

Research Article

A Novel Financial Risk Early Warning Strategy Based on Decision Tree Algorithm

Lili Tong¹ and Guoliang Tong ²

¹Economic & Management College, West Anhui University, Lu'an 237012, Anhui, China

²Law School, West Anhui University, Lu'an 237012, Anhui, China

Correspondence should be addressed to Guoliang Tong; 10000099@wxc.edu.cn

Received 9 November 2021; Revised 14 December 2021; Accepted 15 December 2021; Published 7 January 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Lili Tong and Guoliang Tong. 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.

This paper requires a lot of assumptions for financial risk, which cannot use all of the data and is often limited to financial data; and in the past, most early warning models for financial crises did not, so they could not track the fluctuation and change trend of financial indicators. A decision tree algorithm model is used to propose a financial risk early warning method. Enterprises have suffered as a result of the financial crisis, and some have even gone bankrupt. Any financial crisis, on the other hand, has a gradual and deteriorating course. As a result, it is critical to track and monitor the company's financial operations so that early warning signs of a financial crisis can be identified and effective measures taken to mitigate the company's business risk. This paper establishes a financial early warning system to predict financial operations using the decision tree algorithm in big data. Operators can take measures to improve their enterprise's operation and prevent the failure of the embryonic stage of the financial crisis, to avoid greater losses after discovering the bud of the enterprise's financial crisis, and to avoid greater losses after discovering the bud of the enterprise's financial crisis. This prediction can be used by banks and other financial institutions to help them make loan decisions and keep track of their loans. Relevant businesses can use this signal to make credit decisions and effectively manage accounts receivable; CPAs can use this early warning information to determine their audit procedures, assess the enterprise's prospects, and reduce audit risk. As a result, the principle of steady operation should guide modern enterprise management. Prepare emergency plans in advance of a business risk or financial crisis to resolve the financial crisis and reduce the financial risk.

1. Introduction

In the fierce market competition environment, many enterprises fall into financial difficulties or even bankruptcy due to poor management [1]. Enterprise financial risk means that enterprises may encounter some problems in the process of financial work, and the existence of these problems is likely to affect the normal operation and development of enterprises [2]. Enterprise financial risk early warning means that enterprises can alert these financial risks before encountering such problems, so as to help enterprises find the signs of possible risks in the process of financial work in advance [3]. In the process of enterprise financial work, financial risk early warning has always been the content that enterprise management leaders attach great

importance to. Once an enterprise has financial risk, it is likely to cause serious losses to the enterprise's economy and seriously lead to the rupture of the enterprise's capital chain, making the enterprise face the risk of bankruptcy [4]. Therefore, it is important to strengthen the early warning of enterprise financial risk [5]. The quality of enterprises is related to the stability and development of the securities market. How to build an effective financial crisis early warning system for various interest groups of enterprises to help enterprises accurately predict financial risks and reduce losses is more important. In the economic field, some scholars have also done some research on financial risk early warning based on decision tree, but the data is simply balanced, and the research on algorithm is not comprehensive enough [6]. However, in the process of enterprise

development, there is often a lack of effective risk early warning mechanism and risk control system scheme. Generally, financial risk problems are analyzed through post disposal, qualitative and static. At present, the research on financial risk early warning system is very lacking.

Decision tree algorithm is one of many data mining algorithms. With the rapid development of information technology [7], the related theory and application are becoming more and more mature. The decision tree does not need to assume a priori probability distribution and has good flexibility and robustness; not only discrete and continuous numerical samples, but also “semantic data” can be used. The generated rule set has simple structure, strong interpretability, and high computational efficiency. It can effectively suppress sample noise and attribute deletion. The disadvantages are as follows: the classification rules are complex and it has overfitting phenomenon.

This paper constructs a financial risk [8] early warning model based on the C4.5 algorithm of decision tree [9], uses the key financial indicators generated by the analysis of financial risk analysis system as the attribute set of decision tree, obtains the preliminary financial risk early warning model by analyzing the sample data of financial indicators, and then adds and integrates the representative branches of decision tree model by post-pruning, repeated experiments are carried out to verify the algorithm model, and final financial risk early warning model is obtained [10].

2. Related Work

Financial risk exposure is frequently the final stage of overall risk exposure, in which all potential problems manifest themselves as financial risk. In fact, data mining [11] can be used to retrieve data that can predict potential financial risks at an early stage. Because potential problems usually manifest themselves in nonfinancial or difficult-to-quantify indicators (such as moral hazard and agency cost), there are currently few studies of this type. In addition to financial indicators, the balanced scorecard concept is based on a set of theories [12].

According to consulting and analyzing a large amount of relevant literature on the occurrence, formation, harm, and treatment of financial risk, this paper makes a comprehensive analysis on the problems existing in the early warning management in the current financial risk management, in order to study the financial risk early warning through more scientific and practical research methods, and make the research results improve in practical application. In terms of processing methods, it mainly studies the theory and method system of financial risk early warning by using the algorithm of decision tree and other relevant methods on the basis of financial analysis theory, category unbalanced analysis theory, and early warning decision theory.

At present, there are many views on the definition of financial risk, and there is still no unified standard [13].

The literature holds that financial distress is the obstruction of enterprises in performing their obligations, which is embodied in four forms: insufficient liquidity, insufficient rights and interests, debt arrears, and insufficient funds [14].

According to the literature, companies in financial distress only include companies that have gone bankrupt, become insolvent, or have been liquidated for the benefit of creditors. It is considered that financial risk not only is defined as bankruptcy, but also includes debt default, bank overspending, failure to pay preferred stock dividends, and so on. The literature holds that bankruptcy is only a single event in a series of potential failures of companies with financial crisis tendency [15, 16].

Financial risk exposure is frequently the final stage of overall risk exposure, in which all potential problems manifest themselves as financial risk [17]. In fact, data mining [11] can be used to retrieve data that can predict potential financial risks at an early stage. Because potential problems usually manifest themselves in nonfinancial or difficult-to-quantify indicators (such as moral hazard and agency cost), there are currently few studies of this type. In addition to financial indicators, the balanced scorecard concept is based on a set of theories [12].

The document [18] is unable to pay the creditor's debts when due and is applied for bankruptcy by the creditor. The creditor's application for bankruptcy does not mean that the company is in fact bankrupt. Bankruptcy must be determined by the court's judgment. Before the court declares bankruptcy, the creditor and the debtor may reach a bankruptcy settlement agreement and reach a settlement agreement on the debtor's delay in paying off debts and reducing the amount of debts. After the court determines, the bankruptcy proceedings can be terminated [19].

The book net assets and shareholders' equity of the company are negative [20]. It shows that the company's assets are less than liabilities, and the risk of failure to repay debts is high; at least it is insolvent on the book. However, it does not mean that the company's financial crisis is serious, because many companies have some assets that have not been reasonably evaluated, and the actual value of these assets will greatly exceed the book value [21]. In any case, as long as the shareholders' equity is negative, it is considered that the company's operation is poor and the financial risk is large. The company is unable to perform the interest and principal repayment of the due contract. Once creditors apply to the court for bankruptcy, the company will fall into bankruptcy crisis [22].

In the application process of financial crisis early warning model, there are usually two kinds of errors: misjudging a crisis company as a healthy company and a stable company as a crisis company. The costs of these two errors are generally different. However, in the current domestic research, we usually assume that the cost of these two types of errors is the same, which will obviously affect the practicability of the model. This paper does not generally define a specific sign such as bankruptcy or financial risk, but classifies the financial risk of listed companies in different periods of operation and development into each link of the early warning chain according to the form of the early warning chain and selects the corresponding financial risk form as the early warning target according to the company's financial status and its specific link in the early warning chain [23].

3. Principle of Correlation Algorithm

3.1. Analysis of Financial Crisis Early Warning and Risk Control Based on Decision Tree. Generally speaking, financial risks include capital structure and cash flow risk, accounting and process risk, and accounting and financial reporting risk. The financial crisis studied in this paper only considers the capital structure and cash flow risk (or liquidity risk) [24]. More specifically, we will analyze from the perspective of the project. If the cash flow of the enterprise is less than a certain critical value, the enterprise is unable to obtain loans in time, or there is insufficient financing, we think the enterprise is in financial crisis. Decision tree is a typical machine learning classification algorithm. Each internal node represents the test of the input attribute, the branch derived from the node represents the output of the test, and each leaf node represents the category or class distribution [25].

Decision tree is a hierarchical structure composed of nodes and directed edges. It generally contains three kinds of nodes [26]:

- (1) Root node: there are no in edges, but there are zero or more out edges
- (2) Internal node: there is exactly one in edge and two or more out edges
- (3) Leaf node or endpoint: there is exactly one in edge but no out edge

Starting from the root node of the tree, the test conditions are used to verify the records, and the appropriate branches are selected according to the test results. Follow this branch to another internal node, use new test conditions, or reach a leaf node. After reaching the leaf node, the class item name represented by the leaf node will be assigned to the inspection record.

3.2. Principle Overview of Decision Tree Algorithm. Using the decision tree information mining method, find the most representative attribute in the sample set [10], construct the node of the decision tree, then construct the node of the decision tree according to the attribute value, and analyze the type of the node. If it is an internal node, the node will continue to construct the next node of the decision tree. This is the establishment of decision tree. S sample data exists in set s , and M representative values are set for attribute standards of different categories. D has different categories X_i ($i = 1, \dots, n$). Let S_i be the number of samples in X . Then, for the sample category, the expected information required for differentiation can be expressed as

$$I(S_1, S_2, \dots, S_m) = - \sum_{i=1}^m p_i \log_2(p_i). \quad (1)$$

In the above formula, P_i is the probability that any sample belongs to a set and is expressed in S_i/s . Let attribute y have v different values $\{A_1, \dots, A_V\}$. S can be divided into v subsets $\{S_1, \dots, S_V\}$ by attribute y , in which the value of

samples in special S_j on attribute y is A_j . Let S_{ij} be the number of samples of class X_i in subset S_j ; then the entropy of attribute y is

$$E(Y) = \sum_{j=1}^v \frac{s_{1j} + \dots + s_{mj}}{s} I(s_{1j}, s_{2j}, \dots, s_{mj}). \quad (2)$$

The highest classification accuracy can be obtained by calculating the minimum value of information entropy. For a given subset S_j , $p_{ij} = s_{ij}/s_j$ is the probability that the sample in S_j belongs to S_i . And the information gain on attribute y is

$$\text{Gain}(Y) = I(s_1, \dots, s_m) - EY. \quad (3)$$

Therefore, the smaller the value of entropy, the greater the information gain. In terms of practical application, ID3 algorithm is a very valuable classification tool. As a typical case of machine learning in the field of artificial intelligence, it has the advantages of simple algorithm compilation and strong classification ability.

The amount of information $I(X_i)$ of the financial risk early warning event X_i can be measured as

$$I(x_i) = - \log_2 p(x_i), \quad (4)$$

where $p(x, I)$ represents the probability of occurrence of event X_i .

Suppose there are m incompatible events $X = \{x_1, x_2, \dots, x_m\}$. If only one event x occurs, the average information can be measured as

$$I(x_1, x_2, \dots, x_m) = \sum_{i=1}^m I(x_i) = - \sum_{i=1}^m P(x_i) \log_2 p(x_i). \quad (5)$$

Assuming that x takes a finite number of variables (x_1, x_2, \dots, x_m), the information entropy of X can be defined as

$$E(x) = - \sum_{i=1}^m P(x_i) \log_a p(x_i). \quad (6)$$

Among them, the logarithmic base a can be a positive task number, usually $a = 2$, and is specified.

The information gain $\text{gain}(x)$ of attribute x for classification is

$$\text{Gain}(X) = E(S) - E(S|X). \quad (7)$$

Financial gain rate is formed on the basis of information gain. Suppose that attribute X has V different values x_1, x_2, \dots, x_V , and the training sample s is divided into v subsets S_1, S_2, \dots, S_V with attribute X . There is a training sample S_j with value X_j on attribute X . It is assumed that the training sample is segmented based on the value of attribute X . Split_Inf is the concept of entropy, expressed as

$$\text{Split_Info}(x) = - \sum_{j=1}^v p(x_j) \log_2 p(x_j). \quad (8)$$

Then, the information gain rate of the attribute can be expressed as

$$Gain_Ratio(X) = \frac{Gain(X)}{Split_Info(X)}. \quad (9)$$

The core of C4.5 algorithm is to generate corresponding test nodes in the decision tree and then divide samples according to attribute values.

4. Analysis of Financial Crisis Early Warning and Risk Control Based on Decision Tree

Table 1 shows the description of the simulated enterprise risk status. The sample data has been discretized based on the

decision tree algorithm. According to the risk level of the financial indicators, each financial indicator is divided into four levels: excellent (RL-1), normal (rl-2), low risk (rl-3), and high crisis (rl-4).

It can be seen from Table 1 that the number of category attribute risk status, stability (s), and crisis (c), are 17 and 3, respectively, so the amount of information $I(s, c)$ of the event can be measured as follows:

$$\begin{aligned} I(s, c) &= -\frac{17}{20} \log_2 \frac{17}{20} - \frac{3}{20} \log_2 \frac{3}{20} = 0.6098, \\ E(AI) &= -\frac{5}{20} \left(\frac{4}{5} \log_2 \frac{4}{5} + \frac{1}{5} \log_2 \frac{1}{5} \right) - \frac{7}{20} \left(\frac{6}{7} \log_2 \frac{6}{7} + \frac{1}{7} \log_2 \frac{1}{7} \right) \\ &\quad - \frac{7}{20} \left(\frac{7}{5} \log_2 \frac{7}{5} + \frac{1}{5} \log_2 \frac{1}{5} \right) - \frac{1}{20} \left(1 \log_2 \frac{4}{5} + 0 \right) = 0.3876, \\ Split_Info(AI) &= -\frac{5}{20} \log_2 \frac{5}{20} - \frac{7}{20} \log_2 \frac{7}{20} - \frac{1}{20} \log_2 \frac{1}{20} = 1.8898, \\ Gain_Ratio(AI) &= \frac{0.6098 - 0.3876}{1.8898} = 0.1176. \end{aligned} \quad (10)$$

Examine the cash flow ratio of the attribute. When the value is good, the stability is three, and the crisis is zero, the branch can obtain the leaf node stability directly. When the value is normal, the stability is 7, and the crisis is 0, the leaf node crisis can be obtained directly by the branch. The branch must be recalculated when the value is low risk, the stability is 7, and the crisis is 1. When the risk is high, the stability is zero, and the crisis is two, the branch can obtain the leaf node crisis directly. Figure 1 depicts a preliminary decision tree.

As can be seen from Figure 1, the attribute cash flow ratio is the root node, and the training sample data is divided into four subsets: excellent, normal, low risk, and high risk. When the cash flow ratio is low risk, the leaf node cannot be

obtained directly. The result is to be determined and needs to be calculated again.

The following will use recursive algorithm to analyze the entropy and information gain rate of other attributes when the value of attribute cash flow ratio is low risk:

$$I(s, c) = -\frac{7}{8} \log_2 \frac{7}{8} - \frac{1}{8} \log_2 \frac{1}{8} = 0.5436. \quad (11)$$

Attribute financial income A1, and the number of instances with values of excellent, normal, low risk, and high risk is 2, 2, 4, and 0, respectively. When the value is excellent, the stability is 1 and the crisis is 1; when the value is normal, the stability is 2 and the crisis is 0; when the value is low risk, the stability is 4 and the crisis is 0. Then,

$$\begin{aligned} E(AI) &= -\frac{2}{8} \left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right) - \frac{2}{8} \left(\frac{2}{2} \log_2 \frac{2}{2} + 0 \right) \\ &\quad - \frac{2}{8} \left(1 \log_2 \frac{7}{5} + 0 \right) - \frac{1}{2} \left(1 \log_2 \frac{5}{5} + 0 \right) = 0.2500, \\ Split_Info(AI) &= -\frac{2}{8} \log_2 \frac{2}{8} - \frac{2}{8} \log_2 \frac{2}{8} - \frac{4}{8} \log_2 \frac{4}{8} = 1.5000, \\ Gain_Ratio(AI) &= \frac{0.5436 - 0.2500}{1.5000} = 0.1957. \end{aligned} \quad (12)$$

TABLE 1: Simulated enterprise risk status.

Stock abbreviation	Financial income a1	Accounts receivable turnover a8	Quick ratio a19	Cash flow ratio a22	Risk status
Stock 1	RI-3	RI-2	RI-2	RI-2	Stable
Stock 2	RI-2	RI-3	RI-3	RI-2	Stable
Stock 3	RI-1	RI-3	RI-2	RI-2	Stable
Stock 4	RI-2	RI-3	RI-3	RI-3	Stable
Stock 5	RI-1	RI-4	RI-3	RI-4	Stable
Stock 6	RI-3	RI-4	RI-4	RI-3	Crisis
Stock 7	RI-2	RI-1	RI-1	RI-1	Crisis
Stock 8	RI-1	RI-3	RI-1	RI-2	Stable
Stock 9	RI-3	RI-4	RI-3	RI-3	Stable
Stock 10	RI-2	RI-2	RI-2	RI-4	Stable
Stock 11	RI-1	RI-3	RI-2	RI-2	Crisis
Stock 12	RI-1	RI-4	RI-2	RI-3	Stable
Stock 13	RI-2	RI-3	RI-1	RI-2	Stable
Stock 14	RI-3	RI-4	RI-2	RI-3	Stable
Stock 15	RI-2	RI-3	RI-1	RI-2	Stable
Stock 16	RI-1	RI-1	RI-1	RI-1	Stable
Stock 17	RI-3	RI-4	RI-2	RI-3	Stable
Stock 18	RI-1	RI-3	RI-4	RI-1	Stable
Stock 19	RI-3	RI-3	RI-3	RI-3	Stable
Stock 20	RI-2	RI-2	RI-2	RI-2	Stable

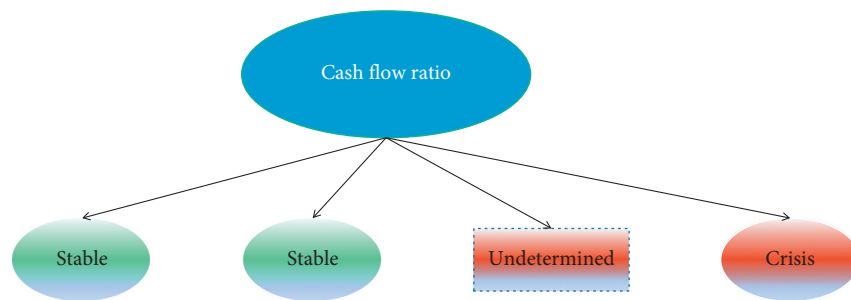


FIGURE 1: C4.5 algorithm initially forms a decision tree.

Through calculation, compared with the attribute information gain rate, it is obvious that the attribute financial income has the largest information gain rate. When the ratio of cash flow ratio is low risk, financial income is selected as the measurement attribute.

Examine the financial income ratio of the attribute. The branch must be recalculated when the value is good, the stability is 1, and the crisis is 1. When the value is normal, the stability is 2, and the crisis is 0, the leaf node crisis can be obtained directly by the branch. The stability is 4 and the crisis is 0 when the value is low risk. Figure 2 depicts the decision tree created in the second step.

As can be seen from Figure 2, the attribute EPS is the inner node, and the training sample data is divided into four subsets: excellent, normal, low risk, and high risk. When the financial income is excellent, the leaf node cannot be obtained directly, and the result is to be determined, which needs to be calculated for the third time.

Using the same method, the entropy and information gain rate of other attributes are analyzed when the value of financial income is excellent. The result is to select the quick ratio as the test attribute. When the value of the test attribute is low risk, the result is stable; for high risk, the

result is crisis. Through the above steps, a complete decision tree is finally obtained, as shown in Figure 3.

The decision tree in Figure 3 uses a specific classification principle to test the data of an unknown mark. If stock X and financial income A1 are excellent, accounts receivable turnover A8 is normal, quick ratio A19 is high risk, and cash flow ratio A22 is low risk; then accounts receivable turnover A8 is normal, quick ratio A19 is high risk, and cash flow ratio A22 is low risk. The risk status is crisis, according to the classification of the constructed decision tree.

If the cash flow ratio is good or normal, the risk situation is stable, according to the decision tree. The risk situation is also stable if the cash flow ratio is low risk and the financial income is normal or low risk. Even if the financial income is excellent, if the cash flow ratio is low risk and the quick ratio is high risk, the risk situation is crisis; if the cash flow ratio is high risk, the risk situation is crisis.

5. Analysis of Experimental Results

In order to verify the classification performance of the decision tree model, first test the data set selected by the decision tree algorithm model, then batch calculate, and

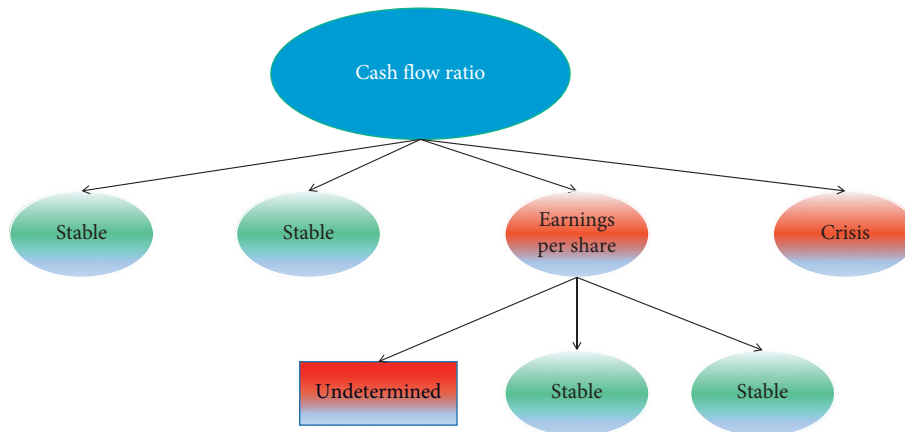


FIGURE 2: Decision tree formed in the second step of C4.5 algorithm.

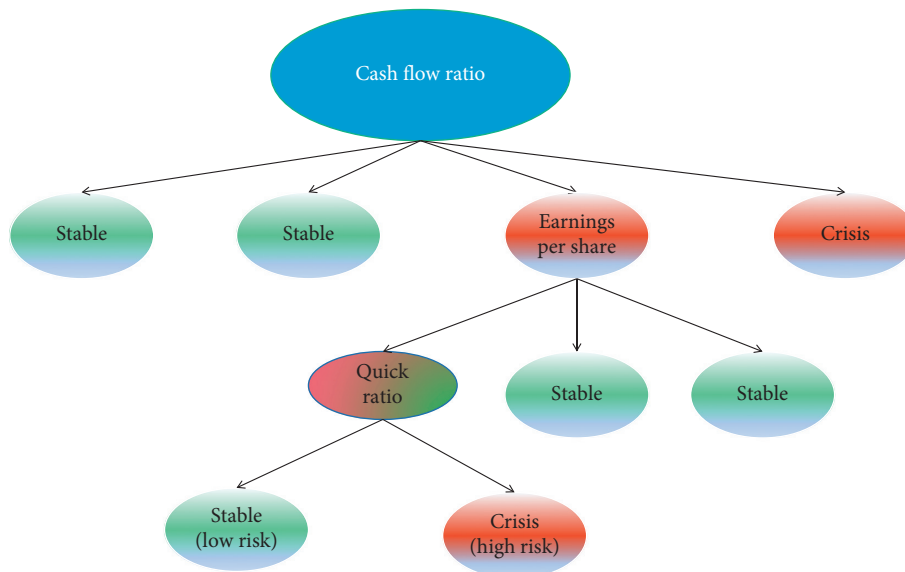


FIGURE 3: C4.5 algorithm forms a complete decision tree.

compare the test results of the combined classifier. The results of each model value of ratio are shown in Figures 4–10:

The statistical line chart of info value shows the split of the five models. Add a trend line to show the line’s approximate trend and the mean value’s approximate range. The number of experiments is represented on the horizontal axis, and the split obtained from each experiment is represented on the vertical axis. The lowest value of the longitudinal axis is set to 0.26 and the highest value is set to 0.34 to cover the split obtained by each model. For the info value and to make it easy to compare the linear distribution, the lowest value is set to 0.26 and the highest value is set to 0.34. The split of ID3 model: the average value of info is the lowest, only about 1.4, and its distribution is also the most dispersed among several models, according to the trend lines of the five figures. The split of the decision tree model has improved as the model has improved. The average value of information has increased slightly, but its distribution is similar to that of the ID3 model.

From the image, compared with the first two, the distribution of the broken line graph of the decision tree model is more concentrated, but its mean value has not improved. In contrast, graph (d) is the split obtained from the decision tree single classifier model. The average value of info (obtained by comparing the Split_Info value with the data obtained under all other sampling rates of the model at 100% sampling rate) is close to 0.32, which is the highest among the first four models, that is, the highest among single classifiers, but it still has little change compared with the dispersion of decision tree. Through the integration of classifiers, the decision tree combines the split of classifiers. The average value of info has been further improved, about 0.33. In terms of discreteness, the g value distribution of the model is very concentrated. It can be seen from the figure that the combined model is more stable. The classification effect of info model is the best, there is little difference between different entropy decision tree models, and the classification effect of

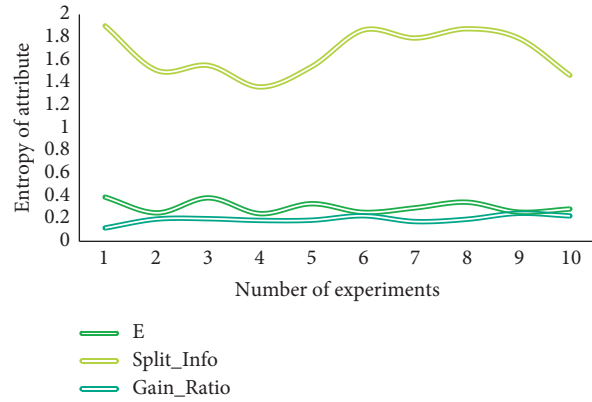


FIGURE 4: Entropy of decision tree model in the first experiment.

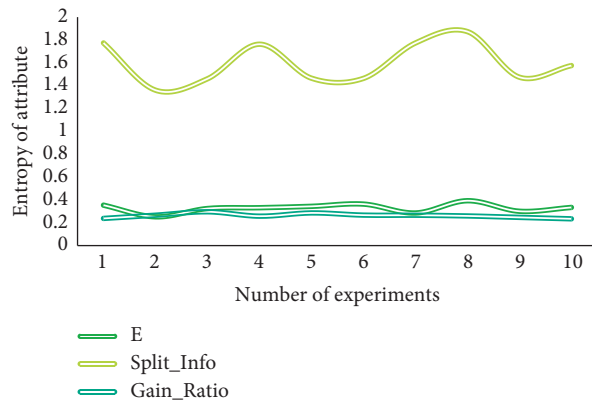


FIGURE 5: Entropy of decision tree model in the second experiment.

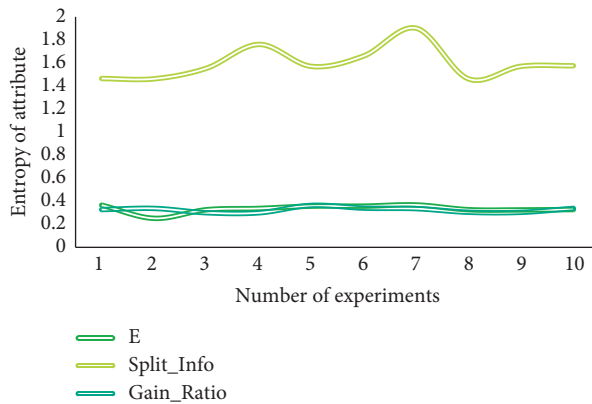


FIGURE 6: Entropy of decision tree model in the third experiment.

decision tree model is the worst. The integration of combined classifiers significantly improves the classification effect of decision tree model, which can show that the decision tree classifier for the classification problem of unbalanced data sets makes the classification effect better.

On the whole, the total classification accuracy of decision tree model is high, but the performance of other data is very poor, which is not ideal for the prediction of unbalanced

bank credit risk [27]; although the distribution of decision tree model in single classifier is relatively concentrated and the classification effect is relatively stable, the low overall classification accuracy makes the model still not ideal. The decision tree combined classifier model of this problem shows great advantages in classification accuracy and stability and has a good effect on the research of credit risk early warning of class unbalanced banks [28].

The results are listed in Tables 2 and 3.

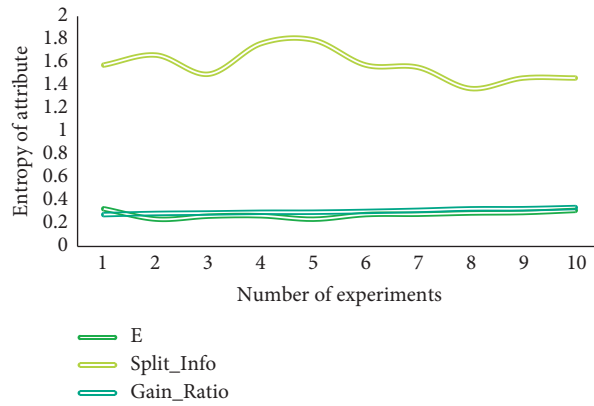


FIGURE 7: Entropy of decision tree model in the fourth experiment.

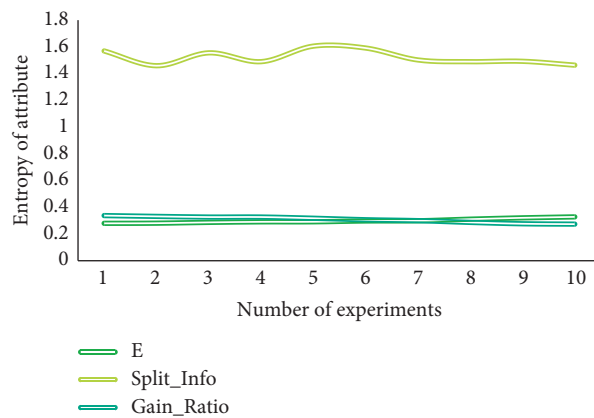


FIGURE 8: Entropy of decision tree model in the fifth experiment.

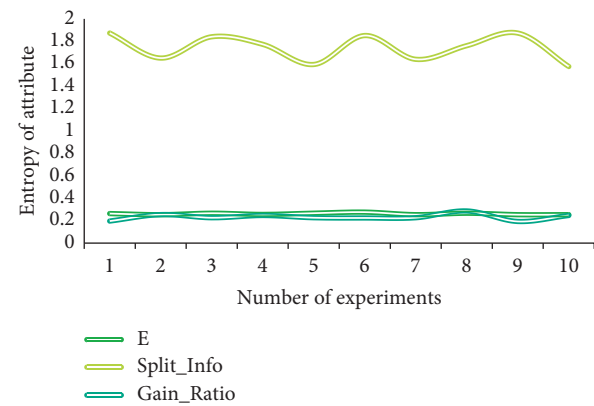


FIGURE 9: Entropy of decision tree model in the fifth experiment.

The mean difference of info is not significant, while the difference of e-means obtained by the decision tree model is not significant. Furthermore, the mean values of any pair of models differ significantly. The difference in info value is not significant, according to the analysis of each pair of models.

The decision tree proposed in this paper is significantly different from other models, particularly the combined decision tree model, demonstrating that the decision tree model proposed in this paper has very superior performance for financial risk early warning data analysis.

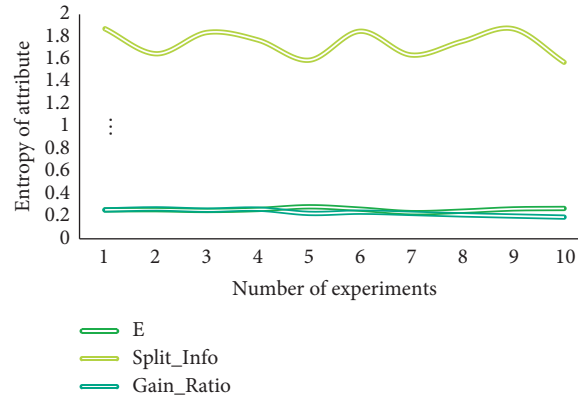


FIGURE 10: Entropy of decision tree model in the fifth experiment.

TABLE 2: Decision tree model Split_Info mean paired sample test.

Entropy model	Risk warning level	ID3	OS-ID3	OUS-ID3
ID3	Stable	0.067	0.046	0.000
OS-ID3	Stable		0.837	0.000
OUS-ID3	Risk			0.000

TABLE 3: Decision tree model Gain Ratio mean paired sample test.

Entropy model	Risk warning level	ID3	OS-ID3	OUS-ID3
ID3	Stable	0.121	0.626	0.000
OS-ID3	Stable		0.157	0.000
OUS-ID3	Risk			0.000

6. Conclusions

People have been unable to use a ratio to obtain enough information to describe all the necessary characteristics of the enterprise, and a single ratio analysis is not only complex, but also difficult for analysts to digest and absorb this information to understand the overall financial situation of the enterprise. By applying the decision tree algorithm financial early warning model to the empirical analysis of enterprises, we can clearly see the effectiveness of the decision tree algorithm early warning model to the Chinese market. The importance of this model is to use multiple variables to obtain information about the financial situation of enterprises. Therefore, it is preferable to combine the financial ratios. In fact, the research on the deterioration of financial situation is not limited to the company. It is more difficult for the company to restructure or reform. It needs a substantial increase in profits to change the current situation. Timely and accurate prediction of financial distress is an essential alarm. Financial risk is the inevitable product of enterprise financial activities, which objectively reflects the operation of enterprises to a certain extent. In the market economy, financial risk exists objectively in every enterprise, especially when the economic system is not perfect, and enterprise financial risk becomes more complex, so the research on enterprise financial risk prevention becomes extremely important.

This paper proposes association rule mining algorithm to analyze enterprise financial risk and decision tree C4.5 algorithm to early warning enterprise financial risk. Although the financial early warning model still lacks a perfect economic theoretical basis and has some problems in methods, it has been highly valued by the practical and academic circles because it plays a good auxiliary role in the decision-making of actual economic activities, even in the current capital market, which is not perfect in all aspects [29].

In the future, the research content of this paper will conduct more in-depth research in the following three aspects:

- (1) Explore more effective decision tree algorithm, and find out the data mining method of difficult-to-quantify indicators such as team cooperation and moral hazard
- (2) Establish a dynamically maintained enterprise financial crisis early warning model according to the time series, mine the financial indicators for a period of time, adjust the desired mining results, find the frequent patterns and rules of financial crisis early warning indicators that meet the requirements, analyze the final mining results, and summarize the real root causes of enterprise financial crisis
- (3) Looking forward to a public enterprise database, the basis of decision tree is to have enough data. Through

the verification of actual data, the research results of this paper will be more meaningful

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] Y. Yang, Z. sun, and J. Lu, "Research on financial risk early warning model based on Z-score index decision tree," *Modern Management*, vol. 7, no. 5, p. 9, 2017.
- [2] J. Liu, L. He, and J. Yuan, "Empirical Study on financial risk early warning of Aerospace Enterprises Based on CHAID model," pp. 60–62, 2021.
- [3] Y. Wang, J. Liu, bingkuan Shi et al., "Research on fund risk early warning based on CHAID algorithm and Bayesian network," *Friends of accounting*, vol. 000, no. 22, pp. 98–102, 2016.
- [4] suting Ye, Y. pan, and Y. Bi, "Research on early warning model of heart disease based on decision tree algorithm. Computer knowledge and technology," *Academic Edition*, vol. 16, no. 19, p. 3.
- [5] P. Yan and J. Wang, "Research on teaching evaluation system based on decision tree algorithm," *Software guide. Educational technology*, vol. 000, no. 9, pp. 19-20, 2009.
- [6] W. Li, wanting Hou, Y. Su et al., "Enterprise financial risk analysis based on data mining," *Hebei enterprise*, no. 11, p. 2, 2020.
- [7] J. Xue, L. Wang, Z. Sun, and C. Xing, "Basic research in diabetic nephropathy health care: a study of the renoprotective mechanism of metformin," *Journal of Medical Systems*, vol. 43, no. 8, p. 266, 2019.
- [8] L. Huang, G. Xie, R. Huang, G. Li, W. Cai, and C. Apostolidis, "Electronic commerce for sustainable rural development: exploring the factors influencing BoPs' ei," *Sustainability*, vol. 13, no. 19, p. 10604, 2021.
- [9] X. Li, Q. Wang, and G. Lin, "Comparative study on enterprise financial early warning model based on machine learning method," *Mall modernization*, vol. 916, no. 7, pp. 156–158, 2020.
- [10] X. Peng, H. Luo, H. Yuan et al., "Research on agreement inventory procurement strategy based on decision tree algorithm," *Jiangxi electric power*, vol. 42, no. 7, pp. 47–50, 2018.
- [11] X. Liu, "Research on China's fiscal risk early warning based on comprehensive improved random forest algorithm," *Computer application and software*, vol. 35, no. 9, pp. 73–78, 2018.
- [12] Z. Huang, Y. Liu, C. Zhan et al., "A novel group recommendation model with two-stage deep learning," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021.
- [13] J. Qiu, "Research on sustainable development strategy of accounting firms -- Based on clustering and decision tree method," *Journal of Nanjing audit institute*, no. 2, pp. 83–89, 2012.
- [14] Y. An, Y. Zhang, and X. Qi, "Application of decision tree algorithm in credit risk early warning of rural banks," *Wireless Internet technology*, vol. 000, no. 1, pp. 163-164, 2015.
- [15] M. Li DeweyP. Han et al., "Research on criminal behavior based on decision tree mining algorithm," *Guangdong Public Security Science and technology*, no. 4, pp. 40–43, 2010.
- [16] H. Hu and H. Hu, "Application of decision tree algorithm in online shopping decision," *Journal of Neijiang Normal University*, no. 2, pp. 24–27, 2015.
- [17] Y. Zeng, "Application of decision tree and support vector machine in corporate financial risk assessment," *Science and Technology of Energetic Materials*, no. 9, pp. 17–19, 2013.
- [18] Y. Qin, Li Ying, Li Li et al., "Financial Risk Analysis and Research of College Students," *Entrepreneurship Projects*, pp. 171-172, 2021.
- [19] Y. Qin, Li Ying, Li Li et al., "Financial risk analysis and research of College Students' entrepreneurship projects," *China market*, no. 7, pp. 171-172, 2015.
- [20] H. Li, "Analysis on the application of decision tree model in university financial early warning system," *Financial and accounting communication: China*, no. 1, p. 2, 2009.
- [21] min Fan, "Analysis and Research on achievement excellence rate based on decision tree algorithm," *Computer Age*, vol. 000, no. 12, pp. 70-71, 2015.
- [22] J. Li and W. Wu, "Research on Evaluation of university scientific research management based on decision tree algorithm," *Science and Technology Bulletin*, vol. 30, no. 3, p. 4, 2014.
- [23] Li Ma and G. Chen, "Research on Application of decision tree algorithm based on data mining," *Agricultural network information*, no. 11, p. 3, 2008.
- [24] J. Liu, L. He, and J. Yuan, "Empirical Study on financial risk early warning of Aerospace Enterprises Based on CHAID model," *Business accounting*, vol. 000, no. 22, pp. 60–62, 2014.
- [25] L. Zhang and X. Jin, "Research on psychological crisis early warning model based on decision tree," *Computer Age*, no. 1, p. 3, 2013.
- [26] Yu Zhang and Z. Zhang, "Research on a customer churn prediction model based on C5," *0 decision tree*, vol. 2021, pp. 89–94, 2015-1.
- [27] homology Lin, "Research on bond default early warning model based on cart bagging algorithm," *National circulation economy*, no. 19, p. 5, 2021.
- [28] C. Xue, "Discussion on CCT score analysis method based on decision tree algorithm," *Contemporary educational practice and teaching research: Electronic Edition*, no. 10, p. 2, 2015.
- [29] H. Wang, "Research on classification algorithm based on decision tree," *Software guide*, no. 17, pp. 136-137, 2007.