SMART LENDER - APPLICANT CREDIBILITY PREDICTION FOR LOAN APPROVAL

TEAM ID:

PNT2022TMID15942

TEAM MEMBERS:

1. RAJESH KUMAR A (927619BEC4161)
2. NAVEEN B (927619BEC4130)
3. NAVEEN K (927619BEC4131)
4. SANJAY G (927619BEC4171)

INDUSTRY MENTORS NAME:

1. NIDHI

FACULTY MENTOR NAME:

1. Dr. SIVANANDAM K

1.INTRODUCTION

A loan is the core business part of banks. The main portion the bank's profit is directly come from the profit earned from the loans. Though bank approves loan after a regress process of verification and testimonial but still there is no surety whether the chosen hopeful is the right hopeful or not. This process takes fresh time while doing it manually. We can prophesy whether that hopeful is safe or not and the whole process of testimonial is automated by machine literacy style. Loan Prognostic is helpful for retainer of banks as well as for the hopeful also.

1.1 PROJECT REVIEW

As the needs of people are increasing, the demand for loans in banks is also frequently getting higher every day. Banks typically process an applicant's loan after screening and verifying the applicant's eligibility, which is a difficult and time-consuming process. In some cases, some applicants default and banks lose capital. The machine learning approach is ideal for reducing human effort and effective decision making in the loan approval process by implementing machine learning tools that use classification algorithms to predict eligible loan applicants. 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.

1.2 PURPOSE

The main purpose of this Project is to compare the Loan prediction Models madeusing various algorithms and choose the best one out of them that can shorten theloan approval time and decrease the risk associated with it. It is done by predicting if the loan can be given to that person based on various parameters like credit score, income, age, marital status, gender, etc. The Prediction model not only helps the applicant but also helps the bank by minimizing the risk and reducing the number of defaulters.

2. LITERATURE SURVEY

Loan Prediction using machine learning model whether it will be safe to allocate the loan to a particular person. This paper has the following sections (i) Collection of Data, (ii) Data Cleaning and (iii) Performance Evaluation. Experimental tests found that the Naïve Bayes model has better performance than other models in terms of loan forecasting. With the enhancement in the banking sector lots of people are applying for bank loans but the bank has its limited assets which it must grant to limited people only, so finding out to whom the loan can be granted which will be a safer option for the bank is a typical process. So, in this

project we try to reduce this risk factor behind selecting the safe person so as to save lots of bank efforts and assets. This is done by mining the Big Data of the previous records of the people to whom the loan was granted before and based on these records/experiences the machine was trained using the machine learning model which give the most accurate result.

2.1 EXISTING PROBLEM

Bank employees check the details of applicant manually and give the loan to eligible applicant. Checking the details of all applicants takes lot of time. The artificial neural network model for predict the credit risk of a bank. The Feed-forward back propagation neural network is used to forecast the credit default. The method in which two or more classifiers are combined together to produce a ensemble model for the better prediction. They used the bagging and boosting techniques and then used random forest technique. The process of classifiers is to improve the performance of the data and it gives better efficiency. In this work, the authors describe various ensemble techniques for binary classification and also for multi class classification.

AUTHOR: Supriya, Pavani, Saisushma

DESCRIPTION: They presented a ML based loan prediction model. The modules in the present approach were data collection and pre-processing, applying the ML models, training followed by testing the data. During the preprocessing stage, the

detection and removal of outliers and imputation removal processing were carried out. In the present method, SVM, DT, KNN and gradient boosting models were employed to predict the possibilities of current status regarding the loan approval process. The conventional 80:20 rule was adopted to split the dataset into training and testing processes. Experimentation concluded that, DT has significantly higher loan prediction accuracy than the other models.

AUTHOR: Vimala Kumari and Vikas

DESCRIPTION: presented a loan prediction model using several Machine Learning (ML) algorithms. The dataset with features, namely, gender, marital status, education, number of dependents, employment status, income, co applicant ‘s income, loan amount, loan tenure, credit history, existing loan status, and property area, are used for determining the loan eligibility regarding the loan sanctioning process. Various ML models adopted in the present method includes, Linear model, Decision Tree (DT), Neural Network and, Random Forest (RF).

AUTHOR: Aboobyda Jafar Hamid and Tarig Mohammed Ahmed

DESCRIPTION: presented a loan risk prediction model based on the data mining techniques, such as Decision Tree (J48), Naïve Bayes (NB) and BayseNet approaches. The procedure followed was training set preparation, building the

model, Applying the model and finally. Evaluating the accuracy. This approach was implemented using Weka Tool and considered a dataset with eight attributes, namely, gender, job, age, credit amount, credit history, purpose, housing, and class. Evaluating these models on the dataset, experimental results concluded that, J48 based loan prediction approach resulted in better accuracy than the other methods.

AUTHOR: Goyal and Kaur

DESCRIPTION: An ensemble technique-based loan prediction procedure for the customers. The sub processes in the present method includes, data collection, filtering the data, feature extraction, applying the model, and finally analysis the results. The various loan prediction procedures implemented in the presentmethod were Random Forest (RF), SVM and Tree model with Genetic Algorithm(TGA).

AUTHOR: Sudhamathy

DESCRIPTION: A risk analysis method in sanctioning a loan for the customers using R package. The various modules include data selection, pre-processing, feature extraction and selection, building the model, prediction followed by the evaluation. The dataset used for evaluation in this method was adopted from UCI repository. To fine tune the prediction accuracy, the pre-processing operation includes the following sub-processes: detection, ranking and removal of outliers, removal of imputation, and balancing of dataset by proportional bifurcation

regarding testing and training process. Further, feature selection process improvesthe prediction accuracy.

2.2 REFERENCES

[1] Kumar Arun, Garg Ishan, Kaur Sanmeet, May-Jun. 2016. Loan Approval Prediction based on Machine Learning Approach, IOSR Journal of Computer Engineering (IOSR-JCE)

[2] Wei Li, Shuai Ding, Yi Chen, and Shanlin Yang, Heterogeneous Ensemble for Default Prediction of Peer-to-Peer Lending in China, Key Laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education, Hefei University of Technology, Hefei 2009, China

[3] Short-term prediction of Mortgage default using ensembled machine learning models, Jesse C.Sealand on july 20, 2018. [4] Clustering Loan Applicants based on Risk Percentage using K-Means Clustering Techniques, Dr. K. Kavitha, International Journal of Advanced Research in Computer Science and Software Engineering.

2.3 PROBLEM STATEMENT DEFINITION

An Exploratory Data Analysis (EDA) regarding the loan prediction procedure based on the client ‘s nature and their requirements. The major factors concentrated during the data analysis were annual income versus loan purpose, customer ‘s trust, loan tenure versus delinquent months, loan tenure versus credit category, loan tenure versus number of years in the current job, and chances for loan repayment versus the house ownership.

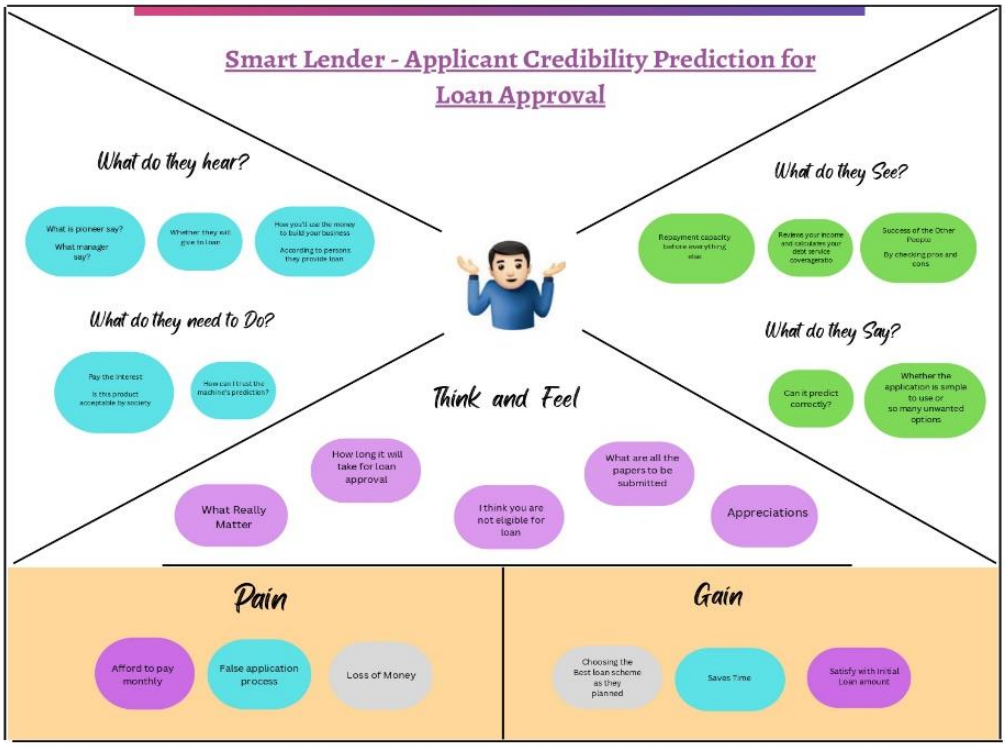
What is the factor may affect the processof loan?	1-Loan type and Loan terms 2-Credit History 3-Dept-to-Income Ratio 4-Employment History5-Down Payment
What is loan Management System?	A loan management system is a digital platform that helps automate every stage of the loan lifecycle, from application to closing.
What is the risk faced by management?	1-Security and Fraud Risk 2-Economic Risk 3-Operational Risk

When does the issue occur?	1-Poor Credit Rating 2-Too Much Debt in an Account3-Not Earning Enough 4-Too Small Deposit
Which factor may affect the Loan Process?	Work Experience Age Income Repay ment

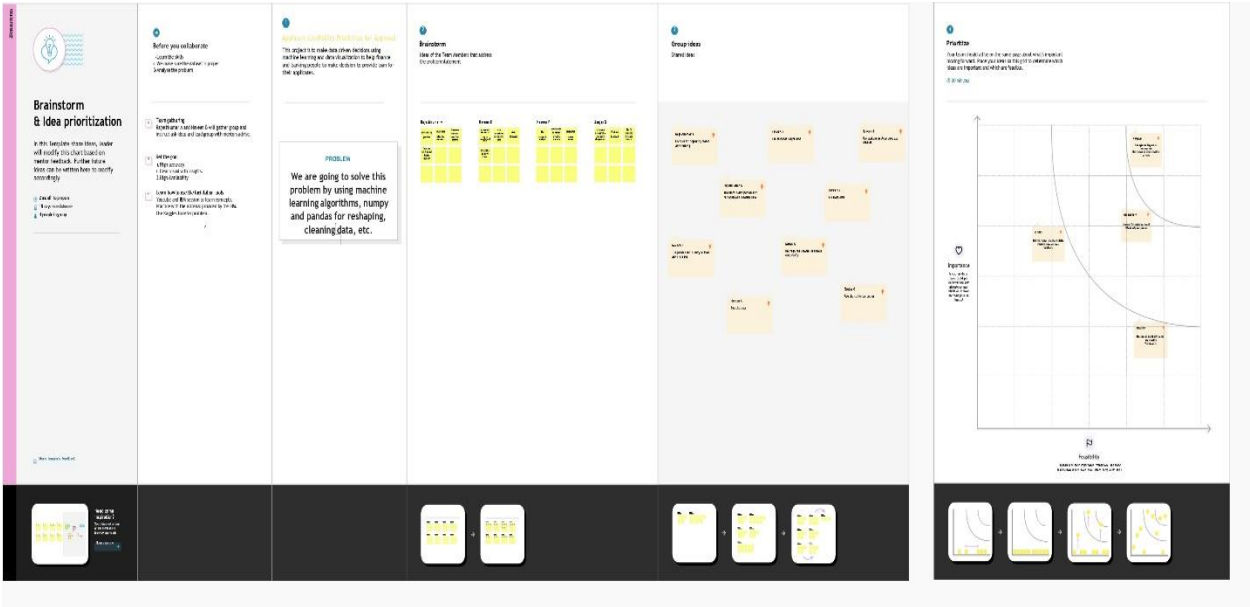
3.IDEATION & PROPOSED SOLUTION

3.1EMPATHY MAP CANVAS

Empathy Map



3.2 IDEATION & BRAINSTORMING



3.3 PROPOSED SOLUTION

To Predict Loan Approval by using Machine Learning. The Solution description is a classification problem in which we need to classify whether the loan will be approved or not. Classification refers to a predictive modeling problem. Based on the results, a modified prediction model will be created to ensure maximum accuracy and performance. The customer first applies for a home loan and after that, the company validates the customer eligibility for the loan. The prediction model not only helps the applicant but also helps the bank by minimizing the risk and reducing the number of defaulters. Social Impact can create an impact on crisis for the Bank and customer. This can be useful in

reducing the time and manpower required to approve loans and filter out the perfect candidates for providing loans results in customer satisfaction. A loan needs to be approved manually by the bank which means that person will be responsible for whether the person is eligible for the loan or not and calculating the risk associated with it. If the banks lose too much money, then it will result in a banking crisis. These banking crisis affects the economy of the country.

S.No.	Parameter	Description
1.	**Problem Statement (Problem to be solved)**	➢ To Predict Loan Approval by using Machine Learning.
2.	**Idea / Solution description**	➢ This is a classification problem in which we need to classify whether the loan will be approved or not. ➢ Classification refers to a predictive modeling problem. ➢ Based on the results, a modified prediction model will be created to ensure maximum accuracy and performance.

3.	**Novelty / Uniqueness**	➢ The customer first applies for a home loan and after that, the company validates the customer eligibility for the loan. ➢ The prediction model not only helps the applicant but also helps the bank by minimizing the risk and reducing the number of defaulters.
4.	**Social Impact / Customer Satisfaction**	▯ These create an impact on crisis forthe Bank and customer. ▯ This can be useful in reducing the time and manpower required to approve loans and filter out the perfect candidates for providing loans results in customer satisfaction ▯ If the banks lose too much money, then it will result in a banking crisis. These banking crisis affects the economy of the country.

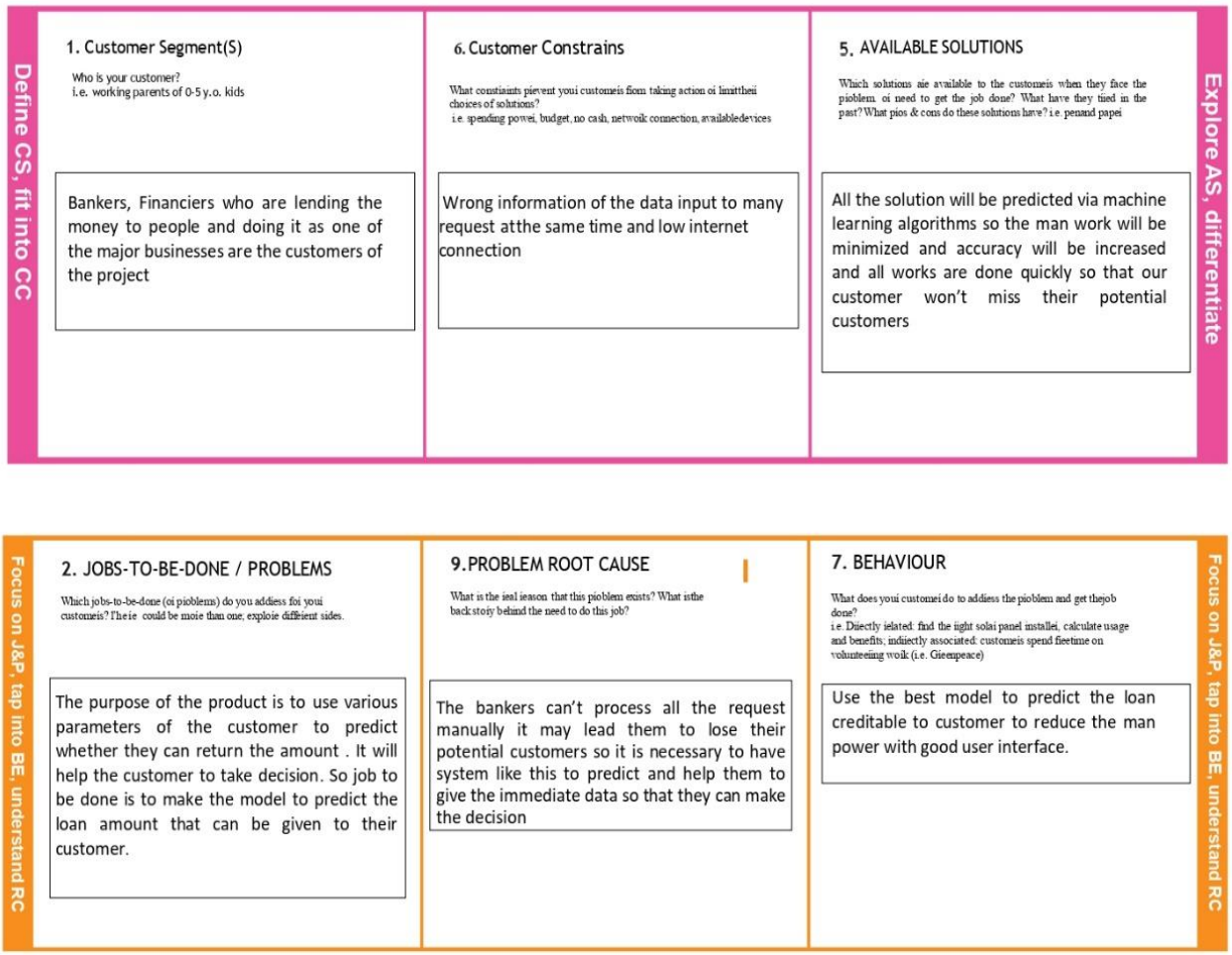
5.	**Business Model (Revenue Model)**	➢ A loan needs to be approved manually by the bank which means that person will be responsible for whether the person is eligible for the loan or not and also calculating the risk associated with it.

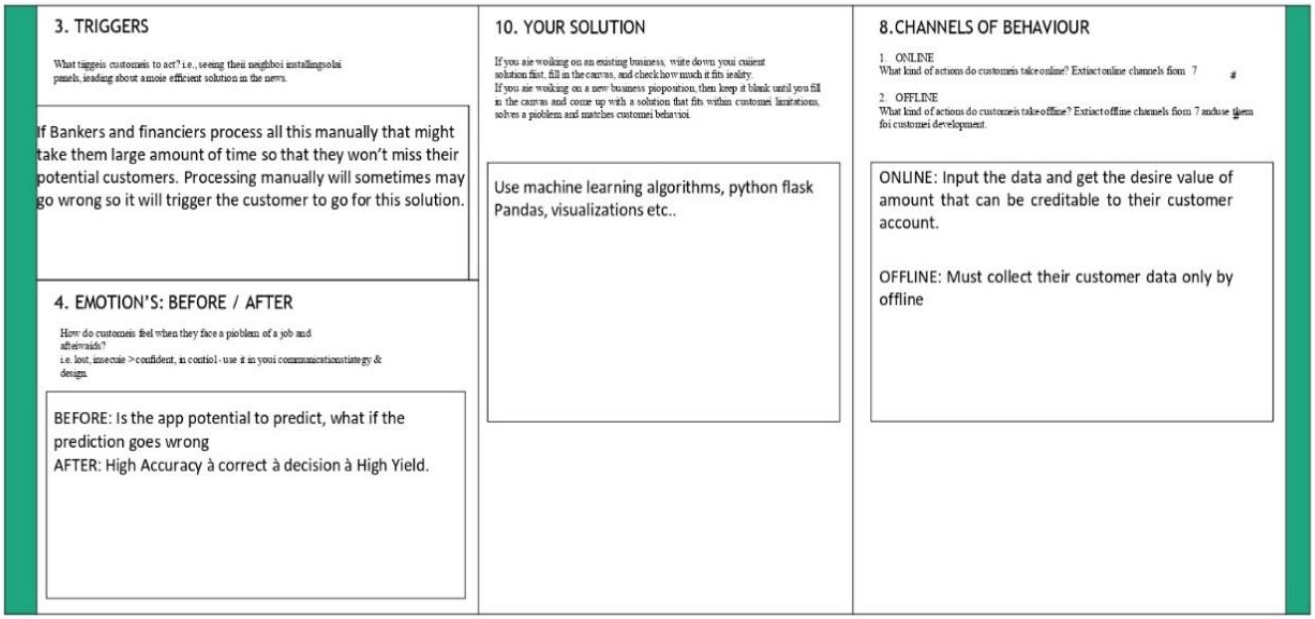
3.4 PROBLEM SOLUTION FIT

Project Title: Applicant Credibility Prediction for Loan Approval

Project Design Phase- I - Solution Fit

Team ID: PNT2022TMID15942





4. REQUIREMENT ANALYSIS

4.1FUNCTIONAL REQUIREMENT

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration via filling online application forms by Bank Official
FR-2	User Confirmation	Confirmation by Document Verification
FR-3	Log in to system	Check Credentials Check Roles of Access.

FR-4	Manage Modules	Manage System AdminsManage Roles of User Manage User permission
FR-5	Check whether details	Customer Detailswhether check Credit score and Employment status
FR-6	Log out	Exit

4.2 NON-FUNCTIONAL REQUIREMENTS

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The prediction model not only helps the applicant but also helps the bank by minimizing the risk and reducing the number of defaulters.
NFR-2	**Security**	Sensitive and private data must be protected from their production in Database server by using the different application results in high security.
NFR-3	**Reliability**	The shared protection achieves a better trade-off between costs and reliability. The model uses dedicated and shared protection schemes to avoid cyber-attack
NFR-4	**Performance**	A data science approach to predict and understand the applicant's profile to minimize the risk of future loan defaults and maximize the predictive capability of deeply understanding the past customer's profile minimizing the risk.

NFR-5	**Availability**	The company wants to automate the loan eligibility process (real-time) based on customer detail provided while filling out online application forms. These details are Gender, Marital Status, Education, number of Dependents, Income, Loan Amount, Credit History, and others.
NFR-6	**Scalability**	Scalability is a major concern for Data science. It has shown that different architectural choices of Machine Learning affect system scalability and find the complete details of the customers.

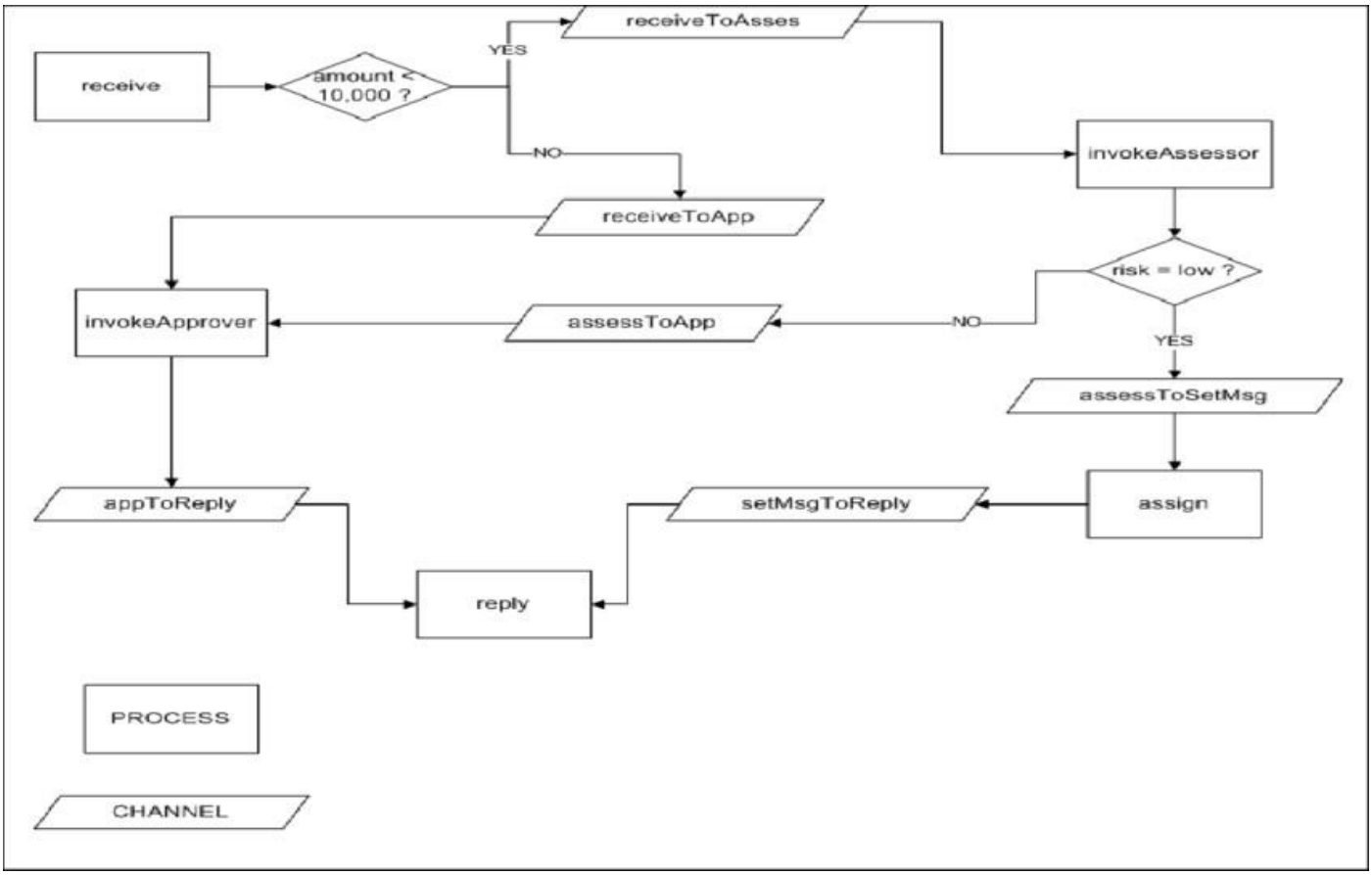
5. PROJECT DESIGN

5.1 DATA FLOW DIAGRAMS

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.

Banking processes use manual procedures to determine whether a borrower is suitable for a loan based on results. Manual procedures were mostly effective, but they were insufficient when there were many loan applications.

At that time, planning would take a long time. As a result, the loan prediction machine learning model can be used to assess a customer's loan status and build strategies.

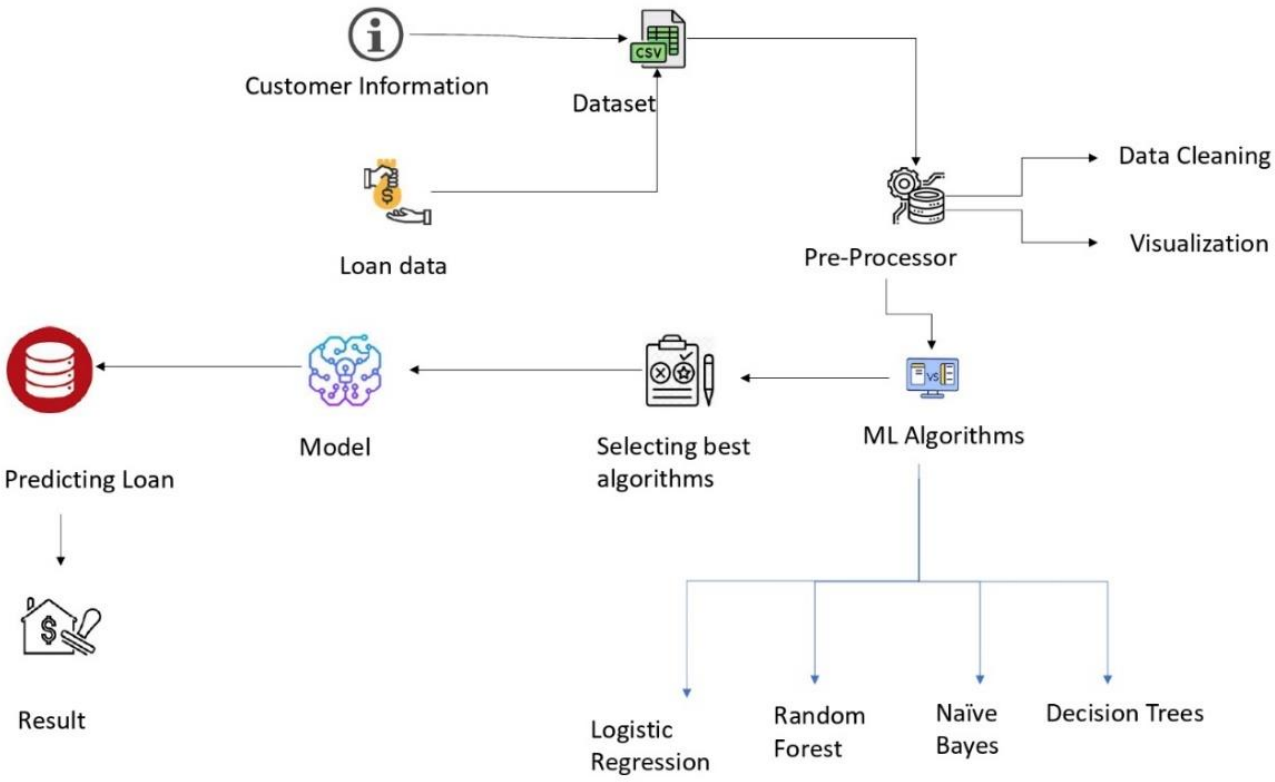


5.2 SOLUTION & TECHNICAL ARCHITECTURE

The dataset is obtained by gathering lot of required datasets and combining them to producea generalized dataset. The dataset thus produced is pre-processed i.e., the dataset is cleaned before doing data visualization. Then

the four algorithms are applied on the same pre- processed dataset and calculated for the best performed algorithms among them.

Then thebest algorithm is used to train the model and test it to check how accurate the algorithm can predict the output. Then we deploy that model to predict if bank loan can be approved or not for a specific candidate.



5.3USER STORIES

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	I can get my speed constraint utilizing climate application	I can get speed restrictions	High	Sprint-1
		USN-2	As a client, I can enroll for the application byentering my email, secret phrase, and confirmingmy secret phrase.	I can get to my account/dashboard	Medium	Sprint-2
		USN-3	As a client, I can increment or diminishing my speed as indicated by the weather conditions change	I can increment or decline my speed	High	Sprint-1
		USN-4	As a client, I could I at any point get my traffic redirection signs relying upon the traffic and the lethal circumstances	I can get to my traffic status ahead in my movement	Medium	Sprint-1
	Login	USN-5	As a client, I can sign out from the dark climate map by entering email and secret key	I can get to the application through my Gmail login	High	Sprint-2
	Interface	USN-6	As a client the connection point ought to be straightforward and effectively open	I can access the point of interaction without any problem	High	Sprint-1
Customer (Web user)	Data generation	USN-7	As a client I utilize open climate application to access the information in regards to the weather conditions changes.	I can get to the information concerning climate through the application	High	Sprint-1
Administrator	Problem solving/ Fault clearance	USN-8	As an in authority charge for the legitimate working of the sign sheets need to keep up with it through occasional observing	Authorities can screen the sign sheets for legitimate working.	Medium	Sprint-2

6. PROJECT PLANNING & SCHEDULING

6.1 SPRINT PLANNING & ESTIMATION

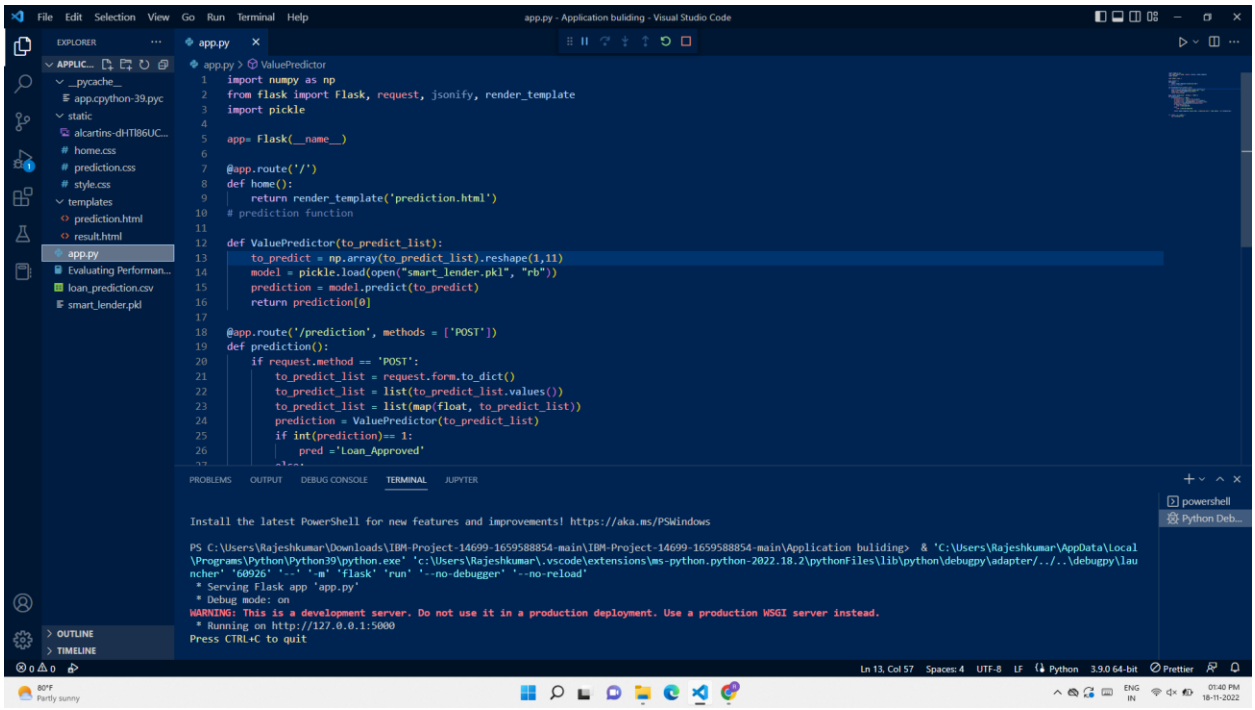
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority
Sprint-1	Data Preparation	USN-1	Collecting loan prediction dataset and pre-processingit	10	High
Sprint-1	Model Building	USN-2	Create an ML model to predict loan Approval	5	Medium
Sprint-1	Model Evaluation	USN-3	Calculate the performance, error rate,and complexity of the ML model and evaluate the dataset based on the parameter that the dataset.	5	Medium
Sprint-2	Model Deployment	USN-4	We need to deploy the model and needto find the prediction.	20	Medium
Sprint-3	Web page (Form)	USN-5	We can use the application by entering the Loan Prediction to analyze or predict the results.	20	Medium
Sprint-4	Dashboard	USN-6	We can predict the Loan Approval status by clicking the submit button and the application will show whether the status is approved or not.	20	High

6.2 SPRINT DELIVERY SCHEDULE

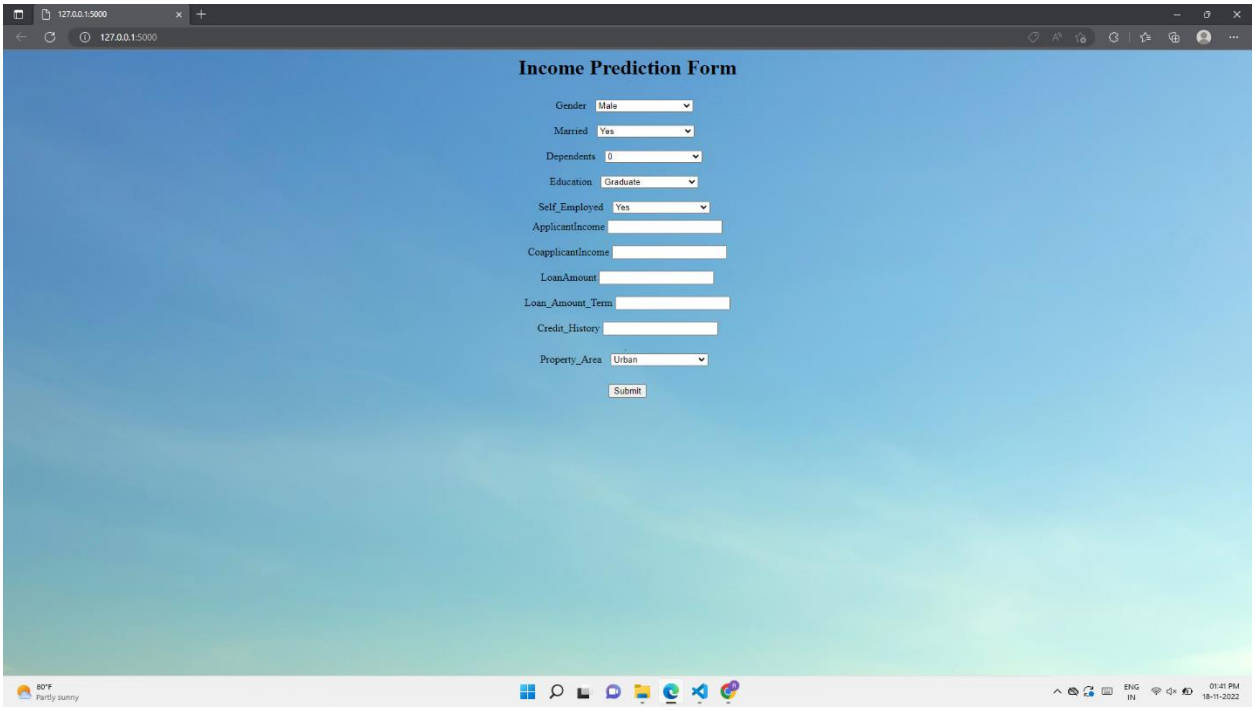
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date	Story Points Completed	Sprint Release Date
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	30 Oct 2022
Sprint-2	20	7 Days	30 Oct 2022	05 Nov 2022	20	06 Nov 2022
Sprint-3	20	7 Days	06 Nov 2022	12 Nov 2022	20	13 Nov 2022
Sprint-4	20	8 Days	13 Nov 2022	20 Nov 2022	20	20 Nov 2022

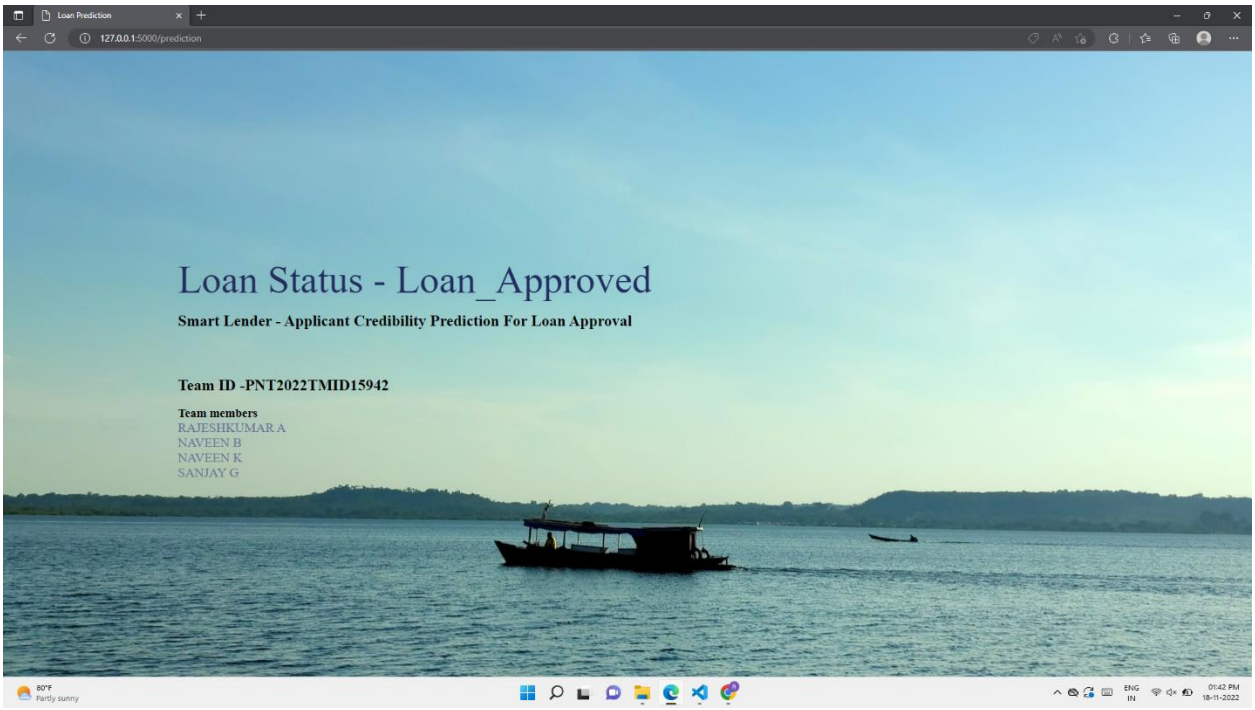
7. CODING & SOLUTIONING

7.1 FEATURE 1



7.2 FEATURE 2


127.0.0.1:5000
Income Prediction Form
Gender Male
Married Yes
Dependents 0
Education Graduate
Self_Employed Yes
ApplicantIncome
CoapplicantIncome
LoanAmount
Loan_Amount_Term
Credit_History
Property_Area Urban
Submit


Loan Prediction
127.0.0.1:5000/prediction
Loan Status - Loan_Approved
Smart Lender - Applicant Credibility Prediction For Loan Approval
Team ID -PNT2022TMID15942
Team members
RAJESHKUMAR A
NAVEEN B
NAVEEN K
SANJAY G

8 TESTING

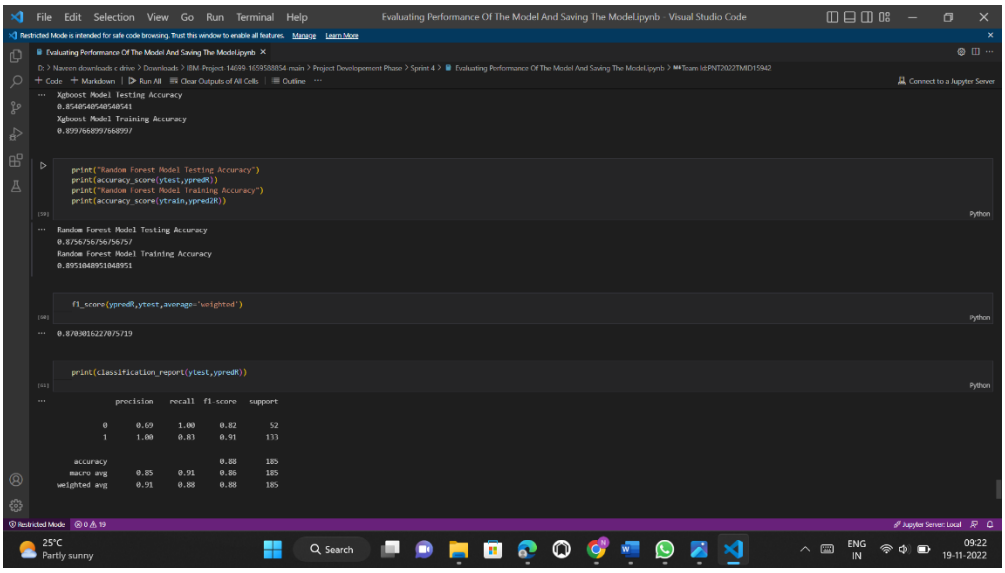
8.1 TEST CASES

USER TEST CASES:

Gender	Male
Married	No
Dependents	0
Education	Graduate
Self-employed	No
Applicant Income	5849
Co-applicant Income	0.0
Loan Amount	0
Loan_Amount_Term	360.0
Credit_History	1.0
Property_Area	urban

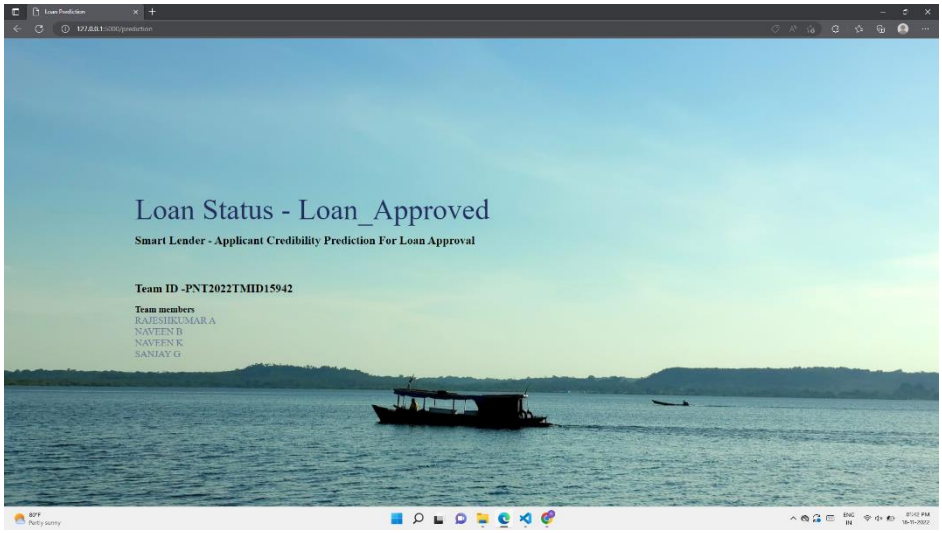
Output:Yes

8.2 USER ACCEPTANCE TESTING



9 RESULTS

9.1 PERFORMANCE METRICS



10 ADVANTAGES

- The loan is not repayable on demand and so available for the term of the loan - generally three to ten years - unless you breach the loan conditions.
- Loans can be tied to the lifetime of the equipment or other assets you are borrowing the money to pay for.

- At the beginning of the term of the loan you may be able to negotiate a repayment holiday, meaning that you only pay interest for a certain amount of time while repayments on the capital are frozen.
- While you must pay interest on your loan, you do not have to give the lender a percentage of your profits or a share in your company.
- Interest rates may be fixed for the term so you will know the level of repayments throughout the life of the loan.
- There may be an arrangement fee that is paid at the start of the loan but not throughout its life. If it is an on-demand loan, an annual renewal fee may be payable.

DISADVANTAGES

- Larger loans will have certain terms and conditions or covenants that you must adhere to, such as the provision of quarterly management information.
- Loans are not very flexible - you could be paying interest on funds you're not using.
- You could have trouble making monthly repayments if your customers don't pay you promptly, causing cashflow problems.

- In some cases, loans are secured against the assets of the business or your personal possessions, eg your home. The interest rates for secured loans may be lower than for unsecured ones, but your assets or home could be at risk if you cannot make the repayments.
- There may be a charge if you want to repay the loan before the end of the loan term, particularly if the interest rate on the loan is fixed.

11 CONCLUSION

From a proper analysis of positive points and constraints on the member, it can be safely concluded that the product is a considerably productive member. This use is working duly and meeting to all Banker requisites. This member can be freely plugged in numerous other systems. There have been mathematics cases of computer glitches, violations in content and most important weight of features is fixed in automated prophecy system, so in the near future the so – called software could be made more secure, trustworthy and dynamic weight conformation. In near future this module of prophecy can be integrated with the module of automated processing system. The system is trained on old training dataset in future software can be made resembling that new testing date should also take part in training data after some fix time.

12 FUTURE SCOPE

In future, this model can be used to compare various machine learning algorithm generated prediction models and the model which will give higher accuracy will be chosen as the prediction model.

13 APPENDIX

Source Code:

https://github.com/IBM-EPBL/IBM-Project-14699-1659588854/tree/main/Project%20Developement%20Phase/Sprint%204

GitHub Link:

https://github.com/IBM-EPBL/IBM-Project-14699-1659588854.git

Project Demo Drive Link:

https://drive.google.com/file/d/1g7Z00hq5dv0Qb2PXIn8w-wtvljBKeyxq/view?usp=sharing