Team Id: NT2022TMID21248

Project Title: Smart Lender- Applicant Credibility Prediction for Loan Approval

Project Report

1. INTRODUCTION

1.1 Project Overview

Despite the fact that our banking system has many products to sell, the main source of income for a bank is its credit line. So, they can earn from interest on the loans they credit. Commercial loans have always been a big part of the banking industry, and lenders are always aiming to reduce their credit risk. Nowadays in the market economy banks play a very crucial role. The profit or loss of a bank is largely influenced by loans, i.e., whether the customers repay the loans or default on them. The banks need to decide whether he/she is a good(non-defaulter) or bad(defaulter) before giving the loans to the borrowers. Among the most important problems to be addressed in commercial loan lending is the borrowers' creditworthiness. The credit risk is defined as the likelihood that borrowers will fail to meet their loan obligations.To predict whether the borrower will be good or bad is a very difficult task for any bank or organization. The banking system uses a manual process for checking whether a borrower is a defaulter or not. No doubt the manual process will be more accurate and effective, but this process cannot work when there are a large number of loan applications at the same time. If there occurs a time like this, then the decision-making process will take a very long time and also lots of manpower will be required. If we are able to do the loan prediction it will be very helpful for applicants and also for the employees of banks. So, the task is to classify the borrower as good or bad i.e., whether the borrower will be able to pay the debts back or not. This can be done with the help of machine learning algorithms.

1.2 Purpose

Loans are the core business of banks. The main profit comes directly from the loan's interest. The loan companies grant a loan after an intensive process of verification and validation. However, they still don't have assurance if the applicant is able to repay the loan with no difficulties. The Company wants to automate the loan eligibility process (real time) based on customer detail provided while filling online application form. These details are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History and others. To automate this process, they have given a problem to identify the customers segments, those are eligible for loan amount so that they can specifically target these customers.

2. LITERATURE SURVEY

2 .1 Existing Problem

Anomaly detection relies on individual's behaviour profiling and works by detecting any deviation from the norm. When it is used for online banking fraud detection, it suffers from three disadvantages. First, for an individual, the historical behaviour data are often too limited for profiling his/her behaviour pattern. Second, because of the heterogeneous nature of transaction data, there is no uniform treatment to various attribute values, which will become a potential barrier for development of the model and for further usage. Third, the transaction data are highly skewed, and it becomes a challenge for utilizing the label information effectively. Anomaly detection often suffers from poor generalization ability and a very high false alarm rate. We argue that individuals' limited historical data for behaviour profiling and fraud data's highly skewed nature could account for this defect. Since it is straightforward to use information from other similar individuals, similarity measurement itself becomes a great challenge due to heterogeneous nature of attribute values.

Disadvantages

1) They had proposed a mathematical model and machine learning algorithms were not used.
2) Class Imbalance problem was not addressed and the proper measure were not taken.

2.2 Refernces

[1] Arun Kumar, Ishan Garg, and Sanmeer Kaur, "Loan Approval Prediction Using Machine Learning Approach," 2018.

[2] K. Hanumantha Rao, G. Srinivas, A. Damodhar, and M. Vikas Krishna at International Journal of Computer Science and Telecommunications published an article titled "Implementation of Anomaly Detection Technique Using Machine Learning Algorithms" (Volume2, Issue3, June 2011).

[3] G. Arutjothi and C. Senthamarai, "Prediction of loan status in commercial banks using machine learning classifier," International Conference on Intelligent Sustainable Systems (ICISS), 2017.

[4] "AzureML based analysis and prediction of loan applicants creditworthy," by Alshouiliy K, Alghamdi A, and Agrawal D P I n 2020, Third International conference on information and computer technologies.

[5] "Developing prediction model of loan risk in banks using data mining Machine Learning and Applications," Hamid A J and Ahmed T M, 2016.

[6] M. Li, A. Mickel, and S. Taylor "Should this loan be approved or denied?" published a paper in the Journal of Statistics Education in 2018.

[7] A. Vinayagamoorthy, M. Somasundaram, and C. Sankar, "Impact of Personal Loans Offered by Banks and Non-Banking Financial Companies in Coimbatore City," 2012.

[8] M. Cary Collins, Ph.D., and Frank M. Guess, Ph.D., MIT's Information Quality Conference, 2000, "Improving information quality in loan approval processes for fair lending and fair pricing."

[9] Arun Kumar, Ishan Garg, and Sanmeet Kaur, "Loan approval prediction based on machine learning approach," National Conference on Recent Trends in Computer Science and Information Technology, 2016.

[10] Sivasree M S and Rekha Sunny T, "Loan Credibility Prediction System Using Decision Tree Algorithm," IJERT, Vol. 4 Issue 09, September-2015.

2.3 Problem Statement

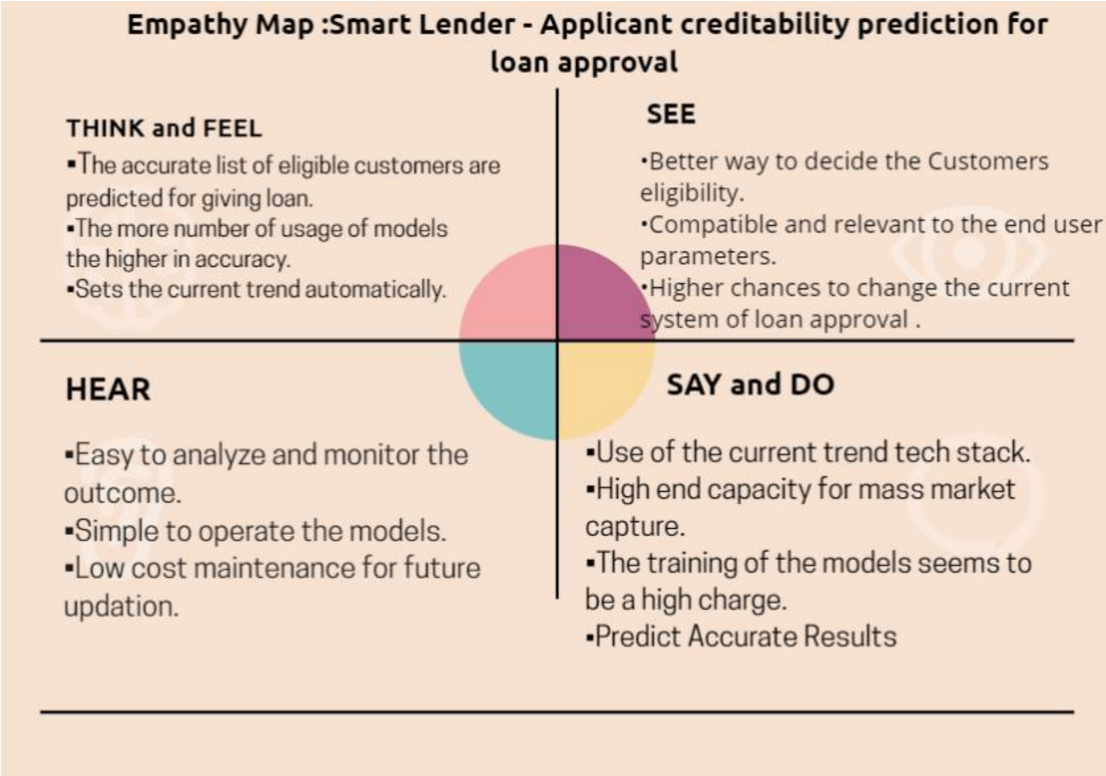
One of the most important factors which affect our country’s economy and financial condition is the credit system governed by the banks. The process of bank credit risk evaluation is recognized at banks across the globe. “As we know credit risk evaluation is very crucial, there is a variety of techniques 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.

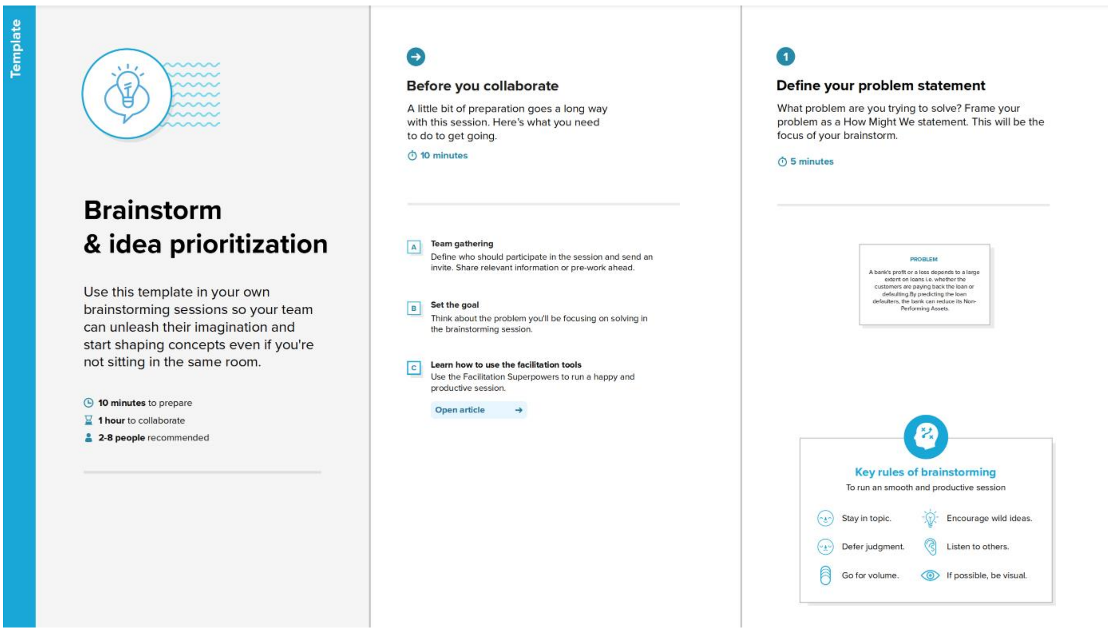
We will be using classification algorithms such as Decision tree, Random forest, KNN, and xgboost. We will train and test the data with these algorithms. From this best model is selected and saved in pkl format. We will be doing flask integration and IBM deployment.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas



3.2 Ideation & Brainstorming



2

Brainstorm

Write down any ideas that come to mind that address your problem statement.

10 minutes

TIP

You can select a sticky note and hit the pencil [switch to sketch] icon to start drawing!

HARIHARAN

- Based on dependents like friends and family
- Based on customer's age
- Customer's capacity to handle their expenses
- Based on Collateral they submit

MADHAVAN

- Based on their employment
- Purpose of loan request
- Based on borrowers age,value of property and current interest rates.
- Customer's financial stability plays a crucial role

GURUPRASAD

- Based on Previous Loan History
- Based on how much a person has income per year
- Recent credit activity of borrower
- Based on a persons criminal record verification

PRAVEEN

- CIBIL or credit score is one of the deciding factor
- Types of bank accounts they hold their money
- Based on previous Credit history
- Based on the person's repayment capacity

Based on Co-Applicant income

3

Group ideas

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

20 minutes

TIP

Add customizable tags to sticky notes to make it easier to find, browse, organize, and categorize important ideas as themes within your mural.

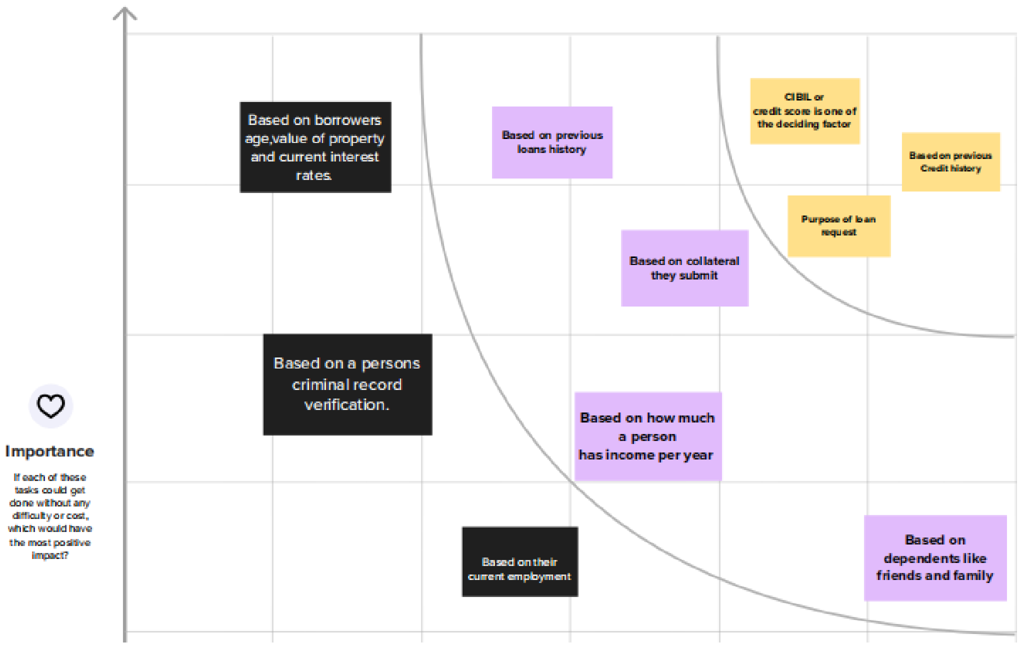
- Based on dependents like friends and family
- Based on a persons criminal record verification.
- Based on the person's repayment capacity
- Based on their employment
- Based on borrowers age,value of property and current interest rates.
- Customer's financial stability plays a crucial role
- CIBIL or credit score is one of the deciding factor
- Purpose of loan request
- Customer's capacity to handle their expenses
- Based on Collateral they submit
- Based on previous Credit history
- Based on how much a person has income per year
- Based on Previous Loan History

4

Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

20 minutes



3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	A bank is a financial institution licensed to receive deposits and make loans needs a way to verify and trust the customer details and their documents for getting loan because they need an trustable customer with proper assets ,cash flow, documents and background who can repay the loan amount and interest on time.To reduce the manual work in the banking sector a model is designed to analyse whether an individual is fit enough to avail the loan or not. The main objective is to predict whether a new applicant granted the loan or not using machine learning models trained on the historical data set. The application approved or not approved depends upon the historical data of the candidate by the system. The historical data of candidates was used to build a machine learning model using different classification algorithms.
2.	Idea / Solution description	A Machine learning model must be developed to predict the credit defaulters. This model must be trained on previous Loan approval data and their manual credibility checked data. This can be then used to predict the applicant's credibility automatically. The customer only need to enter the details,the loan approval status is then predicted automatically and quickly. The property documents of the customer need to be submitted and the customer should agree to the terms and conditions of the bank. The loan approval will also depend on the CIBIL score of the customer. Automatic calculation of interest rate and repayment date based on loan amount.

3.	Novelty / Uniqueness	In this model, the previous manually checked credibility is taken as training data. Once trained it will take Data on Loan history, Financial status and stability, Family status and Co-applicant Credibility as inputs and will provide a Boolean value output for credibility. Provide customer ratings and reviews for understanding the customer. Adding digital signature of the customer on agreement of the terms and conditions. Provides data security & the customer details will not be shared to the third party. Instant Loan approval status
4.	Social Impact / Customer Satisfaction	This model mostly predicts the credibility of a loan applicant accurately, automatically in less time compared to conventional manual checking. This socially helps banks to identify credible loan applicants thus also reduces the loss factor of the Lender (usually Bank). It also speeds up the loan sanctioning process, thus helping the applicants too.
5.	Business Model (Revenue Model)	Credit risk modelling is a method used by lenders to determine the risk involved in providing loan to a particular applicant by analyzing various attributes such as applicant income, co-applicant income, education status, credit history and employment status. Credit risk is the measure of creditworthiness of a borrower. By the help of past data trends for loans provided for the applicants, we can use machine learning algorithms to predict whether a particular applicant might be provided with a loan or not
6.	Scalability of the Solution	• It can be provided as software as aservice. (webpage) • Both borrower and Lender can use this software. • Any type of customer can predict theirloan approval without any discrimination. • Can use this software anytime and anywhere. • This system is easily scalable and efficient. • Easy and user friendly software for all.

3.4 Problem Solution Fit

Problem-Solution fit canvas 2.0

Purpose / Vision To save Bank staffs time in calculating the loan amount and helping users

Define CS, fit into CC

1. CUSTOMER SEGMENT(S) CS

Who is your customer?
i.e. working parents of 0-5 y.o. kids

- Bank staffs and loan applicants

6. CUSTOMER CONSTRAINTS CC

What constraints prevent your customers from taking action or limit their choices of solutions? i.e. spending power, budget, no cash, network connection, available devices.

- Lack of real time data
- Issues in network connectivity.
- Accuracy of the prediction

5. AVAILABLE SOLUTIONS AS

Which solutions are available to the customers when they face the problem or need to get the job done? What have they tried in the past? What pros & cons do these solutions have? i.e. pen and paper is an alternative to digital notetaking

- Calculating the amount manually.
- Contact the bank staff.
- Going with the help of the intermediates.

Explore AS, differentiate

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

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

Which jobs-to-be-done (or problems) do you address for your customers? There could be more than one; explore different sides.

- Analyze various factors that reduce the automation process.
- Accurate prediction of loan amount for customers.
- To make the loan process easier.

9. PROBLEM ROOT CAUSE RC

What is the real reason that this problem exists?
What is the back story behind the need to do this job?
i.e. customers have to do it because of the change in regulations.

- Previous loan history.
- Manually calculating mismatch amount.
- Customers have to wait for long time to get approval.

7. BEHAVIOUR BE

What does your customer do to address the problem and get the job done?
i.e. directly related: find the right solar panel installer, calculate usage and benefits; indirectly associated: customers spend free time on volunteering work (i.e. Greenpeace)

- By using quality websites for finding loan amount.
- Mobile application for predicting amount.
- Less time for loan approval.

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

Identify strong TR & EM

3. TRIGGERS TR

What triggers customers to act? i.e. seeing their neighbour installing solar panels, reading about a more efficient solution in the news.

Takes more time in calculating the loan amount

4. EMOTIONS: BEFORE / AFTER EM

How do customers feel when they face a problem or a job and afterwards?
i.e. lost, insecure > confident, in control - use it in your communication strategy & design.

Before: Frustrated

After: Satisfied

10. YOUR SOLUTION SL

If you are working on an existing business, write down your current solution first, fill in the canvas, and check how much it fits reality.
If you are working on a new business proposition, then keep it blank until you fill in the canvas and come up with a solution that fits within customer limitations, solves a problem and matches customer behaviour.

Accurately predict loan amount in the bank sectors using ML algorithms like SVM Classifier, Random Forest, Logistic Regression based on the input given by the user

8. CHANNELS of BEHAVIOUR CH

8.1 ONLINE

What kind of actions do customers take online? Extract online channels from #7

- To check the loan amount applicable to the customer in websites.

8.2 OFFLINE

What kind of actions do customers take offline? Extract offline channels from #7 and use them for customer development.

- Calculate correct amount by the bank staffs.

Extract online & offline CH of BE

4. REQUIREMENT ANALYSIS

4.1 Functional Requirements

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through entering details such as name,email ,password,phone number,etc.
FR-2	User Login	Login using the registered email id and password.
FR-3	Model Building	Build various machine learning model to predict Applicant Credibility and compare them.
FR-4	Check Details	Get the user details and display if the user has credibility for loan approval or not.
FR-5	Integration	Integrate the front end and the developed ML model using Flask.
FR-6	Alert Message	Notify the user through email or phone regarding the loan approval.

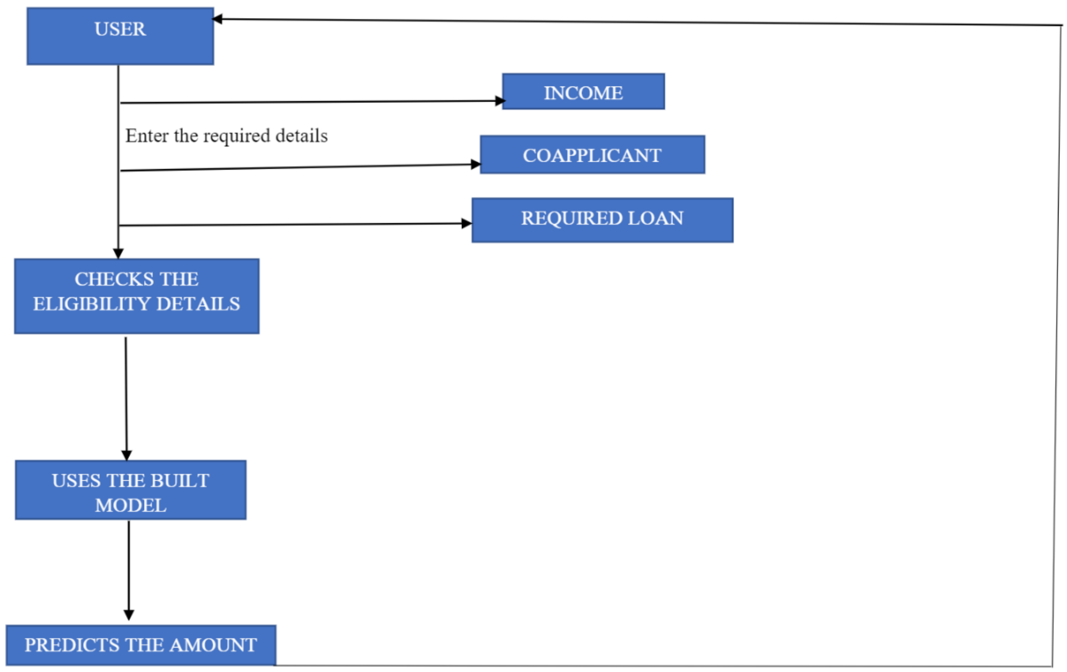
4.2 Non-Functional Requirements

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

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The application will be easy to use with a neat andclean interface. Any valid user data must be accepted for prediction.
NFR-2	**Security**	Safe encryption of data is done to ensure customer data safety. Alert message must be sent to the users.
NFR-3	**Reliability**	Loan approval of applicant must be predicted accurately and the result must be reliable.
NFR-4	**Performance**	The performance and interface must be user friendly.
NFR-5	**Availability**	All banks, financial institutions and customers will beable to use the application. Anyone who has the valid bank account.
NFR-6	**Scalability**	The application is very scalable and runs across operating systems & It must be able to handle increase in the number of users.

5.PROJECT DESIGN

5.1 Data Flow Diagrams



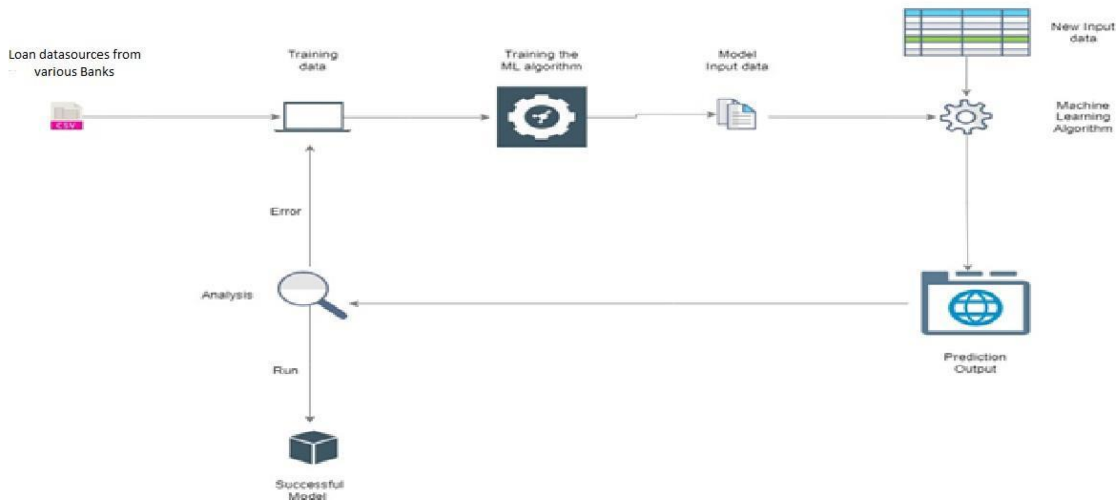
5.2 Solution & Technical Architecture

Solution Architecture:

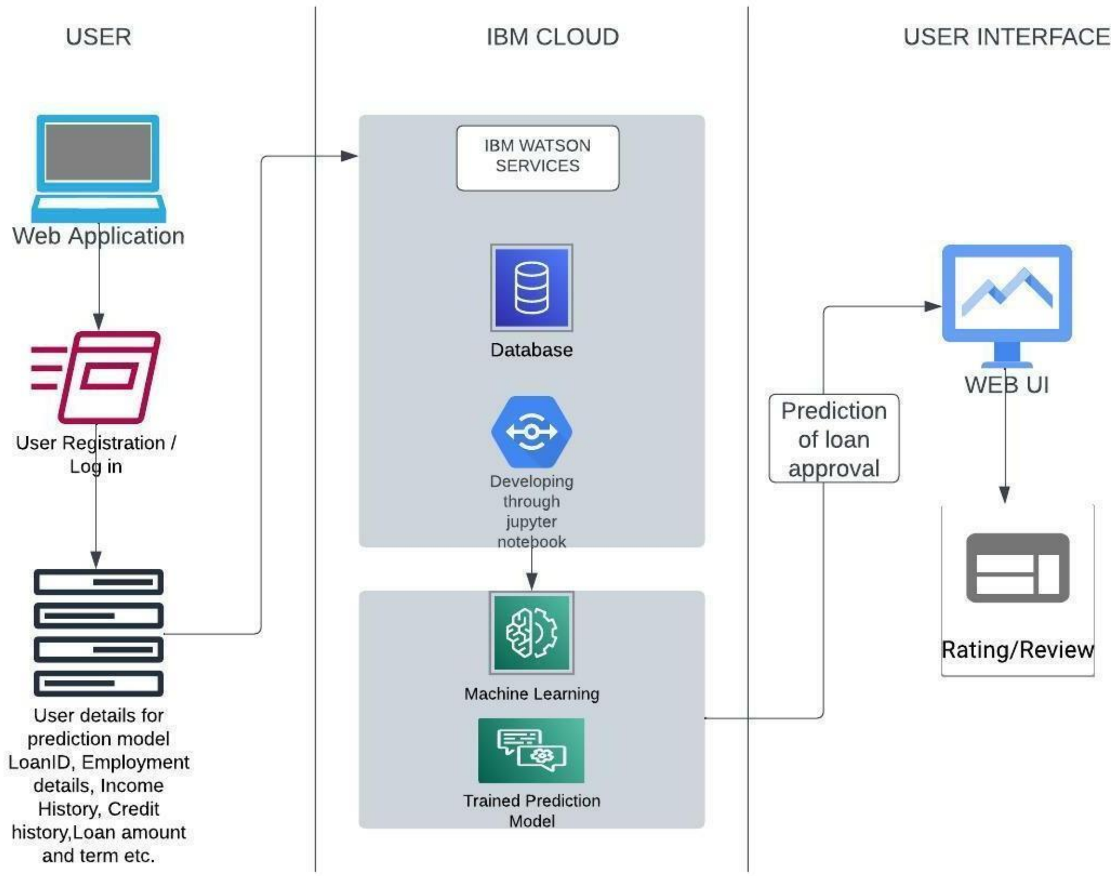
Solution architecture for the university admit eligibility predictor description as follows as:

- First the data sources of various Banks are collected with the opensource platform integration.
- The data's are pre processed and cleaned for the training.
- The machine learning algorithm is decided to train the datasets.
- Model input data is generated by the continuous training.
- For the upgradation of the prediction model the new dataset's are integrated at the regular intervals for trending Banks lists.
- The models are trained and tested with the higher accuracy and precision until the final successful model is build.
- As the final output prediction is done by the desired output

TITLE : SMART LENDER - APPLICANT CREDABILITY PREDICTION FOR LOAN APPROVAL



Technical Architecture



5.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through Gmail		Medium	Sprint-1
	Login	USN-4	As a user, I can log into the application by entering email & password		High	Sprint-1
	Dashboard	USN-5	As a user, I can see my previous loan status		High	Sprint-1
Customer (Web user)	Form Details	USN-6	I need to enter the details in a good looking UI form.		low	Sprint-2
	Displaying amount	USN-7	The predicted loan amount should be displayed as easily readable one.	Easy to notice the loan amount.	High	Sprint-1
		USN-8	Apply suitable Font & styles to the amount		Medium	Sprint-2
	Status	USN-9	As a user, I have to see the status of my loan application.			
Customer Care Executive	Unstable condition	USN-10	I need to contact the executive in case of some unusual situations.		Medium	Sprint-1
Administrator	Updation & change of policies	USN-11	Admin have to update the application regularly as user friendly and need to change the policies if needed.		High	Sprint-1

6 PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

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 byentering my email, password, and confirming my password.	2	High	Madhavan V P & Praveen C P
Sprint-1		USN-2	As a user, I will receive confirmation email onceI have registered for the application.	1	High	Guruprasad M & Hariharan GV
Sprint-1	Login	USN-3	As a user, I can log into the application byentering email & password.	3	Medium	Guruprasad M & Hariharan G V
Sprint-2	Update Profile	USN-4	As a user, after logging in i will have to update my profile by providing all the required details.	2	Medium	Madhavan V P & Praveen CP
Sprint-3	history	USN-5	As user I can see my previous loan details	3	High	Madhavan V P & Praveen C P
Sprint-1	Collecting details	USN-6	I need to enter the details in a good looking UI form	3	medium	Guruprasad M & Hariharan G V
Sprint-3	prediction	USN-7	The predicted Loan amount should be displayed as a easily readable one.	5	High	Guruprasad M & Hariharan G V
Sprint-1	Output	USN-8	Apply suitable styles to the required fonts	5	High	Madhavan V P & Praveen C P
Sprint-4	Logout	USN-9	As a user after the all the process over then I can log out the application	2	Low	Guruprasad M & Hariharan G V
Sprint-4	Status	USN-10	As a user, I have to see the status of my loan application.	2	Medium	Madhavan V P & Praveen C P

6.2 Sprint Delivery Schedule

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	11	8 Days	24 Oct 2022	31 Oct 2022		31 Oct 2022
Sprint-2	5	5 Days	02 Nov 2022	06 Nov 2022		06 Nov 2022
Sprint-3	8	6 Days	07 Nov 2022	12 Nov 2022		12 Nov 2022
Sprint-4	6	6 Days	14 Nov 2022	19 Nov 2022		19 Nov 2022

6.3 Reports from JIRA

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

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

Our Project velocity

Sprint-1 = 11/6 = 1.833

Sprint-2 = 18/6 = 3

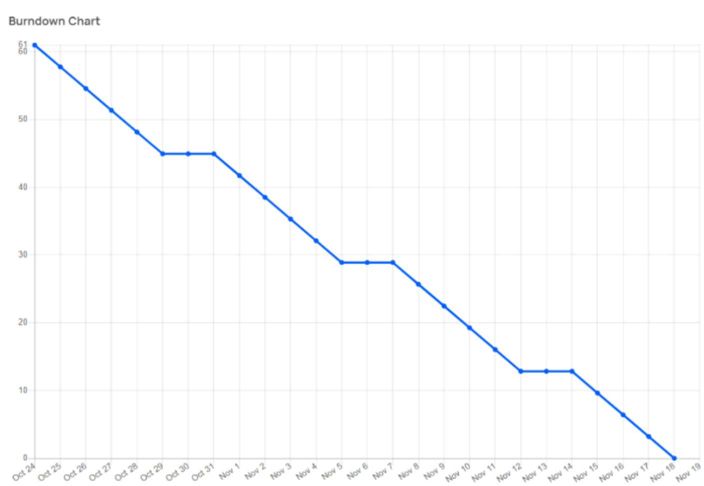
Sprint-3 = 16/6 = 2.67

Sprint-4 = 16/6 = 2.67

Total Velocity = 61/24 = 2.54

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 and Solutioning

Dataset used

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	Coapplicant_Income	LoanAmount	Loan	Loan_Amount_Term	Cred
0	LP001002	Male	No	0.0	Graduate	No	52690	0.0	100000	100000.0	12	
1	LP001003	Male	Yes	1.0	Graduate	No	45830	15080.0	18000	18000.0	18	
2	LP001005	Male	Yes	0.0	Graduate	Yes	3000	0.0	85000	85000.0	18	
3	LP001006	Male	Yes	0.0	Not Graduate	No	25830	23580.0	100000	100000.0	18	
4	LP001008	Male	No	0.0	Graduate	No	60000	0.0	150000	150000.0	18	

Importance of independent variables:

```
Variable: ApplicantIncome      Importance: 0.72561
Variable: Coapplicant_Income   Importance: 0.25648
Variable: LoanAmount           Importance: 0.00396
Variable: Emi                  Importance: 0.00352
Variable: Loan_Amount_Term     Importance: 0.00251
Variable: Dependents           Importance: 0.00246
Variable: Yes                  Importance: 0.00116
Variable: Female               Importance: 0.00107
Variable: Male                 Importance: 0.00088
Variable: No                   Importance: 0.00077
Variable: Yes                  Importance: 0.0007
Variable: No                   Importance: 0.00065
Variable: Credit_History       Importance: 0.00017
Variable: Graduate             Importance: 3e-05
Variable: Not Graduate         Importance: 3e-05
```

Random Forest code:

```
from sklearn.ensemble import RandomForestRegressor
# Instantiate model with 1000 decision trees
rf = RandomForestRegressor(n_estimators = 1000, random_state=50)
# Train the model on training data
rf.fit(x_train,y_train);

predictions=rf.predict(x_test)
# Calculate the absolute errors
errors = abs(predictions -y_test)
# Print out the mean absolute error (mae)
print('Mean Absolute Error:', round(np.mean(errors), 2), 'degrees.')

# Calculate mean absolute percentage error (MAPE)
mape = 100 * (errors / y_test)
```

# Calculate and display accuracy
accuracy = 100 - np.mean(mape)
print('Accuracy:', round(accuracy, 2), '%.')

Decision Tree:

```
In [60]: from sklearn.tree import DecisionTreeRegressor

         regressor = DecisionTreeRegressor()

         regressor.fit(x_train,y_train)

         y_pred = regressor.predict([[0,50000,42000,200000,60,0,3000,0,1,0,1,0,1,1,0]])

         print("Predicted Amount:",y_pred)

         Predicted Amount: [147930.]
```

KNN:

```
In [59]: from sklearn.neighbors import KNeighborsRegressor

         knn = KNeighborsRegressor(n_neighbors=5)

         knn.fit(x_train, y_train)

         print("Predicted Amount:",knn.predict([[0,50000,42000,200000,60,0,3000,0,1,0,1,0,1,1,0]]))

         Predicted Amount: [140325.]
```

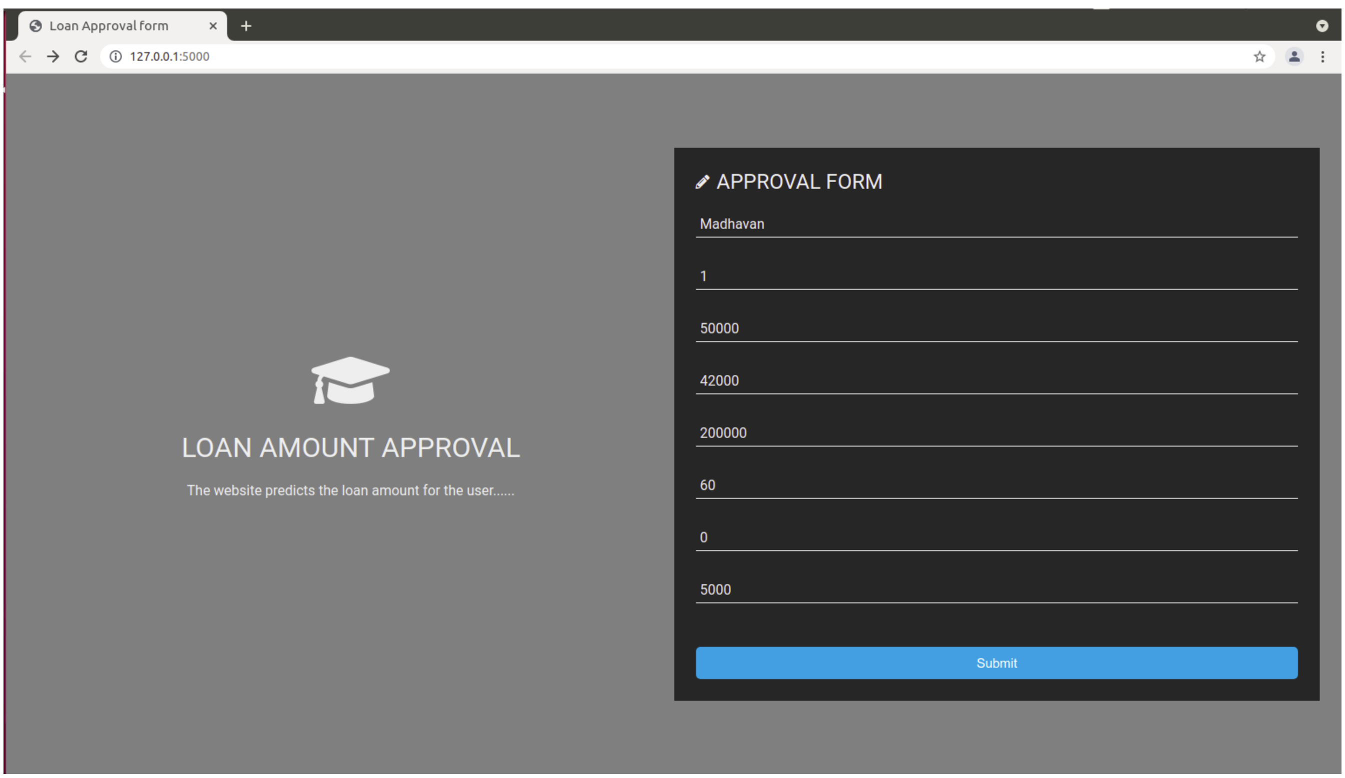
XGBoost:

```
In [56]: from sklearn.ensemble import GradientBoostingRegressor

         xg = GradientBoostingRegressor()

         xg.fit(x_train, y_train)

         print("Predicted Amount:",xg.predict([[0,50000,42000,200000,60,0,3000,0,1,0,1,0,1,1,0]]))

         Predicted Amount: [133185.90245476]
```

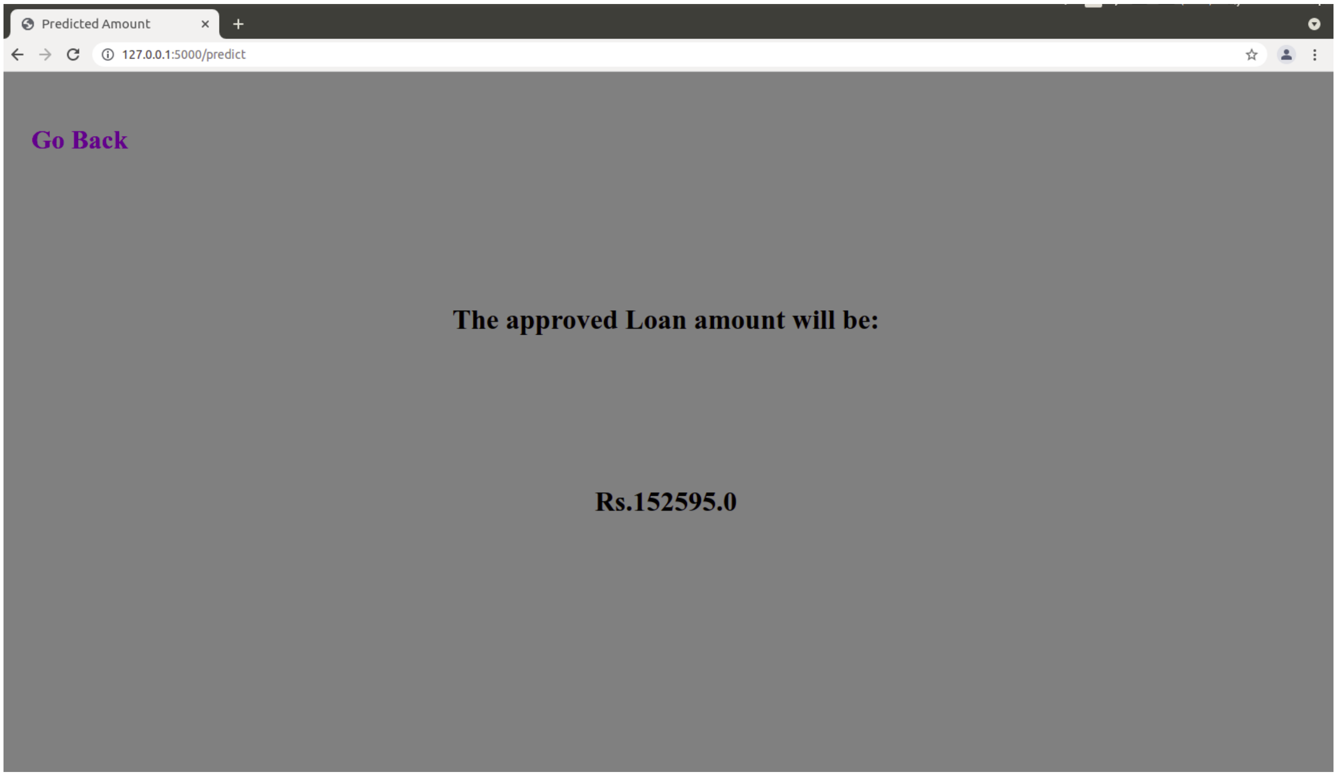
Model Comparison:

```
In [57]: print("Random forest Accuracy:",r2_score(y_test,rf.predict(x_test)))

         print("Decision Tree Accuracy",r2_score(y_test,regressor.predict(x_test)))

         print("KNN Accuracy:",r2_score(y_test,knn.predict(x_test)))

         print("XGboost Accuracy:",r2_score(y_test,xg.predict(x_test)))

Random forest Accuracy: 0.8201031194388251
Decision Tree Accuracy 0.8784230271667762
KNN Accuracy: 0.678718208574209
XGboost Accuracy: 0.8565868979070638
```

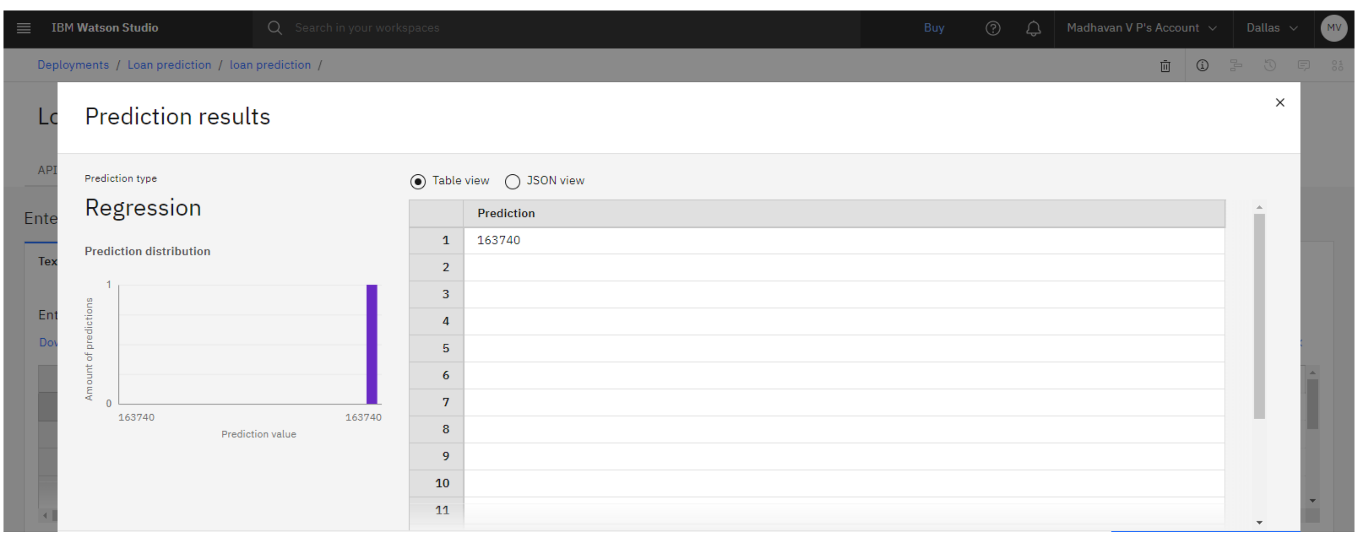
Loan approval form:



Predicted Value:



ML Model on IBM Cloud:



8. TESTING

8.1 Test Cases

Loan _ID	Ge n der	Mar ried	Depen d ents	Educa ti on	Self-employ ed	Applican t Income	Co-applicant Income	Loan Amou nt	Loan Amount Term	Credit History	Propert y Area	Loan Status
LP0 0 1002	Ma l e	No	0	Gradu ate	No	5849	0		360	1	Urban	Y
LP0 0 1003	Ma l e	Yes	1	Gradu ate	No	4583	1508	128	360	1	Rural	N
LP0 0 1005	Ma l e	Yes	0	Gradu ate	Yes	3000	0	66	360	1	Urban	Y
LP0 0 1006	Ma l e	Yes	0	Not Gradu ate	No	2583	2358	120	360	1	Urban	Y
LP0 0 1008	Ma l e	No	0	Gradu ate	No	6000	0	141	360	1	Urban	Y
LP0 0 1011	Ma l e	Yes	2	Gradu ate	Yes	5417	4196	267	360	1	Urban	Y
LP0 0 1013	Ma l e	Yes	0	Not Gradu ate	No	2333	1516	95	360	1	Urban	Y
LP0 0 1014	Ma l e	Yes	3	Gradu ate	No	3036	2504	158	360	0	Semiur ba n	N
LP0 0 1018	Ma l e	Yes	2	Gradu ate	No	4006	1526	168	360	1	Urban	Y
LP0 0 1020	Ma l e	Yes	1	Gradu ate	No	12841	10968	349	360	1	Semiur ba n	N

8.2 User Acceptance Testing

1. Purpose of Document

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

2. Defect Analysis

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

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	10	4	2	3	20
Duplicate	1	0	3	0	4
External	2	3	0	1	6
Fixed	11	2	4	20	37
Not Reproduced	0	0	1	0	1
Skipped	0	0	1	1	2
Won't Fix	0	5	2	1	8
Totals	24	14	13	26	77

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	7	0	0	7
Client Application	51	0	0	51
Security	2	0	0	2
Outsource Shipping	3	0	0	3
Exception Reporting	9	0	0	9
Final Report Output	4	0	0	4
Version Control	2	0	0	2

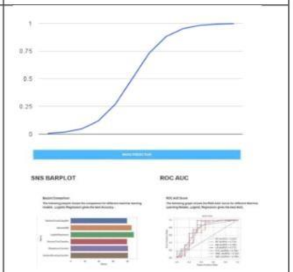
Page 2 of 2

9. RESULTS

9.1 Performance Metrics

Model Performance Testing:

Project team shall fill the following information in model performance testing template.

S.No	Parameter	Values	Screenshot
1.	Metrics	**Regression Model:** MAE - , MSE - , RMSE - , R2 score -**Classification Model:** Confusion Matrix - , Accuray Score- & Classification Report -	
2.	Tune the Model	Hyperparameter Tuning - Validation Method -	

For the purpose of predicting the loan approval status of the applied customer, we have chosen the machine learning approach to study the bank dataset. We have applied various machine learning algorithms to decide which one will be the best for applying on the dataset to get the result with the highest accuracy. 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.

We also 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. This 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.

10. ADVANTAGES AND DISADVANTAGES

Advantages:

The benefits of loan tech servicing software for lenders include:

Eliminating human error

It's no secret, that calculations are something that algorithms handle better than we, humans. In a lending system, there are just too many variables, which is why it is error-prone. The best loan servicing software, however, is created to completely rule out any errors, which is, undoubtedly, beneficial from every standpoint.

Preventing delays in payment

Not being able to collect a debt is something that most lenders are especially wary of. However, if they leverage a traditional loan management approach, they may not see it coming. Loan servicing systems, on the other hand, integrate analytic modules capable of detecting even the most subtle fluctuations in clients' credibility and preventing payment delays in a timely manner.
Saving time

Loan management requires a great level of meticulousness and attention to detail. As a rule, a full-fledged team is required to deal with every aspect of a loan process. Needless to say, loan management carried out manually and based on paperwork takes up a lot of time. A digital lending system, on the other hand, automates the routines and enables your team to dedicate time to other important tasks.

Automated reporting

Automated report generation is another invaluable feature offered by a digital loan servicing platform. Accounting, tax reports, and invoices are often requested by regulatory bodies, borrowers and investors. These high urgency reports should be provided on demand, and contain information, which is 100% accurate. Loan tracking software enables lenders to quickly generate reports of different types and submit them urgently, in the required formats.

Increased revenue

This stems from all of the above: an automated loan processing system enables lenders to process more applications, assign and manage more loans, and see them all the way through closing all while detecting scams and preventing delays. The staff is free to oversee the process and focus on client relationships and look for new business opportunities. This enables financial companies to gain a distinct competitive edge and increase revenue.

Disadvantages

Accessibility

An organization looking to build loan software may not have enough on-premise infrastructure capacities to ensure its non-disruptive operation, updates, and support. Scaling during peak workloads and handling an increase in the number of users and subscriptions may also be quite challenging. Using cloud infrastructure is best to ensure optimal scalability and availability.

Servicing different loan types

The more types of loans your money lending software is capable of servicing, the better. Lending applications that have a wide range of use cases, will surely attract more users than apps targeting only one specific loan type. A loan Tech software to create loan app estimation, for example, may have a broad range of applications from student loan tech calculations to estimating business loans and mortgages.
Centralized data storage

Every stage of the lending process involves working with customer data. The best loan servicing software stores this data in centralized storage accessible during every loan processing stage. A legacy loan management system, on the other hand, uses a siloed approach to data storage, which makes loan processing more laborious and lengthier.

Integrated credit assessment capabilities

Modern loan servicing software for private lenders should be able to instantly connect with credit bureaus and any other bodies responsible for credibility assessment. Such platforms should receive regular credit data updates and leverage big data analytics to assess the trustworthiness of applicants. The client's social media activity, for example, can be a valid source of alternative assessment of credibility.
Automation of routine processes

Using robotic process automation to streamline simple rule-based processes is another must-have feature of a loan management platform. Automation accelerates loan origination and processing and accounts for increasing client satisfaction. On top of that, it helps to avoid human error.
In-built analytic modules

Leveraging artificial intelligence (AI) and big data is another hallmark of excellent loan servicing software for lenders. Not only does it help to generate reports but also enables companies to evaluate market trends, detect patterns in customer behavior and come up with new products and offerings

.

Third-party integration

Another feature that most organizations find especially attractive in a loan processing system, is its capability to integrate with other enterprise software. ERP and CRM solutions are capable of enriching the lending system with data and insights. Systems integrating lending modules with software for remote sales personnel are also enjoying high popularity among lenders.
Security

Finance company software works with classified and highly sensitive data, and for both lenders and customers, security is a matter of paramount importance. An excellent lending system should possess advanced security capabilities, and ensure the highest level of customer, data, and network protection.

11. CONCLUSION

In the debate over which supervised learning model to use for credit risk assessment, we have come to the conclusion that support vector machines can outperform other tree- based models or regression models if the setup of the experiment is similar to that of ours. Furthermore, in the debate over which dimensionality reduction technique to use, our model has shown us that recursive feature elimination with cross-validation can outperform models based on principal component analysis. For future improvements we would like to use more current data and from different sources for illustrating a better understanding of the trends present in this field. Datasets similar to the above-mentioned experiments from previous works will be used to test this model for better comparison. It has been mentioned that in order to reduce computational cost and complexity we have omitted the idea of using neural networks. But as we are looking forward to work with even larger amount of data, we would like to make a comparative analysis using neural networks as well. It is a known fact that neural networks tend to perform better with large datasets and we would like to test this hypothesis in our future works. Again, as we are also discussing the contributions of feature selection/extraction techniques, we would like to implement other dimensionality reduction techniques such as genetic algorithm, univariate feature selection methods, tree-based feature selections etc. to gauge their performances and further improve the efficiency of the credit lending sector. Therefore, this paper can be concluded with the statement that this model illustrates an interesting approach in identifying loan defaulters in this ever-changing economy. Using the dataset from Lending Club our model has brought about remarkable results which in turn can play a major role in assessing the credit risk of borrowers, assist credit lending institutions and enable financial institutions to keep operating in a transparent and profitable wa

12. FUTURE SCOPE

In this section, based on various performance metrics, a comparative analysis will be made of all the generated models. A precise classifier is the backbone of any machine learning model. Four supervised algorithms: Support vector machine (SVM), Logistic Regression (LR), Extreme Gradient Boosting (XGB) and Random Forest (RF) have been selected for the analysis. The hyperparameters of these algorithms will be tuned using GridSearchCV to select the best set of values for each model. The results will be discussed in two categories and will be illustrated in both a graphical and tabular manner. Firstly the models will be evaluated on a holdout test set using a train test split. Then another comparative analysis will be made of the same models but using a 5 fold cross- validation and GridSearchCV.

Z-score has been chosen over normalization (min-max scaling) for scaling the features. Classifiers such as support vector machine, logistic regression or neural network prefer standardization over normalization. Additionally, this paper proposes to use such feature extraction methodologies where maximizing the variance is highly preferred. This can be achieved using standardization. Furthermore GridSearchCV has been used to optimize the hyperparameters of each classifier. Studies done in perfectly show the effectiveness of GridSearchCV in maximizing the performance of classifiers.

13. APPENDIX

Source Code

loan.py

```
from flask import Flask,render_template, request, redirect, url_for, session
import joblib
app = Flask(__name__)

@app.route('/')
def index():
	return render_template('loan.html')

@app.route('/predict',methods =['GET', 'POST'])
def predict():
	name=request.form['name']
	dep=int(request.form['dep'])
	apin=int(request.form['appin'])
	coap=int(request.form['coappin'])
	la=int(request.form['lamt'])
	lat=int(request.form['lamttrm'])
	ch=int(request.form['ch'])
	emi=int(request.form['emi'])
	x=[[dep,apin,coap,la,lat,ch,emi,0,1,0,1,0,1,1,0]]
	model=joblib.load('dt.pkl')
	#amt=model.predict(x)[0]
	return render_template('predict.html',predict=1000000)


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

loan.html

```
<!DOCTYPE html>
<html>
  <head>
    <title>Educational registration form</title>
    <link rel="stylesheet" href="https://use.fontawesome.com/releases/v5.4.1/css/all.css" integrity="sha384-
5sAR7xN1Nv6T6+dT2mhtzEpVJvfS3NScPQTrOxhwjIuvcA67KV2R5Jz6kr4abQsz"
crossorigin="anonymous">
    <link href="https://fonts.googleapis.com/css?family=Roboto:300,400,500,700" rel="stylesheet">
    <style>
      html, body {
      min-height: 100%;
      }
      body, div, form, input, select, p {
      padding: 0;
      margin: 0;
      outline: none;
      font-family: Roboto, Arial, sans-serif;
      font-size: 16px;
      color: #eee;
      }
      body {
      background: url("/uploads/media/default/0001/01/b5edc1bad4dc8c20291c8394527cb2c5b43ee13c.jpeg")
no-repeat center;
      background-size: cover;
      }
      h1, h2 {
      text-transform: uppercase;
      font-weight: 400;
      }
      h2 {
      margin: 0 0 0 8px;
      }
      .main-block {
      display: flex;
      flex-direction: column;
      justify-content: center;
      align-items: center;
      height: 100%;
      padding: 25px;
      background: rgba(0, 0, 0, 0.5);
      }
      .left-part, form {
      padding: 25px;
      }
      .left-part {
      text-align: center;
      }
      .fa-graduation-cap {
```

```
font-size: 72px;
}
form {
background: rgba(0, 0, 0, 0.7);
}
.title {
display: flex;
align-items: center;
margin-bottom: 20px;
}
.info {
display: flex;
flex-direction: column;
}
input, select {
padding: 5px;
margin-bottom: 30px;
background: transparent;
border: none;
border-bottom: 1px solid #eee;
}
input::placeholder {
color: #eee;
}
option:focus {
border: none;
}
option {
background: black;
border: none;
}
.checkbox input {
margin: 0 10px 0 0;
vertical-align: middle;
}
.checkbox a {
color: #26a9e0;
}
.checkbox a:hover {
color: #85d6de;
}
.btn-item, button {
padding: 10px 5px;
margin-top: 20px;
border-radius: 5px;
border: none;
background: #26a9e0;
text-decoration: none;
font-size: 15px;
font-weight: 400;
color: #fff;
```

```
      }
      .btn-item {
      display: inline-block;
      margin: 20px 5px 0;
      }
      button {
      width: 100%;
      }
      button:hover, .btn-item:hover {
      background: #85d6de;
      }
      @media (min-width: 568px) {
      html, body {
      height: 100%;
      }
      .main-block {
      flex-direction: row;
      height: calc(100% - 50px);
      }
      .left-part, form {
      flex: 1;
      height: auto;
      }
      }
    </style>
  </head>
  <body>
    <div class="main-block">
      <div class="left-part">
        <i class="fas fa-graduation-cap"></i>
        <h1>Loan amount Approval</h1>
        <p>The website predicts the loan amount for the user......</p>
      </div>
      <form action="/predict">
        <div class="title">
          <i class="fas fa-pencil-alt"></i>
          <h2>Approval Form</h2>
        </div>
        <div class="info">
          <input class="fname" type="text" name="name"id="name"  placeholder="Full name">
          <input type="text" name="dep" id="dep" placeholder="Dependents(in numbers)">
          <input type="text" name="appin" id="appin" placeholder="Applicant Income">
              <input type="text" name="coappin" id="coappin" placeholder="Co-Applicant Income">
              <input type="text" name="lamt" id="lamt" placeholder="Loan Amount">
              <input type="text" name="lamttrm" id="lamttrm" placeholder="Loan Amount term">
              <input type="text" name="ch" id="ch" placeholder="Credit History">
              <input type="text" name="emi" id="emi" placeholder="EMI">
        </div>
        <button type="submit" href="#">Submit</button>
      </form>
    </div>
```

```
  </body>
</html>
```

predict.html

```
<!DOCTYPE html>
<html>
  <head>
    <title>Predicted Amount</title>
    <link rel="stylesheet" href="https://use.fontawesome.com/releases/v5.4.1/css/all.css" integrity="sha384-5sAR7xN1Nv6T6+dT2mhtzEpVJvfS3NScPQTrOxhwjIuvcA67KV2R5Jz6kr4abQsz" crossorigin="anonymous">
    <link href="https://fonts.googleapis.com/css?family=Roboto:300,400,500,700" rel="stylesheet">
    <style>

   body {
      background-color: grey;
      background-size: cover;
      }
h1
{
padding-top:150px;
}

a{
text-decoration:none;
font-color:red;
font-size:30px;
padding-top:30px;
padding-left:30px;
}

#gb
{
font-color:green;
}

  </style>
  </head>
  <body>
<a href="#" id="gb"><h3><b>Go Back</a></b></h3>
<center>
            <h1>The approved Loan amount will be:</h1>
            <h1>Rs.{{predict}}</h1>
</center>
  </body>
</html>
```

Flask Code with Ml model on IBM Cloud:

```
from flask import Flask,render_template, request, redirect, url_for, session

import requests
app = Flask(__name__)


API_KEY = "UFiKYlV2AoWKo8kRnD98FYnw1kbFf1pjM9tjqIRhQplY"
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.route('/')
def index():
        return render_template('loan.html')

@app.route('/predict',methods =['GET', 'POST'])
def predict():
        name=request.form['name']
        dep=int(request.form['dep'])
        apin=int(request.form['appin'])
        coap=int(request.form['coappin'])
        la=int(request.form['lamt'])
        lat=int(request.form['lamttrm'])
        ch=int(request.form['ch'])
        emi=int(request.form['emi'])
        x=[[dep,apin,coap,la,lat,ch,emi,0,1,0,1,0,1,1,0]]
        #Smodel=joblib.load('dt.pkl')
        #amt=model.predict(x)[0]

        payload_scoring = {"input_data": [{"field":
[['dep','apin','coap','la','lat','ch','emi','0','1','0','1','0','1','1','0']], "values":x}]}

        response_scoring = requests.post('https://us-
south.ml.cloud.ibm.com/ml/v4/deployments/53c0255e-5264-4c30-8c99-
7d4f79629557/predictions?version=2022-11-16', json=payload_scoring,headers={'Authorization': 'Bearer '
+ mltoken})
        print(response_scoring)
        predictions = response_scoring.json()
        predict = predictions['predictions'][0]['values'][0][0]
        print("Final prediction :",predict)
        return render_template('predict.html',predict=predict)


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

GITHUB LINK : https://github.com/IBM-EPBL/IBM-Project-24148-1659938655