Global Journal of Modeling and Intelligent Computing (GJMIC)

Vol. 3, No. 2, July, 2023 ISSN: 2767-1917

LOAN ELIGIBILITY PREDICTION USING MACHINE LEARNING: A COMPARATIVE APPROACH

Sonali Kumari, Siksha 'O' Anusandhan University, India, (chaturvediswarna864@gmail.com)
Debasish Swapnesh Kumar Nayak, Siksha 'O' Anusandhan University, India (swapnesh.nayak@gmail.com)
Tripti Swarnkar, Siksha 'O' Anusandhan University, India (triptiswarnkar@soa.ac.in)

ABSTRACT

Machine learning (ML) algorithms can bring revolution in the research field in almost all areas. Processes in numerous industries, including finance, real estate, security, and genomics, are being transformed by machine learning (ML) algorithms. One of the major impediments in the banking sector is the loan approval process. Modern tools like ML models help accelerate, streamline, and increase the precision of loan approval procedures. It will benefit both the client and the bank in terms of time and manpower required for loan eligibility prediction. The entire work is cantered on a classification problem and is a form of supervised learning in which it is important to determine whether the loan will be approved or not. Also, it is a predictive modelling problem where a class label is predicted from the input data for a specific sample of input data. In this work, we deployed various ML algorithms to identify the loan approval status and compare the performance of implemented models. The implemented models will attempt to predict our target column on the test dataset using information from the loan eligibility prediction dataset obtained from Kaggle, which includes features like loan amount, number of dependents, and education. The parameters like accuracy, confusion matrix, ROC curve, and precision are measured for specific models whose performance is significant.

Keywords: Machine Learning (ML), Supervised Learning, Loan Eligibility Prediction (LEP), Kaggle, Real estate.

1. INTRODUCTION

The banking sector is a crucial component of any economy. Banks serve as intermediaries between savers and borrowers and are responsible for providing financial services to individuals, businesses, and governments. One of the most important services provided by banks is loans. Loans are essential for the growth of the economy. They allow individuals and businesses to invest in new ventures, buy homes and cars, and make other purchases that they would not be able to afford without borrowing. Loans also help businesses to expand their operations, hire more employees, and ultimately contribute to the growth of the economy. In addition to providing access to credit, banks also play a critical role in managing risk. Banks carefully evaluate loan applications to determine the creditworthiness of the borrower and assess the risks associated with the loan. This helps to ensure that loans are made to borrowers who are likely to repay them, minimizing the risk of default. Loan policies are designed by banks and loans are sanctioned based on applicant status as per policies. Banks create lending policies, and loans are approved depending on an applicant's standing under such policies. Banks often only authorize loans following a thorough assessment of the applicant's situation, either through meticulously checking submitted copies or through direct checking of the applicant's assets. Yet, there is no assurance that the applicant chosen from all of the applicants is the best candidate or not. Many writers used various data mining techniques to automate the loan verification and confirmation procedure. The applicant's status can be predicted using machine learning by processing all of the applicant's attributes.

Machine learning can help automate loan eligibility prediction by analyzing vast amounts of data and identifying patterns and trends that may not be easily detected by humans. One of the main benefits of machine learning is that it can learn from past data and use that knowledge to predict future outcomes. By training machine learning algorithms on historical loan data and outcomes, banks can develop models that can accurately predict loan eligibility based on various factors such as credit scores, income, employment history, and other relevant factors. The machine learning models can also help identify critical risk factors that can be used to refine the loan application process by identifying high-risk applications and prompting the human review. By making loan decisions more quickly and accurately, eliminating the need for manual review, and making personalized loan recommendations,

machine learning can also enhance the overall client experience. Overall, machine learning can enhance the speed and accuracy of loan processing while also enhancing the client experience.

In this paper, we examined the performance of various machine learning algorithms, including Knearest neighbor, Logistic Regression, Linear Regression, Decision Tree, Naïve Bayes, SVM, and Random Forest. The analysis of these algorithms leads to the suggestion of an original solution in this area. The deployment of various models and the model architecture is shown in Figure 1.

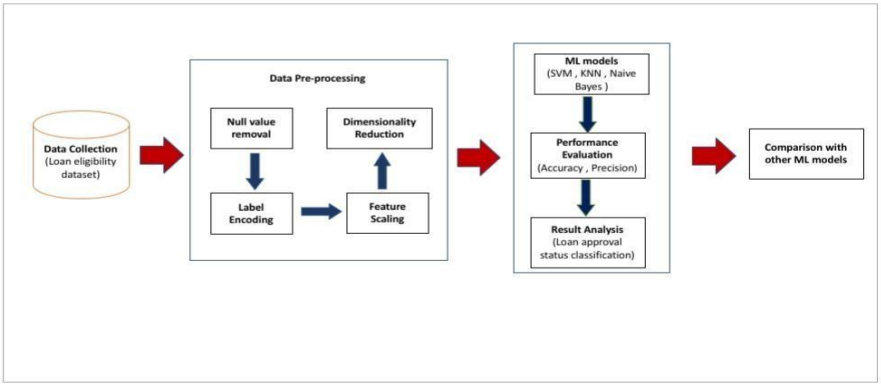

Figure 1: Proposed pipeline of the Model

2. RELATED WORK

Kumar et al. (2022) discussed the traditional loan approval process and how they are unsafe for the bank. They discussed data mining and machine learning approaches to automate the loan approval process so that it will save time and resources. This project aims to replace the current loan approval process in the banking industry.
The authors in [2], compared six different machine learning algorithms based on precision, recall, and accuracy. The objective of this work is to predict if a loan will be approved or not. Random Forest displayed the most accuracy in their study, at 95.55%, whereas Logistic Regression displayed the lowest accuracy.
The authors in [3], discussed how feature selection is important in a predictive model. They used several methods and approaches for the same. They also suggested that a linear neural network can be used instead of a linear regression model to take advantage of the expressive power of a neural network.
The authors in [4], discussed that the cascade of Deep Learning network and Support vector machine improves accuracy by 9%. In this study, the authors employed DNN to convert low-dimensional input data into high-dimensional output features, which were later used to train an SVM credit risk classification model.

3. METHODOLOGY

Loan eligibility prediction systems in the banking sector can be beneficial for both customers and the bank. Machine learning approaches can ease the whole process as they will be cost-effective, and efficient and will save time. In this work, important characteristics that are necessary for predicting loan eligibility are discussed. The dataset used in this study was gathered from a public repository. Training and testing datasets are created after preprocessing the data to increase its overall quality. Machine learning models are developed using training data, while test data are used to assess the model's performance. Decision trees, random forests, support vector machines, K-nearest neighbors, Naïve Bayes, logistic regression, and linear regression are all used, and their effectiveness in predicting loan eligibility is evaluated [5, 8, 9].

4. IMPLEMENTATION

Jupyter notebook is a potential implementation option. It is interactive and features a user-friendly environment for creating machine learning models.

I. Dataset

The dataset for predicting loan eligibility is gathered from the Kaggle public repository. Thirteen (13) characteristics and 614 records make up the dataset. Loan id, Gender, married, dependents, education, self-employment, applicant income, co-applicant income, loan amount, loan term, credit history, property area, and loan status are some of the features which are shown in Figure 2.

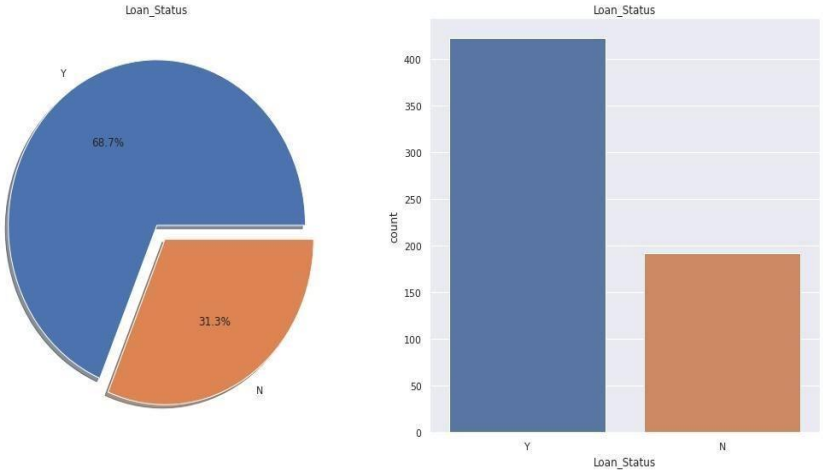

Figure 2: Data count of Loan status divided into two classes Yes (Y) and No (N)

II. 4.2 Data Pre-processing

For the models to perform well, the dataset needs to be preprocessed. In this work, we removed all the null values from the dataset and filled all the missing values using the bfill() and fill() methods. In the next step, we used label encoding along with feature scaling using standard scalar and in the last step we used Linear Discriminant Analysis (LDA) for dimensionality reduction [6-8].

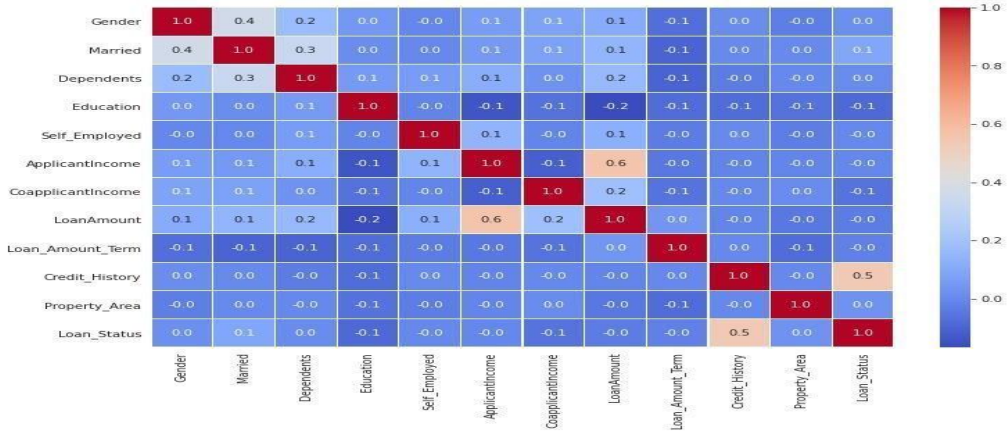

Figure 3: The correlation among all the attributes in the dataset

III. Machine Learning Models

We employed seven machine learning methods in this study: Linear Regression, Logistic Regression, Support Vector Machine, Naïve Bayes, Decision Tree, K-nearest Neighbor, and Random Forest.

5. RESULT

In the banking sector, the loan has always been the greatest source of income for a bank but the loan approval process consumes a lot of time if done manually. To automate this whole process we used machine learning models. Based on the accuracy, we evaluated the performance of seven various machine learning models. Linear regression has the lowest accuracy of just 51%, while the Random Forest classifier outscored all six other models, scoring the greatest accuracy of 90.71%. Table 1 summarizes the overall machine learning models' average accuracy.

Table1: Accuracy of all the implemented models

Classifier	**Accuracy**
Linear Regression	51.21%
KNN	54.09%
Logistic regression	74.83%
Decision Tree	78.62%
SVM	82.23%
Naïve Bayes	85.96%
Random Forest	***90.71%***

6. FUTURE WORK AND CONCLUSION

This work concluded that the dataset is incomplete and still lacks some feature vectors. No classifier was able to perform better than 90.71% since the subspace of the input space that we were trying to generalize has unknown extra dimensions (Random Forest classifier). More feature vectors must be produced in the future if comparable research is carried out to produce the dataset utilized in this study so that the classifiers can build a better understanding of the issue at hand. For improved accuracy and performance in the future, our work aims to employ a machine learning model with some deep learning techniques like CNN. The deployment of DL models may also reduce computational time as it requires fewer manual pre-processing tasks.

REFERENCES

Kumar, C. N., Keerthana, D., Kavitha, M., & Kalyani, M. (2022). Customer Loan Eligibility Prediction Using Machine Learning Algorithms in Banking Sector. *In 2022 7th International Conference on Communication and Electronics Systems (ICCES)*, 1007-1012.

Min Sue park, Hwijae son ,Chongseok hyun & Hyung ju hwang (2021). Explainability of Machine Learning Models for Bankruptcy, *Prediction in IEEE* .

Mohan Kumar, M., Amuthakkani, S. & Jeyamala,. (2016). Comparative analysis of decision tree algorithms for the prediction of eligibility of a man for availing bank loan. Age, 19, 60.

Awodele, O, Alimi, S, Ogunyolu, O, Solanke, O, Iyawe S. & Adegbie F (2022). Cascade of Deep Neural Network And Support Vector Machine for Credit Risk Prediction in 2022 5th Information Technology for Education and Development (ITED).

Reddy, C. S., Siddiq, A. S. & Jayapandian, N. (2022). Machine Learning based Loan Eligibility Prediction using Random Forest Model. *In 2022 7th International Conference on Communication and Electronics Systems (ICCES)* ,1073-1079.

Sheikh, M. A., Goel, A. K. & Kumar, T. (2020). An approach for prediction of loan approval using machine learning algorithm. *In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 490-494.

Singh, V., Yadav, A., Awasthi, R., & Partheeban, G. N. (2021). Prediction of modernized loan approval system based On machine learning approach. *In 2021 International Conference on Intelligent Technologies (CONIT*), 1-4.

Ugochukwu .E. Orji ,Chikodili .H. Ugwuishiwu, Joseph. C. N. Nguemaleu & Peace. N. Ugwuanyi (2022). Machine Learning Models For Predicting Bank Loan Eligibility in IEEE 2022 Nigerian.