

# *Application of Data Mining Technology in Financial Evaluation in the Era of Big Data*

Shichao Chen<sup>1,a,\*</sup>, Shiyu Zou<sup>2,b</sup>

<sup>1</sup>*Division of Scientific Research and Planning, Bengbu College of Technology and Business, Bengbu, Anhui, China*

<sup>2</sup>*State-owned Assets Supervision and Administration Commission of Jiujiang, Jiujiang, Jiangxi, China*

<sup>a</sup>7457238@qq.com, <sup>b</sup>zoushiyu88@126.com

\*Corresponding author

**Keywords:** Financial Analysis; Big Data; Data Mining; Funding Review; Data Processing

**Abstract:** Modern financial analysis can also be called financial statement analysis. It was originally proposed by American bankers and was used to examine the repayment ability of enterprises applying for loans in the early days. Current financial statements are not only used by companies to summarize previous operating results and assess their financial status, but are also widely used to predict their future turnover, providing very reliable financial data for corporate executives, investors and banks. However, with the development of time, the scale of the enterprise is increasing day by day, the data that needs to be processed for financial analysis is also increasing day by day, and its working hours are also increasing, and it is increasingly unable to meet some needs of the company. The emergence of big data technologies has given birth to fresh optimism for financial research at this moment. While solving data processing problems, data mining technology can extract the required data from the massive data generated in the daily activities of the enterprise, which further improves the accuracy of financial analysis. In this article, a more in-depth assessment of organizations' financial analytical abilities was conducted against the backdrop of the fast growth of big data and data mining technologies. It was intended to integrate data mining technology into the daily financial analysis of enterprises, optimize modern financial analysis, shorten the time required for enterprise financial analysis and improve the accuracy of financial analysis. Finally, the paper compared the performance indicators of traditional financial analysis and financial analysis based on data mining technology through a series of experiments. The results showed that the financial analysis ability based on data mining technology was improved by about 63.4% compared with the traditional financial analysis ability.

## 1. Introduction

With the advent of the big data age and the ongoing advancement of data mining technology, more and more enterprises have begun to study the feasibility and practical operation plan of integrating data mining technology into financial analysis. The analysis of enterprise financial

statements based on data mining technology can not only improve the utilization rate of financial information, but also greatly reduce the repetitive work of enterprise financial staff and improve their work efficiency.

With the continuous growth of social economy, the market competition faced by enterprises is becoming more and more fierce. In order to improve their core competitiveness, an enterprise can only continuously strengthen its own strength in all aspects and its financial management capabilities. Financial analysis can not only help enterprises to comprehensively and scientifically examine the financial situation of enterprises, but also improve the overall operation level of enterprises. Barrette conducted a financial analysis of the profitability of a sawmilling business, comparing the production and value of dead trees with different degradation standards in the business, the wood chips produced alone or with wood, and the current market value. The results showed that the use of dead wood and wood pellets to produce pulp is very profitable [1]. Deependra N, in order to study the economic feasibility of generating electricity from renewable resources in rural areas, proposed a method to determine the financial relevance of solar photovoltaic power generation. Using minimum cost and reliability constraints for optimal sizing, this study contributed to PV power reinforcement policies to provide reliable power guarantees in rural areas [2]. In order to compare the cost of Cardiac Rehabilitation (CR) before the new crown epidemic and during the new crown epidemic, Melbostad H used the micro-cost method to calculate the medical expenses of cardiac rehabilitation. The main costs are labor time and consumables costs. The results of the study showed that the average cost of cardiac rehabilitation for patients during the new crown epidemic increased slightly, and the estimated cost of cardiac rehabilitation during the new crown epidemic was higher than the expected income [3]. Halim M A surveyed the financial performance of three banks over several years, and the results showed that compared with private commercial banks, the performance of state-owned commercial banks was lower. If state-owned commercial banks want to improve their performance, they should control costs and pay more attention to efficiency and profitability when making loans [4]. In order to reduce the economic loss caused by bacterial contamination of platelets, Kacker S used a computational model of Markov's decision tree to evaluate the financial loss caused by all the variants of the implementation guidelines from the perspective of the hospital, and the results showed that a large number of delayed samples is an efficient and competitive scheme [5]. In order to reduce the raw material supply cost of cellulosic biorefinery, Shooshtarian A analyzed the growth rate and cost of hybrid poplar, and predicted two economic return modes, proving that hybrid poplar can be used as a low-cost feedstock for refineries [6]. Ma Y used financial statement analysis to study the listing of a Chinese medicine company, and calculated its product performance, basic functions, financial indicators and corporate valuation. Based on the above indicators to analyze the development prospects of the company, the conclusion shows that the company can be listed on the New Third Board [7]. The current financial analysis can play a good role in promoting the development of enterprises, and the continuous development of enterprises has also helped the further development of the theory of financial analysis.

As one of the hot research issues in the field of artificial intelligence and database, data mining has attracted much attention from the industry. Data mining can be widely used in engineering design, scientific development, market analysis and business management and other fields. Yan XS built a database with a large number of basic signals from the data in the financial statements, and used the bootstrapping method to evaluate the impact of data mining on abnormal situations, and found that the data processing ability of data mining can accurately predict stocks with large arbitrage restrictions higher rate [8]. In order to accurately predict the changes of stock prices and bring considerable profits to investors, Namahoot KS established three effective prediction models to predict the changes of stock price indexes, and compared the performance of the three models in

the daily stock market [9]. Li XF has studied the online loan default risk model based on data mining technology. The purpose is to study a data mining method to obtain the default risk from the daily behavior characteristics of the lender to reduce the non-performing loan rate of online loans and bring profits to enterprises [10]. In order to solve the problem of how to choose an appropriate credit scoring model for financial risk control, Protopapadakis E tested the feasibility of selecting a reliable credit scoring model with different algorithms, and finally proved that data mining technology can correctly select a credit scoring model based on data [11]. In order to find an effective method to detect suspicious transactions, Lokanan ME used statistical techniques to mine and analyze suspicious transactions. The final result proved that data mining technology can effectively detect suspicious transactions [12]. In order to prove that data mining can be applied to audit programs, Werner M used data mining algorithms to generate process models by analyzing event logs. This model can provide accurate information about related calculations [13]. Milutinovi S used data mining to study accountants' attitudes towards regulation. The study of data clusters can provide insight into the opinions of accountants with high feature similarity [14]. Although the current theoretical research progress of data mining technology is very rapid, due to the high cost of computer computing power, the large-scale popularization of this technology is still far away.

This paper deeply studied the application of data mining technology in the analysis of enterprise financial statements, and used it to collect data related to enterprise assets in the daily operation of enterprises, and conducted centralized statistics and calculations on them. Finally, the data was calculated using algorithms related to financial statement analysis. Optimizing the operation mode of the enterprise improves the work efficiency of the financial department of the enterprise, so that the senior management of the enterprise can make decisions on the next trend of the enterprise based on a more accurate financial situation forecast, which brings higher competitiveness to the enterprise. Financial analysis based on data mining technology can bring more timely and accurate financial status to enterprises, so that enterprises can further develop.

## **2. Evaluation of Data Mining Technology in the Era of Big Data**

### **(1) Big data technology**

The 21st century is an era of exponential growth of various types of data. Mobile Internet, social software and the Internet of Things have further expanded the scope of Internet use. In the new era, people are also unknowingly enjoying the convenience brought by various technologies at the expense of the leakage of their daily life trajectories. At the same time, everyone is naked on the Internet. Companies rely on big data technology to collect various characteristic tags of each consumer, and use this to make decisions about the direction of the company's next-generation products. It can be said that the arrival of the era of big data poses new challenges to people's ability to grasp information, and it also provides a foundation for human beings to gain a deeper and more comprehensive insight into the world and things [15].

Big data is the massive and diverse data that emerges at high speed. Generally speaking, it is a very large and complex data set. Big data has the characteristics of high speed, diversification, value and authenticity, that is, it covers many types of data, and the transmission speed in memory is very fast. Finally, the data reliability must be guaranteed before the intrinsic value of the data can be mined from the huge data. At present, the operation of all large technology companies in the world is inseparable from big data technology. Companies relying on big data technology can get higher returns than before at lower costs.

The core idea of big data is to predict the progress of things [16]. Usually, big data is mistaken for a technology similar to artificial intelligence, which simulates human thinking about the operation of things and predicts the next development of things. However, in fact, on the contrary, it

is not required to simulate human thinking to calculate things, but to more objectively predict the results of things by performing mathematical calculations on massive amounts of relevant data. For example, it can objectively judge whether a pedestrian collides with a vehicle through the walking trajectory, speed and distance from the vehicle on the road. The specific operation process of big data is shown in Figure 1.

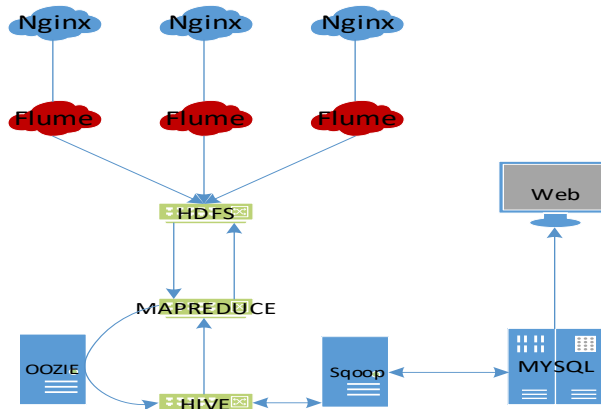


Figure 1: Schematic diagram of the detailed operation process of big data technology

(2) Data mining technology

Data mining technology is a workflow that extracts potentially valuable information from huge and fuzzy actual data, which requires that the data source must be authentic and valid [17]. Generally speaking, data mining technology can be regarded as a method for deep data analysis.

The methods of using data mining technology to analyze data in daily business activities mainly include classification, variation and deviation analysis, association rules, clustering and characteristics, etc. These methods analyze the data in the database from different angles [18]. The classification is first to find the same characteristics of each group of data in the database, and divide them into different categories according to these characteristics. The main purpose is to make a one-to-one correspondence between a specific data or data group in the database and a certain feature through data analysis. The change and deviation analysis is to study the valuable difference between the analysis object and the reality by analyzing the data, which is generally used in the crisis management of enterprises. The purpose of association rules is to describe a certain relationship between various data groups in the database. For example, in the database of enterprise customer management, some objective connections between certain customers can be found from a large amount of customer data through association rules, which is convenient for the enterprise to make the next marketing decision. The data mining process is shown in Figure 2.

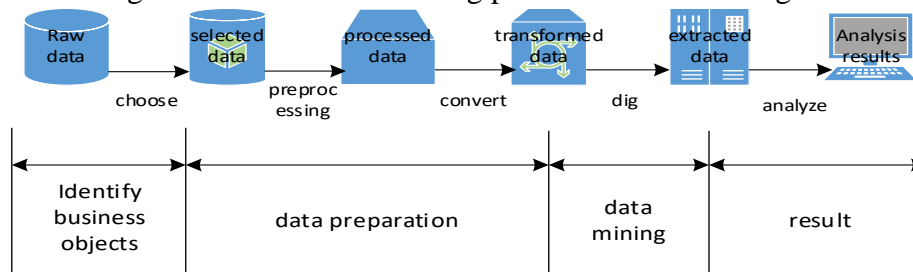


Figure 2: Schematic diagram of the data mining process

The analysis of corporate financial statements based on data mining technology can not only help corporate executives to obtain real-time corporate financial status and other operating conditions, but also make more accurate predictions of consumer preferences, so that corporate

executives can make important decisions. At the same time, it also facilitates the staff of the financial department of the enterprise, improves the efficiency of financial analysis, enables the financial personnel to work more efficiently, and provides a solid foundation for the enterprise to obtain greater competitiveness.

### 3. Evaluation on Financial Statement

The current trend of global economic integration has become more and more obvious, and each enterprise is facing unprecedented challenges and opportunities. Enterprises not only need to compete with their domestic counterparts, but also with foreign companies in the same field [19].

Modern financial analysis is mainly based on the financial report of the enterprise or the operating data of other enterprises, and uses professional methods to systematically study and evaluate the operation and financial status of the enterprise. Generally speaking, corporate financial analysis is mainly to summarize the company's past operating conditions, measure the company's current financial and revenue situation, and make general predictions about the company's future development prospects based on the first two. Through the financial analysis of the enterprise, people can have a more comprehensive understanding of the current situation of the enterprise's funds, revenue capacity and solvency [20]. Through the analysis of these indicators, the level of business operators can be clearly understood, which provides a good data basis for improving business management capabilities. The financial statement analysis process is shown in Figure 3.

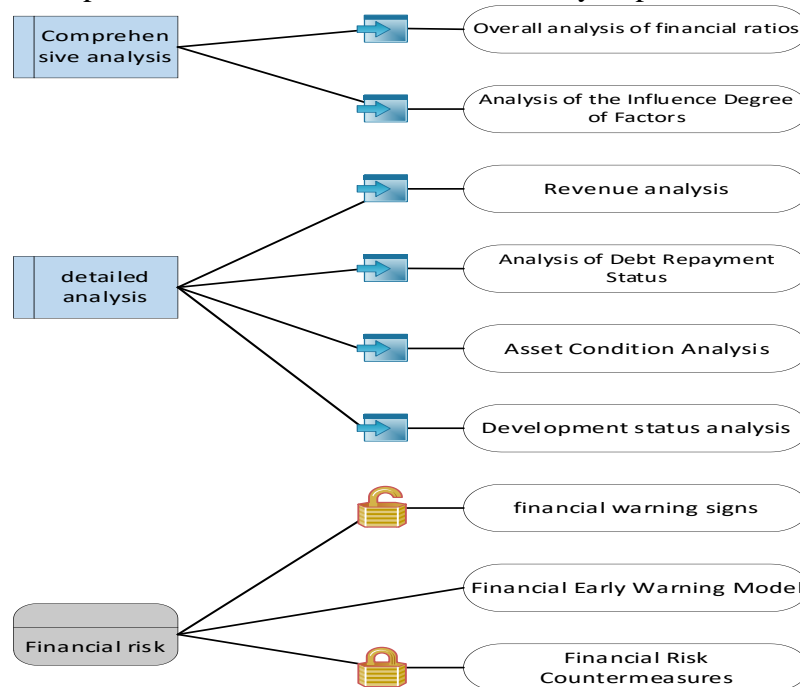


Figure 3: Flowchart of financial statement analysis

Financial statement refers to the form that is compiled regularly in accordance with the rules stipulated in the financial statement on the basis of the daily accounting data. The financial statement analysis is to analyze the financial statement in a fixed period. First of all, it analyzes it by combining point and surface, and uses the method of comparison in the accounting field to conduct a comprehensive analysis of the three tables of the company's profit and loss statement, asset statement and cash flow statement, so as to comprehensively judge the current situation of the company.

#### 4. Big Data and Data Mining Technology Algorithm

The article studies the feasibility and advantages of integrating data mining technology into financial statement analysis under the current technical background. Several common algorithms in data mining are used to optimize and upgrade the capabilities of data collection, analysis, judgment and prediction in financial analysis, which better improves the efficiency and accuracy of corporate financial analysis.

This section would analyze several algorithms of the data mining technology used in this paper. The first is the C4.5 algorithm. The C4.5 algorithm is a decision tree algorithm for classification, which selects classification attributes through information gain. The algorithm first calculates the data gain rate, as shown in Formula (1).

$$SplitInfo_A(S) = - \sum_{j=1}^m \frac{|S_j|}{|S|} \log_2 \frac{|S_j|}{|S|} \quad (1)$$

The data set  $S$  is divided into  $m$  sub-data sets by the attribute value of  $A$ ,  $|S_j|$  represents the number of samples in the  $j$ th sub-data set, and  $|S|$  represents the total number of samples in the entire data set before. The information gain calculation formula after splitting is shown in Formula (2).

$$InfoGain(S, A) = E(S) - E_A(S) \quad (2)$$

Then the efficiency of the information gain of the sample is calculated, as shown in Formula (3).

$$InfoGainRation(S, A) = \frac{InfoGain(S, A)}{SplitInfo_A(S)} \quad (3)$$

After that, the pruning method is used to calculate the error rate before and after the C4.5 algorithm to judge whether the algorithm correctly classifies the data, as shown in Formula (4).

$$ErrorRatio = \frac{\sum_{i=1}^L e_i + 0.5L}{\sum_{i=1}^L n_i} \quad (4)$$

Next, the Bernoulli distribution is used to calculate the number of misjudgments of the subtree, and the mean value of the number of misjudgments of the subtree can be obtained, as shown in Formula (5).

$$ErrorMean = ErrorRatio \times \sum_{i=1}^L n_i \quad (5)$$

At the same time, Formula (6) can be used to calculate the standard deviation of the number of misjudgments of the subtree.

$$ErrorSTD = \sqrt{ErrorRatio \times \sum_{i=1}^L n_i \times (1 - ErrorRatio)} \quad (6)$$

Then all subtrees are replaced with leaf nodes, and the false positive rate is calculated by Formula (7).

$$ErrorRatio' = \frac{e' + 0.5}{n'} \quad (7)$$

The formula represents the core calculation idea of the C4.5 algorithm, and the information gain rate of the data can be accurately calculated through the above formula. Next, the algorithm analysis of the support vector machine is carried out, and the nonlinear classification can be realized by introducing the support vector machine. First, the distance of the function between the given dataset and the two planes is calculated, as shown in Formula (8).

$$\gamma_i = y_i(w^T x_i + b) \quad (8)$$

This interval can represent the correctness and reliability of the classification. Next, Formula (9) is used to calculate the geometric interval.

$$\gamma_i = y_i\left(\frac{w}{\|w\|} \cdot x_i + \frac{b}{\|w\|}\right) \quad (9)$$

Then, the dual problem is solved by the Lagrangian function, and the solution of the most primitive problem is obtained, as shown in Formula (10).

$$L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^m a_i y_i(w^T x_i + b) + \sum_{i=1}^m a_i \quad (10)$$

Finally, the courtship problem is calculated by the given linearly separable training data set, and the decision tree function of dividing plane and classification is obtained, and its calculation formula is shown in Formulas (11) and (12).

$$W = \sum_{i=1}^m a_i^* y_i x_i \quad (11)$$

$$b^* = \frac{1}{|S|} \sum_S (y_s - \sum_{i \in S} a_i^* y_i(x_i^T x_s)) \quad (12)$$

Then comes a detailed introduction to the Expectation Maximum (EM) algorithm. The EM algorithm is essentially an iterative optimization algorithm that can infer its characteristics from a fuzzy dataset parameter. Firstly, the likelihood function is logarithmized, and secondly, each event is represented by the probability of the joint distribution of different categories. The formula is shown in Formula (13).

$$\ln L(\theta) = \sum_{i=1}^n \ln \sum_{j=1}^m p(x_i, z^{(j)}; \theta) \quad (13)$$

At the same time, in order to facilitate the derivation of subsequent formulas, the distribution function  $Q_i(z)$  must satisfy the conditions of Formula (14).

$$\sum_{j=1}^m Q_i(z^{(j)}) = 1, Q_i(z^{(j)}) \geq 0 \quad (14)$$

At this time, the Formula (15) can be obtained by simplifying the distribution function  $Q_i(z)$  after embedding, and the Formula (15) can evaluate the small bound of the likelihood function.

$$J(z, Q) = \sum_{i=1}^n \sum_{j=1}^m Q_i(z^{(j)}) \ln \frac{p(x_i, z^{(j)}; \theta)}{Q_i(z^{(j)})} \quad (15)$$

Next, the expectation of the hidden variable is evaluated. First, Formula (16) is obtained with the value of  $c$  in Formula (14).

$$Q_i(z^{(j)}) = p(z^{(j)} | x_i; \theta) \quad (16)$$

Formula (16) can calculate the posterior probability in the EM algorithm, and at the same time obtain an expected value of the hidden variable in the  $E$  step. The M-step maximum likelihood function is then solved using Formula (17).

$$\ln L(\theta) = \sum_{i=1}^n \sum_{j=1}^m Q_i(z) \ln \frac{p(x_i, z; \theta)}{Q_i(z)} \quad (17)$$

The above is the core computing idea of EM algorithm. Next, the naive Bayesian model would be analyzed in detail. The first is the discrete calculation method, as shown in Formula (18).

$$P(x_j | c_i) = \frac{|X_{c_i x_j}|}{|X_{c_i}|} \quad (18)$$

Secondly,  $X_{c_i}$  represents the sample of category  $c$ ,  $x$  is a set of data in the sample of this

category, and  $x_j$  represents the first attribute of the data set. Then, according to Formula (19), the probability of a certain data set under a certain condition is calculated.

$$P(x_j|c_i) = \frac{1}{\sqrt{2\pi}\sigma_{ci}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma_{ci}^2}\right) \quad (19)$$

Then the given set of data set  $x$  and the above results are brought into the multivariate normal distribution function, and the class conditional probability can be obtained, and the formula is shown in Formula (20).

$$P(x_j|c_i) = (2\pi)^{-m/2} \cdot |\sigma_{ci}|^{-1/2} \cdot \exp\left[-\frac{1}{2}(x - \mu)\right] \quad (20)$$

Finally, for the case where the conditional probability of a certain value class is 0, the Laplace correction is used to solve it, which is shown in Formula (21).

$$P(x_j|c_i) = \frac{|x_{ci}|+1}{|c_i|+N_i} \quad (21)$$

The above algorithms are some core calculation formulas of several main algorithms commonly used in data mining technology, which can analyze and calculate a large amount of data. In this way, some data with higher value can be extracted from the huge data to help decision makers better adjust their own marketing strategies.

## 5. Experiment on Financial Evaluation Based on Data Mining

This section studies the application of data mining technology in the financial analysis of modern enterprises in the era of big data. Through a series of experiments to verify the performance improvement of financial analysis that integrates data mining, the international research and development progress of related technologies is investigated. In addition, the development progress of the current global data mining technology was studied, and some large enterprises were modeled. Referring to its published economic statements, it uses the financial analysis mode integrating data mining technology and the traditional financial analysis mode to analyze and predict the enterprise, and finally compare the calculation results of the virtual model with the actual results. Among them, the market share of various algorithms of data mining technology in recent years is shown in Figure 4.

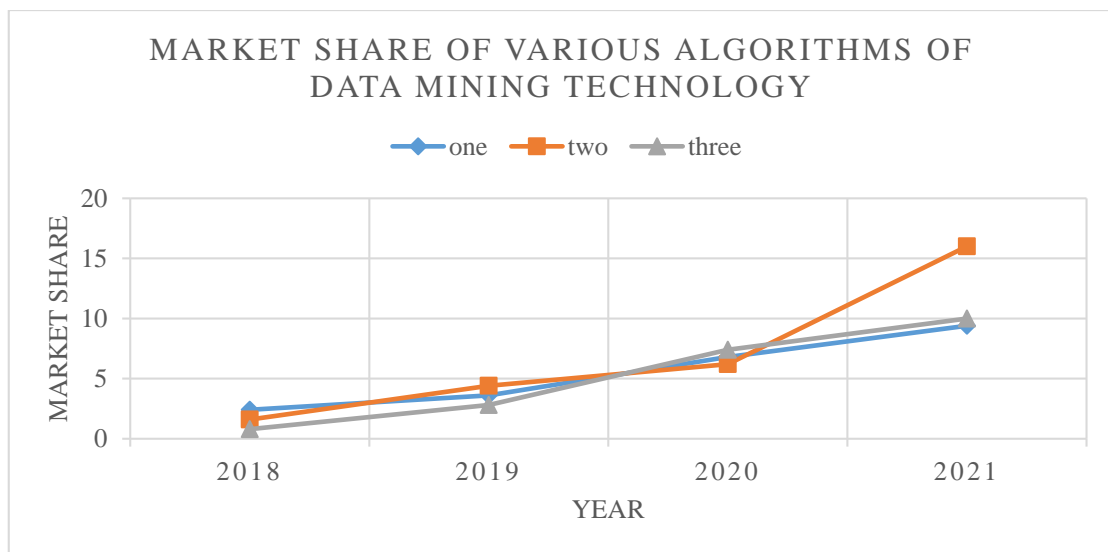


Figure 4: Market share of various algorithms of data mining technology in recent years



From Figure 4, the market shares of three different algorithm models in data mining from 2018 to 2021 can be seen, indicating that the market shares of these three algorithms in data mining technology has been steadily increasing in recent years. Among them, the second data mining algorithm has achieved a leap in market share in 2021 due to the continuous improvement of its algorithm research and development progress, which also shows that data mining technology as a whole still has a very huge room for development.

Financial analysis can be said to be a comprehensive marginal subject developed on the basis of enterprise economic analysis, financial management and accounting. It has a complete theoretical system. There are many methods of financial analysis, such as trend analysis, which compares the same indicators in multiple periods of corporate financial reports to clarify the trend of increase or decrease in each index of the enterprise, thereby reflecting the overall development trend of the enterprise; the ratio analysis method is an analysis method that uses the ratio of two related indexes in the financial statements of the enterprise to indicate the financial status and operating results of the enterprise. The market share of various calculation models of financial analysis technology in recent years is shown in Figure 5.

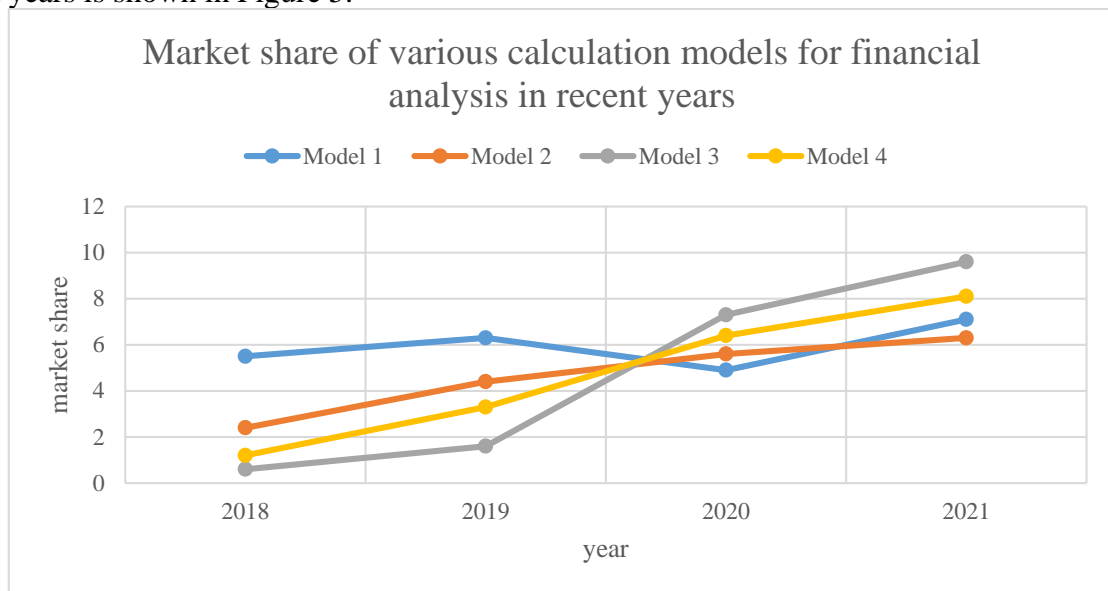


Figure 5: Market share of various calculation models for financial analysis in recent years

Figure 5 shows the development trend of financial analysis calculation models in recent years. Among them, the market share of Model 1 would decrease in 2020 because the calculation model is too old, but its market share would continue to increase after the calculation model of Model 1 is updated in the following year. In addition, the market share of other financial analysis and calculation models has been steadily increasing with the continuous development of Internet technology. Combined with Figure 4, with the continuous integration of the financial analysis calculation model with the data mining technology, there is still a lot of room for development in the future.

With the rise and rapid development of Internet technology, some emerging information technologies have been continuously penetrated into various fields, making the operation mode of various fields have undergone great changes. Among them, the application of data mining technology to the daily work of enterprise financial management has made data mining technology into a new financial data processing method. The extraction, analysis and transformation of financial data that previously required different financial personnel are all handled by data mining technology, which greatly saves the daily work time of financial personnel. In the meanwhile, under

the background of the rapid development of big data technology, enterprises also put forward new requirements for financial analysis methods, that is, some repetitive calculations are required to be done by the computer, and the financial personnel of the enterprise are only responsible for the judgment of some highly subjective options. The performance comparison diagram of financial analysis integrated with data mining technology and traditional financial analysis is shown in Figure 6.

It can be seen from Figure 6 that the processing ability of the financial analysis model based on data mining technology for four different types of sample data has been improved compared with the processing ability of traditional financial analysis for sample data. The processing capability of the analysis for the fourth type of sample data is greatly improved. On the whole, the performance of financial analysis based on data mining is much higher than that of the original financial analysis model in the analysis of 1, 3 and 4 types of financial sample data, and the ability to process the second type of financial data is comparable. Based on this performance difference, the article finds that the performance of financial analysis based on data mining technology is improved by about 63.4% compared with the traditional financial analysis model. With the advancement of data mining technologies, it is believed that the financial analysis calculation model based on data mining technology still has a relatively development prospect in the future.

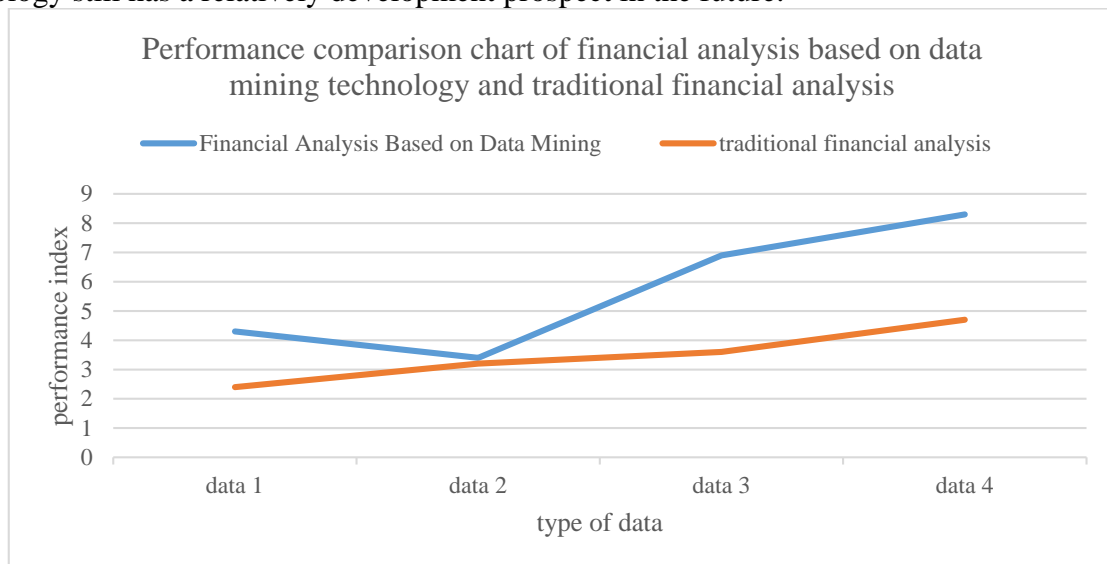


Figure 6: Schematic diagram of the comparison of financial analysis performance indexes before and after optimization

## 6. Conclusions

The development of big data technology not only brings new opportunities for traditional financial analysis of enterprises, but also has a great impact on the accounting industry. From the professional point of view of enterprise development, in the context of the rapid development of big data technology, this paper studied the possibility of integrating data mining technology into enterprise financial analysis to replace part of the repetitive manual work. Through a series of modeling experiments, the financial analysis performance of the data mining technology was compared to the traditional financial analysis performance of the enterprise. It not only verified the possibility of their mutual combination, but also provided new ideas for more and more enterprise financial departments to improve the work efficiency of financial staff. However, due to the fact that most small and medium-sized enterprises currently have insufficient infrastructure for data mining related work, financial analysis technology based on big data can still only be used in some large

enterprises, but it is believed that over time, some infrastructures of enterprises are gradually improved, and financial analysis combined with data mining technology would be used by more and more enterprises.

## Acknowledgement

This work was supported in part by the Key Scientific Research Project of Universities in Anhui Province under Grant KJ2021A1237, in part by the Anhui Provincial Higher Education Institutions' Quality Project under Grant 2022jyxm949.

## References

- [1] Barrette, Thiffault, Achim. A financial analysis of the potential of dead trees from the boreal forest of eastern Canada to serve as feedstock for wood pellet export[J]. *APPL ENERG*, 2017, 27(198):410-425.
- [2] Deependra N, Sagar K, Samundra G. Optimal sizing and financial analysis of a stand-alone SPV-micro-hydropower hybrid system considering generation uncertainty[J]. *International Journal of Low-Carbon Technologies*, 2021, 22(4):16-26.
- [3] Melbostad H, Savage P, Mahoney K. Financial Analysis of Cardiac Rehabilitation and the Impact of COVID-19.[J]. *Journal of cardiopulmonary rehabilitation and prevention*, 2021, 2 (5):308-314.
- [4] Halim MA, Islam MN, Khan AG. Comparative Financial Analysis of Conventional and Islamic Banks of Developing Countries[J]. *Business Administration Studies*, 2020, 30(4):34-42.
- [5] Kacker S, Katz LM, Ness PM. Financial analysis of large-volume delayed sampling to reduce bacterial contamination of platelets[J]. *Transfusion*, 2020, 47(5):997-1002.
- [6] Shooshtarian A, Anderson JA, Armstrong GW. Growing hybrid poplar in western Canada for use as a biofuel feedstock: A financial analysis of coppice and single-stem management[J]. *Biomass and Bioenergy*, 2018, 21 (113):45-54.
- [7] Ma Y. Feasibility Study on Listing of Pharmaceutical Enterprises Based on Financial Analysis: Take Company Y as an Example [J]. *Open Journal of Business and Management*, 2021, 41(3):1325-1337.
- [8] Yan XS, Zheng L. Fundamental Analysis and the Cross-Section of Stock Returns: A Data-Mining Approach[J]. *Review of Financial Studies*, 2017, 91(4):1382-1423.
- [9] Namahoot KS, Jantasri V. Forecasting the stock exchange of Thailand using data mining techniques[J]. *International Journal of Electronic Finance*, 2021, 38(4):211-242.
- [10] Li XF, Zhang C, Lin X C. Research on Default Risk of Peer-To-Peer Online Lending Based on Data Mining Algorithm[J]. *Journal of Computer Science*, 2020, 56(2):83-100.
- [11] Protopapadakis E, Niklis D, Doumpos M. Sample selection algorithms for credit risk modelling through data mining techniques [J]. *International Journal of Data Mining Modelling and Management*, 2019, 12(2):10-13.
- [12] Lokanan ME. Data mining for statistical analysis of money laundering transactions[J]. *Journal of Money Laundering Control*, 2019, 69(4):753-763.
- [13] Werner M. Financial process mining-Accounting data structure dependent control flow inference[J]. *International Journal of Accounting Information Systems*, 2017, 87(25):57-80.
- [14] Milutinovi S, Grljevic O. Analysis of Accountants' Attitudes on Regulation Using Data Mining[J]. *Ekonomika Preduzeca*, 2020, 33(56):341-353.
- [15] Yang Y. Balanced Scheduling Method for Big Data of Network Traffic Based on Set-pair Analysis Strategy[J]. *Journal of Physics: Conference Series*, 2022, 42(1):57-65.
- [16] Nikolay T, Vadim D. Treatment strategy for patients with chronic coronary disease: surgical data versus big data—who is right?[J]. *European Journal of Cardio-Thoracic Surgery*, 2022, 62(4):5-10.
- [17] Chen HT, Gong C D, Xiong C. Study on the rules of TCM syndromes and Chinese herbal medicines about IgAN based on data mining[J]. *Journal of Hainan Medical University*, 2021, 51(12):38-42.
- [18] Gurusubramani S, Mouleeswaran S K, Srinivas P. A Data Centre Configurable Data Mining Document Management Information System[J]. *Journal of Physics: Conference Series*, 2021, 28(4):82-95.
- [19] Haryati U, Irawan, Maswar. Application of mulch and soil ameliorant for increasing soil productivity and its financial analysis on shallots farming in the upland[J]. *IOP Conference Series Earth and Environmental Science*, 2021, 44(1):121-125.
- [20] Liu H. Financial Risk Intelligent Early Warning System of a Municipal Company Based on Genetic Tabu Algorithm and Big Data Analysis. *International Journal of Information Technologies and Systems Approach*, 2022, 15(3): 1-14.