

Research Article

Research on Early Warning of Financial Crisis of Listed Companies Based on Random Forest and Time Series

Chi Zhang , Huaigong Zhong, and Aiping Hu

Nanjing Audit University Jinshen College, Nanjing, Jiangsu 210046, China

Correspondence should be addressed to Chi Zhang; zhangchi@naujsc.edu.cn

Received 8 October 2021; Revised 5 November 2021; Accepted 12 November 2021; Published 10 February 2022

Academic Editor: Ateeq Rehman

Copyright © 2022 Chi Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The global economy has entered a new normal, and the economic environment is evolving at a rapid pace. This requires the establishment of a financial crisis early warning system that can be dynamically analyzed based on historical data information. To address this research objective, this study proposes a k -fold random forest algorithm combined with a time series analysis model as an early warning algorithm for corporate financial crises. The algorithm takes advantage of the ability of the time series analysis model to make short-term forecasts of historical data and uses the time series analysis model to make forecasts of newly constructed financial index data. The k -fold random forest is used to analyze the financial situation of the predicted financial data and achieve the purpose of dynamic financial crisis early warning. The experimental results show that the prediction accuracy of the financial crisis early warning model based on the random forest algorithm and time series is 89%, which indicates that the model is effective and feasible.

1. Introduction

With China's economic development entering a new normal and the government encouraging "mass innovation and entrepreneurship," the increasingly competitive market environment has created increasing difficulties for enterprises to operate and develop. At the same time, the "Internet+" thinking is constantly evolving and has been applied to products and services in various industries, promoting the deep integration of the Internet and traditional industries. To stand firm in the rapidly developing market environment, each enterprise must strengthen its risk control ability, grasp the company's financial situation in real-time, and improve the level of management control.

In addition, most of the two domestic stock exchanges are the places where Chinese companies choose to be listed, and to protect the rights of investors and reduce their exposure to listed companies that have already experienced financial crises or other abnormal conditions, the Security and Exchange Commission (SEC) has introduced a special treatment system for stocks (special treatment, ST) and a delisting risk warning (*ST) system which helps in optimization of the

allocation of resources in the capital market. ST or *ST companies will be subjected to more stringent regulation, and more complete information disclosure in the next years will be subjected to stricter regulation and more and more complete information disclosure, which will have a huge negative impact on the company's share price, financing cost, and image [1]. If a company fails to improve its situation and turn around after a financial crisis, it will face the crisis of delisting warning by the SEC, and it may cause the company to face a bankruptcy crisis. For most companies, there will be a large number of current debts between companies, and when one of them has a financial crisis and is unable to pay its debts in time or loses its ability to pay, it will affect the financial arrangement or asset structure of the company. To prevent losses to small and medium-sized investors, the impact on the relevant enterprises or financial institutions and thus the stability of financial markets, investors, corporate decision-makers, and financial institutions should strengthen the control of information on corporate financial crises.

A company's financial crisis is bound to evolve from one small problem to another and the financial crisis of a company. There are other nonfinancial factors in addition to

financial ones. Usually, the changes in financial data reflect this gradual process information, so it is possible to analyze the financial data through certain algorithms and design a system that can scientifically reflect the company's financial situation, thus guiding business decision-makers to set the right guidelines to improve business activities and prevent such problems in time [2].

With the development of artificial intelligence and machine learning technology, theories and applications of data mining are becoming more and more mature, and timely information acquisition and analysis before the occurrence of the financial crisis have gradually become the primary way to make financial crisis prediction, and relevant data analysis algorithms are gradually used in various industries and have achieved very good application results. Therefore, the main research direction of this study is to apply the data mining method to the early warning of enterprise financial crisis and to design a dynamic early warning model of enterprise financial crisis.

The rest of the journal is organized according to the following pattern. Related works are discussed in Section 2. Methodology is discussed under Sections 3 and 4. Experiments and results are mentioned in Section 5, and the study is concluded in Section 6.

2. Related Work

In the early 20th century, scholars in Europe and the United States began to study financial early warning models, and since the markets of capitalist countries are more complete than global markets, many influential results have been achieved in the direction of financial early warning research and practice.

In 1932, Luca et al. [3] used a univariate analysis model for financial early warning research, which is the earliest study of financial early warning models and has some research significance. This study showed that the only indicators that can best discern the financial position of a firm in the selected sample are the net shareholder equity rate and the equity ratio. Kimbrough et al. [4] proposed that the working capital/total assets indicator can well predict financial risk. Dimitrios et al. [5] proposed that the three pairs of financial indicators, shareholder equity/assets and liabilities, working capital/total capital, and current assets/current liabilities, have the strongest predictive power for financial crises. Wei et al. [6] proposed that the three pairs of financial indicators with the strongest predictive power: net income/total assets, total debt/total assets, and cash flow/total liabilities.

Gladwin [7] conducted a study of corporate financial statement data for 1998 using a univariate analysis model and compared four financial ratio indicators, showing that the current ratio and debt ratio had the best early warning effect.

In 1968, Nan and Ying [8] first used a multivariate analysis model for financial early warning research, in which he selected 33 bankrupt firms and 33 normal firms to conduct experiments, and extracted 5 financial indicators with predictive power from 22 financial indicators, all of

which were selected to have the lowest misjudgment rate, thus establishing the Z-score model. Altman's Z-score model has been widely used to date.

In a study by Appiahene et al. [9], whether the Z-score model can effectively determine the financial crisis of listed enterprises in China was investigated, and the results showed that the Z-score model has an outstanding prediction effect for certain types of enterprises. In 2015, Fabling and Grimes [10] established a multivariate probability ratio regression model in conducting a study on financial crisis prediction. In the same year, Burlon et al. [11] improved the Z-score model based on the crowd search algorithm, and the experimental results showed that this model could predict the financial crisis better and more ideally. In 2018, Tang et al. [12] combined the improved fruit fly optimization algorithm and the Z-score model to build a financial early warning model, and the results showed that the model had the good predictive ability.

In 1977, Lee et al. [13] used the first multivariate logistic regression model to predict financial crises in the banking system. 2058 normal firms and 105 bankrupt firms were selected by Ramos-Perez et al. [14], from which nine financial indicators were extracted and a prediction model was developed. The final four indicators with the highest predictive power were identified: financial structure, liquidity, firm size, and operating performance.

Rose et al. [15] conducted a study using financial data of 70 groups of companies for the past 5 years and analyzed the variability of 21 of these indicators, finally concluding that the return on net assets had the best predictive effect. Cohen and Keren [16] added nonfinancial indicators to the study of financial crisis early warning and used the logistic model to test the relationship between the financial crisis and corporate governance.

With the development of data mining technology and artificial intelligence, many scholars have started to use data mining and machine learning methods in the study of financial early warning models, which can take into account compared to traditional statistical methods. Xu et al. [17] used a neural network model to conduct a study of financial crisis early warning of enterprises and compared the neural network model with a multivariate discriminant analysis model, and the results showed that neural network models have significantly higher predictive effects than multivariate discriminant analysis models. Li et al. [18] selected corporate data from Taiwan for forecasting research and constructed multivariate discriminant analysis model, logistic regression model, and neural network model, respectively, and the results showed that the neural network model outperformed the other two models if the sample data did not satisfy a specific assumption. In a study by Tang et al. [19], a BP (back propagation) neural network algorithm was used to forecast 136 listed companies, and it was shown that the improvement of input variables by the efficacy coefficient method effectively improved the accuracy of the NN model and enhanced the forecast accuracy.

Genetic algorithms are methods that simulate genetic selection and the laws of evolution in nature to search for optimal solutions in a large and complex conceptual space.

Genetic algorithms provide a general framework for solving problems that do not depend on the specific domain of the problem and for prediction of corporate crises can be based on multiple aspects of parameter optimization, which can be conditioned to discriminate and extract rules based on qualitative variables. Cheng [20] conducted a study on financial crisis prediction models based on moderate financial indicators and genetic algorithms, and the results showed that this type of model has a high prediction accuracy.

Haardle and Schaafer constructed an early warning model based on a support vector machine and compared it with the neural network model and multivariate discriminant analysis model, and the results showed that the prediction performance of the support vector machine is better compared with other models. In their study, Tian et al. [21] analyzed the financial data of A-share listed companies in China with the help of the SVM model and the nearest domain method and showed that the SVM model is prone to overwarning, while the best domain method is relatively robust. Chen [22] constructed an early warning model for financial crisis based on the adaptive PSO-SVM model in their study, and the study showed that the method is good at dealing with scalability problems and also has good ability to.

Zhang and Hu [23] used bagging and boosting methods in their study to construct financial crisis early warning models and compared them with neural network models, and the results showed that the integrated learning algorithm outperformed the neural network algorithm. Wang and Wang [24] used a multiple voting approach to integrate multiple classifiers to build a financial crisis early warning model, and their experimental results showed that the accuracy of the integrated model is higher than that of the single classifier. After that, Chun Zhao studied financial crisis early warning using association rules and time series in his Ph.D. thesis, and Yuan Wang and Jianhui Yang combined SVM and AdaBoost methods to construct an LCFC early warning model based on SVM-AdaBoost with good prediction results. Mozeika [25] conducted a study on financial early warning of listed manufacturing companies based on random forest and artificial neural networks, and the results showed that the model has better accuracy compared with the neural network model, multivariate logistic regression model, and univariate model.

3. Improvement of the Random Forest Algorithm Based on Financial Data

3.1. Random Forest Algorithm. Since the out-of-bag (OOB) data of the random forest cannot reach 36% of the original data in practice, which is actually used to measure the prediction error, the lack of test data in calculating the out-of-bag error will lead to the low accuracy of the feature importance measure, which leads to the splitting attribute not being the best choice in creating the decision tree, and thus, the classification accuracy of the algorithm is insufficient. Therefore, the amount of out-of-bag data needs to be improved to enhance the accuracy of the feature importance

TABLE 1: Experimental dataset of algorithm performance.

| Dataset | Total number of samples | Number of attributes |
|---------|-------------------------|----------------------|
| Blood | 748 | 5 |
| Iris | 150 | 4 |
| Yeast | 1484 | 9 |

TABLE 2: Comparison of algorithm performance and experimental accuracy values.

| Dataset | Accuracy of random forest classification | Classification accuracy of the KrF algorithm |
|---------|--|--|
| Blood | 0.563 | 0.613 |
| Iris | 0.892 | 0.886 |
| Yeast | 0.795 | 0.823 |

TABLE 3: Out-of-bag data when the best analysis effect is achieved.

| Dataset | Random forest OOB number | Number of OOBs of the KrF algorithm |
|---------|--------------------------|-------------------------------------|
| Blood | 87 | 145 |
| Iris | 36 | 70 |
| Yeast | 439 | 588 |

measure and to improve the classification accuracy of the random forest. In this study, we propose a k -fold random forest (KRF) algorithm by combining the idea of k -fold cross-validation to improve the sample sampling of random forest.

The KRF algorithm flow is described as follows:

Step1: divide the sample set N into K equal parts, draw one set from the K sample sets as out-of-bag data, and draw L training sample sets from the remaining $K-1$ sets by the randomized put-back sampling method, and the number of each sample set is the same as the number of the original sample sets.

Step2: randomly select m attributes ($m < M$) from the attribute set M of the sample and train the splitting of the decision tree based on the best feature variables among the selected variables

Step3: learn the selected L training sample sets, generate one CART for each sample, and generate L decision trees in total

Step4: combine the classification results of L decision trees and determine the final classification result by the simple majority voting method

3.2. Algorithm Analysis

3.2.1. Algorithm Performance Analysis Experiments. In order to not to lose the generality, the performance of the improved algorithm is examined by using the index of the strength of the classification performance of the random forest algorithm as the evaluation criterion. In this study, three different datasets from the unique client identifier (UCI) database which is a collection of databases and data generators are used as experimental datasets (Table 1)

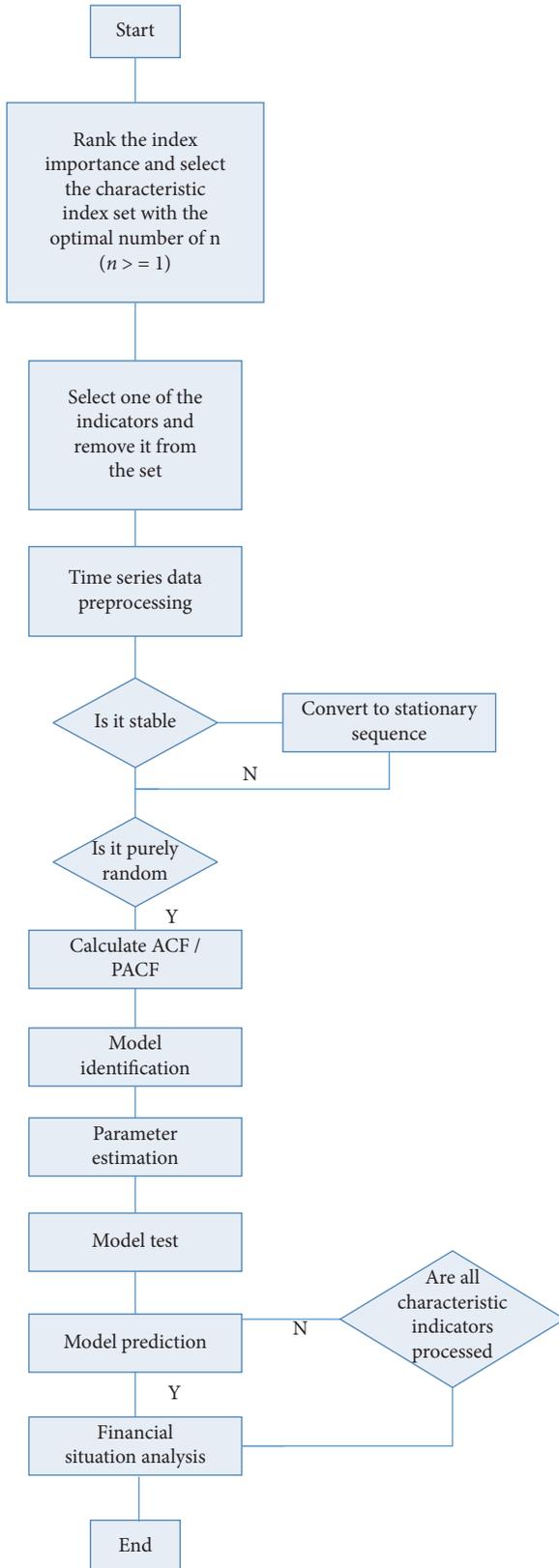


FIGURE 1: Flowchart of the financial crisis early warning algorithm based on random forest and time series.

TABLE 4: Algorithm accuracy.

| Accuracy | Positive class accuracy | Negative class accuracy |
|----------|-------------------------|-------------------------|
| 89.312% | 91.772% | 86.852% |

to simulate the random forest algorithm and the KRF algorithm.

To reflect the real performance of the algorithm more accurately, in the simulation experiment, the experiment of each dataset is repeated 500 times, with 4/5 of the original dataset as the sample dataset and 1/5 as the test set. The result is represented by the mean of 500 experimental results.

3.2.2. Analysis of Experimental Results. After testing on three datasets, the classification accuracy of each algorithm is given in Table 2.

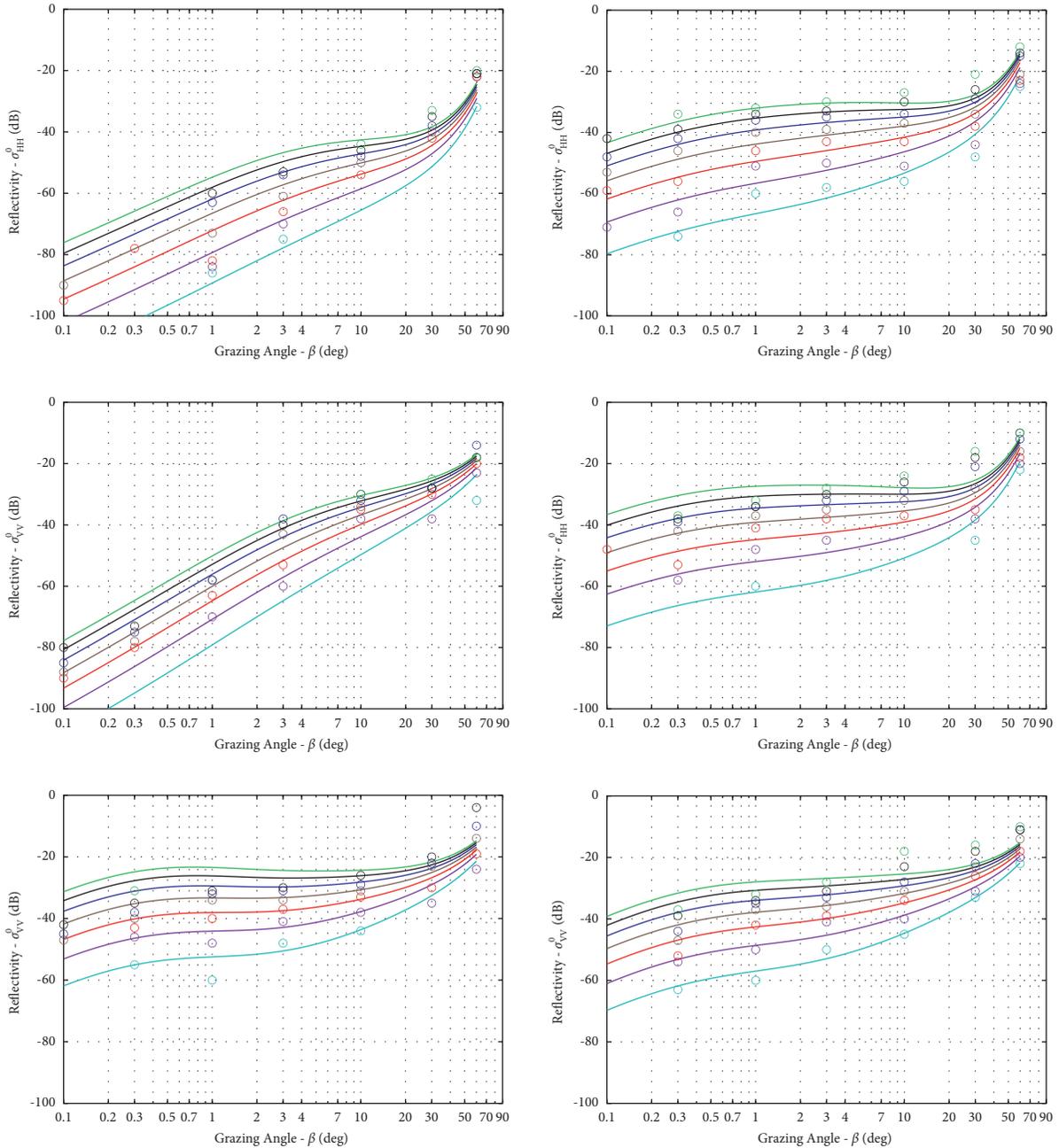
As given in Table 2, the experimental results on the three datasets show that the accuracy of the KrF algorithm is improved compared with the traditional random forest algorithm, which shows that the improved algorithm has better classification performance and that the split attribute selected from the feature set is the optimal result. Out-of-bag data when the best analysis effect is achieved are given in Table 3.

It can be seen from Table 3 that the accuracy of the KrF algorithm is slightly higher than that of the traditional random forest algorithm when the accuracy reaches the highest value. The basis of the k -fold random forest algorithm is reasonable, and it also shows that the characteristic importance measurement result of the KrF algorithm is more accurate.

4. Financial Crisis Early Warning Algorithm Based on Random Forest and Time Series

4.1. Algorithm Flowchart. Usually, in a financial crisis early warning, we should first judge whether the enterprise's finance is in a crisis and finally carry out specific early warning operations according to the judgment results. The early warning instructions are generally issued at the upper layer of the formed system. According to the specific requirements, various early warning operations are made, such as sending warning emails to the management and sending text messages to the responsible personnel. Therefore, this study does not point out what kind of early warning method should be selected but make a judgment on whether it is in a financial crisis. Combined with all the previous research contents, the research ideas of financial crisis early warning are as follows.

First, by improving the random forest algorithm, the k -fold random forest algorithm is proposed. The index importance in the financial data is calculated through the error rate of out-of-pocket data (the direct payment that may or may not be reimbursed later from a 3rd party source), the index importance ranking is realized, and the ranking results



[[PARMS resize(1),pos(50,50),size(200,200),bgcol(156)]]figure 2: Crisis coefficient.

are verified. Finally, the financial status classification experiments are carried out on a different number of index feature sets, and the number of index feature sets with the best classification effect is judged according to the change of accuracy.

Second, according to the steps of time series modeling, the stationarity and randomness of the selected sample data are tested, the nonstationary series (having statistical properties changing with time) are transformed into stationarity (having statistical properties that do not vary in time) by the difference method by taking the differences of consecutive observations, then the stationarity of the difference series is verified again, and the time series model and corresponding order are

determined. Finally, the rationality of the selected model is proved by testing the goodness of the model data fitting.

Third, through multiple modeling, determine the time series model for the selected six financial indicators, respectively, predict the values in the third and fourth quarters of 2018, and finally test the accuracy of the prediction results of the six indicators to prove that the prediction results are effective.

Finally, we continue to classify the prediction results by a k -fold random forest algorithm to judge whether the financial situation of enterprises is good to achieve the purpose of dynamic early warning. The algorithm flowchart is shown in Figure 1.

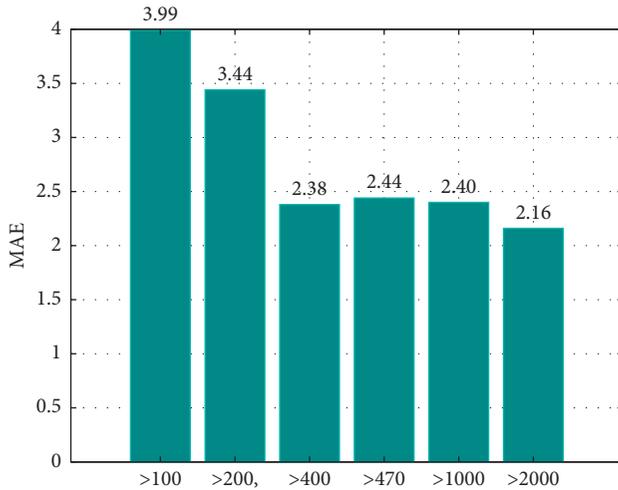


FIGURE 3: Enterprise crisis prediction deviation in multiple stages.

4.2. Experimental Analysis. After completing the modeling and prediction of the time series model, it is still necessary to further improve the prediction results to achieve the purpose of financial crisis early warning. According to the data of the third and fourth quarters of 2018 of the six financial indicators predicted in Section 4.2, combined with the k -fold random forest algorithm, the financial situation analysis is carried out. The accuracy of the financial situation analysis of each enterprise obtained from the experiment is given in Table 4. The results show that the most accurate prediction accuracy of the financial crisis early warning algorithm based on random forest and time series is 89.312%, and the result is reasonable. The prediction ability can be used as a method to judge the financial crisis early warning of enterprises.

5. Experiment

The crisis coefficient is usually a grazing angle (the 90° compliment to angle of incidence) function with the enterprise economy as the parameter, as shown in Figure 2. Such a curve shows a typical behavior, that is, it increases rapidly at a very low sea clutter reflection angle, then in the plateau area at the middle angle, and finally rises rapidly as the grazing angle approaches 90° . To illustrate this behavior of the newly proposed empirical model, some results that have been shown are repeated here, but as a function of the grazing angle. When verifying these results using different distributions, the empirical model is based on different distribution representations, with only two tables at 30° and 60° . Therefore, a better match can be expected at a lower angle.

The enterprise crisis prediction service is evaluated based on reverse test execution. This test has been applied to selected activities according to the criteria, as described in the previous section. In addition, to avoid biased results and overfitting, the whole evaluation process runs in the form of simulation in time. On this basis, the difference between the predicted enterprise crisis and the actual enterprise crisis is the mean absolute error (MAE) [17]. Specifically, MAE measures a set of errors in prediction without considering

their direction, in the sense of the origin of the error. It is the average value of the absolute difference between samples, and the predicted and actual observed individual differences have the same weight. The corresponding mathematical formula is given by the following equation:

$$\text{MAE} = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|. \quad (1)$$

Here, n is the total of the observed values and Y is the actual value and the predicted value. MAE values range from 0 to ∞ , independent of the error direction, as mentioned earlier. These are the results of negative orientation (i.e., the lower the price, the better).

Results are based on the above descriptive analysis, and the average MAE calculates the indicators of multiple activity segments, especially for economic quantities greater than 100, 200, 400, 470, 1000, and 2000, as shown in Figure 3. Figure 3 shows that the minimum error value is about 2.4% points, reaching more than 400 economic activities participated in 40 activities.

6. Conclusions

This study presents an early warning algorithm of enterprise financial crisis based on the combination of the k -fold random forest algorithm and time series analysis model. The algorithm uses the ability of the time series analysis model to make a short-term prediction of historical data and uses the time series analysis model to predict the newly constructed financial index data. Using k -fold random forest to analyze the financial situation of the predicted financial data, the purpose of dynamic financial crisis early warning is realized. The experimental results show that the prediction accuracy of the financial crisis early warning model based on the random forest algorithm and time series is 89%, which shows that the model is effective and feasible.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the General Project of Philosophy and Social Science Research in Colleges and Universities in Jiangsu Province Construction of manufacturing cost management model system based on value chain theory (2020SJA2258).

References

- [1] X. Zhang and W. Dai, "An empirical study of China's financial risk early warning based on PSO-ahp method," *International Journal of Smart Home*, vol. 10, no. 1, pp. 71–80, 2016.

- [2] R. Ye, N. An, Y. Xie, K. Luo, and Y. Lin, "An empirical study on the equity performance of China's health insurance companies during the COVID-19 pandemic—based on cases of dominant listed companies," *Frontiers in Public Health*, vol. 9, Article ID 663189, 2021.
- [3] G. D. Luca, G. Riviuccio, and P. Zuccolotto, "Combining random forest and copula functions: a heuristic approach for selecting assets from a financial crisis perspective," *Intelligent Systems in Accounting, Finance & Management*, vol. 17, no. 2, pp. 91–109, 2010.
- [4] S. O. Kimbrough, C. Chou, Y.-T. Chen, and H. Lin, "On developing indicators with text analytics: exploring concept vectors applied to English and Chinese texts," *Information Systems and E-Business Management*, vol. 12, no. 3, pp. 385–415, 2014.
- [5] G. Dimitrios, C. Ioannis, S. Elias, and S. Koukoulasa, "A Random Forest-Cellular Automata modeling approach to explore future land use/cover change in Attica (Greece), under different socio-economic realities and scales," *Science of The Total Environment*, vol. 646, pp. 320–335, 2018.
- [6] Y. Wei and Y. Zhang, "Study on the relationship between tourist influence and residents' sense of community: an empirical study from changdao county in shandong Province," *Chinese Journal of Population Resources and Environment*, vol. 10, no. 1, pp. 101–104, 2021.
- [7] C. H. Gladwin, "Are their public benefits to private use of fertilizer in Africa," *Sendrom*, vol. 14, no. 4, pp. 120–124, 2002.
- [8] Z. M. Nan and H. S. Ying, "Research of personal loans based on random forests and support vector machines," *Advances in Information Ences & Service Ences*, vol. 4, no. 9, pp. 348–355, 2012.
- [9] P. Appiahene, Y. M. Missah, and U. Najim, "Predicting bank operational efficiency using machine learning algorithm: comparative study of decision tree, random forest, and neural networks," *Advances in Fuzzy Systems*, vol. 2020, no. 5, p. 12, Article ID 8581202, 2020.
- [10] R. Fabling and A. Grimes, "Over the hedge: do exporters practice selective hedging?" *Journal of Futures Markets*, vol. 35, no. 4, pp. 1–3, 2015.
- [11] L. Burlon and M. Vilalta-Buf, "Technical progress, retraining cost and early retirement," *Temi di discussione (Economic working papers)*, vol. 32, no. Supplement s1, pp. 218–238, 2014.
- [12] Z. Tang, Z. Mei, W. Liu, and Y. Xia, "Identification of the key factors affecting Chinese carbon intensity and their historical trends using random forest algorithm," *Journal of Geographical Sciences*, vol. 30, no. 05, pp. 56–69, 2020.
- [13] T. K. Lee, J. H. Cho, D. S. Kwon, and S. Y. Sohn, "Global stock market investment strategies based on financial network indicators using machine learning techniques," *Expert Systems with Applications*, vol. 117, pp. 228–242, 2019.
- [14] E. Ramos-Perez, P. J. Alonso-González, and J. J. Núñez-Velázquez, "Forecasting volatility with a stacked model based on a hybridized Artificial Neural Network," *Expert Systems with Applications*, vol. 129, pp. 1–9, 2019.
- [15] S. E. Rose, K. L. McMahon, A. L. Janke et al., "Diffusion indices on magnetic resonance imaging and neuropsychological performance in amnesic mild cognitive impairment," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 77, no. 10, pp. 1122–1128, 2006.
- [16] A. Cohen and D. Keren, "Individual values and social exchange variables: examining their relationship to and mutual effect on in-role performance and organizational citizenship behavior," *Group & Organization Management*, vol. 33, no. 4, pp. 425–452, 2008.
- [17] Q. Xu, Q. Guo, C. X. Wang et al., "Network differentiation: a computational method of pathogenesis diagnosis in traditional Chinese medicine based on systems science," *Artificial Intelligence in Medicine*, vol. 2021, no. 7724, Article ID 102134, 2021.
- [18] M. Li, H. Jiang, Y. Hao et al., "A systematic review on botany, processing, application, phytochemistry and pharmacological action of Radix Rehmanniae," *Journal of Ethnopharmacology*, vol. 285, Article ID 114820, 2022.
- [19] Y. Tang, X. Li, Y. Chen, Y. Zhong, A. Jiang, and C. Wang, "High-accuracy classification of attention deficit hyperactivity disorder with l2,1-norm linear discriminant analysis and binary hypothesis testing," *IEEE Access*, vol. 8, pp. 56228–56237, 2020.
- [20] K. Cheng, "Smart rural financial innovation based on 5G network and internet of things," *Microprocessors and Microsystems*, vol. 6, no. 3, Article ID 103500, 2020.
- [21] H. Tian, V. Grover, J. Zhao, and Y. Jiang, "The differential impact of types of app innovation on customer evaluation," *Information & Management*, vol. 57, no. 7, Article ID 103358, 2020.
- [22] J. Chen, "Simulation of heterogeneous database encryption algorithm based on B/S structure," *Computer Simulation*, vol. 38, no. 04, pp. 236–239+262, 2021.
- [23] R. Zhang and Z. Hu, "Comparative research on active learning of big aata based on mapreduce and spark," *Microprocessors and Microsystems*, vol. 6, no. 2, Article ID 103425, 2020.
- [24] L. Wang and Y. Wang, "Supply chain financial service management system based on block chain IoT data sharing and edge computing," *AEJ-Alexandria Engineering Journal*, vol. 61, no. 1, pp. 147–158, 2021.
- [25] D. Mozeika, *Integrated Personal Finance Management System for Managing Cash Flow*, Flo Free LLC, Wall, NJ, USA, 2019.