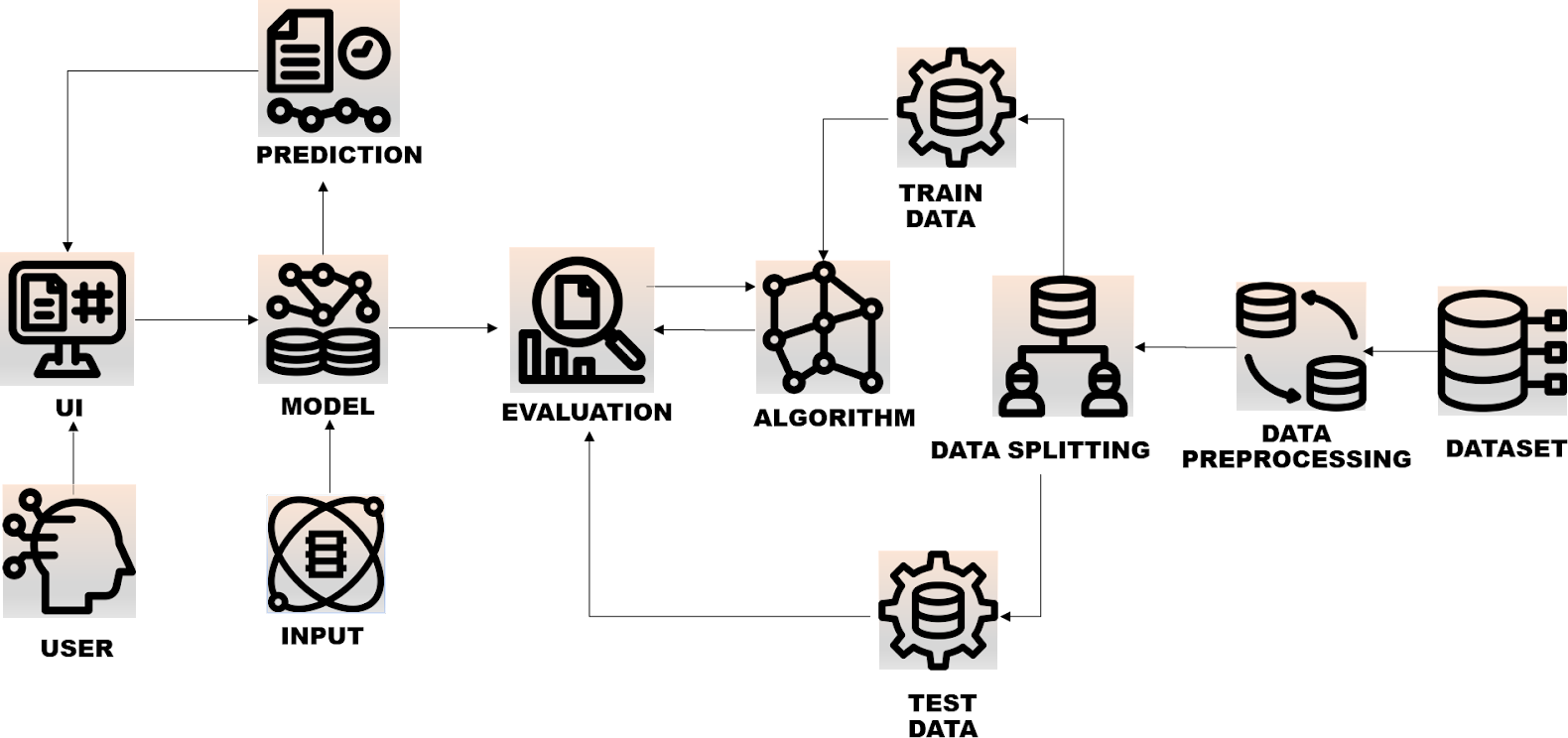
# 1.INTRODUCTION

* 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.



# 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.

# 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.

* + 1. 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.
    2. **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.
    3. 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.

# 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.

# 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.

# 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

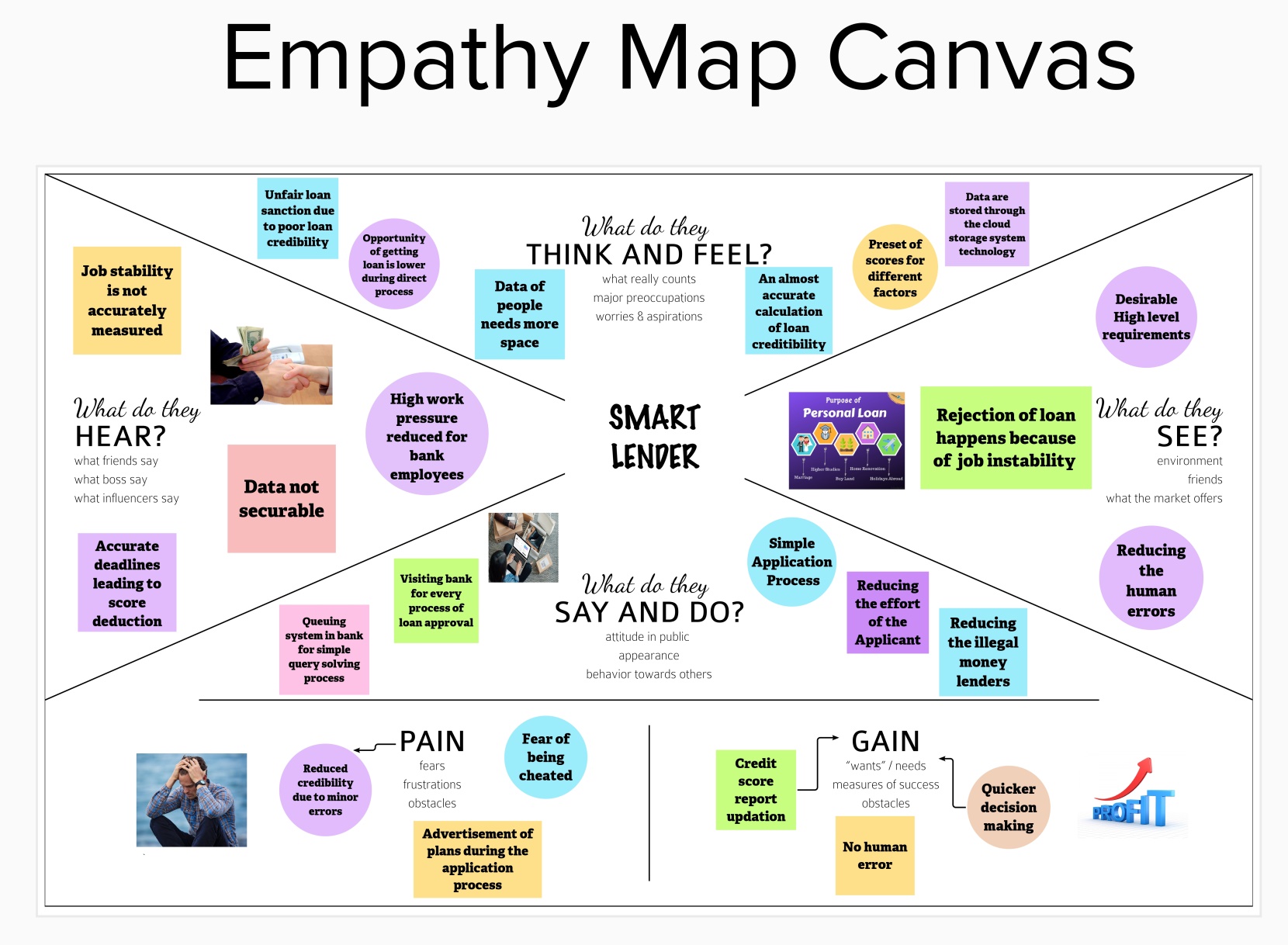
* + 1. Loan default prediction using decision trees and random forest: A comparative study-IOP Conference Series Mehul Madaan et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1022 012042
    2. AFRAH KHAN, EAKANSH BHADOLA, ABHISHEK KUMAR and NIDHI SINGH - LOAN APPROVAL PREDICTION MODEL A COMPARATIVE ANALYSIS | Advances and Applications in Mathematical Sciences Volume 20, Issue 3, January 2021

# 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.

# 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-7035-](https://github.com/IBM-EPBL/IBM-Project-7035-1658845620/blob/main/PROJECT%20DESIGN%20AND%20PLAN/IDEATION%20PHASE/Empathy%20map.pdf) [1658845620/blob/main/PROJECT%20DESIGN%20AND%20PLAN/IDEATION%](https://github.com/IBM-EPBL/IBM-Project-7035-1658845620/blob/main/PROJECT%20DESIGN%20AND%20PLAN/IDEATION%20PHASE/Empathy%20map.pdf)**

**[20PHASE/Empathy%20map.pdf](https://github.com/IBM-EPBL/IBM-Project-7035-1658845620/blob/main/PROJECT%20DESIGN%20AND%20PLAN/IDEATION%20PHASE/Empathy%20map.pdf)**

# Ideation & Brainstorming

****

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

**[https://github.com/IBM-EPBL/IBM-Project-7035-](https://github.com/IBM-EPBL/IBM-Project-7035-1658845620/blob/main/PROJECT%20DESIGN%20AND%20PLAN/IDEATION%20PHASE/BRAINSTORM%20AND%20IDEATION.pdf) 31658845620/blob/main/PROJECT%20DESIGN%20AND%20PLAN/IDEATION%20PHASE/BRAINSTORM%20AND%20IDEATION.p**

# 3.Proposed Solution:

The solution that is desired is focused on solving the problems that exist in the current situation and the parameters will describes the results that are expected.

**TEAMID : PNT2022TMID37097**

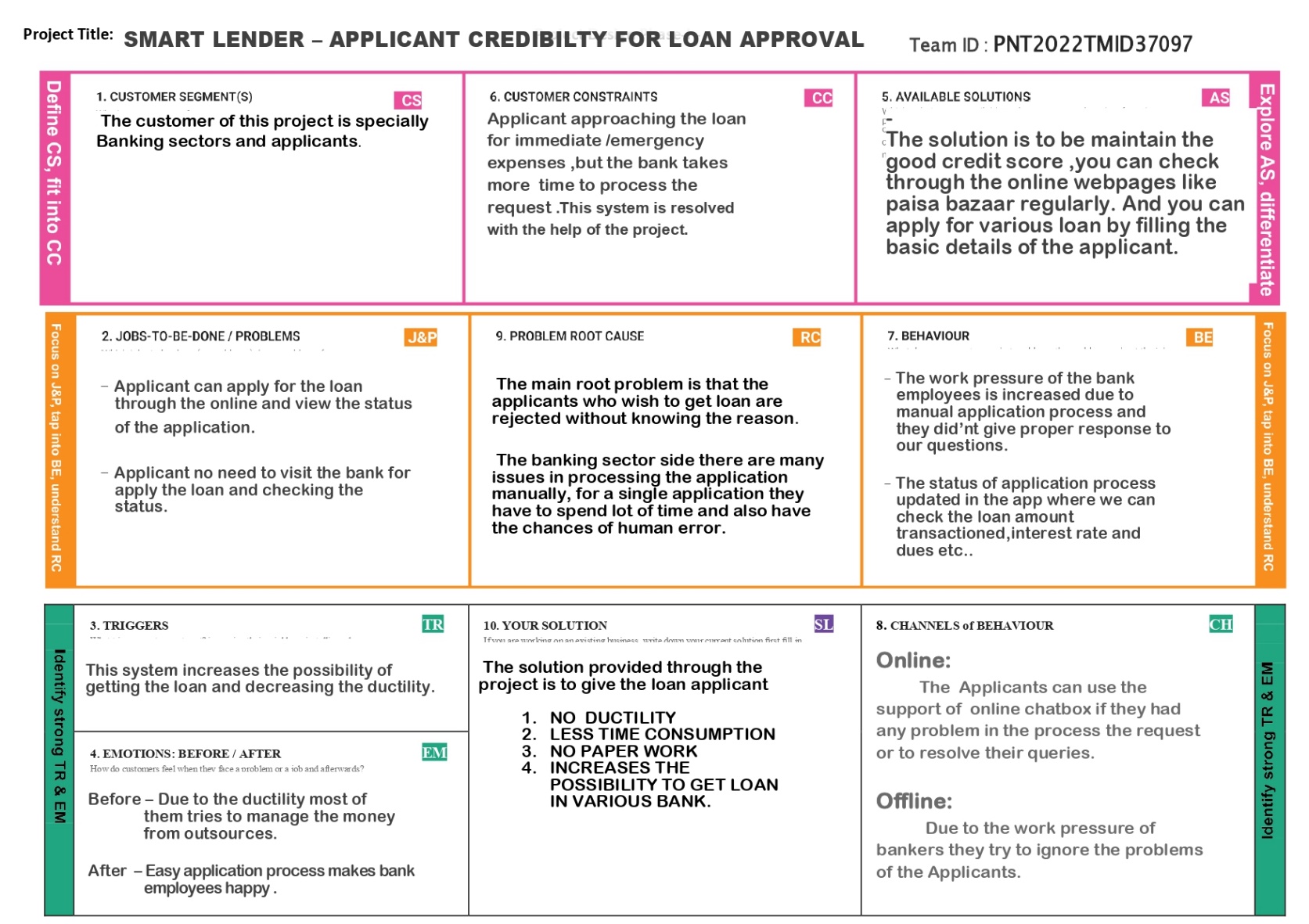
**PROJECT NAME : SMART LENDER-APPLICANT CREDIBILITY PREDICTION FOR**

**LOAN APPROVAL**

|  |  |  |
| --- | --- | --- |
| S.NO. | PARAMETERS | DESCRIPTION |
| 1) | Problems Statement | * Less accurate evaluation. * More resources are used. * More room for human errors. * High time consumption. * Customer service will be compromised. * Tracking can be difficult. |
| 2) | Solution | * More accurate evaluation. * Usage of resources will be reduced. * Reduces human errors. * Highly scalable and provide data driven decisions to stakeholder and higher authority. * Plenty of time will be saved. * Customer service will be improved. * Tracking gets easier.   We will be using classification algorithms such as Decision tree, Random Forest, KNN, and xgboost to achieve higher accuracy in predicting the model. We will train and test the data  with these algorithms, tune by hyperparameter tunning. From this the above ideas are implemented. |
| 3) | Novelty / Uniqueness | As soon as the essential data are provided, the model will predict whether to approve the loan or not - By use of transfer  learning. |
| 4) | Social Impact / Customer Satisfaction | One of the most important factors which affect our country’s economy and financial condition is the credit system governed by the banks. 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. |

|  |  |  |
| --- | --- | --- |
| 5) | Business Model (Revenue Model) | This model can be developed by minimum cost at the same time it will provide the peak performance, higher accuracy and the result will be more effective than traditional techniques. |
| 6) | Scalability of the Solution | Banks need not to go through the background verification process of the applicant by using this model. The model will predict the customers data and their attributes like salary, credit score, etc. |

**4.Problem Solution fit**

****

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

**[https://github.com/IBM-EPBL/IBM-Project-7035-](https://github.com/IBM-EPBL/IBM-Project-7035-1658845620/blob/main/PROJECT%20DESIGN%20AND%20PLAN/PROJECT%20DESIGN%20PHASE-I/PROBLEM%20SOLUTION%20FIT.pdf) [1658845620/blob/main/PROJECT%20DESIGN%20AND%20PLAN/PROJECT%20DESIGN%20PH](https://github.com/IBM-EPBL/IBM-Project-7035-1658845620/blob/main/PROJECT%20DESIGN%20AND%20PLAN/PROJECT%20DESIGN%20PHASE-I/PROBLEM%20SOLUTION%20FIT.pdf) [ASE-I/PROBLEM%20SOLUTION%20FIT.pdf](https://github.com/IBM-EPBL/IBM-Project-7035-1658845620/blob/main/PROJECT%20DESIGN%20AND%20PLAN/PROJECT%20DESIGN%20PHASE-I/PROBLEM%20SOLUTION%20FIT.pdf)**

# REQUIREMENT ANALYSIS

**Functional Requirements:**

Following are the functional requirements of the proposed solution.

|  |  |  |
| --- | --- | --- |
| **FR**  **No.** | **Functional Requirement (Epic)** | **Sub Requirement (Story / Sub-Task)** |
| FR-1 | Applicant Registration | Registration through Form Registration through Gmail |
| FR-2 | Applicant Confirmation | Confirmation via Email Confirmation via OTP |
| FR-3 | Applicant data export process | A Process where applicant admin take member data from bank to be analysed in web applicant |
| FR-4 | Applicant Data Management | Web portal admin is able to add, edit, or erase applicant data or user data |
| FR-5 | Bank Data Management | Web Portal admin is able to add, edit, or erase Bank data or eligibility criteria |
| FR-6 | Data Retention | * Proposed application system handle archival, retrieval, and retention of historical data. * Provide sufficient details concerning these |
| FR-7 | Compliance requirements | Application might need to comply with several regulations. These could include data privacy, information security, transparency, and various other aspects. Document them in detail. |

**Non-functional Requirements:**

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

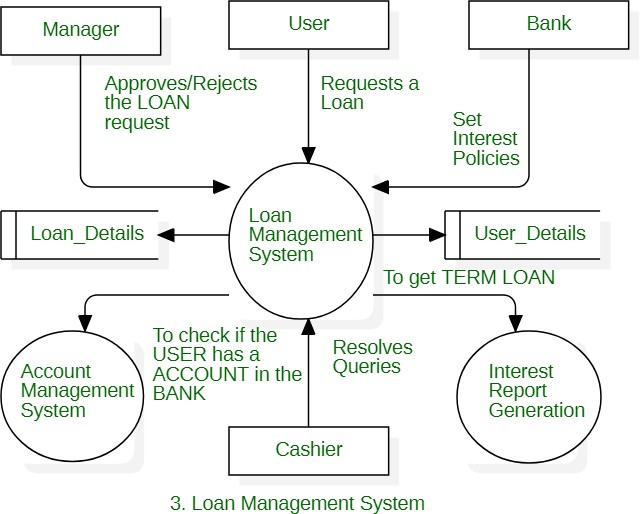
|  |  |  |
| --- | --- | --- |
| **FR No.** | **Non-Functional Requirement** | **Description** |
| NFR-1 | **Usability** | The error rate of users submitting their details at the checkout page mustn’t exceed 10 percent. |
| NFR-2 | **Security** | Authorization access scenarios and definitions, hand- over procedures for Student records between University |
| NFR-3 | **Reliability** | Physical and technical environment constraints including integration and portability concerns |
| NFR-4 | **Performance** | * Speed, reliability, and capacity of the system. * The landing page supporting 5,000 users per hour must provide 6 second or less response time in a |

|  |  |  |
| --- | --- | --- |
|  |  | Chrome desktop browser, including the rendering of  text and images and over an LTE connection. |
| NFR-5 | **Availability** | The web dashboard must be available to US and IND users 99.98 percent of the time every month during  business hours EST & IST. |
| NFR-6 | **Cultural and Political** | Cultural, political, and legal factors affecting the system including best-practice standards, government regulations, professional standards,  location specific policies and practices |
| NFR-7 | **Scalability** | * Horizontal scaling is provided by adding more machines to the pool of servers. * Vertical scaling is achieved by adding more CPU and RAM to the existing machines. |
| NFR-8 | **Portability** | A program running on Windows 10 must be able to run on Windows 11 without any change in its  behaviour and performance |

# 5.PROJECT DESIGN

# 5.1 DATA FLOW DIAGRAM

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



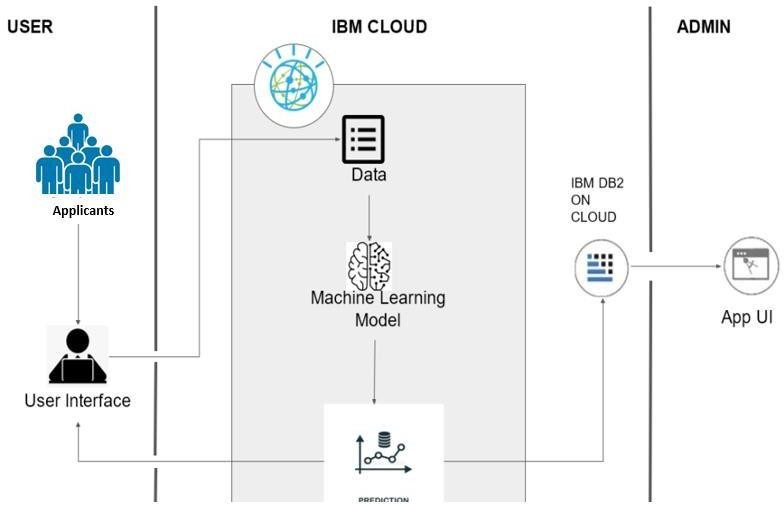
**Loan management system collects the data from the bank system regarding the applicants credit and activities in detail , which data are stored in the ibm cloud.**

***5.2 User Stories***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **User Type** | **Functional Requirement**  **(Epic)** | **User Story Number** | **User Story / Task** | **Acceptance criteria** | **Priority** | **Release** |
| Customer | Landing page | USN-1 | As a user, I can view the details about the bank | I can access the bank’s landing page | Medium | Sprint-1 |
|  |  | USN-2 | As a user, I can view the details and loans  offered by the bank | I can get the details of  bank | Medium | Sprint-1 |
|  |  | USN-3 | As a user, I can see the interests and other loan related information | I can see the description in detailed manner | Low | Sprint-1 |
|  |  | USN-4 | As a user, I can view the latest news of the  bank | I can ensure the loan  getting confidence | Medium | Sprint-1 |
|  |  | USN-5 | As a user, I can fill the contact form for queries | I can fill submit the contact form | Low | Sprint-2 |
|  |  | USN-6 | As a user, I can register for the application by  entering my email, password, and confirming my password | I can access my account / dashboard | High | Sprint-1 |
|  |  | USN-7 | As a user, I will receive confirmation email once I have registered for the application | I can receive confirmation email & click confirm | High | Sprint-2 |
|  |  | USN-8 | As a user, I can see testimonials of applicants who got loan in the bank | I can access the testimonials | Medium | Sprint-1 |
|  | Login | USN-9 | As a user, I can log into the application by entering email & password | I can login | High | Sprint-1 |
|  |  | USN-10 | As a user, I can see the loans offered by the bank | I can see detailed information | Medium | Sprint-2 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **User Type** | **Functional**  **Requirement (Epic)** | **User Story Number** | **User Story / Task** | **Acceptance criteria** | **Priority** | **Release** |
|  |  | USN-11 | As a user, I can check the required details to be needed | I can check required detail | High | Sprint-2 |
|  |  | USN-12 | As a user ,I can fill my profile for priority purpose | I can fill my profile | High | Sprint-3 |
|  |  | USN-13 | As a user, I can fill the applications of the loans required | I can update my application | Medium | Sprint 3 |
|  |  | USN-14 | As a user, I can see my credit score | I can see my credit score | Medium | Sprint-3 |
|  |  | USN-15 | As a user, I can see my application history | I can see my application history | Medium | Sprint-3 |
| Administrator | Landing page | USN-16 | As an administrator, I shall update the news about the bank | I can check if the update is reflected or not | High | Sprint-4 |

# Solution & Technical Architecture:



# PROJECT PLANNING & SCHEDULING 6.1Sprint Planning & Estimation:

|  |  |  |
| --- | --- | --- |
| **TITLE** | **DESCRIPTION** | **DATE** |
| Literature survey and information gathering | Collect the relevant information on project use case, refer the existing solution, technical papers, research publications,… | 23 September 2022 |
| Prepare empathy map | Prepare empathy map canvas to capture the user pains, and gains followed by preparation of problem statements | 23 September 2022 |
| Ideation | List the ideas by organising the brainstorm session and prioritize the top 3 ideas based on the  feasibility and importance | 23 September 2022 |
| Proposed solution | Prepare the proposed solution document , which includes the novelty , feasibility of idea , business model , social impact ,  scalability of solution , etc,.. | 26 September 2022 |
| Problem solution fit | Prepare problem solution fit document | 13 October 2022 |
| Solution architecture | Prepare solution architecture document | 13 October 2022 |
| Customer journey | Prepare customer journey maps to understand the user interactions and experience with the application | 17 October 2022 |
| Functional Requirement | Prepare functional requirement document | 14 October 2022 |
| Data flow diagram | Draw the data flow diagram and submit for review | 18 October 2022 |
| Technology Architecture | Prepare the technology architecture diagram | 18 October 2022 |
| Prepare milestone and activity list | Prepare milestone and activity list for the project | 2 November 2022 |
| Project development –Delivery of sprint-1,2,3,4 | Develop the code and submitting what we have done or completed | On progress |

**Activity List:**

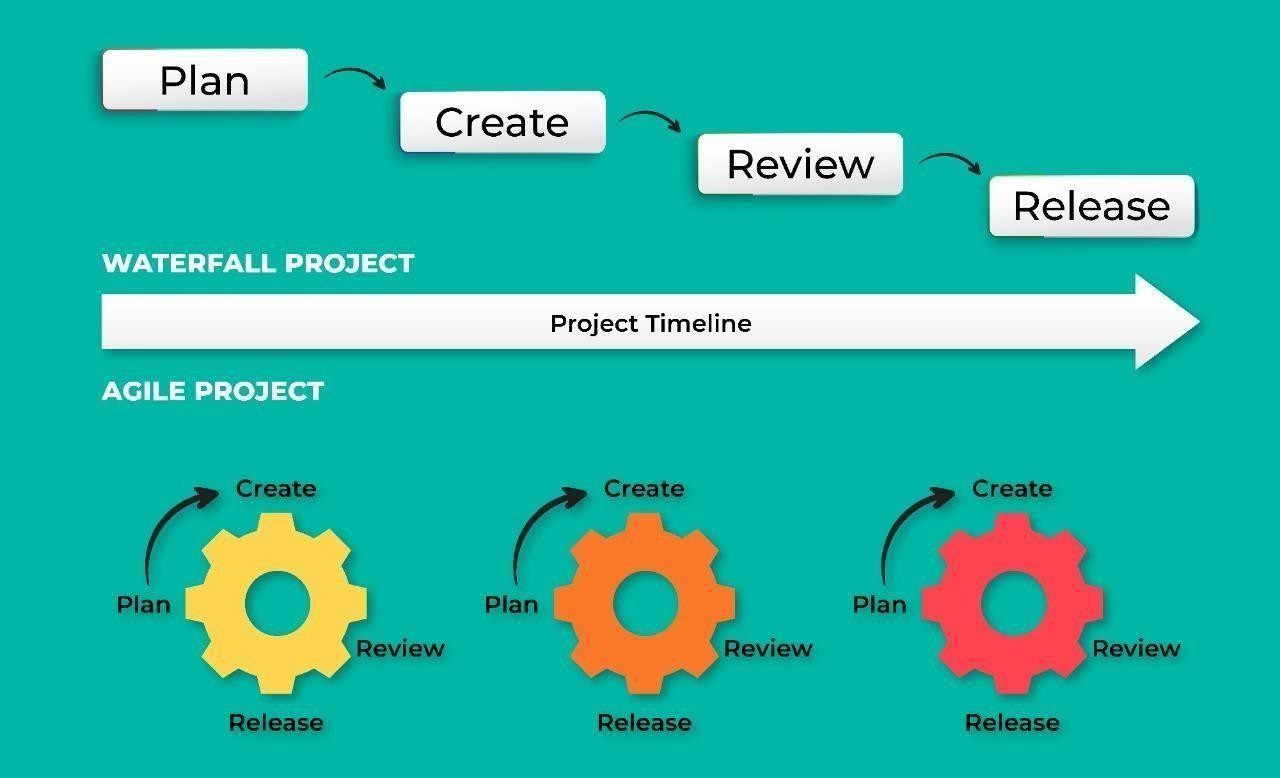
In Project Management Planning is an important task to scheduling the phrase of the project to the Team Member.

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***



# Sprint Delivery Schedule

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sprint** | **Functional requirement [epic]** | **User story number** | **User story/task/activity** | **Story points** | **Priority** | **Team members** |
| Sprint-1 | Registration | USN-1 | User can register for the application by entering his or her email, password, and confirming the password | 5 | High | Anandharaman, Karthik, Barathvaj, Brajesh choudhary, Abinesh kannan |
| Sprint-1 |  | USN-2 | User will receive conformation email or message once registered for  the application | 3 | High | Anandharaman, Karthik, Barathvaj, Brajesh choudhary |
| Sprint-1 | Login | USN-3 | Enter the username and password to login to the application | 2 | High | Anandharaman, Karthik |
| Sprint-2 | Dashboard | USN-4 | User can visualization of the loan given data for a specific time period | 3 | Medium | Anandharaman, Karthik, Barathvaj, Brajesh choudhary, Abinesh kannan |
| Sprint-2 |  | USN-5 | User can change his/her password and can view the account details and search history | 5 | High | Karthik, Barathvaj, Brajesh choudhary,  Abinesh kannan |
| Sprint-3 | Support | USN-6 | User can give the feedback on the accuracy of the prediction and on  the user interface | 5 | High | Karthik, Abinesh kannan |
| Sprint-3 |  | USN-7 | Responds to user queries via email | 2 | Medium | Anandharaman, Karthik |
| Sprint-3 |  | USN-8 | The team must respond immediately to the  queries based on the priority | 5 | High | Anandharaman, Karthik, Barathvaj,  Brajesh choudhary. |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Sprint-4 | Core Function | USN-9 | User can enter their details and informations | 8 | High | Anandharaman, Abinesh kannan |
| Sprint-4 |  | USN-10 | Prediction of eligibility and displaying of result | 2 | Medium | Anandharaman, Karthik, Barathvaj, Brajesh  choudhary |
| Sprint-4 |  | USN-11 | The website is response on all the device and the screen sizes | 5 | High | Anandharaman, Karthik, Barathvaj, Brajesh choudhary, Abinesh kannan |

**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)** | **Sprit Release Date(Actual)** |
| Sprint-1 | 10 | 6 Days | 30 Oct 2022 | 04 Nov 2022 | --- | 05 Nov 2022 |
| Sprint-2 | 07 | 5 Days | 03 Nov 2022 | 07 Nov 2022 | --- | 08 Nov 2022 |
| Sprint-3 | 12 | 6 Days | 08 Nov 2022 | 13 Nov 2022 | --- | 14 Nov 2022 |
| Sprint-4 | 15 | 5 Days | 14 Nov 2022 | 18 Nov 2022 | --- | 19 Nov 2022 |

# Reports from JIRA:

# CODING & SOLUTIONING

* 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>

<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>

## index.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;

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">

<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;

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>

<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>

<footer>

</footer>

</div>

</div>

</body>

</html>

* 1. **Feature 2 *app.py*** 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()

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()

# TESTING

* 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)

# User Acceptance Testing

### 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).

### 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 |

### 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 |

1. **RESULTS**
   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 |

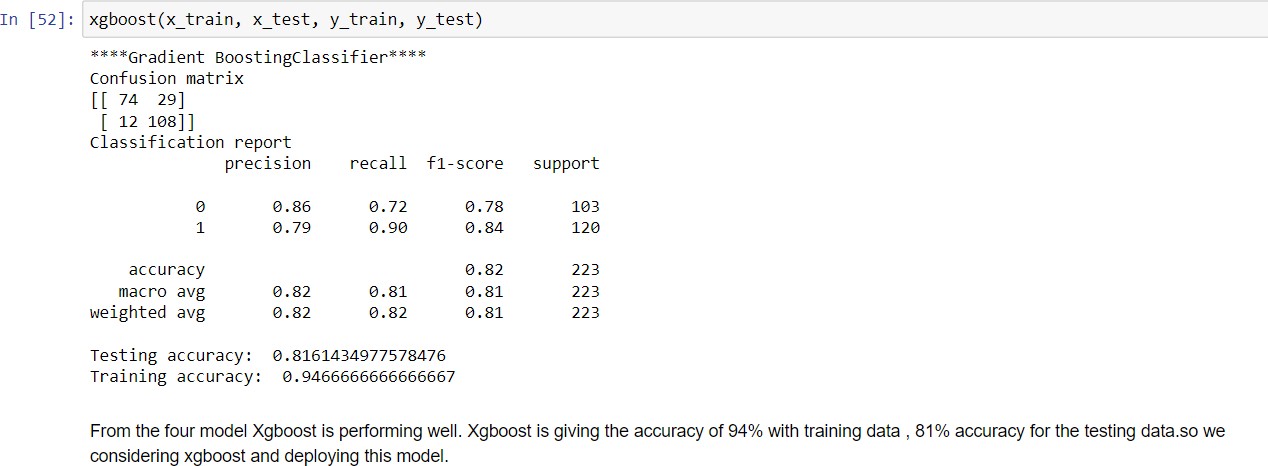


Fig 1 - Metrics

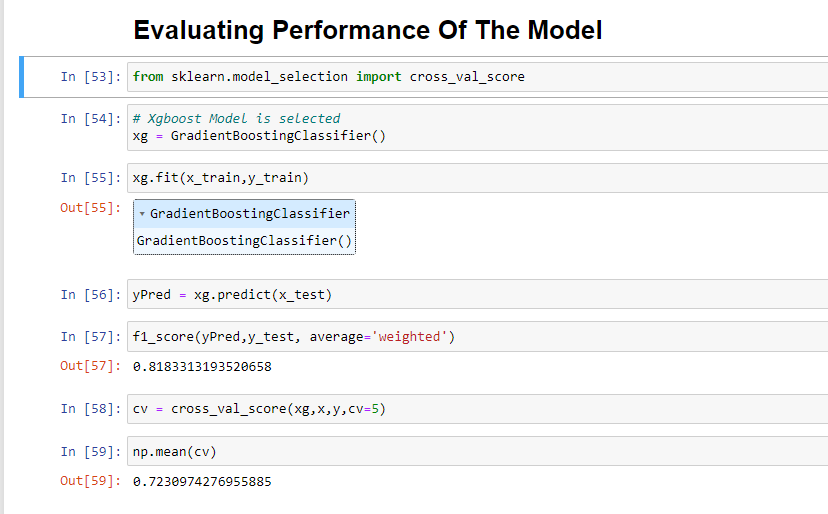


Fig 2 - Tune the Model

1. **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.

1. **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

1. **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.

1. **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=dr](https://drive.google.com/file/d/1czFG2lEquzTjOMJ1wqObRGRTy6ZWrLTo/view?usp=drivesdk)**

**[ivesdk](https://drive.google.com/file/d/1czFG2lEquzTjOMJ1wqObRGRTy6ZWrLTo/view?usp=drivesdk)**