Corporate Accounting Management Risks Integrating Improved Association Rules and Data Mining

HAIYAN LI
Claro M Recto Academy of Advanced Studies,
Lyceum of the Philippines University,
PHILIPPINES

Abstract: - With the development of the times, enterprises need to face more data in operational decision-making. Traditional data analysis strategies cannot handle the growing amount of data well, and the accuracy of analysis will also decrease when faced with uneven data types. The research uses a corporate accounting management risk analysis technology that combines big data algorithms and improved clustering algorithms. This method combines big data processing ideas with a clustering algorithm that incorporates improved weighting parameters. The results show that on the data sets DS1, DS2, and DS3, the NMI values of the GMM algorithm are all 0; while the NMI values of the MCM algorithm correspond to 0.9291, 0.9088 and 0.8881 respectively. At the same time, the Macro-F1 values of the Verify2 algorithm correspond to 0.9979, 0.9501, and 0.9375 respectively, and the recognition accuracy of the data remains above 85%. In the running time comparison, when the number of samples in the data set reaches 5,000, the calculation time of the Verify2 algorithm remains within 5 seconds. In terms of practical application results, the study selected the profitability risk indicators of 40 companies for analysis. After conducting risk ratings, it can be seen that companies No. 5, 6, 7, and 39 have the highest risk levels, and companies No. 33 and 34 have the highest risk levels. The lowest level. After conducting risk assessments on the 40 selected listed companies, the risk level of net asset income of each company remained at level 5, and the risk level of earnings per share remained at level 3. The above results show that this technology has good performance in terms of calculation accuracy and calculation time, can assess enterprise risks, and can provide data support for enterprise operation decisions.

Key-Words: - Data mining; Big data; Clustering algorithm; Risk assessment; Association rules; Enterprise operation decisions.

Received: January 12, 2024. Revised: April 14, 2024. Accepted: June 16, 2024. Published: July 22, 2024.

1 Introduction

With the development of economic globalization and computer technology, the market economy has entered a new stage, [1]. More and more market data need to be mined, analyzed, and processed. Enterprises that can take advantage of the new environment must have excellent data processing and analysis capabilities, [2], [3]. Accounting management risks not only affect the financial health of enterprises but also directly affect their sustainable development and market competitiveness. Therefore, accurately identifying and effectively controlling risks in accounting management has become an important issue in enterprise management. Traditional accounting management risk analysis mainly relies on the analysis of financial statements and internal auditing. Although these methods can to some extent reveal the financial situation of enterprises, they have obvious shortcomings in handling large amounts of complex data, predicting potential risks, and exploring deep-seated risk factors, [4]. As enterprises need to consider more and more data in the risk decision-making process, traditional data analysis methods cannot effectively handle complex data information, and it is difficult to solve the problem of data unevenness and data type. In recent years, various artificial intelligence algorithms have developed rapidly. Among them, improved association rules can improve the efficiency and accuracy of data processing through optimization algorithms. The powerful ability of data mining technology to process big data makes this method suitable for dealing with complex enterprises. More advantages when managing problems, [5]. The application of this method not only helps enterprises discover and respond to financial risks promptly, but also promotes the optimization of internal controls and scientific management decisions. Against this background, this experiment proposes a new method that integrates improved association rules and data mining technology, aiming to analyze corporate accounting management risks more comprehensively and in-depth. In the process, the data analysis process was first set up and the risk assessment indicators for corporate accounting management were selected; then a data mining algorithm was selected to process the corporate data, and then a cluster analysis algorithm and a big data algorithm were combined to complete the analysis of the corporate data. Effective mining of accounting management-related data; finally, the effectiveness and reliability of the proposed method are verified through performance testing and application analysis.

There are two innovative points in the research. The first point is that the research uses a combination of big data and clustering algorithms. The second point is that the study combines enterprise risk assessment with accounting work experience and conducts risk assessment on enterprises from a professional perspective.

The main content of the research is divided into four parts. The first part is a summary of the current domestic and foreign research on corporate accounting management risk technologies related to big data and improved clustering algorithms; the second part is an introduction to the corporate accounting management method proposed by the laboratory that integrates improved association rules and data mining; the third part is the performance analysis and application effect testing of the constructed algorithm; the fourth part is a summary of the methods and results of the entire article, and also analyzes the development direction of future research.

2 Literature Review

Management risk research can reduce corporate risks and increase industry profits. With the development of new technologies, management risk assessment research has attracted more and more attention from scholars. [6], used the K-Means data mining method when classifying poverty lines according to counties/cities in North Sumatra Province to understand the poverty risk in cities. The research results provide information for the economic allocation of the North Sumatra government to further overcome the poverty problem in the region. [7], used the particle swarm optimization algorithm when studying management risks in public sector organizations. The research results indicate that RM issues are not well integrated at the MACS level, a thorough cultural change is still needed, and future RM research must provide empirical data on integration in practice to reduce management risks. [8], used the R bibliometric application to process and analyze data when studying the development trends of environmental accounting published in domestic and foreign journals, Research results show that the most popular keywords at the moment are energy, environment, and assessment. [9], investigated how students view Facebook's help in accounting learning from aspects such as ease of use, usefulness, attitudes towards Facebook usage activities, and student performance. The findings indicate that student performance and course learning outcomes are most likely to improve when students actively participate through the course Facebook group. [10], used a method of readjusting learning and teaching strategies when studying the impact and response measures of COVID-19 in accounting education. The findings indicate that it identifies issues that need to be addressed during the recovery and redesign phases of crisis management and sets a new research agenda for accounting education research.

[11], used data mining methods when summarizing the use of traditional Chinese medicine to preserve ejection fraction in the treatment of heart failure. The database was established using Microsoft Excel 2019, and then the apriori algorithm and hclust function were used in R-Studio (version 4.0.3) for association rule analysis and hierarchical clustering analysis respectively. Research results show that the treatment methods for this disease are to replenish qi, warm yang, activate blood circulation, and diuresis. Astragalus and salvia are the basic compatibility of traditional Chinese medicine. [12], used data mining methods when studying the prescription patterns of different dosage forms of Chinese herbal medicine in the treatment of rheumatoid arthritis (RA) and their impact on immune and inflammatory indicators. Each prescribed herbal medicine was quantified and standardized against the knowledge base to build a database of RA treatment formulas. The research results show that immune and inflammatory indicators have been significantly improved after treatment with traditional Chinese medicine granules and decoction pieces, and there is a long-term correlation between comprehensive evaluation indicators and intervention measures. [13], conducted a systematic and comprehensive review of various data mining tasks and techniques when studying research trends in the field of data mining. The research results introduce various practical applications of data mining and challenges and problems faced by the field of data mining research. When studying the main impact of e-learners'

satisfaction on the e-learning process, [14], used data mining technology to identify relevant factors that affect student learning outcomes. The research results illustrate the impact of e-learning on student performance. When studying spatial data mining components, scholars such as [15], adopted the extended adjacency spatial clustering method based on density and grid and used middleware technology to complete an agricultural geographic information system based on MapXtreme. The research results show that this method solves the problem of agricultural informatization and improves the optimization performance.

Through analysis of existing literature, we found that although traditional accounting management methods can identify and control risks to a certain extent, they have limitations in processing large-scale data sets, predicting future trends, and revealing deep correlations. In contrast, improved association rules and data mining technology can more comprehensively identify risks and predict possible risk trends through efficient data analysis capabilities, thereby providing a more scientific basis for corporate decision-making. In addition, current research also emphasizes the important role of data mining technology in discovering hidden patterns, identifying abnormal behaviors, and optimizing internal controls of enterprises. Combined with improved association rules, this approach can improve the effectiveness of enterprise risk management while ensuring the accuracy and efficiency of data analysis. Given this, the experiment proposes a corporate accounting management risk assessment method that integrates association rules and data mining to further analyze the application of this method in different types of enterprises and diversified business environments, and how to better integrate these technologies into the enterprise's risk management framework to provide more effective support for the sustainable development and risk control of enterprises.

3 Model Construction of Risk Analysis Technology that Combines Big Data Algorithms and Cluster Analysis Algorithms

With the development of society, the pressure of data analysis faced by enterprises in the decision-making process is gradually increasing. The study designed data mining technology that combines big data and clustering algorithms to solve this problem. The study introduces the operating ideas of the risk assessment model and the selection criteria of risk factors. According to the characteristics of corporate accounting data, a more suitable improved clustering algorithm is selected to process these data, and the design ideas of the improved clustering algorithm are introduced.

3.1 The Construction Idea of Risk Analysis Technical Model and the Selection Method of Risk Assessment Indicators

For a long time, accounting work has focused on financial accounting and ignored the importance of management accounting. With the development of global economic integration, complex market information has brought greater pressure to corporate decision-making. At the same time, the development of Internet-related technologies also provides support for accounting management work, [16]. Management accounting work based on data mining can conduct value analysis on massive data and provide information support for corporate decision-making. In data mining work, research is based on big data methods and combined with cluster analysis algorithms to improve the performance of data analysis algorithms. During the analysis process, due to differences in data types and sources, data analysis algorithms need to combine multiple processing algorithms to build a complete application process. The study set up the data analysis process based on the common data types in corporate accounting work. The specific flow chart is shown in Figure 1.

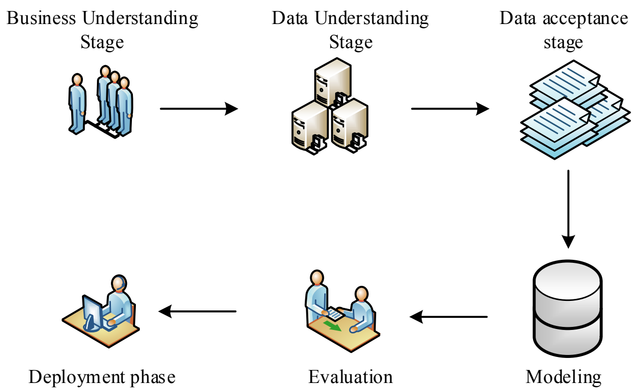

Fig. 1: Workflow diagram of data mining

In Figure 1, the study divides the data mining process into the business understanding stage, data understanding stage, data preparation stage, model establishment stage, model evaluation stage, and algorithm deployment stage. Among them, the main purpose of the business understanding stage is to set analysis indicators based on the objective nature of the problem and initially judge the number and type of influencing factors. The main purpose of the data understanding stage is to quantitatively analyze the indicators set in the business understanding stage and convert various types of abstract data into quantifiable standards to facilitate subsequent data classification work, [17]. The main work of the data preparation stage is to sort out the collected quantitative data, filter out erroneous data, and reduce the dimensionality of the data to improve the speed of data processing. In the subsequent model establishment and evaluation stages, the data is sorted to establish an analysis model of the data, and then through multiple iterations, the actual fit of the model is gradually improved. Finally, in the algorithm deployment stage, the results obtained from data mining are analyzed with the actual situation and finally compiled into a report or management accounting report. In the selection of enterprise risk analysis indicators, we studied the selection criteria of enterprise financial capabilities as indicators, and finally formulated four indicators. Among them, the profitability indicators of the enterprise are shown in Figure 2.

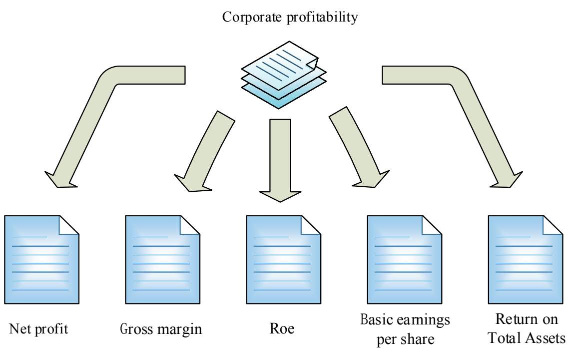

Fig. 2: Enterprise profitability indicators

Figure 2 briefly describes the influencing factors of a company's profitability. Corporate profitability reflects a company's ability to create profits. Among them, net profit margin and gross profit margin reflect the company's ability to generate income within a certain period, the company's ability to resist risks, and the company's operational fault tolerance. The subsequent three indicators describe the investment value of the company, will affect the company's financing ability, and reflect the company's development potential. Afterward, the study set up the solvency indicators of the enterprise. The specific solvency indicators are shown in Table 1.

Table 1. Debt paying ability indicators of enterprises

Financial index	Define _
Quick ratio	The ratio of all current assets minus inventory to current liabilities
Current ratio	The ratio between all current assets and all current liabilities
Cash ratio	The ratio of all cash plus securities to current liabilities
Asset liability ratio	The ratio between total liabilities and total assets
Interest coverage ratio	The ratio of current operating profit to interest expense of the enterprise

Table 1, the solvency of an enterprise is divided into long-term solvency and short-term solvency. Although the solvency will not affect the operation of the enterprise, it will affect the credit of the enterprise, [18]. Indirectly reduce the financing ability of enterprises and increase the difficulty of enterprise development. After setting evaluation indicators for the company's profitability and solvency, the study also analyzed the company's operational capabilities. The specific operational capability evaluation indicators are shown in Figure 3.

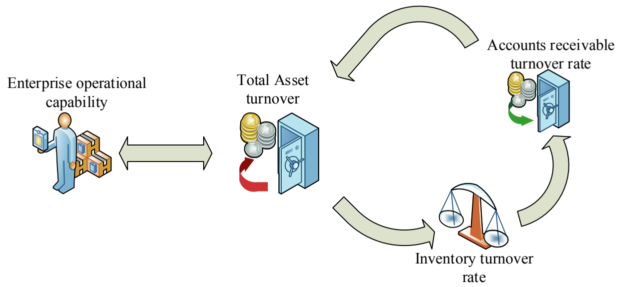

Fig. 3: Operational capability indicators of enterprises

Figure 3, the study uses the company's turnover capacity as an evaluation index to measure the quality of the company's operating capabilities. The liquidity of an enterprise reflects its good risk tolerance and good credit in the market. Therefore, it can show better resilience when encountering operational risks. Finally, the study also included the growth ability of enterprises as one of the evaluation indicators of enterprise risk management. Among them, the various evaluation factors of specific enterprise growth capability indicators are shown in Figure 4.

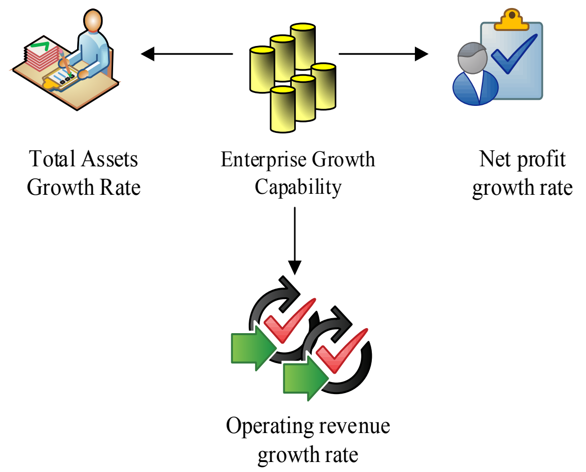

Fig. 4: Growth capability indicators of enterprises

Figure 4, the assessment of corporate growth capabilities reflects the company's ability to continuously exaggerate its development potential through business activities. Among the enterprise growth capability indicators, the total assets growth rate reflects the overall growth trend of the enterprise. The revenue growth rate reflects the current growth rate of the company and also reflects the increase in the number of products sold and the number of services provided by the company. Net profit growth reflects the growth rate of the company's disposable funds and the error tolerance of risk decisions.

3.2 Algorithm Design Ideas and Key Parameter Solution Methods for Risk Assessment Technology

After completing the setting of enterprise risk assessment indicators, it is necessary to select a data mining algorithm to process enterprise data. The study uses a combination of cluster analysis algorithm and big data algorithm to complete the data mining task in enterprise risk management analysis, [19]. Since accounting risk management belongs to non-uniform data, the research uses a Spark-based non-uniform data clustering algorithm to perform data mining, [20]. Traditional clustering algorithms require more iterative calculations when performing cluster division tasks, but Spark changes the calculation mode of traditional clustering algorithms and reduces the computational complexity. The specific parallel computing process is shown in Figure 5.

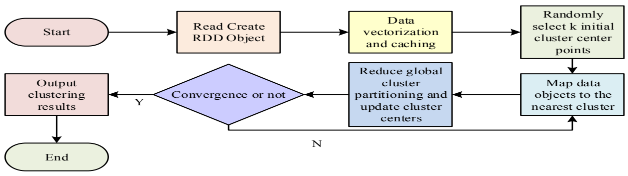

Fig. 5: Improved clustering algorithm

Figure 5, the parallel processing idea can be mainly divided into six stages. In the first stage, data is imported locally and converted into RDD objects to divide task nodes. In the second stage, the data is initialized and the representative points of the initial clusters are obtained, and then the representative points are imported into all working nodes. In the third stage, the sample points are divided into clusters based on the center point. In the fourth stage, the global cluster center node is updated, and after the returned data is obtained, the global cluster center point is calculated. In the fifth stage, the results are analyzed based on actual needs and it is decided whether it is necessary to continue iteration. In the sixth stage, the clustering results are output and the evaluation conclusions of corporate accounting risk management are drawn. In the process of constructing a probability model for non-uniform data, there are often differences in cluster densities. To characterize this difference, it is necessary to design a j probability density function based on the attributes of the sample. The specific form is shown in formula (1).

$$p(x_j, v_{kj}, w_{kj}, \sigma_k) = \frac{1}{\sqrt{2\pi}\,\frac{\sigma}{\sqrt{w_{kj}}}} \exp\left(-\frac{\left(x_j - v_{kj}\right)^2}{2\frac{\sigma_k^2}{w_{kj}}} \right) \quad (1)$$

In formula (1), C_k represents the variance of the cluster. Because in actual situations, the probability of an attribute is often multi-dimensional. To simulate this situation, the study estimates the probability of the vector through the product of a set of marginal distribution variables based on the formula (1). The form of the density function changes as shown in formula (2).

$$p(x_j, v_{kj}, w_{kj}, \sigma_k) = \prod_{j=1}^{D} p(x_j, v_{kj}, w_{kj}, \sigma_k) \quad (2)$$

Formula (2), $P(x)$ represents the probability density of any sample in the cluster. After obtaining the probability density change of any sample, it is also necessary to consider that the same data may be included in multiple clusters. Therefore, constraints need to be set, as shown in formula (3).

$$\begin{cases} \forall k : \alpha_k > 0 \\ \sum_{k=1}^{K} \alpha_k = 1 \end{cases} \tag{3}$$

Formula (3), a_k is the size mark of the cluster. In the process of data processing, the size of the cluster needs to represent the weight information of each cluster. Therefore, the model of non-uniform data should be rewritten as shown in formula (4).

$$L(\square) = \prod_{k=1}^{K} \prod_{x \in C_k} \alpha_k \times p(x, v_k, w_k, \sigma_k) \tag{4}$$

Formula (4), is the set of group parameters. Based on the above model, the clustering problem can be transformed into DB to obtain optimized parameters to maximize the weighted likelihood. The specific form of the data processing model should be as shown in formula (5),

$$\max J_1(\square) = \sum_{k=1}^{K} |C_k| \ln \alpha_k + \sum_{k=1}^{K} \sum_{x \in C_k} \ln p(x, v_k, w_k, \sigma_k) \tag{5}$$

Formula (5), the model undergoes logarithmic transformation based on the original model. Put the reality obtained after transformation into formula (1) and formula (2), you can remove the constant terms in the formula and rewrite the formula. The specific form of the rewritten formula is shown in formula (6),

$$\max J_2(\square) = -\sum_{k=1}^{K} |C_k| \ln \alpha_k + \frac{1}{2} \sum_{k=1}^{K} |C_k| \sum_{j=1}^{D} \ln \frac{\sigma_k^2}{w_{kj}} + \frac{1}{2} \sum_{k=1}^{K} \sum_{j=1}^{D} \sum_{x \in C_k} \frac{w_{jk}(x_j - v_{kj})^2}{\sigma_k^2} \tag{6}$$

Formula (6), is the optimization objective function of the algorithm and is a constant term. Differences between different clusters are represented by differences in constant terms. In the clustering algorithm, it is necessary to constrain the value of the feature weight of the algorithm. Among them, the specific form of the traditional feature weight constraint is shown in formula (7),

$$\begin{cases} \sum_{j=1}^{D} w_{kj} = 1 \\ 0 \le w_{kj} \le 1 \\ j = 1, 2, \cdots, D \end{cases} \tag{7}$$

Formula (7), although the traditional clustering algorithm's method of constraining feature weights can well solve the problem of sample feature information loss, it will affect the algorithm's effect on feature selection. Therefore, the study adopts an improved feature weight constraint formula, the specific form of which is shown in formula (8),

$$\begin{cases} \prod_{j=1}^{D} w_{kj} = 1 \\ 0 < w_{kj} \\ j = 1, 2, \cdots, D \end{cases} \tag{8}$$

Formula (8), the weight constraint used effectively amplifies the differences between features and improves the classification accuracy of the algorithm. Afterwards, the Lagrange multiplier method is used to introduce the constraints of the feature weight parameters into the objective function, and the form of the objective function can be rewritten as shown in formula (9),

$$J(\square)_2 = J(\square)_1 + \sum_{k=1}^{K} \lambda_k (1 - \prod_{j=1}^{D} w_{kj}) + \eta (1 - \sum_{k=1}^{K} \alpha_k) \tag{9}$$

Formula (9), is the Lagrange multiplier. Since the above objective function is a nonlinear function, it is difficult to obtain the global optimal solution. Therefore, it is necessary to rewrite the formula and obtain each parameter separately. Fix the parameters w, v, x and obtain the parameters G. The calculation formula of the parameters is as shown in formula (10),

$$\begin{cases} z = \arg_k \max \dfrac{\alpha_k G(x_i)}{\sum_{j=1}^{D} \alpha_j G(x_i)} \\ G(x_i) = \prod_{j=1}^{D} \dfrac{\sqrt{w_{kj}}}{2\pi\sigma_k} \exp\left(\dfrac{-w_{kj}(x_{ij} - v_{kj})^2}{2\sigma_k^2} \right) \end{cases} \tag{10}$$

Formula (10), by comparing the probability of each Gaussian component, the sample is divided into the cluster with the highest probability, and the parameters are obtained. After calculating the parameters α_k, the solution of the parameters is shown in formula (11),

$$\alpha_k = \frac{|C_k|}{N} \tag{11}$$

Formula (11), σ_k^2 is the number of samples in the cluster, and is w_{ij} the total number of samples, The variance expression of each cluster can be obtained through the different number of samples between each cluster in non-uniform data (12),

$$\sigma_k^2 = \frac{\sum_{j=1}^{D} \sum_{x_i \in C_k}^{K} w_{kj}(x_{ij} - v_{kj})^2}{D * |C_k|} \tag{12}$$

Formula (12), after obtaining the variance expression of each cluster. Fixed parameters w, x, calculated parameters v, and the expression of the parameters can be obtained as shown in formula (13),

$$v_{kj} = \sum_{x_i \in C_k} x_{ij} / |C_k| \tag{13}$$

In v, x formula (13), by fixing the parameters and calculating the parameters w , the parameter expression (14) can be obtained,

$$w_{kj} = \left(|C_k| + \lambda_k\right) \frac{\sigma_k^2}{X_{kj}} \tag{14}$$

Formula (14), the specific expression of the parameters is as shown in formula (15),

$$\begin{cases} X_{kj} = \sum_{x_i \in C_k} (x_{ij} - v_{kj})^2 \\ \lambda_k = \left(\frac{1}{\sigma_k^2} (\prod_{j=1}^{D} \frac{1}{X_{kj}})^{\frac{1}{D}} \right) - |C_k| \end{cases} \tag{15}$$

Formula (15), after limiting the value range of the parameters and obtaining each parameter separately. The clustering algorithm is constructed.

4 Performance Analysis and Application Effects of Risk Assessment Technology Combining Big Data Algorithms and Improved Clustering Algorithms

In actual production and life, enterprise decision-making often has to face complex data types and huge amounts of data. To better deal with these problems, the research uses a non-uniform clustering algorithm to complete the problems of data mining and data processing. To study the data processing capabilities of the clustering algorithm, the study used a random number generation function to generate three data sets. By changing the data dimensions in the data set and setting the variance to enhance the data dispersion, we simulate the situation of uneven data types in actual situations. Among them, the specific situation of the data set is shown in Table 2.

Table 2. Characteristics of synthetic datasets

Datasets	Clusters	Dimensions	Variance
DS1	20:50:00	10	0.21:0.14
DS2	2000:100	50	0.90:0.64
DS3	5000:200	100	1.64:1.34

Table 2, the three synthetic data gradually increase in sample number and data complexity to examine the data processing performance of the improved clustering algorithm. To verify the performance of the improved clustering algorithm, the study selected the GMM algorithm and MCN algorithm to compare the performance of the improved clustering algorithm. To measure the algorithm performance of the improved algorithm, the specific comparison results are shown in Table 3.

Table 3. Clustering results of different algorithms

Measure	Datasets	MCM	GMM	Verify2
Macro-F1	DS1	0.4237	0.5595	0.9979
	DS2	0.4132	0.5331	0.9501
	DS3	0.4477	0.5148	0.9375
NMI	DS1	0.9291	0	0.0058
	DS2	0.9088	0	0.0147
	DS3	0.8881	0	0.0320

In Table 3, on the three data sets of DS1, DS2, and DS3, the F1 values of the Verify2 algorithm are 0.9979, 0.9501, and 0.9375 respectively; the F1 values of the GMM algorithm are 0.5595, 0.5331, and 0.5148 respectively; the F1 values of the MCM algorithm correspond to 0.4237, 0.4132 and 0.4477. In addition, the NMI values of the GMM algorithm on the three data sets are 0, while the NMI values of the MCM algorithm on the three data sets correspond to 0.9291, 0.9088, and 0.8881 respectively; in addition, the NMI values of the Verify2 algorithm correspond to 0.0058, 0.0147, and 0.0320 respectively. This shows that the improved clustering algorithm shows higher recognition accuracy, and the recognition accuracy of the data remains above 85%. Since the amount of market data is generally large, to better demonstrate the operation of the three algorithms on synthetic data sets. The study tested the robustness of the three algorithms and tested the change curve of the calculation time by gradually increasing the number of samples. The calculation time consumption of different algorithms was statistically calculated. The specific situation is shown in Figure 6.

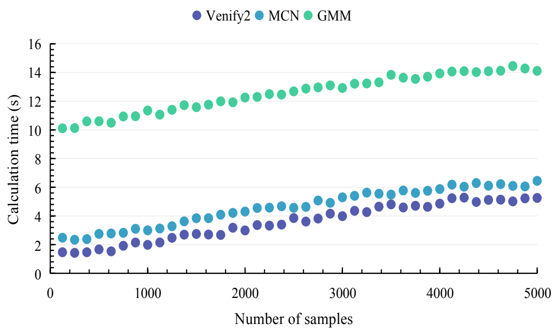

Fig. 6: Running time of different algorithms on the dataset

In Figure 6, as the amount of data increases, the running time of the three algorithms shows an increase in varying degrees. When the amount of data increases to 5000, the running time of the three algorithms has a maximum value. At this time, the corresponding running times of the GMM, MCM, and Verify2 algorithms are 14.54s, 6.89s, and 4.98s respectively. Throughout the entire experiment, the running time of the Verify2 algorithm has been lower than that of the other two algorithms. When the amount of data reaches 5,000, the running time of the Verify2 algorithm is always less than 5 seconds. The reason why the calculation time of the GMM algorithm is significantly lower than that of the Verify2 algorithm is that the Verify2 algorithm uses a spectral clustering algorithm, which makes the improved clustering algorithm save more time when performing matrix operations and is less affected by matrix operations. The comparison shows that the Verify2 algorithm took a shorter time during the experiment, but the amount of sample data processed was equivalent to the other two methods. This shows that when the Verify2 algorithm is applied to the enterprise accounting management system, the system runs faster and more efficiently, processes data faster than the other two algorithms, and has strong data processing capabilities. Then, non-uniform clustering is combined with big data algorithms to deal with the prediction process of corporate accounting risks. The study combines the 16 corporate risk indicators mentioned above, including gross profit margin, net profit margin, return on equity, basic earnings per share, return on total assets, quick ratio, current ratio, cash ratio, asset-liability ratio, and interest coverage ratio., total asset turnover rate, accounts receivable turnover rate, inventory turnover rate, total asset growth rate, net profit growth rate, and operating income growth rate, respectively, are set to the 16 English letters of ap, To explore the correlation between each influencing factor and enterprise risk, the Pearson coefficient between each risk indicator is shown in Figure 7.

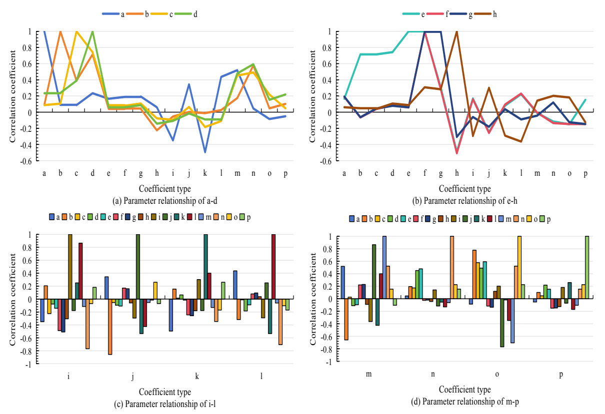

Fig. 7: Correlation coefficients of different risk indicators

Figure 7, by analyzing and processing the correlation of risk factors, the multi-dimensional situation of each risk factor can be obtained. Since the risk factors at this time are generally discretized, although the specific values of each factor can be obtained, the risk level reflected by the data is not yet clear. Therefore, it is also necessary to transform risk factors from discrete data into continuous data through relevant financial data indicators. After completing the data processing of the influencing factors, the risk factors are brought into the improved non-uniform clustering algorithm to analyze and predict enterprise risks. To reflect the performance of accounting risk management technology, the study selected the profitability risk indicators of 40 companies for analysis and conducted statistics and risk ratings on their gross profit margin, net profit margin, net income from assets, and earnings per share. The specific values and risk rating data of interest rate and net interest rate are shown in Figure 8.

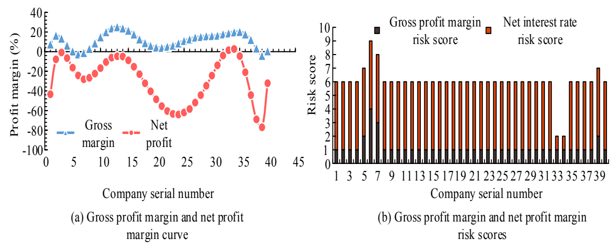

Fig. 8: Comparison of changes in gross profit margin and net profit margin before and after the transformation

Figure 8(a) shows the changes in gross profit margin and net profit margin curves. It can be found that as the company's serial number changes, the values of net profit margin and gross profit margin change in waves. During the change in net interest rate, Company No. 39 has the highest net interest rate value, which is as high as 4.56; Company No. 23 has the smallest net interest rate value, which is -62.85. During the change in gross profit margin, Company 13 has the largest gross profit margin, with a value as high as 25.89; Company No. 6 has the smallest gross profit margin, with a value of -3.44. Figure 8(b) shows the changes in gross profit margin and net profit margin risk scores. It can be seen that experiments convert data into specific risk levels. After conducting the risk rating, it can be seen that companies No. 5, 6, 7, and 39 have the highest risk levels, while companies No. 33 and 34 have the lowest risk levels. After mining and processing corporate revenue data, the risk management technology integrated with big data reflects the company's risk profile. Next, the experiment selected 40 companies and analyzed their net asset income and earnings per share. The specific results are shown in Figure 9.

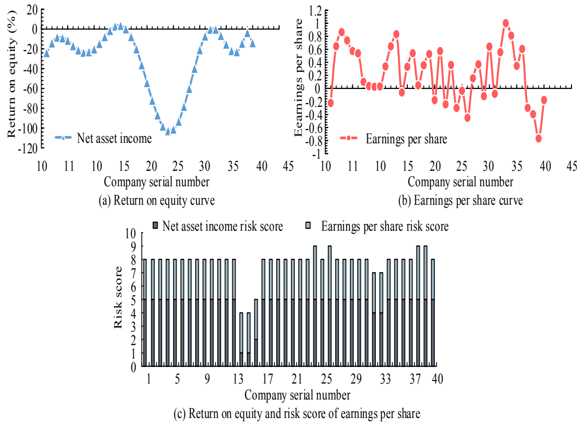

Fig. 9: Radar chart of net asset income and earnings per share before and after restructuring

In Figure 9, it can be seen that companies No. 24, 26, 38, and 39 have higher income risks, while companies No. 14 and 15 have lower income risks. By analyzing the income data of each enterprise, a clear risk level can be provided as a data reference for enterprise investment and risk avoidance from an accounting perspective. For example, Company No. 39 not only has a high risk of gross profit margin and net profit rate but also has a high-risk level of net asset income and earnings per share, making it a very dangerous investment object. The shares of companies No. 14 and 15 have higher returns, which means they are better investment targets. The gross profit margin and net income risk levels of Company No. 33 are low, which means that the company is making steady profits. Given this, it can be seen that data mining technology that combines big data and clustering algorithms can provide information support for enterprise risk assessment and quantitative scoring of various enterprise data.

5 Conclusion

With the development of computer technology and the acceleration of global economic processes, enterprises need to face increasing competitive risks. Traditional risk assessment methods cannot adapt well to larger data structures and uneven data types. Therefore, the study uses accounting risk management technology that combines big data algorithms and non-uniform clustering algorithms to provide data support for corporate risk decisions. Experimental results show that on the three data sets of DS1, DS2, and DS3, the maximum F1 values of the GMM algorithm and the MCM algorithm are 0.5595 and 0.4477 respectively; while the maximum F1 value of the Verify2 algorithm is 0.9979. At the same time, on the three data sets, the NMI value of the GMM algorithm is 0. In the comparison of different running times, as the amount of data continues to increase, the running time of the Verify2 algorithm has always been at the minimum. When the sample data volume reaches 5,000, the running time of the Verify2 algorithm still stays within 5 seconds. The experiment selected 40 companies as research objects and analyzed their profitability risk indicators. When gross profit margin and net profit margin are used as influencing factors, it can be seen that companies No. 5, 6, 7, and 39 have the highest risk levels, and companies No. 33 and 34 have the lowest risk levels. When the data related to net asset income and earnings per share are used as influencing factors, it can be seen that companies No. 24, 26, 38, and 39 have higher income risks, while companies No. 14 and 15 have lower income risks. The higher F1 value shows that the proposed method has superior evaluation ability and high recognition accuracy. The short running time shows that the method proposed in the experiment has excellent data processing capabilities and strong robustness. In addition, by analyzing and processing the correlation of risk factors, the multi-dimensional situation of each risk factor can be obtained. Comprehensive research results show that the accounting risk management technology that combines big data algorithms and non-uniform clustering algorithms has higher performance in terms of calculation speed than traditional algorithms, has better data processing and data analysis capabilities, and can provide enterprises with Provide risk level assessment function. However, at the end of the experiment, specific processes in the application and management of big data methods were proposed based on actual enterprise cases. However, when faced with different industries and problems, due to different data acquisition methods and differences in data composition and structure, specific problems still need to be developed. Correspond to specific analysis rather than generalizations. At the same time, due to other objective reasons, it was not possible to select more companies and a longer period to better reflect the analysis effect of the data mining algorithm, which can be used as a follow-up research direction.

References:

[1] W. Haoxiang and S. Smys, Big data analysis and perturbation using data mining algorithm, *Journal of Soft Computing Paradigm*, Vol. 3, No. 1, 2021, pp. 19-28.

[2] N. S. Amin, P. Shivakumara, T. X. Jun, K. Y. Chong, D. L. L. Zan, and R. Rahavendra, An augmented reality-based approach for designing interactive food menu of restaurant using android, *Artificial Intelligence and Applications*, Vol. 1, No. 1, 2023, pp. 26-34.

[3] M. Barma and U. M. Modibbo, Multiobjective mathematical optimization model for municipal solid waste management with economic analysis of reuse/recycling recovered waste materials, *Journal of Computational and Cognitive Engineering*, Vol. 1, No. 3, 2022, pp. 122-137.

[4] T. Mahmood and Z. Ali, Analysis of Maclaurin symmetric mean operators for managing complex interval-valued q-Rung orthopair fuzzy setting and their applications, *Journal of Computational and Cognitive Engineering*, Vol. 2, No. 2, 2023, pp. 98-115.

[5] Y. Fang, B. Luo, T. Zhao, D. He, B. B. Jiang, and Q. L. Liu, ST-SIGMA: Spatio-temporal semantics and interaction graph aggregation for multi-agent perception and trajectory forecasting, *CAAI Transactions on Intelligence Technology*, Vol. 7, No. 4, 2022, pp. 744-757.

[6] M. A. Hanafiah and A. Wanto, Implementation of data mining algorithms for grouping poverty lines by district/city in North Sumatra, *IJISTECH (International Journal of Information System and Technology)*, Vol. 3, No. 2, 2020, pp. 315-322.

[7] E. Bracci, T. Mouhcine, and T. Rana, Risk management and management accounting control systems in public sector organizations: a systematic literature review, *Public Money and Management*, Vol. 42, No. 6, 2022, pp. 395-402.

[8] M. Taqi, A. S. Rusydiana, N. Kustiningsih, and I. Firmansyah, Environmental accounting: A scientometric using biblioshiny, International *Journal of Energy Economics and Policy*, Vol. 11, No. 3, 2021, pp. 369-380.

[9] R. Othman and N. M. Zambi, Social media as a learning tool in cost and management accounting, *ANP Journal of Social Science and Humanities*, Vol. 2, No. 2, 2021, pp. 39-46.

[10] A. Sangster, G. Stoner, and B. Flood, Insights into accounting education in a COVID-19 world, *Accounting Education*, Vol. 29, No. 5, 2020, pp. 431-562.

[11] H. X. Guo, J. R. Wang, G. C. Peng, P. Li, and M. J. Zhu, A data mining-based study on medication rules of Chinese herbs to treat heart failure with preserved ejection fraction, *Chinese Journal of Integrative Medicine*, Vol. 28, No. 9, 2022, pp. 847-854.

[12] H. Dan, L. Jian, X. Ling, J. Xie, Q. Zhu, P. Chen, Z. Shen, Q. Meng, and H. Wang, Data mining study on prescription patterns of different dosage forms of Chinese herbal medicines for treating and improving immune-inflammatory indices in patients with rheumatoid arthritis, *Chinese Journal of Integrative Medicine*, Vol. 28, No. 3, 2022, pp. 1-8.

[13] M. K. Gupta and P. Chandra, A comprehensive survey of data mining, *International Journal of Information Technology*, Vol. 12, No. 4, 2020, pp. 1243-1257.

[14] I. L. H. Alsammak, A. H. Mohammed, and I. S. Nasir, E-learning and COVID-19: predicting student academic performance using data mining algorithms, *Webology*, Vol. 19, No. 1, 2022, pp. 3419-3432.

[15] H. Si, C. Sun, and H. Qiao, Application of improved multidimensional spatial data mining algorithm in agricultural informationization, *Journal of Intelligent and Fuzzy Systems*, Vol. 38, No. 2, 2020, pp. 1359-1369.

[16] S. Shakya, A self-monitoring and analyzing system for solar power station using IoT and data mining algorithms, *Journal of Soft Computing Paradigm*, Vol. 3, No. 2, 2021, pp. 96-109.

[17] W. Haoxiang and S. Smys, Big data analysis and perturbation using data mining algorithm, *Journal of Soft Computing Paradigm (JSCP)*, Vol. 3, No. 1, 2021, pp. 19-28.

[18] S. Choudhuri, S. Adeniye, and A. Sen, Distribution alignment using complement entropy objective and adaptive consensus-based label refinement for partial domain adaptation, *Artificial Intelligence and Applications*, Vol. 1, No. 1, 2023, pp. 43-51.

[19] M. A. Hanafiah and A. Wanto, Implementation of data mining algorithms for grouping poverty lines by District/City in North Sumatra, *IJISTECH (International*

Journal of Information System and Technology), Vol. 3, No. 2, 2020, pp. 315-322.

[20] P. Durana, V. Krastev, and K. Buckner, Digital twin modeling, multi-sensor fusion technology, and data mining algorithms in cloud and edge computing-based Smart city environments, *Geopolitics, History, and International Relations*, Vol. 14, No. 1, 2022, pp. 91-106.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)
The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself
No funding was received for conducting this study.

Conflict of Interest
The authors have no conflicts of interest to declare.