Improve the Loan Pricing Model Based on the Calculation and Analysis of the Credit Decision Data of Small, Medium and Micro Enterprises

Yang Xu[1](✉) and Zhitao Zhuang[2]

[1] School of Information Science, Beijing Forestry University, Beijing 100083, China
xuyang_2018@bjfu.edu.cn

[2] School of Economics and Management, Beijing Forestry University, Beijing 100083, China

Abstract. In recent years, small, medium and micro enterprises have gradually shown an increasingly important position in my country's economic development. However, in reality, due to their relatively small scale and lack of mortgage assets, the establishment of a credit risk evaluation system for small, medium and micro enterprises has become a bank. Urgent problems. This paper first establishes a credit risk model and a bank's profit maximization model, uses the TOPSIS method to score 123 small, medium and micro enterprises with credit records in a bank, and then introduces the RAROC model loan pricing model to construct the bank's profit maximization model, using xgboost. The classification model predicts the credit rating of the 302 small, medium and micro enterprises that the bank has no credit history, and solves the optimal credit strategy. Finally, taking the COVID-19 as an example, the impact of the epidemic on each industry is quantified, and then it is put into the credit risk model as a sudden risk factor multiplier to measure the impact of sudden risk on the full sample, and then calculate and adjust the new credit strategy. This article deeply analyzes the credit decision-making mechanism of small, medium and micro enterprises, and has corresponding reference value.

Keywords: TOPSIS · Credit Risk · Improved RAROC Model · Xgboost

1 Introduction

In recent years, small, medium and micro enterprises have gradually shown an increasingly important position in my country's economic development. In the "Guiding Opinions of the People's Bank of China on Further Strengthening Financial Services for Small, Medium and Micro Enterprises (Yinfa [2020] No. 120)" [13] issued in May 2020, it is clearly stated that actively encourage and implement the resumption of work and production of small, medium and micro enterprises Credit support policy.

Compared with large enterprises, the financing methods of small, medium and micro enterprises are relatively simple and rely heavily on bank loans [7]. However, due to the small scale of small, medium and micro enterprises and lack of mortgage assets, banks are

N. Radojević et al. (Eds.): ICAID 2022, AHIS 7, pp. 492–510, 2023.
https://doi.org/10.2991/978-94-6463-010-7_52

cautious about this type of business. Therefore, banks tend to provide loans to enterprises with strong strength and high operational stability through existing information such as credit policies and existing transaction information of enterprises, and at the same time give certain interest rate concessions to enterprises with high reputation and low credit risk. The bank first evaluates the ability of small, medium and micro enterprises to repay the funds and the existing credit risks and other factors, based on which to determine whether to lend, and then further determine the specific loan quota, interest rate and loan period and other credit strategies.

In foreign research, Fengzhi Zhang [3] proposed that big data should be used to solve the problems of SMEs in defining property rights and hierarchical privacy, so as to meet their financing needs. Foluso Olugbenga Aribaba [4] and others used cross-sectional survey research methods to study the role of deposit bank loan instruments in providing funds for Nigerian SMEs. Liu Xing-yu [9] and others deeply analyzed the financing status of China's small and medium-sized enterprises, the reasons for financing difficulties and the countermeasures to solve the financing difficulties of small and medium-sized enterprises.

Most of the existing domestic research focuses on how to evaluate the credit risk index system and the relationship between banks and enterprises. Sun Yuchen [12] built a logistic regression default rate measurement model to calculate the expected default rate of small, medium and micro enterprises under different credit levels. Xue Huayi [15] and others constructed a restricted dependent variable model and a multiple logistic regression model to conduct an empirical analysis on the pricing of credit assets of small, medium and micro enterprises. Xu Lihe [14] based on the survey data of enterprises in Guangdong Province in 2018, classified corporate financing channels into three categories: formal credit, informal credit, and supply chain credit financing, and compared and analyzed the differences in the impact of the three on the multi-dimensional innovation of enterprises. Mechanism. Guo Na [6] and others cut from the perspective of bank-enterprise relationship, using questionnaire survey data on the financing status of SMEs to study the issue of bank-enterprise relationship and SME credit constraints.

In response to the above-mentioned problems, this article first quantitatively analyzes the credit risk of 123 companies with credit records of a certain bank, and gives the bank's credit strategy for these companies when the total annual credit is fixed; and then the credit of 302 companies without credit records Quantitative analysis of risks is carried out, and the bank's credit strategy for these companies when the total annual credit is 100 million yuan is given; finally, considering that the production and operation and economic benefits of different companies may be affected differently by some unexpected factors, Give the bank's credit adjustment strategy when the total annual credit is 100 million yuan.

2 Indicator Symbol and Description

2.1 Data Sources

The data in this article comes from the relevant data of 123 companies with credit records in a bank in question C of the 2020 National College Students Mathematical Modeling

Contest (http://www.mcm.edu.cn/). There are a total of 210947 input invoice information records and 162484 output invoice information records in 123 small, medium and micro enterprises. Among them, the company information includes company code, company name, reputation rating, and whether it is in breach of contract; input invoice information includes company code, invoice number, invoice date, seller unit code, amount, tax, total price and tax, and invoice status; output invoice information Including company code, invoice number, invoice date, purchaser unit code, amount, tax, total price and tax, and invoice status.

2.2 Indicator Symbol and Description

The symbols and corresponding indicators used in this study are shown in Table 1.

Table 1. Formatting sections, subsections and subsubsections.

Symbol	Description
X_1	Input amount
X_2	Input tax
X_3	Total input price tax
X_4	Output amount
X_5	Output tax
X_6	Total output price tax
X_7	Tax payable = output tax-input tax
X_8	Number of input = number of input invoices of a single company
X_9	Number of output = number of invoices output by a single company
X_{10}	Number of voided invoices
X_{11}	Negative invoice count
Industry	Industry Scoring
C	Reputation level (A, B, C, D)
S	Risk prediction score
Q	Annual total credit
r	Credit rate
l	Customer Churn Rate
elp	Expected loss rate
ulp	Unexpected loss rate
k	Customer retention factor
T	Sudden risk factor

3 Establish a Credit Risk Model and a Bank's Profit Maximization Model

First of all, this article constructs risk impact indicators by processing the given enterprise information (industry category, reputation level, default, etc.) and enterprise bill transaction data (invoicing and sales amount, tax, number of invoicing and sales transactions, etc.). Then use principal component analysis to reduce the dimensionality, retain the influential indicators, and then use the TOPSIS method to score each company. This article sets a credit risk threshold. Once it falls below this threshold, this article recommends that banks not provide loans for them. Then, this paper introduces the improved RAROC model loan pricing model to construct the bank's profit maximization model, quantifies the expected future loss caused by the risk as the current cost, and then optimizes the entire bank's profit maximization model. Finally, considering the long-term development of the bank and customer retention, this study sets the minimum retention rate for customers with different credit ratings, and further optimizes and adjusts the bank's revenue as a constraint, so as to solve the optimal credit strategy for these 123 companies.

3.1 Model Establishment

1) *Credit Risk Model.* We first use the principal component analysis method (PCA) to reduce the dimension of the impact indicators that affect credit risk constructed in the previous article, and then use the TOPSIS method to predict the risk score of each company.

Principal component analysis method for dimensionality reduction.

Suppose the n-time observation data matrix of the original variable $X_1, X_2, X_3 ... X_p$ is:

$$X = \begin{bmatrix} x_{11} \; x_{12} \; ... \; x_{1p} \\ x_{21} \; x_{22} \; ... \; x_{2p} \\ x_{31} \; x_{32} \; ... \; x_{3p} \end{bmatrix} = \begin{bmatrix} X_1 \; X_2 \; ... \; X_p \end{bmatrix} \tag{1}$$

The data matrix is standardized by the center of the column, the standardized matrix is recorded as X.

Calculate the correlation coefficient matrix R, $R = \left(r_{ij}\right)_{p \times p}$

$$r_{ij} = \frac{\sum_{k=1}^{n} \left(x_{kj} - \overline{x_i}\right)\left(x_{kj} - \overline{x_j}\right)}{\sqrt{\sum_{k=1}^{n} \left(x_{kj} - \overline{x_i}\right)^2 \left(x_{kj} - \overline{x_j}\right)^2}}, r_{ij} = r_{ji}, r_{ii} = 1 \tag{2}$$

Calculate the characteristic equation of R:

$$\det(R - \lambda E) = 0, \lambda_1 \geq \lambda_2 \geq \lambda_p > 0 \tag{3}$$

Determine the number of principal components:

$$\frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{p} \lambda_i} \geq \alpha \tag{4}$$

α is the cumulative contribution rate, generally greater than or equal to 80%
Calculate m corresponding unit eigenvectors:

$$\beta_1 = \begin{bmatrix} \beta_{11} \\ \beta_{21} \\ \cdots \\ \beta_{p1} \end{bmatrix}, \beta_2 = \begin{bmatrix} \beta_{12} \\ \beta_{22} \\ \cdots \\ \beta_{p2} \end{bmatrix}, \ldots\ldots \beta_m = \begin{bmatrix} \beta_{1m} \\ \beta_{2m} \\ \cdots \\ \beta_{pm} \end{bmatrix} \tag{5}$$

Calculate the principal components:

$$Z_i = \beta_{1i}X_1 + \beta_{2i}X_2 + \ldots\ldots + \beta_{pi}X_p, i = 1, 2, \ldots, m \tag{6}$$

TOPSIS risk scoring.

Select evaluation indicators. Assuming another evaluation object and evaluation indicators, the original data matrix can be expressed as:

$$X = \left(X_{ij}\right)_{n\times m} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1m} \\ x_{21} & x_{22} & \ldots & x_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ x_{n1} & x_{n2} & \ldots & x_{nm} \end{bmatrix}_{n\times m}, i = 1, 2, \ldots, n, j = 1, 2, \ldots, m \tag{7}$$

Normalize the original data, and the normalized data constructs a standardized decision matrix.

Determine ideal solution X^+ and negative ideal solution X^-:

$$X^+ = (X_{\max 1}, X_{\max 2}, \ldots, X_{\max m}), X^- = (X_{\min 1}, X_{\min 2}, \ldots, X_{\min m}) \tag{8}$$

Calculate the distance between the i-th evaluation object and the positive and negative ideal solution:

$$D_i^+ = \sqrt{\sum_{j=1}^{m} \left(D_{\max j} - D_{ij}\right)^2}, D_i^- = \sqrt{\sum_{j=1}^{m} \left(D_{\min j} - D_{ij}\right)^2} \tag{9}$$

Calculate the relative closeness of the evaluation object to the ideal solution:

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-}, i = 1, 2, \ldots, n \tag{10}$$

Sort according to the size of S_i. The larger the value of S_i, the closer the unit is to the ideal state, that is, the lower the risk and the better the evaluation effect. We set a risk threshold, companies above this threshold, we choose to give loans, companies below this threshold, we believe that the credit risk is too high, so we recommend that banks not handle credit business.

- Determination of loan amount
  For a certain annual total credit amount, we consider the risk prediction score obtained by the credit risk model to calculate the loan quota allocation of the lending company. Among the companies that can obtain bank credit, the risk prediction score calculated by the credit model reflects the company's exposure risk and the optimal loan amount to a certain extent. Therefore, in the n companies that have passed the risk threshold screening, we set the risk score of each company as S_i, and the total annual bank credit is Q. Therefore, the loan amount Q_i allocated by each company is calculated by the following company.

$$Q_i = \frac{S_i}{\sum_{i=1}^{n} S_i} \cdot Q \tag{11}$$

- *Credit Risk Model.* Since the 1990s, the international banking community has begun to use risk-adjusted return on capital (RAROC) as an important indicator of bank performance. RAROC is a profitability indicator that fully considers risks. After deduction, the expected loss (similar to the current cost) is removed from the income of the numerator, and the denominator is economic capital, which is the unexpected loss, the ratio of the two is the real return on capital that the bank ultimately pursues [8].

From the perspective of risk management, RAROC allocates economic capital to each loan, which cushions the unexpected losses caused by unexpected risks to the bank. From the perspective of performance appraisal, RAROC records expected losses as current costs and considers unexpected risk factors, which can measure the true operating performance and risks of commercial banks in the future for a long time [11].

Due to the relatively small scale of small, medium and micro enterprises and few asset mortgages, there are certain default risks and unexpected risks. Therefore, we learn from the RAROC model idea, and introduce expected losses and unexpected losses into the bank's profit maximization model when solving the bank's profit maximization, so as to improve the authenticity and accuracy of the bank's optimal credit strategy.

The traditional RAROC model is shown in Eq. 12:

$$\begin{aligned} RAROC &= \frac{Risk - adjusted\ return}{Economic\ capital} \\ &= \frac{Total\ Revenue\text{-}Capital\ Cost\text{-}Operating\ Cost\text{-}Expected\ Loss}{Economic\ capital} \end{aligned} \tag{12}$$

According to the specific circumstances of this article, we do not consider bank capital costs and operating costs. The total income is the total bank loan interest income, the calculation of expected losses and economic capital (unexpected losses). We obtain coverage of manufacturing and real estate by consulting existing literature. The expected loss rate and unexpected loss rate of ten industries including the construction industry and the construction industry [2]. The adjusted model is shown in formula 13:

$$\begin{aligned} & Bank\ Adjusted\ Risk - Return\ Ratio = \frac{Risk - Adjusted\ Return}{Economic\ Capital} \\ & = \frac{Loan\ Income - Expected\ Loss}{Economic\ Capital} \\ & = \sum \frac{Q_i \cdot r \cdot f(r, l) - EL}{EC} \\ & = \sum \frac{Q_i \cdot r \cdot f(r, l) - PD \cdot LGD \cdot EAD}{EC} \\ & = \sum \frac{Q_i \cdot r \cdot f(r, l) - Q_i \cdot elp}{Q_i \cdot ulp} \\ & = \sum \frac{r \cdot f(r, l) - elp}{ulp} \end{aligned} \tag{13}$$

$$f(r, l) = 1 - \ln\left(1 + e^{-l}\right) \tag{14}$$

We assume that there is a certain functional relationship between customer churn rate and interest rate:

$$l = \alpha + \beta_1 \cdot r + \beta_2 \cdot r^2 \tag{15}$$

Incorporate formula 15 into formulas 14 and 13, simplifying to get the final bank income model, as shown in formula 16:

$$R = \sum \frac{r \cdot \left[1 - \ln\left(1 + e^{-\left(\alpha + \beta_1 \cdot r + \beta_2 \cdot r^2\right)}\right)\right] - elp}{ulp} \times Q_i \tag{16}$$

Therefore, the bank's profit maximization model is:

$$\begin{aligned} & MAX\left\{\sum \frac{r \cdot \left[1 - \ln\left(1 + e^{-\left(\alpha + \beta_1 \cdot r + \beta_2 \cdot r^2\right)}\right)\right] - elp}{ulp} \times Q_i\right\} \\ & s.t \begin{cases} 10 \le Q_i = \dfrac{S_i}{\sum\limits_{i=1}^{n} S_i} \cdot Q \le 100 \\ 4\% \le r \le 15\% \end{cases} \end{aligned} \tag{17}$$

3) *Interest rate adjustment constraints-retain customers.* When we solve the problem of maximizing bank revenue, we only consider the immediate impact of customer churn rate and incorporate it into the model. Although the bank revenue obtained at this time is the largest, it does not consider the issue of customer retention from a long-term perspective. When handling credit business for small, medium and micro enterprises, we hope to keep more high-reputation and high-strength customers to

protect the bank's future operating income. Therefore, we introduce an interest rate adjustment model. For customers with different credit levels, we set the corresponding target customer retention rate is added as a constraint to the model for optimization. Although the result obtained at this time is not the maximum value of the bank's income statement target, it not only keeps the bank's revenue at a high level, but also guarantees customer retention to a certain extent.

$$\begin{cases} r_i \cdot k_A \cdot S_i \leq l_i \\ r_i \cdot k_B \cdot S_i \leq l_i \\ r_i \cdot k_C \cdot S_i \leq l_i \end{cases} \tag{18}$$

3.2 Model Solution

1) *Credit risk model solution*

Principal component analysis and dimensionality reduction.

For small, medium and micro enterprises, the company's scale and the stability of the supply-sale relationship have a crucial impact on the company's ability to resist risks. Therefore, we initially screened a total of 19 indicators in two dimensions of company size and company stability for principal component analysis (Table 2).

Import all variables into SPSS for principal component analysis. First, perform KMO test. The test results show that further principal component analysis can be performed.

For the dimensions of company scale and the dimensions of supply and marketing stability, three principal components were finally determined, and dimensionality reduction was successfully achieved. We rename the newly generated variables to F1, F2, F3, F4, F5, F6 (Table 3).

- TOPSIS method risk scoring solution
  In order to measure the credit risk of each company, we mainly consider four factors: company size, stability of supply and marketing relations, industry vitality, and credit rating. We obtained six indicators to measure the company's size and the stability of the supply and marketing relationship through the principal component analysis in the previous article. For the industry vitality and reputation level, we quantified it through the method of rating and integrated it into the model for solution.
- Industry vitality quantification
  Based on company information, we classify all sample companies by industry, judge the strength and development prospects of the industry according to the transaction

Table 2. Principal component analysis indicators

Company size dimension	Dimensions of Supply and Marketing Stability
Max(X_1), Max(X_2), Max(X_4), Max(X_5), Mean(X_1), Mean(X_2), Mean(X_4), Mean(X_5), X_7, X_8, X_9	Var(X_1), Var(X_2), Var(X_3), Var(X_4), Var(X_5), Var(X_1), Max(X_8), Max(X_9)

Table 3. Dimensionality reduction results of principal component analysis

F1	F2	F3	F4	F5	F6
9.66782	−0.31288	0.57145	10.4204	2.57705	0.30029
−1.40715	−0.41042	7.16688	−0.0297	−0.96369	8.81793
−0.2592	0.3399	4.34307	−0.0792	0.08982	0.0185
3.54711	4.11905	0.10399	1.80958	1.32987	−0.39897
0.64552	2.24855	0.90619	−0.4971	2.88146	−0.13052
−0.27615	1.79763	0.78899	−1.2061	3.87116	0.44292
−0.58142	1.04521	2.533	−0.282	1.5344	0.77001
0.29191	−0.61243	4.48509	−0.0262	−0.60886	3.65468
−0.0308	−0.40412	1.21083	−0.0194	−0.33568	−0.08416
−0.52948	2.18831	−0.33385	−1.2733	3.02556	0.0723
−0.16763	1.11632	−0.04409	−0.3771	0.90069	−0.06491
0.51813	3.16247	−0.69187	−0.3661	1.75717	−0.26718
−0.4159	0.96317	2.46675	−0.092	−0.15961	3.88082

Table 4. Industry Classification Table

Category	
1	Mining Industry
2	Manufacturing
3	Electricity, Gas and Water Production and Supply Industry
4	Construction Industry
5	Transportation, Storage and Postal Industry
6	Information Transmission, Computer Service and Software Industry
7	Wholesale and Retail Trade
8	Real Estate Industry
9	Leasing and Business Services
10	Water Conservancy, Environment and Public Facilities Management Industry

volume of companies in different industries, and incorporate it into the model as one of the impact indicators for evaluating credit risk.

First of all, we classify the company by industry based on the company information given, and divide the company into 10 industries in total (Table 4).

Table 5. Industry Scoring Table

Category		Industry Scoring
1	Mining Industry	11.60059
2	Manufacturing	10.01827
3	Electricity, Gas and Water Production and Supply Industry	8.700444
4	Construction Industry	10.73055
5	Transportation, Storage and Postal Industry	13.53383
6	Information Transmission, Computer Service and Software Industry	8.178417
7	Wholesale and Retail Trade	8.23642
8	Real Estate Industry	11.60059
9	Leasing and Business Services	5.800296
10	Water Conservancy, Environment and Public Facilities Management Industry	11.60059

We use the credit rating of companies included in a single industry as an indicator of the vitality of the industry, and use this to score the industry. We assign scores to each company based on its reputation level.

Next, we calculate the score of each industry according to Eq. 19.

$$Industry = \frac{n_A \cdot 1 + n_B \cdot \frac{2}{3} + n_C \cdot \frac{1}{3} + n_D \cdot 0}{n_A + n_B + n_C + n_D}$$
$$n_i = Number\ of\ companies\ with\ different\ levels\ of\ reputation\ (i = A, B, C, D) \tag{19}$$

We standardized the industry scores to make the total score 1. In order to facilitate visualization, we expanded them all ten times to make the total score 100 points (Table 5).

- Credit rating quantification
  In order to quantify the company's reputation levels, we assign corresponding values to the four types of reputation levels of A, B, C, and D, and then bring them into the model for the next step of solving. The specific assignments are shown in Table 6.
  In summary, we have completed the quantification of indicators for all variables, and then use the TOPSIS method to predict the credit risk, get the credit risk of each company, and finally reserve 92 companies to give loans (See Appendix Table 2 for specific results).

Table 6. Credit Rating Scale (Quantification of Credit Rating)

Credit rating	A	B	C	D
Assigned score	100	75	50	25

Table 7. Credit Strategy

Enterprise code	Percentage of credit line	Credit rate
E1	0.039365335	0.04
E2	0.028538166	0.04
E3	0.016572111	0.04
E4	0.023872933	0.04
E5	0.016983308	0.04

Credit strategy solving
According to the TOPSIS forecast result and the bank's profit maximization model after adjusting the interest rate, we can obtain the ratio $\frac{S_i}{\sum_{i=1}^{n} S_i}$ of each credit line to the total annual credit. As shown in Table 7.

2) Solving the Bank's Profit Maximization
First, we introduce an improved RAROC model to incorporate the expected loss and unexpected loss of the bank into the model. According to the expected loss rate and unexpected loss rate of the industry given in the article by Feng Li (2014) [2], we take it as the known quantity is brought into the model (Table 8).

Then, based on the data in Annex III, we fit the customer churn rate and interest rate of different reputation levels to find that the two satisfy the $l = \alpha + \beta_1 \cdot r + \beta_2 \cdot r^2$ function relationship.

By calculating the fit, we obtained the size of each coefficient and drew the function curve (Fig. 1).

$$\begin{aligned} l_A &= -76.4101 \cdot r^2 + 21.9844 \cdot r - 0.6971 \\ l_B &= -67.9331 \cdot r^2 + 20.2072 \cdot r - 0.6504 \\ l_C &= -63.9422 \cdot r^2 + 19.5693 \cdot r - 0.6393 \end{aligned} \tag{20}$$

After calculating the above results, we use Lingo to solve the bank's profit maximization model. The results show that almost all corporate credit rates are at a critical value of 15%. Therefore, we introduce an adjusted bank revenue model. For customers with different credit ratings, we should pursue the maximization of bank revenue while ensuring that customers are not lost, instead of just focusing on the current bank's revenue maximization.

Table 8. Industry expected loss rate and unexpected loss rate

Industry	Expected loss rate elp	Unexpected loss rate ulp
1	0.68%	7.23%
2	0.68%	7.58%
3	0.72%	7.34%
4	0.28%	4.53%
5	0.71%	7.27%
6	1.21%	8.67%
7	0.63%	7.21%
8	0.55%	6.18%
9	1.18%	10.00%
10	0.05%	1.65%

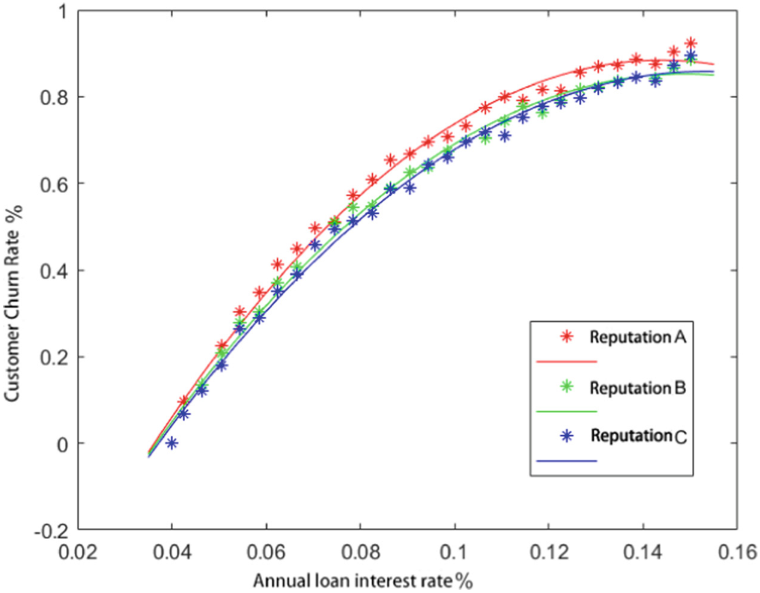

Fig. 1. The relationship between credit interest rate and customer churn rate

We set a minimum customer retention rate for each credit rating and use this to optimize our credit strategy. Finally, the corresponding interest rate under the proportion of each company's credit line is obtained, which is the credit strategy. (See Appendix Table 3 for specific results) (Table 9).

Table 9. Comparison table of reputation level and customer retention rate coefficient k

Credit rating	A	B	C
k	0.7	0.6	0.5

4 Build Companys' Reputation Rating

In order to quantitatively analyze the credit risk of the bank's 302 companies with no credit records, and give the bank's credit strategy for these companies when the total annual credit is 100 million yuan, this article first uses the given data to make a credit rating. Based on the credit risk model, the xgboost classification model is used to predict the reputation level of each company, and then the calculated reputation level and existing risk factors are used to bring the credit risk model into the credit risk model to estimate the size of the risk, and finally bring in the given credit to solve the optimal credit strategy of the annual total credit amount.

4.1 Model Establishment

1) *Reputation rating.* For the 302 companies that need quantitative analysis, because these companies have no historical credit records and do not know the reputation level of each company, we need to rate the reputation level of each company before quantifying the credit risk, and then perform credit risk assessment.

   Given that there are only 123 corporate samples with known reputation ratings, the current mainstream neural network algorithms based on big data training are difficult to learn features from this small data, and there is a great risk of overfitting; at the same time, the classification rules for a single decision tree model are relatively simple, The use is limited; in general, under the condition of small amount of data and multiple features, the integrated tree model is often better than neural network and single decision tree, and the classification performance is good. Therefore, we choose xgboost, an ensemble learning algorithm based on a decision tree model, to predict the credit rating of companies without credit records.
2) *Adjust Credit Strategy.* After passing the credit rating, we obtained the credit rating information of each company. Next, the principal component analysis method in question 1 is used to reduce the dimensionality of the influencing factors of companies without credit records from the two dimensions of company size and the stability of supply and marketing relationships, and finally bring them into the credit risk model and the bank's profit maximization model. Each company scores risks, and then obtains the optimal solution with an annual total credit of 100 million through the bank's profit maximization model.

Table 10. Xgboost model variable.

Variables					
Var(X1)	Var(X2)	Var(X3)	Var(X4)	Var(X5)	Var(X6)
Std(X1)	Std(X2)	Std(X3)	Std(X4)	Std(X5)	Std(X6)
Max(X1)	Max(X2)	Max(X3)	Max(X4)	Max(X5)	Max(X6)
Min(X1)	Min(X2)	Min(X3)	Min(X4)	Min(X5)	Min(X6)
Mean(X1)	Mean(X2)	Mean(X3)	Mean(X4)	Mean(X5)	Mean(X6)
Sum(X3)	Sum(X6)	Sum(X7)	Sum(X8)	Sum(X9)	
X10	X11	Max(X8)	Max(X9)		

Table 11. Corporate Credit Rating

Enterprise code	Credit rating
E128	B
E129	A
E130	A
E131	C

4.2 Model Solution

1) *Reputation rating solution.* We use the xgboost classification model in the sklearn library in Python machine learning to predict the reputation rating of 302 companies without credit records. For 123 samples with known repg_rate (learning rate), controls the running speed and the degree of fitting; max_length (the maximum depth of the tree), is used to control the degree of fitting to the sample. The trained model is tested on the test set, and the accuracy of predicting the reputation rating model is about 68.42%, which is good. There are 69, 119, 59, and 55 companies with credit ratings of A, B, C, and D respectively (Table 10).

   The specific calculation results are shown in Appendix Table 13 (Table 11).

2) Adjust Credit Strategy solution (Figs. 2 and 3).

- Principal component analysis and dimensionality reduction
  We continue to use a total of 19 indicators in the two dimensions of company size and company stability for principal component analysis.

  All the variables are imported into SPSS for principal component analysis and KMO test. The results show that further principal component analysis can be performed.

  For the dimensions of company scale and the dimensions of supply and marketing stability, three principal components are finally determined to achieve dimensionality

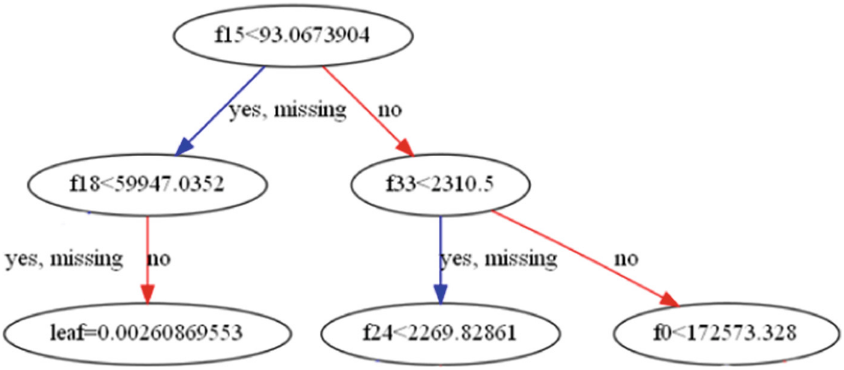

Fig. 2. Single decision tree structure (partial)

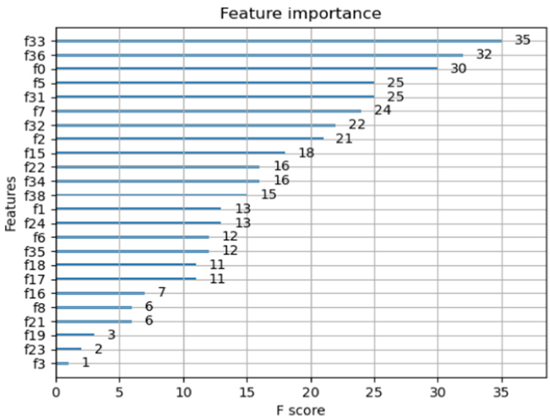

Fig. 3. Feature importance

reduction. We rename the newly generated variables to f1, f2, f3, f4, f5, f6 (see Appendix Table 5). The quantitative results of industry vitality are consistent with the results of the previous solution.

Based on the data solved above, we estimate the credit risk model of companies without credit records, and use the TOPSIS method to predict the risk of each company. See Appendix Table 6 for specific scores.

- New credit strategy solution
  According to the TOPSIS forecast results and the bank's profit maximization model after adjusting the interest rate, we bring in the total annual credit of 100 million yuan, and solve the credit line of each company. And get the corresponding interest rate, and finally get 220 companies to give loans. The specific results are shown in Table 12.
- The adjustment of the model strategy by emergent factors
  Taking into account unexpected risk factors, companies may encounter emergencies such as natural disasters during production and operation. When this happens, how

Table 12. The adjusted credit strategy

Enterprise code	credit limit	Credit rate
E128	0.004866249	0.052533066
E129	0.005220079	0.041250539
E130	0.004195696	0.051321898
E131	0.005438113	0.054769316
E132	0.00468004	0.046010516

should companies adjust their credit strategies on the basis of the xgboost classification model? This article takes the COVID-19 as an example to quantify the impact of the epidemic on each industry, and then put it as a sudden risk factor multiplier into the xgboost classification model to measure the impact of sudden risk on the entire sample, and then calculate and adjust the new credit strategy.

4.3 Model Building

Taking the COVID-19 epidemic as an example, we regard the impact of the epidemic as a sudden situation and believe that it will affect the entire industry. However, due to different industrial and economic structures, different industries have different resistance to the impact of the epidemic and the speed of recovery. Therefore, we quantify the industry losses caused by the impact of the epidemic on different industries as sudden risk factors, and put them into the credit risk model T as a multiplier:

$$S_i' = T_i \cdot S_i \tag{21}$$

Then the new credit risk model is brought into the bank's profit maximization model for optimization.

4.4 Model Solving

Through literature search, we selected the data on the profitability of enterprises in different industries in 2020 shown in the "Impact of the COVID-19 Epidemic on Enterprises and Suggestions for Countermeasures-A Quick Questionnaire Survey Report on Thousand Enterprises" [1]. Choose the loss rate of the first quarter year-on-year as the value of the sudden risk factor T_i (Table 13).

Then, the new S_i' is brought into the bank's profit maximization model for optimization and solution, and the adjusted optimal credit strategy is obtained, and finally 196 companies are granted loans. The specific results are shown in Table 14.

Table 13. Industry Sudden Risk Factor Table

Category		Industry Sudden Risk Factors
1	Mining Industry	0.611
2	Manufacturing	0.639
3	Electricity, Gas and Water Production and Supply Industry	0.325
4	Construction Industry	0.606
5	Transportation, Storage and Postal Industry	0.571
6	Information Transmission, Computer Service and Software Industry	0.625
7	Wholesale and Retail Trade	0.679
8	Real Estate Industry	0.643
9	Leasing and Business Services	0.556
10	Water Conservancy, Environment and Public Facilities Management Industry	0.438

Table 14. Credit Strategy

Enterprise code	credit limit	Credit rate
E130	0.005051958	0.046334029
E131	0.006547929	0.049663484
E132	0.005635148	0.041538848
E133	0.005786093	0.040455199
E134	0.004508737	0.062031522

5 Model Evaluation and Improvement

5.1 Advantages of the Model

We extracted many important factors during data preprocessing, such as the industry sector of the company. This factor can directly establish a relationship with the emergency situation of the third question. At the same time, extracting a large number of relevant factors can also reduce the follow-up corporate reputation. The over-fitting phenomenon caused by too small data size in the prediction of degree level.

We have established the relationship between the bank credit policy and the company's scale, the stability of the supply and marketing relationship, the company's reputation level and the company's industry sector, and scored based on this. The level of the score can accurately and clearly determine whether each company can finally Loan and determine the loan amount.

Introducing an improved RAROC model, including expected losses into the current cost deduction, and at the same time including unexpected losses into the model, which improves the authenticity and reliability of the bank's rate of return.

The customer retention rate coefficient is introduced to ensure that the bank can retain more customers while minimizing the loss of profits, and consider the development of the bank from a long-term perspective.

5.2 Disadvantages of the Model

There is inevitably a correlation between the influencing factors, and the interaction between the influencing factors has not been fully considered.

Due to the small size of the data, it is not suitable to use neural networks, and the predicted value after we use xgboost and constantly adjust the parameters is still low, so that the quantitative analysis of credit risk in the second question still needs to be improved.

The quantification of credibility levels is relatively lacking in objectivity.

6 Conclusion

This paper first establishes the credit risk model and bank profit maximization model. The TOPSIS method was used to evaluate 123 msmes with credit history in banks. Then, RAROC loan pricing model and XGBoost model are introduced to establish the bank profit maximization model. The classification model predicts the credit level of 302 small, medium and micro enterprises with no credit history and determines the optimal credit strategy. Finally, the COVID-19 pandemic is used as an example to quantify the impact of the pandemic on various sectors. Then it is included into the credit risk model as the multiplier of emergency risk factors. It measures the impact of emergency risk on the entire sample. It then calculates and adjusts the new credit strategy. This paper deeply analyzes the credit decision-making mechanism of micro, small and medium-sized enterprises, and gives the corresponding reference value.

References

1. Chinese Entrepreneur Survey System Research Group (2020) The impact of the COVID-19 epidemic on enterprises and countermeasures and suggestions-a rapid questionnaire survey report on thousands of enterprises. Econ Circ 3:80–96
2. Li F (2014) Research on industry loan pricing based on RAROC model--taking China construction bank as an example. Zhejiang Gongshang University
3. Zhang F (2020) Research on the problems and countermeasures of big data solutions for financing difficulties of small and medium sized enterprises. J Soc Sci Human 2(8)
4. Aribaba, FO, Ahmodu A-LO, Oladele R, Yusuff AS, Olaleye RB (2019) The role of deposit money banks' loan facilities in financing small and medium-scale enterprises in Nigeria. Acad J Interdiscip Stud 8(2)
5. Guo M, Zhang B, Zhao L, Zhang Y (2020) Activity semantic recognition method based on joint features and XGBoost. Comput Appl 1–8

6. Guo N, Fan S, Li K (2020) Research on credit constraints of my country's small and medium-sized enterprises from the perspective of bank-enterprise relationship—analysis of the questionnaire on financing of small and medium-sized enterprises. Invest Res 39(6):26–35
7. Hao Y (2013) Research on credit issues of small and medium-sized commercial banks based on game analysis. Nanjing University
8. Li Z (2014) Research on loan pricing of commercial banks in my country under the background of interest rate marketization. Shandong University of Finance and Economics
9. Liu X (2018) Analysis of financing situation and countermeasures of small and medium sized enterprises in China. Int J Soc Sci Univ 1(1)
10. Qu X (2014)Comprehensive evaluation of environmental pollution in China's industrial industry-an empirical analysis based on Topsis. Indust Econ Res 4:51–59+102
11. Sun X (2013) Research on the loan pricing of small and medium-sized commercial banks in my country-based on the RAROC model. Jilin University
12. Yuchen S (2021) Research on the bank's optimal credit strategy for small and medium-sized enterprises under information asymmetry—based on the logistic regression default rate measurement model. Financ Develop Res 6:78–84
13. The People's Bank of China (2020) China Banking Regulatory Commission, Development and Reform Commission, Ministry of Finance, Ministry of Finance, State Administration of Market Supervision, Securities Regulatory Commission, Foreign Exchange Bureau, Guiding Opinions on Further Strengthening Financial Services for Small, Medium and Micro Enterprises-Yinfa No. 120. http://www.safe.gov.cn/hunan/2020/0605/1414.html
14. Lihe X, Qing L (2020) The influence of the structure of credit sources on the multi-dimensional innovation activities of SMEs. Financ Res 46(7):19–34
15. Huayi X, Delun M, Dalu F (2016) An empirical study on the pricing of credit assets of commercial banks—based on the perspective of the allocation of credit assets pricing of small, medium and micro enterprises. Financ Theory Pract 8:1–5