International Journal of Business and Economic Sciences Applied Research

IJBESAR
ijbesar.af.duth.gr

Credit Risk Modelling for Small and Medium-Sized Enterprises in Zimbabwe

Mbakisi Dube†[1], Zivai Gumbo[1], Saiding Munyala[1], Noble J. Malunguza[1]

[1] Department of Actuarial, Insurance and Risk Management Sciences, National University of Science and Technology, Zimbabwe

ARTICLE INFO

Article History

Received 12 December 2025
Accepted 29 March 2026

JEL Classifications
G21,B12,C38,G17, C53,D81

Keywords:
Credit Risk Modelling, Credit Risk Assessment, Machine Learning, Small and Medium Sized Enterprises (SMEs)

ABSTRACT

Purpose:
The main aim of the study was to determine a context-specific credit risk assessment framework that integrates both traditional financial metrics and alternative data sources to better evaluate the creditworthiness of Zimbabwean SMEs. It also identified key factors affecting credit risk for SMEs in Zimbabwe by incorporating both financial and non-financial data.

Design/methodology/approach:
We employed machine learning algorithms which were Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), and AdaBoost (AB). The data was obtained from the loan database of an SME banking division of a commercial bank of Zimbabwe consisting of 52,750 loan applicants over the 5-year period from 2018 to 2022. The features of the credit dataset included capital structure, financial history, profitability, liquidity, growth potential, industry characteristics, management quality, social media engagement, macroeconomic environment, customer concentration, credit history, firm age, collateral availability and whether the applicant defaulted on the loan or not. We performed data preprocessing and cleaning, feature development, hyper parameter selection and cross validation. A split ratio of 80% for the training set and 20% for the testing set was used, followed by an evaluation of the model based on the following performance metrics: classification accuracy, precision, recall (sensitivity), F1-score and the Receiver Operating Characteristic - Area Under the Curve (ROC-AUC).

Findings:
We find that traditional financial indicators such as profitability, liquidity and leverage, non-financial factors—including collateral availability, cash flow stability, management quality, and macroeconomic conditions—play a significant role in shaping credit risk profiles. Non-traditional data sources such as firm characteristics, supplier–buyer relationships, and social media activity can provide deeper insights into SMEs' operational performance and risk exposure. Incorporating these data sources alongside traditional financial information can significantly enhance the prediction of defaults.

Research limitations/implications:
This study was confined to one Zimbabwean bank, this represents a narrow focus since the Zimbabwean banking industry has 343 players as at 30 September 2025, (Reserve Bank of Zimbabwe (2025)). Also, since we base our research on Zimbabwe, it implies that the findings of this study cannot be generalised to all developing countries.

Originality/value:
This study contributes to the theory by providing an enhanced credit risk assessment framework that integrates traditional financial indicators and alternative data sources for Zimbabwean banks when determining the credit risk of SMEs. This will improve access to credit by SMEs in Zimbabwe and in jurisdictions with similar economic environments as those found in Zimbabwe. By providing reliable credit risk assessment methods it increases the financial inclusion of SMEs.

†Corresponding Author: Mbakisi Dube
e-mail: mbakisi.dube@nust.ac.zw

ORCID:0000-0002-4329-8543

DOI: 10.25103/ijbesar.182.05

1. Introduction

Small and Medium-sized Enterprises (SMEs) form the foundation of Zimbabwe's economy, performing an important task of stimulating economic growth, employment, and social development. According to the latest government statistics, there are 1,954,202 owners of Micro, Small and Medium Enterprises (MSMEs) in Zimbabwe, employing 7.05 million individuals, generating an estimated USD 14.2 billion in annual turnover (FinScope, 2021). These enterprises span a diverse range of sectors, including agriculture, manufacturing, trade, and services, and are crucial in fostering innovation, competition, and entrepreneurship.

The significance of the SME sector in Zimbabwe is well-documented in existing literature. Dlamini and Schutte (2020) emphasises that SMEs are the primary source of job creation, with the potential to alleviate poverty and income inequality, especially in rural and marginalised communities in Zimbabwe. Similarly, Makanyeza and Dzvuke (2015) explain the importance of SMEs to the promotion of inclusive and sustainable economic development, as they are more likely to employ local resources and engage with local supply chains. Furthermore, Njanike (2020) states that a thriving SME sector can contribute to the diversification of the Zimbabwean economy, reducing its reliance on a few dominant industries and making it more resilient to external shocks since it is known that 99% of business enterprises in developing countries are SMEs. Matsongoni and Mutambara (2021) state that the informal sector which is predominantly driven by SMEs is the largest absorber of labour in Zimbabwe, implying that it is contributing towards the livelihoods of many Zimbabweans,

SMEs represent a vital sector in Zimbabwe but nonetheless it faces several challenges that prevent it attaining its full potential. A critical impediment is the lack of access to credit and finance, this is generally considered as the most restrictive inhibition that limits growth and development of SMEs in Zimbabwe (Karedza et al., 2014). Inadequate financing restricts the ability of SMEs to invest in new technologies, expand their operations, and capitalise on emerging market opportunities.

1.1 Problem Statement

The limited access to credit and finance experienced by SMEs in Zimbabwe can be attributed to a range of interrelated factors. Firstly, SMEs often lack the collateral and credit history required by traditional financial institutions to secure loans (Manyanga et al., 2023). The high-risk perception associated with SMEs, coupled with information asymmetries, makes them less attractive to lenders compared to larger, established corporates (Dhlandhlara, 2019).

Furthermore, the volatile economic and political environment in Zimbabwe has exacerbated the challenges faced by SMEs in accessing credit. Periods of hyperinflation, currency instability and policy uncertainty have heightened the risk profile of SMEs leading financial institutions to adopt more conservative lending practices (Karedza et al., 2014). In most jurisdictions including Zimbabwe there is a lack of well-developed credit infrastructure such as credit bureaus and collateral registries further compounding the information asymmetries between SMEs and lenders (Dhlandhlara, 2019; Muriithi, 2017). This financing gap has constrained the ability of SMEs to invest in innovation, expand their operations and ultimately realise their full economic potential.

Consequently, the shortcomings of the current credit risk assessment approaches have contributed to the persistent financing gap faced by SMEs in Zimbabwe. Financial institutions constrained by the perceived high risk associated with the SME segment have often resorted to conservative lending practices further reducing access to credit for SMEs in Zimbabwe. This study presents a novel, context-specific credit risk assessment framework that integrates both traditional financial metrics and alternative data sources to better evaluate the creditworthiness of Zimbabwean SMEs. This is achieved by. The model is then validated by using real-world data from Zimbabwean SMEs.

1.2 Significance of the study

By integrating both traditional and alternative data sources, this study will expand the existing theoretical and empirical understanding of credit risk modelling approaches that can effectively capture the nuances of SME lending in emerging market settings. By addressing the critical gap in existing literature this study contributes to the advancement of both academic and practitioner-oriented understandings of credit risk assessment for SMEs. The finding we have generated will inform policy decisions, guide the development of targeted interventions and ultimately support the growth and financial inclusion of the SME sector.

In this study we provide a roadmap towards implementing a machine learning based approach to credit risk assessment that incorporates alternative data by Zimbabwean banks when lending to SMEs. We also provide a comparison of the machine learning algorithms that can be used in order to provide operational guidance to banks in emerging markets, with Zimbabwe acting as a case study. Furthermore, we give several practical recommendations to policymakers, banks and SMEs that if implemented will support the growth and financial inclusion of the SME sector. To the best of our knowledge such a study has never been undertaken, hence the novelty of the present study's findings.

The rest of the paper is structured in the following manner: Section 2 presents a review of literature. Section 3 contains the methodology. Section 4 contains results and analysis including the model performance metrics. Section 5 contains the conclusions and recommendations.

DOI: 10.25103/ijbesar.182.05

2. Review of Literature

2.1 Previous studies

Some studies have performed credit risk assessment modelling that include alternative data using machine learning. Recently, Jiang et al. (2026) utilise data about micro enterprises obtained from an internet bank to assess credit risk using alternative data. They categorise the alternative data as either historical credit data or behavioral data. They proceeded to employ a random forest model on the data categories and perform credit risk assessment. The behavioral data based models are shown to perform better than historical credit risk data based models.

In another recent study, Ahmad (2026) provides a wide ranging analysis of the uptake of credit risk assessment models that use alternative data. The aim of the study was to explore how these approaches have improved the accuracy of credit risk prediction and the broader inclusion of smaller players in the financial sector. They demonstrate that credit risk assessment via the use of alternative data should be complemented by changing its emphasis to transparency and socio-economic equity.

In their study Lee et al. (2026) demonstrate how retail transaction data can be used for credit risk assessment. The data is obtained from a Peruvian company. In their study they combine customer loyalty data with credit card repayment data and juxtapose it with each individual's financial history obtained from the Peruvian financial records. Using machine learning algorithms, they calculate credit scores for individuals with and without a credit history. Their study finds that using retail transaction data improves the credit rating of people who do not have credit histories from 16% to be approximately between 31% and 48%.

3. Methodology

3.1 Introduction

This section outlines the methodological approach used to develop and evaluate a machine learning model for credit risk prediction. An accurate and interpretable machine learning credit risk model was created that can be used by financial institutions to better assess the creditworthiness of loan applicants in Zimbabwe by SMEs.

3.1 Data collection

The data used in this study was obtained from the loan database of an SME banking division of a commercial bank of Zimbabwe. The dataset includes information on 52,750 loan applicants over the 5-year period from 2018 to 2022. The loans that were in default constituted 22% of the portfolio and 78% was fully repaid. The features of the credit dataset included capital structure, financial history, profitability, liquidity, growth potential, industry characteristics, management quality, social media engagement, macroeconomic environment, customer concentration, credit history, firm age, collateral availability and whether the applicant defaulted on the loan or not.

3.2 Data pre-processing

The first step in the data pre-processing phase involved addressing missing values within the dataset. A thorough examination revealed that several attributes had an insignificant number of missing observations, ranging from 0.5% to 2% of the total data. To handle these missing values, a combination of techniques was employed:

- For numeric variables the missing values were imputed using the mean of the available data for that feature.
- For categorical variables the missing values were imputed using the most frequent category (mode) for that feature.
- In cases where the missing data would have exceeded 25% for a particular variable, the feature was considered for removal from the analysis to avoid introducing excessive bias.

3.3 Feature development

After the initial data cleaning we conducted feature development to create additional variables that could potentially improve the predictive power of the credit risk model. This process involves the following steps:

- Derivation of Ratios: Several financial ratios were calculated from the available data, such as the current ratio, debt-to-equity ratio, and return on assets, to capture the financial health and performance of the SMEs.
- Temporal Feature Creation: Variables representing the duration of the SME's operation (firm age) and the timeliness of supplier payments (supplier payment delay) were engineered to incorporate the temporal aspects of the business.
- Categorical Variable Encoding: Categorical features, such as business sector and management experience, were encoded using one-hot encoding to enable their inclusion in the machine learning models.

DOI: 10.25103/ijbesar.182.05

3.4 Model development

3.4.1 Variable selection

The selection of variables for the machine learning-based credit risk assessment model for Zimbabwean SMEs was informed by a comprehensive review of the existing literature on SME financing and credit risk modeling. The key variables to be considered, based on the insights from the literature, are found in Table 3.1.

- Hyperparameter selection: Hyperparameters were selected using a mix of default settings and systematic tuning. For example, SVM hyperparameters (C, gamma, kernel) were optimised using GridSearchCV, while KNN used grid search to identify the optimal number of neighbors (e.g., k=9k=9k=9). In contrast, Logistic Regression and Random Forest were largely implemented with default parameters, with feature selection applied to Logistic Regression via RFECV.
- Cross-validation approach: Beyond the initial train–test split (70/30), cross-validation was applied in specific stages. RFECV used StratifiedKFold (cv=2) for feature selection, and GridSearchCV implicitly employed cross-validation during hyperparameter tuning (e.g., for SVM and KNN). However, there was no consistent, unified cross-validation strategy applied across all models.
- Justification of model choice: The selected models—Logistic Regression, SVM, KNN, and Random Forest—provide a diverse set of learning approaches: linear (LR), kernel-based (SVM), distance-based (KNN), and ensemble methods (RF). This allows for robust comparison and captures different data structures. Additionally, SMOTE was used to address class imbalance, which is critical in credit risk classification problems.

Table 3.1: Model variables

Financial variables	• Liquidity ratios (e.g., current ratio and quick ratio) • Profitability ratios (e.g., return on assets and net profit margin) • Leverage ratios (e.g., debt-to-equity and debt-to-assets): • Cash flow measures (e.g., cash flow from operations and cash flow coverage ratio)
Non-financial variables	• Firm characteristics (e.g., firm age, industry sector, legal structure and customer concentration) • Management quality (e.g., education, experience and decision-making practices) • Collateral and guarantees • Business environment
Alternative data variables	• Digital footprint (e.g., online sales, social media activity and mobile money transactions) • Supply chain relationships (e.g., supplier payment history and customer concentration) • Entrepreneurial characteristics (e.g., owner's credit history and educational background,

Table 3.2 provides a description of the characteristics of each variable used in our models.

Table 3.2: Variable characteristics and description

Variable	Category	Type	Calculation	Additional Notes
Loan Status (Default)	Target Variable	Categorical (Binary)	1 = Default, 0 = Non-default	Dependent variable
Liquidity Ratio (Current Ratio)	Financial	Continuous	Current Assets / Current Liabilities	Measures short-term solvency

DOI: 10.25103/ijbesar.182.05

Quick Ratio	Financial	Continuous	(Current Assets – Inventory) / Current Liabilities	More conservative liquidity measure
Return on Assets (ROA)	Financial	Continuous	Net Income / Total Assets	Profitability indicator
Net Profit Margin	Financial	Continuous	Net Profit / Revenue	Operational efficiency
Debt-to-Equity Ratio	Financial	Continuous	Total Debt / Shareholders' Equity	Capital structure risk
Debt-to-Assets Ratio	Financial	Continuous	Total Debt / Total Assets	Overall leverage
Cash Flow from Operations	Financial	Continuous	Net Cash from Operating Activities	Core cash generation
Cash Flow Coverage Ratio	Financial	Continuous	Operating Cash Flow / Total Debt	Ability to service debt
Revenue Growth Rate	Financial	Continuous	(Current Revenue – Previous Revenue) / Previous Revenue	Growth potential proxy
Firm Age	Non-Financial	Continuous (can be ordinal)	Current Year – Year Established	Proxy for stability
Industry Sector	Non-Financial	Categorical	Encoded (e.g., one-hot)	Captures industry risk differences
Customer Concentration	Non-Financial / Alternative	Continuous	% of revenue from top customers	High concentration = higher risk
Legal Structure	Non-Financial	Categorical	Encoded (e.g., LLC, Corporation)	Impacts liability and governance
Management Experience	Non-Financial	Ordinal	Years of experience or score index	Proxy for management quality
Management Education Level	Non-Financial	Ordinal	Encoded (e.g., High school=1, Degree=2, etc.)	Human capital indicator
Collateral Availability	Non-Financial	Categorical (Binary/Ordinal)	1 = Collateral available, 0 = none (or value-based ratio)	Reduces lender risk
Business Environment Index	Non-Financial	Continuous	Composite index (e.g., regulatory, market conditions)	Macro/business climate
Macroeconomic Indicator (e.g., GDP growth)	Non-Financial	Continuous	External data (e.g., % GDP growth, inflation rate)	Systemic risk factor
Credit History Score	Alternative / Financial	Continuous	Credit bureau score or repayment index	Key risk predictor

DOI: 10.25103/ijbesar.182.05

Previous Defaults	Alternative	Categorical (Binary/Count)	1 = prior default, 0 = none (or count)	Strong predictor of default
Social Media Engagement	Alternative	Continuous	Metrics (followers, engagement rate, activity index)	Proxy for business activity
Online Sales Volume	Alternative	Continuous	Total online transaction value	Digital footprint
Mobile Money Transactions	Alternative	Continuous	Transaction volume/frequency	Financial behavior proxy
Supplier Payment History	Alternative	Ordinal	Payment timeliness score (e.g., late/on-time)	Supply chain reliability
Customer Reviews / Ratings	Alternative	Continuous/Ordinal	Average rating score	Market perception

3.2 Model validation
The following validation techniques will be employed to ensure the reliability, robustness and generalisability of the proposed credit risk model.

3.2.1 Model training

- Data loading and pre-paration:The data was loaded into a Pandas data-frame, which is a popular data structure in Python for working with tabular data. The data-frame included the 13 features (Liquidity, Profitability, Leverage, Firm Age, Management Experience, Collateral Availability, Customer concentration, Supplier Payment Delay, Industry Index, Macroeconomic Index, Business Growth Potential Index, Social Media Engagement Index, and Credit History) and the target variable 'Defaulted' (0 for non-defaulted, 1 for defaulted).
- Train-test split:The pre-processed data is split into training and testing sets using the train_test_split function from scikit-learn. The training set is used to train the machine learning model, while the testing set is used to evaluate the model's performance on unseen data. A split ratio of 80% for the training set and 20% for the testing set was used based on the requirements of this research. Stratified splitting was used to ensure that the class distribution (defaulted vs. non-defaulted loans) is maintained in both the training and testing sets, which is particularly important when dealing with imbalanced datasets.

3.2.2 Model performance
The evaluation of the model is based on a set of performance metrics. We used the following performance metrics: classification accuracy, precision, recall (sensitivity), F1-score and the Receiver Operating Characteristic - Area Under the Curve (ROC-AUC).

4 Model results and analysis

4.1 Introduction
This section presents the results and analysis for the machine learning-based credit risk model for SMEs in Zimbabwe. The section is structured as follows: exploratory analysis, feature importance analysis and model interpretation and discussion.

4.2 Exploratory analysis
To gain a comprehensive understanding of the dataset and identify potential relationships between the variables, exploratory data analysis (EDA) was conducted. The insights gained from exploratory data analysis were used in carrying out feature selection. The study was able to identify the most relevant variables, understand the underlying data characteristics, and detect any potential issues or biases within the dataset.

4.2.1 Summary statistics
The summary statistics were calculated, including mean, median, standard deviation, and range, for each numerical variable to understand their distribution and characteristics. These results are presented in Table 4.1.

DOI: 10.25103/ijbesar.182.05

Table 4.1: Summary statistics

Variable	Mean	Median	Standard Dev	Min	Max
Current Ratio	1.78	1.65	0.82	0.32	5.21
Return on Assets	0.09	0.08	0.06	-0.15	0.33
Debt-to-Equity Ratio	1.42	1.21	0.95	0.08	6.74
Firm Characteristics	8.3	7	5.2	4	15
Management Experience (years)	11.4	10	6.8	2	35
Collateral Availability %	65.2	70	23.1	10	100
Average Monthly Online Sales	18,450.00	15,320.00	11,230.00	2,100.00	92,800.00
Supplier Payment Delay (days)	42	38	16	7	102

The descriptive statistics reveal important information about the characteristics of the SME credit data. For example, the average current ratio is 1.78, indicating that on average, SMEs in the dataset have current assets that are 1.78 times greater than their current liabilities. The average return on assets is 0.09, suggesting a moderate level of profitability among the SMEs. The debt to equity ratio has a minimum value of 0.08 and a maximum of 6.74 with an average of 1.42 an appropriate level considering the economic volatility historically prevalent in Zimbabwe.

4.2.2 Measure of skewness on the numerical variables

To assess the skewness of the numerical variables in the credit risk assessment model we calculated the skewness statistic for each variable. Skewness is a measure of the symmetry of a distribution, with a value of 0 indicating a perfectly symmetric distribution. Table 4.2 presents the skewness values for the numerical variables in the dataset.

Table 4.2

Variable	Skewness
Current Ratio	1.02
Return on Assets	0.41
Debt-to-Equity Ratio	1.78
Firm Characteristics	1.34
Management Experience (years)	0.92
Collateral Availability %	-0.59
Average Monthly Online Sales	2.51
Supplier Payment Delay (days)	0.71

Results in Table 4.2 indicate that Current Ratio has a skewness value of 1.02 which indicates a positive skew, meaning the distribution has a longer right tail. This implies that most SMEs have current ratios at the lower end of the distribution, with a few outliers having significantly higher current ratios. Also, the debt-to-equity ratio possesses a skewness value of 1.78 that represents a strong positive skew, indicating that most SMEs have low debt-to-equity ratios.

DOI: 10.25103/ijbesar.182.05

4.2.3 Correlation analysis

Table 4.3 presents the correlation matrix for the numerical variables.

Table 4.3 Correlation matrix

	Current Ratio	Return on Assets	Debt-to-Equity Ratio	Firm Age (years)	Management Experience (years)	Collateral Availability %	Average Monthly Online Sales	Supplier Payment Delay (days)
Current Ratio	1.00	0.31	-0.45	0.21	0.18	0.29	0.24	-0.13
Return on Assets	0.31	1.00	-0.37	0.15	0.22	0.19	0.27	-0.11
Debt-to-Equity Ratio	-0.45	-0.37	1.00	-0.16	-0.13	-0.24	-0.19	0.09
Firm Age (years)	0.21	0.15	-0.16	1.00	0.34	0.12	0.17	-0.08
Management Experience (years)	0.18	0.22	-0.13	0.34	1.00	0.14	0.13	-0.07
Collateral Availability %	0.29	0.19	-0.24	0.12	0.14	1.00	0.21	-0.10
Average Monthly Online Sales	0.24	0.27	-0.19	0.17	0.13	0.21	1.00	-0.12
Supplier Payment Delay (days)	-0.13	-0.11	0.09	-0.08	-0.07	-0.10	-0.12	1.00

The results in Table 4.3 indicate a current ratio that has a moderate positive correlation of 0.31 with return on assets and a moderate negative correlation of -0.45 with debt-to-equity ratio. This suggests that SMEs with higher current ratios tend to have higher profitability and lower leverage.

Firm age has a weak positive correlation of 0.21 with current ratio, of 0.15 with return on assets and of 0.17 average monthly online sales. This implies that firms with strong firm characteristics like growth potential tend to have slightly better liquidity, profitability, and online sales performance.

4.3 Feature analysis

We conducted a feature analysis for Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), and AdaBoost (AB). Tables 4.4 to 4.8 display coefficients which are also referred to as feature importance values. Feature importance values greater than 0.15 identify the most important variables. The feature-important analysis identifies firm characteristics, return on assets, current ratio, and debt-to-equity ratio as the most significant predictors of credit risk. Collateral availability and average monthly online sales have moderate importance, whilst management experience and supplier payment delay the lowest importance.

Table 4.4 Logistic Regression (LR) feature importance analysis

Variable	Coefficient
Current Ratio	0.72
Return on Assets	1.34
Debt-to-Equity Ratio	-0.59
Firm Characteristics	1.51
Management Experience (years)	0.11
Collateral Availability %	0.44
Average Monthly Online Sales	0.27
Supplier Payment Delay (days)	-0.09

DOI: 10.25103/ijbesar.182.05

Table 4.5 Support Vector Machine (SVM) feature importance analysis

Variable	Coefficient
Current Ratio	0.21
Return on Assets	0.24
Debt-to-Equity Ratio	0.18
Firm Characteristics	0.27
Management Experience (years)	0.07
Collateral Availability %	0.13
Average Monthly Online Sales	0.11
Supplier Payment Delay (days)	0.05

Table 4.6 K-Nearest Neighbors (KNN) feature importance analysis

Variable	Coefficient
Current Ratio	0.19
Return on Assets	0.22
Debt-to-Equity Ratio	0.16
Firm Characteristics	0.29
Management Experience (years)	0.08
Collateral Availability %	0.14
Average Monthly Online Sales	0.12
Supplier Payment Delay (days)	0.06

Table 4.7 Random Forest (RF) feature importance analysis

Variable	Coefficient
Current Ratio	0.20
Return on Assets	0.23
Debt-to-Equity Ratio	0.17
Firm Characteristics	0.31
Management Experience (years)	0.08
Collateral Availability %	0.15
Average Monthly Online Sales	0.13
Supplier Payment Delay (days)	0.07

Table 4.8 AdaBoost (AB) feature importance analysis

Variable	Coefficient
Current Ratio	0.18
Return on Assets	0.21
Debt-to-Equity Ratio	0.16
Firm Characteristics	0.10
Management Experience (years)	0.07
Collateral Availability %	0.13
Average Monthly Online Sales	0.11
Supplier Payment Delay (days)	0.06

DOI: 10.25103/ijbesar.182.05

4.4 Model performance and evaluation

We now focus on the metrics of accuracy, precision, recall, and F1-score. The results for each metric are presented in Table 4.9 providing a comparison of the performance of the models.

Table 4.9 Model performance

Metric	Logistic Regression	SVM	KNN	Random Forest	AdaBoost
Accuracy	0.83	0.82	0.78	0.85	0.81
Precision	0.81	0.80	0.76	0.83	0.79
Recall	0.85	0.84	0.08	0.87	0.83
F-1 Score	0.83	0.82	0.78	0.85	0.81

4.4.1 Overall model performance

The random forest model demonstrated the best performance across all the evaluated metrics. The logistic regression and SVM models also showed strong performance, with the AdaBoost and KNN models comparatively being the weakest. Further evaluation of the model's performance using different subsets of the data or ensemble techniques may achieve improved performance.

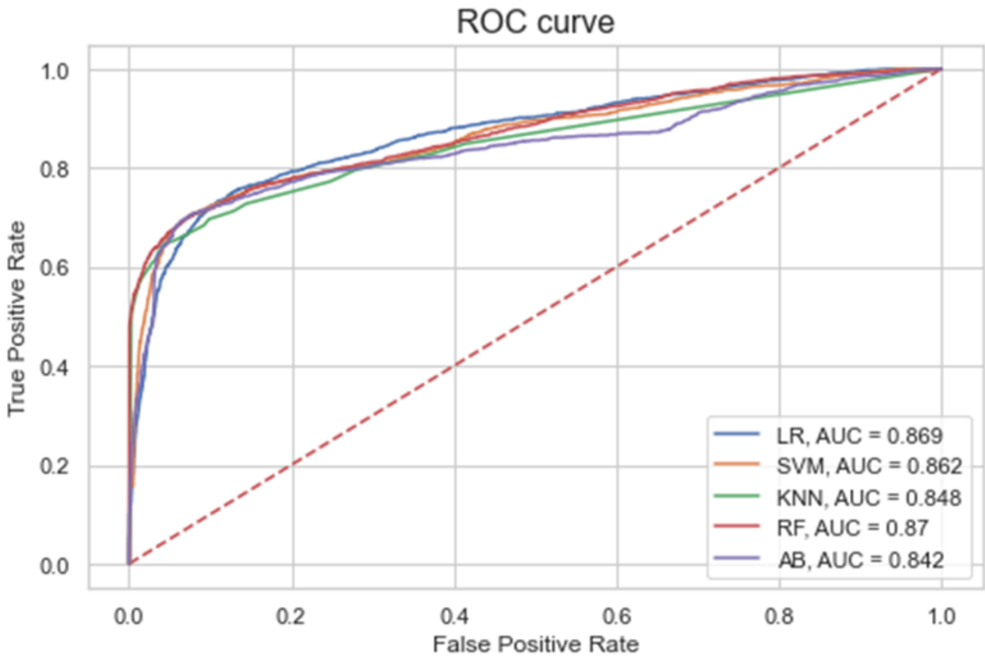
4.4.2 ROC-AUC analysis

Accuracy is often used to evaluate the performance of classification models. However, it encounters limitations when carrying out credit risk assessment. Accuracy measures the overall proportion of correct predictions but is incapable of providing a complete picture of the model's ability to differentiate between credit-worthy and high-risk borrowers (Malik et al., 2024; Dong, Liu and Tham, 2024; Wang, 2024; Wu et al., 2025).

In credit risk assessment, the cost of misclassification can be quite high. For example, incorrectly classifying a high-risk borrower as a low-risk borrower (false negative) could lead to substantial financial losses for the lender (Wang, 2024; Wu et al., 2025). Conversely, incorrectly classifying a low-risk borrower as high-risk (false positive) could result in missed lending opportunities and reduces financial inclusivity (Wang, 2024; Wu et al., 2025).

ROC-AUC when compared to accuracy is a more comprehensive metric that evaluates the model's ability to distinguish between these two classes. It is not affected by the specific classification threshold. ROC-AUC measures the trade-off between the true positive rate (sensitivity) and the false positive rate (1 - specificity) across all possible classification thresholds (Wang, 2024; Wu et al., 2025). Figure 4.1 contains the ROC curve.

Figure 4.1ROC curve



The Random Forest model emerges as the best performer, with an ROC-AUC score of 0.87. This indicates that the ensemble nature of the Random Forest model allows it to effectively handle the complexity of the data and provide highly accurate predictions for credit risk assessment. This suggests that the model can accurately rank the instances based on their predicted probabilities of default, making it a suitable choice for credit risk assessment.

The Logistic Regression model also exhibits a strong ability to distinguish between credit-worthy and high-risk borrowers, with an ROC-AUC score of 0.869. The KNN model also demonstrates good performance in discriminating between credit risk classes, with an ROC-AUC score of 0.848. This indicates that the model can effectively capture the complex, non-linear relationships in the data and provide reliable predictions for credit risk assessment.

4.4.3 Interpretation of model findings

The results are broadly consistent with prior credit risk literature, which typically finds that ensemble methods outperform linear models (Xu, (2025), Miliūnaitė, (2023), Wang et al. (2020)). The strong performance of Random Forest aligns with studies showing that tree-based methods better capture complex borrower behavior. The superior performance of Random Forest in your analysis can be attributed to several factors.

- Non-linear relationships: It effectively captures complex, non-linear interactions between variables (e.g., between liquidity, leverage, and credit history) that linear models like Logistic Regression cannot.
- Interaction effects: Random Forest automatically models interactions between predictors without explicit specification.
- Robustness to noise and overfitting: Through bootstrapping and averaging (bagging), it reduces variance and improves generalisation. Other research studies used the random forest model when performing credit risk assessment and had similar findings (Xu, (2025); Miliūnaitė, (2023); Wang et al. (2020)).

The comprehensive analysis of the credit risk assessment model's performance using various machine learning approaches has provided valuable insights into the key drivers of credit risk management for SMEs in Zimbabwe.

4.4.4 Model interpretation and discussion

The feature importance analysis across the Logistic Regression, SVM, KNN, Random Forest, and AdaBoost models consistently identified the following variables as the most significant predictors of credit risk:

- Return on Assets: SMEs with higher profitability, as measured by return on assets, were less likely to default on their loans. This finding aligns with previous studies that have emphasised the importance of financial performance indicators in credit risk assessment (OECD, 2025).
- Current Ratio: SMEs with a higher current ratio, indicating stronger liquidity and ability to meet short-term obligations, were likely to be classified as low-risk borrowers. This supports the notion that liquidity is a crucial factor in creditworthiness (D'Amato, 2020; Rusu and Roman, 2022; Li et al., 2025; López-Gracia and Sogorb-Mira, 2008).
- Debt-to-Equity Ratio: Higher debt-to-equity ratios, which signify higher financial leverage, were associated with increased credit risk. This is consistent with the findings of other studies who highlighted the importance of capital structure in predicting SME defaults (Feng, Li and Peng 2021; Paeleman et al., 2024).

In addition to these financial performance indicators, the analysis also revealed the significance of other firm-level characteristics:

- Firm characteristics: SMEs with higher growth potential, higher customer concentration and greater firm age were found to have the lowest credit risk profile, suggesting that the longevity and experience of a business can be an important factor in credit risk assessment. This aligns with other studies who noted the role of firm age in reducing information asymmetries and improving access to credit (Pillay, 2024; Garcia-Martinez et al., 2023; Ciampi, et al., 2021; Altman, Sabato and Wilson, 2010).
- Collateral availability: SMEs with higher collateral availability were found likely to be classified as low-risk borrowers. This emphasises the importance of collateral as a risk-mitigating factor, as demonstrated in previous studies (Ciampi, 2021; Pillay, 2024). Lack of collateral is a challenge commonly faced by SMEs in Zimbabwe impeding their access to credit facilities.

Also, the alternative data source variables have a medium contribution towards credit rating. Evidence of this is the consistent middle rating of monthly online sales when compared to traditional financial and non-financial variables. However, this variable outperforms other variables such as management experience. This points to a need to include alternative data sources as a means of improving the accuracy of credit risk assessments and predictions.

DOI: 10.25103/ijbesar.182.05

4.4.5 Credit risk assessment for SMEs with inadequate collateral in Zimbabwe

We suggest that when assessing the credit risk for SMEs applying for loans, without collateral, in Zimbabwe various firm characteristics beyond just the financial ratios of the borrower should also be taken into consideration. Factors such as firm age, growth potential, type of industry, macroeconomic conditions and customer concentration can provide important insights into the creditworthiness of the borrower.

Firm age is a significant indicator of a company's stability and creditworthiness. Older firms typically have a better track record, more established business processes and stronger relationships with suppliers and customers, making them less risky borrowers. Younger firms, on the other hand, may face higher uncertainty and a higher risk of failure which can increase the credit risk associated with lending to them (Pillay, 2024; Garcia-Martinez et al., 2023; Ciampi, et al., 2021; Altman, Sabato and Wilson, 2010).

The growth potential of a firm is also an important consideration in credit risk assessment. Companies with strong growth prospects, as indicated by factors such as industry outlook, market share, and innovation, may be better positioned to generate the cash flow needed to repay their loans, even in the absence of collateral (Abbasi and Tamoradi, 2020). Conversely, firms with limited growth potential may struggle to generate sufficient income to service their debt obligations.

The type of industry in which a firm operates can also have a significant impact on its credit risk. Some industries may be more volatile or cyclical, making them riskier investments, while others may be more stable and resilient. Incorporating industry-specific factors, such as market trends, competitive landscape and regulatory environment, can help lenders better understand the risks associated with lending to a particular company.

Macroeconomic conditions can also influence a firm's creditworthiness. Economic factors, such as GDP growth, inflation, interest rates, and unemployment, can affect a company's sales, profitability, and cash flow, thereby impacting its ability to repay its loans. Lenders should closely monitor the macroeconomic environment and adjust their credit risk assessments accordingly.

Finally, the concentration of a firm's customer base can have a significant impact when assessing credit risk. Companies with a high degree of customer concentration may be more vulnerable to the loss of a major client, which could severely impact their cash flow and ability to repay their loans. Lenders should carefully evaluate the diversity and stability of a borrower's customer base to better understand the associated risks.

By considering these firm-level characteristics in addition to financial ratios, lenders can gain a more comprehensive understanding of the credit risk associated with a particular borrower, even in the absence of collateral. This holistic approach to credit risk assessment can help lenders make more informed decisions and minimise the risk of default in their loan portfolios.

5. Conclusion and recommendations

5.1 Conclusions

The study provides a comprehensive assessment of the key factors influencing SME creditworthiness in Zimbabwe. It finds that, in addition to traditional financial indicators such as profitability, liquidity and leverage, non-financial factors—including collateral availability, cash flow stability, management quality, and macroeconomic conditions—play a significant role in shaping credit risk profiles. This highlights the importance of adopting a more holistic approach to credit risk evaluation that extends beyond conventional financial metrics.

The findings also emphasise the growing importance of alternative data in improving credit risk assessment. Non-traditional data sources such as firm characteristics, supplier–buyer relationships, and social media activity can provide deeper insights into SMEs' operational performance and risk exposure. Incorporating these data sources alongside traditional financial information can significantly enhance the prediction of defaults. At the same time, the study underscores the strong impact of Zimbabwe's macroeconomic, where inflation, and policy uncertainty can undermine SME performance and reduce the reliability of traditional credit models.

Another key issue identified is the role of information asymmetry and institutional weaknesses in limiting SME access to finance. The lack of comprehensive credit data, effective collateral registries, and well-functioning credit bureaus makes it difficult for lenders to accurately assess borrower risk. Addressing these gaps is essential to improving credit allocation. Strengthening financial infrastructure, improving data availability, and reducing information asymmetries would enable lenders to make more informed decisions and expand lending to SMEs.

Based on these findings, several recommendations are proposed. Policymakers should strengthen credit infrastructure, implement credit guarantee schemes, improve financial literacy, and create an enabling regulatory environment while promoting alternative financing mechanisms. Financial institutions are encouraged to adopt context-specific credit risk models that integrate alternative data, diversify lending products, leverage technology, and enhance engagement with SMEs. Meanwhile, SME support agencies should focus on capacity building, data collection, financial education, and fostering partnerships and networks. Collectively, these measures can improve access to finance and support the growth and sustainability of SMEs in Zimbabwe.

DOI: 10.25103/ijbesar.182.05

5.2 Suggestions for further studies

Further studies could consider the potential of integrating qualitative data, such as management interviews, industry benchmarks and expert assessments into the credit risk assessment process to capture other soft factors that may influence SME creditworthiness in Zimbabwe. We could also conduct a longitudinal study to examine the evolution of credit risk profiles among Zimbabwean SMEs over time, capturing the impact of macroeconomic and political changes. This could boost the effectiveness of policy interventions in mitigating default risks. Another research study direction could assess the viability and potential impact of innovative financing solutions. These solutions could include crowdfunding platforms, supply chain financing schemes, and specialised SME lending programs. These can help address the credit access challenges faced by Zimbabwean SMEs.

5.3 Limitations of the study

This study was confined to one Zimbabwean bank, this represents a narrow focus since the Zimbabwean banking industry has 343 players as at 30 September 2025, (Reserve Bank of Zimbabwe (2025)). Also, since we base our research on Zimbabwe, it implies that the findings of this study cannot be generalised to all developing countries.

5.4 Declarations

All authors declare that they have no conflicts of interest. There was no breach of ethical rules and guidelines. All materials and information were not subject to confidentiality restrictions.

References

Abbasi, E. and Tamoradi, A., 2020. *The effect of customers concentration on company risks.* Iranian Journal of Finance, 4(2), pp.19–39. doi:10.22034/ijf.2020.227184.1118.

Ahmad, J., 2026. Data-Driven Underwriting: Leveraging Alternative Data Sources Responsibly. *International Journal of Emerging Trends in Computer Science and Information Technology*, 7(1), pp.92-96.

Altman, E.I., Sabato, G. and Wilson, N., 2010. The value of non-financial information in SME risk management. *Journal of Credit Risk*, *6*(2), pp.95-127.

Bielecki, T.R. and Rutkowski, M., 2013. *Credit risk: modeling, valuation and hedging*. Springer Science & Business Media.

Ciampi, F., Giannozzi, A., Marzi, G. and Altman, E.I., 2021. Rethinking SME default prediction: a systematic literature review and future perspectives. *Scientometrics*, *126*(3), pp.2141-2188.

D'Amato, A., 2020. Capital structure, debt maturity, and financial crisis: empirical evidence from SMEs. *Small Business Economics*, *55*, pp.919–941.

Dhlandhlara, M., 2019. *Barriers to accessing innovative financing techniques for small and medium enterprises (SMEs) in Zimbabwe: The case of SMEs in Harare.* Doctoral dissertation.

Dlamini, B. and Schutte, D.P., 2020. An overview of the historical development of Small and Medium Enterprises in Zimbabwe. *Small Enterprise Research*, *27*(3), pp.306-322.

Dong, H., Liu, R. and Tham, A.W., 2024. Accuracy comparison between five machine learning algorithms for financial risk evaluation. *Journal of Risk and Financial Management*, *17*(2), p.50.

Feng, C., Li, Z. and Peng, Z., 2021. The impact of banking competition on firm credit risk and leverage. *Sage Open*, *11(4)*, p.21582440211061529.

FinScope, 2022. *Micro, small and medium enterprises (MSME) survey highlights.* Technical Report.

Garcia-Martinez, L.J., Kraus, S., Breier, M. and Kallmuenzer, A., 2023. Untangling the relationship between small and medium-sized enterprises and growth: a review of extant literature. *International Entrepreneurship and Management Journal*, *19*(2), pp.455-479.

Jiang, M., Shi, J., Zheng, Y. and Zhou, W., 2026. The Role of Alternative Data in Micro-Enterprises' Credit Risk Assessment in China—Empirical Evidence Based on Machine Learning. *Journal of Behavioral and Experimental Finance*, p.101154.

Jrad, M., 2023. Examining collateral prerequisites for small and medium-sized business loans. *International Journal of Membrane Science and Technology*, *10*(2), pp.1906-1922.

Karedza, G., Sikwila, M.N., Mpofu, T. and Makurumidze, S., 2014. *An analysis of the obstacles to the success of SMEs in Chinhoyi Zimbabwe.* [no journal or publisher info provided].

Lee, J.Y., Yang, J. and Anderson, E.T., 2026. Who benefits from alternative data for credit scoring? Evidence from Peru. *Journal of Marketing Research*, 63(1), pp.105-126.

Li, R., Chen, C., Han, Z. and Wang, Y., 2025. Targeted monetary policy, SMEs' loan availability, and corporate investment: evidence from China. *Small Business Economics*, pp.1-27.

López-Gracia, J. and Sogorb-Mira, F., 2008. Testing trade-off and pecking order theories financing SMEs. *Small Business Economics*, *31*(2), pp.117-136.

Makanyeza, C. and Dzvuke, G., 2015. The influence of innovation on the performance of small and medium enterprises in Zimbabwe. *Journal of African Business*, *16*(1-2), pp.198-214.

Malik, P., Chourasia, A., Pandit, R., Bawane, S. and Surana, J., 2024. Credit risk assessment and fraud detection in financial transactions using machine learning. *Journal of Electrical Systems*, *20*(3s), pp.2061-2069.

Manyanga, W., Kanyepe, J., Chikazhe, L. and Manyanga, T., 2023. The effect of debt financing on the financial performance of SMEs in Zimbabwe. *Cogent Social Sciences*, 9(2), p.2282724.

Matsongoni, H. and Mutambara, E., 2021. Challenges faced by the informal small to medium enterprises—a case study of the manufacturing sector in Zimbabwe. *International Journal of Entrepreneurship*, 25, pp.1–17.

Miliūnaitė, L., 2023. Evaluating the credit risk of SMEs using artificial intelligence, financial and alternative data (Doctoral dissertation, Kauno technologijos universitetas.).

Muriithi, S.M., 2017. African small and medium enterprises (SMEs) contributions, challenges and solutions. *European Journal of Research and Reflection in Management Sciences*, *5*(1), pp.36–48.

Njanike, K., 2019. The factors influencing SMEs growth in Africa: A case of SMEs in Zimbabwe. In: *Regional Development in Africa*. London: IntechOpen.

OECD (2025). *OECD Financing SMEs and Entrepreneurs Scoreboard: 2025 Highlights*. OECD Publishing. Available at: https://www.oecd.org/en/publications/oecd-financing-smes-and-entrepreneurs-scoreboard-2025-highlights_64c9063c-en.html [Accessed 17 Jul. 2025].

Paeleman, I., Guenster, N., Vanacker, T. and Siqueira, A.C.O., 2024. The consequences of financial leverage: Certified B Corporations' advantages compared to common commercial firms. *Journal of Business Ethics*, *189*(3), pp.507-523.

Pillay, S., 2024. An Introduction to the "SME Rating", and Specific "SME Credit Rating" Concepts. In *The Next Frontier in SME Ratings: Improving Access to Credit and Business Opportunities for the World's Largest Business Segment* (pp. 27-54). Cham: Springer Nature Switzerland.

Reserve Bank of Zimbabwe (2025) *Banking sector report for the period ended 30 September 2025*. Harare: Reserve Bank of Zimbabwe.

Rusu, V.D. and Roman, A., 2022. The relationship between financing decision of SMES and their performance. In *Business Development and Economic Governance in Southeastern Europe: 13th International Conference on the Economies of the Balkan and Eastern European Countries (EBEEC), Pafos, Cyprus, 2021* (pp. 353-367). Cham: Springer International Publishing.

Wang, Y., Zhang, Y., Lu, Y. and Yu, X., 2020. A Comparative Assessment of Credit Risk Model Based on Machine Learning—a case study of bank loan data. *Procedia Computer Science*, 174, pp.141-149.

Wang, Z. (2024). *Artificial Intelligence and Machine Learning in Credit Risk Assessment: Enhancing Accuracy and Ensuring Fairness*. Open Journal of Social Sciences, 12(11), pp.1–12. Available at: https://www.scirp.org/pdf/jss20241211_221769377.pdf [Accessed 17 Jul. 2025].

Wu, M., Liu, Y., Liu, Y., and Li, Y. (2025). *Preserving AUC Fairness in Learning with Noisy Protected Groups*.arXivpreprintarXiv:2505.18532. Available at: https://arxiv.org/html/2505.18532v1 [Accessed 17 Jul. 2025]

Xu, Z., 2025. Credit Scoring Using Alternative Data Sources: A Machine Learning Approach. *Academic Journal of Computing & Information Science*, 8(4), pp.56-64.

DOI: 10.25103/ijbesar.182.05