Smart Lender - Applicant Credibility Prediction for Loan Approval

Submitted by:

1) Jagadeesh A (**Team Leader**)

2) Arul Samuel I

3) Badri Narayan J

4) Balakumaran M

1.INTRODUCTION

1.1 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 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.

1.2 PURPOSE

In the present scenario, a loan needs to be approved manually by a representative of 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. As it is done by a human it is a time consuming process

and is susceptible to errors. If the loan is not repaid, then it accounts as a loss to the bank and banks earn most of their profits by the interest paid to them. If the banks lose too much money, then it will result in a banking crisis. These banking crisis affects the economy of the country. So it is very important that the loan should be approved with the least amount of error in risk calculation while taking up as the least time possible. So a loan prediction model is required that can predict quickly whether the loan can be passed or not with the least amount of risk possible.

2.LITERATURE SURVEY

1. PREDICTION OF LOAN STATUS IN COMMERCIAL BANK USING MACHINE LEARNING CLASSIFIER

AUTHORS: G. Arutjothi,Dr. C. Senthamarai

Banking Industry always needs a more accurate predictive modelling system for many issues. Predicting credit defaulters is a difficult task for the banking industry. The loan status is one of the quality indicators of the loan. It doesn't show everything immediately, but it is a first step of the loan lending process. The loan status is used for creating a credit scoring model. The credit scoring model is used for accurate analysis of credit data to find defaulters and valid customers. The objective of this paper is to create a credit scoring model for credit data. Various machine learning techniques are used to develop the financial credit scoring model. In this paper, we propose a machine learning classifier-based analysis model for credit data. We use the combination of Min-Max normalization and K Nearest Neighbor (K-NN) classifier. The objective is implemented using the software package R tool. This proposed model provides the important information with the highest accuracy. It is used to predict the loan status in commercial banks using machine learning classifier.

2. OVERDUE PREDICTION OF BANK LOANS BASED ON LSTM-SVM

AUTHOR: Xin Li, Xianzhong Long, Guozi Sun, Geng Yang, and Huakang Li

In the aspect of bank loans, the accuracy of traditional user loan risk prediction models, such as KNN, Bayesian,DNN, are not benefit from the data growth. This article is based on the work of Overdue Prediction of Bank Loans Based on Deep Neural Network. And we propose to analyze the dynamic behavior of users by LSTM algorithm, and use the SVM algorithm to analyze the user's static data to solve the current prediction problems. This article uses users basic information, bank records, user browsing behavior, credit card billing records, and loan time information to evaluate whether users are delinquent. These static data are the basic input for SVM. For LSTM model, we extract user's recent transaction type from browsing behavior as input to LSTM, to predict the probability of users' overdue behavior. Finally, we calculate the average of the two algorithms as the final result. From the experimental results, this LSTMSVM model shows a great improvement than traditional algorithms.

## 3. THEORITICAL ANALYSIS
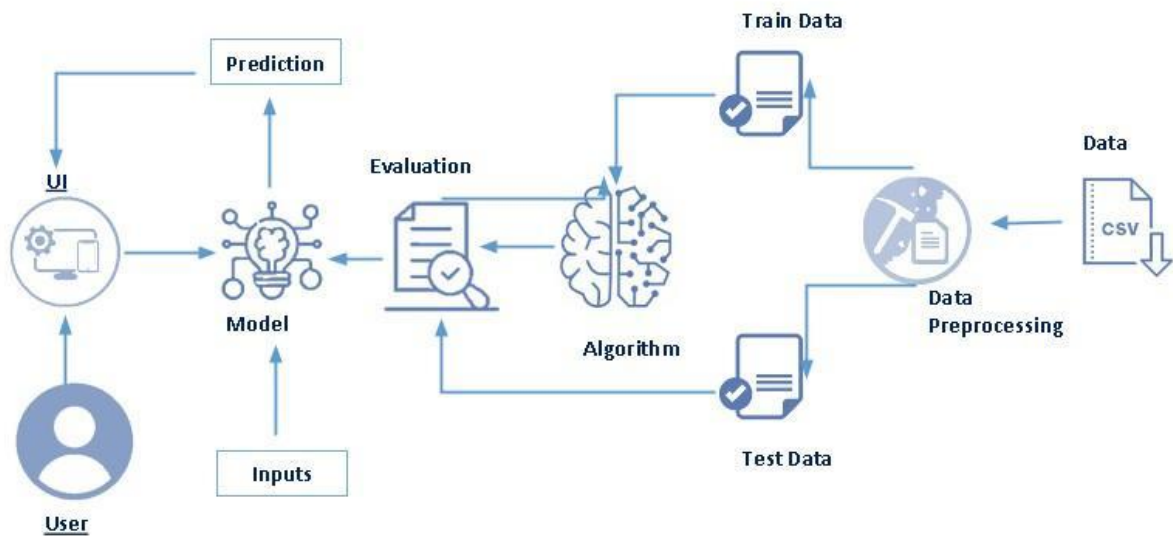
4. EXPERIMENTAL INVESTIGATIONS

A good amount of people seeking loans in India has enlarged to a greater extent and the reasons for it could be many. The Employees working in banking sectors are deprived of knowledge to judge or foresee whether a customer (good or poor) will be able to pay the debt of the loan at the stipulated interest rate. Throughout the financial system, banking institution offer a variety of services, but lines of credit remain their primary and the biggest source of income. Therefore, banking businesses will prosper from the revenue produced on the mortgages they make. Lending, or whether consumers return the money or default on one's loans, impact a banking institutions' financial statement. The banking institutions non-performing investments will be diminished by estimating mortgage. Consequently, more exploration into this event happening is needed. Because detailed estimations are crucial and essential for sufficient service, various methodologies must be assessed and analysed. Our study and work research seeks to provide a comprehensive review of lending estimation systems and structures that employ prediction methods and techniques flourished and developed after recent years. In this study and paper, researchers studied the learning techniques as well as the raw datasets utilized for training and test sets. The

system model's precision is also discussed. Our work also provides a quick overview of a few datasets that can be used to anticipate loan/mortgage analysis. Recent and future trends are also spotlighted.

5. FLOWCHART

- ➤ User is provided with an UI of a form in which they have to fill in the details
- ➤ Depending inputs provided by the user the eligibility status of their loan is displayed on the page
- ➤ Once model analyses the gesture, the prediction with operation applied on image is showcased on the UI. To accomplish this, we have to complete all the activities and tasks listed below:
- ➤ Data Collection:
  - ✓ Collect the dataset or Create the dataset
- ➤ Data Pre-processing:
  - ✓ Import the dataset library.
  - ✓ Checking for null values.
  - ✓ Handling categorical columns.
  - ✓ Balancing the dataset.
  - ✓ Scaling the dataset.
  - ✓ Splitting the data into train and test
- ➤ Model Building:
  - ✓ Apply the train and test dataset Decision Tree model.
  - ✓ Apply the train and test dataset Random Forest Model.
  - ✓ Apply the train and test dataset KNN model.
  - ✓ Compare the models based on metrics and choose the best of it.
- ➤ Application Building:
  - ✓ Build HTML pages.
  - ✓ Build python code.
  - ✓ Run the application
- ➤ Train the model on IBM:
  - ✓ Register for IBM cloud.
  - ✓ Train the ML model on cloud.
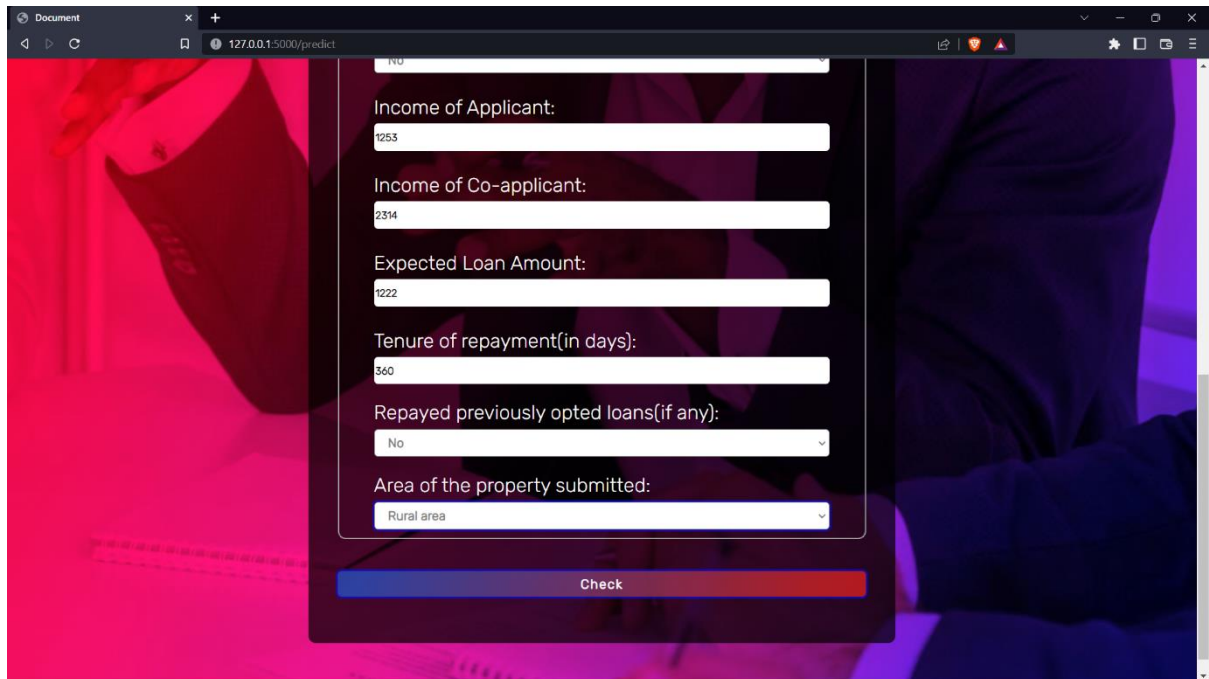 - ✓ Integrate flask app with scoring end point
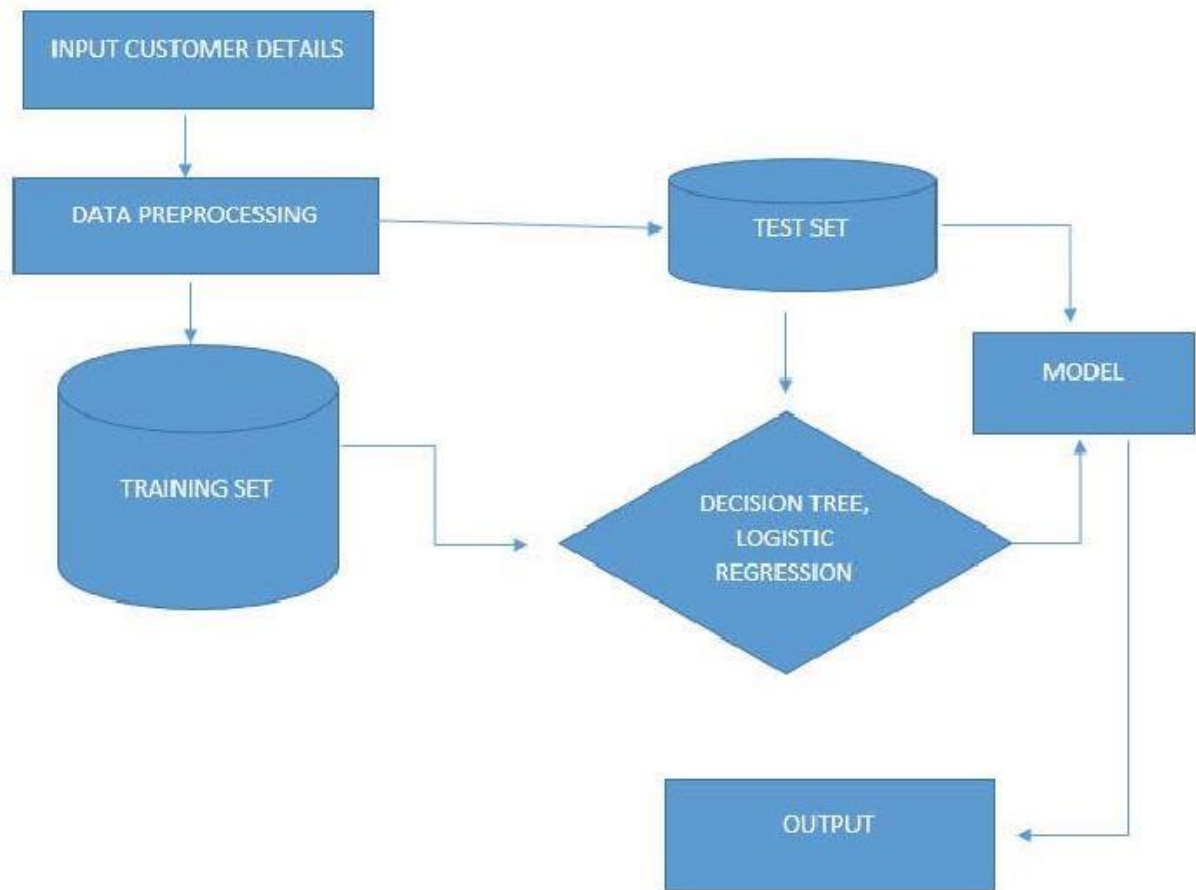
6. RESULT

Final user input and output of the local and IBM deployed flask application:

Homepage of our application website
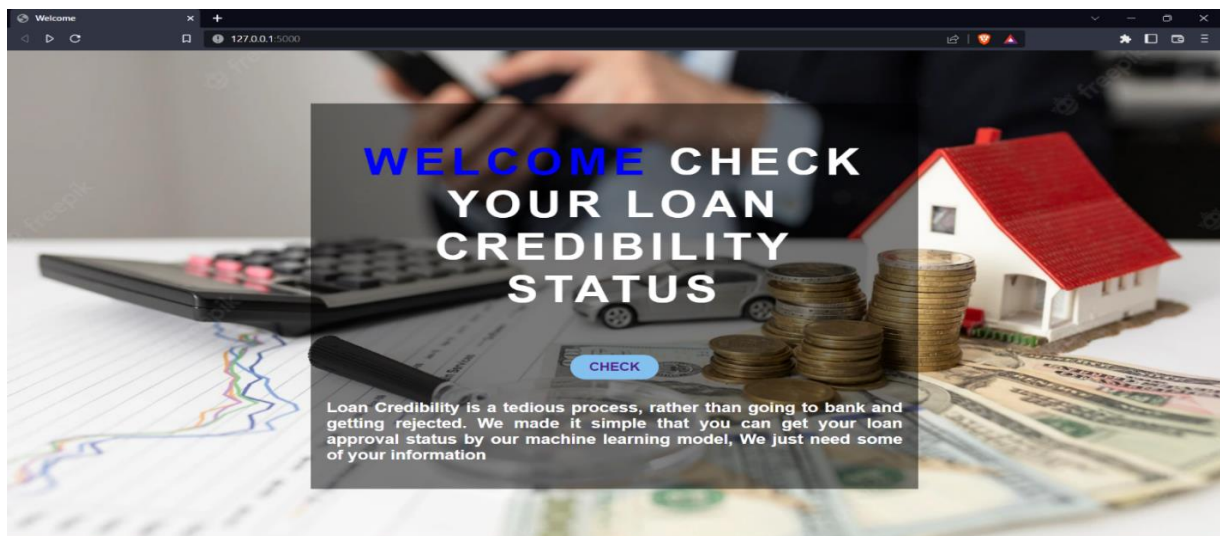
User input and output page of our flask application where the user gets a yes for eligibility of loan application
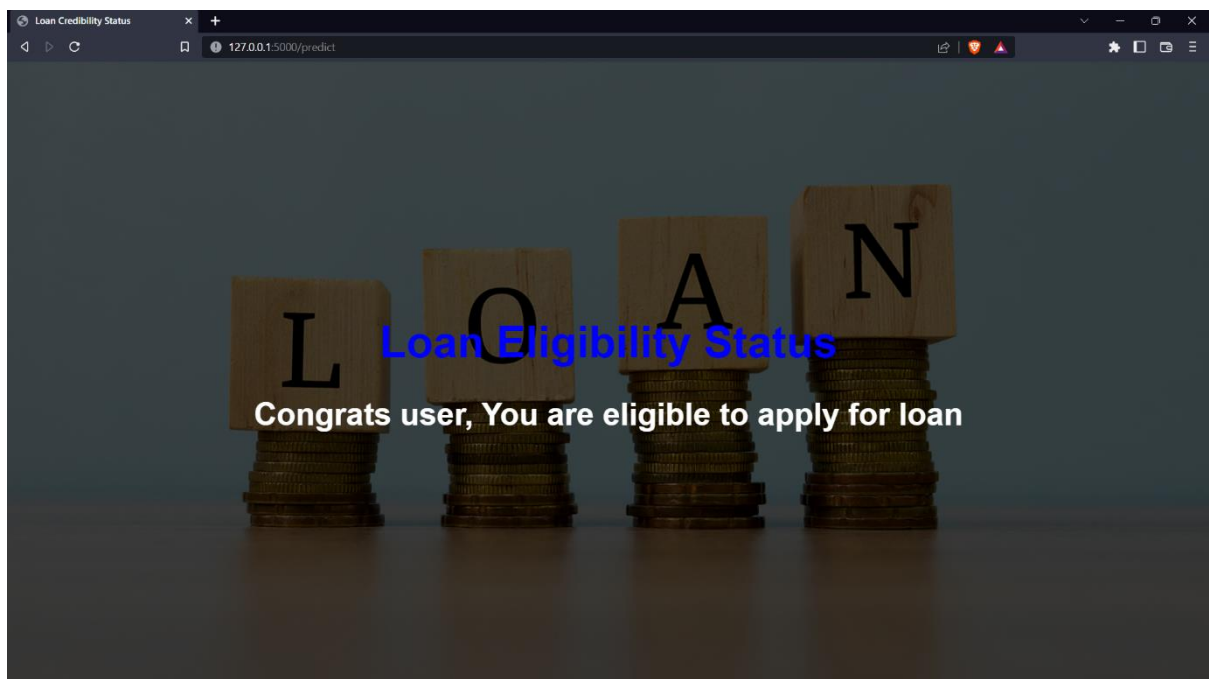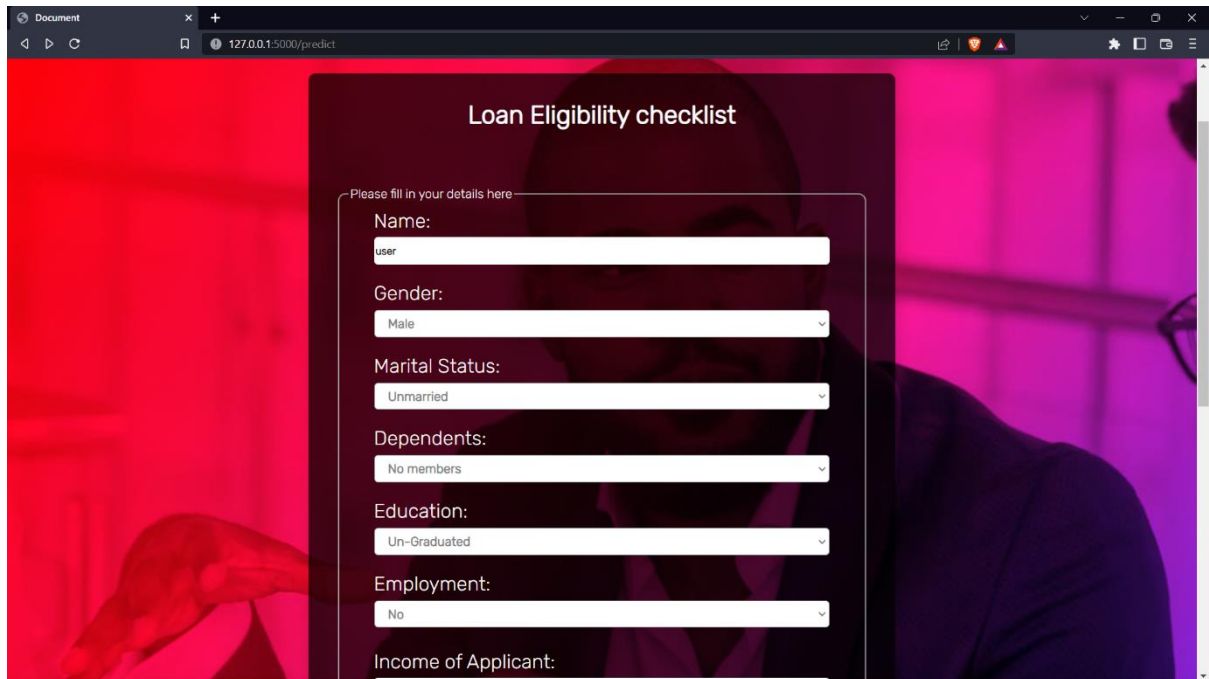
INPUT:

OUTPUT:

User input and output page of our flask application where the user gets a No for eligibility of loan application

INPUT:

OUTPUT:
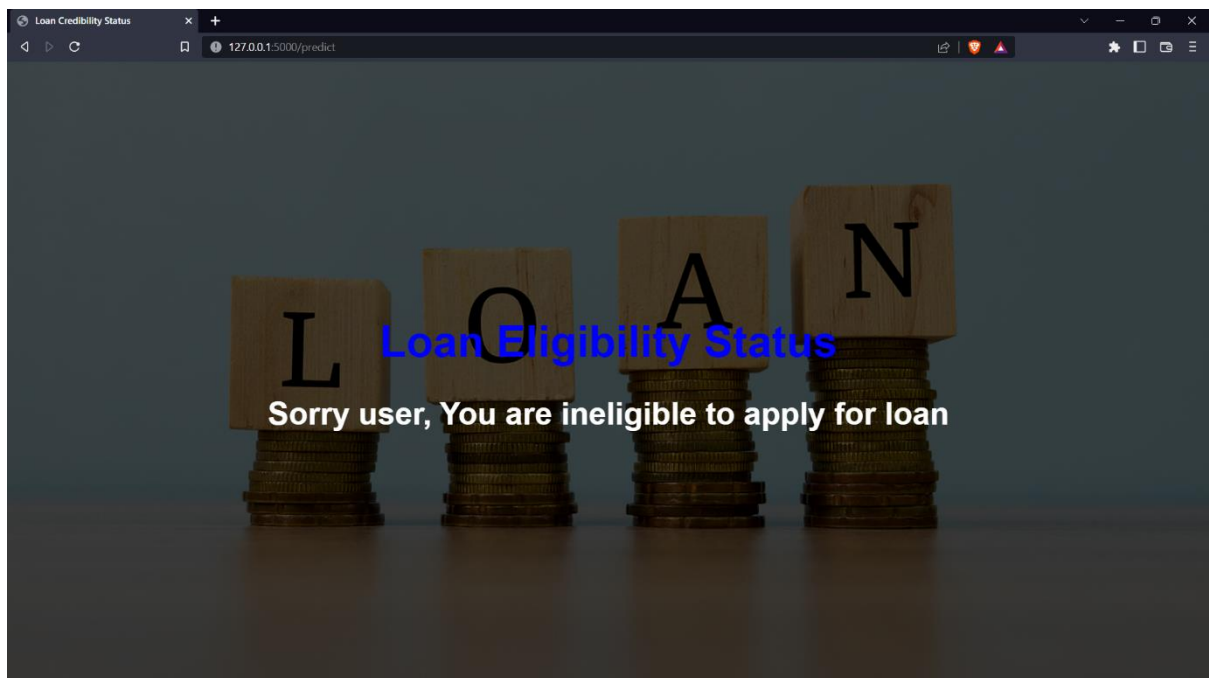
7. ADVANTAGES AND DISADVANTAGES

ADVANTAGES:

Flexibility: A bank loan allows one to repay as per convenience as long as the instalments are regular and timely

Low Interest Rates: Generally, bank loans have the cheapest interest rates. The rates you pay will be cheaper than other types of high interest loans, such as venture capital

Maintain Control: You don't have to give up equity to get a loan from a bank.

DISADVANTAGES:

Requires Profitability: While venture capitalists and angel investors usually take risks to invest in companies that haven't yet proved profitable, banks will take no such risk.

Complicated: Obtaining a bank loan is extremely time consuming. You will be required to fill out excessive paperwork, and the terms of interest will be quite complicated.

Collateral: Regardless of your profitability or how good your credit score happens to be, banks will need some form of collateral. Banks need to protect themselves in the case that you can't make your payments.

8. APPLICATIONS

Loan sanctioning and credit scoring forms a multi-billion dollar industry in the US alone. These companies can earn more by reducing the workload of the employees

The loan approval or rejection has enormous ramifications for both the applicant and the bank, causing possible opportunity costs for both parties. So the users can check for loan eligibility status before getting into bank.

9. CONCLUSIONS

In this project we built a full stack web application which gets some user input required for prediction of loan eligibility criteria. It is designed to match the criteria of human level thinking. It might not be accurate all the time but after application of Machine learning we get rounded accuracy scores for training to be 81% and testing to be 86% based on previously available system's dataset.

This tool can be used to escape the tediousness of the previous process as the user gets to know the status within fraction of seconds after application of input. As this process is online the user don't need to be available at the bank.

Using this system reduces the effort of bank officers and make workers efficiently work in a relaxed manner

10. FUTURE SCOPE

This data science model of ML can be used for the next 10 to 15 years as banks like **Wells Fargo** and **Morgan Stanley** have looked at the use of AI in determining lending risk and developing a loan prediction system in recent years to overcome human bias and delays in the application processing time. Despite using data science in this process, there is still a large amount of manual work involved. Researchers have recently explored the possibility of using deep learning in various aspects of this process.

11. REFERENCES

https://www.projectpro.io/article/loan-prediction-using-machine-learning-project-source-code/632

https://towardsdatascience.com/predict-loan-eligibility-using-machine-learning-models-7a14ef904057

[1] Vaidya and Ashlesha, Predictive and probabilistic approach using logistic regression: Application to prediction of loan approval, 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 2017.

[2] Amin, Rafik Khairul and Yuliant Sibaroni, Implementation of decision tree using C4. 5 algorithm in decision making of loan application by debtor (Case study: Bank pasar of Yogyakarta Special Region), 2015 3rd International Conference on Information and Communication Technology (ICoICT). IEEE, 2015.

[3] Arora, Nisha and Pankaj Deep Kaur, A Bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit risk assessment, Applied Soft Computing 86 (2020), 105936.

[4] Yang, Baoan, et al, An early warning system for loan risk assessment using artificial neural networks, Knowledge-Based Systems 14.5-6 (2001), 303-306.
[5] Metawa, Noura, M. Kabir Hassan and Mohamed Elhoseny, Genetic algorithm based model for optimizing bank lending decisions, Expert Systems with Applications 80 (2017), 75-82.

[6] Hassan, Amira Kamil Ibrahim and Ajith Abraham. "Modeling consumer loan default prediction using ensemble neural networks, 2013 International Conference On Computing, Electrical And Electronic Engineering (ICCEEE). IEEE, 2013.