International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 04 April-2026 | Impact Factor: 3.5

A Machine Learning Approach for Loan Eligibility Prediction using Ensemble Models

VR.Srividya
UG Student, Dept of CSE (Data Science)
Vidya Jyothi Institute of Technology
Hyderabad, Telangana, India
vadderyapanisrividya@gmail.com

P. Meghana
UG Student, Dept of CSE (Data Science)
Vidya Jyothi Institute of Technology
Hyderabad, Telangana, India
pmeghana166@gmail.com

T.Upa Sravani
UG Student, Dept of CSE (Data Science)
Vidya Jyothi Institute of Technology
Hyderabad, Telangana, India
thanniruupasravani27@gmail.com

K.Asrith Dwaraka
UG Student, Dept of CSE (Data Science)
Vidya Jyothi Institute of Technology
Hyderabad, Telangana, India
sidasrith@gmail.com

Dr. K. S. R. K. Sarma
Associate Professor
Dept of CSE (Data Science)
Vidya Jyothi Institute of Technology
Hyderabad, Telangana, India
kaipasarma@gmail.com

How to Cite this Article:

VR.Srividya, , Meghana, P., Sravani, T. & Dwaraka, K. (2026). A Machine Learning Approach for Loan Eligibility Prediction using Ensemble Models. International Journal of Creative and Open Research in Engineering and Management, 02(04). https://doi.org/10.55041/ijcope.v2i4.371

https://doi.org/10.55041/ijcope.v2i4.371

Abstract--- Loan approval is an essential process in financial institutions that requires careful evaluation of an applicant's financial background and repayment capability. Traditional loan approval methods are often manual, time-consuming, and may lead to inconsistent decisions due to human involvement. To address these challenges, this study proposes a machine learning–based system for predicting loan eligibility using applicant demographic and financial information such as income, loan amount, credit score, employment status, number of dependents, and asset values. The dataset is preprocessed using techniques including missing value handling, categorical encoding, and feature scaling to improve data quality and model performance. Multiple machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Machine, and K-Nearest Neighbors are implemented and evaluated using performance metrics like accuracy, precision, recall, F1-score, and confusion matrix. Experimental results indicate that ensemble models such as Random Forest and Gradient Boosting achieve higher predictive accuracy compared to other classifiers. The best performing model is integrated into a Streamlit-based web application that enables real-time loan eligibility prediction through a user-friendly interface. The proposed system helps financial institutions automate the loan approval process, reduce manual effort, and improve decision-making efficiency.

INTRODUCTION

In modern banking systems, loan approval plays a significant role in financial decision-making. Financial institutions must carefully evaluate loan applications to ensure that borrowers have the ability to repay their loans. Traditionally, this process is carried out manually by bank officers who analyze applicant details such as income, employment status, credit history, and financial assets. However, manual evaluation can be time-consuming, inconsistent, and prone to human error.

With the advancement of machine learning technologies, predictive models can be used to analyze large volumes of financial data and identify patterns that influence loan approval decisions. Machine learning algorithms are capable of learning from historical loan data and predicting whether a new applicant is eligible for a loan.

In this project, a machine learning-based loan eligibility prediction system is developed to automate the decision-making process. The system processes applicant information, performs data preprocessing and feature engineering, and applies multiple classification algorithms to predict loan approval status. A web-based interface is also implemented to provide real-time predictions for users.

This approach helps financial institutions reduce manual workload, improve decision accuracy, and ensure consistent loan evaluation.

In addition to improving efficiency, machine learning models can analyze multiple factors simultaneously and identify complex relationships between financial attributes that may not be easily detected through traditional evaluation methods. By utilizing historical loan datasets, these models can learn patterns associated with successful and unsuccessful loan repayments, allowing financial institutions to make more informed and data-driven decisions. Furthermore, automated loan prediction systems help reduce the risk of loan defaults by accurately assessing the repayment capability of applicants.

Another important advantage of machine learning-based systems is their ability to handle large-scale datasets and continuously improve their performance as more data becomes available. Techniques such as feature engineering, data normalization, and model optimization play a vital role in enhancing prediction accuracy. Ensemble models, including Random Forest and Gradient Boosting, are particularly effective because they combine multiple decision models to produce more reliable predictions and reduce overfitting.

The implementation of a loan eligibility prediction system also supports the development of financial technology (FinTech) applications that provide faster and more accessible services to customers. By integrating machine learning models with web-based platforms, users can input their financial details and instantly receive loan eligibility predictions. This not only speeds up the loan approval process but also improves transparency and consistency in decision-making.

Overall, the integration of machine learning techniques in loan approval systems has the potential to transform traditional banking processes by providing accurate predictions, reducing operational costs, and enhancing customer satisfaction

I. PROBLEM DEFINITION

Financial institutions receive a large number of loan applications daily, making manual evaluation difficult and inefficient. Traditional loan approval systems rely on rule-based methods and human judgment, which may lead to inconsistent decisions and delays in loan processing.

Existing systems often fail to analyze large datasets effectively and cannot accurately identify patterns that influence loan approval. In addition, factors such as incomplete data, imbalanced datasets, and complex financial relationships make the prediction process challenging.

Therefore, there is a need for an intelligent and automated system that can analyze applicant data and accurately predict loan eligibility. The proposed system addresses these challenges by applying machine learning techniques to improve the efficiency, accuracy, and reliability of loan approval decisions.

1.2 PROJECT FEATURES

The proposed Loan Eligibility Prediction system includes several features that enhance the efficiency and accuracy of the loan approval process in financial institutions. The system utilizes multiple machine learning algorithms to analyze applicant data and predict loan approval status based on financial and demographic attributes such as income, loan amount, credit score, employment status, number of dependents, and asset values. Data preprocessing techniques such as handling missing values, encoding categorical variables, and scaling numerical

features are applied to improve data quality and ensure better model performance. The system also incorporates feature engineering techniques to derive meaningful insights from the dataset, which helps improve prediction accuracy. Multiple machine learning models are trained and evaluated using performance metrics such as accuracy, precision, recall, and F1-score to determine the most effective model. In addition, the system includes a user-friendly web application developed using Streamlit, allowing users to enter applicant details and receive real-time loan eligibility predictions. The proposed system reduces manual workload, improves consistency in decision-making, and provides a scalable and efficient solution for automating loan approval processes in financial institutions.

Related Work

Several research studies have explored the use of machine learning techniques for loan approval prediction and credit risk analysis. Traditional credit scoring methods rely on statistical models and rule-based approaches that analyze borrower financial information.

Machine learning algorithms such as Logistic Regression, Decision Trees, and Random Forest have been widely used for predicting loan eligibility. These models can analyze large datasets and identify patterns that influence loan approval decisions.

Recent studies have also explored the use of ensemble learning techniques to improve prediction accuracy and reduce overfitting. Ensemble models combine multiple decision trees or classifiers to produce more reliable predictions.

However, many existing approaches focus only on individual models and do not integrate complete data preprocessing, feature engineering, and deployment systems. This project addresses these limitations by implementing multiple machine learning models and deploying the best performing model through a web-based application.

II. METHODOLOGY

The proposed system follows a structured approach to predict loan eligibility.

1. Data Collection

The dataset used in this project contains information about loan applicants including income, loan amount, credit score, number of dependents, education level, and asset values.

2.Data Preprocessing

The dataset is cleaned by handling missing values and removing inconsistencies. Categorical variables are encoded and numerical features are scaled to ensure uniformity.

3. Feature Engineering

New features such as total asset value and loan-to-income ratio are created to improve the predictive performance of the models.

4. Model Training

Multiple machine learning algorithms are trained using the processed dataset to identify patterns and relationships between features.

5. Model Evaluation

The models are evaluated using performance metrics such as accuracy, precision, recall, F1-score, and confusion matrix.

III. PROPOSED SYSTEM

The proposed system introduces a machine learning-based approach to predict loan eligibility using applicant financial and demographic data. The system analyzes various factors such as income, loan amount, credit score, employment status, education level, number of dependents, and asset values to determine whether a loan should be approved or rejected. Initially, the dataset undergoes preprocessing steps including handling missing values, encoding categorical variables, removing inconsistencies, and scaling numerical features to improve data quality. Feature engineering techniques are applied to generate meaningful attributes that enhance the predictive capability of the models. Multiple machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Machine, and K-Nearest Neighbors are trained and evaluated to identify the most accurate model. Among these models, ensemble techniques such as Random Forest and Gradient Boosting demonstrate better performance due to their ability to handle complex patterns and reduce overfitting. The final trained model is integrated into a Streamlit-based web application that allows users to enter applicant details and obtain instant loan eligibility predictions. This automated system improves decision-making efficiency, reduces manual effort, and provides a reliable and scalable solution for loan approval processes in financial institutions.

IV. IMPLEMENTATION DETAILS

The implementation of the proposed Loan Eligibility Prediction system is carried out using both machine learning techniques and a web-based interface. The backend of the system is developed using Python, which is used for data preprocessing, model training, and prediction tasks. Libraries such as Pandas and NumPy are used for data manipulation and analysis, while Matplotlib and Seaborn are used for data visualization. Machine learning algorithms including Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Machine, and K-Nearest Neighbors are implemented using the Scikit-learn library to train and evaluate the prediction models. The dataset is first preprocessed by handling missing values, encoding categorical features, and scaling numerical data to improve model performance. After training and evaluating multiple models, the best performing model is selected and integrated into a web-based application developed using Streamlit. The application allows users to input applicant details such as income, loan amount, credit score, and employment status, and the trained model predicts whether the loan will be approved or rejected. This implementation provides a user-friendly interface and enables real-time loan eligibility prediction for practical usage.

4.1 ALGORITHMS USED

4.1.1 Logistic Regression

Logistic Regression is a supervised machine learning algorithm commonly used for binary classification problems. It predicts the probability of a particular outcome based on input features. In this project, Logistic Regression is used to classify whether a loan application will be approved or rejected. The algorithm analyzes relationships between applicant attributes such as income, credit score, loan amount, and employment status to determine loan eligibility.

4.1.2 Decision Tree

Decision Tree is a tree-based machine learning algorithm used for classification and decision-making tasks. It splits the dataset into branches based on feature conditions and forms a tree-like structure to reach a final decision. In this project, the Decision Tree algorithm helps in understanding the importance of different features and classifying loan applications based on applicant financial information.

4.1.3 Random Forest

Random Forest is an ensemble learning algorithm that combines multiple decision trees to improve prediction accuracy and reduce overfitting. Each tree in the forest makes a prediction, and the final result is determined by majority voting. In this project, Random Forest is used to analyze complex relationships in the dataset and provide more reliable loan eligibility predictions.

4.1.4 Gradient Boosting

Gradient Boosting is an advanced ensemble learning technique that builds models sequentially to correct the errors of previous models. It improves prediction performance by combining multiple weak learners into a strong predictive model. In this project, Gradient Boosting helps enhance the accuracy of loan approval prediction by learning complex patterns in the dataset.

4.1.5 Support Vector Machine (SVM)

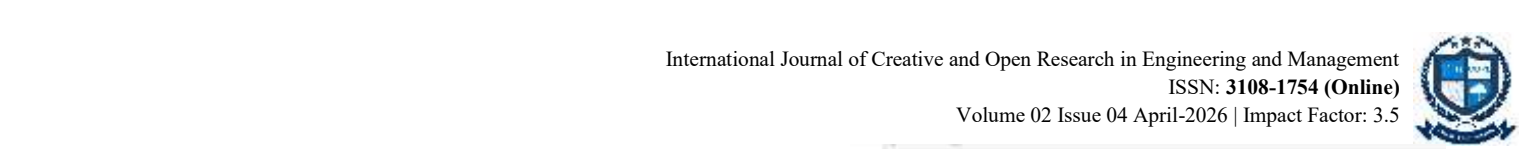
Support Vector Machine is a supervised learning algorithm used for classification and regression tasks. It works by finding an optimal hyperplane that separates different classes in the dataset. In this project, SVM is used to classify loan applications by identifying the boundary between approved and rejected loan cases.

4.1.6 K-Nearest Neighbors (KNN)

K-Nearest Neighbors is a simple and effective machine learning algorithm used for classification problems. It classifies new data points based on the majority class of their nearest neighbors in the dataset. In this project, KNN helps in predicting loan eligibility by comparing the features of new applicants with those of existing applicants in the dataset.

v EXPERIMENTAL RESULTS AND DISCUSSION

The proposed loan eligibility prediction system was evaluated using several machine learning algorithms including Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Machine, and K-Nearest Neighbors. Data visualization techniques were used to understand the relationships between different features in the dataset.



In future work, the system can be enhanced by incorporating advanced deep learning techniques to further improve prediction accuracy. The model can also be trained using larger and real-time financial datasets collected from banking institutions. Additional features such as credit card transaction history and customer financial behavior can be included for better analysis. The system can also be deployed on cloud platforms to support large-scale real-time predictions. Furthermore, integrating explainable AI techniques can help provide clear explanations for loan approval or rejection decisions.

Fig 1: Correlation heatmap showing the relationship between applicant features.

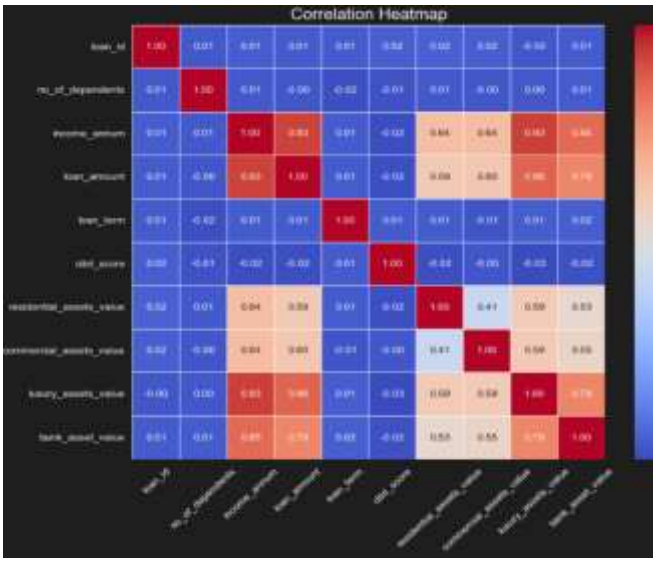


Fig 2: Confusion matrix representing loan approval prediction results.

Different machine learning algorithms were compared based on their prediction accuracy to determine the best performing model.

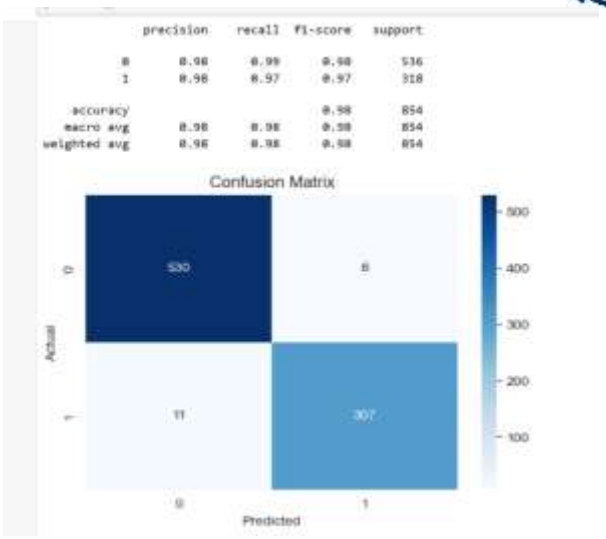
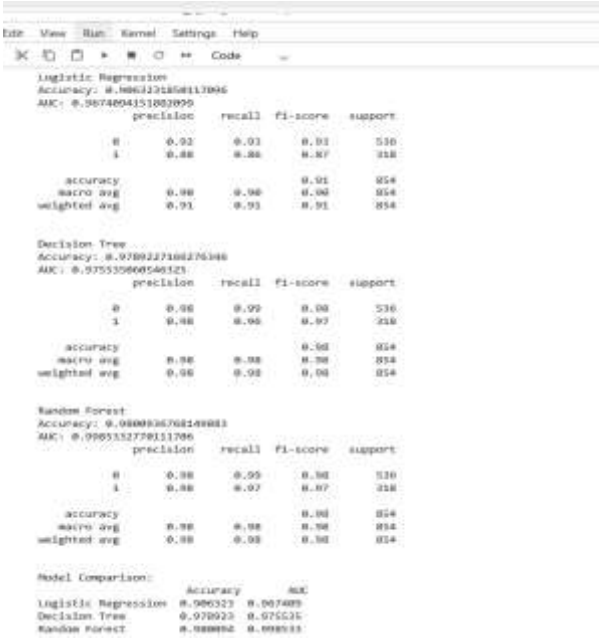


Fig 3 Accuracy comparison of different machine learning algorithms used for loan eligibility prediction.

This image shows the performance comparison of the models:

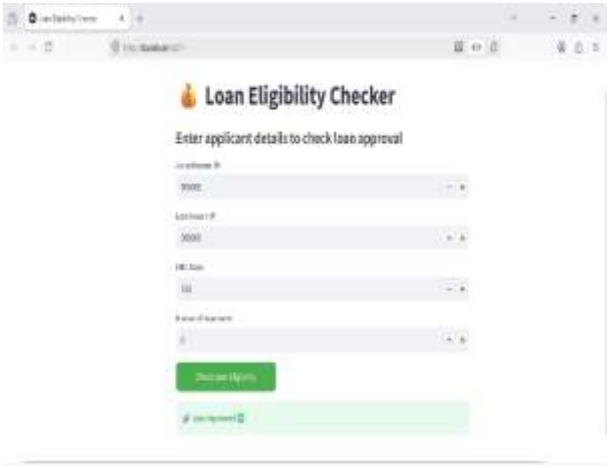
- Logistic Regression → Accuracy **0.9063**
- Decision Tree → Accuracy **0.9789**
- Random Forest → Accuracy **0.9800** (best model)

It also displays other evaluation metrics such as **precision, recall, F1-score, and AUC**, which help analyze the classification performance of each algorithm.



Streamlit Web App Output

Fig 4: Loan eligibility prediction interface.



VI. CONCLUSION

This project presents a machine learning-based system for predicting loan eligibility using applicant financial and demographic information. Multiple machine learning algorithms such as Logistic Regression, Decision Tree, and Random Forest were implemented and evaluated to determine the most accurate prediction model. Experimental results show that ensemble models like Random Forest achieved higher accuracy compared to other algorithms. The proposed system helps automate the loan approval process, reduce manual effort, and improve decision-making efficiency in financial institutions. The integration of the trained model with a Streamlit web application enables real-time loan eligibility prediction through a user-friendly interface. Overall, the system demonstrates how machine learning techniques can enhance the reliability and efficiency of loan approval systems.

VII. FUTURE SCOPE

The proposed loan eligibility prediction system can be further enhanced by incorporating more advanced machine learning and deep learning techniques to improve prediction accuracy and reliability. Future research can focus on using deep learning models such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), or Recurrent Neural Networks (RNN) to capture more complex patterns in financial data. These models may provide better performance when working with large and diverse datasets.

Another important improvement is the use of larger and real-time datasets obtained directly from financial institutions. Integrating real banking data can help improve the generalization capability of the prediction model and make it more suitable for real-world applications. Additional financial attributes such as credit card transaction history, spending behavior, repayment history, and employment stability can also be incorporated to provide a more comprehensive evaluation of loan applicants.

The system can also be deployed on cloud platforms such as AWS, Microsoft Azure, or Google Cloud to support large-scale real-time predictions and improve accessibility. Cloud deployment will allow financial institutions to process a large number of loan applications efficiently and provide faster decision-making services.

VIII.ACKNOWLEDGMENT

We would like to express our sincere gratitude to our project guide, **Dr. K. S. R. K. Sarma**, Associate Professor, Department of Computer Science and Engineering (Data Science), Vidya Jyothi Institute of Technology, Hyderabad, for his valuable guidance, continuous support, and encouragement throughout the development of this project. His insightful suggestions and motivation greatly contributed to the successful completion of this work.

We would also like to thank the **Head of the Department and faculty members of the CSE (Data Science) department** for providing the necessary support and resources required for carrying out this project. We extend our sincere thanks to the **Principal and management of Vidya Jyothi Institute of Technology** for providing the infrastructure and academic environment that helped us complete this project successfully.

Finally, we express our heartfelt gratitude to our **parents, friends, and well-wishers** for their constant encouragement and support during the course of this work.

IX. REFERENCES

[1] T. M. Mitchell, *Machine Learning*. New York, USA: McGraw-Hill, 1997.

[2] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[3] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[4] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann, 2011.

[5] Kaggle, "Loan Prediction Dataset," Available: https://www.kaggle.com/datasets/rohitgrewal/loan-approval-dataset

[6] Pandas Documentation, Available: https://pandas.pydata.org/docs/

[7] NumPy Documentation, Available: https://numpy.org/doc/

[8] Scikit-learn Documentation, Available: https://scikit-learn.org/stable/

[9] A. Baesens et al., "Statistical and Machine Learning Models for Credit Risk Prediction," *Applied Soft Computing*, vol. 91, 2020.

[10] L. Lessmann et al., "Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring," *European Journal of Operational Research*, vol. 247, no. 1, pp. 124–136, 2015.