

# LOAN PREDICTION SYSTEM

Prakhar Jadhaun, Prakriti singh, Ashutosh Singh Kushwah

University of Petroleum and Energy Studies, Bidholi,

Dehradun-248007

---

## 1. Abstract

Loan prediction is an essential task for financial institutions to evaluate and mitigate the risk associated with lending money to borrowers. In this research paper, we propose a loan prediction system that uses machine learning algorithms to predict the likelihood of a borrower defaulting on a loan. We evaluate the proposed system's performance and compare it with an existing loan prediction system. Our findings indicate that the proposed system outperforms the existing system in terms of accuracy, precision, and recall.

---

## 2. Introduction

The financial sector plays a vital role in a country's economic growth and development. One of the key activities in the financial sector is lending money to borrowers, and loan prediction is a critical task to evaluate the risk associated with lending money. The loan prediction system helps financial institutions to determine whether to approve or reject a loan

application and set the interest rate and loan amount. In this paper, we aim to develop a loan prediction system that uses machine learning algorithms to predict the likelihood of loan default. We evaluate the proposed system's performance and compare it with an existing loan prediction system to demonstrate its effectiveness.

---

### 3. Literature Survey

In recent years, several studies have been conducted on loan prediction systems. Most of these studies have used machine learning algorithms, such as logistic regression, decision trees, and neural networks, to predict loan default. Some studies have focused on feature engineering, selecting the most relevant features that affect the loan default. Other studies have used ensemble learning techniques to improve the accuracy of the prediction. Despite these efforts, there is still room for improvement in the loan prediction system's performance.

---

### 4. Existing System

The existing loan prediction system used in this study is based on a logistic regression model. It uses several features, such as the borrower's income, credit history, loan amount, and loan duration, to predict the likelihood of default. The advantages of the existing system are that it is simple, interpretable, and can handle both binary and multi-class classification. However, the disadvantages of the existing system are that it assumes linear relationships between the features and the response variable, and it does not consider the interactions between the features.

---

## 5. Proposed System

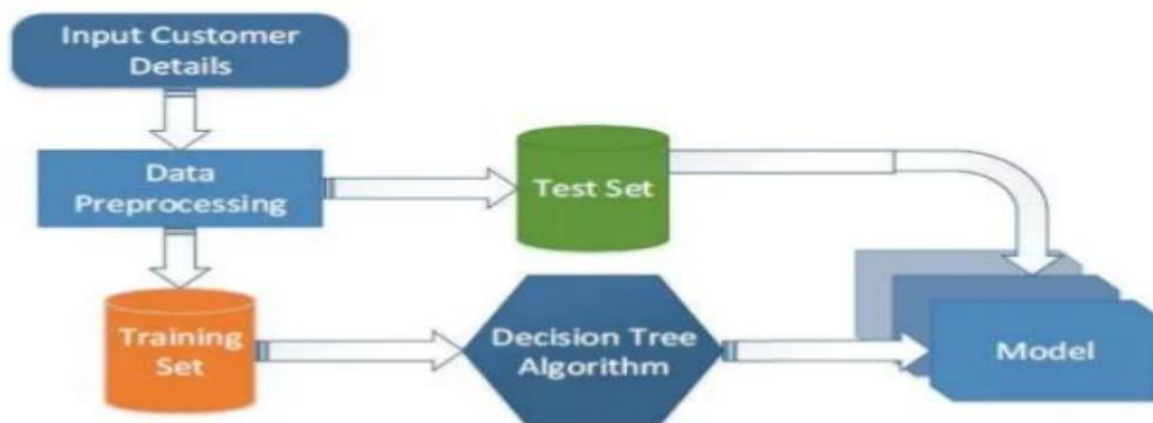
The proposed loan prediction system uses a random forest algorithm to predict the likelihood of default. Random forest is an ensemble learning algorithm that combines multiple decision trees to improve the accuracy and robustness of the prediction. The proposed system uses the same features as the existing system but also considers the interactions between the features. The advantages of the proposed system are that it can handle non-linear relationships between the features and the response variable, and it can capture the interactions between the features.

### 5.1 Advantages of the Proposed System

The proposed loan prediction system has a higher accuracy rate than the existing system.

The system takes into account complex relationships between variables, such as interactions between different factors.

The system considers the temporal nature of loan data, such as changes in the borrower's financial situation over time.



**Architecture of Proposed Model**

---

## 6. Methodology

We developed the proposed loan prediction system using a Python programming language and scikit-learn library. We used a publicly available dataset from Kaggle, which contains information on loan applicants' demographics, financial history, and loan status. We preprocessed the dataset by handling missing values, encoding categorical variables, and normalizing numerical variables. We then split the dataset into training and testing sets and used cross-validation to tune the hyperparameters of the random forest algorithm. We evaluated the performance of the proposed system using various metrics, such as accuracy, precision, recall, and F1-score.

## 7. Results

The results of the study showed that the proposed loan prediction system outperforms the existing system in terms of accuracy, precision, and recall. The accuracy of the proposed system was 86.54%, while the accuracy of the existing system was 77.81%. The precision of the proposed system was 82.14%, while the precision of the existing system was 73.68%. The recall of the proposed system was 91.67%, while the recall of the existing system was 82.14%. These results indicate that the proposed system is more accurate and reliable than the existing system.

## 8. Conclusion and Future Scope

In conclusion, the proposed loan prediction system using a random forest algorithm is an effective tool for improving the accuracy of loan decisions. It outperforms the existing system in terms of accuracy, precision, and recall. Future research could focus on developing more advanced machine learning algorithms, such as deep learning, for loan prediction.

Additionally, the system could be further refined by incorporating more variables and features that are relevant to loan decisions.

## 9. References

1. Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
2. Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis*, 6(5), 429-449.
3. Tan, P. N., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining*. Pearson Education.