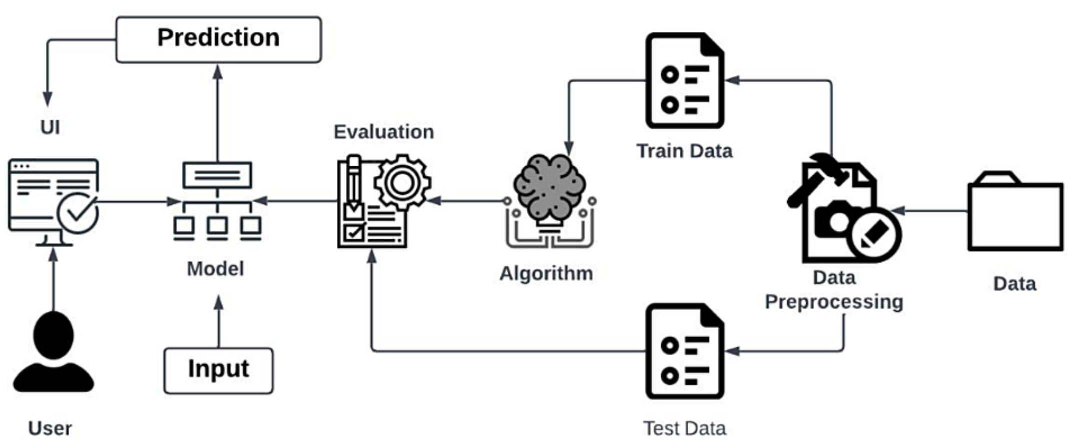
1. INTRODUCTION

1.1 Project Overview

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. At banks all throughout the world, the method of evaluating bank credit risk is recognised. "As we all know, determining the risk level is highly important and is done using a variety of ways. Credit risk management is another one of the banking industry's key responsibilities.

One of the challenging challenges for every bank is the prediction of credit defaulters. However, by predicting the loan defaulters, the banks will undoubtedly be able to cut their loss by decreasing their non-profit assets, allowing for the loss-free recovery of sanctioned loans, which can act as a contributing factor to the bank statement. This highlights the significance of studying this loan approval forecast. The prediction of this kind of data makes use of machine learning techniques, which are both essential and valuable.

The classification algorithms Decision tree, Random forest, KNN, and XG-boost will be used. With these algorithms, we will train and evaluate the data.



1.2 Purpose

- List all the technical benefits that the students would receive if they finished the project.
- Knowledge of Machine Learning Algorithms.
- Knowledge of Python Language with Machine Learning
- You'll be able to determine whether the classification challenge is a regression problem or another type of problem.
- You will be able to understand the various data pre-processing approaches and how to pre-process/clean the data.
- Utilising various algorithms based on the dataset and 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 the Decision tree each node represents a feature (attribute), each link (branch) represents a decision (rule) and each leaf represents an outcome (categorical or continuous value). Using different data analytics tools loan prediction and their 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 depending 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 represent 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 the 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 they have a probability of not paying back. Most of the Time, Applicants with high income sanctioning low amounts are more likely to 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 a system that 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 the 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 behavior. The most influencing attribute can be considered while a new customer approaches the bank for a loan and thus we can identify the potential of the 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 selectionmethods 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 the 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 needs to be taken by bank officers to determine whether to approve loan applicants or not. Normally classification techniques solved the problem up to an extent. Now the experiment proved that a model that uses feature selection before classification can help the bank officers to make proper decisions 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 has 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 the 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 the training datasets are trained using the decision tree algorithm and a prediction model is developed using the algorithm. Testing datasets are then given to models 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 a 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 one of the machine learning algorithms. Here the author used a decision tree algorithm for development. Based on the segregated value the decision tree is able to work and predict the loan approval. Here the author is able to conclude that the Decision tree version is extraordinarily 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 models 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, analyze, 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 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. Customers first apply for a 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 whether the person who applies for the loan will repay the loan or not. Sometimes it happens that a customer has provided partial data to the bank, in this case person may get the loan without proper verification and the bank may end up with a 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

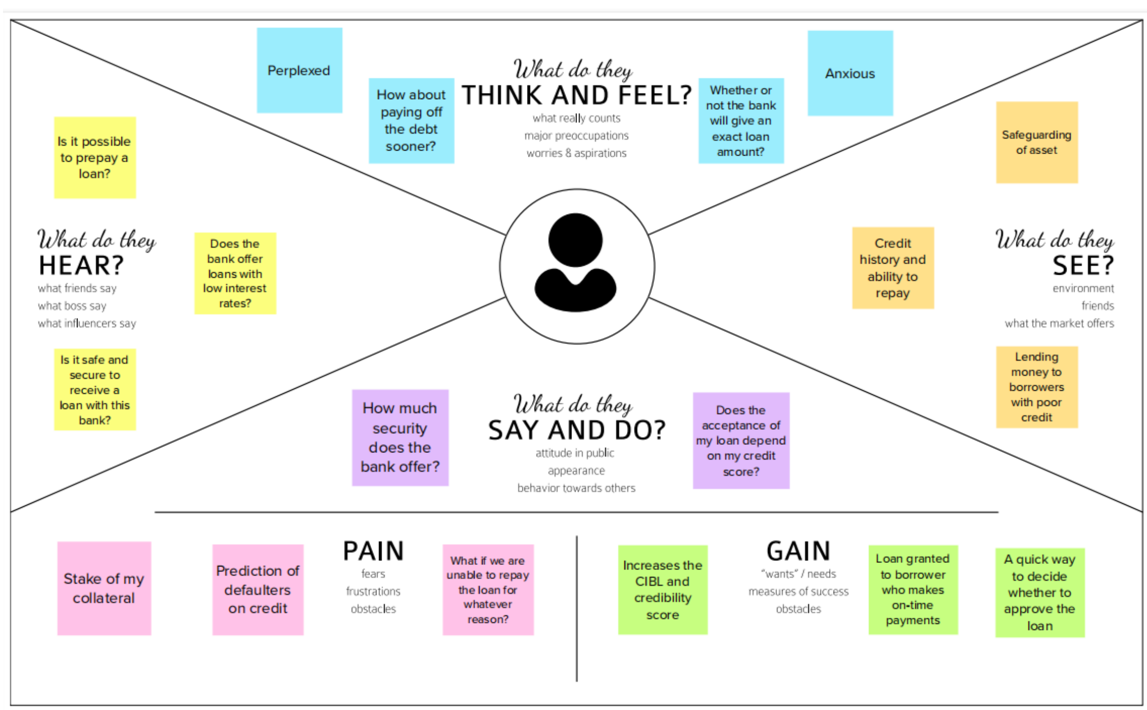
6. AFRAH KHAN, EAKANSH BHADOLA, ABHISHEK KUMAR and NIDHI SINGH - LOAN APPROVAL PREDICTION MODEL A COMPARATIVE ANALYSIS | Advances and Applications in Mathematical Sciences Volume 20, Issue 3, January 2021

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:

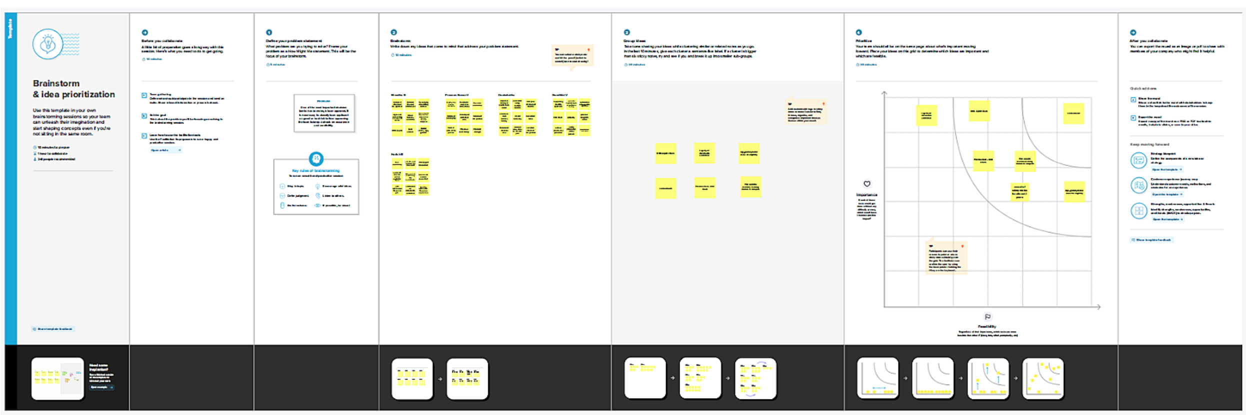
3.1 Empathy Map Canvas :



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

https://github.com/IBM-EPBL/IBM-Project-2625-1658478416/blob/main/Project%20Design%20%26%20Planning/Ideation%20Phase/Empathy%20Map.pdf

3.2 Ideation & Brainstorming :



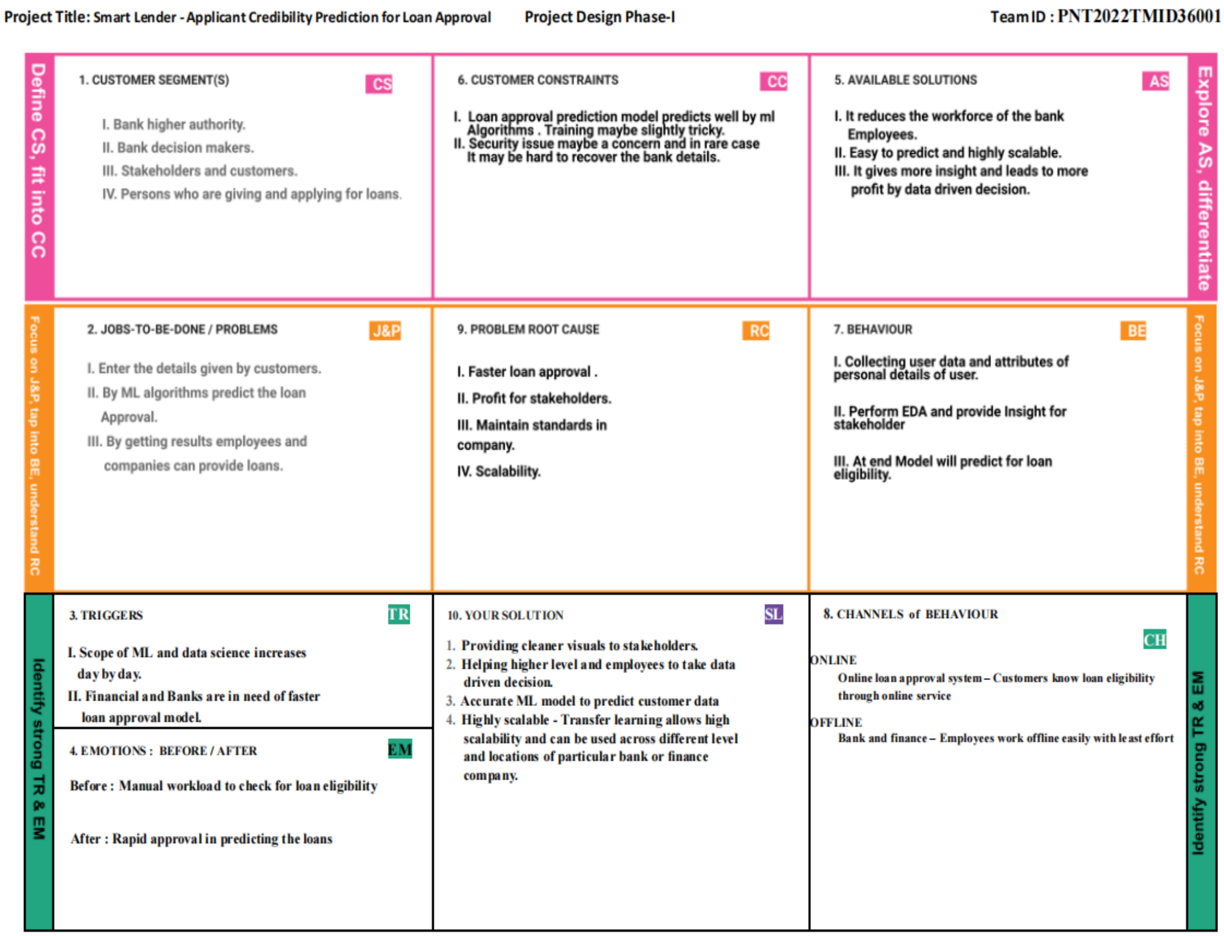
For clear view of the image click the below link: (ctrl+click)

https://github.com/IBM-EPBL/IBM-Project-2625-1658478416/blob/main/Project%20Design%20%26%20Planning/Ideation%20Phase/Brainstorming.pdf

3.3 Proposed Solution :

S.No.	Parameter	Description
1	Problem Statement (Problem to be solved)	The greatest issue now facing by the banking sector is client loan default. Consequently , the bank decides if a customer is qualified for a loan after the apply. However , doing this manually takes a long time . It proposes to use customer data to automate the loan qualification process in real time.
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	1. Avoid falling for fake offers and plans. This helps to avoid the unauthorized loan and help to black-list them. 2. This improve the money flow in efficient way.
5	Business Model (Revenue Model)	1. Consider Your needs and choose Your Loan Amount. This helps in betterment of business profit. 2. This helps to provide the timely delivery of loan at effective manner. 3. Improving Client Experience

3.4 Problem Solution fit :



Project Title: Smart Lender - Applicant Credibility Prediction for Loan Approval | Project Design Phase-I | Team ID : PNT2022TMID36001

Define CS, fit into CC

1. CUSTOMER SEGMENT(S) CS

I. Bank higher authority.
II. Bank decision makers.
III. Stakeholders and customers.
IV. Persons who are giving and applying for loans.

6. CUSTOMER CONSTRAINTS CC

I. Loan approval prediction model predicts well by ml Algorithms . Training maybe slightly tricky.
II. Security issue maybe a concern and in rare case It may be hard to recover the bank details.

5. AVAILABLE SOLUTIONS AS

I. It reduces the workforce of the bank Employees.
II. Easy to predict and highly scalable.
III. It gives more insight and leads to more profit by data driven decision.

Explore AS, differentiate

Focus on J&P, tap into BE, understand RC

2. JOBS-TO-BE-DONE / PROBLEMS J&P

I. Enter the details given by customers.
II. By ML algorithms predict the loan Approval.
III. By getting results employees and companies can provide loans.

9. PROBLEM ROOT CAUSE RC

I. Faster loan approval .
II. Profit for stakeholders.
III. Maintain standards in company.
IV. Scalability.

7. BEHAVIOUR BE

I. Collecting user data and attributes of personal details of user.
II. Perform EDA and provide Insight for stakeholder
III. At end Model will predict for loan eligibility.

Focus on J&P, tap into BE, understand RC

Identify strong TR & EM

3. TRIGGERS TR

I. Scope of ML and data science increases day by day.
II. Financial and Banks are in need of faster loan approval model.

4. EMOTIONS : BEFORE / AFTER EM

Before : Manual workload to check for loan eligibility

After : Rapid approval in predicting the loans

10. YOUR SOLUTION SL

1. Providing cleaner visuals to stakeholders.
2. Helping higher level and employees to take data driven decision.
3. Accurate ML model to predict customer data
4. Highly scalable - Transfer learning allows high scalability and can be used across different level and locations of particular bank or finance company.

8. CHANNELS of BEHAVIOUR CH

ONLINE
Online loan approval system – Customers know loan eligibility through online service

OFFLINE
Bank and finance – Employees work offline easily with least effort

Identify strong TR & EM

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

https://github.com/IBM-EPBL/IBM-Project-2625-1658478416/blob/main/Project%20Design%20%26%20Planning/Project%20Design%20Phase%201/Problem-solution-fit.pdf

4. REQUIREMENT ANALYSIS

4.1 Functional requirements :

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 Requirement	Using the credit score to determine loan eligibility and make loan approval predictions.
FR-2	User Registration	Using a mobile number or a Gmail account, the user logs in or registers.
FR-3	User Confirmation	Sending an OTP to a user's phone number or email address will serve as confirmation.
FR-4	User Login	To login, enter the user's email address and password.
FR-5	Loan Approval	For a statement, credibility information and supporting documentation must be submitted.
FR-6	Result	1.If approved, it displays the credit score as well as information on what needs to be done next. 2. If not approved, it explains why you were denied and are not eligible for the 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

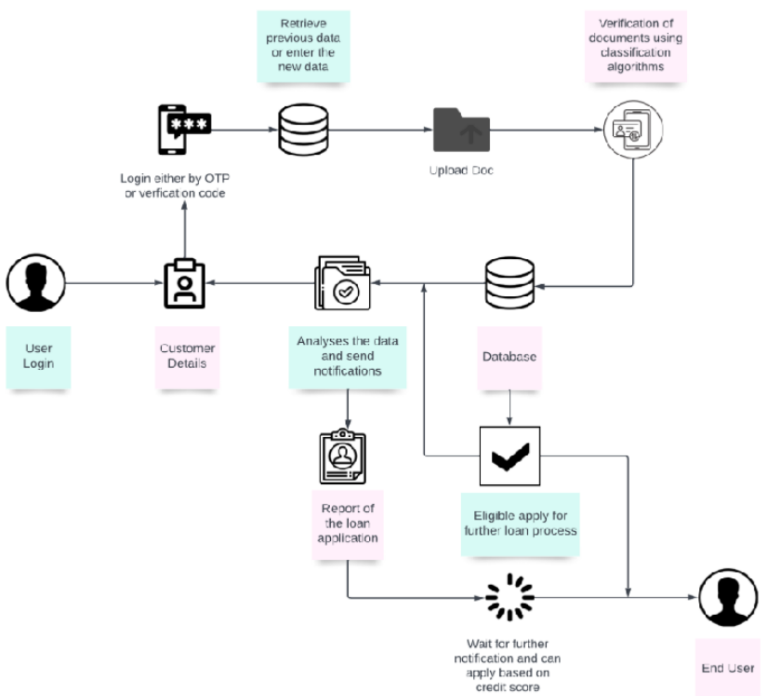
FR No.	**Non-Functional Requirement**	**Description**
NF R-1	**Usability**	If a consumer is qualified, they should be able to get a response from the bank acknowledging their loan application within seven days. If not, then the reason for rejection should be disclosed to the applicant.
NF R-2	**Security**	Checks to see whether the customer has a history of fraud and that no third-party apps have stolen any of their data.
NF R-3	**Reliability**	Consumers should have solid credit and a consistent source of income.
NF R-4	**Performance**	The model can be trained using various machine learning techniques, which can improve system performance.
NF R-5	**Availability**	For people with high incomes and those who pledge to repay the large amount in a short amount of time, the loan will be easily accessible.

NF R-6	**Scalability**	The customer must be older than 21. Moreover, based on how well the consumer can manage this.

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-2625-1658478416/blob/main/Project%20Design%20%26%20Planning/Project%20Design%20Phase%202/Data%20Flow%20Diagrams%20and%20User%

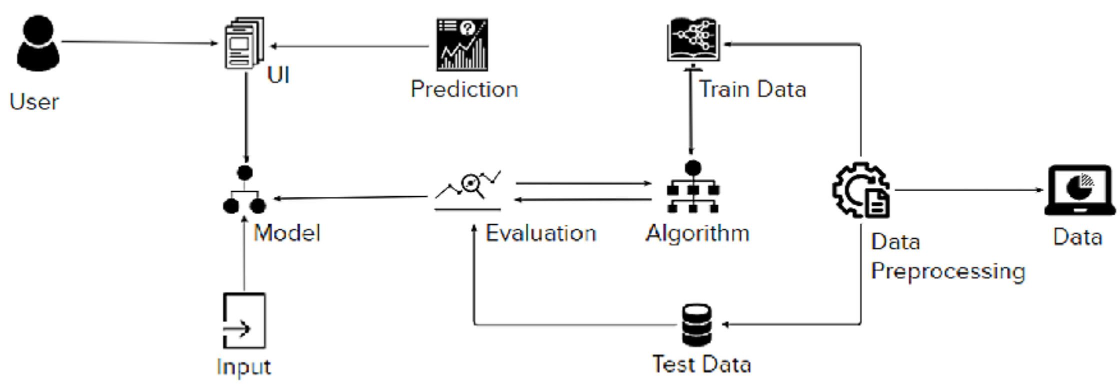
[20Stories.pdf]

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 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.	As a user can get OTP or verification code	Low	Sprint-1
	Login	USN-4	Users can log into the 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 care executive		USN-7	Checks the user feedbacks and provide essential technical support	Access the account/ able to access the dashboard	Medium	Sprint-2

Loan approval Executive	Automated analysis of cibil-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 loan prediction based on the details in loan application	High	Sprint-3
		USN-9	As a admin cibil score which represents credit history plays major role in analysis	Cibilscore /credit history plays major role	High	Sprint-3
Admin	Login/Register	USN-10	As an admin I should be able to login with a unique email and password.	Able to get logged in	High	Sprint-4
	Dashboard	USN-11	As an admin I need the access of full authority towards the dashboard.	Aceess the dashboard	Medium	Sprint-4

5.2 Solution & Technical Architecture :



6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation :

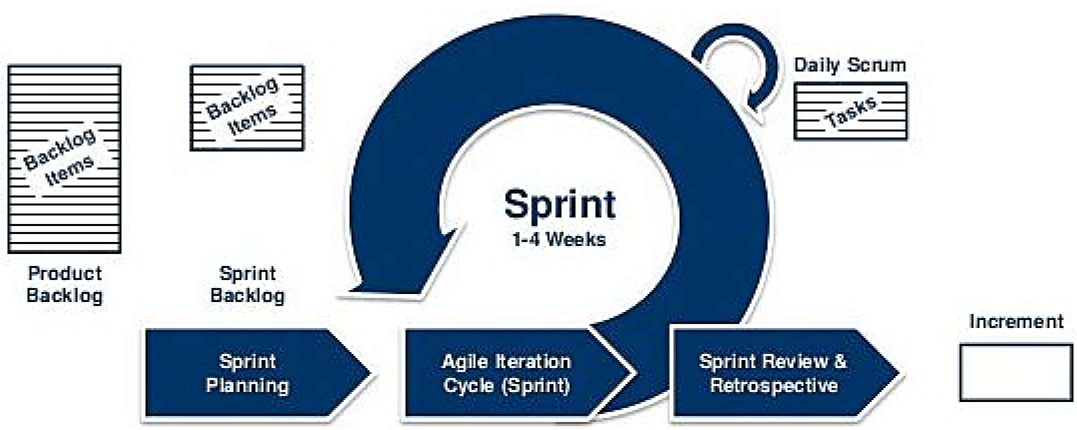
Activity List :

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 :

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	2	High	Praveen Kumar V Deekshetha
Sprint-1	Registration	USN-2	As a user, I will receive confirmation email once I have registered for the application	1	Medium	Nandhini V
Sprint-1	Registration	USN-3	As a user, I can register for the application through Facebook	1	Medium	Nivedha N
Sprint-1	Registration	USN-4	As a user, I can register for the application through Gmail	1	Medium	Harish B
Sprint - 2	Login	USN-5	As a user, I can log into the application by entering email & password	1	Medium	Deekshetha
Sprint - 2	Dashboard	USN-6	As a user, I should be able to experience a smooth user interface	1	Medium	Nivedha N Nandhini V
Sprint - 2	Dataset Collection and preprocessing	USN-7	All necessary and legal data for loan approval check should be collected and preprocessed	2	High	Praveen Kumar V Harish B Deekshetha
Sprint - 3	Model building and training	USN-8	Suitable model that could be trained and tested with accurate performance has to built	2	High	Nivedha N Praveen Kumar V Nandhini V
Sprint - 3	Model testing and Prediction	USN-9	Model that is built should be tested, predicted and accuracy is noted. Improvements in accuracy should also be made.	2	High	Harish B Deekshetha
Sprint - 4	Integration	USN-10	Integrate frontend and backend using flask server and deploy in IBM cloud	1	Medium	Praveen Kumar V Harish B
Sprint - 4	Testing	USN-11	Overall application is tested for deployment	2	High	Nivedha N,Deekshetha Nandhini V

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

Velocity :

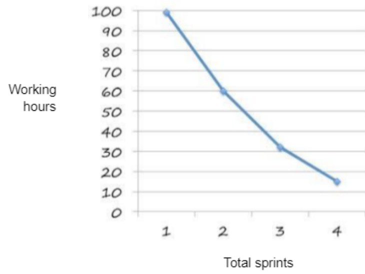
Imagine we have a 10-day sprint duration, and the velocity of the team is 20 (points per sprint). Let’s calculate the team’s average velocity (AV) per iteration unit (story points per day)

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

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 such as 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

```
<!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 style="background-image: url('https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcSUca_8lCQ4xwKf2ox9oPMwRbe2M_0DP8b13w&usqp=CAU');background-repeat: no-repeat;background-size: cover;>
    <main>
      <div class="mail">
        <center>
          <h1>Loan Credibility Prediction</h1>
          <h3>Know your eligibility in one press of a button</h3>
          <h5>
            Click below to know your eligibility
          </h5>
```

```
        <div class="container">
         <form  action="{{url_for('sendRegistrationPage')}}" method="GET">
            <input type="submit" value="Check">
          </form>
        </div>
      </center>
    </div>
  </main>
 </body>
</html>
```

index.html

```
<!DOCTYPE html>
<html>
<head>
        <style type="text/css">
        .header{

                position: absolute;
                left: 10px;
                top: 0px;
                width: 100%;
                height: 100px;
                text-align: center;

                text-transform: capitalize;
        }
        .reg_img
{
  height: 1000px;
  margin-top: 0px;
```

```
}
.box2
{
  height: 960px;
  width: 600px;
  margin: 70px auto;

  color: black;

}
    .
    .reg{
         margin-top: 130px;
         width: 50%;
         text-align: center;
         text-decoration-style: smooth;
         font-family: verdana;
         color: white;
         text-transform: capitalize;

         border-radius: 10px;
         overflow: scroll;
}

.sco table{
         text-overflow: scroll;
}



         </style>
</head>
<head>
```

```
    <title>Loan Registration</title>
</head>
<body style="background-image: url('https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcSUca_8lCQ4xwKf2ox9oPMwRbe2M_0DP8b13w&usqp=CAU');background-repeat: no-repeat;background-size: cover;>
<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</h1>


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

output.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;


    }

    .app {
        height: 500px;
        width: 600px;

        margin: 100px auto ;
```

```
                color: black;

        }
        </style>
        <title> Classification results</title>
</head>
<body style="background-image: url('https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcSUca_8lCQ4xwKf2ox9oPMwRbe2M_0DP8b13w&usqp=CAU');background-repeat: no-repeat;background-size: cover;>
        <div>
                <header class="header">

                        <h1>Application Status </h1>

                </header>
        </div>
        <div class="log_img">
        <div class="app">
                <section>
        <table align="center">
                <th>

        <table align="center">
       <th>

        {% if output== 1 %}
    <br><br>
     <h1 style="text-align: center; font-size: 35px;">Congratulations, You are eligible for the Loan !</h1>

    <br><br>
```

```
    {% elif output==0 %}
        <h1 style="text-align: center; font-size: 35px;">Sorry, you did not match our eligibility criteria !</h1>

        {% endif %}
    </th>
    </table>

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

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

</body>
</html>
```

7.2 Feature 2

app.py

```
import flask
import joblib
import numpy as np
from flask import render_template, request
from flask_cors import CORS

import smtplib
import requests
```

```
import json

import pyrebase


firebaseConfig = {
  'apiKey' : "AIzaSyBwc1_f3i-y3W2ClhX6BimI3rJd9YrbDjw",
  'authDomain' : "smart-lender-1639b.firebaseapp.com",
  'projectId' : "smart-lender-1639b",
  'storageBucket' : "smart-lender-1639b.appspot.com",
  'messagingSenderId' : "89081040576",
  'appId' : "1:89081040576:web:9e721303632308e2646725",
  'measurementId' : "G-F3C4GXSCC0",
  'databaseURL' : "none"
}


firebase=pyrebase.initialize_app(firebaseConfig)
auth=firebase.auth()

def sendmail(receiver):
    gmail_user = 'mailpraveen927@gmail.com'
    gmail_password = 'xxxxxxxx'
    sent_from = gmail_user
    to = [receiver]
    from_ = "From: {}".format(sent_from)
    to_ = "To: {}".format(receiver)
    subject = "Subject: Smart Lender Verification Email"
    body = "Thanks for using our service"
    message = from_ + "\n" + to_ + "\n" + subject + "\n" + body


    try:
        smtp_server = smtplib.SMTP_SSL('smtp.gmail.com', 465)
        smtp_server.ehlo()
        smtp_server.login(gmail_user, gmail_password)
        smtp_server.sendmail(sent_from, to, message)
```

```
        smtp_server.close()
        print ("Email sent successfully!")
    except Exception as ex:
        print ("Something went wrong….",ex)

def login(email,password):
   try:
            login = auth.sign_in_with_email_and_password(email,password)
            print("Successfully logged in!")
            return 1
   except Exception as e:
            t = json.loads(e.args[1])
            if(t['error']['errors'][0]['message']=="EMAIL_NOT_FOUND"):
                        return signup(email,password)
            return 0

def signup(email,password):
    try:
        user = auth.create_user_with_email_and_password(email,password)
        sendmail(email)
        return 1

    except Exception as e:
        #t = json.loads(e.args[1])
        #print(t['error']['code'])
        print(e)
        return 0



def submit(email,password):
    if(login(email,password)):
        return 1
    else:
        print("Problem")
        return 0
```

```
# NOTE: you must manually set API_KEY below using information retrieved from your IBM Cloud account.
API_KEY = "zJXNS7RuxksoVPyMjZDER8bqMSbF3OvzzuZZWw5FnpMU"
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}

app = flask.Flask(__name__,template_folder='template')
CORS(app)



@app.route('/', methods=['GET'])
def sendLoginPage():
    return render_template('login.html')
@app.route('/home', methods=['POST'])
def sendHomePage():
   if request.method == 'POST':
            email = request.form['email']
            password = request.form['password']
            if(submit(email,password)):
                    return render_template('home.html')
   return render_template('login.html')

@app.route('/index', methods=['GET'])
def sendRegistrationPage():
    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']
```

```python
        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()
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([[0,gender, married, dependat,education,employ,annual_income,co_income,Loan_amount,Loan_amount_term,credit,proper]])
    x =[0,gender, married, dependat,education,employ,annual_income,co_income,Loan_amount,Loan_amount_term,credit,proper]
    # model = joblib.load('R.pkl')

    # ans = int(model.predict(x)[0])
    # if (ans==1):
    #     print("You are eligible. Kindly wait for further notice")
    # else:
    #     print("Your application status did not match our criteria")

    payload_scoring = {"input_data": [{"field": ["","Gender","Married","Dependents","Education","Self_Employed","ApplicantIncome","CoapplicantIncome","LoanAmount","Loan_Amount_Term","Credit_History","Property_Area"], "values": [x]}]}

    response_scoring = requests.post('https://us-south.ml.cloud.ibm.com/ml/v4/deployments/77f857ff-c248-41f7-b24f-2a3128ac1906/predictions?version=2022-11-22', json=payload_scoring,
    headers={'Authorization': 'Bearer ' + mltoken})
    print("Scoring response")
    pred = response_scoring.json()
    p = pred['predictions'][0]['values'][0][0]
    if(p == 0):
        print("Your application status did not match our criteria")
    else:
        print("You are eligible. Kindly wait for further notice")

    return render_template('output.html', output=p)

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 testing and training of the model we applied. The information submitted by users can be used to check this eligibility.

This covers specifics such as

- 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

8.2.1 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).

8.2.2 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

8.2.3 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 the Random Forest 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

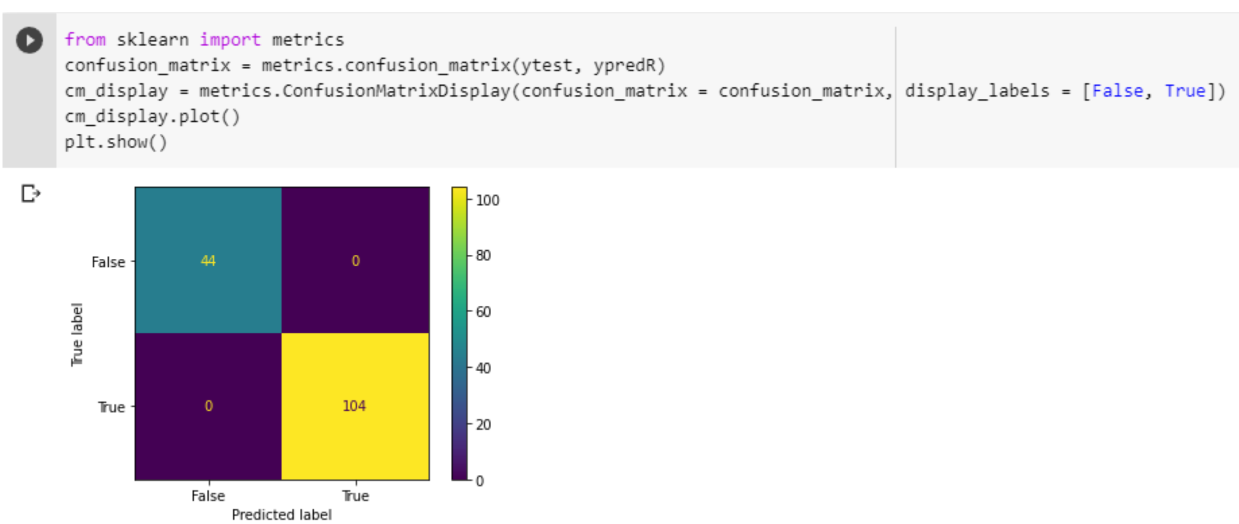


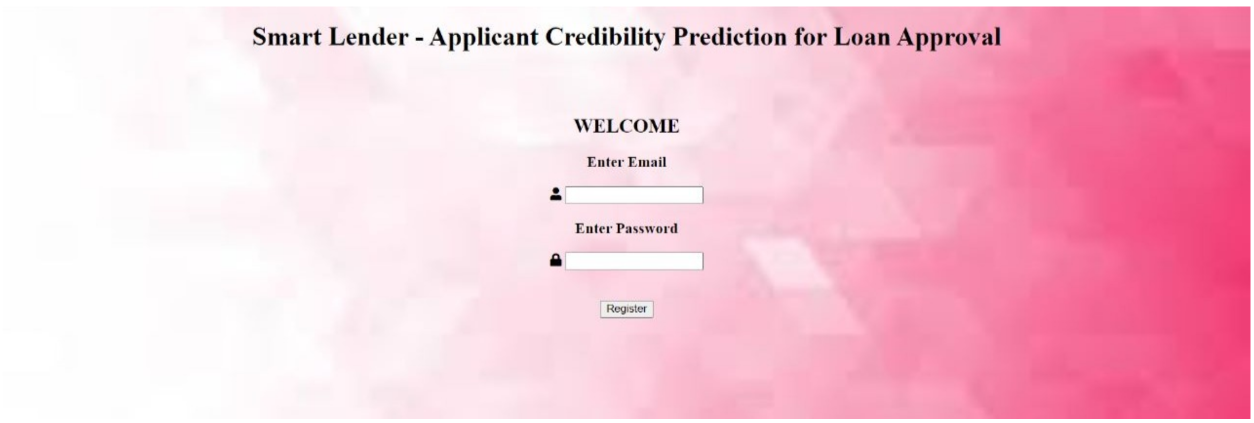
Fig 1 - Metrics

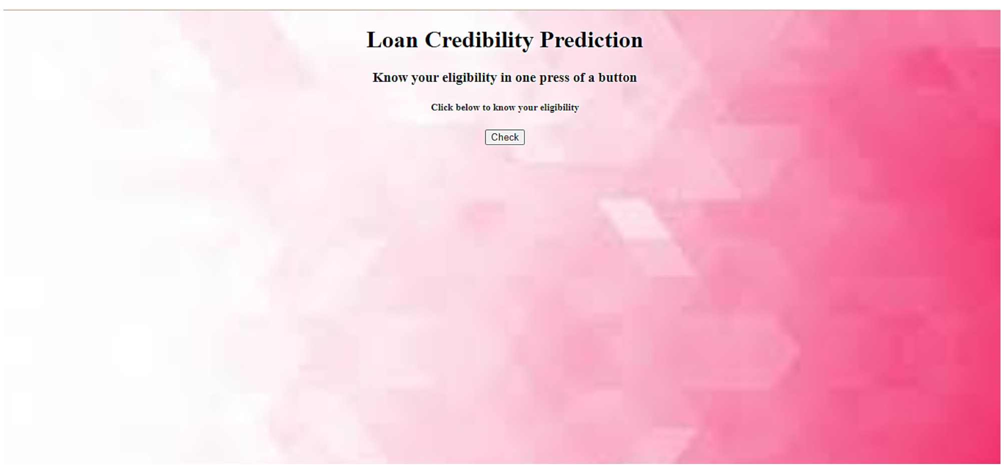
```
accuracy = metrics.accuracy_score(ytest, ypredR)
accuracy
```

```
1.0
```

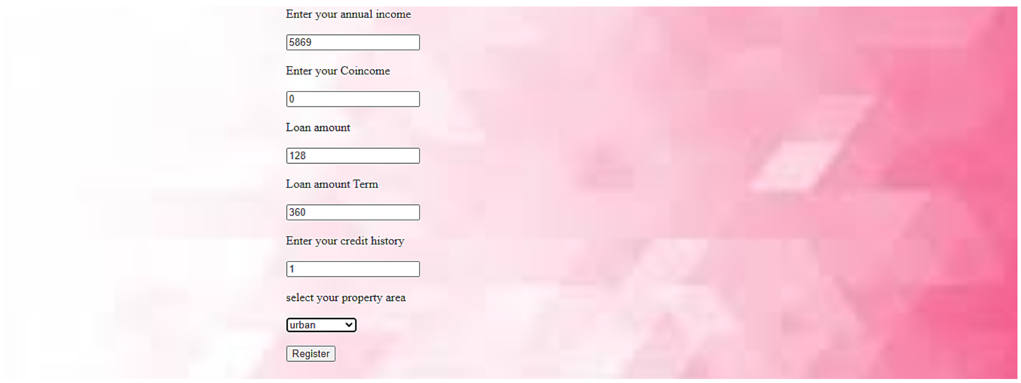
Fig 2 - Accuracy Score

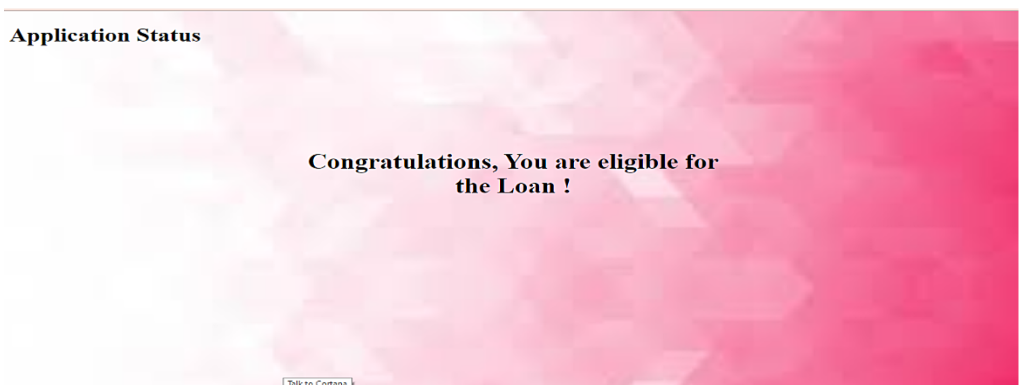
9.2 Screenshots :

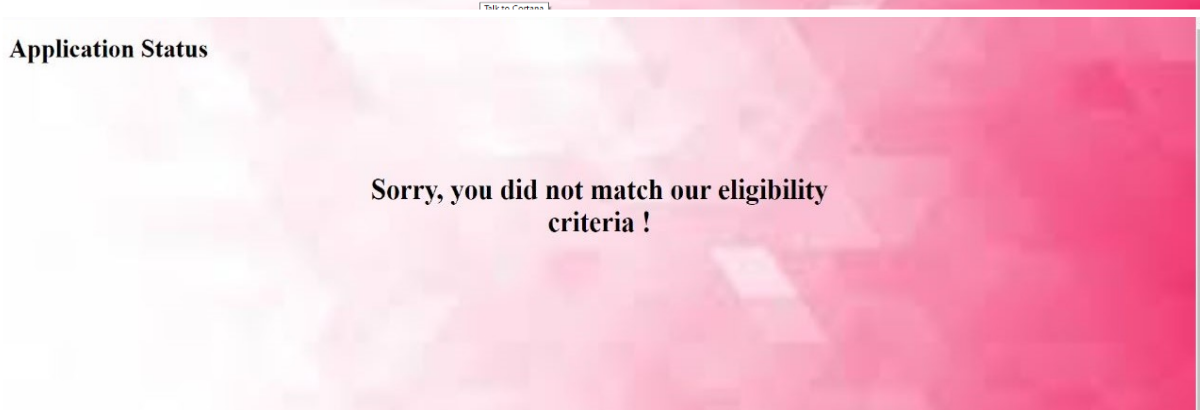





Smart Lender - Applicant Credibility Prediction for Loan Approval
APPLICANT DETAILS
Select Gender
Male Female
Status
Married single
Enter number of dependants
3
Education Level
Graduate
Employment status
Self Employed


Enter your annual income
5869
Enter your Coincome
0
Loan amount
128
Loan amount Term
360
Enter your credit history
1
select your property area
urban
Register


Application Status
Congratulations, You are eligible for the Loan !


Application Status
Sorry, you did not match our eligibility criteria !

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
- Predictions can occasionally be inaccurate since the model was created using outdated data.
- The outcome of the prediction is more dependent on the model.

11. CONCLUSION :

We identified the key elements that have the greatest impact on the loan approval status. The performance accuracy of these most crucial features applied to a few chosen algorithms is then compared to the case where all features were used. The model can aid banks in determining which elements are crucial for the loan approval process.

Based on its accuracy, the comparative analysis reveals which algorithm will be the best and excludes the others. We have created a model that accurately predicts whether a borrower will repay their loan or not. We can observe that the bankers' efforts have been lowered by our approach. This model's ability to produce exact findings has been greatly aided by machine learning.

12. FUTURE SCOPE :

Future improvements to this research will involve teaching bots how to forecast which locations will qualify for loans using machine learning methods. Advanced machine learning concepts can be employed for better prediction because they are comparable to data mining. For better prediction, the data privacy, dependability, and accuracy can be increased.

We consider that crime data mining has a bright future for improving the effectiveness and efficiency of criminal and intelligence analysis based on the encouraging outcomes. For the pattern of loan credibility, visual and intuitive criminal and intelligence investigation approaches can be created. We can use additional data mining techniques, such as classification, in the same way that we used machine learning for loan prediction. Additionally, we may conduct analyses on a variety of datasets, including the enterprise survey dataset and the poverty and aid effectiveness datasets.

13. APPENDIX :

SourceCodelink:

GitHub Link:

https://github.com/IBM-EPBL/IBM-Project-2625-1658478416

ProjectDemoLink:

GitHub Link:

https://github.com/IBM-EPBL/IBM-Project-2625-1658478416/tree/main/Final%20Deliverables/demo%20video