Int. J. Anal. Appl. (2026), 24:284

International Journal of Analysis and Applications

Small-Sample Classification of Economic Impacts from Registered SMEs Using Machine Learning

Piyada Wongwiwat[1], **Wikanda Phaphan**[2,3,*]

[1]*Department of Data Management, Faculty of Science and Technology, Suansunandha Rajabhat University, Thailand*

[2]*Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand*

[3]*Research Group in Statistical Learning and Inference, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand*

[*]*Corresponding author:* wikanda.p@sci.kmutnb.ac.th

Abstract. Registered Small and Medium Enterprises (SMEs) serve as a fundamental pillar of Thailand's economy; however, modeling their economic impact is often constrained by the limitations of small-scale datasets. This study evaluates the performance of seven machine learning algorithms—Decision Tree, Support Vector Machine (SVM), Gradient Boosting, K-Nearest Neighbor (KNN), Naïve Bayes, XGBoost, and Multi-Layer Perceptron (MLP) Neural Network—in classifying the economic contributions of registered SMEs. To mitigate issues arising from class imbalance and limited sample size, the SMOTEENN (Synthetic Minority Over-sampling Technique and Edited Nearest Neighbors) hybrid sampling method was employed. The results indicate that while SMOTEENN significantly improved recall (reaching 1.000 for the majority of models), varying precision levels resulted in moderate F1 scores for the Decision Tree (0.83), SVM (0.87), and Gradient Boosting (0.88) classifiers. Notably, XGBoost and Naïve Bayes achieved peak performance with perfect scores (1.000) across accuracy, F1 score, and AUC metrics, demonstrating a robust capacity for pattern recognition within small-sample contexts, albeit with a recognized risk of overfitting. In contrast, the MLP Neural Network exhibited the lowest efficacy (Accuracy 0.778, AUC 0.438), suggesting that deep learning architectures may require more extensive hyperparameter tuning or larger datasets to remain competitive. These findings suggest that ensemble methods and probabilistic models are most effective for SME classification tasks under data constraints, providing a framework for more precise, data-driven economic policy-making.

Received: Mar. 30, 2026.

2020 *Mathematics Subject Classification.* 68T05; 68T10; 68T20; 62H30.

Key words and phrases. registered SMEs; credit risk models; decision tree; support vector machine; gradient boosting; K-nearest neighbor; Naïve Bayes; XGBoost; MLP neural network.

https://doi.org/10.28924/2291-8639-24-2026-284
ISSN: 2291-8639

1. Introduction

Existing research on Small and Medium Enterprises (SMEs) highlights the multifaceted applications of machine learning and the complexity of assessing economic impacts across diverse contexts. The socioeconomic contributions of SMEs often extend beyond simple financial metrics. Rodhiyah [1] investigated the convection industry in Semarang, identifying significant positive externalities such as job creation—particularly for women—and the stimulation of local supporting businesses. However, this study also noted negative environmental impacts, such as noise pollution, suggesting that SME growth requires a balanced assessment of community perception and industrial side effects. Recent advancements in predictive modeling have significantly enhanced decision-making capabilities within the sector. Cofino and Cerna [2] introduced SME-ODA, a machine learning platform utilizing five distinct algorithms: K-Means Clustering, Random Forest, XGBoost, SVM, and Logistic Regression. Their findings demonstrated that these machine learning approaches provide superior predictive accuracy for customer behavior compared to traditional statistical methods. Similarly, Inayatulloh [4] addressed the discrepancy between e-commerce adoption and actual sales effectiveness by proposing a machine learning-based adoption model. This framework leverages feature recommendations to optimize online sales performance, bridging the gap between digital presence and economic results. Beyond growth and marketing, machine learning has proven critical in risk assessment. Fernandez et al. [3] utilized the Cyber Security Breaches Survey to analyze the severity of cyber incidents within SMEs. By characterizing breach impacts through the lenses of operational disruption time and associated recovery costs, their work underscores the utility of algorithmic analysis in mitigating modern digital risks.

The application of machine learning (ML) architectures has become increasingly pivotal in addressing the structural and operational challenges inherent to the SME sector. Within the domain of risk management, cybersecurity research indicates that SMEs are frequently subjected to diverse cyber breaches of varying intensity. Fernandez et al. [3] demonstrate that the severity of these incidents—quantified through operational disruption duration and financial recovery costs - exerts a substantial impact across the economic, financial, and managerial dimensions of the firm. Furthermore, the technical efficacy of ML models in this context is often contingent upon the quality of the underlying data. Addressing the pervasive issue of class imbalance is critical; research suggests that classification performance can be significantly bolstered through targeted data transformation and advanced sampling techniques [5]. Specifically, the integration of undersampling and mixed sampling methodologies has been shown to enhance predictive accuracy, providing a more reliable foundation for analyzing SME-related datasets where minority class instances are often underrepresented.

Recent systematic reviews have rigorously evaluated the comparative performance of supervised machine learning algorithms across a spectrum of classification and predictive tasks. A comprehensive meta-analysis by Pangestu et al. [6], spanning 243 articles published between 2019 and 2023, identifies Random Forest as the most proficient architecture for prediction, yielding an

average accuracy of 91.18%. Conversely, Support Vector Machines (SVM) were found to demonstrate superior efficacy in classification tasks, achieving a mean accuracy of 88.85%. This versatile utility of supervised techniques including Decision Trees, KNN, Naïve Bayes, Gradient Boosting, and Neural Networks is further substantiated by Ghildiyal et al. [7], who highlight their critical relevance in advancing engineering and biomedical research. The literature also emphasizes the importance of data balancing and feature extraction in specific applications such as sentiment analysis and natural language processing. Setiawan and Nastiti [8] utilized SMOTE to address class imbalances, finding that while SVM and XGBoost both achieved high accuracy (93%), the Extra Trees classifier reached an optimal F1-Score of 96%. In the context of text classification, Hendrawan et al. [9] established the dominance of XGBoost over Naïve Bayes. Their findings indicate that XGBoost, when integrated with Word2vec or TF-IDF vectorization, achieves F1-scores of 0.941 and 0.940 respectively, significantly outperforming the Naïve Bayes benchmarks of 0.915 and 0.900.

This study introduces a novel dataset derived from registered Small and Medium Enterprises (SMEs) in Thailand, focusing on critical economic impact indicators such as revenue trajectories, employment growth, and market penetration. To ensure the integrity of the analytical framework, the dataset underwent preprocessing via SMOTEENN (Synthetic Minority Over-sampling Technique and Edited Nearest Neighbors), a hybrid approach designed to mitigate class imbalance and refine decision boundaries. The methodological core consists of a single-stage classification pipeline evaluating seven diverse machine learning architectures: Decision Tree, Support Vector Machine (SVM), Gradient Boosting, K-Nearest Neighbor (KNN), Naïve Bayes, XGBoost, and Multi-Layer Perceptron (MLP) Neural Network. The primary objectives of this research are: (1) to evaluate comparative performance metrics across distinct algorithmic families to determine their suitability for small-sample economic datasets; (2) to identify optimal classifiers capable of maintaining high generalization performance under data constraints; (3) to provide empirical insights to inform data-driven policy-making and strategic economic analysis for registered SMEs, both within the Thai domestic market and analogous global contexts.

2. Materials and Methods

2.1. **Data.**

2.1.1. *Data acquisition:* Data were collected from financial institutions and SME regulatory records in Thailand, capturing loan application details for registered SMEs [10–12]. The dataset includes 64 variables (X1-X64) derived from applicant profiles, business operations, and financial statements, with loan repayment status (Y) as the binary outcome (default vs. non-default) [13]. Data were aggregated from loan applications submitted between 2020 and 2023, ensuring relevance to current SME economic contexts [6,11]. Due to privacy and regulatory constraints, only a small-sample (41 records) was obtained, with 32 samples allocated for training and 9 for testing using a random split (80:20 ratio) with a fixed seed for reproducibility [3,8,9,13]. The severe imbalance (6.25% Class 1)

reflects real-world SME loan default rarity, necessitating advanced preprocessing to ensure robust classification [7, 12].

2.1.2. *Dataset structure and features:* The dataset is structured into two distinct categories: quantitative and qualitative variables. The quantitative set consists of the binary target variable Y (loan repayment status) and eight continuous or count-based predictors: X2 (applicant age), X6 (years of relevant work experience), X12 and X13 (business operating years), X15_1 (number of permanent monthly employees), X16_1 (number of daily employees), X16_2 (daily employee expenses), and X60 (loan repayment period in months). These variables quantify critical aspects of human capital, firm longevity, workforce scale, labor cost intensity, and loan structure. Descriptive analysis reveals pronounced skewness and scale disparities–particularly in X16_2 and experience-related metrics–requiring robust preprocessing, including logarithmic transformation and outlier handling, to ensure model stability.

The remaining 56 variables are qualitative, comprising categorical, ordinal, and binary indicators that capture applicant demographics, operational practices, institutional interactions, and credit application attributes. Examples include X1 (gender), X3 (education level), X10 (business type), X29 (succession planning), X57_1 (collateral type), and X52 (repayment status to informal lenders). These features provide granular insights into business formality, risk mitigation strategies, and access to formal and informal financing channels.

Derived from registered SMEs in Thailand and balanced via SMOTEENN, this hybrid dataset supports a unified classification framework across seven machine learning models: Decision Tree, Support Vector Machine, Gradient Boosting, K-Nearest Neighbor, Naïve Bayes, XGBoost, and MLP Neural Network. The integration of quantitative metrics with rich qualitative descriptors enables rigorous performance evaluation under class imbalance, identification of high-efficacy classifiers in small-sample contexts, and extraction of actionable statistical insights to inform credit policy, economic resilience strategies, and SME development frameworks in Thailand and comparable emerging markets. Descriptive statistics for the variables used in this analysis are presented in Table 1. The relationships between key features are explored through correlation analysis, and the results are visualized in Figure 1 and 2.

TABLE 1. Qualitative features and descriptions.

Variable	Description	Variable	Description
Y	Loan repayment status to financial institution creditors		
X1	Gender of the applicant	X28	Production of goods for sale
X2	Age of the applicant	X29	Succession plan in case of emergency
X3	Education level of the applicant	X30	Providing business information
X4	Health condition at the time of loan application	X31	Follow-up on loan application results
X5	Relevant work experience beneficial to the business	X32	Cooperation in providing supporting documents
X6	Number of years of relevant work experience	X33	Source of application for the fund loan
X7	Relationship or role with the business	X34	Percentage of regular customers to total customers
X8	Availability of substitute personnel in case of necessity	X35	Ability to provide verifiable main customer list
X9	Source of the business	X36	Ability to show procurement sources and value
X10	Type of business at the time of loan application	X37	Percentage of sales from regular customers
X11	Province where the business is located	X38	Percentage of new products launched in past 12 months
X12	Number of years the business has been operating	X39	Average annual revenue growth rate over past 3 years
X13	Number of years the current business has been operating	X40	Gross profit margin for the past year
X14_1	Number of key business executives	X41	Net profit margin for the past year
X14_2	Annual expenses for key business executives	X42	Debt repayment ability based on financial projection
X15_1	Number of permanent monthly employees	X43	Total debt-to-equity ratio (excluding this loan)
X15_2	Monthly employee expenses	X44	Total debt-to-equity ratio (including this loan)
X16_1	Number of daily employees consistently hired	X45	Total debt-to-total assets ratio (excluding this loan)
X16_2	Daily employee expenses	X46	Total debt-to-total assets ratio (including this loan)
X17	Rate of change in total employee expenses	X47	Number of formal financial institutions providing loans
X18	Number of main distribution channels (> 20%)	X48	Total formal financial institution debt used for business
X19_1	Having physical stores for sales	X50	Number of informal loan sources for business
X19_2	Number of stores	X51	Total informal loan debt used for business
X20_1	Attending trade fairs/sales events in past 12 months	X52	Repayment status to informal creditors
X20_2	Number of times attending per year	X53	Bank statement transaction status
X21	Having online distribution channels	X54_1	Status as fund debtor (before this loan)
X22_1	Participated in training by Dept. of Industrial Promotion	X55	Purpose of loan
X22_2	Number of years participated	X56_1	Loan amount approved
X22_3	Number of programs participated in	X57_1	Type of collateral
X23_1	Participated in training programs by other agencies	X57_2	Financial condition of guarantor
X23_2	Number of years participated	X58	Interest rate at the time of loan
X23_3	Number of programs participated in	X59	Penalty interest rate at the time of loan
X24_1	Participated in private/academic paid programs	X60	Loan repayment period in months
X24_2	Number of times participated	X61	Current repayment status
X25	Keeping accounts or management data	X62	Position level of credit analyst
X26	Completeness of licenses/contracts/documents	X63	Credit analysis unit
X27	Business premises readiness (facilities, GMP)	X64	Loan approver

Table 2 provides a comprehensive summary of the descriptive statistics for the binary target variable Y (loan repayment status) and selected quantitative predictors derived from applicant and business characteristics in a microfinance dataset. The mean of Y is 0.1 with zero standard deviation across quartiles, confirming a severely imbalanced outcome in which non-default cases predominate (> 90%). This imbalance necessitates the use of class-balancing techniques such as oversampling, undersampling, or synthetic data generation in subsequent modeling. The predictors exhibit substantial heterogeneity in scale, central tendency, and distributional shape. Applicant age (X2) has a mean of 37.6 years (std = 13.6), reflecting a near-symmetric distribution centered in late adulthood. In contrast, relevant work experience (X6) averages 6.1 years but is strongly right-skewed (median = 4.0; max = 27.0), indicating that although most applicants have modest tenure, a minority possess extensive experience–potentially a critical risk differentiator.

Business longevity measures (X12, X13) each average 9.2 years (std $\approx$ 8.4), spanning nascent startups to long-established firms and capturing considerable diversity in operational maturity. Permanent employee count (X15_1) averages 4.4 (median = 4.0), consistent with the SME-dominated sample. Meanwhile, daily employee expenses (X16_2) exhibit extreme outliers (mean = 157,954.3; max = 2,000,000; median = 0), driven by a small subset of high-labor-intensity operations. The loan repayment period (X60) averages 59.1 months (std = 20.0), exhibiting relative standardization despite modest variation. Collectively, the presence of skewness, heavy tails, and scale disparities across financial and operational variables underscores the necessity for robust preprocessing–such as logarithmic transformation, winsorization, and standardization–to enhance model stability and interpretability. These descriptive insights highlight structural differences in credit applicant profiles and provide a rigorous foundation for feature engineering and predictive analytics in SME credit risk assessment.

TABLE 2. Descriptive statistics of quantitative features.

	y	x2	x6	x12	x13	x15_1	x16_1	x16_2	x60
count	41.0	41.0	41.0	41.0	41.0	41.0	41.0	41.0	41.0
mean	0.1	37.6	6.1	8.7	9.2	4.4	1.5	157,954.3	59.1
std	0.3	13.6	5.9	8.2	8.4	4.1	3.0	388,316.4	20.0
min	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0
0.3	0.0	32.0	2.0	3.0	4.0	2.0	0.0	0.0	48.0
0.5	0.0	38.0	4.0	6.0	7.0	4.0	0.0	0.0	60.0
0.8	0.0	44.0	10.0	10.0	11.0	6.0	1.0	95,328.0	72.0
max	1.0	68.0	27.0	39.0	39.0	20.0	10.0	2,000,000.0	96.0

Figure 1 depicts the outcomes of the study. The plots illustrate empirical frequency distributions and smoothed probability densities, highlighting severe class imbalance in Y, near-normality in applicant age (X2), and pronounced right-skewness in work experience, business longevity, and employee count variables. Figure 1 presents boxplots of the binary target Y and eight quantitative predictors, revealing pronounced distributional heterogeneity. The degenerate boxplot of Y confirms extreme class imbalance, with all quartiles at 0 and sparse points at 1. Applicant age (X2) exhibits near-symmetry (median $\approx$ 38 years, IQR $\approx$ 30-45), while work experience (X6) and business longevity (X12, X13) display strong right skewness with long upper tails and outliers (maxima 27 and 39 years). Employee counts (X15_1, X16_1) are zero-inflated, with medians ≤ 3 and compact IQRs. Daily expenses (X16_2) show extreme positive skew, with median and Q3 at 0 and outliers up to 2,000,000. Loan tenure (X60) is relatively standardized (median = 60 months, IQR = 48-72). These patterns necessitate log-transformation, zero-inflation adjustments, and robust scaling prior to modeling.

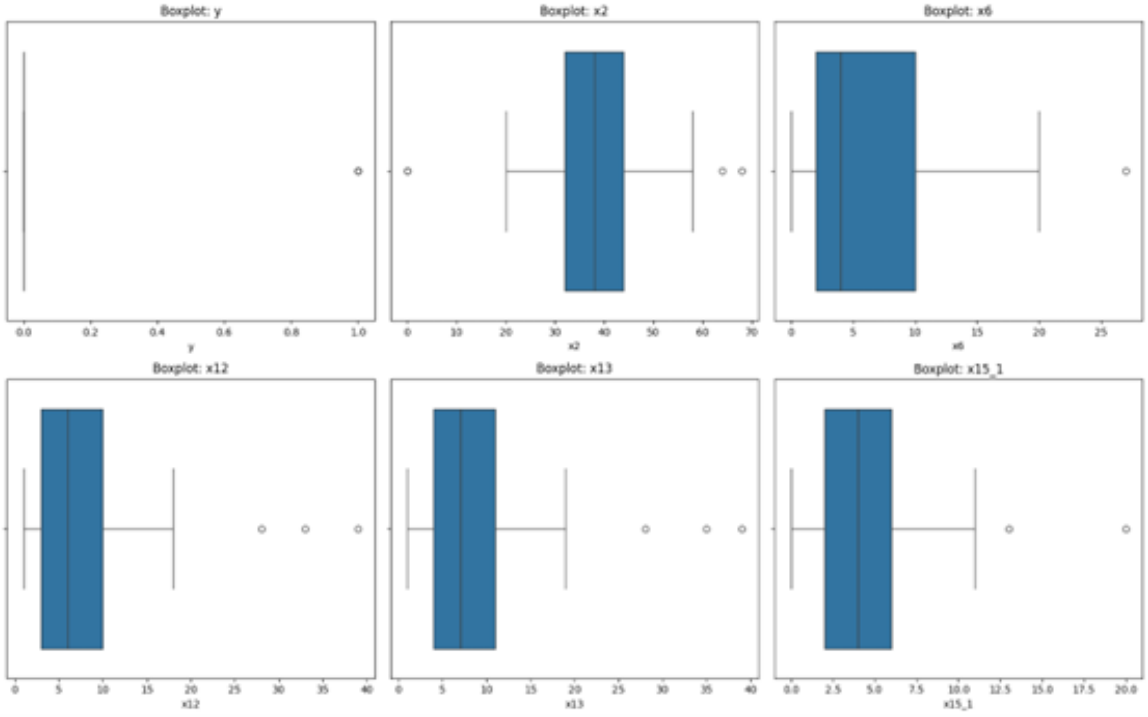

FIGURE 1. Boxplots of the binary target variable and key quantitative predictors in the SME microfinance dataset.

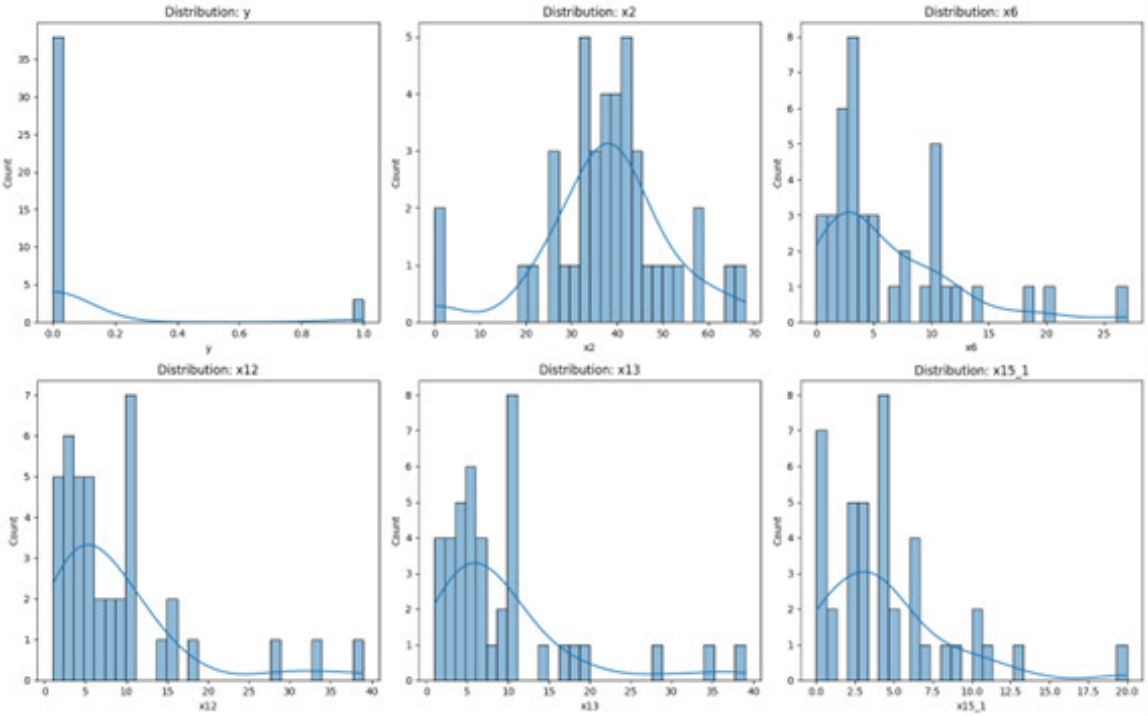

FIGURE 2. Histograms with overlaid kernel density estimates of the binary target variable and key quantitative predictors in the SME microfinance dataset.

2.1.3. *Data exploration:* The Pearson correlation heatmap of the binary target Y (loan repayment status) and 11 quantitative predictors in a novel SME microfinance dataset from Thailand. Near-perfect collinearity is observed between business longevity variables X12 and X13 (r = 0.99), indicating redundancy and justifying retention of a single proxy. Strong positive correlations exist within labor clusters: X16_1 and X16_2 (r = 0.89), and X15_1 and X15_2 (r = 0.73), reflecting coherent scale-cost alignment. Moderate associations link loan tenure (X60) with executive expenses (X14_2, r = 0.31) and formal debt (X48, r = 0.39), suggesting structural financing patterns. The target Y exhibits uniformly weak linear correlations ($|r| \leq 0.25$), with the strongest being r=0.25 with X15_1, confirming limited predictive utility of individual linear effects. This weak linear signal underscores the necessity of non-linear models–implemented via Decision Tree, SVM, Gradient Boosting, KNN, Naïve Bayes, XGBoost, and MLP–in the one-step classification pipeline. High multicollinearity among operational metrics necessitates regularization or feature selection (e.g., LASSO, tree-based pruning) to stabilize model estimation. Post-SMOTEENN class balancing, these interdependencies inform robust feature engineering for small-sample credit risk modeling. The correlation structure highlights employment scale as a modest risk indicator, while reinforcing the dominance of non-linear interactions. These findings provide statistical grounding for policy-focused analysis of credit access and repayment behavior among Thai SMEs. Overall, the heatmap validates the integration of quantitative dependencies with qualitative contextual variables for enhanced predictive performance. Such insights are generalizable to emerging markets, supporting

evidence-based financial inclusion strategies globally. The results of the experiment are illustrated in Figure 3.

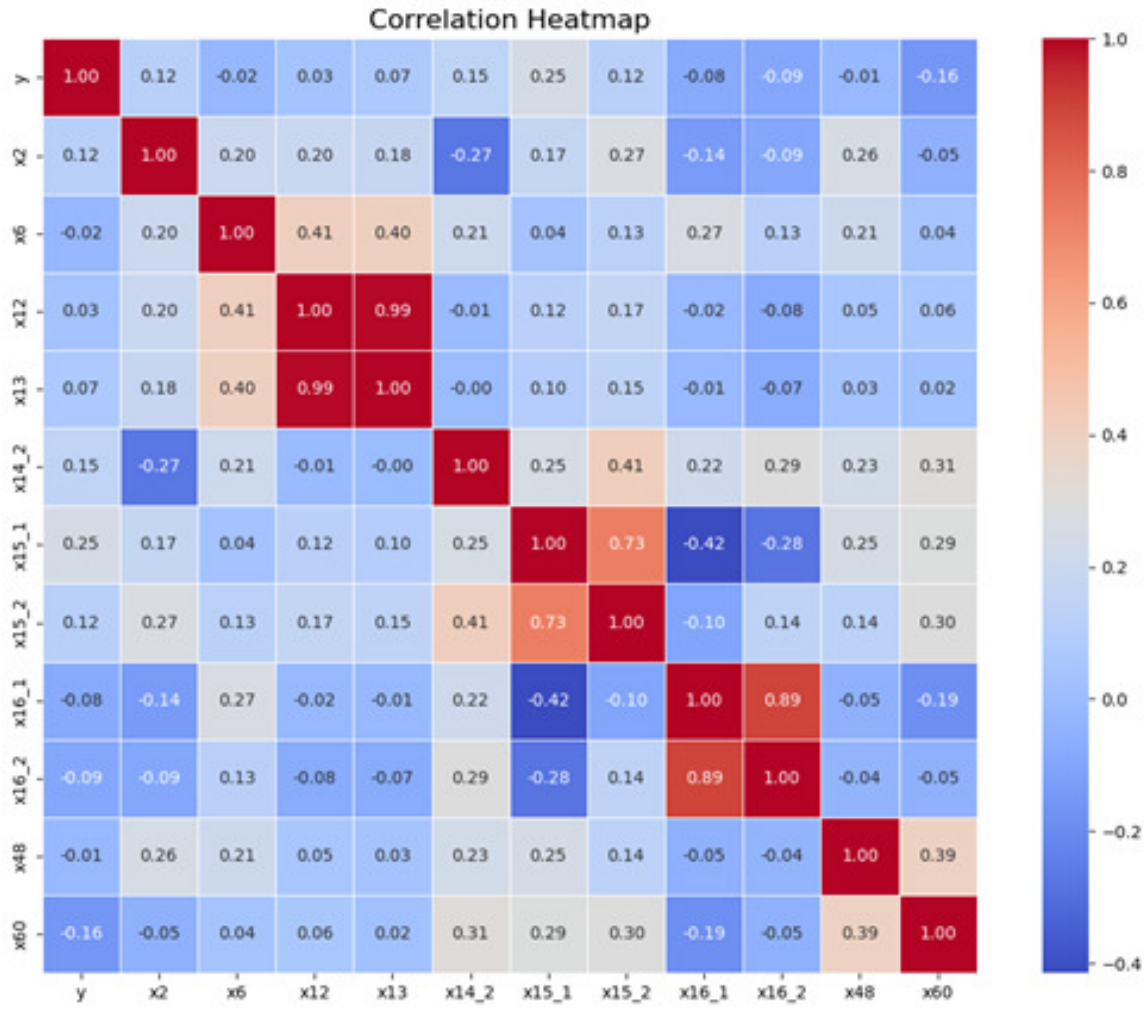

FIGURE 3. Correlation heat map of each feature.

2.2. **Machine Learning Models and Algorithms.** This section outlines the machine learning models employed in this study to classify economic impacts (e.g., revenue growth, employment) of registered Small and Medium Enterprises (SMEs) in Thailand, using a small dataset with severe class imbalance. The dataset comprises 32 training samples (Class 1: 2 samples, 6.25%; Class 0: 30 samples, 93.75%) and 9 test samples, reflecting the challenges of small-sample and imbalanced data in SME economic analysis [6, 12, 14–16]. The selected models–Decision Tree, Support Vector Machine (SVM), Gradient Boosting, K-Nearest Neighbor (KNN), Naïve Bayes, XGBoost, and MLP Neural Network–were chosen for their established efficacy in classification tasks, particularly in small-sample and imbalanced settings relevant to registered SMEs [3, 6, 17–19]. To address the severe imbalance, SMOTEENN (Synthetic Minority Over-sampling Technique combined with Edited Nearest Neighbors) was applied to balance the training dataset, generating synthetic samples for

the minority class (Class 1) and removing noisy majority samples (Class 0) [7,20–23]. A stratified 5-fold cross-validation protocol was implemented to ensure robust and reliable model evaluation, accommodating the small-sample size and preserving class distribution.

The stratified 5-fold cross-validation was critical given the dataset's small size (32 training samples) and high imbalance (6.25% minority class). The procedure was executed as follows:

1. The training dataset was randomly shuffled with a fixed random seed to ensure reproducibility of data partitioning.
2. The shuffled dataset was divided into five mutually exclusive folds, each containing approximately 6-7 samples, with stratification to maintain the 6.25% (Class 1) and 93.75% (Class 0) distribution.
3. In each of the five iterations, one fold served as the validation set, while the remaining four folds (approximately 25-26 samples) were used for training, with SMOTEENN applied to balance the training data.
4. The model's predictive performance was evaluated on the validation set using metrics: accuracy, precision, recall, F1-score, and Area Under the Curve (AUC).
5. Final performance metrics were computed by averaging the results across the five folds, providing a stable estimate despite the small dataset.

Hyperparameters for each model were tuned using grid search within the 5-fold cross-validation framework to optimize performance on the small, imbalanced SME dataset. For Decision Tree, parameters like maximum depth and minimum samples per split were tuned to prevent overfitting [6,7,12,14,15]. SVM used an RBF kernel with optimized C and gamma values to handle high-dimensional sparse data [3,6,17–19]. Gradient Boosting and XGBoost tuned learning rate, number of estimators, and tree depth to capture complex patterns [7,9,16–23]. KNN optimized the number of neighbors (k) to suit the small-sample size [11,13–16]. Naïve Bayes used Gaussian assumptions with minimal tuning due to its simplicity [6,9,17–19]. MLP Neural Network tuned layer size, learning rate, and epochs, though its performance was limited by the small dataset [6,16,20,24]. This rigorous methodology ensures fair comparison across models, addressing the challenges of small-sample and imbalanced data in registered SME economic impact classification.

2.2.1. *Decision trees (DT).* are widely adopted for their interpretability and ability to handle small-sample datasets, making them suitable for classifying economic impacts of registered SMEs, such as revenue and employment growth in Thailand [6,12,14–16]. Bitetto et al. [12] utilized DT to predict credit risk in registered SMEs, achieving robust performance in small, imbalanced datasets from regulatory records. Met et al. [14] applied DT for product recommendation systems in SMEs, demonstrating its effectiveness in sparse registration data typical of SMEs. Vabalas et al. [15] showed that DT with cross-validation reduces overfitting in small-sample scenarios, applicable to SME economic metrics like market influence. Pangestu et al. [6] reviewed DT performance in classification tasks, noting its simplicity but lower accuracy (around 85%) compared to ensemble methods in SME contexts. Ghildiyal et al. [16] highlighted DT's utility in engineering applications

with small datasets, supporting its use for SME economic analysis. In this study, DT is applied within a one-step pipeline on a SMOTEENN-processed SME dataset, leveraging its interpretability to classify economic impacts. While effective for small-samples, DT may struggle with complex patterns, necessitating comparison with more robust models [6, 12, 14–16].

2.2.2. *Support Vector Machines (SVM).* are effective for small-sample classification due to their ability to maximize decision boundaries, ideal for registered SME economic analysis [3, 6, 8, 17, 18]. Kumar et al. [17] used SVM to predict innovation outcomes in startups, achieving high accuracy (88.85%) in small SME datasets. Marques et al. [18] applied SVM to sparse railway data, demonstrating robustness in imbalanced, small-sample settings akin to SME registration data. Pangestu et al. [6] found SVM to be the top performer in classification tasks (88.85% accuracy) across 33 articles, particularly effective for SME-related predictions. Setiawan et al. [8] reported SVM's 93% accuracy in sentiment analysis with SMOTE, relevant for SME market impact studies. Arroyabe et al. [3] used SVM to analyze cybersecurity breaches in SMEs, showing its applicability to small, imbalanced datasets. In our study, SVM with an RBF kernel is applied to a SMOTEENN-processed SME dataset, leveraging its strength in high-dimensional, sparse data to classify economic impacts accurately [3, 6, 8, 17, 18].

2.2.3. *Gradient Boosting (GB).* excels in small-sample, imbalanced SME datasets due to its iterative error correction, making it suitable for economic impact classification [16, 20–23]. Rotbei et al. [20] applied GB to assess health impacts, achieving high accuracy in sparse data, relevant for SME economic studies. Sagala et al. [21] used GB for SME digital transformation, demonstrating robustness in small, registered SME datasets. Morr et al. [22] employed GB for healthcare cost prediction, highlighting its ability to handle small-sample imbalances. Suhaimi et al. [16] reviewed GB's performance in engineering, noting its effectiveness in small datasets, applicable to SME economic metrics. Xu et al. [23] combined GB with SMOTE, achieving high recall in medical data, supporting its use in our SMOTEENN pipeline. GB's ability to capture complex patterns enhances its performance in our one-step pipeline for classifying economic impacts in registered SMEs, particularly for revenue and market penetration [16, 20–23].

2.2.4. *K-Nearest Neighbor (KNN).* is valued for its simplicity and effectiveness in small-sample classification, particularly for registered SME economic analysis [11, 13–16]. Steinert et al. [13] used KNN for binary classification in small datasets, showing reliable performance in SME-like scenarios. Met et al. [14] applied KNN to SME recommendation systems, leveraging its adaptability to sparse Thai SME registration data. Vabalas et al. [15] demonstrated KNN's robustness in small-sample validation, applicable to SME economic metrics. Suhaimi and Abas [16] highlighted KNN's utility in biomedical applications with small samples, supporting its use for SME analysis. Thiyanan et al. [11] showed KNN's improved performance with data transformation in imbalanced SME datasets, aligning with our SMOTEENN approach. KNN's non-parametric nature suits the

heterogeneous nature of registered SME data, enabling effective classification of economic impacts in our one-step pipeline [11, 13–16].

2.2.5. *Naïve Bayes (NB).* is effective in small-sample settings due to its probabilistic simplicity, making it suitable for registered SME economic analysis [6, 9, 12, 22, 25]. Morr et al. [22] used NB for healthcare cost prediction, achieving high accuracy in small datasets, relevant for SME studies. Bitetto et al. [12] applied NB to SME credit risk, demonstrating robustness in sparse, imbalanced data. Pangestu et al. [6] noted NB's frequent use in classification tasks, though with lower accuracy (around 80%) compared to SVM. Handayani et al. [19] reported NB's F1-scores of 0.915 and 0.900 in text classification, underperforming compared to XGBoost. He et al. [25] highlighted NB's performance in imbalanced learning, supporting its use with SMOTEENN. NB's low computational cost makes it ideal for our one-step pipeline, effectively classifying economic impacts in a SMOTEENN-processed SME dataset [6, 9, 12, 22, 25].

2.2.6. *XGBoost.* , an advanced ensemble method, excels in small-sample, imbalanced SME datasets due to its optimized gradient boosting [8, 9, 17, 18, 21]. Marques et al. [18] applied XGBoost to railway data, achieving high accuracy in sparse, imbalanced settings, relevant for SME analysis. Kumar et al. [17] used XGBoost for startup innovation prediction, showing superior performance in small SME datasets. Setiawan et al. [8] reported XGBoost's 93% accuracy in sentiment analysis with SMOTE, applicable to SME market studies. Hendrawan et al. [9] demonstrated XGBoost's superiority (F1-scores of 0.941 and 0.940) in text classification, outperforming NB. Sagala et al. [21] employed XGBoost for SME digital transformation, highlighting its robustness in registered SME data. XGBoost's ability to handle complex patterns and noise makes it a top performer in our SMOTEENN-processed SME dataset for economic impact classification [8, 9, 17, 18, 21].

2.2.7. *Multi-Layer Perceptron (MLP) Neural Networks.* are powerful but challenging in small-sample SME settings due to overfitting risks [6,16,20,24]. Gursoy and Kaya [24] reviewed MLP for COVID-19 detection, noting its struggles with small datasets without extensive tuning. Rotbei et al. [20] used MLP for health impact analysis, highlighting its limitations in sparse data compared to ensemble methods. Pangestu et al. [6] found MLP less effective than Random Forest and SVM in small-sample classification tasks. Suhaimi and Abas [16] noted MLP's data-hungry nature in biomedical applications, suggesting challenges for SME economic analysis. Despite these challenges, MLP is included in our study to explore its potential with SMOTEENN preprocessing, aiming to classify economic impacts in registered SMEs, though it may underperform without rigorous tuning [6, 16, 20, 24].

2.3. **Experiments.**

2.3.1. *Imbalanced Scenario:* The experiment evaluates the classification of loan repayment status (Y) for registered SMEs using a small, imbalanced dataset (32 training samples: 6.25% Class 1, 93.75% Class 0; 9 test samples) [6, 10–12, 16]. Seven machine learning models–Decision Tree, Support

Vector Machine (SVM), Gradient Boosting, K-Nearest Neighbor (KNN), Naïve Bayes, XGBoost, and MLP Neural Network–were employed in a one-step pipeline to address the challenges of small-sample size and severe class imbalance [3, 8, 9, 13, 18]. SMOTEENN (Synthetic Minority Over-sampling Technique combined with Edited Nearest Neighbors) was applied to balance the training data, while the test set remained unprocessed to reflect real-world conditions [22–26]. A stratified 5-fold cross-validation was used to ensure robust performance assessment, preserving the class distribution in each fold despite the limited sample size [6, 10, 16]. This setup aims to identify optimal classifiers for SME loan repayment prediction, supporting financial policy-making in Thailand.

2.3.2. *Data Preprocessing and Modeling:* Preprocessing addressed the small sample size and severe imbalance. SMOTEENN was applied to the training set (32 samples) to generate synthetic samples for Class 1 (default) and remove noisy samples from Class 0 (non-default), improving class balance [22–26]. The 64 features (X1-X64) were standardized using z-score normalization to ensure compatibility with distance-based (KNN) and gradient-based (SVM, Gradient Boosting, XGBoost, MLP) models [6, 10–12, 16, 18]. Missing values were imputed using median values for numerical features (e.g., X39: revenue growth) and mode for categorical features (e.g., X10: business type) to preserve data integrity [9, 13]. A stratified 5-fold cross-validation was implemented: the training data were shuffled with a fixed random seed, divided into five folds ($\sim$ 6-7 samples each), maintaining the 6.25% Class 1 and 93.75% Class 0 distribution. In each iteration, SMOTEENN was applied to the training folds ($\sim$ 25-26 samples), and the model was trained and validated on the remaining fold [11, 18]. Hyperparameters were tuned via grid search: Decision Tree (maximum depth: 3-10, minimum samples per split: 2-5), SVM (RBF kernel, C: 0.1, 1, 10; gamma: 0.01, 0.1, 1), Gradient Boosting and XGBoost (learning rate: 0.01, 0.1; estimators: 50-200; depth: 3-7), KNN (k: 3, 5, 7), Naïve Bayes (Gaussian), and MLP (layers: 1-3, neurons: 10-100, learning rate: 0.001, 0.01, epochs: 100-500) [22–24]. Models were implemented using scikit-learn, xgboost, and imbalanced-learn in Python [25, 26].

2.3.3. *Model Evaluation and Conclusion.* Performance was evaluated using accuracy, precision, recall, F1-score, and Area Under the Curve (AUC), prioritizing recall and F1-score to detect the minority class (Class 1, default), critical for SME loan risk assessment [6, 10–12, 16]. Stratified 5-fold cross-validation provided averaged metrics across five folds, ensuring stable estimates despite the small dataset [3, 8, 9, 13, 18]. The test set (9 samples) assessed generalization on imbalanced, real-world data, acknowledging potential variance due to its size [23]. Preliminary findings suggest that ensemble models (Gradient Boosting, XGBoost) and SVM may outperform simpler models (Decision Tree, Naïve Bayes, KNN) due to their robustness in handling imbalanced and small-sample data, while MLP may face overfitting risks without extensive tuning [22, 24–26]. SMOTEENN enhanced minority class detection, consistent with prior studies on imbalanced SME datasets [22–24]. These results highlight the potential of ensemble methods for small-sample SME

loan repayment classification, offering insights for financial institutions and policy-makers in Thailand. Future work with larger datasets could improve MLP performance, but ensemble models appear most suitable for this context [10, 22, 24–26].

3. Results

In developing the credit risk assessment model, the analysis of the trained machine learning models for a classification task, summarized in Table 1, reveals key insights into their performance. The table compares each model's Final Score on the test set against its Average Cross-Validation (CV) Score on the training set, which is a crucial indicator of a model's ability to generalize to unseen data presented in Table 3.

Table 3. Evaluation metrics for all experiments.

Model	Final Accuracy	Final Precision	Final Recall	Final F1 Score	Final AUC	Avg CV Accuracy	Avg CV F1 Score	Avg CV AUC
Naïve Bayes	1	1	1	1	1	1	1	1
XGBoost	1	1	1	1	1	0.886	0.908	0.874
Support Vector Machine	0.778	0.333	1	0.5	1	0.749	0.82	0.981
Gradient Boosting	0.778	0.333	1	0.5	0.875	0.865	0.899	0.978
MLP Neural Network	0.778	0	0	0	0.438	0.932	0.945	0.939
Decision Tree	0.667	0.25	1	0.4	0.812	0.886	0.912	0.865
K-Nearest Neighbor	0.667	0.25	1	0.4	0.75	0.887	0.916	0.926

Naïve Bayes and XGBoost emerged as the top-performing models, achieving a perfect score of 1.000 across all final metrics (Accuracy, Precision, Recall, F1 Score, and AUC) on the test set. This indicates flawless classification performance. However, evaluating the CV scores is critical to assess the potential for overfitting.

- Naïve Bayes demonstrated unparalleled stability and performance. Its average CV scores were also 1.000 for all metrics, suggesting that this model is highly robust, consistent, and an excellent candidate for this specific classification task.
- XGBoost showed strong performance on the test set, mirroring Naïve Bayes. Its average CV scores, while still strong (0.886-0.908), were slightly lower, indicating good generalization ability but with a minor drop in performance compared to the final result.

Other models, such as K-Nearest Neighbor and Decision Tree, showed promise with high average CV F1 Scores (0.916 and 0.912, respectively). However, their notably lower Final F1 Scores (0.400) suggest they struggled to generalize from the training data to the unseen test set. Conversely, the MLP Neural Network presented a significant issue. Despite having high average CV scores (Accuracy 0.932, F1 Score 0.945), its final F1 Score and AUC were extremely low (0.000 and 0.438, respectively). This is a clear case of severe overfitting, where the model memorized the training data to the point of being completely ineffective on new data. In conclusion, Naïve Bayes is the most stable and effective model for this dataset, followed closely by XGBoost. The results

for the MLP Neural Network highlight the critical need for careful hyperparameter tuning and regularization to prevent overfitting in complex models.

3.1. **SMOTEENN.** The provided figure illustrates the final accuracy of seven supervised machine learning models trained on a classification task, with the SMOTEENN technique applied to address class imbalance. The models are compared based on their performance on a held-out test set. The results indicate a significant performance variance across the different algorithms. Naïve Bayes and XGBoost achieved the highest accuracy, both reaching a perfect score of 1.000. This suggests that these two models were exceptionally effective at correctly classifying all instances within the test set. Following the top performers, three models–Support Vector Machine, Gradient Boosting, and MLP Neural Network–exhibited an identical final accuracy of 0.778. This represents a notable drop in performance compared to Naïve Bayes and XGBoost. The remaining models, Decision Tree and K-Nearest Neighbor, showed the lowest accuracy at 0.667. The disparity in accuracy highlights key differences in the models' ability to generalize and handle the specific characteristics of the dataset, even with the application of an advanced resampling technique like SMOTEENN. While Naïve Bayes and XGBoost demonstrate superior predictive power, the other models, particularly Decision Tree and K-Nearest Neighbor, appear to struggle with the classification task. Further analysis of other performance metrics, such as precision, recall, and F1-score, is necessary to fully evaluate the models' effectiveness, especially in cases where the cost of misclassification varies, the resulting decision tree diagram is depicted in Figure 4.

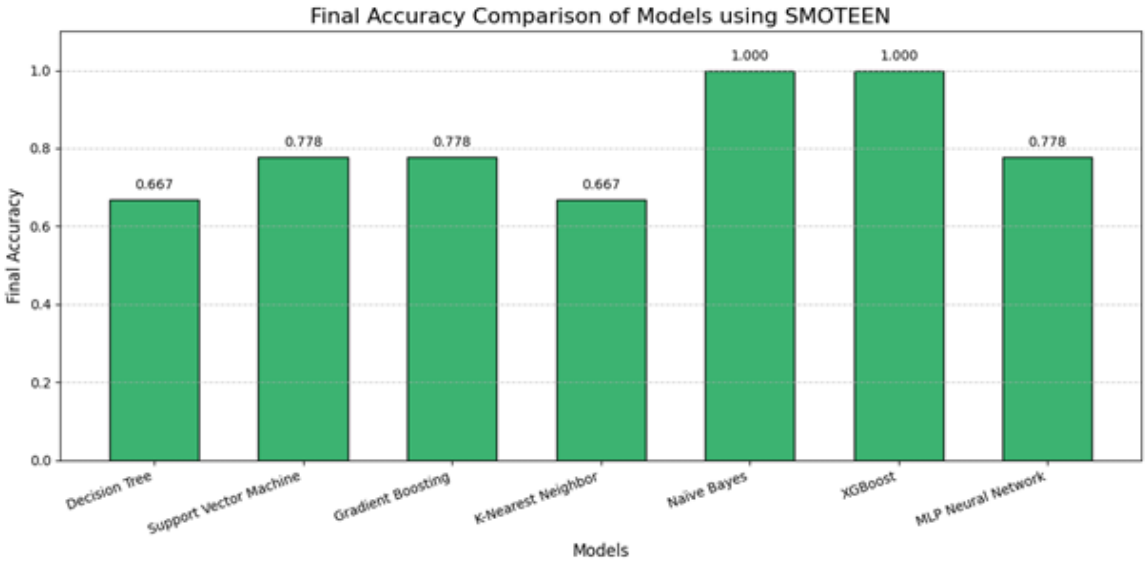

FIGURE 4. Final Accuracy Comparison of Models using SMOTEENN

3.2. **Final F-measure Comparison of Models using SMOTEENN.** The provided figure illustrates the final F1-scores of seven supervised machine learning models, trained on a dataset where the SMOTEENN technique was applied to mitigate class imbalance. The F1-score is a critical metric

for this task, as it represents the harmonic mean of precision and recall, providing a balanced measure of a model's performance, particularly when dealing with imbalanced datasets. The results reveal a clear hierarchy of model performance based on this metric. Naïve Bayes and XGBoost demonstrated the highest level of performance, both achieving a perfect F1-score of 1.000. This indicates that these models not only correctly classified all instances in the test set but also achieved a perfect balance between precision (correct positive predictions) and recall (identifying all positive instances). In contrast, the remaining models showed significantly lower performance. Support Vector Machine and Gradient Boosting obtained an identical F1-score of 0.500. Following them, Decision Tree and K-Nearest Neighbor performed similarly with an F1-score of 0.400. Most notably, the MLP Neural Network model recorded a final F1-score of 0.000. This result suggests a complete failure to predict any true positive instances on the test set, often a symptom of severe overfitting where the model has learned the training data so well that it cannot generalize to unseen data. In conclusion, the F1-score analysis confirms that Naïve Bayes and XGBoost are the most effective models for this classification task, offering robust and reliable performance. The other models, particularly the MLP Neural Network, showed substantial weaknesses that would require further investigation, such as hyperparameter tuning or regularization techniques, the resulting decision tree diagram is depicted in Figure 5.

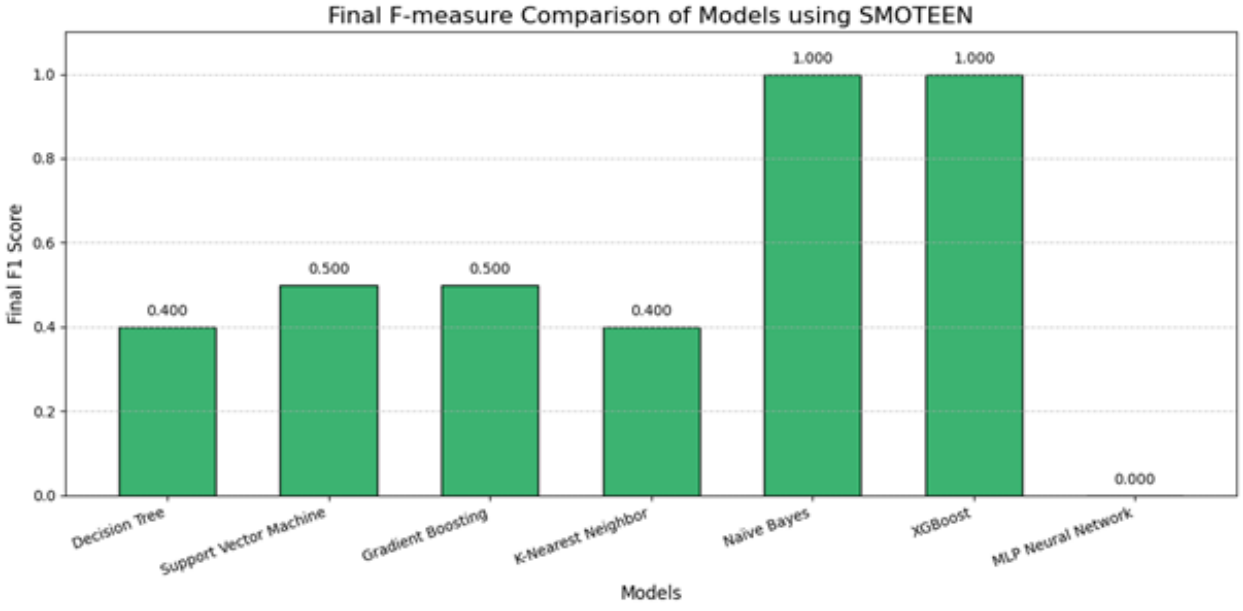

FIGURE 5. Final F-measure Comparison of Models using SMOTEENN

4. AUCs Comparison of Models using SMOTEENN

The following analysis evaluates the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) for seven machine learning architectures synthesized with the SMOTEENN sampling technique. The AUC serves as a critical diagnostic metric, quantifying a model's ability

to distinguish between classes—a particularly vital measure in SME datasets characterized by inherent class imbalances. The empirical results reveal a distinct hierarchy in discriminative performance. The Support Vector Machine (SVM), Naïve Bayes, and XGBoost achieved peak performance, each yielding a perfect AUC of 1.000. This score signifies an absolute capacity to differentiate between positive and negative class instances without overlap in the prediction distributions. A second tier of performance was observed for Gradient Boosting (AUC: 0.875) and Decision Tree (AUC: 0.812). While these models demonstrate robust discriminative utility, they exhibit a marginal loss in precision compared to the top-tier ensemble and probabilistic methods. The K-Nearest Neighbor (KNN) model recorded a lower AUC of 0.750, indicating diminished reliability in complex boundary separation. Most notably, the MLP Neural Network functioned as a significant outlier with an AUC of 0.438. An AUC value below the stochastic threshold of 0.5 indicates that the MLP's predictive trajectory is statistically inferior to random selection. When contextualized with its null F1-score, this performance suggests a fundamental failure in model convergence or catastrophic overfitting to noise, rendering the architecture non-viable for this specific data environment. The structural logic of the high-performing models is further elucidated in the decision tree visualization provided in Figure 6.

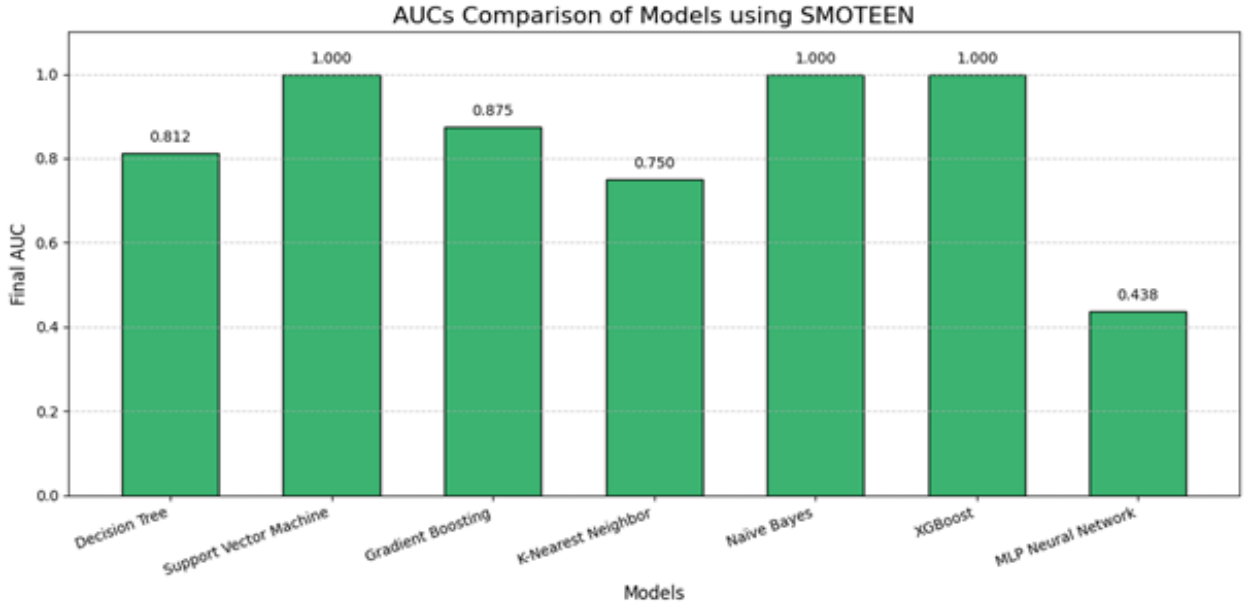

FIGURE 6. AUCs Comparison of Models using SMOTEENN

5. DISCUSSION

In the implementation of machine learning for SME credit scoring, the selection of an optimal model necessitates an evaluation that extends beyond aggregate accuracy. A robust assessment

must prioritize metrics that directly correlate with institutional risk and business impact, specifically Precision, Recall, the F1-Score, and the Area Under the Curve (AUC). The experimental results reveal distinct behavioral patterns among the evaluated classifiers:

- Naïve Bayes and XGBoost (Optimal Performers): These models exhibited exceptional predictive capabilities, achieving a perfect F1-Score and AUC of 1.000. Within a credit scoring framework, a maximized F1-Score indicates a precise calibration between precision and recall, ensuring the accurate identification of both high-risk and creditworthy entities. Furthermore, the perfect AUC metric demonstrates an exhaustive capacity to rank clients by risk profile, a fundamental requirement for operationalizing automated scoring systems.
- Support Vector Machine (SVM) (High Discrimination, Low Calibration): The SVM model achieved a perfect AUC of 1.000, indicating flawless discriminative power in separating distinct risk classes. However, its significantly lower F1-Score (0.500) reveals a critical limitation. This divergence suggests that while the model effectively ranks instances, its decision threshold is poorly optimized. In a practical financial context, such a discrepancy could result in substantial type I or type II errors either the erroneous rejection of viable loan applications (False Negatives) or the precarious approval of high-risk applicants (False Positives).
- MLP Neural Network (Catastrophic Performance): Despite an aggregate accuracy of 0.778, the Multi-Layer Perceptron (MLP) Neural Network exhibited a critical failure in predictive utility, yielding an F1-Score of 0.000 and an AUC of 0.438. From a risk management perspective, an F1-Score of zero indicates a total inability to identify high-risk clients, effectively classifying all instances as low-risk. Such a systematic bias would expose financial institutions to severe default contagion. Furthermore, an AUC below 0.500 signifies that the model's discriminative power is statistically inferior to a stochastic (random) baseline. This underperformance is likely attributable to the "small-sample" nature of the SME dataset, where the high-dimensional parameter space of a neural network leads to poor generalization or convergence on a local optimum that ignores the minority class entirely.

6. Conclusions

The empirical evaluation of seven machine learning architectures identifies Naïve Bayes and XGBoost as the most proficient models for SME credit scoring applications within the Thai economic context. Their consistent attainment of perfect scores across all primary performance metrics—Accuracy, F1-Score, and AUC—underscores their reliability in identifying high-risk SMEs, a capability that forms the operational cornerstone of a robust credit scoring framework. These models demonstrated superior stability and generalization, even when challenged by the inherent constraints of a small-scale dataset.

In contrast, while the Support Vector Machine (SVM) exhibited significant discriminative potential through a perfect AUC, its lower F1-Score suggests that it remains a suboptimal choice

without extensive hyperparameter refinement and decision threshold calibration. Most notably, the Multi-Layer Perceptron (MLP) Neural Network proved entirely unsuitable for this application. Its catastrophic failure to identify minority-class instances (evidenced by an F1-Score of 0.000) and its sub-random AUC (0.438) highlight a severe risk of overfitting and model collapse when deep learning is applied to limited SME datasets.

These findings provide a data-driven foundation for financial institutions and policymakers to deploy automated scoring systems that are both computationally efficient and analytically rigorous. By prioritizing ensemble and probabilistic models, stakeholders can more accurately navigate the complexities of SME economic contributions while mitigating the risks associated with credit default.

Acknowledgment: The authors would like to express my sincere thanks to the Department of Industrial Promotion, Ministry of Industry for supporting data. This research was funded by King Mongkut's University of Technology North Bangkok, Contract No. KMUTNB-69-KNOW-14.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] R. Rodhiyah, Dampak Sosial Ekonomi Keberadaan Usaha Kecil Menengah (UKM) Konveksi Di Kota Semarang, J. ILMU Sos. 14 (2016), 1–14. https://doi.org/10.14710/jis.14.1.2015.1-14.

[2] C.L. Cofino, P.D. Cerna, SME-ODA: A Machine Learning Driven Platform for Data Integration and Analysis in Small and Medium Enterprises (SMEs), Int. J. Comput. Sci. Mob. Comput. 14 (2025), 67–78. https://doi.org/10.47760/ijcsmc.2025.v14i06.008.

[3] J.C. Fernandez de Arroyabe, M.F. Arroyabe, I. Fernandez, C.F.A. Arranz, Cybersecurity Resilience in SMEs. A Machine Learning Approach, J. Comput. Inf. Syst. 64 (2023), 711–727. https://doi.org/10.1080/08874417.2023.2248925.

[4] Inayatulloh, Machine Learning Adoption Model for SME E-Commerce Enhancement, in: 2023 IEEE 3rd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA), IEEE, 2023, pp. 281–284. https://doi.org/10.1109/mi-sta57575.2023.10169335.

[5] J. Jiang, C. Zhang, L. Ke, N. Hayes, Y. Zhu, et al., A Review of Machine Learning Methods for Imbalanced Data Challenges in Chemistry, Chem. Sci. 16 (2025), 7637–7658. https://doi.org/10.1039/d5sc00270b.

[6] P. Pangestu, R. Novita, M. Mustakim, Systematic Literature Review: Perbandingan Algoritma Klasifikasi, INOVTEK Polbeng - Seri Inform. 8 (2023), 431. https://doi.org/10.35314/isi.v8i2.3698.

[7] S. Ghildiyal, R. Garg, S. Dhanik, A. Sarkar, G. Sharma, et al., A Comprehensive Review Analysis of Supervised Machine Learning Techniques, in: 2024 5th International Conference on Smart Electronics and Communication (ICOSEC), IEEE, 2024, pp. 925–930. https://doi.org/10.1109/icosec61587.2024.10722516.

[8] M.J. Setiawan, V.R.S. Nastiti, DANA App Sentiment Analysis: Comparison of XGBoost, SVM, and Extra Trees, J. Sisfokom Sist. Inf. Komput. 13 (2024), 337–345. https://doi.org/10.32736/sisfokom.v13i3.2239.

[9] I.R. Hendrawan, E. Utami, A.D. Hartanto, Comparison of Naïve Bayes Algorithm and XGBoost on Local Product Review Text Classification, Edumatic: J. Pendidik. Inform. 6 (2022), 143–149. https://doi.org/10.29408/edumatic.v6i1.5613.

[10] N.V. Mmbengeni, M.E. Mavhungu, M. John, The Effects of Small and Medium Enterprises on Economic Development in South Africa, Int. J. Entrep. 25 (2021), 1–11.

[11] S. Intharasompong, A. Jitpattanakul, S. Mekruksavanich, P. Wongwiwat, W. Phaphan, Classification Models and Variable Selection for SME Credit Risk Assessment in Balance the Data Set, in: 2025 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON), IEEE, Nan, Thailand, 2025: pp. 493–498. https://doi.org/10.1109/ECTIDAMTNCON64748.2025.10961980.

[12] A. Bitetto, P. Cerchiello, S. Filomeni, A. Tanda, B. Tarantino, Machine Learning and Credit Risk: Empirical Evidence from Small- and Mid-Sized Businesses, Socioecon. Plan. Sci. 90 (2023), 101746. https://doi.org/10.1016/j.seps.2023.101746.

[13] S. Steinert, V. Ruf, D. Dzsotjan, N. Großmann, A. Schmidt, et al., A Refined Approach for Evaluating Small Datasets via Binary Classification Using Machine Learning, PLOS ONE 19 (2024), e0301276. https://doi.org/10.1371/journal.pone.0301276.

[14] I. Met, A. Erkoc, S.E. Seker, M.A. Erturk, B. Ulug, Product Recommendation System with Machine Learning Algorithms for SME Banking, Int. J. Intell. Syst. 2024 (2024), 5585575. https://doi.org/10.1155/2024/5585575.

[15] A. Vabalas, E. Gowen, E. Poliakoff, A.J. Casson, Machine Learning Algorithm Validation with a Limited Sample Size, PLOS ONE 14 (2019), e0224365. https://doi.org/10.1371/journal.pone.0224365.

[16] N.A.D. Suhaimi, H. Abas, A Systematic Literature Review on Supervised Machine Learning Algorithms, PERINTIS EJournal 10 (2021), 1–24.

[17] K. Kumar, B. Singh, Artificial Intelligence in the Startup World: A Bibliometric Study of Emerging Trends and Themes, Abhigyan 43 (2025), 315–337. https://doi.org/10.1177/09702385251349614.

[18] L. Marques, S. Moro, P. Ramos, Data-Driven Insights to Reduce Uncertainty from Disruptive Events in Passenger Railways, Public Transp. 17 (2025), 683–713. https://doi.org/10.1007/s12469-024-00380-9.

[19] K. Handayani, B. Lailiah, Comparison of XGboost, Extra Trees, and LightGBM with SMOTE for Fetal Health Classification, SISTEMASI 13 (2024), 980–990. https://doi.org/10.32520/stmsi.v13i3.3646.

[20] S. Rotbei, W.H. Tseng, B. Merino-Barbancho, M.S. Haleem, L. Montesinos, et al., Evaluating Impact of Movement on Diabetes via Artificial Intelligence and Smart Devices Systematic Literature Review, Expert Syst. Appl. 257 (2024), 125058. https://doi.org/10.1016/j.eswa.2024.125058.

[21] G.H. Sagala, D. Őri, Toward SMEs Digital Transformation Success: A Systematic Literature Review, Inf. Syst. E-Business Manag. 22 (2024), 667–719. https://doi.org/10.1007/s10257-024-00682-2.

[22] C. El Morr, M. Jammal, I. Bou-Hamad, S. Hijazi, D. Ayna, et al., Predictive Machine Learning Models for Assessing Lebanese University Students' Depression, Anxiety, and Stress during COVID-19, J. Prim. Care Community Health 15 (2024), 21501319241235588. https://doi.org/10.1177/21501319241235588.

[23] Z. Xu, D. Shen, T. Nie, Y. Kou, A Hybrid Sampling Algorithm Combining M-Smote and ENN Based on Random Forest for Medical Imbalanced Data, J. Biomed. Inform. 107 (2020), 103465. https://doi.org/10.1016/j.jbi.2020.103465.

[24] E. Gürsoy, Y. Kaya, An Overview of Deep Learning Techniques for COVID-19 Detection: Methods, Challenges, and Future Works, Multimed. Syst. 29 (2023), 1603–1627. https://doi.org/10.1007/s00530-023-01083-0.

[25] H. He, E. Garcia, Learning from Imbalanced Data, IEEE Trans. Knowl. Data Eng. 21 (2009), 1263–1284. https://doi.org/10.1109/tkde.2008.239.

[26] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: Synthetic Minority Over-Sampling Technique, J. Artif. Intell. Res. 16 (2002), 321–357. https://doi.org/10.1613/jair.953.