Journal of Systems Engineering and Information Technology (JOSEIT)
Volume 4, No. 1, April (2025), Pages 7-13
e-ISSN: 2829-310X | 10.29207/joseit.

Comparison and Optimization of Parallel Clustering Algorithms for Chinese A-Share Stock Segmentation Based on Financial Indicators

Hai Mo[1], Niu Yihan[2], Zhang Yuejin[3]
[1]School of Information, Central University of Finance and Economics, Beijing 100081, China
[23]Kunming Branch, Shanghai Pudong Development Bank, Kunming 650000, China

[*]*hai.mo09@gmail.com*

Abstract. This study presents a novel application of parallel clustering algorithms to segment stocks in the Chinese A-share market based on financial indicators. Using the Hadoop platform and Mahout software library, we implemented and compared the performance of the K-means and fuzzy K-means algorithms across five distance measures: Euclidean, squared Euclidean, Manhattan, cosine, and Tanimoto. The analysis utilized 15 financial indicators from 2,544 listed companies to reflect profitability, solvency, growth capability, asset management quality, and shareholder profitability. The experimental results demonstrate that for stock financial data clustering, the K-means algorithm with Tanimoto distance yields optimal execution efficiency and clustering quality, whereas the fuzzy K-means algorithm performs best with squared Euclidean distance. However, the K-means algorithm proved to be more effective overall, successfully categorizing 1,483 stocks into 26 meaningful segments compared to only 511 stocks in 27 segments using fuzzy K-means. The resulting stock segmentation framework divides the market into eight comprehensive categories based on investment value and security, thereby providing investors with practical guidance for stock selection. Our approach enables investors to understand the fundamental characteristics of each stock segment, discern their distinctive features, and identify undervalued stocks with appreciative potential. This study represents the first application of parallel big data clustering algorithms to segment the entire Chinese A-share market, offering significant practical value for investment decision-making.

Keywords: *stock market segmentation; financial indicators; K-means clustering; big data analytics; parallel algorithms*

How to cites:
Hai Mo, Niu Yihan, & Zhang Yuejin. (2025). Comparison and Optimization of Parallel Clustering Algorithms for Chinese A-Share Stock Segmentation Based on Financial Indicators. Journal of Systems Engineering and Information Technology (JOSEIT), 4(1). https://doi.org/10.29207/joseit.v4i1.6535

Received by the Editor: April 12, 2025
Final Revision: April 24, 2025
Published: 2025-04-30

SDGs contributed:

1. Introduction

The stock market, a critical nexus linking listed companies and investors, plays a pivotal role in the financial system [1], [2]. In the context of China's evolving stock market, the selection of investment-worthy stocks is a pivotal concern for investors, as it directly impacts their financial interests. However, it is imperative to acknowledge the numerous factors that exert influence on the stock market, including, but not limited to, political developments, economic policies, and other external factors [3]. The operating performance of listed companies offers a certain degree of insight into the investment value of stocks.

Financial indicators serve as relative indicators for enterprises to summarize and evaluate their financial status and operating results [4], [5]. Consequently, financial indicators that can reflect the operating performance of listed companies are selected and stocks are divided into reasonable segments according to these financial indicators [6]. This approach enables investors to accurately comprehend and discern the prevailing characteristics of stocks, ascertain the scope of investments, and predict stock price trends by examining the aggregate price level of each category. Consequently, investors can select the opportune investment timing.

The clustering technique divides stocks in the market according to specific characteristics and obtains a guiding stock segment classification [7], [8]. This facilitates investors' selection of stocks from an appropriate classification for investment according to their needs [9]. The guiding effect of the clustering results for investors primarily encompasses the following: First, it enables investors to comprehend the fundamental characteristics and overall status of each stock segment, facilitating the initial segmentation of sections into those with exceptional and average performance. Second, it facilitates the discernment of the characteristics of each segment, such as

profitability and growth, based on selected financial indicators, thereby assisting investors in evaluating the investment value of stocks. To ascertain the equilibrium price of a given segment and identify stocks that are lower than this price due to market factors, and regard them as stocks with the best value, ascertain the equilibrium price of a given segment, and identify stocks that are lower than this price. The equilibrium price of a given sector is obtained, and stocks below this price owing to market factors are identified and regarded as stocks with room for appreciation and relatively low investment risk.

This paper proposes a novel utilization of K-means and fuzzy K-means algorithms [10], [11], [12], [13], which are drawn from the Mahout software library, a parallel clustering algorithm designed for the management of voluminous datasets on the Hadoop platform [14]. The financial indicators of nearly 2,600 publicly traded companies were the basis of these analyses. Through a series of experiments, we sought to compare the efficiency and quality of the K-means and fuzzy K-means algorithms in the context of clustering financial indicators within the Hadoop environment [15], [16], [17], [18].

Our objective is to identify a parallelized clustering method suitable for large-scale financial indicators. The ultimate goal of our study was to develop stock segmentation based on financial indicators. This segmentation was designed to empower investors by providing them with a comprehensive understanding of the overall characteristics of stocks. This, in turn, enables investors to accurately identify top-performing stocks and potential stocks within each segment and across the entire segment. Consequently, investors can make the optimal investment decisions. This facilitates investors' optimal investment decision-making process.

2. Related work

Academic literature pertaining to the application of cluster analysis in the domain of stock market sector analysis is generally subdivided into two broad categories.

2.1 Establishment of clustering indicator system

Research has introduced the cluster analysis method into the analysis of securities investment by examining the basic dimensions of stocks, such as industry, company, profitability, and growth factors, and establishing a more comprehensive evaluation index system to measure the degree of similarity of the sample stocks. They then determined the scope and value of investment through a cluster analysis model. Empirical evidence has demonstrated the efficacy and practicality of this method in guiding securities investment decisions [19]. Other studies have developed a stock investment value evaluation index set containing 15 indicators in five aspects [20]. This index set was obtained using fuzzy clustering and indicator screening and was applied to all listed companies. This study provides empirical evidence for the application of data mining technology to stock value investment.

The selection of these 15 financial indicators was deliberate and comprehensive, drawing on both established financial theory and empirical research in the Chinese market context. These indicators were selected across five critical dimensions of corporate financial health: profitability, solvency, growth capability, asset management quality, and shareholder profitability. This multidimensional approach ensures holistic evaluation of the fundamental characteristics of each stock. Profitability indicators were included, as they directly reflect a company's ability to generate earnings relative to its expenses, which fundamentally drives stock valuation.

Solvency metrics were selected to assess companies' capacity to meet long-term obligations, which are particularly important in China's highly leveraged corporate environment. Growth capability indicators provide forward-looking measures that signal potential future performance, a crucial consideration for investment decisions in China's rapidly evolving markets. Asset management quality metrics are incorporated to evaluate operational efficiency, revealing how effectively companies utilize their resources to generate revenue. Finally, shareholder profitability indicators were selected, as they directly align with investor interests, measuring how effectively companies translate business success into returns for equity holders.

2.2 Selection and implementation of clustering methods

Li et al employed a data clustering approach to analyze the financial performance of 31 high-tech companies listed on the stock market [21]. They selected five indicators that reflect the comprehensive profitability of these companies and used a clustering process in SAS software, known as "Cluster," to identify four classes that align with their actual financial status and operating conditions. Yingrui et al. employed three clustering algorithms, K-means, Kohonen, and TwoStep, in Clementine to analyze the clustering of over 800 stocks in China's A-share market [22]. This study utilized 13 financial indicators that reflect the five aspects of listed companies as a clustering index system. The results indicated that the two-step clustering method yielded superior analytical results in the stock clustering analysis. The superiority of the two-step clustering method was further substantiated by its consistently superior performance in analyzing the stock market.

In summary, most domestic studies on the application of clustering algorithms in stock sector analysis utilize a limited number of financial indicators, typically less than ten, as the clustering index system. These studies also

employ a relatively small sample size, often less than 50 stocks, for clustering analysis. In contrast, Reference employed an optimized ant colony clustering algorithm to cluster more than 1,800 stocks. However, they did not utilize a parallel clustering algorithm for big data in their analysis. In this study, we first apply the parallel clustering algorithm for big data to stock segmentation of the entire Chinese A-share market. We then implemented the clustering algorithm in the Mahout algorithm library by using the Hadoop platform. Finally, we perform parallelized clustering on all stocks in the current A-share market using more than 10-dimensional financial indices to obtain the segmentation of the entire A-share market. This is of great practical significance for investors when making reasonable investment decisions. The findings of this study are of significant practical value for investors seeking to make informed investment decisions.

3. Results and Discussion

The experimental data presented in this paper are derived from the June 2014 financial data of all listed companies obtained from the financial indicator analysis database of Chinese listed companies in the Cathay Pacific database. The dataset comprises 2,544 samples, with each containing 15 financial indicators, thereby yielding 15 dimensions. A clustering experiment was conducted to standardize the values of each indicator. The experimental environment was a pseudo-distributed Hadoop platform built on an AliCloud server, including a master node and a slave node.

Table 1. Parameter settings of clustering algorithm

Arithmetic	Number of Clusters	Number of Iterations	Convergence Threshold	Fuzzy Parameter
K-means	20	50	0.5	/
Fuzzy K-means	20	50	0.5	2

Two clustering algorithms, K-means and fuzzy K-means, were utilized to cluster the financial indicators under five different distance measures: Euclidean distance, squared Euclidean distance, Manhattan distance, cosine distance, and Tanigamoto distance. Ultimately, the intercluster density and intracluster density [9(] were computed for each experiment and employed as an evaluation metric for the quality of the clusters). The results of these experiments are presented in Tables 2, 3, and 4.A comprehensive summary of the experimental results is presented in Tables 2 and 3, respectively.

Table 2. Clustering results of K -means with different choices of distance algorithms

Distance Metrics	Number of Iterations	Convergence Time/Min	Intercluster Density	(Math.) Intracluster Density
Euclidean distance	20	6.57	0.093	0.553
Square the Euclidean distance	15	4.951	0.091	0.554
Manhattan Distance	17.3	5.631	0.099	0.544
Cosine Distance	1.7	0.758	0.077	0.555
Tanimoto Distance	4.3	1.602	0.045	0.557

As demonstrated in Tables 2 and 3, for this financial indicator dataset, the fuzzy K-means algorithm requires fewer iterations and exhibits shorter execution times than the K-means algorithm for all distance measure options. This finding suggests that the fuzzy k-means algorithm converges more rapidly and exhibits higher execution efficiency. A comparison of the five distance algorithms revealed that both the fuzzy K-means and K-means algorithms converged after one to two iterations with high execution efficiency when the cosine distance was selected as the distance metric.

However, it should be noted that this distance algorithm does not consider the length of the two vectors and focuses exclusively on the direction from the origin to the two points. An analysis of the clustering outputs of the two algorithms that select the cosine distance as the distance parameter reveals that the distribution of the number of stocks in the 50 classifications is extremely disparate, with some classes containing hundreds of stocks and others containing only a few stocks. The findings of this study indicate that cosine distance is not a viable distance parameter for clustering the financial indicators of stocks. This is because these results offer no practical guidance.

Table 3. Clustering results of fuzzy K -means with different choices of distance algorithms

Distance Metrics	Number of Iterations	Convergence Time/Min	Intercluster Density	(Math.) Intracluster Density
Euclidean distance	3.7	1.484	0.12	0.549
square the Euclidean distance	3.7	1.477	0.05	0.581
Manhattan Distance	5	1.929	0.124	0.544
cosine distance	1	0.585	0.115	0.573
Tanimoto Distance	7.7	2.85	0.063	0.489

After excluding the cosine distance algorithm, the clustering results of the K-means algorithm with different distance measures were compared. The K-means algorithm was found to have the lowest number of iterations and shortest execution time, as well as the smallest density between clusters and the largest density within clusters when choosing the Tanimoto distance as a parameter of the distance algorithm. This finding indicates that the K-means algorithm has the highest execution efficiency and the best clustering quality when choosing the Tanimoto distance for clustering calculations for the dataset of financial indexes.

The clustering calculation indicated that the K-means algorithm exhibited the highest execution efficiency and optimal clustering quality. A subsequent analysis of the clustering output of this experiment reveals that there is an average number of stocks in the 50 classifications, which is instructive for practical application and can be used as the result of the stock segmentation of financial indicators. A comparison of the clustering results of the fuzzy K-means algorithm, where the squared Euclidean distance was employed as the distance metric, revealed the highest execution efficiency and optimal clustering quality. Although the fuzzy K-means algorithm does not result in the same degree of hard clustering as the K-means algorithm, it does lead to the formation of overlapping clusters. This results in a more uneven cluster division compared to the K-means algorithm.

In summary, for this stock financial data clustering experiment, the K-means algorithm should select the Tanimoto distance algorithm as the distance parameter, whereas the fuzzy K-means algorithm should select the squared Euclidean distance algorithm as the distance parameter. The aforementioned analysis indicates that the Tanimoto distance is selected as the distance parameter for the K-means algorithm, whereas the fuzzy K-means algorithm uses the squared Euclidean distance as the distance parameter to cluster the stock financial indicators.

The specific process comprises the following steps: the Clusterdump class in Mahout is utilized for reading the clustering results, and the output is saved as a text file. The next step involves identifying the ticker symbols corresponding to each piece of data based on the financial data. Finally, the number of stocks with the same cluster number, that is, the number of stocks within each segment, should be counted. Given the practical significance of the clustering results, classifications with more than 30 stocks and less than 100 stocks in each class (26 classes in total) were filtered out from the clustering results of the K-means algorithm, and classifications with more than 10 stocks and less than 100 stocks in each class (27 classes in total) were filtered out from the clustering results of the fuzzy K-means algorithm. The standard deviation and average standard deviation of each financial indicator in each category are calculated to reflect the profitability, solvency, asset management quality, growth ability, and shareholder profitability of listed companies.

The calculation results are presented in Tables 4 and 5. In accordance with the principle of stock financial indicator segmentation, companies of the same type are expected to exhibit similar business conditions. That is to say, the standard deviation of the financial indicators of listed companies within the same category should be less than the overall standard deviation of 1. A smaller value indicates a higher degree of similarity in the financial data of the stocks within the same category. The calculation results of the average value and standard deviation of each financial indicator reveal that both the K-means algorithm and the fuzzy K-means algorithm have only one classification whose average standard deviation exceeds 1, whereas the average standard deviation of the other classifications is less than 1. This finding suggests that the results of the division of the two algorithms were more reasonable. However, a comparison of the division results obtained by the K-means algorithm with those of the fuzzy K-means algorithm revealed that only the average standard deviation of the indicators of the five classes exceeded that of the fuzzy K-means algorithm.

Table 4. Fuzzy K -means clustering results standard deviation of financial indicators

Form	Number Of Shares/Each	Profitability	Growth Capacity	Shareholder Profitability	Solvency	Quality Of Asset Management	Average Standard Deviation
1	20	0.226	0.053	0.053	0.063	0.191	0.117
2	13	0.113	0.111	0.111	0.038	0.235	0.122
3	12	0.168	0.047	0.047	0.179	0.150	0.118
4	17	0.256	0.386	0.386	0.311	0.374	0.343
5	11	0.168	0.104	0.104	0.068	0.262	0.141
6	16	0.218	0.115	0.115	0.206	0.253	0.182
7	16	0.161	0.228	0.228	0.751	0.267	0.327
8	10	0.139	1.026	1.026	0.083	0.257	0.506
9	20	0.186	0.146	0.146	0.480	0.139	0.219
10	14	0.180	0.055	0.055	0.032	0.237	0.112
11	17	0.431	0.123	0.123	0.081	0.250	0.201
12	20	10.133	7.617	7.617	10.451	7.635	8.690
13	33	0.153	0.433	0.433	0.141	0.253	0.283
14	29	0.126	0.095	0.095	0.404	0.295	0.203
15	12	0.173	0.048	0.048	0.388	0.236	0.179

16	34	0.209	0.215	0.215	0.266	0.793	0.340
17	16	0.162	0.053	0.053	0.534	0.132	0.187
18	19	0.210	0.053	0.053	0.147	0.151	0.123
19	22	0.175	0.694	0.694	0.421	0.356	0.468
20	38	0.296	1.631	1.631	0.117	0.244	0.784
21	21	0.455	0.055	0.055	0.046	0.243	0.171
22	12	0.130	0.238	0.238	0.055	0.285	0.189
23	17	0.781	0.058	0.058	0.456	0.379	0.346
24	22	0.188	0.078	0.078	0.044	0.235	0.125
25	11	0.108	0.068	0.068	0.116	0.323	0.137
26	16	0.400	0.275	0.275	0.157	0.359	0.293
27	23	0.237	0.117	0.117	0.320	0.335	0.225

Table 5. K-means clustering results standard deviation of financial indicators

Form	Number Of Shares/Each	Profitability	Growth Capacity	Shareholder Profitability	Solvency	Quality Of Asset Management	Average Standard Deviation
1	50	0.085	0.077	0.275	0.062	0.117	0.123
2	77	0.172	0.093	0.363	0.060	0.161	0.170
3	89	0.185	0.166	0.444	0.073	0.271	0.228
4	44	0.159	0.045	0.152	0.048	0.120	0.105
5	64	0.071	0.069	0.147	0.057	0.084	0.086
6	79	0.084	0.088	0.151	0.075	0.087	0.097
7	58	0.193	0.084	1.852	0.159	0.486	0.555
8	76	0.278	0.107	0.297	0.117	0.256	0.211
9	62	0.069	0.045	0.135	0.106	0.088	0.088
10	43	0.105	0.057	0.160	0.043	0.088	0.091
11	30	0.047	0.029	0.096	0.039	0.051	0.052
12	35	0.053	0.038	0.079	0.039	0.050	0.052
13	31	0.070	0.031	0.097	0.154	0.057	0.082
14	87	0.123	0.068	0.217	0.102	0.098	0.122
15	57	0.058	0.043	0.166	0.055	0.063	0.077
16	48	0.109	0.108	0.244	0.070	0.139	0.134
17	52	0.070	0.047	0.196	0.037	0.095	0.089
18	77	0.112	0.079	0.192	0.078	0.113	0.115
19	41	0.066	0.061	0.097	0.040	0.082	0.069
20	33	0.093	0.045	0.211	0.146	0.101	0.119
21	67	0.221	0.174	0.588	0.091	0.679	0.351
22	54	0.074	0.028	0.105	0.036	0.075	0.064
23	55	0.094	0.048	0.157	0.094	0.091	0.097
24	59	0.043	0.038	0.096	0.039	0.061	0.055
25	35	0.196	0.082	0.480	2.096	2.965	1.164
26	80	0.057	0.058	0.093	0.055	0.065	0.066

The total number of stocks included in the valid divisions obtained is 1,483, whereas the fuzzy k-means algorithm has only 511 stocks. This finding indicates that the K-means algorithm is more effective than the fuzzy k-means algorithm in clustering stock financial data. Consequently, this study employs the squared Euclidean distance as the clustering outcome of the K-means algorithm utilizing the distance metric method. This approach was further employed as the ultimate outcome of the categorization of Chinese listed companies' stock segments, as illustrated in Table 6.

Table 6. Clustering results of K-means algorithm for stock financial data

Form	Hallmark	Share Class
Category 1	High investment value, high investment security	(7,8,13)
Category 2	High investment value, general investment security	(3,4,9,11,16,21)
Category 3	Higher investment value, lower investment security	(1, 5, 6, 12, 22, 23)
Category 4	General investment value, high investment security	(25)
Category 5	General investment value, general investment safety	(26, 18, 20, 2)
Category 6	Average investment value, low investment security	(14, 15, 19)
Category 7	Has low investment value, average investment safety	(17)
Category 8	Has a lower investment value, lower investment safety	(10, 24)

4. Conclusions

This paper proposes an analysis of all companies listed on the Chinese stock market. The analysis is carried out using the K-means algorithm in Mahout and the Fuzzy K-means algorithm in Mahout, with different distance measures. Distance measurements reflect the profitability, solvency, growth, asset management quality, and shareholder profitability of all companies listed on the Chinese stock market. The results are succinctly summarized in five dimensions: profitability, solvency, growth, asset management quality, and shareholder profitability. Fifteen financial indicators were grouped and analyzed based on the number of iterations, execution time, and time between clusters from the two clustering algorithms with different distance algorithms. The resulting data were then compared and analyzed in terms of the number of iterations, execution time, intercluster density, and intracluster density using different distance algorithms. The findings of this comprehensive analysis, when combined with the actual clustering results, led to the identification of a suitable clustering algorithm for stock financial data. This study identifies the appropriate distance measure and clustering algorithm for clustering stock financial data, as well as a combination of clustering algorithms. The experimental results of this combination yielded a distance measurement method and a combination of clustering algorithms for the stocks.

References

[1] Y. Li, P. Ni, and V. Chang, "Application of deep reinforcement learning in stock trading strategies and stock forecasting," *Computing*, vol. 102, no. 6, pp. 1305–1322, Jun. 2020, doi: 10.1007/s00607-019-00773-w.

[2] G. Kumar, S. Jain, and U. P. Singh, "Stock Market Forecasting Using Computational Intelligence: A Survey," *Archives of Computational Methods in Engineering*, vol. 28, no. 3, pp. 1069–1101, May 2021, doi: 10.1007/s11831-020-09413-5.

[3] W. Zhu, M. Li, C. Wu, and S. Liu, "Comprehensive financial health assessment using Advanced machine learning techniques: Evidence based on private companies listed on ChiNext," *PLoS One*, vol. 19, no. 12, p. e0314966, Dec. 2024, doi: 10.1371/journal.pone.0314966.

[4] C. Zhao, X. Yuan, J. Long, L. Jin, and B. Guan, "Financial indicators analysis using machine learning: Evidence from Chinese stock market," *Financ Res Lett*, vol. 58, 2023, doi: 10.1016/j.frl.2023.104590.

[5] S. Iacuzzi, "An appraisal of financial indicators for local government: a structured literature review," *Journal of Public Budgeting, Accounting and Financial Management*, vol. 34, no. 6, 2021, doi: 10.1108/JPBAFM-04-2021-0064.

[6] P. Hájek, "Combining bag-of-words and sentiment features of annual reports to predict abnormal stock returns," *Neural Comput Appl*, vol. 29, no. 7, pp. 343–358, Apr. 2018, doi: 10.1007/s00521-017-3194-2.

[7] J. Irani, N. Pise, and M. Phatak, "Clustering Techniques and the Similarity Measures used in Clustering: A Survey," *Int J Comput Appl*, vol. 134, no. 7, 2016, doi: 10.5120/ijca2016907841.

[8] M. Steinbach, G. Karypis, and V. Kumar, "A Comparison of Document Clustering Techniques," *KDD workshop on text mining*, vol. 400, 2000, doi: 10.1109/ICCCYB.2008.4721382.

[9] P. Kumar and A. Kanavalli, "A Similarity based K-Means Clustering Technique for Categorical Data in Data Mining Application," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 2, 2021, doi: 10.22266/ijies2021.0430.05.

[10] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf Sci (N Y)*, vol. 622, 2023, doi: 10.1016/j.ins.2022.11.139.

[11] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," 2020. doi: 10.3390/electronics9081295.

[12] M. B. Ferraro, "Fuzzy k-Means: history and applications," *Econom Stat*, vol. 30, 2024, doi: 10.1016/j.ecosta.2021.11.008.

[13] F. Nie, X. Zhao, R. Wang, X. Li, and Z. Li, "Fuzzy K-Means Clustering with Discriminative Embedding," *IEEE Trans Knowl Data Eng*, vol. 34, no. 3, 2022, doi: 10.1109/TKDE.2020.2995748.

[14] R. Anil *et al.*, "Apache Mahout: Machine Learning on Distributed Data ow Systems," *Journal of Machine Learning Research*, vol. 21, 2020.

[15] X. T. Li and X. H. Duan, "A K-Means Clustering Algorithm for Early Warning of Financial Risks in Agricultural Industry," *Security and Communication Networks*, vol. 2022, 2022, doi: 10.1155/2022/3751539.

[16] F. Zhang, Y. Ding, and Y. Liao, "Financial Data Collection Based on Big Data Intelligent Processing," *International Journal of Information Technologies and Systems Approach*, vol. 16, no. 3, 2023, doi: 10.4018/IJITSA.320514.

[17] M. Mallikarjuna and R. P. Rao, "Application of data mining techniques to classify world stock markets," *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 1, 2020, doi: 10.30534/ijeter/2020/09812020.

[18] F. Baser, O. Koc, and A. S. Selcuk-Kestel, “Credit risk evaluation using clustering based fuzzy classification method,” *Expert Syst Appl*, vol. 223, 2023, doi: 10.1016/j.eswa.2023.119882.
[19] P. D’Urso, L. De Giovanni, R. Massari, R. L. D’Ecclesia, and E. A. Maharaj, “Cepstral-based clustering of financial time series,” *Expert Syst Appl*, vol. 161, p. 113705, Dec. 2020, doi: 10.1016/j.eswa.2020.113705.
[20] A. Q. Md *et al.*, “Novel optimization approach for stock price forecasting using multi-layered sequential LSTM,” *Appl Soft Comput*, vol. 134, p. 109830, Feb. 2023, doi: 10.1016/j.asoc.2022.109830.
[21] T. Li, G. Kou, Y. Peng, and P. S. Yu, “An Integrated Cluster Detection, Optimization, and Interpretation Approach for Financial Data,” *IEEE Trans Cybern*, vol. 52, no. 12, pp. 13848–13861, Dec. 2022, doi: 10.1109/TCYB.2021.3109066.
[22] Z. Yingrui, Z. Chen, and L. Zhongxin, “Dynamic Analysis and Community Recognition of Stock Price Based on a Complex Network Perspective,” *SSRN Electronic Journal*, 2022, doi: 10.2139/ssrn.4090744.