ISSN: 2710-0588

Research on Random Forest Model and Support Vector Machine Model in Financial Risk Early Warning of Listed Companies

Jingting Liu[1], Haixin Zhang[2], Yu Han[1], Qian Zheng[1]

[1] School of Accountancy, Anhui University of Finance & Economics, Bengbu, Anhui 233030, China

[2] School of International Trade and Economics, Anhui University of Finance & Economics, Bengbu, Anhui 233030, China

Abstract

On October 30, 2023, the Central Financial Work Conference was held in Beijing. One of the important contents of the meeting was to prevent and control financial market risks. Because the sustainable development of a company is closely related to the financial market , the company faces increasing financial risk challenges , and it is particularly urgent to establish and optimize an effective financial risk early warning mechanism. The establishment of a financial crisis early warning model can not only enable company managers to adjust the company's development strategy and direction in a timely manner, improve the company's performance, and reduce the company's financial risks, but also predict corporate financial crises, which can help investors evaluate the company's potential value and change the direction of investment. This paper takes the establishment of random forest model as the starting point . First, we understand the theoretical framework of financial risk of listed companies; second, we explain the theoretical basis and advantages of establishing random forest model; then, we deeply explore the theory and empirical analysis of risk indicators; then, we use support vector machine in the specific application of financial risk warning; finally , we get the theoretical analysis and empirical significance of its feature importance, as well as the limitations and future research directions of the method under the fusion of random forest model and support vector machine.

Keywords

Financial Risk Early Warning; Random Forest Model; SVM Model.

1. Introduction

With the implementation of the post-epidemic economic revitalization, the macroeconomic policies in 2023 are gradually relaxed. The company has entered the post-epidemic era, and the financial risk warning work will be further attached great importance. On April 29, 2024, the State Financial Supervision and Administration Bureau put forward the "Guiding Opinions on Promoting the Standardized and Healthy Development of Enterprise Group Financial Companies and Improving the Quality and Effectiveness of Supervision". The guiding opinions aim to effectively prevent and resolve the risks of enterprise group financial companies, further promote financial companies to stick to their main responsibilities and businesses, conscientiously implement national macroeconomic policies, adhere to differentiated, characteristic and professional development, play a necessary and beneficial supplementary role in serving social and economic development, and enhance the sustainability of financial support for the real economy. The real economy is an important part of economic development. As the backbone of the real economy, the steady development of the company is crucial to the sustainable development of the economy. Therefore, financial risk warning should not be

ISSN: 2710-0588

underestimated. However, with the continuous development of the global real economy, the company's operation has gradually shown unstable characteristics. In 2018, Kangmei Pharmaceutical went bankrupt due to serious financial risk problems such as financial fraud and broken capital chain; during the outbreak of the epidemic, unprecedented corporate bankruptcy occurred in the financial market. The successive corporate bankruptcy incidents have attracted the attention of many academic experts to corporate bankruptcy research and the attention of management to financial risk warning. Therefore, it is necessary to analyze and study financial risk warning and financial risk warning models to reduce corporate bankruptcy risks and maintain the stability of corporate sustainable development.

Although the scale of Chinese companies is generally small and the development level is uneven, they are developing rapidly and have great development potential, and are the mainstay of my country's sustainable economic development. At present, many early warning models for corporate financial risks have been proposed. Therefore, this paper selects the random forest model to evaluate corporate financial risks. First, the theory of financial risk indicators is deeply explored and empirically analyzed; then the application of random forest model and support vector machine in financial risk early warning is explained, and finally a relatively accurate conclusion is drawn.

2. Literature Review

Research on financial risk warning of listed companies. Zhou Yanwei (2018) constructed a support vector machine model and concluded that the financial risk warning of listed companies is closely related to the listed company's debt repayment ability, profitability, operating ability, and company growth ability; Duan Yixue, Peng Xuemei, and Fang Kuangnan (2024) conducted a study on listed insurance companies. The net asset return rate, net cash flow from operating activities per 100 yuan of premiums, and net asset change rate can be obtained by the random forest financial diagnosis method. They have the function of identifying financial crises; Zhang Fei (2024) combined the rough set with the random forest algorithm and concluded that the characteristics of financial risk warning are divided into three levels of indicators. The first-level indicators are internal financial factors of the enterprise, non-financial factors of the enterprise, and external macro factors of the enterprise. The second and third-level indicators will detail the first-level indicators.

Research on the application of models for early warning of financial risks of listed companies. In 1964, Dalkey and Helmer proposed the expert survey method, which analyzes the internal and external environment of enterprise organization experts to identify the causes and signs of financial difficulties of the enterprise, so as to predict the possibility of financial risks; Zhou Yanwei (2018) used the TOPSIS method to improve the accuracy of the model under the premise of establishing a support vector machine risk early warning model; Duan Yixue, Peng Xuemei, and Fang Kuangnan (2024) found through multiple comparisons that the accuracy of the random forest financial diagnosis method in financial early warning of insurance companies is much higher than other models; Zhang Fei (2024) concluded that combining rough sets with random forest models can improve the accuracy of financial risk prediction of environmental protection companies in the process of using random forest models to predict financial risks.

3. Financial Risk Warning for Listed Companies based on Random Forest Model

3.1. Theoretical Basis and Advantages of Random Forest Model

Random Forest (RF) is an ensemble learning method proposed by Breiman. It constructs multiple decision trees and integrates their prediction results to improve the accuracy and

ISSN: 2710-0588

robustness of the overall model . Compared with the traditional single decision tree, random forest enhances the generalization ability of the model by introducing randomness. Specifically, each tree randomly samples the data set during training (i.e., bootstrap sampling) and randomly selects features based on this, which is called random feature selection. This method reduces the model's dependence on a single feature, thereby reducing the risk of overfitting.

Another advantage of random forests is their robustness to outliers and noise. Since each tree is trained independently, even if some trees are affected by outliers, other trees can still provide accurate predictions. In addition, random forests are also able to assess the importance of features, which helps identify the most critical financial indicators for early warning of financial risks.

3.2. Theoretical Framework of Financial Risk of Listed Companies

The theoretical framework of financial risk early warning for listed companies is a comprehensive theoretical system, which is based on a comprehensive analysis of the company's financial situation to identify, evaluate and respond to various risk factors that may affect the company's financial health. This framework not only covers the definition and classification of financial risks, but also includes the causes of risks, transmission mechanisms and risk management strategies.

This theoretical framework is built on theories from multiple disciplines including accounting, finance, management and behavioral finance. Accounting provides basic tools such as financial ratio analysis, finance assesses market risk through models such as the capital asset pricing model (CAPM) and modern portfolio theory (MPT), and behavioral finance focuses on the impact of investor psychology and behavioral biases on financial risk.

Risk identification is the first step in the early warning process and involves identifying internal and external factors that may lead to financial losses. Risk assessment quantifies the likelihood and impact of these risks through qualitative and quantitative methods, including sensitivity analysis, stress testing and Monte Carlo simulation.

3.3. Theoretical and Empirical Analysis of Financial Risk Indicators

3.3.1. In-depth Discussion of Theoretical Indicators

Financial risk indicators are key components in the theoretical framework, and they provide a quantitative perspective on a company's financial health. Theoretical indicators include not only traditional financial ratios, but may also include non-financial indicators based on accounting standards, such as corporate governance structure, internal control quality, etc.

3.3.2. Extension of Empirical Analysis

Empirical analysis uses historical data to test the relationship between these indicators and financial risk. This usually involves collecting financial data from a large number of listed companies and using statistical and econometric methods to conduct regression analysis, time series analysis, and panel data analysis to verify the effectiveness and robustness of theoretical indicators.

3.4. Application of Random Forest Model in Financial Risk Early Warning

3.4.1. Detailed Description of Model Construction

The random forest model improves the accuracy and robustness of predictions by integrating multiple decision trees. In financial risk warning, the model can consider multiple financial indicators at the same time and capture the complex relationships and nonlinear characteristics between them. Model construction includes steps such as data preprocessing, feature selection, model training, and parameter tuning.

ISSN: 2710-0588

3.4.2. In-depth Analysis of Predictive Capabilities

The model's predictive power comes from its ability to handle high-dimensional data and nonlinear problems. The random forest model can be used to predict the distribution of features through bootstrap sampling and random selection of features.

To reduce the risk of overfitting and improve the generalization ability of the model.

3.4.3. Model Optimization Strategy

Parameter tuning is a key step in the model building process. Methods such as cross-validation are needed to determine the optimal parameter settings, such as the number of trees, the minimum number of samples for splitting, the maximum depth of the tree, etc., to improve the performance of the model.

3.5. Theoretical Analysis and Empirical Significance of Feature Importance

Feature importance analysis helps to understand the logic behind model predictions and identify the most critical indicators for financial risk prediction. Theoretical analysis shows that feature importance analysis can reveal the contribution of different financial indicators to risk prediction, thus providing a theoretical basis for risk management.

In empirical research, feature importance analysis can not only enhance the interpretability of the model, but also guide the formulation of risk management strategies. By identifying key risk indicators, investors and management can monitor and control risks more effectively.

3.6. Limitations of the Model and Future Research Directions

Although the random forest model shows great potential in financial risk early warning, its "black box" characteristics limit the interpretability of the model. Future research can explore how to combine qualitative analysis and model interpretability tools, such as SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations), to improve the transparency and credibility of the model.

3.7. Conclusion

The random forest model provides a new perspective and tool for early warning of financial risks of listed companies. Through theoretical analysis and feature importance evaluation, the model helps to identify key financial risk indicators and provides a data-driven approach for risk management. However, in order to improve the practicality and credibility of the model, future research needs to conduct more in-depth exploration in terms of model interpretability and cross-domain application.

4. Application of Support Vector Machine (SVM) in the Research of Financial Risk Early Warning of Listed Companies

4.1. Basic Introduction and Theoretical Background of Support Vector Machine (SVM) Model

Model principle and mathematical expression: Support vector machine is a supervised learning model used for classification and regression analysis. It distinguishes different categories by finding an optimal hyperplane in the feature space. The objective function of SVM can be expressed as maximizing the margin while minimizing the classification error. Mathematically, SVM solves the following optimization problem:

$$\min w , b (2 / 1 // w // 2) \text{ subject to } yi(w \cdot xi+b) \geq 1-\xi i, i=1,...,n \text{subject to } yi (w \cdot xi + b) \geq 1- \xi i , i =1,..., n \ \xi i \geq 0, i=1,...,n \ \xi i \geq 0, i =1,..., n$$

ISSN: 2710-0588

Among them, w is the weight vector, b is the bias term, yiyi is the true label of the i-th sample, xi is the feature vector, and ξiξi is a slack variable used to deal with incompletely separable data.

Application of kernel techniques: A key feature of SVM is that it maps data into a high-dimensional space through a kernel function to handle nonlinear separable problems. The choice of kernel function has an important impact on the performance of the model. Commonly used kernel functions include linear kernel, polynomial kernel, radial basis function (RBF) kernel, etc.

4.2. Theoretical and Empirical Analysis of Financial Risk Indicators

4.2.1. In-depth Discussion of Theoretical Indicators

Financial risk indicators are key tools for assessing a company's potential risks. These indicators usually include liquidity indicators, debt-paying ability indicators, profitability indicators, and cash flow indicators. Theoretical analysis requires an in-depth exploration of how these indicators reflect the company's financial health and their relationship with financial risk.

4.2.2. Extension of Empirical Analysis

Empirical analysis uses historical data to test the relationship between these indicators and financial risk. This usually involves collecting financial data from a large number of listed companies and using statistical and econometric methods to perform regression analysis, time series analysis, and panel data analysis. The results of empirical research help verify the effectiveness and robustness of theoretical indicators.

4.2.3. Application of Support Vector Machine Model in Financial Risk Early Warning

1) Detailed description of model construction

In financial risk early warning, the SVM model can distinguish different risk categories by finding the optimal hyperplane in the high-dimensional feature space. SVM is particularly suitable for processing financial data with complex boundaries because it can effectively handle nonlinear relationships through kernel techniques.

2) In-depth analysis of predictive capabilities

The predictive power of the SVM model comes from its ability to find the largest margin in the feature space to split the data. In financial risk early warning, this means that the model can accurately identify risk and non-risk states even when there are complex interactions between data features.

3) Model optimization strategy

Parameter tuning is a key step in the SVM model building process. Cross-validation and other methods are needed to determine the optimal parameter settings, such as the penalty parameter C, kernel function type and its parameters, etc.

4.2.4. Theoretical Analysis and Empirical Significance of Feature Importance

1) In-depth theoretical analysis

Although SVM does not directly provide feature importance measures, feature importance can be indirectly assessed by analyzing support vectors and kernel techniques. Support vectors are data points that are on the boundary of the interval or correctly classified but violate the interval rule, and they are crucial in defining the optimal hyperplane.

2) Expansion of empirical significance

In empirical research, by identifying support vectors, we can determine which financial indicators are most critical to financial risk early warning. This helps improve the interpretability of the model and provide decision support for the formulation of risk management strategies.

3) Exploration of future research directions

ISSN: 2710-0588

Future research can explore how to improve the interpretability of the SVM model in financial risk early warning, such as through visualization techniques or post-processing methods to explain the model's decision-making process. In addition, research can focus on the integration of SVM with other machine learning techniques to improve the accuracy and adaptability of the early warning system.

In summary, the application of support vector machine model in the early warning of financial risks of listed companies shows its potential in classification and prediction. Through in-depth theoretical analysis and empirical research, SVM model can provide a powerful tool for financial risk management, helping investors and management to better understand and deal with potential financial risks.

5. Conclusion

The most important value of this article lies in the use of a new method - random forest model + support vector machine (SVM) to identify and early warn the financial crisis of listed companies. The support vector machine itself is a threshold method that helps us confirm that financial indicators have early warning value; the random forest model can improve the accuracy of early warning and reveal the importance, threshold effect and joint effect of indicators. Therefore, there is continuity and consistency between the two methods. Listed companies can achieve more efficient and rapid early warning and supervision of financial anomalies through "random forest model + support vector machine". This means that driven by artificial intelligence, the future financial risk warning level of listed companies will be significantly improved.

The "random forest model + support vector machine" in this article is relatively simple to construct, avoiding large errors caused by high model complexity. The model is highly accurate and can well predict the company's risk of bankruptcy due to financial risks. However, due to the lack of sufficient empirical analysis in this article, the interpretability is poor and further in-depth research and practice are needed.

Acknowledgments

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

References

[1] Zhou Yanwei. Research on Financial Risk Early Warning of Listed Companies Based on SVM Model[D]. Master's Thesis of Shanghai Normal University, 2018.

[2] Zhang Fei. Construction of financial risk warning model for environmental protection enterprises based on rough sets and random forests [J]. Chinese Agricultural Accounting, 2024(8):51-54.

[3] Duan Yixue, Peng Xuemei, Fang Kuangnan. Research on financial early warning of Chinese property insurance companies - based on random forest financial diagnosis method [J]. Insurance Research, 2024(4):20-33.