JAISEN: **Journal of Advanced Information Systems and Engineering**
Vol. 2, 1 (February, 2026), pp 72-78
ISSN: - EISSN: - , DOI: -

Comparative Analysis of Advanced Machine Learning Models for SME Credit Risk Classification

Rizkiandi Farma Saputra[1], Susandri Susandri[2], Ahmad Zamsuri[3]

[1] Master of Computer Science, Lancang Kuning University; rizkiandifarmasaputra@gmail.com
[2] Master of Computer Science, Lancang Kuning University; susandri@unilak.ac.id
[3] Master of Computer Science, Lancang Kuning University; ahmadzamsuri@unilak.ac.id

ARTICLE INFO

Keywords:

Credit Risk Classification;
Machine Learning;
Xgboost;
Sme Financing;
Predictive Analytics.

Article history:

Received 2026-01-23
Revised 2026-02-12
Accepted 2026-02-25

ABSTRACT

This study addresses the growing need for accurate and reliable credit risk classification in Small and Medium Enterprises (SMEs), which play a vital role in economic development but remain vulnerable to financial instability and non-performing loans. The purpose of this research is to comparatively evaluate the performance of advanced machine learning models in multi-class SME credit risk classification and to identify the most influential predictors affecting creditworthiness. A quantitative experimental approach was employed using 2,000 SME debtor records from 2020–2024, incorporating 17 financial and behavioral variables. The study implemented five supervised learning algorithms Logistic Regression, Random Forest, Support Vector Machine (RBF), XGBoost, and Artificial Neural Network combined with data preprocessing, feature selection, and 5-fold cross-validation. Model performance was assessed using imbalance-aware metrics, including F1-Macro and ROC-AUC, alongside statistical validation using the Friedman test. The results demonstrate that ensemble-based methods outperform traditional models, with XGBoost achieving the highest predictive performance (F1-Macro = 0.92; ROC-AUC = 0.95) and showing statistically significant superiority. Feature importance analysis reveals that Debt-to-Income Ratio, Credit Tenure, and Internal Credit Score are the most influential predictors, aligning with established financial risk theory. In conclusion, this study confirms the effectiveness of ensemble machine learning models in improving SME credit risk classification and highlights their potential integration into automated decision-support systems to enhance risk management, reduce non-performing loans, and support data-driven financial decision-making.

Corresponding Author:
Rizkiandi Farma Saputra
Master of Computer Science, Lancang Kuning University; rizkiandifarmasaputra@gmail.com

1. INTRODUCTION

Small and Medium Enterprises (SMEs) constitute a critical pillar of Indonesia's economic structure, contributing significantly to employment absorption, regional economic resilience, and post-pandemic recovery. Despite their strategic role, SMEs remain highly vulnerable to credit risk due to limited financial literacy, fluctuating cash flows, inadequate collateral structures, and asymmetric information between borrowers and financial institutions. Recent empirical reports indicate increasing Non-Performing Loan (NPL) ratios within the SME segment, reflecting inefficiencies in traditional credit assessment mechanisms (Hanif & Widawati, 2024; Cahyani & Nasrullah, 2025).

Conventional credit scoring approaches—primarily based on linear statistical models such as Logistic Regression—have been widely implemented in financial institutions. However, these methods often struggle to capture complex non-linear interactions among financial indicators, behavioral variables, and macroeconomic uncertainties (Kurniawati, 2021; Nurhalizah & Ardianto, 2024). The growing availability of structured financial datasets and computational capabilities has encouraged the adoption of machine learning (ML) techniques in predictive credit risk modeling.

Machine learning provides superior flexibility in modeling high-dimensional data, handling multicollinearity, and identifying hidden patterns within borrower behavior. Recent studies have explored various ML algorithms, including Support Vector Machine, Random Forest, Naïve Bayes, and XGBoost, for credit scoring and risk prediction (Givari et al., 2022; Damanik et al., 2023; Aqil & Karim, 2025). Ensemble-based methods, particularly boosting algorithms, have demonstrated promising results in financial classification tasks due to their sequential error correction mechanism and improved generalization performance (Nasution et al., 2025; Kandi & Garc, 2025).

However, despite the expanding literature, several research limitations remain evident. First, many previous studies focus on binary classification (default vs. non-default), whereas real-world SME financing often requires multi-class risk categorization (e.g., Lancar, Berisiko Tinggi, Macet). Second, limited research provides statistically validated comparative benchmarking across multiple advanced algorithms within the Indonesian SME context. Third, performance evaluation in prior works frequently relies solely on accuracy metrics without addressing class imbalance issues, potentially leading to biased interpretation of predictive performance (Ulfah et al., 2023; Dzulhijjah et al., 2021).

These gaps indicate the need for a comprehensive, statistically robust, and imbalance-aware comparative evaluation of advanced machine learning models for multi-class SME credit risk classification. Addressing this gap is crucial for enhancing the reliability and operational relevance of predictive credit scoring systems. Therefore, this study aims to:

1. Compare the predictive performance of five supervised machine learning algorithms—Logistic Regression, Random Forest, Support Vector Machine (RBF), XGBoost, and Artificial Neural Network—in multi-class SME credit risk classification.
2. Validate model performance differences statistically using the Friedman test and post-hoc analysis.
3. Identify the most influential financial predictors contributing to SME credit risk classification.

The contributions of this research are threefold. First, it provides a statistically validated multi-model benchmarking framework for SME credit risk classification under imbalance-aware evaluation metrics (F1-Macro and ROC-AUC). Second, it extends the literature by implementing multi-class risk categorization within the Indonesian SME financing context. Third, it offers practical insights for financial institutions to integrate ensemble-based predictive analytics into automated decision-support systems.

By strengthening methodological rigor and empirical validation, this study contributes to the advancement of data-driven credit risk management and intelligent financial information systems.

2. METHODS

This research employs a quantitative experimental approach using secondary SME credit data collected from 2020–2024. The dataset consists of 2,000 observations with 17 variables, including financial indicators, credit behavior, and categorical risk labels (Lancar, Berisiko Tinggi, Macet). Data preprocessing includes cleaning, normalization, feature encoding, imbalance handling, and train-test split (80:20), which is a common workflow in predictive analytics pipelines (Kurniawan et al., 2022; Nurhalizah & Ardianto, 2024). Feature selection was conducted using ANOVA F-test for numerical variables and Chi-Square for categorical variables to retain statistically relevant predictors for classification tasks (Kurniawan et al., 2022). Model validation was performed using 5-fold cross-validation to obtain stable and unbiased performance estimates (Kurniawan et al., 2022; Suhadolnik & Ueyama, 2023).

The evaluated algorithms include Logistic Regression, Random Forest, Support Vector Machine (RBF kernel), XGBoost, and Artificial Neural Network, consistent with comparative supervised learning studies and credit scoring applications (Nasution et al., 2025; Givari et al., 2022; Syafi'i et al., 2022; Damanik et al., 2023; Novianti et al., 2023). Hyperparameter tuning was performed using Grid Search optimization to improve generalization, while controlling overfitting risk, as emphasized in recent ML model development studies (Susandri et al., 2025). Evaluation metrics include Accuracy, Precision, Recall, F1-Macro, ROC-AUC, and Friedman statistical test to determine significant differences among models. In particular, F1-Macro and ROC-AUC were used to provide imbalance-aware evaluation and threshold-independent discrimination, which are recommended for credit risk settings with skewed class distributions (Dzulhijjah et al., 2021; Ulfah et al., 2023; Givari et al., 2022).

Model comparison was statistically validated using the Friedman non-parametric test to assess ranking differences across classifiers. Post-hoc Nemenyi analysis was applied to identify pairwise significant differences at $\alpha = 0.05$. This approach ensures robustness and reduces the risk of random performance bias in multi-model evaluation.

3. FINDINGS AND DISCUSSION

The experimental results are presented systematically to address the research objectives: (1) comparative model performance, (2) statistical validation, and (3) identification of influential predictors in SME credit risk classification.

3.1. Class Distribution and Evaluation Strategy

The dataset exhibits moderate class imbalance, where the "Macet" category represents the largest proportion, followed by "Lancar" and "Berisiko Tinggi." Such distribution may bias accuracy-based evaluation, as models tend to favor the majority class (Dzulhijjah et al., 2021; Ulfah et al., 2023). Therefore, F1-Macro was selected as the primary evaluation metric to ensure balanced performance assessment across all classes, particularly under skewed credit risk distributions (Givari et al., 2022; Ulfah et al., 2023).

The imbalance level remains moderate, thus not requiring extreme resampling techniques; however, performance interpretation prioritizes recall and F1-Macro to avoid underestimating minority classes, as recommended in imbalanced credit classification studies (Dzulhijjah et al., 2021; Givari et al., 2022).

Table 1. Class Distribution of SME Credit Risk Dataset

Risk Category	Frequency	Percentage
Lancar	610	30.5%
Berisiko Tinggi	500	25.0%
Macet	890	44.5%
Total	2000	100%

3.2. Comparative Model Performance

The comparative evaluation results are summarized in Table 2. The results demonstrate that ensemble-based models outperform linear classifiers, particularly in non-linear financial datasets (Syafi'i et al., 2022; Givari et al., 2022; Kandi & Garc, 2025). XGBoost achieved the highest F1-Macro (0.92) and ROC-AUC (0.95), indicating superior discriminatory capability across multi-class classification settings, consistent with findings in comparative ML evaluation research (Nasution et al., 2025; Damanik et al., 2023).

Random Forest also showed strong stability across cross-validation folds, confirming the robustness of bagging-based ensemble methods in credit scoring applications (Givari et al., 2022). Logistic Regression, although stable, exhibited lower predictive power, suggesting limitations in modeling complex feature interactions and non-linear relationships commonly found in SME financing data (Kurniawati, 2021; Nurhalizah & Ardianto, 2024).

Table 2. Comparative Performance of Machine Learning Models

Model	Accuracy	F1-Macro	ROC-AUC
Logistic Regression	0.86	0.84	0.88
Random Forest	0.91	0.90	0.93
SVM (RBF)	0.89	0.87	0.91
XGBoost	0.93	0.92	0.95
ANN	0.88	0.86	0.89

The superiority of boosting mechanisms may be attributed to their sequential error correction strategy, which iteratively minimizes residual loss and enhances generalization performance under complex non-linear feature interactions.

3.3. ROC Curve Analysis

The ROC Curve comparison further illustrates the superiority of ensemble methods. ROC-AUC is widely recommended for evaluating threshold-independent classification performance in financial risk prediction (Ulfah et al., 2023; Xu, 2024).

Figure 1(a) shows that XGBoost consistently maintains the largest area under the curve across classification thresholds. Figure 1(b) highlights its strong true positive rate performance while maintaining low false positive rates, reflecting enhanced separability between default and non-default classes.

These findings confirm that boosting-based ensemble techniques provide enhanced discrimination capability in creditworthiness prediction, as similarly observed in recent ML-based financial modeling studies (Kandi & Garc, 2025; Nasution et al., 2025).

3.4. Statistical Validation Using Friedman Test

To validate performance differences statistically, the Friedman test was conducted. Non-parametric statistical comparison methods are recommended when evaluating multiple machine learning classifiers across the same dataset due to their robustness against distributional assumptions (Givari et al., 2022).

The null hypothesis (H_0) states that there is no significant performance difference among models. The Friedman test yielded a statistically significant result ($p < 0.05$), indicating meaningful performance differences among the evaluated classifiers.

Post-hoc Nemenyi analysis further identifies XGBoost as statistically superior compared to Logistic Regression and ANN, while its difference with Random Forest remains marginal but favorable. This statistical validation strengthens the reliability of the comparative findings and eliminates potential bias from random performance fluctuations, aligning with rigorous evaluation frameworks in predictive analytics research (Kurniawan et al., 2022; Nurhalizah & Ardianto, 2024).

3.5. Feature Importance Analysis

Feature importance analysis (Figure 2) reveals the most influential predictors, namely debt-to-income ratio, credit tenure, internal credit score, and total loan amount.

The dominance of leverage-related indicators confirms classical financial risk theory, which emphasizes repayment capacity and financial burden as primary determinants of default probability (Hanif & Widawati, 2024; Indah Sucianty & Suria Manda, 2022; Alfian, 2024). Specifically:

1. Debt-to-Income Ratio directly measures borrower solvency capacity, frequently identified as a strong predictor of delinquency (Hanif & Widawati, 2024).
2. Credit Tenure reflects repayment exposure duration and risk accumulation over time (Andrianto & Nurjanah, 2023).
3. Internal Credit Score captures historical behavioral patterns and borrower discipline (Kurniawati, 2021).

These findings align with previous empirical research in financial risk modeling and SME financing analysis, reinforcing the theoretical consistency of the results (Irham et al., 2024; Juhainah, 2025).

3.6. Discussion in Relation to Research Objectives

This study addressed three primary objectives. First, regarding model performance, XGBoost achieved superior results across F1-Macro and ROC-AUC metrics, consistent with findings from comparative supervised learning benchmarks (Nasution et al., 2025; Syafi'i et al., 2022). Second, in explaining why XGBoost outperforms other models, its boosting mechanism sequentially corrects prediction errors, enabling more effective handling of non-linear interactions and imbalanced datasets, which are common characteristics of financial risk data (Kandi & Garc, 2025; Givari et al., 2022). Third, the findings are consistent with prior studies demonstrating that ensemble methods outperform linear classifiers in credit risk prediction and scoring systems (Damanik et al., 2023; Aqil & Karim, 2025; Novianti et al., 2023). Compared to studies that rely solely on Logistic Regression or Naïve Bayes, this research provides broader comparative evidence within the SME context and incorporates statistical validation, thereby offering stronger inferential support.

3.7. Practical Implications

From a practical perspective, implementing ensemble-based machine learning systems can namely enhance early risk detection, reduce exposure to non-performing loans, improve automated credit scoring accuracy, support digital financial inclusion strategies.

The integration of machine learning into managerial and fintech decision-making frameworks has been empirically associated with improved risk mitigation, enhanced predictive accuracy, and more dynamic portfolio monitoring capabilities in credit risk environments (Asmoro & Sriyono, 2025; Cahyani & Nasrullah, 2025; Riskandy & Tjahyanto, 2025). Financial institutions may integrate XGBoost-based predictive models into automated decision-support frameworks to improve loan approval consistency and reduce credit portfolio deterioration (Gumar, 2025; Riskandy & Tjahyanto, 2025).

3.8. Future Research Directions

Future research may explore several important directions, including time-to-default modeling, real-time deployment within enterprise systems, the integration of explainability techniques such as SHAP analysis, and the development of hybrid ensemble architectures. These extensions are expected to further enhance predictive reliability and support stronger regulatory compliance in financial institutions, particularly within dynamic and post-pandemic SME financing environments (Hanif & Widawati, 2024; Paz et al., 2025).

4. CONCLUSION

This study demonstrates that advanced ensemble-based machine learning algorithms significantly enhance SME credit risk classification performance compared to conventional linear approaches. Among evaluated models, XGBoost achieved the most optimal predictive performance in terms of F1-Macro and ROC-AUC under moderate class imbalance conditions. Statistical validation using the Friedman test confirms that performance differences among classifiers are significant. Feature importance analysis reinforces classical credit risk theory by highlighting leverage and repayment-related indicators as dominant predictors. The findings provide both empirical and practical contributions to predictive analytics in SME financing. Future research should explore time-to-default modeling, explainability integration, and enterprise-level real-time deployment frameworks to strengthen operational scalability and regulatory compliance.

REFERENCES

Andrianto, T., & Nurjanah, Y. (2023). Pengalaman usaha dan omzet usaha terhadap kelancaran angsuran pembiayaan ultra mikro. *JIakes: Jurnal Ilmiah Akuntansi Kesatuan, 11*(2). https://doi.org/10.37641/jiakes.v11i1.1656

Aqil, I., & Karim, A. (2025). Analisis prediksi risiko pembiayaan pada Bank Baitul Maal Wat Tamwil (BMT) UGT Nusantara menggunakan algoritma Support Vector Machine. *Jurnal NAAFI, 1*(6), 856–866. https://doi.org/10.62387/naafi.v1i5.266

Cahyani, N. D., & Nasrullah, N. (2025). Evaluasi kebijakan penyaluran kredit dan risiko kredit pada pemberian kredit UMKM di Bank BRI Cabang Somba Opu. *Jurnal Bisnis dan Kewirausahaan, 14*(2), 196–209. https://doi.org/10.37476/jbk.v14i2.5123

Damanik, J. R., Fauzi, R., & Hamami, F. (2023). Implementasi algoritma klasifikasi Naïve Bayes untuk klasifikasi credit scoring pada platform peer-to-peer lending. *Journal of Software Engineering and Computer Science (JOSYC), 4*(4), 880–890. https://doi.org/10.47065/josyc.v4i4.4059

Dzulhijjah, D. A., Herlambang, M. B., & Haifan, M. (2021). Implementasi framework CRISP-DM untuk proses data mining aplikasi credit scoring PT. XYZ. *Jurnal Teknologi Informasi dan Ilmu Komputer, 8*(2), 238–251.

Givari, M. R., Sulaeman, M. R., & Umaidah, Y. (2022). Perbandingan algoritma SVM, Random Forest dan XGBoost untuk penentuan persetujuan pengajuan kredit. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 16*(2), 141–149.

Gumar, S. (2025). Assessment of SME loan portfolio quality Kazakhstan: Empirical analysis from 2013 to 2023. *Eurasian Journal of Economics and Business Studies, 69*(3), 66–80. https://doi.org/10.47703/ejebs.v69i3.549

Hanif, M. S., & Widawati, A. S. (2024). Analisis determinan kredit macet pada UMKM di Indonesia. *Jurnal Riset Entrepreneurship, 7*(1), 31. https://doi.org/10.30587/jre.v7i1.7036

Irham, M., Mutia, A., & Ramli, F. (2024). Pengaruh literasi keuangan dan mitigasi risiko terhadap ketahanan pembiayaan UMKM. *Jurnal Manajemen Keuangan dan Perbankan, 17*(1), 52–66.

Juhainah. (2025). Manajemen risiko keuangan dalam menunjang stabilitas pembiayaan UMKM. *Jurnal Ekonomi dan Keuangan Syariah, 2*(1), 79–88.

Kandi, K., & Garc, A. (2025). Evaluating the performance of deep convolutional neural networks and support vector regression for creditworthiness prediction in the financial sector. *International Journal of Artificial Intelligence, 28*(76), 66–84. https://doi.org/10.4114/intartif.vol28iss76pp66-84

Kurniawan, A., Rifa', A., Nafis, M. A., Sefrida, N., & Patria, H. (2022). Pemilihan metode predictive analytics dengan machine learning untuk analisis dan strategi peningkatan. *Indonesian Journal of Applied Statistics, 5*(1). https://doi.org/10.13057/ijas.v5i1.55483

Kurniawati, E. T. (2021). Pengembangan sistem credit scoring pembiayaan UMKM. *Jurnal Sistem Informasi Bisnis, 17*(3), 609–616.

Nasution, N., Hasan, M. A., & Bakri Nasution, F. (2025). Predicting heart disease using machine learning: An evaluation of Logistic Regression, Random Forest, SVM, and KNN models on the UCI Heart Disease dataset. *IT Journal Research and Development, 9*(2), 140–150. https://doi.org/10.25299/itjrd.2025.17941

Novianti, T., Mandati, S. A., & Andana, E. K. (2023). Peningkatan evaluasi risiko kredit menggunakan Decision Tree C4.5. *Journal of Manufacturing in Industrial Engineering & Technology, 2*(2), 1–9. https://doi.org/10.30651/mine-tech.v2i2.21749

Nurhalizah, R. S., & Ardianto, R. (2024). Analisis supervised dan unsupervised learning pada machine learning: Systematic literature review. *Jurnal Informatika, 18*(1), 61–72.

Paz, Á., Crawford, B., Monfroy, E., Barrera-García, J., Fritz, Á. P., Soto, R., Cisternas-Caneo, F., & Yáñez, A. (2025). Machine learning and metaheuristics approach for individual credit risk assessment: A systematic literature review. *Applied Artificial Intelligence Review, 1*(1), 1–46.

Permadi, R. I., & Wulandari, P. (2024). Penilaian kredit untuk UKM menggunakan hybrid BWM dan TOPSIS pada bank syariah di Indonesia. *Jurnal Cahaya Mandalika, 5*(1), 141–153.

Riskandy, Y. H., & Tjahyanto, A. (2025). Integrasi analytic hierarchy process (AHP)–machine learning yang dinamis dalam prediksi risiko kredit. *Jurnal Locus Penelitian dan Pengabdian, 4*(10), 9448–9455. https://doi.org/10.58344/locus.v4i10.4459

Susandri, S., Zamsuri, A., Nasution, N., Efendi, Y., & Alwan, H. B. (2025). The mitigating overfitting in sentiment analysis insights from CNN-LSTM hybrid models. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, 24*(2), 297–308. https://doi.org/10.30812/matrik.v24i2.4742

Syafi'i, N., Nurdiawan, O., & Dwilestari, G. (2022). Penerapan machine learning untuk menentukan kelayakan kredit menggunakan metode support vector machine. *Jurnal Sistem Informasi dan Manajemen, 10*(2).

Trinh, L. T. (2025). A comparative analysis of consumer credit risk models in peer-to-peer lending. *Journal of Economics, Finance and Administrative Science, 29*(58), 346–365. https://doi.org/10.1108/JEFAS-04-2021-0026

Ulfah, M., Saragih, T. H., Kartini, D., Mazdadi, M. I., & Abadi, F. (2023). Penerapan MWMOTE untuk mengatasi ketidakseimbangan kelas pada klasifikasi risiko kredit. *Jurnal Informatika Polinema*, 477–486.

Xu, T. (2024). Fraud detection in credit risk assessment using supervised learning algorithms. *Journal of Financial Technology and Data Science, 12*(2), 30–36.