Journal of Artificial Intelligence Research, vol. 1, no.2, pp. 1-14, NOVEMBER 2024

An Analysis of Credit Risk Prediction for Small and Micro Enterprises

Wei Hu[a], Yuhuan Wu[a] and Ziting Yang[a]

[a]School of Information Resource Management, Renmin University of China, Beijing, China

ARTICLE INFO

Keywords:
SMEs
credit risk
machine learning
data analysis

ABSTRACT

Digital inclusive finance has emerged as a significant catalyst for the high-quality development of small and micro enterprises (SMEs). This study, grounded in credit risk prediction theory, develops a comprehensive profiling and predictive model for SMEs, offering insights into innovative mechanisms by which inclusive finance can support their sustainable growth. Utilizing an extensive literature review, along with experimental modeling based on publicly available data, the study explores two approaches to feature construction. By employing diverse algorithms, it builds predictive models and proposes tailored policy recommendations to enhance inclusive financial practices. The credit prediction model facilitates targeted financial support and innovation management strategies for SMEs, contributing a fresh perspective to advancing the quality and effectiveness of digital inclusive finance.

1. Introduction

China's economy is currently at a pivotal stage, with efforts focused on consolidating recovery and promoting sustained growth. Central to this economic development are the numerous small and micro enterprises (SMEs), which serve as a vital foundation. President Xi Jinping has emphasized the importance of supporting SMEs, underscoring the need to address the real challenges they face. In response, the Ministry of Finance and the State Administration of Taxation have continuously issued tax policy announcements aimed at facilitating SME financing by offering tax incentives. SMEs play a critical role in China's economic structure, and fostering their growth has become a key national priority.

On the one hand, there are nearly 50 million SMEs in China, providing significant employment opportunities and contributing nearly two-thirds of the country's income tax. They are a driving force in national economic development and the building of a modern socialist state. On the other hand, SMEs have faced ongoing challenges. Studies show that during the pandemic, nearly 85% of SMEs lacked sufficient cash flow to sustain operations for more than three months, and the implementation of tax relief policies has been less effective than anticipated, leading to negative sentiment among businesses [38]. Additionally, changes in interpersonal interactions have made it difficult for businesses to resume normal operations [27]. While SMEs have received considerable attention and support, particularly in terms of financing, much room for improvement remains.

Over the years, both domestic and international scholars have made significant strides in studying the development of SMEs. One line of research has focused on the classification and characteristics of SMEs, analyzing data related to their basic information, operational status, and financing needs to extract key features for categorization. This research has revealed substantial differences in market competition, innovation capacity, and growth potential among SMEs. Furthermore, different types of SMEs exhibit varying characteristics regarding financing needs and credit risk. Another line of research has concentrated on SME financing needs and credit risk, revealing common issues such as insufficient credit limits, high financing costs, and a lack of collateral. Credit risk, in particular, has been shown to be closely linked to a company's operational status, credit rating, and the type of collateral provided. Overall, substantial progress has been made in profiling SMEs and analyzing their credit risk [28].

Nevertheless, research on SME profiling and their financing needs still has several limitations: (1) The potential of FinTech in SME financing has not been fully realized. As the market and environment continue to evolve, the credit risk associated with SMEs is also changing, making the development of accurate risk prediction models an ongoing challenge. (2) Existing studies lack effective analytical models for enterprises with incomplete information disclosure or those where data acquisition is difficult. Therefore, this study raises the following research questions: (1) How can we lower the barriers to entry and utilize more accessible information to develop a credit risk prediction model for SMEs that is both interpretable and practically useful? (2) How can we design appropriate credit and subsidy policies, and how can FinTech be applied to enhance the development of SMEs?

This study makes the following contributions: (1) It develops a machine learning-based credit risk prediction method using publicly available financial statement indicators, highlighting the accessibility of data, the interpretability of key metrics, and model performance. Specifically, this study selects and compares two types of indicators–static financial indicators and core growth rates–to assess their effectiveness in predicting credit risk. Four machine learning models are compared: logistic regression, neural networks, k-nearest neighbors (KNN), and support vector machines (SVM). Ultimately, this research proposes a FinTech-driven solution for managing SME credit risk, offering a novel perspective. (2) Based on the findings and existing literature,

DOI: https://doi.org/10.70891/JAIR.2024.110004
ISSN of JAIR: 3078-5529

this study provides three key recommendations to promote the healthy and sustainable growth of SMEs: recommendations on credit policies, government subsidies, and FinTech strategies. The structure of this paper is as follows: the first section introduces the topic, followed by a literature review in the second section. The third section covers research design and sample selection, the fourth section presents the model experiments, the fifth section offers policy recommendations, and the sixth section concludes the study. This article improves the flow, academic tone, and precision of the text while retaining the original meaning.

2. Literature Review

In today's competitive and uncertain economic environment, small and micro enterprises (SMEs) play a pivotal role in driving innovation, job creation, and societal prosperity. At the 10th Lujiazui Forum in June 2018, Yi Gang, Governor of the People's Bank of China, emphasized that SMEs contributed to 80% of employment, 70% of patent rights, over 60% of GDP, and more than 50% of national tax revenue. However, the COVID-19 pandemic has intensified the challenges SMEs face, particularly in market competition and financing difficulties, necessitating targeted strategies for sustainable development [39]. This paper aims to investigate issues related to SME financing and risk prediction, providing new approaches and methods to support their growth.

The term "SMEs" was coined by economist Professor Lang Xianping and includes small enterprises, microenterprises, family workshop businesses, and individual proprietors. SMEs account for more than 99% of all businesses in China, making them a critical part of the country's economic landscape and a driving force for economic growth. Academic research on SMEs primarily focuses on areas such as improving innovation capacity, promoting transformation and upgrading, addressing financing challenges, and predicting risk. This paper focuses on SME financing and risk prediction.

2.1. Research on SME Financing

Existing research on the financing difficulties of SMEs is extensive and thorough. Jin and Zhang identified several challenges that commercial banks face when supporting SMEs, such as limited loan amounts, inadequate risk mitigation measures, the high vulnerability of SMEs, and the relatively low predictability of their risks [13]. They advocated for commercial banks to develop business operations and risk management systems tailored to the unique characteristics of SMEs to support their development while maximizing profits. Shi brought four major financing models–P2P lending, crowdfunding, e-commerce, and Internet direct banking–that could efficiently and suitably meet the financing needs of SMEs in China [25]. Yan analyzed reasons of financing difficulties for SMEs through a survey of over 100 entrepreneurs in Xi'an, a key city in the Belt and Road Initiative, and proposed targeted solutions [32]. Jiang examined factors influencing SME financing behavior in Guangdong Province, including operational performance and industry type [12]. Ye et al. conducted a quantitative analysis and developed a dynamic equilibrium model, concluding that inclusive Internet finance could increase the number of SMEs by 619.91% and their output by 2.72% [35]. Xu et al. suggested that balancing economic benefits and social responsibility could result in mutual benefits between fixed sales revenue and employee wages when loan amounts are constrained, offering new insights into policy design [29]. Yang and Zhang emphasized the importance of digital finance for economic growth and the healthy development of SMEs, using firms listed on the New Third Board as examples, and argued that digital finance could prevent financial crises caused by imbalances in capital structure and liquidity [34]. Cheng and Yang, through a 2015 survey of SMEs, demonstrated that company size, R&D, and core business areas were key factors for financial institutions when reviewing loan applications, with no evidence of discrimination in the lending process [6]. Xu et al. found that despite bank loans being available as a finance, SMEs often bear additional debts, indicating that bank loans and other forms of debt are not perfect substitutes [30].

2.2. Research on SME Risk Prediction

Various methods have been applied to assign credit risk. Gu et al. used imbalanced sample processing techniques to ensure balanced representation and employed different algorithms for early credit risk warnings, improving financing efficiency [9]. Chen and Wang developed a credit default estimation model based on incomplete information of corporate integrating cash flow, default boundaries, and actual cash flow distributions to accurately estimate the likelihood of SME defaults [4]. Li et al. built a multi-layer fuzzy comprehensive evaluation system for credit assessment, validating the effectiveness through case studies of dozens of companies [17]. Feng et al. highlighted the higher credit acceptance rates of urban and high-skilled labor-intensive companies influencing the need for SMEs to expand employment and attract educated talent [7]. Lu et al. employed convolutional neural networks (CNN) to create an evolving risk model, addressing complex financial decision-making and regulatory challenges in the supply chain industry while enhancing feature recognition capabilities [22]. Brancati, using specific business loan data, found that SMEs can benefit from the accumulated knowledge of lending banks, and building strong relationships with banks helps SMEs overcome financial barriers to innovation [2]. Zhang and Wang used quantitative methods to show that smaller, non-state-owned SMEs receive less financial support and tend to perform poorly in economic opportunities, indicating a need for more refined support policies [42]. Stevenson et al. applied deep learning and natural language processing (NLP) to extract information from over 60,000 texts, creating a text-based credit risk approach demonstrating how the results could impact business performance [26]. Liang et al., using theories of information asymmetry and signaling, identified three dimensions of information communication in SME crowdfunding and argued better information descriptions are crucial to crowdfunding success [19]. He and Shen argued that the declining efficiency of capital in the real economy is a significant cause of financial misallocation, and that increasing economic policy uncertainty reduces the support digital finance can provide to the real economy, highlighting the need for accurate SME credit risk assessment [10].

2.3. Implications and Gaps in Existing Research

The aforementioned studies highlight the necessity and significance of researching SME financing and risk prediction, yielding insights. However, there remains a lack of comprehensive research in these areas, indicating that further investigation is urgently needed.

3. Credit Risk Prediction Model Building and Features Selection

Credit risk prediction for MSMEs (Micro and Small Enterprises) has been one of the research hot spots in the field of finance over the past decades. The research goal in this field is to construct accurate and reliable models to help financial institutions and policy makers better understand, assess and control the credit risk of MSMEs. In order to achieve this goal, researchers have actively explored different methods and techniques over the past decades, and have accumulated some experience in the screening of enterprise characteristics and the iterative application of models.

The first is the automation of the enterprise credit risk prediction approach. Bank credit operations are an important method used for corporate credit risk prediction. Traditional credit operations need to be implemented by specialized credit bureaus and require the processing, handling, and reporting of financial and trading information on targeted customers [24]. Traditional credit collection relies on the manual collection of corporate information and requires the review of that information. This review process is usually time-consuming, inaccurate, and does not reflect dynamic trends in business operations. 2015 was the first year of online credit development, and automated online credit became possible due to the real-time generation of large amounts of data and FinTech's ability to analyze large amounts of data. As automated risk prediction can optimize the credit-granting process, improve the efficiency of risk identification for MSMEs [20], and reduce financial risks caused by information asymmetry [14]. Therefore, enterprise risk prediction based on automation is increasingly becoming the development trend of risk control business in the financial industry.

The second is the selection of credit data. In past studies, scholars have extensively examined a variety of firm characteristics to determine their association with credit default. Among them, accounting information in financial statements is the most traditional and widely used credit data for corporate credit risk analysis. However, there is still no consistent answer on how to screen data in financial statements for risk prediction [20]. The information on characteristics other than the individual firms has also been used as a basis for risk prediction. Characteristic information outside the individual enterprise is also used for risk prediction data selection, including the enterprise's industry type, size, age, ownership structure and so on. Some studies have shown that industries such as finance [23] and real estate [15] may be more susceptible to macroeconomic fluctuations and therefore have different performance in terms of credit risk. In recent years, different corporate characteristics have been used for credit risk prediction of MSMEs. The use of text mining technology on the textual information disclosed in the annual reports of listed companies can effectively forecast the financial risk of A-share listed companies, but MSMEs are not subject to the strict disclosure system of listed companies, and the data disclosure is incomplete and unstandardized, so this kind of data is not well used for the risk analysis of MSMEs. Based on the LS-SVM model, a study utilizes the transaction information of MSMEs in provincial branches of a state-owned commercial bank recorded by the commercial bank, and the results of the study have a certain degree of accuracy and advantage. A study uses the non-financial attributes in the credit database of a local government's MSMEs to describe the characteristics of the enterprise in terms of its employees, composition structure, and historical behaviors, and the generalization results are poor.

In summary, for the problem of enterprise credit risk prediction, at the technical level the risk monitoring system is moving towards automation with the development of science and technology, while at the business level, there are still different opinions on how to select enterprise characteristics to build a picture. It is undeniable that past research has provided an important theoretical and empirical foundation for credit risk prediction of MSMEs based on important financial statement indicators. However, with the continuous changes in the financial market and corporate environment, accurate prediction of credit risk is still a challenging task. Moreover, most of the existing studies conducted credit risk assessment experiments on listed companies with more complete disclosure, and there are fewer studies on the credit risk of MSMEs with incomplete disclosure and high difficulty in obtaining disclosure, and the experimental results are poor. Therefore, this study aims to further explore and improve the credit risk prediction model for MSMEs on the basis of previous work to better meet the needs of financial institutions and policy makers.

The innovation of this study is that some machine learning risk prediction methods based on publicly available financial statement indicators are constructed, and this study has three features. First, the accessibility of public financial statements, companies listed on the New Third Board are also required to disclose public financial data, which makes this method more operational and provides a wider range of services for financial institutions; second, the financial statement indicators are highly interpretable. In the field of machine learning and data mining, interpretability is defined

Table 1
Enterprise risk term selection

Announcement code	Type of announcement
1605	Reports on financial risk categories such as debt default
1606	Substantial loss, material damage, or substantial liability
1607	Announcement of major assets being seized, impounded, frozen and other
1613	Insolvency and liquidation announcements

Table 2
Sample positive and negative example data

Type of examples	Number
Positive examples (no credit risk)	3848
Negative examples (with credit risk)	222

as the ability to explain or in order to present understandable terms to humans [8]. It includes the interpretability of the model's independent variables and the interpretability of the model learning process. This study ensures the former. Thirdly, the study compares models with different performances, and this comparison helps to identify the most suitable model for a particular context, thus improving the accuracy and usefulness of credit risk prediction. This innovative point provides financial practitioners with more choices and helps them to make informed decisions based on specific situations.

3.1. Data Selection

The data sources in this section are divided into two parts. First, we selected the accounting data of 4,070 SMEs listed on NSS as of July 1, 2023 from the wind database. Second, we manually screened a total of 1,140 risk announcements of all risk announcements disclosed in the information of the National Small and Medium-sized Enterprises Stock Transfer System (NSSETS) from June 30, 2020 to July 1, 2023, and the screening conditions are shown in Table 1 of the selection of enterprise risk items.

After that, the risk announcement and the risky enterprises were corresponded, and the correspondence found that there were 269 enterprises with credit risk in total among 4070 SMEs. Next, our process of data preprocessing, we completed the removal of some high null rate indicators based on Python programming, and filled in some enterprise accounting data based on the mean value filling method for the missing important indicator items.

3.2. Selection of Important Features

The public financial statements of enterprises contain rich information about enterprise portraits. After Table 2 joining the 4,070 pieces of data, there are 70 data items covering the financial data of the enterprise since three years, including the three-year performance of the enterprise in terms of assets and liabilities, cash flow, and operating income. In order to achieve the interpretability of the model, these indicators should be screened appropriately. In this part, from the static financial indicators, core growth rate of two types of indicators, screening two types of features, comparing the training effect of the two types of features under different models, and giving a new type of solution for the use of FinTech in the new situation of inclusive financial risk control.

3.2.1. Selection of Characteristics Based on Key Financial Analysis Indicators

Multi-period financial accounting statements are an important tool for investment decision makers to access and assess the financial position of an enterprise. These statements provide a variety of financial metrics, including a balance sheet and a statement of profit and loss, that are used to reflect the financial health of a business. Some of the core metrics can be additionally derived through calculations that provide important clues to insight into the financial condition and soundness of the business, including current ratio, debt ratio, debt service coverage, and so on. They play a key role in financial analysis, and investors and analysts often rely on these metrics to assess the financial risk and value of a business.

Academic findings suggest that these core financial indicators are important for assessing the financial soundness and future performance of an enterprise. For example, total assets reflect a firm's size and asset base, total liabilities reveal a firm's debt burden, and equity attributable to shareholders of the parent company represents owners' equity. The dynamic changes and growth rates of these indicators can help analysts better understand the business operations. In addition, cash and cash equivalents are critical to a business's liquidity and ability to pay. These core financial indicators provide investors and analysts with a powerful tool in assessing the financial position of a business, informing decision-making and helping to reveal potential risks and opportunities. In this paper, seven types of characteristics based on important financial analysis indicators are selected as follows.

3.2.2. Selection of Features Based on Core Growth Rate Indicators of Enterprises

Multi-period financial accounting statements are a key source of information for investment decision makers to obtain and assess the financial health of an enterprise. They contain a range of financial indicators that reflect the business performance of the enterprise, with a particularly striking feature being the growth rate of each financial indicator. These growth rate indicators, such as the growth rate of total

Table 3
Characteristics of classes based on significant financial analysis indicators

Indicator name	Calculation method	Meaning
Current ratio	Current assets / Current liabilities	A measure of a firm's short-term solvency, reflecting whether the firm is able to service its short-term debt from current assets. Generally, a higher current ratio indicates better short-term solvency.
Gearing	(Total liabilities / Total assets) × 100%	Indicates the proportion of a firm's total assets that are backed by debt funding, and it measures the degree of financial leverage of the firm. Higher debt ratios may indicate that firms are taking on greater debt risk.
Sales margin	(Net profit / Sales revenue) × 100%	Indicates the percentage of profit a business realizes per unit of product sold or service provided. A higher sales margin indicates that the business realizes a higher profit on sales.
Total asset margin	(Net profit / Total assets) × 100%	Measures the profitability realized out of the total assets of a business. It reflects the overall business performance of the firm and a higher profitability of total assets is usually considered good.
Short-term debt-servicing capacity	(Current assets - Inventories) / Current liabilities	Indicates the ability of a firm to service its short-term debt without regard to inventory. Higher short-term debt service capacity is usually considered beneficial.
Long-term debt-servicing capacity	(Long-term liabilities / Equity attributable to parent company) × 100%	Assesses the ability of the enterprise to service its long-term debt. Lower long-term debt servicing capacity may indicate that the enterprise is less financially stable in the long term.
Total asset turnover	Sales revenue / Total assets	Indicates the sales revenue generated by an enterprise per unit of total assets. A higher total asset turnover ratio indicates that the enterprise is utilizing its assets more efficiently.
Fixed asset turnover	Sales revenue / Fixed assets	Evaluates the efficiency of the utilization of an enterprise's fixed assets, which measures the output of fixed assets held by the enterprise for production and sale.

revenue and net profit, provide investors with an intuitive understanding of the business growth of an enterprise. These growth rates can be used not only to analyze an enterprise's historical performance, but also as a basis for predicting future performance.

The results of academic research indicate that the growth of some financial indicators has a better ability to reflect the business situation of an enterprise. Academic studies have shown that the growth rates of financial indicators have significant information content for assessing the business situation and future performance of enterprises. Some key growth rate indicators, such as net profit growth rate and total revenue growth rate, are usually regarded as important indicators of a firm's profitability and market competitiveness. For example, a high net profit growth rate usually implies that a company is highly profitable, while a low net profit growth rate may mean that a company is facing competitive pressure or other operational challenges [11]. The growth rate of total revenue is also an important indicator of profitability and market competitiveness. Similarly, the growth rate of total revenue can reflect changes in firms' sales, and a high growth rate may suggest an increase in market share or product popularity. These findings emphasize the importance of growth rate metrics in business analysis, which investors and analysts can monitor to better understand a company's operating dynamics and future potential. This is not only valuable for investment decision makers, but also helps academics to delve deeper into the relationship between financial indicators and corporate performance, providing an important reference point for financial research. In this paper, 11 core growth rate indicators based on enterprises are selected as follows, as the first type of alternative feature set.

4. Forecast Analysis of Credit Risk for Micro and Small Enterprises

4.1. Selection of Significant Models

4.1.1. Logistic Regression

The logistic regression model, first proposed by Joseph Berkson in 1944, is a widely used statistical model in dichotomous problems. The core function of the model is

Table 4
Selection of core enterprise growth rate indicators

Indicator name	Calculation method	Meaning
Operating profit ($ million) growth rate (%)	(Operating profit for the current period - Operating profit for the previous period)/Operating profit for the previous period × 100%	A measure of the percentage increase in profitability generated by a business's operations.
Total profit ($ million) growth rate (%)	(Total profit for the current period - Total profit for the previous period)/Total profit for the previous period×100%	A measure of the percentage increase in the overall profitability of a business.
Net income (including minority interests) ($ million) growth rate (%)	(Net profit for the current period - Net profit for the previous period) / Net profit for the previous period × 100%	A measure of the percentage increase in the net profitability of a business, including minority interests.
Net cash flows from operating activities ($ million) growth rate (%)	(Net cash flows from operating activities for the current period - Net cash flows from operating activities for the prior period)/Net cash flows from operating activities for the prior period × 100%	A measure of the percentage increase in cash flows from operating activities of a business.
Net cash flows from investing activities ($ million) growth rate (%)	(Net cash flows from investing activities in the current period - Net cash flows from investing activities in the previous period)/Net cash flows from investing activities in the previous period × 100%	A measure of the percentage increase in the cash flows a business receives from investing activities.
Net cash flows from financing activities ($ million) growth rate (%)	(Net cash flows from financing activities in the current period - Net cash flows from financing activities in the previous period)/Net cash flows from financing activities in the previous period×100%	A measure of the percentage increase in cash flows received by a business from financing activities.
Net increase in cash and cash equivalents ($ million) growth rate (%)	(Net increase in cash and cash equivalents for the current period - Net increase in cash and cash equivalents for the previous period)/Net increase in cash and cash equivalents	Measures the percentage increase in the net increase in cash and equivalents of a business.
Total assets ($ million) growth rate (%)	(Total assets for the current period - Total assets for the previous period)/Total assets for the previous period ×100%	Measures the percentage increase in the total asset size of a business.
Total liabilities ($ million) growth rate (%)	(Total current-period liabilities - Total prior-period liabilities)/Total prior-period liabilities × 100%	Measures the percentage increase in the total liabilities of a business.
Equity attributable to shareholders of the parent company (in millions of dollars) growth rate (%)	(Equity attributable to shareholders of the parent company for the current period - Equity attributable to shareholders of the parent company for the previous period)/Equity attributable to shareholders of the parent company for the previous period × 100%	Measures the percentage increase in equity attributable to shareholders of the parent company.
Total shareholders' equity (including minority interests) ($ million) growth rate (%)	(Total shareholders' equity for the current period - Total shareholders' equity for the previous period)/Total shareholders' equity for the previous period × 100%	Measures the percentage increase in total shareholders' equity, including minority interests.

usually represented as the following mathematical formula:

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)}} \quad (1)$$

Logistic regression has been widely used in financial risk construction. Previous studies have successfully used logistic regression models for debt risk prediction of companies. In this paper, we will use the logistic regression model to build a debt risk prediction model based on the financial data of MSMEs. We will collect the financial statement data of the enterprise, including the income statement, balance sheet, and cash flow statement, as the input features of the model. Through the logistic regression model, we will be able to estimate the probability of debt default for MSMEs, which will provide strong support for risk management and decision making.

4.1.2. Neural Networks

Neural network is the core mathematical model of deep learning. Deep learning refers to the machine learning process of obtaining a deep network structure containing multiple layers based on sample data through certain training methods [16]. Neural networks have become a powerful tool in financial risk prediction, especially for debt risk assessment of MSMEs. This model is inspired by the biological nervous system, which consists of multiple interconnected neurons that can learn and adapt to a variety of complex nonlinear relationships. During the model training process for datasets with classification labels, the model parameters are obtained by adjusting the parameters to minimize the loss function. During multilayer feature extraction, the input values are normalized by a nonlinear function, and different nonlinear functions have different model training speeds, of which the typical two nonlinear functions are formulated as:

$$R = \frac{1}{1+e^{y}} \tag{2}$$

$$R = \max(0, y) \tag{3}$$

In financial risk construction, neural networks can be used to learn and capture nonlinear relationships to extract potential risk signals from various financial features. We will use neural network modeling in this paper to train the model with large-scale financial data of micro, small and medium-sized enterprises (MSMEs) to predict debt risk. And compare it with other models.

4.1.3. K-Nearest Neighbor Algorithm

The K-Nearest Neighbor (KNN) algorithm is a non-parametric statistical method used primarily for classification as well as regression [1]. The KNN algorithm is one of the simplest of all machine learning algorithms, this instance based on algorithm is very simple and effective in itself and is an inert learning algorithm. It classifies the sample points into different categories in the sample space based on the ordering of the distance between the sample points and the sample centroid as a classification method.

KNN models have a wide range of applications in various classification learning tasks. In this paper, we will further explore and extend the application of KNN models to more accurately predict the debt risk of MSMEs. We will do this by collecting corporate financial data from a number of MSMEs, including key information such as profit and loss statements and balance sheets, and inputting these data into the KNN model for analysis.

4.1.4. Support Vector Machines

Support vector machines (SVM) is a machine learning method based on statistical learning theory, VC dimension theory and structural risk minimization principle. It shows many unique advantages in solving small samples, nonlinear and high-dimensional pattern recognition problems, and largely overcomes the problems of "dimensional catastrophe" and "over-learning". For example, given a set of training samples$(x_i, y_i), i = 1, 2, \ldots, l, x \in R^n, y \in \{0, 1\}$, and the hyperplane is $(w \cdot x) + b = 0$, in order to correctly categorize all the samples and have the categorization intervals, it is required that it satisfies the following constraints: $y[w \cdot x_i + b] \geq 1, i = 1, 2, \ldots l$. The problem of solving the parameter is solved under the precondition:

$$\min \phi(w) = \frac{1}{2}\|w\|^2 \tag{4}$$

SVM models also have better performance efficacy in classification learning tasks. In this paper, we will further explore the application of SVM models to more accurately predict the debt risk of MSMEs. We will do this by collecting corporate financial data from a number of MSMEs, including key information such as profit and loss statements, balance sheets, etc., and inputting these data into the SVM model for analysis.

4.1.5. Random Forest Prediction

A random forest is a set of decision tree classifiers $\{h(X, \theta_k), k = 1, 2, \ldots, K\}$. The random forest is an integrated classifier composed of a set of decision tree classifiers, where θ_k is a random vector obeying an independent homogeneous distribution, K denotes the number of decision trees in the random forest, and each decision tree classifier decides the optimal classification result by voting under a given independent variable X. Random forest is a classifier that integrates many decision trees together, if the decision tree is regarded as an expert in the classification task, the random forest is many experts working together to classify a certain task.

Specifically, a decision tree is a supervised learning algorithm. It is applied to categories and continuous input (feature) and output (predictor) variables. Tree-based methods divide the feature space into a series of rectangles and then place a simple model (like a constant) for each rectangle. Conceptually, they are simple and effective. As an example of a lending company determining the credit risk of its customers, consider a simple data set of customers of a lending company, where we are given the inquired account balances, credit histories, years of tenure, and prior loan status of all customers. The relevant task is to predict whether a customer's risk level is plausible. A screening process to confirm the final customer risk level by ranking different customer characteristics is the basic process of decision trees.

Random forest prediction models also have better performance efficacy in classification learning tasks. In this paper, we will further explore the application of the random forest model to more accurately predict the debt risk of MSMEs. We will do this by collecting corporate financial data from a number of MSMEs, including key information such as profit and loss statements, balance sheets, etc., and inputting these data into the random forest model for analysis.

4.1.6. Naive Bayes

Bayes classification is a statistical classification method, which is a class of algorithms that utilize knowledge of probabilistic statistics for classification. In many occasions, Naive Bayes (NB) classification algorithm can be comparable with decision tree and neural network classification algorithms, the algorithm can be applied to large databases, and the method is simple, high classification accuracy and speed.

Let there be a sample data set $D = \{d_1, d_2, \ldots, d_n\}$, the set of feature attributes corresponding to the sample set

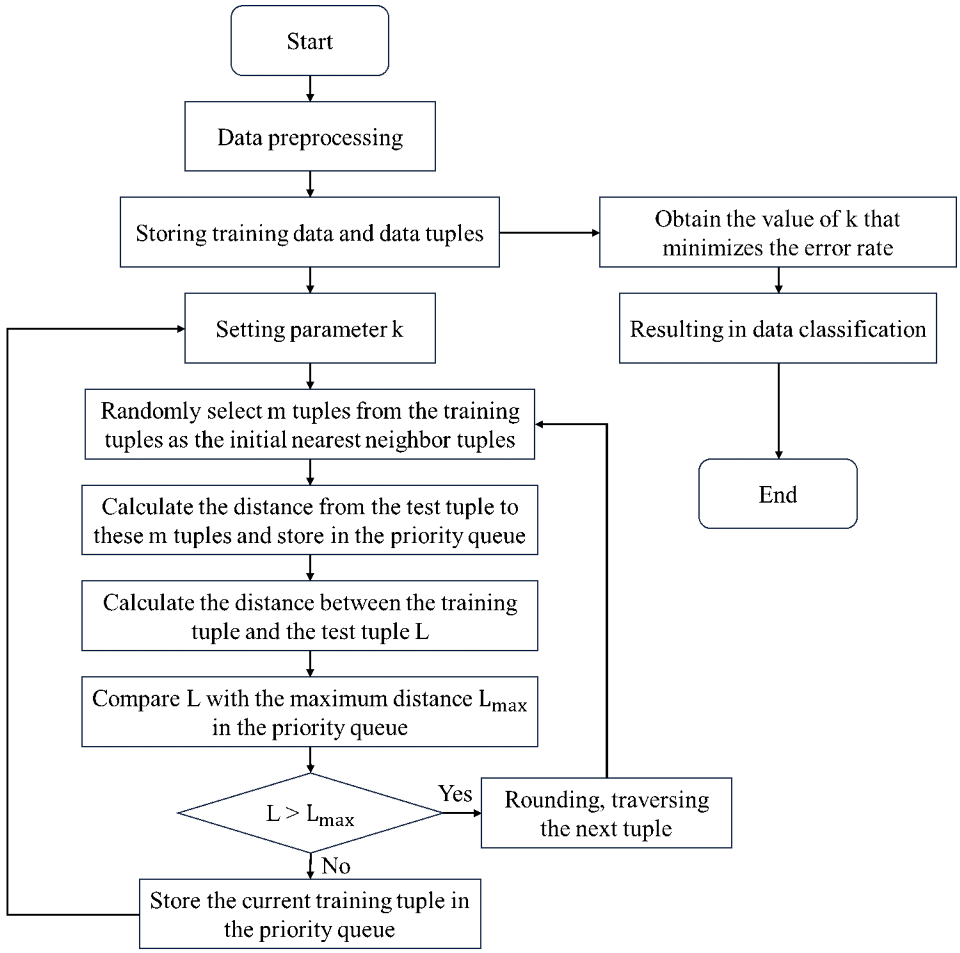

Figure 1: Flowchart of KNN algorithm [41]

is $X = \{x_1, x_2, \dots, x_n\}$, and the class variable is $Y = \{y_1, y_2, \dots, y_m\}$, which means D can be divided into m classes, where $x_1, x_2, \dots, x_n$ are randomly independent of each other, then the prior probability of Y is $P_prior = P(Y)$, the posterior probability of Y is $P_post = P(Y|X)$, then the posterior probability can be obtained by the naive Bayes algorithm, and the posterior probability can be calculated from the prior probability, the evidence $P(X)$, and the class conditional probability $P(X|Y)$ calculated:

$$P(X|Y) = \frac{P(Y)P(X|Y)}{P(X} \tag{5}$$

Naive Bayes is based on the fact that the features are independent of each other, and the above equation can be further expressed as the following equation given a category of y:

$$P(X|Y = y) = \prod_{i=1}^{d} P(x_i|Y = y) \tag{6}$$

The posterior probability can be calculated from the above two equations as:

$$P_{post} = P(Y|X) = \frac{P(y)\prod_{i=1}^{d} P(x_i|Y)}{P(X)} \tag{7}$$

Since the size of $P(X)$ is fixed, it is sufficient to compare only the numerator part of the above equation when comparing posterior probabilities. Thus it can be obtained that a sample data belongs to the category y_i. The naive Bayes formula for this is as follows:

$$P(y_i|x_1, x_2, \dots, x_d) = \frac{P(y_i)\prod_{j=1}^{d} P(x_j|y_i)}{\prod_{j=1}^{d} P(x_j)} \tag{8}$$

Naive Bayes classification also has better performance efficacy in classification learning tasks. In this paper, we will further explore the application of naive Bayes classification to more accurately predict the debt risk of MSMEs. We will do this by collecting corporate financial data from a number of MSMEs, including key information such as profit and loss statements, balance sheets, etc., and inputting these data into the naive Bayes classification for analysis.

4.2. Model Performance Metrics

Receiver Operating Characteristic (ROC) analysis originated as a signal detection method during World War II, and was first applied to process radar images to describe the trade-off between the hit rate and false alarm rate of detected

Table 5
Model AUC performance in the case characterized by the core growth rate indicator

Model	AUC value
Logistic regression	0.511291
Neural network	0.598586
K-nearest neighbor algorithm	0.542536
Support vector machine	0.535550
Random forest prediction	0.649460
Naive Bayes	0.493862

Table 6
Model performance based on the characteristics of key financial indicators

Model	AUC value
Logistic regression	0.648554
Neural network	0.689273
K-nearest neighbor algorithm	0.561692
Support vector machine	0.498382
Random forest prediction	0.667812
Naive Bayes	0.579380

signals. Since then, the method has been widely used in medical diagnosis, and in 1989, Spackman introduced it into the field of machine learning to evaluate classifier performance. The evaluation of ROC curve is the process of analyzing and then comparing ROCs. The most important metric for ROC curve evaluation is the AUC metric, i.e., the area under the curve of ROC. As a single quantitative index of ROC curve, it is a kind of evaluation index on the whole domain of ROC curve. In practice it reflects the classification ability of the classifier very well, so this metric is also most widely used for the comparison of ROC curves. AUC can take a value between 0 and 1, but usually takes a value ranging from 0.5 to 1. The closer the AUC is to 1, the better the performance of the classifier.

In this section, we use logistic regression, k-nearest neighbor model, support vector machine and neural network models for credit risk prediction of sample MSMEs. By changing the threshold for judging whether the model outputs positive or negative examples, a basic knowledge of the performance of the above four models can be obtained.

4.2.1. Model Performance Metrics Characterized by Core Growth Rate Indicators

Naive Bayes obtained an AUC value of 0.493 in this case, which is slightly lower than 0.5. This suggests that naive Bayes has relatively poor classification performance when utilizing the core growth rate features for classification, with an AUC value close to that of random classification. This could mean that the core growth rate data itself is not sufficient to effectively distinguish between positive and negative samples, or that a more sophisticated model is needed to capture the relationships. Random forest demonstrated relatively high performance in this category, with an AUC value of 0.649, significantly higher than logistic regression. This may reflect the fact that random forests are better able to categorize samples with high dimensional data when dealing with core growth rate features. Thus, random forest prediction may be a promising modeling option for credit decision problems that rely on core growth rate data.

The results of these AUC values suggest that financial institutions and policy makers need to remain cautious when utilizing core growth rate data. Logistic regression performs poorly in this case, which may require further data preprocessing, feature engineering, or more sophisticated modeling to improve classification accuracy. In contrast, the relatively stellar performance of random forests provides a potential method for financial institutions to better utilize core growth rate data to improve the accuracy of credit decisions. This finding also implies that core growth rate data may play a more important role in small business credit risk assessment, but requires appropriate modeling methods to fully exploit its information.

4.2.2. Model Performance Metrics Characterized by Important Financial Analysis Indicators

The models characterized by important financial analysis metrics present different levels of performance in terms of AUC values. Compared to the previous section's characteristics, the models trained in this section have better results. The logistic regression model performs poorly when the growth rate indicator is used as a feature value, but in this case a relatively high AUC value of 0.649 is obtained. This indicates that logistic regression is an effective modeling choice for credit risk assessment of small businesses based on financial data, with an AUC value close to 1, which demonstrates that the model exhibits a high level of accuracy in classifying positive and negative samples. On the other hand, the neural network model also shows excellent performance with an AUC value of 0.680, which is slightly higher than the random forest (AUC = 0.667812). This may reflect the potential of neural networks in capturing complex nonlinear relationships in data, especially in the financial domain where data often contain multilevel and multivariate features. This finding suggests that neural network modeling may be a promising option when deeper data mining is required.

The results of this study are important guidance for credit risk prediction of MSMEs. The level of AUC value reflects the accuracy and degree of generalization of different models for the classification of positive and negative samples. Therefore, financial institutions can choose appropriate models to improve the accuracy of credit decision-making according to their needs and data contexts. In addition, the results of this study imply that financial institutions should pay attention to data quality and feature selection to ensure that the data used for model training is more informative and helps to improve the AUC value. In conclusion, the results of AUC values provide the financial industry with powerful tools that are expected to improve credit decision making and

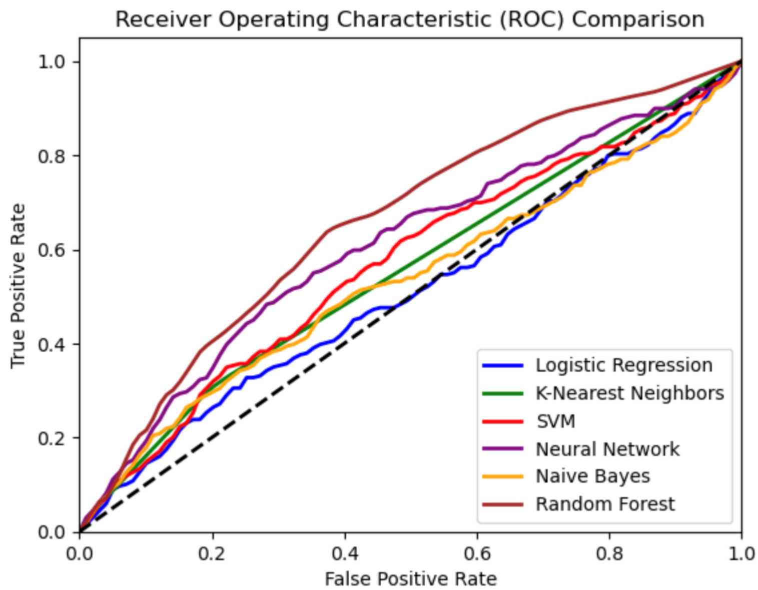

Figure 2: Comparison of model ROC curves for cases characterized by core growth rate indicators

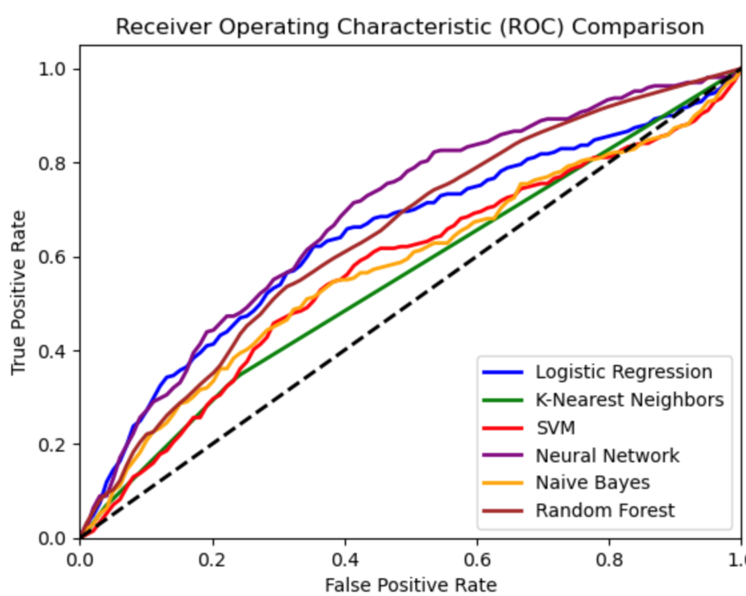

Figure 3: Comparison of model ROC curves in the case characterized by important financial indicators

reduce potential risks, while promoting the financing needs of small businesses.

4.3. Conclusion of the Section

The comparative effectiveness of the two types of features under the same model reveals an important phenomenon in small business credit risk assessment. First, by analyzing the models trained with important property analysis metrics as features, we observe that the neural network model and the random forest perform well in this scenario, with AUC values of 0.689 and 0.667, respectively, indicating the better performance of these models for credit risk assessment based on financial data. The logistic regression model has a large performance gap under different feature sets, while the neural networks all exhibit higher AUC values, showing their advantages in capturing complex nonlinear relationships.

At the same time, models based on core growth rate indicators performed poorly in most cases, with significantly lower AUC values than models characterized by significant property analysis indicators. This may suggest that for small business credit risk assessment, significant financial indicators provide more reliable and sensitive information that helps the models capture potential risks more accurately.

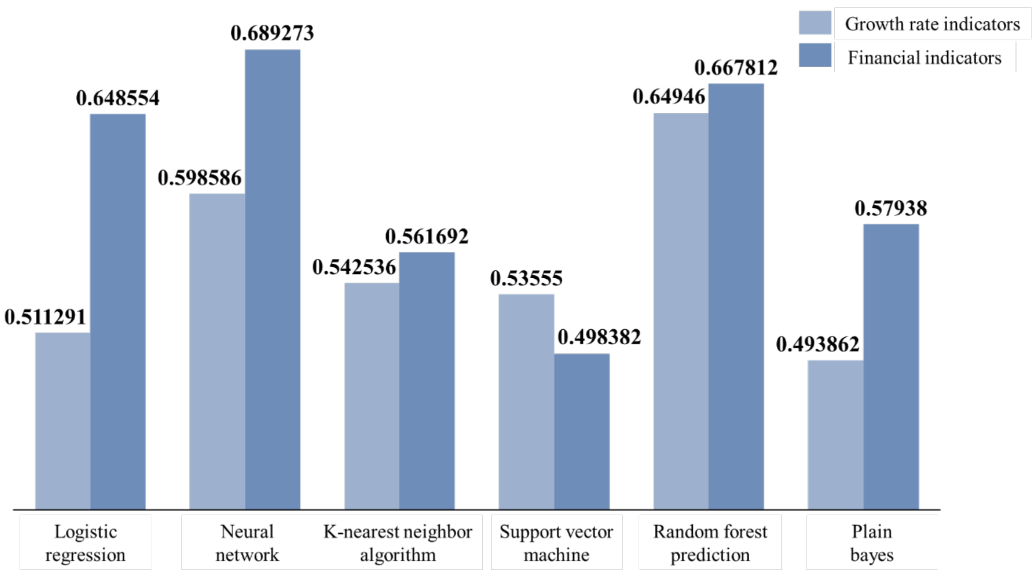

Figure 4: Comparison of model AUC values in the case of two types of indicator characteristics

When we consider the comparative effectiveness of the same type of features under different models, logistic regression shows relatively good performance in both feature cases, while neural networks perform best when characterized by important property analysis metrics. This highlights the adaptability of different models to different types of data and the importance of model selection in a given context.

From a practical application point of view, these results provide useful insights for small business credit risk assessment. When deciding which models and features to adopt, financial institutions should weigh the performance differences between the two. Based on the actual situation and available data, financial institutions can choose the type of model and features that suit their needs to improve the accuracy of credit risk assessment. In addition, this emphasizes the importance of continuing research and development of model selection and feature engineering methods in different contexts to better serve the needs of financial operations.

Taken together, these comparative results highlight the value of data science in small business credit risk assessment, providing financial institutions with additional tools and information to optimize their credit decision-making process and mitigate potential risks while meeting the financing needs of their customers.

5. Policy Analysis and Recommendations

The report of the 19th National Congress of the Communist Party of China explicitly highlights the importance of supporting small and micro enterprises (SMEs), particularly through inclusive finance, to enhance their operational and financial capabilities and foster a sustainable business environment. To further advance the high-quality development of inclusive finance, the State Council released the Implementation Opinions on Promoting the High-Quality Development of Inclusive Finance in September 2023. This document emphasizes the adoption of new development principles, improvement of the inclusive financial system, and enhancement of financial services to support the real economy. It aims to promote broader and more accessible financial services to enable the sustainable development of SMEs. According to the National Financial Regulatory Administration, key measures outlined in the Implementation Opinions include expanding and enhancing financial services for SMEs, increasing support for key sectors, and improving financial institutions' ability to serve SMEs.

SMEs are a driving force behind economic development, yet they face significant challenges in their survival and growth. On one hand, SMEs are inherently vulnerable and bear high trial-and-error costs, making it difficult for them to withstand sudden disruptions. During the pandemic, many SMEs faced operational difficulties, struggled to reopen, and even had to suspend production. On the other hand, innovation is crucial for the high-quality development of SMEs [3]. However, SMEs often struggle to attract top talent, face obstacles in accessing innovation resources, and have difficulty measuring innovation returns, which makes it challenging to achieve stable, long-term innovation outcomes. Furthermore, their limited personnel often face constraints in time and capacity [31]. The core issue hindering SME development is the difficulty and high cost of financing, which severely restricts their ability to govern effectively and innovate. While the Implementation Opinions boost morale by promoting digital inclusive finance and optimizing the social support system for SMEs, several issues remain and require further attention. This paper recommends that the National Financial Regulatory Administration and other relevant departments take targeted measures to address these challenges and ensure the continued high-quality development of SMEs in China.

The challenges facing SME development under the framework of inclusive finance can be categorized as follows: (1) Insufficient development of digital inclusive finance: First, the aggregation of enterprise information is often delayed and incomplete, with a lack of comprehensive and highly visual information platforms. Additionally, relevant institutions are not sufficiently timely in monitoring and mitigating risks related to SME development. Although the government and related departments have introduced policies to promote technological empowerment in inclusive finance, there is a lack of strong measures to encourage the practical application of these products. Second, the coordination and sharing of digital resources remain limited, and information asymmetry persists. Barriers exist between the government, financial institutions, upstream and downstream markets, and consumers in terms of accessing and sharing information about SME operations and development. This results in delays in financial support and market recognition of SMEs' competitive advantages, reducing consumer confidence in SME products and services. Moreover, some SMEs struggle to adapt their development strategies to these challenges [33]. (2) Limited financing channels and high financing costs: Currently, SMEs primarily rely on bank loans and other indirect financing channels for financial support. However, due to their unique capital structures and development characteristics, SMEs face technical challenges and sustainability issues in securing financing. At the same time, direct financing systems, such as equity financing, are underdeveloped. For example, the equity financing market is relatively lagging, making it difficult for SMEs to raise funds through the issuance of stocks or bonds. (3) Lack of targeted and practical policy implementation: While relevant departments emphasize the expansion and improvement of financial services for SMEs, the specific "quantity", "scope", and "quality" of these services remain undefined. Moreover, current financial services and products do not adequately meet the diverse needs of SMEs at different stages of development. Additionally, there are insufficient measures in place to govern the privacy of enterprise and customer information [18].

Based on the exploration and practical analysis of credit risk prediction models in this study, the author proposes the establishment of an inclusive financial service system that deepens fiscal-financial-technology collaboration ("fiscal + financial + technology"). Information integration and sharing are critical to overcoming the final obstacles of inclusive finance. Predictive models, supported by emerging AI technologies, can comprehensively reveal the state of SME development and support regulatory efforts from multiple stakeholders. The specific recommendations are as follows: (1) Government-led fiscal guidance and digital governance: The government should strengthen fiscal guidance and enhance digital governance to promote SME growth in profitability and innovation, thereby improving their overall development quality [36]. This includes providing connection and bridging services across the entire industrial chain to meet the sales and service needs of SMEs [5], such as offering government orders and supporting export businesses. Additionally, the government should increase efforts to attract top talent and foster innovation capacity within SMEs [37]. Talent is critical for innovation, and the government should promote SMEs by enhancing recruitment efforts, offering economic incentives, and improving the wages and social status of technical personnel. A mechanism for regularly evaluating the results of talent training within enterprises should also be established to ensure the continuous improvement of their innovation capabilities [21]. (2) Providing diversified financial products and financing channels: SMEs often suffer from concentrated decision-making power and weak management capacity, and their production is typically focused on emerging industries and simple manufacturing. They are particularly vulnerable to external factors such as cooling consumer demand and rising raw material prices, making the rapid provision of high-quality financial support essential. To address these issues, a new generation of inclusive financial support systems should provide SMEs with diversified information to simplify financing procedures, reduce costs, and improve the efficiency of financial services. The government should also guide and mobilize private capital to support SME development and further develop a government-backed financing guarantee system. Additionally, efforts should be made to strengthen the innovation and development of financial credit products to address the current inadequacies in financial services for SMEs.

6. Conclusion

The high-quality development of small and micro enterprises (SMEs) within the framework of inclusive finance remains a critical challenge in today's economic landscape. This paper has identified key challenges in SME development, constructed a predictive model, and successfully established fundamental principles for forecasting SME credit risk. Additionally, it provides policy analysis grounded in the characteristics of SME growth and the practical results of the model, with recommendations for enhancing policy support and implementing financial products. Based on data analysis, the paper also offers insights into potential SME development trends.

It is evident that inclusive finance will continue to play a pivotal role for an extended period. By examining the core factors influencing SME growth, this study broadens the practical framework for data-driven enterprise management. The findings offer significant value for accurately identifying key growth indicators in SMEs and effectively supporting their development, contributing to smarter management practices. However, it should be noted that the data utilized in this study was sourced from third-party databases, which limited the variety of data available. Furthermore, the proposed predictive model and policy recommendations have yet to be tested in practice.

Despite these limitations, this research presents actionable suggestions and methodologies based on real enterprise data, building upon established SME development practices. Future studies will focus on expanding the data types used, further refining the model, and continuing collaborations with other researchers. The aim is to extend the model's applicability and develop more comprehensive methodologies for fostering SME growth using predictive models [40].

Acknowledgement

The authors declare that there is no funding and no conflict of interest.

References

[1] Altman, N.S., 1992. An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician 46, 175–185.
[2] Brancati, E., 2015. Innovation financing and the role of relationship lending for SMEs. Small Business Economics 44, 449–473.
[3] Cao, Y., You, J., Shi, Y., Hu, W., 2022. Studies on improving the performance of small and micro enterprises through green innovation. Problemy Ekorozwoju 17, 151–161.
[4] Chen, T., Wang, S., 2023. Incomplete information model of credit default of micro and small enterprises. International Journal of Finance and Economics 28, 2956–2974.
[5] Chen, Z., Haider, M.Z., Luo, N., Yu, S., Yuan, X., Zhang, Y., Noreen, T., 2023. Reputation-based partition scheme for IoT security. Security and Privacy 6, e287.
[6] Cheng, C., Yang, L., 2022. What drives the credit constraints faced by Chinese small and micro enterprises? Economic Modelling 113, 105898.
[7] Feng, R., Shen, C., Guo, Y., 2024. Digital finance and labor demand of manufacturing enterprises: Theoretical mechanism and heterogeneity analysis. International Review of Economics and Finance 89, 17–32.
[8] Freiesleben, T., König, G., Molnar, C., Tejero-Cantero, Á., 2024. Scientific inference with interpretable machine learning: Analyzing models to learn about real-world phenomena. Minds and Machines 34, 32.
[9] Gu, Z., Lv, J., Wu, B., Hu, Z., Yu, X., 2024. Credit risk assessment of small and micro enterprise based on machine learning. Heliyon 10, e27096.
[10] He, G., Shen, L., 2021. Whether digital financial inclusion can improve capital misallocation or not: A study based on the moderating effect of economic policy uncertainty. Discrete Dynamics in Nature and Society 2021, 4912836.
[11] Ishikawa, A., Fujimoto, S., Mizuno, T., 2011. Shape of growth-rate distribution determines the type of non-gibrat's property. Physica A: Statistical Mechanics and Its Applications 390, 4273–4285.
[12] Jiang, M., 2017. Analysis of factors influencing financing behaviors of China's small and micro enterprises-taking small and micro enterprises in guangdong province as an example. Agro Food Industry Hi-Tech 28, 3089–3092.
[13] Jin, Y., Zhang, S., 2019. Credit rationing in small and micro enterprises: A theoretical analysis. Sustainability 11, 1330.
[14] Jumadinova, J., Dasgupta, P., 2015. Automated pricing in a multiagent prediction market using a partially observable stochastic game. ACM Transactions on Intelligent Systems and Technology 6, 1–25.
[15] Kanno, M., 2020. Credit risk assessment in real estate investment trusts: A perspective on blockholding and lending networks. International Review of Financial Analysis 71, 101556.
[16] LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
[17] Li, J., Zhang, Q., Yan, F., Zhong, M., 2017. A cloud model-based multi-level fuzzy comprehensive evaluation approach for financing credit of scientific and technological small-medium enterprises. Journal of Difference Equations and Applications 23, 443–456.
[18] Li, Z., Fan, Z., Liu, J., Wang, L., Meng, X., 2022. Large-scale App privacy governance. Frontiers of Engineering Management 9, 640–652.
[19] Liang, X., Hu, X., Jiang, J., 2020. Research on the effects of information description on crowdfunding success within a sustainable economy–the perspective of information communication. Sustainability 12, 650.
[20] Liu, T., Yang, L., 2024. Financial risk early warning model for listed companies using BP neural network and rough set theory. IEEE Access 12, 27456–27464.
[21] Liu, Y., Zhong, W., 2018. Education and training management model and incentive mechanism based on core employees of scientific and technological small and micro enterprises. Educational Sciences: Theory and Practice 18, 2699–2706.
[22] Lu, J., Chen, X., 2022. Risk model of financial supply chain of Internet of Things enterprises: A research based on convolutional neural network. Computer Communications 183, 96–106.
[23] Müller, K., Verner, E., 2023. Credit allocation and macroeconomic fluctuations. The Review of Economic Studies 91, 3645–3676.
[24] Rubin, G.M., Overstreet, G.A., Beling, P., Rajaratnam, K., 2013. A dynamic theory of the credit union. Annals of Operations Research 205, 29–53.
[25] Shi, Z., 2017. Research on the financing problems of small and micro enterprises under the background of Internet finance. Agro Food Industry Hi-Tech 28, 2435–2439.
[26] Stevenson, M., Mues, C., Bravo, C., 2021. The value of text for small business default prediction: A deep learning approach. European Journal of Operational Research 295, 758–771.
[27] Sun, H., Liu, X., Du, Z., Wu, Y., Zhang, H., Xu, X., 2022. Heterogeneous influence of individuals' behavior on mask efficacy in gathering environments. Frontiers of Engineering Management 9, 550–562.
[28] Wang, W., Yu, S., Bekele, T.M., Kong, X., Xia, F., 2017. Scientific collaboration patterns vary with scholars' academic ages. Scientometrics 112, 329–343.
[29] Xu, B., Costa-Climent, R., Wang, Y., Xiao, Y., 2020. Financial support for micro and small enterprises: Economic benefit or social responsibility? Journal of Business Research 115, 266–271.
[30] Xu, N., Yuan, Y., Rong, Z., 2022a. Depressed access to formal finance and the use of credit card debt in Chinese SMEs. China Economic Review 72, 101758.
[31] Xu, R., Zhang, T., Zhang, Q., 2022b. Investigating the effect of online and offline reputation on the provision of online counseling services: A case study of the Internet hospitals in China. Frontiers of Engineering Management 9, 563–576.
[32] Yan, B., 2017. Financing strategy of small and micro enterprises in Xi'an on the background of "The belt and road". Agro Food Industry Hi-Tech 28, 1665–1668.
[33] Yan, H.B., Li, Z., 2022. Review of sentiment analysis: An emotional product development view. Frontiers of Engineering Management 9, 592–609.
[34] Yang, L., Zhang, Y., 2020. Digital financial inclusion and sustainable growth of small and micro enterprises–evidence based on China's new third board market listed companies. Sustainability 12, 3733.
[35] Ye, B., Yuan, J., Guan, Y., 2022. Internet finance, financing of small and micro enterprises and the macroeconomy. Emerging Markets Finance and Trade 58, 2851–2866.
[36] Ye, X., Su, X., Yao, Z., Dong, L.a., Lin, Q., Yu, S., 2023. How do citizens view digital government services? Study on digital government service quality based on citizen feedback. Mathematics 11, 3122.
[37] Yu, S., Alqahtani, F., Tolba, A., Lee, I., Jia, T., Xia, F., 2022a. Collaborative team recognition: A core plus extension structure. Journal of Informetrics 16, 101346.
[38] Yu, S., He, S., Cai, Z., Lee, I., Naseriparsa, M., Xia, F., 2022b. Exploring public sentiment during COVID-19: A cross country analysis. IEEE Transactions on Computational Social Systems 10, 1083–1094.
[39] Yu, S., Qing, Q., Zhang, C., Shehzad, A., Oatley, G., Xia, F., 2021a. Data-driven decision-making in COVID-19 response: A survey. IEEE Transactions on Computational Social Systems 8, 1016–1029.

[40] Yu, S., Xia, F., Zhang, C., Wei, H., Keogh, K., Chen, H., 2021b. Familiarity-based collaborative team recognition in academic social networks. IEEE Transactions on Computational Social Systems 9, 1432–1445.
[41] Zhang, M.L., Zhou, Z.H., 2007. ML-KNN: A lazy learning approach to multi-label learning. Pattern Recognition 40, 2038–2048.
[42] Zhang, W., Wang, Z., 2023. Can SMEs benefit equally from supportive policies in China? PLoS ONE 18, e0280253.