

FORECASTING LOAN APPROVAL USING MACHINE LEARNING MODEL

Omkar Shete, Anshuman Pratyush, Tejas Agrawal, Dr Chaya Jadhav

Dept. of Computer Science Engineering

Dr. D.Y. Patil Institute of Technology, Pune, Maharashtra,

Abstract - Banks are making a significant part of their profits through loans. Loan approval is a very important process for banking organizations. It is tough to predict the possibility of payment of loans by the customers because there is an increasing rate of loan defaults and the banking authorities are finding it more difficult to correctly access loan requests and tackle the risks of people defaulting on loans. In this case, the right prediction would be very beneficial using some machine learning algorithm. In this work, we use a machine learning technique such as a Random Forest classifier that will predict the person who is reliable for a loan. This work's primary objective is to predict whether the loan approval to a specific individual is safe or not.

Key Words: Machine learning, Random Forest classifier, classification, big data, Artificial Intelligence.

1.INTRODUCTION

In many regions of the world, banks, home finance companies, and certain NBFCs deal with different loan kinds such as mortgages, personal loans, business loans, etc. These businesses operate in rural, semi-urban, and urban settings. Following a customer's loan application, these businesses confirm that the customer is eligible to receive the loan or not. So by using a machine learning algorithm, this system offers a way to automate this process. In order to apply for a loan, the consumer will complete out an online form. This form includes information about the applicant's sex, marital status, qualifications, dependents' specifics, annual income, loan amount, and credit history, among other things. Previous research in this era has shown that there are so many methods to study the problem of controlling loan default. But as the right predictions are very important for the maximization of profits, it is essential to study the nature of the different methods and their comparison. A very important approach in predictive analytics is used to study the problem of predicting loan defaulters:

The models are compared on the basis of the performance measures such as sensitivity and specificity. The final results have shown that the model produce different results. Model is marginally better because it includes variables (personal attributes of customer like age, purpose, credit history, credit amount, credit duration, etc.) other than checking account information (which shows wealth of a customer) that should be taken into account to calculate the probability of default on loan correctly.

2.LITERATURE SURVEY

1. The authors of this paper have described how the banking industry has developed. Many people are applying for bank loans, but because banks have finite resources, they can only lend to a certain number of applicants. As a result, a typical process involves deciding who the bank can lend to and who will be a safer option for it. Using machine learning, it is possible to automate the process of determining whether a loan should be authorised or not. This is done by searching Big Data for prior records of the borrowers who have received loans in the past, after which machine learning techniques are used to train the computer based on these records/experiences. There are several ways to analyse the issue of loan default control, according to earlier research in this era. However, as making accurate predictions is essential for maximising profits, it is crucial to look into and contrast the different approaches. This study discusses and evaluates a variety of machine learning methods that have been applied in the past [1].
2. The author of this essay has described how technology has advanced in the banking industry. Also, there have been numerous improvements in the financial industry. Every day, there are more applications to approve loans. When choosing an applicant for loan approval, they must take into account a few bank policies. The bank has to choose which is ideal for approval based on a few criteria.

To manually check out each person before recommending them for loan approval is difficult and dangerous. Based on the prior performance of the person to whom the loan amount was previously accredited, we utilise a machine learning technique in this study to forecast the person who is trustworthy for a loan. The main goal of this activity is to determine whether or not approving a loan to a certain person is safe. [3]

3. In this essay, the author has discussed how estimation or assessment of default on a debt is an important process that banks should carry out to help them determine whether or not a loan applicant can be a defaulter at a later stage so that they can process the application and decide whether or not to approve the loan. The result drawn from these analyses aids banks and other financial organisations in reducing losses and ultimately increasing the quantity of credits. As a result, it is crucial to build a model that will consider all of an applicant's characteristics and produce a result pertaining to the application in question. In today's technologically advanced world, criminal activity is carried out using all possible ways to borrow the funds from their illegal operations. The rising problem of bank distress within the economic system is reflected in the rising amount of bad debts coming from commercial banks' lending. In order to assist the banks make better judgements in the future, we have employed data mining techniques to forecast the potential defaulters using a dataset that comprises information on house loan applications. [4]
4. In order to be able to profit from the interest on those loans that they credit, the author of this paper has described credit scores. The profitability or loss of a bank is mostly determined by the loans it makes, namely whether or not its clients are making their loan repayments. The bank can lower its non-performing assets by foreseeing loan defaulters. Therefore, by assessing a customer's chance of defaulting on a loan, a logistic regression technique makes it simple to identify the ideal clients to target for loan approval. The model suggests that a bank should consider all of a customer's traits, which are crucial when deciding whether to issue credit rather than only targeting wealthy consumers. [5]
5. The author of this paper has explained how the financial world has changed and how many people are now applying for bank advances. However, banks only have a limited amount of resources that they can allocate to specific people, so it is common

interaction to determine who can be granted credit in order to make the bank feel safer. Therefore, in this activity, we work to lessen the risk associated with selecting the covered person in order to conserve a significant amount of bank resources and efforts. This is accomplished by analysing historical records of those to whom advances have previously been granted, and based on these records and interactions, a machine was built using an AI model that produces the most accurate results. The main goal of this essay is to predict whether or not relegating the advance to a certain person will be safeguarded. There are four sections in this essay. Data gathering, AI model comparisons based on the data, and data collection. Testing (iv) Training the framework using the most motivating model. [6]

6. In this essay, the author explains how predicting consumer loans is a problem that affects all retail banks during the course of their existence. If done correctly, it may save a lot of man hours when a retail bank closes. If the business decides to partially automate the loan acceptance process (in real time), it will do so using the information that customers enter while filling out an online application. Gender, Marital Status, Education, Dependents, Income, Loan Amount, Credit History, and others are some of these minor factors. They have made it difficult to identify the consumer groups that are permitted for the whole loan amount in order to automate this approach so that they may more precisely target these clients. Predicting whether or not a loan would be accepted is necessary. In a classification task, we must forecast distinct values depending on a predetermined set of independent variables. Our goal is to develop a machine learning model to categorise a dataset obtained from Kaggle with the highest level of accuracy possible. Using Python assistance on a Jupyter notebook, the random forest classification algorithm exhibits the highest degree of accuracy when categorising provided online loan candidates. [8]

3.METHODOLOGY

To deal with the problem, we developed automatic loan prediction using machine learning techniques. We will train the machine with previous dataset. so, machine can analyse and understand the process. Then machine will check for eligible applicant and give us result.

1. **Dataset Collection:** A data set is a collection of related, discrete items of related data that may be accessed individually or in combination or managed as a whole entity. The entire dataset is in un-processed form

consists of categorical data and descriptions. This dataset is real-time data which illustrates loan administration experience.

2. **Data Pre-processing:** Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine-learning model. Data pre-processing is a required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

3. **Feature Engineering and Data Splitting:** Feature Engineering involves feature scaling and feature selection. Through Machine learning pipelines feature engineering had performed using normalization and standardization techniques.

4. **Random Forest classifier:** Random Forest Classifier is a machine learning algorithm that combines multiple decision trees to make predictions. It creates an ensemble of decision trees, each trained on a random subset of the training data and features. During prediction, each tree independently generates a prediction, and the final prediction is determined through majority voting or averaging. Random Forest Classifier is known for its robustness, ability to handle high-dimensional data, and resistance to overfitting. It is widely used for classification tasks, providing accurate and reliable predictions in various domains.

4.SYSTEM ARCHITECTURE

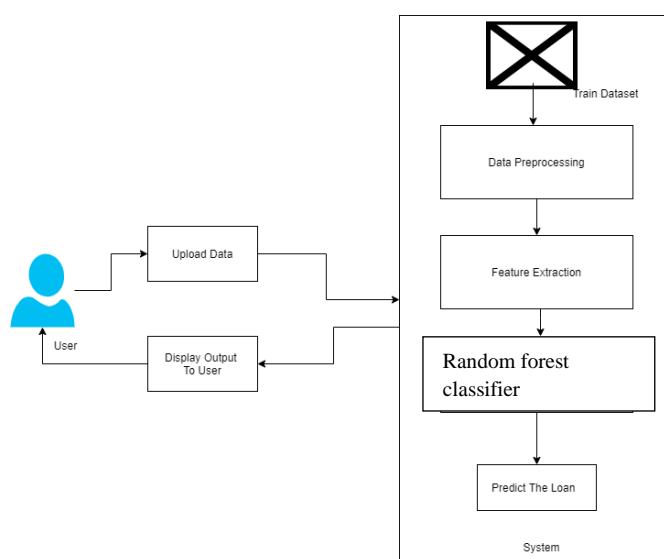


Fig 1. Architecture Diagram

4.ALGORITHM

Random forest classifier: Random Forest is a supervised learning algorithm which is used for both classification as well as regression. But however, it is mainly used for classification problems. As we know that a forest is made up of trees and more trees means more robust forest. Similarly, random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result.

Pros and Cons of Random Forest

PROS: -

- It overcomes the problem of overfitting by averaging or combining the results of different decision trees.
- Random forests work well for a large range of data items than a single decision tree does.
- Random forest has less variance than single decision tree.

CONS: -

- Complexity is the main disadvantage of Random Forest algorithms.
- Construction of Random forests are much harder and time-consuming than decision trees.
- More computational resources are required to implement Random Forest algorithm.

5.FLOWCHART

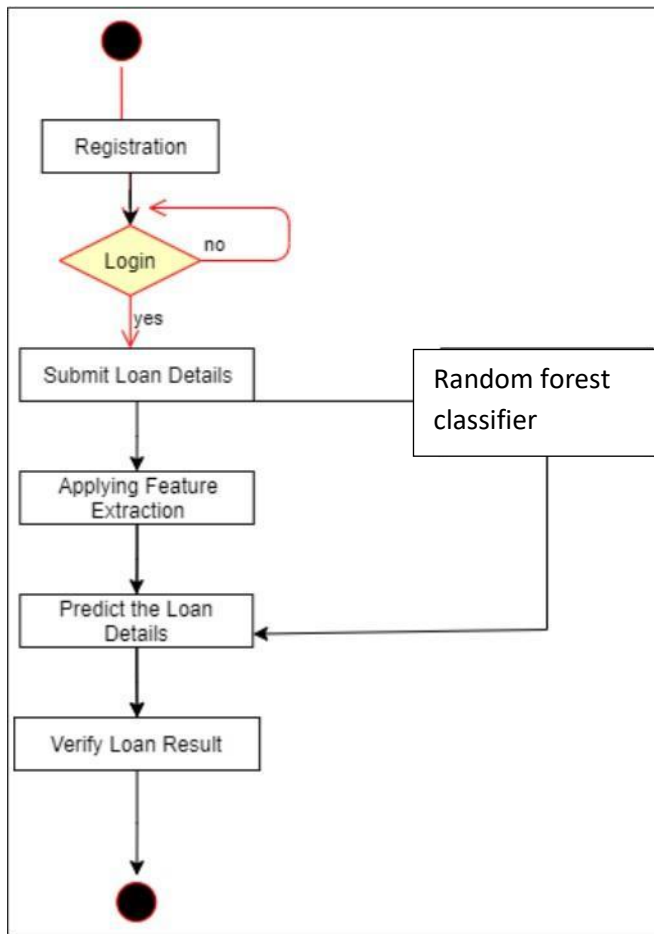


Fig 2. Flowchart

6.SYSTEM SPECIFICATION

- Processor: Intel Core i5
- Speed: 2.80 GHz
- RAM: 8GB
- Hard disk: 40GB

Software Specifications

- Operating system Windows 7 or further.
- IDE: Anaconda
- Coding Language: Python

Result

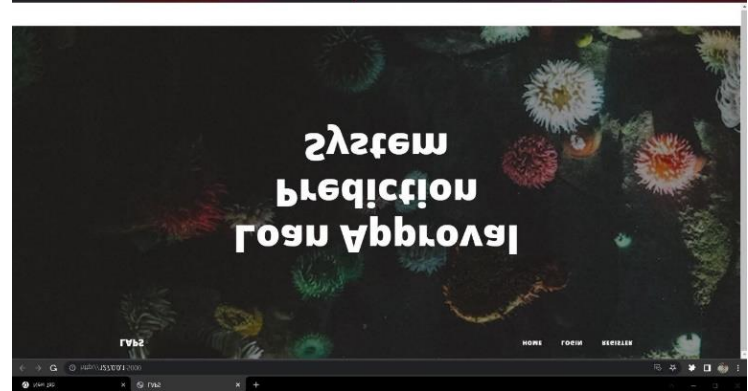


Fig 3 GUI Main Page

GUI main page: - This is the main start page of the project. It consists of login and register options, if users are already registered, they can choose login option otherwise they can use register option to register themselves.

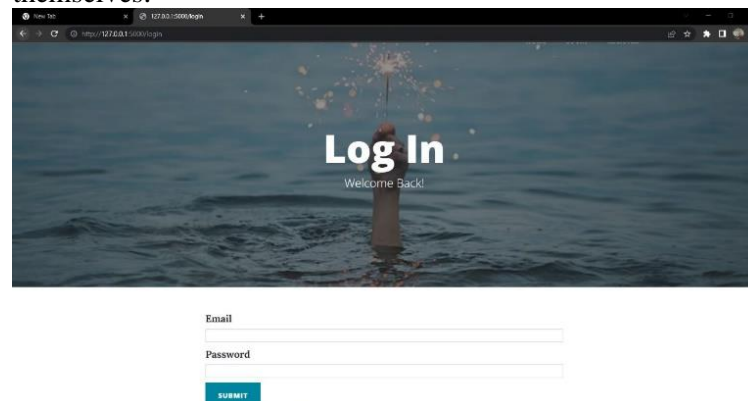


Fig 4 Login Page

Login Page: - This is the login page through which users can access their account by entering their username and password.

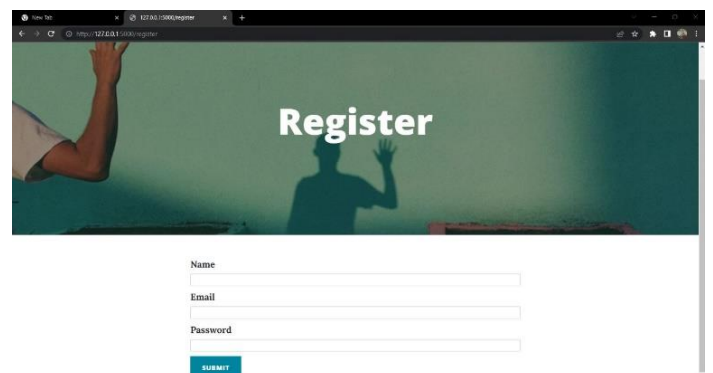


Fig 5 Registration Page

Registration Form: This is registration form which first time users can use to register themselves by entering the required details.

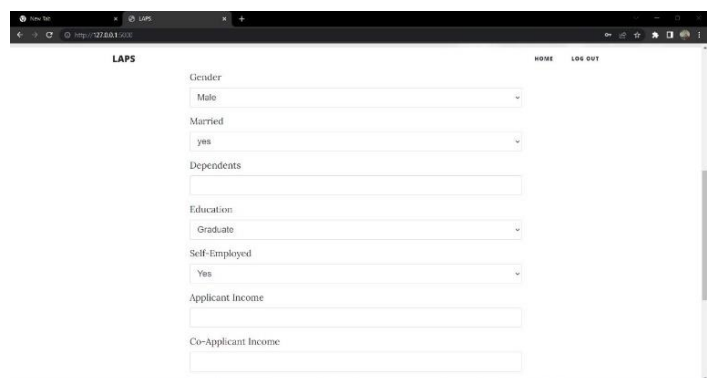
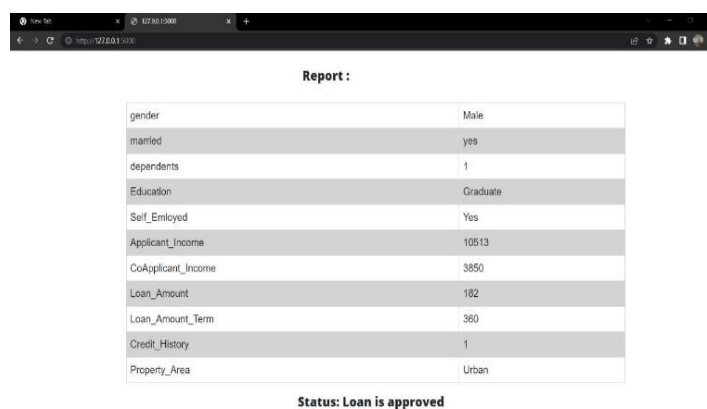


Figure 6 Input.



Report :	
gender	Male
married	yes
dependents	1
Education	Graduate
Self_Employed	Yes
Applicant_Income	10513
CoApplicant_Income	3850
Loan_Amount	182
Loan_Amount_Term	360
Credit_History	1
Property_Area	Urban

Status: Loan is approved

Fig 7 Input output

Input Output - This the credit analysis page, here the user enters his/her information and the machine learning (Random Forest classifier) model predicts the outcome. The outcome is either the loan will be approved or not. Blue bar text is for positive result (i.e., approved) and red bar text is for negative result (i.e., rejected).

7.CONCLUSION AND FUTURE WORK

This technique is used to forecast whether or not a candidate's loan application will be granted. The model also assists banks in determining whether or not a candidate is qualified or not.

The current model is trained on historical datasets. In the future, we will update the datasets dynamically over time.

The accuracy of the model will be improved significantly after training it with extensive datasets.

8.REFERENCES

- [1]. Ritika Purswani, Sakshi Verma, Yash Jaiswal, Prof. Surekha M, "Loan Approval Prediction using Machine Learning: A Review", IEEE 2020.
- [2]. Pidikiti Supriya, Myneedi Pavani, Nagarapu Saisushma, Namburi Vimala Kumari, K Vikas, "Loan Prediction by using Machine Learning Models", International Journal of Engineering and Techniques - Volume 5 Issue 2, Mar-Apr 2019
- [3]. Anshika Gupta, Vijay Pant, Sudhanshu Kumar, Pravesh Kumar Bansal, "Bank Loan Prediction System using Machine Learning", Conference: 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), September 2021.
- [4]. Bhoomi Patel, Harshal Patil, Jovita Hembram, Shree Jaswal. "Loan Default Forecasting using Data Mining", International conference for emerging technology, 2020.
- [5]. Mohammad Ahmad Sheikh, Amit Kumar Goel, Tapas Kumar. "An Approach for Prediction of Loan Approval using Machine Learning Algorithm". IEEE 2020.
- [6]. Yash Divate, Prashant Rana, Pratik Chavan. "Loan Approval Prediction Using Machine Learning". International Research Journal of Engineering and Technology (IRJET), Volume: 08 Issue:05 | May 2021
- [7]. Vishal Singh, Ayushman Yadav, Rajat Awasthi. "Prediction of Modernized Loan Approval System Based on Machine Learning Approach". IEEE 2021.
- [8]. L. Udhyam Bhanu, Dr. S. Narayana. "Customer Loan Prediction using Supervised Learning Techniques". International journal of scientific and Research publications, 2021.
- [9]. Nikhil Bansode, Adarsh Verma, Abhishek Sharma, Varsha Bhole. "Predicting Loan Approval using ML". International research journal of Engineering and Technology, 2021.
- [10]. Subhiksha R, Vaishnavi L, Shalini B, N. Manikandan. "Bank loan Approval Prediction using Data Science". International journal for research in applied science and engineering technology. 2022.