Smart Lender - Applicant Credibility Prediction For Loan Approval

NALAIYA THIRAN - PROJECT REPORT

PROJECT ID:PNT2022TMID00843

Submitted by

Annamalai P [211419104013]

Ashwin Kumar A [211419104027]

Chriswin Joel Anto D [211419104048]

Abishek Benjamin Raj D [211419104003]

In partial fulfillment for the award of the degree

Of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

PANIMALAR ENGINEERING COLLEGE, CHENNAI-600123.

(AN AUTONOMOUS INSTITUTION , AFFILIATED TO ANNA UNIVERSITY)

NOVEMBER 2022

PANIMALAR ENGINEERING COLLEGE, CHENNAI-600123.

(AN AUTONOMOUS INSTITUTION , AFFILIATED TO ANNA UNIVERSITY)

BONAFIDE CERTIFICATE

Certified that this project report

"Smart Lender - Applicant Credibility Prediction For Loan Approval - PNT2022TMID00843"

is the bonafide work of

Annamalai P (211419104013)

Ashwin Kumar A (211419104027)

Chriswin Joel Anto D (211419104048)

Abishek Benjamin Raj D (211419104003)

who carried out the NALAIYA THIRAN project work under the supervision.

Nidhi
INDUSTRY MENTOR
IBM

Dr.N.Pughazendi
FACULTY MENTOR
Department of CSE
Panimalar Engineering College

INDEX

CHAPTER 1

INTRODUCTION

1.1 Project Overview

The main business of practically all banks is the distribution of loans. The majority of a bank's assets are directly attributable to the revenue generated by the loans that the bank disbursed. In a banking environment, the main goal is to place one's assets in trustworthy hands. Even if many banks and financial institutions today accept loans after a lengthy process of verification and validation, there is still no guarantee that the chosen applicant is the most deserving candidate among all applicants. The main source of risk that the banking industry faces is defaults. This can be greatly mitigated by using data mining techniques like categorization and prediction.

There are mainly two objectives that are to be achieved through these techniques. They are:

1) Identification of the relevant attributes that signal the capacity of borrowers to pay back the loan.
2) Determining the best model(s) to evaluate credit risk

One of the most effective methods to accomplish this goal is the Decision Tree Induction Algorithm. By using the model that has been created, it will be possible to assess credit risk more accurately, which could improve how capital is allocated to the bank. Through this method, we can determine whether a specific application is secure or not, and the entire feature validation process is automated using machine learning.

1.2 Purpose

The aim of this Project is to provide a quick, immediate and easy way to choose the deserving applicants. It can provide special advantages to the bank. The Loan Prediction System can automatically calculate the weight of each feature taking part in loan processing and on new test data the same features are processed with respect to their associated weight. A time limit can be set for the applicant to check whether his/her loan can be sanctioned or not. Jumping to certain applications allows for priority-based review using the Loan Prediction System. The governing authority of a bank or financial corporation is the only audience for this paper. The entire prediction process is carried out in private, and no stakeholders can affect how it is processed. Results for a certain Loan Id can be sent to other bank departments, allowing them to respond to applications in the most effective way possible. This facilitates the completion of other formalities by all other departments**.**

CHAPTER 2

LITERATURE SURVEY

2.1 Existing Problem

One of the most significant elements affecting the economy and financial health of our nation is the credit system controlled by the banks. Additionally, one of the primary roles of the banking industry is to manage credit risk. But one of the most challenging duties for every bank is the forecast of loan defaulters.

2.2 Problem Definition

When banks must give loans to consumers who are in need of money, this issue arises. However, by anticipating loan defaulters, banks may be able to cut their loss by decreasing their non-profit assets. Therefore, bank staff must find a means to carry out this task so that sanctioned loans may be recovered without incurring any losses.

As they are highly important and helpful in the forecast of these sorts of data, machine learning techniques may be applied to do such classifications of the credit defaulters. The ML algorithms will be trained and evaluated on the pre processed dataset. Finally, the application chooses and uses the best model.

CHAPTER 3

IDEATION AND PROPOSED SOLUTION

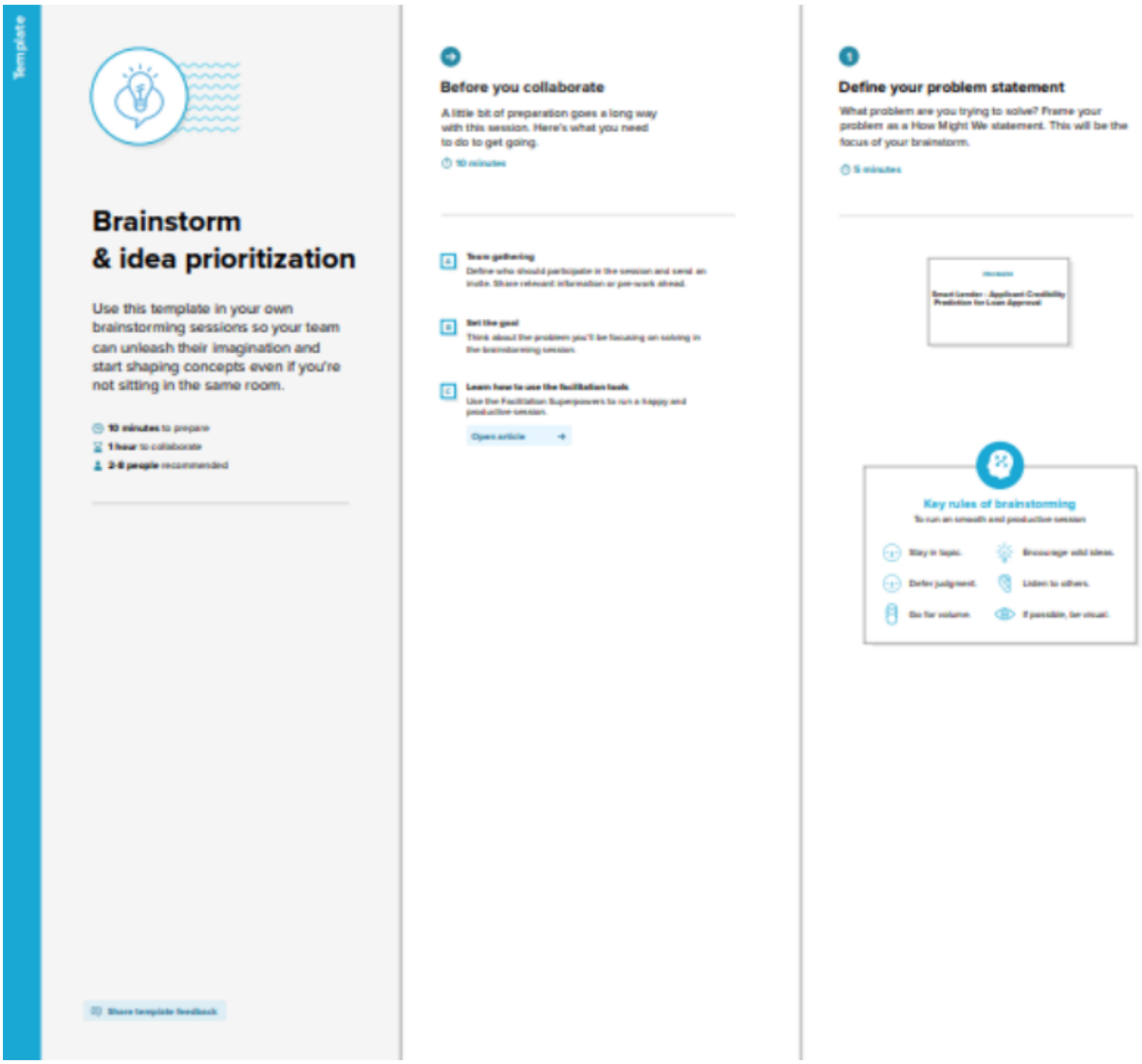
3.1 Empathy Map

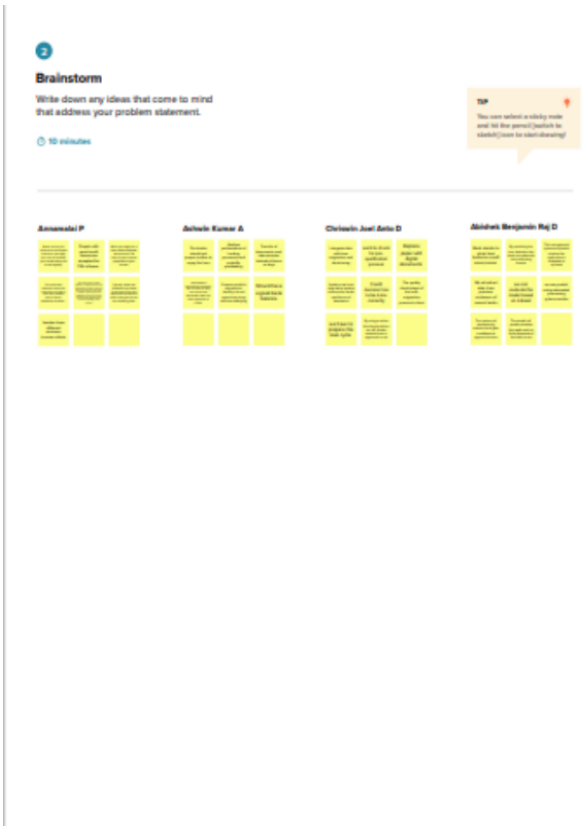
The primary purpose of the empathy map is to bridge the understanding of the user and developer. Figure 3.1 represents the empathy map for the Smart lender applicant credibility prediction for loan approval.

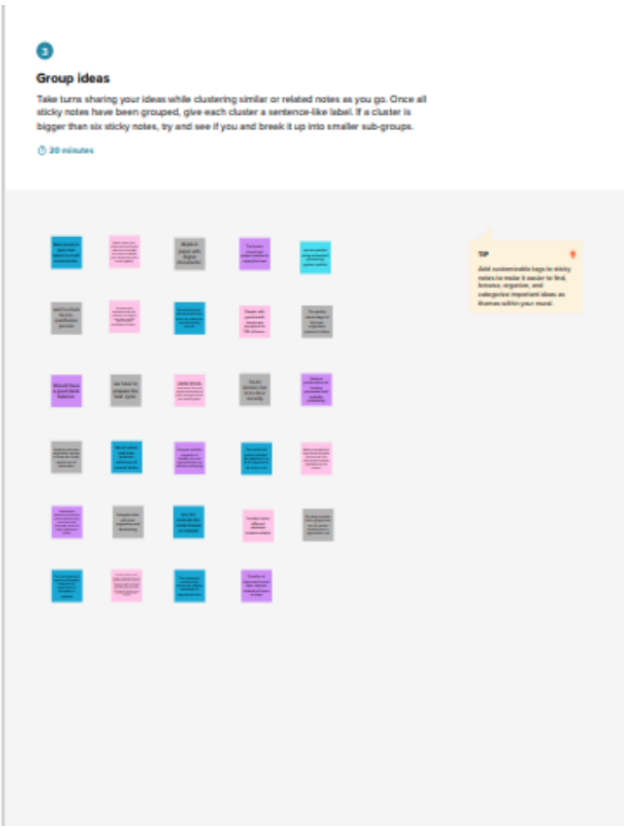
Figure 3.1 – Empathy Map

3.2 Ideation and Brainstorming

This is often the most exciting stage in a project, because during Ideation and brainstorming, the aim is to generate a large quantity of ideas that the team can then filter and cut down into the best, most practical, or most innovative ones to inspire new and better design solutions and products. Figure 3.2 shows the stages of ideation and brainstorming for the machine learning based vehicle performance analyzer.




2
Brainstorm
Write down any ideas that come to mind that address your problem statement.
10 minutes
TIP
Aanandalal P
Ashwin Kumar A
Chrisolin Joel Anto D
Abishek Benjamin Raj D


3
Group ideas
Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.
20 minutes
TIP

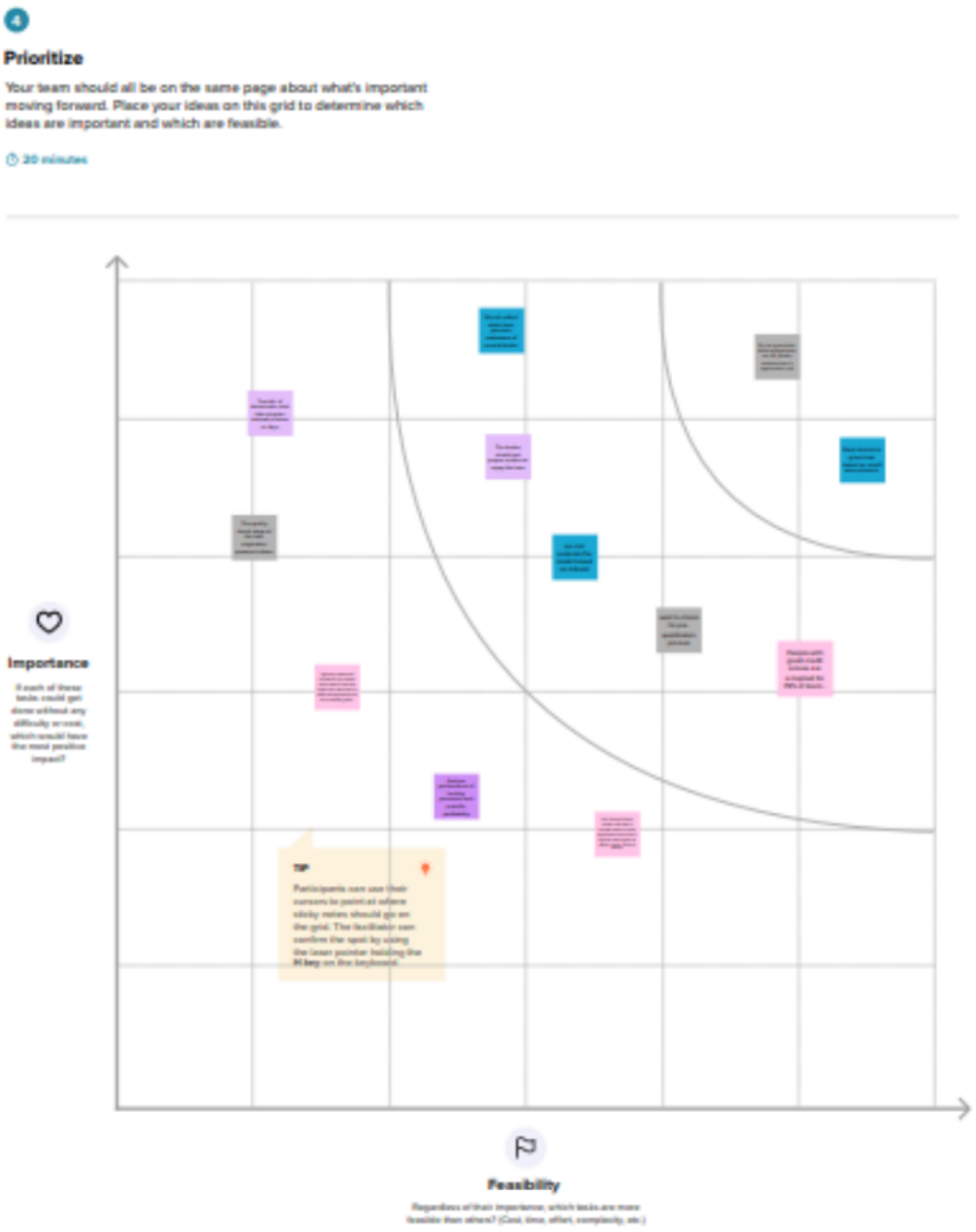


Figure 3.2 – Ideation & Brainstorming

3.3 Proposed Solution

S.No.	Parameter	Description
1	Problem Statement (Problem to be solved)	People approach banks to fulfill their needs via bank loans. This practice has been increasing day by day across the globe, especially for business, education, marriage, agriculture, etc. But several people take advantage and misuse the facilities given by the banks, so banks realize that retaining customers and preventing fraud should be a strategic policy for healthy competition. One of the important factors affecting the economic and financial condition of our country is the credit system operated by banks. Bank credit risk evaluation is recognized in banks all over the world. There are various methods used for risk level calculation. As we know, credit risk assessment is very crucial. Every day, many people apply for loans, but not all of these applicants are trustworthy, and not all can be approved. We heard about many cases where people could not repay the loan amount, causing the bank to suffer a huge loss. So, the main source of income of any bank from its customers is their credit line. By using applied data science techniques and machine learning algorithms, we will check the credit score of the person and predict whether the loan is approved or not. This makes the loan approval process very easy

2	Idea / Solution description	● Customer property documentation must be submitted, and the customer must accept the bank's terms and conditions. ● It will Provide captcha security. ● Automatic interest rate and repayment schedule determination based on loan amount. ● A machine learning algorithm will be utilized to build a reliable and effective ● A computer programme that determines whether a person is qualified for a loan based on a variety of factors (including gender, education level, number of dependents, marital status, employment, credit score, loan amount, and others). ● In order to increase the accuracy of predictive data and data mining applications, ensemble modeling is a technique that involves running two or more related but distinct models and then merging the results into a single score. ● Ensemble approaches in machine learning combine a number of algorithms to improve prediction accuracy. ● KNN, Decision tree, Random forest, and Xgboost are some examples of the various ML models that can be employed.. ● Various effective machine learning methods can be used to forecast a customer's loan eligibility.

3	Novelty / Uniqueness	● Secures data by offering. The customer's information won't be given to a third party. ● Instant Loan approval status. ● The main benefits of Ensemble models are: Better Forecasting ,More Constant model , Better results ,Reduces error. All these factors make the project unique.
4	Social Impact / Customer Satisfaction	● Nowadays bank play a vital role in the market economy. The success or failure of an organization largely depends on the industry's ability to evaluate credit risk. Banks have many products to sell in our banking system, but their main source of income is their credit lines. ● The bank can minimize its Non-Performing Assets by forecasting loan defaulters. ● Secure storage of customer details. ● Easy and fast loan approval process for the customer.
5	Business Model (Revenue Model)	● The bank can minimize its Non-Performing Assets by forecasting loan defaulters. ● Automation of the loan approval processes opens new financing opportunities for small businesses and individuals. ● They can charge the processing fees and service fees from customers. ● They can generate revenue by referencing.

6	Scalability of the Solution	● Any type of customer can predict their loan approval without any discrimination. ● This system is easily scalable and efficient. ● It can be provided as software as a service. Both borrower and Lender can use this

3.4 Problem Solution Fit

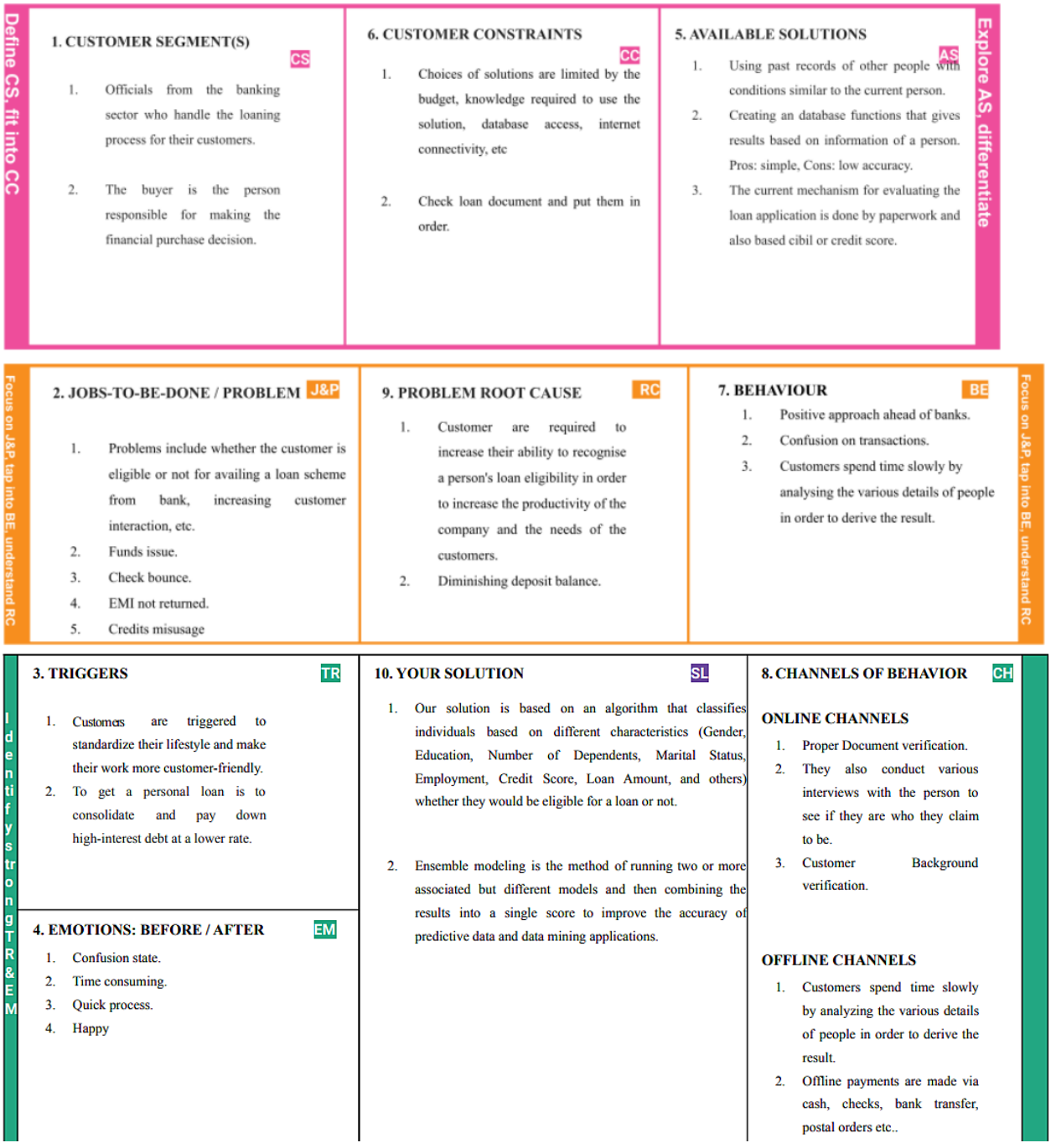


Figure 3.4 – Problem Solution Fit

CHAPTER 4

REQUIREMENT ANALYSIS

4.1 Functional Requirements

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Requirement	To check the loan eligibility using the credit score, prediction for loan approval
FR-2	User Confirmation	Through one time verification and using captcha etc..,.
FR-3	Profile Updation	The user can update their profile when there is need to add any add-ons to it
FR-4	User Registration	The user gets login/signup using Gmail account or by using mobile number
FR-5	User Authentication	The OTP/verification code the user gets authenticated and OTP is used for mobile number registration
FR-6	Feedback Evaluation	The user provided feedbacks are used for evaluation of app performance and Updation is made

		over that

Table 4.1 – Functional Requirements

Non-functional Requirements:

Following are the non-functional requirements of the proposed solution.

FR No.	Non-Functional Requirement	Sub Requirement (Story / Sub-Task)
NFR-1	Usability	This application is mainly used to analyse cibil score and predict the eligibility for users to avail for loan approval by following community guidelines
NFR-2	Security	It uses OTP and verify code verification for each user and uses and hybrid security features over internet to safely maintain the updated documents of user

NFR-3	Reliability	Maintaining the app up to date for reliant future, durability and efficiency of the mobile app by releasing patch fix and software updates
NFR-4	Performance	It has a user-friendly interface and can check multiple persons cibil score parallels irrespective of server traffic. It stores the data collected over in an efficient database
NFR-5	Availability	It is platform independent and it is available where the users are able to wish it want to be. Depending upon the user requirements all services get offered
NFR-6	Scalability	Provides accurate prediction for user eligibility by using highly efficient algorithms and testing all the documents uploaded by the user at highly efficient rate

Table 4.2 – Non-Functional Requirements

CHAPTER 5

PROJECT DESIGN

5.1 Dataflow Diagram

A Data Flow Diagram (DFD) is a traditional visual representation of how information flows within a system. A neat and clear DFD can thus depict the right amount of the system requirements graphically. It not only shows how data enters and leaves the system, but also what changes the information and where the data is stored. Figure 5.1 represents the DFD for the given project.

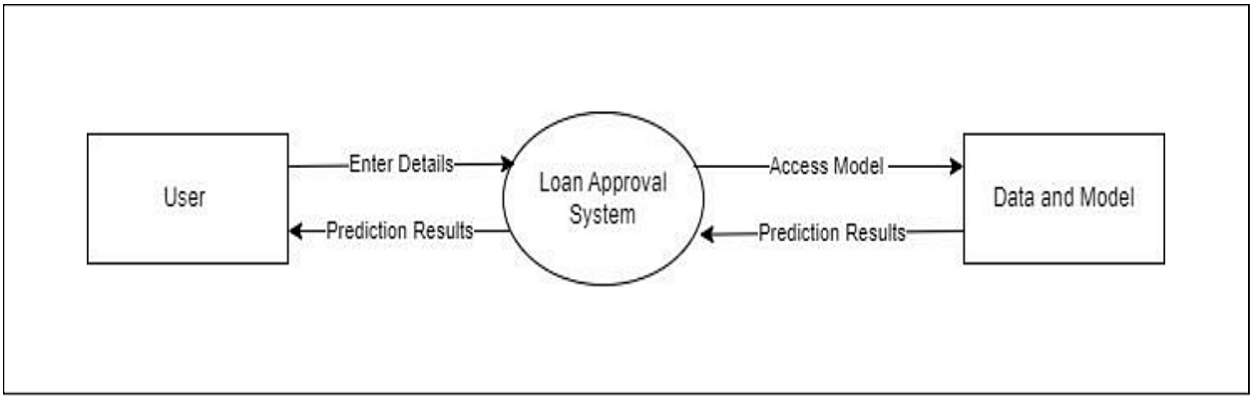


Figure 5.1 – Dataflow Diagram

5.2 Technical Architecture

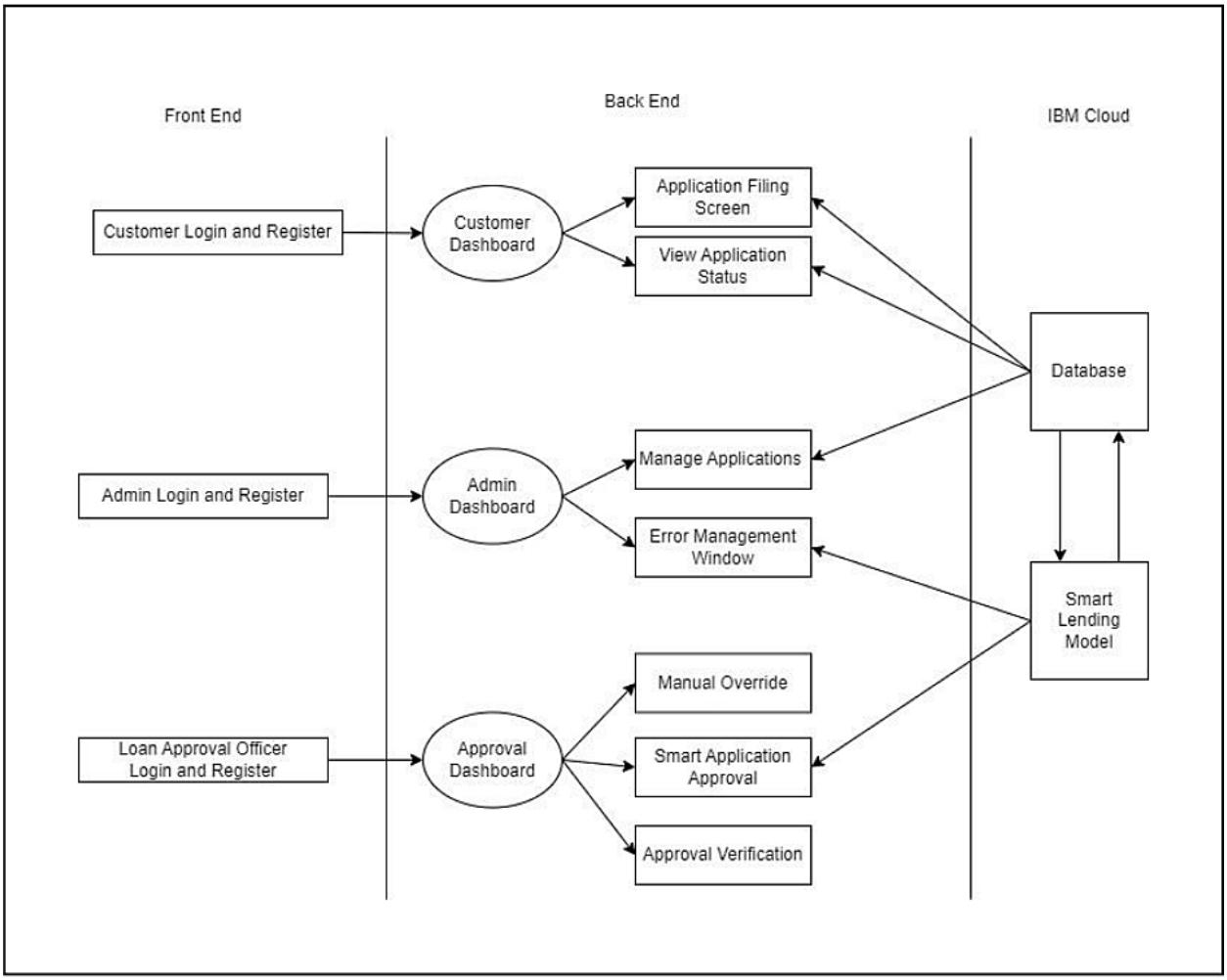


Figure 5.2 Technical Architecture

5.2.1 Component and Technologies

S.No	Component	Description	Technology
1	User Interface	How user interacts with application e.g. Web UI, Mobile App, Chatbot etc	HTML, CSS, JavaScript / Flask
2	Application Filing	Converting the csv file to python object	Python
3	View Application Status	Pre-Processing and normalizing the data to get accurate results	Python
4	Manage Applications	Database Service on Cloud	IBM DB2, IBM Cloudant etc.
5	Error Management Window	File storage requirements	IBM Block Storage or Other Storage Service or Local Filesystem
6	Manual Override	Long short-term memory (LSTM) is an artificial neural network. Unlike standard feedforward neural networks, LSTM has feedback connections GRU Gated recurrent units is like a long short-term memory (LSTM) with a forget gate, but has fewer parameters than LSTM, as it lacks an output gate.	Object Recognition Model, etc

7	Smart Application Approval	Application Deployment on Local System / Cloud Local Server Configuration: 2.5Ghz processor, 8GB RAM Cloud Server Configuration: 4 GB GPU	Local, Cloud Foundry, Kubernetes, etc.
8	Application Verification	Verification of application and documents provided by applicant	JavaScript
9	File Storage	Network File System(NFS)	IBM Cloud File Storage
10	Machine Learning	Automated decision making for loan approval and giving details for decision	XGBoost, Random Forest
11	Infrastructure (Sever/Cloud)	Default	Flask

Table 5.2.1 – Components and Technologies

5.2.2 Application Characteristics

S.No	Characteristics	Description	Technology
1	Open-Source Frameworks	Flask is used to host the website. Scikit, NumPy and TensorFlow are all open-source python machine learning frameworks	Scikit, NumPy, TensorFlow, Flask
2	Security Implementations	OpenSSL is a program and library that supports many different cryptographic operations, including : symmetric key encryption, public/private key pair generation, pubic key encryption, hash functions	OpenSSL Encryption

3	Scalable Architecture	Since the application servers can be deployed on many machines. Also, the database doesn't make longer connections with every client-it only requires connections from a smaller number of applications servers. It improves data integrity	3 Tier Architecture
4	Availability	Decentralized storage and distribution along with web application approach make the service highly available	IBM Cloud File Storage, MySQL Online
5	Performance	Long term header expiration Cacheable AJAX Cookie Free Domain Compress gzip components	AJAX, CDN

Table 5.2.2 – Application Characteristics

5.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	As a user I can enter Gmail and set a password	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can get a code for confirmation	High	Sprint-1

		USN-3	Registration as a user can be confirmed using OTP or verification code	Aa a user can get OTP or verification code	Low	Sprint-1
	Login	USN-4	Users can login to web/mobile interface by storing or using the registered login credentials	Able to login	Medium	Sprint-1
		USN-5	As a user, I can log into the application by entering email & password	Can be able to login using Gmail	Medium	Sprint-1
	Dashboard	USN-6	As a user, I should be able to login the profile or status dashboard	Able to access dashboard account	Medium	Sprint-2

Customer (Web user)		USN-7	Checks the user feedback and provide essential technical support	Access the account/able to access the dashboard		Sprint-2
Loan Approval Executive	Automated analysis of score	USN-8	As a loan approval officer, I can make decisions by checking and monitoring all the feeded applications and getting to a prediction	Get a decision for loan prediction based on the details provided in the loan application	High	Sprint-3
		USN-9	As a admin cibil score which represents credit history plays major role in analysis	Cibil score/credit history place major role	High	Sprint-3

Admin	Login/Register	USN-10	As an admin I should be able to login with the unique email and password	Able to logged in	High	Sprint-4
	Dashboard	USN-11	As an admin I need the access of full authority towards the dashboard	Access the dashboard	Medium	Sprint-4

Table 5.3 – User Stories

CHAPTER 6

PROJECT PLANNING AND SCHEDULING

Sprint	Functional Requirement (Epic)	User Story Number	User Story/ Task	Story Points	Priority	Team Members
Sprint-1	Dashboard	USN - 1	As a user, I need to read the details given in the form.	2	High	Annamalai, Ashwin, Abishek, Anto
Sprint-2	Dashboard	USN - 1	As a user, after reading all the information I need to click on predict	2	Low	Annamalai, Ashwin, Abishek, Anto
Sprint-3	Details enter page	USN - 2	As a user, I need to enter all the required details in the corresponding fields.	2	Medium	Annamalai, Ashwin, Abishek, Anto
Sprint-3	Validation	USN - 3	Validates the truthfulness of entered information	1	High	Annamalai, Ashwin, Abishek, Anto

Sprint-4	Prediction	USN - 4	Prediction based on the details given by the user is done.	2	Medium	Annamalai, Ashwin, Abishek, Anto
Sprint-4	Display Result	USN - 5	Final resulting page is displayed.	2	Low	Annamalai, Ashwin, Abishek, Anto

Table 6.1 - Sprint Planning

6.2 Sprint Delivery Schedule

Sprint	Story Points	Duration (Days)	Sprint Start Date	Story Points Completed	Sprint Release Date
Sprint 1	20	6	24 Oct 2022	20	29 Oct 2022
Sprint 2	20	6	31 Oct 2022	16	05 Oct 2022
Sprint 3	20	6	07 Nov 2022	18	12 Nov 2022
Sprint 4	20	6	14 Nov 2022	15	19 Nov 2022

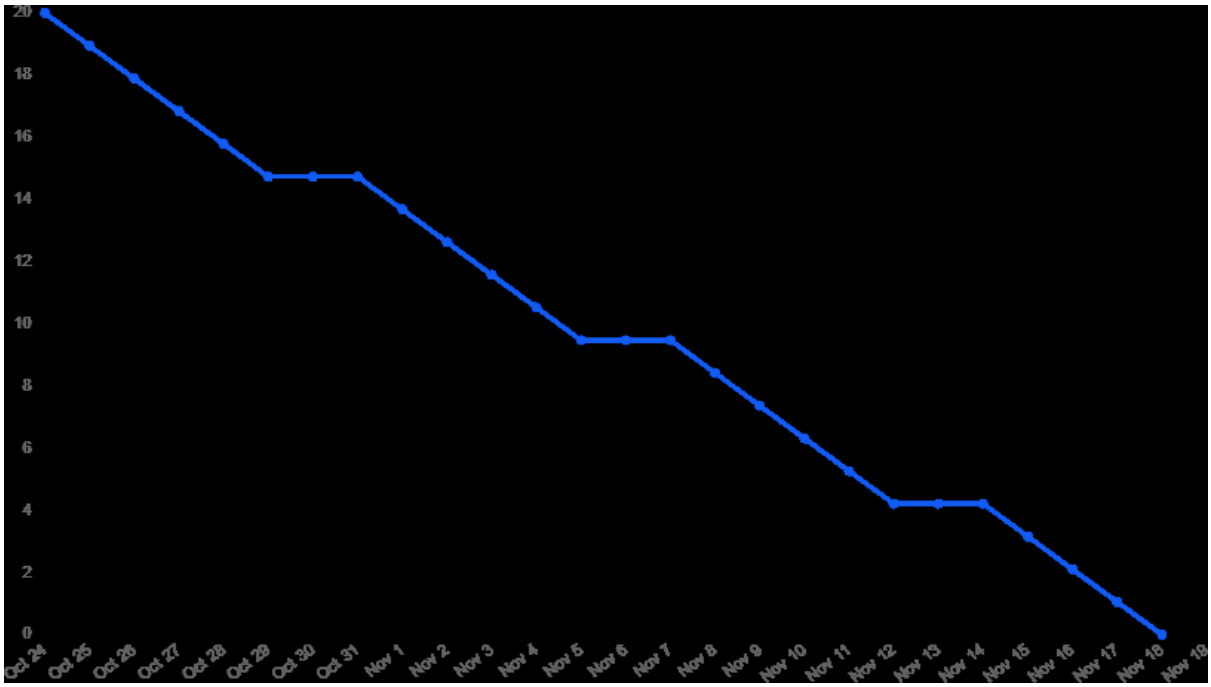
Table 6.2 - Sprint Delivery Schedule

6.3 Reports from JIRA

Velocity: Imagine we have a 10-day sprint duration, and the velocity of the team is 20 (points per sprint). Let's calculate the team's average velocity (AV) per iteration unit (Story points per day):

$$AV = \frac{sprint\ duration}{velocity} = \frac{20}{10} = 2$$

Burndown Chart: A burndown chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such as Scrum. However, burn down charts can be applied to any project containing measurable progress over time.



CHAPTER 7

CODING AND SOLUTION

7.1 Feature 1

FR No.,	Feature	Description
FR - 1	Applicant credibility prediction	Showing the predicted result of applicant credibility

Table 7.1 - Description for Feature 1

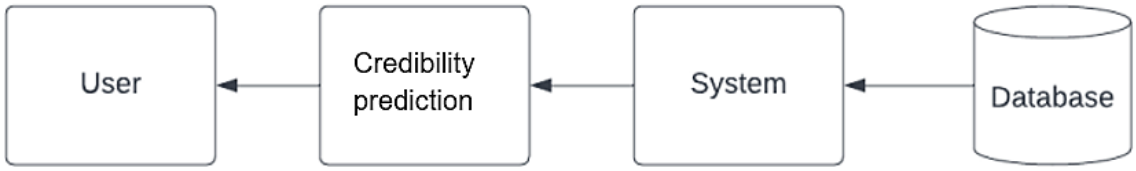


Figure 7.1 – Data Flow Diagram for Feature 1

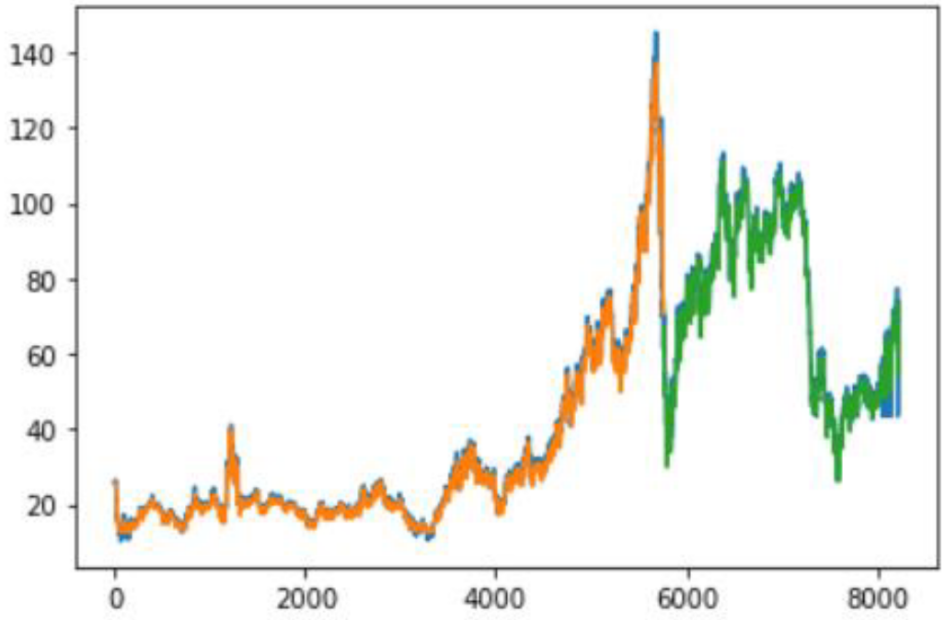
CHAPTER 8

TESTING

8.1 Test Cases

An applicant's details make up the test cases. The test examples are given to the model, and the predicted and actual loan status are compared. The model's performance is examined using the loss metric. Figure 8.1 displays the outcomes of the tests. The bottom blue line displays the actual loan standing The prediction made using the training data is shown by the orange lines. The forecast based on testing data is represented by the green line.

Figure 8.1 - Test Cases Run



8.2 User Acceptance Testing

This report shows the number of test cases that have passes, failed and untested.

Section	Total Cases	Not Tested	Fail
Model Prediction	8	0	0
Front End	8	0	0

CHAPTER 9

RESULTS

9.1 Performance Metrics

Root Mean Square Error (RMSE) is the performance metric used to evaluate the model. Both the prediction on the training data and the testing data are used to calculate RMSE. The model is more accurate when the RMSE score is lower. Figure 9.1 displays the results of the RMSE. It is evident from figure 9.1 that the RMSE are relatively low. This shows that the model is operating more effectively and that the forecasts are rather accurate.

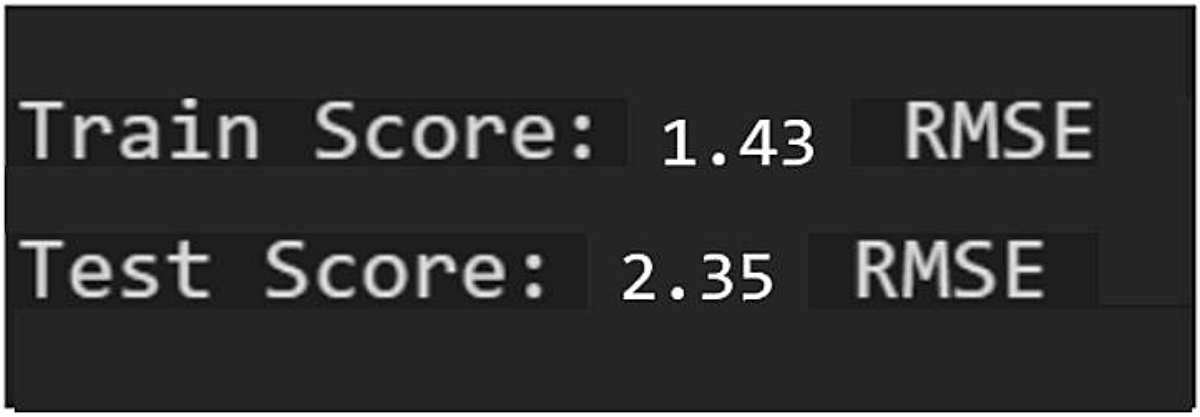


Figure 9.1 - Performance Metrics

CHAPTER 10

PROS AND CONS

10.1 Pros

- The application we have created is user friendly
- This application is flexible as user can predict the credibility of the applicant easily.

10.2 Cons

As the data provided by the user to model for prediction is true information of the applicant means the prediction result is inconsistent.

CHAPTER 11

CONCLUSION

The XGBoost model is used by the prediction system. Both the training and testing data have relatively low RMSE scores. This demonstrates the model's high level of accuracy. The flask framework is used to deliver a website, allowing users to interact with the model. The user can predict the applicant's credibility with this aid. They can thereby collect the sanctioned loans with the fewest losses possible. They don't have to spend much work making loan acceptance decisions; they can leave that to the applicant. The application's outcomes satisfy the clients (bankers).

CHAPTER 12

FUTURE WORKS

Currently, the model cannot provide consistent results for inaccurate data. Therefore, we intend to provide a feature that allows users to verify the accuracy of the applicant data within the application itself. The outcome of the prediction is therefore more trustworthy.

CHAPTER 13

APPENDIX

13.1 Source Code

```
from flask import render_template,Flask,request
import numpy as np
import pickle
import requests

# # NOTE: you must manually set API_KEY below using information retrieved from your IBM Cloud
account.
# API_KEY = "QCFx9k6096PpFRR8lVJGTkcy-XFNT4pj1SlbTvglDMw7"
# token_response = requests.post('https://iam.cloud.ibm.com/identity/token', data=
{"apikey":API_KEY, "grant_type": 'urn:ibm:params:oauth:grant-type:apikey'})
# mltoken = token_response.json()["access_token"]

# header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' + mltoken}

API_KEY = "QCFx9k6096PpFRR8lVJGTkcy-XFNT4pj1SlbTvglDMw7"
token_response = requests.post('https://iam.cloud.ibm.com/identity/token', data=
{"apikey":API_KEY, "grant_type": 'urn:ibm:params:oauth:grant-type:apikey'})
mltoken = token_response.json()["access_token"]

header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' + mltoken}
token_response = requests.post('https://iam.cloud.ibm.com/identity/token', data=
{"apikey":API_KEY, "grant_type": 'urn:ibm:params:oauth:grant-type:apikey'})
mltoken = token_response.json()["access_token"]


app= Flask(__name__, template_folder='Template')
scale = pickle.load(open('D:\Downloads\IBM-Project-3414-1658559917-main/model.pkl','rb'))

@app.route('/')
def home():
    return render_template('index.html')
@app.route('/predict.html')
def formpg():
    return render_template('predict.html')
@app.route('/submit',methods = ['POST'])
def predict():
    gender,married,depend,education,self_emp,applicant_income,co_income,loan_amount,loan_term,
    credit_history,property_area = [x for x in request.form.values()]
    if gender == 'Male':
        gender = 1
    else:
        gender = 0

    if married == 'Yes':
        married = 1
    else:
        married = 0

    if education == 'Graduate':
        education = 0
    else:
        education = 1

    if self_emp == 'Yes':
        self_emp = 1
    else:
        self_emp = 0
```

```
if education == 'Graduate':
    education = 0
else:
    education = 1

if self_emp == 'Yes':
    self_emp = 1
else:
    self_emp = 0

if depend == '3+':
    depend = 3

applicant_income = int(applicant_income)
applicant_income = np.log(applicant_income)
loan_amount = int(loan_amount)
loan_amount = np.log(loan_amount)

if credit_history == 'Yes':
    credit_history = 1
else:
    credit_history = 0

if property_area == 'Urban':
    property_area = 2

elif property_area == 'Rural':
    property_area = 0
else:
    property_area = 1

features = [[gender,married,education,self_emp,applicant_income,co_income,loan_amount,
loan_term,credit_history,property_area]]
con_features = np.array(features)

sf = con_features.tolist()

payload_scoring = {"input_data": [{"fields": ['gender','married','education','self_emp',
'applicant_income','co_income','loan_amount','loan_term','credit_history','property_area'],
"values": sf}]}

response_scoring = requests.post('https://eu-de.ml.cloud.ibm.com/ml/v4/deployments/
c5826276-fada-4224-8e4d-4bc5e2050b56/predictions?version=2022-11-18', json=payload_scoring,
headers={'Authorization': 'Bearer ' + mltoken})
prediction = response_scoring.json()

print(prediction)
predict = prediction['predictions'][0]['values'][0][0]



#prediction = model.predict(scale_features)
```

```
    #prediction = model.predict(scale_features)
    if predict == 0:
        return render_template('submit.html', prediction_text ='You are eligible for loan')
    else:
        return render_template('submit.html',prediction_text = 'Sorry you are not eligible for
        loan')


if __name__ == "__main__":
    app.run(debug=True)
```

13.2 GitHub and Project Demo Link

Github link: https://github.com/IBM-EPBL/IBM-Project-3414-1658559917

Project demo link: https://drive.google.com/file/d/1rDzvcslOlITlK31dBEqUGEE0ofVs-kSv/view?usp=share_link