



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 3, March 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.379



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

ML Based Loan Approval Prediction System A Novel Approach

Dr. MK Jayanthi Kannan¹, A.R Nithej², Padi Akhil³, Pavan Gowda⁴, Prashant Pareek⁵

¹Professor, School of Engineering and Technology, Jain University Bangalore, India

UG Student, Department of Information Science and Engineering School of Engineering and Technology, Jain University Bangalore, India ^{2,3,4,5}

ABSTRACT: Loan approval system is important in banks in order to reduce the loss and approve loans only for eligible customers, who are able to repay the loan amount. Various studies can see in this area and many studies are still focusing on this problem as give assets on safe hand in very important for any bank. However, the performance of previous studies is good, but the accuracy can be still increased. The main objective of this study is to increase the performance of loan prediction system. This study is focusing on different machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, K nearest neighbors, Artificial neural network, Naive Bayes, Adaboost, and Voting classifier to predict the loan approval.

KEYWORDS: Decision tree, random forest, support vector machine, linear models, neural networks, adaboost, loan approval

I. INTRODUCTION

Loan is very essential for the current world. Various types of platform are available in the world to take loan on low rate, where some are private, and few are govt. platform. Normally, to give loan, banks are focus on the CIBIL (Credit Information Bureau (India) Limited) score, credit score along with background of applicant to give a loan. However, it has been seen in the history that applicant took loan but didn't repay it. So, to save the bank assets from the defaulter become a necessity of the world as directly or indirectly, banks are important in the growth of any country (Arun, Ishan, & Sanmeet, 2016). Various research research has been done in this area to safe the bank as well as the growth of the country. However, the performance of previous studies is good, but the accuracy can be still increased (Vangaveeti, Venna, Kidambi, Marneni, & Maganti, 2020). Many machine learning algorithms has been used to predict the loan and some of them shown very good performance. Such as: logistic regression (LR), decision tree (DT), Artificial neural network (ANN), K-Nearest Neighbour (KNN), Naïve Bayes (NB), Support vector machine (SVM), and few algorithms under Ensemble techniques like Random Forest (RF), Adaboost, Voting classifier (VC). Apart from predicting, few works are completed to do analysis of the background details of the applicant by using P2P and Exploratory Data Analysis (EDA). To build the loan prediction system, dataset always play a very important role. Apart from the dataset, pre-processing, and classification is also important. Various types of features are presented in the dataset, to filter the dataset and convert into proper manner, is a part of pre-processing. In the last, model will use the dataset to train the system where many machine learning algorithms will use to complete the training part. This study is divided into six parts: introduction, literature, dataset, methodology, result and conclusion. Details of each step such as: literature, dataset, methodology, result are discussed properly in the perspective to loan prediction. In the last, the performance of this study is compared with other studies.

II. RELATED WORK

A study was conducted to predict whether a borrower will default on a loan is of significant concern to platforms and investors in online P2P lending (Jiang, Wang, Wang, & Ding, 2017[1]. A two-stage method designed to select an effective feature set containing both soft and hard information along with P2P was used to complete the study. An empirical analysis using realword data from a major P2P lending platform in China shows that the proposed method can improve loan default prediction performance compared with existing methods based only on hard information. This study introduced a topic model to extract valuable features from the descriptive text concerning loans and construct four default prediction models to demonstrate the performance of these features for default prediction.

Another study was also completed to create a credit scoring model for credit data and loan approval status Arutjothi and Senthamarai in the year 2017 [2]. Authors used MinMax Normalization and K-Nearest Neighbour (KNN) to conduct

this research and proposed a system that showed good performance along with useful information. The performance of this paper using logistic regression as a tool, this paper specifically delineates about whether or not loan for a set of records of an applicant will be approved.

Similarly, Vaidya in 2017 [3] and Xiaojun Ma et al. [4] in 2018 used logistic regression, and P2P, Data cleaning, Default rate, LightGBM algorithm, XGboost algorithm, respectively, to predict the loan approval status. The performance of this study was good, and the system was able to predict that applicants will be approved for loan or not. One study focused on various parameters that should be approved or not by Jency, Sumathi, & Sri in 2019[5]. To do analysis, Exploratory Data Analysis (EDA) was used. After the analysis, it can say that short term loans are chosen mostly by the clients. Another analysis was also conducted to find out the relationship between the Italian bank that, within a bank, approves a loan and the subsequent performance of the loan by Calcagnini, Cole, Giombini, & Grandicelli, [6] in 2018. P2P was used to complete this study.

Similarly, in 2020, a study by Tejaswini, Kavya, Ramya, Triveni, & Maddumala [7] in 2020 focussed on the prediction whether the loan in terms of banking investment is in safe hand or not, by using Logistic Regression (LR), Decision Tree (DT), Random Forest (RF) algorithm. As compared to LR and RF, DT has performed well in terms of accuracy. Furthermore in 2020, there was another study done by Vangaveeti et al [8] in 2020, using Logistic Regression algorithm for the prediction to provide a speedy, immediate and simple approach to pick the deserving applicants and the performance was good based on accuracy.

Another study titled ‘Should This Loan be Approved or Denied?’: A Large Dataset with Class Assignment Guidelines’ by Min Li, Amy Mickel, Stanley Taylor [9] in the year 2018 Logistic Regression was used and In this article, a large and rich dataset from the U.S. Small Business Administration (SBA) and an accompanying assignment designed to teach statistics as an investigative process of decision making are presented.

Entropy method of constructing a combined model for improving loan default prediction in 2019 [10] by Yihen Li used Random forest, logistic regression, artificial neural network the experimental results reveal that the proposed combined model outperforms the two base models on four evaluation metrics.

Again, EDA was used to do analysis that the person is getting loan is reliable or not and after that Logistic Regression, Decision Tree, SVM, Naïve Bayes also used to predict the status of the loan by Blessie & Rekha, [11] in 2019. The performance of the Logistic Regression, Decision Tree, SVM, and Naïve Bayes was 78.91%, 71.92%, 65.27%, and 80.42%, respectively. According to this literature, various algorithms of machine learning has been applied to predict the loan approval. Moreover, several algorithms such as supervised, and ensemble algorithm can be also used to increase the performance to loan approval system.

III. PROPOSED ALGORITHM

A. Algorithms used:

- Logistic regression
- Decision Tree
- Random forest classifier
- K-Nearest Neighbor
- Naïve Bayes
- AdaBoost algorithm
- Linear SVM
- Polynomial SVM
- Wavelet SVM

B. Description of the Proposed Algorithm:

Logistic regression is a supervised learning classification to predict the probability of a target variable

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. The decisions or the test are performed on the basis of features of the given dataset.

Random forest classifier creates a set of decision trees from randomly selected subset of training set. It then aggregates the votes from different decision trees to decide the final class of the test object.

K-Nearest Neighbor is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.

Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.

Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.

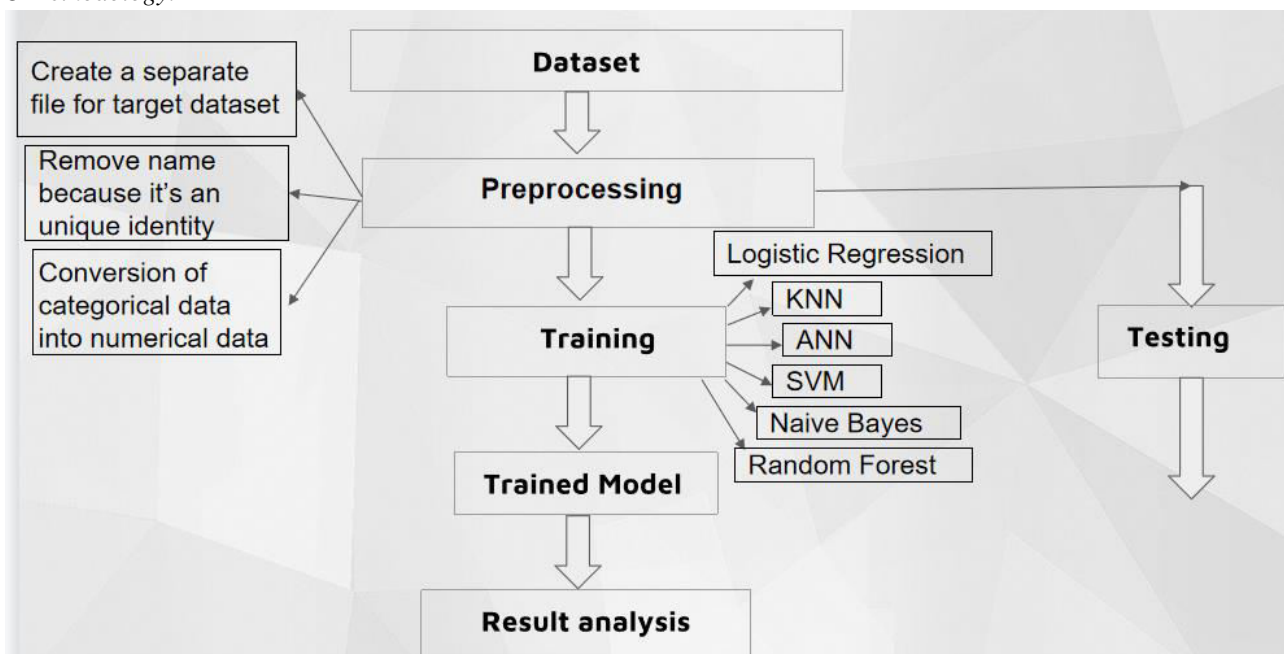
AdaBoost algorithm, short for Adaptive Boosting, is a Boosting technique that is used as an Ensemble Method in Machine Learning. Voting is one of the simplest way of combining the predictions from multiple machine learning algorithms. We can train data set using different algorithms and ensemble then to predict the final output. The final output on a prediction is taken by majority vote accordingly.

Linear SVM, Linear SVM is Super classifier. Generally linear SVM is used in biclassification problems, for example, the problem setting. Where there are two classes coming into consideration

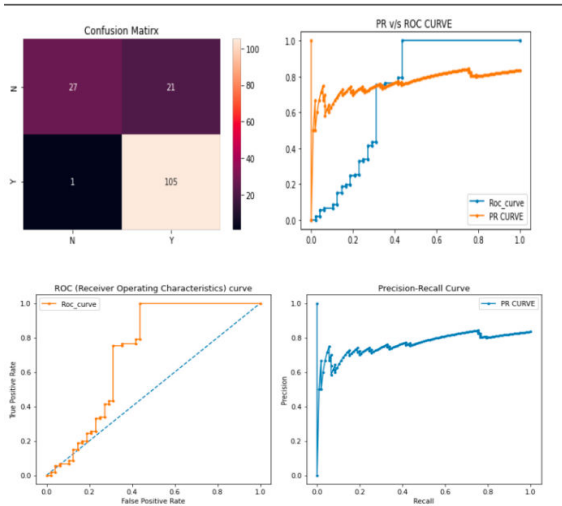
Polynomial SVM, polynomial kernel function is a normal kernel function that is commonly used with SVM in machine learning. It represents the relation of vectors in a feature space over polynomials of actual variables by allowing learning on non linear variable linear models.

Wavelet SVM Wavelet variance measures variability in the form of equivalently or signal by scale, variability trends, in a signal over frequency

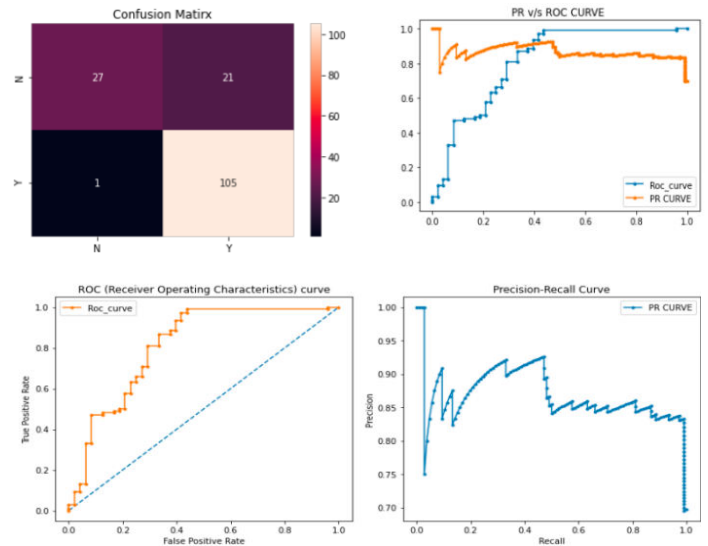
C Methodology:



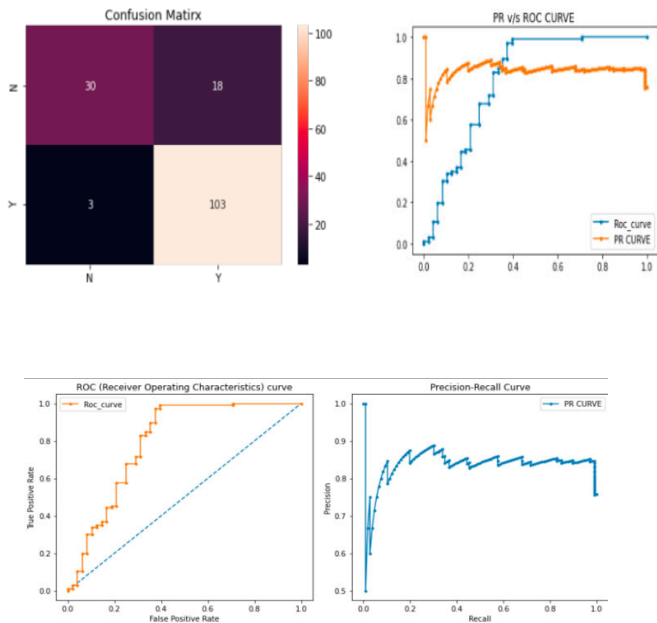
IV. SIMULATION RESULTS



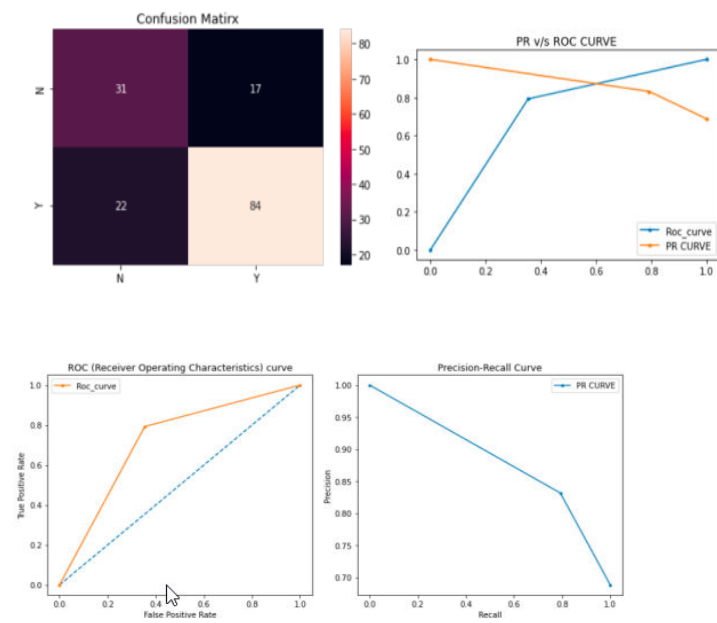
linear SVM



logistic regression



Naive Bayes



Decision Tree



ALGORITHMS	PERFORMANCE OF THE PREVIOUS STUDY (%)	PERFORMANCE OF THE PROPOSED STUDY (%)
Logistic Regression	78.91	86
Decision Tree	71.92	74
SVM (RBF)	65.27	86
Naïve Bayes	80.42	86

In the previous study, Logistic regression, Decision tree, Support Vector Machine (RBF), and Naive Bayes got a performance of 78.91%, 71.92%, 65.27%, and 80.42%, respectively

It can be seen in the comparison of results, the proposed study performed well and provided a higher performance of 86% ,74%,86%,86% by using Logistic regression, Decision tree, Support Vector Machine (RBF), and Naive Bayes.

V. CONCLUSION AND FUTURE WORK

In this paper, we presented a loan approval prediction system that uses machine learning algorithms to predict the likelihood of loan approval based on the borrower's credit history, income, and other relevant factors. We reviewed several related studies on loan approval prediction using machine learning algorithms and presented a use case and flow diagrams for our loan approval prediction system. Our system has the potential to improve the efficiency and accuracy of the loan approval process, reduce the risk of default, and ultimately benefit both the financial institution and the borrower.

REFERENCES

1. XiaojunMa, JinglanSha, DehuaWang, YuanboYu, QianYang, XueqiNiu , ‘Loan default prediction by combining soft information extracted from descriptive text in online peer-to-peer lending’, Annals of Operation research 266, 511-529(2018).
2. G Arutjothi , C.Senthamarai, ‘Prediction of loan status in commercial bank using machine learning classifier’, IEEE conference December 2017, DOI: 10.1109/ISS1.2017.8389442 .
3. Ashlesha Vaidya ‘Predictive and probabilistic approach using logistic regression: Application to prediction of loan approval’ IEEE July 2017, DOI: 10.1109/ICCCNT.2017.8203946.
4. XiaojunMa, JinglanSha, DehuaWang, YuanboYu, QianYang, XueqiNiu , ‘Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning’, Electronic Commerce Research and Application volume 31, september-october 2018, page 24-39.
5. X Francis Jency, V.P. Sumathi ‘An Exploratory Data Analysis for Loan Prediction’, International Journal of recent technology and engineering, Volume-7, issue - 4s, November 2018.
6. Giorgio Calcagnini, Rebel Cole, Germana Giombini & Gloria Grandicelli ‘Hierarchy of bank loan approval and loan performance’ Economia Politica, 35, 935-954, 2018.
7. J.Tejaswini, T. Mohana Kavya, R. Devi Naga Ramya, P. Sai Triveni Venkata Rao Maddumala ‘Accurate loan approval prediction based on machine learning approach’, Journal of engineering sciences, vol -11 issue – 4, 2020.
8. Sai Aparna Vangaveeti, Naga Likitha Venna, Prasanna Naga Sri RamyaYajamanam Kidambi, Harika Marneni, Naga Satish Kumar Maganti ‘Logistic regression based on loan approval prediction’, Journal of Composition Theory, vol – 13, issue – 5, 2020.
9. Min Li, Amy Mickel, Stanley Taylor ‘Should This Loan be Approved or Denied?’: A Large Dataset with Class Assignment Guidelines’, Journal of statistics education, vol- 26, issue – 1, page 55-66 , 2018.
10. Yiheng Li, Weidong Chen ‘Entropy method of constructing a combined model for improving loan default prediction’, Journal of the operation research society, volume 72, issue- 5, 2021.
11. E. Chandra Blessie, R. Rekha ,‘Exploring the Machine Learning Algorithm for prediction of loan sanctioning process’, International Journal of Innovative Technology and Exploring Engineering, , Volume-9 Issue-1 , 2019.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor
Impact Factor: 8.165



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

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



www.ijircce.com

Scan to save the contact details