Enhancing Loan Approval Prediction Using Ensemble Machine Learning Techniques Through Comprehensive Model Comparison and Performance Evaluation Analysis

Qing Tan[1,*]

[1]Computer Science Department, Athabasca University, Canada

ABSTRACT

Loan approval prediction is a crucial task in the financial sector, as it directly impacts risk management and decision-making processes. This study aims to enhance the accuracy of loan approval prediction by applying ensemble machine learning techniques and comparing their performance with a baseline model. The dataset used in this study contains borrower demographic, financial, and employment-related attributes, and missing values were handled using a deletion method to ensure data consistency. Several models were implemented, including Logistic Regression as the baseline model, as well as ensemble methods such as Random Forest, Gradient Boosting, and Voting Classifier. The models were evaluated using multiple performance metrics, including Accuracy, Precision, Recall, F1-Score, and ROC-AUC. The experimental results show that ensemble models consistently outperform the baseline model across all evaluation metrics. Random Forest achieved the highest ROC-AUC, indicating superior discriminative capability, while the Voting Classifier provided the best balance between precision and recall, resulting in the highest F1-Score. In addition, feature importance analysis revealed that CreditScore, Income, and Employment Type are the most influential factors in loan approval decisions. These findings demonstrate that ensemble learning methods are effective in improving predictive performance and can provide reliable support for loan approval decision-making systems.

Keywords Loan Approval Prediction, Ensemble Learning, Random Forest, Classification Models, Machine Learning

INTRODUCTION

Loan approval prediction is a critical task in the financial sector, as it directly affects risk management, profitability, and operational efficiency of financial institutions [1]. Accurate assessment of borrower eligibility enables lenders to minimize default risk while ensuring that creditworthy applicants are not unfairly rejected. Traditionally, loan approval decisions have relied on manual evaluation and rule-based systems, which are often time-consuming and subject to human bias. In addition, these conventional approaches are limited in their ability to capture complex interactions among multiple factors such as credit score, income level, employment status, and demographic characteristics.

With the rapid advancement of data availability and computational capabilities, machine learning techniques have been widely adopted to

Submitted: 18 August 2025
Accepted: 30 October 2025
Published: 23 May 2026

Corresponding author
Qing Tan,
qingt@athabascau.ca

Additional Information and Declarations can be found on page 156

DOI: 10.47738/jdmdc.v3i2.65

How to cite this article: Q. Tan, "Enhancing Loan Approval Prediction Using Ensemble Machine Learning Techniques Through Comprehensive Model Comparison and Performance Evaluation Analysis," *J. Digit. Mark. Digit. Curr.*, vol. 3, no. 2, pp. 144-158, 2026.

improve the accuracy and efficiency of loan approval prediction [2]. Various models such as Logistic Regression, Decision Trees, and Support Vector Machines have been extensively used due to their simplicity and interpretability. However, these baseline models often struggle to capture non-linear relationships and complex feature interactions present in real-world financial data. As a result, their predictive performance may be limited, particularly in datasets with high variability and imbalance.

To address these limitations, ensemble learning methods have emerged as state-of-the-art approaches in classification tasks [3]. Techniques such as Random Forest, Gradient Boosting, and Voting Classifier combine multiple models to improve predictive accuracy and robustness. These methods are capable of reducing variance, handling non-linear relationships, and improving generalization performance. As a result, ensemble models have demonstrated superior performance in various domains, including credit risk assessment and financial decision-making.

Several research gaps can be identified in the existing literature. First, many studies focus on a single model or a limited set of models, without providing a comprehensive comparison between baseline and multiple ensemble approaches under the same experimental conditions [4]. Second, the issue of class imbalance, which is common in loan approval datasets, is often not adequately addressed, potentially leading to biased model performance. Third, there is limited emphasis on model interpretability, particularly in identifying key features that influence loan approval decisions.

Therefore, this study aims to enhance loan approval prediction by implementing and comparing baseline and ensemble machine learning models within a unified framework. Specifically, Logistic Regression is used as a baseline model and compared with ensemble methods, including Random Forest, Gradient Boosting, and Voting Classifier. The models are evaluated using multiple performance metrics such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC to ensure a comprehensive assessment. In addition, feature importance analysis is conducted to identify the most influential factors in loan approval decisions. The findings of this study are expected to provide valuable insights for improving predictive performance and supporting more reliable decision-making in financial institutions [5].

Literature Review and Related Works

Machine learning techniques have been widely applied in financial decision-making, particularly in loan approval and credit risk assessment. Traditional classification models such as Logistic Regression have been commonly used due to their simplicity and interpretability, especially in structured financial datasets [6], [7]. These models provide a baseline for comparison but are often limited in their ability to capture complex and non-linear relationships among variables, which are common in real-world financial data [8]. As a result, their predictive performance may not be optimal in more complex scenarios.

To overcome these limitations, tree-based models such as Decision Trees and Random Forest have gained significant attention. Random Forest, in particular, has demonstrated strong performance in classification tasks due to its ability to reduce overfitting and improve generalization by aggregating multiple decision trees [9], [10]. Similarly, boosting-based methods such as Gradient Boosting and AdaBoost have been shown to enhance predictive accuracy by sequentially correcting the errors of weak learners [11], [12]. These approaches have proven effective in handling high-dimensional data and capturing intricate patterns in financial datasets.

More recently, ensemble learning techniques have emerged as state-of-the-art approaches for improving model performance in loan approval prediction. Methods such as Voting Classifier and Stacking combine multiple base models to leverage their individual strengths, resulting in more robust and accurate predictions [13], [14]. These ensemble approaches have been successfully applied in various financial applications, demonstrating superior performance compared to single models, particularly in terms of accuracy and robustness [15]. In addition, ensemble methods are capable of handling non-linear relationships and reducing model variance, making them suitable for complex classification tasks.

Another important aspect in loan approval prediction is the issue of class imbalance, which is commonly observed in financial datasets. Imbalanced data can lead to biased models that favor the majority class, reducing the effectiveness of predictions for minority classes [16]. Several studies have highlighted the importance of using appropriate evaluation metrics such as F1-score and ROC-AUC to better assess model performance in such scenarios [17]. Furthermore, techniques such as resampling and cost-sensitive learning have been proposed to address imbalance, although not all studies incorporate these methods comprehensively [18].

In addition to predictive performance, model interpretability has become an important consideration in financial applications. Understanding the key factors that influence loan approval decisions is essential for transparency and regulatory compliance [19]. Feature importance analysis, particularly in tree-based models, has been widely used to identify influential variables such as credit score, income, and employment status [20]. These insights not only enhance model explainability but also provide valuable information for decision-makers in financial institutions.

Despite the significant advancements in machine learning for loan approval prediction, several research gaps remain. Many existing studies focus on a limited number of models or do not provide a comprehensive comparison between baseline and multiple ensemble techniques under a unified framework. In addition, the combined evaluation of performance metrics and interpretability is often lacking. Therefore, this study aims to address these gaps by providing a systematic comparison of baseline and ensemble models while also analyzing feature importance to support more transparent and effective decision-making.

Methodology

Research Framework

This study adopts a structured machine learning workflow to develop and evaluate predictive models for loan approval. The research process consists of several stages, including data collection, data preprocessing, feature transformation, model development, performance evaluation, and model interpretation. The overall research framework is illustrated in figure 1.

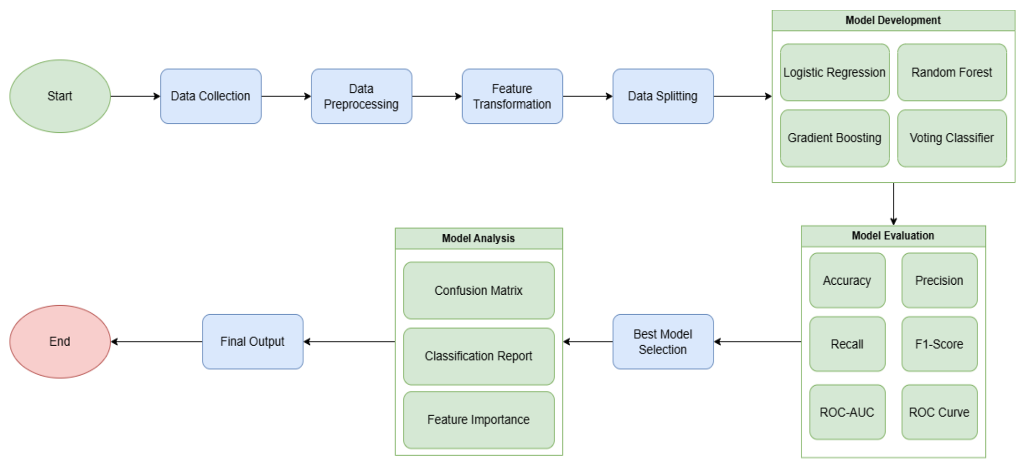

Figure 1 Research framework of the proposed study

Initially, the dataset is loaded and explored to understand its structure, distribution, and data quality. Subsequently, preprocessing techniques are applied to handle missing values and transform the data into a suitable format for machine learning algorithms. After preprocessing, multiple models, including both baseline and ensemble methods, are trained using a consistent pipeline. The models are then evaluated using multiple performance metrics to ensure a comprehensive comparison. Finally, the best-performing model is analyzed further through confusion matrix evaluation and feature importance analysis to provide interpretability and practical insights.

Dataset Description

The dataset used in this study contains borrower-related information that is relevant for predicting loan approval decisions, combining both financial and demographic attributes. It includes a total of nine input features, which are divided into numerical and categorical variables. The numerical features consist of Age, Income, LoanAmount, CreditScore, and YearsExperience, which represent the applicant's financial condition, creditworthiness, and professional background. These features are essential for assessing an individual's ability to repay a loan. In addition, the dataset includes categorical features such as Gender, Education, City, and EmploymentType, which provide contextual information about the applicant's socio-economic profile and employment stability. The target variable, LoanApproved, is a binary classification label where a value of 1 indicates that the loan application is approved and a value of 0 indicates that it is not approved. This combination of features enables the model to learn patterns that distinguish between approved and non-approved applicants based on both quantitative financial indicators and qualitative demographic factors.

Data Preprocessing

Data preprocessing was performed to ensure data quality and consistency. Missing values were handled using the deletion method by removing all rows containing null values. The dataset was then divided into numerical and categorical features. Numerical features were standardized using the following transformation:

$$x' = \frac{x - \mu}{\sigma} \quad (1)$$

X is the original value, μ is the mean, and σ is the standard deviation.

Categorical features were transformed using one-hot encoding, where each category is represented as a binary vector. All preprocessing steps were implemented using a unified pipeline to ensure consistency between training and testing data.

Data Splitting

The dataset was divided into training and testing sets using an 80:20 ratio to ensure a sufficient amount of data for both model learning and evaluation. Specifically, 80% of the data was used to train the models, allowing them to learn underlying patterns and relationships among features, while the remaining 20% was reserved for testing to evaluate model performance on unseen data. Stratified sampling was applied during the splitting process to preserve the original distribution of the target variable LoanApproved in both subsets. This is particularly important because the dataset is imbalanced, with a higher proportion of non-approved loan instances compared to approved ones. By maintaining this distribution, the models are trained and evaluated on data that accurately reflects real-world conditions, which helps prevent biased learning and ensures that performance metrics provide a reliable assessment of the models' generalization capability.

Model Development

This study implements both baseline and ensemble models. Logistic Regression models the probability of class membership using the sigmoid function:

$$P(y = 1 \mid x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)}} \tag{2}$$

Random Forest is an ensemble of decision trees constructed using bootstrap sampling and feature randomness. The final prediction is obtained by majority voting:

$$\hat{y} = mode(h_1(x), h_2(x), \dots, h_n(x)) \tag{3}$$

Gradient Boosting builds models sequentially by minimizing a loss function:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \tag{4}$$

$h_m(x)$is the weak learner and $\gamma_m$is the learning rate.

The Voting Classifier combines predictions from multiple models using soft voting:

$$\hat{y} = arg\ max \sum_{i=1}^{n} w_i P_i(y \mid x) \tag{5}$$

$P_i(y \mid x)$is the predicted probability from model $i$and $w_i$is the weight assigned to each model.

Model Evaluation

Model performance was evaluated using several metrics.

Accuracy is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

Precision is defined as:

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

Recall is defined as:

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

F1-score is defined as:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

ROC-AUC represents the area under the ROC curve, which evaluates the trade-off between true positive rate and false positive rate.

Best Model Analysis

The best-performing model was selected based on the highest ROC-AUC score, as this metric provides a robust evaluation of the model's ability to distinguish between approved and non-approved loan applications across different classification thresholds. Following the selection process, a detailed performance analysis was conducted using a confusion matrix, which provides a comprehensive breakdown of prediction outcomes in terms of true positives, true negatives, false positives, and false negatives. This analysis enables a deeper understanding of how well the model performs in correctly identifying both classes and highlights potential risks such as incorrectly approving ineligible applicants or rejecting eligible ones. In addition to performance evaluation, feature importance analysis was performed to identify the most influential variables contributing to the prediction. This step is essential for interpreting the model's decision-making process and understanding which features have the greatest impact on loan approval outcomes, thereby enhancing the transparency and practical relevance of the model in real-world financial applications.

Implementation Tools

The implementation of this study was carried out using the Python programming language due to its flexibility and extensive support for data science and machine learning tasks. Several well-established libraries were utilized to support different stages of the workflow. The panda's library was used for data loading, cleaning, and manipulation, allowing efficient handling of structured datasets. NumPy was employed for numerical computations and array operations, which are essential for efficient data processing. The scikit-learn library served as the core framework for building and evaluating machine learning models, providing tools for preprocessing, model training, pipeline construction, and performance evaluation. For visualization purposes, matplotlib and seaborn were used to generate informative plots such as bar charts, ROC curves, heatmaps, and feature importance graphs. Together, these libraries provide a comprehensive and integrated environment for developing, evaluating, and visualizing machine learning models in a reproducible and efficient manner.

Algorithm 1: Loan Approval Prediction Using Ensemble Learning

Input: Dataset Dcontaining features and target variable

Output: Best model M^*, prediction results, and feature importance

Process:

Start

Load dataset $D = \{(x_i, y_i)\}_{i=1}^{N}$

Remove missing values:

$$D' \leftarrow \{(x_i, y_i) \in D \mid x_i \ has\ no\ null\ values\}$$

Separate features and target:

$$X \leftarrow \{x_i\}, y \leftarrow \{y_i\}$$

Split features into numerical and categorical:

$$X = X_{num} \cup X_{cat}$$

Normalize numerical features:

$$x' = \frac{x - \mu}{\sigma}$$

Encode categorical features using one-hot encoding:

$$X'_{cat} \leftarrow OHE(X_{cat})$$

Combine transformed features:

$$X' \leftarrow [X'_{num}, X'_{cat}]$$

Split dataset into training and testing sets:

$$(X_{train}, X_{test}, y_{train}, y_{test}) \leftarrow StratifiedSplit(X', y)$$

Define model set:

$$M = \{Logistic\ Regression, Random\ Forest, Gradient\ Boosting, Voting\ Classifier\}$$

For each model $M_j \in M$:

Train model $M_j$using X_{train}, y_{train}

Predict $\hat{y}_j \leftarrow M_j(X_{test})$

Compute probability $p_j \leftarrow P(y = 1 \mid X_{test})$

Calculate evaluation metrics:

Accuracy $= \frac{TP+TN}{TP+TN+FP+FN}$

Precision $= \frac{TP}{TP+FP}$

Recall $= \frac{TP}{TP+FN}$

F1-score $= 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$

ROC-AUC $= AUC(y_{test}, p_j)$

End For

Select best model:

$$M^* = arg\ max\ (ROC - AUC)$$

Generate confusion matrix for M^*:

$$CM = [TN\ FP\ FN\ TP\ ]$$

Extract feature importance:

$$FI \leftarrow Importance(M^*)$$

Return M^*, predictions, evaluation results, and feature importance

End

Results

Dataset Overview

The dataset used in this study initially consisted of 5,000 instances with 10 features, including demographic, financial, and employment-related attributes. During the initial exploration, several features were found to contain missing values, particularly in Income (196 missing values), CreditScore (194 missing values), and Education (198 missing values). To ensure data quality and consistency, a deletion method was applied by removing all records containing missing values. As a result, the dataset was reduced to 4,430 instances, representing approximately 88.6% of the original data. Although this approach simplifies preprocessing by eliminating incomplete records, it may also lead to the loss of potentially useful information.

The target variable, LoanApproved, exhibits an imbalanced class distribution. After preprocessing, 3,377 instances, equivalent to approximately 76%, belong to the non-approved class, while 1,053 instances, or approximately 24%, belong to the approved class. This imbalance indicates that the dataset is skewed toward the majority class, which may influence model performance by biasing predictions toward non-approval outcomes. Therefore, evaluation metrics beyond accuracy, such as precision, recall, F1-score, and ROC-AUC, are necessary to provide a more reliable assessment of model performance in handling imbalanced data.

Model Performance Evaluation

The performance of both baseline and ensemble models was evaluated using multiple metrics, including Accuracy, Precision, Recall, F1-Score, and ROC-AUC, to ensure a comprehensive assessment of classification performance. The results, as presented in table 1, indicate that ensemble models consistently outperform the baseline Logistic Regression model across all evaluation metrics, demonstrating their ability to capture more complex patterns in the data. Among the evaluated models, Random Forest achieved the highest ROC-AUC score of 0.9515, indicating superior capability in distinguishing between approved and non-approved loan applications. In addition, the Voting Classifier achieved the highest F1-Score of 0.9220, reflecting the best balance between precision and recall, which is particularly important in imbalanced classification problems. These findings highlight the effectiveness of ensemble learning methods in improving predictive performance compared to traditional single models.

Table 1 Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Random Forest	0.9628	0.9495	0.8910	0.9193	0.9515
Voting Classifier	0.9639	0.9497	0.8957	0.9220	0.9462
Gradient Boosting	0.9628	0.9450	0.8957	0.9197	0.9447
Logistic Regression	0.8939	0.7773	0.7773	0.7773	0.9210

ROC Curve Analysis

The ROC curve comparison is illustrated in figure 2, which provides a visual evaluation of each model's ability to distinguish between approved and non-approved loan applications across different classification thresholds. From figure 2, it can be observed that the Random Forest model achieves the highest ROC-AUC value, indicating the strongest discriminative performance among all models. The Voting Classifier and

Gradient Boosting models also demonstrate high performance, with ROC curves that closely follow that of Random Forest, suggesting that they are similarly effective in capturing underlying patterns in the data. In contrast, Logistic Regression exhibits a comparatively lower ROC curve, reflecting its limited ability to model complex, non-linear relationships. Overall, the results confirm that ensemble methods provide superior classification capability and more reliable predictive performance compared to the baseline model.

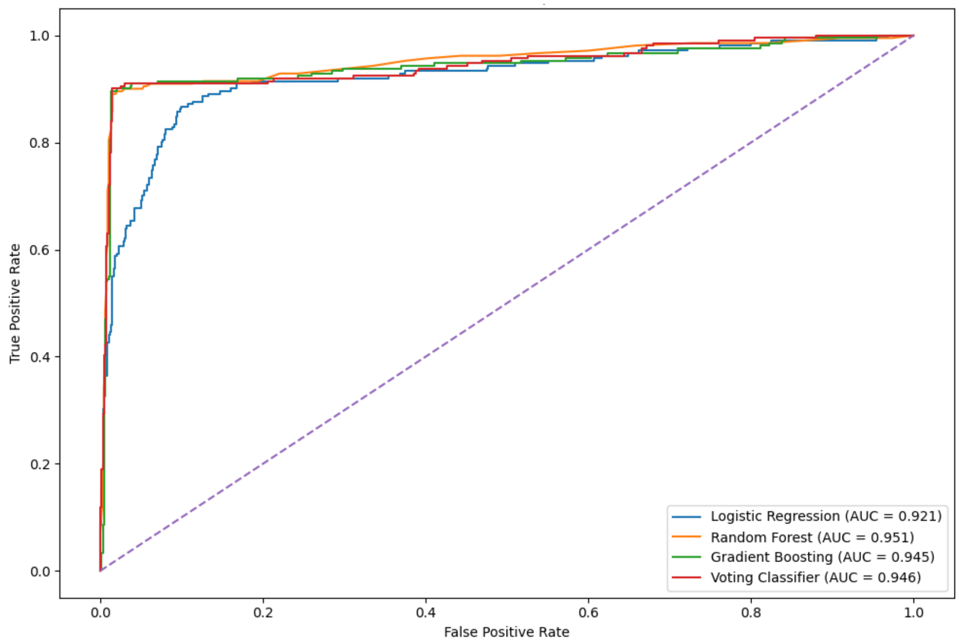

Figure 2 **ROC curve comparison of all models**

Best Model Performance (Random Forest)

Based on the highest ROC-AUC score, Random Forest was selected as the best-performing model, demonstrating superior ability in distinguishing between approved and non-approved loan applications. The confusion matrix of the Random Forest model, as shown in figure 3, provides a detailed breakdown of classification outcomes. The model correctly classified 665 instances as true negatives and 188 instances as true positives, while producing only 10 false positives and 23 false negatives. These results indicate that the model achieves a high level of accuracy and maintains a strong balance between correctly identifying approved and non-approved cases. The relatively low number of false positives suggests that the model is effective in minimizing the risk of approving ineligible applicants, which is critical in financial decision-making. At the same time, the low number of false negatives indicates that only a small portion of eligible applicants are incorrectly rejected. Overall, this performance demonstrates that the Random Forest model is reliable for loan approval prediction and is capable of supporting accurate and risk-aware decision processes.

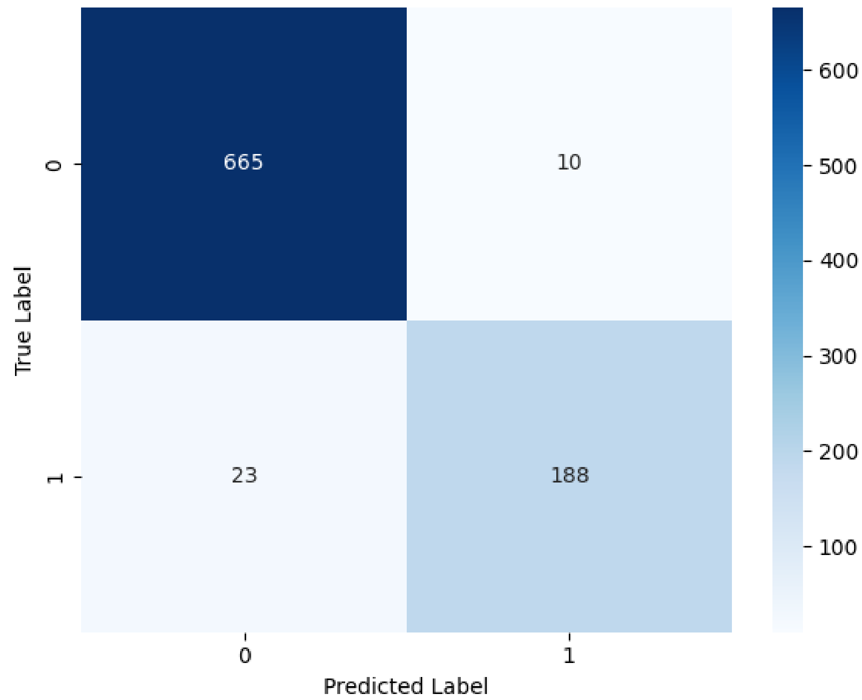

Figure 3 **Confusion matrix of the Random Forest model**

Feature Importance Analysis

Feature importance derived from the Random Forest model is illustrated in figure 4, providing insights into the relative contribution of each feature in the prediction process. As shown in figure 4, CreditScore is the most influential feature with a significantly higher importance value compared to all other variables, indicating that it plays a dominant role in determining loan approval outcomes. This is followed by Income, which reflects the applicant's financial capacity, and Employment Type, which represents job stability and reliability of income sources. Other features such as LoanAmount, Age, and YearsExperience contribute to the model to a lesser extent but still provide meaningful information in the decision-making process. The relatively lower importance of categorical features such as City and Education suggests that demographic factors have less influence compared to financial attributes. Overall, these findings indicate that financial credibility and income stability are the primary determinants in loan approval decisions, aligning with real-world lending practices where risk assessment is heavily based on an applicant's ability to repay the loan.

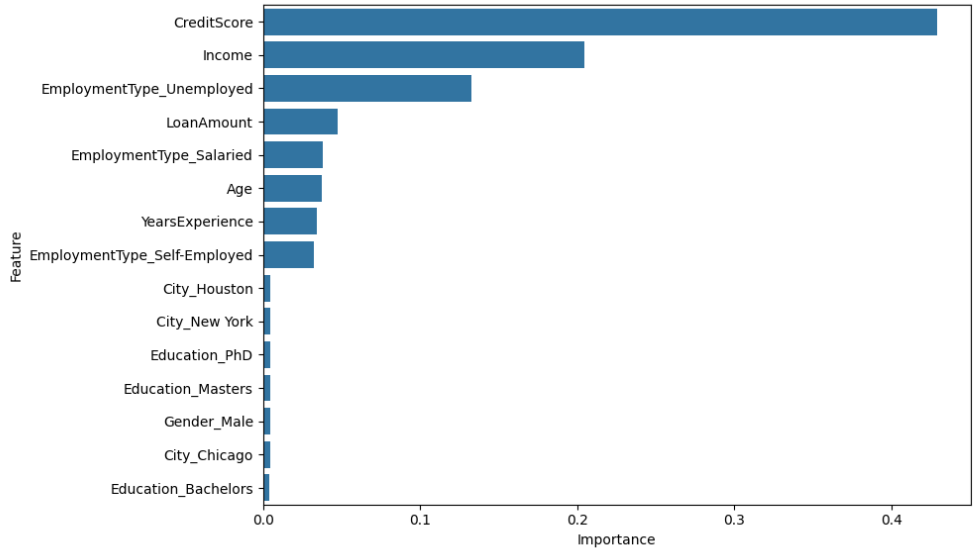

Figure 4 Top 15 feature importances from Random Forest

Discussion

The results of this study clearly demonstrate that ensemble learning methods provide substantial improvements in loan approval prediction compared to the baseline Logistic Regression model. The superior performance of Random Forest, Gradient Boosting, and Voting Classifier across all evaluation metrics indicates their ability to model complex and non-linear relationships within the dataset. This is particularly important in financial data, where interactions between variables such as credit score, income, and employment status are often not linear. Random Forest achieved the highest ROC-AUC, indicating strong discriminative capability, while the Voting Classifier achieved the highest F1-score, reflecting a better balance between precision and recall. These findings suggest that ensemble approaches not only improve predictive accuracy but also provide more stable and reliable performance when dealing with imbalanced classification problems. In practical applications, this means that ensemble models are more suitable for decision support systems in loan approval processes compared to traditional single models.

Further analysis of the confusion matrix and feature importance provides deeper insights into model behavior and decision-making. The Random Forest model produced a low number of false positives and false negatives, indicating its effectiveness in minimizing both financial risk and missed opportunities. This balance is crucial in real-world lending scenarios, where incorrect approvals can lead to financial loss and incorrect rejections can reduce customer acquisition. Additionally, feature importance analysis revealed that CreditScore, Income, and Employment Type are the most influential variables in determining loan approval outcomes. This aligns with established financial principles, where creditworthiness and income stability are key indicators of repayment ability. However, the study also has limitations, particularly the use of the deletion method for handling missing values, which reduces dataset size and may remove useful information. The presence of class imbalance may also influence model performance, although multiple evaluation metrics were used to provide a more comprehensive assessment. Future research can address these limitations by applying advanced preprocessing techniques, incorporating additional ensemble methods, and performing hyperparameter optimization to further enhance model performance and generalizability.

Conclusion

This study focused on enhancing loan approval prediction by applying ensemble machine learning techniques and comparing their performance with a baseline Logistic Regression model. The results demonstrate that ensemble models, including Random Forest, Gradient Boosting, and Voting Classifier, consistently achieve superior performance across multiple evaluation metrics, indicating their effectiveness in capturing complex patterns within the dataset. Random Forest achieved the highest ROC-AUC, reflecting strong discriminative ability in distinguishing between approved and non-approved loan applications, while the Voting Classifier provided the best balance between precision and recall, as indicated by the highest F1-score. In addition, the analysis of feature importance revealed that CreditScore, Income, and Employment Type are the most influential factors in the decision-making process, emphasizing the importance of financial credibility and income stability in loan approval. The low number of misclassifications produced by the best-performing model further confirms its reliability for practical applications in financial decision support systems. Despite these promising results, the study is limited by the use of a deletion method for handling missing values and the presence of class imbalance, which may affect model generalizability. Therefore, future research is recommended to explore more advanced preprocessing techniques, incorporate additional ensemble algorithms, and apply hyperparameter optimization to further improve predictive performance and robustness.

Declarations

Author Contributions

Conceptualization: Q.T.; Methodology: Q.T.; Software: Q.T.; Validation: Q.T.; Formal Analysis: Q.T.; Investigation: Q.T.; Resources: Q.T.; Data Curation: Q.T.; Writing Original Draft Preparation: Q.T.; Writing Review and Editing: Q.T.; Visualization: Q.T.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009.

[2] D. J. Hand and W. E. Henley, "Statistical classification methods in consumer credit scoring: A review," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 160, no. 3, pp. 523–541, Sep. 1997, doi: 10.1111/j.1467-985X.1997.00078.x.

[3] S. Lessmann, B. Baesens, H.-V. Seow, and L. C. Thomas, "Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research," *European Journal of Operational Research*, vol. 247, no. 1, pp. 124–136, Nov. 2015, doi: 10.1016/j.ejor.2015.05.030.

[4] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[5] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001, doi: 10.1214/aos/1013203451.

[6] H. Liu, "Feature selection," in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds. Boston, MA, USA: Springer, 2011, doi: 10.1007/978-0-387-30164-8_306.

[7] N. M. Nasrabadi, "Pattern recognition and machine learning," *Journal of Electronic Imaging*, vol. 16, no. 4, p. 049901, Oct. 2007, doi: 10.1117/1.2819119.

[8] N. Soltani Halvaiee and M. K. Akbari, "A novel model for credit card fraud detection using artificial immune systems," *Applied Soft Computing*, vol. 24, no. November, pp. 40–49, Nov. 2014, doi: 10.1016/j.asoc.2014.06.042.

[9] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, vol. 2016, no. August, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.

[10] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, Mar. 1986, doi: 10.1007/BF00116251.

[11] H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.

[12] J. Davis and M. Goadrich, “The relationship between precision-recall and ROC curves,” in *Proc. 23rd Int. Conf. Machine Learning (ICML)*, vol. 2006, no. June, pp. 233–240, 2006, doi: 10.1145/1143844.1143874.

[13] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” *Journal of Artificial Intelligence Research*, vol. 16, no. June, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[14] A. K. Sharma, R. K. Dubey, P. Goswami, and N. Dandotiya, “Transforming credit risk management with machine learning in the financial sector,” in *Proc. 2024 4th Int. Conf. Advancement in Electronics & Communication Engineering (AECE)*, Ghaziabad, India, vol. 2024, no. March, pp. 1323–1328, 2024, doi: 10.1109/AECE62803.2024.10911170.

[15] S. Finlay, *Credit Scoring, Response Modeling, and Insurance Rating: A Practical Guide to Forecasting Consumer Behavior*. London, U.K.: Palgrave Macmillan, 2012.

[16] S. Lessmann, B. Baesens, H.-V. Seow, and L. C. Thomas, “Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research,” *European Journal of Operational Research*, vol. 247, no. 1, pp. 124–136, Nov. 2015, doi: 10.1016/j.ejor.2015.05.030.

[17] M. H. Ismail, T. R. Razak, N. M. Noor, and A. A. Aziz, “Comparative analysis of machine learning algorithms through visualizations,” in *Proc. 2023 IEEE 8th Int. Conf. Recent Advances and Innovations in Engineering (ICRAIE)*, Kuala Lumpur, Malaysia, vol. 2023, no. March, pp. 1–5, 2023, doi: 10.1109/ICRAIE59459.2023.10468098.

[18] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, “A study of the behavior of several methods for balancing machine learning training data,” *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, pp. 20–29, Jun. 2004, doi: 10.1145/1007730.1007735.

[19] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” *arXiv preprint arXiv:1705.07874*, vol. 2017, no. May, pp. 1-10, 2017, doi: 10.48550/arXiv.1705.07874.

[20] J. Brownlee, *Machine Learning Mastery With Python: Understand Your Data, Create Accurate Models, and Work Projects End-to-End*. 2016.