The Economics and Finance Letters
2025 Vol. 12, No. 4, pp. 814-831
ISSN(e): 2312-430X
ISSN(p): 2312-6310
DOI: 10.18488/29.v12i4.4622

A machine learning approach for credit risk assessment of SMEs: Evidence from Morocco

Jalila Boumhidi[1+]
Abellatif Marghich[2]

[1,2]*University Sidi Mohamed Ben AbdellahFez., Morocco.*
[1]*Email: jalila.boumhidi@usmba.ac.ma*
[2]*Email: abdellatif.marghich@usmba.ac.ma*

(+ Corresponding author)

Article History
Received: 15 September 2025
Revised: 18 November 2025
Accepted: 9 December 2025
Published: 24 December 2025

Keywords
Logistic regression
Machine learning
Random forest
Small and medium-sized enterprises
XG boost.

JEL Classification:
C53.

ABSTRACT

Assessing the credit risk of small and medium-sized enterprises (SMEs) has become increasingly complex as borrower profiles are diverse and often non-linear. Traditional rating methods, still widely used in practice, struggle to capture this variability, which can limit their reliability in modern financial contexts. The objective of this study is to evaluate whether machine-learning techniques can provide more accurate and operationally useful tools for SME credit-risk assessment. Using a dataset of 124 Moroccan SMEs, containing financial, behavioral, and transactional variables, we applied three supervised classification models: logistic regression, random forest, and XGBoost, to predict contract defaults. The models are assessed through standard performance metrics including accuracy, precision, recall, F1-score, and AUC. Results demonstrate that XGBoost provides the strongest detection of defaults, eliminating false negatives in our test set and making it especially suitable for loss-minimization contexts. By contrast, random forest achieves the highest discrimination between risky and non-risky profiles (AUC = 0.93), offering a balanced solution for operational scoring. Logistic regression, while less accurate, retains value for its interpretability and transparency. Overall, the findings highlight that ensemble methods, particularly XGBoost, can significantly improve the reliability of SME credit-risk evaluation. These results provide practical insights for financial institutions seeking to minimize default risk while also advancing the integration of artificial intelligence into credit-risk management frameworks.

Contribution/Originality: This study contributes to the literature by systematically benchmarking logistic regression, random forest, and XGBoost for SME credit risk prediction using real-world data. It incorporates underexplored variables such as tax declarations, invoicing, and structural indicators. The findings provide new insights into balancing accuracy, interpretability, and robustness in financial risk assessment.

1. INTRODUCTION

Over the past few years, the rapid growth of digital technologies and the increasing convergence of technology and finance have led to a major upheaval in the financial sector. This change has resulted in a rise in credit demand, particularly through consumer financing channels, and has undervalued the importance of effective models for assessing credit risk. The risk of credit remains a significant challenge for lending institutions, especially commercial banks. This is because it indicates how likely it is that borrowers will be unable to meet their repayment obligations. Traditionally, rating systems and qualitative evaluations are used to assess solvency. Although these methods have long been employed to evaluate credit risk, they are often limited by their reliance on static historical data and may not account for changes in the borrower's behavior and broader economic conditions. In an environment characterized by financial uncertainty and shifting market conditions, the ability to accurately assess risk is more critical than ever. Strengthening underwriting practices and refining loan approval processes are essential not only to minimize exposure to default but also to ensure the long-term health of credit portfolios.

In this complex environment, SMEs occupy a special position in the real economy, but they continue to face difficulties in accessing finance. This is partly due to incomplete financial statements, limited collateral, and weak communication with credit institutions. When lenders assess credit risk, they often confront increased uncertainty because the information imbalance makes it harder to obtain a clear picture of the borrower's financial health. Although traditional analysis tools are still used, they have limitations: they rely on linear statistical logic that does not reflect the diversity of SMEs' business models or the rapidly changing nature of their environment. In this context, methods based on machine learning are entirely valid. By utilizing a broader range of data, they help to better understand economic dynamics and improve predictions of borrower behavior.

This study examines how different machine learning models classify firms according to their creditworthiness. The aim is to distinguish between companies that are likely to meet their financial commitments and those with a higher probability of default by analyzing a wide range of business variables. Choosing an appropriate model for credit risk assessment is essential, since incorrect classification can expose lenders to substantial losses. A direct comparison is therefore required to determine which method is most effective in predicting loan defaults among SMEs. The analysis focuses on three established models: logistic regression, random forests, and XGBoost. Their performance is measured with accuracy, precision, recall, F1 score, and AUC. The study seeks to improve the reliability of credit evaluations and to provide financial institutions with tools that can support lending decisions in practice.

At a broader level, the findings are expected to strengthen credit assessment methods and provide banks with approaches that encourage more consistent decision-making. As Inekwe (2016) points out, financial distress among SMEs has wide-reaching implications, not only for creditors but also for employee welfare and entrepreneurial development. This underscores the broader economic importance of improving credit risk assessment tools for small businesses.

Much of the existing research has concentrated on large firms and banking portfolios, often relying on traditional balance-sheet ratios (Addo, Guegan, & Hassani, 2018; Lessmann, Baesens, Seow, & Thomas, 2015). In contrast, this work directs attention to small and medium-sized enterprises (SMEs). Beyond standard financial ratios, the analysis also draws on operational and fiscal variables such as invoicing patterns, reported tax levels, and sales concentration elements that are seldom part of traditional scoring models. This makes the study one of the first to compare logistic regression, random forests, and XGBoost on SME-specific data. By doing so, it contributes to the literature on credit risk modeling and illustrates how these methods can be applied in settings where financial information is incomplete or fragmented.

2. LITERATURE REVIEW

In today's financial world, credit risk assessment is no longer just a routine task; it is a strategic priority. The digital transformation of banking and the growing complexity of borrower profiles have made this process even more challenging. It is very difficult to determine if individuals and businesses are financially stable when the available data is often broken, inconsistent, or changes over time. Although logistic regression and other traditional tools have long been used as benchmarks, particularly in institutional settings, they are not very effective when dealing with nonlinear patterns and anomalous behaviors that are frequently observed in small and medium-sized businesses (SMEs). These limitations have led to an increased adoption of AI, especially machine learning, in scientific research. By integrating various data sources, these techniques enable the modeling of complex relationships among variables and enhance prediction accuracy.

Lessmann et al. (2015) study, which contrasts 41 classification techniques used for creditworthiness assessment, is a positive example in this regard. They show that ensemble methods, such as Random Forest, perform better at prediction than traditional statistical models. Several studies have demonstrated the superiority of ensemble methods, particularly when dealing with noisy or semi-structured data.

Zhao et al. (2015) studied the use of neural networks for credit risk. They used a multilayer perceptron with only one hidden layer. A basic model achieved 87% accuracy, demonstrating that even simple designs can sometimes outperform traditional statistical methods.

In the context of alternative lending, Byanjankar, Heikkilä, and Mezei (2015) applied a neural network model to predict credit risk in peer-to-peer lending markets, demonstrating that machine learning techniques can outperform traditional methods even when borrower profiles are limited or non-standard an insight that supports their applicability in SME credit evaluation.

Addo et al. (2018) conducted research on how traditional models can be combined with deep learning. Their study focused on predicting business failure. In this context, they found that decision trees, especially when enhanced with modern techniques, still offer strong predictive power. These models are easier to understand than deep networks, which makes them especially useful for analysts in real-world situations. Often, traditional credit scoring systems consider small and medium-sized businesses (SMEs) to be riskier and less appealing.

Xia, Xu, Wei, Wei, and Tang (2023) studied credit risk among small and medium-sized businesses (SMEs) in the context of supply chains. They tested various machine learning models, such as decision trees, logistic regression, support vector machines, and random forests. According to their study, the random forest model outperformed other models, which is interesting. It was more accurate and had fewer misclassifications. This indicates that the random forest is a highly effective method for assessing credit risk, particularly for small and medium-sized businesses.

Kou et al. (2021) proposed a new approach for predicting the bankruptcy of small and medium-sized businesses by utilizing transaction data instead of traditional financial records. Their research indicates that the most effective ensemble models incorporate a variety of methods, such as logistic regression, Random Forest, and XGBoost, combined through a two-step variable integration process. Additionally, their findings highlight that behavioral indicators, including cash flows and business relationships, can serve as highly valuable predictors in bankruptcy prediction models.

Medianovskyi, Malakauskas, Lakstutiene, and Yahia (2022) explored the use of interpretable machine learning models for predicting financial distress among SMEs, combining predictive accuracy with transparency through SHAP values, an approach that aligns with the growing demand for explainable AI in credit-risk management.

Classical studies, such as the one by Svabova, Michalkova, Durica, and Nica (2020), used Amadeus data from 2016 to 2018 and well-known methods like logistic regression and discriminant analysis on Slovakian PMEs. Their model was more than 90% accurate, which demonstrates that even simple statistical methods can be effective when the data is well-structured.

Wang and Guedes (2025) examined failure prediction for both SMEs and new ventures in Portugal using a near-census sample of 229,855 SMEs and 101,645 new ventures over 2010-2018. Their results reveal that while age and size consistently discriminate failure risk across both groups, the specific financial predictors differ markedly between SMEs and new ventures, and the SME model achieves higher classification accuracy. This finding underlines the necessity of treating distinct firm segments separately when developing credit-risk or failure-prediction tools.

Dastile, Celik, and Potsane (2020) conducted a systematic literature review of 74 primary studies in credit scoring and report that ensemble classifiers outperform single classifiers, but note that interpretability and class imbalance remain major challenges in machine learning-based credit scoring.

Shi, Tse, Luo, D'Addona, and Pau (2022) conducted a comprehensive review of machine learning-driven credit risk modeling, identifying key algorithmic trends, performance challenges, and the increasing importance of explainability and real-time data integration.

Altman and Sabato (2007) developed one of the first SME-specific credit scoring models using U.S. data, demonstrating that traditional models designed for large firms are less effective for SMEs. Their logistic regression approach identified firm size, liquidity, and leverage as key predictors of default, emphasizing the importance of segment-specific modeling frameworks.

In a study on Malaysian SMEs, Abdullah, Ahmad, Zainudin, and Rus (2019) applied logistic regression to predict financial distress using financial ratios and found that while some predictors were significant, the model's performance was constrained by the limited depth of available accounting data. Their findings highlight the need to explore alternative data sources and modeling techniques to improve predictive reliability, especially within SME contexts.

Nguyen, Nguyen, Le, and Nguyen (2019) analyzed the financial aspects of publicly traded real estate companies in Vietnam and showed how the ROA, ROE, and asset turnover rate can be used to determine a company's likelihood of bankruptcy.

Moscatelli, Parlapiano, Narizzano, and Viggiano (2020) evaluated a broad set of machine learning algorithms on banking data for corporate default forecasting, demonstrating that ensemble and kernel-based methods significantly outperform traditional approaches in terms of predictive accuracy and robustness findings that are also transferable to SME risk modeling.

Other researchers have also used data from Moody's database, which contains information on both the economy as a whole and accounting data. Provenzano et al. (2020). They demonstrated through out-of-sample validation that machine learning models are effective in predicting corporate failure.

Modina, Pietrovito, Gallucci, and Formisano (2023) analyzed a large sample of Italian SMEs and found that incorporating bank–firm relationship indicators, such as credit line usage and loan overruns, significantly improves default prediction compared to models based solely on accounting data. Their findings emphasize the value of combining transactional and behavioral variables an approach also reflected in our use of tax, invoicing, and structural indicators.

Gupta, Wilson, Gregoriou, and Healy (2014) found that the internationalisation of SMEs significantly influences the performance of credit risk models, suggesting that firm-level structural features can impact predictive outcomes. This supports the integration of variables such as market concentration and customer dependency in credit scoring frameworks, as implemented in the current study.

Mhlanga (2021) examined the role of machine learning and artificial intelligence in improving credit risk assessment in emerging economies, emphasizing how alternative data and AI can enhance financial inclusion for SMEs. This supports the case for expanding beyond traditional financial variables to evaluate underbanked firms.

El Qadi, Trocan, Diaz-Rodriguez, and Frossard (2023) compared an AI model with Tinubu's rating system and reported that the AI approach was more effective at assessing long-term risk, though it produced more cautious evaluations. This contrast highlights the different ways algorithmic and human assessments approach credit risk.

Finally, Huang et al. (2023) introduced non-financial variables, often overlooked in credit scoring, to predict default among Chinese SMEs. Their research combines new models, such as natural language processing algorithms, and concludes that ensemble models, especially XGBoost, are sufficiently robust to handle large, diverse data sets.

Hossain et al. (2025) examined several common algorithms, including linear regression, neural networks, random forests, and gradient boosting models such as XGBoost and LightGBM, to determine which were most effective at predicting credit risk in banks. Their study indicates that XGBoost achieves the highest accuracy at 88.7%, with LightGBM performing very closely. This demonstrates that ensemble methods are highly effective when applied to structured banking data. Li, Zhang, and Wang (2021) introduced a modified gradient boosting approach (G-XGBoost) tailored for small sample credit datasets and showed that it can outperform standard XGBoost in terms of recall and robustness, further supporting the suitability of ensemble methods even with limited data.

Berrada, Barramou, and Alami (2024) proposed a machine learning-based approach for corporate loan default prediction, demonstrating that ensemble techniques such as XGBoost and random forests can significantly outperform traditional models in predictive accuracy. Their findings further validate the applicability of these models in structured lending contexts, including SMEs.

We do use this method of comparison in our work, but there are some significant differences. It is clear that it is intended for small and medium-sized enterprises (SMEs), whose needs and financial circumstances differ greatly from those of individuals who take out bank loans. By contrast, our dataset includes real-world tax, business, and behavioral variables, while Hossain et al. (2025) only use standard accounting data. Finally, we examine the model's robustness (via the rates of false positives and false negatives) and its interpretability, which are both critical for making scoring easier to use in banks and other financial organizations. Table 1 presents a summary of the main empirical studies that have compared traditional and machine learning methods for credit-risk assessment. It highlights the diversity of algorithms applied (logistic regression, random forest, neural networks, XGBoost) and their relative performance across various contexts and datasets. Ciampi, Giannozzi, Marzi, and Altman (2021) conducted a systematic review of over 100 studies on SME default prediction and emphasized the need to incorporate alternative data sources and machine learning techniques to improve model performance. Their findings support a shift toward AI-based methods that are better suited to capturing the complexity of SME risk, particularly in the wake of financial shocks like the COVID-19 crisis. Similarly, Cheraghali and Molnár (2024) conducted a comprehensive methodology-focused review of 145 studies on SME default prediction, highlighting major gaps in variable selection, validation practices, and model benchmarking. Their findings underscore the importance of using robust techniques such as cross-validation, feature selection, and comparative model testing principles that are reflected in the current study's approach to evaluating machine learning models for SME credit risk.

Table 1. Summary of related work in SME credit-risk prediction.

Article	Year	Journal	Main conclusion
Lessmann et al. (2015)	2015	European Journal of Operational Research	Ensemble methods, especially Random Forest, outperform single traditional models for credit scoring.
Zhao et al. (2015)	2015	Expert systems with applications	A simple MLP achieves high accuracy in credit scoring, demonstrating that neural networks are a viable option.
Addo et al. (2018)	2018	Risks	Tree-based models remain strong and explainable; deep learning adds value in some contexts.
Svabova et al. (2020)	2020	Sustainability	With well-structured data, simple statistical models (LR/LDA) can exceed 90% accuracy.
Kou et al. (2021)	2021	Decision support systems	Using behavioral data with two-stage feature selection improves SME bankruptcy prediction; ensembles perform best.
Nguyen et al. (2019)	2019	Investment management and financial innovations	ROA, ROE, and asset turnover significantly explain bankruptcy risk in listed real estate firms.
Provenzano et al. (2020)	2020	arXiv	Out-of-sample validation confirms that machine learning-based strategies are effective for predicting failures.
El Qadi et al. (2023)	2022	Signal image and video processing	Explainable AI (e.g., SHAP) aligns model outputs with expert knowledge; AI scoring is accurate yet conservative.
Xia et al. (2023)	2023	Sustainability	For supply chain SMEs, Random Forest provides the best performance among classifiers.
Huang et al. (2025)	2025	Applied sciences	Non-financial signals combined with advanced models improve SME default prediction.
Huang et al. (2024)	2024	Frontiers in artificial intelligence and applications	SHAP with XGBoost highlights sales volume, valid invoices, and declared tax as key predictors of SME credit risk.
Hossain et al. (2025)	2025	American journal of engineering and technology	Gradient boosting models (XGBoost) provide the best accuracy (88.7%) for banking credit-risk prediction.

Alongside the main contributions already discussed, several other studies have recently addressed similar issues in credit risk assessment, particularly for small and medium-sized enterprises. For instance, Khandani, Kim, and Lo

(2010) showed that including behavioral transaction-level data can significantly improve the accuracy of default predictions, especially when combined with machine learning techniques like random forests. Baesens, Setiono, Mues, and Vanthienen (2003) compared traditional scoring models with data mining methods and found that advanced models generally offer more accurate results. Probst and Boulesteix (2018) emphasized the importance of tuning hyperparameters, especially in ensemble models, to improve model performance in financial applications. Interpretability is also gaining attention. Ribeiro, Singh, and Guestrin (2016) introduced the LIME framework, which helps explain complex models, while Doshi-Velez and Kim (2017) argued that interpretability is essential when applying AI in sensitive areas like credit scoring. Chen, Liang, and Wang (2022) also stressed the need for fairness and transparency in AI governance in finance. Other researchers have focused on the role of data structure and behavioral patterns.

Liu, Zhang, and Fan (2022) combined graph neural networks with XGBoost to capture relationships between firms, thereby improving SME default prediction. As Du Jardin (2009) emphasizes, the predictive power of credit-risk models heavily depends on the careful selection of relevant variables. This study follows a similar logic by choosing explanatory features that are both theoretically grounded and empirically validated, including tax ratios, billing behavior, and structural concentration metrics.

3. MATERIALS AND METHODS

3.1. Data Collection

The database used in this work was taken from a public GitHub repository that was made available as part of a project about credit decisions for small and medium-sized businesses. The main advantage of this source is that it is easy to access and has a clear structure, which makes it a good basis for testing supervised classification methods. There are several quantitative variables derived from the business activities of companies, such as sales, amounts billed, and average purchase size. Most of these data are converted to a logarithmic scale to reduce the effect of extreme values. Aside from these "business" characteristics, the database also includes structural indicators, such as the Gini coefficient or the HHI index, which provide insights into the concentration of the economy within the client portfolio.

For contract defaults, the dependent variable is set to 1, and for non-defaults, it is set to 0. The goal is to determine the probability that a company will terminate its contract early by using the data that is currently available. We are developing a scoring system to assess the risk associated with each observation.

3.2. Justification for the Selection of Explanatory Variables

Choosing the right explanatory variables is a crucial step in building a strong predictive model. The goal is to keep the most relevant indicators that can explain the phenomenon being studied, which in this case is the risk of contract default in small and medium-sized businesses. Adding unnecessary or redundant variables might make the model heavier and less effective by adding noise or encouraging overfitting. The variables were chosen based on their fit with economic theory and how well they have been tested in scientific literature.

3.2.1. Variables Related to Commercial Activity

From a financial perspective, commercial activity indicators such as sales volume, number of invoices issued, and average transaction amount are traditionally considered direct indicators of a company's operational strength. A steady and regular activity tends to show that cash flows are stable and that short-term commitments can be met more easily. On the other hand, sudden changes or low volume can indicate liquidity problems or that the company is too dependent on certain clientele.

Huang et al. (2024) conducted a study on a sample of 425 Chinese SMEs that confirms this: using SHAP values on an XGBoost model, they demonstrate that variables such as the total amount of sales, the number of valid invoices, and the declared tax amounts are among the ten most important factors in the model. These results support the idea that billing data is a strong indicator of potential failures, especially in a small business setting where financial statements are sometimes incomplete or difficult to interpret.

3.2.2. Tax Variables

In the real world of small and medium-sized businesses, taxes are much more important than what accounting aggregates might suggest. When a business has to pay many taxes, even if it is doing well, it might quickly lose its ability to keep enough cash on hand. The average tax rate on bills and the tax-to-price ratio are two ways to measure this pressure practically. These measures indicate the portion of a business's income that the government collects before the business can cover its fixed costs or generate a profit.

This aspect is not only intuitive but also supported by research, such as Delis, Galariotis, Iosifidi, and Ongena (2025). Their careful study clearly shows that businesses subject to higher levels of taxation will experience a decrease in net cash flow, an increase in debt costs, and a higher default risk, especially among small businesses that are more vulnerable to cash flow problems. The clear link between tax pressure and financial fragility justifies including both factors in a credit risk model.

3.2.3. Structural Variables

When assessing the credit risk of small and medium-sized enterprises (SMEs), it is essential to consider factors beyond traditional financial indicators, including structural elements related to their economic environment. The Gini coefficient and the Herfindahl-Hirschman Index (HHI) are two objective measures used to evaluate the dependence of a business on a limited number of partners. These indicators provide insights into the distribution or concentration of income and help identify situations where businesses may be vulnerable. For example, a high dependence on a small number of customers or suppliers can increase susceptibility to issues such as contract breaches, partner bankruptcies, or industry shocks. Such dependence can lead to less predictable cash flows and reduced flexibility, particularly in unstable economic conditions. Incorporating these structural variables into a risk scoring model allows for a more comprehensive assessment by considering the company's resilience, not solely its performance metrics.

3.2.4. Behavioral Indicator

The idea behind including the ratio of canceled amounts on invoices is that companies' transactional behaviors can reveal structural weaknesses that are difficult to see in other ways. A high number of cancellations can indicate that a business is unstable, such as when customer relationships are strained, mistakes frequently occur in internal processes, or when contracts are often broken.

In the context of small and medium-sized businesses, these signals are obvious. These systems, which are often less formal, show their conflicts in operational details before the financial indicators start to decline. These variables make all of their sense in this kind of dynamic; they provide access to a level of granularity that is frequently lacking in traditional financial statements.

3.3. Data Preparation

Even though the database was well-organized, it needed to be cleaned up before the results could be trusted. A first examination found several unusual or inconsistent values, which were manually fixed to maintain the sample's quality.

To lessen the effects of statistical dispersion, several variables that measure economic activity (sales, invoices, average amounts) were changed to a logarithmic scale. This change is meant to reduce the impact of extreme values and make the distributions more uniform across companies of different sizes, which is a key requirement in a comparison analysis. This choice also aligns with a more nuanced economic interpretation, as it reduces the scale of the differences.

There weren't many qualitative factors, but they were encoded using a basic binary method to avoid any implicit bias. As for the target variable, which was coded as either "default" or "not default," it had a slightly uneven distribution. This imbalance stayed within limits that do not call for a resampling-type adjustment.

The final sample was split randomly, with 80 for the learning phase and 20 for the validation phase. This distribution allows us to test how well the models perform on data they have not seen before, while also ensuring that the procedure can be replicated. Table 2 presents the description of all variables used in the SME credit-risk dataset, including commercial, tax, structural, and behavioral indicators. It also specifies their data types and example values to clarify their operational meanings within the model.

Table 2. Description of variables used in the SME credit-risk dataset.

Variable	Description	Type	Example value
Company_Code	Unique identifier for each company	Categorical	E105
Sales_Volume_Log	Total sales volume (Logarithmic scale)	Numeric	14.91
Invoiced_Quantity_Log	Total quantity invoiced (Logarithmic scale))	Numeric	5.83
Avg_Sales_Amount_Log	Average amount per invoiced sale (Logarithmic scale)	Numeric	9.07
Purchase_Size_Log	Average purchase size (Logarithmic scale)	Numeric	5.76
Tax_Price_Ratio	Ratio between tax and price	Numeric	0.1379
Avg_Tax_Rate	Average tax rate on invoices	Numeric	0.1599
Downstream_Customers	Total number of downstream customers	Numeric	8
Negative_Invoice_Ratio	Ratio of invoices with a negative amount during the period	Numeric	0.0047
Period_Customers	Number of downstream customers during a given period	Numeric	8
Gini_Coefficient	Gini coefficient measures sales concentration	Numeric	0.9913
Seller_HHI_Index	Herfindahl–Hirschman Index (HHI) for seller market concentration	Numeric	0.0006
Contract_Default	Indicates whether a contract default occurred (1 = default, 0 = no).	Binary (Yes/No)	1

3.4. Model Development

For the modeling phase, we chose to test three different algorithms. The objective was to evaluate older and newer methods to see which one works best for credit scoring. The models selected are logistic regression, random forest, and XGBoost. We started with the logistic regression model due to its simplicity and the interpretability of its outputs. Although it is a linear model, it remains quite useful for credit rating work, especially because it allows you to connect each variable directly to the probability of a default. This link between the explanatory variables and the likelihood of an event (such as breaking a contract) makes the decision-making process easier to understand.

Actually, Huang et al. (2024), point out how useful this approach is because it can provide a clear probabilistic estimate of the predictions.

Similarly, Hossain et al. (2025) note that the clarity of logistic regression makes it a useful benchmark for evaluating more complex methods. Within our set of models, Random Forest emerged as a strong alternative because it combines robustness with flexibility. The idea is to make many separate decision trees, each trained on a different subsample of the data. Next, we combine the predictions from these trees, typically through a majority vote.

This makes the model more useful, even when the data is noisy or the relationships between variables are difficult to understand. Hossain et al. (2025) note that this approach reduces the risk of overfitting while maintaining a balance between accuracy and stability. In credit scoring, XGBoost is now widely used, as it increases predictive performance by building decision trees sequentially, each one correcting the errors of the previous stage.

Chen and Guestrin (2016) discuss this method, which significantly enhances model accuracy, especially when relationships between variables are complex or non-linear. We selected XGBoost for our study because it is fast and capable of handling heterogeneous data, which are essential qualities for multivariate analysis. XGBoost differs from other models due to its built-in regularization, which helps reduce overfitting and improves the model's generalization ability. Numerous recent studies have supported this method.

Rao, Liu, and Goh (2023) discussed the efficacity of the XGBoost model in assessing auto credit risk, while Liu et al. (2022) analyzed its performance. These results show that this method is useful for looking into problems with scoring reliability, especially for small and medium-sized entreprises.

To compare how well the models performed, we split the data into two groups. We used 80% of the data to train the models and the remaining 20% to test the predictions on new data. We applied a search grid and cross-validation to improve the performance of the algorithms most affected by internal parameters. This rigorous approach enabled us to make fair comparisons between the different models, taking into account their ability to adapt to the specific credit risk of SMEs.

As well as evaluating the overall performance of each algorithm, we extracted information at the level of individual variables to make the models interpretable. For logistic regression, this involved reporting the estimated coefficients (β) and transforming them into odds ratios (OR). These values quantify the marginal effect of each explanatory variable on the probability of contract default. For the Random Forest and XGBoost models, we computed feature importance scores, which summarize the relative contribution of each predictor to the model's classification accuracy. Table 3 presents the coefficients and odds ratios associated with the logistic regression model. Table 4 reports the feature importance values for the XGBoost model, while Table 5 shows the feature importance values for the Random Forest model. Figures 1 and 2 provide a graphical representation of these importance values for easier interpretation.

Table 3. Logistic regression coefficients and odds ratios.

Variable	Coefficient (β)	Odds Ratio (OR)
Average tax rate on invoices	0.389	1.475
Tax-to-price ratio	0.307	1.359
Ratio of canceled invoice amounts	0.241	1.273
Gini coefficient	0.131	1.140
Number of invoices (Log)	0.052	1.053
Sales volume (Log)	0.040	1.040
Number of downstream customers	-0.0004	0.9996
Average invoiced sales amount (Log)	-0.012	0.988
Seller HHI index	-0.129	0.879

Table 4. Feature importance for the XGBoost model.

Variable	Importance
Sales volume (Logarithmic)	0.316
Seller HHI index	0.230
Tax-to-price ratio	0.123
Ratio of canceled invoice amounts	0.103
Number of invoices (Logarithmic)	0.097
Average invoiced sales amount (Logarithmic)	0.082
Number of downstream customers	0.049
Average tax rate on invoices	0.000
Gini coefficient	0.000

Table 5. Feature importance for the random forest model.

Variable	Importance
Sales volume (Logarithmic)	0.229
Seller HHI index	0.182
Average invoiced sales amount (Logarithmic)	0.109
Number of invoices (Logarithmic)	0.099
Gini coefficient	0.084
Average tax rate on invoices	0.083
Tax-to-price ratio	0.079
Number of downstream customers	0.068
Ratio of canceled invoice amounts	0.068

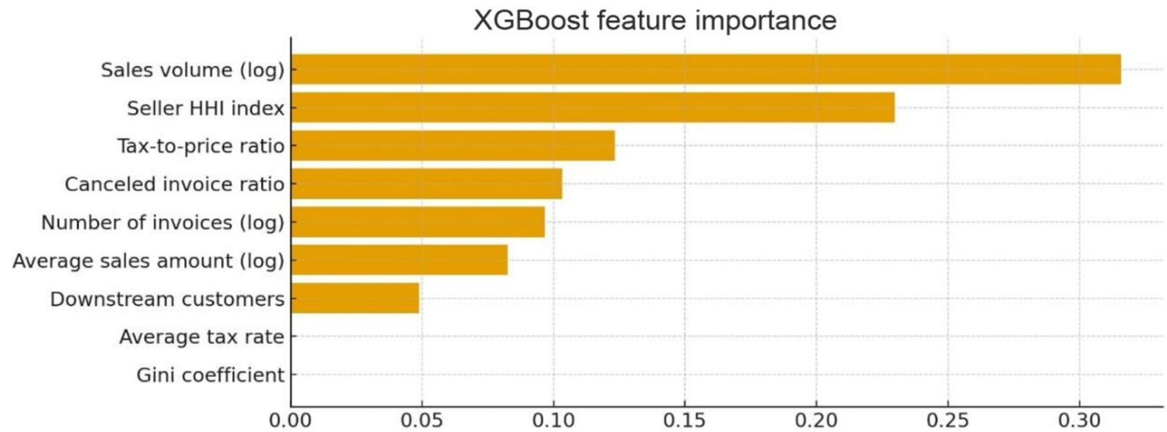

Figure 1. Feature importance values for the XGBoost model.

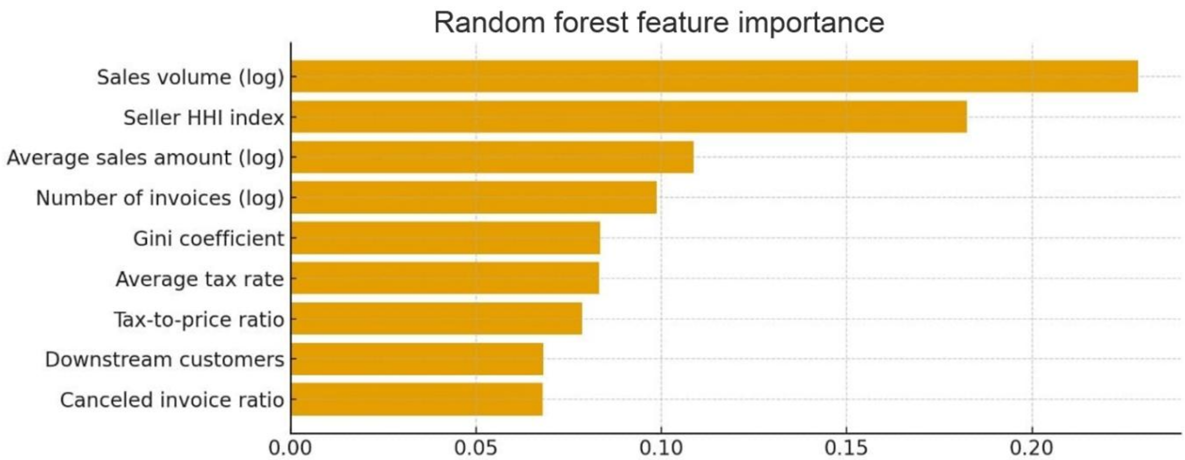

Figure 2. Feature importance values for the random forest model.

3.5. Model Evaluation

After training the models, their performance was evaluated using the test set and several indicators to obtain a comprehensive and balanced view of the results. The goal was not only to calculate the overall percentage of accurate predictions but also to determine the extent to which each model was able to identify contract default cases, which represented the minority class and, therefore, the main target.

A preliminary calculation was conducted to determine accuracy, which is frequently employed as a starting point. However, when there is an imbalance between the classes, this metric can provide a partial view: a model can have a high overall accuracy while missing the default cases every time. We used additional metrics, including accuracy, recall, and the F1-score, which combines the first two and offers a better picture of how well the model performs on the positive class.

We built a confusion matrix for each method to make the analysis more precise. It helps identify what kinds of mistakes were made false positives and false negatives and determine if a model tends to favor the majority class over detecting breaks. Finally, the AUC (area under the ROC curve) was calculated as a general measure of discrimination to assess how well the model could distinguish between risky companies and others, regardless of the decision threshold used.

Finding the model that performs consistently and satisfies credit risk requirements is the aim of this comparison to determine which is best for making predictions. Real-world applications of this model include customer portfolio management and decision-making systems.

4. RESULTS

This section presents the results of the three models tested in this study: logistic regression, random forest, and XGBoost. They all used the same dataset, which was split in half for training and testing. We examined the performance using various metrics, including accuracy, recall, precision, F1 score, AUC, and the confusion matrix. These metrics enable us to evaluate both the overall ability of the models to correctly predict cases and their sensitivity to contract violation cases, which is the primary focus here. To ensure a fair comparison, all three models were evaluated under identical settings without extensive hyperparameter tuning. A comparison table summarizes the results, followed by a more detailed analysis that highlights the strengths and weaknesses of each approach in the context of credit scoring. Table 6 presents the comparative performance metrics for the three models—logistic regression, random forest, and XGBoost based on accuracy, precision, recall, F1-score, AUC, and confusion matrix results.

Table 6. Performance of models on the test dataset.

Model	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	AUC (%)	FP	FN	TP	TN
Logistic regression	64	71	64	67	74	3	6	14	2
Random forest	84	84	84	84	93	2	2	18	3
XGBoost	88	90	88	86	68	3	0	20	2

Figure 3 illustrates the overall performance of the three models across the main evaluation metrics, showing that XGBoost achieves the highest predictive accuracy, followed by Random Forest and Logistic Regression.

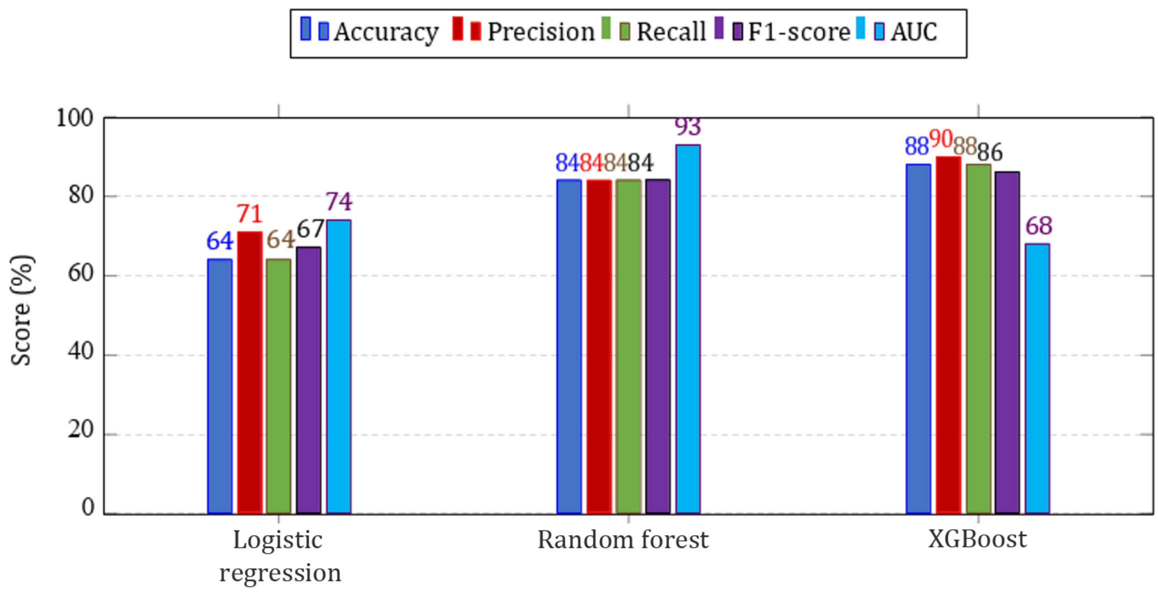

Figure 3. Model performance on the test dataset.

Beyond overall performance metrics, an examination of Tables 3–5 and Figures 1–2 highlights how the explanatory variables contribute to credit risk assessment. In the logistic regression, fiscal indicators such as the average tax rate on invoices and the tax-to-price ratio exert the strongest positive effects, confirming that higher taxation levels significantly increase the risk of default. Behavioral variables, particularly the ratio of cancelled invoice amounts, also play a decisive role by signaling transactional instability that often precedes financial distress. Conversely, the seller's HHI index displays a negative association, underlining the protective effect of diversification against overreliance on a limited customer base. The non-linear approaches (Random Forest and XGBoost) further validate these findings while providing complementary perspectives.

Both models rank sales volume and the number of invoices among the most influential explanatory variables, indicating that consistent commercial activity is closely tied to repayment capacity. Structural concentration (HHI) and irregular billing behavior (cancellations) also emerge prominently, reinforcing their importance across methodological settings. Taken together, these results suggest that fiscal burden, operational irregularities, and structural dependence constitute the core dimensions of SME credit risk, regardless of the algorithm employed.

5. DISCUSSION

Accuracy: The overall accuracy of the predictions allows us to assess how well each model distinguishes between companies that are failing and those that are not. XGBoost is the most accurate in our sample, with an accuracy rate of 88%. Random Forest follows with 84%, and logistic regression is significantly less accurate at 64%. This result indicates that XGBoost is generally the most suitable model for reliable classification.

Precision: The precision is the percentage of correctly identified risky businesses among all those predicted to be such. XGBoost has a high score of 90% here, which keeps the number of false positives low. Random Forest (84%) and logistic regression (71%) are next, but logistic regression makes more mistakes in its alerts.

Recall: The recall test checks your ability to correctly identify all real cases of defaults.

XGBoost performs again with 88%, identifying all failing companies without missing any (0 false negatives). Random Forest achieves 84%, whereas logistic regression is at 64%, indicating it is more likely to miss risky cases.

The F1-Score, which combines accuracy and recall, allows us to evaluate the balance between how well alerts are detected and their reliability. XGBoost still has an advantage with a score of 86%, followed by Random Forest (84%) and logistic regression (67%). These results demonstrate that XGBoost maintains a strong balance between sensitivity and accuracy.

AUC (Area Under Curve): The AUC demonstrates the ability of each model to differentiate between the two classes. Random Forest obtains the highest score (93%) in this instance, emphasizing its exceptional capacity to distinguish between risky and non-risky profiles. By contrast, XGBoost records a lower AUC of 68%, suggesting a less distinct decision boundary. Logistic regression, with an AUC of 74%, occupies an intermediate position between the two.

Confusion matrix analysis: Examining the confusion matrices helps to better understand the results. XGBoost stands out because it correctly identifies all the cases of default (no false negatives), with 20 failing companies correctly classified and only 3 false positives. This shows that it is very reliable for cases at risk. Random Forest has an interesting balance, with a more even distribution: 18 true positives, 3 true negatives, and only 4 errors in total (2 false positives and 2 false negatives). Logistic regression, on the other hand, has 6 type II errors (false negatives), which can be a problem when identifying defaults is a priority. The low number of true negatives (2) also indicates difficulty in identifying low-risk firms. Figure 4 illustrates the confusion matrix for the logistic regression model, showing the distribution of correctly and incorrectly classified cases, with several false negatives highlighting its lower sensitivity to defaults.

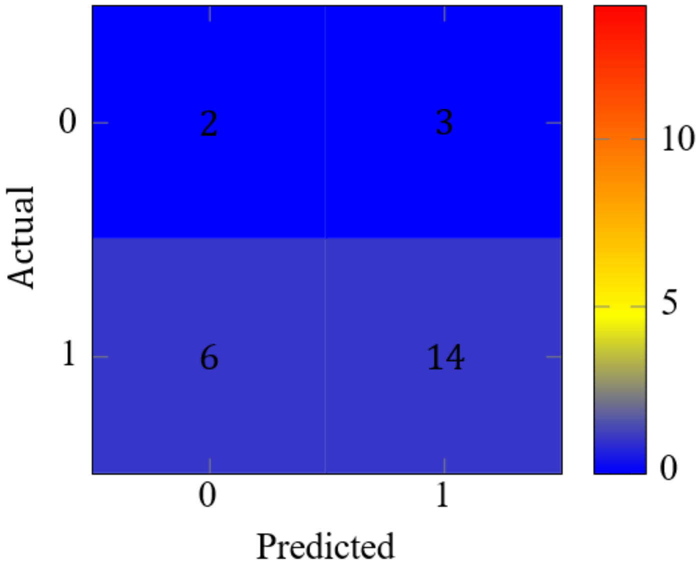

Figure 4. Confusion matrix for logistic regression.

Figure 5 illustrates the confusion matrix for the XGBoost model, indicating that it successfully identifies all default cases (zero false negatives), confirming its strong predictive power in detecting high-risk SMEs.

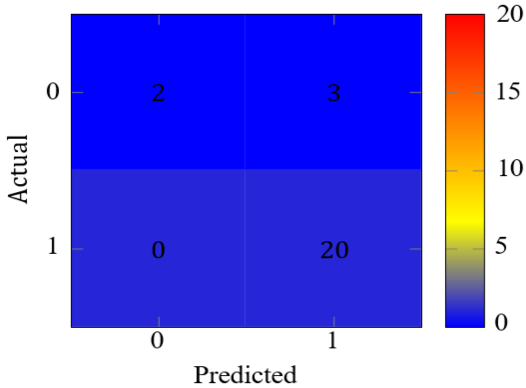

Figure 5. Confusion matrix for XGBoost.

Figure 6 illustrates the confusion matrix for the Random Forest model, demonstrating a balanced trade-off between false positives and false negatives, confirming its stability and discrimination capacity.

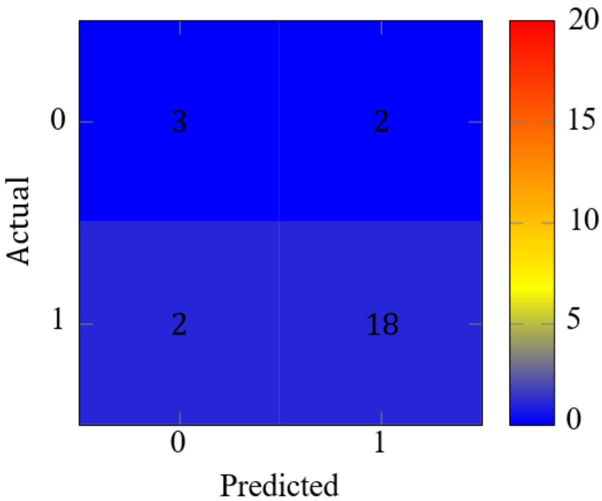

Figure 6. Confusion matrix for random forest.

5.1. Comparative Study

The results indicate significant differences among the three models used: logistic regression, random forest, and XGBoost. Each algorithm has its own profile, both statistically and in terms of potential application in decision-making.

The majority of raw metrics indicate that XGBoost is superior, with high recall and accuracy scores of 88% and 90%, respectively, and no failure cases overlooked. This demonstrates a good sense of direction.

On the other hand, the relatively low AUC value (68%) indicates that the decision boundary is less distinct. In other words, the model is effective at identifying risky cases but has difficulty distinguishing between risky and non-risky profiles.

The random forest, on the other hand, acts more consistently. Although it shows somewhat worse precision and recall performance, it has the best AUC (93%) and indicates a better division between risky and non-risky businesses. This model seems to be better at managing mistakes since it reduces both missed opportunities and false alarms, which can be helpful in a cautious but effective decision-making process.

Logistic regression is behind in comparison. It produces a greater number of errors when classifying failing businesses (6 false negatives) and continues to struggle to accurately distinguish between the two groups. It is still useful because it is simple and easy to read, but its limits become clear when the data has a more complicated structure.

In general, these results indicate that the choice of model is not solely based on its accuracy. It depends on the trade-off you choose between sensitivity and specificity. If you want to reduce costly mistakes in both directions (missing relevant instances and over-predicting), the random forest stands out as a very strong alternative for evaluating SMEs.

6. CONCLUSIONS

The aim of the study was to evaluate the efficacy of machine learning models in predicting the credit risk of small and medium-sized enterprises (SMEs). The adoption of data-driven predictive algorithms is not only a technical improvement but also a necessary response to the increasing demand for credit risk management that is more responsive, flexible, and reliable in financial systems where conventional scoring techniques are still dependent on rigid criteria, incomplete financial statements, or subjective assessments. This research is a valuable addition to the developing body of literature that investigates the intersection of credit risk assessment and artificial intelligence, with a particular focus on small and medium-sized enterprises (SMEs), which frequently encounter financing constraints despite their significance to economic development and employment.

Our empirical analysis illustrates that distinct performance patterns are produced by various models. XGBoost has been recognized as a highly effective instrument for default detection, demonstrating exceptional recall and the ability to identify every observed failure case. This ability to reduce Type II errors (missed defaults) renders it especially advantageous in credit risk management, where the consequences of undetected defaults can be catastrophic. In contrast, Random Forest demonstrated a robust ability to discriminate between solvent and defaulting firms, achieving a strong AUC score and offering an appealing compromise between specificity and sensitivity. Despite its relative simplicity, logistic regression continues to be pertinent due to its interpretability and transparency. In practical terms, this implies that classical statistical models continue to provide explanatory clarity that is frequently required in regulatory or managerial decision-making contexts, despite the fact that ensemble methods outperform traditional approaches in predictive accuracy.

These results are consistent with the findings of Hossain et al. (2025) in the banking sector, who also noted that XGBoost consistently outperformed competing models in terms of accuracy and recall. The external validity of our study is bolstered by the convergence of results across various domains, which implies that ensemble-based algorithms may serve as a widely applicable standard for financial risk modeling. However, our findings illustrate the importance of selecting a model that is consistent with the institutional objectives, in addition to its predictive capabilities. For instance, XGBoost is the optimal choice when the primary objective is to reduce costly errors of omission. Random Forest is a reliable compromise when interpretability and balance are prioritized. Logistic regression serves as a baseline benchmark when transparency is unavoidable.

This study makes a significant contribution to the literature by systematically benchmarking classical and modern machine learning methods for the prediction of SME risk and by incorporating nontraditional variables, such as tax declarations, invoicing, and structural indicators. From an academic perspective, this research provides empirical evidence concerning the trade-offs among model interpretability, robustness, and accuracy for financial institutions and policymakers. Such evidence is especially relevant for SME financing, where lending decisions are often based on incomplete information and considerable uncertainty. The findings highlight the importance of facilitating the integration of AI-based credit risk models into regulatory and institutional frameworks. When used alongside conventional scoring methods, AI-based models can enhance access to finance for small and medium-sized enterprises (SMEs) without undermining financial stability. For instance, XGBoost is effective in reducing missed defaults, while Random Forest strikes a balance between predictive accuracy and interpretability. These qualities make both models valuable for designing risk assessment tools that promote SME growth while safeguarding lenders. Integrating such methods into credit policies and guarantee frameworks can promote greater inclusiveness and strengthen financial stability. However, the results should be interpreted carefully. The study draws only on 124 Moroccan SMEs, and this narrow focus limits the extent to which the conclusions can be generalized beyond that context. Because the dataset is tied to a single country and period, the external validity is restricted. Future research should therefore test the models on larger and more diverse samples to determine whether the findings hold elsewhere.

While the results offer solid recommendations in this context, further research using larger and cross-country samples is necessary to strengthen the generalizability of the conclusions. The models do not yet incorporate temporal dynamics or macroeconomic disruptions, which are recognized as factors that can affect default risk. Secondly, although ensemble methods such as Random Forest and XGBoost achieve high levels of accuracy, their "black-box" nature can restrict their adoption in regulatory environments that necessitate transparent explanations of risk drivers. Lastly, logistic regression's predictive potential is reduced when dealing with complex, high-dimensional data, despite the fact that it provides interpretability.

Future research should address these limitations by extending the dataset across sectors and jurisdictions, incorporating longitudinal and macroeconomic variables, and utilizing sophisticated explainability frameworks like SHAP or LIME to improve transparency. It could also be beneficial to investigate cost-sensitive learning approaches, as these methods explicitly account for the asymmetric costs of misclassification in credit risk. Additionally, the disparity between predictive performance and interpretability could be bridged by experimenting with hybrid models that combine machine learning power with statistical clarity, such as Bayesian-additive trees or explainable boosting machines. The implementation of temporal validation schemes, such as sliding windows or rolling forecasts, would further facilitate a dynamic understanding of the evolution of risk over time.

In summary, the incorporation of artificial intelligence into the credit risk assessment of small and medium-sized enterprises (SMEs) achieves more than just an increase in predictive accuracy. It is indicative of a broader paradigm shift in financial decision-making, moving away from rigid, one-size-fits-all scoring systems toward more adaptive, transparent, and context-sensitive frameworks. This study makes both practical and theoretical contributions to the field of risk management by illustrating the strengths and trade-offs of multiple algorithms. Ultimately, the knowledge gained from this research may facilitate the development of financing systems that are not only more efficient but also more equitable, better suited to the needs of SMEs, and more capable of managing the uncertainties of contemporary financial markets.

Funding: This study received no specific financial support.
Institutional Review Board Statement: Not applicable.
Transparency: The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.
Data Availability Statement: The corresponding author can provide the supporting data of this study upon a reasonable request.
Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: Both authors contributed equally to the conception and design of the study. Both authors have read and agreed to the published version of the manuscript.

REFERENCES

Abdullah, N. A. H., Ahmad, A. H., Zainudin, N., & Rus, R. M. (2019). Predicting financially distressed small-and medium-sized enterprises in Malaysia. *Global Business Review, 20*(3), 627-639.

Addo, P. M., Guegan, D., & Hassani, B. (2018). Credit risk analysis using machine and deep learning models. *Risks, 6*(2), 38. https://doi.org/10.3390/risks6020038

Altman, E. I., & Sabato, G. (2007). Modelling credit risk for SMEs: Evidence from the US market. *Abacus, 43*(3), 332–357. https://doi.org/10.1111/j.1467-6281.2007.00234.x

Baesens, B., Setiono, R., Mues, C., & Vanthienen, J. (2003). Using neural network rule extraction and decision tables for credit-risk evaluation. *Management Science, 49*(3), 312-329.

Berrada, R. I., Barramou, F., & Alami, B. O. (2024). Towards a machine learning-based model for corporate loan default prediction. *International Journal of Advanced Computer Science and Applications, 15*(3), 565–573. https://doi.org/10.14569/IJACSA.2024.0150357

Byanjankar, A., Heikkilä, M., & Mezei, J. (2015). *Predicting credit risk in peer-to-peer lending: A neural network approach.* Paper presented at the 2015 IEEE Symposium Series on Computational Intelligence (SSCI), Cape Town, South Africa. IEEE

Chen, J., Liang, X., & Wang, Q. (2022). Artificial intelligence governance in finance: Transparency, fairness, and accountability. *AI & Society, 37*(1), 89–102.

Chen, T., & Guestrin, C. (2016). *Xgboost: A scalable tree boosting system.* Paper presented at the Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining.

Cheraghali, H., & Molnár, P. (2024). SME default prediction: A systematic methodology-focused review. *Journal of Small Business Management, 62*(6), 2847-2905. https://doi.org/10.1080/00472778.2023.2277426

Ciampi, F., Giannozzi, A., Marzi, G., & Altman, E. I. (2021). Rethinking SME default prediction: A systematic literature review and future perspectives. *Scientometrics, 126*(3), 2141-2188. https://doi.org/10.1007/s11192-020-03856-0

Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing, 91*, 106263. https://doi.org/10.1016/j.asoc.2020.106263

Delis, M. D., Galariotis, E., Iosifidi, M., & Ongena, S. (2025). Corporate taxes and entrepreneurs' income: A credit channel. *Journal of Corporate Finance, 93*, 102805. https://doi.org/10.1016/j.jcorpfin.2025.102805

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608.* https://doi.org/10.48550/arXiv.1702.08608

Du Jardin, P. (2009). Bankruptcy prediction models: How to choose the most relevant variables. *Bankers, Markets & Investors, 98*, 39–46.

El Qadi, A., Trocan, M., Diaz-Rodriguez, N., & Frossard, T. (2023). Feature contribution alignment with expert knowledge for artificial intelligence credit scoring. *Signal, Image and Video Processing, 17*(2), 427-434. https://doi.org/10.1007/s11760-022-02239-7

Gupta, J., Wilson, N., Gregoriou, A., & Healy, J. (2014). The effect of internationalisation on modelling credit risk for SMEs: Evidence from UK market. *Journal of International Financial Markets, Institutions and Money, 31*, 397-413. https://doi.org/10.1016/j.intfin.2014.05.001

Hossain, S., Sajal, A., Jamee, S. S., Tisha, S. A., Siddique, M. T., Obaid, M. O., . . . Haque, M. S. U. (2025). Comparative analysis of machine learning models for credit risk prediction in banking systems. *The American Journal of Engineering and Technology, 7*(04), 22-33. https://doi.org/10.37547/tajet/Volume07Issue04-04

Huang, B., Zhao, F., Tian, M., Zhang, D., Zhang, X., Wang, Z., . . . Chen, B. (2024). Explainability of machine learning in credit risk assessment of SMEs. *Artificial Intelligence and Human-Computer Interaction, 385*, 165-176.

Huang, H., Li, J., Zheng, C., Chen, S., Wang, X., & Chen, X. (2025). Advanced default risk prediction in small and medum-sized enterprises using large language models. *Applied Sciences, 15*(5), 2733. https://doi.org/10.3390/app15052733

Huang, S., Dong, L., Wang, W., Hao, Y., Singhal, S., Ma, S., . . . Patra, B. (2023). Language is not all you need: Aligning perception with language models. *Advances in Neural Information Processing Systems, 36*, 72096-72109.

Inekwe, J. N. (2016). Financial distress, employees' welfare and entrepreneurship among SMEs. *Social Indicators Research, 129*(3), 1135-1153. https://doi.org/10.1007/s11205-015-1164-6

Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance, 34*(11), 2767-2787. https://doi.org/10.1016/j.jbankfin.2010.06.001

Kou, G., Xu, Y., Peng, Y., Shen, F., Chen, Y., Chang, K., & Kou, S. (2021). Bankruptcy prediction for SMEs using transactional data and two-stage multiobjective feature selection. *Decision Support Systems, 140*, 113429. https://doi.org/10.1016/j.dss.2020.113429

Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research, 247*(1), 124-136. https://doi.org/10.1016/j.ejor.2015.05.030

Li, X., Zhang, Y., & Wang, Z. (2021). G-XGBoost: A modified gradient boosting approach for small sample credit datasets. *Journal of Financial Risk Management, 15*(3), 112-128.

Liu, J., Zhang, S., & Fan, H. (2022). A two-stage hybrid credit risk prediction model based on XGBoost and graph-based deep neural network. *Expert Systems with Applications, 195*, 116624. https://doi.org/10.1016/j.eswa.2022.116624

Medianovskyi, K., Malakauskas, A., Lakstutiene, A., & Yahia, S. B. (2022). *Interpretable machine learning for SME financial distress prediction. In Laouar, M.R., Balas, V.E., Lejdel, B., Eom, S., Boudia, M.A. (Eds).* Paper presented at the 12th International Conference on Information Systems and Advanced Technologies "ICISAT 2022". ICISAT 2022. Lecture Notes in Networks and Systems. Springer, Cham.

Mhlanga, D. (2021). Financial inclusion in emerging economies: The application of machine learning and artificial intelligence in credit risk assessment. *International Journal of Financial Studies, 9*(3), 39. https://doi.org/10.3390/ijfs9030039

Modina, M., Pietrovito, F., Gallucci, C., & Formisano, V. (2023). Predicting SMEs' default risk: Evidence from bank-firm relationship data. *The Quarterly Review of Economics and Finance, 89*, 254-268. https://doi.org/10.1016/j.qref.2023.04.008

Moscatelli, M., Parlapiano, F., Narizzano, S., & Viggiano, G. (2020). Corporate default forecasting with machine learning. *Expert Systems with Applications, 161*, 113567. https://doi.org/10.1016/j.eswa.2020.113567

Nguyen, V. C., Nguyen, T. N., Le, T. T. O., & Nguyen, T. T. (2019). Determining the impact of financial performance factors on bankruptcy risk: An empirical study of listed real estate companies in Vietnam. *Investment Management & Financial Innovations, 16*(3), 307-318. https://doi.org/10.21511/imfi.16(3).2019.27

Probst, P., & Boulesteix, A.-L. (2018). To tune or not to tune the number of trees in random forest. *Journal of Machine Learning Research, 18*(181), 1-18.

Provenzano, A. R., Trifiro, D., Datteo, A., Giada, L., Jean, N., Riciputi, A., . . . Nordio, C. (2020). Machine learning approach for credit scoring. *arXiv preprint arXiv:2008.01687*. https://doi.org/10.48550/arXiv.2008.01687

Rao, C., Liu, Y., & Goh, M. (2023). Credit risk assessment mechanism of personal auto loan based on PSO-XGBoost Model. *Complex & Intelligent Systems, 9*(2), 1391-1414. https://doi.org/10.1007/s40747-022-00854-y

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *" Why should i trust you?" Explaining the predictions of any classifier.* Paper presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

Shi, S., Tse, R., Luo, W., D'Addona, S., & Pau, G. (2022). Machine learning-driven credit risk: A systemic review. *Neural Computing and Applications*, *34*(17), 14327-14339. https://doi.org/10.1007/s00521-022-07472-2

Svabova, L., Michalkova, L., Durica, M., & Nica, E. (2020). Business failure prediction for Slovak small and medium-sized companies. *Sustainability*, *12*(11), 4572. https://doi.org/10.3390/su12114572

Wang, W., & Guedes, M. J. (2025). Firm failure prediction for small and medium-sized enterprises and new ventures. *Review of Managerial Science*, *19*(7), 1949-1982. https://doi.org/10.1007/s11846-024-00742-4

Xia, Y., Xu, T., Wei, M.-X., Wei, Z.-K., & Tang, L.-J. (2023). Predicting chain's manufacturing SME credit risk in supply chain finance based on machine learning methods. *Sustainability*, *15*(2), 1087. https://doi.org/10.3390/su15021087

Zhao, Z., Xu, S., Kang, B. H., Kabir, M. M. J., Liu, Y., & Wasinger, R. (2015). Investigation and improvement of multi-layer perceptron neural networks for credit scoring. *Expert Systems with Applications*, *42*(7), 3508-3516. https://doi.org/10.1016/j.eswa.2014.12.006