

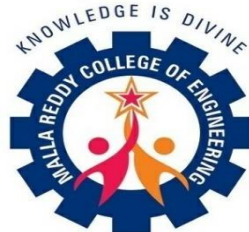
A
TECHNICAL SEMINAR
REPORT ON
MACHINE LEARNING PROJECT
ON
PREDICTION OF LOAN

submitted in the partial fulfillment of the requirement for the award of the degree of

BACHELOR OF
TECHNOLOGY IN
COMPUTER SCIENCE AND ENGINEERING
BY

T.VENKAT PRAVEEN

18Q91A05N5



DEPARTMENT OF COMPUTER SCIENCE AND
ENGINEERING

MALLA REDDY COLLEGE OF
ENGINEERING

(Affiliated To JNTU, Hyderabad, Approved By AICTE, New Delhi, Accredited by NBA)

Maisammaguda, Dhulapally (Post Via Kompally), Secunderabad,

Telangana-500100

2021-22



MALLA REDDY COLLEGE OF ENGINEERING

(Affiliated To JNTU, Hyderabad , Approved By AICTE , New Delhi, Accredited by NBA

Maisammaguda, Dhulapally, (Post Via Kompally), Secunderabad

Telangana State500100

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Technical Seminar titled “**MACHINE LEARNING PROJECT ON PREDICTION OF LOAN**” is bonafide work done by **Mr. TEEDA.VENKAT PRAVEEN, 18Q91A05N5** in the partial fulfillment of the award of the degree in **Bachelor of Technology** in **Computer Science and Engineering** submitted to the Jawaharlal Nehru Technological University Hyderabad during the academic year **2021-22**.

Seminar Supervisor

Mr. Sayyad. Rasheeduddin

Assistant Professor

Head of the Department

Mrs. Ch. Vijaya Kumari

Assoc. Professor

DECLARATION

I VENKAT PRAVEEN with **Reg No: 18Q91A05N5** hereby declare that, Technical Seminar report entitled “**MACHINE LEARNING PROJECT ON PREDICTION OF LOAN**” done by me is submitted in the partial fulfillment of the requirements for the award of degree of **BACHELOR OF TECHNOLOGY** in **COMPUTER SCIENCE AND ENGINEERING** is my original work.

The Results embodied in this project report have not been submitted to any other University or institute for the award of any degree or diploma.

Signature of the Candidate

VENKAT PRAVEEN

18Q91A05N5

.....

Date :

Place :Hyderabad

ACKNOWLEDGEMENT

I am indeed indebt to our Head of the Department **Mrs.Ch.Vijaya Kumari**,Associate Professor, for giving her valuable suggestions and constant encouragement which helped us in completing this technical seminar successfully.

I would like to thank our Seminar coordinator **Mr. Sayyad Rasheeduddin**, M.Tech,[Ph.D] Assistant Professor, for his technical guidance, constant encouragement and support in carrying out my Seminar. We also extend our thanks to our Teaching and Non-Teaching members for their cooperation during our course.

VENKAT PRAVEEN
18Q91A05N5

CONTENTS

- **LIST OF TABLES**
- **LIST OF ACRONYMS**

CHAPTER 1

INTRODUCTION

OBJECTIVE

PROBLEM DEFINITION

SCOPE OF PROJECT

EXTERNAL INTERFACE REQUIREMENTS

SOFTWARE REQUIREMENTS

HARDWARE REQUIREMENTS

OPERATING ENVIRONMENT

CHAPTER 2

LITERATURE SURVEY

INTRODUCTION

EXISTING SYSTEM

PROPOSED SYSTEM

MODULES

CHAPTER 3

DESIGN

INTRODUCTION

UML DIAGRAMS

CHAPTER 4

IMPLEMENTATION

INTRODUCTION

DESIGN AND IMPLEMENTATION CONSTRAINTS

SCREENSHOTS OF PROJECT IMPLEMENTATION

STEP BY STEP IMPLEMENTATION PROCESS

CHAPTER 5

TESTING

SOFTWARE TESTING

SCREENSHOTS OF TESTING

CHAPTER 6

CONCLUSION

REFERENCES

CODING

LIST OF TABLES

| # | Title of System Feature | Description | Priority | Functional Requirements |
|----|-------------------------|---|-----------|--|
| 1. | Login Page | This page is used to login and maintain the security by Authenticating the user details. | High 9 | <ul style="list-style-type: none"> It only accepts with user id and password. For an invalid user id or password ,it gives an error message. |
| 2. | Admin Page | This page is used to show the data/details of any respective customer who apply for loan. | High 9 | All the details of customers would be displayed. |
| 3. | Home Page | This Page behaves as a user interface to display the featuresof the database and guides the user to enter the details for Prediction. | High 8 | It provides the attributes for any respective user for prediction of loan. |

LIST OF ACRONYMS

| | |
|-------------|----------------------------------|
| HTML | Hypertext Markup Language |
| CSS | Cascading Style Sheets |
| ER | Entity Relationship |
| DB | Data base |

CHAPTER 1

Introduction

OBJECTIVE

Objective of this project :

- firstly, Collection of data.
- Secondly, Preprocessing and cleaning of data.
- then, Modelling with various algorithms to obtain the most suitable algorithm for the respective case.
- Backend
- Developing user friendly UI and integrating with the backend.
- We will be having the data collected from the respective users accessible in the admin portal which can be used to improve the model.

PROBLEM DEFINITION

So we all know that the usual process for loan eligibility is too long like the customers first apply for the loan and then the bank will check for the customer eligibility to offer him/her loan or not. However, the usual process takes a lot of time. Hence, we want to automate and make the process of loan eligibility as fast as possible based on their respective information. So in this case lots of time can be reduced and loan eligibility criteria can be approved in the proper manner.

SCOPE OF PROJECT

This Project is mainly aimed to develop a user-friendly environment between a respective customer and respective bank. It decreases the loan time and using the (ML) machine learning models it can easily predict the loan approval based

on the respective customer information and for that we will be designing a web application using flask which is a python framework and saving all customers details in the database which we are using is MySQL.

EXTERNAL INTERFACE REQUIREMENTS

Software Requirements

1. Py-charm community
2. OS: Windows XP
3. PHP Triad (PHP5.6, MySQL, Apache, and PhpMyAdmin)

Hardware Requirements

1. Intel core i5 Processor
2. 1GB RAM
3. 20GB HARDDISK

Non-Functional Requirements

- Performance -

It actually performed well on Google Chrome as it requires a low amount of storage and ram. Its performance is mainly based on the internet connection. If there is a fast internet connection, then it gives the output fast otherwise it depends on the speed of the internet.

- Reliability-

This web application is reliable in nature, and it only goes down whenever there is no stable internet connection.

- Availability-

This web application is easily available for access on web browsers. It is user-friendly in appearance.

- Security-

As this web application is developed accordingly by keeping various security parameters in mind therefore it secures the data of the respective user.

- Maintainability-

This web application is maintained easily.

- Portability-

This web application is easily moved to another browser because of the cross-browser compatibility code used, therefore the app is portable.

OPERATING ENVIRONMENT

Python – Programming language

Frontend – HTML

Backend – MY SQL

PyCharm – Tool

Software Interfaces

The software is based on application interface, and it requires the following technical skills

1. HTML

HTML stands for Hypertext Markup Language (HTML) which is a standard markup language used for creating web pages and web applications. With the help of Cascading Style Sheets (CSS) and JavaScript, it forms a whole layout page for the World Wide Web. The Web browsers will receive HTML documents from a web server or from local storage and will provide the respective documents into multimedia web pages. HTML helps us to show the structure of a web page with appropriate design and manage the work.

2. CSS

CSS stands for Cascading Style Sheets (CSS) which is a style sheet language used for showing the design of a respective document which is written in a markup language like HTML. CSS is a style sheet technology of the World Wide Web, including JavaScript and HTML.

3. FLASK

Flask is a mini web framework that provides us with some tools, libraries and some technologies to construct a light-weight web applications using python. Flask was first developed by Armin Ronacher who leads an international group of python enthusiasts (POCCO). Flask is known for its mini web framework and is also easy to use compared to Django. Flask is used vastly for its easy process.

CHAPTER 2

Literature Survey

INTRODUCTION

In literature Survey we are going to look into the details of my existing Project and some left over drawbacks of the proposed System. We are going to improve the performance and efficiency of the new System.

EXISTING SYSTEM

Resource : A survey on Model for Loan Prediction Anchal Goyal [1], Ranpreet Kaur [2] Research Scholar [1], Assistant Proffesor [2] Department of Computer Science and Engineering RIMT –IET (PTU), Mandi Gobindgarh Punjab – India.

The existing web application will not have details of respective Customer , which is one of the major drawback and the data collected from respective Customer may be in danger because of data breaching. Because nowadays everyone wants their data to be Confidential.

So, long before a gentleman named Dr.K.Mishra proposed a method in which two or more classifiers are combined together to produce an ensemble model for the better prediction. So, they used the bagging and boosting techniques and then also used random forest techniques. The process of classifiers is to improve the performance of the data and it gives better efficiency. Even though there are some web applications for predictions of loan, but they don't have the attributes that I have in my project and also there's no safety on data of respective customer .So, keeping this in mind I have given credentials page for more safety and to avoid data leak.

PROPOSED SYSTEM

This project is a Flask based web app that deploys a machine learning model which predicts if the loan can be approved or not based on certain attributes and also previous financial history. . Each Customer must Sign up first for knowing the result of loan approval.

Sign up is mandatory for all the Customers. Even there is a login credential for Admin. The data given by the users/customers will be available in the admin portal which can be used for further analysis and betterment of the my model. The data will be safely stored. The data can also be used by the banks to extend their loan campaigns.

MODULES

MODULE 1 : ADMIN

Admin maintains all the details of respective Customer. He manages all the data , which will be safely stored in Admin Portal.

MODULE 2 : USER LOGIN

It only accepts with user id and password. For an invalid user id or password ,it gives an error message. This page is used to login and maintain the security by Authenticating the user details.

MODULE 3 : HOME PAGE

It provides the attributes for any respective user for prediction of loan. This Page behaves as a user interface to display the features of the database and guides the user to enter the details for Prediction.

CHAPTER 3

Design

INTRODUCTION

Today's generation many people rely on bank loans to fulfill their needs. Hence, the rate of loan applications increases with a very fast speed in recent years. So, the risk is always involved in approval of loans.

The banking officials/Bank Staffs are very conscious about the payment of the loan amount by its respective customers. Even after taking a lot of precautions and analyzing the respective loan applicant data, the loan approval decisions are always not correct. Firstly, a Loan is a sum of money that one or more individuals or companies borrow from banks or other financial institutions, So as to financially manage planned or unplanned events. There are various attributes taken into consideration to approve or disapprove a loan.

My project aims to develop a web app that will help the respective user to know if their loan can be approved or not. AI is an emerging technology nowadays. Hence, the application of AI solves many problems of the real world. (ML) Machine Learning is an AI technique which is very much useful in prediction systems. It creates a model from training data. While making the prediction of loan the model

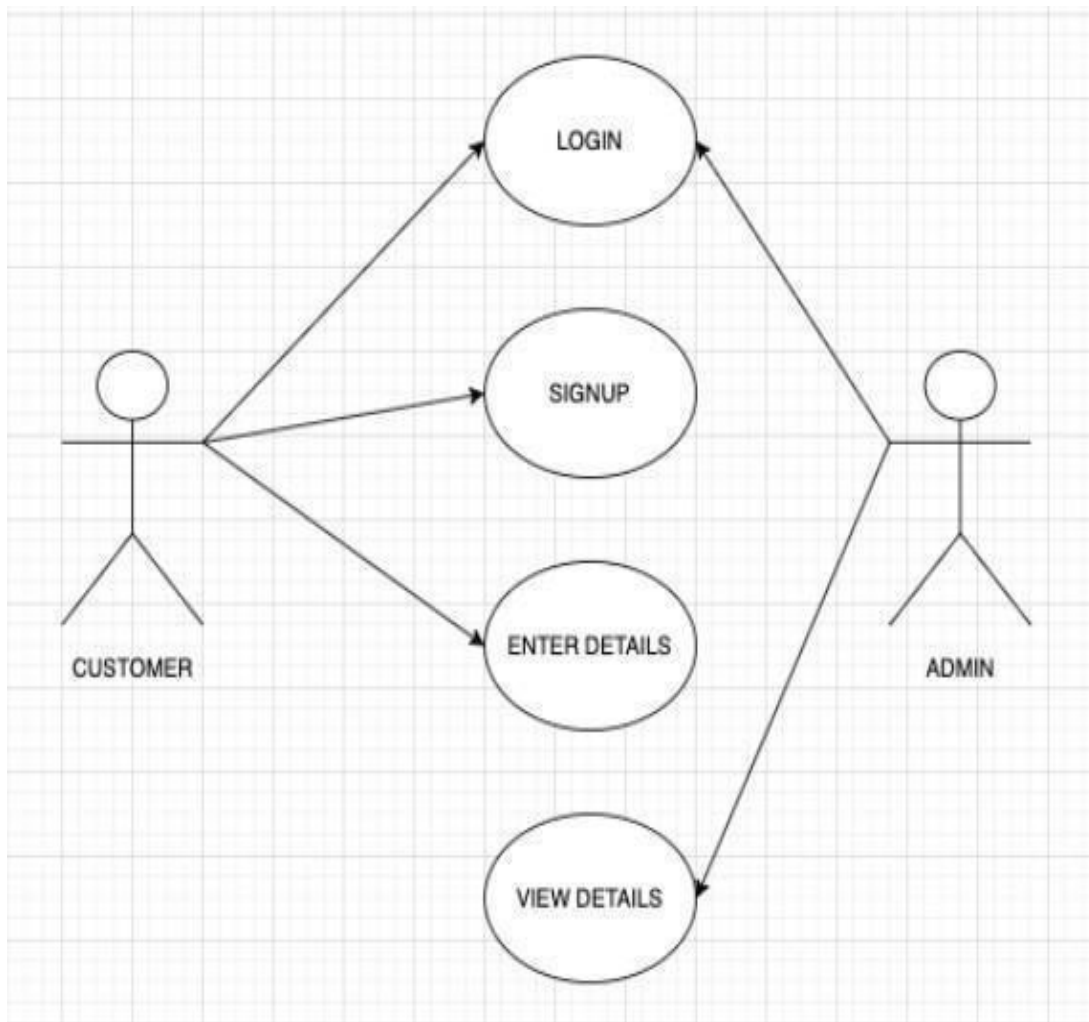
which is developed by training algorithm (which is machine learning) is used. The machine learning algorithm will train the system using a fraction of the data available and test the remaining data.

UML DIAGRAMS

LIST OF FIGURES

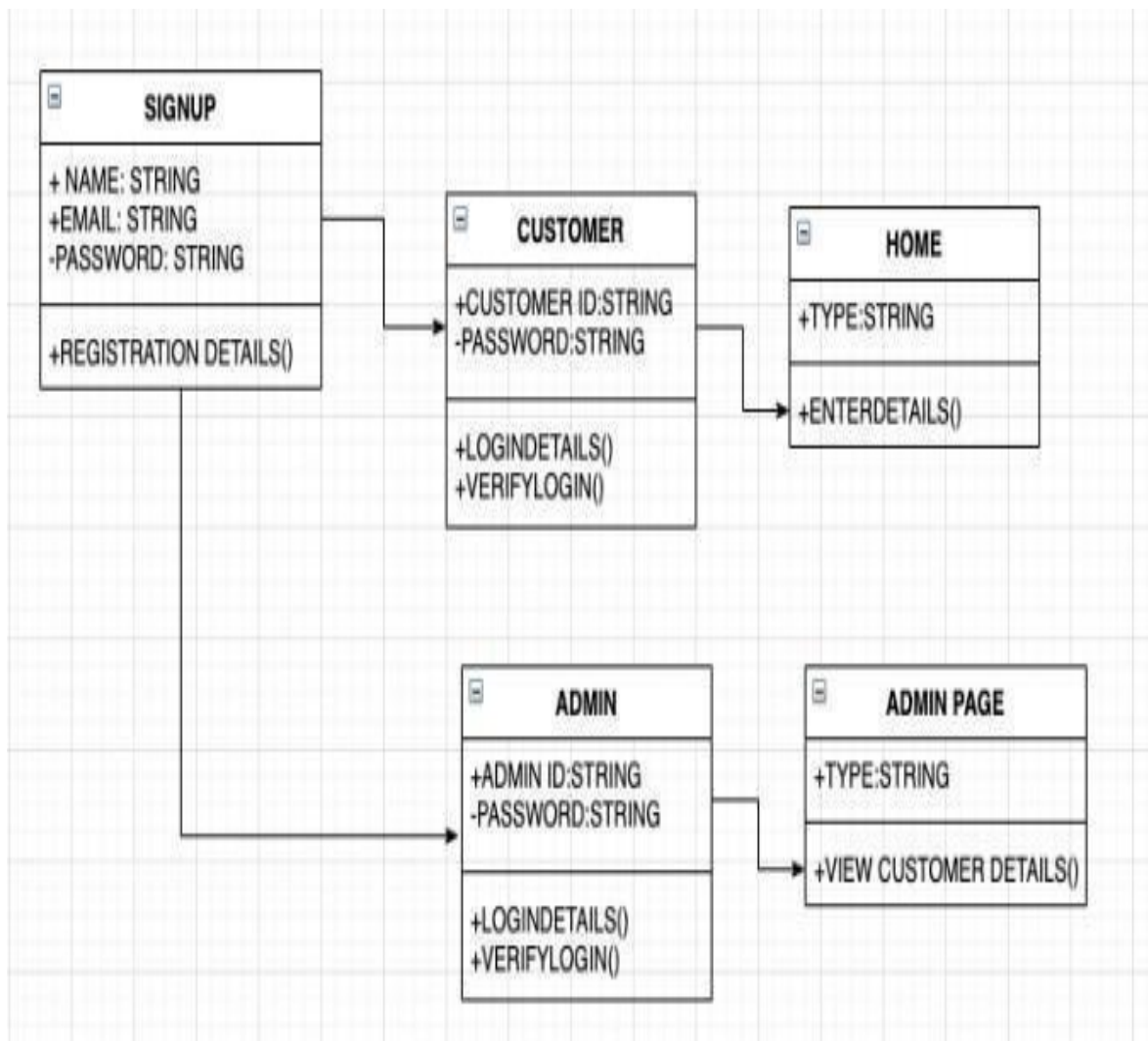
- USE CASE DIAGRAM
- CLASS DIAGRAM
- ENTITY RELATIONSHIP DIAGRAM
- ARCHITECTURE DIAGRAM
- PIPE FILTER ARCHITECTURE DIAGRAM
- COLLABORATION ARCHITECTURE
DIAGRAM

Use Case Diagram



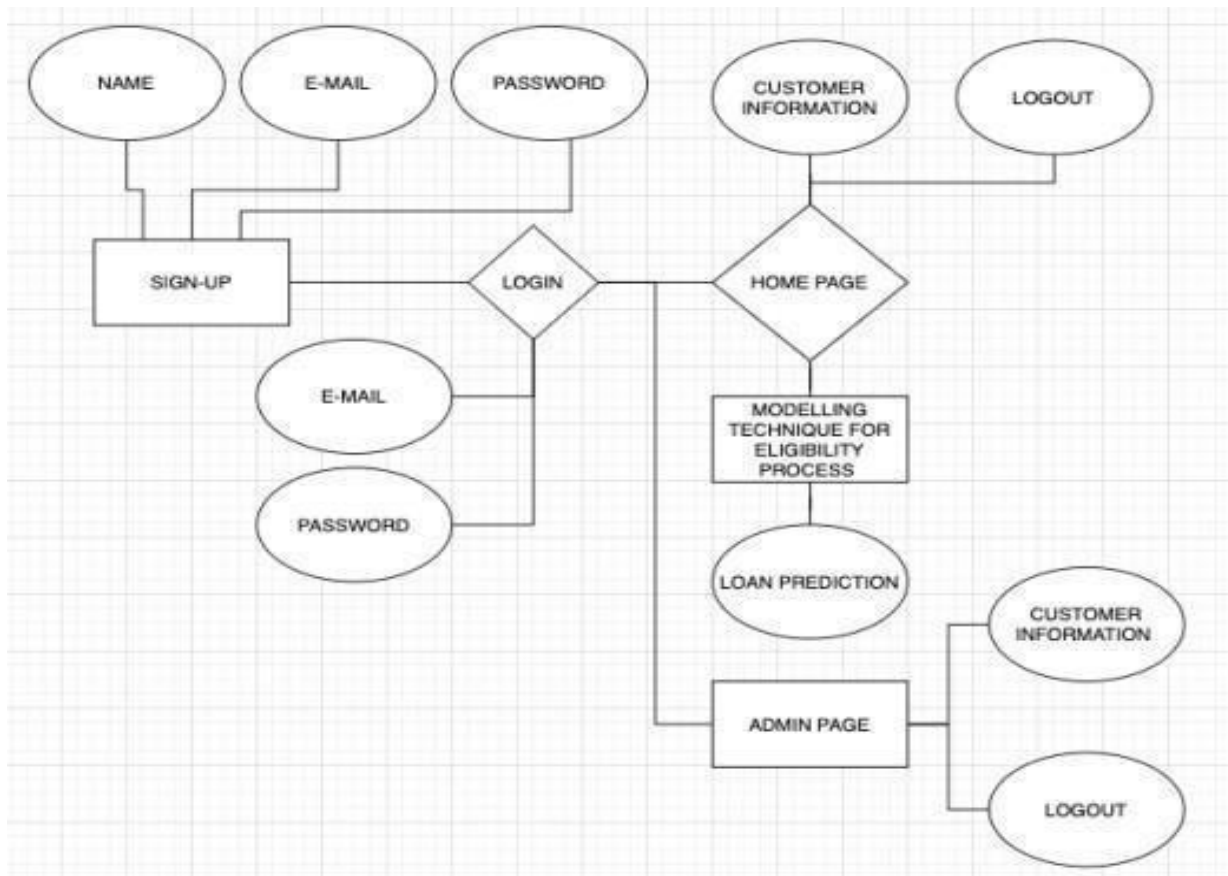
From the above diagram we can understand that, Customer will be first directed to Signup page then enter the details and after the process, the respective customer must have to login and then he/she will be directed to Homepage. The whole details can be viewed by Admin.

Class Diagram



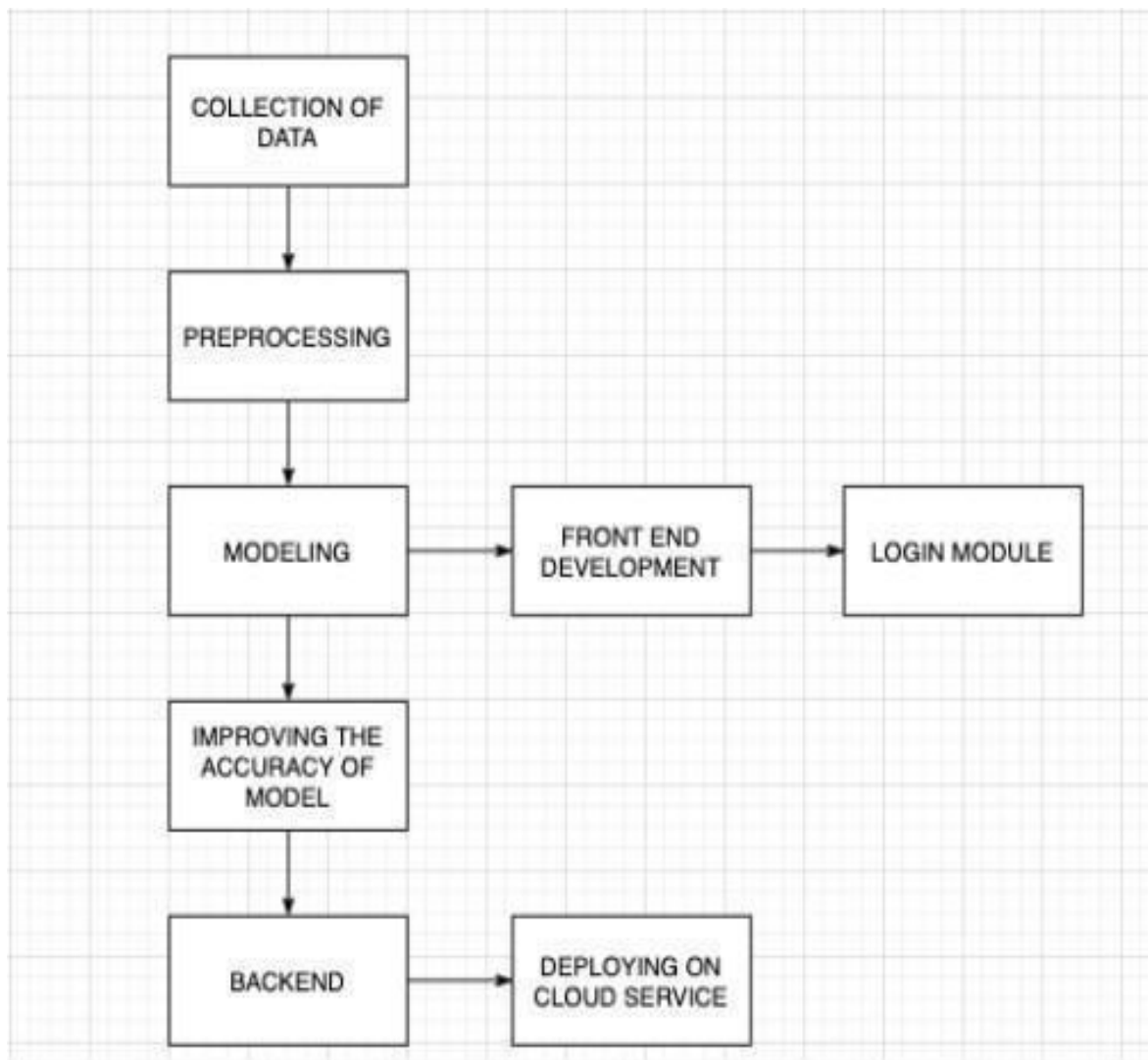
As you can see the Class diagram illustrates us with three different Sections the upper Section is name of the Class ,Middle Section will be attributes of Class, and the bottom Section includes class Operators.

Entity Relationship Diagram



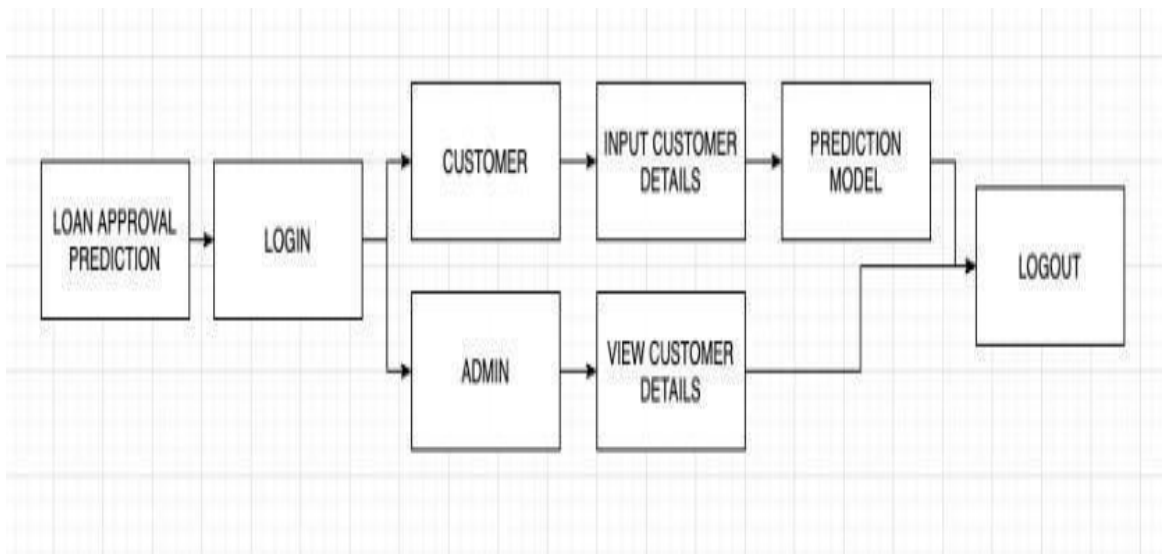
The diagram shows us the relation between entities as follows there are three entities in the above diagram and their attributes are also shown and the shape with rhombus classifies us the relation.

Architecture diagram



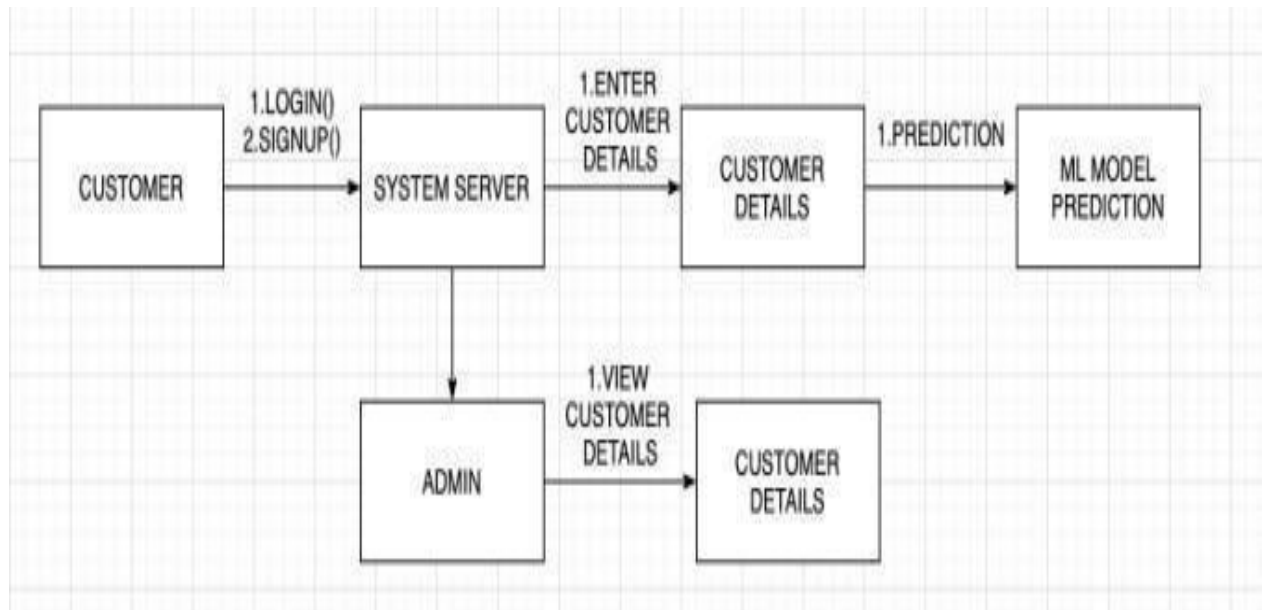
It shows the abstract and also the overall outline: relationships, Constraints and boundaries between components. From the above diagram we can understand the whole process from collection of data to the final stage (i.e.) deploying it on Cloud Service.

Pipe filter architecture Diagram



Pipe and Architecture Diagram shows the design pattern that allows for the whole Process of a System. Customer will be having a separate page and the Admin page will be having Separate page , but they both will be having their respective pages and they can logout after the work is Completed.

Collaboration architecture Diagram



The Steps from the above diagram are firstly, the respective Customer needs to Sign up and then Login then he/she will be directed to Homepage, then enter the details and the model Predicts the Loan by ML Algorithm and result will be provided as Loan can be approved / not approved. The whole Process details can be viewed by Admin.

CHAPTER 4

Implementation

INTRODUCTION

So , we all know Implementation is the most important part in Succeeding a Successful System and also giving the users confidence that the new system is effective. So, in this project I have implemented two main stages, where one is for Customers (i.e.) we and one is for admin.

In this project each program is tested individually and the dataset used for this is taken from Kaggle .The system that has been developed is accepted and proved to be satisfactory for the user.

DESIGN AND IMPLEMENTATION CONSTRAINTS

All modules are coded thoroughly based on requirements from software organization. The software is designed in such a way that the user can easily interact with the screen.

SCREENSHOTS OF PROJECT IMPLEMENTATION

MAIN PAGE:

[Home](#) [Contact](#) [Logout](#)

LOAN APPROVAL

ENTER THE BELOW DETAILS

| | | |
|----------------------|----------------------|----------------------|
| Gender | Married | Dependents |
| <input type="text"/> | <input type="text"/> | <input type="text"/> |
| Education | Self Employed | Applicant Income |
| <input type="text"/> | <input type="text"/> | <input type="text"/> |

127.0.0.1:5000/home#home

CONTACT PAGE :

CONTACT

Email

admin@gmail.com

| | | |
|--|---------------|------------------|
| Gender | Married | Dependents |
| Education | Self Employed | Applicant Income |
| Coapplicant Income | Loan Amount | Loan Amount Term |
| Credit History | | Property Area |
| <input type="button" value="Predict Probability"/> | | |

Loan cannot be approved

STEP BY STEP IMPLEMENTATION PROCESS

- As soon as we run the project on command prompt and also run the Xampp Server on the other hand it will provide us link .
- By clicking the link a new page will be visible .
- Then home page will be visible where you find a Sign up box and Sign in box.
- Firstly a respective Customer need to Sign up for next Step. As Sign up is mandatory for each and every Customer.

- After Sign up the respective Customer will be directly landed to Home page where he have to enter details for his loan approval.
- The details are also mandatory as with out details the web application cannot Predict the Output.
- Hence details are : Gender
 - Married Status
 - Dependents
 - Education

Continuation of details : Employment Status

- Applicants Income
- Co-applicants Income
- Loan Amount
- Loan Amount term
- Credit History
- Property Area

- So as shown , above are the details should be provided by the respective Customers to know the Loan Status.
- Then there will be a message generated with red highlighted text showing “LOAN CAN BE APPROVED” and “LOAN CANNOT BE APPROVED”.

There's also a admin page where he/she can see all the details of respective Customer and data is saved in MYSQL Database, which will be used for betterment of my model.

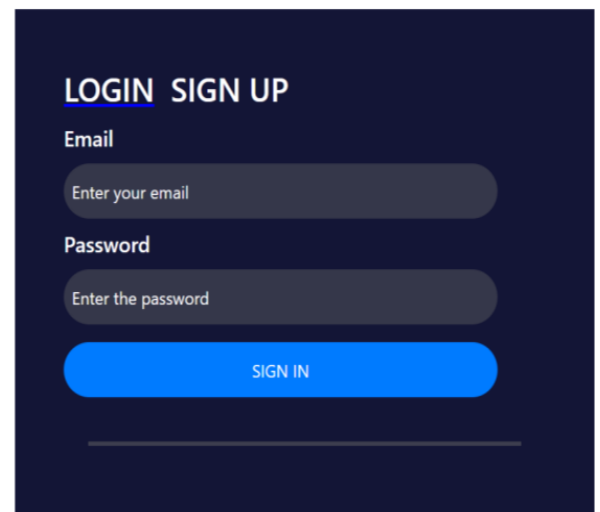
Chapter 5

Testing

SCREENSHOTS OF TESTING

LOGIN PAGE:

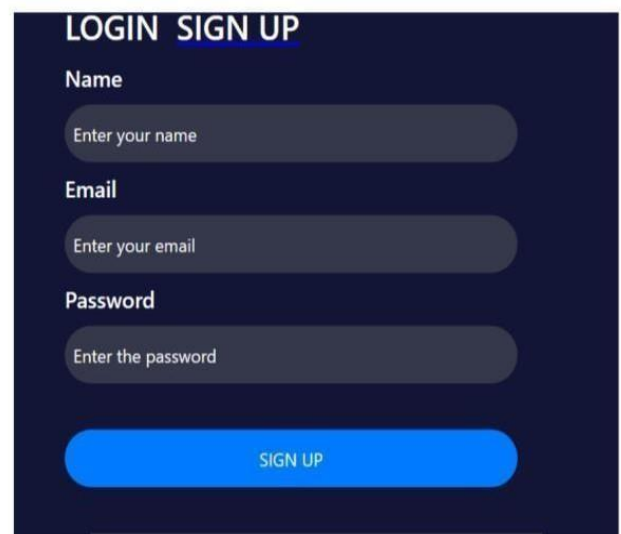
LOAN APPROVAL



A screenshot of a login page with a dark blue background. At the top, there are two links: "LOGIN" (underlined in blue) and "SIGN UP". Below these links are two input fields: "Email" with the placeholder text "Enter your email" and "Password" with the placeholder text "Enter the password". Below the password field is a blue button labeled "SIGN IN". A horizontal line is visible at the bottom of the form area.

SIGN UP PAGE:

LOAN APPROVAL



A screenshot of a sign up page with a dark blue background. At the top, there are two links: "LOGIN" and "SIGN UP" (underlined in blue). Below these links are three input fields: "Name" with the placeholder text "Enter your name", "Email" with the placeholder text "Enter your email", and "Password" with the placeholder text "Enter the password". Below the password field is a blue button labeled "SIGN UP". A horizontal line is visible at the bottom of the form area.

TEST CASE CUSTOMER:

LOAN APPROVAL

[LOGIN](#) [SIGN UP](#)

Name


test

Email

test@gmail.com

Password

....



SIGN UP

MAIN HOME PAGE:

[Home](#) [Contact](#) [Logout](#)

LOAN APPROVAL

ENTER THE BELOW DETAILS

| | | |
|----------------------|----------------------|----------------------|
| Gender | Married | Dependents |
| <input type="text"/> | <input type="text"/> | <input type="text"/> |
| Education | Self Employed | Applicant Income |
| <input type="text"/> | <input type="text"/> | <input type="text"/> |

127.0.0.1:5000/home#home

CONTACT PAGE:

CONTACT

Email

admin@gmail.com

THE RESPECTIVE USER DETAILS AND THEIR PREDICTION:

| | | |
|--------------------------------|---------------|------------------|
| Gender | Married | Dependents |
| | | |
| Education | Self Employed | Applicant Income |
| | | In \$ |
| Coapplicant Income | Loan Amount | Loan Amount Term |
| In \$ | In \$ | In days |
| Credit History | | Property Area |
| | | |
| <div>Predict Probability</div> | | |

Loan cannot be approved

CHAPTER 6

Conclusion & Future Work

So hereby , I conclude with confidence that the Logistic Regression model is extremely efficient and gives a better result when compared to other models . As, the accuracy is more in Logistic Regression model. It works correctly and fulfils all the requirements. This system properly and accurately calculates the result. It predicts the loan is approved or rejected by the loan applicant or customer very accurately.

REFERENCES :

<https://flask.palletsprojects.com/en/1.1.x/>

<https://www.pythonanywhere.com/forums/>

<https://help.pythonanywhere.com/pages/Flask/>

<https://www.kaggle.com/>

- **A survey on Model for Loan Prediction Anchal Goyal [1], Ranpreet Kaur [2] Research Scholar [1], Assistant Proffesor [2] Department of Computer Science and Engineering RIMT –IET (PTU), Mandi Gobindgarh Punjab – India.**

CODING

index.html

```
DOCTYPE html>
html lang="en">
head>

```


Admin page:

```
DOCTYPE html>
html lang="en">
head>
  <meta charset="UTF-8">
  <title>Admin</title>
  <div class="topnav">
    <a href="/admin">Home</a>
    <a style="..." href="/logout">Logout</a>
  </div>
</style>

  .topnav {
background-color: #e9e3dc;
overflow: hidden;

/* Style the links inside the navigation bar */
topnav a {
float: left;
color: #1f242e;
text-align: center;
padding: 14px 16px;
text-decoration: none;
```

Login page

```
DOCTYPE html>
html>
head>
<link rel="stylesheet" href="https://stackpath.bootstrapcdn.com/bootst
<script src="https://code.jquery.com/jquery-3.5.1.slim.min.js" integr
<script src="https://cdn.jsdelivr.net/npm/popper.js@1.16.1/dist/umd/p
<script src="https://stackpath.bootstrapcdn.com/bootstrap/4.5.2/js/bc
<style>
    .mainpg {
background-color: white;
background-size: cover;
    }
    .maincont {
background-color: rgb(19 21 54);
margin-top: 10vh;
height: 80vh;
width: 100%;
    }
    .cont {
background-color: white;
height: 50vh;
width: 100%;
    }
```

```
import os
```

```
import numpy as np
```

```
import pandas as pd
```

```
import warnings
```

```
Plotting of libraries
```

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

```
Libraries
```

```
from sklearn.preprocessing import LabelEncoder
```

```
from sklearn.model_selection import train_test_split
```

```
from sklearn.metrics import confusion_matrix
```

```
from sklearn.metrics import classification_report
```

```
from sklearn.metrics import accuracy_score
```

```
Models
```

```
from xgboost import XGBClassifier
```

```
from sklearn.tree import DecisionTreeClassifier
```

```
from sklearn.ensemble import RandomForestClassifier
```

```
from sklearn.linear_model import LogisticRegression
```

```
print(f"training set (row, col): {datatest.shape}\n\ntesting set (row, col): {datatrain.sha
```

```
training set (row, col): (367, 12)
```

```
testing set (row, col): (614, 13)
```

```
#column information
```

```
datatrain.info(verbose=True, null_counts=True)
```

```
(class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 614 entries, 0 to 613
```

```
Data columns (total 13 columns):
```

| # | Column | Non-Null Count | Dtype |
|---|-------------------|----------------|---------|
| 0 | Loan_ID | 614 non-null | object |
| 1 | Gender | 601 non-null | object |
| 2 | Married | 611 non-null | object |
| 3 | Dependents | 599 non-null | object |
| 4 | Education | 614 non-null | object |
| 5 | Self_Employed | 582 non-null | object |
| 6 | ApplicantIncome | 614 non-null | int64 |
| 7 | CoapplicantIncome | 614 non-null | float64 |

```
#this column is not needed, so I drop it for both test and train datasets
datatrain.drop('Loan_ID',axis=1,inplace=True)
datatest.drop('Loan_ID',axis=1,inplace=True)
#checking the new shapes
print(f"training set (row, col): {datatrain.shape}\n\ntesting set (row, col): {datatest
```

```
training set (row, col): (614, 12)
```

```
testing set (row, col): (367, 11)
```

```
#missing values in decsending order
datatrain.isnull().sum().sort_values(ascending=False)
```

| | |
|-------------------|----|
| Credit_History | 50 |
| Self_Employed | 32 |
| LoanAmount | 22 |
| Dependents | 15 |
| Loan_Amount_Term | 14 |
| Gender | 13 |
| Married | 3 |
| Loan_Status | 0 |
| Property_Area | 0 |
| CoapplicantIncome | 0 |

```
#converting categorical values to numbers

to_numeric = {'Male': 1, 'Female': 2,
              'Yes': 1, 'No': 2,
              'Graduate': 1, 'Not Graduate': 2,
              'Urban': 3, 'Semiurban': 2, 'Rural': 1,
              'Y': 1, 'N': 0,
              '3+': 3}

# adding the new numeric values from the to_numeric variable to both datasets
datatrain = datatrain.applymap(lambda lable: to_numeric.get(lable) if lable in to_numeric else 0)
datatest = datatest.applymap(lambda lable: to_numeric.get(lable) if lable in to_numeric else 0)

# convertind the Dependents column
Dependents_ = pd.to_numeric(datatrain.Dependents)
Dependents__ = pd.to_numeric(datatest.Dependents)

# dropping the previous Dependents column
datatrain.drop(['Dependents'], axis = 1, inplace = True)
datatest.drop(['Dependents'], axis = 1, inplace = True)
```

USE IT REAL

]

Self_E
Applica
Coapplicant
Loan_Amo
Cred
Prop
Loan
De

correlation table for a more detailed analysis:

```
#correlation table  
corr = datatrain.corr()  
corr.style.background_gradient(cmap='coolwarm').set_precision(2)
```

| | Gender | Married | Education | Self_Employed | ApplicantIncome | CoapplicantIncome | LoanAmount | Loan_Amount_Term | Cr |
|---------------|--------|---------|-----------|---------------|-----------------|-------------------|------------|------------------|-----|
| Gender | 1.00 | 0.36 | -0.05 | -0.00 | -0.06 | -0.08 | -0.11 | 0.07 | -0 |
| Married | 0.36 | 1.00 | -0.01 | 0.00 | -0.05 | -0.08 | -0.15 | 0.10 | -0 |
| Education | -0.05 | -0.01 | 1.00 | 0.01 | -0.14 | -0.06 | -0.17 | -0.07 | -0 |
| Self_Employed | -0.00 | 0.00 | 0.01 | 1.00 | -0.13 | 0.02 | -0.11 | 0.03 | 0.1 |

Accuracy conclusion:

```
score = [DT_SC, RF_SC, XGB_SC, LR_SC]  
Models = pd.DataFrame({  
    'n_neighbors': ["Decision Tree", "Random Forest", "XGBoost", "Logistic Regression"],  
    'Score': score})  
Models.sort_values(by='Score', ascending=False)
```

| | n_neighbors | Score |
|---|---------------------|----------|
| 3 | Logistic Regression | 0.832432 |
| 2 | XGBoost | 0.821622 |
| 1 | Random Forest | 0.778378 |
| 0 | Decision Tree | 0.681081 |