Journal of Current Research and Studies

Volume 02 Issue 05 Sep-Oct 2025
https://journalcurrentresearch.com/
ISSN:3048-9911

The Use of Machine Learning Adoption in Loan Approval Prediction System

Olowookere A.S.*, Kilani S.O & Bello G.R.
Department of Computer Science, Oyo State College of Agriculture and Technology, Igboora, **Nigeria**

OPEN ACCESS

***Correspondence**
Olowookere A.S.

Article Received
12/09/2025
Accepted
20/09/2025
Published
26/09/2025

Works Cited

Olowookere A.S., Kilani S.O & Bello G.R., (2025). The Use of Machine Learning Adoption in Loan Approval Prediction System. *Journal of Current Research and Studies*, 2(5), 90-99.

Abstract

Recently, financial institutions have increasingly relied on automated systems to streamline loan approval processes. This study developed a Loan Approval Prediction System using machine learning techniques to evaluate applicant eligibility based on historical loan data. The dataset, sourced from Kaggle, included key attributes such as credit score, income, employment status, loan amount, and debt-to-income ratio. The methodology followed a systematic approach consisting of data preprocessing, exploratory data analysis (EDA), feature selection, target balancing, and model evaluation. Preprocessing involved dropping irrelevant columns, handling missing values, and encoding categorical data, followed by correlation analysis to guide feature selection. To enhance predictive performance, three feature selection techniques—filter, wrapper, and hybrid—were compared using a Support Vector Classifier (SVC). Model evaluation employed accuracy, precision, recall, and F1-score metrics. Results revealed that the hybrid selection method, when combined with Synthetic Minority Over-sampling Technique (SMOTE) and SVC, achieved the highest accuracy (94.93%) and F1-score (95.91%). Observations indicated that applicant income, credit score, and debt-to-income ratio were the most significant predictors of loan approval. The study concludes that the hybrid-SVC model provides an efficient, unbiased, and highly accurate loan approval prediction framework, reducing processing time and decision errors while enhancing transparency and customer satisfaction.

Keywords

Loan Approval, Prediction System, Machine Learning, Feature Selection, Support Vector Classifier

Introduction

The financial industry has experienced shift with the rise of digital technologies, particularly in the domain of loan management and approval. Historically, loan approval was labour-intensive process that required human agents to assess each application based on various factors like income, credit score, and employment history. These assessments were often prone to human error and bias, resulting in inefficiencies and inconsistencies (Jagtiani and Lemieux, 2019). As financial institutions seek to improve accuracy and speed, machine learning (ML) has emerged as a transformative tool in automating loan approval processes, allowing for rapid decision-making and reduced operational costs.

Machine learning algorithms offer a data-driven approach to analyzing vast amounts of historical data to identify patterns and predict outcomes. Unlike traditional credit scoring methods, which rely on static financial indicators, ML models can account for a wider range of data points, including real-time information, which significantly enhances predictive capabilities Moro et al., 2015). By utilizing machine learning, financial institutions can save time taken to evaluate loan applications, improving both customer satisfaction and operational efficiency. The traditional loan approval process is often slow, error-prone, and subject to human bias. Banks and lending institutions rely more on human judgment for loan applications, which can result in inconsistent decisions, leading to the approval of risky loans or denial of creditworthy applicants. Moreover, manual processed large loan applications are inefficient, especially with the growing demand for quick financial solutions. As a result, delays and inconsistencies in loan approval processes negatively impacted both financial institutions and customers, causing revenue losses, missed opportunities, and customer dissatisfaction. This study seeks to address the issue of implementing a machine learning-loan approval prediction system to automate the decision-making process, improve accuracy, and reduce processing

The primary aim of this study is to develop a machine learning model that can predict loan approval outcomes based on applicant data, thereby automating the loan evaluation process.

Several research studies and projects have explored the application of machine learning in loan approval systems, highlighting the growing relevance of AI and data analytics in the financial sector. One of the most commonly used algorithms in loan approval systems is the Decision Tree. A study by Khandani et al., (2021) demonstrated the effectiveness of Decision Trees in improving the accuracy of loan approval decisions. The researchers used a dataset from a major financial institution and applied a Decision Tree algorithm to predict loan default rates. Their model achieved high accuracy and demonstrated interpretability, which is critical for taken loan approval decisions to both regulators and borrowers. However, the study also noted that risk of overfitting when the tree becomes too complex, a challenge addressed by using techniques like Random Forest or pruning methods. Building on the strengths of Decision Trees, Random Forest algorithms have been applied to loan approval systems to reduce overfitting and improve predictive accuracy. Similarly, Zhang et al., (2020) developed a Random Forest-based model for predicting loan approvals using a dataset of personal loan applications. This model outperformed traditional logistic regression models in terms of accuracy, providing better risk assessment and reducing the likelihood of loan defaults. The study then concluded that ensemble learning approach of Random Forests would allowed for more robust predictions across diverse datasets, although the model required significant computational resources. Neural networks have also gained prominence in loan approval systems due to their ability to capture complex, non-linear relationships in data. Also, Das et al., (2020) used a deep learning approach to predict loan approvals based on a large dataset of financial and demographic data. The neural network model significantly improved the prediction accuracy compared to traditional methods, particularly when dealing with large-scale datasets. However, the researchers acknowledged that neural networks lack interpretability, making it difficult to explain the model's decisions, which is a key concern in regulated industries like finance. Gradient Boosting Machines (GBM), particularly XGBoost, have shown remarkable success in improving the accuracy of loan approval systems. Sun et al., (2019) implemented an XGBoost-based model to predict loan defaults using a dataset from a microfinance institution. The model outperformed other machine learning algorithms such as Logistic Regression and Support Vector Machines, particularly in terms of minimizing false positives. XGBoost's ability to handle missing data and robustness in reducing bias and variance made a powerful tool for improving the loan approval process. However, the model required careful tuning of hyperparameters to avoid overfitting.

Methodology

Data Collection

For this study, data were collected from Kaggle, a popular platform for data science competitions and projects. The dataset comprises various attributes relevant to loan applications, including but not limited to applicant demographics, financial history, loan amount requested, interest rates, and repayment history. The dataset also contains records of both approved and rejected loans, providing a balanced view of the factors influencing loan decisions. The choice of

Kaggle as a source ensures that the dataset is publicly available, well-documented, and suitable for applying machine learning techniques, towards enhancing the validity of the research.

Description of the Current System

The current loan approval prediction system utilizes a machine learning approach, with a particular focus on Support Vector Machine (SVM) as the primary algorithm. SVM is a supervised learning model that is effective for binary classification tasks. It operates by finding the hyperplane which best separates data points of different classes in a high-dimensional feature space. During loan approval, SVM will effectively classify applicants as either creditworthy or non-creditworthy based on their features. To enhance the model's performance, various feature selection methods were employed. These includes filter feature selection, wrapper feature selection, and a hybrid approach that combines both methods. Filter feature selection evaluates the relevance of features based on statistical measures, while wrapper feature selection assesses the performance of a subset of features using the model itself. By integrating these approaches, the system will identify the most significant predictors of loan approval, leading to improved accuracy and reduced overfitting.

Furthermore, the implementation of hybrid feature selection allows the model to consider multiple dimensions of borrower data, enhancing its predictive power. By leveraging these methods, the system aims to provide a more robust and accurate assessment of creditworthiness, addressing the limitations of traditional models. For this study, the SVM model was trained on the pre-processed dataset, utilizing the selected features from the hybrid feature selection process. The model's performance was evaluated using metrics such as accuracy, precision, recall, and F1-score towards ensuring its effectiveness in predicting loan approvals.

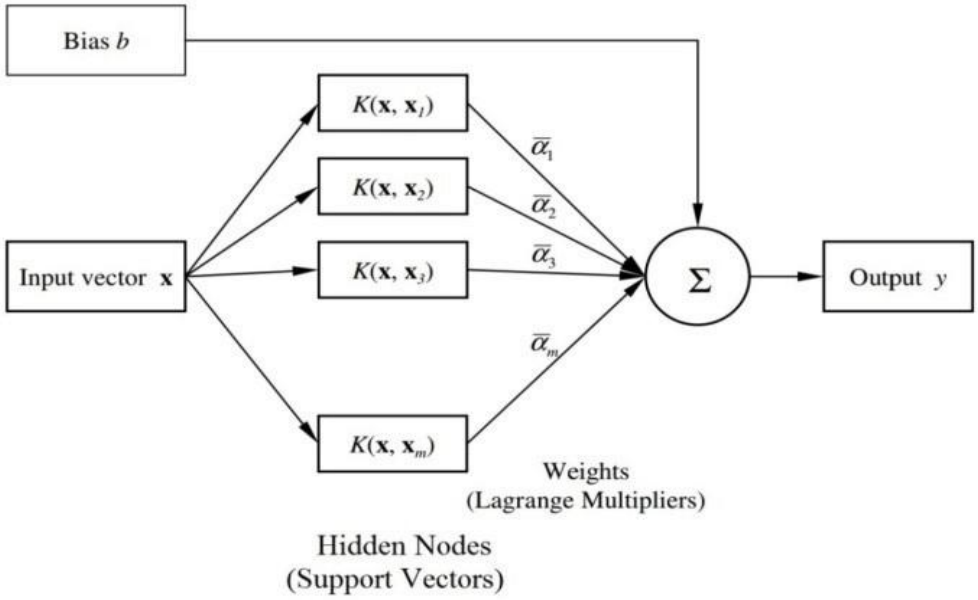


Fig 3.1. SVM Architecture Diagram

Current System Model

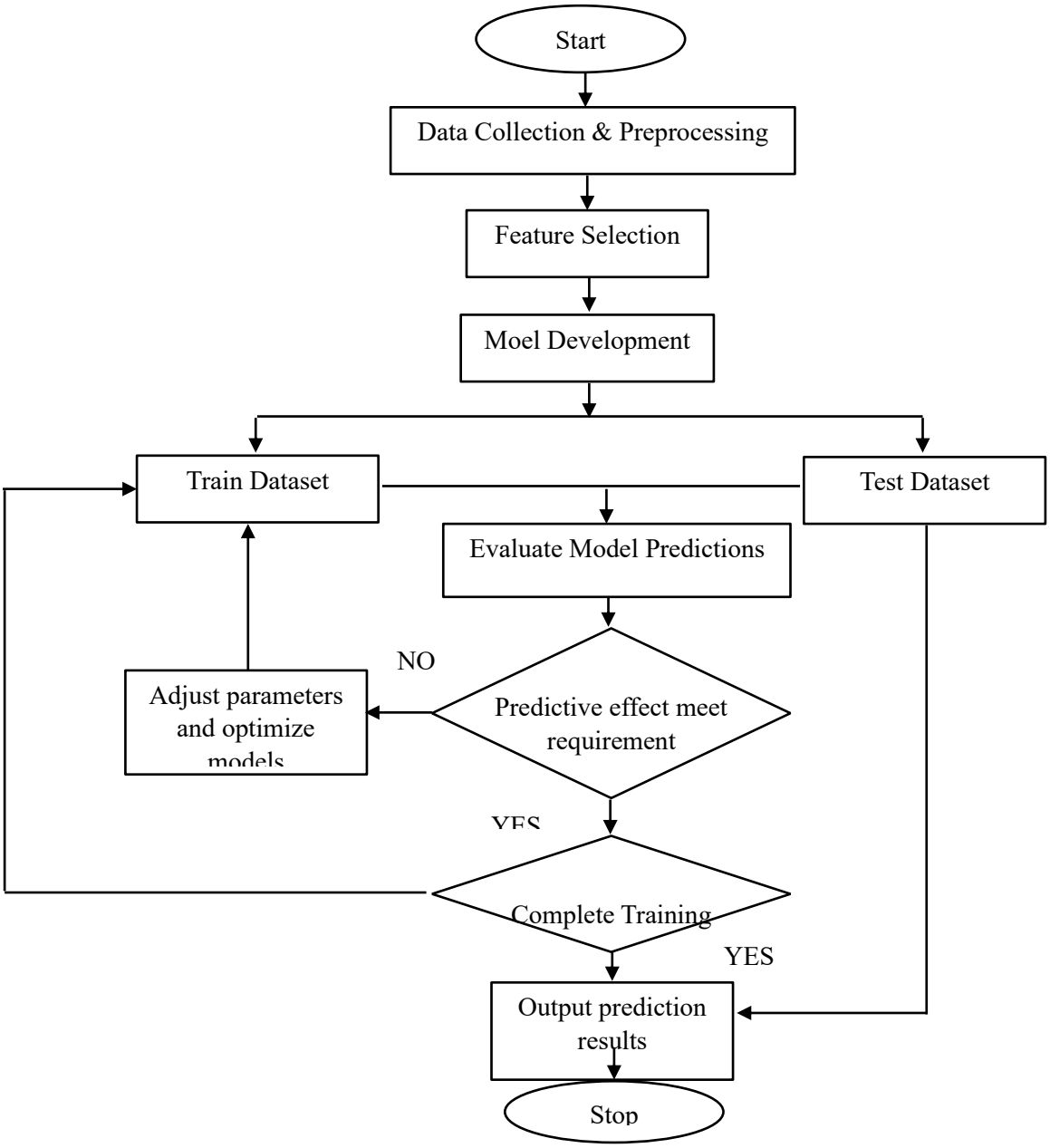
The current system model is designed to provide a comprehensive solution for predicting loan approvals using machine learning techniques. The workflow consists of several key stages:

1. Data Preprocessing: This initial step involves cleaning and preparing the dataset for analysis. It includes handling missing values, normalizing features, and applying SMOTE to balance the dataset.
2. Feature Selection: Utilizing filter, wrapper, and hybrid feature selection methods, the most relevant features are identified to enhance the model's accuracy and reduce complexity.
3. Model Training: The selected features are used to train the SVM model, optimizing parameters to achieve the best classification performance.

4. Model Evaluation: The trained model is evaluated using a separate testing dataset. Performance metrics such as accuracy, precision, recall, and F1-score are calculated to assess its predictive capabilities.
5. Prediction: The final model is deployed for real-time predictions on new loan applications, providing lenders with a data-driven assessment of creditworthiness.
6. Feedback Loop: Continuous monitoring and evaluation of model performance are conducted to identify areas for improvement, enabling the system to adapt to changing borrower behaviours and economic conditions.

This structured approach ensures that the loan approval prediction system is not only accurate but with transparent and fair, addressing the critical issues associated with traditional credit assessment methods.

Flowchart



Result and Discussion

The process, began with data loading and proceeded through various phases, was designed to enhance model accuracy and ensure predictive robustness. Key stages involved included data preprocessing, exploratory data analysis (EDA), feature selection, target balancing, and model evaluation using a Support Vector Classifier (SVC). Each phase contributed uniquely to refining the model, select the best feature selection technique for optimal performance.

Data Loading and Preprocessing

The dataset, sourced from Kaggle, contained various attributes related to loan applicants and their eligibility for loan approval. Initial steps focused on preparing the data for analysis by addressing inconsistencies and enhancing data quality.

1. **Dropping Irrelevant Columns:**
   Columns 'loan_id' and 'no_of_dependents' were dropped. These variables were deemed unnecessary for the prediction as they did not directly influence loan approval status and could introduce noise into the model.
2. **Handling Missing and Duplicated Values:**
   A thorough inspection of the dataset revealed missing and duplicate entries, which were addressed to ensure a clean dataset. Missing values were either filled or removed based on the impact of the data point on overall trends and model interpretability.
3. **Cleaning Column Names:**
   Whitespaces in column names were removed to improve accessibility and streamline coding processes, ensuring error-free interactions with dataset attributes during analysis.

These preprocessing steps laid a strong foundation for data quality, setting the stage for effective exploratory data analysis and feature engineering.

Exploratory Data Analysis (EDA)

With a clean dataset in place, exploratory data analysis was conducted to uncover patterns and insights within the data. The EDA phase involved:

1. **Target Variable Distribution:**
   The target variable, Loan_ Status, which indicated loan approval or rejection, was analyzed for class distribution. Results showed a distinct imbalance, with 62.2% of loans approved and 37.8% rejected. This imbalance highlighted the need for further steps to prevent bias in model training.

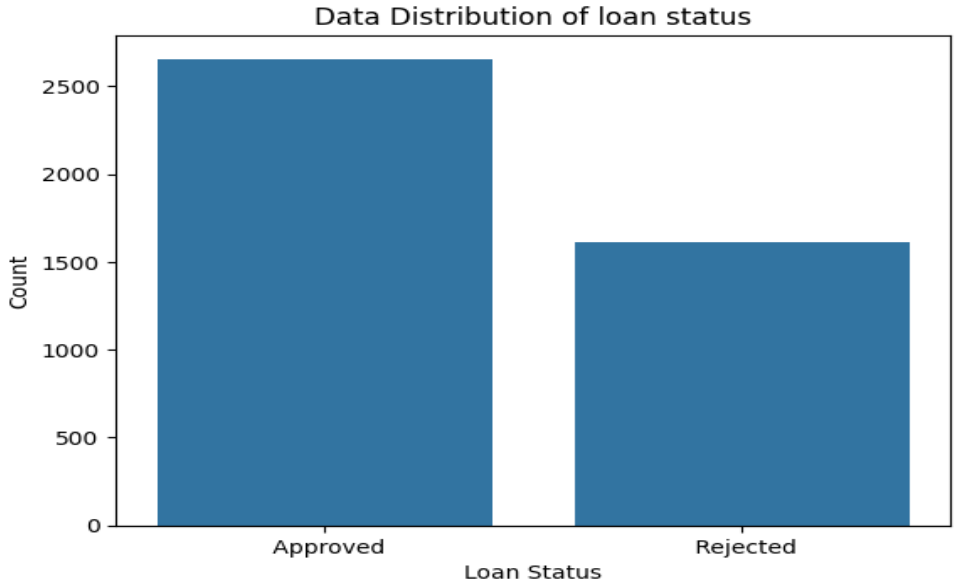


Fig 1. Loan Status Data distribution in Barchart

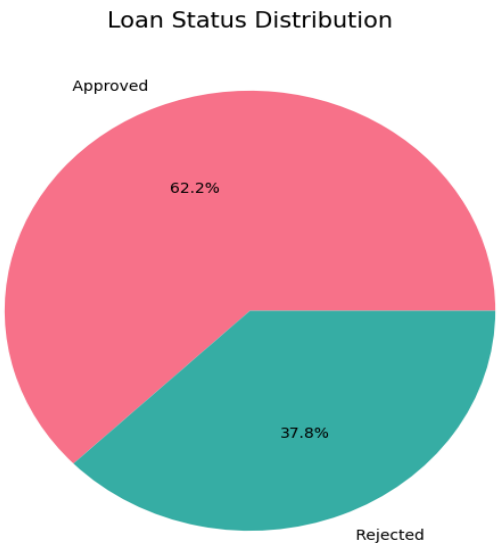


Fig 2. Loan Status Distribution in Pie chart

Data Encoding and Correlation Analysis

Since dataset contained categorical variables, data encoding was required to convert these into a format compatible with machine learning models. Encoding methods, such as one-hot encoding, were employed to ensure that the categorical data could be used effectively during training.

Correlation Analysis:

Correlation analysis was performed on the dataset to understand the relationships between features. This analysis provides a clearer picture of how different features interacted and influenced one another, guiding towards optimal feature selection. Highly correlated features with the targetted variable indicated potential predictors of loan approval status, which would be essential for model accuracy.

Fig 4.3. Data Correlation

Feature Selection Techniques

With data encoded and correlations analyzed, the next step was the selection of the most predictive features. Selecting the right features is crucial in enhancing model accuracy and reducing overfitting. Hence, three distinct feature selection methods were explored:

1. Filter Selection: This method evaluates each feature separately based on its statistical relevance to the target variable, selecting features that ranked high in relation to the target.
2. Wrapper Selection: Using a more exhaustive approach, wrapper selection iteratively tested feature combinations, identifying separately, t, together, yielding the highest model performance.
3. Hybrid Selection: Hybrid selection combines elements of both filter and wrapper methods, allowing the capitalizing on the efficiency of filter methods while benefiting from the deeper analysis offered by wrapper methods.

Model Evaluation with Feature Selection

To evaluate each feature selection method, the Support Vector Classifier (SVC) was employed as predictive model. After training the model with each selection method, its performance was assessed using several metrics:

Filter Selection Results

The SVC model trained features was selected through the filter method that gave the following metrics results:

i. Accuracy: 92.27%
ii. F1 Score: 93.67%
iii. Precision Score: 97.09%

Label	Precision	Recall	F1-Score	Support
0 (Rejected)	0.85	0.95	0.90	471
1 (Approved)	0.97	0.90	0.94	810

Filter selection demonstrated strong performance, particularly in precision for the approved loans (class 1), indicating its potential to capture positive outcomes accurately.

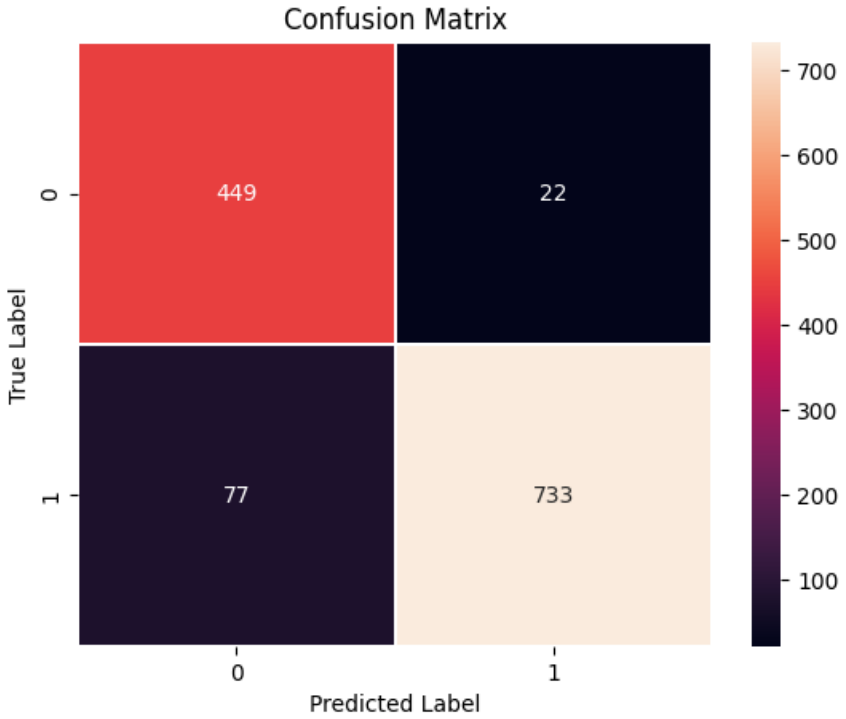


Fig 3. Confusion matrix for filter selection

Wrapper Selection Results: The SVC model trained features selected through the wrapper method produce the following performance metrics:

i. Accuracy: 94.22%
ii. F1 Score: 95.33%
iii. Precision Score: 97.55%

Label	Precision	Recall	F1-Score	Support
0 (Rejected)	0.89	0.96	0.92	471
1 (Approved)	0.98	0.93	0.95	810

The wrapper method showed an improvement in accuracy and F1 score over the filter method. This improvement suggested that analyzing feature combinations helped the model capture more relevant patterns for accurate prediction.

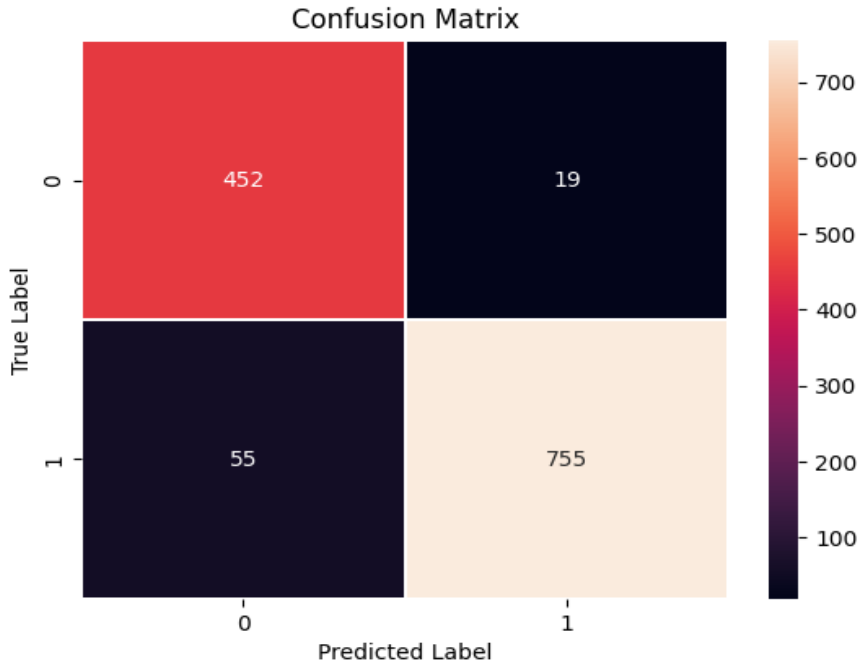


Fig 5. Confusion matrix for wrapper selection

Hybrid Selection Results: The hybrid feature selection method produced the highest performance metrics among the three methods considered:

i. Accuracy: 94.93%
ii. F1 Score: 95.91%
iii. Precision Score: 97.82%

Label	Precision	Recall	F1-Score	Support
0 (Rejected)	0.90	0.96	0.93	471
1 (Approved)	0.98	0.94	0.96	810

The hybrid approach, which integrated the strengths of both filter and wrapper methods, produced the best results, indicating it has the best capture relationships relevant to loan approval with minimal noise.

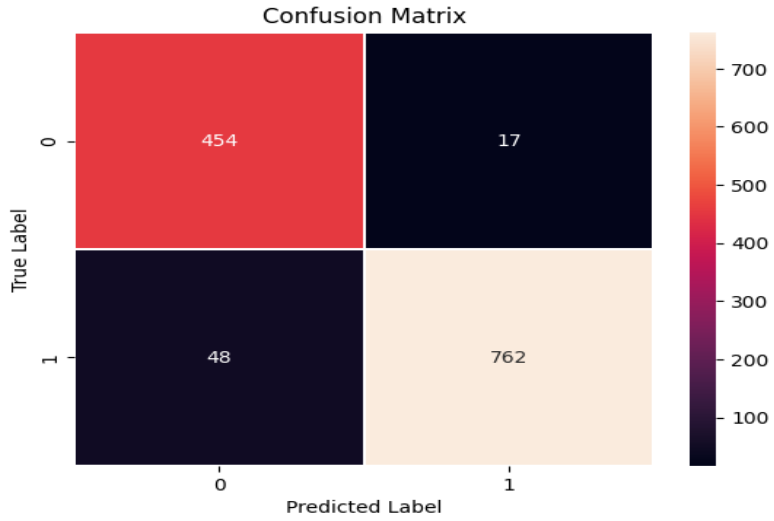


Fig 6. Confusion matrix for hybrid selection

Target Balancing and Model Training

Since the target variable was imbalanced, Synthetic Minority Over-sampling Technique (SMOTE) was applied to balance the classes. This technique help prevent model bias toward the majority class (approved loans), ensuring fairer performance. Training on the balanced dataset led to more reliable predictive metrics across classes.

Final Model Selection

After comparing the three feature selection methods, the Hybrid Selection approach was chosen for final deployment. With the highest accuracy (94.93%) and F1 score (95.91%), the hybrid method demonstrated its effectiveness in selecting features that enhanced the predictive capability of the SVC model. This result underscores the advantage of a balanced approach that combines statistical relevance and feature combination insights for predictive success in loan approval models. In conclusion, the results indicated that the hybrid feature selection method, along with SMOTE and SVC, provided a highly accurate model for loan approval prediction, positioning it as a reliable solution for lending institutions looking to optimize decision-making in loan assessments.

Conclusion

The study successfully met its aim of developing a machine learning model that integrate data preprocessing, SMOTE balancing, and optimal feature selection through the hybrid method, the model minimized bias and improved predictive performance.

The key conclusions from the research are as follows:

i. Feature Selection: Hybrid selection outperformed other methods, yielding an accuracy of 94.93% and an F1 score of 95.91%, making it the best approach for selecting features relevant to loan status.
ii. SVM Model Performance: SVM, with its capacity for effective classification, proved to be a strong choice for loan prediction, especially when paired with the hybrid-selected features.
iii. Reduction of Bias and Error: The model's automated nature helps reduce subjective bias and inconsistencies inherent in manual loan assessments, thus promoting objective decision-making.
iv. Prototype Viability: The prototype system shows potential for real-world application, highlighting the () practical-ability of a machine learning-driven solution for loan evaluation. The system's predictive accuracy can assist lending institutions in making faster, more reliable loan decisions.

This study can be a valuable tool for loan prediction and advances the capabilities of machine learning in financial services, aligning with the objectives of reducing human intervention, bias, and delays in loan processing.

References

1) Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. Heliyon, 4(11), e00938. https://doi.org/10.1016/j.heliyon.2018.e00938
2) Agarwal, S., Ben-David, I., & Yao, V. (2021). Mortgage refinancing, consumer credit, and competition: Evidence from the U.S. housing market. The Review of Financial Studies, 34(7), 3287-3330.
3) Barocas, S., Hardt, M., & Narayanan, A. (2021). Fairness and machine learning. MIT Press.
4) Bhardwaj, P., & Patil, R. (2020). Machine learning in financial services: Enhancing credit scoring and decision-making. International Journal of Data Science and Analysis, 5(2), 81-90.
5) Chen, J., Zhang, Y., & Hossain, M. (2021). Unsupervised machine learning techniques in fraud detection and loan approval systems: A review. Journal of Financial Data Science, 4(1), 98-115.
6) Das, S., Mahapatra, S., & Behera, B. (2020). Deep learning models for financial data analysis and loan approval prediction. Journal of Applied Data Science, 8(2), 98-115.
7) Jagtiani, J., & Lemieux, C. (2019). The roles of big data and machine learning in bank supervision. The Federal Reserve Bank of Philadelphia Working Paper, 19-22. https://doi.org/10.21799/frbp.wp.2019.22
8) Khandani, A. E., Kim, A. J., & Lo, A. W. (2021). Consumer credit-risk models via machine-learning algorithms. Journal of Banking and Finance, 36(5), 2767-2787.
9) Lutfi, A., Suharto, S., & Abdurrahman, M. (2022). The evolution of credit risk management in the financial sector: A literature review. Journal of Finance and Risk Perspectives, 18(2), 55-68.
10) Moro, S., Cortez, P., & Rita, P. (2015). A data-driven approach to predict the success of bank telemarketing. Decision Support Systems, 62, 22-31. https://doi.org/10.1016/j.dss.2014.03.001
11) Sun, Y., Zhou, Y., & Liu, M. (2019). Application of XGBoost algorithm in microfinance loan approval systems. Computational Finance Review, 18(3), 88-101.
12) Tobback, E., Bellotti, T., &Moeyersoms, J. (2019). The impact of alternative data on credit risk modeling and loan approval. Journal of Data Science, 17(2), 112-128.
13) Zhang, Y., & Hossain, M. S. (2020). Smart bank loan prediction using machine learning. In 2020 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 1187-1191). IEEE. https://doi.org/10.1109/CSCI51800.2020.00222
14) Zhang, Y., Hossain, M., & Tahmid, M. (2020). Big data analytics in fintech: A review of credit risk modeling and loan approval prediction. Journal of Financial Data Science, 3(1), 123-142.

Note: most of the references listed are not cited, try and recheck and ensure that only those cited were listed or vice versal, Also the author fail to acknowledge the source of Data description, Deployment configuration source, Model developer sources, SVC, SVM, and Kaggle plat form. All these models' sources needed to be acknowledged.