E-ISSN 2621–086X

Indonesian Journal of Applied Statistics
Vol. 7, No. 2, pp. 112-120, November 2024
Journal homepage: https://jurnal.uns.ac.id/ijas/index
Doi: 10.13057/ijas.v7i2.94456

Grouping Indonesian Province Farmers' Term of Trade Using Dynamic Time Warping

Imtikhanah Anis Mahmudi, Rohmatul Fajriyah*
Master Program in Statistics, Universitas Islam Indonesia, Yogyakarta, Indonesia

Corresponding author: rfajriyah@uii.ac.id

Submitted: 25-Oct-2024 **Revised:** 15-Nov-2024 **Accepted:** 02-Dec-2024

Abstract. *This study employs dynamic time warping (DTW) to analyze the farmer's terms of trade (FTT) across 34 provinces in Indonesia, aiming to identify patterns and cluster similarities in time series data. DTW is recognized for its effectiveness in measuring flexible similarities under time distortions, making it particularly suitable for time series classification across various fields. The FTT is utilized to assess farmers' purchasing power by comparing the prices they receive for their products to the prices they pay for goods and services. K-Medoid clustering techniques were applied to group provinces based on their DTW distances, revealing three distinct clusters. The silhouette score indicates that three clusters as the optimum cluster for the FTT data. The findings show that the first and third clusters have low mean of FTT and the second cluster has the highest mean FTT. These indicates disparities in farmers' income and purchasing power across regions where the government needs to enhance agricultural strategies and improve economic conditions for farmers in the first and third clusters.*

Keywords: *Clustering, dynamic time warping, farmers term of trade, k-medoid.*

1. INTRODUCTION

Dynamic time warping (DTW) is probably the most popular distance measure for time series data, because it captures flexible similarities under time distortions [1]. Time series classification is a part of machine learning with many applications. It is characterized by the he temporal structure of the data, where standard machine learning algorithms cannot be used as data are correlated [2].

Research topic in time series classification is increasing research topic due to its wide range of applications. Many algorithms have been proposed, including the algorithm that extract the feature from timeseries. Fulcher and Jones [3] explain that feature-based methods allow for a better understanding of the time series properties that are important for classification. This can help in gaining insights into the underlying patterns and trends in the data. This method allow for a comparison of different feature sets and their performance on the same dataset, which can help in selecting the most suitable feature set for a given problem.

Dynamic time warping (DTW) is known to be a robust, outlier-insensitive alternative to other distance measures such as Euclidean distance or Manhattan distance [4,5]. DTW is particularly useful for aligning time series data that have different lengths or that exhibit temporal distortions, such as phase shifts or time warping [6]. This is a powerful tool for time series analysis that can help in accurately comparing and classifying time series data.

In the context of DTW, clustering can be used to group regions or provinces in Indonesia based on their time series data. The DTW method has been widely adopted by researchers in Indonesia across various fields, including agriculture [7], economic [8,9], poverty studies [10], computer science [11], healthcare [12], and the linkage between these areas [13]. By clustering regions or provinces based on their time series data, decision-makers can gain valuable insights into the similarities and differences between different regions, allowing them to make informed decisions and develop appropriate policies.

The Indonesian government's goal is to improve the well-being of all its people by developing different sectors, especially agriculture. Many people in rural areas depend on farming to make a living, then it's important for this sector to boost economic growth, increase farmers' incomes, and reduce poverty. To track how well farmers are doing, we need a way to measure their purchasing power. One common way to do this is by using the farmer’s terms of trade (FTT). FTT compares two things: the prices farmers get when they sell their products (It) and the prices they pay for goods and services (Ib). The ‘It’ tells us how much farmers are earning, while the ‘Ib’ shows what farmers need to spend on both their daily needs and production costs. The FFT helps to understand how well farmers can trade what they produce for the things they need to live and continue farming [1].

Central Bureau of Statistics [1] define farmer is a person who manage an agricultural business including food crops, horticulture crops, smallholding estate crops, animal husbandry and fishery at his own risk for sale, either as farm owner or farm worker. Price received by farmers (It) is the farm gate price which is the average of producer prices of agricultural products not including the transportation and packaging costs into the selling prices. The multiplication of average price and volume of sale will show the total income received by farmers. Price paid by farmers is the average of retail prices of goods and services needed by farmers, either for household consumption or production process [1]. The equation to get the farmers’ term of tradev (FTT or NTP in Bahasa Indonesia) can be expressed in equation 1.

$$NTP\ or\ FTT = \frac{It}{Ib} \times 100 \tag{1}$$

In general, the value of FTT more than 100 show that farmers experience a rise in term of trade. The value of FTT equal with 100 give information that farmers experience a stable of term of trade and when the value of FTT less than 100 show that farmers experience a fall in their terms of trade when the price that farmers paid increase at a faster rate than the price received by farmers. According to the Central Bureau of Statistics, the index of prices received by farmers (It) reflects the price changes of agricultural products that farmers produce. This index also serves as supplementary data for calculating the gross domestic product (GDP) or gross regional domestic product (GRDP) for the agricultural sector. On the other hand, the index of prices paid by farmers (Ib) tracks price changes for goods and services consumed by rural households. The production cost index provides insight into the price fluctuations of inputs necessary for farming activities. By analyzing the FTT, it is possible to assess whether farmers' income growth can keep pace with rising expenses. Essentially, FTT measures the balance between the cost of production and farmers' purchasing power for goods and services.

In the context of economic development, it is necessary to classify farmer exchange rate indexes from each sector of farmer exchange rates so that farmer exchange rate groups are grouped using cluster analysis [14]. Previous studies have applied clustering techniques, such as self-organizing maps (SOM) and K-affinity propagation (K-AP) [15], as well as K-means [16], to group provinces in Indonesia based on FTT. In this paper, the FTT for 34 provinces in Indonesia will be analyzed using DTW, similar to [17] and the grouping technique is K-medoids instead of K-means.

2. MATERIALS

2.1. Dynamic Time Warping

The distance measure utilized in this research is DTW, which is employed to calculate and identify the optimal path between two time series data [6]. The DTW algorithm efficiently gauges the similarity between two time series data, minimizing the impact of time lags and distortions, enabling detection even in different phases.

Mizutani and Dreyfus [18] explained that given an observed pattern with synthesized template patterns, the task is to calculate the optimum alignment among them, for which the similarity is quantified as a flexible distance measure realized by DTW. The task is considered as finding the *constrained minimum-cost path problem*, which is a problem that can be solved by dynamic programming (DP).

Given two time series data $Q = q_1, q_2, q_3, \cdots, q_m$ of size and $C = c_1, c_2, c_3, \cdots, c_m$ of the size n, a matrix of size $m \times n$ is formed. The value d_{ij} represents the distance between q_i and c_j [9]. The calculation of the d_{ij} value can be expressed in equation 2.

$$d_{ij} = \left(q_i - c_j\right)^2 \tag{2}$$

The value of each cell is calculated using DP by taking the minimum cost path from adjacent cells. This results in the recurrence relation can be expressed in equation 2.

$$D(i,j) = d_{ij} + \min(D(i-1,j), (D(i,j-1), D(i-1,j-1) \tag{3}$$

or equation 3.

$$d_{DTW} = (Q,C) = min\sqrt{\sum d_{ij}} \tag{4}$$

By calculating values from the base cases (i.e., the edges of the matrix) and using these in the computation of neighboring cells, DTW finds the lowest cost path between the two sequences.

Many researchers have praised the use of DTW in various fields, such as speech recognition [19,20], sport (posture recognition) in [21], health [2,22,23,24], road surveillance [25], object movement, marketing [26], food and quality assurance [2]. But there are some drawback that users needs to be aware. Research of [27,28,29] have discussed some limitation of DTW. The findings of [27] encouraged people to apply it in many areas after their empirical research results dispel the three myths of DTW, namely length of time series, the constraints, and the speed. The limitation of DTW for two dimensional characteristic movements has been addressed by [28]. Their modification shows an improved performance compared to other methods. Another research, [29], has proposed the modification of traditional DTW, adaptively constraints DTW, and showed that it can be implemented in both of clustering and classification.

2.2 K-medoid Cluster analysis

Cluster analysis is a technique aimed at grouping objects or individuals into clusters based on the similarity of the object characteristics. The primary objective of this analysis is to enhance similarity within each cluster while reducing similarity between different clusters. Cluster analysis is descriptive in nature, does not involve inferential statistics, and lacks a statistical foundation that permits generalizations from samples to populations [30]. Cluster analysis, categorizing it into several clustering methods, including: partitional clustering or distance-based clustering, hierarchical clustering, density-based clustering, and categorical data clustering [31]. K-means and k-medoids are popular methods used in customer segmentation analysis based on purchasing patterns. However, K-medoids is specifically designed to address the weaknesses of k-means, which is susceptible to the influence of outlier data [32].

The K-medoids algorithms, which is used in here, are described in [33] as following

a. Select randomly k from n data as *medoid*
b. Calculate the distance of each data to the existing medoid, by using a valid distance measure (Euclidean, Manhattan atau Minkowski) and assign it to the closest medoid
c. Calculate the total cost (sum of all distance from all data points to the medoids)
d. Select a random point as the new medoid and swap it with the previous medoi. Repeat steps b and c.
e. If the total cost of the new medoid is smaller than that of the previous medoid, make the new medoid permanent and repeat step d.
f. If the total cost of the new medoid is greater than the cost of the previous medoid, undo the swap and repeat step d.
g. The Repetitions are continued until no change is encountered with new medoids to cluster data points.

2.3. Silhouette Coefficient

The silhouette coefficient is one measure of accuracy that can be used in determining the accuracy of time series grouping. In addition, the silhouette coefficient is used to determine the grouping quality. The silhouette score consider as bad classification if the value is 0.00-0.25. The silhouette score consider as weak classification if the value is 0.26-0.50. The silhouette score consider as good classification if the value greater than 0.51-0.70. It is consider as strong classification when the value is 0.71-1.00 [17]. With the following equation 4.

$$s(i) = \frac{b(i) - a(i)}{max[a(i), b(i)]} \tag{4}$$

3. RESEARCH METHODS

The dataset utilized in this study originates from the official website of the Indonesian Central Bureau of Statistics (BPS). Specifically, it pertains to the FTT across provinces and years in Indonesia. There are 34 provinces and based on data from January 2020 to September 2024. The variables in this study are:

1. The FTT values for each province and across time.
2. The time (monthly) since January 2020 to September 2024.

4. RESULTS AND DISCUSSION

Before clustering the dataset, we described the data visually as in Figure 1. Figure 1 shows that some provinces have FTT increased highly, moderately and lowly. Then clustering process is begun by computing the Silhouette score. From Table 1, the data set optimally can be clustered into two or three clusters. For some analysis purpose we choose to have 3 clusters.

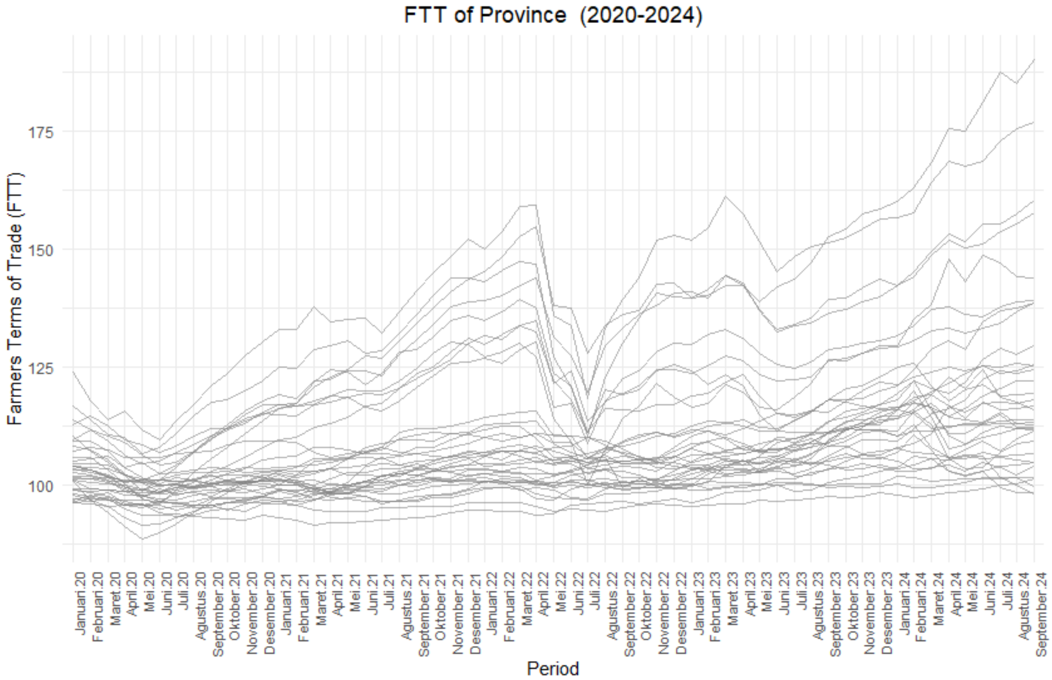

Figure 1. Visualization of FTT timeseries data

Table 1. Silhouette score

Clusters	**Silhouette**
2	0.657398
3	0.60221
4	0.348138
5	0.348992

Based on Table 1, then we have the first cluster includes 14 provinces, the second cluster comprises 8 provinces, and the third cluster consists of 14 provinces as shown in Table 2. Figure 2 summarizes the clusters, with the main line illustrating the fluctuations of FTT over time since 2020, which suggest occurred post-COVID-19. The figure demonstrates that the second cluster exhibits the highest mean, as highlighted in Table 3. Besides, the third cluster was the lowest.

Table 2. Cluster member

Clusters	**Province**
1	Sumatera Barat, Sumatera Selatan, Lampung, Jawa Tengah, Jawa Timur, Nusa Tenggara Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara
2	Sumatera Utara, Riau, Jambi, Bengkulu, Kep. Bangka Belitung, Kalimantan Barat, Kalimantan Timur, Sulawesi Barat
3	Kep. Riau, DKI Jakarta, Jawa Barat, DI Yogyakarta, Banten, Bali, Nusa Tenggara Timur, Gorontalo, Maluku, Maluku Utara, Papua Barat, Papua

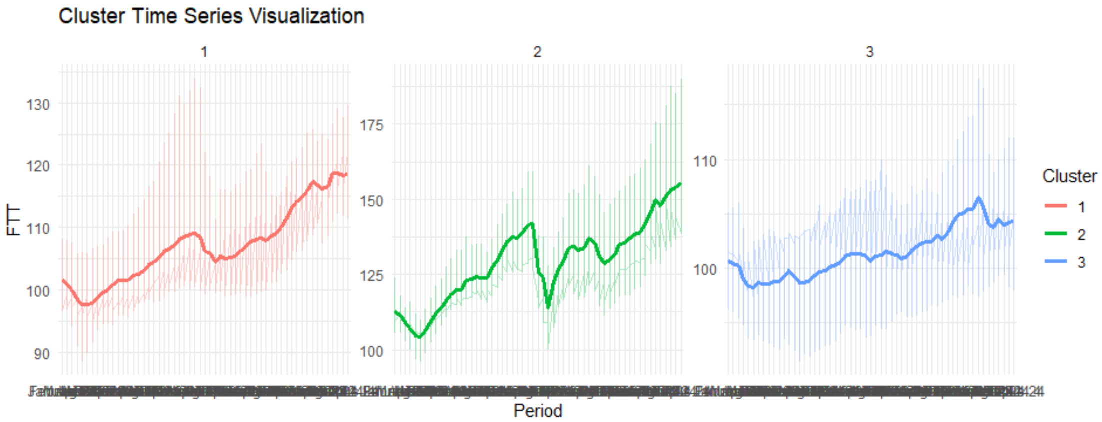

Figure 2. Visualization of FTT time series data

Figure 2 illustrates that the second cluster experiences a fluctuating period, as evidenced by a steep graph in mid-2022, but it quickly improves by 2024. Cluster 1 demonstrates a noticeable increase, with no drastic decline observed in mid-2022. In contrast, Cluster 3 exhibits the lowest average among all clusters but subjectively not as fluctuate as other cluster. The provinces within Cluster 3 need to implement different strategies, as the exchange rate of agricultural products relative to the cost of production and the consumption of goods and services indicates challenges in this group.

Table 3. Cluster summary

Cluster	Mean	Deviation Standard	Min	Max
1	107	7.81	88.6	134
2	129	16.5	96.1	190
3	101	4.23	91.5	117

Figure 3 presents a boxplot summarizing the clusters, revealing that the second cluster exhibits the highest mean and standard deviation. This suggests that provinces within the second cluster enjoy comparatively the better selling prices for their products, indicating that farmers are experiencing an improvement in their terms of trade. This occurs when the average prices received by farmers rise at a faster pace than the average prices they pay, or when the prices they receive decrease at a slower rate than the prices they pay.

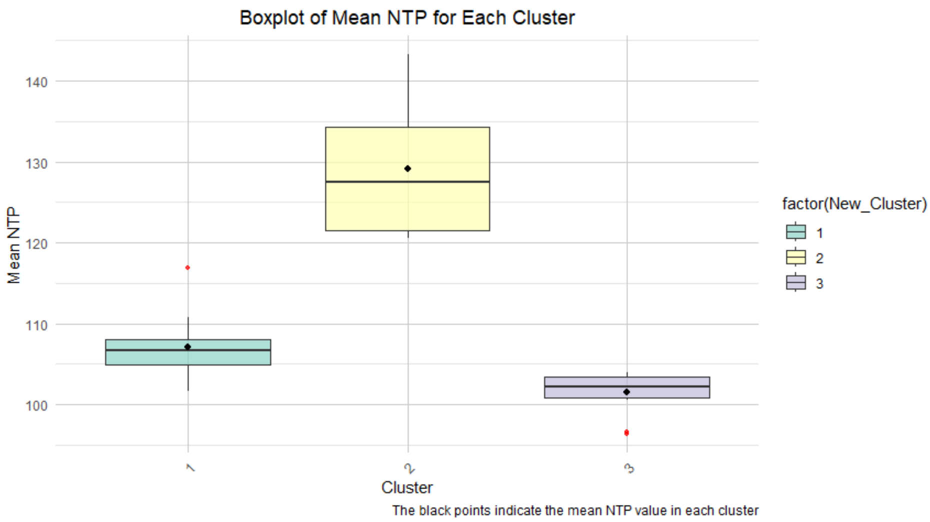

Figure 3. Boxplot of NTP/FTT cluster summary

Figure 3 also describes that the government should pay more attention to farmers in clusters 1 and 3, especially cluster 3. Because during the 4 years period their FTT significantly lower than farmers which are in cluster 2 provinces.

5. CONCLUSIONS

In summary, this study employs dynamic time warping (DTW) to analyze fluctuations in farmer's terms of trade (FTT) across various provinces, effectively grouping them based on calculated distances. The results reveal three distinct clusters: the first cluster includes 14 provinces, the second cluster consists of 8 provinces, and the third cluster encompasses 14 provinces. The second cluster demonstrates the highest mean FTT, indicative of better selling prices for farmers, suggesting an improvement in their terms of trade. In contrast, the third cluster displays the lowest average FTT, indicating a need for strategic interventions in these provinces related to agricultural product pricing and production costs. The visual representations, including time series and boxplots, further illustrate the dynamics of FTT fluctuations and provide insights into the performance of each cluster, particularly in the context of post-COVID-19 economic conditions.

The conclusions drawn from the research contribute to the existing body of scientific knowledge by highlighting the impact of economic conditions, particularly in a post-COVID-19 context, on agricultural pricing dynamics. In future research prospects, this study opens avenues for further investigation into the underlying factors affecting FTT across different provinces. Future studies could explore the impact of specific agricultural policies, market access, and local economic conditions on FTT variability. Additionally, applying similar methodologies to other economic indicators could enhance our understanding of agricultural economics and inform more effective policy decisions aimed at improving farmers' livelihoods.

Acknowledgements

We would like to express our heartfelt gratitude to the Department of Statistics at Universitas Islam Indonesia for providing the necessary guidance and support throughout this research. The invaluable insights and encouragement from the department members greatly enhanced our understanding of statistical methodologies and their application in this study. We also appreciate the collaborative environment fostered within the department, which allowed us to engage in meaningful discussions and refine our analytical skills. This research would not have been possible without the strong foundation laid by the department, and we are truly grateful for the opportunity to be part of such an esteemed institution.

REFERENCES

[1] Statistics Indonesia, *Statistik Nilai Tukar Petani 2023*, vol. 30. Indonesia: Badan Pusat Statistik, 2023.

[2] P. Senin, “Dynamic time warping algorithm review,” *Science (80-.).*, vol. 2007, no. December, pp. 1–23, 2008, [Online]. Available: http://129.173.35.31/~pf/Linguistique/Treillis/ReviewDTW.pdf.

[3] B. D. Fulcher and N. S. Jones, “Highly comparative feature-based time-series classification,” *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 12, pp. 3026–3037, 2014, doi: 10.1109/TKDE.2014.2316504.

[4] K. Bringmann, N. Fischer, I. van der Hoog, E. Kipouridis, T. Kociumaka, and E.

Rotenberg, “Dynamic dynamic time warping,” *Proc. Annu. ACM-SIAM Symp. Discret. Algorithms*, vol. 2024-Janua, pp. 208–242, 2024, doi: 10.1137/1.9781611977912.10.

[5] A. D. Munthe, “Penerapan clustering time series untuk menggerombolkan provinsi di indonesia berdasarkan nilai produksi padi,” *J. Litbang Sukowati Media Penelit. dan Pengemb.*, vol. 2, no. 2, p. 11, 2019, doi: 10.32630/sukowati.v2i2.61.

[6] I. S. Narendra and M. Muhajir, “Analysis of dynamic time warping in the development of gross regional domestic product Yogyakarta,” *J. Ris. Inform.*, vol. 4, no. 4, pp. 397–406, 2022, doi: 10.34288/jri.v4i4.432.

[7] D. Setiawan and A. Zahra, “Pengelompokan kemiskinan di Indonesia menggunakan time series based clustering,” *Inferensi*, vol. 6, no. 1, p. 83, 2023, doi: 10.12962/j27213862.v6i1.14969.

[8] S. Putri, A. N., Satyahadewi, N., Aprizkiyandari, “Pengelompokan provinsi di Indonesia menggunakan time series clustering pada sektor ekspor nonmigas,” *Jambura J. Math.*, vol. 6, no. 1, pp. 16–22, 2024, doi: https://doi.org/10.37905/jjom.v6i1.21921.

[9] C. Dinata, D. Puspitaningrum, and E. Erna, “Implementasi teknik dynamic time warping (DTW) pada aplikasi speech to text,” *J. Tek. Inform.*, vol. 10, no. 1, pp. 49–58, 2018, doi: 10.15408/jti.v10i1.6816.

[10] C. Cindy, C. Cynthia, V. Vito, D. Sarwinda, B. D. Handari, and G. F. Hertono, “Cluster analysis on dengue incidence and weather data using k-medoids and fuzzy c-means clustering algorithms (case study: spread of dengue in the DKI Jakarta Province),” *J. Math. Fundam. Sci.*, vol. 53, no. 3, pp. 466–486, 2021, doi: 10.5614/j.math.fund.sci.2021.53.3.9.

[11] D. Miljkovic and P. Vatsa, “On the linkages between energy and agricultural commodity prices: A dynamic time warping analysis,” *Int. Rev. Financ. Anal.*, vol. 90, no. July, 2023, doi: 10.1016/j.irfa.2023.102834.

[12] R. A. J. and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. Upper Saddle River, 2007.

[13] I. Ayundari, D. Statistika, F. Matematika, and S. Data, “Penentuan zona musim di Mojokerto,” vol. 2, no. 2, 2019.

[14] D. Amelia, G. Kholijah, and U. Jambi, “Analisis cluster pengelompokan provinsi di Indonesia berdasarkan sub sektor nilai tukar petani,” *J. Demogr. Soc. Transform.*, vol. 3, no. 1, pp. 1–12, 2023, [Online]. Available: https://www.researchgate.net/publication/378686148_Analisis_Cluster_Pengelompokan_Provinsi_di_Indonesia_Berdasarkan_Sub_Sektor_Nilai_Tukar_Petani.

[15] S. H. Hastuti, W. P. Nurmayanti, and A. A. Saputri, “Penerapan metode clustering self organizing maps (SOM) dan k-affinity propagation (K-AP) dalam mengelompokkan nilai tukar petani di Indonesia 2022,” *Var. J. Stat. Its Appl.*, vol. 5, no. 1, pp. 79–88, 2023, doi: 10.30598/variancevol5iss1page79-88.

[16] M. M. Na’im, “Penggunaan k-means dalam pengelompokan provinsi di Indonesia berdasarkan ntp (nilai tukar petani),” Universitas Nahdlatul Ulama Sunan Giri, 2024.

[17] L. P. W. Adnyani and P. R. Sihombing, “Analisis cluster time series dalam pengelompokan provinsi di Indonesia berdasarkan nilai PDRB,” *J. Bayesian J. Ilm. Stat. dan Ekon.*, vol. 1, no. 1, pp. 47–54, 2021, doi: 10.46306/bay.v1i1.5.

[18] E. Mizutani, & S. Dreyfus, “On using dynamic programming for time warping in pattern recognition,” *Information Sciences*, *580*, 684-704, 2021. https://doi.org/10.1016/j.ins.2021.08.075

[19] Y. Y. Yu, P. P. Wu, K. Mengersen, W. Hobbs, “Classifying ball trajectories in invasion sports using dynamic time warping: A basketball case study,” PLoS One, Oct 20;17(10):e0272848, 2022, doi: 10.1371/journal.pone.0272848.

[20] S. Jiang, & Z. Chen, “Application of dynamic time warping optimization algorithm in speech recognition of machine translation,” *Heliyon*, *9*(11), e21625, 2023, https://doi.org/10.1016/j.heliyon.2023.e21625

[21] C. Niu, “The application of improved DTW algorithm in sports posture recognition,” *Systems and Soft Computing*, *6*, 200163, 2024,

https://doi.org/10.1016/j.sasc.2024.200163.

[22] C. Serantoni, A. Riente, A. Abeltino, G. Bianchetti, M. Maria De Giulio, S. Salini, A. Russo, F. Landi, M. De Spirito, & G. Maulucci, "Integrating Dynamic Time Warping and K-means clustering for enhanced cardiovascular fitness assessment," *Biomedical Signal Processing and Control*, *97*, 106677, 2024, https://doi.org/10.1016/j.bspc.2024.106677

[23] T. Watase, Y. Omiya, S. Tokuno, "Severity classification using dynamic time warping-based voice biomarkers for patients with covid-19: Feasibility Cross-Sectional Study," *JMIR Biomed Eng*, Nov 6;8:e50924, 2023, doi: 10.2196/50924. PMID: 37982072; PMCID: PMC10631492.

[24] S. Lee, "Application of dynamic time warping algorithm for pattern similarity of gait," *Journal of Exercise Rehabilitation*, *15*(4), 526, 2019, https://doi.org/10.12965/jer.1938384.192

[25] S. K. Sharma, H. Phan, J. Lee, "An application study on road surface monitoring using DTW based image processing and ultrasonic sensors," *Appl. Sci.*, *10*, 4490, 2020, https://doi.org/10.3390/app10134490

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

[27] C. A. Ratanamahatana, and E. Keogh, "Everything you know about dynamic time warping is wrong," *Third Workshop on Mining Temporal and Sequential Data, in conjunction with the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004),* August 22-25, 2004 - Seattle, WA, 2004.

[28] Choi, Hyo-rim & Kim, Taeyong., "Modified dynamic time warping based on direction similarity for fast gesture recognition," *Mathematical Problems in Engineering*, 1-9, 2018, 10.1155/2018/2404089.

[29] H. Li, J. Liu, Z. Yang, R. W. Liu, K. Wu, & Y. Wan, "Adaptively constrained dynamic time warping for time series classification and clustering," *Information Sciences*, Volume 534, Pages 97-116, 2020, https://doi.org/10.1016/j.ins.2020.04.009.

[30] W. C. Hair, Joseph F., Anderson, Rolph E., Black, *Multivariate Data Analysis*, 7th ed. America: Pearson, 2014.

[31] B. Malik, A., & Tuckfield, *Applied unsupervised learning with R: Uncover hidden relationships and patterns with K-Means clustering, hierarchical clustering, and PCA*. UK: y Packt Publishing Ltd., 2019.

[32] N. F. Fahrudin and R. Rindiyani, "Comparison of k-medoids and k-means algorithms in segmenting customers based on RFM criteria," *E3S Web Conf.*, vol. 484, 2024, doi: 10.1051/e3sconf/202448402008.

[33] R. Fajriyah, and A. M. Imtikhanah, *Kesehatan Masyarakat Indonesia 2013 : SBRC Series Analisis Data Kesehatan 1.01*. Edited by Winoto, Darmawan E. Eureka Media Aksara, 2023.