



---

## **Loan Approval Prediction Using Machine Learning**

*Goliya Bhavani*

*Student, Rajam, Vizianagaram, 532127, india.*

---

### **ABSTRACT**

In modern world loans are the major requirement for organizations and individuals alike. By this only, banks get major part of total profit. Due to significant amount of financial constraints, the activity of loan taking has become more. Loan approval process is a challenging task for any financial institution as they need to go through several documents of a person and it becomes plenty when they need to go through several customers. This is a classification problem which defines whether the bank should consider the application for loan approval or not. Prediction of loan approval can be done by various parameters like credit scoring, income, age, marital status, gender, etc. Prediction model is used to classify the customers based on the documents submitted. So that the customer who are eligible and have all the documents required are separated and processed to the next step in loan approval. The aim of this is quick, immediate and easy way to choose the better loan applicants. Machine learning techniques which are useful in predicting the loan approval to a person are Logistic Regression, Random Forest, Naive Bayes. These techniques are useful in reducing the human effort and make efficient way to find the eligible candidates.

**Keywords:** *Loan Approval, Classification, Prediction, Machine Learning, Logistic Regression, Random Forest, Naive Bayes.*

---

### **1. Introduction**

Banks Loans have become an important source of external financing for firms and households due financial constraints to develop firms and business. The most profit for the bank comes only from loans however the increase of loan lending is associated with a number of risks, such as risk of defaulting or credit risk, which is linked to the inability of the borrower to pay back the loan at the agreed time and conditions. Banks have limited goods, so it is essential to choose the right applicant who repays the loan within the time limits. Selecting the right candidate is the main responsibility of the bank. When process is done manually many problems arise in choosing the right person for the approval. Bank should choose correct one otherwise the bank has to face financial trouble and lack of profits. Banks aim is to invest their asset in the safe hands. If a bank is providing the loan to a person the bank should think about for what purpose they are taking, will they repay the amount they took, are their documents are valid or not, etc. Thus we need some machine learning algorithms which chooses the applicant automatically. These machine learning models supports both the employee and applicants. The primary purpose of this model is to find the right one and reduce the selection time for choosing. There are several ways to predict but data mining is important to note that because it takes the previous records of customers and train the system to predict the approval. Some of the methods used here to predict are Naïve bayes, Logistic Regression, Support Vector Machine, Classification, Random Forest. So based on the accuracy of these used algorithms we can predict the loan approval easily. Based on the machine learning we have two types of datasets, one is training dataset and the other is test dataset. Here every model will works on some variables and tries to give us the result whether to approve the loan to a person or not. The core objective of this paper is to create a less complex system for prediction of loan model.

---

### **2. Literature Survey**

#### **In Paper [1]:**

The paper focuses on the use of machine learning techniques to predict the safety of granting loans to individuals. It highlights the importance of selecting the right applicants to minimize risk and optimize the allocation of bank assets. The paper is divided into four sections: data collection, comparison of machine learning models, training of the system, and testing. The authors collected data on loan amount, loan amount term, credit history, property area, and loan status. The paper also discusses the advantages of the loan prediction system. It provides a quick and easy way to choose deserving applicants, automates the validation process, and allows for setting time limits for loan approval. The system is private and cannot be altered by stakeholders. Results can be sent to various departments for appropriate action.

#### **In Paper [2]:**

This paper includes a benchmark study by Moscato, Picariello, and Sperlí, which compared various machine learning approaches for credit score prediction. Additionally, Sheikh, Goel, and Kumar proposed an approach for the prediction of loan approval using machine learning. These studies provide a foundation for understanding the application of machine learning in the context of loan approval and credit score prediction. Furthermore, the

paper discusses the importance of understanding customer behaviour and the limitations of machine learning models in comprehensively understanding complex customer behaviour. The paper also emphasizes the significance of machine learning models in verifying the credibility of loan customers based on credit history, which is indicative of future repayment behaviour. This aligns with the goal of the study to create a machine learning model for bank stakeholders to assess the credibility of loan customers based on credit history and previous loan settlement patterns.

**In Paper [3]:**

Machine learning (ML) techniques have proven to be useful in predicting outcomes for large amounts of data. In this study, three ML algorithms, Logistic Regression (LR), Decision Tree (DT), and Random Forest (RF), are applied to predict the loan approval of customers. The experimental results indicate that the accuracy of the Decision Tree algorithm is better compared to Logistic Regression and Random Forest approaches. The use of ML algorithms in loan approval prediction can provide a more accurate predictive modeling system for banking organizations, helping them make informed decisions and improve loan recovery rates. Limitation are lack of comparison with other algorithms, lack of consideration for further engineering.

**In Paper [4]:**

The paper explores the use of various machine learning algorithms for predicting loan approval, including logistic regression, random forest, decision tree, support vector machine, and naive bayes. It highlights that the random forest algorithm outperforms the other algorithms in terms of accuracy for loan approval prediction. The paper includes a detailed methodology for building the prediction model using machine learning algorithms. It compares the accuracy of different machine learning algorithms in the context of loan approval prediction.

**In Paper [5]:**

Machine learning (ML) techniques have proven to be useful in predicting outcomes for large amounts of data. In this study, three ML algorithms, Logistic Regression (LR), Decision Tree (DT), and Random Forest (RF), are applied to predict the loan approval of customers.

The experimental results indicate that the accuracy of the Decision Tree algorithm is better compared to Logistic Regression and Random Forest approaches. The use of ML algorithms in loan approval prediction can provide a more accurate predictive modeling system for banking organizations, helping them make informed decisions and improve loan recovery rates. The limitations are lack of comparison with other algorithms, lack of consideration for further engineering.

---

**3. Methodology:**

Firstly, the loan approval dataset, containing 614 rows and 13 columns classified into two classes (Approve and Reject), is acquired. The data is then cleaned, prepared, and manipulated to ensure accuracy in the model's performance. This process is crucial as real-world data often contains unorganized, missing, or noisy elements, and having a clean dataset is essential for the accuracy of the model. Additionally, feature engineering is employed to create new input or target features from existing features, aiming to better represent the machine learning problem to the model and improve its accuracy. Feature scaling, a method used to normalize the range of independent variables or features of data, is also applied to the dataset. The data is split into 70% for training and 30% for testing, and machine learning algorithms such as Naive Bayes Classifier, Gradient Boosting, Support Vector Machine, and K-Nearest Neighbor are used for model training and evaluation. Overall, the data preprocessing involves acquiring, cleaning, and manipulating the loan approval dataset, as well as employing feature engineering, feature scaling, and model training and evaluation.

SVM and Random Forest were found to have the best accuracy, both achieving 88.89% accuracy, with Random Forest ultimately being selected for implementation due to its reliability and robustness. The Decision Tree algorithm was used, and its accuracy was improved by implementing more decision trees as a Random Forest classifier, which increased the accuracy to 88.89%. KNN, Logistic Regression, and Gaussian NB provided accuracies of 84.44%, 86%, and 80% respectively, with Gradient Boosting performing the worst with an accuracy of 77.78%. The study also discusses the K-Nearest Neighbour (KNN) algorithm, Decision Tree, and Gradient Boosting, highlighting their characteristics and applications in classification problems. Overall, the paper evaluates and compares the performance of various machine learning algorithms, ultimately selecting Random Forest for implementation due to its high accuracy and reliability. The following shows a brief explanation about the prediction of loan.



#### 4. Discussion:

This paper presents a machine learning model for predicting loan approval, aiming to assist banks in making safer lending decisions. The study utilizes a dataset of previous loan records to train and evaluate six machine learning classification models: Decision Trees, Random Forest, Support Vector Machine, Linear Models, Neural Network. The evaluation parameters used in the study include accuracy, precision, recall, F1 score, area under the receiver operating characteristic (ROC) curve, and mean decrease accuracy for Random Forest models. These parameters are essential for assessing the performance of the machine learning models in predicting loan approval. The results and discussion of the study highlight the effectiveness of the machine learning models in predicting loan approval. The models are evaluated based on their ability to accurately identify safe loan applicants, thereby reducing the risk for banks and optimizing the allocation of assets. The study concludes that the machine learning model is highly efficient and meets the requirements of banking institutions, providing a valuable tool for automating the loan approval process. Overall, the paper provides insights into the application of machine learning in the banking sector, specifically in the context of loan approval prediction. The results and discussion emphasize the potential benefits of using machine learning models to enhance the efficiency and accuracy of loan approval processes, ultimately benefiting both banks and loan applicants.



#### 5. Conclusion:

Based on the project, it can be concluded that machine learning is a promising tool for checking the credibility of a customer who is taking a loan based on credit history. We have used different models for comparing them and arrive at the best model for the particular problem. Different models used are

SVM, Random Forest, Decision tree, KNN, Logistic Regression, Gaussian NB, Gradient Boosting. The best accuracy was given by SVM and Random Forest with a 88.89 accuracy. Finally Random Forest, SVM is selected for implementation since it is an ensemble technique and it is more reliable and more robust than other algorithms.

To improve the accuracy of the model, additional data can be collected and exploring more feature engineering techniques can be done. Moreover, it is important to note that customer behaviour is a complex one to be understood completely, as there are many other factors which can't be taken into consideration. As such, the limitations of machine learning models need to be kept in consideration.

---

## 6. References:

---

- [1] Arun, K., Ishan, G., & Sanmeet, K. (2016). Loan approval prediction based on machine learning approach. *IOSR J. Comput. Eng.*, 18(3), 18-21.
- [2] Yaraswini, P., Arunasri, P., Pratyusha, Y., Reddy, P. S., & Kumari, S. Analysis and Forecasting of bank loan approval data using machine learning algorithms.
- [3] Kadam, A. S., Nikam, S. R., Aher, A. A., Shelke, G. V., & Chandgude, A. S. (2021). Prediction for loan approval using machine learning algorithm. *International Research Journal of Engineering and Technology (IRJET)*, 8(04).
- [4] Mella, N. V. V. P., & Sai, R. R. LOAN APPROVAL PREDICTION
- [5] Tejaswini, J., Kavya, T. M., Ramya, R. D. N., Triveni, P. S., & Maddumala, V. R. (2020). Accurate loan approval prediction based on machine learning approach. *Journal of Engineering Science*, 11(4), 532-532.
- [6] Sarkar, A. (2021). Machine learning techniques for recognizing the loan eligibility. *International Research Journal of Modernization in Engineering Technology and Science*, 3(12), 1135-1142.