Project Report Format

INTRODUCTION
- Project Overview
- Purpose

LITERATURE SURVEY
- Existing problem
- References
- Problem Statement Definition

IDEATION & PROPOSED SOLUTION
- Empathy Map Canvas
- Ideation & Brainstorming
- Proposed Solution
- Problem Solution fit

REQUIREMENT ANALYSIS
- Functional requirement
- Non-Functional requirements

PROJECT DESIGN
- Data Flow Diagrams
- Solution & Technical Architecture
- User Stories

PROJECT PLANNING & SCHEDULING
- Sprint Planning & Estimation
- Sprint Delivery Schedule
- Reports from JIRA

CODING & SOLUTIONING (Explain the features added in the project along with code)
- Feature 1
- Feature 2
- Database Schema (if Applicable)

TESTING
- Test Cases
- User Acceptance Testing

RESULTS
- Performance Metrics

ADVANTAGES & DISADVANTAGES

CONCLUSION

FUTURE SCOPE

APPENDIX

Source Code

GitHub & Project Demo Link

INTRODUCTION

1.1 Project overview

One of the most important factors which affect our country's economy and financial condition is the credit system governed by the banks. The process of bank credit risk evaluation is recognized at banks across the globe. "As we know credit risk evaluation is very crucial, there are a variety of techniques used for risk level calculation. In addition, credit risk is one of the main functions of the banking community

1.2 Purpose

The prediction of credit defaulters is one of the difficult tasks for any bank. But by forecasting the loan defaulters, the banks definitely may reduce their loss by reducing their non-profit assets, so that recovery of approved loans can take place without any loss and it can play as the contributing parameter of the bank statement. This makes the study of this loan approval prediction important. Machine Learning techniques are very crucial and useful in the prediction of these types of data.

The aim of this project is to predict the probability of credit default based on credit card owner's characteristics and payment history.By finding the credit

defaulters, the bank employees can decide whether to provide loan for the candidates.

LITERATURE SURVEY

1.3 Existing problem

The prediction of credit card fraud defaulters is one of the difficult tasks for any bank.In addition,credit card risk is one the main functions of the banking community.The prediction of credit defaulters is one of the difficult tasks for any bank. But by forecasting the loan defaulters, the banks definitely may reduce their loss by reducing their non-profit assets, so that recovery of approved loans can take place without any loss and it can play as the contributing parameter of the bank statement. This makes the study of this loan approval prediction important. Machine Learning techniques are very crucial and useful in the prediction of these types of data.

1.4 References

https://ieeexplore.ieee.org/document/9432308

https://ieeexplore.ieee.org/abstract/document/8123782

https://ieeexplore.ieee.org/document/9121114

https://ieeexplore.ieee.org/document/8776942/authors#authors

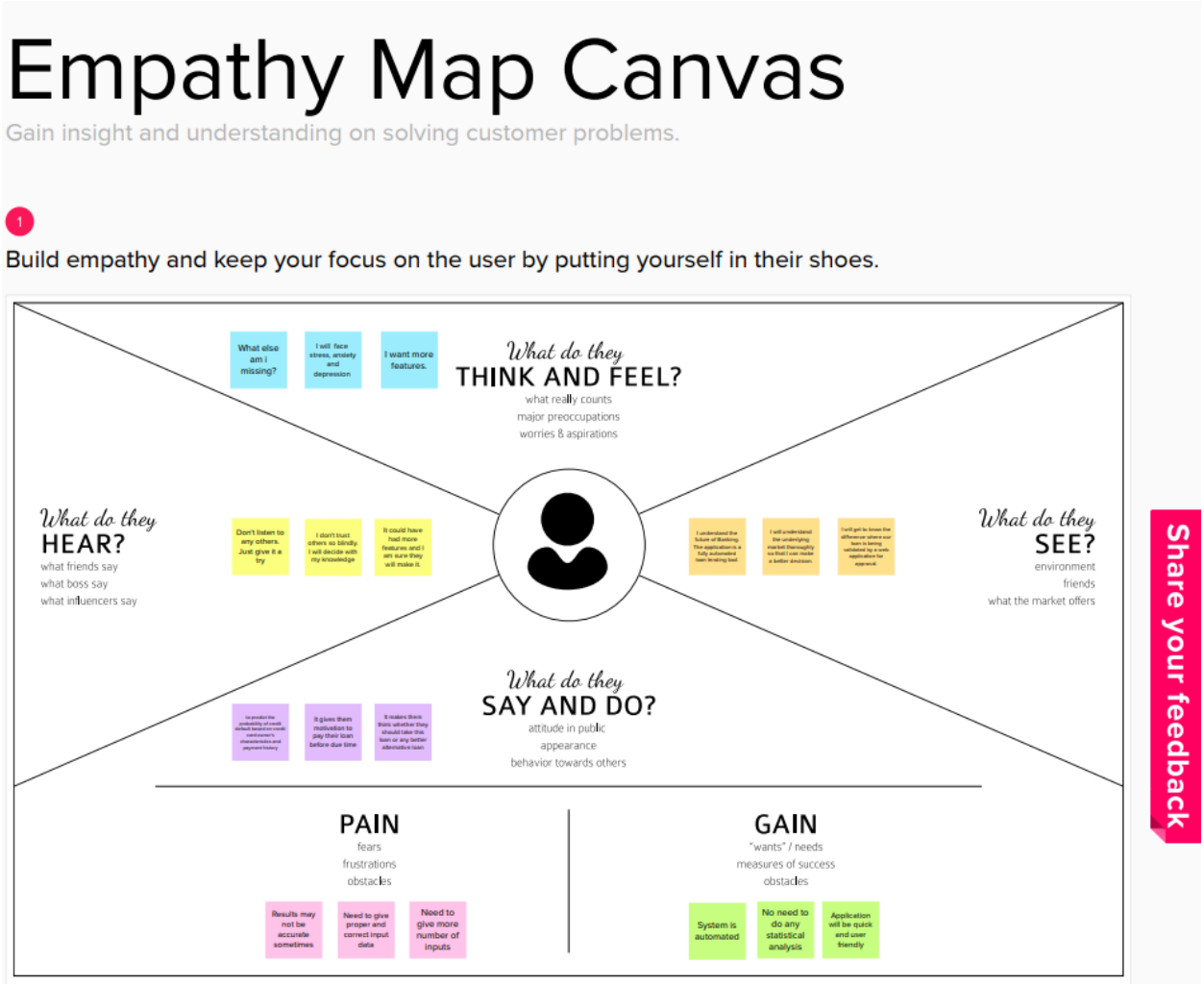
1.5 Problem Statement Definition

The credit system controlled by banks is one of the most significant variables that has an impact on the economy and financial situation of our nation. Credit risk management is another one of the banking industry's key responsibilities.

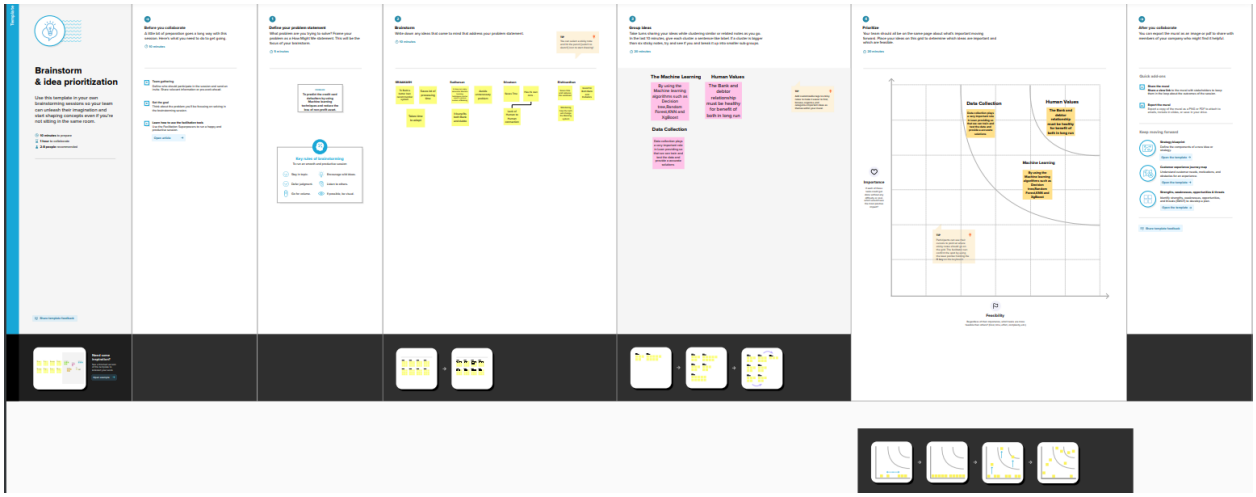
The aim of this project is to predict the probability of credit default based on credit card owner's characteristics and payment history.By finding the credit defaulters, the bank employees can decide whether to provide loan for the candidates.

IDEATION & PROPOSED SOLUTION

1.6 Empathy Map Canvas



1.7 Ideation & Brainstorming



1.8 Proposed Solution

Project Design Phase-I
Proposed Solution Template

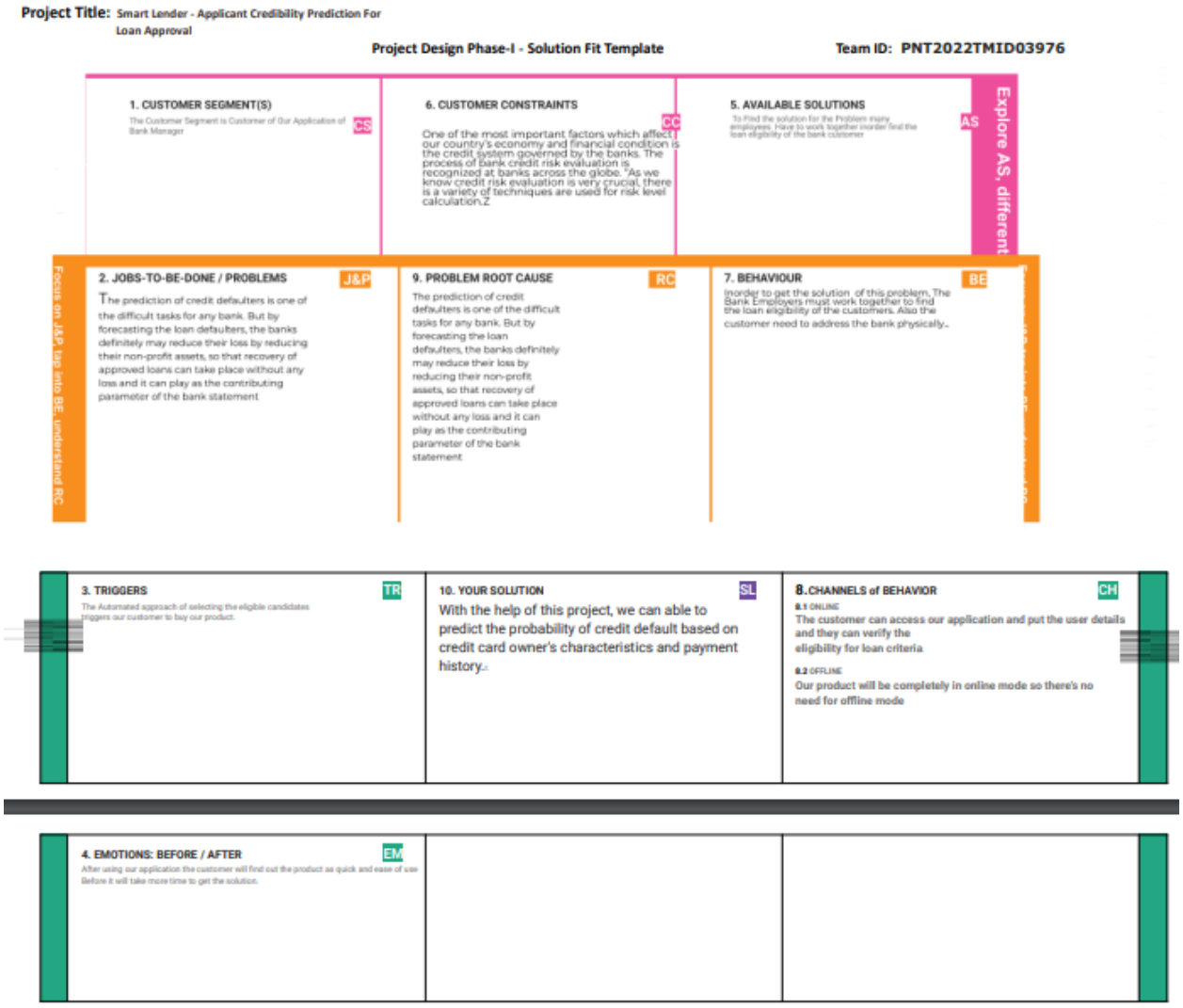
Date	19 September 2022
Team ID	PNT2022TMID03976
Project Name	Project - xxx
Maximum Marks	2 Marks

Proposed Solution Template:

Project team shall fill the following information in the proposed solution template.

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	The prediction of credit defaulters is one of the difficult tasks for any bank. But by forecasting the loan defaulters, the banks definitely may reduce their loss by reducing their non-profit assets, so that recovery of approved loans can take place without any loss and it can play as the contributing parameter of the bank statement.
2.	Idea / Solution description	The Solution is to build an application by which we can predict the probability of credit default based on credit card owner's characteristics and payment history.
3.	Novelty / Uniqueness	This Project is implemented using the Machine Learning Algorithms with improved accuracy of finding the results
4.	Social Impact / Customer Satisfaction	The Social impact is Ease of use,Economically friendly and quick in decision making etc.
5.	Business Model (Revenue Model)	The Revenue can be achieved by a better Market adopting this technology with better features.
6.	Scalability of the Solution	This application can be further improved by adding new features and thus provide scalability.

1.9 Problem Solution fit



REQUIREMENT ANALYSIS

1.10 Functional & non functional requirements

Project Design Phase-II
Solution Requirements (Functional & Non-functional)

Date	30 October 2022
Team ID	PNT2022TMID03976
Project Name	Smart Lender - Applicant Credibility Prediction for Loan Approval
Maximum Marks	4 Marks

Functional Requirements:

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	Web UI	A user interface hosted in web to easily interact with the applicants.
FR-2	Uploading Loan Approval Documents	Uploading the loan application document to verify applicant credibility.
FR-3	Uploading Applicants ID proof Documents	Applicants ID proofs like copy of PAN card, Address proof document and Identity proof documents are uploaded for checking the applicants credibility.
FR-4	Database Management	Applicants loan application data will be saved in the database and will be used for future reference and all the other documents are deleted after processing.
FR-5	Reporting	Predicting applicants credibility for loan approval using given data and generatingthe report for the applicant.
FR-6	Internet Connectivity	User should have a stable internet connection to access the functionality of our project via web application.

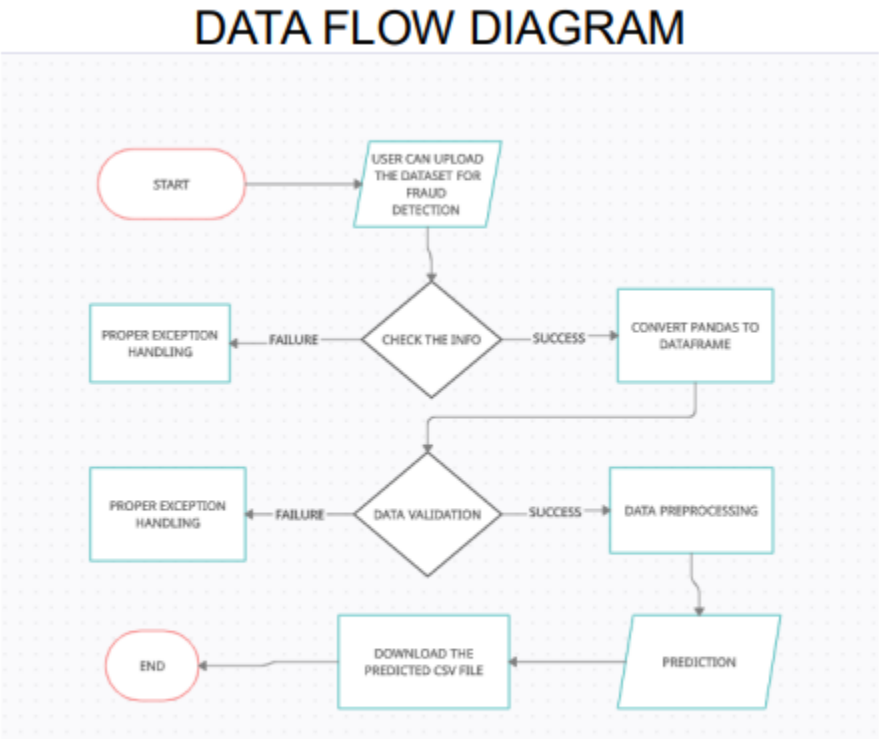
Non-functional Requirements:

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

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The whole system can be accessed through web application. Hence it is very easy to use.
NFR-2	**Security**	The user data will be stored in a database so the user's data is secured.
NFR-3	**Reliability**	As the data is stored in a database, the data cannot be manipulated externally so it is highly reliable.
NFR-4	**Performance**	Application effectively compares user given parameters with the required dataset. Hence performance would be considerably good.
NFR-5	**Availability**	Application is active all the time so the user can avail it anytime.
NFR-6	**Scalability**	Application can be used in any kind of operating system either in small or large OS so the scalability is very high.

PROJECT DESIGN

1.11 Data Flow Diagrams

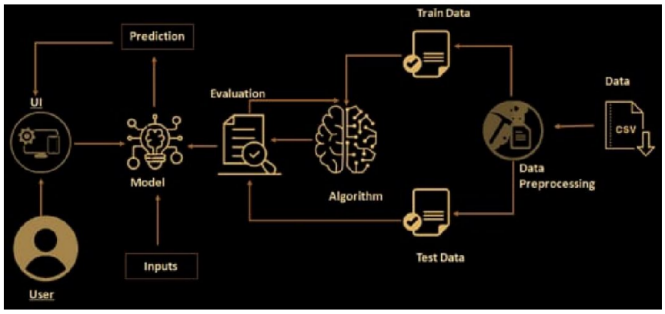


1.2 Solution & Technical Architecture

Project Design Phase-II
Technology Stack (Architecture & Stack)

Date	15 October 2022
Team ID	PNT2022TMID03976
Project Name	Small Lender-Applicant Credibility Prediction for Loan Approval
Maximum Marks	4 Marks

Technical Architecture:



1.3 User Stories

A user interface hosted on the web to easily interact with the applicants. FR-2 Uploading Loan Approval Documents Uploading the loan application document to verify applicant credibility. FR-3 Uploading Applicants ID proof Documents Applicants ID proofs like copy of PAN card, Address proof document and Identity proof documents are uploaded for checking the applicants credibility. FR-4 Database Management Applicants loan application data will be saved in the database and will be used for future reference and all the other documents are deleted after processing. FR-5 Reporting Predicting applicants credibility for loan approval using given data and generating the report for the applicant. FR-6 Internet Connectivity User should have a stable internet connection to access the functionality of our project via web application.

PROJECT PLANNING & SCHEDULING

Sprint Planning & Estimation

Project Planning Phase

Project Planning Template (Product Backlog, Sprint Planning, Stories, Story points)

Date	18 October 2022
Team ID	PNT2022TMID03976
Project Name	Project - Smart Lender -Applicant Credibility Prediction for Loan Approval
Maximum Marks	4 Marks

Product Backlog, Sprint Schedule, and Estimation (4 Marks)

Use the below template to create product backlog and sprint schedule

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority
Sprint-1	Get Data From Source	USN-1	First step is collecting the data set of the clients from source.	1	Medium
Sprint-1	Raw Data Validation	USN-2	We need to classify the data set as Good_data_Folder and Bad_Data_Folder	1	High
Sprint-2	Data Set Insertion Into DataBase & Exploratory Data analysis	USN-3	A Sqlite DataBase is created and all the files in Good_data_Folder is inserted.Then the data is analyzed using statistics and various visualizations is plotted.	2	High
Sprint-2	Data Pre Processing	USN-4	Data cleaning is done by removing unwanted parameters.	1	Medium
Sprint-3	Hyper parameter tuning and Model Building	USN-5	The data is tested on various algorithms and best parameters is derived from Grid Search	2	High
Sprint-3	Flask Setup	USN-6	Flask Framework is created. when the user uploads the data model will predict the output.	1	Medium
Sprint-4	Default prediction and saving the output in CSV file	USN-7	After predicting the best accuracy model the output is stored as CSV file.	1	Medium
Sprint-4	Deployment	USN-8	Deploying the model on IBM studio	1	Medium

Sprint Delivery Schedule

Project Planning Phase

Project Planning Template (Product Backlog, Sprint Planning, Stories, Story points)

Date	18 October 2022
Team ID	PNT2022TMID03976
Project Name	Project - Smart Lender- Applicant Credibility Prediction for Loan Approval
Maximum Marks	8 Marks

Product Backlog, Sprint Schedule, and Estimation (4 Marks)

Use the below template to create product backlog and sprint schedule

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Get Data From Source	USN-1	First step is collecting the data set of the clients from source.	5	Medium	S.SRI AAKASH SUDHARSAN B SRIVATSAN G RISHI VARDHAN M
Sprint-1	Raw Data Validation	USN-2	We need to classify the data set as Good_data_Folder and Bad_Data_Folder	5	High	S.SRI AAKASH SUDHARSAN B SRIVATSAN G RISHI VARDHAN M
Sprint-2	Data Set Insertion Into DataBase & Exploratory Data analysis	USN-3	A Sqlite DataBase is created and all the files in Good_data_Folder is inserted.Then the data is analyzed using statistics and various visualizations is plotted.	5	High	S.SRI AAKASH SUDHARSAN B SRIVATSAN G RISHI VARDHAN M
Sprint-2	Data Pre Processing	USN-4	Data cleaning is done by removing unwanted parameters.	5	High	S.SRI AAKASH SUDHARSAN B SRIVATSAN G RISHI VARDHAN M
Sprint-3	Hyper parameter tuning and Model Building	USN-5	The data is tested on various algorithms and best parameters is derived from Grid Search	5	Medium	S.SRI AAKASH SUDHARSAN B SRIVATSAN G RISHI VARDHAN M

Project Tracker, Velocity & Burndown Chart: (4 Marks)

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date	Sprint Release Date (Actual)
Sprint 1	10	6 days	24 Oct 2022	29 Oct 2022	5	0
Sprint 2	10	6 days	01 Nov 2022	7 Nov 2022	5	0
Sprint 3	10	6 days	08 Nov 2022	14Nov 2022	5	0
Sprint 4	10	6 days	14 Nov 2022	17 Nov 2022	5	0

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=Speed velocity/duration=10/6=1.667

Burndown Chart:

A burn down 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.

https://www.visual-paradigm.com/scrum/scrum-burndown-chart/
https://www.atlassian.com/agile/tutorials/burndown-charts

CODING AND SOLUTIONING

The following features are added in this project:

1. User-friendly UI: The UI is designed is such a way that each application user is able to understand the working of the application.The application is not complex and simple to use.
2. Error Handling and Event Logs: The Errors will be encountered, an explanation will be displayed as to what went wrong during the training of the model.An error will be defined as anything that falls outside the normal and intended usage.Proper Exception handling and event logs are implemented which maintains the modularity of the code in the backend training.The Example of such a method is shown below

```
def handle_imbalanced_data(x,y):
    try:
        rdsmple = SMOTETomek() #Using Oversampling balancing the dataset
        x_sampled, y_sampled = rdsmple.fit_resample(x, y)
        return x_sampled, y_sampled
    except Exception as e:
        raise Exception()
```

TESTING

1. Test Cases:

The Project is designed and implemented in modular fashion with various modules where each module is responsible for performing different tasks.Therefore different Unit test cases had been implemented in each module of the project

Test Case Description	Prerequisite	Expected Result
Verify whether the Events Logging is working perfectly	1. Project Events logging module should be developed	Event logging is working for every test cases
Verify whether the data validation module is working is pipelined manner	numpy ,pandas, sqlite are needed to be installed	The program should validate the dataset completely and load the valid data into the database
Verify whether the data preprocessing steps are implemented in designed manner	numpy, pandas, matplotlib and seaborn	The validated data should be preprocessed in the proper steps.
Verify whether the parameter tuning and model selection steps are implemented in correct order.	pandas, scikit-learn, machine Learning algorithms	The module should perform all the hyperparameter tuning and model selection techniques on the preprocessed data and finally generate the model.

2. User Acceptance testing:

Test Case Description	Prerequisite	Expected Result
Verify whether the Application URL is accessible to the user	1. Application URL should be defined	Application URL should be accessible to the user
Verify whether the Application loads completely for the user when the URL is accessed	1. Application URL is accessible 2. Application is deployed	The Application should load completely for the user when the URL is accessed
Verify whether user is able to edit all input fields	1.Application is accessible	User should be able to edit all input fields
Verify whether user is presented with recommended results on clicking submit button	1.Application is accessible	User should be presented with recommended results on clicking submit
Verify whether the recommended results are in accordance to the user data	1.Application is accessible	The recommended results should be in accordance to the user data
Verify whether user has options to download the recommended results as well	1.Application is accessible	User should have options to download the recommended results

PERFORMANCE METRICS

Performance is going to be very important for this project. Since the bank employers will use this system for predicting the credit card defaulters, the accuracy should be as high as possible.Therefore the major performance metrics used in this project are ROC- AUC Score, Precision, Recall, F1-Score and accuracy which is best suited for binary classification.The Performance of

random forest is better when compared to other algorithms.There the model is build using random forest.The classification report of random forest is as follows

Classification Report

	precision	recall	f1-score	support
0	0.83	0.85	0.84	6797
1	0.85	0.82	0.83	6841
accuracy			0.84	13638
macro avg	0.84	0.84	0.84	13638
weighted avg	0.84	0.84	0.84	13638

Confusion Matrix

```
[[5800  997]
 [1225 5616]]
```

ADVANTAGES AND DISADVANTAGES

1. Advantages:
    - Application is user friendly and simple to use.
    - The model has been trained after the hyperparameter and model selection techniques therefore the performance is better.
    - Proper Error handling and Event logging mechanisms are implemented so the debugging is easy.
    - The Application is completely in modular fashion therefore the application is highly scalable which is the major advantage in this application.
2. Disadvantages:
    - The model can be trained on even more dataset than this.
    - More features can be added this this application.

CONCLUSION:

This Machine Learning based system will predict the probability of credit default based on the credit card owner's characteristics and payment history.Therefore the bank employees can decide whether or not to give loan for the customers.

source code link: https://github.com/sudharsanbaskars/smart-loan-lender

Project link:-https://github.com/IBM-EPBL/IBM-Project-21897-1659795590/tree/main/Final%20Deliverables/smart-loan-lender-main

Demo link:-

https://www.kapwing.com/videos/6378865139435c01e174bab0