



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 4, April 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Survey on Bank Loan Prediction System using Machine Learning

Guide: Prof. R.L. Ghule, Gaikwad Diksha Dilip, Kolekar Gauri Dnyanoba, Pawar Ashwini Dilip

Department of Computer Engineering, S.B. Patil College of Engineering, Indapur, India

ABSTRACT—The quality of life for humans has improved thanks to technology. We intend to provide something fresh and unique every day. Every issue we face has a solution, and we have machines that support our way of life and help us feel relatively complete in the banking industry. Before the loan amount is approved, the sector contender receives proofs and/or backup. The system's historical data about the candidate determines whether or not the application is granted. Numerous people ask for loans every day in the banking industry, yet the bank's resources are constrained. In this situation, a classes-function algorithm would be quite helpful in making the correct forecast. Examples include classifiers for support vector machines, logistic regression, and random forests. The number of loans, or whether the client or customer is repaying the loan, determines the bank's profit or loss. For the banking industry, loan recovery is of utmost importance. In the banking industry, the improvement process is crucial. A machine learning model was developed utilizing various categorization techniques using the historical data of the candidates. Using machine learning models trained on the historical data set, the primary goal of this work is to predict whether a new applicant will be approved for a loan or not.

KEYWORDS—: Loan Dataset, Random Forest, Django.

I. INTRODUCTION

People prefer to apply for loans online because of the daily data growth brought on by the banking industry's digitization. As a typical tool for information analysis, artificial intelligence (AI) is receiving more and more attention. People from diverse industries are using AI calculations to solve problems based on their knowledge of the relevant industries. When approving loans, banks are having a lot of trouble[1]. The bank workers oversee a large number of applications every day, which is difficult and increases the likelihood of mistakes. The majority of banks make money through loans, however it can be dangerous to select eligible clients from the pool of applicants. A bank could suffer a severe loss due to one error[2].

Almost all banks' main line of operation is loan distribution. Out of all applicants, one will receive a loan as a result of this project. A quick, impartial procedure that saves the bank time involves prioritising each applicant for review. The timely completion of all other client formalities by the bank authorities benefits the customers. The nicest thing is that it works well for both applicants and banks.

With the help of this mechanism, specific applications might be granted priority consideration[3].

II. RELATED WORK

As per paper[1] -The quality of life for humans has improved thanks to technology. We intend to provide something fresh and unique every day. In the banking industry, the candidate receives proofs/backup prior to acceptance of the loan amount. We have machines to support our life and make us somewhat complete. The system's historical data about the candidate determines whether or not the application is granted. Numerous people ask for loans every day in the banking industry, yet the bank's resources are constrained. A classes-function algorithm would be very helpful in this situation if the proper prediction could be made.

As per paper[2]- Kaggle data is gathered for analysis and forecasting. The various performance measures have been estimated using logistic regression models. On the basis of performance metrics like sensitivity and specificity, the models are compared. The final results have demonstrated that the model generates various results. The model is only



slightly superior because it incorporates variables (personal characteristics of the customer, such as age, purpose, credit history, credit amount, and credit duration) in addition to checking account information, which reveals the customer's wealth, that should be considered in order to accurately calculate the probability of loan default.

As per paper[3]- This paper builds a prototype model that the organisation might use to decide whether to approve or refuse the loan. In order to increase the accuracy of defaulters in banks based on probability, writers proposed a clustering approach. Utilizing R's implementation of the KNN method, the experimental findings were achieved. Authors in certain states made a problem statement regarding the issue of class inequality. Different solutions to the class imbalance issue are discussed.

As per paper[4]-Information Gain, Particle Swarm Optimization (PSO), and Genetic Algorithm are all discussed in this study. All machine learning models using raw data and subsequently all models employing each feature selection approach had their outcomes compared. The findings reveal that the scores significantly increased after feature selection methods were used. The Decision tree (Pruned) method turns out to get the best outcomes.

As per paper[5]-In this paper, create a decision tree-based high performance predictive model for loan acceptance prediction. Experiments were conducted using a variety of tree approaches, from the simplest and most understandable decision tree to the most intricate random forests. Due to the highlight coupled and complex feature space, which resulted in an overly simplified tree that was problematic to use, the results showed insufficient performance with regard to simplified decision trees.

As per paper[6]-To assist the provident fund centre in better managing asset liquidity, we have also established the early loan payback class. We put out a model based on the stacking ensemble method and compared it to six different models, including the Cart decision tree, Random Forest, AdaBoost, XGBoost, CatBoost, and LightGBM. Following experiments, the accuracy, recall, F1, AUC, and kappa performance evaluation metrics for our stacking model show a significant improvement.

As per paper [7]-In this paper, Due to a number of issues, the lender is unable to repay the loan on time; this creates issues with the delayed cash flow and makes managing the liquidity of assets more challenging. In order to help lenders manage the risks of early loan payback and loan overdue, the provident fund centre must assess the user's loan repayment behaviour. According to our investigation, the majority of China's provident fund centres are unable to forecast users' loan repayment behaviour using data mining and machine learning on user data.

As per paper[8]-In this paper, several algorithms were put into practise to forecast loan defaulters. Logistic Regression, Random Forest, Gradient Boosting, and CatBoost Classifier were used to achieve the best results. Compared to Logistic Regression, the Gradient Boosting procedure produces outcomes that are superior to or equal to those. Multiple models are used in the procedure known as gradient boosting. Given that the variables' interpretations are ambiguous, it is uncommon to throw out a variable. On the other hand, even if doing so reduces the overall model accuracy and predictive ability, it is common practise to exclude variables while fitting logistic regression.

As per paper[9]-The accuracy of conventional user loan risk prediction models, including KNN, Bayesian, and DNN, does not gain from the expansion of data in the context of bank loans. Based on the research of Overdue Prediction of Bank Loans Based on Deep Neural Network, this article was created. To address the existing prediction issues, we also suggest using the LSTM algorithm to study users' dynamic behaviour and the SVM algorithm to examine users' static data. In order to determine whether users are in default, this article analyses user demographic data, bank records, browsing habits, credit card billing histories, and loan repayment data. The fundamental input for SVM is these static data.

As per paper[10]-In this paper, We suggest an analysis approach for credit data based on machine learning classifiers. We employ the Min-Max normalisation and K-Nearest Neighbor (K-NN) classifier in combination. The software package R tool is used to carry out the goal. This suggested model delivers crucial data with the maximum level of precision. In commercial banks, it is used to forecast the status of loans using a machine learning classifier.



|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204197 |

III. EXISTING SYSTEM/OPEN ISSUES

When approving loans, banks are having a lot of trouble. The bank workers oversee a large number of applications every day, which is difficult and increases the likelihood of mistakes. The majority of banks make money through loans, however it can be dangerous to select eligible clients from the pool of applicants.

It places emphasis on giving each component a varied weight, however in reality loans might occasionally be issued solely on the basis of one powerful factor, making this technique ineffective. Both applicants and bank employees benefit greatly from loan prediction.

IV. CONCLUSION

In order to decrease human interference and boost productivity, the rapidly expanding IT sector of today needs to develop new technology and upgrade existing technology. Anyone looking to apply for a loan or the banking system will use this model. It will be very beneficial for managing banks. It is abundantly obvious from the data analysis that it lessens all fraud committed at the time of loan acceptance. Everyone values their time highly, thus by doing this, not only the bank but also the applicant's wait time will be shortened.

REFERENCES

- [1] Vishal Singh ,AyushmanYadav, RajatAwasthi, N.Partheeban“Prediction of Modernized Loan Approval System Based on Machine Learning Approach”(IEEE-2021)
- [2] Mohammad Ahmad Sheikh, Amit Kumar Goel, Tapas Kumar,An Approach for Prediction of Loan Approval using Machine Learning Algorithm,(ICESC 2020)
- [3] P. Maheswari, CH. V. Narayana, Predictions of Loan Defaulter - A Data Science Perspective
- [4] SujoyBarua, DivyaGavandi, PoojaSangle, LeenaShinde, JyotiRamteke ,Swindle: Predicting the Probability of Loan Defaults using CatBoost Algorithm
- [5] Mohamed Alaradi ,SawsanHilal, Tree-Based Methods for Loan Approval(IEEE-2020)
- [6] LilingKe, Chenhao Li, Tao Zang, Zhengongcai, Jin Wen, Renju Wang, FengQian, Hong Tang,“Loan Repayment Behavior Prediction of Provident Fund Users Using a StackingBased Model” 1, IEEE-2021
- [7] Lili Lai, “Loan Default Prediction with Machine Learning TechniqueCCNS-2020
- [8] Bhoomi Patel, HarshalPatil, JovitaHemram, Shree Jaswal, Loan Default Forecasting using Data Mining,INCETJun 5-7, 2020,
- [9] Xin Li, Xianzhong Long, Guozi Sun, Geng Yang, HuakangLi“Overdue Prediction of Bank Loans Based on LSTMSVM”2018 IEEE
- [10]G. Arutjothi, Dr. C. Senthamarai“Prediction of Loan Status in Commercial Bank using Machine Learning Classifier”, (ICISS 2017) IEEE



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.423



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 **6381 907 438** **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details