Credit Decision of Small and Micro-sized Businesses Based on Big Data Analysis

Wenjing Wang[1,2,*]

[1]North-Chiang Mai University,Chiang Mai,50000,Thailand
[2]Department of Basic Courses,Shandong Institute of Commerce&Technology , Jinan, 250103, Shandong, China
g636501002@northcm.ac.th

Abstract
In this paper, we select big data information such as credit risk, credit policy and transaction bill information of small and micro businesses , screen enterprises that do not meet the evaluation requirements, establish credit indicators, and evaluate the strength of enterprises. Based on the big data analysis, this paper establishes a correlation regression analysis model, evaluates the relationship between credit and transaction information of enterprises, and gives the optimal credit decision scheme model. By collecting the impact of COVID-19 epidemic on different types of enterprises, the small and medium-sized enterprises selected as the research object in this paper are divided into six categories, and their different impact indexes are obtained,Then, 302 enterprises are classified according to their size, and their resistance to sudden factors is established,The possible sudden factors are analyzed, and the corresponding adjustment suggestions for their credit strategies in the post-epidemic era are given.

Keywords:*Small And Micro Sized Businesses, Big Data Analysis, Credit Decision, Risk AssessmentModel, Post-Epidemic Era*

1.INTRODUCTION

In recent years, small and medium-sized enterprises have gradually developed and become an indispensable part of national economic development. Therefore, in order to ensure the normal operation of enterprises and provide them with funds, banks should issue loans to them. According to the credit risk of enterprises [2], banks determine whether to lend, and decide their credit strategies such as loan amount, interest rate and term according to the clear credit policy and transaction bill information of lending enterprises, and give preferential interest rates to enterprises with high reputation and low credit risk.

For the bank, how to evaluate the enterprise credit risk through the enterprise information, input invoice and output invoice information data, and formulate the enterprise credit strategy when the annual total credit is fixed.

Due to the differences in production and operation products and scale, enterprises will have economic fluctuations due to some unexpected factors, and their business operations and economic benefits will be impacted. In 2020, for example, the outbreak of the COVID-19 epidemic has made it possible to analyze the blows that different industries and different types of enterprises will suffer, and to analyze different influencing factors. Through the credit risk of small and medium-sized enterprises and the impact of unexpected factors on each enterprise, this paper gives the bank-to-enterprise credit adjustment strategy [3].

2.CREDIT RISK ASSESSMENT

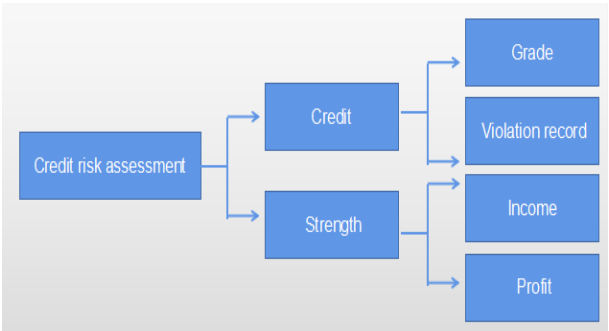

Figure 1. Credit risk assessment hierarchy

A. El-Hashash et al. (Eds.): IEIT 2022, ACSR 100, pp. 501–506, 2023.
https://doi.org/10.2991/978-94-6463-058-9_82

2.1.Enterprise reputation

The reputation of an enterprise should be judged according to its reputation rating and violation records.

The existing data are the credit ratings of small and medium-sized enterprises, which are A, B, C and D, and are assigned by the membership function of large Cauchy distribution.

$$f(x)=\begin{cases}\left[1+\alpha(x-\beta)^{-2}\right]^{-1}, 1\le x\le 3\\ a\ln x+b, 3\le x\le 5\end{cases} \tag{1}$$

undetermined coefficient: a、b、α、β.

$f(5)=1$, When the credit rating is "good", the grade is 1, that is, When the grade is "pass", the grade is 0.8, When the grade is "poor", the grade is 0.01,

$\alpha=1.1086$

$\beta=0.8942$

$a=0.3915$

$b=0.3699$

A,B,C,D,correspond, $\{1 \quad 0.9126 \quad 0.8 \quad 0.5245\}$.

It can be concluded that among the 123 enterprises, the credit rating of A is 1, that of B is 0.9126, that of C is 0.8, and that of D is 0.5245. The higher the score of enterprise reputation grade, the higher the reputation.

Then, sort out the known default situations. Excel is used to screen out the default situations of enterprises with various credit grades [6], and then the "Yes" and "No" in the default are assigned, and the range standard method is used for standardization.

There are two cases of breach of contract: [Yes, No], which can be set as 0,1. As shown in the following table:

Table 1. Default assignment

Break a contract	yes	No
Evaluation	1	0

Standardization of credit rating:

$$A_j'=\frac{A_j-\min\limits_{1\le j\le 123}A_j}{\max\limits_{1\le j\le 123}A_j-\min\limits_{1\le j\le 123}A_j}=\frac{A_j-0.5245}{1-0.5245},(j=1,2,3\Lambda,123) \tag{2}$$

Standardize "Yes, No" of default:

$$B_j'=\frac{B_j-\min\limits_{1\le j\le 123}B_j}{\max\limits_{1\le j\le 123}B_j-\min\limits_{1\le j\le 123}B_j}=\frac{B_j-0}{1-0},(j=1,2,3\Lambda,123) \tag{3}$$

Enterprise reputation is the sum of credit rating and default,Because the credit rating, i.e., the proportion of default, is unknown, constant C(0≤c≤1) is used to express the credit rating and the criteria for judging whether or not to default. The reputation of different enterprises can be expressed as:

$$C_j=cA_j'+(1-c)B_j',(0\le c\le 1, j=1,2,3\Lambda,123) \tag{4}$$

Assuming that the weight c=0.6, the total score of enterprise reputation can be calculated. As shown in Table 2:

Table 2. Total score of enterprise reputation

Enterprise	E1	E2	E3	E4	E5	E6	E7	E8	E9
Grade	1	1	0.8	0.8	0.913	1	1	1	1
Standardization	1	1	0.579	0.579	0.816	1	1	1	1
Break a contract	1	1	1	1	1	1	1	1	1
Total	1	1	0.748	0.748	0.89	1	1	1	1
Enterprise	E10	E11	E12	E13	E14	E15	E16	E17	E18
Grade	1	1	1	1	0.913	0.913	1	0.913	1
Standardization	1	1	1	1	0.816	0.816	1	0.816	1
Break a contract	1	1	1	1	1	1	1	1	1
Total	1	1	1	1	0.89	0.89	1	0.89	1

3.ENTERPRISE STRENGTH

Enterprise strength, from the input invoice and output invoice, the total input and output of each enterprise can be calculated [7].

Because the total amount of input and output can be calculated and the total profit can be calculated through the information of input invoice and output invoice in Annex 1, the output amount and total profit of an enterprise may be regarded as two indicators to judge the strength of an enterprise.

Firstly, the data of two maximal indexes are standardized.

$$X_j=\frac{x_j-\min\limits_{1\le j\le 123}x_j}{\max\limits_{1\le j\le 123}x_j-\min\limits_{1\le j\le 123}x_j}=\frac{x_j-14300}{4698633440-14300},(j=1,2,3,\Lambda,123) \tag{5}$$

$$Y_j = \frac{y_j - \min\limits_{1\le j\le 123} y_j}{\max\limits_{1\le j\le 123} y_j - \min\limits_{1\le j\le 123} y_j} = \frac{y_j - (-1939308588)}{1891956370 - 1939308588}, (j = 1,2,3,\Lambda,123) \tag{6}$$

The strength of different enterprises can be expressed as:

$$D_j = dX_j + (1-d)Y_j, (0 \le d \le 1, j = 1,2,3,\Lambda,123) \tag{7}$$

It is assumed that under the condition that the invoicing information of input and output items given in Appendix 1 is accurate. Output amount represents the purchase and sales frequency of an enterprise in a period of time, so the larger the output amount of an enterprise, the larger the scale of the enterprise, so it may be assumed that the output amount accounts for a larger proportion of the strength of the enterprise than the profit. If d=0.6, the comprehensive strength of different enterprises can be calculated.

After weighting the output amount and total profit, the total score of comprehensive strength of 123 enterprises is obtained [8].

Add the obtained total credit score and total strength score to establish a credit risk model:

$$Z_j = \frac{1}{2}C_j + \frac{1}{2}D_j, (j = 1,2,3\Lambda,123) \tag{8}$$

The total credit risk assessment of enterprises is the sum of credit score and strength assessment. In the credit risk model, the higher the credit value, the higher the credit, the higher the total score of strength and the stronger the strength.

Therefore, it can be concluded that the greater the z value, the lower the risk.

4.BANK CREDIT DECISION-MAKING

Banks lend to enterprises with strong strength and stable relationship between supply and demand. Therefore, before making a credit policy, it is necessary to select and remove the enterprises that will not lend according to their strength and the relationship between supply and demand, and then carry out policy analysis [4].

First of all, according to the principle that banks do not lend to enterprises with credit rating of D, enterprises with credit rating of D are screened out

Then, it analyzes the stability of the relationship between supply and demand of enterprises. A sales bill is an invoice issued by an enterprise for the buyer when selling products. Input bill is an invoice issued by the seller when an enterprise purchases goods. Therefore, it can be seen from the ratio of total output bills to total input bills whether the business operation is stable or not. When the ratio is equal to 1, it means that the amount of sales in recent years is offset, and when it is near 1, it means that the amount of sales in recent years is almost offset. If the ratio is far lower than 1 and close to 0, it means that the sales volume is far lower than the input amount, and the relationship between supply and demand is unstable [5]. Therefore, we might as well assume that the ratio of the amount of output notes to the amount of input notes is more than 0.90, which means that the relationship between supply and demand is stable. If the number is too large or too small, it can be removed.

Table 3. Suspicious data of unstable relationship between supply and demand

Enterprise	Income	Output	Output/Income
E2	162487965	62629	0.00038543
E17	142782019	299622	0.00209845
E12	95447012	923965	0.00968039
E16	292748	213556685	729.489817
E104	280	270563	966.296428
E69	3250	3748394	1153.35
E95	1138	2380080	2091.458699
E42	11769	26235453	2229.19984
E68	740	5957078	8050.10540

The following functions can be defined:

$$x_G(x) = \begin{cases} 1, x \in G; \\ 0, x \notin G. \end{cases} \tag{9}$$

Lending enterprises are "1", and non-lending enterprises are "0".

As the bank's annual total credit is fixed, no specific value is given. Therefore, it is impossible to specify the specific amount when determining the loan amount of enterprises. Therefore, it is advisable to set the bank's annual total credit as the unit "1", and determine the loan amount according to the ratio of the output amount of each enterprise to the output amount of all enterprises.

Using the ratio of output amount and credit risk as the credit risk model, the output amount, that is, the loan amount, is the credit risk.With matlab software.

$$y_1 = -0.0092 + 0.0411x \tag{10}$$

According to the data integration analysis of loan annual interest rate and customer churn rate, the higher the bank loan annual interest rate, the greater the customer churn rate. That is, the greater the bank's benefits, the greater the customer churn rate. Moreover, in the lost enterprises, when the loan annual interest rate

is fixed, the loss rate of customers with high credit rating is greater. Because the relationship between the three credit ratings and the annual interest rate of loans is equivalent, and there is not much difference between them, a balance can be struck between the interests of banks and the churn rate of the three rated customers [1].

The analysis shows that there is a positive curve relationship between customer churn rate and loan annual interest rate, and there is little difference among enterprises with credit rating of A, B and C,Therefore, the relationship model between customer churn rate and loan annual interest rate can be established by taking the average value of three grades of churn rate:

$$y_2 = ax^2 + bx + c \tag{11}$$

Calculate with MATLAB software, substitute data, and get:

$$a = 0.1393, b = -0.0123, c = 0.0446 \tag{12}$$

Get: $$y_2 = -0.0123x^2 + 0.0446x + 0.1393 \tag{13}$$

$$y_1 = \begin{cases} 10, (x_1 > 0.000165) \\ -0.0092 + 0.0411x_1, (0.000165 \le x_1 \le 0.15899) \\ 100, (x_1 < 0.15899) \end{cases} \tag{14}$$

$$y_2 = \begin{cases} 4\%, (x_2 > 0.2) \\ 0.1393 - 0.0123x_2^2 + 0.0446x, (0.2 \le x_2 \le 0.8) \\ 15\%, (x_2 < 0.8) \end{cases} \tag{15}$$

y_2 X is the relationship model between customer churn rate and loan annual rate,It can be seen from this model that when customer churn rate is greater than 0.20, the loan annual rate is 4%,When the customer churn rate is less than 0.8, the loan annual interest rate is 15%.

The value of preferential interest rate obtained by the enterprise may be set as the sum of credit risk and credit score, and then the preferential interest rate is obtained according to the preferential parameters,If the preferential parameters are greater than 1, it means that the enterprise has small credit risk and strong strength, so it is given preferential treatment; Enterprises with preferential parameters less than 1 are weak and do not meet the lower limit of the lowest preferential interest rate, so preferential interest rates will not be granted.

Establish a linear regression model, solve it with MATLAB, and get:

$$y = -1.8609 + 1.5309x \tag{16}$$

Calculate the preferential interest rate of each enterprise according to the model.

y_1 X is the relationship model between credit risk and credit risk,It can be seen from this model that when credit risk is greater than 0.000165, the credit risk is 1 million,When the credit risk is less than 0.15899, the output amount is 1 million.

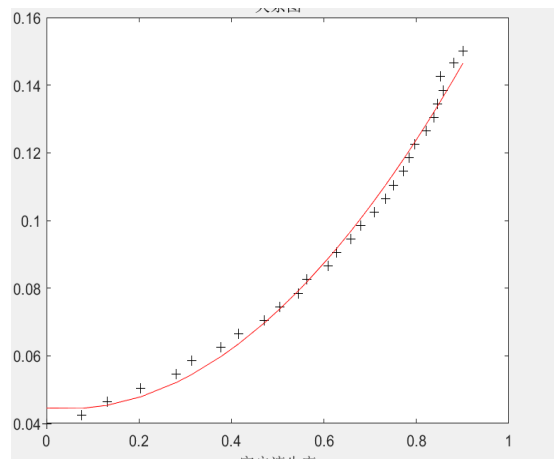

Figure 2. Relationship between customer churn rae and loan annual interest rate

5.HOW TO IMPROVE CREDIT DECISION-MAKING IN THE POST-EPIDEMIC ERA

In different development periods, enterprises will suffer losses in different degrees due to sudden factors. For example, during the Spring Festival this year, enterprises will face great challenges.

Data show that the enterprises most seriously affected by the epidemic are hotels, restaurants, tourism, intermediary services, building materials and real estate industries. During the shutdown period, enterprises generally face relatively large comprehensive cost pressures such as rent, wages, taxes and fees; The pressure of resignation is stable, but the pressure of layoffs is greater; The price of raw materials has generally increased; Problems such as insufficient inventory and difficulty in finding alternative suppliers.

Impact of epidemic situation on different enterprises. Tourism enterprises are most affected by this epidemic, and the whole tourism industry has been affected. Travel agencies suffered a total loss, small and medium-sized travel agencies faced closure, hotels suffered suspension of business, and flights were cancelled,Nowadays, tourism has suffered, and scenic

spots have been required to be closed completely, thus unable to generate income. Automobile enterprises are faced with the shortage of parts supply and the interruption of automobile supply chain. Operation of network car is limited.

Due to the different scope and ways of business operation, production and sales of enterprises in different industries, sudden factors have different influences on different industries.

The enterprises in the data are divided into six categories: catering, manufacturing, science and technology, culture, medical care and logistics. Establish an impact model:

$$y_4 = 1.87x_1 + 1.48x_2 + 1.16x_3 + 0.71x_4 + 0.46x_5 + 0.34x_6 \quad (17)$$

Different types of enterprises have different ways of operation, decision-making and income,Generally speaking, enterprises with large scale of operation are more obviously affected by the butterfly effect of sudden factors, and the investment and operation scale of self-employed households are relatively the smallest, and the degree of sudden factors is also the smallest; The investment and operation scale of limited liability company is relatively large, and the operation scale of limited company is relatively the largest, which is also affected by unexpected factors.

Therefore, 302 enterprises are divided into three categories: limited company, limited liability company and self-employed households, and a size model of resistance ability of different types of enterprises to unexpected factors is established.

As follows:

$$y_5 = 1.29x_1 + 1.29x_2 + 0.42x_3 \quad (18)$$

Then, the possible unexpected factors are analyzed, which are mainly considered according to the unexpected factors that often appear in the historical and big data analysis: epidemic situation, public opinion and improper management of senior executives.

The influence of unexpected factors on the production, operation and economic benefits of each enterprise is obtained.

Establish pairwise comparison matrix according to three indexes:

$$y_3 = 0.14x_1 + 0.39x_2 + 0.47x_3 \quad (19)$$

Finally, according to the resistance of enterprises in different industries to sudden factors, the resistance of different types of enterprises to sudden factors, and the impact of sudden factors on the production, operation and economic benefits of enterprises, the credit strategy is adjusted accordingly.

6.CONCLUSION

The model can be used flexibly, and a new model can be established for one's own use,However, due to the huge data, the accuracy of each quantity can not be guaranteed in the data quantification process, and there will be certain errors. There is subjectivity in establishing the model to a certain extent, and it is difficult to find out the relationship among the factors affecting strength, interest rate and loan amount. The multi-factor analysis model is used to improve, and the four factors of credit, strength, loan amount and interest rate are analyzed by multi-factors. And get the relationship between them.

REFERENCES

[1] Bo Huang, Qing-Pu Zhang and Yun-Quan Hu, "Research on credit risk management of the state-owned commercial bank," [J] 2005 International Conference on Machine Learning and Cybernetics, 2005, pp. 4038-4043 Vol. 7.

[2] Han Zhonggeng, Practical Course of Mathematical Modeling, [M], Higher Education Press, 2020.9

[3] KHAN, K, ZHAO, Huawei, ZHANG, Han, et al.The Impact of COVID-19 Pandemic on Stock Markets: An Empirical Analysis of World Major Stock Indices [J] The Journal of Asian Finance, Economics and Business, 2020.(7) 463-474

[4] Qinghua He,Credit risk analysis and preventive measures of large and medium-sized enterprises in He Qinghua Commercial Bank [J] Modern Economic Information, 2019(12) 369-376

[5] Rui Xie, Rui Liu, Xian-Bei Liu, et al.Evaluation of SMEs' Credit Decision Based on Support Vector Machine-Logistics Regression [J] Journal of Mathematics, vol. 2021, Article ID 5541436, 10 pages, 2021. https://doi.org/10.1155/2021/5541436.

[6] Weasion Wei, Gu Li, Ran Qingli, Thoughts on the Recruitment of Civil Servants, [J] Journal of Engineering Mathematics, 2004.21(z1) 137-141

[7] Zhou, L., Wu, K., Liu, H. et al. CIRD-F: Spread and Influence of COVID-19 in China, [J] Shanghai Jiaotong Univ. (Sci.) 2020.(15) 147-156

[8] Zhang Renyu, Wu Guoqiang, Economics under Butterfly Effect, [J] Research on Commodity and Quality Theory, 2011(6)24-25.