

Retraction

Retracted: Big Data Analytics and Discrete Choice Model for Enterprise Credit Risk Early Warning Algorithm

Security and Communication Networks

Received 1 August 2023; Accepted 1 August 2023; Published 2 August 2023

Copyright © 2023 Security and Communication Networks. 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.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] J. Yu, "Big Data Analytics and Discrete Choice Model for Enterprise Credit Risk Early Warning Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 3272603, 13 pages, 2022.

Research Article

Big Data Analytics and Discrete Choice Model for Enterprise Credit Risk Early Warning Algorithm

Jiangbo Yu 

Business School, Luoyang Normal University, Luoyang 471934, China

Correspondence should be addressed to Jiangbo Yu; yujiangbo@lynu.edu.cn

Received 14 October 2021; Revised 4 December 2021; Accepted 9 December 2021; Published 5 January 2022

Academic Editor: Chinmay Chakraborty

Copyright © 2022 Jiangbo Yu. 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.

A business credit risk early warning algorithm based on big data analysis and discrete selection model is presented to address the issues of poor sample fitting performance, long warning time, and low warning accuracy that plague the traditional enterprise credit risk early warning algorithm. A-share listed enterprises in China were chosen as the credit data source for screening the samples based on big data analysis. After screening, financial failure firms were coupled, and paired samples were created. The credit risk variables, which included financial and corporate governance characteristics, were chosen based on the created samples. The enterprise financial risk submodel and the nonfinancial risk submodel were built based on the enterprise credit risk variables, and the financial and nonfinancial index scores of enterprise customers were evaluated separately to develop a discrete choice model of enterprise credit risk. The algorithm's sample fitting performance was employed to achieve early warning of corporate credit risk. The algorithm based on big data analytics and discrete choice model is compared to the traditional method in order to verify its validity. The findings of the experiment reveal that the algorithm's sample fitting performance is superior to the traditional one, making it more suitable for enterprise credit risk early warning. The proposed model depicts 85% accuracy.

1. Introduction

Credit growth is to promote investment, production, and consumption. Traditional financing products such as medium- and long-term loans (mostly for infrastructure projects) and commercial-paper financing are still the major components of new loans for various enterprises at the moment. Many company construction projects have relatively easy access to a substantial number of bank credit money due to the trust in government credit. However, due to the unpredictability of many enterprise infrastructure projects' profitability or a lack of government experience, there are numerous hazards [1], including project rushes, project returns, and cost uncertainty. Because credit cards are revolving credit lines, lenders and investors have more possibilities for actively monitoring and managing them than other types of retail loans, such as mortgages. As a result, maintaining credit card portfolios might be a significant source of revenue for financial institutions. Better risk management might result in annual savings of hundreds of millions of

dollars for financial organizations. For financial institutions, better risk management could lead to financial revenue of hundreds of thousands of dollars. Lenders, for example, could reduce their exposure by cutting or freezing credit lines on accounts that are likely to default. Early credit risk prediction is necessary because effective application of the above risk management measures necessitates banks' ability to identify accounts that are likely to default.

These risks could cause issues during the implementation phase, necessitating a comprehensive risk assessment and a project cost budget. At the same time, due to the nature of public interests, many projects are not heavily commercialized, putting their future earnings in jeopardy. Furthermore, large-scale projects are typically expensive, take a long time to build, and have a significant payback period, during the operation phase; it is difficult to ensure debt-paying ability. As a result, company credit risk is becoming more complex and expansive, posing a slew of new and challenging criteria for enterprise credit risk management. The risk warning system can help businesses improve their resilience, adaptability, and

competitiveness, as well as preventing crises from forming or germinating in the first place.

This paper expanded the current research status and relevance of financial risk early warning, as well as the development background, current position, and future difficulties, based on reviewing and assessing prior research works. This paper proposes a novel technique that can predict the potential risk related to finances and protect the organization from potential credit risks. The detailed sections are arranged as follows: Section 2 introduces the related works; Section 3 discusses the sample screening structure and variable selection process; Section 4 discusses the experimental results of big data analytics and discrete choice model for enterprise credit risk early warning algorithm; Section 5 is the conclusion.

2. Related Works

According to relevant experts, financial intermediary services are reasonably mature, and the driving factor behind their promotion is to increase value. Bayesian model to investigate the early income mechanism in order to help organizations predict and reduce the risk of loan transactions in [2] is used. Although the system may detect early signs of business credit risk, it suffers from a lack of precision. In [3] the authors proposed an enterprise credit risk assessment method based on improved genetic algorithm, which satisfies the adaptability of improved algorithm to corporate credit risk. The algorithm can accurately forewarn the credit risk of enterprises, but it takes a long time.

Increased value can be obtained by expanding the creation of value-added services and effectively lowering various expenses, i.e., increasing income while lowering expenditure [4]. Collaboration between the supply chain and financial service providers will be varied, boost the value of both parties, and advance the financial industry to a higher level [4]. The linear regression method is used by related researchers in loan risk evaluation research. The linear regression model's underlying conceit is to create a regression equation based on individual variables in order to calculate the likelihood that consumer credit performance is "excellent" [5] but many flaws in parametric statistical methods have been identified by researchers.

Through balance sheets, credit business, and other channels, commercial banks and other financial institutions form a complicated network interaction. Risk identification, risk assessment, risk early warning, and risk treatment are the key components of financial risk early warning job, which may be further split into financial risk organization form, indicator system, and prediction method. When banking is subjected to internal and external shocks that cause debt failures, volatility risks are transferred through interbank credit channels, and risk spillovers affect the activities of other banks and financial institutions, posing systemic hazards [6].

Researchers in related fields have also studied enterprise credit, starting from the perspective of organizational security and forecasting techniques for risk management. In [4] the BP neural network model to evaluate the project risk

is used. The model can effectively avoid overtraining and overfitting and has good generalization ability. Compared with the fuzzy theory, the influence of human factors is avoided. In [5] the authors discussed the comprehensive gray correlation of computational languages and their numbers and the enterprise credit as examples to analyze and compare and finally verified the practicability, rationality, and effectiveness of this method. But the above algorithm has the problem of poor sample fitting performance.

Considering the sample fitting performance of enterprise credit risk, this paper proposes an enterprise credit risk early warning algorithm based on big data analytics and discrete choice model.

3. Basic Definitions

3.1. Sample Screening Structure and Variable Selection

3.1.1. Sample Screening. Based on big data analytics, A-share listed companies in China were selected as sample sources, and nonfinancial listed companies that were specially treated (ST) for the first time for financial reasons were used as samples of financial failure companies. On this basis, small- and medium-sized listed companies and large listed companies are screened by industry [7, 8]. The industry division standards of listed companies are mainly according to the "Guidelines for the Classification of Listed Companies" issued by the Securities Futures Commission. Large, medium, and small listed enterprises are screened according to the "Interim Provisions on Standards of Small and Medium-Sized Enterprise" jointly issued by the State Economic and Trade Commission and another three ministries, and the "Supplementary Standards for the Division of Large, Medium, and Small Nonindustrial Enterprises" issued by the State-Owned Assets Supervision and Administration Commission. The "Interim Provisions on Standards of Small and Medium-Sized Enterprise" are shown in Table 1.

The "Supplementary Standards for the Division of Large, Medium, and Small Nonindustrial Enterprises" are shown in Table 2 [8].

The industry classifications in the "Guidelines for the Classification of Listed Companies" are not exactly the same as those in the "Interim Provisions on Standards of Small and Medium-Sized Enterprise" and the "Supplementary Standards for the Division of Large, Medium, and Small Nonindustrial Enterprises." Inconsistencies in the sector are quite easy to create. Some large-scale industries are more precisely matched according to their secondary industries to eliminate these industry matching errors [9, 10]. The sample was further screened as follows:

- (i) Companies with significant asset reorganization during the period from $(t-5)$ to $(t-2)$ were excluded from the ST companies, and companies with missing data were excluded as well.
- (ii) A total of 132 samples from ST companies were left, including 80 small and medium-sized listed enterprises and 52 large listed enterprises.

TABLE 1: Interim Provision on Standards of Small and Medium-Sized Enterprise.

Industry	Number of people engaged	Sales volume	Total assets	Explain
Industry	300–2000	30 million–3 billion	40 million–4 billion	Medium-sized enterprises must meet the lower limit of three indicators at the same time, and the rest are small enterprises.
Construction business	600–300	30 million–3 billion	40 million–4 billion	Medium-sized enterprises must meet the lower limit of three indicators at the same time, and the rest are small enterprises.
Wholesale business	100–200	30 million–3 billion	—	Medium-sized enterprises must meet the lower limit of three indicators at the same time, and the rest are small enterprises.
Retail	100–500	10–150 million	—	Medium-sized enterprises must meet the lower limit of three indicators at the same time, and the rest are small enterprises.
Communications and transportation industry	500–3000	30 million–3 billion	—	Medium-sized enterprises must meet the lower limit of three indicators at the same time, and the rest are small enterprises.
Postal industry	400–1000	30 million–3 billion	—	Medium-sized enterprises must meet the lower limit of three indicators at the same time, and the rest are small enterprises.
Accommodation and catering	400–800	30–150 million	—	Medium-sized enterprises must meet the lower limit of three indicators at the same time, and the rest are small enterprises.

TABLE 2: Supplementary Standards for the Division of Large, Medium, and Small Nonindustrial Enterprises.

Industry	Index name	Company	Large	Medium-sized	Small-scale
Agriculture, forestry, animal husbandry, and fishery	Number of people employed sales volume	People ten thousand RMB	≥ 3000	500–3000	< 500
			≥ 15000	1000–15000	< 1000
Warehousing industry	Number of people employed sales volume	People ten thousand RMB	≥ 500	100–500	< 100
			≥ 15000	1000–15000	< 1000
Estate	Number of people employed sales volume	People ten thousand RMB	≥ 200	100–200	< 100
			≥ 15000	1000–15000	< 1000
Finance	Number of people employed sales volume	People ten thousand RMB	≥ 500	100–500	< 100
			≥ 50000	5000–50000	< 5000
Geological exploration and water conservancy environmental management	Number of people employed sales volume	People ten thousand RMB	≥ 2000	600–2000	< 600
			≥ 20000	2000–20000	< 2000
Sports and entertainment industry	Number of people employed sales volume	People ten thousand RMB	≥ 600	200–600	< 200
			≥ 15000	3000–15000	< 3000
Information transmission industry	Number of people employed sales volume	People ten thousand RMB	≥ 400	100–400	< 100
			≥ 30000	3000–30000	< 3000
Computer service and software industry	Number of people employed sales volume	People ten thousand RMB	≥ 300	100–300	< 100
			≥ 30000	3000–30000	< 3000
Leasing industry	Number of people employed sales volume	People ten thousand RMB	≥ 300	100–300	< 100
			≥ 15000	1000–15000	< 1000
Business and technology services	Number of people employed sales volume	People ten thousand RMB	≥ 400	100–400	< 100
			≥ 15000	1000–15000	< 1000
Resident service industry	Number of people employed sales volume	People ten thousand RMB	≥ 800	200–800	< 200
			≥ 15000	1000–15000	< 1000
Other enterprises	Number of people employed sales volume	People ten thousand RMB	≥ 500	100–500	< 100
			≥ 15000	1000–15000	< 1000

Table 3 gives the statistics of the ST samples.

For the purpose of determining the starting point of financial failure, the first two years of ST are used. That is, if the year of ST is t year, then $(t-2)$ year is the starting point of financial failure [11, 12]. The warning capability of the loss in the first three years was the subject of inquiry, that is, the

warning capability in the year of $(t-5)$, $(t-4)$, and $(t-3)$ regarding whether there will be a financial failure in the $(t-2)$ year, taking into consideration the timeliness of financial failure warnings. The first two years of ST are the starting point of financial failure when choosing a financial failure point.

TABLE 3: Statistics of ST samples.

CSRC classification	2015	2016	2017	2018	2019	Total	Proportion
Agriculture, forestry, animal husbandry, and fishery	0	1 (0/1)	1 (0/1)	2 (0/2)	0	4 (0/4)	3.03
Extractive industry	0	3 (0/3)	2 (0/2)	0	0	5 (0/5)	3.79
Manufacturing industry	11 (4/7)	20 (7/13)	25 (9/16)	13 (6/7)	10 (3/7)	79 (29/50)	59.9
Communications and transportation industry	0	0	1 (0/1)	0	0	1 (0/1)	0.76
Information technology industry	0	4 (0/4)	3 (1/2)	0	1 (1/0)	8 (2/6)	6.06
Wholesale and retail	0	3 (0/3)	2 (2/0)	1 (1/0)	0	6 (3/3)	4.55
Estate	0	11 (8/3)	3 (2/1)	2 (2/0)	3 (2/1)	19 (14/5)	14.4
Social service industry	1 (1/0)	2 (1/1)	1 (1/0)	0	0	4 (3/1)	3.03
Cultural communication industry	0	1 (0/1)	0	0	0	1 (0/1)	0.76
Production and supply of electricity, gas, and water	1 (0/1)	0	2 (1/1)	1 (0/1)	1 (0/1)	5 (1/4)	3.79
Total	13 (5/8)	45 (16/29)	40 (16/24)	19 (9/10)	15 (6/9)	132 (52/80)	100

Note. The figures in brackets are the number of St large companies and St small and medium-sized companies, respectively.

3.1.2. Sample Construction. After the sample screening, the financially failed companies were paired to construct paired samples. The non-ST samples paired with the ST sample are derived from the companies that never have an ST in the sample period and have excluded the IPO (Initial Public Offerings). Because the existing standards have a multidimensional definition of the enterprise scale, it is easy to have mismatch asset scale, which will cause the inhomogeneity in paired samples [13]. In order to reduce this mismatch, from the three-dimensional characteristics: asset size, sales revenue, and number of employees, 80 small and medium-sized enterprises and 52 large enterprises were paired in a ratio of 1 : 2 in the same year and in the same industry. The time of pairing is the first two years of the financial failing firm before ST, in the ($t-2$) year. After pairing, a total of 396 samples were obtained, including 240 small and medium-sized enterprises (SMEs) and 156 large enterprises (LEs) [14]. Then, 180 samples were randomly selected from the SMEs group as the estimated samples to build the model, and the remaining 60 were used as verification samples to test the warning effect of the model.

3.1.3. Variable Selection. Based on the constructed samples, corporate credit risk variables were selected, including financial variables and corporate governance variables [15, 16]. 28 financial ratios were selected from seven categories of financial indicators, which reflect the financial leverage structure, solvency, profitability, operating capacity, growth potential, investment income level, and cash flow status of the listed company, as shown in Table 4.

In order to eliminate industry differences in financial ratios, the industry median adjustments were made on an annual basis for all 28 financial ratios, as follows:

$$R_{it} = \frac{X_{it}}{X_{igt}}, \quad (1)$$

where X_{igt} represents the median of financial ratio i in the industry g in the t year, X_{it} represents the median of the industry, and R_{it} represents the annual adjustment value of financial ratio. All financial ratio data comes from the Wind database [17].

The corporate governance variables investigate the impact of corporate governance characteristics on financial failures mainly from the ownership structure, actual controller type, board structure, and executive incentives as shown in Figure 1.

Table 5 provides a full overview of the above corporate governance characteristics. The data for all corporate governance variables is derived from the ‘‘Listed Corporate Governance Structure Database’’ in the CCER China Economic and Financial Database [18].

3.2. Construction of Warning Model of Enterprise Credit Risk

3.2.1. Construction of Enterprise Financial Risk Submodel. The enterprise financial risk submodel was built based on the enterprise credit risk variables to provide a high-precision data source for the enterprise credit risk warning model. To eliminate the correlation between financial indicators, the 21 secondary financial indicators were first subjected to principal component analysis (PCA). Then the extracted principal components were used to construct a submodel of corporate financial risk [19].

The specific operation of principal component analysis is as follows:

- (1) A data file has been created. Numerical variables $X_1, X_2, X_3, \dots, X_{21}$ were defined. The collected 21 scalar values were then standardized.
- (2) Principal component analysis was performed on the standardized data. The results are shown in Table 6. The cumulative variance contribution rate of the seven principal components has reached 84.62%, according to the cumulative variance contribution rate table and the principal component analysis matrix. From the 8th one, Eigen values are all less than 1. It was observed that 7 principal components effectively reflect 84.62% of the information in the indicator system, substantially simplifying the research process compared to 21 indicators components.
- (3) The principal component score was calculated. According to the cumulative variance contribution

TABLE 4: Selection of financial variables.

Financial variables and symbols	Financial variables and symbols	Financial variables and symbols
Asset liability ratio (X_1)	Operating profit/total operating revenue (X_{11})	Earnings per share (X_{21})
Equity ratio (X_2)	Return on equity (X_{12})	Net assets per share (X_{22})
Current assets ratio (X_3)	Total assets ratio of retained earnings (X_{13})	Total cash to assets ratio (X_{23})
Current liability ratio (X_4)	Total assets turnover rate (X_{14})	Net cash flow growth rate (X_{24})
Working capital/total assets (X_5)	Turnover rate of accounts receivable (X_{15})	Accounts payable/total assets (X_{25})
Current ratio (X_6)	Accounts payable turnover (X_{16})	Accounts payable/current assets (X_{26})
Quick ratio (X_7)	Inventory turnover (X_{17})	Current market value debt ratio (X_{27})
Cash flow liability ratio (X_8)	Fixed assets turnover rate (X_{18})	P/E ratio (X_{28})
Interest cover (X_9)	Growth rate of total assets (X_{19})	
Return on total assets (X_{10})	Operating revenue growth rate (X_{20})	

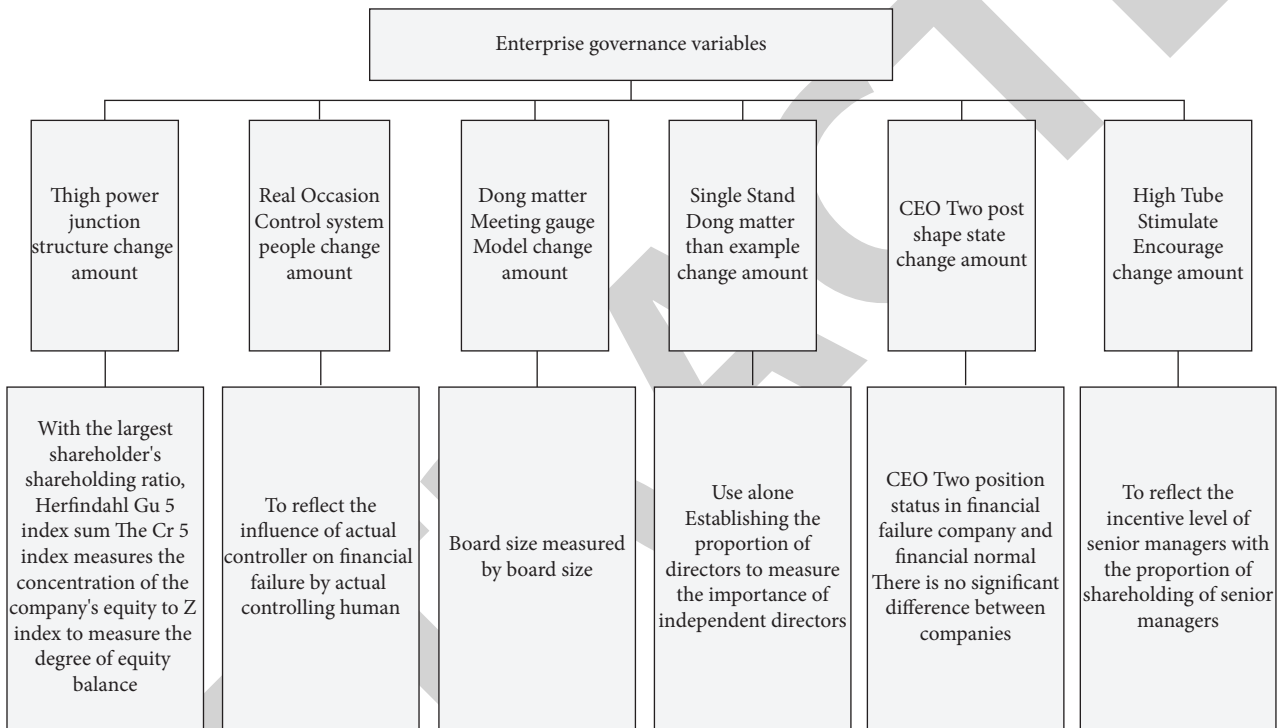


FIGURE 1: Composition of enterprise governance variables.

TABLE 5: Enterprise governance variables and their meanings.

Variable type	Variable name and symbol	Specific meaning
Ownership structure	Shareholding ratio of the largest shareholder (G_1)	Number of shares/total share capital held by the largest shareholder
	CR_5 index (G_2)	The sum of the shareholding ratios of the top five major shareholders
	Herfindahl_5 index (G_3)	Sum of the squares of the shareholding ratio of the top five major shareholders
	Z index (G_4)	Number of shares held by the first largest shareholder/number of shares held by the second largest shareholder
Type of actual controller	Actual controller (G_5)	If the last controlling shareholder belongs to state-owned holding, take 1; otherwise take 0
Board structure	Board size (G_6)	Number of board members (natural logarithm)
	Proportion of independent directors (G_7)	Total number of independent directors/total number of directors
Executive incentive	CEO status (G_8)	The chairman and general manager shall be one person, otherwise 0
	Shareholding ratio of senior executives (G_9)	Total shares/total share capital held by senior executives at the end of the year

TABLE 6: Analysis results.

Component	Initial Eigen values			Extraction sums of squared loadings		
	Total	Variance contribution rate (%)	Cumulative (%)	Total	Variance contribution rate (%)	Cumulative (%)
1	5.298	26.228	26.228	5.298	25.228	25.228
2	3.599	17.136	42.364	3.599	17.136	42.364
3	2.681	12.768	55.133	2.681	12.768	55.133
4	2.048	9.752	64.885	2.048	9.752	64.885
5	1.597	7.606	72.491	1.597	7.606	72.491
6	1.501	7.147	79.638	1.501	7.147	79.638
7	1.046	4.982	84.620	1.046	4.982	84.620
8	0.670	3.193	87.812			
9	0.549	2.613	90.426			
10	0.524	2.495	92.921			
11	0.437	2.081	95.003			
12	0.390	1.855	96.858			
13	0.261	1.244	98.101			
14	0.166	0.790	98.891			
15	0.109	0.519	99.410			
16	0.063	0.302	99.712			
17	0.37	0.178	99.890			
18	0.13	0.064	99.954			
19	0.007	0.034	99.987			
20	0.003	0.013	100.000			
21	1.29E-016	6.14E-016	100.000			

rate table, the Eigen values of the coefficient matrix of the seven principal components are

$$\begin{cases} \lambda_1 = 5.298, \\ \lambda_2 = 3.599, \\ \lambda_3 = 2.681, \\ \lambda_4 = 2.048, \\ \lambda_5 = 1.597, \\ \lambda_6 = 1.501, \\ \lambda_7 = 1.046. \end{cases} \quad (2)$$

For values in each column in the principal component analysis matrix, they are divided by $\sqrt{\lambda_1}$, $\sqrt{\lambda_2}$, $\sqrt{\lambda_3}$, $\sqrt{\lambda_4}$, $\sqrt{\lambda_5}$, $\sqrt{\lambda_6}$, and the unit eigenvector corresponding to each Eigen value was obtained. Thus, seven principal components ($F_1, F_2, F_3, F_4, F_5, F_6$, and F_7) can be expressed [20–22], and the standardized variables can be used to determine each principal component's final score.

Then, using a BP neural network, a submodel of corporate financial risk was created. Figure 2 depicts the individual steps.

3.2.2. Construction of Nonfinancial Risk Submodel. A nonfinancial submodel was constructed based on corporate credit risk variables. First, the target level O: nonfinancial risk score was set. Then, the criterion level C: there are enterprise scale score, shareholding structure, and audit opinion. Last, measure level P: there are total enterprise assets, annual income, the proportion of the largest shareholder, the shareholding ratio of the two major shareholders,

and the type of audit opinion. Because the target level and the measure level are linked by the relationship between the primary and secondary indicators, considering that the total assets and the annual income of the enterprise will also affect the audit opinion, the logical relationships between different levels are shown in Figure 3.

The importance level between each indication can be assessed using the hierarchical structure model of nonfinancial indicators to produce the judgment matrix of each level. The particular results are presented in Figure 4 after clicking “calculating result” in the yaahp software.

After obtaining the proportion of nonfinancial indicators, a nonfinancial submodel can be constructed as given in the following equation:

$$S = (0.3339 \cdot Q_z) + (0.3339 \cdot Q_w) + (0.1602 \cdot V) + (0.1187 \cdot D_1) + (0.0533 \cdot D_2), \quad (3)$$

where S represents the comprehensive score of nonfinancial indicators, Q_z represents the total assets of the enterprise, Q_w represents the annual income of the enterprise, V represents the type of audit opinion, D_1 represents the shareholding ratio of the largest shareholder, and D_2 represents the shareholding ratio of the two major shareholders.

3.2.3. Construction of Enterprise Credit Risk Early Warning Model. The input variables must be determined initially when creating a model. The output variables can be calculated using the input variables. The enterprise comprehensive credit score prediction based on the discrete choice model is shown in Figure 5.

The enterprise financial indicator submodel and the nonfinancial indicator submodel are currently being built, and the enterprise customer's financial and nonfinancial

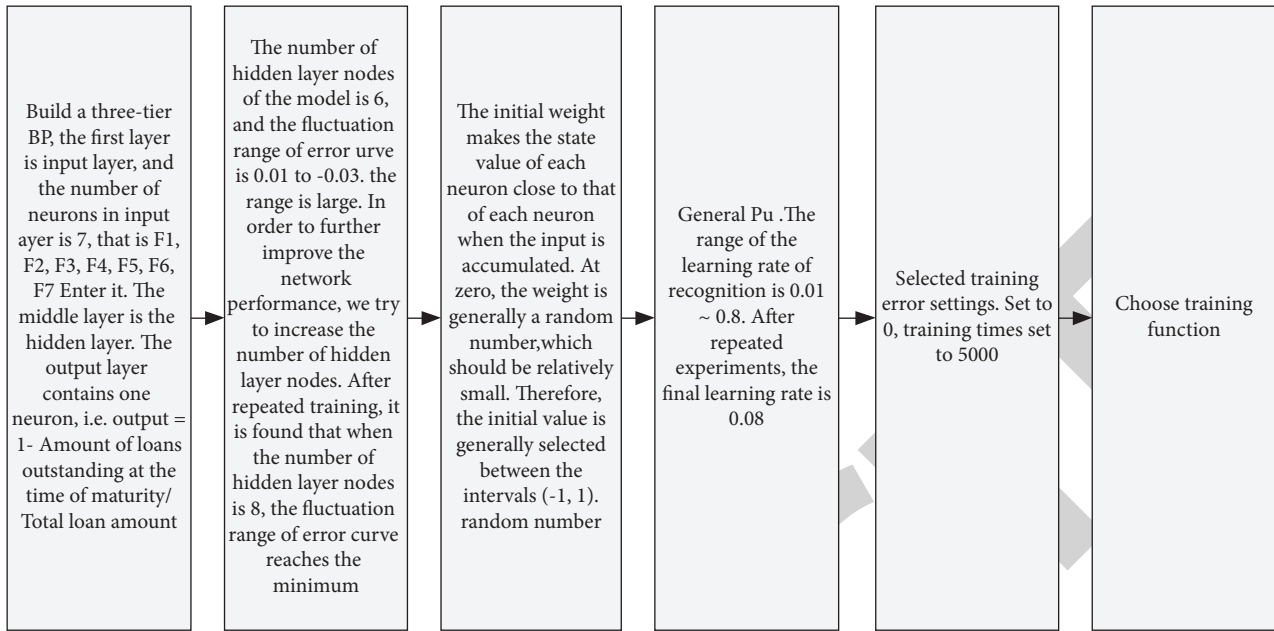


FIGURE 2: Steps of model construction.

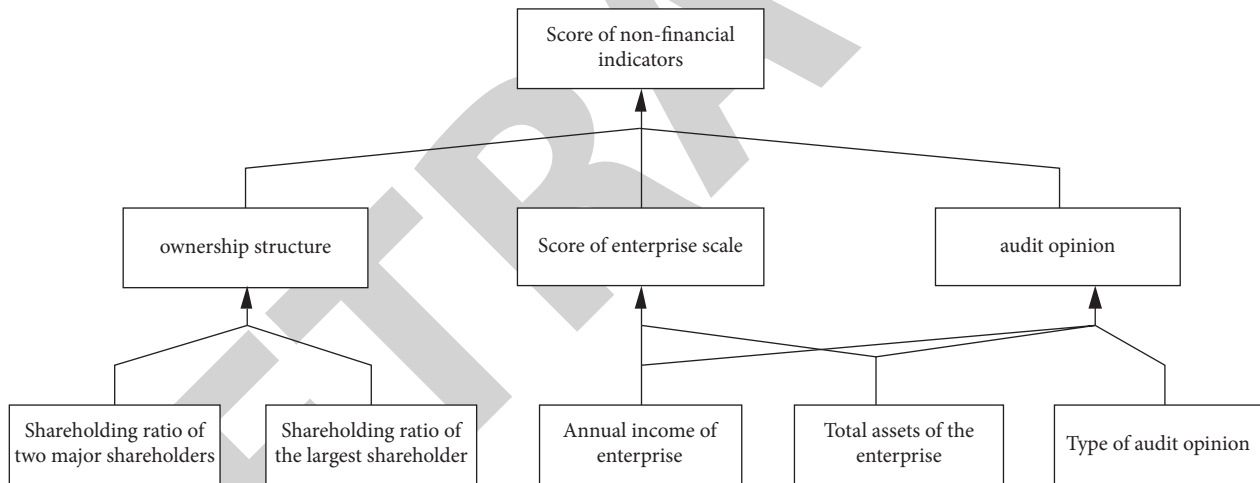


FIGURE 3: Hierarchy model of nonfinancial indicators.

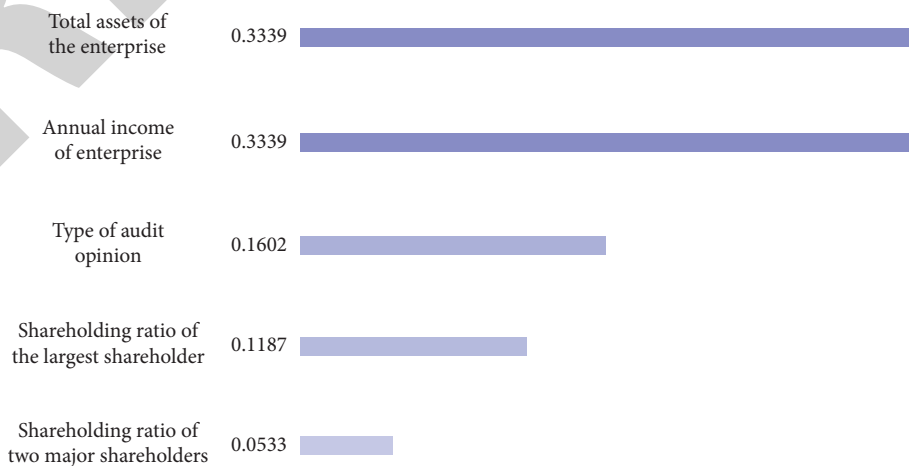


FIGURE 4: Proportion of nonfinancial indicators.

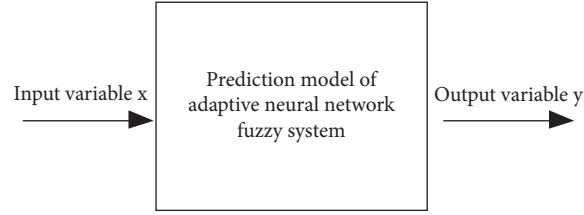


FIGURE 5: Prediction of enterprise comprehensive credit score.

indicator scores can be obtained using the two models. In this way, an enterprise credit risk early warning model based on discrete choice model can be constructed. The warning model has two input variables, namely, the financial indicator score and the nonfinancial indicator score of the enterprise. It has one output: “1-customer default probability.” Default probability refers to the possibility that the borrower can repay the loan principal and interest or fulfill the relevant obligations according to the contract within a certain period of time in the future, and it is inversely proportional to credit risk. Usually, the probability of default refers to the one-year default probability [23].

3.3. Design of Enterprise Credit Risk Early Warning Algorithm.

The sample fitting design enterprise credit risk early warning algorithm is based on the created enterprise credit risk early warning model. The specific steps of the enterprise credit risk early warning algorithm are as follows:

- (1) The relative importance of each indicator was quantified: The importance of the credit risk indicator was scored. After removing the top and lowest scores, the average score of the left is the indicator's score [24, 25].
- (2) A warning judgment matrix was constructed: If a_{ij} represents the relative importance scale of elements a_i and a_j to elements in the previous level, then a positive reciprocal judgment matrix can be constructed as follows:

$$A = (a_{ij})_{n \times m} = \begin{bmatrix} a_1 & a_1 & \dots & a_1 \\ a_2 & a_2 & \dots & a_n \\ \frac{a_2}{a_1} & \frac{a_2}{a_2} & \dots & \frac{a_2}{a_n} \\ \frac{a_1}{a_1} & \frac{a_1}{a_2} & \dots & \frac{a_1}{a_n} \\ \dots & \dots & \dots & \dots \\ \frac{a_n}{a_1} & \frac{a_n}{a_2} & \dots & \frac{a_n}{a_n} \end{bmatrix}_{n \times m}, \quad (4)$$

where A represents the positive reciprocal judgment matrix, and n and m represent constants, respectively.

Based on the positive reciprocal judgment matrix, the enterprise credit risk early warning judgment

matrix was constructed, including two input early warning judgment matrices as given in matrices 5 and 6:

$$R_3 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix},$$

$$R_4 = \begin{bmatrix} 1 & \frac{1}{2} & 2 & 1 & \frac{1}{6} \\ 2 & 1 & 4 & 2 & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} & 1 & \frac{1}{2} & \frac{1}{8} \\ 1 & \frac{1}{2} & 2 & 1 & \frac{1}{6} \\ 6 & 4 & 8 & 6 & 1 \end{bmatrix}, \quad (5)$$

where R_3 and R_4 represent two input warning judgment matrices.

An output warning judgment matrix is as follows:

$$G = \begin{bmatrix} 1 & 1 & 2 & \frac{1}{3} & 2 \\ 1 & 1 & 2 & \frac{1}{3} & 2 \\ \frac{1}{2} & \frac{1}{2} & 1 & \frac{1}{4} & 1 \\ 3 & 3 & 4 & 1 & 4 \\ \frac{1}{2} & \frac{1}{2} & 1 & \frac{1}{4} & 1 \end{bmatrix}, \quad (6)$$

where G represents the output warning judgment matrix.

TABLE 7: Specific information of experimental subjects.

Features	A area	B area	C area	D area	E area	Total
<i>Industry attribute</i>						
Agriculture, forestry, animal husbandry, and fishery	0	1	1	4	0	6
Mining industry	0	10	3	5	3	21
Manufacturing industry	2	85	48	48	162	345
Construction business	0	0	1	3	3	7
Wholesale and retail	0	0	0	3	2	5
Leasing business service industry	0	0	0	1	0	1
Transportation, warehousing, and postal services	0	0	0	0	2	2
Production and supply of electric gas and water	0	0	0	0	2	2
Other industries	0	7	1	6	11	25
Total	2	103	54	70	185	414
<i>Types of enterprises</i>						
State-owned enterprise	0	0	0	0	3	3
Collective enterprise	0	4	0	3	10	17
Private enterprise	2	99	54	67	172	394
Total	2	103	54	70	185	414
<i>Enterprise age</i>						
Within 5 years	2	43	42	20	73	180
6–15 years	0	48	10	43	93	194
16–25 years	0	10	2	5	15	32
More than 26 years	0	2	0	2	4	8
Total	2	103	54	70	185	414
<i>Credit rating</i>						
Class A	0	11	9	7	29	56
Class AA	1	74	15	37	91	218
Class AAA	1	18	30	26	65	140
Total	2	103	54	70	185	414

A fuzzy layer warning judgment matrix is as follows:

$$R_5 = \begin{bmatrix} 1 & 3 & 3 & 1 & 4 \\ \frac{1}{3} & 1 & 1 & \frac{1}{3} & 2 \\ \frac{1}{3} & 1 & 1 & \frac{1}{3} & 2 \\ 1 & 3 & 3 & 1 & 4 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} & 1 \end{bmatrix}, \quad (7)$$

where R_5 represents the fuzzy layer warning judgment matrix.

- (3) The two input warning judgment matrices are utilized to create one output warning judgment matrix and one fuzzy layer warning judgment matrix.

The enterprise credit risk early warning was realized based on the aforementioned judgment matrix. This will assist in preventing the crisis and ensuring the company's steady development, allowing the company to grow sustainably.

4. Experimental Researches

In order to ensure the validity of the experimental results, compare the algorithm in this paper with the algorithms proposed by asset-pricing model [2] algorithm and two-stage econometric approach proposed in [3]. The sample fit,

time used, and accuracy of three algorithms were compared. The sample fit is judged through the volatility of the sample fitting curve. If the volatility of the sample fitting curve is small, the sample is proved to have good fitting performance.

4.1. Experimental Process. Enterprise credit risk warning algorithm based on big data analytics and discrete choice model was used to conduct experiment on enterprise credit risk early warning. The data utilized in the experiment are all from my sea data set, and the data is analyzed using the online data analysis software MOA (an experimental tool for massive online analysis). 414 enterprises from 5 regions were selected as experimental objects. The specific information is shown in Table 7.

Further, 414 experimental enterprises were subdivided and analyzed in their asset scale, number of employees, and sales revenue. The relevant statistical results are shown in Table 8.

In order to ensure the validity of the experimental results, compare the algorithm in this paper with the algorithms proposed in [2, 3]. The sample fit, time used, and accuracy of three algorithms were compared. The sample fit is judged through the volatility of the sample fitting curve. If the volatility of the sample fitting curve is small, the sample is proved to have good fitting performance.

4.1.1. Sample Fit. Figure 6 depicts the sample fit comparison findings of the algorithm in this paper with the algorithms presented by [2, 3].

TABLE 8: Statistical description of experimental samples.

	“A” area	“B” area	“C” area	“D” area	“E” area	Total
<i>Asset scale (unit: 10000 yuan)</i>						
0–200 (contain)	0	5	0	2	2	9
200–500 (contain)	0	4	0	7	11	22
500–1000 (contain)	0	25	8	12	46	91
1000–3000 (contain)	1	61	31	23	71	187
3000–5000 (contain)	0	4	7	11	22	44
5000 above	1	4	8	15	33	61
Total	2	103	54	70	185	414
<i>Number of employees (unit: No.)</i>						
0–50 (contain)	0	24	17	15	27	83
50–200 (contain)	2	65	33	28	98	226
200–500 (contain)	0	11	4	14	35	64
500–100 (contain)	0	3	0	10	23	36
1000 above	0	0	0	3	2	5
Total	2	103	54	70	185	414
<i>Sales revenue (unit: 10000 yuan)</i>						
0–200 (contain)	0	8	0	4	2	14
200–500 (contain)	0	4	1	8	7	20
500–1000 (contain)	0	18	2	5	7	32
1000–3000 (contain)	0	40	17	22	50	129
3000–5000 (contain)	1	8	15	11	33	68
5000 above	1	25	19	20	86	151
Total	2	103	54	70	185	414

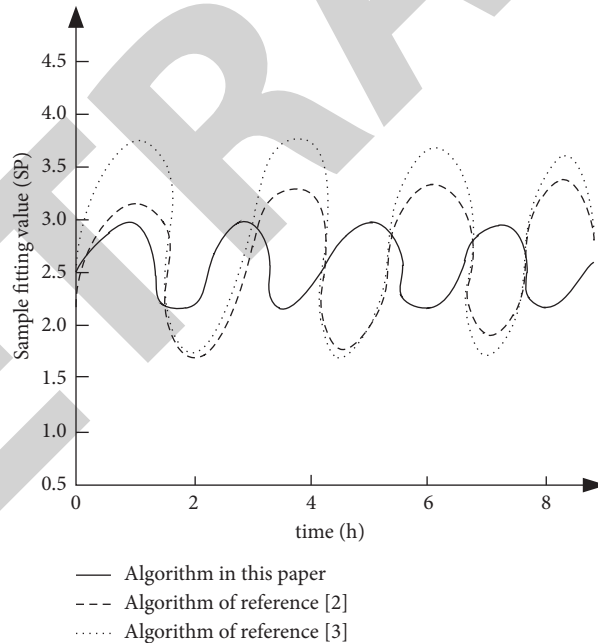


FIGURE 6: Experimental comparison results of sample fit.

According to Figure 6, the sample fitting curve of the algorithm in this paper has less volatility, which proves that its sample fit is better than those of the asset-pricing model [2] algorithm and two-stage econometric approach proposed in [3].

4.1.2. Time Used. A comparison was done between the time needed by the method in this paper and the time used by

algorithms presented by [2, 3] based on the number of enterprises necessary to compute, and the results are displayed in Figure 7.

According to the experimental results in Figure 7, the enterprise credit risk early warning algorithm based on big data analytics and discrete choice model takes less time, and its curve gradually flattens after the number of company's reaches 1000 and does not change until the number of enterprises reaches 2000. It depicts that when the number of

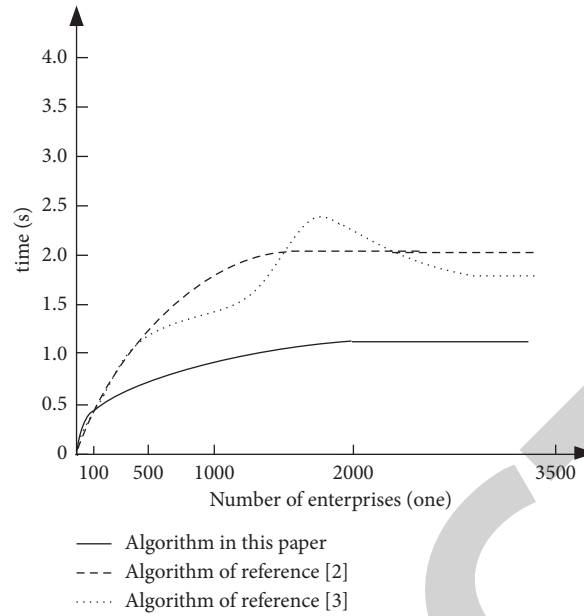


FIGURE 7: Experimental comparison results of time used.

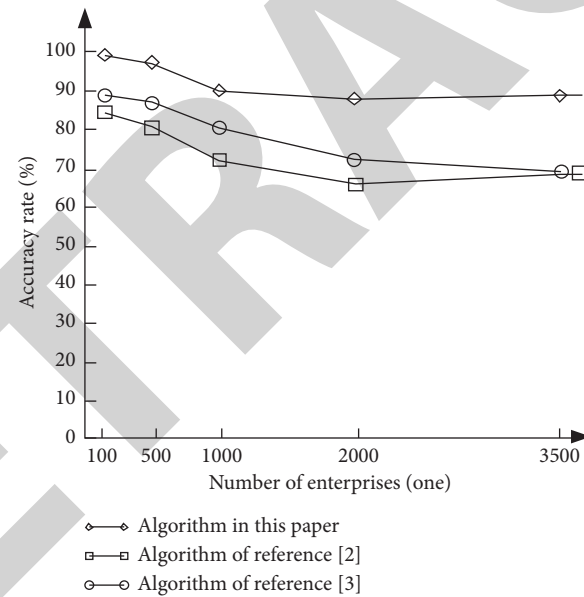


FIGURE 8: Experimental comparison results of accuracy.

enterprises is greater than 100, the enterprise credit risk early warning algorithm in this paper uses less time than the other two algorithms. This is because the algorithm in this paper firstly filters the enterprise credit risk before constructing the sample. When there are a significant number of enterprises that have set aside sufficient time for the prompt prevention of corporate credit risk, this step allows the algorithm to make the credit risk warning in a shorter amount of time.

4.1.3. Accuracy. The accuracy rate of three algorithms is used to assess the accuracy of their calculating results. The accuracy of the algorithm outcomes is high if the accuracy rate is high and stable. The algorithm’s outputs, on the other

hand, have a poor level of accuracy. Figure 8 depicts the comparing results.

Figure 8 shows that the algorithm in this paper has a higher accuracy and is more stable. Therefore, the enterprise credit risk early warning algorithm based on big data analytics and discrete choice model has higher accuracy. This is related to the sample fit experiment in 3.2.1. The higher the sample fitting performance, the higher the accuracy rates of the algorithm and the higher the accuracy of the algorithm results. Because the algorithm in this paper improves the sample fitting performance, the accuracy of the algorithm results is also improved, which makes the enterprise credit risk early warning more accurate, reduces unnecessary credit risk prevention, and avoids waste of resources.

5. Conclusions

The early warning system can protect an enterprise from the biggest losses and bankruptcy. Hence, some smart techniques need to be devised in order to warn the enterprise regarding credit risk. The early warning system for enterprise credit risk in this study is able to evaluate properly the potential credit risk of the futuristic credit related activities. The proposed mechanism is applied on the dataset and it shows good results in terms of sample fitting performance, complexity, and accuracy. The system can warn in advance regarding the potential credit risks and it has a higher accuracy as compared to the conventional techniques. It introduces a novel technique for early warning of company credit risk, as well as technical assistance to help businesses to increase their business without having fears of losses and competitiveness. With the influence of enterprise credit risk early warning on current social stability factors, the algorithm should have extensive application space. The proposed technique shows higher accuracy and takes minimal time in producing results and the data is also fitted properly. The accuracy obtained by the proposed algorithm proves the viability of the suggested method in this paper. However, the technique presented in this paper is one-sided in its application, and therefore it is not appropriate for all enterprises. As a result, the next study will concentrate on broadening the algorithm's reach.

Data Availability

The data used to support the findings of this study are available on request.

Conflicts of Interest

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

Funding

This study did not receive any funding in any form.

References

- [1] D. Xiao, "International experience for SME credit risks identification: based on Monitoring-Cashflow method," in *Proceedings of the 2011 International Conference on Management and Service Science*, pp. 1–4, Bangkok, Thailand, May 2011.
- [2] M. Wang, J. Yu, and Z. Ji, "Credit risk assessment of high-tech enterprises based on RSNCL-ANN ensemble model," in *Proceedings of the 2018 International Conference on Math and Art Intelligence (ICMAI)*, pp. 73–78, Chongqing, China, March 2018.
- [3] J. Galindo and P. Tamayo, "Credit risk assessment using statistical and machine learning: basic methodology and risk modeling applications," *Computational Economics*, vol. 15, no. 1/2, pp. 107–143, 2000.
- [4] X. Xiao-si, C. Ying, and R. Ruo-en, "Studying on forecasting the enterprise bankruptcy based on SVM," in *Proceedings of the 2006 International Conference on Management Science and Engineering*, pp. 1041–1045, Lille, France, October 2006.
- [5] N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, "Explainable machine learning in credit risk management," *Computational Economics*, vol. 57, no. 1, pp. 203–216, 2021.
- [6] R. S. Kenett and S. Salini, "Modern analysis of customer satisfaction surveys: comparison of models and integrated analysis," *Applied Stochastic Models in Business and Industry*, vol. 27, no. 5, pp. 465–475, 2011.
- [7] M. Kaur and S. Kadam, "Bio-inspired workflow scheduling on HPC platforms," *Tehnički Glasnik*, vol. 15, no. 1, pp. 60–68, 2021.
- [8] A. Ono, R. Hasumi, and H. Hirata, "Differentiated use of small business credit scoring by relationship lenders and transactional lenders: evidence from firm-bank matched data in Japan," *Journal of Banking and Finance*, vol. 42, pp. 371–380, 2014.
- [9] L. Zhang, H. Hu, and D. Zhang, "A credit risk assessment model based on SVM for small and medium enterprises in supply chain finance," *Financial Innovation*, vol. 1, no. 14, 2015.
- [10] M. Kaur, "FastPGA based scheduling of dependent tasks in grid computing to provide QoS to grid users," in *Proceedings of the 2016 International Conference on Internet of Things and Applications (IOTA)*, pp. 418–423, Pune, India, January 2016.
- [11] N. Yoshino and F. Taghizadeh-Hesary, "Analysis of credit ratings for small and medium-sized enterprises: evidence from Asia," *Asian Development Review*, vol. 32, no. 2, pp. 18–37, 2015.
- [12] H. S. Bhamra, C. Dorion, A. Jeanneret, and M. Weber, "Low inflation: high default risk and high equity valuations," *Social Science Electronic Publishing*, vol. 21, pp. 46–70, 2018.
- [13] T. Madjia, A. Amir-Reza, D. C. Debora, and P. Maryam, "An artificial neural network and Bayesian network model for liquidity risk assessment in banking," *Neurocomputing*, vol. 275, pp. 2525–2554, 2018.
- [14] K. W. Li, "Analyzing the TFP performance of Chinese industrial enterprises," *The Singapore Economic Review*, vol. 63, pp. 194–217, 2018.
- [15] C. Zhao-Quan, C. Guang-Cai, X. Li-Ning, Y. Jing-Hui, and T. Xu, "Evaluating hedge fund downside risk using a multi-objective neural network," *Journal of Visual Communication and Image Representation*, vol. 59, pp. 433–438, 2019.
- [16] L. Jiang, S. R. Sakhare, and M. Kaur, "Impact of industrial 4.0 on environment along with correlation between economic growth and carbon emissions," in *International Journal of System Assurance Engineering and Management*, Springer Science and Business Media LLC, Berlin, Germany, 2021.
- [17] S. Yu, G. Chi, and X. Jiang, "Credit rating system for small businesses using the K-S test to select an indicator system," *Management Decision*, vol. 57, no. 1, pp. 229–247, 2019.
- [18] L. Chen, V. Jagota, and A. Kumar, "Research on optimization of scientific research performance management based on BP neural network," *International Journal of System Assurance Engineering and Management*, 2021.
- [19] K. Mahajan, U. Garg, and M. Shabaz, "CPIDM: a clustering-based profound iterating deep learning model for HSI segmentation," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 7279260, 12 pages, 2021.
- [20] A. Fronzetti Colladon and E. Remondi, "Using social network analysis to prevent money laundering," *Expert Systems with Applications*, vol. 67, pp. 49–58, 2017.
- [21] J. Gupta, N. Wilson, A. Gregoriou, and J. Healy, "The effect of internationalisation on modelling credit risk for SMEs:

- evidence from UK market,” *Journal of International Financial Markets, Institutions and Money*, vol. 31, no. 1, pp. 397–413, 2014.
- [22] A. Belhadi, S. S. Kamble, V. Mani, I. Benkhati, and F. E. Touriki, “An ensemble machine learning approach for forecasting credit risk of agricultural SMEs’ investments in agriculture 4.0 through supply chain finance,” in *Annals of Operations Research*, Springer Science and Business Media LLC, Berlin, Germany, 2021.
- [23] S. A. Lundqvist and A. Vilhelmsson, “Enterprise risk management and default risk: evidence from the banking industry,” *Journal of Risk and Insurance*, vol. 85, no. 1, pp. 127–157, 2018.
- [24] C. Yan, X. Fu, W. Wu, S. Lu, and J. Wu, “Neural network based relation extraction of enterprises in credit risk management,” in *Proceedings of the 2019 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 1–6, Kyoto, Japan, March 2019.
- [25] G. Wang and J. Ma, “A hybrid ensemble approach for enterprise credit risk assessment based on support vector machine,” *Expert Systems with Applications*, vol. 39, no. 5, pp. 5325–5331, 2012.