

## Research Article

# Establishment of a Financial Crisis Early Warning System for Domestic Listed Companies Based on Three Decision Tree Models

Gang Wang,<sup>1</sup> Keming Wang ,<sup>2</sup> Yingying Zhou ,<sup>1</sup> and Xiaoyan Mo<sup>3</sup>

<sup>1</sup>School of Business, Guangdong University of Foreign Studies, Guangzhou 510006, China

<sup>2</sup>Center for Accounting, Finance and Institutions, Sun Yat-Sen Business School, Sun Yat-Sen University, Guangzhou 510275, China

<sup>3</sup>School of English and Education, Guangdong University of Foreign Studies, Guangzhou 510006, China

Correspondence should be addressed to Keming Wang; wangkm@mail.sysu.edu.cn

Received 25 May 2020; Accepted 2 June 2020; Published 5 October 2020

Guest Editor: Weilin Xiao

Copyright © 2020 Gang Wang 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 financial crisis is a realistic problem that the general enterprise must encounter in the process of financial management. Due to the impact of the COVID-19 and the Sino-US trade war, domestic companies with unsound financial conditions are at risk of shutdowns and bankruptcies. Therefore, it is urgently needed to study the financial warning of enterprises. In this study, three decision tree models are used to establish the financial crisis early warning system. These three decision tree models include C50, CART, and random forest decision trees. In addition, the ROC curve was used for comprehensive evaluation of the accuracy analysis of the model to confirm the predictive ability of each model. This result can provide reference for domestic financial departments and provide financial management basis for the investing public.

## 1. Introduction

With the continuous improvement of enterprise development capabilities, diversified investment has become an important mode for many enterprises to carry out investment activities, which plays an important supporting role in promoting the continuous development and growth of enterprises. Despite the increasing number and scale of diversified investments by enterprises, there are still some companies that lack awareness of financial risks, resulting in a series of financial risks and the failure of diversified investments. In the current situation of the COVID-19 epidemic and the Sino-US trade war environment, it is particularly important for investors and relevant government departments to grasp the viability of enterprises. This requires enterprises to deeply understand the series of financial risks they may face and take practical and effective measures to vigorously strengthen the early warning and prevention of financial risks and promote the diversified investment of enterprises to achieve tangible results.

Financial crisis early warning models are commonly based on the listed company's financial data, business methods, and published financial information. Most of them use mathematics or statistics to establish models for prediction and analysis, so as to discover the financial crisis in the process of enterprise management, give early warning to the operation and management of listed companies before the crisis, and then, take measures to prevent the crisis from happening. These prevent the crisis from causing harm to enterprises and effectively safeguard the interests of relevant financial departments, enterprises, and the investing public.

The result of financial crisis early warning is not simply whether a financial crisis has occurred. There will be some fraudulent accounting and financial matters in the financial data disclosed by listed companies. The packaging and whitewashing of the company's financial statements will seriously distort the information and make the predicted results inconsistent with the actual results. Like this type of enterprise, although the early warning results do not have the risk of financial crisis, it deserves our special attention.

While most of the current domestic literature use mathematical models to build financial crisis early warning systems, this study uses three decision tree models to establish financial crisis early warning systems. These three decision tree models include C50, CART, and random forest decision trees. In addition, in this study, the ROC curve was used for comprehensive evaluation on the accuracy analysis of the model to confirm the predictive ability of each model. This result can provide reference for domestic financial departments and provide financial management basis for the investing public.

The structure of the following part of the article is as follows: The second part is about the domestic and foreign financial crisis warning-related literature, providing background information for this part. In the third part, this article will discuss three decision tree models, including C50, CART, and random forest decision tree, analyze the financial crisis prediction of 168 listed enterprises, and use modeling and cross-validation methods to test the accuracy of the model. In the last part, conclusions and recommendations will be made for the overall study.

## 2. Literature Review

Since the 1960s, many scholars have conducted in-depth exploration of the prediction model of financial crisis in enterprises, the most famous of which is Altman's *Z*-score model (*Z*-ScoreModel). Our country's research started in 1996. Zhou et al. [1] established an *F*-score model with cash flow. Coincidentally, Yang and Xu et al. [2] also established a *Y*-score model based on the *Z*-score model, using principal component analysis Empirical Research. Zhang [3] used discriminant analysis methods to establish functions to analyze financial crisis enterprises and nonfinancial crisis enterprises. In 2005, Yang and Huang [4] used the BP artificial neural network to establish the early warning model, which increased the prediction accuracy. Xu and Chen [5] summarized the previous results and created four new financial early warning models, including the ideal distance discriminant model, the closest distance discriminant model, the principal component discriminant model of minimum deviation, and the fuzzy comprehensive evaluation model and used empirical evidence. The analysis verifies the abovementioned four models and puts forward the conclusion that the current financial early warning model needs to be combined with qualitative indicators.

In the research of the past three years, in 2018, Wang [6] used principal component analysis and innovative thinking to select 11 characteristic financial ratios with strong *E*-correlation of financial variables to establish a linear model and obtained remarkable results. There are significant differences between the main characteristic variables of financial risk companies and financial health companies. Lin [7] first proposed to take ST companies and non-ST companies with a 1 : 2 ratio as the research object, built a number of financial crisis early warning indicator systems, used nonparametric tests to eliminate nonsignificant indicators, then extracted the principal components as explanatory variables through principal component analysis, so as to

establish a comprehensive evaluation function of the financial situation to divide the financial situation twice, and finally, build an Ologit financial crisis early warning model. Lin [8] found a new way to study the construction and significance of the financial crisis early warning index system for listed companies. Based on the principles of scientificity, effectiveness, and availability, 23 indicators were selected to build a financial crisis early warning system for listed companies, which provided more basis for studying the selection of financial early warning indicators. In the same year, Yang [9] also conducted research on indicators. Based on the existing domestic research results, he selected financial indicators with a high discrimination rate to construct a new financial risk assessment system for listed companies. In terms of model innovation, Ou [10] innovatively used the factor analysis method to build an early warning model of financial risk for real estate companies and, then, conducted financial risk research and analysis on real estate companies. Hu [11] used the *F*-point model to conduct an empirical analysis of listed companies in the construction machinery industry, proving that the *F*-point model is still a more effective early warning method in this industry.

In 2019, in terms of model innovation, Song [12] innovatively used deep learning to build a neural network model for prediction, and the prediction accuracy rate reached more than 72%. Zhang [13] innovatively established the Aalen additive model to predict the financial distress of listed enterprises, especially the time-varying characteristics of influencing factors, based on the negative net profit of enterprises for two consecutive years and the survival analysis method, so as to find the dynamic situation of financial risks of enterprises. In terms of indicator innovation, Xiong [14] added four types of nonfinancial indicators, including earnings management degree indicators, market price indicators, governance structure indicators, and auditor-related indicators on the basis of traditional financial indicators. He reconstructed the financial risk warning model and built a financial risk early warning model based on the Logit model. The research results show that adding nonfinancial indicators to the traditional early warning model can improve the early warning accuracy of the model.

In 2020, in terms of model innovation, Wu [15] took the listed companies on China's GEM as the research object, and used the Twin-SVM to construct the financial crisis early warning model for the nonequilibrium sample characteristics composed of different financial conditions of the enterprise. The model is not only superior in prediction accuracy to other models but also significantly superior in the robustness of prediction. The generalization performance of the two subindustries of manufacturing and information transmission and software and information technology service industry is also significantly superior to the rest several models. Wang [16] developed an early warning model of financial risk of punishment constraint based on cluster analysis. In the empirical analysis of listed enterprises in the *a*-share market, the new model showed good classification effect and robustness. According to the

current development of science and technology, Li [17] proposed that enterprises need to establish a financial risk early warning system that conforms to the development trend of big data. Taking big data and finance as the fundamental entry point, this paper actively finds out the deficiencies of the early warning system of enterprise financial risk under the background of big data and aims to improve it. Lian [18] focused on the analysis of the existing problems of financial risk warning and prevention under the diversified investment of enterprises and drew the conclusion that the focus should be on the innovation of the financial risk warning concept and the improvement of the financial risk prevention system, management mode, and monitoring mechanism.

From the review presented above, it can be seen that most of the current domestic and foreign financial crisis early warning literature use mathematical equations or statistical methods to establish financial crisis early warning systems and compare various predictive capabilities using mathematical equations or statistical methods for analysis. According to the previous literature results, many scholars pointed out that the accuracy of the mathematical model is not high and can only make linear prediction of numerical values, and the prediction results of nonlinear numerical trends are relatively inaccurate. This study adopts the data mining method to establish a financial early warning system with three decision tree models, which is innovative.

### 3. Research Methods and Indicators

Decision tree is a common data mining method. Its basic idea is to continuously make decisions based on certain characteristics according to human thinking and, finally, derive the most suitable classification. Each node represents a sample set with certain characteristics, and the nodes are directly generated according to some characteristics of the samples. However, among the classification prediction methods, decision tree is one of the most typical data mining methods. It is often used in combination with other data mining methods. For example, fuzzy decision tree is a mixed fuzzy theory and decision tree and rough set decision tree is a mixed rough set theory and decision tree. Mixing different methods can improve the classification and prediction ability of the decision tree. The decision tree will establish a tree-like structure, which is composed of root nodes, child nodes, and category leaf nodes. If the decision tree stops growing, each piece of data representing the sample data will be processed, and there will be no unprocessed data.

**3.1. C50 Decision Tree.** C50 decision tree has a wide range of applications and is suitable for discrete and continuous type of numerical or category prediction. On the branch criterion, the category type is based on the information gain ratio, while the continuous type is branched on the basis of variance reduction. In the branch method, categorical variables adopt multivariate branch and continuous variables adopt binary branch.

**3.2. CART Decision Tree.** The application of CART decision tree is also very common. Similar to C50, it is also suitable for discrete and continuous numerical or category prediction. The binary recursive segmentation technique is adopted in the branching criterion. The segmentation of each sample set by CART algorithm is the GINI coefficient calculation, and the smaller the value is, the more reasonable the segmentation is. CART decision tree divides the current sample set into two subsample sets, so that each nonleaf node of the growing decision tree has only binary branches. Therefore, the decision tree generated by CART algorithm is a binary tree with simple structure.

**3.3. Random Forest.** This study mentions a very special decision tree called random forest, which is a classifier that contains multiple decision trees, and its output category is determined by the mode of the category of the individual tree output. The algorithm of the random forest is as follows: First,  $N$  is the number of training samples, and  $M$  is the number of feature variables. Then, the number of feature variables  $m$  is used to determine the decision result of a node on the decision tree, which should be much smaller than  $M$ . Then, we sample  $N$  times (Bootstrap) from  $N$  batches of training data in the way of put back sampling to form a data set and use the undrawn samples to make prediction and evaluate the error. For each node,  $m$  feature variables are randomly selected, and the decision of each node in the decision tree is based on these feature variables. According to the  $m$  feature variables, we calculate the best segmentation method. Note that each tree will grow completely without pruning.

**3.4. Index Selection and Data Sources.** This study uses data from domestic companies listed on the Shanghai Stock Exchange and Shenzhen Stock Exchange, deducting companies with incomplete data and using a two-to-one approach, which is two-thirds of normal enterprise and a third of the crisis, a total of 168 data. These explanatory variables ( $X$ ) data are based on the turnover rate (receivables turnover rate ( $X1$ ) and total assets turnover rate ( $X2$ )), growth rate (main business income growth rate ( $X3$ ) and total asset growth rate ( $X4$ )), and rate of return (return on equity ( $X5$ ) and rate of return on assets ( $X6$ )), and other indicators are collected. The narrative statistics is shown in Table 1. Among them, the crisis enterprise (ST) included by the explained variable ( $Y$ ) refers to the enterprise that has suffered losses for two consecutive years and is specially treated, which is represented by the value of 1 and the normal enterprise by the value of 0. In addition, in data processing, we divided all 168 data into 6 groups of data, and each group of data has 28 data. In the decision tree modeling, 5 groups of 140 data are used to establish a decision tree model, and a group of 28 data is substituted into the decision tree model to test the accuracy of the model for cross validation. Therefore, there were 6 sets of data, that is, the first group was modeled by 1–5 groups and the sixth group was tested. The second group adopted 2–6 group modeling and the first group, testing. The third group adopted 3–1 group modeling and the second

TABLE 1: Sample data description statistics.

	X1	X2	X3	X4	X5	X6	Y
Maximum	5147.7006	9.0348	8748.3663	3702.8692	988.2194	115.8828	1.0000
Minimum	0.2053	0.0223	-73.3733	-45.2444	-72.7279	-42.4491	0.0000
Average	126.6895	0.6820	167.5339	93.5607	145.3563	15.0707	0.3362
Standard deviation	595.8717	1.0842	978.7711	436.6452	194.7686	22.2904	0.4694

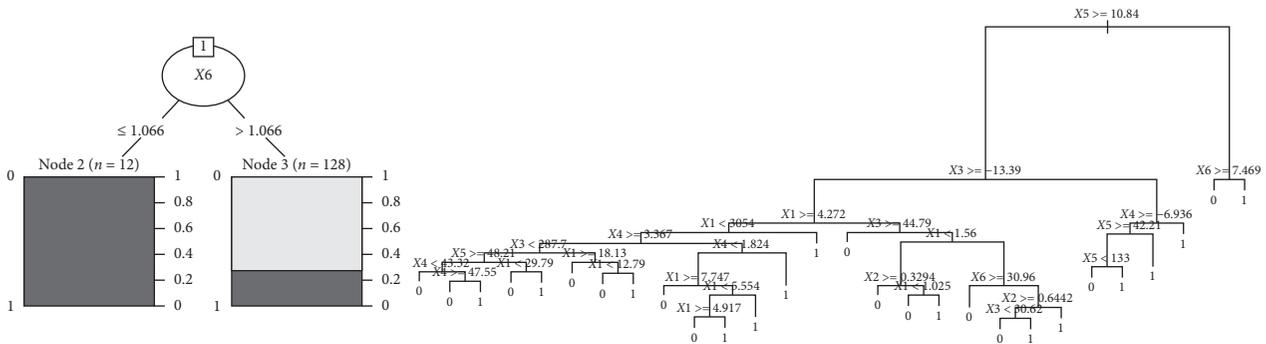


FIGURE 1: The first set of data modeling tree diagram.

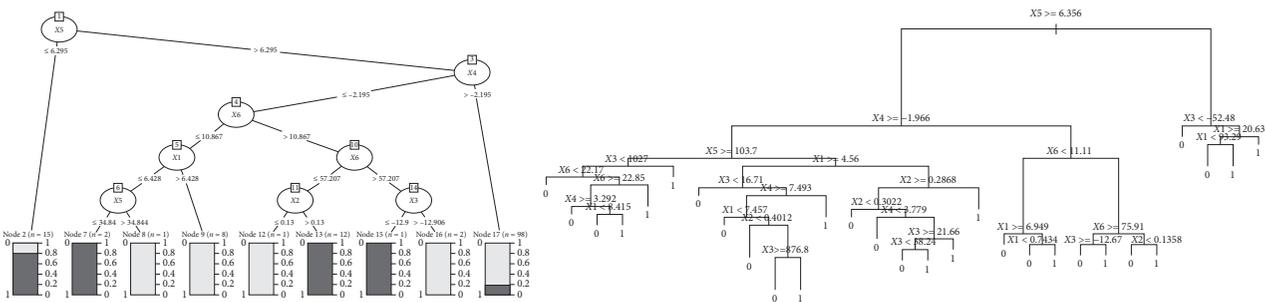


FIGURE 2: The second set of data modeling tree diagram.

group, testing. The fourth group adopted 4–2 group modeling and the third group, testing. The fifth group adopted group 5–3 modeling and group 4, testing. The sixth group adopted the 6–4 group modeling and the fifth group, testing. In this way, 168 pieces of data would be collected under the prediction of the three decision tree models, which could provide us with the accuracy of further analysis of the financial warning model of the three decision trees.

#### 4. Empirical Analysis

4.1. Establishment of the Financial Early Warning Model of Three Decision Trees. In this study, three decision tree models were used to establish the domestic enterprise financial crisis early warning system, and it was divided into 6 groups and 5 groups of data to establish the decision tree model. First of all, we carried out the establishment of three decision tree models, and these models include C50 decision tree, CART decision tree, and random forest. The modeling tree diagram can be seen in Figures 1–6.

Through the tree diagram, we can find that the classification of the decision tree in each group is not the same, but the classification key factors of each model in each group are

generally consistent. X1 appeared as a key factor to create the category 1 round, X5 appeared as a key factor to create the category 2 rounds, and X6 appeared as a key factor to create the category 3 rounds. They are, respectively, corresponding to the accounts receivable turnover ratio (X1), rate of return (return on equity (X5), and return on assets (X6)) and fully show that the three indicators are main factors to measure whether the listed companies are of the financial crisis. The error rate of overall modeling and testing of decision tree is shown in Table 2.

From Table 2, it is found that the classification accuracy of the C50 decision tree is not as good as CART and random forest in all training stages, and the classification error of C50 in the test stage is mostly lower than that of CART and random forest. In other words, during the testing phase, the C50 decision tree performed relatively well. If we ignore the type of model and only look at the sum of the training stage and the test stage, it is found from the table that the classification error of the training stage is lower than that of the test stage. If we observe the sum of the three classification errors, we will find that the classification error of C50 is higher than that of CART and random forest, but the difference is not obvious. Therefore, the ROC curve is needed to further explore the accuracy of the three models.

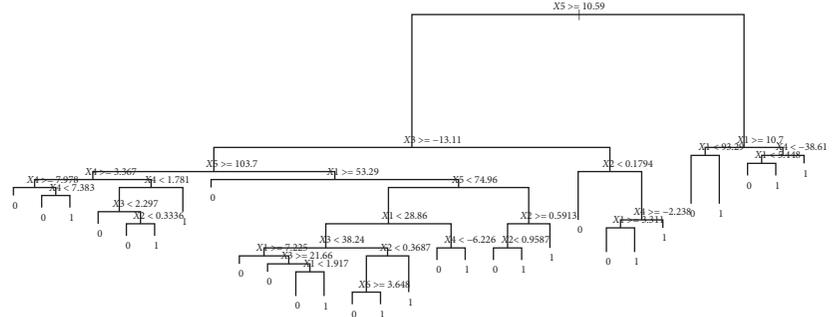
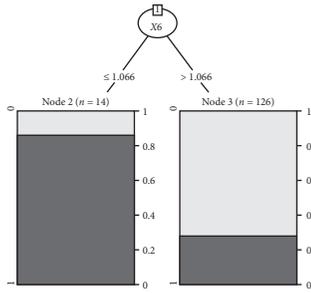


FIGURE 3: The third set of data modeling tree diagram.

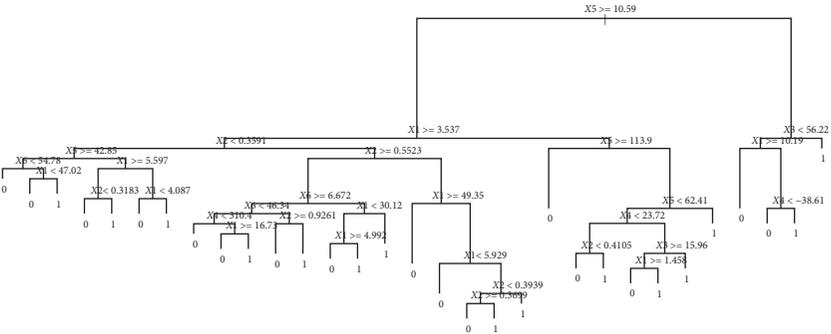
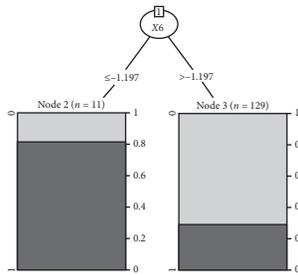


FIGURE 4: The fourth set of data modeling tree diagram.

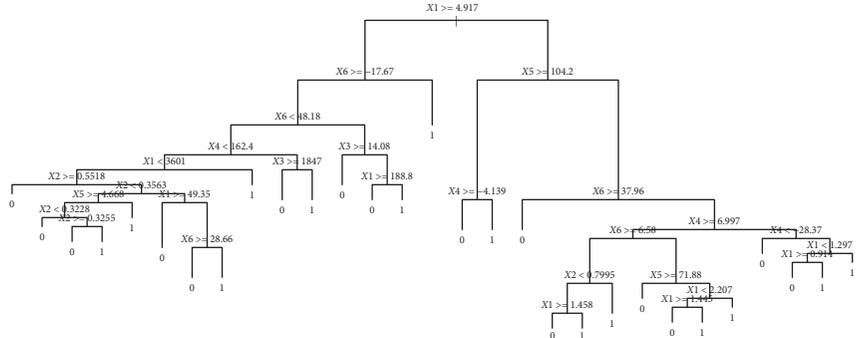
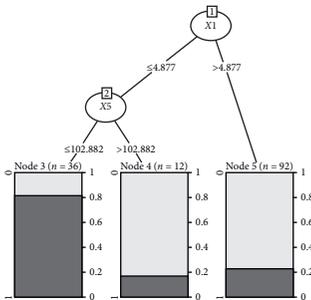


FIGURE 5: The fifth set of data modeling tree diagram.

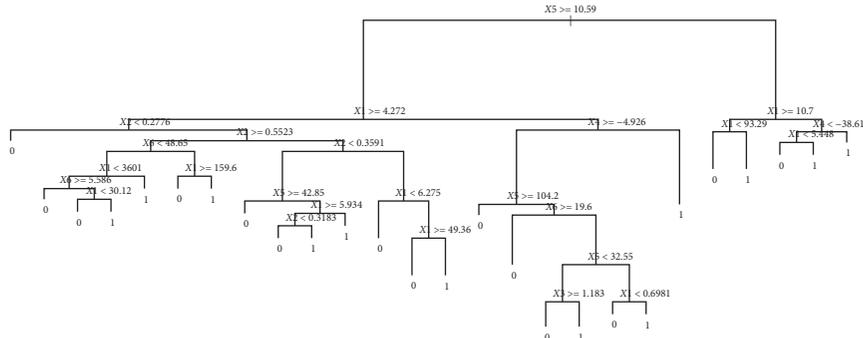
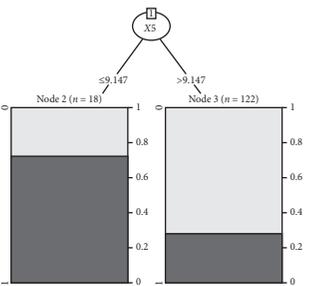


FIGURE 6: The sixth set of data modeling tree diagram.

TABLE 2: Misclassification rate of each group during training and testing.

Group	Model	C50	CART	Random forest	Total
Group 1 to group 5, training and group 6, testing	Training	0.242	0.014	0.021	0.277
	Testing	0.429	0.571	0.464	1.464
Group 2 to group 6, training and group 1, testing	Training	0.164	0	0	0.164
	Testing	0.464	0.571	0.571	1.606
Group 3 to group 1, training and group 2, testing	Training	0.264	0.021	0.014	0.299
	Testing	0.321	0.464	0.321	1.106
Group 4 to group 2, training and group 3, testing	Training	0.279	0.014	0.014	0.307
	Testing	0.285	0.25	0.25	0.785
Group 5 to group 3, training and group, 4 testing	Training	0.25	0.014	0.014	0.278
	Testing	0.321	0.286	0.321	0.928
Group 6 to group 4, training and group 5, testing	Training	0.279	0	0	0.279
	Testing	0.214	0.25	0.286	0.75
Total		3.512	2.455	2.276	8.243

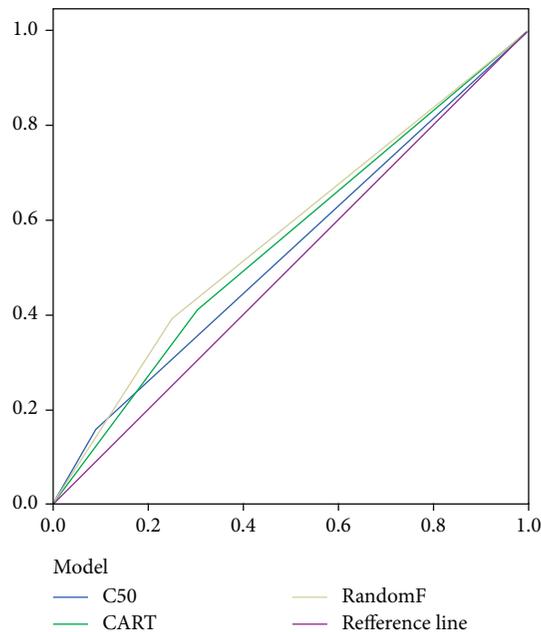


FIGURE 7: ROC curves of classified prediction results of three decision tree early warning models. Note: the x-axis represents specificity and the y-axis represents sensitivity.

**4.2. ROC Curve of Classified Prediction Results.** Figure 7 is the ROC curve of the financial warning model of 168 domestic listed enterprises, which is cross verified by the data of group 6. Bradley (1997) pointed out that the larger the area of the reference line and the area under the curve, the more accurate the classification ability of the model. It is obvious from the figure that the blue line represents the C50 model, the green line represents the CART, the yellow line represents the random forest, and the purple line represents the reference line. The area under the curve of random forest is larger than that of C50 and CART decision tree, which means that the classification ability of random forest is more accurate.

According to the output results of ROC curve analysis in Table 3, the area under the Random F curve (AUC) is 571, which is higher than that of the C50 and CART model, so it has a good ability of early warning and detection.

TABLE 3: Area under curve of three models.

Test result variable	Area	Standard error <sup>a</sup>	Asymp Sig. <sup>b</sup>	95% confidence interval	
				Lower limit	Upper limit
C50	0.536	0.048	0.451	0.441	0.630
CART	0.554	0.048	0.258	0.460	0.647
RandomF	0.571	0.048	0.132	0.478	0.665

Test result variables: C50, CART, and RandomF have at least one knot between the positive and negative actual state groups. Statistics can be skewed.

## 5. Conclusions

The main aim of this research is to discuss how to use decision trees to establish a domestic enterprise financial

early warning model. Especially in the current situation of the COVID-19 epidemic and the Sino-US trade war environment, it is very important for investors and relevant government departments to grasp the viability of enterprises. This study uses the financial data of 168 domestic enterprises in 2016 and divides the data into 6 groups, with 5 groups for modeling, 1 group for testing, and cross-validation to verify the classification accuracy of the 3 decision models. First, it can be concluded from the research results that the random forest model highlights its good classification and prediction ability in comparison. Secondly, after research, it is concluded that the three indicators of accounts receivable turnover rate, return on equity and return on assets are important factors to measure whether a listed company is experiencing financial crisis. In the future, enterprises and investors can attach importance to these indexes when referring to official data.

The disadvantage of this study is that this study only uses three decision tree models, and there are many methods in data mining that are very suitable for building financial early warning models, such as the neural network, Bayesian classification, and especially, the artificial neural network that is with a number of different species such as the fall risk transmission neural network, neural network and support vector machine(SVM), and gray neural network. Therefore, this study suggests that a similar neural network can be used to establish an enterprise financial crisis early warning system in the future.

## Data Availability

All enterprise index 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 study was supported by the National Social Science Fund of China Key Research Project (Project No. 17ZDA086): Research on Reforms and Innovations of Monitoring System in State-Owned Enterprises.

## References

- [1] S. Zhou, J. Yang, and P. Wang, "On the early warning analysis of financial crisis—F-score model," *Accounting Research*, vol. 8, no. 8, pp. 8–11, 1996.
- [2] S. Yang and W. Xu, "Financial early warning model of listed companies—an empirical study of Y score model," *China Soft Science*, vol. 1, no. 1, pp. 56–60, 2003.
- [3] L. Zhang, "Discriminant model for early warning analysis of financial crisis," *Quantitative Economic Technical Economic Research*, vol. 3, no. 3, pp. 49–51, 2000.
- [4] S. Yang and L. Huang, "Listed company financial early warning model based on BP neural network," *System Engineering Theory and Practice*, vol. 1, no. 1, pp. 12–18, 2005.
- [5] W. Xu and D. Chen, "Financial risk early warning modeling principles and several new early warning models," *Statistics and Decision*, vol. 8, no. 8, pp. 150–153, 2016.
- [6] J. Wang, "The establishment of financial early warning mathematical model of listed companies," *Commercial Accounting*, vol. 12, pp. 135–136, 2018.
- [7] Q. Lin, "Research on early warning of financial crisis of listed companies based on ologit model," *Management*, vol. 12, no. 12, pp. 95–97, 2018.
- [8] D. Lin, "Construction and Significance of Early Warning Index System for Financial Crisis of Listed Companies."
- [9] Y. Yang, "Research on information disclosure of listed companies and construction of financial risk early warning system," *Modern Business*, vol. 1, pp. 151–152, 2018.
- [10] G. Ou, "Research on early warning of financial risk of real estate enterprises based on factor analysis method," *Social Scientist*, vol. 9, no. 56, pp. 56–63, 2018.
- [11] S. Hu, "Financial early warning analysis based on F-score model\_taking listed companies in the construction machinery industry as an example," *Financial Management and Capital Operation*, vol. 22, no. 33, pp. 33–36, 2018.
- [12] G. Song, "Research on the financial risk warning model of listed companies based on deep learning," *Value Engineering*, vol. 1, no. 1, pp. 53–56, 2019.
- [13] M. Zhang, "ST forecast of Chinese listed companies based on aalen additive model," *Journal of Systems Management*, vol. 28, no. 1, p. 10, 2019.
- [14] Y. Xiong, "Research on financial risk early warning of listed companies based on F-score," *Risk Management*, vol. 1, no. 27, pp. 111–115, 2019.
- [15] Q. Wu, "Identification and early warning of financial crisis of listed companies on the growth enterprise market," *Finance and Accounting Monthly*, vol. 2, no. 56, pp. 56–64, 2020.
- [16] X. Wang, "Penalty-constrained financial risk early warning model based on cluster analysis," *Finance and Economics*, vol. 2, no. 35, pp. 153–156, 2020.
- [17] D. Li, "Research on early warning of corporate financial risks in the era of big data," *Operation and Management*, vol. 1, pp. 116–117, 2020.
- [18] Y. Lian, "Research on financial risk early warning and prevention under enterprise diversified investment," *Economic Management Space*, vol. 1, pp. 131–132, 2020.