

# Research Article Forecasting Credit Risk of SMEs in Supply Chain Finance Using Bayesian Optimization and XGBoost

## Chen Zhang 问 and Xinmiao Zhou 问

Business School, Ningbo University, Ningbo 315211, China

Correspondence should be addressed to Xinmiao Zhou; zhouxinmiao@nbu.edu.cn

Received 13 September 2022; Revised 14 June 2023; Accepted 22 November 2023; Published 11 December 2023

Academic Editor: Juan Frausto-Solis

Copyright © 2023 Chen Zhang and Xinmiao Zhou. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supply chain finance plays a crucial role as a financing channel for small- and medium-sized enterprises (SMEs). However, issues such as financial problems and credit defaults have led to disruptions in this channel. To address credit risk control in SME financing within the field of supply chain finance, this paper focuses on a sample of 506 equipment manufacturing companies listed on the SME board of the Shenzhen Stock Exchange from 2016 to 2020. Taking into consideration, the overall risks faced by these enterprises, the study establishes seven first-level indicators and identifies 84 candidate second-level indicators. Partial correlation and variance analysis are then used for the first round of indicator screening, followed by the use of a BP neural network for the second round of selection. As a result, a system of 26 indicators for supply chain financial risk is constructed. The XGBoost model is employed to evaluate the constructed risk index system, while SVM and random forest models are used as comparison models. Bayesian optimization is utilized for parameter tuning of the three models. Empirical results demonstrate that the BO-XGBoost model reduces prediction errors in comparison to the control models. Furthermore, statistical tests reveal that the predicted values of the BO-XGBoost model significantly differ from those of the other control models. Compared to other individual models, the BO-XGBoost model exhibits increased accuracy in credit risk prediction and a significant discriminative effect. These findings highlight the effectiveness of constructing an efficient risk indicator system and utilizing Bayesian optimization for parameter tuning in XGBoost to better differentiate between risky and normal enterprises, thereby minimizing default losses. The research results underscore the advantages of employing Bayesian optimization in XGBoost, which can be applied in credit default prediction for SMEs and serves as a valuable tool in financial risk management and control.

## 1. Introduction

Small- and medium-sized enterprises (SMEs) are recognized as the key drivers of economic growth [1]. According to statistics from the Asia-Pacific Economic Cooperation, SMEs in the Asia-Pacific region, including China, play a crucial role in promoting national technological innovation and employment. SMEs contribute over 50% to GDP and over 60% to employment, making them significant contributors to the economy, especially in China.

In the current economic and financial landscape, smalland micro-enterprises, as well as private enterprises, face challenges in obtaining financing due to the difficulty of fund providers to assess the risk level of loans without credit ratings. This reduces the willingness of fund providers to lend and results in higher lending rates to cover the risk margin premium due to incomplete and asymmetric information [2]. Additionally, from the perspective of domestic enterprise loans based on enterprise scale and guarantee, smaller enterprises tend to have a higher proportion of mortgage and pledged loans issued by banks.

Despite many SMEs having sound production and operational processes and long-term cooperative relationships with larger enterprises, they often lack recognized mortgaged assets and formal financial statement data, leading to reduced willingness of banks to lend to SMEs. Faced with these challenges, supply chain finance (SCF) has emerged as a potential solution. The schematic diagram of SCF is shown in Figure 1.

SCF is centered around the core enterprises in the supply chain, leveraging the strong creditworthiness of these core



FIGURE 1: Supply chain finance business.

enterprises to facilitate funding for upstream and downstream enterprises. This allows for credit to be extended based on the financial status of these upstream and downstream enterprises, verifying the authenticity and stability of their business transactions with the core enterprises. Through credit granting, upstream and downstream enterprises in the supply chain can access low-cost bank credit, ultimately leading to increased income generation.

However, traditional SCF is vulnerable to credit risk due to incomplete information and moral hazard arising from asymmetric information [3]. Banks may be hesitant to engage in relevant business with upstream and downstream SMEs or may require excessively high loan interest rates as a safety margin. Additionally, due to the high value, fast frequency, and wide geographical scope of the supply chain's upstream and downstream trade chain, obtaining real, effective, timely, and costeffective information through traditional information collection and processing methods is challenging.

As a result, controlling the credit risk of SMEs is of utmost importance. Credit evaluation has become a crucial tool in financial risk management. Establishing an evaluation index framework and utilizing forecasting models based on historical data of financing enterprises can effectively assess and control risks associated with credit in SCF.

#### 2. Related Research

2.1. Connotation of SCF. The literature on the definition and connotation of SCF can be categorized into two main perspectives: financial orientation and supply chain orientation [4–6]. Different scholars have varying definitions of SCF. For instance, Wuttke et al. [7] defined SCF as a series of short-term financial solutions, including reverse factoring, which target financial flows and aim to improve working capital and reduce costs for purchasing companies and their suppliers. Pfohl and Gomm [8] expanded the scope of SCF by including inventory, accounts payable, and accounts receivable, and incorporate working capital into the overall framework of SCF [7]. Furthermore, some scholars further broaden the scope of SCF beyond the management of working capital, encompassing the capital flow and information flow of the entire supply chain [8, 9].

2.2. The Development Model of SCF. SCF involves the integration of capital flow and supply chain management to provide commercial trade service products and financing services for enterprises at each transaction link. It aims to optimize the availability and cost of funds within the supply chain, with the buyer as the core. SCF typically starts with the core enterprise in the supply chain, providing funds to upstream suppliers facing financing difficulties, mitigating corruption risks in the supply chain, or extending bank credit to downstream distributors to enhance the purchasing power of the core enterprise's commercial credit.

SCF is often positioned as the core of the accounts receivable offtake business [10]. Upstream and downstream transactions, such as orders, inspections, invoices, and payment notices, are incorporated into various financing methods, including order financing, postinspection financing, accounts receivable financing, and appointment payment financing, which are integrated into the SCF process. Banks are linked with the core enterprise and upstream and downstream enterprises, providing flexible financial products and services through this financing model. The supply chain of a specific commodity typically involves the purchase of raw materials, the manufacture of intermediate and final products, and the delivery of products to consumers through sales channels, linking suppliers, manufacturers, distributors, retailers, and end users as a whole. With the advancement of fintech and the optimization of banks' strategies in the field of SCF, an online and intelligent innovative business model has emerged -an online SCF integrated platform, as shown in Figure 2. This platform utilizes big data, artificial intelligence, Internet of Things, and other technologies to provide comprehensive financial services, such as credit granting, guarantee, settlement, and wealth management.

2.3. Risk Measurement of SCF. The integration of supply chain logistics, cash flow, and information flow in the financial supply chain has the potential to replace the traditional evaluation of SMEs based on ontology. However, the reluctance of core manufacturers to fully disclose economic activity information to upstream and downstream manufacturers makes it difficult to verify the authenticity of supply chain trade. This lack of transparency across the enterprise in the supply chain and across banks results in increased risks and moral hazards that are challenging to track operationally. As a result, banks often rely on the value of inventories and receivables for risk control.

To address these challenges, researchers have conducted extensive research on the entire supply chain market and proposed various mathematical models aimed at reducing supply chain financial risks associated with raw materials, product markets, and production inventories. However, as



FIGURE 2: Online supply chain finance integrated platform.

the supply chain financial service model continues to expand, corresponding risks are also increasing, particularly in light of the continuous downward trend of the global economy. This has led to a significant increase in financial risks within the supply chain, which has become exponentially challenging to manage.

The advancement of computing science and machine learning methods has led to a significant increase in the application of various machine learning techniques in enterprise credit risk assessment [11]. For instance, Jiang et al. [12] utilized support vector machine (SVM) in credit risk assessment, Rtayli and Enneya [13] employed random forest classifier and SVM to detect fraud risk, Chang et al. [14] utilized XGBoost to evaluate credit risk of financial institutions, and Guo [15] developed a loan risk assessment algorithm based on backpropagation (BP) neural network. Teles et al. [16] compared Bayesian networks with artificial neural networks for predicting recovery value in credit operations. Shen et al. [17] proposed a new integrated model that combines classifier optimization technology for personal credit risk assessment. Furthermore, Lappas and Yannacopoulos [18] incorporated empirical knowledge from experts, and Du et al. [19] utilized a two-stage mixed model to measure financial credit risk. These diverse approaches highlight the growing use of machine learning in credit risk assessment in recent years.

One challenge in credit scoring is the potential presence of nonlinear effects in the data. To address this issue, Dumitrescu et al. [20] developed a penalty logic tree regression (PLTR) method to study and account for possible nonlinear effects in credit score data. This approach incorporates information from decision trees to enhance the performance of logistic regression (LR). Similarly, Abedin et al. [21] simultaneously applied SVMs and a P-neural network based CDP algorithm to predict customer bankruptcy in financial risk management.

However, one problem in credit risk assessment is the issue of imbalanced data sets, where the number of outstanding loans is typically lower than the number of paid loans. This poses challenges for risk assessment models. To address this, Khemakhem and Boujelbene [22] utilized a SVM model with multicores, while Hou et al. [23] employed a multiclass imbalance weighting mechanism to mitigate the impact of imbalanced constraints on financial data categories. Zhang et al. [24] used a firefly algorithm optimized SVM, while Luo et al. [25] adopted a kernelless quadric SVM model to explore hyperparameter optimization in risk management.

Furthermore, the small sample size of financial risk data can result in poor training model performance, as it may not achieve the desired accuracy or effectively distinguish risks. Additionally, data access can be restricted due to data protection regulations. In response to these challenges, Li et al. [26] and Liu et al. [27] applied generative adversarial network to construct a credit risk model for small and microenterprises, proposing a method based on the XGBoost model. However, different quality data samples in enterprise financial data can cause errors and other issues during modeling. Levy and Baha [28] compared the classification performance of LR and linear discriminant analysis models on financial data samples from SMEs.

Credit risk management also poses challenges in the context of SCF. The use of big data technology enables effective credit warning and prevention in the realm of Internet finance. Qi et al. [29] used weighted K-nearest neighbors to explore risk factors affecting the development of Internet finance. Liu and Huang [30] integrated a SVM model to assess risk in SCF and combined it with noise reduction methods to enhance credit assessment accuracy. Building upon concepts from supply chain management budget models, Zhao and Li [31] employed a BP neural network model to discuss the main factors influencing the financial impact of SMEs and the benefits of supply chain budgeting in addressing expenditure issues for SMEs. Additionally, Wu et al. [32] used neural networks, while Hosseini and Khaled [33] applied a hybrid ensemble and analytic hierarchy process approach to conduct risk assessment on SCF and credit enterprise supplier selection.

In current literature, however, credit evaluation research is mostly focused on personal credit or corporate credit without considering SCF. There are limited studies on SMEs' financing under the supply chain financial model, and the research on supply chain financial risk control is often simplistic, with credit risk system indicators that may not be well-founded. Therefore, there is a need to establish a more comprehensive credit risk control index and framework to accurately measure credit risk in SCF. Additionally, the methods used in literature for model construction often lack reasonable hyperparameter adjustment and optimization, leading to inaccurate credit risk predictions. Improved methods for hyperparameter tuning are necessary to obtain more precise results.

The work of this paper is as follows: (1) First, screening supply chain risk indicators through partial correlation coefficient and variance analysis. (2) Then, evaluating the impact of individual indicators on prediction accuracy using BP neural network, gradually removing redundant indicators that would reduce accuracy, and constructing a financial credit risk indicator system for supply chain. (3) Applying XGBoost and Bayesian optimization (BO) to evaluate the constructed financial credit risk indicator system for supply chain and draw corresponding conclusions.

#### 3. Introduction to Theoretical Basis

3.1. BP Neural Network. BP neural network is used in the screening of indicators. This paper adopts a three-layer BP neural network model, which consists of an input layer, a hidden layer, and an output layer. The input layer is responsible for receiving input information from the candidate indicators and passing it to the neurons in the hidden layer. The hidden layer serves as an internal information processing layer and is responsible for information transformation. The hidden layer connects to the output layer, and after further processing, it completes the forward propagation process of a learning iteration. The nodes in the hidden layer utilize the sigmoid function as activation function, which is defined as  $f(x) = 1/(1 + e^{-x})$ . The input and output nodes can be activated using the sigmoid function as well. The output value for the hidden layer nodes is given as follows:

$$Q_j = f\left(\sum W_{ij} \cdot x_i + b_j\right),\tag{1}$$

and the output value of the network for the output layer nodes is given as follows:

$$Y_k = f(W_{jk} \cdot Q_j + b_k).$$
<sup>(2)</sup>

In Equations (1) and (2), f represents the sigmoid function, b represents the bias of the neural unit, and  $W_{ij}$  represents the weight of the influence of the input node  $x_i$  on the hidden layer node  $x_j$ .

*3.2. XGBoost.* XGBoost is a powerful boosting-based machine learning algorithm that can handle both classification and regression problems. It works by iteratively training multiple weak learners to achieve a strong learner [34]. The training flowchart of XGBoost is illustrated in Figure 3.

XGBoost implements an additive model based on a collection of classification and regression trees (CART). Each time a new decision tree is added as a base learner, it is trained to fit the residual of the previous prediction. The final prediction of the model is obtained by summing up the predictions of all the decision trees. In XGBoost, the objective function consists of a loss function and a regularization term, which is an improvement over the gradient boosting decision tree algorithm. The loss function is enhanced using secondorder Taylor expansion to incorporate first and second derivatives, resulting in effective control of model overfitting and improved prediction accuracy. The addition model of CART trees in XGBoost can be expressed as follows:



FIGURE 3: XGBoost training flowchart.

$$\widehat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F,$$
(3)

where  $\hat{y}_i$  represents the predicted value of the *i*th sample, *K* represents the number of trees,  $x_i$  represents the *i*th sample data of the input,  $f_k(x_i)$  represents the *k*th decision tree, and  $f_k$  is a function in the tree set space *F*.

The XGBoost regression model is optimized to the objective function during the training process, the original model remains unchanged in each iteration, a new function is added to the *t*th iteration model, and the sum of the newly generated decision tree and all previous trees is predicted. The residuals are fitted, and the iterative process is expressed as follows:

$$\begin{split} \widehat{y}_{i}^{(0)} &= 0\\ \widehat{y}_{1}^{(1)} &= f_{1}(x_{i}) = \widehat{y}_{i}^{(0)} + f_{1}(x_{i})\\ \widehat{y}_{2}^{(2)} &= f_{1}(x_{i}) + f_{2}(x_{i}) = \widehat{y}_{i} + f_{2}(x_{i}) \\ \cdots \\ \widehat{y}_{i}^{(t)} &= \sum_{K=1} f_{k}(x_{i}) = \widehat{y}_{i}^{(t-1)} + f_{t}(x_{i}) \end{split}$$
(4)

The objective function expression of XGBoost is as follows:

$$Obj = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \sum_{K=1}^{K} \Omega(f_k) = \sum_{i=1}^{N} l \Big\{ y_i, \hat{y}_i^{(t-1)} + f_i(x_i) \Big\} + \sum_{K=1}^{K} \Omega(f_k).$$
(5)

Among them, the first part is the loss function, which can describe the error between the predicted value and the real value, and the second part is the regular term, which can effectively control the complexity of the model building tree structure and prevent overfitting. The regularization term is expressed as follows:

Predicted category Negatives (0) Positives (1) True False Negatives (0) negatives positives Actual category (TN) (FP) False True Positives (1) positives negatives (FN) (TP)

FIGURE 4: Confusion matrix.

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2, \qquad (6)$$

where  $\gamma$  and  $\lambda$  represent the weighting factor, *T* is the number of leaf nodes, and  $\omega$  represents the weight of leaf nodes.

XGBoost used the Taylor expansion method to optimize the objective function. Finally, the optimal objective function of XGBoost is given as follows:

$$Obj = -\frac{1}{2} \sum_{i=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma T.$$
(7)

*3.3. Bayesian Optimization.* BO is a global optimization algorithm based on probability distribution [35]. Aiming at the hyperparameter optimization of the XGBoost model, in the decision space of a set of hyperparameters, BO constructs a probability model for the function to be optimized  $f : x \to R^d$ , and further uses this model to select the following an evaluation point, iterative loop in turn to obtain the optimal solution of hyperparameters:

$$x^* = \operatorname{argmin}_{x \in \gamma \subseteq \mathbb{R}^d} f(x), \tag{8}$$

where  $x^*$  is the optimal hyperparameter combination,  $\chi$  is the decision space, and f(x) is the objective function.

The main core steps of the BO algorithm are the priori function and the acquisition function. The prior function uses the Gauss regression process, based on Bayes' theorem, to convert the prior probability model into a posterior probability distribution. The acquisition function uses the probability of improvement (PI) to select the next evaluation point.

A Gaussian process belongs to a nonparametric model and is also a collection of random variables, determined by the mean function and the kernel function, namely:

$$f(x) \sim gp(m(x), k(x, x')), \tag{9}$$

where m(x) = E[f(x)] is the mean function and k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))] is the covariance function.

In the problem of XGBoost hyperparameter optimization, establish a sample data set D = (X, y) of hyperparameters, where  $X = (x_1, x_2, \dots, x_t)$ . For the training set,  $y = \{f(x_1), f(x_2), \dots, f(x_t)\}$  is the set of continuous functions f(x).



FIGURE 5: Distribution of samples.

If a new sample  $x_{t+1}$  is added and the covariance matrix is updated, denoted as *K*, then the joint Gaussian distribution can be expressed as follows:

$$\begin{bmatrix} f_{1t} \\ f_{t+1} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \mathbf{K} & \mathbf{k} \\ \mathbf{k}^T & \mathbf{k}(\mathbf{x}_{t+1}, \mathbf{x}_{t+1}) \end{bmatrix}\right), \tag{10}$$

where  $k = [k(x_{t+1}, x_1) \quad k(x_{t+1}, x_2) \quad \cdots \quad k(x_{t+1}, x_t)]$ , the posterior probability of  $f_{t+1}$  can be further obtained distributed:

$$P(f_{t+1}|D_{1t}, x_{t+1}) \sim N(u, \sigma^2)$$
  

$$u = k^T K^{-1} f_{1t}$$
  

$$\sigma^2 = k(x_{t+1}, x_{t+1}) - k^T K^{-1} k$$
(11)

PI is used as the sampling function, and the expression is given as follows:

$$f_{\rm PI}(x) = \varPhi\left(\frac{u(x) - f(x^+) - \xi}{\sigma(x)}\right),\tag{12}$$

where  $\Phi(\cdot)$  is the CDF of normal distribution, u(x) and  $\sigma(x)$  are the mean and variance of the objective function,

respectively,  $f(x^+)$  is the current optimal objective function value, and  $\xi$  is the parameter.

3.4. Evaluation Metrics. The evaluation of the model needs appropriate evaluation indicators to measure. This paper uses four evaluation indicators: error, accuracy,  $F_1$  score, and receiver operating characteristic (ROC) to evaluate the performance of BO-XGBoost and the other control models (SVM, RF, XGBoost, BO-SVM, and BO-RF). Based on the combination of the real category and the predicted category, the confusion matrix of the classification results can be formed, as shown in Figure 4:

$$SE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y_i})^2, \qquad (13)$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \widehat{y_i}|, \qquad (14)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN},$$
 (15)

$$Precision = \frac{TP}{TP + FP},$$
 (16)



FIGURE 6: Candidate risk index system.

TABLE 1: Partial correlation coefficient and $F$ value of op	perating capacity.
--	--------------------

	1. RTR	2. RDs	3. IT	 12. CCR	13. CAT	14. DSOI	F	Remains
1. RTR	1	0.998	0.073	 -0.399	-0.077	-0.525	0.498	+
2. RD	0.998	1	-0.026	 -0.349	-0.074	-0.528	0.221	-
3. IT	0.073	-0.026	1	 -0.358	0.507	-0.315	NA	+
13. CAT	-0.077	-0.074	0.507	 0.701	1	0.023	0.015	_
14. DSOI	-0.525	-0.528	-0.315	 0.436	0.023	1	NA	+

*Note.* "NA" means not to calculate the *F* value, "+" means keep the index, and "-" means delete the index. Bold values indicate the maximum partial correlation coefficient in each row.

$$\operatorname{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}},\tag{17}$$

$$F_1 = 2 \times \frac{\text{Recall} \times \text{precision}}{\text{Recall} + \text{precision}}.$$
 (18)

## 4. Construction of Credit Risk Indicator System for Supply Chain Finance

4.1. Data Description and Standardization. This study focuses on equipment manufacturing enterprises listed on the SME board of the Shenzhen Stock Exchange from 2016 to 2020,

TABLE 2: Profitability index screening results.

	Accuracy	d (%)	Filter
ТАР		-1.27	_
ROPTCE		-4.73	-
OPR		10.32	+
NPMS	A0 = 88.73%	-8.57	_
RORTA	A1 = 32.7%	-9.22	_
OPR	A = 80.6%	18.25	+
NFA		23.72	+
ROTAR		-13.53	_
GPM		2.44	+

*Note.* "+" stands for retained index, "-" stands for deletion index, "A0" is the correct rate of samples without credit risk, "A1" is the correct rate of samples with credit risk, and "A" is the correct rate of all samples.

using data obtained from the Wind database. To accurately reflect the characteristics of the supply chain, information such as the official website of the enterprise and the company's annual report is utilized to identify the supply chain in which each SME operates. The study includes a total of 505 supply chain samples, consisting of 273 samples from the machinery manufacturing industry, 124 samples from the electronics manufacturing industry, and 109 samples from other manufacturing industries, as illustrated in Figure 5.

In order to better process the data, the indicators of each enterprise are standardized, let  $X_{ij}$  be the *i*th sample of *j*th values and the processing methods are as follows:

$$Y_{ij} = \frac{X_{ij} - \min\{X_{ij}\}}{\max\{X_{ij}\} - \min\{X_{ij}\}}.$$
(19)

We mainly measure whether the enterprise has the risk of default from the seven major first-level indicators of *profitability*, *solvency*, *development capability*, *operating capability*, *cash flow platform service* and *macroenvironment* indicators. Second, through the evaluation results of the information disclosure of the listed company, whether the company has a breach of contract is given qualified and unqualified. The specific candidate risk indicator system is shown in Figure 6.

#### 4.2. First Indicator Screening

4.2.1. Partial Correlation Analysis. Let  $r_{ij}$  be the correlation coefficient between the *i*th index and the *j*th index,  $x_i$  and  $x_j$  be the *i*th of the *j*th sample standardized scores of the *i* and *j* indicators,  $\overline{x_i}$  and  $\overline{x_j}$  are the average of the *i* and *j* indicators, respectively, as given in the following equation:

$$\mathbf{r}_{ij} = \frac{\sum_{i,j=1}^{n} (x_i - \overline{x_i}) \left( x_j - \overline{x_j} \right)}{\sqrt{\sum_{i,j=1}^{n} (x_i - \overline{x_i})^2 \sum_{i,j=1}^{n} \left( x_j - \overline{x_j} \right)^2}}.$$
 (20)

Let *R* be the matrix composed of the correlation coefficient  $r_{ij}$  of the *i* index and the *j* index, and then, we obtain the following equation:

TABLE 3: Number of indicators and correct rate after step-by-step screening.

	All	First trim	Second trim
Removed	0	38	20
Remains	84	46	26
A0	_	76.34%	85.17%
A1	_	68.27%	75.30%
А	-	72.26%	82.83%

*Note.* "A0" is the correct rate of samples without credit risk, "A1" is the correct rate of samples with credit risk, and "A" is the correct rate of all samples.

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}.$$
 (21)

A correlation coefficient of 0.5 is considered a moderate degree of correlation. Therefore, in this paper, the critical value of the partial correlation coefficient is determined to be 0.6. When the two indicators are  $|R_{ih}| \ge 0.6$ , the indicator that reflects relatively little default information and will be deleted.

4.2.2. Calculation of *F* Value. Among the two indicators of  $|R_{ih}| \ge 0.6$ , in order to avoid subjectively deleting the indicators that have a significant impact on the credit risk status, the *F* value is used to delete the indicators with less ability to distinguish credit risks and delete the evaluation index with smaller *F* value.

Let  $F_i$  be the F value of the *i*th indicator,  $\overline{x_i}$  is the normalized mean of the *i*th indicator,  $\overline{x_i}^1$  is the risk-free firm's standardized mean of *i*th indicators,  $\overline{x_i}^0$  is the standardized mean of the *i*th indicator of risky enterprises,  $n_1$  represents the total number of samples without credit risk,  $n_0$  represents the total number of samples with credit risk, and y = 0 and y = 1 represent the enterprise without credit risk and with credit risk, respectively, then  $F_i$  is given as follows:

$$F_{i} = \frac{(\overline{x_{i}}^{1} - \overline{x_{i}})^{2} + (\overline{x_{i}}^{0} - \overline{x_{i}})^{2}}{\frac{1}{n_{1} - 1} \sum_{y=0} (x_{i} - \overline{x_{i}}^{1})^{2} + \frac{1}{n_{0} - 1} \sum_{y=1} (x_{i} - \overline{x_{i}}^{0})^{2}}, \qquad (22)$$

where  $\overline{x_i} = \frac{1}{n} \sum_{i=1}^n x_i$ ,  $\overline{x_i}^1 = \frac{1}{n_1} \sum_{y=0} x_i$ , and  $\overline{x_i}^0 = \frac{1}{n_0} \sum_{y=1} x_i$ . The *F* value reflects the ability of different indicators to

The F value reflects the ability of different indicators to discriminate the default state. The larger the F value, the stronger the ability of the indicator to reflect the default information.

4.2.3. Index Screening Steps. Step 1: Calculate the index partial correlation coefficient of the secondary indicators layer and find the index of  $|R_{ih}| \ge 0.6$ .

*Step 2*: Substitute the risk indicators of  $|R_{ih}| \ge 0.6$  into Equation (21) to calculate the *F* value.

Step 3: According to the partial correlation coefficient between the indicators and the F value, among the two

		The second	
Types of index	Indicator name	Definition or calculation	Variable symbol
Label	ST	Company stock is special treatment	1 or 0
	Net fixed assets	Total fixed asset purchase price + capital improvements) – (accumulated depreciation + fixed asset liabilities)	X1
Profitability	Operating profit ratio	Operating profit/net sales	X2
	Return on equity(ROE)	Net income/shareholders' equity	X3
	Gross profit margin	(Net sales – cost of goods sold)/net sales	X4
	Current ratio(CR)	Current assets/current liabilities	X5
	Quick ratio(QR)	(Current assets – inventories – prepaid expenses)/current liabilities	X6
solvency	Cash ratio	Cash and cash equivalents/current liabilities	X7
	Debt-to-assets ratio	Total debt/total assets	X8
	Main business's increasing rate of income	(Current period revenue – prior period revenue)/prior period revenue	6X
Development capability	Net profit margin ratio	Net profit/revenue	X10
	Net assets growth rate	(Current period net assets – prior period net assets)/prior period net assets	X11
	Cash flow from operating activities	Funds from operations + changes in working capital	X12
	Current assets	Current assets are all the assets of a company that are expected to be sold or used as a result of standard business operations over the next year	X13
Cash flow	Current liabilities	Current liabilities are a company's debts or obligations that are due to be paid to creditors within 1 year.	X14
	Cash flow to sales ratio	Operating cash flow/net sales	X15
	Return on assets(ROA)	Net income/average total assets	X16
	Operating cash flow to net profit ratio	Operating cash flow/net profit	X17
	Operating cash flow ratio	Operating cash flow/current liabilities	X18
Oneroting conscience	Receivables turnover ratio	The receivables turnover ratio measures the efficiency with which a company is able to collect on its receivables or the credit it extends to customers	61X
Optianing capacity	Inventory turnover	Inventory turnover refers to the amount of time that passes from the day an item is purchased by a company until it is sold	X20
	Days sales of inventory	Inventory turnover refers to the amount of time that passes from the day an item is purchased by a company until it is sold	X21
	Perfect order percentage	The perfect order rate measures how many orders you ship without incident, where incidents are damaged goods, inaccurate orders, or late shipments	X22
Platform as a service	Total orders	The number of all orders generated by the platform	X23
	Billing data	Billing data means data relating to the charges for consumption of services which have been paid for or otherwise	X24
	GDP deflator	GDP deflator is a measure of the level of prices of all new, domestically produced, final goods, and services in an economy in a year	X25
Macroenvironment	Consumer price index	CPI is a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services	X26

TABLE 4: Enterprise supply chain financial risk indicators.

Hyperparameters	Definition description	Value
learning_rate	Learning rate	0.875
max_depth	Maximum depth of each tree	15
min_child_weight	Minimum sum of instance weight needed in a child	1
subsample	Subsample ratio of instances for each tree	0.8245
gamma	Minimum loss reduction required to make a split during tree building	0
colsample_bytree	Subsample ratio of features for each tree	0.726

TABLE 5: Hyperparameters obtained by Bayesian optimization.



FIGURE 7: Model prediction error.

indicators whose partial correlation coefficient is greater than 0.6, delete the indicator with the smaller F value.

4.3. Rescreening of Indicators through Neural Network. Suppose the correct rate of neural network discrimination is A, and the number of samples that match the output result with the actual result is k, then: A = k/n. After the neural network deletes each index one by one, calculate the correct rate of discrimination and calculate one by one. Let d as the change rate of A. If the correct rate is higher than the original correct rate, the indicator will be deleted. The specific steps are as follows:

*Step 1*: Bring the data into the neural network and calculate the correct rate of judgment under all indicators *A*.

Step 2: Delete the *i*th indicator, and calculate the correct rate  $A'_i$  of the neural network model at this time:  $d_i = A'_i - A$  If  $d_i > 0$ , it proves that after deleting this indicator, the correct rate of neural network judgment is improved. On the contrary, the correct rate of judgment decreases.

Step 3: Sort the indicators with  $d_i > 0$  and find the indicator that has the greatest impact on the correct rate, namely  $\max_{1 \le i \le p} (d_i)$ , and delete the indicator.

Step 4: Repeat steps 2 and 3, deleting the indicators of  $d_i > 0$  in turn, until all  $d_i < 0$ .



FIGURE 8: Summary of model prediction accuracy.

4.4. Construction of Credit Risk Evaluation Index System. Standardized data were screened based on partial correlation coefficient-F value. The 84 second-level indicators in the table are classified according to the first-level indicators to calculate the partial correlation coefficient and the calculated value of F value. Taking the first-level indicator operating capacity as an example, there are 14 second-level indicators. Its partial correlation coefficient and F value calculation are listed in Table 1. Compare the F values of the two, and delete the index category with the smaller F value.

Next, we put all the secondary indicator training sets under the same primary indicator into the neural network model. By gradually deleting individual indicators, the degree of change in the accuracy of the model in the test set is judged. We take the first-level indicator *profitability* as an example, and the empirical results are shown in Table 2.

After the partial correlation–variance screening of the credit risk indicator system, on this basis, the neural network is used to gradually delete the indicator screening method with the largest reduction in the accuracy rate each time, and finally the number of remaining indicators are got. In a total of 84 secondary indicators, 38 indicators reflecting information duplication are deleted through partial correlation–variance, 20 indicators are deleted through neural network, and the remaining 26 indicators are shown in Table 3.

The supply chain risk index system presented in Table 4 can mainly reflect three aspects. First, the repayment ability willingness of SMEs are revealed according to the *operating ability*, *profitability*, *solvency*, and *development capability*. Second, through *platform as a service*, it reveals whether the company has stable orders and sustainable development capabilities. Finally, through the *macroenvironment*, it reveals the development stability of SMEs in the supply chain environment.

#### 5. Supply Chain Finance Credit Risk Assessment

*5.1. XGBoost Parameter Optimizations*. After the XGBoost hyperparameter optimization range is determined, the main steps of BO hyperparameter tuning are as follows:

Step 1. Build the BO-XGBoost supply chain financial credit evaluation prediction model and obtain the random initialization point of the hyperparameters within the parameter range according to the determined hyperparameters to be optimized  $X = [x_1, x_2, ..., x_n]$ , input the experimental data,



FIGURE 9: ROC curve with different models.

train the BO-XGBoost model, the response value of the objective function is  $Y = [f(x_1), f(x_2), \dots, f(x_n)]$ . Set up the Gaussian regression process.

*Step 2.* Based on the acquisition function PI, select the next hyperparameter combination sampling point  $x_{n+1}$  from the Gaussian regression model.

*Step 3.* The hyperparameter value of the XGBoost model is a new hyperparameter combination, and the model is trained to obtain the credit default prediction value of the risk model. If the error between the predicted value and the standard value meets the target requirements, the optimal hyperparameter combination is output.

Step 4. If the error between the predicted credit default value and the standard value fails to meet the target requirements, add  $(x_{n+1}, f(x_{n+1}))$  to the Gaussian regression model. Repeat steps 2 and 3 to find a hyperparameter combination that meets the target requirements, and stop the iteration.

The hyperparameters of the XGBoost model were tuned using BO, while the default values were used for the remaining parameters. The optimal hyperparameters of the XGBoost model are summarized in Table 5.

5.2. *Model Error Comparison.* Since the predicted results are presented in the form of probabilities. In this paper, the predicted value of the predicted probability of 0.5 is classified

as 1, that is, there will be supply chain financial risks, and the value of the predicted probability of less than 0.5 is classified as no supply chain financial risks will occur. According to the model prediction error in Figure 7, in the same running time, the MAE and MSE of the RF model without optimization are the lowest. However, in the performance of each model using the BO method, the MAE and MSE of the model both decreased, indicating the effectiveness of the BO algorithm and the BO-XGBoost model is the best performer. MAE and MSE reach 37.77% and 30.25%, respectively, which are much lower than the values of other models.

5.3. Prediction Accuracy Comparison. In this paper, the AUC value is the main optimization objective, and the accuracy and  $F_1$  are used for auxiliary comparison to verify the BO parameters for SVM, random forest, and XGBoost, respectively. The results we get are shown in Figure 8.

In the SVM, RF, and XGBoost models, the AUC values are similar, but the BO-XGBoost model has the highest AUC value. Additionally, BO has resulted in improved AUC values for SVM, RF, and XGBoost models. As shown in Figure 8, although the AUC values of BO-RF and BO-XGBoost are nearly identical, BO-XGBoost still performs better. XGBoost has a prediction accuracy of 91.4%, whereas the accuracy of other models without BO optimization is below 90%. These findings highlight the importance of BO in enhancing model performance. BO-XGBoost has demonstrated improvements of 3.4% in accuracy, 1.28% in AUC, and 2.03% in F1 score compared to XGBoost. The ROC curve in Figure 9 further validates the performance enhancement of the XGBoost model with BO. These empirical results indicate that the supply chain risk indicator system, after multiple indicator selections, effectively covers various aspects of supply chain risk occurrences. Moreover, appropriate parameter tuning through model optimization can further enhance the accuracy of predicting risk enterprises.

#### 6. Conclusion

SCF plays a crucial role in financing for SMEs. However, credit risks are common due to incomplete financial information. In this paper, a two-step approach involving partial correlation coefficient, variance analysis, and BP neural network is used to screen candidate indicators and construct a supply chain financial risk indicator system. Additionally, XGBoost and BO methods are employed to build a hybrid model called BO-XGBoost for predicting credit default in SMEs. The main conclusions of this study are given below.

It was determined that effective construction of corporate financial indicators, taking into account factors, such as profitability, solvency of core companies, the degree of transactions between SMEs and core companies, and the external macroenvironment faced by the supply chain, is essential to fully reflect the status of enterprises in SCF. Moreover, the prediction results were significantly improved when using hyperparameters obtained from BO compared to the default configuration parameters of XGBoost. When compared to SVM and RF models, the BO-XGBoost hybrid model demonstrated excellent predictive ability for credit risk of SME financing in SCF. The proposed combined model of XGBoost and BO has the potential to enhance the generalization ability of risk prediction, resulting in better performance in terms of prediction accuracy and the ability to effectively distinguish between risky companies and normal companies.

While this paper proposes a risk indicator construction method and demonstrates the good performance of the BO-XGBoost hybrid model for SME credit risk prediction in SCF, there are still challenges that need to be addressed in future research. First, the weight of core enterprises in collaborating with SMEs should be considered when analyzing SME defaults in SCF, as the role of core enterprises can significantly impact SME credit risk. Second, further refinement and improvement of the credit evaluation index system for SMEs in the SCF environment, tailored to the unique characteristics of the industry, can enhance the accuracy of credit risk prediction. Last, more efficient and faster optimization algorithms should be developed to reduce computational overhead of the BO-XGBoost hybrid model. Future research should focus on developing targeted risk models and optimizing algorithms to enhance the accuracy, efficiency, and applicability of credit risk prediction in the context of SCF for SMEs.

#### **Data Availability**

Data will be made available upon request.

#### **Conflicts of Interest**

The authors declare that they have no conflicts of interest.

### Acknowledgments

This research was supported by the National Social Science Foundation of China (21BJY002, 20AZD033), National Natural Science Foundation of China (72373078), and China Scholarship Council.

#### References

- S. C. Gherghina, M. A. Botezatu, A. Hosszu, and L. N. Simionescu, "Small and medium-sized enterprises (SMEs): the engine of economic growth through investments and innovation," *Sustainability*, vol. 12, no. 1, Article ID 347, 2020.
- [2] N. Yoshino and F. Taghizadeh-Hesary, "Optimal credit guarantee ratio for small and medium-sized enterprises' financing: evidence from Asia," *Economic Analysis and Policy*, vol. 62, pp. 342–356, 2019.
- [3] D. Mhlanga, "Financial inclusion in emerging economies: the application of machine learning and artificial intelligence in credit risk assessment," *International Journal of Financial Studies*, vol. 9, no. 3, Article ID 39, 2021.
- [4] L. M. Gelsomino, R. Mangiaracina, A. Perego, and A. Tumino, "Supply chain finance: a literature review," *International Journal* of *Physical Distribution & Logistics Management*, vol. 46, no. 4, 2016.
- [5] H. Song, K. Yu, and Q. Lu, "Financial service providers and banks' role in helping SMEs to access finance," *International Journal of Physical Distribution & Logistics Management*, vol. 48, no. 1, pp. 69–92, 2018.
- [6] R. Pellegrino, N. Costantino, and D. Tauro, "Supply chain finance: a supply chain-oriented perspective to mitigate commodity risk and pricing volatility," *Journal of Purchasing and Supply Management*, vol. 25, no. 2, pp. 118–133, 2019.
- [7] D. A. Wuttke, C. Blome, K. Foerstl, and M. Henke, "Managing the innovation adoption of supply chain finance—empirical evidence from six European case studies," *Journal of Business Logistics*, vol. 34, no. 2, pp. 148–166, 2013.
- [8] H.-C. Pfohl and M. Gomm, "Supply chain finance: optimizing financial flows in supply chains," *Logistics Research*, vol. 1, no. 3-4, pp. 149–161, 2009.
- [9] E. Hofmann and O. Belin, *Supply Chain Finance solutions*, Springer-Velag, Berlin, Heidelberg, 2011.
- [10] X. Zhao, K. H. Yeung, Q. Huang, and X. Song, "Improving the predictability of business failure of supply chain finance clients by using external big dataset," *Industrial Management & Data Systems*, vol. 115, no. 9, pp. 1683–1703, 2015.
- [11] J. Chen, T. Cai, W. He et al., "A blockchain-driven supply chain finance application for auto retail industry," *Entropy*, vol. 22, no. 1, Article ID 95, 2020.
- [12] H. Jiang, W.-K. Ching, K. F. C. Yiu, and Y. Qiu, "Stationary mahalanobis kernel SVM for credit risk evaluation," *Applied Soft Computing*, vol. 71, pp. 407–417, 2018.

- [13] N. Rtayli and N. Enneya, "Selection features and support vector machine for credit card risk identification," *Procedia Manufacturing*, vol. 46, pp. 941–948, 2020.
- [14] Y.-C. Chang, K.-H. Chang, and G.-J. Wu, "Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions," *Applied Soft Computing*, vol. 73, pp. 914–920, 2018.
- [15] Y. Guo, "Credit risk assessment of P2P lending platform towards big data based on BP neural network," *Journal of Visual Communication and Image Representation*, vol. 71, Article ID 102730, 2020.
- [16] G. Teles, J. J. P. C. Rodrigues, R. A. L. Rabê, and S. A. Kozlov, "Artificial neural network and bayesian network models for credit risk prediction," *Journal of Artificial Intelligence and Systems*, vol. 2, no. 1, pp. 118–132, 2020.
- [17] F. Shen, X. Zhao, Z. Li, K. Li, and Z. Meng, "A novel ensemble classification model based on neural networks and a classifier optimisation technique for imbalanced credit risk evaluation," *Physica A: Statistical Mechanics and its Applications*, vol. 526, Article ID 121073, 2019.
- [18] P. Z. Lappas and A. N. Yannacopoulos, "A machine learning approach combining expert knowledge with genetic algorithms in feature selection for credit risk assessment," *Applied Soft Computing*, vol. 107, Article ID 107391, 2021.
- [19] G. Du, Z. Liu, and H. Lu, "Application of innovative risk early warning mode under big data technology in internet credit financial risk assessment," *Journal of Computational and Applied Mathematics*, vol. 386, Article ID 113260, 2021.
- [20] E. Dumitrescu, S. Hué, C. Hurlin, and S. Tokpavi, "Machine learning for credit scoring: improving logistic regression with non-linear decision-tree effects," *European Journal of Operational Research*, vol. 297, no. 3, pp. 1178–1192, 2022.
- [21] M. Z. Abedin, C. Guotai, S. Colombage, and F.-E. Moula, "Credit default prediction using a support vector machine and a probabilistic neural network," *Journal of Credit Risk*, vol. 14, no. 2, pp. 1–27, 2018.
- [22] S. Khemakhem and Y. Boujelbene, "Support vector machines for credit risk assessment with imbalanced datasets," *International Journal of Data Mining, Modelling and Management*, vol. 10, no. 2, pp. 171–187, 2018.
- [23] W.-H. Hou, X.-K. Wang, H.-Y. Zhang, J.-Q. Wang, and L. Li, "A novel dynamic ensemble selection classifier for an imbalanced data set: an application for credit risk assessment," *Knowledge-Based Systems*, vol. 208, Article ID 106462, 2020.
- [24] H. Zhang, Y. Shi, X. Yang, and R. Zhou, "A firefly algorithm modified support vector machine for the credit risk assessment of supply chain finance," *Research in International Business* and Finance, vol. 58, Article ID 101482, 2021.
- [25] J. Luo, X. Yan, and Y. Tian, "Unsupervised quadratic surface support vector machine with application to credit risk assessment," *European Journal of Operational Research*, vol. 280, no. 3, pp. 1008–1017, 2020.
- [26] J. Li, H. Liu, Z. Yang, and L. Han, "A credit risk model with small sample data based on G-XGBoost," *Applied Artificial Intelligence*, vol. 35, no. 15, pp. 1550–1566, 2021.
- [27] J. Liu, S. Zhang, and H. Fan, "A two-stage hybrid credit risk prediction model based on XGBoost and graph-based deep neural network," *Expert Systems with Applications*, vol. 195, Article ID 116624, 2022.
- [28] A. Levy and R. Baha, "Credit risk assessment: a comparison of the performances of the linear discriminant analysis and the logistic regression," *International Journal of Entrepreneurship* and Small Business, vol. 42, no. 1/2, pp. 169–186, 2021.

- [29] M. Qi, Y. Gu, and Q. Wang, "Internet financial risk management and control based on improved rough set algorithm," *Journal of Computational and Applied Mathematics*, vol. 384, Article ID 113179, 2021.
- [30] Y. Liu and L. Huang, "Supply chain finance credit risk assessment using support vector machine-based ensemble improved with noise elimination," *International Journal of Distributed Sensor Networks*, vol. 16, no. 1, pp. 1–10, 2020.
- [31] J. Zhao and B. Li, "Credit risk assessment of small and medium-sized enterprises in supply chain finance based on SVM and BP neural network," *Neural Computing and Applications*, vol. 34, no. 15, pp. 12467–12478, 2022.
- [32] Y. Wu, X. Li, Q. Liu, and G. Tong, "The analysis of credit risks in agricultural supply chain finance assessment model based on genetic algorithm and backpropagation neural network," *Computational Economics*, vol. 60, no. 4, pp. 1269–1292, 2021.
- [33] S. Hosseini and A. A. Khaled, "A hybrid ensemble and AHP approach for resilient supplier selection," *Journal of Intelligent Manufacturing*, vol. 30, no. 1, pp. 207–228, 2019.
- [34] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785– 794, ACM, 2016.
- [35] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the human out of the loop: a review of bayesian optimization," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 148–175, 2016.