

Research Article

Financial Data Analysis and Application Based on Big Data Mining Technology

Jinfeng Cheng 

School of Finance and Accounting, Henan Industry and Trade Vocational College, ZhengZhou 451191, China

Correspondence should be addressed to Jinfeng Cheng; chengjinfeng@hngm.edu.cn

Received 19 April 2022; Accepted 27 May 2022; Published 25 June 2022

Academic Editor: Le Sun

Copyright © 2022 Jinfeng Cheng. 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.

We provide a brief overview of the connotation and characteristics of data mining technology in the era of big data, analyze the feasibility of data mining technology in business management from the economic and technical perspectives, and propose specific application suggestions according to the content and requirements of business management. This paper describes in detail the principles and steps of using the weighted plain Bayesian algorithm and the decision tree algorithm to analyze students' performance; firstly, we need to obtain the plain Bayesian analysis model of college students' learning literacy in physical education and the C4.5 graduation literacy analysis model, and then use certain rules to combine the weighted plain Bayesian algorithm and the decision tree algorithm to obtain the WNB-C4.5 college students' learning literacy analysis model. In addition, in the prediction of financial risks, the classification scheme can be used in the judgment of violation of regulations, but the most used classification scheme is the decision tree. Experiments show that the effectiveness of this scheme in data mining for financial companies is increased by 2% compared to the benchmark method.

1. Introduction

With the rapid development of Internet, cloud computing, and Internet of Things (IoT) technologies in recent years, modern society has gradually stepped into the era of informatization and data-oriented environment [1]. In the development of enterprises, production and operation activities will generate a lot of data and the trend of explosive growth [2]. Comprehensive retrieval, analysis, and application of data can lay a good foundation for the formulation of scientific decision-making, and data information has gradually become an important factor affecting the development capacity of enterprises [3, 4].

With the continuous development of data mining technology, researchers have been expanding data mining technology, which makes its application research fields become more and more extensive [5]. At present, a large number of data mining techniques are successfully applied in many fields such as medical and health care, national

defense science and technology, education and teaching, enterprise applications, and communication industry, which are widely concerned by researchers [6].

For example, in the area of intelligent decision support system, few researchers [7, 8] researched and designed an intelligent decision support system based on data warehouse, OLAP, and data mining methods, and also researched and designed a new intelligent decision architecture framework. In terms of data warehouse applications, Liu et al. [9] researched and implemented a management system applicable to customer data analysis based on data warehousing and data mining techniques, and the combination of the two techniques reflects the advantages of analyzing historical data and is widely used in the mobile communication industry. Few researchers [10, 11] analyzed the intelligent financial decision support system of the ZT Group by combining three key mining techniques, namely association rules, fuzzy methods, and unstructured data mining techniques, and also adopted a function mapping approach to achieve improved

efficiency of operations in response to the shortcomings of the above three techniques. Similar analysis of intelligent decision support system based on data mining technology has many other worthy examples [12].

With the wide application of data mining techniques, most universities also apply the techniques commonly used in data mining to their daily educational teaching activities. Cui and Yan [13] designed and implemented an efficient grade analysis system based on data mining. The system adopts the grade analysis method of data mining, which can quickly and efficiently uncover valuable potential information hidden in a large amount of grade data and help university academic staff to comprehensively analyze students' grades. In [14], data mining technology is applied to the university reader borrowing query and analysis management system, the association rule mining technique is studied in depth, and the classical Apriori algorithm is analyzed, while the Apriori algorithm is improved, thus improving the efficiency of the algorithm to a great extent.

In the metrology business processing, the traditional metrology business often fails to extract valuable data information quickly and effectively when dealing with huge data, which restricts the metrology business management decisions [15]. The use of data mining technology, by building data mining models and data warehouses can effectively handle the huge amount of data generated in metrology business, thus reducing the errors in metrology business to within the standard range and improving the efficiency of metrology business. In [16], the application of data mining technology in WAP business operation is described. By analyzing and comparing the advantages and disadvantages characteristics of various data mining methods, the association algorithm is finally selected to mine the access logs generated by WAP, and some practical optimization solutions are proposed for the performance of data warehouse and data mining.

The deep application of data mining techniques is also widely involved in medical applications. Xu et al. [17] improved the Apriori algorithm by analyzing classical association rule mining algorithms such as AIS algorithm, FP-Growth, and Apriori algorithm, and proposed an array-based mining association rule DRA algorithm, which greatly improves the operation efficiency because the DRA algorithm does not need to generate candidate sets. In [18], a design idea of a data mining system for TCM cases with gastric pain was proposed, and the application of classical association algorithm in TCM cases with gastric pain was effectively verified by mining the medication pattern in 1221 cases for treating a certain disease using the Apriori algorithm.

From the above literature review, it is easy to find that data mining technology is now involved in almost every aspect of people's daily production life, and it is also used in intelligent decision support systems, higher education institutions, metrology business processing, mobile dream network data business, medical business, and other fields with increasing maturity [19, 20]. Based on the respective characteristics of classical Apriori association rule, clustering algorithm, and decision tree algorithm in data mining technology, we decided to use the above three data mining

algorithms to realize the analysis of the enterprise's financial data so as to uncover the potential value information in the enterprise's financial data and provide a reliable decision basis for the enterprise's leadership.

2. System Business Requirements Analysis

2.1. System Process Analysis. The system process analysis mainly describes the execution process of a core business in the main functional modules of the system. Since the financial management system has more functions and the accompanying business process is also relatively large, in view of this, this chapter will focus on analyzing the original financial card management process in the financial management system. The original financial fixed assets card management specification process is shown in Figure 1.

- Step 1: login to the system with the minimum open month.
- Step 2: enter the "Fixed Assets" management operation and enter the original card node; locate an original card at the same time and copy the original card operation.
- Step 3: select the fixed assets category.
- Step 4: enter the items of this fixed assets master card.
- Step 5: save the card.
- Step 6: select the attached card.
- Step 7: make changes to the selected supplementary card, add another supplementary card, and enter the contents again.
- Step 8: save the card.

Through the above eight steps, you can realize the original financial card data entry workflow.

The general ledger of enterprise assets is an accounting of enterprise fixed assets according to certain classification standards in a certain period of all economic operations, the original value of the assets, accumulated depreciation, net value (provision for impairment, net) in a three-column format of debit, credit, and balance of the summary to reflect the changes in their value of the pages of the account. The flow chart of general ledger management is shown in Figure 2.

The general ledger process includes the initial balance entry and after the trial balance, the initial accounts can be created. The general ledger manager can then create some account vouchers and documents based on the initial accounts, and by signing and stamping on the postpayment vouchers, eventually form transfer vouchers for year-end account review and audit role, and finally form bookkeeping methods for year-end transfer.

3. Data Mining Technology in the Era of Big Data

Big data mining technology is an important constituent element of knowledge discovery to analyze data with computer algorithms. In a large number of databases, the

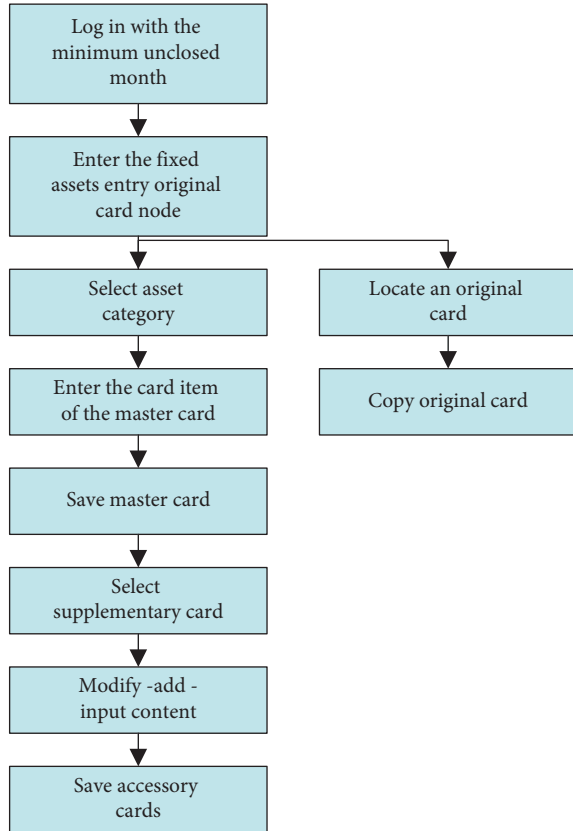


FIGURE 1: Original financial card management process.

required data is obtained, and the data is appropriately transformed, mined, and utilized to obtain valuable information. Generally speaking, the object of big data mining is basically structured, semistructured, or other structured data. The process of data mining is mainly data selection \rightarrow data mining \rightarrow data analysis (see Figure 3).

4. Financial Analysis Method Based on Weighted Multiple Random Decision Trees

The classification problem of financial data is completed by adding a random decision tree scheme to the model, as shown in Figure 4.

The criticality of the attributes in the financial data warehouse varies under different mining objectives, so the criticality of each attribute should be analyzed quantitatively when establishing the decision tree. The current schemes that are often used to confirm the criticality of attributes are the discriminant matrix-based scheme and the information entropy-based scheme. In this study, we use the discriminant matrix scheme to evaluate the importance of attributes. In addition, since financial data are highly specialized, it is not possible to reflect the actual importance of an attribute by relying only on the discriminant matrix, so this project adds artificial weights to modify and intervene in the discriminant matrix to make the calculation of attribute weights more accurate [21, 22].

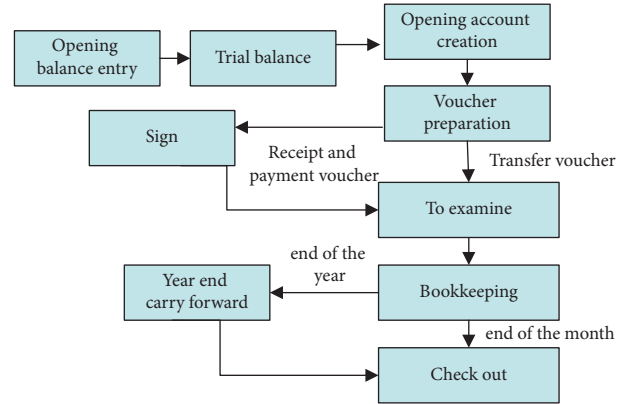


FIGURE 2: General ledger management process.

4.1. *Defining the Resolution Matrix.* A diagonal matrix of $|u| \times |u|$. Each of these terms is defined as

$$C_{ij} = \begin{cases} \{\alpha \in A | \alpha(x_i) \neq \alpha(x_j)\} d(x_i) \neq d(x_j), d(x) \in D \\ \phi \quad d(x_i) = d(x_j), d(x) \in D \end{cases} \quad (1)$$

The number of occurrences and the importance of the attributes in the discrimination matrix are positively correlated; and the shorter the data item with the attribute present, the more critical the attribute.

4.2. *Calculating Attribute Weights for Financial Data.* Initialize all $a_i \in A$ such that $w(a_i) = 0$.

For each term of the diagonal matrix in the resolution matrix C_{jk} calculate

$$w(a_i) = w(a_i) + |C_{jk}|, \quad a_i \in C_{jk}, 0 < k < j = |U|. \quad (2)$$

In the above equation, $|A|$ is the base of all attributes and $|C_{jk}|$ is the base of the discrimination matrix C_{jk} .

After the system presents the weights, it is possible to manually correct the weights in the system, so it is necessary to add the correction coefficients $w^l(a_i)$, $-1 < w^l(a_i) < 1$, if you want to increase the weight of a_i by setting $w^l(a_i)$ to a positive value, and the opposite by setting it to a negative value, then the weight of attribute a_i is $W_{ai} = w(a_i) + w^l(a_i)$.

5. Experimental Validation

This validation data are derived from the financial statistics of more than 1400 company customers who have worked with a commercial bank, and the period of validation data are uniform from 2013 to 2016. The financial information data tables are divided into attributes based on the bank's transaction database, so the financial information data tables provided by the bank can be transformed into 24 attributes that clearly show the financial situation of the company, as presented in Table 1.

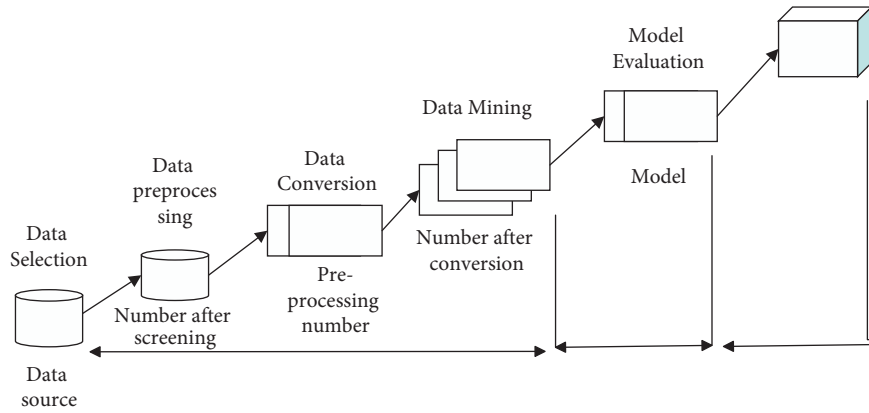


FIGURE 3: Flow of big data mining.

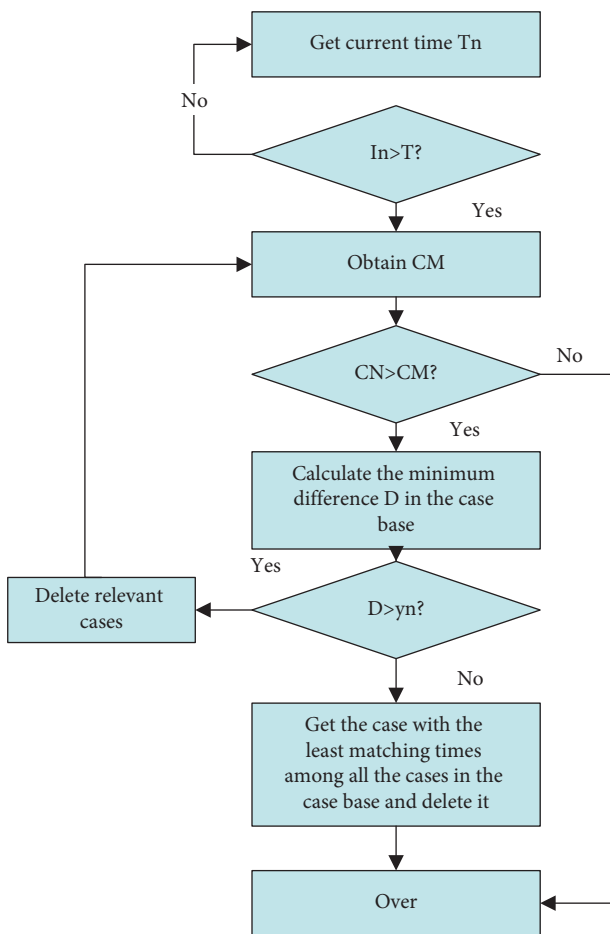


FIGURE 4: Case deletion process.

Due to the actual situation of the company in 2017 and the indicators related to the company, experts in finance classify the company risk into four categories: large, large, small, and small. In this case, companies with high risk are those that will go bankrupt from 2015 to 2017; companies with high risk are those that will default; companies with low risk are those that will not default but their financial situation will deteriorate, and companies with low risk have a

normal financial situation and will not default. The results of the study showed that the best way to apply the decision is to build 10 random decision trees. Therefore, in this study, a total of 10 randomized decision trees were built from the analyzed data because the decision trees were built in a randomized manner, and a total of 5 trials were conducted to verify the stability of the decision trees [23, 24].

The remaining 300 data are test data. The training data were used to build a random decision tree, and the completed decision tree was tested using the test data to finally document the classification accuracy of the decision tree. The experimental results are presented in Table 2 and Figure 5.

The results of the experiment show that this randomized decision tree algorithm classifies companies with large risk, large risk, small risk, and small risk with improved classification accuracy, which has been determined by bank staff to be a practical reference for predicting bank risk. However, because of the small number of large risk data in the training dataset, this branch is not sufficiently trained, making the stochastic decision tree algorithm less accurate than the other branches for large risk classification [25].

Each time, using the same training and validation data, the C4.5 algorithm is applied to classify the risk level, and the final results are presented in Table 3 and Figure 6.

The results of the experiments show that this randomized decision tree algorithm improves the classification accuracy for the risk level of large, risk level of large, risk level of small, and risk level of small by a considerable amount. Similarly, it can be seen that because the number of data with large risk level in the training dataset is relatively small, this class of branches is not trained sufficiently. The accuracy of the C4.5 algorithm is significantly lower for the risky branches compared to the other branches. This is shown in Figure 7.

From Figure 7, we can see that the accuracy of the randomized decision tree algorithm is higher than that of the C4.5 algorithm, which is about 10% higher.

In order to improve the correct rate, 300 data with high risk level are added to the training data set because the training of high risk level is not sufficient. The number of training data with large risk is ensured by replacing the

TABLE 1: Table of experimental attributes.

Attribute code	Attribute	Calculation formula
A1	Return on assets	$(\text{Total profit} + \text{finance costs}) / (\text{Total assets} + \text{total assets of the previous period}) * 2$
A2	Gearing ratio	$\text{Total liabilities} / \text{Total assets}$
A3	Net profit on total assets	$\text{Net Income} / (\text{Total assets} + \text{total assets of the previous period}) * 2$
A4	Return on assets	$\text{Net income} / (\text{Total shareholders' equity} + \text{total shareholders' equity of the previous period}) * 2$
A5	Operating income net profit ratio	$\text{Profit from main business} / \text{income from main business}$
A6	Quick ratio	$(\text{Total current assets} - \text{net inventory}) / \text{total current liabilities}$
A7	Current ratio	$\text{Total current assets} / \text{total current liabilities}$
A8	Fixed assets ratio	$\text{Total fixed assets} / \text{total assets}$
A9	Inventory turnover ratio	$\text{Cost of main business} / (\text{net inventory} + \text{net inventory of previous period}) * 2$
A10	Interest cover multiplier	$(\text{Net profit} + \text{income tax} + \text{finance costs}) / \text{financial costs}$
A11	Total assets turnover ratio	$\text{Income from main business} / (\text{Total assets} + \text{total assets of the previous period}) * 2$
A12	Working capital to total assets ratio	$(\text{Total current assets} - \text{Total current liabilities}) / \text{Total assets}$
A13	Cash from main business ratio	$\text{Cash flow from operating activities} / \text{income from main business}$
A14	Accounts receivable turnover ratio	$\text{Income from main business} / (\text{accounts receivable} + \text{prior period accounts receivable}) * 2$
A15	Fixed assets turnover ratio	$\text{Revenue from main business} / (\text{Total fixed assets} + \text{total fixed assets of the previous period}) * 2$
A16	Accounts receivable turnover ratio	$\text{Income from main business} / (\text{total fixed assets} + \text{total fixed assets of the previous period}) * 2$
A17	Capital adequacy ratio	$\text{Total shareholders' equity} / \text{Total assets}$
A18	Inventory current liability ratio	$\text{Net inventory} / \text{Total liquidity liabilities}$
A19	Cash flow to current liabilities ratio	$\text{Total cash flow from operating activities} / \text{Current liabilities}$
A20	Net income growth rate	$\text{Net profit for the period} / \text{Net profit for the previous period}$
A21	Operating profit growth rate	$\text{Operating profit for the period} / \text{Operating profit for the previous period}$
A22	Main revenue growth rate	$\text{Income from main business for the period} / \text{Income from main business for the previous period}$
A23	Net assets growth rate	$\text{Net assets for the period} / \text{Net assets for the previous period}$
A24	Debt capital ratio	$\text{Total liabilities} / \text{Total shareholders' equity}$

TABLE 2: Comparison of the correct classification rate of multiple stochastic decisions.

Verification times	Small risk %	Less risky %	Risky %	High risk %
1	87.95	78.58	72.71	54.33
2	88.36	79.23	73.98	53.35
3	89.21	80.03	72.55	45.99
4	88.25	79.65	73.39	58.39
5	86.59	78.54	73.38	52.25
Average	88.01	78.61	73.68	52.63

original random sampling with stratified sampling, in which the initial data are stratified by small, small, large, and large risk, and then random sampling is used for each stratum. The classification results using the stratified random sampling method are presented in Table 4 and Figures 8 and 9.

From the above figure, we can see that the correct rate of using the stratified sampling method with high risk is 10% higher than that of the random sampling method with high risk. The underlying reason is that 300 risky data are added to the training data set, which provides more samples for the stratified sampling. Therefore, the number of samples in the training data determines whether the decision tree classification is correct or not, and if the number of samples is

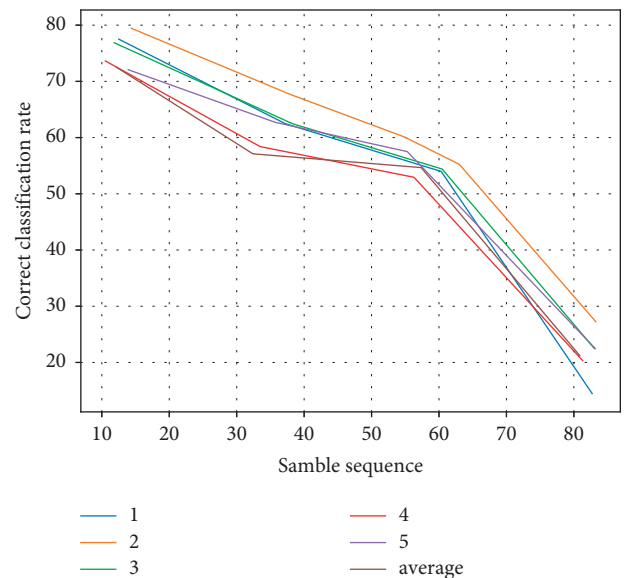


FIGURE 5: Comparison of the correct classification rate of multiple randomized decision trees.

large enough, the decision tree classification will be more correct.

TABLE 3: Comparison of C4.5 classification accuracy.

Verification times	Small risk %	Less risky %	Risky %	High risk %
1	72.58	65.35	6.31	35.59
2	73.36	66.78	62.19	37.12
3	74.55	66.52	66.37	34.39
4	73.98	65.35	62.98	42.86
5	75.91	66.29	63.28	33.98
Average	74.01	66.22	62.58	36.59

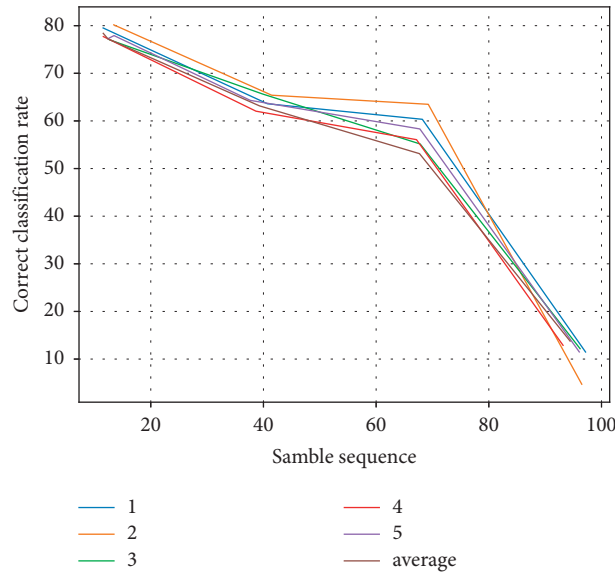


FIGURE 6: Comparison of the correct classification rate of C4.5 algorithm.

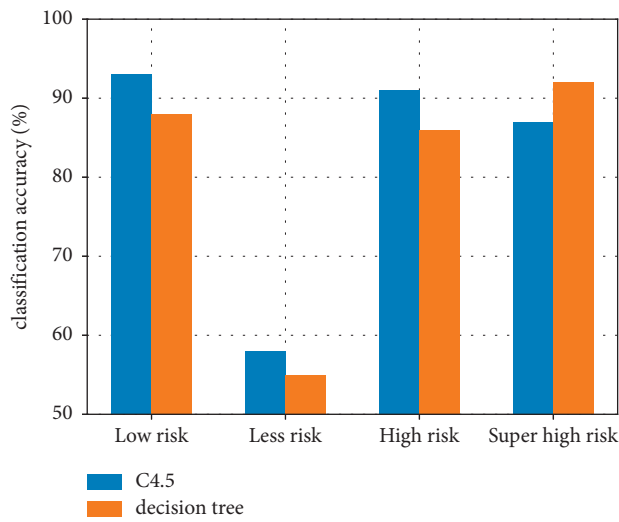


FIGURE 7: Comparison of the classification accuracy of the two algorithms.

TABLE 4: Comparison of the correct rate of stratified sampling for multiple random decision trees.

Verification times	Small risk %	Less risky %	Risky %	High risk %
1	89.33	86.36	77.55	71.03
2	90.32	85.22	76.53	71.98
3	91.39	88.26	78.96	71.65
4	88.36	86.12	79.32	70.11
5	88.96	88.69	79.89	75.97
Average	88.98	84.98	78.03	71.56

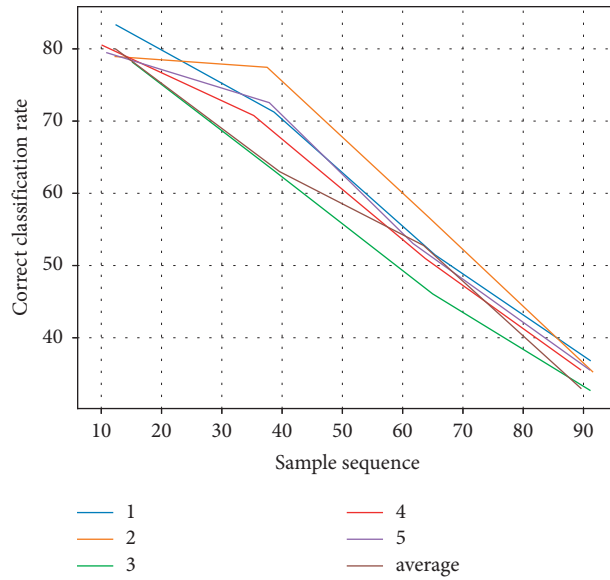


FIGURE 8: Comparison of the correct classification rate for stratified sampling of multiple random decision trees.

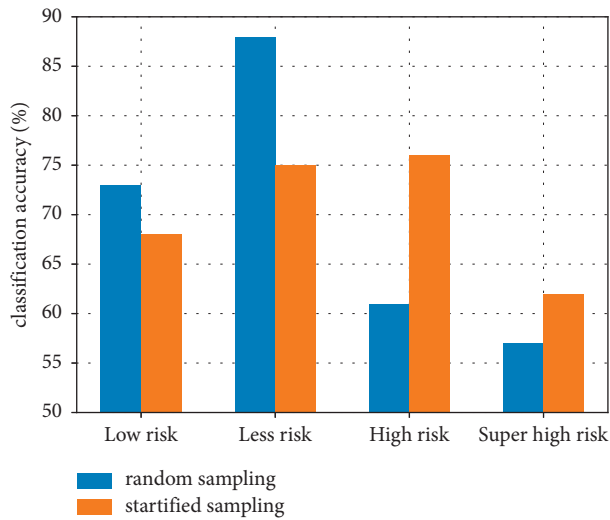


FIGURE 9: Comparison of the correct classification rate of random sampling and stratified sampling.

6. Conclusion

In the era of big data, the content of enterprise financial analysis has increased and the complexity of work is higher. The reasonable application of data mining technology can reduce the work pressure of financial personnel and can improve the quality and efficiency of financial analysis, so it is recommended to promote the use. Good foundation to play the role of data support. During the enterprise cost efficiency accounting, data mining technology can be applied to analyze the association of a certain type of cost and another directly unrelated cost. If it has high correlation characteristics, it needs to be integrated into the process of project budgeting and decision-making to improve the accuracy of cost-benefit accounting.

Data Availability

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

Conflicts of Interest

The author declares that they have no conflicts of interest regarding this work.

References

- [1] H. Shang, D. Lu, and Q. Zhou, "Early warning of enterprise finance risk of big data mining in internet of things based on fuzzy association rules," *Neural Computing & Applications*, vol. 33, no. 9, pp. 3901–3909, 2021.
- [2] D. Li, L. Deng, and Z. Cai, "Statistical analysis of tourist flow in tourist spots based on big data platform and DA-HKRVM algorithms," *Personal and Ubiquitous Computing*, vol. 24, no. 1, pp. 87–101, 2020.
- [3] A. Mohamed, M. K. Najafabadi, Y. B. Wah, E. A. K. Zaman, and R. Maskat, "The state of the art and taxonomy of big data analytics: view from new big data framework," *Artificial Intelligence Review*, vol. 53, no. 2, pp. 989–1037, 2020.
- [4] T. Xie, R. Liu, and Z. Wei, "Improvement of the fast clustering algorithm improved by K-means in the big data," *Applied Mathematics and Nonlinear Sciences*, vol. 5, no. 1, pp. 1–10, 2020.
- [5] J. L. Hung, W. He, and J. Shen, "Big data analytics for supply chain relationship in banking," *Industrial Marketing Management*, vol. 86, pp. 144–153, 2020.
- [6] F. Wang, M. Li, Y. Mei, and W. Li, "Time series data mining: a case study with big data analytics approach," *IEEE Access*, vol. 8, pp. 14322–14328, 2020.
- [7] S. Khanra, A. Dhir, and M. Mäntymäki, "Big data analytics and enterprises: a bibliometric synthesis of the literature," *Enterprise Information Systems*, vol. 14, no. 6, pp. 737–768, 2020.
- [8] R. Hou, Y. Kong, B. Cai, and H. Liu, "Unstructured big data analysis algorithm and simulation of Internet of Things based on machine learning," *Neural Computing & Applications*, vol. 32, no. 10, pp. 5399–5407, 2020.
- [9] C. Liu, Y. Feng, D. Lin, L. Wu, and M. Guo, "Iot based laundry services: an application of big data analytics, intelligent logistics management, and machine learning techniques," *International Journal of Production Research*, vol. 58, no. 17, pp. 5113–5131, 2020.
- [10] F. Arena and G. Pau, "An overview of big data analysis," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 4, pp. 1646–1653, 2020.
- [11] P. An, Z. Wang, and C. Zhang, "Ensemble unsupervised autoencoders and Gaussian mixture model for cyberattack detection," *Information Processing & Management*, vol. 59, no. 2, 102844 pages, 2022.
- [12] L. Wang and C. A. Alexander, "Big data analytics in medical engineering and healthcare: methods, advances and challenges," *Journal of Medical Engineering & Technology*, vol. 44, no. 6, pp. 267–283, 2020.
- [13] Z. Cui and C. Yan, "Deep integration of health information service system and data mining analysis technology," *Applied Mathematics and Nonlinear Sciences*, vol. 5, no. 2, pp. 443–452, 2020.
- [14] D. Balios, "The impact of Big Data on accounting and auditing," *International Journal of Corporate Finance and Accounting (IJCFA)*, vol. 8, no. 1, pp. 1–14, 2021.
- [15] J. Li, Z. Zhou, J. Wu et al., "Decentralized on-demand energy supply for blockchain in internet of things: a microgrids approach," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 6, pp. 1395–1406, Dec, 2019.
- [16] W. Duan, J. Gu, M. Wen, G. Zhang, Y. Ji, and S. Mumtaz, "Emerging technologies for 5G-IoV networks: applications, trends and opportunities," *IEEE Network*, vol. 34, 2020.
- [17] Z. Xu, C. Cheng, and V. Sugumaran, "Big data analytics of crime prevention and control based on image processing upon cloud computing," *Journal of Surveillance, Security and Safety*, vol. 1, no. 1, pp. 16–33, 2020.
- [18] M. Mohammadpoor and F. Torabi, "Big Data analytics in oil and gas industry: an emerging trend," *Petroleum*, vol. 6, no. 4, pp. 321–328, 2020.
- [19] P. Galetsi, K. Katsaliaki, and S. Kumar, "Big data analytics in health sector: theoretical framework, techniques and prospects," *International Journal of Information Management*, vol. 50, pp. 206–216, 2020.
- [20] O. I. Khalaf and G. M. Abdulsahib, "Design and performance analysis of wireless IPv6 for data exchange," *Journal of Information Science and Engineering*, vol. 37, pp. 1335–1340, 2021.
- [21] U. Srilakshmi, N. Veeraiah, Y. Alotaibi, S. A. Alghamdi, O. I. Khalaf, and B. V. Subbayamma, "An improved hybrid secure multipath routing protocol for MANET," *IEEE Access*, vol. 9, pp. 163043–163053, 2021.
- [22] T. Nguyen, R. G. Gosine, and P. Warrian, "A systematic review of big data analytics for oil and gas industry 4.0," *IEEE Access*, vol. 8, pp. 61183–61201, 2020.
- [23] R. R. Nadikattu, "Research on data science, data analytics and big data," *International Journal of Engineering Science*, vol. 9, no. 5, pp. 99–105, 2020.
- [24] M. Q. Shabbir and S. B. W. Gardezi, "Application of big data analytics and organizational performance: the mediating role of knowledge management practices," *Journal of Big Data*, vol. 7, no. 1, p. 47, 2020.
- [25] S. V. Novikov, "Data science and big data technologies role in the digital economy," *TEM Journal*, vol. 9, no. 2, pp. 756–762, 2020.