ISSN: 2710-0170

Research on machine learning and financial risk early warning of listed companies based on random forest model

Jingting Liu, Xueting Zhou, Yue Wei, Liyuan Lu

School of Accountancy, Anhui University of Finance & Economics, Bengbu Anhui 233030, China.

Abstract

This paper explores the application of machine learning technology based on random forest model in financial risk early warning of listed companies. By analyzing the data, we constructed a random forest model to identify potential financial risk factors. The study found that the random forest model showed high accuracy and stability in dealing with financial risk early warning issues. The model can effectively identify key indicators that affect the company's financial status and provide timely risk warning signals for investors and management. In addition, we also discussed the limitations of the model and the direction of future research, aiming to provide listed companies with a more accurate and effective financial risk early warning system, thereby helping companies to identify and respond to potential financial challenges as early as possible.

Keywords

Random forest model, decision tree, machine learning, financial risk warning.

1. Introduction

In the context of global economic integration, the financial health of listed companies directly affects the stability of the capital market and the confidence of investors. However, financial crises are often hidden and sudden. Once they break out, they will not only cause significant losses to related companies, but may also affect the entire industry and even the entire economic system. Therefore, early identification and early warning of financial risks have become the focus of common concern for corporate managers, investors and regulators. In recent years, with the rapid development of big data and machine learning technologies, using these technologies to warn of financial risks of listed companies has become a hot topic of research.

Traditional financial risk warning methods, such as ratio analysis and multivariate discriminant analysis, can reveal signs of financial crisis to a certain extent, but are often limited by the linear assumptions of the model and its ability to handle complex financial relationships. The random forest model, as an important algorithm in the field of machine learning, provides a new perspective for financial risk warning with its nonlinear decision tree integration, powerful feature selection capabilities and adaptability to high-dimensional data. The model constructs multiple decision trees, each of which independently classifies or regresses the data, and finally obtains the prediction results by voting or averaging. It can not only process a large number of financial indicators, but also automatically identify key financial risk factors, thereby improving the accuracy and timeliness of warnings.

This paper will build a financial risk early warning model based on random forests, verify its prediction performance, and explore the limitations and improvement directions of the model in practical applications. In addition, this paper will also compare and analyze the differences between the random forest model and traditional early warning methods to provide listed companies, investors and regulators with a more scientific and effective financial risk early

ISSN: 2710-0170

warning tool. Through the research of this paper, it is hoped that it can provide valuable reference and guidance for theoretical research and practical operations in related fields, and promote the healthy and stable development of the capital market.

2. Literature Review

Research on machine learning of random forest models and financial risks of listed companies. In 2019, scholar Ma Xuhui proposed that risks in the financial market are everywhere, and the operations of listed companies are affected by various factors. Therefore, the research on the financial risks of listed companies is also of great significance. Traditional research methods based on financial statements, such as the five-factor Z-score model, are of great significance in the field of financial risk prediction, but they still have their limitations. It is proposed to use decision tree models and random forest models for financial early warning of listed companies. Scholar Shan Yulu (2020) proposed two integrated algorithms: Stacking model and SVM-Logistic model based on parallel connection, and obtained the conclusion that the support vector machine model has the best prediction performance among the four base models; the prediction effects of the two integrated models are better than the individual base models; the prediction effect of the Stacking model is slightly better than the SVM-Logistic model. Liang Chuangwei (2021) conducted a comprehensive evaluation of the early warning effects of different models by constructing random forest models and Ada Boost algorithms, and finally proposed that the early warning results of random forest in the integrated learning method are the most accurate among all models. Huang Xiaowei (2023) summarized the commonly used machine learning models and their applications in financial early warning, and finally pointed out the possible future research directions of machine learning in financial early warning. It further proves that effective financial early warning helps stakeholders judge the operating conditions of listed companies, make decisions and avoid risks. Machine learning models can quickly process large amounts of data, mine valuable information and draw conclusions, and can efficiently judge and warn the financial status of listed companies.

3. Early warning model based on machine learning

Effective financial early warning helps to judge the company's financial status and operating results to avoid risks. This requires the early warning model to be able to adapt well to the company's operating environment, dynamically obtain relevant information, and ensure the accuracy of financial early warning. Machine learning can quickly and accurately capture information, process large amounts of data, and draw conclusions. The more common machine learning algorithms in financial distress early warning research include neural networks, decision trees, and some ensemble learning methods.

3.1. Neural Networks

The basic building block of a neural network model is a neuron. A neuron receives some inputs, performs weighted processing on these inputs, and generates outputs through an activation function. The output of the neuron is then used as the input of the next layer of neurons, forming a network structure. The training process of a neural network is to continuously adjust the weights and biases of each neuron so that the neural network can more accurately predict the output results. In 1943 , McCulloch and Pitts produced the first neural network based on their artificial neurons. Since then, many studies have used the neural network model and the results show that the model is superior to the traditional early warning model.

3.2. Decision Tree

A decision tree is a prediction model with good prediction capabilities. It is a classification and regression model based on a tree structure. It gradually divides the data through a series of

ISSN: 2710-0170

judgment conditions to form a tree-shaped decision rule for prediction. If the decision tree has an overfitting problem, it can be solved by pruning, that is, cutting off some nodes from the generated tree. Adding pruning to the structure of decision tree generation is called pre-pruning. The idea is to determine whether the generalization performance of the decision tree will be improved if the node is divided before each node is divided . If the division can be performed, or not , the division operation is not performed, and the node is a leaf node . The idea of post-pruning is to first generate a complete decision tree according to the generation steps of the decision tree, and estimate each non- leaf node of the decision tree from the bottom to the top to see whether removing the node can improve the generalization performance of the decision tree. If so, replace the subtree with a leaf node.

3.3. Random Forest

Random forest is an ensemble learning method that consists of multiple decision trees. It improves the accuracy of prediction by combining the prediction results of multiple decision trees through voting or averaging. Random forest is known as "a method that represents the state of the art in ensemble learning". Previous studies have shown that the performance of random forest models is better than other technologies. Shrivastava et al. discussed the feasibility of random forest algorithm in predicting the bankruptcy of Indian companies , and pointed out in particular that random forest analyzes the importance of features based on the likelihood of the company getting into trouble, which helps explain the connection between the tendency of corporate bankruptcy and its characteristics.

In the face of various types of enterprise data, the use of machine learning can effectively improve the quality and efficiency of prediction. The ability of machine learning can also be upgraded according to the continuous update of data. Combining various factors, this article will focus on analyzing and processing the financial risk warning of listed companies by establishing a random forest model.

4. Analysis of model prediction results

First, we built a random forest model based on the financial statements of listed companies, and constructed relevant financial indicators by studying relevant literature on financial distress. These data and whether listed companies have financial risks were used as inputs to the model. Based on these data and sample labels , we built three models, namely neural network model, decision tree model and random forest model . From the empirical results , the combination with the boosting algorithm does improve the effect of the decision tree . However, the random forest model is better than the simple decision tree model in various indicators. At the same time, the random forest model has the highest accuracy and recall rate, so that the financial risks of most listed companies can be identified. The neural network model is also used to classify the text of financial risk identification based on a small number of keywords and simple rules, and it is integrated with the random forest model to improve the precision and accuracy of the model.

5. Conclusion

In this study, we explored in depth how to use the advanced machine learning technology of the random forest model to provide early warning of the financial risks of listed companies. Through detailed data analysis and model verification, we confirmed the excellent ability of the random forest model in processing complex financial data and identifying potential risk signals. This model can not only accurately predict the financial status of listed companies, but also effectively distinguish different risk levels, providing valuable decision-making basis for investors and management. Future research directions are recommended to focus on further

ISSN: 2710-0170

optimization of the model, such as combining deep learning technology to explore deeper feature relationships, and introducing real-time data streams to realize a dynamic early warning system to adapt to the rapidly changing market environment. At the same time, the integration of cross-domain knowledge, such as combining macroeconomic indicators, industry characteristics and corporate strategies, will help build a more comprehensive and accurate early warning system.

The results of this study not only promote the research in the field of machine learning and financial risk early warning in academia, but also provide effective tools for the practical world, helping to improve risk management efficiency and promote the healthy and stable development of listed companies. We expect that with the continuous advancement of technology, machine learning will play a more critical role in financial risk early warning and contribute to the sustainable development of the economy.

Acknowledgements

This work is supported by Anhui University of Finance & Economics 2024 Undergraduate Research innovation fund project fund, Project number: XSKY24039ZD.

References

[1] Ma Xuhui. Research on intelligent identification of financial risks of listed companies based on machine learning[D]. Master's thesis of Nanjing University , 2019

[2] Shan Yulu. Research on early warning of financial distress of listed companies based on machine learning[D]. Master's thesis of Shandong University , 2020

[3] Liang Chuangwei. Research on early warning of financial distress of listed companies based on machine learning[D]. Master's thesis of Zhongyuan University of Technology , 2021

[4] Huang Xiaowei. A review of research on financial early warning of listed companies based on machine learning [J]. Modern Commerce and Industry, 2023 (3): 138-139.