JISKA (Jurnal Informatika Sunan Kalijaga)
Vol. 11, No. 2, MAY, 2026, Pp. 225 – 236
ISSN: 2527 – 5836 (print) | 2528 – 0074 (online)

Optimizing Financial Risk Prediction for Loan Approval Decisions

Khalifatur Rauf [(1)*], Adi Cahyo Kuswijayanto [(2)], Ella Kristiantini Susan [(3)]
Department of Information Technology, Universitas Teknologi Yogyakarta, Yogyakarta, Indonesia
e-mail : {blograuf.kr,asasuke77adi,sekarlangit02}@gmail.com.
* Corresponding author.
This article was submitted on 23 January 2026, revised on 15 April 2026, accepted on 15 April 2025, and published on 25 May 2026.

Abstract

Accurate financial risk prediction is essential for effective loan approval decision-making, particularly in data-driven financial systems. This study investigates the influence of feature selection strategies on the performance of machine learning models for loan approval prediction using the publicly available Kaggle "Financial Risk for Loan Approval" synthetic dataset, which contains 20,000 applications. Experiments evaluated multiple feature selection paradigms, including filter-based, wrapper-based, embedded, and PCA-informed approaches across six classification models using stratified 10-fold cross-validation and imbalance-aware metrics. The results show that feature selection consistently improves predictive robustness and minority class recognition. Contrary to assumptions favoring complex models, Logistic Regression combined with Lasso regularization achieved the best overall predictive performance, yielding an ROC-AUC of 99.41% and an F1-score of 91.72%. Embedded feature selection methods provided the most favorable balance between accuracy and computational efficiency. These findings indicate that the effectiveness of feature selection depends heavily on its interaction with model complexity, providing empirical guidance for designing robust, interpretable financial risk prediction systems.

Keywords: Financial Risk Prediction, Loan Approval, Feature Selection, Machine Learning, Classification

Abstrak

Prediksi risiko keuangan yang akurat sangat penting untuk pengambilan keputusan persetujuan pinjaman yang efektif, khususnya dalam sistem keuangan berbasis data. Penelitian ini mengkaji pengaruh strategi seleksi fitur terhadap kinerja model machine learning dalam memprediksi persetujuan pinjaman menggunakan dataset sintetis publik Kaggle "Financial Risk for Loan Approval" yang berisi 20.000 aplikasi. Eksperimen mengevaluasi berbagai paradigma seleksi fitur, termasuk metode berbasis filter, wrapper, embedded, serta pendekatan berbasis PCA, pada enam model klasifikasi dengan menggunakan stratified 10-fold cross-validation dan metrik yang mempertimbangkan ketidakseimbangan data. Hasil penelitian menunjukkan bahwa seleksi fitur secara konsisten meningkatkan ketahanan prediktif (predictive robustness) serta kemampuan mengenali kelas minoritas. Berlawanan dengan asumsi yang mengunggulkan model kompleks, Logistic Regression yang dikombinasikan dengan regularisasi Lasso justru menghasilkan kinerja prediktif terbaik secara keseluruhan, dengan nilai ROC-AUC sebesar 99,41% dan F1-score sebesar 91,72%. Metode seleksi fitur embedded memberikan keseimbangan yang optimal antara akurasi dan efisiensi komputasi. Temuan ini menunjukkan bahwa efektivitas seleksi fitur sangat bergantung pada interaksinya dengan kompleksitas model, serta memberikan panduan empiris dalam merancang sistem prediksi risiko keuangan yang andal dan mudah diinterpretasikan.

Kata Kunci: Prediksi Risiko Keuangan, Persetujuan Pinjaman, Seleksi Fitur, Pembelajaran Mesin, Klasifikasi

1. INTRODUCTION

Accurate risk assessment is fundamental to financial decision-making in lending because incorrect loan approvals can lead to capital losses and reduced economic growth. With the growth of digital finance and the availability of large volumes of borrower data, machine learning models have become central tools for predicting financial risk and loan approval outcomes. These models aim to classify applicants as likely to repay or default, allowing lenders to balance risk and profitability (Ileberi et al., 2024; Chang et al., 2024; Dumitrescu & Hurlin, 2021; Lessmann et al., 2021; Noriega et al., 2023). Machine learning approaches have shown improved predictive performance over traditional statistical models by capturing complex nonlinear relationships in financial data (Chang et al., 2024).

Despite this progress, the performance of machine learning models is highly dependent on the quality of input variables. High-dimensional financial datasets often contain redundant or weakly informative features, which can degrade predictive accuracy, increase computational complexity, and obscure model interpretability. Feature selection is therefore a critical preprocessing step that improves model efficiency and generalizability by retaining only relevant predictors (Ileberi et al., 2024; Jemai & Zarrad, 2023; Bulut & Arslan, 2024; Ayari et al., 2026). Recent studies in credit risk assessment confirm that models incorporating feature selection methods tend to outperform those trained on raw high-dimensional datasets, particularly when coupled with robust classifiers such as gradient boosting or ensemble methods (Chang et al., 2024; Kaur et al., 2023; Muangthanang et al., 2024).

Class imbalance is another pervasive challenge in financial risk prediction. Loan default or approval datasets frequently contain a disproportionate number of instances in one class relative to the other, leading to biased models that favor the majority class. Resampling techniques such as SMOTE or combined oversampling and cleaning methods have been used to mitigate imbalance, improving model sensitivity to minority classes without sacrificing overall accuracy (Aruleba & Sun, 2025; Sáez et al., 2020; Branco et al., 2022). These preprocessing strategies demonstrate that both feature selection and class balance adjustments are essential for producing reliable classifiers in lending environments.

In the literature, a diverse range of feature selection techniques has been applied to credit risk and loan approval prediction. Filter-based methods such as information gain and univariate statistical metrics offer computational simplicity and have been successfully integrated with ensemble classifiers in predictive frameworks (Ileberi et al., 2024; Jemai & Zarrad, 2023). Meanwhile, studies in Indonesian machine learning literature have shown that feature selection techniques such as Information Gain can significantly improve classification performance in supervised learning tasks, particularly when applied to classical algorithms such as Naïve Bayes and K-Nearest Neighbor (Norhalimi & Siswa, 2022). PCA-based dimensionality reduction and hybrid approaches have also been proposed to reduce feature space while maintaining explanatory power in credit scoring tasks (Bulut & Arslan, 2024; Abdi & Williams, 2023; Yang et al., 2022). However, unsupervised approaches like PCA are rarely used directly as feature selectors in supervised classification because they prioritize variance rather than predictive relevance, which raises concerns about interpretability in decision-critical domains such as lending (Abdi & Williams, 2023; Molnar, 2022; Quan & Sun, 2024).

Although prior research has examined feature selection techniques within various machine learning frameworks for financial risk prediction, most studies remain fragmented and focus primarily on post-approval default prediction rather than ex-ante loan approval decisions. In contrast, loan approval systems require predictive models that assess applicant reliability before credit is granted, making early-stage decision support more practically relevant. Additionally, while feature selection and class imbalance handling have been widely studied independently, there is still a limited systematic comparison across multiple feature selection paradigms under consistent experimental settings. This gap is particularly important because financial datasets are typically high-dimensional and imbalanced, and model performance can vary significantly depending on preprocessing choices (Ileberi et al., 2024; Jemai & Zarrad, 2023; Bulut & Arslan, 2024; Kaur et al., 2023; Lessmann et al., 2021).

To address these limitations, this study systematically evaluates feature selection methods for ex-ante loan approval decision-making under imbalanced learning conditions. Specifically, the research investigates how different feature selection paradigms, including filter-based, wrapper-based, embedded, and PCA-informed approaches, affect predictive performance, robustness to class imbalance, and computational efficiency. These methods are evaluated across multiple machine learning classifiers to provide a comprehensive understanding of their relative strengths and weaknesses in practical lending environments (Chang et al., 2024; Dumitrescu & Hurlin, 2021; Noriega et al., 2023; Aruleba & Sun, 2025; Sáez et al., 2020). The research is guided by the following questions:

RQ1: How do different feature selection paradigms influence predictive performance and minority-class detection in imbalanced ex-ante loan approval datasets?

RQ2: Which combinations of feature selection methods and machine learning classifiers provide the best balance between predictive accuracy, interpretability, and computational efficiency for loan approval decision-making?

This study offers three main contributions. First, it provides a comprehensive empirical comparison of filter, wrapper, embedded, and PCA-informed feature selection techniques within a unified experimental framework. Second, it highlights the effectiveness of regularized linear models (particularly Logistic Regression with L1 regularization) as both accurate and highly interpretable solutions in imbalanced loan approval settings. Third, it introduces a balanced evaluation framework that incorporates predictive metrics, minority-class sensitivity, and computational cost to better reflect real-world deployment constraints.

2. METHODS

2.1 Dataset Description

This study uses a publicly available synthetic dataset titled "Financial Risk for Loan Approval," obtained from Kaggle (Zoppelletto, 2025). The dataset is designed to emulate real-world loan application scenarios while avoiding ethical and privacy risks associated with proprietary financial records. It contains 20,000 observations and 36 original attributes representing demographic information, employment status, income characteristics, credit behavior, and loan-related variables. The target variable LoanApproved is binary, indicating whether a loan application is approved or rejected. While the use of a synthetic dataset circumvents privacy regulations, it serves as a limitation of this study, as the data may not perfectly capture the erratic macroeconomic fluctuations and unobserved borrower behaviors present in real-world lending environments.

The approval class represents approximately 23.9% of the dataset, resulting in a moderately imbalanced class distribution. Such an imbalance is consistent with practical lending environments and presents a challenge for classification models that may otherwise favor majority class predictions. Synthetic financial datasets have been increasingly adopted in recent research on financial machine learning due to their ability to enable reproducibility and methodological comparison without regulatory constraints (Ileberi et al., 2024; Chang et al., 2024).

2.2 Data Preprocessing

Data preprocessing is conducted to ensure numerical stability and compatibility across machine learning models. Categorical attributes are transformed using one-hot encoding for nominal variables and ordinal encoding for education level, preserving inherent ordering where applicable. Temporal attributes derived from application timestamps are decomposed into year, month, and day components to capture potential seasonality effects. Continuous variables with heterogeneous scales are standardized using z-score normalization to prevent dominance of large magnitude features in distance-based and gradient-based classifiers.

To address class imbalance, the Synthetic Minority Over-sampling Technique is applied exclusively to training folds during cross-validation. This strategy generates synthetic minority-class samples in the feature space while preventing information leakage into the test data. Recent studies confirm that applying resampling only on training partitions improves generalization and avoids optimistic bias in evaluation (Aruleba & Sun, 2025; Sáez et al., 2020). For the 10-Fold Cross Validation, the dataset was partitioned into 90% training and 10% testing splits for each fold. SMOTE was applied exclusively to the training folds using a default setting of k = 5 nearest neighbors to prevent synthetic data leakage into the validation phase.

2.3 Feature Selection Techniques

Multiple feature selection paradigms are examined to capture complementary notions of feature relevance. Prior to feature selection, an intercorrelation analysis is conducted to address multicollinearity among input variables. A pairwise Pearson correlation matrix is computed, and for any pair of features exhibiting an absolute correlation coefficient above 0.85, one feature is removed based on lower univariate relevance to the target variable. This procedure reduces redundancy and stabilizes subsequent model estimation.

Filter-based feature selection evaluates each variable independently of the learning model. Linear dependency between individual features and the target variable is measured using the Pearson correlation coefficient:

$$r = \frac{cov(X, y)}{\sigma_X \sigma_y} \quad (1)$$

where X denotes an input feature and y represents the loan approval label. Features with low absolute correlation values are considered weak predictors. Additionally, univariate statistical significance is assessed using the ANOVA F statistic, which measures the ratio of interclass to intraclass variance. For discrete non-negative attributes, the chi-square test is employed to quantify deviations from independence between features and the target variable.

Wrapper-based selection is implemented using recursive feature elimination, which iteratively trains a classifier and removes the least informative features based on model performance. This method captures feature interactions but incurs a higher computational cost, making it more suitable for offline analysis than for real-time deployment (Jemai & Zarrad, 2023).

Embedded feature selection integrates feature selection within the learning process. Logistic regression with L1 regularization is employed to enforce sparsity by minimizing the objective function:

$$L = -\log likelihood + \lambda \sum |\beta_j| \quad (2)$$

where β_j represents feature coefficients and λ controls the regularization strength. Features with coefficients shrunk to zero are excluded from the model. Tree-based embedded importance is also examined using ensemble classifiers, where feature relevance is measured through impurity reduction across decision nodes, enabling the capture of nonlinear dependencies.

Unsupervised feature evaluation is explored using Principal Component Analysis (PCA). Instead of directly using principal components as inputs, which reduces interpretability, this study employs a PCA-informed feature ranking approach. Specifically, absolute feature loadings across principal components are examined, and original variables contributing most strongly to the components that collectively explain 95% of total variance are retained. This allows dimensionality reduction while preserving interpretability of original financial variables, addressing a known limitation of unsupervised transformations in decision-critical applications (Abdi & Williams, 2023; Yang et al., 2022).

2.4 Classification Models

To examine whether the impact of feature selection is consistent across learning paradigms, this study evaluates a diverse set of supervised classification models widely used in financial risk prediction. Logistic Regression (LR) is employed as a baseline linear classifier due to its extensive use in credit scoring and regulatory-compliant decision systems. Its probabilistic formulation and coefficient interpretability make it particularly suitable for benchmarking the effects of feature reduction on model transparency and stability (Chang et al., 2024; Dumitrescu & Hurlin, 2021).

Nonlinear and ensemble-based classifiers are included to capture complex interactions among financial attributes. Random Forest (RF) and Gradient Boosting (GB) models are selected because of their strong empirical performance on structured financial data and their robustness to noise and multicollinearity. Random Forest constructs multiple decorrelated decision trees to reduce variance, while Gradient Boosting iteratively improves weak learners to minimize classification loss. Recent studies demonstrate that these ensemble methods consistently outperform linear models in credit risk prediction when feature interactions are present (Ileberi et al., 2024; Kaur et al., 2023).

Support Vector Machines (SVM-s) are used to evaluate margin based classification in high-dimensional feature spaces. By maximizing the separation margin between classes, Support Vector Machines are effective in modeling complex decision boundaries, particularly when kernel functions are employed. Distance-based classification is represented by the K Nearest Neighbors (KNN) algorithm, which assigns class labels based on proximity in feature space. This model is highly sensitive to feature scaling and dimensionality, making it well-suited for assessing the effectiveness of feature selection techniques (Jemai & Zarrad, 2023). Finally, a Multilayer Perceptron (MLP) is included to represent neural network-based learning. As a feedforward architecture with nonlinear activation functions, the multilayer perceptron is capable of approximating complex nonlinear relationships but is more susceptible to overfitting, thereby providing insight into how feature reduction influences generalization performance (Aruleba & Sun, 2025). Hyperparameters for all models are optimized using grid search within training folds to ensure a fair and consistent comparison.

2.5 Evaluation Strategy and Metrics

Model evaluation is conducted using Stratified 10-Fold Cross-Validation to ensure that the proportions of approved and rejected loan applications are preserved across all training and testing splits. This validation strategy provides a reliable estimate of generalization performance while reducing variance caused by random data partitioning. To prevent information leakage, all preprocessing steps, including feature selection and class imbalance handling, are applied exclusively within the training folds prior to model fitting, following best practices for imbalanced classification (Sáez et al., 2020; Chicco & Jurman, 2023).

Given the imbalanced nature of loan approval data, reliance on accuracy alone can lead to misleading conclusions. Threshold independent metrics are therefore emphasized. The area under the receiver operating characteristic curve measures the model’s ability to discriminate between approved and rejected applications across all decision thresholds. In addition, the area under the precision-recall curve (PR-AUC) is reported, as it provides a more informative assessment when the minority class is of primary interest in imbalanced classification problems (Chicco, Tötsch, & Jurman, 2021; Powers, 2020).

In addition to ROC-AUC and F1-score, the Matthews Correlation Coefficient (MCC) is included as a primary robustness metric. MCC is particularly suitable for imbalanced datasets because it considers all four outcomes of the confusion matrix and produces a high score only when the classifier performs well across both classes simultaneously (Chicco & Jurman, 2023; Chicco et al., 2021). MCC is defined as:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{3}$$

Threshold-dependent metrics, including precision, recall, and F1-score, are also reported to support interpretability and comparison with prior studies. Precision reflects the reliability of positive predictions, while recall measures the model's ability to correctly identify approved applications. The F1-score provides a harmonic balance between these two measures. Finally, computational efficiency is evaluated using training and inference times, measured in seconds, on a standard consumer-grade CPU. These metrics reflect practical operational constraints in real-world lending systems where models may require frequent retraining and low-latency predictions.

3. RESULTS AND DISCUSSION

The experimental results demonstrate that feature selection plays a critical role in improving financial risk prediction for loan approval decisions. The overall performance of all classification models across experimental configurations is summarized in Table 3 in the Appendix, which provides a comprehensive view of predictive behavior under different feature selection strategies. Across models, feature selection consistently improves imbalance-aware metrics, indicating that removing redundant or weakly informative variables enhances class separability and model stability. Similar effects have been reported in recent financial risk prediction studies, in which dimensionality reduction improves robustness to class imbalance (Ileberi et al., 2024; Kaur et al., 2023). Models trained without feature selection tend to exhibit higher variance and reduced sensitivity to the minority approval class, particularly in complex and high-dimensional feature spaces.

A more focused comparison of classification models under their respective best-performing feature selection scenarios is presented in Table 1. The results indicate that Logistic Regression (LR) with Lasso regularization achieves the highest overall predictive performance, yielding an ROC-AUC of 99.4120 and an F1-score of 91.7223. This demonstrates that a well-regularized linear model can outperform more complex nonlinear and ensemble-based approaches under the evaluated feature selection configurations.

Table 1 Model Performance Under Best Feature Selection

Model	Best Scenario	Feature Count	*ROC-AUC*	*F1-Score*	Time (s)
LR	Lasso Regularization	17	**99.4120**	91.7223	**0.4842**
SVM	Lasso Regularization	17	99.3491	91.3816	86.7808
MLP	Lasso Regularization	17	99.3563	**92.2253**	89.6565
GB	Lasso Regularization	17	98.5702	87.5701	49.6329
RF	Lasso Regularization	17	98.0925	84.8553	27.1863
KNN	Recursive Feature Elimination	6	94.3190	80.2697	0.8644

In comparison, ensemble models such as Gradient Boosting and Random Forest achieve strong but lower maximum performance, with F1-scores of 87.57 and 84.85, respectively. These findings suggest that although ensemble methods are generally effective for tabular financial data with nonlinear relationships, their advantage may be reduced when feature selection has already removed redundancy and when the underlying data structure is largely linear. This observation aligns with recent studies highlighting that model performance in financial prediction tasks is highly dependent on feature representation quality rather than model complexity alone (Chang et al., 2024; Aruleba & Sun, 2025).

From a financial risk modeling perspective, the strong performance of Logistic Regression with Lasso is particularly important. The L1 regularization mechanism enforces sparsity by shrinking irrelevant coefficients to zero, resulting in a highly interpretable model structure. This characteristic is especially valuable in loan approval systems where regulatory compliance requires transparent and explainable decision-making processes. The ability to isolate a small subset of meaningful financial predictors enhances both predictive performance and decision interpretability in credit risk settings.

The comparative trends observed in Table 1 are visually reinforced in Figure 1, which illustrates the ROC-AUC and F1-score achieved by each classification model under its best-performing scenario. The figure clearly shows that Logistic Regression maintains the highest overall performance, while ensemble models such as Gradient Boosting and Random Forest remain competitive but do not surpass the linear model in this experimental setting. This behavior is consistent with prior studies emphasizing that predictive performance in financial datasets is strongly influenced by feature quality and selection strategy, sometimes more than model complexity (Jemai & Zarrad, 2023).

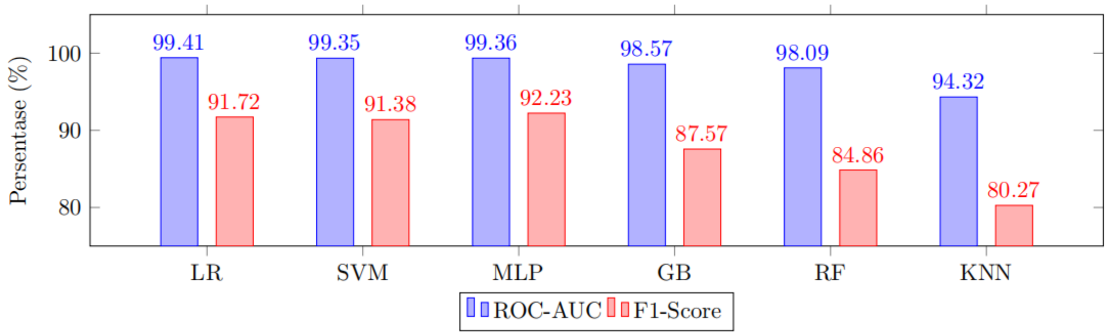

Figure 1 ROC-AUC and F1-Score Comparison Across Models

The influence of feature selection strategies is further examined at the method level in Table 2, which compares feature selection approaches in terms of both predictive performance and computational efficiency. Embedded feature selection methods demonstrate a strong balance between accuracy and efficiency, achieving competitive ROC AUC and F1 scores with relatively low computational overhead. Wrapper-based methods yield great predictive improvements but incur higher training costs, which may limit their applicability in time-sensitive or large-scale lending systems. Filter-based methods provide modest but consistent improvements with minimal computational burden, making them suitable for rapid model prototyping and baseline enhancement. These trade-offs are widely acknowledged in recent comparative analyses of feature selection techniques in financial machine learning (Kaur et al., 2023; Dumitrescu & Hurlin, 2021).

Table 2 Feature Selection Performance and Efficiency Comparison

Feature Selection Methods	***Avg. ROC-AUC***	***Avg. F1-Score***	**Avg. Time (s)**	**Efficiency vs Baseline**
None (Baseline)	97.6804	86.3900	117.80	**-**
Feature-Target Correlation	94.4497	77.4711	56.28	+52.3% Faster
Inter-Feature Correlation	96.9720	83.5463	126.15	-7.09% Slower
Lasso Regularization	**98.4394**	**88.3166**	62.90	+46.6% Faster
Principal Component Analysis	97.7177	86.4588	107.61	+8.6% Faster
Recursive Feature Elimination	96.4065	81.3499	43.74	+62.8% Faster
Tree-Based Random Forest	96.4081	81.3846	43.22	**+63.3% Faster**
Univariate ANOVA F-test	94.4497	77.4711	55.71	+52.7% Faster
Univariate Chi Square	94.5543	77.7776	50.67	+57.0% Faster

The relative effectiveness of these feature selection strategies is summarized in Figure 2, which presents the average ROC AUC and F1 score achieved by each method across all classification models. The figure highlights that embedded and PCA informed approaches achieve more stable performance across models, while filter based methods show greater variability. This result supports recent findings that stability across classifiers is an important indicator of feature selection robustness (Abdi & Williams, 2023; Yang et al., 2022).

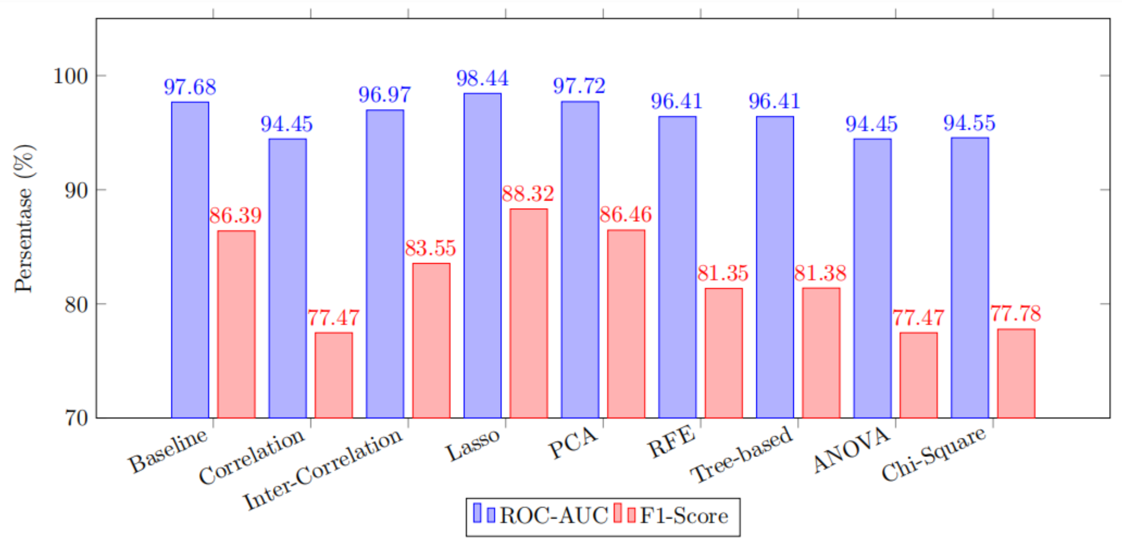

Figure 2 Average ROC-AUC and F1-Score Across Feature Selection Methods

The PCA-informed feature contribution approach exhibits an intermediate behavior between supervised and unsupervised selection strategies. While PCA is traditionally applied as a dimensionality reduction technique rather than a feature selector, the back-tracing of feature loadings used in this study enables the identification of original variables that meaningfully contribute to the variance structure. Models trained on PCA-informed feature subsets achieve performance comparable to supervised filter methods and, in some cases, approach embedded selection performance. Similar PCA interpretability strategies have recently been proposed to balance dimensionality reduction and explainability in financial applications (Abdi & Williams, 2023).

From an evaluation perspective, the results confirm that relying solely on accuracy would obscure meaningful differences in model behavior under class imbalance. Models with similar accuracy values often exhibit substantial divergence in recall and F1 score, underscoring the importance of imbalance-aware metrics for loan approval prediction. The improvements observed in ROC AUC and F1 score following feature selection indicate enhanced discriminative power and improved recognition of the minority class, which are critical for responsible lending decisions. Recent studies also emphasize that ROC AUC and F1 score provide more reliable insight for imbalanced financial classification tasks (Chicco & Jurman, 2023; Chicco, Tötsch, & Jurman, 2021).

Overall, the results demonstrate that no single feature selection method is universally optimal across all models. Instead, the effectiveness of feature selection depends on the interaction between the selection strategy, the classification model, and the computational constraints of deployment. By systematically evaluating feature selection methods using both performance and efficiency criteria, this study provides empirical insights to guide the design of robust and interpretable financial risk prediction systems for loan approval decisions.

4. CONCLUSIONS

This study addressed the problem of how feature selection influences the effectiveness of financial risk prediction for loan approval decisions. The analysis demonstrates that feature selection is not merely a preprocessing step, but a determining factor in improving predictive robustness, particularly under class imbalance conditions commonly observed in lending data. Models trained with appropriate feature selection exhibit stronger discriminative power and improved recognition of minority classes compared to models trained on the full feature set.

The results further indicate that the impact of feature selection is model-dependent. While ensemble-based classifiers and nonlinear models demonstrate strong performance under certain configurations, they do not consistently outperform optimized linear models in this study. Specifically, Logistic Regression combined with Lasso regularization emerges as the best-performing method, achieving the highest ROC-AUC and F1-scores while maintaining excellent computational efficiency (0.48 seconds) and strong interpretability. This makes it particularly suitable for regulatory-compliant loan approval systems where transparency and efficiency are critical requirements.

Embedded feature selection methods provide a practical balance between predictive performance and computational efficiency by integrating feature selection directly into the learning process. Wrapper-based methods can yield competitive performance but at a significantly higher computational cost. In contrast, PCA-informed feature contribution analysis shows that unsupervised variance structure can support dimensionality reduction, although interpretability at the feature level remains essential in financial decision-making contexts.

From a broader perspective, these findings support recent evidence that effective financial risk prediction requires joint consideration of feature relevance, model complexity, and evaluation metrics rather than reliance on a single modeling component. The observed limitations of accuracy as a standalone metric further reinforce the importance of imbalance-aware evaluation for responsible loan approval decision-making.

A limitation of this study is the use of a synthetic dataset, which, while beneficial for reproducibility and controlled experimentation, may not fully capture the complexity and temporal shifts present in real-world financial environments. Future research is recommended to validate the proposed findings using real banking datasets, explore hybrid feature selection strategies that combine supervised and unsupervised approaches, and further investigate MCC-optimized threshold tuning to enhance robustness under severe class imbalance conditions.

REFERENCES

Abdi, H., & Williams, L. J. (2023). Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 15(1), e1609. https://doi.org/10.1002/wics.1609
Aruleba, I., & Sun, Y. (2025). Enhanced credit risk prediction using deep learning and hybrid resampling techniques. Machine Learning with Applications, Article 100692. https://doi.org/10.1016/j.mlwa.2024.100692
Ayari, H., Guetari, P. R., & Kraïem, N. (2026). Machine learning powered financial credit scoring: A systematic literature review. Artificial Intelligence Review, 59, 13. https://doi.org/10.1007/s10462-025-11416-2
Branco, P., Torgo, L., & Ribeiro, R. P. (2022). A survey of predictive modeling on imbalanced domains. ACM Computing Surveys, 54(2), Article 31. https://doi.org/10.1145/3439720
Bulut, C., & Arslan, E. (2024). Comparison of the impact of dimensionality reduction and data splitting on classification performance in credit risk assessment. Artificial Intelligence Review, 57, 252. https://doi.org/10.1007/s10462-024-10904-1
Chang, V., Sivakulasingam, S., Wong, S. T. W., Ganatra, M. A., & Luo, J. (2024). Credit risk prediction using machine learning and deep learning: A study on credit card customers. Risks, 12(11), 174. https://doi.org/10.3390/risks12110174
Chicco, D., & Jurman, G. (2023). The advantages of the Matthews correlation coefficient over F1 score and accuracy in binary classification evaluation. BMC Genomics, 24, 6. https://doi.org/10.1186/s12864-023-09150-3
Chicco, D., Tötsch, N., & Jurman, G. (2021). The Matthews correlation coefficient is more informative than F1 score in binary classification. BMC Genomics, 22, 486. https://doi.org/10.1186/s12864-021-07779-6

Dumitrescu, E. I., & Hurlin, C. (2021). Machine learning for credit risk modeling. Review of Finance, 25(3), 775–808. https://doi.org/10.1093/rof/rfaa036

Ileberi, E., Sun, Y., & Wang, Z. (2024). A machine learning-based credit risk prediction engine system using a stacked classifier and a filter-based feature selection method. Journal of Big Data, 11, 23. https://doi.org/10.1186/s40537-024-00882-0

Jemai, J., & Zarrad, A. (2023). Feature selection engineering for credit risk assessment in retail banking. Information, 14(3), 200. https://doi.org/10.3390/info14030200

Kaur, H., Pannu, H. S., & Malhi, A. K. (2023). A systematic review on imbalanced classification in financial risk prediction. Applied Sciences, 13(4), 2107. https://doi.org/10.3390/app13042107

Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2021). Benchmarking state-of-the-art classification algorithms for credit scoring. European Journal of Operational Research, 290(3), 682–699. https://doi.org/10.1016/j.ejor.2020.08.055

Molnar, C. (2022). Interpretable machine learning (2nd ed.). Open-access online book. https://christophm.github.io/interpretable-ml-book/

Norhalimi, M., & Siswa, T. A. Y. (2022). Optimasi seleksi fitur information gain pada algoritma Naïve Bayes dan K-Nearest Neighbor. JISKA (Jurnal Informatika Sunan Kalijaga), 7(3), 237–255. https://doi.org/10.14421/jiska.2022.7.3.237-255

Powers, D. M. W. (2020). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. Journal of Machine Learning Technologies, 2(1), 37–63. https://doi.org/10.48550/arXiv.2010.16061

Sáez, J. A., Luengo, J., Stefanowski, J., & Herrera, F. (2020). SMOTE-IPF: Addressing noisy and borderline examples in imbalanced classification. Information Sciences, 512, 1430–1449. https://doi.org/10.1016/j.ins.2019.10.004

Yang, L., Zhang, Y., & Chen, X. (2022). Interpretable dimensionality reduction for financial data analysis. Journal of Computational Finance, 26(2), 45–67. https://doi.org/10.21314/JCF.2022.414

Muangthanang, C., Mungsing, S., & Chirawichitcha, N. (2024). Credit risk prediction model using feature engineering and machine learning techniques. International Scientific Journal of Engineering and Technology, 8(1), 19–26. https://doi.org/10.25126/isjet.202512499

Noriega, J. P., Rivera, L. A., & Herrera, J. A. (2023). Machine learning for credit risk prediction: A systematic literature review. Data, 8(11), 169. https://doi.org/10.3390/data8110169

Quan, J., & Sun, X. (2024). Credit risk assessment using the factorization machine model with feature interactions. Humanities and Social Sciences Communications, 11, 234. https://doi.org/10.1057/s41599-024-02700-7

Zoppelletto, L. (2025). Financial Risk for Loan Approval [Dataset]. Kaggle. https://www.kaggle.com/datasets/lorenzozoppelletto/financial-risk-for-loan-approval

JISKA (Jurnal Informatika Sunan Kalijaga)
Vol. 11, No. 2, MAY, 2026, Pp. 225 – 236
ISSN: 2527 – 5836 (print) | 2528 – 0074 (online)

APPENDIX

Table 3 Model Performance for Loan Approval Prediction

Scenario	Feat.	AUC-ROC	AUC-PR	Acc.	Prec.	Rec.	F1-score	Time (s)
GB + Correlations	6	95.2446	88.094	90.045	81.3547	75.7113	78.4297	33.6102
GB + Correlations + SMOTE	6	95.2419	88.0922	87.73	68.9488	88.5565	77.5284	51.7689
GB + Inter-Correlation	62	98.3682	95.499	93.94	89.7688	84.2887	86.9332	82.7696
GB + Inter-Correlation + SMOTE	62	98.3748	95.4532	93.325	82.5886	91.3808	86.7521	156.5975
GB + Lasso	17	98.5702	95.9694	94.24	90.4888	84.8536	87.5701	49.6329
GB + Lasso + SMOTE	17	98.5458	95.8636	93.39	81.6637	93.3473	87.1072	87.4989
GB + Baseline	69	98.567	96.0083	94.38	90.793	85.1674	87.8771	109.5282
GB + Baseline + SMOTE	69	98.5783	95.9849	93.72	83.4479	92.0084	87.5125	200.6795
GB + PCA	50	98.5771	96.0188	94.345	90.7127	85.1046	87.8038	112.2054
GB + PCA + SMOTE	50	98.586	96.0143	93.875	83.8494	92.1548	87.7991	187.2207
GB + RFE	6	96.9869	91.3746	91.75	84.3362	80.5021	82.3507	33.9066
GB + RFE + SMOTE	6	96.9424	91.1102	90.205	73.9121	91.2343	81.6595	51.4997
GB + Tree-based	6	96.9871	91.3756	91.75	84.3362	80.5021	82.3507	32.896
GB + Tree-based + SMOTE	6	96.9424	91.1102	90.205	73.9121	91.2343	81.6595	52.3619
GB + Anova-F	6	95.2446	88.094	90.045	81.3547	75.7113	78.4297	33.355
GB + Anova-F + SMOTE	6	95.2419	88.0922	87.73	68.9488	88.5565	77.5284	51.515
GB + Chi-Square	6	95.354	88.6472	90.585	82.854	76.4435	79.5128	23.5362
GB + Chi-Square + SMOTE	6	95.3843	88.9009	87.96	69.6308	88.0753	77.764	37.1518
KNN + Correlations	6	91.8137	79.0383	88.845	78.8917	72.8243	75.7341	0.8145
KNN + Correlations + SMOTE	6	91.0713	73.2506	85.655	65.5214	84.3933	73.7681	1.2351
KNN + Inter-Correlation	62	90.3166	74.7095	86.92	81.9787	58.0544	67.9626	5.997
KNN + Inter-Correlation + SMOTE	62	89.3077	63.872	74.765	48.5741	94.2259	64.0983	11.3562
KNN + Lasso	17	96.1946	88.683	92.13	88.5656	77.0293	82.3862	3.8062
KNN + Lasso + SMOTE	17	95.6695	83.5461	89.53	71.9159	92.2385	80.814	7.5674
KNN + Baseline	69	91.9951	78.7612	88.495	84.4489	63.5774	72.5354	7.6677
KNN + Baseline + SMOTE	69	91.1081	68.162	77.535	51.6509	94.8954	66.8868	11.9684
KNN + PCA	50	92.0066	79.1652	88.565	84.5047	63.8912	72.7585	6.3397
KNN + PCA + SMOTE	50	91.4885	69.8346	79.54	54.1968	93.1381	68.517	9.8264
KNN + RFE	6	94.319	83.9882	90.73	81.6776	78.9331	80.2697	0.8644
KNN + RFE + SMOTE	6	93.5728	78.3605	88.525	70.7235	88.7448	78.7118	1.2142
KNN + Tree-based	6	94.319	83.9882	90.73	81.6776	78.9331	80.2697	1.273
KNN + Tree-based + SMOTE	6	93.5728	78.3605	88.525	70.7235	88.7448	78.7118	1.1779
KNN + Anova-F	6	91.8137	79.0383	88.845	78.8917	72.8243	75.7341	0.8109
KNN + Anova-F + SMOTE	6	91.0713	73.2506	85.655	65.5214	84.3933	73.7681	1.1702
KNN + Chi-Square	6	91.887	79.956	89.32	79.749	74.1632	76.8486	1.1583
KNN + Chi-Square + SMOTE	6	91.2023	74.5797	86.045	66.3639	84.477	74.3222	1.3282
LR + Correlations	6	95.5583	88.8782	87.925	69.0271	89.7699	78.042	0.791
LR + Correlations + SMOTE	6	95.5619	88.9027	88.075	69.3727	89.728	78.2465	0.7547
LR + Inter-Correlation	62	99.0217	97.2176	94.315	83.3252	95.3138	88.9131	2.0468
LR + Inter-Correlation + SMOTE	62	99.0167	97.2113	94.7	85.1191	94.3305	89.4859	2.8306
LR + Lasso	17	99.412	98.1936	95.815	87.06	96.9247	91.7223	0.4842
LR + Lasso + SMOTE	17	99.4137	98.1976	96.06	88.4357	96.1088	92.1063	1.2491
LR + Baseline	69	99.5604	98.6742	96.215	88.2381	97.1339	92.4686	5.4497
LR + Baseline + SMOTE	69	99.554	98.6578	96.645	90.404	96.1925	93.2052	5.5032
LR + PCA	50	99.5335	98.5971	96.04	87.7082	97.0502	92.1399	3.4322
LR + PCA + SMOTE	50	99.5275	98.5806	96.425	89.7663	96.0042	92.7776	5.0609
LR + RFE	6	97.257	92.1592	90.31	73.933	91.8828	81.9283	0.3354
LR + RFE + SMOTE	6	97.2522	92.1561	90.455	74.4168	91.569	82.1002	0.7633
LR + Tree-based	6	97.257	92.1592	90.31	73.933	91.8828	81.9283	0.3496
LR + Tree-based + SMOTE	6	97.2522	92.1561	90.455	74.4168	91.569	82.1002	1.3035
LR + Anova-F	6	95.5583	88.8782	87.925	69.0271	89.7699	78.042	0.3487
LR + Anova-F + SMOTE	6	95.5619	88.9027	88.075	69.3727	89.728	78.2465	0.7429
LR + Chi-Square	6	95.5886	89.3766	87.47	68.0726	89.6862	77.3902	0.3663
LR + Chi-Square + SMOTE	6	95.6035	89.4494	87.655	68.4955	89.5816	77.6264	1.5385
MLP + Correlations	6	95.5486	88.8952	90.335	82.0543	76.2971	79.0486	16.7086
MLP + Correlations + SMOTE	6	95.3999	88.6595	88.05	69.5658	89.0167	78.0869	80.7459
MLP + Inter-Correlation	62	98.3873	95.4675	94.06	87.6954	87.4477	87.5659	163.2442
MLP + Inter-Correlation + SMOTE	62	98.382	95.4146	93.975	87.4791	87.3431	87.398	224.1517
MLP + Lasso	17	99.3563	98.0406	96.285	92.316	92.1548	92.2253	89.6565
MLP + Lasso + SMOTE	17	99.2986	97.8959	96.025	90.5126	93.2218	91.8188	146.6927
MLP + Baseline	69	99.426	98.297	96.37	92.4765	92.3431	92.4045	139.2362
MLP + Baseline + SMOTE	69	99.4263	98.2996	96.35	92.5333	92.1757	92.3513	155.5665
MLP + PCA	50	99.3594	98.1153	96.13	92.0811	91.7155	91.8871	132.0004

Scenario	Feat.	AUC-ROC	AUC-PR	Acc.	Prec.	Rec.	F1-score	Time (s)
MLP + PCA + SMOTE	50	99.3711	98.1395	96.2	92.1884	91.9038	92.0397	166.012
MLP + RFE	6	97.2282	92.0576	92.15	84.663	82.0502	83.3262	19.2293
MLP + RFE + SMOTE	6	97.1874	91.957	89.785	72.4794	92.4268	81.2319	42.5215
MLP + Tree-based	6	97.2292	92.0585	92.14	84.1934	82.6778	83.4089	17.2364
MLP + Tree-based + SMOTE	6	97.1986	92.0084	89.975	72.9853	92.2594	81.4834	40.0622
MLP + Anova-F	6	95.5486	88.8952	90.335	82.0543	76.2971	79.0486	16.7122
MLP + Anova-F + SMOTE	6	95.3999	88.6595	88.05	69.5658	89.0167	78.0869	79.5963
MLP + Chi-Square	6	95.5833	89.4642	90.725	83.2023	76.6736	79.8012	19.896
MLP + Chi-Square + SMOTE	6	95.473	89.2951	88.03	69.6649	88.5356	77.9582	55.3249
RF + Correlations	6	94.6999	86.7447	89.725	81.9102	73.1799	77.2939	24.979
RF + Correlations + SMOTE	6	94.5651	86.106	88.68	73.0275	83.4728	77.8974	41.1757
RF + Inter-Correlation	62	97.6177	93.1051	92.315	90.0714	76.318	82.6017	27.0861
RF + Inter-Correlation + SMOTE	62	97.2667	92.0261	91.86	82.8985	83.0962	82.9931	62.5804
RF + Lasso	17	98.0925	94.4686	93.15	89.9695	80.3138	84.8553	27.1863
RF + Lasso + SMOTE	17	98.0284	94.2313	92.98	82.5878	89.5397	85.9123	52.2578
RF + Baseline	69	97.6469	93.24	92.48	89.0216	78.2008	83.2505	32.9947
RF + Baseline + SMOTE	69	97.473	92.6332	92.185	81.8653	86.5063	84.109	74.252
RF + PCA	50	97.7832	93.5737	92.64	89.2124	78.7657	83.6501	35.7593
RF + PCA + SMOTE	50	97.6043	93.0422	92.34	81.9586	87.1548	84.4695	77.6754
RF + RFE	6	96.6107	89.8448	91.245	83.8435	78.5146	81.0816	23.3076
RF + RFE + SMOTE	6	96.5029	89.231	90.58	76.5125	87.4477	81.6079	39.4548
RF + Tree-based	6	96.5818	89.8589	91.21	83.7171	78.5146	81.0219	23.2792
RF + Tree-based + SMOTE	6	96.5384	89.3434	90.63	76.5023	87.8033	81.7503	39.5321
RF + Anova-F	6	94.6999	86.7447	89.725	81.9102	73.1799	77.2939	24.8748
RF + Anova-F + SMOTE	6	94.5651	86.106	88.68	73.0275	83.4728	77.8974	42.5193
RF + Chi-Square	6	94.4813	87.1666	90.19	82.9207	74.2887	78.3501	19.1333
RF + Chi-Square + SMOTE	6	94.4682	86.4353	88.76	73.3252	83.3473	77.9991	30.4101
SVM + Correlations	6	94.3023	82.6079	87.695	68.4918	89.8954	77.7436	143.6857
SVM + Correlations + SMOTE	6	94.3891	82.8723	87.78	68.7152	89.749	77.8334	279.1073
SVM + Inter-Correlation	62	98.834	96.6451	94.195	83.5012	94.4142	88.6149	270.0807
SVM + Inter-Correlation + SMOTE	62	98.7706	96.4405	94.74	87.4205	91.1506	89.237	505.0545
SVM + Lasso	17	99.3491	97.9304	95.6	86.0428	97.4477	91.3816	86.7808
SVM + Lasso + SMOTE	17	99.3416	97.9139	95.965	88.4101	95.6904	91.9002	202.0215
SVM + Baseline	69	99.4218	98.2544	95.775	87.2503	96.4435	91.6115	246.7759
SVM + Baseline + SMOTE	69	99.4078	98.204	96.33	90.82	94.1841	92.4674	424.0326
SVM + PCA	50	99.3939	98.1623	95.65	86.769	96.5481	91.3933	187.153
SVM + PCA + SMOTE	50	99.3811	98.118	96.225	90.4401	94.1841	92.27	368.6557
SVM + RFE	6	96.4714	87.005	89.41	71.2884	93.3054	80.8167	104.1266
SVM + RFE + SMOTE	6	96.5471	87.3649	89.625	71.8105	93.2008	81.1143	207.6492
SVM + Tree-based	6	96.4714	87.005	89.41	71.2884	93.3054	80.8167	103.8213
SVM + Tree-based + SMOTE	6	96.5471	87.3649	89.625	71.8105	93.2008	81.1143	205.3745
SVM + Anova-F	6	94.3023	82.6079	87.695	68.4918	89.8954	77.7436	142.4074
SVM + Anova-F + SMOTE	6	94.3891	82.8723	87.78	68.7152	89.749	77.8334	274.5136
SVM + Chi-Square	6	94.7675	85.9683	87.93	69.3574	88.7657	77.8583	144.4931
SVM + Chi-Square + SMOTE	6	94.8582	86.1955	87.965	69.4466	88.7238	77.8997	273.6462