1. 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 is a variety of techniques are used for risk level calculation. In addition, credit risk is one of 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.

We will be using classification algorithms such as Decision tree, Random forest, KNN, and XG-boost. We will train and test the data with these algorithms.
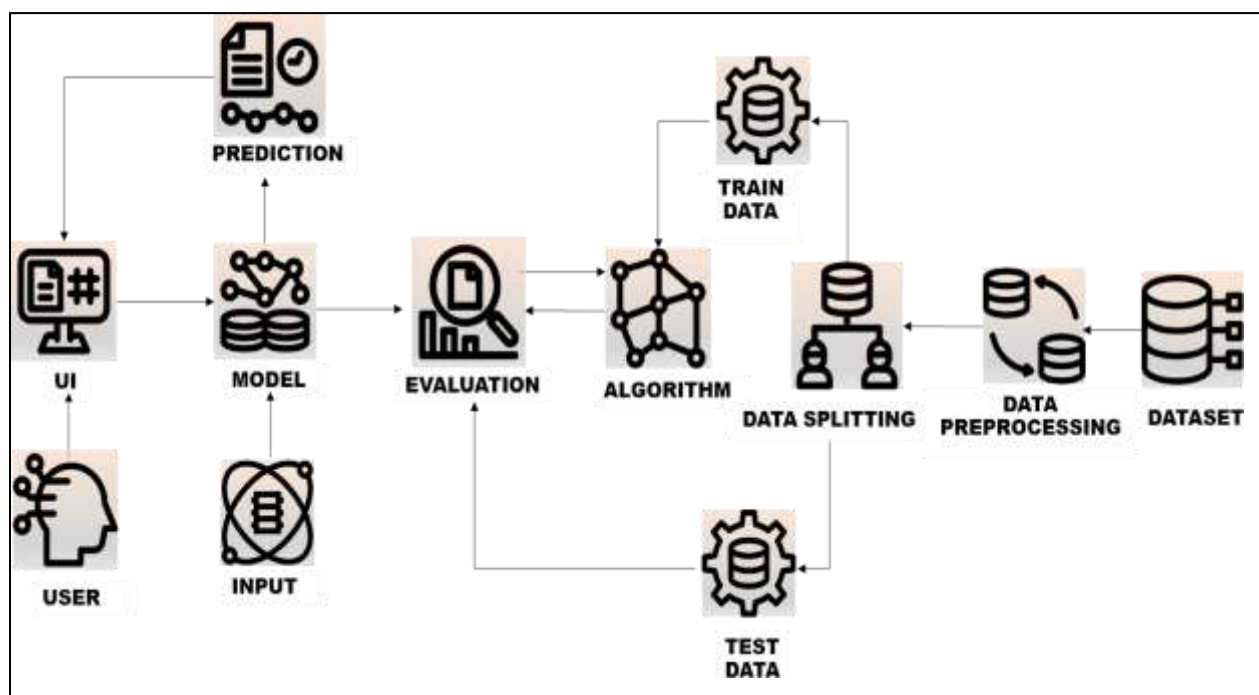
1.2 Purpose

- Write what are all the technical aspects that students would get if they complete this project.
- Knowledge of Machine Learning Algorithms.
- Knowledge of Python Language with Machine Learning
- You'll be able to understand the problem to classify if it is a regression or a classification kind of problem.
- You will be able to know how to pre-process/clean the data using different data pre-processing techniques.
- Applying different algorithms according to the dataset and based on visualization.
- Real-Time Analysis of Project
- Building ease of User Interface (UI)
- Navigation of ideas towards other projects(creativeness)
- Knowledge of building ML models.
- How to build web applications using the Flask framework.

2. LITERATURE SURVEY

[1] **Ms. Kathe Rutika Pramod** uses the decision tree for the loan prediction. In Decision tree each node represents a feature (attribute), each link (branch) represents a decision (rule) and each leaf represents an outcome (categorical or continues value). Using different data analytics tools loan prediction and there severity can be forecasted. In this process it is required to train the data using different algorithms and then compare user data with trained data to predict the nature of loan. Several R functions and packages were used to prepare the data and to build the classification model. The work proves that the R package is an efficient visualizing tool that applies data mining techniques. Using R Package, customer's data analysis can be done and depends on that bank can sanction or reject the loan. In real time customers data sets may have many missing and imputed data which needs to be replaced with valid data generated by making use of the available completed data. The dataset has many attributes that define the credibility of the customers seeking for several types of loan. The values for these attributes can have outliers that do not fit into the regular range of data. DT is a supervised learning algorithm used to solve classification and regression problems too. Here, DT uses tree representation to solve the prediction problem, i.e., external node and leaf node in a tree represents attribute and class labels respectively. The analytical process started from data cleaning and processing, Missing value imputation with mice package, then exploratory analysis and finally model building and evaluation. The best accuracy on public test set is 0.811. This brings some of the following insights about approval. Applicants with Credit history not passing fails to get approved, Probably because that they have a probability of a not paying back. Most of the Time, Applicants with high income sanctioning low amount is to more likely get approved which make sense, more likely to pay back their loans. Some basic characteristic gender and marital status seems not to be taken into consideration by the company.

[2] **Shubham Nalawade, Suraj Andhe, Siddhesh Parab, Prof. Amruta Sankhe** proposed system includes a web application with a model trained by using machine learning algorithms deployed in it. There are a total 11 fields in the form which the user needs to fill. The dataset that we have used for training the model also includes 11 attributes. This dataset is pre-processed before using it for training the model. The pre-processing is done by replacing the null values in the dataset with mean and mode method and replacing the string values with 1 and 0 using label encoder. Then the dataset was divided into two parts: train and test. 90% of the dataset is used for training purposes and 10% is used for testing the accuracy that the model will give for different algorithms. After splitting the dataset different algorithms were applied and each of them gave different accuracy. The best we got was from Logistic Regression i.e., 88%. Once the model is trained a pickle file is created of the model. When the client wants to predict his/her loan approval the client has to first fill a form by visiting our web application. After filling the form, the user has to just click on the MAKE PREDICTION button and depending on the pickle file or the model that we have trained it will give the result as whether the loan of the customer will be approved or not. As we have also done the comparison of different machine learning algorithms in terms of their accuracy. The web application also includes a bar plot graph of the comparison of algorithms, insights of the dataset that we have used for training the model. This system will make it easier for the banks or organizations to do the job of loan approval prediction. Here author compared different machine learning algorithms for the Property Loan dataset; they are Random Forest, Naive Bayes, Logistic Regression and K Nearest Neighbors. The Logistic Regression algorithm gave the best accuracy (88.70%). Following this approach, we found that apart from the logistic regression, the rest of the algorithms performed satisfactory in terms of giving out the accuracy. The accuracy range of the rest of the algorithms were from 75% to 85%. Whereas the logistic regression gave us the best possible accuracy (88.70%) after the comparative study of all the algorithms.

[3] **Soni P M, Varghese Paul** introduces a new hybrid feature selection algorithm using wrapper method and fisher score method. The new algorithm is termed as wrapper-fisher feature selection algorithm. In this work, LCPS uses a wrapperfisher feature selection algorithm to select the most significant features which will improve the accuracy of Random Forest (RF) classification. After studying various past data from the bank it is possible to identify several attributes that can influence the customer behaviour. The most influencing attribute can be considered while a new customer approaches the bank for loan and thus we can identify the potential of customer. Here by enabling the bank officers to identify fraud applicants by using the final application of this research work. The accuracy level considerably increased after feature selection

methods were applied to the classifier. The proposed algorithm had produced better accuracy than existing methods. Experiments on standard data sets proved that the proposed algorithm for loan credibility prediction system outperforms many other feature selection methods. , a novel hybrid feature selection approach is proposed to predict the loan repayment capability behavior of a customer in a cost effective way. Complex set of decision making are need to be taken by bank officers to determine whether to approve loan applicants or not. Normally classification technique solved the problem up to an extent. Now the experiment proved that a model that use feature selection before classification can help the bank officers to take proper decision more accurately. This proposed methodology will protect the bank from further misuse, fraud applications etc by identifying the customers whose repayment capability status is risky especially in the co-operative banking sector. The experiment proved that the classification accuracy have considerably increased after feature selection. The proposed algorithm had produced better accuracy than existing methods. Experiments on standard data sets proved that the proposed algorithm for loan credibility prediction system outperforms many other feature selection methods.

[4] In **Dr.AMIT KUMAR GOEL** proposed model for loan prediction, Dataset is split into training and testing data. After then training datasets are trained using the decision tree algorithm and a prediction model is developed using the algorithm. Testing datasets are then given to model for the prediction of loan. The motive of this paper is to predict the defaults who will repay the loan or not. Various libraries like pandas, numpy have been used. After the loading of datasets, Data preprocessing like missing value treatment of numerical and categorical is done by checking the values. Numerical and categorical values are segregated. Outliers and frequency analysis are done. developed a prediction model for Loan sanctioning which will predict whether the person applying for loan will get loan or not. The major objective of this project is to derive patterns from the datasets which are used for the loan sanctioning process and create a model based on the patterns derived in the previous step. This model is developed by using the one of the machine learning algorithms. Here the author used decision tree algorithm for development. Based on the segregated value the decision tree able to work and predict the loan approval. Here author is able to conclude that Decision tree version is extraordinary efficient and gives a higher end result. Developed a model which can easily predict that the person will repay its loan or not. we can see our model has reduced the efforts of bankers. Machine learning has helped a lot in developing this model which gives precise results.

[5] **Mehul Madaan** used two machine learning algorithms, the Random Forest and Decision Trees to work out a model for loan prediction and credit risk assessment. The results of both the model are shown below with their classification report and confusion matrix to get a better understanding of the accuracy and other scores of the two models. This paper aimed to explore, analyse, and build a machine learning algorithm to correctly identify whether a person, given certain attributes, has a high probability to default on a loan. This type of model could be used by Lending Club to identify certain financial traits of future borrowers that could have the potential to default and not pay back their loan by the designated time. The Random Forest Classifier provided us with an accuracy of 80% while the Decision Tree method provided us with an accuracy of 73%. Hence, the Random Forest model appears to be a better option for such kind of data. Lending Club must be careful when identifying potential borrowers who fit certain criteria. For example, borrowers who do not own a home and are applying for a small business or wedding loan, this could be a negative combination that results in the borrower defaulting on a loan. One of the drawbacks is simply the limited number of people who defaulted on their loan in the 8 years of data (2007-2015). We could use an updated data frame that consists of the next 3 years' values (2015-2018) and see how many of the current loans were paid off, defaulted, or even charged off. Then, these new data points can be used for prediction or and training new models for better and more accurate results. Since the algorithm puts some of the non-defaulters in the default class, we might want to look further into this issue to help the model accurately predict capable borrowers.

[6] In the paper presentation of **AFRAH KHAN, EAKANSH BHADOLA, ABHISHEK KUMAR and NIDHI SINGH**, It will be comparing different prediction models and deduce their limitations as well as advantages. Since all the research papers used different sets of data to infer the accuracy and for cross validation of data, the authors have used the same data for all the models which will give a clearer view on their performance and lead to a better comparison of the same. On the basis of the results, a modified prediction model will be created to ensure maximum accuracy and performance. The predictive models

based on Logistic Regression, Decision Tree and Random Forest, give the accuracy as 80.945%, 93.648% and 83.388% whereas the cross-validation is found to be 80.945%, 72.213% and 80.130% respectively. This shows that for the given dataset, the accuracy of model based on decision tree is highest but random forest is better at generalization even though it's cross validation is not much higher than logistic regression.

2.1 Existing problem :

They have presence across all urban and rural areas. Customer first apply for home loan after that company validates the customer eligibility for loan. However doing this manually takes a lot of time. Hence it wants to automate the loan eligibility process (real time) based on customer information and verify their documents. So the final thing is to identify the factors/ customer segments that are eligible for taking loan.

2.2 Existing System:

Banks need to analyze for the person who applies for the loan will repay the loan or not. Sometime it happens that customer has provided partial data to the bank, in this case person may get the loan without proper verification and bank may end up with loss. Bankers cannot analyze the huge amounts of data manually, it may become a big headache to check whether a person will repay its loan or not. It is very much necessary to know the person getting loan is going in safe hand or not. So, it is pretty much important to have a automated model which should predict the customer getting the loan will repay the loan or not.
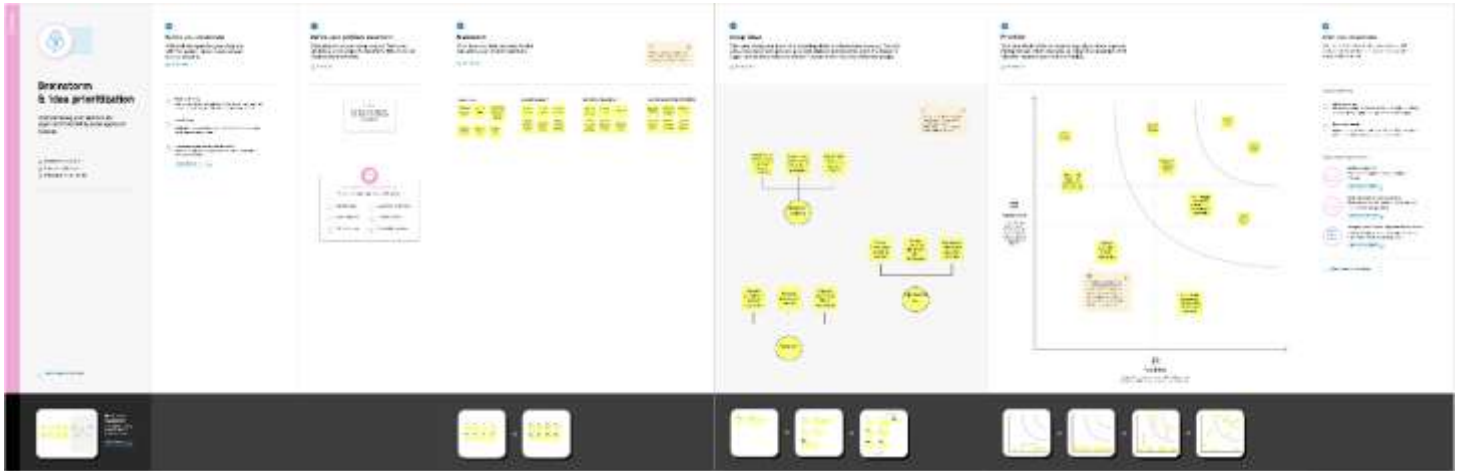
2.3 References :

[1] Ms. Kathe Rutika Pramod Information Technology Engineering SVIT, Nashik Maharashtra, India a An Approach For Prediction Of Loan Approval Using Machine Learning Algorithm-2021 IJCRT | Volume 9, Issue 6 June 2021

[2] Shubham Nalawade, Suraj Andhe, Siddhesh Parab, Prof. Amruta Sankhe- Loan Approval Prediction-Loan Approval Prediction -2021 IJCRT | Volume: 09 Issue: 04 | Apr 2022

[3] Soni P M, Varghese Paul- Algorithm For the Loan Credibility Prediction System- International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1S4, June 2019

[4] Dr.AMIT KUMAR GOEL, M.Tech., Ph.D - LOAN PREDICTION SYSTEM - APRIL / MAY-2020

[5] Loan default prediction using decision trees and random forest: A comparative study-IOP Conference Series Mehul Madaan et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1022 012042

[6] AFRAH KHAN, EAKANSH BHADOLA, ABHISHEK KUMAR and NIDHI SINGH - LOAN APPROVAL PREDICTION MODEL A COMPARATIVE ANALYSIS | Advances and Applications in Mathematical Sciences Volume 20, Issue 3, January 2021

2.4 Problem Statement Definition :

The credit system governed by the banks is one of the most important factors which affect our country's economy and financial condition. Also, credit risk is one of the main functions of the banking community. People approach banks to fulfill their needs by taking 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 by giving the fake document to the bank, so banks realize that retaining customers and preventing fraud should be a strategic policy for healthy competition. By using applied data science techniques and machine learning algorithms, we will verify the documents and check the credit score of the person and predict whether the loan is approved or not.

3. IDEATION & PROPOSED SOLUTION
3.1Empathy Map Canvas:

For clear view of the image click the below link: (ctrl+click)

https://github.com/IBM-EPBL/IBM-Project-26575-1660029860/blob/main/Ideathon%20Phase/Empathimap.pdf

3.2 Ideation & Brainstorming
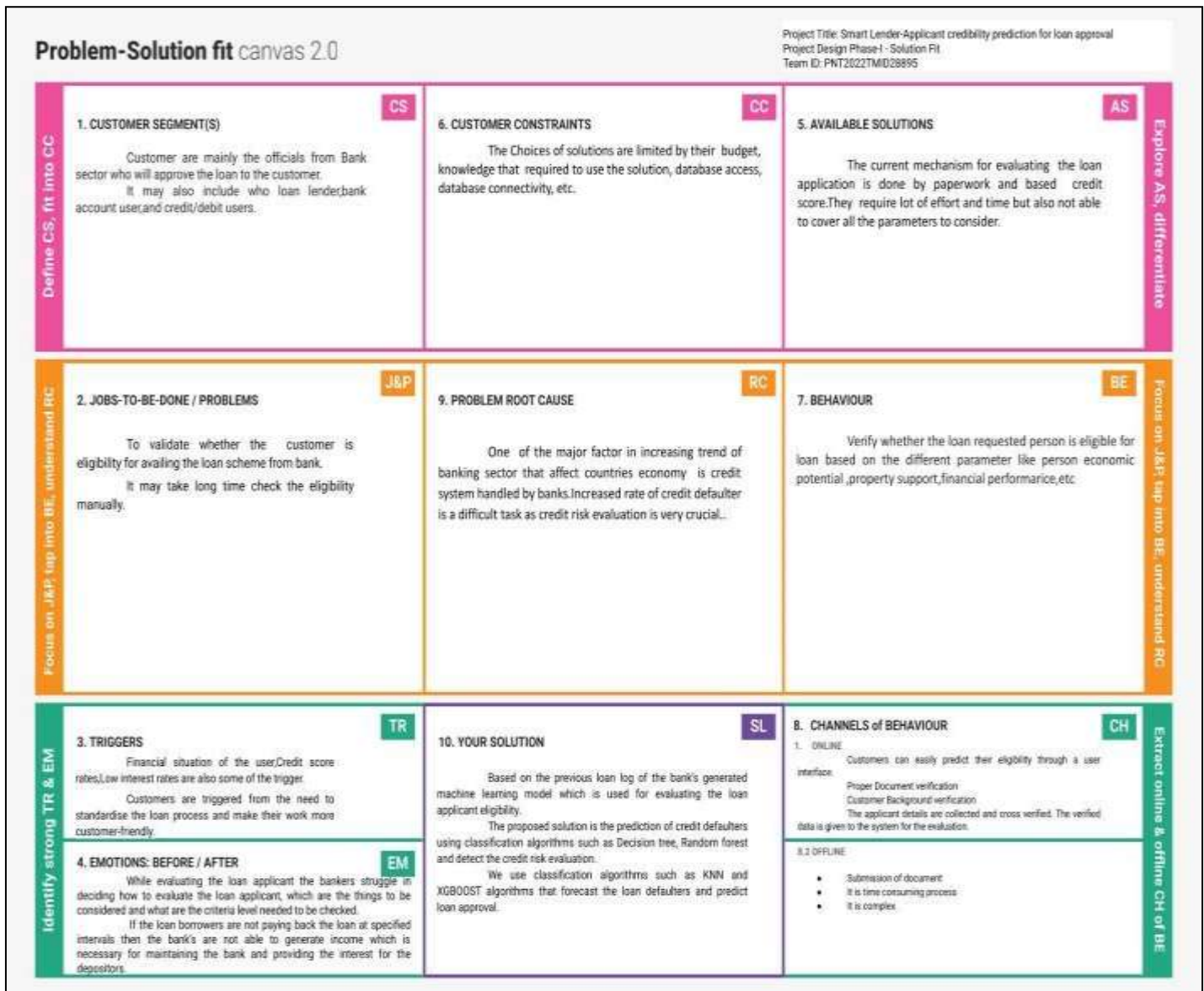

For clear view of the image click the below link: (ctrl+click)

https://github.com/IBM-EPBL/IBM-Project-26575-1660029860/blob/main/Ideathon%20Phase/Brainstrom.pdf

3.3 Proposed Solution:

S.No.	Parameter	Description
1	Problem Statement (Problem to be solved)	They have presence across all urban and rural areas. Customer first apply for home loan after that company validates the customer eligibility for loan. However doing this manually takes a lot of time. Hence it wants to automate the loan eligibility process (real time) based on customer information and verify their documents. So the final thing is to identify the factors/ customer segments that are eligible for taking loan
2	Idea / Solution Description	The interface helps the customer to predict the Applicant Credibility Prediction for Loan Approval using machine learning models
3	Novelty / Uniqueness	1. Verify the Time taken for Loan disposal 2. Check user bank credit history 3. Provide every banks loan process information 4. Add banks interests comparison features.
4	Social Impact / Customer Satisfaction	a. Avoid falling for fake offers and plans. This helps to avoid the unauthorized loan and help to black-list them. b. This improve the money flow in efficient way.
5	Business Model (Revenue Model)	a. Consider Your needs and choose Your Loan Amount. This helps in betterment of business profit. b. This helps to provide the timely delivery of loan at effective manner. c. Improving Client Experience

3.4 Problem Solution fit

Problem-Solution fit canvas 2.0

Project Title: Smart Lender-Applicant credibility prediction for loan approval
Project Design Phase-I - Solution Fit
Team ID: PNT2022TMID28895

1. CUSTOMER SEGMENT(S) — CS

Customer are mainly the officials from Bank sector who will approve the loan to the customer.
It may also include who loan lender,bank account user,and credit/debit users.

6. CUSTOMER CONSTRAINTS — CC

The Choices of solutions are limited by their budget, knowledge that required to use the solution, database access, database connectivity, etc.

5. AVAILABLE SOLUTIONS — AS

The current mechanism for evaluating the loan application is done by paperwork and based credit score.They require lot of effort and time but also not able to cover all the parameters to consider.

2. JOBS-TO-BE-DONE / PROBLEMS — J&P

To validate whether the customer is eligibility for availing the loan scheme from bank.
It may take long time check the eligibility manually.

9. PROBLEM ROOT CAUSE — RC

One of the major factor in increasing trend of banking sector that affect countries economy is credit system handled by banks.Increased rate of credit defaulter is a difficult task as credit risk evaluation is very crucial..

7. BEHAVIOUR — BE

Verify whether the loan requested person is eligible for loan based on the different parameter like person economic potential ,property support,financial performarice,etc

3. TRIGGERS — TR

Financial situation of the user,Credit score rates,Low interest rates are also some of the trigger.
Customers are triggered from the need to standardise the loan process and make their work more customer-friendly.

4. EMOTIONS: BEFORE / AFTER — EM

While evaluating the loan applicant the bankers struggle in deciding how to evaluate the loan applicant, which are the things to be considered and what are the criteria level needed to be checked.
If the loan borrowers are not paying back the loan at specified intervals then the bank's are not able to generate income which is necessary for maintaining the bank and providing the interest for the depositors.

10. YOUR SOLUTION — SL

Based on the previous loan log of the bank's generated machine learning model which is used for evaluating the loan applicant eligibility.
The proposed solution is the prediction of credit defaulters using classification algorithms such as Decision tree, Random forest and detect the credit risk evaluation.
We use classification algorithms such as KNN and XGBOOST algorithms that forecast the loan defaulters and predict loan approval.

8. CHANNELS of BEHAVIOUR — CH

1. ONLINE
Customers can easily predict their eligbility through a user interface.
Proper Document verification
Customer Background verification
The applicant details are collected and cross verified. The verified data is given to the system for the evaluation.

8.2 OFFLINE
- Submission of document
- It is time consuming process
- It is complex

For clear view of the image click the below link: **(ctrl+click)**

https://github.com/IBM-EPBL/IBM-Project-26575-1660029860/blob/main/Project%20Design%20Pase-I/Problem%20solution%20fit.pdf

4. REQUIREMENT ANALYSIS

4.1 Functional requirement:

Following are the functional requirements of the proposed solution.

- ❖ A functional requirement defines a function of a system or its component, where a function is described as a specification of behaviour between inputs and outputs.
- ❖ It specifies "what should the software system do?"
- ❖ It is mandatory
- ❖ Defined at a component level
- ❖ Usually easy to define
- ❖ Helps you verify the functionality of the software

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Bank Statement	Customer should submit the 6 months bank statement when taking decision on loan application.
FR-4	Proof of Identity	Customer can submit their Aadhaar card, PAN card, Passport, Voters ID card, Driving License as proof of identification.
FR-5	Income proof	Customer are required to show some income proof such as Salary slips, Employer affidavits, Income tax returns, Bank statements, Form 16, etc
FR-6	Debt-to-income	DTI is the gross of all monthly debt payments divided by the gross monthly income, calculated as a percentage.
FR-7	Credit Score and History	Lenders will look into your credit score to see how credit worthy you are. The closer your credit score is to 900, the higher are the chances that the lender will approve your loan application.
FR-8	Origination Fee	It is typically 0.5% to 1% of the loan amount and is charged by a lender as compensation for processing a loan application.
FR-9	Collateral	It allows the customer to pledge any asset to seek a loan.

4.2 Non-Functional requirements:

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

❖ A non-functional requirement defines the quality attribute of a software system
❖ It places constraint on "How should the software system fulfil the functional requirements?"
❖ It is not mandatory
❖ Applied to system as a whole
❖ Usually more difficult to define
❖ Helps you verify the performance of the software

FR No.	Non-Functional Requirement	Description
NFR-1	Usability	Customers can use the application in almost all the web browsers. Application is with good looking and detailed UI, which makes it more friendly to use.
NFR-2	Security	Customers are asked to create an account for themselves using their email which is protected with an 8 character-long password, making it more secure
NFR-3	Reliability	Customers can raise their queries and will be replied with a valid reply, as soon as possible, making the application even more reliable and trust-worthy.
NFR-4	Performance	Customers will have a smooth experience while using the application, as it is simple and is well optimised.
NFR-5	Availability	Application is available 24/7 as it is hosted on IBM Cloud
NFR-6	Scalability	In future, may be cross-platform mobile applications can be developed as the user base grows.

5. PROJECT DESIGN

5.1 Data Flow Diagrams

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 rightamount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.

For clear view of the image click the below link: **(ctrl+click)**
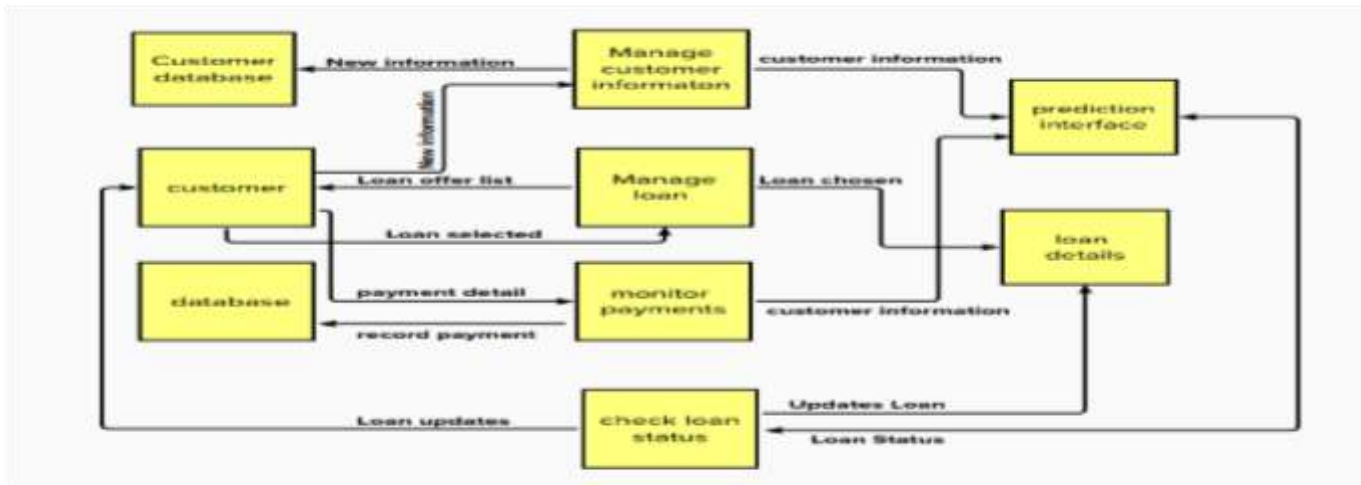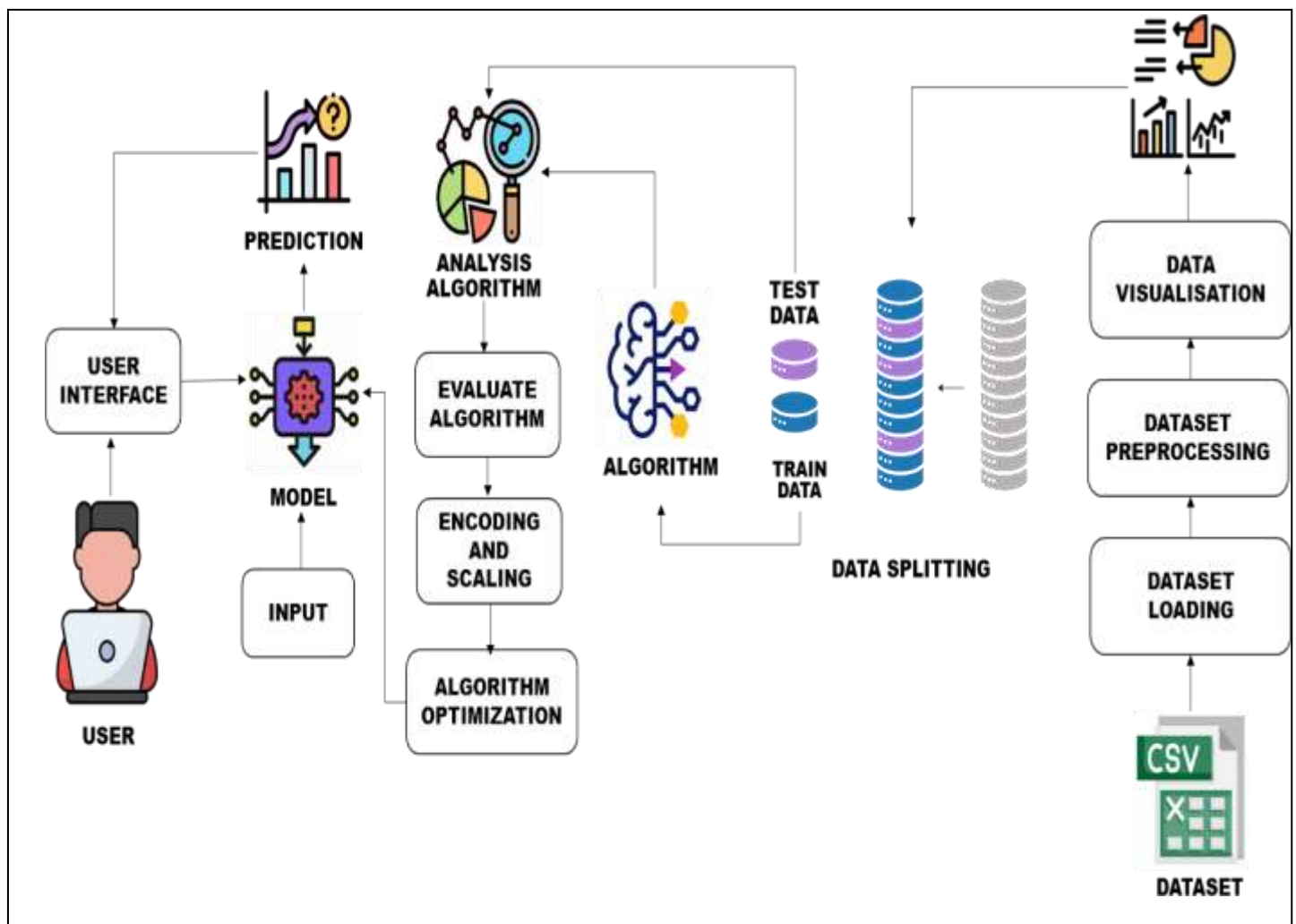
https://github.com/IBM-EPBL/IBM-Project-26575-1660029860/blob/main/Project%20Design%20pase-II/DATA%20FLOW%20DIAGRAMS%20.pdf

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 byentering my email, password, and confirming my password.	I can access my account /dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation emailonce I have registered for the application	I can receive confirmationemail & click confirm	High	Sprint-1
		USN-3	Automatically,the information are stored in theCustomer database		Low	Sprint-3
		USN-4	As a user, I can register for the applicationthrough Gmail		Medium	Sprint-2
	Login	USN-5	As a user, I can log into the application byentering email & password	I can login to the application using login credentials	High	Sprint-1
Customer (Webuser)		USN-6	After logging in ,I have to selected loan chosen,then the loan offer list is displayedfrom the Loan Offers Database.	I have select the types ofloan	High	Sprint-1
		USN-7	The payment information has to be given thenthe moniter payments stores the information.		Medium	Sprint-2
		USN-8	The record payments are stored in thepayment database		Low	Sprint-3
Administrator		USN-9	The web application gives the loan status andthe upadates the loan to the customer and the updated loan is also stored in Loan Details in Database	Then I have shown withthe loan status	High	Sprint-1

5.2 Solution & Technical Architecture:

5.3 User Stories

For clear view of the image click the below link: **(ctrl+click)**

https://github.com/IBM-EPBL/IBM-Project-26575-1660029860/blob/main/Project%20Design%20pase-II/CUSTOMER%20%20JOURNEY%20MAP.pdf
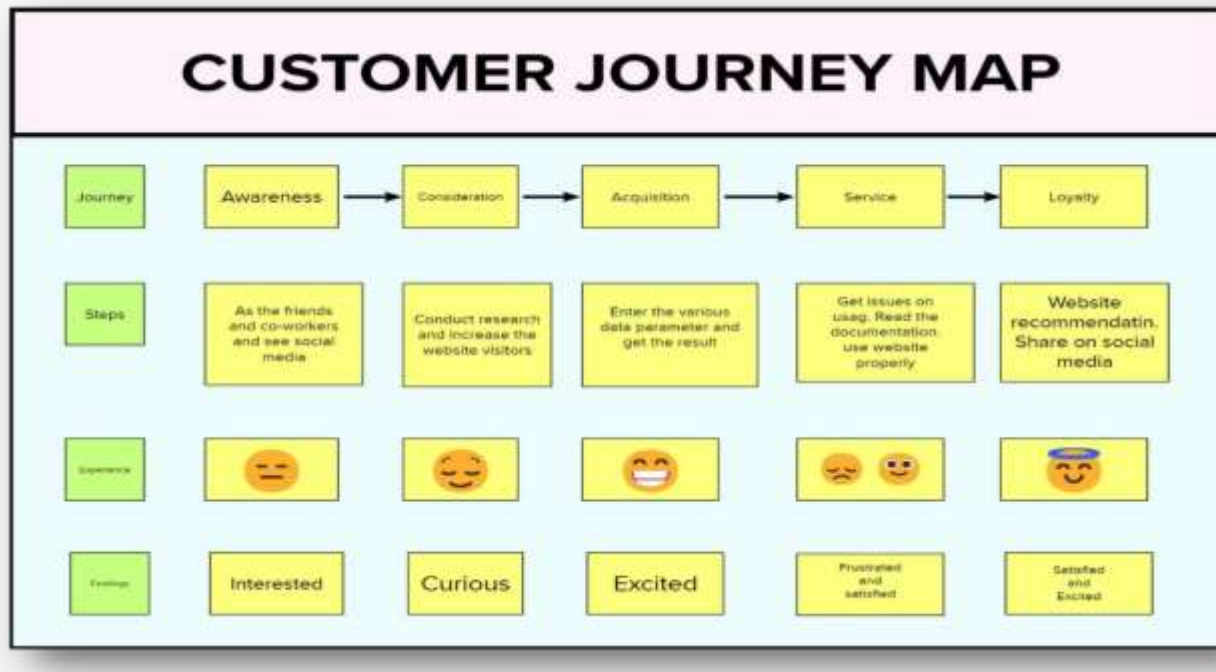
6. PROJECT PLANNING & SCHEDULING

6.1Sprint Planning & Estimation:

Activity List:

In Project Management Planning is an important task to scheduling the phrase of the project to the Team Member.

In this Activity can shows the various activity are allocated and Done by the Team Members!

In Project we can Split into the Four Step of Phrases are

Phrase 1: Information Collection and Requirement Analysis

Phrase 2: Project Planning and Developing Modules

Phrase 3: Implementing the High Accuracy Deep Learning Algorithm to Perform

Phrase 4: Deploying the Model on Cloud and Testing the Model and UI Performance

Agile Methodology for Activity Planning
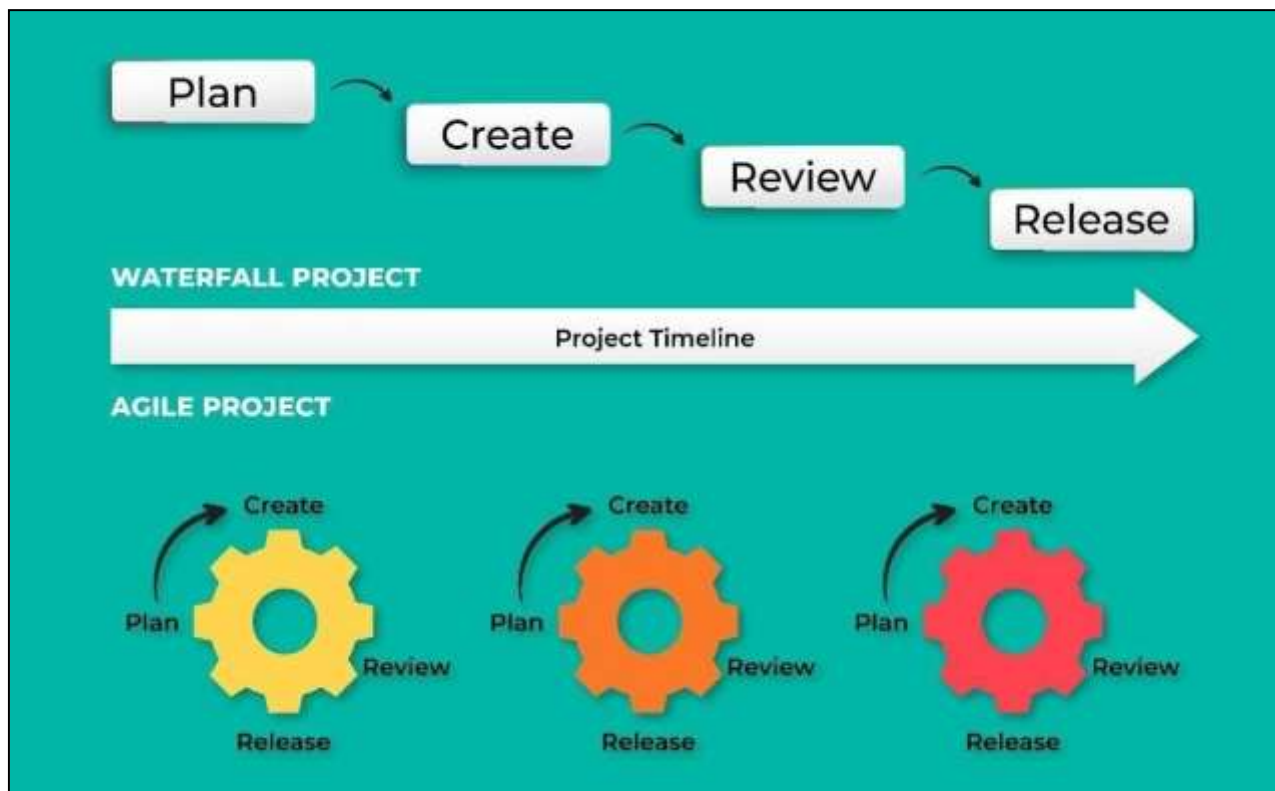
6.2 Sprint Delivery Schedule

Release	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint – 1	Forms	USN – 1	As a user, I can enter the data which I have and also the data which thewebsite asks to me	6	VeryHigh	VIGNESH A, RATHESH PRAVEEN P
Sprint – 3	Prediction	USN – 2	As I have given the data into the webpage, now the data can be predictedfor the loan avail	4	Medium	ULAGESWAR AN T, JAYARAM J
Sprint – 4	Deployment of the Webpage in Cloud	USN – 3	As a user, I require global access to theweb page as a user	3	Low	VIGNESH A, RATHESH PRAVEEN P
Sprint – 4	Deployment of AI modelin the cloud	USN – 4	Model would be running on the Cloud	3	Low	ULAGESWAR AN T, JAYARAM J
Sprint – 2	Model building	USN – 5	I require an ML model that cancategorise Credit defaulters	5	High	VIGNESH A, RATHESH PRAVEEN P
Sprint – 3	User Interface building	USN – 6	As a User, I need a medium to enter mydata	4	Medium	ULAGESWAR AN T, JAYARAM J

Project Tracker, Velocity & Burndown Chart:

| 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	20	6 Days	24 Oct 2022	29 Oct 2022	6	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	6	6 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	6	13 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	6	19 Nov 2022

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)

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

6.3 Reports from JIRA:

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 suchas Scrum. However, burn down charts can be applied to any project containing measurable progress over time.
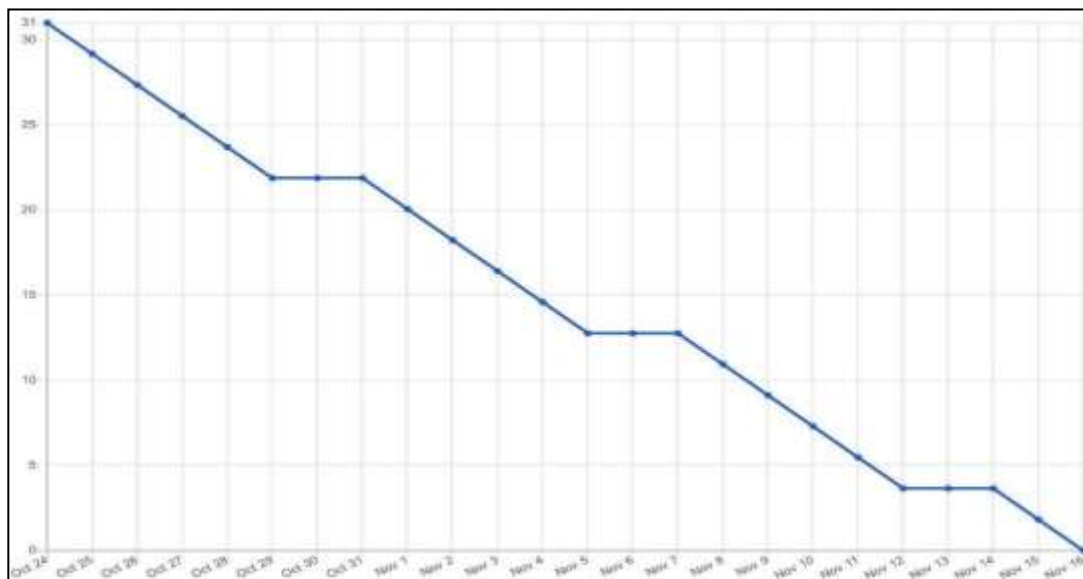
7. CODING & SOLUTIONING

7.1 Feature 1

home.html

```html
<!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=1.0" />
    <link rel="stylesheet" href="style.css" type="text/css" />
    <title>Loan Predictor</title>
  </head>

  <body>
    <main>
      <div class="mail">
       <center>
        <h1>Loan Prediction</h1>
        <h3>Find your Loan Eligibility</h3>
        <h5>
          Click below button and fill the details to know your Loan
          Eligibility.
        </h5>

        <div class="container">
         <a href="index.html">
           <button class="btn" data-hover="Loan Predictor">
            <div>Click to Check</div>
           </button>
         </a>
        </div>
       </center>
      </div>
    </main>
  </body>
</html>
```

```html
<!DOCTYPE html>
<html>
<head>
    <style type="text/css">
    .header{
            background: linear-gradient(pink);
            background:-webkit-linear-gradient(pink);
            background:-moz-linear-gradient(pink);
            background:-o-linear-gradient(pink);
            position: absolute;
            left:  0px;
            top: 0px;
            width: 100%;
            height: 100px;
            text-align: center;

            text-transform: capitalize;
    }
    .reg_img
{
 height: 1000px;
 margin-top: 0px;
 background-image: url("../static/33.jpeg");

}
.box2
{
 height: 960px;
 width: 600px;
 background-color: black;
 margin: 70px auto;
 opacity: .7;
 color: white;
 padding: 20px;
}
  .
   .reg{
    margin-top: 130px;
```

```css
        width: 50%;
    text-align: center;
    text-decoration-style: smooth;
    font-family: verdana;
    color: white;
    text-transform: capitalize;
    background: rgb(10, 34, 56);
    border-radius: 10px;
    overflow: scroll;
}

.sco table{
    text-overflow: scroll;
}


    </style>
</head>
<head>
    <title>Loan Registration using Machine Learning </title>
</head>
<body>
<header>
    <div class="header", id="ls">

            <h1>Smart Lender - Applicant Credibility Prediction for Loan Approval 💰</h1>
    </div>
</header>

<div class="reg_img">

            <div class="box2">

                    <h1 style="text-align: center; font-size: 25px;">APPLICANT DETAILS
FORM</h1>


            <form action="{{url_for('prediction')}}" method="POST">
```

```html
<br><br>
<label>Select Gender</label>
<br><br>
<input type="radio" value="Male" name="gender">Male
<input type="radio" value="Female" name="gender">Female
<br><br>
<label>Status</label>
<br><br>
<input type="radio" value="Married" name="status">Married
<input type="radio" value="Single" name="status">single
<br><br>
<label>Enter number of dependants</label>
<br><br>
<input type="text" name="dependants", placeholder="3">
<br><br>
<label>Education Level</label>
<br><br>
<select name = "education">
        <option value="Graduate"   value="Graduate">Graduate</option>
        <option     value="Not     Graduate"      value="Not     Graduate">  Not
Graduate</option>
</select>
<br><br>
<label>Employment status</label>
<br><br>
<select name="employ">
        <option value="No">Employed</option>
        <option value="Yes"> Self Employed</option>
        <option value="Yes">I do both</option>
        <option value="No">None of above</option>
</select>
<br><br>
<label>Enter your annual income </label>
<br><br>
<input type="text" name="aincome", placeholder="5849">
<br><br>
<label>Enter your Coincome </label>
<br><br>
<input type="text" name="coincome", placeholder="0">
```

```html
            <br><br>
            <label>Loan amount </label>
            <br><br>
            <input type="text" name="Lamount", placeholder="128">
            <br><br>
            <label>Loan amount Term </label>
            <br><br>
            <input type="text" name="Lamount_term", placeholder="360">
            <br><br>
            <label>Enter your credit history </label>
            <br><br>
            <input type="text" name="credit", placeholder="1">
            <br><br>
            <label>select your property area </label>
            <br><br>
            <select name = "property_area">
            <option value="urban">urban</option>
            <option value="Semiurban">Semi urban</option>
            <option value="Rural">Rural</option>
            </select>
            <br><br>
            <input type="submit" name="sumbit" value="Register">
        </form>
    </div>

  </div>

  <footer>
      <div>

      </div>
  </footer>

  </body>
  </html>
```

```html
<!DOCTYPE html>
<html>
<head>
    <style type="text/css">
            .header{
            background: linear-gradient( white);
            background:-webkit-linear-gradient( white);
            background:-moz-linear-gradient( white);
            background:-o-linear-gradient( white);
            position: absolute;
            left:  0px;
            top: 0px;
            width: 100%;
            height: 100px;
            text-align: center;
            text-transform: capitalize;
    }
    .log_img
    {
    height: 650px;
    margin-top: 0px;
    background-image: url("../static/22.jpeg");

    }

    .app {
            height: 500px;
            width: 600px;
            background-color: rgba(0, 0, 0, 0.945);
            margin: 100px auto ;
            opacity: .7;
            color: white;
            padding: 20px;
    }
    </style>
    <title> Classification results</title>
</head>
```

```html
<body>
    <div>
        <header class="header">

            <h1>The Application result </h1>

        </header>
    </div>
    <div class="log_img">
    <div class="app">
        <section>
    <table align="center">
        <th>
    <h1> Your information has been Submitted Please wait </h1>
    <table align="center">
      <th>

    {% if output== 1 %}
    <br><br>
    <h1    style="text-align:    center;    font-size:    35px;font-family:    Lucida
Console;">Congratulations 👍👍👍 You are eligible for the Loan 💰💰💰</h1>
    <img src="../static/88.jfif",width=500px,height=600px>
    <br><br>

    {% elif output==0 %}
    <h1 style="text-align: center; font-size: 35px;font-family: Lucida Console;">We are sad
😞😞😞 to inform that you are not eligible for the Loan</h1>
    <img src="../static/77.png",width=500px,height=600px>
    {% endif %}
    </th>
    </table>

    </th>
    </table>
        </section>
    </div>

    <div>
```

```html
        <footer>

        </footer>
    </div>
  </div>


  </body>
  </html>
```

7.2 Feature 2
app.py
```python
import flask
import joblib
import numpy as np
from flask import render_template, request
from flask_cors import CORS
app = flask.Flask(__name__)
CORS(app)

@app.route('/', methods=['GET'])

def sendHomePage():
    return render_template('index.html')


@app.route('/output', methods = ['POST'])

def prediction():
    if request.method == 'POST':
        gender = request.form['gender']
        married = request.form['status']
        dependat = request.form['dependants']
        education = request.form['education']
        employ = request.form['employ']
        annual_income = request.form['aincome']
        co_income = request.form['coincome']
        Loan_amount = request.form['Lamount']
        Loan_amount_term = request.form['Lamount_term']
        credit = request.form['credit']
        proper = request.form['property_area']

    gender = gender.lower()
    married= married.lower()
    education = education.lower()
```

```python
        employ = employ.lower()
        proper = proper.lower()
        if(employ=='yes'):
                employ = 1
        else:
                employ = 0
        if(gender=='male'):
                gender = 1
        else:
                gender = 0
        if (married=='married'):
                married=1
        else:
                married=0
        if (proper=='rural'):
                proper=0
        elif (proper=='semiurban'):
                proper=1
        else:
                proper=2
        if (education=='graduate'):
                education=0
        else:
                education=1

        dependat = int(dependat)
        annual_income = int(annual_income)
        co_income = int(co_income)
        Loan_amount = int(Loan_amount)
        Loan_amount_term = int(Loan_amount_term)
        credit = int(credit)
        x =np.array([[gender, married,
dependat,education,employ,annual_income,co_income,Loan_amount,Loan_amount_term,credit
,proper]])
        model = joblib.load('Forest.pkl')
        ans = int(model.predict(x)[0])
        if (ans==1):
                print("Congratulations your eligble for this Loan")
        else:
                print("We are sad to inform that your request has not been accepted")
        return render_template('output.html', output=ans)
if __name__ == '__main__':
        app.debug = True
        app.run()
```

8. TESTING

8.1 Test Cases

For checking the loan application, We have two testcase

- Eligible
- Not Eligible

This is based on the training and testing the model we used in our application.

This eligibility can be checked by using the details entered by the users. This includes the details like

- Gender
- Status
- Dependants
- Education
- Employ
- Income
- Co-income(additional income)
- Loan amount
- Loan amount term(in days)
- Credit history
- Property area(type of location)

8.2 User Acceptance Testing

1. *Purpose of Document*

The purpose of this document is to briefly explain the test coverage and open issues of the project - **Smart Lender - Applicant Credibility Prediction for Loan Approval** at the time of the release to User Acceptance Testing (UAT).

2. *Defect Analysis*

This report shows the number of resolved or closed bugs at each severity level, and how they were resolved

Resolution	Severity 1 (High)	Severity 2 (Moderate)	Severity 3 (Low)	Subtotal
By Design	1	3	2	6
Duplicate	1	0	3	4
External	2	3	0	5

Fixed	4	6	4	14
Not Reproduced	0	0	1	1
Totals	8	12	10	30

3. *Test Case Analysis*

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

Section	Total Cases	Not Tested	Fail	Pass
Print Engine	6	0	0	6
Client Application	16	0	0	16
Security	2	0	0	2
Exception Reporting	3	0	0	3
Final Report Output	4	0	0	4
Version Control	1	0	0	1

9. RESULTS

9.1 Performance Metrics:

In our project we used XG-Boost model for prediction.

S.No.	Parameter	Values	Screenshot
1.	Metrics	**Classification Model:** Confusion Matrix , Accuray Score- & Classification Report	Fig 1
2.	Tune the Model	Hyperparameter Tuning Validation Method	Fig 2

```
In [52]: xgboost(x_train, x_test, y_train, y_test)

****Gradient BoostingClassifier****
Confusion matrix
[[ 74  29]
 [ 12 108]]
Classification report
              precision    recall  f1-score   support

           0       0.86      0.72      0.78       103
           1       0.79      0.90      0.84       120

    accuracy                           0.82       223
   macro avg       0.82      0.81      0.81       223
weighted avg       0.82      0.82      0.81       223

Testing accuracy:  0.8161434977578476
Training accuracy:  0.9466666666666667
```

From the four model Xgboost is performing well. Xgboost is giving the accuracy of 94% with training data , 81% accuracy for the testing data.so we considering xgboost and deploying this model.

Fig 1 - Metrics

Evaluating Performance Of The Model

```
In [53]: from sklearn.model_selection import cross_val_score

In [54]: # Xgboost Model is selected
         xg = GradientBoostingClassifier()

In [55]: xg.fit(x_train,y_train)
Out[55]: ▾ GradientBoostingClassifier
         GradientBoostingClassifier()

In [56]: yPred = xg.predict(x_test)

In [57]: f1_score(yPred,y_test, average='weighted')
Out[57]: 0.8183313193520658

In [58]: cv = cross_val_score(xg,x,y,cv=5)

In [59]: np.mean(cv)
Out[59]: 0.7230974276955885
```

Fig 2 - Tune the Model

10. ADVANTAGES & DISADVANTAGES:

ADVANTAGES:

- Fast and highly accurate result

- Easy handling of the problem

- Less risk and more convenient to use

- Reliablity is pretty high

- Better choice for responsive result

- Better user interface

DISADVANTAGES:

- Machine Learning model in general is little complex

- Prediction sometime not reliable because the model is build on the old data.

- The prediction result is more depend on the model.

11. CONCLUSION:

We determined the most important features that influence the loan approval status. These most important features are then used on some selected algorithms and their performance accuracy is compared with the instance of using all the features. The model can help the banks in figuring out which factors are important for the loan approval procedure. The comparative study makes us clear about which algorithm will be the best and ignores the rest, based on their accuracy.

We have developed a model which can easily predict that the person will repay its loan or not. we can see our model has reduced the efforts of bankers. Machine learning has helped a lot in developing this model which gives precise results

12. FUTURE SCOPE:

Future enhancement of this research work on training bots to predict the loan eligibility areas by using machine learning techniques. Since, machine learning is similar to data mining advanced concept of machine learning can be used for better prediction. The data privacy, reliability, accuracy can be improved for enhanced prediction

From the encouraging results, we believe that crime data mining has a promising future for increasing the effectiveness and efficiency of criminal and intelligence analysis. Visual and intuitive criminal and intelligence investigation techniques can be developed for loan credibility pattern. As we have applied machine learning technique of data mining for loan prediction we can also perform other techniques of data mining such as classification. Also we can perform analysis on various dataset such as enterprise survey dataset, poverty dataset, aid effectiveness dataset, etc.

13. APPENDIX:

Source Code link:

https://drive.google.com/drive/folders/1d2RZUmBB1VoaScmhWF2i17NY06mF84KZ

GitHub Link:

https://github.com/IBM-EPBL/IBM-Project-7035-1658845620

Project Demo Link:

https://drive.google.com/file/d/1czFG2lEquzTjOMJ1wqObRGRTy6ZWrLTo/view?usp=drivesdk