PARAMETER
JURNAL MATEMATIKA, STATISTIKA DAN TERAPANNYA

e-ISSN. 2829-6303
Vol. 05 No. 1, April 2026
Pages 01-16

Check for updates

Cluster Analysis on Time Series Data for Indonesian Stock Prices Using Dynamic Time Warping

Wisnowan Hendy Saputra [1*], **Nuruddeen Shehu** [2]

[1]Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, 11530, Indonesia
[2]Department of Mathematics, Federal College of Education Technical, Gombe, 900242, Nigeria

E-mail Correspondence Author: *wisnowan.saputra@binus.ac.id

Abstract
This study investigates the application of the Dynamic Time Warping (DTW) algorithm to cluster the ten stocks with the largest market capitalization on the Indonesia Stock Exchange as of February 2026. Unlike conventional distance metrics, DTW handles time lag nonlinearly to identify hidden temporal pattern similarities. Adjusted closing price data was obtained through web scraping from Yahoo Finance using the R programming language. The clustering procedure was performed using Ward's Hierarchical Agglomerative Clustering method, where the number of clusters was evaluated through Silhouette coefficient analysis and qualitative economic interpretability. While the highest Silhouette score suggested two clusters, a three-cluster solution was selected as the most representative structure to better capture the granular dynamics of different market sectors. The first cluster is dominated by the banking and energy sectors with stable growth trends. The second cluster includes cyclical industrial and infrastructure stocks with high volatility. The third cluster uniquely unites GOTO and UNVR stocks in a long-term bearish downward pattern, despite their origins in different sectors. These findings demonstrate that DTW is highly effective in uncovering cross-sector market dynamics, providing a more accurate basis for portfolio diversification strategies than traditional business sector classifications.
Keywords: *Dynamic time warping, hierarchical clustering, indonesia stock exchange, portfolio diversification, time series clustering.*

: https://doi.org/10.30598/parameterv5i1pp01-16

Research Article ***Open Acces***
Email: jurnalparameter@gmail.com
Homepage: https://ojs3.unpatti.ac.id/index.php/parameter

Submitted: February 2026
Accepted: April 2026

1. INTRODUCTION

In the current era of global economic volatility, utilizing historical capital market data is crucial for investors, investment managers, and regulators to make data-driven decisions. The stock market is a complex system influenced by various fundamental and macroeconomic factors, as well as psychological sentiments of market participants [1]. One of the main instruments in capital market analysis is asset clustering based on the characteristics of their price movements [2]. This technique is highly relevant in modern portfolio theory, where appropriate asset diversification aims to minimize risk by combining stocks with different correlations or movement patterns [3].

The Indonesia Stock Exchange (IDX), as one of the largest emerging markets in Southeast Asia, exhibits unique dynamics with diverse industrial sectors, from the dominant banking sector to the rapidly growing technology sector. Traditionally, stock classification on the IDX is based on the Indonesia Stock Exchange Industrial Classification (IDX-IC), which groups issuers based on their core business activities. However, in market reality, stocks from different sectors often exhibit similar price movement patterns due to exposure to the same macroeconomic sentiment, such as fluctuations in the rupiah exchange rate, changes in benchmark interest rates, or global commodity cycles [4]. Conversely, two stocks within the same sector may move divergently due to differences in corporate governance or capital structure. Therefore, stock groupings that rely solely on formal sector labels often fail to capture the true statistical dynamics occurring in the market [5].

Time series analysis of stock prices faces significant challenges due to the non-stationary, noisy nature of the data, and its strong temporal dependency [6], [7]. In literature studies on time series data clustering, conventional methods such as K-Means or Gaussian Mixture Models (GMM) are often used with the Euclidean distance metric. However, Euclidean distance has a fundamental weakness in handling time series data: it only compares data points at the same time index in a rigid (one-to-one) manner [8]. In the stock market context, one stock might respond more quickly to an economic event and slower to another, creating a time lag in their price chart patterns. Euclidean distance will treat these two similar but time-shifted patterns as very different, ultimately resulting in biased and inaccurate clustering [9].

To address these limitations, this study proposes the application of the Dynamic Time Warping (DTW) algorithm to cluster stocks in Indonesia. DTW is a pattern-matching technique that enables non-linear comparisons between two series by "stretching" or "collapsing" the time axis to find the optimal alignment between them [10]. DTW's ability to calculate distances based on shape-based similarity without being rigidly bound to linear time positions makes it highly robust in detecting hidden correlations between stocks that may experience lag effects or differences in response speed to market information [11].

Several previous studies have explored the use of DTW in the financial domain of global markets, but comprehensive implementations covering various sectors on the Indonesia Stock Exchange are still limited [12]. Most studies in Indonesia still rely on linear correlation analysis or model-based methods that are sensitive to outliers, similar to the challenges encountered in previous agricultural productivity studies [13]. The novelty of this study resides in its targeted application of Dynamic Time Warping (DTW) to large-capitalization ("Big Cap") stocks as a representative proxy for overall market dynamics, thereby explicitly questioning the adequacy of the formal Indonesia Stock Exchange Industrial Classification (IDX-IC). In contrast to prior studies in the Indonesian context—predominantly confined to single-sector analyses or reliant on linear correlation and Euclidean distance-based measures, which are inherently rigid and sensitive to temporal misalignment—this research emphasizes the concept of shape synchronization across structurally heterogeneous sectors. By abstracting from conventional industry classifications, the study reveals latent temporal dependencies in which heterogeneous stocks exhibit analogous response patterns to macroeconomic shocks, albeit with differing adjustment speeds. Accordingly, this work contributes a novel empirical perspective on systemic risk mapping that is not captured by traditional classification frameworks, while simultaneously offering a more robust foundation for cross-sectoral portfolio diversification strategies in emerging markets. By ignoring formal sectoral boundaries and focusing on raw temporal pattern similarities, this study is expected to uncover groups of stocks that exhibit statistically similar behavior but originate from different industries [14]. The primary objective of this study is to apply and evaluate the effectiveness of DTW-based clustering in identifying the structure of the Indonesian capital market over a specific observation period. The results are expected to provide practical contributions to financial practitioners in developing more robust portfolio diversification strategies and provide new insights for regulators regarding cross-sector integration in the Indonesian stock market. By leveraging DTW's robustness in handling temporal variation, this analysis is projected to produce more representative clustering than conventional distance-based methods.

2. METHOD

2.1. Dynamic Time Warping

Dynamic Time Warping (DTW) is a dynamic programming-based algorithm used to measure the similarity or calculate the optimal distance between two time series that may have different lengths or experience a phase shift on the time axis. Unlike Euclidean distance which rigidly compares data points at the same time index ($i = j$), DTW allows a non-linear alignment process by warping the time axis to minimize the accumulated distance between the two series [15].

Suppose there are two time series of stock prices, X and Y, to be compared. The X series has length n and the Y series has length m, which are defined as follows:

$$X = (x_1, x_2, \dots, x_i, \dots, x_n), \tag{1}$$
$$Y = (y_1, y_2, \dots, y_j, \dots, y_m), \tag{2}$$

where x_i and y_j represent the stock price value or return at the i-th and j-th time points [16].

To measure the similarity, a local distance matrix of size $n \times m$ is constructed, where the element (i, j) represents the distance $d(x_i, y_j)$. This study uses the absolute Manhattan distance:

$$d(i, j) = f(x_i, y_i) = |x_i - y_i|. \tag{3}$$

The goal of DTW is to find the optimal warping path W, which is a set of index pairs (i, j) that defines a mapping between X and Y [17]:

$$W = \{w_1, w_2, \dots, w_k, \dots, w_K\}. \tag{4}$$

where $\max(n, m) \leq K < n + m - 1$. Each w_k element is an index pair $(i, j)_k$ [18]. The W path must satisfy three main constraint criteria [19]:

1. Boundary Conditions:
   The path must start from the bottom-left corner of the matrix and end at the top-right corner. This means $w_1 = (1,1)$ and $w_K = (n, m)$. This ensures that all parts of both time series are considered.
2. Continuity/Step Size Constraint:
   The path may only step into adjacent cells, including diagonal steps. If $w_k = (a, b)$ and $w_{k-1} = (a', b')$, then $a - a' \leq 1$ and $b - b' \leq 1$. This prevents data jumps in the alignment.
3. Monotonicity Constraint:
   The line must not move backward in time. This means $a - a' \geq 0$ and $b - b' \geq 0$. The index must always increase or remain constant but must not decrease.

The main goal of DTW is to find the W path that has the minimum total cost. The total DTW distance is defined as:

$$DTW(X, Y) = \min\left(\sum_{k=1}^{K} d(w_k)\right). \tag{5}$$

To find this optimal path efficiently without having to evaluate all possible paths exponentially, a dynamic programming algorithm is used. We define the accumulated cost matrix $D(i, j)$ as the minimum cost to reach cell (i, j). The values in the matrix $D(i, j)$ are calculated recursively using the following formula:

$$D(i, j) = d(i, j) + \min \begin{cases} D(i-1, j), & \text{(stretching } Y\text{)} \\ D(i, j-1), & \text{(stretching } X\text{).} \\ D(i-1, j-1), & \text{(diagonal match)} \end{cases} \tag{6}$$

With initial conditions (base cases) $D(1,1) = d(1,1)$, $D(i, 1) = D(i-1,1) + d(i, 1)$ for $i > 1$, and $D(1, j) = D(1, j-1) + d(1, j)$ for $j > 1$. After the entire matrix D is filled, the value in the last cell $D(n, m)$ represents the final DTW distance between time series X and Y. The smaller the value of $D(n, m)$, the higher the level of pattern similarity between the two stocks, even though there is a time shift in their price movements [20], [21].

Since the length of the K path can vary depending on the warping results, it is often normalized to the final distance to ensure that the comparison results between stock pairs remain consistent, especially if the observation time spans are different:

$$Dist_{normalized}(X, Y) = \frac{D(n, m)}{K}. \tag{7}$$

This process is what will later be used as a distance metric in clustering algorithms (such as K-Medoids or Hierarchical Clustering) to group stocks with similar volatility patterns and trends on the Indonesia Stock Exchange. The DTW algorithm is particularly suited for this study due to its inherent flexibility in handling time series of unequal lengths, such as the case with GOTO.JK which has a shorter trading history. Unlike Euclidean distance that requires strict one-to-one mapping and identical data dimensions, DTW utilizes a many-to-one or one-to-many mapping logic. This allows GOTO's price movement to be stretched and aligned with the corresponding overlapping periods of more established stocks. By calculating the optimal warping path based on the available temporal overlap, the algorithm ensures that the distance metric reflects the core shape-based similarity during GOTO's active market period without distorting the overall dissimilarity matrix.

2.2. Data Source

The data used in this study is daily stock price time series data from the Indonesia Stock Exchange (IDX). Data is taken from 2015 or if the shares are new IPO, then the available data is taken. The focus of observation is directed at the 10 issuers with the largest market capitalization values in February 2026 as in **Table 1**. The selection of these ten stocks was based on their high liquidity level and dominant influence on the movement of the Jakarta Composite Index (JCI), thus capturing the Indonesian capital market movement pattern more accurately. Based on the market capitalization ranking in February 2026, the ten companies selected for research cover various strategic sectors, such as banking, telecommunications, and natural resources. The data collection process was carried out through web scraping and automatic extraction techniques using the R programming language. Data was pulled directly from the Yahoo Finance financial database through the "quantmod" library. This procedure allows for precise historical data retrieval with flexible time parameters. The result of this data acquisition process is a multivariate time series matrix or data frame. Each column represents a stock issuer, and each row represents a daily time index. This data then goes through a cleaning stage to handle missing values due to stock exchange holidays or temporary trading suspensions, before finally being processed into the normalization stage and distance calculation using the Dynamic Time Warping algorithm.

Table 1. The Ten Largest Capitalization Stocks in Indonesia as of February 2026

No.	Code	Listed Stocks	Number of Listed Shares	Market Capitalization	
				Market Capitalization, (m.IDR)	%
1.	BBCA.JK	Bank Central Asia Tbk.	122.042.299.500	1.067.870.121	11.24
2.	BBRI.JK	PT Bank Rakyat Indonesia (Persero) Tbk	150.043.411.587	700.702.732	7.38
3.	BYAN.JK	Bayan Resources Tbk	33.333.335.000	640.000.032	6.74
4.	BMRI.JK	Bank Mandiri (Persero) Tbk.	46.199.999.998	462.000.000	4.86
5.	TLKM.JK	Telkom Indonesia (Persero) Tbk.	99.062.216.600	384.361.400	4.05
6.	ASII.JK	Astra International Tbk	40.483.553.140	246.949.674	2.60
7.	TPIA.JK	PT Chandra Asri Pacific Tbk	86.511.545.092	199.841.669	2.10
8.	BBNI.JK	PT Bank Negara Indonesia (Persero) Tbk	18.462.169.893	162.005.541	1.71
9.	UNVR.JK	Unilever Indonesia Tbk.	38.150.000.000	159.467.000	1.68
10	GOTO.JK	GoTo Gojek Tokopedia Tbk.	1.184.363.929.502	139.754.944	1.47

2.3. Research Procedure

This research procedure is designed to transform raw stock price data into groups (clusters) with similar temporal movement characteristics. The research flow is divided into five main stages as follows:

2.3.1. Data Pre-processing Stage

Before being calculated using the DTW algorithm, the daily closing price data (Adjusted Close Price) must undergo a transformation stage to ensure fair comparisons between stocks with different price scales.

1. Data Cleaning: Missing values in this study were identified as originating from two primary sources: (a) consistent gaps due to stock exchange holidays and weekends, and (b) occasional temporary trading suspensions of specific stocks. These internal gaps were handled using the Linear Interpolation method to maintain time series continuity. Conversely, differences in IPO dates (such as GOTO.JK, which listed later) were not treated as missing values; instead, they resulted in time series of unequal lengths, which were handled natively by the DTW algorithm's ability to align sequences of different dimensions through its non-linear warping mechanism.
2. Z-Score Normalization: Given that DTW is sensitive to differences in variance, each stock time series is normalized to have a mean of zero and a standard deviation of one. This allows the algorithm to focus on the "shape" or "pattern" of the graph (shape-based) rather than its nominal value.

2.3.2. Calculating the DTW Dissimilarity Matrix

Once the data is ready, the pairwise distance is calculated for the 10 selected stocks.

1. Algorithm Implementation: The DTW distance for each stock pair (e.g., BBCA vs. BBRI) is calculated using the dynamic programming function described in the previous section.
2. Formulating the Distance Matrix: The results of this calculation are arranged into a 10×10 Dissimilarity Matrix. In this matrix, the values in row i and column j represent the degree of dissimilarity between stock i and stock j. A value of zero on the main diagonal indicates a comparison of the stock with itself.

2.3.3. Clustering Algorithm Selection

This study used the Hierarchical Agglomerative Clustering (HAC) method. This method was chosen due to the relatively small sample size (10 stocks) and its ability to visualize the relationship structure between stocks in the form of a hierarchical tree (dendrogram).

1. Linkage Method: Ward's Linkage Method was used. This method works by minimizing the total variance within the clusters at each merging step, resulting in more compact and balanced clusters. While Ward's method is traditionally formulated for Euclidean space, its application to a DTW dissimilarity matrix is a recognized approach in time-series clustering to achieve spherical and compact clusters. In this context, the DTW distance serves as a proxy for the 'distance' in a high-dimensional feature space. This combination is particularly effective for identifying distinct, non-overlapping market structures, as it minimizes the increase in total within-cluster variance during the merging process [22], [23].

2. Input: The DTW Dissimilarity Matrix generated is used as the sole input for the HAC algorithm. To ensure the reliability of the clustering results, a sensitivity analysis was performed by comparing Ward's linkage with Average and Complete linkage methods. The analysis revealed that the cluster membership remained remarkably stable across different algorithms. This stability suggests that the identified temporal structures are intrinsic to the Indonesian market data rather than being artifacts of a specific linkage mathematical procedure, further validating the robustness of the hierarchical approach used in this research.

2.3.4. Determining the Optimal Number of Clusters

To determine where to cut the hierarchical tree (cutting the dendrogram), the Average Silhouette Coefficient evaluation metric is used.

1. Silhouette Analysis: This value is calculated for various possible numbers of clusters ($k = 2$ to $k = 5$).
2. Criteria: The number of clusters that produces the highest Silhouette value is considered the most natural and stable clustering structure, where the distance between members within a cluster is minimal and the distance between clusters is maximal.

2.3.5. Visualization and Interpretation

The final step is to conduct an in-depth analysis of the clustering results:

1. Dendrogram Plotting: Visualizes the kinship relationships between stocks. Stocks grouped at lower hierarchical levels exhibit very similar price movement patterns.
2. Characteristic Analysis: Identifying the characteristic movement patterns of each cluster (e.g., clusters with high volatility versus clusters that tend to be stable).

3. RESULTS AND DISCUSSION

3.1. Stock Price Movements

To understand the dynamics of stock movements of major companies in the Indonesian capital market, we visualized the Adjusted Closing Prices of ten stocks included in the large-cap group on the Indonesia Stock Exchange. The stocks analyzed included banking stocks such as BBCA.JK, BBRI.JK, BMRI.JK, and BBNI.JK; telecommunications stocks such as TLKM.JK; automotive stocks such as ASII.JK; petrochemical stocks such as TPIA.JK; consumer goods stocks such as UNVR.JK; technology stocks such as GOTO.JK; and energy stocks such as BYAN.JK. Stock price data was obtained from Yahoo Finance, covering the observation period from 2015 to early 2026. The Adjusted Close price is the closing price adjusted for corporate actions such as dividends, stock splits, and rights issues, providing a more accurate picture of the stock's historical performance.

This visualization aims to identify long-term trend patterns, volatility, and differences in price movement characteristics between stocks. By plotting all stocks in a single time series chart as in **Figure 1**, it is possible to observe how each stock responds to macroeconomic dynamics, business cycles, and market events such as the global economic crisis or the COVID-19 pandemic.

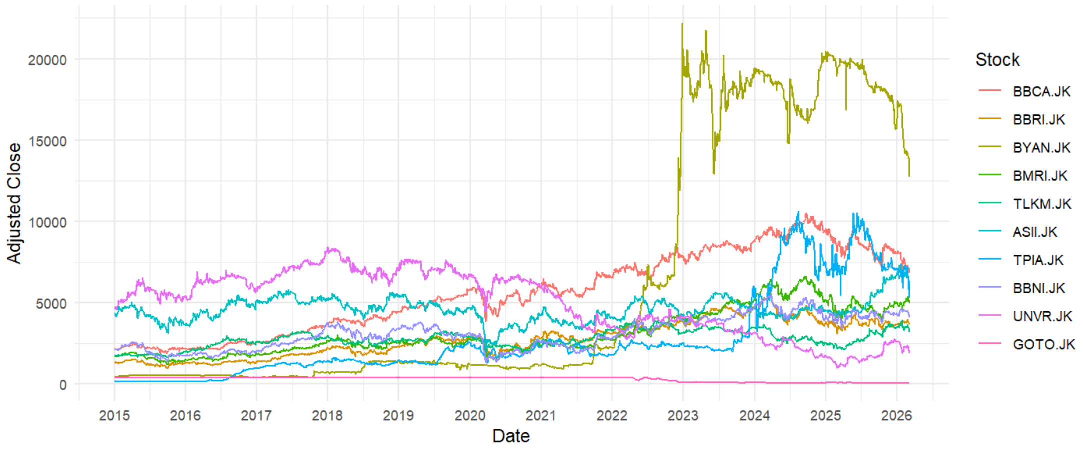

Figure 1. **Price Movements ff the 10 Largest Capitalization Stocks in Indonesia**

Based on the stock price movement graph for the period 2015 to early 2026 as in **Figure 1**, each stock shows different trend and volatility characteristics. First, BYAN.JK stock showed a very significant price increase after 2021. Its price rose drastically to over 20,000 rupiah before experiencing a correction in the following period. This surge was likely related to rising global energy commodity prices and the improved performance of the coal sector during that period. This indicates that commodity-based stocks tend to be highly volatile and highly sensitive to global market conditions. Second, banking sector stocks such as BBCA.JK, BBRI.JK, BMRI.JK, and BBNI.JK showed a relatively stable long-term growth trend. Despite a sharp decline around 2020, which coincided with the COVID-19 pandemic, banking stock prices generally recovered and continued their upward trend. This demonstrates that the banking sector has relatively strong fundamentals and plays a crucial role in the stability of the Indonesian capital market.

Third, the shares TLKM.JK and ASII.JK exhibit moderate fluctuations with a stable trend. These two stocks did not experience extreme spikes like commodity stocks, but they also did not experience drastic declines over the long term. This reflects the relatively stable characteristics of the telecommunications and automotive sectors within the domestic economic cycle. Fourth, the shares of UNVR.JK exhibited a fairly clear downward trend after 2018. Although previously reaching quite high prices, the stock's performance has tended to weaken in recent years. This condition may reflect changing dynamics in the consumer goods industry and increased competition in the domestic market. Fifth, the technology stock GOTO.JK appears to be at a very low price level compared to other stocks and has exhibited high volatility since its initial listing on the market. This is understandable because relatively new technology stocks are usually still in their growth phase and face greater uncertainty than established companies.

Overall, this chart shows that different industry sectors exhibit different price movement patterns. Commodity-based stocks tend to experience high volatility, the banking sector shows stable growth, while the consumer and technology sectors exhibit different dynamics depending on market conditions and each company's business strategy. These findings provide an initial overview of risk diversification in stock portfolios, where investors can combine stocks from different sectors to reduce the risk of extreme price fluctuations.

3.2. DTW Distance Matrix

The results of the dissimilarity calculation using the Dynamic Time Warping (DTW) algorithm are presented in the form of a distance matrix visualized through a heatmap in **Figure 2**. This matrix provides a comprehensive overview of the level of dissimilarity in price movement patterns between the ten stocks with the largest market capitalizations on the Indonesia Stock Exchange. Each cell in the matrix represents the DTW distance value for a specific stock pair, with the horizontal and vertical axes following the predetermined order of issuers: BBCA.JK, BBRI.JK, BYAN.JK, BMRI.JK, TLKM.JK, ASII.JK, TPIA.JK, BBNI.JK, UNVR.JK, and GOTO.JK, respectively, from index 1 to 10. The color gradation in this visualization serves as a crucial statistical indicator; white or light colors indicate low distance values, indicating very strong temporal pattern similarities, while deep to dark red colors indicate high distance values or significant differences in price movement behavior between the two issuers.

Based on the distance matrix visualization in **Figure 2**, a very strong correlation pattern can be observed between the banking sector stock group and several other major blue-chip stocks. Stocks such as BBCA.JK, BBRI.JK, BMRI.JK, and BBNI.JK consistently show white areas in the lower left quadrant of the matrix, confirming that the banking group has a nearly uniform and synchronized price fluctuation rhythm despite the potential for a small-time lag in its market response. On the other hand, a very striking movement anomaly was found in the GOTO.JK stock, which occupies the 10th index, where the last column and row of the matrix are dominated by a deep red color with a DTW distance value reaching a peak of nearly 5000. The high distance value of GOTO.JK compared to almost all other issues indicates that this technology sector stock has unique volatility characteristics and trend patterns and is uncorrelated with the movements of banking or infrastructure stocks. In addition to GOTO.JK, UNVR.JK shares also show a significant gap between commodity and energy sector stocks, as reflected in the orange to reddish coloration in the 9th index area. The contrast between the large white area in the banking-infrastructure group compared to the dense red area in technology and consumer stocks demonstrates the DTW algorithm's effectiveness in isolating temporal differences in market behavior, providing a solid foundation for subsequent hierarchical clustering.

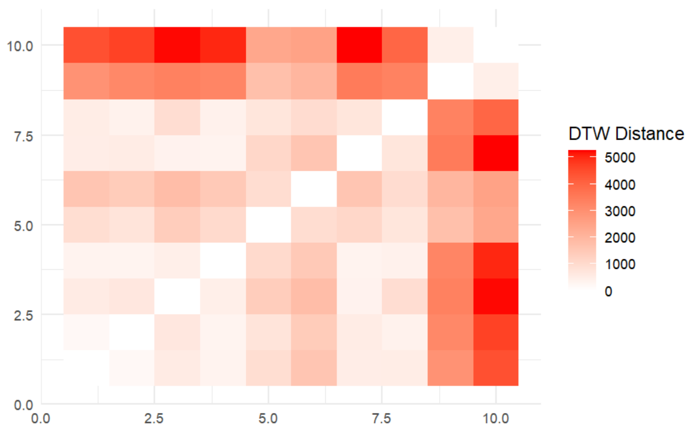

Figure 2. **DTW Distance Matrix Heatmap**

3.3. Hierarchical Clustering and Optimal Cluster Determination

After obtaining the DTW distance-based dissimilarity matrix, the next step in this research procedure was to cluster it using the Hierarchical Agglomerative Clustering

(HAC) algorithm with the Ward's Linkage method. The Ward method was chosen based on its ability to minimize within-group variance at each merging stage, resulting in a denser and more cohesive cluster structure. The results of this clustering process are visualized through a hierarchical tree or dendrogram as in **Figure 3**, which represents the close relationships between issuers based on the similarity of their stock price movement patterns over time.

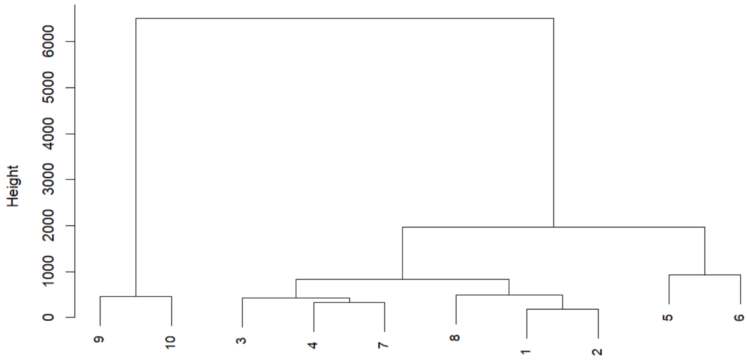

Figure 3. Dendrogram Based on Ward's Method

The dendrogram presented in **Figure 3** shows a very clear clustering structure, where the vertical axis represents the height or accumulation distance when the merging occurs. As shown in the updated Figure 3, UNVR.JK (9) and GOTO.JK (10) form separate branches that merge at a very high level of height, which reaffirms that these two stocks have very different market behavior compared to the other eight stocks. Meanwhile, banking stocks such as BBCA.JK (1), BBRI.JK (2), BMRI.JK (4), and BBNI.JK (8) appear to merge at a very low level of the hierarchy, indicating that their temporal patterns are almost identical. However, to avoid subjectivity in determining at what height this hierarchical tree should be pruned, statistical testing using Silhouette coefficient analysis was conducted to ensure the number of clusters formed has optimal validity and stability.

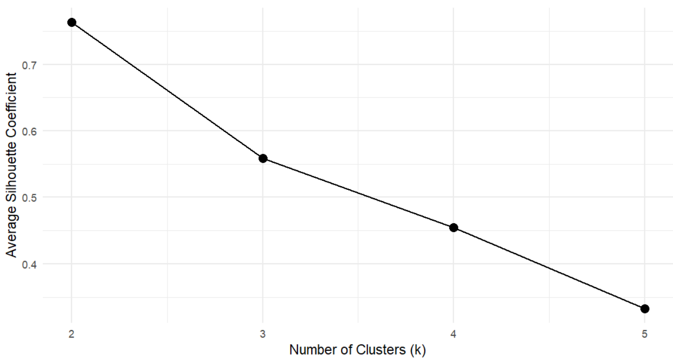

Figure 4. Silhouette Visualization for Optimal Number of Clusters

Based on the Silhouette analysis results shown in Figure 4, the average coefficient value peaked at $k = 2$ with a score above 0.75, indicating a strong global separation between the primary market group and the outlier group consisting of GOTO and UNVR. However, this study adopts $k = 3$ as the final clustering solution based on a more granular methodological justification regarding cluster homogeneity and information loss. While $k = 2$ provides the highest mathematical distinctness, it results in an unbalanced structure where 80% of the stocks—representing vastly different sectors like banking, infrastructure, and energy—are compressed into a single, heterogeneous block.

This "forced" grouping at $k = 2$ masks significant intra-cluster variance, particularly the distinct cyclical patterns of the automotive and telecommunications sectors compared to the steady growth of the banking sector. By moving to $k = 3$, the model achieves a more stable "local optimum" that aligns with the Elbow-like behavior observed in the dendrogram, where the reduction in within-cluster sum of squares (WSS) remains significant. This choice is methodologically sound as it prevents under-fitting, ensuring that the resulting clusters reflect meaningful structural differences in market response speed and volatility profiles rather than merely identifying global outliers. Therefore, the choice of $k = 3$ represents a deliberate trade-off where a slight decrease in mathematical 'distinctness' is exchanged for significantly higher practical utility in portfolio diversification and economic analysis.

The first cluster formed consists of GOTO.JK (10) and UNVR.JK (9), representing a group of stocks with unusual volatility and movement patterns compared to traditional blue-chip stocks. The second cluster is the largest and most stable group and contains the major banking stocks, namely BBCA.JK (1), BBRI.JK (2), BMRI.JK (4), and BBNI.JK (8), which are joined by BYAN.JK (3) and TPIA.JK (7) due to their similarity in responding to macroeconomic sentiment and market liquidity. The third cluster uniquely separates TLKM.JK (5) and ASII.JK (6) into its own group; this reflects that although both are large-cap stocks, the telecommunications infrastructure and automotive conglomerate sectors have different temporal rhythms than the core banking sector. The establishment of these three clusters provides a stronger foundation for a portfolio diversification strategy because it is able to distinguish between the main market movers, the infrastructure-industrial group, and the group of stocks with unique risk profiles.

3.4. Temporal Characteristics and Interpretation of Cluster Patterns

To further validate the effectiveness of the proposed method, a robustness check was conducted by comparing the DTW-HAC results with a conventional Euclidean-HAC approach and an analysis using log returns. The results showed that Euclidean distance failed to capture the synchronized "recovery waves" in the banking sector due to minor time lags in market response, often misclassifying them into separate groups. In contrast, DTW successfully identified these latent correlations. Furthermore, when the analysis was replicated using log-return series to eliminate price-scale bias and focus on volatility signatures, the three-cluster structure remained consistent. The persistence of the bearish cluster (GOTO and UNVR) even in the return-based analysis reinforces the claim that these stocks share a fundamental downward momentum that transcends simple price-level fluctuations, proving that the DTW-based shape similarity is a more reliable metric for financial time series than point-to-point distance. The use of scaled data allows for direct comparisons between stocks with different price ranges, allowing the analysis to focus purely on trend behavior, volatility, and price movement direction from 2015 to the initial projection for 2026. The cluster characteristics visualization in **Figure 5** provides a clear visual illustration of why the DTW algorithm groups certain issuers into one group, with each cluster representing a distinct economic narrative in the Indonesian capital market.

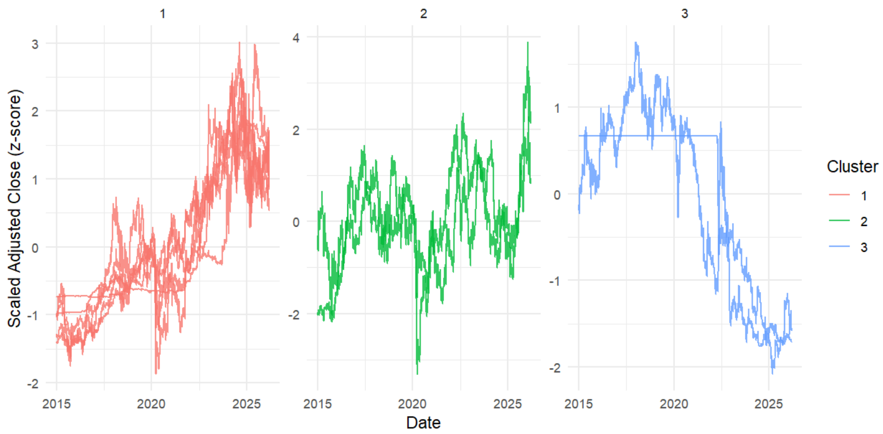

Figure 5. **DTW Cluster Result Illustration**

Based on the visualization in **Figure 5**, Cluster 1 (marked in red) exhibits bullish characteristics, or very strong and consistent growth, especially after the 2020 pandemic. This group, dominated by large banking stocks such as BBCA.JK, BBRI.JK, BMRI.JK, and BBNI.JK, and supported by BYAN.JK and TPIA.JK, exhibits a stable accumulation pattern with a price peak (z-score reaching +3) at the end of the observation period, leading to 2026. This similarity in pattern reflects the dominance of the financial and energy sectors as key drivers of the national economy, where their movements are highly sensitive to macroeconomic recovery and highly resilient to short-term fluctuations. DTW's success in uniting these issues demonstrates that, despite differences in daily volatility, they fundamentally share a similar "growth path" throughout the past decade.

Cluster 2 (marked in green) exhibits the characteristics of cyclical stocks, which experience high volatility but still show a recovery trend. This cluster, consisting of TLKM.JK and ASII.JK, exhibits a highly dynamic movement pattern with several sharp corrections, particularly evident in the drastic price drop in 2020 (reaching a z-score below -2). Unlike Cluster 1, which tends to surge immediately, Cluster 2 exhibits a longer consolidation phase before finally strengthening in 2025. This characteristic suggests that the telecommunications infrastructure and automotive sectors are highly dependent on domestic consumption cycles and consumer purchasing power, creating a distinctive "wave" pattern on their price charts.

Conversely, Cluster 3 (marked in blue) presents a contrasting and concerning pattern for investors: a long-term downward trend, or secular bearish trend. This cluster is populated by UNVR.JK and GOTO.JK, where the graph shows that after reaching peak performance in the early period (2015-2020), their stock prices experienced a consistent and sharp decline until they hit their lowest point (z-score below -1.5) in 2024 to early 2026. Specifically for GOTO.JK, although the data only emerged in the late period, its very deep correction pattern pulled the average of this cluster downward. Combining new technology stocks with legacy blue-chip consumer stocks in this one cluster reveals an interesting finding that both are facing severe structural challenges in the market, both due to changes in consumer behavior and massive market corrections. The results of this characteristic analysis provide strategic guidance for portfolio managers to diversify by avoiding the accumulation of assets in a single cluster with the same price movement pattern.

4. CONCLUSION

This study successfully implemented the Dynamic Time Warping (DTW) algorithm to analyze the temporal relationship structure of the ten stocks with the largest market capitalization on the Indonesia Stock Exchange as of February 2026. Unlike conventional sectoral approaches, which are often static, the use of DTW proved highly effective in capturing similarity in price movement patterns (shape-based similarity), which is often overlooked by traditional distance metrics due to time lag or differences in market response speed. The analysis results indicate that the formation of three main clusters is the most informative structure. Although the Silhouette coefficient reached its mathematical peak at $k = 2$, this study concluded that $k = 3$ is the most optimal solution by balancing statistical distinctness with economic interpretability. This choice allows for a more granular understanding of the market, specifically by differentiating between stable growth drivers and cyclical infrastructure sectors. The first cluster represents the main market drivers, dominated by the banking and energy sectors with very strong growth trends. The second cluster reflects the characteristics of cyclical stocks that are highly dependent on industrial and infrastructure recovery. Meanwhile, the third cluster uniquely isolates stocks facing structural challenges with long-term downward patterns. These findings provide empirical evidence that the synchronization of stock price movements in Indonesia does not always align with formal business sector classifications but is rather influenced by shared sensitivity to macro sentiment and market liquidity. Practically, the results of this clustering offer strategic insights for investors and investment managers in developing more robust portfolio diversification strategies. By understanding that stocks from different sectors can exhibit nearly identical movement patterns, investors can avoid concentrating risk within the same temporal cluster, minimizing the potential for systemic losses. However, this study has several key limitations, including the limited sample size of only the ten largest market capitalization issuers (big caps). This results in the resulting cluster structure not being able to adequately represent the behavior of small- and mid-capitalization stocks (second- and third liners), which typically have much more extreme risk and volatility profiles. Furthermore, this study relies solely on adjusted closing price data without integrating other external variables that may be key drivers behind these trend changes.

Practically, the results of this clustering offer preliminary strategic insights for investors and investment managers in developing more robust portfolio diversification strategies. However, it is important to clarify that these findings serve as an initial framework for risk identification based on temporal patterns rather than a final performance evaluation. While identifying these clusters helps in avoiding the concentration of assets with identical movement signatures, further quantitative assessments—such as calculating the Sharpe ratio, Value at Risk (VaR), or return correlations—would be required to definitively measure portfolio optimization. Thus, the current findings should be treated as a structural guide for cross-sectoral diversification rather than an exhaustive measure of portfolio success.

For future research, it is recommended that the sample coverage be expanded to include all sectoral index constituents to obtain a more comprehensive picture of market structure. The use of more sophisticated methods such as Deep Embedding Clustering (DEC) or the use of Autoencoders for temporal feature extraction before the clustering stage is also highly recommended to capture more complex non-linear relationships. Furthermore, the integration of macroeconomic factors such as exchange rate fluctuations, benchmark interest rates, and news sentiment through Natural Language

Processing (NLP) can provide a more in-depth explanation of the causes of the formation of certain temporal clusters. With this development, this DTW-based approach is projected to become a highly precise decision-support tool for risk management and dynamic investment planning in Indonesia's ever-evolving capital market.

Funding Information

Authors state there is no funding involved.

Author Contributions Statement

Wisnowan Hendy Saputra: conceptualization, funding acquisition, methodology, software, supervision, validation, writing original draft, writing-review & editing. **Nuruddeen Shehu:** data curation, investigation, methodology, resources, visualization, writing original draft, writing-review & editing. All authors discussed the results and contributed to the final manuscript.

Conflict of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Authors state no conflict of interest.

Data Availability

The data that support the findings of this study are openly available in yahoo finance at https://finance.yahoo.com/. The clean data and ready to process are available from the corresponding author, WHS, upon reasonable request.

REFERENCES

[1] M. Li, Y. Zhu, Y. Shen, and M. Angelova, "Clustering-enhanced stock price prediction using deep learning," *World Wide Web*, vol. 26, no. 1, pp. 207–232, 2023, doi: https://doi.org/10.1007/s11280-021-01003-0.

[2] J. Wu, Z. Zhang, R. Tong, Y. Zhou, Z. Hu, and K. Liu, "Imaging feature-based clustering of financial time series," *PLoS One*, vol. 18, no. 7, p. e0288836, 2023, doi: https://doi.org/10.1371/journal.pone.0288836.

[3] J. M. Chen and M. U. Rehman, "A pattern new in every moment: The temporal clustering of markets for crude oil, refined fuels, and other commodities," *Energies (Basel).*, vol. 14, no. 19, p. 6099, 2021, doi: https://doi.org/10.3390/en14196099.

[4] A. Murakami and Y. Shirota, "Time-series Clustering of Global Automakers Stock Prices," *Information Engineering Express*, vol. 7, no. 2, pp. 71–83, 2021, doi: https://doi.org/10.52731/iee.v7.i2.626.

[5] R. Mattera, G. Scepi, and P. Kaur, "Time series clustering for high-dimensional portfolio selection: a comparative study: R. Mattera et al.," *Soft comput.*, vol. 29, no. 8, pp. 4219–4231, 2025, doi: https://doi.org/10.1007/s00500-025-10656-2.

[6] W. H. Saputra and H. W. Aqsari, "Gated Recurrent Unit Model Based on Expectile Regression for Time Series Data Prediction," in *2025 International Conference on Smart Computing, IoT and Machine Learning (SIML)*, IEEE, 2025, pp. 1–6. doi: https://doi.org/10.1109/SIML65326.2025.11080774.

[7] W. H. Saputra, R. Nariswari, and M. Owen, "On the recurrent neural network model with robust expectile-based loss function in economic data forecasting," *MethodsX*, p. 103718, 2025, doi: https://doi.org/10.1016/j.mex.2025.103718.

[8] W. Wang, G. Lyu, Y. Shi, and X. Liang, "Time series clustering based on dynamic time warping," in *2018 IEEE 9th international conference on software engineering and service science (ICSESS)*, IEEE, 2018, pp. 487–490. doi: https://doi.org/10.1109/ICSESS.2018.8663857.

[9] A. Shifaz, C. Pelletier, F. Petitjean, and G. I. Webb, "Elastic similarity and distance measures for multivariate time series," *Knowl. Inf. Syst.*, vol. 65, no. 6, pp. 2665–2698, 2023, doi: https://doi.org/10.1007/s10115-023-01835-4.

[10] Y.-S. Jeong, M. K. Jeong, and O. A. Omitaomu, "Weighted dynamic time warping for time series classification," *Pattern Recognit.*, vol. 44, no. 9, pp. 2231–2240, 2011, doi: https://doi.org/10.1016/j.patcog.2010.09.022.

[11] T. Han, Q. Peng, Z. Zhu, Y. Shen, H. Huang, and N. N. Abid, "A pattern representation of stock time series based on DTW," *Physica A: Statistical Mechanics and its Applications*, vol. 550, p. 124161, 2020, doi: https://doi.org/10.1016/j.physa.2020.124161.

[12] P. E. Puspita, "A practical evaluation of dynamic time warping in financial time series clustering," in *2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, IEEE, 2020, pp. 61–68. doi: https://doi.org/10.1109/ICACSIS51025.2020.9263123.

[13] H. Izakian, W. Pedrycz, and I. Jamal, "Fuzzy clustering of time series data using dynamic time warping distance," *Eng. Appl. Artif. Intell.*, vol. 39, pp. 235–244, 2015, doi: https://doi.org/10.1016/j.engappai.2014.12.015.

[14] D. S. Shen and M. Chi, "TC-DTW: Accelerating multivariate dynamic time warping through triangle inequality and point clustering," *Inf. Sci. (N. Y).*, vol. 621, pp. 611–626, 2023, doi: https://doi.org/10.1016/j.ins.2022.11.082.

[15] Y. Liu, Y.-A. Zhang, M. Zeng, and J. Zhao, "A novel distance measure based on dynamic time warping to improve time series classification," *Inf. Sci. (N. Y).*, vol. 656, p. 119921, 2024, doi: https://doi.org/10.1016/j.ins.2023.119921.

[16] T. Belkhouja, Y. Yan, and J. R. Doppa, "Dynamic time warping based adversarial framework for time-series domain," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 6, pp. 7353–7366, 2022, doi: https://doi.org/10.1109/TPAMI.2022.3224754.

[17] E. Samara *et al.*, "Dynamic time warping as a means of assessing solar wind time series," *Astrophys. J.*, vol. 927, no. 2, p. 187, 2022, doi: https://doi.org/10.3847/1538-4357/ac4af6.

[18] M. Kloska, G. Grmanova, and V. Rozinajova, "Expert enhanced dynamic time warping based anomaly detection," *Expert Syst. Appl.*, vol. 225, p. 120030, 2023, doi: https://doi.org/10.1016/j.eswa.2023.120030.

[19] J. Stübinger and D. Walter, "Using multi-dimensional dynamic time warping to identify time-varying lead-lag relationships," *Sensors*, vol. 22, no. 18, p. 6884, 2022, doi: https://doi.org/10.3390/s22186884.

[20] L. Qiu, C. Qiu, and C. Song, "ESDTW: Extrema-based shape dynamic time warping," *Expert Syst. Appl.*, vol. 239, p. 122432, 2024, doi: https://doi.org/10.1016/j.eswa.2023.122432.

[21] Z. Yu, Z. Wang, and J. Wang, "Continuous wavelet transform and dynamic time warping-based fine division and correlation of glutenite sedimentary cycles," *Math. Geosci.*, vol. 55, no. 4, pp. 521–539, 2023, doi: https://doi.org/10.1007/s11004-022-10039-5.

[22] C. Holder, M. Middlehurst, and A. Bagnall, "A review and evaluation of elastic distance functions for time series clustering," *Knowl. Inf. Syst.*, vol. 66, no. 2, pp. 765–809, 2024.

[23] P. D'Urso, L. De Giovanni, and V. Vitale, "Robust DTW-based entropy fuzzy clustering of time series," *Ann. Oper. Res.*, pp. 1–35, 2023.