A Comparative Study of Machine Learning Algorithms for Predicting Loan Default and Eligibility

Supreeth S Athreyas

Department of Computer Science and Engineering, Jyothy Institute of Technology, Bengaluru, India, supreethathreyas7@gmail.com

Thanmai B K

Department of Computer Science and Engineering, Jyothy Institute of Technology, Bengaluru, India, thanmaibk@gmail.com

Varshini Kashyap S

Department of Computer Science and Engineering, Jyothy Institute of Technology, Bengaluru, India, varshinikashyap.s@gmail.com

Sukesh S Bairy

Department of Computer Science and Engineering, Jyothy Institute of Technology, Bengaluru, India, sukeshbairy@gmail.com

Harish K

Department of Computer Science and Engineering, Jyothy Institute of Technology, Bengaluru, India, harish.k@jyothyit.ac.in

Abstract: Loosening up credits to corporates and individuals for the smooth working of creating economies as is INDIA unpreventable. As a growing number of customers apply for credits in the banks and non-banking financial companies (NBFC), it is really pursuing banks and NBFCs with confined financing to contraption a standard objective and safe framework to credit money to its borrowers for their financial necessities. Also, starting late NBFC inventories have persevered through a basic ruin similarly as the stock expense. It has added to a sickness that has also spread to other money related stocks, horribly impacting the benchmark of late. In this paper, an endeavour is made to assemble the risk related to picking the sensible person who could repay the credit on time thus keeping the bank's non-performing assets (NPA) on hold. This is refined by dealing with the records of the customer who secured credits from the bank into a readied man-made intelligence model which could yield a definite result. The prime point of convergence of the paper is to choose whether or not it will be ensured to allocate the loan to a particular person.

Keywords: Loan; Credit; Non-Banking Financial Company (NBFC); Eligibility; Default

I. INTRODUCTION

Cash raising and advancing for land, purchaser, home advance, and associations' advances are the central bit of almost every bank's strategy. Advancing money to wrong customers shape the huge wellspring of credit peril. A huge segment of the bank's assets comes clearly from the advantage got from the bank's credits. The monetary associations' face, at any rate, twofold test to perceive the possible cognizant defaulters from the up-and-comers and the uneven nature of few bank labourers who have been at the induction of architects of defaulting associations for quite a while. The basic goal of the monetary organization is to safely contribute their capital. In the current circumstance, various NBFCs and banks underwrite credits after an undeniable check and approval measure, regardless, it remains uncertain whether the contender picked is the exemplary right of the generally large number of candidates.

Artificial Intelligence AI is an emerging technology nowadays. The application of AI solves many problems in the real world. Machine Learning is an AI technique that is very useful in prediction systems. Through this strategy, we can predict whether particular candidates are secure and the computer-based intelligence methodology automates the entire pattern of approval. We consider it an essential model of machine learning. It makes a model from training data. While making the forecast the model which is created via a training algorithm (which is Machine Learning) is utilized. The machine learning algorithm prepared the framework utilizing a small amount of the information accessible and tested the excess information.

The AI methods can be applied on an example test information first and afterward can be utilized in making expectation related choices. This paper applied the AI approaches in tackling the loan endorsement issue of the banking area. This paper tries to guarantee that the meriting clients can be immediately chosen effortlessly which offers different advantages to the bank itself. This strategy will gauge the weight consequently of every standard that takes an interest in the advance handling and cycle the equivalent concerning the related load of the new test information. The borrower can set a time to check whether the credit is authorized. This framework can avoid the consecutive check measure and could lead to a particular highlight be checked based on need.

II. LITERATURE SURVEY

Much work has been done by researchers in the field of Loan Eligibility Prediction, depending on their research area and field of interest. The related work in the field of banking sector as well as machine learning, shows that researchers have proposed and implemented various machine learning methods, but the comparative study among these methods is still lacking for as far as Eligibility Prediction is concerned. The work done hence proves to be a novel approach while considering the results and finding various machine learning algorithms for Eligibility Prediction.

An Exploratory Data Analysis for Loan Prediction based on the nature of the clients was proposed by X.Francis Jency, V.P.Sumathi and Janani Shiva Sri [1]. This paper's main purpose is to classify and analyse the nature of applicants for loans. Depending on certain factors, this paper classifies the customers. Classification is performed using analyses of exploratory data which is a technique for analysing data sets that summarizes the main features with visual methods.

A Survey on Ensemble model for Loan Prediction was proposed by Anchal Goyal and Ranpreet Kaur [2] which includes techniques of bagging, boosting, adaboost and random forest. In this paper a technique is proposed in which at least two classifiers are consolidated to make a group model for better forecast. Through an ensemble model which analyzes the different models and chooses the best model for our data which lets the organization settle on the correct decision for the client's loan application.

Loan Prediction Analysis using a Decision Tree was proposed by Nikhil Madane and Siddharth Nanda [3]. It emphasizes on Decision trees and Ensemble methods. In this paper, they have reviewed credit scoring of mortgage loans and the criteria that cause an applicant to be rejected. Implementation of a loan prediction system that helps companies make the right decision to approve or reject customers ' loan requests.

Prediction of Loan Approval using Machine Learning was proposed by Rajiv Kumar, Vinod Jain, PremSagar Sharma, Shashank Awasthi and Gopal Jha [4]. It incorporates Logistic Regression, Decision Trees and Random Forest Algorithms. The results have shown that the prediction accuracy is 93.04%, 95% and 92.53% for LR, DT and RF algorithms respectively. Among the three, accuracy of the DT algorithm is best for prediction of loans.

Exploring the Machine Learning Algorithm for Prediction the Loan Sanctioning Process was proposed by E. Chandra Blessie and R. Rekha [5]. It focuses on Logistic Regression, SVM, Decision Tree, and Naïve Bayes classifiers to make the loan prediction. It shows that the Naïve Bayes model is extremely efficient and gives a better result when compared to other models. It fulfills all requirements of bankers and can be connected to many other systems.

A Study on Predicting Loan Default based on the Random Forest Algorithm was projected by Lin Zhua, Dafeng Quia, Daji Ergua, Cai Yinga and Kuiyi Liub [6] . In this paper, the random forest algorithm is adopted to create a model for predicting loan and the results are compared with other three algorithms of logistic regression, decision tree and support vector machines. It proves that the random forest algorithm performs outstanding than the other three algorithms in the prediction of loan default and incorporates a strong ability of generalization.

Prediction of Loan Status in Commercial Bank using Machine Learning Classifier proposed by G. Arutjothi and C. Senthamarai, [7] have used Min-Max normalization and KNN classifiers for prediction of loan approval status in banks. The machine learning based prediction system was applied on commercial banks. The paper concludes that the machine learning approach is very useful in loan status prediction.

Accurate Loan Approval Prediction based on ML Approach was proposed by J. Tejaswini, T. Mohana Kavya, R. Devi Naga Ramya, P. Sai Triveni and Venkata Rao Maddumala [8]. It uses Logistic Regression, Decision Tree and Random Forest for predicting the loan approval. It is highly Efficient Component and Compatible to other systems.

A Novel Optimized Classifier For the Loan Repayment Capability Prediction System was proposed by Soni P M and Varghese Paul [9]. Random Forest Algorithm , R programming Language and Weka Tool have been used. Two different models before optimization and after optimization were generated in R and found that the tuning of parameters leads to better accuracy. The comparison of models in weka and R proved that weka produces considerably good results.

Credit Risk Prediction Based on Machine Learning Methods was proposed by Yu Li [10]. It uses Logistic Regression and XGBoost Algorithm for predicting loan eligibility. By comparing the model discrimination, model interpretability and model stability of logistic regression model and XGBoost model, it can be inferred that the model discrimination and model stability of XGBoost model are significantly higher than that of logistic regression model, which can effectively improve the identification ability of personal fast credit risk.

III. EXISTING SOLUTION

Singular advances can be a brisk strategy to get a genuinely essential cash blend, which can be used to hold your home rebuild, emergency clinical charges, start an endeavor, or even take an event. In any case, getting an individual credit from a bank incorporates the essential development of a customer check, which determines the end result of your development application. While applying for credit, the borrower requirements to give a letter to the bank, affirming them to run your record as a customer. Banks will survey your repayment history with others and the proportion of commitment you have as of now. The bank by then overviews your pay and finds out

your commitment organization incorporation extent. The path toward applying for a credit is repetitive moreover, monotonous. It involves the following steps usually:

1. Applicant fills a loan application form.

2. Hands it over to the bank.

3. Bank checks the Credit score and report of the concerned applicant.

4. Low credit score leads to loan rejection.

5. High credit score leads to eligibility check basis documentation.

6. Non eligibility leads to rejection.

The most critical aspect of the whole process is that the applicant must have a significant credit history. But there are cases where the applicant is a first-time borrower who does not have a credit history. During such cases, there is a high possibility that he/she will be denied the loan.

IV. CONCLUSION

In this paper, we did a comparative study of Machine Learning models used to predict the loan eligibility and approval.   We used 3 algorithms namely Logistic Regression, Decision tree and Random Forest on a sample dataset to predict the outcomes. Among the three the accuracy of Random Forest was found to be promising. In Future the Random Forest algorithm can be applied on other datasets available for loan approval to further investigate its accuracy.

This brings some of the following insights about credit advance approval:

1. Loan applications that don't pass rules are generally not affirmed, likely since they are more likely to not payback.

2. More often than not, low-income candidates are bound to get loan endorsement, which makes sense and those applicants are bound to repay their advances.

REFERENCES

[1] Yu Li (2019). Credit Risk Prediction Based on Machine Learning Methods. The 14th International Conference on Computer Science & Education (ICCSE 2019)

[2] Soni P M, Varghese Paul (2019). A Novel Optimized Classifier For the Loan Repayment Capability Prediction System. Third International Conference on Computing Methodologies and Communication (ICCMC 2019)J. Tejaswini ,T. Mohana Kavya , R. Devi Naga Ramya , P. Sai Triveni Venkata Rao Maddumala (2020). Accurate Loan Approval Prediction based on ML Approach. Journal Of Engineering Sciences 2020

[3] G. Arutjothi, Dr. C. Senthamarai (2017). Prediction of Loan Status in Commercial Bank using Machine Learning Classifier. Proceedings of the International Conference on Intelligent Sustainable Systems (ICISS 2017)

[4] Nikhil Madane and Siddharth Nanda (2019). Loan Prediction Analysis using a Decision Tree. Journal of the Gujarat Research SocietyRajiv Kumar , Vinod Jain , PremSagar Sharma , Shashank Awasthi , Gopal Jha (2019). Prediction of Loan Approval using

Machine Learning. International Journal of Advanced Science and Technology

[5] E.Chandra Blessie, R. Rekha (2019). Exploring the Machine Learning Algorithm for Prediction of the Loan Sanctioning Process. International Journal of Innovative Technology and Exploring Engineering (IJITEE)

[6] X.Francis Jency, V.P.Sumathi, Janani Shiva Sri (2018). An Exploratory Data Analysis for Loan Prediction Based on Nature of the Clients. International Journal of Recent Technology and Engineering (IJRTE)

[7] Lin Zhua, Dafeng Qiua, Daji Ergua, Cai Yinga, Kuiyi Liu (2019). A study on predicting loan default based on the random forest algorithm. 7th International Conference on Information Technology and Quantitative Management (ITQM 2019)

[8] Anchal Goyal, Ranpreet Kaur (2016). A survey on Ensemble Model for Loan Prediction. International Journal of Engineering Trends and Applications (IJETA)

[9] Zoran Ereiz (2019). Predicting Default Loans using Machine Learning (OptiML). 27th Telecommunications Forum TELFOR 2019

[10] Ya-Qi Chen, Jianjun Zhang, Wing W. Y. NG (2018). Loan Default Prediction Using Diversified Sensitivity Undersampling. Proceedings of the 2018 International Conference on Machine Learning and Cybernetics, Chengdu, China 2018