BAREKENG: Journal of Mathematics and Its Applications
March 2025 Volume 19 Issue 1 Page 0163–0172
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss1pp0163-0172

ANALYSING MARKET DYNAMICS: REVEALING OBSCURED PATTERNS IN LQ45 STOCKS (2021-2023) USING WARD'S HIERARCHICAL CLUSTERING

Giovanny Theotista[1*], Margareta Febe[2], Michael Sannova Ryan[3]

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

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

ABSTRACT

Article History:

Received: 30th April 2024
Revised: 25th August 2024
Accepted: 4th September 2024
Published: 13th January 2025

Keywords:

Hierarchical Clustering;
Indonesian Stock Market;
Ward Clustering.

This study aimed to address the instability of the Indonesian stock market from 2021 to 2023 by analyzing the LQ45 index, a critical indicator of economic robustness and corporate performance. Hierarchical Ward clustering was employed to categorize LQ45 stocks based on fundamental metrics such as Return, Volume, Price, Price-Earnings Ratio (PER), Earnings Per Share (EPS), and Dividends. Data preprocessing involved feature creation, Max-Abs scaling for normalization, and binary encoding of categorical variables. The optimal number of clusters was identified using dendrograms, revealing two primary clusters: one focusing on core materials and the other on financial services, alongside other industry-specific clusters. This method, characterized by its ability to minimize variance within clusters and determine natural groupings without predefined assumptions, provided valuable insights for financial advisors, policymakers, and investors. The findings offer practical guidance for optimizing decision-making, minimizing risks, and leveraging opportunities within the Indonesian stock market during a period of significant economic uncertainty. By employing this strategy, investors and traders can gain a comprehensive understanding of the current condition of the stock market, offering a thorough comprehension of the connections between equities and the operational and financial issues currently under scrutiny.

How to cite this article:

G. Theotista, M. Febe and M. S. Ryan., "ANALYSING MARKET DYNAMICS: REVEALING OBSCURED PATTERNS IN LQ45 STOCKS (2021-2023) USING WARD'S HIERARCHICAL CLUSTERING," *BAREKENG: J. Math. & App.*, vol. 19, iss. 1, pp. 0163-0172, March, 2025.

1. INTRODUCTION

The dynamic and influential status of the Indonesian stock market in the global financial arena has attracted the interest of analysts and investors around the globe. An essential element of this market is the LQ45 index, which tracks the performance of forty-five actively traded stocks on the IDX. In constructing this index, consideration is given to liquidity on the stock market, sound financial health, potential for growth, market capitalization, and frequency of trading, among other aspects. To ensure currency, offer significant insights into the nation's economy as a whole, and detect nascent patterns and investment prospects, the index's constituents are revised every six months [1].

Amidst the period of instability spanning from 2021 to 2023, the LQ45 index assumes a pivotal role in evaluating the economic robustness of Indonesia and the operational effectiveness of its largest corporations. It furnishes market participants with invaluable indicators as they navigate the Indonesian stock market [2]. The COVID-19 pandemic has further intensified global economic and political unrest, which has had substantial ramifications for the Indonesian stock market. This includes notable volatility in commodity prices, alterations in governmental strategies, and disturbances in the worldwide supply chain.

In the past, various methods have been employed to evaluate market instability, each with its own set of strengths and limitations. Traditional approaches such as volatility analysis using standard deviation and variance, Value at Risk (VaR), and econometric models like GARCH and ARIMA have provided insights into market behaviour by measuring fluctuations in stock prices and estimating potential losses under normal market conditions. However, these methods often fall short of capturing the full scope of market instability, particularly during periods of extreme economic events or structural changes. They tend to rely heavily on historical data and assume that past patterns will continue in the future, which may not always be true, especially in the face of unprecedented global crises like the COVID-19 pandemic. Additionally, these methods may not adequately account for non-linear relationships and tail risks, which are crucial for understanding market dynamics during times of heightened uncertainty. As a result, there is a growing interest in exploring alternative approaches, such as clustering analysis, which can offer a more flexible and data-driven means of identifying patterns and trends in complex financial environments.

Clustering analysis becomes a potent instrument in this particular context for deciphering intricate data patterns and correlations within the stock market [3]. Through the application of fundamental metrics including Return, Volume, Price, Price-Earnings Ratio (PER), Earnings Per Share (EPS), and Dividends, investors and analysts can acquire substantial knowledge regarding the complexities of the market [4].

Ward's hierarchical clustering is a widely used agglomerative clustering method that minimizes within-cluster variance at each merging step, producing compact and homogeneous clusters. Compared to other methods, Ward's method offers advantages such as creating balanced clusters, providing clear interpretability through dendrograms, and demonstrating robustness to noise. It also does not require predefining the number of clusters, unlike k-means. Foundational work by [5] laid the groundwork for this method, with further comprehensive overviews provided by [6], [7], and [8]. [9] say that Ward's hierarchical clustering is one of the simple clustering methods.

By applying the hierarchical Ward clustering technique to analyze LQ45 stocks from 2021 to 2023, this study offers a novel approach to understanding stock ownership and market activities amidst significant economic events. By incorporating a diverse range of financial indicators, including but not limited to Return, Volume, Price, Price-Earnings Ratio (PER), Earnings Per Share (EPS), and Dividends, the research provides a comprehensive perspective on stock grouping and behaviour during a period marked by notable economic shifts. This holistic understanding enhances the ability of financial advisors, policymakers, and investors to make informed decisions, navigate potential risks, and capitalize on opportunities within the Indonesian stock market landscape. In summary, this research project aims to pioneer a more nuanced understanding of the Indonesian stock market dynamics, facilitating more accurate investment decision-making in a dynamic financial environment.

2. RESEARCH METHODS

Ward's Agglomerative Clustering is a hierarchical clustering technique used to identify natural groupings within a dataset by minimizing the variance within each cluster at every step of the merging process. In this research, Ward's method is applied to analyze the stocks in the LQ45 index on the Indonesian Stock Exchange (IDX) from 2021 to 2023, utilizing financial indicators such as Return, Volume, Price, Price-Earnings Ratio (PER), Earnings Per Share (EPS), and Dividends. The process begins with each stock as its cluster and iteratively merges clusters that result in the smallest increase in the total within-cluster variance. This continues until all stocks are grouped into a single cluster, providing a comprehensive view of the stock market structure.

The advantage of using Ward's method lies in its ability to create balanced and homogeneous clusters, which enhances interpretability and provides clearer insights into stock performance and behaviour. Unlike other clustering methods, such as k-means, Ward's method does not require prior specification of the number of clusters, making it more flexible and adaptable to the data's natural structure. Additionally, the use of a dendrogram to visualize the clustering process helps determine the optimal number of clusters by examining where significant increases in variance occur. This makes Ward's Agglomerative Clustering particularly useful for analyzing complex financial data, allowing investors and analysts to identify underlying patterns and trends in the LQ45 stocks during a period of economic volatility.

From [10], [11], [12], and [13], we know that hierarchical clustering is a technique in cluster analysis that aims to group data into several clusters based on their level of similarity. Ward's Agglomerative clustering, a popular Hierarchical clustering method, integrates clusters based on similarity thresholds, starting with each data point as an independent cluster. The Ward technique is commonly employed in the field of aggregate clustering. Joe H. Ward created the Ward technique in 1963, which aims to minimize overall variation by clustering data [14]. This process is accomplished by minimizing the number of variations after clustering. Therefore, the fundamental goal of Ward is to provide uniformity inside each established cluster. The equation to calculate the distance between two clusters in the Ward method is as follows:

$$\Delta(A,B) = \sum_{i\in A\cup B}||x_i - m_{A\cup B}||^2 - \sum_{i\in A\cup B}||x_i - m_A||^2 - \sum_{i\in B}||x_i - m_B||^2 \quad (1)$$

Let $\boldsymbol{m_j}$ be the centre of the cluster $\boldsymbol{j}$ and $\Delta(\boldsymbol{A},\boldsymbol{B})$ be the difference between the Sum Square Error (SSE) of the combined cluster and the sum of SSE of the two clusters before merging.

2.1 Steps of The Research

The utilization of the Ward method in the agglomerative clustering algorithm generally encompasses the subsequent stages:

1. Initially, each data point is regarded as a cluster.
2. The process of computing distance matrices among all collaborative cluster partners.
3. The process of connecting two clusters that exhibit minimal overall variation after their merger.
4. The diagonal should be maximized while the prior gaps should be minimized until just one remaining cluster remains.

2.2 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	Code	Company name
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

No	Code	Company name
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 price, 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 date (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	1700	12,191,600	⋯	720	16,090,900
ADRO	Energy	910.9	1,059.87	1455	110,366,200	⋯	2380	124,776,700
AKRA	Energy	129.17	194.00	654	109,114,000	⋯	1475	34,881,900
AKRA	Energy	129.17	52.14	800	1,572,500	⋯	1475	34,881,900
AMRT	Consumer Defensive	79.2	134.98	2190	1,040,753,400	⋯	2930	28,228,000
ANTM	Basic Materials	168.19	0.00	3516.35	2,032,849	⋯	1705	83,597,900
ARTO	Financial Services	1.7	83.80	6225	34,840,200	⋯	2900	7,581,900
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
UNVR	Consumer Defensive	130.13	516.00	7475	8,203,000	⋯	3530	6,004,400

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

2.3 Stock Price

In the context of the stock market, the term "stock price" denotes the final value at which different shares are exchanged among the different participants in the market. The fluctuation of stock values is directly caused by the dynamic interaction between market demand and supply, which naturally exhibits volatility. This leads to the volatility of stock values. Analysing stock prices necessitates a thorough understanding of price fluctuations, the recognition of patterns like support and resistance, and the utilisation of additional technical indicators to make well-informed decisions. The utilisation of these components facilitates the decision-making process.

2.4 Stock Price Return s

Logarithmic returns also referred to as log returns, are commonly used in finance to measure the percentage change in the value of a financial instrument over a specific period. Log returns are frequently preferred over basic arithmetic returns in many financial models due to its mathematical properties, such as additivity and symmetry. **[15]**
The logarithmic return of a stock price can be calculated using the following equation:

$$R_{t+\Delta t,t} = \ln \frac{S_{t+\Delta t}}{S_t} \tag{2}$$

The generated number represents the logarithmic deviation of the stock's return during the chosen analysis period **[16]**.

The previously indicated computation provides a systematic measurement of the percentage change in the stock price during a specific time frame. Consequently, it is a highly valuable tool for examining financial data, especially in the realm of volatility modelling, risk reduction, and financial strategizing. **[17]**

2.5 Trading volume

We follow [18] and [19] to describe trading volume, which is a determinant that diminishes the quantity of currency transacted within a specified period. A reduction in interest may be indicated by a low volume, whereas an increase in interest in equities is often indicated by a large volume. Trading volume is a commonly employed method by technical analysts to validate train pricing and evaluate train potential.

2.6 EPS

According to [18] and [19], Earnings per share (EPS) is a metric used to quantify a company's total earnings. When assessing the viability of a company's ongoing profitability, the aforementioned figure is of paramount significance. Investors often view a persistent upward trend in earnings per share (EPS) as a favourable indication of future growth possibilities.

2.7 PER

According to [18] and [19], the price-to-earnings ratio (PER) pertains to the correlation between the prevailing stock price and earnings per share (EPS). This specific indicator bears considerable significance within the domain of fundamental financial analysis. The price-to-earnings ratio (PER) is employed as a metric for assessing the valuation of a stock, representing the valuation that an investor is willing to assign to each dollar of a company's shares. A high price-to-earnings ratio may suggest a significant valuation, whilst a low ratio may suggest a reduced valuation.

2.8 Dividend

According to [18] and [19], dividend is the regular distribution of a portion of a corporation's financial resources to its shareholders. A highly prevalent practice in the corporate realm is the dissemination of dividends, which involves the allocation of the company's profits. An escalating and steady stream of dividends may have a substantial effect on investors seeking a lucrative return on their investment. Moreover, dividends possess the capacity to significantly influence both the overall well-being and financial stability of a company.

3. RESULTS AND DISCUSSION

Data preparation and modelling techniques will be explored in this section using Python 3.1 and Excel from Microsoft. The pre-processing data are to create the desired feature, encoding, creating a new feature (PER), scaling, and feature selection. Furthermore, this encompasses the processes that are implemented to generate particular characteristics. After the data pre-treatment stage is over, Ward's Hierarchical Clustering is used to cluster the data. To make the most of the reference clusters, Python 3.1 employs the Dendrogram. The next steps, after the modelling process is finished, include analysing and interpreting the results and giving the investor suggestions.

3.1 Data Preprocessing: Feature Importance

In the first step of our process, we will calculate the Average Trading Volume by adding together the daily trading volume for the years 2021 through 2023 that we have collected. 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 $PER = \frac{Stock\ Price}{EPS}$. We use the price at date: 12/29/2023. The next step is to determine the average return on investment as well as the standard deviation of return on investment in order to evaluate the historical performance of the asset and the volatility of the asset. In order to account for the additive nature of log normal returns, daily returns will be computed using log normal returns.

3.2 Data Preprocessing: Binary Encoding

Binary encoding is a method used to convert categorical data into binary values. [20] The binary code, consisting of 0s and 1s and dispersed across numerous columns, is allocated to each category. The technique being discussed is appropriate for variables with a high degree of variability. Binary encoding is a method that can be divided into multiple separate stages. Initially, the relevant elements for the encoding process are determined. Subsequently, the subsequent action is allocating numerical values to each of the categories. To attain the desired length, these integers are later converted into binary representations. Finally, new columns are created to represent each bit in binary form.

3.3 Data Preprocessing: Scaling

Scaling is a procedure that decreases the volume of data to a more manageable size by restricting the potential range of data. [21] Normalization is sometimes used as a synonym for scaling. Specifically, clustering techniques such as Ward's Agglomerative Clustering can take advantage of this decrease in variability to generate more precise results. The Ward's Agglomerative Clustering technique is highly responsive to changes in the number of dimensions of the features, as it relies on Euclidean distance. This places it in a position that is more susceptible to these swings. When considering clustering, it is important to take into account the impact of traits with larger ranges. Consequently, scaling is necessary to ensure that each feature contributes equally. Before commencing the scaling procedure, it is important to do assessments on the attributes of the numerical data, encompassing the identification of any outliers. Boxplots are a method that can be used to identify outliers among other techniques.

Preprocessing is essential to ensure that the data is suitably prepared for modelling and analysis, hence facilitating the creation of accurate and valuable outcomes.

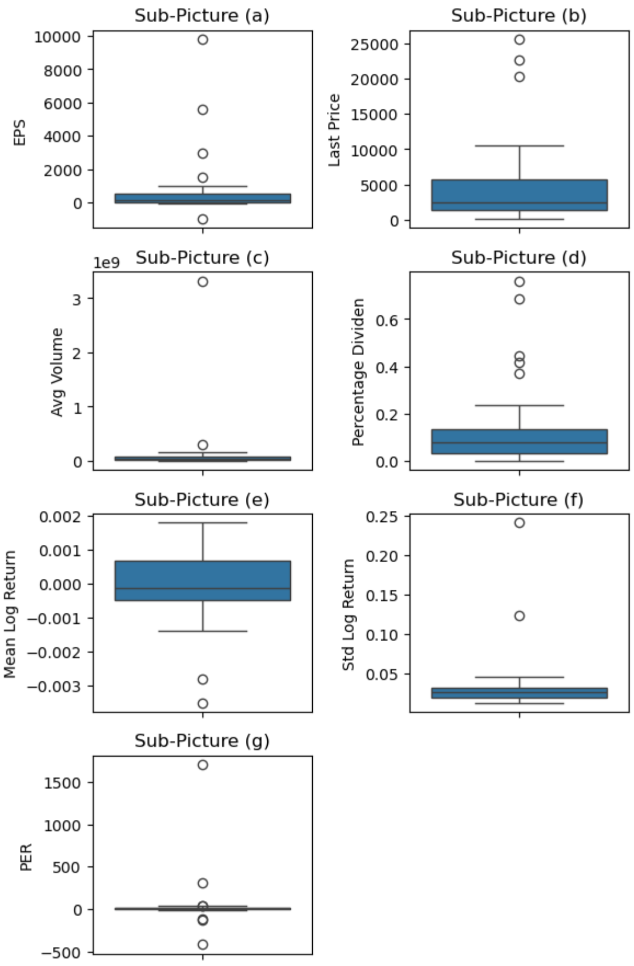

Figure 1. Boxplot Normality Test
(a) EPS, (b) Stock Price at the end of 2023, (c) Average Volume Trade, (d) Percentage Dividend, (e) Mean Lognormal ROI, (f) Standard Deviation Lognormal ROI, (g) PER

The method of analysis chosen was Absolute Maximum Scaling due to the presence of outliers in all of the data items. Absolute maximum scaling, also known as absolute normalisation or absolute maximum scaling, is a pre-processing technique where each value in a series is divided by the value that represents the highest value in that series. The objective is to normalise the data such that the highest absolute value in the series is either 1 or -1. This will be achieved by altering the data. This strategy is commonly used in the field of machine learning and data analysis to ensure that variables with varied scales do not disproportionately influence the output of certain algorithms.

The subsequent equation outlines the procedure for determining the absolute maximum scaling value:

$$x_i' = \frac{x_i}{|x_{max}|} \quad (3)$$

3.4 Ward's Agglomerative Clustering

This stage will include using ward linkage to perform agglomerative clustering. The analysis will utilise the following characteristics: earnings per share (EPS), end-of-year price (2023), average volume of purchases (2021-2023), percentage dividend (2021-2023), mean return (2021-2023), standard return (2021-2023), end-of-year price-to-earnings ratio (2023), and binary encoding of binary dummy variables. When using Agglomerative Clustering, the first step is to establish the optimal value of k, which is most accurately represented by a dendrogram.

Based on references [22], a dendrogram is a tree diagram used in hierarchical clustering analysis to illustrate how objects or data are grouped based on similarity. A dendrogram illustrates the hierarchical structure of data and how clusters are gradually formed and merged. The X-axis represents data or clusters, while the Y-axis represents the level of similarity or merging distance. By utilizing a dendrogram, the appropriate number of clusters can be determined by examining the level of significant merging distances.

The grouping is represented graphically by the dendrogram. It is often drawn backwards, beginning at similarity 0 and the last cluster including all of the items. The final cluster separates into its three parent clusters at the similarity where three clusters were combined to form it, and so on. This data representation is the most precise and reliable. Here is the outcome of the dendrogram visualisation:

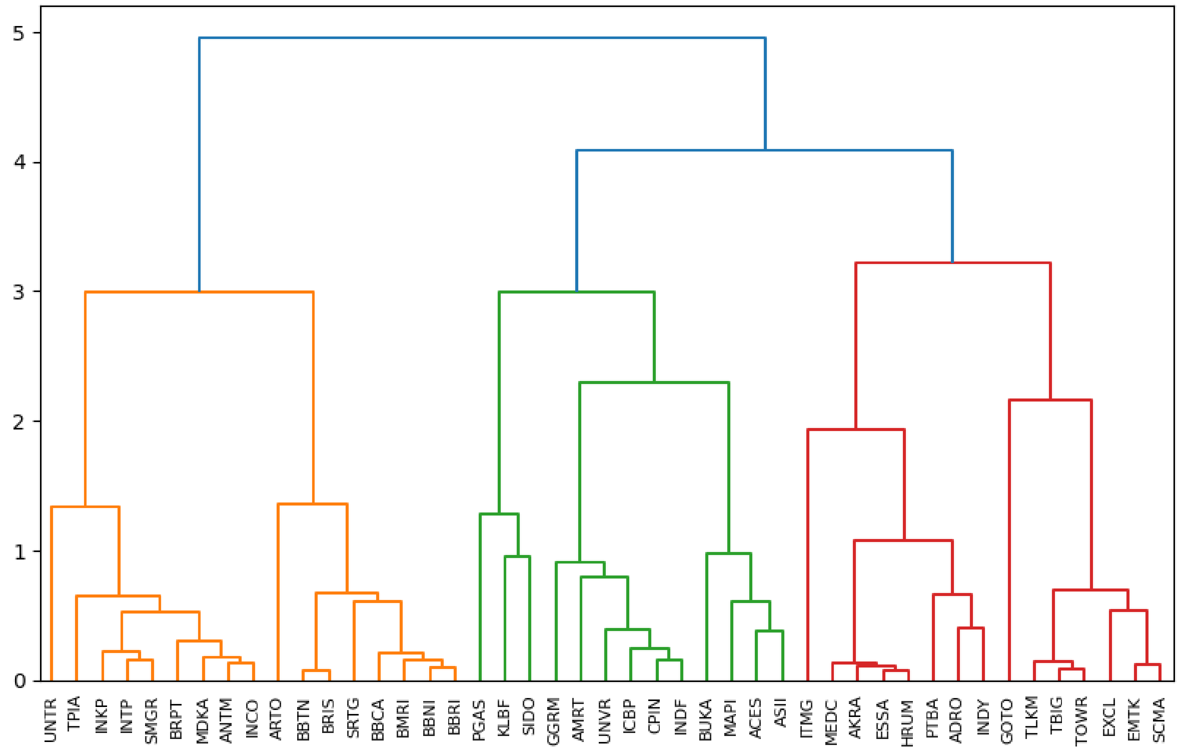

Figure 2. Dendrogram

Through the use of the dendrogram as represented in [23], we are able to examine the changes that have occurred in the clusters to ascertain the most appropriate value of k. Based on the dendrogram analysis, the optimal number of clusters is determined to be three, as there is a significant merging distance between the similarity levels of 5 and 4.0. This suggests that dividing the data into three clusters provides the most meaningful grouping, capturing the natural structure within the data. When using $k = 3$, the dataset is effectively segmented into three distinct clusters, each containing data points with high internal similarity. This clustering approach allows for a clearer understanding of the data, revealing important patterns and relationships that would otherwise be less apparent, and provides valuable insights for further analysis or decision-making.

Table 3. Ward's Cluster Each Company

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

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

Based on the research findings, the stock clusters reveal distinct sector-based groupings:

1. Cluster 0 primarily comprises stocks from the Energy, Communication Services, and Technology sectors. This cluster reflects a concentration in industries that are often associated with significant technological advancement and innovation. Stocks in these sectors may share common characteristics related to their growth potential, technological dependencies, and market dynamics. Investors interested in sectors with rapid technological changes and evolving consumer communication needs might find Cluster 0 particularly relevant.

2. Cluster 1 is predominantly characterized by stocks from the Financial Services and Basic Materials sectors. This grouping suggests a focus on industries involved in financial transactions, investment, and fundamental raw materials required for various industrial processes. Stocks in these sectors may exhibit traits such as sensitivity to economic cycles, regulatory impacts, and commodity price fluctuations. Investors looking for stability and fundamental investment opportunities might find Cluster 1 appealing due to its emphasis on essential financial services and materials.

3. Cluster 2 encompasses stocks from sectors that are not included in the aforementioned clusters, representing a diverse array of industries. This cluster includes companies from sectors with unique characteristics or emerging trends not yet fully captured by the other clusters. The inclusion of these stocks provides a broader perspective on market opportunities and risks, highlighting the importance of considering various industry sectors when analyzing investment portfolios.

4. CONCLUSIONS

In conclusion, by employing the hierarchical Ward clustering technique to analyze LQ45 stocks from 2021 to 2023, this study advances our understanding of stock behaviour and market activities during significant economic events. The application of diverse financial indicators, including Return, Volume, Price, Price-Earnings Ratio (PER), Earnings Per Share (EPS), and Dividends, offers a comprehensive view of stock groupings and their responses to economic shifts.

The findings reveal distinct clusters of stocks: Cluster 0 includes those from the Energy, Communication Services, and Technology sectors; Cluster 1 focuses on Financial Services and Basic Materials; and Cluster 2 comprises stocks from various other sectors. This clustering approach highlights sector-specific trends and behaviours, providing valuable insights for financial advisors, policymakers, and investors.

Understanding these clusters allows for more tailored investment strategies, better risk management, and enhanced decision-making. By navigating the Indonesian stock market's dynamics with this novel approach, stakeholders can more effectively capitalize on opportunities and address potential risks in a rapidly changing financial environment.

ACKNOWLEDGMENT

The research is funded by the 2023/2024 grant from Bunda Mulia University.

REFERENCES

[1] F. Idah, R. Siti, and R. Handayani, "THE INFLUENCE OF GOOD CORPORATE GOVERNANCE PRACTICE ON THE STOCK PRICE (Study on Company of LQ45 Index in Indonesia Stock Exchange during 2012-2016)," 2017.

[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. S. Bini and T. Mathew, "Clustering and Regression Techniques for Stock Prediction," *Procedia Technology*, vol. 24, pp. 1248–1255, 2016, doi: 10.1016/j.protcy.2016.05.104.

[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] J. H. Ward, "Hierarchical Grouping to Optimize an Objective Function," *J Am Stat Assoc*, vol. 58, no. 301, pp. 236–244, Mar. 1963, doi: 10.1080/01621459.1963.10500845.

[6] T. Hastie, J. Friedman, and R. Tibshirani, *The Elements of Statistical Learning*. New York, NY: Springer New York, 2001. doi: 10.1007/978-0-387-21606-5.

[7] L. Ferreira and D. B. Hitchcock, "A comparison of hierarchical methods for clustering functional data," *Commun Stat Simul Comput*, vol. 38, no. 9, pp. 1925–1949, Oct. 2009, doi: 10.1080/03610910903168603.

[8] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview, II," *WIREs Data Mining and Knowledge Discovery*, vol. 7, no. 6, Nov. 2017, doi: 10.1002/widm.1219.

[9] A. J. Parker and A. S. Barnard, "Selecting Appropriate Clustering Methods for Materials Science Applications of Machine Learning," *Adv Theory Simul*, vol. 2, no. 12, Dec. 2019, doi: 10.1002/adts.201900145.

[10] F. Nielsen, "Hierarchical Clustering," in *Introduction to HPC with MPI for Data Science*, Springer, 2016, pp. 195–211. doi: 10.1007/978-3-319-21903-5_8.

[11] P. K. Kimes, Y. Liu, D. N. Hayes, and J. S. Marron, "Statistical Significance for Hierarchical Clustering," *Biometrics*, vol. 73, no. 3, pp. 811–821, Sep. 2017, doi: 10.1111/biom.12647.

[12] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview, <scp>II</scp>," *WIREs Data Mining and Knowledge Discovery*, vol. 7, no. 6, Nov. 2017, doi: 10.1002/widm.1219.

[13] 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.

[14] N. Randriamihamison, N. Vialaneix, and P. Neuvial, "Applicability and Interpretability of Ward's Hierarchical Agglomerative Clustering With or Without Contiguity Constraints," *J Classif*, vol. 38, no. 2, pp. 363–389, Jul. 2021, doi: 10.1007/s00357-020-09377-y.

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

[16] G. Pernagallo and B. Torrisi, "An empirical analysis on the degree of Gaussianity and long memory of financial returns in emerging economies," *Physica A: Statistical Mechanics and its Applications*, vol. 527, Aug. 2019, doi: 10.1016/j.physa.2019.121296.

[17] 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.

[18] 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.

[19] 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

[20] 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.

[21] 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.

[22] Z. Zhang, F. Murtagh, S. Van Poucke, S. Lin, and P. Lan, "Hierarchical cluster analysis in clinical research with heterogeneous study population: highlighting its visualization with R," *Ann Transl Med*, vol. 5, no. 4, pp. 75–75, Feb. 2017, doi: 10.21037/atm.2017.02.05.

[23] M. Forina, C. Armanino, and V. Raggio, "Clustering with dendrograms on interpretation variables," *Anal Chim Acta*, vol. 454, no. 1, pp. 13–19, Mar. 2002, doi: 10.1016/S0003-2670(01)01517-3.