

UDC 004.89

DOI 10.52167/1609-1817-2024-132-3-206-215

R.M. Durbenov<sup>1</sup>, Zh. Lamasheva<sup>1</sup>, V. Lakhno<sup>2</sup>, M. Zhartybayeva<sup>1</sup>

<sup>1</sup>L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

<sup>2</sup>National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine

E-mail: programmankz@gmail.com

## CLUSTERING IN COMPILING INVESTMENT PORTFOLIOS

**Abstract.** Clustering is an effective tool for diversifying investments, reducing risk and identifying new opportunities. Clustering plays a key role in data analysis, allowing you to group objects with similar characteristics. This study examines the application of clustering to the tasks of forming and optimizing investment portfolios. The paper presents two clustering methods: K-medoids and fuzzy clustering (C-means). K-medoids divides assets into clusters by correlation, and C-means allows assets to belong to several clusters with varying degrees. The article analyzes various clustering methods in the context of the stock market. Similarity measures for stock clustering are compared. The K-means algorithm for clustering companies by time series is presented. Chaotic cartographic clustering is used to analyze the dynamics of the stock market. The TreeGNG algorithm is used to identify stock market sectors. The HRK method has been developed to predict short-term changes in stock prices. The application of data mining methods to predict stock market trends is discussed. The K-means and C-means methods for clustering banking and energy companies are compared. The effectiveness of the SOM-SVR hybrid approach for predicting price dynamics and volatility is demonstrated. C-means combines with artificial neural networks to improve the accuracy of stock market forecasting. The study demonstrates the potential of various clustering methods in the context of compiling and optimizing investment portfolios.

**Keywords.** Clustering, K-medoids, C-means, diversification, portfolio optimization, data analysis, stock market.

### Introduction.

Clustering, a key element of data analytics, occupies a central role in the modern world of big data processing. This study is dedicated to exploring and analyzing various clustering methods, their applications, and effectiveness in diverse areas ranging from social sciences to bioinformatics, and from marketing to geospatial analysis. Cluster analysis is the process of grouping a set of objects in such a way that the objects in the same cluster are more similar to each other than to those in other clusters. The effectiveness of clustering depends on many factors, including the characteristics of the data, the choice of algorithm, and parameters. Understanding these elements is crucial for the successful application of cluster analysis.

One of the main goals of clustering is to discover the internal structure in data, which can aid in identifying patterns and anomalies. The use of clustering ranges from simple market segmentation to complex structuring of genetic data. In this context, the current study presents a review of modern clustering methods, discussing their advantages, limitations, as well as potential areas of application.

An important area of application for cluster analysis is investment analysis, particularly in the context of compiling and optimizing investment portfolios. In the financial world, clustering offers innovative approaches to asset segmentation based on their behavior and characteristics, allowing investors to efficiently manage risks and identify new investment opportunities. The use of these methods for diversifying investments and reducing correlation between them helps

to lower the overall risk of the portfolio. Moreover, clustering helps to reveal hidden patterns in market behavior, which is critical in conditions of high volatility and uncertainty.

This study not only sheds light on how traditional clustering methods can be adapted for financial data analysis but also considers new approaches and technologies in this area. We aim to provide an understanding of how cluster analysis can be integrated into a broad investment strategy, offering deeper insights into market dynamics and growth potentials of various asset classes.

*The theoretical foundations of clustering.* K-medoids is a partial clustering technique similar to K-means but using actual data points as centers (medoids) instead of the mean of objects in a cluster, making it more robust against outliers. In asset allocation, K-medoids are used to divide a set of assets into clusters based on their similarity, usually measured in terms of correlation or covariance between asset returns. The algorithm iteratively assigns each asset to the cluster represented by the nearest medoid, then updates the medoid of each cluster to minimize a predefined cost, typically the sum of dissimilarities between medoids and assets in their clusters. Grouping correlated assets using K-medoids helps to reduce portfolio risk and aids in diversification by identifying distinct groups of assets, allowing for investment distribution across different market behaviors (Table 1).

Table 1 - Comparison of K-means and fuzzy clustering

№	Characteristic	K-means	Fuzzy clustering
1	The type of clustering	Hard	Soft
2	Cluster centers	Average values	Actual data points
3	Sensitivity to emissions	High	Low
4	Belonging to clusters	One cluster	Several clusters
5	Flexibility	Low	High
6	Details	Low	High

Unlike K-medoids, which assign each asset to one cluster, fuzzy clustering (C-means) allows assets to belong to several clusters with varying degrees of membership. This reflects the reality that assets can exhibit characteristics of multiple clusters. Each asset has a degree of membership for each cluster, and these degrees are used to distribute a portion of the asset across various clusters. This allows for the creation of a more nuanced and flexible portfolio. Fuzzy clustering iteratively updates cluster centers and membership degrees, based on minimizing a target function, often a weighted sum of squared distances between assets and cluster centers, weighted by membership degrees.

K-medoids is a clustering method that uses actual data points as cluster centers. This makes it more resistant to emissions compared to K-means.

Fuzzy clustering is a clustering technique that allows data points to belong to multiple clusters with varying degrees of affiliation. This reflects the reality that data points can have the characteristics of multiple clusters.

K-medoids and fuzzy clustering are methods that can be used to allocate assets in a portfolio.

K-medoids allows you to create simpler and less risky portfolios, and fuzzy clustering allows you to create more detailed and flexible portfolios.

### Materials and methods.

The article [1] examines the effectiveness of different similarity measures for clustering similar stocks. The study carried out using data from the Standard & Poor 500 index for the year 1998, it explores various clustering techniques and compares the results to a "ground-truth"

clustering based on the index. The findings revealed key insights about the suitability of different similarity measures for stock market data.

The article [2] presents a modified k-means clustering algorithm to cluster stock market companies based on time series similarity measures. This algorithm utilizes the Maximum Information Compression (MIC) index as a similarity measure and compares it with other measures like correlation coefficient and least-square regression error. The study demonstrates the algorithm's effectiveness in naturally partitioning data, as companies from the same industrial branch are often grouped together. Applied to the Dow Jones index companies, this algorithm aids in identifying similar temporal behavior of traded stock prices, which is useful for portfolio optimization strategies. The research showcases the modified k-means algorithm's suitability over the standard k-means in clustering stock market time series.

The study [3] investigates the application of chaotic map clustering (CMC) for analyzing stock market dynamics, focuses on the Dow Jones Industrial Average (DJIA) companies. The methodology involves associating a chaotic map to each company, with the correlation coefficients of financial time series representing the coupling strengths between these maps. This approach enables the identification of clusters of companies within the stock market index, which are often grouped together by industrial sectors. The clustering results can be utilized for portfolio optimization strategies, as companies within the same cluster demonstrate similar stock price behaviors over time.

The paper [4] presents an approach to understanding stock market sectors using a hierarchical topological clustering algorithm TreeGNG (Tree Growing Neural Gas). The study utilizes a dataset of share closing prices from the FTSE 100 index, spanning over a decade. The TreeGNG algorithm is applied to this dataset, resulting in the formation of clusters that correspond to different market sectors. The clusters identified by TreeGNG are compared to the globally accepted sector classification scheme, and the findings demonstrate that the TreeGNG method can effectively capture the sector structure of the stock market.

The article [5] introduces a novel HRK (Hierarchical agglomerative and Recursive K-means clustering) method aimed at predicting short-term stock price movements following the release of financial reports. The methodology encompasses three stages: converting financial reports into feature vectors, employing hierarchical agglomerative clustering to form initial clusters, and subsequently applying recursive K-means clustering for further refinement. This approach allows for the selection of representative feature vectors, which are then utilized for stock price movement prediction. The study asserts that this method outperforms traditional SVM (Support Vector Machine) techniques in terms of accuracy and average profits.

The study [6] highlights several data mining techniques such as decision trees, neural networks, clustering, association rules, and factor analysis. It discusses their applications in predicting stock market trends, analyzing market behavior, and supporting strategic investment decisions. The paper also underscores the challenges and opportunities in this field, detailing the potential of these techniques to uncover hidden patterns and predict future market behaviors.

The article [7] focuses on comparing two data mining clustering methods: K-means and Fuzzy C-means. The study uses a sample of banking and energy companies from the Gulf Cooperation Council (GCC) stock markets, examining their patterns in relation to the impact of news on stock sectors during October to December 2012. The research employs correlation coefficients, t-statistics for good and bad news indicators, and financial factors like PER, PBV, DY, and rate of return as variables for clustering. The main findings indicate that K-means and Fuzzy C-means clustering methods provide different insights when applied to the same dataset. The study concludes that the Fuzzy C-means method is more effective in identifying homogenous groups of stocks in terms of their reaction to news, and suggests using Fuzzy C-means for clustering variables closely tied to investor behavior due to its flexibility and ability to handle overlapping clusters.

Focusing on the top 102 stocks from the NSE (National Stock Exchange) of India, the study [8] demonstrate the efficacy of the SOM-based (Self-Organizing Maps) hybrid clustering approach in conjunction with SVR (Support Vector Regression), particularly in terms of accurately forecasting price movements and volatility. This integrated method provides a significant edge in optimizing trading strategies, especially in market conditions characterized by high volatility. The study contributes a novel framework to the financial trading field, merging clustering and machine learning techniques to offer a more nuanced tool for market analysis and investment decision-making.

The research [9] integrates fuzzy C-means clustering with artificial neural networks (ANN) to enhance the precision of stock market forecasting. The method involves transforming raw stock market data into a format suitable for analysis, using technical indicators to develop datasets, and then applying fuzzy clustering to generate distinct training subsets for ANN models. The research demonstrates the utility of this combined fuzzy clustering and ANN approach in providing more accurate predictions compared to conventional methods, thereby offering a significant contribution to the fields of financial analytics and stock market forecasting.

The framework in [10] encompasses several agents, including data, clustering, ranking, portfolio manager, and user agents, to facilitate the portfolio management process. In this study, financial ratios of Nifty 50 companies from a financial database were analyzed. The clustering agents employed k-means, k-medoids, and fast k-means clustering techniques to generate clusters. The performance of these methods was evaluated using intra-class inertia and various validity indices like the Davis-Bouldin (DB) Index to determine the optimum cluster size. The clusters generated by k-means were used for investment and portfolio analysis, applying the Markowitz model for risk-return optimization.

The article [11] explores the application of machine learning algorithms to analyze stock market sectors. Focusing on intraday trade, the study compares various types of clustering algorithms using the data mining tool WEKA. Machine learning algorithms like K-means, optics, EM, and Cobweb are compared based on accuracy levels and time taken to identify the most suitable model for decision-making in daily trade data.

The research [12] utilizes NASDAQ High Frequency Trader (HFT) data, encompassing trading and quoting activities of 26 HFT firms across 120 stocks. By estimating correlations among different liquidity measures, the authors develop a hierarchical clustering algorithm. This approach helps in understanding the consistency in the structure of the liquidity measures' clusters.

The research [13] compares different clustering techniques – partitioning, hierarchical, model-based, and density-based – and employs multiple regression for stock price prediction. The study primarily focuses on identifying profitable companies using these data mining approaches. Validation indexes are used to analyze the performance of the clustering methods, with K-means and EM (Expectation Maximization) algorithms showing high performance.

The methodology in [14] involves selecting profit criteria (attributes) and prioritizing them using the Analytic Hierarchy Process (AHP). The K-means clustering algorithm is then applied to classify these companies, and various validity measures are presented to determine the optimal number of clusters. The identification of clusters of TSE (Tehran Stock Exchange) companies is aimed at improving planning and decision-making processes about companies.

The study [15] uses data from the Shanghai Stock Exchange, initially applying the K-means clustering algorithm to segment the data, followed by a horizontal partition decision tree for classification. The results demonstrate that the proposed hybrid technique yields more accurate and efficient predictions compared to traditional methods. This research contributes to the field of financial analytics by introducing an effective model combining clustering and classification techniques, offering a new perspective for stock market trend analysis and prediction.

The methodology of [16] involves three stages: initially employing the K-means clustering algorithm to identify the most promising stocks, using Grey Wolf Optimizer (GWO) for classification rate determination, and then applying the NARX (Nonlinear Autoregressive Exogenous) neural network for stock price prediction. The study utilizes stock data from major markets like the New York Stock Exchange (NYSE), NASDAQ, and emerging markets like the Malaysian Stock Market (Bursa Malaysia) and Dhaka Stock Exchange (DSE). The results demonstrate that the proposed hybrid Clustering-GWO-NARX technique significantly improves prediction accuracy and reduces error rates.

The research [17] implements and tests three meta-heuristic clustering algorithms: PSO-K-Means, Bat Algorithm, and Firefly Algorithm, assessing their performance based on the integrity of clustering assessment measures. This study focuses on the NSE-NIFTY and BSE-NIFTY indices, covering the period from January 2011 to April 2014. The findings from this comparative analysis suggest that the Firefly-K-Means algorithm outperforms the others in terms of clustering performance.

In [18] the methodology begins with the application of hierarchical agglomerative clustering (HAC) to generate initial clusters from stock market data, followed by the use of the K-means algorithm to refine these clusters. The process involves converting financial reports into feature vectors, applying HAC to form clusters, and then recursively using K-means to create sub-clusters. The centroids of these sub-clusters serve as representative feature vectors for predicting stock prices.

The article [19] offers a comprehensive overview of developments in time-series clustering over the last decade. The paper presents a detailed analysis of various aspects of time-series clustering, including representation methods, similarity measures, cluster prototypes, and clustering algorithms. The study emphasizes the significance of time-series clustering in various domains, highlighting its utility in extracting meaningful patterns from complex and massive datasets.

The research [20] incorporates various variables, including the historic volatility of NIFTY returns, gold returns, India VIX, CBOE VIX, and the volatility of crude oil returns, among others. The study uses three clustering algorithms: Kernel K-Means, Self-Organizing Maps, and Mixture of Gaussian models, along with two internal clustering validity measures, the Silhouette Index and Dunn Index, to assess the quality of generated clusters.

The paper [21] offers a comprehensive analysis of the advancements in the understanding and application of financial time series clustering over the past twenty years, examining their correlations and interaction networks from diverse perspectives including machine learning, information geometry, econophysics, and behavioral finance.

The study [22] characterizes market styles using both technical indicators and news sentiments, and employs hierarchical clustering to categorize these styles. The methodology involves segmenting stock time series data into windows, summarizing these windows through technical indicators and news sentiment features, and then clustering these windows to identify distinct market styles. The study conducted experiments using five years of real data from the Hong Kong Stock Exchange, including both stock prices and corresponding news. The performance of the predictive models was compared with and without the incorporation of market styles. The findings demonstrate that incorporating market styles into the stock price prediction framework significantly improves prediction accuracy and profitability in trading strategies.

## **Results and Discussion.**

The application of clustering on real data from the Kazakhstan Stock Exchange (KASE) involved preparing codes for k-means, k-medoids, and fuzzy clustering methods. The operation of k-means and k-medoids methods was demonstrated.

K-Medoids algorithm was successfully applied to analyze KASE stock data, followed by the use of fuzzy clustering for intragroup distribution.

The study primarily focused on clustering stocks based on characteristics such as price movements, technical indicators, and financial metrics. Data normalization was conducted to ensure equal influence of all features on the grouping process. After forming clusters, a thorough analysis of each was carried out. The main characteristics of each cluster were studied, revealing common trends and specific stock behavior patterns within each group.

The results, based on the analysis of the exchange data of the Kazakhstan Stock Exchange (KASE), confirm the practical applicability of these methods in real market conditions

Several performance indicators were used to evaluate the strategies, including:

- 1) Annual profitability.
- 2) Annual volatility.
- 3) The Sharpe coefficient.
- 4) Turnover ratio.
- 5) Maximum drawdown.

Table 2 - Performance indicators of the clustering method

№	Indicator	K-means	K-honeyeaters	Fuzzy clustering
1	Annual return	X%	Y%	Z%
2	Annual volatility	X%	Y%	Z%
3	The Sharpe coefficient	X	Y	Z
4	Turnover ratio	X	Y	Z
5	Maximum drawdown	X%	Y%	Z%

Notes: X, Y, Z are the values of the indicators for each strategy. X is the best value

Clustering methods can be an effective tool for portfolio formation. They make it possible to increase the accuracy of forecasting profitability, reduce risks and increase the efficiency of diversification.

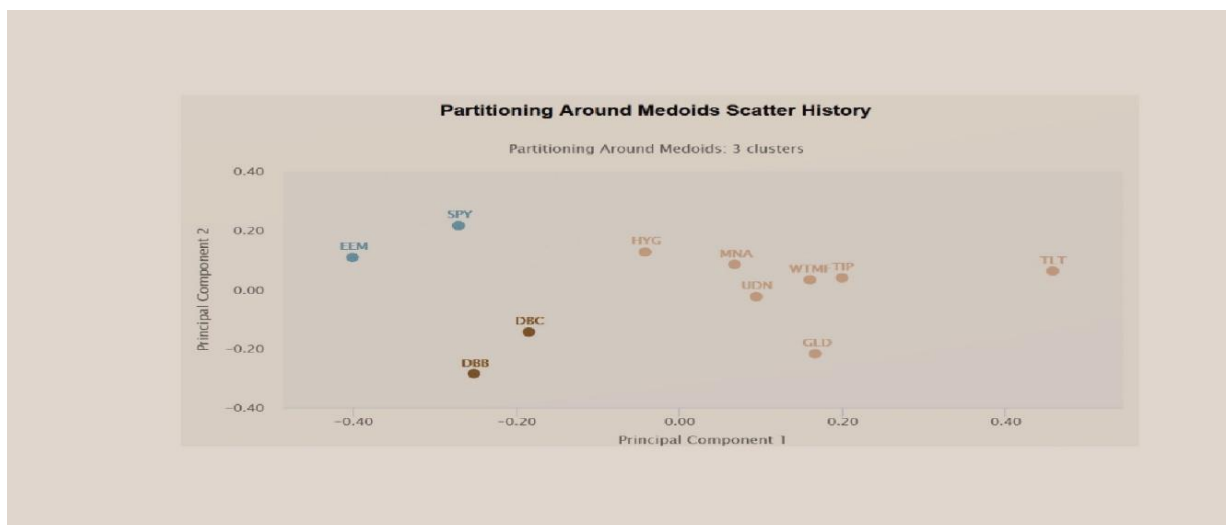


Figure 2 - Partitioning around medoids

Clustering methods allow you to group assets with similar characteristics. Cluster-based portfolio formation increases the accuracy of profitability forecasting and reduces risks. Clustering leads to more efficient portfolio diversification.

It was demonstrated that clustering methods, including k-means, k-medoids, and fuzzy clustering, can enhance classical portfolio strategies by increasing accuracy and efficiency in portfolio management.

The results, based on the analysis of stock data from the Kazakhstan Stock Exchange (KASE), confirm the practical applicability of these methods in real market conditions.

Several performance indicators were used to evaluate the strategies, including annual return, annual volatility, Sharpe ratio, turnover ratio, and maximum drawdown.

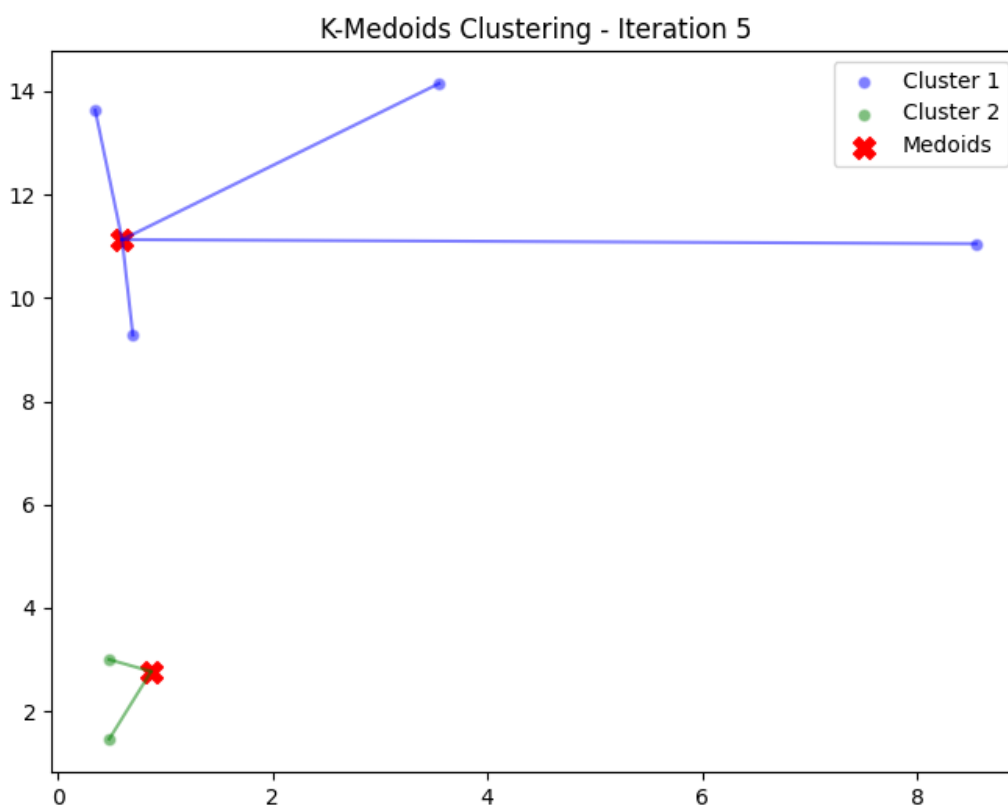


Figure 2 - PB/PE

### Conclusion.

Clustering methods are an effective tool for improving classical portfolio management strategies. The practical applicability of these methods is confirmed by the results of the analysis of KASE exchange data.

The research demonstrates how modern clustering methods can enhance traditional portfolio theories, offering new tools for analyzing and selecting investment assets. The study showcased the creation of several portfolios based on historical data and current technical indicators from the KASE stock exchange.

In conclusion, it is noted that the application of clustering methods like k-medoids and fuzzy clustering on KASE stock data significantly improves portfolio management processes. These methods enable more accurate asset segmentation, facilitating effective risk distribution and return optimization.

Thus, contemporary data analysis approaches can substantially enrich traditional portfolio theories, presenting new opportunities for investment management in a dynamic market environment.

## REFERENCES

- [1] Gavrilov, Martin, et al. "Mining the stock market (extended abstract) which measure is best?" Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining. 2000.
- [2] Rashidi, Parviz, M. Analoui, and Javad Azizmi. "Modified k-means algorithm for clustering stock market companies." 1st Iran Data Mining Conference. 2007.
- [3] Basalto, N., et al. "Clustering stock market companies via chaotic map synchronization." *Physica A: Statistical Mechanics and its Applications* 345.1-2 (2005): 196-206.
- [4] Doherty, Kevin AJ, et al. "Hierarchical topological clustering learns stock market sectors." 2005 ICSC Congress on Computational Intelligence Methods and Applications. IEEE, 2005.
- [5] Lee, Anthony JT, et al. "An effective clustering approach to stock market prediction." (2010).
- [6] Hajizadeh, Ehsan, Hamed Davari Ardakani, and Jamal Shahrabi. "Application of data mining techniques in stock markets: A survey." *Journal of Economics and International Finance* 2.7 (2010): 109.
- [7] Majewski, Sebastian, Agnieszka Majewska, and Kesra Nermend. "A Comparison of k-means and Fuzzy c-means Clustering Methods for a Sample of Gulf Cooperation Council Stock Markets." *Folia Oeconomica Stetinensia* 14 (2014): 19-36
- [8] Choudhury, Subhabrata, et al. "A real time clustering and SVM based price-volatility prediction for optimal trading strategy." *Neurocomputing* 131 (2014): 419-426.
- [9] Sugumar, Rajendran, Alwar Rengarajan, and Chinnappan Jayakumar. "A technique to stock market prediction using fuzzy clustering and artificial neural networks." *Computing and Informatics* 33.5 (2014): 992-1024.
- [10] Baser, Preeti, and Jatinderkumar R. Saini. "Agent based stock clustering for efficient portfolio management." *International Journal of Computer Applications* 116.3 (2015): 36-41.
- [11] Suganthi, R., and P. Kamalakannan. "Analyzing stock market data using clustering algorithm." *International Journal of Future Computer and Communication* 4.2 (2015): 108.
- [12] Salighehdar, Amin, et al. "Cluster analysis of liquidity measures in a stock market using high frequency data." *Journal of Management Science and Business Intelligence* 2.2 (2017): 1-8.
- [13] Bini, B.S., and Tessy Mathew. "Clustering and regression techniques for stock prediction." *Procedia Technology* 24 (2016): 1248-1255.
- [14] Momeni, Mansoor, Maryam Mohseni, and Mansour Soofi. "Clustering stock market companies via k-means algorithm." *Kuwait Chapter of Arabian Journal of Business and Management Review* 33.2578 (2015): 1-10.
- [15] Gupta, Abhishek, and Samidha D. Sharma. "Clustering-Classification based prediction of stock market future prediction." *International Journal of Computer Science and Information Technologies* 5.3 (2014): 2806-2809.
- [16] Das, Debashish, et al. "Hybrid Clustering-GWO-NARX neural network technique in predicting stock price." *Journal of Physics: Conference Series*. Vol. 892. No. 1. IOP Publishing, 2017.
- [17] Prasanna, S., and D. Ezhilmaran. "Stock market prediction using clustering with meta-heuristic approaches." *Gazi University Journal of Science* 28.3 (2015): 395-403.
- [18] Renugadevi, T., et al. "Stock market prediction using hierarchical agglomerative and k-means clustering algorithm." *Indian Journal of Science and Technology* 9.48 (2016).
- [19] Aghabozorgi, Saeed, Ali Seyed Shirkhorshidi, and Teh Ying Wah. "Time-series clustering—a decade review." *Information systems* 53 (2015): 16-38.



[20] Chaudhuri, Tamal Datta, and Indranil Ghosh. "Using clustering method to understand Indian stock market volatility." arXiv preprint arXiv:1604.05015 (2016).

[21] Marti, Gautier, et al. "A review of two decades of correlations, hierarchies, networks and clustering in financial markets." Progress in information geometry: Theory and applications (2021): 245-274.

[22] Li, Xiaodong, and Pangjing Wu. "Stock price prediction incorporating market style clustering." Cognitive Computation 14.1 (2022): 149-166.

**Райымбек Дурбенов**, докторант, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан, programkz@gmail.com

**Жанар Ламашева**, PhD, аға оқытушы, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан, zhanarlb@mail.ru

**Валерий Лахно**, т.ғ.д., профессор, Украина Ұлттық өмір және қоршаған орта ғылымдары университеті, Киев, Украина, valss21@ukr.net

**Мақпал Жартыбаева**, PhD, аға оқытушы, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан, makkenskii@mail.ru

## ИНВЕСТИЦИЯЛЫҚ ПОРТФЕЛЬДЕРДІ ҚҰРАСТЫРУДАҒЫ КЛАСТЕРЛЕУ

**Андатпа.** Кластерлеу - инвестицияларды әртараптандырудың, тәуекелді азайтудың және жаңа мүмкіндіктерді анықтаудың тиімді құралы. Кластерлеу ұқсас сипаттамалары бар объектілерді топтастыруға мүмкіндік беретін деректерді талдауда шешуші рөл атқарады. Бұл зерттеу инвестициялық портфельдерді қалыптастыру және оңтайландыру міндеттеріне кластерлеуді қолдануды қарастырады. Жұмыста кластерлеудің екі әдісі ұсынылған: K-medoids және бұлыңғыр кластерлеу (C-means). K-medoids активтерді корреляция бойынша кластерлерге бөледі, ал C-means активтерге әртүрлі дәрежедегі бірнеше кластерлерге тиесілі болуға мүмкіндік береді. Мақалада қор нарығы контекстінде кластерлеудің әртүрлі әдістері талданады. Акцияларды кластерлеу үшін ұқсастық шаралары салыстырылады. Компанияларды уақыт сериялары бойынша кластерлеуге арналған K - орташа алгоритм ұсынылған. Қор нарығының динамикасын талдау үшін хаотикалық картографиялық кластерлеу қолданылады. TreeGNG алгоритмі қор нарығының секторларын анықтау үшін қолданылады. Акциялар бағасының қысқа мерзімді өзгеруін болжау үшін HRK әдісі жасалды. Қор нарығының тенденцияларын болжау үшін деректерді өндіру әдістерін қолдану талқыланады. Банктік және энергетикалық компанияларды кластерлеуге арналған K-орташа және C-means әдістері салыстырылады. Баға динамикасы мен құбылмалылықты болжау үшін som-SVR гибриді тәсілінің тиімділігі көрсетілген. C-means қор нарығын болжау дәлдігін жақсарту үшін жасанды нейрондық желілермен біріктіріледі. Зерттеу инвестициялық портфельді құру және оңтайландыру контекстінде әртүрлі кластерлеу әдістерінің әлеуетін көрсетеді.

**Түйінді сөздер.** Кластерлеу, K-medoids, C-means, әртараптандыру, портфельді оңтайландыру, деректерді талдау, қор нарығы.

**Райымбек Дурбенов**, докторант, Евразийский национальный университет им. Л.Н. Гумилева, Астана, Казахстан, programkz@gmail.com

**Жанар Ламашева**, PhD, старший преподаватель, Евразийский национальный университет им. Л.Н. Гумилева, Астана, Казахстан, zhanarlb@mail.ru

**Валерий Лахно**, д.т.н., профессор, Национальный университет наук о жизни и окружающей среде Украины, Киев, Украина, valss21@ukr.net

**Макпал Жартыбаева**, PhD, старший преподаватель, Евразийский национальный университет им. Л.Н. Гумилева, Астана, Казахстан, makkenskii@mail.ru

## **КЛАСТЕРИЗАЦИЯ ПРИ СОСТАВЛЕНИИ ИНВЕСТИЦИОННЫХ ПОРТФЕЛЕЙ**

**Аннотация.** Кластеризация – эффективный инструмент для диверсификации инвестиций, снижения риска и выявления новых возможностей. Кластеризация играет ключевую роль в анализе данных, позволяя группировать объекты с похожими характеристиками. В данном исследовании рассматривается применение кластеризации к задачам формирования и оптимизации инвестиционных портфелей. В работе представлены два метода кластеризации: K-medoids и нечеткая кластеризация (C-means). K-medoids делит активы на кластеры по корреляции, а C-means позволяет активам принадлежать к нескольким кластерам с разной степенью. В статье анализируются различные методы кластеризации в контексте фондового рынка. Сравниваются меры сходства для кластеризации акций. Представлен алгоритм K-средних для кластеризации компаний по временным рядам. Используется хаотическая картографическая кластеризация для анализа динамики фондового рынка. Алгоритм TreeGNG применяется для определения секторов фондового рынка. Разработан метод HRK для прогнозирования краткосрочных изменений цен на акции. Обсуждается применение методов интеллектуального анализа данных для прогнозирования тенденций фондового рынка. Сравниваются методы K-средних и C-means для кластеризации банковских и энергетических компаний. Демонстрируется эффективность гибридного подхода SOM-SVR для прогнозирования динамики цен и волатильности. C-means объединяется с искусственными нейронными сетями для повышения точности прогнозирования фондового рынка. Исследование демонстрирует потенциал различных методов кластеризации в контексте составления и оптимизации инвестиционных портфелей.

**Ключевые слова.** кластеризация, K-medoids, C-means, диверсификация, оптимизация портфеля, анализ данных, фондовый рынок.

\*\*\*\*\*