Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 8520094, 8 pages
https://doi.org/10.1155/2021/8520094

Research Article

Intelligent Prediction Mathematical Model of Industrial Financial Fraud Based on Data Mining

Xiuqin Geng and Dawei Yang

Shandong Polytechnic, Jinan 250104, China

Correspondence should be addressed to Dawei Yang; 1103954838@qq.com

Received 13 June 2021; Revised 25 June 2021; Accepted 10 July 2021; Published 4 August 2021

Academic Editor: Gengxin Sun

The essence of enterprise financial modeling is to use mathematical models to classify and sort out all kinds of enterprise information according to the main line of value creation and on this basis to complete the analysis, prediction, and value evaluation of enterprise financial situation. A reasonable financial model is also an effective means to reduce financial fraud. In this paper, a financial fraud identification model is constructed based on empirical data. In the process of model construction, the primary feature set is selected according to the financial fraud motivation theory, and then, the original feature set is obtained by Mann–Whitney test on the primary feature set, and the final fraud identification feature set is selected from the original feature set by using Relief and Boruta algorithms. Finally, based on the final fraud identification feature set, the data algorithms such as decision tree, logistic regression, support vector machine, and random forest are used to identify financial fraud. The experimental results show that the combination of financial fraud identification features constructed by the Relief algorithm and random forest model has the best recognition effect. The evaluation indexes of the G mean value and the F value were 75.86% and 78.33%, respectively.

1. Introduction

In the process of enterprise production and operation, enterprise financial management has the function of making enterprise managers understand the operation status of enterprises in time, providing a decision-making basis for enterprise managers, and standardizing the rationality and legitimacy in the process of enterprise operation. With the expansion of the scale of enterprise financial management, we need to change the backward traditional manual way of financial management, through the establishment of financial data-related models and optimization methods to better grasp the business dynamics of enterprises.

Financial fraud [1, 2] is a kind of behavior where the management of a company deliberately manipulates financial information to conceal the true assets and liabilities, operating results, and cash flow of the company in order to achieve some improper purposes and then induces the users of financial statements to make wrong economic decisions based on false financial information. PricewaterhouseCoopers' research report shows that, in recent years, the number of enterprise financial fraud cases has increased year by year. In the past two years, about half of American organizations suffered from financial fraud. In the past few years, the losses caused by financial statement fraud and asset misappropriation in various regions of the world have increased year by year, with a total of about US $3.7 trillion worldwide, and the fraud will cause the company's revenue loss of nearly 5% in the current year.

As an independent third party, auditors are responsible for the reasonable assurance of whether there are material misstatements due to fraud or errors in the financial statements of enterprises. Therefore, improving the ability of auditors to identify financial fraud is of great significance to curb financial fraud and reduce the losses caused by financial fraud. In the digital information environment, data audit has appeared; the new audit mode takes the original data in the audited database as the audit object, establishes the audit

intermediate table through the collection, sorting, and analysis of the original data, and then constructs the model for data analysis by using data mining technologies such as classification, clustering, association analysis, and outlier detection.

In the era of big data, in the face of the explosive growth of data, data mining is more and more widely used by virtue of the ability to find favorable patterns and trends from data sets. Data mining is a process of discovering useful patterns and trends from large data sets. With the deepening of the research on data mining theory, data mining technology has become increasingly diverse. According to the purpose, data mining can be divided into three categories: classification, clustering, and association rule analysis. In audit research, the data mining technology which is often used by the majority of scholars is mainly a classification algorithm.

As an important tool of data processing and information mining, mathematical modeling and data mining are paid more and more attention by audit theory researchers. In the process of modeling, a lot of mathematical theories are involved, such as optimization theory, probability theory, and quantitative statistics. There are many very important mathematical models in accounting and financial management, such as the capital asset pricing model [3], portfolio model [4], securities valuation model [5], and Black-Scholes option pricing model [6]. Dai et al. [7] studied the companies that restate financial statements due to violation of generally accepted accounting principles and found that the sensitivity of option portfolio held by senior managers to stock price has a significant positive correlation with financial statements. Liu et al. [8] studied the correlation between ordinary employees' stock compensation and corporate financial statement fraud. Glancy et al. [1] used the data model to study a large number of financial fraud enterprises and found that the financial fraud enterprises have different degrees of external financing needs during the period of false reporting. By using the data mining method, Albrecht et al. [9] found that, in the period of corporate financial fraud, the number of insiders selling stocks and exercising stock appreciation rights is larger than in other periods. By constructing a mathematical model, Burnes et al. [10] found that the bonus plan is usually directly related to the management's income and has a lower limit; it makes the management easily affected by the game psychology. Khachatryan et al. [11] used an association rule algorithm to study the relationship between financial leverage and corporate fraud and found that financial fraud companies have higher financial leverage than nonfinancial fraud companies.

A large number of studies have found that there is a systematic relationship between financial characteristics and financial fraud. In some cases, financial characteristics are considered to reflect the occurrence mechanism of financial fraud. Louzada et al. [12] proposed a mathematical model to predict the false income. The model found that the percentage change of total assets is positively correlated with the false income, and the percentage change of the number of employees is negatively correlated with the false income. Based on the mathematical model, Tarjo et al. [2] found that when the growth of earnings per share slows down, it shows that they are facing the negative impact of financial performance, and there is a great possibility of financial fraud. Bose et al. [13] used the nonparametric Mann–Whitney test to test 23 financial indicators. The test results showed that 17 financial indicators had significant differences in the samples of judging the possibility of financial report fraud and nonfinancial report fraud. At the same time, the results of Mann–Whitney nonparametric test are further verified by the classification model of logistic regression and artificial neural network algorithm. Sun et al. [14] conducted a t-test on the identification characteristics of fraud risk factors and constructed a fraud identification model by using a support vector machine algorithm and logistic regression analysis.

It can be seen from the above analysis that the mathematical model and data mining technology can extract a large number of information from the financial information and nonfinancial information provided by customers' business activities, which cannot be obtained by the existing audit evidence collection methods. It is of great benefit to improve the audit efficiency and audit effect.

2. Feature Selection of Financial Fraud Identification

The company's fictitious profits, misappropriation of assets, and improper accounting treatment will directly affect the financial statements, which will lead to the abnormality of the company's statement items and various financial indicators calculated according to the statement items. Therefore, some statement items and financial indicators can become the identification attributes of enterprise financial fraud to a certain extent. For those identification attributes that are useful for financial fraud identification, we call them relevant features. Feature selection [15] is the process of selecting relevant feature subsets from the constructed fraud recognition feature set.

Feature selection is an important branch of machine learning. That is, a candidate subset is generated in the initial feature set, and the correlation is evaluated by using the evaluation function. Based on the evaluation results, the next candidate subset is generated, and then the evaluation function is used to evaluate it; the process is repeated until a better feature subset cannot be found.

As shown in Figure 1, feature selection usually takes three processes: firstly, the candidate feature subset is generated by subset search, then the subset goodness-of-fit is evaluated by the selected evaluation function, and finally, a threshold for the evaluation function is set. When the value of the evaluation function reaches the threshold, it can stop searching and output the optimal feature subset.

Subset search is the first key step of feature selection and the process of generating the best candidate subset. According to different subset search patterns, the search algorithm can be divided into a complete search, heuristic search, and random search [16].

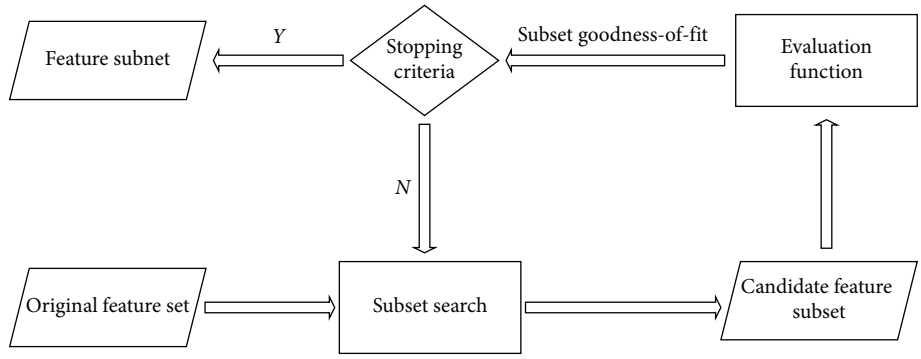

Figure 1: The basic process of feature selection.

The idea of a complete search is to traverse all possible feature subsets in the initial feature set to find feature subsets containing all important information. This method is suitable for the case that the number of features in the original feature set is small.

Heuristic search is divided into the forward search, backward search, bidirectional search, and case-based search. The case-based algorithm mainly includes the Relief algorithm [17]. The algorithm uses distance measure as an evaluation function. Firstly, the user sets a parameter and initializes the weight of each feature to 0, then, an instance is randomly selected from the training set, its near-hit neighbors and near-miss neighbors are calculated according to Euclidean distance, and the weight of each feature is updated by using the near-hit and near-miss neighbors; after all feature weights are calculated, all features whose weights are greater than a certain threshold are selected.

In heuristic search, the forward search algorithm and backward search algorithm often fall into the trap of local optimal value when searching the optimal candidate subset. The random algorithm generates a candidate subset randomly and then performs a forward search algorithm or backward search algorithm on this candidate subset. It makes up for the defect that the forward search algorithm and backward search algorithm cannot jump out of the local optimal value. The random algorithms commonly used in subset search are randomly generated sequence selection algorithm [18], simulated annealing algorithm [19], and genetic algorithm [20].

Subset evaluation is the second key step in feature selection. In this step, the candidate subset goodness-of-fit is evaluated according to the set evaluation function, and whether to stop the subset search process is decided according to the candidate subset goodness-of-fit. The essence of subset evaluation is to evaluate the difference between the current candidate subset's partition of the training data set and the real partition of the training data set. The smaller the difference is, the better the current candidate subset will be. When the difference between the partition of the new candidate subset and the real partition of the training data set reaches the minimum, the candidate subset is stopped. The current candidate subset is the feature subset that contains all the important information.

We introduce the whole idea of the feature selection algorithm from two aspects of subset search and subset evaluation. However, due to the differences in search methods and evaluation methods used in subset search and subset evaluation, feature selection methods can be roughly divided into filtering, packaging, and embedded.

A Relief algorithm is a filtering feature selection algorithm. Firstly, an example x_j is randomly selected from the training data set $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$; the nearest neighbor $x_{j,nh}$ is found from the sample set similar to x_j; it is called a near-hit neighbor. Then, the nearest neighbor $x_{j,nm}$ is found from the sample set different from x_j; it is called near-miss neighbor. Finally, the correlation statistic component of the corresponding attribute u can be calculated as follows:

$$\vartheta^u = \sum_j -\operatorname{diff}\left(x_j^u, x_{j,nh}\right)^2 + \operatorname{diff}\left(x_j^u, x_{j,nm}\right)^2, \quad (1)$$

where x_j^u represents the value of sample x_j on feature attribute u.

The above calculation process is repeated for different samples, and then, the estimation results of different samples are averaged to get the relevant statistical components of different attributes u; the larger the component value is, the stronger the classification ability is. The smaller the component value is, the weaker the classification ability is. When the component value is greater than the set threshold, the feature attribute is selected to classify the target.

Boruta algorithm [21] is a package based on a random forest classification algorithm. Its idea is consistent with the basic idea of random forest classifier; that is, by increasing the randomness of the system and collecting the results from the random sample set, the misleading influence of random fluctuation and correlation can be reduced. Boruta algorithm mainly includes the following steps:

Step 1: copy all real features in the data set to create shadow features

Step 2: randomize the newly added shadow features to eliminate the correlation between them

Step 3: a random forest classifier is trained on the extended data set, and the calculated Z-score is collected

Step 4: find out the maximum Z-score value of the shadow feature, and then record a hit for each real feature whose Z-score is higher than the maximum Z-score value, which means that the real feature is more important than the shadow feature

Step 5: the same bilateral tests as maximum Z-score value are performed for real features whose importance cannot be determined

Step 6: attributes that are less important than shadow features are considered unimportant and are permanently removed from the original set

Step 7: the significant higher importance than shadow feature is regarded as important, and delete all shadow attributes

Step 8: repeat this process until the importance of all attributes is assigned, or the algorithm has reached the previously set limit of random forest operation

According to the existing relevant researches [22, 23], this paper selects the appropriate financial indicators from nine categories of profitability, operating capacity, development capacity, per share index, ratio structure, solvency, risk level, disclosure of financial indicators, and cash flow analysis as the features of financial fraud identification.

The feature selection of financial fraud identification is divided into three steps. First of all, the independent sample T-test and Mann–Whitney test are used to test nine categories of financial indicators. Through these two methods, we can screen out the indicators that can significantly distinguish financial fraud companies from nonfinancial fraud companies. However, the independent sample T-test method requires that the indicators obey the normal distribution. Kolmogorov–Smirnov normality test is carried out for the nine categories of indicators, assuming that all the financial indicators follow the normal distribution, and the test results show that the significant values are less than 0.05, which indicates that these indicators do not follow the normal distribution. Therefore, all indicators are tested by Mann–Whitney test, assuming that there is no significant difference between these nine categories of financial indicators and whether the companies are financial fraud. The test results show that the significant value of 75 indicators is less than 0.05, which shows that these 75 financial indicators have a significant role in identifying financial fraud samples. This paper selects these 75 financial indicators as the original feature set of financial fraud identification. Then, Boruta and Relief feature selection algorithms are used to further screen the primary features to reduce the feature dimension and improve the model adaptability of fraud identification features.

Boruta algorithm divides the primary features into three index sets: Confirmed, Rejected, and Tentative. All 18 indicators in Confirmed are chosen as the identification indicators of the fraud classification model. The result of feature selection is shown in Table 1.

Table 1: The result of feature selection based on the Boruta algorithm.

Feature	Feature name
F1	Interest coverage ratio
F2	Cash flow rate
F3	Asset-liability ratio
F4	Total assets net profit margin
F5	Return on invested capital
F6	Weighted average return on net assets after deducting loss
F7	Ratio of receivables to income
F8	Accounts receivable turnover
F9	Total assets turnover
F10	Growth rate of total profit
F11	Sustainable growth rate
F12	Growth rate of owner's equity
F13	Financial leverage
F14	Comprehensive leverage
F15	Basic earnings per share after deducting nonrecurring profit and loss
F16	Retained earnings per share
F17	Net cash flow from operating activities per share
F18	Comprehensive tax rate

These indicators mainly reflect the company's profitability, operation ability, development ability, debt-paying ability, and cash flow status. From the perspective of solvency, the feature set selected by the Boruta algorithm pays more attention to the solvency of business activities to corporate debt and the solvency of business achievements to interest. From the perspective of operational capacity, the feature set selected by the Boruta algorithm pays more attention to the turnover efficiency of enterprise accounts receivable. From the perspective of development capability, the feature set selected by the Boruta algorithm pays more attention to the internal growth power of enterprises in the future. From the perspective of risk level, the feature set screened by the Boruta algorithm considers that the financial risk and operational risk of an enterprise have an effect on the identification of financial fraud.

In the process of feature selection using the Relief algorithm, the selection threshold is set to 0, the features with weight greater than 0 in the original feature set are retained, and the features with weight less than 0 in the original feature set are discarded. Finally, 17 categories of fraud identification feature indicators are obtained. The result of feature selection is shown in Table 2.

These 17 indicators evaluate the financial situation of enterprises from seven dimensions: solvency, profitability, operation ability, development ability, risk level, per share index, and tax burden. Compared with the feature set screened by the Boruta algorithm, the feature set screened by the Relief algorithm thinks that the interest-paying ability of enterprises is lack of recognition degree to identify financial fraud. According to the Relief algorithm, the management of cost and expense, the loss of asset impairment, and the growth of sales expenses have a better recognition degree for the identification of financial fraud.

Table 2: The result of feature selection based on the Relief algorithm.

Feature	Feature name
F1	Cash flow rate
F2	Equity multiplier
F3	Return on equity
F4	Cost profit margin
F5	Asset impairment loss income ratio
F6	Ratio of accounts receivable to income
F7	Accounts receivable turnover
F8	Business cycle
F9	Total assets turnover
F10	Growth rate of return on equity
F11	Growth rate of total profit
F12	Growth rate of sales expenses
F13	Sustainable growth rate
F14	Comprehensive leverage
F15	Undistributed profit per share
F16	Net cash flow from operating activities per share
F17	Comprehensive tax rate

3. Construction of Industrial Financial Fraud Identification Model

The financial fraud identification model is a two-classification model based on a classification algorithm. The common evaluation indexes for the performance of the model are error rate and accuracy. Error rate refers to the proportion of samples with the wrong classification in the total number of samples, while accuracy refers to the proportion of samples with correct classification in the total number of samples. Although the error rate and accuracy are very common, their practicability is not high. In order to better judge the accuracy of the financial fraud identification model, this paper selects the confusion matrix to evaluate the performance of the model. The confusion matrix [24] is an important tool to evaluate the performance of the classification model. It can reflect the number of correct classification and wrong classification of each category in the sample. For the two-classification task of enterprise financial fraud identification, the combination of the real categories of sample enterprises and the prediction categories of fraud identification model can be divided into true positive (TP), false positive (FP), true negative (TN), and false negative (FN). The confusion matrix of classification results is shown in Table 3.

In Table 3, 1 represents the fraud sample, and 0 represents the nonfraud sample.

According to the confusion matrix, some other indicators, shown in Table 4, are designed to evaluate the classification effect, including accuracy, sensitivity, and specificity.

In addition to the above indicators, F value and G mean value are often used. These two indexes give the comprehensive performance evaluation of the fraud identification model.

F value [25] is a comprehensive consideration of sensitivity and accuracy, and its calculation formula is defined as follows:

Table 3: Structure of confusion matrix.

		Prediction category	
		1	0
Actual category	1	TP	FN
	0	FP	TN

Table 4: Indicators of evaluation metrics based on confusion matrix.

Evaluation metrics	Formula
Accuracy	(TP + TN)/(TP + FP + TN + FN)
Sensitivity	TP/(TP + FN)
Precision	TP/(TP + FP)
Specificity	TN/(TN + FP)

$$F = \frac{\left(\rho^2 + 1\right) \times P \times S}{\rho^2 \times P + S}, \tag{2}$$

where S represents the sensitivity of the model, P represents the precision of the model, and ρ is the parameters for adjusting accuracy and sensitivity weights.

If accuracy and sensitivity are considered equally important, then $\rho = 1$. When evaluating the performance of the fraud model, the larger the F value is, the better the performance of the model is.

G mean value is a comprehensive measure of sensitivity and specificity, and it is also a comprehensive index used to evaluate the performance of the model. Its calculation formula is defined as follows:

$$G = \sqrt{S \times M}, \tag{3}$$

where M represents the specificity of the model.

When evaluating the performance of the fraud model, the larger the G value is, the better the performance of the model is.

Support vector machine, decision tree, logistic regression, and random forest were used to build financial fraud recognition models, and the recognition effects of different models were evaluated.

Based on the CSMAR database, this paper obtains 257 listed companies' consolidated financial statements with fictitious profits or assets from 2010 to 2019 as the fraud samples. At the same time, according to the selection principle of control samples, the corresponding number of control samples is selected according to the ratio of 1 : 1. According to the feature selection of fraud identification samples, this paper preprocesses the original samples and the feature samples filtered by the Boruta algorithm and Relief algorithm. 70% of the preprocessed data set is used as a training set and 30% as a test set.

In the fraud identification model experiment, 5-cross validation is used. Due to the instability of classification model, 10 running results are selected for each classification model. The mean value represents the running result of each model, shown in Table 5.

From the results of the above four fraud recognition models, the original feature set samples have good

Table 5: The mean running result of each model.

Classification model	Specificity (%)	Sensitivity (%)	Accuracy (%)	G (%)	F (%)
Decision tree	69.02	65.01	66.79	66.06	66.01
Logistic regression	68.97	69.98	70.05	68.96	69.12
Random forest	76.30	75.17	76.08	75.31	75.27
Support vector machine	72.18	78.92	76.37	74.96	76.16

recognition results in the support vector machine model. The values of G mean and F reach 74.96% and 76.16%, respectively. However, there are 75 fraud features in the original feature set, which is the main reason for the high efficiency of model recognition. Therefore, it is necessary to reduce the dimension of the original feature set to find the feature set with fewer fraud identification features and better model results.

Because of the large number of dimensions in the original feature set, it is not easy to extract and apply the fraud features. Therefore, the initial feature set selected by the Boruta algorithm is used as the feature set of fraud identification. Each classification model selects 10 running results, and the mean value represents the running result of each model, shown in Table 6.

From the results of the above four fraud recognition models, the feature set samples screened by the Boruta algorithm have good recognition results in the random forest model, and the values of G mean and F reach 74.26% and 74.31%, respectively. The number of fraud identification features in the feature set screened by the Boruta algorithm is reduced from 75 to 18, which reduces the dimension of the original fraud identification feature set. However, the identification effect of the feature set samples screened by the Boruta algorithm is not as good as that of the original fraud identification feature set.

The feature set screened by the Boruta algorithm cannot reduce the dimension of the fraud recognition feature and keep the good recognition efficiency of the fraud recognition model. Therefore, the feature set screened by the Relief algorithm is further used as the feature of fraud identification. Each classification model selects 10 running results; the mean value represents the running result of each model, shown in Table 7.

From the results of the above four fraud identification models, it can be seen that the feature set samples screened by the Relief algorithm have good identification results in the random forest model, and the G mean and F reach 75.86% and 78.33%, respectively. Compared with the original fraud identification feature set, the number of fraud identification features in the feature set screened by the Relief algorithm is reduced from 75 to 17, which reduces the dimension of the original fraud identification feature set; the overall recognition effect of the feature set samples screened by the Relief algorithm is better than that of the original fraud recognition feature set samples.

In conclusion, the random forest model has the best performance among the four fraud identification models. From the perspective of financial fraud identification features, the comprehensive identification performance of the fraud identification features selected by the Relief algorithm in the random forest model reaches 78.33%, which can best reflect the differences between fraudulent enterprises and nonfraudulent enterprises. The contribution of this paper is mainly reflected in two aspects: one is to combine the prior knowledge of fraud identification with feature selection algorithm to select the feature set of financial fraud identification based on the Relief algorithm and Boruta algorithm; the other is to verify the above two kinds of fraud identification features by building a financial fraud identification model. It is found that the set of financial fraud recognition features selected based on the Relief algorithm has the best recognition performance.

4. Discussion

It has rich theoretical and practical significance to study the application of data mining and mathematical model in financial fraud identification. On the one hand, it enriches the theoretical system of financial fraud audit; on the other hand, it provides new ideas and methods for financial fraud audit practice. In this paper, the prior knowledge and feature selection algorithm of financial fraud identification are used to study the characteristics of financial fraud identification, and the financial fraud identification model is established based on data mining technology.

Compared with the existing related research results, the research results of this paper are mainly reflected in two aspects. One is to combine the prior knowledge of fraud identification with a feature selection algorithm to select the feature set suitable for industrial financial fraud identification. The other is to verify two kinds of fraud identification features by constructing the financial fraud identification model.

Through this study, it is found that fraudulent enterprises have weak solvency, high debt risk, and strong willingness to finance, and the cash flow generated by operating activities is lower than other normal enterprises in the same industry. Financial indicators such as cash flow ratio, equity multiplier, and net cash flow per share from operating activities are good features for fraud identification. The assets of fraudulent enterprises are in poor condition and slow turnover, and their profitability and growth ability are lower than other normal enterprises in the same industry. The inventory turnover rate, accounts receivable turnover rate, return on net assets, growth rate of return on net assets, sustainable growth rate, and other financial indicators are good fraud identification characteristics. The growth rate of costs and expenses of fraudulent enterprises is higher than that of normal enterprises in the same industry, and the comprehensive tax burden of enterprises is also lower than

Table 6: The mean running result of each model based on the feature set selected by the Boruta algorithm.

Classification model	Specificity (%)	Sensitivity (%)	Accuracy (%)	G (%)	F (%)
Decision tree	66.32	65.72	66.58	66.25	66.17
Logistic regression	66.02	75.93	71.16	70.76	71.02
Random forest	71.27	76.31	74.38	74.26	74.31
Support vector machine	70.09	73.58	71.96	71.72	71.35

Table 7: The mean running result of each model based on the feature set selected by the Relief algorithm.

Classification model	Specificity (%)	Sensitivity (%)	Accuracy (%)	G (%)	F (%)
Decision tree	62.51	68.02	65.28	64.21	66.78
Logistic regression	80.06	60.99	70.76	69.88	68.92
Random forest	73.72	78.29	75.97	75.86	78.33
Support vector machine	85.07	58.62	63.26	56.37	67.06

that of other normal enterprises. The cost-profit rate, asset impairment loss rate, sales expense growth rate, comprehensive tax rate, and other financial indicators are good fraud identification characteristics.

5. Conclusions

This paper mainly studies the application of mathematical models and data mining technology in financial fraud identification. Based on the prior knowledge and feature selection algorithm of financial fraud identification, the financial fraud identification features are studied, and the financial fraud identification model is established based on logistic regression, decision tree, support vector machine, random forest, and other data mining technologies. This paper attempts to provide an effective analysis and prediction method for auditors to improve their ability to identify fraud risks. In future research, we will focus on the parameter adjustment of the model to further improve the recognition performance of the model.

Data Availability

The basic data used in this paper are downloaded from the online public data set: China Stock Market & Accounting Research Database https://www.gtarsc.com/

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by a grant from Shandong Scientific Research Project of China (no. J18RB132).

References

[1] F. H. Glancy and S. B. Yadav, "A computational model for financial reporting fraud detection," *Decision Support Systems*, vol. 50, no. 3, pp. 595–601, 2011.

[2] A. Tarjo and N. Herawati, "Application of beneish M-score models and data mining to detect financial fraud," *Procedia — Social and Behavioral Sciences*, vol. 211, pp. 924–930, 2015.

[3] V. Vovk and G. Shafer, "The game-theoretic capital asset pricing model," *International Journal of Approximate Reasoning*, vol. 49, no. 1, pp. 175–197, 2014.

[4] Z. Qin, S. Kar, and H. Zheng, "Uncertain portfolio adjusting model using semiabsolute deviation," *Soft Computing*, vol. 20, no. 2, pp. 1–9, 2016.

[5] L. A. Zhe and B. Aba, "Normalized nonconformity measures for automated valuation model," *Expert Systems with Applications*, vol. 180, no. 1, Article ID 115165, 2017.

[6] P. Roul and V. M. K. P. Goura, "A compact finite difference scheme for fractional black-scholes option pricing model," *Applied Numerical Mathematics*, vol. 166, pp. 40–60, 2021.

[7] L. Dai, Z. Fu, and Z. Huang, "Option pricing formulas for uncertain financial market based on the exponential ornstein-uhlenbeck model," *Journal of Intelligent Manufacturing*, vol. 28, no. 3, pp. 597–604, 2017.

[8] Y. Liu and D. S. Wang, "Symmetry analysis of the option pricing model with dividend yield from financial markets," *Applied Mathematics Letters*, vol. 24, no. 4, pp. 481–486, 2011.

[9] C. Albrecht, D. Holland, R. Malagueño, S. Dolan, and S. Tzafrir, "The role of power in financial statement fraud schemes," *Journal of Business Ethics*, vol. 131, no. 4, pp. 803–813, 2015.

[10] D. Burnes, C. R. Henderson, C. Sheppard, R. Zhao, K. Pillemer, and M. S. Lachs, "Prevalence of financial fraud and scams among older adults in the United States: a systematic review and meta-analysis," *American Journal of Public Health*, vol. 107, no. 8, p. 1295, 2017.

[11] D. Khachatryan and B. Muehlmann, "Determinants of successful patent applications to combat financial fraud," *Scientometrics*, vol. 111, no. 3, pp. 1–31, 2017.

[12] F. Louzada and A. Ara, "Bagging k-dependence probabilistic networks: an alternative powerful fraud detection tool," *Expert Systems with Applications*, vol. 39, no. 14, pp. 11583–11592, 2012.

[13] I. Bose, S. Piramuthu, and M. J. Shaw, "Quantitative methods for detection of financial fraud," *Decision Support Systems*, vol. 50, no. 3, pp. 557-558, 2011.

[14] N. Sanaz and S. Mehdi, "Cost-sensitive payment card fraud detection based on dynamic random forest and k-nearest neighbors," *Expert Systems with Applications*, vol. 110, no. 11, pp. 381–392, 2018.

[15] G. Sun and S. Bin, "Router-level internet topology evolution model based on multi-subnet composited complex network model," *Journal of Internet Technology*, vol. 18, no. 6, pp. 1275–1283, 2017.

[16] S. Bin and G. Sun, "Optimal energy resources allocation method of wireless sensor networks for intelligent railway systems," *Sensors*, vol. 20, no. 2, p. 482, 2020.

[17] N. Amjady, A. Daraeepour, and F. Keynia, "Day-ahead electricity price forecasting by modified relief algorithm and hybrid neural network," *IET Generation, Transmission & Distribution*, vol. 4, no. 3, pp. 432–444, 2010.

[18] S. Bin, G. Sun, N. Cao et al., "Collaborative filtering recommendation algorithm based on multi-relationship social network," *Computers, Materials & Continua*, vol. 60, no. 2, pp. 659–674, 2019.

[19] C. Takeang and A. Aurasopon, "Multiple of hybrid lambda iteration and simulated annealing algorithm to solve economic dispatch problem with ramp rate limit and prohibited operating zones," *Journal of Electrical Engineering & Technology*, vol. 14, no. 1, pp. 111–120, 2019.

[20] G. Sun, C. C. Chen, and S. Bin, "Study of cascading failure in multisubnet composite complex networks," *Symmetry*, vol. 13, no. 3, p. 523, 2021.

[21] M. B. Kursa and W. R. Rudnicki, "Feature selection with boruta package," *Journal of Statistical Software*, vol. 36, no. 11, pp. 1–13, 2010.

[22] T. J. Mock and J. L. Turner, "Auditor identification of fraud risk factors and their impact on audit programs," *International Journal of Auditing*, vol. 9, no. 1, pp. 59–77, 2005.

[23] J. Tang, K. E. Karim, and B. Cooper, "Financial fraud detection and big data analytics — implications on auditors' use of fraud brainstorming session," *Managerial Auditing Journal*, vol. 34, no. 3, pp. 324–337, 2019.

[24] D. Simon and D. L. Simon, "Analytic confusion matrix bounds for fault detection and isolation using a sum-of-squared-residuals approach," *IEEE Transactions on Reliability*, vol. 59, no. 2, pp. 287–296, 2010.

[25] G. Tian, S. Zhou, G. Sun, and C. C. Chen, "A novel intelligent recommendation algorithm based on mass diffusion," *Discrete Dynamics in Nature and Society*, vol. 2020, Article ID 4568171, 9 pages, 2020.