Research Article

Research on the Application of DM Technology with RF in Enterprise Financial Audit

Yinhui Hao [ID][1] **and Fuqiang Qiu**[2]

[1]*School of Accounting, Tongling University, Tongling, Anhui 244000, China*
[2]*School of Electrical Engineering, Tongling University, Tongling, Anhui 244000, China*

Correspondence should be addressed to Yinhui Hao; haoyinhui125@126.com

Data mining (DM), as a new technology in the information age, is applied to modern audit work, which is more effective than traditional audit methods. In view of the problems existing in traditional tax audit methods, such as the huge amount of audit data, limited knowledge and experience of auditors, and difficult tracking of audit data, this paper uses computer-aided audit technology to collect, clean up, convert, and analyze data, comprehensively uses data warehouse technology, pattern recognition method, data analysis method, and anomaly detection theory as research methods, and makes a comprehensive study on tax affairs. Then, a random forest (RF) algorithm is used to establish the classification and identification model of audit risk. Second, based on the RF algorithm, the audit early warning framework of accounts receivable and payable in enterprise financial sharing mode is constructed, and the financial data and business data in enterprise financial sharing mode are extracted by using big data technology. The comparison of the results shows that the RF model has higher prediction accuracy and better robustness, which can better improve the antirisk ability of listed companies in China.

1. Introduction

The development of modern information technology has had a great impact on audit: paperless transactions make it difficult to grasp audit clues. Online data processing leads to the shift of internal control evaluation focus. Diversified trading methods lead to changes in modern auditing methods [1]. The modern risk-oriented audit requires auditors to concentrate limited audit resources on high-risk links, which puts higher demands on auditors and audit work, an important way for enterprises of all sizes in China to raise funds [2]. The launch of China's Growth Enterprise Market (GEM) meets the inherent requirements of China's response to the international financial crisis and promotes steady and rapid economic development and meets the inherent needs of accelerating the transformation of economic development mode and cultivating and developing strategic emerging industries. It is a powerful supplement to the main board market [3, 4]. As the main body of the capital market, listed companies' operating conditions directly affect the development of the capital market, so it is of great significance to study the financial early warning of listed companies.

At present, data mining (DM) technology has been widely used in finance, retail, medical care, telecommunications, and other aspects and has become an effective method and way to utilize information resources [5–7]. The emergence of data warehouse and DM has created favorable conditions for solving this problem. Data warehouse is developed based on database. This mechanism of analyzing a large amount of data and making relevant decisions is called online analytical processing. Using modern information technologies such as big data and cloud computing, the types and quantity of audit data collected have changed greatly, which will have a great impact on the traditional audit mode [8]. Combinatorial classification technology in DM is also called the set classification method. Combinatorial classification technology combines several single classification models represented by decision tree (DT) model to form a classifier to classify the same dataset, and the final

classification result is obtained by voting the results of each DT [9]. If the audit risk cannot be effectively identified, it will affect the quality of audit work and make certified public accountants unable to express correct audit opinions, which will undoubtedly reduce the authority of audit reports and trigger the "trust crisis" of the public for third-party audit institutions.

DM, as an advanced information technology, can help auditors screen out representative audit samples in a very short time, which reduces audit risk to a certain extent. Only a comprehensive audit can effectively eliminate the risk of sampling audit, and the application of DM software makes it possible to comprehensively examine the overall data of the audited object [10]. At present, the theory of DM is quite mature, and it is gradually applied to many fields. Random forest (RF) algorithm can classify high-dimensional datasets, select the importance of features, and has high noise tolerance, high operation efficiency, and accuracy, which is convenient and fast. In this paper, to fully consider the openness and dynamics of the financial sharing environment, we studied and solved the key scientific problem of accounts receivable and payable audit early warning in enterprise financial sharing center, used RF theory to construct the audit early warning framework of accounts receivable and payable from the perspective of classification, worked out the implementation framework combined with evidence theory, and verified the correctness and effectiveness of the evaluation method in case analysis.

First, this paper expounds the computer audit model and then constructs an audit risk identification model based on data mining, which is optimized by using random forest algorithm. In the fourth chapter, the actual effect of this algorithm is verified. Finally, the conclusions and suggestions for future development are put forward.

2. Related Work

The process of auditors obtaining audit clues, determining doubtful points, and eliminating and implementing them through data analysis is essentially the process of finding problems and detecting them. DM technology is used to obtain the patterns and rules contained in the data, to discover the anomalies of economic business. Literature [11] has the following definition of audit risk in the academic research of audit procedure: the audit risk refers to the possibility that audit practitioners give inappropriate audit opinions after taking necessary financial audit procedures for the financial statements of the audited units but fail to find the existing substantive errors in their statements. Literature [12] pointed out that audit risk not only includes the risk that the auditors issue inappropriate audit opinions due to misstatement, procedural defects, and other reasons in the traditional sense but also includes the risk of civil and criminal liability caused by this. Literature [13] found that because enterprises are in the market competition, there is a great correlation between price fluctuation and audit risk. Literatures [14, 15] measured the audit risk with audit fees and found that variables such as accounts receivable level,

enterprise assets scale, and number of subsidiaries are significantly correlated with audit risk. In [16], the empirical evidence shows that effective corporate governance can significantly reduce the level of audit risk. Among them, the audit risk is negatively correlated with the nature of equity, and the audit risk is positively correlated with whether the two positions are integrated. Literature [17] uses the fuzzy comprehensive evaluation method to analyze the P2P online lending platform and makes audit strategies accordingly. Literature [18] holds that fuzzy comprehensive evaluation method and other methods are suitable audit risk assessment methods for accounting firms in auditing practice.

By using the association rule analysis technology in DM, there is a certain relationship between the number of fixed assets of the audited entity and daily expenses such as property insurance transfer fees and depreciation expense. Based on this, it is possible to find out whether there is behavior of off-balance sheet assets. The data in the accounting statements of enterprises will change correspondingly with the changes of business operations of enterprises. Usually, the data changes of main items in the accounting statements have certain regularity. If the data changes are abnormal, there may be false elements in the accounting statements. Literature [19] classifies the possible growth evaluation indicators and holds that there are assets, employees, market share, output, profit, and operating income. In [1], the DM method is used for early warning of corporate financial risks. Based on four independent research methods of early warning of financial risks, these methods are combined in different ways, and three mixed models are established, and then, these methods are empirically analyzed. The verification results show that, under the same conditions, the mixed mode is obviously superior to the single method mode. Literature [20] mainly considers the solvency, profitability, and development potential of enterprises and comprehensively evaluates the growth of GEM listed companies through relevant analysis methods. From the perspective of improving the financial level and economic efficiency of enterprises, literature [21]introduces how to choose the financial sharing business according to the actual situation of enterprises and then explains the importance of the financial sharing mode to process transformation and proves this advantage of the financial sharing mode through case analysis. Literature [22] holds that, due to the low rate of correspondence and reply of accounts receivable, the audit of accounts receivable should be an alternative audit procedure, and the main methods and key points needing attention in the alternative audit process are studied. In [23], using Granger causality test to the financial early warning index system, the weight of early warning index is determined, and the financial audit early warning index is designed. It is concluded that China's financial audit is in the middle warning level, among which the direct hidden risk has the greatest influence on the financial audit. Literature [24] proposed that, after constructing DTs, DTs should be sorted according to their classification ability, and DTs with strong classification ability should be selected to construct new RF. In literature [25], using RF algorithm, the common articles in the

author's place are classified, and the experimental results show that this method can achieve good results.

Scholars at home and abroad have done a lot of research on the growth of enterprises, and it is relatively mature to build an index system that affects the growth of enterprises based on financial indicators. RF algorithm can classify high-dimensional data sets, select the importance of features, and has high noise tolerance, high operation efficiency, and accuracy and is convenient and fast, so it is widely used. In view of the financial sharing mode, this paper introduces RF algorithm into the implementation scheme of accounts receivable and payable audit early warning in the financial sharing center of group enterprises and provides feasible operational suggestions for group enterprises.

3. Research Method

3.1. Computer Audit Model. The data needed for audit is often analytical summary data, while the database often stores all kinds of detailed operational data, which results in the phenomenon of rich data but poor information in the database. Too much detailed data stored in the database will not only affect the efficiency of audit analysis but also be unfavorable for auditors to focus on auditing useful information. Most DM tools only analyze the data once and store the analysis results in a separate temporary file. These historical analysis results are rarely used or even not used in the future audit analysis, which is not conducive to the application of historical experience or the storage and update of knowledge.

The audit provided by the traditional computer audit model has the limitation of nonreal-time, and because of its static and difficult adjustment, it greatly limits the development of audit business. Therefore, with the gradual opening of the market and the increasingly fierce market competition, the audit industry urgently needs to establish a new computer audit model.

The audit model designed in this paper can run as an independent system or a subsystem of enterprise internal audit system. Its core functions are mainly composed of three parts: human-computer interaction module, data acquisition module, and prediction analysis module. The functional structure of the system is shown in Figure 1.

Man-machine interaction module is the key part of computer audit model, which runs through the whole process of computer audit, and it affects the use effect of computer audit model. Man-machine interaction module is the interface between users and computers and plays an important role in transmitting commands. Man-machine interaction module is a window of the whole model.

However, DM method can go deep into the underlying data of the audited unit to find out the hidden audit fraud. DM technology includes statistical analysis, neural network analysis, DT analysis, association rules, genetic algorithm, and other methods. Different mining methods can solve different audit problems, and the auditors should choose appropriate mining tools for different audit requirements.

The fundamental task of this subsystem is to obtain the data of the previously mentioned channels widely and
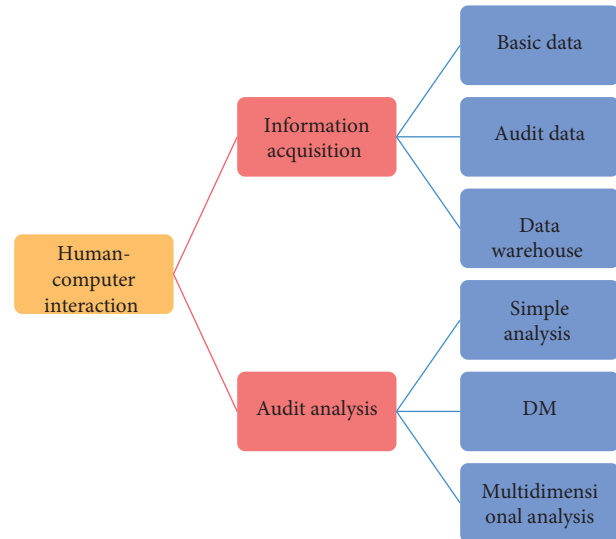

FIGURE 1: Function module of new audit model.

effectively, so that the data information about the audited units can flow into the auditing units continuously, and all kinds of audit information can be obtained widely, quickly, and accurately. At the same time, the data warehouse presents these processed data to the auditors in a multiangle and multilevel manner according to the requirements of the auditors and in various forms such as crosstab and histogram. The data stored in the data warehouse is integrated, consistent, and high-quality data, which brings a lot of convenience to the follow-up audit work.

3.2. Audit Risk Identification Model Based on DM. At present, when faced with audit problems, enterprises are more willing to train internal auditors through recruitment because of the confidentiality of enterprise data. When the internal audit ability of enterprises is insufficient, they will look for the accounting firms in the market to consult and get help. At present, there is no clear and specific civil liability judgment and penalty laws and regulations in the relevant civil laws, which makes it difficult to carry out the criminal implementation of the accounting profession in the process of government accountability, which leads to insufficient inspection and supervision of audit work within the profession. Some auditors with low professional ethics will use legal loopholes to help enterprises avoid taxation, but they violate laws and regulations and cause the loss of national finance.

Usually, accounting firms will adopt the bidding method to undertake audit business. By strengthening the understanding and analysis of undertaking audit business in advance, the mistakes in the audit process can be effectively reduced. In the industry, the audit committee plays the role of coordinating and managing customers and accounting firms, maintaining objectivity in communication to a high degree, and being less interfered by external factors. From the whole industry, the audit quality can be effectively improved, and the audit risk can be reduced through management and communication.

Because there are many financial experts who are familiar with the commonly used analysis indicators, with the powerful computing power of computers, we can draw many financial analysis indicators and measure the overall rationality of the statements, so that certified public accountants can grasp the audited financial statements at a certain height and thus control the audit risk more effectively.

RF model is a theory of statistical learning method. Its basic ideas can also be summarized as follows.

First, the bootstrap resampling method is used to extract n samples ($n < N$) from N original training samples, and the capacity of each sample is the same as the number of samples in the academician training set. Then, n classification and regression trees (CART) DT models are established for n samples to obtain n classification results.

Finally, according to these n classification results, voting records are made and the final classification results are determined. The classification decision is shown as follows:

$$H(x) = \text{argmax}_Y \sum_{i=1}^{k} I(h_i(x) = Y). \qquad (1)$$

$H(x)$ represents the combined classification model, h_i is a single DT classification model, Y represents the output variable (that is, the target variable), and $I(\cdot)$ is an illustrative function.

Outliers usually exist in areas with low density, and the outlier calculation value of some coordinate points is based on the reciprocal of the surrounding density. Density can be defined as the reciprocal of the average distance from n coordinate points to k nearest neighbors, and its calculation formula is shown as follows:

$$\text{Density} = \frac{1}{\sum_{j=1}^{n} \left(|x_i - x_j| + |y_i - y_j| \right)/2n}. \qquad (2)$$

In the formula, n represents the total number of point sets, and the point set is defined as (x_i, y_i).

This function uses the variant of Euclidean distance to find the nearest k neighbors in any of its cases. The variant of Euclidean distance can be applied to data sets containing both nominal variables and numerical variables. The specific calculation formula is as follows:

$$d(x, y) = \sqrt{\sum_{i=1}^{p} \delta_i(x_i, y_i)}, \qquad (3)$$

where the $\delta_i(\cdot)$ function of is the distance between two values of the variable i.

After the audit risk identification is determined to adopt RF method, the samples are input into the identification model for training and identification according to a certain process, which is the main content of the overall model construction process in this subsection. See Figure 2 for details.

The number of variables m selected for each branch decision of DT and the number of DTs n in RF model are two important parameters that affect the performance of the model. These two parameters need to be adjusted timely according to the actual situation. After many experiments, the number of variables m in this paper is selected as 5, and the number of DT is 50 at intervals.

3.3. Application of Enterprise Financial Audit Early Warning Model Based on RF Algorithm.

Financial activities refer to the fund movement in the process of enterprise reproduction, that is, the activities of raising, using and distributing funds. Financial risk refers to the risk that enterprises will suffer economic blow to a certain extent because of unreasonable business activities when they are engaged in financial activities, which will also lead to the competitiveness of enterprises in their industries not as good as before but will be eliminated by the market in the long run.

Whether an enterprise can survive in the industry largely depends on its operational ability. If the mechanism of enterprise management is backward, this defect will affect the future development of the enterprise, causing great harm to the enterprise and even financial crisis. Financial risks of the same nature have many influencing factors, while financial risks of different natures have interaction or correlation, which leads to the complexity and variability of financial risks. Therefore, risks do not exist in a single form, but financial risks of different natures often coexist in the business management and financial activities of enterprises.

Because RF algorithm can handle complicated and diversified data, this paper designs an audit early warning model of accounts receivable and payable based on RF. The audit early warning model mainly realizes the data collection of accounts receivable and payable, evaluates the risk level, and outputs the audit risk early warning signal through the audit early warning model, thus improving the audit efficiency and quality and reducing the audit risk.

Take the designed audit early warning indicators as the object, and the training set is $M = \{M_1, M_2, \cdots, M_n\}$, $M_i = \{M_{i1}, M_{i2}, \cdots, M_{in}\}$, where $n = 12$. The cumulative weight of each feature is calculated as follows:

$$W_j^{i+1} = W_j^i + \frac{di - ff(j, x, p(x))}{m} - \frac{di - ff(j, x, o(x))}{m}. \qquad (4)$$

In formula (1), i represents the sampling sample; j represents the value range of audit early warning features [1, 14]; $o(x)$ represents the nearest neighbor sample of the same kind; $p(x)$ represents the nearest neighbor samples of different classes.

Let the partial correlation degree of function f on variable x_j be the expected value of function f on all other variables except x_j. Then, the partial correlation degree of the function f on x_j is

$$f_j(x_j) = E_{x(-j)}[f(x)]. \qquad (5)$$

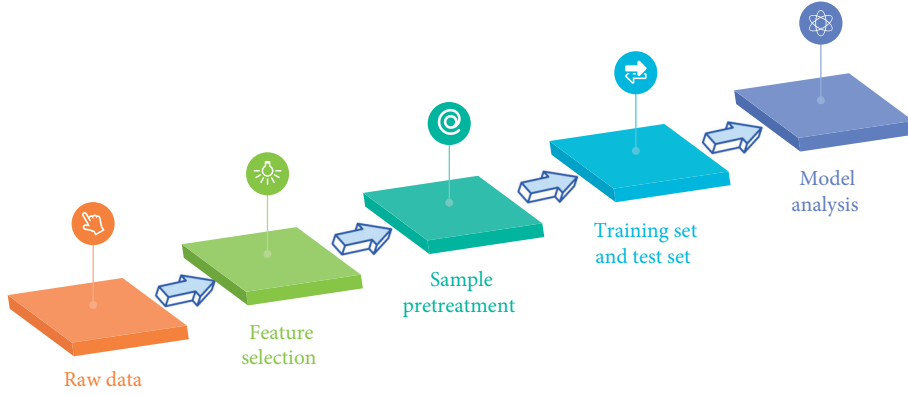Avoid excessive compression of large coefficients, and the penalty function is

FIGURE 2: Overall flowchart of audit risk identification model.

$$P_{MCP}(|\beta|; \lambda, a) = \begin{cases} \lambda|\beta| - \dfrac{|\beta|^2}{2a}, & |\beta| \le a\lambda, \\[3mm] \dfrac{a\lambda^2}{2}, & |\beta| > a\lambda. \end{cases} \tag{6}$$

The parameter $a > 0$ is the undetermined parameter.

It can be seen that, at $|\beta| \le a\lambda$, the first derivative of minimax concave penalties (MCP) function decreases with the increase of $|\beta|$, and the penalty function increases more slowly with the increase of $|\beta|$, $|\beta|$. When $|\beta| = a\lambda$ is used, the first derivative of the penalty function is 0; that is, no penalty is generated for larger coefficients.

Because there are obvious industry differences among listed companies, it is necessary to eliminate the influence of incomparable factors among industries on index values before establishing financial early warning model for listed companies. To make the data comparable, first, each index is standardized by industry. In this paper, bootstrap resampling technique is used to calculate the mean and standard deviation of each index. The forward processing is carried out as follows:

$$y_{ij} = \frac{1}{|x_{ij} - k|}. \tag{7}$$

In the formula, k represents the moderate value, which takes the average value of the j index of the industry where the i-th object is located. After the moderate index is normalized, all indexes are standardized by industry.

The design process of the audit early warning model is as follows (see Figure 3 for details):

(1) The collection of accounts payable information includes not only the branches and subsidiaries but also the historical data of the industry.

(2) The design of index system should guarantee the three principles of comprehensiveness, importance, and desirability of indexes.

(3) The design of audit early warning model should determine the index feature weight, early warning threshold and optimal parameters.

(4) Collect information data to train and test the model, and adjust the early warning threshold and optimized parameters according to the training results.

(5) Use actual information data to test the model.

First, the enterprise financial sharing center establishes the audit early warning index system according to the key concerns of accounts receivable and payable of each subsidiary. Then, we collected audit data, including accounts receivable and payable related information of each branch in the financial sharing center and historical data of the industry. We selected random forest characteristics, optimized the parameters, and set audit early warning threshold again. Finally, the model is tested with actual data. If the result is greater than the threshold range, an audit warning signal will be sent out.

4. Results Analysis and Discussion

After the audit early warning model has been trained and validated, it is only necessary to adjust the optimal parameters and early warning threshold according to the actual situation of the enterprise. At this time, only the authenticity and comprehensiveness of the collected data have great influence on the audit early warning output. The high abstraction and generalization of the inner and essence of data is also the sublimation of the understanding of data from sensibility to rationality. By using DM technology, auditors start from the original data, go deep into the detailed data to find evidence, and through in-depth analysis of the data, find and discover the data rules, to find abnormal phenomena.

Figures 4 and 5 show the parameters of the initial cluster center and the final cluster center in the model based on RF algorithm, in which the smaller the distance between the cluster centers under the same feature indicates that the feature can best represent its classification results; otherwise, the opposite is true.

The numerical difference between the initial cluster center and the final cluster center is very small, which is also closely related to the convergence of the model after one iteration. The more the iterations, the larger the distance between the two, and the smaller the opposite.
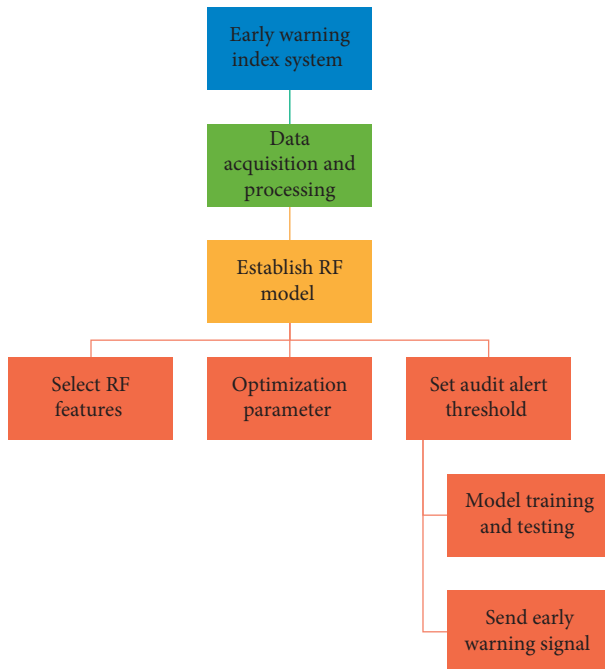
Figure 3: Design process of audit early warning model.

Figure 4: Initial clustering center.

Figure 5: Final clustering center.

Figure 6 shows the data chart of variance analysis based on RF, which shows the difference of clustering index and significance corresponding to each feature of each enterprise under the tax situation.

However, for the significance of the most important observation quantity, see if its value is less than 0.05. If the conditions are met, it can be concluded that the statistic quantity has good variance consistency. Through hypothesis testing requirements, this index can be used as a direct standard to evaluate whether an enterprise is honest or not. Because they will affect the accuracy and effectiveness of the built model to a certain extent and will make the model produce coupling and poor generalization performance, the next research can focus on the in-depth and analysis of this large block.

After training and testing 1285 pieces of data, set 1000 as the DT number of RF identification model and carry out corresponding tests to get Figure 7.

It can be found that, of 786 training samples, 123 samples were incorrectly identified, with an accuracy rate of 84.35%. Of 285 overall samples, 292 samples were incorrectly identified, with an accuracy rate of 79.6%. Of the 37 test samples, 81 samples were incorrectly identified, with an accuracy rate of 77.28%.

As a group of matrices, the confusion matrix has the same number of rows and columns, which can be randomly expanded from two rows and two columns to N rows and N columns. Among them, the column is used to show the real category of samples, and the row shows the classification results predicted by the classification model. By calculating the confusion matrix, the number of correctly classified and wrongly classified samples can be analyzed and calculated, and the number of different classified wrong samples can also be obtained. The confusion matrix table has been
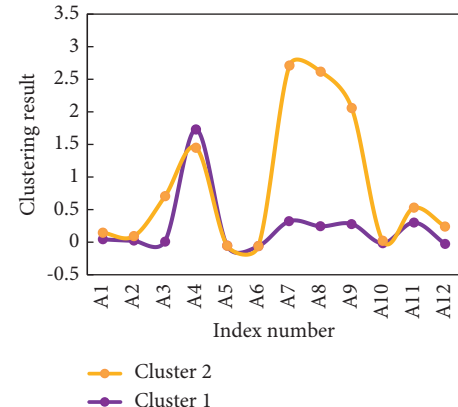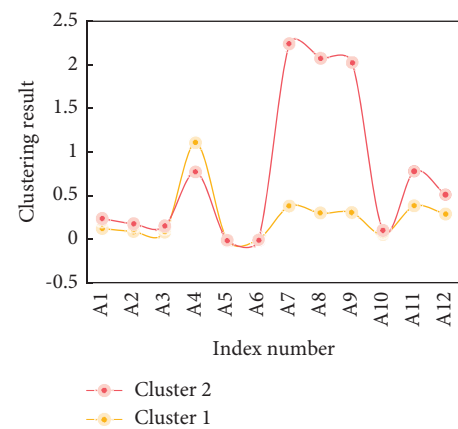
attached to each identified model, and each evaluation index is directly calculated to obtain Figure 8.

As can be seen from Figure 8, the accuracy rates of the three models are all above 80%, which shows that these models are effective in identifying audit risks. The performance is the best among the three models. For audit risk identification, the previously mentioned indicators need to pay special attention to the false negative rate index, which indicates that the model classifies high-risk samples into low-risk samples, which will cause auditors to respond improperly to audit risks, thus issuing wrong audit opinions, leading to audit failure and serious consequences.

This paper randomly selects the accounts receivable and payable audit data of a certain quarter of an enterprise as the forecast set and the accounts receivable and payable audit data of the remaining six quarters as the training set and the accounts receivable and payable audit data of the last quarter as the verification set. In this paper, when the average feature weight of a feature is less than 0.05, the feature will be eliminated. The average feature weight ranking diagram is shown in Figure 9.

From the audit warning interval of each feature in Figure 9, it can be seen that except for the accounts receivable subledger balance, accounts payable balance, other accounts receivable, other accounts payable, accounts
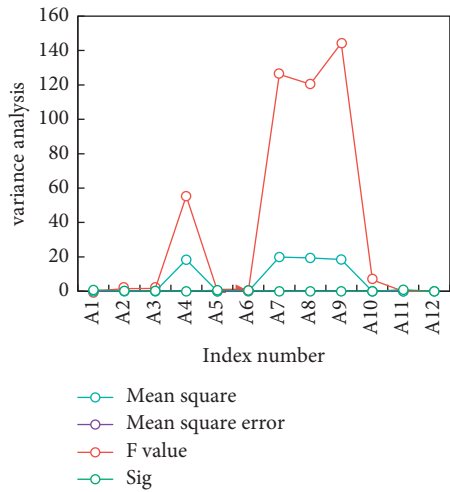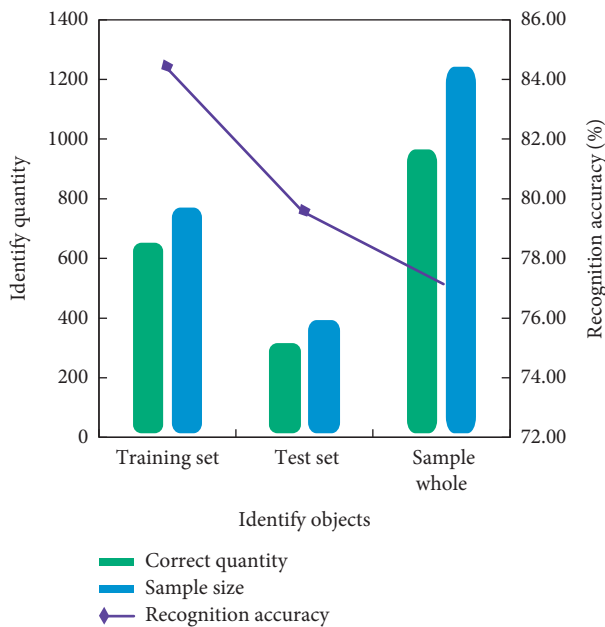
Figure 6: Variance analysis.

Figure 7: Identification result of RF model.

Figure 8: Identification model comparison.

Figure 9: Feature average weight ranking.

receivable aging, and accounts payable aging, these six features are of higher importance, which is consistent with the reality. There may be identification differences between other accounts receivable and other payables, and there may be anomalies between accounts receivable and accounts payable due to their older ages, which are all important points to be paid attention to in the audit process of accounts receivable and accounts payable.

Figure 10 can be obtained by sorting out the judgment accuracy of the test sets of all the above models.

From the comparison, in the prediction of training set or test set, the oversampling RF has little difference with the traditional RF in total judgment accuracy and nonenterprise judgment accuracy. Comparing the RF model with pure financial indicators with the RF model with nonfinancial indicators, it is found that the judgment accuracy of the
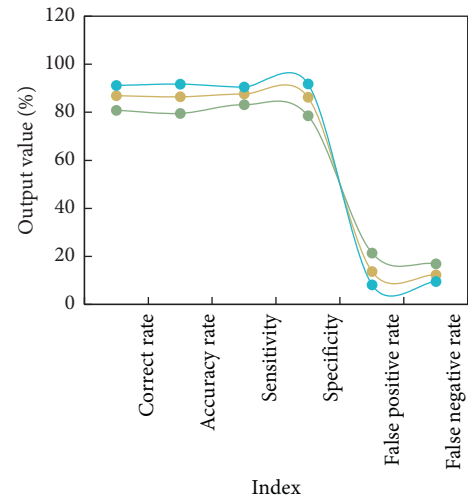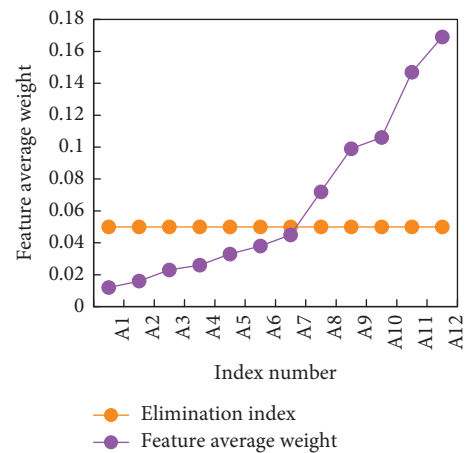
model has been significantly improved after adding nonfinancial indicators.

On the one hand, the innovation and development of industrial enterprises need a good innovation atmosphere and conditions, and the government achieves its goal by publicizing innovation consciousness and organizing corresponding activities, but such measures will take a long time to achieve visible results. On the other hand, talents are the main driving force to lead the innovation and development of industrial enterprises, so the introduction and cultivation of talents are fundamental for the development of enterprises, because talents are related to the management efficiency, production efficiency, and sales ability of the technical development team of enterprises, which can directly or indirectly improve the management level and profitability of enterprises.

Enterprises themselves should take risk prevention measures and improve the internal management
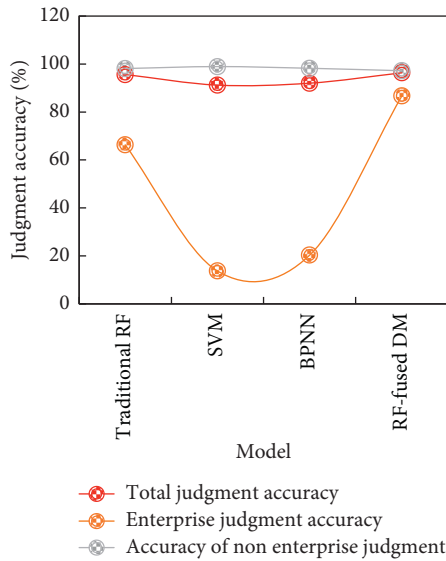
FIGURE 10: Comparison of test set judgment accuracy.

mechanism, to reduce credit risk and make fund suppliers more willing to provide funds, and enterprises can also increase the sources of funds. However, at the same time, they should also control the capital structure to prevent excessive financial leverage, which will lead to debt repayment crisis. The government can also support the capital needs of industrial enterprises by setting up corresponding institutions or designating corresponding policies and guide financial institutions to provide financial services for the real economy, so that financial funds can flow to the real economy departments, so that the financing channels of industrial enterprises can be smoother.

The development and application of DM technology does not require auditors to go to the site all the time to audit, which is the basis of establishing internal audit informatization. In the follow-up audit process, the focus should be on whether relevant departments correct major risk nodes in a timely manner. Otherwise, the responsibility is found, and the reason is analyzed. In addition, follow-up audit work should be guided by the principles of risk orientation and cost-effectiveness, and the risk of enterprises, the feasibility of implementing control measures, and the importance of time arrangement should be analyzed. The abnormal situation of data in the database is found out. By comparing and analyzing the data, it can find abnormal situations in the data, quickly warn, and give corresponding processing suggestions, so that enterprise decision makers can use it in the actual operation and management of enterprises.

5. Conclusion

At present, facing the globalization of information technology, computer audit, as an application result of information technology, is also developing rapidly. Facing the globalization, networking, and paperless environment of information, the information system can only keep pace with the times and be constantly updated, which makes the computer audit have no choice but to constantly enrich and improve itself to minimize the audit risk. The purpose of this paper is to study the application of DM technology in enterprise financial audit through DM technology. The RF identification model of single audit risk identification model is constructed, and the significance of false negative rate in audit risk identification is emphatically analyzed. Among 786 training samples, 123 samples were wrongly identified, with an accuracy rate of 84.35%. Among 285 samples, 292 samples were wrongly identified, with an accuracy rate of 79.6%. Among the 37 test samples, 81 samples were incorrectly identified, with an accuracy rate of 77.28%. The analysis results show that the support vector machine recognition model is the best in every index. Then, the RF algorithm is used to establish the audit early warning model of accounts receivable and payable, and finally, the audit early warning results are fed back. Combined with the characteristics of enterprises, it is necessary to strengthen the supervision and management of accounts receivable and accounts payable, as well as audit and early warning, and improve the quality of internal audit to provide feasible operational suggestions for enterprises.

Follow-up audit work should follow the principle of risk orientation and cost-effectiveness, analyze the risks of enterprises, the feasibility of implementing control measures, and the importance of time arrangement, and find out the abnormal situation of data in the database. By comparing and analyzing the data, we can find the abnormal situation in the data, give a quick warning, and give corresponding treatment suggestions.

Data Availability

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

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] Y. Yang, Y. Cao, and C. Yang, "Innovation method for centralised management and control mode of enterprise financial audit under the background of Industry 4.0," *International Journal of Technology Management*, vol. 84, no. 3/4, p. 197, 2020.

[2] H. Zhou, "Innovation strategy of enterprise's financial audit informatisation in the era of Industry 4.0," *International Journal of Technology Management*, vol. 84, no. 3/4, p. 157, 2020.

[3] N.-T. Kuo and C.-F. Lee, "Investor legal protection, capitalized development costs, and audit fees: a cross-country

analysis," *Journal of International Financial Management & Accounting*, vol. 29, no. 1, pp. 57–82, 2018.

[4] D. Dobija and K. Puawska, "Linking diversity on audit committees and financial reporting quality in Poland," *European Journal of International Management*, vol. 1, no. 1, p. 1, 2019.

[5] L. Li, Y. Feng, and L. Li, "Big data audit based on financial sharing service model," *Journal of Intelligent and Fuzzy Systems*, vol. 39, no. 6, pp. 8997–9005, 2020.

[6] R. Zheng, J. Jiang, X. Hao, and W. F. Y. Ren, "bcBIM: a blockchain-based big data model for bim modification audit and provenance in mobile cloud," *Mathematical Problems in Engineering*, vol. 2019, Article ID 5349538, pp. 1–13, 2019.

[7] Y. Cao and A. R. Y. Wang, "An unexpected cause of pulmonary hypertension," *European Heart Journal - Case Reports*, vol. 4, no. 6, pp. 1-2, 2020.

[8] S. Wu, "Internet public informatioan text data mining and intelligence influence analysis for user intent understanding," *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 4, pp. 1–8, 2019.

[9] O. Bogma, O. Vialets, L. Dukhnovska, and N. H. Klymash, "Automated control system as a tool for ensuring financial and economic security of the enterprise," *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*, vol. 5, no. 6, pp. 142–147, 2020.

[10] S. Lu and Caishigang, "Study on papermaking enterprise financial accounting data analysis model based on decision tree and optimized SVM algorithm," *Paper Asia*, vol. 34, no. 6, pp. 57–62, 2018.

[11] N. Kanungsukkasem and T. Leelanupab, "Financial latent dirichlet allocation (FinLDA): feature extraction in text and data mining for financial time series prediction," *IEEE Access*, vol. 7, pp. 71645–71664, 2019.

[12] L. Ouyang, Y. Yuan, and F.-Y. Wang, "Learning markets: an AI collaboration framework based on blockchain and smart contracts," *IEEE Internet of Things Journal*, vol. 13, no. 99, p. 1, 2020.

[13] T. Beridze, A. Cherep, Z. Baranik, and V. I. Korenyev, "Analysis of the regression model of the enterprise's financial activity by research on residual error," *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*, vol. 34, no. 2, pp. 193–197, 2021.

[14] Y. Mao, Z. Huang, C. Feng, and H. Q. J. Chen, "An early warning method of distribution system fault risk based on data mining," *Mathematical Problems in Engineering*, vol. 2020, no. 3, Article ID 8880661, pp. 1–10, 2020.

[15] X. Xu, X. Yin, and X. Chen, "A large-group emergency risk decision method based on data mining of public attribute preferences," *Knowledge-Based Systems*, vol. 163, pp. 495–509, 2019.

[16] H. Wang and Q. Liang, "Risk analysis and route optimization of dangerous goods transportation based on the empirical path set," *Journal of Advanced Transportation*, vol. 2020, no. 99, Article ID 8838692, pp. 1–13, 2020.

[17] H. Zhao, H. Yu, D. Li, T. Mao, and H. Zhu, "Vehicle accident risk prediction based on AdaBoost-SO in VANETs," *IEEE Access*, vol. 7, pp. 14549–14557, 2019.

[18] S. Liu, L. Sun, S. Zhu, and J. X. W. Li, "Operation strategy optimization of desulfurization system based on data mining," *Applied Mathematical Modelling*, vol. 81, no. May, pp. 144–158, 2020.

[19] P.-W. Chi and C.-L. Lei, "Audit-free cloud storage via deniable attribute-based encryption," *IEEE Transactions on Cloud Computing*, vol. 6, no. 2, pp. 414–427, 2018.

[20] Y. Gong, Z. Niu, and T. Bai, "Societal risk acceptance criteria for gas distribution pipelines based on incident data from the United States," *Journal of Loss Prevention in the Process Industries*, vol. 63, p. 104002, 2020.

[21] S. Lu, G. Lin, H. Liu, C. Ye, H. Que, and Y. Ding, "A weekly load data mining approach based on hidden markov model," *IEEE Access*, vol. 15, p. 1, 2019.

[22] D. Zhu, H. Zhang, Y. Sun, and H. Qi, "Injury risk prediction of aerobics athletes based on big data and computer vision," *Scientific Programming*, vol. 2021, no. 1, Article ID 5526971, pp. 1–10, 2021.

[23] N. Restrepo, J. M. Uribe, and D. Manotas, "Financial risk network architecture of energy firms," *Applied Energy*, vol. 215, pp. 630–642, 2018.

[24] L.-T. Zhao, S.-Q. Guo, and Y. Wang, "Oil market risk factor identification based on text mining technology," *Energy Procedia*, vol. 158, pp. 3589–3595, 2019.

[25] T. Hu, P. Yi, J. Lan, and Y. P. Hu, "ACST: audit-based compromised switch tolerance for enhancing data plane robustness in software-defined networking," *Computer Networks*, vol. 161, pp. 264–280, 2019.