Eurasian Research Unit
Eurasian Campus - Sri Lanka

Journal of Eurasian Social Sciences (JESS)

Hierarchical and Dynamic Time Warping-Based Unsupervised Clustering of S&P SL20 Companies in the Colombo Stock Exchange, Sri Lanka

***[1]Ruwantha Perera, [2] Dulanji Samarakoon, [3] Dilushika Anton Michael, [4] Yasmi Rajapaksha, [5] Kavindi Vinurika, [6] Gihashi Liyanage**

[1,2,3,4,5,6.] Department of Statistics & Computer Science, Faculty of Science, University of Kelaniya, Sri Lanka

ARTICLE INFO

Article History

Received: 31st March 2026
Revised: 25th April 2026
Published Online: 1st July 2026

Keywords:

Colombo Stock Exchange (CSE), Dynamic Time Warping, Hierarchical Clustering, S&P SL20 Index, Unsupervised Learning

***Corresponding Author**

Email: ruwanthap2000@gmail.com

How to cite: *Perera, R., Samarakoon, D., Michael, D.A., Rajapaksha, Y., Vinurika, K. and Liyanage, G. (2026) 'Hierarchical and Dynamic Time Warping-Based Unsupervised Clustering of S&P SL20 Companies in the Colombo Stock Exchange, Sri Lanka', Journal of Eurasian Social Sciences, 1(1), pp. 67–74.*

ABSTRACT

The Colombo Stock Exchange plays a crucial role in Sri Lanka's financial system. However, the conventional industry-based classification of S&P SL20 companies may fail to capture deeper relationships within the market. This study introduces a data-driven framework to reveal non-traditional sector groupings using Dynamic Time Warping and hierarchical clustering. Daily trading data from S&P SL20 constituents, from August 2024 to July 2025, were collected and pre-processed to derive key features including daily returns, volatility, momentum, and relative volume. Each series was normalised to ensure comparability. DTW, which aligns time series with non-linear temporal variations, was applied to capture similarities in market dynamics across companies, and agglomerative hierarchical clustering with average linkage was used to form clusters. Internal validation through Silhouette, Davies–Bouldin, and Calinski-Harabasz indices revealed robust clustering structures. A two-cluster solution denoted Richard Pieris as an outlier, while a five-cluster solution uncovered more nuanced relationships, grouping firms such as Vallibel One, Central Finance, LB Finance, and Hemas Holdings despite differing industry labels. Sampath and Commercial Banks showed strong temporal co-movements, whereas Lanka Lubricants and RCL exhibited distinctly unique behaviours. These findings imply that stock price co-movements in Sri Lanka are shaped not only by sectoral affiliation but also by broader forces such as liquidity shocks, momentum trends, and investor sentiment. By identifying hidden interconnections, this research contributes a novel perspective to Sri Lankan financial analysis and complements existing studies that focus on volatility or firm-level fundamentals. In practice, the results imply that investors can enhance Portfolio allocation and reduce risk by considering clusters of firms with similar trading dynamics rather than relying solely on Standard industry groupings.

1. Introduction

In the context of increasingly interconnected financial systems, the adequacy of conventional industry-based classification frameworks for explaining equity market behaviour has come under growing scrutiny. Particularly within emerging markets, where informational asymmetries, liquidity constraints, and behavioural biases are more pronounced, stock price dynamics often reflect complex and non-linear interdependencies that transcend sectoral boundaries. The Colombo Stock Exchange (CSE), and specifically the S&P SL20 Index as its principal large-cap benchmark, provides a pertinent empirical setting to examine such dynamics within a developing economy undergoing structural, political, and macroeconomic transitions. As digital technologies accelerate the speed of information dissemination and amplify investor reactions, financial markets are becoming more sensitive to sentiment-driven and externally induced shocks, making traditional analytical approaches increasingly limited in capturing the full spectrum of market behaviour.

Conventional asset pricing and portfolio construction techniques predominantly rely on sectoral or fundamental classifications to interpret market movements and manage risk. While these approaches offer a structured framework, they often fail to detect latent co-movement patterns that arise from shared macroeconomic exposures, synchronised investor sentiment, and rapid information diffusion in the digital era. In practice, firms operating in distinct industries may exhibit strong behavioural alignment due to common responses to policy changes, geopolitical developments, or global economic signals. Such hidden interdependencies are frequently overlooked within static classification systems, thereby constraining a comprehensive understanding of systemic risk propagation and collective market dynamics. This limitation is particularly significant in emerging markets like Sri Lanka, where relatively concentrated investor bases and evolving institutional frameworks can intensify herd behaviour and information cascades.

To address these challenges, this study adopts a data-driven and methodologically advanced approach by employing Dynamic Time Warping (DTW) in conjunction with hierarchical clustering techniques to identify endogenous groupings of equities based on the similarity of their price trajectories. Unlike conventional correlation-based methods, DTW captures temporal misalignments and non-linear relationships in financial time series, enabling a more flexible and robust representation of market dependencies. By uncovering hidden behavioural clusters within the S&P SL20, the study contributes both methodologically and conceptually by integrating unsupervised machine learning with insights from behavioural finance and market microstructure. Importantly, this research is positioned within the broader discourse on digital-era transformations in financial systems, highlighting how advanced analytical tools can enhance the understanding of market interconnectedness, support evidence-based policymaking, and promote economic resilience in emerging economies. Through this interdisciplinary perspective, the study bridges quantitative financial modelling with social science inquiry, offering insights into how collective behaviour, information flows, and technological advancements shape financial stability in the Eurasian context.

To uncover these hidden connections, data-driven approaches have gained importance in financial research. Dynamic Time Warping (DTW) captures similarities in stock price movements by aligning time series with potential temporal distortions, while hierarchical clustering groups companies based on these co-movements rather than conventional industry labels. Together, these methods provide an alternative way of understanding how firms are interconnected in the Sri Lankan equity market.

The purpose of this research is not only to identify such non-traditional sector groupings in the S&P SL20, but also to give investors practical awareness. By suggesting that individuals spread their investments into smaller amounts across different clusters of companies, the study highlights diversification as a strategy for reducing risk and improving future gains. In doing so, it aims to bridge the gap between technical financial analysis and useful advice for investors.

2. Literature Review

The stock market plays a vital role in the capital generation of a nation. With increased financial integration, factoring in the regional and international dynamics shared in stock markets assists investors in portfolio diversification decisions and assists policy makers in implementing appropriate policies during a period (Samarawickrama and Pallegedara, 2023). The Colombo Stock Exchange is the capital stock market in Sri Lanka, which consists of 20 highly capitalised firms (A Study on Colombo Stock Exchange via Statistical Analysis of Financial Data, n.d.).

The S&P SL20 Index was initiated on 18 July 2012 and was launched in Colombo on 26 July 2012. The S&P Sri Lanka 20 seeks to be comprised of liquid and tradable stocks for easy and cost-effective replication as trading instruments, with possible application as index funds and Exchange-Traded Funds (ETFs) (Perera, Dissanayake and Jayasundara, 2016).

Although listed stocks provide investors with opportunities to earn profits by taking advantage of price fluctuations in the stock market, such volatility also introduces significant uncertainty, as stock prices can rise or fall unexpectedly. This means that while some investors may benefit from favourable movements, the same volatility increases the overall risk of dealing with a particular firm, since sharp and unpredictable price changes often signal instability in the company's financial position, market performance, or external environment.

Investors are looking for a reward for their investment. Significant attention to the earnings information of a company and the stock price's behavioural patterns is essential to make a better investment decision (Riyath and Madushan, n.d.) So, to identify the pattern of the market, many researchers try to fit some models. But those models are mostly not successful because of the volatility of the market (Perera, Dissanayake and Jayasundara, 2016).

So, some researchers try to find some groups of the market with the same pattern. Shu Bin use New York Stock Market data for S&P 500 companies' historical stock price movement, and their historical return on assets and asset turnover ratios are used as dissimilarity metrics for K-means clustering. References (Bin, n.d.) and also (A Study on Colombo Stock Exchange via Statistical Analysis of Financial Data, n.d.) use an effective clustering method, HRK (Hierarchical agglomerative and Recursive K-means clustering), to predict the short-term stock price movements. Likewise, researchers use several clustering methods to identify similar patterns of global stock market companies.

Previous studies on stock market clustering have mainly focused on traditional clustering techniques such as K-means clustering, which generally assume static relationships among companies. However, stock market data are highly dynamic and time-dependent in nature, where price movements and behavioural patterns change continuously over time. Because of this, conventional distance-based clustering methods may not adequately capture similarities in temporal patterns among companies. To address this limitation, the present study applies Dynamic Time Warping (DTW) together with hierarchical clustering to identify more meaningful clusters based on the movement patterns of stock prices. In addition, although several international studies have explored DTW-based clustering in financial markets, very limited attention has been given to the Sri Lankan context. In particular, no previous study has examined the clustering behaviour of S&P SL20 companies using this dynamic approach. Therefore, this research aims to fill that gap by applying DTW-based hierarchical clustering to the Sri Lankan stock market and identifying similarities among leading listed companies based on their time-series behaviour.

3. Methodology

3.1. Data Collection

The dataset for this study was constructed using the Sri Lanka 20 (S&P SL20) index constituents, which represent the top blue-chip companies listed on the Colombo Stock Exchange. To ensure a robust and representative analysis, all twenty constituent firms of the SL20 index were included in the study. These firms are Central Finance Company PLC (CFIN), Ceylon Cold Stores PLC (CCS), Chevron Lubricants Lanka PLC (LLUB), Commercial Bank of Ceylon PLC (COMB), DFCC Bank (DFCC), Dialog Axiata PLC (DIAL), Hatton National Bank PLC (HNB), Hayleys PLC (HAYL), Hemas Holdings PLC (HHL), John Keells Holdings PLC (JKH), Lanka IOC PLC (LIOC), LB Finance PLC (LFIN), LOLC Finance PLC (LOFC), LOLC Holdings PLC (LOLC), Melstacorp PLC (MELS), National Development Bank PLC (NDB), Nations Trust Bank PLC (NTB), Royal Ceramics Lanka PLC (RCL), Sampath Bank PLC (SAMP), and Vallibel One PLC (VONE).

The data were collected on a daily frequency over the period August 1, 2024, to July 31, 2025, providing one full year of trading records. For each company, we obtained standard open, high, low, close, and volume variables, which formed the basis for subsequent feature engineering. For the entire series, time-series clustering was chosen as the methodological approach, as it is well-suited for grouping firms based on their global temporal behaviour (Liao, 2005).

3.2. Data Preprocessing

Before clustering, the dataset was pre-processed to ensure comparability across variables. First, all company records were sorted by date and aligned to a common trading calendar. To enhance the explanatory power of raw data, we engineered the following features in addition to raw open and close prices:

$$\text{Daily return } (r_t) = \frac{Close_t - Close_{t-1}}{Close_{t-1}} \tag{1}$$

$$\text{High-Low range} = High_t - Low_t \tag{2}$$

$$\text{relative volume} = \frac{Volume_t}{long\ run\ Average} \tag{3}$$

$$\text{Momentum} = Close_t - Open_t \tag{4}$$

$$\text{Volatility} = \frac{High_t - Low_t}{Close_t} \tag{5}$$

$$\text{Volatility with k period} = \sqrt{\frac{1}{k}\sum_{i=0}^{k-1}(r_{t-i} - \bar{r})^2} \tag{6}$$

In addition to retaining the open and close prices, as well as the company name and date as identifiers. These variables capture both return dynamics, volatility, momentum, and trading activity, providing a more comprehensive representation of market behaviour. To remove the influence of differing scales across companies, each series was then z-normalised by subtracting its mean and dividing by its standard deviation. This transformation, widely recommended in time-series clustering, ensures that similarity is determined primarily by the shape of the series rather than its magnitude (Javed, Lee and Rizzo, 2020).

3.3. Clustering

To measure similarity between companies, we employed Dynamic Time Warping (DTW), a dynamic programming method that aligns two sequences by allowing non-linear shifts along the time axis (Chiba and Sakoe, 1978; Berndt and Clifford, n.d.). DTW was computed separately for each of the five features, and the mean of these distances was taken as the overall dissimilarity measure. To construct clusters, we applied agglomerative hierarchical clustering with average linkage, which is well-suited for non-Euclidean distances such as DTW. Unlike Ward's method, which assumes Euclidean geometry, average linkage provides theoretically consistent results with DTW (Murtagh and Legendre, 2014).

To determine the optimal number of clusters, multiple internal validation indices were applied:

- Silhouette Score: Measures how well-separated each firm is from other clusters (Rousseeuw, 1987).
- Calinski–Harabasz Index: Evaluates the ratio of between-cluster to within-cluster dispersion (Calinski and Harabasz, 1974).
- Davies–Bouldin Index: Assesses average cluster similarity, with lower values indicating better separation (Davies and Bouldin, 1979).

The evaluation was conducted across solutions with different clusters. Based on the average of metric ranks. Cluster assignments were extracted by cutting the dendrogram at the chosen level, and validation was carried out using both internal measures and visualisation techniques. Multidimensional scaling (MDS) was applied to embed the DTW distance matrix into two dimensions, enabling visual inspection of group separations (Borg and Groenen, 1997).

4. Results

When examining the variables, notable differences emerged in Volume and Momentum across firms over time. As illustrated in Fig. 1, John Keells consistently recorded the highest trading volumes on several occasions, while LOLC Finance and Dialog Axiata also displayed notable peaks during specific periods. Volume is a key indicator of market activity and liquidity: higher trading volume generally reflects stronger investor interest and the ease with which shares can be bought or sold without drastically affecting the price.

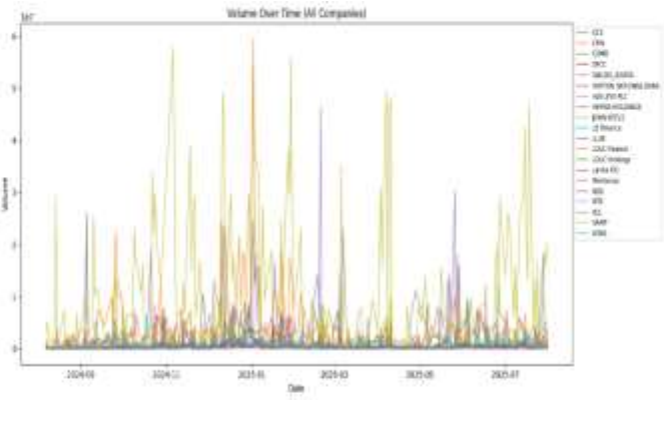

Figure 1: Volume over time

Similarly, Momentum displayed substantial variation among firms. Momentum, defined as the difference between closing and opening prices, captures the strength of price trends. High momentum indicates strong upward or downward movements, often accompanied by higher volatility and risk, whereas low momentum suggests relatively stable price behaviour. As shown in Fig. 2, LOLC Holdings demonstrated instances of high momentum, implying greater price fluctuations and, consequently, higher risk exposure for investors.

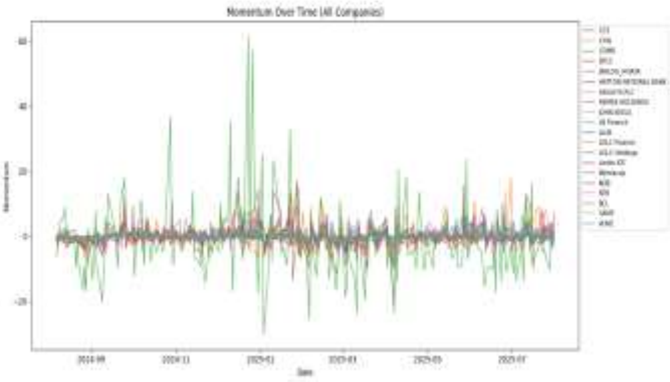

Figure 2: Momentum over time

Applying the DTW hierarchical clustering framework to the S&P SL20 dataset (1st August 2024 to 31st July 2025), the evaluation metrics indicated different optimal cluster numbers, as shown in Fig. 3. Both the Silhouette index (maximum value) and Davies-Bouldin index (minimum value) suggested two clusters as the optimal solution, whereas the Calinski-Harabasz index (maximum value) favoured five clusters.

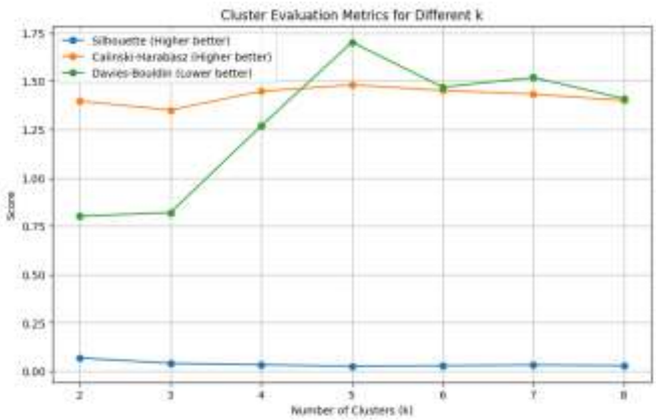

Figure 3: Evaluation metrics for DTW hierarchical clustering

The dendrogram in Fig. 4 provides a visual representation of the hierarchical clustering process applied to the S&P SL20 dataset using DTW. Each company is represented as a leaf on the horizontal axis, while the vertical axis indicates the distance at which clusters are merged. Companies that are joined at lower distances exhibit greater similarity in their time-series behaviour, whereas those merged at higher levels are relatively dissimilar. For instance, firms such as SAMP and COMB are clustered together at a relatively low distance, suggesting that their stock price dynamics share closer temporal patterns. In contrast, companies like RCL and LLUB are linked at much higher distances, showing that their stock movements behave very differently from other companies.

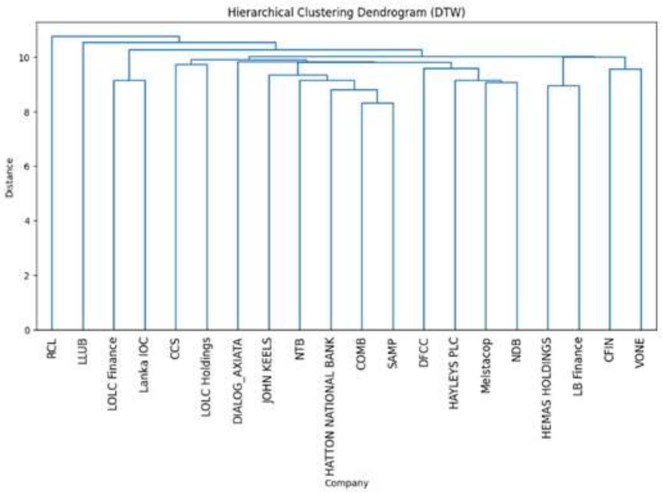

Figure 4: Dendrogram showing the hierarchical clustering of the S&P SL20

The two-cluster solution suggested by the Silhouette and Davies-Bouldin indices shows that RCL stands apart from the other companies. When three clusters are considered, LLUB separates from the previous grouping and forms a distinct cluster. With four clusters, LOLC Finance and Lanka IOC break away and form another cluster. Under the five-cluster solution indicated by the Calinski-Harabasz index (Fig. 5), VONE, CFIN, LB Finance, and Hemas Holdings form a new group. Fig. 5 illustrates the five-cluster solution, displaying the distribution of companies in a two-dimensional representation.

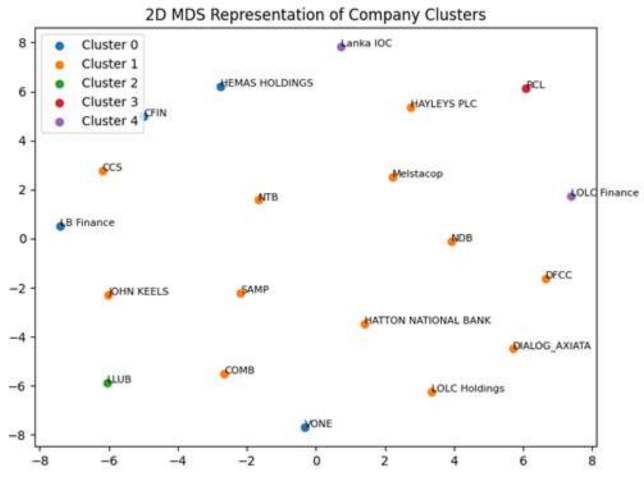

Figure 5: Five-cluster solution of the S&P SL20 dataset in a 2D representation.

When six clusters are considered, the previous group further divides into two subgroups, with LB Finance and Hemas Holdings forming one cluster and VONE and CFIN forming another. At the seven-cluster solution, CCS and LOLC Holdings separate into their own cluster. Expanding to eight clusters, Dialog Axiata emerges as a distinct cluster, standing apart from the other firms.

5. Discussion

This study applied Dynamic Time Warping (DTW) based hierarchical clustering to the S&P SL20 companies, uncovering hidden connections among firms that conventional industry categories fail to capture. The results showed that banks such as Sampath Bank (SAMP) and Commercial Bank (COMB) shared close temporal dynamics, while companies like Richard Pieris (RCL) and Lanka Lubricants (LLUB) exhibited highly distinct behaviours, forming their own clusters. Evaluation metrics suggested both a two-cluster solution, where RCL diverged from the rest, and a five-cluster solution with more granular subgroups such as Vallibel One (VONE), Central Finance (CFIN), LB Finance, and Hemas Holdings. These findings suggest that Sri Lankan companies can move together due to broader market forces such as liquidity shocks or investor sentiment, rather than just industry affiliation. Compared with earlier CSE studies that focused primarily on volatility and earnings as drivers of stock price behavior (Javed, Lee and Rizzo, 2020), (Riyath and Madushan, n.d.), this research highlights the added value of clustering in revealing non-traditional groupings with practical implications for portfolio diversification.

Globally, clustering has been widely used to detect stock market patterns, such as Shu Bin's application of K-means clustering to S&P 500 firms, which grouped stocks based on both price movements and financial ratios (Bin, n.d.). Similarly, the HRK method has been applied for short-term stock price prediction (Samarawickrama and Pallegedara, 2023). The present findings are consistent with these global studies, showing that clustering reveals latent structures that traditional time-series or volatility models cannot capture. For Sri Lankan investors, this implies that diversification strategies should consider clusters of firms with distinct dynamics, rather than relying solely on industry-based categories. By extending clustering methodologies to the S&P SL20, this study contributes new insights to local financial research, bridging the gap between technical analysis and actionable guidance for investors.

6. Conclusion

This study investigated the interrelationships among S&P SL20 companies listed on the Colombo Stock Exchange using Dynamic Time Warping (DTW) based hierarchical clustering. By analyzing daily trading data over one year, we identified both broad and fine-grained cluster structures, showing that firms can exhibit similar market behaviors even when they belong to different industries. The results highlighted distinct clusters such as Richard Pieris (RCL) and Lanka Lubricants (LLUB), which stood apart from the main group, as well as smaller clusters formed by finance and diversified holdings companies. These patterns demonstrate that stock market co-movements in Sri Lanka are not always aligned with traditional industry classifications but are influenced by liquidity, momentum, and investor sentiment.

The findings add to the growing body of literature on market clustering, extending global methods to the Sri Lankan context. Compared with earlier studies that primarily examined volatility and firm-level earnings to explain stock price movements, this research shows that clustering provides a complementary and powerful tool for identifying hidden sectoral relationships. The study highlights that diversification across identified clusters, rather than industries, may help investors reduce risk and improve portfolio performance.

From a practical perspective, this dynamic clustering approach can be useful for investors and traders seeking to diversify their portfolios. By allocating investments across clusters identified by time-dependent market behavior, investors may reduce non-systematic risk and improve the risk-return balance. For example, capital can be distributed across two-way, three-way, or four-way cluster groupings depending on the level of diversification desired, rather than relying only on traditional industry-based classification. This strategy can support more efficient portfolio construction for S&P SL20 companies and can also be extended to other listed firms in the Colombo Stock Exchange, providing a broader framework for dynamic risk management and investment decision-making. Future research could expand the dataset beyond one year, incorporate macroeconomic variables, or compare alternative clustering algorithms to further validate and enhance these insights.

7. Recommendations

This study can be extended by applying the proposed framework to a larger set of companies in the Colombo Stock Exchange to obtain a more comprehensive market analysis. In addition, the same approach can be applied to other international stock markets such as the New York Stock Exchange, Shanghai Stock Exchange, and the Japanese stock market to test its wider applicability. Unlike previous studies that mainly use static clustering methods such as K-means, this framework focuses on dynamic clustering based on time-dependent behavior, which can provide more realistic market groupings. Future studies may also include sector-level analysis and multi-market comparisons to further improve the usefulness of the model. In practice, this approach can help investors improve portfolio diversification, manage risk more effectively, and support better investment decisions.

References

Wayamba University of Sri Lanka, Sri Lanka, I. D. W. Samarawickrama, A. Pallegedara, and Wayamba University of Sri Lanka, Sri Lanka, “A Long-Term Relationship Between Sri Lankan Stock Market and Global Stock Markets: A Time Series Analysis,” presented at the SLIIT International Conference on Advancements in Sciences and Humanities 2023, Dec. 2023, pp. 247–252. doi: 10.54389/avnd4327.

U. Perera, R. Dissanayake, and M. Jayasundara, “Assessing the Impact of S&P SL20 Index Construction on Listed Companies in Colombo Stock Exchange (CSE),” Int. J. Econ. Finance, vol. 8, no. 7, p. 159, July 2016, doi: 10.5539/ijef.v8n7p159.

M. I. M. Riyath and H. C. N. Madushan, “Impact of Firms’ Earnings on Stock Prices: Special Reference to the S&P Sri Lanka Companies”.

S. Bin, “K-Means Stock Clustering Analysis Based on Historical Price Movements and Financial Ratios”.

T. Warren Liao, “Clustering of time series data—a survey,” Pattern Recognit., vol. 38, no. 11, pp. 1857–1874, Nov. 2005, doi: 10.1016/j.patcog.2005.01.025.

A. Javed, B. S. Lee, and D. M. Rizzo, “A benchmark study on time series clustering,” Mach. Learn. Appl., vol. 1, p. 100001, Sept. 2020, doi: 10.1016/j.mlwa.2020.100001.

S. CHIBA and H. SAKOE, “Dynamic Programming Algorithm Optimisation for Spoken Word Recognition,” IEEE Trans. Acoust. SPEECH SIGNAL Process., vol. ASSP-26, Feb. 1978.

D. J. Berndt and J. Clifford, “Using Dynamic Time Warping to Find Patterns in Time Series”.

F. Murtagh and P. Legendre, “Ward’s Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward’s Criterion?,” J. Classif., vol. 31, no. 3, pp. 274–295, Oct. 2014, doi: 10.1007/s00357-014-9161-z.

P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” J. Comput. Appl. Math., vol. 20, pp. 53–65, Nov. 1987, doi: 10.1016/0377-0427(87)90125-7.

T. Calinski and J. Harabasz, “A dendrite method for cluster analysis,” Commun. Stat. - Theory Methods, vol. 3, no. 1, pp. 1–27, 1974, doi: 10.1080/03610927408827101.

D. L. Davies and D. W. Bouldin, “A Cluster Separation Measure,” IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979, doi: 10.1109/TPAMI.1979.4766909.

Borg and P. Groenen, Modern Multidimensional Scaling. in Springer Series in Statistics. New York, NY: Springer New York, 1997. doi: 10.1007/978-1-4757-2711-1.

