International Journal of Trend in Research and Development, Volume 9(5), ISSN: 2394-9333
www.ijtrd.com

An Empirical Study on the Credit Evaluation Model of Logistics Enterprises

Gui Junhan
Beijing Wuzi University, Beijing, China

***Abstract*—**Enterprise credit construction is not only related to the long-term development and lasting vitality of enterprises, but also an important means to promote the entire industry to enhance credit awareness and improve comprehensive competitiveness, and is the basic guarantee for the healthy operation of the national economy. With the vigorous development of China's modern logistics industry, in the face of fierce market competition, it has become an urgent problem for logistics companies to improve their credit brand, expand their credit reputation, and enable them to obtain more business opportunities in the operation of the modern logistics market. Therefore, this paper will draw on the successful experience of credit evaluation in other industries, combine the current situation of the logistics enterprise credit evaluation index system, and establish a set of logistics enterprise credit evaluation index system according to the principles of credit evaluation system construction. In this paper, Logistic regression, decision tree and random forest method are used to establish a credit evaluation model of logistics enterprises. After research, it is found that the three models have better prediction effects on the credit evaluation of logistics enterprises, and the random forest model has the best applicability in the credit risk evaluation of listed companies.

Keywords*—Enterprise Credit Evaluation; Logistic Regression; Decision Tree; Random Forest*

I. INTRODUCTION

Credit is the lifeblood of an enterprise's survival and development, and an "intangible asset" with extraordinary value. Enterprise credit construction is not only related to the long-term development and lasting vitality of enterprises, but also an important means to promote and promote the whole industry to enhance credit awareness and improve comprehensive competitiveness. With the vigorous development of China's modern logistics industry, in the face of fierce market competition, it has become an urgent problem to be solved how logistics enterprises can improve their credit brand, expand their credit reputation, and enable enterprises to obtain more business opportunities in the operation of the modern logistics market.

Data mining technology has developed rapidly in recent years. It can mine useful information from massive data, and has the advantages of high efficiency, flexibility and wide applicability. Therefore, it has been widely used in many fields and achieved remarkable results. Many scholars have begun to try to apply data mining technology to analyze and predict in the field of credit evaluation, and have come to the conclusion that data mining method is better than traditional statistical measurement method. Commonly used data mining methods are: decision tree, logistic regression, random forest, discriminant analysis, neural network, etc.

Although the credit evaluation methods in developed countries are constantly evolving, the credit analysis models are innovating. However, at present, the exploration of the construction and development of the integrity system of logistics enterprises in my country is still in its infancy, and the credit system of the entire industry has not yet been established. Therefore, this paper will draw on the successful experience of credit evaluation in other industries, and combine the status quo of the logistics enterprise credit evaluation index system to establish a set of logistics enterprise credit evaluation index system. In this paper, Logistic regression, decision tree and random forest method are used to establish a credit evaluation model of logistics enterprises.

II. LITERATURE REVIEW

In terms of the construction of enterprise credit evaluation indicators: Gao Lingyun has built a small and medium-sized enterprise credit rating index system including market evaluation, basic quality evaluation, business management evaluation, financial evaluation, technological innovation evaluation and credit activity records. On the basis of analyzing the factors affecting the credit risk of small and medium-sized enterprises, Wu Yan selected 20 indicators from six aspects: solvency, profitability, operation and management, capital utilization, credit status and development potential, and constructed a credit system for small and medium-sized enterprises. Due to the short development history of small and medium-sized enterprises in my country, the imperfect financial system, and the serious distortion of corporate financial information, it is not feasible to simply use financial indicators in practice. Many scholars have recognized the important supplementary role of non-financial indicators. Jiang Lingmin proposed a quantitative analysis model of financial factors from three perspectives of industry risk, operational risk and management risk, which solved the problem of too much subjectivity in the analysis of non-financial factors.

In terms of research methods: Xia Liming established a credit risk evaluation model for small and medium-sized enterprises based on supply chain finance by using the grey analytic hierarchy process and the one-time threshold method. Wei Yun uses a combination of clustering and factor analysis to comprehensively evaluate the credit risk of SMEs. Liang Qi comprehensively uses financial indicators and corporate governance indicators, and uses robust logistic regression considering extreme value samples to construct a financial failure early warning model. Hu Haiqing used the support vector machine algorithm (SVM) to establish a credit risk assessment model, and compared it with the credit risk assessment model established by the BP neural network algorithm.

In general, the shortcomings of the current research are mainly reflected in the following three aspects: First, there are few researches that consider enterprise credit risk evaluation based on the logistics industry, and more are to evaluate the credit risk from the small and medium-sized enterprises themselves. Second, the current logistics enterprise evaluation index construction methods are different, and the current industry does not have a comprehensive, scientific and authoritative index system. Third, the current corporate credit assessment methods are rich, but there are few literature studies on the comparative research of each model.

III. THE CONSTRUCTION OF INDEX SYSTEM

The three-dimensional credit theory adopted in this paper is a popular innovative credit theory system in our country. It was proposed by Professor Wu Jingmei of Renmin University of China, and provided a basic theoretical framework for the construction of the credit evaluation index system of my country's logistics enterprises. The three-dimensional credit theory includes three dimensions: integrity, compliance, and compliance, covering credit activities in three aspects: the basic quality of credit subjects, social activities, and economic activities. Therefore, this paper selects 21 indicators from five aspects of enterprise peripheral risk, solvency, operating ability, profitability, and development ability. The indicator system is shown in Table 1.Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

TABLE I. INDICATOR SYSTEM

First-level indicator	Secondary indicators
Enterprise Peripheral Risk	The number of administrative penalties and environmental penalties in the past two years The number of untrustworthy and abnormal operations of surrounding enterprises in the past two years The number of administrative and environmental protection penalties of surrounding enterprises in the past two years
Solvency	current ratio quick ratio equity ratio cash flow debt ratio assets and liabilities equity Multiplier
Management capacity	inventory turnover accounts receivable turnover current asset turnover fixed asset turnover total asset turnover
Profitability	return on equity return on assets net profit
Development ability	operating income growth rate net profit growth rate total asset growth rate net asset growth rate

IV. MODEL BUILDING

A. Data sources

This paper selects listed and NEEQ logistics companies as the research objects, and according to the principle of data availability, determines the number of samples that conform to the untrustworthy behavior of logistics companies from 2000 to 2019. The ratio of enterprises to normal enterprises, according to the ratio of 1:2, selects 217 normal credit samples in the same accounting period as the untrustworthy samples. The sample financial data comes from Guotaian database and Wind database. This paper takes whether the logistics enterprise is untrustworthy as the dependent variable (Y=0 means that the enterprise is not untrustworthy, Y=1 means that the enterprise is untrustworthy), and uses the selected 21 indicators as independent variables to conduct research.

B. Data preprocessing

1) Handling of missing values and outliers

Through data review, it is found that there are missing values in financial variables such as current ratio, quick ratio, and property right ratio, and the proportion of missing values for each variable is less than 20%. Therefore, this paper uses the commonly used mean imputation method to supplement missing values. The outliers are processed by elimination method.

2) Standardized processing

In the constructed evaluation index system, since the units of variables are inconsistent, this paper processes the variable data through Z-value standardization to eliminate the dimensional relationship between variables.

C. Model selection

1) Logistic model

The main purpose of logistic regression is to establish a model to find out the functional relationship between the probability of an event and each variable. Assuming an event occurs, Y=1, otherwise Y=0. P is the probability that the event occurs, 1-P is the probability that the event does not occur, P can be expressed as a linear function of the explanatory variable X, and the logistic model changes the equation to:

$$\text{Logit(p)} = \ln\left(\frac{P}{1-P}\right) = Y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \varepsilon$$

Then the probability of occurrence of the event returned by the logistic model is:

$$P = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}$$

The logistic model generally sets the threshold to 0.5, and the classification result can be expressed as:

$$\text{Label} = \begin{cases} 1 & P > 0.5 \\ 0 & P < 0.5 \end{cases}$$

2) Decision tree

The basic principle of decision tree is to find the best classification criteria by recursive cutting method, and then finally form rules. It can summarize decision rules from some data with numerous features and labels, and present these rules in a tree-like structure. The decision tree algorithm uses training data to establish a decision tree model according to the principle of minimizing the loss function, which usually includes three steps: feature selection, decision tree generation and decision tree pruning.

3) Random forest

Random forest is one of the decision tree learning methods. The decision tree learning method is a classification method composed of a tree structure, and the class probability is calculated in each branch. The decision tree starts with the root node and gradually builds up sub-trunks. The internal nodes of the sub-trunks are also called leaves. Each node represents a related index. The samples are classified according to the index, and the layers are progressive. Although the decision tree is easy to build and has a wide range of applications, it still has major defects. In order to improve the accuracy of decision trees, scholars have used iterative and bagging techniques. Random forest can be regarded as an enhanced version of bagging technology. By constructing a large number of random decision trees for classification, training multiple tree models, and obtaining the final classification result.

D. Comparison of models

In this paper, 70% of the sample data is randomly selected as the training set, and the remaining 30% of the sample data is used as the test set for evaluating the model performance.

Taking the initial indicators as model variables, the R4.0.4 software was used to build the model. The classification evaluation standard is an important indicator to verify whether a classifier can classify accurately. The evaluation standards of the classification algorithm are: accuracy, precision, recall and F-value. In this paper, the confusion matrix (Table 2) is obtained according to the prediction results of the data test set, and the above four evaluation criteria are calculated to compare the pros and cons of different models.

TABLE II. TABLE STYLES

	A sample of dishonesty	Unfaithful sample
Samples predicted to be untrustworthy	a	b
Predicted samples that are not discredited	c	d

The accuracy rate refers to the proportion of correctly classified samples to the total number of samples, reflecting the overall performance of a classifier. The calculation formula is as follows:

$$\text{Accuracy} = \frac{a+d}{a+b+c+d}$$

The accuracy rate refers to the ratio of the number of correctly classified positive samples to the number of samples determined by the classifier as positive samples, which reflects the reliability of a model. The calculation formula is as follows:

$$\text{Precision} = \frac{a}{a+c}$$

The recall rate refers to the ratio of the number of correctly classified positive samples to the number of real positive samples, which reflects the effectiveness of the model. The calculation formula is as follows:

$$Recall = \frac{a}{a+b}$$

The F-value is the harmonic mean of precision and recall, and is calculated as follows:

$$F = \frac{2 * Precision * Recall}{Precision + Recall}$$

It can be seen from Table 3 that the random forest model has the highest accuracy rate, precision rate, recall rate and F-value, and the accuracy rate is 83.7%, indicating that the model prediction effect is better. The random forest model can calculate the influence of the variable on the impurity of each node in the classification tree to determine the importance of the variable, as shown in Fig 1, so it is more suitable for the credit evaluation model of logistics enterprises. This also shows that random forests are not sensitive to multivariate collinearity, and are more robust in modeling missing data and unbalanced data, with the advantage that the number of independent variables can be as many as several thousand.

TABLE III. MODEL COMPARISON

	Accuracy	precision	recall	F-value
Logistic Regression	0.6817	0.977	0.693	0.811
Decision tree	0.727	0.841	0.757	0.797
Random forest	0.837	0.954	0.827	0.886

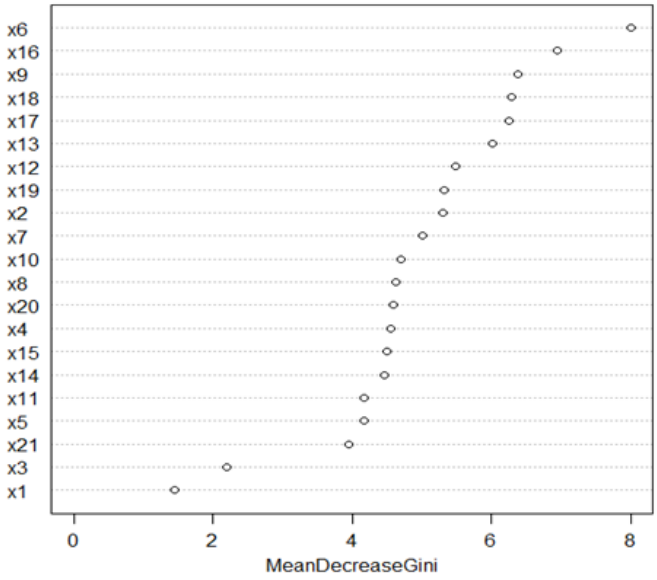

Fig. 1. Indicator importance

CONCLUSION

This paper selects 330 listed logistics companies in my country to conduct empirical research. Taking the company's ST as the criterion for poor credit status, a Logistic regression model, a decision tree model and a random forest model were established to classify and predict the credit status of listed companies. The final result shows:

(1) The two models have better prediction effects on the last company's credit evaluation, and the random forest model has the highest accuracy of 83.7%, which shows that the random forest model has the best applicability in the credit risk evaluation of listed companies.

(2) From the random forest model, it can be seen that the property right ratio, total net interest rate of total assets, inventory turnover rate, net profit growth rate, operating income growth rate, and fixed asset turnover rate are more important and have a greater impact on corporate credit.

Acknowledgment

Thanks to my teacher for her careful guidance, her profound knowledge has given me profound inspiration. I will continue to work hard in the next research career and continue to move forward in this research direction.

References

[1] Gao Lingyun, Cheng Min, Xu Haijun. Research on the construction of credit evaluation system for small and medium-sized enterprises in my country [J]. Journal of East China Jiaotong University, 2004(06): 90-93.
[2] Wu Yan. The establishment of credit evaluation system for small and medium-sized enterprises in my country [J]. Research on Science and Technology Management, 2005(09):205-206.
[3] Li Liqun. Analysis of non-financial factors in credit decision-making [J]. Financial Theory and Practice, 2004(09): 24-26.
[4] Jiang Lingling. A method of commercial bank customer credit rating based on fuzzy judgment [J]. Business Research, 2006(22):120-123.DOI:10.13902/j.cnki.syyj.2006.22.036.
[5] Wang Xuan, Guo Weichen. Analysis and Research on Credit Evaluation of Third-Party Logistics Enterprises [J]. Development Research, 2011(04):129-133.
[6] Wei Yun, Zhu Quancong. Research on credit risk evaluation model and application of small and medium-sized enterprises based on cluster analysis and factor analysis [J]. Credit Information, 2018,36(04):32-35.
[7] Sheng Xia, Li Bin, Zhang Di. Prediction of changes in credit evaluation of listed companies based on data mining [J]. Statistics and Decision, 2016(15):159-162.
[8] Liao Wei. Research and Application of Third-Party Logistics Enterprise Credit Model [D]. Hunan University, 2013.
[9] Guo Xiaobo, Wang Wanting, Zhou Xin. Effectiveness Analysis of Credit Risk Identification Factors for Small and Medium-sized Enterprises in my country: Based on the Credit Data of Small and Medium-sized

Enterprises in Beijing [J]. International Finance Research, 2011(04): 62-67.

[10] Xia Liming, Zeng Lele, Meng Li. Research on the Construction of Credit Risk Evaluation Model for Small and Medium Enterprises——Based on the Perspective of Supply Chain Finance [J]. Operation and Management, 2011(11):103-105.DOI:10.16517/j. cnki.cn12-1034/f.2011.11.001.

[11] Liang Qi, Guo Xinwei, Shi Ning. Research on the Early Warning of Financial Failure of Small and Medium-sized Listed Companies in my country—Based on a Comprehensive Investigation of Financial Indicators and Corporate Governance Indicators [J]. Economic Management, 2012, 34(03): 123-132 .DOI:10.19616/j.cnki.bmj.2012.03.016.

[12] Hu Haiqing, Zhang Lang, Zhang Daohong. Research on Credit Risk Assessment of Small and Medium Enterprises from the Perspective of Supply Chain Finance: A Comparative Study Based on SVM and BP Neural Network [J]. Management Review, 2012, 24(11): 70- 80.DOI:10.14120/j.cnki.cn11-5057/f.2012.11.015.

[13] Wang Xinhong, Chen Yanjie, Zhong Weizhou. Construction of credit risk evaluation index system for small and medium-sized enterprises in my country [J]. Journal of Northwest Normal University (Social Science Edition), 2009, 46(04): 129-132. DOI: 10.16783/j .cnki.nwnus.2009.04.026.

[14] Shen Fei, Wang Xifu, Hu Pujuan, Yu Yacen. Research on the Design of Credit Evaluation Information System for Modern Logistics Enterprises in my country [J]. Logistics Technology, 2009, 28(01): 130-132.