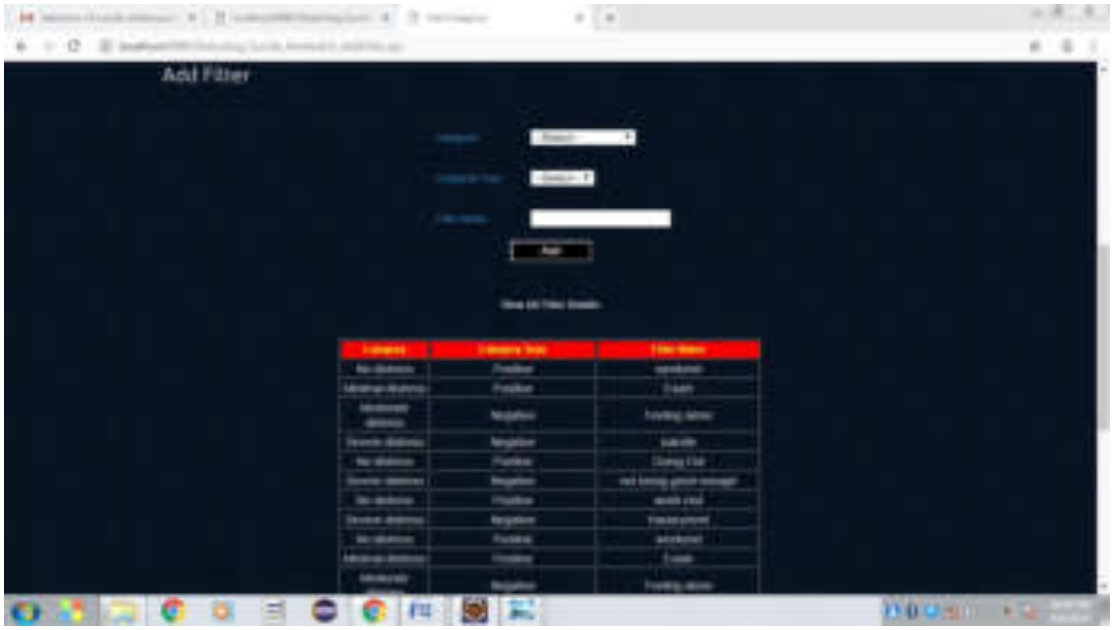
5.3.14 User waiting for friend to accept request

Simultaneously the user can see the status of his request whether it is accepted or waiting for approval.

5.3.15 Friend accepts the request

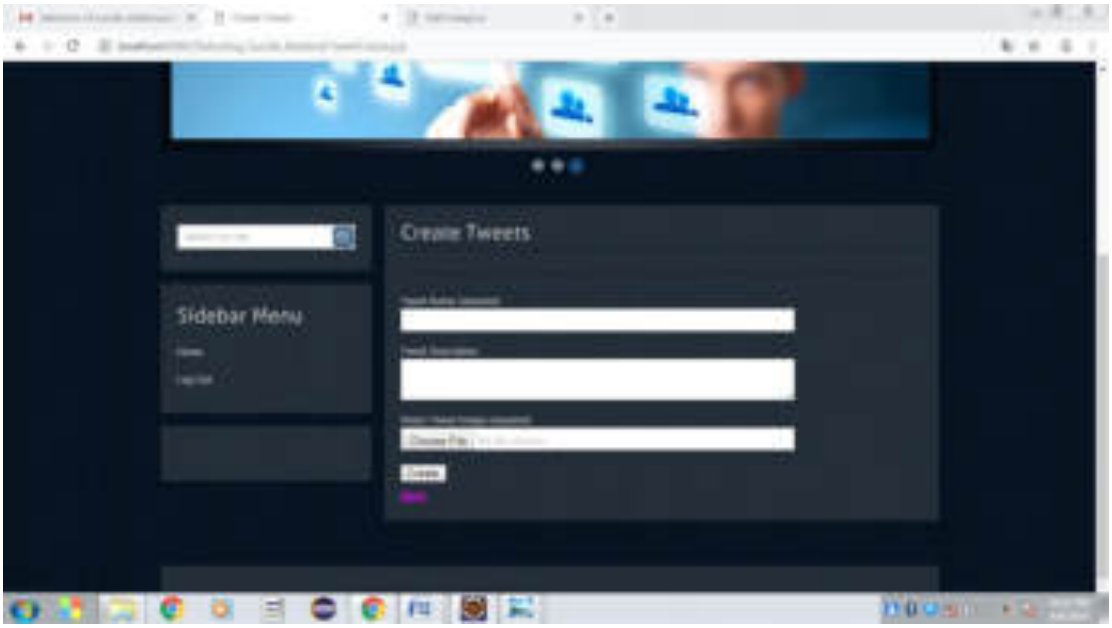
If the Friend Request was accepted, the accepted request can be seen in the dashboard.

5.3.16 Admin adds filter



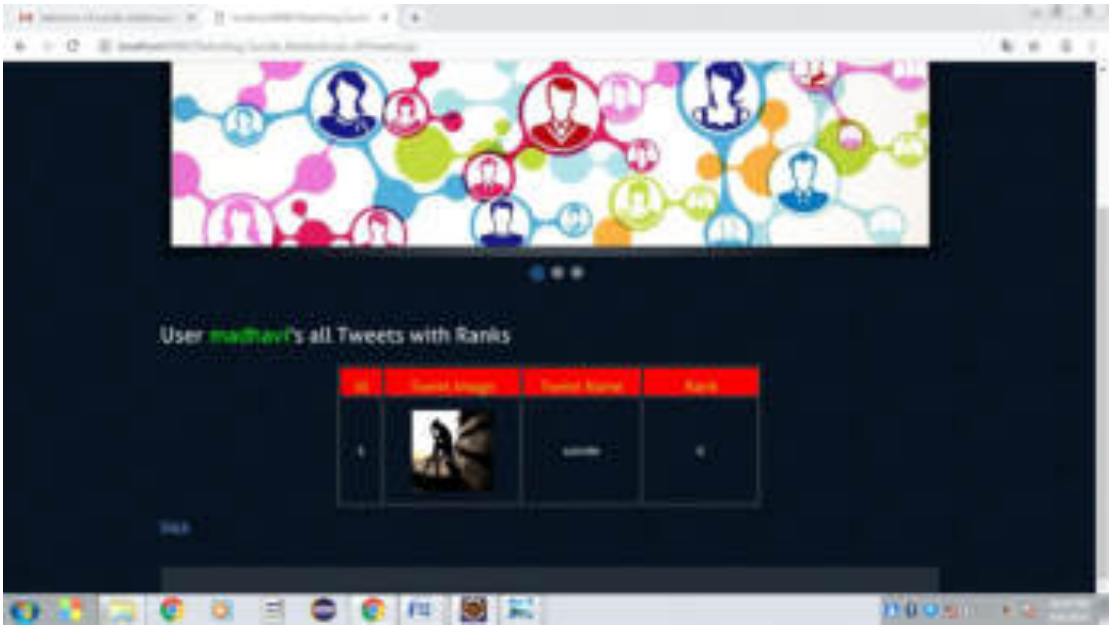
If the admin intended to filter out the the suicide related posts then he can make use of this widget. The Admin can filter out the posts based on the various categories.

5.3.17 User creates the tweets



This is the something like the news feed which includes all the posts, Tags done by the user. Users can create Tweets with respect to the mood swing or stress levels.

5.3.18 User's rank page



In case of any User Tweet or post depicting any irregular behavior or unusual behavior then those type of tweets are filtered out based on the pre-designed criteria scale.

5.3.19 Admin views all user's tweets

If any user posts which matches the criteria of the suicide or by few keywords then all the posts will be shortlisted and a final analysis is done by the admin.

5.3.20 Final analysis done by admin

The Final Analysis will be carried out by the admin and necessary rescue operation will be performed to save the user.

6. CODING

6.1 User profile

```
<style type="text/css">
<!--
.style1 {color: #FF0000}
-->
</style>
<table width="547" border="2" align="center"  cellpadding="0" cellspacing="0"  >
<%@ include file="connect.jsp" %>
<%@ page import="org.bouncycastle.util.encoders.Base64"%>
<%
String user=request.getParameter("value");
String s1,s2,s3,s4,s5,s6,s7,s8,s9,s10,s11,s12,s13;
 int i=0;
 try
{
 String query="select * from user where name='"+user+"'";
Statement st=connection.createStatement();
ResultSet rs=st.executeQuery(query);
if ( rs.next() )
{
i=rs.getInt(1);
s1=rs.getString(2);
s2=rs.getString(4);
```

```
//s6=rs.getString(10);

//if(s6.equals("waiting")){

//        s1=new String(Base64.encode(s1.getBytes()));

//        s2=new String(Base64.encode(s2.getBytes()));

//}

%>

<tr>

<td width="226" rowspan="6" ><div style="margin:10px 13px 10px 13px;" ><a class="#" id="img1" href="#" >

<input name="image" type="image" src="uProfilePic.jsp?id=<%=i%>" style="width:200px; height:200px;" />

</a></div></td>

</tr>

<tr>

<td width="129" height="62" valign="middle" bgcolor="#FFFFFF" style="color: #2c83b0;"><div align="left" class="style4 style14 style19 style24" style="margin-left:20px;">E-Mail</div></td>

<td  width="182" valign="middle" height="62"><div align="left" class="style23 style1" style="margin-left:20px;">

<%out.println(s2);%>

</div></td>

</tr>

<tr>

<td width="129" height="52" valign="middle" bgcolor="#FFFFFF" style="color: #2c83b0;"><div align="left" class="style4 style14 style19 style24" style="margin-left:20px;">Mobile</div></td>

<td  width="182" valign="middle" height="52"><div align="left" class="style23 style1" style="margin-left:20px;">
```

<%out.println(s3);%>

</td>

</tr>

<tr>

<td width="129" height="52" align="left" valign="middle" bgcolor="#FFFFFF" style="color: #2c83b0;"> Address </td>

<td width="182" align="left" valign="middle" height="52">

<%out.println(s4);%>

</td>

</tr>

<tr>

<td width="129" height="56" align="left" valign="middle" bgcolor="#FFFFFF" style="color: #2c83b0;"> DOB </td>

<%out.println(s5);%>

</td>

</tr>

<%

}

connection.close();

}

catch(Exception e)

{

```
out.println(e.getMessage());
}
%>
</table>
```

6.2 Admin profile

```
<%@ include file="connect.jsp" %>
<%
try {
        String id=request.getParameter("id");
        String str = "Authorized";
        Statement st1 = connection.createStatement();
        String query1 ="update user set status='"+str+"' where id="+id+" ";
        st1.executeUpdate (query1);
        connection.close();
        response.sendRedirect("ViewAllUsers.jsp");
   }
  catch(Exception e)
  {
     out.println(e.getMessage());
   }
   %>
```

6.3 View all users

```
<!DOCTYPE html PUBLIC "-// W3C //DTD XHTML 1.0 Transitional //EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">

<%@ include file="connect.jsp"%>

<html xmlns="http://www.w3.org/1999/xhtml">

<style type="text/css">

<!--

.style25 {color: #CCCC33}

-->

</style>

<head>

<title>View All Users</title>

<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />

<link href="css/style.css" rel="stylesheet" type="text/css" />

<script type="text/javascript" src="js/cufon-yui.js"></script>

<style type="text/css">

<!--

.style16 {font-size: 24px;

font-weight: bold;

}

.style21 {color: #0066CC; font-size: 18px; font-weight: bold; }

-->

</style>

</head>

<body>
```

```
<div class="main">
<div class="header">
<div class="header_resize">
<div class="text">
<h1><a href="#" class="style26"></a>Detection of suicide-related posts in Twitter data streams</h1>
</div>
<div class="menu_nav">
<ul>
<li><strong><a href="index.html">Home</a></strong></li>
<ul>
<li><a href="TSLogin.jsp"><strong>Tweet Admin </strong></a></li>
</ul>
</div>
<div class="clr"></div>
<div class="slider">
<div id="coin-slider"><a href="#"><img src="images/slide1.jpg" width="935"
height="307" alt="" /></a><a href="#"><img src="images/slide2.jpg" width="935"
<div class="clr"></div>
</div>
<div class="content">
<div class="content_resize">
<div class="post_content">
<h2 align="center">
<table width="842" border="1" align="center">
<tr>
```

```
<td width="143" height="32" align="left" valign="bottom" bgcolor="#FFFFFF"><div align="center" class="style21 style27">User Image </div></td>

<td width="148" align="left" valign="bottom" bgcolor="#FFFFFF"><div align="center" class="style21 style27">E-Mail</div></td>

</tr>

<%

        String s1="",s2="",s3="",s4="",s5="",s6="",s7="",s8="",s9,s10,s11,s12,s13="";

        int i=0,j=0;

 try

{

String query="select * from user ";

Statement st=connection.createStatement();

ResultSet rs=st.executeQuery(query);

while ( rs.next() )

{

i=rs.getInt(1);

s1=rs.getString(2);

s2=rs.getString(5);

%>

<tr>

<td><div align="center" class="style22">

<input  name="image" type="image" src="images1.jsp?imgid=<%=i%>" width="100" height="100" alt="Submit" />

</input>

</div></td>
```

```
<td align="center"  valign="middle"><div  align="center"   class="style22
style25"><%=s1%></div></td>

<%

if(s5.equalsIgnoreCase("waiting"))

{

 %>

<td  align="center"  valign="middle"><div align="center" class="

style22"><a href="admin_Status.jsp?id=<%=i%>">waiting</a></div></td>

<%

}

else

{

%>

<td  width="17"  align="center" valign=" middle"><  div    align="center"
class="style22"><%=s5%></div></td>

<%

}

%>

</tr>

<%

 }

 connection.close();

 }

 catch(Exception e)

{
```

```
out.println(e.getMessage());
}
%>
</table>
</h2>
</div>
</div>
<p> </p>
<p> </p>
<p> </p>
<p><a href="TSMain.jsp" class="style24">Back</a></p>
</div>
</div>
<div class="fbg">
<div class="fbg_resize">
<div class="clr"></div>
</div>
</div>
<div class="footer">
<div class="footer_resize">
<div style="clear:both;"></div>
</div>
<div align=center></div>
</body>s
</html>
```

6.4 Create Tweets

```
<!DOCTYPE   html PUBLIC " -//W3C //DTD  XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">

<html xmlns="http://www.w3.org/1999/xhtml">

<head>

<title>Create Tweets </title>

<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />

<link href="css/style.css" rel="stylesheet" type="text/css" />

<link rel="stylesheet" type="text/css" href="css/coin-slider.css" />

<script type="text/javascript" src="js/cufon-yui.js"></script>

<script type="text/javascript" src="js/coin-slider.min.js"></script>

<style type="text/css">

<!--

.style1 {font-size: 26px}

.style2 {

        font-size: 26px;

        font-weight: bold;

        font-family: Arial, Helvetica, sans-serif;

        color: #FFFFFF;

}

.style4 {

        font-size: 26px;

        font-weight: bold;

}

.style5 {
```

```
        color: #FF00FF;
        font-size: 14px;
}
.style6 {color: #FFFFFF}
-->
</style>
</head>
<body>
<div class="main">
<div class="header">
<div class="header_resize">
<div class="text">
<h1><a href="#" class="style1">Detection of suicide-related posts in Twitter data streams</a></h1>
</div>
<div class="menu_nav">
<ul>
<li><strong><a href="index.html">Home</a></strong></li>
<li class="active"><strong><a href="UserLogin.jsp"><strong>User</strong></a></strong></li>
<li><a href="TSLogin.jsp"><strong>Tweet Admin </strong></a></li>
</ul>
</div>
<div class="clr"></div>
<div class="slider">
<div id="coin-slider"><a href="#"><img src="images/slide1.jpg" width="935"
height="307" alt="" /></a></div>
```

```
<div class="clr"></div>
</div>
<div class="clr"></div>
</div>
<div class="content">
<div class="content_resize">
<div class="mainbar">
<div class="article">
<h2 class="style2">Create Tweets </h2>
<p class="infopost style6"> </p>
<form action="TweetCreateAuthen.jsp" method="post" id="" enctype="multipart/form-data">
<ol>
<li>
<label for="tweet">Tweet  Name (required)</label>
<input id="tweet" name="tweet" class="text" />
</li>
<li>
<label for="description">Tweet Description </label>
<textarea id="description" name="description" rows="3" cols="50"></textarea>
</li>
<li>
<label for="pic">Select Tweet Image (required)</label>
<input type="file" id="pic" name="pic" class="text" />
</li>
<li><br />
```

<input name="submit" type="submit" value="Create" />

Back

</form>

<form id="formsearch" name="formsearch" method="post" action="#">

<input name="editbox_search" class="editbox_search" id="editbox_search" maxlength="80" value="Search our ste:" type="text" />

<input name="button_search" src="images/search.gif" class="button_search" type="image" />

</form>

Home

Log Out

<h2 class="star"> </h2>

```
<div class="clr"></div>
</div>
<div class="fbg">
<div class="fbg_resize">
<div class="clr"></div>
</div>
<div align=center></div>
</body>
</html>
```

6.5 View All Suicide Related Posts

```
<! DOCTYPE html     PUBLIC " -// W3C // DTD  XHTML 1.0  Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<style type="text/css">
<!--
.style31 {font-size: 26px}
.style38 {color: #FFFFFF}
-->
</style>
<head>
<title>A_View All Suicide-related Posts</title>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<link href="css/style.css" rel="stylesheet" type="text/css" />
<link rel="stylesheet" type="text/css" href="css/coin-slider.css" />
```

```
<script type="text/javascript" src="js/cufon-yui.js"></script>
<style type="text/css">
<!--
.style1 {font-size: 30px}
        font-weight: bold;
}
.style24 {font-size: 16px}
-->
</style>
</head>
<body>
<div class="main">
<div class="text">
<h1><a href="#" class="style1 style31">Detection of suicide-related posts in Twitter data streams</a></h1>
</div>
<div class="menu_nav">
<ul>
<li><strong><a href="index.html"><strong>Home</strong></a></strong></li>
<li><a href="UserLogin.jsp"><strong>User</strong></a></li>
<li class="active"><a href="TSLogin.jsp"><strong>Tweet Admin </strong></a></li>
</ul>
</div>
<div class="clr"></div>
<div class="slider">
```

```
<div id="coin-slider"><a href="#"><img src="images/slide1.jpg" width="935"

height="307" alt="" /></a></div>

<div class="clr"></div>

</div>

</div>

</div>

<div class="content">

<h2 align="center" class="style16 style38">Detection of suicide-related posts....Micro Blogs... </h2>

<div class="clr"></div>

<h2 align="center" class="style6"><span class="style16 style32">View All No Distress</span> Details
</h2>

<iframe width="750" height="300" src="A_No_distress.jsp"></iframe>

<p> </p>

<a href="TSMain.jsp" class="style24">Back</a></div>

</div>

<div class="fbg">

<p> </p>

</div><div class="footer">

<div class="footer_resize">

<div style="clear:both;"></div>

</div>

<div align=center></div>

</body>

</html>
```

7. TESTING AND VALIDATION

7.1 Introduction

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

7.2 Design of Test Cases and Scenarios

The following are the Testing Methodologies:

- Unit Testing.
- Integration Testing
- User Acceptance Testing
- Output Testing
- Validation Testing

Unit Testing

Unit testing focuses verification effort on the smallest unit of Software design that is the module. Unit testing exercises specific paths in a module's control structure to ensure complete coverage and maximum error detection. This test focuses on each module individually, ensuring that it functions properly as a unit. Hence, the naming is Unit Testing.

During this testing, each module is tested individually and the module interfaces are verified for the consistency with design specification. All important processing path are tested for the expected results. All error handling paths are also tested.

Integration Testing

Integration testing addresses the issues associated with the dual problems of verification and program construction. After the software has been integrated a set of high order tests are conducted. The main objective in this testing process is to take unit tested modules and builds a program structure that has been dictated by design.

The following are the types of Integration Testing:

- **Top Down Integration**

This method is an incremental approach to the construction of program structure. Modules are integrated by moving downward through the control hierarchy, beginning with the main program module. The module subordinates to the main program module are incorporated into the structure in either a depth first or breadth first manner.

In this method, the software is tested from main module and individual stubs are replaced when the test proceeds downwards.

2. Bottom-up Integration

This method begins the construction and testing with the modules at the lowest level in the program structure. Since the modules are integrated from the bottom up, processing required for modules subordinate to a given level is always available and the need for stubs is eliminated. The bottom up integration strategy may be implemented with the following steps:

- The low-level modules are combined into clusters into clusters that perform a specific Software sub-function.
- A driver (i.e.) the control program for testing is written to coordinate test case input and output.
- The cluster is tested.
- Drivers are removed and clusters are combined moving upward in the program structure

The bottom up approaches tests each module individually and then each module is module is integrated with a main module and tested for functionality.

User Acceptance Testing

User Acceptance of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changes wherever required. The system developed provides a friendly user interface that can easily be understood even by a person who is new to the system.

Output Testing

After performing the validation testing, the next step is output testing of the proposed system, since no system could be useful if it does not produce the required output in the specified format. Asking the users about the format required by them tests the outputs generated or displayed by the system under consideration. Hence the output format is considered in 2 ways – one is on screen and another in printed format.

Validation Testing

The process of evaluating software during the development process or at the end of the development process to determine whether it satisfies specified business requirements.

Validation Testing ensures that the product actually meets the client's needs. It can also be defined as to demonstrate that the product fulfills its intended use when deployed on appropriate environment.

7.3 Validation

Validation checks are performed on the following fields.

Text Field:

The text field can contain only the number of characters lesser than or equal to its size. The text fields are alphanumeric in some tables and alphabetic in other tables. Incorrect entry always flashes and error message.

Numeric Field:

The numeric field can contain only numbers from 0 to 9. An entry of any character flashes an error messages. The individual modules are checked for accuracy and what it has to perform. Each module is subjected to test run along with sample data. The individually tested modules are integrated into a single system. Testing involves executing the real data information is used in the program the existence of any program defect is inferred from the output. The testing should be planned so that all the requirements are individually tested.

A successful test is one that gives out the defects for the inappropriate data and produces and output revealing the errors in the system.

Preparation of Test Data:

Taking various kinds of test data does the above testing. Preparation of test data plays a vital role in the system testing. After preparing the test data the system under study is tested using that test data. While testing the system by using test data errors are again uncovered and corrected by using above testing steps and corrections are also noted for future use.

Using Live Test Data:

Live test data are those that are actually extracted from organization files. After a system is partially constructed, programmers or analysts often ask users to key in a set of data from their normal activities. Then, the systems person uses this data as a way to partially test the system. In other instances, programmers or analysts extract a set of live data from the files and have them entered themselves.

It is difficult to obtain live data in sufficient amounts to conduct extensive testing. And, although it is realistic data that will show how the system will perform for the typical processing requirement, assuming that the live data entered are in fact typical, such data generally will not test all combinations or formats that can enter the system. This bias toward typical values then does not provide a true systems test and in fact ignores the cases most likely to cause system failure.

Using Artificial Test Data:

Artificial test data are created solely for test purposes, since they can be generated to test all combinations of formats and values. In other words, the artificial data, which can quickly be prepared by

a data generating utility program in the information systems department, make possible the testing of all login and control paths through the program.

The most effective test programs use artificial test data generated by persons other than those who wrote the programs. Often, an independent team of testers formulates a testing plan, using the systems specifications.

The package "Virtual Private Network" has satisfied all the requirements specified as per software requirement specification and was accepted.

User Training

Whenever a new system is developed, user training is required to educate them about the working of the system so that it can be put to efficient use by those for whom the system has been primarily designed. For this purpose the normal working of the project was demonstrated to the prospective users. Its working is easily understandable and since the expected users are people who have good knowledge of computers, the use of this system is very easy.

Maintenance

This covers a wide range of activities including correcting code and design errors. To reduce the need for maintenance in the long run, we have more accurately defined the user's requirements during the process of system development. Depending on the requirements, this system has been developed to satisfy the needs to the largest possible extent. With development in technology, it may be possible to add many more features based on the requirements in future. The coding and designing is simple and easy to understand which will make maintenance easier.

Testing strategy:

A strategy for system testing integrates system test cases and design techniques into a well planned series of steps that results in the successful construction of software. The testing strategy must co-operate test planning, test case design, test execution, and the resultant data collection and evaluation .A strategy for software testing must accommodate low-level tests that are necessary to verify that a small source code segment has been correctly implemented as well as high level tests that validate major system functions against user requirements.

Software testing is a critical element of software quality assurance and represents the ultimate review of specification design and coding. Testing represents an interesting anomaly for the software. Thus, a series of testing are performed for the proposed system before the system is ready for user acceptance testing.

System Testing:

Software once validated must be combined with other system elements (e.g. Hardware, people, and database). System testing verifies that all the elements are proper and that overall system function performance is achieved. It also tests to find discrepancies between the system and its original objective, current specifications and system documentation.

Unit Testing:

In unit testing different are modules are tested against the specifications produced during the design for the modules. Unit testing is essential for verification of the code produced during the coding phase, and hence the goals to test the internal logic of the modules. Using the detailed design description as a guide, important Conrail paths are tested to uncover errors within the boundary of the modules. This testing is carried out during the programming stage itself. In this type of testing step, each module was found to be working satisfactorily as regards to the expected output from the module.

In Due Course, latest technology advancements will be taken into consideration. As part of technical build-up many components of the networking system will be generic in nature so that future projects can either use or interact with this. The future holds a lot to offer to the development and refinement of this project.

8. CONCLUSION

In this project, we designed and evaluated a novel approach to monitor the mental health of a user on Twitter. Building off existing research, we worked to translate and quantify suicide warning signs in an online context (user-centric and post-centric behavioral features). In particular, we focused on detecting distress-related and suicide-related content and developed two approaches to score a tweet: an NLP-based approach and a more traditional machine learning text classifier.

To detect changes in emotional well-being, we considered a Twitter user's activity as a stream of observations and applied a martingale framework to detect change points within that stream. Our experiments show that our NLP text-scoring approach successfully separates out tweets exhibiting distress-related content and acts as a powerful input into the martingale framework. While the martingale values "react" to changes in online speech, the change point detection method needs improvement. We were able to detect the true change point for one validation case, but the approach needs to be more robust with respect to parameter setting and positive changes in speech.

For future research, we plan to further explore the impact of martingale parameters on the change detection effectiveness. We also hope to expand the approach to include image processing and other social media outlets in order to assess the effectiveness in other settings. Another interesting perspective is to consider more fine-grained emotion classes such as anger, sadness, fear, etc., instead of considering four levels of distress. However, overall, we believe our initial work presents an innovative approach to detecting suicide-related content in a text stream setting.

9. REFERENCES

1. American Psychiatric Association, “Practice guideline for the assessment and treatment of patients with suicidal behaviors,” Amer. J. Psychiatry vol. 160, no. 11,pp. 1-60,2003.
2. American Foundation for Suicide Prevention (AFSP). [Online].Available: https://afsp.org
3. B. O’Dea, S. Wan, P. J. Batterham, et al., “Detecting suicidality on twitter,” Internet Interventions, vol. 2, no. 2, pp. 183–188, 2015.
4. D. Rickwood, F. P. Deane, C. J. Wilson, et al., “Young people’shelp seeking for mental health problems,” Aust. e-J. Adv. MentalHealth, vol. 4, no. 3, pp. 218–251, 2005.
5. G. Coppersmith, et al.,“Exploratory analysis of social media prior to a suicide attempt,” in Proc. 3rd Workshop Comput. Linguistics Clinical Psychol., 2016, pp. 106–117.
6. J. Jashinsky, S. H. Burton, C. L. Hanson, et al., “Tracking suicide risk factors through Twitter in the US,” J. Crisis Intervention Suicide Prevention, vol. 35, no. 1, pp. 51–59, 2014.
7. J. Robinson, G. Cox, E. Bailey, et al., “Social media and suicide prevention: A systematic review,” Early Intervention Psychiatry,vol. 10, no. 2, pp. 103–121, 2016.
8. J. F. Gunn and D. Lester, “Twitter postings and suicide: Ananalysis of the postings of a fatal suicide in the 24 hours prior to death,” Suicidologi, vol. 17, no. 3, pp. 28–30, 2012.
9. M. De Choudhury, E. Kiciman, M. Dredze, et al., “Discovering shifts to suicidal ideation from mental health content in socialmedia,” in Proc. 2016 CHI Conf. Human Factors Comput. Syst.,San Jose, CA, USA, 2016, pp. 2098–2110.
10. M. De Choudhury, S. Counts, and E. Horvitz, “Predicting postpartum changes in emotion and behavior via social media,”inProc. SIGCHI Conf. Human Factors Comput. Syst., New York,NY, USA, 2013, pp. 3267–3276.
11. M. D. Rudd, A. L. Berman, T. E. Joiner, et al., “Warning signs for suicide: Theory, research, and clinical applications,” Suicide Life- Threatening Behav., vol. 36, no. 3, pp. 255–262, 2006.
12. M. T. Lehrman, C. O. Alm, and R. A. Proano, “Detecting distressed and non-distressed affect states in short forum texts,”inProc. 2012 Workshop Lang. Social Media, Montreal, QC, Canada,2012, pp. 9–18.
13. M. De Choudhury, M. Gamon, S. Counts, et al., “Predicting depression via social media,” in Proc. 7th Int. AAAI Conf.Weblogs Social Media, Boston, MA, USA, 2013, pp. 128–137.
14. “Mental health action plan 2013–2020,” World Health Organization, Geneva, Switzerland,2013. [Online]. Available:ht_tp://www_.who.int/mental_health/publications/action_plan/en

15. M. Moreno, L. Jelenchick, K. Egan, et al., “Feeling bad onFacebook: depression disclosures by college students on a socialnetworking site,” Depression Anxiety, vol. 28, no. 6, pp. 447–455,2011.
16. P. Corrigan, “How stigma interferes with mental health care,”Amer. Psychologist, vol. 59, no. 7, pp. 614–525, 2004.
17. “Preventing suicide: A global imperative,” World Health Organization, Geneva, Switzerland, 2014. [Online]. Available:ht_tp://www_.who.int/mentalhealth/suicideprevention/world_report_2014/en
18. S. R. Braithwaite, C. Giraud-Carrier, J. West, et al., “Validating machine learning algorithms for twitter data against establishedmeasures of suicidality,” JMIR Mental Health, vol. 3, no. 2, 2015,Art. no. e21.
19. S. S. Rude, E. M. Gortner, and J. W. Pennebaker, “Language use of depressed and depression-vulnerable college students,” Cogn. Emotion, vol. 18, no. 8, pp. 1121–1133, 2004.
20. V. Kailasam and E. Samuels, “Can social media help mental health practitioners prevent suicides?” Current Psychiatry, vol. 14, no. 2, pp. 37–51, 2015.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"ANDROID ATTENDANCE SYSTEM"

A Project report submitted to Jawaharlal Nehru Technological University Kakinada in partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING.

By

P. Lalithamma (17H71A0521)

A. Neelima (17H71A0528)

K. Mounika (17H71A0524)

P. Yamini Lakshmi (17H71A0558)

Under the Guidance of
Ms K. VinayaSree, Assistant Professor
Department of Computer Science & Engineering

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Autonomous)
Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada
Accredited with NAAC 'A' Grade, NBA Accredited B.Tech(CSE, ECE, EEE, ME). An ISO 9001:2008 Certified Institute
Kanchikacherla - 521180, Krishna Dist., A.P, India. Phone No.: 08676-273535, 273623

Devineni Venkata Ramana & Dr. Hima Sekhar

MIC College of Technology

(Autonomous)

CERTIFICATE

This is to certify that the project work entitled **"ANDROID ATTENDANCE SYSTEM"** is a bonafied work carried out by Ms. P. Lalithamma (17H71A0521), Ms. A. Neelima (17H71A0528), Ms. K. Mounika (17H71A0524), Ms. P. Yamini Lakshmi (17H71A0558) in partial fulfilment for the award of the degree of Bachelor of Technology in Computer Science and Engineering of Jawaharlal Nehru Technological University, Kakinada during the year 2020-2021. It is certified that all corrections/ suggestions indicated for assessment have been incorporated in the report. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the above degree.

Project Guide	**Head of the Department**
K. VinayaSree	**Dr D. Prasad**
Examiner	**Principal**
	Dr. K. Srinivasa Rao

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Autonomous)
Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada
Accredited with NAAC 'A' Grade, NBA Accredited B.Tech(CSE, ECE, EEE, MEC) An ISO 9001:2008 Certified Institute
Kanchikacherla - 521180, Krishna Dist., A.P. India. Phone No.: 08678-273333, 273433

DECLARATION

Hereby we, who carried out the project on "**Android Attendance System"**, declare that the matter included in this project report is a genuine work done by us and has not, been submitted to this university or any other university/institute for the fulfilment of the requirement of the degree.

P. Lalithamma (17H71A0521)

A. Neelima (17H71A0528)

K. Mounika (17H71A0524)

P. Yamini Lakshmi (17H71A0558)

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
(Autonomous)
Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada
Accredited with NAAC 'A' Grade, NBA Accredited B.Tech(CSE, ECE, EEE, MEC), An ISO 9001:2008 Certified Institute
Kanchikacherla - 521180, Krishna Dist., A.P. India. Phone No: 08678-273233, 273823

ACKNOWLEDGMENT

We are grateful to Ms **K. VinayaSree**, Assistant Professor, **Department of Computer Science and Engineering**, whose guidance has been immense all throughout the project work.

Our sincere thanks to **Dr. D. Prasad**, **Head of the Department**, **Computer Science and Engineering**, for his valuable guidance and timely advices in the completion of our project.

We express our heartfelt gratitude to **Dr. K. Srinivasa Rao**, **Principal** for permitting us to take up our project work and to complete the project successfully.

We would like to express our sincere and heartful thanks to all the lecturers of the department for their continuous cooperation, which has given us the cogency to build up adamant aspiration over the completion of our project.

Finally, we thank one and all who directly and indirectly helped us to complete our project successfully.

List of contents

ABSTRACT

Student attendance system is the system of tracking the attendance of the student on the basis of presence in class. Successful industries, schools, and universities begin by engaging students and making sure that they will come regularly so the attendance rate becomes very important. In this paper, a smart student attendance system is designed and implemented based on the android operating system. In compression with other traditional attendance systems, the proposed system provides a faster, cheaper and reachable system for online student attendance and generates the attendance report automatically.

List of figures

List of Tables

List Of Output Screens

1. INTRODUCTION

1.1 Definition:

Attendance Management System is a software developed for daily student attendance in schools, colleges and institutes. It facilitates access to the attendance information of a particular student in a particular class. The information is sorted by the operators, which will be provided by the teacher for a particular class. This system will also help in evaluating attendance eligibility criteria of a student.

1.2 Purpose

The purpose of developing attendance management systems is to computerize the traditional way of taking attendance. Another purpose for developing this software is to generate the report automatically at the end of the session or in between of the session.

1.3 Scope

The following project has much scope both in present as well as in future. In the present situation the system can be accessed both in lappies and mobile devices. The scope of the project is any mobile device which has access to it's server on which the project has been deployed, i.e. the project is developed as a web application, and it will work for a particular institute. But later on the project can be modified to operate for many institutes. In the future the system can be automated using students' fingerprints.

1.4 Overview

Our project is online attendance management system and the objectives are:

*. Reduces admin work by integrating the details of the students of all the departments into a single database. Manual work for information retrieval on attendance becomes less as the work becomes digitized.

*. Easy access for students because they can view their attendance and make up for the shortage of attendance accordingly. It is also time saving as manual work is less and chances of errors are less.

*. It eliminates duplicate data entry in time and attendance entries Auto-generation of various types of reports of student attendance.

2. PROJECT DESCRIPTION

2.1 Problem Definition

This system will reduce manual work and avoid redundant data. By maintaining the attendance manually, then efficient reports cannot be generated. The system can generate an efficient weekly, consolidated report based on the attendance. As the attendances are maintained in registers it has been a tough task for admin and staff to maintain for a long time. Instead the software can keep long and retrieve the information when needed.

2.2 Project Overview

Attendance Management System basically has two main modules for proper functioning: Admin module has rights for creating any new entry of faculty and student details. Users have the rights of making daily attendance, generating reports. Attendance reports can be taken by giving details of student details, date, and class.

2.3 Product Perspective

The product Attendances Management system is an independent product and does not depend on any other product or system. The product will automate various tasks associated with handling student details and better organizing the stored information and optimum performance, thus helping the Colleges to ensure smooth working of these processes.

2.4 Product Functions

Our system has two types of accessing modes:-

i. Administrator

ii. User

* Teacher

* Student

(i) Administrator

Administrators have rights to manage student details, add a new student, provide register number for all students, assign each student a course etc., Administrator can update his profile, and also can give help to the teachers and students.

(ii) User:

There are two users:

* Student:

Students do the login and see profile, View Total Attendance, Apply Leave, Complain etc.

* Faculty:

Add students, view the student details and take attendance, Generate Report and View Total Attendance.

2.5 User Characteristics

This software gives access to two kinds of users.

1. Administrator:

The administrators have features to add, delete and modify information stored in the database.

2. Authorized User:

Teaching staff have access to view the data stored in the database and can update the student's attendance in the form of formatted reports. Students have access to view the data stored in the database.

2.6 Constraints

Interface is only in English; no other language option is available. User can login with his assigned username and password, no guest facilities are available.

2.7 Assumptions and Dependencies

We assume that the Office personnel do all the data entry based and the correct values obtained from forms and registers. We assume that the computers that will use the software will be part of the college LAN. Users with administrator access should be careful in deleting or modifying any information knowingly or unknowingly which will lead to inconsistency of the database. The end users of this software are assumed to have a basic level of computer knowledge i.e. point and click.

3. SYSTEM ANALYSIS

3.1 Introduction

Analysis can be defined as breaking up any whole so as to find out their nature, function etc. It defines design as to make preliminary sketches of; to sketch a pattern or outline for plan. To plan and carry out especially by artistic arrangement or in a skillful wall. System analysis and design can be characterized as a set of techniques and processes, a community of interests, a culture and an intellectual orientation. The various tasks in the system analysis include the following.

Understanding applications.

Planning.

Scheduling.

Developing candidate solution.

Performing trade studies.

Performing cost benefit analysis.

Recommending alternative solutions.

Supervising, installing and maintaining the system.

This system manages the analysis of the report creation and develops manual entry of the student attendance. First design the students entry form , staff allocation and time table allocation forms. This project will help the attendance system for the department calculate percentage and reports for eligibility criteria of examination .The application attendance entry system will provide a flexible report for all students.

3.2 Existing System

Existing system is a manual entry for the students. Here the attendance will be carried out in the hand written registers. It will be a tedious job to maintain the record for the user. The human effort is more here. The retrieval of the information is not as easy as the records are maintained in the hand written registers. This application requires correct feed on input into the respective field. Suppose the wrong inputs are entered, the application resists working. so the user finds it difficult to use.

Disadvantages of Existing System

- **Not User Friendly:**
    - The existing system is not user friendly because the retrieval of data is very slow and data is not maintained efficiently.
- **Difficulty in report generating:**
    - We require more calculations to generate the report so it is generated at the end of the session and the student does not get a single chance to improve their attendance.
- **Manual control:**
    - All calculations to generate reports are done manually so there is greater chance of errors.
- **Lots of paperwork:**
    - Existing system requires a lot of paperwork. Loss of even a single register/record led to a difficult situation because all the papers are needed to generate the reports.
- **Time consuming:**
    - Every work is done manually so we cannot generate reports in the middle of the session or as per the requirement.

3.3 Proposed System

To overcome the drawbacks of the existing system, the proposed system has been evolved. This project aims to reduce the paperwork and save time to generate accurate results from the student's attendance. The system provides the best user interface. The efficient reports can be generated by using this proposed system.

Characteristics of the Proposed System

- **User Friendly:**

    The proposed system is user friendly because the retrieval and storing of data is fast and data is maintained efficiently. Moreover the graphical user interface is provided in the proposed system, which allows users to deal with the system very easily.

- **Reports are easily generated:**

  Reports can be easily generated in the proposed system so users can generate the report as per the requirement (monthly) or in the middle of the session. Users can give the notice to the students so he/she becomes regular.

- **Very less paperwork:**

  The proposed system requires very less paper work. All the data is feted into the computer immediately and reports can be generated through computers. Moreover, work becomes very easy because there is no need to keep data on papers.

- **Computer operator control:**

  Computer operator control will be there so no chance of errors. Moreover, storing and retrieving information is easy. So work can be done.

3.4 Feasibility Study

Feasibility analysis begins once the goals are defined. It starts by generating broad possible solutions, which are possible to give an indication of what the new system should look like. This is where creativity and imagination are used. Analysts must think up new ways of doing things- generate new ideas. There is no need to go into the detailed system operation yet. The solution should provide enough information to make reasonable estimates about project cost and give users an indication of how the new system will fit into the organization. It is important not to exert considerable effort at this stage only to find out that the project is not worthwhile or that there is a need to significantly change the original goal. Feasibility of a new system means ensuring that the new system, which we are going to implement, is efficient and affordable.

3.4.1 Importance

Feasibility studies allow companies to determine and organize all of the necessary details to make a business work. A feasibility study helps identify logistical problems, and nearly all business-related problems, along with the solutions to alleviate them. Feasibility studies can also lead to the development of marketing strategies that convince investors or a bank that investing in the business is a wise choice.

3.4.2 Components

There are several components of a feasibility study:

- **Description**

  A layout of the business, the products and/or services to be offered and how they will be delivered.

- **Market feasibility**
  - describes the industry, the current and future market potential, competition, sales estimations and prospective buyers.
- **Technical feasibility**
  - Lays out details on how a good or service will be delivered, which includes transportation, business location, technology needed, materials and labour.
- **Financial feasibility**
  - A projection of the amount of funding or start-up capital needed, what sources of capital can and will be used, and what kind of return can be expected on the investment.
- **Organizational feasibility**
  - A definition of the corporate and legal structure of the business; this may include information about the founders, their professional background and the skills they possess necessary to get the company off the ground operational.

There are various types of feasibility to be determined. They are

3.4.3 Economic Feasibility

Definition

Economic feasibility is also known as COST/BENEFIT ANALYSIS. Analysis of a project's costs and revenues in an effort to determine whether or not it is logically possible to complete the purpose of an economic feasibility study (EFS) is to demonstrate the net benefit of a proposed project for accepting or disbursing electronic funds/benefits, taking into consideration the benefits and costs to the agency, other state agencies, and the general public as a whole. The EFS is composed of two required forms:

· Business Case

· Cost Benefit Analysis

How Is The Project Economically Feasible?

- The system being developed is economic with respect to School or College's point of view. It is cost effective in the sense that it has eliminated the paper work completely.
- The system is also time effective because the calculations are automated which are made at the end of the month or as per the user requirement.
- The result obtained contains minimum errors and are highly accurate as the data is required.
- Development of this application is highly economically feasible. The only thing to be done is making an environment with effective supervision.

3.4.4 Behavioural Feasibility

The system working is quite easy to use and learn due to its simple but attractive interface. Users require no special training for operating the system.

3.4.5 Technical Feasibility

The technical requirement for the system is economic and it does not use any other additional Hardware and software. Technical evaluation must also assess whether the existing systems can be upgraded to use the new technology and whether the organization has the expertise to use it.

3.4.6 Operational Feasibility

The system working is quite easy to use and learn due to its simple but attractive interface. Users require no special training for operating the system. Technical performance includes issues such as determining whether the system can provide the right information for the Department personnel student details, and whether the system can be organized so that it always delivers this information at the right place and on time using intranet services. Acceptance revolves around the current system and its personnel.

4. REQUIREMENTS SPECIFICATION

4.1 Hardware Requirements

- RAM : 4GB or more
- HARD DISK: 500GB free
- Any smartphone

4.2 Software Requirements

- Operating System: Windows 10
- Front Design: Android Studio
- Front-End Language: Android
- Back-End Language: Java
- Databases: MySQL

4.3 Functional Requirements

Attendance Management System involves the following functions - Easily track attendance information of students. Quickly produce an attendance bulletin. The functions are:

1. Add Student
2. Add Faculty
3. Add Attendance
4. View Total Attendance
5. Attendance per Student
6. Leave
7. Generate Report

4.4 Non-Functional Requirements

4.4.1 Performance

Easy tracking of records and updating can be done. All the requirements relating to performance characteristics of the system are specified in the section below. There are two types of requirements.

1. Static Requirements
2. Dynamic Requirements

1. Static Requirements:

These requirements do not impose any constraints on the execution characteristics of the System. They are:

Number of Terminals:

The software makes use of an underlying database that will reside at the server, while the front end will be available online to the administrative and departmental computers as well as students and teachers.

Number of Users:

The number of users may vary, as this software finds applications in almost all departments of the organization.

2. Dynamic Requirements:

These specify constraints on the execution characteristics of the system. They typically include response time and throughout the system. Since these factors are not applicable to the proposed software, it will suffice if the response time is high and the transactions are carried out precisely and quickly.

4.4.2 Reliability

The software will not be able to connect to the centralized database in the event that the college LAN fails or in the event of the server being down due to a hardware or software failure.

4.4.3 Availability

The software will be available only to authorized users of the colleges like teachers to mark the students' attendance, students to view their enrolled course, admin to add an update students records.

4.4.4 Security:

The security requirements deal with the primary security. The software should be handled only by the administrator and authorized users. Only the administrator has the right to assign permission like creating new accounts and generating passwords. Only authorized users can access the system with username and password. 10 4.4.5 Maintainability Backups for databases are available.

4.4.5 Portability:

The Software is a windows-based application and is built in VB.Net and MYSQL so it is platform independent and is independent of the operating system.

4.5 Design Constraints:

This software provides security. The login form prevents the system from being misused by unauthorized users. Only an authorized operator will be granted rights to modify as per requirements. This software is also reliable and fault tolerant. The system developed is designed to handle invalid inputs. Since reliability is a major area of concern the system has a backup to avoid data loss. The user should know the programming language very well that is used to develop software.

5. SOFTWARE DESCRIPTION

5.1 Android Studio

Android Studio is the official Integrated Development Environment (IDE) for Android app development. Android Studio offers more features that enhance your productivity when building Android apps, such as:

- A flexible Gradle-based build system
- A fast and feature-rich emulator
- A unified environment where you can develop for all Android devices
- Apply Changes to push code and resource changes to your running app without restarting your app
- Code templates and GitHub integration to help you build common app features and import sample code
- Extensive testing tools and frameworks
- Lint tools to catch performance, usability, version compatibility, and other problems
- C++ and NDK support
- Built-in support for Google Cloud Platform, making it easy to integrate Google Cloud Messaging and App Engine

5.1.1 Project structure

Each project in Android Studio contains one or more modules with source code files and resource files. Types of modules include:

- Android app modules
- Library modules
- Google App Engine modules

By default, Android Studio displays your project files in the Android project . This view is organized by modules to provide quick access to your project's key source files.

In the Window,
All the build files are visible at the top level under **Gradle Scripts** and each app module contains the following folders:

- **manifests**: Contains the AndroidManifest.xml file.
- **java**: Contains the Java source code files, including JUnit test code.
- **res**: Contains all non-code resources, such as XML layouts, UI strings, and bitmap images.
- The **toolbar** lets you carry out a wide range of actions, including running your app and launching Android tools.
- The **navigation bar** helps you navigate through your project and open files for editing. It provides a more compact view of the structure visible in the **Project** window.
- The **editor window** is where you create and modify code. Depending on the current file type, the editor can change. For example, when viewing a layout file, the editor displays the Layout Editor.
- The **tool window bar** runs around the outside of the IDE window and contains the buttons that allow you to expand or collapse individual tool windows.
- The **tool windows** give you access to specific tasks like project management, search, version control, and more. You can expand them and collapse them.
- The **status bar** displays the status of your project and the IDE itself, as well as any warnings or messages.

Steps for developing the Android Project are as follows:

Step 1: Open your Android Studio after installing

Step 2: Click on "Start a New Android Project" and choose the activity we want the project to be. For me, I'm selecting an empty activity.

Step 3: Configure your project by changing the project's name and the desired language the user wants to code with. Android Studio supports two languages, i.e., Java and Kotlin. The user can also choose their desired API, like which android version they want to run, and then click on finish.

Step 4: Wait until your Gradle build finishes successfully and find your java and .xml file from folders of the app.

Step 5: Click onto your .xml file to see the initial layout of the file like the given image. You can drag and drop from the palette's items and change your layout by dragging and dropping.

Step 6: At last, Run your file by the run button on the top right corner in your emulator.

5.2 JAVA

JAVA was developed by James Gosling at **Sun Microsystems** Inc in 1991, later acquired by Oracle Corporation. It is a simple programming language. Java makes writing, compiling, and debugging programming easy. It helps to create reusable code and modular programs.

Java is a class-based, object-oriented programming language and is designed to have as few implementation dependencies as possible. A general-purpose programming language made for developers to *write once run anywhere* that is compiled Java code can run on all platforms that support Java. Java applications are compiled to bytecode that can run on any Java Virtual Machine. The syntax of Java is similar to c/c++.

It is used for:

- Mobile applications (specially Android apps)
- Desktop applications
- Web applications
- Web servers and application servers
- Games
- Database connection
- And much, much more!

5.2.1 Why Use Java?

- Java works on different platforms (Windows, Mac, Linux, Raspberry Pi, etc.)
- It is one of the most popular programming language in the world
- It is easy to learn and simple to use
- It is open-source and free
- It is secure, fast and powerful
- It has a huge community support (tens of millions of developers)
- Java is an object oriented language which gives a clear structure to programs and allows code to be reused, lowering development costs
- As Java is close to C++ and C#, it makes it easy for programmers to switch to Java or vice versa

5.3 MySql

5.3.1 Introduction:

MySQL is a Relational Database Management System (RDBMS). RDBMS means R--DB--MS.

R stands for Relational.

DB stands for Database.

A repository for the information store. The data in a database is organized into tables, and each table is organized into rows and columns. Each row in a table is called a record. A record may contain several pieces (called fields) of information, and each column in a table is known as a field.

MS stands for Management System, the software that allows you to insert, retrieve, modify, or delete records.

R stands for Relational, indicating a particular kind of DBMS that is good at relating information stored in one table to information stored in another table by looking for elements common to each of them. Relational DBMS has the advantage of efficient storage, and retrieval mechanisms for data, and uses a normalization process during design of RDBMS.

5.3.2 Features of MySql

Speed:

Of course, the speed at which a server side program runs depends primarily on the server hardware. Given that the server hardware is optimal, MySQL runs very fast. It supports clustered servers for demanding applications.

Ease of use:

MySQL is a high-performance, relatively simple database system. From the beginning, MySQL has typically been configured, monitored, and managed from the command line. However, several MySQL graphical interfaces are available as described below:

MySQL Administrator:

This tool makes it possible for administrators to set up, evaluate, and tune their MySQL database server. This is intended as a replacement for mysqladmin.

MySQL Query Browser:

Provides database developers and operators with a graphical database operation interface. It is especially useful for seeing multiple query plans and result sets in a single user interface.

Configuration Wizard:

Administrators can choose from a predefined list of optimal settings, or create their own.

MySQL System Tray:

Provides Windows-based administrators a single view of their MySQL instance, including the ability to start and stop their database servers.

Cost:

MySQL is available free of cost. MySQL is an "Open Source" database. MySQL is part of the LAMP (Linux, Apache, MySQL, PHP / Perl / Python) environment, a fast growing open source enterprise software stack. More and more companies are using LAMP as an alternative to expensive proprietary software stacks because of its lower cost, reliability, and documentation.

Query Language Support:

MySQL understands standards based SQL (Structured Query Language).

Capability:

Many clients can connect to the server at the same time. Clients can use multiple databases simultaneously. You can access MySQL using several interfaces such as command-line clients, Web browsers.

Connectivity and security:

MySQL is fully networked, and databases can be accessed from anywhere on the Internet, so you can share your data with anyone, anywhere. The connectivity could be achieved with Windows programs by using ODBC drivers. By using the ODBC connector to MySQL, any ODBC-aware client application (for example, Microsoft Office, report writers, Visual Basic) can connect to MySQL.

Portability:

MySQL runs on many varieties of UNIX, as well as on other non-UNIX systems, such as Windows and OS/1. MySQL runs on hardware from home PCs to high-end server. MySQL can be installed on Windows XP, Windows Server 1003, Red Hat Fedora Linux, Debian Linux, and others.

MySQL is Open Source Software:

Open Source means that it is possible for anyone to use and modify. Anybody can download MySQL from the Internet and use it without paying anything. Anybody so inclined can study the source code and change it to fit their needs. MySQL uses the GPL (GNU General Public License) to define what you may and may not do with the software in different situations. If you feel uncomfortable with the GPL or need to embed MySQL into a commercial application you can buy a commercially licensed version from us.

5.3.3 Why use MySql:

MySQL is very fast, reliable, and easy to use. MySQL also has a very practical set of features developed in very close cooperation with our users. MySQL was originally developed to handle very large databases much faster than existing solutions and has been successfully used in highly demanding production environments for several years. Though under constant development, MySQL today offers a rich and very useful set of functions. The connectivity, speed, and security make MySQL highly suited for accessing databases on the Internet.

5.3.4 The technical features of MySql

MySQL is a client/server system that consists of a multi-threaded SQL server that supports different back ends, several different client programs and libraries, administrative tools, and several programming interfaces. We also provide MySQL as a multi-threaded library which you can link into your application to get a smaller, faster, easier to manage product. MySQL has a lot of contributed software available.

6. MODULES DESCRIPTION

Modules Description

The system should be designed in such a way that only authorized people should be allowed to access some particular modules. The records should be modified by only administrators and no one else. The user should always be in control of the application and not the vice versa. The user interface should be consistent so that the user can handle the application with ease and speed. The application should be visually, conceptually clear.

6.1 Administrator module

Student Details

This module deals with the allocation of roll no and personal details for the new batch. It will generate personal details and academic details of the students.

Faculty Details

It helps to allot the subject and the subject code to the particular staff. It provides the facility to have a username and password to the staff.

Report details

Reports can be taken by daily, weekly and consolidate: Weekly report get all hour details of attendance starting date to ending date and display the status Consolidate report get all student attendance details starting date to ending date status help for the eligibility criteria of the student to attend the examination.

6.2 Faculty module

Attendance details

It assists the staff to mark attendance to the students for their subject. This will authenticate the staff before making the entry.

Report details

Weekly reports get details of attendance from starting date to ending date and display the status. Consolidated reports get all student attendance details from starting date to ending date status help for the eligibility criteria of the student to attend the examination.

6.3 Student module

Generate report; get details of attendance from starting date to ending date and display the status.

7. SOFTWARE DESIGN

7.1 Introduction

The UML stands for Unified modelling language, is a standardized general-purpose visual modelling language in the field of Software Engineering. It is used for specifying, visualizing, constructing, and documenting the primary artefacts of the software system.

UML Diagrams help in designing and characterizing, especially those software systems that incorporate the concept of Object orientation. It describes the working of both the software and hardware systems.

7.2 Types of UML Diagrams

UML Diagrams can be classified into following categories,

1. Structural Diagrams

Structure diagrams show the things in the modelled system. In a more technical term, they show different objects in a system.

- Class Diagram
- Component Diagram
- Deployment Diagram
- Object Diagram
- Package Diagram
- Profile Diagram
- Composite Structure Diagram

2. Behavioural Diagrams

Behavioural diagrams show what should happen in a system. They describe how the objects interact with each other to create a functioning system.

- Use Case Diagram
- Activity Diagram
- State Machine Diagram
- Sequence Diagram
- Communication Diagram
- Interaction Overview Diagram
- Timing Diagram

7.2.1 Class Diagram

The class diagram depicts a static view of an application. It represents the types of objects residing in the system and the relationships between them. A class consists of its objects, and also it may inherit from other classes. A class diagram is used to visualize, describe, document various different aspects of the system, and also construct executable software code.

It shows the attributes, classes, functions, and relationships to give an overview of the software system. It constitutes class names, attributes, and functions in a separate compartment that helps in software development. Since it is a collection of classes, interfaces, associations, collaborations, and constraints, it is termed as a structural diagram.

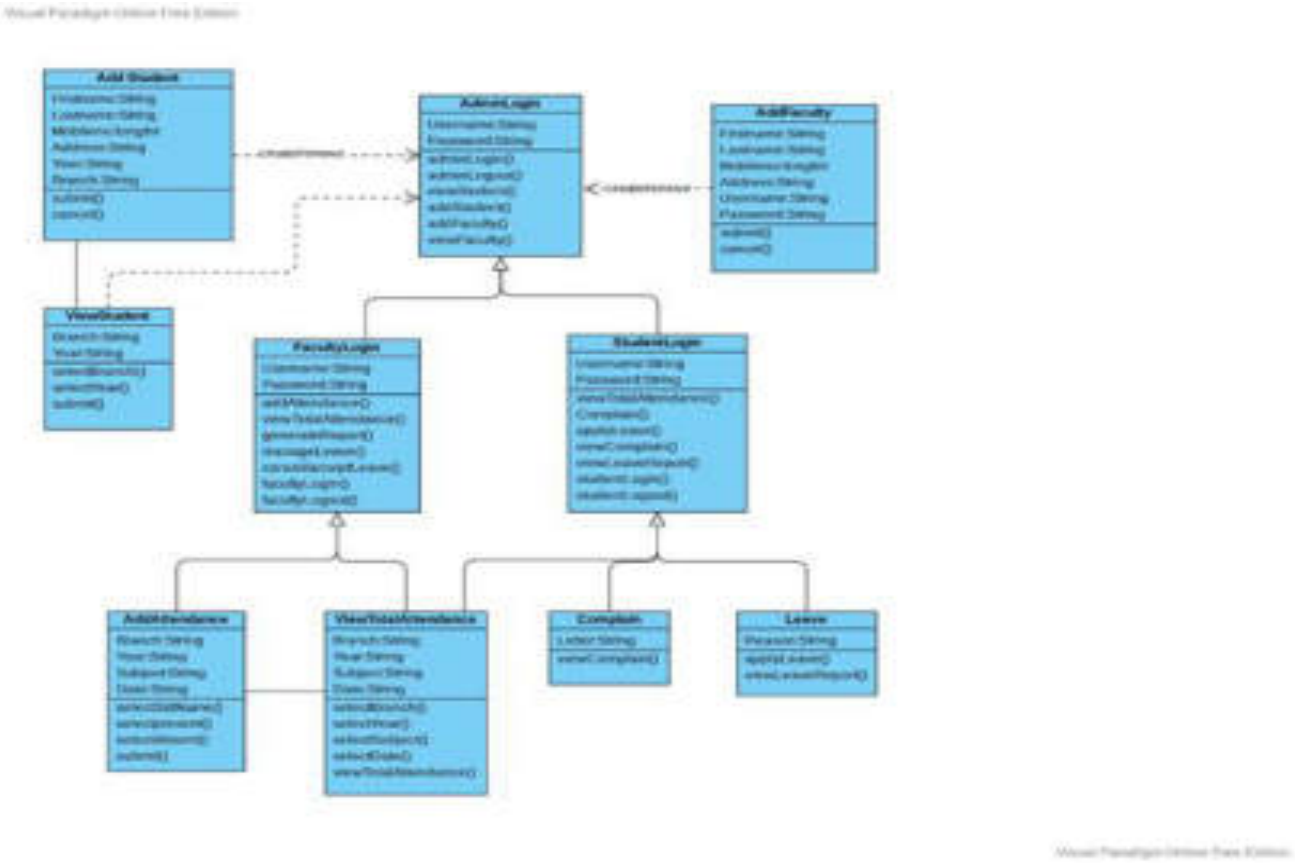


Fig 1: Class Diagram

7.2.2 Use Case Diagram

A use case diagram is used to represent the dynamic behaviour of a system. It encapsulates the system's functionality by incorporating use cases, actors, and their relationships. It models the tasks, services, and functions required by a system/subsystem of an application. It depicts the high-level functionality of a system and also tells how the user handles a system.

Following are the purposes of a use case diagram given below:

1. It gathers the system's needs.
2. It depicts the external view of the system.
3. It recognizes the internal as well as external factors that influence the system.
4. It represents the interaction between the actors.

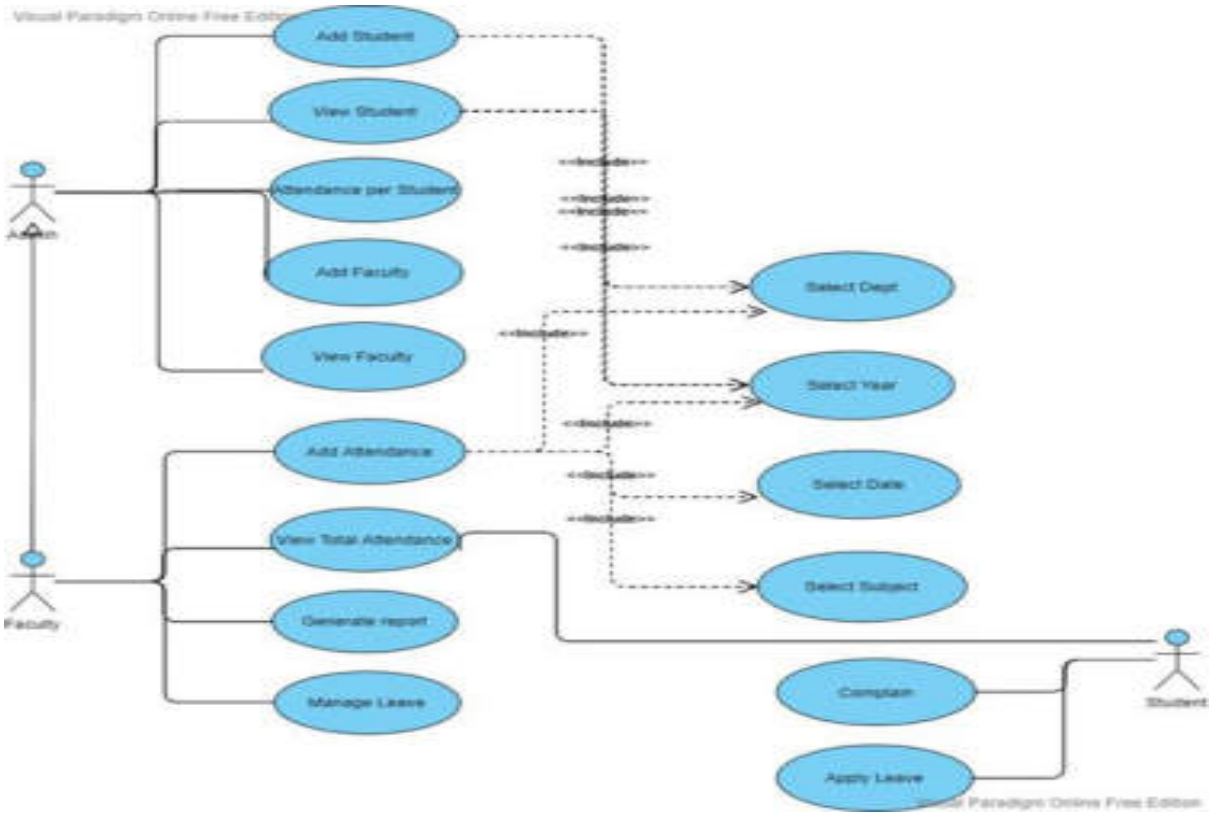


Fig 2: Use Case Diagram

7.2.3 Sequence Diagram

The sequence diagram represents the flow of messages in the system and is also termed as an event diagram. It helps in envisioning several dynamic scenarios. It portrays the communication between any two lifelines as a time-ordered sequence of events, such that these lifelines took part at the run time. In UML, the lifeline is represented by a vertical bar, whereas the message flow is represented by a vertical dotted line that extends across the bottom of the page. It incorporates the iterations as well as branching.

Purpose of a Sequence Diagram

1. To model high-level interaction among active objects within a system.
2. To model interaction among objects inside a collaboration realizing a use case.
3. It either models generic interactions or some certain instances of interaction.

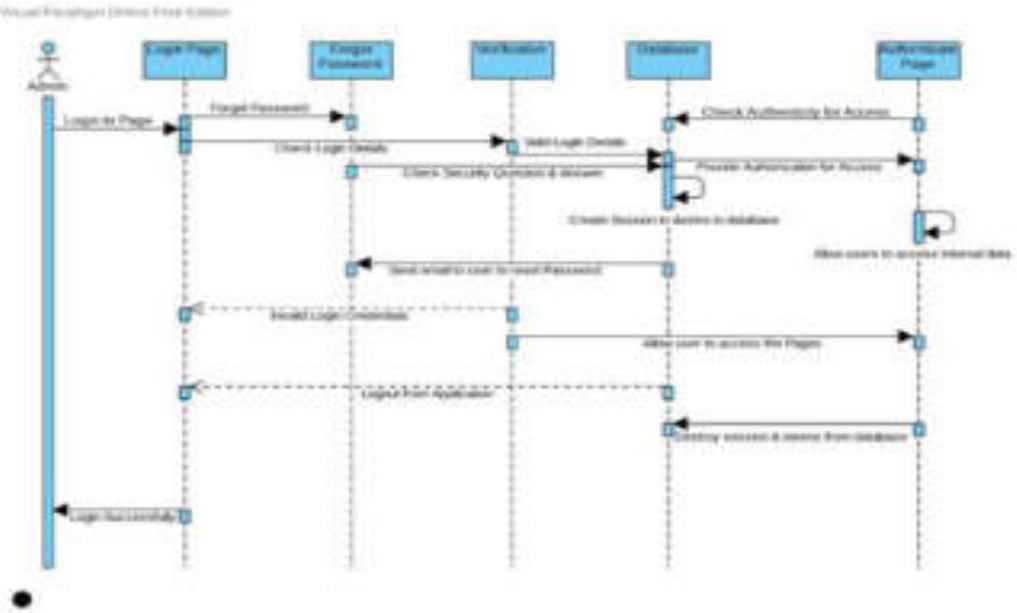


Fig 3: Sequence Diagram

7.2.4 Activity Diagram

In UML, the activity diagram is used to demonstrate the flow of control within the system rather than the implementation. It models the concurrent and sequential activities.

The activity diagram helps in envisioning the workflow from one activity to another. It puts emphasis on the condition of flow and the order in which it occurs. The flow can be sequential, branched, or concurrent, and to deal with such kinds of flows, the activity diagram has come up with a fork, join, etc.

It is also termed as an object-oriented flowchart. It encompasses activities composed of a set of actions or operations that are applied to model the behavioural diagram.

Admin Activity Diagram

Admin Activity Diagram represents Granting a user admin privileges allows them to edit the page, and to add other users as editors and admins.

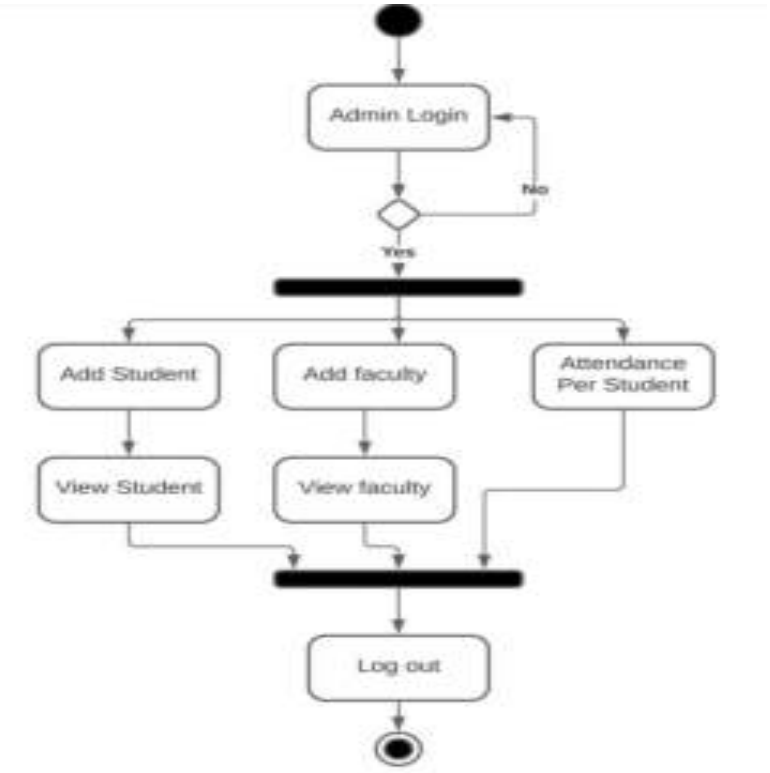


Fig 4: Admin Activity Diagram

Faculty Activity Diagram

The following diagram represents the Activity UML Diagram of faculty management system which shows the flows between the activity of teaching, attendance, report, leave.

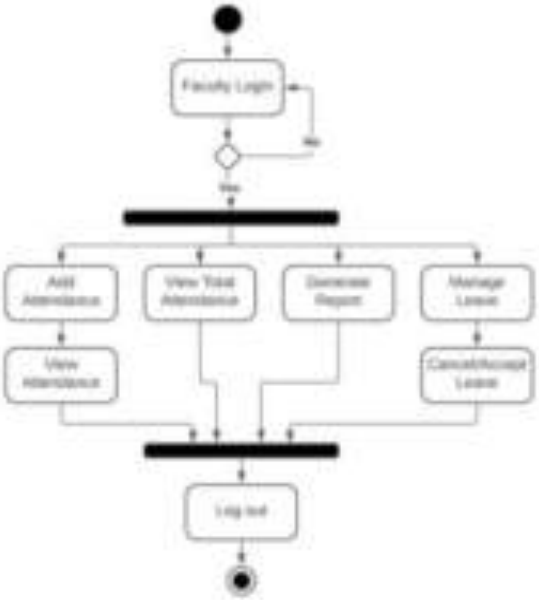


Fig 5: Faculty Activity Diagram

Student Activity Diagram

The following diagram represents the Activity UML Diagram of student which shows the operations performed that are view attendance, applying for leave and they can view reports ete.

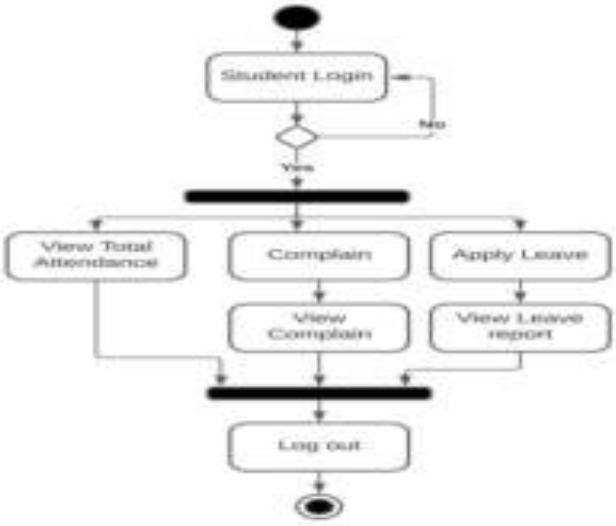


Fig 6: Student Activity Diagram

7.2.5 Component Diagram

A component diagram is used to break down a large object-oriented system into the smaller components, so as to make them more manageable. It models the physical view of a system such as executables, files, libraries, etc. that resides within the node.

It visualizes the relationships as well as the organization between the components present in the system. It helps in forming an executable system. A component is a single unit of the system, which is replaceable and executable. The implementation details of a component are hidden, and it necessitates an interface to execute a function. It is like a black box whose behaviour is explained by the provided and required interfaces.

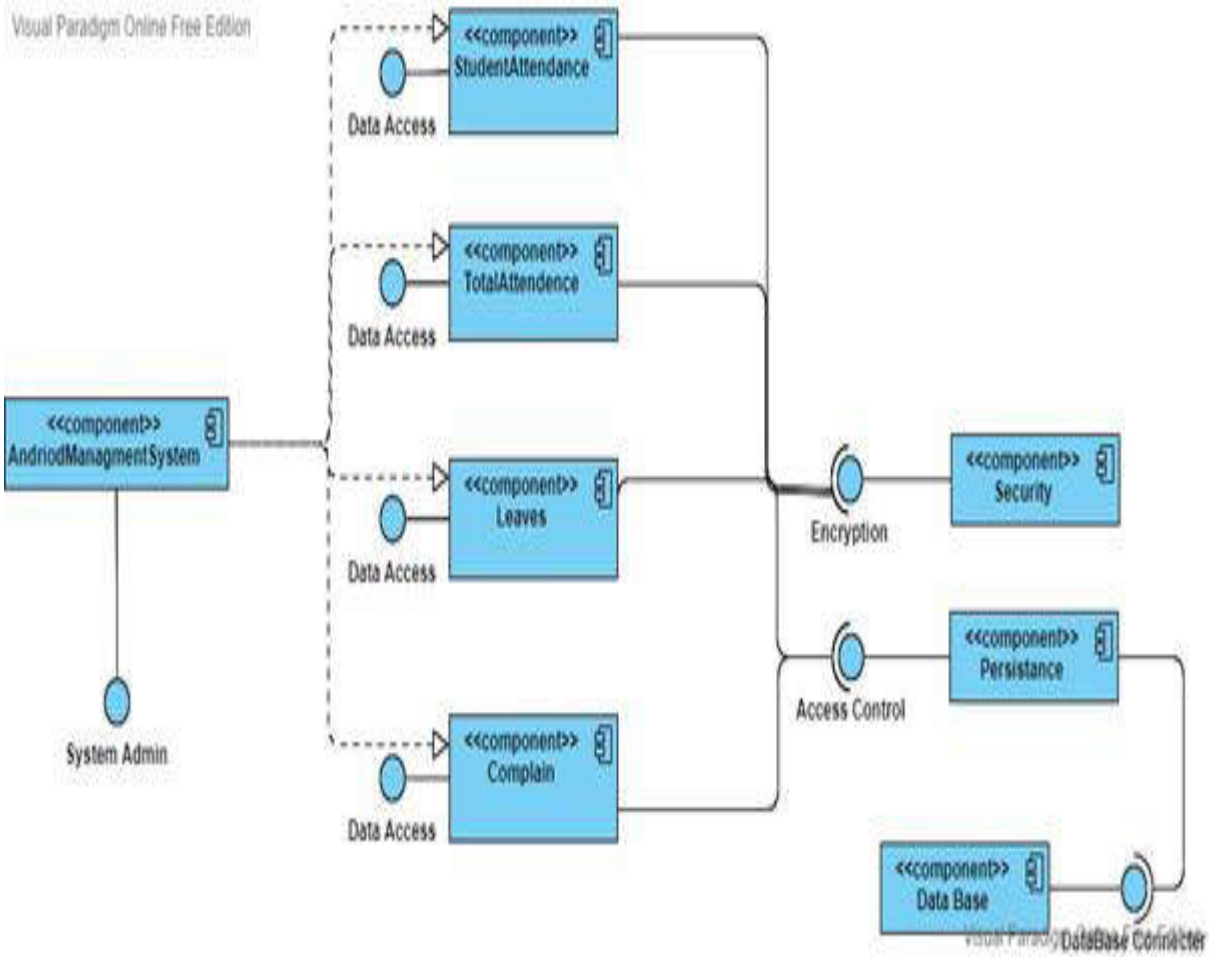


Fig 7: Component Diagram

7.2.6 Deployment Diagram

The deployment diagram visualizes the physical hardware on which the software will be deployed. It portrays the static deployment view of a system. It involves the nodes and their relationships.

It ascertains how software is deployed on the hardware. It maps the software architecture created in design to the physical system architecture, where the software will be executed as a node. Since it involves many nodes, the relationship is shown by utilizing communication paths.

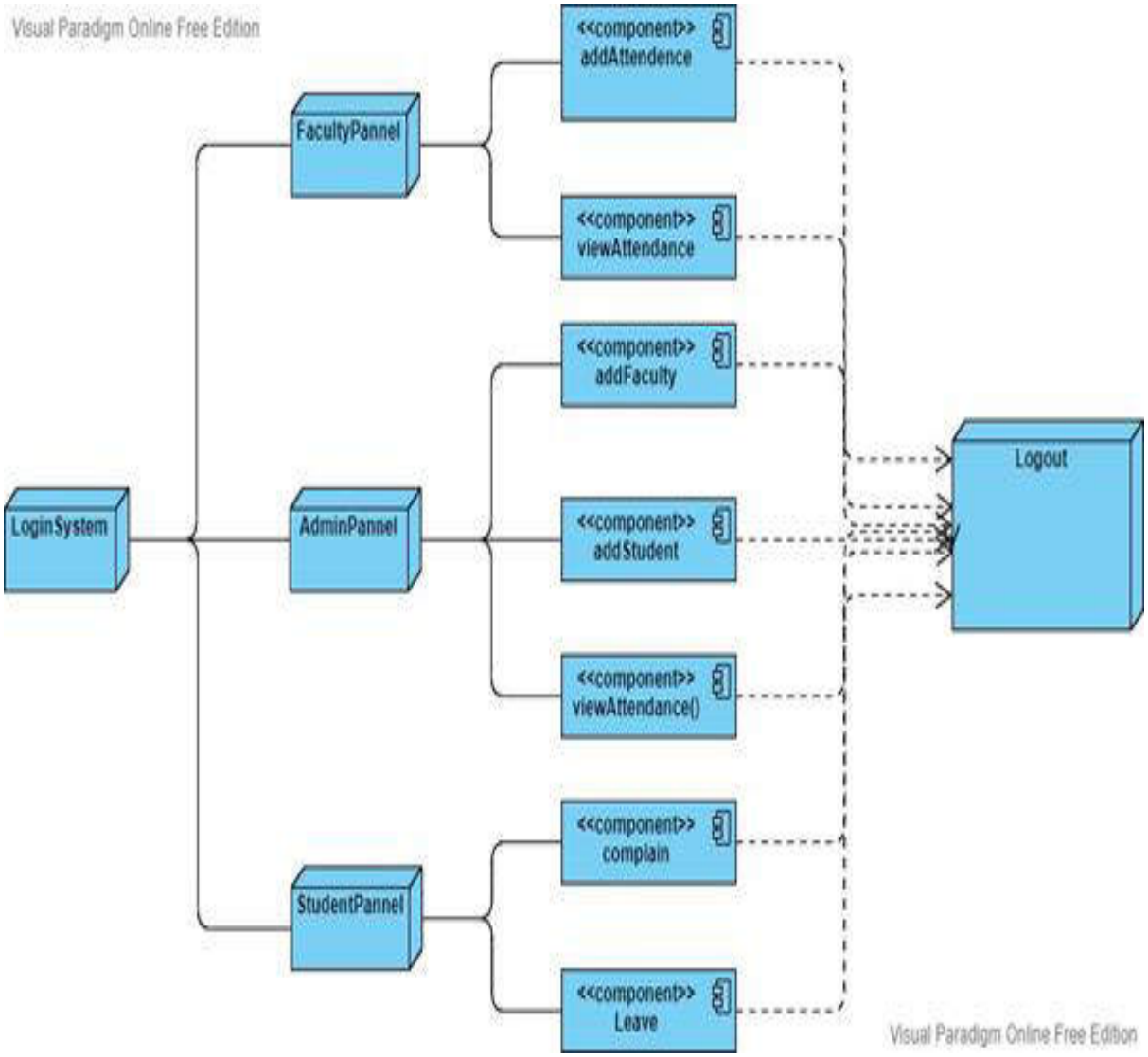


Fig 8: Deployment Diagram

7.2.7 Data Flow Diagram

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It can be manual, automated, or a combination of both.

It shows how data enters and leaves the system, what changes the information, and where data is stored.

7.2.7-1: 0-level DFD

It is also known as the fundamental system model, or context diagram that represents the entire software requirement as a single bubble with input and output data denoted by incoming and outgoing arrows. Then the system is decomposed and described as a DFD with multiple bubbles.

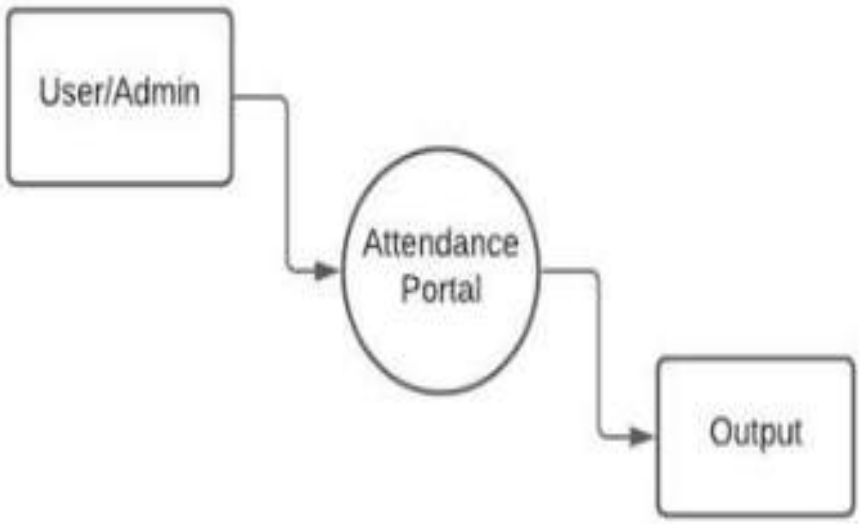


Fig 9: 0-level DFD

7.2.7-2: 1-level DFD

In 1-level DFD, a context diagram is decomposed into multiple bubbles/processes. In this level, we highlight the main objectives of the system and break down the high-level process of 0-level DFD into subprocesses.

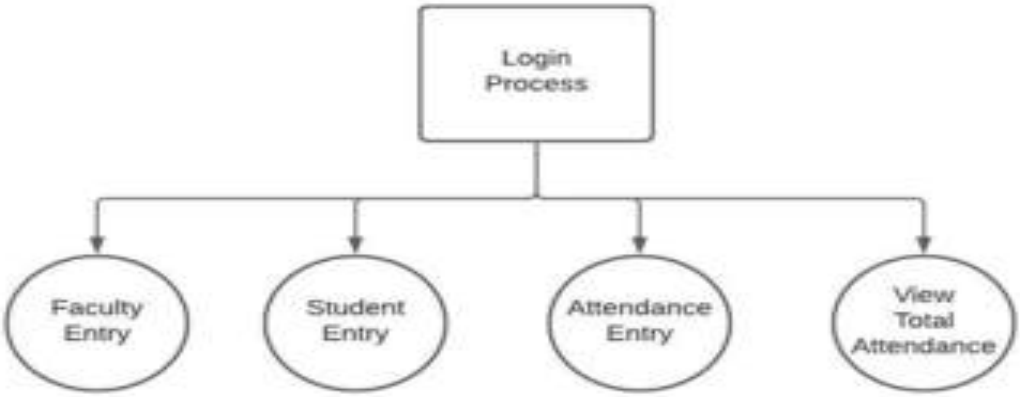


Fig 10: 1-level DFD

7.2.7-3: 2-level DFD

2-level DFD goes one process deeper into parts of 1-level DFD. It can be used to project or record the specific/necessary detail about the system's functioning.

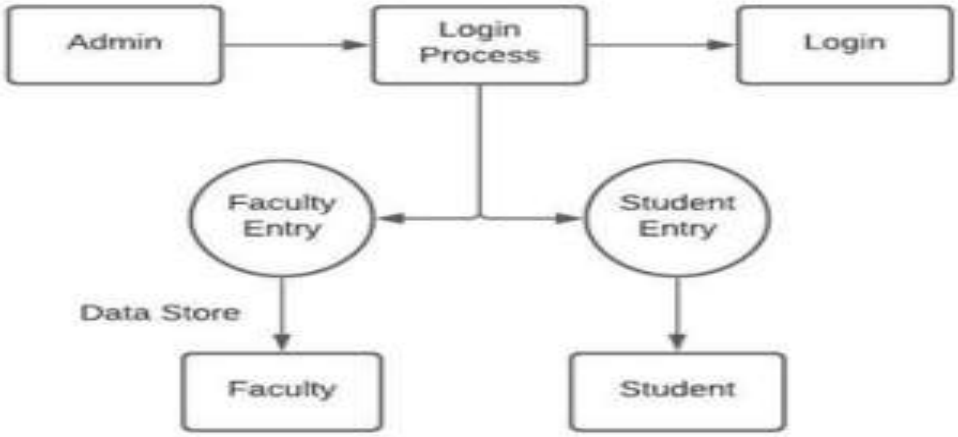


Fig 11: 2-level DFD

7.2.7-4: 3-level DFD

The level 3 DFD is the further decomposition of level 2 processes into sub-processes which again give the detailed description of the data flow in each of the processes shown below.

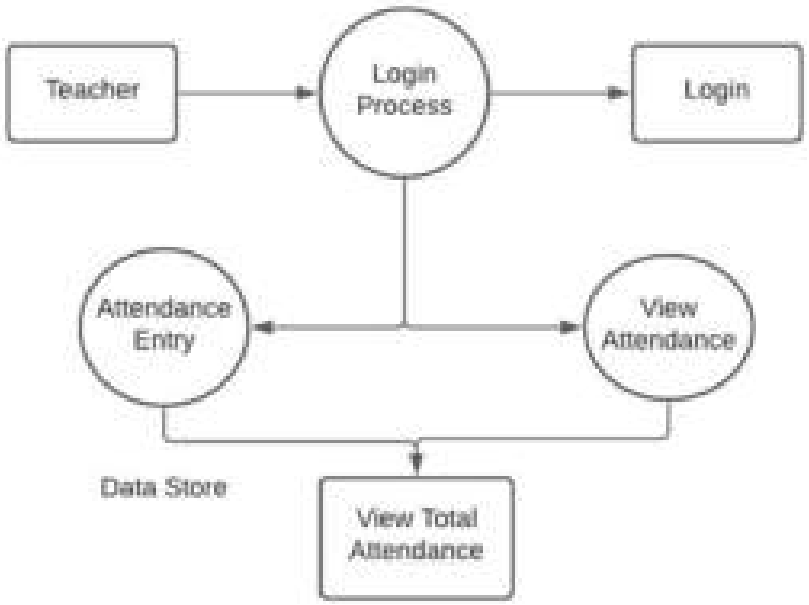


Fig 12: 3-level DFD

7.8 Input Design

Input design is part of overall system design that requires special attention designing input data is to make the data entered easy and free from errors. The input forms are designed using the controls available in Android Studio.

Input design is the process of converting the user originated inputs to a computer based format. A system user interacting through a workstation must be able to tell the system whether to accept the input to produce reports. The collection of input data is considered to be the most expensive part of the system design. Since the input has to be planned in such a manner so as to get relevant information, extreme care is taken to obtain pertinent information .

This project first will enter the input of allocation forms. It will be created on student details form and subject entry form, time table form .It will help to calculate subject wise attendance system.

7.9 Output Design

Output design of this application “Student Attendance management system” generally refers to the results and information that are generated by the system for many end-users; output is the main reason for developing the system and the basis on which they evaluate the usefulness of the application.

The output is designed in such a way that it is attractive, convenient and informative. Forms are designed with various features, which make the console output more pleasing.

As the outputs are the most important sources of information to the users, better design should improve the system’s relationships with us and also will help in decision making. Form design elaborates the way output is presented and the layout available for capturing information.

One of the most important factors of the system is the output it produces. This system refers to the results and information generated. Basically the output from a computer system is used to communicate the result of processing to the user.

Attendance management system to show the report subject wise attendance maintained by staff. Taken as a whole report obtained on administrator privileges only.

8. SOURCE CODE

activity_main.xml

```
<RelativeLayout xmlns:android="http://schemas.android.com/apk/res/android"
        xmlns:tools="http://schemas.android.com/tools"
        android:layout_width="match_parent"
        android:layout_height="match_parent"
        android:paddingBottom="@dimen/activity_vertical_margin"
        android:paddingLeft="@dimen/activity_horizontal_margin"
        android:paddingRight="@dimen/activity_horizontal_margin"
        android:paddingTop="@dimen/activity_vertical_margin"
        android:background="@drawable/clg"
        tools:context=".MainActivity" >
        <TextView
                android:id="@+id/textView1"
                android:layout_width="395dp"
                android:layout_height="wrap_content"
                android:layout_alignParentLeft="true"
                android:layout_alignParentTop="true"
                android:layout_marginLeft="28dp"
                android:layout_marginTop="208dp"
                android:text="ATTENDANCE SYSTEM"
                 android:textAlignment="center"
                 android:textAppearance="?android:attr/textAppearanceLarge"
                 android:textColor="@android:color/holo_red_dark"
                 android:textSize="25dp"
                 android:textStyle="bold" />
        <Button
                android:id="@+id/buttonstart"
                android:layout_width="match_parent"
                android:layout_height="wrap_content"
                android:layout_centerHorizontal="true"
                android:layout_centerVertical="true"
```

```
            android:background="@drawable/roundedbutton"
            android:text="Start" />
        <TextView
            android:id="@+id/textView"
            android:layout_width="174dp"
            android:layout_height="wrap_content"
            android:layout_alignParentLeft="true"
            android:layout_alignParentTop="true"
            android:layout_marginLeft="115dp"
            android:layout_marginTop="157dp"
            android:text="WELCOME TO"
            android:textAlignment="center"
            android:textAppearance="?android:attr/textAppearanceLarge"
            android:textColor="@android:color/holo_red_dark"
            android:textSize="25dp"
            android:textStyle="bold" />
</RelativeLayout>
```

login.xml

```
<RelativeLayout xmlns:android="http://schemas.android.com/apk/res/android"
        xmlns:tools="http://schemas.android.com/tools"
        android:layout_width="match_parent"
        android:layout_height="match_parent"
        android:paddingBottom="@dimen/activity_vertical_margin"
        android:paddingLeft="@dimen/activity_horizontal_margin"
        android:paddingRight="@dimen/activity_horizontal_margin"
        android:paddingTop="@dimen/activity_vertical_margin"
        android:background="@drawable/clg"
        tools:context=".MainActivity" >
        <TextView
            android:id="@+id/textView1"
            android:layout_width="wrap_content"
            android:layout_height="wrap_content"
            android:layout_alignParentLeft="true"
```

```
        android:layout_alignParentTop="true"
        android:layout_marginLeft="82dp"
        android:layout_marginTop="43dp"
        android:text="Login here.."
        android:textAppearance="?android:attr/textAppearanceLarge" />
    <EditText
        android:id="@+id/editTextpassword"
        android:layout_width="229dp"
        android:layout_height="44dp"
        android:layout_below="@+id/textView3"
        android:layout_alignLeft="@+id/textView3"
        android:layout_marginLeft="0dp"
        android:layout_marginTop="17dp"
        android:background="@drawable/roundedtextview"
        android:ems="10"
        android:inputType="textPassword" />
    <TextView
        android:id="@+id/textView3"
        android:layout_width="115dp"
        android:layout_height="36dp"
        android:layout_below="@+id/editTextusername"
        android:layout_alignLeft="@+id/editTextusername"
        android:layout_marginLeft="4dp"
        android:layout_marginTop="14dp"
        android:text="Password"
        android:textAppearance="?android:attr/textAppearanceMedium" />
    <TextView
        android:id="@+id/textView2"
        android:layout_width="108dp"
        android:layout_height="46dp"
        android:layout_below="@+id/textView1"
        android:layout_alignLeft="@+id/editTextusername"
        android:layout_marginLeft="3dp"
        android:layout_marginTop="131dp"
```

```
            android:text="Username"
            android:textAppearance="?android:attr/textAppearanceMedium" />
        <EditText
            android:id="@+id/editTextusername"
            android:layout_width="230dp"
            android:layout_height="50dp"
            android:layout_below="@+id/textView2"
            android:layout_marginTop="-6dp"
            android:layout_marginLeft="70dp"
            android:background="@drawable/roundedtextview"
            android:ems="10">
            <requestFocus />
        </EditText>
        <Spinner
            android:id="@+id/spinnerloginas"
            android:layout_width="wrap_content"
            android:layout_height="wrap_content"
            android:layout_alignLeft="@+id/textView2"
            android:layout_alignBottom="@+id/textView2"
            android:layout_alignParentRight="true"
            android:layout_marginLeft="1dp"
            android:layout_marginRight="49dp"
            android:layout_marginBottom="77dp" />
         <Button
            android:id="@+id/buttonlogin"
            android:layout_width="match_parent"
            android:layout_height="wrap_content"
            android:layout_alignRight="@+id/spinnerloginas"
            android:layout_alignParentBottom="true"
            android:layout_marginBottom="50dp"
            android:layout_marginLeft="50dp"
            android:background="@drawable/roundedbutton"
            android:text="Login" />
</RelativeLayout>
```

menu.xml

```
<RelativeLayout xmlns:android="http://schemas.android.com/apk/res/android"
        xmlns:tools="http://schemas.android.com/tools"
        android:layout_width="match_parent"
        android:layout_height="match_parent"
        android:background="@drawable/clg"
        android:paddingBottom="@dimen/activity_vertical_margin"
        android:paddingLeft="@dimen/activity_horizontal_margin"
        android:paddingRight="@dimen/activity_horizontal_margin"
        android:paddingTop="@dimen/activity_vertical_margin"
        tools:context=".MainActivity" >
         <Button
                android:id="@+id/buttonViewstudent"
                 android:layout_width="match_parent"
                android:layout_height="wrap_content"
                android:layout_below="@+id/buttonaddstudent"
                android:background="@drawable/roundedbutton"
                android:layout_marginTop="18dp"
                android:text="View student" />
        <Button
                android:id="@+id/buttonlogout"
                android:layout_width="match_parent"
                android:layout_height="wrap_content"
                android:layout_alignLeft="@+id/buttonviewfaculty"
                android:layout_alignParentBottom="true"
                android:background="@drawable/roundedbutton"
                android:text="Logout" />
        <Button
                android:id="@+id/buttonviewfaculty"
                android:layout_width="match_parent"
                android:layout_height="wrap_content"
                android:layout_above="@+id/buttonlogout"
                android:layout_alignLeft="@+id/buttonaddfaculty"
```

```
            android:layout_marginBottom="16dp"
            android:background="@drawable/roundedbutton"
            android:text="View faculty" />
        <Button
            android:id="@+id/buttonaddfaculty"
            android:layout_width="match_parent"
            android:layout_height="wrap_content"
            android:layout_above="@+id/buttonviewfaculty"
            android:layout_alignLeft="@+id/buttonViewstudent"
            android:layout_marginBottom="14dp"
            android:background="@drawable/roundedbutton"
            android:text="Add faculty" />
        <Button
            android:id="@+id/buttonaddstudent"
            android:layout_width="match_parent"
            android:layout_height="wrap_content"
            android:layout_alignLeft="@+id/buttonViewstudent"
            android:layout_alignParentTop="true"
            android:layout_marginTop="22dp"
            android:background="@drawable/roundedbutton"
            android:text="Add student" />
        <Button
            android:id="@+id/attendancePerStudentButton"
            android:layout_width="match_parent"
            android:layout_height="wrap_content"
            android:layout_below="@+id/buttonViewstudent"
            android:layout_centerHorizontal="true"
            android:layout_marginTop="29dp"
            android:background="@drawable/roundedbutton"
            android:text="Attendance Of Student" />
</RelativeLayout>
```

studentmenu.xml

```
<?xml version="1.0" encoding="utf-8"?>
<LinearLayout xmlns:android="http://schemas.android.com/apk/res/android"
    xmlns:tools="http://schemas.android.com/tools"
    android:orientation="vertical"
    android:layout_width="match_parent"
    android:layout_height="match_parent"
    android:background="@drawable/clg"
    tools:context=".MainActivity">
    <Button
        android:id="@+id/buttoncomplain"
        android:layout_width="187dp"
        android:layout_height="37dp"
        android:layout_marginTop="200dp"
        android:layout_marginLeft="100dp"
        android:text="Complain"
        android:background="@drawable/roundedbutton"/>
    <Button
        android:id="@+id/buttonleave"
        android:layout_width="187dp"
        android:layout_height="37dp"
        android:layout_marginTop="30dp"
        android:layout_marginLeft="100dp"
        android:text="Leave"
        android:background="@drawable/roundedbutton" />
    <Button
        android:id="@+id/buttonAttendance"
        android:layout_width="187dp"
        android:layout_height="37dp"
        android:layout_marginTop="30dp"
        android:layout_marginLeft="100dp"
        android:text="View Attendance"
        android:background="@drawable/roundedbutton"/>
```

```
        <Button
                android:id="@+id/buttonlogout"
                android:layout_width="187dp"
                android:layout_height="37dp"
                android:layout_marginLeft="100dp"
                android:layout_marginTop="30dp"
                android:text="Logout"
                android:background="@drawable/roundedbutton"/>
</LinearLayout>
```

LoginActivity.java

```
package com.android.attendance.activity;
import android.app.Activity;
import android.content.Intent;
import android.graphics.Color;
import android.os.Bundle;
import android.text.TextUtils;
import android.view.Menu;
import android.view.View;
import android.view.View.OnClickListener;
import android.widget.AdapterView;
import android.widget.AdapterView.OnItemSelectedListener;
import android.widget.ArrayAdapter;
import android.widget.Button;
import android.widget.EditText;
import android.widget.Spinner;
import android.widget.TextView;
import android.widget.Toast;
import com.android.attendance.bean.FacultyBean;
import com.android.attendance.bean.StudentBean;
import com.android.attendance.context.ApplicationContext;
import com.android.attendance.db.DBAdapter;
import com.example.androidattendancesystem.R;
```

```
public class LoginActivity extends Activity {
	Button login;
	EditText username,password;
	Spinner spinnerloginas;
	String userrole;
	private String[] userRoleString = new String[] { "admin", "faculty", "student"};
	@Override
	protected void onCreate(Bundle savedInstanceState) {
		super.onCreate(savedInstanceState);
		setContentView(R.layout.login);
		login =(Button)findViewById(R.id.buttonlogin);
		username=(EditText)findViewById(R.id.editTextusername);
		password=(EditText)findViewById(R.id.editTextpassword);
		spinnerloginas=(Spinner)findViewById(R.id.spinnerloginas);
		spinnerloginas.setOnItemSelectedListener(new OnItemSelectedListener() {
		@Override
		public void onItemSelected(AdapterView<?> arg0, View view, int arg2, long
arg3) {
		// TODO Auto-generated method stub
		((TextView) arg0.getChildAt(0)).setTextColor(Color.WHITE);
		userrole =(String) spinnerloginas.getSelectedItem();

	}
@Override
	public void onNothingSelected(AdapterView<?> arg0) {
	  // TODO Auto-generated method stub
	}
 });
ArrayAdapter<String> adapter_role = new ArrayAdapter<String>(this,
android.R.layout.simple_spinner_item, userRoleString);
	adapter_role
	.setDropDownViewResource(android.R.layout.simple_spinner_dropdown_item);
	spinnerloginas.setAdapter(adapter_role);
	login.setOnClickListener(new OnClickListener() {
```

```
        @Override
        public void onClick(View v) {
       // TODO Auto-generated method stub
       if(userrole.equals("admin"))
       {
         String user_name = username.getText().toString();
         String pass_word = password.getText().toString();
         if (TextUtils.isEmpty(user_name))
         {
           username.setError("Invalid User Name");
         }
         else if(TextUtils.isEmpty(pass_word))
         {
           password.setError("enter password");
         }
         else
         {
           if(user_name.equals("admin") & pass_word.equals("admin123")){
           Intent intent =new Intent(LoginActivity.this,MenuActivity.class);
           startActivity(intent);
           Toast.makeText(getApplicationContext(), "Login successful",
Toast.LENGTH_SHORT).show();
           }
        else{
             Toast.makeText(getApplicationContext(), "Login failed",
Toast.LENGTH_SHORT).show();
           }
         }
    }
    else if(userrole.equals("faculty"))
    {
        String user_name = username.getText().toString();
        String pass_word = password.getText().toString();
        if (TextUtils.isEmpty(user_name))
```

```
            {
              username.setError("Invalid User Name");
            }
            else if(TextUtils.isEmpty(pass_word))
            {
              password.setError("enter password");
            }
            DBAdapter dbAdapter = new DBAdapter(LoginActivity.this);
            FacultyBean facultyBean = dbAdapter.validateFaculty(user_name, pass_word);
             if(facultyBean!=null)
            {
              Intent intent = new
Intent(LoginActivity.this,AddAttandanceSessionActivity.class);
              startActivity(intent);
             ((ApplicationContext)LoginActivity.this.getApplicationContext()).setFacultyBean(f
acultyBean);
              Toast.makeText(getApplicationContext(), "Login successful",
Toast.LENGTH_SHORT).show();
            }
            else
            {
              Toast.makeText(getApplicationContext(), "Login failed",
Toast.LENGTH_SHORT).show();
            }
        }
         else
         {
            String user_name = username.getText().toString();
            String pass_word = password.getText().toString();
           if (TextUtils.isEmpty(user_name))
            {
              username.setError("Invalid User Name");
            }
            else if(TextUtils.isEmpty(pass_word))
```

```
            {
              password.setError("enter password");
            }
            DBAdapter dbAdapter = new DBAdapter(LoginActivity.this);
            StudentBean studentBean = dbAdapter.validateStudent(user_name, pass_word);
            if(studentBean!=null)
            {

                    Intent  inte = new Intent(LoginActivity.this,StudentMenuActivity.class);

                  startActivity(inte);
                      ((ApplicationContext)LoginActivity.this.getApplicationContext()).setStude
ntBean(studentBean);
              Toast.makeText(getApplicationContext(), "Login successful",
Toast.LENGTH_SHORT).show();
            }
           else          {
              Toast.makeText(getApplicationContext(), "Login failed",
Toast.LENGTH_SHORT).show();
            }
          }
     }
    });
}
 @Override
 public boolean onCreateOptionsMenu(Menu menu) {
    // Inflate the menu; this adds items to the action bar if it is present.
    getMenuInflater().inflate(R.menu.main, menu);
    return true;
 }
}
```

AddAttandanceSessionActivity.java

```
package com.android.attendance.activity;
import java.util.ArrayList;
import java.util.Calendar;
import com.android.attendance.bean.AttendanceBean;
import com.android.attendance.bean.AttendanceSessionBean;
import com.android.attendance.bean.FacultyBean;
import com.android.attendance.bean.StudentBean;
import com.android.attendance.context.ApplicationContext;
import com.android.attendance.db.DBAdapter;
import com.example.androidattendancesystem.R;
import android.app.Activity;
import android.app.DatePickerDialog;
import android.app.Dialog;
import android.content.Intent;
import android.graphics.Color;
import android.os.Bundle;
import android.text.TextUtils;
import android.view.Menu;
import android.view.View;
import android.view.View.OnClickListener;
import android.widget.AdapterView;
import android.widget.AdapterView.OnItemSelectedListener;
import android.widget.ArrayAdapter;
import android.widget.Button;
import android.widget.DatePicker;
import android.widget.EditText;
import android.widget.ImageButton;
import android.widget.Spinner;
import android.widget.TextView;
import android.widget.Toast;
public class AddAttandanceSessionActivity<AddAttandanceActivity> extends Activity {
private ImageButton date;
```

```
private Calendar cal;
private int day;
private int month;
private int dyear;
private EditText dateEditText;
Button submit;
Button viewAttendance;
Button viewTotalAttendance;
Button logout;
Spinner spinnerbranch,spinneryear,spinnerSubject;
String branch = "cse";
String year = "SE";
String subject = "SC";
private String[] branchString = new String[] { "cse","ece","eee","civil","mech","it"};
private String[] yearString = new String[] {"SE","TE","BE","FE"};
private String[] subjectSEString = new String[] {"SC","MC"};
private String[] subjectTEString = new String[] {"GT","CN"};
private String[] subjectBEString = new String[] {"DS","NS"};
private String[] subjectFinal = new String[] {"M3","DS","M4","CN","M5","NS"};
AttendanceSessionBean attendanceSessionBean;
@Override
  protected void onCreate(Bundle savedInstanceState) {
        super.onCreate(savedInstanceState);
        setContentView(R.layout.add_attandance);
        logout =(Button)findViewById(R.id.buttonlogout);
        spinnerbranch=(Spinner)findViewById(R.id.spinner1);
        spinneryear=(Spinner)findViewById(R.id.spinneryear);
        spinnerSubject=(Spinner)findViewById(R.id.spinnerSE);
        logout.setOnClickListener(new OnClickListener() {
        @Override
         public void onClick(View view) {
        Intent inte=new Intent(getBaseContext(),LoginActivity.class);
        startActivity(inte);
      }
```

```
    });
    ArrayAdapter<String> adapter_branch = new
ArrayAdapter<String>(this,android.R.layout.simple_spinner_item, branchString);
    adapter_branch.setDropDownViewResource(android.R.layout.simple_spinner_dropdown
_item);
    spinnerbranch.setAdapter(adapter_branch);
    spinnerbranch.setOnItemSelectedListener(new OnItemSelectedListener() {
      @Override
      public void onItemSelected(AdapterView<?> arg0, View view,
          int arg2, long arg3) {
        // TODO Auto-generated method stub
        ((TextView) arg0.getChildAt(0)).setTextColor(Color.WHITE);
        branch =(String) spinnerbranch.getSelectedItem();
      }
     @Override
      public void onNothingSelected(AdapterView<?> arg0) {
        // TODO Auto-generated method stub
      }
    });
    ///.....................spinner2
    ArrayAdapter<String> adapter_year = new ArrayAdapter<String>(this,
android.R.layout.simple_spinner_item, yearString);
    adapter_year.setDropDownViewResource(android.R.layout.simple_spinner_dropdown_it
em);
    spinneryear.setAdapter(adapter_year);
    spinneryear.setOnItemSelectedListener(new OnItemSelectedListener() {
      @Override
      public void onItemSelected(AdapterView<?> arg0, View view,
          int arg2, long arg3) {
        // TODO Auto-generated method stub
        ((TextView) arg0.getChildAt(0)).setTextColor(Color.WHITE);
        year =(String) spinneryear.getSelectedItem();
        Toast.makeText(getApplicationContext(), "year:"+year,
Toast.LENGTH_SHORT).show();
```

```
        }
        @Override
        public void onNothingSelected(AdapterView<?> arg0) {
          // TODO Auto-generated method stub
        }
      });

      ArrayAdapter<String> adapter_subject = new ArrayAdapter<String>(this,
android.R.layout.simple_spinner_item, subjectFinal);
     adapter_subject.setDropDownViewResource(android.R.layout.simple_spinner_dropdown_
item);
      spinnerSubject.setAdapter(adapter_subject);
      spinnerSubject.setOnItemSelectedListener(new OnItemSelectedListener() {
        @Override
        public void onItemSelected(AdapterView<?> arg0, View view,
            int arg2, long arg3) {
          // TODO Auto-generated method stub
          ((TextView) arg0.getChildAt(0)).setTextColor(Color.WHITE);
          subject =(String) spinnerSubject.getSelectedItem();
        }
        @Override
        public void onNothingSelected(AdapterView<?> arg0) {
          // TODO Auto-generated method stub
        }
      });
      date = (ImageButton) findViewById(R.id.DateImageButton);
      cal = Calendar.getInstance();
      day = cal.get(Calendar.DAY_OF_MONTH);
      month = cal.get(Calendar.MONTH);
      dyear = cal.get(Calendar.YEAR);
      dateEditText = (EditText) findViewById(R.id.DateEditText);
      date.setOnClickListener(new OnClickListener() {
        @Override
          public void onClick(View arg0) {
```

```
        showDialog(0);
      }
    });
    submit=(Button)findViewById(R.id.buttonsubmit);
    submit.setOnClickListener(new OnClickListener() {
      @Override
      public void onClick(View arg0) {
        AttendanceSessionBean attendanceSessionBean = new AttendanceSessionBean();
        FacultyBean
bean=((ApplicationContext)AddAttandanceSessionActivity.this.getApplicationContext()).get
FacultyBean();
        attendanceSessionBean.setAttendance_session_faculty_id(bean.getFaculty_id());
        attendanceSessionBean.setAttendance_session_department(branch);
        attendanceSessionBean.setAttendance_session_class(year);
                attendanceSessionBean.setAttendance_session_date(dateEditText.getText().to
String());
        attendanceSessionBean.setAttendance_session_subject(subject);
        DBAdapter dbAdapter = new DBAdapter(AddAttandanceSessionActivity.this);
        int sessionId= dbAdapter.addAttendanceSession(attendanceSessionBean);
        ArrayList<StudentBean>
studentBeanList=dbAdapter.getAllStudentByBranchYear(branch, year);
       ((ApplicationContext)AddAttandanceSessionActivity.this.getApplicationContext()).set
StudentBeanList(studentBeanList);
        Intent intent = new
Intent(AddAttandanceSessionActivity.this,AddAttendanceActivity.class);
        intent.putExtra("sessionId", sessionId);
        startActivity(intent);
      }
    });
      viewAttendance=(Button)findViewById(R.id.viewAttendancebutton);
      viewAttendance.setOnClickListener(new OnClickListener() {
        @Override
        public void onClick(View arg0) {
        AttendanceSessionBean attendanceSessionBean = new AttendanceSessionBean();
```

```
        FacultyBean
bean=((ApplicationContext)AddAttandanceSessionActivity.this.getApplicationContext()).get
FacultyBean();
        attendanceSessionBean.setAttendance_session_faculty_id(bean.getFaculty_id());
        attendanceSessionBean.setAttendance_session_department(branch);
        attendanceSessionBean.setAttendance_session_class(year);
       attendanceSessionBean.setAttendance_session_date(dateEditText.getText().toString());
        attendanceSessionBean.setAttendance_session_subject(subject);
        DBAdapter dbAdapter = new DBAdapter(AddAttandanceSessionActivity.this);
                ArrayList<AttendanceBean> attendanceBeanList =
dbAdapter.getAttendanceBySessionID(attendanceSessionBean);
        ((ApplicationContext)AddAttandanceSessionActivity.this.getApplicationContext()).set
AttendanceBeanList(attendanceBeanList);

        Intent intent = new
Intent(AddAttandanceSessionActivity.this,ViewAttendanceByFacultyActivity.class);
        startActivity(intent);
             }
    });
         viewTotalAttendance=(Button)findViewById(R.id.viewTotalAttendanceButton);
         viewTotalAttendance.setOnClickListener(new OnClickListener() {
           @Override
            public void onClick(View arg0) {
        AttendanceSessionBean attendanceSessionBean = new AttendanceSessionBean();
        FacultyBean
bean=((ApplicationContext)AddAttandanceSessionActivity.this.getApplicationContext()).get
FacultyBean();
        attendanceSessionBean.setAttendance_session_faculty_id(bean.getFaculty_id());
        attendanceSessionBean.setAttendance_session_department(branch);
        attendanceSessionBean.setAttendance_session_class(year);
        attendanceSessionBean.setAttendance_session_subject(subject);
        DBAdapter dbAdapter = new DBAdapter(AddAttandanceSessionActivity.this);
                ArrayList<AttendanceBean> attendanceBeanList =
dbAdapter.getTotalAttendanceBySessionID(attendanceSessionBean);
```

```
        ((ApplicationContext)AddAttandanceSessionActivity.this.getApplicationContext()).set
AttendanceBeanList(attendanceBeanList);
                    Intent intent = new
Intent(AddAttandanceSessionActivity.this,ViewAttendanceByFacultyActivity.class);
          startActivity(intent);
                }
    });
  }
  @Override
  @Deprecated
  protected Dialog onCreateDialog(int id) {
    return new DatePickerDialog(this, datePickerListener, dyear, month, day);
  }
  private DatePickerDialog.OnDateSetListener datePickerListener = new
DatePickerDialog.OnDateSetListener() {
    public void onDateSet(DatePicker view, int selectedYear,
     int selectedMonth, int selectedDay) {
       dateEditText.setText(selectedDay + " / " + (selectedMonth + 1) + " / "
           + selectedYear);
    }
  };
}
```

9. DATABASE TABLES

9.1 Login Table:

SNO	Name	Type
1	usertype	varchar(100)
2	username	varchar(100)
3	password	varchar(20)

9.2 Student Table:

SNO	Name	Type
1	firstname	varchar(100)
2	lastname	varchar(100)
3	mobileno	numeric(10)
4	address	varchar(100)
5	username	varchar(100)
6	password	varchar(20)
7	department	Text
8	year	varchar(30)

9.3 Faculty Table:

SNO	Name	Type
1	firstname	varchar(100)
2	lastname	varchar(100)
3	mobileno	numeric(10)
4	address	varchar(100)
5	username	varchar(100)
6	password	varchar(20)

9.4 Attendance Table:

SNO	Name	Type
1	department	Text
2	year	varchar(30)
3	subject	varchar(10)
4	date	date

10. SYSTEM TESTING

10.1 Introduction

Once source code has been generated, software must be tested to uncover (and correct) as many errors as possible before delivery to the customer. Our goal is to design a series of test cases that have a high likelihood of finding errors. To uncover the errors software techniques are used. These techniques provide systematic guidance for designing tests that Exercise the internal logic of software components, and Exercise the input and output domains of the program to uncover errors In program function, behavior and performance. Internal program logic is exercised using ―White box test case design Techniques. Software requirements are exercised using ―block box test case Design techniques. In both cases, the intent is to find the maximum number of errors with the Minimum amount of effort and time.

10.2 Testing Methodologies

A strategy for software testing must accommodate low-level tests that are necessary to verify that a small source code segment has been correctly implemented as well as high-level tests that validate major system functions against customer requirements. A strategy must provide guidance for the practitioner and a set of milestones for the manager. Because the steps of the test strategy occur at a time when deadline pressure begins to rise, progress must be measurable and problems must surface as early as possible. Following testing techniques are well known and the same strategy is adopted during this project testing.

10.2.1 Unit testing

Unit testing focuses verification effort on the smallest unit of software design- the software component or module. The unit test is white-box oriented. The unit testing is implemented in every module of the student attendance management System. by giving correct manual input to the system, the data are stored in the database and retrieved. If you want a required module to access input or get the output from the End user. Any error will accrue the time will provide a handler to show what type of error will accrue.

10.2.2 System testing

System testing is actually a series of different tests whose primary purpose is to fully exercise the computer-based system. Below we have described the two types of testing which have been taken for this project. it is to check all modules worked on input basis. If you want to change any values or inputs will change all information. so specified input is must.

10.2.3 Performance Testing

Performance testing is designed to test the run-time performance of software within the context of an integrated system. Performance testing occurs throughout all steps in the testing process. Even at the unit level, the performance of an individual module may be assessed as white-box tests are conducted. This project reduces attendance table codes. It will generate reports fast, no need to have extra time or waiting for results. Entered correct data will show result in a few milliseconds. Just used the low memory of our system. Automatically do not get access to another software. Get user permission and access to other applications.

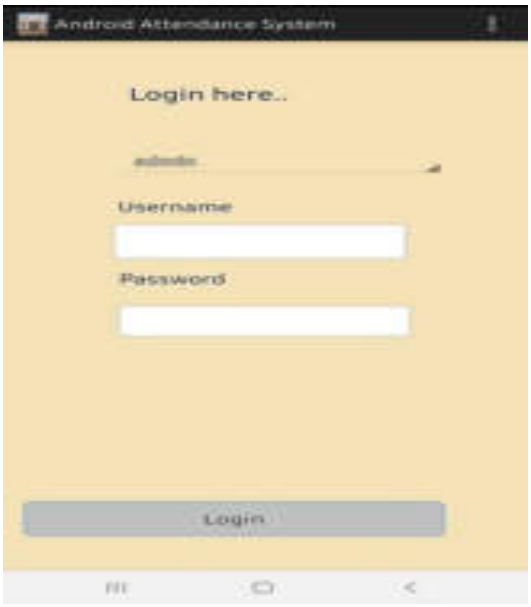
11. SNAPSHOTS

Home Page:

A home page is the main page of an application. The term also refers to one or more pages always shown in a browser when the application starts up. In this case, it is also known as the start page.

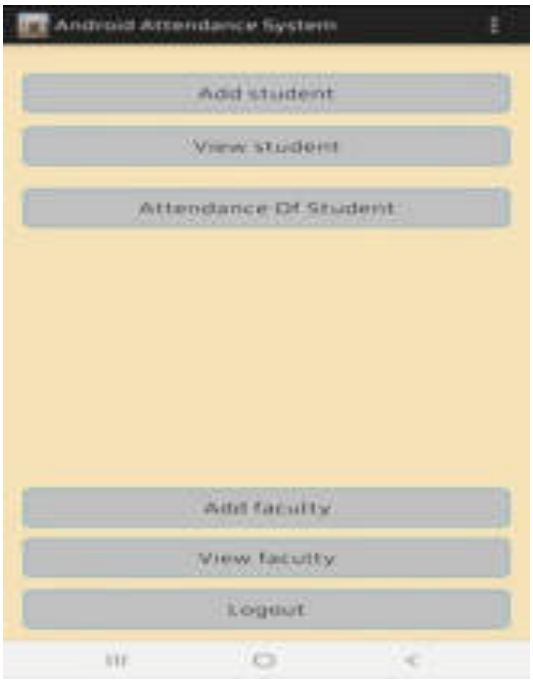
Login Page:

A login page is an application page or an entry page to an application that requires user identification and authentication, regularly performed by entering a username and password combination. Logins may provide access to an entire site or part of a website.



Admin Page:

When you create a Page, you automatically become the Page's admin. Only an admin can assign roles and change others' roles.

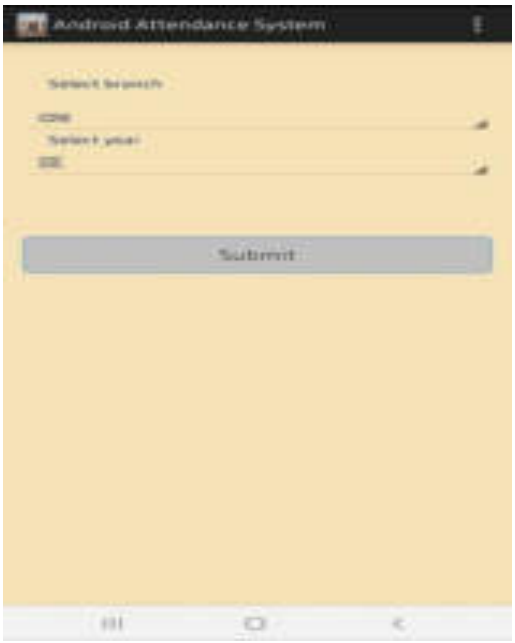


Add Student Page:

Only the admin of the application can add the student into the database using some attributes.

View Student Page:

Students who are registered in the application, their details can be viewed by providing some information such as their branch and year.



Add Faculty Page:

Only the admin of the application can add the faculty into the database using some attributes.

View Faculty Page:

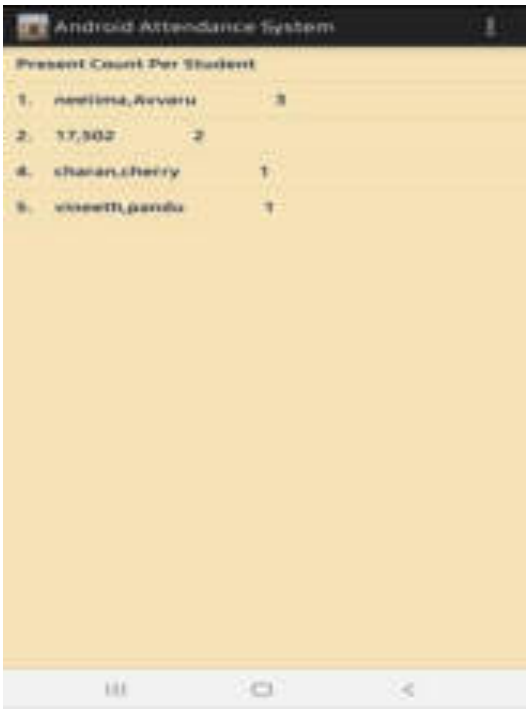
Teachers who are registered in the application, their details can be viewed.

Faculty Page:

Faculty members can login and can have access to view the data stored in the database and can update the student's attendance in the form of formatted reports.

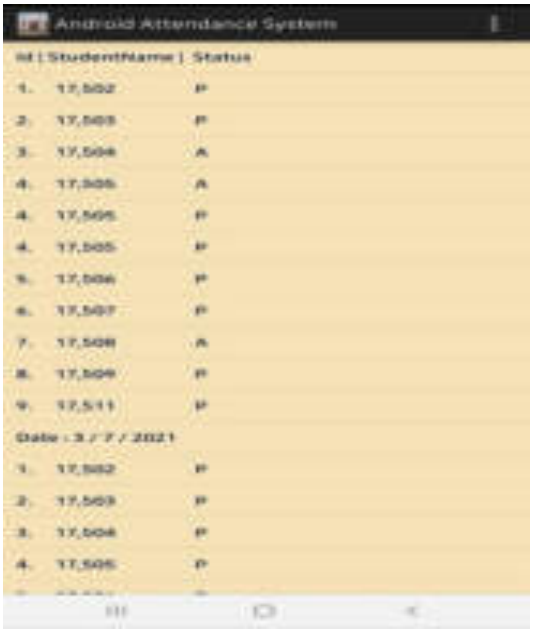
Attendance Page:

Attendance sheet is an official document for listing those attending a meeting, class, course etc, and this page shows only current day attendance.



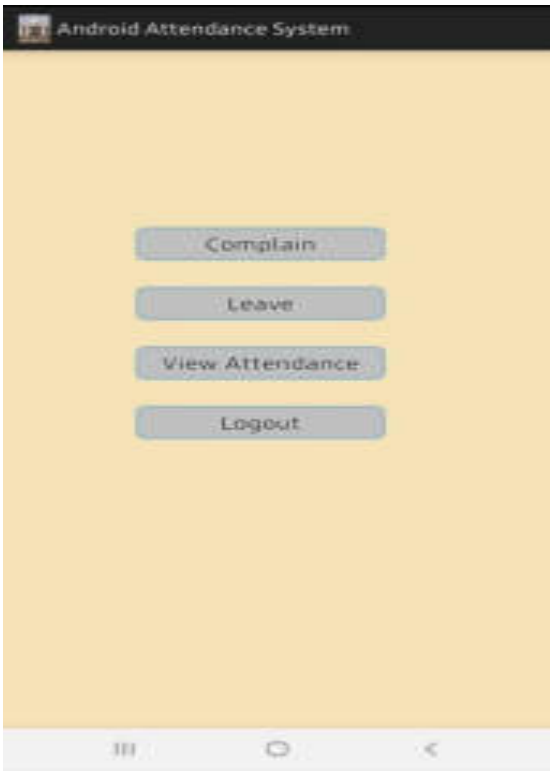
Total Attendance Page:

The following screen represents the total attendance of a student which displays all days attendance.



Student Page:

A student portal is a commonly used phrase to describe the login page where students can provide a username and password to gain access to View the Attendance, Apply Leave etc.



12. SYSTEM MAINTENANCE

Software maintenance is far more than finding mistakes. Provision must be made for environment changes which may affect either the computer, or other parts of the computer based systems. Such activity is normally called maintenance.

It includes both the Improvement of the system functions and the corrections of faults which arise during the operation of a new system.

It may involve the continuing involvement of a large proportion of computer Department resources. The main task may be to adapt existing systems in a changing environment.

Backup for the entire database files are taken and stored in storage devices like Flash drives, pen drives and disks so that it is possible to restore the system at the earliest. If theirs is a breakdown or collage, then the system gives provision to restore database files.

Storing data in a Separate secondary device leads to an effective and efficient maintenance of the system. The nominated person has sufficient knowledge of the organization's computer passed proposed change.

13. CONCLUSION

13.1 Conclusion

The Attendance Management System is developed using Visual Basic.NET fully meets the objectives of the system which it has been developed. The system has reached a steady state where all bugs have been eliminated. The system is operated at a high level of efficiency and all the teachers and users associated with the system understand its advantage. The system solves the problem. It was intended to solve the requirement specification.

13.2 Scope for future development

The project has a very vast scope in future. The project can be implemented on intranet in future. Project can be updated in near future as and when requirement for the same arises, as it is very flexible in terms of expansion. With the proposed software of database Space Manager ready and fully functional the client is now able to manage and hence run the entire work in a much better, accurate and error free manner.

14. REFERENCES

1. Mohd Heikal Husin, "Smart Attendance Systems Using Technologies", October, 2009.

2. Rakhi Joshi, "Android Based Smart Learning and Attendance Management System", June 2015.

3. Ganiyu R. A., "Mobile Operating Systems and Application Development Platforms", July10, 2014.

4. Gary Sims, "I want to develop Android Apps ", January 18, 2016

5. "AndroidArchitecture",http://www.tutorialspoint.com/ andoird/andoird_ architecture.html.

6. Juliana Puentes Giraldo, "programming language definition", 6 May, 2014.

7. Matthew David, "Mobile app development tools" ,24 Nov2015.

8. Neeraj Sharma, "Database Fundamentals " First EditionNovember 2010.

9. Mark Matthews,"MySQL and Java Developer'sGuide",5th edition march 2003

10. Usha Rani& Ramya Krishna, "Overview of Android For User Applications", November 2014.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

CONTENT ANALYSIS OF PDF DOCUMENT USING NATURAL LANGUAGE PROCESSING

Submitted

to

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
COMPUTER SCIENCE AND ENGINEERING

Submitted by

D.Venkata Sai Pravallika (17H71A05B6)

L.Meghana(17H71A0586)

G.K.S.S.M.Koushik(17H71A0577)

G.Anuradha (17H71A0562)

Under the Esteemed Guidance

of

J.Venkata Krishna

Associate Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273628, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited B.Tech (CSE|ECE|ME)
Accredited by NAAC A With Grade

2020-21

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
Accredited by NAAC With Grade A

CERTIFICATE

This is to certify that the Main Project entitled **"CONTENT ANALYSIS OF PDF DOCUMENT USING NATURAL LANGUAGE PROCESSING"** is a bonafide work carried out by **Ms D.Venkata Sai Pravallika (17H71A05B6), Ms L.Meghana (17H71A0586), Mr G.K.S.S.M.Koushik(17H71A0577), Ms G.Anuradha(17H71A0562)** in partial fulfilment for the award of degree of Bachelor of Technology in **Computer Science and Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

J.Venkata Krishna
Project Guide

D. Prasad
Head of the Department

Examiner

Dr K. Srinivas
Principal

ACKNOWLEDGEMENT

We would like to take this opportunity to express our deepest appreciation to the following people for their valuable contributions and assistance with this Project.

Initially, our would like to thank our project supervisor, **Mr J.Venkata Krishna, Associate professor, Department of Computer Science and Engineering** for the guidance and support, especially for the valuable ideas and knowledge shared to us throughout the Project.

We have the immense pleasure in expressing our thanks and deep sense of gratitude to**, Mr. D. Prasad, Head of the Department, Computer Science and Engineering** for extending necessary facilities and support for the completion of the Project.

We whole sincerely acknowledge **Dr. K. Srinivas, Principal and Prof D. PANDURANGA RAO, CEO** for giving opportunity to take up the Project work. We also extend our thanks to all faculty members of **Computer Science and Engineering**, for their valuable guidance and encouragement.

We would like to extend our warm appreciation to all our friends for sharing us their knowledge, valuable contributions and help with this Project.

Finally, our special thanks go to our families for their continuous support and help throughout my academic years and for their continual support and encouragement for the completion of the project.

D.Venkata Sai Pravallika (17H71A05B6)

L.Meghana(17H71A0586)

G.K.S.S.M.Koushik(17H71A0577)

G.Anuradha (17H71A0562)

DECLARATION

We **Ms D.Venkata Sai Pravallika(17H71A05B6), Ms.L.Meghana (17H71A0586), Ms.G.Anuradha (17H71A0562), Mr. K.S.S.M Koushik (17H71A0577)**of the Main-Project "**CONTENT ANALYSIS OF PDF DOCUMENT USING NATURAL LANGUAGE PROCESSING"**, hereby declare that the matter embodied in this Project is the genuine work done by us and has not been submitted either to this university or to any other university/institute for the fulfilment of the requirement of any course of study.

D.Venkata Sai Pravallika(17H71A05B6)

L. Meghana(17H71A0586)

G.Anuradha(17H71A0562)

G.K.S.S.M Koushik(17H71A0577)

ABSTRACT

Documents have become an essential part of human life, but the problem arises when someone wants a certain information in a particular document, it is a huge time consuming task and the task becomes even more difficult if the document is not organized. This is where the proposed system comes in, the user can upload any document no matter how the document is structured, the system reads the document and provides relevant answers to any question the user asks. This system uses high end to end Natural Language processing(NLP) algorithms and well pre-trained models to provide answers to any kind of query. Initially, the document goes through pre-processing and the pre-processed text is stored in a document store called Elasticsearch, it is also used to retrieve the documents at very high speed as it uses semantic search to search relevant information based on keywords and the model uses Bidirectional Encoder Representations from Transformers(BERT) to better understand the document and provide human-like answers to the user. The model is then implemented in the web platform where the user can upload the file and ask the question, and the model provides the relevant answer along with the context where the answer is taken.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

ML	-	Machine Learning
AI	-	Artificial Intelligence
NLP	-	Natural Language Processing
NLU	-	Natural Language Understanding
BERT	-	Bidirectional Encoder Representations from Transformers
NLTK	-	Natural Language Toolkit
LDA	-	Latent Dirichlet Allocation
RP	-	Random Projections
HDP	-	Hierarchical Dirichlet Process
NER	-	Named Entity Recognition
NLG	-	Natural Language Generation
ASP	-	Active Server Page
NSP	-	Next Sentence Prediction
CLS	-	Classification
MLM	-	Masked Language Modelling
SEP	-	Separate Segment
TF-IDF	-	Term Frequency - Inverse Document Frequency
BM25	-	Best Matching - 25
SQuAD	-	Stanford Question Answering Dataset

CHAPTER 1

INTRODUCTION

1.1 .OVERVIEW OF THE PROJECT

This project is an end-to-end question answering system for large document collections. Recent advances in NLP have enabled the application of QA to real world settings and is designed to be the bridge between research and industry.

It involves the following steps:

1. Preprocessing

2. Document Store

3. Retriever

4. Reader

1. Pre-processing

Pre-processing includes a suite of tools to extract text from different file types, normalize white space, and split text into smaller pieces to optimize retrieval. These data preprocessing steps can have a big impact on the systems performance and effective handling of data.

The PreProcessor takes one of the documents created by the converter as input, performs various cleaning steps and splits them into multiple smaller documents.

2.Document Store :

The document store acts as a database that stores text and metadata, provides them to the retriever at the query time.

3.Retriever

The Retriever is a lightweight filter that can quickly go through the full document store and pass a set of candidate documents to the Reader. It is a tool for sifting out the obvious negative cases, saving the Reader from doing more work than it needs to and speeding up the querying process.

4.Reader

The Reader, also known as Open-Domain QA systems in Machine Learning speak, is the core component that enables people to find the answers that are needed. Haystack's Readers are built on the latest transformer based language models, strong in their grasp of semantics, sensitive to syntactic structure, and state-of-the-art in QA tasks like SQuAD and Natural Questions.

1.2 SCOPE AND OBJECTIVE

SCOPE

The scope of this model is to provide a question answer system for any document with high performance in a more scalable approach and to provide better accuracy with high end to end natural language processing algorithms.

OBJECTIVE

People go through many documents each day, but when someone want a certain information in a particular document, it is a huge time consuming task and the task becomes even more difficult if the document is not organized. This is where the system comes in, the user can upload any document no matter how the document is structured, the

system reads the document and provides relevant answers to any question the user asks from the document. This system uses high end to end Natural Language processing(NLP) algorithms and well pre-trained models to provide answers to the queries.

CHAPTER 2

SYSTEM ANALYSIS

2.1 INTRODUCTION

Literature review forms the most important part of all scientific research. As a systematic investigation to find conclusions and to achieve facts, every scientific research builds on existing knowledge.

Unless one wants to change the wheel, precise awareness on the extent of wisdom on a subject is necessary to carry on research that adds value to the area. Literature review for scientific research can be defined as a survey of scientific papers, scholarly articles, and all other systematic scientific sources similar to a particular problem, field of study, theory, to include a description, summary, and critical evaluation of a concept, school of thought, or ideas related to the research question is Tested.

In extension, the literature review familiarizes the author to the extent of knowledge in their area. When represented as a part of the paper, it establishes to the readers, the author's depth of understanding and knowledge of their field subject. The literature is primarily scrutinized to identify gaps in the knowledge of the field source. This gap is further explored during research to establish latest facts or theories that provide value to the field.

2.2 LITERATURE SURVEY

TITLE 1: Target-Dependent Sentiment Classification With BERT.

AUTHORS: Z. Gao, A. Feng, X. Song and X. Wu.

Research on machine assisted text analysis follows the rapid development [1] of digital media, and sentiment analysis is among the applications .

Traditional sentiment analysis methods require complex feature engineering, and embedding representations have dominated leaderboards for a long time. However, the context-independent nature limits their representative power in rich context, hurting performance in Natural Language Processing (NLP) tasks. Bidirectional Encoder Representations from Transformers (BERT), among other pre-trained language models, beats existing best results in eleven NLP tasks (including sentence-level sentiment classification) by a large margin, which makes it the new baseline of text representation. As a more challenging task, fewer applications of BERT have been observed for sentiment classification at the aspect level. they implement three target-dependent variations of the BERTbase model, with positioned output at the target terms and an optional sentence with the target built in. Experiments on three data collections show that the TD-BERT model achieves new state-of-the-art performance, in comparison to traditional feature engineering methods, embedding-based models and earlier applications of BERT. With the successful application of BERT in many NLP tasks, the experiments try to verify if its context-aware representation can achieve similar performance improvement in aspect-based sentiment analysis. Surprisingly, coupling it with complex neural networks that used to work well with embedding representations does not show much value, sometimes with performance below the vanilla BERT-FC implementation. On the other hand, incorporation of target information shows stable accuracy improvement, and the most effective way of utilizing that information is displayed through the experiment.

TITLE 2: An Analysis of the Emotional Evolution of Large-Scale Internet Public Opinion Events Based on the BERT-LDA Hybrid Model.

AUTHORS: X. Tan, M. Zhuang, X. Lu and T. Mao. The purpose of this article is to analyse the [2] emotional evolution of the netizens in reaction to the events of the Anti-ELAB (Anti-Extradition LaAmendment Bill)

movement in Hong Kong. The authors attempt to investigate evolving laws of large-scale Internet public opinion events and provide relevant agencies with a theoretical basis for a public opinion response mechanism. On the basis of improving the Bidirectional Encoder Representations from Transformers (BERT) pre-training task, add in-depth pre-training tasks, and based on the optimisation results of the LDA topic embedding, they integrate deeply with the LDA model to dynamically present the fine-grained public sentiment of the event. Through the collection of large-scale text data related to the Anti-ELAB Movement from a well-known forum in Hong Kong, a BERT-LDA hybrid model for large-scale network public opinion analysis is constructed in a complex context. Through empirical analysis, the authors calculate and reveal the emotional change process of netizens and opinion leaders in the three transition stages of the Anti-ELAB Movement with the evolution of the topic word as the core by visualisation. They also analyse the emotional distribution and evolution trend of public opinion under the `text topic', and deeply analyse the character and role of opinion leaders in Anti-ELAB public opinion events. The improved BERT-LDA model or sentiment classification AUC value exceeds 99.6% in the sentiment classification task for the Anti-ELAB Movement.

TITLE 3: A Hybrid BERT Model That Incorporates Label Semantics via Adjustive Attention for Multi-Label Text Classification..

AUTHORS: L. Cai, Y. Song, T. Liu and K. Zhang

The multi-label text classification task aims [3] to tag a document with a series of labels. Previous studies usually treated labels as symbols without semantics and ignored the relation among labels, which caused information loss. In this paper, the authors show that explicitly modeling label semantics can improve multi-label text classification. The authors propose a hybrid neural network model to simultaneously take advantage of both label semantics and

fine-grained text information. Specifically, they utilize the pre-trained BERT model to compute context-aware representation of documents. Furthermore, incorporating the label semantics in two stages. First, a novel label graph construction approach is proposed to capture the label structures and correlations. Second, they propose a neoteric attention mechanism-adjustive attention to establish the semantic connections between labels and words and to obtain the label-specific word representation. The hybrid representation that combines context-aware feature and label-special word feature is fed into a document encoder to classify. Experimental results on two publicly available datasets show that thier model is superior to other state-of-the-art classification methods.

TITLE 4: A Question Answering-Based Framework for One-Step Event Argument extraction.

AUTHOR: Y. Zhang et al

Event argument extraction, which aims [4] to identify arguments of specific events and label their roles, is a challenging subtask of event extraction. Previous approaches solve this problem in a two-stage manner that first extracts named entities as argument candidates and then determines their roles. However, many nested entities may be missed or wrongly predicted during the argument candidate extraction procedure, which substantially affects the performance of the downstream classifier. In this paper, the authors propose a novel one-step question answering based framework, which performs argument candidate extraction and argument role classification simultaneously to mitigate the error propagation problem in conventional two-stage methods. Since the conventional question answering based framework cannot be applied directly to this task, they design a Question Answering based Sequence Labeling (QA-SL) model to tackle inexistent argument roles and multiple argument token spans. Moreover, considering the overwhelming number of parameters in question answering based neural network models and the relatively small size of event extraction corpus, they fine-tune the

pre-trained model from BERT to mitigate the data scarcity problem. Extensive experiments demonstrate the benefits of the proposed method, leading to a competitive performance compared with state-of-the-art methods. To the best of knowledge, this is the first work to cast event argument extraction as a question answering task.

TITLE 5: Cross-Lingual Passage Re-Ranking With Alignment Augmented Multilingual BERT.

AUTHORS: D. Chen, S. Zhang, X. Zhang and K. Yang

The task of Cross-lingual Passage Re-ranking (XPR) aims[5] to rank a list of candidate passages in multiple languages given a query, which is generally challenged by two main issues: (1) the query and passages to be ranked are often in different languages, which requires strong cross-lingual alignment, and (2) the lack of annotated data for model training and evaluation. In this article, the authors propose a two-stage approach to address these issues. At the first stage, they introduce the task of Cross-lingual Paraphrase Identification (XPI) as an extra pre-training to augment the alignment by leveraging a large unsupervised parallel corpus. This task aims to identify whether two sentences, which may be from different languages, have the same meaning. At the second stage, they introduce and compare three effective strategies for cross-lingual training. To verify the effectiveness of their method, they construct an XPR dataset by assembling and modifying two monolingual datasets. Experimental results show that their augmented pre-training contributes significantly to the XPR task. Besides, they directly transfer the trained model to test on out-domain data which are constructed by modifying three multi-lingual Question Answering (QA) datasets. The results demonstrate the cross-domain robustness of the proposed approach.

TITLE 6: Natural Language Statistical Features of LSTM-Generated Texts

AUTHORS: M. Lippi, M. A. Montemurro, M. Degli Esposti and G. Cristadoro DETECTION OF QUALITY OF LEAF

Long short-term memory (LSTM) networks have recently shown remarkable performance [6] in several tasks that are dealing with natural language generation, such as image captioning or poetry composition. Yet, only few works have analyzed text generated by LSTMs in order to quantitatively evaluate to which extent such artificial texts resemble those generated by humans. The authors compared the statistical structure of LSTM-generated language to that of written natural language, and to those produced by Markov models of various orders. In particular, they characterized the statistical structure of language by assessing word-frequency statistics, long-range correlations, and entropy measures. Their main finding is that while both LSTM- and Markov-generated texts can exhibit features similar to real ones in their word-frequency statistics and entropy measures, LSTM-texts are shown to reproduce long-range correlations at scales comparable to those found in natural language. Moreover, for LSTM networks, a temperature-like parameter controlling the generation process shows an optimal value-for which the produced texts are closest to real language-consistent across different statistical features investigated.

TITLE 7: Graph-Based Methods for Natural Language Processing and Understanding—A Survey and Analysis

AUTHORS: M. T. Mills and N. G. Bourbakis. This survey and analysis presents the functional components, performance, and maturity of graph-based methods [7] for natural language processing and natural language understanding and their potential for mature products. Resulting capabilities from the methods surveyed include summarization, text entailment, redundancy reduction, similarity measure, word sense induction and disambiguation, semantic relatedness, labeling (e.g., word sense), and novelty detection. Estimated scores for accuracy, coverage, scalability, and performance are derived from each method. This survey and analysis, with

tables and bar graphs, offers a unique abstraction of functional components and levels of maturity from this collection of graph-based methodologies.

TITLE 8: Machine Reading Comprehension Framework Based on Self-Training for Domain Adaptation.

AUTHORS: H. -G. Lee, Y. Jang and H. Kim

Machine reading comprehension (MRC) is a type of question answering mechanism in which a computer reads documents and answers related questions[8]. The accuracies of recent MRC systems surpass those of humans. However, most MRC systems exhibit significant performance deteriorations when domains are changed. Hence, the authors propose a self-training framework for MRC. The proposed framework is composed of a pseudo-answer extractor, a pseudo-question generator, and an MRC system. In the source domain, components are pre trained using an MRC training dataset. In the target domain, the performance of the pseudo-question generator and MRC system is improved through a mutual self-training scheme. During the mutual self-training, the pseudo-question generator provides new training data to the MRC system and obtains rewards from the MRC system for reinforcement learning. In experiments with a Wikipedia domain (source domain) and civil affair domain (target domain), an MRC system based on the proposed self-training scheme demonstrates better performances than that based on automatic data augmentation.

CHAPTER 3
SYSTEM DESIGN

3.1 INTRODUCTION

Design is a multi-step that focuses on data structure software architecture, procedural details, algorithm and interface between modules. The design process also translates the requirements into presentation of software that can be accessed for quality before coding begins. Systems design could be seen as the application of systems theory to product development. There is some overlap with the disciplines of systems analysis, systems architecture and systems engineering. Until the 1990s, systems design had a crucial and respected role in the data processing industry. In the 1990s, standardization of hardware and software resulted in the ability to build modular systems.

The increasing importance of software running on generic platforms has enhanced the discipline of software engineering. The architectural design of a system emphasizes the design of the system architecture that describes the structure, behavior and more views of that system and analysis. It is meant to satisfy specific needs and requirements of a business or organization through the engineering of a coherent and well-running system.

They needed to be able to standardize their work into a formal discipline with proper methods, especially for new fields like information theory, operations research and computer science in general. Computer software design change continuously as new methods; better analysis and border understanding evolved. Software design is at relatively early stage in its revolution. Therefore, software design methodology lacks the depth, flexibility and quantitative nature that are normally associated with more classical engineering disciplines. However techniques for software designs do exit, criteria for design qualities are available and design notation can be applied.

3.2 EXISTING SYSTEM

In the existing system, the question answering is implemented using different algorithms such as BERT, AllenNLP and so on. It is a simple system that works on text data of utmost 500 words and provides answers to queries from it.

EXISTING SYSTEM ADVANTAGES

1. Provides better accuracy due to less number of data.
2. Requires less computation power and is easy to implement.

EXISTING SYSTEM DRAWBACKS

1. Not every user has the data in the particular format which the system inputs.
2. The system can only provide queries to certain domain specific queries.
3. The existing system is restricted to limited input data.

3.3 PROPOSED SYSTEM

Inputs any kind of document regardless of domain and size, from the user and provides answers to the queries by training on the data with existing pre-trained models.

PROPOSED SYSTEM ADVANTAGES

1. It can provide answers to any kind of documents no matter the size.
2. Better scalability can be achieved than the existing one.

3.4 SYSTEM ALGORITHM

BERT

BERT relies on a Transformer (the attention mechanism that learns contextual relationships between words in a text). A basic Transformer consists of an encoder to read the text input and a decoder to produce a prediction for the task. Since BERT's goal is to generate a language representation model, it only needs the encoder part. The input to the encoder for BERT is a sequence of tokens, which are first converted into vectors and then processed in the neural network. But before processing can start, BERT needs the input to be massaged and decorated with some extra metadata:

1. Token embeddings: A [CLS] token is added to the input word tokens at the beginning of the first sentence and a [SEP] token is inserted at the end of each sentence.
2. Segment embeddings: A marker indicating Sentence A or Sentence B is added to each token. This allows the encoder to distinguish between sentences.
3. Positional embeddings: A positional embedding is added to each token to indicate its position in the sentence.

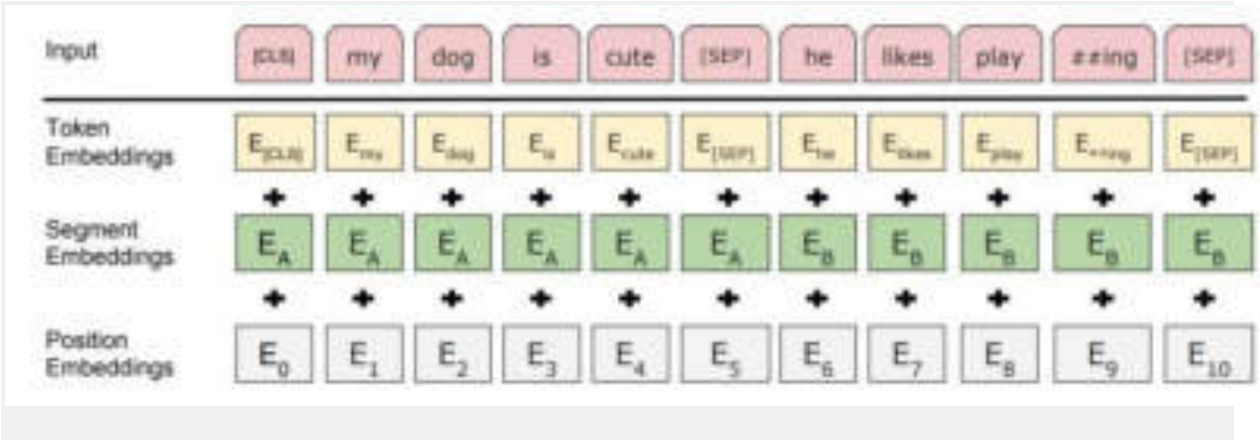


FIG 3.1 : BERT Embeddings

Essentially, the Transformer stacks a layer that maps sequences to sequences, so the output is also a sequence of vectors with a 1:1 correspondence between input and output tokens at the same index. BERT does not try to predict the next word in the sentence. Training makes use of the following two strategies:

1. Masked LM (MLM)

The idea here is "simple": Randomly mask out 15% of the words in the input — replacing them with a [MASK] token — run the entire sequence through the BERT attention based encoder and then predict only the masked words, based on the context provided by the other non-masked words in the sequence. However, there is a problem with this naive masking approach — the model only tries to predict when the [MASK] token is present in the input, while the model tries to predict the correct tokens regardless of what token is present in the input. To deal with this issue, out of the 15% of the tokens selected for masking:

- 80% of the tokens are actually replaced with the token [MASK].
- 10% of the time tokens are replaced with a random token.
- 10% of the time tokens are left unchanged.

While training the BERT loss function considers only the prediction of the masked tokens and ignores the prediction of the non-masked ones. This results in a model that converges much more slowly than left-to-right or right-to-left models.

2.Next Sentence Prediction (NSP)

In order to understand *relationship* between two sentences, BERT training process also uses next sentence prediction. A pre-trained model with this kind of understanding is relevant for tasks like question answering. During training the model gets as input pairs of sentences and it learns to predict if the second sentence is the next sentence in the original text as well.

BERT separates sentences with a special [SEP] token. During training the model isfed with two input sentences at a time such that:

- 50% of the time the second sentence comes after the first one.
- 50% of the time it is a a random sentence from the full corpus.

BERT is then required to predict whether the second sentence is random or not, with the assumption that the random sentence will be disconnected from the first sentence:

To predict if the second sentence is connected to the first one or not, basically the complete input sequence goes through the Transformer based model, the output of the [CLS] token is transformed into a 2×1 shaped vector using a simple classification layer, and the IsNext-Label is assigned using softmax.

The model is trained with both Masked LM and Next Sentence Prediction together.

3.1 SYSTEM ARCHITECTURE

The following figure shows the flow of the project at first, data is collected from the user and passed into the file converter which converts the uploaded file to text document and preprocessing is applied on it. Then the text is uploaded intothe document store. The required data is retrieved from the document store based on the keyword search of user's query, then the retrieved data is fed to the reader, here the system uses BERT to understand the context and give the answer to the user.

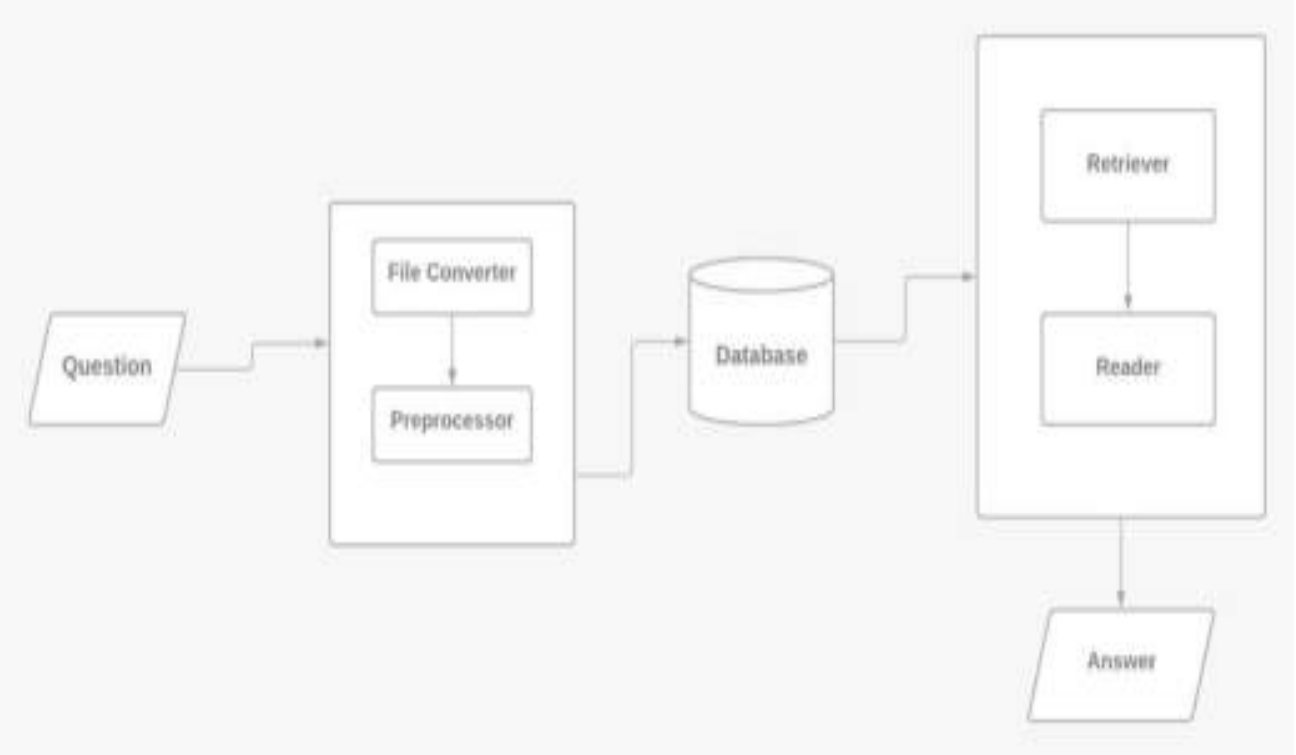


FIG 3.2 - ARCHITECTURE DIAGRAM

3.2 UML DIAGRAMS

ACTIVITY DIAGRAM

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams are intended to model both computational and organizational processes, well as the data flows intersecting with the relating activities.

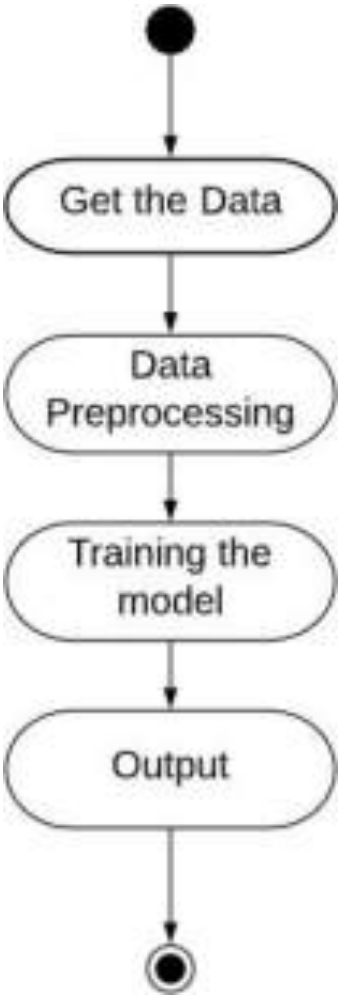


FIG 3.3 - ACTIVITY DIAGRAM

STATE DIAGRAM:

A state diagram is used to represent the condition of the system or part of the system at finite instances of time. It's a behavioral diagram and it represents the behavior using finite state transitions. State diagrams are also referred to as State machines and State-chart Diagrams

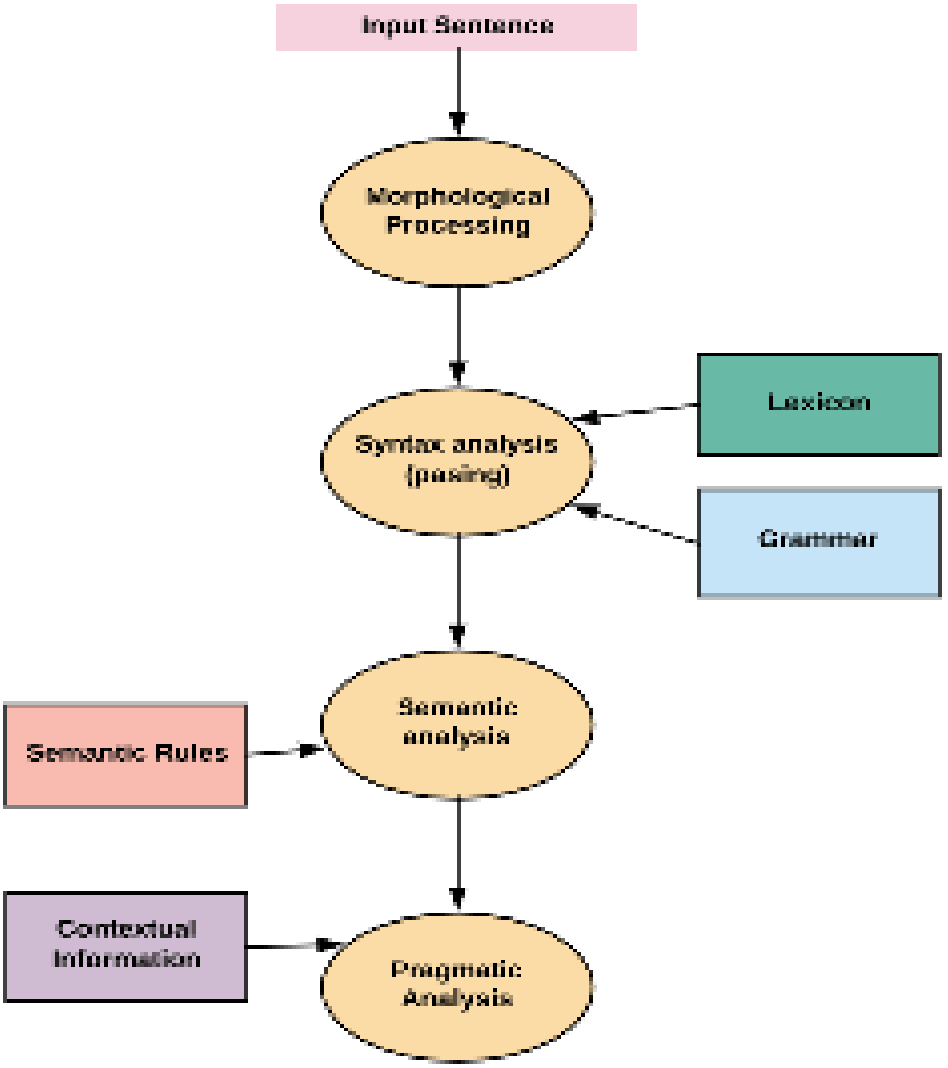


FIG : 3.4.1 –STATE DIAGRAM

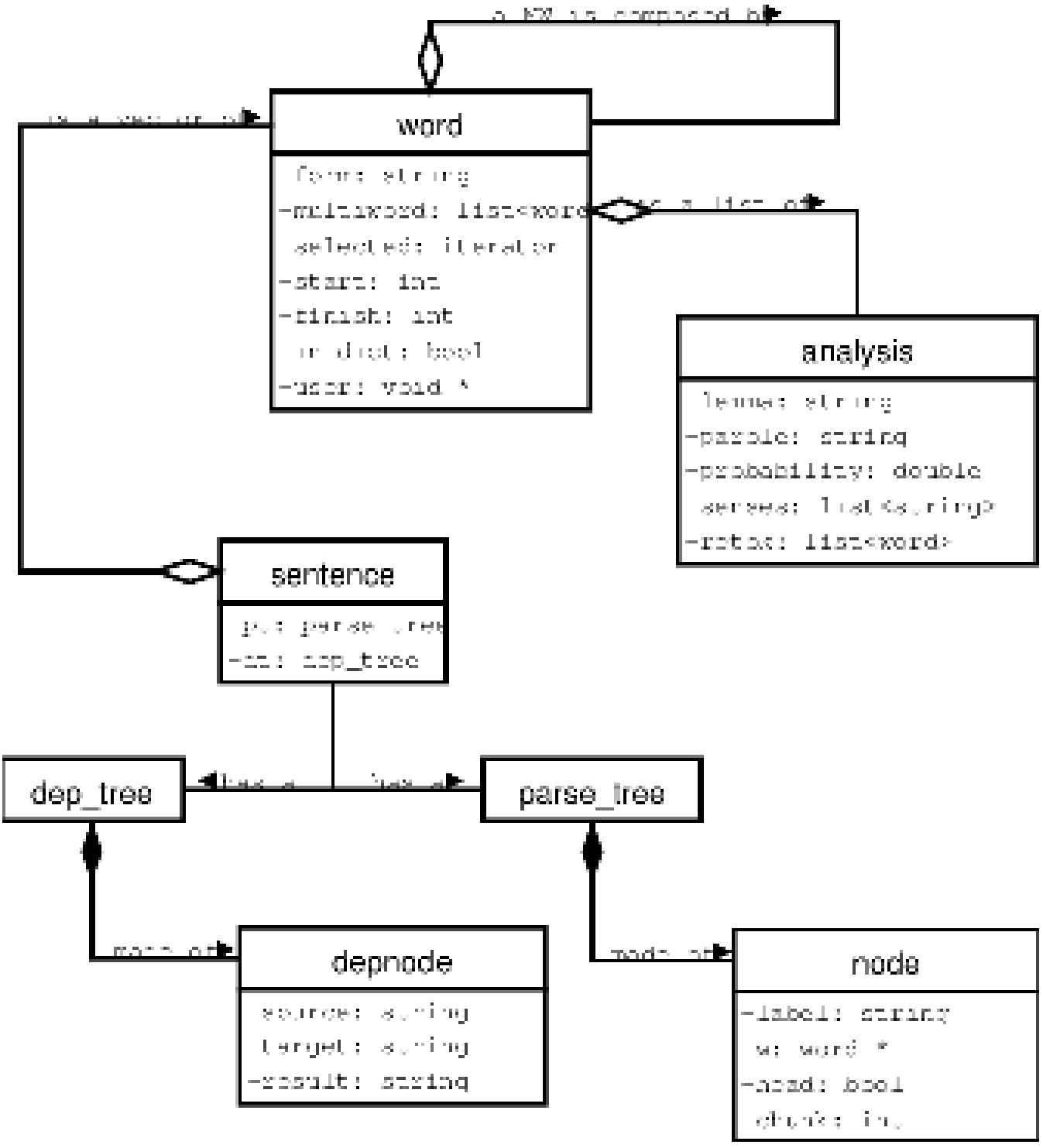


FIG :3.4.2 STATE DIAGRAM

COMPONENT DIAGRAM :

A component provides the set of required interfaces that a component realizes or implements. These are the static diagrams of the unified modeling language. Component diagrams are used to represent the working and behavior of various components of a system.

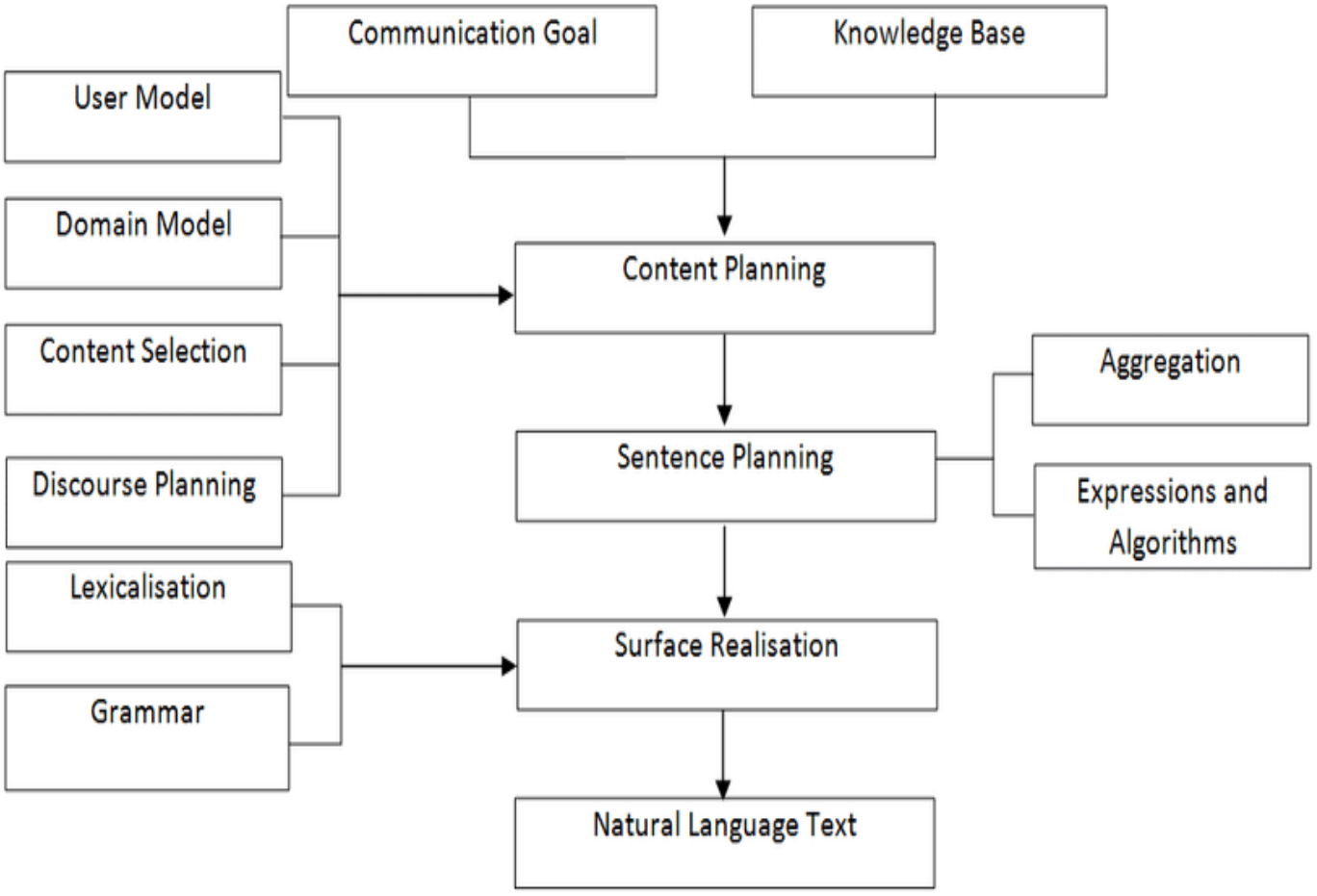


FIG :3.6 COMPONENT DIAGRAM

USE CASE DIAGRAM:

A use case diagram is a graphical depiction of a user's possible interactions with a system. A use case diagram shows various use cases and different types of users the system has and will often be accompanied by other types of diagrams as well. The use cases are represented by either circles or ellipses. The actors are often shown as stick figures.

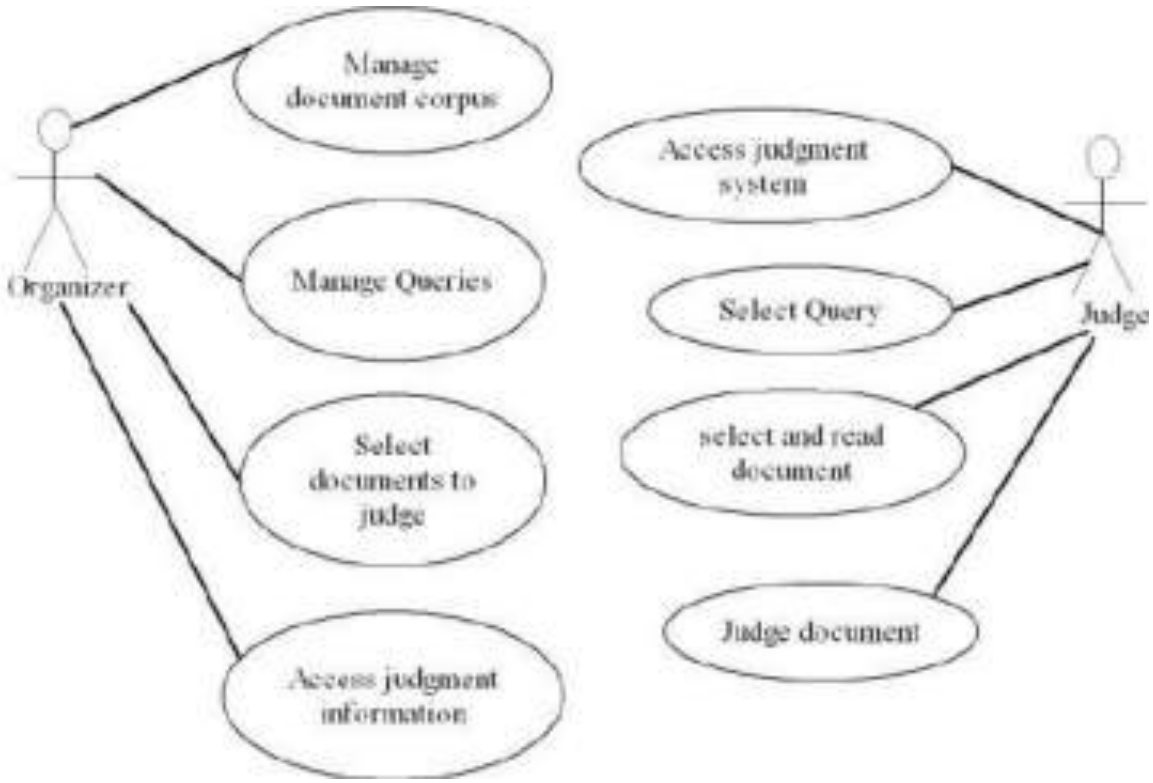


FIG :3.5 USE CASE DIAGRAM

3.3 HARDWARE USED

System	:	Pentium IV 2.4 GHz
Hard Disk	:	40GB
RAM	:	4 GB

3.4 SOFTWARE USED

Operating system	: Ubuntu 18.04
IDE	:Visual Studio Coding
Language	: Python

CHAPTER 4

TESTING AND IMPLEMENTATION

4.1 NATURAL LANGUAGE PROCESSING LIBRARIES USED

Thanks to the development of useful NLP libraries, today, NLP is finding applications across the various parallels of the industrial landscape. In fact, NLP has now become an integral part of Deep Learning development. Extracting valuable information from free text is essential for developing chatbots, patent research & analysis, voice/speech recognition, patient data processing, and querying image content, among other use cases of NLP.

The fundamental aim of NLP libraries is to simplify text preprocessing. A good NLP library should be able to correctly convert free text sentences into structured features (for example, cost per hour) that can easily be fed into ML or DL pipelines. Also, an NLP library should have a simple-to-learn API, and it must be able to implement the latest and greatest algorithms and models efficiently. These are the most important libraries used in the model:

- Natural Language Toolkit(NLTK)
- spaCy
- Pipeline
- Pattern
- Scikit Learn

1. Natural Language Toolkit (NLTK)

NLTK is one of the leading platforms for building Python programs that canwork with human language data. It presents a practical introduction to programming for language processing.

4. spaCy

spaCy is an open-source NLP library in Python. It is designed explicitly for production usage – it lets people develop applications that process and understand huge volumes of text.

5. Pipeline

The pipelines are a great and easy way to use models for inference. These pipelines are objects that abstract most of the complex code from the library, offering a simple API dedicated to several tasks, including Named Entity Recognition, Masked Language Modeling, Sentiment Analysis, Feature Extractionand Question Answering.

6. Pattern

Pattern is a text processing, web mining, natural language processing, machine learning, and network analysis tool for Python. It comes with a host of tools for data mining (Google, Twitter, Wikipedia API, a web crawler, and an HTML DOM parser), NLP (part-of-speech taggers, n-gram search, sentiment analysis, WordNet), ML (vector space model, clustering, SVM), and network analysis by graph centrality and visualization.

7. Scikit Learn

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning

and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistent interface in Python. This library, which islargely written in Python, is built upon NumPy, SciPy and Matplotlib.

4.2 SYSTEM IMPLEMENTATION

MODULES:

- ➢ Collecting Dataset
- ➢ Preprocessing
- ➢ Document Store
- ➢ Retriever
- ➢ Reader
- ➢ Web Interface

MODULE DESCRIPTION:

Collecting Dataset

The model uses a pre-trained dataset known as SQuAD which is provided by Stanford, when it comes to the question answering system, one cannot fail to mention the Stanford Question Answer Dataset (SQuAD). There are plenty of QA or MRC datasets available (Natural Questions, MS MARCO, CoQA, etc.), but SQuAD is one of the most used.

- SQuAD 1.1 (2016) contains 100,000 question-answer pairs (536 articles);
- SQuAD 2.0 (2018) adds 53,775 new unanswerable questions about the same paragraphs;
- Dataset collection: the top 10000 articles of English Wikipedia, 536 articles were sampled uniformly at random; 23,215 paragraphs for the 536 articles covering a wide range of topics, from musical celebrities to abstract concepts

1. Pre-processing

- It is the gathering of task related information based on some targeted variables to analyse and produce some valuable outcome.
- However, some of the data may be noisy, i.e. may contain inaccurate values, incomplete values or incorrect values.
- Hence, it is necessary to process the data before analysing it and coming to the results.
- Data pre-processing can be done by data cleaning, data transformation, data selection.
- Data cleaning includes Fill in missing values, smooth noisy data, identify or remove outliers, and resolve inconsistencies
- Data transformation may include smoothing, aggregation, generalization, transformation which improves the quality of the data.
- Data selection includes some methods or functions which allow to select the useful data for the system.

2. Document Store

The document store acts as a database that stores text and metadata, provides them to the retriever at the query time. Elasticsearch is the database this project uses. Elasticsearch (ES) is an open-source, broadly-distributable, readily-scalable, enterprise-grade search engine and database. Accessible through an extensive and elaborate API, Elasticsearch can power extremely fast searches that support data discovery applications.

3. Retriever

The Retriever is a lightweight filter that can quickly go through the full document store and pass a set of candidate documents to the Reader. It is a tool for

sifting out the obvious negative cases, saving the Reader from doing more work than it needs to and speeding up the querying process.

TF-IDF is a commonly used baseline for information retrieval that exploits two key intuitions, documents that have more lexical overlap with the query are more likely to be relevant, and words that occur in fewer documents are more significant than words that occur in many documents.

Given a query, a tf-idf score is computed for each document as follows:

score = tf*idf

Where:

- tf is how many times words in the query occur in that document.
- idf is the inverse of the fraction of documents containing the word.

In practice, both terms are usually log normalised.

In the model, BM25 is used, which is a variant of TF-IDF, it is a retrieval method that does not need a neural network for indexing. It improves upon its predecessor in two main aspects, It saturates tf after a set number of occurrences ofthe given term in the document, and It normalises by document length so that shortdocuments are favoured over long documents if they have the same amount of word overlap with the query.

4. Reader

The Reader, also known as Open-Domain QA systems in Machine Learningspeak, is the core component that enables people to find the answers that are needed. Haystack's Readers are built on the latest transformer based language models, strong in their grasp of semantics, sensitive to syntactic structure, and state-of-the-art in QA tasks like SQuAD and Natural Questions.

5. Web Interface

The whole model would not have any use if it cannot be used by someone, in order for someone to use the model they need a proper interface whether its a web application or mobile application.

Therefore, a web interface is introduced here in order for an user to use the application. The web interface is implemented using Flask, it is a micro web platform written in Python it uses html file as a template.

In this web application, there is an interface where the user can upload the document and after uploading the user will be redirected to a page where he can ask the query and then redirected to the answer page.

4.1 TESTING

4.3.1 TESTING PROCESS

Everyone makes mistakes and if left unchecked, some of these mistakes canlead to failures or bugs that can be very expensive to recover from. Testing the code helps to catch these mistakes or avoid getting them into production in thefirst place. Testing therefore is very important in software development. Used effectively, tests help to identify bugs, ensure the quality of the product and to verify that the software does what it is meant to do.

4.3.2 TYPES OF TESTINGS

4.3.2.1 Unit Testing

Unit testing includes the design of test belongings that validate that the internal program logic is operative properly, and that program input produces validoutputs. All choice branches and internal code flow should be authorized. It is the

testing of separate software units of the request. It is done after the close of an individual unit before integration. This is a structural testing, that relies on data of its structure and is invasive. Unit tests achieve basic tests at factor level and test a specific commercial process, application, and/or system formation. Unit tests ensure that each single path of business process completes accurately to the documented provisions and contains clearly defined inputs and probable results.

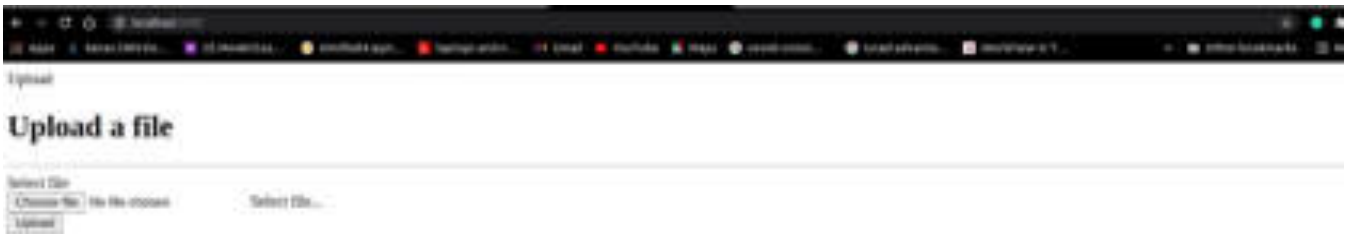


FIG 4.1 - UNIT TESTING HTML

The html front end file is run before integrating to the backend to check whether the file can be uploaded or not and the test shows the file can be uploaded successfully.

FIG 4.2 - UNIT TESTING MODEL

FIG 4.3 - UNIT TESTING BACKEND

The above figure represents the output of the backend without the web interfaceand the highest probability of 0.86 is achieved.

4.3.2.2 Integration Testing

Integration tests are calculated to test integrated software components to regulate if they actually run as one program.

In this model, three things are integrated: database, backend and web interface. The database server must be run independently and backend should connect to that database by the port.

FIG 4.4 - INTEGRATION TESTING DATABASE

The database server is run here and connected to the server by accessing the localhost address from the backend and the database is connected at localhost:9200/

FIG 4.5 - INTEGRATION TESTING CONNECTION

4.2 RESULTS AND DISCUSSIONS

Overall, the algorithm gives better results than expected, successfully training the documents and predicting the output.

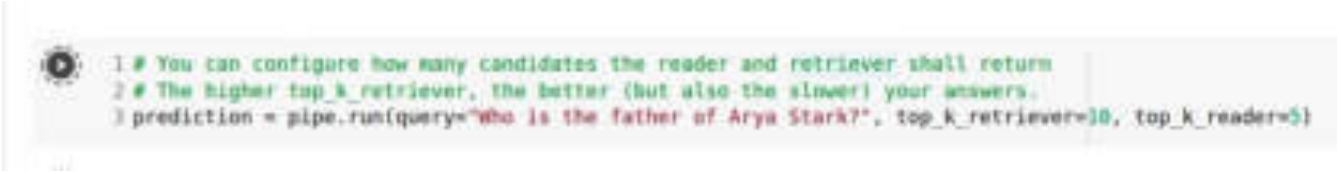


Fig 4.6 - Question

A text file was tested which consists of all the game of thrones scripts and when a particular question is asked it gives the following output.

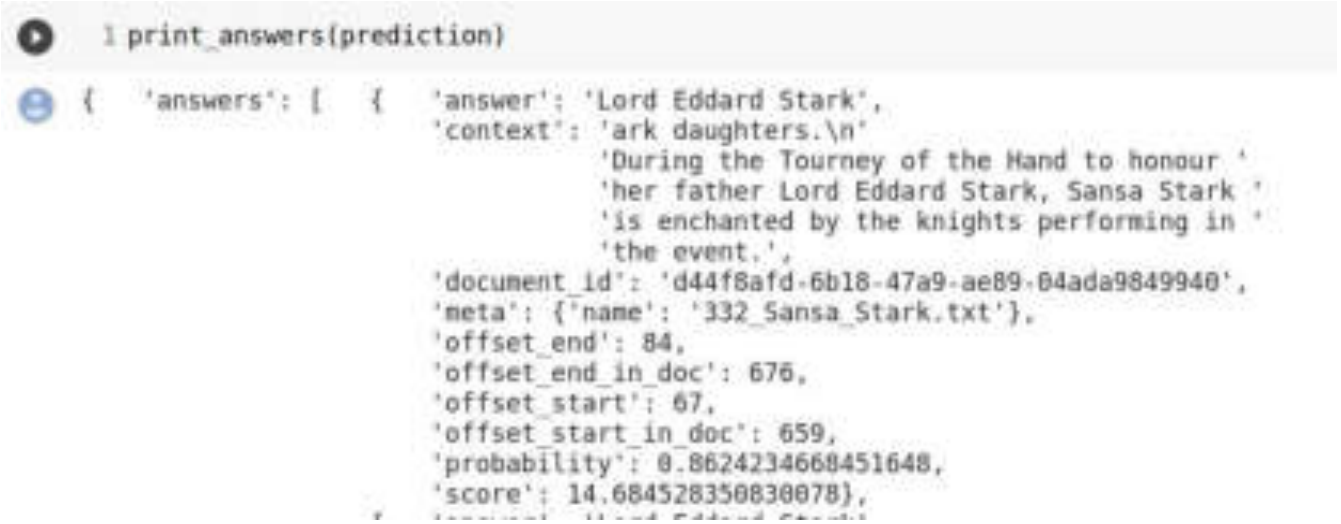


Fig 4.7 - Output

The model has successfully predicted the answer and the context where the answer is taken from with the probability of 0.86 which is a pretty decent for BERT based NLP tasks.

CHAPTER 5

CONCLUSION

5.1 CONCLUSION

This project is a huge advantage for people who deal with huge documents everyday as it can provide answers to queries from any document regardless of domain and size of the document. The project is also implemented in a more efficient and scalable manner for future enhancements and could achieve better results from future advanced NLP algorithms.

5.2 FUTURE ENHANCEMENT

The project has a very vast scope in future. The project can be implemented based on more advanced Natural Language Processing algorithms in future. Project can be updated in the near future as and when requirement for the same series, as it is very flexible in terms of expansion. The proposed system can provide answers to any kind of documents if the required computation is provided,in a much better, accurate manner.

APPENDIX A : SOFTWARE DESCRIPTION

A.1 PYTHON TECHNOLOGY

Python is an interpreted, object-oriented programming language similar to PERL, that has gained popularity because of its clear syntax and readability. Python is said to be relatively easy to learn and portable, meaning its statements can be interpreted in a number of operating systems, including UNIX-based systems, Mac OS, MS-DOS, OS/2, and various versions of Microsoft Windows
98. Python was created by Guido van Rossum, a former resident of the Netherlands, whose favourite comedy group at the time was Monty Python'sFlying Circus. The source code is freely available and open for modification and reuse. Python has a significant number of users. A notable feature of Python is its indenting of source statements to make the code easier to read. Python offers dynamic data type, ready-made class, and interfaces to many system calls and libraries. It can be extended, using the C or C++ language. Python can be used as the script in Microsoft's Active Server Page (ASP) technology.

A.2 PYTHON PLATFORM

Apart from Windows, Linux and MacOS, C, Python implementation runs on
21 different **platforms**. Iron Python is a .NET framework based **Python** implementation and it is capable of running in both Windows, Linux and in other environments where .NET framework is available.

A.3 NATURAL LANGUAGE PROCESSING – OVERVIEW

Natural Language Processing, or **NLP** is a subfield of Artificial Intelligence research that is focused on developing models and points of interaction between humans and computers based on natural language. This includes text, but also speech-based systems.

Computer scientists and researchers have been studying this topic for entire decades, but only recently has it become a hot topic again, a situation made possible by recent breakthroughs in the research community.

A.3.1 WHAT IS NATURAL LANGUAGE PROCESSING?

Natural Language Processing (NLP) is the technology used to help machines to understand and learn text and language. With NLP data scientists aim to teach machines to understand what is said and written to make sense of the human language. It is used to apply machine learning algorithms to text and speech.

Fully understanding and representing the meaning of language is an extremely difficult goal. One of the major challenges to developing NLP applications is computers most likely need structured data, but as far as human speech is concerned it is unstructured and often ambiguous.

The rules that direct the information using natural languages are not easy forcomputers to understand and translate, especially to sense the tone for example if someone uses a sarcastic remark to pass information. This means computers have to comprehensively understand the meaning of words as well as the intention or emotion behind the words. Unlike programming, human languages are ambiguous which make them complex and hard to learn.

NLP has primarily two aspects: natural language understanding (NLU) or natural language interpretation (NLI) (i.e. human to machine) and natural language generation (NLG) (i.e. machine to human). In simple words, one can say that NLGis the inverse of NLU (broadly called as NLP). Natural language generation(NLG) is when software automatically transforms data into written narrative.

A.3.2 NATURAL LANGUAGE PROCESSING – TASKS

A.3.2.1 Stemming

Stemming is the process of reducing the words(generally modified or derived) to their word stem or root form. The objective of stemming is to reduce related words to the same stem even if the stem is not a dictionary word.

A.3.3.2 Lemmatization

Lemmatization is the process of reducing a group of words into their lemma or dictionary form. It takes into account things like POS(Parts of Speech), the meaning of the word in the sentence, the meaning of the word in the nearby sentences etc. before reducing the word to its lemma.

A.3.3.3 Word Embeddings

Word Embeddings is the name of the techniques which are used to represent Natural Language in vector form of real numbers. They are useful because of computers' inability to process Natural Language. So these Word Embeddings capture the essence and relationship between words in a Natural Language using real numbers. In Word Embeddings, a word or a phrase is represented in a fixed dimension vector of length say 100.

A.3.3.4 Part-Of-Speech Tagging

In Simplistic terms, Part-Of-Speech Tagging is the process of marking up words in a sentence as nouns, verbs, adverbs, and so on.

A.3.3.5 Named Entity Disambiguation

Named Entity Disambiguation is the process of identifying the mentions of entities in a sentence.

A.3.3.6 Named Entity Recognition

Named Entity Recognition is the task of identifying entities in a sentence and classifying them into categories like a person, organisation, date, location, time etc

A.3.3.7 Sentiment Analysis

Sentiment Analysis is a broad range of subjective analysis which usesNatural Language processing techniques to perform tasks such as identifying the sentiment of a customer review, positive or negative feeling in a sentence, judging mood via voice analysis or written text analysis etc

A.3.3.8 Semantic Text Similarity

Semantic Text Similarity is the process of analysing similarity between two pieces of text with respect to the meaning and essence of the text rather than analysing the syntax of the two pieces of text. Also, similarity is different from relatedness.

A.3.3.9 Language Identification

Language identification is the task of identifying the language in which the content is in. It makes use of statistical as well as syntactic properties of the language to perform this task. It may also be considered as a special case of text classification

APPENDIX B : SAMPLE CODING

upload-file.html

```
<body style="background-color:#a3ff9d;" >
{% block title %}Upload{% endblock %}
{% block main %}
<div class="container">
  <div class="row">
    <div class="col">
<h1 align="center">Upload a file</h1>
      <hr>
<form action="/upload-file" method="POST" enctype="multipart/form-data">
<div class="form-group">

          <div class="custom-file">       
            <b>Select a file :</b>
              <input type="file" class="custom-file-input b1 " name="file" id="file">
              <h1> <div class="custom-file-label" for="file">click here to up load :
<button type="submit" class="btn btn-primary b1">Upload</button></div></h1>

          </div>
        </div>
```

```
</form>
</div>
  </div>
</div>
{% endblock %}
<style>
  label{
    display: inline-block;
    background-color: indigo;
    color: white;
    padding: 0.5rem;
    font-family: sans-serif;
    border-radius: 0.3rem;
    cursor: pointer; margin-
    top: 1rem;
  }


  .b1 {
    background-color: #4CAF50;
    border: none;
```

```
  padding: 16px 32px;
  text-align: center; text-
  decoration: none;display:
  inline-block; font-size:
  16px; margin: 4px 2px;
  transition-duration: 0.4s;
  cursor: pointer;
}

.b1 {
  background-color: white;

  border: 2px solid #4CAF50;
}
.b1:hover {
  background-color: #4CAF50;color:
  white;
}
```

```
 b{
  font-size: 30px;
 }
</style>
</body>
```

Question.html

```
<body style="background-color:#a3ff9d;" >{% block title %}Upload{% endblock
%}
{% block main %}
<div class="container"  >
  <div class="row">
    <div class="col">
<h1 align="center" >ASK A QUESTION</h1>
      <hr><br><br>
<form action="/qa" method="POST" enctype="multipart/form-data">
<div class="form-group" >
          <!-- <label>Select file</label> -->
          <div class="custom-file" >        
            <b>Enter your question here :</b>   
```

```
                        <input type="text" class="custom-file-input" name="question"
id="question">    
            <button type="submit" class="btn btn-primary b1" >submit</button>
            <br>
            <br>
            <!-- <label class="custom-file-label" for="file">Select file...</label> -->
          </div>
        </div>


<br>
<br>
<br>
<table style="width:100%">
{%for i in answer%}


  <tr>
    <td> <p style = 'border:1 px solid green';>{{i}}</p></td>


  </tr>
  {%endfor%}
</table>
```

```
 <br>


</form>
</div>
  </div>
</div>
{% endblock %}
<style>
  table, th, td { border:
    1px solid ;
  }
  .custom-file-input {
    background-color: #4CAF50; /* Green */border:
    none;
    color: white; padding:
    16px 100px;


    text-decoration: none;


    font-size: 16px;
    margin: 4px 2px
    transition-duration: 0.4s;


  }
```

```
.custom-file-input
  { background-color: white;


  border: 2px solid #4CAF50;
}
.b1 {
  background-color: #4CAF50;
  border: none;


  padding: 16px 32px;
  text-align: center; text-
  decoration: none;display:
  inline-block; font-size:
  16px; margin: 4px 2px;
  transition-duration: 0.4s;
  cursor: pointer;
}
```

```
.b1 {
  background-color: white;

  border: 2px solid #4CAF50;
}
.b1:hover {
  background-color: #4CAF50;color:
  white;
}
b{
 font-size: 30px;

}

</style>
</body>
```

QA.py

```
from haystack import Finder
from haystack.preprocessor.cleaning import clean_wiki_text
from haystack.preprocessor.utils import convert_files_to_dicts, fetch_archive_from_http
```

```
from haystack.reader.farm import FARMReader

from haystack.reader.transformers import TransformersReaderfrom

haystack.utils import print_answers

from haystack.document_store.elasticsearch import ElasticsearchDocumentStorefrom

haystack.file_converter.pdf import PDFToTextConverter

from haystack.preprocessor.preprocessor import PreProcessorfrom

haystack.retriever.sparse import ElasticsearchRetriever from

haystack.pipeline import ExtractiveQAPipeline


document_store = ElasticsearchDocumentStore(host="localhost", username="",password="",
index="document")

converter = PDFToTextConverter(remove_numeric_tables=True,
valid_languages=["en"])

model = "deepset/roberta-base-squad2"

reader = TransformersReader(model, use_gpu=0)def

response(query):

    doc_pdf = converter.convert(file_path="data/abc.pdf", meta=None)


    # This is a default usage of the PreProcessor.

    # Here, it performs cleaning of consecutive whitespaces

    # and splits a single large document into smaller documents.


     # Each document is up to 1000 words long and document breaks cannot fall inthe middle of
sentences
```

```
    # Note how the single document passed into the document gets split into 5smaller
documents


    preprocessor =

        PreProcessor( clean_empty_line

        s=True, clean_whitespace=True,

        clean_header_footer=False,

        split_by="word",

        split_length=100,

        split_respect_sentence_boundary=True

    )

    docs = preprocessor.process(doc_pdf) print(f"n_docs_input:

    1\nn_docs_output: {len(docs)}")


    """for f_name, f_path in zip(filenames, filepaths):#

        Optional: Supply any meta data here

          # the "name" field will be used by DPR if embed_title=True, rest is customand can be
named arbitrarily

        cur_meta = {"name": f_name, "category": "a" }
```

```
    # Run the conversion on each file (PDF -> 1x doc)d =
    converter.convert(f_path, meta=cur_meta)

    # clean and split each dict (1x doc -> multiple docs)d =
    processor.process(d)
        docs.extend(d)"""

document_store.write_documents(docs)

retriever = ElasticsearchRetriever(document_store)

pipe = ExtractiveQAPipeline(reader, retriever)

prediction = pipe.run(query=query, top_k_retriever=10, top_k_reader=3)#
# for i in str(print_answers(prediction, details = 'minimal')):#     return
i
     # return str(ans = []
# for i in prediction['answers']:# if
i['probability']>=0.4:
#           ans.append(i['answer'])
```

```
    # return str(ans)print_answers(prediction, details = 'minimal'))k =
    prediction['answers']
    return ("Answer:",k[1]['answer']),("Context:", k[1]['context'])
```

App.py

```
from flask import Flask, jsonify, request, render_template, redirect, make_response
import os

from flask.helpers import url_for
import hay

app = Flask(____name___)
app.config["FILE_UPLOADS"] = "/home/teja/haystack/data"@app.route("/",
methods=["GET", "POST"])
def start():
    return render_template('upload-file.html')

@app.route("/upload-file", methods=["GET", "POST"])def
upload_file():
    if request.method == "POST":
```

```
        if request.files:
            file = request.files["file"]print(file)
            file.filename = "abc.pdf" file.save(os.path.join(app.config["FILE_UPLOADS"],
            file.filename))# return redirect(request.url)
    return render_template("question.html")


@app.route("/ask-question", methods=["GET", "POST"])def
question():
    return render_template('question.html',)
@app.route("/qa", methods = ["GET", "POST"])def qa():
    a=request.form['question']ans=
    hay.response(a)
    return render_template('question.html',answer=ans)



if___name____== '___main___':
    app.run(debug=True)
```

OUTPUT/RESULT: SAMPLE SCREENSHOTS

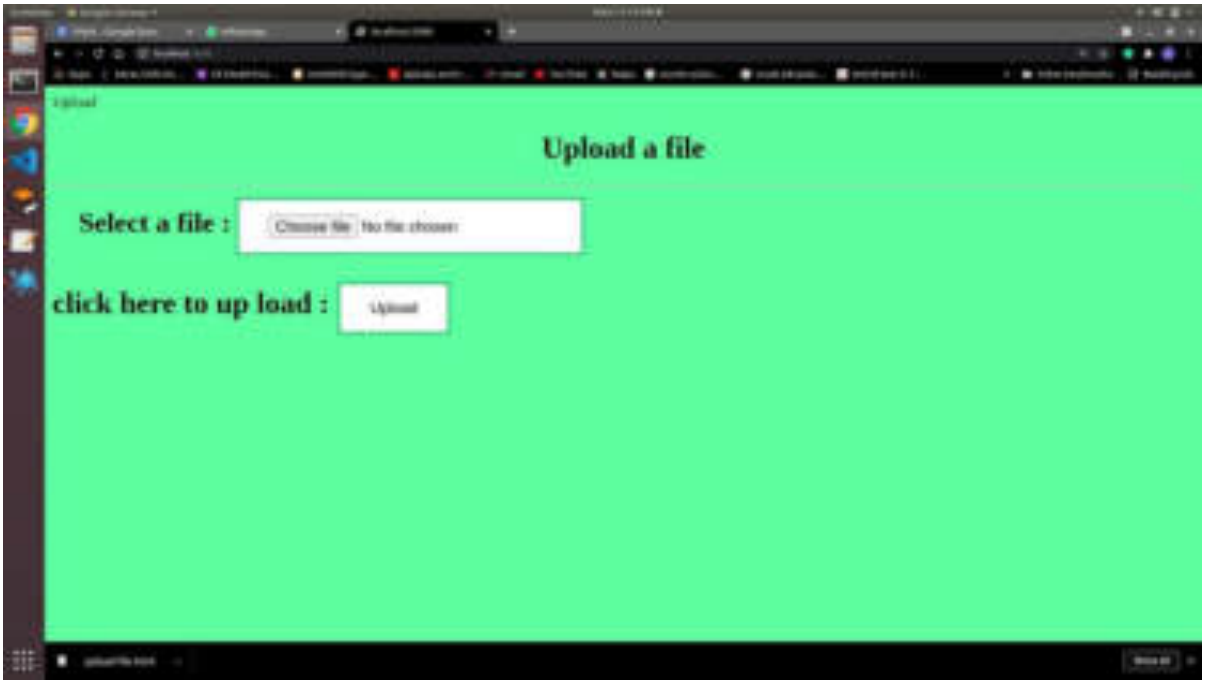


FIG 8.1 - SAMPLE SCREENSHOT 1

In the above screenshot 1, the user can upload any kind of file he wants to queryfrom and then the user is redirected to a page where he can ask his query. The document can be any type such as pdf, docx, or text file regardless of the size.

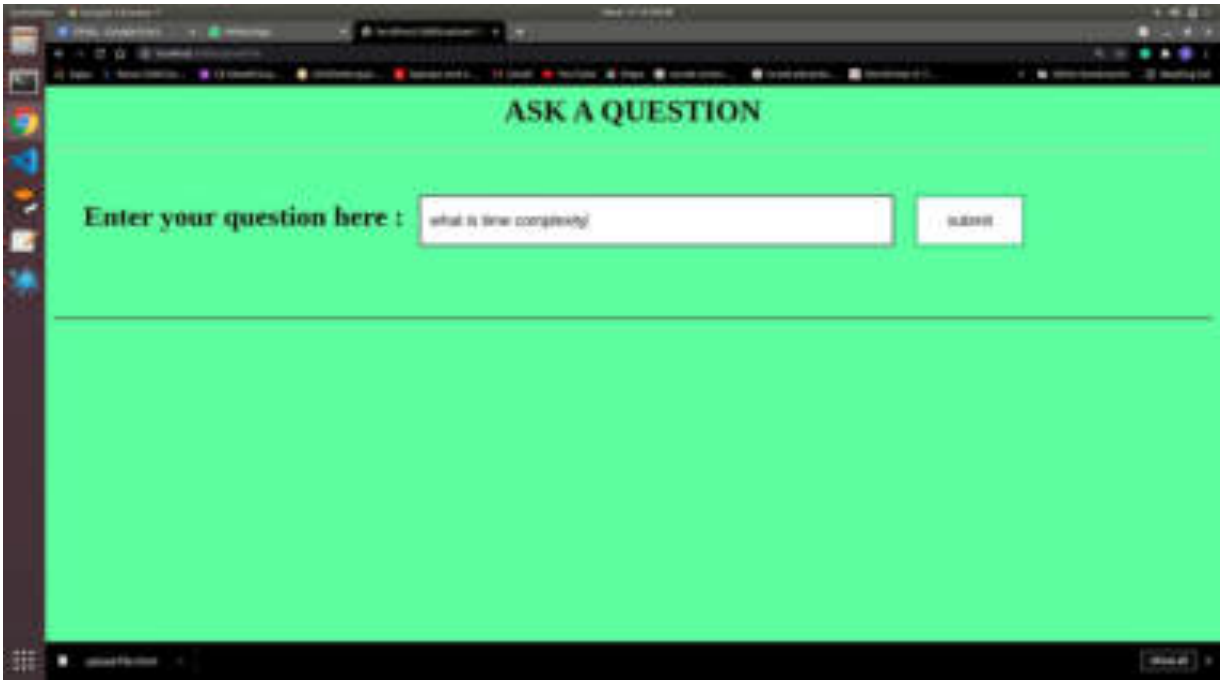


FIG 8.2 - SAMPLE SCREENSHOT 2

The screenshot 2 represents a page where the user can ask the query and the userwill be redirected to the answer page.

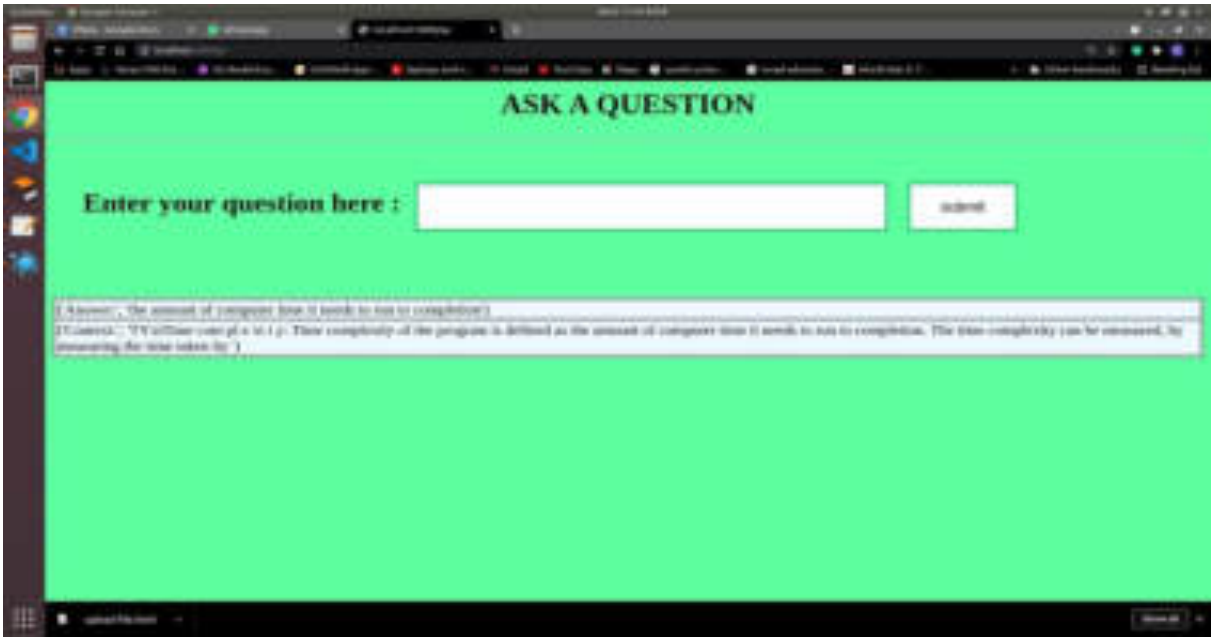


FIG 8.3 SAMPLE SCREENSHOT 3

In the above screenshot, an answer is provided along with the context for the queryasked by the user.

CHAPTER 9

REFERENCES

[1] Z. Gao, A. Feng, X. Song and X. Wu, "Target-Dependent Sentiment Classification With BERT," in IEEE Access, vol. 7, pp. 154290-154299, 2019, doi: 10.1109/ACCESS.2019.2946594.

[2] X. Tan, M. Zhuang, X. Lu and T. Mao, "An Analysis of the Emotional Evolution of Large-Scale Internet Public Opinion Events Based on the BERT-LDA Hybrid Model," in IEEE Access, vol. 9, pp. 15860-15871, 2021, doi: 10.1109/ACCESS.2021.3052566.

[3] L. Cai, Y. Song, T. Liu and K. Zhang, "A Hybrid BERT Model That Incorporates Label Semantics via Adjustive Attention for Multi-Label Text Classification," in IEEE Access, vol. 8, pp. 152183-152192, 2020, doi: 10.1109/ACCESS.2020.3017382.

[4] D. Chen, S. Zhang, X. Zhang and K. Yang, "Cross-Lingual Passage Re-Ranking With Alignment Augmented Multilingual BERT," in IEEE Access, vol. 8, pp.

[5] Y. Zhang et al., "A Question Answering-Based Framework for One-Step Event Argument Extraction," in IEEE Access, vol. 8, pp. 65420-65431, 2020, doi: 10.1109/ACCESS.2020.2985126.

[6] M. Lippi, M. A. Montemurro, M. Degli Esposti and G. Cristadoro, "Natural Language Statistical Features of LSTM-Generated Texts," in IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 11, pp. 3326-3337, Nov. 2019, doi: 10.1109/TNNLS.2019.2890970.

[7] M. T. Mills and N. G. Bourbakis, "Graph-Based Methods for Natural Language Processing and Understanding—A Survey and Analysis," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 44, no. 1, pp. 59-71, Jan. 2014, doi: 10.1109/TSMCC.2012.2227472.

[8] H. -G. Lee, Y. Jang and H. Kim, "Machine Reading Comprehension Framework Based on Self-Training for Domain Adaptation," in IEEE Access, vol. 9, pp. 21279-21285, 2021, doi: 10.1109/ACCESS.2021.3054912.

[9] M. -Y. Day and Y. -L. Kuo, "A Study of Deep Learning for Factoid Question Answering System," *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*, Las Vegas, NV, USA, 2020, pp. 419-424, doi: 10.1109/IRI49571.2020.00070.

[10] A. F. Adoma, N. -M. Henry and W. Chen, "Comparative Analyses of Bert, Roberta, Distilbert, and Xlnet for Text-Based Emotion Recognition," *2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, Chengdu, China, 2020, pp. 117-121, doi: 10.1109/ICCWAMTIP51612.2020.9317379.

[11] J. Sai Sharath and R. Banafsheh, "Question Answering over Knowledge Base using Language Model Embeddings," 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 2020, pp. 1-8, doi: 10.1109/IJCNN48605.2020.9206698

[12] N. T. M. Trang and M. Shcherbakov, "Vietnamese Question Answering System f rom Multilingual BERT Models to Monolingual BERT Model," 2020 9th International Conference System Modeling and Advancement in Research Trends (SMART), Moradabad, India, 2020, pp. 201-206, doi: 10.1109/SMART50582.2020.9337155.

L. Munikar, S. Shakya and A. Shrestha, "Fine-grained Sentiment Classification using BERT," 2019 Artificial Intelligence for Transforming

Business and Society (AITB), Kathmandu, Nepal, 2019, pp. 1-5, doi:10.1109/AITB48515.2019.8947435

[13] D. V. Vekariya and N. R. Limbasiya, "A Novel Approach for Semantic Similarity Measurement for High Quality Answer Selection in Question Answering using Deep Learning Methods," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 518-522, doi: 10.1109/ICACCS48705.2020.9074471.

[14] M. -Y. Day and Y. -L. Kuo, "A Study of Deep Learning for Factoid Question Answering System," 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI), Las Vegas, NV, USA, 2020, pp. 419-424, doi: 10.1109/IRI49571.2020.00070.

"Corn_(maize)___Healthy","Corn_(maize)___Northern_Leaf_Blight","Grape___Black_rot"

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

AUTOMOBILE DATA MANAGEMENT AND BILLING SYSTEM

Submitted to

Jawaharlal Nehru Technological University Kakinada

in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
COMPUTER SCIENCE AND ENGINEERING

Submitted by

V.RASAGNA(17H71A0599) **P.NUTHANA(17H71A0590)**

R.V.V.NAVEEN KUMAR(17H71A05B5) **M.RAHUL(17H71A0594)**

Under the Esteemed Guidance of

Mr.J.Venkata Krishna

B.Tech,M.Tech(Ph.D)

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-21

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by NAAC With Grade A

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"AUTOMOBILE DATA MANAGEMENT AND BILLING SYSTEM"** is a bonafide work carried out by **V.RASAGNA(17H71A0599), P.NUTHANA(17H71A0590), R.V.V.NAVEEN KUMAR(17H71A05B5), M.RAHUL (17H71A0594)** in partial fulfilment for the award of degree of Bachelor of Technology in **Computer Science and Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

Mr.J.Venkata Krishna
Project Guide

D.Prasad
Head of the Department

Examiner

Dr K Srinivas
PRINCIPAL

ACKNOWLEDGEMENT

We would like to take this opportunity to express our deepest appreciation to the following people for their valuable contributions and assistance with this Project.

Initially, we would like to thank our project supervisor,**Mr.J.VenkataKrishna, B.tech,M.Tech(Ph.D), Department of Computer Science and Engineering** for the guidance and support, especially for the valuable ideas and knowledge shared to us throughout the Project.

We have the immense pleasure in expressing our thanks and deep sense of gratitude to**, Mr. D. Prasad, Head of the Department, Computer Science and Engineering** for extending necessary facilities and support for the completion of the Project.

We whole sincerely acknowledge **Dr. K. Srinivas, Principal and Prof D. PANDURANGA RAO, CEO** for giving opportunity to take up the Project work. We also extend our thanks to all faculty members of **Computer Science and Engineering**, for their valuable guidance and encouragement.

We would like to extend our warm appreciation to all our friends for sharing us their knowledge, valuable contributions and help with this Project.

Finally, our special thanks go to our family for their continuous support and help throughout our academic years and for their continual support and encouragement for the completion of the project.

DECLARATION

We **V.RASAGNA(17H71A0599), P.NUTHANA(17H71A0590), R.V.V.NAVEEN KUMAR(17H71A05B5), M.RAHUL (17H71A0594)** of the Main-Project "**AUTOMOBILE DATA MANAGEMENT AND BILLING SYSTEM**", hereby declare that the matter embodied in this Project is the genuine work done by us and has not been submitted either to this university or to any other university/institute for the fulfilment of the requirement of any course of study.

PROJECT MEMBERS

V.RASAGNA	**(17H71A0599)**
P.NUTHANA	**(17H71A0590)**
R.V.V.NAVEEN KUMAR	**(17H71A05B5)**
M.RAHUL	**(17H71A0594)**

CONTENTS

ABSTRACT

The Automobile Data Management and Billing System deals with automobile spare parts data processing like stock of available products,to add new products,to update price of products,to delete products and to calculate total price of all products purchased by the customer. It is designed with a goal to making the existing system more informative, reliable, fast and easier. Many stores have a lot difficulty to manage thebusiness,customer ordering ,supplying the items and the bill payment process. By using pen and paper it is difficult to calculate the total bills of all the customers .The manual System of the Automobile store are maintained by handwritten documents/records which is very untidy and clumsy process the owner has difficult to known about profit and loss information. Less efficiency has a great impact on the productivity in order to increase sales,it is very important to maintain data up-to-date.

To overcome these limitation of the existing system, the proposed system has been evolved. The main aim of the project is to reduce the paper work and saving time to generate accurate results for the customers. It's easy for the user to meet the challenges of managing their records and retrieval of useful information. The system revolves around the menu of the items available. The project is very flexible because it allows the user to get information in various possible categories. Everything is done dynamically. All the details are stored into appropriate database. The entered product details can be modified, deleted and viewed by the administrator. Finally the efficient reports can be generated by using this proposed system.

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

1.1 Introduction

The Automobile Data Management and Billing System is used to manage the stock of items in the Automobile store, and also focuses on calculating the total bill of items purchased by the customer in the Automobile store and it also sends a whatsapp message of bill details after generating the bill to the customer. In this implementation of Billing system the bill calculation includes all the items of the Automobile spare parts like rods,wheels,nuts,lights,horns etc.

Automobile Billing Systems provides the information about the items available in the Automobile store in order to create customer friendly atmosphere, generally by displaying the menu the user have the clear picture about the things he/she needs and it saves lot of time for customer and once the shopping is done the customer is ready to take the bill , the bill will be displayed as item name cost and the total bill is displayed.

Automobile Data Management and Billing System focuses Mainly on Managing the stock of Products in the store such as displaying the available products,updates the price and quantity of the products,searching for items and also possible to delete the unavailable products in the store.

The Automobile Data Management and Billing System have many ad vantages because it displays the menu and calculates the bill under each section we can avoid any mathematical errors, the customer haves the clarity regarding the items taken and the amount charged per each item, with this we can also store the data of the customer information, stock sold per day which helps to maintain the stock necessity.

1.2 Problem Statement:

In this first we create a role's in the Automobile store. we can even search and display the customer records. And we will create a bills file of the customer update their bills into this file and can also generate those records. Next verify whether the has paid the bill or not. if not we will send a notification to them regarding that issue. The project basically deals with the billing system.

1.3 Scope:

Our project has a big scope to do. We can:

1.Calculate the bill.

2.Give the bill to the customer.

3.Store products and their prices.

4.Can see the report of all the products stored in the super market.

This project is basically developed as a solution to the drawbacks of existing system. This project can be used as a real-world application and by any organization.

1.4 Purpose:

To make software fast in processing., with good user interface so that the user can change it should be used for a long time without error.

1.5 Existing System:

The Automobile Data Management and Billing System is to deals with Automobile automation and it includes both purchasing a selling of items. It is designed with a goal to making the existing system more informative, reliable, fast and easier. Many stores have a lot difficult to manage the business and customer ordering and supplying the items and the bill payment process. By using manual calculation is difficult to accountant. The manual System the Automobile store are inefficiencies Large records-books have to be maintained where relevant and irrelevant information has to be stored which is very untidy and clumsy process the owner has difficult to known about profit and loss information. Less efficiency has a great impact on the productivity of any human being keeping the data up-to-date.

- There is no consistency as the data may not be updated on time.
- Feasibility is reduced.
- Less reliability.
- Security is not provided, and anyone can access.
- Prioritization of records is difficult.
- Difficult to maintain.
- As everything is done manually its slow process.
- No timely acknowledgement services.

1.6 Proposed system:

The proposed system is developed to overcome the limitations of the existing system, the proposed system has been evolved. The main aim of the project is to reduce the paper work and saving time to generate accurate results for the customers. It's easy for the user to meet the challenges of managing their records and retrieval of useful information. The system revolves around the menu of the items available. The project is very flexible because it allows the user to get information in various possible categories. Everything is done dynamically. All the details are stored into appropriate database. The entered product details can be modified, deleted and reported to the administrator. This system provides the best user interface. Finally the efficient reports can be generated by using this proposed system. The objectives of the proposed system are as follows:

- Easy to use, effective and efficient
- Accurate results.
- Easy maintenance.
- Fast access.
- More feasibility.
- More secure.
- Error reduction.
- Reduction in the use of paper.
- Reduction in man power.

ANALYSIS

2.1 Requirement Specifications:

Software Requirements:

Operating system	:	Windows 10
Programming Language	:	Python
Database Technology	:	MySQL
IDE	:	PyCharm

Hardware Requirements:

Processor	:	I3 and above
RAM	:	4GB RAM (Preferable)
Hard disk	:	8 GB or more.

2.2 Functional Requirements:

Modules

Admin:

This module is considered only when there is a requirement of safety and security by the User. Only after the login process, the rest of the application is made available to the user.

To login, user has to first register manually by providing desired user-ID and password. Provided user-IDs and passwords by the users are maintained in a database. Then the user logins in to the application by giving user-ID and password provided.

1.To provide Security

2.To manage the stock in the automobile store.

Managing of products includes ;

1. adding new products.
2. Update the price.
3. Update the quantiy.
4. Display all the products.
5. Search for products.

6. Delete unavailable products.

User Login Sequence will be as follows:

- User registration by providing user-ID and password.
- User will be prompted to enter user credentials.
- If the user enters correct credentials, then he gets access.

ValidateCashier:

This module is considered only when there is a requirement of safety and security by the Cashier. Only after the login process, the rest of the application is made available to the cashier. To login, cashier has to first register manually by providing desired cashir-ID and password. Provided cashier-IDs and passwords by the users are maintained in a database. Then the cashier logins in to the application by giving cahier-ID and password provided.

- To provide security.
- To calculates bill of all items purchased by customer.
- Sends a message to customer about bill details.

Billing:

This is one of the module which is used to calculate the bill to all the items purchased by the customer in the Automobile store.It is provided with fuctions like getting price of the products,getting quantity of the products and also calculation of the bill.Another important function is to update the quantity of the corresponding products after selling.The bill Calculation is done by the using product id which is the primary key that is used for unique identification of the available products.

Message:

Message module is used to send a whatsapp message to the customer about bill details.Message is sent by using time zone,hour and minute of the country.

2.3 Non-Functional Requirements:

- The user interface should be easy for users to operate without additional training.
- The response time should be fast.
- The System should allow large number of concurrent users.

2.4 Software models:

The different types of project models available to develop a software are:

- Classical waterfall model
- Iterative waterfall model
- Agile model
- Incremental model
- RUP Process model
- Spiral model
- Prototype model

2.5 Feasibility Study:

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

Economic Feasibility

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus, the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

Technical Feasibility

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

Social Feasibility

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

DESIGN

3.1 UML DIAGRAMS :

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object-oriented computer

software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML. The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems. The UML is a very important part of developing objects-oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

3.1.2 GOALS:

The Primary goals in the design of the UML are as follows:

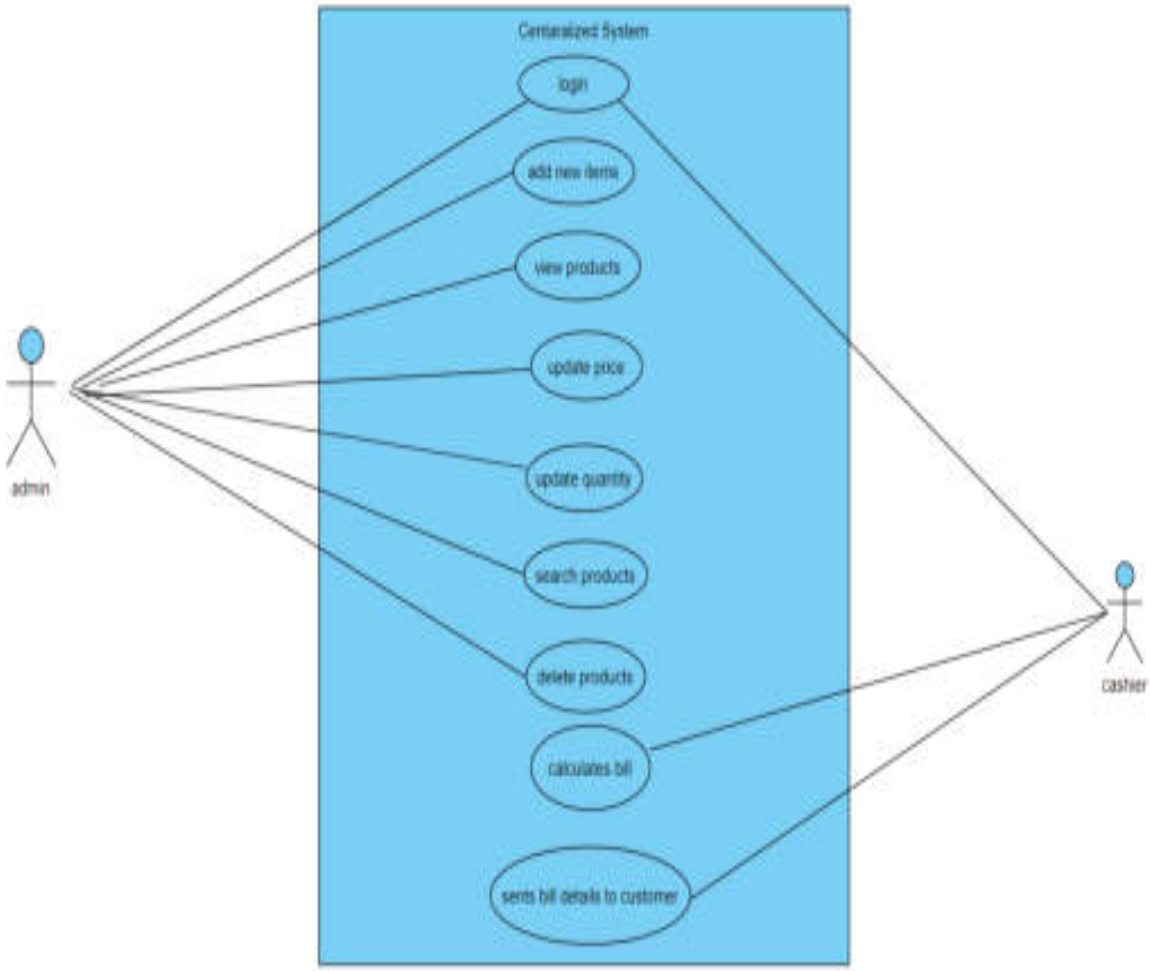
1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and

exchange meaningful models.

2. Provide extensibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding the modeling language.
5. Encourage the growth of OO tools market.
6. Support higher level development concepts such as collaborations, frameworks, patterns and components.
7. Integrate best practices.

3.2 USE CASE DIAGRAM:

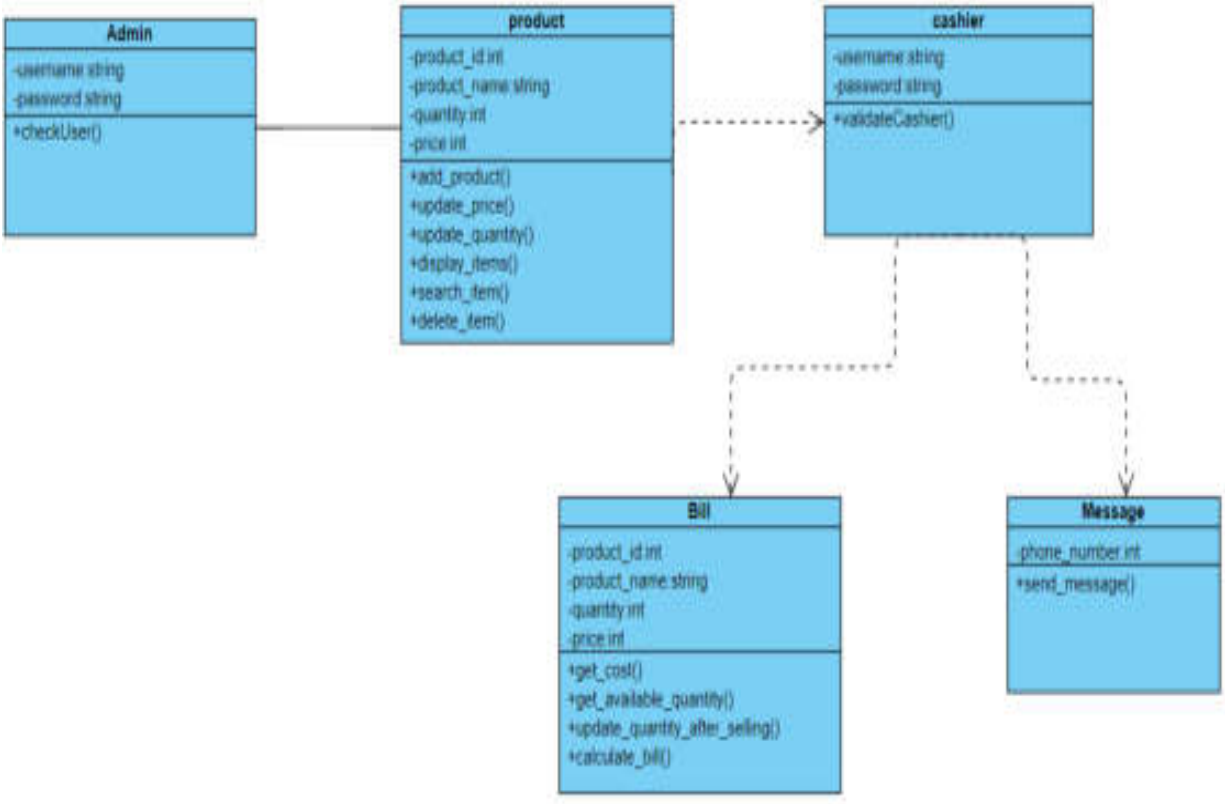
A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



3.3 CLASS DIAGRAM:

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static

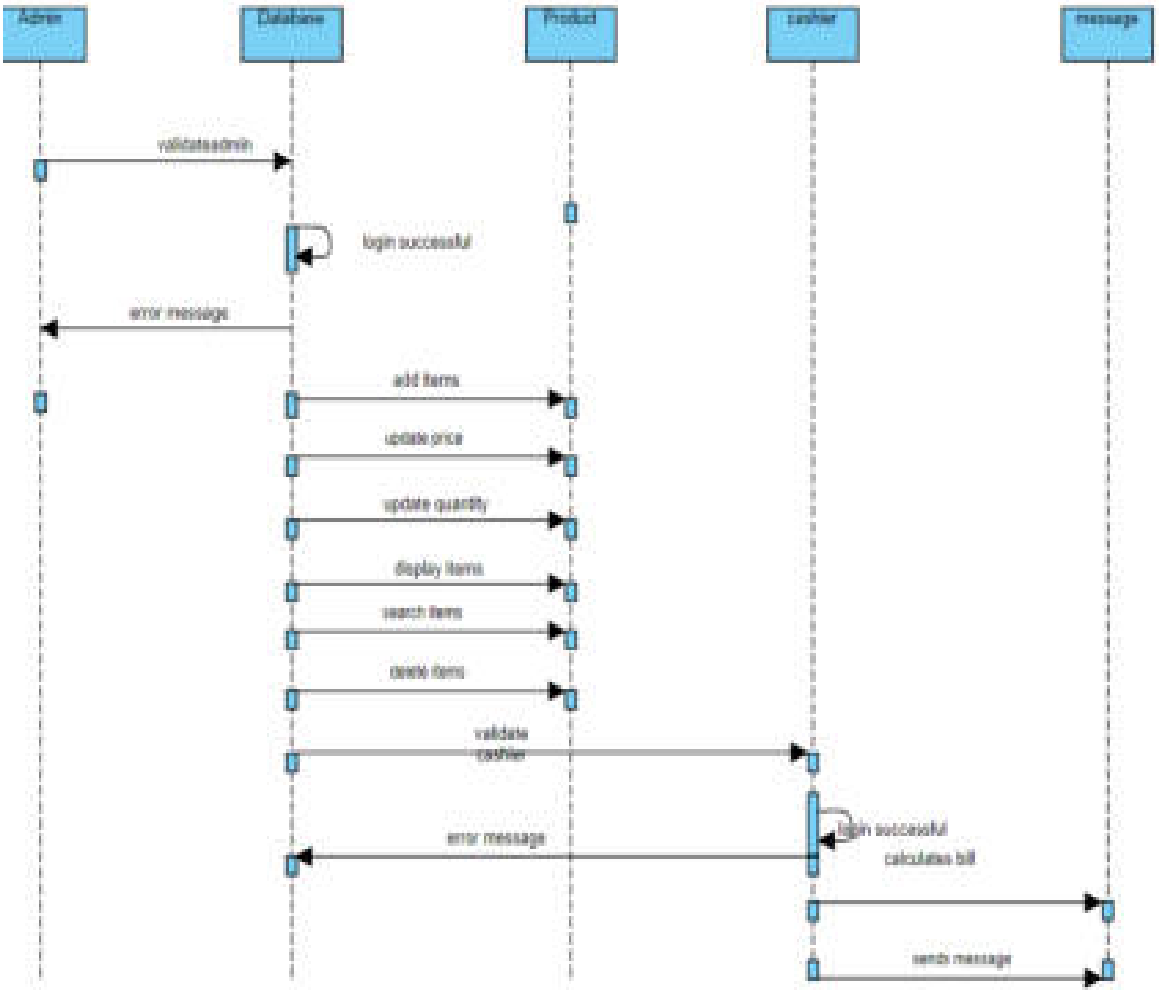
structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.



3.4 SEQUENCE DIAGRAM:

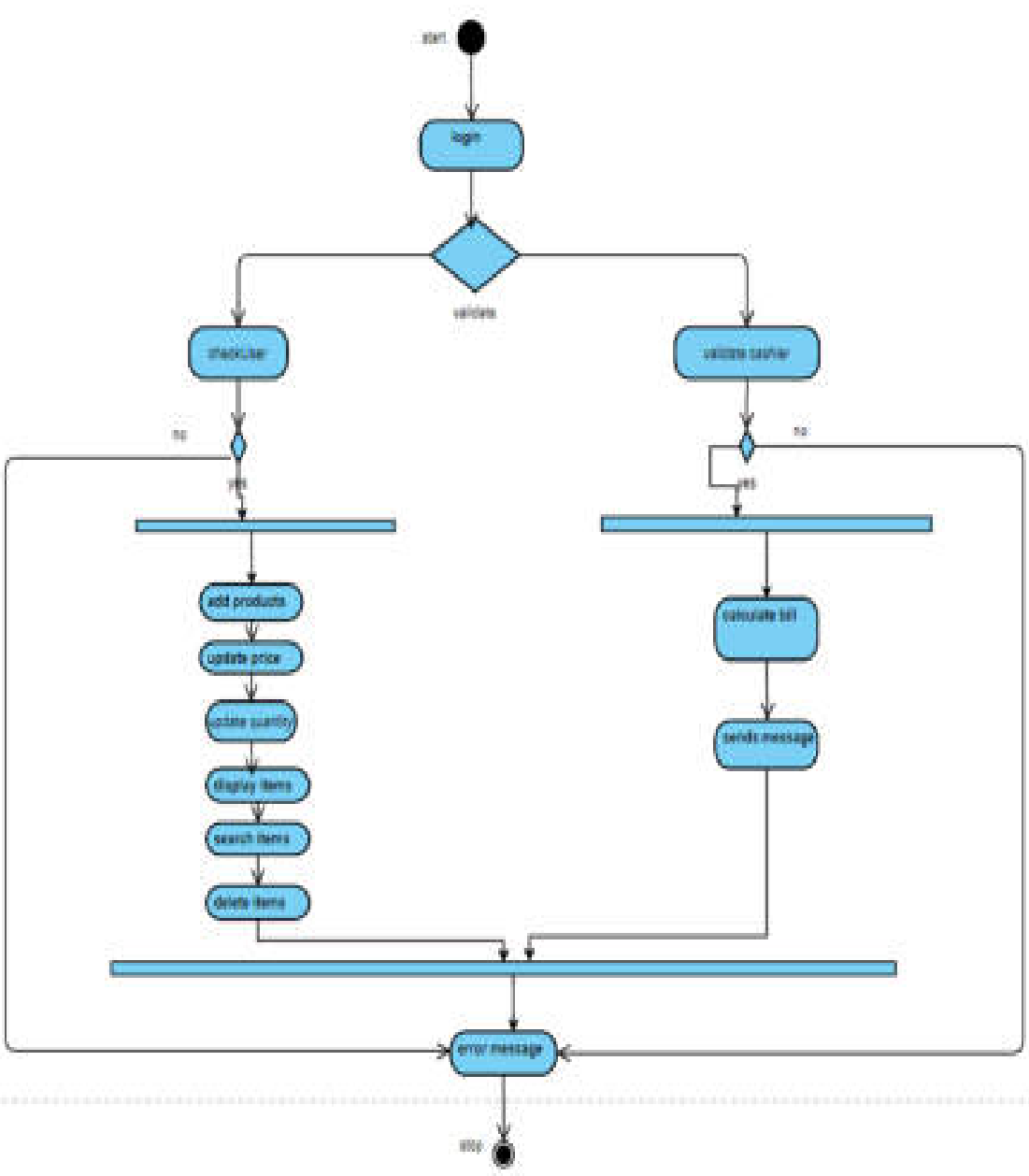
A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows

how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.



3.5 ACTIVITY DIAGRAM:

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.



IMPLEMENTATION

4.1 Introduction and Module Description:

PROJECT MODULES

4.1.1 Name of the Module-1: Admin Module

This module is considered only when there is a requirement of safety and security by the User. Only after the login process, the rest of the application is made available to the user.

To login, user has to first register manually by providing desired user-ID and password. Provided user-IDs and passwords by the users are maintained in a database. Then the user logins in to the application by giving user-ID and password provided.

1.To provide Security

2.To manage the stock in the automobile store.

Managing of products includes ;

1. adding new products.
2. Update the price.
3. Update the quantiy.
4. Display all the products.
5. Search for products.
6. Delete unavailable products.

4.1.2 Name of the Module-2: ValidateCashier Module

This module is considered only when there is a requirement of safety and security by the Cashier. Only after the login process, the rest of the application is made available to the cashier. To login, cashier has to first register manually by providing desired cashir-ID and password. Provided cashier-IDs and passwords by the users are maintained in a database. Then the cashier logins in to the application by giving cahier-ID and password provided.

- To provide security.
- To calculates bill of all items purchased by customer.
- Sends a message to customer about bill details.

4.1.3 Name of the Module-3: Billing Module

This is one of the module which is used to calculate the bill to all the items purchased by the customer in the Automobile store.It is provided with fuctions like getting price of the products,getting quantity of the products and also calculation of the bill.Another important function is to update the quantity of the corresponding products after selling.The bill Calculation is done by the using product id which is the primary key that is used for unique identification of the available products.

4.1.4 Name of the Module-4: Message Module

Message module is used to send a whatsapp message to the customer about bill details.Message is sent by using time zone,hour and minute of the country.

4.2 OVERVIEW OF THECHNOLOGIES USED

4.2.1 Python Technology:

Python is an interpreted high-level general-purpose programming language. ... It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented and functional programming.

The Python Programming Language

Python is dynamically-typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library.

Guido van Rossum began working on Python in the late 1980s, as a successor to the ABC programming language, and first released it in 1991 as Python 0.9.0. Python 2.0 was released in 2000 and introduced new features, such as list comprehensions and a garbage collection system using reference counting. Python 3.0 was released in 2008 and was a major revision of the language that is not completely backward-compatible and much Python 2 code does not run unmodified on Python 3. Python 2 was discontinued with version 2.7.18 in 2020.

Python consistently ranks as one of the most popular programming languages.

Features in Python:

- Easy to code.
- Free and Open Source.
- Object-Oriented Language.
- GUI Programming Support.
- High-Level Language.
- Extensible feature.
- Python is Portable language.
- Python is Integrated language.

What can Python do?

- Python can be used on a server to create web applications.
- Python can be used alongside software to create workflows.
- Python can connect to database systems. It can also read and modify files.
- Python can be used to handle big data and perform complex mathematics.
- Python can be used for rapid prototyping, or for production-ready software development.

Why Python?

- Python works on different platforms (Windows, Mac, Linux, Raspberry Pi, etc).
- Python has a simple syntax similar to the English language.
- Python has syntax that allows developers to write programs with fewer lines than some other programming languages.

- Python runs on an interpreter system, meaning that code can be executed as soon as it written. This means that prototyping can be very quick.

- Python can be treated in a procedural way, an object-oriented way or a functional way.

4.2.2 Database Technology - MySQL:

What Is a Database?

A database is an organized collection of structured information, or data, typically stored electronically in a computer system. A database is usually controlled by a **database management system (DBMS)**. Together, the data and the DBMS, along with the applications that are associated with them, are referred to as a database system, often shortened to just database.

Data within the most common types of databases in operation today is typically modeled in rows and columns in a series of tables to make processing and data querying efficient. The data can then be easily accessed, managed, modified, updated, controlled, and organized. Most databases use structured query language (SQL) for writing and querying data.

What is database software?

Database software is used to create, edit, and maintain database files and records, enabling easier file and record creation, data entry, data editing, updating, and reporting. The software also handles data storage, backup and reporting, multi-access control, and security. Strong database security is especially important today, as data theft becomes more frequent. Database software is sometimes also referred to as a “database management system” (DBMS).

Database software makes data management simpler by enabling users to store data in a structured form and then access it. It typically has a graphical interface to help create and manage the data and, in some cases, users can construct their own databases by using database software.

What is a database management system (DBMS)?

A database typically requires a comprehensive database software program known as a database management system (DBMS). A DBMS serves as an interface between the database and its end users or programs, allowing users to retrieve, update, and manage how the information is organized and optimized.

A DBMS also facilitates oversight and control of databases, enabling a variety of administrative operations such as performance monitoring, tuning, and backup and recovery.Some examples of popular database software or DBMSs include MySQL, Microsoft Access, Microsoft SQL Server, FileMaker Pro, Oracle Database, and dBASE.

What is a MySQL database?

MySQL is an open source relational database management system based on SQL. It was designed and optimized for web applications and can run on any platform. As new and different requirements emerged with the internet, MySQL became the platform of choice for web developers and web-based applications. Because it's designed to process millions of queries and thousands of transactions, MySQL is a popular choice for ecommerce businesses that need to manage multiple money transfers. On-demand flexibility is the primary feature of MySQL.

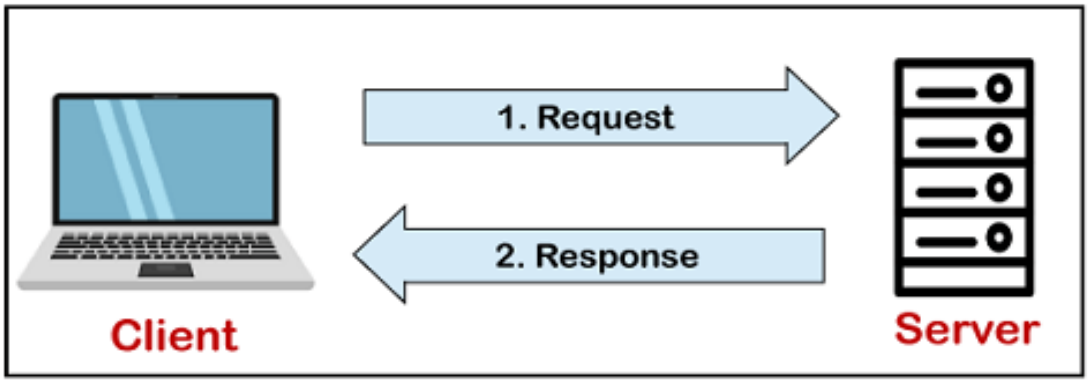
MySQL is the DBMS behind some of the top websites and web-based applications in the world, including Airbnb, Uber, LinkedIn, Facebook, Twitter, and YouTube.

MySQL is a Relational Database Management System (RDBMS) software that provides many things, which are as follows:

- It allows us to implement database operations on tables, rows, columns, and indexes.
- It defines the database relationship in the form of tables (collection of rows and columns), also known as relations.
- It provides the Referential Integrity between rows or columns of various tables.
- It allows us to updates the table indexes automatically.
- It uses many SQL queries and combines useful information from multiple tables for the end-users.

How MySQL Works?

MySQL follows the working of Client-Server Architecture. This model is designed for the end-users called clients to access the resources from a central computer known as a server using network services. Here, the clients make requests through a graphical user interface (GUI), and the server will give the desired output as soon as the instructions are matched. The process of MySQL environment is the same as the client-server model.



The core of the MySQL database is the MySQL Server. This server is available as a separate program and responsible for handling all the database instructions, statements, or commands. The working of MySQL database with MySQL Server are as follows:

1. MySQL creates a database that allows you to build many tables to store and manipulate data and defining the relationship between each table.
2. Clients make requests through the GUI screen or command prompt by using specific SQL expressions on MySQL.
3. Finally, the server application will respond with the requested expressions and produce the desired result on the client-side.

A client can use any MySQL GUI. But, it is making sure that your GUI should be lighter and user-friendly to make your data management activities faster and easier. Some of the most widely used MySQL GUIs are MySQL Workbench, SequelPro, DBVisualizer, and the Navicat DB Admin Tool. Some GUIs are commercial, while some are free with limited functionality,.

DATABASE

TABLES:

5.1 Automobile_stock_db Table:

Columnname	Type	Null	key
Product id	int	No	primary
Product name	Varchar	No	-
price	int	No	-
Quantity	int	No	-

5.2 User_Credentials Table:

Column name	Type	Null	key
username	varchar	No	primary
password	varchar	No	-

5.3 Validate_cashier Table:

Column name	Type	Null	key
username	varchar	No	primary
password	varchar	No	-

CODE

Main.py:

```
import class1

import billing
import user
import validateCashier

class1_product = class1.Product()
billing_bill = billing.Bill()

ch = 'y'

choiceMain = input("Press \n1:Manage Products \n2:Cashier")
if choiceMain == '1':
    valid = user.checkUser()
    if (valid):
        while ch != 'n':
            choice1 = input(
                "Press \n1:To add new items \n2:To update price \n3:To update quantity \n4:TO display all the products \n5:To search the item \n6:To delete the item\n"
            )

            if (choice1 == '1'):
                class1_product.add()

            elif choice1 == '2':
                class1_product.update_price()

            elif choice1 == '3':
                class1_product.update_quantity()

            elif choice1 == '4':
```

```
            class1_product.display()

        elif choice1 == '5':
            class1_product.search()

        else:
            class1_product.delete()
        ch = input("Do you want to do one more operation(y/n) : ")

    else:
        print("Invalid Username or Password")

elif choiceMain == '2':
    valid = validateCashier.checkCashier()
    if (valid):
        while ch != 'n':
            billing_bill.calculate_bill()
            ch = input("Do you want to do one more operation(y/n) : ")
    else:
        print("Invalid Username or Password")
```

Class1.py:

```
import mysql.connector

class Product:

    def add(self):
        ch = 'y'
        while ch != 'n':
            product_id = input("Enter the Product Id:")
            product_name = input("Enter the Product Name:")
            price = input("Enter the Price:")
            quantity = input("Enter the Quantity:")
```

```
        mydb = mysql.connector.connect(
            host="localhost",
            user="root",
            password="mypassword",
            database="project")
        mycursor = mydb.cursor()

        insert = ("insert into automobile_stock_db(product_id,product_name,price,quantity)"
                "values (%s,%s,%s,%s)"
                )
        data = (product_id, product_name, price, quantity)
        mycursor.execute(insert, data)
        mydb.commit()

        print("Data Successfully Inserted into DB")

        ch = input("Do you want to add annother item (y/n)")


def update_price(self):
    ch1 = 'y'
    while ch1 != 'n':
        product_id = input("Enter the Product Id:")
        price = input("Enter the Price:")
        mydb = mysql.connector.connect(
            host="localhost",
            user="root",
            password="mypassword",
            database="project")
        mycursor = mydb.cursor()

        update_price = "update automobile_stock_db set price=%s where product_id=%s"
        values = (price, product_id)
```

```
            mycursor.execute(update_price, values)
            mydb.commit()
            print(mycursor.rowcount, "records successfully updated")

            ch1 = input("Do you want to update price of annother item (y/n)")



    def update_quantity(self):
        ch2 = 'y'
        while ch2 != 'n':
            product_id = input("Enter the product Id:")
            quantity = input("Enter the quantity:")
            mydb = mysql.connector.connect(
                host="localhost",
                user="root",
                password="mypassword",
                database="project")
            mycursor = mydb.cursor()

            update_quantity = "update automobile_stock_db set quantity=%s where
product_id=%s"
            values = (quantity, product_id)

            mycursor.execute(update_quantity, values)
            mydb.commit()
            print(mycursor.rowcount, "records successfully updated")

            ch2 = input("Do you want to update quantity of annother item (y/n)")



    def display(self):
        mydb = mysql.connector.connect(
```

```
        host="localhost",
        user="root",
        password="mypassword",
        database="project")

        mycursor = mydb.cursor()

        mycursor.execute("SELECT * FROM automobile_stock_db")

        myresult = mycursor.fetchall()

        print("(PID , ITEM , PRICE , QUANTITY)")
        for x in myresult:
            print(x)



    def search(self):
        ch3 = 'y'
        while ch3 != 'n':
            product_id = input("Enter the Product Id:")
            mydb = mysql.connector.connect(
                host="localhost",
                user="root",
                password="mypassword",
                database="project")
            mycursor = mydb.cursor()

            search = "select * from automobile_stock_db where product_id=%s"
            values = (product_id,)

            mycursor.execute(search, values)
            result = mycursor.fetchall()
            print("(PID , ITEM , PRICE , QUANTITY)")
            print(result)
```

```
            print("Records successfully fetched")

            ch3 = input("Do you want to search annother item (y/n)")



    def delete(self):
        ch4 = 'y'
        while ch4 != 'n':
            product_id = input("Enter the Product Id to be deleted:")
            mydb = mysql.connector.connect(
                host="localhost",
                user="root",
                password="mypassword",
                database="project")
            mycursor = mydb.cursor()

            delete = "delete from automobile_stock_db where product_id=%s"
            values = (product_id,)

            mycursor.execute(delete, values)
            mydb.commit()
            print(mycursor.rowcount, "record(s) successfully deleted")
            ch4 = input("Do you want to update annother item (y/n)")
```

User.py:

```
import mysql.connector

conn = mysql.connector.connect(
    host="localhost",
    user="root",
    password="mypassword",
    database="project")
```

```
cur = conn.cursor()



def checkUser():
    userName = input("Enter User Name :")
    password = input("Enter Password :")

    mypassword_queue = []

    cur.execute("select password from user_credentials where username='{}'".format(userName))
    rows = cur.fetchall()
    for row in rows:
        for x in row:
            mypassword_queue.append(x)
    if password in mypassword_queue:
        return True
    return False
```

validateCashier.py:

```
import mysql.connector

conn = mysql.connector.connect(
    host="localhost",
    user="root",
    password="mypassword",
    database="project")
cur = conn.cursor()



def checkCashier():
    userName = input("Enter User Name :")
    password = input("Enter Password :")

    mypassword_queue = []
```

```
    cur.execute("select password from validate_cashier where username='{}'".format(userName))
    rows = cur.fetchall()
    for row in rows:
        for x in row:
            mypassword_queue.append(x)
    if password in mypassword_queue:
        return True
    return False
```

Billing.py:

```
    import mysql.connector
    import message

    conn = mysql.connector.connect(
        host="localhost",
        user="root",
        password="mypassword",
        database="project")
    cur = conn.cursor()

    class Bill:

        def get_cost(self,product_id):
            cur.execute("select * from automobile_stock_db where
product_id='{}'".format(product_id))
            rows = cur.fetchone()
            return (rows[2]) # it will return price

        def get_available_quantity(self,product_id):
            cur.execute("select * from automobile_stock_db where
product_id='{}'".format(product_id))
```

```
            rows = cur.fetchone()
            return (rows[3])  # it will return quantity

        def update_quantity_after_selling(self,product_id,quantity,check_quantity):
            new_quantity=check_quantity-quantity
            cur.execute("update automobile_stock_db set quantity='{}' where
product_id='{}'".format(new_quantity,product_id))
            conn.commit()

        def calculate_bill(self):
            ch = 'y'
            bill = 0
            product_id = []
            quantity = []
            p_price = []
            t_price = []
            items = 0
            while ch != 'n':
                pid = int(input("enter product id"))
                qty = int(input("enter quantity"))
                check_quantity = self.get_available_quantity(pid)
                if qty <= check_quantity:
                    items = items + 1
                    self.update_quantity_after_selling(pid,qty,check_quantity)
                    price = self.get_cost(pid)
                    product_id.append(pid)
                    quantity.append(qty)
                    p_price.append(price)
                    t_price.append(price * qty)
                    bill += qty * price
                    ch = input("do you want to buy another")
                else:
                    ch='n'
                    print("Quantity Not available for product id",pid)
```

```
        if items != 0:
            print("--------------------------------------------------------------------------------")
            print("                              BILLING                                 ")
            print("--------------------------------------------------------------------------------")
            print("pid | quantity | price per single item | Total item price for purchased quantity")
            print("--------------------------------------------------------------------------------")
            loop = 0
            while loop < items:
                print(str(product_id[loop]) + "      " + str(quantity[loop]) + "           " + str(
                    p_price[loop]) + "                    " + str(t_price[loop]))
                loop = loop + 1
            print("--------------------------------------------------------------------------------")

            print("                         Total Bill :    ", bill)
            print("--------------------------------------------------------------------------------")
            message.send_message(bill)
```

Message.py:

```
import pywhatkit
from datetime import datetime
import pytz

# it will get the time zone of the specified location
IST = pytz.timezone('Asia/Kolkata')



def send_message(total_bill):
    phonenumber = input("Enter Customer Mobile Number")

    hour = datetime.now(IST).hour
    minute = (datetime.now(IST).minute) + 1
```

```
try:
    msge = "      Welcome to Ebazar Shopping  \n  Your total purchased amount is " + str(
        total_bill) + "  \n            THANK YOU  \n       Please visit again!!"
    pywhatkit.sendwhatmsg(f"+91{phonenumber}", msge, hour, minute)
    print("Successfully Sent!")

except:
    print("An Unexpected Error occured while sending whatsapp message!")
```

OUTPUT SCREENS

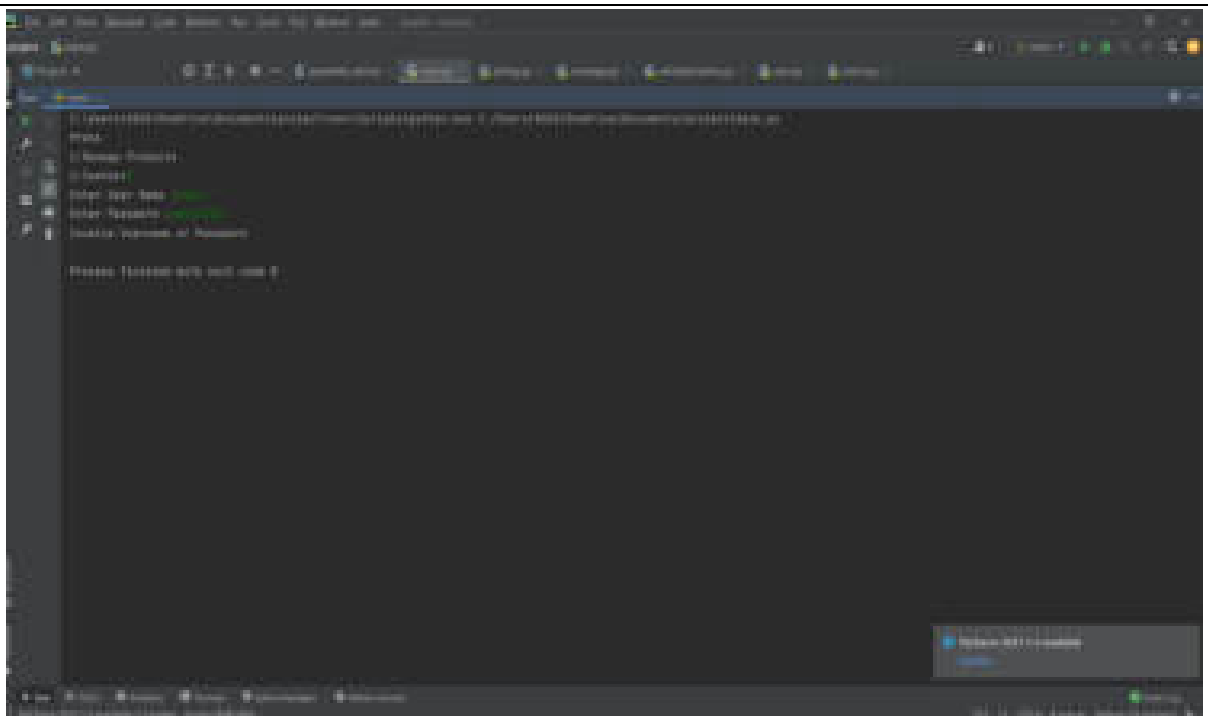


Fig.7.1 Output screen for Admin login

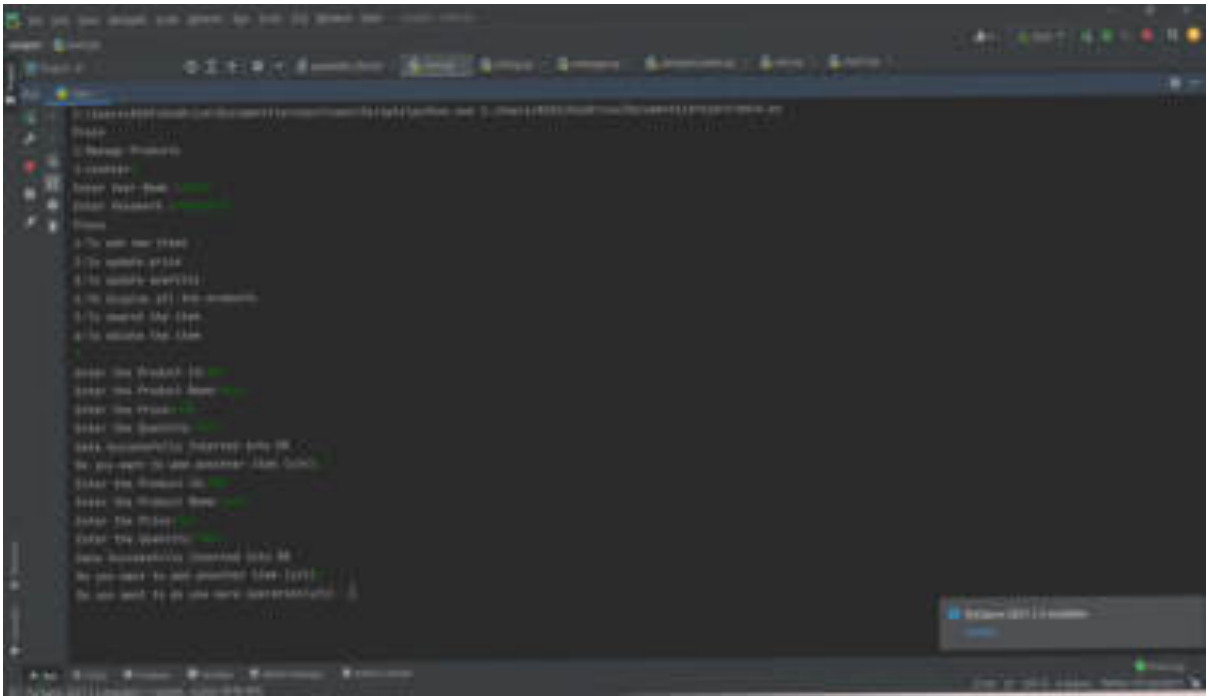


Fig.7.2 Output screen for adding new products.

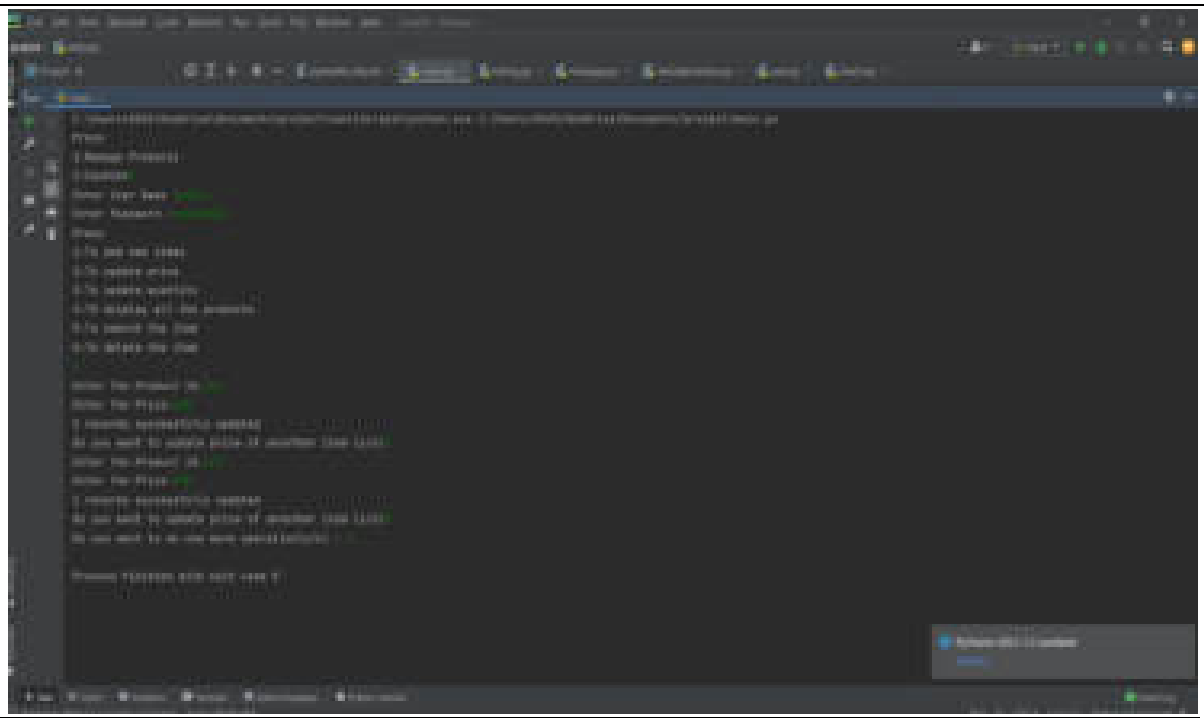


Fig.7.3 Output screen for updating price of products.

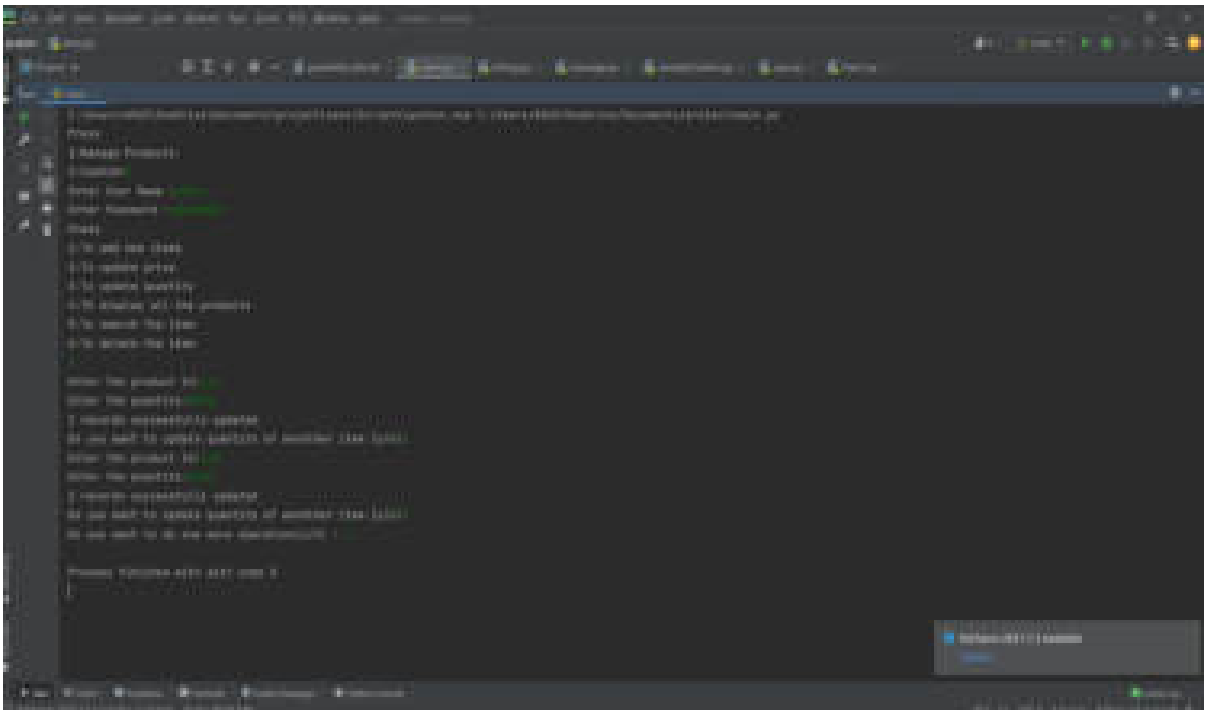


Fig.7.4 Output screen for updating quantity of products.

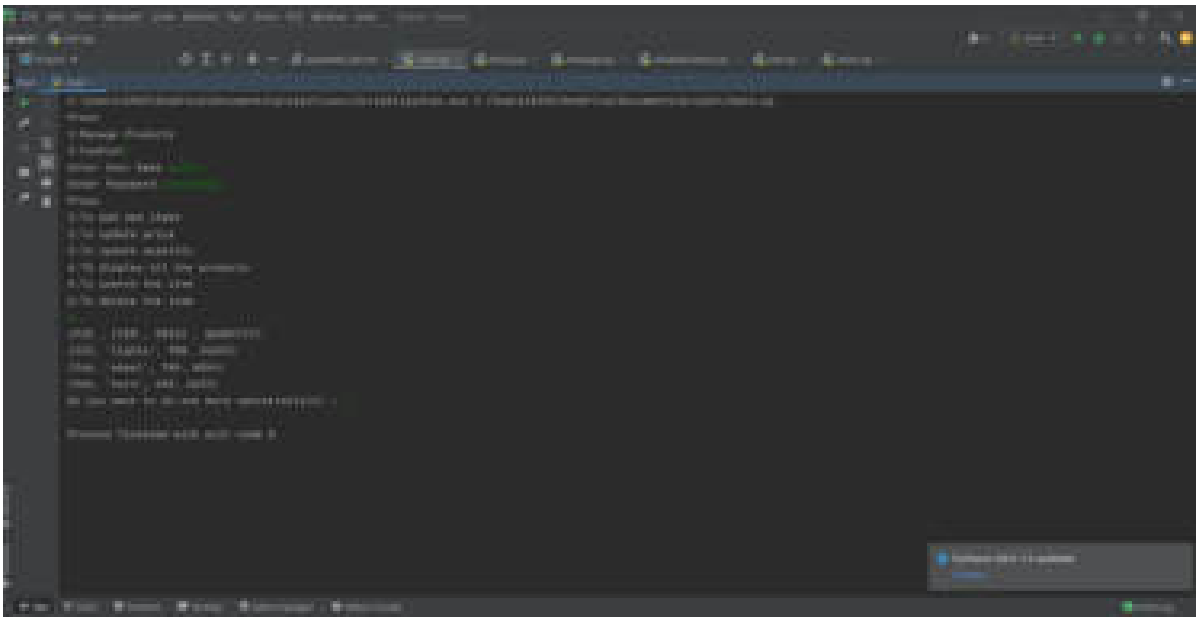


Fig.7.5 Output screen for displaying products.

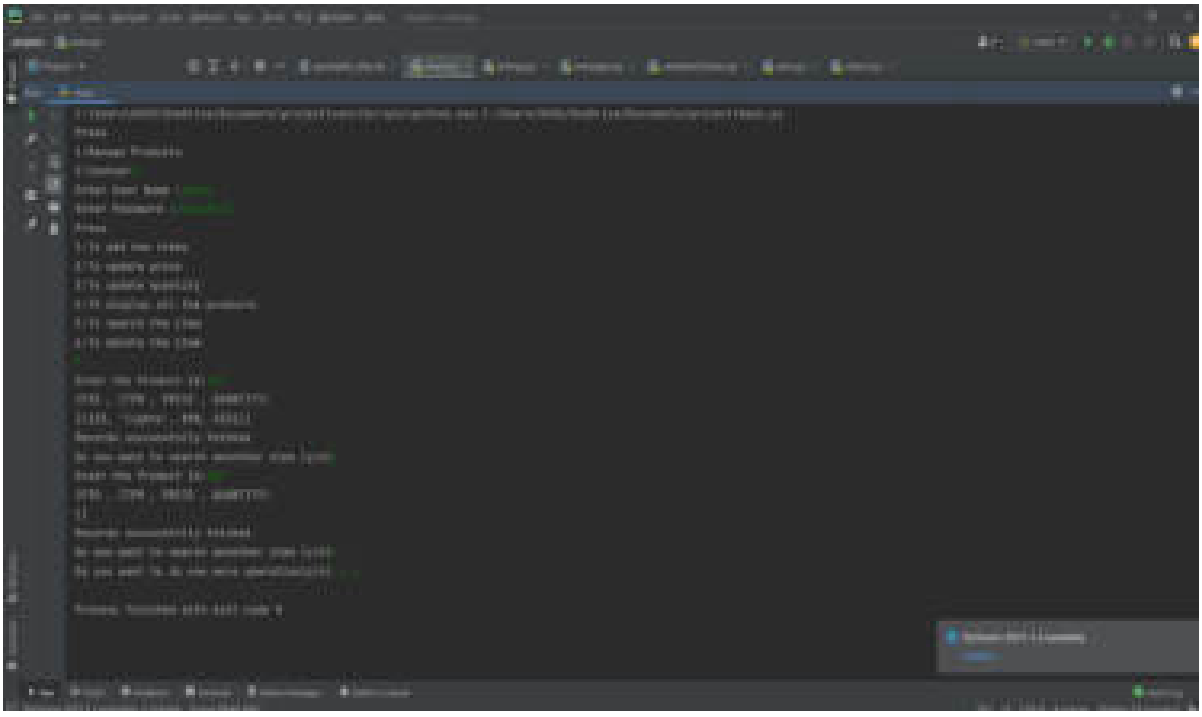


Fig.7.6 Output screen for searching products.

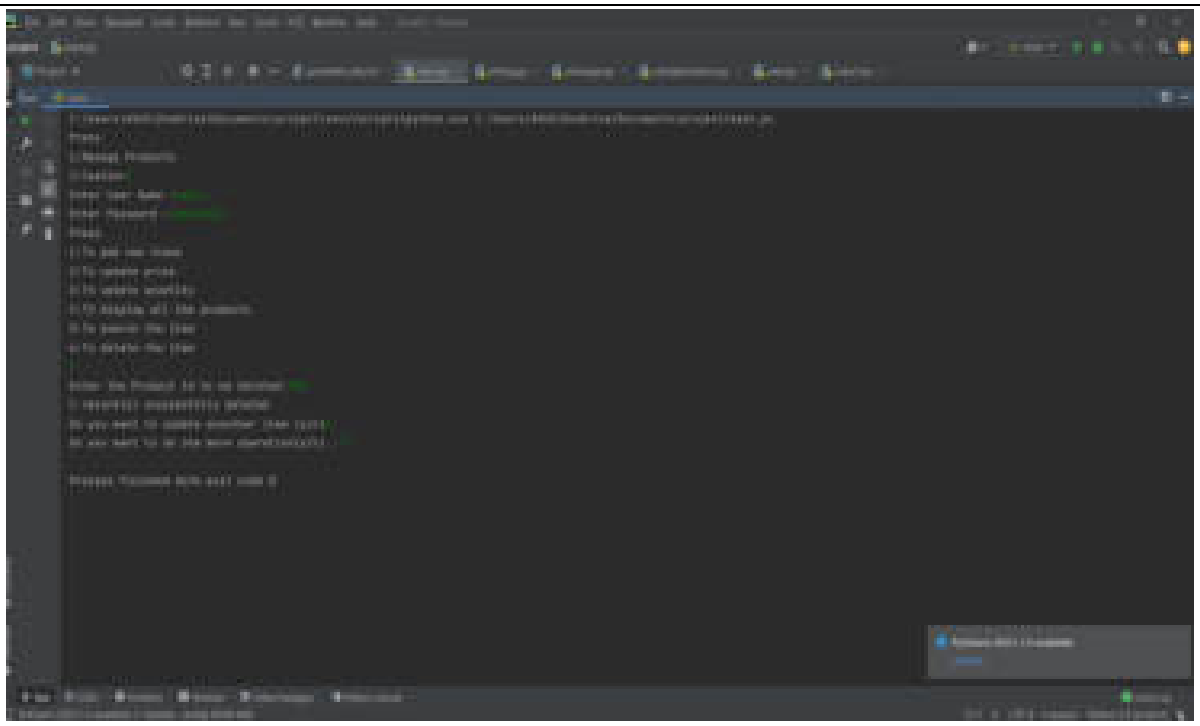


Fig.7.7 Output screen for deleting products.

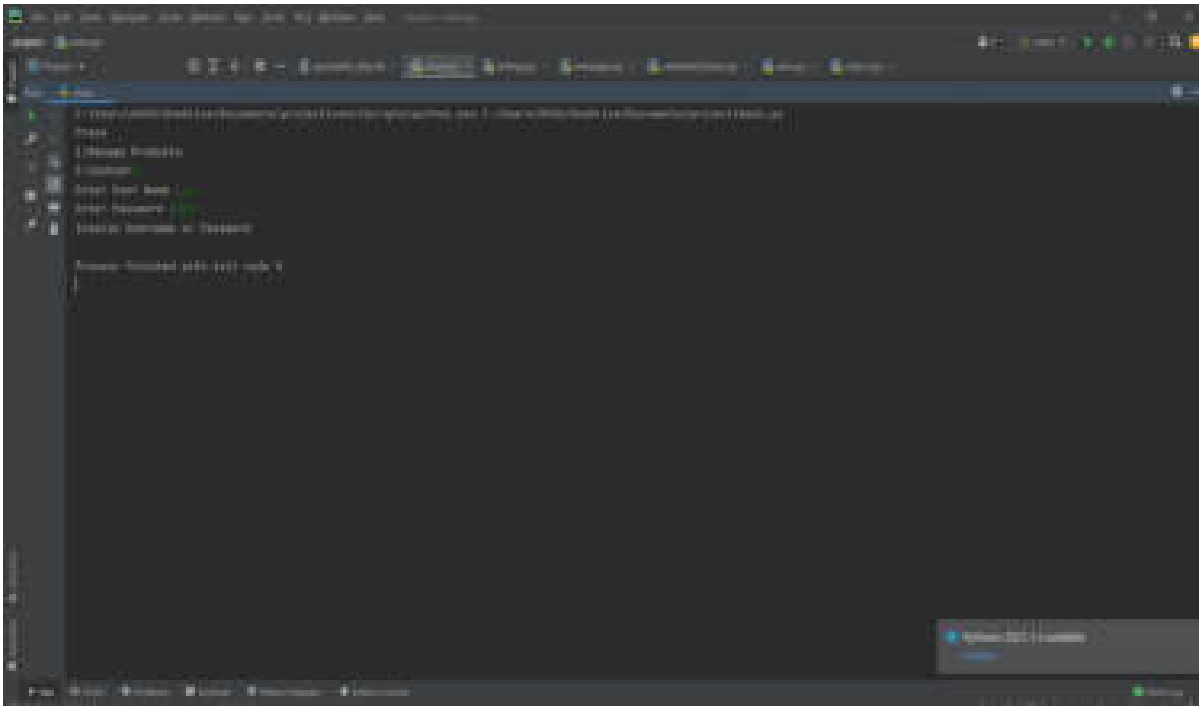


Fig.7.8 Output screen for cashier login.

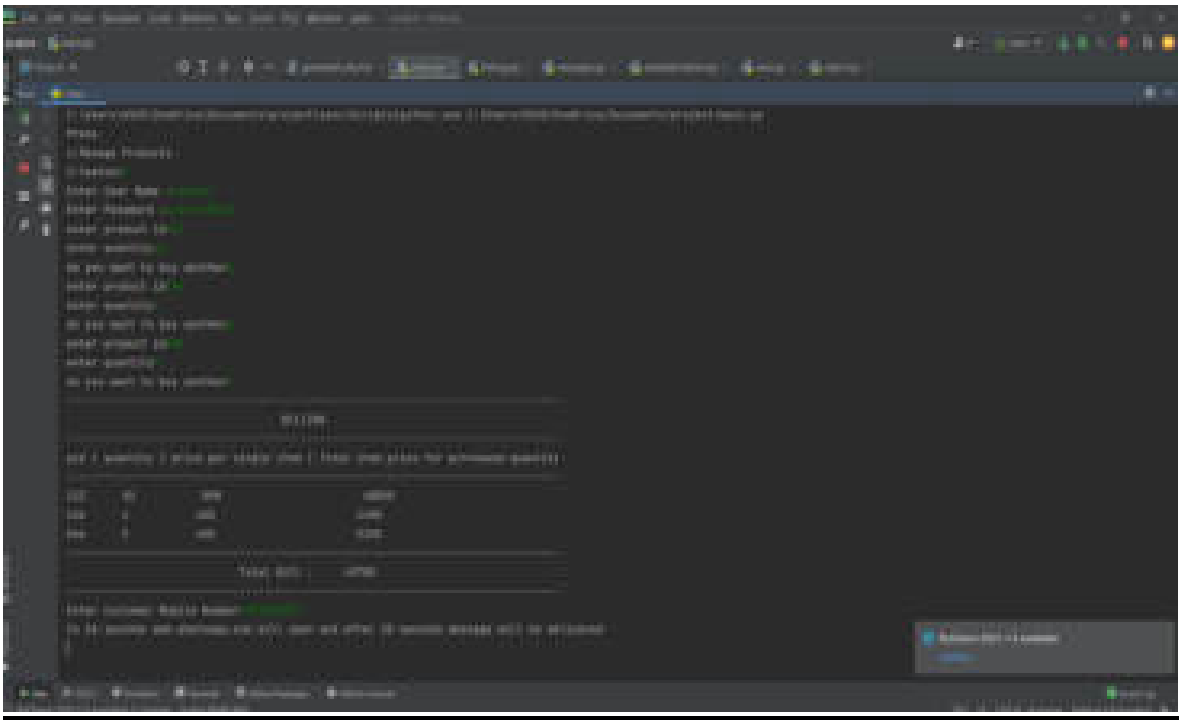


Fig.7.9 Output screen for calculation of bill.

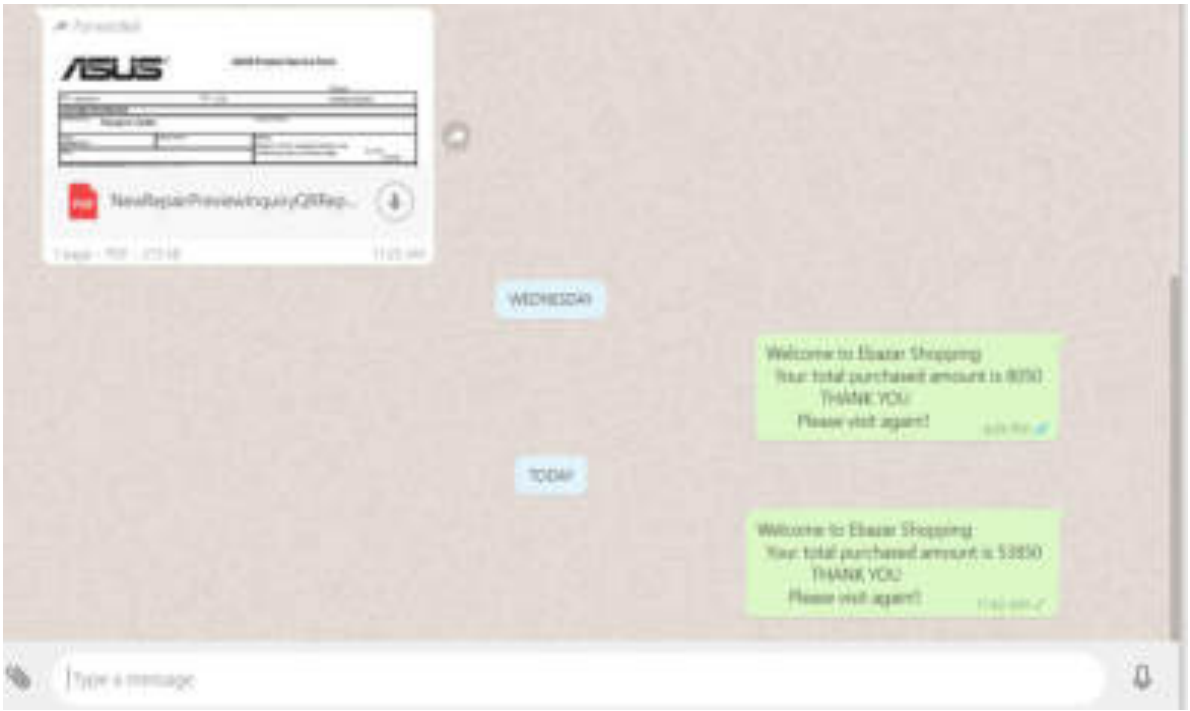


Fig.7.10 Output screen for bill details sent to whatsapp.

TESTING

8.1 INTRODUTION

TESTING

The testing phase is an important part of software development. It is the puterized system which will help in automating a process of finding errors and missing operations and also a complete verification to determine whether the objectives are met and the user requirements are satisfied.

Software testing is carried out in three steps:

1. The first includes unit testing, where in each module is tested to provide its correctness, validity and also determine any missing operations and to verify whether the objectives have been met. Errors are noted down and corrected immediately. Unit testing is the important and major part of the project. So errors are rectified easily in particular module and program clarity is increased. In this project entire system is divided into several modules and is developed individually. So unit testing is conducted to individual modules.
2. The second step includes Integration testing. It need not be the case, the software whose modules when run individually and showing perfect results, will also show perfect results when run as a whole. The individual modules are clipped under this major module and tested again and are verified for the results. This is due to poor interfacing, which may results in data being lost across an interface. A module can have inadvertent, adverse effect on any other or on the global data structures, causing serious problems.
3. The final step involves validation and testing which determines the software functions as the user expected. Here also some modifications are done and finally in the completion of the project it is satisfied fully by the end user.

8.2 Maintenance and Environment

AS the number of computer based systems, grieve libraries of computer software began to expand.House developed projects produced tones of thousand soft program source statements. Software products purchased from the outside added hundreds of thousands of new statements. A dark cloud appeared on the horizon. All of these programs, all of those source statements-had to be corrected when false were detected, modified as user requirements changed, or adapted to new hardware that was purchased. These activities were collectively called software Maintenance.

The maintenance phase focuses on change that is associated with error correction, adaptations required as the software's environment evolves, and changes due to enhancements brought by changing customer requirements. Four types of changes are encountered during the maintenance phase.

- Correction
- Adaption
- Enhancement
- Prevention

CORRECTION

Even with the best quality assurance activities is lightly that the customer will uncover defects in the software. Corrective maintenance changes the software to correct defects. Maintenance is a set of software Engineering activities that occur after software has been delivered to the customer and put into operation. Software configuration management is a set of tracking and control activities that began when a software project begins and terminates only when the software is taken out of the operation.

We may define maintenance by describing four activities that are undertaken after a program is released for use:

Corrective Maintenance

Adaptive Maintenance

Perfective Maintenance or Enhancement

Preventive Maintenance or reengineering

ADAPTATION

Over time, the original environment (E>G., CPU, operating system, business rules, external product characteristics) for which the software was developed is likely to change. Adaptive maintenance results in modification of the software to accommodate change to its external environment.

ENHANCEMENT

As software is used, the customer/user will recognize additional functions that will provide benefit. Perceptive maintenance extends the software beyond its original function requirements.

PREVENTION

Computer software deteriorates due to change, and because of this, preventive maintenance, often called software re engineering, must be conducted to enable the software to serve the needs of its end users. In essence, preventive maintenance makes changes to computer programs so that they can be more easily corrected, adapted, and enhanced. Software configuration management (SCM) is an umbrella activity that is applied throughout the software process.

8.3 TESTCASES

Test Case Identification Number: TC1

1	Purpose of the testcase	To check weather admin can successfully logged in or not
2	Pre-condition	Program should be launched
3	Inputs	Username and password of admin
4	Expected Outputs	Login successful
5	Post Conditions	Displays operation details
6	Outputs	Shows operations details
7	Pass/Fail	pass

Table:8.1 Test case for admin login.

Test Case Identification Number: TC2

1	Purpose of the testcase	To check weather cashier can successfully logged in or not
2	Pre-condition	Program should be launched
3	Inputs	Username and password of cashier
4	Expected Outputs	Login successful
5	Post Conditions	Asks for product id to calculate bill.
6	Outputs	Asks for pid to calculate bill.
7	Pass/Fail	pass

Table:8.2 Test case for cashier login.

Test Case Identification Number: TC3

1	Purpose of the testcase	To check weather admin can successfully perform operations on stock.
2	Pre-condition	Program should be launched
3	Inputs	Choose any of the operation.
4	Expected Outputs	Performs operations
5	Post Conditions	Successfully performs operations.
6	Outputs	Performs operations.
7	Pass/Fail	pass

Table:8.3 Test case for managing products.

Test Case Identification Number: TC4

1	Purpose of the testcase	To calculate bill of alla the items by cahier.
2	Pre-condition	Program should be launched
3	Inputs	Asks for pid unique identity.
4	Expected Outputs	Calculate bill.
5	Post Conditions	Asks to continue for another operation.
6	Outputs	Calculates bill.
7	Pass/Fail	pass

Table:8.4 Test case for calculating bill by cashier.

Test Case Identification Number: TC5

1	Purpose of the testcase	To check weather customer recieves message or not.
2	Pre-condition	Program should be launched
3	Inputs	Customer mobile number.
4	Expected Outputs	Message successfully sent.
5	Post Conditions	Message is sent to customer whatsapp.
6	Outputs	Message is successfully sent.
7	Pass/Fail	pass

Table:8.5 Test case for message sending.

CONCLUSION

CONCLUSION

In conclusion, Automobile Data Management and Billing System has to improve with making appropriate effort to stop the rising problem to all manual data entering operation in order to enhance the operation of such data management. This project can be used to aid all stores that still operating manually have been successfully developed.

FUTURE ENHANCEMENT

FUTURE ENHANCEMENTS

This application is developed with a intent to easily handle large amounts of data that avoids paper work and calculates the bill of all the items dynamically which is faster and secure. The next enhancement is to improve the scope of the project to bring awareness of the application and make the all users to use the application.

BIBILOGRAPHY

BIBILOGRAPHY

The following books were referred during the analysis and execution of the project.

Python Programming

By John Zelle

MySQL Connector/Python Revealed

By Jesper Wisborg Krogh

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"STOCK MARKET PREDICTION USING MACHINE LEARNING"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
COMPUTER SCIENCE AND ENGINEERING

Submitted by

N. Kavya (17H71A0578)

R. Durga Prasad (17H71A0567)

G. Bala Kumar (17H71A0563)

G. Ravi Teja (17H71A05A0)

Under the Esteemed Guidance of
Ms Mareswaramma

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by NAAC With Grade A

2020-21

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited B.Tech (CSE|ECE|ME)
Accredited by NAAC With Grade A

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"STOCK MARKET PREDICTION USING MACHINE LEARNING"** is a bonafide work carried out by **N. Kavya (17H71A0578), R. Durga Prasad (17H71A0567), G. Bala Kumar (17H71A0563), G. Ravi Teja (17H71A05A0)** in partial fulfilment for the award of degree of Bachelor of Technology in **Computer Science and Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

P. Mareswaramma
Project Guide

D. Prasad
Head of the Department

Examiner

Dr K Srinivas
PRINCIPAL

ACKNOWLEDGEMENT

We would like to take this opportunity to express our deepest appreciation to the following people for their valuable contributions and assistance with this Project.

Initially, we would like to thank our project supervisor, Ms. P. Mareswaramma**, Assistant Professor, Department of Computer Science and Engineering** for the guidance and support, especially for the valuable ideas and knowledge shared to me throughout the Project.

We have the immense pleasure in expressing our thanks and deep sense of gratitude to**, Mr. D. Prasad, Head of the Department, Computer Science and Engineering** for extending necessary facilities and support for the completion of the Project.

We whole sincerely acknowledge **Dr. K. Srinivas, Principal and Prof D. PANDURANGA RAO, CEO** for giving opportunity to take up the Project work. We also extend our thanks to all faculty members of **Computer Science and Engineering**, for their valuable guidance and encouragement.

We would like to extend our warm appreciation to all my friends for sharing us their knowledge, valuable contributions and help with this Project.

Finally, our special thanks to our family for their continuous support and help throughout our academic years and for their continual support and encouragement for the completion of the project.

DECLARATION

We N. Kavya, R. Durga Prasad, G. Bala Kumar, G. Ravi Teja of the Main-Project "**STOCK MARKET PREDICTION USING MACHINE LEARNING**", hereby declare that the matter embodied in this Project is the genuine work done by us and has not been submitted either to this university or to any other university/institute for the fulfilment of the requirement of any course of study.

N. Kavya (17H71A0578)

R. Durga Prasad (17H71A0567)

G. Bala Kumar (17H71A0563)

G. Ravi Teja (17H71A05A0)

CONTENTS

LIST OF FIGURES

ABSTRACT

Stock market prediction is that the scene of trying to complete the long-run value of company stock. Analysis of the stock price we take the price. By using the neural network, we develop a model. Within the neural network, we use a recurrent neural network that remembers each and each information through time.

LSTM networks maintain to stay contextual information of inputs by associate a loop that permits information to travel from one step to the following. These loops make recurrent neural networks seem magical. Each x train has the last 60 days value for the present-day y data. This can add more accurate results when a measure to existing stock price prediction algorithms. The network is trained and evaluated for accuracy with various sizes of knowledge, and also the results are tabulated.

In Stock Market Prediction, the aim is to predict the future value of the financial stocks of a company**.** The recent trend in stock market prediction technologies is the use of machine learning which makes predictions based on the values of current stock market indices by training on their previous values**.** Machine learning itself employs different models to make prediction easier and authentic**.** The paper focuses on the use of Regression and LSTM based Machine learning to predict stock values**.** Factors considered are open, close, low, high and volume**.**

CHAPTER-1

INTRODUCTION

The stock market is basically an aggregation of various buyers and sellers of stock. A stock (also known as shares more commonly) in general represents ownership claims on business by a particular individual or a group of people. The attempt to determine the future value of the stock market is known as a stock market prediction. The prediction is expected to be robust, accurate and efficient. The system must work according to the real-life scenarios and should be well suited to real-world settings.

A correct prediction of stocks can lead to huge profits for the seller and the broker. Frequently, it is brought out that prediction is chaotic rather than random, which means it can be predicted by carefully analyzing the history of respective stock market. Machine learning is an efficient way to represent such processes. It predicts a market value close to the tangible value, thereby increasing the accuracy. Introduction of machine learning to the area of stock prediction has appealed to many researches because of its efficient and accurate measurements.

The vital part of machine learning is the dataset used. The dataset should be as concrete as possible because a little change in the data can perpetuate massive changes in the outcome. In this project, supervised machine learning is employed on a dataset obtained from Yahoo Finance. This dataset comprises of following five variables: open, close, low, high and volume. Open, close, low and high are different bid prices for the stock at separate times with nearly direct names. The volume is the number of shares that passed from one owner to another during the time period. The model is then tested on the test data.

Regression and LSTM models are engaged for this conjecture separately. Regression involves minimizing error and LSTM contributes to remembering the data and results for the long run. Finally, the graphs for the fluctuation of prices with the dates (in case of Regression based model) and between actual and predicted price (for the LSTM based model) are plotted.

The system is also expected to take into account all the variables that might affect the stock's value and performance. There are various methods and ways of implementing the prediction system like Fundamental Analysis, Technical Analysis, Machine Learning, Market Mimicry, and Time series aspect structuring. With the advancement of the digital era, the prediction has moved up into the technological realm. The most prominent and promising technique involves the use of Artificial

Neural Networks, Recurrent Neural Networks that is basically the implementation of machine learning. Machine learning involves artificial intelligence which empowers the system to learn and improve from past experiences without being programmed time and again.

Traditional methods of prediction in machine learning use algorithms like Backward Propagation, also known as Backpropagation errors. Lately, many researchers are using more of ensemble learning techniques. It would use low price and time lags to predict future highs while another network would use lagged highs to predict future highs. These predictions were used to form stock prices. Stock market price prediction for short time windows appears to be a random process. The stock price movement over a long period of time usually develops a linear curve. People tend to buy those stocks whose prices are expected to rise in the near future. The uncertainty in the stock market refrain people from investing in stocks. Thus, there is a need to accurately predict the stock market which can be used in a real-lifescenario.

The methods used to predict the stock market includes a time series forecasting along with technical analysis, machine learning modelling and predicting the variable stock market. The datasetsof the stock market prediction model include details like the closing price opening price, the data andvarious other variables that are needed to predict the object variable which is the price in a given day.The previous model used traditional methods of prediction like multivariate analysis with a predictiontime series model. Stock market prediction outperforms when it is treated as a regression problem butperforms well when treated as a classification.

The aim is to design a model that gains from the marketinformation utilizing machine learning strategies and gauge the future patterns in stock value development. The Support Vector Machine (SVM), LSTM & etc. can be used for both classification and regression. It has been observed that LSTM are more used in classification based problem like ours. The LSTM technique, we plot every single data component as a point in n-dimensional space (where n is the number of features of the dataset available) with the value of feature being the value of a particular coordinate and, hence classification is performed by finding the hyper plane that differentiates the two classes explicitly.

1.2 PROBLEM STATEMENT

The Stock Market Prediction task is interesting as well as divides researchers and academics into two groups those who believe that we can devise mechanisms to predict the market and those who believe that the market is efficient and whenever new information comes up the market absorbs it by correcting itself, thus there is no space for prediction.

1.3 PURPOSE

The main purpose of prediction markets is the aggregation of beliefs over an unknown future outcome. Because they incorporate a wide variety of thoughts and opinions, prediction markets have proven to be quite effective as a prognostic tool. Thus, these markets can directly advise important policy decisions, by giving more accurate estimates of the aggregate consequences of those decisions. Prediction can be made for varying reasons including hedging against undesired events, insurance purpose or pure speculation. But it's clear today that the potential impact of this concept could go far beyond betting.

Individuals are also enabled to take advantage of proprietary information on a future event or outcome and turn it into a profit without revealing the source or content of the information. Thus, Prediction markets allow for the aggregation of information that would usually not be shared and allows for more accurate predictions.

1.4 MOTIVATION

The motivation of our project is:

- To determine whether the stock market price of a firm remains extremely fluctuating or stable by using current and previous data.
- To determine characteristics and analysis of stock price change in a given period of time.
- Decipher trends on real time change in prices of stocks in companies.
- To help the investors to invest wisely in stocks and help them to make profits on those investments.

CHAPTER-2

LITERATURE SURVEY

2.1 RELATED WORK

Survey of Stock Market Prediction Using Machine Learning Approach The stock market prediction has become an increasingly important issue in the present time. One of the methods employed is technical analysis, but such methods do not always yield accurate results. So it is important to develop methods for a more accurate prediction. Generally, investments are made using predictions that are obtained from the stock price after considering all the factors that might affect it.

The technique that was employed in this instance was a regression. Since financial stock marks generate enormous amounts of data at any given time a great volume of data needs to undergo analysis before a prediction can be made. Each of the techniques listed under regression has its own advantages and limitations over its other counterparts. One of the noteworthy techniques that were mentioned was linear regression.

The way linear regression models work is that they are often fitted using the least squares approach, but they may alternatively be also being fitted in other ways, such as by diminishing the "lack of fit" in some other norm, or by diminishing a handicapped version of the least squares loss function. Conversely, the least squares approach can be utilized to fit nonlinear models.

2.2 MACHINE LEARNING APPROACH IN STOCK MARKET PREDICTION

The vast majority of the stockbrokers while making the prediction utilized the specialized, fundamental or the time series analysis. Overall, these techniques couldn't be trusted completely, so there emerged the need to give a strong strategy to financial exchange prediction. To find the best accurate result, the methodology chose to be implemented as machine learning and AI along with supervised classifier. Results were tried on the binary classification utilizing SVM classifier with an alternate set of a feature list.

The greater part of the Machine Learning approach for taking care of business issues had their benefit over factual techniques that did exclude AI, despite the fact that there was an ideal procedure for specific issues. Swarm Intelligence optimization method named Cuckoo search was most easy to accommodate the parameters of SVM. The proposed hybrid CS-SVM strategy exhibited the performance to create increasingly exact outcomes in contrast with ANN. Likewise, the CS-SVM display performed better in the forecasting of the stock value prediction. Prediction

stock cost utilized parse records to compute the predicted, send it to the user, and autonomously perform tasks like buying and selling shares utilizing automation concept. Naïve BayesAlgorithm was utilized.

2.3 CORPORATE COMMUNICATION NETWORK AND STOCK PRICE MOVEMENTS: INSIGHTSFROMDATA MINING

This paper tries to indicate that communication patterns can have a very significant effect on an organization's performance. This paper proposed a technique to reveal the performance of a company. The technique deployed in the paper is used to find the relationships between the frequencies of email exchange of the key employees and the performance of the company reflected in stock values. In order to detect association and non-association relationships, this paper proposed to use a data mining algorithm on a publicly available dataset of Enron Corp. The Enron Corporation was an energy, commodities, and services company based in Houston, Texas whose stock dataset is available for public use.

2.4 IMPACT OF FINANCIAL RATIOS AND TECHNICAL ANALYSIS ON STOCK PRICE PREDICTION USING RANDOM FORESTS

The use of machine learning and artificial intelligence techniques to predict the prices of the stock is an increasing trend. More and more researchers invest their time every day in coming up with ways to arrive at techniques that can further improve the accuracy of the stock prediction model. Due to the vast number of options available, there can be n number of ways on how to predict the price of the stock, but all methods don't work the same way. The output varies for each technique even if the same dataset is being applied.

In the cited paper the stock price prediction has been carried out by using the random forest algorithm is being used to predict the price of the stock using financial ratios form the previous quarter. This is just one way of looking at the problem by approaching it using a predictive model, using the random forest to predict the future price of the stock from historical data.

However, there are always other factors that influence the price of the stock, such as sentiments of the investor, public opinion about the company, news from various outlets, and even events that causethe entire stock market to fluctuate. By using the financial ratio along with a model that can effectively analyze sentiments the accuracy of the stock price prediction model can be increased.

H. Gunduz, Z. Cataltepe and Y. Yaslan predicted stock prices using deep neural network techniques. Similarly, M. Billah, S. Waheed and A. Hanifa suggested further improvements to stock prediction using neural networks through the use of a training algorithm

which they designed on their own. K. V. Sujatha and S. M. Sundaram suggested insightful techniques on handling non-normal situations which may often arise during the working of the system and cause disruptions or lead to inaccurate predictions.

Liu, Liao and Y. Ding conducted similar work and designed a model for applying LSTM to stock prediction with lots of scope for improvements to prediction accuracy. K. A. Althelaya, E. M. El-Alfy and S. Mohammed further contributed to the field by staging experiments and simulations to assess the feasibility of applying deep learning techniques to prediction of stock prices.

Wilson and Sharda studied prediction firm bankruptcy using neural networks and classical multiple discriminant analysis, where neural networks performed significantly better than multiple discriminant analysis. Min and Lee were doing prediction of bankruptcy using machine learning. They evaluated methods based on Support Vector Machine, multiple discriminant analysis, logistic regression analysis, and three-layer fully connected back-propagation neural networks. Their results indicated that support vector machines outperformed other approaches.

Lee was trying to predict credit rating of a company using support vector machines. They used various financial indicator and ratios such as interest coverage ratio, ordinary income to total assets, Net income to stakeholders' equity, current liabilities ratio, etc. and achieved accuracy of around 60%. Predicting credit rating of the companies were also studied using neural networks achieving accuracy between 75% and 80% for the United States and Taiwan markets.

Tsai and Wang did a research where they tried to predict stock prices by using ensemble learning, composed of decision trees and artificial neural networks. They created dataset from Taiwanese stock market data, taking into account fundamental indexes, technical indexes, and macroeconomic indexes. The performance of Decision Tree + Artificial Neural Network trained on Taiwan stock exchange data showed F- score performance of 77%. Single algorithms showed F-score performance up to 67%.

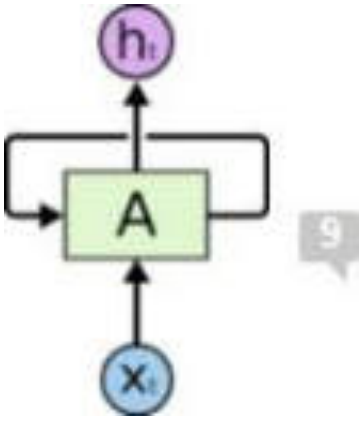
Kim and Han used a genetic algorithm to transform continuous input values into discrete ones. The genetic algorithm was used to reduce the complexity of the feature space. This paper proposes a novel evolutionary computing method called a genetic quantum algorithm. Genetic Quantum Algorithm is based on the concept and principles of quantum computing such as qubits and superposition of states. Instead of binary, numeric, or symbolic representation, by adopting bit chromosome as a representation Genetic Quantum Algorithm can represent a linear superposition of solutions due to its probabilistic representation. As genetic operators, quantum gates are employed for the search of the best solution.

2.5 ALGORITHMS

RECURRENT NEURAL NETWORKS

Humans don't start their thinking from scratch every second. As you read this essay, you understand each word based on your understanding of previous words. You don't throw everything away and start thinking from scratch again. Your thoughts have persistence.

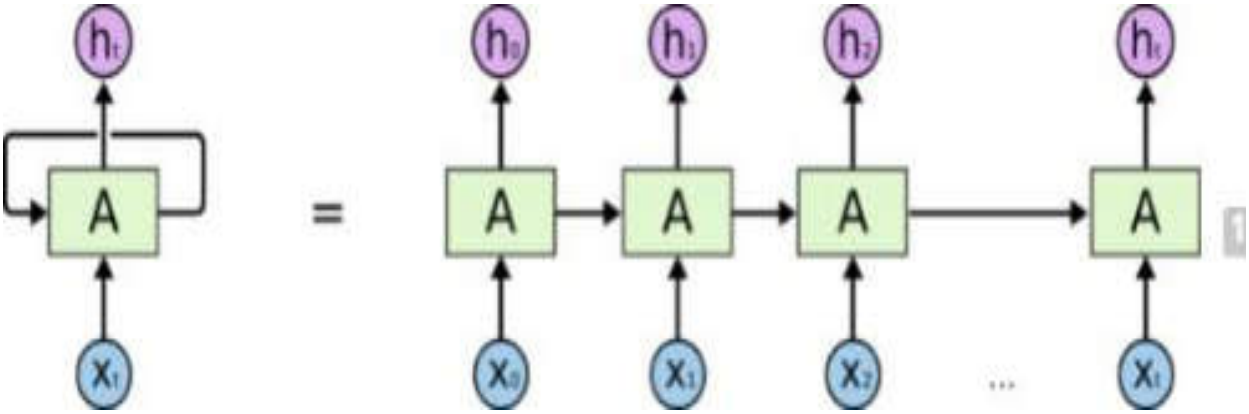
Traditional neural networks can't do this, and it seems like a major shortcoming. For example, imagine you want to classify what kind of event is happening at every point in a movie. It's unclear how a traditional neural network could use its reasoning about previous events in the film to inform later ones. Recurrent neural networks address this issue. They are networks with loops in them, allowing information to persist.



Recurrent Neural Networks have loops.

In the above diagram, a chunk of neural network, AA, looks at some input xtxt and outputs a value htht. A loop allows information to be passed from one step of the network to the next.

These loops make recurrent neural networks seem kind of mysterious. However, if you think a bitmore, it turns out that they are not all that different than a normal neural network. A recurrent neural network can be thought of as multiple copies of the same network, each passing a messageto a successor. Consider what happens if we unroll the loop:



An unrolled recurrent neural network

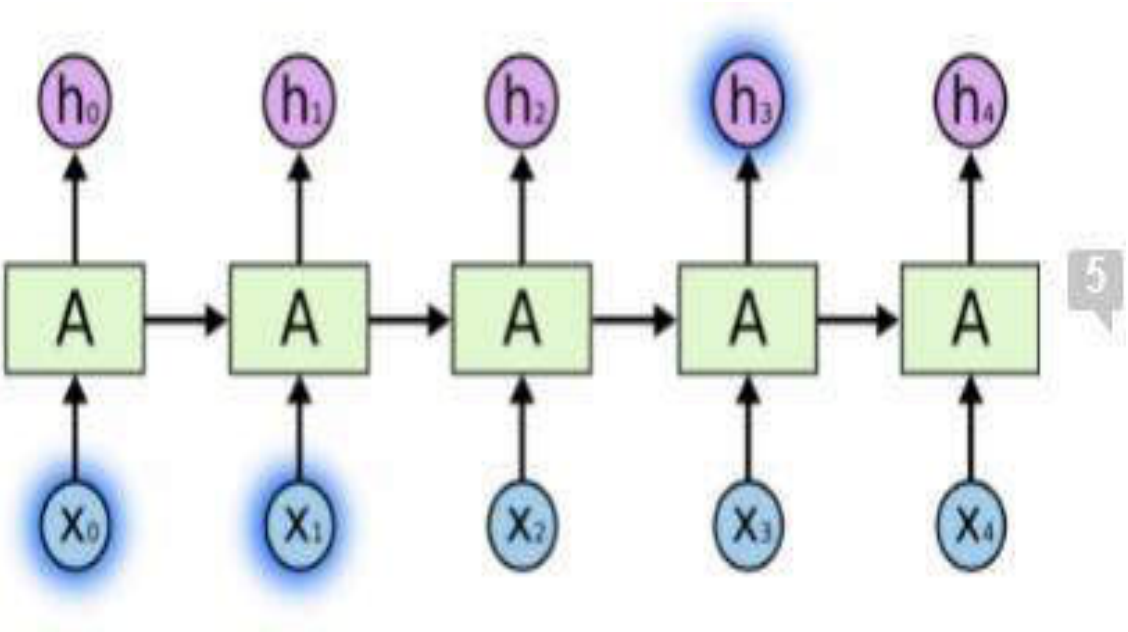
This chain-like nature reveals that recurrent neural networks are intimately related to sequencesand lists. They're the natural architecture of neural network to use for such data. And they certainly are used! In the last few years, there have been incredible success applying RNNs to a variety of problems: speech recognition, language modelling, translation, image captioning… The list goes on. I will leave discussion of the amazing feats one can achieve with RNNs to Andrej Karpathy excellent blog post, The Unreasonable Effectiveness of Recurrent Neural Networks. But they really are pretty amazing.

Essential to these successes is the use of "LSTMs," a very special kind of recurrent neural network which works, for many tasks, much better than the standard version. Almost all exciting results based on recurrent neural networks are achieved with them. It's these LSTMs that this essay will explore.

THE PROBLEM OF LONG-TERM DEPENDENCIES

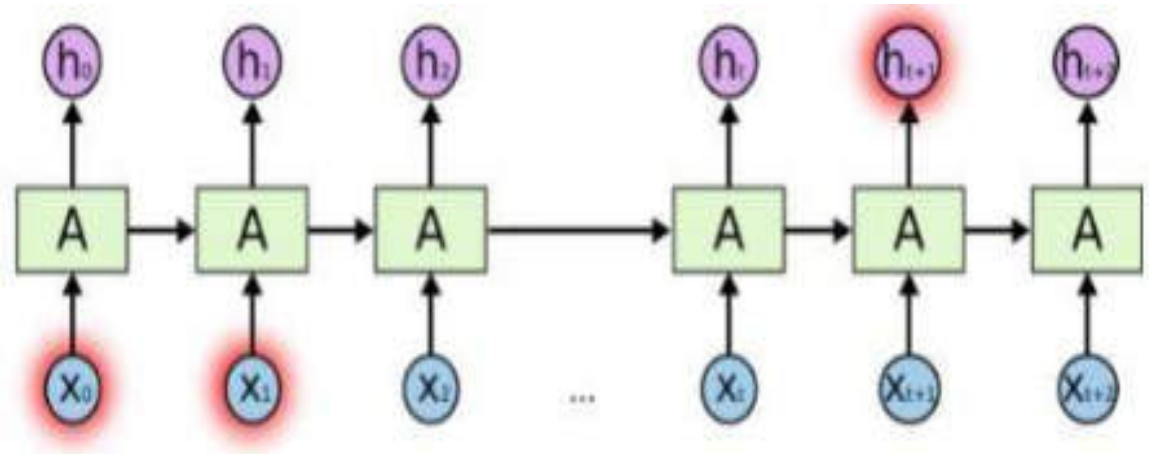
One of the appeals of RNNs is the idea that they might be able to connect previousinformation to the present task, such as using previous video frames might inform the understanding of the present frame. If RNNs could do this, they'd be extremely useful.But can they? It depends.

Sometimes, we only need to look at recent information to perform the present task. For example,consider a language model trying to predict the next word based on the previous ones. If we aretrying to predict the last word in "the clouds are in the sky," we don't need any further context – it's pretty obvious the next word is going to be sky. In such cases, where the gap between the relevant information and the place that it's needed is small, RNNs can learn to use the past information.



But there are also cases where we need more context. Consider trying to predict the last word inthe text “I grew up in France… I speak fluent French.” Recent information suggests that the nextword is probably the name of a language, but if we want to narrow down which language, we needthe context of France, from further back. It’s entirely possible for the gap between the relevantinformation and the point where it is needed to become very large.

Unfortunately, as that gap grows, RNNs become unable to learn to connect the information.



In theory, RNNs are absolutely capable of handling such “long-term dependencies.” A humancould carefully pick parameters for them to solve toy problems of this form. Sadly, in practice, RNNs don’t seem to be able to learn them. The problem was explored in depth by Hoch Reiter(1991) [German] and Bengio, et al. (1994), who found some pretty fundamental reasons why it might be difficult.

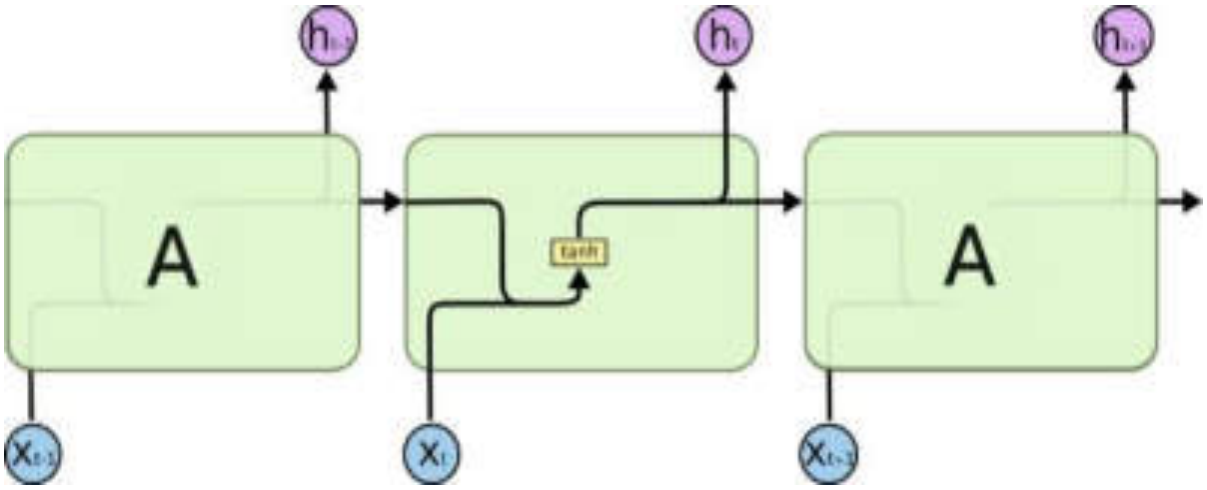
Thankfully, LSTMs don’t have this problem!

LSTM NETWORKS

Long Short Term Memory networks – usually just called “LSTMs” – are a special kind of RNN, capable of learning long-term dependencies. They were introduced by Hoch Reiter & Schmidhuber (1997), and were refined and popularized by many people in following work.1 they work tremendously well on a large variety of problems, and are now widely used.

LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically their default behavior, not something they struggle to learn!

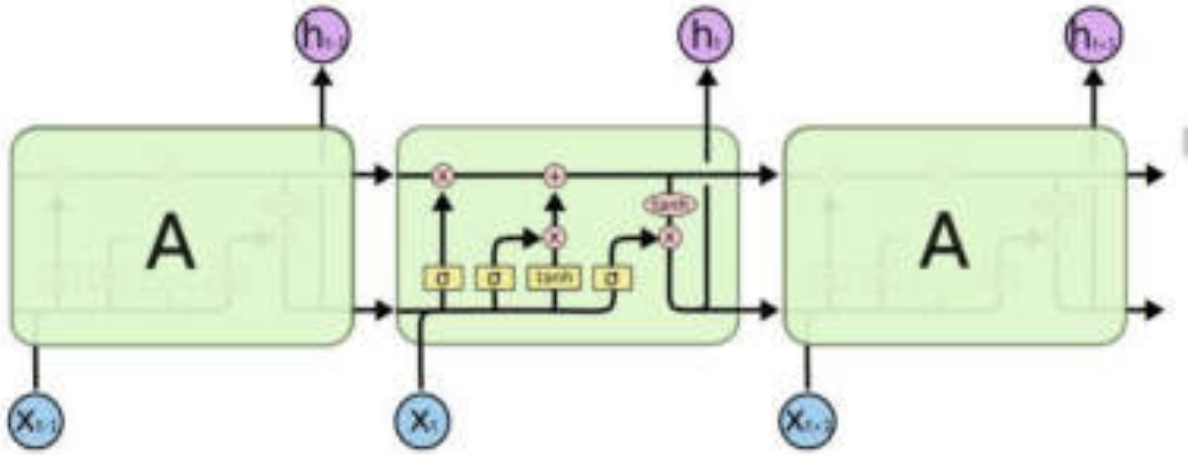
All recurrent neural networks have the form of a chain of repeating modules of neural



network. Instandard RNNs, this repeating module will have a very simple structure, such as a single tanh layer.

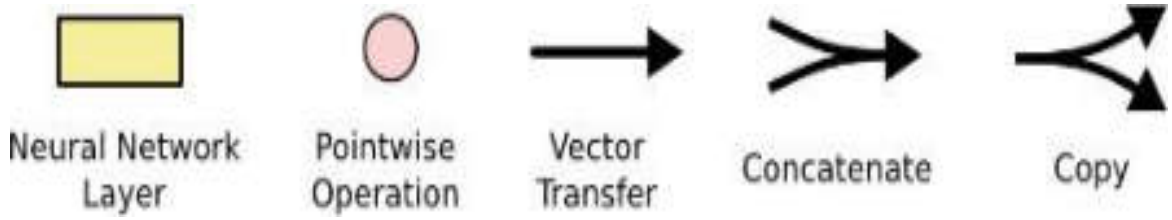
The repeating module in a standard RNN contains a single layer.

LSTMs also have this chain like structure, but the repeating module has a different structure. Instead of having a single neural network layer, there are four, interacting in a very special way.



The repeating module in an LSTM contains four interacting layers.

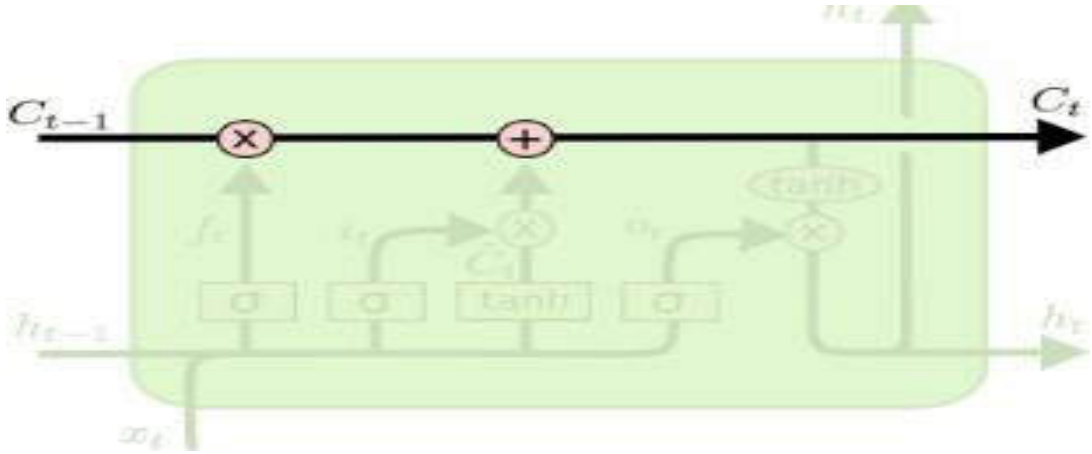
Don’t worry about the details of what’s going on. We will walk through the LSTM diagram step by step later. For now, let’s just try to get comfortable with the notation we will be using.



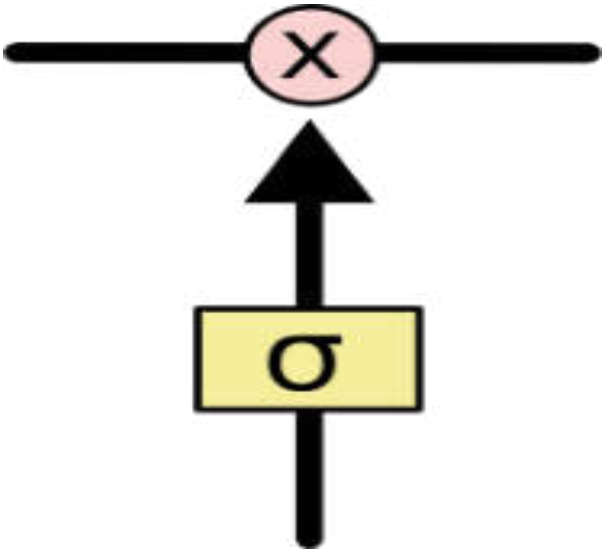
In the above diagram, each line carries an entire vector, from the output of one node to the inputsof others. The pink circles represent pointwise operations, like vector addition, while the yellowboxes are learned neural network layers. Lines merging denote concatenation, while a line forkingdenote its content being copied and the copies going to different locations.

THE CORE IDEA BEHIND LSTMS

The key to LSTMs is the cell state, the horizontal line running through the top of the diagram. The cell state is kind of like a conveyor belt. It runs straight down the entire chain, with only someminor linear interactions. It’s very easy for information to just flow along it unchanged.



The LSTM does have the ability to remove or add information to the cell state, carefully regulatedby structures called gates. Gates are a way to optionally let information through. They are composed out of a sigmoid neuralnet layer and a pointwise multiplication operation.



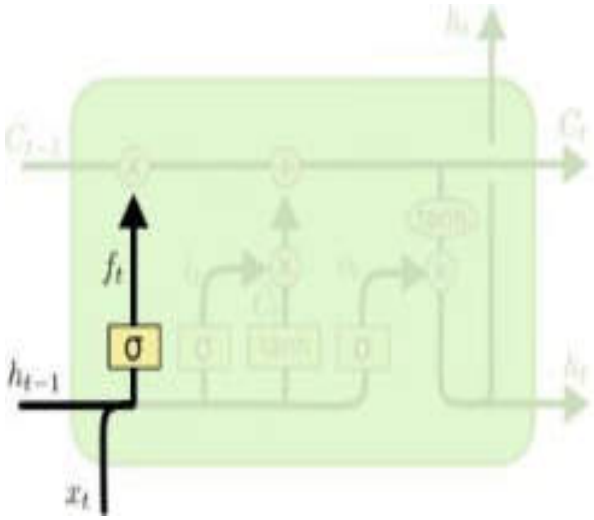
The sigmoid layer outputs numbers between zero and one, describing how much of each component should be let through. A value of zero means “let nothing through,” while a value ofone means “let everything through!”

An LSTM has three of these gates, to protect and control the cell state.

STEP-BY-STEP LSTM WALK THROUGH

The first step in our LSTM is to decide what information we’re going to throw away from the cell state. This decision is made by a sigmoid layer called the “forget gate layer.” It looks at ht−1ht−1 and xtxt, and outputs a number between 00 and 11 for each number in the cell state Ct−1Ct−1. 11 represents “completely keep this” while a 00 represents “completely get rid of this.”

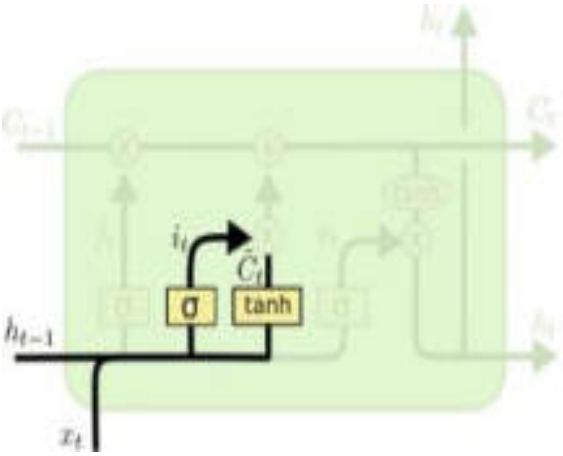
Let’s go back to our example of a language model trying to predict the next word based on all the previous ones. In such a problem, the cell state might include the gender of the present subject, so that the correct pronouns can be used. When we see a new subject, we want to forget the gender of the old subject.



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

The next step is to decide what new information we're going to store in the cell state. This has twoparts. First, a sigmoid layer called the "input gate layer" decides which values we will update.

Next, a tanh layer creates a vector of new candidate values, C~tC~t, that could be added to thestate. In the next step, we will combine these two to create an update to the state. In the example of our language model, we'd want to add the gender of the new subject to the cellstate, to replace the old one we're forgetting.

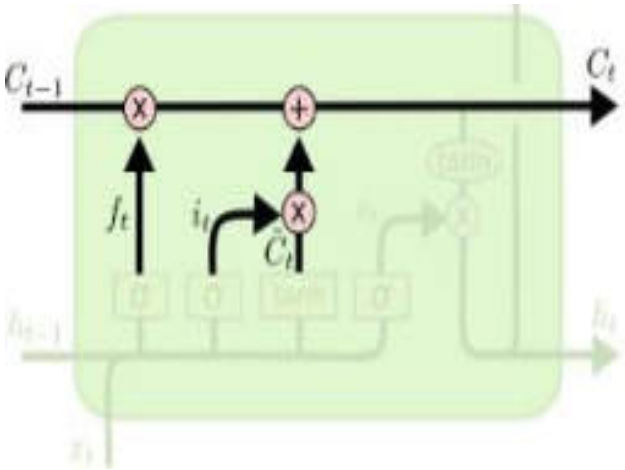


$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

It is now time to update the old cell state, Ct−1Ct-1, into the new cell state CtCt. The previous steps already decided what to do, we just need to actually do it.

We multiply the old state by ftft, forgetting the things we decided to forget earlier. Then we add it∗C~tit∗C~t. This is the new candidate values, scaled by how much we decided to update each state value.

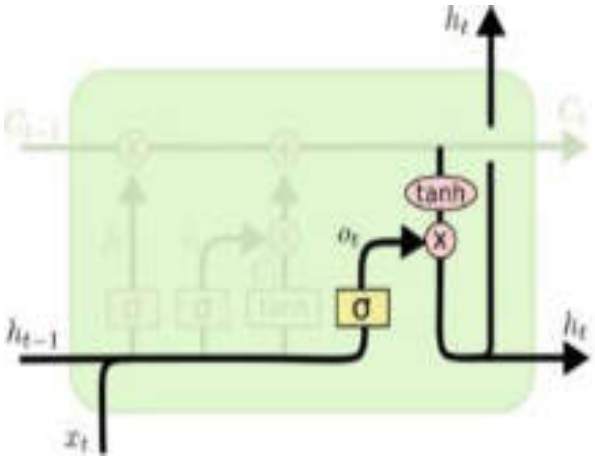
In the case of the language model, this is where we'd actually drop the information about the oldsubject's gender and add the new information, as we decided in the previous steps.



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Finally, we need to decide what we're going to output. This output will be based on our cell state,but will be a filtered version. First, we run a sigmoid layer which decides what parts of the cellstate we're going to output. Then, we put the cell state through tanh (to push the values to be between −1−1 and 11) and multiply it by the output of the sigmoid gate, so that we only outputthe parts we decided to.

For the language model example, since it just saw a subject, it might want to output informationrelevant to a verb, in case that's what is coming next. For example, it might output whether thesubject is singular or plural, so that we know what form a verb should be conjugated into if that'swhat follows next.



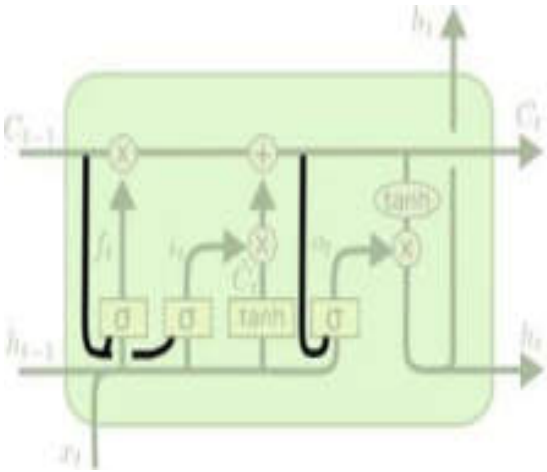
$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$

VARIANTS ON LONG SHORT TERM MEMORY

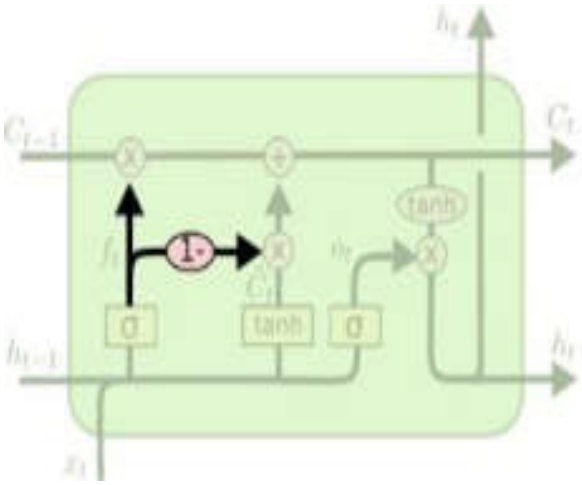
What I have described so far is a pretty normal LSTM. But not all LSTMs are the same as the above. In fact, it seems like almost every paper involving LSTMs uses a slightly different version. The differences are minor, but it's worth mentioning some of them.

One popular LSTM variant, introduced by Gers & Schmidhuber (2000), is adding "peephole connections." This means that we let the gate layers look at the cell state.



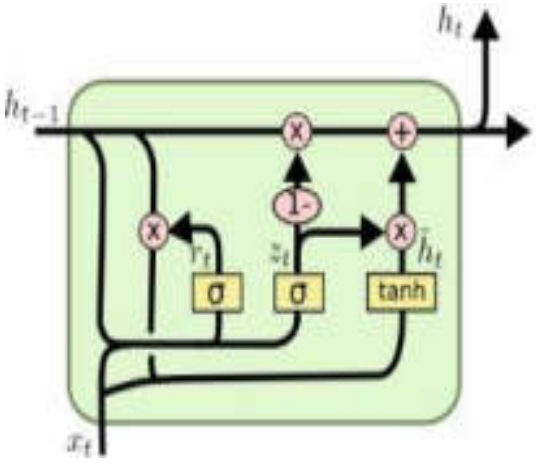
$$f_t = \sigma\left(W_f \cdot [C_{t-1}, h_{t-1}, x_t] + b_f\right)$$
$$i_t = \sigma\left(W_i \cdot [C_{t-1}, h_{t-1}, x_t] + b_i\right)$$
$$o_t = \sigma\left(W_o \cdot [C_t, h_{t-1}, x_t] + b_o\right)$$

The above diagram adds peepholes to all the gates, but many papers will give some peepholes and not others. Another variation is to use coupled forget and input gates. Instead of separately deciding what to forget and what we should add new information to, we make those decisions together. We only forget when we're going to input something in its place. We only input new values to the state when we forget something older.



$$C_t = f_t * C_{t-1} + (1 - f_t) * \tilde{C}_t$$

A slightly more dramatic variation on the LSTM is the Gated Recurrent Unit, or GRU, introduced by Cho, et al. (2014). It combines forget and input gates into a single "update gate."It also merges the cell state and hidden state, and makes some other changes. The resulting model is simpler than standard LSTM models, and has been growing increasingly popular.



$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

These are only a few of the most notable LSTM variants. There are lots of others, like Depth GatedRNNs by Yao, et al. (2015). There's also some completely different approach to tackling long- term dependencies, like Clockwork RNNs by Koutnik, et al. (2014).

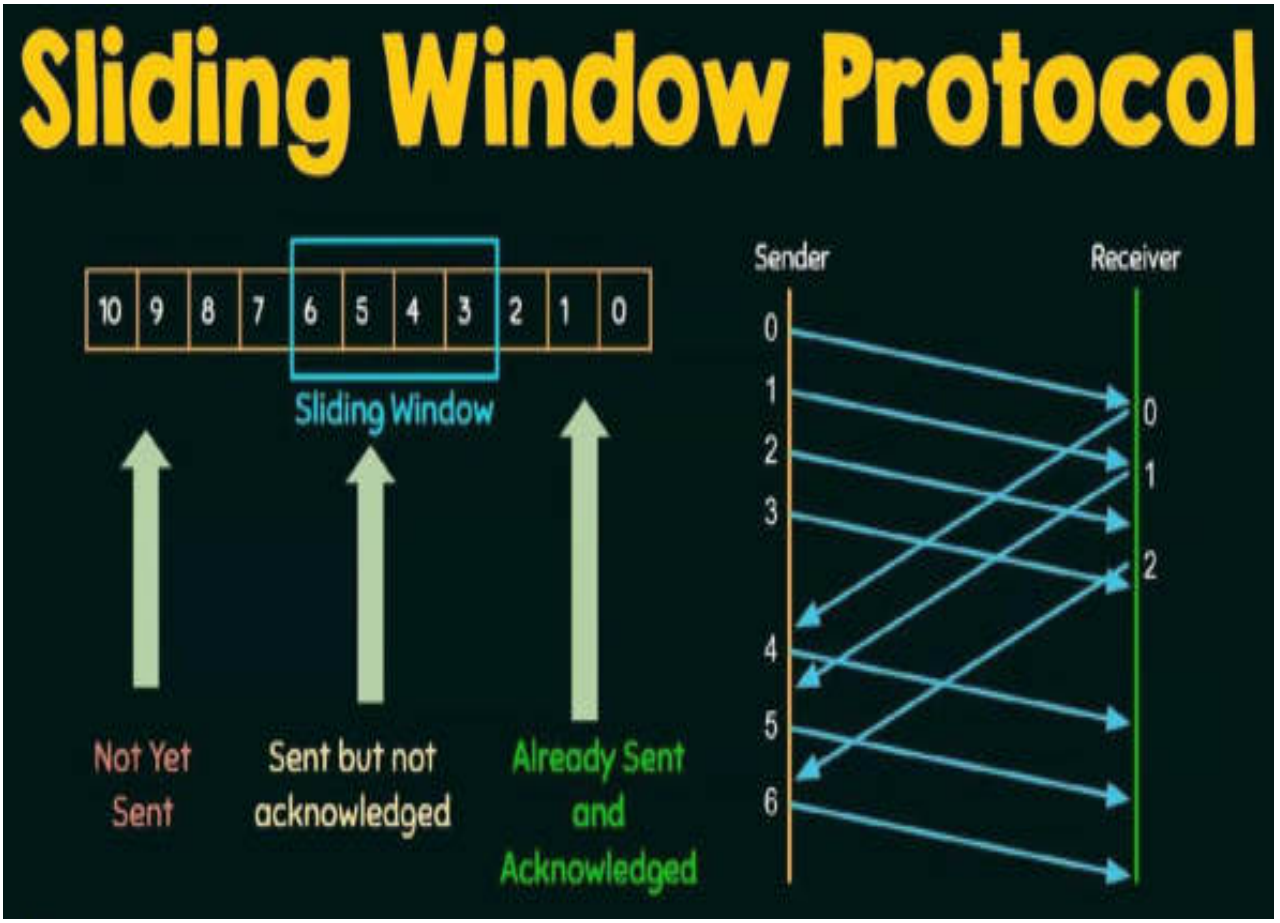
Which of these variants is best? Do the differences matter? Greff, et al. (2015) do a nice comparison of popular variants, finding that they're all about the same. Jozefowicz, et al. (2015) tested more than ten thousand RNN architectures, finding some that worked better than LSTMs on certain tasks.

CHAPTER-3

SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

The problem with estimating the stock price will remain a problem if a better stock market prediction algorithm is not proposed. Predicting how the stock market will perform is quite difficult. The movement in the stock market is usually determined by the sentiments of thousands of investors. Stock market prediction, calls for an ability to predict the effect of recent events on the investors. These events can be political events like a statement by a political leader, a piece of news on scam etc. It can also be an international event like sharp movements in currencies and commodity etc. All these events affect the corporate earnings, which in turn affects the sentiment of investors. It is beyond the scope of almost all investors to correctly and consistently predict these hyper parameters. All these factors make stock price prediction very difficult. Once the right data is collected, it then can be used to train a machine and to generate a predictive result.



DISADVANTAGESOF THE EXISTING SYSTEM

- The existing system fails when there are rare outcomes or predictors, as the algorithm is based on bootstrap sampling.
- The previous results indicate that the stock price is unpredictable when the traditional classifier is used.
- The existence system reported highly predictive values, by selecting an appropriate time period for their experiment to obtain highly predictive scores.
- The existing system does not perform well when there is a change in the operating environment.
- It doesn't focus on external events in the environment, like news events or social media.
- It exploits only one data source, thus highly biased.
- The existing system needs some form of input interpretation, thus need of scaling.
- It doesn't exploit data pre-processing techniques to remove inconsistency and incompleteness of the data.

3.2PROPOSED SYSTEM

In this proposed system, we focus on predicting the stock values using machine learning algorithms like Random Forest and Support Vector Machines. We proposed the system "Stock market price prediction" we have predicted the stock market price using the random forest algorithm. In this proposed system, we were able to train the machine from the various data points from the past to make a future prediction. We took data from the previous year stocks to train the model. We majorly used two machine-learning libraries to solve the problem.

The first one was numpy, which was used to clean and manipulate the data, and getting it into a form ready for analysis. The other was scikit, which was used for real analysis and prediction. The data set we used was from the previous years stock markets collected from the public database available online, 80 % of data was used to train the machine and the rest 20 % to test the data.

The basic approach of the supervised learning model is to learn the patterns and relationships in the data from the training set and then reproduce them for the test data. We used the python pandas library for data processing which combined different datasets into a data frame. The tuned up data frame allowed us to prepare the data for feature extraction.

The data frame features were date and the closing price for a particular day. We used all these features to train the machine on random forest model and predicted the object variable, which is the price for a given day.

We also quantified the accuracy by using the predictions for the test set and the actual values. The proposed system touches different areas of research including data pre-processing, random forest, and so on.

3.2.1 ADVANTAGES OF PROPOSED SYSTEM

- For long time lag problems. LSTM can handle noise and continuous values.
- Memory for long time periods.
- LSTM provide us with a large range of parameters such as learning rates, input and output biases. Hence, no need for fine adjustments.
- The complexity to update each weight is reduced to O (1) with LSTMs.

CHAPTER-4

MODULE IDENTIFICATION

The various modules of the project would be divided into the segments as described.

4.1DATA COLLECTION

Data gathering is a highly fundamental module and an early project phase. It usually covers the appropriate dataset collection. The data set utilized in the forecast of the market needs to be used for filtering on the basis of several factors. The data collecting also adds additional external data to improve the dataset. Our data consist primarily of stock prices from the preceding year.

4.2DATA PREPROCESSING

Data pre-processing is component of data mining that requires raw data to be transformed into a more cohesive structure. Raw data is usually unreliable or partial with numerous mistakes. In the pre-processing of data, missing values are checked, categorical values were sought, the data set was divided into training and test set and the feature is ultimately scaled to restrict the range of variables in order to compare them in a similar environment.

4.3 DATA SCORING

Data measurement is called the process of applying a prediction model to a collection of data. The LSTM algorithm is the method to handle the data set. LSTM consists of an ensemble technique, typically used for classification and regression. We obtain fascinating outcomes on the basis of the learning models.

CHAPTER-5

SYSTEM REQUIREMENTS

5.1 HARDWARE SYSTEM CONFIGURATION:-

- Processor - i3
- RAM - 4 GB (min)
- Hard Disk - 250 GB
- Key Board - Standard Windows Keyboard
- Mouse - Two or Three Button Mouse
- Monitor - SVGA(Super Video Graphics Array)

5.2 SOFTWARE REQUIREMENTS:

- Operating System - Windows XP
- Coding Language - Python
- Back End - DBSQL

CHAPTER-6

PRELIMINARY INVESTIGATION

The first and foremost strategy for development of a project starts from the thought of designing a mail enabled platform for a small firm in which it is easy and convenient of sending and receiving messages, there is a search engine, address book and also including some entertaining games. When it is approved by the organization and our project guide the first activity, i.e. Preliminary investigation begins. The activity has three parts:

- **Request Clarification**
- **Feasibility Study**
- **Request Approval**

6.1 REQUEST CLARIFICATION

After the approval of the request to the organization and project guide, with an investigation being considered, the project request must be examined to determine precisely what the system requires.

Here our project is basically meant for users within the company whose systems can be interconnected by the Local Area Network (LAN). In today's busy schedule man need everything should be provided in a readymade manner. So taking into consideration of the vastly use of the net in day to day life, the corresponding development of the portal came into existence.

6.2 FEASIBILITY ANALYSIS

An important outcome of preliminary investigation is the determination that the system request is feasible. This is possible only if it is feasible within limited resource and time. The different feasibilities that have to be analyzed are

- **Operational Feasibility**
- **Economic Feasibility**
- **Technical Feasibility**

6.2.1 OPERATIONAL FEASIBILITY

Operational Feasibility deals with the study of prospects of the system to be developed. This system operationally eliminates all the tensions of the Admin and helps him in effectively tracking the project progress. This kind of automation will surely reduce the time and energy, which previously consumed in manual work. Based on the study, the system is proved to be operationally feasible.

6.2.2 ECONOMIC FEASIBILITY

Economic Feasibility or Cost-benefit is an assessment of the economic justification for a computer based project. As hardware was installed from the beginning & for lots of purposes thus the cost on project of hardware is low. Since the system is a network based, any number of employees connected to the LAN within that organization can use this tool from at any time. The Virtual Private Network is to be developed using the existing resources of the organization. So the project is economically feasible.

6.2.3 TECHNICAL FEASIBILITY

According to Roger S. Pressman, Technical Feasibility is the assessment of the technical resources of the organization. The organization needs IBM compatible machines with a graphical web browser connected to the Internet and Intranet. The system is developed for platform Independent environment. Java Server Pages, JavaScript, HTML, SQL server and WebLogic Server are used to develop the system. The technical feasibility has been carried out. The system is technically feasible for development and can be developed with the existing facility.

6.3 REQUEST APPROVAL

Not all request projects are desirable or feasible. Some organization receives so many project requests from client users that only few of them are pursued. However, those projects that are both feasible and desirable should be put into schedule. After a project request is approved, it cost, priority, completion time and personnel requirement is estimated and used to determine where to add it to any project list. Truly speaking, the approval of those above factors, development works can be launched.

CHAPTER-7

SYSTEM DESIGN AND DEVELOPMENT

7.1 INPUT DESIGN

Input Design plays a vital role in the life cycle of software development, it requires very careful attention of developers. The input design is to feed data to the application as accurate aspossible. So inputs are supposed to be designed effectively so that the errors occurring while feeding are minimized. According to Software Engineering Concepts, the input forms or screens are designedto provide to have a validation control over the input limit, range and other related validations.

This system has input screens in almost all the modules. Error messages are developedto alert the user whenever he commits some mistakes and guides him in the right way so that invalidentries are not made. Let us see deeply about this under module design.

Input design is the process of converting the user created input into a computer- based format. The goal of the input design is to make the data entry logical and free from errors. Theerror is in the input are controlled by the input design. The application has been developed in user- friendly manner. The forms have been designed in such a way during the processing the cursor is placed in the position where must be entered. The user is also provided with in an option to select an appropriate input from various alternatives related to the field in certain cases.

Validations are required for each data entered. Whenever a user enters an erroneous data, error message is displayed and the user can move on to the subsequent pages after completing all the entries in the current page.

7.2 OUTPUT DESIGN

The Output from the computer is required to mainly create an efficient method of communication within the company primarily among the project leader and his team members, in other words, the administrator and the clients. The output of VPN is the system which allows the project leader to manage his clients in terms of creating new clients and assigning new projects to them, maintaining a record of the project validity and providing folder level access to each client on the user side depending on the projects allotted to him. After completion of a project, a new project may be assigned to the client. User authentication procedures are maintained at the initial stages itself. A new user may be created by the administrator himself or a user can himself register as a new user but the task of assigning projects and validating a new user rests with the administrator only.

The application starts running when it is executed for the first time. The server has to be started and then the internet explorer in used as the browser. The project will run on the local area network so the server machine will serve as the administrator while the other connected systems can act as the clients. The developed system is highly user friendly and can be easily understood by anyone using it even for the first time.

7.3 SOFTWARE REQUIREMENT SPECIFICATIONS

Software requirements specification is a rigorous assessment of requirements before the more specific system design stages, and its goal is to reduce later redesign. It should also provide a realistic basis for estimating product costs, risks, and schedules. Used appropriately, software requirements specifications can help prevent software project failure. The software requirements specification document lists sufficient and necessary requirements for the project development. To derive the requirements, the developer needs to have clear and thorough understanding of the products under development.

7.3.1 LANGUAGE USED

Python (version 3.6) is a general purpose, dynamic, high level and interpreted programming language. It supports Object Oriented programming approach to develop applications. It is simple and easy to learn and provides lots of high-level data structures.

- Python's syntax and dynamic typing with its interpreted nature, makes it an ideal language for scripting and rapid application development.
- Python is easy to learn yet powerful and versatile scripting language which makes it attractive for Application Development.
- Python supports multiple programming pattern, including object oriented, imperative and functional or procedural programming styles.
- Python is not intended to work on special area such as web programming. That is why it is known as multipurpose because it can be used with web, enterprise, 3D CAD etc.

7.3.2 FRAMEWORKS USED:

TensorFlow (version 2.4.3)

TensorFlow is an end-to-end open source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML powered application.

TensorFlow can train and run deep neural networks for handwritten digit classification, image recognition, word embeddings, recurrent neural networks, sequence-to-sequence models for machine translation, natural language processing, and PDE (partial differential equation) based simulations. Best of all, TensorFlow supports production prediction at scale, with the same modelsused for training.

Keras (version 2.3.0)

Keras is a deep learning API written in Python, running on top of the machine learning platform TensorFlow. It was developed with a focus on enabling fast experimentation. Being able togo from idea to result as fast as possible is key to doing good research.

Keras is:

- **Simple** -- but not simplistic. Keras reduces developer cognitive load to free you to focuson the parts of the problem that really matter.
- **Flexible** -- Keras adopts the principle of progressive disclosure of complexity: simpleworkflows should be quick and easy, while arbitrarily advanced workflows should be*possible* via a clear path that builds upon what you've already learned.
- **Powerful** -- Keras provides industry-strength performance and scalability: it is usedby organizations and companies including NASA, YouTube, or Way Mo

Django (version 3.0.5)

Django belongs to the full-stack Python framework. It is built by experienced developers, Django is a high level Python web framework which allows rapid, clean and pragmatic design development. Django handles much of the complexities of web development, so you can focus on writing your app without a need to reinvent the wheel. It’s free and open source.

To map objects to database table, Django uses ORM and the same is used to transfer from onedatabase to other. It works with mostly all important databases like Oracle, MySQL, PostgreSQL, SQLite, etc.

There are numerous websites in the industry which uses Django as their primary framework forbackend development.

Features of Django

Some of the exemplary features of this Python web framework are

- URL routing
- Authentication

- Database schema migrations
- ORM (Object-relational mapper)
- Template engine
- Excellent documentation and high scalability. Used by Top MNCs and Companies, such as Instagram, Disqus, Spotify, YouTube, Bitbucket, Dropbox, etc. and the list is never-ending.
- Easiest Framework to learn, rapid development and Batteries fully included.

7.3.3 LIBRARIES USED

Pandas

Pandas is an open-source Python Library providing high-performance data manipulation andanalysis tool using its powerful data structures. The name Pandas is derived from the word Panel Data – an Econometrics from Multidimensional data.

In 2008, developer Wes McKinney started developing pandas when in need of high performance,flexible tool for analysis of data prior to Pandas, Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data — load, prepare, manipulate, model, and analyze.

Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

Pandas deals with the following three data structures −

- Series
- Data Frame
- Panel

These data structures are built on top of Numpy array, which means they are fast.

Key Features of Pandas

- Fast and efficient Data Frame object with default and customized indexing.
- Tools for loading data into in-memory data objects from different file formats.
- Data alignment and integrated handling of missing data.
- Reshaping and pivoting of date sets.

- Label-based slicing, indexing and sub setting of large data sets.
- Columns from a data structure can be deleted or inserted.
- Group by data for aggregation and transformations.
- High performance merging and joining of data.
- Time Series functionality.

Sklearn

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modelling including classification, regression, and clustering and dimensionality reduction via a consistence interface inPython. This library, which is largely written in Python, is built upon **Numpy, SciPy** and **Matplotlib**.

The best way to represent data in Scikit-learn is in the form of tables. A table represents a 2-D grid of data where rows represent the individual elements of the dataset and the columns represents the quantities related to those individual elements.

7.3.4 PACKAGES

Pandas-datareader

The Pandas-datareader is a sub package that allows one to create a data frame from various internet data sources, currently including:

- Yahoo! Finance
- Google Finance
- St. Louis FED (FRED)
- Kenneth French's data library
- World Bank
- Google Analytics

7.3.5 API USED

Yahoo finance (Y finance)

The Yahoo Finance API is a range of libraries/APIs/methods to obtain historical and real time data for a variety of financial markets and products, as shown on YahooFinance- https://finance.yahoo.com/.

Some of the offerings include market data on Cryptocurrencies, regular currencies, stocks andbonds, fundamental and options data, and market analysis and news. Yahoo used to have its own official API but this was shut down in 2017.

Features of Yahoo finance API

- Free
- Impressive range of data
- Quick and easy to set yourself up
- Simple

7.3.6 DATABASES

SQLite

SQLite is an in-process library that implements a self-contained, server less, zero- configuration, transactional SQL database engine. It is a database, which is zero-configured, whichmeans like other databases you do not need to configure it in your system.

SQLite is one of the fastest-growing database engines around, but that's growth in terms ofpopularity, not anything to do with its size.

SQLite engine is not a standalone process like other databases, you can link it statically or dynamically as per your requirement with your application. SQLite accesses its storage filesdirectly. The source code for SQLite is in the public domain.

7.3.7 IDE USED

Visual Studio

Visual Studio is an **Integrated Development Environment (IDE)** developed by Microsoft to develop GUI (Graphical User Interface), console, Web applications, web apps, mobile apps, cloud, and web services, etc. With the help of this IDE, you can create managed code as well as native code. It uses the various platforms of Microsoft software development software like Windows store, Microsoft Silverlight, and Windows API, etc. It is not a language- specific IDE as you can use this to write code in C#, C++, VB (Visual Basic), Python, JavaScript,and many more languages. It provides support for 36 different programming languages. It is available for Windows as well as for mac OS.

Features of Visual Studio are:

- Support for multiple programming languages
- Intelligence
- Cross Platform Support
- Extensions and Support
- Repository
- Web-Support
- Hierarchy Structure
- Improving code
- Terminal support
- Multi projects
- Git Support
- Commenting

Advantages of Visual Studios

- Cross-platform support
    - Windows
    - Linux
    - Mac
- Light-weight
- Robust Architecture
- Intelli-Sense
- Freeware: Free of Cost- probably the best feature of all for all the programmers out there,even more for the organizations.
- Many users will use it or might have used it for desktop applications only, but it alsoprovides great tool support for Web Technologies like; HTML, CSS, JSON.

Visual Studio Scope

The most common languages are:

- C#
- Visual Basic
- Java-Script
- R
- XML
- Python
- CSS
- GO

DATA FLOW DIAGRAM:

DFD is the abbreviation for **Data Flow Diagram**. The flow of data of a system or a process is represented by DFD. It also gives insight into the inputs and outputs of each entity and the process itself. DFD does not have control flow and no loops or decision rules are present. Specific operations depending on the type of data can be explained by a flowchart. Data Flow Diagram can be represented in several ways. The DFD belongs to structured-analysis modeling tools. Data Flow diagrams are very popular because they help us to visualize the major steps and data involved in software-system processes.

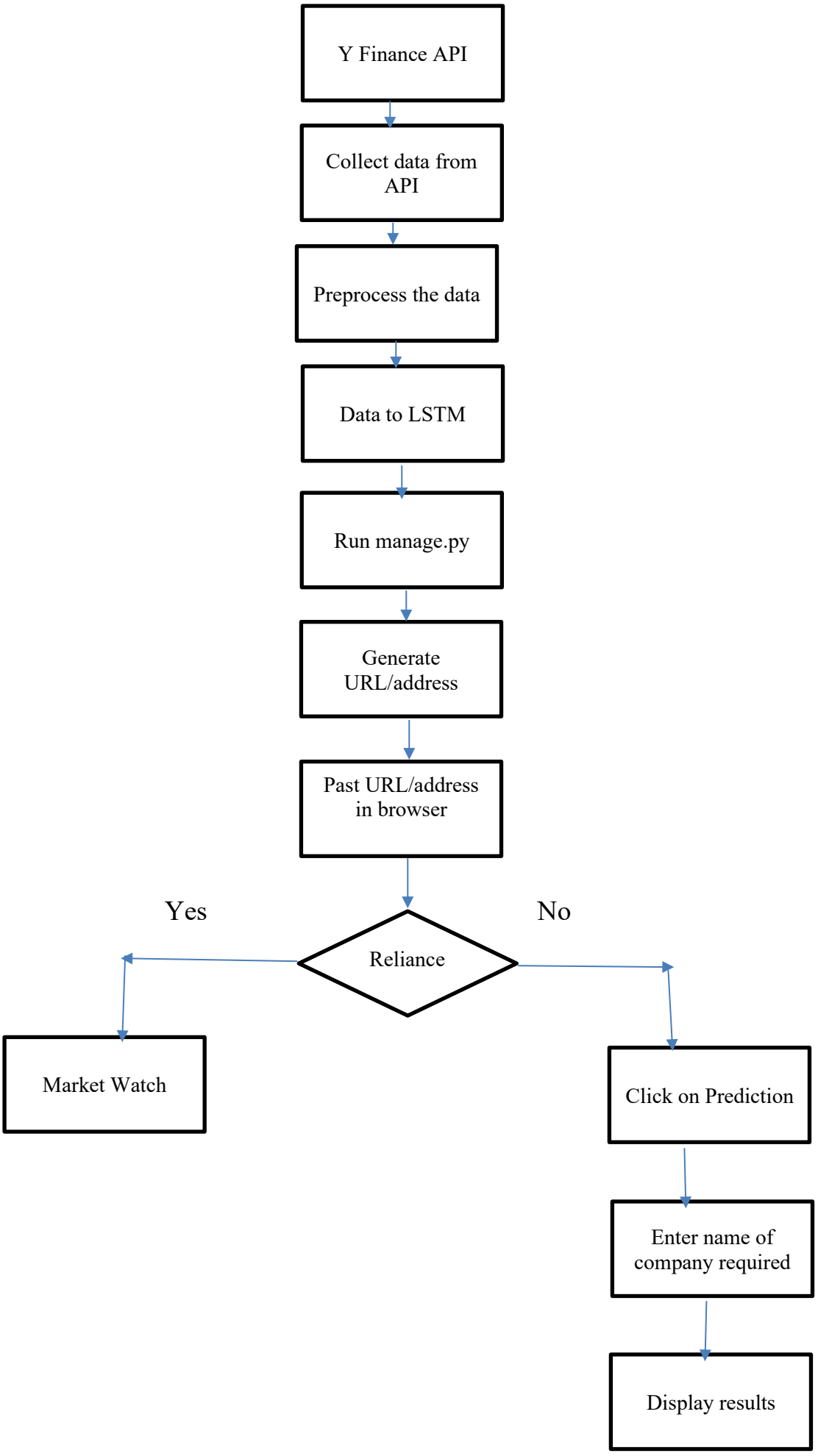
The Data Flow Diagram has 4 components:

- Process
- Data flow
- Warehouse
- Terminator

RULES FOR CREATING DFD

- The name of the entity should be easy and understandable without any extra assistance (like comments).
- The processes should be numbered or put in ordered list to be referred easily.
- The DFD should maintain consistency across all the DFD levels.
- A single DFD can have maximum processes up to 9 and minimum 3 processes.

DIAGRAM


Y Finance API
Collect data from API
Preprocess the data
Data to LSTM
Run manage.py
Generate URL/address
Past URL/address in browser
Reliance
Yes
No
Market Watch
Click on Prediction
Enter name of company required
Display results

CHAPTER-8

SYSTEM STUDY

8.1 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

8.1.1 ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

8.1.2 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical

resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement; as only minimal or null changes are required for implementing this system.

8.1.3 SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

CHAPTER-9

IMPLEMENTATION

9.1 FRONT END

Django: Django is a Python-based web framework that allows you to quickly create efficient web applications. It is also called batteries included framework because Django provides built-in features for everything including Django Admin Interface, default database – SQLlite3, etc. When you're building a website, you always need a similar set of components: a way to handle user authentication (signing up, signing in, signing out), a management panel for your website, forms, a way to upload files, etc. Django gives you ready-made components to use and that too for rapid development.

HTML

HTML stands for Hyper Text Markup Language. It is the standard markup language for creating Web pages.HTML describes the structure of a Web page. It consists of a series of elements. These elements tell the browser how to display the content.HTML elements label pieces of content such as "this is a heading", "this is a paragraph", "this is a link", etc.

CSS

Cascading **S**tyle **S**heets, fondly referred to as CSS, is a simple design language intended to simplify the process of making web pages presentable.

CSS handles the look and feel part of a web page. Using CSS, you can control the color of the text, the style of fonts, the spacing between paragraphs, how columns are sized and laid out, what background images or colors are used, layout designs and variations in display for different devices and screen sizes as well as a variety of other effects.

CSS is easy to learn and understand but it provides powerful control over the presentation of an HTML document. Most commonly, CSS is combined with the markup languages HTML or XHTML.

9.2 BACK END

Django: Django belongs to the full-stack Python framework. It is built by experienced developers, Django is a high level Python web framework which allows rapid, clean and pragmatic design development. Django handles much of the complexities of web

development, so you can focus on writing your app without a need to reinvent the wheel. It's free and open source.

To map objects to database table, Django uses ORM and the same is used to transfer from one database to other. It works with mostly all important databases like Oracle, MySQL, PostgreSQL, SQLite, etc.

There are numerous websites in the industry which uses Django as their primary frameworkfor backend development.

SQLite: SQLite is an in-process library that implements a self-contained, serverless, zero-configuration, transactional SQL database engine. It is a database, which is zero-configured, which means like other databases you do not need to configure it in your system. SQLite engine is not a standalone process like other databases, you can link it statically or dynamically as per your requirementwith your application. SQLite accesses its storage files directly.

9.3CODESNIPETS

LSTM CODE

```
def lstm_prediction(se, stock_symbol):
        import pandas as pd
        import numpy as np


        def fetch_stock_data(se, stock_symbol):
                """fetch stock data"""
                from pandas_datareader import data as pdr
                import yfinance as yf
                yf.pdr_override()
                if se == 'NSE': stock_symbol += ".NS"
```

```
        return pdr.get_data_yahoo(stock_symbol, period="5y")



"""LSTM model development"""

from sklearn.preprocessing import MinMaxScaler

from keras.models import Sequential

from keras.layers import Dense, Dropout, LSTM


og_df = fetch_stock_data(se, stock_symbol)

todataframe = og_df.reset_index(inplace=False)


#to print the info of the http://127.0.0.1:8000/OG dataset

print("\n<----------Info of the OG dataset---------->")

print(todataframe.info())

print("<---------------------------------------------------------------->\n")


#dataframe creation

seriesdata = todataframe.sort_index(ascending=True, axis=0)

new_seriesdata = pd.DataFrame(index=range(0,len(todataframe)),columns=['Date','Close'])

for i in range(0,len(seriesdata)):

    new_seriesdata['Date'][i] = seriesdata['Date'][i]

    new_seriesdata['Close'][i] = seriesdata['Close'][i]
```

```
        #setting the index again
        new_seriesdata.index = new_seriesdata.Date
        new_seriesdata.drop('Date', axis=1, inplace=True)
        #creating train and test sets this comprises the entire data's present in the dataset
        myseriesdataset = new_seriesdata.values
        totrain = myseriesdataset
        #converting dataset into x_train and y_train
        scalerdata = MinMaxScaler(feature_range=(0, 1))
        scale_data = scalerdata.fit_transform(myseriesdataset)
        x_totrain, y_totrain = [], []
        length_of_totrain=len(totrain)
        for i in range(60,length_of_totrain):
            x_totrain.append(scale_data[i-60:i,0])
            y_totrain.append(scale_data[i,0])
        x_totrain, y_totrain = np.array(x_totrain), np.array(y_totrain)
        x_totrain = np.reshape(x_totrain, (x_totrain.shape[0],x_totrain.shape[1],1))
        #LSTM neural network
        lstm_model = Sequential()
        lstm_model.add(LSTM(units=50,                                    return_sequences=True,
input_shape=(x_totrain.shape[1],1)))
        lstm_model.add(LSTM(units=50))
        lstm_model.add(Dense(1))
```

```
        lstm_model.compile(loss='mean_squared_error', optimizer='adadelta')
        lstm_model.fit(x_totrain, y_totrain, epochs=3, batch_size=1, verbose=2)
        #predicting next data stock price
        myinputs = new_seriesdata[len(new_seriesdata) - (100) - 60:].values
        myinputs = myinputs.reshape(-1,1)
        myinputs  = scalerdata.transform(myinputs)
        tostore_test_result = []
        for i in range(60,myinputs.shape[0]):
            tostore_test_result.append(myinputs[i-60:i,0])
        tostore_test_result = np.array(tostore_test_result)
        tostore_test_result =
np.reshape(tostore_test_result,(tostore_test_result.shape[0],tostore_test_result.shape[1],1))
        myclosing_priceresult = lstm_model.predict(tostore_test_result)
        myclosing_priceresult = scalerdata.inverse_transform(myclosing_priceresult)



        #Combining og and predicted dataset for end result.
        datelist = pd.date_range(pd.datetime.now().date(), periods=101)[1:]
        predicted_df = pd.DataFrame(myclosing_priceresult, columns=['Close'], index=datelist)
        result_df = pd.concat([og_df, predicted_df])[['Close']]
        result_df = result_df.reset_index(inplace=False)
        result_df.columns = ['Date', 'Close']
```

```
#to print the info of the END RESULT dataset

print("\n<----------------------Info of the RESULT dataset---------------------->")

print(result_df.info())

print("<------------------------------------------------------------------------>\n")

# import matplotlib.pyplot as plt



# plt.plot(result_df['Close'])



# plt.show()



def get_json(df):

    """ Small function to serialise DataFrame dates as 'YYYY-MM-DD' in JSON """

    import json

    import datetime

    def convert_timestamp(item_date_object):

        if isinstance(item_date_object, (datetime.date, datetime.datetime)):

            return item_date_object.strftime("%Y-%m-%d")



    dict_ = df.to_dict(orient='records')



    return json.dumps(dict_, default=convert_timestamp)
```

```
        return get_json(result_df)
```

Views.py

```
from django.shortcuts import render, redirect
from pred_app.lstm_prediction import *

# ---------- MAIN WEB PAGES ----------
def redirect_root(request):
    return redirect('/pred_app/index')

def index(request):
        return render(request, 'pred_app/index.html')

def pred(request):
    return render(request, 'pred_app/prediction.html')

def contact(request):
        return render(request, 'pred_app/contact.html')

def search(request, se, stock_symbol):
        import json
        predicted_result_df = lstm_prediction(se, stock_symbol)
        return render(request, 'pred_app/search.html', {"predicted_result_df": predicted_result_df})
```

Manage.py

```
#!/usr/bin/env python
```

```
"""Django's command-line utility for administrative tasks."""
import os
import sys


def main():
    os.environ.setdefault('DJANGO_SETTINGS_MODULE', 'predstocks.settings')
    try:
        from django.core.management import execute_from_command_line
    except ImportError as exc:
        raise ImportError(
            "Couldn't import Django. Are you sure it's installed and "
            "available on your PYTHONPATH environment variable? Did you "
            "forget to activate a virtual environment?"
        ) from exc
    execute_from_command_line(sys.argv)


if __name__ == '__main__':
    main()
```

Base.html

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" style="background-image:
url(https://wallpapercave.com/wp/wp2800193.jpg);">
   <head>
      {% load static %}
```

```
<title>
   PredStocks
</title>
<!-- get the required files from 3rd party sources -->
<link href="http://fonts.googleapis.com/css?family=Roboto" rel="stylesheet" type="text/css">
   <meta content="text/html; charset=utf-8" http-equiv="Content-Type"/>
   <link href="{% static 'pred_app/css/style.css' %}" rel="stylesheet" type="text/css"/>
   <!-- Bootstrap -->
   <link href="{% static 'pred_app/css/bootstrap.css' %}" rel="stylesheet" type="text/css">
      <link href="{% static 'pred_app/fonts/font-awesome/css/font-awesome.css' %}"
rel="stylesheet" type="text/css">
         <link href="{% static 'pred_app/css/menusearch.css' %}" rel="stylesheet"
type="text/css">
            <!-- Stylesheet==================================================
-->
            <link href="{% static 'pred_app/css/style.css' %}" rel="stylesheet" type="text/css">
               <link href="{% static 'pred_app/css/responsive.css' %}" rel="stylesheet"
type="text/css">
                  <script src="{% static 'pred_app/js/modernizr.custom.js' %}"
type="text/javascript">
                  </script>
                  <link
href="http://fonts.googleapis.com/css?family=Raleway:500,600,700,100,800,900,400,200,300"
rel="stylesheet" type="text/css">
                     <link href="http://fonts.googleapis.com/css?family=Playball" rel="stylesheet"
type="text/css">
                        <script src="{% static 'pred_app/js/jquery.1.11.1.js' %}"
type="text/javascript">
                        </script>
                        <script src="{% static 'pred_app/js/jquery-1.10.2.min.js' %}"
type="text/javascript">
                        </script>
                        <script src="{% static 'pred_app/js/handlebars.js' %}">
                        </script>
                        <script src="http://code.jquery.com/jquery-migrate-1.2.1.min.js"
type="text/javascript">
                        </script>
```

```
                    </link>
                </link>
            </link>
        </link>
    </link>
</link>
</head>
<body>
    <!-- jQuery (necessary for Bootstrap's JavaScript plugins) -->
    <script src="https://ajax.googleapis.com/ajax/libs/jquery/1.11.1/jquery.min.js">
    </script>
    <script src="{% static 'pred_app/js/jquery.1.11.1.js'%}" type="text/javascript">
    </script>
    <!-- Include all compiled plugins (below), or include individual files as needed -->
    <script src="{% static 'pred_app/js/bootstrap.js'%}" type="text/javascript">
    </script>
    <!-- Javascripts
================================================== -->
    <script src="{% static 'pred_app/js/typeahead.js'%}">
    </script>
    <script src="{% static 'pred_app/js/search.js'%}">
    </script>
    <div class="header" style="background-color: rgb(40, 40, 40); color: white;">
        <div class="logo">
            <h1>
                <a href="index" onmouseover="mouseOver()" onmouseout="mouseOut()">
                    <span id="preds_logo">
                        Pred
                    </span>
                    <span id="stocks_logo">
                    Stocks
                </span>
                    <br>
                        <small>
                            One stop solution to smart prediction !
```

```
                </small>
              </br>
            </a>
          </h1>
        </div>
        <div class="menu_nav">
          <ul>
            {% block nav %}{% endblock %}
          </ul>
        </div>
        <div class="clr">
        </div>
        <div class="hbg">
          <img alt="" class="fl" height="400px" src="{% static 'pred_app/img/header_b.jpg' %}"
width="100%"/>
        </div>
        <div class="clr">
        </div>
      </div>
      {% block content %}{% endblock %}
      <div class="clr">
      </div>
      <div class="fbg">
        <div class="fbg_resize">
          <div class="col c1">
            <h2>
              <span>
                Image Gallery
              </span>
            </h2>
            <a href="#">
              <img alt="" height="58" src="{% static 'pred_app/img/image2.jpg' %}" width="58"/>
            </a>
            <a href="#">
              <img alt="" height="58" src="{% static 'pred_app/img/image3.jpg' %}" width="58"/>
            </a>
            <a href="#">
```

```
            <img alt="" height="58" src="{% static 'pred_app/img/image4.jpg' %}" width="58"/>
          </a>
          <a href="#">
            <img alt="" height="58" src="{% static 'pred_app/img/image5.jpg' %}" width="58"/>
          </a>
          <a href="#">
            <img alt="" height="58" src="{% static 'pred_app/img/image6.jpg' %}" width="58"/>
          </a>
          <a href="#">
            <img alt="" height="58" src="{% static 'pred_app/img/image7.jpg' %}" width="58"/>
          </a>
        </div>
        <div class="col c3" style="width:620px">
          <h2>
            <span>
              About Us
            </span>
          </h2>
```

The stock market can be a confusing place for the average investor. But you don't need to be a hedge-fund manager or a billionaire to beat Wall Street at its own game.

We seek to help the ordinary individual investor make intelligent investments by providing information that gives you an edge.

PredStocks uses state of the art software to predict stock prices and generate trading "Buy-Sell" signals that will help you make your decision.

```
        </div>
      </div>
    </div>
  </body>
</html>
<div class="clr">
</div>
<div class="footer">
  <div class="footer_resize">
    <p class="lf">
      © Copyright
      <a href="#">
        MyWebSite
```

```
            </a>
            .
        </p>
        <div class="clr">
        </div>
    </div>
</div>
<script type="text/javascript">
    function mouseOver() {
  document.getElementById("stocks_logo").style.color = "#fe001a";
  document.getElementById("preds_logo").style.color = "#97c950";
}
function mouseOut() {
  document.getElementById("stocks_logo").style.color = "#97c950";
  document.getElementById("preds_logo").style.color = "#fe001a";
}
</script>
```

Contact.html

```
{% extends "pred_app/base.html" %}
{% load static %}
{% block nav %}
<li>
    <a href="index">
        Home
    </a>
</li>
<li>
    <a href="pred">
        Prediction
    </a>
</li>
<li class="active">
    <a href="contact">
        Contact Us
    </a>
</li>
{% endblock nav %}
```

```
{% block content %}
<div class="content">
   <div class="article">
      <h2>
         <span>
            To Contact Us
         </span>
      </h2>
      <div class="clr">
      </div>
   </div>
   <div class="article">
      <h2>
         <span>
            Send us
         </span>
         mail
      </h2>
      <div class="clr">
      </div>
      <form action="#" id="sendemail" method="post">
         <ol>
            <li>
               <label for="name">
                  Name (required)
               </label>
               <input class="text" id="name" name="name"/>
            </li>
            <li>
               <label for="email">
                  Email Address (required)
               </label>
               <input class="text" id="email" name="email"/>
            </li>
            <li>
               <label for="website">
                  Website
```

```
                </label>
                <input class="text" id="website" name="website"/>
            </li>
            <li>
                <label for="message">
                    Your Message
                </label>
                <textarea cols="50" id="message" name="message" rows="8">
                </textarea>
            </li>
            <li>
                <input class="send" id="imageField" name="imageField" src="{% static
'pred_app/img/submit.gif' %}" type="image"/>
            </li>
        </ol>
    </form>
  </div>
</div>
{% endblock content %}
```

Index.html

```
{% extends "pred_app/base.html" %}
{% block nav %}
<li class="active">
  <a href="index">
    Home
  </a>
</li>
<li>
  <a href="pred">
    Prediction
  </a>
</li>
<li>
  <a href="contact">
    Contact Us
  </a>
```

```
</li>
{% endblock nav %}
{% block content %}
<h1>Market Watch</h1>

<!-- TradingView Widget BEGIN -->
<div class="tradingview-widget-container">
  <script type="text/javascript" src="https://s3.tradingview.com/tv.js"></script>
  <script type="text/javascript">
  new TradingView.widget(
  {
  "width": "100%",
  "height": "610",
  "symbol": "BSE:RELIANCE",
  "interval": "D",
  "timezone": "Asia/Kolkata",
  "theme": "Dark",
  "style": "2",
  "locale": "in",
  "toolbar_bg": "#f1f3f6",
  "enable_publishing": false,
  "allow_symbol_change": true,
  "news": [
    "headlines"
  ],
  "container_id": "tradingview_197fc"
}
  );
  </script>
</div>
<!-- TradingView Widget END -->
<br>
    <br>
        {% endblock content %}
    </br>
</br>
```

```
{% extends "pred_app/base.html" %}
{% load static %}
{% block nav %}
<li>
   <a href="index">
      Home
   </a>
</li>
<li class="active">
   <a href="pred">
      Prediction
   </a>
</li>
<li>
   <a href="contact">
      Contact Us
   </a>
</li>
{% endblock nav %}
{% block content %}
<div class="content">
   <img alt="" height="180"
src="https://thelinknewspaper.ca/images/articles/Volume_36/Special/_resized/mediadem.polls.Laura
Lalonde.gif" width="500"/>
   <div class="sidebar">
      <div class="gadget">
         <br>
            <br>
               <div class="clr">
               </div>
               <form action="http://www.google.com/search" target="_blank" method="get">
                  <input maxlength="255" name="q" size="25" type="text" value=""/>
                  <input type="submit" value="Search Web"/>
               </form>
            </br>
         </br>
      </div>
```

```
    </div>
    <div class="form-group col-sm-6 col-sm-offset-1">
        <h2 align="center">
            Get Your Predictions !
        </h2>
        <form align="left" class="" id="search">
            <input class="form-control typeahead" data-url2="{% static 'pred_app/data/stock2.json' %}"
data-url="{% static 'pred_app/data/stock.json' %}" id="typeahead" placeholder="Search Stock
Symbol..." type="text" usr-url=""/>
        </form>
    </div>
    <br>
        <br>
            <br>
                <br>
                    <br>
                        <br>
                            <br>
                            </br>
                        </br>
                    </br>
                </br>
            </br>
        </br>
    </br>
</div>
{% endblock content %}
```

Search.html

```
{% extends "pred_app/base.html" %}
{% load static %}
{% block nav %}
<li>
    <a href="/pred_app/index">
```

```
        Home
    </a>
</li>
<li class="active">
    <a href="/pred_app/pred">
        Prediction
    </a>
</li>
<li>
    <a href="/pred_app/contact">
        Contact Us
    </a>
</li>
{% endblock nav %}
{% block content %}
<!-- Styles -->
<style>
#chartdiv {
  width: 100%;
  height: 500px;
  max-width: 100%;
}
</style>

<!-- Resources -->
<script src="https://www.amcharts.com/lib/4/core.js"></script>
<script src="https://www.amcharts.com/lib/4/charts.js"></script>
<script src="https://www.amcharts.com/lib/4/themes/material.js"></script>
<script src="https://www.amcharts.com/lib/4/themes/animated.js"></script>

<!-- Chart code -->
<script>
am4core.ready(function() {

// Themes begin
am4core.useTheme(am4themes_material);
am4core.useTheme(am4themes_animated);
```

```
// Themes end

// Create chart
var chart = am4core.create("chartdiv", am4charts.XYChart);
chart.padding(0, 15, 0, 15);

// Load external data
var jsonObject = JSON.parse('{{ predicted_result_df | escapejs}}');
chart.data = jsonObject

// the following line makes value axes to be arranged vertically.
chart.leftAxesContainer.layout = "vertical";

// uncomment this line if you want to change order of axes
//chart.bottomAxesContainer.reverseOrder = true;

var dateAxis = chart.xAxes.push(new am4charts.DateAxis());
dateAxis.renderer.grid.template.location = 0;
dateAxis.renderer.ticks.template.length = 8;
dateAxis.renderer.ticks.template.strokeOpacity = 0.1;
dateAxis.renderer.grid.template.disabled = true;
dateAxis.renderer.ticks.template.disabled = false;
dateAxis.renderer.ticks.template.strokeOpacity = 0.2;
dateAxis.renderer.minLabelPosition = 0.01;
dateAxis.renderer.maxLabelPosition = 0.99;
dateAxis.keepSelection = true;
dateAxis.minHeight = 30;

dateAxis.groupData = true;
dateAxis.minZoomCount = 5;

// these two lines makes the axis to be initially zoomed-in
// dateAxis.start = 0.7;
// dateAxis.keepSelection = true;

var valueAxis = chart.yAxes.push(new am4charts.ValueAxis());
valueAxis.tooltip.disabled = true;
```

```
valueAxis.zIndex = 1;
valueAxis.renderer.baseGrid.disabled = true;
// height of axis
valueAxis.height = am4core.percent(65);

valueAxis.renderer.gridContainer.background.fill = am4core.color("#000000");
valueAxis.renderer.gridContainer.background.fillOpacity = 0.05;
valueAxis.renderer.inside = true;
valueAxis.renderer.labels.template.verticalCenter = "bottom";
valueAxis.renderer.labels.template.padding(2, 2, 2, 2);

//valueAxis.renderer.maxLabelPosition = 0.95;
valueAxis.renderer.fontSize = "0.8em"

var series = chart.series.push(new am4charts.LineSeries());
series.dataFields.dateX = "Date";
series.dataFields.valueY = "Close";
series.tooltipText = "{valueY.value}";
series.name = "MSFT: Value";
series.defaultState.transitionDuration = 0;

var valueAxis2 = chart.yAxes.push(new am4charts.ValueAxis());
valueAxis2.tooltip.disabled = true;
// height of axis
valueAxis2.height = am4core.percent(35);
valueAxis2.zIndex = 3
// this makes gap between panels
valueAxis2.marginTop = 30;
valueAxis2.renderer.baseGrid.disabled = true;
valueAxis2.renderer.inside = true;
valueAxis2.renderer.labels.template.verticalCenter = "bottom";
valueAxis2.renderer.labels.template.padding(2, 2, 2, 2);
//valueAxis.renderer.maxLabelPosition = 0.95;
valueAxis2.renderer.fontSize = "0.8em"

valueAxis2.renderer.gridContainer.background.fill = am4core.color("#000000");
valueAxis2.renderer.gridContainer.background.fillOpacity = 0.05;
```

```
chart.cursor = new am4charts.XYCursor();

var scrollbarX = new am4charts.XYChartScrollbar();
scrollbarX.series.push(series);
scrollbarX.marginBottom = 20;
scrollbarX.scrollbarChart.xAxes.getIndex(0).minHeight = undefined;
chart.scrollbarX = scrollbarX;



/**
 * Set up external controls
 */

// Date format to be used in input fields
var inputFieldFormat = "yyyy-MM-dd";

document.getElementById("b1m").addEventListener("click", function() {
  resetButtonClass();
  var max = dateAxis.groupMax["day1"];
  var date = new Date(max);
  date.setMonth(date.getMonth() - 1);

  dateAxis.zoomToDates(
    date,
    new Date(max)
  );
  //this.className = "amcharts-input amcharts-input-selected";
});



document.getElementById("b3m").addEventListener("click", function() {
  resetButtonClass();
  var max = dateAxis.groupMax["day1"];
  var date = new Date(max);
```

```
  date.setMonth(date.getMonth() - 3);

  dateAxis.zoomToDates(
    date,
    new Date(max)
  );
  //this.className = "amcharts-input amcharts-input-selected";
});

document.getElementById("b6m").addEventListener("click", function() {
  resetButtonClass();
  var max = dateAxis.groupMax["day1"];
  var date = new Date(max);
  date.setMonth(date.getMonth() - 6);

  dateAxis.zoomToDates(
    date,
    new Date(max)
  );
  //this.className = "amcharts-input amcharts-input-selected";
});


document.getElementById("b1y").addEventListener("click", function() {
  resetButtonClass();
  var max = dateAxis.groupMax["week1"];
  var date = new Date(max);
  date.setFullYear(date.getFullYear() - 1);

  dateAxis.zoomToDates(
    date,
    new Date(max)
  );
  //this.className = "amcharts-input amcharts-input-selected";
});
```

```
document.getElementById("bytd").addEventListener("click", function() {
  resetButtonClass();
  var date = new Date(dateAxis.max);
  date.setMonth(0, 1);
  date.setHours(0, 0, 0, 0);
  dateAxis.zoomToDates(date, new Date(dateAxis.max));
  //this.className = "amcharts-input amcharts-input-selected";
});

document.getElementById("bmax").addEventListener("click", function() {
  resetButtonClass();
  dateAxis.zoom({start:0, end:1});
  //this.className = "amcharts-input amcharts-input-selected";
});

function resetButtonClass() {
  var selected = document.getElementsByClassName("amcharts-input-selected");
  for(var i = 0; i < selected.length; i++) {
    selected[i].className = "amcharts-input";
  }
}

dateAxis.events.on("selectionextremeschanged", function() {
  updateFields();
});

dateAxis.events.on("extremeschanged", updateFields);

function updateFields() {
  var minZoomed = dateAxis.minZoomed +
am4core.time.getDuration(dateAxis.mainBaseInterval.timeUnit, dateAxis.mainBaseInterval.count) *
0.5;
  document.getElementById("fromfield").value = chart.dateFormatter.format(minZoomed,
inputFieldFormat);
  document.getElementById("tofield").value = chart.dateFormatter.format(new
```

```
Date(dateAxis.maxZoomed), inputFieldFormat);
}

document.getElementById("fromfield").addEventListener("keyup", updateZoom);
document.getElementById("tofield").addEventListener("keyup", updateZoom);

var zoomTimeout;
function updateZoom() {
  if (zoomTimeout) {
    clearTimeout(zoomTimeout);
  }
  zoomTimeout = setTimeout(function() {
    resetButtonClass();
    var start = document.getElementById("fromfield").value;
    var end = document.getElementById("tofield").value;
    if ((start.length < inputFieldFormat.length) || (end.length < inputFieldFormat.length)) {
      return;
    }
    var startDate = chart.dateFormatter.parse(start, inputFieldFormat);
    var endDate = chart.dateFormatter.parse(end, inputFieldFormat);

    if (startDate && endDate) {
      dateAxis.zoomToDates(startDate, endDate);
    }
  }, 500);
}

}); // end am4core.ready()
</script>

<!-- HTML -->
    <div id="controls" style="width: 100%; overflow: hidden;">
    <div style="float: left; margin-left: 15px;">
    From: <input type="text" id="fromfield" class="amcharts-input" />
    To: <input type="text" id="tofield" class="amcharts-input" />
    </div>
    <div style="float: right; margin-right: 15px;
```

```
    <button id="b1m" class="amcharts-input">1m</button>
    <button id="b3m" class="amcharts-input">3m</button>
    <button id="b6m" class="amcharts-input">6m</button>
    <button id="b1y" class="amcharts-input">1y</button>
    <button id="bytd" class="amcharts-input">YTD</button>
    <button id="bmax" class="amcharts-input">MAX</button>
    </div>
    </div>
<div id="chartdiv"></div>
{% endblock content %}
```

CHAPTER-10

SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

10.1 TYPES OF TESTS

10.1.1 UNIT TESTING

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive.

Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results. Unit testing is usually conducted as part of a combined code and unit test phaseof the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

10.1.1.1 TEST STRATEGY AND APPROACH

Field testing will be performed manually and functional tests will be written in detail.

10.1.1.2 TEST STRATEGIES

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

10.1.1.3 FEATURES TO BE TESTED

- Verify that the entries are of the correct format
- No duplicate entries should be allowed.

- All links should take the user to the correct page.

.

10.1.2 INTEGRATION TESTING

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concernedwith the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components. Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

The following are the types of Integration Testing**:**

10.1.1.4 TOP DOWN INTEGRATION

This method is an incremental approach to the construction of program structure. Modules are integrated by moving downward through the control hierarchy, beginning with the main program module. The module subordinates to the main program module are incorporated into the structure in either a depth first or breadth first manner.

In this method, the software is tested from main module and individual stubs are replaced when the test proceeds downwards.

10.1.1.5 BOTTOM UP INTEGRATION

This method begins the construction and testing with the modules at the lowest levelin the program structure. Since the modules are integrated from the bottom up, processing required for modules subordinate to a given level is always available and the need for stubs is eliminated. The bottom up integration strategy may be implemented with the following steps:

- The low-level modules are combined into clusters into clusters that perform a specific Software sub-function.
- A driver (i.e.) the control program for testing is written to coordinate test case input and output.

- The cluster is tested.
- Drivers are removed and clusters are combined moving upward in the program structure

The bottom up approaches tests each module individually and then each module is module is integrated with a main module and tested for functionality.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

10.1.2 FUNCTIONAL TESTING

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

10.1.3 SYSTEM TESTING

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

10.1.4 WHITE BOX TESTING

White Box Testing is a testing in which in which the software tester hasknowledge of the inner workings, structure and language of the software, or at least its purpose.It is purpose. It is used to test areas that cannot be reached from a black box level.

10.1.5 BLACK BOX TESTING

Black Box Testing is testing the software without any knowledge of the innerworkings, structure or language of the module being tested. Black box tests, as most other kindsof tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in whichthe software under test is treated, as a black box .you cannot "see" into it. The test provides inputs and responds to outputs without considering how the software works.

10.1.6 ACCEPTANCE TESTING

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

10.2 TESTING METHODOLOGIES

The following are the Testing Methodologies:

- **Unit Testing.**
- **Integration Testing.**
- **User Acceptance Testing.**
- **Output Testing.**
- **Validation Testing.**

10.2.1 UNIT TESTING

Unit testing focuses verification effort on the smallest unit of Software design that is the module. Unit testing exercises specific paths in a module's control structure to ensure complete coverage and maximum error detection. This test focuses on each module individually, ensuring that it functions properly as a unit. Hence, the naming is Unit Testing.

During this testing, each module is tested individually and the module interfaces are verified for the consistency with design specification. All-important processing path are tested for the expected results. All error handling paths are also tested.

10.2.2 INTEGRATION TESTING

Integration testing addresses the issues associated with the dual problems of verification and program construction. After the software has been integrated a set of highorder tests are conducted. The main objective in this testing process is to take unit tested modules and builds a program structure that has been dictated by design.

10.2.3 USER ACCEPTANCE TESTING

User Acceptance of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changes wherever required. The system developed provides a friendly user interface that can easily be understood even by a person who is new to the system.

10.2.4 OUTPUT TESTING

After performing the validation testing, the next step is output testing of the proposed system, since no system could be useful if it does not produce the required output in the specified format. Asking the users about the format required by them tests the outputs generated or displayed by the system under consideration. Hence the output format is considered in 2 ways – one is on screen and another in printed format.

10.2.5 VALIDATION CHECKING

Validation checks are performed on the following fields.

10.2.5.1 TEXT FIELD

The text field can contain only the number of characters lesser than or equal to its size. The text fields are alphanumeric in some tables and alphabetic in other tables. Incorrect entry always flashes and error message.

10.2.5.2 NUMERIC FIELD

The numeric field can contain only numbers from 0 to 9. An entry of any character flashes an error messages. The individual modules are checked for accuracy and what it has to

perform. Each module is subjected to test run along with sample data. The individually tested modules are integrated into a single system. Testing involves executing the real data information is used in the program the existence of any program defect is inferred from the output. The testing should be planned so that all the requirements are individually tested.

A successful test is one that gives out the defects for the inappropriate data and produces and output revealing the errors in the system.

10.3 PREPARATION OF TEST DATA

Taking various kinds of test data does the above testing. Preparation of test data plays a vital role in the system testing. After preparing the test data the system under study is tested using that test data. While testing the system by using test data errors are again uncovered and corrected by using above testing steps and corrections are also noted for future use.

10.3.1 USING LIVE TEST DATA

Live test data are those that are actually extracted from organization files. After a system is partially constructed, programmers or analysts often ask users to key in a set of data from their normal activities. Then, the systems person uses this data as a way to partially test the system. In other instances, programmers or analysts extract a set of live data from the files and they have entered themselves.

It is difficult to obtain live data in sufficient amounts to conduct extensive testing. And, although it is realistic data that will show how the system will perform for the typical processing requirement, assuming that the live data entered are in fact typical, such data generally will not test all combinations or formats that can enter the system. This bias toward typical values then does not provide a true systems test and in fact ignores the cases most likely to cause system failure.

10.3.2 USING ARTIFICIAL TEST DATA

Artificial test data are created solely for test purposes, since they can be generated to test all combinations of formats and values. In other words, the artificial data, which can quickly be prepared by a data generating utility program in the information systems department, make possible the testing of all login and control paths through the program.

The most effective test programs use artificial test data generated by persons other than those who wrote the programs. Often, an independent team of testers formulates a testing plan, usingthe systems specifications.

The package “Virtual Private Network” has satisfied all the requirements specified as per software requirement specification and was accepted.

10.4 USER TRAINING

Whenever a new system is developed, user training is required to educate them about the working of the system so that it can be put to efficient use by those for whom the system has been primarily designed. For this purpose the normal working of the project was demonstrated to the prospective users. Its working is easily understandable and since the expected users are people who have good knowledge of computers, the use of this system is very easy.

10.5 MAINTAINENCE

This covers a wide range of activities including correcting code and design errors. To reduce the need for maintenance in the long run, we have more accurately defined the user’s requirements during the process of system development. Depending on the requirements, this system has been developed to satisfy the needs to the largest possible extent. With development in technology, it may be possible to add many more features based on the requirements in future. The coding and designing is simple and easy to understand which will make maintenance easier.

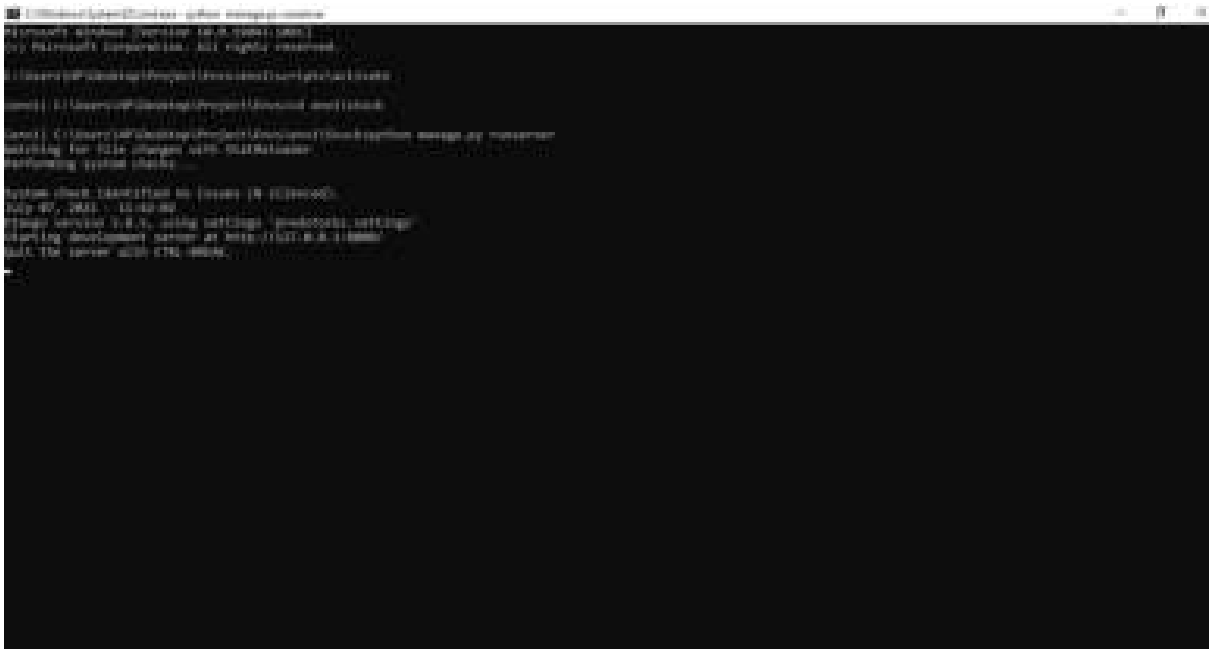
10.6 TESTING STRATEGY

A strategy for system testing integrates system test cases and design techniques into a well-planned series of steps that results in the successful construction of software. The testing strategy must co-operate test planning, test case design, test execution, and the resultant data collection and evaluation. A strategy for software testing must accommodate low-level tests that are necessary to verify that a small source code segment has been correctly implemented as well as high level tests that validate major system functions against user requirements.

Software testing is a critical element of software quality assurance and represents the ultimatereview of specification design and coding.

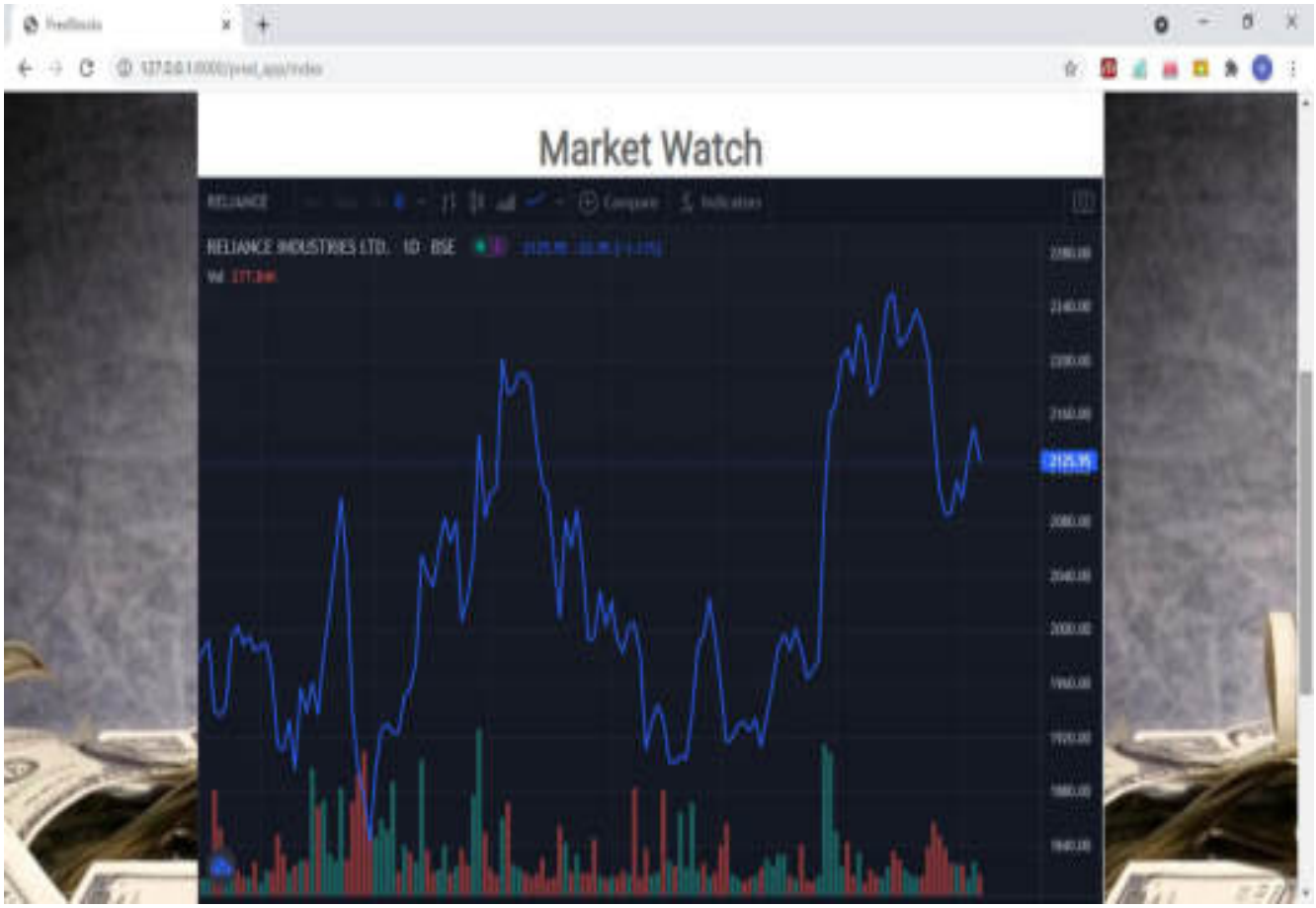
CHAPTER-11

EXECUTION AND OUTPUT

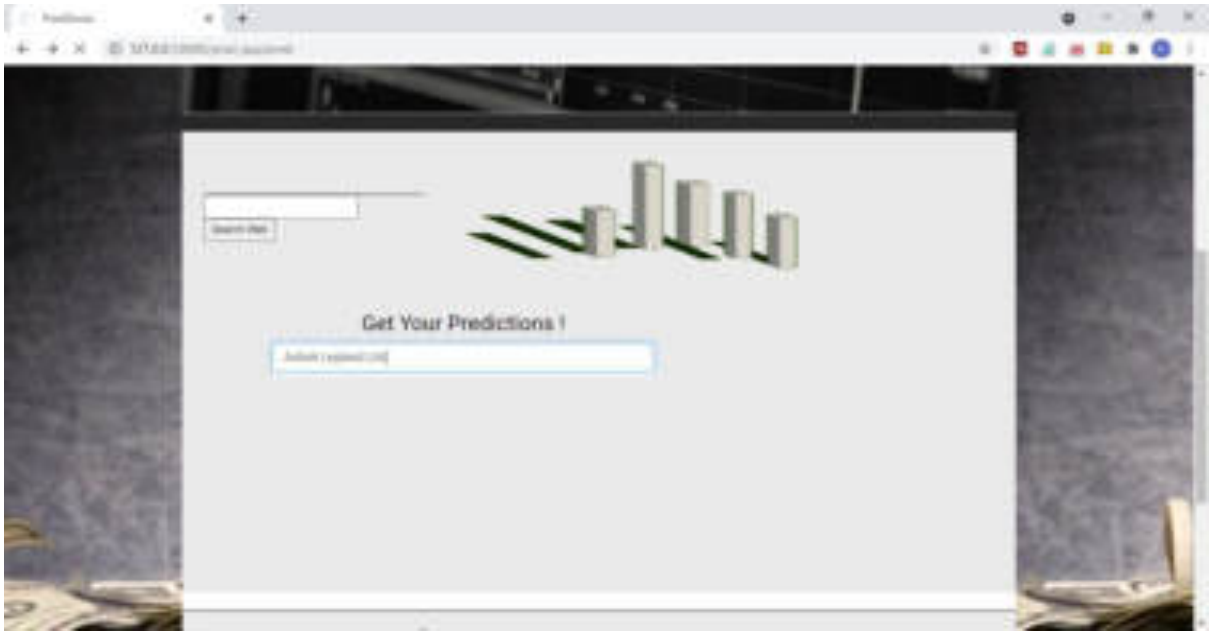


Scripts in the virtual environment need to be activated first, then go to the folder stock which contains all the code and then runserver. After the execution we will get a address or url. After this paste address or url in any browser to get result.

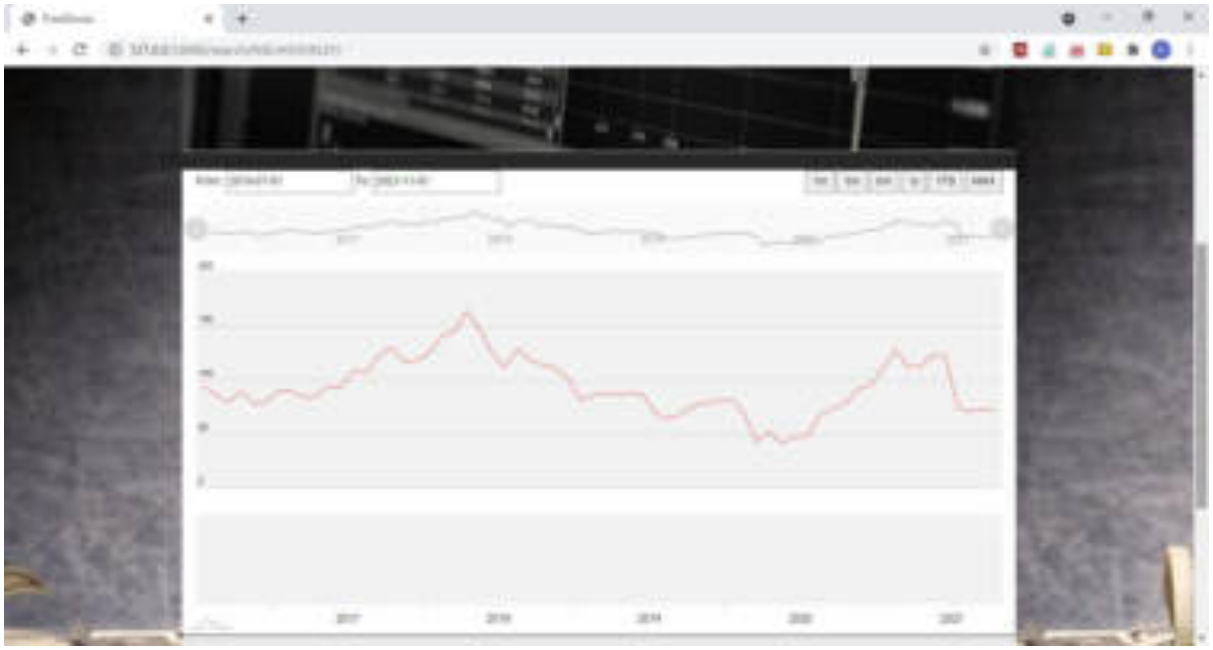
After pasting the url the above home page is diplayed which helps us to choose the options needed like prediction and contact us.



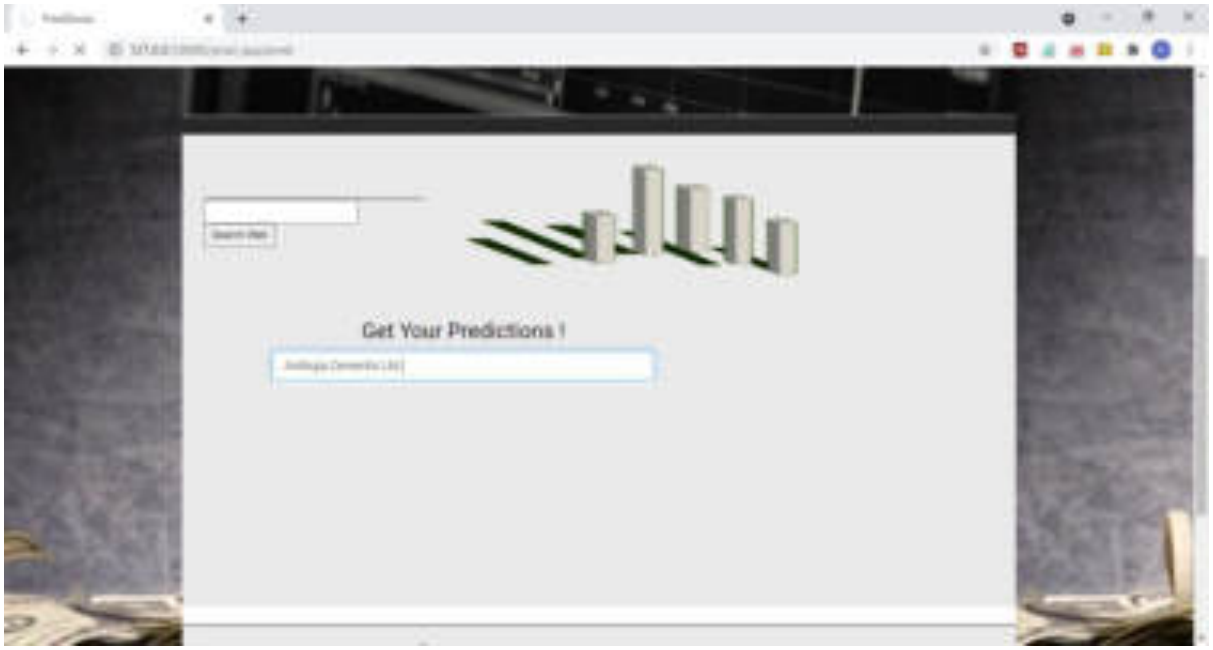
The above output is the stock price prediction of Reliance Industries LTD based on the real time data. In the home page the market watch gives the above result.



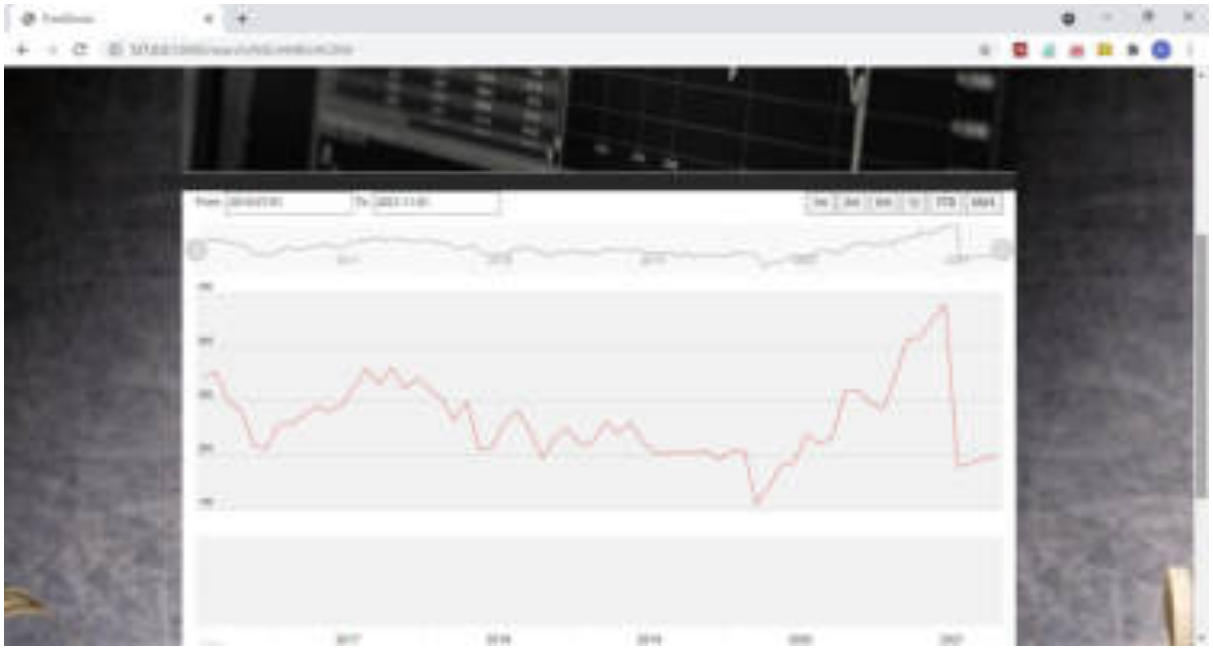
When we select the predict option in home page then we will get the above page. The rectangle box under the Get Your Predictions is used to select the company name for which the stock price need to be predicted. It predicts based on historical data.



The above graph is the stock price values of Ashok Layland LTD. We can also select the duration from which month to which month or which year to which year it should predict by using from and to options in the above page.



Similarly, The rectangle box under the Get Your Predictions is used to select the company name for which the stock price need to be predicted. It predicts based on historical data. The list of companies for which we can predict here are Ashok Layland Ltd, Ambuja Cements Ltd, Apollo tyres, Reliance Ltd, Amar Raja Batteries Ltd, Adithya Birla, Asian Paints etc.



The above graph is the stock price values of Ambuja Cements Ltd. We can also select the duration from which month to which month or which year to which year it should predict by using from and to options in the above page.

CHAPTER-12

CONCLUSION

This paper was an attempt to determine the future prices of the stocks of a company with greater accuracy and reliability using machine learning techniques. The primary contribution of the researchers being the application of the novel LSTM Model as a means of determining the stock prices. Both the techniques have shown an improvement in the accuracyof predictions, thereby yielding positive results with the LSTM model proving to be more efficient.

The results are quite promising and has led to the conclusion that it is possible to predict stock market with more accuracy and efficiency using machine learning techniques. In the future, the accuracy of the stock market prediction system can be further improved by utilizing a much bigger dataset than the one being utilized currently. Furthermore, other emerging models of Machine Learning could also be studied to check for the accuracy rate resulted by them.

Sentiment analysis though Machine Learning on how news affects the stockprices of a company is also a very promising area. Other deep learning based models can alsobe used for prediction purposes.

CHAPTER-13

FUTURE ENHANCEMENTS

- Potential improvement can be made to the data collection and analysis method.

- Future research can be done with possible improvement such as more refined data and more accurate algorithm.

- Along with this research would also like to study and implement economic growth model for stock market prediction and the analysis of how economic growth model will effect in stock market prediction in comparision to linear regression model and with specialized machine learning techniques.

- Since the data is very large displaying the result (prediction graph) is taking minimum 5-10 minutes. So it should be minimized in the future.

CHAPTER-14

REFERENCES

[1] M. Usmani , S. H. Adil, K. Raza and S. S. A. Ali, "Stock market prediction using machine learning techniques," 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, 2016, pp. 322-327.

[2] K. Raza, "Prediction of Stock Market performance by using machine learning techniques," 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT), Karachi, 2017, pp. 1-1.

[3] H. Gunduz, Z. Cataltepe and Y. Yaslan, "Stock market direction prediction using deep neural networks," 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya, 2017, pp. 1-4.

[4] M. Billah, S. Waheed and A. Hanifa, "Stock market prediction using an improved training algorithm of neural network," 2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), Rajshahi, 2016, pp. 1-4.

[5] H. L. Siew and M. J. Nordin, "Regression techniques for the prediction of stock price trend," 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), Langkawi, 2012, pp. 1-5.

[6] K. V. Sujatha and S. M. Sundaram, "Stock index prediction using regression and neural network models under non normal conditions," INTERACT-2010, Chennai, 2010, pp. 59-63.

[7] S. Liu, G. Liao and Y. Ding, "Stock transaction prediction modelling and analysis based on LSTM," 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, 2018, pp. 2787-2790.

[8] T. Gao, Y. Chai and Y. Liu, "Applying long short term memory neural networks for predicting stock closing price," 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2017, pp. 575-578.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"IMAGE CAPTIONING USING DEEP LEARNING"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
COMPUTER SCIENCE AND ENGINEERING

Submitted by

G.LEENA (17H71A0583)

T.GIRIDHAR NAGA SAI (17H71A0582)

V.LAKSHMIPRASANNA (17H71A0582)

B.MAHESH BABU(17H71A0584)

Under the Esteemed Guidance of

Ms.Mareshwaramma

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited B.Tech (CSE|ECE|ME)
Accredited by NAAC A With Grade

2020-21

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC With Grade A

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"Image Captioning Using Deep Learning"** is a bonafide work carried out by **G.Leena (17H71A0583), V.Lakshmi Prasanna (17H71A0582), T.Giridhar Naga Sai (17H71A0570), B.Mahesh Babu (17H71A0584)** in partial fulfilment for the award of degree of Bachelor of Technology in **Computer Science and Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

P. Mareshwaramma
Project Guide

D. Prasad
Head of the Department

Examiner

Dr K Srinivas
PRINCIPAL

ACKNOWLEDGEMENT

We would like to take this opportunity to express my deepest appreciation to the following people for their valuable contributions and assistance with this Project.

Initially, we would like to thank my project supervisor, **Ms.Mareshwaramma, M.tech, Department of Computer Science and Engineering** for the guidance and support, especially for the valuable ideas and knowledge shared to me throughout the Project.

We have the immense pleasure in expressing my thanks and deep sense of gratitude to**, Mr. D. Prasad, Head of the Department, Computer Science and Engineering** for extending necessary facilities and support for the completion of the Project.

We whole sincerely acknowledge **Dr. K. Srinivas, Principal and Prof D. PANDURANGA RAO, CEO** for giving opportunity to take up the Project work. We also extend my thanks to all faculty members of **Computer Science and Engineering**, for their valuable guidance and encouragement.

We would like to extend my warm appreciation to all my friends for sharing us their knowledge, valuable contributions and help with this Project.

Finally, my special thanks go to my family for their continuous support and help throughout my academic years and for their continual support and encouragement for the completion of the project.

DECLARATION

We **G.Leena (17H71A0583), V.Lakshmi Prasanna (17H71A0582), T.Giridhar Naga Sai (17H71A0570), B.Mahesh Babu (17H71A0584)** of the Main-Project "**Image Captioning Using Deep Learning**", hereby declare that the matter embodied in this Project is the genuine work done by us and has not been submitted either to this university or to any other university/institute for the fulfilment of the requirement of any course of study.

G.LEENA **(17H71A0583)**

V.LAKSHMI PRASANNA **(17H71A0582)**

T.GIRIDHAR NAGA SAI **(17H71A0570)**

B.MAHESH BABU **(17H71A0584)**

CONTENTS

ABSTRACT

Every day, we encounter a large number of images from various sources such as the internet, news articles, document diagrams and advertisements. These sources contain images that viewers would have to interpret themselves. Most images do not have a description, but the human can largely understand them without their detailed captions. However, machine needs to interpret some form of image captions if humans need automatic image captions from it.

Ever since researchers started working on object recognition in images, it became clear that only providing the names of the objects recognized does not make such a good impression as a full human-like description. As long as machines do not think, talk, and behave like humans, natural language descriptions will remain a challenge to be solved. In this project, Image Captioning using deep learning, is the process of generation of textual description of an image. This helps to summarize the details in the picture without the interference of a person and only using Deep Learning as a leverage.

CHAPTER 1

INTRODUCTION

1.1 MOTIVATION

In the past few years, computer vision in image processing area has made significant progress, like Image Classification and Object Detection. Benefiting from the advances of image classification and object detection, it becomes possible to automatically generate one or more sentences to understand the visual content of an image, which is the problem known as Image Captioning.

1.2 PURPOSE

As long as machines do not think, talk, and behave like humans, natural language descriptions will remain a challenge to be solved. Traditional object detection and image classification task is needed to identify objects within the image. With the help of the task Image Captioning, which is not just identifying the objects but also identifying the relationship between them and total scene understanding of the image.After understanding the scene it is also required to generate a human like description of that image

1.3 PROBLEM STATEMENT

Given a new image, an image captioning algorithm should output a description about this image at a semantic level. For example, in the following figure, the input image consists of people, boards and the waves. In the bottom, there is a sentence describing the content of the image— the objects emerging in the image, the action and the scene are all described in this sentence.

A couple of people riding waves on top of boards.

1.4 BACKGROUND

For the image captioning task, humans can easily understand the image content and express it in the form of natural language sentences according to specific needs; however, for computers, it requires the integrated use of image processing, computer vision, natural language processing and other major areas of research results. The challenge of image captioning is to design a model that can fully use image information to generate more human-like rich image descriptions. The meaningful description generation process of high level image semantics requires not only the understanding of objects or scene recognition in the image, but also the ability to analyze their states, understand the relationship among them and generate a semantically and syntactically correct sentence. It is currently unclear how the brain understands an image and organizes the visual information into a caption. Image captioning involves a deep understanding of the world and which things are salient parts of the whole.

Image Captioning is performed by following key tasks in order. At first features are extracted. After proper extraction of features different objects from an image are detected, after that the relationship between objects are to be identified (i.e. if objects are cat and grass it is to be identified that if cat is on grass). Once objects are detected and relationships are identified now it is required to generate the text description, i.e. Sequence of words in orderly form that they make a good sentence according to the relationship between the image objects. This involves both Deep Learning and Natural language processing technologies.

Using these networks various different methods are developed to perform image captioning in various different domains. However, still, there is room for the machine to make capable enough to generate descriptions like a human.

1.5 SCOPE

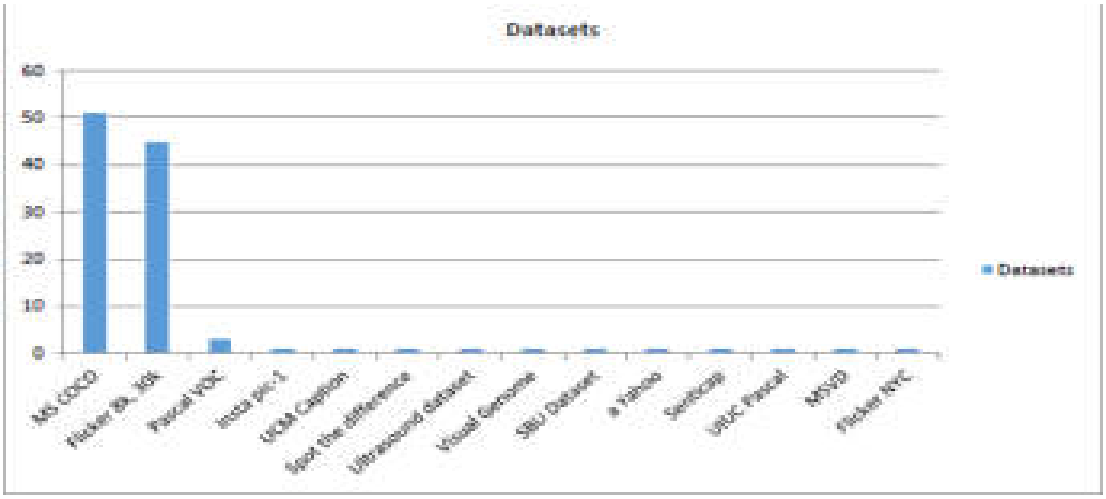
Generating complete and natural image descriptions automatically has large potential effects, such as titles attached to news images, descriptions associated with medical images, text-based image retrieval, information accessed for blind users, human-robot interaction. These applications in image captioning have important theoretical and practical research value. Therefore, image captioning is a more complicated but meaningful task in the age of artificial intelligence.

CHAPTER 2

LITERATURE SURVEY

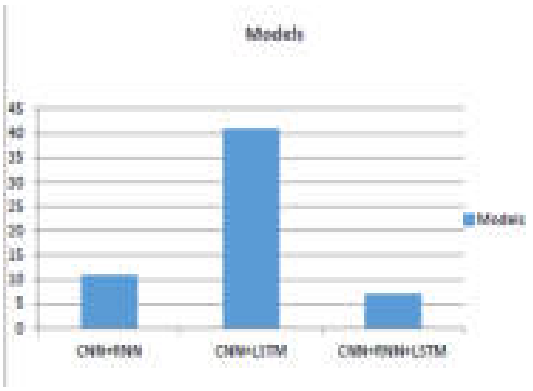
Image captioning is a very recent and growing research problem nowadays. Day by day various new methods are being introduced to achieve satisfactory results in this field. However, there are still lots of attention required to achieve results as good as a human. Image captioning is defined as the process of generating captions or textual descriptions for images based on the contents of the image. It is a machine learning task that involves both natural language processing (for text generation) and computer vision (for understanding image contents).

To develop an Image captioning model, there are many datasets on which this model can be trained on for acquiring some best results. In literature most common used data sets are MSCOCO and flicker 8k and 30k. Moreover for a text description of specific task like in medical or traffic movement description their own dedicated datasets are created.

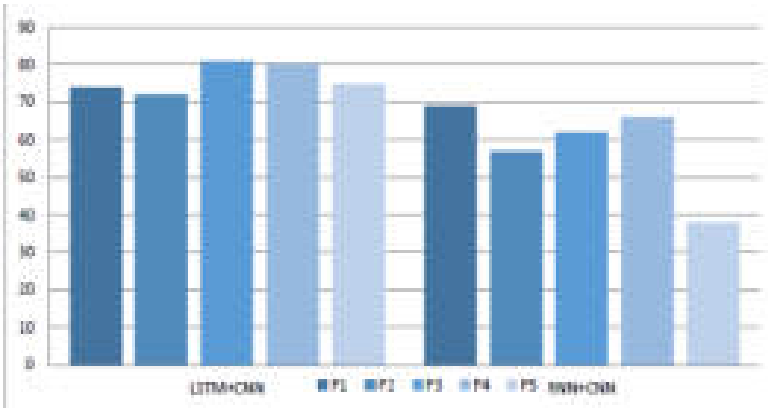


Deep learning network used for images is Convolution neural network. CNN has been proved best to map image data into output variable. There are various prebuilt model that take advantage from this feature of CNN i.e. RCNN faster RCNN etc. These models are used for object detection and localization in images which is very necessary task in image captioning since it's not just classification task and understanding image contents is necessary. Once image data is understood there is need of predicting the sequence of words to generate the text for that particular image. For sequence prediction two most famous networks are Recurrent Neural Network (RNN) and long short term memory (LSTM). For image captioning generation task CNN is either used

with RNN or LSTM where CNN is used for understanding image contents and RNN or LSTM for text description generation.

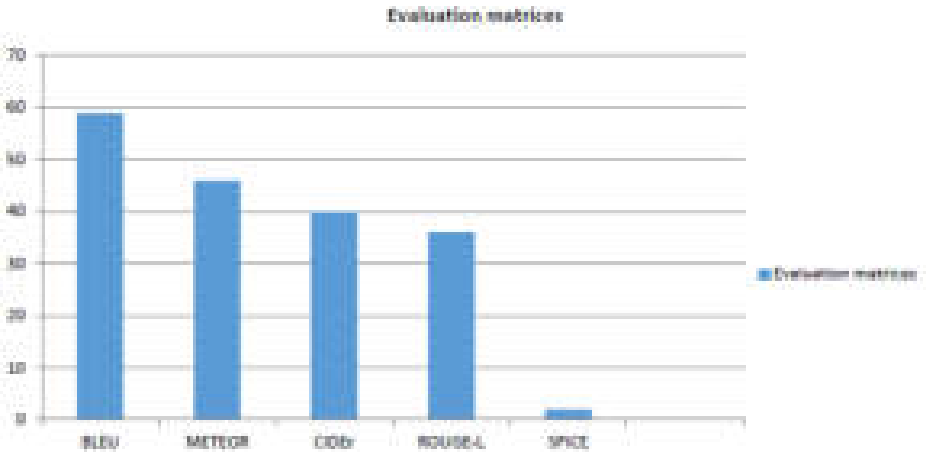


DeepLearning Models used for Image captioning Models



Comparison of best scores achieved

Evaluating the trained model is quiet difficult task in image captioning for this purpose various evaluation matrices are created. Most common evaluation mechanisms found in literature are BLEU, ROUGE-L, CIDEr, METEOR, and SPICE. It is found that BLEU score is most popular method of evaluation used by almost all of the studies. BLEU stands for bilingual evaluation understudy. It is an evaluation mechanism widely use in text generation. It is a mechanism for comparing the machine generated text with one or more manually written text. So basically it summarizes that how close a generated text is to an expected text. BLEU score is majorly prevalent in automated machine translation but it can be also used in image captioning, text summarization, speech recognition etc.



CHAPTER 3

SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

For generating captions for the given image, at first features are extracted.After proper extraction of features different objects from an image are detected, after that the relationship between objects are to be identified (i.e. if objects are cat and grass it is to be identified that if cat is on grass). Once objects are detected and relationships are identified now it is required to generate the text description, i.e. Sequence of words in orderly form that they make a good sentence according to the relationship between the image objects.

For Feature extraction, Convolutional Neural Networks (CNNs) were used, and for the generation of captions after extracting features Recurrent Neural Networks (RNNs) were used in this system.

A recurrent neural network is one type of Artificial Neural Network (ANN) and is used in application areas of Natural Language Processing (NLP) and Speech Recognition. An RNN model is designed to recognize the sequential characteristics of data and thereafter using the patterns to predict the coming scenario. In RNN, the output from the previous steps is fed into the input of the current state. For instance, to predict the next letter of any word or to predict the next word of the sentence, there is a need to remember the previous letters or the words and store them in some form of memory.

3.1.1 DISADVANTAGES:

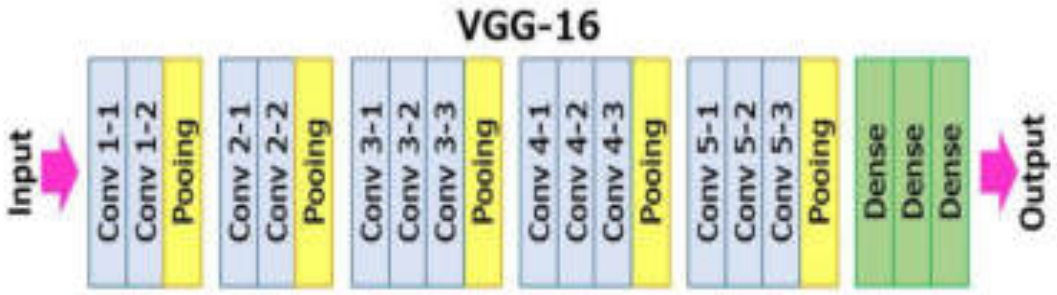
1. Leads to Vanishing Gradient or Exploding Gradient problems.

2. Results in Long-term dependencies.

3. Not much accurate

3.2 PROPOSED SYSTEM

For Image caption generation, first extraction of features of an image must be done. We are using a pre-trained model to interpret the content of the photos.There are many models to choose from. In this case, we will use the Oxford Visual Geometry Group, or VGG, model that won the ImageNet competition in 2014. This is called VGG16 model, because it contains 16 Convolutional layers.



For caption generation, we are using a type of RNN model i.e., LSTM (Long-Short Term Memory) model. LSTM networks are an extension of recurrent neural networks (RNNs) mainly introduced to handle situations where RNNs fail.LSTM's have a nature of remembering information for a long periods of time, which is also their default behavior. LSTM can retain important information over time using memory cells.

This network model has three important gates additional to the one operation in RNN (cell state). These four gates Input gate, Forget gate and Output gate helps in remembering the information that is most needed and forgetting the information which is no longer needed.

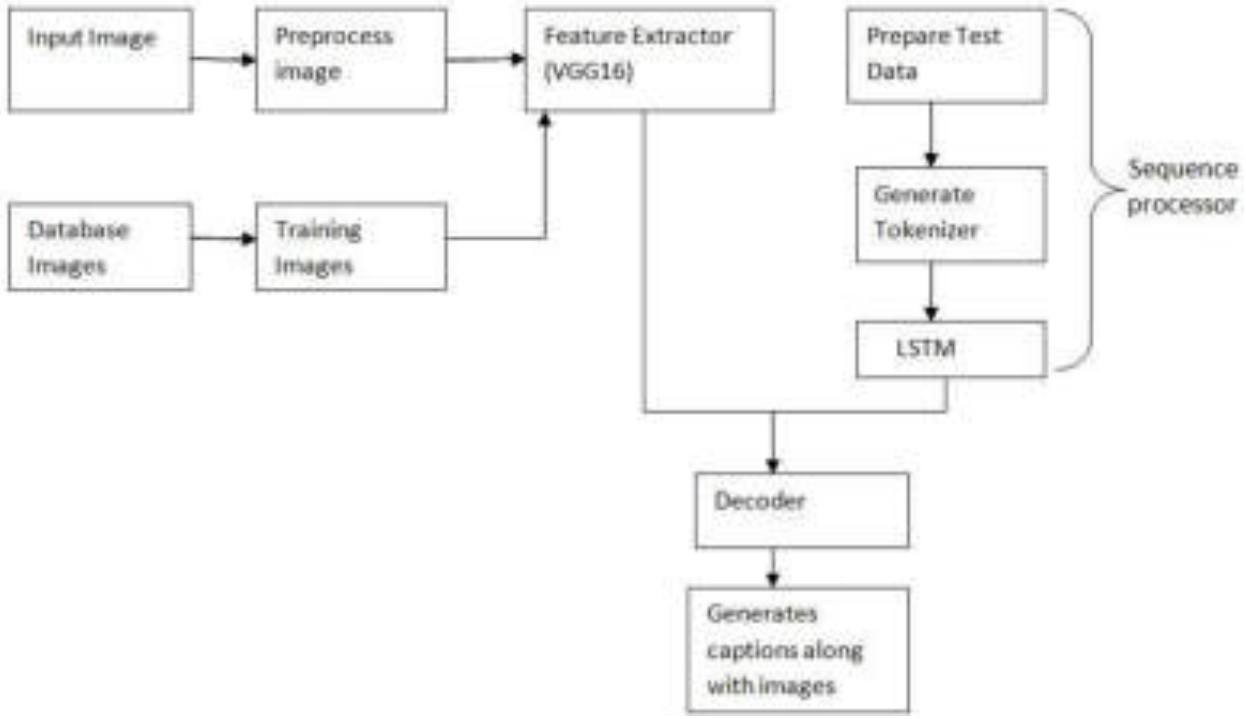
3.2.1 ARCHITECTURE:

Proposed System architecture consists of,

Photo Feature Extractor: This is a 16-layer VGG model pre-trained on the ImageNet dataset. We have pre-processed the photos with the VGG model (without the output layer) and will use the extracted features predicted by this model as input.

Sequence Processor: This is a word embedding layer for handling the text input, followed by a Long Short-Term Memory (LSTM) recurrent neural network layer.

Decoder: Both the feature extractor and sequence processor output a fixed-length vector. These are merged together and processed by a Dense layer to make a final prediction.



3.2.2 ADVANTAGES:

1. Long term dependencies are solved.
2. Vanishing gradient problem is solved.
3. Results in accuracy and efficiency.

CHAPTER 4

SYSTEM DESIGN

4.1 SOFTWARE REQUIREMENT SPECIFICATIONS

Software requirements specification is a rigorous assessment of requirements before the more specific system design stages, and its goal is to reduce later redesign. It should also provide a realistic basis for estimating product costs, risks, and schedules. Used appropriately, software requirements specifications can help prevent software project failure. The software requirements specification document lists sufficient and necessary requirements for the project development. To derive the requirements, the developer needs to have clear and thorough understanding of the products under development.

4.1.1 LANGUAGE USED :

Python (version 3.8.5) is a general purpose, dynamic, high level and interpreted programming language. It supports Object Oriented programming approach to develop applications. It is simple and easy to learn and provides lots of high-level datastructures.

- Python is easy to learn yet powerful and versatile scripting language which makes it attractive for Application Development.
- Python's syntax and dynamic typing with its interpreted nature, makes it an ideal language for scripting and rapid applicationdevelopment.
- Python supports multiple programming pattern, including object oriented, imperative and functional or procedural programmingstyles.
- Python is not intended to work on special area such as web programming. That is why it is known as multipurpose because it can be used with web, enterprise, 3D CADetc.

4.1.2 FRAMEWORKS:

TensorFlow (version 2.4.3)

TensorFlow is an end-to-end open source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML powered application. TensorFlow can train and run deep neural networks for handwritten digit

classification, image recognition, word embeddings, recurrent neural networks, sequence-to-sequence models for machine translation, natural language processing, and PDE (partial differential equation) based simulations. Best of all, TensorFlow supports production prediction at scale, with the same models used for training.

Keras (version 2.3.0)

Keras is a deep learning API written in Python, running on top of the machine learning platform TensorFlow. It was developed with a focus on enabling fast experimentation. Being able to go from idea to result as fast as possible is key to doing good research.

Keras is:

- **Simple** -- but not simplistic. Keras reduces developer cognitive load to free you to focus on the parts of the problem that really matter.
- **Flexible** -- Keras adopts the principle of progressive disclosure of complexity: simple workflows should be quick and easy, while arbitrarily advanced workflows should be *possible* via a clear path that builds upon what you've already learned.
- **Powerful** -- Keras provides industry-strength performance and scalability: it is used by organizations and companies including NASA, YouTube, or Waymo

4.1.3 IDEs (Integrated Development Environment):

JupyterNotebook:

Thisisaweb-based,interactivecomputingnotebookenvironment.We can edit and run human-readable docs while describing the dataanalysis.The Jupyter Notebook is an open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text. Jupyter Notebook is maintained by the people at ProjectJupyter.

Jupyter Notebooks are a spin-off project from the IPython project, which used tohave an IPython Notebook project itself. The name, Jupyter, comes from the core supported programming languages that it supports: Julia, Python, and R. Jupyter ships with the IPython kernel, which allows you to write your programs in Python, but there are currently over 100 other kernels that you can alsouse.

The Jupyter Notebook is not included with Python, so if you want to try it out, you will need to install Jupyter. There are many distributions of the Python language. This article will focus on just two of them for the purposes of installing Jupyter Notebook. The most popular is CPython, which is the reference version of Python that you can get from their website. It is also assumed that you are usingPython.

Spyder Editor :

Spyder is a powerful scientific environment written in Python, for Python, and designed by and for scientists, engineers and data analysts. It offers a unique combination of the advanced editing, analysis, debugging, and profiling functionality of a comprehensive development tool with the data exploration, interactive execution, deep inspection, and beautiful visualization capabilities of a scientific package.

Beyond its many built-in features, its abilities can be extended even further via its plugin system and API. Furthermore, Spyder can also be used as a PyQt5 extension library, allowing you to build upon its functionality and embed its components, such as the interactive console, in your own software.

4.1.4 LIBRARIES :

Numpy :

NumPy is the fundamental package for scientific computing with Python. It contains among other things:

- a powerful N-dimensional arrayobject
- sophisticated (broadcasting) functions
- tools for integrating C/C++ and Fortrancode
- useful linear algebra, Fourier transform, and random numbercapabilities
- Besides its obvious scientific uses, NumPy can also be used as an efficient multi- dimensional container of generic data. Arbitrary data-types can be defined. This allows NumPy to seamlessly and speedily integrate with a wide variety ofdatabases.
- NumPy is licensed under the BSD license, enabling reuse with fewrestrictions.

NLTK :

NLTK (Natural Language Toolkit) is a suite that contains libraries and programs for statistical language processing. It is one of the most powerful NLP (Natural Language Processing) libraries, which contains packages to make machines understand human language and reply to it with an appropriate response.

NLTK is suitable for linguists, engineers, students, educators, researchers, and industry users alike. NLTK is available for Windows, Mac OS X, and Linux. Best of all, NLTK is a free, open source, community-driven project.

NLTK has been called “a wonderful tool for teaching, and working in, computational linguistics using Python,” and “an amazing library to play with natural language.”

4.2 UML DIAGRAMS

Use-oriented techniques are widely used in software requirement analysis and design. Use cases and usage scenarios facilitate system understanding and provide a common language for communication. This paper presents a scenario-based modeling technique and discusses its applications. In this model, scenarios are organized hierarchically and they capture the system functionality at various abstraction levels including scenario groups, scenarios, and sub-scenarios. Combining scenarios or sub-scenarios can form complex scenarios. Data are also separately identified, organized, and attached to scenarios. This scenario model can be used to cross check with the UML model. It can also direct systematic scenario-based testing including test case generation, test coverage analysis with respect to requirements, and functional regression testing.

4.2.1 CLASS DIAGRAM

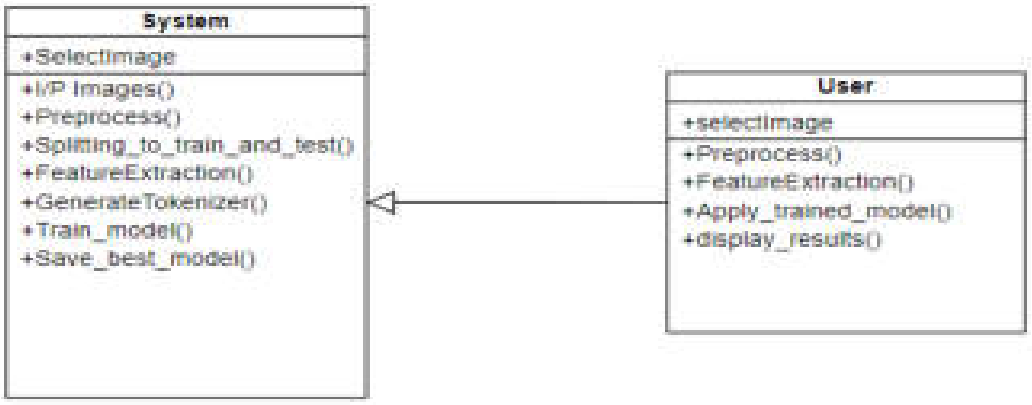
Model, objects are entities that combine state (i.e., data), behavior (i.e., procedures, or methods) and identity (unique existence among all other objects). The structure and behavior of an object are defined by a class, which is a definition, or blueprint, of all objects of a specific type. An object must be explicitly created based on a class and an object thus created is considered to be an instance of that class. An object is similar to a structure, with the addition of method pointers, member access control, and an implicit data member which locates instances of the class (i.e. actual objects of that class) in the class hierarchy (essential

for runtime inheritance features)

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, and the relationships between the classes.

The class diagram is the main building block in object oriented modeling. It is used both for general conceptual modeling of the semantics of the application, and for detailed modeling translating the models into programming code. The classes in a class diagram represent both the main objects and or interactions in the application and the objects to be programmed. In the class diagram these classes are represented with boxes which contain the two parts:

- The upper part holds the name of theclass.
- The middle part contains the attributes of theclass.
- The lower part contains the operations of theclass.



4.2.2 USECASE DIAGRAM

A Use Case Diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those usecases.

The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted. Interaction among actors is not shown on the use case diagram. If this interaction is essential to a

coherent description of the desired behavior, perhaps the system or use case boundaries should be re-examined. Alternatively, interaction among actors can be part of the assumptions used in the usecase.

Use cases:

A use case describes a sequence of actions that provide something of measurable value to an actor and is drawn as a horizontal ellipse.

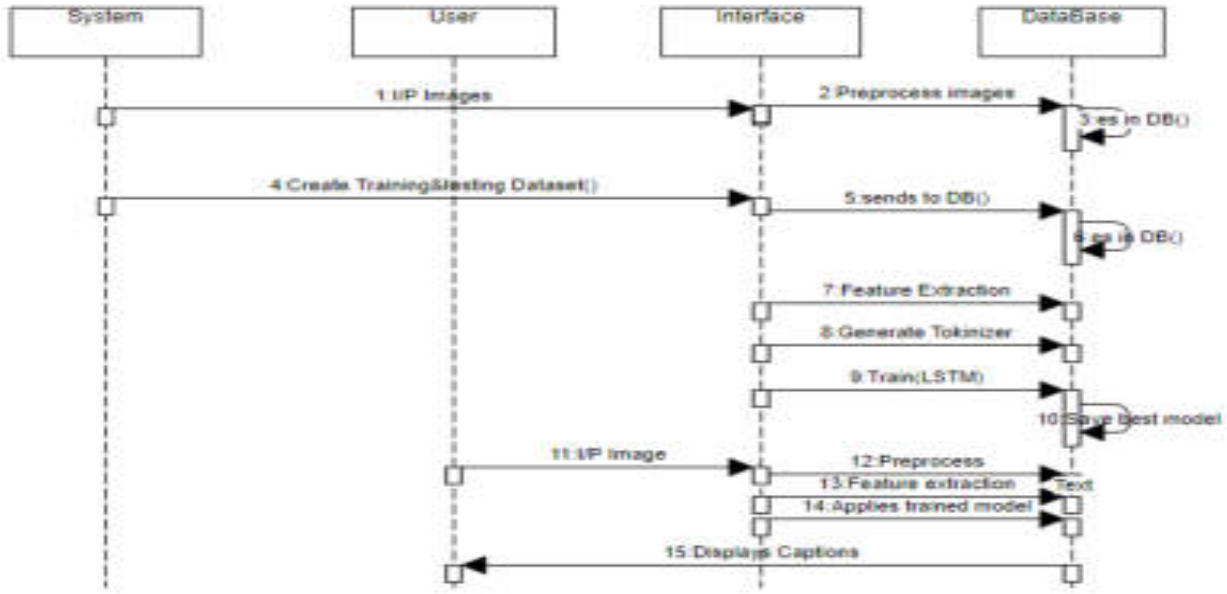
Actors

An actor is a person, organization, or external system that plays a role in one or more interactions with the system.

4.2.3 SEQUENCE DIAGRAM

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart.

Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams. A sequence diagram shows, as parallel vertical lines (lifelines), different processes or objects that live simultaneously, and, as horizontal arrows, the messages exchanged between them, in the order in which they occur. This allows the specification of simple runtime scenarios in a graphical manner. If the lifeline is that of an object, it demonstrates a role. Note that leaving the instance name blank can represent anonymous and unnamed instances. In order to display interaction, messages are used. These are horizontal arrows with the message name written above them. Solid arrows with full heads are synchronous calls, solid arrows with stick heads are asynchronous calls and dashed arrows with stick heads are return messages.

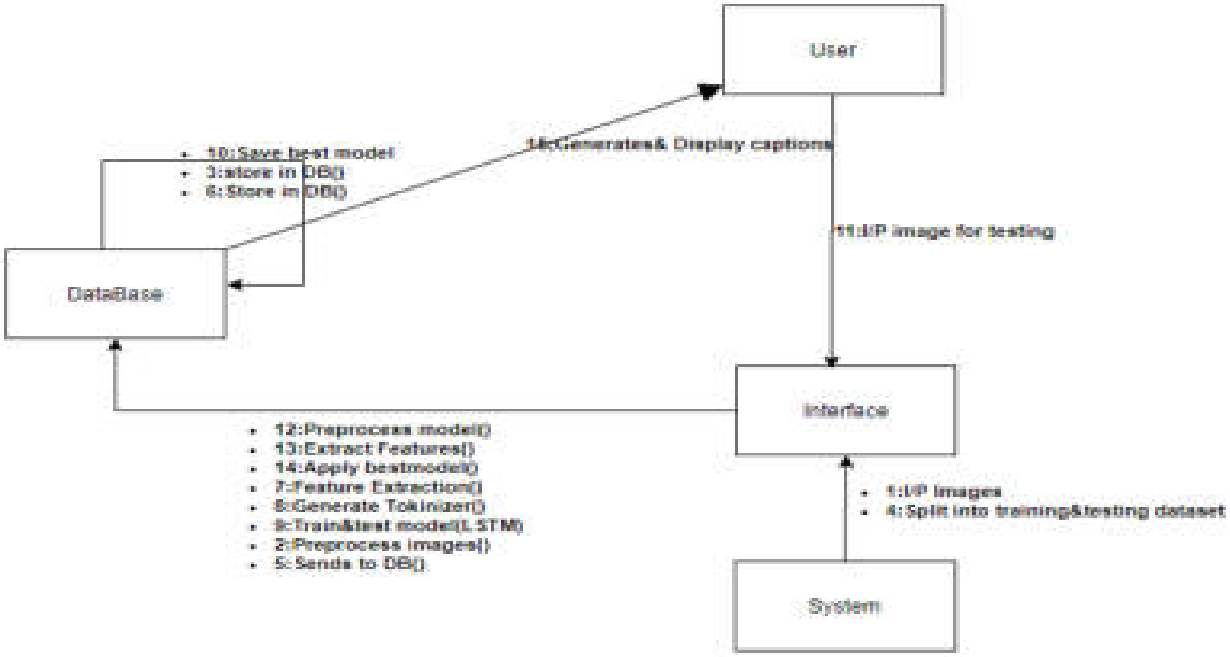


4.2.4 COLLABORATION DIAGRAM

A Sequence diagram is dynamic, and, more importantly, is time ordered. A Collaboration diagram is very similar to a Sequence diagram in the purpose it achieves; in

other words, it shows the dynamic interaction of the objects in a system. A distinguishing feature of a Collaboration diagram is that it shows the objects and their association with other objects in the system apart from how they interact with each other. The association between objects is not represented in a Sequence diagram.

A Collaboration diagram is easily represented by modeling objects in a system and representing the associations between the objects as links. The interaction between the objects is denoted by arrows. To identify the sequence of invocation of these objects, a number is placed next to each of thesearrows.

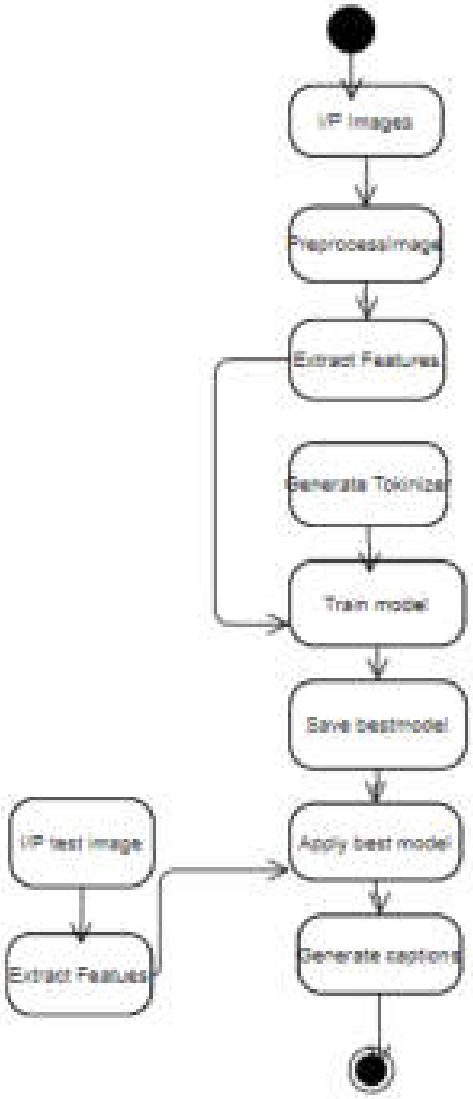


4.2.5 ACTIVITY DIAGRAM

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control. Activity diagrams are constructed from a limited repertoire of shapes, connected with arrows. The most important shape types:

- rounded rectangles represent activities;diamonds representdecisions

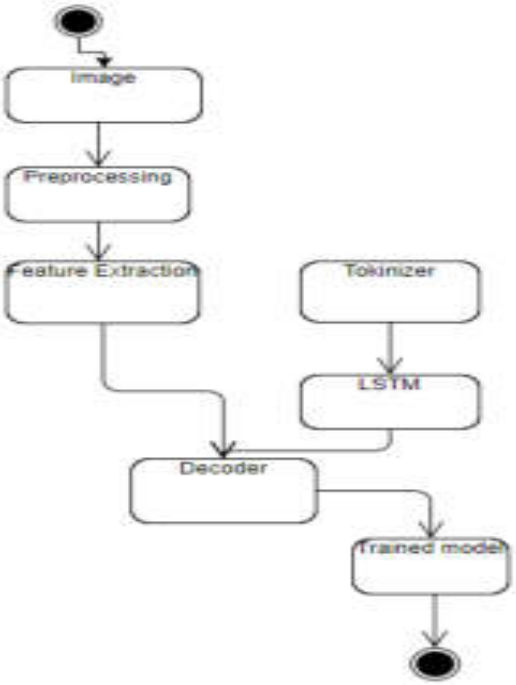
- bars represent the start (split) or end (join) of concurrentactivities
- a black circle represents the start (initial state) of theworkflow
- An encircled black circle represents the end (finalstate).



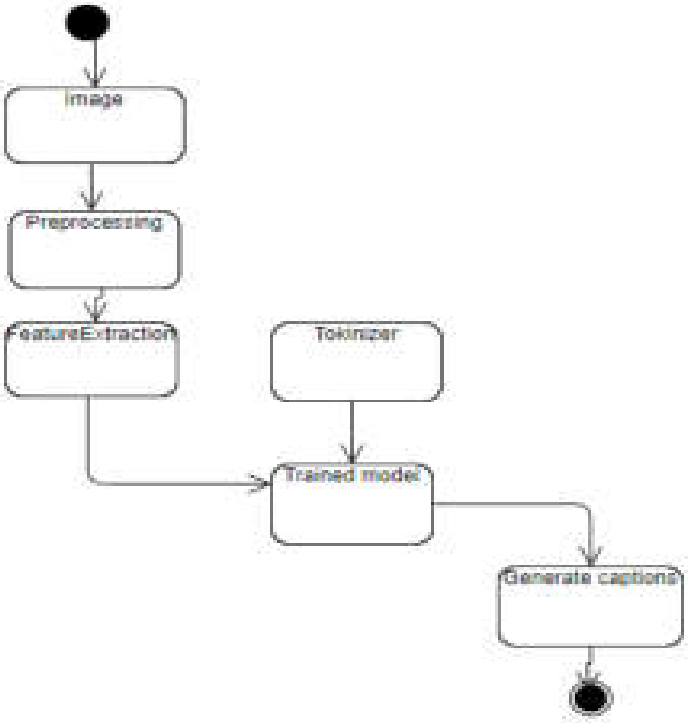
4.2.6 STATE CHART DIAGRAM

A state diagram, also called a state machine diagram or state chart diagram, is an illustration of the states an object can attain as well as the transitions between those states in the Unified Modeling Language. A state diagram resembles a flowchart in which the initial state is represented by a large black dot and subsequent states are portrayed as boxes with rounded corners.

STATE CHART DIAGRAM FOR SYSTEM

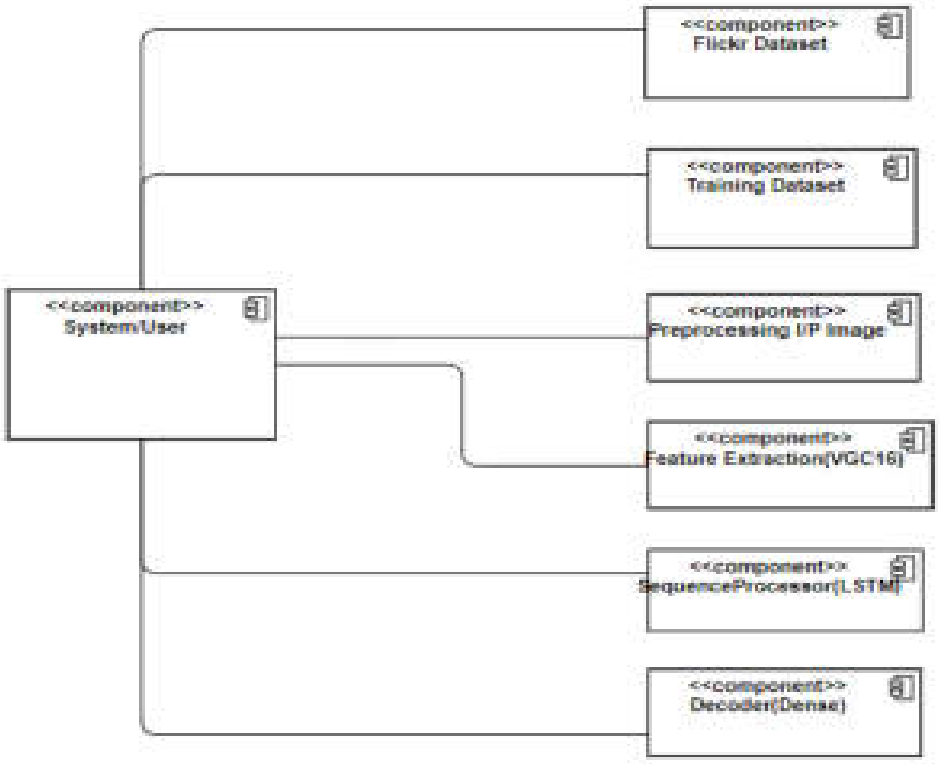


STATE CHART DIAGRAM FOR USER



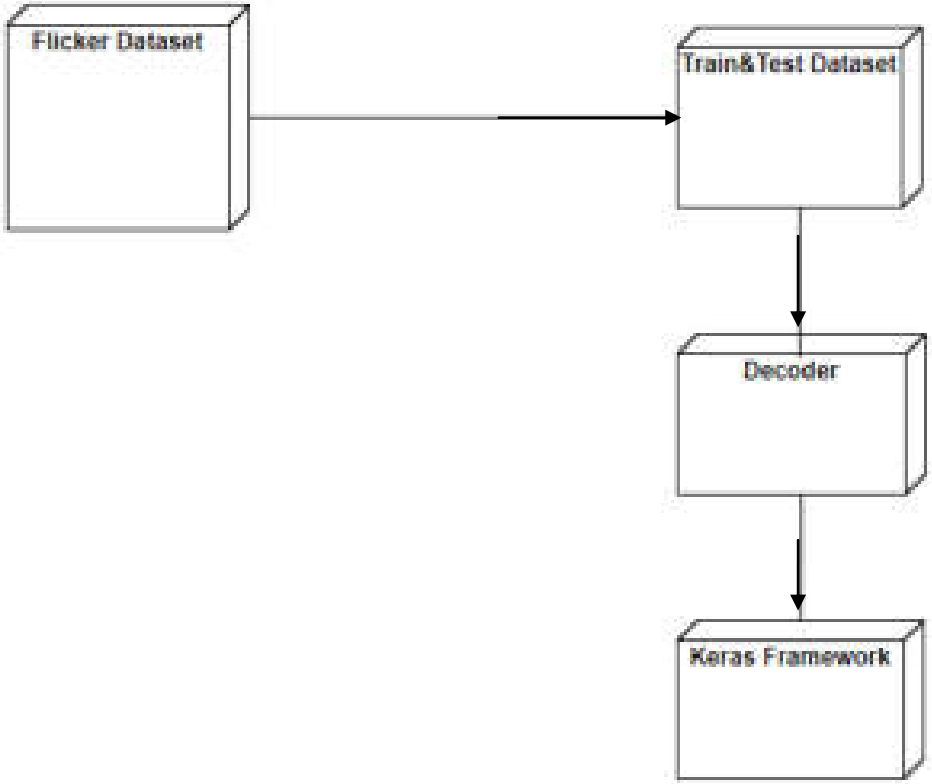
4.2.7 COMPONENT DIAGRAM

A component contains a set of collaborating classes. Each class within a component has been fully elaborated to include all attributes and operations that are relevant to its implementation. As part of the design elaboration, all interfaces (messages) that enable the classes to communicate and collaborate with other design classes must also be defined. To accomplish this, the designer begins with the analysis model and elaborates analysis classes (for components that relate to the problem domain) and infrastructure classes(or components that provide support services for the problem domain).



4.2.8 DEPLOYMENT DIAGRAM

Deployment diagrams are used to visualize the topology of the physical components of a system where the software components aredeployed.So deployment diagrams are used to describe the static deployment view of a system.



CHAPTER 5

IMPLEMENTATION

5.1 MODULES

5.1.1 THE DATASET

The dataset we are using is Flickr8k Dataset. The datasets are of the following size

- **Flickr8k_Dataset.zip** (1 Gigabyte) An archive of all photographs.
- **Flickr8k_text.zip** (2.2 Megabytes) An archive of all text descriptions for photographs.

After downloading the datasets and extracting them. Extracted directories are

- **Flickr8k_Dataset**: Contains 8092 photographs in JPEG format.
- **Flickr8k_text**: Contains a number of files containing different sources of descriptions for the photographs.

The dataset has a pre-defined training dataset (6,000 images), development dataset (1,000 images), and test dataset (1,000 images).

5.1.2. PHOTO FEATURE EXTRACTION

For feature extraction, weuse a pre-trained model to interpret the content of the photos.There are many models to choose from. In this case, we will use the Oxford Visual Geometry Group, or VGG, model that won the ImageNet competition in 2014. This model is called VGG16, because of the 16 convolutional layers.

Keras provides this pre-trained model directly. The first time we use this model, Keras will download the model weights from the Internet, which are about 500 Megabytes. We could use this model as part of a broader image caption model. The problem is, it is a large model and running each photo through the network every time we want to test a new language model configuration (downstream) is redundant.

Instead, we pre-compute the “photo features” using the pre-trained model and save them to file. We can then load these features later and feed them into our model as the interpretation of a given photo in the dataset. It is no different to running the photo through the full VGG model; it is

just we will have done it once in advance.This is an optimization that will make training our models faster and consume less memory.

The function named *extract_features()* that, given a directory name, will load each photo, prepare it for VGG, and collect the predicted features from the VGG model. The image features are a 1-dimensional 4,096 element vector.The function returns a dictionary of image identifier to image features.We call this function to prepare the photo data for testing our models, then save the resulting dictionary to a file named '*features.pkl*'.

5.1.3. PREPROCSSING TEXT DATA

The dataset contains multiple descriptions for each photograph and the text of the descriptions requires some minimal cleaning. First, we load the file containing all of the descriptions. Each photo has a unique identifier. This identifier is used on the photo filename and in the text file of descriptions.

The function *load_descriptions*() that, given the loaded document text, will return a dictionary of photo identifiers to descriptions. Each photo identifier maps to a list of one or more textual descriptions. Next, we need to clean the description text. The descriptions are already tokenized and easy to work with.

We cleaned the text in the following ways in order to reduce the size of the vocabulary of words

- Convert all words to lowercase.
- Remove all punctuation.
- Remove all words that are one character or less in length (e.g. 'a').
- Remove all words with numbers in them.

The *clean_descriptions()* function that, given the dictionary of image identifiers to descriptions, steps through each description and cleans the text. Once cleaned, we can summarize the size of the vocabulary.Ideally, we want a vocabulary that is both expressive and as small as possible. A smaller vocabulary will result in a smaller model that will train faster.

The *save_descriptions()* function that, given a dictionary containing the mapping of identifiers to descriptions and a filename, saves the mapping to file. Running the example first prints the number of loaded photo descriptions (8,092) and the size of the clean vocabulary (8,763 words).

Finally, we saved the dictionary of image identifiers and descriptions to a new file named '*descriptions.txt'*, with one image identifier and description per line.

5.1.4 DEVELOPING DEEP LEARNING MODEL

5.1.4.1. Load Data:

First, we load the prepared photo and text data so that we can use it to fit the model.We are going to train the data on all of the photos and captions in the training datasetThe train and development dataset have been predefined in the*Flickr_8k.trainImages.txt* and *Flickr_8k.devImages.txt* files respectively, that both contain lists of photo file names. From these file names, we can extract the photo identifiers and use these identifiers to filter photos and descriptions for each set.

The function *load_clean_descriptions()* loads the cleaned text descriptions from '*descriptions.txt*' for a given set of identifiers and returns a dictionary of identifiers to lists of text descriptions.The model we will develop will generate a caption given a photo, and the caption will be generated one word at a time.

Next, we can load the photo features for a given dataset. The function named *load_photo_features()* that loads the entire set of photo descriptions, then returns the subset of interest for a given set of photo identifiers.The *create_tokenizer()* function that will fit a Tokenizer given the loaded photo description text.

5.1.4.2. Define Model :

We will describe the model in three parts:

Photo Feature Extractor. This is a 16-layer VGG model pre-trained on the ImageNet dataset. We have pre-processed the photos with the VGG model (without the output layer) and will use the extracted features predicted by this model as input.

Sequence Processor. This is a word embedding layer for handling the text input, followed by a Long Short-Term Memory (LSTM) recurrent neural network layer.

Decoder. Both the feature extractor and sequence processor output a fixed-length vector. These are merged together and processed by a Dense layer to make a final prediction.

- The Photo Feature Extractor model expects input photo features to be a vector of 4,096 elements. These are processed by a Dense layer to produce a 256 element representation of the photo.
- The Sequence Processor model expects input sequences with a pre-defined length (34 words) which are fed into an Embedding layer that uses a mask to ignore padded values. This is followed by an LSTM layer with 256 memory units.
- Both the input models produce a 256 element vector. Further, both input models use regularization in the form of 50% dropout. This is to reduce overfitting the training dataset, as this model configuration learns very fast.
- The Decoder model merges the vectors from both input models using an addition operation. This is then fed to a Dense 256 neuron layer and then to a final output Dense layer that makes a Softmax prediction over the entire output vocabulary for the next word in the sequence.

The function named *define_model()* defines and returns the model ready to be fit.

5.1.4.3. Train Model :

The first step is to define a function that we can use as the data generator. The function *data_generator()* will be the data generator and will take the loaded textual descriptions, photo features, tokenizer and max length. Finally, we can use the *fit_generator()* function on the model to train the model with this data generator.

We trained the model for 20 epochs and simply save the model after each training epoch. The models with the lowest loss is considered as the best model and that model is used for evaluation and testing.

5.1.5. EVALUATION OF MODEL

Once the model is fit, we evaluate the skill of its predictions on the holdout test dataset. We will evaluate a model by generating descriptions for all photos in the test dataset and evaluating those predictions with a standard cost function. We will generate predictions for all photos in the test dataset and in the train dataset.

The function named *evaluate_model()* will evaluate a trained model against a given dataset of photo descriptions and photo features. The actual and predicted descriptions are collected and evaluated collectively using the corpus BLEU score that summarizes how close the generated text is to the expected text.

BLEU scores are used in text translation for evaluating translated text against one or more reference translations. We compared each generated description against all of the reference descriptions for the photograph. We then calculate BLEU scores for 1, 2, 3 and 4 cumulative n-grams.The NLTK Python library implements the BLEU score calculation in the *corpus_bleu()* function. A higher score close to 1.0 is better, a score closer to zero is worse.

5.1.6. GENERATION OF TOKENIZER

We need the Tokenizer for encoding generated words for the model while generating a sequence, and the maximum length of input sequences, used when we defined the model. With the encoding of text, we can create the tokenizer and save it to a file so that we can load it quickly whenever we need it without needing the entire Flickr8K dataset. We can create the Tokenizer as before and save it as a pickle file *tokenizer.pkl.*

5.1.7. TESTING THE MODEL

First, we must load the Tokenizer from *tokenizer.pkl.* Next, we must load the photo for which we want to describe and extract the features. We could do this by re-defining the model and adding the VGG-16 model to it, or we can use the VGG model to predict the features and use them as inputs to our existing model. We used a modified version of the *extract_features()* function used during data preparation, but adapted to work on a single photo.We can then generate a description using the *generate_desc()* function defined when evaluating the model.

5.2 CODE DESIGN CHARACTERISTICS

5.2.1 FRONT END

Tkinter: Python has a lot of GUI Frameworks, but Tkinter is the only framework that's built into the Python standard library. Tkinter has several strengths. It's **cross-platform**, so the same code works on Windows, macOS, and Linux. Visual elements are rendered using native operating system elements, so applications built with Tkinter look like they belong on the platform where they're run.

5.2.2 BACK END

Python : To develop Machine Learning or Deep Learning projects or applications, the mostly used language across the world is Python. This is because of the libraries and frameworks that are provided by this language. For programming complex models, Python is the one language that makes it easier to understand and build models.

5.3 CODE SNIPPETS

a. Photo Feature Extraction

```
from os import listdir
from pickle import dump
from keras.applications.vgg16 import VGG16
from keras.preprocessing.image import load_img
from keras.preprocessing.image import img_to_array
from keras.applications.vgg16 import preprocess_input
from keras.models import Model

# extract features from each photo in the directory
def extract_features(directory):
	# load the model
	model = VGG16()
	# re-structure the model
	model = Model(inputs=model.inputs, outputs=model.layers[-2].output)
	# summarize
	print(model.summary())
	# extract features from each photo
	features = dict()
	for name in listdir(directory):
		# load an image from file
		filename = directory + '/' + name
```

```
            image = load_img(filename, target_size=(224, 224))
            # convert the image pixels to a numpy array
            image = img_to_array(image)
            # reshape data for the model
            image = image.reshape((1, image.shape[0], image.shape[1], image.shape[2]))
            # prepare the image for the VGG model
            image = preprocess_input(image)
            # get features
            feature = model.predict(image, verbose=0)
            # get image id
            image_id = name.split('.')[0]
            # store feature
            features[image_id] = feature
            print('>%s' % name)
    return features

# extract features from all images
directory = 'C:/Project_b4/Flicker8k_Dataset'
features = extract_features(directory)
print('Extracted Features: %d' % len(features))
# save to file
dump(features, open('features.pkl', 'wb'))
```

b. Preprocessing text data

```
import string

# load doc into memory

def load_doc(filename):

        # open the file as read only

        file = open(filename, 'r')

        # read all text

        text = file.read()

        # close the file

        file.close()

        return text


# extract descriptions for images

def load_descriptions(doc):

        mapping = dict()

        # process lines

        for line in doc.split('\n'):

                # split line by white space

                tokens = line.split()

                if len(line) < 2:

                        continue

                # take the first token as the image id, the rest as the description
```

```
            image_id, image_desc = tokens[0], tokens[1:]
            # remove filename from image id
            image_id = image_id.split('.')[0]
            # convert description tokens back to string
            image_desc = ' '.join(image_desc)
            # create the list if needed
            if image_id not in mapping:
                mapping[image_id] = list()
            # store description
            mapping[image_id].append(image_desc)
    return mapping

def clean_descriptions(descriptions):
    # prepare translation table for removing punctuation
    table = str.maketrans('', '', string.punctuation)
    for key, desc_list in descriptions.items():
        for i in range(len(desc_list)):
            desc = desc_list[i]
            # tokenize
            desc = desc.split()
            # convert to lower case
            desc = [word.lower() for word in desc]
            # remove punctuation from each token
```

```
                desc = [w.translate(table) for w in desc]
                # remove hanging 's' and 'a'
                desc = [word for word in desc if len(word)>1]
                # remove tokens with numbers in them
                desc = [word for word in desc if word.isalpha()]
                # store as string
                desc_list[i] =  ' '.join(desc)

# convert the loaded descriptions into a vocabulary of words
def to_vocabulary(descriptions):
        # build a list of all description strings
        all_desc = set()
        for key in descriptions.keys():
                [all_desc.update(d.split()) for d in descriptions[key]]
        return all_desc
# save descriptions to file, one per line
def save_descriptions(descriptions, filename):
        lines = list()
        for key, desc_list in descriptions.items():
                for desc in desc_list:
                        lines.append(key + ' ' + desc)
        data = '\n'.join(lines)
        file = open(filename, 'w')
```

```
        file.write(data)
        file.close()
filename = 'C:/Project_b4/Flickr8k.token.txt'
# load descriptions
doc = load_doc(filename)
# parse descriptions
descriptions = load_descriptions(doc)
print('Loaded: %d ' % len(descriptions))
# clean descriptions
clean_descriptions(descriptions)
# summarize vocabulary
vocabulary = to_vocabulary(descriptions)
print('Vocabulary Size: %d' % len(vocabulary))
# save to file
save_descriptions(descriptions, 'descriptions.txt')
```

c. Developing deep learning model—Defining & Training model

```
from numpy import array
from pickle import load
from keras.preprocessing.text import Tokenizer
from keras.preprocessing.sequence import pad_sequences
from keras.utils import to_categorical
from keras.models import Model
```

```
from keras.layers import Input
from keras.layers import Dense
from keras.layers import LSTM
from keras.layers import Embedding
from keras.layers import Dropout
from keras.layers.merge import add

# load doc into memory
def load_doc(filename):
	# open the file as read only
	file = open(filename, 'r')
	# read all text
	text = file.read()
	# close the file
	file.close()
	return text

# load a pre-defined list of photo identifiers
def load_set(filename):
	doc = load_doc(filename)
	dataset = list()
	# process line by line
	for line in doc.split('\n'):
```

```
            # skip empty lines
            if len(line) < 1:
                continue
            # get the image identifier
            identifier = line.split('.')[0]
            dataset.append(identifier)
    return set(dataset)

# load clean descriptions into memory
def load_clean_descriptions(filename, dataset):
    # load document
    doc = load_doc(filename)
    descriptions = dict()
    for line in doc.split('\n'):
        # split line by white space
        tokens = line.split()
        # split id from description
        image_id, image_desc = tokens[0], tokens[1:]
        # skip images not in the set
        if image_id in dataset:
            # create list
            if image_id not in descriptions:
                descriptions[image_id] = list()
```

```
                # wrap description in tokens
                desc = 'startseq' + ' '.join(image_desc) + 'endseq'
                # store
                descriptions[image_id].append(desc)
    return descriptions

# load photo features
def load_photo_features(filename, dataset):
    # load all features
    all_features = load(open(filename, 'rb'))
    # filter features
    features = {k: all_features[k] for k in dataset}
    return features

# covert a dictionary of clean descriptions to a list of descriptions
def to_lines(descriptions):
    all_desc = list()
    for key in descriptions.keys():
        [all_desc.append(d) for d in descriptions[key]]
    return all_desc

# fit a tokenizer given caption descriptions
def create_tokenizer(descriptions):
```

```
    lines = to_lines(descriptions)
    tokenizer = Tokenizer()
    tokenizer.fit_on_texts(lines)
    return tokenizer

# calculate the length of the description with the most words
def max_length(descriptions):
    lines = to_lines(descriptions)
    return max(len(d.split()) for d in lines)

# create sequences of images, input sequences and output words for an image
def create_sequences(tokenizer, max_length, desc_list, photo, vocab_size):
    X1, X2, y = list(), list(), list()
    # walk thrn ough each description for the image
    for desc in desc_list:
        # encode the sequence
        seq = tokenizer.texts_to_sequences([desc])[0]
        # split one sequence into multiple X,y pairs
        for i in range(1, len(seq)):
            # split into input and output pair
            in_seq, out_seq = seq[:i], seq[i]
            # pad input sequence
            in_seq = pad_sequences([in_seq], maxlen=max_length)[0]
```

```
                # encode output sequence
                out_seq = to_categorical([out_seq], num_classes=vocab_size)[0]
                # store
                X1.append(photo)
                X2.append(in_seq)
                y.append(out_seq)
    return array(X1), array(X2), array(y)

# define the captioning model
def define_model(vocab_size, max_length):
    # feature extractor model
    inputs1 = Input(shape=(4096,))
    fe1 = Dropout(0.5)(inputs1)
    fe2 = Dense(256, activation='relu')(fe1)
    # sequence model
    inputs2 = Input(shape=(max_length,))
    se1 = Embedding(vocab_size, 256, mask_zero=True)(inputs2)
    se2 = Dropout(0.5)(se1)
    se3 = LSTM(256)(se2)
    # decoder model
    decoder1 = add([fe2, se3])
    decoder2 = Dense(256, activation='relu')(decoder1)
    outputs = Dense(vocab_size, activation='softmax')(decoder2)
```

```
	# tie it together [image, seq] [word]
	model = Model(inputs=[inputs1, inputs2], outputs=outputs)
	# compile model
	model.compile(loss='categorical_crossentropy', optimizer='adam')
	# summarize model
	model.summary()
	#plot_model(model, to_file='model.png', show_shapes=True)
	return model

# data generator, intended to be used in a call to model.fit_generator()
def data_generator(descriptions, photos, tokenizer, max_length, vocab_size):
	# loop for ever over images
	while 1:
		for key, desc_list in descriptions.items():
			# retrieve the photo feature
			photo = photos[key][0]
			in_img, in_seq, out_word = create_sequences(tokenizer, max_length, desc_list, photo, vocab_size)
			yield [in_img, in_seq], out_word

# load training dataset (6K)
filename = 'C:/Project_b4/Flickr_8k.trainImages.txt'
train = load_set(filename)
```

```
print('Dataset: %d' % len(train))

# descriptions

train_descriptions=load_clean_descriptions('C:/Project_b4/descriptions.txt', train)

print('Descriptions: train=%d' % len(train_descriptions))

# photo features

train_features = load_photo_features('C:/Project_b4/features.pkl', train)

print('Photos: train=%d' % len(train_features))

# prepare tokenizer

tokenizer = create_tokenizer(train_descriptions)

vocab_size = len(tokenizer.word_index) + 1

print('Vocabulary Size: %d' % vocab_size)

# determine the maximum sequence length

max_length = max_length(train_descriptions)

print('Description Length: %d' % max_length)


# define the model

model = define_model(vocab_size, max_length)

# train the model, run epochs manually and save after each epoch

epochs = 20

steps = len(train_descriptions)

for i in range(epochs):

        # create the data generator
```

```
        generator = data_generator(train_descriptions, train_features, tokenizer, max_length,
vocab_size)
        # fit for one epoch
        model.fit_generator(generator, epochs=1, steps_per_epoch=steps, verbose=1)
        # save model
        model.save('model_' + str(i) + '.h5')
```

d. evaluation of model

```
from numpy import argmax
from pickle import load
from keras.preprocessing.text import Tokenizer
from keras.preprocessing.sequence import pad_sequences
from keras.models import load_model
from nltk.translate.bleu_score import corpus_bleu

# load doc into memory
def load_doc(filename):
        # open the file as read only
        file = open(filename, 'r')
        # read all text
        text = file.read()
        # close the file
        file.close()
```

```
        return text


# covert a dictionary of clean descriptions to a list of descriptions
def to_lines(descriptions):
        all_desc = list()
        for key in descriptions.keys():
                [all_desc.append(d) for d in descriptions[key]]
        return all_desc


# load a pre-defined list of photo identifiers
def load_set(filename):
        doc = load_doc(filename)
        dataset = list()
        # process line by line
        for line in doc.split('\n'):
                # skip empty lines
                if len(line) < 1:
                        continue
                # get the image identifier
                identifier = line.split('.')[0]
                dataset.append(identifier)
        return set(dataset)
```

```
# load clean descriptions into memory
def load_clean_descriptions(filename, dataset):
	# load document
	doc = load_doc(filename)
	descriptions = dict()
	for line in doc.split('\n'):
		# split line by white space
		tokens = line.split()
		# split id from description
		image_id, image_desc = tokens[0], tokens[1:]
		# skip images not in the set
		if image_id in dataset:
			# create list
			if image_id not in descriptions:
				descriptions[image_id] = list()
			# wrap description in tokens
			desc = 'startseq' + ' '.join(image_desc) + 'endseq'
			# store
			descriptions[image_id].append(desc)
	return descriptions

# load photo features
def load_photo_features(filename, dataset):
```

```
    # load all features
    all_features = load(open(filename, 'rb'))
    # filter features
    features = {k: all_features[k] for k in dataset}
    return features

# fit a tokenizer given caption descriptions
def create_tokenizer(descriptions):
    lines = to_lines(descriptions)
    tokenizer = Tokenizer()
    tokenizer.fit_on_texts(lines)
    return tokenizer

# calculate the length of the description with the most words
def max_length(descriptions):
    lines = to_lines(descriptions)
    return max(len(d.split()) for d in lines)

# map an integer to a word
def word_for_id(integer, tokenizer):
    for word, index in tokenizer.word_index.items():
        if index == integer:
            return word
```

```
        return None


# generate a description for an image
def generate_desc(model, tokenizer, photo, max_length):
        # seed the generation process
        in_text = 'startseq'
        # iterate over the whole length of the sequence
        for i in range(max_length):
                # integer encode input sequence
                sequence = tokenizer.texts_to_sequences([in_text])[0]
                # pad input
                sequence = pad_sequences([sequence], maxlen=max_length)
                # predict next word
                yhat = model.predict([photo,sequence], verbose=0)
                # convert probability to integer
                yhat = argmax(yhat)
                # map integer to word
                word = word_for_id(yhat, tokenizer)
                # stop if we cannot map the word
                if word is None:
                        break
                # append as input for generating the next word
                in_text += ' ' + word
```

```
            # stop if we predict the end of the sequence
            if word == 'endseq':
                break
    return in_text

# evaluate the skill of the model
def evaluate_model(model, descriptions, photos, tokenizer, max_length):
    actual, predicted = list(), list()
    # step over the whole set
    for key, desc_list in descriptions.items():
        # generate description
        yhat = generate_desc(model, tokenizer, photos[key], max_length)
        # store actual and predicted
        references = [d.split() for d in desc_list]
        actual.append(references)
        predicted.append(yhat.split())
    # calculate BLEU score
    print('BLEU-1: %f' % corpus_bleu(actual, predicted, weights=(1.0, 0, 0, 0)))
    print('BLEU-2: %f' % corpus_bleu(actual, predicted, weights=(0.5, 0.5, 0, 0)))
    print('BLEU-3: %f' % corpus_bleu(actual, predicted, weights=(0.3, 0.3, 0.3, 0)))
    print('BLEU-4: %f' % corpus_bleu(actual, predicted, weights=(0.25, 0.25, 0.25, 0.25)))

# load training dataset (6K)
```

```
filename = 'C:/Project_b4/Flickr_8k.trainImages.txt'

train = load_set(filename)

print('Dataset: %d' % len(train))

# descriptions

train_descriptions = load_clean_descriptions('descriptions.txt', train)

print('Descriptions: train=%d' % len(train_descriptions))

# prepare tokenizer

tokenizer = create_tokenizer(train_descriptions)

vocab_size = len(tokenizer.word_index) + 1

print('Vocabulary Size: %d' % vocab_size)

# determine the maximum sequence length

max_length = max_length(train_descriptions)

print('Description Length: %d' % max_length)


# prepare test set

# load test set

filename = 'C:/Project_b4/Flickr_8k.testImages.txt'

test = load_set(filename)

print('Dataset: %d' % len(test))

# descriptions

test_descriptions = load_clean_descriptions('descriptions.txt', test)

print('Descriptions: test=%d' % len(test_descriptions))

# photo features
```

```
test_features = load_photo_features('features.pkl', test)

print('Photos: test=%d' % len(test_features))


# load the model

filename = 'model_19.h5'

model = load_model(filename)

# evaluate model

evaluate_model(model, test_descriptions, test_features, tokenizer, max_length)
```

e. generate tokenizer

```
from keras.preprocessing.text import Tokenizer

from pickle import dump


# load doc into memory

def load_doc(filename):

        # open the file as read only

        file = open(filename, 'r')

        # read all text

        text = file.read()

        # close the file

        file.close()

        return text


# load a pre-defined list of photo identifiers

def load_set(filename):
```

```
	doc = load_doc(filename)
	dataset = list()
	# process line by line
	for line in doc.split('\n'):
		# skip empty lines
		if len(line) < 1:
			continue
		# get the image identifier
		identifier = line.split('.')[0]
		dataset.append(identifier)
	return set(dataset)

# load clean descriptions into memory
def load_clean_descriptions(filename, dataset):
	# load document
	doc = load_doc(filename)
	descriptions = dict()
	for line in doc.split('\n'):
		# split line by white space
		tokens = line.split()
		# split id from description
		image_id, image_desc = tokens[0], tokens[1:]
		# skip images not in the set
		if image_id in dataset:
			# create list
			if image_id not in descriptions:
				descriptions[image_id] = list()
```

```
                    # wrap description in tokens
                    desc = 'startseq ' + ' '.join(image_desc) + ' endseq'
                    # store
                    descriptions[image_id].append(desc)
    return descriptions

# covert a dictionary of clean descriptions to a list of descriptions
def to_lines(descriptions):
    all_desc = list()
    for key in descriptions.keys():
        [all_desc.append(d) for d in descriptions[key]]
    return all_desc

# fit a tokenizer given caption descriptions
def create_tokenizer(descriptions):
    lines = to_lines(descriptions)
    tokenizer = Tokenizer()
    tokenizer.fit_on_texts(lines)
    return tokenizer

# load training dataset (6K)
filename = 'C:/Project_b4/Flickr_8k.trainImages.txt'
train = load_set(filename)
print('Dataset: %d' % len(train))
# descriptions
train_descriptions = load_clean_descriptions('C:/Project_b4/descriptions.txt', train)
print('Descriptions: train=%d' % len(train_descriptions))
```

```
# prepare tokenizer
tokenizer = create_tokenizer(train_descriptions)
# save the tokenizer
dump(tokenizer, open('tokenizer.pkl', 'wb'))
```

f. Testing with GUI

```
from keras.preprocessing import image
from tkinter import *
from PIL import ImageTk, Image
from tkinter import filedialog
from sklearn.preprocessing import LabelBinarizer
from keras.optimizers import Adam
from keras.preprocessing.image import img_to_array
from pickle import load
from numpy import argmax
from keras.preprocessing.sequence import pad_sequences
from keras.applications.vgg16 import VGG16
from keras.preprocessing.image import load_img
from keras.applications.vgg16 import preprocess_input
from keras.models import Model
from keras.models import load_model
# extract features from each photo in the directory
def extract_features(filename):
        # load the model
        model = VGG16()
        # re-structure the model
        model = Model(inputs=model.inputs, outputs=model.layers[-2].output)
```

```
	# load the photo
	image = load_img(filename, target_size=(224, 224))
	# convert the image pixels to a numpy array
	image = img_to_array(image)
	# reshape data for the model
	image = image.reshape((1, image.shape[0], image.shape[1], image.shape[2]))
	# prepare the image for the VGG model
	image = preprocess_input(image)
	# get features
	feature = model.predict(image, verbose=0)
	return feature

# map an integer to a word
def word_for_id(integer, tokenizer):
	for word, index in tokenizer.word_index.items():
		if index == integer:
			return word
	return None

# generate a description for an image
def generate_desc(model, tokenizer, photo, max_length):
	# seed the generation process
	in_text = 'startseq'
	# iterate over the whole length of the sequence
	for i in range(max_length):
		# integer encode input sequence
		sequence = tokenizer.texts_to_sequences([in_text])[0]
```

```
            # pad input
            sequence = pad_sequences([sequence], maxlen=max_length)
            # predict next word
            yhat = model.predict([photo,sequence], verbose=0)
            # convert probability to integer
            yhat = argmax(yhat)
            # map integer to word
            word = word_for_id(yhat, tokenizer)
            # stop if we cannot map the word
            if word is None:
                  break
            # append as input for generating the next word
            in_text += ' ' + word
            # stop if we predict the end of the sequence
            if word == 'endseq':
                  break
      return in_text
width=256
height=256
root = Tk()
root.geometry("550x300+300+150")
root.resizable(width=True, height=True)
def openfn():
   filename = filedialog.askopenfilename(title='open')
   return filename
def open_img():
   x = openfn()
```

```
tokenizer = load(open('tokenizer.pkl', 'rb'))
    # pre-define the max sequence length (from training)
    max_length = 32
    # load the model
    model = load_model('C:/Project_b4/model_19.h5')
    # load and prepare the photograph
    photo = extract_features(x)
    # generate description
    description = generate_desc(model, tokenizer, photo, max_length)
    print(description)
    global label
    label.pack_forget()
    label = Label( root, text="Prediction : "+description)
    label.pack()
    img = Image.open(x)
    img = img.resize((250, 250), Image.ANTIALIAS)
    img = ImageTk.PhotoImage(img)
    global panel
    panel.pack_forget()
    panel = Label(root, image=img)
    panel.image = img
    panel.pack()
label = Label(root)
panel=Label(root)
btn = Button(root, text='Select Image',fg='white', bg='DarkSlateGray', command=open_img).pack()
root.configure(background='DodgerBlue')
root.mainloop()
```

CHAPTER 6

TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, subassemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

6.1. TYPES OF TESTS

6.1.1. Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expectedresults.

6.1.2. Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box. you cannot "see" into it. The test provides inputs and responds to outputs without considering how the software work

6.1.3. User Acceptance Testing

User Acceptance of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changes wherever required. The system developed provides a friendly user interface that can easily be understood even by a person who is new to the system.

6.1.4. OutputTesting

After performing the validation testing, the next step is output testing of the proposed system, since no system could be useful if it does not produce the required output in the specified format. Asking the users about the format required by them tests the outputs generated or displayed by the system under consideration. Hence the output format is considered in 2 ways – one is on screen and another in printed format.

.

6.2. TEST CASES

6.2.1. TEST_CASE_ID_01

Test Case Description	To check if the image is taken as input
Input to the model	Image is given as input
Expected Output	Generates text descriptions along with input image
Actual Output	Generates text descriptions along with input image
Pass/Fail	PASS

6.2.2 TEST_CASE_ID_02

Test Case Description	To check if anything other than image is given
Input to the model	Text file is given as input
Expected Output	Displays an error
Actual Output	Displays an error
Pass/Fail	PASS

CHAPTER 7

EXPERIMENTAL RESULTS

7.1. PHOTO FEATURE EXTRACTION

VGG16 LAYERS

Model: "model_1"

Layer (type)	Output Shape	Param #
input_2 (InputLayer)	[(None, 224, 224, 3)]	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
fc1 (Dense)	(None, 4096)	102764544
fc2 (Dense)	(None, 4096)	16781312

Total params: 134,260,544
Trainable params: 134,260,544
Non-trainable params: 0

7.2. FEATURES EXTRACTED

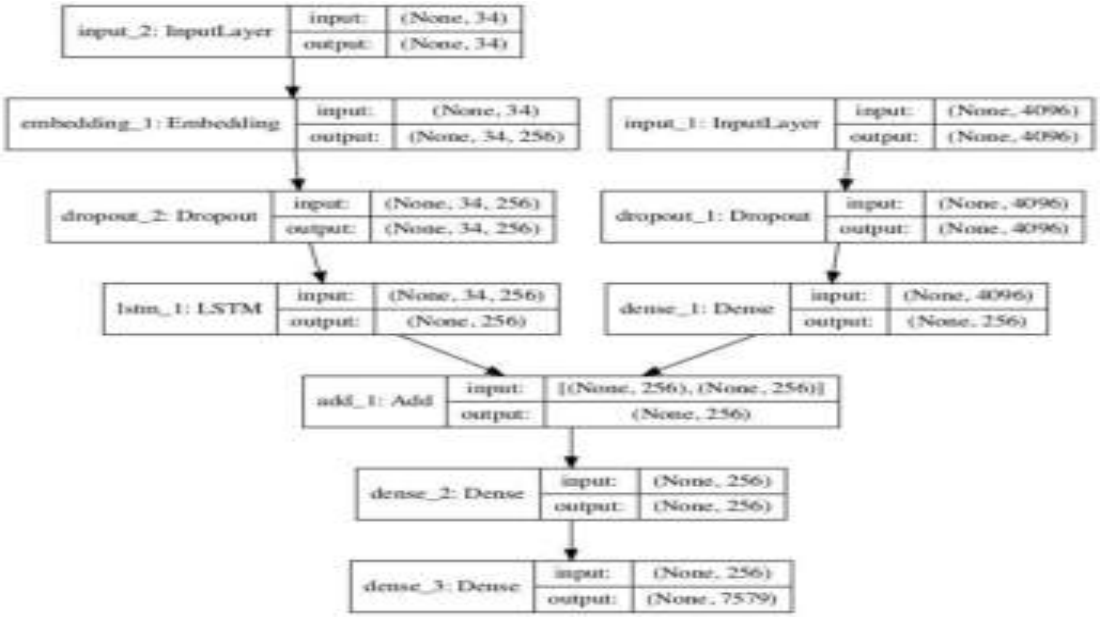
```
>976392326_082dafc3c5.jpg
>97731718_eb7ba71fd3.jpg
>977856234_0d9caee7b2.jpg
>978580450_e862715aba.jpg
>979201222_75b6456d34.jpg
>979383193_0a542a059d.jpg
>98377566_e4674d1ebd.jpg
>985067019_705fe4a4cc.jpg
>987907964_5a06a63609.jpg
>989754491_7e53fb4586.jpg
>989851184_9ef368e520.jpg
>990890291_afc72be141.jpg
>99171998_7cc800ceef.jpg
>99679241_adc853a5c0.jpg
>997338199_7343367d7f.jpg
>997722733_0cb5439472.jpg
Extracted Features: 8091
```

7.2.3 PREPROCESSING TEXT DATA

```
Loaded: 8092
Vocabulary Size: 8763
```

7.2.4. DEVELOPING DEEP LEARNING MODEL – LOAD, DEFINE, TRAIN

ARCHITECTURE OF THE MODEL



TRAINING MODEL

```
Dataset: 6000
Descriptions: train=6000
Photos: train=6000
Vocabulary Size: 10400
Description Length: 32
Model: "model_3"
```

Layer (type)	Output Shape	Param #	Connected to
input_7 (InputLayer)	[(None, 32)]	0	
input_6 (InputLayer)	[(None, 4096)]	0	
embedding_2 (Embedding)	(None, 32, 256)	2662400	input_7[0][0]
dropout_4 (Dropout)	(None, 4096)	0	input_6[0][0]
dropout_5 (Dropout)	(None, 32, 256)	0	embedding_2[0][0]
dense_6 (Dense)	(None, 256)	1048832	dropout_4[0][0]
lstm_2 (LSTM)	(None, 256)	525312	dropout_5[0][0]
add_2 (Add)	(None, 256)	0	dense_6[0][0] lstm_2[0][0]
dense_7 (Dense)	(None, 256)	65792	add_2[0][0]
dense_8 (Dense)	(None, 10400)	2672800	dense_7[0][0]

```
Total params: 6,975,136
Trainable params: 6,975,136
Non-trainable params: 0
```

```
6000/6000 [==============================] - 3244s 539ms/step - loss: 5.9942
6000/6000 [==============================] - 3000s 500ms/step - loss: 4.4698
6000/6000 [==============================] - 3097s 516ms/step - loss: 4.1241
6000/6000 [==============================] - 2899s 483ms/step - loss: 3.9463
6000/6000 [==============================] - 3045s 508ms/step - loss: 3.7835
6000/6000 [==============================] - 3098s 516ms/step - loss: 3.6809
6000/6000 [==============================] - 3064s 511ms/step - loss: 3.6065
6000/6000 [==============================] - 2985s 497ms/step - loss: 3.5456
6000/6000 [==============================] - 2981s 497ms/step - loss: 3.4953
6000/6000 [==============================] - 2943s 491ms/step - loss: 3.4550
6000/6000 [==============================] - 3301s 550ms/step - loss: 3.4166
6000/6000 [==============================] - 3341s 557ms/step - loss: 3.3787
6000/6000 [==============================] - 3506s 584ms/step - loss: 3.3510
6000/6000 [==============================] - 3232s 539ms/step - loss: 3.3313
6000/6000 [==============================] - 2853s 475ms/step - loss: 3.3024
6000/6000 [==============================] - 2977s 496ms/step - loss: 3.2862
6000/6000 [==============================] - 2974s 496ms/step - loss: 3.2681
6000/6000 [==============================] - 3018s 503ms/step - loss: 3.2585
6000/6000 [==============================] - 2982s 497ms/step - loss: 3.2413
6000/6000 [==============================] - 2964s 494ms/step - loss: 3.2321
```

7.2.4. EVALUATION OF MODEL

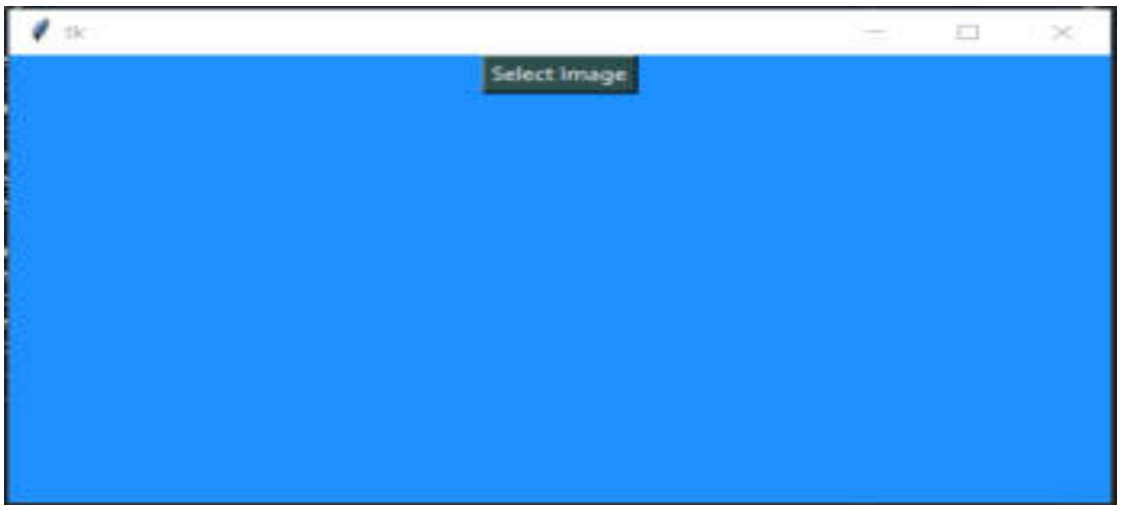
```
Dataset: 6000
Descriptions: train=6000
Vocabulary Size: 10400
Description Length: 32
Dataset: 1000
Descriptions: test=1000
Photos: test=1000
BLEU-1: 0.128364
BLEU-2: 0.063430
BLEU-3: 0.041126
BLEU-4: 0.012124
```

7.2.5 GENERATE TOKENIZER

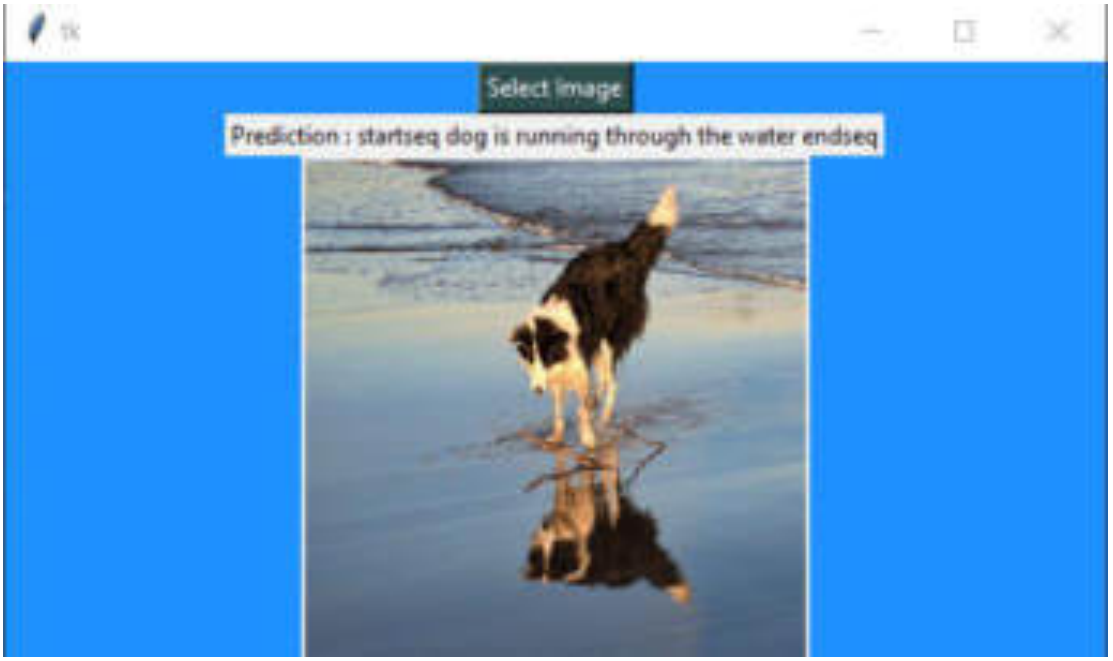
```
Dataset: 6000
Descriptions: train=6000
```

7.2.6. TESTING WITH GUI

INPUT IMAGE



OUTPUT



CHAPTER 8

APPLICATIONS

1. Recommendations in Editing Applications

The image captioning model automates and accelerates the close captioning process for digital content production, editing, delivery, and archival. Well-trained models replace manual efforts for generating quality captions for images as well as videos.

2. Assistance for Visually Impaired

The advent of ML solutions like image captioning is a boon for visually impaired people who are unable to comprehend visuals. With AI-powered image caption generator, image descriptions can be read out to visually impaired, enabling them to get a better sense of their surroundings.

3. Media and Publishing Houses

The media and public relations industry circulate tens of thousands of visual data across borders in the form of newsletters, emails, etc. The image captioning model accelerates subtitle creation and enables executives to focus on more important tasks.

4. Social Media Posts

For social media, artificial intelligence is moving from discussion rooms to underlying mechanisms for identifying and describing terabytes of media files. It enables community administrators to monitor interactions and analysts to formulate business strategies.

CHAPTER 9

CONCLUSION

Image captioning has made significant advances in recent years. Recent work based on deep learning techniques has resulted in a breakthrough in the accuracy of image captioning. The text description of the image can improve the content-based image retrieval efficiency, the expanding application scope of visual understanding in the fields of medicine, security, military and other fields, which has a broad application prospect. At the same time, the theoretical framework and research methods of image captioning can promote the development of the theory and application of image annotation and visual question answering (VQA), cross media retrieval, video captioning and video dialog, which has important academic and practical application value

CHAPTER 10

FUTURE ENHANCEMENTS

A lot of Enhancements can be done on this model to improve the accuracy and to identify images and the scenarios in order to generate better results.

- This model is built on Flickr-8k Dataset, which consists of only 8096 images, on which the training is done. This can be further improved by using either Flickr-30k Dataset or MS COCO-300k Dataset.

- This model is trained in a space of 8GB RAM, and can be improved by training this model in a larger space to improve the scores achieved by the BLEU scores.

- Changing the model architecture for generating the captions and more hyper parameter tuning can be done.

REFERENCES

[1]. Murk Chohan, Adil Khan, Muhammad Saleem Mahar, Saif Hassan, Abdul Ghafoor and Mehmood Khan, “Image Captioning using Deep Learning: A Systematic Literature Review” International Journal of Advanced Computer Science and Applications(IJACSA), 11(5), 2020. http://dx.doi.org/10.14569/IJACSA.2020.0110537

[2]. K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, “Bleu: a method forautomatic evaluation of machine translation,” in Proceedings of the 40thannual meeting on association for computational linguistics. Association for Computational Linguistics, 2002, Conference Proceedings, pp.311–318.

[3]. Image Captioning Based on Deep Neural NetworksShuang Liu, Liang Bai, Yanli Hu and Haoran Wang, MATEC Web Conf., 232 (2018) 01052

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"CUSTOMER SEGMENTATION USING MACHINE LEARNING"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
COMPUTER SCIENCE AND ENGINEERING

Submitted by

Ch. Sri Lakshmi (17H71A05B1) **P.Siva Krishna (17H71A05A9)**

G.Yamini (17H71A05B9) **B.Swetha (17H71A05B3)**

Under the Esteemed Guidance of

Ms.Mareshwaramma

(Assistant Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273621, Fax: 08678 - 273569
E mail: dvrhmic@mictech.ac.in, Website: www.mictech.ac.in

2020-21

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE)(ECE)(ME)
Accredited by
NAAC A
With Grade

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"Customer Segmentation Using Machine Learning"** is a bonafide work carried out by **Ch.SriLakshmi(17H71A05B1) P.SivaKrishna(17H71A05A9),G.Yamini(17H71A05B9),B.Swetha(17H71A05B3)** in partial fulfilment for the award of degree of Bachelor of Technology in **Computer Science and Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

Ms.Mareshwaramma
Project Guide

D. Prasad
Head of the Department

Examiner

Dr K Srinivas
PRINCIPAL

ACKNOWLEDGEMENT

We would like to take this opportunity to express my deepest appreciation to the following people for their valuable contributions and assistance with this Project.

Initially, we would like to thank our project supervisor, **Ms.Mareshwaramma, Assistant Professor, Department of Computer Science and Engineering** for the guidance and support, especially for the valuable ideas and knowledge shared to me throughout the Project.

We have the immense pleasure in expressing our thanks and deep sense of gratitude to**, Mr. D. Prasad, Head of the Department, Computer Science and Engineering** for extending necessary facilities and support for the completion of the Project.

We whole sincerely acknowledge **Dr. K. Srinivas, Principal and Prof D. PANDURANGA RAO, CEO** for giving opportunity to take up the Project work. We also extend our thanks to all faculty members of **Computer Science and Engineering**, for their valuable guidance and encouragement.

We would like to extend our warm appreciation to all my friends for sharing us their knowledge, valuable contributions and help with this Project.

Finally, our special thanks go to our family for their continuous support and help throughout our academic years and for their continual support and encouragement for the completion of the project.

DECLARATION

We **Ch.Sri Lakshmi (17H71A05B1) , SivaKrishna (17H71A05A9), G.Yamini (17H71A05B9), B.Swetha(17H71A05B3)** of the Main-Project "**Customer Segmentation Using Machine Learning**", hereby declare that the matter embodied in this Project is the genuine work done by us and has not been submitted either to this university or to any other university/institute for the fulfilment of the requirement of any course of study.

Ch Sri Lakshmi (17H71A05B1)

P Siva Krishna (17H71A05A9)

G Yamini (17H71A05B9)

B.Swetha(17H71A05B3)

CONTENTS

ABSTRACT

The project is aimed at developing application for “Customer Segmentation Using Machine Learning”. Customer segmentation and pattern extraction is one of the key aspects of business decision support system. In this emerging market it is very difficult to maintain customer base. To overcome this difficulty every business has to focus on customer segmentation.

Customer Segmentation is one of the most important applications of unsupervised learning. Customer Segmentation is the process of division of customer base into several groups of individuals that share a similarity in different ways that are relevant to marketing such as gender, age, interests, and miscellaneous spending habits.

In this machine learning project, we will make use of K-means clustering which is the essential algorithm for clustering unlabeled dataset. Companies that deploy customer segmentation are under the notion that every customer has different requirements and require a specific marketing effort to address them appropriately. However if you need to know your best customer, this is the best methodology.

Keywords: K-means, Machine Learning, unsupervised data

List of Figures

Chapter 1

INTRODUCTION

1.1 Introduction

Customer segmentation is the method of dividing customers into groups or clusters on the common characteristics. The most common way in which business segment their customer data such as gender,age,income expenditure. Customer segmentation plays a crucial role for supporting business decisions.

Due to increase in customers it has become so complicated to maintain a customer base. To overcome such complexity business has to focus on customer segmentation.

In Machine Leaning, algorithms are of two types supervised and unsuper- vised machine learning algorithms. Customer segmentation using machine learning uses k-means clustering algorithm which comes under the unsupervised machine learning in which unlabeled data is classified into clusters.

In this project the Mall customers dataset is used on which clusters will be done based on the Annual Income and spending score of the customers. The Mall customers dataset contains features such as Custid,age,Annual income, spending score and so on. Based on this dataset, algorithm will build a model which is used to make clusters or segments.

1.2 Motivation

Customer segmentation is the practice of dividing a company's customers into groups that reflect similarity among customers in each group. The goal of segmenting customers is to decide how to relate to customers in each segment in order to maximize the value of each customer to the business.

1.3 Purpose

Customer Segmentation is the process of division of customer base into several groups of individuals that share a similarity in different ways that are relevant to marketing such as gender, age, interests, and miscellaneous spending habits.

Companies that deploy customer segmentation are under the notion that every customer has different requirements and require a specific marketing effort to address them appropriately. Companies aim to gain a deeper approach of the customer they are targeting. Therefore, their aim has to be specific and should be tailored to address the requirements of each and every individual customer. Furthermore, through the data collected, companies can gain a deeper understanding of customer preferences as well as the requirements for discovering valuable segments that would reap them maximum profit. This way, they can strategize their marketing techniques more efficiently and minimize the possibility of risk to their investment.

The technique of customer segmentation is dependent on several key differentiators that divide customers into groups to be targeted. Data related to demographics, geography, economic status as well as behavioral addressing the various segments.

1.4 Problem Statement

Customer Segmentation is one the most important applications of unsupervised learning. Using clustering techniques, companies can identify the several segments of customers allowing them to target the potential user base. In this machine learning project, we will make use of k-means clustering which is the essential algorithm for clustering unlabeled dataset.

In this project the Mall Customers dataset is used on which clusters will be done based on the Annual Income and Spending score of the customers. Based on the Mall Customers dataset, algorithm will build a model which is used to make clusters or segments.

1.5 Background

Customer segmentation has the potential to allow marketers to address each customer in the most effective way. Using the large amount of data available on customers (and potential customers), a customer segmentation analysis allows marketers to identify discrete groups of customers with a high degree of accuracy based on demographic, behavioral and other indicators.

Since the marketer's goal is usually to maximize the value (revenue and/or profit) from each customer, it is critical to know in advance how any particular marketing action will influence the customer. Ideally, such "action-centric" customer segmentation will not focus on the short-term value of a marketing action, but rather the long-term Customer Lifetime Value(CLV) impact that

such a marketing action will have. Thus, it is necessary to group, or segment, customers according to their CLV.

1.6 Scope of the Project

Every segment has a specific behavior. Business decision making will vary from behavior to behavior. Because of this algorithm consistent and effective business strategies will improve business of an organization.

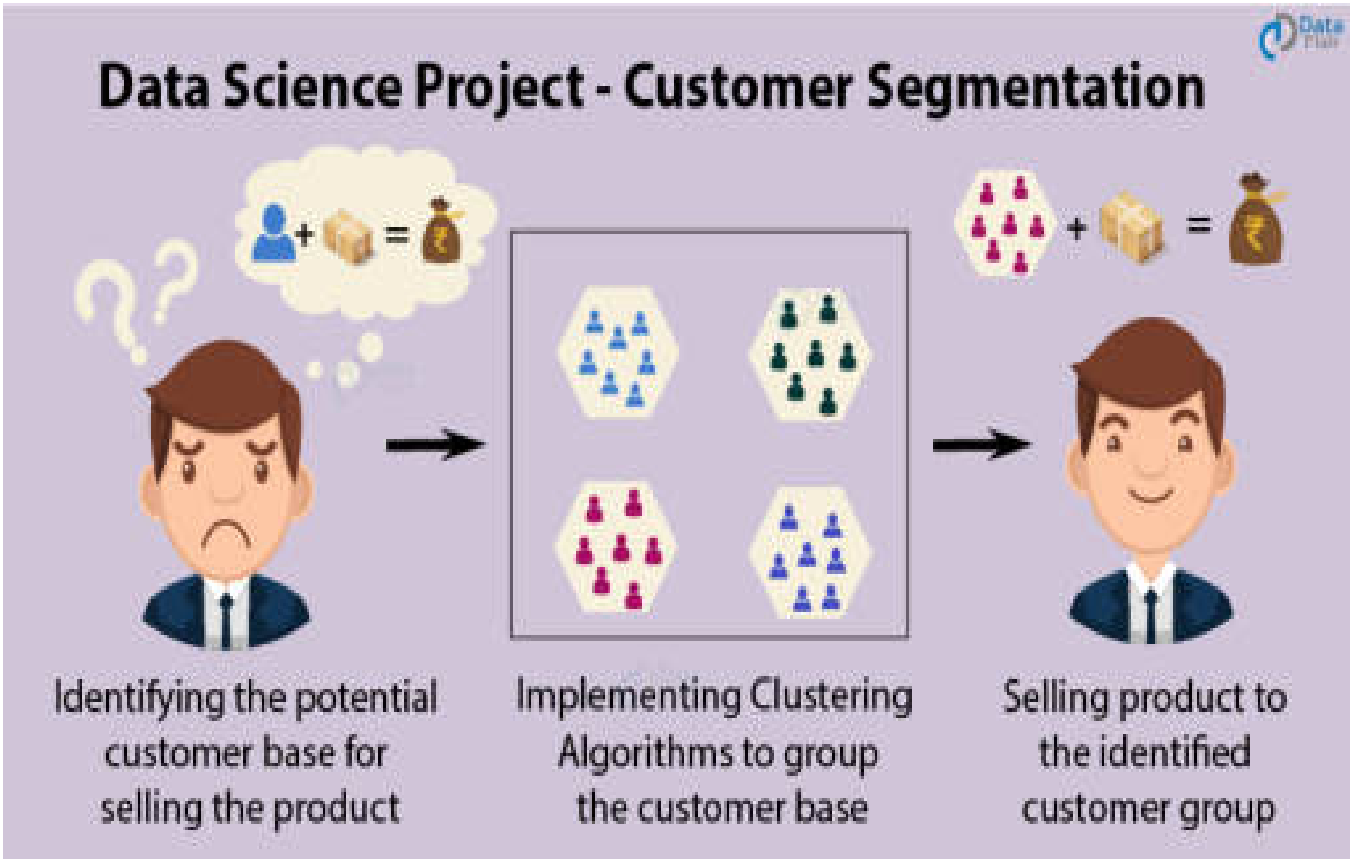


Figure 1.1: **Need of clustering**

Chapter 2

LITERATURE SURVEY

Raquel Florez-Lopez et al.,[1] describes the importance of customer relationship management (CRM) has increased in previous years due to a competitive environment forcing the company to implement a dynamic management of clients to obtain higher profits and to acquire a higher market share than its competitors. One of the most important CRM decisions for retailers is the design of efficient direct marketing politics, especially, if a new product or service is launched into the market. The development of a CRM program needs to fulfill several successive stages, customer segmentation and the measurement of the success of the CRM program being two of the most critical steps.

Market segmentation has traditionally developed through statistical techniques such as PCA, DA, or logistical regression .These models consider restrictive a priority hypothesis that are not fulfilled in presence of high volumes of data and multiple-correlated features, thus feature selection methods and alternative clusterization techniques must be considered to obtain robust segmentation. In this article, the author proposed to combine statistical and machine learning techniques for market segmentation; in particular, the performance of decision trees (both univariety and oblique variants) is analyzed. Although several metrics have been developed to measure the success of the CRM program, they are partial solutions that must be integrated in an overall measure.

Mahboubeh Khajvand et al.,[2] describes customer segmentation is one of the customer lifetime value(CLV) applications which focused on in current study. The author clustered customers into segments ac- cording to RFM and Extended Recency , Frequency, Monetary(RFM) parameters using K-means Algorithm. Clustering customers into different groups helps decision-makers identify market segments more clearly and thus develop more effective marketing and sale strategies for customer retention. Since the RFM weights differ with the characteristics of industry, Analytical Hierarchy Process (AHP) method was applied to know the relative importance of RFM variables based on expert point of view . Then CLV rank was assigned to each segment on the basis of the CLV value. The current value provides financial viewpoint and potential value indicates cross-selling opportunities.

K.Maheswari [3] describes that online shopping does not allow the customers to touch the product. They have to attract the customers by showing images, photos and other displays. Data mining is a most powerful tool to discover knowledge from the database. In this work, the data set is normalized to produce improved results. Minimum was calculated using scale () function in R

software. Clustering was applied on normalized data set. In this paper, choosing of right number of clusters was performed. Future work concentrates more advanced algorithms to perform this work.

Sahar F.Sabbe [4] describes that this study tries to present a bench mark for the most widely used state of the arts for churn classification. The accuracy of the models was taken on a public dataset of customers in Telecom company. Based on the findings of this study, learning techniques are recommended as both Random forest and Ad boost model have given the best accuracy. However, the study can be extended by including hybrid models and deep learning models. The other performance models can be used for performance evaluation. Models can also develop against different dataset from different domains.

J.du Toit et al., [5] describes the Silhouette evaluation suggests that using correlation as a distance metric gives the simplest result. The Euclidean distance metric was thought of as the slight outperformance was not enough to warrant the added procedure complexness. It was shown that the calculated range of clusters for each algorithm provided smart theoretical results. However, extra segmentation info was extracted by choosing two extra clusters, as urged by domain expertise. The calculated clusters corresponded well with the expected customer or load teams. Depending on the location and type of customer in the area, the cluster assignments can be confirmed.

According to Charles W. Lamb and Carl McDaniel (2003,), the first step in segmenting markets is to “select a market or product category for study”. It may be a market in which the firm has already occupied a new but related market or product category, or a totally new one. The second step is to “choose a basis or bases for segmenting the market”. This step requires managerial insight, creativity and market knowledge. There are no scientific procedures for selecting segmentation variables. However, a successful segmentation plan must produce market segments which meet the four basic criteria: “substantiality, identifiably, accessibility, and responsiveness”.

The third step is “selecting segmentation descriptors”. After choosing one or more bases, the marketer must select the segmentation descriptors. Descriptors identify the specific segmentation variables to use. The fourth one is to “profile and analyze segments”. The analysis should include the segment’s size, expected growth, purchase frequency, current brand usage, brand loyalty, and long-term sales and profit potential.

This information can then be used to rank potential market segments by profit opportunity, risk, consistency with organizational task and objectives, and other factors which are important to the company. The fifth step is to “select target markets”. This step is not a part of the segmentation process but a natural result of it. It is a major decision that affects and often directly determines the firm’s marketing mix. The last one is “designing, implementing and maintaining appropriate

marketing mixes". The marketing mix has been described as product, distribution, promotion and price strategies which are used to bring about mutually satisfying relationships with target markets. Roger Best (1990) proposes a framework for implementing a market segmentation strategy. He suggests a set of sequential steps to be taken in a needs-based segmentation process the primary benefit of needsbased segmentation is that segments are created around specific customer needs.

The goal is to determine what observable demographics and behaviors differentiate one segment from another in order to make need-based market segmentation actionable.

Chapter 3

SYSTEM ANALYSIS

3.1 Existing System

Segmentation is done using rule-based model in which they used only one attribute to classify the customer base into clusters. Every rule should be defined and changed according to their purpose manually that consumes more time.

Rule Based segmentation examines user behaviour and segments user groups based on a set of rules. The specific 'rules' change from one school of thought to another, but for the most part they're a mix between common sense and past experience, meant to predict intent. For example, in e-commerce, users who abandoned a cart are a very common rule based segment.

Disadvantages:

- Lot of manual work: The RB system demands deep knowledge of the domain as well as a lot of manual work.
- Time consuming: Generating rules for a complex system is quite challenging and time consuming.
- Less learning capacity: Here, the system will generate the result as per the rules so the learning capacity of the system by itself is much less.
- Complex domains: If an application that you want to build is too complex, building the RB system can take lot of time and analysis. Complex pattern identification is a challenging task in the RB approach.

.

3.2 Proposed System

In this project we are using the popular K-Means clustering algorithm to segment customers. This technique is relatively easy to implement and yet it allowed us to gather tons of informations from my data and unearth interesting patterns of behavior in the customer base.

K-Means clustering is part of the Unsupervised Learning modeling family, a set of techniques used to find patterns in data that has not been labeled, classified or categorized. As this method does not require to have a targets for clustering, it can do good job in the exploratory phase of customer segmentation.

Each cluster is represented by its center, corresponding to the mean of elements assigned to the cluster. In this project we use the Mall Customers dataset and Elbow method is used to select the k-value. Within cluster sum of squares values is calculated using the standard values in the elbow method.

Advantages:

- Improves channel of distribution
- More customer retention
- Enhances competitiveness
- Cost is considerably less
- We can form clusters for Large data sets
- Easy to implement when compared to the previous rule based model
- Manual work is eliminated
- We will more accurate values when compared to the previous method

3.3 Project Domain

3.3.1 Machine learning

Machine Learning (ML) is the Field of study that gives computers the capability to learn without being explicitly programmed**. It** is the logical investigation of calculations and measurable models that PC frameworks use to play out a particular undertaking without utilizing un-equivocal guidelines, depending on examples and induction. It is viewed as a sub set of man-made brain-power. Machine learning calculations manufacture a scientific model dependent on test data, known as ”training data”, so as to settle on expectations or choices without being unequivocally modified to play out the undertaking. Machine learning calculations are utilized in a wide way of utilizations, for example, email shifting, where it is trouble- some or infeasible to build up a customary calculation for adequately playing out the undertaking. In this we have Supervise and Un-supervised Learning Models.

In supervised learning, the algorithm “learns” from the training dataset by iteratively making predictions on the data and adjusting for the correct answer. While supervised learning models tend to be more accurate than unsupervised learning models, they require upfront human intervention to label the data appropriately. For example, a supervised learning model can predict how long your commute will be based on the time of day, weather conditions and so on. But first, you’ll have to train it to know that rainy weather extends the driving time.

Unsupervised learning models, in contrast, work on their own to discover the inherent structure of unlabeled data.Note that they still require some human intervention for validating output variables. For example, an unsupervised learning model can identify that online shoppers often purchase groups of products at the same time.

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, uncovering key insights within data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase, requiring them to assist in the identification of the most relevant business questions and subsequently the data to answer them

The learning system of a machine learning algorithm has three main parts.

1. **A Decision Process**: In general, machine learning algorithms are used to make a prediction or classification. Based on some input data, which can be labelled or unlabeled, your algorithm will produce an estimate about a pattern in the data.
2. **An Error Function**: An error function serves to evaluate the prediction of the model. If there are known examples, an error function can make a comparison to assess the accuracy of the model.
3. **An Model Optimization Process**: If the model can fit better to the data points in the training set, then weights are adjusted to reduce the discrepancy between the known example and the model estimate. The algorithm will repeat this evaluate and optimize process, updating weights autonomously until a threshold of accuracy has been met.

Few examples of machine learning we might encounter every day:

Speech Recognition: It is also known as automatic speech recognition (ASR), computer speech recognition, or speech-to-text, and it is a capability which uses natural language processing (NLP) to process human speech into a written format. Many mobile devices incorporate speech recognition into their systems to conduct voice search—e.g. Siri—or provide more accessibility around texting.

Customer Service: Online chatbots are replacing human agents along the customer journey. They answer frequently asked questions (FAQs) around topics, like shipping, or provide personalized advice, cross-selling products or suggesting sizes for users, changing the way we think about customer engagement across websites and social media platforms. Examples include messaging bots on e-commerce sites with virtual agents, messaging apps, such as Slack and Facebook Messenger, and tasks usually done by virtual assistants and voice assistants.

Computer Vision: This AI technology enables computers and systems to derive meaningful information from digital images, videos and other visual inputs, and based on those inputs, it can take action. This ability to provide recommendations distinguishes it from image recognition tasks. Powered by convolutional neural networks, computer vision has applications within photo tagging in social media, radiology imaging in healthcare, and self-driving cars within the automotive industry.

Recommendation Engines: Using past consumption behavior data, AI algorithms can help to discover data trends that can be used to develop more effective cross-selling strategies. This is used to make relevant add-on recommendations to customers during the checkout process for online retailers.

Automated stock trading: Designed to optimize stock portfolios, AI-driven high-frequency trading platforms make thousands or even millions of trades per day without human intervention

3.3.2 Clustering

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. In simple words, the aim is to segregate groups with similar traits and assign them into clusters.

Cluster analysis, or clustering, is an unsupervised machine learning task. It involves automatically discovering natural grouping in data. Unlike supervised learning (like predictive modeling), clustering algorithms only interpret the input data and find natural groups or clusters in feature space.

Clustering techniques apply when there is no class to be predicted but rather when the instances are to be divided into natural groups. A cluster is often an area of density in the feature space where examples from the domain (observations or rows of data) are closer to the cluster than other clusters. The cluster may have a center (the centroid) that is a sample or a point feature space and may have a boundary or extent. Clustering can be helpful as a data analysis activity in order to learn more about the problem domain, so-called pattern discovery or knowledge discovery.

For example:

The phylogenetic tree could be considered the result of a manual clustering analysis.
Separating normal data from outliers or anomalies may be considered a clustering problem.
Separating clusters based on their natural behavior is a clustering problem, referred to as market segmentation.

Clustering can also be useful as a type of feature engineering, where existing and new examples can be mapped and labeled as belonging to one of the identified clusters in the data. Evaluation of identified clusters is subjective and may require a domain expert, although many clustering-specific quantitative measures do exist. Typically, clustering algorithms are compared academically on synthetic datasets with pre-defined clusters, which an algorithm is expected to discover.

Clustering can be helpful as a data analysis activity in order to learn more about the problem domain, so-called pattern discovery or knowledge discovery.

For example:

The phylogenetic tree could be considered the result of a manual clustering analysis.
Separating normal data from outliers or anomalies may be considered a clustering problem.
Separating clusters based on their natural behavior is a clustering problem, referred to as market segmentation.

Clustering can also be useful as a type of feature engineering, where existing and new examples can be mapped and labeled as belonging to one of the identified clusters in the data. Evaluation of identified clusters is subjective and may require a domain expert, although many clustering-specific quantitative measures do exist. Typically, clustering algorithms are compared academically on synthetic datasets with pre-defined clusters, which an algorithm is expected to discover.

The scikit-learn library provides a suite of different clustering algorithms to choose from. A list of 10 of the more popular algorithms is as follows:

- Affinity Propagation
- Agglomerative Clustering
- BIRCH
- DBSCAN
- K-Means
- Mini-Batch K-Means
- Mean Shift
- OPTICS
- Spectral Clustering
- Mixture of Gaussians

Each algorithm offers a different approach to the challenge of discovering natural groups in data. There is no best clustering algorithm, and no easy way to find the best algorithm for your data without using controlled experiments.

Figure3.1: **process of segmenting**

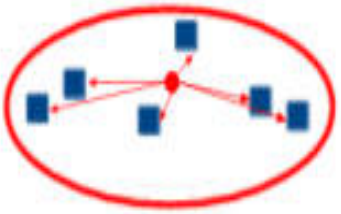
K-Means Clustering

K-means algorithm is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance betweenthe data points and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster.

K-means algorithm explores for a preplanned number of clusters in an unlabelled multidimensional dataset, it concludes this via an easy interpretation of how an optimized cluster can be expressed. You can take the centre as a data point that outlines the means of the cluster, also it might not possibly be a member of the dataset. In simple terms, k-means clustering enables us to cluster the data into several groups by detecting the distinct categories of groups in the unlabelled datasets by itself, even without the necessity of training of data.

It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabeled dataset on its own without the need for any training. It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters.

It tells us how far the points within a cluster are. So, inertia actually calculates the sum of distances of all the points within a cluster from the centroid of that cluster**.** We calculate this for all the clusters and the final inertial value is the sum of all these distances. This distance within the clusters is known as intracluster distance. So, inertia gives us the sum of intracluster distances.



Intra cluster distance

Now, For the value of inertia for a good cluster, We want the points within the same cluster to be similar to each other. Hence, the distance between them should be as low as possible**.**

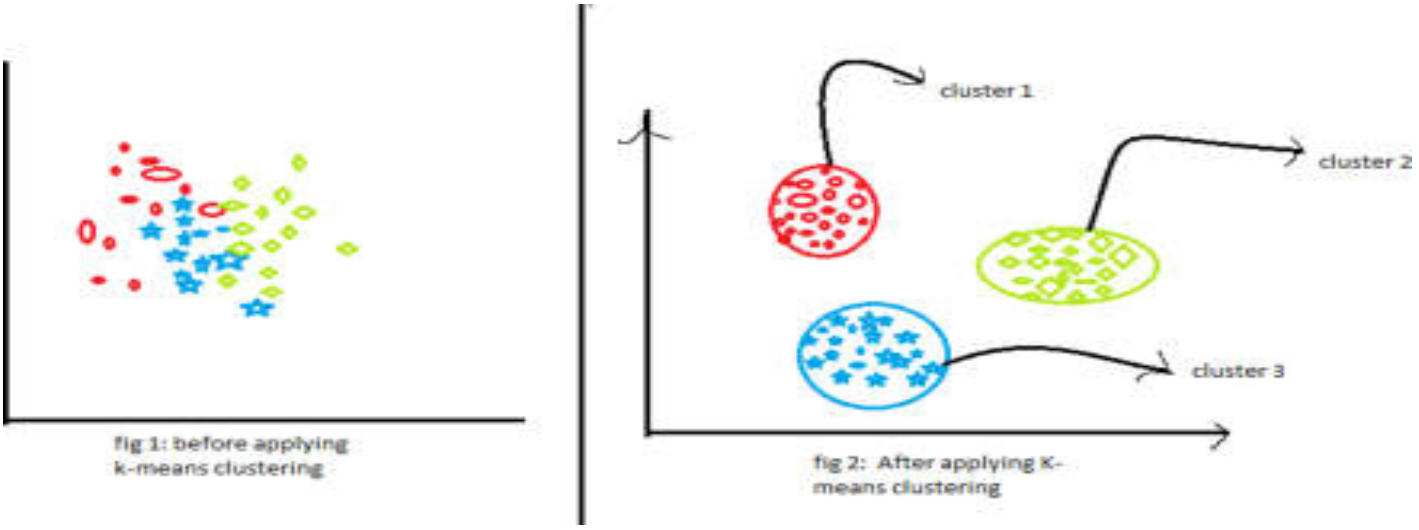


Figure 3.2: **k-means clustering process**

Diagrammatic Implementation of K Means Clustering

STEP 1: Let's choose number k of clusters, i.e., K=2, to segregate the dataset and to put them into different respective clusters. We will choose some random 2 points which will act as centroid to form the cluster.

STEP 2: Now we will assign each data point to a scatter plot based on its distance from the closest K-point or centroid. It will be done by drawing a median between both the centroids. Consider the below image:

STEP 3: points left side of the line is near to blue centroid, and points to the right of the line are close to the yellow centroid. The left one Form cluster with blue centroid and the right one with the yellow centroid.

STEP 4:repeat the process by choosing a new centroid. To choose the new centroids, we will find the new center of gravity of these centroids, which is depicted below :

STEP 5: Next, we will reassign each data point to the new centroid. We will repeat the same process as above (using a median line). The yellow data point on the blue side of the median line will be included in the blue cluster.

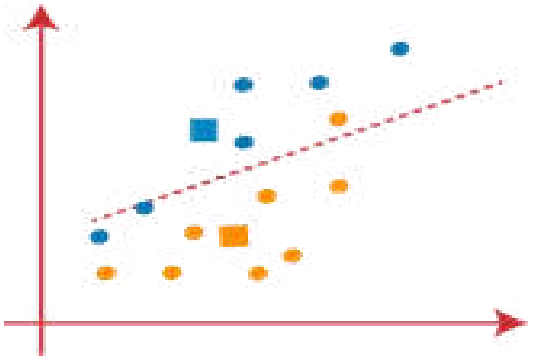


Figure 3.3: **clustering step 1**

STEP 6: As reassignment has taken place, so we will repeat the above step of finding new centroids.

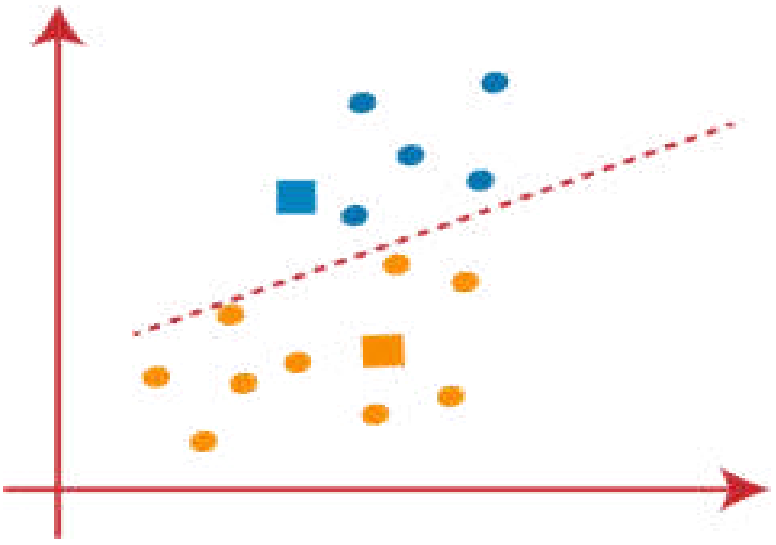


Figure 3.4: Clustering step-2

STEP 7: We will repeat the above process of finding the center of gravity of centroids, as being depicted below

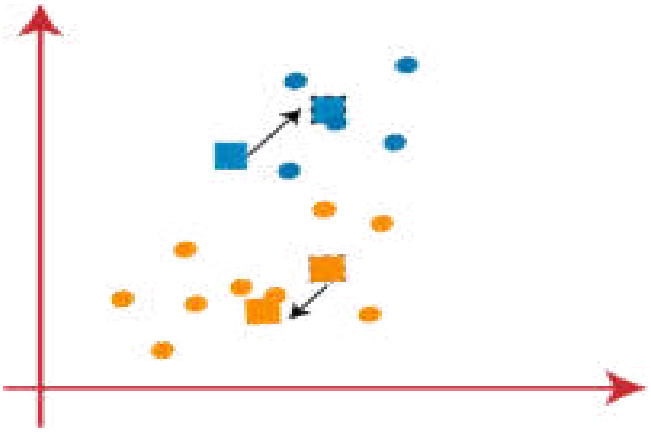


Figure 3.5: Clustering step-3

STEP 8: After Finding the new centroids we will again draw the median line and reassign the data points, like the above steps.

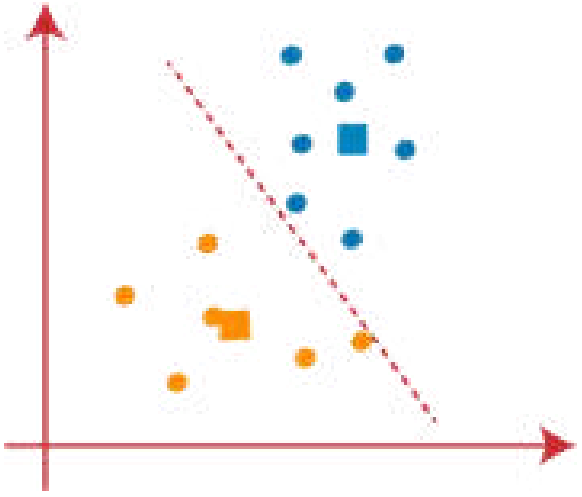


Figure 3.6: Clustering step-4

STEP 9: We will finally segregate points based on the median line, such that two groups are being formed and no dissimilar point to be included in a single group.

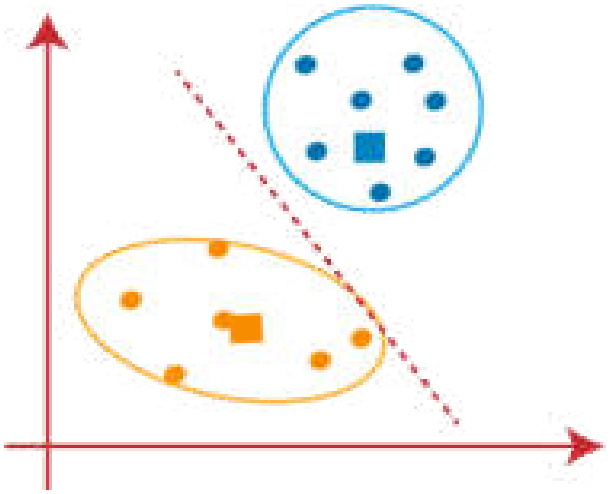


Figure 3.7: Clustering step-5

The final Cluster being formed are as Follows:

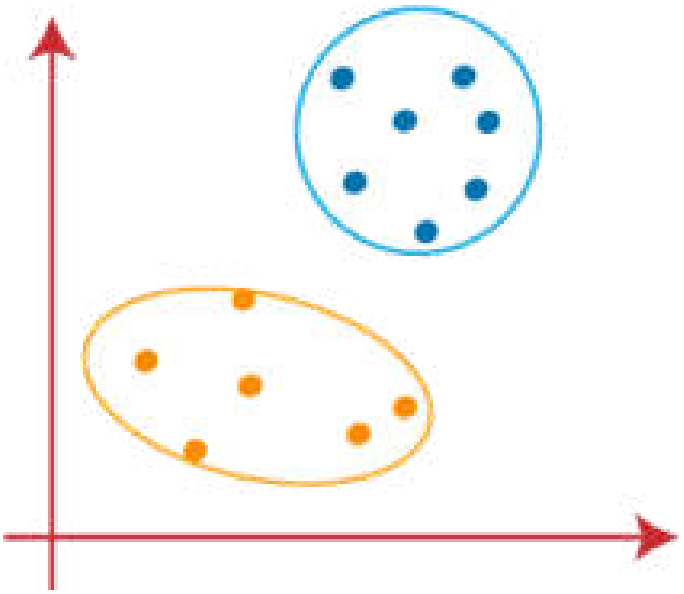


Figure 3.8: Clustering step-6

CLV-Focused Customer Segmentation

Of course, it is always easier to make assumptions and use “gut feelings” to define rules which will segment customers into logical groupings, e.g., customers who came from a particular source, who live in a particular location or who bought a particular product/service. However, these high-level categorizations will seldom lead to the desired results**.**

It is obvious that some customers will spend more than others during their relationship with a company. The best customers will spend a lot for many years. Good customers will spend modestly over a long period of time, or will spend a lot over a short period of time. Others won’t spend too much and/or won’t stick around too long**.**

The right approach to segmentation analysis is to segment customers to groups based on predictions regarding their total future value to the company, with the goal of addressing each group (or individual) in the way most likely to maximize that future, or lifetime, value.

Customer Segmentation Models

Accurate customer segmentation involves tracking dynamic changes, and frequently updating new data. Although segmenting customers according to their CLV is the recommended approach, there are many types of customer segmentation models. Some of the more common types are segmentation via cluster analysis, RFM segmentation, and longevity. Some marketers might even combine one or more segmentation models in order to reach their goals.

No matter the types of segmentation models marketers decide to use, they all require marketers to create groupings of customers to serve as a first step in segmenting the customer base. Usually this will result in marketers having a series of tiers for each type of segmentation model. Marketers can then mix different tiers across models to create more defined segments.

For example, mixing the highest tier of customers based on an RFM model and combining it with a low longevity tier will result in marketers having a segment of highly active, newly acquired customers.

Customer Segmentation and Machine Learning

An additional approach to customer segmentation is leveraging machine learning algorithms to discover new segments. Different to marketer-designed segmentation models, as the ones described above, machine learning customer segmentation allows advanced algorithms to surface insights and groupings that marketers might find difficulty discovering on their own.

Furthermore, marketers that create a feedback loop between the segmentation model and campaign results will have ever improving customer segments. In these cases, the machine learning model will be not only able to refine its definition of segments, but also be able to identify if a specific subset of the segment is outperforming the rest, optimizing marketing performance**.**

Customer segmentation vs. market segmentation

Companies can use marketing automation software to define and create customer segments. The customer segments can be based on demographic data, psychographic data and activity-based data such as actions that users took on a website. Companies use marketing automation software to configure, schedule and execute campaigns for particular customer segments.

Customer segmentation is different from market segmentation. An example of market segmentation is grouping customers by the products or services they purchase. A company may perform market segmentation based on distinct lines of business such as software, professional services and training. The company can then allocate resources to each market segment and employ separate marketing and activities to each.

Key Steps in a Needs-Based Market Segmentation Process:

Market segmentation strategy is an adaptive strategy. It consists of the operation of the market with the purpose of selecting one or more market segments which the organisation can target through the development of specific marketing mixes that adapt to particular market need.

Steps in Segmentation Process Description:

1. Needs-Based Segmentation : Group customers into segments based on similar needs and benefits sought by customer in solving a particular consumption problem.

2. Segment Identification : For each needs-based segment, determine which demographics, lifest--yles, and usage behaviors make the segment distinct and identifiable.

3. Assess Segment Attractiveness : Using predetermined segment attractiveness criteria, determine the overall attractiveness of each segment.

4. Evaluate Segment Profitability : Determine segment profitability (net marketing contribution).

5. Segment Positioning : For each segment, create a "value proposition" and product-price position--ning strategy based on that segment's unique customer needs and characteristics.

6. Segment "Acid Test" : Test the attractiveness of each segment's positioning strategy.

7. Marketing-Mix Strategy : Expand segment positioning strategy to include all aspects of the marketing mix: product, price, promotion, place, and people.

3.4 Benefits of customer segmentation

At the expansion stage, executing a marketing strategy without any knowledge of how your target market is segmented is akin to firing shots at a target 100 feet away—while blindfolded. The likelihood of hitting the target is a matter of luck more than anything else.Without a deep understanding of how a company’s best current customers are segmented, a business often lacks the market focus needed to allocate and spend its precious human and capital resources efficiently.

Further more, a lack of best current customer segment focus can lead to diffused go-to-market and product development strategies that hamper a company’s ability to fully engage with its

target segments. Together, all of those factors can ultimately impede a company's growth. If best current customer segmentation is done right, however, the business benefits are numerous. In order to identify the best current customer segments, the process is broken down into five clear steps, from setting up your project to performing customer data analysis, executing data collection, conducting customer segment analysis and prioritization, and incorporating the results into your organizational strategy.

While recognizing that being able to identify your best current customer segment can help your business is important, it is meaningless unless you act on it, or if you engage in segmentation activities that are more distracting than helpful. To be effective, you need to execute a best current customer segmentation process that is driven by a clearly defined set of objectives and outputs, and is backed by all of the company's relevant stakeholders. This guide will help you accomplish those tasks. The systematic and scientific data collection and analysis processes laid out in this guide might seem complicated, but they are not impossible to manage. Like almost any initiative, you simply need to ensure that key players and shareholders remain focused on their specific roles and responsibilities, and work collectively to achieve a clearly defined set of goals and objectives.

Chapter 4

SYSTEM DESIGN

4.1 Software requirements specifications

Software requirements specification is a rigorous assessment of requirements before the more specific system design stages, and its goal is to reduce later redesign. It should also provide a realistic basis for estimating product costs, risks, and schedules. Used appropriately, software requirements specifications can help prevent software project failure. The software requirements specification document lists sufficient and necessary requirements for the project development. To derive the requirements, the developer needs to have clear and thorough understanding of the products under development.

4.1.1 Language Used

Python is a cross-functional, maximally interpreted language that has lots of advantages to offer. The object-oriented programming language is commonly used to streamline large complex data sets. Over and above, having a dynamic semantics plus unmeasured capacities of RAD(rapid application development), Python is heavily utilized to script as well. There is one more way to apply Python – as a coupling language.

Another Python's advantage is high readability that helps engineers to save time by typing fewer lines of code for accomplishing the tasks. Being fast, Python jibes well with data analysis. And that's due to heavy support; availability of a whole slew of open-source libraries for different purposes, including but not limited to scientific computing. Therefore, it's not surprising at all that it's claimed to be the preferred programming language for data science. There is a scope of unique features provided that makes Python a-number-one option for data analysis

Python (version 3.8.5) is a general purpose, dynamic, high level and interpreted programming language. It supports Object Oriented programming approach to develop applications. It is simple and easy to learn and provides lots of high-level datastructures.

- .Python is easy to learn yet powerful and versatile scripting language which makes it attractive for Application Development.
- Python's syntax and dynamic typing with its interpreted nature, makes it an ideal language for scripting and rapid application development.

Python supports multiple programming pattern, including object oriented, imperative and functional or procedural programming

4.1.2 Tools – Anaconda

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda® distribution that allows you to launch applications and easily manage conda packages, environments, and channels without using command-line commands. Navigator can search for packages on Anaconda.org or in a local Anaconda Repository. It is available for Windows, macOS, and Linux.

In order to run, many scientific packages depend on specific versions of other packages. Data scientists often use multiple versions of many packages and use multiple environments to separate these different versions.

The command-line program conda is both a package manager and an environment manager. This helps data scientists ensure that each version of each package has all the dependencies it requires and works correctly.

Navigator is an easy, point-and-click way to work with packages and environments without needing to type conda commands in a terminal window. You can use it to find the packages you want, install them in an environment, run the packages, and update them – all inside Navigator.

The following applications are available by default in Navigator:

- JupyterLab
- Jupyter Notebook
- Spyder
- PyCharm
- VSCode
- Glueviz
- Orange 3 App
- RStudio
- Anaconda Prompt (Windows only)

The simplest way to run code with navigator is with Spyder. From the Navigator Home tab, click Spyder, and write and execute your code. You can also use Jupyter Notebooks the same way. Jupyter Notebooks are an increasingly popular system that combine your code, descriptive text, output, images, and interactive interfaces into a single notebook file that is edited, viewed, and used in a web browser.

Running Python in Spyder IDE (Integrated development environment)

1. On Navigator's Home tab, in the Applications pane on the right, scroll to the Spyder tile and click the Install button to install Spyder.
2. Launch Spyder by clicking Spyder's Launch button.
3. In the new file on the left, delete any placeholder text, then type or copy/paste print("Hello Anaconda")
4. In the top menu, click File - Save As and name your new program hello.py.
5. Run your new program by clicking the triangle Run button.
6. You can see your program's output in the bottom right Console pane.

4.1.3 IDE Used- Spyder

It is always necessary to have interactive environments to create software applications and this fact becomes very important when you work in the fields of Data Science, engineering, and scientific research. The Python Spyder IDE has been created for the same purpose.

Spyder is an open-source cross-platform IDE. The Python Spyder IDE is written completely in Python. It is designed by scientists and is exclusively for scientists, data analysts , and engineers. It is also known as the Scientific Python Development IDE and has a huge set of remarkable features which are discussed below.

Features of Spyder

- Customizable Syntax Highlighting
- Availability of breakpoints (debugging and conditional breakpoints)
- Interactive execution which allows you to run line, file, cell, etc.
- Run configurations for working directory selections, command-line options, current/ dedicated/ external console, etc
- Can clear variables automatically (or enter debugging)
- Navigation through cells, functions, blocks, etc can be achieved through the Outline Explorer
- It provides real-time code introspection (The ability to examine what functions, keywords, and classes are, what they are doing and what information they contain)

- Automatic colon insertion after if, while, etc
- Supports all the IPython magic commands
- Inline display for graphics produced using Matplotlib
- Also provides features such as help, file explorer, find files, etc

Editor: Spyder's multi-language Editor integrates a number of powerful tools right out of the box for an easy to use, efficient editing experience. The Editor's key features include syntax highlighting , real-time code and style analysis , on-demand completion, call tips and go-to-definition features , a function/class browser, horizontal and vertical splitting, and much more.

Console: The IPython Console allows you to execute commands and interact with data inside IPython interpreters. Any Ipython Console, whether external or started by Spyder, supports:

- Automatic code completion
- Real-time function calltips
- Full GUI integration with the enhanced Spyder Debugger.
- The Variable Explorer, with GUI-based editors for many built-in and third-party Python objects.
- Display of Matplotlib graphics in Spyder's plots pane, if the Inline backend is selected under Preferences ‣ IPython console ‣ Graphics ‣ Graphics backend, and inline in the console if Mute inline plotting is unchecked under the Plots pane's options menu.

Plots: plots pane shows the static figures and images created during your session. It will show you plots from the IPython Console, produced by your code in the Editor or generated by the Variable Explorer allowing you to interact with them in several ways.

Debugging: In Spyder, debugging is supported through integration with the enhanced ipdb debugger in the IPython Console. This allows breakpoints and the execution flow to be viewed and controlled right from the Spyder GUI, as well as with all the familiar IPython console commands.

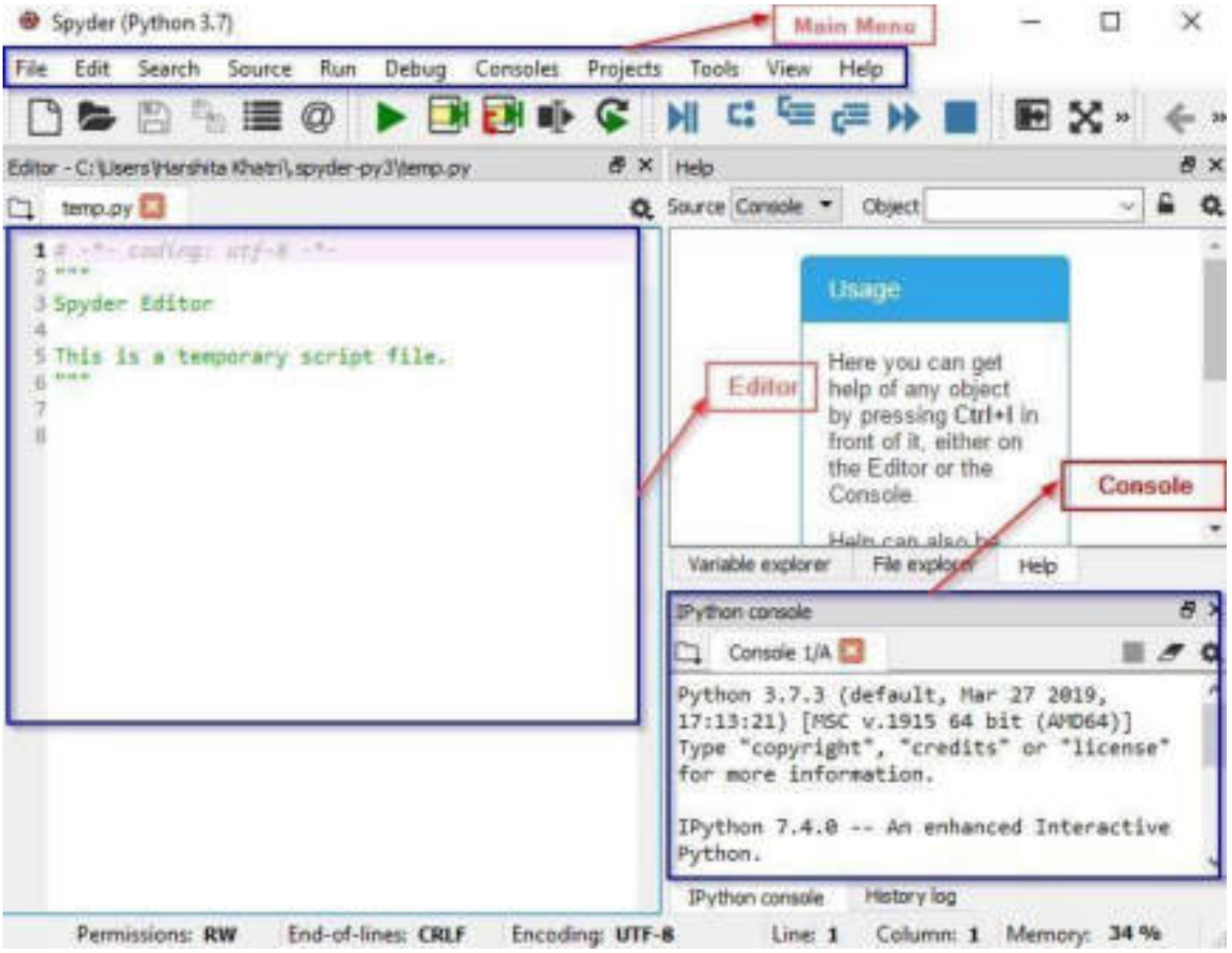


Figure 4.1 : Spyder IDE

4.1.4 Libraries Used

1. **Numpy :** NumPy is the fundamental package for scientific computing with Python. It contains among otherthings:

   - A powerful N-dimensional array object
   - sophisticated (broadcasting) functions
   - tools for integrating C/C++ and Fortrancode
   - useful linear algebra, Fourier transform, and random numbercapabilities
   - Besides its obvious scientific uses, NumPy can also be used as an efficient multi- dimensional container of generic data. Arbitrary data-types can be defined. This allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

2. **Pandas :** Pandas comes handy as it was developed specifically for data extraction and preparation. It provides high-level data structures and wide variety tools for data analysis. It provides many inbuilt methods for groping, combining and filtering data.
3. **Matplotlib:** Matpoltlib is a very popular Python library for data visualization. Like Pandas, it is not directly related to Machine Learning. It particularly comes in handy when a programmer wants to visualize the patterns in the data. It is a 2D plotting library used for creating 2D graphs and plots. A module named pyplot makes it easy for programmers for plotting as it provides features to control line styles, font properties, formatting axes, etc. It provides various kinds of graphs and plots for data visualization, viz., histogram, error charts, bar chats, etc,
4. **Seaborn**: It is a data visualization library built on top of matplotlib and closely integrated with pandas data structures in Python. Visualization is the central part of Seaborn which helps in exploration and understanding of data.
5. **Sklearn**: Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python. This library, which is largely written in Python, is built upon NumPy, SciPy and Matplotlib.

4.2 UML Diagrams

Use-oriented techniques are widely used in software requirement analysis and design. Use cases and usage scenarios facilitate system understanding and provide a common language for communication. This paper presents a scenario-based modeling technique and discusses its applications. In this model, scenarios are organized hierarchically and they capture the system functionality at various abstraction levels including scenario groups, scenarios, and sub-scenarios. Combining scenarios or sub-scenarios can form complex scenarios. Data are also separately identified, organized, and attached to scenarios. This scenario model can be used to cross check with the UML model. It can also direct systematic scenario-based testing including test case generation, test coverage analysis with respect to requirements, and functional regression testing.

4.2.1 Class Diagram

Model, objects are entities that combine state (i.e., data), behavior (i.e., procedures, or methods) and identity (unique existence among all other objects). The structure and behavior of an object are defined by a class, which is a definition, or blueprint, of all objects of a specific type. An object must be explicitly created based on a class and an object thus created is considered to be an instance of that class. An object is similar to a structure, with the addition of method pointers, member access control, and an implicit data member which locates instances of the class (i.e. actual objects of that class) in the class hierarchy (essential for runtime inheritance features)

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, and the relationships between the classes.

The class diagram is the main building block in object oriented modeling. It is used both for general conceptual modeling of the semantics of the application, and for detailed modeling translating the models into programming code. The classes in a class diagram represent both the main objects and or interactions in the application and the objects to be programmed. In the class diagram these classes are represented with boxes which contain the two parts:

- The upper part holds the name of theclass.
- The middle part contains the attributes oftheclass.
- The lower part contains the operations oftheclass

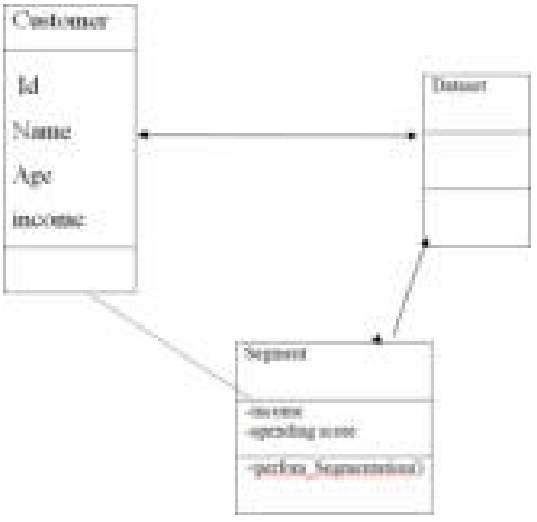


Figure 4.2: **Class Diagram**

4.2.2 Use Case Diagram

A Use Case Diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use- case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those usecases.

The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted. Interaction among actors is not shown on the use case diagram. If this interaction is essential to a coherent description of the desired behavior, perhaps the system or use case boundaries should be re-examined. Alternatively, interaction among actors can be part of the assumptions used in the usecase.

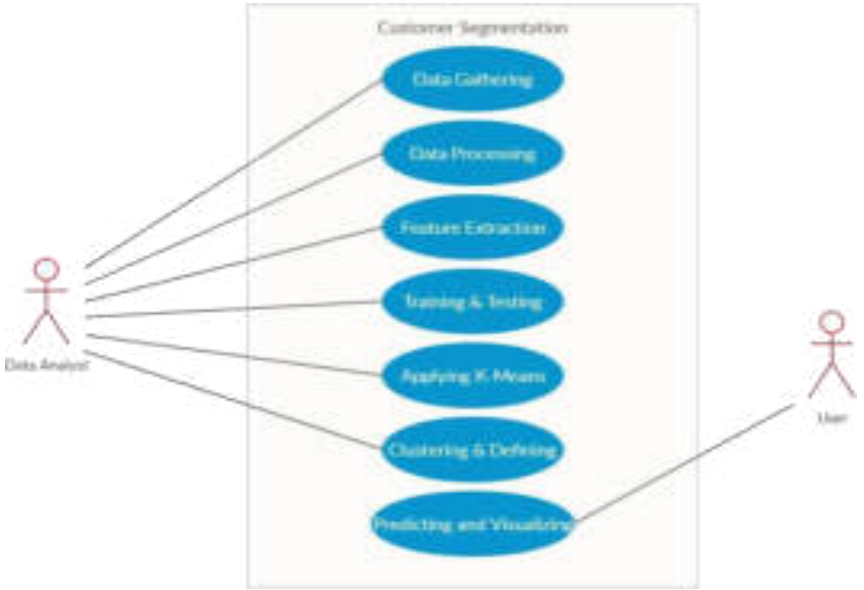
Use cases:

A use case describes a sequence of actions that provide something of measurable value to an actor and is drawn as a horizontal ellipse.

Actors:

An actor is a person, organization, or external system that plays a role in one or more interactions with the system.

Figure 4.3: **Use Case Diagram**



4.2.3 Sequence Diagram

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart.

Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams. A sequence diagram shows, as parallel vertical lines (lifelines), different processes or objects that live simultaneously, and, as horizontal arrows, the messages exchanged between them, in the order in which they occur. This allows the specification of simple runtime scenarios in a graphical manner. If the lifeline is that of an object, it demonstrates a role.

Note that leaving the instance name blank can represent anonymous and unnamed instances. In order to display interaction, messages are used. These are horizontal arrows with the message name written above them. Solid arrows with full heads are synchronous calls, solid arrows with stick heads are asynchronous calls and dashed arrows with stick heads are return messages.

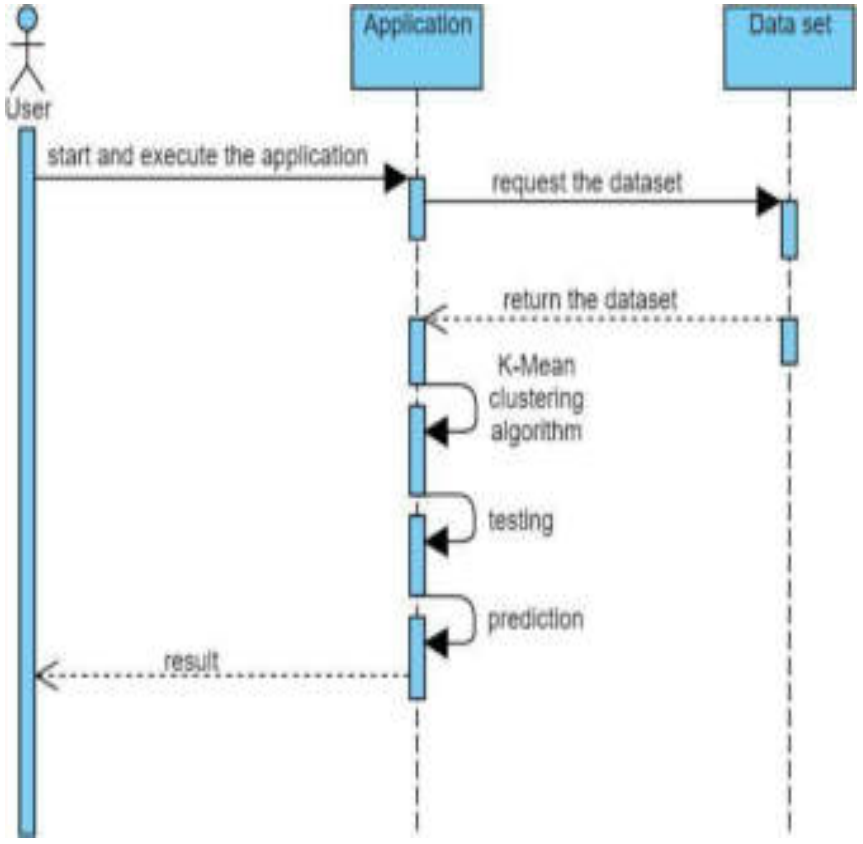


Figure 4.4: **Sequence Diagram**

4.2.4 Data Flow Diagram

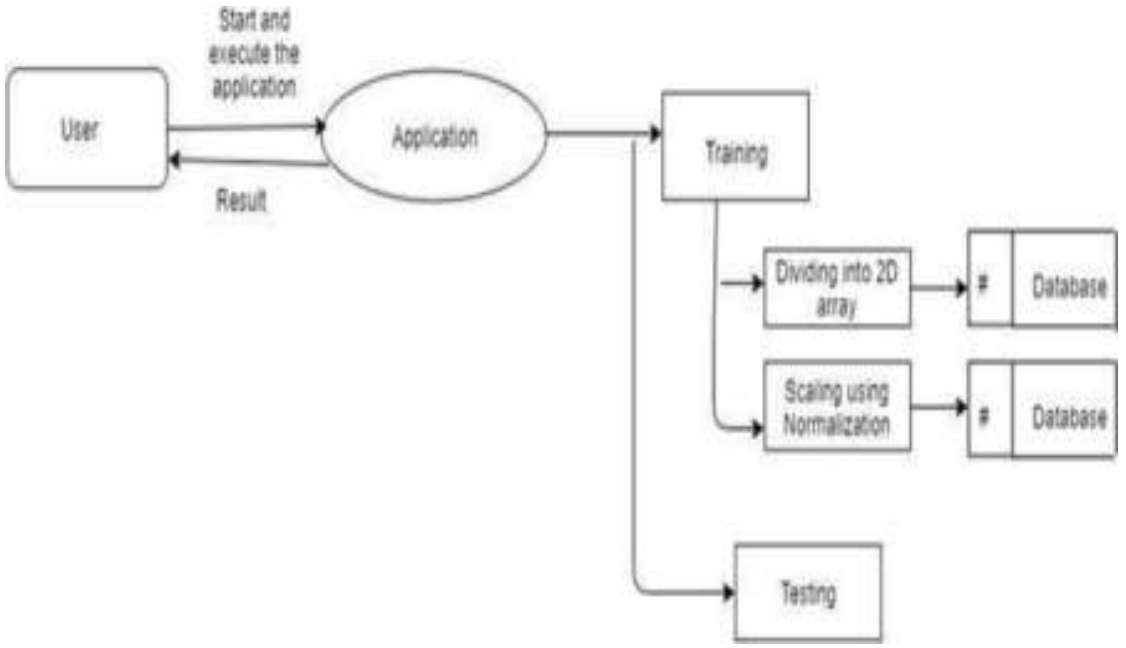


Figure 4.5: **Data Flow Diagram**

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It can be manual, automated, or a combination of both. It shows how data enters and leaves the system, what changes the information, and where data is stored.The objective of a DFD is to show the scope and boundaries of a system as a whole. It may be used as a communication tool between a system analyst and any person who plays a part in the order that acts as a starting point for redesigning a system. The DFD is also called as a data flow graph or bubble chart.

1. All names should be unique. This makes it easier to refer to elements in the DFD.
2. Remember that DFD is not a flow chart. Arrows is a flow chart that represents the order of events; arrows in DFD represents flowing data. A DFD does not involve any order of events.
3. Suppress logical decisions. If we ever have the urge to draw a diamond-shaped box in a DFD, suppress that urge! A diamond-shaped box is used in flow charts to represents decision points with multiple exists paths of which the only one is taken. This implies an ordering of events, which makes no sense in a DFD.

4. Do not become bogged down with details. Defer error conditions and error handling until the end of the Analysis.

This level explains the process of the system in a very detailed man- ner. In first level DFD (Generation of individual fields): how data flows through individual process in it are shown. In second detailed level DFD: how data flows through the system to form a detailed de- scription of the individual processes After starting and executing the application, training the dataset is done by using dividing into 2Darray and scaling using normalization algorithms, and then testing is done.

4.2.5 Architecture Diagram:

The block diagram is basially used for a higher level, less detailed description aimed at understanding the overall concepts and less at understanding the implementation.

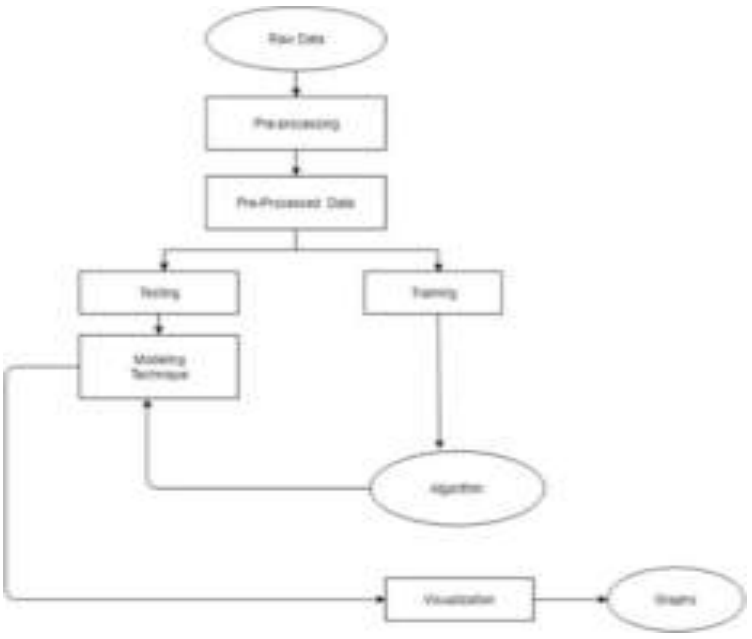


Figure 4.6: **Architecture Diagram**

A simple architecture diagram (UML) helps system designers and developers visualize the high-level structure of their system or application to ensure it meets their users' needs. It can also help describe patterns that are used throughout the design. Similar to a blueprint, teams can use our simple architecture diagram template as a guide to easily discuss and improve their products.

Any real-world system is used by different users. The users can be developers, testers, business people, analysts, and many more. Hence, before designing a system, the architecture is made with different

perspectives in mind. The most important part is to visualize the system from the perspective of different viewers. The better we understand the better we can build the system.

UML plays an important role in defining different perspectives of a system. These perspectives are −

- Design
- Implementation
- Process
- Deployment

Software architecture is all about how a software system is built at its highest level. It is needed to think big from multiple perspectives with quality and design in mind. The software team is tied to many practical concerns, such as:

- The structure of the development team.
- The needs of the business.
- Development cycle.
- The intent of the structure itself.

Software architecture provides a basic design of a complete software system. It defines the elements included in the system, the functions each element has, and how each element relates to one another. In short, it is a big picture or overall structure of the whole system, how everything works together.

To form an architecture, the software architect will take several factors into consideration:

- What will the system be used for?
- Who will be using the system?
- What quality matters to them?
- Where will the system run?

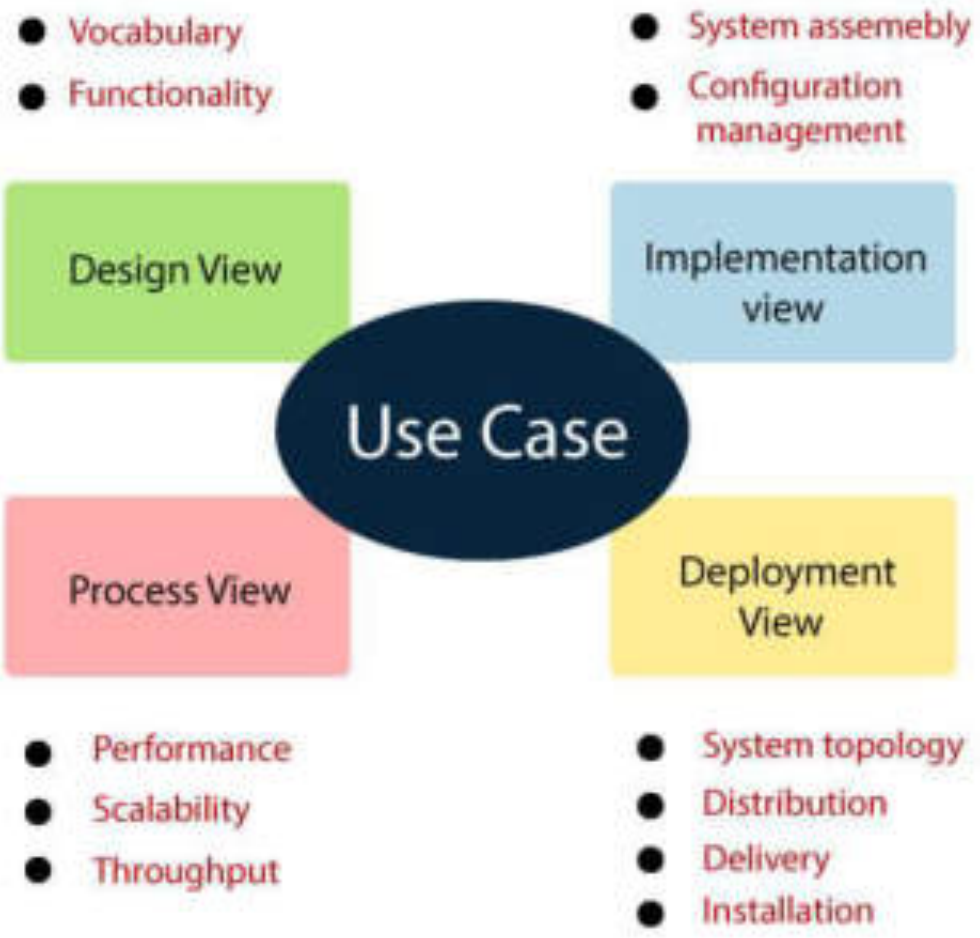
The architect plans the structure of the system to meet the needs like these. It is essential to have proper software architecture, mainly for a large software system. Having a clear design of a complete system as a starting point provides a solid basis for developers to follow.

Each developer will know what needs to be implemented and how things relate to meet the desired needs efficiently. One of the main advantages of software architecture is that it provides high productivity

to the software team. The software development becomes more effective as it comes up with an explained structure in place to coordinate work, implement individual features, or ground discussions on potential issues. With a lucid architecture, it is easier to know where the key responsibilities are residing in the system and where to make changes to add new requirements or simply fixing the failures.

In addition, a clear architecture will help to achieve quality in the software with a well-designed structure using principles like separation of concerns; the system becomes easier to maintain, reuse, and adapt. The software architecture is useful to people such as software developers, the project manager, the client, and the end-user. Each one will have different perspectives to view the system and will bring different agendas to a project. Also, it provides a collection of several views. It can be best understood as a collection of five views:

1. Use case view
2. Design view
3. Implementation view
4. Process view
5. Development view



Use case view

- It reveals the requirements of the system.
- It is a view that shows the functionality of the system as perceived by external actors.
- With UML, it is easy to capture the static aspects of this view in the use case diagrams, whereas its dynamic aspects are captured in interaction diagrams, state chart diagrams, and activity diagrams.

Design View

- It is a view that shows how the functionality is designed inside the system in terms of static structure and dynamic behavior.
- It captures the vocabulary of the problem space and solution space.
- With UML, it represents the static aspects of this view in class and object diagrams, whereas its dynamic aspects are captured in interaction diagrams, state chart diagrams, and activity diagrams.

Implementation View

- It is the view that represents the organization of the core components and files.
- It primarily addresses the configuration management of the system?s releases.
- With UML, its static aspects are expressed in component diagrams, and the dynamic aspects are captured in interaction diagrams, state chart diagrams, and activity diagrams.

Process View

- It is the view that demonstrates the concurrency of the system.
- It incorporates the threads and processes that make concurrent system and synchronized mechanisms.
- It primarily addresses the system's scalability, throughput, and performance.
- Its static and dynamic aspects are expressed the same way as the design view but focus more on the active classes that represent these threads and processes.

Deployment View

- It is the view that shows the deployment of the system in terms of physical architecture.
- It includes the nodes, which form the system hardware topology where the system will be executed.
- It primarily addresses the distribution, delivery, and installation of the parts that build the physical system.

Hardware Specifications

- RAM: 2GB or above
- Processor: i3 or above
- Processor speed: 2.4GHz
- Hard disk: 500GB

Software Specifications

- Programming Language: Python 3.7
- Tools: Anaconda – Spyder IDE
- Operating system: Windows

CHAPTER 5

MODULE DESCRIPTION

Customer segmentation procedures

Customer segmentation, also called consumer segmentation or client segmentation, procedures include:

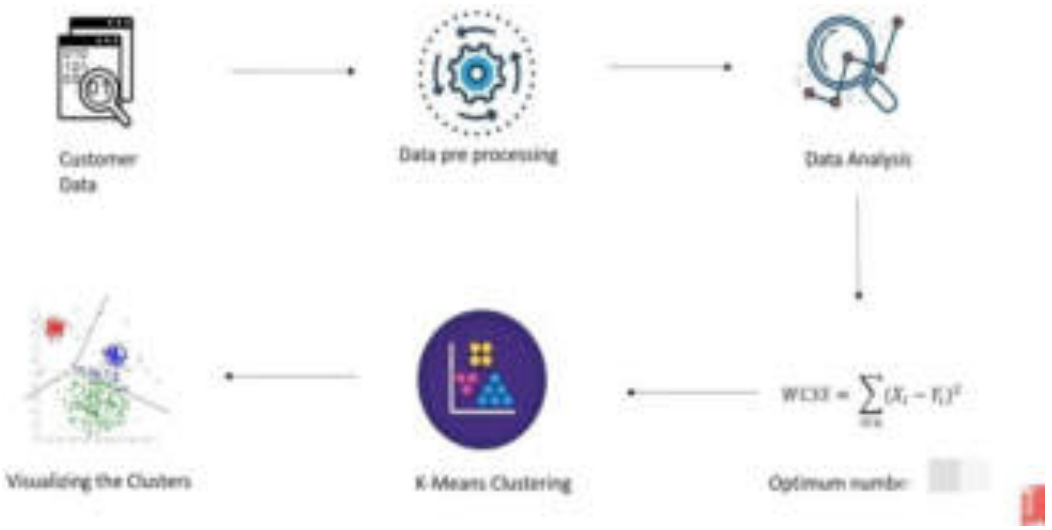
- Deciding what data will be collected and how it will be gathered Collecting data and integrating data from various sources Developing methods of data analysis for segmentation
- Establishing effective communication among relevant business units (such as marketing and customer service) about the segmentation Implementing applications to effectively deal with the data and respond to the information it provides.

Modules

The project contains the following main modules:

- Gathering of Data
- Data Preprocessing
- Feature Extraction
- Appling K-means Algorithms.
- Clustering the data and defining all clusters
- Visualization and Market strategy suggested

Figure 5.1 : Modules



Gathering of Data

Firstly, we will define our libraries like NumPy, pandas, matplotlib and sklearn in the code. NumPy used for matrix multiplications, Pandas is used for data manipulation, Matplotlib for visualization of data and sklearn is for Machine learning Algorithms.

Next step is to import the Mall Customers dataset in the code by using Pandas module. The dataset contains features like custid, Annual Income, Spending Score and so on. After importing the data, we should check for the extreme values and missing values in all the Data features of the Mall customers dataset.

Data Preprocessing

Data preprocessing is a process of preparing the raw data of Mall customers suitable for a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So, for this, we use data pre-processing task. In the data Pre-processing step, data can be preprocessed before us- ing the whole data for the model. It includes removing of extreme values and filling up of the missing values either with mean or me- dian or based on the customer requirements in all the features of the dataset. In this process also includes scaling of the data which means making the whole data comes under the one scale rather than multiple scales.

Feature Extraction

In the Mall customers dataset includes multiple features like custid, Annual Income, Spending score and so on. By using all those features, the accuracy of the model to predict the cluster is low. To increase the accuracy of the model we should have to select the important or relevant features by using which we can get the accurate prediction of cluster to which item belongs to.

Feature extraction is a process of reduction by which first set of raw data is reduced to more manageable groups for processing. This process is useful when you need to reduce the number of resources needed for processing the data without losing important or relevant information. A characteristic of these large datasets is a large number of variables that require a lot of computing resources to process. It is the name for methods that select or combine variables into features, effectively decreasing the amount of data that must be processed, still accurately and completely describing the original dataset.

It can also reduce the amount of redundant data for a given dataset. Also, the reduction of the data and the machine's efforts in building features facilitate the speed of learning and generalization steps in the machine learning process.

Benefits of feature extraction are:

a) Reduces Overfitting: It means whenever all the features of the Mall

Customers dataset are used to create a model then for training data the accuracy rate will be high whereas for testing data accuracy rate will be low so that the model is overfitted to training data of Mall Customers. So, by using feature extraction can reduce Overfitting.

b)Improves Accuracy: By using important or relevant features the accuracy rate of model gets improved

c) Reduces Training Time: By using feature extraction we can reduce the Unrelated features of the dataset which also reduces the training dataset, thus the

processing time of training data set to the model gets reduced.

Appling K-means Algorithm

By using cluster module which is in sklearn module in python we will apply K-means clustering algorithm to the training dataset to create a machine learning model which will

predict item's cluster. K-means is one of the unsupervised learning algorithms that solve the well-known clustering problem.

The procedure follows easy way to classify a given data set through a certain number of clusters fixed a priority.

The main idea is to define the k-centroids, one for each cluster. These centroids should be placed as much as possible far away from each other. At this point we need to recalculate k-new centroids as barycentre of the clusters resulting from previous step. A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are to be done.

Clustering the data and defining all clusters

In business decision support system, clusters play vital role. Cluster is the group of customers who shares their similarities. We are given a data set of Mall Customers, with certain features, and values for these features. The task is to categorize those items into clusters. To achieve this, we will use the K-Means clustering algorithm, an unsupervised learning algorithm. By applying clustering algorithm, we will get k clusters, then every cluster has its own label which is assigned to the data of respective cluster.

Every cluster's behavior is different from others. The behavior of every cluster is defined by the distance between the centroid and the customer data. The items which matches with the cluster's behavior will be in that respective clusters. In such way required number of clusters are defined.

Visualization and Market strategy suggested

To visualize the clusters of the dataset we need to map them onto the graph by using matplotlib module which is present in the python. Every cluster's behavior is different from others. Based on the behavior of the cluster the decision is made which comes under the market strategy suggested.

Chapter 6

IMPLEMENTATION

Implementation is the phase of the venture when the plan is transformed out into a working framework. Along these lines it very well may be viewed as the most basic stage in accomplishing a new one and in giving the client, certainty that the new framework will work and be effective.

The execution arrange includes cautious arranging, examining of the current framework and its requirements knowing of techniques to accomplish changeover and assessment of changeover strategies.

6.1 Input

This task include preparing a Machine learning model to record the Annual Income and Spending score of the customers. The dataset incorporates spending activity of 200 unique people. Here, the dataset of Mall customers is taken from Kaggle.

Initially the data set will have five attributes, customer id, name, gender, annual income, Spending score .

The CSV file with all these values in shown below.

	A	B	C	D	E	F	G
1	CustomerI	Gender	Age	Annual Inc	Spending Score (1-100)		
2	1	Male	19	15	39		
3	2	Male	21	15	81		
4	3	Female	20	16	6		
5	4	Female	23	16	77		
6	5	Female	31	17	40		
7	6	Female	22	17	76		
8	7	Female	35	18	6		
9	8	Female	23	18	94		
10	9	Male	64	19	3		
11	10	Female	30	19	72		
12	11	Male	67	19	14		
13	12	Female	35	19	99		
14	13	Female	58	20	15		
15	14	Female	24	20	77		
16	15	Male	37	20	13		
17	16	Male	22	20	79		
18	17	Female	35	21	35		
19	18	Male	20	21	66		
20	19	Male	52	23	29		
21	20	Female	35	23	98		
22	21	Male	35	24	35		
23	22	Male	25	24	73		
24	23	Female	46	25	5		
25	24	Male	31	25	73		
26	25	Female	54	28	14		
27	26	Male	29	28	82		
28	27	Female	45	28	32		
29	28	Male	35	28	61		
30	29	Female	40	29	31		
31	30	Female	23	29	87		

Mall_Customers

Figure6.1: **Data set inputed**

6.2 Output Design

This Output will be a plotted graph with clusters and each cluster will include a centroid.

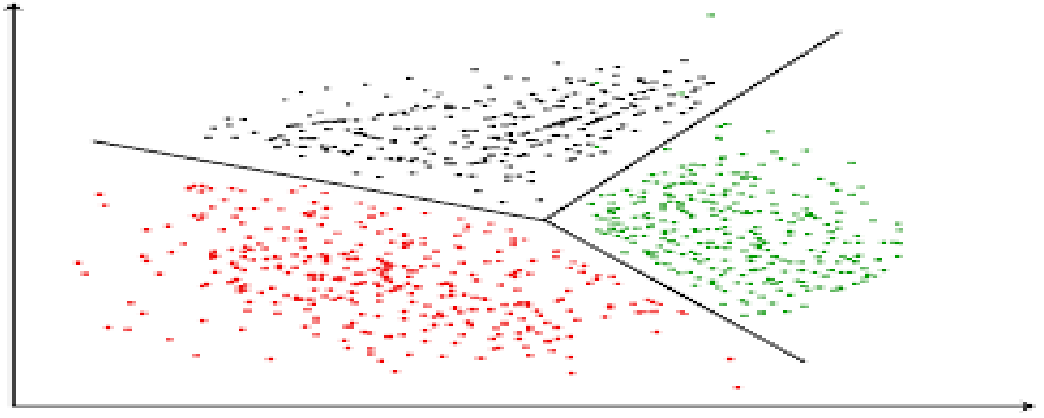


Figure 6.2: Clustered Output

6.3 Elbow Method

The Elbow Method is one of the most popular methods to determine this optimal value of k. In the Elbow method, we are actually varying the number of clusters (K) from 1 – 10. For each value of K, we are calculating WCSS (Within-Cluster Sum of Square).

WCSS is the sum of squared distance between each point and the centroid in a cluster. When we plot the WCSS with the K value, the plot looks like an Elbow. As the number of clusters increases, the WCSS value will start to decrease. WCSS value is largest when K = 1.

When we analyze the graph we can see that the graph will rapidly change at a point and thus creating an elbow shape. From this point, the graph starts to move almost parallel to the X-axis. The K value corresponding to this point is the optimal K value or an optimal number of clusters.

6.4 K-Means Algorithm

It is the simplest and commonly used iterative type unsupervised learning algorithm. In this, we randomly initialize the **K** number of centroids in the data (the number of k is found using the Elbow method) and iterates these centroids until no change happens to the position of the centroid. Let's go through the steps involved in K means clustering for a better understanding.

1) Select the number of clusters for the dataset (K)
2) Select K number of centroids randomly
3) By calculating the Euclidean distance or Manhattan distance assign the points to the nearest centroid, thus creating K groups
4) Now find the original centroid in each group
5) Again reassign the whole data point based on this new centroid, then repeat step 4 until the position of the centroid doesn't change.

Finding the optimal number of clusters is an important part of this algorithm. A commonly used method for finding optimal K value is Elbow Method.

6.5 WCSS Calculation

WCSS is the sum of squares of the distances of each data point in all clusters to their respective centroids.

$$\text{WCSS}(k) = \sum_{j=1}^{k} \sum_{\mathbf{x}_i \in \text{cluster } j} \|\mathbf{x}_i - \bar{\mathbf{x}}_j\|^2,$$

where $\bar{\mathbf{x}}_j$ is the sample mean in cluster j

The idea is to minimise the sum. Suppose there are n observation in a Given dataset and we specify n number of clusters (k = n) then WCSS will become zero since data points themselves will act as centroids and the distance will be zero and ideally this forms a perfect cluster, however this doesn't make any sense as we have as many clusters as the observations.

Thus there exists a threshold value for K which we can find using the Elbow point graph. We can find the optimum value for K using an Elbow point graph. We randomly initialize the K-Means algorithm for a range of K values and will plot it against the WCSS for each K value.

The within-cluster sum of squares is a measure of the variability of the observations within each cluster. In general, a cluster that has a small sum of squares is more compact than a cluster that has a large sum of squares.

Clusters that have higher values exhibit greater variability of the observations within the cluster. As the number of observations increases, the sum of squares becomes larger. Therefore, the within- cluster sum of squares is often not directly comparable across clusters with different numbers of observations.

6.6 Implementation steps

1. Importing relevant libraries

First we import all the required libraries for our project.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
```

2. Data collection and analysis

Then we import the data set of our project.

```
# loading the data from csv file to a Pandas DataFrame
customer_data = pd.read_csv('/content/Mall_Customers.csv')
```

```
# first 5 rows in the dataframe
customer_data.head()
```

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

3. Checking for Missing values

Then we check for all the missing values in the dataset.

```
# checking for missing values
customer_data.isnull().sum()
```

```
CustomerID                0
Gender                    0
Age                       0
Annual Income (k$)        0
Spending Score (1-100)    0
dtype: int64
```

Here, there are no missing or null values. So we can proceed further with the dataset.

4.Feature extraction

We won't get accurate results using all the attributes. So we will proceed with only 2 columns, Annual income and spending score.

```
[[ 15  39]
 [ 15  81]
 [ 16   6]
 [ 16  77]
 [ 17  40]
 [ 17  76]
 [ 18   6]
 [ 18  94]
 [ 19   3]
 [ 19  72]
 [ 19  14]
 [ 19  99]
 [ 20  15]
 [ 20  77]
 [ 20  13]
 [ 20  79]
 [ 21  35]
 [ 21  66]
 [ 23  29]
 [ 23  98]
 [ 24  35]
```

5. Training the K-means Clustering Model

```
kmeans = KMeans(n_clusters=5, init='k-means++', random_state=0)

# return a label for each data point based on their cluster
Y = kmeans.fit_predict(X)

print(Y)
```

```
[3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3
 1 3 1 3 1 3 0 3 1 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 2 4 2 0 2 4 2
 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4
 2 4 2 4 2 4 2 4 2 4 2 4 2 4 2]
```

6.7 Sample Code

```
#importing necessary modules
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import sklearn
from sklearn.cluster import KMeans


# loading the data from csv file to a Pandas DataFrame
customer_data = pd.read_csv('c:/Users/srilu/Desktop/Python_project/Mall_Customers.csv')

# first 5 rows in the dataframe
customer_data.head()
```

```
        # finding the number of rows and columns
        customer_data.shape

# getting some informations about the dataset

   customer_data.info()
         # checking for missing
  values
  customer_data.isnull().sum()
    X = customer_data.iloc[:,[3,4]].values

 # finding wcss value for different number of  clusters
wcss = []
  for i in range(1,11):
  kmeans = KMeans(n_clusters=i, init='k-means++', random_state=42) kmeans.fit(X)
  wcss.append(kmeans.inertia_)


  #Training K-Means
  kmeans = KMeans(n_clusters=5, init='k-means++', random_state=0)


   # return a label for each data point based on their cluster
   Y = kmeans.fit_predict(X)


  # plotting all the clusters and their Centroids
  plt.figure(figsize=(8,8))
  plt.scatter(X[Y==0,0], X[Y==0,1], s=50, c='green', label='Cluster 1')
  plt.scatter(X[Y==1,0], X[Y==1,1], s=50, c='red', label='Cluster 2')
  plt.scatter(X[Y==2,0], X[Y==2,1], s=50, c='yellow', label='Cluster 3')
  plt.scatter(X[Y==3,0], X[Y==3,1], s=50, c='violet', label='Cluster 4')
  plt.scatter(X[Y==4,0], X[Y==4,1], s=50, c='blue', label='Cluster 5')
  # plot the centroids
  plt.scatter(kmeans.cluster_centers_[:,0], kmeans.cluster_centers_[:,1], s=100, c='cyan',
  label='Centroids')
```

```
plt.title('Customer Groups')
plt.xlabel('Annual Income')
plt.ylabel('SpendingScore')
 plt.show()
```

Chapter 7

TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, subassemblies, assemblies and/or a finished product.

It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

7.1 Types of testing

Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive.

Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expectedresults.

Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent.

Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document.

It is a testing in which the software under test is treated, as a black box. you cannot "see" into it. The test provides inputs and responds to outputs without considering how the software work

User Acceptance Testing

User Acceptance of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changes wherever required. The system developed provides a friendly user interface that can easily be understood even by a person who is new to the system.

Output Testing

After performing the validation testing, the next step is output testing of the proposed system, since no system could be useful if it does not produce the required output in the specified format. Asking the users about the format required by them tests the outputs generated or displayed by the system under consideration. Hence the output format is considered in 2 ways – one is on screen and another in printed format.

10 tests used to assess the quality of segmentation useful.

1. Segmentation needs to be practical

Segments need to be sufficient in size to make targeting them worthwhile your effort and they need to be in a position to engage with your offer. Having too many segments will also make managing

your strategies tedious. Too few and it won't give you the granularity you need. Between 5 and 8 segments tends to work best in many cases. In order to devise coherent strategies and develop long-term relationships, the segments should also be mutually exclusive – each individual can only sit in one segment.

2. Segments must be discernably different

When segmenting your audience (or potential audience) the first thing is to make sure the segments have discernably different needs. If one segment has no distinguishing differences from another or they all respond in a similar way, you will end up with homogenised strategies, which defeats the point of segmentation.

3. Don't confuse behavioural clusters with segments

In Arts marketing for example, the Box Office database holds detailed information on behavioural patterns (performances selected, seat choices, frequency, party size, planning horizons, geography and so on). However this only describes WHAT your segments are CURRENTLY doing. It can't tell you WHY they do these things. People may frequently move cluster but they shouldn't frequently move segment. To be really effective, segments should be defined by WHAT each seeks to get out of the experience.

4. Use attitudinal segmentation

Not all C2DE families share the same needs and motivations, the same applies to 'young people' or BAMEs. Using demographics as a basis for a segmentation system makes for simple evaluation, but can't give you everything you need to inform planning. Attitudinal segmentation provides you with deep insight into your audiences.

Understanding what drives audiences, how they want to be made to feel and what motivates their involvement, enables you to identify and effectively target groups of individuals with shared values. This puts you in a position to influence their behaviour and deliver deeply satisfying experiences.

5. Segmentation should lend itself to differentiated strategies and campaigns

The same production or exhibition or product can appeal to a number of different segments for different reasons. One may seek primarily social outcomes, another driven to learn something new and a third be looking for emotional resonance. These three segments will recognise different benefits from the same work. They will respond to different messages via different platforms and be influenced by different factors.

Differentiated campaigns with messages optimized to resonate with a particular segment are more effective than a generic message aimed at everyone but speaking strongly to no one.

6. It should lend itself to development of products and services

As well as increasing your ability to reach segments, understanding what they want to get out of the experiences you offer, puts you in a position to the best experience and to shape the audience journey. Front of house ambience, customer service, catering, learning opportunities and merchandise can all be developed to cater for priority segment needs.

7. Remember to consider everyone in your priority segments

With limited resources the easy, obvious but potentially disastrous solution is to contact just those who are most actively and most recently engaged. For a time this might work but rinse and repeat this narrow selection process for any length of time and the result will be chronic audience-underdevelopment.

8. Great segmentation doesn't live in the marketing department

A segmentation system delivered by marketing to other departments will often fail to become embedded within the organisation. Segmentation needs to involve everyone and belong to everyone.

In order to be immediately and intuitively recognisable and credible across departments the segmentation process has to be iterative with everyone involved in identifying then and fleshing them out. This way it will provide a common language for talking about audiences – bridging the understanding of marketers, programmers, front of house, learning, fundraising, hospitality – putting audiences at the centre of the conversation.

9. Build it into your ongoing research

Segmentation, the gift that keeps on giving… whatever research you are conducting or data you are analysing, building your segments into this means you immediately have a more subtle and granular understanding of the outcomes.

10. You need to be able to monitor and evaluate your success

This should perhaps be my first, not last point. As with all evaluation, it needs to be considered from the get-go. Make sure you've got mechanisms in place to monitor what works best, where the greatest return on investment lies, who is responding, which messages resonate with which audience segments via which platforms.

This means you can adjust your strategies and change your messaging as you go. Segment evaluation also helps you prioritize and plan developments based on how well you are currently meeting the needs of your segments – a virtuous circle of improvement.

7.2 Test Cases:

1. Giving a sample data set with 12 random customer values and exploring the clusters formed.

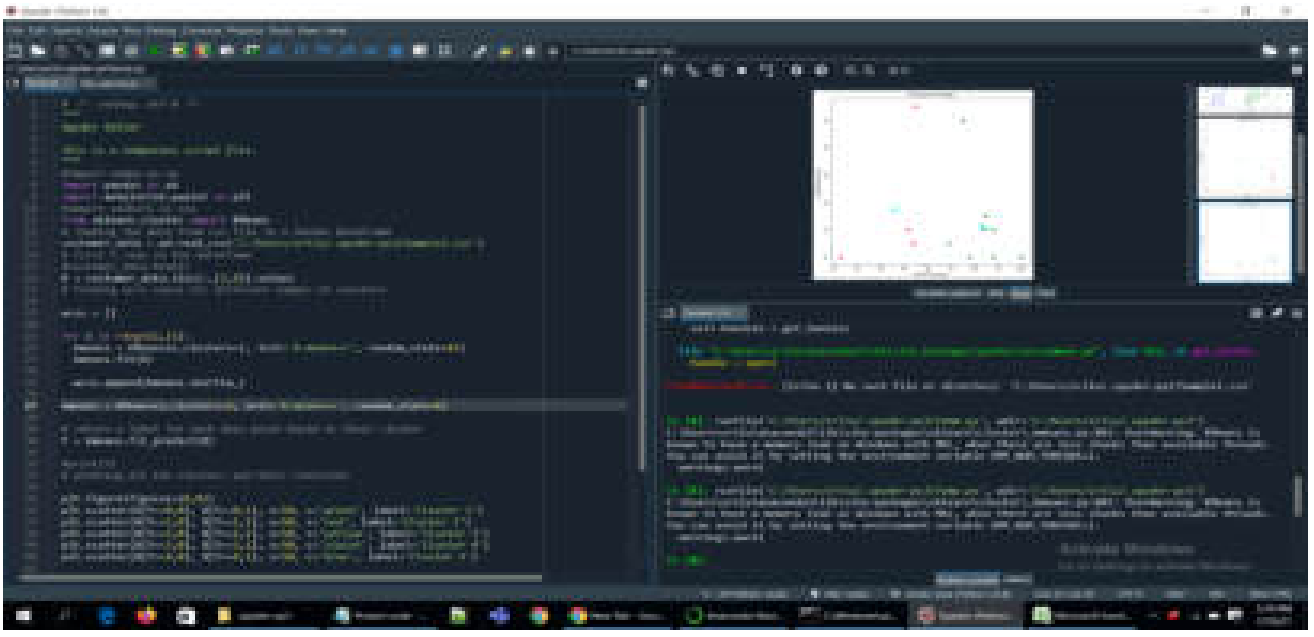


Figure7.1: **Test case1**

2. Giving the number of clusters to be formed as one for the mall Customers dataset.

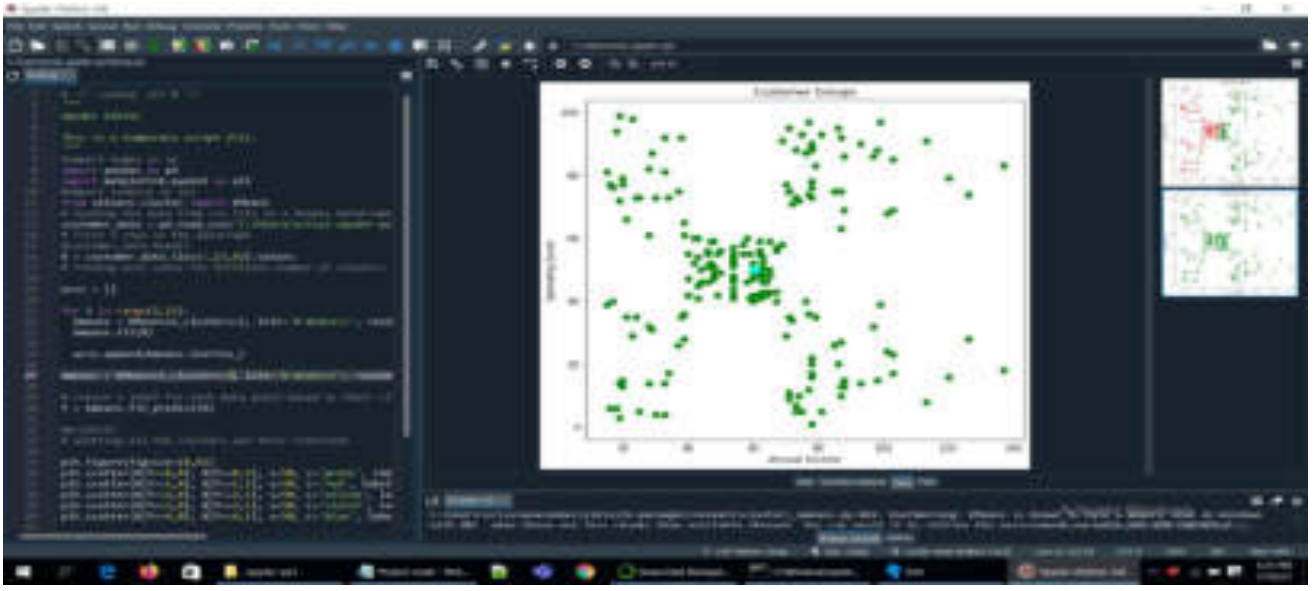


Figure7.2: **Test Case2**

As we can see above, if we give the number of clusters as one then all values in the dataset comes under a single cluster. So the results are not accurate.

3. Giving the number of clusters to be formed as two for the mall Customers dataset.

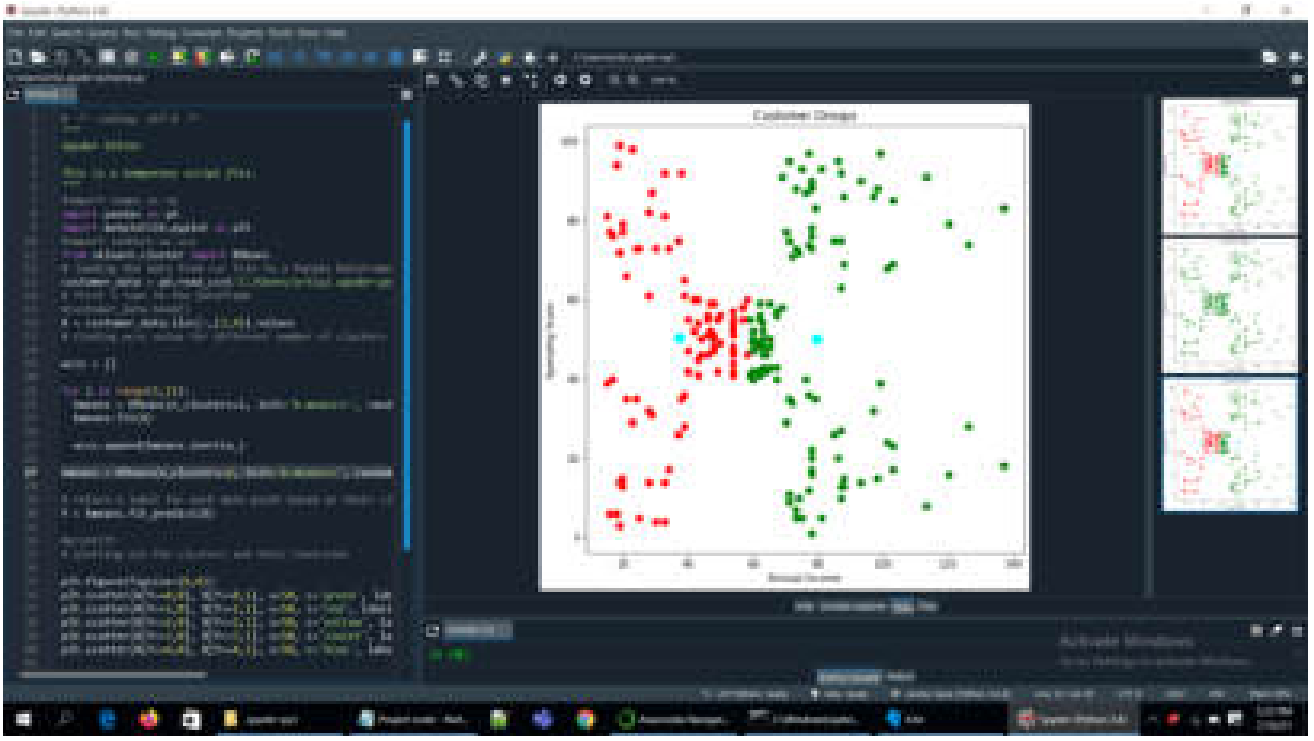


Figure7.3: **Test Case3**

As we can see above, if we give the number of clusters as two then all values in the dataset are catagerized two two clusters and thus will have two centroids.

4. Giving the number of clusters to be formed as three for the mall Customers dataset.

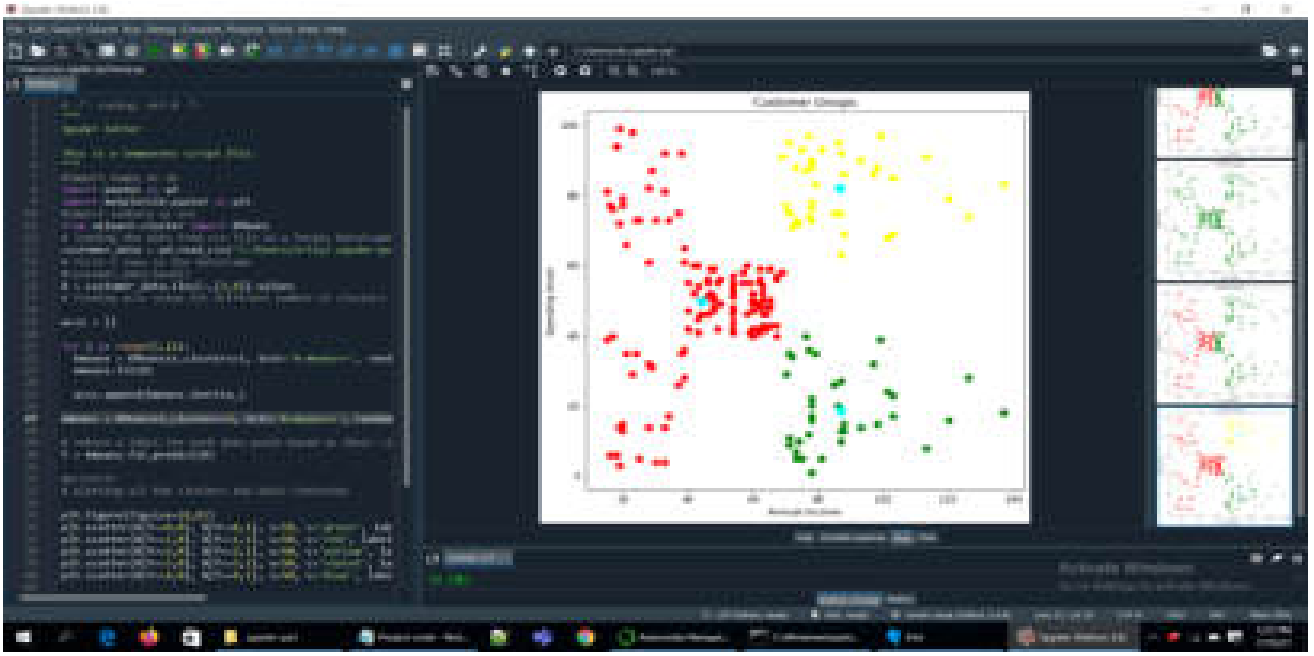


Figure7.4: **Test Case4**

As we can see in the above figure , if we give the number of clusters as three then all values in the will form three clusters and thus will have three centroids.
Similarly we can test the output with different values, but the accurate result will be for the value given by the elbow point graph i.e; five in this dataset.

Chapter 8

RESULTS AND DISCUSSIONS

Efficiency of the Proposed System

As this method does not require to have a target, it can be of great help in the exploratory phase of customer segmentation. Each cluster is represented by its center, corresponding to the mean of elements assigned to the cluster. This technique is easy to implement and yet it allowed us to gather tons of information from the data.

Benefits of customer segmentation

By enabling companies to target specific groups of customers, a customer segmentation model allows for the effective allocation of marketing resources and the maximization of cross- and up-selling opportunities. When a group of customers is sent personalized messages as part of a marketing mix that is designed around their needs, it's easier for companies to send those customers special offers meant to encourage them to buy more products.

Customer segmentation can also improve customer service and assist in customer loyalty and retention. As a by-product of its personalized nature, marketing materials sent out using customer segmentation tend to be more valued and appreciated by the customer who receives them as opposed to impersonal brand messaging that doesn't acknowledge purchase history or any kind of customer relationship.

Results

.

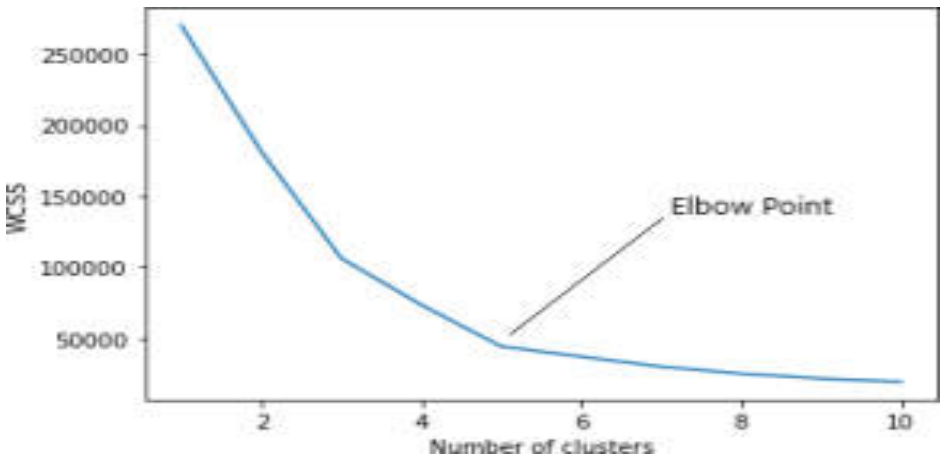


Figure 8.1: **Shows K value selection**

This is the elbow point graph which shows the k-value selection. For the above-given graph, the optimum value for K would be 5.

As we can see that with an increase in the number of clusters the WCSS value decreases. We select the value for K on the basis of the rate of decrease in WCSS. For example, from cluster 1 to 2 to 3 in the above graph we see a sudden and huge drop in WCSS. After 5 the drop is minimal and hence we chose 5 to be the optimal value for K.

One major drawback of K-Means clustering is the random initialisation of centroids. The formation of clusters is closely bound by the initial position of a centroid. The random positioning of the centroids can completely alter clusters and can result in a random formation.

The solution is K-means++. K-Means++ is an algorithm that is used to initialise the K-Means algorithm. The algorithm is as follows:

1. Choose one centroid uniformly at random from among the datapoints.
2. For each data point say x, compute D(x), which is the distance between x and the nearest centroid that has already been chosen.
3. Choose one new data point at random as a new centroid, using a weighted probability distribution where a point x is chosen with probability proportional to D(x)2.
4. Repeat Steps 2 and 3 until K centres have been chosen.
5. Proceed with standard k-means clustering.

Figure 8.2: **Output cluster**

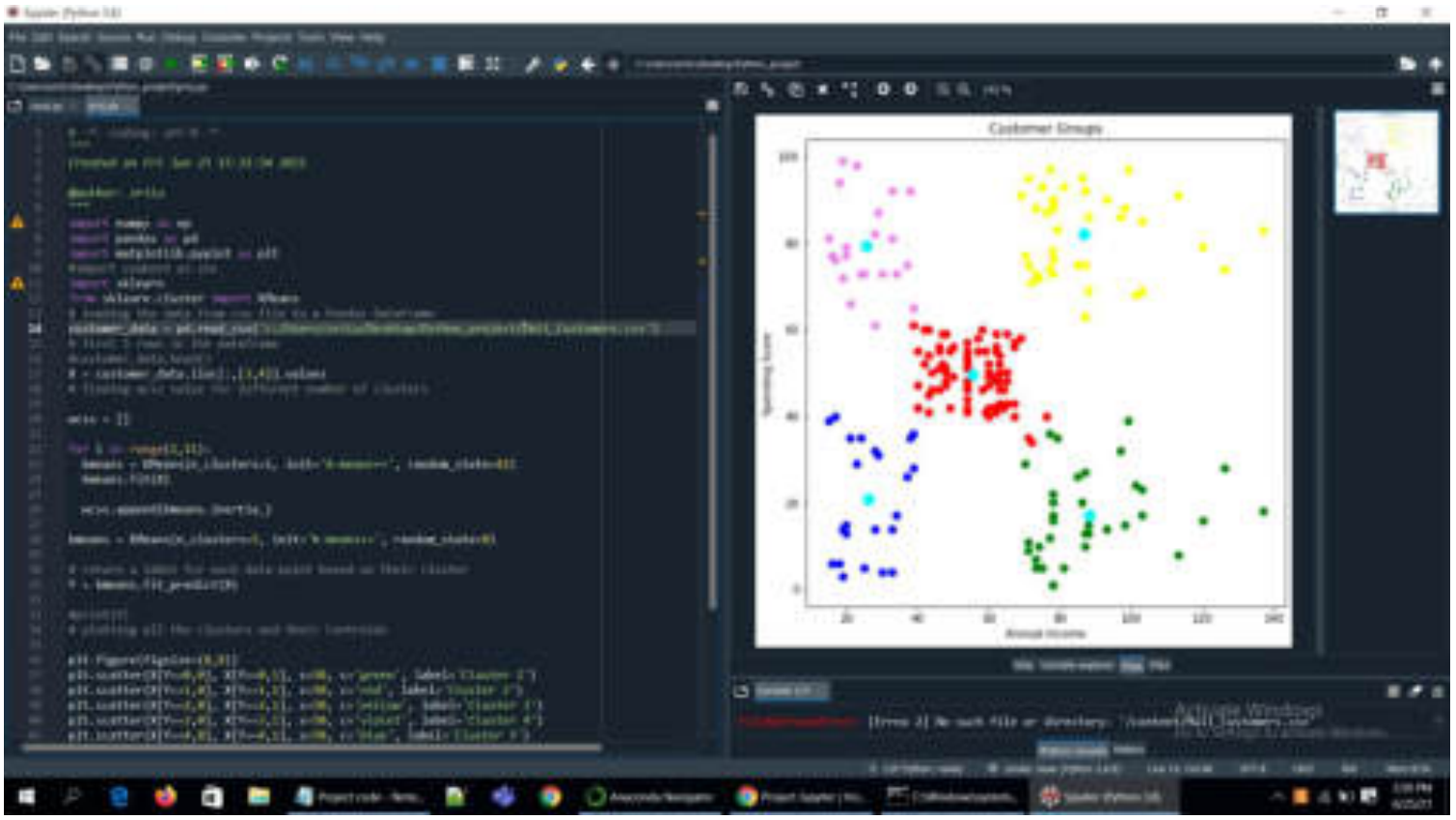


Figure 8.3: **Result Screenshot**

CHAPTER 9

APPLICATIONS

K-Means algorithm is very popular and used in a variety of applications such as market segmentation, document clustering, image segmentation and image compression, etc. The goal usually when we undergo a cluster analysis is either:

1. Get a meaningful intuition of the structure of the data we're dealing with.

2. Cluster-then-predict where different models will be built for different subgroups if we believe there is a wide variation in the behaviors of different subgroups. An example of that is clustering patients into different subgroups and build a model for each subgroup to predict the probability of the risk of having heart attack.

Demographic segmentation: It is one of the most popular and commonly used types of market segmentation. It refers to statistical data about a group of people. Demographic segmentation groups consumers by information you might see in a Census, such as: Age, gender, income, education, nationality, family size, etc.

Behavioral/Psychographic: These types of segmentation commonly look at lifestyle and personality characteristics that influence purchasing and consumption decisions.

Geographic: Geographic segmentation groups consumers based on location. This might include examining consumers by country, region or city, or taking a deeper dive to determine if they are in an urban, rural, or suburban area.

Combination: Often times, it makes sense to use multiple types of segmentation to capture the best picture of your customer. When there is an absence of customer data, businesses may turn to reviewing demographic and geographic information to get a basic understanding of their best customers, or work with an analytics partner to leverage enhanced customer segmentation insights.

Chapter 10

CONCLUSION AND FUTURE ENHANCEMENTS

Conclusion

This proposed algorithm is mainly useful to divide the whole customer base into the groups of the customers based on the similarities. The customer base that used in the project is Mall customers dataset. This dataset is Tabular data consist of rows. Each row is consider as customer.

By using the K-means Algorithm dividing the customers into the segments or groups of the customers. Customer segmentation helps to make business decisions based on the segments of the customers. Every segment has a specific behavior. Business decision making will vary from behavior to behavior. Because of this algorithm consistent and effective business strategies will improve business of an organization.

Future Enhancements

This algorithm can further be implemented to know various business patterns and thus can help to make the business strategy more efficient and reliable.

REFERENCES

[1] Raquel Florez-Lopez.,Juan Manuel Ramon-Jeronimo (2009).”Marketing Segmentation Through Machine Learning Models An Approach Based on Customer Relationship Management and Customer Profitability Account- ing. Social Science Computer Review”, 27(1), 96-117.

[2] Mahboubeh Khajvand , Kiyana Zolfaghar , Sarah Ashoori , Somayeh Al- izadeh . (2010). ”Estimating customer lifetime value based on RFM analy- sis of customer purchase behavior: case study”. Procedia Computer Science 3 (2011) 57–63.

[3]K. Maheswari. (2019).” Finding Best Possible Number of Clusters using K- Means Algorithm.International Journal of Engineering and Advanced Tech- nology (IJEAT) ISSN: 2249 – 8958”, Volume-9 Issue-1S4.

[4]Sahar F. Sabbeh(2018). ”Machine-Learning Techniques for Customer Re- tention: A Comparative Study. International Journal of Advanced Computer Science and Applications”, Vol. 9, No. 2,.

[5]J. du Toit, R. Davimes, A. Mohamed, K. Patel, and J. M. Nye. (2016).” Cus- tomer Segmentation Using Unsupervised Learning on Daily Energy Load Profiles,Journal of Advances in Information Technology”, Vol. 7, No. 2,.

[6]Puwanenthiren Premkanth, ”Market Segmentation and Its Impact on Cus- tomer Satisfaction with Especial Reference to Commercial Bank of Ceylon PLC. Global Journal of Management and Business Research Publisher:

Global Journals Inc. 2012. Print ISSN: 0975-5853”, Volume 12 Issue 1

[7]T.Nelson Gnanaraj, Dr.K.Ramesh Kumar N.Monica.” Survey on mining clusters using new k-mean algorithm from structured and unstructured data. International Journal of Advances in Computer Science and Technology.
2007”. Volume 3, No.2 .

[8] Vaishali R. Patel1 and Rupa G. Mehta. ”Impact of Outlier Removal and Normalization Approach in Modified k-Means Clustering Algorithm. IJCSI International Journal of Computer Science Issues”, Vol. 8, Issue 5, No 2, September 2011 ISSN : 1694-0814 .

[9] Zenthimal V (1988)” Customer perception of price quality and valuek means end model journal of marketing”, vol 2, pp 2-22 .

[10] Baker, W.E. And Sinkula, J.M. (1999) ”The Synergistic Effect Of Mar- ket Orientation And Learning Orientation On Organizational Performance, Journal Of Academy Of Marketing Science”, Vol. 27 .

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"GENDER DETECTION WITH OPENCV"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
COMPUTER SCIENCE ENGINEERING

Submitted by

G.GEETHA (17H71A0569)
K.PAVANI (17H71A0591)
Y.SRAVAN KUMAR(17H71A05B0)
B.LAKSHMI (17H71A0580)

Under the Esteemed Guidance of
Mrs.V.SRI LAKSHMI
Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in
2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"Gender Detection with Opencv"** is a bonafide work carried out by **Geetha.G (17H71A0569), Sravan Kumar.Y (17H71A05B0), Pavani.K (17H71A0591), Lakshmi.B(17H71A0580)** in partial fulfillment for the award of degree of Bachelor of Technology in **Computer science and Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

V. Sri Lakshmi
Project Guide

D.Prasad
Head of the Department

Examiner

Dr. K. Srinivas
Principal

ACKNOWLEDGEMENT

We are grateful to **Mrs. V. SRI LAKSHMI, Associate Professor, Department of Computer Science and Engineering,** whose guidance has been immense all throughout the project work.

Our sincere thanks to **Mr. D. PRASAD, Head of the Department, ComputerScience and Engineering,** for his valuable guidance and timely advice in the completion of our project.

We express our heartfelt gratitude to **Dr. K. SRINIVAS, Principal, DVR & Dr.HS MIC College of Technology**, for permitting us to take up our project work and to complete the project successfully.

We would like to express our sincere and heart full thanks to all the lecturers of the department for their continuous cooperation, which has given us the cogency to build up adamant aspiration over the completion of our project.

Finally, we thank one and all who directly and indirectly helped us to complete our project successfully.

DECLARATION

We carried out the project on “GENDER DETECTION WITH OPENCV”, hereby declared that the matter included in this project is a genuine work done by us and has been submitted to this institute for the fulfillment of the requirement of the degree.

we also declare that this project is a result of our own effort and that has not been copied from anyone and we have taken only citations from the sources which are mentioned in the references.

PROJECT TEAM MEMBERS

G. GEETHA

Y. SRAVAN KUMAR

K. PAVANI

B. LAKSHMI.

ABSTRACT

Humans are capable of determining an individual's gender relatively easily using facial attributes. Although it is challenging for machines to perform the same task, in the past decade incredible strides have been made in automatically making prediction from face image. The project identifies or detects the gender from the given face images. The tools used involve Convolutional Neural Network along with programming language like Python. The project has been motivated by problems like lack of security, frauds, child molestation, robbery, criminal identification.

CONTENT

LIST OF FIGURES

LIST OF TABLES

1.INTRODUCTION

Automatically predicting demographic information such as gender from face images is becoming increasingly significant for law enforcement and intelligence agencies. Humans are capable of determining an individual's gender relatively easily using facial attributes. Although it is challenging for machines to perform the same task, in the past decade incredible strides have been made in automatically predicting gender from the face. The complexity of predicting demographic information depends on the type of demographic category being predicted and the availability of adequate dataset. The project involves identifying or detecting the gender from the given face images. The project uses Deep Learning Technology where Convolutional Neural Network(CNN) acts as classifier. CNN is used in applications where both classification speed and maximum accuracy is considered important unlike Neural Networks which focuses on classification speed.

This python project enables us to detect the gender of a person. Computer vision help us to study the patterns and provide the result .but the views of computer limit itself to study the high definition characteristics of human beings. There are various methods to solve the problem of accurately estimating gender of a person. The first approach relies on manually extracting features such as size of the head ,position of eyes, length of nose. Another approach is based on end-to –end deep learning models. The two methods can be combined to form a mixed approach. most modern methods use deep learning approach.

To train gender models it is necessary to have labeled datasets. There are many publicly available datasets with gender annotations. another option is to create datasets manually by labeling face images. The motivation behind this thesis was to build an application for gender classification using a model that is useful for real life predictions. In this thesis we focus on deep learning end-to-end methods. The convolutional neural networks algothims are used for model training and prediction.

2.SYSTEM ANALYSIS

2.1 EXISTING SYSTEM:

The traditional way of gender detection is done through human interaction. By using facial attributes humans determine the gender. This process of identification is a slow process and need large amount of manual interaction in case of large data sets.

DISADVANTAGES OF EXISTING SYSTEM:

1. When the data is very big in number it is not possible to classify at a required short period of time

2. It is a slow and time consuming process.

2.2 PROPOSED SYSTEM:

1. The proposed approach is used to detect gender without human intervention.
2. The whole process of the Framework for the method is explained in figure 1. The entire process is classified into four stages had been a, namely data acquisition, data pre-processing, Build the machine learning model, and predict the gender of person in the image with the trained model.

2.3 ADVANTAGES OF PROPOSED SYSTEM:

1. It can be used for surveillance at banks and residential areas
2. This can be used to overcome the frauds that can take place during voting i.e.,can be used for voter identification.

3.REQUIRMENT ANALYSIS

3.1 PRELIMINARY INVESTIGATION

The first and foremost strategy for development of a project starts from the thought of designing a mail enabled platform for a small firm in which it is easy and convenient of sending and receiving messages, there is a search engine, address book and also including some entertaining games. When it is approved by the organization and our project guide the first activity, i.e. preliminary investigation begins. The activity has three parts:

- Request Clarification
- Feasibility Study
- Request Approval

3.2 REQUEST CLARIFICATION

After the approval of the request to the organization and project guide, with an investigation being considered, the project request must be examined to determine precisely what the system requires.

Here our project is basically meant for users within the company whose systems can be interconnected by the Local Area Network(LAN). In today's busy schedule man need everything should be provided in a readymade manner. So taking into consideration of the vast use of the net in day to day life, the corresponding development of the portal came into existence.

3.3 FEASIBILITY ANALYSIS

An important outcome of preliminary investigation is the determination that the system request is feasible. This is possible only if it is feasible within limited resources and time. The different feasibilities that have to be analyzed are

- Operational Feasibility
- Economic Feasibility
- Technical Feasibility

OPERATIONAL FEASIBILTY

Operational Feasibility deals with the study of prospects of the system to be developed. This system operationally eliminates all the tensions of the Admin and helps him in effectively tracking the project progress. This kind of automation will surely reduce the time and energy, which previously consumed in manual work. Based on the study, the system is proved to be operationally feasible.

ECONOMIC FEASIBILTY

Economic Feasibility or Cost-benefit is an assessment of the economic justification for a computer based project. As hardware was installed from the beginning & for lots of purposes thus the cost on project of hardware is low. Since the system is a network based, any number of employees connected to the LAN within that organization can use this tool from at any time. The Virtual Private Network is to be developed using the existing resources of the organization. So the project is economically feasible.

TECHNICAL FEASIBILTY

According to Roger S. Pressman, Technical Feasibility is the assessment of the technical resources of the organization. The organization needs IBM compatible machines with a graphical web browser connected to the Internet and Intranet. The system is developed for platform Independent environment. Java Server Pages, JavaScript, HTML, SQL server and WebLogic Server are used to develop the system. The technical feasibility has been carried out. The system is technically feasible for development and can be developed with the existing facility.

3.4 FUNCTIONAL REQUIRMENTS

In software engineering, a **functional requirement** defines a function of a software system or its component. A function is described as a set of inputs, the behavior, and outputs (see also software). Functional requirements may be calculations, technical details, data manipulation and processing and other specific functionality that define what a system is supposed to accomplish. Behavioral requirements describing all the cases where the system uses them

are expressed in the form **system shall do <requirement>**. The plan for implementing functional requirements is detailed in the system design. In requirements engineering, functional requirements specify particular results of a system. Functional requirements drive the application architecture of a system. A requirements analyst generates use cases after gathering and validating a set of functional requirements. The hierarchy of functional requirements is: **user/stakeholder request -> feature -> use case -> business rule.**

Functional requirements drive the application architecture of a system. A requirements analyst generates use cases after gathering and validating a set of functional requirements. Functional requirements may be technical details, data manipulation and other specific functionality of the project is to provide the information to the user.

3.5 NON FUNCTIONAL REQUIRMENTS

In systems engineering and requirements engineering, a **non-functional requirement** is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific behaviors

The project non-functional requirements include the following.

- Updating Work status.
- Problem resolution.
- Error occurrence in the system.
- Customer requests.

Availability: A system's "availability" or "uptime" is the amount of time that is operational and available for use. It's related to is the server providing the service to the users in displaying images. As our system will be used by thousands of users at any time our system must be available always. If there are any cases of updation they must be performed in a short interval of time without interrupting the normal services made available to the users.

Efficiency: Specifies how well the software utilizes scarce resources: CPU cycles, disk space, memory, bandwidth etc. All of the above mentioned resources can be effectively used by performing most of the validations at client side and reducing the workload on server by using

JSP instead of CGI which is being implemented now.

during the design, development, testing and deployment of the system. New modules can be easily integrated to our system without disturbing the existing modules or modifying the logical database schema of the existing applications.

Portability: Portability specifies the ease with which the software can be installed on all necessary platforms, and the platforms on which it is expected to run. By using appropriate server versions released for different platforms our project can be easily operated on any operating system, hence can be said highly portable.

Scalability: Software that is scalable has the ability to handle a wide variety of system configuration sizes. The nonfunctional requirements should specify the ways in which the system may be expected to scale up (by increasing hardware capacity, adding machines etc.). Our system can be easily expandable. Any additional requirements such as hardware or software which increase the performance of the system can be easily added. An additional server would be useful to speed up the application.

Integrity: Integrity requirements define the security attributes of the system, restricting access to features or data to certain users and protecting the privacy of data entered into the software. Certain features access must be disabled to normal users such as adding the details of files, searching etc which is the sole responsibility of the server. Access can be disabled by providing appropriate logins to the users for only access.

Usability: Ease-of-use requirements address the factors that constitute the capacity of the software to be understood, learned, and used by its intended users. Hyperlinks will be provided for each and every service the system provides through which navigation will be easier. A system that has high usability coefficient makes the work of the user easier.

Performance: The performance constraints specify the timing characteristics of the software.

4.TECHNOLOGIES USED IN THE PROJECT

What things you need to install the software and how to install them:

1. Python3.6
    - This setup requires that your machine has python 3.6 installed on it. you can refer to this url https://www.python.org/downloads/ to download python. Once you have python downloaded and installed, you will need to set up PATH variables (if you want to run a python program directly, detailed instructions are below in *how to run the software section*). To do that check this: https://www.pythoncentral.io/add-python-to-path-python-is-not-recognized-as-an-internal-or-external-command/.
    - Setting up PATH variable is optional as you can also run the program without it and more instructions are given below on this topic.
2. Second and easier option is to download jupyter notebook on command prompt using the command pip install jupyter notebook.
3. You will also need to download and install below packages after you install python and jupyter notebook from the steps above

Packages used are:

- Tensorflow
- Keras
- opencv

- Use below commands in command prompt to install packages.
- pip install tensor flow

  pip install num py

  pip install keras

Computer Vision

It is a field that includes processing, analyzing and understanding images in general high dimensional data from the real world in order to produce numerical and symbolic information or it is a technology of science and machines that see it obtain information from images.

Deep Learning

Deep learning is a powerful set of techniques for learning using neural networks. Neural network is a beautiful biologically inspired program paradigm which enables a computer to learn from data.These are learning algorithms.

Neural Network

Neural networks reflect the behavior of the human brain, allowing computer programs to recognize patterns and solve common problems in the fields of AI, machine learning, and deep learning. Neural networks, also known as artificial neural networks (ANNs) or simulated neural networks(SNNs), are a subset of machine learning and are at the heart of deep learning algorithms. Their name and structure are inspired by the human brain, mimicking the way that biological neurons signal to one another.

The three important types of neural networks that form the basis for most pre-trained models in deeplearning:

1. Artificial Neural Networks(ANN)
2. Convolution Neural Networks(CNN)
3. Recurrent Neural Networks(RNN)

Convolutional Neural Network

Pioneer of Convolutional Neural Network is Yann Lecun.He is the director of Facebook AI Research Group. He Built the first convolutional Neural Network called LeNet in1988.LeNetwas used for character recognition tasks like reading zip codes and digits.

CNN also known as ConvNet. A convolutional neural network is a feed-forward neural network that is generally used to analyze visual images by processing data with grid-like topology. A convolutional neural network is used to detect and classify objects in an image. In CNN,every image is represented in the form of an array of pixel values.

Layers in a Convolutional Neural Network

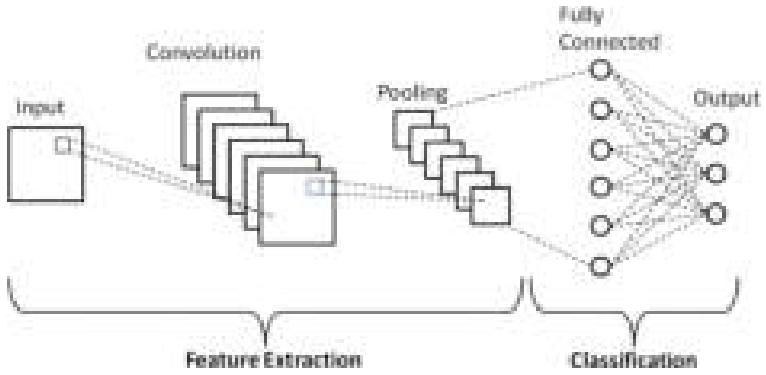
A convolution neural network has multiple hidden layers that help in extracting information from an image.The four important layers in CNN are:

1.Convolution Layer

2.ReLU Layer

3.Pooling Layer

4.Fully Connected



Input layer:

It Accepts the pixels of the Image as input in the form of arrays.

Hidden Layers:

Carry out features extraction by performing certain calculations & manipulations. These are multiple hidden layers like convolution layer, ReLU layer, Pooling layer etc. that performs feature extraction.

Convolution Layer:

This is the first step in the process of extracting valuable features from an image. This layer uses a matrix filter and performs convolution operation to detect patterns in the image. A convolution layer has several filters that perform the convolution operation. Every image is considered as a matrix of pixel values. Consider the following 5x5 image whose pixel values are either 0 or 1.There's also a filter matrix with a dimension of 3x3.

ReLU Layer:

ReLU stands for the rectified linear unit. Once the feature maps are extracted, the next step is to move them to a ReLU layer. ReLU performs an elementwise operation and sets all the negative pixels to 0.It introduces non-linearity to the network, and the generated output is a rectified

feature map. Its activation function is applied to the convolution layer to get a rectified feature map of the image.Below is the graph of a ReLU function:

Pooling Layer:

Pooling is a down-sampling operation that reduces the dimensionality of the feature map. The rectified feature map now goes through a pooling layer to generate a pooled feature map.Also uses multiple filters to detect edges, corners, eyes, features, etc. Different kind of pooling operations can be applied out of which some are explained below.

Fully Connected Layer(FC):

Fully Connected layers in a neural network are those layers where all the inputs from one layer are connected to every activation unit of the next layer. The output from the convolutional layers represents high-level features in the data. While that output could be flattened and connected to the output layer, adding a fully-connected layer is a (usually) cheap way of learning non-linear combinations of these features. The fully connected layer operates on a flattened input where each input is connected to all neurons.

OpenCV

OpenCV is a cross-platform library using which we can develop real-time computer vision applications. It mainly focuses on image processing, video capture and analysis including features like face detection and object detection.All the OpenCV array structures are converted to and from Numpy arrays. This also makes it easier to integrate with other libraries that use Num py such as Sci Py and Mat plot lib. In this, all OpenCV data types are preserved as such.Later, OpenCV came with both cv and cv2. Now, in the latest releases, there is only the cv2 module,and cv is as ub class inside cv2 .You need to call import cv2. cv as cv to access it.

Tensor Flow

Tensor Flow is a mathematical computation library for training and building you machine learning and deep learning model with simple to use high level APIs. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML-powered applications. Tensor Flow was originally developed by researchers and engineers working on the Google Brain team.

5.SYSTEM REQUIREMENTS SPECIFICATIONS

SOFTWARE REQUIREMENTS

- Software: Python IDLE
- Version:3.6
- Numpy, Tensor flow, keras
- Operating System: windows

HARDWARE REQUIREMENTS

- Processor: IntelcoreI5
- Ram2 :GB
- Hard Disk:500GB or More
- A simple enhanced windows keyboard

6.SYSTEM DESIGN

OVERVIEW OF SYSTEM DESIGN:

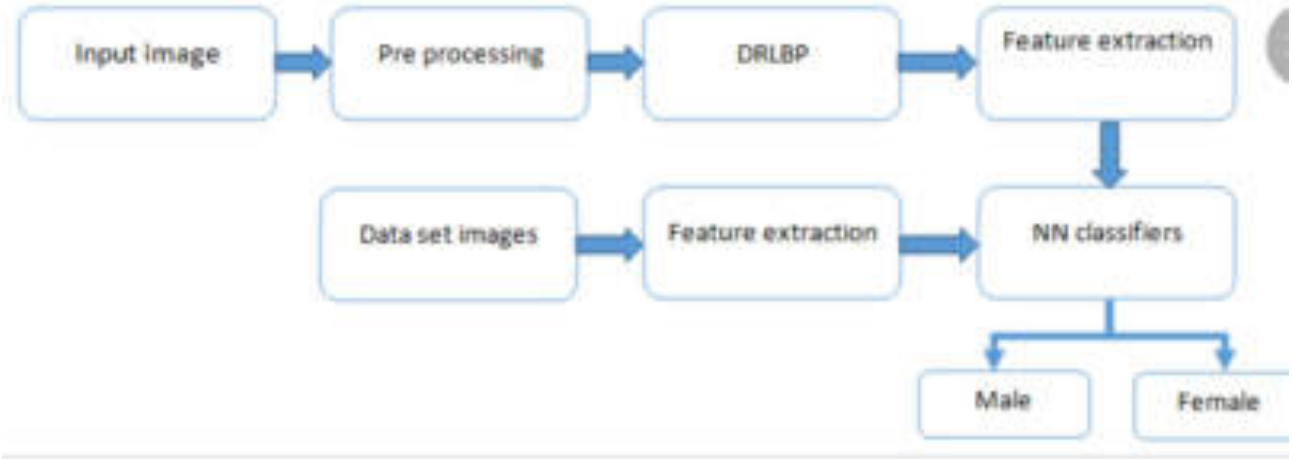


Fig 1: overview of system design

BLOCK DIAGRAM:

Fig 2:Block diagram

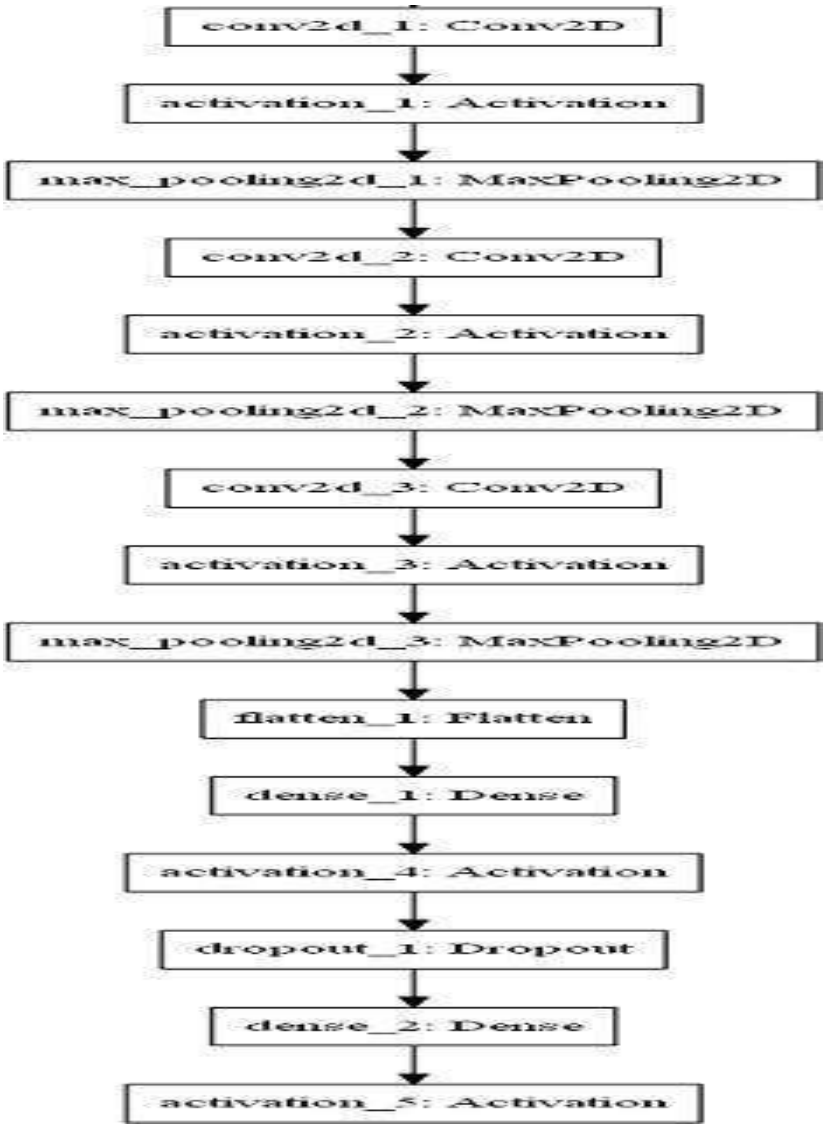


Fig 3: Model for training data

As shown in Fig. 2. The input given is in the form of a face image to the pre-processing unit. The pre-processing unit analyzes the image features based on the algorithms. Data preprocessing for machine learning is a technique that is used to convert the raw data into a clean data set i.e whenever the data is gathered from different sources, it is collected in raw format and raw format is not appropriate for the analysis. After preprocessing the data, the model is trained using this clean dataset. An unknown image is then inputted to predict the gender of the unknown image.

As shown in Fig. 2.The model consists of five stacks of layers. The layers include Convolutional layer, Activation layer, Max Pooling layer, Flatten layer, Dense layer and Dropout layer. The first three stacks consist of Convolutional layer, Activation layer and Max Pooling layer. The activation layer used is Rectified Linear Unit(ReLu). The fourth stack consists of the Flatten layer, Dense layer and Activation layer. After the first three layers, the output is in 3D and needs to be converted to 1D form. The Flatten layer is used to convert the 3D output to 1D format. The Dense layer is also known as Fully Connected layer. It is used to convert the matrix into a list form at and all the nodes are connected to each other. The Activation layer used is Rectified Linear Unit(ReLu).The fifth stack consists of Dropout layer, Dense layer and Activation layer

Dropout layer is used to remove the duplicate images present in the data set to avoid systems undergoing over-fitting. The Activation layer used is Sigmoid. The sigmoid classification is best for binary classification problems such as gender. Generally, the network is trained using a larger and domain related dataset. After the convergence of the network parameters, an extra training step is done to optimize the network weights using in-domain data. This allows the system to apply convolutional neural networks on small training sets.

7. DESIGN DIAGRAMS

Scenario Based Modeling:

Use-oriented techniques are widely used in software requirement analysis and design. Use cases and usage scenarios facilitate system understanding and provide a common language for communication. This paper presents a scenario-based modeling technique and discusses its applications. In this model, scenarios are organized hierarchically and they capture the system functionality at various abstraction levels including scenario groups, scenarios, and sub-scenarios. Combining scenarios or sub-scenarios can form complex scenarios. Data are also separately identified, organized, and attached to scenarios. This scenario model can be used to cross check with the UML model. It can also direct systematic scenario-based testing including test case generation, test coverage analysis with respect to requirements, and functional regression testing.

Use Case Diagram:

A use case describes a sequence of actions that provide something of measurable value to an actor and is drawn as a horizontal ellipse.

Actors: An actor is a person, organization, or external system that plays a role in one or more interactions with the system.

System boundary boxes (optional):

A rectangle is drawn around the use cases, called the system boundary box, to indicate the scope of the system. Anything within the box represents functionality that is in scope and anything outside the box is not.

Four relationships among use cases are used often in practice.

a. **Include:**

In one form of interaction, a given use case may include another. "Include is a Directed Relationship between two use cases, implying that the behavior of the included use case is inserted into the behavior of the including use case.

The first use case often depends on the outcome of the included use case. This is useful for extracting truly common behaviors from multiple use cases into a single description. The notation is a dashed arrow from the including to the included use case, with the label "«include»". There are no parameters or return values. To specify the location in a flow of events in which the base use case includes the behavior of another, you simply write include followed by the name of the use case you want to include, as in the following flow for track order.

b. Extend:

In another form of interaction, a given use case (the extension) may extend another. This relationship indicates that the behavior of the extension use case may be inserted in the extended use case under some conditions. The notation is a dashed arrow from the extension to the extended use case, with the label "«extend»". Modelers use the «extend» relationship to indicate use cases that are "optional" to the base use case.

c. Generalization:

In the third form of relationship among use cases, a generalization/specialization relationship exists. A given use case may have common behaviors, requirements, constraints, and assumptions with a more general use case. In this case, describe them once, and deal with it in the same way, describing any differences in the specialized cases. The notation is a solid line ending in a hollow triangle drawn from the specialized to the more general use case following the standard generalization notation.

d. Associations:

Associations between actors and use cases are indicated in use case diagrams by solid lines. An association exists whenever an actor is involved with an interaction described by a use case. Associations are modeled as lines connecting use cases and actors to one another, with an optional arrowhead on one end of the line. The arrowhead is often used to indicate the direction of the initial invocation of the relationship or to indicate primary actor.

Identified Use Cases

The “user model view” encompasses a problem and solution from the preservative of those individuals whose problem the solution addresses. The view presents the goals and objectives of the problem owners and their requirements of the solution. This view is composed of “use case diagrams”. These diagrams describe the functionality provided by a system to external integrators. These diagrams contain actors, use cases, and their relationships.

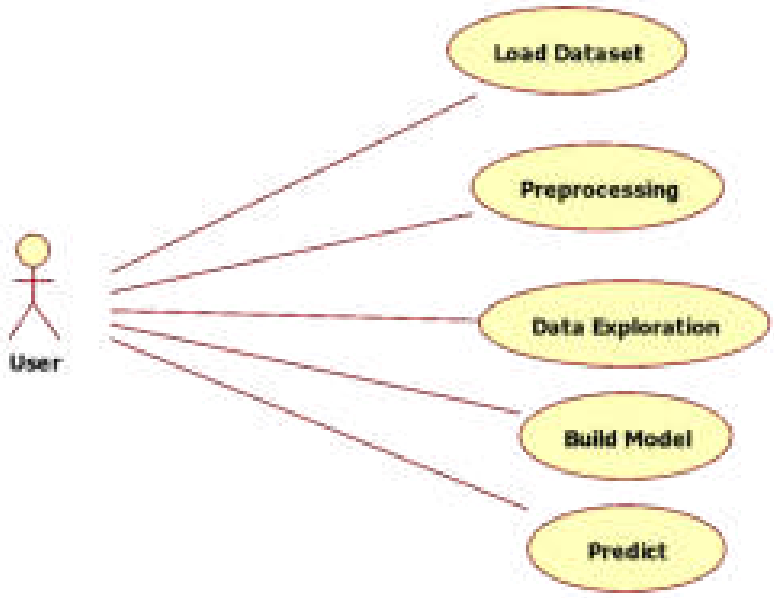


Fig: 4.1 Use case diagram

Class diagram:

Model objects are entities that combine state (i.e., data), behavior (i.e., procedures, or methods) and identity (unique existence among all other objects). The structure and behavior of an object are defined by a class, which is a definition, or blueprint, of all objects of a specific type. An object must be explicitly created based on a class and an object thus created is considered to be an instance of that class. An object is similar to a structure, with the addition of method pointers, member access control, and an implicit data member which locates instances of the class (i.e. actual objects of that class) in the class hierarchy (essential for runtime inheritance features)

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, and the relationships between the classes.

The class diagram is the main building block in object oriented modeling. It is used both for general conceptual modeling of the semantics of the application, and for detailed modeling translating the models into programming code. The classes in a class diagram represent both the main objects and or interactions in the application and the objects to be programmed. In the class diagram these classes are represented with boxes which contain the two parts:

- The upper part holds the name of the class.
- The middle part contains the attributes of the class.
- The lower part contains the operations of the class.

Fig: 4.2 Class diagram

Sequence Diagram:

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart.

Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams. A sequence diagram shows, as parallel vertical lines (lifelines), different processes or objects that live simultaneously, and, as horizontal arrows, the messages exchanged between them, in the order in which they occur. This allows the specification of simple runtime scenarios in a graphical manner. If the lifeline is that of an object, it demonstrates a role. Note that leaving the instance name blank can represent anonymous and unnamed instances. In order to display interaction, messages are used. These are horizontal arrows with the message name written above them. Solid arrows with full heads are synchronous calls, solid arrows with stick heads are asynchronous calls and dashed arrows with stick heads are return messages. This definition is true as of UML 2, considerably different from UML 1.x.

Activation boxes, or method-call boxes, are opaque rectangles drawn on top of lifelines to represent that processes are being performed in response to the message (Execution Specifications in UML).

Objects calling methods on themselves use messages and add new activation boxes on top of any others to indicate a further level of processing. When an object is destroyed (removed from memory), an X is drawn on top of the lifeline, and the dashed line ceases to be drawn below it (this is not the case in the first example though). It should be the result of a message, either from the object itself, oranother.

A message sent from outside the diagram can be represented by a message originating from a filled-in circle (found message in UML) or from a border of sequence diagram (gate in UML)

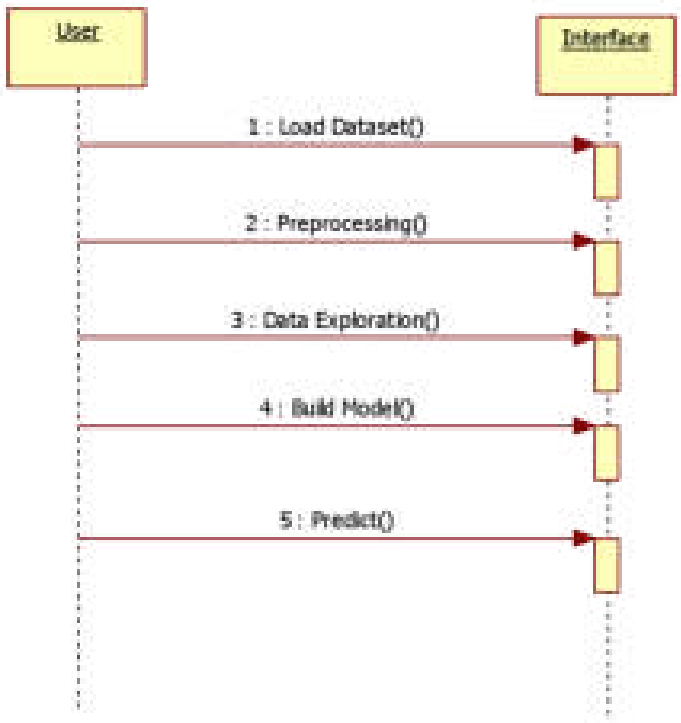


Fig: 4.3 Sequence diagram

4.1 Collaboration Diagram:

A Sequence diagram is dynamic, and, more importantly, is time ordered. A Collaboration diagram is very similar to a Sequence diagram in the purpose it achieves; in other words, it shows the dynamic interaction of the objects in a system. A distinguishing feature of a Collaboration diagram is that it shows the objects and their association with other objects in the system apart from how they interact with each other. The association between objects is not represented in a Sequence diagram.

A Collaboration diagram is easily represented by modeling objects in a system and representing the associations between the objects as links. The interaction between the objects is denoted by arrows. To identify the sequence of invocation of these objects, a number is placed next to each of these arrows.

Defining a Collaboration Diagram:

A sophisticated modeling tool can easily convert a collaboration diagram into a sequence diagram and the vice versa. Hence, the elements of a Collaboration diagram are essentially the same as that of a Sequence diagram.

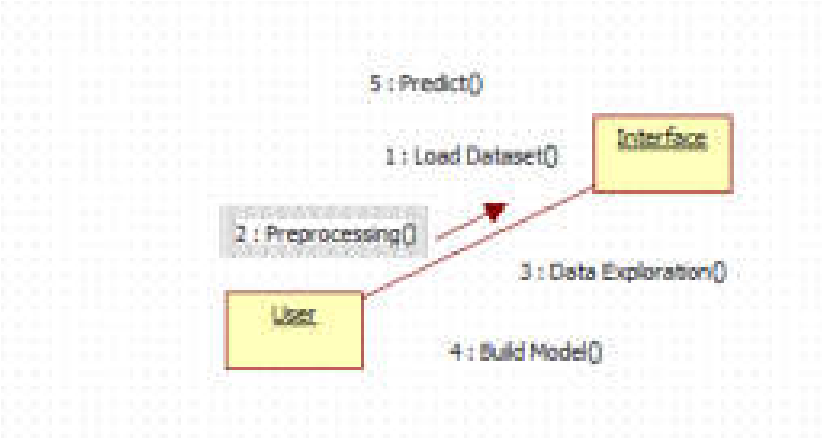


Fig: 4.4 Collaboration diagram

Flow-Oriented Modeling

4.2 Activity Diagram:

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control. Activity diagrams are constructed from a limited repertoire of shapes, connected with arrows. The most important shape types:

- diamonds represent decisions;
- bars represent the start (split) or end (join) of concurrent activities;
- a black circle represents the start (initial state) of the workflow;
- An encircled black circle represents the end (finalstate).

Arrows run from the start towards the end and represent the order in which activities happen. However, the join and split symbols in activity diagrams only resolve this for simple cases; the meaning of the model is not clear when they are arbitrarily combined with the decisions orloops.

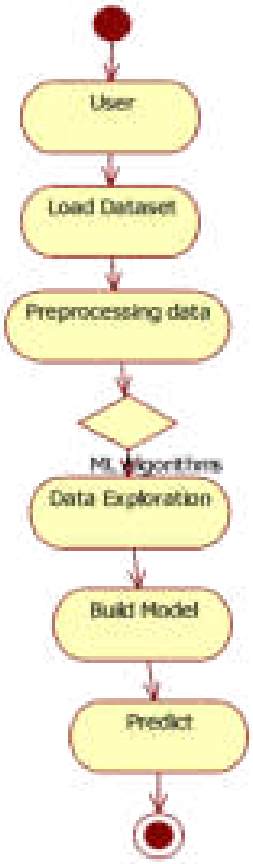


Fig: 4.5 Activity diagram

State Chart Diagram:

Objects have behaviors and states. The state of an object depends on its current activity or condition. A state chart diagram shows the possible states of the object and the

ransitions that cause a change in state. A state diagram, also called a state machine diagram or state chart diagram, is an illustration of the states an object can attain as well as the transitions between those states in the Unified Modeling Language. A state diagram resembles a flowchart in which the initial state is represented by a large black dot and subsequent states are portrayed as boxes with rounded corners. There may be one or two horizontal lines through a box, dividing it into stacked sections. In that case, the upper section contains the name of the state, the middle section (if any) contains the state variables and the lower section contains the actions performed in that state. If there are no horizontal lines through a box, only the name of the state is written inside it. External straight lines, each with an arrow at one end, connect various pairs of boxes. These lines define the transitions between states. The final state is portrayed as a large black dot with a circle around it. Historical states are denoted as circles with the letter H inside.

State diagram:

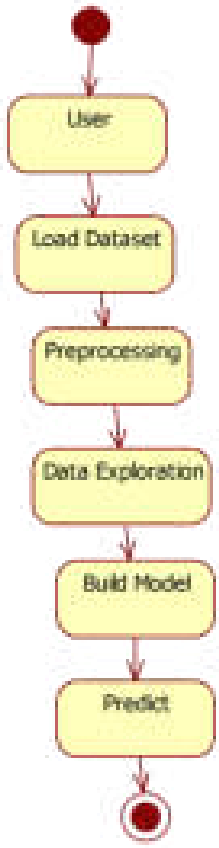


Fig 4.6 state diagram

8.IMPLEMENTATION AND SOURCE CODE

To start with the project, the first step that needs to be done is data collection. Datasets play an important in deep learning as it is used to train the system to get the required output. Some datasets are available publicly while some are not.The UTK Face is a publicly available dataset which can be used for gender classification. This dataset has images of people of different age, gender and ethnicity. In the project,the dataset images are input into the algorithm to identify the gender. The UTK Face dataset is used to train the model to perform gender classification. The input given is in the form of a face image and the image features are analyzed based on the algorithm.An unknown image is then inputted to predict the gender of the same.An output will be generated which will contain the gender prediction of the unknown face image. The UTK Face dataset is categorized into 'Train' and 'Validation', each of which contains 'Male'and 'Female'. The project is trained using 8,000 images in each class and validated using 1,000 images in each class.The model is trained using ConvNet (Convolutional Neural Network) consisting of 5 layers. The CNN consists of many hidden layers such as Convolutional layer, ReLu layer, Max Pooling layer, Fully Connected layer, etc. Using these layers,the input face image is converted into weights and saved in'.h5' format.These weights are then used to predict an unknown image.The average accuracy achieved in the project is 90%.

The training program includes data augmentation.Data augmentation means increasing the number of images in the dataset because plentiful high-quality information is the key to significant machine learning models. Foremost, training examples need to be augmented via a variety of random transformations, so that the model would never see twice the exact same picture and this helps prevent overfitting there by generalizing the model in a better way.

Fig 5:Data Augmentation Output

Convnet is the right tool for image classification. Data augmentation is a way to fight overfitting, however, it is not enough since the augmented samples are still highly correlated. The main focus for fighting overfitting must be the entropic capability i.e the abundant information that the model is allowed to store. A model which can store a lot of information has the potential to be more accurate by leveraging more features in comparison to a model that can only store a few features. But the former is also more At risk as it will start storing irrelevant features whereas a model that can only store a few features will have to focus on the most significant features found within the data. The one which stores fewer features are more likely to be truly relevant and are easy to generalize.The setup for the project is as follows:

- 16000 training examples(8000 per class)
- 2000 validation examples (1000 per class)

Training Dataset:The training dataset is a set of examples employed to train the model i.e. to fit the parameters. Most of the approaches used for training the samples tend to overfit if the dataset is not increased and used in variety.

Validation Dataset:A validation dataset is also called the 'development dataset'or'devset'and is used to fit the hyper parameters of the classifier. It is necessary to have a validation dataset along with training and test data set because it helps avoid over fitting. The ultimate goal is to choose a network performing the best on unseen data hence we use a validation dataset which is independent of the training dataset.

Test Dataset:The test dataset is not dependent on the training or validation dataset. If a model is fitting both the training dataset as well as test dataset then it can be said that minimum overfitting has taken place. The test dataset is the dataset which is only used to test the performance of the classifier or model.The test dataset is employed to check the performance characteristics like the accuracy, loss, sensitivity, etc.

Here the class, .flow from directory() is used to generate groups of image data(and their labels)directly from jpgs in the respective folders and these generators can then be used to train the model. An epoch is one complete representation of the dataset to be learned by the learning machine. One epoch is when the entire dataset is passed both forward and backward through the neural network only once.This approach gives a validation accuracy of 0.81-0.94 after 10 epochs.

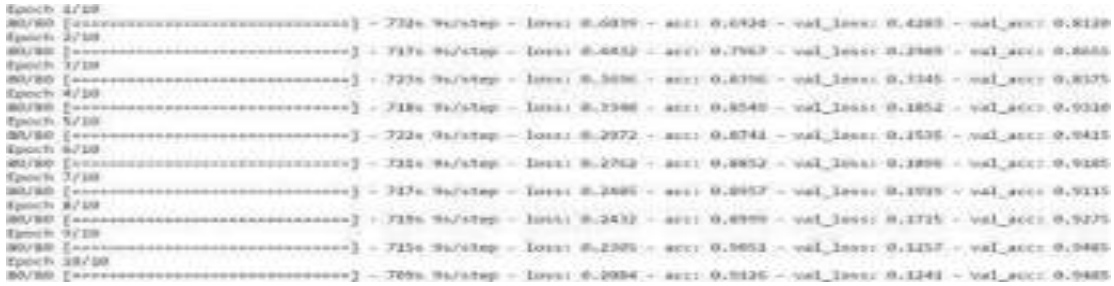


Fig-6:Accuracy and Loss in Each Epoch

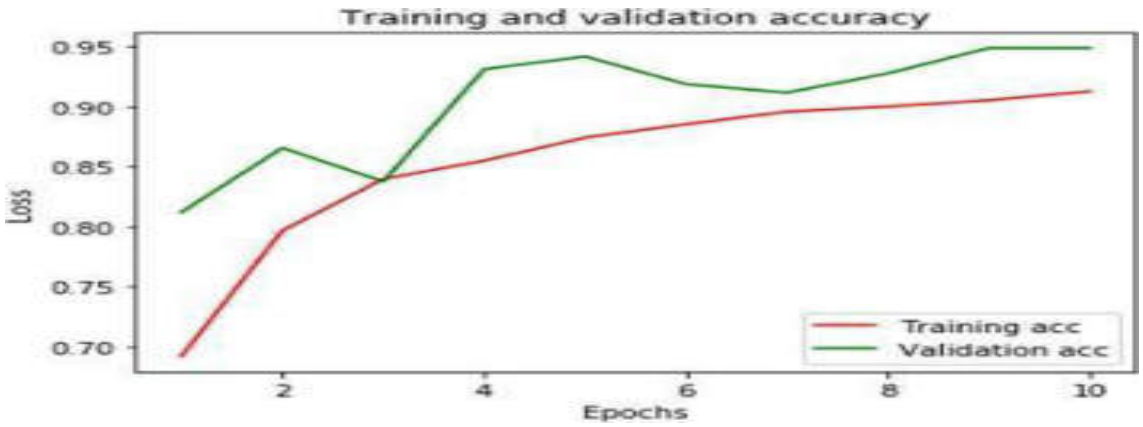


Fig-7:Training and Validation Accuracy Graph

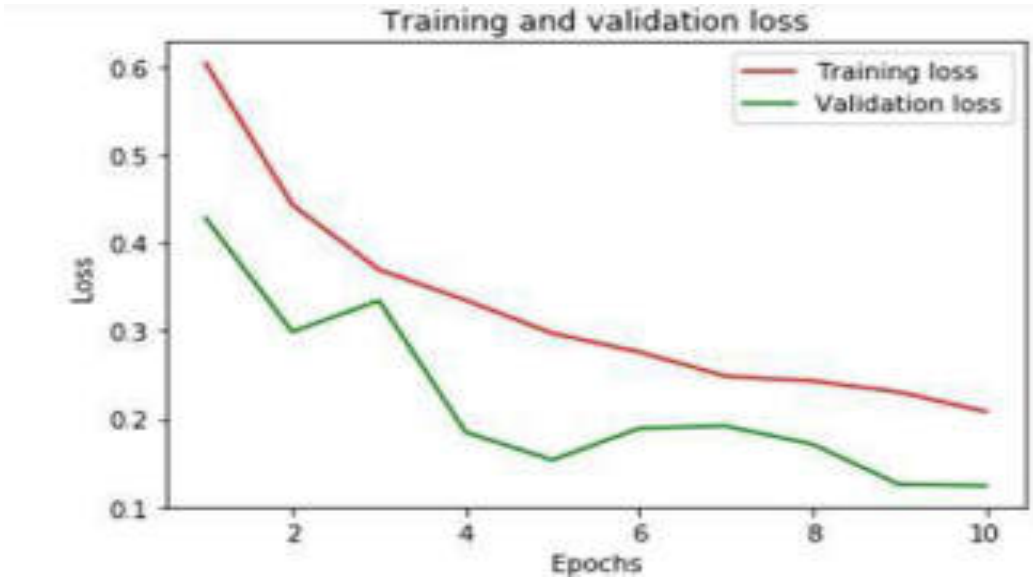


Fig-8:Training and Validation Loss Graph

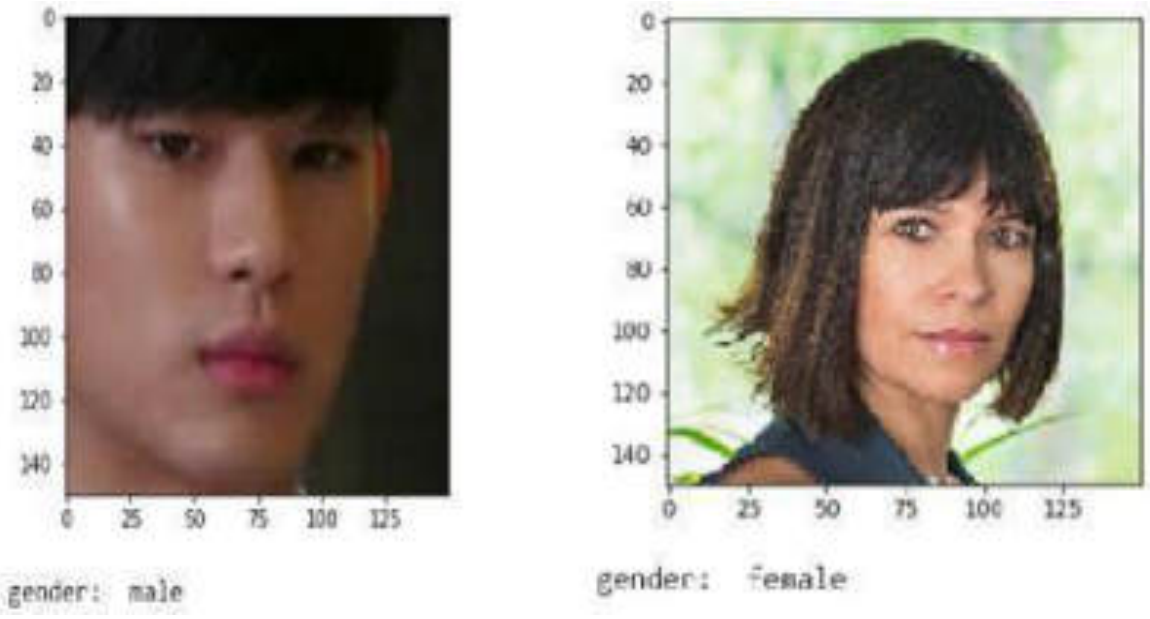


Fig-9:Output Prediction

SOURCE CODE

There will be 3 modules in the Project

1.Training the data

2.Applying CNN model

3.Gender Prediction by Using Webcam

1. Train.py || Preprocessing data

```
import matplotlib
matplotlib.use("Agg")
from keras.preprocessing.image import ImageDataGenerator
from keras.optimizers import Adam
from keras.preprocessing.image import img_to_array
from keras.utils import to_categorical
from keras.utils import plot_model
from sklearn.model_selection import train_test_split
from model.smaller vggnet import SmallerVGGNet
import matplotlib.pyplot as plt
import numpy as np
import argparse
import random
import cv2
import os
import glob

# handle command line arguments
ap = argparse.ArgumentParser()
ap.add_argument("-d", "--dataset", required=True,
        help="path to input dataset (i.e., directory of images)")
ap.add_argument("-m", "--model", type=str, default="gender_detection.model",
        help="path to output model")
ap.add_argument("-p", "--plot", type=str, default="plot.png",
        help="path to output accuracy/loss plot")
args = ap.parse_args()

# initial parameters
```

```
epochs = 100
lr = 1e-3
batch_size = 64
img_dims = (96,96,3)

data = []
labels = []

# load image files from the dataset
image_files = [f for f in glob.glob(args.dataset + "/**/*", recursive=True) if not os.path.isdir(f)]
random.seed(42)
random.shuffle(image_files)

# create ground-truth label from the image path
for img in image_files:

    image = cv2.imread(img)

    image = cv2.resize(image, (img_dims[0],img_dims[1]))
    image = img_to_array(image)
    data.append(image)

    label = img.split(os.path.sep)[-2]
    if label == "woman":
        label = 1
    else:
        label = 0

    labels.append([label])

# pre-processing
data = np.array(data, dtype="float") / 255.0
labels = np.array(labels)

# split dataset for training and validation
(trainX, testX, trainY, testY) = train_test_split(data, labels, test_size=0.2,
```

```
                                    random_state=42)
trainY = to_categorical(trainY, num_classes=2)
testY = to_categorical(testY, num_classes=2)

# augmenting data set
aug = ImageDataGenerator(rotation_range=25, width_shift_range=0.1,
                         height_shift_range=0.1, shear_range=0.2, zoom_range=0.2,
                         horizontal_flip=True, fill_mode="nearest")

# build model
model = SmallerVGGNet.build(width=img_dims[0], height=img_dims[1], depth=img_dims[2],
                            classes=2)

# compile the model
opt = Adam(lr=lr, decay=lr/epochs)
model.compile(loss="binary_crossentropy", optimizer=opt, metrics=["accuracy"])

# train the model
H = model.fit_generator(aug.flow(trainX, trainY, batch_size=batch_size),
                        validation_data=(testX,testY),
                        steps_per_epoch=len(trainX) // batch_size,
                        epochs=epochs, verbose=1)

# save the model to disk
model.save(args.model)

# plot training/validation loss/accuracy
plt.style.use("ggplot")
plt.figure()
N = epochs
plt.plot(np.arange(0,N), H.history["loss"], label="train_loss")
plt.plot(np.arange(0,N), H.history["val_loss"], label="val_loss")
plt.plot(np.arange(0,N), H.history["acc"], label="train_acc")
plt.plot(np.arange(0,N), H.history["val_acc"], label="val_acc")

plt.title("Training Loss and Accuracy")
```

```
plt.xlabel("Epoch #")
plt.ylabel("Loss/Accuracy")
plt.legend(loc="upper right")

# save plot to disk
plt.savefig(args.plot)
```

2.Smallervgnet.py || Training the Model using CNN

```
from keras.models import Sequential
from keras.layers.normalization import BatchNormalization
from keras.layers.convolutional import Conv2D
from keras.layers.convolutional import MaxPooling2D
from keras.layers.core import Activation
from keras.layers.core import Flatten
from keras.layers.core import Dropout
from keras.layers.core import Dense
from keras import backend as K


class SmallerVGGNet:
    @staticmethod
    def build(width, height, depth, classes):
        model = Sequential()
        inputShape = (height, width, depth)
        chanDim = -1

        if K.image_data_format() == "channels_first":
            inputShape = (depth, height, width)
            chanDim = 1
```

```
model.add(Conv2D(32, (3,3), padding="same", input_shape=inputShape))
model.add(Activation("relu"))
model.add(BatchNormalization(axis=chanDim))
model.add(MaxPooling2D(pool_size=(3,3)))
model.add(Dropout(0.25))

model.add(Conv2D(64, (3,3), padding="same"))
model.add(Activation("relu"))
model.add(BatchNormalization(axis=chanDim))
model.add(Conv2D(64, (3,3), padding="same"))
model.add(Activation("relu"))
model.add(BatchNormalization(axis=chanDim))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.25))

model.add(Conv2D(128, (3,3), padding="same"))
model.add(Activation("relu"))
model.add(BatchNormalization(axis=chanDim))
model.add(Conv2D(128, (3,3), padding="same"))
model.add(Activation("relu"))
model.add(BatchNormalization(axis=chanDim))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.25))

model.add(Flatten())
model.add(Dense(1024))
model.add(Activation("relu"))
model.add(BatchNormalization())
model.add(Dropout(0.5))

model.add(Dense(classes))
model.add(Activation("sigmoid"))

return model
```

3.Detect_gender_webcam.py || Detecting gender By using webcam

```
# import necessary packages
from keras.preprocessing.image import img_to_array
from keras.models import load_model
#from keras.utils import get_file
import numpy as np
import argparse
import cv2
import os
import cvlib as cv
# load model
model = load_model('gender_detection.model')

# open webcam
webcam = cv2.VideoCapture(0)

if not webcam.isOpened():
    print("Could not open webcam")
    exit()

classes = ['man','woman']

# loop through frames
while webcam.isOpened():

    # read frame from webcam
    status, frame = webcam.read()

    if not status:
        print("Could not read frame")
        exit()

    # apply face detection
    face, confidence = cv.detect_face(frame)

    print(face)
```

```
print(confidence)

# loop through detected faces
for idx, f in enumerate(face):

    # get corner points of face rectangle
    (startX, startY) = f[0], f[1]
    (endX, endY) = f[2], f[3]

    # draw rectangle over face
    cv2.rectangle(frame, (startX,startY), (endX,endY), (0,255,0), 2)

    # crop the detected face region
    face_crop = np.copy(frame[startY:endY,startX:endX])

    if (face_crop.shape[0]) < 10 or (face_crop.shape[1]) < 10:
        continue

    # preprocessing for gender detection model
    face_crop = cv2.resize(face_crop, (96,96))
    face_crop = face_crop.astype("float") / 255.0
    face_crop = img_to_array(face_crop)
    face_crop = np.expand_dims(face_crop, axis=0)

    # apply gender detection on face
    conf = model.predict(face_crop)[0]
    print(conf)
    print(classes)

    # get label with max accuracy
    idx = np.argmax(conf)
    label = classes[idx]

    label = "{}: {:.2f}%".format(label, conf[idx] * 100)

    Y = startY - 10 if startY - 10 > 10 else startY + 10
```

```
        # write label and confidence above face rectangle
        cv2.putText(frame, label, (startX, Y),  cv2.FONT_HERSHEY_SIMPLEX,
                0.7, (0, 255, 0), 2)

    # display output
    cv2.imshow("gender detection", frame)

    # press "Q" to stop
    if cv2.waitKey(1) & 0xFF == ord('q'):
        break
# release resources
webcam.release()
cv2.destroyAllWindows()
```

OUTPUT SCREEN:

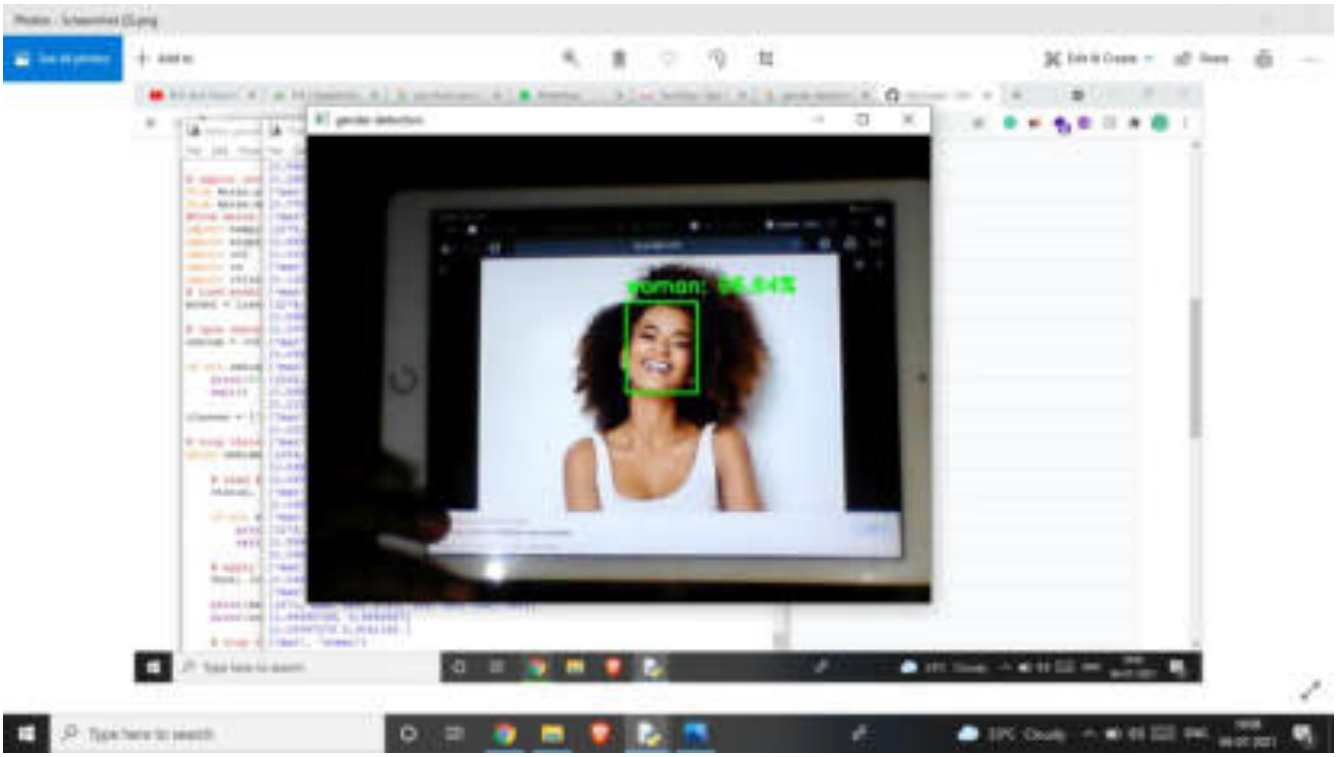


Fig 10.1:Output screen1

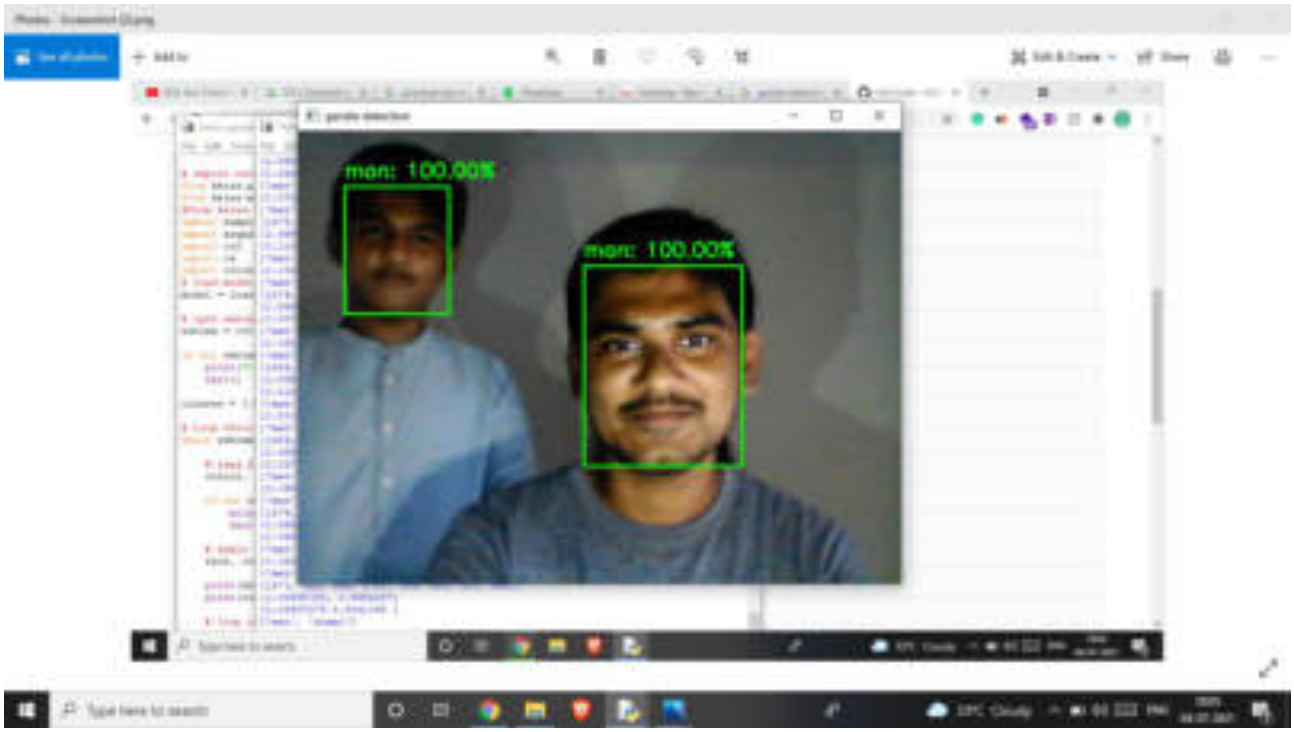


Fig 10.2:Output screen2

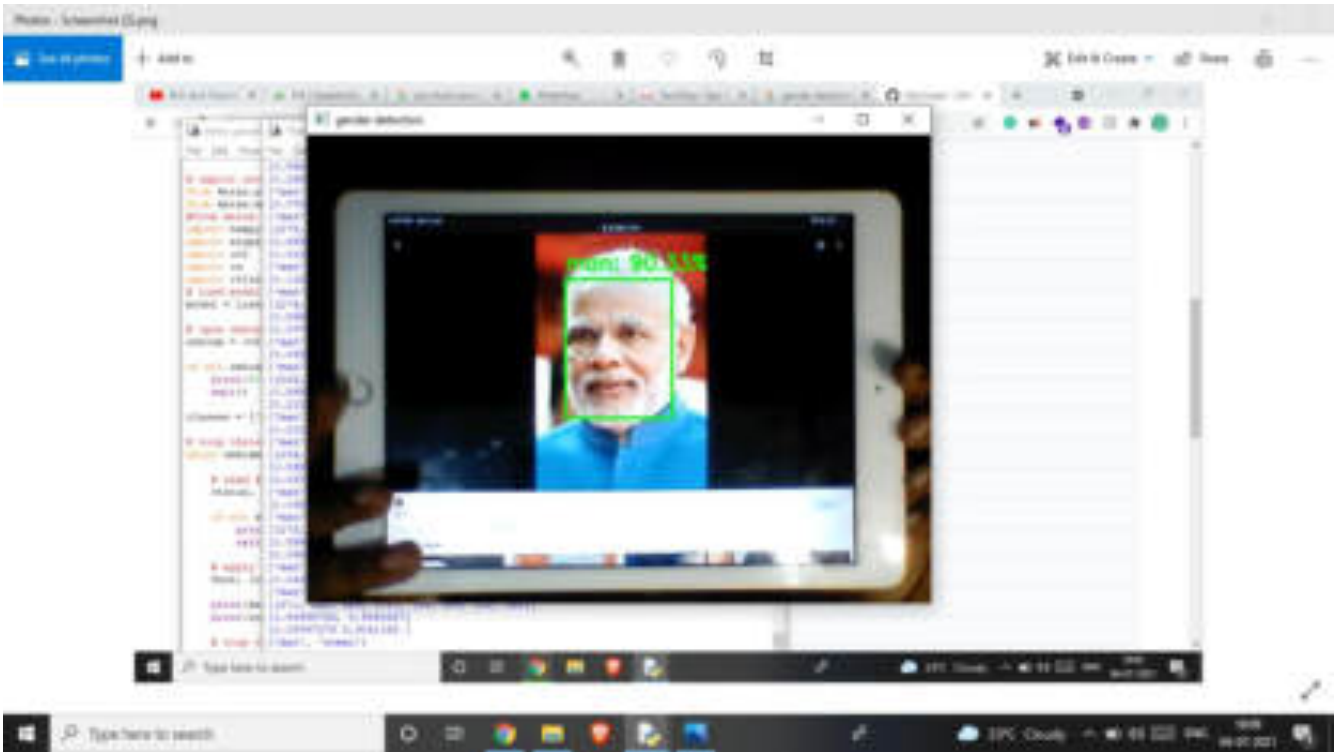


Fig 10.3: Output screen3

9.SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, subassemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of tests. Each test type addresses a specific testing requirement.

9.1TYPES OF TESTS

Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application. It is done after the completion of an individual unit before integration. This is a structural testing that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfactory, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

InvalidInput : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems/Procedures : interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identifying Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

White Box Testing

White Box Testing is testing in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It has a purpose. It is used to test areas that cannot be reached from a black box level.

Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a test in which the software under test is treated as a black box. you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects. The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

Test Results

All the test cases mentioned above passed successfully. No defects encountered.

9.2TESTING METHODOLOGIES

The following are the Testing Methodologies:

- UnitTesting.
- Integration Testing.
- User AcceptanceTesting.
- Output Testing.
- Validation Testing

Unit testing focuses verification effort on the smallest unit of Software design that is the module. Unit testing exercises specific paths in a module's control structure to ensure complete coverage and maximum error detection. This test focuses on each module individually, ensuring that it functions properly as a unit. Hence, the naming is Unit Testing.

During this testing, each module is tested individually and the module interfaces are verified for the consistency with design specification. All the important processing path are tested for the expected results. All error handling paths are also tested.

Integration Testing

Integration testing addresses the issues associated with the dual problems of verification and program construction. After the software has been integrated a set of high order tests are conducted. The main objective in this testing process is to take unit tested modules and build a program structure that has been dictated by design.

The following are the types of Integration Testing:

- **Top Down Integration**

This method is an incremental approach to the construction of program structure. Modules are integrated by moving downward through the control hierarchy, beginning with the main program module. The module subordinates to the main program module are incorporated into the structure in either a depth first or breadth first manner.

In this method, the software is tested from the main module and individual stubs are replaced when the test proceeds downwards.

- **Bottom-up Integration**

This method begins the construction and testing with the modules at the lowest level in the program structure. Since the modules are integrated from the bottom up, processing required for modules subordinate to a given level is always available and the need for stubs is eliminated. The bottom up integration strategy may be implemented with the following steps:

- The low-level modules are combined into clusters into clusters that perform a specific Softwaresub-function.
- A driver (the control program) for testing is written to coordinate test case input & output

- Drivers are removed and clusters are combined moving upward in the program structure

The bottom up approach tests each module individually and then each module is integrated with a main module and tested for functionality.

User Acceptance Testing

User Acceptance of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changes wherever required. The system developed provides a friendly user interface that can easily be understood even by a person who is new to the system.

Output Testing

After performing the validation testing, the next step is output testing of the proposed system, since no system could be useful if it does not produce the required output in the specified format. Asking the users about the format required by them tests the outputs generated or displayed by the system under consideration. Hence the output format is considered in 2 ways – one is on screen and another in printed format.

9.3 TEST CASES

TEST CASES

Test Case Identification Number:TC1 Test case for male image

1	Purpose of the test case	To check whether Prediction for male is correct or not
2	Pre-condition	Training model should be done
3	Inputs	Male image
4	Expected Outputs	It should display the male pop up with accuracy rate.
5	Post Conditions	Check whether predicted image is accurate or not
6	Outputs	Male(99%)
7	Pass/Fail	pass

Table .1: Test case for Male image

Test Case Identification Number:TC2

1	Purpose of the test case	To check whether Prediction for female is correct or not
2	Pre-condition	Training model should be done
3	Inputs	Female image
4	Expected Outputs	It should display the male pop up with accuracy rate.
5	Post Conditions	Check whether predicted image is accurate or not
6	Outputs	Female(99%)
7	Pass/Fail	pass

Table .1: Test case for Female image

Test Case Identification Number:TC3

1	Purpose of the test case	To check whether Prediction for female and female is correct or not
2	Pre-condition	Training model should be done
3	Inputs	Image with two men
4	Expected Outputs	It should display the male pop up with accuracy rate.
5	Post Conditions	Check whether predicted image is accurate or not
6	Outputs	Male, Male(99%)
7	Pass/Fail	pass

Table .3: Test case for Male, Male image

Test Case Identification Number:TC4

1	Purpose of the test case	To check whether Prediction for male and male is correct or not
2	Pre-condition	Training model should be done
3	Inputs	Image of two women
4	Expected Outputs	It should display the male pop up with accuracy rate.
5	Post Conditions	Check whether predicted image is accurate or not
6	Outputs	Female, Female(99%)
7	Pass/Fail	pass

Table .4: Test case for Female, Female image.

10. CONCLUSION

Convolutional Neural Network , a supervised machine learning algorithm gives accurate and better results as compared to other algorithms. For gender classification, the model is trained on the pre-processed data and hence is able to determine the gender of the face image. The categories used for gender classification are: male and female. This approach gives an average validation accuracy of 90% after 10 epochs for gender.The variance of validation accuracy is not that high because only less validation samples are used. More the number of samples, the more will be the accuracy of the model. The accuracy of the system can be increased by increasing the images in the dataset, changing the small ConvNet to VGG16 architecture, using bottle neck features etc.

10.FUTURE WORKS

Upon changing the dataset the same model can be trained to predict motion, age, ethnicity etc. The gender classification can be used to predict gender in uncontrolled real time scenarios such as railway stations, banks, bus stops, airports, etc. For example, depending upon the number of male and female passengers on the railway station, restrooms can be constructed to ease the travelling.

11.REFERENCES

- E. Makinen and R. Ralsamo, "Evaulation of Gender Classification Methods with Automatically Detected and Aligned Faces,"IEEE Transactions on Pattern Analaysis and Machine Intelligence,vol.30,no. 3,pp.541547,2008.
- S. U. Rehman, S. Tu, y. Huang, and Z. Yang,Face recognition :A Novel Un-supervised Convolutional Neural Network Method, IEEE International Conference of Online analysis Computing Science[ICOCAS],2016.
- N. Srinivas, H. Atwal, D. C. Rose, g. Mahalingam, K. Ricanek,and D. S. Bolme,Age ,Gender And Fine-Grained Ethinicity Prediction Using Convolutational Neural Networks for the East Asian
- N. Jain, S. Kumar, A. p. Shamsolomoali, and M. Zareapoor,Hybrid Deep Neural Networks for Face Emotion Recognition ,Pattern Recognition Letters,2018
- G. Levi ,and T. Hassner,"Age and Gender Classification Using Convolutional Neural Networks,"IEEE Worksship on Analysis and Modelling of Faces and Gestures on cv and Pattern Recognition (CVPR),Boston,2015.
- S. Turabzadeb, H. Meng ,R. M. Swash ,M. Pleva and J.Juhar, Realtime Emotional State Detection From Facial Expression On Embeeded Devices ,Seventh International Conference On Innovative Computing Technology (INTECH)2017.
- A. Dehghan ,E. G. Oritiz , G. Shu, and S. Z. Masood ,Dager:Deep Age,Gender and Emotion Recognition Using Convolutional Neural Networks ,arXiv preprint arXiv:1702.04280,2017.
- A. Krizhevsky, I. Sutskever,and G. E. Hinston, ImageNet Classification with deep convolutional neural networks ,Communications of the ACM,vol. 60 ,no. 6

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"PLAGIARISM CHECKER"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
COMPUTER SCIENCE AND ENGINEERING

Submitted by

J. Suhasini (17H71A05B2)

G. Ramya Tejaswini (17H71A0598)

Ch. Anupama (17H71A0561)

G. Siri (17H71A05A8)

Under the Esteemed Guidance of

Mrs.V.SRI LAKSHMI

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC A
With Grade

2020-21

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited B.Tech (CSE|ECE|ME)
Accredited by NAAC With Grade A

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"Plagiarism Checker"** is a bonafide work carried out by **J.Suhasini (17H71A05B2), G.Ramya Tejaswini (17H71A0598), Ch.Anupama (17H71A0561), G.Siri (17H71A05A8)** in partial fulfilment for the award of degree of Bachelor of Technology in **Computer Science and Engineering of Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

V. Sri Lakshmi
Project Guide

D. Prasad
Head of the Department

Examiner

Dr K Srinivas
PRINCIPAL

ACKNOWLEDGEMENT

We would like to take this opportunity to express my deepest appreciation to the following people for their valuable contributions and assistance with this Project.

Initially, we would like to thank my project supervisor, **V. Sri Lakshmi, Assistant Professor, Department of Computer Science and Engineering** for the guidance and support, especially for the valuable ideas and knowledge shared to me throughout the Project.

We have the immense pleasure in expressing my thanks and deep sense of gratitude to**, Mr. D. Prasad, Head of the Department, Computer Science and Engineering** for extending necessary facilities and support for the completion of the Project.

We whole sincerely acknowledge **Dr. K. Srinivas, Principal and Prof D. PANDURANGA RAO, CEO** for giving opportunity to take up the Project work. We also extend my thanks to all faculty members of **Computer Science and Engineering**, for their valuable guidance and encouragement.

We would like to extend my warm appreciation to all my friends for sharing us their knowledge, valuable contributions and help with this Project.

Finally, my special thanks go to my family for their continuous support and help throughout my academic years and for their continual support and encouragement for the completion of the project.

DECLARATION

We **J.Suhasini, G.Ramya Tejaswini, Ch.Anupama, G.Siri** of the Main-Project "**Plagiarism Checker**", hereby declare that the matter embodied in this Project is the genuine work done by us and has not been submitted either to this university or to any other university/institute for the fulfilment of the requirement of any course of study.

J.Suhasini (17H71A05B2)

G.RamyaTejaswini(17H71A0598)

Ch.Anupama (17H71A0561)

G.Siri (17H71A05A8)

ABSTRACT

Plagiarism is unethical theift of intellectual property.

Nowadays plagiarism has become a major issue in the literary world mostly students use search engines for plagiarized material and use that at an academic level.The main reason behind plagiarism is it's the easiest and quickest way to accomplish literary tasks like academic papers, articles or essays. Sometimes students could not sustain the pressure of work and deadlines for the projects, so they prefer to steal someone else's work instead of creating original content.At academic level students often use plagiarized content instead of creating their own. So there are plagiarism checkers for students that easily identify the duplicate content.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The simple meaning of plagiarism is cheating, but if you check the real meaning of plagiarism in the dictionary then it will give somewhat like “the practice of taking someone else’s work or ideas and passing them off as one’s own”

Thus, Plagiarism is stealing or copying someone else’s intellectual work (can be an idea, literary work or academic work, etc.) and representation it as your own work without giving credit to the creator.

This is a really simple plagiarism checker .You just input the text you want to search for and it returns a link if there is any similar text.

- When the document to be checked is fed to these software checkers, they break the text into smaller chunks of phrases and search for similarity into the search engines. On encountering similar content, the tool marks the phrase or words that are plagiarized.

- Teachers can use this plagiarism checker in essay writing competition ,assignement writing or in any hacktons to detect whether their students content is copied from any website or not.

- With using this plagiarism checker if the content is copied then they are nullified if not they are awarded the respective marks.

- Nowadays plagiarism has become a most serious and major issue of the literary world. Most victims who fall in the trap of plagiarism are students, because students have to produce a number of essays and research papers within the given deadline, they don’t want to lose marks and sometimes it became difficult to sustain the pressure of work.

- Hence, students go for an easy and less time-consuming way out to complete their assignments and apply a simple formula of copy-pasting other writer’s original content and present as their own work or present the previous content as new and original content. That’s why plagiarism checker has become essential for academies, educational institutions and universities. It is also important for promoting original content writings.

11.BACKGROUND

The etymology of the word plagiarism is from the word "plagiarius" meaning "kidnapper,seducer,plunderer..".

However,the first time it was used in the context of literature was sometimes around 80 AD by the Roman poet Martial. At that time, poets were expected to be able to recite keywork by other authors. However, when Martial learned that another poet, Fidentinus, was reciting his works and taking credit for them, Martial chose to respond.

But Martial didn't have the option of going to the courts. Modern copyright law wouldn't exist for another 1600 years and there was no legal remedy available. Instead, Martial wrote a series of verses about Fidentinus, essentially creating a diss track about him.

In one of those verses, Martial referred to Fidentinus as a "plagiarus," essentially calling him a kidnapper.

Interestingly though, Martial wasn't concerned about the use of his work without attribution. Instead, he was more concerned about the lack of payment, essentially saying in one verse,"If you're willing that they be called mine, I'll send you the poems for free.If you want them to be called yours, buy this one, so that they won't be mine."

This isn't a surprise considering that ghostwriting was a common way for Roman poets to earn a living. It would be quite some time before the focus on plagiarism turned to authorship and originality.

1.2 PURPOSE OF THE PROJECT

Plagiarism Checker helps in detecting plagiarism in thesis, dissertations, assignments, and blogs. It is helpful for students and teachers both.

1.3 PROBLEM STATEMENT

The objectives of any plagiarism checker include to find the similarities in the text and ensure that the document is original. It further implies that no part of the document is copied from other writer's work. ... A plagiarism checker is a nice way for checking the originality of a document.

1.4 SCOPE OF THE PROJECT

The following steps can help you to effectively define the scope of a project.

Step 1: Identify the project needs

When you are clearly able to identify the needs of a project, you are more likely to set a sound benchmark from the beginning.
Understanding the 'what and why' of a project will enable you to set specific goals and objectives. It also sets the groundwork for what tasks are to follow and how they are to be performed.

Step 2: Confirm the objectives and goals of the Project

The basis of the project scope should entail your goals and objectives to be one that follows a SMART guideline. That is, to be Specific, Measurable and Achievable.
It should also be Realistic and completed within a specific Timeframe.

Specific–This involves stating accurately what the project wants to achieve. That is, what, why and how these will be done. Clarity will reduce the chances of ambiguities and misunderstandings.

Measurable –Are your goals and objectives able to provide feedback and be accountable for?

Achievable –Can your project's goals and objectives be achieved, given the resources on hand?

Realistic – Are the goals and objectives easy to deliver, especially if you face problems or complications. Will these reduce the overall quality of the project's outcome and cause running over budget and not meeting the set deadlines.

Time Frame – Can your project goals and objectives be met within the allocated time frame? Is it a key criterion to meet these deadlines?

Step 3: Project Scope description

You as a leader, need to be clear about the features and functioning required for your product or service.

For example, you are building a website. You need a list that provides how you will build your website, the type of branding required and so on. In other words, what certain qualities will increase achieving your project's success.

Step 4: Expectations and acceptance

Successful projects are ones that take into account the satisfaction of the enduser. Whether they meet the end-users expectations and accept the product, service or process. The end-users could be your customers or your internal team.

For customers, this includes pricing, value, and quality of products/services as well as availability, delivery and return policies. For employees, this includes the effectiveness and efficiency of new operational processes. Ultimately, your project scope is one that should be attuned to giving better outcomes to whoever your end users may be.

Step 5: Identify constraints

There are always roadblocks to achieving what you were set out to do. When being aware of possible limitations along the way, it can help you minimize problems that may delay or constrain your ability to achieve your project's outcome.

These can be caused by dynamic environmental conditions (internal and external), technological glitches and/or lack of resources. Communicating such problems with your team early on and taking steps to overcome these hurdles will reduce delays in project completion and keep spending within budget. Whether these are based on assumptions or uncertainty, analyzing their impact throughout the projects timeline further reduces the risk of failure.

Step 6: Identify necessary changes

It is always best to avoid reworking the scope of your project, as it means investing in more time, money and resources**.**

However, at times these changes are inevitable and necessary .Limit changes by taking on the perspectives of customers, stakeholders, and employees involved in the project. This minimizes disagreements later on.

2.LITERATURE SURVEY

Academic institutions often have clear-cut zero-tolerance policies for students caught plagiarizing their research and essays. This is because of the widespread incidence of plagiarism among higher education students, which presents an increasing concern for school administrators and faculty who are serious in defending academic integrity (MacLennan, 2018). However, it is a bit different for research professionals submitting their work to journals for publication where even self-plagiarism is not allowed. Self-plagiarism is when the author reuses previously published content and passes it off as new without telling the reader where the work previously appeared (Roig, 2011).

In a plagiarism study by Enders et al. (2004), they found that only 19% of journal editors had formal plagiarism policies within their organizations. As such, journal editors can respond to instances of plagiarism in a number of ways. In the survey, 70.9% said that notifying the original author as soon as possible is definitely an appropriate response to a clear case of plagiarism (Enders, Walter & Hoover, Gary, 2004). However, 42% said informing the plagiarists' dean and other superiors is not likely an appropriate response. Moreover, 51% said that informing the public about the incident is also not likely an appropriate response.

Another plagiarism study by Karabag & Berggren (2012) tried to relate the issue to papers that were retracted due to intellectual dishonesty. They observed that science journals are more active in retracting papers than business and economic journals. For instance, they found more than 700 papers retracted from ScienceDirect, a database of scientific journals from 1985 to 2012. In contrast, they only found seven retraction notices from Emerald Management Journals. This led them to conclude that "the leading business and economics journals' response to academic dishonesty and plagiarism has been slow (Karabag and Berggren**,** 2012)".

One prominent case of plagiarism involved Harvard Law professor Charles Ogletree. He admitted to copying word for word six paragraphs from the work of another law professor to his book (Pacia, 2004). Ogletree issued a public apology, after which an internal investigation found he committed an honest mistake (Torres, 2004). This decision was criticized by The Harvard Crimson as imposing double standards on students and professors caught plagiarizing (The Crimson Staff, 2004).

In another case, Bindu Ganga, the director of training at Argosy University-Chicago was found to have plagiarized her doctoral research project about lying (Newbart, 2006). Ganga was fired from her job and stripped of her Doctor of Psychology degree (Newbart, 2007). This was after allegations of plagiarism from a student were confirmed by Turnitin, which showed her work was roughly 45% similar to other published works.

These unfortunate cases illustrate the importance of ensuring one's research is plagiarism-free to avoid damage to one's professional reputation and career.

3.SYSTEM ANALYSIS

System analysis is the process of studying a given system to find how a system functions its components and the interactions. This system analysis as analyst may come to know a given business logic functions and its objective to find better solution as well as drawbacks if any and details of improvements to enhance the system performance.

3.1 EXISTING SYSTEM:

3.1.1DUPLICHECKER

Duplichecker is a plagiarism checker tool that can find instances of duplicate content.

It scans the internet for the same sentences, phrases, or paragraphs you've got on your website, and points you to external URLs that have the same content.

3.1.2 ADVANTAGES

The advantage of a Duplichecker is that it is free of cost and you can easily use it online from the internet. Even if it is a free service, still it is the most trustworthy service that you can use to check plagiarism. Through this application, you can also check stuff relevant to other software. It not only checks plagiarism but also provides you editing services for any mistake or blunder so that you will have perfect writing. It also checks grammatical errors, detects spelling mistakes. It tells you the right way for placing the references. The most common and reliable example of plagiarism duplichecker is Viper. This is the highly used software to detect the places of plagiarism in any work.

3.1.3 DISADVANTAGES

With some advantages there are also some disadvantages such as the plagiarism tool can accurately detect plagiarism when the text is comprised of seven or more than seven words but it cannot work as well with the text of smaller words. And because of this, we cannot say that it is an infallible application. However, it plays a good role when working with barefaced cases of plagiarism. Although, it cannot completely find the mistakes it can help up to some extent.

With the help of a plagiarism finder, the authors and writers may check that their efforts are still original or someone has stolen it and made a duplicate of it

3.2 PROPOSED SYSTEM

3.2.1 PLAGIARISM CHECKER

- Plagiarism checker is a online plagiarism detection tool and help to find the duplicate content.
- It takes input has 2 ways 1.Copy and paste
  2.Upload a file

Plagiarism checker is a online plagiarism detection tool and helps to find the duplicate content. It takes the input as text and searches a similar text is present in google or not. If similar text is present then returns the link of similar text.

3.2.2 ADVANTAGES

- It is simple.
- It is easy to use.
- It can accurately detect plagiarism when the text is comprised of not only longer words but also smaller words.
- There is no limit for plagiarism scans.
- No payments, No subscriptions.
- Any user can access for free.

3.3 OPERATIONAL ENVIRONMENT

The operating environment defines about the software requriments and hardware requriments

The Software Requriments that are used in the project are:

Operating System : Windows Family

IDE : Python 3

The Hardware Requriments that are used in the project are:

- 64-bit cpu
- 4 GB ram.
- 5 GB free disk space.

3.4 SYSTEM FEATURE

The System Features of Plagiarism Checker on the given text data it analyze and retrieve the URL's and then check similarity ratio.

3.4.1 DESCRIPTION AND PRIORITY

Plagiarism Checker is an ongoing field of research in industry's marketing technology. The priority given for each requriment in Plagiarism Checker is considered to be very high. Every requriment is implemented in the sequential order and in evitable.

3.5FUNCTIONAL REQURIMENTS

The Functional Requriments of Plagiarism Checker are as follows:

Req_ID_1: Extraction of Text.

Req_ID_2: Pre-processing of Text.

Req_ID_3: Creation of URL'S.

Req_ID_4: Comparing the Similarity.

3.5.1 Req_ID_1 Extraction of Text

Here, The input that is given by user whether it be text or file. That data is taken and send it to preprocess.

3.5.2 Req_ID_2 Pre-processing of Text

The Data or text is preprocessed by dividing the pharses into chuncks.This divided text is sent to other units.

3.5.3 Req_ID_3 Creation of URL's

By using the module google search ,this data or text is processed .This google search module uses search method and searches from the similar data in google. And the URL's that are given by the google search module are stored**.**

3.5.4 Req_ID_4 Comparing the Similarity

These stored URL's data is retrieved by using python module called Beautifulsoup. This module retrive the data and checks the similarity between input data and retrieved data. And finally it displays the percentage of similarity on the screen.

3.5 PROPOSED SYSTEM ARCHITECTURE

System architecture refers to the placement of these software components on physical machines. Two closely related components can be co-located or placed on different machines. The location of components will also impact performance and reliability.

The resulting architectural style ultimately determines how components are connected, data is exchanged, and how they all work together as a coherent system**.**

The below figure shows the clear idea how does plagiarism checker work. The step by step process

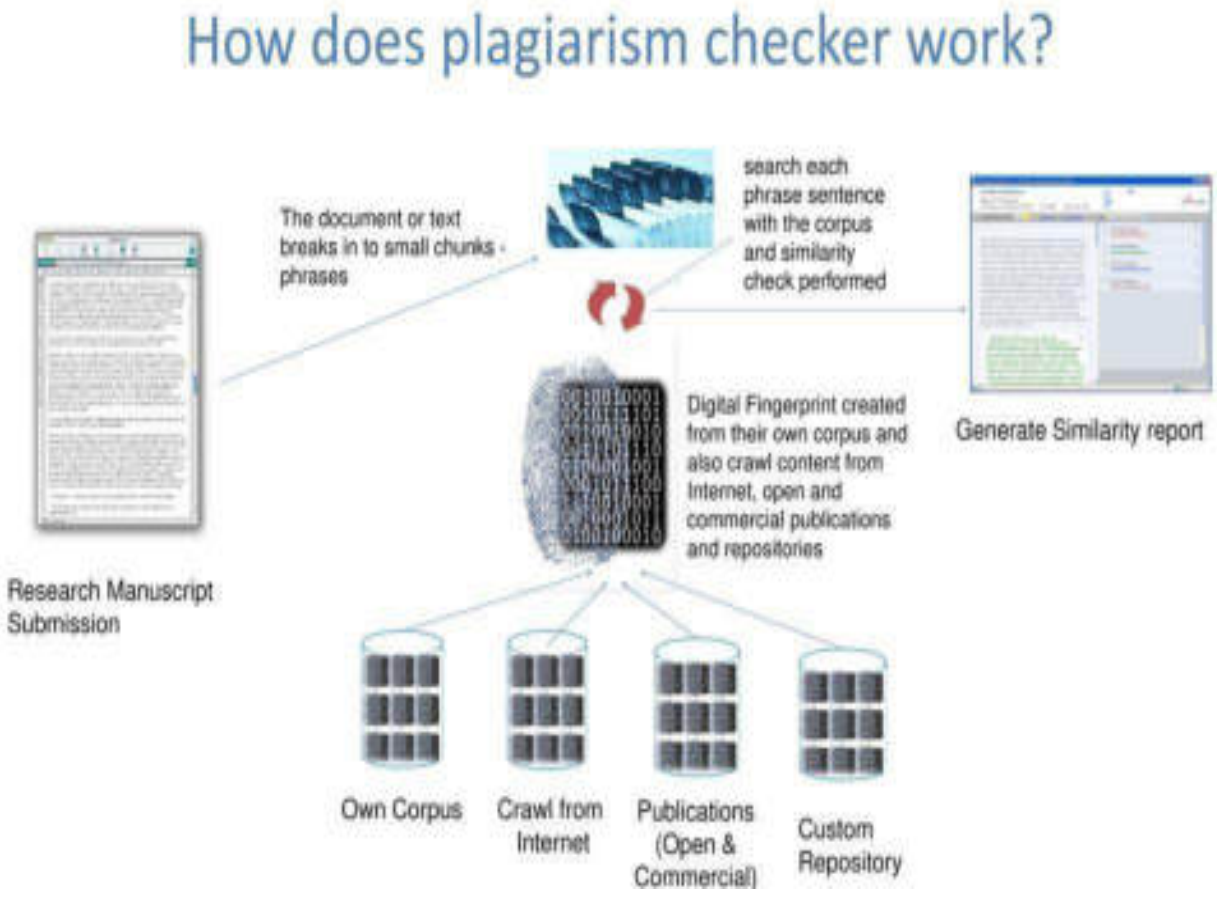


Fig:3.1 Proposed System Architecture

3.6 REAL TIME EXAMPLE

Helpful for Educational Industry

Such a site is helpful for the educational industry. Because within some seconds, students are increasingly dependent online world to complete their investigation projects. However, call that sheer laziness or maybe laxity; they pun intended, the pain of changing the content. Instead of connected with collecting data and crafting their paper within their own thoughts, they just copypaste the content and make a so-called outstanding report and also project as their own. At initial glance, such a written report looks really impressive. However, when you put it on a plagiarism checker, the awful truth stares inside your face.

Copy-pasting online has to turn into a threat nowadays. It doesn't just obstruct the ones originality within the individual exactly who indulges in such a react, but furthermore taints the writing means. To maintain your sanctity connected with writing whole, plagiarism checker zero cost online helps you to catch copycats.

3.7 FLOWCHART

A flowchart is a visual representation of the sequence of steps and decisions needed to perform a process. Each step in the sequence is noted within a diagram shape. Steps are linked by connecting lines and directional arrows. This allows anyone to view the flowchart and logically follow the process from beginning to end.

A flowchart is a powerful business tool. With proper design and construction, it communicates the steps in a process very effectively and efficiently.

The below picture depicts the flow of the system.

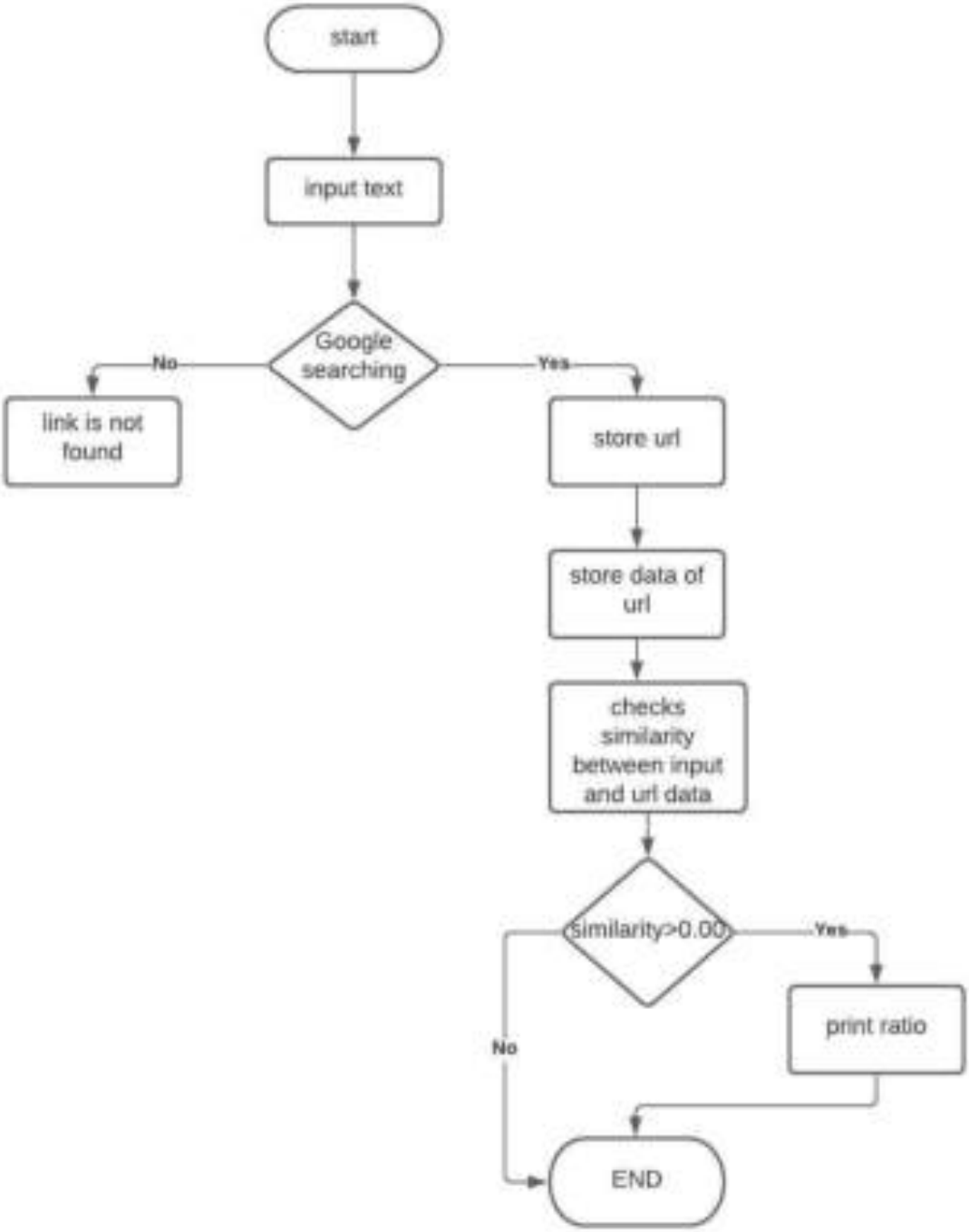


Fig:3.2 Flow chart of Plagiarism Checker

4.SYSTEM DESIGN

4.1 INTRODUCTION

Design is a multi-step that focuses on data structure software architecture, procedural details, algorithm and interface between modules. The design process also translates the requirements into presentation of software that can be accessed for quality before coding begins. Systems design could be seen as the application of systems theory to product development. There is some overlap with the disciplines of systems analysis, systems architecture and systems engineering. Until the 1990s, systems design had a crucial and respected role in the data processing industry. In the 1990s, standardization of hardware and software resulted in the ability to build modular systems.

The increasing importance of software running on generic platforms has enhanced the discipline of software engineering. The architectural design of a system emphasizes the design of the system architecture that describes the structure, behavior and more views of that system and analysis. It is meant to satisfy specific needs and requirements of a business or organization through the engineering of a coherent and well-running system.

They needed to be able to standardize their work into a formal discipline with proper methods, especially for new fields like information theory, operations research and computer science in general. Computer software design change continuously as new methods; better analysis and border understanding evolved. Software design is at relatively early stage in its revolution. Therefore, software design methodology lacks the depth, flexibility and quantitative nature that are normally associated with more classical engineering disciplines. However techniques for software designs do exit, criteria for design qualities are available and design notation can be applied.

4.2 UML DIAGRAMS:

The Unified Modelling Language(UML) is a standard language for writing software blueprints. The UML may be used to visualize, specify, construct, and document the artifacts of a software intensive system.

The UML is only a appropriate for modelling systems ranging from enterprise information system distributed Web-based applications even to hard real time embedded systems . It is a very expensive language, addressing all the views needed and then deploy such systems . The UML is only a language and so is just one part of software development method. The UML is process independent, although optimally it should be used in a process that is use case driven, architecturecentric, iterative and incremental.

4.2.1 CLASS BASED DIAGRAM

CLASS DIAGRAM

Class diagram is a static diagram. It represents the static view of an application. Class diagram is not only used for visualizing, describing, and documenting different aspects of a system but also for constructing executable code of the software application.

Class diagram describes the attributes and operations of a class and also the constraints imposed on the system. The class diagrams are widely used in the modeling of objectoriented systems because they are the only UML diagrams, which can be mapped directly with object-oriented languages.

Class diagram shows a collection of classes, interfaces, associations, collaborations, and constraints. It is also known as a structural diagram.

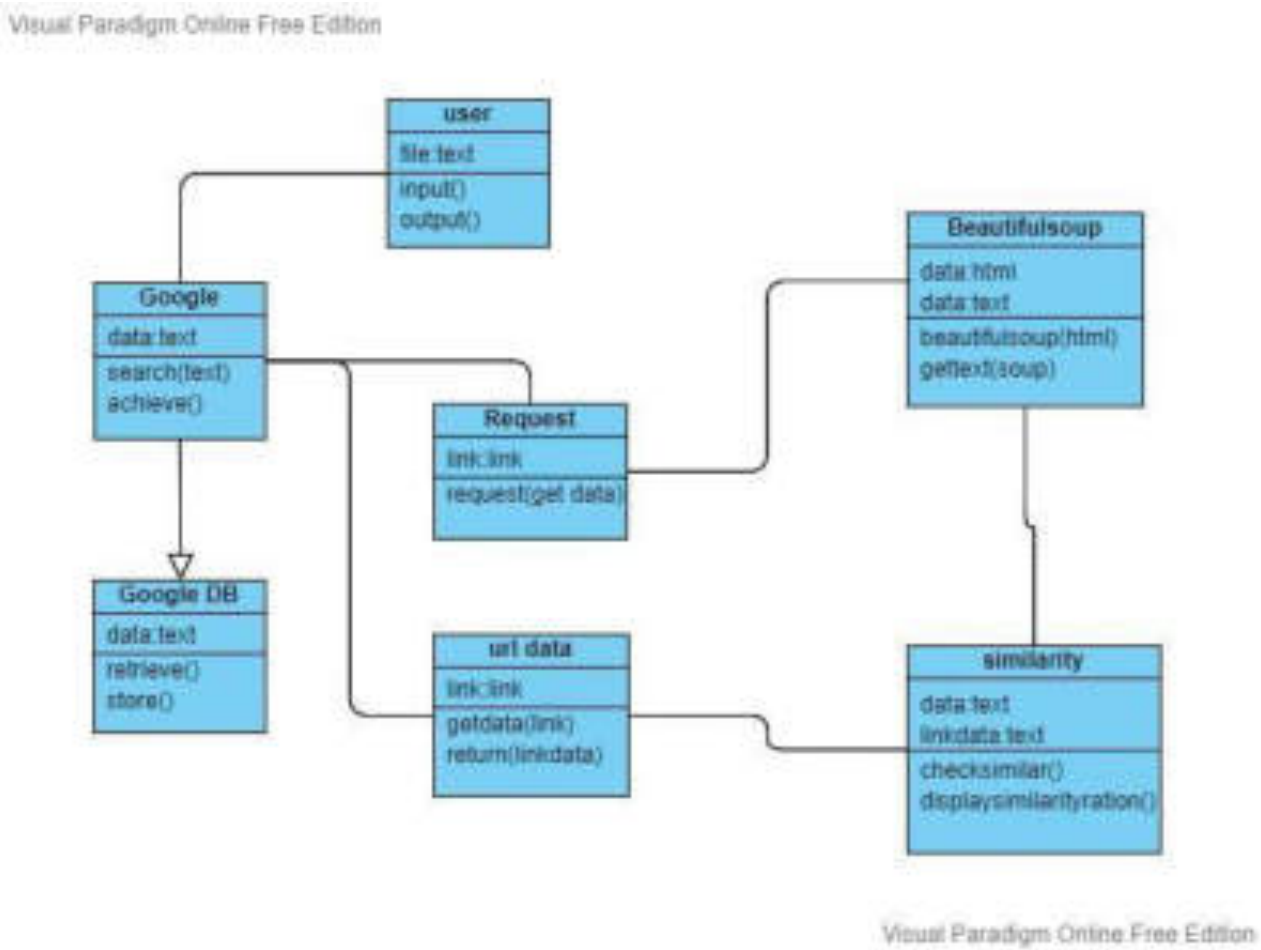


Fig:4.1 Class Diagram of Plagiarism Checker

4.2.2 USE CASE DIAGRAM

To model a system, the most important aspect is to capture the dynamic behavior. Dynamic behavior means the behavior of the system when it is running/operating.

Only static behavior is not sufficient to model a system rather dynamic behavior is more important than static behavior. In UML, there are five diagrams available to model the dynamic nature and use case diagram is one of them. Now as we have to discuss that the use case diagram is dynamic in nature, there should be some internal or external factors for making the interaction.

These internal and external agents are known as actors. Use case diagrams consists of actors, use cases and their relationships. The diagram is used to model the system/subsystem of an application. A single use case diagram captures a particular functionality of a system.

Hence to model the entire system, a number of use case diagrams are used.

Purpose of Use Case Diagrams

The purpose of use case diagram is to capture the dynamic aspect of a system. However, this definition is too generic to describe the purpose, as other four diagrams (activity, sequence, collaboration, and Statechart) also have the same purpose. We will look into some specific purpose, which will distinguish it from other four diagrams.

Use case diagrams are used to gather the requirements of a system including internal and external influences. These requirements are mostly design requirements. Hence, when a system is analyzed to gather its functionalities, use cases are prepared and actors are identified.

When the initial task is complete, use case diagrams are modelled to present the outside view.

In brief, the purposes of use case diagrams can be said to be as follows –

- Used to gather the requirements of a system.
- Used to get an outside view of a system.
- Identify the external and internal factors influencing the system.
- Show the interaction among the requirements are actors.

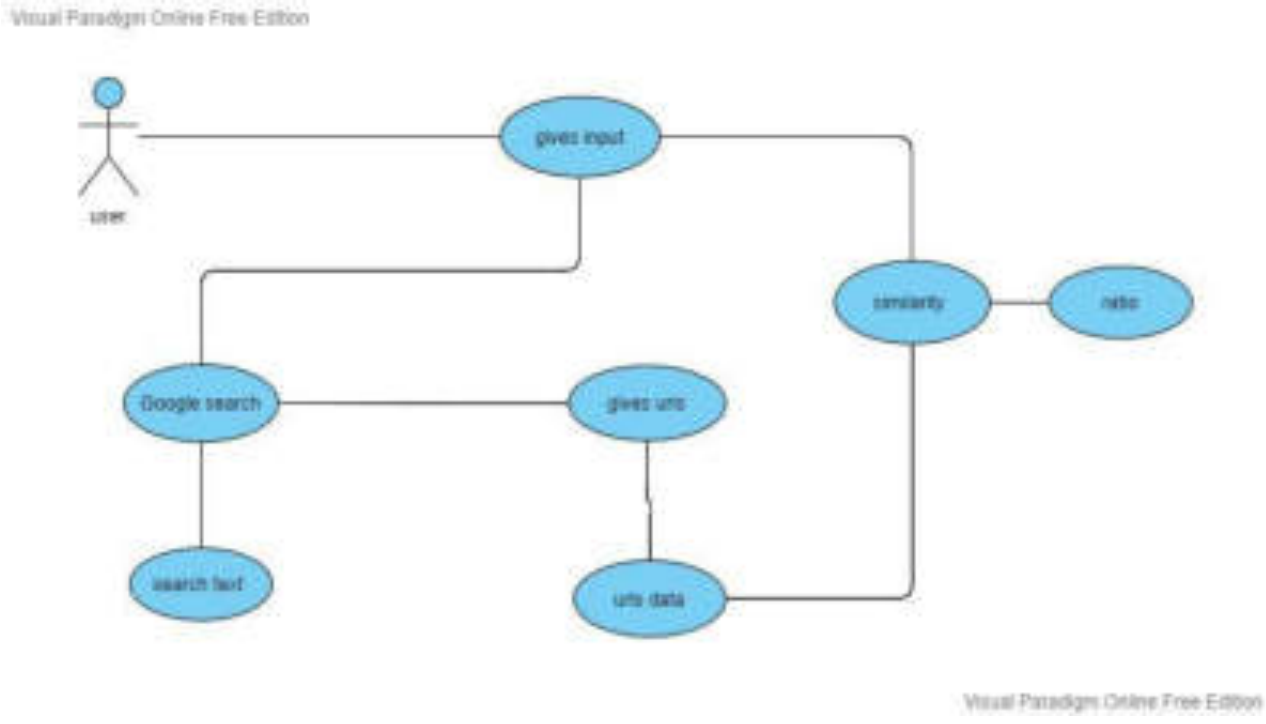


Fig:4.2 Use Case Diagram of Plagiarism Checker

4.2.3 SEQUENCE DIAGRAM

A sequence diagram simply depicts interaction between objects in a sequential order i.e. the order in which these interactions take place. We can also use the terms event diagrams or event scenarios to refer to a sequence diagram. Sequence diagrams describe how and in what order the objects in a system function. These diagrams are widely used by businessmen and software developers to document and understand requirements for new and existing systems**.**

Sequence Diagram Notations –

Actors – An actor in a UML diagram represents a type of role where it interacts with the system and its objects. It is important to note here that an actor is always outside the scope of the system we aim to model using the UML diagram.

Lifelines – A lifeline is a named element which depicts an individual participant in a sequence diagram. So basically each instance in a sequence diagram is represented by a lifeline. Lifeline elements are located at the top in a sequence diagram. The standard in UML for naming a lifeline follows the following format – Instance Name : Class Name.

Messages – Communication between objects is depicted using messages. The messages appear in a sequential order on the lifeline. We represent messages using arrows. Lifelines and messages form the core of a sequence diagram.

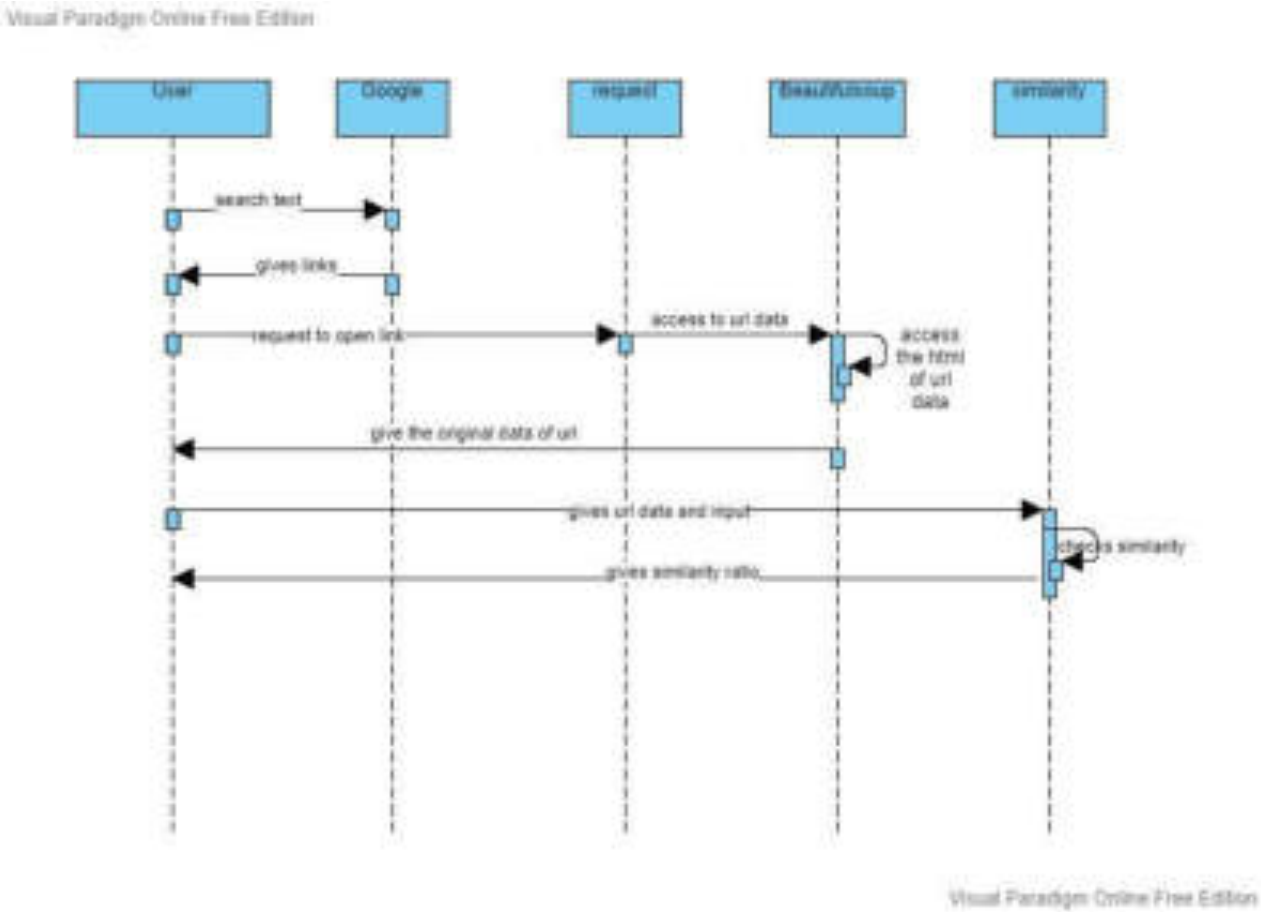


Fig:4.3 Sequence Diagram of Plagiarism Checker

4.2.4 COLLABORATION DIAGRAM

The collaboration diagram is used to show the relationship between the objects in a system. Both the sequence and the collaboration diagrams represent the same information but differently. Instead of showing the flow of messages, it depicts the architecture of the object residing in the system as it is based on object-oriented programming. An object consists of several features. Multiple objects present in the system are connected to each other. The collaboration diagram, which is also known as a communication diagram, is used to portray the object's architecture in the system.

Notations of a Collaboration Diagram

Following are the components of a component diagram that are enlisted below:

Objects: The representation of an object is done by an object symbol with its name and class underlined, separated by a colon.

In the collaboration diagram, objects are utilized in the following ways:

The object is represented by specifying their name and class.

It is not mandatory for every class to appear.

A class may constitute more than one object.

In the collaboration diagram, firstly, the object is created, and then its class is specified.

To differentiate one object from another object, it is necessary to name them.

Actors: In the collaboration diagram, the actor plays the main role as it invokes the interaction. Each actor has its respective role and name. In this, one actor initiates the use case.

Links: The link is an instance of association, which associates the objects and actors. It portrays a relationship between the objects through which the messages are sent. It is represented by a solid line. The link helps an object to connect with or navigate to another object, such that the message flows are attached to links.

Messages: It is a communication between objects which carries information and includes a sequence number, so that the activity may take place. It is represented by a labeled arrow, which is placed near a link. The messages are sent from the sender to the receiver, and the direction must be navigate in that particular direction.The receiver must understand the message

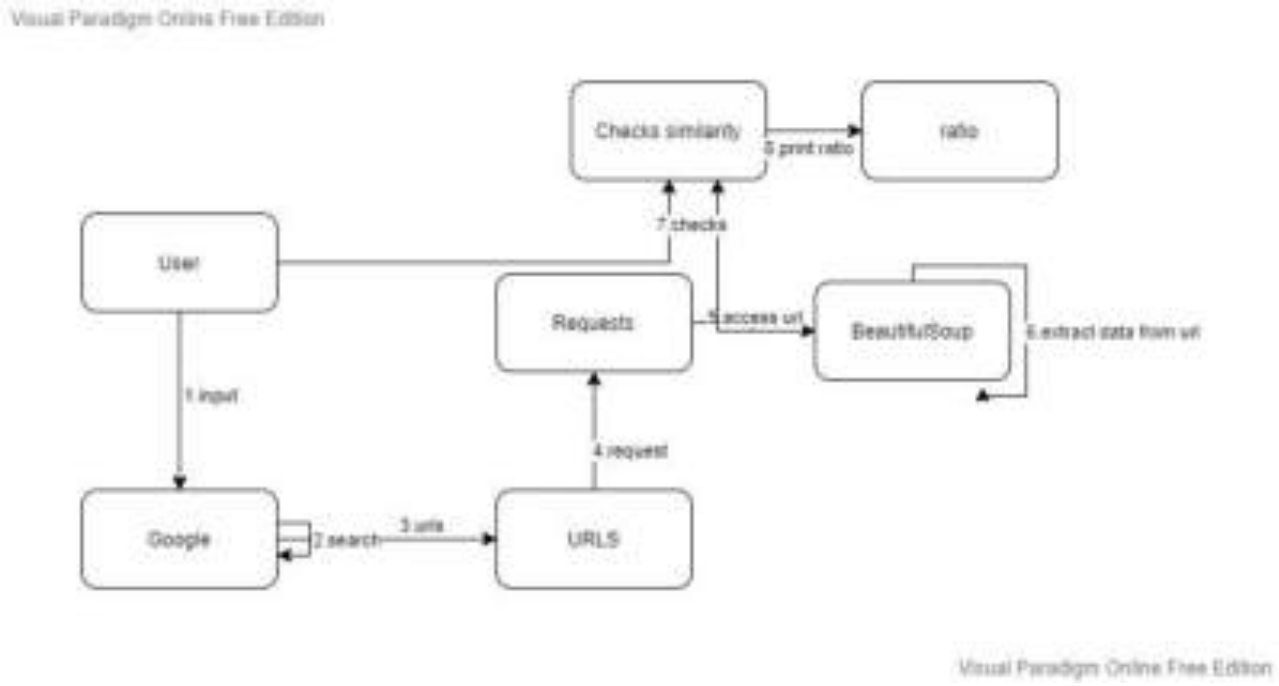


Fig:4.4 Collaboration Diagram of Plagiarism Checker

4.2.5 ACTIVITY DIAGRAM

Activity diagram is another important diagram in UML to describe the dynamic aspects of the system.

Activity diagram is basically a flowchart to represent the flow from one activity to another activity.

The activity can be described as an operation of the system.The control flow is drawn from one operation to another. This flow can be sequential, branched, or concurrent. Activity diagrams deal with all type of flow control by using different elements such as fork, join, etc

Purpose of Activity Diagrams

The basic purposes of activity diagrams is similar to other four diagrams. It captures the dynamic behavior of the system. Other four diagrams are used to show the message flow from one object to another but activity diagram is used to show message flow from one activity to another.

Activity is a particular operation of the system. Activity diagrams are not only used for visualizing the dynamic nature of a system, but they are also used to construct the executable system by using forward and reverse engineering techniques. The only missing thing in the activity diagram is the message part.

It does not show any message flow from one activity to another. Activity diagram is sometimes considered as the flowchart. Although the diagrams look like a flowchart, they are not. It shows different flows such as parallel, branched, concurrent, and single.

The purpose of an activity diagram can be described as –

- Draw the activity flow of a system.
- Describe the sequence from one activity to another.
- Describe the parallel, branched and concurrent flow of the system

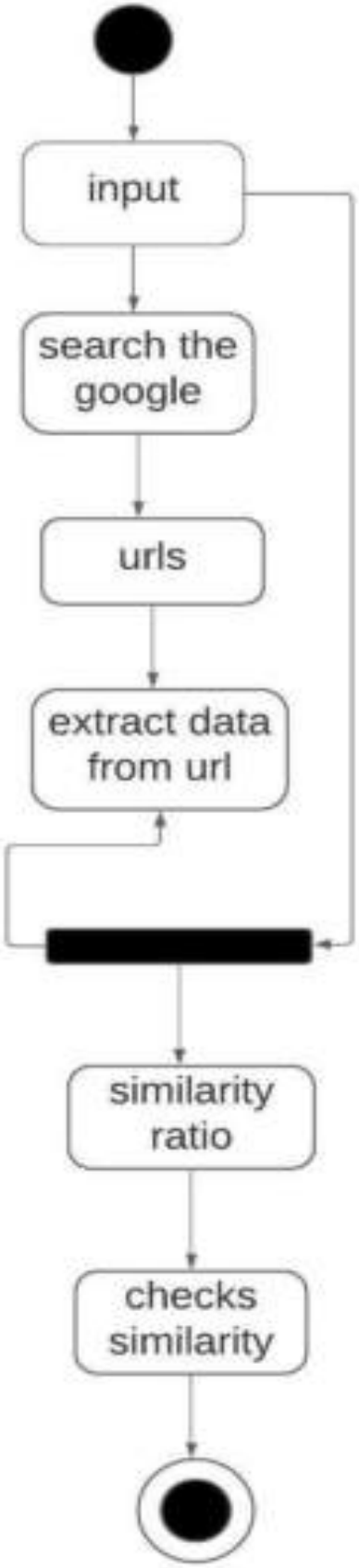


Fig:4.5 Activity diagram

4.2.6 STATE CHART DIAGRAM

The name of the diagram itself clarifies the purpose of the diagram and other details. It describes different states of a component in a system. The states are specific to a component/object of a system.

A Statechart diagram describes a state machine. State machine can be defined as a machine which defines different states of an object and these states are controlled by external or internal events.

Activity diagram explained in the next chapter, is a special kind of a Statechart diagram. As Statechart diagram defines the states, it is used to model the lifetime of an object.

Purpose of Statechart Diagrams

Statechart diagram is one of the five UML diagrams used to model the dynamic nature of a system. They define different states of an object during its lifetime and these states are changed by events. Statechart diagrams are useful to model the reactive systems. Reactive systems can be defined as a system that responds to external or internal events.

Statechart diagram describes the flow of control from one state to another state. States are defined as a condition in which an object exists and it changes when some event is triggered. The most important purpose of Statechart diagram is to model lifetime of an object from creation to termination.

Statechart diagrams are also used for forward and reverse engineering of a system. However, the main purpose is to model the reactive system.

Following are the main purposes of using Statechart diagrams –

- To model the dynamic aspect of a system.
- To model the life time of a reactive system.
- To describe different states of an object during its life time.
- Define a state machine to model the states of an object.

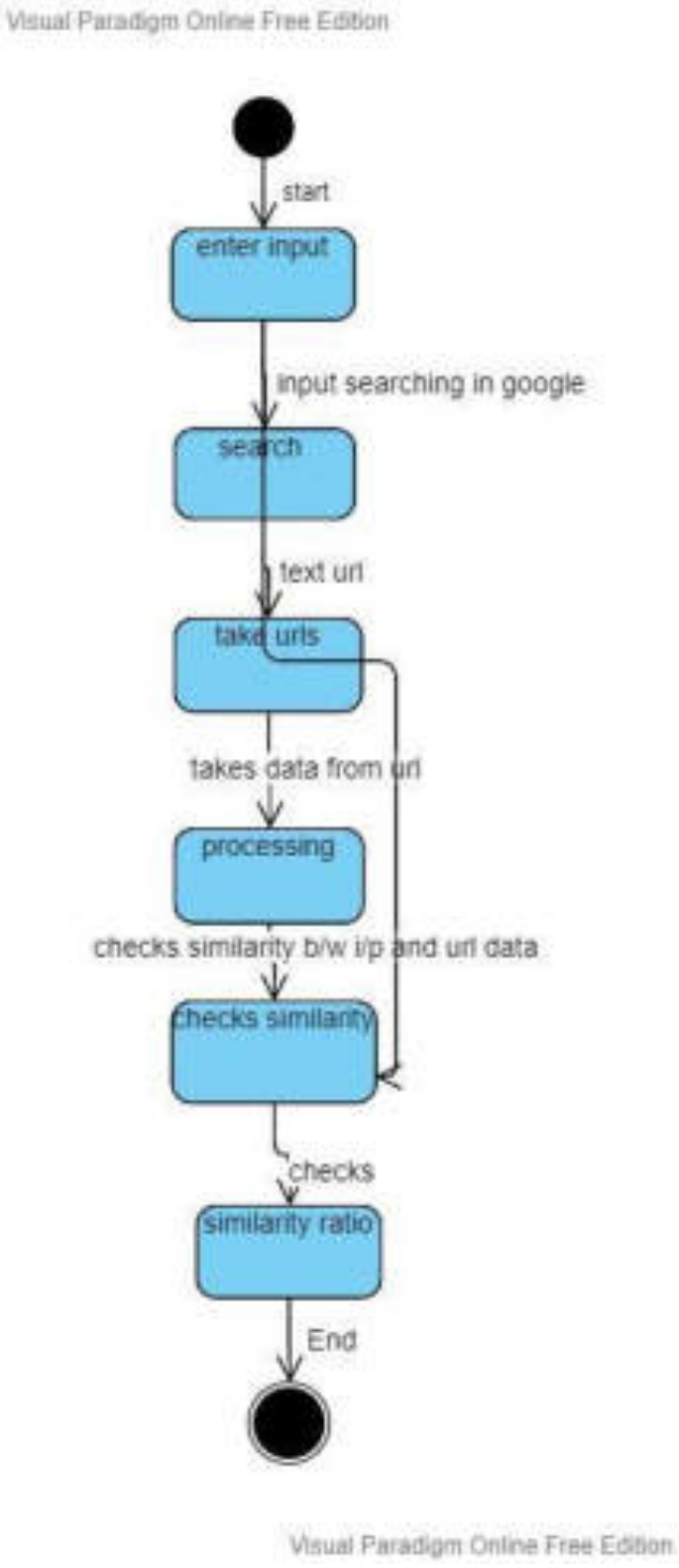


Fig:4.6 State Chart Diagram of Plagiarism Checker.

4.2.7 COMPONENT DIAGRAM

Component diagrams are different in terms of nature and behavior. Component diagrams are used to model the physical aspects of a system. Now the question is, what are these physical aspects?

Physical aspects are the elements such as executables, libraries, files, documents, etc. which reside in a node.

Component diagrams are used to visualize the organization and relationships among components in a system. These diagrams are also used to make executable systems.

Purpose of Component Diagrams

Component diagram is a special kind of diagram in UML. The purpose is also different from all other diagrams discussed so far. It does not describe the functionality of the system but it describes the components used to make those functionalities.

Thus from that point of view, component diagrams are used to visualize the physical components in a system. These components are libraries, packages, files, etc.

Component diagrams can also be described as a static implementation view of a system. Static implementation represents the organization of the components at a particular moment.

A single component diagram cannot represent the entire system but a collection of diagrams is used to represent the whole.

The purpose of the component diagram can be summarized as –

- Visualize the components of a system.
- Construct executables by using forward and reverse engineering.
- Describe the organization and relationships of the components.

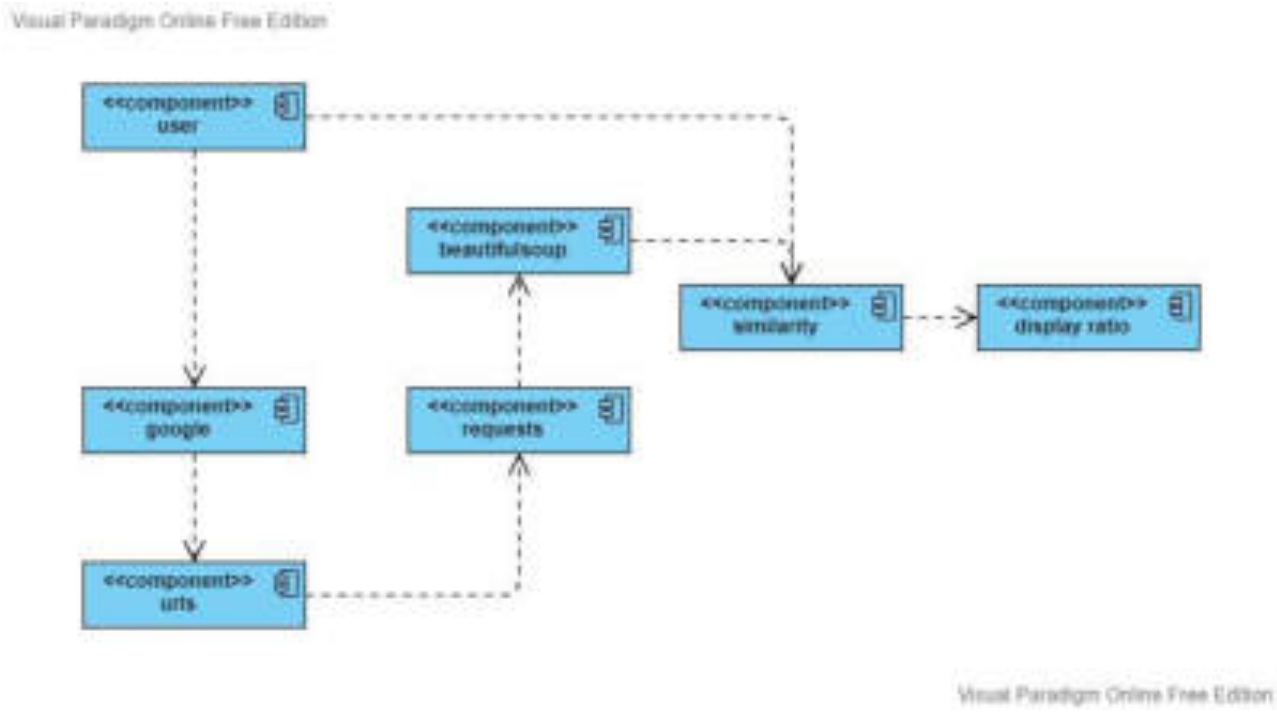


Fig:4.7 Component Diagram of Plagiarism Checker.

4.2.8 DEPOLYMENT DIAGRAM

Deployment diagrams are used to visualize the topology of the physical components of a system, where the software components are deployed.

Deployment diagrams are used to describe the static deployment view of a system. Deployment diagrams consist of nodes and their relationships.

Purpose of Deployment Diagrams

The term Deployment itself describes the purpose of the diagram. Deployment diagrams are used for describing the hardware components, where software components are deployed. Component diagrams and deployment diagrams are closely related.

Component diagrams are used to describe the components and deployment diagrams shows how they are deployed in hardware.

UML is mainly designed to focus on the software artifacts of a system. However, these two diagrams are special diagrams used to focus on software and hardware components.

Most of the UML diagrams are used to handle logical components but deployment diagrams are made to focus on the hardware topology of a system. Deployment diagrams are used by the system engineers.

The purpose of deployment diagrams can be described as –

- Visualize the hardware topology of a system.
- Describe the hardware components used to deploy software components.
- Describe the runtime processing nodes.

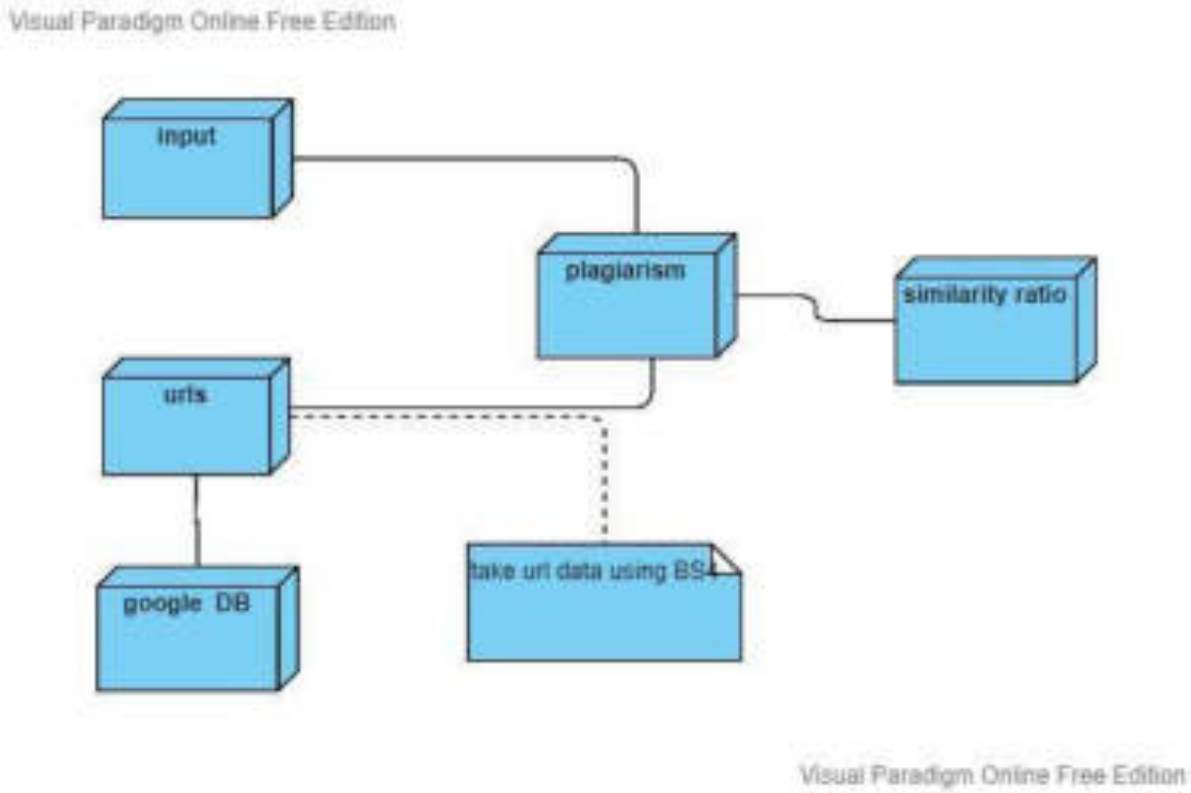


Fig:4.8 Deployment Diagram of Plagiarism Checker

5. IMPLEMENTATION

5.1 TECHNOLOGIES

5.1.1 INTRODUCTION TO PYTHON

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

Often, programmers fall in love with Python because of the increased productivity it provides. Since there is no compilation step, the edit-test-debug cycle is incredibly fast. Debugging Python programs is easy: a bug or bad input will never cause a segmentation fault. Instead, when the interpreter discovers an error, it raises an exception. When the program doesn't catch the exception, the interpreter prints a stack trace. A source level debugger allows inspection of local and global variables, evaluation of arbitrary expressions, setting breakpoints, stepping through the code a line at a time, and so on. The debugger is written in Python itself, testifying to Python's introspective power. On the other hand, often the quickest way to debug a program is to add a few print statements to the source: the fast edit-test-debug cycle makes this simple approach very effective.

5.1.2 PYTHON ENVIRONMENT

By default, any Python interpreter that you've installed runs in its own global environment, which is not specific to any one project. For example, if you just run python (Windows) or python3 (macOS/Linux) at a new command prompt, you're running in that interpreter's global environment. Accordingly, any packages that you install or uninstall affect the global environment and all programs that you run within that context.

Although working in the global environment is an easy way to get started, that environment will, over time, become cluttered with many different packages that you've installed for different projects. Such clutter makes it difficult to thoroughly test an application against a specific set of packages with known versions, which is exactly the kind of environment you'd set up on a build server or web server.

For this reason, developers often create a virtual environment for a project. A virtual environment is a subfolder in a project that contains a copy of a specific interpreter. When you activate the virtual environment, any packages you install are installed only in that environment's subfolder. When you then run a Python program within that environment, you know that it's running against only those specific packages. Be aware that if you're not using a virtual environment, and you have multiple versions of Python installed and set in the path environment variable, you might need to specify the Python interpreter to use in the terminal for installing packages to the global environment.

5.1.3 WHY WE USE PYTHON FOR PROGRAMMING

Python's expansive library of open source data analysis tools, web frameworks, and testing instruments make its ecosystem one of the largest out of any programming community.

Python is an accessible language for new programmers because the community provides many introductory resources. The language is also widely taught in universities and used for working with beginner-friendly devices such as the Raspberry Pi.

5.1.4 IS PYTHON PROGRAMMING AN EASY LANGUAGE TO LEARN

Python is the easiest programming language to learn. As compared with other languages, python code is small and uses with less lines of code, which is perfectly easy to learn, execute and understand. The syntax of python is very easy to know, learn & implement as compared with other programming languages.

Many program developers use this language to develop websites, create algorithms, or perform any other tasks. And as many MNC’s look for a candidate who has advanced knowledge about Python, people are moving in advancing their future in python and are derived towards learning it.

So, if you are one of them or you want to secure your future by learning python the 3RI Technologies is the best of all. Here you will get to hone your knowledge from basic to advanced level.

5.1.5 APPLICATIONS OF PYTHON PROGRAMMING

1. Web Development

Python is widely used in web development. It makes use of simple code to build a beautiful and functional website. It has related libraries and modules like HTTP, XML, JSON, IMAP, POP, FTP, etc. These help in storage in databases, content management, and interfacing with internet protocols.

It also has frameworks integrated with it like Django, Flask, Pyramid, and Bottle. In addition, Python also comes with packages like Beautiful Soup, Twisted Python, and Feedparser.

2. Artificial Intelligence and Machine Learning

Many AI and ML models are built using Python. Data analysts use it too. The fundamental cause for this is the availability of many tools and libraries specific to these applications. Some of them include:

- Pandas for Data Manipulation and Analysis.
- Numpy and SciPy for mathematical computations.
- Seaborn and Matplotlib for visualization.
- Scikit learn, Tensor Flow, Keras for algorithms

3. Game Development

One of the interesting applications of the Python Programming language is to build games like egg catcher, snake game, etc.

Python comes with packages like Pygame and Pykyra to build awesome games. It facilitates the development of 3D games with respective libraries. It is also used for scripting in many game engines.

Famous games like Battlefield 2, World of Tanks, Toontown Online, and Civilization 4 are also built using Python.

4. Desktop GUI Development

Its simplicity and platform independence nature make it a good choice for GUI applications. Though Tkinter is the standard library for GUI development, there are other supporting libraries like:

- Wx Python
- Kivy
- PyQt , PySide
- PyGUI
- Delphi
- GTK+

5. Enterprise/ Business Applications

Business applications are developed keeping in mind the comfort and ease of usage by the organization. Python's features like scalability, extensibility, and readability make it the best choice for business applications.

The following used Python:

- Odoo: This is an all-in-one business management software for enterprising purposes.
- Tryton: general-purpose, high level, third-tier business application

6. Software Development

Python supports the development of software with its libraries and packages. It is used for building, testing, management, and so on. For example

- Scons help in build control
- Buildbot and Apache are used for autonomous and continuous testing.
- Roundup and Trac test and manage errors.

7. Educational Purposes

Python helps beginners get introduced to the Programming environment. It is because of its simple syntax and beginner-friendly nature. Another reason could be its scope and reach in the development community.

8. Operating Systems Development

Many operating systems are using Python as a backbone. Generally, it is used along with C. Some of them are as follows:

- Linux based Ubuntu's Ubiquity Installer
- Anaconda Installer of Red Hat Enterprises

9. Language Development

Yes! Python being one of the programming languages helps in building other languages. These have similar syntax object models and indentation as compared to Python. Its architecture is taken as inspiration to build:

- Swift
- Boo
- Cobra
- CoffeeScript
- OCaml

10. Networking Programming

Besides these high-level facilities, it also supports low-level network programming. The framework TwistedPython serves for this purpose. It also avails a socket interface for ease.

11.Scientific and Numeric Computations

Python can also serve in doing scientific and numeric calculations. Thanks to its concerned libraries that makes these calculations easier. Apps like Abacus and FreeCAD are built using Python. Some of the libraries that can do these calculations are

- NumPy
- SciPy
- Pandas
- Natural language Toolkit

12. Web Scraping

Nowadays, Python is being widely used for web scraping purposes. We can access the huge data over the internet using these web scraping techniques. And this data is useful in different ways like for analysis purposes.

Some of the tools that are available in Python for scraping are:

- BeautifulSoup
- Python Request
- Scrapy
- MechanicalSoup
- Selenium

13. Image Processing and Graphic Designing

Using Python, one can process images using libraries like Opencv and Scikit Image. These have become popular recently for image analysis, especially in the research field.

Python is also used in animation fields. Many 2D applications like GIMP, Paint Shop Pro and Scribus and 3D animation applications like Blender, Maya, Lightwave, and 3ds Max use Python at the backend.

14.Text Processing

Text processing means automatic processing of electronic text. It involves tasks like searching, formatting, generating content, and filtering based on text.

You would have heard of NLP (Natural Language Processing). If you have not heard, you would have at least used it. The suggestions our mobile's keyboard gives is because of NLP. This is one of the text processing applications using Python.

15. Data Science and Data Analysis

Data Science and Data Analysis are two of the booming fields in the market. This involves analysis of a huge amount of data, finding relations and doing future predictions. Python modules like NumPy, Pandas, Matplotlib and Seaborn serve this purpose.

16. Audio and Video Applications

We can use Python to build audio and video player applications. Also these apps built using Python show better performance compared to that of the other ones. Some examples include CPlay and TimePlayer.

17. Integration to Embedded Systems

Embedded Systems and IoT are becoming popular and necessary in many places. These use Embedded C as the backend language. Since Python is built on C Programming language, it can communicate with these devices. One such famous device that uses Python is Raspberry Pi.

18. Building CAD Apps

CAD (Computer aided design) has become a common tool for engineers and architects. They use it to creare designs, which needs high end tools to build. Python with its user-friendly features serves for this purpose. One of the CAD applications that uses Python is Fandango.

19. Access to Database

Python can also integrate with databases. Python has:

- Interfaces to database management systems like MYSQL, Oracle, MS SQL server, PostgreSQL, and so on.
- Standard Database API.
- Connection to object databases like ZODB and Durus

20. Robotics

In robotics, Python is a popular language used by both hobbiestics and companies. One of its uses include Raspberry Pi to build robotic models. Python is also used in companies for robotic process automation (RPA) to build working arms and other models.

21. Console based Apps

Python serves as a suitable choice to build console-based applications because of its REPL (ReadEval-Print Loop) principle. Also, the presence of libraries that can handle command-line arguments adds to this.

Some advanced libraries in Python can also be used to build console-based Applications.

5.1.6 PYTHON ENVIRONMENT SETUP

Python is available on a wide variety of platforms including Linux and Mac OS X. Let's understand how to set up our Python environment.

Local Environment Setup

Open a terminal window and type "python" to find out if it is already installed and which version is installed.

- Unix (Solaris, Linux, FreeBSD, AIX, HP/UX, SunOS, IRIX, etc.)
- Win 9x/NT/2000
- Macintosh (Intel, PPC, 68K)
- OS/2
- DOS (multiple versions)
- PalmOS
- Nokia mobile phones
- Windows CE
- Acorn/RISC OS
- BeOS
- Amiga
- VMS/OpenVMS
- QNX
- VxWorks
- Pison
- Python has also been ported to the Java and .NET virtual machines

Getting Python

The most up-to-date and current source code, binaries, documentation, news, etc., is available on the official website of Python https://www.python.org/

You can download Python documentation from https://www.python.org/doc/. The documentation is available in HTML, PDF, and PostScript formats.

Installing Python

Python distribution is available for a wide variety of platforms. You need to download only the binary code applicable for your platform and install Python.

If the binary code for your platform is not available, you need a C compiler to compile the source code manually. Compiling the source code offers more flexibility in terms of choice of features that you require in your installation.

Here is a quick overview of installing Python on various platforms –
Unix and Linux Installation
Here are the simple steps to install Python on Unix/Linux machine.

- Open a Web browser and go to https://www.python.org/downloads/.
- Follow the link to download zipped source code available for Unix/Linux.
- Download and extract files.
- Editing the Modules/Setup file if you want to customize some options.
- run ./configure script
- make
- make install

This installs Python at standard location /usr/local/bin and its libraries at /usr/local/lib/pythonXX where XX is the version of Python.

Windows Installation

Here are the steps to install Python on Windows machine.

Open a Web browser and go to https://www.python.org/downloads/.

Follow the link for the Windows installer python-XYZ.msi file where XYZ is the version you need to install.

To use this installer python-XYZ.msi, the Windows system must support Microsoft Installer 2.0. Save the installer file to your local machine and then run it to find out if your machine supports MSI.

Run the downloaded file. This brings up the Python install wizard, which is really easy to use. Just accept the default settings, wait until the install is finished, and you are done.

Macintosh Installation

Recent Macs come with Python installed, but it may be several years out of date. See http://www.python.org/download/mac/ for instructions on getting the current version along with extra tools to support development on the Mac. For older Mac OS's before Mac OS X 10.3 (released in 2003), MacPython is available.

Jack Jansen maintains it and you can have full access to the entire documentation at his website − http://www.cwi.nl/~jack/macpython.html. You can find complete installation details for Mac OS installation.

Setting up PATH

Programs and other executable files can be in many directories, so operating systems provide a search path that lists the directories that the OS searches for executables.

The path is stored in an environment variable, which is a named string maintained by the operating system. This variable contains information available to the command shell and other programs.

The path variable is named as PATH in Unix or Path in Windows (Unix is case sensitive; Windows is not).

In Mac OS, the installer handles the path details. To invoke the Python interpreter from any particular directory, you must add the Python directory to your path.

5.17 PYTHON DATA TYPES

In programming, data type is an important concept.

Variables can store data of different types, and different types can do different things.

Python has the following data types built-in by default, in these categories:

- Text Type: str
- Numeric Types: int, float, complex
- Sequence Types: list, tuple, range
- Mapping Type: dict
- Set Types: set, frozenset
- Boolean Type: bool
- Binary Types: bytes, bytearray, memoryview.

5.2 CODE

5.2.1 STRING OR TEXT AS INPUT

```
from googlesearch import search

import requests

from bs4 import BeautifulSoup

from difflib import SequenceMatcher

 import validators

import ssl

import sys

from requests.packages.urllib3.exceptions import InsecureRequestWarning
```

```
def google_scrape(url):
        headers={
            'User-Agent':'chrome/64.0.3282.186(Windows NT 10.0; WOW64; rv:62.0)
Gecko/20100101 Firefox/100.0',
        }
  requests.packages.urllib3.disable_warnings(InsecureRequestWarning)
html=requests.get(url,verify=False,headers=headers)
#html.encoding('utf-8')
soup=BeautifulSoup(html.text, 'html.parser')
 return soup.text
queryf=[]
i=1
query=input('enter a string:')
queryf=query.split()
if len(queryf)>=5:
    for url in search(query,10):
        valid=validators.url(url)
        if valid==True:
            a=google_scrape(url)
            f=open("texttest.txt","w",errors='ignore')
            f.write(a)
            file=open("texttest.txt","r")
            data=file.readlines()
            str2=''.join(data)
```

```
similarity=SequenceMatcher(None,query,str2).ratio()
if similarity>0.0:
      print(f"the contents of {url} are {similarity} common")
       i=i+1
       f.truncate(0)
else:  print("enter a string that contains greater than equal to 5 words")
```

5.2.2 FILE AS INPUT

```
from googlesearch import search
import requests
from bs4 import BeautifulSoup
 from difflib import SequenceMatcher
import validators
import ssl
import sys
from requests.packages.urllib3.exceptions import InsecureRequestWarning
def google_scrape(url):
        headers={
            'User-Agent':’ chrome/64.0.3282.186 (Windows NT 10.0; WOW64; rv:62.0)
        Gecko/20100101  Firefox/100.0',
           }
requests.packages.urllib3.disable_warnings(InsecureRequestWarning)
html=requests.get(url,verify=False,headers=headers)
#html.encoding('utf-8')
 #html.fromstring(req.content.decode(encoding="utf-8", errors="ignore"))
```

```
soup=BeautifulSoup(html.text, 'html.parser')
return soup.text
i=1
f1=open(input("enter a file name:"),'r')
text1=f1.readlines()
str1=''.join(text1)
f1.close()
if len(str1)>=5:
        for url in search(str1,10):
            valid=validators.url(url)
             if valid==True:
                a=google_scrape(url)
                 f=open("filetest.txt","w",errors='ignore')
                  f.write(a)
                  file=open("filetest.txt","r")
                  data=file.readlines()
                  str2=''.join(str1)
        similarity=SequenceMatcher(None,str1,str2).ratio()
        if similarity>0.0:
                  print(f"the contents of {url} are {similarity} common")
                  i=i+1
                  f.truncate(0)
        else:
                  print("enter a string that contains greater than equal to 5 words")
```

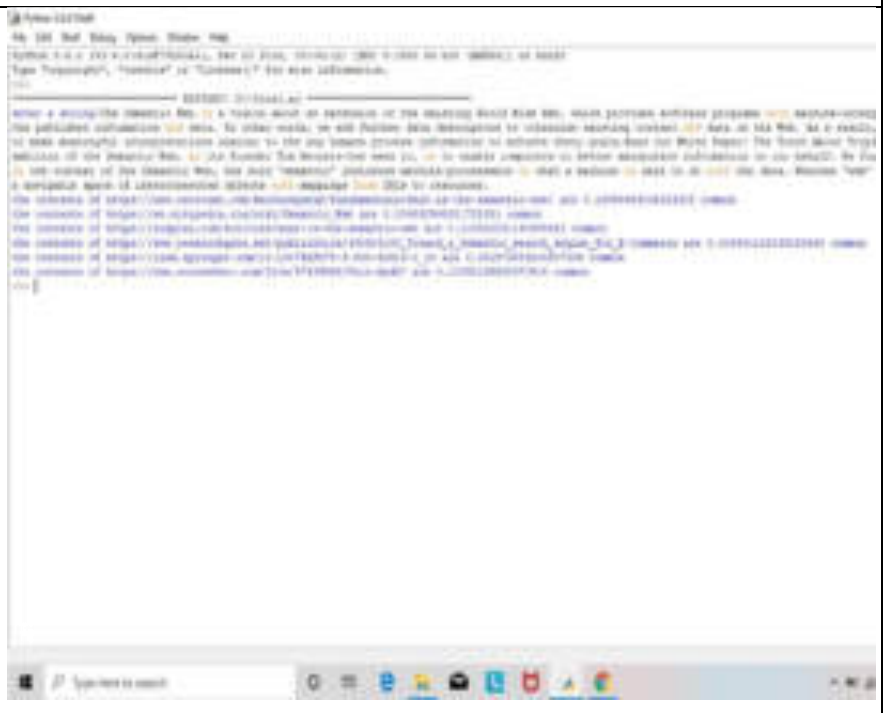
5.2.3 TWO FILES AS INPUT

```
from difflib import SequenceMatcher
with open(input("enter a file name1:")) as f1, open(input("enter a file name 2:")) as f2:
    f1_data=f1.read()     f2_data=f2.read()
    similarity=SequenceMatcher(None,f1_data,f2_data).ratio()
    print(f"the contents are {similarity*100}% common")
```

6. SYSTEM TESTING

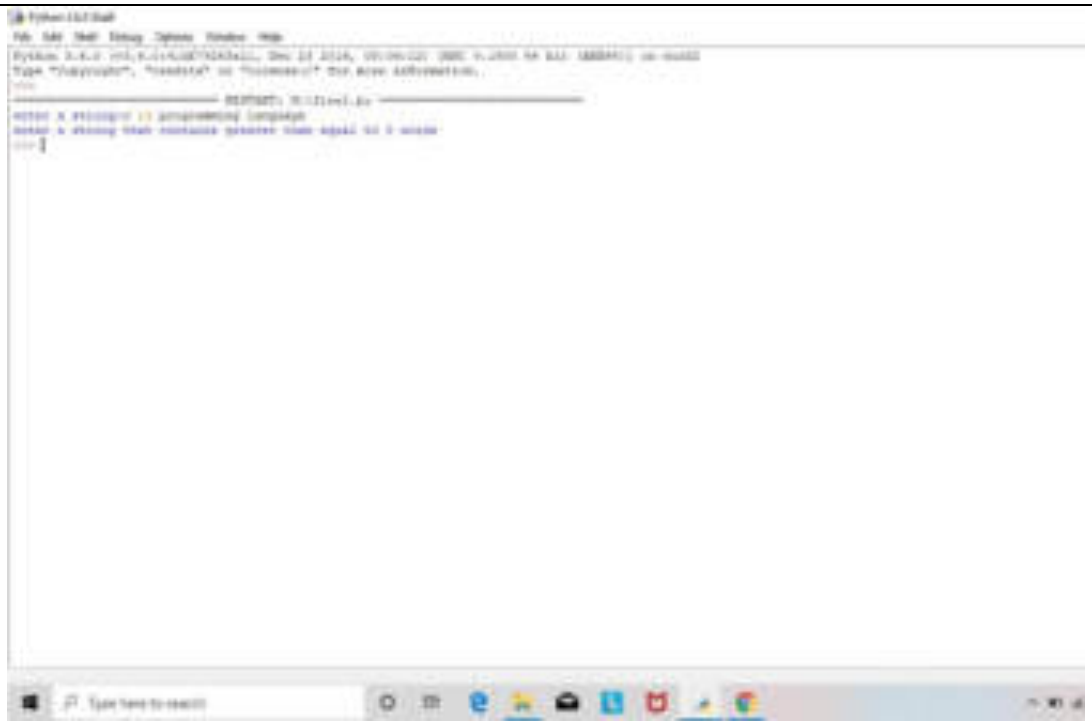
6.1 TEST CASES:

Test_case_id: 01

Description	When we give the input as string with greater than or equal to 5 Words it produces the URL's with similarity ratio.
Input	Give input as string with greater than 5 words.
Expected Output	Produces the URL's with similarity ratio .
Actual Output	Produces the URL's with similarity ratio.
Remarks	Pass.
Screen Shot	

Testcase: 6.1 Test case describes about URL's with similarity ratio of input string with length greater than or equal to 5.

Test_case_id: 02

Descripti on	When we give the input as string with less than words it produces output as enter a string that contains greater than equal to 5 words.
Input	Given input as “c is programming language”.
Expected Output	Enter the string that contains greater than equal to 5 words.
Actual Output	Enter the string that contains greater than equal to 5 words.
Remarks	Pass
Screen Shot	

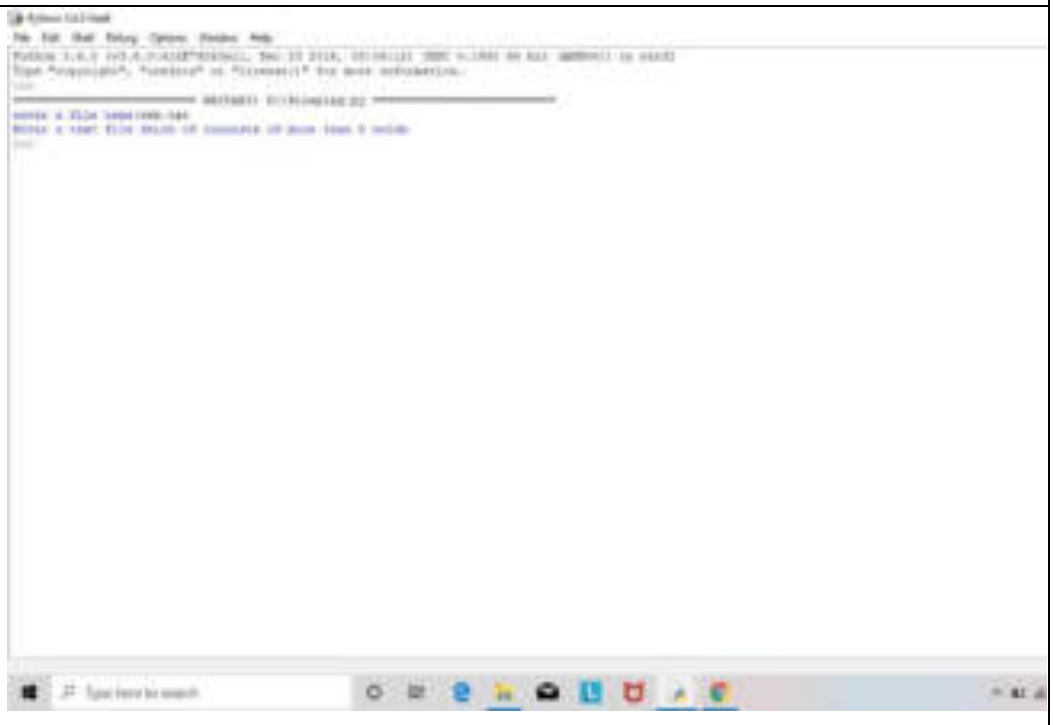
Testcase: 6.2 Test case describes input as string length less than 5.

Test_case_id: 03

Descripti on	When we give the input as file with greater than or equal to 5 Words it produces the links with similarity ratio.
Input	Given input file “inptest.txt”.
Expected Output	Produces the URL’s with similarity ratio.
Actual Output	Produces the URL’s with similarity ratio.
Remarks	Pass.
Screen Shot	

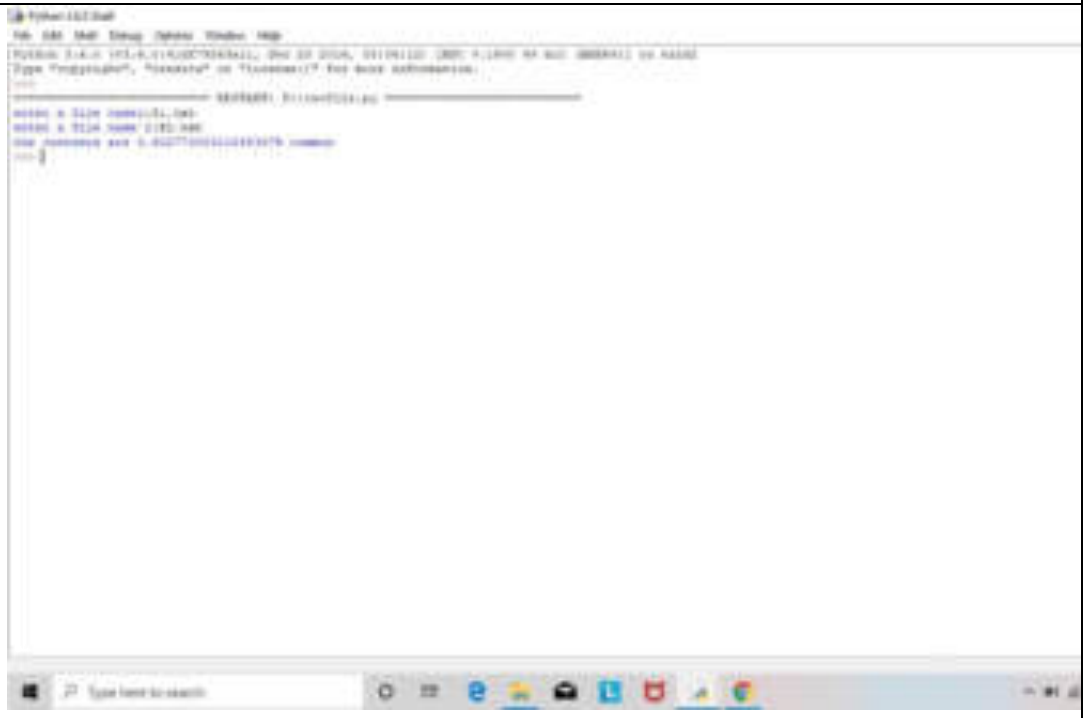
Testcase: 6.3 Test case describes about URL’s with similarity ratio of input file with length greater than or equal to 5.

Test_case_id: 04

Description	When we give the input as file with less than words it produces output as enter a string that contains greater than equal to 5 words.
Input	Give input file “web.txt”.
Expected Output	Enter the string that contains greater than equal to 5 words
Actual Output	Enter the string that contains greater than equal to 5 words
Remarks	Pass
Screen Shot	

Testcase: 6.4 Test case describes input file with length less than 5.

Test_case_id: 05

Descripti on	When we give the input as 2 files it produces the ratio of similarity.
Input	Give input files “f1.txt”. “f2.txt”.
Expected Output	Produces as the similarity ratio.
Actual Output	Produces as the similarity ratio.
Remarks	Pass
Screen Shot	

Testcase: 6.5 Test case describes input as two files.

7. EXPERIMENTAL RESULTS

7.1 SCREEN SHOTS

7.1.1 SIMILARITY RATIO OF INPUT STRING

When we give the input as string with greater than or equal to 5 Words it produces the URL's with similarity ratio.

Input: The input is given as string to retrieve the URL's with Similarity ratio of the string**.**

Output: It retrieve the URL's with Similarity ratio.

Fig 7.1 Represents the output for input string

7.1.2 SIMILARITY RATIO OF INPUT FILE

When we give the input as file with greater than or equal to 5 Words it produces the links with similarity ratio

Input: The input is given as file to retrieve the URL's with Similarity ratio of the file.

Output: It retrieve the URL's with Similarity ratio.

Fig 7.1 Represents the output for input file

7.1.3 SIMILARITY RATIO OF TWO FILES

When we give the input as file with less than words it produces output as enter a string that contains greater than equal to 5 words.

Input: The input is given as two files to retrieve Similarity ratio of the two files.

Output: It retrieve the Similarity ratio.

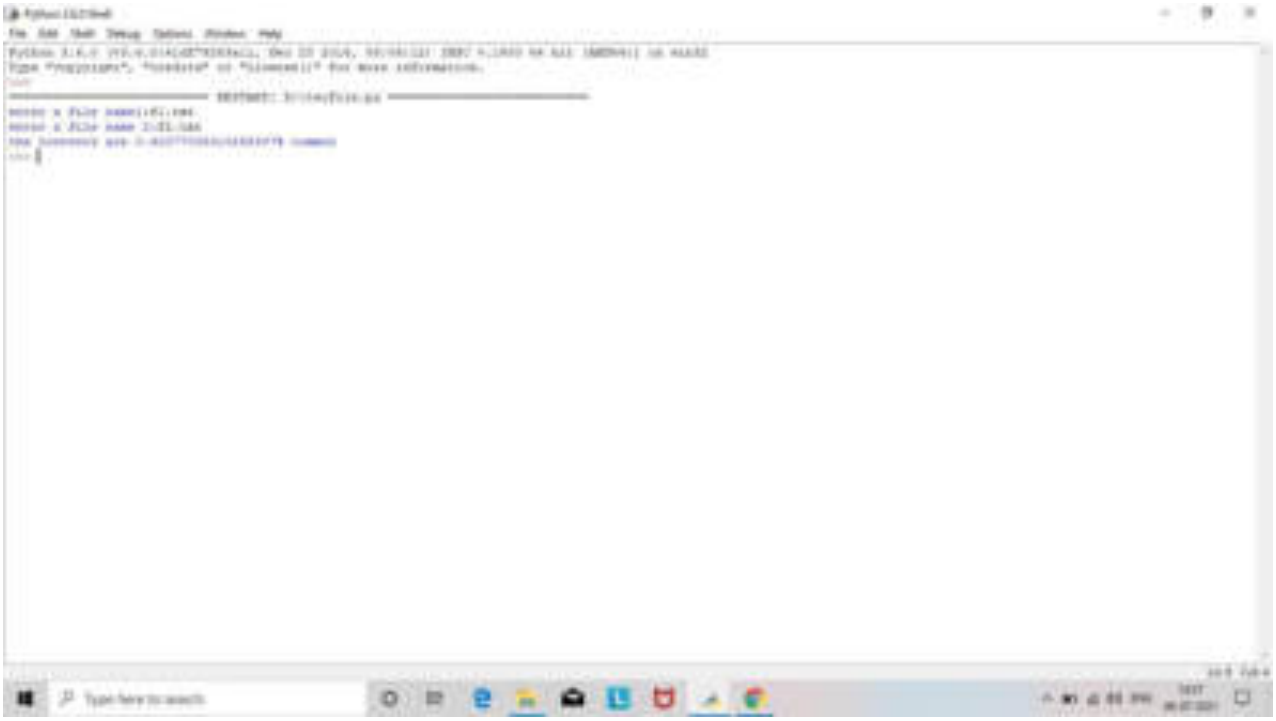


Fig 7.1 Represents the output for input two files

8.CONCLUSION AND FUTURE SCOPE

8.1 CONCLUSION

Nowadays plagiarism has become a most serious and major issue of the literary world. Most victims who fall in the trap of plagiarism are students, because students have to produce a number of essays and research papers within the give deadline, they don't want to lose marks and sometimes it became difficult to sustain the pressure of work.

Hence, students go for an easy and less time consuming way out to complete their assignments and apply a simple formula of copy-pasting other writer's original content and present as their own work or present the previous content as new and original content. That's why Plagiarism checker software has becomes essential for academies, educational institutions and universities. It is also important for promoting original content writings .

As the above discuss, different cultures, different forms of pressure and procrastination all are the main causes of why people choose to plagiarize. Therefore, students' willingness, value tendency and source of stresses can directly influence the percentage of students' plagiarism. Learning citation, paraphrasing and summarizing also can help people to avoid plagiarism but the problem is not every international students are able to handle these methods. The result is students should be a better academic writer if they want to avoid plagiarism. If those English speaking countries can solve the above problems and enable students to understand the value of doing assignments, less plagiarism will appear in the future.

8.2 FUTURE SCOPE

In future it is very useful because the technologies are improving the students are becoming very lazy day by day and they always wants to copy the contents. So by improving this technology we can strictly prohibit the plagiarism.

I'd like to examine three sectors that could benefit from a similar plagiarism checker product – art, music, and video. All of these industries currently have a number of outstanding legal cases and disputes over plagiarized content. With the right technology, plagiarism in all of these fields could be minimized through detection and checking prior to distribution

9.REFERENCES

[1] G. Whale, “Plague : plagiarism detection using program structure,” Dept. of Computer Science Technical Report 8805, University of NSW,Kensington, Australia, 2008.

[2] M. J. Wise, “Detection of Similarities in Student Programs: YAP'ing may be Preferable to Plague'ing,” ACM SIGSCE Bulletin (proc. Of 23rd SIGCSE Technical Symp.), 2002.

[3] P. Lutz, M. Guido, and M. Phlippsen, “JPlag: Finding plagiarisms among a set of programs,”Fakultätfür Informatik Technical Report 2000-1, Universität Kalrsruhe, Karlsruhe, Germany, 2000. International Journal of Computer Theory and Engineering Vol. 4, No. 2, April 2012 187.

[4] M. J. Wise, “YAP3: improved detection of similarities in computer program and other texts,” Proc. of SIGCSE’96 Technical Symposium,2006.

[5] L. Prechelt, G. Malpohl, and M. Philippsen, “Finding plagiarisms among a set of programs with JPlag,” Journal of Universal Computer Science, 2008.

[6] D. Gitchell and N. Tran, “Sim: a utility for detecting similarity in computer programs,” the 30th SIGCSE Technical Symposium on Computer Science Education, 2006.

[7] M. Mozgovoy, K. Fredriksson, and D. White, “Fast plagiarism detection system,” Lecture Notes in Computer Science, 2005.

[8] J. A. Faidhi and S. K. Robinson, “An empirical approach for detecting program similarity and plagiarism within a university programming environment,” computer education, 2007.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

Security Provision Through Fraud Detection

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
COMPUTER SCIENCE AND ENGINEERING

Submitted by

I Tejasree (17H71A05B4)

I Navya (17H71A0581)

S Sailusha (17H7A05A6)

G Santhoshi (17H71A05A7)

Under the Esteemed Guidance of

Ms. Srilakshmi

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC A
With Grade

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"Security Provision Through Fraud Detection"** is a bonafide work carried out by **I. Tejasree (17H71A05B4), I. Navya (17H71A0581), S. Sailusha(17H71A05A6), G. Santhoshi(17H71A05B4)** in partial fulfilment for the award of degree of Bachelor of Technology in **Computer Science and Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

Ms. Srilakshmi
Project Guide

Mr.D. Prasad
Head of the Department

Examiner

Dr K Srinivas
PRINCIPAL

ACKNOWLEDGEMENT

We would like to take this opportunity to express my deepest appreciation to the following people for their valuable contributions and assistance with this Project.

Initially, we would like to thank my project supervisor, **Ms. Srilakshmi, Assistant Professor, Department of Computer Science and Engineering** for the guidance and support, especially for the valuable ideas and knowledge shared to me throughout the Project.

We have the immense pleasure in expressing my thanks and deep sense of gratitude to**, Mr. D. Prasad, Head of the Department, Computer Science and Engineering** for extending necessary facilities and support for the completion of the Project.

We whole sincerely acknowledge **Dr. K. Srinivas, Principal and Prof D. PANDURANGA RAO, CEO** for giving opportunity to take up the Project work. We also extend my thanks to all faculty members of **Computer Science and Engineering**, for their valuable guidance and encouragement.

We would like to extend my warm appreciation to all my friends for sharing us their knowledge, valuable contributions and help with this Project.

Finally, my special thanks go to my family for their continuous support and help throughout my academic years and for their continual support and encouragement for the completion of the project.

DECLARATION

We **I. Tejasree (17H7A05B4), I. Navya (17H71A0581), S. Sailusha(17H71A05A6), G. Santhoshi (17H7A05A7)**of the Main-Project "**Security Provision Through Fraud Detection**", hereby declare that the matter embodied in this Project is the genuine work done by us and has not been submitted either to this university or to any other university/institute for the fulfilment of the requirement of any course of study.

I Tejasree (17H71A05B4)

I.Navya(17H71A0581)

S Sailusha(17H71A05A6)

G Santhoshi (17H71A05A7)

Abstract

Our project is to provide the security for the personal/private information and with hold the leakage of the data by giving perfect security and access information.

Using ML for the interaction or inter communication, it can be helpful in identifying the frauds and notify to prevent the intruders.

Our approach proposed the overall process of detecting mobile payment fraud based on machine learning, supervised and unsupervised method to detect fraud and process large amounts of financial data

Fraud Detection Machine Learning Algorithms Using Logistic Regression: Logistic Regression is a supervised learning technique that is used when the decision is categorical. It means that the result will be either 'fraud' or 'non-fraud' if a transaction occurs.

An important factor in the decision-making process to allocate resources to a public vs. private cloud is the fine-tuned control available in private cloud environments. In private clouds, additional levels of control and supplemental protection can compensate for other limitations of private cloud deployments and may contribute to a practical transition from monolithic server-based data centers.

Mobile payment fraud is the unauthorize use of mobile transaction through identity theft or credit card stealing to fraudulently obtain money. Mobile payment fraud is the fast growing issue through the emergence of smartphone and online transition services. In the real world, highly accurate process in mobile payment fraud detection is needed since financial fraud causes financial loss. Therefore, our approach proposed the overall process of detecting mobile payment fraud based on machine learning, supervised and unsupervised method to detect fraud and process large amounts of financial data. Moreover, our approach performed sampling process and feature selection process for fast processing with large volumes of transaction data and to achieve high accuracy in mobile payment detection.measure and ROC curve are used to validate our proposed model.

SECURITY PROVISION THROUGH FRAUD DETECTION

Chapter 1

INTRODUCTION

1.1 Introduction

Our project is to provide the security for the personal/private information and with hold the leakage of the data by giving perfect security and access information.

Using ML for the interaction or inter communication, it can be helpful in identifying the frauds and notify to prevent the intruders.

An important factor in the decision-making process to allocate resources to a public vs private cloud is the fine-tuned control available in private cloud environments. In private clouds, additional levels of control and supplemental protection can compensate for other limitations of private cloud deployments and may contribute to a practical transition from monolithic server-based data centers.

Mobile payment fraud is the unauthorized use of mobile transaction through identity theft or credit card stealing to fraudulently obtain money.

Mobile payment fraud is the fast growing issue through the emergence of smartphone and online transaction services.

In the real world, highly accurate process in mobile payment fraud detection is needed since financial fraud causes financial loss.

Therefore, our approach proposed the overall process of detecting mobile payment fraud based on machine learning, supervised and unsupervised method to detect fraud and process large amounts of financial data. Credit card fraud is a huge ranging term for theft and fraud committed using or involving at the time of payment by using this card. The purpose may be to purchase goods without paying, or to transfer unauthorized funds from an account. Credit card fraud is also an add on to identity theft. As per the information from the United States Federal Trade Commission, the theft rate of identity had been holding stable during the mid 2000s, but it was increased by 21 percent in 2008. Even though credit card fraud, that crime which most people

associate with ID theft, decreased as a percentage of all ID theft complaints In 2000, out of 13 billion transactions made annually, approximately 10 million or one out of every 1300 transactions turned out to be fraudulent.

1.2 Motivation

This model can further be improved with the addition of more algorithms into it. However, the output of these algorithms needs to be in the same format as the others. Once that condition is satisfied, the modules are easy to add as done in the code. This provides a great degree of modularity and versatility to the project.

More room for improvement can be found in the dataset. As demonstrated before, the precision of the algorithms increases when the size of dataset is increased. Hence, more data will surely make the model more accurate in detecting frauds and reduce the number of false positives. 0.05% (5 out of every 10,000) of all monthly active accounts was fraudulent. Today, fraud detection systems are introduced to control one-twelfth of one percent of all transactions processed which still translates into billions of dollars in losses. Credit Card Fraud is one of the biggest threats to business establishments today. However, to combat the fraud effectively, it is important to first understand the mechanisms of executing a fraud.

Card numbers generally the Primary Account Number (PAN) are often reprinted on the card, and a magnetic stripe on the back contains the data in machine-readable format. It contains the following Fields:

- Name of card holder
- Card number
- Expiration date
- Verification/CVV code
- Type of card

FUTURE ENHANCEMENTS

While we couldn’t reach out goal of 100% accuracy in fraud detection, we did end up creating a system that can, with enough time and data, get very close to that goal. As with any such project, there is some room for improvement here.

The very nature of this project allows for multiple algorithms to be integrated together as modules and their results can be combined to increase the accuracy of the final result.

This model can further be improved with the addition of more algorithms into it. However, the output of these algorithms needs to be in the same format as the others. Once that condition is satisfied, the modules are easy to add as done in the code. This provides a great degree of modularity and versatility to the project.

More room for improvement can be found in the dataset. As demonstrated before, the precision of the algorithms increases when the size of dataset is increased. Hence, more data will surely make the model more accurate in detecting frauds and reduce the number of false positives.

Chapter 2

LITERATURE SURVEY

Fraud act as the unlawful or criminal deception intended to result in financial or personal benefit. It is a deliberate act that is against the law, rule or policy with an aim to attain unauthorized financial benefit.

Numerous literatures pertaining to anomaly or fraud detection in this domain have been published already and are available for public usage. A comprehensive survey conducted by Clifton Phua and his associates have revealed that techniques employed in this domain include data mining applications, automated fraud detection, adversarial detection. In another paper, Suman, Research Scholar, GJUS&T at Hisar HCE presented techniques like Supervised and Unsupervised Learning for credit card fraud detection. Even though these methods and algorithms fetched an unexpected success in some areas, they failed to provide a permanent and consistent solution to fraud detection.

A similar research domain was presented by Wen-Fang YU and Na Wang where they used Outlier mining, Outlier detection mining and Distance sum algorithms to accurately predict fraudulent transaction in an emulation experiment of credit card transaction data set of one certain commercial bank. Outlier mining is a field of data mining which is basically used in monetary and internet fields. It deals with detecting objects that are detached from the main system i.e. the transactions that are not genuine. They have taken attributes of customers behaviour and based on the value of those attributes they have calculated that distance between the observed value of that attribute and its predetermined value.

Unconventional techniques such as hybrid data mining/complex network classification algorithm is able to perceive illegal instances in an actual card transaction data set, based on network reconstruction algorithm that allows creating representations of the deviation of one

instance from a reference group have proved efficient typically on medium sized online transaction.

There have also been efforts to progress from a completely new aspect. Attempts have been made to improve the alert- feedback interaction in case of fraudulent transaction.

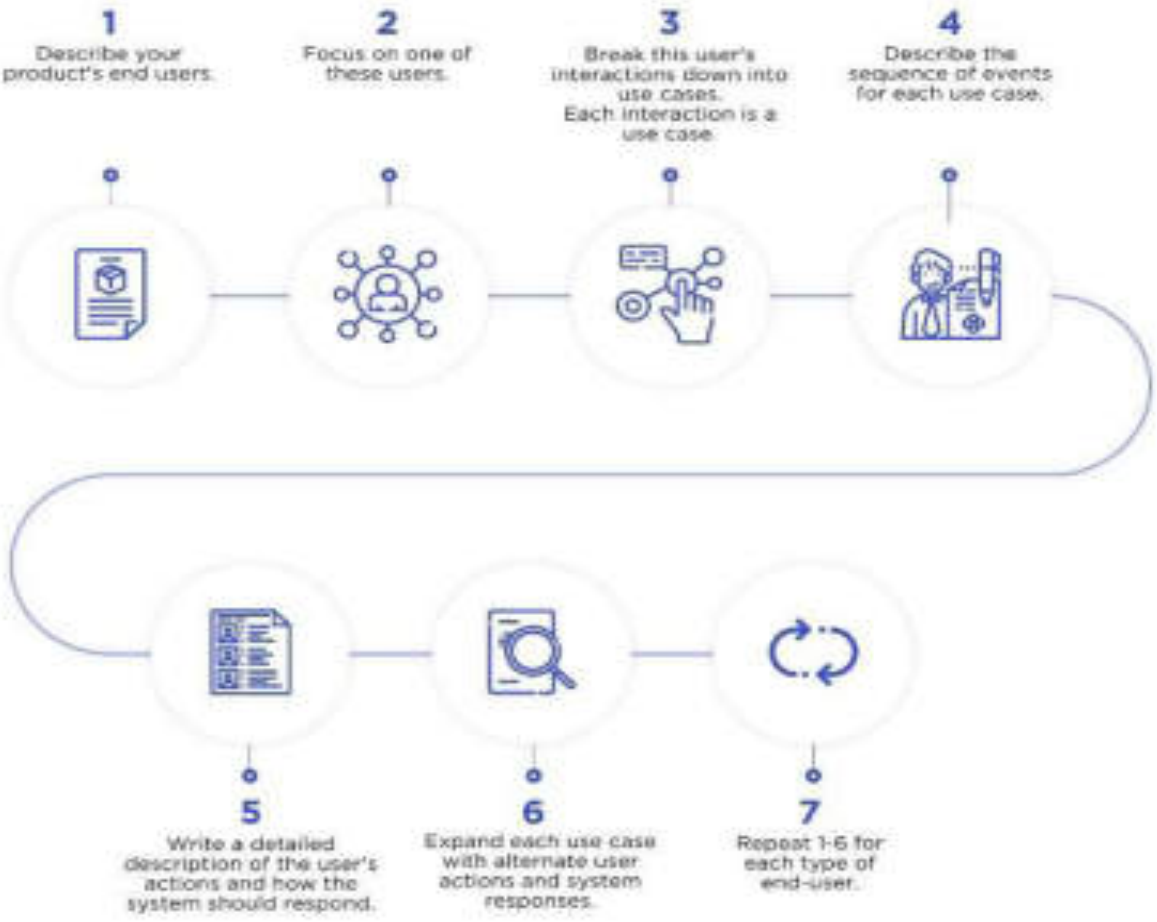
In case of fraudulent transaction, the authorized system would be alerted and a feedback would be sent to deny the ongoing transaction.

Artificial Genetic Algorithm, one of the approaches that shed new light in this domain, countered fraud from a different direction. It proved accurate in finding out the fraudulent transactions and minimizing the number of false alerts. Even though, it was accompanied by classification problem with variable misclassification costs.

Multiple Supervised and Semi-Supervised machine learning techniques are used for fraud detection [8], but we aim is to overcome three main challenges with card frauds related dataset i.e., strong class imbalance, the inclusion of labelled and unlabelled samples, and to increase the ability to process a large number of transactions.

Different Supervised machine learning algorithms [3] like Decision Trees, Naive Bayes Classification, Least Squares Regression, Logistic Regression and SVM are used to detect fraudulent transactions in real-time datasets. Two methods under random forests [6] are used to train the behavioural features of normal and abnormal transactions. They are Random-tree-based random forest and CART-based. Even though random forest obtains good results on small set data, there are still some problems in case of imbalanced data. The future work will focus on solving the above-mentioned problem. The algorithm of the random forest itself should be improved.

Performance of Logistic Regression, K-Nearest Neighbour, and Naïve Bayes are analysed on highly skewed credit card fraud data where Research is carried out on examining meta-classifiers and meta-learning approaches in handling highly imbalanced credit card fraud data. Through supervised learning methods can be used there may fail at certain cases of detecting the fraud cases. A model of deep Auto-encoder and restricted Boltzmann machine (RBM) [2] that can construct normal transactions to find anomalies from normal patterns.


1
Describe your product's end users.
2
Focus on one of these users.
3
Break this user's interactions down into use cases. Each interaction is a use case.
4
Describe the sequence of events for each use case.
5
Write a detailed description of the user's actions and how the system should respond.
6
Expand each use case with alternate user actions and system responses.
7
Repeat 1-6 for each type of end-user.

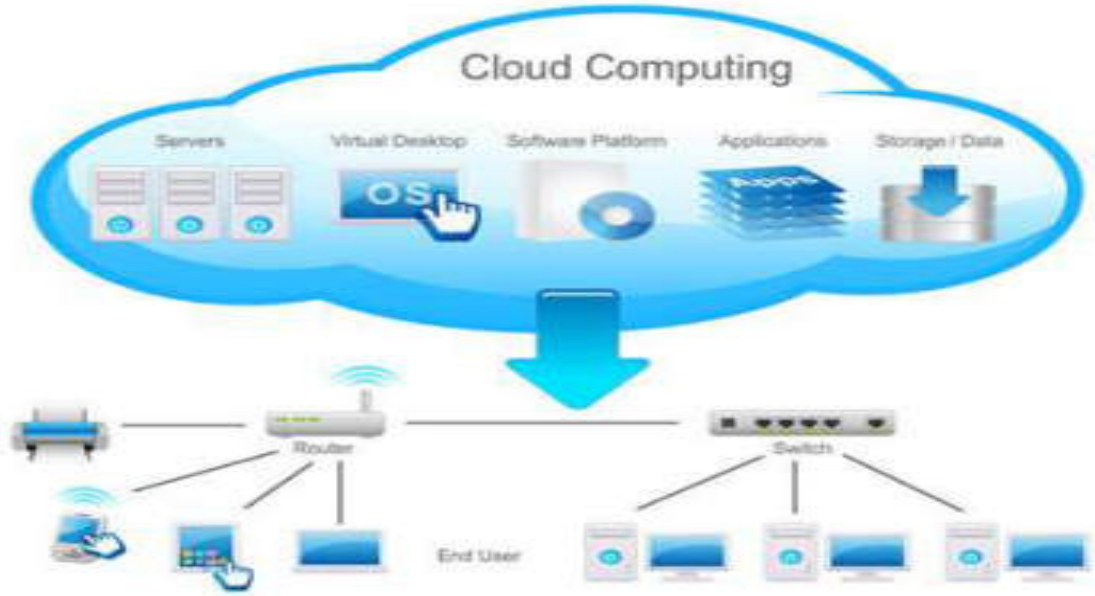
Chapter 3

CLOUD COMPUTING

Project domain

Cloud computing is the on-demand availability of computer system resources, especially data storage (cloud storage) and computing power, without direct active management by the user. The term is generally used to describe data centers available to many users over the Internet. Large clouds, predominant today, often have functions distributed over multiple locations from central servers. If the connection to the user is relatively close, it may be designated an edge server.

Clouds may be limited to a single organization (enterprise clouds), or be available to multiple organizations (public cloud).



Cloud computing relies on sharing of resources to achieve coherence and economies of scale.

Advocates of public and hybrid clouds note that cloud computing allows companies to avoid or minimize up-front IT infrastructure costs. Proponents also claim that cloud computing allows enterprises to get their applications up and running faster, with improved manageability and less maintenance, and that it enables IT teams to more rapidly adjust resources to meet fluctuating and unpredictable demand, providing the burst computing capability: high computing power at certain periods of peak demand.

Cloud providers typically use a "pay-as-you-go" model, which can lead to unexpected operating expenses if administrators are not familiarized with cloud-pricing models.

The availability of high-capacity networks, low-cost computers and storage devices as well as the widespread adoption of hardware virtualization, service-oriented architecture and autonomic

and utility computing has led to growth in cloud computing. As of 2017, most cloud computers run a Linux-based operating system.

Across the broader market, discussions about cloud security have focused primarily on the customer side of the equation. Even as cloud providers continue to devote the resources necessary to ensure that customer data is secure, they can't overlook the fact that some of their own customers could be a threat.

Fraud manifests in the cloud in several ways, according to experts. Typically, fraudsters use a stolen credit card to procure virtual machine (VM) instances or platform services on which they build their operations – among them phishing schemes, money-transfer scams, identity theft and malware.

"[You] can go get a fraudulent credit card, a good one – it'll be working, but it'll be stolen – for less than a dollar," Rowell said. "So, think about how the cloud enables [criminals]. All they have to do is sign up online and they can have a server in five minutes for less than a buck, and it's a throwaway identity."

In a joint investigation in 2012, researchers from McAfee Labs and Guardian Analytics uncovered a massive, cloud-based banking fraud operation that attempted to bilk an estimated $78 million from account holders in Europe, Latin America and the United States. The investigation, dubbed "Operation High Roller" because of the criminals' focus on high-balance accounts, found the scheme's success hinged on the resource availability and automation in the cloud, as opposed to a single host computer.

"With no human participation required, each attack moves quickly and scales neatly," investigators wrote in a report.

In some cases, criminals skip the stolen credit cards altogether and instead crack into a legitimate customer's account, hijacking the VMs to use for their own fraudulent activities. Cybercriminals are also looking to Infrastructure as a Service to provide vast amounts of on-demand processing power to launch distributed-denial-of-service attacks, according to Raj Samani, vice president and chief technology officer of McAfee Inc.'s EMEA operations.
Service offered via the internet are generalized as cloud computing and is categorized as IaaS, PaaS, FaaS, and SaaS, it is mainly divided into two types I,e, Deployment and Service models where the deployment model consists of the cloud computing services like Public Cloud, Private Cloud, Private and Hybrid which is used for applications like blog hosting, creation of a fresh application, Video, and Audio stream, etc.

Types of Cloud Computing

There are mainly two types of Cloud Computing which are:

- Deployment Model

- Service Model

The following kinds of cloud computing may be available based on the Three types of deployment model:

1. Public
2. Community Cloud
3. Private
4. Hybrid

Public cloud

The public cloud is described as computing services provided through the public internet by third-party suppliers, making them accessible to those who want to use them or buy them. It can be free or on-demand for customers to pay for the cycles, storage or bandwidth they consume per usage.

Example: Sun Cloud, AWS, Microsoft Azure,

Community cloud

A particular group of customers from organizations with shared issues can only use cloud infrastructure. It may be owned, operated, managed and run by one or more of the communal organizations, a third party or a mixture of them.

Private Cloud

A private cloud is a cloud computing system in which IT services are supplied for the specialized use of one organization over private IT facilities**.** A single organization operates the cloud infrastructure only. It can be run on-site or off-site by the organization or a third party**.** Private cloud terms are often employed interchangeably with the virtual private cloud (VPC). Technically speaking, a VPC is a private cloud that uses the infrastructure of a third-party cloud provider, while an inner cloud is enforced.

Example: AWS, VMware

Hybrid Cloud

A hybrid cloud is a computer environment that incorporates a government and private cloud to share information and apps. Organizations achieve flexibility and the computer capacity of a government cloud for fundamental and insensitive computing functions whilst safe behind a corporate firewall for business-critical apps and information.

Example:

When delicate data is engaged, Federal agencies opt for personal clouds

The following kinds of cloud computing may be available depending on the type of service model:

- LAAS
- PAAS
- SAAS
- FAAS

LaaS

LaaS is a cloud computing model in which companies' external cloud providers provide and manage virtualized infrastructure for businesses. With LaaS, businesses can outsource Internet-based storage, servers, data centers and network elements with the same functionality as local infrastructure. Some examples of LaaS's extensive use include automated policies such as support, clustering, inner networking, restoration, tracking, hosting the site, etc. The provider is responsible for server and storage building, networking of firewalls/security and physical data centers. The supplier is in charge of server and storage construction, firewall networking and safety and physical information centers.

PaaS

PaaS builds on LaaS. Cloud vendors are providing computing resources here to create and test apps, including cloud and hardware infrastructure parts, such as middleware and operating systems. In addition to constructing and keeping the required equipment, the PaaS environment allows cloud users to install and house information sets, development instruments, and company assessment software. Some of the main players with a PayPal system include CloudBees, Salesforce.com, Engine, Heroku, Google App, Bluemix.

SaaS

Software as a service (SaaS) is a distribution system of software where a third-party supplier hosts and provides apps to clients over the Internet. The SaaS host application management model is similar to ASP, in which the provider hosts the software of customers and delivers it over the internet to approved end-users. The provider provides network-based customers with a unique copy of an application that was specially designed by a provider for the SaaS distribution in the software on-demand SaaS model. The source code for the Application is the same for all clients and is implemented by all clients when fresh functions or functionalities are developed.

FaaS

FaaS stands for the Functions as a server. FaaS brings another layer of PaaS abstraction, making designers fully insulated from all the things under their software in the stack. Installation blocks of code in narrow functionality and set them to be activated by a certain event rather than the hazards associated with virtual servers, containers and application runtimes. Until an event happens, FaaS applications don't consume LaaS resources, lowering the charges.

Cloud Service Providers

1Google:

It has become a synonym for the word "search". People are noticed often saying that "Just Google it and you will find everything. But it is not the only thing Google provide as

service it also provides us the cloud services like: - G-mail, Google docs, Picasa, Google Analytics, Google Ad words and Ad sense.

2 Microsoft:

It provides its own platform for providing a set of cloud services offered to users and application developers. Services run in Microsoft Data centre. Services provided by Microsoft are: - Windows Azure, SQL Azure, Windows Azure App Fabric and Windows Azure Marketplace.

3 Amazon Web Services (AWS):

It provides a cloud computing platform for all business sizes. AWS helps business organization to choose their own computing platform as in need of the organization and pay for what they use.

Services provided AWS are:- Amazon Elastic compute cloud, Amazon Simple Storage services, Amazon Virtual Private Cloud, Amazon Cloud front, Amazon Relational Database and Amazon Simple Queue services.

Applications of Cloud Computing

Cloud Computing is one of the most dominant field of computing resources online because sharing and management of resources is easy using cloud. These properties have made it an active component in the following fields

1 E-Learning

It is a new trend in the field of education that provides an attractive environment for students, faculty members, and researchers. Students, faculty members, researchers can connect to the cloud of their organization and access data and information from there.

2 Enterprise resource planning (ERP)

Use of Cloud in ERP comes into existence when the business of any organization grows. The work of managing applications, human resources, payroll etc becomes expensive and complex. To overcome it service providers can install ERP in the cloud itself.

3 E-Governance Cloud

computing can improve the functioning of a government by improving the way it provides the services to its citizens, institutions and cooperation with other governments. This can be done by expanding the availability of environment, making environment more scalable and customized. It also cut off the burden of managing, installing and upgrading applications.

3.1 Detecting cloud fraud

Fraud detection and prevention is a delicate dance for cloud providers, which must balance customer privacy concerns with the need to snuff out illegal activities, according to John Howie, chief operating officer of the CSA and former head of data center security for a large cloud provider.

“Cloud providers have built up these very sophisticated, accurate and successful antifraud systems, and they’ve invested a lot of time and energy in it,” Howie said. “They monitor how the customers use the service without monitoring their data – instead, [they look] for patterns of activity that are indicators of [fraud].”

Once fraudulent behavior is detected, providers alert law-enforcement agencies and will even notify their competitors of patterns through anonymous forums the CSA hosts, Howie said. Sharing this information “has already developed very tangible results,” he noted.

But technology can only go so far:

The market is filled with software that can identify irregular behavior, such as a new customer who provisioned a VM that sent an unusually large number of emails, Howie said. The most effective line of defense is an antifraud team with the training to determine whether a credit card used to procure services is stolen and how to follow up with the customer, he added.

“A lot of fraud or potential fraud is really caught in that stage. It’s not high tech,” Howie said. “It’s really about customer management.”

Some argue it’s still unclear whether an abundance of safeguards is the solution.

“As a cloud provider, how onerous do you want to become?” asked McAfee’s Samani. “If every five minutes you’re [sending] an email saying, ‘You’re not allowed to do that because we think it’s malicious’ … [customers] are going to turn around and say, ‘This is too much effort. The cloud isn’t easy for us.’ It becomes a double-edged sword.”

Algorithm steps:

Step 1: Read the dataset.

Step 2: Random Sampling is done on the data set to make it balanced.

Step 3: Divide the dataset into two parts i.e., Train dataset and Test dataset.

Step 4: Feature selection are applied for the proposed models.

Step 5: Accuracy and performance metrics has been calculated to know the efficiency for different algorithms.

Step6: Then retrieve the best algorithm based on efficiency for the given dataset.

Decision Tree Algorithm:

Decision tree is a type of supervised learning algorithm (having a pre-defined target variable) that is mostly used in classification problems. It works for both categorical and continuous input and output variables. In this technique, we split the population or sample into two or more homogeneous sets (or sub-populations) based on most significant splitter / differentiator in input variables.

TYPES OF DECISION TREE

1. Categorical Variable Decision Tree:

   Decision Tree which has categorical target variable then it called as categorical variable decision tree.

2. Continuous Variable Decision Tree:

   Decision Tree has continuous target variable then it is called as Continuous Variable Decision Tree

Performance metrics:

The basic performance measures derived from the confusion matrix. The confusion matrix is a 2 by 2 matrix table contains four outcomes produced by the binary classifier. Various measures such as sensitivity, specificity, accuracy and error rate are derived from the confusion matrix.

Accuracy:

Accuracy is calculated as the total number of two correct predictions(A+B) divided by the total number of the dataset(C+D).It is calculated as (1-error rate).

Accuracy=A+B/C+D

Whereas,

A=True Positive

B=True Negative

C=Positive

D=Negative

Error rate:

Error rate is calculated as the total number of two incorrect predictions(F+E) divided by the total number of the dataset(C+D).

Error rate=F+E/C+D

Whereas,

E=False Positive

F=False Negative

C=Positive

D=Negative

Sensitivity:

Sensitivity is calculated as the number of correct positive predictions(A) divided by the total number of positives(C).

$$Sensitivity=A/C$$

Specificity:

Specificity is calculated as the number of correct negative predictions(B) divided by the total number of negatives(D).

$$Specificity=B/D.$$

Accuracy, Error-rate, Sensitivity and Specificity are used to report the performance of the system to detect the fraud in the credit card. In this paper, three machine learning algorithms are developed to detect the fraud in credit card system. To evaluate the algorithms, 70% of the dataset is used for training and 30% is used for testing and validation. Accuracy, error rate, sensitivity and specificity are used to evaluate for different variables for three algorithms as shown in Table 3. The accuracy result is shown for logistic regression; Decision tree and random forest classifier are 92.7, 95.8, and 97.6 respectively. The comparative results show that the Random forest performs better than the logistic regression and decision tree techniques.

TERMINOLOGY OF DECISION TREE:

1. Root Node: It represents entire population or sample and this further gets divided into two or more homogeneous sets.

2. Splitting: It is a process of dividing a node into two or more sub-nodes.

3. Decision Node: When a sub-node splits into further subnodes, then it is called decision node.

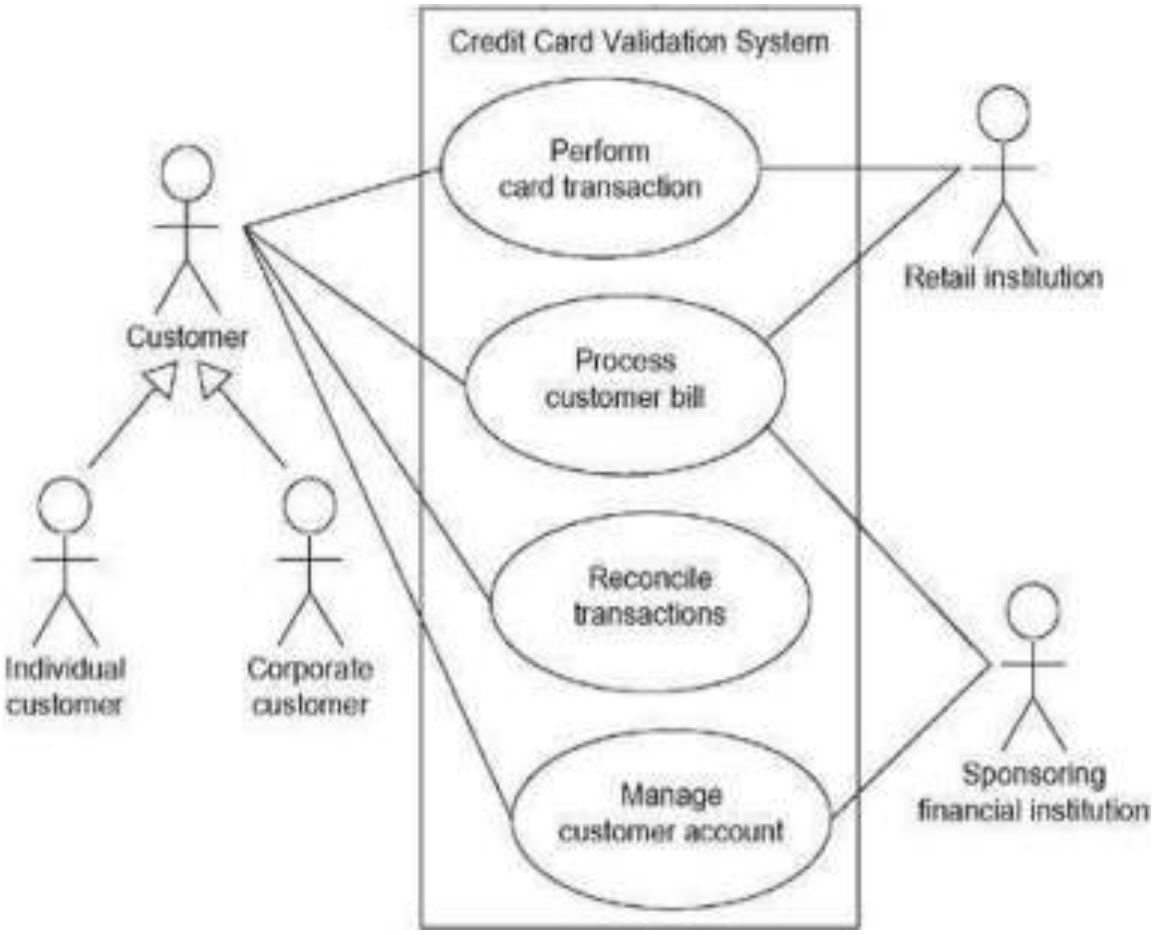
4. Leaf/ Terminal Node: Nodes do not split is called Leaf or Terminal node.

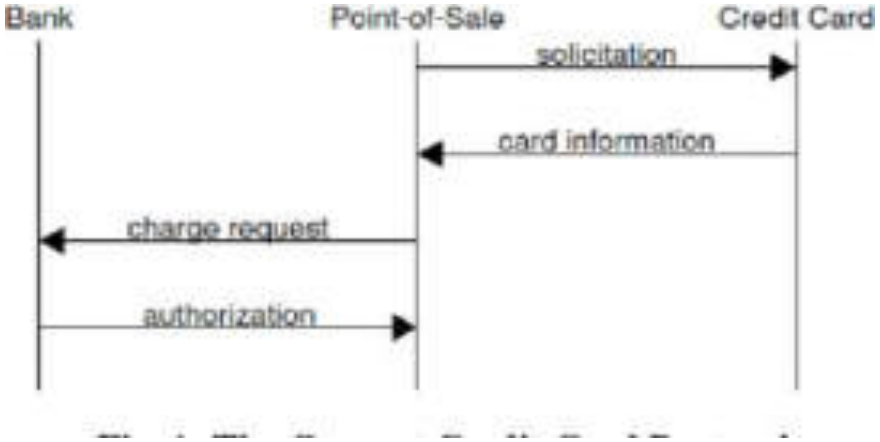
5. Pruning: When we remove sub-nodes of a decision node, this process is called pruning. You can say opposite process of splitting.

6. Branch / Sub-Tree: A sub section of entire tree is called branch or sub-tree.

7. Parent and Child Node: A node, which is divided into sub-nodes is called parent node of sub-nodes where as sub-nodes are the child of parent node.

DESIGN





Testing :

By Passing different types of data sets and run the program .

Implementing:

Linking the engine to the required web/application .

3.2CONSEQUENCES OF FAILURE TO DETECT FRAUD:

Although fraud may not be the gravest security threat cloud providers face, ignoring it jeopardizes their bottom line in several ways.

From a purely financial perspective, any revenue gained from a stolen credit card is likely to evaporate quickly, thanks to the sophisticated fraud detection systems banks and credit card companies now use. The real damage comes from the revenues cloud providers never see from legitimate customers because the hundreds of VMs they would have paid to access have been tied up by the fraudsters.

"[There are] service providers that … do not have adequate fraud measures in place, and they have to be losing insane amounts of money on it," said Dimension Data's Rowell. "It's got to have an immense impact to their profitability as well as just the health and cleanliness of their platform."
Moreover, cloud providers that don't commit resources to fraud detection and prevention could ruin their reputation – and kiss goodbye any chance to engage enterprise customers, Rowell added.

"If you were putting up a storefront, you wouldn't want to hang your shingle beside a shop that says, 'Hey, we're selling stolen credit cards.' No one wants to be associated with that," he said. "It's incumbent on the service provider industry to police fraud. If they're not doing it, they're doing their entire customer base a disservice."

Enterprises are also likely to block IP addresses from which spam and other suspicious activity originate, unintentionally blacklisting the cloud providers that host them.

While there is no legal precedent yet, it's possible that governments and law enforcement agencies may start holding cloud providers criminally or civilly responsible for neglecting to detect and eradicate fraud, said ISACA's Spivey.

"Depending on how big the problem becomes will determine whether regulators or lawmakers start to get more involved," he said. "But if I'm running a store, for instance, and I know people are coming into the store buying and selling drugs, and I never brought it up to people, then law

enforcement is basically going to [conclude] that I enabled this to occur because I let it happen on my premises."

Factors Affecting the Audit Process

Our framework identifies three main factors impacting the audit process:

(i) "institutional forces," which includes elements such as the regulatory regime, standard setting mechanisms, and peer reviews

(ii) "auditor knowledge, training and experience," which represents the intellectual capital the auditors bring to the engagement and

(iii) "auditor incentives," which encompasses elements such as time pressure and fees paid by the client on auditors' motivations to detect fraud.

Institutional Forces

All the experts emphasized the impact that institutional factors can have on auditors' effectiveness at detecting fraud. For example, one expert stated that standard setters are reluctant to require new procedures that may be effective at detecting fraud because they fear it may elevate auditors' responsibilities for detecting fraud. The underlying concern for this reaction is that if auditors have more requirements for detecting fraud, they may be held to a higher legal liability standard if they fail to detect fraud. Another expert suggested that auditing standards do not require effective procedures for detecting fraud and are therefore not focused on fraud detection. The experts also touched upon the effect of audit structure, with one noting that audit procedures designed to detect fraud are not integrated into the audit methodology.

Knowledge, Training, and Experience

All four experts noted that auditors' lack training in fraud detection methods or fraud investigation techniques. They agreed that auditors are not effectively trained to detect or recognize fraud. One expert noted that fact patterns suggesting that fraud exists (i.e., fraud schemes) are unfamiliar to many auditors because they have not been trained in this area and because fraud is a rare event. Auditors may lack adequate training in fraud detection methods or fraud investigation techniques.

Auditor Incentives

The specific incentive issues mentioned by the fraud experts included conflicts of interest resulting from being paid by the client. One expert observed that this conflict will lead auditors to subtly avoid testing areas where they suspect fraud exists. Similarly, one expert mentioned that auditors become advocates for their clients and therefore lose the ability to objectively evaluate fraudulent accounting methods. Conflicts of interest can be contrasted with other incentive-related concerns noted by our experts. For example, an expert noted that time and fee budgets cause auditors to reduce costs by doing less quality or quantity of audit testing than necessary to detect fraud or to use staff with less expertise than is optimal for detecting fraud. These pressures are also believed to lead auditors to do the minimum to meet the letter of the law as specified in auditing standards as opposed to meeting the spirit of the standard such as searching for evidence to detect fraud.

Chapter 4

4.FRAUD DETECTION

Fraud detection is a set of activities undertaken to prevent money or property from being obtained through false pretenses. Fraud detection is applied to many industries such as banking or insurance. In banking, fraud may include forging checks or using stolen credit cards. Other forms of fraud may involve exaggerating losses or causing an accident with the sole intent for the payout.

With an unlimited and rising number of ways someone can commit fraud, detection can be difficult to accomplish. Activities such as reorganization, downsizing, moving to new information systems or encountering a cybersecurity breach could weaken an organization's ability to detect fraud. This means techniques such as real-time monitoring for frauds is recommended. Organizations should look for fraud in financial transactions, location, devices used, initiated sessions and authentication systems.

Fraud can be committed in a number of different ways and in a number of different settings. For example, fraud can be committed in banking, insurance, government and in healthcare sectors.

One common type of fraud in banking is customer account takeover, where someone illegally gains access to a victim's bank account using bots. Other examples of fraud in banking include the use of malicious applications, the use of false identities, money laundering, credit card fraud and mobile fraud.

Fraud in insurance can include premium diversion fraud, which is the embezzlement of insurance premiums; or frees churning, which is excessive trading by a stockbroker to maximize commissions. Other forms of insurance fraud include asset diversion, workers compensation, car accident, stolen or damaged car, and house fire fraud. The motive behind all insurance fraud is financial profits.

Government fraud is committing fraud against federal agencies such as the departments of Health and Human Services, Transportation, Education, or Energy. Types of government fraud include billing for unnecessary procedures, overcharging for items that cost much less, providing old equipment when billing for new or reporting hours worked for a worker that does not exist.

Healthcare fraud includes drug fraud and medical fraud, as well as encompassing some insurance fraud. Healthcare fraud is committed when someone defrauds an insurer or government health care program.

According to the Cambridge Dictionary, fraud is "the crime of getting money by deceiving people." It's as old as humanity. Ever since people started exchanging goods and services, there has been a risk of one party scamming the other. And there has always been a risk of a third party scamming both the seller and the buyer. With the development and expansion of e-commerce, fraud has taken on new forms and become more powerful than ever. As the scale of e-shopping, online banking, and online insurance increases, fraudsters take full advantage of every weak spot in every system they can find. Quite often, before professionals can patch up a system, sensitive data is stolen and millions of dollars are lost. Fraud has turned into a major issue and an uncontrolled expenditure for e-commerce retailers on a global level.

Preventing, detecting, and eliminating fraud are some of the primary concerns of the e-commerce and banking industries at present. One of the most promising means for achieving them are machine learning development services.

Fraud detection is a process that detects and prevents fraudsters from obtaining money or property through false means. It is a set of activities undertaken to detect and block the attempt of fraudsters from obtaining money or property fraudulently. Fraud detection is prevalent across banking, insurance, medical, government, and public sectors, as well as in law enforcement agencies.

Random Forest:

Random forest is a tree based algorithm which involves building several trees and combining with the output to improve generalization ability of the model. This method of

combining trees is known as an ensemble method. Ensembling is nothing but a combination of weak learners (individual trees) to produce a strong learner. Random Forest can be used to solve regression and classification problems. In regression problems, the dependent variable is continuous. In classification problems, the dependent variable is categorical.

WORKING OF RANDOM FOREST:

Bagging Algorithm is used to create random samples. Data set D1 is given for n rows and m columns and new data set D2 is created for sampling n cases at random with replacement from the original data. From dataset D1,1/3rd of rows are left out and is known as Out of Bag samples. Then, new dataset D2 is trained to this models and Out of Bag samples is used to determine unbiased estimate of the error. Out of m columns, M << m columns are selected at each node in the data set. The M columns are selected at random. Usually, the default choice of M, is m/3 for regression tree and M is sqrt(m) for classification tree. Unlike a tree, no pruning takes place in random forest i.e, each tree is grown fully. In decision trees, pruning is a method to avoid over fitting. Pruning means selecting a sub tree that leads to the lowest test error rate. Cross validation is used to determine the test error rate of a sub tree. Several trees are grown and the final prediction is obtained by averaging or voting.

Algorithm steps for finding the best algorithm

Step 1: Import the dataset

Step 2: Convert the data into data frames format

Step3: Do random oversampling using ROSE package

Step4: Decide the amount of data for training data and testing data

Step5: Give 70% data for training and remaining data for testing.

Step6: Assign train dataset to the models

Step7: Choose the algorithm among 3 different algorithms and create the model

Step8: Make predictions for test dataset for each algorithm

Step9: Calculate accuracy for each algorithm

Step10: Apply confusion matrix for each variable

Step11: Compare the algorithms for all the variables and find out the best algorithm.

4.1 HOW FRAUD DETECTION WORKS:

Fraudulent activities include money laundering, cyberattacks, fraudulent banking claims, forged bank checks, identity theft, and many such illegal practices. As a result, organizations implement modern fraud detection and prevention technologies and risk management strategies to combat growing fraudulent transactions across diverse platforms.

These techniques apply adaptive and predictive analytics (i.e., machine learning) to create a fraud risk score along with real-time monitoring of fraudulent events. This allows continuous monitoring of transactions and crimes in real-time. It also helps decipher new and sophisticated preventive measures via automation.

4.2TYPES OF FRAUD DETECTION:

Fraud detection generally involves data analysis-based techniques. These techniques are broadly categorized as statistical data analysis techniques and artificial intelligence or AI-based techniques. Let’s understand both in detail.

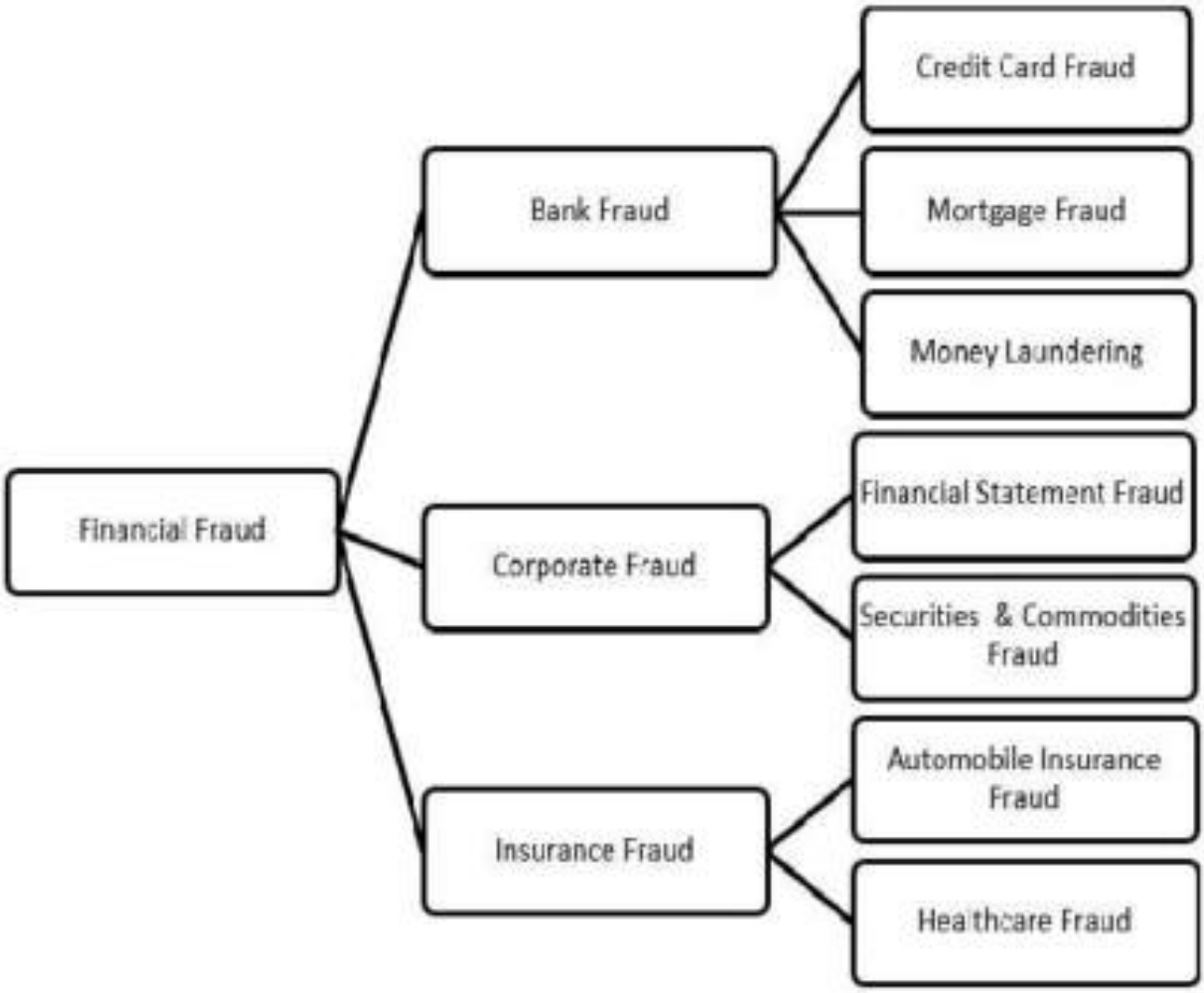


Fig. 1 Types of Common Fraud [8]

Chapter 5
STATISTICAL DATA ANALYSIS TECHNIQUES

Statistical data analysis for fraud detection performs various statistical operations such as fraud data collection, fraud detection, and fraud validation by conducting detailed investigations. These techniques are further subdivided into the following types:

5.1 STATISTICAL PARAMETER CALCULATION:

Statistical parameter calculation refers to the calculation of various statistical parameters such as averages, quantiles, performance metrics, and probability distributions for fraud-related data

collected during the data capturing process.

Anomaly and fraud detection is a multi-billion-dollar industry. According to a **Nilson Report**, the amount of global credit card fraud alone was USD 7.6 billion in 2010. In the UK fraudulent credit card transaction losses were estimated at more than USD 1 billion in 2018. To counter these kinds of financial losses a huge amount of resources are employed to identify frauds and anomalies in every single industry.

In data science, “Outlier”, “Anomaly” and “Fraud” are often synonymously used, but there are subtle differences. An “outliers’ generally refers to a data point that somehow stands out from the rest of the crowd. However, when this outlier is completely unexpected and unexplained, it becomes an anomaly. That is to say, all anomalies are outliers but not necessarily all outliers are anomalies. In this article, however, I am using these terms interchangeably.

There are numerous reasons why understanding and detecting outliers are important. As a data scientist when we make data preparation we take great care in understanding if there is any data point unexplained, which may have entered

erroneously. Sometimes we filter completely legitimate outlier data points and remove them to ensure greater model performance.

There is also a huge industrial application of anomaly detection. Credit card fraud detection is the most cited one but in numerous other cases anomaly detection is an essential part of doing business such as detecting network intrusion, identifying instrument failure, detecting tumor cells etc.

A range of tools and techniques are used to detect outliers and anomalies, from simple statistical techniques to complex machine learning algorithms, depending on the complexity of data and sophistication needed. The purpose of this article is to summarise some simple yet powerful statistical techniques that can be readily used for initial screening of outliers. While complex algorithms can be inevitable to use, sometimes simple techniques are more than enough to serve the purpose.

Z-score

$$Z\text{-score} = \frac{x - mean}{Standard\ Deviation}$$

Z-score is probably the simplest one yet an useful statistical measure for anomaly detection. In a statistical distribution, Z-score tells you how far is a given data point from the rest of the crowd. Technically speaking, Z-score measures how many standard deviations away a given observation is from the mean. A Z-score of 2 means that the data point is 2 standard deviation far from the mean.

To calculate Z-score you only need two parameters: mean and standard deviation, and they are fairly easy to get in any programming language (for example in Python you just apply the function `describe()` to a data frame).

Once these two parameters are obtained, Z-score for any given data point is calculated using the following simple formula:

5.2 REGRESSION ANALYSIS:

Regression analysis allows you to examine the relationship between two or more variables of interest. It also estimates the relationship between independent and dependent variables. This helps understand and identify relationships between several fraud variables, which further helps in predicting future fraudulent activities. These predictions are based on the usage patterns of fraud variables in a potentially fraudulent use case.

5.3PROBABILITY DISTRIBUTION AND MODELS:

In this technique, models and probability distributions of various business fraudulent activities are mapped, either in terms of different parameters or probability distributions.

5.4DATA MATCHING:

Data matching is used to compare two sets of collected data (i.e., fraud data). The process can be carried out either based on algorithms or programmed loops. In addition, data matching is used to remove duplicate records and identify links between two data sets for marketing, security, or other purposes.

Your personal information may be shared internally and with other agencies such as the Cabinet Office, and may be used in data matching exercises, for the prevention and detection of fraud.

How we use your personal information

We are required by law to protect the public funds that we administer. Information provided to us may be shared with other bodies responsible for auditing or administering public funds or where undertaking a public function, in order to prevent and detect fraud.

National Fraud Initiative (NFI)

The Cabinet Office is responsible for carrying out the National Fraud Initiative (NFI) data matching exercise which is intended to detect fraudulent and erroneous payments from the public purse. The work is designed to help reduce the level of many types of fraud that can occur including housing benefit, council tax reduction, payroll, pension, licensing and housing tenancy fraud. Data matching involves comparing computer records held by one body against other computer records held by the same or another body to see how far they match. This is usually personal information. Computerised data matching allows potentially fraudulent claims and payments to be identified. Where a match is found it may indicate that there is an inconsistency which may require further investigation. No assumption can be made as to whether there is fraud, error or other explanation until an investigation is carried out. The Cabinet Office currently requires us to participate in this NFI data matching exercise to assist in the prevention and detection of fraud. We are required to provide particular sets of data to the Cabinet Office for matching for each exercise, and these are set out by the Cabinet Offices. This does not require the consent of the individuals concerned under

the Data Protection Act 1998. Details of this data matching can be found on the Data specification Gov.uk website. Once the Cabinet Office has completed the data matching exercise, the results are shared with us, we then investigate any inconsistencies to determine whether there is or has been fraud taking place. The use of data by the Cabinet Office in a data matching exercise is carried out with statutory authority under its powers in Part 6 of the Local Audit and Accountability Act 2014. It does not require the consent of the individuals concerned under the Data Protection Act 1998. Data matching by the Cabinet Office is subject to a Code of Practice. Further information on the Cabinet Officer's legal powers and the reasons why it matches particular information are provided on the Gov.uk website.

Other fraud prevention and detection work

The primary objective of the council's Counter Fraud strategy is to ensure the prevention of fraud and corrupt acts and to ensure that any instances of these are investigated and dealt with effectively. As part of this strategy we conduct a programme of pro-active counter fraud reviews into transactions and records held across different business areas. They are designed specifically to identify unusual, erroneous or potentially fraudulent transactions. We may also use information including personal information to prevent and detect fraud in any of our systems and may supply information to credit reference agencies or other external bodies for such purposes.

Chapter 6

AI BASED TECHNIQUES

Deploying AI for fraud prevention has helped companies enhance their internal security and streamline business processes. Through improved efficiency, AI has emerged as an essential technology to prevent fraud at financial institutions. AI-based fraud detection techniques include the following methods.

6.1 DATA MINING:

Data mining for fraud detection and prevention classifies, clusters, and segments the data and automatically finds associations and rules in the data that may signify interesting patterns, including those related to fraud.

6.2. NEURAL NETWORKS:

Neural networks under fraud detection perform classification, clustering, generalization, and forecasting of fraud-related data that can be compared against conclusions that are raised in internal audits or formal financial document.

6.3MACHINE LEARNING:

Fraud detection with machine learning becomes possible due to the ability of ML algorithms to learn from historical fraud patterns and recognize them in future transactions. Machine learning either uses supervised or unsupervised learning methods.

In supervised learning, a random subsample of all records is manually classified as either 'fraudulent' or 'non-fraudulent'. In unsupervised learning, on the other hand, methods search for common patterns (i.e., fraudulent) and correlations in the raw data, and predictions are built without additional labeling.

6.4 PATTERN RECOGNITION:

Pattern recognition algorithms detect approximate classes, clusters, or patterns of suspicious behavior, either automatically (unsupervised) or manually (supervised). Other techniques such as link analysis, Bayesian networks, decision theory, and sequence matching are also used for fraud detection purposes.

Pattern recognition is the process of recognizing patterns by using a Machine Learning algorithm. Pattern recognition can be defined as the classification of data based on knowledge already gained or on statistical information extracted from patterns and/or their representation. Pattern recognition is the ability to detect arrangements of characteristics or data that yield information about a given system or data set. Predictive analytics in data science work can make use of pattern recognition algorithms to isolate statistically probable movements of time series data into the future. In a technological context, a pattern might be recurring sequences of data over time that can be used to predict trends, particular configurations of features in images that identify objects, frequent combinations of words and phrases for natural language processing (NLP), or particular clusters of behaviour on a network that could indicate an attack — among almost endless other possibilities. In IT, pattern recognition is a branch of Machine Learning that emphasizes the recognition of data patterns or data regularities in a given scenario. Pattern recognition involves classification and cluster of patterns.

Features of Pattern Recognition:

- Pattern recognition completely rely on data and derives any outcome or model from data itself
- Pattern recognition system should recognise familiar pattern quickly and accurate

- Recognize and classify unfamiliar objects very quickly
- Accurately recognize shapes and objects from different angles
- Identify patterns and objects even when partly hidden
- Recognise patterns quickly with ease, and with automaticity
- Pattern recognition always learn from data

Training and Learning Models in Pattern Recognition

Training and Learning is the building block model of Pattern Recognition. Learning is a phenomena through which a system gets trained and becomes adaptable to give result in an accurate manner. Learning is the most important phase as how well the system performs on the data provided to the system depends on which algorithms used on the data.

The model need to undergo from two phases and dataset is divided into two categories, one which is used in training the model and called as Training set and the other is used in testing the model after training called as Testing set.

Training set

Training set is used to build a model. It consists of the set of images which are used to train the system. Training rules and algorithms used give relevant information on how to associate input data with output decision. The system is trained by applying these algorithms on the dataset, all the relevant information is extracted from the data and results are obtained. Generally, 80-85% of the data of the dataset is taken for training data.

Testing set

Testing data is used to test the system. It is the set of data which is used to verify whether the system is producing the correct output after being trained or not. Generally, 20% of the data of the dataset is used for testing. Testing data is used to measure the accuracy of the system.

Chapter 7

APPLICATIONS OF FRAUD DETECTION

Fraud detection is of paramount importance for banks and other companies that deal with a significant number of financial transactions and are therefore at higher risk of suffering from financial fraud. However, other sectors such as ecommerce companies, credit card companies, electronic payment platforms, and B2C fintech companies also need to employ fraud detection to prevent or limit financial fraud.

The most common applications of fraud detection include account-related fraud and payment and transaction fraud. Account fraud is further divided into new account fraud and account takeover fraud. In new account fraud, new accounts are created by using fake identities. Such frauds can be identified by using the patterns of various devices and session indicators for detecting fake identities.

Account theft frauds occur when a hacker obtains products and services by using another person's existing account. In order to prevent this, session, device, and behavioral biometrics of the user can be computed and scored to verify an account. In addition, analyzing user journeys for behavioral patterns can help detect account takeovers before they cause any financial harm.

Payment fraud is any kind of false or illegal transaction that is carried out by a cybercriminal. The perpetrator cheats the victim of money, personal property, interest, or sensitive information.

This category further includes unauthorized transactions fraud, stolen merchandise fraud, and false requests for refund fraud.

7.1 EXAMPLE:

BANKING AND FINANCIAL SERVICES:

As the digital trend has been gaining traction worldwide, frauds have been increasing with the rising number of online and ATM transactions. The most common types of banking frauds are:

7.2 API FRAUD:

Payment Services Directive 2 (PSD2) mandates certain European financial institutions to open up their services via application programming interfaces (APIs). This creates a new attack surface.

7.3 STOLEN/FAKE CREDIT CARD FRAUD:

A fake card is generated based on a user's card information that scammers manage to gain access to. There are numerous ways that scammers use to do this, with card skimming being the most common. Credit card skimming is a technique where the scammer attaches a small device to the transaction machine that cannot be easily noticed.

7.4 WEBSITE CLONING:

Website cloning is one of the most popular methods among scammers to fleece people of their money. As the name suggests, cybercriminals first create a 'clone' site of the original website. Next, they create a trap intended to get unsuspecting victims to visit the cloned site. This is usually done via links shared through emails, text messages, or social media posts, asking unsuspecting users to click on them.

7.5 ATM FRAUD:

ATM fraud is described as a fraudulent activity where criminals use another person's ATM card to withdraw money instantly from that account. Different types of ATM frauds include card shimming, card skimming, card trapping, and keyboard jamming, to name a few.

17

Chapter 8

MACHINE LEARNING

The term Machine Learning was coined by Arthur Samuel in 1959, an American pioneer in the field of computer gaming and artificial intelligence, and stated that "it gives computers the ability to learn without being explicitly programmed".

And in 1997, Tom Mitchell gave a "well-posed" mathematical and relational definition that "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

8.1CLASSIFICATION OF MACHINE LEARNING:

Machine learning implementations are classified into three major categories, depending on the nature of the learning "signal" or "response" available to a learning system which is as follows:

8.1.1. SUPERVISED LEARNING:

When an algorithm learns from example data and associated target responses that can consist of numeric values or string labels, such as classes or tags, in order to later predict the correct response when posed with new examples comes under the category of Supervised learning. This approach is indeed similar to human learning under the supervision of a teacher. The teacher provides good examples for the student to memorize, and the student then derives general rules from these specific examples.

8.1.2. UNSUPERVISED LEARNING:

Whereas when an algorithm learns from plain examples without any associated response, leaving to the algorithm to determine the data patterns on its own. This type of algorithm tends to restructure the data into something else, such as new features that may represent a class or a new series of un-correlated values. They are quite useful in providing humans with insights into the meaning of data and new useful inputs to supervised machine learning algorithms.

As a kind of learning, it resembles the methods humans use to figure out that certain objects or events are from the same class, such as by observing the degree of similarity between objects. Some recommendation systems that you find on the web in the form of marketing automation are based on this type of learning.

8.1.3.REINFORCEMENT LEARNING:

When you present the algorithm with examples that lack labels, as in unsupervised learning. However, you can accompany an example with positive or negative feedback according to the solution the algorithm proposes comes under the category of Reinforcement learning, which is connected to applications for which the algorithm must make decisions (so the product is prescriptive, not just descriptive, as in unsupervised learning), and the decisions bear consequences. In the human world, it is just like learning by trial and error.

Errors help you learn because they have a penalty added (cost, loss of time, regret, pain, and so on), teaching you that a certain course of action is less likely to succeed than others. An interesting example of reinforcement learning occurs when computers learn to play video games by themselves.

In this case, an application presents the algorithm with examples of specific situations, such as having the gamer stuck in a maze while avoiding an enemy. The application lets the algorithm know the outcome of actions it takes, and learning occurs while trying to avoid what it discovers to be dangerous and to pursue survival. You can have a look at how the company Google DeepMind has created a reinforcement learning program that plays old Atari's video games. When watching the video, notice how the program is initially clumsy and unskilled but steadily improves with training until it becomes a champion.

8.1.4.SEMI-SUPERVISED LEARNING:

Where an incomplete training signal is given: a training set with some (often many) of the target outputs missing. There is a special case of this principle known as Transduction where the entire set of problem instances is known at learning time, except that part of the targets are missing.

8.2MACHINE LEARNING USED IN FRAUD DETECTION

The concept behind using machine learning in fraud detection is that fraudulent transactions have specific features that legitimate transactions do not. Based on this assumption, machine learning algorithms detect patterns in financial operations and decide whether a given transaction is legitimate. Machine learning fraud detection algorithms are way more effective than humans. They can process a raft of information faster than a team of the best analysts ever could. What's more, ML algorithms can spot patterns that seem unrelated or go unnoticed by a human. By exploring and studying tons of cases of fraudulent behavior, ML algorithms determine the most stealthy fraudulent patterns and remember them forever.

8.3ML HANDLES OVERLOAD WELL:

As online fraud becomes omnipresent, it also gets more sophisticated. Both fraudsters and businesses use cutting-edge technologies and race to get a step ahead. In such a stressful environment, companies need to analyze way more information than they can handle to fight fraud.

Even when a business hires a team of top data scientists, they cannot chase fraudulent attempts as fast as they happen. Fraud detection machine learning models come to the rescue, being able to work 24/7 and analyze enormous amounts of data at the snap of a finger.

8.4 HOW DOES MACHINE LEARNING IN FRAUD DETECTION WORK:

To detect fraud, a machine learning model first needs to collect data. The model analyzes all the data gathered, segments, and extracts the required features from it. Next, the machine

To detect fraud, a machine learning model first needs to collect data. The model analyzes all the data gathered, segments, and extracts the required features from it. Next, the machine learning model receives training sets that teach it to predict the probability of fraud. Finally, it creates fraud detection machine learning models.

The first step, data input, differs for ML and humans. Whereas humans struggle to comprehend massive amounts of data, such a task is a piece of cake for ML. The more data an ML model receives, the better it can learn and polish its fraud detection skills.

Feature extraction is the next step. At this point, features describing good customer behavior and fraudulent behavior are added. These features usually include (but are not limited to) the customer's location, identity, orders, network, and chosen payment method. Based on the complexity of the fraud detection system, the list of investigated features can differ.

Next, a training algorithm is launched. In a nutshell, this algorithm is a set of rules that an ML model has to follow when deciding whether an operation is legitimate or fraudulent. The more data a business can provide for a training set, the better the ML model will be.

Finally, when the training is over, the company receives a fraud detection machine learning model suitable for their business. This model can detect fraud in next to no time with high accuracy. To be effective in credit card fraud detection, a machine learning model needs to be constantly improved and updated. Payment fraud detection can be eliminated for a while using ML. But sooner or later, fraudsters will come up with new tricks to game the system unless you keep it updated.

Chapter 9

PYTHON

Python is a popular programming language. It was created by Guido van Rossum, and released in 1991.

It is used for Web development (server-side), Software development, Mathematics, System scripting.

9.1 What Can Python Do:

Python can be used on a server to create web applications.

Python can be used alongside software to create workflows.

Python can connect to database systems. It can also read and modify files.

Python can be used to handle big data and perform complex mathematics.

Python can be used for rapid prototyping, or for production-ready software development.

9.2 Why Python

- Python works on different platforms (Windows, Mac, Linux, Raspberry Pi, etc).

- Python has a simple syntax similar to the English language.

- Python has syntax that allows developers to write programs with fewer lines than some other programming languages.

- Python runs on an interpreter system, meaning that code can be executed as soon as it is written. This means that prototyping can be very quick.

- Python can be treated in a procedural way, an object-oriented way or a functional way.

- The most recent major version of Python is Python 3, which we shall be using in this tutorial. However, Python 2, although not being updated with anything other than security updates, is still quite popular.

- I this tutorial Python will be written in a text editor. It is possible to write Python in an Integrated Development Environment, such as Thonny, Pycharm, Netbeans or Eclipse which are particularly useful when managing larger collections of Python files.

9.3 PYTHON COMPARED TO OTHER PROGRAMMING LANGUAGES:

- Python was designed for readability, and has some similarities to the English language with influence from mathematics.

- Python uses new lines to complete a command, as opposed to other programming languages which often use semicolons or parentheses.

Python relies on indentation, using whitespace, to define scope; such as the scope of loops, functions and classes. Other programming languages often use curly-brackets for this purpose.

9.4 FRAUD DETECTION IN PYTHON:

Frauds are really in many transactions. We can apply machine learning algorithms to lies the past data and predict the possibility of a transaction being a fraud transaction. In our example we will take credit card transactions, analyse the data, create the features and labels and finally apply one of the ML algorithms to judge the nature of transaction as being fraud or not. Then we will find out the accuracy, precision as well as f-score of the model we are chosen.

An important issue confronting retailers and other businesses today is the preponderance of credit card fraud. This issue recently hit home, as my son was a victim a week prior to me writing this.

We can apply machine learning to help detect credit card fraud, but there is a bit of a problem in that the vast majority of transactions are perfectly legitimate, which reduces a typical model's sensitivity to fraud.

Chapter 10
CREDIT CARDFRAUD DETECTION WITH MACHINE

LEARNING IN PYTHON:

Assume that you are employed to help a credit card company to detect potential fraud cases so that the customers are ensured that they won't be charged for the items they did not purchase. You are given a dataset containing the transactions between people, the information that they are fraud or not, and you are asked to differentiate between them. This is the case we are going to deal with. Our ultimate intent is to tackle this situation by building classification models to classify and distinguish fraud transaction.

10.1 CLASSIFICATION

Steps Involved:

- Import he required packages into our python environment.
- Importing the data.
- Processing the data to our needs and Exploratory Data Analysis.
- Feature Selection and Data Split.
- Building six types of classification models.

Evaluating the created classification models using the evaluation metrics. We are using python for this project because it is really effortless to make use of a bunch of methods, has an extensive amount of packages for machine learning, and can be learned easily. In recent days, the job market for python is seamlessly higher than any other programming language and companies like Netflix are using python for data science and many other applications. With that, let's dive into the coding part.

Sample code

1. Importing Necessary Libraries

To start, let's print out the version numbers of all the libraries we will be using in t
libraries correctly and ensures that this tutorial will be reproducible.

```
import sys
import numpy
import pandas
import matplotlib
import seaborn
import scipy

print('Python: {}'.format(sys.version))
print('Numpy: {}'.format(numpy.__version__))
print('Pandas: {}'.format(pandas.__version__))
print('Matplotlib: {}'.format(matplotlib.__version__))
print('Seaborn: {}'.format(seaborn.__version__))
print('Scipy: {}'.format(scipy.__version__))
```

```
Python: 3.6.6 |Anaconda custom (64-bit)| (default, Jun 28 2018, 11:2
Numpy: 1.15.1
Pandas: 0.23.4
Matplotlib: 2.2.3
Seaborn: 0.9.0
Scipy: 1.1.0
```

```
# import the necessary packages
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

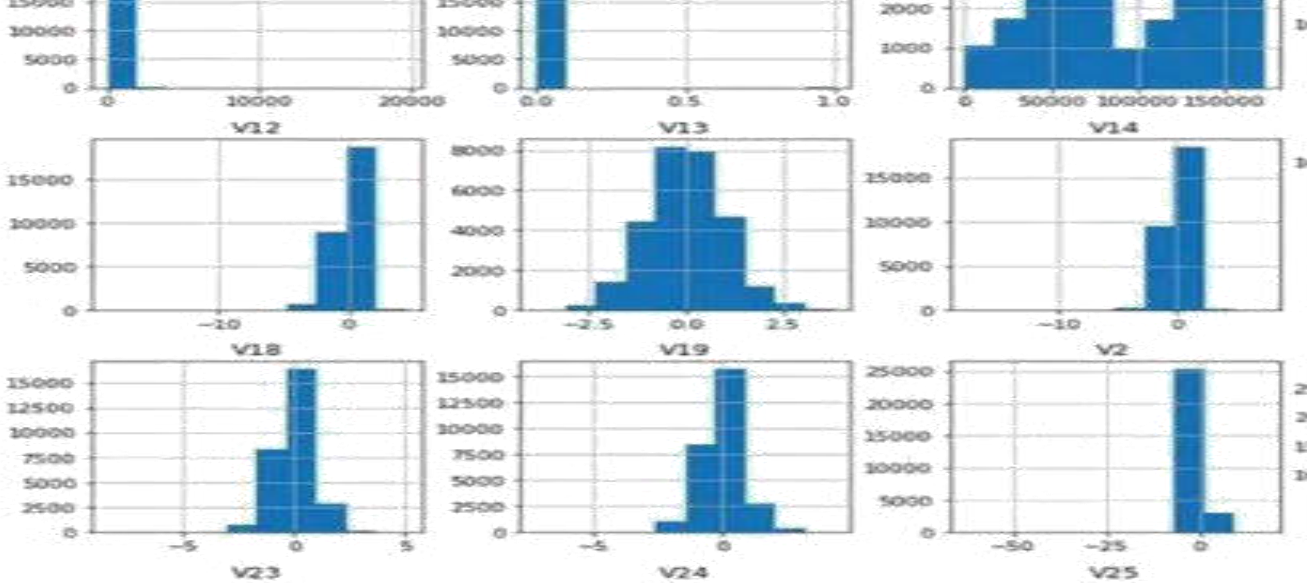
```
# Load the dataset from the csv file using pandas
data = pd.read_csv('creditcard.csv')
```

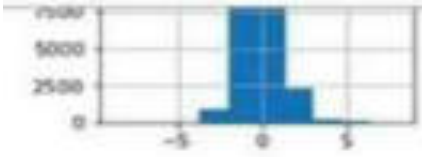
```
# Start exploring the dataset
print(data.columns)
```

```
Index(['Time', 'V1', 'V2', 'V3', 'V4', 'V5', 'V6', 'V7', 'V8', 'V9', 'V10',
       'V11', 'V12', 'V13', 'V14', 'V15', 'V16', 'V17', 'V18', 'V19', 'V20',
       'V21', 'V22', 'V23', 'V24', 'V25', 'V26', 'V27', 'V28', 'Amount',
       'Class'],
      dtype='object')
```

```
# Print the shape of the data
data = data.sample(frac=0.1, random_state = 1)
print(data.shape)
print(data.describe())

# V1 - V28 are the results of a PCA Dimensionality reduction to protect user
```

```
(28481, 31)
                Time            V1            V2            V3            V4
count   28481.000000  28481.000000  28481.000000  28481.000000  28481.000000
mean    94705.035216     -0.001143     -0.018290      0.000795      0.000350
std     47584.727034      1.994661      1.709050      1.522313      1.420003
min         0.000000    -40.470142    -63.344698    -31.813586     -5.266509
25%     53924.000000     -0.908809     -0.610322     -0.892884     -0.847376
50%     84551.000000      0.031139      0.051775      0.178943     -0.017692
75%    139392.000000      1.320048      0.792685      1.035197      0.737312
max    172784.000000      2.411499     17.418649      4.069865     16.715537
```


V12
V13
V14
V18
V19
V2
V23
V24
V25

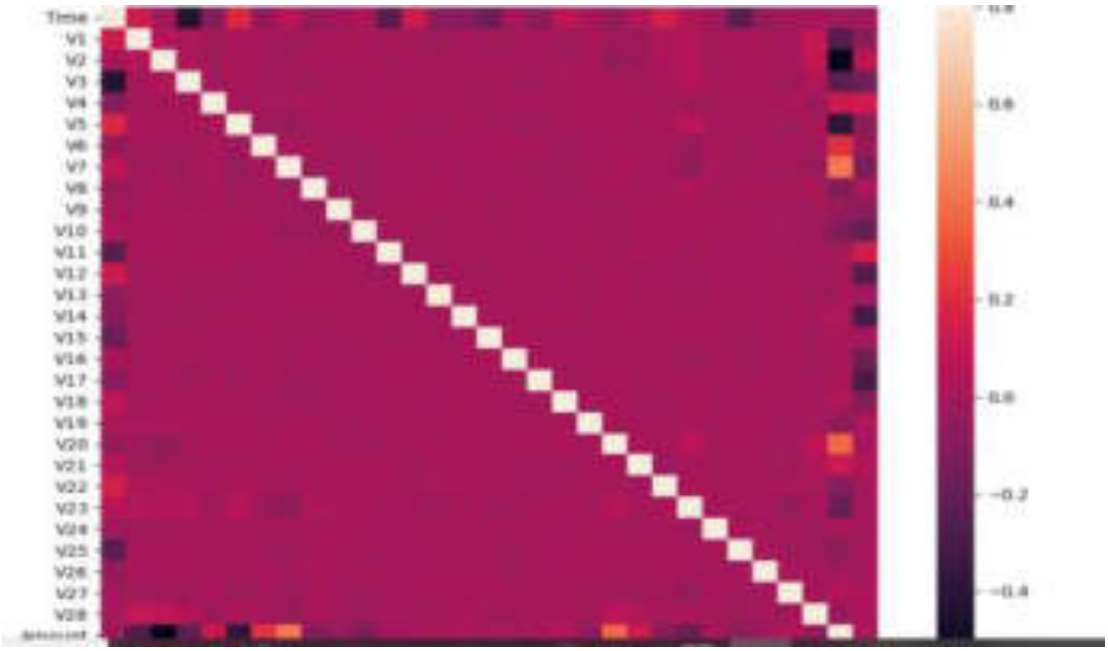



```
# Determine number of fraud cases in dataset

Fraud = data[data['Class'] == 1]
Valid = data[data['Class'] == 0]

outlier_fraction = len(Fraud)/float(len(Valid))
print(outlier_fraction)

print('Fraud Cases: {}'.format(len(data[data['Class'] == 1])))
print('Valid Transactions: {}'.format(len(data[data['Class'] == 0])))
```

```
0.0017234102419808666
Fraud Cases: 49
Valid Transactions: 28432
```

```
# Correlation matrix
corrmat = data.corr()
fig = plt.figure(figsize = (12, 9))

sns.heatmap(corrmat, vmax = .8, square = True)
plt.show()
```

```
C:\fd>python completeprogram.py
0.0017234102419808666
Fraud Cases: 49
Valid Transactions: 28432

C:\fd>
```



REFERENCES

Credit Card Fraud Detection Based on Transaction Behaviour -by John Richard D. Kho, Larry A. Vea published by Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017

CLIFTON PHUA1, VINCENT LEE1, KATE SMITH1 & ROSS GAYLER2 A Comprehensive Survey of Data Mining-based Fraud Detection Research published by School of Business Systems, Faculty of Information Technology, Monash University, Wellington Road,

Clayton, Victoria 3800, Australia

Survey Paper on Credit Card Fraud Detection by Suman , Research Scholar, GJUS&T Hisar HCE, Sonepat published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 3, March 2014

Research on Credit Card Fraud Detection Model Based on Distance Sum by Wen-Fang YU and Na Wang published by 2009 International Joint Conference on Artificial Intelligence

Credit Card Fraud Detection through Parenclitic Network Analysis- By Massimiliano Zanin, Miguel Romance, Regino Criado, and SantiagoMoral published by Hindawi Complexity Volume 2018,
Article ID 5764370, 9 pages

Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy published by IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 29, NO. 8, AUGUST 2018

Credit Card Fraud Detection-by Ishu Trivedi, Monika, Mrigya, Mridushi published by International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 1, January 2016

David J.Wetson,David J.Hand,M Adams,Whitrow and Piotr Jusczak Plastic Card Fraud Detection using Peer Group Analysis Springer, Issue 2008.

A. C. Bahnsen, A. Stojanovic, D. Aouada, B. Ottersten, "Cost sensitive credit card fraud detection using Bayes minimum risk", Machine Learning and Applications (ICMLA). 2013 12th International Conference, vol. 1, pp. 333-338, 2013.

B.Meena, I.S.L.Sarwani, S.V.S.S.Lakshmi," Web Service mining and its techniques in Web Mining" IJAEGT,Volume 2,Issue 1 , Page No.385-389.

F. N. Ogwueleka, "Data Mining Application in Credit Card Fraud Detection System", Journal of Engineering Science and Technology, vol. 6, no. 3, pp. 311-322, 2011.

G. Singh, R. Gupta, A. Rastogi, M. D. S. Chandel, A. Riyaz, "A Machine Learning Approach for Detection of Fraud based on SVM", International Journal of Scientific Engineering and Technology, vol. 1, no. 3, pp. 194-198, 2012, ISSN ISSN: 2277-1581.

Conclusion

Here I conclude that Our project main motive is to provide security for the information by detecting the frauds.

Credit card fraud is without a doubt an act of criminal dishonesty. This article has listed out the most common methods of fraud along with their detection methods and reviewed recent findings in this field. This paper has also explained in detail, how machine learning can be applied to get better results in fraud detection along with the algorithm, pseudocode, explanation its implementation and experimentation results.

While the algorithm does reach over 99.6% accuracy, its precision remains only at 28% when a tenth of the data set is taken into consideration. However, when the entire dataset is fed into the algorithm, the precision rises to 33%. This high percentage of accuracy is to be expected due to the huge imbalance between the number of valid and number of genuine transactions.

Since the entire dataset consists of only two days transaction records, its only a fraction of data that can be made available if this project were to be used on a commerial scale. Being based on machine learning algorithms, the program will only increase its efficiency over time as more data is put into it.

In this paper, Machine learning technique like Logistic regression, Decision Tree and Random forest were used to detect the fraud in credit card system. Sensitivity, Specificity, accuracy and error rate are used to evaluate the performance for the proposed system. The accuracy for logistic

regression, Decision tree and random forest classifier are 90.0, 94.3, and 95.5 respectively. By comparing all the three method, found that random forest classifier is better than the logistic regression and decision tree.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

Kakinada

"WEB AUTOMATION USING SELENIUM"

A Project Report submitted to

Jawaharlal Nehru Technological University Kakinada

in partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY
IN
COMPUTER SCIENCE AND ENGINEERING

By

Sai Manideep P (17H71A0566) **Sai Sandeep K (17H71A05A4)**
Guna Mallika T (17H71A0572) **Vinay Babu D (17H71A05B7)**

Under the esteemed guidance of
Mr. Mohammed Eliaz
Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Devineni Venkata Ramana & Dr. Hima Sekhar
MIC College of Technology
Kanchikacherla - 521 180. Phone : 08678-273535
(Approved by AICTE & Affiliated to JNTU-Kakinada)

CERTIFICATE

This is to certify that the project entitled "WEB AUTOMATION USING SELENIUM" which is a bonafide work carried out by Sai Manideep P(17H71A0566), Guna Mallika T (17H771A0572), Sai Sandeep K(17H71A05A4), Vinay Babu D(17H71A05B7) in partial fulfillment for the award of degree of BACHELOR OF TECHNOLOGY in COMPUTER SCIENCE AND ENGINEERING by JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA during the academic year 2017-2021.It is certified that all the corrections and suggestions indicated for internal assessment have been incorporated in the report. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the above degree.

Project Guide
Mr. MOHAMMED ELIAZ

Head of the Department
Mr. D. PRASAD

Examiner 1

Examiner 2

ACKNOWLEDGEMENT

It gives us immense pleasure to thank our guide **Mr. MOHAMMED ELIAZ, Assistant Professor in the Department of Computer Science and Engineering** for being his source of constant inspiration, precious guidance. Without his close monitoring and valuable suggestions this work wouldn't have taken this shape for providing the background knowledge to deal with the problem.

We have immense pleasure in expressing our thanks and deep sense of gratitude to **Mr. D. PRASAD, Professor & Head of the Department of Computer Science and Engineering** for extending necessary facilities for the completion of Project.

With immense respect, we express our sincere gratitude to **Mr. K. Srinivasa Rao, Principal** for permitting us to take up our project and to complete the project successfully.

We would like to express our sincere thanks to all faculty members of the department for their cooperation which has given us the cogency to build-up adamant aspiration over the completion of our project to complete the project successfully in time.

Finally, we thank one and all who directly and indirectly helped us to complete the project successfully.

Project Associates
Sai Manideep P (17H71A0566)
Guna Mallika T (17H71A0572)
Sai Sandeep K (17H71A05A4)
Vinay Babu D (17H71A05B7)

DECLARATION

We, **SAI MANIDEEP P, GUNA MALLIKA T, SAI SANDEEP K, VINAY BABU D** bearing register numbers 17H71A0566,17H771A0572,17H71A05A4,17H71A05B7 do hereby declare that this project entitled "**WEB AUTOMATION USING SELENIUM**" was carried out by us under the esteemed guidance of Mr. MOHAMMED ELIAZ. This project work is submitted to **JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA**, in fulfilment of requirement for the award of **Bachelor of Technology in Computer Science and Engineering** during the academic year **20202021.**

Project Associates
Sai Manideep P (17H71A0566)
Guna Mallika T (17H71A0572)
Sai Sandeep K (17H71A05A4)
Vinay Babu D (17H71A05B7)

TABLE OF CONTENTS

ABSTRACT

Think of all the clicking and typing you do at your desk everyday involving web sites and web-based applications getting automated. This concept of making everything automated is known as **WEB AUTOMATION.**

Software companies are more committed to produce more quality software systems with lower costs. Software tested by a separate testing team, more quality is inbuilt. Earlier, Manual testing was done and now a days due to more complexity in the software to be developed, there is a need for automation of testing. Automated software testing can increase the depth and scope of tests to help improve software quality. Automated software testing can reduce the time to run repetitive tests from days to hours. A time savings that translates directly into cost savings.

Using automated testing tools like Selenium, GUI testing and Cross Browser testing and Web Automation is done more effectively.

In this project we perform Web Automation using Selenium Web Driver Python. The webbased application is tested with selenium web driver with Python code. Cross Browser testing is done with the various leading browsers to check the performance of applications as expected.

Selenium:

Selenium is a powerful tool for controlling web browser through program. It is functional for all browsers, works on all major OS and its scripts are written in various languages i,e Python, Java, C#, Ruby etc. We will be working with python.

CHAPTER – 1

TESTING

1.1 INTROUCTION

Think of all the clicking and typing you do at your desk everyday involving web sites and web-based applications getting automated. This concept of making everything automated is known as **WEB AUTOMATION.**

Software tests have to be repeated often during development cycles to ensure quality. Every time source code is modified software tests should be repeated. Manually repeating these tests is costly and time consuming.

Automated software testing can increase the depth and scope of tests to help improve software quality. Lengthy tests that are often avoided during manual testing can be run unattended. They can even be run on multiple computers with different configurations.
They are much faster than manual tests. Automated software testing can reduce the time to run repetitive tests from days to hours. A time savings that translates directly into cost savings.

Using automated testing tools like Selenium, GUI testing and Cross Browser testing and Web Automation is done more effectively.

1.2 TESTING

Testing is one of the main phase in the Software Development Life Cycle.
Software Testing is evaluation of the software against requirements gathered from users and system specifications. It is a method to check whether the actual software product matches expected requirements and to ensure that software product is defect free. Testing is conducted at the phase level in software development life cycle or at module level in program code. It involves execution of software system components using manual or automated tools to evaluate one or more properties of interest. Software testing comprises Validation and Verification.

Software Validation

Validation is the process of examining whether or not the software satisfies the user requirements. It is carried out at the end of the SDLC. If the software matches requirements for which it was made, it is validated.

- Validation ensures the product under development is as per the user requirements.
- Validation answers the question – "Are we developing the product which attempts all that the user needs from this software?".
- Validation emphasizes on user requirements.

Software Verification

Verification is the process of confirming if the software is meeting the business requirements, and is developed adhering to the proper specifications and methodologies.

- Verification ensures the product being developed is according to design specifications.
- Verification answers the question– "Are we developing this product by firmly following all design specifications?"
- Verifications concentrate on the design and system specifications. Target of the test are -
- **Errors** - These are actual coding mistakes made by developers. In addition, there is a difference in output of software and desired output, is considered as an error.
- **Fault** - When error exists fault occurs. A fault, also known as a bug, is a result of an error which can cause a system to fail.
- **Failure** - failure is said to be the inability of the system to perform the desired task. Failure occurs when fault exists in the system.

1.2.1 IMPORTANCE OF TESTING

Software Testing is important because if there are any bugs or errors it can be identified early and solved before delivery of the software product. Properly tested software product ensures reliability, security and high performance which furthers results in time saving, cost effectiveness and customer satisfaction.

The few main reasons why the testing is more important:

- Helps in saving money
- Security
- Satisfaction of customer
- Quality of the product
- Enhancing the developing process
- Easy while adding the new features
- Determining the performance of software

1.2.2 NEED OF TESTING

Testing is important because software bugs could be expensive or even dangerous. Software bugs can potentially cause monetary and human loss, and history is full of such examples.

- In April 2015, the Bloomberg terminal in London crashed due to a software glitch affecting more than 300,000 traders on financial markets. It forced the government to postpone a 3bn pound debt sale.
- Starbucks was forced to close about 60 percent of stores in the U.S and Canada due to software failure in its POS system. At one point, the store served coffee for free as they were unable to process the transaction.

- Vulnerability in Windows 10. This bug enables users to escape from security sandboxes through a flaw in the win32k system.
- In 2015 fighter plane F-35 fell victim to a software bug, making it unable to detect targets correctly.
- In April of 1999, a software bug caused the failure of a $1.2 billion military satellite launch, the costliest accident in history.
- In May of 1996, a software bug caused the bank accounts of 823 customers of a major U.S. bank to be credited with 920 million US dollars.

1.2.3 ADVANTAGES OF TESTING

- **Cost-Effective:** It is one of the important advantages of software testing. Testing any IT project on time helps you to save your money for the long term. In case if the bugs caught in the earlier stage of software testing, it costs less to fix.
- **Security:** It is the most vulnerable and sensitive benefit of software testing. People are looking for trusted products. It helps in removing risks and problems earlier.
- **Product quality:** It is an essential requirement of any software product. Testing ensures a quality product is delivered to customers.
- **Customer Satisfaction:** The main aim of any product is to give satisfaction to their customers. UI/UX Testing ensures the best user experience.

1.2.4 TESTING-ITs TYPES

Typically testing is divided into two categories:

- Functional Testing
- Non-Functional Testing or Performance Testing

Functional Testing types include:

- **Unit Testing:** It is a type of software testing where individual units or components of a software are tested.
- **Integration Testing:** It is defined as a type of testing where software modules are integrated logically and tested as a group.
- **System Testing**: It is a type of software testing that is performed on a complete integrated system to evaluate the compliance of the system with the corresponding requirements.

- **Sanity Testing:** It is performed to ensure that the code changes that are made are working as properly. Sanity testing is a stoppage to check whether **testing** for the build can proceed or not
- **Smoke Testing:** It is a software testing process that determines whether the deployed software build is stable or not.
- **Interface Testing:** It is defined as a software testing type which verifies whether the communication between two different software systems is done correctly. A connection that integrates two components is called interface. Testing of these connecting services or interfaces is referred to as Interface Testing.
- **Regression Testing:** It is defined as a type of software testing to confirm that a recent program or code change has not adversely affected existing features.
- **Beta/Acceptance Testing:** It is an opportunity for real users to use a product in a production environment to uncover any bugs or issues before a general release. Beta testing is the final round of testing before releasing a product to a wide audience.

Non-functional Testing types include:

- **Performance Testing:** Performance testing is a testing measure that evaluates the speed, responsiveness and stability of a computer, network, software program or device under a workload.
- **Load Testing:** Load testing generally refers to the practice of modeling the expected usage of a software program by simulating multiple users accessing the program concurrently.
- **Stress Testing:** Stress Testing is a type of software testing that verifies stability & reliability of software application.
- **Volume Testing:** is a type of Software Testing, where the software is subjected to a huge volume of data. It is also referred to as flood testing.
- **Security Testing:** Security testing is a process intended to reveal flaws in the security mechanisms of an information system that protect data and maintain functionality as intended.
- **Compatibility Testing:** it is performed on an application to check its compatibility

- **Install Testing:** Most software systems have installation procedures that are needed before they can be used for their main purpose. Testing these procedures to achieve an installed software system.
- **Recovery Testing:** Recovery testing is the activity of testing how well an application is able to recover from crashes, hardware failures and other similar problems.
- **Reliability Testing:** It is a testing technique that relates to test the ability of a software to function and given environmental conditions that helps in uncovering issues in the software design and functionality.
- **Usability Testing:** Usability testing refers to evaluating a product or service by testing it with representative users.
- **Compliance Testing:** It is also known as Conformance testing is a nonfunctional testing technique which is done to validate whether the system developed meets the organization's prescribed standards or not.
- **Localization Testing:** It is a type of software testing that is performed to verify the quality of a product for a specific culture or locale.

1.2.5 AUTOMATE TESTING vs MANUAL TESTING

AUTOMATED TESTING	MANUAL TESTING
● It is more reliable as it is performed by human error, so it is not reliable	● It is not accurate at all times due to tools
● It is faster than human approach	● It is time consuming
● Investment required for testing tools	● Investment required for human resources
● Coding knowledge is required	● No Coding knowledge required
● Difficult to introduce value-added	●Supports value-added services

Differences between Automated Testing and Manual Testing

1.2.6 BLACK BOX TESTING

Black Box Testing is a software testing method in which the functionalities of software applications are tested without having knowledge of internal code structure, implementation details and internal paths. Black Box Testing mainly focuses on input and output of software applications and it is entirely based on software requirements and specifications. It is also known as Behavioral Testing.

fig 1.1 Black Box Testing

The above Black-Box can be any software system you want to test. For Example, an operating system like Windows, a website like Google, a database like Oracle or even your own custom application. Under Black Box Testing, you can test these applications by just focusing on the inputs and outputs without knowing their internal code implementation.

WHITE BOX TESTING

White Box Testing is software testing technique in which internal structure, design and coding of software are tested to verify flow of input-output and to improve design, usability and security.

In white box testing, code is visible to testers so it is also called Clear box testing, Open box testing, Transparent box testing, Code-based testing and Glass box testing. It is one of two parts of the Box Testing approach to software testing. Its counterpart, Blackbox testing, involves testing from an external or end-user type perspective.

On the other hand, White box testing in software engineering is based on the inner workings of an application and revolves around internal testing. The term "WhiteBox" was used because of the see-through box concept. The clear box or White Box name symbolizes the ability to see through the software's outer shell (or "box") into its inner workings. Likewise, the "black box" in "Black Box Testing" symbolizes not being able to see the inner workings of the software so that only the end-user experience can be tested.

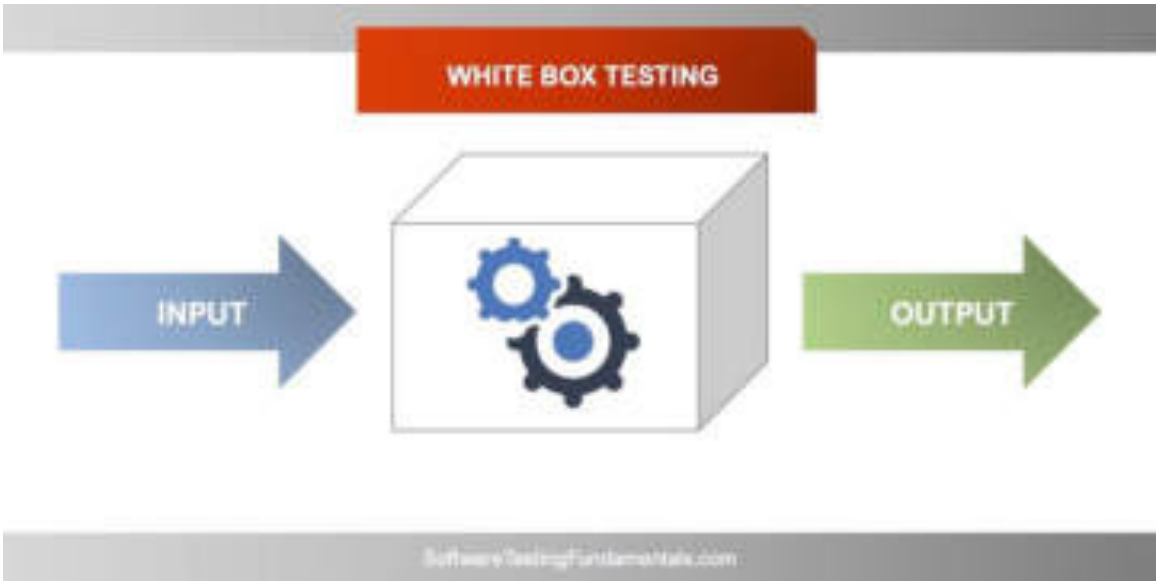


Fig 1.2 White Box Testing

CHAPTER 2

SOFTWARE REQUIREMENT SPECIFICATION

2.1 EXISTING SYSTEM

- There are many existing solutions for the Web Automation. To perform web Automation there are many tools which are already in use.
- Some of the tools which are being used are

Commercial Testing Tools

- HP UTF(QTP)
- HP Mercury WinRunner
- IBM Rational Functional Tester

Open Source Testing Tools

- Selenium
- Watir
- Sikuli

Fig 1.3 Tools for Automation

2.2 PROPOSED SYSTEM

Selenium is a powerful tool for controlling web browser through program. It is functional for all browsers works on all major operating systems and are also written in various languages. i.e. python, java, C#, Ruby etc.

Features of Selenium:

- Free and Open source
- It has a large user base & helping community
- Cross-browsing capability
- Platform compatibility
- Support for different programming languages.

CHAPTER – 3

SELENIUM

3.1INTRODUCTION

Selenium is an Open Source tool is used for automating the test carried out on web browsers. (web applications are tested using any web browser).

Fig 1.4 Selenium

Why Selenium for Automation Testing?

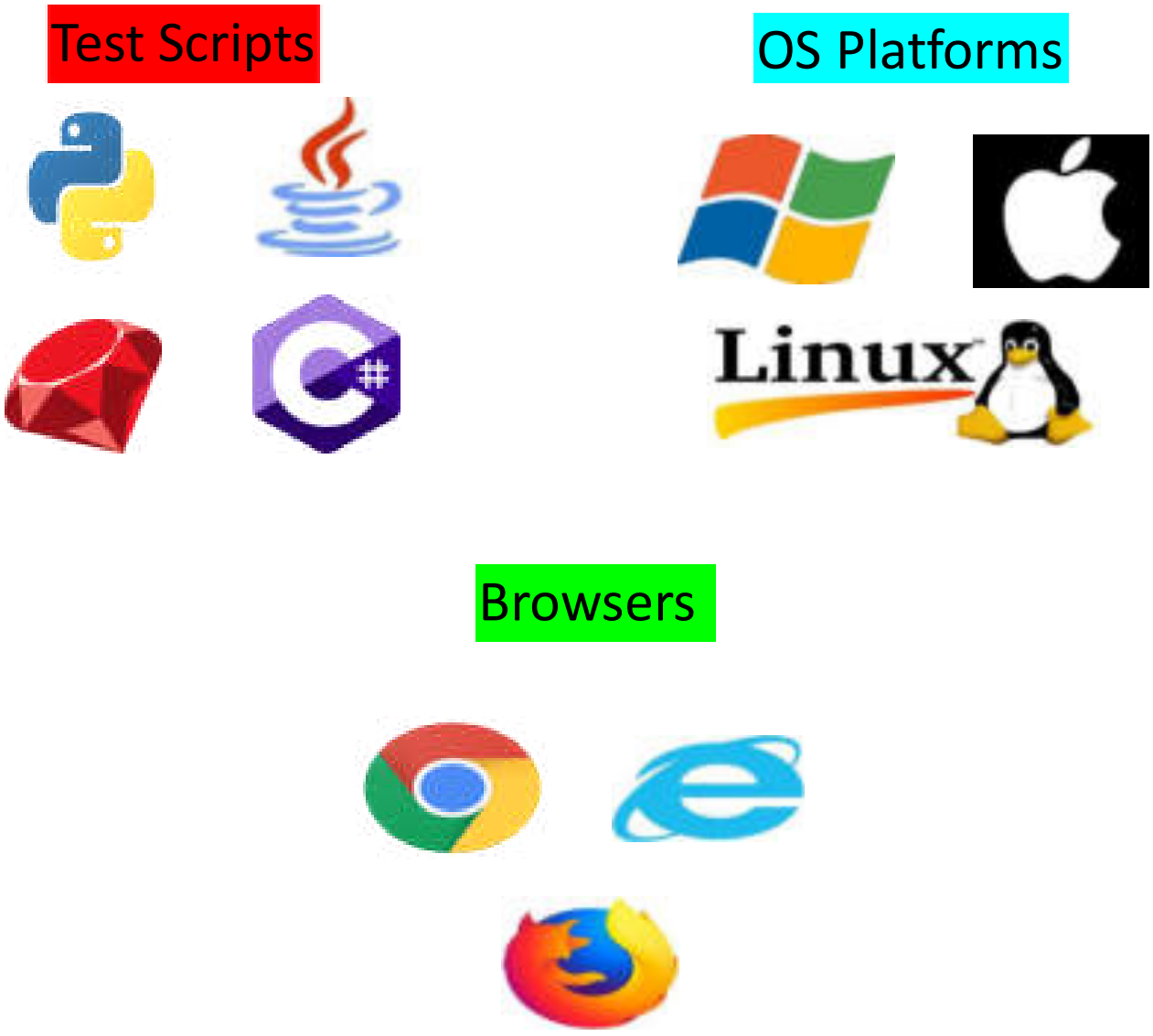


Fig 1.5 Selenium Supports

It is also integrated with tools like testNG & Junit for managing the test cases and generating the reports and it can also be integrated with maven Jenkins and docker to achieve continous testing.

These are some of reasons that keep selenium at the top when compared to other Automation tools.

Selenium Python bindings provides a simple API to write functional/acceptance tests using Selenium WebDriver. Through Selenium Python API you can access all functionalities of Selenium WebDriver in an intuitive way.

Selenium Python bindings provide a convenient API to access Selenium WebDrivers like Firefox, Chrome, Remote etc. The current supported Python versions are 3.5 and above.

The four components of Selenium are:

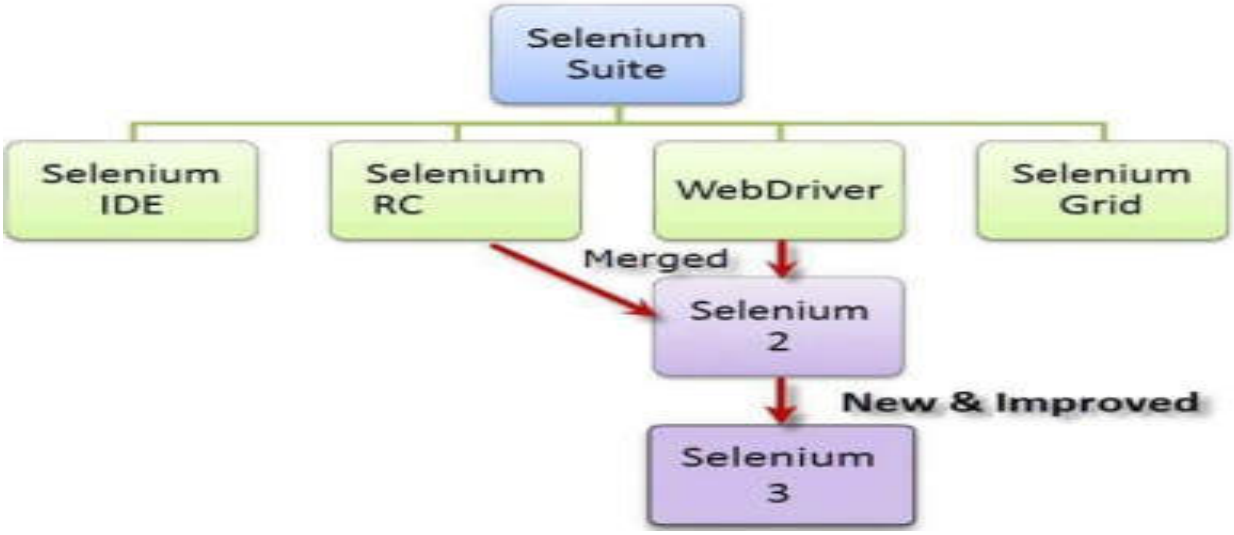


Fig 1.6 Selenium and its components

Selenium Components

- Selenium 1.x (ThoughWorks Inc.)

  Selenium IDE

  Selenium RC

- Selenium 2.x (Google Inc.)

  Selenium WebDriver

  Selenium Grid

SELENIUM IDE:

- Selenium IDE (Integrated Development Environment) is the major tool in the Selenium Suite. It is a complete integrated development environment (IDE) for Selenium tests. It is implemented as a Firefox Add-On and as a Chrome Extension. It allows for recording, editing and debugging of functional tests. It was previously known as Selenium Recorder. Selenium-IDE was originally created by Shinya Kasarani and donated to the Selenium project in 2006. Selenium IDE was previously little-maintained. Selenium IDE began being actively maintained in 2018.

- Scripts may be automatically recorded and edited manually providing autocompletion support and the ability to move commands around quickly. Scripts are recorded in Selenese, a special test scripting language for Selenium. Selenese provides commands for performing actions in a browser (click a link, select option) and for retrieving data from the resulting pages

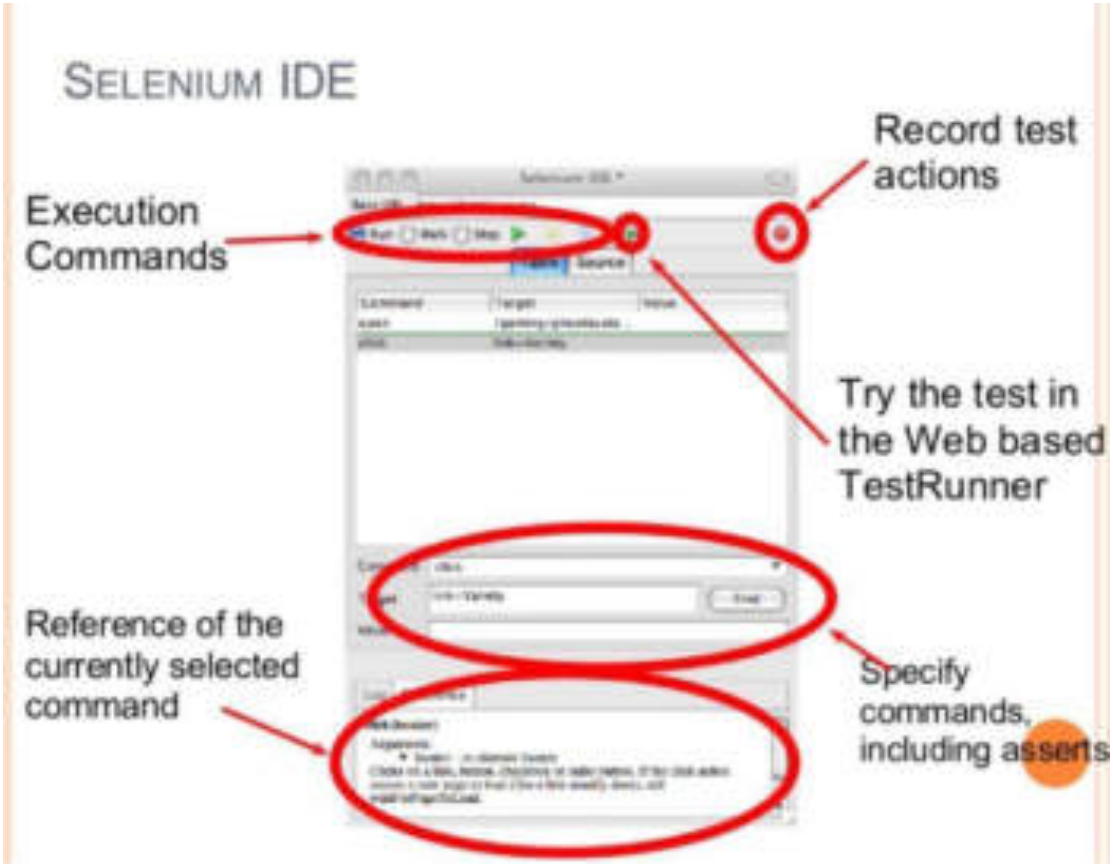


Fig 1.7 Selenium IDE

SELENESE:

Selenese is the set of commands which are used in Selenium IDE. These set of commands are used to test our web applications.

General Selenese Commands:

- Clicking a link – click or clickAndWait commands
- Entering values – type command
- Selecting options from a drop-down list box – select command
- Clicking check boxes or radio buttons – click command

Types:

- Actions - Used for performing operations
- Assertions - Used for checkpoints
- Accessors - Used for storing a value in a part variable

SELENIUM RC:

Selenium Remote Control (RC) is a server, written in Java, that accepts commands for the browser via HTTP. RC makes it possible to write automated tests for a web application in any programming language, which allows for better integration of Selenium in existing unit test frameworks. To make writing tests easier, the Selenium project currently provides client drivers for PHP, Python, Ruby, .NET, Perl and Java. The Java driver can also be used with JavaScript (via the Rhino engine). An instance of selenium RC server is needed to launch the html test case – which means that the port should be different for each parallel run. However, for Java/PHP test cases only one Selenium RC instance needs to be running continuously.

Fig 1.8 Selenium RC Architecture

SELENIUM WEB DRIVER:

Selenium WebDriver is the successor to Selenium RC. Selenium WebDriver accepts commands (sent in Selenese, or via a Client API) and sends them to a browser. This is implemented through a browser-specific browser driver, which sends commands to a browser and retrieves results. Most browser drivers actually launch and access a browser application (such as Firefox, Google Chrome, Internet Explorer, Safari, or Microsoft Edge); there is also an HtmlUnit browser driver, which simulates a browser using the headless browser HtmlUnit.

Selenium WebDriver does not need a special server to execute tests. Instead, the WebDriver directly starts a browser instance and controls it. However, Selenium Grid can be used with WebDriver to execute tests on remote systems (see below). Where possible, WebDriver uses native operating system level functionality rather than browser-based JavaScript commands to drive the browser. This bypasses problems with subtle differences between native and JavaScript commands, including security restrictions.

Fig 1.9 Selenium Web Driver

SELENIUM GRID:

Selenium Grid is a server that allows tests to use web browser instances running on remote machines. With Selenium Grid, one server acts as the hub. Tests contact the hub to obtain access to browser instances. The hub has a list of servers that provide access to browser instances (WebDriver nodes), and lets tests use these instances. Selenium Grid allows running tests in parallel on multiple machines and to manage different browser versions and browser configurations centrally (instead of in each individual test).

The ability to run tests on remote browser instances is useful to spread the load of testing across several machines and to run tests in browsers running on different platforms or operating systems. The latter is particularly useful in cases where not all browsers to be used for testing can run on the same platform.

Advantages of Selenium

QTP and Selenium are the most used tools in the market for software automation testing. Hence it makes sense to compare the pros of Selenium over QTP.

Selenium	QTP
Selenium is an open-source tool.	QTP is a commercial tool and there is a cost involved in each one of the licenses.
Can be extended for various technologies that expose DOM.	Limited add-ons and needs add-ons for each one of the technologies.
Has capabilities to execute scripts across different browsers.	Can run tests in specific versions of Firefox , IE, and Chrome.
Can execute scripts on various operating systems.	Works only with Windows.

Supports mobile devices.	Supports mobile devices with the help of third-party tools.
Executes tests within the browser, so focus is NOT required while script execution is in progress.	Needs Focus during script execution, as the tool acts on the browser (mimics user actions).
Can execute tests in parallel with the use of Selenium Grids.	QTP cannot execute tests in parallel, however integrating QTP with QC allows testers to execute in parallel. QC is also a commercial tool.

Disadvantages of Selenium

Let us now discuss the pitfalls of Selenium over QTP.

Selenium	QTP
Supports only web based applications.	Can test both web and desktop applications.
No feature such as Object Repository/Recovery Scenario	QTP has built-in object repositories and recovery scenarios.
No IDE, so the script development won't be as fast as QTP.	More intuitive IDE; automation can be achieved faster.
Cannot access controls within the browser.	Can access controls within the browser such as favorites bar, backward, and forward buttons.

No default test report generation.	Default test result generation within the tool.
For parameterization, users has to rely on the programming language.	Parameterization is built-in and easy to implement.

3.2 INSTALLING PYTHON BINDINGS FOR SELENIUM

Use pip to install the selenium package. Python 3 has pip available in the standard library. Using pip, you can install selenium like this:

```
pip install selenium
```

You may consider using virtualenv to create isolated Python environments. Python 3 has venv which is almost the same as virtualenv.

You can also download Python bindings for Selenium from the PyPI page for the selenium package. and install manually.

3.3 INSTRUCTIONS FOR WINDOWS USERS

1.Install Python 3 using the MSI available in python.org download page.

2.Start a command prompt using the cmd.exe program and run the pip command as given below to install selenium

```
C:\Python39\Scripts\pip.exe install selenium
```

Now you can run your test scripts using Python. For example, if you have created a Selenium based script and saved it inside C:\my_selenium_script.py, you can run it like this:

```
C:\Python39\python.exe C:\my_selenium_script.py
```

3.4 DRIVERS

Selenium requires a driver to interface with the chosen browser. Firefox, for example, requires a geckodriver, which needs to be installed before the below examples can be run. Make sure it's in your PATH, e. g., place it in

/usr/bin or /usr/local/bin.

Failure to observe this step will give you an error selenium.common.exceptions.

WebDriverException:

Message: 'geckodriver' executable needs to be in PATH.

Other supported browsers will have their own drivers available. Links to some of the more popular browser drivers follow.

Chrome:	https://sites.google.com/a/chromium.org/chromedriver/downloads
Edge:	https://developer.microsoft.com/en-us/microsoft-edge/tools/webdriver/
Firefox:	https://github.com/mozilla/geckodriver/releases
Safari:	https://webkit.org/blog/6900/webdriver-support-in-safari-10/

3.5 DOWNLOADING SELENIUM SERVER

Selenium server is a Java program. Java Runtime Environment (JRE) 1.6 or newer version is recommended to run Selenium server.

You can download Selenium server 2.x from the download page of selenium website. The file name should be like this:

selenium-server-standalone-2.x.x.jar

You can always download the latest 2.x version of Selenium server.

If Java Runtime Environment (JRE) is not installed in your system, you can download the JRE from the Oracle website. If you are using a GNU/Linux system and have root access in your system, you can also use your operating system instructions to install JRE.

If java command is available in the PATH (environment variable), you can start the Selenium server using this command:

java -jar selenium-server-standalone-2.x.x.jar

Replace 2.x.x with the actual version of Selenium server you downloaded from the site.

If JRE is installed as a non-root user and/or if it is not available in the PATH (environment variable), you can type the relative or absolute path to the java command. Similarly, you can provide a relative or absolute path to the Selenium server jar file. Then, the command will look something like this:

/path/to/java -jar /path/to/selenium-server-standalone-2.x.x.jar

CHAPTER-4

SYSTEM DESIGN

4.1 Selenium Architecture:

The following figure gives the architecture followed by the Selenium. It uses any type of languages for writing Scripts and converts through JSON which is wire protocol over HTTP. Then by using Browser drivers the scripts convert to Browsers and displays desired output by using HTTP over HTTP Server.

Fig 2.0 Selenium Architecture

4.2 UML Diagrams:

Use case Diagram:

The purpose of use case diagram is to capture the dynamic aspect of a system. However, this definition is too generic to describe the purpose, as other four diagrams (activity, sequence, collaboration, and Statechart) also have the same purpose. We will look into some specific purpose, which will distinguish it from other four diagrams.

Use case diagrams are used to gather the requirements of a system including internal and external influences. These requirements are mostly design requirements. Hence, when a system is analyzed to gather its functionalities, use cases are prepared and actors are identified.

In brief, the purposes of use case diagrams can be said to be as follows −

- Used to gather the requirements of a system.
- Used to get an outside view of a system.
- Identify the external and internal factors influencing the system.
- Show the interaction among the requirements are actors.

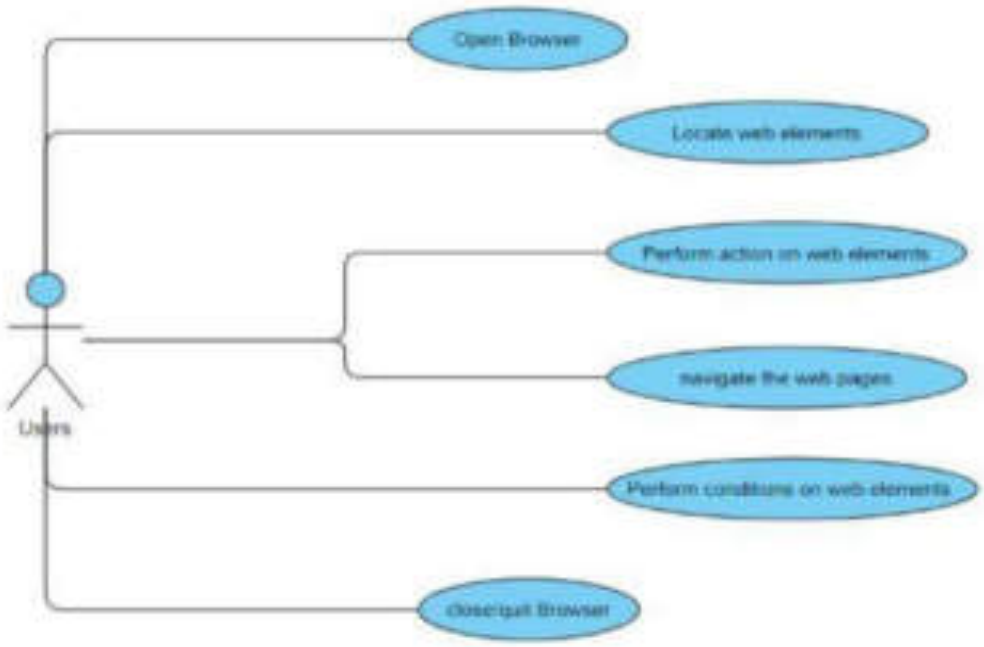


Fig 2.1 Sequence diagram for Web Automation

Sequence Diagram:

Sequence Diagrams are interaction diagrams that detail how operations are carried out. They capture the interaction between objects in the context of a collaboration. Sequence Diagrams are time focus and they show the order of the interaction visually by using the vertical axis of the diagram to represent time what messages are sent and when.

Sequence Diagrams captures:

- the interaction that takes place in a collaboration that either realizes a use case or an operation (instance diagrams or generic diagrams)
- high-level interactions between user of the system and the system, between the system and other systems, or between subsystems (sometimes known as system sequence diagrams)

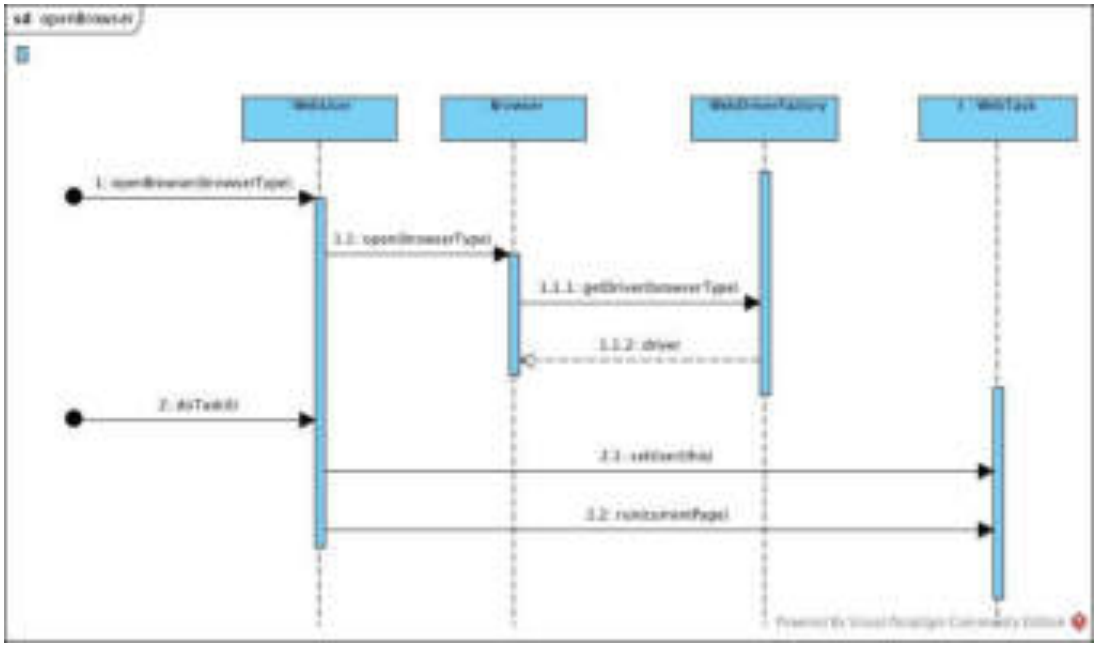


Fig 2.2 Sequence diagram for Web Automation

4.3 SOFTWARE REQUIREMENTS

- **Operating System** - Windows 7/8/10
- **Programming Language** - Python
- **IDE** - PyCharm
- **Testing Tools** – Selenium

4.4 HARWARE REQUIREMENTS:

- **Processor -** intel i5
- **Ram** - 4GB
- **Hard disk** - 500 GB

CHAPTER-5

NAVIGATION COMMANDS

5.1 INTRODUCTION

The first thing you'll want to do with WebDriver is navigate to a link. The normal way to do this is by calling get method.

```
driver.get("http://www.google.com")
```

WebDriver will wait until the page has fully loaded (that is, the onload event has fired) before returning control to your test or script. "Be aware that if your page uses a lot of AJAX on load then WebDriver may not know when it has completely loaded". If you need to ensure such pages are fully loaded then you can use waits.

5.2 INTERACTING WITH THE PAGE

Just being able to go to places isn't terribly useful. What we'd really like to do is to interact with the pages, or, more specifically, the HTML elements within a page. First of all, we need to find one. WebDriver offers a number of ways to find elements. For example, given an element defined as:

```
<input type="text" name="passwd" id="passwd-id" />
```
you

could find it using any of:

```
element = driver.find_element_by_id("passwd-id")
element =

driver.find_element_by_name("passwd")
element =

driver.find_element_by_xpath("//input[@id='passwd-id']")
element =

driver.find_element_by_css_selector("input#passwd-id")
```
You can

also look for a link by its text, but be careful! The text must be an exact match! You should also be careful when using XPATH in WebDriver. If there's more than one element that matches the query, then only the first will be returned. If nothing can be found, a NoSuchElementException will be raised.

WebDriver has an "Object-based" API; we represent all types of elements using the same interface. This means that although you may see a lot of possible methods you could invoke when you hit your IDE's auto-complete key combination, not all of them will make sense or be valid. Don't worry! WebDriver will attempt to do the Right Thing, and if you call a method that makes no sense ("setSelected()" on a "meta" tag, for example) an exception will be raised.

So, you've got an element. What can you do with it? First of all, you may want to enter some text into a text field:

```
element.send_keys("some text")
```

You can simulate pressing the arrow keys by using the "Keys" class:

```
element.send_keys(" and some", Keys.ARROW_DOWN)
```

It is possible to call send_keys on any element, which makes it possible to test keyboard shortcuts such as those used on Gmail. A side-effect of this is that typing something into a text field won't automatically clear it. Instead, what you type will be appended to what's already there. You can easily clear the contents of a text field or textarea with the clear method: element.clear()

5.3FILLING FORMS

We've already seen how to enter text into a textarea or text field, but what about the other elements? You can "toggle" the state of the drop down, and you can use "setSelected" to set something like an OPTION tag. Dealing with SELECT tags isn't too bad: **element = driver.**

find_element_by_xpath("//select[@name='name']")

all_options = element.find_elements_by_tag_name("option")

for option in all_options:

print ("Value is: %s" % option.get_attribute("value"))

option.click()

This will find the first "SELECT" element on the page, and cycle through each of its OPTIONs in turn, printing out their values, and selecting each in turn.

As you can see, this isn't the most efficient way of dealing with SELECT elements. WebDriver's support classes include one called a "Select", which provides useful methods for interacting with these:

from selenium.webdriver.support.ui import Select

select = Select(driver.find_element_by_name('name'))

select.select_by_index(index)

select.select_by_visible_text("text")

select.select_by_value(value)

WebDriver also provides features for deselecting all the selected options:

```
select = Select(driver.find_element_by_id('id'))

select.deselect_all()
```

This will deselect all OPTIONs from that particular SELECT on the page.

Suppose in a test, we need the list of all default selected options, select class provides a property method that returns a list:

```
select = Select(driver.find_element_by_xpath("//select[@name='name']"))

all_selected_options = select.all_selected_options
```

To get all available options:

```
options = select.options
```

Once you've finished filling out the form, you probably want to submit it. One way to do this would be to find the "submit" button and click it:

```
# Assume the button has the ID "submit" :)

driver.find_element_by_id("submit").click()
```

Alternatively, WebDriver has the convenience method "submit" on every element. If you call this on an element within a form, WebDriver will walk up the DOM until it finds the enclosing form and then calls submit on that. If the element isn't in a form, then the NoSuchElementException will be raised:

```
element.submit()
```

5.4 MOVING BETWEEN WINDOWS AND FRAMES

It's rare for a modern web application not to have any frames or to be constrained to a single window. WebDriver supports moving between named windows using the "switch_to_window" method:

driver.switch_to_window("windowName")

All calls to the driver will now be interpreted as being directed to the particular window. But how do you know the window's name? Take a look at the javascript or link that opened it:

Click here to open a new window

Alternatively, you can pass a "window handle" to the "switch_to_window()" method. Knowing this, it's possible to iterate over every open window like so: **for handle in driver.window_handles:**

driver.switch_to_window(handle)

You can also swing from frame to frame (or into iframes): **driver.switch_to_frame("frameName")**

It's possible to access subframes by separating the path with a dot, and you can specify the frame by its index too. That is:

driver.switch_to_frame("frameName.0.child") would go to the frame named "child" of the first subframe of the frame called "frameName". All frames are evaluated as if from *top*.

Once we are done with working on frames, we will have to come back to the parent frame which can be done using:

driver.switch_to_default_content()

5.5 POPUP DIALOGS

Selenium WebDriver has built-in support for handling popup dialog boxes. After you've triggered action that would open a popup, you can access the alert with the following:

alert = driver.switch_to.alert

This will return the currently open alert object. With this object, you can now accept, dismiss, read its contents or even type into a prompt. This interface works equally well on alerts, confirms, prompts. Refer to the API documentation for more information.

5.6 NAVIGATION

Earlier, we covered navigating to a page using the "get" command

(driver.get("http://www.example.com"))

As you've seen, WebDriver has a number of smaller, task-focused interfaces, and navigation is a useful task. To navigate to a page, you can use get method: **driver.get("http://www.example.com")**

To move backward and forward in your browser's history:

driver.forward()

driver.back()

driver.refresh()

Please be aware that this functionality depends entirely on the underlying driver. It's just possible that something unexpected may happen when you call these methods if you're used to the behavior of one browser over another.

5.7 COOKIES

Before moving to the next section of the tutorial, you may be interested in understanding how to use cookies. First of all, you need to be on the domain that the cookie will be valid for:

Go to the correct domain

driver.get("http://www.example.com") Now set the cookie.

This one's valid for the entire domain **cookie**

= {'name' : 'foo', 'value' : 'bar'}

driver.add_cookie(cookie)

CHAPTER-6

LOCATING ELEMENTS

6.1 INTRODUCTION

There are various strategies to locate elements in a page. You can use the most appropriate one for your case. Selenium provides the following methods to locate elements in a page:

- **find_element_by_id**
- **find_element_by_name**
- **find_element_by_xpath**
- **find_element_by_link_text**
- **find_element_by_partial_link_text**
- **find_element_by_tag_name**
- **find_element_by_class_name**
- **find_element_by_css_selector**

To find multiple elements (these methods will return a list):

- **find_elements_by_name**
- **find_elements_by_xpath**
- **find_elements_by_link_text**
- **find_elements_by_partial_link_text**
- **find_elements_by_tag_name**
- **find_elements_by_class_name**
- **find_elements_by_css_selector**

Apart from the public methods given above, there are two private methods which might be useful for locating page elements:

- **find_element**
- **find_elements**

Example usage:

```
fromselenium.webdriver.common.byimport   By

driver.find_element(By.XPATH,'//button[text()="Some text"]')

driver.find_elements(By.XPATH, '//button')
```

These are the attributes available for By class:

```
ID = "id"

XPATH = "xpath"

LINK_TEXT  = "link text"

PARTIAL_LINK_TEXT  = "partial link text"

NAME = "name"

TAG_NAME = "tag name"

CLASS_NAME = "class name"

CSS_SELECTOR  = "css selector"
```

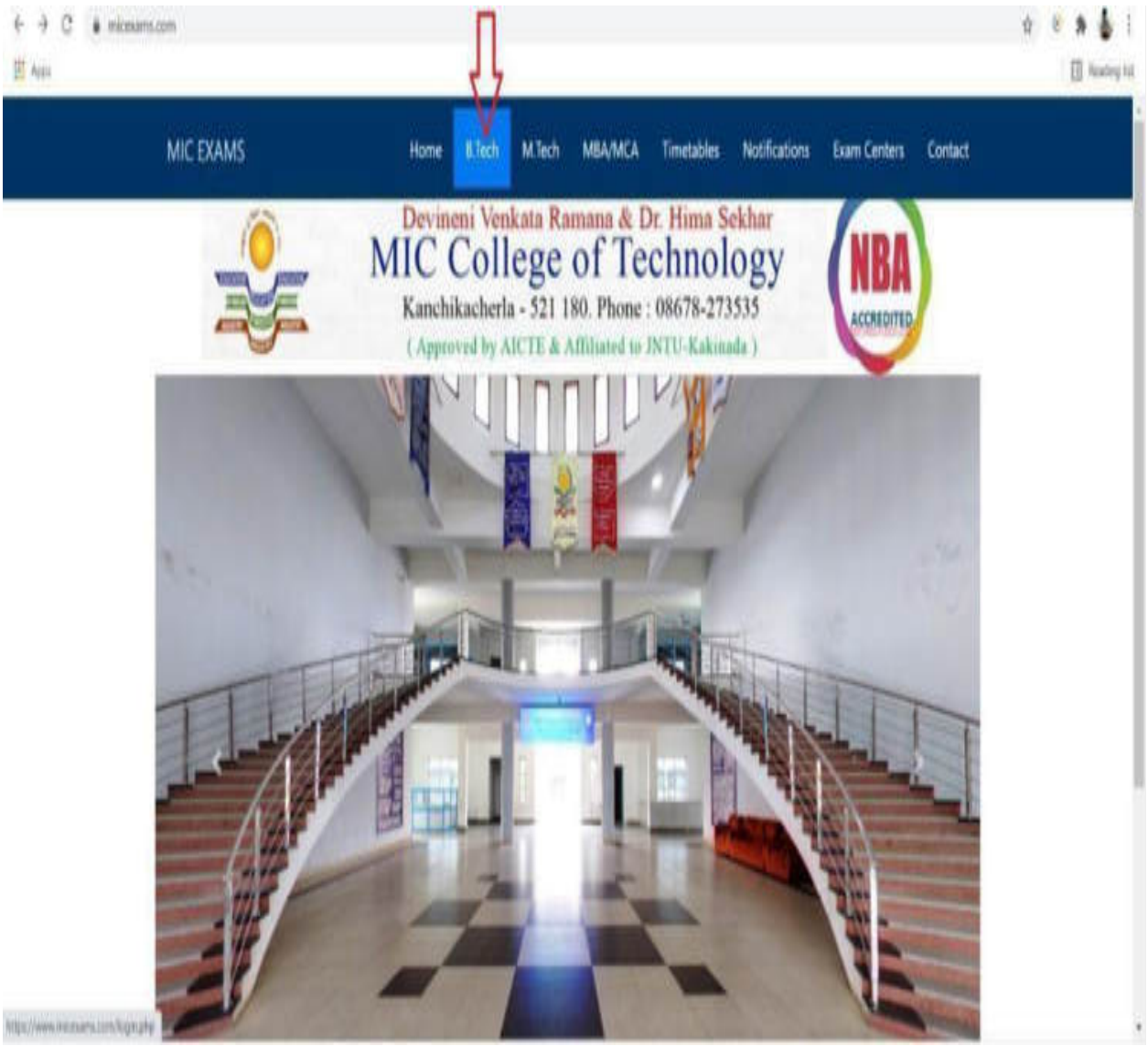


Fig 2.3 Locator-Web Elements

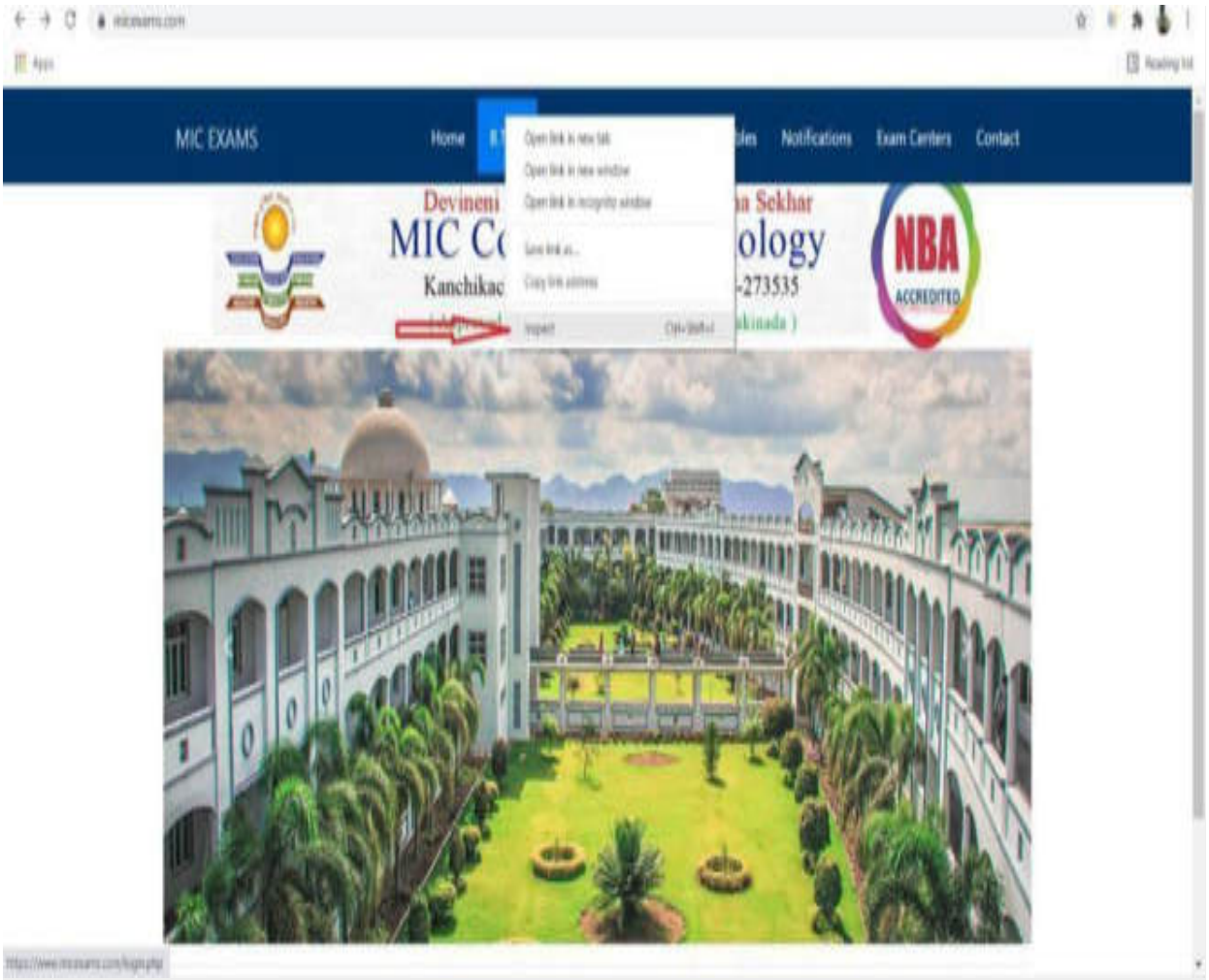


Fig 2.4 Locators

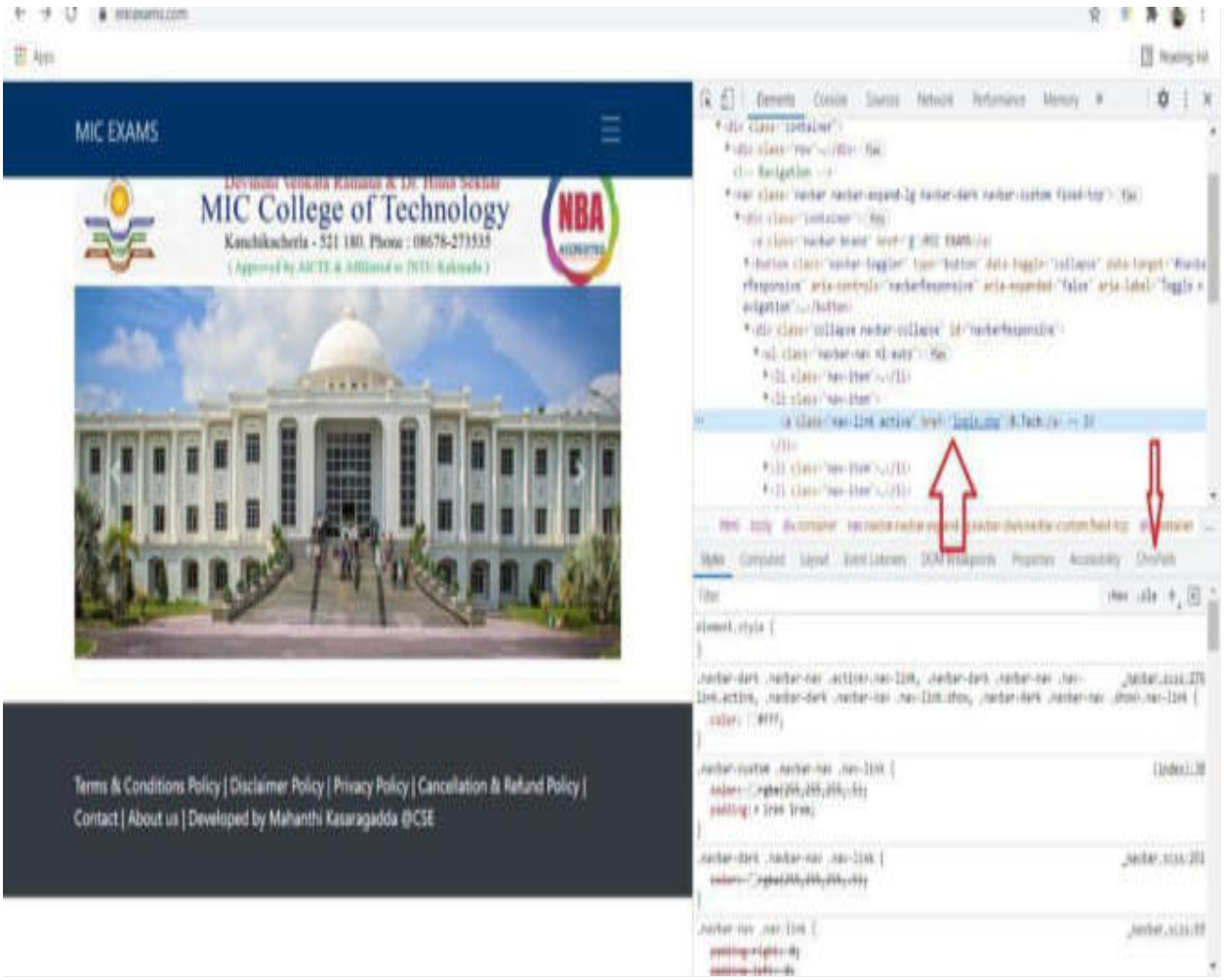


Fig 2.5 Locators-Inspect

6.2 LOCATING BY ID

Use this when you know the id attribute of an element. With this strategy, the first element with a matching id attribute will be returned. If no element has a matching id attribute, a NoSuchElementException will be raised.
For instance, consider this page source:

```
<html>

<body>

<form id="loginForm">

<input name="username" type="text" />

<input name="password" type="password" />

<input name="continue" type="submit" value="Login" />

</form>

</body>

</html>
```

The form element can be located like this:

```
login_form = driver.find_element_by_id('loginForm')
```

6.3 LOCATING BY NAME

Use this when you know the name attribute of an element. With this strategy, the first element with a matching name attribute will be returned. If no element has a matching name attribute, a NoSuchElementException will be raised.
For instance, consider this page source:

```
<html>

<body>

<form id="loginForm">

<input name="username" type="text" />

<input name="password" type="password" />

<input name="continue" type="submit" value="Login" />

<input name="continue" type="button" value="Clear" />

</form>

</body>

</html>
```

The username & password elements can be located like this:

```
username = driver.find_element_by_name('username')

password = driver.find_element_by_name('password')
```

This will give the "Login" button as it occurs before the "Clear" button:

```
continue = driver.find_element_by_name('continue')
```

6.4 LOCATING BY XPATH

XPath is the language used for locating nodes in an XML document. As HTML can be an implementation of XML (XHTML), Selenium users can leverage this powerful language to target elements in their web applications. XPath supports the simple methods of locating by id or name attributes and extends them by opening up all sorts of new possibilities such as locating the third checkbox on the page.

One of the main reasons for using XPath is when you don't have a suitable id or name attribute for the element you wish to locate. You can use XPath to either locate the element in absolute terms (not advised), or relative to an element that does have an id or name attribute. XPath locators can also be used to specify elements via attributes other than id and name.

Absolute XPaths contain the location of all elements from the root (html) and as a result are likely to fail with only the slightest adjustment to the application. By finding a nearby element with an id or name attribute (ideally a parent element) you can locate your target element based on the relationship. This is much less likely to change and can make your tests more robust.

For instance, consider this page source:

```
<html>
<body>
<form id="loginForm">
<input name="username" type="text" />
<input name="password" type="password" />
<input name="continue" type="submit" value="Login" />
<input name="continue" type="button" value="Clear" />
</form>
</body>
</html>
```

The form elements can be located like this:

login_form = driver.find_element_by_xpath("/html/body/form[1]")

login_form = driver.find_element_by_xpath("//form[1]")

login_form = driver.find_element_by_xpath("//form[@id='loginForm']")

1. Absolute path (would break if the HTML was changed only slightly)
2. First form element in the HTML
3. The form element with attribute id set to loginForm

The username element can be located like this:

username = driver.find_element_by_xpath("//form[input/@name='username']")

username = driver.find_element_by_xpath("//form[@id='loginForm']/input[1]")

username = driver.find_element_by_xpath("//input[@name='username']")

1. First form element with an input child element with name set to username
2. First input child element of the form element with attribute id set to loginForm
3. First input element with attribute name set to username

The "Clear" button element can be located like this:

clear_button

driver.find_element_by_xpath("//input[@name='continue'][@type='button']")

clear_button = driver.find_element_by_xpath("//form[@id='loginForm']/input[4]")

1. Input with attribute name set to continue and attribute type set to button
2. Fourth input child element of the form element with attribute id set to loginForm

These examples cover some basics, but in order to learn more, the following references are recommended:

- W3Schools XPath Tutorial
- W3C XPath Recommendation
- XPath Tutorial - with interactive examples.

Here is a couple of very useful Add-ons that can assist in discovering the XPath of an element:

- xpath Finder - Plugin to get the elements xpath.
- XPath Helper - for Google Chrome

6.5 LOCATING HYPERLINKS BY LINK TEXT

Use this when you know the link text used within an anchor tag. With this strategy, the first element with the link text matching the provided value will be returned. If no element has a matching link text attribute, a NoSuchElementException will be raised.

For instance, consider this page source:

```
<html>
<body>
<p>Are you sure you want to do this?</p>
<a href="continue.html">Continue</a>
<a href="cancel.html">Cancel</a>
```

6.6 LOCATING ELEMENTS BY TAG NAME

Use this when you want to locate an element by tag name. With this strategy, the first element with the given tag name will be returned. If no element has a matching tag name, a NoSuchElementException will be raised.

For instance, consider this page source:

```
<html>

<body>
<h1>Welcome</h1>
<p>Site content goes here.</p>
</body> </html>
```

The heading (h1) element can be located like this:

```
 heading1 = driver.find_element_by_tag_name('h1')
</body> </html>
```

The continue.html link can be located like this:

```
continue_link = driver.find_element_by_link_text('Continue')

continue_link = driver.find_element_by_partial_link_text('Conti')
```

6.7 LOCATING ELEMENTS BY CLASS NAME

Use this when you want to locate an element by class name. With this strategy, the first element with the matching class name attribute will be returned. If no element has a matching class name attribute, a NoSuchElementException will be raised.
For instance, consider this page source:

```
<html>

<body>

<p class="content">Site content goes here.</p>

</body>

</html>
```

The “p” element can be located like this:

```
content = driver.find_element_by_class_name('content')
```

6.8 LOCATING ELEMENTS BY CSS SELECTORS

Use this when you want to locate an element using CSS selector syntax. With this strategy, the first element matching the given CSS selector will be returned. If no element matches the provided CSS selector, a NoSuchElementException will be raised.
For instance, consider this page source:

```
<html>

<body>

<p class="content">Site content goes here.</p>

</body>

</html>
```

The “p” element can be located like this:

```
content = driver.find_element_by_css_selector('p.content')
```

CHAPTER – 7

CONDITIONAL COMMANDS

These commands are used to find that some elements are present or not, displayed or not.

The following are types. They are:

- **Is_displayed()** - It is used for any type of web elements.
- **Is_enabled()** - It is used for any type of web elements.
- **Is_selected()** - It is used for only checkboxes & radio buttons.

CHAPTER – 8

IMPLEMENTATION

8.1 TECHNOLOGIES

8.1.1 PYTHON

Guido van Rossum in 1991 and further developed by the Python Software Foundation. It was designed with an emphasis on code readability, and its syntax allows programmers to express their concepts in fewer lines of code.

Python is a programming language that lets you work quickly and integrate systems more efficiently.

Python is an easy to learn, powerful programming language. It has efficient high-level data structures and a simple but effective approach to object-oriented programming. Python's elegant syntax and dynamic typing, together with its interpreted nature, make it an ideal language for scripting and rapid application development in many areas on most platforms.

The Python interpreter and the extensive standard library are freely available in source or binary form for all major platforms from the Python Web site, https://www.python.org/, and may be freely distributed. The same site also contains distributions of and pointers to many free third party Python modules, programs and tools, and additional documentation.

The Python interpreter is easily extended with new functions and data types implemented in C or C++ (or other languages callable from C). Python is also suitable as an extension language for customizable applications.

For a description of standard objects and modules, see The Python Standard Library. The Python Language Reference gives a more formal definition of the language. To write extensions in C or C++, read Extending and Embedding the Python Interpreter and Python/C API Reference Manual. There are also several books covering Python in depth.

Python 2 vs Python 3

In most of the programming languages, whenever a new version releases, it supports the features and syntax of the existing version of the language, therefore, it is easier for the projects to switch in the newer version. However, in the case of Python, the two versions Python 2 and Python 3 are very much different from each other.

A list of differences between Python 2 and Python 3 are given below:

- Python 2 uses **print** as a statement and used as print "something" to print some string on the console. On the other hand, Python 3 uses **print** as a function and used as print("something") to print something on the console.
- Python 2 uses the function raw_input to accept the user's input. It returns the string representing the value, which is typed by the user. To convert it into the integer, we need to use the int() function in Python. On the other hand, Python 3 uses input() function which automatically interpreted the type of input entered by the user. However, we can cast this value to any type by primitive functions(int(), str(), etc.).
- In Python 2, the implicit string type is ASCII, whereas, in Python 3, the implicit string type is Unicode.
- Python 3 doesn't contain the xrange() function of Python 2. The xrange() is the variant of range()function which returns a xrange object that works similar to Java iterator. The range() returns a list for example the function range(0,3) contains 0,1,2.
- There is also a small change made in Exception handling in Python 3. It defines a keyword **as** while is necessary to be used.

Why learn Python?

Python provides many useful features to the programmer. These features make it most popular and widely used language. We have listed below few-essential feature of Python.

- Easy to use and Learn
- Expressive Language

- Interpreted Language
- Object-Oriented Language
- Open Source Language
- Extensible
- Learn Standard Library
- GUI Programming Support
- Integrated
- Embeddable
- Dynamic Memory Allocation
- Wide Range of Libraries and Frameworks

Where is Python used?

Python is a general-purpose, popular programming language and it is used in almost every technical field. The various areas of Python use are given below.

- Data Science
- Date Mining
- Desktop Applications
- Console-based Applications
- Mobile Applications
- Software Development
- Artificial Intelligence
- Web Applications
- Enterprise Applications
- 3D CAD Applications
- Machine Learning
- Computer Vision or Image Processing Applications.
- Speech Recognitions

Beginning with Python programming:

1) Finding an Interpreter:

Before we start Python programming, we need to have an interpreter to interpret and run our programs. There are certain online interpreters like **https://ide.geeksforgeeks.org/**, http://ideone.com/ or http://codepad.org/ that can be used to run Python programs without installing an interpreter.

***Windows*:** There are many interpreters available freely to run Python scripts like IDLE (Integrated Development Environment) that comes bundled with the Python software downloaded from **http://python.org/**.

***Linux*:** Python comes preinstalled with popular Linux distros such as Ubuntu and Fedora. To check which version of Python you're running, type "python" in the terminal emulator. The interpreter should start and print the version number.

***macOS*:** Generally, Python 2.7 comes bundled with macOS. You'll have to manually install Python 3 from **http://python.org/**.

2) First Python Code:

Just type in the following code after you start the interpreter.

```
# Scripts Begins

Print ("Python Programming")

# Scripts Ends
```

Output:

```
Python Programming
```

Python is a programming language that lets you work quickly and integrate systems more efficiently.

Below are some facts about Python Programming Language:

1. Python is currently the most widely used multi-purpose, high-level programming language.
2. Python allows programming in Object-Oriented and Procedural paradigms.

3. Python programs generally are smaller than other programming languages like Java. Programmers have to type relatively less and indentation requirement of the language, makes them readable all the time.
4. Python language is being used by almost all tech-giant companies like – Google, Amazon, Facebook, Instagram, Dropbox, Uber… etc.
5. The biggest strength of Python is huge collection of standard library which can be used for the following:
    - Machine Learning
    - GUI Applications (like Kivy, Tkinter, PyQtetc.)
    - Web frameworks like Django (used by YouTube, Instagram, Dropbox)
    - Image processing (like OpenCV, Pillow)
    - Web scraping (like Scrapy, BeautifulSoup, Selenium)
    - Test frameworks
    - Multimedia
    - Scientific computing
    - Text processing and many more..

8.1.2 PYTHON IDE(Pycharm):

PyCharm is the most popular IDE used for Python scripting language. This chapter will give you an introduction to PyCharm and explains its features.

PyCharm offers some of the best features to its users and developers in the following aspects −

- Code completion and inspection
- Advanced debugging
- Support for web programming and frameworks such as Django and Flask

Features of PyCharm:

Besides, a developer will find PyCharm comfortable to work with because of the features mentioned below −

Code Completion:

PyCharm enables smoother code completion whether it is for built in or for an external package.

SQLAlchemy as Debugger:

You can set a breakpoint, pause in the debugger and can see the SQL representation of the user expression for SQL Language code.

Git Visualization in Editor:

When coding in Python, queries are normal for a developer. You can check the last commit easily in PyCharm as it has the blue sections that can define the difference between the last commit and the current one.

Code Coverage in Editor:

You can run **.py** files outside PyCharm Editor as well marking it as code coverage details elsewhere in the project tree, in the summary section etc.

Package Management:

All the installed packages are displayed with proper visual representation. This includes list of installed packages and the ability to search and add new packages.

Local History:

Local History is always keeping track of the changes in a way that complements like Git. Local history in PyCharm gives complete details of what is needed to rollback and what is to be added.

Refactoring:

Refactoring is the process of renaming one or more files at a time and PyCharm includes various shortcuts for a smooth refactoring process.

User Interface of PyCharm Editor:

The user interface of PyCharm editor is shown in the screenshot given below. Observe that the editor includes various features to create a new project or import from an existing project.

Path to install Selenium in Pycharm

Settings -> Python -> Projectname -> Python Interpreter -> Selenium install

Pycharm Installation:

Steps involved:

You will have to follow the steps given below to install PyCharm on your system. These steps show the installation procedure starting from downloading the PyCharm package from its official website to creating a new project.

Step 1:

Download the required package or executable from the official website of PyCharm https://www.jetbrains.com/pycharm/download/#section=windowsHere you will observe two versions of package for Windows as shown in the screenshot given below –

Fig 2.6 Download PyCharm

Note that the professional packages involves all the advanced features and comes with free trail for few days and the user has to buy a licensed key for activation beyond the trail period. Community package is for free and can be downloaded and installed as and when required. It includes all the basic features need for installation.

Step 2:

Download the community package (executable file) onto your system and mention a destination folder as

Shown below –

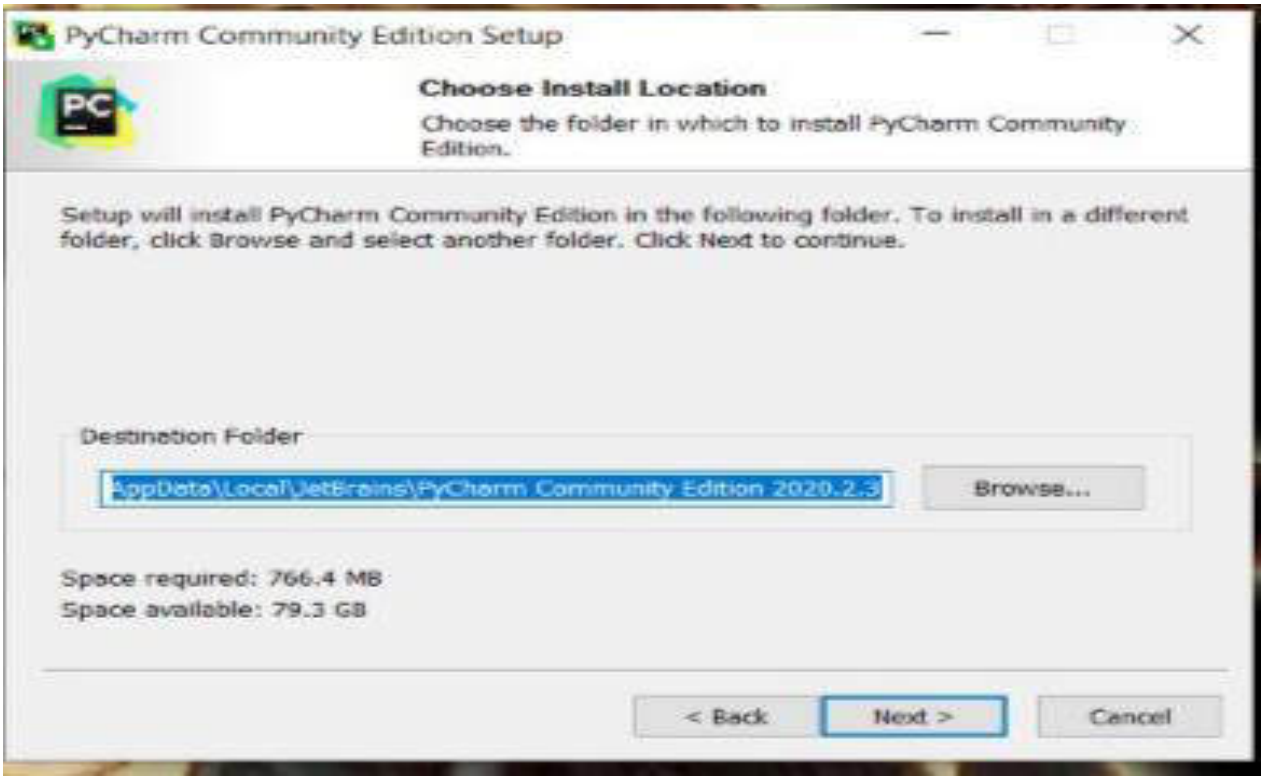


Fig 2.7 setup at destination folder

Step 3:

Now, begin the installation procedure similar to any other software package.

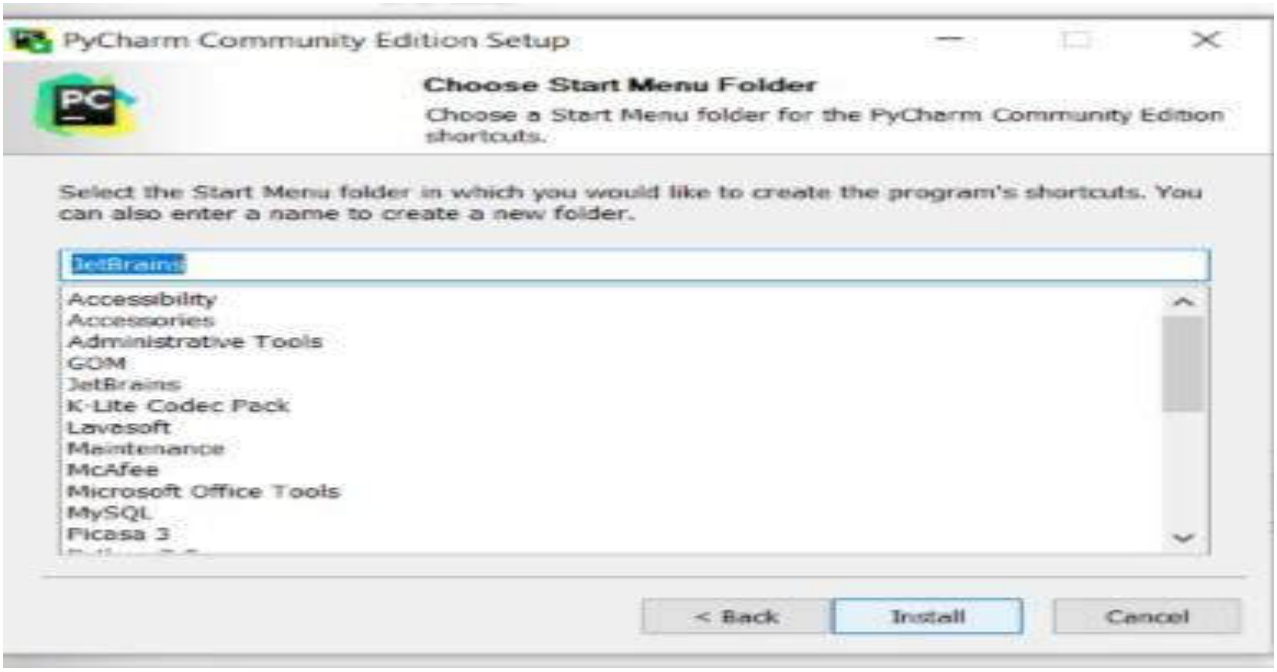


Fig 2.8 Use of Software

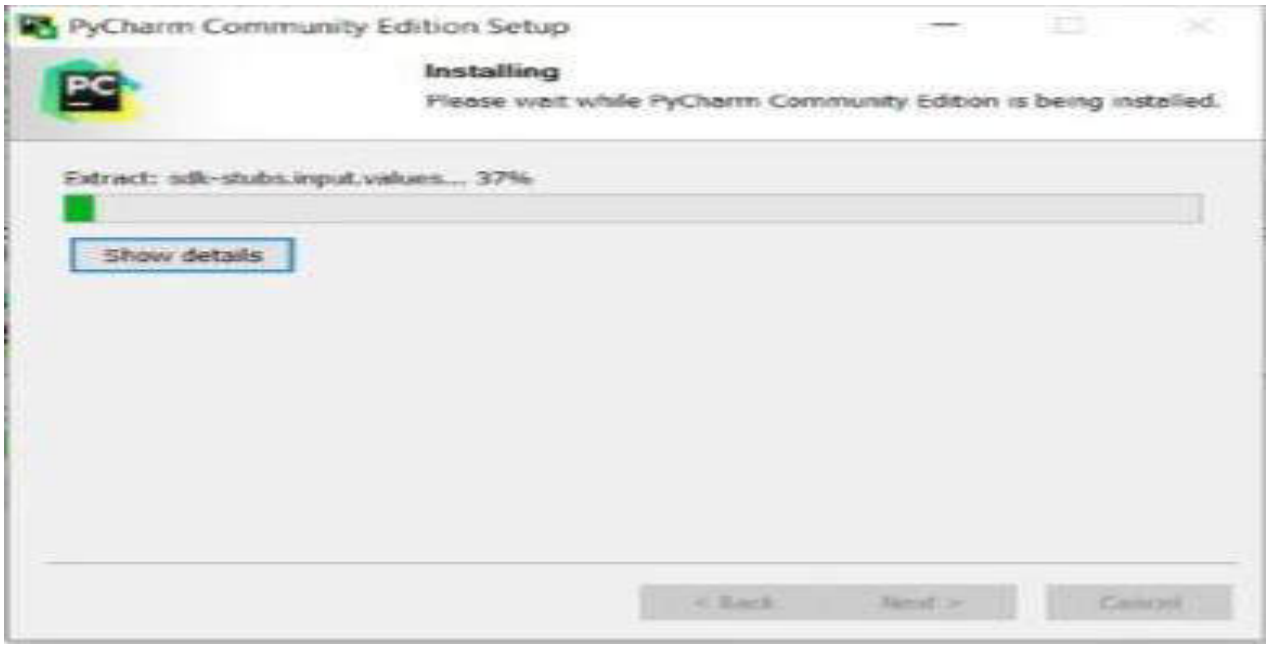


Fig 2.8 Installing

Step 4:

Once the installation is successful. PyCharm asks you to import settings of the existing package if any

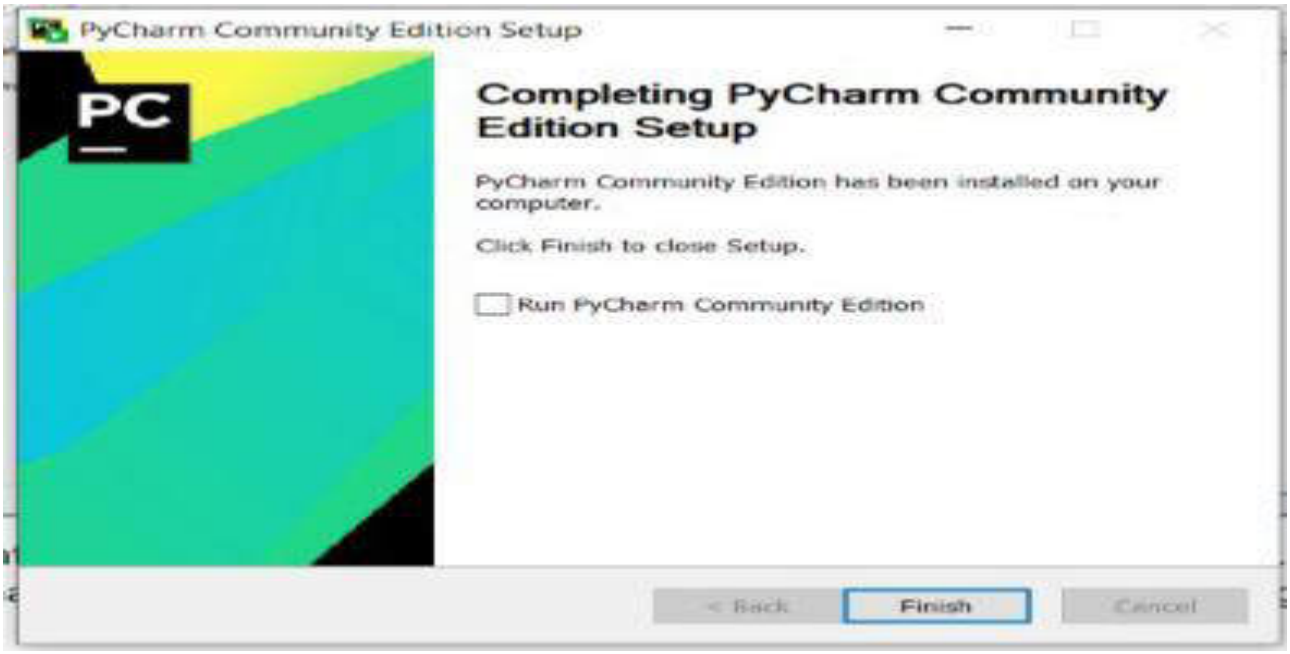


Fig 2.9 Installation successful

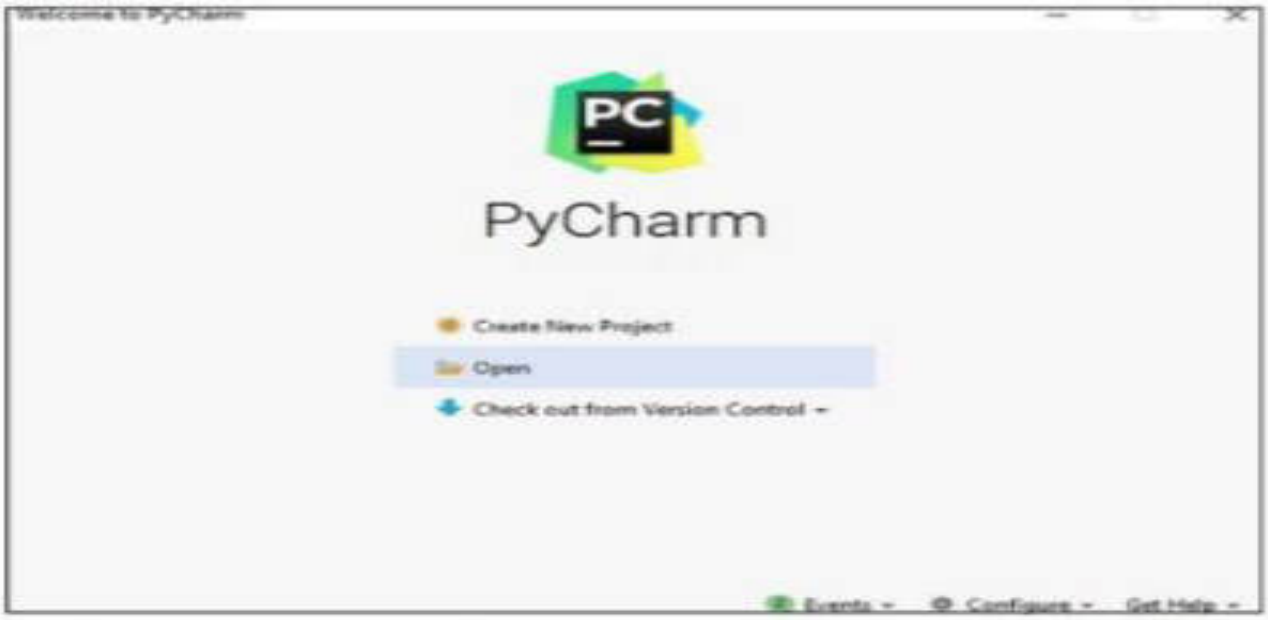


Fig 3.0 Open PyCharm

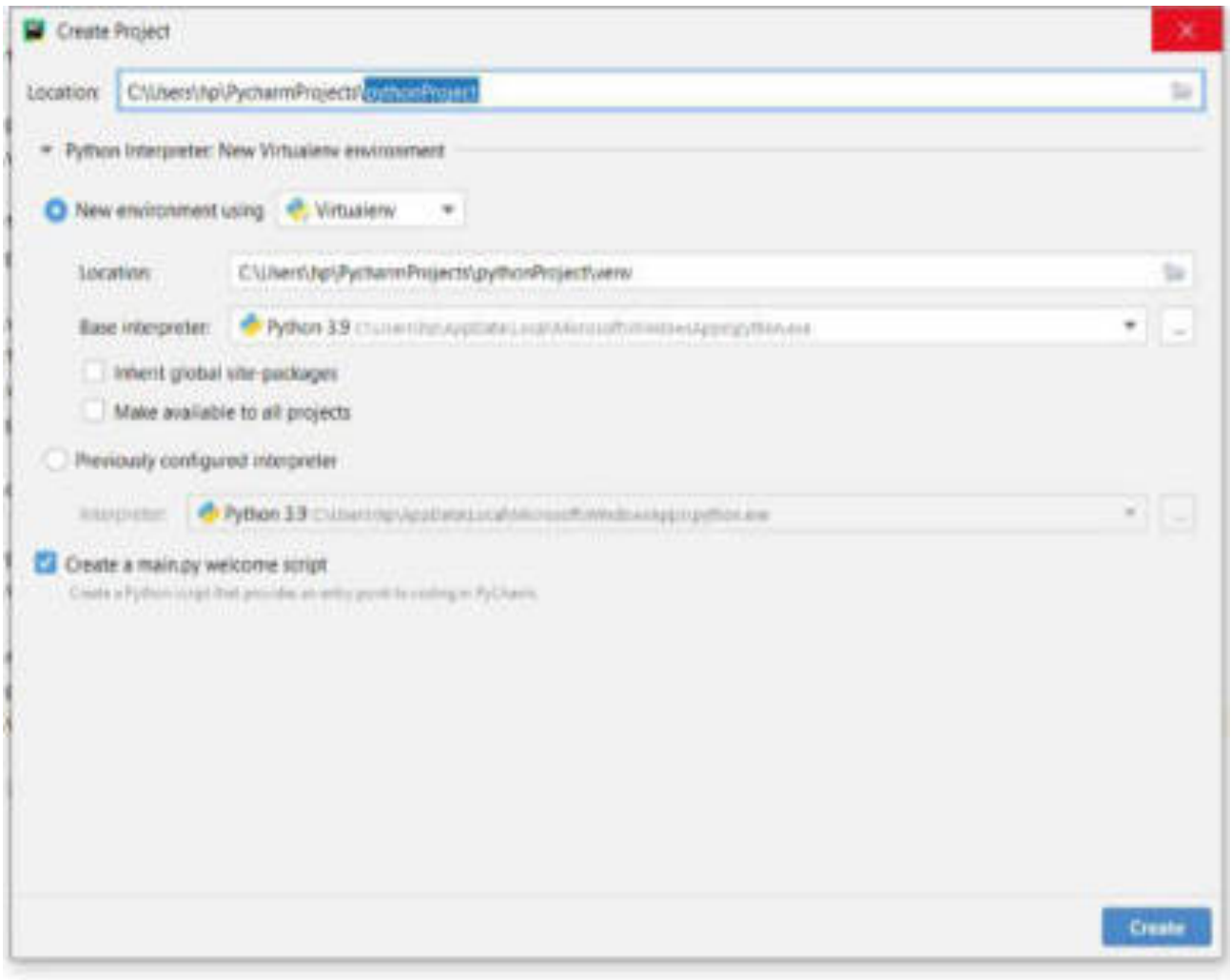


Fig 3.1 Execute the Project

8.2 BROWSER DRIVER DOWNLOAD:

To download the Drivers we follows the below path:

SeleniumHQ.org -> downloads -> browsers

Firefox - GeckoDriver version specified with OS

Google - ChromeDriver version specified with OS

IE - IEDriverServer version specified with OS

There are other Drivers like

1) Opera Driver
2) Android Driver
3) HTML unit Driver
4) Edge Driver
5) Safari Driver
6) Iphone driver
7) Remote web Driver

8.3 The Seven Basic Steps of Selenium Tests:

- Create a WebDriver instance.
- Navigate to a Web page.
- Locate an HTML element on the Web page.
- Perform an action on an HTML element.
- Anticipate the browser response to the action.
- Run tests and record test results using a test framework.
- Conclude by quitting the browser and web driver

8.4 CODING:

1) Opening of WEB Application in Multiple Browsers.

Cross-Browsing testing: It is a type of functional test to check that our web application works as expected in different browsers. Ex Chrome, Firefox, IE, Edge etc.

Fig 3.2 Cross Browsering

CODE:

```
from selenium import webdriver

#Chrome driver

driver =
webdriver.Chrome(executable_path="C:\Drivers\chromedriver_win32\chromedriver.exe"
)

#Firefox driver

driver = webdriver.Firefox(executable_path="C:\Drivers\geckodriver-
v0.29.1win64\geckodriver.exe")

#InternetExplorer driver

driver =
webdriver.Ie(executable_path="C:\Drivers\IEDriverServer_Win32_2.40.0\IEDriverServer
.exe")

#URL to open the page

driver.get("https://www.hotstar.com")

driver.maximize_window()   #Maximize the page window

print(driver.title)  #title of the page

print(driver.current_url)  #url of the page

print(driver.page_source)  #HTML code for the page

driver.close()    #close the current page

driver.quit()      #closes all the pages
```

2) Accessing of WEB elements on page by using Locators

Locators: Locators are the basic building blocks of selenium WebDriver, as it is used for identification of web element/object from the web pages and also these locators provides simple and effective way to interact with web pages using simple commands.

Locating elements in WebDriver is done by using the findElement() and findElements() methods

CODE:

```
from selenium import webdriver
from selenium.webdriver.common.keys import Keys

driver =
webdriver.Chrome(executable_path="C:\Drivers\chromedriver_win32\chromedriver.exe"
)

driver.get("https://www.amazon.com")

# Using id locator

driver.find_element_by_id("student-link").click()

driver.find_element_by_id("twotabsearchtextbox").send_keys("phones")

# Using name locator

driver.find_element_by_name("field-keywords").send_keys("selenium")

driver.find_element_by_name("Go").click()

driver.maximize_window()

# Using linktext locator

driver.find_element_by_link_text("Log In").click()

# Using css_selector locator

driver.find_element_by_css_selector("").send_keys("python")
```

 # *Using partiallinttext locator*

```
driver.find_element_by_partial_link_text("Log").click()
```

 # *Using Xpath locator*

```
driver.find_element_by_xpath("//header/div[@id='navbar']/div[@id='navmain']/div[1]/a[1]/span
[1]").click()

driver.close()
```

3) Navigation of WEB pages.

CODE:

```
from selenium import webdriver

from selenium.webdriver.common.keys import Keys

import time driver =

webdriver.Chrome(executable_path="C:\Drivers\chromedriver_win32\chromedriver.exe"
)

driver.maximize_window()

driver.get("http://www.amazon.com/")  #first page

time.sleep(5)

print(driver.title)

driver.get("http://www.flipkart.com/") #second page

time.sleep(5)

print(driver.title)
```

```
driver.back()

time.sleep(5)

print(driver.title)

driver.forward()

time.sleep(5)

print(driver.title)

driver.close()
```

4) Use of Conditional Commands

CODE:

```
from selenium import webdriver

import time

#Chrome driver
driver =
webdriver.Chrome(executable_path="C:\Drivers\chromedriver_win32\chromedriver.exe"
)

driver.get("http://www.coursera.com/")

driver.maximize_window()

driver.find_element_by_link_text("Log In").click()

time.sleep(2)

el = driver.find_element_by_name("email")

print(el.is_displayed())
```

```
print(el.is_enabled())

el.send_keys("psmd1999@gmail.com")

ps = driver.find_element_by_name("password")

print(el.is_displayed())

print(el.is_enabled())

ps.send_keys("Psmd1999@")

time.sleep(2)

driver.find_element_by_xpath("//button[contains(text(),'Login')]").click()

driver.get("www.goole.com")

w=driver.find_element_by_xpath("//body/div[7]/div[3]/div[1]/div[2]/div[1]/div[2]/div[1]/div
[1]/div[1]/div[1]/div[1]/div[1]/div[1]/div[1]/div[1]/div[1]/div[1]/div[1]/span[1]")

print(w.is_selected())

q = driver.find_element_by_name("r52e6ee7ae84041e7a0dc9690c293df3a")

print(q.is_selected()) driver.close()
```

CHAPTER-9

CONCLUSION

When developing web software, the ultimate goal of the tester or developer is to ensure that the application is tested often and thoroughly. More often than not, creating automated test scripts is the best way to be sure that this goal is accomplished. In particular, the developer wants to be sure to create maintainable test scripts that will last through the many changes that applications undergo.

An Open Source test tool, Selenium IDE has many advantages, including an easy to use record and playback tool, and the ability to test scripts inside of the browser. Finally, we are going to automate a web browser or any web-based application by writing selenium script.

CHAPTER – 10

BIBLIOGRAPHY

WEBSITES:

- https://www.seleniumHQ.org
- https://www.guru99.com/introduction-to-selenium.html
- https://www.geeksforgeeks.org/selenium-python-tutorial

TEXT BOOKS:

- Selenium with Python: Automation Testing with Python by Anand Hooda.
- Let's Learn Selenium Using Python Kindle Edition by David Mayank Johri.
- Selenium WebDriver automation tool training with Java code and programs, eclipse, TestNG, POM framework, Practical Guide, Tips and Tricks, example test case, step by step tutorial: Automation Testing by Pratik Raju.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

SECURE DATA SHARING IN CLOUD COMPUTING USING RS-IBE ALGORITHMS

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
COMPUTER SCIENCE AND ENGINEERING

Submitted by

P NANDINI (17H71A0588)

Y MAMATHA (17H71A0585)

B NARENDRA (17H71A0589)

P PRABHU KUMAR(17H71A0593)

Under the Esteemed Guidance of
MD. ELIAZ
ASSISTANT PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERIN

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|EEE|ME)
Accredited by
NAAC With Grade A

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"SECURE DATA SHARING IN CLOUD COMPUPTING USING RS-IBE ALGORITHM"** is a bonafide work carried out by **MS.P NANDINI (17H71A0588), MS. Y MAMATHA (17H71A0585), MR.B NARENDRA (17H71A0589), MR. P PRABHU KUMAR (17H71A0593)** in partial fulfilment for the award of degree of Bachelor of Technology in **Computer Science and Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

MD. Eliaz
Project Guide

D. Prasad
Head of the Department

Examiner

Dr K Srinivas Rao
PRINCIPAL

ACKNOWLEDGEMENT

We would like to take this opportunity to express my deepest appreciation to the following people for their valuable contributions and assistance with this Project.

Initially, we would like to thank my project supervisor, **Mr.MD Eliaz, Assistant Professor, Department of Computer Science and Engineering** for the guidance and support, especially for the valuable ideas and knowledge shared to me throughout the Project.

We have the immense pleasure in expressing my thanks and deep sense of gratitude to**, Mr. D. Prasad, Head of the Department, Computer Science and Engineering** for extending necessary facilities and support for the completion of the Project.

We whole sincerely acknowledge **Dr. K. Srinivas, Principal and Prof D. PANDURANGA RAO, CEO** for giving opportunity to take up the Project work. We also extend my thanks to all faculty members of **Computer Science and Engineering**, for their valuable guidance and encouragement.

We would like to extend my warm appreciation to all my friends for sharing us their knowledge, valuable contributions and help with this Project.

Finally, my special thanks go to my family for their continuous support and help throughout my academic years and for their continual support and encouragement for the completion of the project.

DECLARATION

We **P NANDINI (17H71A0588), Y MAMATHA (17H71A0585), B NARENDRA (17H71A0589), P PRABHU KUMAR (17H71A0593),** of the Main-Project "**SECURE DATA SHARING IN CLOUD COMPUTING USING RS-IBE ALGORITHM**", hereby declare that the matter embodied in this Project is the genuine work done by us and has not been submitted either to this university or to any other university/institute for the fulfilment of the requirement of any course of study.

P NANDINI (17H71A0588)
Y MAMATHA (17H71A0585)
B NARENDRA (17H71A0589)
P PRABHU KUMAR (17H71A0593)

Abstract

Cloud computing provides a flexible and convenient way for data sharing, which brings various benefits for both the society and individuals. But there exists a natural resistance for users to directly outsource the shared data to the cloud server since the data often contain valuable information. Thus, it is necessary to place cryptographically enhanced access control on the shared data. Identity-based encryption is a promising crypto graphical primitive to build a practical data sharing system. However, access control is not static. That is, when some user's authorization is expired, there should be a mechanism that can remove him/her from the system. Consequently, the revoked user cannot access both the previously and subsequently shared data. To this end, we propose a notion called revocable-storage identity-based encryption (RS-IBE), which can provide the forward/backward security of cipher text by introducing the functionalities of user revocation and cipher text update simultaneously. Furthermore, we present a concrete construction of RS-IBE, and prove its security in the defined security model. The performance comparisons indicate that the proposed RS-IBE scheme has advantages in terms of functionality and efficiency, and thus is feasible for a practical and cost-effective data-sharing system. Finally, we provide implementation results of the proposed scheme to demonstrate its practicability.

Index

1.1 INTRODUCTION TO PROJECT

CLOUD computing is a paradigm that provides massive computation capacity and huge memory space at a low cost. It enables users to get intended services irrespective of time and location across multiple platforms (e.g., mobile devices, personal computers), and thus brings great convenience to cloud users. Among numerous services provided by cloud computing, cloud storage service, such as Apple's iCloud, Microsoft's Azure and Amazon's S3, can offer a more flexible and easy way to share data over the Internet, which provides various benefits for our society. However, it also suffers from several security threats, which are the primary concerns of cloud users.

ORGANIZATION PROFILE

Software Solutions is an IT solution provider for a dynamic environment where business and technology strategies converge. Their approach focuses on new ways of business combining IT innovation and adoption while also leveraging an organization's current IT assets. Their work with large global corporations and new products or services and to implement prudent business and technology strategies in today's environment.

RANGE OF EXPERTISE INCLUDES:

- Software Development Services
- Engineering Services
- Systems Integration
- Customer Relationship Management
- Product Development
- Electronic Commerce
- Consulting
- IT Outsourcing

We apply technology with innovation and responsibility to achieve two broad objectives:

- Effectively address the business issues our customers face today.
- Generate new opportunities that will help them stay ahead in the future.

THIS APPROACH RESTS ON:

- A strategy where we architect, integrate and manage technology services and solutions - we call it AIM for success.
- A robust offshore development methodology and reduced demand on customer resources.
- A focus on the use of reusable frameworks to provide cost and times benefits.

They combine the best people, processes and technology to achieve excellent results - consistency. We offer customers the advantages of:

SPEED:

They understand the importance of timing, of getting there before the competition. A rich portfolio of reusable, modular frameworks helps jump-start projects. Tried and tested methodology ensures that we follow a predictable, low - risk path to achieve results. Our track record is testimony to complex projects delivered within and evens before schedule.

EXPERTISE:

Our teams combine cutting edge technology skills with rich domain expertise. What's equally important - they share a strong customer orientation that means they actually start by listening to the customer. They're focused on coming up with solutions that serve customer requirements today and anticipate future needs.

A FULL SERVICE PORTFOLIO:

They offer customers the advantage of being able to Architect, integrate and manage technology services. This means that they can rely on one, fully accountable source instead of trying to integrate disparate multi vendor solutions.

SERVICES:

Xxx is providing its services to companies which are in the field of production, quality control etc with their rich expertise and experience and information technology they are in best position to provide software solutions to distinct business requirements.

1.2 PURPOSE OF THE PROJECT

It seems that the concept of revocable identity-based encryption (RIBE) might be a promising approach that fulfills the aforementioned security requirements for data sharing.

RIBE features a mechanism that enables a sender to append the current time period to the cipher text such that the receiver can decrypt the cipher text only under the condition that he/she is not revoked at that time period.

A RIBE-based data sharing system works as follows:

Step 1: The data provider (e.g., David) first decides the users (e.g., Alice and Bob) who can share the data. Then, David encrypts the data under the identities Alice and Bob, and uploads the cipher text of the shared data to the cloud server.

Step 2: When either Alice or Bob wants to get the shared data, she or he can download and decrypt the corresponding cipher text. However, for an unauthorized user and the cloud server, the plaintext of the shared data is not available.

Step 3: In some cases, e.g., Alice's authorization gets expired, David can download the cipher text of the shared data, and then decrypt-then-re-encrypt the shared data such that Alice is prevented from accessing the plaintext of the shared data, and then upload the re-encrypted data to the cloud server again.

1.3Existing System:

Boneh and Franklin first proposed a natural revocation way for IBE. They appended the current time period to the cipher text, and non-revoked users periodically received private keys for each time period from the key authority.

Boldyreva, Goyal and Kumar introduced a novel approach to achieve efficient revocation. They used a binary tree to manage identity such that their RIBE scheme reduces the complexity of key revocation to logarithmic (instead of linear) in the maximum number of system users.

Subsequently, by using the aforementioned revocation technique, Libert and Vergnaud proposed an adaptively secure RIBE scheme based on a variant of Water's IBE scheme.

Chen et al. constructed a RIBE scheme from lattices.

Disadvantage of Existing System:

Unfortunately, existing solution is not scalable, since it requires the key authority to perform linear work in the number of non-revoked users. In addition, a secure channel is essential for the key authority and non-revoked users to transmit new keys.

However, existing scheme only achieves selective security.

This kind of revocation method cannot resist the collusion of revoked users and malicious non-revoked users as malicious non-revoked users can share the update key with those revoked users.

Furthermore, to update the cipher text, the key authority in their scheme needs to maintain a table for each user to produce the re-encryption key for each time period, which significantly increases the key authority's workload.

1.4 Proposed System Algorithms

Our RS-IBE scheme uses the same binary tree structure introduced by Boldyreva, Goyal and Kumar [20] to achieve efficient revocation. To describe the revocation mechanism, we first present several notations. Denote by ε the root node of the binary tree BT , and Path(η) the set of nodes on the path from ε to the leaf node η (including ε and η). For a non-leaf node θ, we let θl and θr stand for its left and

right child, respectively. Given a time period t and revocations list RL, which is comprised of the tuples (ηi,ti) indicating that the node ηi was revoked at time period ti, the algorithm KUNodes(BT ,RL,t) outputs the smallest subset Y of nodes of BT such that Y contains an ancestor for each node that is not revoked before the time period t.

Advantage of Proposed System:

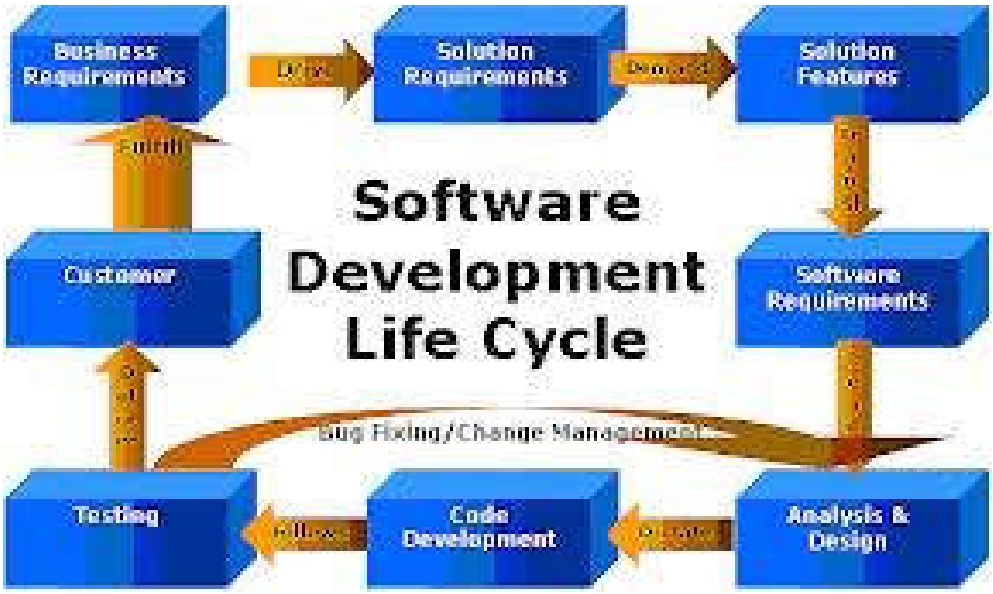
We provide formal definitions for RS-IBE and its corresponding security model;
We present a concrete construction of RS-IBE.
The proposed scheme can provide confidentiality and backward/forward2 secrecy simultaneously
We prove the security of the proposed scheme in the standard model, under the decisional ℓ-Bilinear Diffie-Hellman Exponent (ℓ-BDHE) assumption. In addition, the proposed scheme can withstand decryption key exposure
The procedure of ciphertext update only needs *public information*. Note that no previous identity-based encryption schemes in the literature can provide this feature;
The additional computation and storage complexity, which are brought in by the forward secrecy, is all upper bounded by O(log(T)2), where T is the total number of time periods.

SYSTEM ANALYSIS

2.1 INTRODUCTION

Software Development Life Cycle:-

There is various software development approaches defined and designed which are used/employed during development process of software, these approaches are also referred as "Software Development Process Models". Each process model follows a particular life cycle in order to ensure success in process of software development.



Requirements

Business requirements are gathered in this phase. This phase is the main focus of the project managers and stake holders. Meetings with managers, stake holders and users are held in order to determine the requirements. Who is going to use the system? How will they use the system? What data should be input into the system? What data should be output by the system? These are general questions that get answered during a requirements gathering phase. This produces a nice big list of functionality that the system should provide, which describes functions the system should perform, business logic that processes data, what data is stored and used by the system, and how the user interface should work. The overall result is the system as a whole and how it performs, not how it is actually going to do it.

Design

The software system design is produced from the results of the requirements phase. Architects have the ball in their court during this phase and this is the phase in which their focus lies. This is where the details on how the system will work is produced. Architecture, including hardware and software, communication, software design (UML is produced here) are all part of the deliverables of a design phase.

Implementation

Code is produced from the deliverables of the design phase during implementation, and this is the longest phase of the software development life cycle. For a developer, this is the main focus of the life cycle because this is where the code is produced. Implementation my overlap with both the design and testing phases. Many tools exists (CASE tools) to actually automate the production of code using information gathered and produced during the design phase.

Testing

During testing, the implementation is tested against the requirements to make sure that the product is actually solving the needs addressed and gathered during the requirements phase. Unit tests and system/acceptance tests are done during this phase. Unit tests act on a specific component of the system, while system tests act on the system as a whole.

So in a nutshell, that is a very basic overview of the general software development life cycle model. Now let's delve into some of the traditional and widely used variations.

SDLC METHDOLOGIES

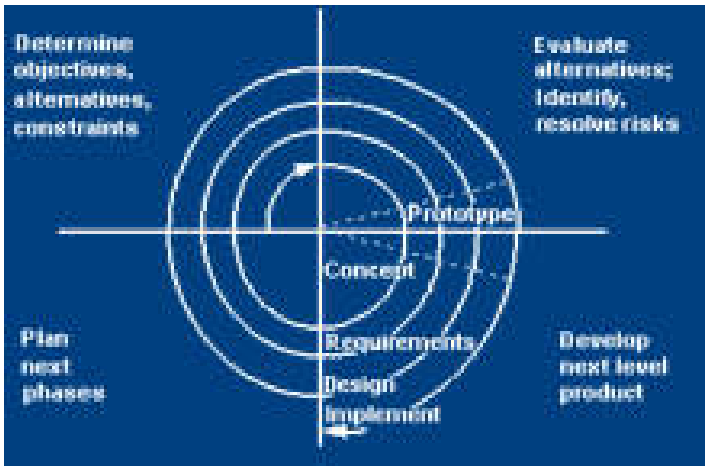
This document play a vital role in the development of life cycle (SDLC) as it describes the complete requirement of the system. It means for use by developers and will be the basic during testing phase. Any changes made to the requirements in the future will have to go through formal change approval process.

SPIRAL MODEL was defined by Barry Boehm in his 1988 article, "A spiral Model of Software Development and Enhancement. This model was not the first

model to discuss iterative development, but it was the first model to explain why the iteration models.

As originally envisioned, the iterations were typically 6 months to 2 years long. Each phase starts with a design goal and ends with a client reviewing the progress thus far. Analysis and engineering efforts are applied at each phase of the project, with an eye toward the end goal of the project.

The following diagram shows how a spiral model acts like:



The steps for Spiral Model can be generalized as follows:

- The new system requirements are defined in as much details as possible. This usually involves interviewing a number of users representing all the external or internal users and other aspects of the existing system.
- A preliminary design is created for the new system.
- A first prototype of the new system is constructed from the preliminary design. This is usually a scaled-down system, and represents an approximation of the characteristics of the final product.
- A second prototype is evolved by a fourfold procedure:

1. Evaluating the first prototype in terms of its strengths, weakness, and risks.
2. Defining the requirements of the second prototype.
3. Planning an designing the second prototype.
4. Constructing and testing the second prototype.

- At the customer option, the entire project can be aborted if the risk is deemed too great. Risk factors might involved development cost overruns, operating-cost miscalculation, or any other factor that could, in the customer's judgment, result in a less-than-satisfactory final product.
- The existing prototype is evaluated in the same manner as was the previous prototype, and if necessary, another prototype is developed from it according to the fourfold procedure outlined above.
- The preceding steps are iterated until the customer is satisfied that the refined prototype represents the final product desired.
- The final system is constructed, based on the refined prototype.
- The final system is thoroughly evaluated and tested. Routine maintenance is carried on a continuing basis to prevent large scale failures and to minimize down time.

2.2 STUDY OF THE SYSTEM

In the flexibility of uses the interface has been developed a graphics concepts in mind, associated through a browser interface. The GUI's at the top level has been categorized as follows

1. Administrative User Interface Design
2. The Operational and Generic User Interface Design

The administrative user interface concentrates on the consistent information that is practically, part of the organizational activities and which needs proper

authentication for the data collection. The Interface helps the administration with all the transactional states like data insertion, data deletion, and data updating along with executive data search capabilities.

The operational and generic user interface helps the users upon the system in transactions through the existing data and required services. The operational user interface also helps the ordinary users in managing their own information helps the ordinary users in managing their own information in a customized manner as per the assisted flexibilities.

Modules Involved

Modules:

1. Data confidentiality

2. Backward secrecy

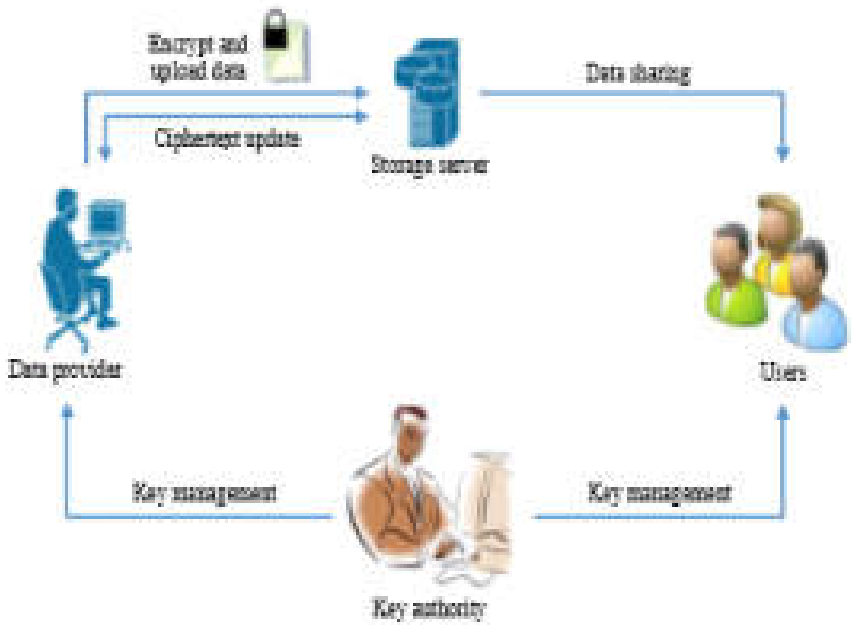
3. Forward secrecy

Data confidentiality: Unauthorized users should be prevented from accessing the plaintext of the shared data stored in the cloud server. In addition, the cloud server, which is supposed to be honest but curious, should also be deterred from knowing plaintext of the shared data.

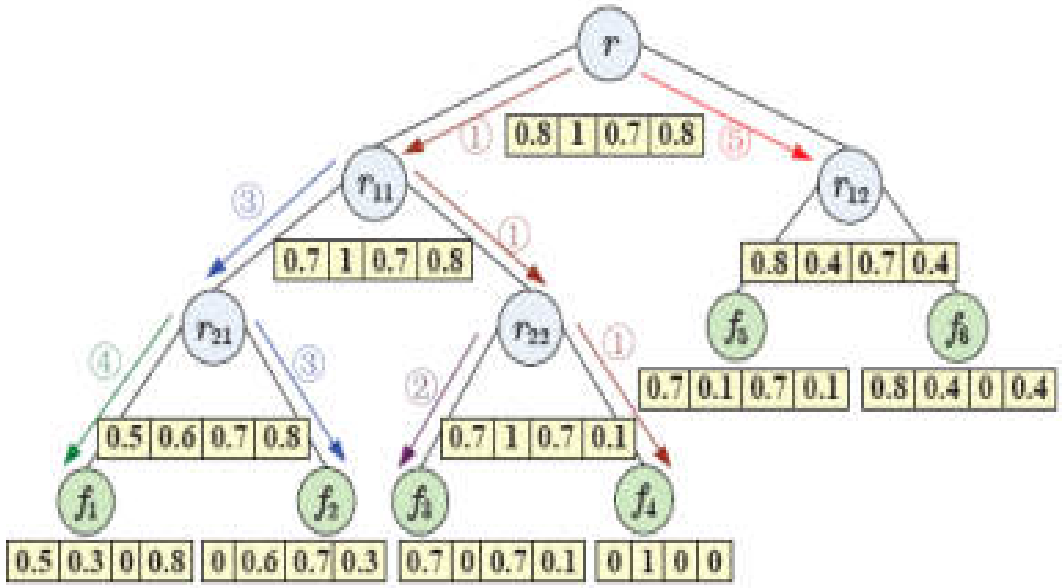
• **Backward secrecy:** Backward secrecy means that, when a user's authorization is expired, or a user's secret key is compromised, he/she should be prevented from accessing the plaintext of the subsequently shared data that are still encrypted under his/her identity.

• **Forward secrecy:** Forward secrecy means that, when a user's authority is expired, or a user's secret key is compromised, he/she should be prevented from accessing the plaintext of the shared data that can be previously accessed.

System Architecture:



Tree-Based Index with the Document Collection



We construct a special keyword balanced binary tree as the index, and propose a "Greedy Depth-first Search" algorithm to obtain better efficiency than linear search.

2.3 HARDWARE&SOFTWARE SPECIFICATIONS

Hardware Requirements:

Processor - intel i3 or above

Speed - 3.5 GHz

RAM - 1GB (min)

Hard Disk - 40 GB

Software Requirements:

- Operating System :Windows 7 or above
- Application Server : IIS 7.0
- Front End : ASP.Net and HTML
- Database : SQL SERVER
- Database Connectivity : ADO.Net. and JAVA

2.4 PROPOSED SYSTEM

It seems that the concept of revocable identity-based encryption (RIBE) might be a promising approach that fulfills the aforementioned security requirements for data sharing.

RIBE features a mechanism that enables a sender to append the current time period to the ciphertext such that the receiver can decrypt the ciphertext only under the condition that he/she is not revoked at that time period.

A RIBE-based data sharing system works as follows:
Step 1: The data provider (e.g., David) first decides the users (e.g., Alice and Bob) who can share the data. Then, David encrypts the data under the identities Alice and Bob, and uploads the ciphertext of the shared data to the cloud server.
Step 2: When either Alice or Bob wants to get the shared data, she or he can download and decrypt the corresponding ciphertext. However, for an unauthorized user and the cloud server, the plaintext of the shared data is not available.
Step 3: In some cases, e.g., Alice's authorization gets expired, David can download the ciphertext of the shared data, and then decrypt-then-re-encrypt the shared data such that Alice is prevented from accessing the plaintext of the shared data, and then upload the re-encrypted data to the cloud server again.

2.5Functional requirements

Data provider can Login in system.

Data provider Upload the data based on encryption format.

Admin allows the user to share the data in data-storing center.

Admin can provide the permissions into Users.

User can access the data from data-storing center.

Admin revoke the permissions also.

Non-Functional Requirements

- Non-functional requirements describe user-visible aspects of the system that are not directly related to functionality of the system.

  User Interface

- A menu interface has been provided to the user to be user friendly.

Performance Constraints

- Requests should be processed within no time.
- Users should be authenticated for accessing the requested data

Error Handling and Extreme Conditions

- In case of User Error, the System should display a meaningful error message to the user, such that the user can correct his Error.
- The high level components in proposed system should handle exceptions that occur while connecting to various database servers, IO Exceptions etc.

Quality Issues

- Quality issues mainly refers to reliable, available and robust system developing the proposed system the developer must be able to guarantee the reliability transactions so that they will be processed completely and accurately.
- The ability of system to detect failures and recovery from those failures refers to the availability of system.

2.6 PROCESS MODEL USED WITH JUSTIFICATION

ACCESS CONTROL FOR DATA WHICH REQUIRE USER AUTHENTICATION

The following commands specify access control identifiers and they are typically used to authorize and authenticate the user (command codes are shown in parentheses)

USER NAME (USER)

The user identification is that which is required by the server for access to its file system. This command will normally be the first command transmitted by the user after the control connections are made (some servers may require this).

PASSWORD (PASS)

This command must be immediately preceded by the user name command, and, for some sites, completes the user's identification for access control. Since password information is quite sensitive, it is desirable in general to "mask" it or suppress type out.

SOFTWARE REQUIREMENTS SPECIFICATION

INTRODUCTION:

Purpose: The main purpose for preparing this document is to give a general insight into the analysis and requirements of the existing system or situation and for determining the operating characteristics of the system.

Scope: This Document plays a vital role in the development life cycle (SDLC) and it describes the complete requirement of the system. It is meant for use by the developers and will be the basic during testing phase. Any changes made to the requirements in the future will have to go through formal change approval process.

DEVELOPERS RESPONSIBILITIES OVERVIEW:

The developer is responsible for:

- Developing the system, which meets the SRS and solving all the requirements of the system?
- Demonstrating the system and installing the system at client's location after the acceptance testing is successful.
- Submitting the required user manual describing the system interfaces to work on it and also the documents of the system.
- Conducting any user training that might be needed for using the system.
- Maintaining the system for a period of one year after installation.

4.1. FUNCTIONAL REQUIREMENTS

OUTPUT DESIGN: Outputs from computer systems are required primarily to communicate the results of processing to users. They are also used to provides a permanent copy of the results for later consultation. The various types of outputs in general are:

- External Outputs, whose destination is outside the organization

- Internal Outputs whose destination is within organization and they are the
- User's main interface with the computer.
- Operational outputs whose use is purely within the computer department.
- Interface outputs, which involve the user in communicating directly.

OUTPUT DEFINITION

The outputs should be defined in terms of the following points:

- Type of the output
- Content of the output
- Format of the output
- Location of the output
- Frequency of the output
- Volume of the output
- Sequence of the output

It is not always desirable to print or display data as it is held on a computer. It should be decided as which form of the output is the most suitable.

INPUT DESIGN

Input design is a part of overall system design. The main objective during the input design is as given below:

- To produce a cost-effective method of input.
- To achieve the highest possible level of accuracy.
- To ensure that the input is acceptable and understood by the user.

INPUT STAGES:

The main input stages can be listed as below:

- Data recording
- Data transcription
- Data conversion
- Data verification
- Data control
- Data transmission
- Data validation
- Data correction

INPUT TYPES:

It is necessary to determine the various types of inputs. Inputs can be categorized as follows:

- External inputs, which are prime inputs for the system.
- Internal inputs, which are user communications with the system.
- Operational, which are computer department's communications to the system?
- Interactive, which are inputs entered during a dialogue.

INPUT MEDIA:

At this stage choice has to be made about the input media. To conclude about the input media consideration has to be given to;

- Type of input
- Flexibility of format
- Speed

- Accuracy
- Verification methods
- Rejection rates
- Ease of correction
- Storage and handling requirements
- Security
- Easy to use
- Portability

Keeping in view the above description of the input types and input media, it can be said that most of the inputs are of the form of internal and interactive. As

Input data is to be the directly keyed in by the user, the keyboard can be considered to be the most suitable input device.

ERROR AVOIDANCE

At this stage care is to be taken to ensure that input data remains accurate form the stage at which it is recorded up to the stage in which the data is accepted by the system. This can be achieved only by means of careful control each time the data is handled.

ERROR DETECTION

Even though every effort is make to avoid the occurrence of errors, still a small proportion of errors is always likely to occur, these types of errors can be discovered by using validations to check the input data.

DATA VALIDATION: Procedures are designed to detect errors in data at a lower level of detail. Data validations have been included in the system in almost every area where there is a possibility for the user to commit errors. The system will not accept invalid data. Whenever an invalid data is keyed in, the system immediately prompts the user and the user has to again key in the data and the system will accept the data only if the data is correct. Validations have been included where necessary.

The system is designed to be a user friendly one. In other words the system has been designed to communicate effectively with the user. The system has been designed with popup menus.

USER INTERFACE DESIGN: It is essential to consult the system users and discuss their needs while designing the user interface:

USER INTERFACE SYSTEMS CAN BE BROADLY CLASIFIED AS:

1. User initiated interface the user is in charge, controlling the progress of the user/computer dialogue. In the computer-initiated interface, the computer selects the next stage in the interaction.
2. Computer initiated interfaces

In the computer initiated interfaces the computer guides the progress of the user/computer dialogue. Information is displayed and the user response of the computer takes action or displays further information.

USER_INITIATED INTERGFACES: User initiated interfaces fall into tow approximate classes:

1. Command driven interfaces: In this type of interface the user inputs commands or queries which are interpreted by the computer.
2. Forms oriented interface: The user calls up an image of the form to his/her screen and fills in the form. The forms oriented interface is chosen because it is the best choice.

COMPUTER-INITIATED INTERFACES: The following computer – initiated interfaces were used:

1. The menu system for the user is presented with a list of alternatives and the user chooses one; of alternatives.
2. Questions – answer type dialog system where the computer asks question and takes action based on the basis of the users reply.

Right from the start the system is going to be menu driven, the opening menu displays the available options. Choosing one option gives another popup menu with more options. In this way every option leads the users to data entry form where the user can key in the data.

ERROR MESSAGE DESIGN: The design of error messages is an important part of the user interface design. As user is bound to commit some errors or other while designing a system the system should be designed to be helpful by providing the user with information regarding the error he/she has committed.

This application must be able to produce output at different modules for different inputs.

PERFORMANCE REQUIREMENTS:

Performance is measured in terms of the output provided by the application.

Requirement specification plays an important part in the analysis of a system. Only when the requirement specifications are properly given, it is possible to design a system, which will fit into required environment. It rests largely in the part of the users of the existing system to give the requirement specifications because they are the people who finally use the system. This is because the requirements have to be known during the initial stages so that the system can be designed according to those requirements. It is very difficult to change the system once it has been designed and on the other hand designing a system, which does not cater to the requirements of the user, is of no use.

The requirement specification for any system can be broadly stated as given below:

- The system should be able to interface with the existing system
- The system should be accurate
- The system should be better than the existing system

The existing system is completely dependent on the user to perform all the duties.

5. LITERATURE SURVEY

5.1 INTRODUCTION

Firstly, outsourcing data to cloud server implies that data is out control of users. This may cause users' hesitation since the outsourced data usually contain valuable and sensitive information. Secondly, data sharing is often implemented in an open and hostile environment, and cloud server would become a target of attacks. Even worse, cloud server itself may reveal users' data for illegal profit. Thirdly, data sharing is not static. That is, when a user's authorization gets expired, he/she should no longer possess the privilege of accessing the previously and subsequently shared data. Therefore, while outsourcing data to cloud server, users also want to control access to these data such that only those currently authorized users can share the outsourced data.

5.2 HISTORY

The concept of identity-based encryption was introduced by Shamir and conveniently instantiated by Boneh and Franklin [14]. IBE eliminates the need for providing a public key infrastructure (PKI). Regard less of the setting of IBE or PKI, there must be an approach to revoke users from the system when necessary, e.g., the authority of some user is expired or the secret key of some user is disclosed. In the traditional PKI setting, the problem of revocation has been well studied and several techniques are widely approved, such as certificate revocation list or appending validity periods to certificates. However, there are only a few studies on revocation in the setting of IBE.

5.3 PURPOSE

Firstly, outsourcing data to cloud server implies that data is out control of users. This may cause users' hesitation since the outsourced data usually contain valuable and sensitive information.

Data confidentiality: Unauthorized users should be prevented from accessing the plaintext of the shared data stored in the cloud server. In addition, the cloud server, which is supposed to be honest but curious, should also be deterred from knowing plaintext of the shared data.

• **Backward secrecy:** Backward secrecy means that, when a user's authorization is expired, or a user's secret key is compromised, he/she should be prevented from

accessing the plaintext of the subsequently shared data that are still encrypted under his/her identity.

• **Forward secrecy:** Forward secrecy means that, when a user's authority is expired, or a user's secret key is compromised, he/she should be prevented from accessing the plaintext of the shared data that can be previously accessed.

5.1 INTRODUCTION TO .NET FRAMEWORK

The **Microsoft .NET Framework** is a software technology that is available with several Microsoft Windows operating systems. It includes a large library of pre-coded solutions to common programming problems and a virtual machine that manages the execution of programs written specifically for the framework. The .NET Framework is a key Microsoft offering and is intended to be used by most new applications created for the Windows platform.

The pre-coded solutions that form the framework's Base Class Library cover a large range of programming needs in a number of areas, including user interface, data access, database connectivity, cryptography, web application development, numeric algorithms, and network communications. The class library is used by programmers, who combine it with their own code to produce applications.

Programs written for the .NET Framework execute in a software environment that manages the program's runtime requirements. Also part of the .NET Framework, this runtime environment is known as the Common Language Runtime (CLR). The CLR provides the appearance of an application virtual machine so that programmers need not consider the capabilities of the specific CPU that will execute the program. The CLR also provides other important services such as security, memory management, and exception handling. The class library and the CLR together compose the .NET Framework.

Principal design features Interoperability

Because interaction between new and older applications is commonly required, the .NET Framework provides means to access functionality that is implemented in programs that execute outside the .NET environment. Access to COM components is provided in the System.Runtime.InteropServices and System.EnterpriseServices namespaces of the framework; access to other functionality is provided using the P/Invoke feature.

Common Runtime Engine

The Common Language Runtime (CLR) is the virtual machine component of the .NET framework. All .NET programs execute under the supervision of the CLR, guaranteeing certain properties and behaviors in the areas of memory management, security, and exception handling.

Base Class Library

The Base Class Library (BCL), part of the Framework Class Library (FCL), is a library of functionality available to all languages using the .NET Framework. The BCL provides classes which encapsulate a number of common functions, including file reading and writing, graphic rendering, database interaction and XML document manipulation.

Simplified Deployment

Installation of computer software must be carefully managed to ensure that it does not interfere with previously installed software, and that it conforms to security requirements. The .NET framework includes design features and tools that help address these requirements.

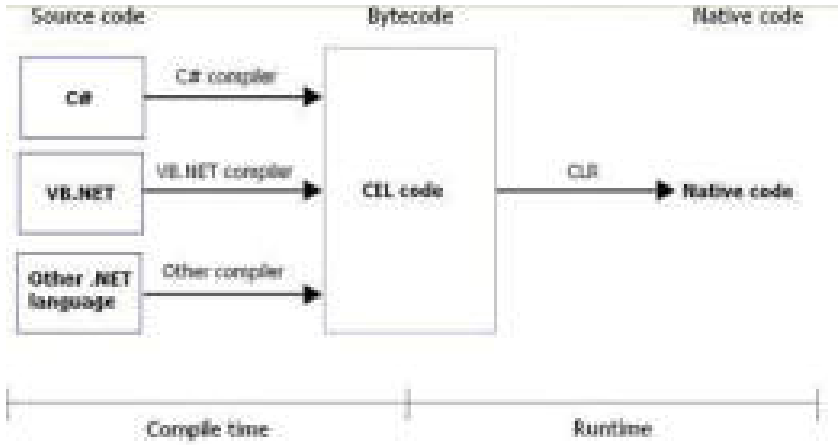
Security

The design is meant to address some of the vulnerabilities, such as buffer overflows, that have been exploited by malicious software. Additionally, .NET provides a common security model for all applications.

Portability

The design of the .NET Framework allows it to theoretically be platform agnostic, and thus cross-platform compatible. That is, a program written to use the framework should run without change on any type of system for which the framework is implemented. Microsoft's commercial implementations of the framework cover Windows, Windows CE, and the Xbox 360. In addition, Microsoft submits the specifications for the Common Language Infrastructure (which includes the core class libraries, Common Type System, and the Common Intermediate Language), the C# language, and the C++/CLI language to both ECMA and the ISO, making them,

Architecture



Visual overview of the Common Language Infrastructure (CLI)

Common Language Infrastructure

The core aspects of the **.NET framework** lie within the Common Language Infrastructure, or **CLI**. The purpose of the CLI is to provide a language-neutral platform for application development and execution, including functions for exception handling, garbage collection, security, and interoperability. Microsoft's implementation of the CLI is called the **Common Language Runtime** or **CLR**.

Assemblies

The intermediate CIL code is housed in .NET assemblies. As mandated by specification, assemblies are stored in the Portable Executable (PE) format, common on the Windows platform for all DLL and EXE files. The assembly consists of one or more files, one of which must contain the manifest, which has the metadata for the assembly. The complete name of an assembly (not to be confused with the filename on disk) contains its simple text name, version number, culture, and public key token. The public key token is a unique hash generated when the assembly is compiled, thus two assemblies with the same public key token are guaranteed to be identical from the point of view of the framework. A private key can also be specified known only to the creator of the assembly and can be used for strong naming and to guarantee that the assembly is from the same author when a new version of the assembly is compiled (required to add an assembly to the Global Assembly Cache).

Metadata

All CLI is self-describing through .NET metadata. The CLR checks the metadata to ensure that the correct method is called. Metadata is usually generated by language compilers but developers can create their own metadata through custom attributes. Metadata contains information about the assembly, and is also used to implement the reflective programming capabilities of .NET Framework.

Security

.NET has its own security mechanism with two general features: Code Access Security (CAS), and validation and verification. Code Access Security is based on evidence that is associated with a specific assembly. Typically the evidence is the source of the assembly (whether it is installed on the local machine or has been downloaded from the intranet or Internet). Code Access Security uses evidence to determine the permissions granted to the code. Other code can demand that calling code is granted a specified permission. The demand causes the CLR to perform a call stack walk: every assembly of each method in the call stack is checked for the required permission; if any assembly is not granted the permission a security exception is thrown.

When an assembly is loaded the CLR performs various tests. Two such tests are validation and verification. During validation the CLR checks that the assembly contains valid metadata and CIL, and whether the internal tables are correct. Verification is not so exact. The verification mechanism checks to see if the code does anything that is 'unsafe'. The algorithm used is quite conservative; hence occasionally code that is 'safe' does not pass. Unsafe code will only be executed if the assembly has the 'skip verification' permission, which generally means code that is installed on the local machine.

.NET Framework uses appdomains as a mechanism for isolating code running in a process. Appdomains can be created and code loaded into or unloaded from them independent of other appdomains. This helps increase the fault tolerance of the application, as faults or crashes in one appdomain do not affect rest of the application. Appdomains can also be configured independently with different security privileges. This can help increase the security of the application by isolating potentially unsafe code. The developer, however, has to split the application into sub domains; it is not done by the CLR.

Class library

Namespaces in the BCL
System
System. CodeDom
System. Collections
System. Diagnostics
System. Globalization
System. IO
System. Resources
System. Text
System.Text.RegularExpressions

Microsoft **.NET Framework** includes a set of standard **class libraries**. The class library is organized in a hierarchy of namespaces. Most of the built in APIs are part of either System.* or Microsoft.* namespaces. It encapsulates a large number of common functions, such as file reading and writing, graphic rendering, database interaction, and XML document manipulation, among others. The .NET class libraries are available to all .NET languages. The .NET Framework class library is divided into two parts: the **Base Class Library** and the **Framework Class Library**.

The **Base Class Library** (BCL) includes a small subset of the entire class library and is the core set of classes that serve as the basic API of the Common Language Runtime. The classes in mscorlib.dll and some of the classes in System.dll and System.core.dll are considered to be a part of the BCL. The BCL classes are available in both .NET Framework as well as its alternative implementations including .NET Compact Framework, Microsoft Silverlight and Mono.

The **Framework Class Library** (FCL) is a superset of the BCL classes and refers to the entire class library that ships with .NET Framework. It includes an expanded set of libraries, including WinForms, ADO.NET, ASP.NET, Language Integrated Query, Windows Presentation Foundation, Windows Communication Foundation among others. The FCL is much larger in scope than standard libraries for languages like C++, and comparable in scope to the standard libraries of Java.

Memory management

The .NET Framework CLR frees the developer from the burden of managing memory (allocating and freeing up when done); instead it does the memory management itself. To this end, the memory allocated to instantiations of .NET types (objects) is done contiguously from the managed heap, a pool of memory managed by the CLR. As long as there exists a reference to an object, which might be either a direct reference to an object or via a graph of objects, the object is considered to be in use by the CLR. When there is no reference to an object, and it cannot be reached or used, it becomes garbage. However, it still holds on to the memory allocated to it. .NET Framework includes a garbage collector which runs periodically, on a separate thread from the application's thread, that enumerates all the unusable objects and reclaims the memory allocated to them.

The .NET Garbage Collector (GC) is a non-deterministic, compacting, mark-and-sweep garbage collector. The GC runs only when a certain amount of memory has been used or there is enough pressure for memory on the system. Since it is not guaranteed when the conditions to reclaim memory are reached, the GC runs are non-deterministic. Each .NET application has a set of roots, which are pointers to objects on the managed heap (*managed objects*). These include references to static objects and objects defined as local variables or method parameters currently in scope, as well as objects referred to by CPU registers. When the GC runs, it pauses the application, and for each object referred to in the root, it recursively enumerates all the objects reachable from the root objects and marks them as reachable. It uses .NET metadata and reflection to discover the objects encapsulated by an object, and then recursively walk them. It then enumerates all the objects on the heap (which were initially allocated contiguously) using reflection. All objects not marked as reachable are garbage. This is the *mark* phase. Since the memory held by garbage is not of any consequence, it is considered free space. However, this leaves chunks of free space between objects which were initially contiguous. The objects are then *compacted* together, by using memcpy to copy them over to the free space to make them contiguous again. Any reference to an object invalidated by moving the object is updated to reflect the new location by the GC. The application is resumed after the garbage collection is over.

The GC used by .NET Framework is actually *generational*. Objects are assigned a *generation*; newly created objects belong to *Generation 0*. The objects that survive a garbage collection are tagged as *Generation 1*, and the Generation 1 objects that survive another collection are *Generation 2* objects. The .NET Framework uses up to Generation 2 objects. Higher generation objects are garbage collected less

frequently than lower generation objects. This helps increase the efficiency of garbage collection, as older objects tend to have a larger lifetime than newer objects. Thus, by removing older (and thus more likely to survive a collection) objects from the scope of a collection run, fewer objects need to be checked and compacted.

Versions

Microsoft started development on the .NET Framework in the late 1990s originally under the name of Next Generation Windows Services (NGWS). By late 2000 the first beta versions of .NET 1.0 were released.

The **.NET Framework** stack.

Overview of .NET Framework release history

Generation	Version number	Release date	Development tool	Distributed with
1.0	1.0.3705.0	2002-02-13	Visual Studio .NET	N/A
1.1	1.1.4322.573	2003-04-24	Visual Studio .NET 2003	Windows Server 2003
2.0	2.0.50727.42	2005-11-07	Visual Studio 2005	Windows Server 2003 R2
3.0	3.0.4506.30	2006-11-06	Expression Blend	Windows Vista, Windows Server 2008
3.5	3.5.21022.8	2007-11-19	Visual Studio 2008	Windows 7, Windows Server 2008 R2
4.0	4.0.30319.1	2010-04-12	Visual Studio 2010	N/A
4.5	4.5.50709.17929	2012-08-15	Visual Studio 2012	Windows 8, Windows Server 2012

Client Application Development

Client applications are the closest to a traditional style of application in Windows-based programming. These are the types of applications that display windows or forms on the desktop, enabling a user to perform a task. Client applications include applications such as word processors and spreadsheets, as well as custom business applications such as data-entry tools, reporting tools, and so on. Client applications usually employ windows, menus, buttons, and other GUI elements, and they likely access local resources such as the file system and peripherals such as printers. Another kind of client application is the traditional ActiveX control (now replaced by the managed Windows Forms control) deployed over the Internet as a Web page. This application is much like other client applications: it is executed natively, has access to local resources, and includes graphical elements.

In the past, developers created such applications using C/C++ in conjunction with the Microsoft Foundation Classes (MFC) or with a rapid application development (RAD) environment such as Microsoft® Visual Basic®. The .NET Framework incorporates aspects of these existing products into a single, consistent development environment that drastically simplifies the development of client applications.

The Windows Forms classes contained in the .NET Framework are designed to be used for GUI development. You can easily create command windows, buttons, menus, toolbars, and other screen elements with the flexibility necessary to accommodate shifting business needs.

For example, the .NET Framework provides simple properties to adjust visual attributes associated with forms. In some cases the underlying operating system does not support changing these attributes directly, and in these cases the .NET Framework automatically recreates the forms. This is one of many ways in which the .NET Framework integrates the developer interface, making coding simpler and more consistent.

Server Application Development

Server-side applications in the managed world are implemented through runtime hosts. Unmanaged applications host the common language runtime, which allows your custom managed code to control the behavior of the server.

This model provides you with all the features of the common language runtime and class library while gaining the performance and scalability of the host server.

Server-side managed code

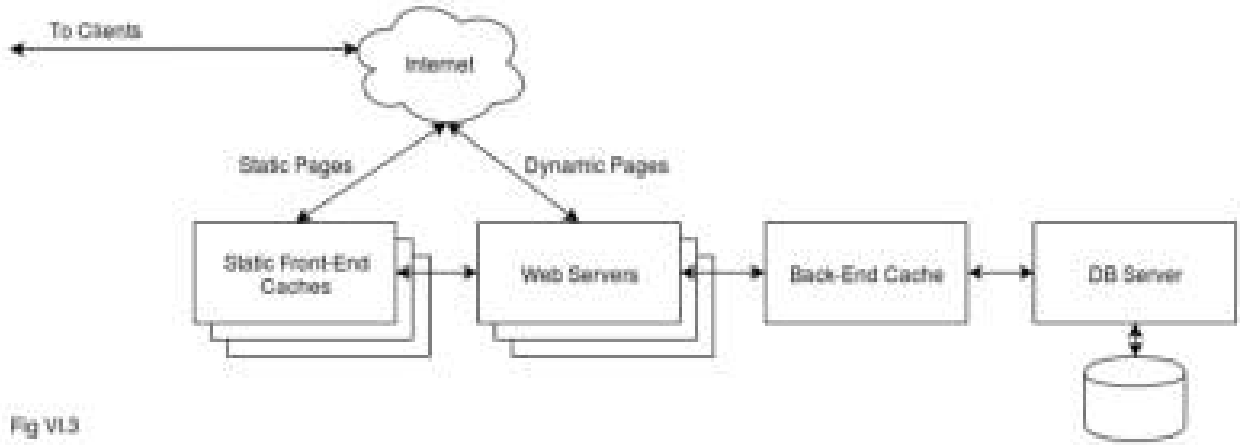


Fig VI.3

ASP.NET is the hosting environment that enables developers to use the .NET Framework to target Web-based applications. However, ASP.NET is more than just a runtime host; it is a complete architecture for developing Web sites and Internet-distributed objects using managed code. Both Web Forms and XML Web services use IIS and ASP.NET as the publishing mechanism for applications, and both have a collection of supporting classes in the .NET

5.2 C#.NET

The Relationship of C# to .NET

C# is a new programming language, and is significant in two respects:

- It is specifically designed and targeted for use with Microsoft's .NET Framework (a feature rich platform for the development, deployment, and execution of distributed applications).

- It is a language based upon the modern object-oriented design methodology, and when designing it Microsoft has been able to learn from the experience of all the other similar languages that have been around over the 20 years or so since object-oriented principles came to prominence

One important thing to make clear is that C# is a language in its own right. Although it is designed to generate code that targets the .NET environment, it is not itself part of .NET. There are some features that are supported by .NET but not by C#, and you might be surprised to learn that there are actually features of the C# language that are not supported by .NET like Operator Overloading.

However, since the C# language is intended for use with .NET, it is important for us to have an understanding of this Framework if we wish to develop applications in C# effectively. So, in this chapter

The Common Language Runtime:

Central to the .NET framework is its run-time execution environment, known as the **Common Language Runtime** (**CLR**) or the **.NET runtime**. Code running under the control of the CLR is often termed **managed** code.

However, before it can be executed by the CLR, any source code that we develop (in C# or some other language) needs to be compiled. Compilation occurs in two steps in .NET:

1. Compilation of source code to **Microsoft Intermediate Language** (**MS-IL**)

2. Compilation of IL to platform-specific code by the CLR

At first sight this might seem a rather long-winded compilation process. Actually, this two-stage compilation process is very important, because the existence of the Microsoft Intermediate Language (managed code) is the key to providing many of the benefits of .NET. Let's see why.

Advantages of Managed Code

Microsoft Intermediate Language (often shortened to "Intermediate Language", or "IL") shares with Java byte code the idea that it is a low-level language with a simple syntax (based on numeric codes rather than text), which can be very quickly translated into native machine code. Having this well-defined Universal syntax for code has significant advantages.

Platform Independence

First, it means that the same file containing byte code instructions can be placed on any platform; at runtime the final stage of compilation can then be easily accomplished so that the code will run on that particular platform. In other words, by compiling to Intermediate Language we obtain platform independence for .NET, in much the same way as compiling to Java byte code gives Java platform independence.

You should note that the platform independence of .NET is only theoretical at present because, at the time of writing, .NET is only available for Windows. However, porting .NET to other platforms is being explored (see for example the Mono project, an effort to create an open source implementation of .NET, at http://www.go-mono.com/).

Performance Improvement

Although we previously made comparisons with Java, IL is actually a bit more ambitious than Java byte code. Significantly, IL is always **Just-In-Time** compiled, whereas Java byte code was often interpreted. One of the disadvantages of Java was that, on execution, the process of translating from Java byte code to native executable resulted in a loss of performance (apart from in more recent cases, here Java is JIT-compiled on certain platforms).

Instead of compiling the entire application in one go (which could lead to a slow start-up time), the JIT compiler simply compiles each portion of code as it is called (just-in-time). When code has been compiled once, the resultant native executable is stored until the application exits, so that it does not need to be recompiled the next time that portion of code is run. Microsoft argues that this process is more efficient than compiling the entire application code at the start, because of the likelihood those large portions of any application code will not actually be executed in any given run. Using the JIT compiler, such code will never get compiled.

This explains why we can expect that execution of managed IL code will be almost as fast as executing native machine code. What it doesn't explain is why Microsoft expects that we will get a performance *improvement*. The reason given for this is that, since the final stage of compilation takes place at run time, the JIT compiler will know exactly what processor type the program will run on. This means that it can optimize the final executable code to take advantage of any features or particular machine code instructions offered by that particular processor.

Traditional compilers will optimize the code, but they can only perform optimizations that will be independent of the particular processor that the code will run on. This is because traditional compilers compile to native executable before the software is shipped. This means that the compiler doesn't know what type of processor the code will run on beyond basic generalities, such as that it will be an

x86-compatible processor or an Alpha processor. Visual Studio 6, for example, optimizes for a generic Pentium machine, so the code that it generates cannot take advantages of hardware features of Pentium III processors. On the other hand, the JIT compiler can do all the optimizations that Visual Studio 6 can, and in addition to that it will optimize for the particular processor the code is running on.

Language Interoperability

How the use of IL enables platform independence, and how JIT compilation should improve performance. However, IL also facilitates **language interoperability**. Simply put, you can compile to IL from one language, and this compiled code should then be interoperable with code that has been compiled to IL from another language.

Intermediate Language

From what we learned in the previous section, Intermediate Language obviously plays a fundamental role in the .NET Framework. As C# developers, we now understand that our C# code will be compiled into Intermediate Language before it is executed (indeed, the C# compiler *only* compiles to managed code). It makes sense, then, that we should now take a closer look at the main characteristics of IL, since any language that targets .NET would logically need to support the main characteristics of IL too.

Here are the important features of the Intermediate Language:

- Object-orientation and use of interfaces
- Strong distinction between value and reference types
- Strong data typing
- Error handling through the use of exceptions
- Use of attributes

Support of Object Orientation and Interfaces

The language independence of .NET does have some practical limits. In particular, IL, however it is designed, is inevitably going to implement some particular programming methodology, which means that languages targeting it are going to have to be compatible with that methodology. The particular route that

Microsoft has chosen to follow for IL is that of classic object-oriented programming, with single implementation inheritance of classes.

Besides classic object-oriented programming, Intermediate Language also brings in the idea of interfaces, which saw their first implementation under Windows with COM. .NET interfaces are not the same as COM interfaces; they do not need to support any of the COM infrastructure (for example, they are not derived from I Unknown, and they do not have associated GUIDs). However, they do share with
COM interfaces the idea that they provide a contract, and classes that implement a given interface must provide implementations of the methods and properties specified by that interface.

Object Orientation and Language Interoperability

Working with .NET means compiling to the Intermediate Language, and that in turn means that you will need to be programming using traditional object-oriented methodologies. That alone is not, however, sufficient to give us language interoperability. After all, C++ and Java both use the same object-oriented paradigms, but they are still not regarded as interoperable. We need to look a little more closely at the concept of language interoperability.

An associated problem was that, when debugging, you would still have to independently debug components written in different languages. It was not possible to step between languages in the debugger. So what we *really* mean by language interoperability is that classes written in one language should be able to talk directly to classes written in another language. In particular:

- A class written in one language can inherit from a class written in another language
- The class can contain an instance of another class, no matter what the languages of the two classes are
- An object can directly call methods against another object written in another language
- Objects (or references to objects) can be passed around between methods
- When calling methods between languages we can step between the method calls in the

Debugger, even where this means stepping between source code written in different languages
This is all quite an ambitious aim, but amazingly, .NET and the Intermediate Language have achieved it. For the case of stepping between methods in the debugger, this facility is really offered by the Visual Studio .NET IDE rather than from the CLR itself.

Strong Data Type

One very important aspect of IL is that it is based on exceptionally **strong data typing**. What we mean by that is that all variables are clearly marked as being of a particular, specific data type (there is no room in IL, for example, for the Variant data type recognized by Visual Basic and scripting languages). In particular, IL does not normally permit any operations that result in ambiguous data types.

For instance, VB developers will be used to being able to pass variables around without worrying too much about their types, because VB automatically performs type conversion. C++ developers will be used to routinely casting pointers between different types. Being able to perform this kind of operation can be great for performance, but it breaks type safety. Hence, it is permitted only in very specific circumstances in some of the languages that compile to managed code. Indeed, pointers (as opposed to references) are only permitted in marked blocks of code in C#, and not at all in VB (although they are allowed as normal in managed C++). Using pointers in your code will immediately cause it to fail the memory type safety checks performed by the CLR.

You should note that some languages compatible with .NET, such as VB.NET, still allow some laxity in typing, but that is only possible because the compilers behind the scenes ensure the type safety is enforced in the emitted IL.

Although enforcing type safety might initially appear to hurt performance, in many cases this is far outweighed by the benefits gained from the services provided by .NET that rely on type safety. Such services include:

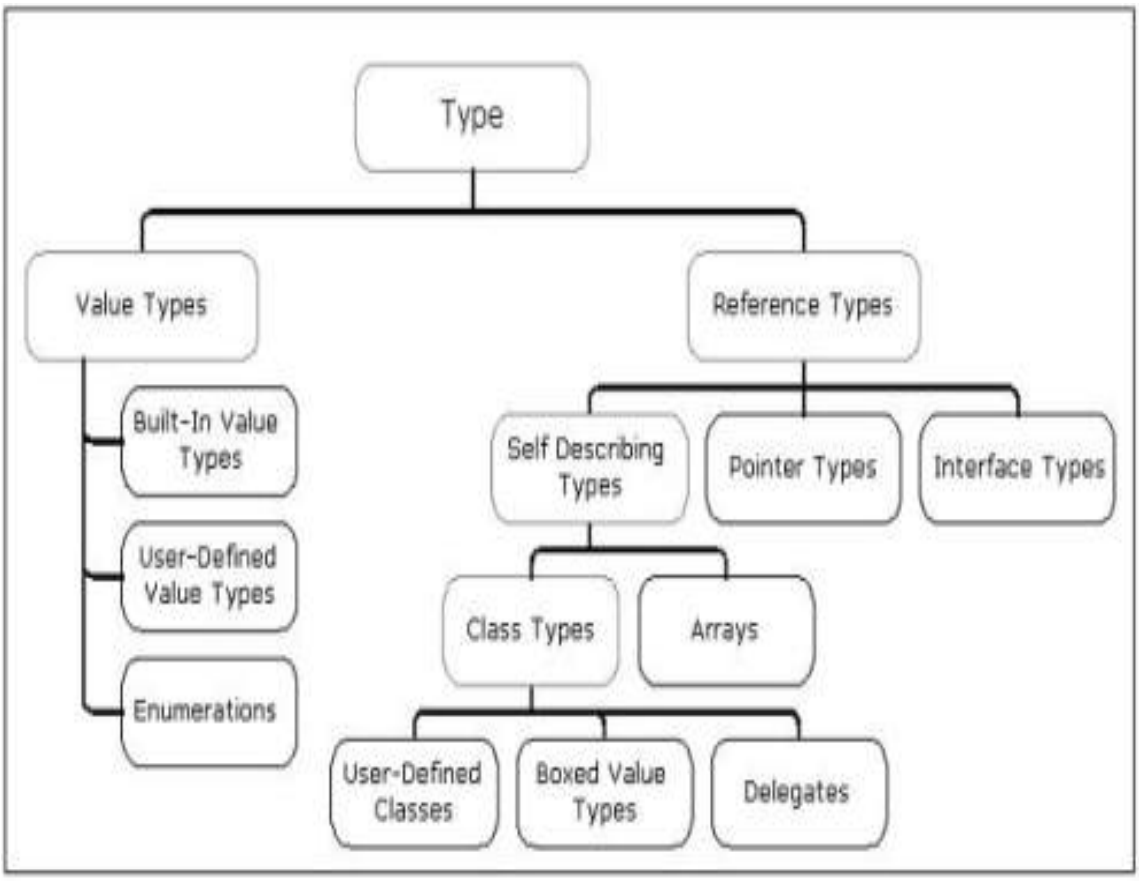
- Language Interoperability
- Garbage Collection

- Security
- Application Domains

Common Type System (CTS)

This data type problem is solved in .NET through the use of the **Common Type System** (**CTS**). The CTS defines the predefined data types that are available in IL, so that all languages that target the .NET framework will produce compiled code that is ultimately based on these types.

The CTS doesn't merely specify primitive data types, but a rich hierarchy of types, which includes well-defined points in the hierarchy at which code is permitted to define its own types. The hierarchical structure of the Common Type System reflects the single-inheritance object-oriented methodology of IL, and looks like this:



Common Language Specification (CLS)

The Common Language Specification works with the Common Type System to ensure language interoperability. The CLS is a set of minimum standards that all compilers targeting .NET must support. Since IL is a very rich language, writers of most compilers will prefer to restrict the capabilities of a given compiler to only support a subset of the facilities offered by IL and the CTS. That is fine, as long as the compiler supports everything that is defined in the CLS.

> It is perfectly acceptable to write non-CLS-compliant code. However, if you do, the compiled IL code isn't guaranteed to be fully language-interoperable.

Garbage Collection

The **garbage collector** is .NET's answer to memory management, and in particular to the question of what to do about reclaiming memory that running applications ask for. Up until now there have been two techniques used on Windows platform for deal locating memory that processes have dynamically requested from the system:

- Make the application code do it all manually
- Make objects maintain reference counts

The .NET runtime relies on the garbage collector instead. This is a program whose purpose is to clean up memory. The idea is that all dynamically requested memory is allocated on the heap (that is true for all languages, although in the case of .NET, the CLR maintains its own managed heap for .NET applications to use). Every so often, when .NET detects that the managed heap for a given process is becoming full and therefore needs tidying up, it calls the garbage collector. The garbage collector runs through variables currently in scope in your code, examining references to objects stored on the heap to identify which ones are accessible from your code – that is to say which objects have references that refer to them. Any objects that are not referred to are deemed to be no longer accessible from your code and can therefore be removed. Java uses a similar system of garbage collection to this.

Security

.NET can really excel in terms of complementing the security mechanisms provided by Windows because it can offer code-based security, whereas Windows only really offers role-based security.

Role-based security is based on the identity of the account under which the process is running, in other words, who owns and is running the process. Code-based security on the other hand is based on what the code actually does and on how much the code is trusted. Thanks to the strong type safety of IL, the CLR is able to inspect code before running it in order to determine required security permissions. .NET also offers a mechanism by which code can indicate in advance what security permissions it will require to run.

.Net Framework Classes

The .NET base classes are a massive collection of managed code classes that have been written by Microsoft, and which allow you to do almost any of the tasks that were previously available through the Windows API. These classes follow the same object model as used by IL, based on single inheritance. This means that you can either instantiate objects of whichever .NET base class is appropriate, or you can derive your own classes from them.

The great thing about the .NET base classes is that they have been designed to be very intuitive and easy to use. For example, to start a thread, you call the Start () method of the Thread class. To disable a Textbox, you set the Enabled property of a Textbox object to false. This approach will be familiar to Visual Basic and Java developers, whose respective libraries are just as easy to use. It may however come as a great relief to C++ developers, who for years have had to cope with such API functions as GetDIBits (), RegisterWndClassEx (), and IsEqualIID

Name Spaces

Namespaces are the way that .NET avoids name clashes between classes. They are designed, for example, to avoid the situation in which you define a class to represent a customer, name your class Customer, and then someone else does the same thing (quite a likely scenario – the proportion of businesses that have customers seems to be quite high).

A namespace is no more than a grouping of data types, but it has the effect that the names of all data types within a namespace automatically get prefixed with the name of the namespace. It is also possible to nest namespaces within each other. For example, most of the general-purpose .NET base classes are in a namespace called System. The base class Array is in this namespace, so its full name is System. Array.

If a namespace is explicitly supplied, then the type will be added to a nameless global namespace.

Creating .Net Application using C#

C# can be used to create console applications: text-only applications that run in a DOS window. You'll probably use console applications when unit testing class libraries, and for creating Unix/Linux daemon processes. However, more often you'll use C# to create applications that use many of the technologies associated with .NET. In this section, we'll give you an overview of the different types of application that you can write in C#.

Creating Windows Forms

Although C# and .NET are particularly suited to web development, they still offer splendid support for so-called "fat client" apps, applications that have to be installed on the end-user's machine where most of the processing takes place. This support is from **Windows Forms**.

In fact, the .NET SDK provides a utility that creates a wrapper for ActiveX controls, so that they can be placed on Windows Forms. As is the case with Web Controls, Windows Control creation involves deriving from a particular class, System.Windows.Forms.Control.

Windows Services

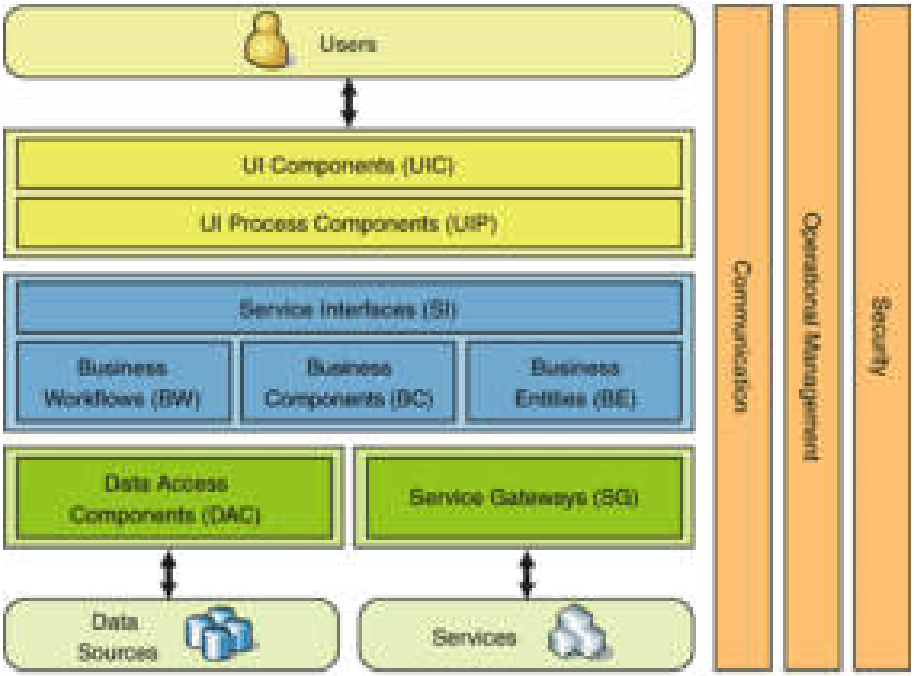
A Windows Service (originally called an NT Service) is a program that is designed to run in the background in Windows NT/2000/XP (but not Windows 9x). Services are useful where you want a program to be running continuously and ready to respond to events without having been explicitly started by the user. A good example would be the World Wide Web Service on web servers, which Listens out for web requests from clients.

It is very easy to write services in C#. There are .NET Framework base classes available in the System.ServiceProcess namespace that handle many of the

boilerplate tasks associated with services, and in addition, Visual Studio .NET allows you to create a C# Windows Service project, which starts you out with the Framework C# source code for a basic Windows service.

The Role of C# In .Net Enterprise Architecture

C# requires the presence of the .NET runtime, and it will probably be a few years before most clients – particularly most home machines – have .NET installed. In the meantime, installing a C# application is likely to mean also installing the .NET redistributable components. Because of that, it is likely that the first place we will see many C# applications is in the enterprise environment. Indeed, C# arguably presents an outstanding opportunity for organizations that are interested in building robust, n-tiered client-server applications.



SYSTEM DESIGN

6.1. INTRODUCTION

Software design sits at the technical kernel of the software engineering process and is applied regardless of the development paradigm and area of application. Design is the first step in the development phase for any engineered product or system. The designer's goal is to produce a model or representation of an entity that will later be built. Beginning, once system requirement have been specified and analyzed, system design is the first of the three technical activities -design, code and test that is required to build and verify software.

The importance can be stated with a single word "Quality". Design is the place where quality is fostered in software development. Design provides us with representations of software that can assess for quality. Design is the only way that we can accurately translate a customer's view into a finished software product or system. Software design serves as a foundation for all the software engineering steps that follow. Without a strong design we risk building an unstable system –

6.2. NORMALIZATION

It is a process of converting a relation to a standard form. The process is used to handle the problems that can arise due to data redundancy i.e. repetition of data in the database, maintain data integrity as well as handling problems that can arise due to insertion, updation, deletion anomalies.

Insertion anomaly: Inability to add data to the database due to absence of other data.

Deletion anomaly: Unintended loss of data due to deletion of other data.

Update anomaly: Data inconsistency resulting from data redundancy and partial update.

Normal Forms: These are the rules for structuring relations that eliminate anomalies.

FIRST NORMAL FORM:

A relation is said to be in first normal form if the values in the relation are atomic for every attribute in the relation. By this we mean simply that no attribute value can be a set of values or, as it is sometimes expressed, a repeating group.

SECOND NORMAL FORM:

A relation is said to be in second Normal form is it is in first normal form and it should satisfy any one of the following rules.

1) Primary key is a not a composite primary key
2) No non key attributes are present
3) Every non key attribute is fully functionally dependent on full set of primary key.

THIRD NORMAL FORM:

A relation is said to be in third normal form if their exits no transitive dependencies.

Transitive Dependency: If two non key attributes depend on each other as well as on the primary key then they are said to be transitively dependent.

The above normalization principles were applied to decompose the data in multiple tables thereby making the data to be maintained in a consistent state.

6.3. E – R DIAGRAMS

- The relation upon the system is structure through a conceptual ER-Diagram, which not only specifics the existential entities but also the standard relations through which the system exists and the cardinalities that are necessary for the system state to continue.
- The entity Relationship Diagram (ERD) depicts the relationship between the data objects. The ERD is the notation that is used to conduct the date modeling activity the attributes of each data object noted is the ERD can be described resign a data object descriptions.
- The set of primary components that are identified by the ERD are
    - ◆ Data object
    - ◆ Relationships
    - ◆ Attributes
    - ◆ Various types of indicators.

The primary purpose of the ERD is to represent data objects and their relationships.

6.4. DATA FLOW DIAGRAMS

A data flow diagram is graphical tool used to describe and analyze movement of data through a system. These are the central tool and the basis from which the other components are developed. The transformation of data from input to output, through processed, may be described logically and independently of physical components associated with the system. These are known as the logical data flow diagrams. The physical data flow diagrams show the actual implements and movement of data between people, departments and workstations. A full description of a system actually consists of a set of data flow diagrams. Using two familiar notations Yourdon, Gane and Sarson notation develops the data flow diagrams. Each component in a DFD is labeled with a descriptive name. Process is further identified with a number that will be used for identification purpose. The development of DFD'S is done in several levels. Each process in lower level diagrams can be broken down into a more detailed DFD in the next level. The lop-level diagram is often called context diagram. It consists a single process bit, which plays vital role in studying the current system. The process in the context level diagram is exploded into other process at the first level DFD.

The idea behind the explosion of a process into more process is that understanding at one level of detail is exploded into greater detail at the next level. This is done until further explosion is necessary and an adequate amount of detail is described for analyst to understand the process.

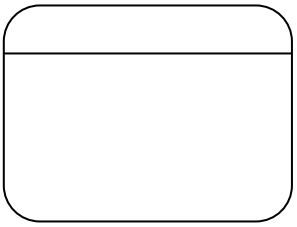
Larry Constantine first developed the DFD as a way of expressing system requirements in a graphical from, this lead to the modular design.

A DFD is also known as a “bubble Chart” has the purpose of clarifying system requirements and identifying major transformations that will become programs in system design. So it is the starting point of the design to the lowest level of detail. A DFD consists of a series of bubbles joined by data flows in the system.

DFD SYMBOLS:

In the DFD, there are four symbols

1. A square defines a source(originator) or destination of system data
2. An arrow identifies data flow. It is the pipeline through which the information flows
3. A circle or a bubble represents a process that transforms incoming data flow into outgoing data flows.
4. An open rectangle is a data store, data at rest or a temporary repository of data



Process that transforms data flow.

Source or Destination of data

Data flow

CONSTRUCTING A DFD:

Several rules of thumb are used in drawing DFD'S:

1. Process should be named and numbered for an easy reference. Each name should be representative of the process.
2. The direction of flow is from top to bottom and from left to right. Data traditionally flow from source to the destination although they may flow back to the source. One way to indicate this is to draw long flow line back to a source. An alternative way is to repeat the source symbol as a destination. Since it is used more than once in the DFD it is marked with a short diagonal.
3. When a process is exploded into lower level details, they are numbered.
4. The names of data stores and destinations are written in capital letters. Process and dataflow names have the first letter of each work capitalized

A DFD typically shows the minimum contents of data store. Each data store should contain all the data elements that flow in and out.

Questionnaires should contain all the data elements that flow in and out. Missing interfaces redundancies and like is then accounted for often through interviews.

SAILENT FEATURES OF DFD'S

1. The DFD shows flow of data, not of control loops and decision are controlled considerations do not appear on a DFD.
2. The DFD does not indicate the time factor involved in any process whether the dataflow take place daily, weekly, monthly or yearly.
3. The sequence of events is not brought out on the DFD.

4. TYPES OF DATA FLOW DIAGRAMS

1. Current Physical
2. Current Logical
3. New Logical
4. New Physical

CURRENT PHYSICAL:

In Current Physical DFD process label include the name of people or their positions or the names of computer systems that might provide some of the overall system-processing label includes an identification of the technology used to process the data. Similarly data flows and data stores are often labels with the names of the actual physical media on which data are stored such as file folders, computer files, business forms or computer tapes.

CURRENT LOGICAL:

The physical aspects at the system are removed as mush as possible so that the current system is reduced to its essence to the data and the processors that transforms them regardless of actual physical form.

NEW LOGICAL:

This is exactly like a current logical model if the user were completely happy with he user were completely happy with the functionality of the current system but had problems with how it was implemented typically through the new logical model will differ from current logical model while having additional functions, absolute function removal and inefficient flows recognized.

NEW PHYSICAL:

The new physical represents only the physical implementation of the new system.

RULES GOVERNING THE DFD'S

PROCESS

1) No process can have only outputs.
2) No process can have only inputs. If an object has only inputs than it must be a sink.
3) A process has a verb phrase label.

DATA STORE

1) Data cannot move directly from one data store to another data store, a process must move data.
2) Data cannot move directly from an outside source to a data store, a process, which receives, must move data from the source and place the data into data store
3) A data store has a noun phrase label.

SOURCE OR SINK

The origin and / or destination of data.

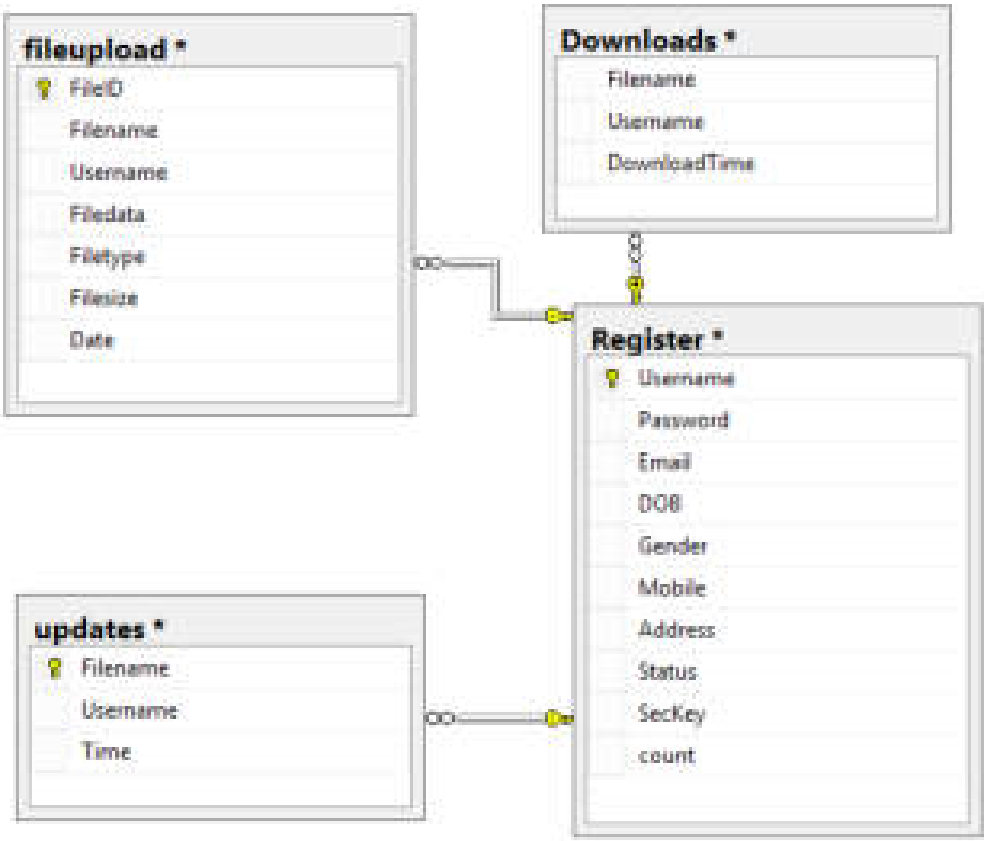
1) Data cannot move direly from a source to sink it must be moved by a process
2) A source and /or sink has a noun phrase land

DATA FLOW

1) A Data Flow has only one direction of flow between symbols. It may flow in both directions between a process and a data store to show a read before an update. The later is usually indicated however by two separate arrows since these happen at different type.
2) A join in DFD means that exactly the same data comes from any of two or more different processes data store or sink to a common location.
3) A data flow cannot go directly back to the same process it leads. There must be at-least one other process that handles the data flow produce some other data flow returns the original data into the beginning process.
4) A Data flow to a data store means update (delete or change).
5) A data Flow from a data store means retrieve or use.

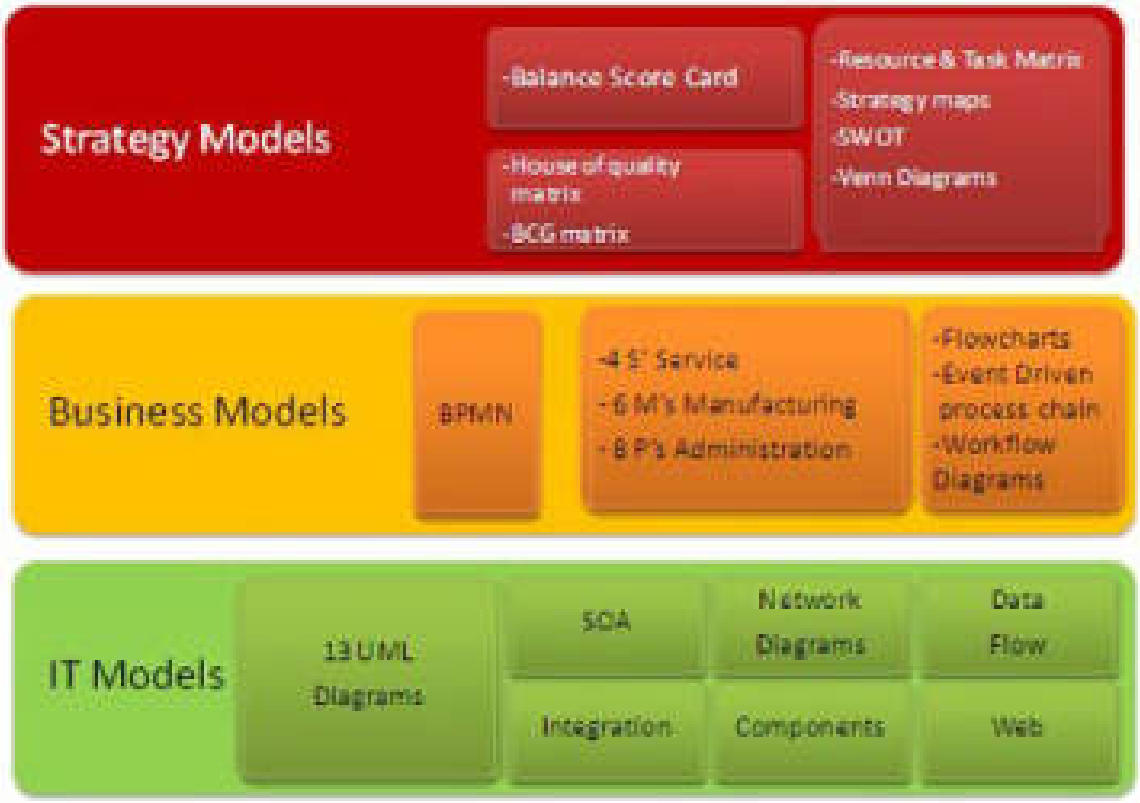
A data flow has a noun phrase label more than one data flow noun phrase can appear on a single arrow as long as all of the flows on the same arrow move together as one package.

ER-DIAGRAM


fileupload *
FileID
Filename
Username
Filedata
Filetype
Filesize
Date
Downloads *
Filename
Username
DownloadTime
Register *
Username
Password
Email
DOB
Gender
Mobile
Address
Status
SecKey
count
updates *
Filename
Username
Time

6.3. UML DIAGRAMS

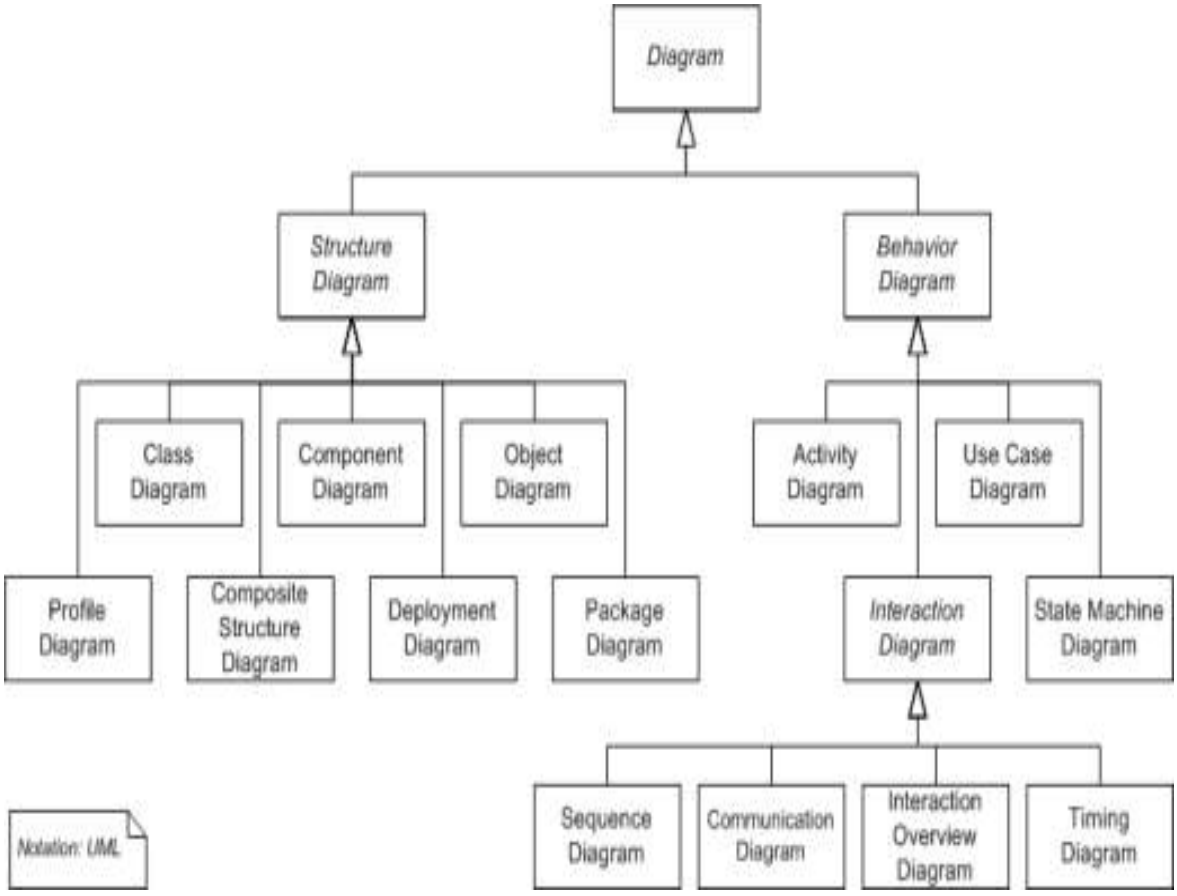
The Unified Modeling Language (UML) is used to specify, visualize, modify, construct and document the artifacts of an object-oriented software intensive system under development. The UML uses mostly graphical notations to express the design of software projects. UML offers a standard way to visualize a system's architectural blueprints, including elements such as:



- actors
- business processes
- (logical) components
- activities

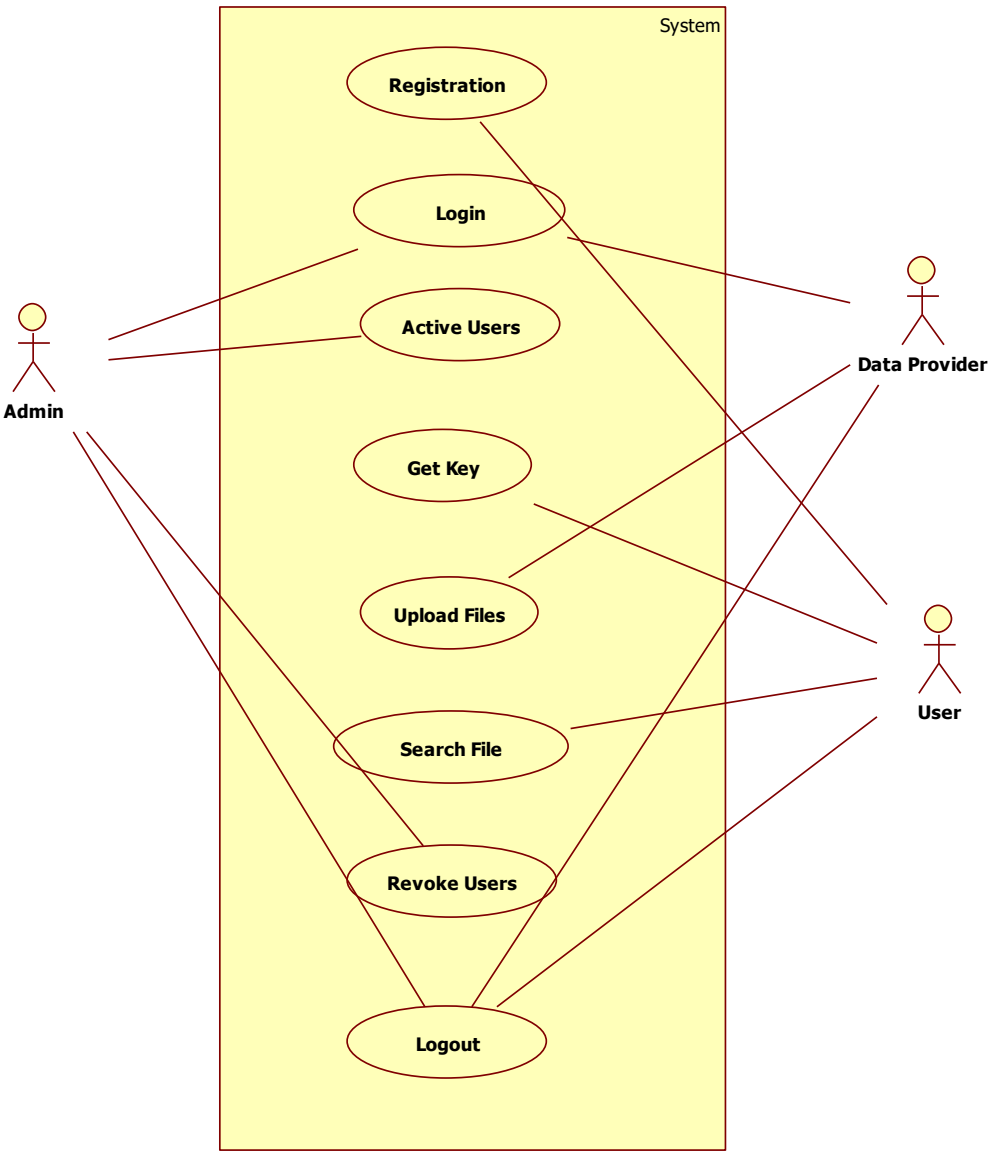
- programming language statements
- database schemas, and
- Reusable software components.

UML Diagrams Overview



UML combines best techniques from data modeling (entity relationship diagrams), business modeling (work flows), object modeling, and component modeling. It can be used with all processes, throughout the software development life cycle, and across different implementation technologies. UML has synthesized the notations of the Booch method, the Object-modeling technique (OMT) and Object-oriented software engineering (OOSE) by fusing them into a single, common and widely usable modeling language. UML aims to be a standard modeling language which can model concurrent and distributed systems.

Use Case Diagram:-



Sequence Diagrams:

From the name *Interaction* it is clear that the diagram is used to describe some type of interactions among the different elements in the model. So this interaction is a part of dynamic behavior of the system.

This interactive behavior is represented in UML by two diagrams known as *Sequence diagram* and *Collaboration diagram*. The basic purposes of both the diagrams are similar.

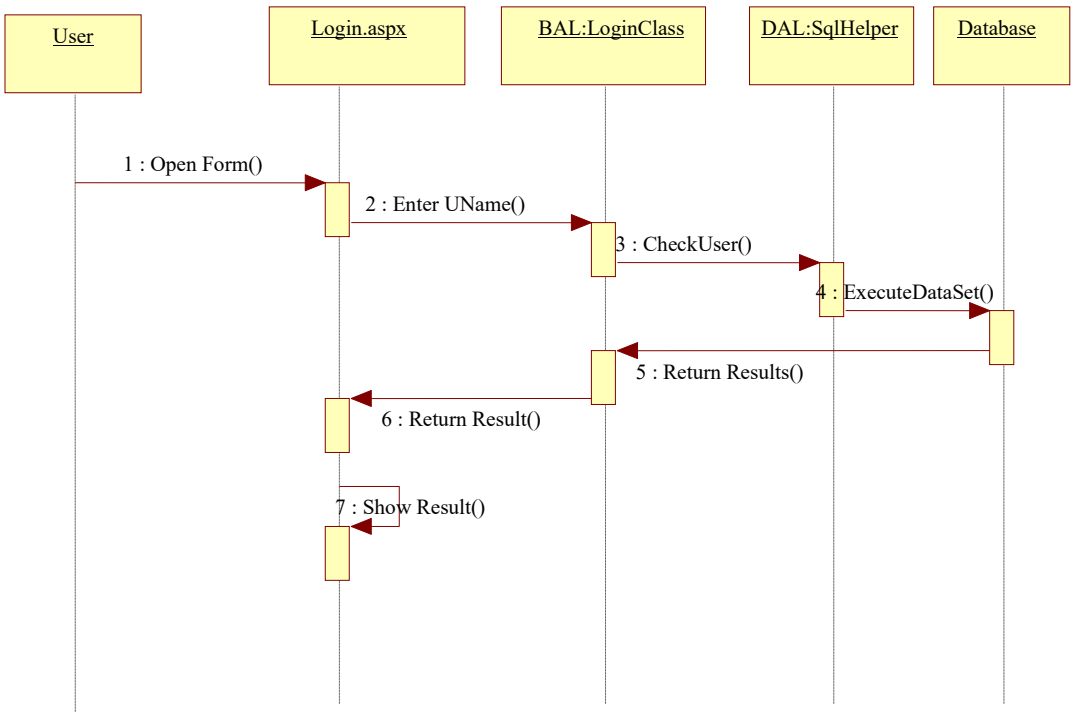
Sequence diagram emphasizes on time sequence of messages and collaboration diagram emphasizes on the structural organization of the objects that send and receive messages.

A sequence diagram is an interaction diagram. From the name it is clear that the diagram deals with some sequences, which are the sequence of messages flowing from one object to another.

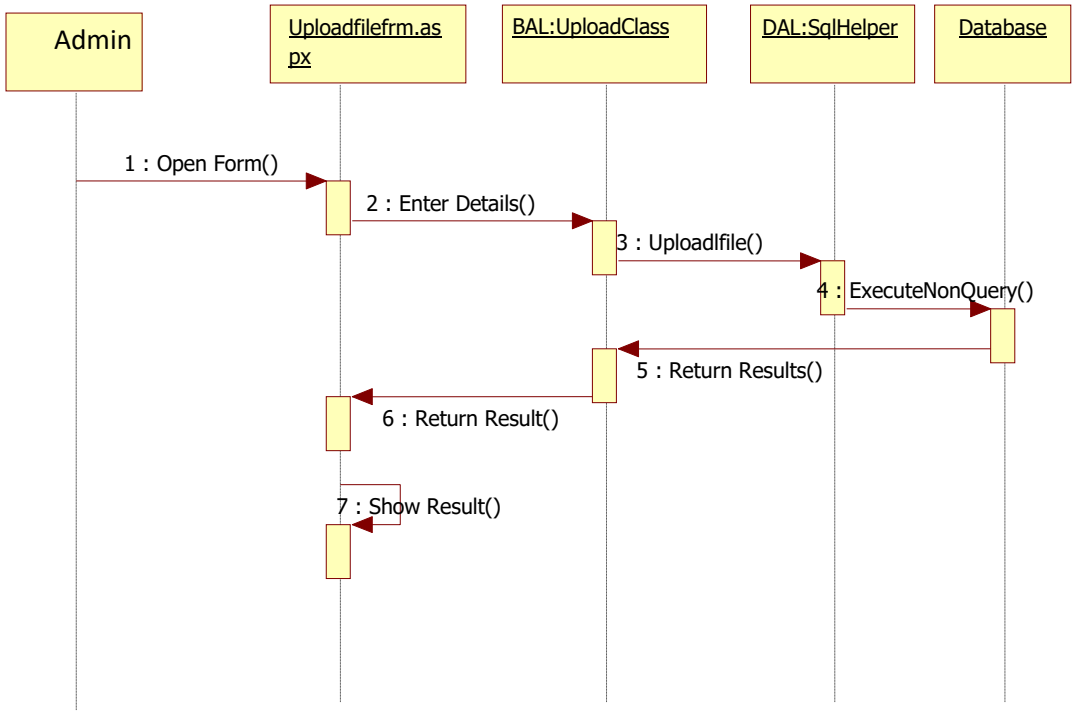
Interaction among the components of a system is very important from implementation and execution perspective.

So Sequence diagram is used to visualize the sequence of calls in a system to perform a specific functionality.

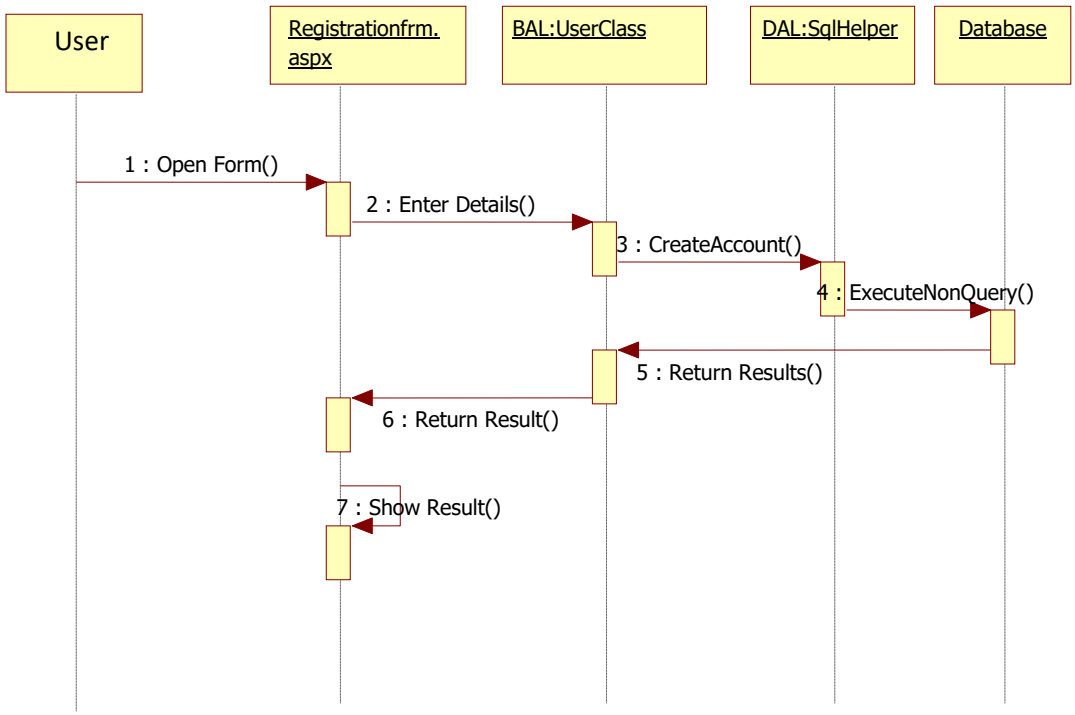
Login Sequence Diagram:



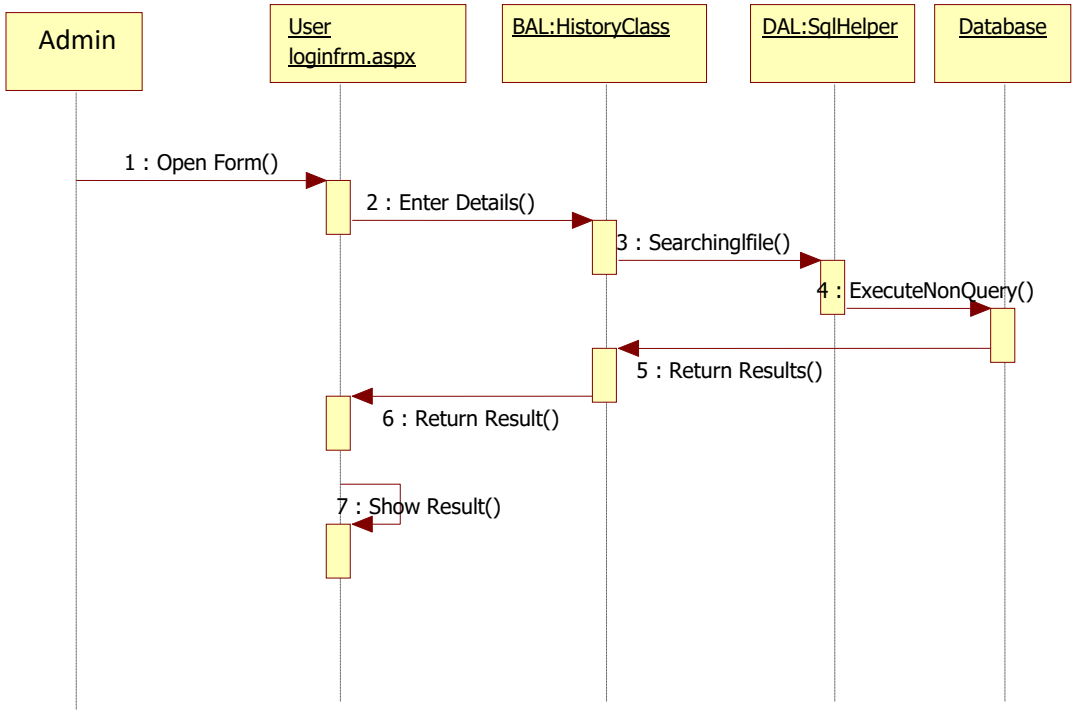
Upload File:


Admin
Uploadfilefrm.aspx
BAL:UploadClass
DAL:SqlHelper
Database
1 : Open Form()
2 : Enter Details()
3 : Uploadlfile()
4 : ExecuteNonQuery()
5 : Return Results()
6 : Return Result()
7 : Show Result()

User Registration Sequence Diagram:



User Login History Sequence Diagram:

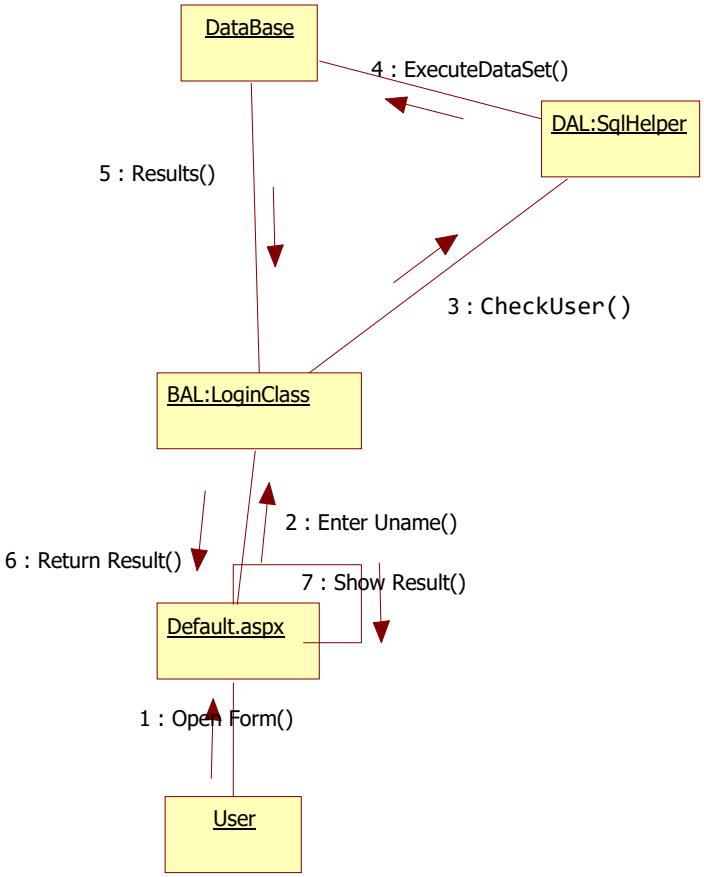


DATA DICTONARY

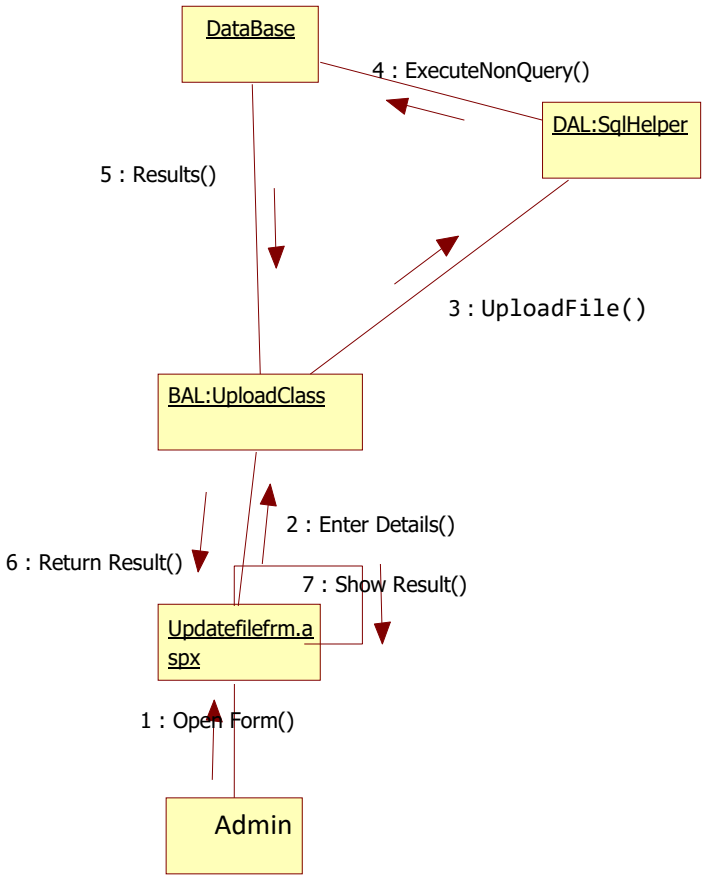
After carefully understanding the requirements of the client the the entire data storage requirements are divided into tables. The below tables are normalized to avoid any anomalies during the course of data entry.

Collaboration Diagram

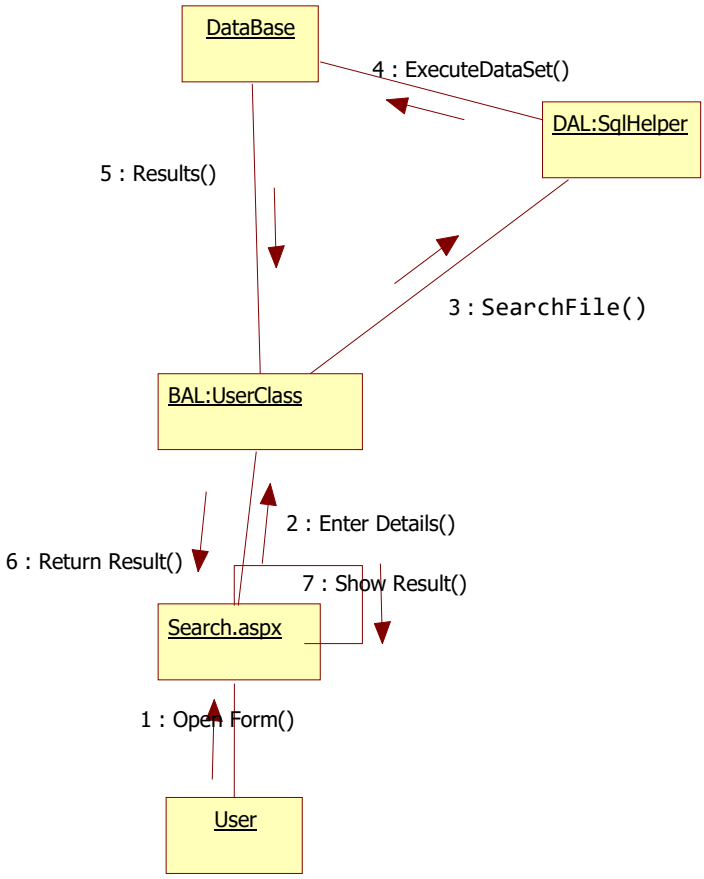
Login Collaboration



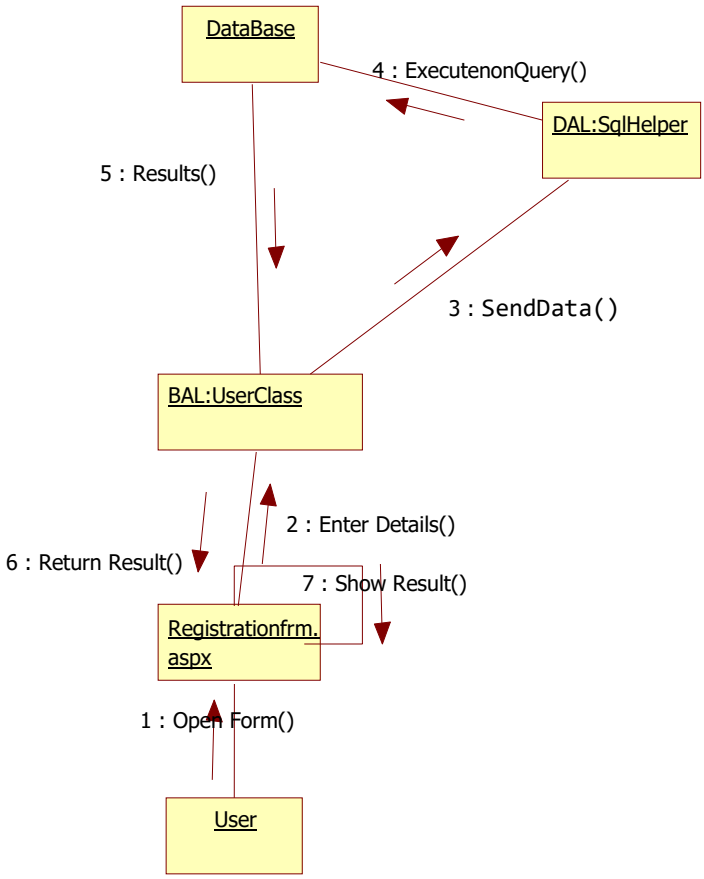
Upload File:-


DataBase
4 : ExecuteNonQuery()
DAL:SqlHelper
5 : Results()
3 : UploadFile()
BAL:UploadClass
2 : Enter Details()
6 : Return Result()
7 : Show Result()
Updatefilefrm.aspx
1 : Open Form()
Admin

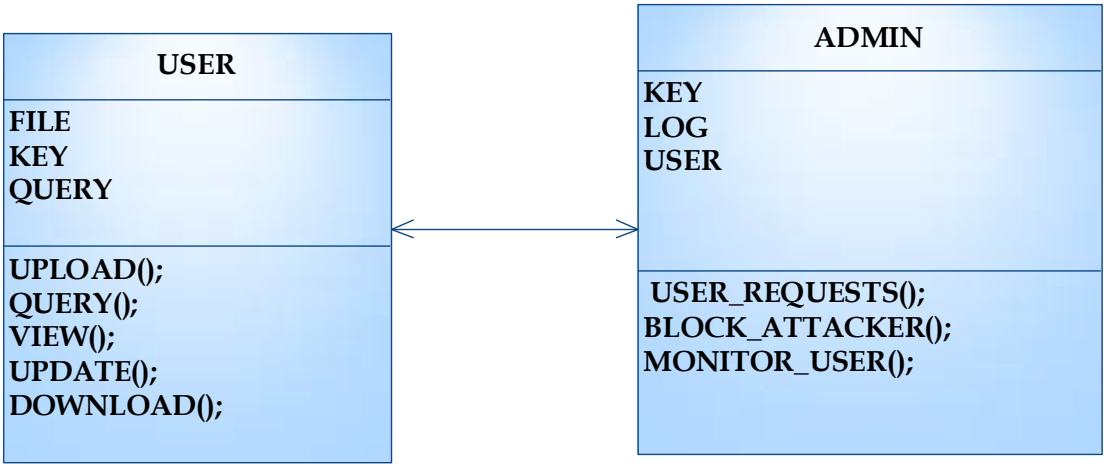
User Serach:-


DataBase
4 : ExecuteDataSet()
DAL:SqlHelper
5 : Results()
3 : SearchFile()
BAL:UserClass
2 : Enter Details()
6 : Return Result()
7 : Show Result()
Search.aspx
1 : Open Form()
User

User Registration:



Class Diagram:



Activity Diagrams-:

Activity diagrams are graphical representations of Workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

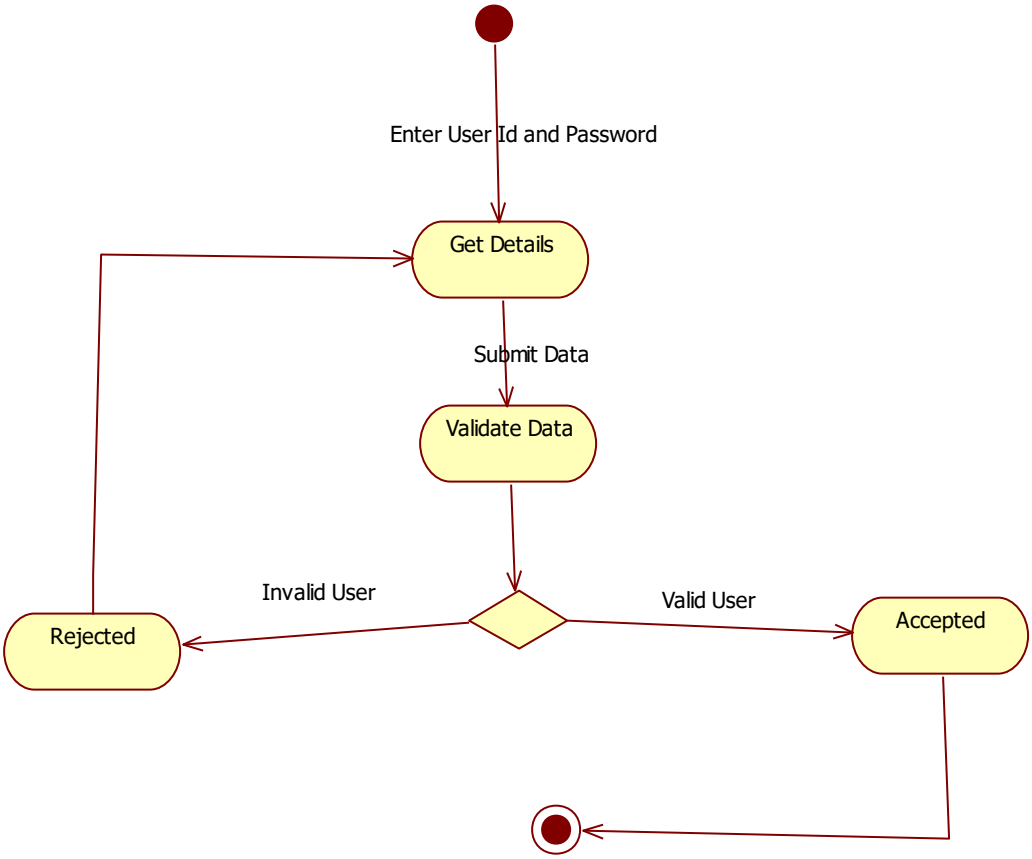
`Activity diagrams are constructed from a limited number of shapes, connected with arrows. The most important shape types:

- *rounded rectangles* represent *activities*;
- *diamonds* represent *decisions*;
- *bars* represent the start (*split*) or end (*join*) of concurrent activities;
- a *black circle* represents the start (*initial state*) of the workflow;
- An *encircled black circle* represents the end (*final state*).

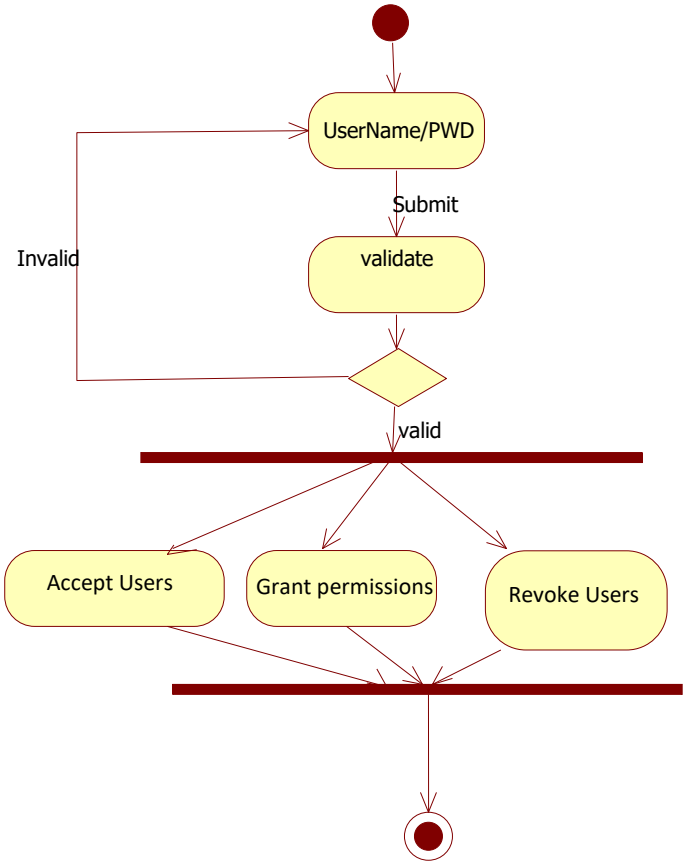
Arrows run from the start towards the end and represent the order in which activities happen.

Hence they can be regarded as a form of flowchart. Typical flowchart techniques lack constructs for expressing concurrency. However, the join and split symbols in activity diagrams only resolve this for simple cases; the meaning of the model is not clear when they are arbitrarily combined with decisions or loops.

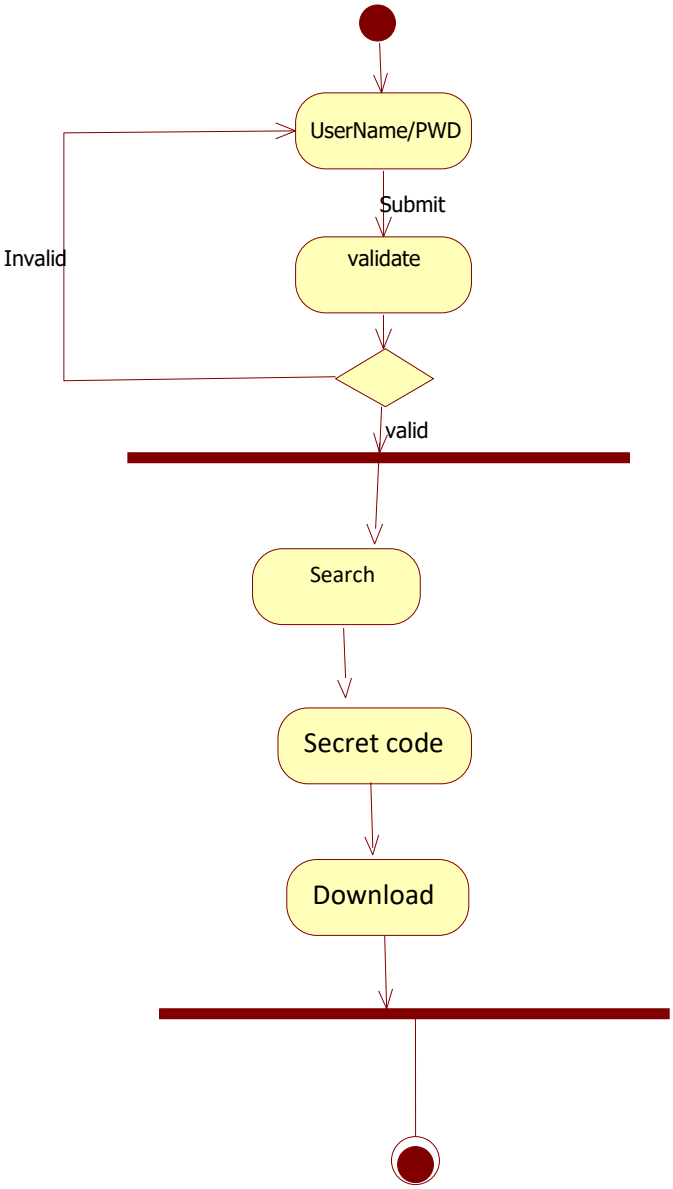
Login Activity



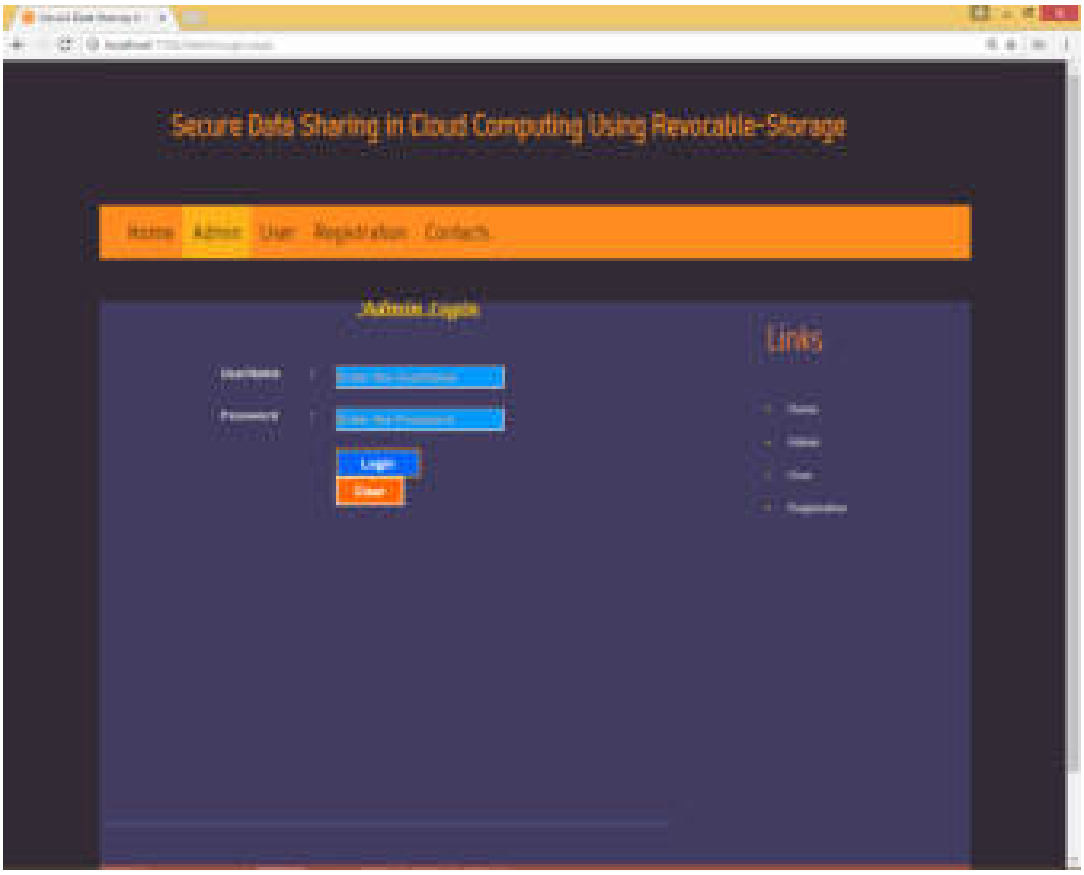
Admin Activity:-

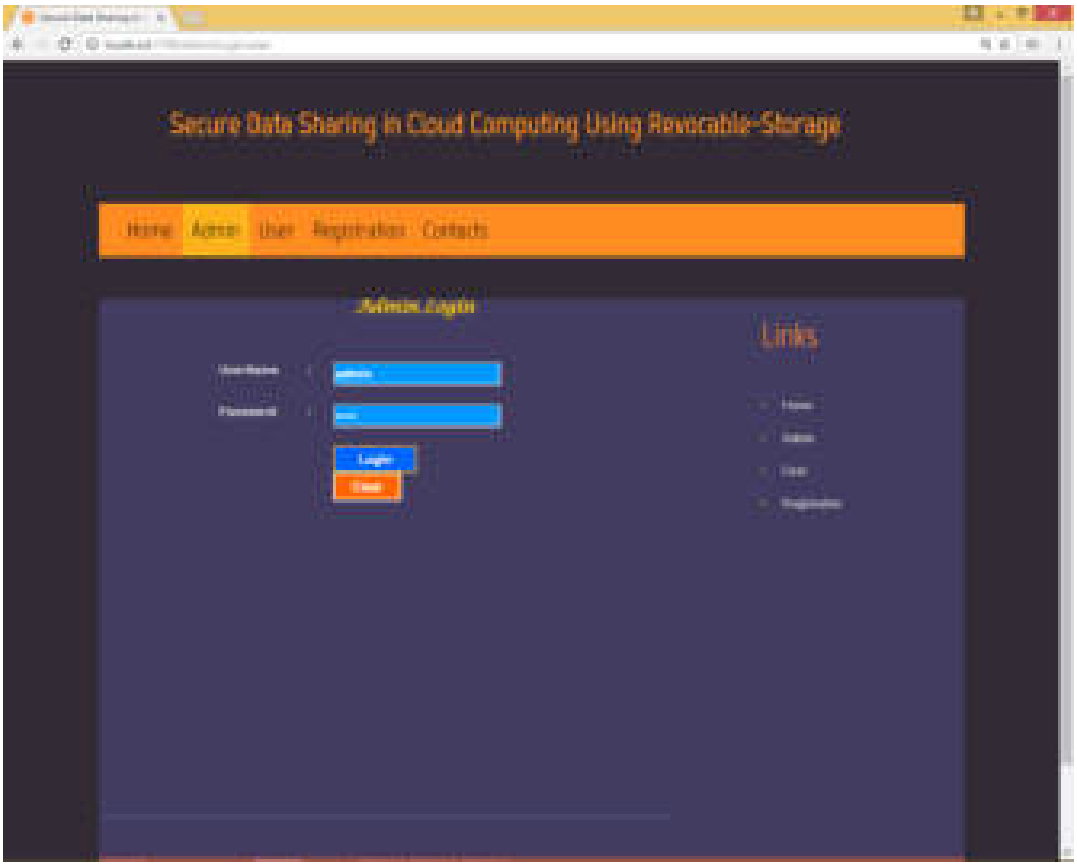


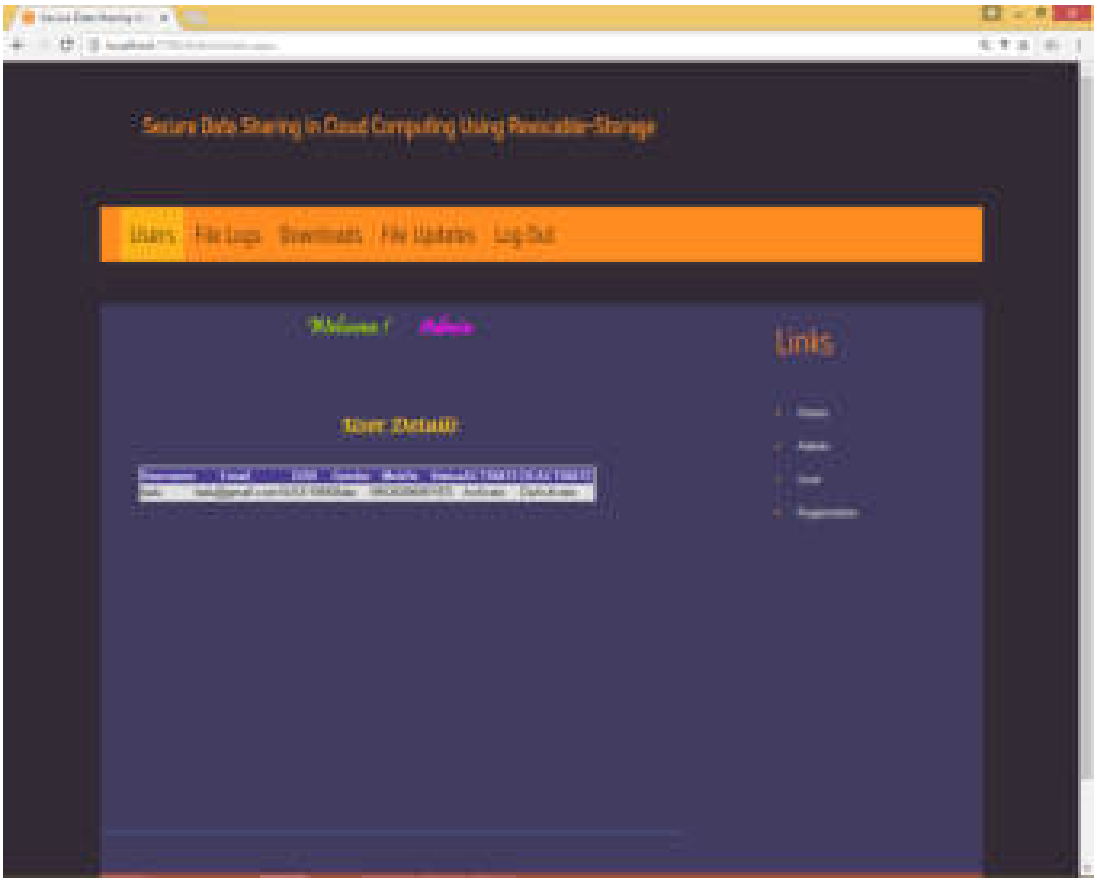
USER Activity:

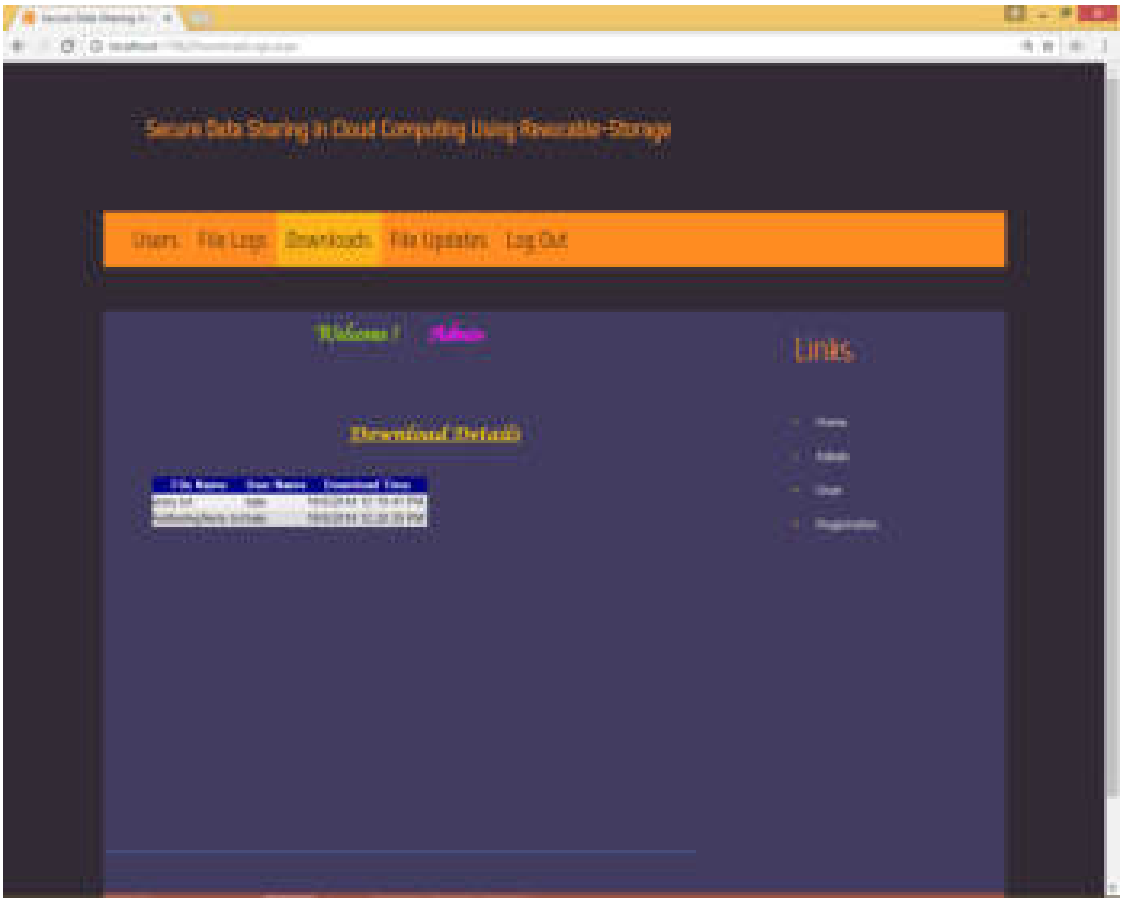


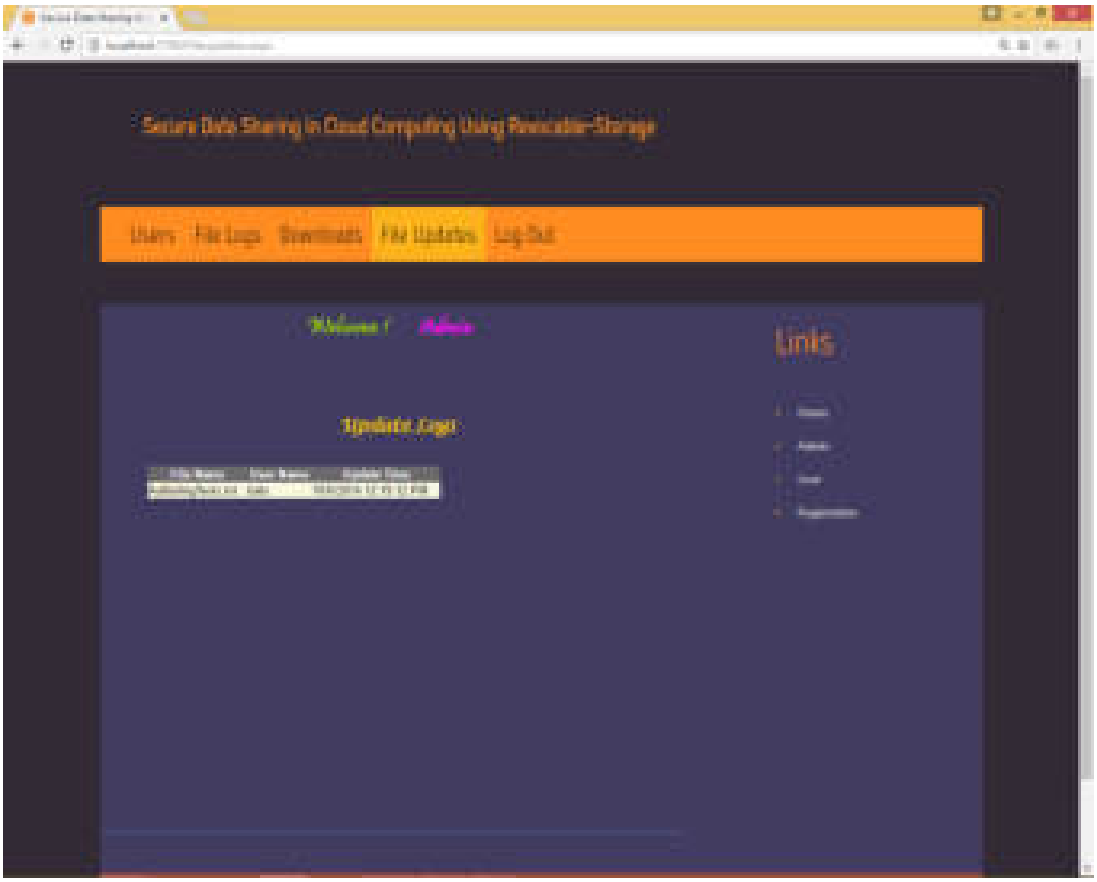
OUTPUT SCREENS


Secure Data Sharing in Cloud Computing Using Revocable-Storage
Home
Admin
User
Registration
Contact
Admin Login
Links


Secure Data Sharing in Cloud Computing Using Revocable-Storage
Home
Admin
User
Registration
Contacts
Admin Login
Links


Secure Data Sharing in Cloud Computing Using Revocable-Storage
Users
File Logs
Downloads
File Updates
Log Out
Links


Secure Data Sharing in Cloud Computing Using Revocable-Storage
Users
File Logs
Downloads
File Updates
Log Out
Welcome !
Admin
Download Details
Links


Secure Data Sharing in Cloud Computing Using Revocable-Storage
Users
File Logs
Downloads
File Updates
Log Out
Welcome !
Admin
Update Logs
Links

SYSTEM TESTING AND IMPLEMENTATION

8.1INTRODUCTION

Software testing is a critical element of software quality assurance and represents the ultimate review of specification, design and coding. In fact, testing is the one step in the software engineering process that could be viewed as destructive rather than constructive.

A strategy for software testing integrates software test case design methods into a well-planned series of steps that result in the successful construction of software. Testing is the set of activities that can be planned in advance and conducted systematically. The underlying motivation of program testing is to affirm software quality with methods that can economically and effectively apply to both strategic to both large and small-scale systems.

8.2. STRATEGIC APPROACH TO SOFTWARE TESTING

The software engineering process can be viewed as a spiral. Initially system engineering defines the role of software and leads to software requirement analysis where the information domain, functions, behavior, performance, constraints and validation criteria for software are established. Moving inward along the spiral, we come to design and finally to coding. To develop computer software we spiral in along streamlines that decrease the level of abstraction on each turn.

A strategy for software testing may also be viewed in the context of the spiral. Unit testing begins at the vertex of the spiral and concentrates on each unit of the software as implemented in source code. Testing progress by moving outward along the spiral to integration testing, where the focus is on the design and the construction of the software architecture.

8.3. UNIT TESTING

Unit testing focuses verification effort on the smallest unit of software design, the module. The unit testing we have is white box oriented and some modules the steps are conducted in parallel.

1. WHITE BOX TESTING

This type of testing ensures that

- All independent paths have been exercised at least once
- All logical decisions have been exercised on their true and false sides
- All loops are executed at their boundaries and within their operational bounds
- All internal data structures have been exercised to assure their validity.

To follow the concept of white box testing we have tested each form .we have created independently to verify that Data flow is correct, All conditions are exercised to check their validity, All loops are executed on their boundaries.

2. BASIC PATH TESTING

Established technique of flow graph with Cyclomatic complexity was used to derive test cases for all the functions. The main steps in deriving test cases were:

Use the design of the code and draw correspondent flow graph.

Determine the Cyclomatic complexity of resultant flow graph, using formula:

V(G)=E-N+2 or

V(G)=P+1 or

V (G) =Number Of Regions

Where V (G) is Cyclomatic complexity,

E is the number of edges,

N is the number of flow graph nodes,

P is the number of predicate nodes.

Determine the basis of set of linearly independent paths.

3. CONDITIONAL TESTING

In this part of the testing each of the conditions were tested to both true and false aspects. And all the resulting paths were tested. So that each path that may be generate on particular condition is traced to uncover any possible errors.

4. DATA FLOW TESTING

This type of testing selects the path of the program according to the location of definition and use of variables. This kind of testing was used only when some local variable were declared. The *definition-use chain* method was used in this type of testing. These were particularly useful in nested statements.

5. LOOP TESTING

In this type of testing all the loops are tested to all the limits possible. The following exercise was adopted for all loops:

- All the loops were tested at their limits, just above them and just below them.
- All the loops were skipped at least once.
- For nested loops test the inner most loop first and then work outwards.
- For concatenated loops the values of dependent loops were set with the help of connected loop.
- Unstructured loops were resolved into nested loops or concatenated loops and tested as above

SYSTEM SCUERITY

9.1 INTRODUCTION

The protection of computer based resources that include hardware, software, data, procedures and people against unauthorized use or natural Disaster is known as System Security.

System Security can be divided into four related issues:

- Security
- Integrity
- Privacy
- Confidentiality

SYSTEM SECURITY refers to the technical innovations and procedures applied to the hardware and operation systems to protect against deliberate or accidental damage from a defined threat.

DATA SECURITY is the protection of data from loss, disclosure, modification and destruction.

SYSTEM INTEGRITY refers to the power functioning of hardware and programs, appropriate physical security and safety against external threats such as eavesdropping and wiretapping.

PRIVACY defines the rights of the user or organizations to determine what information they are willing to share with or accept from others and how the organization can be protected against unwelcome, unfair or excessive dissemination of information about it.

CONFIDENTIALITY is a special status given to sensitive information in a database to minimize the possible invasion of privacy. It is an attribute of information that characterizes its need for protection.

9.3 SECURITY SOFTWARE

It is the technique used for the purpose of converting communication. It transfers message secretly by embedding it into a cover medium with the use of information hiding techniques. It is one of the conventional techniques capable of hiding large secret message in a cover image without introducing many perceptible distortions.

NET has two kinds of security:

- Role Based Security
- Code Access Security

The Common Language Runtime (CLR) allows code to perform only those operations that the code has permission to perform. So CAS is the CLR's security system that enforces security policies by preventing unauthorized access to protected **resources** and **operations**. Using the Code Access Security, you can do the following:

- Restrict what your code can do
- Restrict which code can call your code
- Identify code

CONCLUSION

Cloud computing brings great convenience for people. Particularly, it perfectly matches the increased need of sharing data over the Internet. In this paper, to build a cost-effective and secure data sharing system in cloud computing, we proposed a notion called RS-IBE, which supports identity revocation and ciphertext update simultaneously such that a revoked user is prevented from accessing previously shared data, as well as subsequently shared data. Furthermore, a concrete construction of RS-IBE is presented. The proposed RS-IBE scheme is proved adaptive-secure in the standard model, under the decisional ℓ-DBHE assumption. The comparison results demonstrate that our scheme has advantages in terms of efficiency and functionality, and thus is more feasible for practical applications.

BIBLIOGRAPHY

[1] L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A break in the clouds: towards a cloud definition," ACM SIGCOMM Computer Communication Review, vol. 39, no. 1, pp. 50–55.

[2] iCloud. Apple storage service. [Online]. Available: https://www.icloud.com/

[3] Azure storage service. [Online]. Available: http://www.windowsazure.com/

[4] Amazon. (2014) Amazon simple storage service (amazon s3). [Online]. Available: http://aws.amazon.com/s3/

[5] K. Chard, K. Bubendorfer, S. Caton, and O. F. Rana, "Social cloud computing: A vision for socially motivated resource sharing," Services Computing, IEEE Transactions on, vol. 5, no. 4, pp. 551–563, 2012.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"CLOUD BASED MULTIMEDIA CONTENT PROTECTION SYSTEM"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
COMPUTER SCIENCE AND ENGINEERING

Submitted by

SK. HASEENA(17H71A0574)

A.HARISH(17H71A0573)

K.RAMANADH (17H71A0597)

D.SAI SANDEEP (17H71A05A3)

Under the Esteemed Guidance of

MR MD ELIAZ

ASSISTANT PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC A
With Grade

2020-21

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited B.Tech (CSE|ECE|ME)
Accredited by NAAC With Grade A

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"CLOUD BASED MULTIMEDIA CONTENT PROTECTION SYSTEM"** is a bonafide work carried out by **MS.SK Haseena (17H71A0574), MR.A Harish(17H71A0573), MR.K Ramanadh(17H71A0597), MR.D Sai Sandeep (17H71A05A3)** in partial fulfilment for the award of degree of Bachelor of Technology in **Computer Science and Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

MD.Eliaz
Project Guide

D. Prasad
Head of the Department

Examiner

Dr K Srinivas
PRINCIPAL

ACKNOWLEDGEMENT

We would like to take this opportunity to express my deepest appreciation to the following people for their valuable contributions and assistance with this Project.

Initially, we would like to thank my project supervisor, **Mr.MD Eliaz, Assistant Professor, Department of Computer Science and Engineering** for the guidance and support, especially for the valuable ideas and knowledge shared to me throughout the Project.

We have the immense pleasure in expressing my thanks and deep sense of gratitude to**, Mr. D. Prasad, Head of the Department, Computer Science and Engineering** for extending necessary facilities and support for the completion of the Project.

We whole sincerely acknowledge **Dr. K. Srinivas, Principal and Prof D. PANDURANGA RAO, CEO** for giving opportunity to take up the Project work. We also extend my thanks to all faculty members of **Computer Science and Engineering**, for their valuable guidance and encouragement.

We would like to extend my warm appreciation to all my friends for sharing us their knowledge, valuable contributions and help with this Project.

Finally, my special thanks go to my family for their continuous support and help throughout my academic years and for their continual support and encouragement for the completion of the project.

DECLARATION

We **SK HASEENA(17H71A0574),A HARISH(17H71A0573),K RAMANADH (17H71A0597),D SAI SANDEEP(17H71A05A3),**of the Main-Project "**CLOUD BASED MULTIMEDIA CONTENT PROTECTION SYSTEM**", hereby declare that the matter embodied in this Project is the genuine work done by us and has not been submitted either to this university or to any other university/institute for the fulfilment of the requirement of any course of study.

SK HASEENA(17H71A0574)
A HARISH(17H71A0573)
K RAMANADH(17H71A0597)
D SAI SANDEEP(17H71A05A3)

CONTENTS

ABSTRACT:

We propose a new design for large-scale multimedia content protection systems. Our design leverages cloud infrastructures to provide cost efficiency, rapid deployment, scalability, and elasticity to accommodate varying workloads. The proposed system can be used to protect different multimedia content types, including videos, images, audio clips, songs, and music clips. The system can be deployed on private and/or public clouds. Our system has two novel components: (i) method to create signatures of videos, and (ii) distributed matching engine for multimedia objects. The signature method creates robust and representative signatures of videos that capture the depth signals in these videos and it is computationally efficient to compute and compare as well as it requires small storage. The distributed matching engine achieves high scalability and it is designed to support different multimedia objects. We implemented the proposed system and deployed it on two clouds: Amazon cloud and our private cloud. Our experiments with more than 11,000 videos and 1 million images show the high accuracy and scalability of the proposed system. In addition, we compared our system to the protection system used by YouTube and our results show that the YouTube protection system fails to detect most copies of videos, while our system detects more than 98% of them. This comparison shows the need for the proposed signature method, since the state-of-the-art commercial system was not able to handle videos.

LISTOF FIGURES

INTRODUCTION

Advances in processing and recording equipment of multimedia content as well as the availability of free online hosting sites have made it relatively easy to duplicate copyrighted materials such as videos, images, and music clips. Illegally redistributing multimedia content over the Internet can result in significant loss of revenues for content creators. Finding illegally-made copies over the Internet is a complex and computationally expensive operation, because of the sheer volume of the available multimedia content over the Internet and the complexity of comparing content to identify copies.

We present a novel system for multimedia content protection on cloud infrastructures. The system can be used to protect various multimedia content types, including regular videos, images, audio clips, songs, and music clips. The system can run on private clouds, public clouds, or any combination of public-private clouds. Our design achieves rapid deployment of content protection systems, because it is based on cloud infrastructures that can quickly provide computing hardware and software resources. The design is cost effective because it uses the computing resources on demand. The design can be scaled up and down to support varying amounts of multimedia content being protected. The proposed system is fairly complex with multiple components, including: (i) Crawler to download thousands of multimedia objects from online hosting sites, (ii) Signature method to create representative fingerprints from multimedia objects, and (iii) distributed matching engine to store signatures of original objects and match them against query objects. We propose novel methods for the second and third components, and we utilize off-the-shelf tools for the crawler. We have developed a complete running system of all components and tested it with more than 11,000 videos and 1 million images. We deployed parts of the system on the Amazon cloud with varying number of machines (from 8 to 128), and the other parts of the system were deployed on our private cloud. This deployment model was used to show the flexibility of our system, which enables it to efficiently utilize varying computing resources and minimize the cost, since cloud providers offer different pricing models for computing and network resources. Through extensive experiments with real deployment, we show the high accuracy (in terms of precision and recall) as well as the scalability and elasticity of the proposed system. The contributions of this paper are as follows:

- Complete multi-cloud system for multimedia content protection. The system supports different types of multimedia content and can effectively utilize varying computing resources.
- Novel method for creating signatures for videos. This method creates signatures that capture the depth in stereo content without computing the depth signal itself, which is a computationally expensive process.
- New design for a distributed matching engine for high-dimensional multimedia objects. This design provides the primitive function of finding K-nearest neighbors for large-scale datasets. The design also offers an auxiliary function for further processing of the K neighbors. This two-level design enables the proposed system to easily support different types of multimedia content. For example, in finding video copies, the temporal aspects need to be considered in addition to matching individual frames. This is unlike finding image copies. Our design of the matching engine employs the MapReduce programming model.
- Rigorous evaluation study using real implementation to assess the performance of the proposed system and compare it against the closest works in academia and industry. Specifically, we evaluate the entire end-to-end system with 11,000 videos downloaded from YouTube. Our results show that a high precision, close to 100%, with a recall of more than 80% can be achieved even if the videos are subjected to various transformations such as blurring, cropping, and text insertion. In addition, we compare our system versus the Content ID system used by YouTube to protect videos. Our results show that although the Content ID system provides robust detection of video copies, it fails to detect copies of videos when videos are subjected to even simple transformations such as re-encoding and resolution change. Our system, on the other hand, can detect almost all copies of videos even if they are subjected to complex transformations such as synthesizing new virtual views and converting videos to anaglyph and plus-depth formats.

Furthermore, we isolate and evaluate individual components of our system. The evaluation of the new signature method shows that it can achieve more than 95% precision and recall for stereoscopic content subjected to 15 different video transformations; several of them

are specific to videos such as view synthesis. The evaluation of the distributed matching engine was done on the Amazon cloud with up to 128 machines. The engine was used to manage up to 160 million data points, each with 128 dimensions, extracted from over 1 million images. The results show that our design of the matching engine is elastic and scalable. They also show that our system outperforms the closest object matching system in the literature, called Rank Reduce, by a wide margin in accuracy and it is more efficient in terms of space and computation.

1.1 PROBLEM STATEMENT:

- The problem of protecting various types of multimedia content has attracted significant attention from academia and industry. One approach to this problem is using watermarking, in which some distinctive information is embedded in the content itself and a method is used to search for this information in order to verify the authenticity of the content.

1.2 PURPOSE:

- We propose a new design for large-scale multimedia content protection systems. Our design leverages cloud infrastructures to provide cost efficiency, rapid deployment, scalability, and elasticity to accommodate varying workloads. The proposed system can be used to protect different multimedia content types, including videos, images, audio clips, songs, and music clips.

1.3 MOTIVATION:

- Watermarking requires inserting watermarks in the multimedia objects before releasing them as well as mechanisms/systems to find objects and verify the existence of correct watermarks in them. Thus, this approach may not be suitable

for already-released content without watermarks in them.

1.4 SCOPE OF THE PROJECT:

- The signature method creates robust and representative signatures of videos that capture the depth signals in these videos and it is computationally efficient to compute and compare as well as it requires small storage.
- The distributed matching engine achieves high scalability and it is designed to support different multimedia objects

1. LITERATURE SURVEY

The problem of protecting various types of multimedia content has attracted significant attention from academia and industry. One approach to this problem is using watermarking, in which some distinctive information is embedded in the content itself and a method is used to search for this information in order to verify the authenticity of the content. Watermarking requires inserting watermarks in the multimedia objects before releasing them as well as mechanisms/systems to find objects and verify the existence of correct watermarks in them. Thus, this approach may not be suitable for already-released content without watermarks in them. The watermarking approach is more suitable for the somewhat controlled environments, such as distribution of multimedia content on DVDs or using special sites and custom players. Watermarking may not be effective for the rapidly increasing online videos, especially those uploaded to sites such as YouTube and played back by any video player. Watermarking is not the focus of this paper. The focus of this paper is on the other approach for protecting multimedia content, which is content based copy detection (CBCD). In this approach, signatures (or fingerprints) are extracted from original objects. Signatures are also created from query (suspected) objects downloaded from online sites. Then, the similarity is computed between original and suspected objects to find potential copies. Many previous works proposed different methods for creating and matching signatures. These methods can be classified into four categories: spatial, temporal, color and transform-domain. Spatial signatures (particularly the block-based) are the most widely used. However, their weakness is the lack of resilience against large geometric transformations. Temporal and color signatures are less robust and can be used to enhance spatial signatures. Transform-domain signatures are computationally intensive and not widely used in practice. For more details, see surveys for audio fingerprintingand 2D video fingerprinting. Youtube Content ID, Vobile VDNA and MarkMonitor are some of the industrial examples which use fingerprinting for media protection, while methods such as we can be referred to as the academic state-of-the-art. Unlike previous works, the contribution of this paper is to design a large-scale 1520-9210 (c) 2015 IEEE.

 This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2015.2389628, IEEE Transactions on Multimedia 5 system to find copies that can be used for different types of multimedia content and can leverage multicloud infrastructures to minimize the cost, expedite deployment, and dynamically scale up and down. That is, we design our system such that previous content-based copy detection methods for creating and matching signatures can be implemented within our system. In addition to our cloud-based system, we propose a new method for 3D video fingerprinting, and a new design for the distributed matching engine. The works related to each of these components are summarized in the following subsections. A. 3D Video Signatures Content-based copy detection of 3D videos is a fairly new problem.

The work in computes SIFT points in each view and uses the number of matching SIFT points to verify matches. Comparing all SIFT points in each frame is not practical for large databases due to the storage overhead and search complexity. On the other hand, the work in assumes that the depth maps are given or estimated. Estimating the depth map from stereoscopic videos is quite expensive. The method in is suitable for 3D videos encoded in the video plus depth format, but not for stereoscopic videos. Our proposed method in this paper captures the depth properties without calculating the depth map itself and it is computationally efficient because it does not compare all features in the frame. Although 3D copy detection methods are scarce in the literature, there are many methods available for 2D video copy detection. Hampapur et al. use the temporal features of the video as the signature. Similarly, Tasdemir et al. use motion vectors as the signature for each frame. Some methods use color histograms as signatures. The color histogram signature is prone to global variations in color which are common when recoding video. Another group of methods use interest points of video frames as signature. For example, Liu et al. use local SIFT features as the frame signature. Using gradient information has also shown to be robust to many 2D transformations. All of the above 2D video

fingerprinting methods can be implemented in the proposed system. In addition, while some of these methods can be used for 3D video copy detection, they are designed for 2D videos, and they ignore the information in different views and the depth of 3D videos. This information is important especially in the presence of 3D video transformations such as view synthesis, where views from different viewpoints can be generated using the depth map of the 3D video. When two new views are synthesized, the positioning of each pixel in the frame is changed, and some areas are occluded while other areas become visible. The luminance, gradient, color and even the interest points in 1520-9210 (c) 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information. This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2015.2389628, IEEE Transactions on Multimedia 6 each block can change as well when a new view is synthesized.

Thus, the extracted signature using any of the 2D methods will change accordingly. Therefore, when searching for similar signatures, manipulated versions may not be identified. The importance of using signatures that have some information from the depth signal has been shown. In addition, our experiments and comparisons in this paper show that the state-of-the-art copy detection system used by YouTube (called Content ID) fails to detect many simple transformations made on 3D videos such as re-encoding, conversion to row or column interleaved formats, and creating new virtual views. Based on the available information from the patent describing the Content ID system [10] and our own experiments, we believe that the poor performance of Content ID on 3D videos is because it does not consider any depth information. B. Distributed Matching Engine Unlike many of the previous works, e.g., [3] which designed a system for image matching, our proposed matching engine is general and it can support different types of multimedia objects, including images, 2D videos, and 3D videos. To achieve this generality, we divide the engine into two

main stages. The first stage computes nearest neighbors for a given data point, and the second stage post-processes the computed neighbors based on the object type. In addition, our design supports high-dimensionality which is needed for multimedia objects that are rich in features. Computing nearest neighbors is a common problem in many applications. Our focus in this paper is on distributed techniques that can scale to large datasets. Liao et al. build a multi-dimensional index using R-tree on top of the Hadoop distributed file system (HDFS). Their index, however, can only handle low dimensional datasets—they performed their experiments with two dimensional data. They solve the K nearest neighbors over large datasets using MapReduce. Lu et al. construct a Voronoi-like diagram using some selected pivot objects. They then group the data points around the closest pivots and assign them to partitions, where searching can be done in parallel. The system is also designed for low dimensional datasets; it did not consider data with more than 30 dimensions. In contrast, in our experiments we used images and videos with up to 128 dimensions. Aly et al. propose a distributed system for image retrieval. A major drawback of this system is using a single root machine that directs all query points, which makes it a single point of failure as well as a bottleneck that could slow down the whole system. Our system does not use a central node, and thus it is more robust and scalable. The closest work to ours is the RankReduce system, which implements a distributed LSH (Locality Sensitive Hashing) index on a computing cluster using MapReduce. RankReduce maintains multiple hash This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2015.2389628, IEEE Transactions on Multimedia 7 tables over a distributed cluster, which requires storing multiple replicas of the datasets in hash tables. This incurs significant storage cost and it increases the number of I/O operations. In contrast, our system stores the dataset only once. We compare the proposed matching engine against RankReduce and we show that our system returns more accurate neighbors and it is more efficient

MACHINE LEARNING

Machine learning is a method of data analysis that automates analytical model building. It is a branch of Artificial Intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.

2.1 Evolution of machine learning

Because of new computing technologies, machine learning today is not like machine learning of the past. It was born from pattern recognition and the theory that computers can learn without being programmed to perform specific tasks; researchers interested in artificial intelligence wanted to see if computers could learn from data. The iterative aspect of machine learning is important because as models are exposed to new data, they can independently adapt.

They learn from previous computations to produce reliable,repeatable decisions and results.It's a science that's not new–but one that has gained fresh momentum.

While many machine learning algorithms have been around for a long time, the ability to automatically apply complex mathematical calculations to big data– over and over, faster and faster – is a recent development. Here are a few widely publicized examples of machine learning applications you may be familiar with:

- The heavily hyped, self-driving Google car? The essence of machine learning.
- Does online recommendation offer such as those from Amazon and Netflix? Machine learning applications for everyday life.
- Knowing what customers are saying about you on Twitter? Machine learning combined with linguistic rule creation.
- Fraud detection? One of the more obvious, important uses in our world today.

2.2 Why is machine learning important?

Resurging interest in machine learning is due to the same factors that have made data

mining and Bayesian analysis more popular than ever. Things like growing volumes and

varieties of available data, computational processing that is cheaper and more powerful, and affordable data storage.

All of these things mean it's possible to quickly and automatically produce models that can analyze bigger, more complex data and deliver faster, more accurate results – even on a very large scale. And by building precise models, an organization has a better chance of identifying profitable opportunities – or avoiding unknown risks.

2.3 Popular machine learning methods

Two of the most widely adopted machine learning methods are **supervised learning** and **unsupervised learning** – but there are also other methods of machine learning.

2.3.1 **Supervisedlearning**algorithmsaretrainedusinglabelledexamples,suchas**an** input wherethedesiredoutputisknown.Forexample,apieceofequipmentcouldhavedata points labelledeither“F”(failed)or“R”(runs).Thelearningalgorithmreceivesasetof inputs along with the corresponding correct outputs, and the algorithm learns by comparing its actual output with correct outputs to find errors. It then modifies the model accordingly. Through methods like classification, regression, prediction and gradient boosting, supervised learning uses patterns to predict the values of the label on additional unlabelled data. Supervised learning is commonly used in applications where historical data predicts likely future events. For example, it can anticipate when credit card transactions are likely to be fraudulent or which insurance customer is likely to file a claim.

2.3.2 Unsupervised learning is used against data that has no historical labels. The system is not told the "right answer." The algorithm must figure out what is being shown. The goal is to explore the data and find some structure within. Unsupervised learning works well on transactional data. For example, it can identify segments of customers with similar attributes who can then be treated similarly in marketing campaigns. Or it can find the main attributes that separate customer segments from each other. Popular techniques include self-organizing maps, nearest-neighbour

mapping, k-means clustering and singular value decomposition. These algorithms are also used to segment text topics, recommend items and identify data outliers.

2.3.3 Semi-supervised learning is used for the same applications as supervised learning. But it uses both labelled and unlabelled data for training – typically a small amount of labelled data with a large amount of unlabelled data (because unlabelled data is less expensive and takes less effort to acquire).

This type of learning can be used with methods such as classification, regression and prediction. Semi-supervised learning is useful when the cost associated with labelling is too high to allow for a fully labelled training process. Early examples of this include identifying a person's face on a web cam.

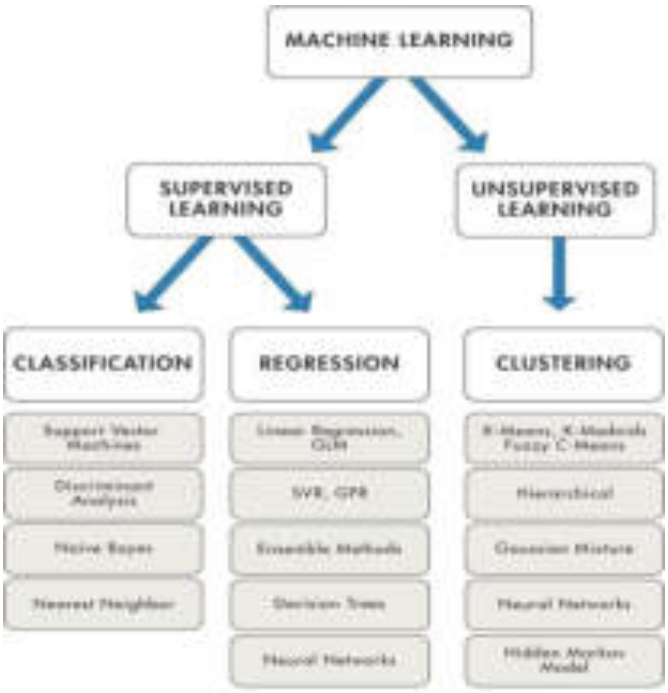


Fig: 2.3.1

2. SYSTEM ANALYSIS

EXISTING SYSTEM:

The problem of protecting various types of multimedia content has attracted significant attention from academia and industry. One approach to this problem is using watermarking, in which some distinctive information is embedded in the content itself and a method is used to search for this information in order to verify the authenticity of the content. Watermarking requires inserting watermarks in the multimedia objects before releasing them as well as mechanisms/systems to find objects and verify the existence of correct watermarks in them. Thus, this approach may not be suitable for already-released content without watermarks in them. The watermarking approach is more suitable for the somewhat controlled environments, such as distribution of multimedia content on DVDs or using special sites and custom players. Watermarking may not be effective for the rapidly increasing online videos, especially those uploaded to sites such as YouTube and played back by any video player. Watermarking is not the focus of this paper.

PROPOSED SYSTEM:

The focus of this paper is on the other approach for protecting multimedia content, which is content based copy detection (CBCD). In this approach, signatures (or fingerprints) are extracted from original objects. Signatures are also created from query (suspected) objects downloaded from online sites. Then, the similarity is computed between original and suspected objects to find potential copies. Many previous works proposed different methods for creating and matching signatures. These methods can be classified into four categories: spatial, temporal, color and transform-domain. Spatial signatures (particularly the block-based) are the most widely used. However, their weakness is the lack of resilience against large geometric transformations. Temporal and color signatures are less robust and can be used to enhance spatial signatures. Transform-domain signatures are computationally intensive and not widely used in practice. For more details, see surveys for audio fingerprinting and 2D video fingerprinting.

Unlike previous works, the contribution of this paper is to design a large-scale system to find copies that can be used for different types of multimedia content and can leverage multi cloud infrastructures to minimize the cost, expedite deployment, and dynamically scale up and

down. That is, we design our system such that previous content-based copy detection methods for creating and matching signatures can be implemented within our system. In addition to our cloud-based system, we propose a new method for 3D video fingerprinting, and a new design for the distributed matching engine. The works related to each of these components are summarized in the following subsections.

SYSTEM SPECIFICATION:

HARDWARE REQUIREMENTS:

- **System** : Pentium IV 2.4 GHz.
- **Hard Disk** : 40 GB.
- **Mouse** : Optical Mouse.
- **Ram** : 2 GB.

SOFTWARE REQUIREMENTS:

- **Operating system** : Windows 7 and above.
- **Coding Language** : Python.
- **Front-End** : Python.
- **Designing** :Html,css,javascript.

MODULES:

There are three modules can be divided here for this project they are listed as below

- User Verification
- File encryption with key
- Resource Transmit
- System analysis

MODULE DESCRIPTION:

- **USER VERIFICATION**

  Every user has to authentication in the form of login and to do that they need to register. Once user register the details and waiting for the approval from the administrator. Administrator is verifying the details and approves their request to login in to the system.

- **FILE ENCRYPTION WITH KEY**

  File is uploaded with the key. The uploaded file is encrypted using Advanced Encryption Standard with the user given key. The keys were used to decrypt as well so it is very important to maintain the secrecy. The encryption and decryption done with the authentication with the both users respectively through trusted authority.

- **RESOURCE TRANSMIT**

  The user has to request for file and they will get the key from admin through protected communications like ssl account. The registered email id only receives the mail. Then they can decrypt with the key they wanted. Then file can be shared to the particular requested user.

- **SYSTEM ANALYSIS**

  The proposed system is analysis with the help of certain graphs. To ensure the quality of the system graphs has been plot. The variety of charts like pie chart, bar chart, spline chart will gives you the brief and clear comparisons among the details.

ARCHITECTURE:

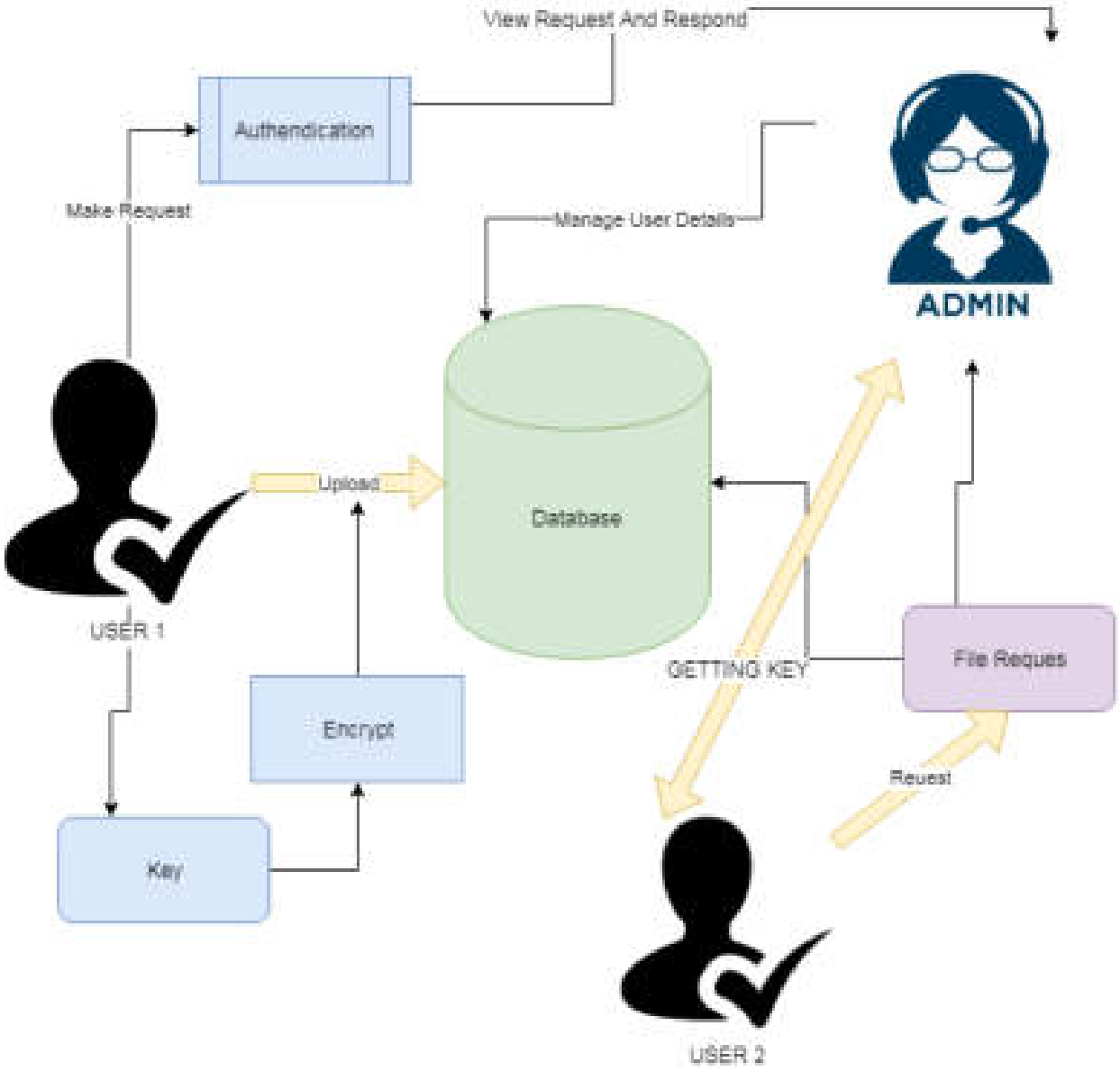


Fig: 3.1.1

SOFTWARE MODEL:

The Waterfall Model was the first Process Model to be introduced. It is also referred to as a **linear-sequential life cycle model**. It is very simple to understand and use. In a waterfall model, each phase must be completed before the next phase can begin and there is no overlapping in the phases.

The Waterfall model is the earliest SDLC approach that was used for software development.

The waterfall Model illustrates the software development process in a linear sequential flow. This means that any phase in the development process begins only if the previous phase is complete. In this waterfall model, the phases do not overlap.

Waterfall Model - Design

Waterfall approach was first SDLC Model to be used widely in Software Engineering to ensure success of the project. In "The Waterfall" approach, the whole process of software development is divided into separate phases. In this Waterfall model, typically, the outcome of one phase acts as the input for the next phase sequentially.

The following illustration is a representation of the different phases of the Waterfall Model.

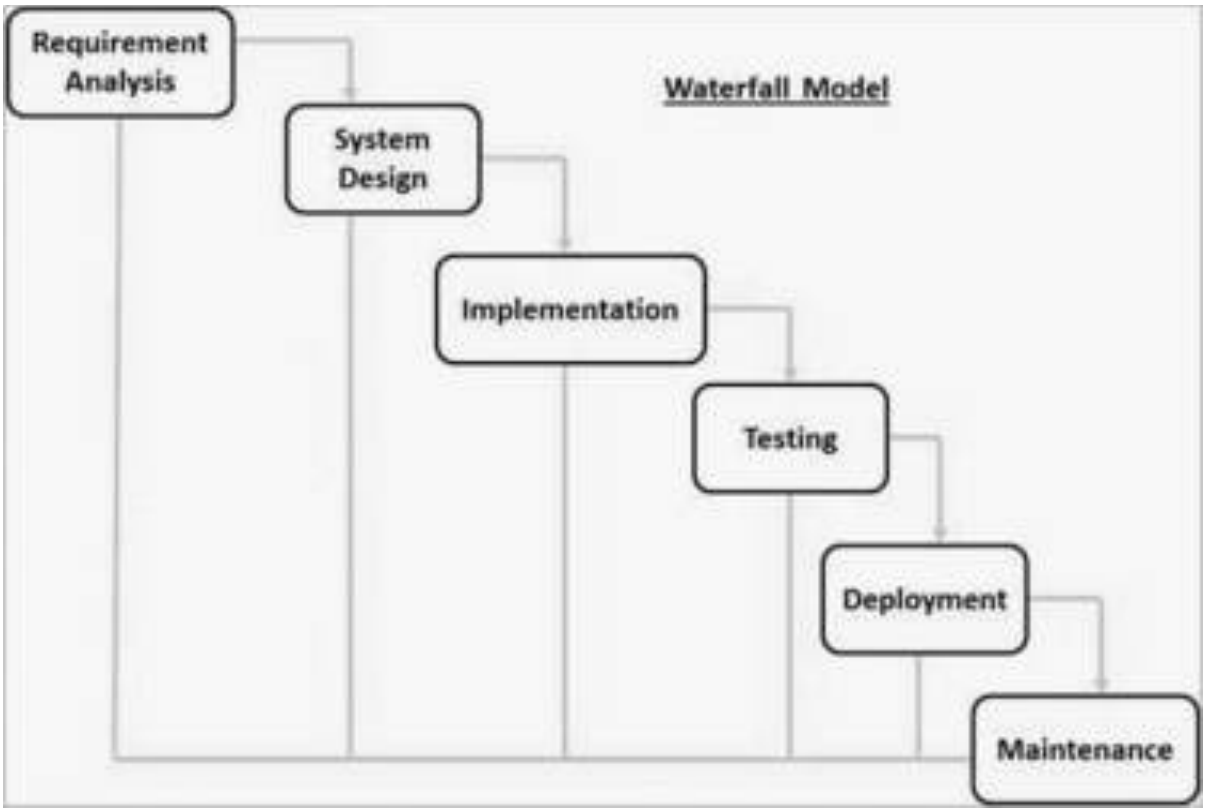


Fig: 3.1.2

The sequential phases in Waterfall model are −

- **Requirement Gathering and analysis** − All possible requirements of the system to be developed are captured in this phase and documented in a requirement specification document.
- **System Design** − The requirement specifications from first phase are studied in this phase and the system design is prepared. This system design helps in specifying hardware and system requirements and helps in defining the overall system architecture.
- **Implementation** − With inputs from the system design, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality, which is referred to as Unit Testing.
- **Integration and Testing** − All the units developed in the implementation phase are integrated into a system after testing of each unit. Post integration the entire system is tested for any faults and failures.

- **Deployment of system** − Once the functional and non-functional testing is done; the product is deployed in the customer environment or released into the market.
- **Maintenance** − There are some issues which come up in the client environment. To fix those issues, patches are released. Also to enhance the product some better versions are released. Maintenance is done to deliver these changes in the customer environment.

FEASIBILITY STUDY:

The first and foremost strategy for development of a project starts from the thought of designing a mail enabled platform for a small firm in which it is easy and convenient of sending and receiving messages, there is a search engine,address book and also including some entertaining games. When it is approved by the organization and our project guide the first activity, i.e., preliminary investigation begins. The activity has three parts:

- **Request Clarification**
- **Feasibility Study**
- **Request Approval**

REQUEST CLARIFICATION:

After the approval of the request to the organization and project guide, with an investigation being considered, the project request must be examined to determine precisely what the system requires.

Here our project is basically meant for users within the company whose systems can be interconnected by the Local Area Network (LAN). In today's busy schedule man need everything should be provided in a readymade manner. So taking into consideration of the vastly use of the net in day to day life, the corresponding development of the portal came into existence.

FEASIBILITY STUDY:

An important outcome of preliminary investigation is the determination that the system request is feasible. This is possible only if it is feasible within limited resource and time. The different

feasibilities that have to be analyzed are

- **Operational Feasibility**
- **Economic Feasibility**
- **Technical Feasibility**

Operational Feasibility:

Operational Feasibility deals with the study of prospects of the system to be developed. This system operationally eliminates all the tensions of the Admin and helps him in effectively tracking the project progress. This kind of automation will surely reduce the time and energy, which previously consumed in manual work. Based on the study, the system is proved to be operationally feasible.

Economic Feasibility:

Economic Feasibility or Cost-benefit is an assessment of the economic justification for a computer based project. As hardware was installed from the beginning & for lots of purposes thus the cost on project of hardware is low. Since the system is a network based, any number of employees connected to the LAN within that organization can use this tool from at anytime. The Virtual Private Network is to be developed using the existing resources of the organization. So the project is economically feasible.

Technical Feasibility:

According to Roger S. Pressman, Technical Feasibility is the assessment of the technical resources of the organization. The organization needs IBM compatible machines with a graphical web browser connected to the Internet and Intranet. The system is developed for platform Independent environment. Java Server Pages, JavaScript, HTML, SQL server and WebLogic Server are used to develop the system. The technical feasibility has been carried out. The system is technically feasible for development and can be developed with the existing facility

REQUEST APPROVAL

Not all request projects are desirable or feasible. Some organization receives so

many project requests from client users that only few of them are pursued. However, those projects that are both feasible and desirable should be put into schedule. After a project request is [i]approved, it cost, priority, completion time and personnel requirement is estimated and used to determine where to add it to any project list. Truly speaking, the approval of those above factors, development works can be launched.

ADVANCED ENCRYPTION STANDARD(AES) ALGORITHM

AES comprises three block ciphers: AES-128, AES-192 and AES-256. Each cipher encrypts and decrypts data in blocks of 128 bits using cryptographic keys of 128-, 192- and 256-bits, respectively. The Rijndael cipher was designed to accept additional block sizes and key lengths, but for AES, those functions were not adopted.

Symmetric (also known as secret-key) ciphers use the same key for encrypting and decrypting, so the sender and the receiver must both know -- and use -- the same secret key. All key lengths are deemed sufficient to protect classified information up to the "Secret" level with "Top Secret" information requiring either 192- or 256-bit key lengths. There are 10 rounds for 128-bit keys, 12 rounds for 192-bit keys and 14 rounds for 256-bit keys -- a round consists of several processing steps that include substitution, transposition and mixing of the input plaintext and transform it into the final output of cipher text.

The AES encryption algorithm defines a number of transformations that are to be performed on data stored in an array. The first step of the cipher is to put the data into an array; after which the cipher transformations are repeated over a number of encryption rounds. The number of rounds is determined by the key length, with 10 rounds for 128-bit keys, 12 rounds for 192-bit keys and 14 rounds for 256-bit keys.

The first transformation in the AES encryption cipher is substitution of data using a substitution table; the second transformation shifts data rows, the third mixes columns. The last

transformation is a simple exclusive or (XOR) operation performed on each column using a different part of the encryption key -- longer keys need more rounds to complete.

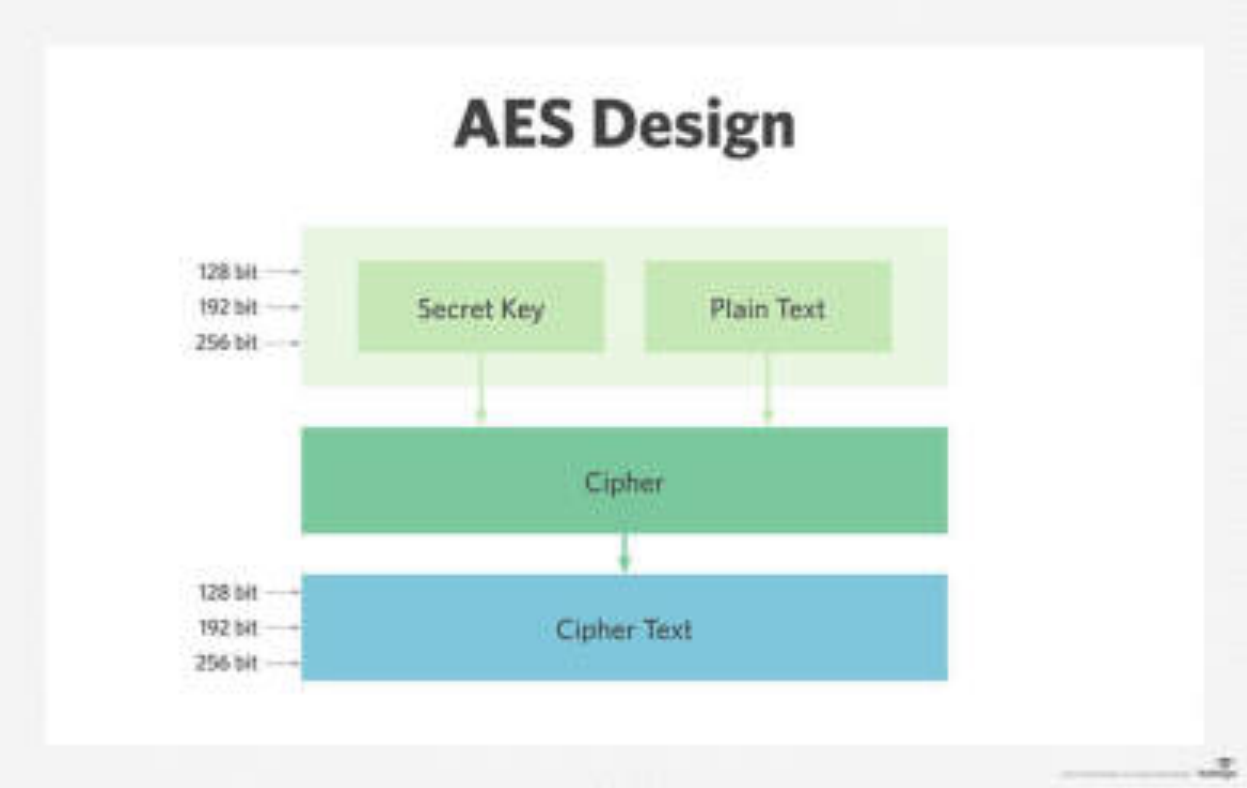


Fig: 3.1.3

3. SYSTEM DESIGN

Systems design is the process of defining elements of a system like modules, architecture, components and their interfaces and data for a system based on the specified requirements. It is the process of defining, developing and designing systems which satisfies the specific needs and requirements of a business or organization.

4.1 UML DIAGRAMS:

UML (Unified Modeling Language) is a general-purpose, graphical modeling language in the field of Software Engineering. UML is used to specify, visualize, construct, and document the artifacts (major elements) of the software system. It was initially developed by Grady Booch, Ivar Jacobson, and James Rumbaugh in 1994-95 at Rational software, and its further development was carried out through 1996. In 1997, it got adopted as a standard by the Object Management Group.

UML (Unified Modeling Language) is a general-purpose, graphical modeling language in the field of Software Engineering. UML is used to specify, visualize, construct, and document the artifacts (major elements) of the software system. It was initially developed by Grady Booch, Ivar Jacobson, and James Rumbaugh in 1994-95 at Rational software, and its further development was carried out through 1996. In 1997, it got adopted as a standard by the Object Management Group.

What is UML

The UML stands for Unified modeling language, is a standardized general-purpose visual modeling language in the field of Software Engineering. It is used for specifying, visualizing, constructing, and documenting the primary artifacts of the software system. It helps in designing and characterizing, especially those software systems that incorporate the concept of Object orientation. It describes the working of both the software and hardware systems.

The UML was developed in 1994-95 by Grady Booch, Ivar Jacobson, and James Rumbaugh at the Rational Software. In 1997, it got adopted as a standard by the Object Management Group (OMG).

The Object Management Group (OMG) is an association of several companies that controls the open standard UML. The OMG was established to build an open standard that mainly supports the interoperability of object-oriented systems. It is not restricted within the boundaries, but it can also be utilized for modeling the non-software systems. The OMG is best recognized for the Common Object Request Broker Architecture (CORBA) standards.

Goals of UML

- Since it is a general-purpose modeling language, it can be utilized by all the modelers.
- UML came into existence after the introduction of object-oriented concepts to systemize and consolidate the object-oriented development, due to the absence of standard methods at that time.
- The UML diagrams are made for business users, developers, ordinary people, or anyone who is looking forward to understand the system, such that the system can be software or non-software.
- Thus it can be concluded that the UML is a simple modeling approach that is used to model all the practical systems.

Characteristics of UML

The UML has the following features:

- It is a generalized modeling language.
- It is distinct from other programming languages like C++, Python, etc.
- It is interrelated to object-oriented analysis and design.
- It is used to visualize the workflow of the system.

- It is a pictorial language, used to generate powerful modeling artifacts.

Conceptual Modeling

Before moving ahead with the concept of UML, we should first understand the basics of the conceptual model.

A conceptual model is composed of several interrelated concepts. It makes it easy to understand the objects and how they interact with each other. This is the first step before drawing UML diagrams.

Following are some object-oriented concepts that are needed to begin with UML:

- **Object:** An object is a real world entity. There are many objects present within a single system. It is a fundamental building block of UML.
- **Class:** A class is a software blueprint for objects, which means that it defines the variables and methods common to all the objects of a particular type.
- **Abstraction:** Abstraction is the process of portraying the essential characteristics of an object to the users while hiding the irrelevant information. Basically, it is used to envision the functioning of an object.
- **Inheritance:** Inheritance is the process of deriving a new class from the existing ones.
- **Polymorphism:** It is a mechanism of representing objects having multiple forms used for different purposes.
- **Encapsulation:** It binds the data and the object together as a single unit, enabling tight coupling between them.

You can also create your own set of diagrams to meet your requirements. Diagrams are generally made in an incremental and iterative way.

There are two broad categories of diagrams and they are again divided into subcategories −

- Structural Diagrams

- Behavioral Diagrams

Structural Diagrams

The structural diagrams represent the static aspect of the system. These static aspects represent those parts of a diagram, which forms the main structure and are therefore stable.

These static parts are represented by classes, interfaces, objects, components, and nodes. The four structural diagrams are −

- Class diagram
- Object diagram
- Component diagram
- Deployment diagram

BEHAVIOURAL DIAGRAMS:

- Behavioral diagrams basically capture the dynamic aspect of a system. Dynamic aspect can be further described as the changing/moving parts of a system.
- UML has the following five types of behavioral diagrams −

➢ Use case diagram

➢ Sequence diagram

➢ Collaboration diagram

➢ State chart diagram

➢ Activity diagram

USECASE DIAGRAM FOR ADMIN:

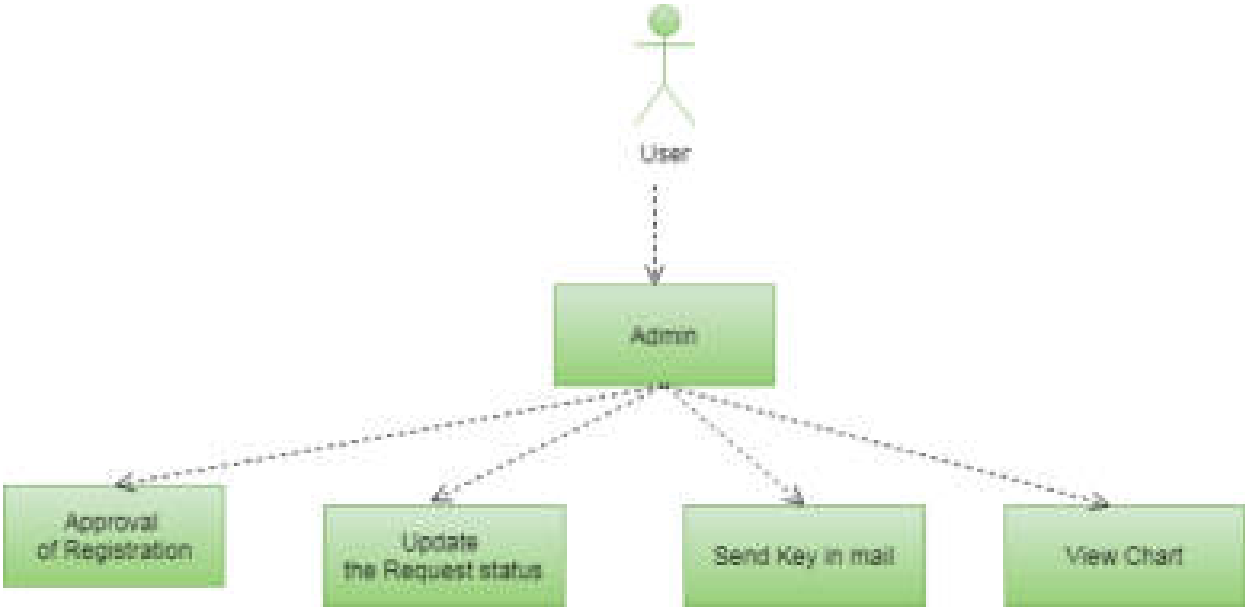


Fig: 4.1.1

USER:

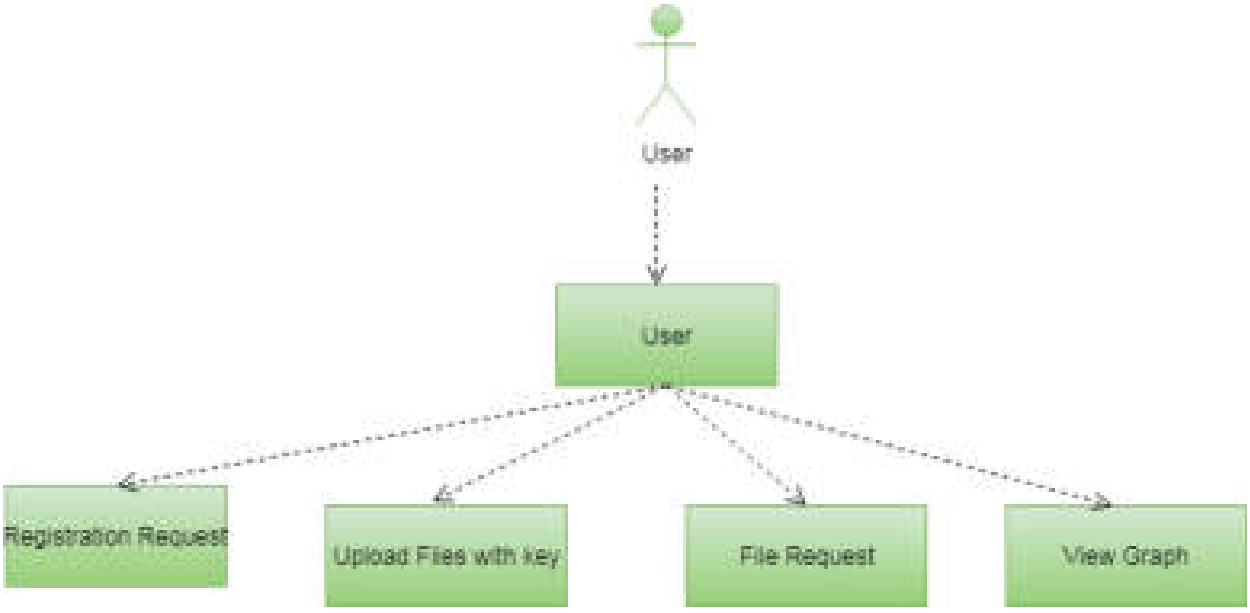


Fig: 4.1.2

CLASS DIAGRAM:

Model, objects are entities that combine state (i.e., data), behavior (i.e., procedures, or methods) and identity (unique existence among all other objects). The structure and behavior of an object are defined by a class, which is a definition, or blueprint, of all objects of a specific type. An object must be explicitly created based on a class and an object thus created is considered to be an instance of that class. An object is similar to a structure, with the addition of method pointers, member access control, and an implicit data member which locates instances of the class (i.e. actual objects of that class) in the class hierarchy (essential for runtime inheritance features)

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, and the relationships between the classes.

The class diagram is the main building block in the object oriented modeling. It is used both for general conceptual modeling of the semantics of the application, and for detailed modeling translating the models into programming code. The classes in a class diagram represent both the main objects and or interactions in the application and the objects to be programmed. In the class diagram these classes are represented with boxes which contain the two parts:

- The upper part holds the name of the class.
- The middle part contains the attributes of the class.
- The lower part contains the operations of the class.

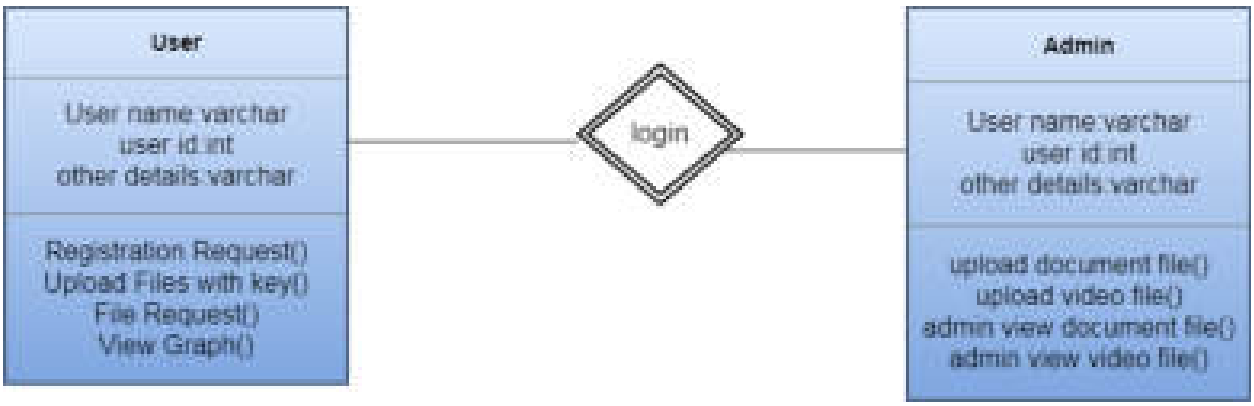


Fig: 4.1.3

Sequence Diagram

The sequence diagram represents the flow of messages in the system and is also termed as an event diagram. It helps in envisioning several dynamic scenarios. It portrays the communication between any two lifelines as a time-ordered sequence of events, such that these lifelines took part at the run time. In UML, the lifeline is represented by a vertical bar, whereas the message flow is represented by a vertical dotted line that extends across the bottom of the page. It incorporates the iterations as well as branching.

Purpose of a Sequence Diagram

1. To model high-level interaction among active objects within a system.
2. To model interaction among objects inside a collaboration realizing a use case.
3. It either models generic interactions or some certain instances of interaction.

SEQUENCE DIAGRAM FOR ADMIN

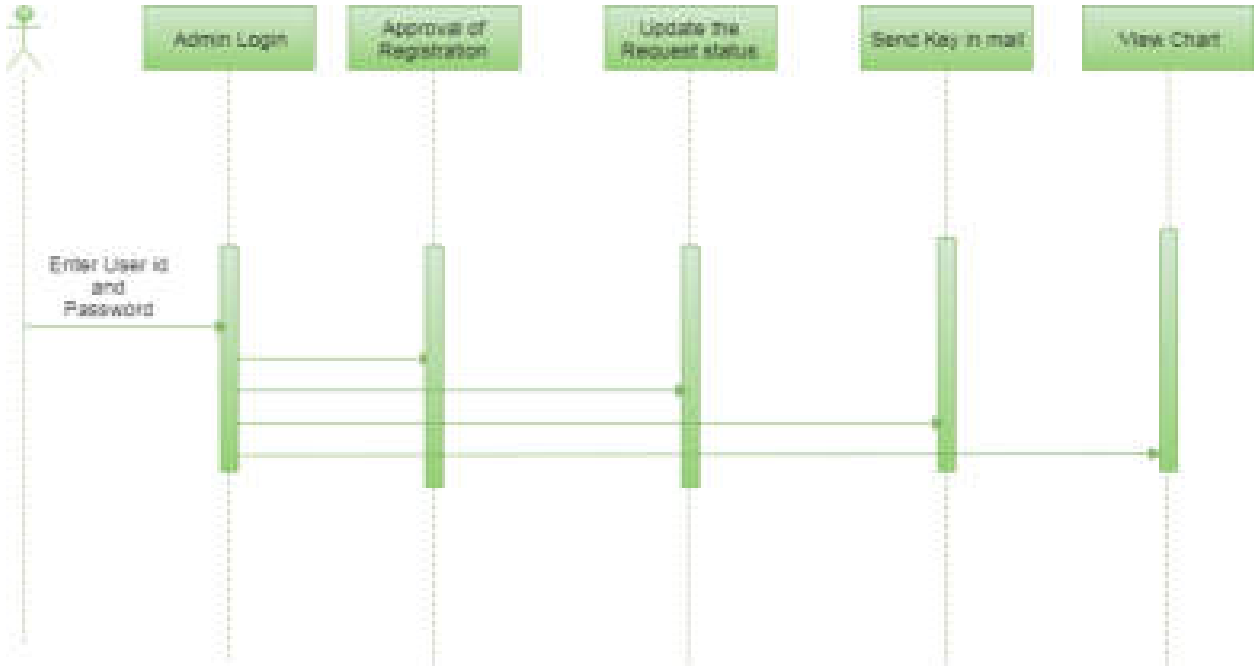


Fig: 4.1.4

USER

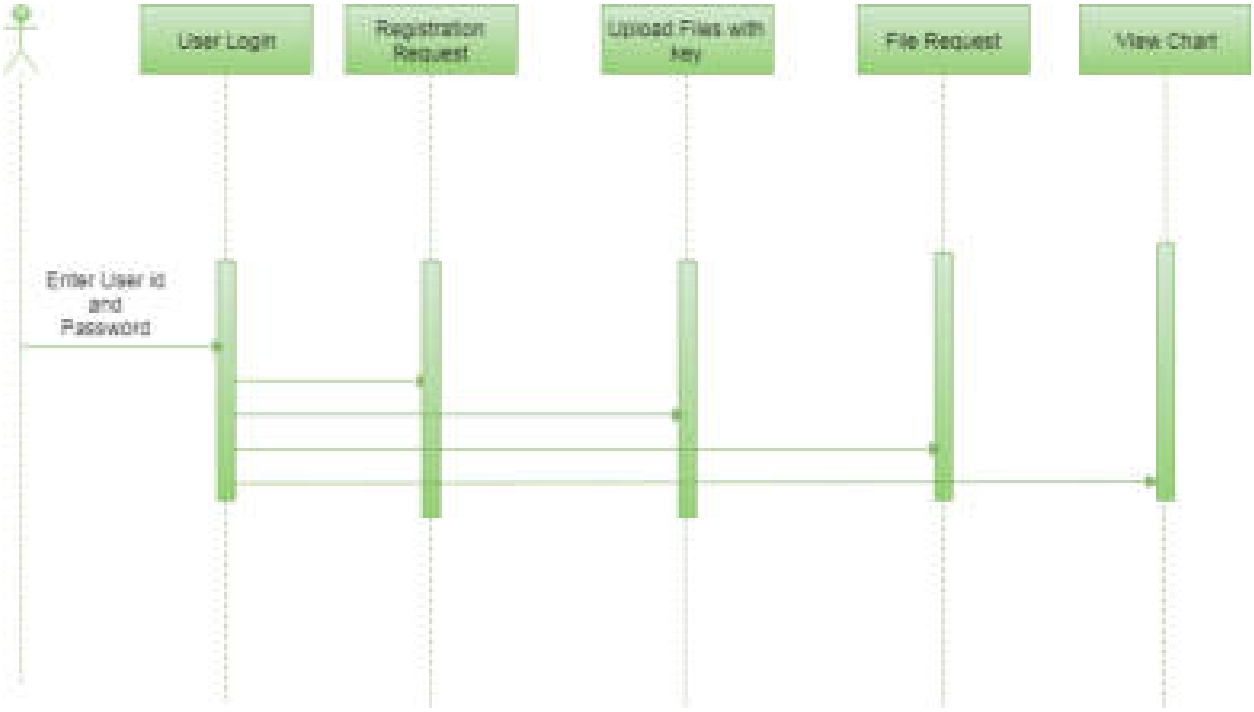


Fig: 4.1.5

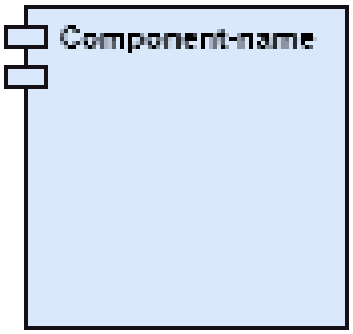
UML Component Diagram

A component diagram is used to break down a large object-oriented system into the smaller components, so as to make them more manageable. It models the physical view of a system such as executables, files, libraries, etc. that resides within the node.

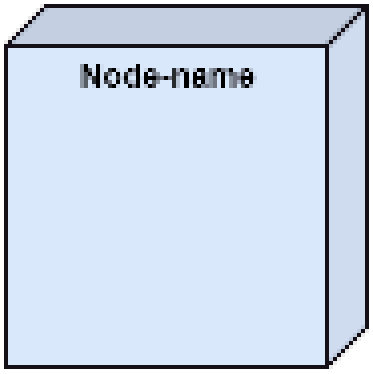
It visualizes the relationships as well as the organization between the components present in the system. It helps in forming an executable system. A component is a single unit of the system, which is replaceable and executable. The implementation details of a component are hidden, and it necessitates an interface to execute a function. It is like a black box whose behavior is explained by the provided and required interfaces.

Notation of a Component Diagram

a) A component



b) A node



Purpose of a Component Diagram

Since it is a special kind of a UML diagram, it holds distinct purposes. It describes all the individual components that are used to make the functionalities, but not the functionalities of the system. It visualizes the physical components inside the system. The components can be a library, packages, files, etc.

The component diagram also describes the static view of a system, which includes the organization of components at a particular instant. The collection of component diagrams represents a whole system.

The main purpose of the component diagram are enlisted below:

1. It envisions each component of a system.
2. It constructs the executable by incorporating forward and reverse engineering.

3. It depicts the relationships and organization of components.

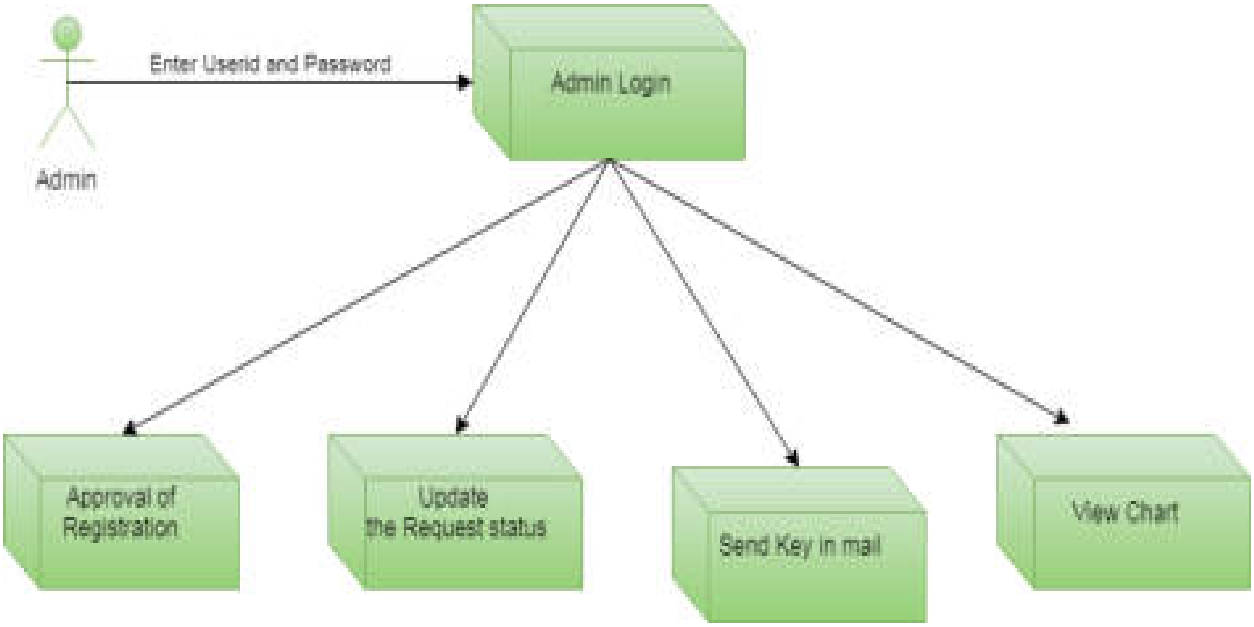


Fig: 4.1.6

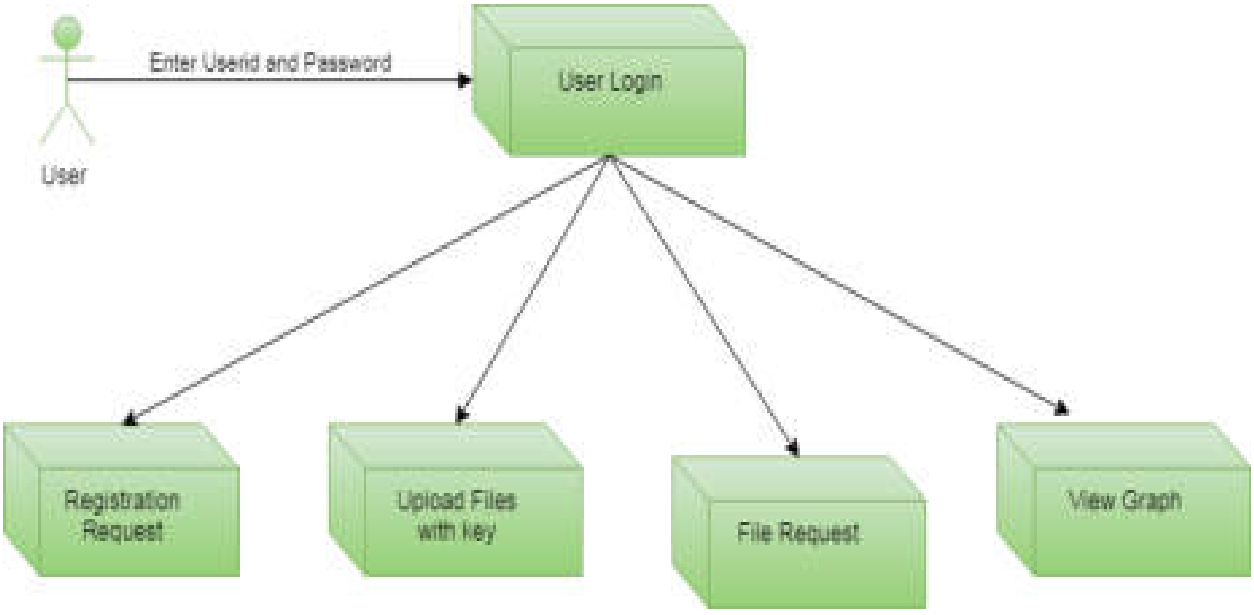


Fig: 4.1.7

UML Activity Diagram

In UML, the activity diagram is used to demonstrate the flow of control within the system rather than the implementation. It models the concurrent and sequential activities.

The activity diagram helps in envisioning the workflow from one activity to another. It put emphasis on the condition of flow and the order in which it occurs. The flow can be sequential, branched, or concurrent, and to deal with such kinds of flows, the activity diagram has come up with a fork, join, etc.

It is also termed as an object-oriented flowchart. It encompasses activities composed of a set of actions or operations that are applied to model the behavioral diagram.

Why use Activity Diagram?

An event is created as an activity diagram encompassing a group of nodes associated with edges. To model the behavior of activities, they can be attached to any modeling element. It can model use cases, classes, interfaces, components, and collaborations.

It mainly models processes and workflows. It envisions the dynamic behavior of the system as well as constructs a runnable system that incorporates forward and reverse engineering. It does not include the message part, which means message flow is not represented in an activity diagram.

It is the same as that of a flowchart but not exactly a flowchart itself. It is used to depict the flow between several activities.

ADMIN

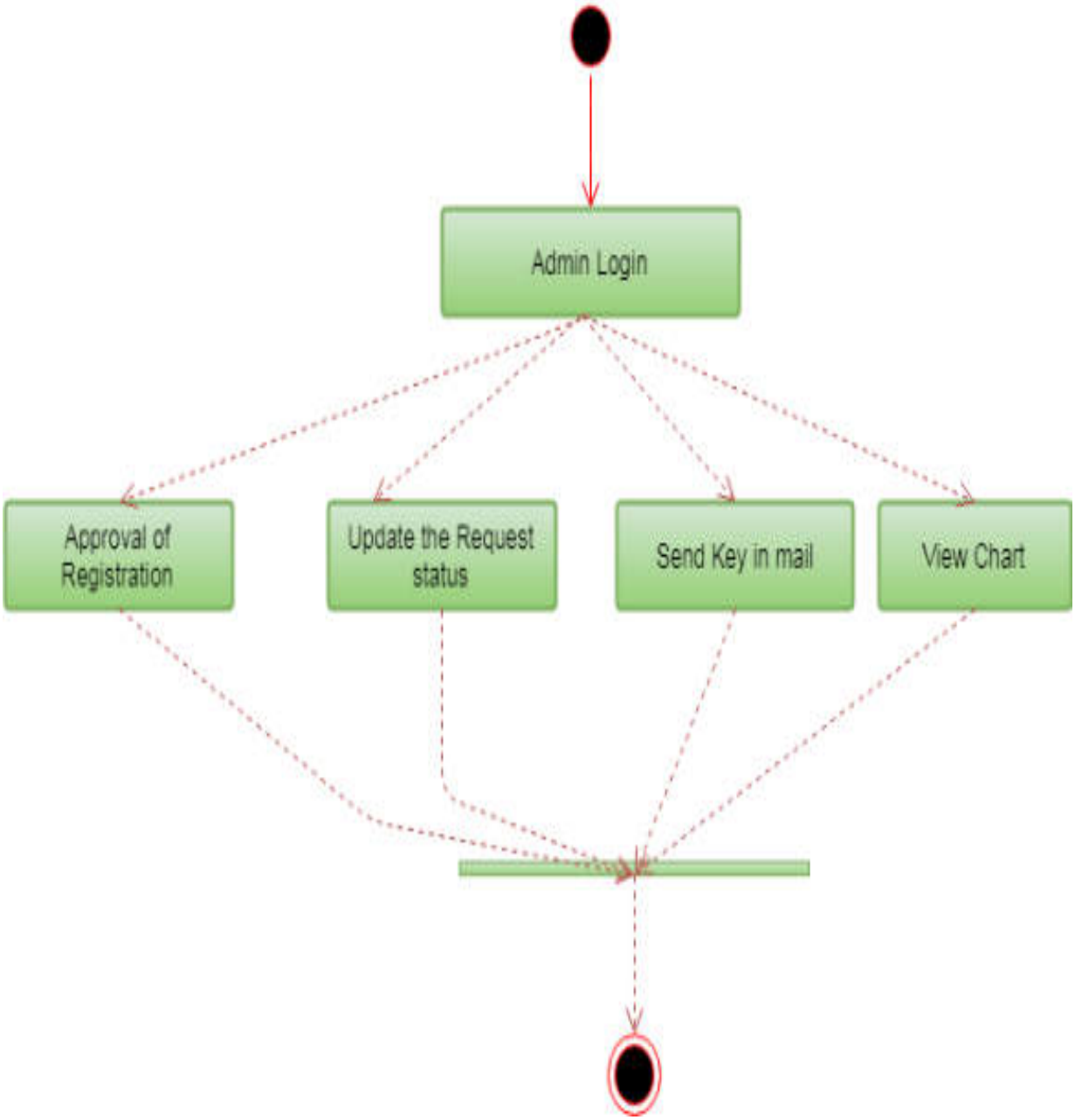


Fig: 4.1.8

USER

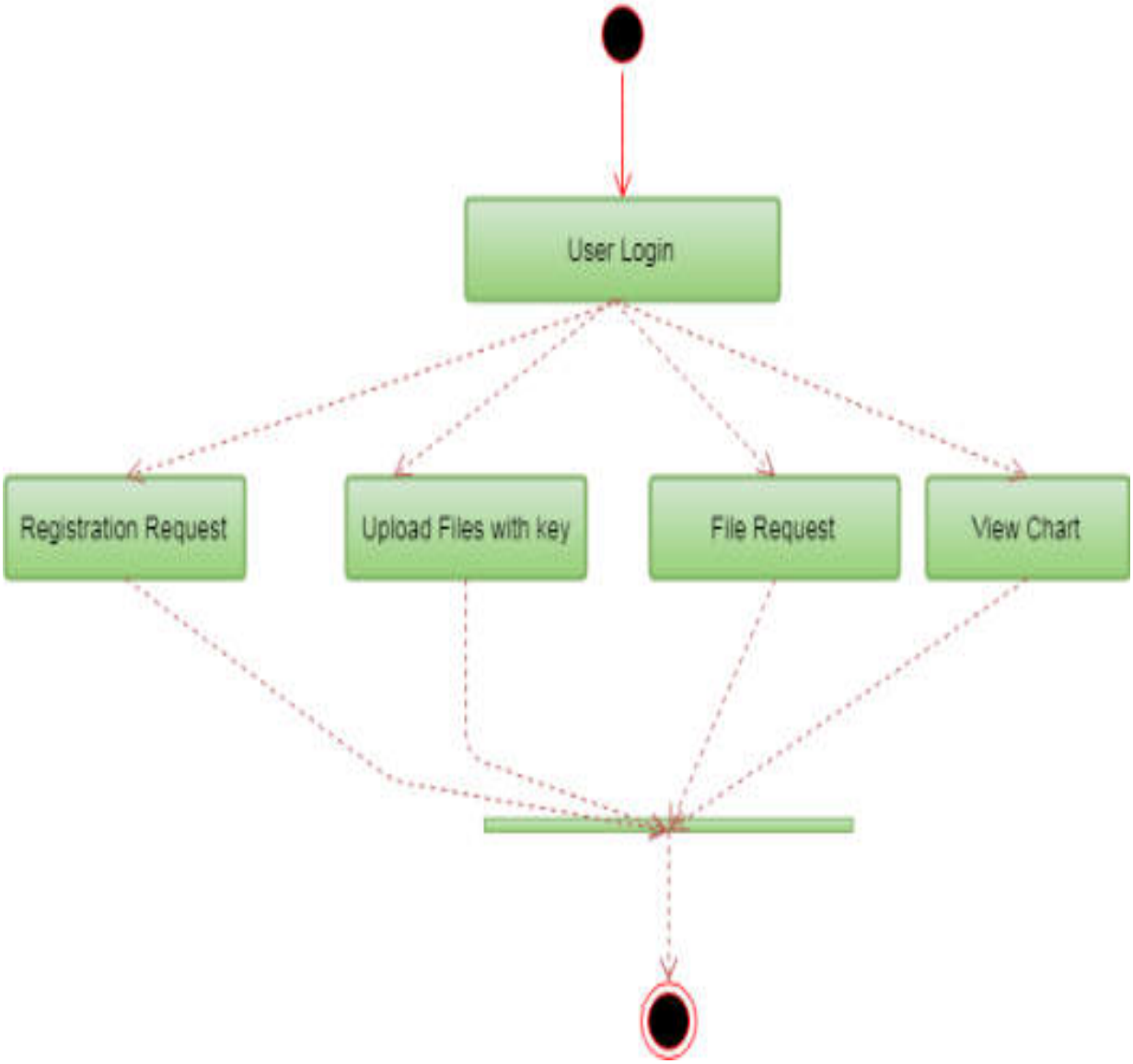

User Login
Registration Request
Upload Files with key
File Request
View Chart

Fig: 4.1.9

5. IMPLEMENTATIONS

5.1 TECHNOLOGIES:

Python:

- Python is a general purpose, dynamic, high level and interpreted programming language. It supports Object Oriented programming approach to develop applications. It is simple and easy to learn and provides lots of high-level data structures.
- Python is easy to learn yet powerful and versatile scripting language which makes it attractive for Application Development.
- Python's syntax and dynamic typing with its interpreted nature, makes it an ideal language for scripting and rapid application development.
- Python supports multiple programming pattern, including object oriented, imperative and functional or procedural programming styles.
- Python is not intended to work on special area such as web programming. That is why it is known as multipurpose because it can be used with web, enterprise, 3D CAD etc.
- We don't need to use data types to declare variable because it is dynamically typed so we can write a=10 to assign an integer value in an integer variable.
- Python makes the development and debugging fast because there is no compilation step included in python development and edit-test-debug cycle is very fast.

Python Features:

- Python provides lots of features that are listed below.

  Easy to Learn and Use: Python is easy to learn and use. It is developer-friendly and high level programming language.

- **Expensive Language**: Python language is more expressive means that it is more

understandable and readable.

- **Interpreted Language**: Python is an interpreted language i.e. interpreter executes the code line by line at a time. This makes debugging easy and thus suitable for beginners.
- **Cross-platform Language**: Python can run equally on different platforms such as Windows, Linux, Unix and Macintosh etc. So, we can say that Python is a portable language.
- **Free and Open Source**: Python language is freely available at official web address. The source-code is also available. Therefore it is open source.
- **Object-Oriented Language**: Python supports object oriented language and concepts of classes and objects come into existence.
- **Extensible**: It implies that other languages such as C/C++ can be used to compile the code and thus it can be used further in our python code.
- **Large Standard Library**: Python has a large and broad library and provides rich set of module and functions for rapid application development.
- **GUI Programming Support**: Graphical user interfaces can be developed using Python.
- **Integrated**: It can be easily integrated with languages like C, C++, JAVA etc.

Python applications:

Python is known for its general purpose nature that makes it applicable in almost each domain of software development. Python as a whole can be used in any sphere of development.

Here, we are specifying applications areas where python can be applied.

- Web Applications:

We can use Python to develop web applications. It provides libraries to handle internet protocols such as HTML and XML, JSON, Email processing, request, beautiful Soup, Feed parser etc. It also provides Frameworks such as Django, Pyramid, Flasketc to design and develop web based applications. Some important developments are: PythonWikiEngines, Pocoo, PythonBlogSoftware etc.

- Desktop GUI Applications:

  Python provides Tk GUI library to develop user interface in python based application. Some other useful toolkits wxWidgets, Kivy, pyqt that are useable on several platforms. The Kivy is popular for writing multi touch applications.

- **Software Development:**

  Python is helpful for software development process. It works as a support language and can be used for build control and management, testing etc.

- Scientific and Numeric:

  Python is popular and widely used in scientific and numeric computing. Some useful library and package are SciPy, Pandas, IPython etc. SciPy is group of packages of engineering, science and mathematics.

- Business Applications:

  Python is used to build business applications like ERP and e-commerce systems. Tryton is a high level application platform.

- Console Based Application:

  We can use Python to develop console based applications. For example: IPython.

- Audio or Video based Applications:

  Python is awesome to perform multiple tasks and can be used to develop multimedia applications. Some of real applications are: TimPlayer, cplay etc.

- 3D CAD Applications:

To create CAD application Fandango is a real application which provides full features of CAD.

- Enterprise Applications:

  Python can be used to create applications which can be used within an Enterprise or an Organization. Some real time applications are: OpenErp, Tryton, Picalo etc.

- Applications for Images:

  Using Python several application can be developed for image. Applications developed are: VPython, Gogh, imgSeeketc.

  HOW TO INSTALL PYTHON (ENVIRONMENT SET UP):

Installation on Windows

Visit the link https://www.python.org/downloads/to download the latest release of Python. In this process, we will install Python 3.6.7 on our Windows operating system.

Double-click the executable file which is downloaded; the following window will open. Select Customize installation and proceed.

Fig: 5.1.1 Python Installation step-1

The following window shows all the optional features. All the features need to be installed and are checked by default; we need to click next to continue.

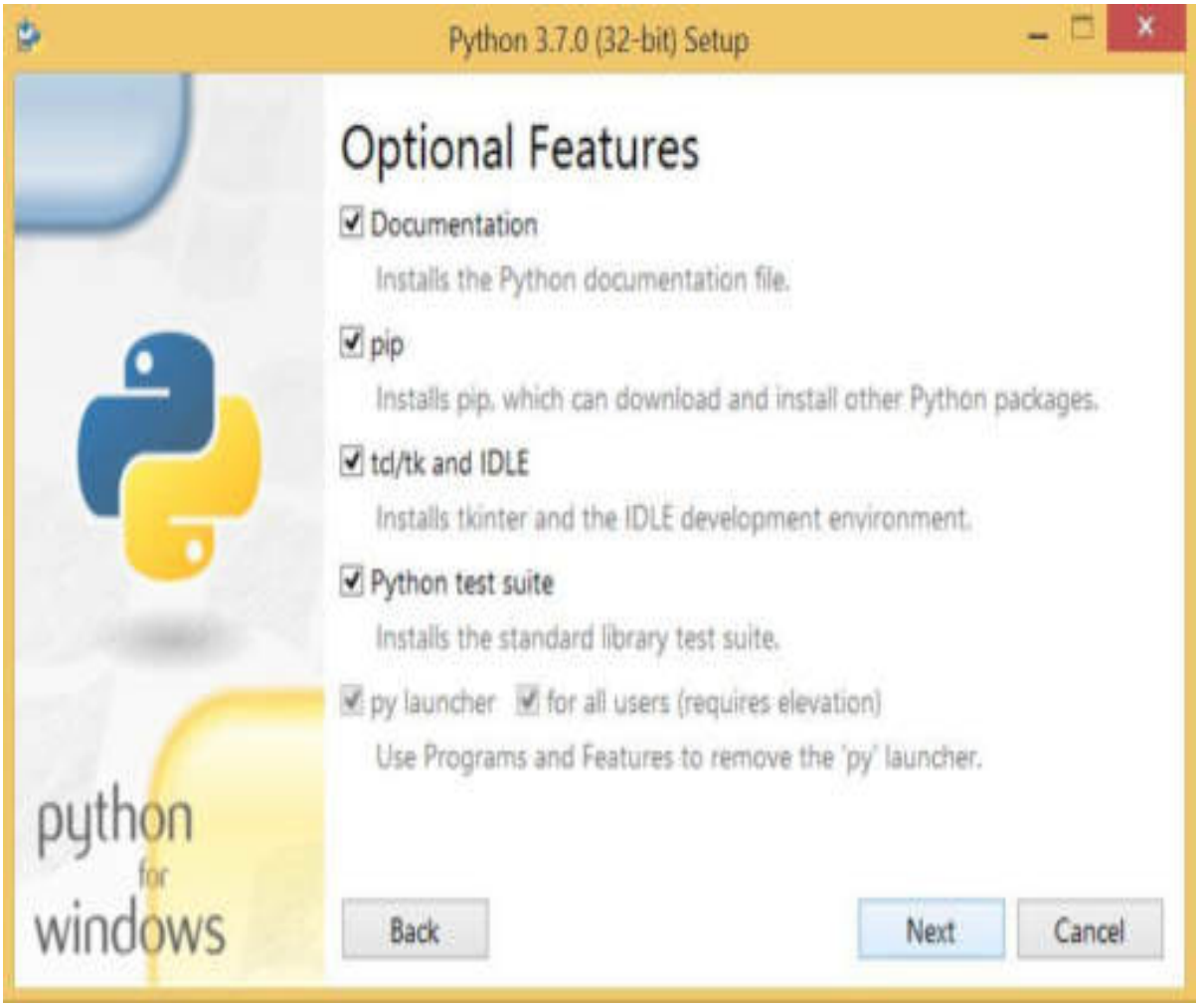


Fig 5.1.2: Python Installation step-2

The following window shows a list of advanced options. Check all the options which you want to install and click next. Here, we must notice that the first check-box (install for all users) must be checked.

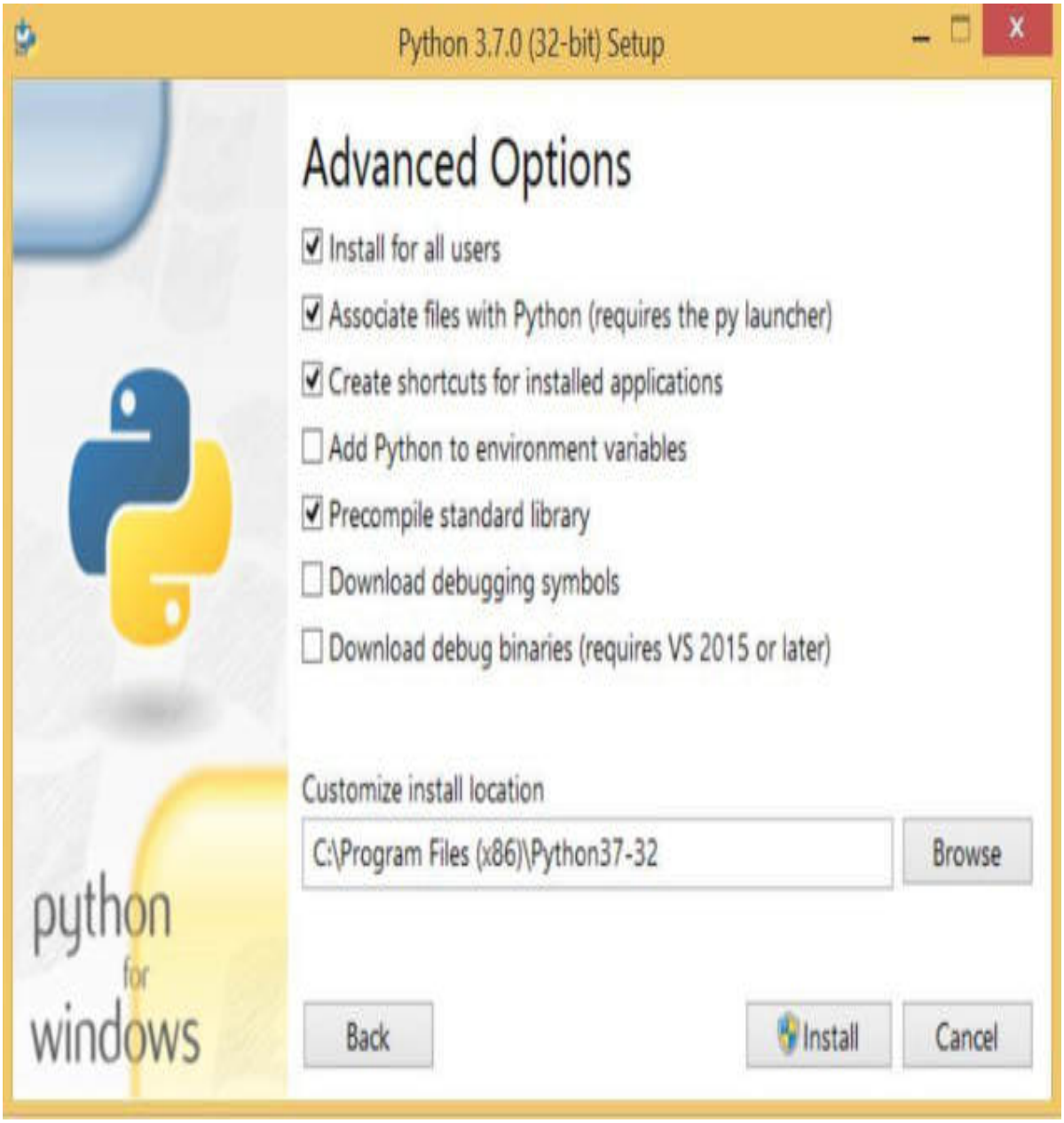


Fig 5.1.3: Python Installation step-3

Now, we are ready to install python-3.6.7.

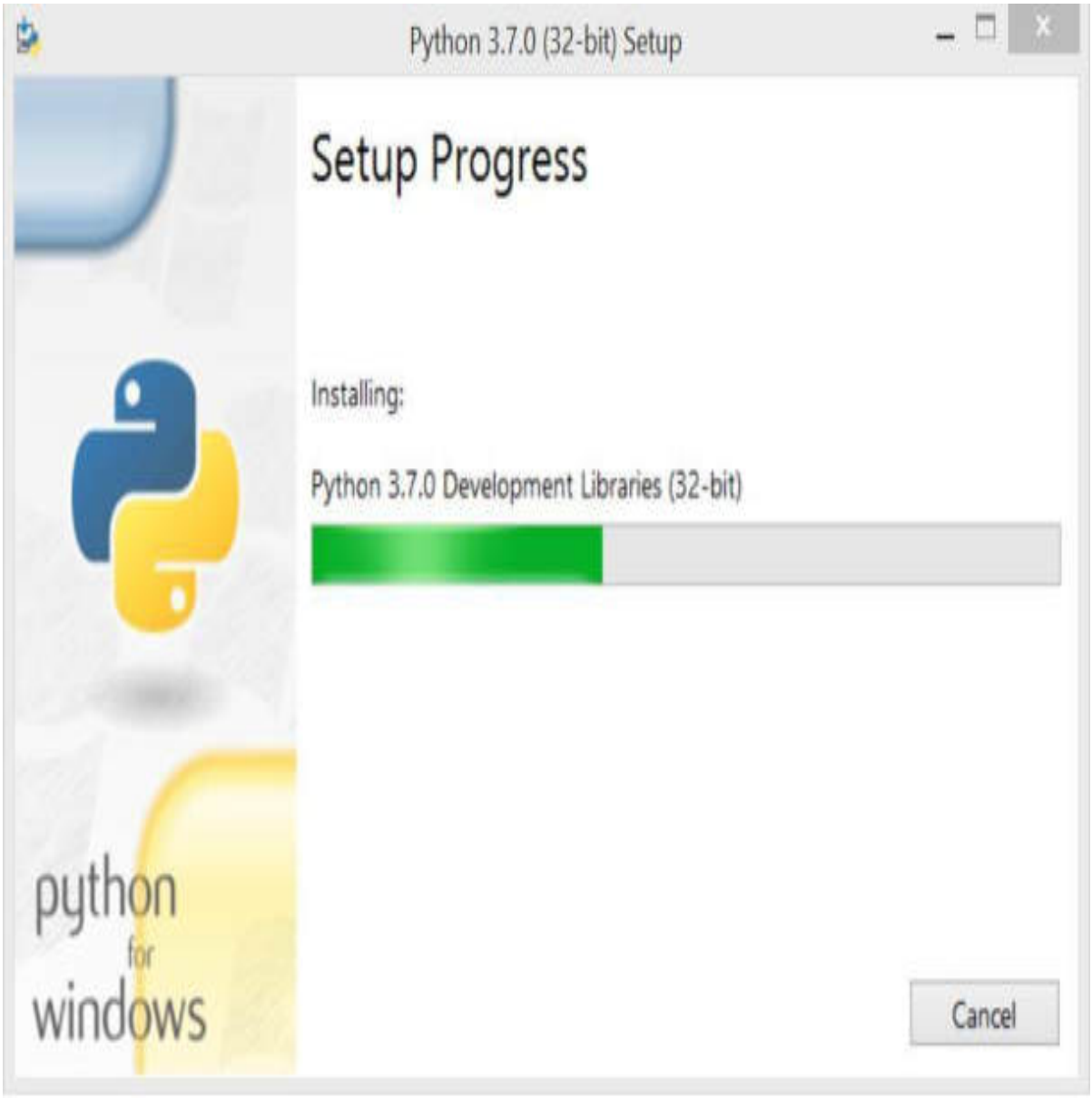


Fig 5.1.4: Python Installation step-4

Now, try to run python on the command prompt. Type the command python in case of python2 or python3 in case of python3. It will show an error as given in the below image. It is because we haven't set thepath.

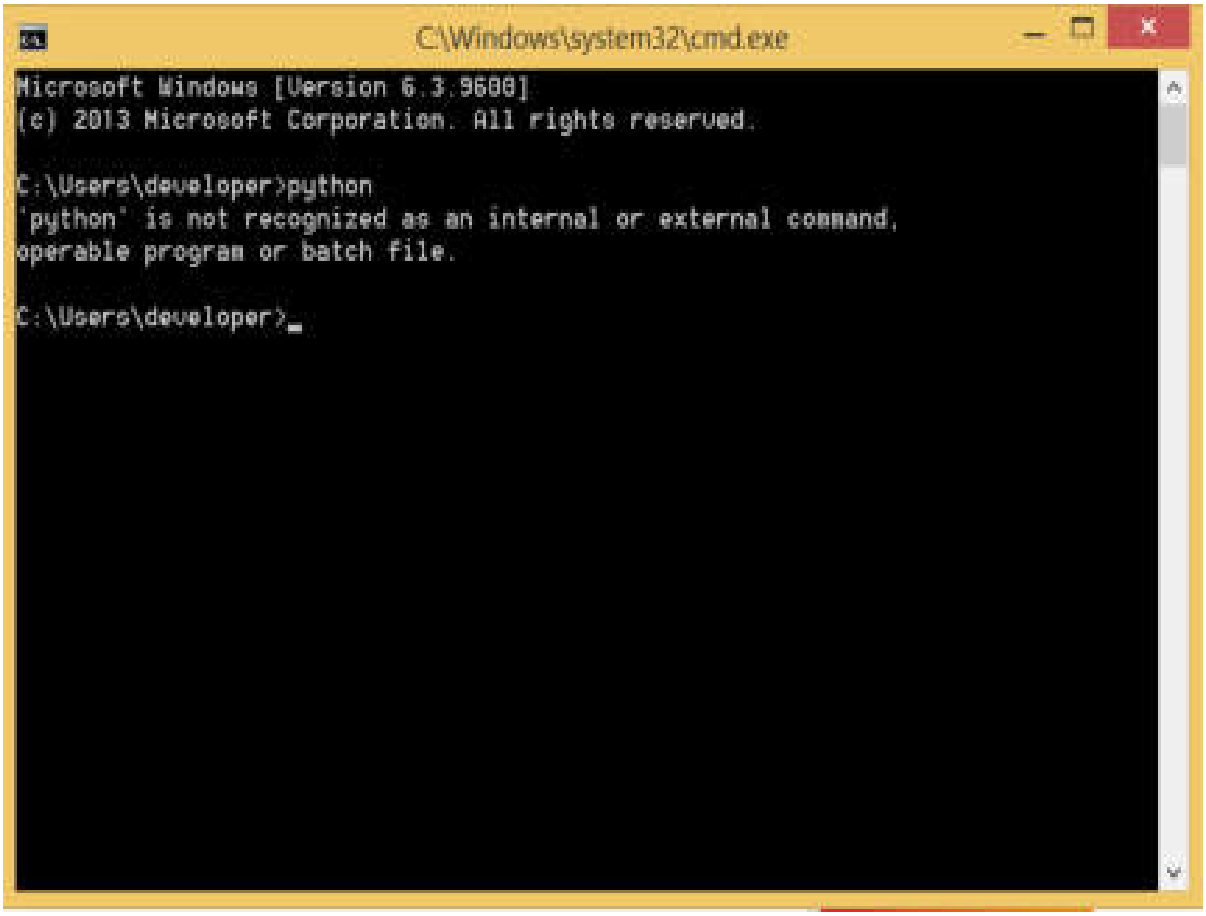


Fig 5.1.5: Python Installation step-5

To set the path of python, we need to the right click on "my computer" and go to Properties→ Advanced → Environment Variables

PYCHARM

PyCharm is the most popular IDE for Python, and includes great features such as excellent code completion and inspection with advanced debugger and support for web programming and various frameworks. PyCharm is created by Czech company, Jet brains which focusses on creating integrated development environment for various web development languages like JavaScript and PHP.

PyCharm is the most popular IDE used for Python scripting language. This chapter will give you an introduction to PyCharm and explains its features.

PyCharm offers some of the best features to its users and developers in the following aspects −

- Code completion and inspection
- Advanced debugging
- Support for web programming and frameworks such as Django and Flask

Features of PyCharm

Besides, a developer will find PyCharm comfortable to work with because of the features mentioned below −

- **Code Completion:** PyCharm enables smoother code completion whether it is for built in or for an external package.
- **SQLAlchemy as Debugger :**You can set a breakpoint, pause in the debugger and can see the SQL representation of the user expression for SQL Language code.
- **Git Visualization in Editor:** When coding in Python, queries are normal for a

developer. You can check the last commit easily in PyCharm as it has the blue sections that can define the difference between the last commit and the current one.

- **Code Coverage in Editor:** You can run .py files outside PyCharm Editor as well marking it as code coverage details elsewhere in the project tree, in the summary section etc.

- **Code Coverage in Editor:** You can run .py files outside PyCharm Editor as well marking it as code coverage details elsewhere in the project tree, in the summary section etc.

- **Package Management:** All the installed packages are displayed with proper visual representation. This includes list of installed packages and the ability to search and add new packages.

- **Local History** is always keeping track of the changes in a way that complements like Git. Local history in PyCharm gives complete details of what is needed to rollback and what is to be added.

- **Refactoring** is the process of renaming one or more files at a time and PyCharm includes various shortcuts for a smooth refactoring process.

- **Wamp Server:** WAMPs are packages of independently-created programs installed on computers that use a Microsoft Windows operating system. Apache is a web server. MySQL is an open-source database. PHP is a scripting language that can manipulate information held in a database and generate web pages dynamically each time content is requested by a browser. Other programs may also be included in a package, such as php My Admin which provides a graphical user interface for the MySQL database manager, or the alternative scripting languages Python or Perl.

Steps Involved

You will have to follow the steps given below to install PyCharm on your system. These steps show the installation procedure starting from downloading the PyCharm package from its official website to creating a new project.

Step 1

Download the required package or executable from the official website of PyCharm https://www.jetbrains.com/pycharm/download/#section=windowsHere you will observe two versions of package for Windows as shown in the screenshot given below −

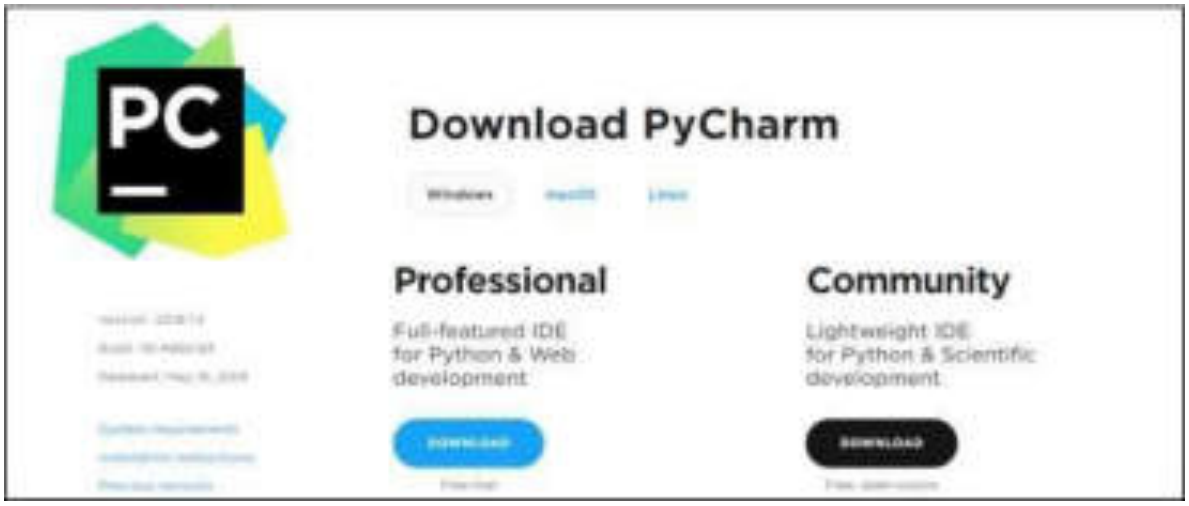


FIG:5.1.6

Note that the professional package involves all the advanced features and comes with free trial for few days and the user has to buy a licensed key for activation beyond the trial period. Community package is for free and can be downloaded and installed as and when required. It includes all the basic features needed for installation. Note that we will continue with community package throughout this tutorial.

Step 2

Download the community package (executable file) onto your system and mention a destination folder as shown below −

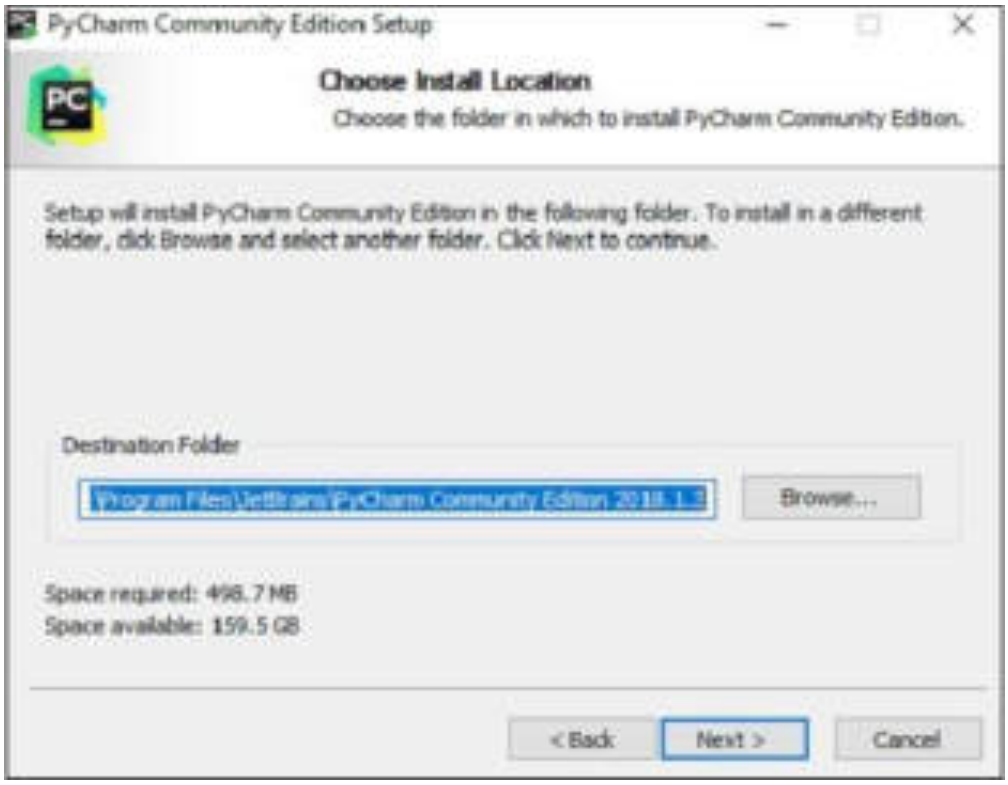


FIG:5.1.7

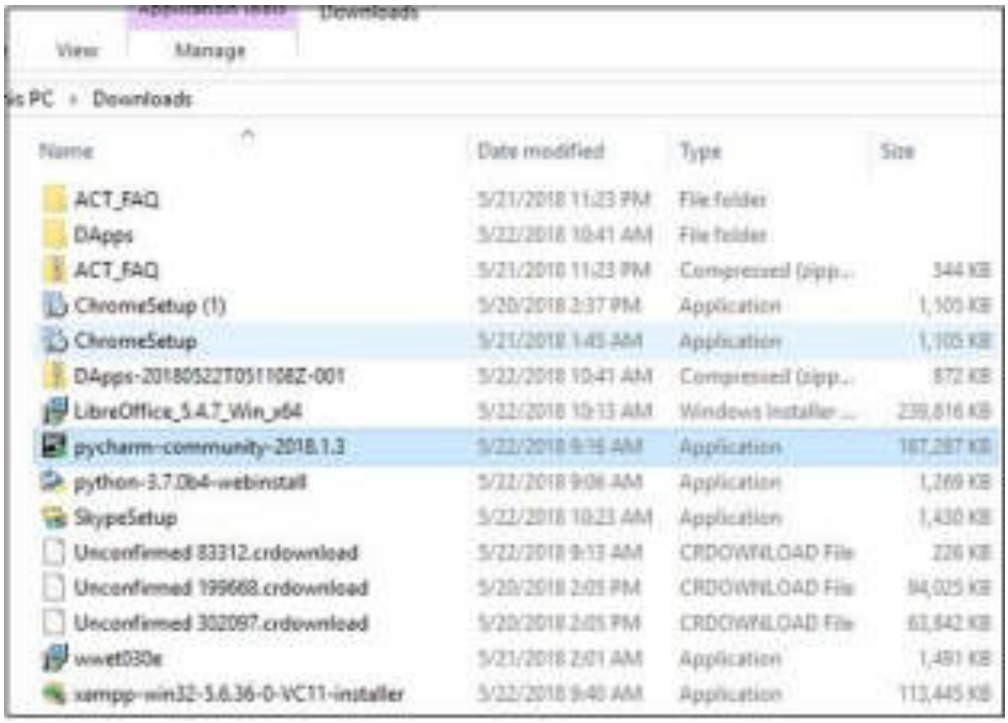


FIG:5.1.8

Step 3

Now, begin the installation procedure similar to any other software package.

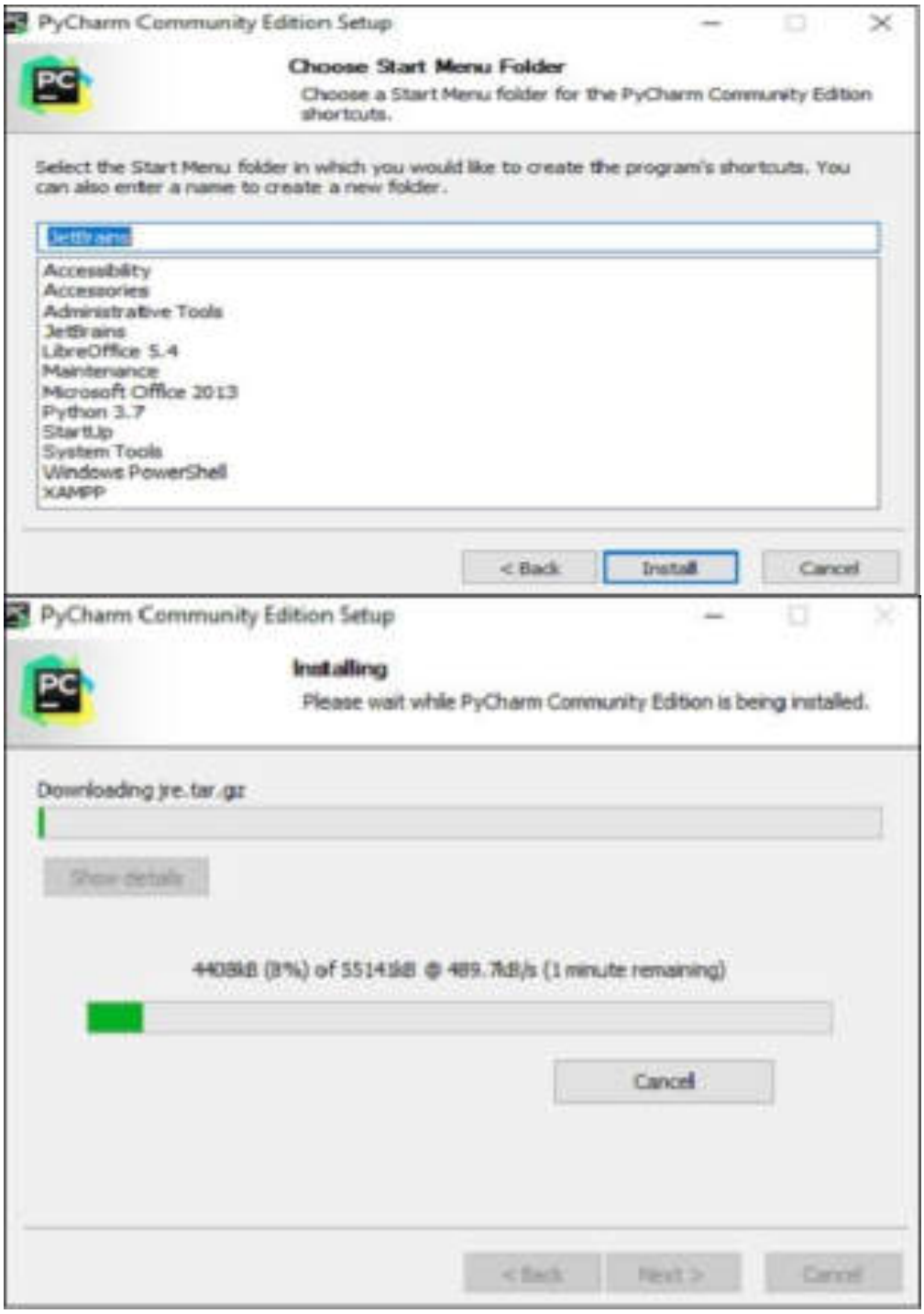


FIG:5.1.9,5.1.10

Step 4

Once the installation is successful, PyCharm asks you to import settings of the existing package if any.

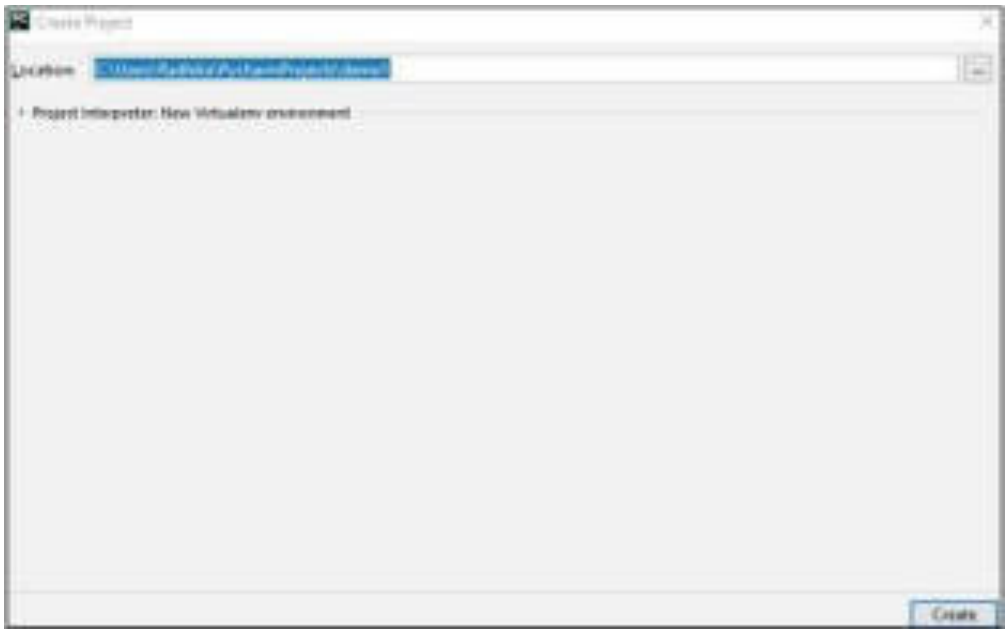


FIG:5.1.11,5.1.12,5.1.13

This helps in creating a new project of Python where you can work from the scratch. Note that unlike other IDEs, PyCharm only focusses on working with projects of Python scripting language.

DJANGO

Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. It’s free and open source.

Django's primary goal is to ease the creation of complex, database-driven websites. Django emphasizes reusability and "pluggability" of components, rapid development, and the principle of don't repeat yourself. Python is used throughout, even for settings files and data models.

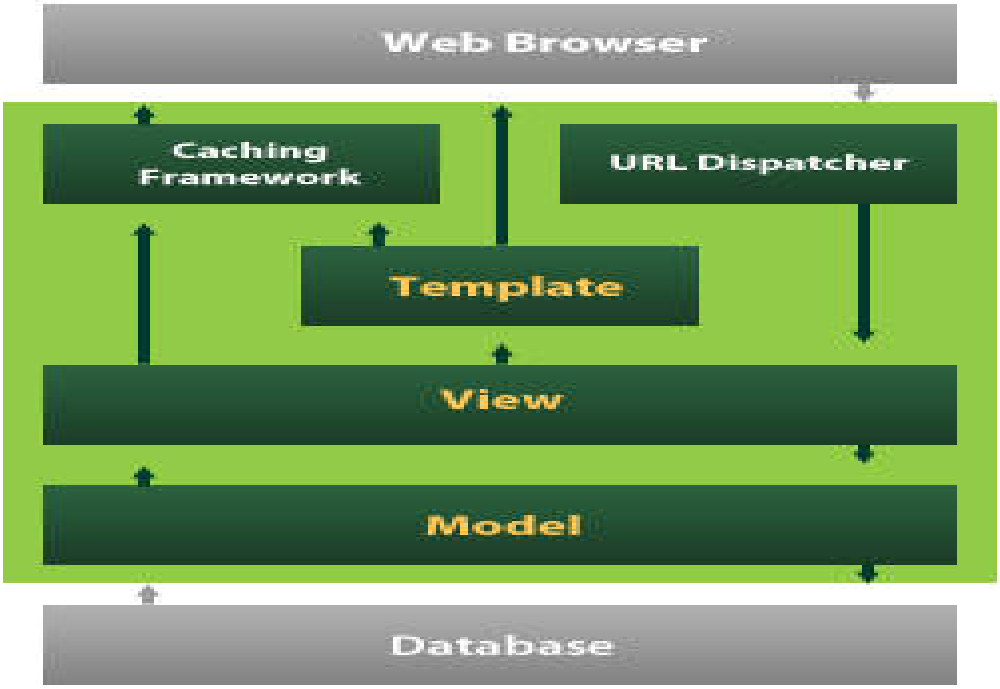


FIG:5.1.14

Django also provides an optional administrative create, read, update and delete interface that is generated dynamically through introspection and configured via admin models

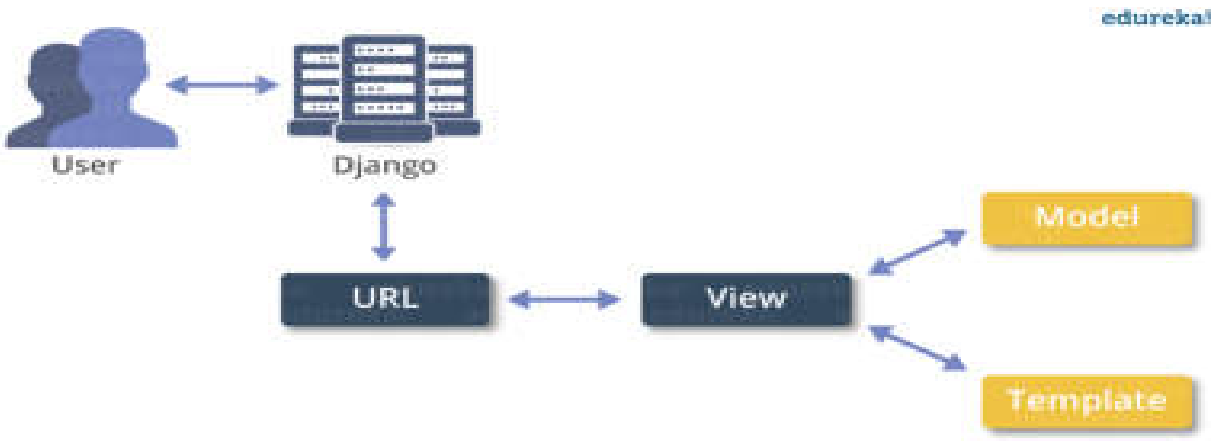


FIG:5.1.15

LIBRARIES:

Matplotlib:

- Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms.
- Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter notebook, web application servers, and four graphical user interface toolkits.
- Matplotlib tries to make easy things easy and hard things possible.
- You can generate plots, histograms, power spectra, bar charts, error charts, scatter plots, etc., with just a few lines of code.
- For simple plotting the pyplot module provides a MATLAB-like interface, particularly when combined with IPython.
- For the power user, you have full control of line styles, font properties, axes properties, etc, via an object oriented interface or via a set of functions familiar to MATLAB users.

Numpy:

- Numpy is the fundamental package for scientific computing with Python. It contains among other things:
  - a powerful N-dimensional array object
  - sophisticated (broadcasting) functions
  - tools for integrating C/C++ and Fortran code
  - useful linear algebra, Fourier transform, and random number capabilities
  - Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined. This allows NumPy to seamlessly and speedily integrate with a wide variety of databases.
  - NumPy is licensed under the BSD license, enabling reuse with few restrictions.

Pandas:

- **History of development**
- In 2008, **pandas** development began at AQR Capital Management. By the end of 2009 it had been open sourced, and is actively supported today by a community of like-minded individuals around the world who contribute their valuable time and energy to help make open source **pandas** possible.
- Since 2015, **pandas** is a Num FOCUS sponsored project. This will help ensure the success of development of **pandas** as a world-class open-source project.
- Timeline
- **2008**: Development of **pandas** started
- **2009**: **pandas** become open source
- **2012**: First edition of **Python for Data Analysis** is published **data alignment** and integrated handling of **missing data**: gain automatic label-based alignment in computations and easily manipulate messy data into an orderly form.
- Flexible **reshaping** and pivoting of datasets.
- Intelligent label-based **slicing**, **fancy indexing**, and **subsetting**of large datasets.
- Columns can be inserted and deleted from data structures for **size mutability**.
- Aggregating or transforming data with a powerful **group by** engine allowing split-apply-combine operations on datasets.
- High performance **merging and joining** of data sets.
- **Hierarchical axis indexing** provides an intuitive way of working with high-dimensional data in a lower-dimensional data structure.
- **Time series**-functionality: date range generation and frequency conversion, moving window statistics, date shifting and lagging. Even create domain-specific time offsets and join time series without losing data.
- **2015**: **pandas** becomes a Num FOCUS sponsored project
- **2018**: First in-person core developer sprint

- Library Highlights
- A fast and efficient **Data Frame** object for data manipulation with integrated indexing.
- Tools for **reading and writing data** between in-memory data structures and different formats: CSV and text files, Microsoft Excel, SQL databases, and the fast HDF5format;
- **data alignment** and integrated handling of **missing data**: gain automatic label-based alignment in computations and easily manipulate messy data into an orderly form.
- Flexible **reshaping** and pivoting of datasets.

- Intelligent label-based **slicing**, **fancy indexing**, and **subsetting** of large datasets.
- Columns can be inserted and deleted from data structures for **size mutability**.
- Aggregating or transforming data with a powerful **group by** engine allowing split-apply-combine operations on datasets.
- High performance **merging and joining** of data sets.
- **Hierarchical axis indexing** provides an intuitive way of working with high-dimensional data in a lower-dimensional data structure.
- **Time series**-functionality: date range generation and frequency conversion,moving window statistics, date shifting and lagging. Even create domain-specific time offsets and join time series without losing data.
- Highly **optimized for performance**, with critical code paths written in Cython or C.
- Python with **pandas** is in use in a wide variety of **academic and commercial** domains, including Finance, Neuroscience, Economics, Statistics, Advertising, Web Analytics, and more.

Scikit-learn:

- **Scikit-learn** (formerly **scikits.learn**and also known as **sklearn**) is a free software machine learning library for the Python programming language.
- It features various classification ,regression and clustering algorithms including

support vector machines,random forests,gradient boosting, *k*-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

- Scikit-learn is largely written in Python, and uses numpy etensively for high- performance linear algebra and array operations.
- Further more some core algorithms are written in Cython to improve performance.
- Support vector machines are implemented by a Cython wrapper around LIBSVM; logistic regression and linear support vector machines by a similar wrapper around LIBLINEAR.

- In such cases, extending these methods with Python may not be possible.
- Scikit-learn integrates well with many other Python libraries, such as matplotlib and plotly for plotting, numpy for array factorization, pandas data frames, skimpy, and many more.
- Scikit-learn is one of the most popular machine learning libraries on GitHub.

SciPy:

- **SciPy** is a free and open-source Python library used for scientific computingand technical computing.
- SciPy contains modules for optimization, linear algebra, integration, interpolation, special functions,FFT,signal and image processing,ODE solvers and other tasks common in science and engineering.
- SciPy builds on the NumPy array object and is part of the NumPy stack which includes tools like Matplotlib, pandas and SymPy, and an expanding set of scientific computing libraries.
- This NumPy stack has similar users to other applications such as MATLAB,GNU Octave, and Scilab.
- The NumPy stack is also sometimes referred to as the SciPystack.
- SciPy is also a family of conferences for users and developers of these tools: SciPy (in the United States), Euro SciPy (in Europe) and SciPy.in (in India).

6.2. SAMPLE CODE

Users.login.html:

```
{% extends 'common/base.html' %}
{% block mainblock %}
{% load staticfiles %}
<style>
    .mainbl{
        height:30em;
        width:50%;
overflow:auto;
position:relative;
        left:5%;
        top:7em;
background:rgba(191,200,173,0.8);
float:left;
    }
    table{
position:relative;
        left:30%;
        top:7em;
        border-spacing:0px;
    }
    td{
text-align:center;
        padding:5px;
    }
    input[type=text],input[type=password],input[type=email]{
        padding:10px;
    }
    input[type=submit]{
        padding-left:30px;
        padding-right:30px;
        padding-top:10px;
        padding-bottom:10px;
text-transform:uppercase;
color:black;
    }
</style>

<div class="mainbl">
<form method="post">
        {% csrf_token %}
<table>
<tr>
<td><p style="color:white">{{message}}</p></td>
<td><p style="color:white;">{{users}}</p></td>
</tr>
<tr>
<td>User Name</td>
<td><input type="text" name="username" placeholder="Enter User Name"></td>
</tr>
<tr>
<td>Password</td>
<td><input type="password" name="password" placeholder="Enter Password"></td>
</tr>
<tr>
<td colspan="2" style="text-align:center;"><input type="submit" name="submit"
value="submit"></td>
</tr>
```

```
<tr>
<td colspan="2" style="text-align:center;"><a href="{% url 'users:registration'
%}">Register Here...</a></td>
</tr>
</table>
</form>
</div>
<imgsrc="{% static 'images/start_now.gif' %}" style="width:40%;margin-
left:750px;margin-top:-160px;">
{% endblock %}
```

Registration.html:

```
{% extends 'common/base.html' %}
{% block mainblock %}

<style>
    input[type=submit]{
    padding-left:30px;
    padding-right:30px;
    padding-top:10px;
    padding-bottom:10px;
text-transform:uppercase;
color:black;
    }
     .mainbl{
    height:40em;
    width:50%;
overflow:auto;
position:relative;
    left:25%;
    top:4em;
background:rgba(191,200,173,0.8);
    }
    table{
position:relative;
        left:30%;
        top:7em;
        border-spacing:0px;
    }
    td{
text-align:center;
    padding:5px;
    }
    input[type=text],input[type=password],input[type=email],input[type=number]{
    padding:10px;
    }
</style>

<div class="mainbl">
<form method="POST">
        {% csrf_token %}
        {% if form.errors %}
    {% for field in form %}
        {% for error in field.errors %}
<div class="alert alert-danger">
<p style="font-size:100px;"><strong>{{ error|escape }}</strong></p>
</div>
        {% endfor %}
    {% endfor %}
```

```
    {% for error in form.non_field_errors %}
<div class="alert alert-danger">
<p style="font-size:100px;"><strong>{{ error|escape }}</strong></p>
</div>
    {% endfor %}
{% endif %}
<table>
<tr>
<td>First Name</td>
<td><input type="text" name="firstname" placeholder="Enter First Name" required></td>
</tr>
<tr>
<td>Last Name</td>
<td><input type="text" name="lastname" placeholder="Enter Last Name" required></td>
</tr>
<tr>
<td>User Name</td>
<td><input type="text" name="username" placeholder="Enter User Name" required></td>
</tr>
<tr>
<td>Password</td>
<td><input type="password" name="password" placeholder="Enter Password" required></td>
</tr>
<tr>
<td>Mobile Number</td>
<td><input type="number" name="mobile" placeholder="Enter Mobile Number"
required></td>
</tr>
<tr>
<td>Email Id</td>
<td><input type="email" name="email" placeholder="Enter Email Id" required></td>
</tr>
<tr>
<td>Locations</td>
<td><input type="text" name="location" placeholder="Enter Location" required></td>
</tr>
<tr>
<td>Date Of Birth</td>
<td><input type="text" name="dob" placeholder="Enter Date Of Birth" required></td>
</tr>
<tr>
<td colspan="2" style="text-align:center;"><input type="submit" name="submit"
value="submit"></td>
</tr>
</table>
</form>
</div>

{% endblock %}
```

users.view.py

```
import base64
import re
from encodings.base64_codec import base64_encode, base64_decode

import os
from random import randint
```

```
import pyAesCryptas pyAesCrypt
import pyaesas pyaes
from django.core.files.storageimport FileSystemStorage
from django.core.mailimport send_mail
from django.db.modelsimport query
from django.httpimport HttpResponse
from django.shortcutsimport render, redirect, get_object_or_404

from cybersecurity.EmailSettingsimport SET_DEFAULT_FROM_EMAIL
from users.decoratorsimport my_login_required
from users.formsimport UserRegistrationForm, MakeQuestionsForms
from users.modelsimport UserRegistration, MakeQuestions, Resources, ResourceRequest


deflogin(request):
if request.method == 'POST':
        username = request.POST.get('username')
        password = request.POST.get('password')
try:
user_object = UserRegistration.objects.get(userName=username,password=password)
except:
user_object = None
if user_objectis not None:
request.session['user_id'] = user_object.id
return redirect('users:home')
return render(request,'users/login.html',)

defregistration(request):
if request.method == 'POST':
'''forms = UserRegistrationForm(request.POST)
        if forms.is_valid():
user_registration = forms.save()
            id = user_registration.pk'''
firstName = request.POST.get('firstname','')
lastName = request.POST.get('lastname','')
userName = request.POST.get('username','')
        password = request.POST.get('password','')
mobileNumber = request.POST.get('mobile','')
emailId = request.POST.get('email','')
        location = request.POST.get('location','')
        dob = request.POST.get('dob','')
usr = UserRegistration.objects.create(firstName=firstName, lastName=lastName,
userName=userName, password=password, mobilenum=mobileNumber, emailId=emailId,
location=location, dob=dob)
if usr:
return redirect('users:regquestions',usr.id)
return render(request, 'users/registration.html')


@my_login_required
defhome(request):
all_resource_objects = Resources.objects.all()
return render(request,'users/home.html',{'all_resource_objects':all_resource_objects})


@my_login_required
defuploadfiles(request):
    message = ''
userid = request.session['user_id']
user_obj = UserRegistration.objects.get(id=userid)
if request.method == 'POST':
        BASE_PATH =
os.path.join(os.path.dirname(os.path.dirname(os.path.abspath(__file__))),
```

```
'assets/media/tmp/')
myfile = request.FILES['myfile']
        key = request.POST.get('password', None)
file_type = request.POST.get('filetype', None)
file_description = request.POST.get('description', None)
resource_id = randint(0,999999)
if file_type == "Image":
valid_extensions = ['.jpeg', '.jpg', '.bmp', '.gif', '.png', '.svg']
eliffile_type == "Video":
valid_extensions = ['.wmv', '.avi', '.flv', '.mov', '.mp4', '.webm', '.mpeg', '.3gp']
eliffile_type == "Audio":
valid_extensions = ['.wav', '.mp3', '.wma', '.mp4a', '.mpc', '.ogg', '.wma', '.oga']
f_path, file_extension = os.path.splitext(os.path.join(BASE_PATH, myfile.name))
if file_extensionin valid_extensionsand not re.search("[!#$%&'*+-
.^_`|~:]",file_description) and not re.search("^[!#$%&'*+-.^_`|~:]$",key):
file_name = myfile.name
fs = FileSystemStorage()
temp_file = fs.save('tmp/'+myfile.name, myfile)
file_size = fs.size('tmp/'+myfile.name)
uploaded_file_url = fs.url(temp_file)
            obj=Resources.objects.filter(file_name=file_name,file_size=file_size)
if obj.count() >0:
                message = 'file already exists'
else:
if uploaded_file_url:
bufferSize = 64 * 1024
pyAesCrypt.encryptFile(os.path.join(BASE_PATH,myfile.name),
os.path.join(BASE_PATH,str(resource_id)), key, bufferSize)
#os.remove(os.path.join(BASE_PATH, myfile.name))
resource_obj = Resources.objects.create(user=user_obj,file_name=
file_name,file_type=file_type,file_description=file_description,
file_size=file_size,file_extension=file_extension,file_key=key,resource_id=resource_id
,
resource=temp_file)
if resource_obj:
return redirect('users:home')
eliffile_extensionnot in valid_extensions:
upcnt = user_obj.uploadcount
if upcnt+1 >5:
user_obj.status = 'blocked'
user_obj.save(update_fields=["status"])
return redirect('users:login')
else:
user_obj.uploadcount = upcnt+1
user_obj.save(update_fields=["uploadcount"])
            message = 'not valid extension'
else:
            message = 'please check description or key'
return
render(request,'users/uploadfiles.html',{'user_obj':user_obj,'message':message})


@my_login_required
defcharts(request,chart_type):
    user_obj=UserRegistration.objects.get(id=request.session['user_id'])
res_obj=Resources.objects.filter(user=user_obj)
return render(request,'users/charts.html',{'res_obj':res_obj})

@my_login_required
defdeletefiles(request,id):
    res = Resources.objects.get(id=id)
res_name = res.resource_id
if res.delete():
```

```
        BASE_PATH =
os.path.join(os.path.dirname(os.path.dirname(os.path.abspath(__file__))),
'assets/media/tmp/')
os.remove(os.path.join(BASE_PATH, res_name))
return redirect('users:home')

@my_login_required
defrequestforfiles(request,id):
user_id = request.session['user_id']
user_object = UserRegistration.objects.get(id=user_id)
resource_object = Resources.objects.get(id=id)
if ResourceRequest.objects.create(user=user_object,resource=resource_object):
return redirect('users:home')

@my_login_required
defrequest_status(request):
reid = request.session['user_id']
reob = UserRegistration.objects.get(id=reid)
req_status = ResourceRequest.objects.filter(user=reob)
return render(request,'users/request_status.html',{'status':req_status})

@my_login_required
defdecrypt_file(request,id):
    message=''
OUT_PATH = 'no'
req_obj = ResourceRequest.objects.get(id=id)
    rid = req_obj.resource.id
res_obj = Resources.objects.get(id=rid)
atts = req_obj.resource.attacks
dows = req_obj.resource.downloads
file_id = req_obj.resource.resource_id
file_name = req_obj.resource.file_name
enc_file = 'assets/media/tmp/'+file_id
dec_file = 'assets/media/tmp/'+file_name
    PATH_OUT =
os.path.join(os.path.dirname(os.path.dirname(os.path.abspath(__file__))),dec_file)
if request.POST.get("decr_btn"):
res_id = request.POST.get('resourceid','')
res_pass = request.POST.get('passkey','')
if res_id == req_obj.resource.resource_id:
if res_pass == req_obj.resource.file_key:
bufferSize = 64 * 1024
BASE_PATH = os.path.join(os.path.dirname(os.path.dirname(os.path.abspath(__file__))),
enc_file)
                BASE_PATH_OUT =
os.path.join(os.path.dirname(os.path.dirname(os.path.abspath(__file__))),
dec_file)
                password = res_pass
pyAesCrypt.decryptFile(BASE_PATH, BASE_PATH_OUT, password, bufferSize)
res_obj.downloads = dows + 1
res_obj.save(update_fields=['downloads'])
return redirect('users:decryptfile',id)
else:
res_obj.attacks = atts + 1
res_obj.save(update_fields=['attacks'])
else:
res_obj.attacks = atts+1
res_obj.save(update_fields=['attacks'])
cac_count = request.session['check_count']
if cac_count<5:
if res_pass == req_obj.resource.file_key:
request.session['check_count']=cac_count+1
message="Please Ensure Your resource id is correct..."
```

```
else:
request.session['check_count'] = cac_count + 1
message='Please Ensure Credntials are right'
else:
del request.session['check_count']
return redirect('users:login')
return
render(request,'users/decrypt_file.html',{'req_obj':req_obj,'message':message,'BASE_PA
TH_OUT':PATH_OUT})

@my_login_required
defdownloadss(request,nam):
res_obj = Resources.objects.get(id=nam)
nam = res_obj.file_name
str =
os.path.join(os.path.dirname(os.path.dirname(os.path.abspath(__file__))),'assets/media
/tmp/'+nam)
    response = HttpResponse(str, content_type='text/plain')
    response['Content-Disposition'] = 'attachment; filename=%s' % nam
return render(request,'users/downs.html',{'str':nam})

@my_login_required
deflogout(request):
del request.session['user_id']
return redirect('users:login')

defthanks(request):
if request.method == 'POST':
return redirect('users:login')
return render(request,'users/thanks.html',)

defregquestions(request,id):
user_obj = get_object_or_404(UserRegistration, pk=id)
if request.method == "POST":
'''forms = MakeQuestionsForms(request.POST)
        if forms.is_valid():
forms.save(commit=False)
forms.user = user_obj
forms.save()'''
question_one = request.POST.get('firstset','')
question_two = request.POST.get('secondset','')
question_thr = request.POST.get('thirdset','')
question_fur = request.POST.get('fourthset','')
question_fiv = request.POST.get('fifthset','')
ans_one = request.POST.get('ans_one','')
ans_two = request.POST.get('ans_two','')
ans_thr = request.POST.get('ans_thr','')
ans_fur = request.POST.get('ans_fur','')
ans_fiv = request.POST.get('ans_fiv','')
que =
MakeQuestions.objects.create(user=user_obj,question_one=question_one,question_two=ques
tion_two,
question_three=question_thr,question_four=question_fur,
question_five=question_fiv,answer_one=ans_one,answer_two=ans_two,
answer_three=ans_thr,answer_four=ans_fur,answer_five=ans_fiv)
if que:
return redirect('users:login')
return render(request,'users/savequestions.html',)

defaskquestions(request,id):
user_object = UserRegistration.objects.get(id=id)
    questions = MakeQuestions.objects.filter(user=user_object)
msg = ''
```

```
if request.method == "POST":
ans_one = request.POST.get('ans_one','')
ans_two = request.POST.get('ans_two','')
ans_three = request.POST.get('ans_three','')
ans_four = request.POST.get('ans_four','')
ans_five = request.POST.get('ans_five','')
try:
            check = MakeQuestions.objects.get(user=user_object)
except:
            check = None
if check:
if check.answer_one == ans_oneand check.answer_two == ans_twoand check.answer_three ==
ans_threeand check.answer_four == ans_fourand check.answer_five == ans_five:
request.session['user_id'] = user_object.id
return redirect('users:home')
else:
msg = 'Wrong Details'
else:
msg = 'error 2'
return render(request,'users/askquestions.html',{'questions':questions,'msg':msg})


defverification(request,id):
#sdi = base64_decode(id)
    #sdi = dec(id)
plaintext = "Hello World!!!!!"
plaintext_bytes = b'Hello World!!!!!'
key = b"CHENNAISUNDAYSYSTEMS1234569870_&"
aes = pyaes.AES(key)
ciphertext = aes.encrypt(plaintext_bytes)
    decrypted = aes.decrypt(ciphertext)
return render(request, 'users/verification.html',{'enc':ciphertext,'dec':decrypted} )
```

Admin login.html:

```
{% extends 'common/base.html' %}
{% block mainblock %}
{% load staticfiles %}
<style>
    .logintbl{
    height:30em;
    width:50%;
overflow:auto;
position:relative;
    left:8%;
    top:7em;
float:left;
background:rgba(191,200,173,0.8);
    }
    table{
position:relative;
        left:35%;
        top:10em;
        border-spacing:0px;
    }
    td{
text-align:center;
    padding:5px;
    }
    input[type=text],input[type=password]{
```

```
    padding:10px;
    }
    input[type=submit]{
    padding-left:30px;
    padding-right:30px;
    padding-top:10px;
    padding-bottom:10px;
text-transform:uppercase;
color:black;
    }
</style>
<div class="logintbl">
<form method="post">
        {% csrf_token %}
<table>
<tr>
<thstyle="color:white;">ADMINISTRATOR LOGIN</th>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><input type="text" name="username" placeholder="Enter User Name"></td>
</tr>
<tr>
<td><input type="password" name="password" placeholder="Enter Password"></td>
</tr>
<tr>
<td style="text-align:center;"><input type="submit" name="submit" value="submit"></td>
</tr>
</table>
</form>
</div>
<imgsrc="{% static 'images/admin.jpg' %}" style="margin-left:250px;margin-top:150px;">
<imgsrc="{% static 'images/adminloading.gif' %}" style="margin-left:250px;margin-
top:50px;">
{% endblock %}
```

Admin.view.py

```
import os

from django.core.mailimport send_mail, EmailMultiAlternatives
from django.db.modelsimport Q, Count
from django.shortcutsimport render, redirect

from cybersecurity.settingsimport DEFAULT_FROM_EMAIL
from users.modelsimport UserRegistration, ResourceRequest, Resources


deflogin(request):
if request.method == 'POST':
        username = request.POST.get('username','')
        password = request.POST.get('password','')
if username == 'admin' and password == 'admin':
return redirect('admins:home')
return render(request,'admins/login.html',)

defhome(request):
resource_objects = Resources.objects.all()
```

```
return render(request,'admins/home.html',{'resource_objects':resource_objects})

defremovefiles(request,file_id):
file_object = Resources.objects.get(id=file_id)
re_id = file_object.resource_id
if file_object.delete():
        BASE_PATH =
os.path.join(os.path.dirname(os.path.dirname(os.path.abspath(__file__))),
'assets/media/tmp/')
os.remove(os.path.join(BASE_PATH, re_id))
return redirect('admins:home')

defblockfiles(request,file_id):
file_object = Resources.objects.get(id=file_id)
file_object.status = 'blocked'
file_object.save(update_fields=["status"])
return redirect('admins:home')

defunblockfiles(request,file_id):
file_object = Resources.objects.get(id=file_id)
file_object.status = 'approved'
file_object.save(update_fields=["status"])
return redirect('admins:home')

defuserslist(request):
users_o = UserRegistration.objects.exclude(status='blocked')
return render(request, 'admins/userslist.html', {'users': users_o})

defremoveusers(request,user_id):
users_o = UserRegistration.objects.get(id=user_id)
users_o.delete()
return redirect('admins:userslist')

defblockusers(request,user_id):
users_o = UserRegistration.objects.get(id=user_id)
users_o.status='blocked'
users_o.save(update_fields=["status"])
return redirect('admins:userslist')

defblockedusers(request):
users_o = UserRegistration.objects.filter(status='blocked')
return render(request, 'admins/blockedusers.html', {'users': users_o})

defunblockusers(request,user_id):
users_o = UserRegistration.objects.get(id=user_id)
users_o.status = 'success'
users_o.count=0
users_o.uploadcount=0
users_o.save(update_fields=["status","count","uploadcount"])
return redirect('admins:blockedusers')

defgraph(request):
chart_objects =
UserRegistration.objects.all().values('status').annotate(total=Count('status'))
success_objects =
UserRegistration.objects.filter(status='success').values('count').annotate(total=Count
('count'))
blocked_objects =
UserRegistration.objects.filter(status='blocked').values('count').annotate(total=Count
('count'))
return
render(request,'admins/graph.html',{'chart_objects':chart_objects,'success_objects':su
ccess_objects,'blocked_objects':blocked_objects})
```

```
def logout(request):
return redirect('admins:login')

def userrequestsforfile(request):
request_objects = ResourceRequest.objects.all()
return
render(request,'admins/userrequestsforfile.html',{'request_objects':request_objects})

def approverequests(request,request_id):
request_obj = ResourceRequest.objects.get(id=request_id)
request_obj.status = 'approved'
subject = "Decrytion Key"
text_content=""
html_content = "<br/><p>RESOURCE
ID:<strong>"+request_obj.resource.resource_id+"</strong><br><br>KEY:<strong>"+request_
obj.resource.file_key+"</strong></p>"
from_mail = DEFAULT_FROM_EMAIL
to_mail = [request_obj.user.emailId]
#if send_mail(subject,message,from_mail,to_mail):
msg = EmailMultiAlternatives(subject, text_content, from_mail, to_mail)
msg.attach_alternative(html_content, "text/html")
if msg.send():
request_obj.save(update_fields=["status"])
return redirect('admins:userrequestsforfile')
return redirect('admins:userrequestsforfile')

def cancelrequests(request,request_id):
request_obj = ResourceRequest.objects.get(id=request_id)
request_obj.status = 'cencelled'
request_obj.save(update_fields=["status"])
return redirect('admins:userrequestsforfile')
```

SYSTEM TESTING

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

Introduction:

In general, software engineers distinguish software faults from software failures. In case of a failure, the software does not do what the user expects. A fault is a programming error that may or may not actually manifest as a failure. A fault can also be described as an error in the correctness of the semantic of a computer program. A fault will become a failure if the exact computation conditions are met, one of them being that the faulty portion of computer software executes on the CPU.

A fault can also turn into a failure when the software is ported to a different hardware

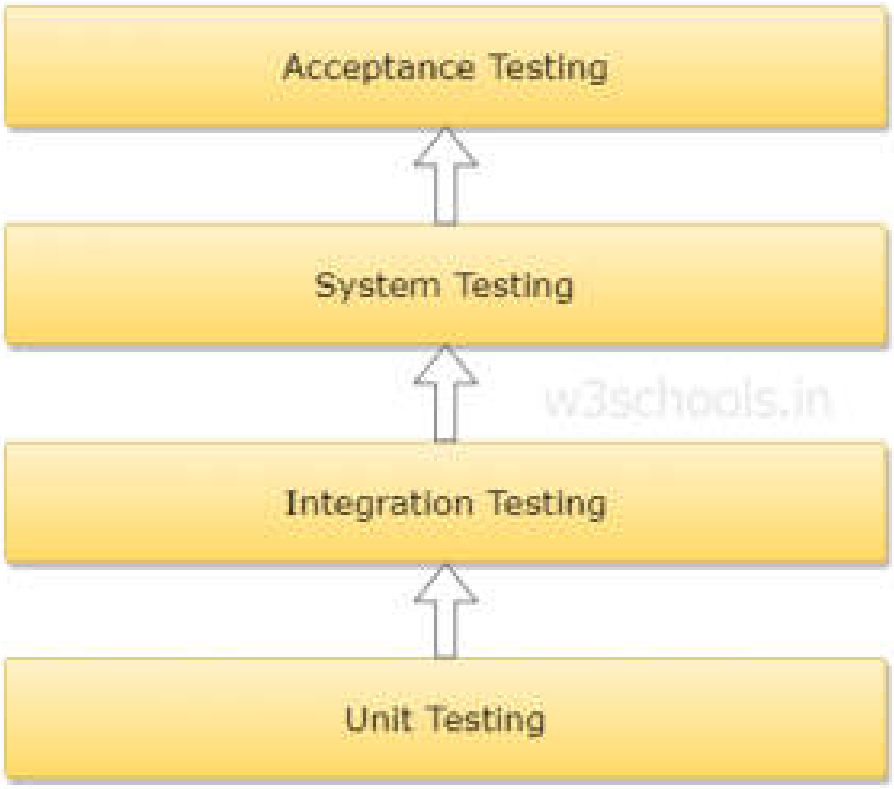
Software testing may be viewed as a sub-field of Software Quality Assurance but typically exists independently (and there may be no SQA areas in some companies). In SQA, software process specialists and auditors take a broader view on software and its development. They examine and change the software engineering process itself to reduce the amount of faults that end up in the code or deliverfaster.

Regardless of the methods used or level of formality involved the desired result of testing is a level of confidence in the software so that the organization is confident that the software has an acceptable defect rate. What constitutes an acceptable defect rate depends on the nature of the software. An arcade video game designed to simulate flying an airplane would presumably have a much higher tolerance for defects than software used to control an actual airliner.

A problem with software testing is that the number of defects in a software product can be very large, and the number of configurations of the product larger still. Bugs that occur infrequently are difficult to find in testing. A rule of thumb is that a system that is expected to function without faults for a certain length of time must have already been tested for at least that length of time. This has severe consequences for projects to write long-lived reliable software.

7.1 Test Cases:

Different levels of testing:



Levels of Testing

TYPES OF TESTS

Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs

accurately to the documented specifications and contains clearly defined inputs and expected results.

Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems/Procedures- interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

Unit Testing

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

8. RESULTS

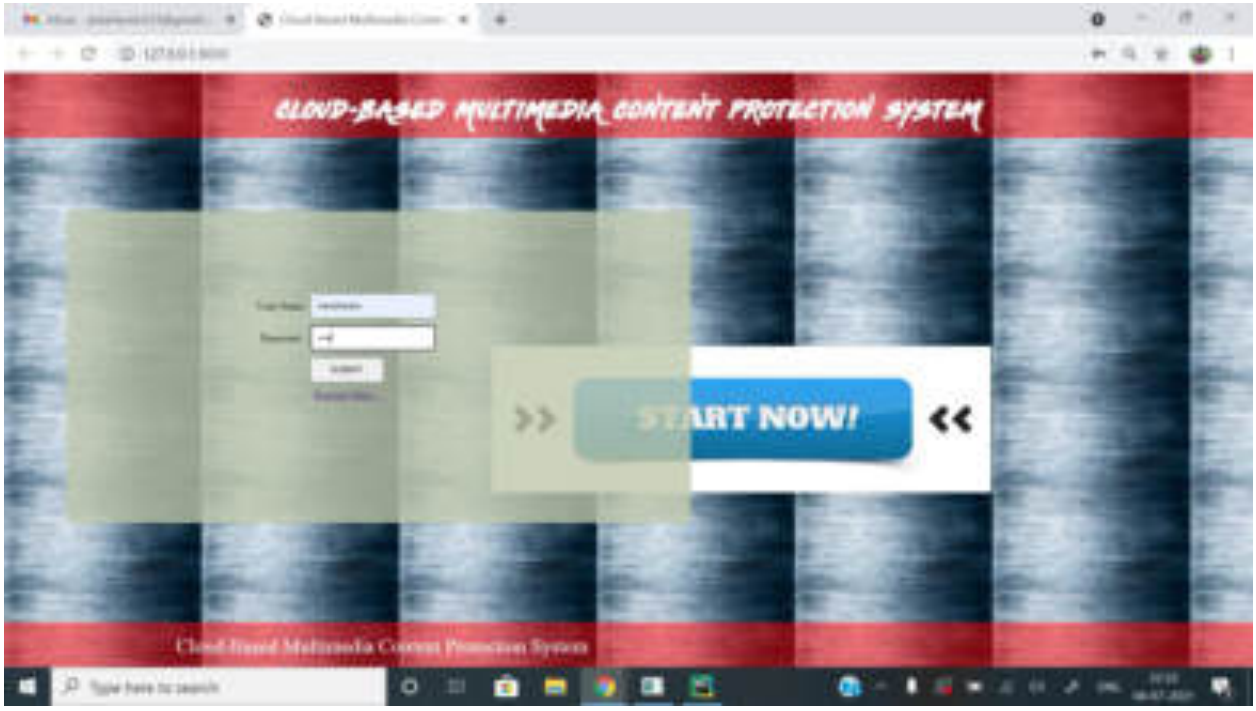


FIG:8.1 User Login Page

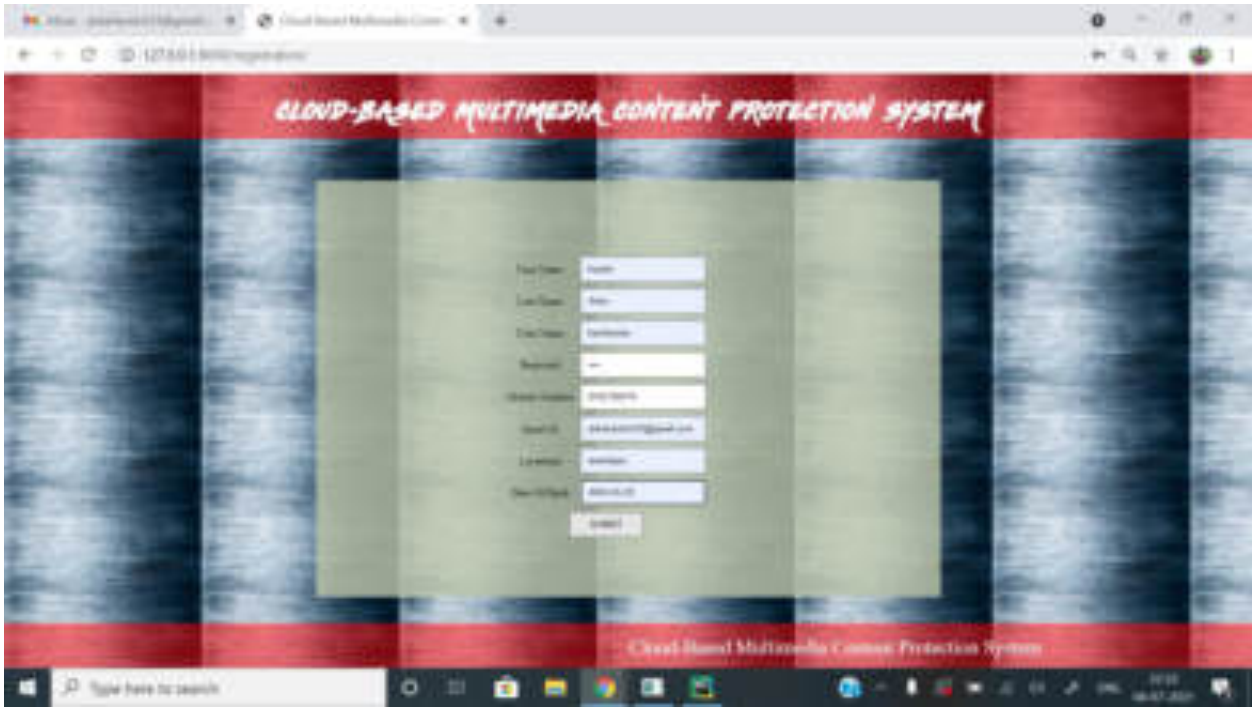


FIG:8.2 User Registration Page

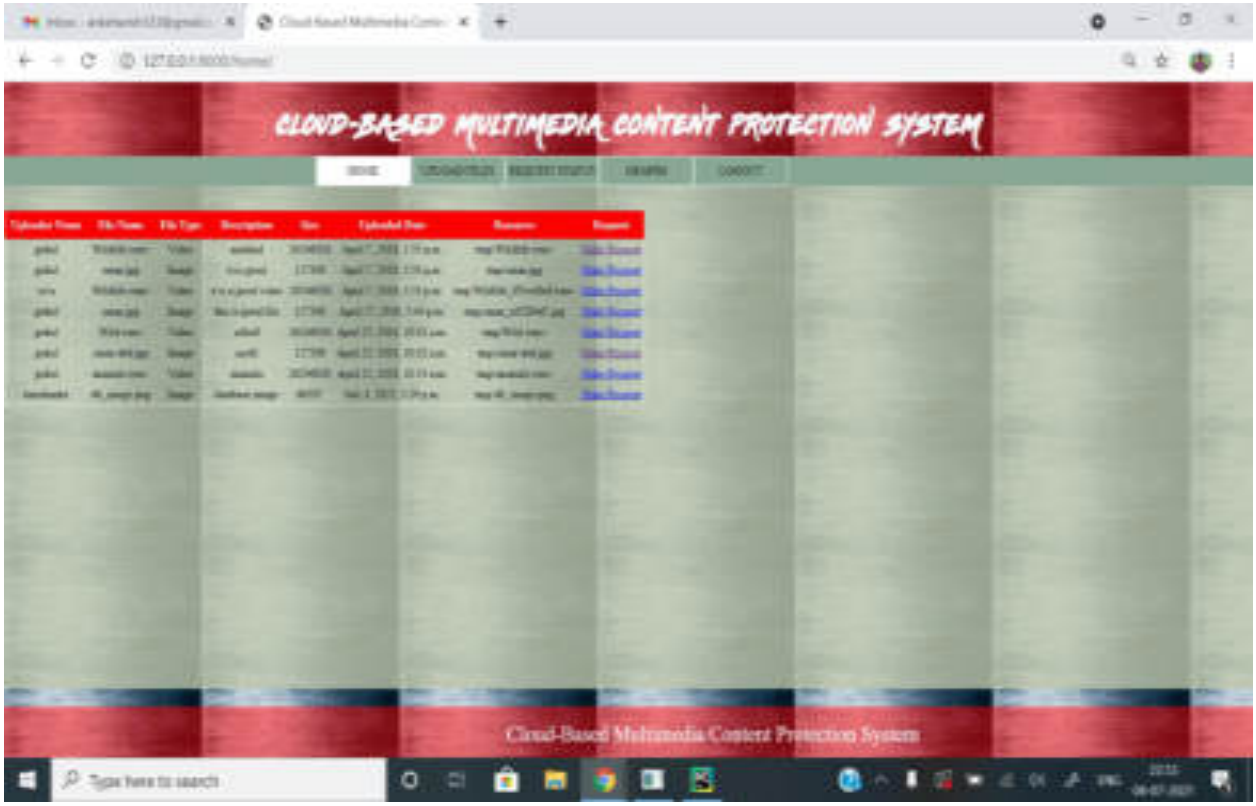


FIG:8.2 User Home Page, It display all multimedia content in cloud. User can send request to access it.

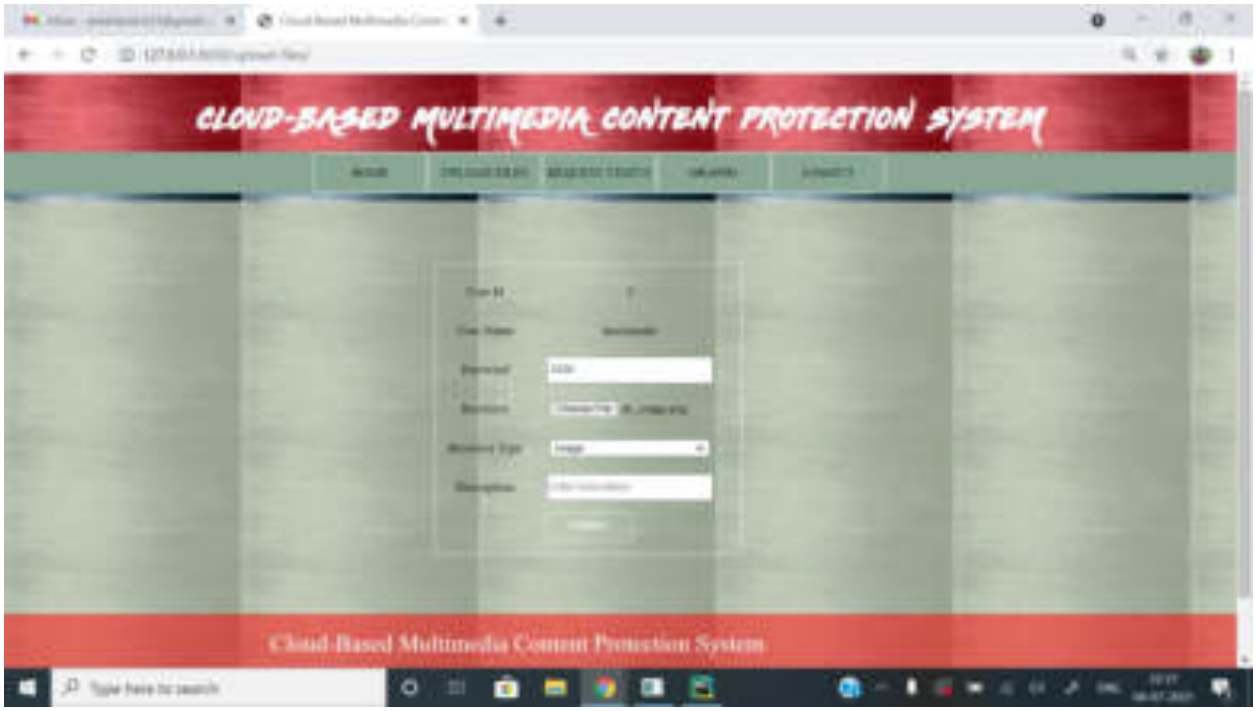


FIG:8.3 User can upload their own media details

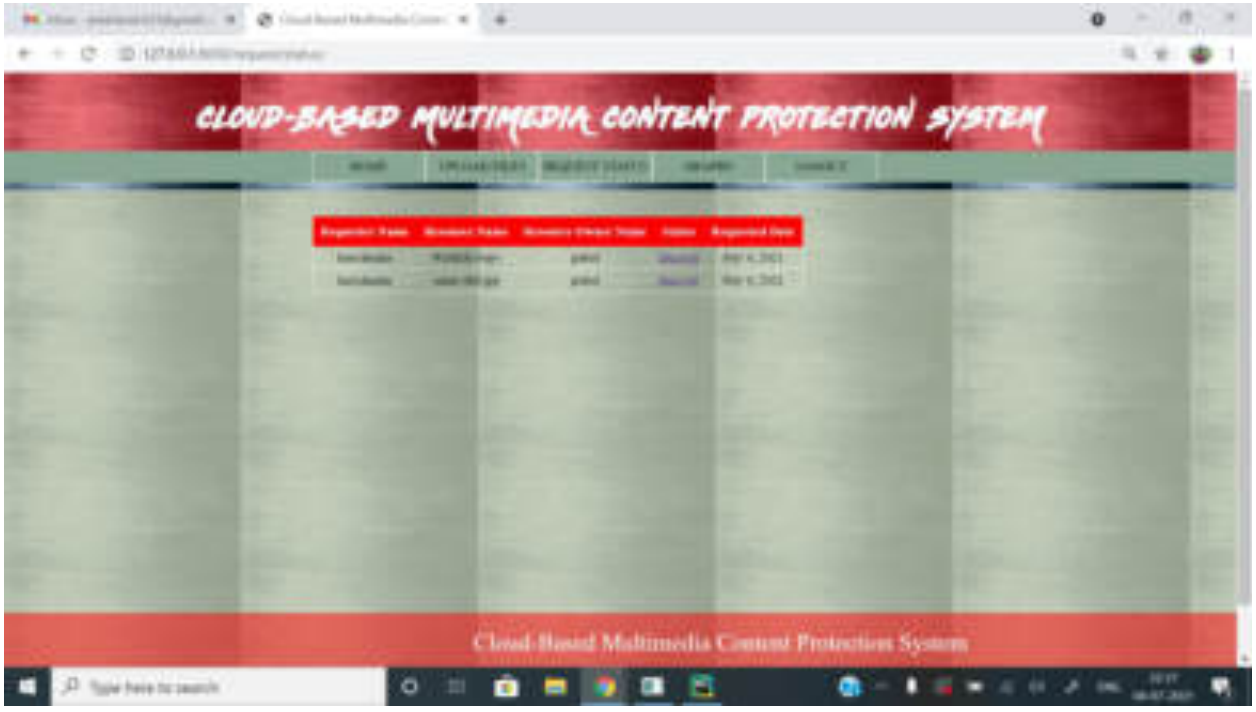


FIG:8.4 Users Request Status page

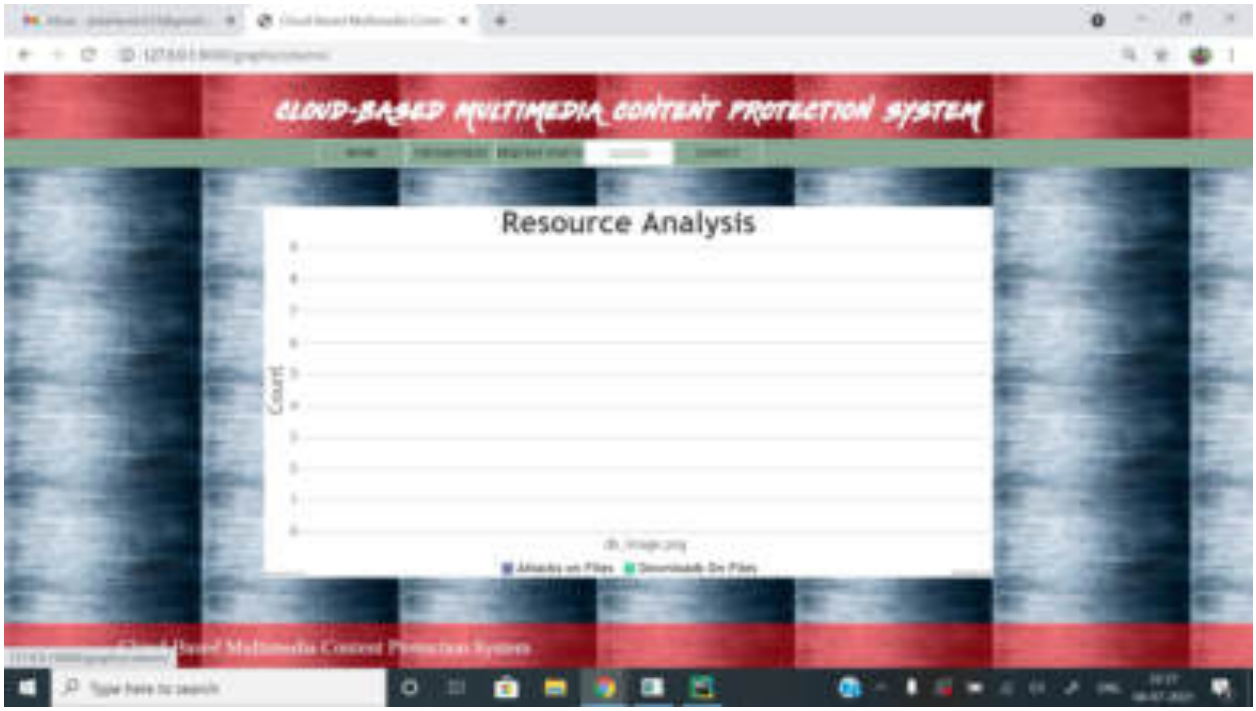


FIG:8.5 Users Resource Analysis page

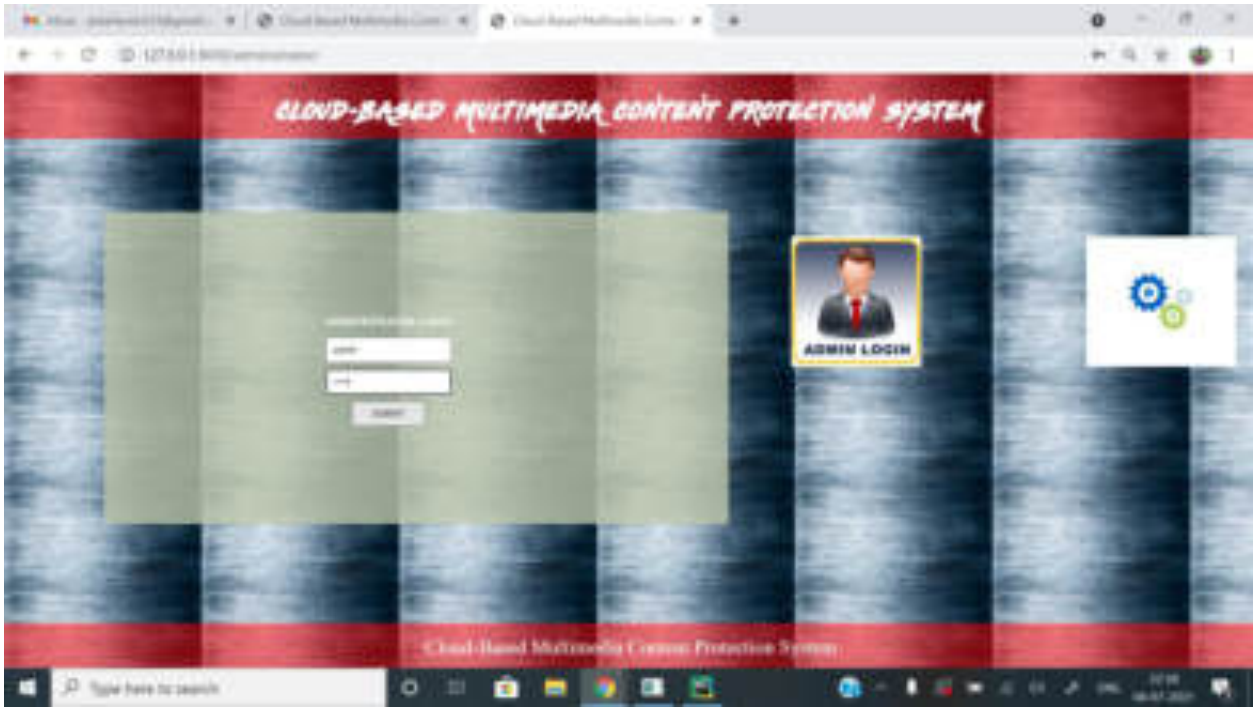


FIG:8.6 Admin login page

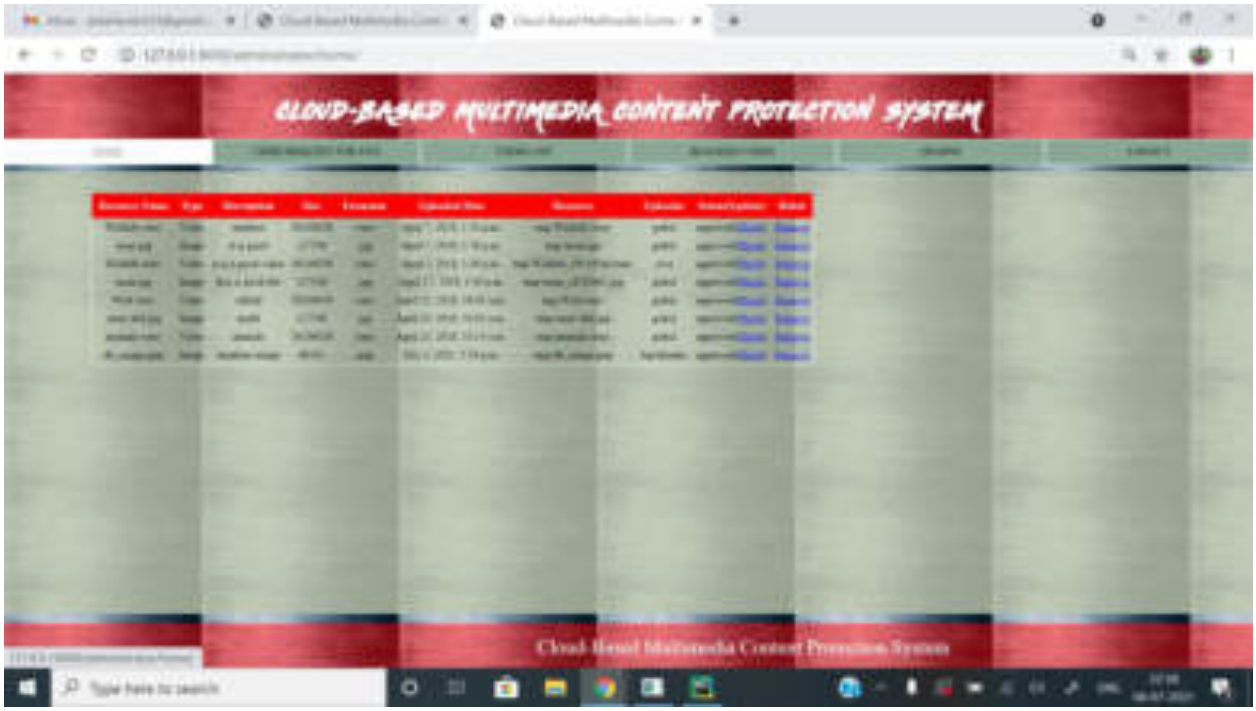


FIG:8.7 Admin home page and User Details

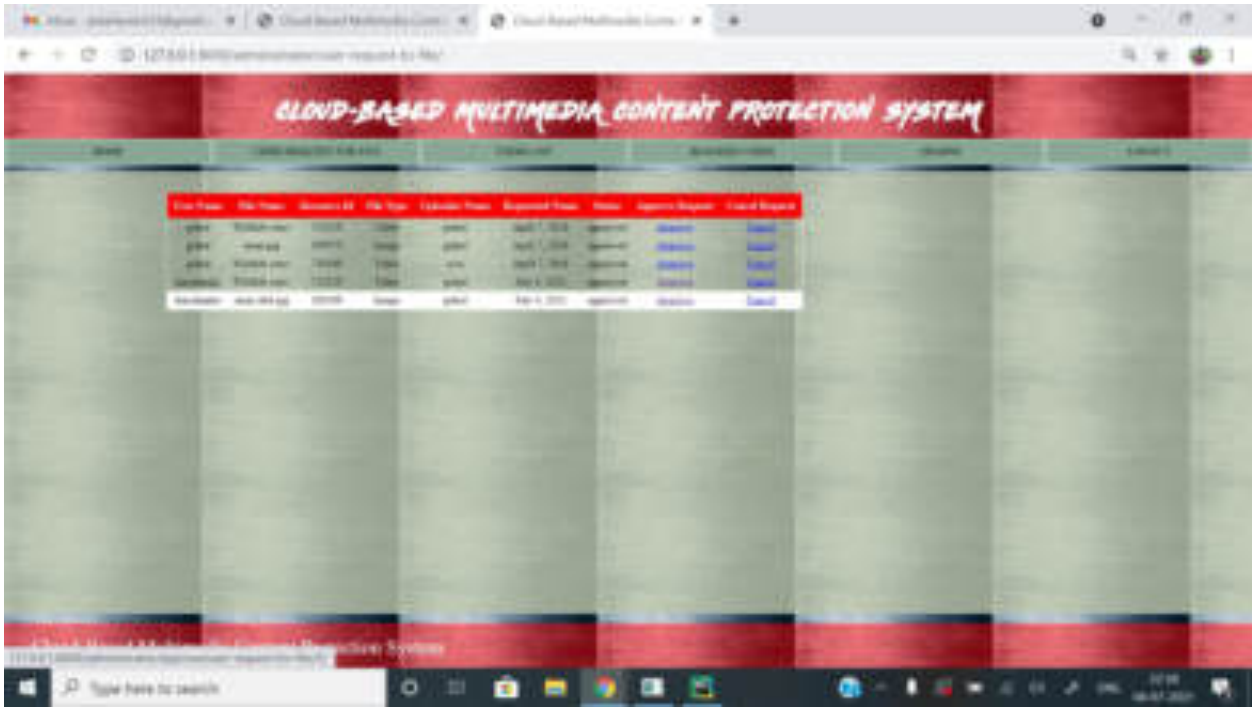


FIG:8.8 Users requested files response page. Whenever it is approved it sends key to mail.

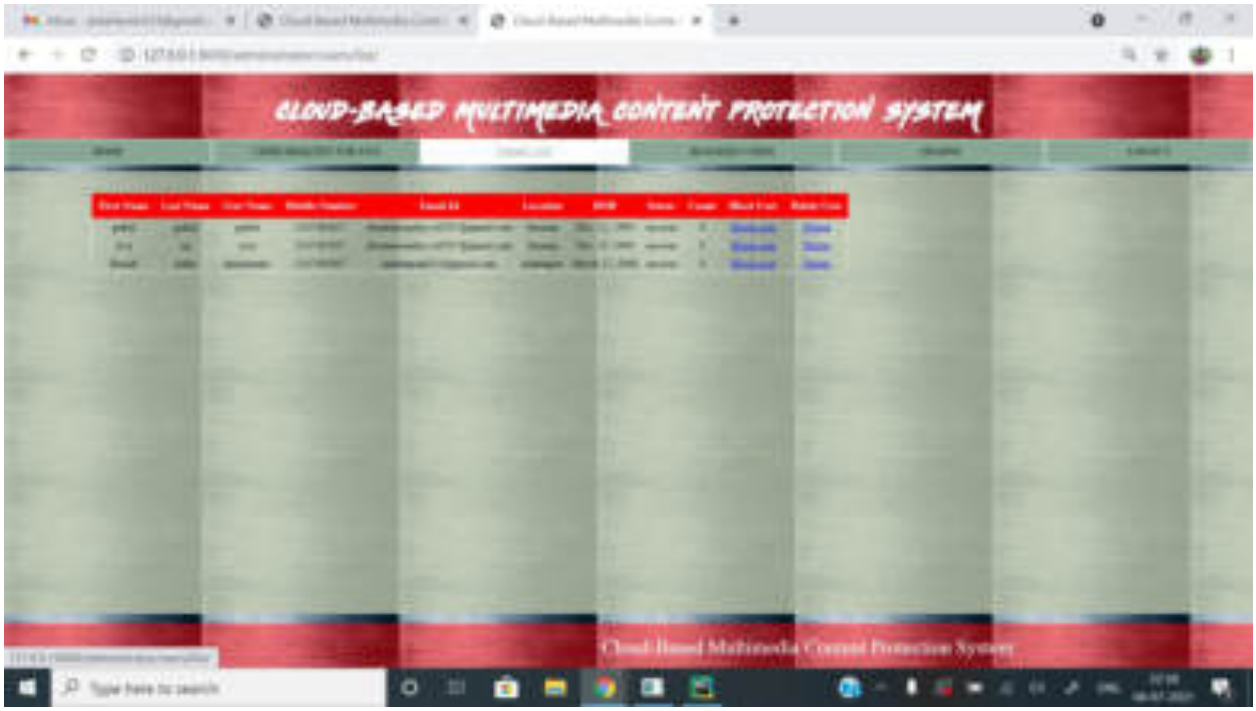


FIG:8.9 Users List in Admin side

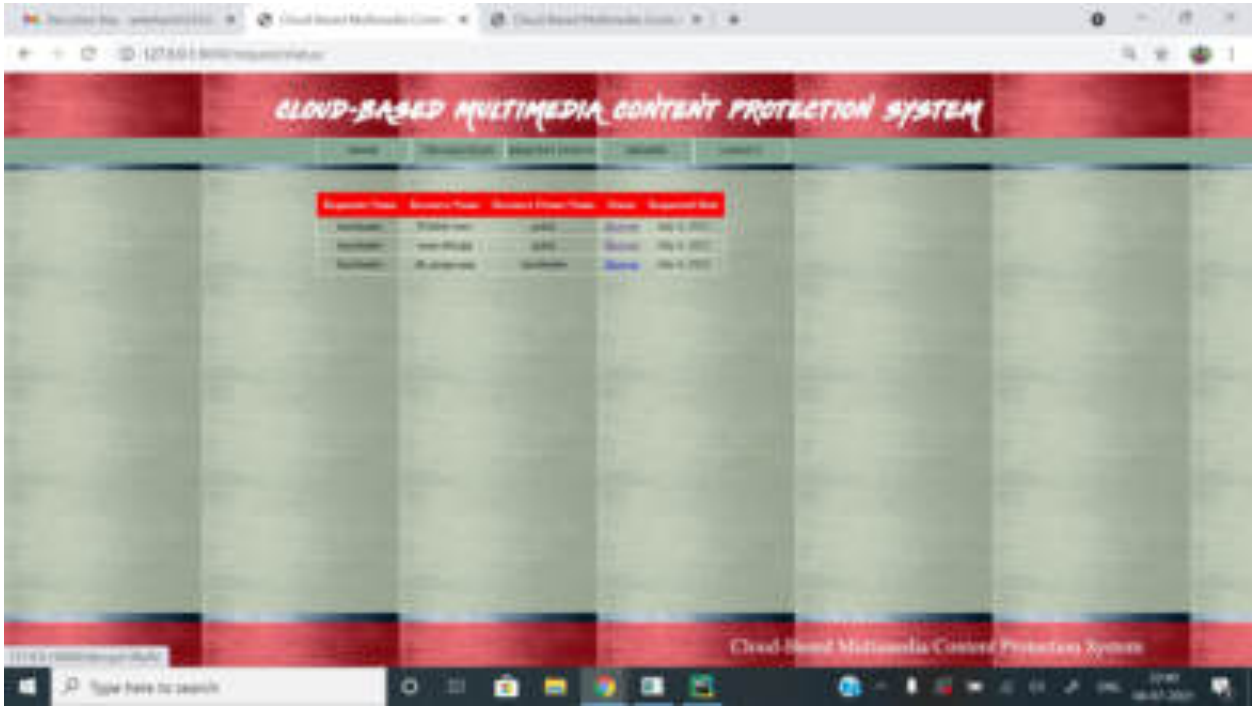


FIG:8.10 Users request status page.

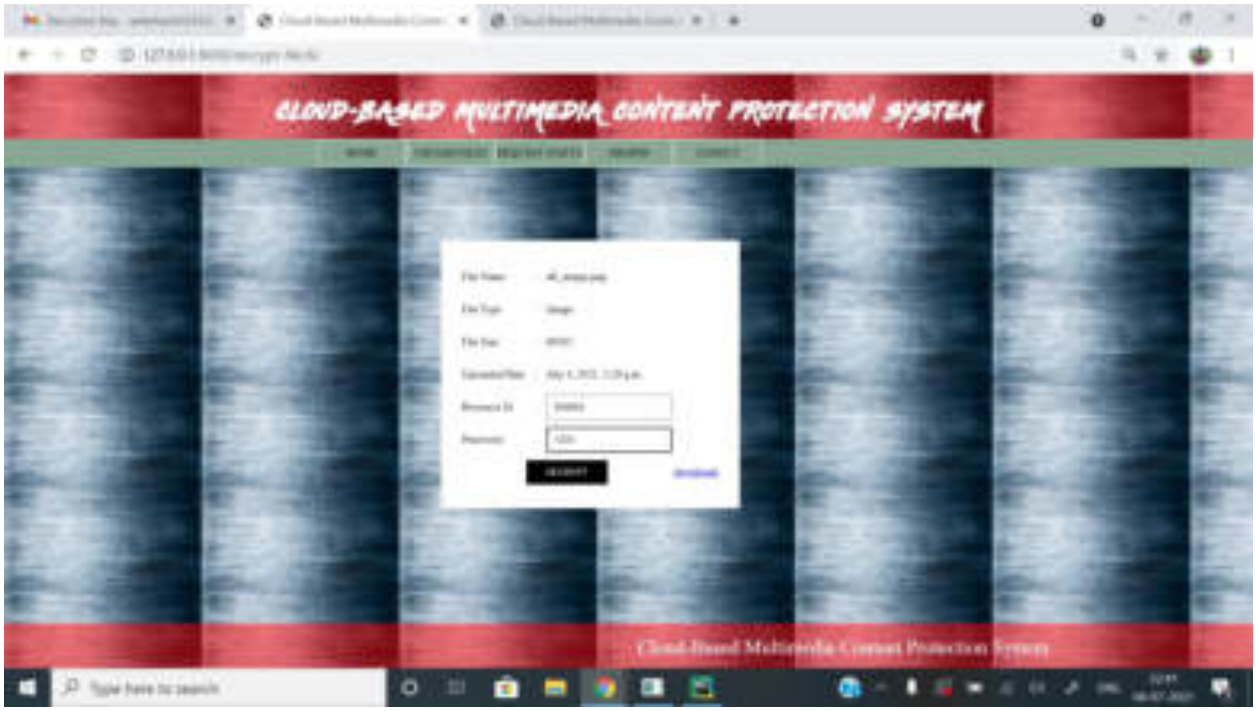


FIG:8.11 user can Encrypt the file with resource id & key. Then download it.

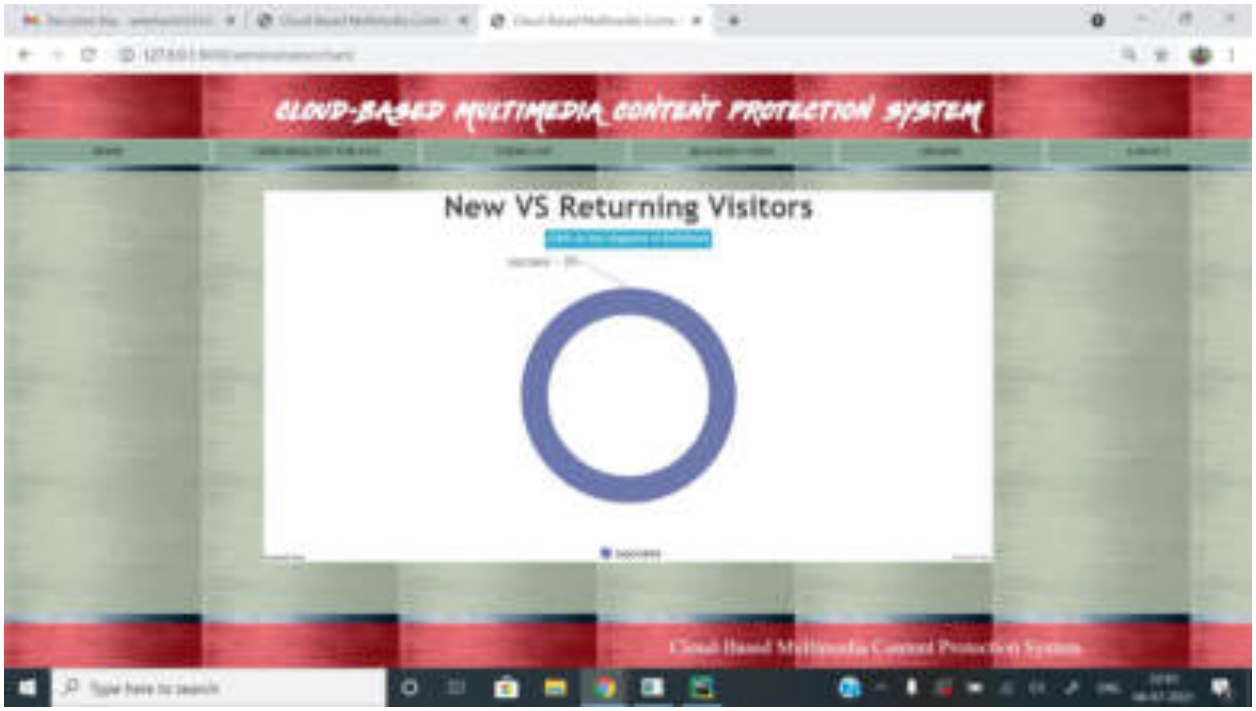


FIG:8.12 Admin Graph

CONCLUSION

Distributing copyrighted multimedia objects by uploading them to online hosting sites such as YouTube can result in significant loss of revenues for content creators. Systems needed to find illegal copies of multimedia objects are complex and large scale. In this paper, we presented a new design for multimedia content protection systems using multi-cloud infrastructures. The proposed system supports different multimedia content types and it can be deployed on private and/or public clouds. Two key components of the proposed system are presented. The first one is a new method for creating signatures of videos. Our method constructs coarse-grained disparity maps using stereo correspondence for a sparse set of points in the image. Thus, it captures the depth signal of the video, without explicitly computing the exact depth map, which is computationally expensive. Our experiments showed that the proposed signature produces high accuracy in terms of both precision and recall and it is robust to many video transformations including new ones that are specific to videos such as synthesizing new views. The second key component in our system is the distributed index, which is used to match multimedia objects characterized by high dimensions. The distributed index is implemented using the Map Reduce framework and our experiments showed that it can elastically utilize varying amount of computing resources and it produces high accuracy. The experiments also showed that it outperforms the closest system in the literature in terms of accuracy and computational efficiency. In addition, we evaluated the whole content protection system with more than 11,000 videos and the results showed the scalability and accuracy of the proposed system. Finally, we compared our system against the Content ID system used by YouTube. Our results showed that: (i) there is a need for designing robust signatures for videos since the current system used by the leading company in the industry fails to detect most modified 3D copies, and (ii) our proposed signature method can fill this gap, because it is robust to many video transformations.

Future scope

The work in this paper can be extended in multiple directions. For example, our current system is optimized for batch processing. Thus, it may not be suitable for online detection of illegally distributed multimedia streams of live events such as soccer games. In live events, only small segments of the video are available and immediate detection of copyright infringement is crucial to minimize financial losses. To support online detection, the matching engine of our system needs to be implemented using a distributed programming framework that supports online processing, such as Spark. In addition, composite signature schemes that combine multiple modalities may be needed to quickly identify short video segments. Furthermore, the crawler component needs to be customized to find online sites that offer pirated video streams and obtain segments of these streams for checking against reference streams, for which the signatures would also need to be generated online. Another future direction for the work in this paper is to design signatures for recent and complex formats of 3D videos such as multi view plus depth. A multi view plus depth video has multiple texture and depth components, which allow users to view a scene from different angles. Signatures for such videos would need to capture this complexity, while being efficient to compute, compare, and store.

REFERENCES

[1] A. Abdelsadek. Distributed index for matching multimedia objects. Master's thesis, Simon Fraser University, Canada, 2014.

[2] A. Abdelsadek and M. Hefeeda. Dimo: Distributed index for matching multimedia objects using mapreduce. In Proc. of ACM Multimedia Systems Conference (MMSys'14), pages 115–125, Singapore, March 2014.

[3] M. Aly, M. Munich, and P. Perona. Distributed Kd-Trees for Retrieval from Very Large Image Collections. In Proc. of British Machine Vision Conference (BMVC), Dundee, UK, August 2011.

[4] J. Bentley. Multidimensional binary search trees used for associative searching. Communications of the ACM, 18(9):509– 517, September 1975.

[5] P. Cano, E. Batle, T. Kalker, and J. Haitsma. A review of algorithms for audio fingerprinting. In Proc. of IEEE Workshop on Multimedia Signal Processing, pages 169–173, St. Thomas, US Virgin Islands, December 2002.

[6] Copyright on Youtube. http://www.youtube.com/yt/copyright/. 1520-9210 (c) 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information. This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2015.2389628, IEEE Transactions on Multimedia 30

[7] J. Dean and S. Ghemawat. Mapreduce: simplified data processing on large clusters. In Proc. of Symposium on Operating Systems Design and Implementation (OSDI'04), pages 137–150, San Francisco, CA, December 2004.

[8] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'09), pages 248–255, Miami, FL, June 2009.

[9] A. Hampapur, K. Hyun, and R. Bolle. Comparison of sequence matching techniques for video copy detection. In Proc. of SPIE Conference on Storage and Retrieval for Media Databases (SPIE'02), pages 194–201, San Jose, CA, January 2002.

[10] S. Ioffe. Full-length video fingerprinting. Google Inc., July 24 2012. US Patent 8229219.

[11] A. Kahng, J. Lach, W. Mangione-Smith, S. Mantik, I. Markov, M. Potkonjak, P. Tucker, H.

Wang, and G. Wolfe. Watermarking techniques for intellectual property protection. In Proc. of the 35th Annual Design Automation Conference (DAC'98), pages 776–781, San Francisco, CA, June 1998.

[12] N. Khodabakhshi and M. Hefeeda. Spider: A system for finding 3d video copies. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 9(1):7:1–7:20, February 2013. [13] S. Lee and C. Yoo. Robust video fingerprinting for content-based video identification. IEEE Transactions on Circuits and Systems for Video Technology, 18(7):983–988, July 2008.

[14] H. Liao, J. Han, and J. Fang. Multi-dimensional index on hadoop distributed file system. In Proc. of IEEE Conference on Networking, Architecture and Storage (NAS'10), pages 240–249, Macau, China, July 2010.

[15] Z. Liu, T. Liu, D. Gibbon, and B. Shahraray. Effective and scalable video copy detection. In Proc. of ACM Conference on Multimedia Information Retrieval (MIR'10), pages 119–128, Philadelphia, PA, March 2010.

[16] J. Lu. Video fingerprinting for copy identification: From research to industry applications. In IS&T/SPIE Electronic Imaging, volume 7254 of SPIE Proceedings, pages 725402–725402. International Society for Optics and Photonics, SPIE, 2009.

[17] W. Lu, Y. Shen, S. Chen, and B. Ooi. Efficient processing of k nearest neighbor joins using mapreduce. Proceedings of the VLDB Endowment (PVLDB), 5(10):1016–1027, June 2012.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"AUTOMATIC ATTENDANCE USING FACE RECOGNITION"

Submitted to

Jawaharlal Nehru Technological University Kakinada
in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
COMPUTER SCIENCE AND ENGINEERING

Submitted by

MALLELA RAJESH	**(17H71A0595)**	**KONDLA MOUNIKA**	**(17H71A0587)**
VYDA SAI PRASANNA	**(17H71A05A2)**	**VANUMU YASHWANTH**	**(17H71A05C0)**

Under the Esteemed Guidance of
Ms. Parveen Pathan, M.Tech.,
Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.

Phone : 08678 - 273535, 273623, Fax: 08678 - 273569

E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr.Hima Sekhar

MIC College of Technology

(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India

Phone : 08678 - 273535, 273623, Fax: 08678 - 273569

E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

CERTIFICATE

This is to certify that the Main Project entitled **"AUTOMATIC ATTENDANCE USING FACE RECOGNITION"** is a bonafide work carried out by **MALLELA RAJESH (17H71A0595), KONDLA MOUNIKA (17H71A0587), VYDA SAI PRASANNA (17H71A05A2) and VANUMU YASHWANTH (17H71A05C0)** in partial fulfillment for the award of degree of Bachelor of Technology in **Computer Science and Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

Parveen Pathan
Project Guide

D.Prasad
Head of the Department

(Dr. K.SRINIVAS)
Principal

Examiner 1

Examiner 2

ACKNOWLEDGEMENT

We would like to take this opportunity to express our deepest appreciation to the following people for their valuable contributions and assistance with this Project.

Initially, we would like to thank our project supervisor **Ms. Parveen Pathan, M.Tech., Department of Computer Science and Engineering** for his guidance and support, especially for his valuable ideas and knowledge provided throughout the Project. His expertise and experience in makes valuable comments and suggestions have been very useful in solving problems encountered during the Project.

We have the immense pleasure in expressing our thanks and deep sense of gratitude **Mr. D.Prasad, Head of the Department, Computer Science and Engineering** for extending necessary facilities for the completion of the Project.

We whole heartedly acknowledge **Dr. K.SRINIVAS, Principal and Prof D. PANDURANGA RAO, CEO** for giving opportunity to execute this Project. We also extend my thanks to all faculty members of **Computer Science and Engineering**, for their valuable guidance and encouragement in this Project.

We would like to extend our warm appreciation to all my friends for sharing us their knowledge, valuable contributions and help with this Project.

Finally, our special thanks to our families for their continuous support and help throughout our academic years and for their continual support and encouragement for this Project.

MALLELA RAJESH (17H71A0595)

KONDLA MOUNIKA (17H71A0587)

VYDA SAI PRASANNA(17H71A05A2)

VANUMU YASHWANTH(17H71A05C0)

DECLARATION

We **MALLELA RAJESH (17H71A0595), KONDLA MOUNIKA (17H71A0587), VYDA SAI PRASANNA (17H71A05A2) and VANUMU YASHWANTH (17H71A05C0)** of the main-Project "**AUTOMATIC ATTENDANCE USING FACE RECOGNITION**", hereby declare that the matter embodied in this Project is the genuine work done by us and has not been submitted either to this university or to any other university/institute for the fulfilment of the requirement of any course of study.

MALLELA RAJESH	**(17H71A0595)**
KONDLA MOUNIKA	**(17H71A0587)**
VYDA SAI PRASANNA	**(17H71A05A2)**
VANUMU YASHWANTH	**(17H71A05C0)**

CONTENTSPAGE NO

ABSTRACT

Facial Recognition is a technology of biometrics has been used in many areas like security systems, human machine interaction and image processing techniques. The main objective of this paper is to offer system that simplify and atomate the process of recording and tracking student attendance in the easier way. We proposed a system called automated attendance management system that uses face recognition method gives solution to the faculty thereby reducing the burden in taking attendance. The system used to calculate attendance automatically by recognizing the facial dimensions. An efficient face Recognition based attendance system has been developed by improving the efficiency of the system and also for the secured attendance. The algorithm used in this system is History of Oriented Gradients. The system is not only detecting the faces but also the distance of the facial characters under varying conditions. The proposed system provides the success rate at face recognition is around 93% to 95% and face identification is 99% and gives better result than the existing methods.

CHAPTER 1

INTRODUCTION

This is a project about Facial Recognition-Based Attendance System for Educational Institutions.In this chapter, the problem and motivation, research objectives, project scope, project contributions & the background information of the project will be discussed in detail.

1.1Problem Statement and Motivation

According to the previous attendance management system, the accuracy of the data collected is the biggest issue. This is because the attendance might not be recorded personally by the original person, in another word, the attendance of a particular person can be taken by a third party without the realization of the institution which violates the accuracy of the data. For example, student A is lazy to attend a particular class, so student B helped him/her to sign for the attendance which in fact student A didn't attend the class, but the system overlooked this matter due to no enforcement practiced. Supposing the institution establish an enforcement, it might need to waste a lot of human resource and time which in turn will not be practical at all. Thus, all the recorded attendance in the previous system is not reliable for analysis usage. The second problem of the previous system is where it is too time consuming. Assuming the time taken for a student to sign his/her attendance on a 3-4 paged name list is approximately 1 minute. In 1 hour, only approximately 60 students can sign their attendance which is obviously inefficient and time consuming. The third issue is with the accessibility of those information by the legitimate concerned party. For an example, most of the parents are very concerned to track their child's actual whereabouts to ensure their kid really attend the classes in college/school. However, in the previous system, there are no ways for the parents to access such information.

Therefore, evolution is needed to be done to the previous system to improve efficiency, data accuracy and provides accessibility to the information for those legitimate party.

1.2Research Objectives:

In order to solve the drawbacks of the previous system stated in 1.1, the existing system will need to evolve. The proposed system will reduce the paperwork where attendance will no longer involve any manual recording. The new system will also reduce the total time needed to do attendance recording. The new system will acquire individual attendance by means of facial recognition to secure data accuracy of the attendance.

The following are objectives of the project:

- To develop a portable Automatic Attendance System which is handy and self-powered.
- To ensure the speed of the attendance recording process is faster than the previous system which can go as fast as approximately 3 second for each student.
- Have enough memory space to store the database.
- Able to recognize the face of an individual accurately based on the face database.
- Allow parents to track their child's attendance.
- Develop a database for the attendance management system.
- Provide a user-friendly interface for admins to access the attendance database and for non-admins (parents) to check their child's attendance by mailing the attendance.
- Allow new students or staff to store their faces in the database by using a GUI.
- Able to show an indication to the user whether the face- recognition process is successful or not.

1.3 Project Scope and Direction

The main intention of this project is to solve the issues encountered in the old attendance system while reproducing a brand new innovative smart system that can provide convenience to the institution. In this project, an application will be developed which is capable of recognising the identity of each individuals and eventually record down the data into a database system. Apart from that, an excel sheet is created which shows the students attendance and is directly mailed to the respected faculty.

The followings are the project scopes:

- The targeted groups of the attendance monitoring system are the students and staff of an educational institution.
- The database of the attendance management system can hold up to 2000 individual's information.
- The facial recognition process can only be done for 1 person at a time.
- An excel sheet is created which contains the student attendance and is mailed to the respected faculty.
- The project has to work under a Wi-Fi coverage area or under Ethernet connection, as the system need to update the database of the attendance system constantly.
- The device on which the application is running is powered up by power bank to improve the portability of the application.

1.4 Impact, Significance and contributions

Many attendance management systems that exist nowadays are lack of efficiency and information sharing. Therefore, in this project, those limitations will be overcome and also further improved and are as follows:

Students will be more punctual on attending classes. This is due to the attendance of a student can only be taken personally where any absentees will be noticed by the system. This can not only train the student to be punctual as well as avoids any immoral ethics such as signing the attendance for their friends.

- The institution can save a lot of resources as enforcement are now done by means of technology rather than human supervision which will waste a lot of human resource for an insignificant process.
- The application can operate on any device at any location as long as there is Wi-Fi coverage or Ethernet connection which makes the attendance system to be portable to be placed at any intended location. For an example, the device can be placed at the entrance of the classroom to take the attendance.
- It saves a lot of cost in the sense that it had eliminated the paperwork completely.
- The system is also time effective in this all calculations are automated in shortthe project is developed to solve the existing issues in the old attendance system.

1.5 Historical development prior to the project

Back in the years, attendance management system in school/colleges was done by manual reporting where the student's attendance was recorded by placing a mark or signature beside their name in a name list to indicate their presence in a particular class. While the staff in the institution willreport their attendance through the punch card machine which also have to be done manually. Later on, some of those attendance systems had evolved into using smart cards to replace signature markings where each students/staff will be required to report their attendance using a smart card embedded with a unique identification chip.

CHAPTER 2

LITERATURE SURVEY

The second research journals “Face Recognition Based Attendance Marking System” (SenthamilSelvi, Chitrakala, Antony Jenitha, 2014) is based on the identification of face recognition to solve the previous attendance system’s issues. This system uses camera to capture the images of the employee to do face detection and recognition. The captured image is compared one by one with the face database to search for the worker’s face where attendance will be marked when a result is found in the face database. The main advantage of this system is where attendance is marked on the server which is highly secure where no one can mark the attendance of other. Moreover, in this proposed system, the face detection algorithm is improved by using the skin classification technique to increase the accuracy of the detection process. Although more efforts are invested in the accuracy of the face detection algorithm, the system is yet not portable. This system requires a standalone computer which will need a constant power supply that makes it not portable. This type of system is only suitable for marking staff’s attendance as they only need to report their presence once a day, unlike students which require to report their attendance at every class on a particular day, it will be inconvenient if the attendance marking system is not portable. Thus, to solve this issue, the whole attendance management system can be developed on an portable module so that it can be work just by executing the python program.

CHAPTER 3

SYSTEM ANALYSIS

3.1 Existing System

There are so many systems available in the present to take the attendance. They are RFID based system ,Finger Print based system, Iris based system and wireless based system and Face recognition using Deep learning, Machine learning etc.

3.1.2 Limitations of Existing System

In these techniques, the RFID based system we observe the drawback of Fraudulent usage and finger print based system having the drawback of Time consuming for students to wait and give their attendance and Iris based having the drawback of Invades the privacy of the user and wireless based having the drawback of Poor performance if topology is bad and finally face recognition using machine learning having the drawback of using complex algorithms to process the data so we difficult to learn and understand the problem and also whenever error is occurred it is difficult to overcome that error.

3.1.3 Drawbacks of Existing System

The Existing System RFID drawbacks are:

- Materials like metal & liquid can impact signal.
- Sometimes not as accurate or reliable as barcode scanners.
- Cost – **RFID** readers can be 10x more expensive than barcode readers.
- Implementation can be difficult & time consuming.

The Existing Finger Print System drawbacks are:

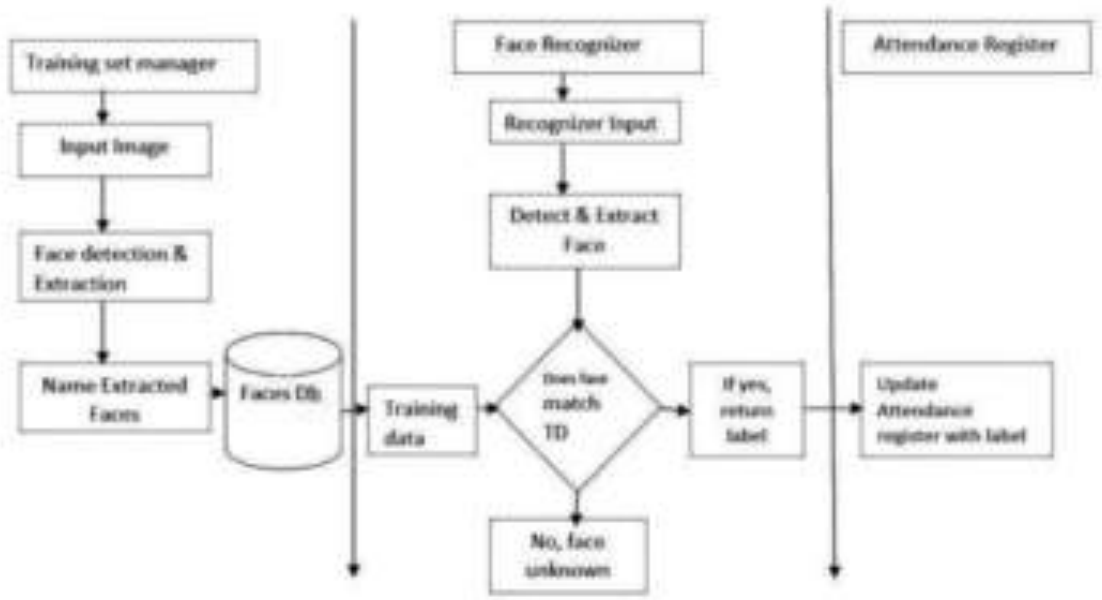
- The sample of DNA can easily be ruined during the process of DNA **fingerprinting**, causing the sample to become completely useless for testing.
- The process itself is complex and tedious, and can give results that may be hard to interpret.

The Existing Systems images are displayed below in order to have a better understanding of automatic attendance system using facial recognition in python.

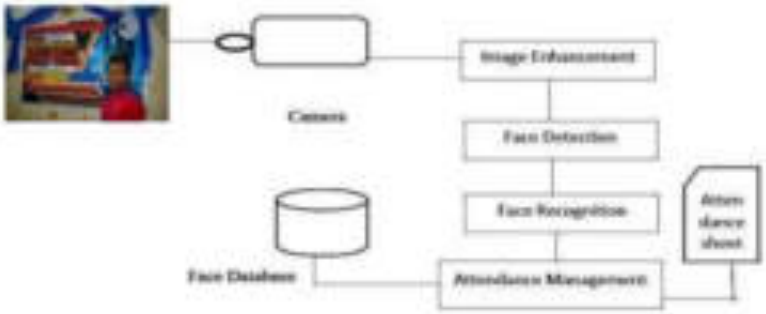
3.2 Proposed System

In our proposed system we are implemented Automatic AttendanceThrough Face Recognition using Python. And by using the python libraries such as Face Recognition library, Open CV, NumPy and History OfOriented Gradients Algorithm which is easy to implement and gives fast and accurate result when compared to existing system. This system creates a good platform for faculty to take the attendance in an easier way, and which helps them to reduce their work and have a long time for delivering their lecture to the students very effectively.

3.2.1 Architecture of Proposed System



The above image is the architecture of the proposed system for facial recognition using python in order to create the most effective system for attendance taking through facial recognition.



3.2.2 Merits of Proposed System

- To increase performance.
- Has less Computational time.
- Easy to implement and Easy to understand
- Fast and Accurate result

CHAPTER 4

SYSTEM DESIGN

4.1 SOFTWARE REQUIREMENT SPECIFICATIONS

Software requirements specification is a rigorous assessment of requirements before the more specific system design stages, and its goal is to reduce later redesign. It should also provide a realistic basis for estimating product costs, risks, and schedules. Used appropriately, software requirements specifications can help prevent software project failure. The software requirements specification document lists sufficient and necessary requirements for the project development. To derive the requirements, the developer needs to have clear and thorough understanding of the products under development.

4.1.1 LANGUAGE USED

Python (version 3.8.5) is a general purpose, dynamic, high level and interpreted programming language. It supports Object Oriented programming approach to develop applications. It is simple and easy to learn and provides lots of high-level datastructures.

- Python is easy to learn yet powerful and versatile scripting language which makes it attractive for Application Development.
- Python's syntax and dynamic typing with its interpreted nature, makes it an ideal language for scripting and rapid application development.
- Python supports multiple programming pattern, including object oriented, imperative and functional or procedural programming styles.
- Python is not intended to work on special area such as web programming. That is why it is known as multipurpose because it can be used with web, enterprise, 3D CAD etc.

4.1.2 IDE's(Integrated Development Environment)

PyCharm is a dedicated Python Integrated Development Environment (IDE) providing a wide range of essential tools for Python developers, tightly integrated to create a convenient environment for productive Python, Web, and data science development.

PyCharm is available in three editions:

- **Community** (free and open-sourced): for smart and intelligent Python development, including code assistance, refactorings, visual debugging, and version control integration.
- **Professional** (paid) : for professional Python, web, and data science development, including code assistance, refactorings, visual debugging, version control integration, remote configurations, deployment, support for popular web frameworks, such as Django and Flask, database support, scientific tools (including Jupyter notebook support), big data tools.
- **Edu** (free and open-sourced): for learning programming languages and related technologies with integrated educational tools.

There is a multitude of reasons for this, including the fact that it is developed by JetBrains, the developer behind the popular IntelliJ IDEA IDE that is one of the big 3 of Java IDEs and the "smartest JavaScript IDE" WebStorm. Having the support for web development by leveraging Django is yet another credible reason.

- There are a galore of factors that make PyCharm one of the most complete and comprehensive integrated development environments for working with the Python programming language.
- Before proceeding further into exploring the know-how of PyCharm i.e., features, installation, and pros & cons, let's first get a brief introduction to PyCharm.

What is PyCharm?

- Available as a cross-platform application, PyCharm is compatible with Linux, macOS, and Windows platforms. Sitting gracefully among the best Python IDEs, PyCharm provides support for both Python 2 (2.7) and Python 3 (3.5 and above) versions.

PyCharm comes with a plethora of modules, packages, and tools to hasten Python development while cutting-down the effort required to do the same to a great extent, simultaneously. Further, PyCharm can be customized as per the development requirements, and personal preferences call for. It was released to the public for the very first time back in February of 2010. In addition to offering code analysis, PyCharm features:

- A graphical debugger
- An integrated unit tester
- Integration support for version control systems (VCSs)
- Support for data science with Anaconda

What is PyCharm used for?

The main reason Pycharm for the creation of this IDE was for Python programming, and to operate across multiple platforms like Windows, Linux, and macOS. The IDE comprises code analysis tools, debugger, testing tools, and also version control options. It also assists developers in building Python plugins with the help of various APIs available. The IDE allows us to work with several databases directly without getting it integrated with other tools. Although it is specially designed for Python, HTML, CSS, and Javascript files can also be created with this IDE. It also comes with a beautiful user interface that can be customized according to the needs using plugins.

What is an IDE?

PyCharm is an extremely popular Python IDE. An Integrated Development Environment or IDE features a code editor and a compiler for writing and compiling programs in one or many programming languages.

Furthermore, an IDE comes with a galore of features that facilitate comprehensive software development. As an IDE allocates different colors to different programming entities, typically known as syntax highlighting, it becomes more accessible to:

- Differentiate between various programming entities, such as a class and a function, and to spot them.
- Look for the wrong keywords.
- Read and comprehend the code.

Most IDEs feature an auto-complete feature that produces suggestions when writing code. This makes writing code more efficient, quick, and less prone to errors and typos. Other standard features offered by a modern IDE are:

- Project editor window for efficiently managing and organizing files necessary for a program/project.
- Inspecting the output of the code written using the output window
- Suggestions for resolving errors and warnings
- A range of modules and packages readily available at a single place

PyCharm Pricing Model

PyCharm is offered in three variants:

- A freemium version dubbed The Community Edition available under the Apache License.
- A commercial version labeled Professional Edition available under a proprietary license.
- A free-to-use educational version dubbed the Edu Edition, aimed for students and professionals interested in learning Python, available under the Apache License.

Features of PyCharm

1. Intelligent Code Editor

PyCharm comes with a smart code editor that facilitates writing high-quality Python code. It offers an enhanced level of code comprehension and readability by means of distinct color schemes for keywords, classes, and functions, i.e., syntax and error highlighting.

In addition to offering the smart code completion feature, the code editor generates instructions for completing the current code. Identifying errors and issues is much more comfortable, along with linter integration and quick fixes.

2. Availability of Integration Tools

PyCharm provides support for integrating a range of tools. These tools vary from helping in enhancing the code productivity to facilitate dealing with data science projects. Some of the most essential integration tools available for PyCharm include:

- Anaconda - A free and open-source Python distribution geared towards scientific computing with simplified package management and deployment.
- IPython - A robust command shell for interactive computing.
- Kite - An AI-powered autocomplete plugin.
- Pylint - A source-code, bug, and quality checker.
- pytest - A framework for writing small tests for Python code.
- WakaTime - A developer dashboard with productivity metrics and automatic time tracking

3. Data Science and Machine Learning [Professional Edition Only]

PyCharm comes with support for scientific libraries, such as Matplotlib and SciPy, to help Python developers accomplish data science and machine learning projects.

4. **Google App Engine [Professional Edition Only]**

Google App Engine, or directly App Engine, is a PaaS and cloud computing platform targeted towards developing and hosting web applications. It offers automatic scaling for web applications. The professional edition of PyCharm provides support for Google App Engine.

5. Integrated Debugging and Testing

An IDE comes with support for debugging and testing programs. To accomplish the same, PyCharm features an integrated Python debugger and integrated unit testing with line-by-line code coverage.

6. Multi-technology Development [Professional Edition Only]

Python developers can also use PyCharm for creating web applications. As such, the Python IDE provides support for popular web technologies, including CoffeeScript, CSS, HTML, JavaScript, TypeScript. Additionally, it also includes support for Cython, template languages, and SQL.

Live editing is also available in PyCharm, i.e., developers can create/modify a web page while pushing it live simultaneously. Hence, changes can be followed directly on a web browser. Building web applications using AngularJS or NodeJS is also available.

7. Project and Code Navigation

The code navigation feature makes it much easier for developers to navigate to a class, function, or file. It also helps in significantly cutting-down effort and time required to edit and enhance the Python code. File structure views and specialized project views are readily available.

The lens mode allows a developer to inspect and debug the entire Python source code thoroughly. With code navigation, locating an element, variable, etc. is done in almost no time. Developers can quickly jump between classes, files, and methods.

8. Refactoring

The refactoring feature in PyCharm helps in improving the internal structure of a Python program without affecting the external performance of the same. Making changes to both local and global variables is efficient and fast.

The extract method is also there to split up extended classes and functions. Other useful code refactoring features include:

- Introduce constant
- Introduce variable
- Pull up
- Push down
- Rename

9. Remote Development

PyCharm allows running, debugging, testing, and deploying applications on remote hosts or virtual machines. For the purpose, the Python IDE offers:

- An integrated SSH terminal
- Docker and Vagrant integration
- Remote interpreters

10. Support for Popular Python Web Frameworks [Professional Edition Only]

PyCharm lets developers leverage Django in their Python development projects. The Python IDE offers the autocomplete feature and generates suggestions for Django.

Debugging code written using Django is also available. PyCharm also provides support for other popular Python frameworks, namely Flask, Pyramid, and web2py.

11. Version Control Systems (VCSs) Integration

In its simplicity, a version control system (VCS) keeps track of the changes made to files, applications, and other sources of information. It can be considered as a database of changes.

PyCharm provides a unified user interface for CVS, Git, Mercurial, Perforce, and Subversion.

Other Pycharm Features

Other than the aforementioned features, PyCharm offers the following capabilities:

- Code generation for generating language-specific code constructs.
- Code reference information for instantly accessing API documentation, hints on using various programming entities, etc
- File templates for creating scripts, stub classes, etc
- Import assistance for importing missing libraries
- Intention actions and quick fixes for optimizing code
- Language-specific tools for developing, running, testing, and deploying applications
- Language injections to work with supported languages inside attributes, tags, or string literals
- Live templates for expanding abbreviations into complicated code constructs

Installing and Setting Up PyCharm

Minimum System Requirements

Memory - 4GB
Storage Space - 2.5GB (main) + 1GB (caches)
Resolution - 1024x768
OS - 64-bit version of macOS 10.11/Microsoft Windows 7 SP1/any Linux distribution supporting Gnome, KDE, or Unity DE

Recommended System Requirements

Memory - 8GB
Storage Space - 5GB of SSD
Resolution - 1920x1080
OS - Any latest 64-bit version of macOS/Microsoft Windows/Linux

One of the significant advantages is that PyCharm brings to the Python development table is easy installation and less setup time. There are three modes of installation available for PyCharm:

1. Standalone Installation

Here is how to install the popular Python IDE in the traditional way:

- Step #01 - Go to https://www.jetbrains.com/pycharm/download.
- Step #02 - Choose a platform among Windows, Mac, and Linux.

For Windows

- Step #03 - Choose the PyCharm edition; Professional (paid), or Community (free).
- Step #04 - Hit the Download button to start downloading the .exe installer.
- Step #05 - Once downloaded, successfully run the installer and follow the setup wizard steps.

4.1.3 LIBRARIES

Face Recognition:

Recognize and manipulate faces from Python or from the command line with the world's simplest face recognition library.Built usingdlib's state-of-the-art face recognitionbuilt with deep learning. The model has an accuracy of 99.38% on the Labelled faces in the wild benchmark.This also provides a simple command line tool that letsyou do face recognition on a folder of images from the command line.The face_recognitionlibrary, created by **Adam Geitgey**, wraps around dlib's facial recognition functionality, making it easier to work with.

Before we can recognize faces in images and videos, we first need to quantify the faces in our training set. Keep in mind that we are not actually training a network here — **the network has already been trained to create 128-d embeddings** on a dataset of ~3 million images.

We certainly *could* train a network from scratch or even fine-tune the weights of an existing model but that is more than likely overkill for many projects. Furthermore, you would need a *lot* of images to train the network from scratch.

Instead, it's easier to use the pre-trained network and then use it to construct 128-d embeddings for each of the 218 faces in our dataset.

Then, during classification, we can use a simple k-NN model + votes to make the final face classification. Other traditional machine learning models can be used here as well.

face_recognition library in Python can perform a large number of tasks:

- Find all the faces in a given image
- Find and manipulate facial features in an image
- Identify faces in images
- Real-time face recognition

The face_recognition library, created by Adam Geitgey, wraps around dlib's facial recognition functionality, and this library is super easy to work with and we will be using this in our code. Remember to install dlib library first before you install face_recognition.

- To install OpenCV, type in command prompt

pip install opencv-python

- I have tried various ways to install dlib on Windows but the easiest of all of them is via Anaconda. First, install Anaconda (here is a guide to install it) and then use this command in your command prompt:

conda install -c conda-forge dlib`or pip install dlib`

- Next to install face_recognition, type in command prompt

pip install face_recognition

- Now that we have all the dependencies installed, let us start coding. We will have to create three files, one will take our dataset and extract face embedding for each face using dlib. Next, we will save these embedding in a file.

- In the next file we will compare the faces with the existing the recognise faces in images and next we will do the same but recognise faces in live webcam feed

- *Extracting features from Face*

- First, you need to get a dataset or even create one of you own. Just make sure to arrange all images in folders with each folder containing images of just one person.

In face recognition itself we are using svm classifier for generating 180+ images at a time lets get to know about that in detail

Machine learning involves predicting and classifying data and to do so we employ various machine learning algorithms according to the dataset.

SVM or Support Vector Machine is a linear model for classification and regression problems. It can solve linear and non-linear problems and work well for many practical problems. The idea of SVM is simple: The algorithm creates a line or a hyperplane which separates the data into classes.

In this blog post I plan on offering a high-level overview of SVMs. I will talk about the theory behind SVMs, it's application for non-linearly separable datasets and a quick example of implementation of SVMs in Python as well. In the upcoming articles I will explore the maths behind the algorithm and dig under the hood.

THEORY

At first approximation what SVMs do is to find a separating line(or hyperplane) between data of two classes. SVM is an algorithm that takes the data as an input and outputs a line that separates those classes if possible.

Lets begin with a problem. Suppose you have a dataset as shown below and you need to classify the red rectangles from the blue ellipses(let's say positives from the negatives). So your task is to find an ideal line that separates this dataset in two classes (say red and blue).

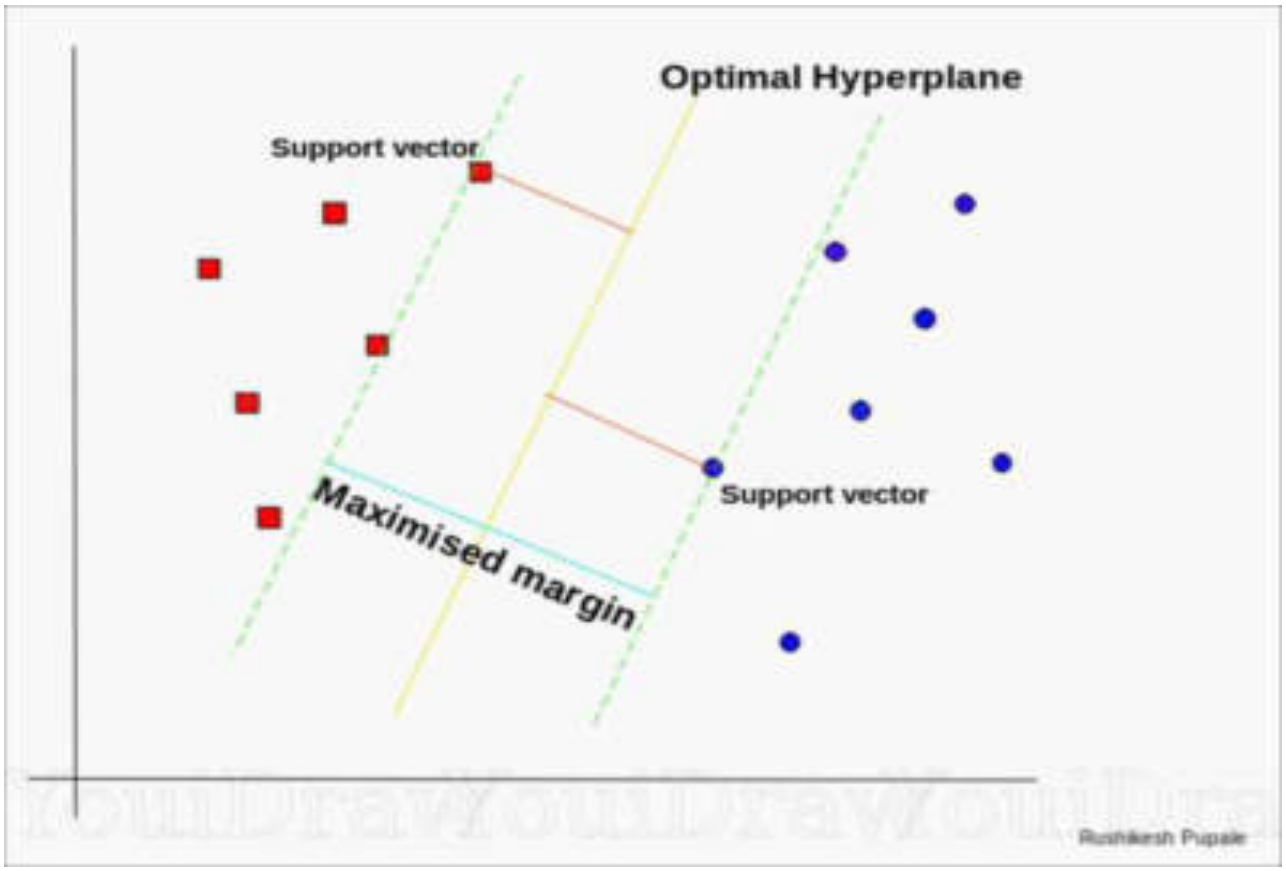
We have two candidates here, the green colored line and the yellow colored line. Which line according to you best separates the data?

If you selected the yellow line then congrats, because thats the line we are looking for. It's visually quite intuitive in this case that the yellow line classifies better. But, we need something concrete to fix our line.

The green line in the image above is quite close to the red class. Though it classifies the current datasets it is not a generalized line and in machine learning our goal is to get a more generalized separator.

SVM's way to find the best line

According to the SVM algorithm we find the points closest to the line from both the classes.These points are called support vectors. Now, we compute the distance between the line and the support vectors. This distance is called the margin. Our goal is to maximize the margin. The hyperplane for which the margin is maximum is the optimal hyperplane.



Optimal Hyperplane using svm algorithm

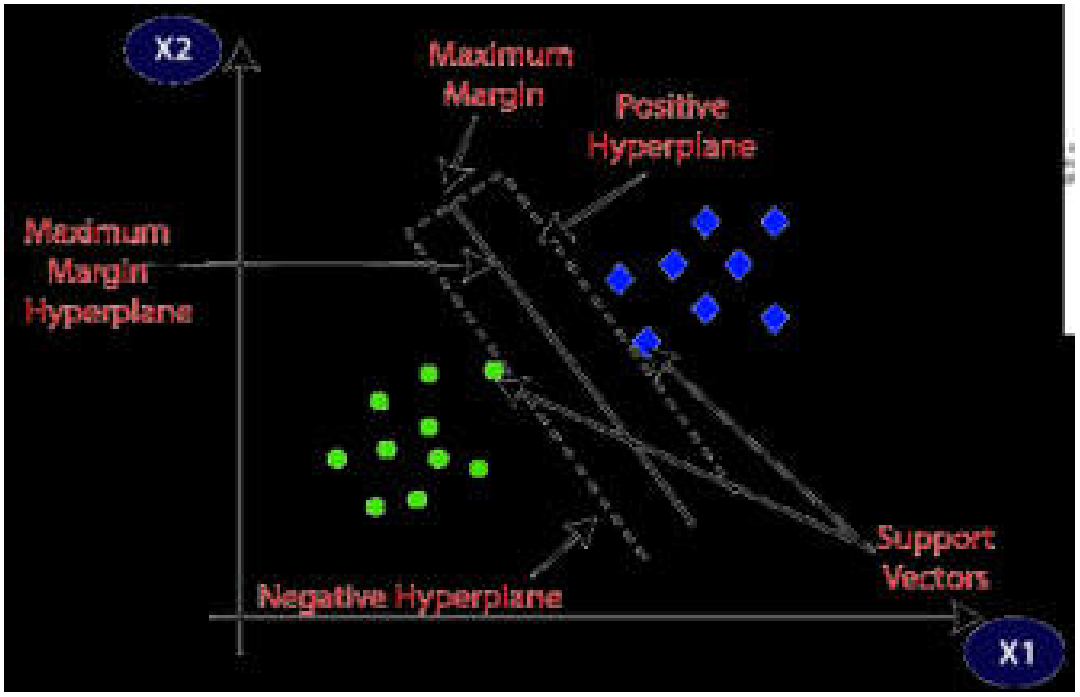
SVMs can be used to solve various real-world problems:

- SVMs are helpful in text and hypertext categorization, as their application can significantly reduce the need for labeled training instances in both the standard inductive and transductive settings.[7] Some methods for shallow semantic parsing are based on support vector machines.[8]
- Classification of images can also be performed using SVMs. Experimental results show that SVMs achieve significantly higher search accuracy than traditional query refinement schemes after just three to four rounds of relevance feedback. This is also

true for image segmentation systems, including those using a modified version SVM that uses the privileged approach as suggested by Vapnik.[9][10]

- Classification of satellite data like SAR data using supervised SVM.[11]
- Hand-written characters can be recognized using SVM.[12][13]
- The SVM algorithm has been widely applied in the biological and other sciences. They have been used to classify proteins with up to 90% of the compounds classified correctly. Permutation tests based on SVM weights have been suggested as a mechanism for interpretation of SVM models.[14][15] Support-vector machine weights have also been used to interpret SVM models in the past.[16] Posthoc interpretation of support-vector machine models in order to identify features used by the model to make predictions is a relatively new area of research with special significance in the biological sciences.

SVM can be of two types:



- **Linear SVM:** Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.
- **Non-linear SVM:** Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.

Hyperplane and Support Vectors in the SVM algorithm:

Hyperplane: There can be multiple lines/decision boundaries to segregate the classes in n-dimensional space, but we need to find out the best decision boundary that helps to classify the data points. This best boundary is known as the hyperplane of SVM.

The dimensions of the hyperplane depend on the features present in the dataset, which means if there are 2 features (as shown in image), then hyperplane will be a straight line. And if there are 3 features, then hyperplane will be a 2-dimension plane.

We always create a hyperplane that has a maximum margin, which means the maximum distance between the data points.

Support Vectors:

The data points or vectors that are the closest to the hyperplane and which affect the position of the hyperplane are termed as Support Vector. Since these vectors support the hyperplane, hence called a Support vector.

Numpy:

NumPy is the fundamental package for scientific computing with Python. It contains among other things:

- a powerful N-dimensional array object
- sophisticated (broadcasting) functions
- tools for integrating C/C++ and Fortran code
- useful linear algebra, Fourier transform, and random number capabilities
- Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined. This allows NumPy to seamlessly and speedily integrate with a wide variety of databases.
- NumPy is licensed under the BSD license, enabling reuse with few restrictions.

NumPy is meant for creating ***homogeneous*** n-dimensional arrays (n = 1..n). Unlike Python lists, all elements of a NumPy array should be of same type. so the following code is not valid if data type is provided

1. NumPy uses much less memory to store data

The NumPy arrays takes significantly less amount of memory as compared to python lists. It also provides a mechanism of specifying the data types of the contents, which allows further optimisation of the code.

As an example, we can create a simple array of six elements using a python list as well as by using numpy.array(...) , The difference in amount of memory occupied by it is quite astounding.

If this difference seems intimidating then prepare to have more. It's possible to do further optimisation on the NumPy arrays if we know the maximum number of individual data we're planning to have.

In the example above, NumPy by default considers these integers as 8 Bytes integers, however, we can provide data types with NumPy arrays if we know the maximum range of the data. For example, we can use 1 Byte integer for storing numbers upto 255 and 2 Bytes integer for numbers upto 65535

The optimized memory allocation is not limited to storing numbers, it also expands to storing strings. However, since NumPy arrays are mostly used for mathematical calculations, string will be stored and used rarely. Nevertheless, it's good to know if the requirement arises

2. Using NumPy for creating n-dimension arrays

An n-dimension array is generally used for creating a matrix or tensors, again mainly for the mathematical calculation purpose. Compare to python list base n-dimension arrays, NumPy not only saves the memory usage, it provide a significant number of additional benefits which makes it easy to mathematical calculations

Here is a list of things we can do with NumPy n-dimensional arrays which is otherwise difficult to do.

The dimensions of the array can be changed at runtime as long as the multiplicity factor produces the same number of elements. For example, a 2 * 5 matrix can be converted into 5 * 2 and a 1 * 4 into 2 * 2. This can be done by calling the NumPy .reshape(...) function on the arrays.

NumPy can also generate a predefined set of number for an array. A predefined set of numbers with steps (including decimal steps can be generated using numpy.arange(...) function. The output of this function will always be a single dimension set of numbers. However, we can use reshape on this output to generate dimension of our choice.

NumPy provides the API for creating n-dimension arrays using pre-filled ones and zeros where all members of the matrix are either zero or one. One of the most probable usage of this is to create a Sparse or Dense matrix for machine learning .

By default, NumPy uses floating point float64 data type for creating ones and zeros, however, the data type can be changed to integer using dtype option as depicted below

```
Importnumpyas
np

np_array=np.zeros(10)
# Generates => [0. 0. 0. 0. 0. 0. 0. 0. 0.
0.]
np_array=np.ones(10)
# Generates => [1. 1. 1. 1. 1. 1. 1. 1. 1.
1.]
np_array=np.zeros((2,5))
"""
Generates the following
[[0. 0. 0. 0. 0.]
[0. 0. 0. 0. 0.]]
"""
np_array=np.ones((2,5))
"""
Generates the following
[[1. 1. 1. 1. 1.]
 [1. 1. 1. 1. 1.]]
"""
# Changing the datatype as Integer
np_array=np.ones((2,5), dtype='Int32')
"""
Generates the following
[[1 1 1 1 1]
 [1 1 1 1 1]]
```

As .reshape(x,y) can convert an array into multi dimensional array, similarly, its possible to create a single dimensional array from any anymulti dimensional array using the function .ravel()

3. Mathematical operations on NumPy n-Dimension Arrays

As stated earlier, NumPy is not only about efficient storing the data, it also makes it extremely easy to perform mathematical operations on it. Any actions on n-dimension arrays behaves exactly similar to mathematical operations.

NumPy n-dimensional arrays makes it extremely easy to perform mathematical operations on it

This is the main USP of NumPy because of which it's widely used in data analytic community. The python lists are nowhere near to what it can do.

4. Finding Elements in NumPy array

While working with data sets there will be times when we need to find specific data from the available data set. Though NumPy provides multiple functions for the same, three of them will be used more often than others. They are where, nonzeroand count_nonzero

Where and nonzero functions returns the index associated with the True statement.

Some important points about Numpy arrays:

- We can create a N-dimensional array in python using numpy.array().
- Array are by default Homogeneous, which means data inside an array must be of the same Datatype. (Note you can also create a structured array in python).
- Element wise operation is possible.
- Numpy array has the various function, methods, and variables, to ease our task of matrix computation.
- Elements of an array are stored contiguously in memory. For example, all rows of a two dimensioned array must have the same number of columns. Or a three dimensioned array must have the same number of rows and columns on each card.

OpenCv:

OpenCV-Python is a library of Python bindings designed to solve computer vision problems.Python is a general purpose programming language started by **Guido van Rossum** that became very popular very quickly, mainly because of its simplicity and code readability. It enables the programmer to express ideas in fewer lines of code without reducing readability.

Compared to languages like C/C++, Python is slower. That said, Python can be easily extended with C/C++, which allows us to write computationally intensive code in C/C++ and create Python wrappers that can be used as Python modules. This gives us two advantages: first, the code is as fast as the original C/C++ code (since it is the actual C++ code working in background) and second, it easier to code in Python than C/C++. OpenCV-Python is a Python wrapper for the original OpenCV C++ implementation.

OpenCV-Python makes use of **Numpy**, which is a highly optimized library for numerical operations with a MATLAB-style syntax. All the OpenCV array structures are converted to and from Numpy arrays. This also makes it easier to integrate with other libraries that use Numpy such as SciPy and Matplotlib.

In the early days of OpenCV, the goals of the project were described as:

- Advance vision research by providing not only open but also optimized code for basic vision infrastructure.
- No more reinventing the wheel.
- Disseminatevision knowledge by providing a common infrastructure that developers could build on, so that code would be more readily readable and transferable.
- Advance vision-based commercial applications by making portable, performance-optimized code available for free – with a license that did not require code to be open or free itself.

Versions of Open CV:

- Deep neural networks (DNN)The first alpha version of OpenCV was released to the public at the IEEE .
- Conference on Computer Vision and Pattern Recognition in 2000, and five betas were released between 2001 and 2005.
- The first 1.0 version was released in 2006. A version 1.1 "pre-release" was released in October 2008.
- The second major release of the OpenCV was in October 2009. OpenCV 2 includes major changes to the C++ interface, aiming at easier, more type-safe patterns, new functions, and better implementations for existing ones in terms of performance (especially on multi-core systems).
- Official releases now occur every six months and development is now done by an independent Russian team supported by commercial corporations.
- In August 2012, support for OpenCV was taken over by a non-profit foundation

OpenCV.org, which maintains a developer and user site.

- On May 2016, Intel signed an agreement to acquire It, a leading developer of OpenCV.

In the Artificial Intelligence field, Computer Vision is one of the most interesting and challenging tasks. Computer Vision acts as a bridge between Computer Software and visualizations. Computer Vision allows computer software to understand and learn about the visualizations in the *surroundings*.

Let us understand an example: Based on the shape, color, and size that determines the fruit. This task is very easy for the human brain but in the Computer Vision pipeline, first, we need to gather the data, then we perform the data processing operations, and then we train and teach the model to understand how to distinguish between the fruits based on its size, shape, and color of the fruit.

Nowadays, various packages are available to perform machine learning, deep learning, and computer vision problems. So far, computer vision is the best module for such complex problems. *OpenCV* is an open-source library. It is supported by different programming languages such as R, Python, etc. It runs probably on most platforms such as Windows, Linux, and macOS.

Advantages Of Open Cv:

1. Open CV is free of cost and an open-source library.

2. Open CV is fast as it is written in C/C++ language as compared to others

3. With less system RAM, OpenCV works better.

4. It supports most of the operating systems like Windows, Linux, and macOS.

In this section, we are going to implement face recognition using OpenCV and Python.

First, let us see what libraries we will need and how to install them:

1. OpenCV

2. dlib

3. Face_recognition

OpenCV is a video and image processing library and it is used for image and video analysis, like facial detection, license plate reading, photo editing, advanced robotic vision, and many more.

The **dlib** library contains our implementation of 'deep metric learning' which is used to construct our face embeddings used for the actual recognition process.

The **face_recognition** library is super easy to work with and we will be using this in our code. First, remember to install dlib library before you install face_recognition.

The face_recognition library, created by Adam Geitgey, wraps around dlib's facial recognition functionality, and this library is super easy to work with and we will be using this in our code. Remember to install dlib library first before you install face_recognition.

To install OpenCV, type in command prompt

pip install opencv-python

I have tried various ways to install dlib on Windows but the easiest of all of them is via Anaconda. First, install Anaconda (here is a guide to install it) and then use this command in your command prompt:

conda install -c conda-forge dlib

Next to install face_recognition, type in command prompt

pip install face_recognition

Now that we have all the dependencies installed, let us start coding. We will have to create three files, one will take our dataset and extract face embedding for each face using dlib. Next, we will save these embedding in a file.

In the next file we will compare the faces with the existing the recognise faces in images and next we will do the same but recognise faces in live webcam feed

Extracting features from Face

First, you need to get a dataset or even create one of you own. Just make sure to arrange all images in folders with each folder containing images of just one person.

Applications of OpenCV:

There are lots of applications which are solved using OpenCV, some of them are listed below:

1.face recognition

2.Automated inspection and surveillance

3.Number of people – count (foot traffic in a mall, etc)

4.Vehicle counting on highways along with their speeds

5.Interactive art installations

6.Anamoly (defect) detection in the manufacturing process (the odd defective products)

7.Street view image stitching

8.Video/image search and retrieval

9.Robot and driver-less car navigation and control

10.Object recognition.

4.1.4 ALGORITHM

HISTORY OF GRADIENTS ALGORITHM

Histogram of Oriented Gradients, or HOG for short, are descriptors mainly used in computer vision and machine learning for object detection. However, we can also use HOG descriptors for quantifying and representing both shape and texture.

HOG features were first introduced by Dalal and Triggs in their CVPR 2005 paper, Histogram of Oriented Gradients for Human Detection. In their work, Dalal and Triggs proposed HOG and a 5-stage descriptor to classify humans in still images.

The 5 stages include:

Stage 1: Normalizing the image prior to description.

Stage 2: Computing gradients in both the x and y directions.

Stage 3: Obtaining weighted votes in spatial and orientation cells.

Stage 4: Contrast normalizing overlapping spatial cells.

Stage 5: Collecting all Histograms of Oriented gradients to form the final feature vector.The most important parameters for the HOG descriptor are the orientations, pixelspercell, and the

cellsperblock . These three parameters (along with the size of the input image) effectively control the dimensionality of the resulting feature vector. We'll be reviewing these parameters and their implications later in this article.

In most real-world applications, HOG is used in conjunction with a Linear SVM to perform object detection. The reason HOG is utilized so heavily is because local object appearance and shape can be characterized using the distribution of local intensity gradients. In fact, these are the exact same image gradients that we learned about in the Gradients lesson, but only now we are going to take these image gradients and turn them into a robust and powerful image descriptor.

We'll be discussing the steps necessary to combine both HOG and a Linear SVM into an object classifier later in this course. But for now just understand that HOG is mainly used as a descriptor for object detection and that later these descriptors can be fed into a machine learning classifier.

HOG is implemented in both OpenCV and scikit-image. The OpenCV implementation is less flexible than the scikit-image implementation, and thus we will primarily used the scikit-image implementation throughout the rest of this course.

What are HOG descriptors used to describe?

HOG descriptors are mainly used to describe the structural shape and appearance of an object in an image, making them excellent descriptors for object classification. However, since HOG captures local intensity gradients and edge directions, it also makes for a good texture descriptor.

The HOG descriptor returns a real-valued feature vector. The dimensionality of this feature vector is dependent on the parameters chosen for the orientations ,pixels_per_cell , and cells_per_block parameters mentioned above.

How do HOG descriptors work?

The cornerstone of the HOG descriptor algorithm is that appearance of an object can be modeled by the distribution of intensity gradients inside rectangular regions of an image:

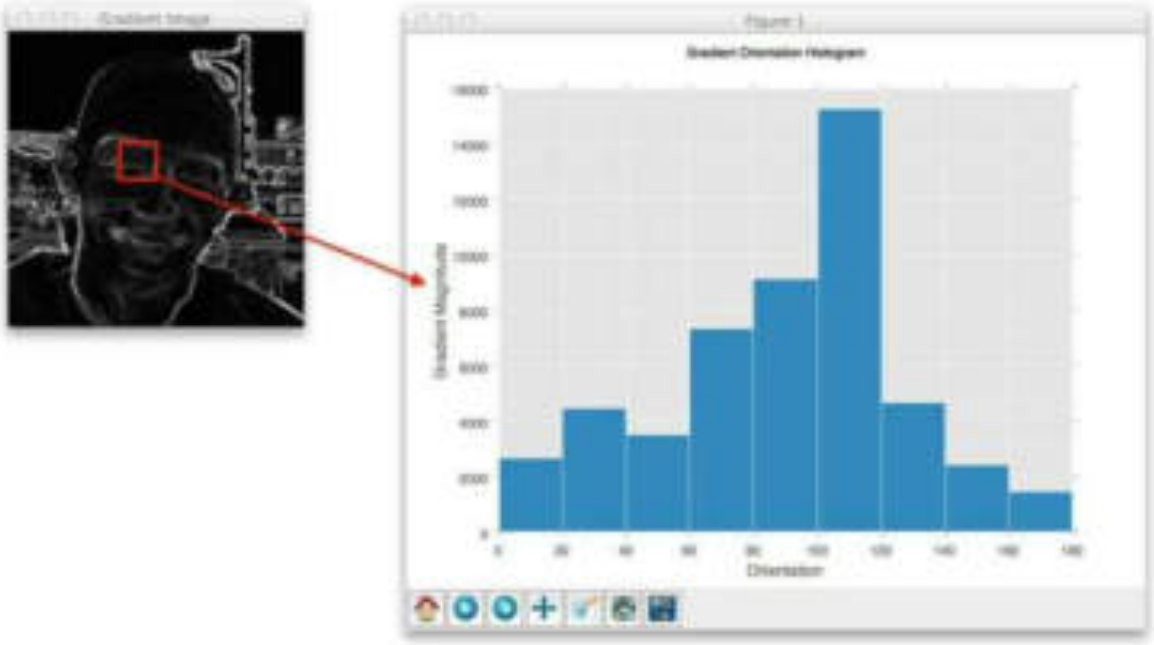


Figure1: An example of computing a histogram of oriented gradients for a particular region of an image

Implementing this descriptor requires dividing the image into small connected regions called cells, and then for each cell, computing a histogram of oriented gradients for the pixels within each cell. We can then accumulate these histograms across multiple cells to form our feature vector.

Dalal and Triggs also demonstrated that we can perform block normalization to improve performance. To perform block normalization we take groups of overlapping cells, concatenate their histograms, calculate a normalizing value, and then contrast normalize the histogram. By normalizing over multiple, overlapping blocks, the resulting descriptor is more robust to changes in illumination and shadowing. Furthermore, performing this type of normalization implies that each of the cells will be represented in the final feature vector multiple times but normalized by a slightly different set of neighboring cells.

Now, let's review each of the steps for computing the HOG descriptor.

Step 1: Normalizing the image prior to description.

This normalization step is entirely optional, but in some cases this step can improve performance of the HOG descriptor. There are three main normalization methods that we can consider:

Gamma/power law normalization:

In this case, we take the \log(p) of each pixel p in the input image. However, as Dalal and Triggs demonstrated, this approach is perhaps an “over-correction” and hurts performance.

Square-root normalization:

Here, we take the sqrt(p) of each pixel p in the input image. By definition, square-root normalization compresses the input pixel intensities far less than gamma normalization. And again, as Dalal and Triggs demonstrated, square-root normalization actually increases accuracy rather than hurts it.

Variance normalization:

A slightly less used form of normalization is variance normalization. Here, we compute both the mean \mu and standard deviation \sigma of the input image. All pixels are mean centered by subtracting the mean from the pixel intensity, and then normalized through dividing by the standard deviation: p' = (p - \mu) / \sigma. Dalal and Triggs do not report accuracy on variance normalization; however, it is a form of normalization that I like to perform and thought it was worth including.

In most cases, it’s best to start with either no normalization or square-root normalization. Variance normalization is also worth consideration, but in most cases it will perform in a similar manner to square-root normalization (at least in my experience).

Step 2:Gradient computation

The first actual step in the HOG descriptor is to compute the image gradient in both the x and y direction. These calculations should seem familiar, as we have already reviewed them in the Gradients lesson.

We’ll apply a convolution operation to obtain the gradient images:

G_{x} = I \star D_{x} and G_{y} = I \star D_{y}

where I is the input image, D_{x} is our filter in the x-direction, and D_{y} is our filter in the y-direction.

Compares the two images and gives the resultant output of the two images

Step 3:Weighted votes in each cell

Now that we have our gradient magnitude and orientation representations, we need to divide our image up into cells and blocks.

A “cell” is a rectangular region defined by the number of pixels that belong in each cell. For example, if we had a 128 x 128 image and defined our pixels_per_cell as 4 x 4, we would thus have 32 x 32 = 1024 cells:

Step 4: Contrast normalization over blocks

To account for changes in illumination and contrast, we can normalize the gradient values locally. This requires grouping the “cells” together into larger, connecting “blocks”. It is common for these blocks to overlap, meaning that each cell contributes to the final feature vector more than once

Here is an example of how to compute HOG descriptors using scikit-image:

```
from skimage import feature
H = feature.hog(logo, orientations=9, pixels_per_cell=(8, 8), cells_per_block=(2, 2),transform_sqrt=True, block_norm="L1")
```

We can also visualize the resulting HOG image:

```
from skimage import exposure
from skimage import feature
import cv2

(H, hogImage) = feature.hog(logo, orientations=9, pixels_per_cell=(8, 8),cells_per_block=(2, 2), transform_sqrt=True, block_norm="L1",visualize=True)
hogImage = exposure.rescale_intensity(hogImage, out_range=(0, 255))
hogImage = hogImage.astype("uint8")

cv2.imshow("HOG Image", hogImage)
```

(2017-11-28) Update for skimage: In scikit-image==0.12 , the normalise parameter has been updated to transform_sqrt . The transform_sqrt performs the exact same operation, only with a different name. If you're using an older version of scikit-image (again, before the v0.12 release), then you'll want to change transform_sqrt to normalise . In scikit-image==0.15 the default value of block_norm="L1″ has been deprecated and changed to block_norm="L2-Hys" . Therefore, for this lesson we'll explicitly specify block_norm="L1″ . Doing this will avoid it switching to "L2-Hys" with version updates without us knowing (and yielding incorrect car logo identification results). You can read about L1 and L2 norms here.

Advantages Of History Of Oriented Gradients

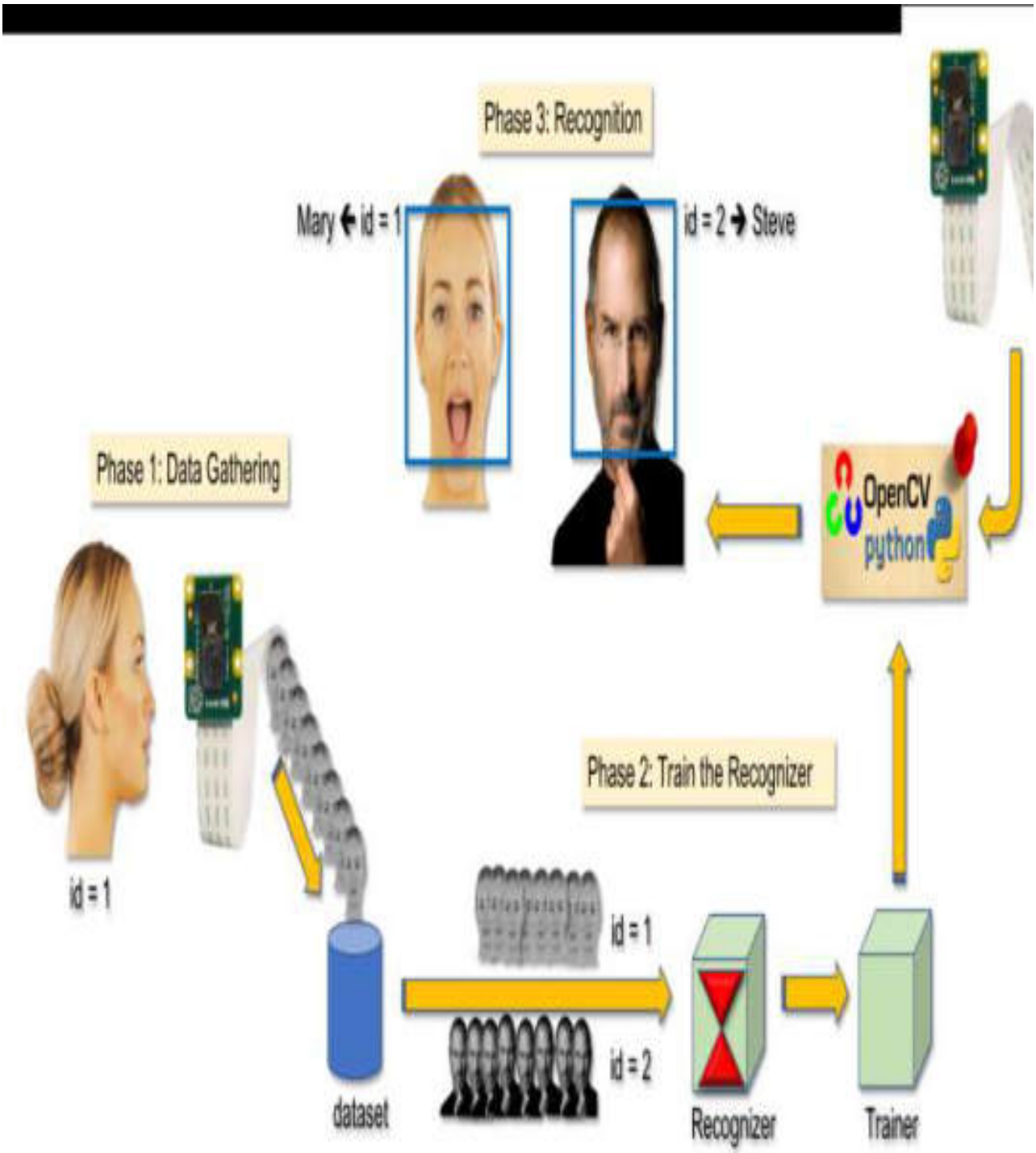
1.Very powerful descriptor.

2.Excellent at representing local appearance.

3.Extremely useful for representing structural objects that do not demonstrate substantial variation in form (i.e. buildings, people walking the street, bicycles leaning against a wall).

4.Very accurate for object classification.

Design Methodology for Face Recognition System



The above shown image just represents our methodology of taking attendance using face recognition and the predefined libraries will tell them how to recognize the face and all.

How Our Proposed System of giving attendance using face recognition will be shown below

4.2 UML DIAGRAMS

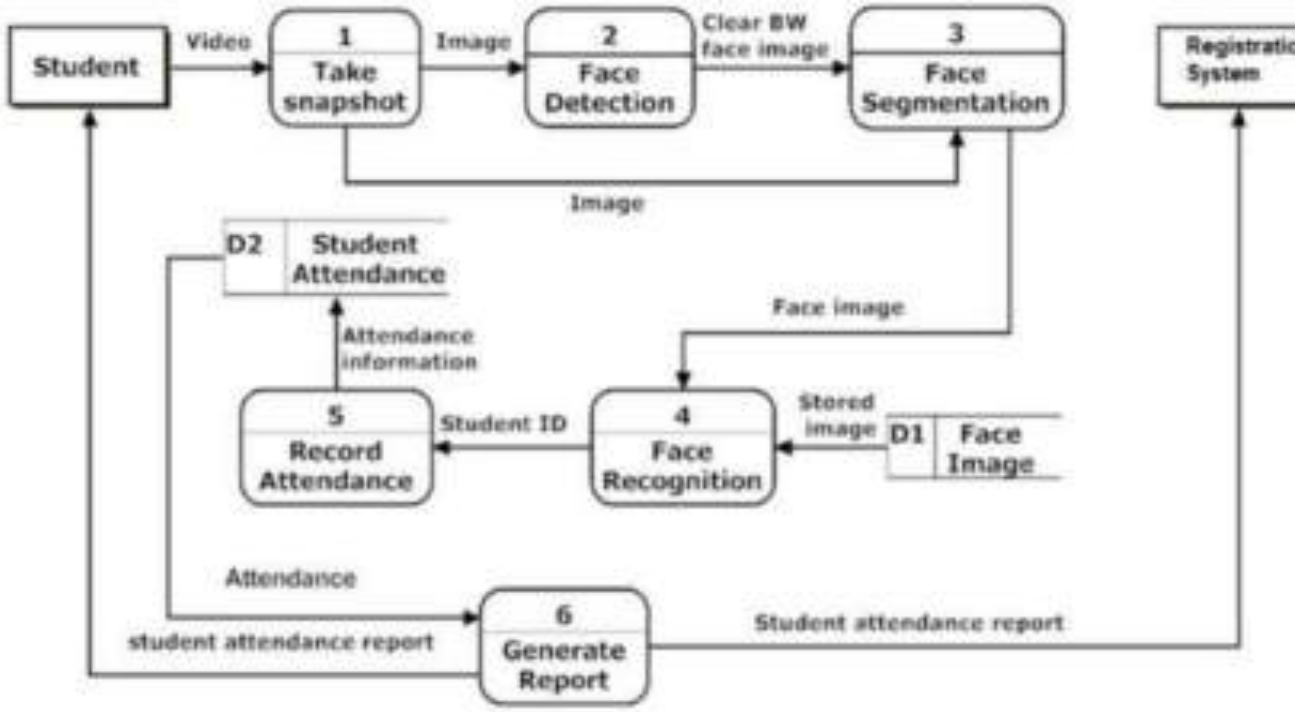
Use-oriented techniques are widely used in software requirement analysis and design. Use cases and usage scenarios facilitate system understanding and provide a common language for communication. This paper presents a scenario-based modeling technique and discusses its applications. In this model, scenarios are organized hierarchically and they capture the system functionality at various abstraction levels including scenario groups, scenarios, and sub-scenarios. Combining scenarios or sub-scenarios can form complex scenarios. Data are also separately identified, organized, and attached to scenarios. This scenario model can be used to cross check with the UML model. It can also direct systematic scenario-based testing including test case generation, test coverage analysis with respect to requirements, and functional regression testing.

4.2.1 CLASS DIAGRAM

Model, objects are entities that combine state (i.e., data), behavior (i.e., procedures, or methods) and identity (unique existence among all other objects). The structure and behavior of an object are defined by a class, which is a definition, or blueprint, of all objects of a specific type. An object

must be explicitly created based on a class and an object thus created is considered to be an instance of that class. An object is similar to a structure, with the addition of method pointers, member access control, and an implicit data member which locates instances of the class (i.e. actual objects of that class) in the class hierarchy (essential for runtime inheritance features) In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, and the relationships between the classes. The class diagram is the main building block in object oriented modeling. It is used both for general conceptual modeling of the semantics of the application, and for detailed modeling translating the models into programming code. The classes in a class diagram represent both the main objects and or interactions in the application and the objects to be programmed. In the class diagram these classes are represented with boxes which contain the two parts:

- The upper part holds the name of the class.
- The middle part contains the attributes of the class.
- The lower part contains the operations of the class.



4.2.2 USECASE DIAGRAM

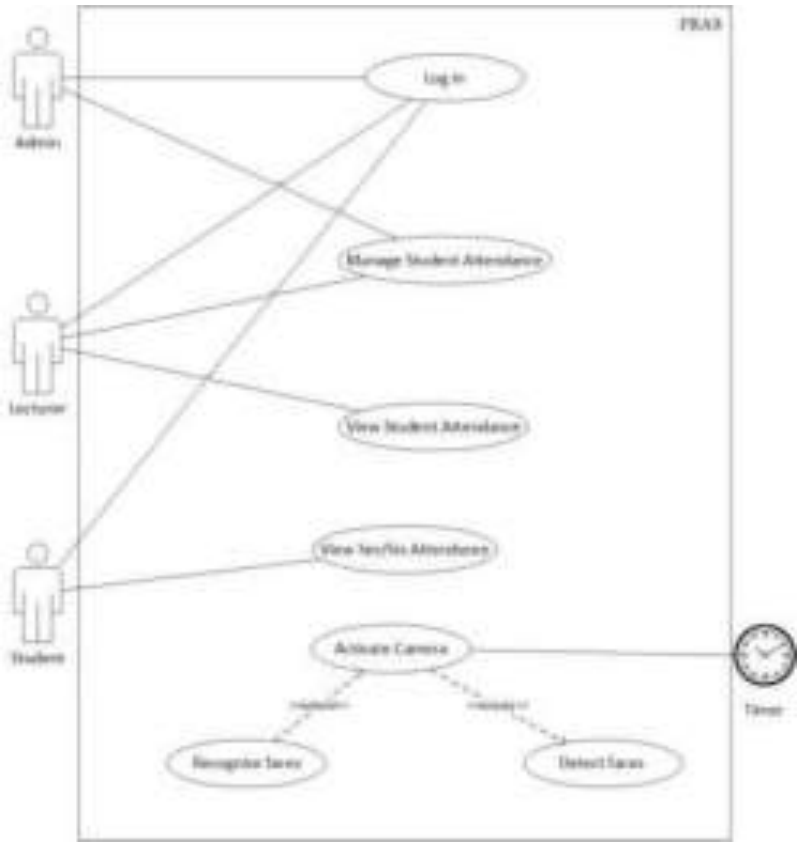
A Use Case Diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those usecases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted. Interaction among actors is not shown on the use case diagram. If this interaction is essential to acoherent description of the desired behavior, perhaps the system or use case boundaries should be re-examined. Alternatively, interaction among actors can be part of the assumptions used in the usecase.

Use cases:

A use case describes a sequence of actions that provide something of measurable value to an actor and is drawn as a horizontal ellipse.

Actors

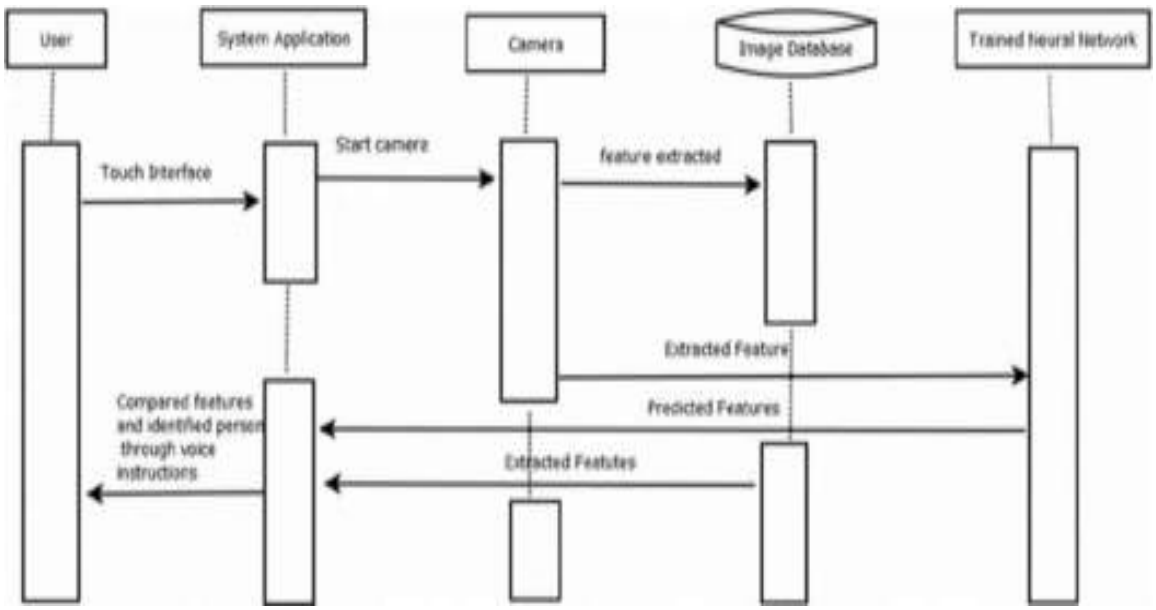
An actor is a person, organization, or external system that plays a role in one or more interactions with the system



4.2.3 SEQUENCE DIAGRAM

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart.

Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams. A sequence diagram shows, as parallel vertical lines (lifelines), different processes or objects that live simultaneously, and, as horizontal arrows, the messages exchanged between them, in the order in which they occur. This allows the specification of simple runtime scenarios in a graphical manner. If the lifeline is that of an object, it demonstrates a role. Note that leaving the instance name blank can represent anonymous and unnamed instances. In order to display interaction, messages are used. These are horizontal arrows with the message name written above them. Solid arrows with full heads are synchronous calls, solid arrows with stick heads are asynchronous calls and dashed arrows with stick heads are return messages.

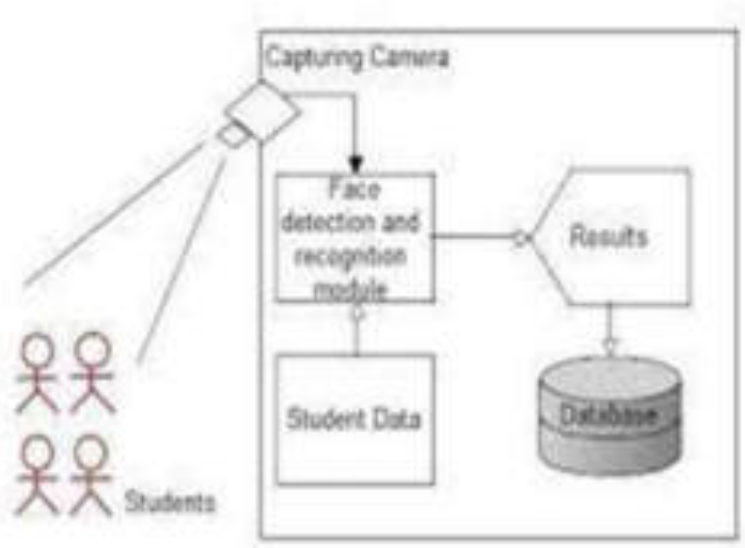


4.2.4 COLLABORATION DIAGRAM

A Sequence diagram is dynamic, and, more importantly, is time ordered. A Collaboration diagram is very similar to a Sequence diagram in the purpose it achieves; in 15 other words, it shows the dynamic interaction of the objects in a system.

A distinguishing feature of a Collaboration diagram is that it shows the objects and their association with other objects in the system apart from how they interact with each other. The association between objects is not represented in a Sequence diagram.

A Collaboration diagram is easily represented by modeling objects in a system and representing the associations between the objects as links. The interaction between the objects is denoted by arrows. To identify the sequence of invocation of these objects, a number is placed next to each of thesearrows

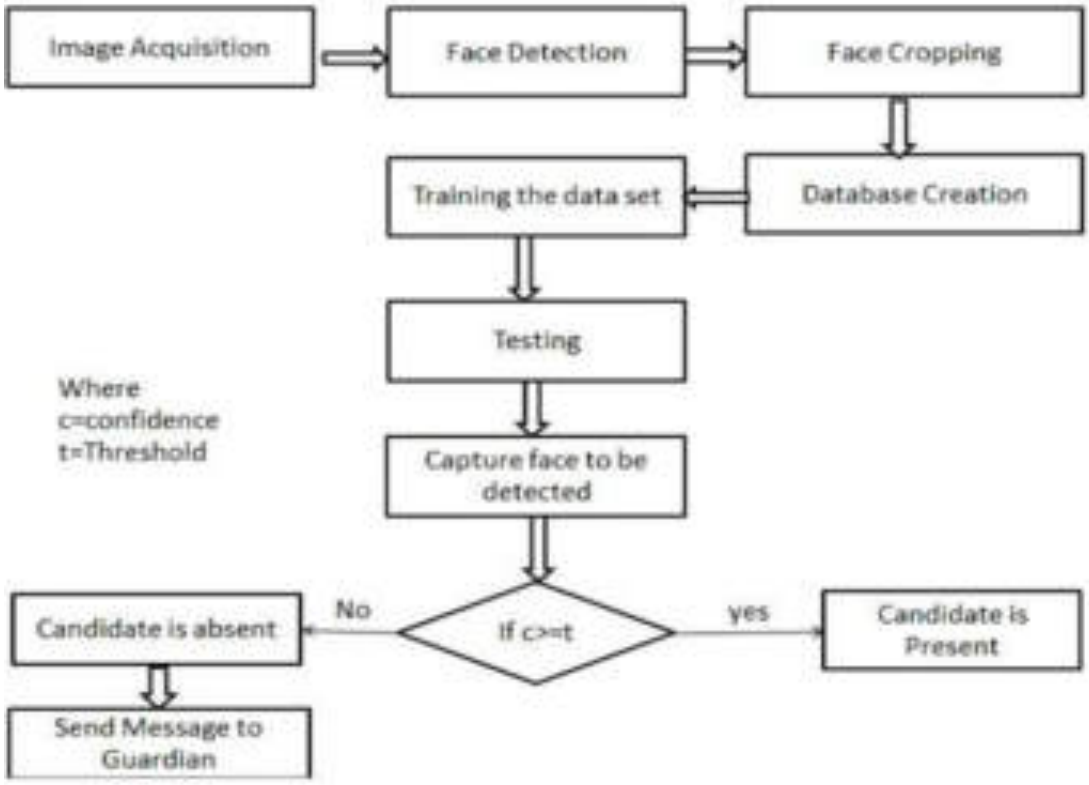


4.2.5 ACTIVITY DIAGRAM

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control. Activity diagrams are constructed from a limited repertoire of shapes, connected with arrows. The most important shape types:

- Rounded rectangles represent activities; diamonds represent decisions .
- Bars represent the start (split) or end (join) of concurrent activities .
- A black circle represents the start (initial state) of the workflow .
- An encircled black circle represents the end (finalstate).

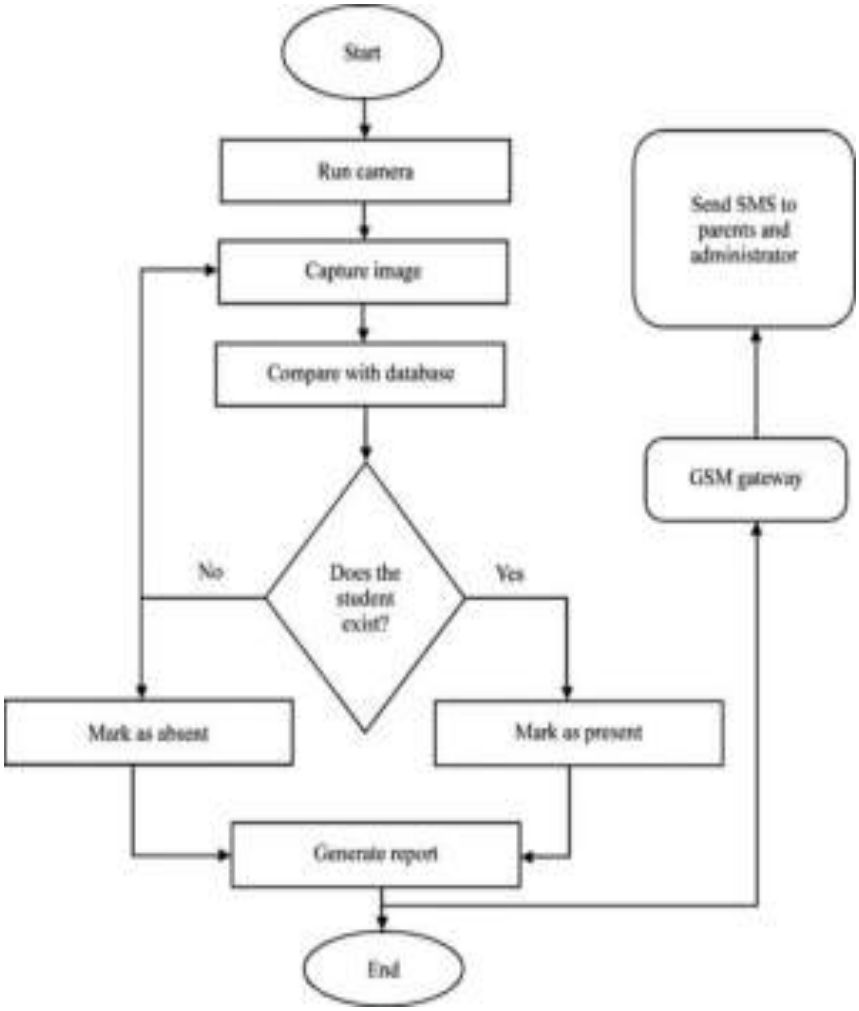
The **basic purposes of activity diagrams** is similar to other four **diagrams**. It captures the dynamic behaviour of the system. Other four **diagrams** are used to show the message flow from one object to another but **activity diagram** is used to show message flow from one **activity** to another.



4.2.6 STATE CHART DIAGRAM

A state diagram, also called a state machine diagram or state chart diagram, is an illustration of the states an object can attain as well as the transitions between those states in the Unified Modeling Language.

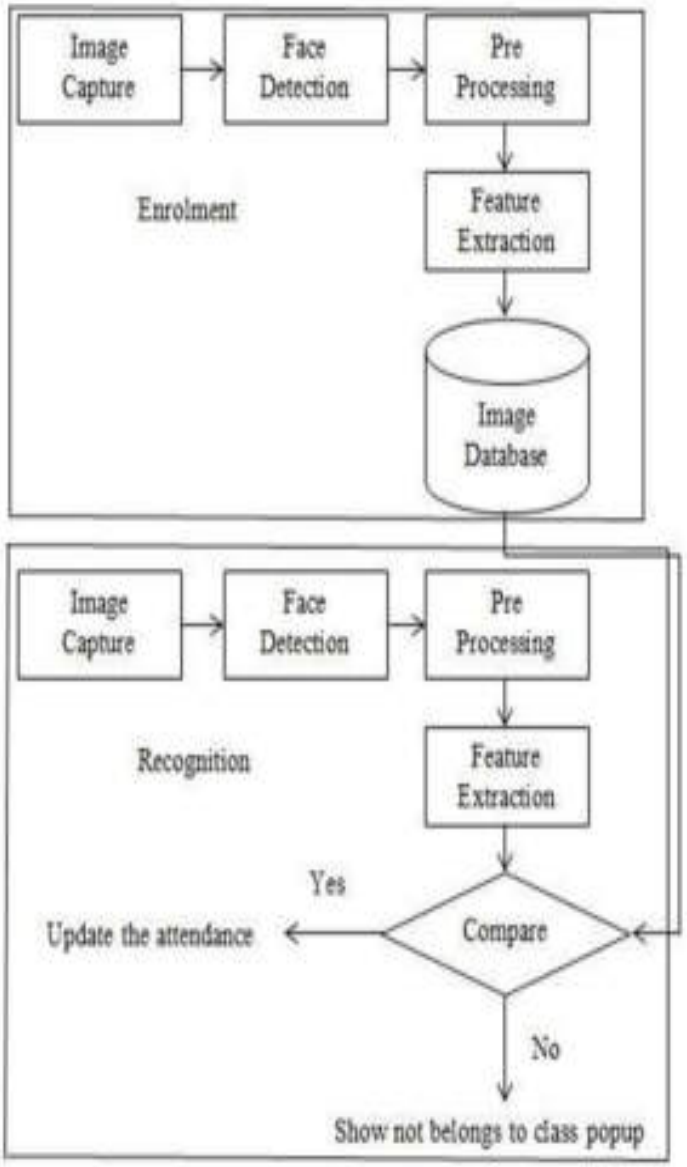
A state diagram resembles a flowchart in which the initial state is represented by a large black dot and subsequent states are portrayed as boxes with rounded corners.



4.2.7 COMPONENT DIAGRAM

A component contains a set of collaborating classes. Each class within a component has been fully elaborated to include all attributes and operations that are relevant to its implementation. As part of the design elaboration, all interfaces (messages) that enable the classes to communicate and collaborate with other design classes must also be defined.

To accomplish this, the designer begins with the analysis model and elaborates analysis classes (for components that relate to the problem domain) and infrastructure classes(or components that provide support services for the problem domain).



4.2.8 DEPLOYMENT DIAGRAM

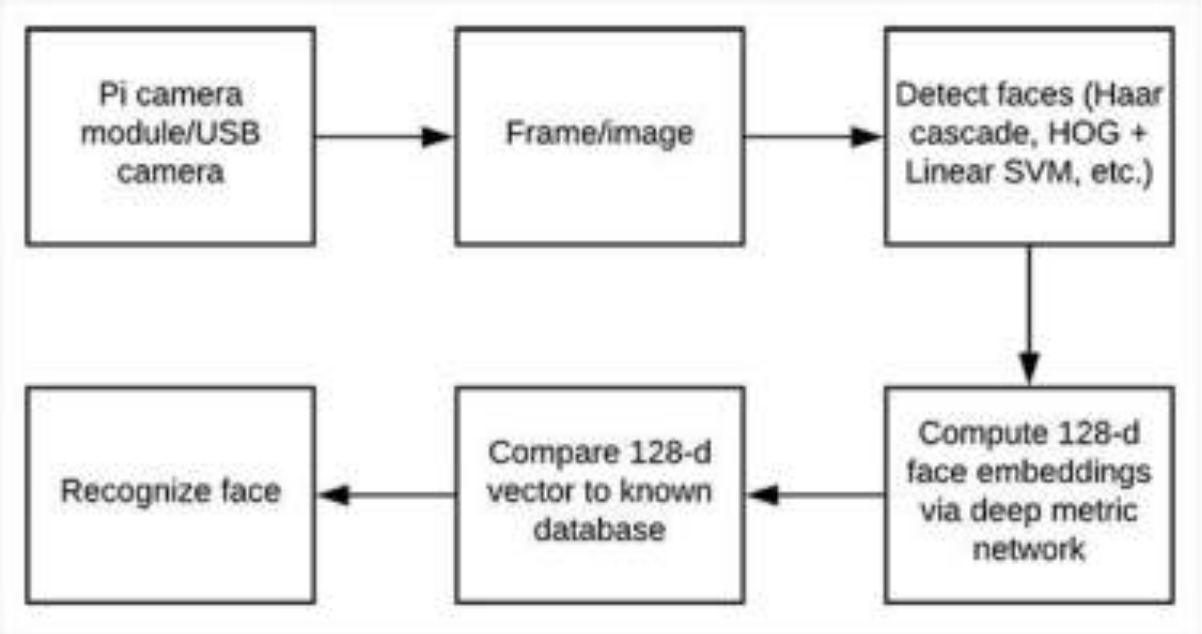
Deployment diagrams are used to visualize the topology of the physical components of a system where the software components aredeployed.So deployment diagrams are used to describe the static deployment view of a system.

The term Deployment itself describes the purpose of the diagram. Deployment diagrams are used for describing the hardware components, where software components are deployed. Component diagrams and deployment diagrams are closely related.

Component diagrams are used to describe the components and deployment diagrams shows how they are deployed in hardware.

UML is mainly designed to focus on the software artifacts of a system. However, these two diagrams are special diagrams used to focus on software and hardware components.

Most of the UML diagrams are used to handle logical components but deployment diagrams are made to focus on the hardware topology of a system. Deployment diagrams are used by the system engineers.



CHAPTER 5

IMPLEMENTATION

5.1 Step 1: Finding all the Faces

The first step in our pipeline is *face detection*. Obviously we need to locate the faces in a photograph before we can try to tell them apart!

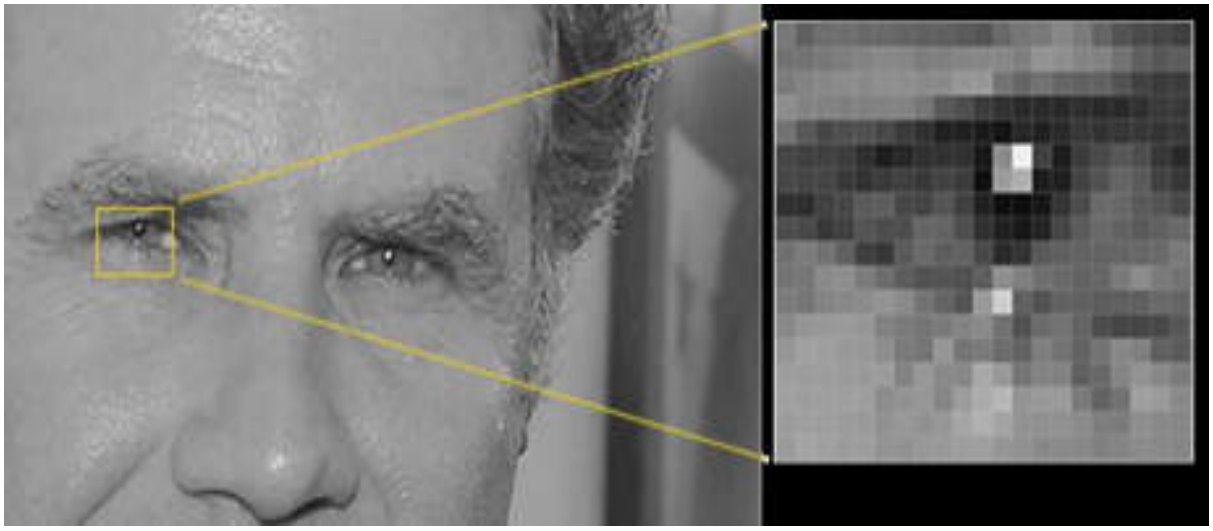
If you've used any camera in the last 10 years, you've probably seen face detection in action:

Face detection is a great feature for cameras. When the camera can automatically pick out faces, it can make sure that all the faces are in focus before it takes the picture. But we'll use it for a different purpose — finding the areas of the image we want to pass on to the next step in our pipeline.

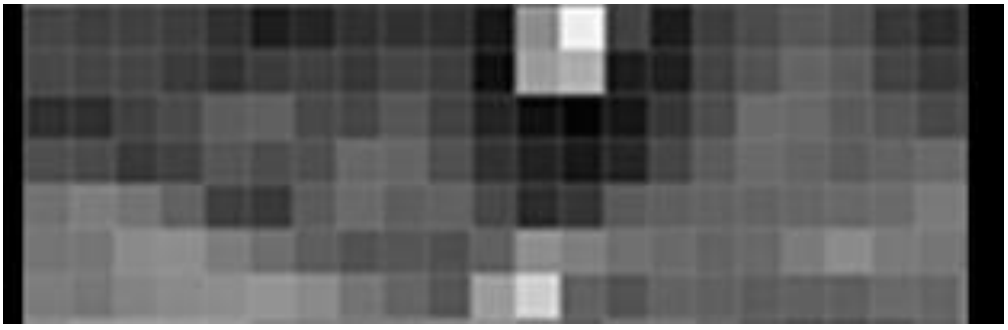
Face detection went mainstream in the early 2000's when Paul Viola and Michael Jones invented a way to detect faces that was fast enough to run on cheap cameras. However, much more reliable solutions exist now. We're going to us a method invented in 2005 called Histogram of Oriented Gradients — or just ***HOG*** for short.

To find faces in an image, we'll start by making our image black and white because we don't need color data to find faces:

Then we'll look at every single pixel in our image one at a time. For every single pixel, we want to look at the pixels that directly surrounding it:



Our goal is to figure out how dark the current pixel is compared to the pixels directly surrounding it. Then we want to draw an arrow showing in which direction the image is getting darker



If you repeat that process for **every single pixel** in the image, you end up with every pixel being replaced by an arrow. These arrows are called *gradients* and they show the flow from light to dark across the entire image:

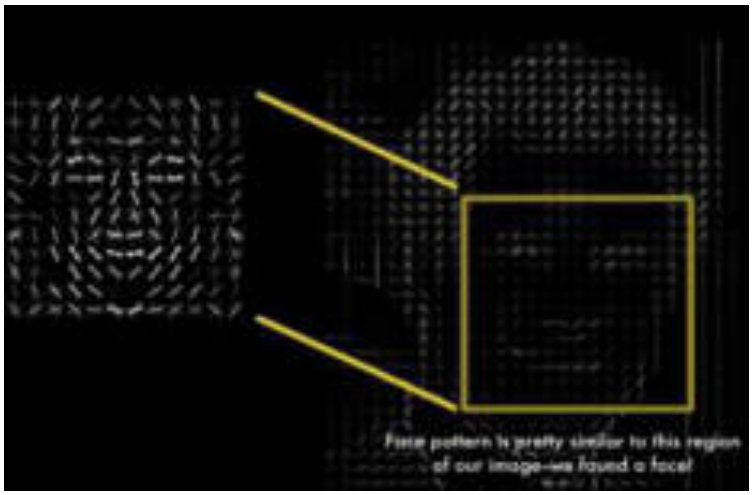
This might seem like a random thing to do, but there's a really good reason for replacing the pixels with gradients. If we analyze pixels directly, really dark images and really light images of the same person will have totally different pixel values. But by only considering the *direction* that brightness changes, both really dark images and really bright images will end up with the same exact representation. That makes the problem a lot easier to solve!

But saving the gradient for every single pixel gives us way too much detail. We end up missing gradient of trees from forest. It would be better if we could just see the basic flow of lightness/darkness at a higher level so we could see the basic pattern of the image.

To do this, we'll break up the image into small squares of 16x16 pixels each. In each square, we'll count up how many gradients point in each major direction (how many point up, point up-right, point right, etc…). Then we'll replace that square in the image with the arrow directions that were the strongest.

The end result is we turn the original image into a very simple representation that captures the basic structure of a face in a simple way:

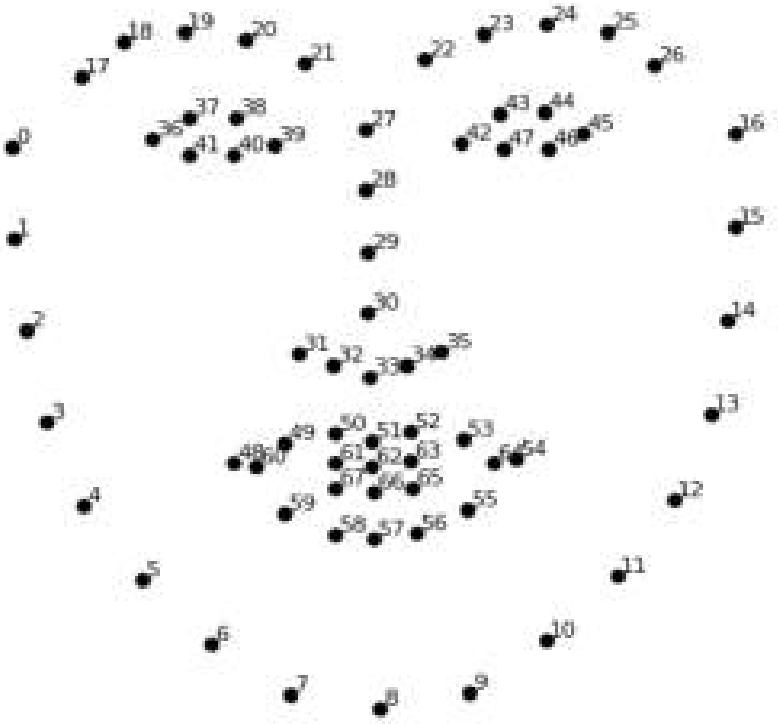
To find faces in this HOG image, all we have to do is find the part of our image that looks the most similar to a known HOG pattern that was extracted from a bunch of other training faces:



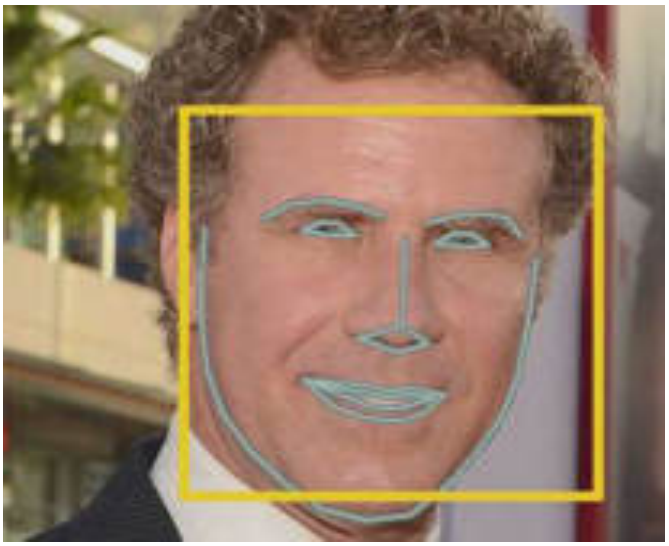
Using this technique, we can now easily find faces in any image:

5.1.2 Step 2: Posing and Projecting Faces

Whew, we isolated the faces in our image. But now we have to deal with the problem that faces turned different directions look totally different toa computer



Taking Face Co-ordinates as



Here if we see the coordinates which are represented at top are clearly represented in face

Now that we know were the eyes and mouth are, we'll simply rotate, scale and shear the image so that the eyes and mouth are centered as best as possible. We won't do any fancy 3d warps because that would introduce distortions into the image. We are only going to use basic image transformations like rotation and scale that preserve parallel lines.

5.1.3 Step 3: Encoding Faces

Now we are to the meat of the problem — actually telling faces apart. This is where things get really interesting!

The simplest approach to face recognition is to directly compare the unknown face we found in Step 2 with all the pictures we have of people that have already been tagged. When we find a previously tagged face that looks very similar to our unknown face, it must be the same person. Seems like a pretty good idea, right?

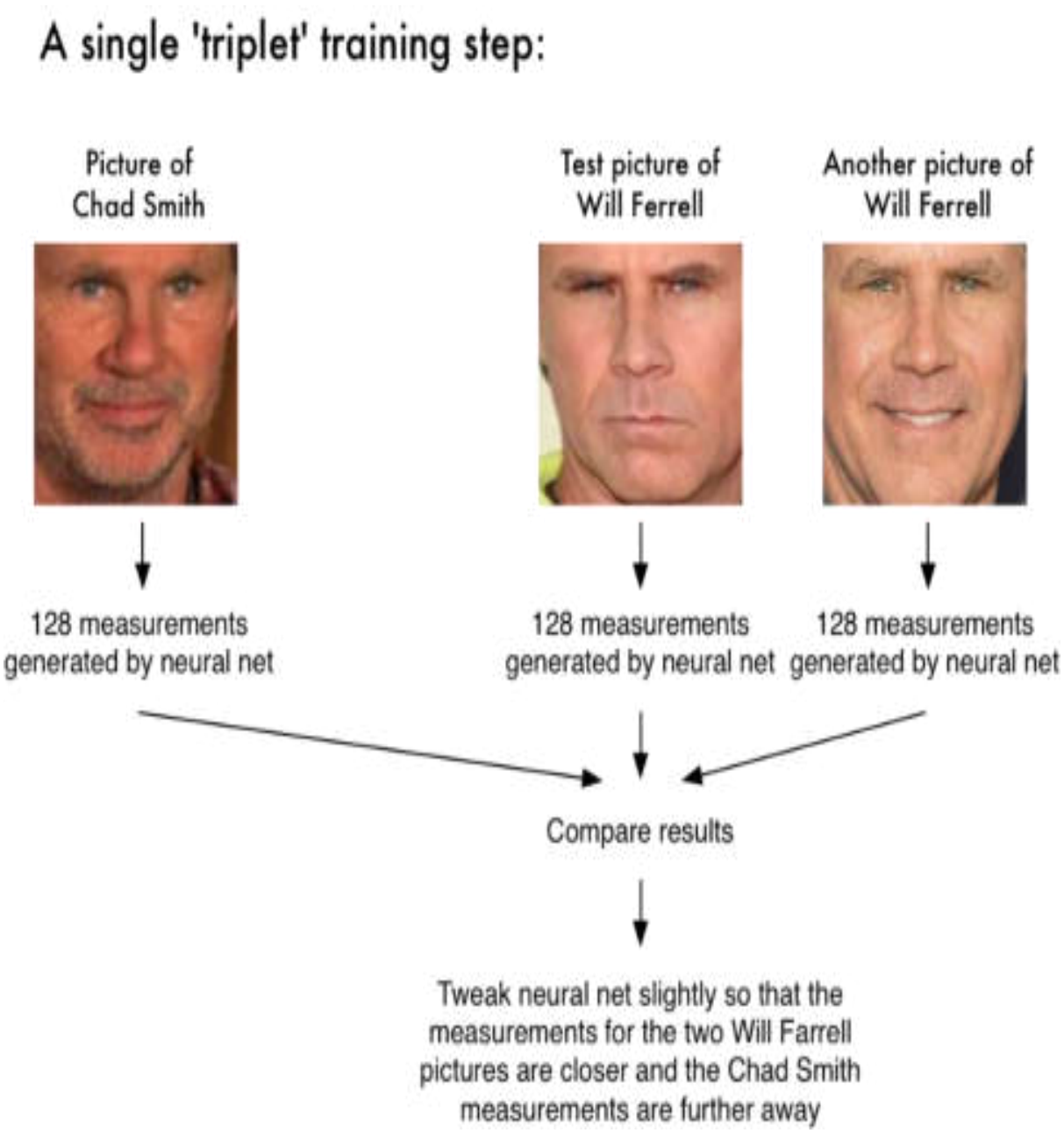
There's actually a huge problem with that approach. A site like Facebook with billions of users and a trillion photos can't possibly loop through every previous-tagged face to compare it to every newly uploaded picture. That would take way too long. They need to be able to recognize faces in milliseconds, not hours.

The training process works by looking at 3 face images at a time:

1 Load a training face image of a known person

2 Load another picture of the same known person

3 Load a picture of a totally different person



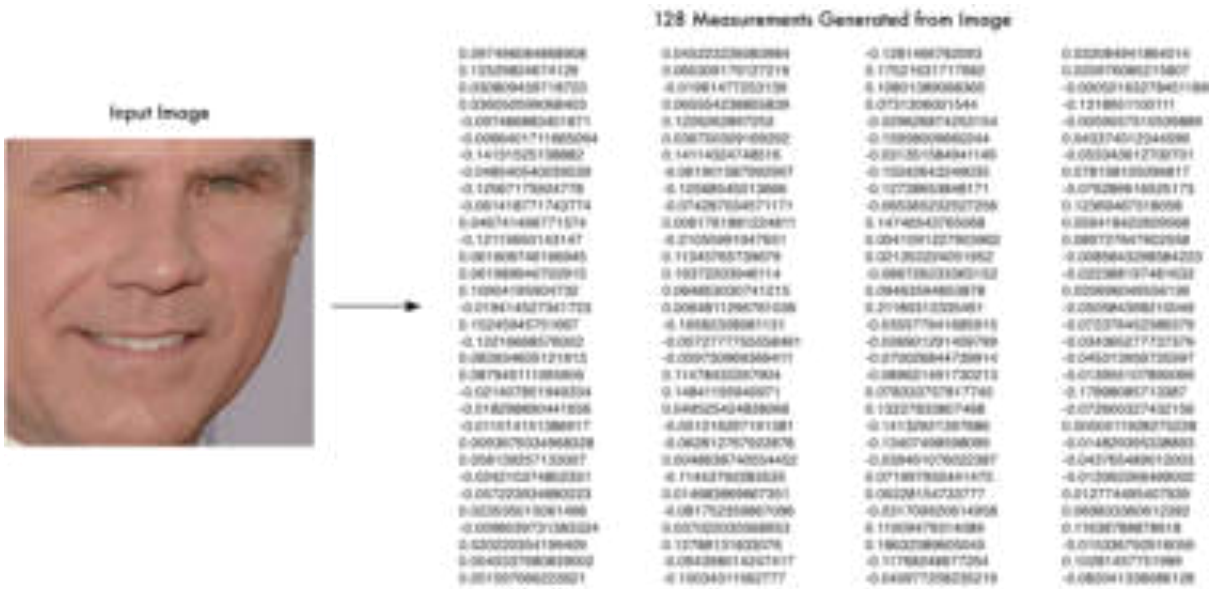
After repeating this step millions of times for millions of images of thousands of different people, the neural network learns to reliably generate 128 measurements for each person. Any ten different pictures of the same person should give roughly the same measurements.

Machine learning people call the 128 measurements of each face an **embedding**. The idea of reducing complicated raw data like a picture into a list of computer-generated numbers comes up a lot in machine learning (especially in language translation). The exact approach for faces we are using [was invented in 2015 by researchers at Google] but many similar approaches exist.

This process of training a convolutional neural network to output face embeddings requires a lot of data and computer power. Even with an expensive [NVidia Telsa video card], it takes [about 24 hours] of continuous training to get good accuracy.

But once the network has been trained, it can generate measurements for any face, even ones it has never seen before! So this step only needs to be done once.

So all we need to do ourselves is run our face images through their pre-trained network to get the 128 measurements for each face. Here's the measurements for our test image:



So what parts of the face are these 128 numbers measuring exactly? It turns out that we have no idea. It doesn't really matter to us. All that we care is that the network generates nearly the same

numbers when looking at two different pictures of the same person.

5.1.4 Step 4: Finding the person's name from the encoding

This last step is actually the easiest step in the whole process. All we have to do is find the person in our database of known people who has the closest measurements to our test image.

You can do that by using any basic machine learning classification algorithm. No fancy deep learning tricks are needed. We'll use a simple linear SVM classifier, but lots of classification algorithms could work.

All we need to do is train a classifier that can take in the measurements from a new test image and tells which known person is the closest match. Running this classifier takes milliseconds. The result of the classifier is the name of the person!

5.1.5 Generation Of Output

Generation of output will be done when once the above step process is completed, then if we keep a person in front of the standard web cam or a camera in which this module is inserted then it clearly recognizes the person and gives the output along with the name which we have saved at the time of saving the image in to the database.

5.1.6 Testing the model

For testing purpose , first we will check whether the model is working or not by making a person stand in front of web cam and then if it recognizes and returns the name along with the perfect time when the person is recognized then we will implement this process on multiple people where they are surrounded in a class and test there whether it all matches the correct person or not and giving the attendance to all persons or not ,this step is done and here comes our third testing process where the main doubt raises to all whether the model which we implemented recognizes the twin persons correctly or not and this process will be same for the twins too hence it returns the correct person face as output.

5.2 CODE DESIGN CHARACTERISTICS

5.2.1 BACK END

Python: To develop Machine Learning or Deep Learning projects or applications, the mostly used language across the world is Python. This is because of the libraries and frameworks that are provided by this language. For programming complex models, Python is the one language that makes it easier to understand and build models.

5.3 CODE SNIPPETS

```
import cv2

import numpy as np

import face_recognition

import os

from datetime import datetime


path='imagesAttendance'

images = []

classNames=[]
```

```
myList = os.listdir(path)
print(myList)
for cl in myList:
curImg = cv2.imread(f'{path}/{cl}')
images.append(curImg)
classNames.append(os.path.splitext(cl)[0])
print(classNames)

def findEncodings(images):
encodeList = []
    for img in images:
img = cv2.cvtColor(img, cv2.COLOR_BGR2RGB)
        encode = face_recognition.face_encodings(img)[0]
encodeList.append(encode)
    return encodeList

def markAttendance(name):
    with open('Attendance.csv','r+') as f:
myDataList = f.readline()
nameList=[]
        for line in myDataList:
            entry = line.split(',')
nameList.append(entry[0])
        if name not in nameList:
```

```
            now = datetime.now()
dtString = now.strftime('%H:%M:%S')
f.writelines(f'\n{name},{dtString}')
encodeListKnown = findEncodings(images)
print('Encoding Complete')
cap =cv2.VideoCapture(0)
while True:
    success, img = cap.read()
imgS = cv2.resize(img,(0,0),None,0.25,0.25)
imgS = cv2.cvtColor(imgS, cv2.COLOR_BGR2RGB)
facesCurFrame = face_recognition.face_locations(imgS)
encodesCurFrame = face_recognition.face_encodings(imgS,facesCurFrame)
    for encodeFace,faceLoc in zip(encodesCurFrame,facesCurFrame):
        matches = face_recognition.compare_faces(encodeListKnown,encodeFace)
faceDis = face_recognition.face_distance(encodeListKnown,encodeFace)
        #print(faceDis)
matchIndex = np.argmin(faceDis)
        if matches[matchIndex]:
            name = classNames[matchIndex].upper()
            #print(name)
            y1,x2,y2,x1 = faceLoc
            y1,x2,y2,x1 = y1*4,x2*4,y2*4,x1*4
            cv2.rectangle(img,(x1,y1),(x2,y2),(0,255,0),2)
            cv2.rectangle(img,(x1,y2-35),(x2,y2),(0,255,0),cv2.FILLED)
```

```
            cv2.putText(img,name,(x1+6,y2-6),cv2.FONT_HERSHEY_COMPLEX,1,(255,255,255),2)
markAttendance(name)
    cv2.imshow('Webcam',img)
    cv2.waitKey(1)
```

CHAPTER 6

TESTING

Testing is the process of detecting errors .Testing performs a very critical rolefor quality assurance and for ensuring the reliability of software .The results of test in gare used later on during maintenance also.

The aim of testing is often to demonstrate that a program works by showingthat it has no errors. The basic purpose of testing phase is to detect the errors that maybe present in the program.

Hence one should not start testing with the intent of showing that a program works, but the intent should be to show that a program doesn't work. Testing is the process of executing a program with the in ten to finding errors.

6.1 TYPES OF TESTS

6.1.1. Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

6.1.2 Integration Testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that

arise from the combination of components.

White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box. you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software work

6.1.3 User Acceptance Testing

User Acceptance of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changes wherever required. The system developed provides a friendly user interface that can easily be understood even by a person who is new to the system.

6.1.4 Output Testing

After performing the validation testing, the next step is output testing of the proposed system, since no system could be useful if it does not produce the required output in the specified format. Asking the users about the format required by them tests the outputs generated or displayed by the system under consideration. Hence the output format is considered in 2 ways – one is on screen and another in printed format.

6.2 TEST CASES

6.2.1 TEST_CASE_ID_01

DESCRIPTION	When we give the input as dataset it produces the accuracy result based on chosen algorithm
INPUT	Give the input data file having single image
EXPECTED OUTPUT	Produces the accuracy result for the dataset tested of FACE RECOGNITION.
ACTUAL OUTPUT	Produces the accuracy result for the dataset tested.
REMARKS	**PASS**

Table1:Test case describes about processing of single image as input

TEST_CASE_ID_02

DESCRIPTION	Whenwegivetheinputdatafileitproducestheaccuracyresultbasedonchosenalgorithm
INPUT	Givetheinputdatafilehavingmultiple images
EXPECTEDOUTPUT	Producestheaccuracyresult forthedatasettestedof FACE RECOGNITION.
ACTUALOUTPUT	Producestheaccuracyresult forthedataset tested.
REMARKS	**PASS**

TEST_CASE_ID_03

DESCRIPTION	When we give the input dataset it produces the accuracy result based on chosen algorithm
INPUT	Give the input dataset having Twins image
EXPECTEDOUTPUT	Produces the accuracy result for the dataset tested of FACE RECOGNITION.
ACTUALOUTPUT	Produces the accuracy result for the dataset tested.
REMARKS	**PASS**

CHAPTER 7

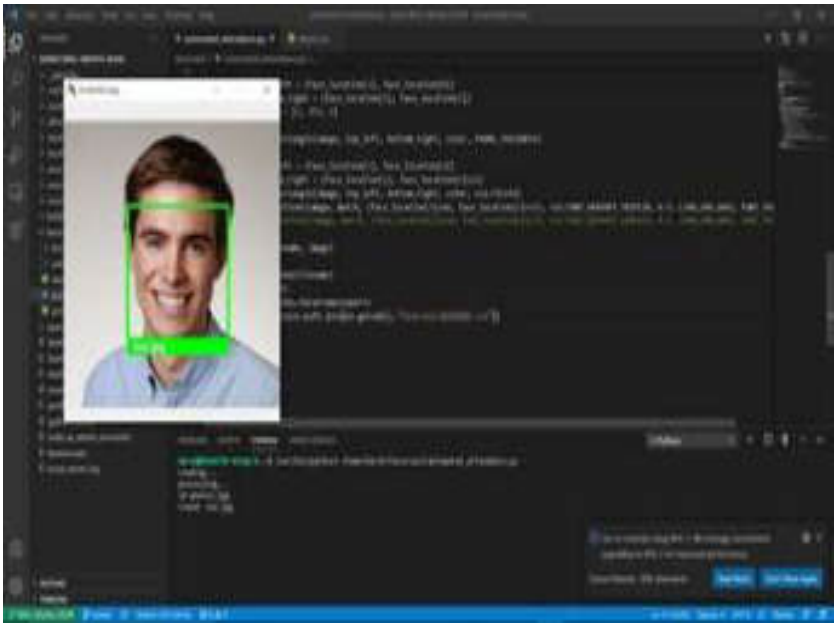
EXPERIMENTAL RESULTS

7.1 DATASETS COLLECTION

For Recognizing a individual person face through our project first we need to have the person image who stand in front of web cam in our data base, so when the person appears in front of the cam based on the encodings it will check with the images in the database and returns the output the name of the person or the name of the image saved at the time of storing in the database. So, we need a dataset to recognize the persons

7.2 Recognizing the Face image

For recognizing and experimenting purpose we took another image and the output screen of the image which we took and the process done for the face recognition will be at backend and hence the output will be displayed to us. When we have a single person it shows the image and compares it with the multiple images in database and returns the output.



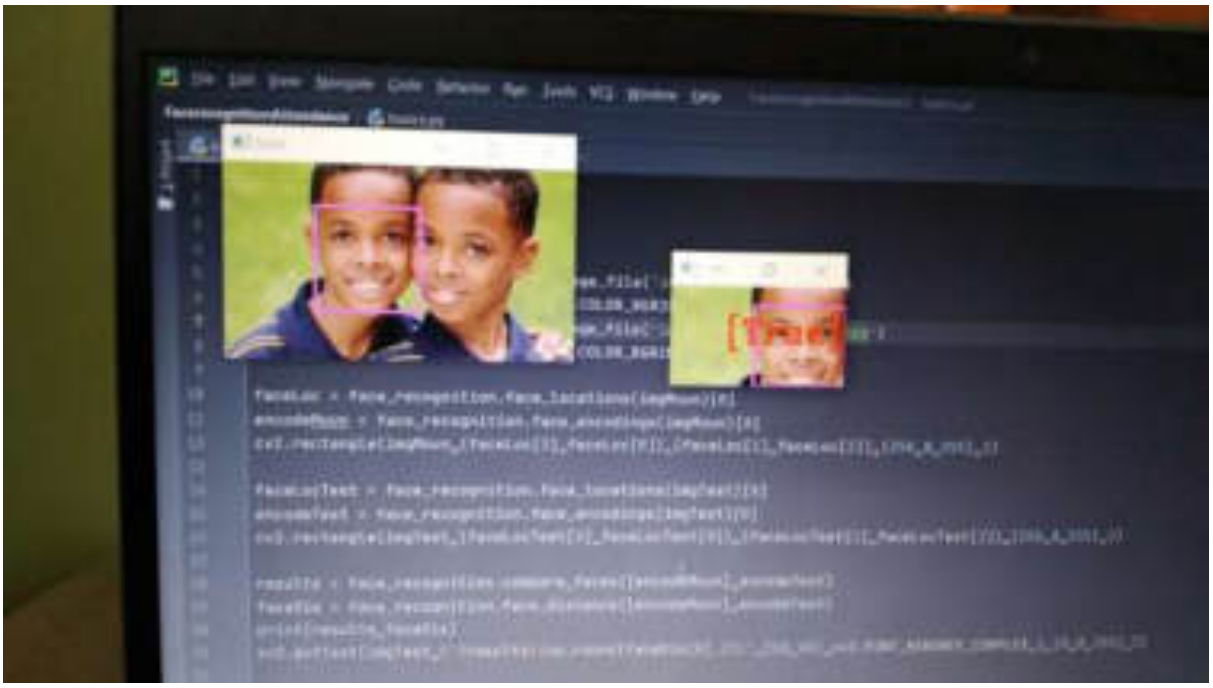
When we have single person in front of camera

If we observe here the total images in the below image are stored in database and based on the images which are stored the module was recognizing the persons according to it and gives the name of the person individually and gives the attendance them eventually when cam recognizes the person

When we have Twins how it detects

We all know that twins will have the same features and they look same even we cant describe

who is the person we looking for when we see the both persons at a time as a human we may mistaken but the module which knows the encodings and measurements or we can call them as coordinates of face will recognize the correct person and returns the output the person whom we needed by this the time will be reduced and no confusion will raise at all.



CHAPTER 8

APPLICATIONS

Face Identification:

Face recognition systems identify people by their face images. Face recognition systems establish the presence of an authorized person rather than just checking whether a valid identification (ID) or key is being used or whether the user knows the secret personal identification numbers (Pins) or passwords. The following are example. To eliminate duplicates in a nationwide voter registration system because there are cases where the same person was assigned more than one identification number. The face recognition system directly compares the face images of the voters and does not use ID numbers to differentiate one from the others. When the top two matched faces are highly similar to the query face image, manual review is required to make sure they are indeed different persons so as to eliminate duplicates.

Access Control:

In many of the access control applications, such as office access or computer logon, the size of the group of people that need to be recognized is relatively small. The face pictures are also caught under natural conditions, such asfrontal faces and indoor illumination. The face recognition system of this application can achieve high accuracy without much co-operation from user. The following are the example. Face recognition technology is used to monitor continuously who is in front of a computer terminal. It allows the user to leave the terminal without closing files and logging out. When the user leaves for a predetermined time, a screen saver covers up the work and disables the mouse & keyboard. When the user comes back and is recognized, the screen saver clears and the previous session appears as it was left. Any other user who tries to logon without authorization is denied.

Security:

Today more than ever, security is a primary concern at airports and for airline staff office and passengers. Airport protection systems that use face recognition technology have been implemented at many airports around the world. The following are the two examples. In October, 2001, Fresno Yosemite International (FYI) airport in California deployed Viisage's face recognition technology for airport security purposes. The system is designed to alert FYl's airport

public safety officers whenever an individual matching the appearance of a known terrorist suspect enters the airport's security checkpoint. Anyone recognized by the system would have further investigative processes by public safety officers. Computer security has also seen the application of face recognition technology. To prevent someone else from changing files or transacting with others when the authorized individual leaves the computer terminal for a short time, users are continuously authenticated, checking that the individual in front of the computer screen or at a user is the same authorized person who logged in.

Image database investigations: Searching image databases of licensed drivers, benefit recipients, missing children, immigrants and police bookings.

General identity verification: Electoral registration, banking, electronic commerce, identifying newborns, national IDs, passports, employee IDs.

Surveillance:

Like security applications in public places, surveillance by face recognition systems has a low user satisfaction level, if not lower. Free lighting conditions, face orientations and other divisors all make the deployment of face recognition systems for large scale surveillance a challenging task. The following are some example of facebased surveillance. To enhance town center surveillance in Newham Borough of London, this has 300 cameras linked to the closed circuit TV (CCTV) controller room. The city council claims that the technology has helped to achieve a 34% drop in crime since its facility. Similar systems are in place in Birmingham, England. In 1999 Visionics was awarded a contract from National Institute of Justice to develop smart CCTV technology.

CHAPTER 9

CONCLUSION

Before the development of this project. There are many loopholes in the process of taking attendance using the old method which caused many troubles to most of the institutions. Therefore,the facial recognition feature embedded in the attendance monitoring system can not only ensure attendance to be taken accurately and also eliminated the flaws in the previous system. By using technology to conquer the defects cannot merely save resources but also reduces human intervention in the whole process by handling all the complicated task to the machine. The only cost to this solution is to have sufficient space in to store all the faces into the database storage. Fortunately, there is such existence of micro SD that can compensate with the volume of the data. In this project, the face database is successfully built. Apart from that, the face recognizing system is also working well.At the end, the system not only resolve troubles that exist in the old model but also provide convenience to the user to access the information collected by mailing the attendance sheet to the respected faculty.

CHAPTER 10

FUTURE ENHANCEMENTS

The main intention of this project is to solve the issues encountered in the old attendance system while reproducing a brand new innovative smart system that can provide convenience to the institution. In this project, an application will be developed which is capable of recognising the identity of each individuals and eventually record down the data into a database system. Apart from that, an excel sheet is created which shows the students attendance and is directly mailed to the respected faculty.

The followings are the project scopes:

- The targeted groups of the attendance monitoring system are the students and staff of an educational institution.
- The database of the attendance management system can hold up to 2000 individual's information.
- The facial recognition process can only be done for 1 person at a time.
- An excel sheet is created which contains the student attendance and is mailed to the respected faculty.
- The project has to work under a Wi-Fi coverage area or under Ethernet connection, as the system need to update the database of the attendance system constantly.
- The device on which the application is running is powered up by power bank to improve the portability of application.

REFERENCES

[1] Diepen, G. (2017). Detecting and tracking a face with Python and OpenCV. [online] www.GuidoDiepen.nl. Available at: https://www.guidodiepen.nl/2017/02/detecting-andtracking-a-face-with-python-and-opencv/ [Accessed 19 Oct. 2017].

[2] Docs.opencv.org. (2017). Face Recognition with OpenCV — OpenCV 2.4.13.4 documentation. [online] Available at: https://docs.opencv.org/2.4/modules/contrib/doc/facerec/facerec_tutorial.html#facedatabase [Accessed 19 Oct. 2017]. Element14.com. (2017).

[3] Community: Raspberry Pi 3. [online] Available at: https://www.element14.com/community/community/raspberrypi/raspberrypi3/content?filterID=contentstatus[published]~objecttype~objecttype[document]&filterID=contentstatus[published]~language~language%5Bcpl%5D [Accessed 21 Oct. 2017].

[4] GitHub. (2017). iashris/Face-Detection-OpenCV. [online] Available at: https://github.com/iashris/Face-Detection-OpenCV/blob/master/FaceDet.py [Accessed 19 Oct. 2017].

[5] Opencv-python-tutroals.readthedocs.io. (2017).

Getting Started with Images — OpenCVPython Tutorials 1 documentation. [online] Available at: http://opencvpythontutroals.readthedocs.io/en/latest/py_tutorials/py_gui/py_image_display/py_image_displa y.html [Accessed 19 Oct. 2017].

[6] Projects.raspberrypi.org. (2017). Cite a Website - Cite This For Me. [online] Available at: https://projects.raspberrypi.org/en/projects/getting-started-with-guis [Accessed 18 Oct. 2017].

[7] Raja, R. (2017). Face Detection using OpenCV and Python.. [online] SuperDataScience - Big Data | Analytics Careers | Mentors | Success. Available at: https://www.superdatascience.com/opencv-face-recognition/ [Accessed 19 Oct. 2017].

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"SIGN LANGUAGE DETECTION USING OPENCV AND DEEP LEARNING"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
COMPUTER SCIENCE AND ENGINEERING

Submitted by

M.AMURTHA(17H71A05B8)

I.J.V.D.SOWMYA(17H71A0576)

P.CHAITANYA KUMAR(17H71A0565)

S.SAI TARUN(17H71A05A5)

Under the Esteemed Guidance of
MS. PARVEEN PATHAN

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

2020-21

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC A
With Grade

2020-21

CERTIFICATE

This is to certify that the Main Project entitled **"SIGN LANGUAGE RECOGNITION USING OPENCV AND DEEP LEARNING"** is a bonafide work carried out by **M.AMURTHA(17H71A05B8),I.J.V.D.SOWMYA(17H71A0576),P.CHAITANYAKUMAR(17H71A0565),S.SAI TARUN(17H71A05A5)** in partial fulfilment for the award of degree of Bachelor of Technology in **Computer Science and Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

Parveen
Project Guide

D. Prasad
Head of the Department

Examiner

Dr K Srinivas
PRINCIPAL

ACKNOWLEDGEMENT

We would like to take this opportunity to express my deepest appreciation to the following people for their valuable contributions and assistance with this Project.

Initially, we would like to thank my project supervisor, **Ms. Parveen Pathan, Assistant Professor, Department of Computer Science and Engineering** for the guidance and support, especially for the valuable ideas and knowledge shared to me throughout the Project.

We have the immense pleasure in expressing my thanks and deep sense of gratitude to**, Mr.D. Prasad, Head of the Department, Computer Science and Engineering** for extending necessary facilities and support for the completion of the Project.

We whole sincerely acknowledge **Dr. K. Srinivas, Principal and Prof D. PANDURANGA RAO, CEO** for giving opportunity to take up the Project work. We also extend my thanks to all faculty members of **Computer Science and Engineering**, for their valuable guidance and encouragement.

We would like to extend my warm appreciation to all my friends for sharing us their knowledge, valuable contributions and help with this Project.

Finally, my special thanks go to my family for their continuous support and help throughout my academic years and for their continual support and encouragement for the completion of the project.

DECLARATION

We **M. Amrutha, I. J. V. D. Sowmya, P. Chaitanya Kumar, S. Sai Tarun** of the Main-Project "**Sign language detection using OpenCV and Deep Learning**", hereby declare that the matter embodied in this Project is the genuine work done by us and has not been submitted either to this university or to any other university/institute for the fulfilment of the requirement of any course of study.

M.AMURTHA(17H71A05B6)

I.J.V.D. SOWMYA (17H71A0576)

P. CHAITANYA KUMAR (17H71A0565)

S.SAI TARUN (17H71A05A5)

CONTENTS

ABSTRACT

Sign language is used by deaf and hard hearing people to exchange information between their own community and with other people. Computer recognition of sign language deals from sign gesture acquisition and continues till text/speech generation. Sign gestures can be classified as static and dynamic. However static gesture recognition is simpler than dynamic gesture recognition but both recognition systems are important to the human community.

The sign language recognition steps are described in this survey. The data acquisition, data preprocessing and transformation, feature extraction, classification and results obtained are examined. Some future directions for research in this area also suggested. To make the better decisions, the use of Machine Learning techniques has increasing rapidly.

One of the most challenging tasks in the education sector is to predict the student academic performance. Every institution has their own criteria for analyzing the performance of students. There are lack of investigating the suitable factors which affect the academic performance. The objective of this project is to predict the student performance using Machine Learning algorithm i.e., KNN & Decision tree by comparing the accuracy of both algorithms. We state the best accurate algorithm for student performance prediction.

LIST OF TABLES

LIST OF FIGURES

1.INTRODUCTION

Motion of any body part like face, hand is a form of gesture. Here for gesture recognition we are using image processing and computer vision. Gesture recognition enables computer to understand human actions and also acts as an interpreter between computer and human. This could provide potential to human to interact naturally with the computers without any physical contact of the mechanical devices. Gestures are performed by deaf and dumb community to perform sign language.

This community used sign language 3 for their communication when broadcasting audio is impossible, or typing and writing is difficult, but there is the vision possibility. At that time sign language is the only way for exchanging information between people. Normally sign language is used by everyone when they do not want to speak, but this is the only way of communication for deaf and dumb community. Sign language is also serving the same meaning as spoken language does.

This is used by deaf and dumb community all over the world but in their regional form like ISL, ASL. Sign language can be performed by using Hand gesture either by one hand or two hands. It is of two type Isolated sign language and continuous sign language. Isolated sign language consists of single gesture having single word while continuous ISL or Continuous Sign language is a sequence of gestures that generate a meaningful sentence. In this report we performed isolated ASL gesture recognition technique.

1.1 About The Project

The intent of this project is to provide a digital autonomous platform to the hearing impaired and the English speakers.

1.2 Sign Language

Deaf people around the world communicate using sign language as distinct from spoken language in their everyday a visual language that uses a system of manual, facial and body movements as the means of communication. Sign language is not an universal language, and different sign languages are used in different countries, like the many spoken languages all over the world. Hundreds of sign languages are in used around the world, for instance, Japanese Sign Language, British Sign Language (BSL), Spanish Sign Language, Turkish Sign Language.

Sign language is a visual language and consists of 3 major components:

- **Fingerspelling** - Used to Spell Words
- **Word Level Sign** - Used for The Majority Of Communication.
- **Non manual features** - Facial expressions, tongue/mouth position and body posture used to modify sign meaning.

Fingerspelling	Word Level Sign	Non manual features
Used To Spell Words	Used For The Majority Of Communication	Facial expressions, tongue/mouth position and

		body posture used to modify sign meaning

1.Types of Communication

We have used American Sign Language (ASL) as a base in our project.American Sign Language is a complex visual spatial language that is used by the deaf community in the United States and English-speaking parts of Canada. It is linguistically complete, natural language. It is native language of many deaf men and women, as well as some hearing children born into deaf families.

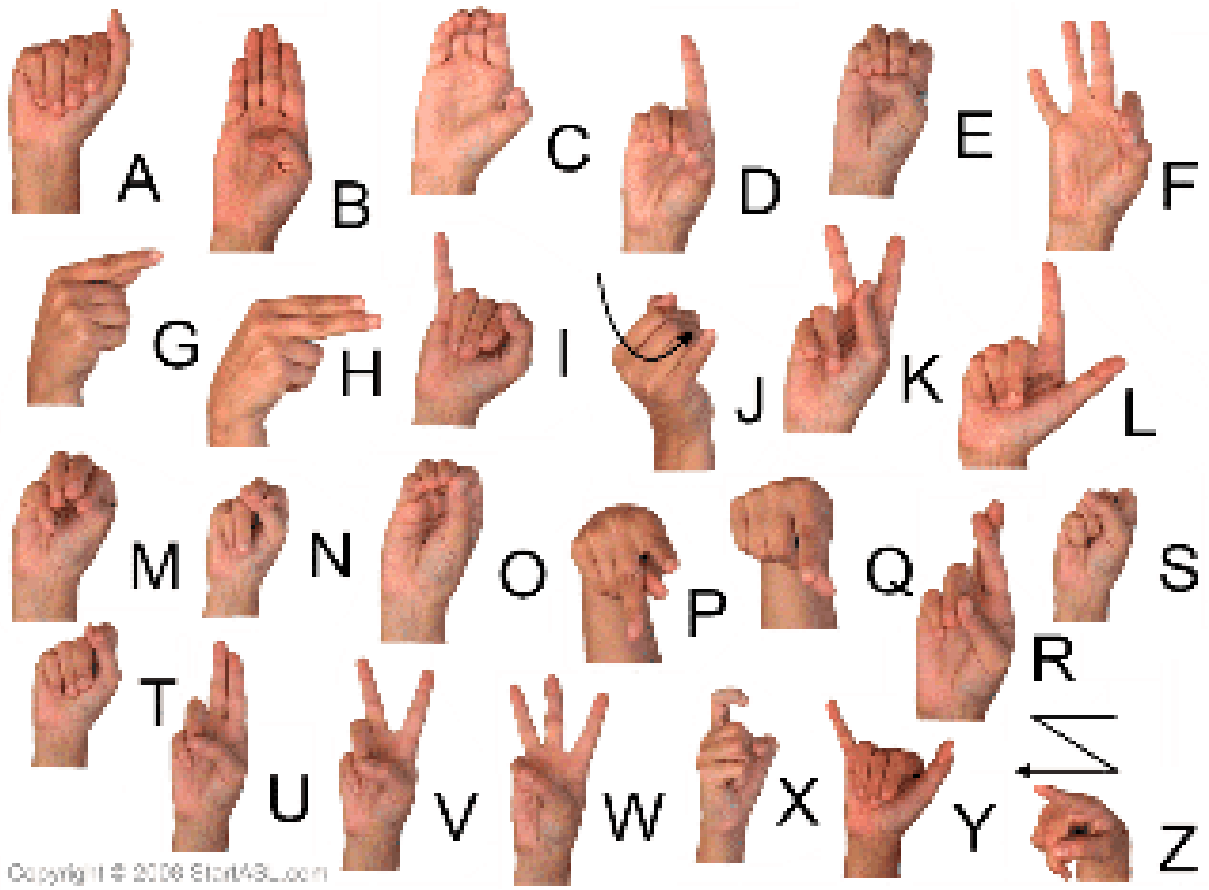


Fig 1: American Sign Language [18]11

1.3 Need For Sign Language Recognition

While sign language is very important to deaf-mute people, to communicate both with normal people and with themselves, is still getting little attention from the normal people. We as the normal people, tend to ignore the importance of sign language, unless there are loved ones who are deaf-mute. One of the solution to communicate with the deaf-mute people is by using the services of sign language interpreter. But the usage of sign language interpreter can be costly. Cheap solution is required so that the deaf-mute and normal people can communicate normally.

1.4 Challenges

One of the major obstacle that we faced was the availability of multiple sign languages present(e.g. ASL, ISL, BSL). Finding a suitable and large enough data set to train our model to obtain an acceptable result was also a hurdle we had to face.

Training the given data set was computationally challenging as well.

1.5 Tools Used

OpenCV:

OpenCV is an open-source C++ library for image processing and computer vision, originally developed by Intel and now supported by Willow Garage. Therefore it is not mandatory for your OpenCV applications to be open or free. It is a library of many inbuilt functions mainly aimed at real time image processing.

CNN (Inception):

Convolutional neural network (ConvNets or CNNs) is one of the main categories to do images recognition, images classifications. Objects detections, recognition faces etc., are some of the areas where CNNs are widely used. CNN image classifications takes input image, process it and classify it under certain categories

Convolutional Neural Network:

- In neural networks, Convolutional neural network (ConvNets or CNNs) is one of the main categories to do images recognition, images classifications. Objects detections, recognition faces etc., are some of the areas where CNNs are widely used.

- Technically, deep learning CNN models to train and test, each input image will pass it through a series of convolution layers with filters (Kernels), Pooling, fully connected layers (FC) and apply SoftMax function to classify an object with probabilistic values between 0 and 1.

- The below figure is a complete flow of CNN to process an input image and classifies the objects based on values.

- There are mainly six layers in any CNN model:
    - Convo layer (Convo + ReLU)
    - Pooling layer
    - Fully connected (FC) layer
    - Batch normalization (bn)
    - Dropout (do)
    - SoftMax/logistic layer

1.6 System configuration

1.6.1 Software requirements:

- **Operating System** : Windows
- **SDK** : OpenCV, CNN
- **Language** : Python

1.6.2 Hardware Requirements:

- **Web Cam** : Good quality
- **Ram** : Minimum 8GB or higher
- **Rom** : 16 GB
- **Other devices** : Monitor, Keyboard, Mouse

2.LITERATURE SURVEY

In the recent years, there has been tremendous research on the hand sign recognition. The technology of gesture recognition is divided into two categories.

A Survey of Hand Gesture Recognition Methods in Sign Language Recognition Sign Language Recognition (SLR) system, which is required to recognize sign languages, has been widely studied for years. The studies are based on various input sensors, gesture segmentation, extraction of features and classification methods.

Due to recent advancement in classification methods, many of the recent proposed works mainly contribute on the classification methods, such as hybrid method and Deep Learning. Based on our review, Hidden Markov Models (HMM) based approaches have been explored extensively in prior research, including its modifications.

2.1 EXISTING SYSTEMS

2.1.1 Vision based:

In vision-based methods computer camera is the input device for observing the information of hands or fingers. The Vision Based methods require only a camera, thus realizing a natural interaction between humans and computers without the use of any extra devices. These systems tend to complement biological vision by describing artificial vision systems that are implemented in software and/or hardware. This poses a challenging problem as these systems need to be background invariant, lighting insensitive, person and camera independent to achieve real time performance. Moreover, such systems must be optimized to meet the requirements, including accuracy and robustness.

The vision-based hand gesture recognition system is shown in fig.:

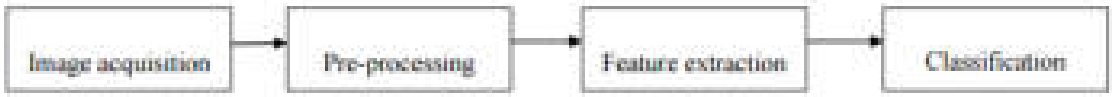


Fig 2: Block Diagram of vision-based recognition system

Vision based analysis, is based on the way human beings perceive information about their surroundings, yet it is probably the most difficult to implement in a satisfactory way.
Several different approaches have been tested so far.

1. One is to build a three-dimensional model of the human hand. The model is matched to images of the hand by one or more cameras, and parameters corresponding to palm orientation and joint angles are estimated. These parameters are then used to perform gesture classification

2. Second one to capture the image using a camera then extract some feature and those features are used as input in a classification algorithm for classification.

2.1.2 Glove Based:

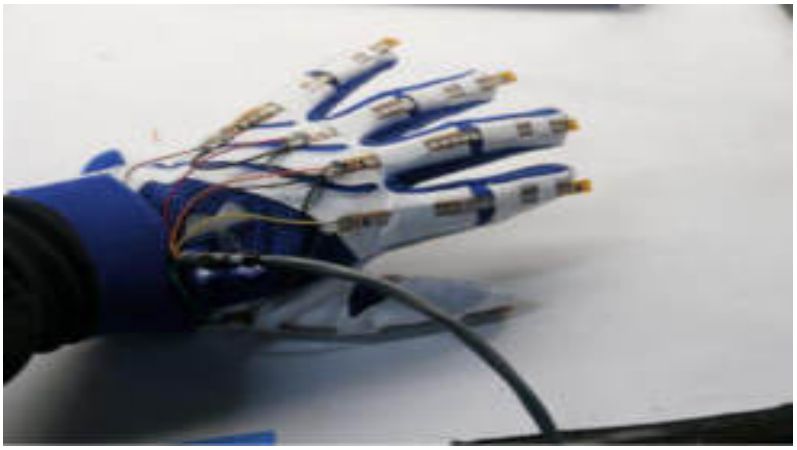


Fig. 3 Hand Gesture Recognition

In glove-based systems, data gloves are used which can achieve the accurate positions of hand gestures as its positions are directly measured. The Data Glove based methods use sensor devices for digitizing hand and finger motions into multiparametric data. The extra sensors make it easy to collect hand configuration and movement.

However, the devices are quite expensive and bring much cumbersome experience to the users some of the earlier gesture recognition systems attempted to identify gestures using glove-based devices that would measure the position and joint angles of the hand. However, these devices are very cumbersome and usually have many cables connected to a computer.

This has brought forth the motivation of using nonintrusive, vision-based approaches for recognizing gestures Also the sensors used for the detection of the sign language and the gesture recognition in the system that are available in the market are quite costly. In computer recognition of spoken language, speech data is captured using a microphone connected to an ADC. Similarly a data capturing device is also required in order to recognize sign language; in this case measuring the position and movement of the signer's hands.

2.1.3 Using Flex Sensors:

In this system an electromechanical robot is designed and controlled using hand gesture in real time. The system is designed on microcontroller platform using Keil and MPLAB tools. Hand gesture recognition is done on the principle of resistance change sensed through flex sensor. These sensors are integrated in a hand gloves from which input to the system is given. The designed system is divided into two sections as transmitter and receiver. The transmitter section will be in hand gloves from which the data is sensed and processed through PIC16F7487 and send serially to the receiver section.

RF technology is used to transmit the data at the receiver section at the frequency of 2.4 GHz. ARM 7 (LPC2148) processor is used to receive the data. Here from the received data, the character is predicted and matched with the closest character from which the character is identified and displayed on LCD. The various case studies is prepared for the designed system and tested in real time.

The proposed system can be used for the various applications such as in unmanned machines, industries, handicapped personnel etc.

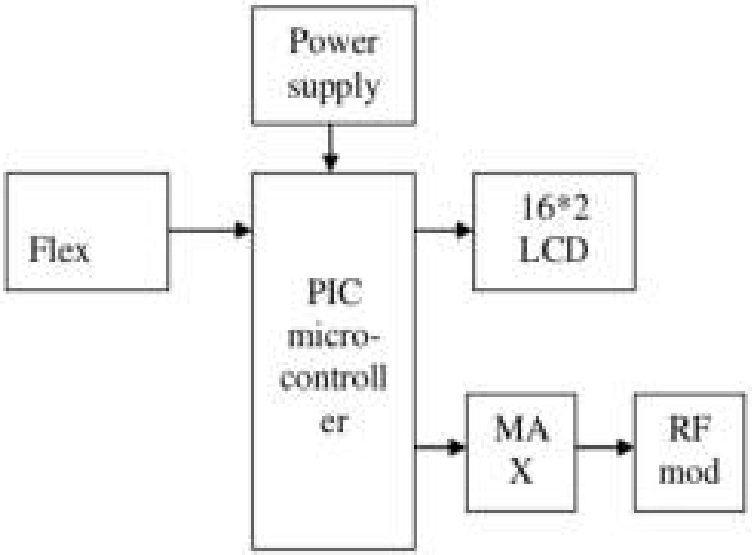


Fig 4: Transmitter section for hand gesture recognition

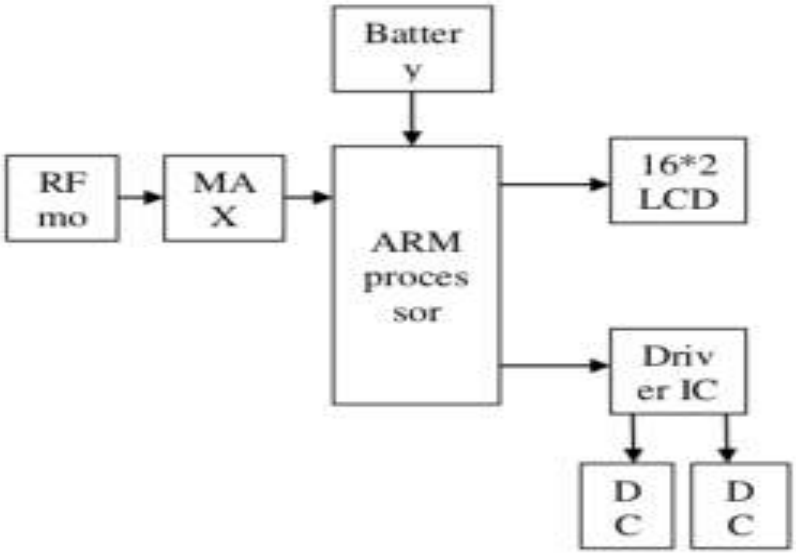


Fig 5: Receiver section for hand gesture recognition

2.2 Proposed system

In our model, we use a simple CNN to detect the sign of our hand. This CNN will recognize our hand sign and outputs the sign showed by us. We used this CNN to reduce the complexity that is faced using the existing models. CNN takes the hand sign as input and detects the sign so as to give the output sign as text/alphabet.

3. CONVOLUTIONAL NEURAL NETWORK(CNN)

Neural networks, as its name suggests, is a machine learning technique which is modeled after the brain structure. It comprises of a network of learning units called neurons. These neurons learn how to convert **input signals** (e.g., picture of a cat) into corresponding **output signals** (e.g., the label "cat"), forming the basis of automated recognition.

A convolutional neural network (CNN, or ConvNet) is a type of feedforward artificial neural network in which the connectivity pattern between its neurons is inspired by the organization of the animal visual cortex.

CNNs have repetitive blocks of neurons that are applied across space (for images) or time (for audio signals etc.). For images, these blocks of neurons can be interpreted as 2D convolutional kernels, repeatedly applied over each patch of the image. For speech, they can be seen as the 1D convolutional kernels applied across time windows. At training time, the weights for these repeated blocks are 'shared', i.e. the weight gradients learned over various image patches are averaged.

There are four main steps in CNN:

- convolution,
- subsampling,
- activation and
- full connectedness.

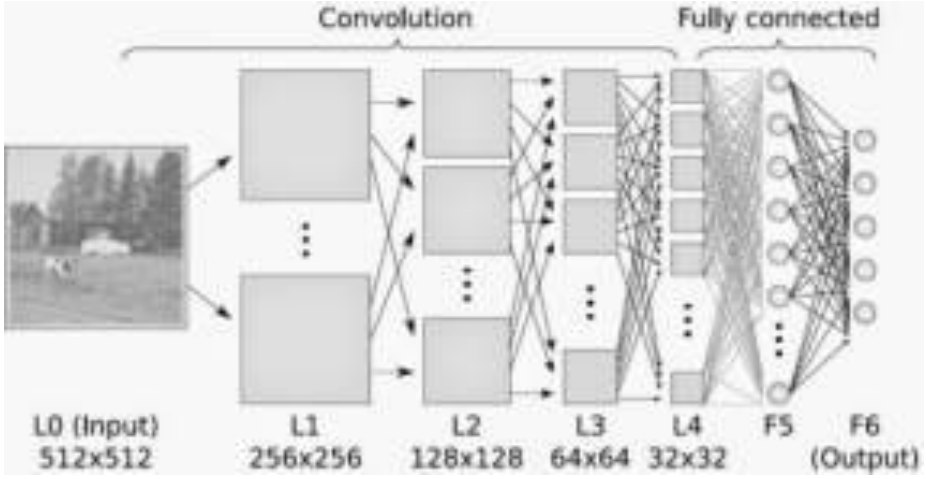


Fig 6: Convolutional neural network

3.1 Types of CNN

3.1.1 Convolution:

The first layers that receive an input signal are called convolution filters. Convolution is a process where the network tries to label the input signal by referring to what it has learned in the past. If the input signal looks like previous cat images it has seen before, the "cat" reference signal will be mixed into, or convolved with, the input signal. The resulting output signal is then passed on to the next layer.

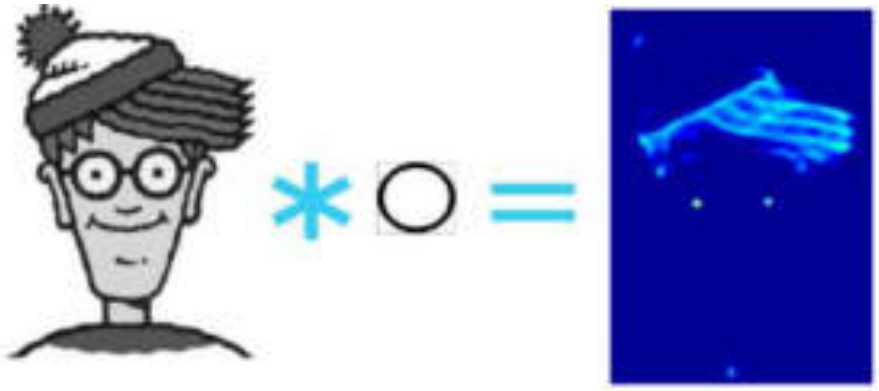


Fig 7: Convolving Wally with a circle filter. The circle filter responds strongly to the eyes.

Convolution has the nice property of being **translational invariant.** Intuitively, this means that each convolution filter represents a feature of interest (e.g., whiskers, fur), and the CNN algorithm learns which features comprise the resulting reference (i.e., cat). The output signal strength is not dependent on where the features are located, but simply whether the features are present. Hence, a cat could be sitting in different positions, and the CNN algorithm would still be able to recognize it.

3.1.2 Subsampling:

Inputs from the convolution layer can be "smoothened" to reduce the sensitivity of the filters to noise and variations. This smoothing process is called **subsampling,** and can be achieved by taking averages or taking the maximum over a sample of the signal. Examples of subsampling methods (for image signals) include reducing the size of the image, or reducing the color contrast across red, green, blue (RGB) channels.

Fig 8: Sub sampling Wally by 10 times. This creates a lower resolution image.

3.1.3 Activation:

The activation layer controls how the signal flows from one layer to the next, emulating how neurons are fired in our brain. Output signals which are strongly associated with past references would activate more neurons, enabling signals to be propagated more efficiently for identification. CNN is compatible with a wide variety of complex activation functions to model signal propagation, the most common function being the Rectified Linear Unit (ReLU), which is favored for its faster training speed.

3.1.4 Fully Connected:

The last layers in the network are fully connected, meaning that neurons of preceding layers are connected to every neuron in subsequent layers. This mimics high level reasoning where all possible pathways from the input to output are considered.

3.1.5 (During Training) Loss:

When training the neural network, there is additional layer called the loss layer. This layer provides feedback to the neural network on whether it identified inputs correctly, and if not, how far off its guesses were. This helps to guide the neural network to reinforce the right concepts as it trains. This is always the last layer during training.

3.2 Layers of CNN:

There are five different layers in CNN

- Input layer
- Convo layer (Convo + ReLU)
- Pooling layer
- Fully connected (FC) layer
- SoftMax/logistic layer
- Output layer

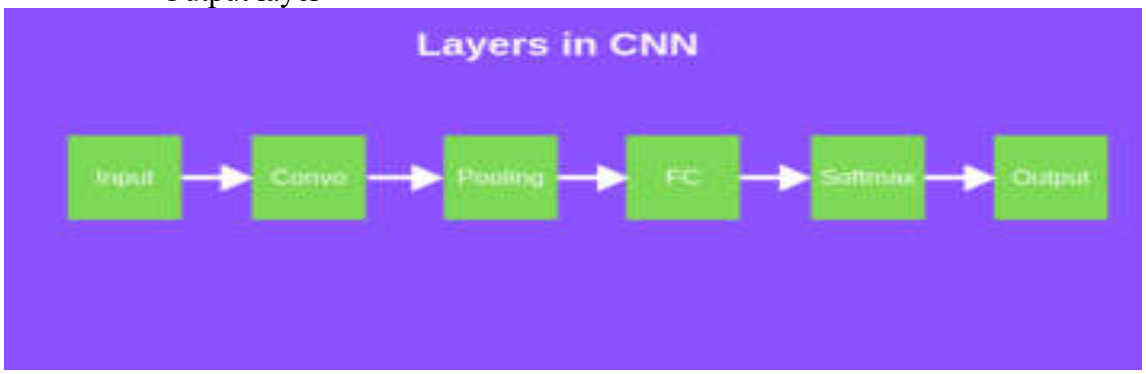


Fig 9. Block Diagram of Layers in CNN

3.2.1 Input Layer

Input layer in CNN should contain image data. Image data is represented by three dimensional matrix as we saw earlier. You need to reshape it into a single column. Suppose you have image of dimension 28 x 28 =784, you need to convert it into 784 x 1 before feeding into input. If you have "m" training examples then dimension of input will be (784, m).

3.2.2 Convo Layer

Convo layer is sometimes called feature extractor layer because features of the image are get extracted within this layer. First of all, a part of image is connected to Convo layer to perform convolution operation as we saw earlier and calculating the dot product between receptive field(it is a local region of the input image that has the same size as that of filter) and the filter. Result of the operation is single integer of the output volume. Then we slide the filter over the next receptive field of the same input image by a Stride and do the same operation again. We will repeat the same process again and again until we go through the whole image. The output will be the input for the next layer.

Convo layer also contains ReLU activation to make all negative value to zero.

3.2.3 Pooling Layer

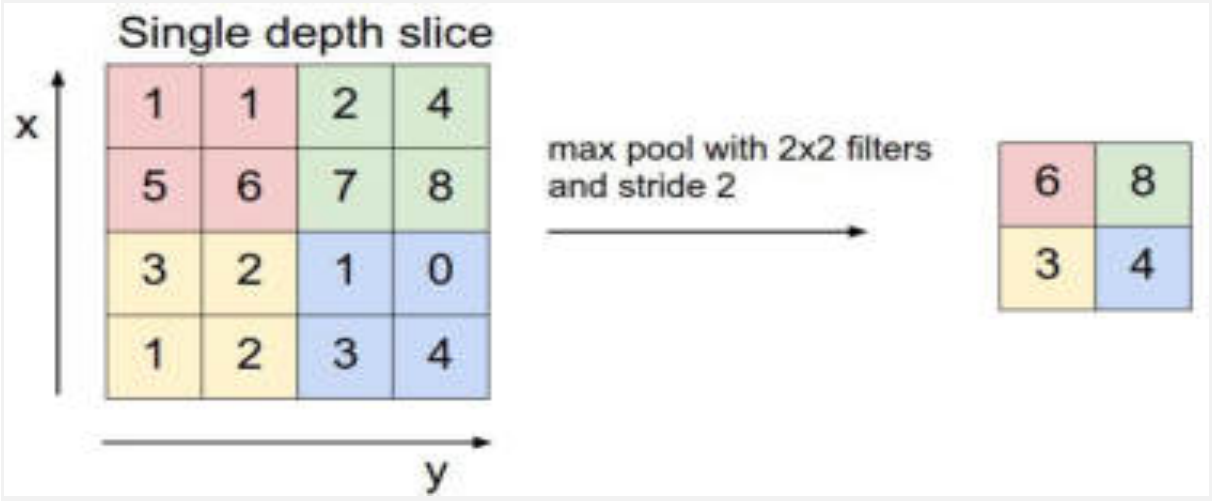


Fig 10. Pooling Layer Calculation

Pooling layer is used to reduce the spatial volume of input image after convolution. It is used

between two convolution layer. If we apply FC after Convo layer without applying pooling or max pooling, then it will be computationally expensive and we don't want it. So, the max pooling is only way to reduce the spatial volume of input image. In the above example, we have applied max pooling in single depth slice with Stride of 2. You can observe the 4 x 4 dimension input is reduce to 2 x 2 dimension.

There is no parameter in pooling layer but it has two hyperparameters — Filter(F) and Stride(S).

In general, if we have input dimension W1 x H1 x D1, then

$W2 = (W1-F)/S+1$

$H2 = (H1-F)/S+1$

$D2 = D1$

Where W2, H2 and D2 are the width, height and depth of output.

3.2.4 Fully Connected Layer (FC)

Fully connected layer involves weights, biases, and neurons. It connects neurons in one layer to neurons in another layer. It is used to classify images between different category by training.

3.2.5 SoftMax / Logistic Layer

Softmax or Logistic layer is the last layer of CNN. It resides at the end of FC layer. Logistic is used for binary classification and softmax is for multi-classification.

3.2.6 Output Layer

Output layer contains the label which is in the form of one-hot encoded.

3.3 APPLICATIONS OF CNN:

Decoding Facial Recognition:

Facial recognition is broken down by a convolutional neural network into the following major components –

- Identifying every face in the picture
- Focusing on each face despite external factors, such as light, angle, pose, etc.
- Identifying unique features19 Comparing all the collected data with already existing data in the database to match a face with a name. A similar process is followed for scene labeling as well.

Analyzing Documents:

Convolutional neural networks can also be used for document analysis. This is not just useful for handwriting analysis, but also has a major stake in recognizers. For a machine to be able to scan an individual's writing, and then compare that to the wide database it has, it must execute almost a million commands a minute.

It is said with the use of CNNs and newer models and algorithms, the error rate has been brought down to a minimum of 0.4% at a character level, though it's complete testing is yet to be widely seen.

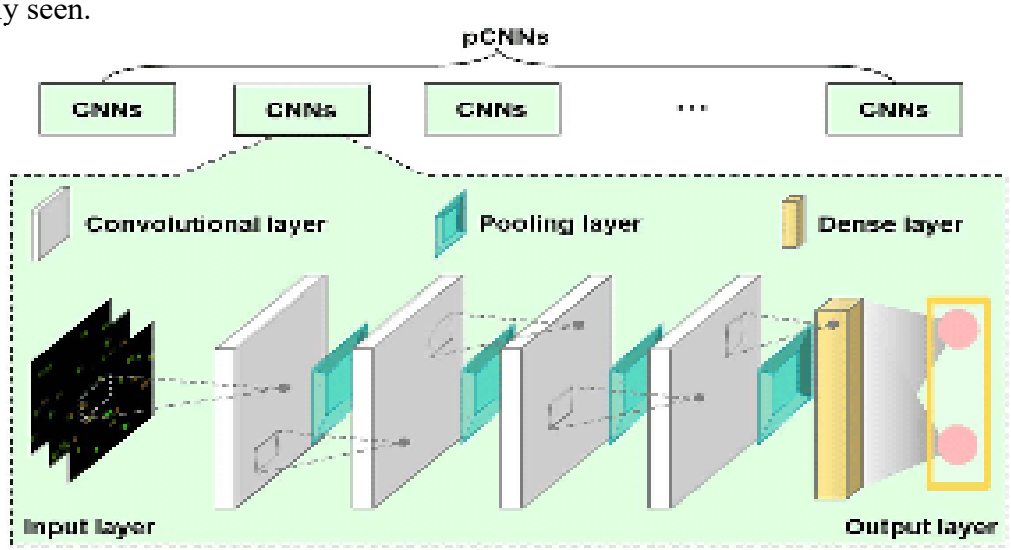


Fig 11: Layers involved in CNN

4.DESIGN

4.1 Dataflow Diagram

The DFD is also known as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of the input data to the system, various processing carried out on these data, and the output data is generated by the system. It maps out the flow of information for any process or system, how data is processed in terms of inputs and outputs.

It uses defined symbols like rectangles, circles and arrows to show data inputs, outputs, storage points and the routes between each destination. They can be used to analyse an existing system or model of a new one. A DFD can often visually "say" things that would be hard to explain in words and they work for both technical and non- technical.

There are four components in DFD:
1. External Entity
2. Process
3. Data Flow
4. data Store

1) External Entity:

It is an outside system that sends or receives data, communicating with the system. They are the sources and destinations of information entering and leaving the system. They might be an outside organization or person, a computer system or a business system. They are known as terminators, sources and sinks or actors. They are typically drawn on the edges of the diagram. These are sources and destinations of the system's input and output. Representation:

2) Process:

It is just like a function that changes the data, producing an output. It might perform computations for sort data based on logic or direct the dataflowbased on business rules.
Representation:

3) Data Flow:

A dataflow represents a package of information flowing between two objects in the data-flow diagram, Data flows are used to model the flow of information into the system, out of the system and between the elements within the system.
Representation:

4) Data Store:

These are the files or repositories that hold information for later use, such as a database table or a membership form. Each data store receives a simple label.
Representation:

4.2 UML DIAGRAMS

UML stands for Unified Modeling Language. Taking SRS document of analysis as input to the design phase drawn UML diagrams. The UML is only language so is just one part of the software development method. The UML is process independent, although optimally it should be used in a process that should be driven, architecture-centric, iterative, and incremental. The UML is language for visualizing, specifying, constructing, documenting the articles in a software- intensive system. It is based on diagrammatic representations of software components.

A modeling language is a language whose vocabulary and rules focus on the conceptual and physical representation of the system. A modeling language such as the UML is thus a standard language for software blueprints.

The UML is a graphical language, which consists of all interesting systems. There are also different structures that can transcend what can be represented in a programming language.

4.2.1 Use Case Diagram:

Use Case during requirement elicitation and analysis to represent the functionality of the system. Use case describes a function by the system that yields a visible result for an actor. The identification of actors and use cases result in the definitions of the boundary of the system i.e., differentiating the tasks accomplished by the system and the tasks accomplished by its environment.

The actors are outside the boundary of the system, whereas the use cases are inside the boundary of the system. Use case describes the behaviour of the system as seen from the actor's point of view. It describes the function provided by the system as a set of events that yield a visible result for the actor.

Fig 12. Use Case Diagram

Purpose of Use Case Diagrams:

The purpose of use case diagram is to capture the dynamic aspect of a system. However, this definition is too generic to describe the purpose, as other four diagrams41 (activity, sequence, collaboration, and Statechart) also have the same purpose. We will look into some specific purpose, which will distinguish it from other four diagrams.

Use case diagrams are used to gather the requirements of a system including internal and external influences. These requirements are mostly design requirements. Hence, when a system is analyzed to gather its functionalities, use cases are prepared and actors are identified.

When the initial task is complete, use case diagrams are modelled to present the outside view.
In brief, the purposes of use case diagrams can be said to be as follows −

- Used to gather the requirements of a system.
- Used to get an outside view of a system.
- Identify the external and internal factors influencing the system.
- Show the interaction among the requirements are actors.

4.2.2 Class Diagram:

Class diagrams model class structure and contents using design elements such as classes, packages and objects. Class diagram describe the different perspective when designing a system-conceptual, specification and implementation. Classes are composed of three things: name, attributes, and operations. Class diagram also display relationships such as containment, inheritance, association etc. The association relationship is most common relationship in a class diagram. The association shows the relationship between instances of classes.

How to Draw a Class Diagram?

- Class diagrams are the most popular UML diagrams used for construction of software applications.
- It is very important to learn the drawing procedure of class diagram.
- Class diagrams have a lot of properties to consider while drawing but here the diagram will be considered from a top level view.
- Class diagram is basically a graphical representation of the static view of the system and represents different aspects of the application.
- A collection of class diagrams represent the whole system.

The following points should be remembered while drawing a class diagram −

- The name of the class diagram should be meaningful to describe the aspect of the system.
- Each element and their relationships should be identified in advance.
- Responsibility (attributes and methods) of each class should be clearly identified For each class, minimum number of properties should be specified, as unnecessary properties will make the diagram complicated.
- Use notes whenever required to describe some aspect of the diagram.
- At the end of the drawing it should be understandable to the developer/coder.
- Finally, before making the final version, the diagram should be drawn on plain paper and reworked as many times as possible to make it correct.

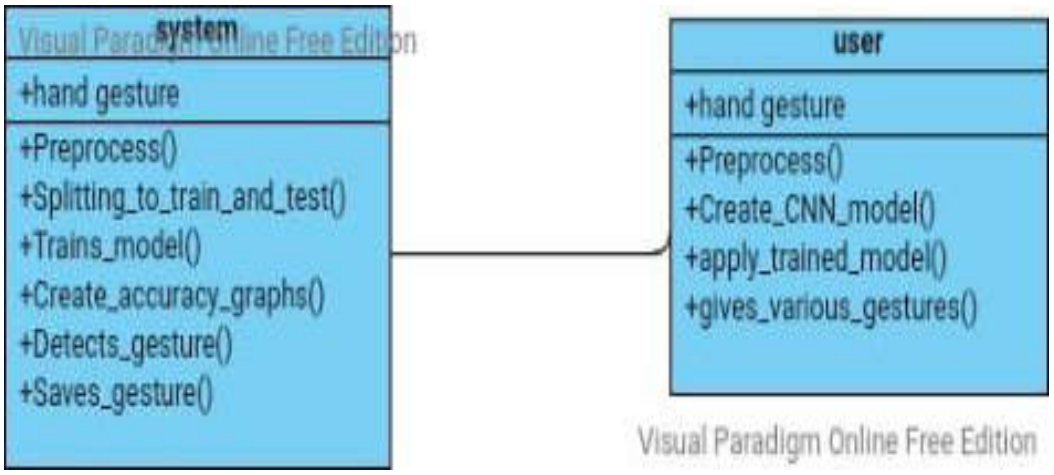


Fig 13. Class Diagram

4.2.3 State Chart:

A state chart diagram describes a state machine which shows the behaviour of classes. It shows the actual changes in state not processes or commands that create those changes and is the dynamic behavior of objects over time by model1ing the life cycle of objects of each class.

It describes how an object is changing from one state to another state.

There are mainly two states in State Chart Diagram:

1. Initial State
2. Final-State.

Some of the components of State Chart Diagram are:

- **State:** It is a condition or situation in life cycle of an object during which it's satisfies same condition or performs some activity or waits for some event.
- **Transition**: It is a relationship between two states indicating that object in first state performs some actions and enters into the next state or event.
- **Event**: An event is specification of significant occurrence that has a location in time and space.

Fig 14.State Chart Diagram

4.2.4 Deployment diagram:

Deployment diagrams are a set of nodes and their relationships. These nodes are physical entities where the components are deployed.

Deployment diagrams are used for visualizing the deployment view of a system. This is generally used by the deployment team.

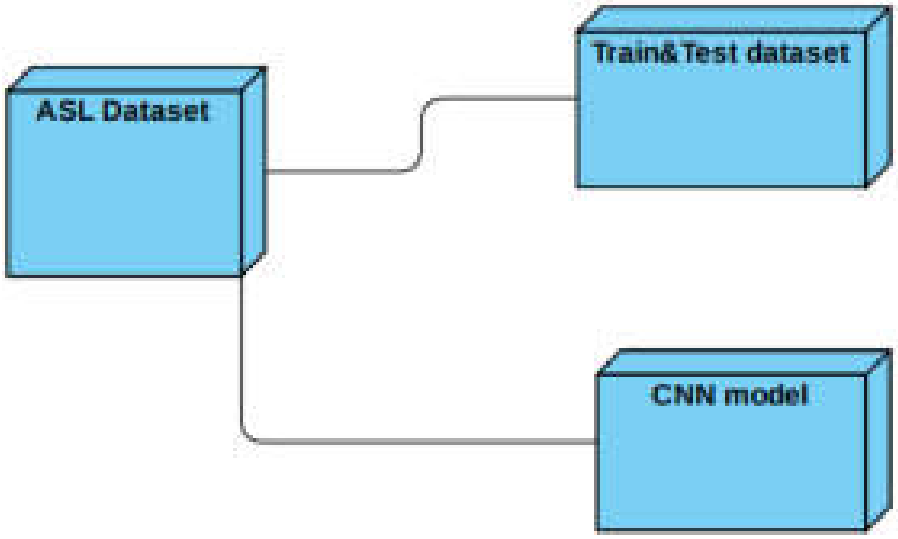


Fig 15. Deployment Diagram

4.2.5 Activity diagram:

Activity diagram describes the flow of control in a system. It consists of activities and links. The flow can be sequential, concurrent, or branched.

Activities are nothing but the functions of a system. Numbers of activity diagrams are prepared to capture the entire flow in a system.

Activity diagrams are used to visualize the flow of controls in a system. This is prepared to have an idea of how the system will work when executed.

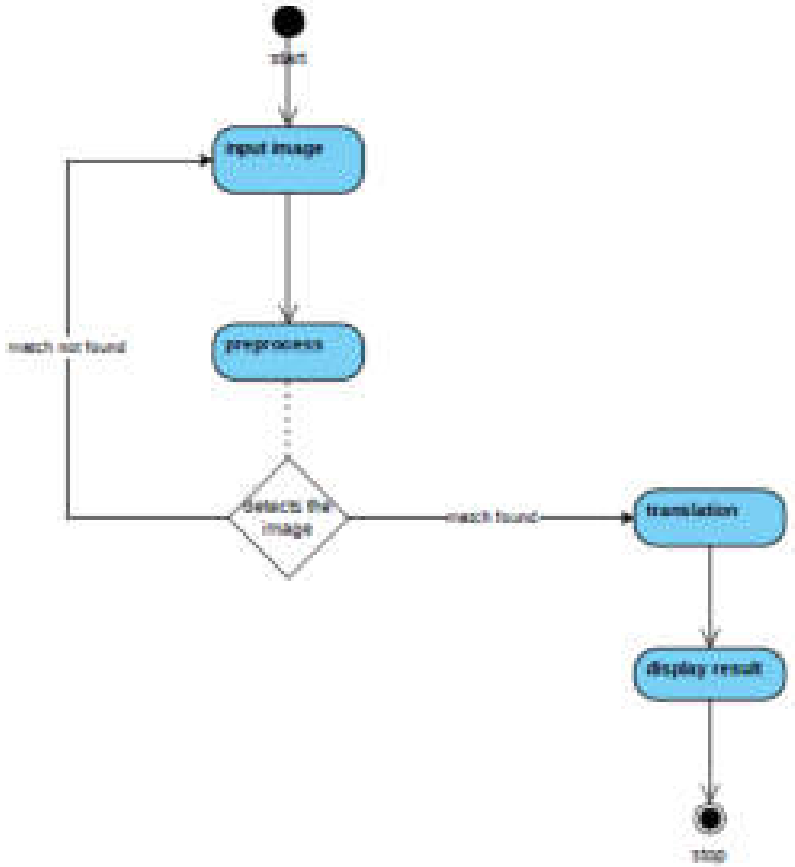


Fig 16. Activity Diagram

4.2.6 Component diagram:

Component diagrams represent a set of components and their relationships. These components consist of classes, interfaces, or collaborations. Component diagrams represent the implementation view of a system.

During the design phase, software artifacts (classes, interfaces, etc.) of a system are arranged in different groups depending upon their relationship. Now, these groups are known as components.

Finally, it can be said component diagrams are used to visualize the implementation.

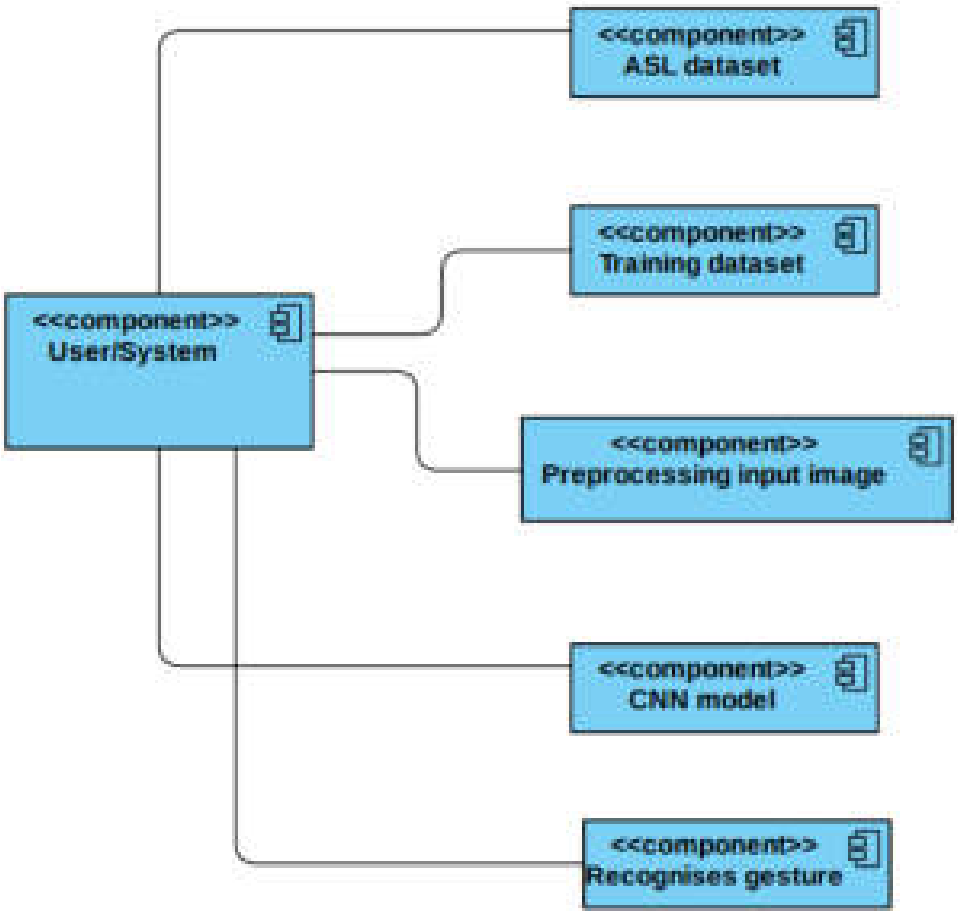


Fig 17. Component Diagram

4.2.7 Object diagram:

Object diagrams can be described as an instance of class diagram. Thus, these diagrams are more close to real-life scenarios where we implement a system.

Object diagrams are a set of objects and their relationship is just like class diagrams. They also represent the static view of the system.

The usage of object diagrams is similar to class diagrams but they are used to build prototype of a system from a practical perspective.

Fig 18. Object Diagram

5. Experiment analysis and results

5.1 Implementation:

In our model, we used simple CNN to detect the sign/gesture, which helps deaf people to communicate with others without any difficulty. We used

- **OpenCV** to read and preprocess the images,
- **argparse** for parsing command line arguments,
- **imutils** for getting proper image paths.

Steps:

- At first, we preprocess the images and save them to the disk.
- Later we create **data.csv** file
- Next, we start to create neural network.
- After creating neural network, we read the data and prepare the train and validation set.
- After completing the preparation, we train the model for specified number of epochs.
- Finally, we execute our model and analyse loss and accuracy vales.

5.2 Dataset description:

We will the **ASL Alphabet** dataset from Kaggle. This is a very large dataset containing 87000 images. All the images are 200×200 in dimension.

In this dataset, there are 29 classes in total. 26 of these classes are letters from A-Z. Then there are three more classes that correspond to *SPACE*, *DELETE*, and *NOTHING*. These three more classes will become really essential if you will want to take on the task to expand the project into something much larger. So, there are 3000 images from each class.

But we will not be using all the images to train our convolutional neural network. It will take much more time and resources. In fact, we will just use a subset of the images as a starting point

5.3 Source Code:

Preprocessing images:

```
import os
import cv2
import random
import numpy as np
import argparse

from tqdm import tqdm

parser = argparse.ArgumentParser()
parser.add_argument('-n','--num-
images',default=1000,type=int,help='number of images to preprocess for each cate
gory')
args = vars(parser.parse_args())

print("Preprocessing {} from each category...".format(args['num_images']))

dir_paths =os.listdir('../input/asl_alphabet_train/asl_alphabet_train/')
dir_paths.sort()

root_path = '../input/asl_alphabet_train/asl_alphabet_train'

for idx, dir_path in tqdm(enumerate(dir_paths),total=len(dir_paths)):
    all_images = os.listdir(f"{root_path}/{dir_path}")
    os.makedirs(f"../input/preprocessed_image/{dir_path}",exist_ok=True)
    for i in range(args['num_images']):
        rand_id = (random.randint(0,2999))
        image = cv2.imread(f"{root_path}/{dir_path}/{all_images[rand_id]}")
        image = cv2.resize(image, (224,224))

        cv2.imwrite(f"../input/preprocessed_image/{dir_path}/{dir_path}{i}.jpg",
image)
print('DONE')
```

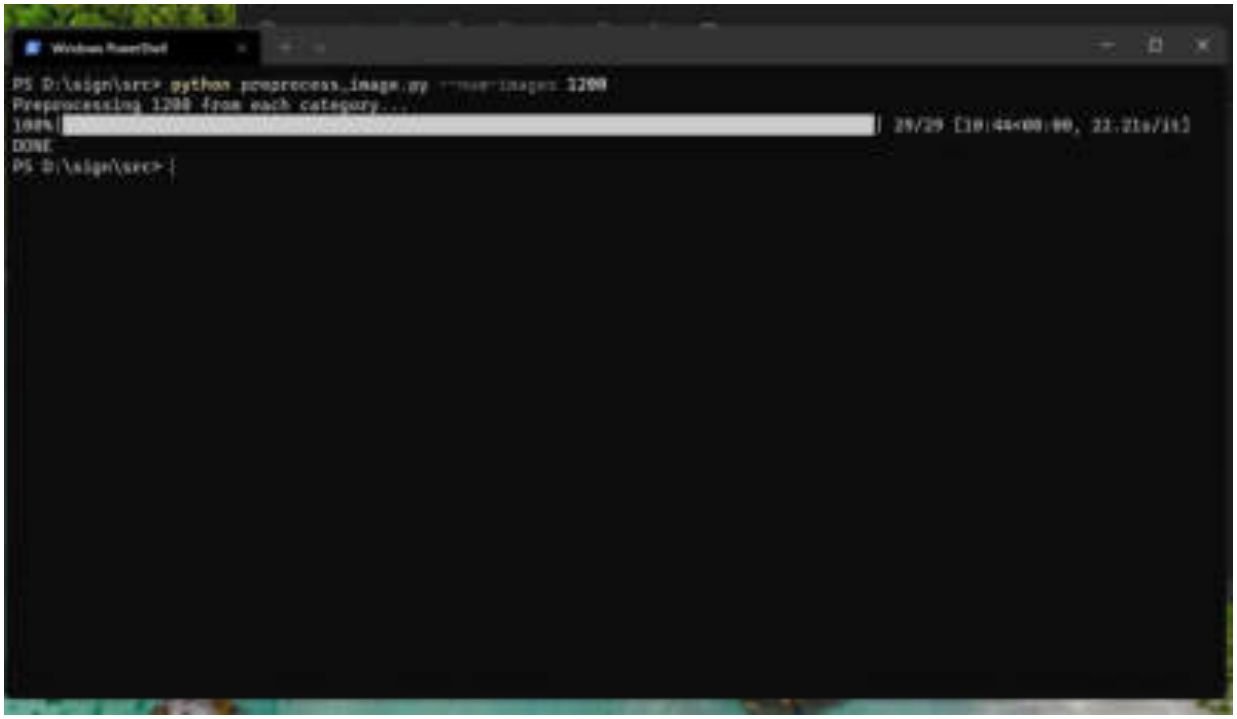


Fig 19. Image while preprocessing

Creating CSV file:

```
import pandas as pd
import numpy as np
import os
import joblib


from sklearn.preprocessing import LabelBinarizer
from tqdm import tqdm
from imutils import paths
# get all the image paths
image_paths = list(paths.list_images('../input/preprocessed_image'))
# create a DataFrame
data = pd.DataFrame()
labels = []
for i, image_path in tqdm(enumerate(image_paths), total=len(image_paths)):
    label = image_path.split(os.path.sep)[-2]
    # save the relative path for mapping image to target
    data.loc[i, 'image_path'] = image_path
    labels.append(label)
labels = np.array(labels)
# one hot encode the labels
lb = LabelBinarizer()
labels = lb.fit_transform(labels)
print(f"The first one hot encoded labels: {labels[0]}")
```

```
print(f"Mapping the first one hot encoded label to its category: {lb.classes_[0]
}")
print(f"Total instances: {len(labels)}")
for i in range(len(labels)):
    index = np.argmax(labels[i])
    data.loc[i, 'target'] = int(index)
# shuffle the dataset
data = data.sample(frac=1).reset_index(drop=True)
# save as CSV file
data.to_csv('../input/data.csv', index=False)
# pickle the binarized labels
print('Saving the binarized labels as pickled file')
joblib.dump(lb, '../outputs/lb.pkl')
print(data.head(5))
```

CNN Model:

```
import torch.nn as nn
import torch.nn.functional as F
import joblib
# load the binarized labels
print('Loading label binarizer...')
lb = joblib.load('../outputs/lb.pkl')

class CustomCNN(nn.Module):
    def __init__(self):
        super(CustomCNN, self).__init__()
        self.conv1 = nn.Conv2d(3, 16, 5)
        self.conv2 = nn.Conv2d(16, 32, 5)
        self.conv3 = nn.Conv2d(32, 64, 3)
        self.conv4 = nn.Conv2d(64, 128, 5)
        self.fc1 = nn.Linear(128, 256)
        self.fc2 = nn.Linear(256, len(lb.classes_))
        self.pool = nn.MaxPool2d(2, 2)
    def forward(self, x):
        x = self.pool(F.relu(self.conv1(x)))
        x = self.pool(F.relu(self.conv2(x)))
        x = self.pool(F.relu(self.conv3(x)))
        x = self.pool(F.relu(self.conv4(x)))
        bs, _, _, _ = x.shape
        x = F.adaptive_avg_pool2d(x, 1).reshape(bs, -1)
        x = F.relu(self.fc1(x))
```

```
        x = self.fc2(x)
        return x
```

Training Model:

```
import pandas as pd
import joblib
import numpy as np
import torch
import random
import albumentations
import matplotlib.pyplot as plt
import argparse
import torch.nn as nn
import torch.nn.functional as F
import torch.optim as optim
import torchvision.transforms as transforms
import time
import cv2
import cnn_models
from tqdm import tqdm
from sklearn.model_selection import train_test_split
from torch.utils.data import Dataset, DataLoader

# construct the argument parser and parse the arguments
parser = argparse.ArgumentParser()
parser.add_argument('-e', '--epochs', default=10, type=int,
    help='number of epochs to train the model for')
args = vars(parser.parse_args())

''' SEED Everything '''
def seed_everything(SEED=42):
    random.seed(SEED)
    np.random.seed(SEED)
    torch.manual_seed(SEED)
    torch.cuda.manual_seed(SEED)
    torch.cuda.manual_seed_all(SEED)
    torch.backends.cudnn.benchmark = True
SEED=42
seed_everything(SEED=SEED)
''' SEED Everything '''
# set computation device
device = ('cuda:0' if torch.cuda.is_available() else 'cpu')
print(f"Computation device: {device}")

# read the data.csv file and get the image paths and labels
df = pd.read_csv('../input/data.csv')
X = df.image_path.values
y = df.target.values
```

```
(xtrain, xtest, ytrain, ytest) = (train_test_split(X, y,
                                test_size=0.15, random_state=42))
print(f"Training on {len(xtrain)} images")
print(f"Validationg on {len(xtest)} images")

# image dataset module
class ASLImageDataset(Dataset):
    def __init__(self, path, labels):
        self.X = path
        self.y = labels
        # apply augmentations
        self.aug = albumentations.Compose([
            albumentations.Resize(224, 224, always_apply=True),
        ])
    def __len__(self):
        return (len(self.X))

    def __getitem__(self, i):
        image = cv2.imread(self.X[i])
        image = self.aug(image=np.array(image))['image']
        image = np.transpose(image, (2, 0, 1)).astype(np.float32)
        label = self.y[i]
        return torch.tensor(image, dtype=torch.float), torch.tensor(label, dtype
=torch.long)

train_data = ASLImageDataset(xtrain, ytrain)
test_data = ASLImageDataset(xtest, ytest)

# dataloaders
trainloader = DataLoader(train_data, batch_size=32, shuffle=True)
testloader = DataLoader(test_data, batch_size=32, shuffle=False)

# model = models.MobineNetV2(pretrained=True, requires_grad=False)
model = cnn_models.CustomCNN().to(device)
print(model)
# total parameters and trainable parameters
total_params = sum(p.numel() for p in model.parameters())
print(f"{total_params:,} total parameters.")
total_trainable_params = sum(
    p.numel() for p in model.parameters() if p.requires_grad)
print(f"{total_trainable_params:,} training parameters.")

# optimizer
optimizer = optim.Adam(model.parameters(), lr=0.001)
# loss function
criterion = nn.CrossEntropyLoss()

# training function
def fit(model, dataloader):
```

```
    print('Training')
    model.train()
    running_loss = 0.0
    running_correct = 0
    for i, data in tqdm(enumerate(dataloader), total=int(len(train_data)/dataloa
der.batch_size)):
        data, target = data[0].to(device), data[1].to(device)
        optimizer.zero_grad()
        outputs = model(data)
        loss = criterion(outputs, target)
        running_loss += loss.item()
        _, preds = torch.max(outputs.data, 1)
        running_correct += (preds == target).sum().item()
        loss.backward()
        optimizer.step()

    train_loss = running_loss/len(dataloader.dataset)
    train_accuracy = 100. * running_correct/len(dataloader.dataset)

    print(f"Train Loss: {train_loss:.4f}, Train Acc: {train_accuracy:.2f}")

    return train_loss, train_accuracy

    #validation function
def validate(model, dataloader):
    print('Validating')
    model.eval()
    running_loss = 0.0
    running_correct = 0
    with torch.no_grad():
        for i, data in tqdm(enumerate(dataloader), total=int(len(test_data)/data
loader.batch_size)):
            data, target = data[0].to(device), data[1].to(device)
            outputs = model(data)
            loss = criterion(outputs, target)

            running_loss += loss.item()
            _, preds = torch.max(outputs.data, 1)
            running_correct += (preds == target).sum().item()

        val_loss = running_loss/len(dataloader.dataset)
        val_accuracy = 100. * running_correct/len(dataloader.dataset)
        print(f'Val Loss: {val_loss:.4f}, Val Acc: {val_accuracy:.2f}')

        return val_loss, val_accuracy

train_loss , train_accuracy = [], []
val_loss , val_accuracy = [], []
start = time.time()
```

```
for epoch in range(args['epochs']):
    print(f"Epoch {epoch+1} of {args['epochs']}")
    train_epoch_loss, train_epoch_accuracy = fit(model, trainloader)
    val_epoch_loss, val_epoch_accuracy = validate(model, testloader)
    train_loss.append(train_epoch_loss)
    train_accuracy.append(train_epoch_accuracy)
    val_loss.append(val_epoch_loss)
    val_accuracy.append(val_epoch_accuracy)
end = time.time()

 # accuracy plots
plt.figure(figsize=(10, 7))
plt.plot(train_accuracy, color='green', label='train accuracy')
plt.plot(val_accuracy, color='blue', label='validataion accuracy')
plt.xlabel('Epochs')
plt.ylabel('Accuracy')
plt.legend()
plt.savefig('../outputs/accuracy.png')
plt.show()

# loss plots
plt.figure(figsize=(10, 7))
plt.plot(train_loss, color='orange', label='train loss')
plt.plot(val_loss, color='red', label='validataion loss')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.legend()
plt.savefig('../outputs/loss.png')
plt.show()
# save the model to disk
print('Saving model...')
torch.save(model.state_dict(), '../outputs/model.pth')
```

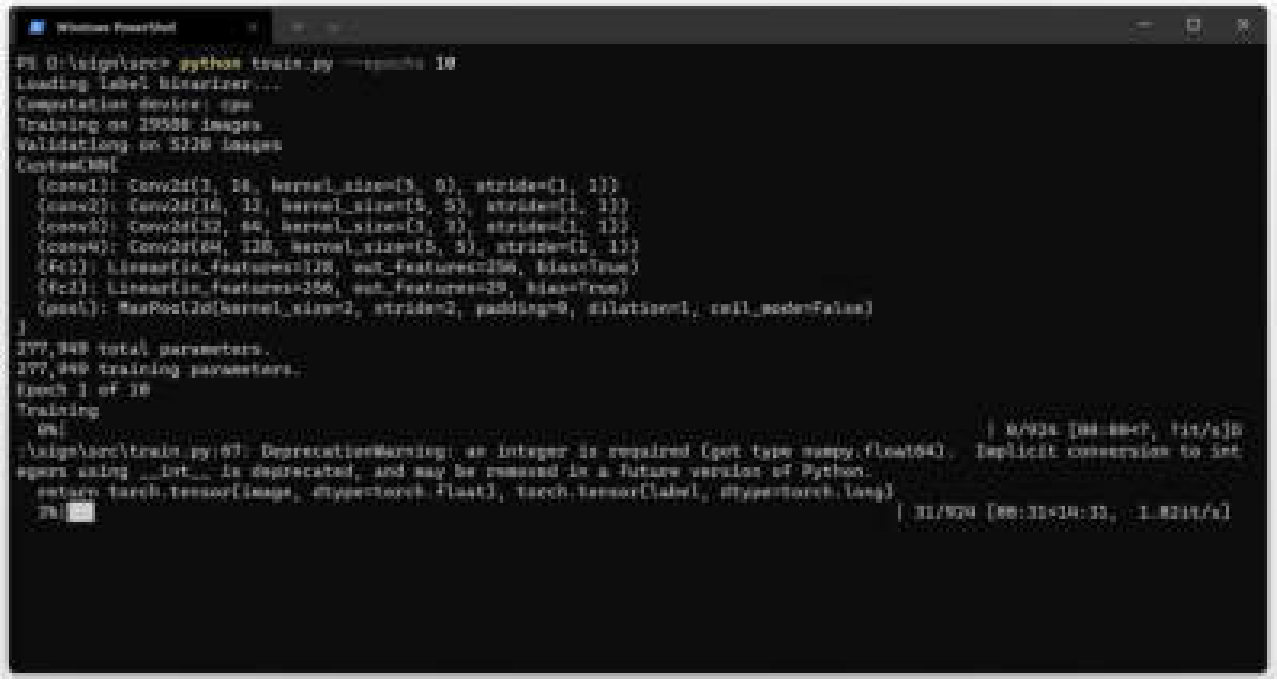


Fig 20. Image while training

Testing the model:

```
import torch
import joblib
import torch.nn as nn
import numpy as np
import cv2
import argparse
import albumentations
import torch.nn.functional as F
import time
import cnn_models
# construct the argument parser and parse the arguments
parser = argparse.ArgumentParser()
parser.add_argument('-i', '--img', default='A_test.jpg', type=str,
    help='path for the image to test on')
args = vars(parser.parse_args())

aug = albumentations.Compose([
                albumentations.Resize(224, 224, always_apply=True),
])

# load label binarizer
lb = joblib.load('../outputs/lb.pkl')
```

```
model = cnn_models.CustomCNN()
model.load_state_dict(torch.load('../outputs/model.pth'))
print(model)
print('Model loaded')

image = cv2.imread(f"../input/asl_alphabet_test/asl_alphabet_test/{args['img']}"
)
image_copy = image.copy()

image = aug(image=np.array(image))['image']
image = np.transpose(image, (2, 0, 1)).astype(np.float32)
image = torch.tensor(image, dtype=torch.float)
image = image.unsqueeze(0)
print(image.shape)

start = time.time()
outputs = model(image)
_, preds = torch.max(outputs.data, 1)
print('PREDS', preds)
print(f"Predicted output: {lb.classes_[preds]}")
end = time.time()
print(f"{(end-start):.3f} seconds")

cv2.putText(image_copy, lb.classes_[preds], (10, 30), cv2.FONT_HERSHEY_SIMPLEX,
0.9, (0, 0, 255), 2)
cv2.imshow('image', image_copy)
cv2.imwrite(f"../outputs/{args['img']}", image_copy)
cv2.waitKey(0)
```

Webcam Testing:

```
import torch
import joblib
import torch.nn as nn
import numpy as np
import cv2
import torch.nn.functional as F
import time
import cnn_models



# load label binarizer
lb = joblib.load('../outputs/lb.pkl')
model = cnn_models.CustomCNN()
model.load_state_dict(torch.load('../outputs/model.pth'))
print(model)
print('Model loaded')

def hand_area(img):
    hand = img[100:324, 100:324]
    hand = cv2.resize(hand, (224,224))
    return hand

cap = cv2.VideoCapture(0)
if (cap.isOpened() == False):
    print('Error while trying to open camera. Plese check again...')
# get the frame width and height
frame_width = int(cap.get(3))
frame_height = int(cap.get(4))
# define codec and create VideoWriter object
out = cv2.VideoWriter('../outputs/asl.mp4', cv2.VideoWriter_fourcc(*'mp4v'), 30,
 (frame_width,frame_height))

# read until end of video



while(cap.isOpened()):
    # capture each frame of the video
    ret, frame = cap.read()
    # get the hand area on the video capture screen
    cv2.rectangle(frame, (100, 100), (324, 324), (20, 34, 255), 2)
    hand = hand_area(frame)
    image = hand
```

```
    image = np.transpose(image, (2, 0, 1)).astype(np.float32)
    image = torch.tensor(image, dtype=torch.float)
    image = image.unsqueeze(0)

    outputs = model(image)
    _, preds = torch.max(outputs.data, 1)




 cv2.putText(frame, lb.classes_[preds], (10, 30), cv2.FONT_HERSHEY_SIMPLEX, 0.9,
 (0, 0, 255), 2)
    cv2.imshow('image', frame)
    out.write(frame)
    # press `q` to exit
    if cv2.waitKey(27) & 0xFF == ord('q'):
        break
# release VideoCapture()
cap.release()
# close all frames and video windows
cv2.destroyAllWindows()
```

5.4 Results:

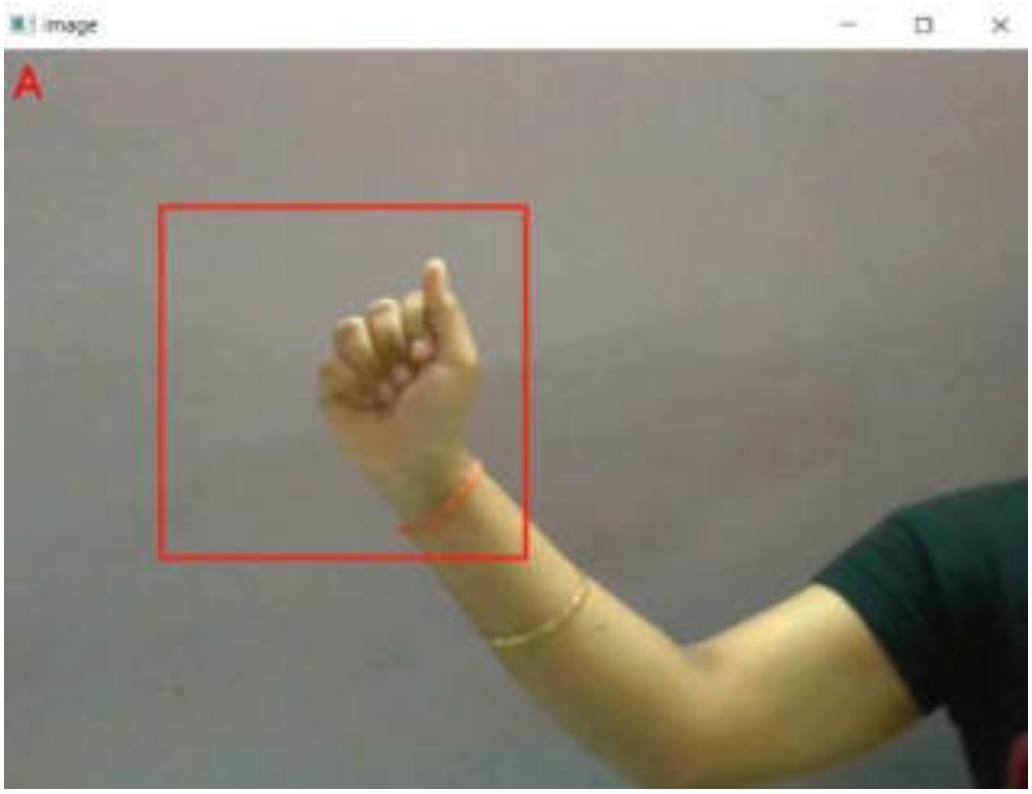


Fig 21. Output-1

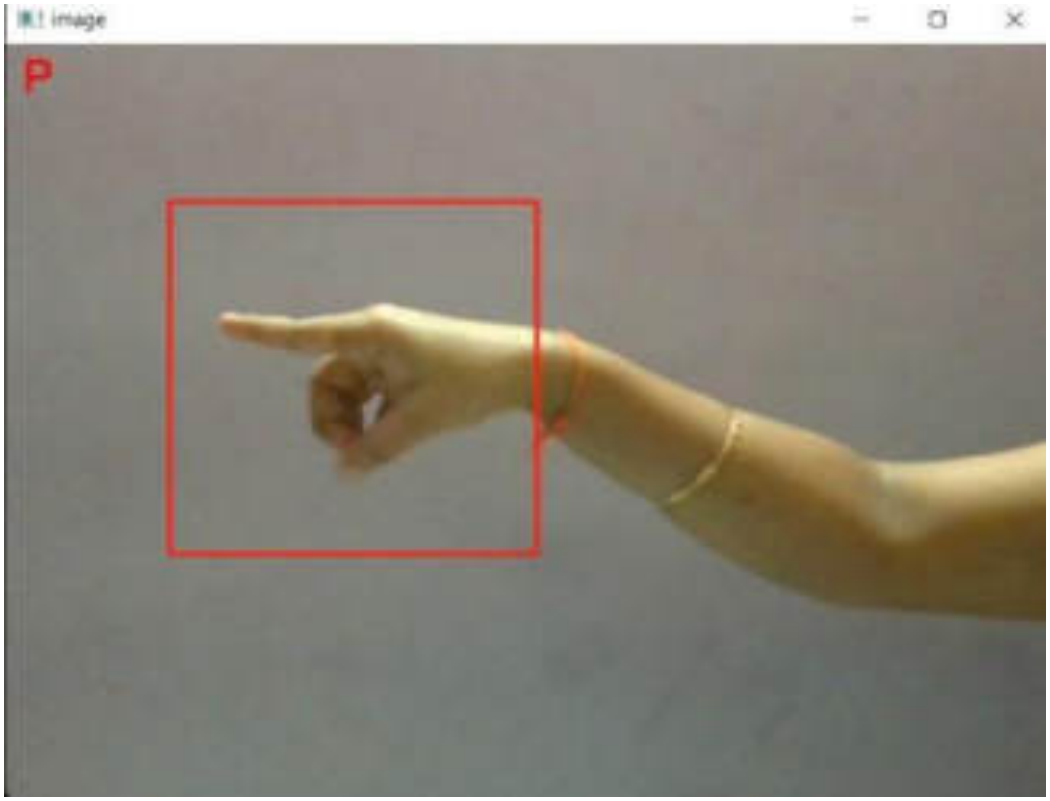


Fig 22. Output-2

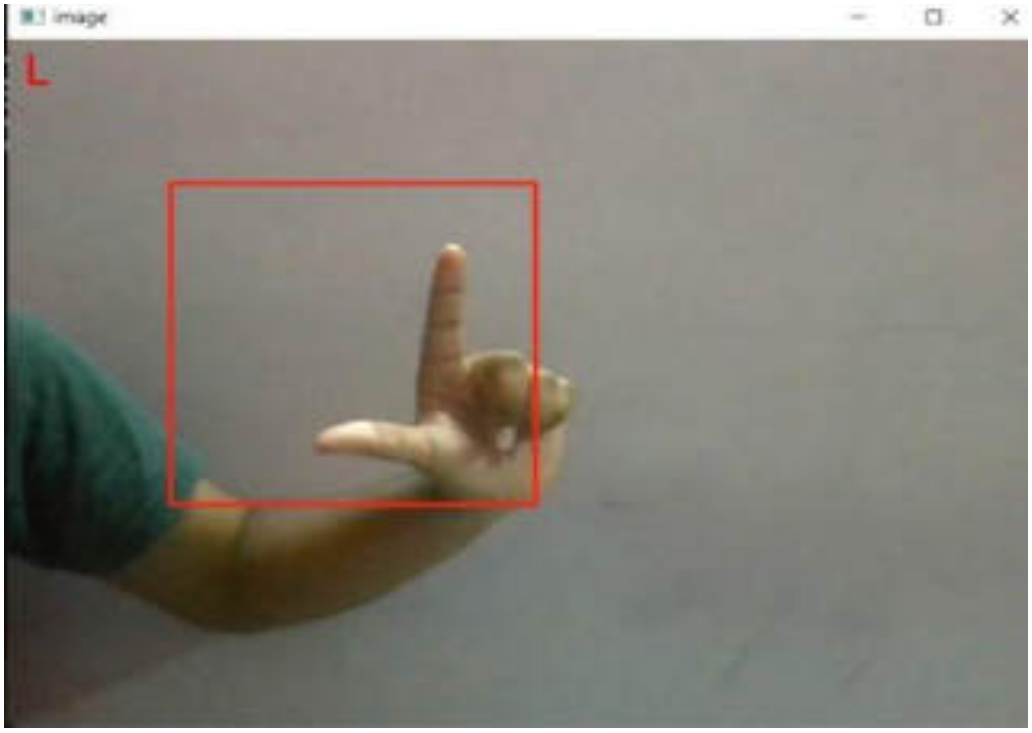


Fig 23. Output-3

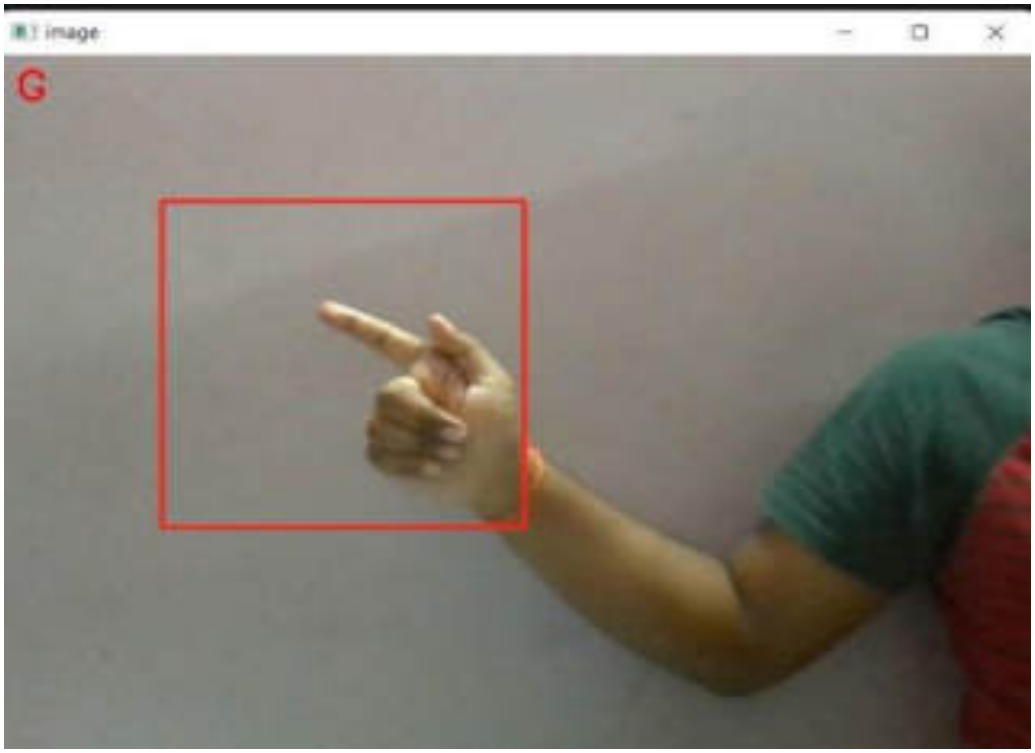


Fig 24. Output-4

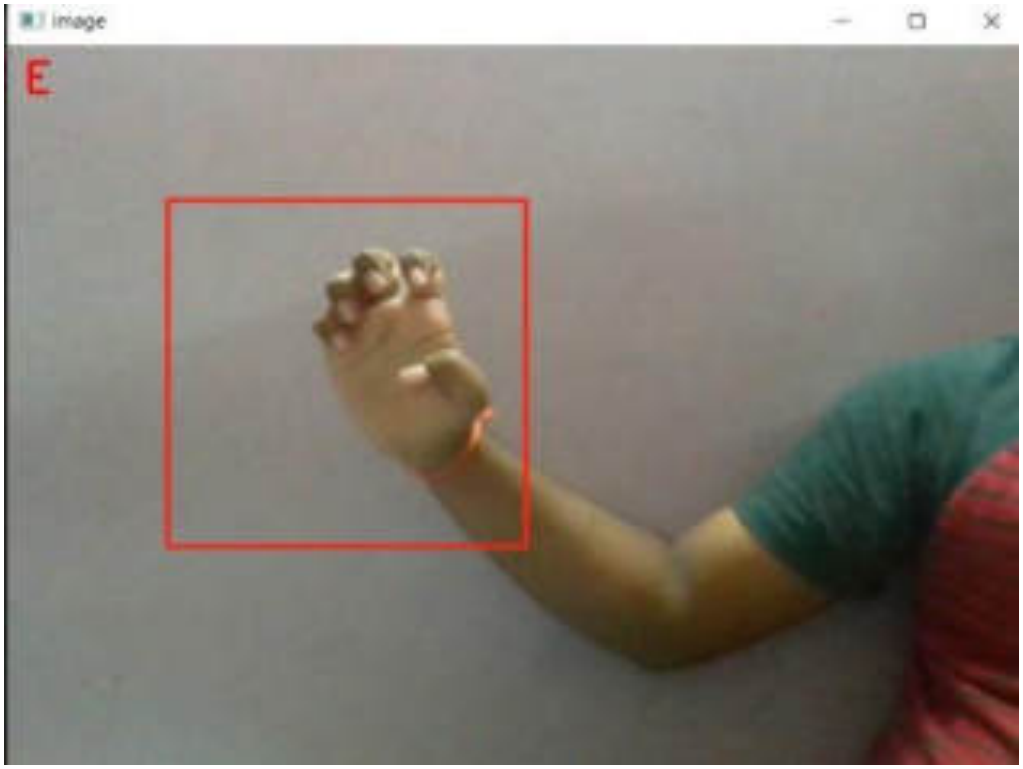


Fig 25. Output-5

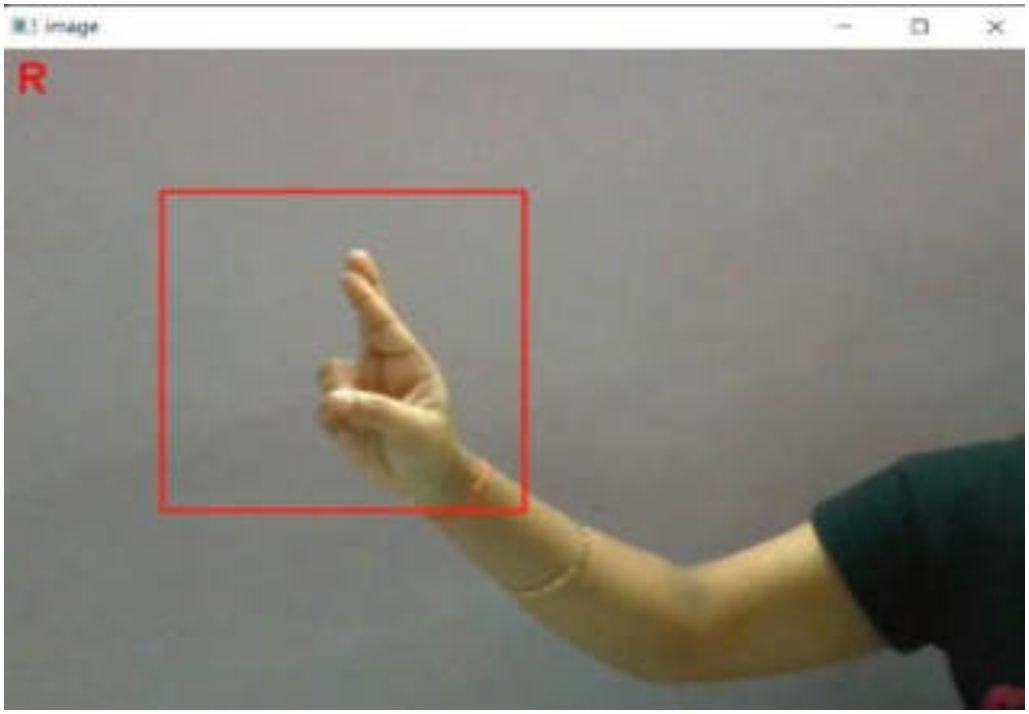


Fig 26. Output-6

5.5 Accuracy graphs:

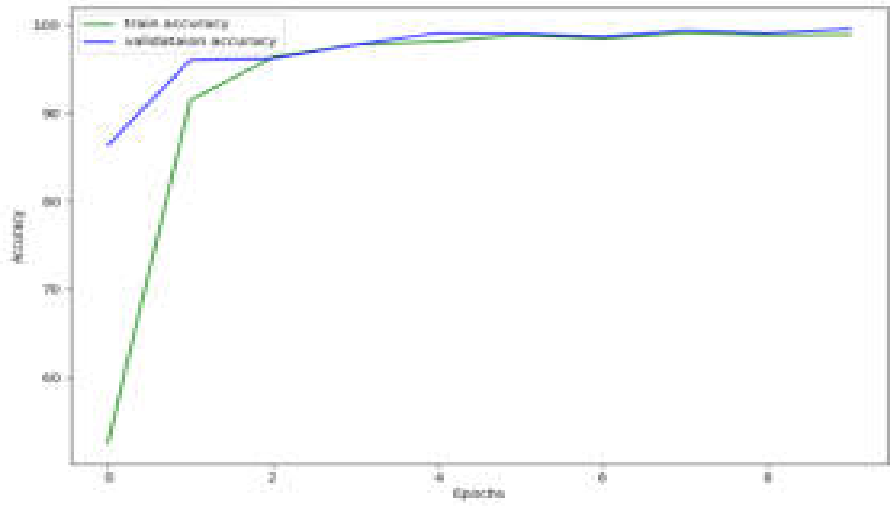


Figure 27. Accuracy graphical plot after training the deep learning neural network

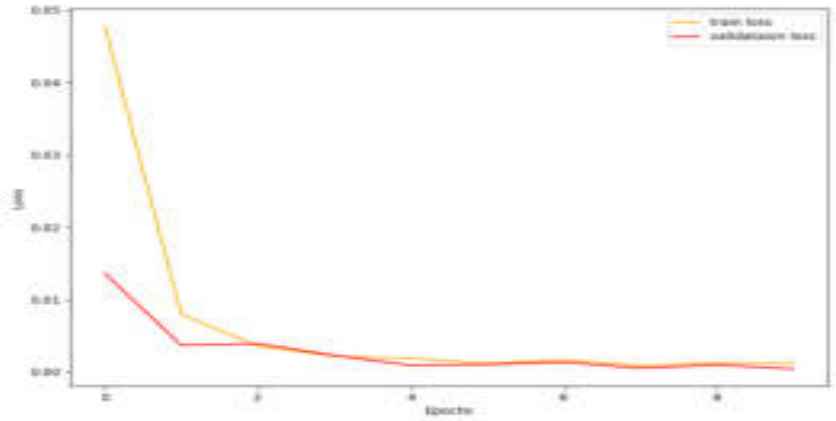


Figure 28. Graphical plot for the loss after training the deep learning neural network

6.Conclusion and Future Work

6.1 Conclusion

The software is designed in such a way that future modification can be done easily. The following conclusions can be deduced from the development
of the project:

- It provides an abstract platform between the user.
- CNN has improved the efficiency as discussed before.
- This application will avoid the manual work.

6.2 Future Works

- Incorporation of other sign languages.
- A mobile application that can do the same.
- And this work can be extended to the applications of Controlling of robot, video gaming etc.
- As we know when no image is supplied or an illegal image is provide then too the CNN would classify the false image into a class. This is the major drwback in using CNN. Reducing false positive would also be a priority work.

BIBLIOGRAPHY

References:

[1] Pratibha Pandey, Vinay Jain, "Hand Gesture Recognition for Sign Language Recognition: A Review", International Journal of Science, Engineering and Technology Research (IJSETR), Volume 4, Issue 3, March 2015.

[2] https://kaggledatasets.com

[3] https://medium.com/machinelearningalgorithms

[4] Wario, Ruth & Nyaga, Casam. (2018). Sign Language Gesture Recognition through Computer Vision.

[5] Vivek Bheda & Dianna Radpour Using Deep Convolutional Networks for Gesture Recognition in American Sign Language (arXiv:1710.06836)

[6] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens & Zbigniew Wojna - Rethinking the Inception Architecture for Computer Vision (arXiv:1512.00567)

[7] http://debuggercafe.com

[8]Haitham Hasan · S. AbdulKareem : Static hand gesture recognition using neural networks.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

"DETECTION OF FAKE ONLINE REVIEWS USING RANDOM FOREST ALGORITHM"

Submitted to
Jawaharlal Nehru Technological University Kakinada
in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY
In
COMPUTER SCIENCE AND ENGINEERING

Submitted by

A.POORNA LAKSHMI PRASANNA(17H71A0592)

P.INDRAJA(17H71A0575)

T.KEDHARNATH(17H71A0579)

B.GREESHMASAI(17H71A0571)

Under the Esteemed Guidance of
Ms. PARVEEN PATHAN
ASSISTANT PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhamic@mictech.ac.in, Website: www.mictech.ac.in

2020-2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
an ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
E mail: dvrhamic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by NAAC With Grade A

2020-2021

CERTIFICATE

This is to certify that the main project entitled **"DETECTION OF FAKE ONLINE REVIEWS USING RANDOM FOREST ALGORITHM"** is a bonafide work carried out by **A. POORNA LAKSHMI PRASANNA (17H71A0592) , P.INDRAJA (17H71A0575) , T.KEDHARNATH (17H71A0579) , B.GREESHMASAI(17H71A0571)** in partial fulfilment for the award of degree of Bachelor of Technology in **Computer Science and Engineering** of **Jawaharlal Nehru Technological University Kakinada** during the year 2020-21.

MS. PARVEEN PATHAN
(Guide Name)

MR.D PRASAD
(Head of the Department)

Examiner

Dr. K.SRINIVAS
(Principal)

ACKNOWLEDGEMENT

We would like to take this opportunity to express my deepest appreciation to the following people for their valuable contributions and assistance with this Project.

Initially, we would like to thank my project supervisor, **MS. PARVEEN PATHAN, Assistant Professor, Department of Computer Science and Engineering** for the guidance and support, especially for the valuable ideas and knowledge shared to me throughout the Project.

We have the immense pleasure in expressing my thanks and deep sense of gratitude to**, Mr. D. Prasad, Head of the Department, Computer Science and Engineering** for extending necessary facilities and support for the completion of the Project.

We whole sincerely acknowledge **Dr. K. Srinivas, Principal and Prof D. PANDURANGA RAO, CEO** for giving opportunity to take up the Project work. We also extend my thanks to all faculty members of **Computer Science and Engineering**, for their valuable guidance and encouragement.

We would like to extend my warm appreciation to all my friends for sharing us their knowledge, valuable contributions and help with this Project.

Finally, my special thanks go to my family for their continuous support and help throughout my academic years and for their continual support and encouragement for the completion of the project.

DECLARATION

We **POORNA LAKSHMI PRASANNA, INDRAJA, KEDHARNATH, GREESHMASAI** of the Main-Project "**DETECTION OF FAKE ONLINE REVIEWS USING RANDOM FOREST ALGORTHM**", hereby declare that the matter embodied in this Project is the genuine work done by us and has not been submitted either to this university or to any other university/institute for the fulfilment of the requirement of any course of study.

A. POORNA LAKSHMI PRASANNA(17H71A0592)
P. INDRAJA(17H71A0575)
T. KEDHARNATH(17H71A0579)
B. GREESHMASAI(17H71A0571)

CONTENTS

CHAPTERS **PAGENO**

LIST OF FIGURES

ABSTRACT

Online reviews have great impact on today's business and commerce. Decision making for purchase of online products mostly depends on reviews given by users. Hence, opportunistic individuals or groups try to manipulate product reviews for their own interests. Generally, the online reviews are given by the users through their accessible accounts like Gmail, Face-book etc. Some approaches like review content based method and behavior feature based method are proposed to detect fake reviews. Content based method mainly focuses on what is written on the review that is the text of the review where behavior based method focuses on country, ip-address, number of posts of the reviewer etc. These proposed approaches are supervised classification models which results in high rate of misclassification, high time complexity and less accuracy. To overcome these disadvantages, we are using Random Forest algorithm to improve our performance of classification. By this, the system is very fast and they offer efficient estimates with more accuracy.

CHAPTER 1

INTRODUCTION

1.INTRODUCTION

1.1 PROBLEM STATEMENT

Technologies are changing rapidly. Old technologies are continuously being replaced by new and sophisticated ones. These new technologies are enabling people to have their work done efficiently. Such an evolution of technology is online marketplace. We can shop and make reservation using online websites. Almost, everyone of us checks out reviews before purchasing some products or services. Hence, online reviews have become a great source of reputation for the companies. Also, they have large impact on advertisement and promotion of products and services. With the spread of online marketplace, fake online reviews are becoming great matter of concern. People can make false reviews for promotion of their own products that harms the actual users. Also, competitive companies can try to damage each others reputation by providing fake negative reviews.

1.1 METHODOLOGY

For detection of fake online reviews, we start with raw text data. We have used a dataset which was already labeled by the previous researchers. We remove unnecessary texts like article and prepositions in the data. Then these text data are converted into numeric data for making them suitable for the classifier. Important and necessary features are extracted and then classification process took place. We have used the dataset called “YELP.” We have used word frequency count, sentiment polarity and length of the review as our features. We have not taken n-gram or parts of speech as features because these are the derived features from bag of words and may cause over-fitting.

1.2 OBJECTIVE OF THEPROJECT

In the proposed project we have used random forest algorithm for detecting fake online reviews. As technology changes, it becomes difficult to detect the fake reviews. With the rise of machine learning, it becomes easy to detect those fake reviewers.

CHAPTER 2
LITERATURE SURVEY

2.LITERATURE SURVEY

A number of studies have been conducted which focused on spam detection in e-mail and on the web, however, only recently have any studies been conducted on opinion spam. Jindal and Liu (2008) have worked on “Opinion Spam and Analysis” and have found that opinion spam is widespread and different in nature from either e-mail or Web spam. They have classified spam reviews into 3 types: Type 1, Type 2 and Type 3. Here Type 1 spam reviews are untruthful opinions that try to mislead readers or opinion mining systems by giving untruthful reviews to some target objects for their own gains. Type 2 spam reviews are brand only reviews, those that comment only on the brand and not on the products. Type 3 spam reviews are not actually reviews, they are mainly either advertisements or irrelevant reviews which do not contain any opinions about the target object or brand. Although humans detect this kind of opinion spam they need to be filtered, as it is a nuisance for the end user. Their investigation was based on 5.8 million reviews and 2.14 million reviewers (members who wrote at least one review) crawled from amazon.com and they have discovered that spam activities are widespread. They have regarded spam detection as a classification problem with two classes, spam and non-spam. And have built machine-learning models to classify a review as either spam or non-spam. They have detected type 2 and type 3 spam reviews by using supervised learning with manually labeled training examples and found that the highly effective model is logistic regression model. However, to detect type 1 opinion spam, they would have had to manually label training examples. Thus they had to use duplicate spam reviews as positive training examples and other reviews as negative examples to build amodel.

In the paper "Finding Deceptive Opinion Spam by Any Stretch of the Imagination" by Ott, et al. 2011, they have given focus to the deceptive opinion spam i.e. the fictitious opinions which are deliberately written to sound authentic so as to deceive the user. The user cannot easily identify this kind of opinion spam. They have mined all 5-star truthful reviews for 20 most famous hotels in Chicago area from trip advisor and deceptive opinions were gathered for the same hotels using amazon mechanical trunk (AMT). They first asked human judges to evaluate the review and then they have automated the task for the same set of reviews, and they found that automated classifiers outperform humans for each metric. The task was viewed as standard text categorization task, psycholinguistic deceptive detection and genre identification. The performance from each approach was compared and they found that the psycholinguistic deceptive detection and genre

identification approach was outperformed by n-gram based text categorization, but a combined classifier of n-gram and psychological deception features achieved nearly 90% cross-validated accuracy. Finally they came into a conclusion that detecting deceptive opinions is well beyond the capabilities of humans. Since then, various dimensions have been explored: detecting individual (Lim et al., 2010) and group spammers (Mukherjee et al., 2012), time-series (Xie et al., 2012) and distributional analysis (Feng et al.,2012a).

Yoo and Gretzel (2009) gather 40 truthful and 42 deceptive hotel reviews and, using a standard statistical test, they have manually compared the psychologically relevant linguistic differences between them. In (Mukherjee, et al., 2013), authors have briefly analyzed "What yelp filter might be doing?" by working with different combination of linguistic features like unigram, bigram, distribution of parts of speech tags and yielding detection accuracy. Authors have found that a combination of linguistic and behavioral features comparatively yielded more accuracy.

Scraping process: As the consumer electronics field has not been studied before, there is not an available dataset to experiment with, so the starting point consists on data collection. Furthermore, Yelp's filter has been used as a reference for labeling reviews as fake or not since, according to its CEO, Yelp's filtering algorithm has evolved over the years to filter fake reviews. Also, this filter has been claimed to be highly accurate. As Yelp shows both trustful and filtered reviews, it is possible not only to extract information about reviews and users but also whether if they had been filtered or not. So, the first step was developing a web scraper to gather the necessary information for the experiments. This web scraper was programmed in Python using Scrapy, a library which offers the possibility of building web crawlers. The resulting corpus was compound by reviews from four important cities of the USA: New York, San Francisco, Los Angeles and Miami. Yelp offers the possibility of searching category businesses in each city, so the scraping process is focused on the pages in which all electronic businesses from the different selected cities appeared. For each of these pages, all businesses appearing on it were selected in order to retrieve their information. Secondly, for each business site all the reviews appearing on it were collected along with their user profile URL. In this step, reviews were also labeled as fake or not according to the Yelp filter. The last step of the corpus creation consisted on gathering all the information appearing on each user profile site through the URLs which were collected in the previous step. Each of those profile sites show all the necessary information to develop the features of the proposed framework.

FAKE FEATURE FRAMEWORK (F3)

In this section we introduce a feature framework Fake Feature Framework (F3) for organizing the

extraction and characterization of features in fake detection. Its definition is inspired on the analysis of previous research, and includes a novel definition of social features. Previous works have classified features into textual, behavioral and product features. Our main contribution consists in providing a more detailed classification of user centric features, taking advantage of the social aspects of a social network such as Yelp.

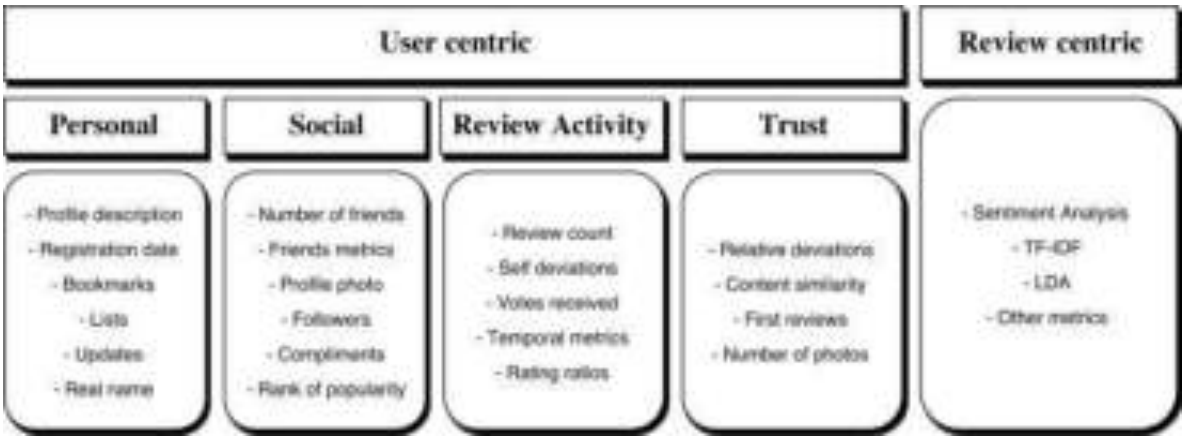


Fig1 Fake Feature Framework (F3)

User centric features consider information related to how users behave in a social network such as Yelp and which information users provide; and are divided into four types: Personal Features (P), Social Features (S), Reviewing Activity Features (RA), and Trusting Features (T). The first type, P, is information related to user profile, such as the self description written by the user; users' businesses subscriptions, known as bookmarks; lists containing several bookmarks; registration date; updates made on self reviews; and the user's real name. This information is available in social review networks, such as Yelp, and can be automatically obtained. The second type, S, is the feature set related to the way the user interacts with other users. In our case, we have identified these features by inspecting the Yelp social features. Our hypothesis is that social activity can help in the classification task, since social features can help to extend the context of the linguistics features. Several works have also proposed this approach in a number of applications, such as sentiment analysis or stance detection. Social features included here are number of friends, metrics gathered from friends such as their number of reviews or friends

number of followers, number of compliments, rank of popularity and the presence of a profile photo. The third type, RA is the feature set related to the way the user posts his reviews. Some examples of reviewing activity features are review counts , self review deviations , number of

received votes (in Yelp there are three types: ‘cool’, ‘funny’ and ‘useful’), number of posted tips (which are shorter reviews that appear on Yelp), maximum number of reviews in a date or specific review counts in temporal windows. Additionally, metrics related to the ratios of positive and negative reviews are included here. A last type, T, is the feature set that aims at pointing out inconsistencies or abnormal behavior in the user review activity. For example, a high content similarity of the reviews of a given user could reveal that said user may be using templates for reviewing. In this regard, rating deviations with respect to other users reviewing the same businesses can also offer greater insight. Other features included in this type are the number of first reviews and number of uploaded photos.

CHAPTER 3

ALGORITHM

3.ALGORITHM

3.1NTRODUCTION TO RANDOM FORESTALGORITHM

Here Random Forest algorithm can use both for classification and regression kind of hitches.

This algorithm is used both for classification and regression.

We are going to learn, how the random forest algorithm works in machine learning for the classification task.

3.2TABLE OFCONTENTS:

- What is Random Forest algorithm?
- Why Random Forest algorithm?
- Random Forest Algorithm real life example
  Random Forest Example
- How Random Forest Algorithm works
  Pseudo code Random Forest creation
  Pseudo code for performing predictions
- Random Forest Algorithm Applications
  Banking
  Medicine
  Stock Market
  E-Commerce
- Advantages of Random Forest Algorithm

3.2.1 RANDOM FORESTALGORITHM

Random forest is a supervised learning algorithm which is used for both classification as well as regression. But however, it is mainly used for classification problems. As we know that a forest is made up of trees and more trees means more robust forest. Similarly, random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result.

3.2.2 WHY RANDOM FORESTALGORITHM?

We can see the following advantages:

- The same random forest algorithm or the random forest classifier can use forboth classification and the regression task.
- Random Forest classifier will handle the missing values.
- When we have more trees in the forest, random forest classifier won't over fit the model.
- Can model the random forest classifier or categorical values also.

3.2.3 RANDOM FOREST ALGORITHM REAL LIFEEXAMPLE:

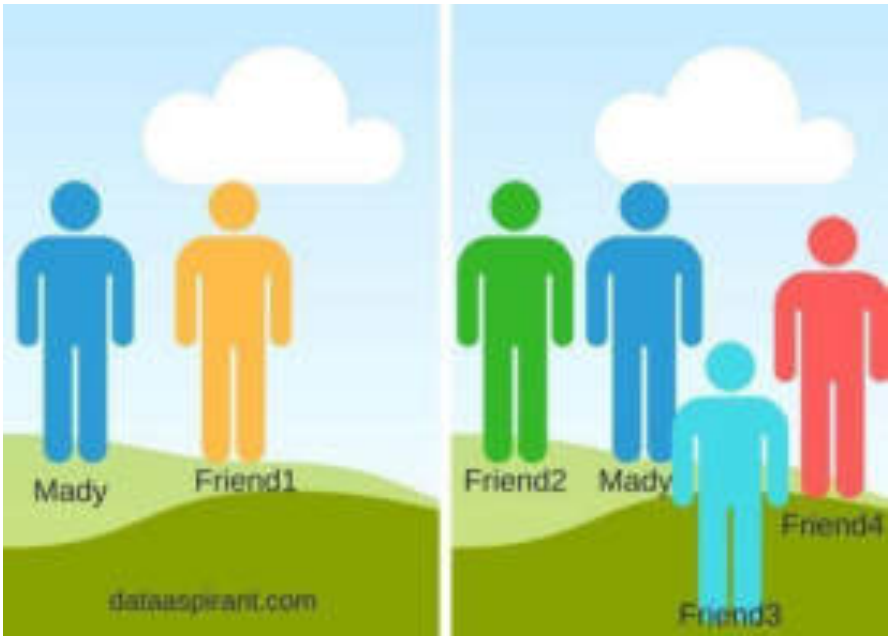


Fig 3.2.3 Random Forest AlgorithmExample

Suppose Mady wants to go to different places that he may like for his two-week vacation, and he asks his friend for advice. His friend will ask where he has been to already, and whether he likes the places that he's visited. Based on Mady's answers, his friend starts to give the recommendation. Here, his friend forms the decision tree. Mady wants to ask more friends for advicebecausehethinksonlyonefriendcannothelphimmakeanaccuratedecision.Sohisother friends also ask him random questions and finally provide an answer. He considers the place with the most votes as his vacation decision. Here, the author provides an analysis for this example. His one friend asked him some questions and gave the recommendation of the best place based on the answers. This is a typical decision tree algorithm approach. The friend created the rules based on

the answers and used the rules to find the answer that matched the rules. Mady's friends also randomly asked him different questions and gave answers, which for Mady are the votes for the place. At the end, the place with the highest votes is the one Mady will select to go. This is the typical Random Forest algorithm approach.

3.2 HOW RANDOM FOREST ALGORITHMWORKS?

The pseudocode for random forest algorithm can spilt into two stages.

- Random forest creation pseudocode.
- Pseudocode to perform prediction from the created random forest classifier.

Random Forest pseudocode:

1. Randomly select "k" features from total "m" features, where k<<m.
2. Among the "k" features, calculate the node "d" using the best spilt point.
3. Repeat 1 to 3 steps until "l" number of nodes has been reached.
4. Build forests by repeating steps 1 to 4 for "n" number of times to create "n" number of trees.

Random Forest prediction pseudocode:

1. To perform prediction using the trained random forest algorithm uses the below pseudocode.
2. Takes the test features and use the rules of each randomly created decision tree to predict the outcome and stores the predicted outcome(target).
3. Calculate the votes for each predicted target.
4. Consider the high voted predicted target as the final prediction from the random forest algorithm.

3.3 RANDOM FOREST ALGORITHMAPPLICATIONS

Fig 3.4 Random Forest Algorithm Applications

The random forest algorithm used in wide varieties applications. Here, we are going to address few of them.

1) Banking
2) Medicine
3) Stock Market
4) E-commerce

1) **Banking:** For the application in banking ,Random Forest algorithm is used to find loyal customers, who mean customers who can take out plenty of loans and pay interest to the bank properly, and fraud customers, which means customers who have bad records like failure to payback a loan on time or have dangerous actions.

2) **Medicine:**Fortheapplicationinmedicine,RandomForestalgorithmcanbeusedtoboth identify the correct combination of components in medicine, and to identify diseases by analysing the patient's medical records.

3) **Stockmarket:**Fortheapplicationinthestockmarket,RandomForestalgorithmcanbe used to identify a stock's behavior and the expected loss or profit.

4) **E-commerce:**Fortheapplicationine-commerce,RandomForestalgorithmcanbeused for predicting whether the customer will like the recommend products, based on the experience of similar customers.

3.4 ADVANTAGES OF RANDOM FORESTALGORITHM:

- The over fitting problem will never come when we use the random forest algorithm.
- The same random forest algorithm can be used for both classification and regression task.
- The random forest algorithm can be used for feature engineering.
- Which means identifying the most important features out of the available features from the training dataset.

DATASET

Totally we use 11 attributes in our dataset called YELP.

- User_id
- Review_id
- Date
- Text
- Business_id
- Stars
- Votes.cool
- Votes.funny
- Votes.useful
- Goodforlunch
- categories

CHAPTER 4
SYSTEM ANALYSIS

4.SYSTEM ANALYSIS

4.1STUDY OF THE SYSTEM

1. Numpy
2. Pandas
3. Scikit-learn

1. Numpy:

Numpy is a general-purpose array-processing package. It affords a high-performance multidimensional array object, and tools for working with these arrays.
It is the fundamental package for scientific computing with Python. It encompasses innumerable features including these important ones:

- A powerful N-dimensional array object
- Tools for integrating C/C++ and Fortran code
- Sophisticated(broadcasting)functions
- Useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, Numpy can also be used as a competent multi-dimensional ampule of generic data. Arbitrary data-types can be identified using Numpy which allows Numpy to seamlessly and speedily assimilate with a wide variety of databases.

2. Pandas:

Pandas is an open-source Python library providing high-performance data manipulation and scrutiny tool using its influential data structures. Python was majority used for data mining and preparation. It had very little contribution concerning data analysis. Pandas solved this enigma. Here by using Pandas, we can succeed five typical steps in the processing and scrutiny of data, regardless of the origin of the data load, prepare, manipulate, model and analyze. Python with Pandas is used in a wide range of fields comprising speculative and commercial domains including finance, economics, statistics, analysis etc.

3. Scikit-Learn:

Scikit-Learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python. It is licensed under a permissive simplified BSD license and is distributed under many Linux distributions, encouraging and commercial use. The library is built upon the SciPy(Scientific Python) that must be installed before you can use scikit-learn. This stack that includes:

- **Numpy:** Base n-dimensional array package.
- **SciPy:** Fundamental library for scientific computing.
- **IPython:** Enhanced interactive console.
- **Pandas:** Data structures and analysis

CHAPTER 5
PROPOSED SYSTEM

5.PROPOSED SYSTEM

5.1SYSTEMARCHITECTURE

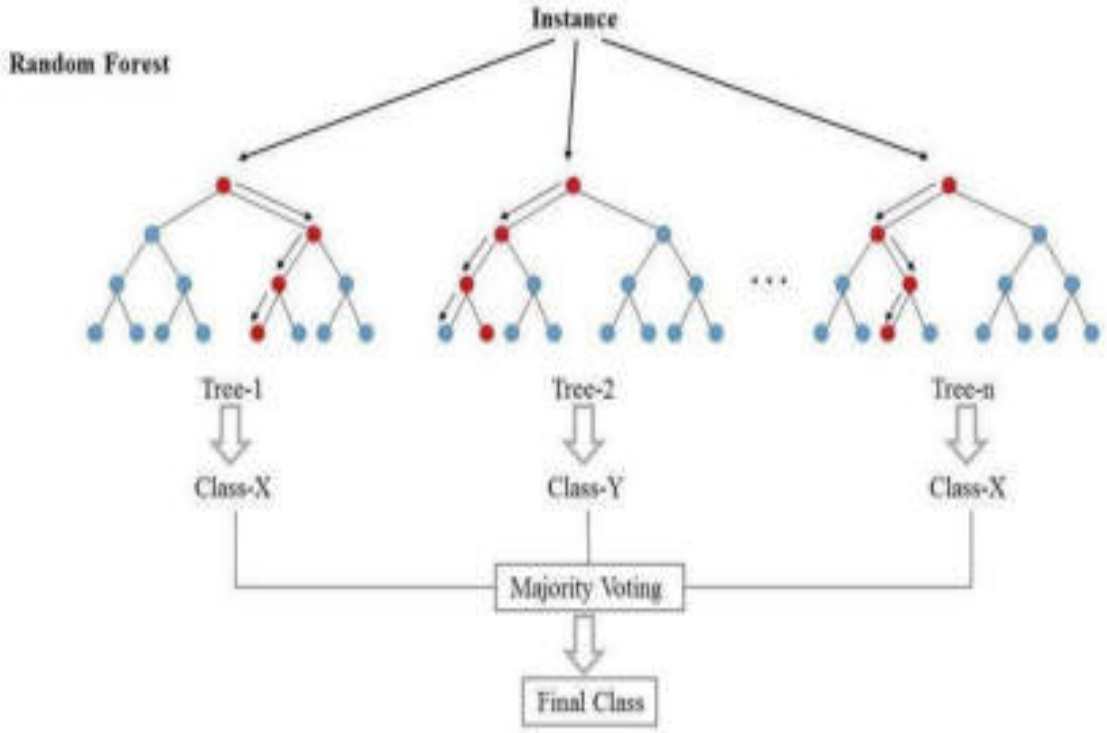


Fig 5.1 System Architecture

5.2FUNCTIONALREQUIREMENTS

User

- Load Data
- Data Analysis
- Data Pre-processing
- Model Building
- Prediction

5.3NONFUNCTIONALREQUIREMENTS

1. Secure access of confidential data. SSL can be used.
2. 24 * 7availability.
3. Better component design to get better performance at peak time.

4. Flexible service based architecture will be highly desirable for future extension.

5.4EXISTING SYSTEM

CONTENT BASEDMETHOD:

Content based methods focus on what is the content of the review. That is the text of the review or what is told in it. Heydarietal have attempted to detect spam review by analyzing the linguistic features of the review. Ottetal used three techniques to perform classification. These three techniques are- genre identification, detection of psycholinguistic deception and text categorization

1) **Genre Identification**: The parts-of-speech (POS) distribution of the review are explored by Ottetal. They used frequency count of POS tags as the features representing the review for classification.
2) **Detection of Psycholinguistic Deception**: The psycholinguistic method approaches to assign psycholinguistic meanings to the important features of are view.
3) **Text Categorization**: Ottetal experimented n-gram that is now popularly used as an important feature in fake review detection.

BEHAVIOR BASED METHOD:

It focuses on the reviewer that includes characteristics of the person who is giving the review. Lim et al. [7] addressed the problem of review spammer detection, or finding users who are the source of spam reviews. People who post intentional fake reviews have significantly different behavior than the normal user. They have identified the following deceptive rating and review behaviors.

Giving unfair rating too often: Professional spammers generally posts more fake reviews than the real ones. Suppose a product has average rating of 9.0 out of 10. But a reviewer has given 4.0 rating. Analyzing the other reviews of the reviewer if we find out that he often gives this type of unfair ratings than we can detect him as a spammer.

Giving good rating to own country's product: Some- times people post fake reviews to promote products of own region. This type of spamming is mostly seen in case of movie reviews. Suppose, in an international movie website an Indian movie have the rating of 9.0 out of10.0,

where most of the reviewers are Indian. This kinds of spamming can be detected using address of the reviewers.

Giving review on a vast variety of product: Each person has specific interests of his own. A person gener- ally is not interested in all types of products. Suppose a person who loves gaming may not be interested in classic literature. But if we find some people giving reviews in various types of products which exceeds the general behavior then we can intuit that their reviews are intentional fake reviews.

Deceptive online review detection is generally considered as a classification problem and one popular approach is to use supervised text classification techniques. These techniques are robust if the training is performed using large datasets of labeled instances from both classes, deceptive opinions (positive instances) and truthful opinions (negative examples).

Disadvantages:

1. Assuring the quality of the reviews is difficult i.e. misclassification rate ishigh.
2. High time complexity.
3. Less accuracy.

5.2 PROPOSEDSYSTEM

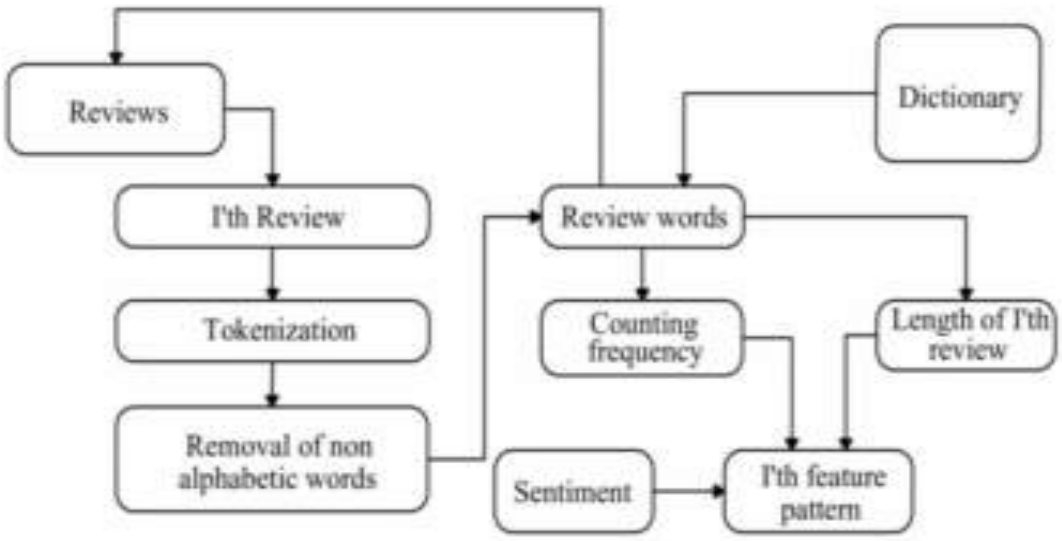


Fig 5.5 Stages of proposed feature extraction process

From the above figure, we can see that, when we are working with i'th review, it's corresponding features are generated in the following procedure.

1. Each review goes through tokenization process first. Then, unnecessary words are removed and candidate feature words are generated.
2. Each candidate feature words are checked against the dictionary and if its entry is available in the dictionary then its frequency is counted and added to the column in the feature vector that correspond to the numeric map of the word.
3. Alongside with counting frequency, the length of the review is measured and added to the feature vector.
4. Finally, sentiment score which is available in the data set is added in the feature vector. We have assigned negative sentiment as zero valued and positive sentiment as some positive valued in the feature vector.

IDENTIFICATION FACTORS CONSIDERED FOR FAKE REVIEW DETECTION

1) Repeated reviews and rating: Same review comment and rating repeatedly for the same product or service.

2) Username with integers: guanine user or buyer must exhibit his/her real name and it should not contain only numbers. As only number indicates spammer since guanine buyer will have name with address in alphanumeric notation.

3) Star (*) Rating only: Just giving star rating doesn't mean review is guanine as customer will also try to say something before giving star rating. Hence just star rating considered asfake.

4) No category Details: actual buyer will post his review by specifying the features of product. So if user posts his review without specifying product category or feature, it is considered as fake.

5) Rating Vs Review Sentiment: If review found to be with highest rating but low sentiment or vice-versa i.e. if user posting high rating but his/her review comments does not match to the rating, such reviews are considered to be fake.

6) Review Length: a genuine buyer will post his review by specifying what did he like or didn't like by addressing the features of product or services. Thus review must be within a predefined length.

Advantages:

1) It provides more accurate results compared to SVM.
2) No mis classification occurs.
3) Easy to use.

5.3 SYSTEMREQUIREMENTS

5.6.1 SOFTWAREREQUIREMENTS

- Operating system : Windows7.
- Coding Language :Python
- Database :MYSQL

5.6.2 HARDAWREREQUIREMENTS

- Hard Disk : 120GB.
- Monitor : 15''LED
- Input Devices : Keyboard, Mouse
- Ram : 1GB

EXAMPLE DATASET

Age	Competition	Type	Profit
old	yes	software	down
old	no	software	down
old	no	hardware	down
mid	yes	software	down
mid	yes	hardware	down
mid	no	hardware	up
mid	no	software	up
new	yes	software	up
new	no	hardware	up
new	no	software	up

EXAMPLE SOLVING

First of all, we will find information gain from the Target attribute which is profit. From the dataset P=5,N=5

$$I(p, n) = \frac{-p}{p+n}\log_2\left(\frac{p}{p+n}\right) - \frac{n}{p+n}\log_2\left(\frac{n}{p+n}\right)$$

$$I(5, 5) = \frac{-5}{5+5}\log_2\left(\frac{5}{5+5}\right) - \frac{5}{5+5}\log_2\left(\frac{5}{5+5}\right)$$

$$I(5, 5) = 1$$

So, the information gain of the overall dataset is equal to 1.

To determine the root node we need to compute the entropy of each attribute.

Entropy (age):

Information gain of each categorical data

age	p_i	n_i	$I(p_i, n_i)$
old	0	3	I (0, 3) = 0
mid	2	2	I (2, 2) = 1
new	3	0	I (3, 0) = 0

Entropy of age:

$$E(A) = \text{Entropy } E(A) = \sum_{i=1}^{v} \frac{p_i+n_i}{p+n} (I(p,n))$$

$$E(A) = \frac{0+3}{5+5}(0) + \frac{2+2}{5+5}(1) + \frac{3+0}{5+5}(0)$$

$$E(A) = 0 + 0.4 + 0$$

$$E(A) = 0.4$$

Gain (Age):

Gain (age) = Information Gain-Entropy (age)

Gain (age) = I (p,n)-Entropy (age)

Gain (age) = 1-0.4

Gain (Age) = 0.6

Entropy (competition):

competition	p_i	n_i	$I(p_i, n_i)$
yes	1	3	I (1, 3) = 0.81
no	4	2	I (4, 2) = 0.91

Entropy of comp:

$$E(A) = \text{Entropy } E(A) = \sum_{i=1}^{v} \frac{p_i+n_i}{p+n} I(p,n))$$

$$E(A) = \frac{1+3}{5+5}(0.81) + \frac{4+2}{5+5}(0.91)$$

E (A) = 0 .8754

Gain(Comp):

Gain (competition) = Information Gain-Entropy (competition)

Gain (competition) = I (p,n)-Entropy (competition)

Gain (competition) = 1-0.8754

Gain (competition) = 0.124515

Entropy (type):

type	p_i	n_i	$I(p_i, n_i)$
software	3	3	I (3, 3) = 1
hardware	2	2	I (2, 2) = 1

Entropy of type:

$$E(A) = \text{Entropy } E(A) = \sum_{i=1}^{v} \frac{p_i + n_i}{p+n} (I(p,n))$$

$$E(A) = \frac{3+3}{5+5}(1) + \frac{2+2}{5+5}(1)$$

$$E(A) = 1$$

Gain (type):

Gain (type) = Information Gain-Entropy (type)

Gain (type) = I (p,n) - Entropy (type)

Gain (type) = 1-1

Gain (type) = 0

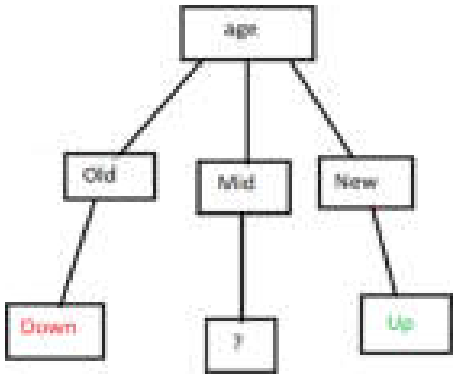
Gain of (age, competition, type)

Gain of age = 0.6

Gain of competition = 0.12

Gain of type = 0

Check the Maximum Gain of all the attributes:



Age	Competition	Type	Profit
mid	yes	software	down
mid	yes	hardware	down
mid	no	hardware	up
mid	no	software	up

From the above dataset, P=2,N= 2

Information Gain:

$$I(p, n) = \frac{-p}{p+n}\text{Log}_2\left(\frac{p}{p+n}\right) - \frac{n}{p+n}\text{Log}_2\left(\frac{n}{p+n}\right)$$

$$I(2, 2) = \frac{-2}{2+2}\text{Log}_2\left(\frac{2}{2+2}\right) - \frac{2}{2+2}\text{Log}_2\left(\frac{2}{2+2}\right)$$

$$I(2, 2) = 1$$

Entropy (competition):

Information gain of each categorical data

competition	p_i	n_i	$I(p_i, n_i)$
yes	2	0	$I(2,0) = 0$
no	0	2	$I(0,2) = 0$

Entropy of comp:

$$E(A) = \text{Entropy } E(A) = \sum_{i=1}^{v} \frac{p_i+n_i}{p+n}(I(p,n))$$

$$E(A) = \frac{2+0}{2+2}(0) + \frac{0+2}{2+2}(0)$$

$$E(A) = 0$$

Gain (Comp):

Gain (competition) = Information Gain-Entropy (competition)

Gain (competition) = I (p ,n)-Entropy (competition)

Gain (competition) = 1-0

Gain (competition) = 1

Entropy (type):

Information gain of each categorical data

type	p_i	n_i	$I(p_i, n_i)$
software	1	1	$I(1,1) = 1$
hardware	1	1	$I(1,1) = 1$

Entropy of type:

$$E(A) = \text{Entropy } E(A) = \sum_{i=2}^{v} \frac{p_i + n_i}{p+n} (I(p,n))$$

$$E(A) = \frac{1+1}{2+2}(1) + \frac{1+1}{2+2}(1)$$

$$E(A) = 1$$

Gain(type):

Gain (type) = Information Gain-Entropy (type)

Gain (type) = I (p,n) - Entropy (type)

Gain (type) = 1-1

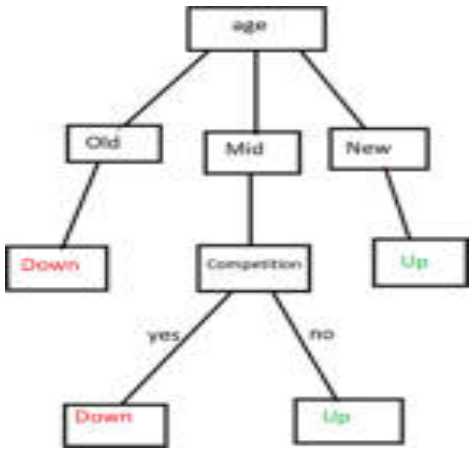
Gain (type) = 0

Gain(competition, type):

Gain of competition = 1

Gain of type = 0

Check the Maximum Gain of all the attributes :


age
Old
Mid
New
Down
competition
Up
yes
no
Down
Up

CHAPTER 6
SYSTEM DESIGN

6.SYSTEM DESIGN

6.1 MODULES

1. DATACOLLECTION
2. DATAPRE-PROCESSING
3. FEATUREEXTRACTION
4. EVALUATION MODEL DATACOLLECTION

1. DATACOLLECTION

Data used in this paper is a set of values from different features. This step is concerned with selecting the subset of all available data that you will be working with. ML problems start with data preferably, lots of data (examples or observations) for which you already know the target answer. Data for which you already know the target answer is called labeleddata.

2. DATA PRE-PROCESSING

Organize your selected data by formatting, cleaning, and sampling from it. Three common data pre processing steps are:

Formatting: The data you have selected may not be in a format that is suitable for you to work with. The data may be in a relational database and you would like it in a flat file, or the data may be in a proprietary file format and you would like it in a relational database or a textfile.

Cleaning: Cleaning data is the removal of fixing of missing data. There may be data instances that are incomplete and do not carry the data you believe you need to address the problem.

These instances may need to be removed. Additionally, there may be sensitive information in some of the attributes and those attributes may need to be anonymized or removed from the data entirely.

Sampling: There may be far more selected data available than you need to work with. More data can be result in much longer running times for algorithms and larger computational and memory requirements. You can take a smaller representative sample of the selected data that may be much faster for exploring and prototyping solutions before considering the whole dataset.

3. FEATUREEXTRACTION

Next thing is to do Feature Extraction is an attribute reduction process. Unlike feature selection, which ranks the existing attributes according to their predictive significance, feature extraction actually transforms the attributes. The transformed attributes or features are linear combinations of the original attributes. Finally our models are trained using Classifier. Some machine learning algorithms were used to classify pre-processed data. The classifiers were Random Forest.

4. EVALUATION MODEL

Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future.

Evaluating model performance with the data used for training is not acceptable in data science because it can easily generate overoptimistic and over fitted models. There are two methods of evaluating models in data, Hold-Out and Cross-Validation. To avoid over fitting, both methods use a test set to evaluate model performance.

Accuracy is defined as the percentage of correct predictions for the test data. It can be calculated easily by dividing the number of correct predictions by the number of total predictions.

- PythonGUI
- Python built-inmodules
    - Numpy
    - Pandas
    - Scipy
    - Sklearn

6.2 PYTHON

Python is a dynamic, high level, free open source and interpreted programming language. It supports object-oriented programming as well as procedural oriented programming.

In Python, we don't need to declare the type of variable because it is a dynamic typed language.

For example, x=10

here x can be anything such as String, int etc

6.2.1 FEATURES OF PYTHON

Easy to code:

Python is high level programming language. Python is very easy to learn language as compared to other language like c, c#, java script, java etc. It is very easy to code in python language and anybody can learn python basic in few hours or days. It is also developer-friendly language.

Free and Open Source:

Python language is freely available at official website and you can download it.

Since, it is open-source, this means that source code is also available to the public. So you can download it as, use it as well as share it.

Object-Oriented Language:

One of the key features of python is Object-Oriented programming. Python supports object oriented language and concepts of classes, objects encapsulation etc.

GUI Programming Support:

Graphical Users interfaces can be made using a module such as PyQt5, PyQt4, wxPython or Tk in python. PyQt5 is the most popular option for creating graphical apps withPython.

High-Level Language:

Python is a high-level language. When we write programs in python, we do not need to remember the system architecture, nor do we need to manage the memory.

Extensible feature:

Python is a **Extensible** language. We can write our some python code into c or c++ language and also we can compile that code in c/c++language.

Python is Portable language:

Python language is also a portable language. For example, if we have python code for windows and if we want to run this code on other platform such as Linux, Unix and Mac then we do not need to change it, we can run this code on any platform.

Python is Integrated language:

Python is also an Integrated language because we can easily integrated python with other language like c, c++etc.

InterpretedLanguage:

Python is an Interpreted Language because python code is executed line by line at a time.like other language c, c++, java etc there is no need to compile python code this makes it easier to debug our code. The source code of python is converted into an immediateform called bytecode.

Large Standard Library

Python has a large standard library which provides rich set of module and functions so you do not have to write your own code for every single thing. There are many libraries present in python for such as regular expressions, unit-testing, web browsers etc.

Dynamically Typed Language:

Python is dynamically-typed language. That means the type (for example- int, double, long etc) for a variable is decided at run time not in advance. because of this feature we don't need to specify the type of variable.

6.2.2 Python Libraries

SciPy

Scipy is a very popular library among Machine Learning enthusiasts as it contains different modules for optimization, linear algebra, integration and statistics. There is a difference between the SciPy library and the SciPy stack. The SciPy is one of the core packages that make up the SciPy stack. SciPy is also very useful for image manipulation.

Numpy

NumPy is a very popular python library for large multi-dimensional array and matrix processing, with the help of a large collection of high-level mathematical functions. It is very useful for fundamental scientific computations in Machine Learning. It is particularly useful for linear algebra, Fourier transform, and random number capabilities. High-end libraries like TensorFlow uses NumPy internally for manipulation of Tensors.

Skikit-learn

Skikit-learn is one of the most popular ML libraries for classical ML algorithms. It is built on top of two basic Python libraries, viz., NumPy and SciPy. Scikit-learn supports most of the supervised and unsupervised learning algorithms. Scikit-learn can also be used for data-mining and data-analysis, which makes it a great tool who is starting out withML.

Pandas

Pandas is a popular Python library for data analysis. It is not directly related to Machine Learning. As we know that the dataset must be prepared before training. In this case, Pandas

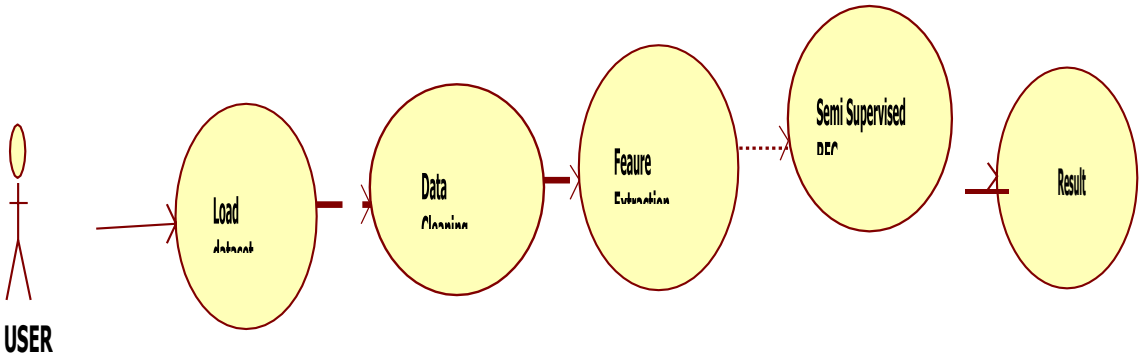
comes handy as it was developed specifically for data extraction and preparation. It provides high-level data structures and wide variety tools for data analysis. It provides many inbuilt methods for groping, combining and filtering data.

6.3 UMLDIAGRAMS

UML stands for **U**nified **M**odeling **L**anguage. Its a rich language to model software solutions, application structures, system behavior and business processes.

USE-CASE DIAGRAM:

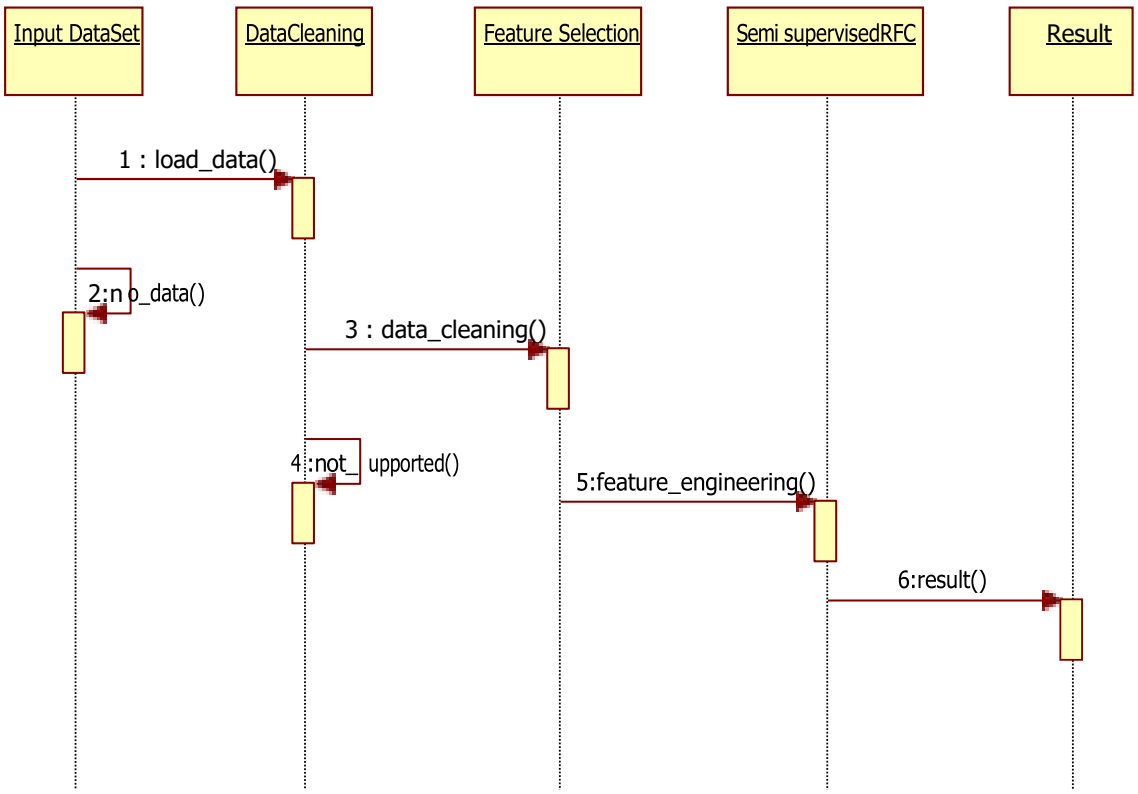
Use case diagrams are a set of use cases, actors, and their relationships. They represent the use case view of a system. A use case represents a particular functionality of a system. Hence, use case diagram is used to describe the relationships among the functionalities and their internal/external controllers. These controllers are known as actors.



SEQUENCE DIAGRAM

A sequence diagram is an interaction diagram. From the name, it is clear that the diagram deals with some sequences, which are the sequence of messages flowing from one object to another.

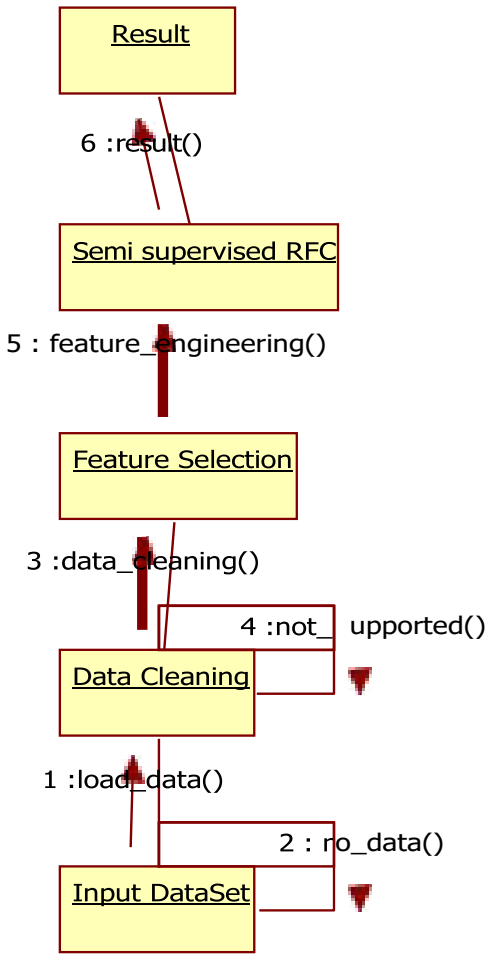
Interaction among the components of a system is very important from implementation and execution perspective. Sequence diagram is used to visualize the sequence of calls in a system to perform a specific functionality.



COLLABORATION DIAGRAM

Collaboration diagram is another form of interaction diagram. It represents the structural organization of a system and the messages sent/received. Structural organization consists of objects and links.

The purpose of collaboration diagram is similar to sequence diagram. However, the specific purpose of collaboration diagram is to visualize the organization of objects and their interaction.

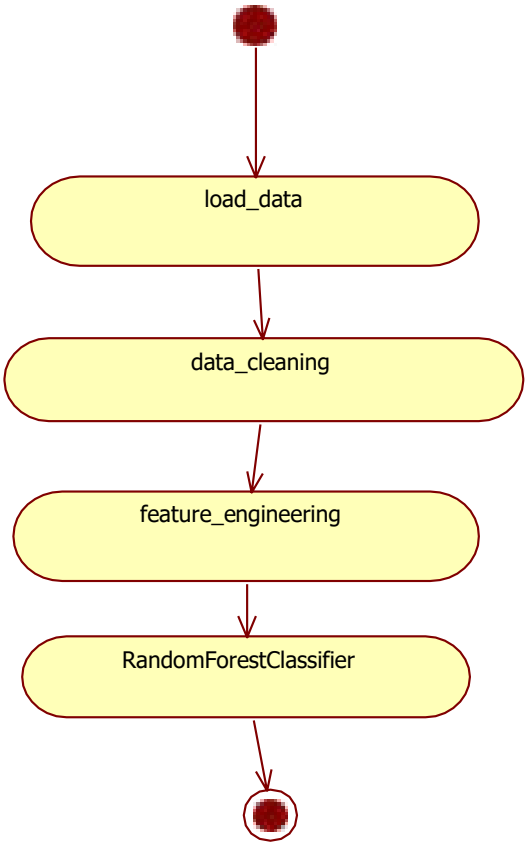


ACTIVITY DIAGRAM

Activity diagram describes the flow of control in a system. It consists of activities and links. The flow can be sequential, concurrent, or branched.

Activities are nothing but the functions of a system. Numbers of activity diagrams are prepared to capture the entire flow in a system.

Activity diagrams are used to visualize the flow of controls in a system. This is prepared to have an idea of how the system will work when executed.



CHAPTER 7
SYSTEM TESTING
&
IMPLEMENTATION

7.SYSTEM TESTING AND IMPLEMENTATION

7.1 INTRODUCTION TO TESTING

Software Testing is critical element of software quality assurance and represents the ultimate review of specification, design and coding. The increasing visibility of software as a system element and attendant costs associated with a software failure are motivating factors for we planned, through testing. Testing is the process of executing a program with the intent of finding an error. The design of tests for software and other engineered products can be as challenging as the initial design of the product itself.

There are basically two types of testing approaches.

Black-Box testing – the specified function that a product has been designed to perform, tests can be conducted that demonstrate each function is fully operated.

White-Box Testing – knowing the internal workings of the product, tests can be conducted to ensure that the internal operation of the product performs according to specifications and all internal components have been adequately exercised.

White Box and Black Box testing methods have been used to test his package. The entire loop constructs have been tested for their boundary and intermediate conditions. The test data was designed with a view to check for all the conditions and logical decisions. Error handling has been taken care of by the use of exceptionhandlers.

7.2TESTINGSTRATEGIES

Testing is a set of activities that can be planned in advanced and conducted systematically. A strategy for software testing must accommodation low-level tests that are necessary to verify that a small source code segment has been correctly implemented as well as high-level tests that validate major system functions against customerrequirements.

Software testing is one element of verification and validation. Verification refers to the set of activities that ensure that software correctly implements as specific function.

Validation refers to a different set of activities that ensure that the software that has been built is traceable to customer requirements.

The main objective of software is testing to uncover errors. To fulfill this objective, a series of test steps unit, integration, validation and system tests are planned and executed. Each test step is accomplished through a series of systematic test technique that assist in the design of test cases.

With such testing step, the level of abstraction with which software is considered is broadened.

Testing is the only way to ensure the quality of software and it is an umbrella activity rather than a separate phase. This is an activity to be performed in parallel with the software effort and one that consists of its own phases of analysis, design, implementation, execution and maintenance.

7.3 UNIT TESTING

Unit testing focuses verification effort on the smallest unit of Software design that is the module. Unit testing exercises specific paths in a module's control structure to ensure complete coverage and maximum error detection. This test focuses on each module individually, ensuring that it functions properly as a unit. Hence, the naming is Unit Testing.

During this testing, each module is tested individually and the module interfaces are verified for the consistency with design specification. All important processing path are tested for the expected results. All error handling paths are also tested.

7.4 INTEGRATION TESTING

Integration testing addresses the issues associated with the dual problems of verification and program construction. After the software has been integrated a set of high order tests are conducted. The main objective in this testing process is to take unit tested modules and builds a program structure that has been dictated by design.

The following are the types of Integration Testing:

1. **TOP DOWN INTEGRATION**

This method is an incremental approach to the construction of program structure. Modules are integrated by moving downward through the control hierarchy, beginning with the main program module. The module subordinates to the main program module are incorporated into the structure in either a depth first or breadth first manner.

In this method, the software is tested from main module and individual stubs are replaced when the test proceeds downwards.

2. **BOTTOM-UPINTEGRATION**

This method begins the construction and testing with the modules at the lowest level in the program

structure. Since the modules are integrated from the bottom up, processing required for modules subordinate to a given level is always available and the need for stubs is eliminated. The bottom up integration strategy may be implemented with the followingsteps:

The low-level modules are combined into clusters into clusters that perform a specific Softwaresub-function.

- A driver (i.e.) the control program for testing is written to coordinate test case input andoutput.
- The cluster istested.
- Drivers are removed and clusters are combined moving upward in the program structure

The bottom up approaches tests each module individually and then each module is module is integrated with a main module and tested for functionality.

7.5USERACCEPTANCETESTING

User Acceptance of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changes wherever required. The system developed provides a friendly user interface that can easily be understood even by a person who is new to the system.

7.6OUTPUTTESTING

After performing the validation testing, the next step is output testing of the proposed system, since no system could be useful if it does not produce the required output in the specified format. Asking the users about the format required by them tests the outputs generated or displayed by the system under consideration. Hence the output format is considered in 2 ways – one is on screen and another in printed format.

7.7IMPLEMENTATION

Implementation is the process of converting a new or revised system design operational one. There are three types of Implementation.

- Implementation of a computer system to replace a manual system. The problems encountered are converting files, training users and verifying printouts forintegrity.
- Implementation of a new computer system to replace an existing one. This is usually a

difficult conversion. If not properly planned there can be manyproblems.

7.8SOURCE CODE

```
import warnings
warnings.filterwarnings("ignore")
import os
import numpy as np
import pandas as pd
from sklearn import metrics

from sklearn.metrics import precision_ score from
sklearn.metrics import recall_score
from sklearn. metrics import f1_score

from sklearn.model_ selection import cross_validate

from sklearn.feature_extraction.text import CountVectorizer from
sklearn.naive_bayes import Multinomial NB
from sklearn.model_ selection import train_test_split

neg_deceptive_folder_path = 'negative_polarity\\deceptive_from_ MTurk\\' neg_true_
folder_path = 'negative_polarity\\truthful_from_Web\\'
pos_deceptive_ folder_path = 'positive_ polarity\\deceptive_ from_ MTurk\\' pos_true_
folder_path = 'positive_ polarit y\\truthful_ from_ Trip Advisor\\' polarity_ class = []
reviews = []
spamity_ class =[] for
i in range(1, 6):
    insideptru = pos_true_ folder_ path + 'fold' + str(i)
    insidepdec = pos_ deceptive_ folder_path + 'fold' + str(i)
  insidentru = neg_true_ folder_ path + 'fold' + str(i)
    insidendec = neg_ deceptive_ folder_path + 'fold' + str(i) pos_list
```

```
= []
for data_file in sorted( os. listdir( insidendec)):
    polarity_ class. append('negative')
    spamity_ class. append( str( data_file. split('_')[ 0])) with
    open(os. path.join(insidendec, data_ file)) as f: contents= f.
    read()
    reviews. append(contents)
    for data_file in sorted( os. listdir( insidentru)): polarity_
    class. append('negative')
    spamity_ class. append(str( data_file.split('_')[ 0])) with
    open(os. path.join(insidentru, data_ file)) as f:contents =
    f. read()
    reviews. append(contents)
    for data_file in sorted( os. listdir( insidepdec)): polarity_
    class. append('positive')
    spamity_ class. append( str( data_file. split('_')[ 0])) with
    open(os. path.join(insidepdec, data_ file)) as f: contents= f.
    read()
    reviews. append(contents)
    for data_file in sorted( os. listdir( insideptru)): polarity_
    class. append('positive')
    spamity_ class. append(str( data_file.split('_')[ 0]))
  with open(os. path. join( insideptru, data_ file)) as
  f:contents = f. read()
```

```
                reviews. append(contents)

               data_fm          =pd. DataFrame({'polarity_
class':polarity_

             class,'review':reviews,'spamity_ class':s pamity_ class})

              data_ fm. loc[ data_ fm['spamity_ class']=='d','spamity_ class']=0

            data_ fm. loc[ data_ fm['spamity_ class']=='t','spamity_ class']=1

            data_x = data_ fm['review']

            data_ y = np. asarray( data_ fm['spamity_ class'],dtype=int)

            X_train, X_ test, y_ train, y_ test = train_ test_split( data_x, data_ y,

            test_size=0. 3) cv = CountVectorizer()

            X_traincv=cv.fit_transform(_ltr

            ain)

            X_testcvcv.transform(X_test)

            nbayes =MultinomialNB()

            nbayes. fit(X_ traincv,y_train)

            y_ predictions = nbayes. predict(

            X_testcv) y_ result = l ist(

            y_predictions)

            yp=["True" if a==1 else "Deceptive" for a in y_ result] X_testlist= li st( X_test)

               output_ fm=pd. DataFrame({'Review':X_ testlist,'True( 1)/Deceptive(0)': yp})

             output_ fm
               print("Accuracy %:",metrics.accuracy_ score( y_ test, y_predictions)*100)

          print("PrecisionScore:",precision_score(y_test,y_predictions,average='micro'))

        print("Recall Score: ", recall_score( y_ test, y_ predictions, average='micro') )

       print("F1 Score: ", f1 _score( y_test, y_ predictions, average='micro') )
```

```
################RFC##############
    import pandas as pd
    import numpy as np
    import matplotlib.pyplot as plt
    import seaborn as sns
    import nltk
    from nltk. corpus import stopwords
    import string
    import math
    from sklearn. feature_extraction.text import Count Vectorizer
    from sklearn. model_ selection import train_test_split, cross_ val_score from
    sklearn. metrics import classification_report
    from sklearn.metrics import confusion_matrix, accuracy_score,
    roc_ auc_score,roc_curve
    #from sklearn. grid_search import GridSearchCV
    from sklearn.model_ selection import GridSearchCV
    # LOADING THE DATASET AND SEEING THE DETAILS
    data = pd. read_ csv('yelp. csv') #
    SHAPE OF THE DATASET
    print("Shape of the dataset:")
    print(data. shape)
    # COLUMN NAMES
    print("Column names:")
     print( data.columns)
    # DATATYPE OF EACH COLUMN
    print("Datatype of each column:")
    print(data. dtypes)
```

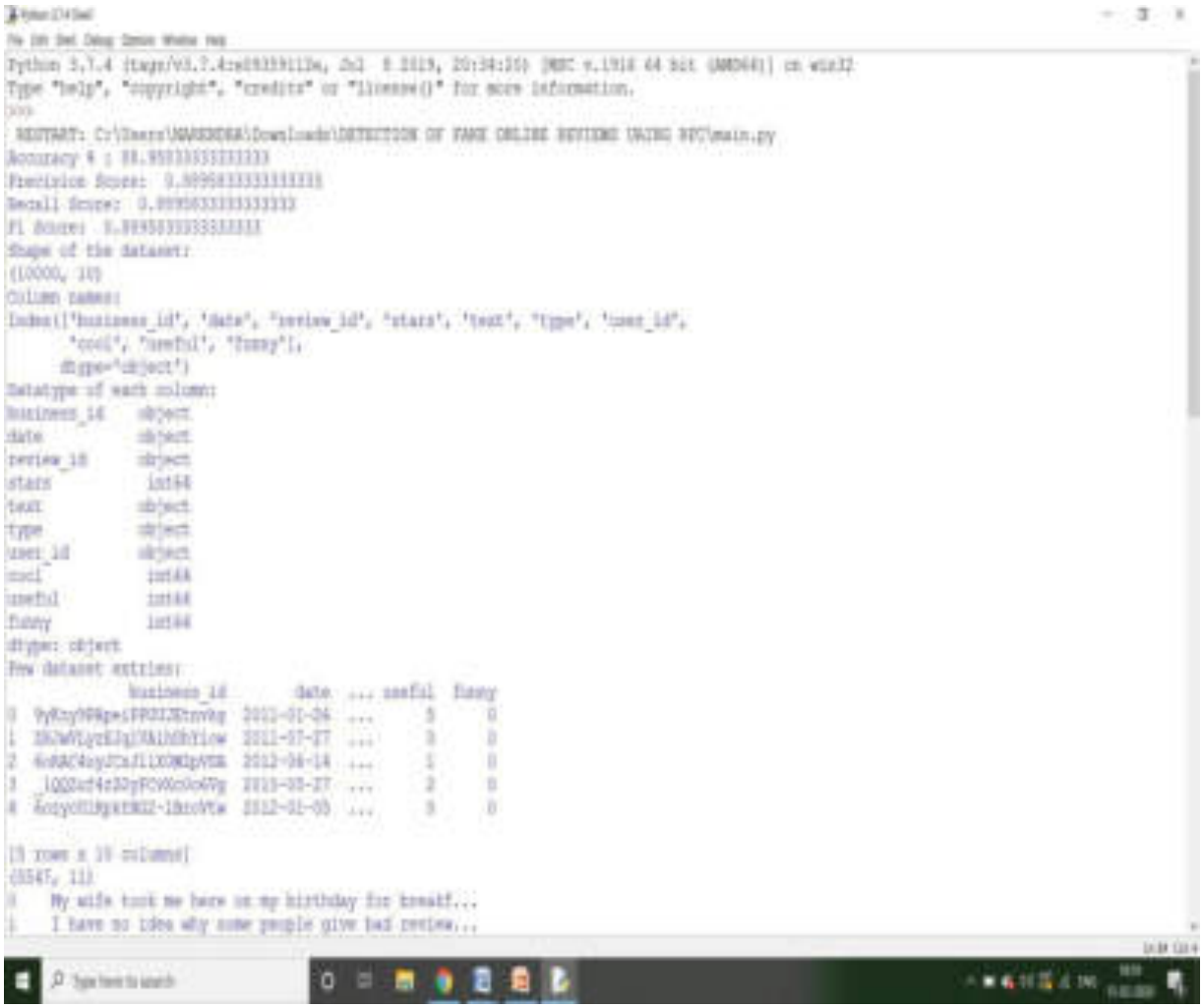
```
# SEEING FEW OF THE ENTRIES
print("Few dataset entries:")
print(data. head())
# DATASET SUMMARY
data. describe( include='all')
data['length'] = data['text']. apply( len) data.
head()
graph = sns. Facet Grid(data=data, col='stars')
graph.map( plt. hist,'length',bins=50, color='blue')
stval = data. groupby('stars').mean()
stval
stval. corr()
data_ classes = data[( data['stars']==1) |(data['stars']==3) |( data['stars']==5)]
data_ classes.head()
print( data_ classes.shape)
# Seperate the dataset into X and Y  for prediction
x = data_classes['text']
y = data_ classes['stars']
print(x.head())
print( y. head())
def text_process(text):
    nopunc = [ char for char in text if char not in string.punctuation]
     nopunc = ''.join(nopunc)
return[word for word in    nopunc.split()
if word.lower() not  in stopwords.words('english')]
```

```
vocab = Count Vectorizer( analyzer=text_process). fit( x)
print(len( vocab.vocabulary_))
r0 = x[ 0]
print( r0 )
vocab0 = vocab.transform([ r0])
print( vocab0 )
print("Getting the words back:")
print( vocab. get_feature_names()[ 19648]) print(
vocab.  get_feature_names()[  10643])  x  =
vocab.transform( x)
#Shape of the matrix:
print("Shape of the sparse matrix: ", x.shape) #Non-
zero occurences:
print("Non- Zero occurences: ",x.nnz) #
DENSITY OF THE MATRIX
density=(x.nnz/(x.shape[0]*x.shape[1]))*100
print("Density of the matrix = ",density)
 x_train,x_test,y_train,y_test                                    =
train_test_split( x, y, test_ size=0. 2, random_state=101)
from sklearn. ensemble import Random Forest Classifier
rmfr = Random ForestClassifier()
rmfr. fit( x_train, y_ train)
predrmfr = rmfr.predict(x_test)
print("Confusion Matrix for Random Forest Classifier:")
print( confusion_matrix( y_test, predrmfr))
```

```
print("Score:", round(accuracy_ score( y_ test,predrmfr)*100,2))
print("Classification Report:", classific ation_ report( y_ test,predrmfr))
############################################################
import pickle
var = input("Please enter the review text you want to verify: ")
print("You entered: " + str( var))
def detecting_ fake_review( var):
#retrieving the best model for prediction call
load_model = pickle.load(open('model.sav', 'rb'))
prediction = load_ model.predict([ var])
  prob = load_model. predict_proba([var])
return ( print("The given statement is ", prediction[ 0]),
 print("The truth probability score is ",prob[ 0][ 1]))
ifname== 'main':
    detecting_ fake_review(var)
```

CHAPTER 9

RESULTS


Python 3.7.4 (tags/v3.7.4:e09359112e, Jul 8 2019, 20:34:20) [MSC v.1916 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
RESTART: C:\Users\NAVEENDRA\Downloads\DETECTION OF FAKE ONLINE REVIEWS USING RFC\main.py
Accuracy 4 : 88.95333333333333
Precision Score: 0.8956333333333333
Recall Score: 0.8956333333333333
F1 Score: 0.8955333333333333
Shape of the dataset:
(10000, 10)
Column names:
Index(['business_id', 'date', 'review_id', 'stars', 'text', 'type', 'user_id',
'cool', 'useful', 'funny'],
dtype='object')
Datatype of each column:
business_id object
date object
review_id object
stars int64
text object
type object
user_id object
cool int64
useful int64
funny int64
dtype: object
Few dataset entries:
business_id date ... useful funny
0 9yKzy9PApeiPPOUJEtnvkg 2011-01-26 ... 5 0
1 ZRJwVLyzEJq1VAihDhYiow 2011-07-27 ... 3 0
2 6oRAC4uyJCsJl1X0WZpVSA 2012-06-14 ... 1 0
3 _1QQZuf4zZOyFCvXc0o6Vg 2010-05-27 ... 2 0
4 6ozycU1RpktNG2-1BroVtw 2012-01-05 ... 3 0
[5 rows x 10 columns]
(5547, 11)
0 My wife took me here on my birthday for breakf...
1 I have no idea why some people give bad review...

3 Ros Oakota, and £0W Chaparral Ooq Part! .
4 Scott is a ! !.
6 Drop that you're and After
Jlaaext,dtype:
0 S
1 S
3 S
4 S
6 S
Jlaastars, nth4

Hy oe on 4y birthday for breakfast and was . The weatherwas perfect trade
g an pleasure Our tress as and our on the . It
looked like the f up pretty kly the get here the better.

moyourselIa and Bloody sas and the best I ' ve ever had. I'at sure they use in
gredients Iron and thea freshwhenyou it was aaz .

MITBIflC the menu looks , bad the white truff f and sas tasty and
lcatae 2 of their bread as and trade the It sas the best •toast• I ' v
e ever had.

can' to hack !

the words

absence
Shape of the sparse : 5547, 3133 6}
: S7
of the 0.
usion flat rim forest Class

[3102

Score 92
f : —scor support


Anyway, I can't wait to go back!
Getting the words back:
Shape of the sparse matrix:
Density of the matrix =
Confusion Matrix for Random Forest Classifier:
Classification Report: precision recall f1-score support
accuracy
macro avg
weighted avg
Please enter the review text you want to verify: My wife and I stayed here in the middle of February 2005. We took the (EL) subway to the hotel from O'Hare. It was very convenient. The lobby was nice and the room was clean and comfortable. I've stayed in many a hotel room, this bed was the most comfortable I've ever slept in. The hotel is located in a great area and close to everything. Walking distance to almost all the interesting sites Chicago has. Would definetly recommend this hotel to anyone visiting this great city.
You entered: My wife and I stayed here in the middle of February 2005. We took the (EL) subway to the hotel from O'Hare. It was very convenient. The lobby was nice and the room was clean and comfortable. I've stayed in many a hotel room, this bed was the most comfortable I've ever slept in. The hotel is located in a great area and close to everything. Walking distance to almost all the interesting sites Chicago has. Would definetly recommend this hotel to anyone visiting this great city.
The given statement is True
The truth probability score is

CHAPTER 9
CONCLUSION

9.CONCLUSION

We have shown supervised text mining technique for detecting fake online reviews. We have combined features from several research works to create a better feature set. Thus, we have been able to increase the accuracy of previous semi- supervised techniques done by Jiten et al. [8].
This ensures that our dataset is labeled well.

OUTPUT PARAMETERS

- Confusion Matrix: It is a Error matrix which is used to visualize the performance ofa classifier.
- Accuracy=(TP+TN)/AllSamples
- Precision=TP/(TP+FP)
- Recall=TP/(TP+FN)
- F1Score=2*(Recall*precision)/(Recall/precision)

Where TP is TruePositives

TN is True Negatives
FP is False Positives
FN is False Negatives

REFERENCES

- https://docs.python.org/3/

- https://stackoverflow.com/questions/tagged/python

- https://www.jpinfotech.org/detection-of-fake-online-reviews-using-semi-supervised-and-supervised-learning/

- https://en.wikipedia.org/wiki/Semi-supervised_learning

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P. India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e mail: derhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited B.Tech (CSE|ECE|EEE|ME)
Accredited by NAAC With Grade A

DEPARTMENT OF INFORMATION TECHNOLOGY

A.Y:2020-21

List of Major Projects

IV B Tech IT VIII Semester Project Batch Wise				
S.No	**Regd No.**	**Name**	**Project Title**	**Guide Name**
A1	17H71A1202	Akhila M	Road Safety Prediction Using Machine Learning Algorithm	Ms. R. Srilakshmi
	17H71A1253	Sri Sahithi M		
	17H71A1213	Gagana Sindhu Ch		
	17H71A1222	Karthik Kumar P		
	17H71A1256	Teja Sri D		
A2	17H71A1235	Navya D	Sign Language Recognition Using Machine Learning	Ms. R. Srilakshmi
	17H71A1223	Kiranmai P		
	17H71A1246	Renuka Andimani		
	17H71A1237	Nikhita Ch		
	17H71A1231	Lakshmisrinivas K		
A3	17H71A1252	Sri Lakshmi P	Student Query Classification System	Mr. P. Narasimharao
	17H71A1204	Asma Khatun P		
	17H71A1207	Bhavana M		
	17H71A1259	Yaswanth Ch		
	17H71A1230	Lakshmi Narayana N		

A4	17H71A1244	Ramya K	An Advanced Android Based Trained Fruit And Vegetable Identification Using Machine Learning For Nutrition's	Mr. K. Mahanthi
	17H71A1206	Bhashitha K		
	17H71A1201	Abdul Rahaman		
	17H71A1245	Ramya P		
A5	17H71A1240	Pravallika S	Image Forgery Detection Using CNN	Mr. P. Narasimharao
	17H71A1221	Kamal Durga Prasad A		
	17H71A1260	Pedda Lakshmi B		
	17H71A1209	Chandana Y		
A6	17H71A1232	Malathi M	Privacy-Aware Personal Data Storage (P-PDS) Using Cloud Computing	Mr. P. Narasimharao
	17H71A1239	Padmaja P		
	17H71A1248	Sai Sri Venkata Sravani D		
	17H71A1212	Dinesh Kumar K		
A7	17H71A1251	Sri Lakshmi J	Spam Detection In Social Media	Ms. T. Bhargavi
	17H71A1233	Mani Harshitha K		
	17H71A1218	Harsha K		
	17H71A1249	Sandeep Iddipilla		
A8	17H71A1227	Lakshmi Lavanya J	Currency Recognisition System Using Image Processing	Ms. T. Bhargavi
	17H71A1220	Jyothi Priyanka M		
	17H71A1205	Bhargavi B		
	17H71A1247	Sai Kiran K		
A9	17H71A1250	Shazma Pathan	Online Product Rating Using Expressions	Ms. T. Bhargavi
	17H71A1208	Chandana V		
	17H71A1238	Omkar Pavan M		
	17H71A1210	Devi Prasanna K		

A10	17H71A1228	Lakshmi Manoja G	Cloud Based Content Security System	Ms. B. V. N. Praveena
	17H71A1229	Lakshmi N		
	17H71A1214	Gayatri K		
	17H71A1215	Geyasri Vagdevi Methuku		
A11	17H71A1243	Raafiya Begum Mohammed	Credit Card Fraud Detection Using Machine Learning Techniques	Ms. B. V. N. Praveena
	17H71A1225	Laila M		
	17H71A1211	Dimple Preethi V		
	17H71A1216	Girija Sankar N		
A12	17H71A1241	Priyanka J	An Advanced Keyword Attacks Over Encrypted Data In Cloud	Mr. K. Mahanthi
	17H71A1217	Guna Goutham V		
	17H71A1219	Juhitha G		
	17H71A1224	Kishore B		
A13	17H71A1257	Venkata Saida Lakshmi P	Employee Payroll System	Ms. B. V. N. Praveena
	17H71A1254	Sri Sai Sneha P		
	17H71A1255	Supraja M		
	17H71A1203	Anusha K		

Signature of Coordinator

(Dr. G.Sai Chaithanya Kumar)

Signature of HoD

(Mr.D.Prasad)

Road Safety Prediction Using Machine Learning Algorithm

A Project Report
Submitted in partial fulfillment of the
requirements for the award of the degree of

Bachelor of Technology

In

Information Technology

By

Akhila M (17H71A1202) **Sri Sahithi M (17H71A1253)**

Gagana Sindhu Ch (17H71A1213) **Karthik Kumar P (17H71A1222)**

Teja Sri D (17H71A1256)

Under the esteemed guidance of

Ms R Srilakshmi

Assistant Professor

Department of Information Technology

Devineni Venkata Ramana & Dr.Hima Sekhar

MIC College of Technology

(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

April, 2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273335, 273623, Fax : 08678 - 273569
e-mail: drvhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE,ECE,EEE,ME)
NAAC
With Grade A

CERTIFICATE

This is to certify that the project report entitled **"Road Safety Prediction Using Machine Learning Algorithm"** submitted by **Akhila M (17H71A1202), Sri Sahithi M (17H71A1253), Gagana Sindhu Ch (17H71A1213), Karthik Kumar P (17H71A1222), Teja Sri D (17H71A1256)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in **Computer Science & Engineering** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Mr. R.SREE LAKSHMI)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K.SRINIVAS)

Dr. K. SRINIVAS, M.Tech., Ph.D (JNTUH)
Professor & Principal
Dept. of Computer Science & Engg.
DVR & Dr. H.S. MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India - Pin: 521180.

Internal Examiner
(Dr. G.SAI CHAITANYA KUMAR)

External Examiner

Sign Language Recognition Using Machine Learning

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Bachelor of Technology

In

Information Technology

By

Navya D (17H71A1235) **Kiranmai P (17H71A1223)**

Renuka Andimani (17H71A1246) **Nikhita Ch (17H71A1237)**

Lakshmisrinivas K (17H71A1231)

Under the esteemed guidance of

Ms R Srilakshmi

Assistant Professor

Department of Information Technology

Devineni Venkata Ramana & Dr.Hima Sekhar

MIC College of Technology

(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

April, 2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS

NBA-Accredited B.Tech (CSE|ECE|EEE|ME)

Accredited by NAAC With Grade A

CERTIFICATE

This is to certify that the project report entitled **"Sign Language Recognition Using Machine Learning"** submitted by **Navya D (17H71A1235), Kiranmai P (17H71A1223), Renuka Andimani (17H71A1246), Nikhita Ch (17H71A1237)**, **Lakshmisrinivas K (17H71A1231)**to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in **Information Technology** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Ms. R.SREE LAKSHMI)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K.SRINIVAS)

Dr. K. SRINIVAS, M.Tech, Ph.D (JNTUH)
Professor, [illegible]
DVR & Dr.HS MIC College of Technology
Kanchikacherla, Krishna Dt,
Andhra Pradesh, India – Pin: 521180

Internal Examiner
(Dr. G.SAI CHAITANYA KUMAR)

External Examiner

Student Query Classification System

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Bachelor of Technology

In

Information Technology

By

Sri Lakshmi P (17H71A1252) **Asma Khatun P (17H71A1204)**
Bhavana M (17H71A1207) **Yaswanth Ch (17H71A1259)**
Lakshmi Narayana N (17H71A1230)

Under the esteemed guidance of
Mr P Narasimharao
Assistant Professor

Department of Information Technology

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

April, 2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273533, 273623, Fax: 08678 - 273569
e-mail: drhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|EEE|ME)
Accredited by
NAAC A
With Grade

CERTIFICATE

This is to certify that the project report entitled "**Student Query Classification System**" submitted by **Sri Lakshmi P (17H71A1252), Asma Khatun P (17H71A1204), Bhavana M (17H71A1207), Yaswanth Ch (17H71A1259), Lakshmi Narayana N (17H71A1230)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in **Information Technology** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Mr. P.NARASIMHA RAO)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K.SRINIVAS)

Dr. K. Srinivas, M.Tech, Ph.D, MISTE
Professor, Dept. of Electrical Engg.
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India - Pin: 521180

Internal Examiner
(Dr. G.SAI CHAITANYA KUMAR)

External Examiner

An Advanced Android Based Trained Fruit And Vegetable Identification Using Machine Learning For Nutrition's

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Bachelor of Technology

In

Information Technology

By

Ramya K (17H71A1244)
Bhashitha K (17H71A1206)
Abdul Rahaman (17H71A1201)
Ramya P (17H71A1245)

Under the esteemed guidance of
Mr K Mahanti
Assistant Professor

Department of Information Technology

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

April, 2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.

AUTONOMOUS
NBA-Accredited
NAAC A With Grade

CERTIFICATE

This is to certify that the project report entitled "**An Advanced Android Based Trained Fruit And Vegetable Identification Using Machine Learning For Nutrition's**" submitted by **Ramya K (17H71A1244), Bhashitha K (17H71A1206), Abdul Rahaman (17H71A1201), Ramya P (17H71A1245)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in **Information Technology** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Mr. K.MAHANTHI)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K.SRINIVAS)

Internal Examiner
(Dr. G.SAI CHAITANYA KUMAR)

External Examiner

Image Forgery Detection Using CNN

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Bachelor of Technology

In

Information Technology

By

Pravallika S (17H71A1240) **Kamal Durga Prasad A (17H71A1221)**
Pedda Lakshmi B (17H71A1260) **Chandana Y (17H71A1209)**

Under the esteemed guidance of
Mr P Narasimharao
Assistant Professor

Department of Information Technology

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

April, 2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e mail: drvhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE)ECE)ME)
Accredited by
NAAC With Grade A

CERTIFICATE

This is to certify that the project report entitled "**Image Forgery Detection Using CNN**" submitted by **Pravallika S (17H71A1240), Kamal Durga Prasad A (17H71A1221), Pedda Lakshmi B (17H71A1260), Chandana Y (17H71A1209)**to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in **Information Technology** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Mr. P.NARASIMHA RAO)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K.SRINIVAS)

Dr. K. S... M.Tech, Ph.D (JNTUH)
Professor ... Electronics & Comm. Engg.
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India – Pin: 521180

Internal Examiner
(Dr. G.SAI CHAITANYA KUMAR)

External Examiner

Privacy-Aware Personal Data Storage (P-PDS) Using Cloud Computing

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Bachelor of Technology

In

Information Technology

By

Malathi M (17H71A1232) **Padmaja P (17H71A1239)**

Sai Sri Venkata Sravani D (17H71A1248) **Dinesh Kumar K (17H71A1212)**

Under the esteemed guidance of

Mr P Narasimharao

Assistant Professor

Department of Information Technology

Devineni Venkata Ramana & Dr.Hima Sekhar

MIC College of Technology

(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

April, 2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone : 08678 - 273535, 273621, Fax: 08678 - 273569
e-mail: dvrhima@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|EEE|ME)
Accredited by NAAC With Grade A

CERTIFICATE

This is to certify that the project report entitled **"Privacy-Aware Personal Data Storage (P-PDS) Using Cloud Computing"** submitted by **Malathi M (17H71A1232), Padmaja P (17H71A1239), Sai Sri Venkata Sravani D (17H71A1248), Dinesh Kumar K (17H71A1212)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in **Information Technology**is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Mr. P.NARASIMHA RAO)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K.SRINIVAS)

Dr. K. SRINIVAS, M.Tech, Ph.D (JNTUH)
Professor, Dept. of Electrical Engg.
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India – Pin: 521180

Internal Examiner
(Dr. G.SAI CHAITANYA KUMAR)

External Examiner

Spam Detection In Social Media

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Bachelor of Technology

In

Information Technology

By

Sri Lakshmi J (17H71A1251) **Mani Harshitha K (17H71A1233)**
Harsha K (17H71A1218) **Sandeep Iddipilla (17H71A1249)**

Under the esteemed guidance of

Ms T Bhargavi

Assistant Professor

Department of Information Technology

Devineni Venkata Ramana & Dr.Hima Sekhar

MIC College of Technology

(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

April, 2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|MC)
Accredited by
NAAC A
With Grade

CERTIFICATE

This is to certify that the project report entitled **"Spam Detection In Social Media"** submitted by **Sri Lakshmi J (17H71A1251), Mani Harshitha K (17H71A1233), Harsha K (17H71A1218), Sandeep Iddipilla (17H71A1249)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in **Information Technology** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Ms.T.BHARGAVI)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K.SRINIVAS)

Dr. K. Srinivas, M.Tech (JNTUH)
Professor, Dept. of Electrical Engg.
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India – Pin: 521180.

Internal Examiner
(Dr. G.SAI CHAITANYA KUMAR)

External Examiner

Currency Recognisition System Using Image Processing

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Bachelor of Technology

In

Information Technology

By

Lakshmi Lavanya J (17H71A1227)
Jyothi Priyanka M (17H71A1220)
Bhargavi B (17H71A1205)
Sai Kiran K (17H71A1247)

Under the esteemed guidance of

Ms T Bhargavi

Assistant Professor

Department of Information Technology

Devineni Venkata Ramana & Dr.Hima Sekhar

MIC College of Technology

(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

April, 2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|EEE|ME)
Accredited by
NAAC A
With Grade

CERTIFICATE

This is to certify that the project report entitled **"Currency Recognisition System Using Image Processing"** submitted by **Lakshmi Lavanya J (17H71A1227), Jyothi Priyanka M (17H71A1220), Bhargavi B (17H71A1205), Sai Kiran K (17H71A1247)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in **Information Technology** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Bhargavi

Supervisor
(Ms. T.BHARGAVI)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K.SRINIVAS)

DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India – Pin: 521180

Sai

Internal Examiner
(Dr. G.SAI CHAITANYA KUMAR)

External Examiner

Online Product Rating Using Expressions

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Bachelor of Technology

In

Information Technology

By

Shazma Pathan(17H71A1250) **Chandana V (17H71A1208)**
Omkar Pavan M (17H71A1238) **Devi Prasanna K (17H71A1210)**

Under the esteemed guidance of
Ms T Bhargavi
Assistant Professor

Department of Information Technology

Devineni Venkata Ramana & Dr.Hima Sekhar

MIC College of Technology

(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

April, 2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P. India
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech: CSE|ECE|EEE|ME|
NAAC A
With Grade

CERTIFICATE

This is to certify that the project report entitled **"Online Product Rating Using Expressions"** submitted by **Shazma Pathan(17H71A1250) Chandana V (17H71A1208),Omkar Pavan M (17H71A1238) ,Devi Prasanna K (17H71A1210)**to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in **Information Technology** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Bhargavi T

Supervisor
(Ms. T.BHARGAVI)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K.SRINIVAS)

Dr. K. SRINIVAS, M.Tech., Ph.D (JNTUH)
Professor, Dept. of Mechanical Engg.
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India - Pin: 521180

Internal Examiner
(Dr. G.SAI CHAITANYA KUMAR)

External Examiner

Cloud Based Content Security System

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Bachelor of Technology

In

Information Technology

By

Lakshmi Manoja G (17H71A1228) **Lakshmi N (17H71A1229)**

Gayatri K (17H71A1214) **Geyasri Vagdevi Methuku(17H71A1215)**

Under the esteemed guidance of

Ms B V N Praveena

Assistant Professor

Department of Information Technology

Devineni Venkata Ramana & Dr.Hima Sekhar

MIC College of Technology

(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

April, 2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone: 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE,ECE,ME)
Accredited by
NAAC A
With Grade

CERTIFICATE

This is to certify that the project report entitled "**Cloud Based Content Security System** " submitted by **T Lakshmi Manoja G (17H71A1228) , Lakshmi N (17H71A1229), Gayatri K (17H71A1214) , Geyasri Vagdevi Methuku (17H71A1215)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science & Engineering is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Ms. B.PRAVEENA)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K.SRINIVAS)

Dr. K. SRINIVAS, M.Tech., Ph.D (JNTUH)
Professor, Dept. of Mechanical Engg.
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India - Pin: 521180

Internal Examiner
(Dr. G.SAI CHAITANYA KUMAR)

External Examiner

Credit Card Fraud Detection Using Machine Learning Techniques

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Bachelor of Technology

In

Information Technology

By

Raafiya Begum Mohammed (17H71A1243) **Laila M (17H71A1225)**
Dimple Preethi V (17H71A1211) **Girija Sankar N(17H71A1216)**

Under the esteemed guidance of
Ms B V N Praveena
Assistant Professor

Department of Information Technology

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

April, 2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phone: 08678 - 273533, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE,ECE,IT,EEE,MEC)
Accredited by
NAAC A
With Grade

CERTIFICATE

*This is to certify that the project report entitled "*__Credit Card Fraud Detection Using Machine Learning Techniques__*" submitted by* **Raafiya Begum Mohammed (17H71A1243), Laila M (17H71A1225), Dimple Preethi V (17H71A1211), Girija Sankar N(17H71A1216)** *to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in* **Information Technology** *is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.*

Supervisor
(Ms.B.PRAVEENA)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K.SRINIVAS)

Dr. K. SRINIVAS, M.Tech., Ph.D (JNTUH)
Professor, Dept. of Electrical Engg.
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India - Pin: 5211**

Internal Examiner
(Dr. G.SAI CHAITANYA KUMAR)

External Examiner

DR. HS MIC COLLEGE OF TECHNOLOGY
Kanchikacherla
Pin : 521180
Ph.273535

An Advanced Keyword Attacks Over Encrypted Data In Cloud

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Bachelor of Technology

In

Information Technology

By

Priyanka J (17H71A1241) **Guna Goutham V (17H71A1217)**

Juhitha G (17H71A1219) **Kishore B (17H71A1224)**

Under the esteemed guidance of

Mr K Mahanti

Assistant Professor

Department of Information Technology

Devineni Venkata Ramana & Dr.Hima Sekhar

MIC College of Technology

(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

April, 2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|EEE|ME)
Accredited by
NAAC A
With Grade

CERTIFICATE

This is to certify that the project report entitled **"An Advanced Keyword Attacks Over Encrypted Data In Cloud"** submitted by **Priyanka J (17H71A1241) , Guna Goutham V (17H71A1217),Juhitha G (17H71A1219) ,Kishore B (17H71A1224)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in **Information Technology** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Mr. K.MAHANTHI)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K.SRINIVAS)

Dr. K. Srinivas, M.Tech, Ph.D (JNTUH)
Professor, Dept. of Electrical Engg.
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India - Pin: 521180

Internal Examiner
(Dr. G.SAI CHAITANYA KUMAR)

External Examiner

Employee Payroll System

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Bachelor of Technology

In

Information Technology

By

Venkata Saida Lakshmi P (17H71A1257) **Sri Sai Sneha P (17H71A1254)**
Supraja M (17H71A1255) **Anusha K (17H71A1203)**

Under the esteemed guidance of
Ms B V N Praveena
Assistant Professor

Department of Information Technology

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

April, 2021

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P. India
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited B.Tech (CSE|ECE|ME)
Accredited by NAAC With Grade A

CERTIFICATE

This is to certify that the project report entitled **"Employee Payroll System"** submitted by **Venkata Saida Lakshmi P (17H71A1257)** ,**Sri Sai Sneha P (17H71A1254), Supraja M (17H71A1255)** **Anusha K (17H71A1203)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in **Computer Science & Engineering** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Mr. K.MAHANTHI)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K.SRINIVAS)

Internal Examiner
(Dr. G.SAI CHAITANYA KUMAR)

External Examiner

DVR & Dr. HS
MIC College of Technology
ISO 9001:2015 Certified Institute
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India
Phones: 08678 - 253535 / 94914 57799 / 73826 16826
E-mail: office@mictech.ac.in, Website: www.mictech.edu.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|MEC)
Accredited by NAAC A With Grade

Department of Computer Applications

MCA MAJOR PROJECTS LIST

A.Y:2020-21

S.No.	Regd No.	Name	Project TITLE	Guide Name
1	19H75F0002	Asmat M	Detection of Online Reviews Using Semi Supervised And Supervised Learning	Mrs. A. Anuradha
2	19H75F0003	Bhavya Sri R	Designing Secure and Efficient biometric-based secure access mechanism for cloud service	Mrs. V Srilakshmi
3	19H75F0004	Gayatri Ch	Cartoonifying image using python	Mrs. R. Srilakshmi
4	19H75F0005	Gopi T	Facial Emotion Recognition using Convolutional Neural Network	Mr. K Mahanthi
5	19H75F0006	Gopi T	Heart Disease Identification Method Using Machine Learning Classification in E-Health care	Ms. K Vinaya Sree
6	19H75F0007	Gopi M	Detection and Classification of Fruit Diseases Using Image processing &Cloud Computing	Mr. J Venkata Krishna
7	19H75F0008	Gunasekhar P	Image Security Using Artificial Neural Networks	Mr. Md Eliaz
8	19H75F0009	Jayathri R	Authenticaticated medical Documents releasing with privacy protection and release control	Mrs.Ch Sabitha
9	19H75F0011	Kishore kumar S	Academic Performance Analysis	Mr. D Varun Prasad
10	19H75F0012	Komali satyavani M	Identification of plants leaf diseases using Machine Learning Algorithms	Mrs. BVN. Praveena

11	19H75F0013	Koteswara Rao M	A Hierarchical attention model for social contextual image recommendations	Mrs. S. Lavanya
12	19H75F0014	Lakshminath V	Stock market trend prediction	Mr. P. Narasimha Rao
13	19H75F0015	Madhavi latha M	Online rental system and Universal Rental capture	Mrs. Parveen Patan
14	19H75F0016	Manoj kumar K	Efficient client side Deduplication of Encrypted data with public auditing in cloud storage	Ms. T. Bhargavi
15	19H75F0017	Manoj kumar K	Venue Booking System	Mr. K. Swamy
16	19H75F0018	Naveen Ch	Combining data owner-side and cloud-side Access control for Encrypted cloud Storage	Mrs. Parveen Patan
17	19H75F0019	Navya N	Two level authentication based secure data transmission over cloud	Mr. K Mahanthi
18	19H75F0020	Pavan kalyan J	Secure Data Transfer and Deletion from Counting Bloom Filter in cloud Computing	Mr. Md Eliaz
19	19H75F0021	Rajesh K	An Efficient data retrival System in cloud for data Security	Mrs.Ch Sabitha
20	19H75F0022	Sandhya vani K	Missing child identification using DL and Multiclass SVM	Mrs. V Lakshmi Chetana
21	19H75F0023	Srilakshmi R	Driver drowsiness monitoring system using visual behavior and machine learning	Ms. K Vinaya Sree
22	19H75F0024	Tahir SK	Student Management System	Mr. J Venkata Krishna
23	19H75F0025	Tejasri K	Performance analysis and evaluation of machine learning algorithm in rainfall prediction	Mrs. S. Lavanya
24	19H75F0026	Tirumala Rao S	A Distributed Trust Evaluation Protocal with privacy protection for intercloud	Mrs. V Lakshmi Chetana

SIGNATURE OF HOD

(D.PRASAD)

Detection of Online Reviews Using Semi Supervised And Supervised Learning

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Asmat Molla (19H75F0002)

Under the esteemed guidance of

Mrs. A. Anuradha
Assistant Professor

Department of Computer Applications

DVR & Dr.HS
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July, 2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist., A.P, India.
Phone: 08678 - 273535, 9491437799, 7382016824
Website: www.mictech.ac.in

AUTONOMOUS
NAAC A

CERTIFICATE

This is to certify that the project report **entitled "Detection of Online Reviews Using Semi Supervised And Supervised Learning "** submitted by **Asmat Molla (19H75F0002)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Mrs. A. Anuradha)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K.SRINIVAS)

Dr. K. SRINIVAS, M.Tech., Ph.D (JNTUH)
Professor, Dept. of Mechanical Engg.
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India - Pin: 521180.

Internal Examiner
(Dr. G. Sai Chaitanya Kumar)

External Examiner

Designing Secure nad Efficient Biometric- based Secure Access Mechanism for Cloud Service

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Ravuri Bhavya Sri(19H75F0003)

Under the esteemed guidance of

Mrs. V Srilakshmi
Assistant Professor

Department of Computer Applications

DVR & Dr.HS
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist. A.P. India
Phone: 08678 - 275535, 9491457799, 7382016824
Website: www.mictech.ac.in

AUTONOMOUS
NAAC A

CERTIFICATE

This is to certify that the project report entitled **"Designing Secure nad Efficient Biometric- based Secure Access Mechanism for Cloud Service"** submitted by **Ravuri Bhavya Sri (19H75F0003)**to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any otherInstitute for the award of any Degree.

Supervisor
(Mrs. V. SRILAKSHMI)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K. SRINIVAS)

Dr. K. Srinivas, M.Tech., Ph.D (JNTUH)
Professor, Dept. of Computer Engg.
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India – Pin: 521180

Internal Examiner
(Dr. G. Sai Chaitanya Kumar)

External Examiner

DVR & DR. HS MIC COLLEGE OF TECHNOLOGY Kanchikacherla Ph: 521180 Ph: 275535 NAAC A GRADE

Cartoonifying Image Using Python

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Chileveru Gayatri(19H75F0004)

Under the esteemed guidance of

Mrs. R Srilakshmi

Assistant Professor

Department of Computer Applications

DVR & Dr.HS

MIC College of Technology

(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist., A.P, India.
Phone: 08678 - 275535, 9491457799, 7382616824
Website: www.mictech.ac.in

CERTIFICATE

This is to certify that the project report entitled **"Cartoonifying Image Using Python"** submitted by **Chileveru Gayatri(19H75F0004)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any otherInstitute for the award of any Degree.

Supervisor

(Mrs. R. SriLakshmi)

Head of the Department

(Mr. D.PRASAD)

Principal

(Dr. K. SRINIVAS)

Dr. K. SRINIVAS, M.Tech., Ph.D (JNTUH)
Professor, Dept. of Mechanical Engg.
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India - Pin: 521180

Internal Examiner

(Dr. G. Sai Chaitanya Kumar)

External Examiner

Facial Emotion Recognition Using Convolutional Neural Network

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Turaka Gopi(19H75F0005)
Under the esteemed guidance of
Mr K Mahanti
Assistant Professor

Department of Computer Applications

DVR & Dr.HS
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist., A.P. India
Phone: 08678 - 273535, 9491457799, 7382614824
Website: www.mictech.ac.in

ACCREDITED NAAC A

CERTIFICATE

This is to certify that the project report entitled **"Facial Emotion RecognitionUsing Convolutional Neural Network"** submitted by **TurakaGopi(19H75F0005)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(K.MAHANTHI)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K. SRINIVAS)

Dr. K. SRINIVAS, M.Tech., Ph.D (JNTUH)
Professor, Dean of Academic Affairs
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.,
Andhra Pradesh, India - Pin: 521180.

Internal Examiner
(Dr. G. Sai Chaitanya Kumar)

External Examiner

Heart Disease Identification Method Using Machine Learning Classification in E-Health Care

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Tammavarapu Gopi(19H75F0006)

Under the esteemed guidance of

Ms. K Vinaya Sree
Assistant Professor

Department of Computer Applications

DVR & Dr.HS
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist., A.P, India
Phone: 08678 - 273535, 9491457709, 7382616824
Website: www.mictech.ac.in

CERTIFICATE

This is to certify that the project report entitled "**Heart Disease Identification Method Using Machine Learning Classification in E-Health Care**"submitted by **Tammavarapu Gopi(19H75F0006)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(.Ms. K. Vinaya Sree B ai)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K. SRINIVAS)

Dr. K. SRINIVAS, M.Tech., Ph.D (JNTUH)
Professor, Dept. of Electrical Engg.
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.,
Andhra Pradesh, India - Pin: 521180.

Internal Examiner
(Dr. G. Sai Chaitanya Kumar)

External Examiner

Detection And Classification of Fruit Diseases Using Image Processing & Cloud Computing

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Mandalapu Gopi(19H75F0007)

Under the esteemed guidance of

Mr. J Venkata Krishna
Assistant Professor

Department of Computer Applications

DVR & Dr.HS
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist., A.P, India.
Phone: 08678 - 273535, 9491457799, 7382616824
Website: www.mictech.ac.in

AUTONOMOUS
NAAC A

CERTIFICATE

This is to certify that the project report entitled "**Detection And Classification of Fruit Diseases Using Image Processing & Cloud Computing**" submitted by **Mandalapu Gopi(19H75F0007)**to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Mr. P. Venkata Krishna)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K. SRINIVAS)

Dr. K. S... M.Tech, Ph.D (JNTUH)
Professor, D...
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India – Pin: 521180.

Internal Examiner
(Dr. G. Sai Chaitanya Kumar)

External Examiner

DVR & Dr. HS MIC COLLEGE OF TECHNOLOGY Kanchikacherla Pin : 521180 Ph: 273535

Image Security Using Artificial Neural Networks

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Pandi Guna Sekhar(19H75F0008)

Under the esteemed guidance of

Mr. Md Eliaz
Assistant Professor

Department of Computer Applications

DVR & Dr.HS
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist. A.P, India
Phone: 08678 - 273535, 9491457799, 7382616824
Website: www.mictech.ac.in

AUTONOMOUS
NAAC A

CERTIFICATE

This is to certify that the project report entitled **"Image Security Using Artificial Neural Networks"** submitted by **Pandi Guna Sekhar(19H75F0008)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Mr. MD. Eliaz)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K. SRINIVAS)

Dr. K. SRINIVAS ... Ph.D (JNTUH)
Professor, ...
DVR & Dr. HS ...
Kanchikacherla, Krishna Dt.,
Andhra Pradesh, India – Pin: 521180

Internal Examiner
(Dr. G. Sai Chaitanya Kumar)

External Examiner

Authenticated Medical Documents Releasing With Privacy Protection and Release Control

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Ratakonda Jayathri(19H75F0009)

Under the esteemed guidance of

Mrs.Ch Sabitha
Assistant Professor

Department of Computer Applications

DVR & Dr.HS
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist., A.P. India
Phone: 08678 - 273535, 9491457799, 7382616824
Website: www.mictech.ac.in

AUTONOMOUS
NAAC A

CERTIFICATE

This is to certify that the project report entitled "**Authenticated Medical Documents Releasing with Privacy Protection and Release Control** " submitted by **Ratakonda Jayathri(19H75F0009)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Sabitha

Supervisor

(Ms. Ch. Sabitha)

Head of the Department

(Mr. D.PRASAD)

Principal

(Dr. K. SRINIVAS)

Dr. K. SRINIVAS, M.Tech, Ph.D (JNTUH)
Professor, Dean of Engineering Engg.
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India – Pin: 521180

Internal Examiner

(Dr. G. Sai Chaitanya Kumar)

External Examiner

Academic Performance Analysis

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Sukari Kishore Kumar(19H75F0009)

Under the esteemed guidance of

Mr. D Varun Prasad

Assistant Professor

Department of Computer Applications

DVR & Dr.HS

MIC College of Technology

(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist., A.P, India.
Phone: 08678 - 273535, 9491457799, 7382616824
Website: www.mictech.ac.in

ACCREDITED NAAC A

CERTIFICATE

This is to certify that the project report entitled "**Academic Performance Analysis**" submitted by **Sukari Kishore Kumar(19H75F0009)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any otherInstitute for the award of any Degree.

Supervisor
(Mr. D. Varun Prasad)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K. SRINIVAS)

Dr. K. SRINIVAS, M.Tech., Ph.D (JNTUH)
Professor, Dept. of Computer Science Engg.
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India – Pin: 521180

Internal Examiner
(Dr. G. Sai Chaitanya Kumar)

External Examiner

Identification of Plants Leaf Diseases Using Machine Learning Algorithms

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Mudunuri Komali Satyavani(19H75F0012)

Under the esteemed guidance of

Mrs. BVN. Praveena
Assistant Professor

Department of Computer Applications

DVR & Dr.HS
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist., A.P, India
Phone: 08678 - 273535, 9491457799, 7382016824
Website: www.mictech.ac.in

AUTONOMOUS
NAAC A with Grade

CERTIFICATE

This is to certify that the project report entitled "**Identification of Plants Leaf Diseases Using Machine Learning Algorithms**" submitted by **Mudunuri Komali Satyavani(19H75F0012)**to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications in Computer Science & Engineering is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Mrs. B V N PRAVEENA)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K. SRINIVAS)

Dr. K. SRINIVAS, M.Tech., Ph.D (JNTUH)
Professor, Dean of Technical Engg.
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India - Pin: 521180

Internal Examiner
(Dr. G. Sai Chaitanya Kumar)

External Examiner

A Hierarchical Attention Model For Social Contextual Image Recommendations

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Markpudi Koteswar Rao(19H75F0013)

Under the esteemed guidance of

Mrs. S. Lavanya
Assistant Professor

Department of Computer Applications

DVR & Dr.HS
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist., A.P, India.
Phone: 08678 - 273535, 9491437799, 7382010824
Website: www.mictech.ac.in

CERTIFICATE

This is to certify that the project report entitled **"A Hierarchical Attention Model For Social Contextual Image Recommendations"** submitted by **Markpudi Koteswar Rao(19H75F0013)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Ms. S.LAVANYA)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K. SRINIVAS)

Dr. K. SRINIVAS, M.Tech, Ph.D (JNTUH)
Professor, Dept. of Computer Science & Engg.
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India – Pin: 521180.

Internal Examiner
(Dr. G. Sai Chaitanya Kumar)

External Examiner

Stock Market Trend Prediction

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Vinukollu Lakshminath(19H75F0014)

Under the esteemed guidance of

Mr. P. Narasimha Rao
Assistant Professor

Department of Computer Applications

DVR & Dr.HS
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist., A.P. India.
Phone: 08678 - 273535, 9491457799, 7382616824
Website: www.mictech.ac.in

AUTONOMOUS
NAAC A

CERTIFICATE

This is to certify that the project report entitled **"Stock Market Trend Prediction"** submitted by **Vinukollu Lakshminath(19H75F0014)**to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any otherInstitute for the award of any Degree.

Supervisor
(Mr. P. NARASIMHA RAO)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K. SRINIVAS)

Dr. K. SRINIVAS, M.Tech., Ph.D (JNTUH)
Professor, Dept. of Electrical Engg.
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India – Pin: 521180.

Internal Examiner
(Dr. G. Sai Chaitanya Kumar)

External Examiner

Online Rental System and Universal Rental Capture

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Muthyam Madhavi Latha(19H75F0015)

Under the esteemed guidance of

Mrs. Parveen Patan
Assistant Professor

Department of Computer Applications

DVR & Dr.HS
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist., A.P, India.
Phone: 08678 - 273535, 9491457799, 7382016824
Website: www.mictech.ac.in

CERTIFICATE

This is to certify that the project report entitled "**Online Rental System and Universal Rental Capture**" submitted by **Muthyam Madhavi Latha(19H75F0015)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications in a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Parveen
Supervisor

(Mrs. PARVEEN PATAN)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K. SRINIVAS)

Dr. K. [illegible]
Professor, Dept. of [illegible]
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India – Pin: 521180

Internal Examiner

(Dr. G. Sai Chaitanya Kumar)

External Examiner

Efficient Client Side Deduplication of Encrypted Data With Public Auditing in Cloud Storage

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Katikala Manoj Kumar(19H75F0016)
Under the esteemed guidance of
Ms. T. Bhargavi
Assistant Professor

Department of Computer Applications

DVR & Dr.HS
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist., A.P, India.
Phone: 08678 - 273535, 9491437799, 7382016824
Website: www.mictech.ac.in

AUTONOMOUS
NAAC A

CERTIFICATE

This is to certify that the project report entitled **"Efficient Client Side Deduplication of Encrypted Data With Public Auditing in Cloud Storage"** submitted by **Katikala Manoj Kumar(19H75F0016)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applicationsin a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any otherInstitute for the award of any Degree.

Bhargavi

Supervisor

(Ms. T. BHARGAVI)

Head of the Department

(Mr. D.PRASAD)

Principal

(Dr. K. SRINIVAS)

Dr. K. SRINIVAS, M.Tech., Ph.D.
Professor,
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India - Pin: 521180

Internal Examiner

(Dr. G. Sai Chaitanya Kumar)

External Examiner

Venue Booking System

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Kola Manoj Kumar(19H75F0017)
Under the esteemed guidance of
Mr. K. Swamy
Assistant Professor

Department of Computer Applications

DVR & Dr.HS
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist., A.P, India.
Phone: 08678 - 273535, 9491457799, 7382016824
Website: www.mictech.ac.in

AUTONOMOUS
NAAC A

CERTIFICATE

This is to certify that the project report entitled "**Venue Booking System**" submitted by **Kola Manoj Kumar(19H75F0017)**to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications in a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

K. Swami

Supervisor

(Mr. K. SWAMI)

Head of the Department

(Mr. D.PRASAD)

Principal

(Dr. K. SRINIVAS)

Dr. K. Srinivas, M.Tech, Ph.D (JNTUH)
Professor & Principal
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.,
Andhra Pradesh, India – Pin: 521180

Internal Examiner

(Dr. G. Sai Chaitanya Kumar)

External Examiner

Combining Data Owner-Side and Cloud-Side Access Control for Encrypted Cloud Stroage

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Chirumamilla Naveen(19H75F0018)

Under the esteemed guidance of

Ms.P Mareswaramma

Assistant Professor

Department of Computer Applications

DVR & Dr.HS

MIC College of Technology

(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist., A.P., India
Phone: 08678 - 273535, 9491437799, 7382016824
Website: www.mictech.ac.in

AUTONOMOUS
NAAC A

CERTIFICATE

This is to certify that the project report entitled "**Combining Data Owner-Side and Cloud-Side Access Control for Encrypted Cloud Stroage**" submitted by **Chirumamilla Naveen(19H75F0018)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications in a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Mrs. P. MARESWARAMMA)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K. SRINIVAS)

Dr. K. SRINIVAS, M.Tech., Ph.D.
Professor, Dept. of Mathematics
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India – Pin: 521180

Internal Examiner
(Dr. G. Sai Chaitanya Kumar)

External Examiner

Two Level Authentication Based Secure Data Transmission Over Cloud

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Navya Nandru(19H75F0019)

Under the esteemed guidance of

Dr.Sai Chaitanya Kumar

Assistant Professor

Department of Computer Applications

DVR & Dr.HS

MIC College of Technology

(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist., A.P. India.
Phone: 08678 - 273535, 9491457799, 7382616824
Website: www.mictech.ac.in

CERTIFICATE

This is to certify that the project report entitled "**Two Level Authentication Based Secure Data Transimission Over Cloud**" submitted by **Navya Nandru(19H75F0019)**to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Application is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Dr. G. SAI CHAITANYA KUMAR)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K. SRINIVAS)

Dr. K. Srinivas
Professor
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India – Pin: 521180.

Internal Examiner
(Dr. G. Sai Chaitanya Kumar)

External Examiner

Secure Data Transfer and Deletion From Counting Bloom Filter in Cloud Computing

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Pavan Kalyan Jillepalli(19H75F0020)

Under the esteemed guidance of

Mr. M Vijay Kumar
Assistant Professor

Department of Computer Applications

DVR & Dr.HS
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist., A.P, India.
Phone: 08678 - 273535, 9491457799, 7382016824
Website: www.mictech.ac.in

AUTONOMOUS
NAAC A

CERTIFICATE

This is to certify that the project report entitled "**Secure Data Transfer and Deletion From Counting Bloom Filter in Cloud Computing**" submitted by **Pavan Kalyan Jillepalli(19H75F0020)**to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Application is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

M. Vijay

Supervisor

(Mr. M. VIJAY KUMAR)

Head of the Department

(Mr. D.PRASAD)

Principal

(Dr. K. SRINIVAS)

Dr. K. SRINIVAS
Professor, Dept. of Computer Science
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India – Pin: 521180

Internal Examiner

(Dr. G. Sai Chaitanya Kumar)

External Examiner

An Efficient Data Retrieval System in Cloud for Data security

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Rajesh kota(19H75F0021)

Under the esteemed guidance of

Mr. D Prasad
Assistant Professor

Department of Computer Applications

DVR & Dr.HS
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist. A.P. India
Phone: 08678 - 273535, 9491457799, 7382016824
Website: www.mictech.ac.in

CERTIFICATE

This is to certify that the project report entitled "**An Efficient Data Retrieval System in Cloud for Data security**" submitted by **Rajesh kota(19H75F0021)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor

(Mr. D. PRASAD)

Head of the Department

(Mr. D.PRASAD)

Principal

(Dr. K. SRINIVAS)

Dr. K. S
Profes
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India – Pin: 52118

Internal Examiner

(Dr. G. Sai Chaitanya Kumar)

External Examiner

Missing Child Identification Using DL and Multiclass SVM

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Sandhya Kasthala(19H75F0022)

Under the esteemed guidance of

Mrs. V Lakshmi Chetana
Assistant Professor

Department of Computer Applications

DVR & Dr.HS
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist., A.P, India
Phone: 08678 - 273535, 9491457799, 7382616824
Website: www.mictech.ac.in

AUTONOMOUS
NAAC A

CERTIFICATE

This is to certify that the project report entitled **"Missing Child Identification Using DL and Multiclass SVM"** submitted by **Sandhya Kasthala(19H75F0022)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Mrs. V. LAKSHMI CHETANA)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K. SRINIVAS)

Dr. K. SRINIVAS
Professor & Principal
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India – Pin: 521180

Internal Examiner
(Dr. G. Sai Chaitanya Kumar)

External Examiner

DVR & Dr. HS MIC COLLEGE OF TECHNOLOGY
Kanchikacherla
Ph. Office: Ph.273535

Driver Drowsiness Monitoring System Using Visual Behaviour and Machine Learning

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Rangiseeti SriLakshmi(19H75F0023)

Under the esteemed guidance of

Ms. K Vinaya Sree Bai
Assistant Professor

Department of Computer Applications

DVR & Dr.HS
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist., A.P, India
Phone: 08678 - 273535, 9491457799, 7382010824
Website: www.mictech.ac.in

AUTONOMOUS
NAAC A

CERTIFICATE

This is to certify that the project report entitled "**Driver Drowsiness Monitoring System Using Visual Behaviour and Machine Learning**" submitted by **Rangiseeti SriLakshmi(19H75F0023)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Application is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Ms. K. VINAYA SREE BAI)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K. SRINIVAS)

Internal Examiner
(Dr. G. Sai Chaitanya Kumar)

External Examiner

Dr. K. [illegible] (JNTUH)
Principal
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.,
Andhra Pradesh, India – Pin: 5211[illegible]

Student Management System

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Shaik Tahir(19H75F0024)

Under the esteemed guidance of

Mrs. V Lakshmi Chetana

Assistant Professor

Department of Computer Applications

DVR & Dr.HS

MIC College of Technology

(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist. A.P, India.
Phone: 08678 - 273535, 9491457799, 7382616824
Website: www.mictech.ac.in

AUTONOMOUS
NAAC A with Grade

CERTIFICATE

This is to certify that the project report entitled "**Student Management System**" submitted by **Shaik Tahir(19H75F0024)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Application is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Mrs. V. LAKSHMI CHETANA)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K. SRINIVAS)

Dr. K. SRINIVAS, Ph.D (JNTUH)
Professor
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India – Pin: 521180

Internal Examiner
(Dr. G. Sai Chaitanya Kumar)

External Examiner

Performance Analysis and Evalution of Machine Learning Algorithm in Rainfall Prediction

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Teja Sri Kapa(19H75F0025)

Under the esteemed guidance of

Mrs. S Lavanya
Assistant Professor

Department of Computer Applications

DVR & Dr.HS
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist., A.P, India.
Phone: 08678 - 273535, 9491457799, 7382616824
Website: www.mictech.ac.in

CERTIFICATE

This is to certify that the project report entitled "**Performance Analysis and Evalution of Machine Learning Algorithm in Rainfall Prediction**" submittedby **Teja Sri Kapa(19H75F0025)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Application is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor

(Ms. S. LAVANYA)

Head of the Department

(Mr. D.PRASAD)

Principal

(Dr. K. SRINIVAS)

Dr. K. [illegible] M.E., Ph.D (JNTUH)
Professor [illegible]
DVR & Dr. HS [illegible] of Techn...
Kanchikacherla, Krishna Dt.
Andhra Pradesh, India – Pin: 52118[illegible]

Internal Examiner

(Dr. G. Sai Chaitanya Kumar)

External Examiner

A Distributed Trust Evaluation protocol with Privacy Protection for Inter Cloud

A Project Report
submitted in partial fulfillment of the
requirements for the award of the degree of

Master of Computer Applications

By

Samudrala Tirumala Rao (19H75F0024)

Under the esteemed guidance of

Mrs. A Anuradha
Assistant Professor

Department of Computer Applications

DVR & Dr.HS
MIC College of Technology
(Autonomous)

Kanchikacherla– 521180, Krishna Dist., Andhra Pradesh

July,2021

DVR & Dr. HS
MIC College of Technology
Kanchikacherla - 521180, Krishna Dist., A.P, India.
Phone: 08678 - 273535, 9491457799, 7382616824
Website: www.mictech.ac.in

AUTONOMOUS
NAAC A

CERTIFICATE

This is to certify that the project report entitled "**A Distributed Trust Evaluation protocol with Privacy Protection for Inter Cloud**" submitted by **Samudrala Tirumala Rao (19H75F0024)** to the DVR & Dr. HS MIC College of Technology in partial fulfillment of the requirements for the award of the Degree of Master of Computer Application is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor
(Mrs. A. ANURADHA)

Head of the Department
(Mr. D.PRASAD)

Principal
(Dr. K. SRINIVAS)

Professor & Principal
DVR & Dr. HS MIC College of Technology
Kanchikacherla, Krishna Dist.
Andhra Pradesh, India – Pin: 521180

Internal Examiner
(Dr. G. Sai Chaitanya Kumar)

External Examiner

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited B.Tech (CSE|ECE|ME)
Accredited by NAAC With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"EVALUATING THE PERFORMANCE OF SHELL AND TUBE HEAT EXCHANGER BY THE ADDITION OF SWIRL "** is a bonafide work carried out by **Vishnu Vardhan G (17H75A0360), Suhail Ahmed Md (17H75A0346), Naga Vamsi N (17H75A0324), Rasheed Sd (16H71A0338), Kishan K (17H75A0312)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"EXPERIMENTAL INVESTIGATION ON TRIBOLOGICAL BEHAVIOUR OF ZA-27 ALLOY REINFORCED WITH BORON NITRIDE AND BORON CARBIDE "** is a bonafide work carried out by **Venkata Sai P (17H75A0356), Ramesh C (17H75A0334), Vamsi Krishna K (17H75A0351), Lokesh V (16H71A0368), Naga Sai A (16H71A0322)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"PREDICTION OF MATERIAL REMOVAL RATE AND SURFACE ROUGHNESS IN WIRE ELECTRICAL DISCHARGE MACHINING PROCESS USING REGRESSION ANALYSIS"** is a bonafide work carried out by **Mahesh Naidu B (17H75A0315), Nagasai P (17H75A0325), Naga Chaitanya V (16H71A0373), Nagendra Babu B (16H71A0325), Rahul B (16H71A0333)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"RESPONSE SURFACE METHODOLOGY ANALYSIS OF TIG WELDED BUTT JOINTS "** is a Bonafede work carried out by **Anil Kumar K (17H75A0303), Mani Kanth V (17H75A0316), Naveen P (17H75A0326), Bhanu Prasad K (16H71A0358), Chandra Sekhar Reddy K (16H71A0361)** to the **Devineni Venkata Ramana & Dr. Hima Sekhar MIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: drvhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"DESIGN AND ANALYSIS OF MACHINE TOOLS ELEMENTS FOR LATHE SPINDLE"** is a bonafide work carried out by **Kondala Rao G (16H71A0367), Suman K 16H71A0347), Sai Tulasiram Ch (16H71A0390), Naga Aravind B (17H75A0320), Sai Babu D (16H71A0340)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: drvhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"DESIGN AND ANALYSIS OF PEDAL OPERATED GRASS CUTTER"** is a bonafide work carried out by **Narasimha Rao D (16H71A0374), Somandeep V (17H75A0344), Poorna Kiran V (17H75A0329), Joy Benson K (14H71A0380), Gopi Krishna M (16H71A0305)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"PERFORMANCE IMPROVEMENT OF PYRAMID SOLAR STILL USING FINS WITH AND WITHOUT PCM "** is a bonafide work carried out by **Ganesh Mani K (17H75A0308), Noushad Shaik (16H71A0327), Saikrishna J (17H75A0341), Bheshman Sridatta K (16H71A0360), Ajay B (16H71A0301)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: drvhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"PERFORMANCE CHARACTERISTICS AND EMISSION ANALYSIS OF TAMANU OIL ON BIODIESEL-DIESEL-METHANOL"** is a bonafide work carried out by **Pavan Kumar K (17H75A0328), Venka Mahesh Kumar Ch (17H75A0352), Upendra Varma M K (17H75A0349), Naveen K (16H71A0377), Srinivas L (16H71A0346)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"DESIGN AND FABRICATION OF MULTIPURPOSE ROVER USING ROCKER BOGIE MECHANISM "** is a bonafide work carried out by **Venkata Ramana A (17H75A0354), Indrababu B (17H75A0311), Venkateswara Rao D (17H75A0358), Pavan Kumar (16H71A0328), Sumanth T (16H71A039)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"STUDY THE STRUCTURAL PROPERTIES OF AA7075 VARYING INPUT PROCESS PARAMETERS BY TIG WELDING"** is a bonafide work carried out by **Sai Akash B (16H71A0339), Kusuma Haranadh M (16H71A0317), Rampandu B (17H75A0335), Sai Krishna V (16H71A0387), Sai Babu B (17H75A0338)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"WELDING SIMULATION ON SS304 AND INCONEL 617 "** is a bonafide work carried out by **Rakesh K (16H71A0335), Dheeraj Krishna Nagendra A (17H75A0306), Surendra Kumar K (16H71A0348), Tariq Ahmed Sk (16H71A0395), Yaswanth Kumar D (17H75A0361)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by NAAC With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"PRODUCTIVITY IMPROVEMENT OF A SOLAR STILL BY SENSIBLE HEAT STORAGE MATERIAL "** is a bonafide work carried out **Manikanth G (17H75A0318), Gopi Kiarn P (17H75A0309), Siva Rama Kumar Reddy G (17H75A0343), Pavan N (16H71A0378), Prem Chand D (15H71A0366)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"DESIGN AND THERMAL ANALYSIS OF A BOILER TUBE WITH INTERNAL HELICAL RIBS"** is a bonafide work carried out by **Sai Krishna B (17H75A0339), Prasanth K (17H75A0331) Guna Sekhar B (16H71A0321), Rama Narasimha Rao O (16H71A0336), Naga Dheeraj K (16H71A0321)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"ESTIMATION AND COSTING OF A THREE PULLEY BLOCK ASSEMBLY"** is a bonafide work carried out by **Sai Bhavani Naga Surya Teja D (16H71A0341), Siva Prasad Reddy Ch (16H71A0345), Ganesh Manikanta S (16H71A0362), Yaswanth Kumar B (16H71A03A4), Naga Vijaya Kumar A (16H71A0324)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited B.Tech (CSE|ECE|ME)
Accredited by NAAC With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **" PERFORMANCE AND EMISSION ANALYSIS OF BIODIESEL ON FOUR STROKE CI ENGINE BY USING TOBACCO OIL WITH PROPANOL"** is a bonafide work carried out by **Ahmed Basheer Sk (17H75A0301), Sri Harsha M (17H75A0345), Raj Kumar K (16H71A0334), Vamsi Krishna K (16H71A0352), Raj Kumar M (16H71A0380)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"STRUCTURE PROPERTY CORRELATION OF MODIFIED AL-MG ALLOY FOR AEROSPACE APPLICATIONS"** is a bonafide work carried out **Naveed Ahamad Sd (16H71A0376), Venkata Pavan Kumar C (17H75A0353), Siva Naga Raju Ch (17H75A0342), Yeswanth Y (16H71A03A5), Jaseem Sk (16H71A0313)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by NAAC With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"DESIGN AND FABRICATION OF PIEZO ELECTRIC SHOES FOR MILITARY APPLICATIONS"** is a bonafide work carried out **Pavan Kumar N (16H71A0329), Sai Srinadh B (17H75A0340), Venkata Sai Krishna Y (16H71A0399), Naga Sri Pavan K (16H71A0323), Mohammad Raffi Sk (16H71A0371)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"DESIGN AND STATIC STRUCTURAL ANALYSIS OF CRANK SHAFT"** is a bonafide work carried out **Gopi Krishna Ch (17H75A0363), Karthik M (17H75A0311), Kalyan Kumar K (16H71A0315), Teja K (17H75A03485), Vignan Teja G (16H71A03A23)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"PERFORMANCE CHARACTERISTICS AND EMISSION ANALYSIS OF TAMANU OIL BIODIESEL-DIESEL-PROPANOL BLENDS "** is a bonafide work carried out **Venkata Sai Kumar T (16H71A0382), Vamsi Krishna G (17H75A0350), Sasank R (16H71A0392), Sandeep Ch (16H71A0391), Venkata Naveen B (16H71A0353)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"FABRICATION AND EXPERIMENTAL EVALUATION OF MECHANICAL PROPERTIES IN HYBRID NATURAL FIBER COMPOSITE MATERIALS (JUTE-NEEM)"** is a bonafide work carried out **Sandeep B (16H71A0343), Hardhik Hari Deep P (16H71A0364), Akhil P (17H75A0302), Uma Maheswar Rao D (16H71A0351), Azhar Akram Md (16H71A0355)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by NAAC With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"DESIGN AND EVALUATION OF CONDENSER AND DIFFUSER OF VCR SYSTEM "** is a bonafide work carried out **Naga Sai Ravindra Sarma M (17H75A0323), Syamprasad K (17H75A0347), Gopala Rao K (16H71A0363), Balnaga Ajay Susanth K (16H71A0357), Sailesh Buddi (16H71A0342)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"REDUCTION OF FRICTION LOSSES IN AUTOMOBILE ENGINE BY ADDING NANO PARTICLES IN BASE OIL (CASTROL SAE 10W30 4T ENGINE OIL)"** is a bonafide work carried out **Prasad D (17H75A0330), Deeva Selva Samuel Das M (17H75A0305), Gopi V (16H71A0306), Sunny B (16H71A0394), Harish K (16H71A0309)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"TRIBOLOGICAL CHARACTERISTICS OF MODIFIED AL-MG ALLOY FOR LIGHT WEIGHT APPLICATIONS"** is a bonafide work carried out **Venkata Ramana E (17H75A0355), Yeswanth Babu N (17H75A0362), Sai Kumar N (16H71A0388), Harish K (16H71A0309)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"DESIGN AND ANALYSIS OF SMALL WIND TURBINE FOR POWER GENERATION THROUGH EXHAUST AIR"** is a bonafide work carried out **Ravi Kumar B (17H75A0336), Uday Kumar K 16H71A0396), Uday Bhaskar M (16H71A0350), Jagadish Babu V (16H71A0312)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: drvhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by NAAC With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"DESIGN AND ANALYSIS OF PRESSURE VESSEL USING COMPOSITE LAYERS AND CHANGING END CONDITIONS"** is a bonafide work carried out **Prudviraj P (17H75A0333), Naga Manikanta T (17H75A0322), Roja Ram P (17H75A0337), Narayana J (16H71A0375)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"DESIGN AND DEVELOPMENT OF VARIOUS LIQUID FUELS FIRED POROUS BURNER"** is a bonafide work carried out **Pardhu m (17H75A0327), Venu Gopal P (17H75A0359), Praveen G (17H75A0332), Manohar Ch (16H71A0370)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by NAAC With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"HEAT TRANSFER ANALYSIS ON FINS WITH DIFFERENT GEOMETRIES"** is a bonafide work carried out **Jwala Venkata Narendra Ch (16H71A0314), Mani Kumar P (17H75A0317), Purdhvi Kumar P (16H71A0332), Chaitanya Praveen Surya Teja (17H75A0304 Madhu Babu G (15H71A03A3)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"EVALUATION OF MECHANICAL PROPERTIES OF TIG WELDED STAINLESS STEEL AND MILD STEEL"**is a bonafide work carried out **Venkatesh K (17H75A0357), Sena Sai Ram K (15H71A0336), Ramesh D (16H71A0337), Asmatullah Sk (16H71A0303), Ganesh M (16H71A0304)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"PERFORMANCE PARAMETERS TESTING OF A TYPICAL TWO WHEELER ENGINE RUNNING ONLY ON CNG "** is a bonafide work carried **out Mohseen Raza Ab (17H75A0319), Naga Durga Jagadish B (15H71A03B1), Rangisetti Gopi (16H75A0303), Dileep Vinay B (17H75A0307), Pavan Ramakotesh K (16H71A0342)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e-mail: drvhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited
B.Tech (CSE|ECE|ME)
Accredited by
NAAC
With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"FABRICATION AND EXPERIMENTAL EVALUATION OF MECHANICAL PROPERTIES IN HYBRID NATURAL FIBER COMPOSITE MATERIALS (RAAVI-NEEM) "** is a bonafide work carried out **Harshavardhan Reddy Ch (16H71A0310), Ravindra Babu J (16H71A0384), Bala Akhil Sai P (16H71A0356), Sai Gokul R (16H71A0385)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal

Devineni Venkata Ramana & Dr.Hima Sekhar
MIC College of Technology
(Approved by AICTE & Permanently Affiliated to JNTUK, Kakinada)
Kanchikacherla - 521180, Krishna Dist, A.P, India.
Phone : 08678 - 273535, 273623, Fax: 08678 - 273569
e mail: dvrhsmic@mictech.ac.in, Website: www.mictech.ac.in

AUTONOMOUS
NBA-Accredited B.Tech (CSE|ECE|ME)
Accredited by NAAC With Grade A

CERTIFICATE

This is to certify that the Major Project entitled **"EXPERIMENTAL INVESTIGATION ON TRIBOLOGICAL BEHAVIOUR OF ZA-27 ALLOY REINFORCED WITH BORON NITRIDE"** is a bonafide work carried out **Kishore Krishnachaitanya k (17H75A0310), Suraj Ch (16H71A0349), Jagadeesh S (17H75A0310), Ramki R (16H71A0383)** to the **Devineni Venkata Ramana & Dr.Hima SekharMIC College of Technology** in partial fulfullment of the award of the Bachelor Degree of Technology in **MECHANICAL ENGINEERING** is a bonafide record of work. The contents of this report, in full or in parts, have not been submitted to any other institute for the award of any degree

Supervisor **Head of the Department**

Principal