BAREKENG: Journal of Mathematics and Its Applications
March 2025 Volume 19 Issue 1 Page 0153–0162
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss1pp0153-0162

UNDERSTANDING LQ45 STOCKS (2021-2023) WITH K-MEANS CLUSTERING

Margareta Febe[1*], Giovanny Theotista[2], Winson[3]

[1]*Accounting Study Program, Faculty of Social and Humaniora, Universitas Bunda Mulia*
[2,3]*Data Science Study Program, Faculty of Technology and Design, Universitas Bunda Mulia*
Jl. Lodan Raya no. 2, Jakarta-14430, Indonesia

*Corresponding author's e-mail: * **margaretafebe@gmail.com***

ABSTRACT

Article History:

Received: 30th April 2024
Revised: 26th August 2024
Accepted: 11th September 2024
Published:13th January 2025

Keywords:

Financial Analysis;
Investment Decision;
K-means Clustering.

The primary aim of this study is to examine the use of K-Means clustering in analyzing LQ45 stocks from 2021 to 2023, utilizing data obtained from the Yahoo Finance platform. The analysis delves into key performance measures such as the price-to-earnings ratio (PER), earnings per share (EPS), dividends, trading volume, and historical return on investment. This technique categorizes stocks with similar characteristics, providing financial analysts, money managers, and investors with valuable insights. The objective of the clustering analysis is to gain a deeper understanding of the relationship between intrinsic stock features and the inherent price volatility of companies. This is accomplished by using historical datasets to conduct stock feature analysis. Mathematics plays a crucial role in the K-Means model by providing the foundational algorithms and statistical methods used to categorize and analyze the data. The study contributes to the field of financial market analysis by demonstrating how understanding group-to-group dynamics can affect investment decisions and offering a more precise representation of large datasets in financial contexts. These findings provide significant insights for individuals involved in financial matters in the stock market, helping to identify potential investment opportunities and reduce risk more effectively.

How to cite this article:

M. Febe, G. Theotista and Winson., "UNDERSTANDING LQ45 STOCKS (2021-2023) WITH K-MEANS CLUSTERING," *BAREKENG: J. Math. & App.*, vol. 19, iss. 1, pp. 0153-0162, March, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

The LQ45 index, a widely recognized benchmark in the Indonesian stock market, comprises 45 stocks carefully selected and updated on a semi-annual basis. These updates are based on several factors, including market capitalization, trading frequency, stock market liquidity, financial health, and growth prospects [1]. Alongside the Composite Index of Indonesia (IHSG), the LQ45 index provides a snapshot of the performance of actively traded stocks on the Indonesia Stock Exchange (IDX), serving as a reliable indicator of market sentiment and a critical tool for guiding investment decisions [2].

The inherent volatility and fluctuations characteristic of the stock market are influenced by numerous factors, such as macroeconomic indicators, geopolitical events, and investor sentiment. Therefore, investors and financial analysts need a comprehensive understanding of these dynamics to make informed decisions. In this context, clustering analysis, particularly K-Means clustering, emerges as a crucial tool to identify patterns and relationships among stocks based on various criteria, such as price, return, trading volume, price-to-earnings ratio (PER), earnings per share (EPS), dividend yield, and other relevant factors [3].

Despite the growing use of advanced analytical methods to detect correlations and patterns in extensive financial datasets, there is a noticeable gap in the literature regarding the application of K-Means clustering to the LQ45 index. Previous studies have demonstrated the effectiveness of K-Means in segmenting stocks based on similar characteristics, facilitating the creation of optimal stock portfolios. However, these studies have primarily focused on markets outside Indonesia or on different stock indices [4], [5]. This research intends to fill this gap by applying K-Means clustering specifically to LQ45 stocks, providing insights that are directly applicable to the Indonesian stock market.

The simplicity, efficiency, and scalability of K-Means clustering make it a particularly advantageous choice for stock analysis. The method's ability to quickly process large datasets and its straightforward implementation make it accessible to both seasoned financial analysts and beginners [6]. Furthermore, K-Means clustering effectively identifies distinct clusters of stocks with similar performance indicators, aiding in the diversification of investments and minimizing risk. Its adaptability allows for integration with other clustering techniques or use as an initial step before more complex analyses, enhancing the overall quality of stock groupings [7].

This study aims to utilize K-Means clustering to analyze the LQ45 stocks from 2021 to 2023. The objective is to provide valuable insights into the risks, returns, and investment opportunities within the Indonesian stock market during this period. By focusing on this specific index, the study addresses the urgent need for more localized, data-driven analysis tools in the digital era, equipping financial professionals and investors with the necessary tools to navigate the complexities of the modern stock market.

2. RESEARCH METHODS

According to [8], [9], and [10], K-Means Clustering is a technique employed to categorize data points into clusters according to their resemblance. During the process of data analysis, it is usual practice to utilize the K-Means Clustering technique to classify data into separate clusters based on shared criteria. When presenting this method in a research or instructional setting, it is essential to effectively communicate the following information to the intended recipients:

K-Means Clustering is a widely employed technique for grouping data into clusters based on the similarities among the clusters within the data. This practice involves homogenizing the data, ensuring that data within each group have similar characteristics, while data across different groups exhibit significant differences.

Before we continue to describe the K-Means Clustering algorithm, we must know about Euclidean distance. Euclidean distance is a measure of the straight-line distance between two points (P_1 and P_2) in Euclidean space. For a two-dimensional space, it's given by:

$$d = \left\|\overrightarrow{P_1 P_2}\right\| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

In three dimensions:

$$d = \left\|\overrightarrow{P_1P_2}\right\| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{2}$$

In n-dimensional space:

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \ldots + (u_n - v_n)^2} \tag{3}$$

Clustering algorithms like K-means use Euclidean distance to determine which cluster (centroid) a data point should belong to. Each cluster is represented by a centroid (mean point), and data points are assigned to the nearest centroid based on Euclidean distance.

The goal of clustering is to group data points that are similar to each other based on their features. Similarity is measured by proximity in the n-dimensional space, where smaller Euclidean distances indicate higher similarity.

The algorithm comprises several crucial steps:

1. The process of selecting centroids starts by first finding K locations to act as the initial centroids for the potential groups. Furthermore, the appropriate number of groupings is predetermined.
2. Clusters are created by allocating data points to centroids based on their closeness. A highly efficient approach to achieve this purpose is to minimize an objective function that is directly associated with distance, with the Euclidean distance being the primary technique.
3. Once the assignment of data points to their respective clusters is completed, the centroids' positions are updated by computing the mean of all data points inside each cluster.
4. The iterative process of repeating steps two and three continues until the positions of the centroids remain unchanged or until the convergence conditions, such as reaching a maximum number of iterations or a minimal number of changes in centroids, are met.

Once the algorithm achieves convergence, it will generate K-clusters, each consisting of its own centroids and assigned data points.

2.1 Data

This study uses data from the LQ45 Index as listed on January 5th, 2024. The LQ45 Index consists of 45 companies that are considered to be the most liquid and have strong financial fundamentals on the Indonesia Stock Exchange (IDX). Below is a table listing the company names and their corresponding codes included in the LQ45 Index:

Table 1. LQ45 Stock Companies

No	Kode	Nama Perusahaan
1	ACES	Ace Hardware Indonesia Tbk.
2	ADRO	Adaro Energy Indonesia Tbk.
3	AKRA	AKR Corporindo Tbk
4	AMRT	Sumber Alfaria Trijaya Tbk.
5	ANTM	Aneka Tambang Tbk.
6	ARTO	Bank Jago Tbk
7	ASII	Astra International Tbk
8	BBCA	Bank Central Asia Tbk.
⋮	⋮	⋮
45	UNVR	Bank Rakyat Indonesia (Persero) Tbk

***Data source**: Yahoo Finance,* accessed on January 5th, 2024.

The following table shows 45 company codes, sectors, EPS, dividends, company stock prices, and volume of trade, which are members of the Indonesian stocks exchange LQ45 taken throughout 2021-2023 on the Yahoo Finance website, a total of 732 work dates (from January 4th, 2021 until December 29th, 2023) which was launched on January 5th, 2024.

Table 2. LQ45 Stock Dataset

Company Code	Sector	EPS	Total Dividend	Date: 1/4/2021		⋯	Date: 12/29/2023	
				Closing Price	Volume of Trade		Closing Price	Volume of Trade
ACES	Consumer Cyclical	46.66	83.80	1,700	12,191,600	⋯	720	16,090,900
ADRO	Energy	910.9	1,059.87	1,455	110,366,200	⋯	2,380	124,776,700
AKRA	Energy	129.17	194.00	654	109,114,000	⋯	1,475	34,881,900
AKRA	Energy	129.17	52.14	800	1,572,500	⋯	1,475	34,881,900
AMRT	Consumer Defensive	79.2	134.98	2,190	1,040,753,400	⋯	2,930	28,228,000
ANTM	Basic Materials	168.19	0.00	3,516.35	2,032,849	⋯	1,705	83,597,900
ARTO	Financial Services	1.7	83.80	6,225	34,840,200	⋯	2,900	7,581,900
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
UNVR	Consumer Defensive	130.13	516.00	7,475	8,203,000	⋯	3,530	6,004,400

Data source*: Yahoo Finance,* accessed on January 5th, 2024.

2.2 Return On Investment

Return on Investment is a quantitative measure applied to analyze the performance of a stock during a given timeframe while accounting for swings in stock price over time. Within the area of financial analysis, the statistic known as return on investment (ROI) is applied to assess the performance of a given fund throughout the designated time frame. Following the requirements of the analysis, return prices have the flexibility to be updated at any time intervals, including weekly, monthly, quarterly, or annual. This helps investors and analysts to acquire insights into the performance of the aforementioned company during various periods.

We follow **[11]** to describe return on investment, which is computed by subtracting the difference between the original and final prices over a given time. This resulting sum is then stated as a percentage of the initial price, known as the return definition. In mathematical terms, the return value for a given period $[t, \Delta t]$, when not divided by S_t, can be expressed as the price at the time t.

$$\hat{R}_{t+\Delta t,t} = \frac{S_{t+\Delta t} - S_t}{S_t} \tag{4}$$

The return on investment is obtained by calculating the natural logarithm of the stock price.

$$R_{t+\Delta t,t} = \ln \frac{S_{t+\Delta t}}{S_t} \tag{5}$$

This definition possesses several advantages, namely meeting the characteristics of addition:

$$\begin{aligned} R_{N,0} &= \ln \frac{S_N}{S_0} = \ln\left(\frac{S_N}{S_{N-1}} \frac{S_{N-1}}{S_{N-2}} \dots \frac{S_1}{S_0}\right) \\ &= \ln \frac{S_N}{S_{N-1}} + \ln \frac{S_{N-1}}{S_{N-2}} + \dots + \ln \frac{S_1}{S_0} = \sum_{i=1}^{N} \hat{R}_{i,i-1} \end{aligned} \tag{6}$$

2.3 Trading volume

According to **[12]** and **[13]**, trading volume is a determinant that diminishes the monetary value transacted within a specified period. A reduction in interest can be inferred from a low volume, while a high volume typically indicates a rise in interest in stocks. Technical analysts often employ trading volume as a method to confirm trend pricing and evaluate the viability of a trend.

2.4 Price-to-Earnings Ratio (PER)

We follow [12] and [13], the PER represents the correlation between the present stock price and the earnings per share. In the field of fundamental financial research, this specific indication is highly significant. The PER is a metric used to assess the valuation of a stock, representing the price an investor is prepared to pay for each dollar of a company's shares. A high price-to-earnings ratio may suggest a significant valuation, while a low ratio may suggest a reduced valuation.

2.5 Earnings Per Share (EPS)

According to [12] and [13], the earnings per share (EPS) of a firm denotes the company's net income divided by the number of outstanding shares. This statistic is essential in assessing the financial sustainability of a business. Investors generally view regularly rising profits per share (EPS) as a favorable sign because it indicates a steady pattern of growth.

2.6 Dividend

We follow [12] and [13], a dividend is the systematic allocation of funds by a company to its shareholders. Often, the company's profits are used to distribute dividends. Investors who are looking for a high return on their investment could be greatly affected by a growing and reliable stream of dividends. Moreover, dividends can significantly influence a company's overall health and financial stability.

3. RESULTS AND DISCUSSION

This part will use Python 3.1 and Microsoft Excel to preprocess and model the data. The preprocessing data are to create the desired feature, encoding, creating a new feature (PER), scaling, and feature selection. After preprocessing the data, the data is clustered using K-Means Clustering followed by the Elbow Method to determine the number of clusters that are ideal using Python 3.1. After modeling, the results are compared and interpreted, as well as recommendations for what investors should do.

3.1 Data Preprocessing: Feature Importance

The primary step in determining the mean trade volume involves estimating the average daily trading volume for the period spanning from 2021 to 2023. Subsequently, the total dividend percentage, where the dividend percentage (D) is the total dividend amount received within a particular time interval divided by the price after the period (the end of the year 2023). It is possible to phrase it as $D = \frac{\sum D_j'}{S_0}$. The next step is to determine PER where $PER = \frac{Stock\ Price}{EPS}$. In order to assess the historical performance of the stock market and the volatility of the stock market, the average daily return and standard deviation of stock returns will be computed. The Price-to-Earnings ratio, also known as the PER ratio, is determined by dividing the closing stock price by the earnings per share, also known as the EPS.

3.2 Data Preprocessing: One Hot Encoding

The process of transforming categorical variables or qualitative data into a format that can be processed by a machine learning model is achieved through the utilization of a technique known as one-hot encoding. This technique is particularly useful in the field of data analysis [14]. Due to the fact that it can be classified into a number of different categories, such as finance, energy, consumption, and others, this study classifies industrial type as a categorical variable.

Prior to undergoing transformation through one-hot encoding, the industries of a company can be represented as a single column with category values such as "Financial Services," "Consumer Defensive," "Basic Materials," and so on. This is done before the transformation process takes place. It is typical for machine learning models to rely on numerical input, which is why it is necessary to convert categorical variables into numerical representation to make algorithmic interpretation easier.

The process that is referred to as "one-hot encoding" entails the construction of individual columns for each distinct category that is contained in the categorical variable. Using the criteria that have been defined, a value of 1 will be assigned to each column if the observation is connected with the relevant category, and a value of 0 will be allocated to each column otherwise. To phrase it another way, each category is given its column and is represented as a binary variable [15].

3.3 Data Preprocessing: Scaling

The transformation of data into a predetermined scale is accomplished through the process of scaling, which is sometimes referred to as normalization [16]. Within the world of clustering algorithms, such as K-Means Clustering, as well as a variety of other machine learning techniques, the procedure that was discussed above carries a significant amount of weight. K-Means Clustering is a technique that classifies data points into clusters by taking into consideration the distance between them. This is accomplished by the utilization of the Euclidean distance metric. There is a substantial amount of sensitivity displayed by the algorithm to variations in scale among several features. The distance of connection will have a bigger dampening impact as a consequence of a characteristic that demonstrates a higher value of rent. When compared to another feature, a feature that possesses a higher value of rent can induce clustering and reduce the influence of other features that exhibit lower value or rents. This is a hypothetical scenario that exists.

Before beginning the scaling process, it is essential to conduct an examination of the characteristics of the numerical data in order to determine whether or not any outliers are among the observations. An overview of the findings of the study can be described as follows:

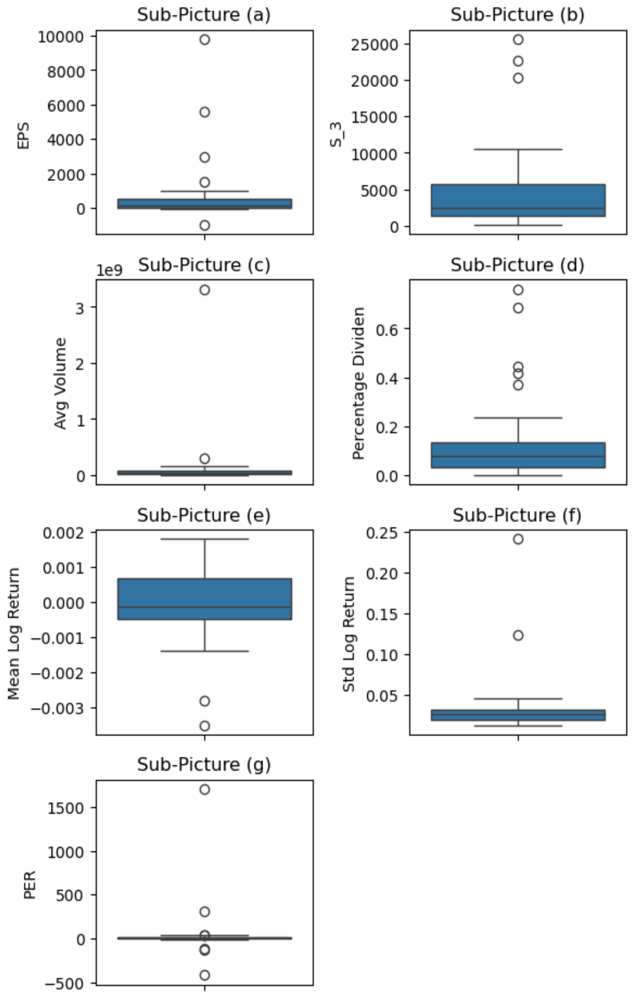

Figure 2. Boxplot Normality Test
(a) EPS, (b) Stock Price at the end of 2023, (c) Average Volume Trade, (d) Percentage Dividend, (e) Mean Lognormal ROI, (f) Standard Deviation Lognormal ROI, (g) PER

In the case of any outliers within the dataset, it is advisable to utilize the technique of Absolute Maximum Scaling [17]. The method of absolute maximum scaling, alternatively referred to as normalizing or absolute maximum scaling, involves dividing each value inside a given set by the absolute maximum value within that set. The purpose is to edit the data in such a way that the greatest number in the appropriate column is either 1 or -1. This strategy is often applied in machine learning and data analysis to guarantee that variable magnitudes of the same function do not result in disproportionately big effects on the output of a limited number of algorithms.

The ideas underlying the absolute maximum scaling can be summarized as follows:

$$x_i' = \frac{x_i}{|x_{max}|} \tag{7}$$

3.4 K-Means Clustering

During this stage, Clustering will be conducted using the following features: 'EPS', 'end-of-year price 2023', 'average volume of purchases 2021-2023', Percentage Dividend 2021-2023', 'mean return 2021-2023', 'standard return 2021-2023', 'PER end-of-year 2023', and a binary variable representing the outcomes of one-hot encoding.

The Elbow Method is a widely used technique in cluster analysis to estimate the optimal number of clusters, as demonstrated in works [18], [19], and [20]. This technique is frequently employed in conjunction with the K-Means algorithm, although it can also be utilized with other clustering techniques. The Elbow technique aims to determine the optimal point at which further addition of clusters ceases to result in a substantial enhancement in the clustering quality.

The Elbow technique consists of the following steps:

1. Compute the WCSS (Within-Cluster Sum of Squares): WCSS refers to the summation of the squared distances between every data point within a cluster and its corresponding centroid (or center) of that cluster. WCSS quantifies the degree to which data points inside a cluster are tightly grouped around the centroid.
2. Experiment with different values of K: Experiment with different values of K (the number of clusters) within the range of $K = 1$ to a specified upper limit (such as 10 or higher). Calculate the appropriate WCSS for each value of K.
3. Plot illustrating the relationship between the Within-Cluster Sum of Squares (WCSS) and the number of clusters (K): Produce a visual representation that demonstrates the correlation between Within-Cluster Sum of Squares (WCSS) and the variable K. The X-axis denotes the number of clusters (K), while the Y-axis indicates the within-cluster sum of squares (WCSS).
4. Determine the Elbow point: The Elbow point refers to the specific point on the graph where the rate of decrease in WCSS (Within-Cluster Sum of Squares) begins to considerably slow down. This fact suggests that the addition of more clusters does not result in a substantial enhancement in the clustering's quality.

After identifying the elbow point, you can choose the ideal number of clusters for your clustering study depending on this point. This facilitates the attainment of optimal clustering results. The following visualization results were achieved using the elbow method:

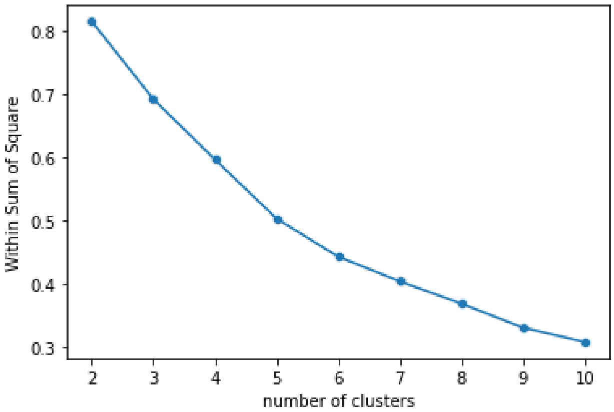

Figure 3. Elbow Method

One might infer that the ideal value for K is 5. Subsequently, a thorough examination is conducted on the clustering results obtained with a value of K equal to five. The following are the outcomes of clustering with a value of K set to five:

Table 3. K-Means Cluster of LQ45 Stock

Company Code Cluster "0"	Company Code Cluster "1"	Company Code Cluster "2"	Company Code Cluster "3"	Company Code Cluster "4"
AMRT	ARTO	ADRO	EMTK	ACES
CPIN	BBCA	AKRA	EXCL	ANTM
GGRM	BBNI	ESSA	SCMA	ASII
ICBP	BBRI	HRUM	TBIG	BRPT
INDF	BBTN	INDY	TLKM	BUKA
UNVR	BMRI	ITMG	TOWR	GOTO
	BRIS	MEDC		INCO
		PTBA		INKP
				INTP
				KLBF
				MAPI
				MDKA
				PGAS
				SIDO
				SMGR
				TPIA
				UNTR

***Data source**: Author's Jupyter Notebook*

The results of this research reinforce the efficacy of K-Means clustering in financial data analysis, aligning well with existing theoretical frameworks and previous empirical studies. The segmentation of the stock group into five distinct clusters, characterized by unique attributes and spanning various industries such as Consumer Defensive, Financial Services, Energy, and Communication Services, highlights the method's ability to discern meaningful patterns within financial datasets. This capability to capture industry-specific performance metrics aligns with previous research, indicating that K-Means clustering effectively groups stocks with similar performance indicators, facilitating more informed investment decisions and contributing to portfolio diversification and risk minimization. By identifying distinct clusters, investors can achieve balanced portfolios that mitigate exposure to sector-specific risks, showcasing the practical application of K-Means clustering in developing diversified stock portfolios.

The study also confirms the scalability and processing efficiency of K-Means, which is crucial for handling extensive financial datasets generated by financial markets. Previous studies have emphasized K-Means' efficiency in managing large datasets, and the current research supports these claims, highlighting the method's robustness in large-scale financial analysis. Furthermore, the adaptability of K-Means allows it to be used in conjunction with other clustering methods or as a preliminary step in more complex analyses, enhancing the overall quality of stock groupings and enabling more refined investment strategies. These findings, consistent with existing literature, reinforce the role of K-Means as a foundational tool in financial analysis and portfolio management, providing empirical evidence of its utility in segmenting stocks across different industries and enhancing the theoretical understanding of clustering in financial contexts.

4. CONCLUSIONS

The clustering analysis reveals that the shares can be classified into five distinct clusters, namely the Consumer Defensive sector, Financial Services sector, Energy sector, Communication Services sector, and other sectors. The conclusion of the analysis can assist investors in making informed investment choices and gaining a deeper understanding of the dynamics of the stock market. Investors may utilize strategic methods to determine the sector that best aligns with their investment goals, based on the findings of the investigation. In order to maximize profitability, it is essential to tailor it based on the unique characteristics of each cluster.

Given the current situation, applying K-Means Clustering analysis to the LQ45 stock provides valuable information about the stock's characteristics and helps make smarter investment decisions in a changing market.

ACKNOWLEDGMENT

This research is financially supported by the grant 2023/2024 provided by Bunda Mulia University.

REFERENCES

[1] I. Sudarman and N. Diana, "The Effect of Financial Ratios on Sharia Stock Prices in Companies in the LQ45 Index 2020-2021," *Jurnal Ilmiah Ekonomi Islam*, vol. 8, no. 1, p. 117, Feb. 2022, doi: 10.29040/jiei.v8i1.4228.

[2] J. Saputri, D. Oktafalia, and M. Muzammil, "Analisis Return Portofolio yang Optimal Pada Saham LQ 45 yang Tercatat di Bursa Efek Indonesia Selama Periode," *Business & Management Journal Bunda Mulia*, vol. 8, no. 1, 2008.

[3] B. Supeno, "Competitive Environment Analysis Through Finance Fundamentals Analysis: LQ45 stock prices on the Indonesia stock exchange," *IOP Conf Ser Earth Environ Sci*, vol. 469, no. 1, p. 012074, Apr. 2020, doi: 10.1088/1755-1315/469/1/012074.

[4] J. Rowena, "Earnings Volatility, Kebijakan Dividen, Dan Pertumbuhan Asset Berpengaruh Terhadap Volatilitas Harga Saham Pada Perusahaan Manufaktur Di BEI Periode," *Jurnal Administrasi Kantor*, vol. 5, no. 2, pp. 231–242, 2017.

[5] D. Wu, X. Wang, and S. Wu, "Construction of stock portfolios based on k-means clustering of continuous trend features," *Knowl Based Syst*, vol. 252, p. 109358, Sep. 2022, doi: 10.1016/j.knosys.2022.109358.

[6] Z. Fang and C. Chiao, "Research on prediction and recommendation of financial stocks based on K-means clustering algorithm optimization," *Journal of Computational Methods in Sciences and Engineering*, vol. 21, no. 5, pp. 1081–1089, Nov. 2021, doi: 10.3233/JCM-204716.

[7] B. B. Nair, P. K. S. Kumar, N. R. Sakthivel, and U. Vipin, "Clustering stock price time series data to generate stock trading recommendations: An empirical study," *Expert Syst Appl*, vol. 70, pp. 20–36, Mar. 2017, doi: 10.1016/j.eswa.2016.11.002.

[8] T. M. Ghazal *et al.*, "Performances of K-Means Clustering Algorithm with Different Distance Metrics," *Intelligent Automation & Soft Computing*, vol. 29, no. 3, pp. 735–742, 2021, doi: 10.32604/iasc.2021.019067.

[9] M.-S. Yang and K. P. Sinaga, "A Feature-Reduction Multi-View k-Means Clustering Algorithm," *IEEE Access*, vol. 7, pp. 114472–114486, 2019, doi: 10.1109/ACCESS.2019.2934179.

[10] P. Govender and V. Sivakumar, "Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019)," *Atmos Pollut Res*, vol. 11, no. 1, pp. 40–56, Jan. 2020, doi: 10.1016/j.apr.2019.09.009.

[11] K. A. Sidarto, M. Syamsuddin, and N. Sumarti, *Matematika Keuangan*, 1st ed. Bandung: ITB Press, 2019.

[12] V. Veny and Y. Gunawan, "Perubahan Harga Saham Dilihat dari Faktor Fundamental Perusahaan Makanan dan Minuman," *Jurnal Akuntansi Bisnis*, vol. 15, no. 1, Feb. 2022, doi: 10.30813/jab.v15i1.2874.

[13] J. H. Wijaya and K. Gusni, "The Influence of Financial Performance against Stock Prices in Companies Listed in LQ45 Index 2012-2016 Period," 2018. [Online]. Available: www.sciencepubco.com/index.php/IJET

[14] A. C. H. Choong and N. K. Lee, "Evaluation of convolutionary neural networks modeling of DNA sequences using ordinal versus one-hot encoding method," in *2017 International Conference on Computer and Drone Applications (IConDA)*, IEEE, Nov. 2017, pp. 60–65. doi: 10.1109/ICONDA.2017.8270400.

[15] M. K. Dahouda and I. Joe, "A Deep-Learned Embedding Technique for Categorical Features Encoding," *IEEE Access*, vol. 9, pp. 114381–114391, 2021, doi: 10.1109/ACCESS.2021.3104357.

[16] V. Sharma, “A Study on Data Scaling Methods for Machine Learning,” *International Journal for Global Academic & Scientific Research*, vol. 1, no. 1, Feb. 2022, doi: 10.55938/ijgasr.v1i1.4.

[17] M. Ahsan, M. Mahmud, P. Saha, K. Gupta, and Z. Siddique, “Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance,” *Technologies (Basel)*, vol. 9, no. 3, p. 52, Jul. 2021, doi: 10.3390/technologies9030052.

[18] D. Marutho, S. Hendra Handaka, E. Wijaya, and Muljono, “The Determination of Cluster Number at k-Mean Using Elbow Method and Purity Evaluation on Headline News,” in *2018 International Seminar on Application for Technology of Information and Communication*, IEEE, Sep. 2018, pp. 533–538. doi: 10.1109/ISEMANTIC.2018.8549751.

[19] A. Sofia, “K-MEANS CLUSTERING USING ELBOW METHOD IN CASE OF DIABETES MILLETUS TYPE II IN INDONESIA,” *Journal of Statistical Methods and Data Science*, vol. 1, no. 2, Apr. 2024, doi: 10.31258/jsmds.v1i2.5.

[20] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, “Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster,” *IOP Conf Ser Mater Sci Eng*, vol. 336, p. 012017, Apr. 2018, doi: 10.1088/1757-899X/336/1/012017.