Accounting, Auditing and Finance (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/accaf.2025.060115
ISSN 2616-3136 Vol. 6 Num. 1

Information Construction of Enterprise Financial Management in Era of Big Data

Jincen Han[1,a,*]

[1]*Nanchong Vocational and Technical College, Nanchong, Sichuan, China*
[a]*595670113@qq.com*
[*]*Corresponding author*

Keywords: Corporate Financial Management; Data Mining; K-Means Clustering; Financial Informatization

Abstract: Under the background of informatization, enterprise financial management is more and more inclined to informatization construction, from the traditional mode of financial management to the direction of modernization and intelligence. Financial management informatization construction is a new form of financial accounting and control organization. It can effectively improve the efficiency of accounting and reduce costs, and also provide rich information for the financial management and operation management of enterprises, which is convenient for enterprises to make decisions. However, in the operation and management practice of corporate finance, there are often many problems that cannot be solved by conventional management methods. This study investigated the current area of research from the angles of economic digitisation and customer analytics, and it offered a foundation for the development of organization managerial accounting digitisation. Through the understanding of the current situation of traditional enterprise financial management, an intelligent management model based on big data was obtained, which can realize the optimization of the company's financial management. On this basis, a method of enterprise forecasting based on optimal initialization clustering algorithm was proposed to analyze the application prospect of enterprise financial management informatization. The traditional clustering algorithm and the improved clustering algorithm were used for 15 iterations of the sample data in the enterprise financial management. The results showed that the traditional algorithm was unstable in iterations, while the improved clustering algorithm was relatively stable. As can be observed, the modified grouping technique has a little validity of the measures value than the best choice.

1. Introduction

Big data technology is constantly being developed and applied, and as a result, new trends are starting to appear in the industries of healthcare, industrial, transport, and finance. The existing technology is finding it ever harder to suit the needs of businesses as the monetary sharing center is used by large and small-sized businesses more and more frequently. Currently, a key possibility for economics and even the financial industry is the application of contemporary digital technologies to

enhance the operation and oversight of global cities.

Big data technology advancement includes the preparation, saving, exhibition, and use of information along to the gathering and analysis of data. The use of big data relies heavily on greater data sources, and as the correctness of the investigation is directly influenced by the richness of the information, it is necessary to first verify the data's validity and integrity. Since datasets are gathered from numerous systems, issues including irregular format, missing data, and noise are frequently present. Data preparation is therefore crucial. Big data is used to apply the outcomes of machine learning to the important business world, hence providing decision making for the enterprise's pertinent operation.

The global data exchange unit and the growth of the insurance market have drawn more attention as the process of digitization has accelerated. Finding usable information buried in a lot of data is called data mining. In order to enhance the conventional clustering method, which can efficiently handle and analyze the financial data of businesses, this paper applied data mining technology.

2. Related Work

To achieve financial informatization, a growing number of major group organizations have begun paying attention and utilize big data. Saygili A T investigates panel processes, consumer rights, release of information, and transparency. His stakeholder-focused administration procedures and profitability as measured by accounting assessment of credit and non-financial enterprises are positively correlated [1]. Moreover, the objectivity and financial accounting knowledge of the audit committee are also represented. Mnif Y addresses the features of the two corporate governance procedures of the panel of directors for compliance [2]. The relationship between financial reporting quality, company must determine, and collapse risk is examined by Silva P in more detail [3]. Liu Y developed ideas for advancing the innovative production methodology coupled to intelligent machines from the perspectives of line up, firms, personalities, finance, etc. in response to various market demands [4]. These studies, however, do not take into account how financial information technology is built.

Stricter rules for financial reporting digital materials have been imposed by the globalized world. Historical accountants should be modernized in order to increase the quality of financial records, the processed and input speed of tax statement, and all three. The data mining-based organization budgeting platform was built and developed by Wei Y. He created the accounting methodology to satisfy the requirements of the organization, which can accomplish the majority of the business and audit face of the business and offer a specific foundation for the firm's budgeting decision [5].By particularly assessing the influence of economic information retrieval, solvency ratio, and movement research methodology in fraud diagnosis, Eko EU assesses the implementation of scientific financial reporting in protecting consumers in Nigerian commercial banks [6]. Kalakech K H employed a descriptive analytical approach and concluded that. Respondents differ in their perceptions of applying data mining concepts in Lebanese bank management operations due to differences in experience, work level and qualifications [7]. Bou-Hamad I demonstrated that data mining techniques outperformed real models under dynamic forecasting schemes for moderately to highly persistent time series through the static and dynamic forecasting capabilities of real data generation processes in financial time series [8]. Li M examined the information retrieval investigation of financial service providers' historical data, measuring the contribution of SMEs to supply chain credit, and outlining how big data analytics might assist SMEs in obtaining supply chain financing through developed online media [9]. These studies all have instructive recommendations, but there is a lack of research on financial data.

3. Corporate Financial Data Processing

3.1. K-Means Clustering Algorithm

In the absence of specific classification, unsupervised learning algorithms are generally used [10]. According to certain classification criteria, the method classifies multiple objects in the data set according to their characteristics, so that the data in the same class has the greatest similarity. This method utilizes the K-Means algorithm, also known as K-means clustering, which finds a representative center in the sample space. It is then divided into a center of mass that is closest to it, and then by iteratively repeating and correcting the center of mass until the specified termination condition is met.

Data mining is a method of solving problems through algorithms using large amounts of data. Guided learning specifically includes three strategies: classification strategy, estimation strategy and prediction strategy, as shown in Figure 1.

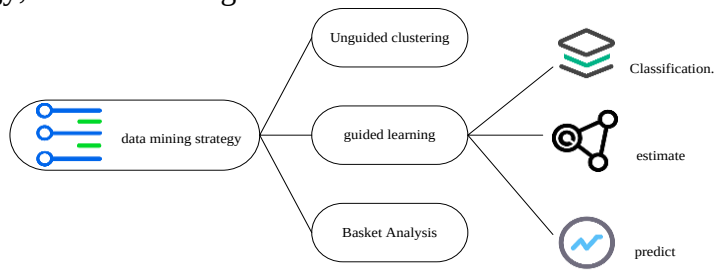

Figure 1 Data mining strategy

Guided learning utilizes existing patterns to distinguish structurally similar objects, including classification, estimation, and prediction [11]. Classification maps data to specific categories by building models and is applied to scenarios such as credit rating. Estimation is similar to classification, but the output is a numerical value, such as used to calculate the bad debt ratio of accounts receivable. Prediction focuses on estimating future events, such as sales forecasts for the next fiscal year [12-13]. In cluster analysis without a preset model, all parameters are independent variables and there are no dependent variables. The knowledge system needs to be reconstructed through the inherent structure of the data itself to assist guided learning [14]. Clustering identifies data features and divides them into multiple subsets for customer segmentation and precision marketing [15]. Shopping basket analysis is used to discover the correlation between data and determine highly correlated products, namely "shopping basket" products, by judging product sales data. Corporate financial analysis uses internal and external information to evaluate operations and financial conditions and support business decisions. Its research originated from the credit review needs of banks in the early 20th century [16]. The main body and emphasis of corporate financial analysis are shown in Figure 2 below:

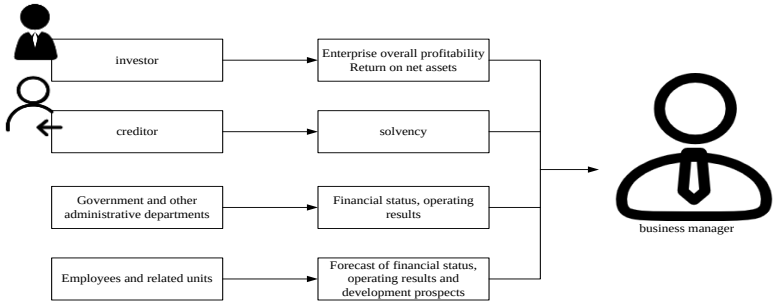

Figure 2 The structure of the main financial entity of the enterprise

Corporate financial analysis has three characteristics: first, it is an important means of measuring the business and financial status of a company, reflecting its debt repayment capacity, operational capacity and profitability, and assisting in decision-making and the formulation of reward and punishment mechanisms [17-18]; second, it helps to discover potential profit and performance growth points and promote the sustainable development of the company; third, it provides a basis for value assessment for external entities such as banks, governments and investors, enabling them to keep abreast of the company's financial status and development prospects [19].

In the K-means algorithm, the distance is used to measure the distance of the sample, so when calculating, it must be ensured that the weights of each variable on the distance should be the same. When multiple variables are calculated simultaneously, the larger the order of magnitude, the greater the impact on the result. To reduce the negative impact by orders of magnitude, the data should be normalized using the following formula:

$$\mathrm{a}_{\mathrm{x}} = \frac{\mathrm{a}_{\mathrm{x}} - a_{\min}}{a_{\max} - a_{\min}} \tag{1}$$

In the formula, a_{x} represents a random data item in the variable, $a_{\min}$ is the smallest one in the variable, and $a_{\max}$ is the largest one. The processed data set is defined as data set A. Each piece of data contains multiple attribute dimensions, and each data point in A represents a corporate financial sample. According to the performance of the company's financial shared center, the performance level is divided into four levels: excellent, good, qualified and unqualified, and the initial cluster centers of the four categories are initialized. According to the principle of minimum distance, the sampling points are corresponding to their closest mass centers, thus forming 4 groups. On this basis, the Euclidean distance formula is adopted, and its calculation formula is as follows:

$$\mathrm{d}_{euc}\left(a_x, a_y\right) = \sqrt{\sum_{k=1}^{n}\left(a_{xk} - a_{yx}\right)^2} \tag{2}$$

After classifying the four groups, the centroid of each group is updated. In the k-means algorithm, the centroid calculation formula is the average vector of all samples of such groups, and the centroid formula of the xth group z_{x} is:

$$\mathrm{z}_{\mathrm{x}} = \frac{1}{\mathrm{m}_x}\sum_{a \in \mathrm{z}_{\mathrm{x}}} a \tag{3}$$

When calculating the average distance d_{xzx} from the sample x to other samples z_{x} of the same type, the shorter the distance, the more the sample x should be allocated to z_{x}, which is defined as:

$$\mathrm{i}(x) = \mathrm{d}_{xzx} \tag{4}$$

Secondly, while defining $\mathrm{j}(x) = \min\left\{D_{xzy}, y \neq x\right\}$, the average distance d_{xzy} is calculated for all samples from sample i to outlier z_{y}. The longer the distance is, the less the sample x can be assigned to other clusters.

Then the silhouette coefficient of the sample x can be defined as:

$$\mathrm{s}(x) = \frac{j(x) - i(x)}{\max\left\{j(x), i(x)\right\}} = \frac{\min\left\{D_{xzy}, y \neq x\right\} - D_{xzx}}{\max\left\{\min\left\{D_{xzx}, y \neq x\right\}, D_{xzx}\right\}} \tag{5}$$

Finally, the average value of the contour coefficient s(x) of each sampling point is taken as the equivalent of the overall clustering result, and its expression is as follows:

$$S = \frac{1}{m}\sum_{x=1}^{m} s(x) \tag{6}$$

3.2. K-Means Clustering Method Based on Optimal Initialization Method

Based on the traditional K-means clustering algorithm, an improved K-means algorithm was further studied. Its advantage is that it avoids the random selection problem of the initial center point, considers the sample density and correlation during the initialization process, reduces the number of iterations, improves the calculation efficiency, and is more suitable for large-scale data processing [20]. The core function of management accounting is to provide strong support for corporate decision-making. The traditional financial analysis model relies on historical data and the subjective judgment of financial personnel, which is difficult to meet the needs in a complex environment. After the introduction of data mining technology, it can actively discover correlations in massive data, build accurate mathematical models, and achieve more efficient decision-making responses. Figure 3 shows the financial data measurement support process under data mining.

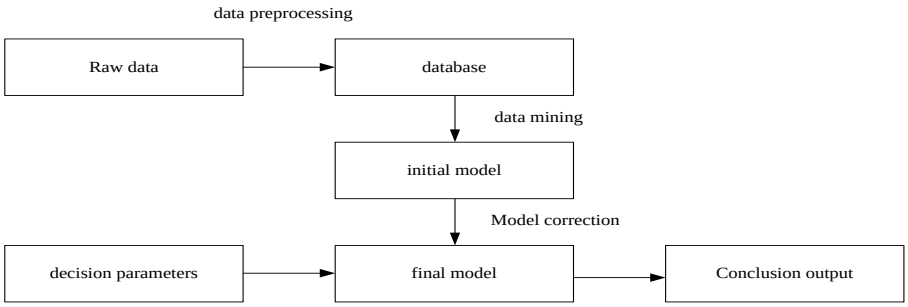

Figure 3 The flow chart of financial data measurement support under data mining

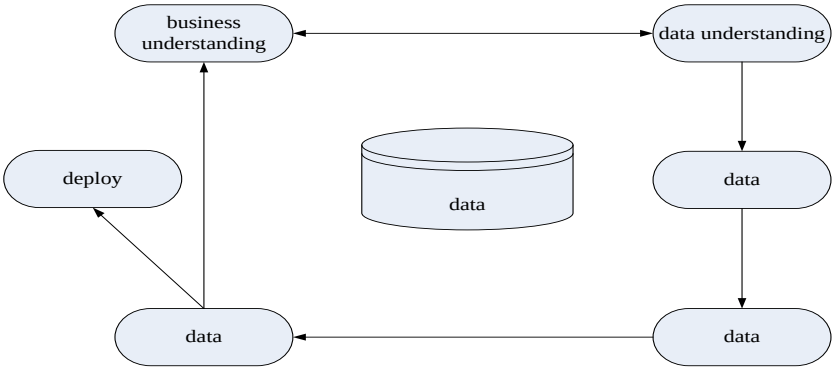

Figure 4 General process of data mining

In the accounting field, the cash flow method is often used to determine the fair value of assets, while the discount rate in traditional valuation methods mainly relies on the subjective judgment of financial personnel. The data mining system automatically extracts and corrects relevant data such as production, sales, and finance of enterprises through technical means. When decision makers have fair value requirements, they only need to enter the correction parameters to obtain more accurate calculation processes and results. In enterprise production, cost accounting needs to clarify the collection and allocation of costs of each product, which is related to budget preparation and is challenging to achieve accuracy. In the era of big data, data mining technology can efficiently

obtain cost-related data, establish multiple data models and connections, and use objective analysis methods to achieve accurate cost calculation and classification control. The success of data mining does not lie in complex operations and modeling itself, but in converting mining results into useful knowledge and guiding actual behavior. Its process includes business understanding, data interpretation, data analysis, modeling, evaluation and deployment. Figure 4 shows its general process.

The data mining technology is used in the basic principle and overall system structure of financial analysis, as shown in Figure 5. First, it is possible to pinpoint the issue that needs to be addressed and to decide what external and internal data are necessary from the problem's perspective. The necessary data is extracted by the machine learning and data financial analysis platform, and after preparation such cleanup and conversions, it is entered into the data warehouse to get ready for the subsequent step of the analysis and mining. It is necessary to examine and then transform the collected data into the knowledge needed by the user. At this stage, the financial analysis proposed in this paper can determine the analysis ideas and concepts of data mining topics, understand data, select models, analyze and evaluate results, and improve the models and analysis results. The stage three is to publish the data, and there are forms of output: both active and passive. According to predetermined user demands, active signal is connected to actively getting results, suggestions, and alarms to users. Once the framework replies to an user query, it transmits the result or notification to the person generated on the administrator's and material user's requests. This is commonly referred to as passive presentation.

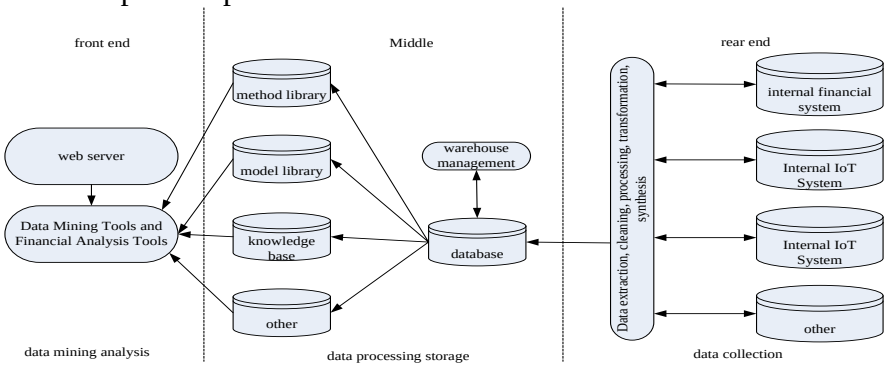

Figure 5 Workflow diagram of data mining financial analysis system

The data warehouse converts a large amount of data from the enterprise level into an overall data view, thus forming a complete data mining system to support decision-making and quantitative analysis requirements. Through the random selection of the original point, the local extreme value problem is effectively solved, and the number of clustering and iteration is reduced, thereby greatly improving the efficiency of the algorithm.

In the improved K-means initialization method, the input multidimensional financial data is first normalized to unify its value range to the interval [0, 1]. The normalized data is combined according to the attribute dimension to form an $m \times 3$ data matrix. Then all data points are sorted according to the distance matrix, and the Euclidean distance between every two data points is calculated to form a distance vector. The distance vector is rearranged in ascending order, and several points with the largest distance are determined as the basis for division, so as to divide the overall data into several initial subsets.

The distance between each data point and the current existing cluster center is calculated and assigned to the nearest category. For each data point in each category, the new cluster center of the category is determined by taking the mean of all its dimensional values. Repeat the above steps until the cluster center no longer changes significantly, and finally obtain a stable clustering result.

This method avoids the uncertainty caused by random initialization in the initial stage and significantly improves the clustering efficiency and stability.

4. Effect of Enterprise Financial Informatization Construction under Big Data

4.1. Pre-experiment Work Preparation

The financial status of a corporate finance company reflects its ability to allocate financial resources and the efficiency of serving the corporate group. This paper analyzes profitability, economic growth, and debt repayment capacity, extracts eight typical indicators as clustering variables, and uses them to evaluate the operating performance of the corporate finance shared center. The sample data are grouped under different clustering methods, and the division effects of the three types of indicators are compared. The one-way ANOVA is used to test whether the differences between the groups are significant, in order to verify the ability of traditional clustering and optimal initialization clustering to distinguish. Table 1 lists the one-way ANOVA used in the different clustering methods. An alternative assumption is that the group averages are not equal or complete. This paper chooses the credibility of 0.05, that is, when the significance is lower than 0.05, the original hypothesis would be rejected, otherwise, the original hypothesis would be selected.

Table 1 Differences between different clustering methods among factors

Clustering Method	Clustered Bianl	Sum Of Square	Mean Square
Traditional Clustering Methods	D1	97.046	32.108
	D2	64.232	21.17
	D3	70.372	23.216
Optimal Initialization Clustering Method	D1	93.261	30.846
	D2	87.171	88.388
	D3	4.98	4.058

4.2. Clustering Results of Enterprise Financial Management Center

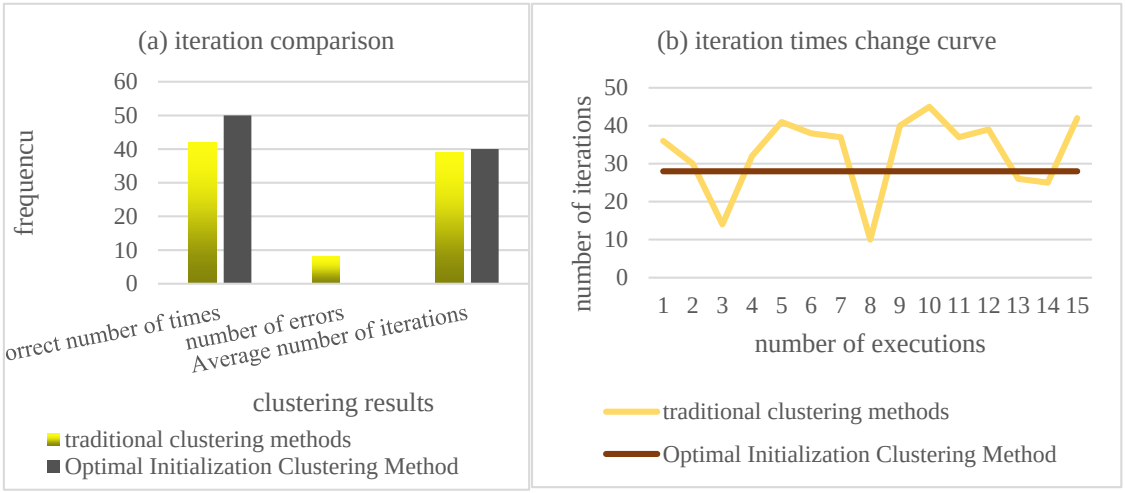

Figure 6 Comparison of algorithm iteration times and correct times

In the traditional clustering method, since the initial clustering center is randomly selected, in the iterative process, the traditional clustering method has the same effect on the clustering center. Therefore, in different situations, the similarity relationship matrix is used. In the traditional

clustering method, different clustering is performed on different local points, thereby improving the stability and computational complexity of the algorithm. The test results are shown in Figure 6.

The optimal initialization clustering algorithm is not only stronger than the traditional clustering algorithm in classification accuracy, but also has a significant improvement in classification stability and real-time performance. The traditional clustering algorithm and the optimal initialization clustering algorithm are used to repeatedly run 50 clustering operations on the sample data selected in the enterprise financial management, and the correct number of clustering results and the number of iterations are counted as shown in Figure 6(a).

The variation curve of the classic similarity measure and the fuzzy system with optimal initialization are shown in Figure 6(b) for the first 15 iterations. The graph shows that the algorithm iterations is extremely unstable when the traditional clustering algorithm is running, and it changes almost every time. With a finite iteration coefficient, the clustering fails when the local minimum points equal 26 and 25. Even when clustering is successful, the number of iterations varies over a large span. In the optimal initialization clustering algorithm, due to the similarity relationship preprocessing, the initial cluster center is determined, and the cluster center influence weight is assigned to each sample. Therefore, the success rate and the number of iterations of the algorithm are extremely stable, and the number of iterations is nearly a quarter less than the average of the traditional clustering algorithm. The convergence value of the optimal initialization clustering algorithm is slightly higher than that of the optimal solution due to the consideration of the weight of the sample on the cluster center.

4.3. Difference between Traditional Clustering Algorithm and Optimal Initialization Clustering Algorithm

As shown in Figure 7, in the traditional K-means clustering algorithm, many iterations are needed most, while the improved K-means clustering algorithm has fewer iterations.

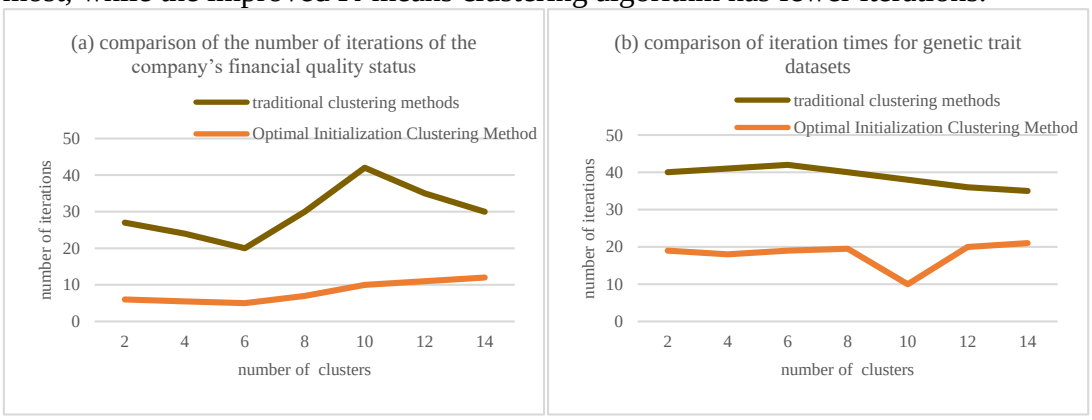

Figure 7 Comparison of the number of iterations of the dataset

Figure 7(a) shows the number of iterations based on the clustering analysis of the financial status of enterprises, and it is found that the improved clustering algorithm is superior to the traditional K-means clustering algorithm in terms of the number of iterations and calculation speed. The results show that there is no obvious difference between the K-means clustering algorithm based on optimal initialization and the traditional clustering algorithm when the number of cluster types is small. When the groups are 6, 8, and 10, the improved clustering effect is better than the traditional clustering algorithm, and the K-means clustering algorithm using the optimal initialization algorithm has better iteration amount and operation speed.

The method applies a small sampling of information compared to the classic approach, therefore

there are no appreciable differences in the analysing and understanding or iterations between the two. The genetic character data is contrasted with the traditional clustering approach in order to fully exploit the benefits of the new clustering. The results are displayed in Figure 7(b).

The improved clustering algorithm has significant differences in the overall average and each factor dimension. The F value obtained by the analysis of difference can be reflected in statistics. Therefore, whether it is traditional K-means clustering or K-means clustering based on optimal initialization, the results can reflect the differences between different categories. The clustering result analysis is shown in Figure 8.

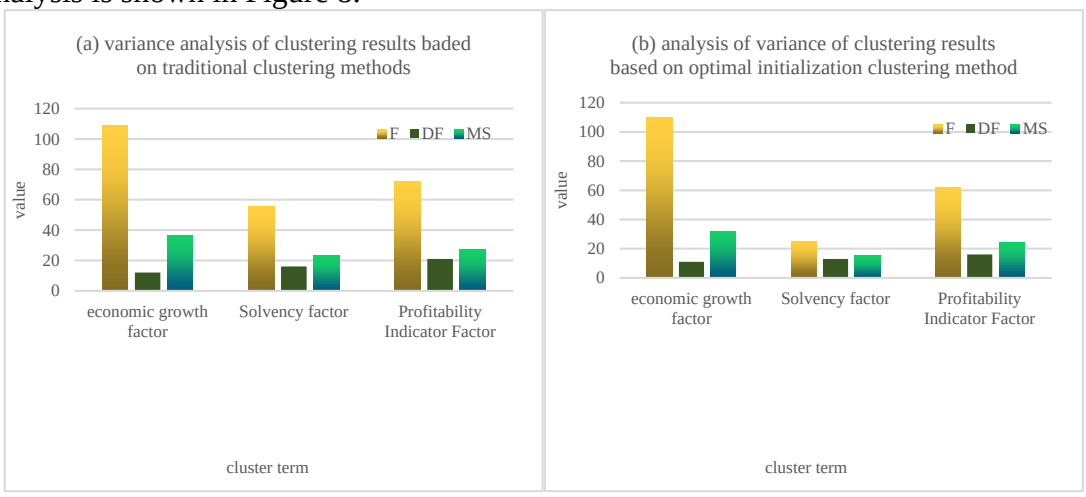

Figure 8 ANOVA of clustering results

One-way analysis of variance was performed on the traditional K-means clustering results. The results of the analysis are shown in Figure 8(a) below, where F is the statistic, DF is the degree of freedom, and MS is the mean square. According to the differences and degrees of freedom in various aspects, the corresponding F value is obtained. The study found that the economic growth factor, solvency factor and income factor showed obvious differences in K-means clustering.

5. Discussions

Existing methods use high-dimensional longitudinal data clustering based on models and regularization techniques, taking into account clustering and variable selection, improving interpretability and analysis efficiency. The improved K-means algorithm reduces processing load and improves clustering quality by optimizing distance calculation and cluster point stability. Empirical tests show that it has good results in practical applications. However, the performance of different algorithms under specific conditions is not universal, and the effect is also limited when facing complex or noisy data. Future research should consider multi-domain applicability and robustness, and explore hybrid methods or integrated algorithms to enhance performance and meet the challenges of real data analysis.

6. Conclusions

Due to proposing a fresh K-means clustering method based on the standard K-means classification method, the cultural Landscape algorithm called is first examined. Then, according to the basic idea of the algorithm and the specific calculation steps, new clustering algorithms are introduced respectively. Finally, the differences in the number of iterations and clustering effects between the improved algorithm and the traditional K-means algorithm are compared. The research shows that the two improved clustering methods can reduce the number of clusters and iterations,

and can effectively improve the clustering effect. There is a significant difference between the traditional K-means clustering and the improved clustering method.

References

[1] Saygili A T,.Saygili E,.Taran A. The effects of corporate governance practices on firm-level financial performance: evidence from Borsa Istanbul Xkury companies. Journal of Business Economics and Management, 2021, 22(4):884-904.

[2] Mnif Y,.Znazen O. Corporate governance and compliance with IFRS 7. Managerial Auditing Journal, 2020, 35(3):448-474.

[3] Silva P. Corporate governance, earnings quality and idiosyncratic crash risk during the 2007-2008 financial crisis. Journal of Multinational Financial Management, 2019, 51(Sep.):61-79.

[4] Liu Y. Research on Development of Mass Customization Service Mode Integrated with Artificial Intelligent Technology. Strategic Study of Chinese Academy of Engineering, 2018, 20(4):118-121.

[5] Wei Y. Research on design and realization of enterprise accounting management platform based on data mining. International Journal for Engineering Modelling, 2018, 31(1):265-272.

[6] Eko E U,.Adebisi A W,.Moses E J. Evaluation of Forensic Accounting Techniques in Fraud Prevention/Detection in the Banking Sector in Nigeria. International Journal of Finance and Accounting, 2020, 9(3):56-66.

[7] Kalakech K H,.El-Mousawi H Y. Impact of Application of Data Mining in the Lebanese Banking Sector. International Journal of Economics and Finance, 2019, 11(8):100-101.

[8] Bou-Hamad I,.Jamali I. Forecasting financial time-series using data mining models: A simulation study. Research in International Business and Finance, 2020, 51(Jan.):101072.1-101072.14.

[9] Li M,.Song H,.Yu K. Big data analytics in digital platforms: how do financial service providers customise supply chain finance?. International Journal of Operations & Production Management, 2021, 41(4):410-435.

[10] Yue H,.Liao H,.Li D, Chen L. Enterprise Financial Risk Management Using Information Fusion Technology and Big Data Mining. Wireless Communications and Mobile Computing, 2021, 2021(1):1-13.

[11] Kazanidis I,.Valsamidis S,.Gounopoulos E, Kontogiannis S. Proposed S-Algo+data mining algorithm for web platforms course content and usage evaluation. Soft Computing, 2020, 24(19):14861-14883.

[12] Kwon O,.Rahmatian S,.Iriberri A, Wu Z. Comparison of Neural Networks and Ordinary Least Squares Models in Forecasting Chines Stock Prices. Journal of International Finance and Economics, 2019, 19(1):17-32.

[13] Wu J,.Shi L,.Lin W P, Tsai SB, Xu G. An Empirical Study on Customer Segmentation by Purchase Behaviors Using a RFM Model and K-Means Algorithm. Mathematical Problems in Engineering, 2020, 2020(6):1-7.

[14] Jang B J,.Kim L H,.Kim K Y. The Effects of the Creating of Shared Value of Hospitality Industry Companies on Corporate Trust and Non-financial Management Performance: Focused on Travel Agencies. Journal of Tourism and Leisure Research, 2021, 33(1):197-216.

[15] Abdulrazzaq Al Chahadah, Ghaleb Awad, El Refae, Amer Qasim. The use of data mining techniques in accounting and finance as a corporate strategic tool: an empirical investigation on banks operating in emerging economies. International journal of economics and business research, 2018, 15(4):442-452.

[16] Xiao Y. Study on Talent Introduction Strategies in Zhejiang University of Finance and Economics Based on Data Mining. International Journal of Statistical Distributions and Applications, 2018, 4(1):20-22.

[17] Lu S, Caishigang. Study on papermaking enterprise financial accounting data analysis model based on decision tree and optimized SVM algorithm. Paper Asia, 2018, 34(6):57-62.

[18] Baral S K,.Parida J K. Corporate Governance and Its Impact on Financial Performance: A Systematic Survey. Asian Journal of Management, 2018, 9(4):1237-1242.

[19] Ondiba P K,.Cheruiyot T,.Sulo T. Effect of Cross Functional Job Rotation on Corporate Financial Performance of Commercial Banks in Kenya. Journal of Financial Risk Management, 2021, 10(2):187-199.

[20] Ali S A,.Yassin M,.Aburaya R. The Impact of Firm Characteristics on Corporate Financial Performance in Emerging Markets: Evidence From Egypt. International Journal of Customer Relationship Marketing and Management, 2020, 11(4):70-89.