Home Loan Eligibility Prediction using Machine Learning

Amritraj Singh[1], Suraj Bhikaji Raut[2], Vaibhav Kadam[3],*Vikas More

Department of Computer Engineering

St. John College of Engineering and Management

Abstract: *In developing countries like India, the home is a dream of most of the people living during a village also as slum in urban areas and bourgeoisie people living in houses on rental basis. For the aspirants of homes, housing loan is golden opportunity to bring their dreams into reality. The housing loan is a crucial issue within the country like India where crores of individuals are struggling for shelter. The banks can play a supporting role during this connection by providing housing loan to several people that are eligible for it. This project aims at the comparative analysis of home loan policies and their implementation by various banks.*

Keywords: **Loan, Shelter, House, Machine Learning.**

1. INTRODUCTION

Loans are the core business of banks; home loans are one among them. Most of the profit comes directly from the loan's interest. The loan companies provide the loan after an intensive process of verification and validation. However, they still don't have assurance if the applicant is in a position to repay the loan with no difficulties. We are building a predictive model to predict if an applicant during a position is able to repay the lending company or not. The main objective of this project is to predict whether assigning the loan to a specific person is going to be safe or not. We have implemented this loan prediction problem using Decision tree algorithm and data cleaning in Python as there are missing values within the dataset.

2. RELATED WORK

Uzair Aslam, Asim Sohail and Hafiz Tariq Aziz et al. [1] stated that the there have been a multitude of techniques which were used to assign credit scores and much research has been done on the topic throughout the years. Unlike before, where the initial models depended on professional opinions for assessing the loan worthiness of an individual, recently focus has shifted towards applying advanced machine learning algorithms and neural networks for credit scoring and risk assessment. These techniques can be further classified into two categories: the traditional statistical techniques and advanced statistical techniques. And the proposed system stated that most of the studies tuned and optimized their models to attain maximum accuracy for predicting credit scores of borrowers, however only a small number of the reviewed studies discussed the repercussions of misclassifications of false negatives. And the proposed idea of the paper is focused more towards false negatives when approaching the loan prediction problem and we know that Random Forest algorithm gives more efficiency than decision tree and logistics regression.

Kumar Arun, Garg Ishan and Kaur Sumeet et al. [2] focused on the main objective of this paper which is to predict whether assigning the loan to particular person will be safe or not. Through this system we can predict whether that particular applicant is safe or not and the whole process of validation of features is automated by machine learning technique. The disadvantage of this model is that it

emphasizes different weights to each factor but in real life sometime loan can be approved on the basis of single strong factor only, which is not possible through

this system. The proposed system of this Paper is to provide quick, immediate and easy way to choose the deserving applicants. It can provide special advantages to the bank. The Loan Prediction System can automatically calculate the weight of each features taking part in loan processing and on new test data same features are processed with respect to their associated weight. A time limit can be set for the applicant to check whether his/her loan can be sanctioned or not. Loan Prediction System allows jumping to specific application so that it can be check on priority basis. This Paper is exclusively for the managing authority of Bank/finance company, whole process of prediction is done privately no stakeholders would be able to alter the processing. And the proposed idea is that the data which is provided by the customer is entered in system to check eligibility for loan approval. In this research paper there are various machine learning Algorithms mentioned which can be used We are going to check each and every algorithm in this paper to check which is efficient the purpose of this research paper is to see what algorithms to use in our system from a proper analysis of positive points and constraints on the component, it can be safely concluded that the product is a highly efficient component.

Anchal Garg and Ranpreet Kaur et al. [3] focused on an effective prediction technique that helps the banker to predict the credit risk for customers who have applied for loan. A prototype is described in the paper which can be used by the organizations for making the correct or right decision for approve or reject the request for loan of the customers. The paper uses three different models (SVM Model, Random Forest Network and Tree Model for Genetic Algorithm) and the Ensemble Model, which combines these three models and analyses the credit risk for optimum results. A neural network with standard topology and a feed-forward neural network with ad hoc connections. Neural network can be used for prediction model. This paper shows that the above two models give optimum results with less error. In the proposed work, eleven machine learning models are constructed which have nine properties that are used to predict the credit risk of customers who have applied for loan. Under different training algorithms, this paper presented an ensemble models for loan predications by using several parameters like Accuracy, Gini, Auc, Roc etc. to do the comparison. The main purpose of this paper is to test the accuracy of models and develop a new model called ensemble model that combine the outputs of the three different models to predict the loan of costumers. Real Coded Genetic Algorithms is used to calculate the feature importance.

Zoran Ereiz et al. [4] focused on the paper which shows credit risk is one of the most important risks to be managed by a financial institution. This case study mainly focuses on credit score and its importance while working on loans, and how it plays a major role while procedure. The management of credit risk of credit portfolios is therefore one the most important tasks for the financial liquidity and stability of banking sector in connection with increased sensitivity of banks to the credit risks. Credit risk management is a matter of predicting the possibility of a default loan. Different financial institutions manage credit risks in different ways. In this paper according to the input data BigML returned 33 candidate models out of which 3 were short-listed: a 64-model 465-node decision forest, A neural network based on 128 evaluated networks, and a logistic regression model. All three models were evaluated using the test data. Both the decision forest and logistic regression

models failed to identify ~5.35% of default loans. Neural network model failed to identify 2.63% of default loans. The ROC curve for the neural network model is 0.7690. At the same time the recall for a default loan using this model is 52.55% with a precision of 15.52% and an accuracy of 82.1% that means this paper concludes that machine learning model (Neural Network) is precise in 82 /100 test. Machine Learning helps to distinguish borrowers who repay loans promptly from those who don't. The significant role of AI and machine learning already play a major role in banking and lending sectors and identify challenges and practical issues such as availability of suitable data, availability of skilled staff etc.

3. PROPOSED SYSTEM

We will be creating a website in that customer first register himself then login by same credential. Enter all details regarding loan eligibility criteria this inputs we used as testing data in the Eligibility model. To build our model we will be taking a dataset from Kaggle and after applying machine learning technique, we predict our result. Hence the more accurate the process is it will give more profit as well as improve future growth of a company.

Below is the proposed system diagram for our website:-

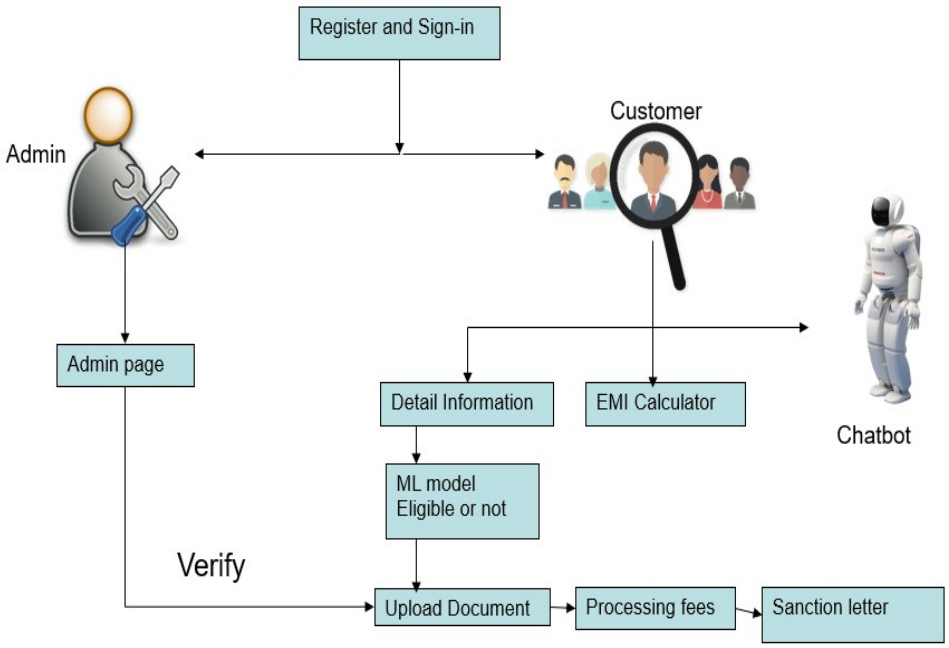


Figure 1. Proposed System

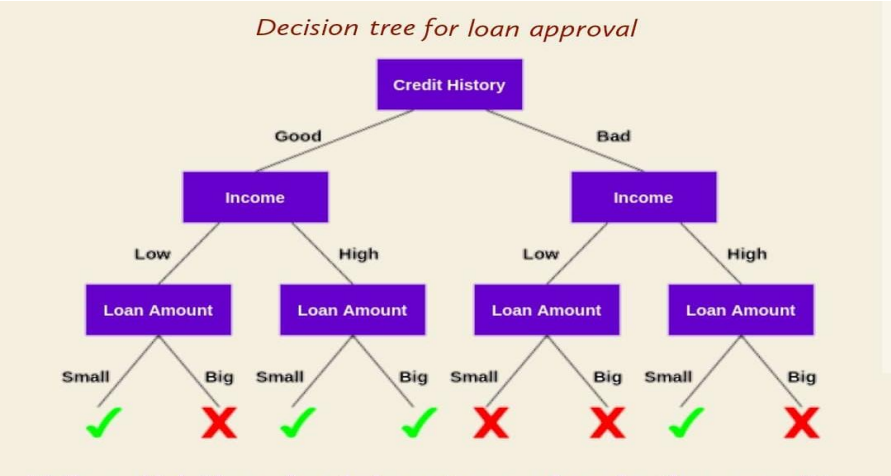
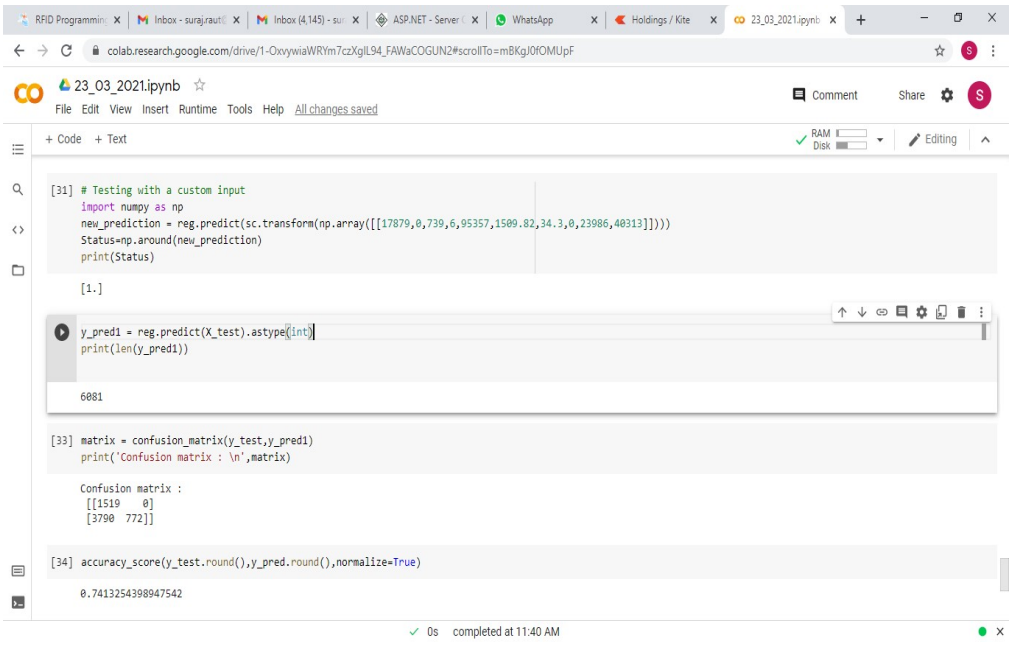


Figure 2. System Architecture

To reduce the manual work in the finance companies. We will be going to create the website which addresses all the major things like customer eligibility using ML, chatbot for customer queries, Processing fees using Instamojo payment gateway and for basic understanding of how much is the loan amount and for that how much interest is applied for how many years accordingly installments are generated. We take documents online itself and admin will verify those documents. If the documents are valid then only further process is carried otherwise customer need to upload it again. Finally, loan sanction letter is generated in soft copy format.

4. RESULT

In this approach, Python programming language is used to implement machine learning algorithms. For training 90 percent data is used and 10 percent data is used for testing. On the basis of this train data set, system analyze rest of 10 percent data and predict the results in term of loan status either accepted or rejected. After completion of the loan form, the user get access to the payment gateway where he has to pay certain amount for as processing fee.



5. CONCLUSION AND FUTURE WORK

In this project, we have implemented a chatbot, payment gateway, and simple website. The data will be saved locally for now but afterward, we are implementing it in the cloud. We will be implementing a mobile application for the same website.

REFERENCES

List and number all bibliographical references that has important contribution on the paper, (if possible, limit to 30, which only are necessary citations are recommended). 9-point Times New Roman, fully justified, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [1]. Do not abbreviate the months. Don't forget to put period (.) at the end of each reference. (See examples below)

6.1. Journal Article

[1] *Uzair Aslam, Asim Sohail and Hafiz Tariq Aziz.* ***(2019)****, "An Empirical Study on Loan Default Prediction Models", "Journal of Computational and Theoretical Nanoscience".*

[2] *Kumar Arun, Garg Ishan and Kaur Sumeet.* ***(2016)****, "Loan Approval Prediction based on Machine Learning Approach", "IOSR Journal of Computer Engineering".*

[3] *Anchal Garg and Ranpreet Kaur.* ***(2016)****, "Loan Prediction Using Ensemble Technique", "International Journal of Advanced Research in Computer and Communication Engineering".*

6.2. Conference Proceedings

[4] *Zoran Ereiz.* ***(2019)****, "Predicting Default Loan Using Machine Learning (OptiML)" , "27th Telecommunications Forum TELFOR 2019".*