BAREKENG: Journal of Mathematics and Its Applications
September 2026 Volume 20 Issue 3 Page 2299-2310
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss3pp2299-2310

DYNAMIC TIME WARPING-BASED FUZZY C-MEANS WITH MULTIDIMENSIONAL SCALING FOR TIME SERIES CLUSTERING

Sri Hidayati[1*], Regita Putri Permata[2], Fidi Wincoko Putro[3]

[1,2]*Data Science Study Program, Telkom University*
[3]*Software Engineering Study Program, Telkom University*
Jln. Ketintang No. 156, Surabaya, Jawa Timur, 60231, Indonesia

*Corresponding author's e-mail: * **srihidayatisi@telkomuniversity.ac.id***

Article Info

Article History:
Received: 15th August 2025
Revised: 8th January 2026
Accepted: 10th March 2026
Available online: 8th April 2026

Keywords:
Dynamic Time Warping;
Fuzzy C-Means;
Multidimensional scaling;
Rainfall;
Time series clustering.

ABSTRACT

Weather refers to atmospheric conditions such as temperature, humidity, air pressure, wind speed, and rainfall, all of which influence human activities. Rainfall is particularly important due to its impact on agriculture and water resource management. This study classifies regions on Java Island based on rainfall patterns using the Fuzzy C-Means algorithm. Rainfall variations are influenced by geographical, topographical, and climatic factors, requiring methods that can capture spatial and temporal changes. Fuzzy C-Means was selected for its ability to manage data uncertainty and overlapping clusters. To measure rainfall pattern similarity between regions, the Dynamic Time Warping (DTW) method was applied. Since DTW is a non-Euclidean metric and incompatible with Fuzzy C-Means, the Multidimensional Scaling (MDS) method was used to convert DTW distance matrices into Euclidean feature vectors. The study used secondary daily rainfall data from NASA (2021–2024). Clustering performance was evaluated using the Silhouette Coefficient, yielding a value of 0.413184, indicating good compactness and separation. Results identified three clusters: low rainfall (Cluster 0), moderate rainfall (Cluster 1), and high rainfall (Cluster 2). ANOVA results confirmed significant differences in average rainfall between clusters, with Tukey HSD tests showing Cluster 2 significantly differs from Clusters 0 and 1, while Clusters 0 and 1 are not significantly different. These findings demonstrate that combining DTW, MDS, and Fuzzy C-Means effectively identifies temporal rainfall patterns and produces statistically meaningful clustering. The spatial distribution of each cluster is visualized using GeoJSON and a database for clearer interpretation.

How to cite this article:

S. Hidayati, R. P. Permata, and F. W. Putro, "DYNAMIC TIME WARPING-BASED FUZZY C-MEANS WITH MULTIDIMENSIONAL SCALING FOR TIME SERIES CLUSTERING", *BAREKENG: J. Math. & App.*, vol. 20, no. 3, pp. 2299-2310, Sep, 2026.

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

Research Article · Open Access

1. INTRODUCTION

Weather is the atmospheric conditions that include temperature, humidity, air pressure, wind speed, and precipitation in a particular area at a specific time. Weather changes can occur in a relatively short period of time and affect human activities, such as agriculture, transportation, and other activities. One of the most important weather elements to consider is rainfall [1]. Rainfall refers to the amount of water that falls to the Earth's surface in liquid or solid form because of the condensation of water vapor in the atmosphere. Understanding rainfall is crucial for planning activities that depend on weather conditions. Rainfall (mm) is the height of rainwater collected on a flat surface, where evaporation does not occur, the water does not seep into the ground, and the water does not flow [2]. Rainfall plays a role in various fields, such as agriculture and water resource management in a region [3]. The importance of rainfall analysis is evident in these two sectors, as both are highly dependent on rainfall patterns for their sustainability and efficiency. In the agricultural sector, rainfall determines the planting time and the types of crops that are suitable to prevent crop failure. Ironically, Java Island, which is the main producer of staple food commodities, has not been able to achieve maximum production levels due to various issues such as climate change. Fluctuations in crop productivity must always be monitored as part of efforts to meet food needs in anticipation of significant population growth [4].

Java Island, as part of the Indonesian Maritime Continent, is one of the regions on the equator that is vulnerable to the impacts of climate change, with highly complex atmospheric and oceanic dynamics that play a significant role in the global climate cycle [5]. Additionally, the climate conditions on Java Island exhibit temporal variations influenced by changes in the rainy and dry seasons. This spatial and temporal diversity makes Java Island an ideal region for analysis. The island also makes a significant contribution to the national agricultural sector, making an understanding of rainfall patterns crucial for supporting food security [6]. A method is needed that can group rainfall data based on similarity in patterns and handle the diversity and uncertainty in the data. Therefore, a clustering approach is used in this study. Clustering is a technique for grouping data into several clusters based on the similarity of characteristics between data points [7]. Rainfall data in the form of a time series is difficult to analyze directly due to seasonal fluctuations and differences in patterns between regions, so selecting the appropriate distance measure is crucial in the clustering process. Therefore, a distance measure more suited to the characteristics of such data is needed. Distance measurement in time series clustering analysis uses the Dynamic Time Warping (DTW) distance.

DTW is an algorithm that determines the optimal bending path between two time series, so that the output consists of bending path values and the distance between the two series. Euclidean distance is a type of distance measurement in cluster analysis used to calculate the distance between data objects and cluster centers [8]. DTW is suitable for data that shows time shifts in phenomenon patterns, such as rainfall, which can be influenced by various external factors. DTW enables nonlinear alignment between two time series, so that rainfall patterns can still be accurately compared even if there are time shifts in events [9]. This study uses the non-hierarchical Fuzzy C-Means method to cluster regions based on rainfall. This method was chosen for its flexibility in handling data with uncertainty and overlapping between clusters. However, there are difficulties in applying DTW distance to the Fuzzy C-Means algorithm, which operates in a vector space with a symmetric distance matrix, such as Euclidean, while Dynamic Time Warping (DTW) is known to be a robust, outlier-insensitive alternative to other distance measures such as Euclidean distance or Manhattan distance. 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 [10].

In this study, to address the problem, Multidimensional Scaling (MDS) was used to convert DTW into feature vectors, so that the MDS results could be applied to the Fuzzy C-Means algorithm with Euclidean distance between points and cluster centers. The Fuzzy C-Means method is a fuzzy logic-based clustering algorithm that offers high flexibility in grouping data patterns [11]. The objective of this study is to analyze the application of the Fuzzy C-Means algorithm using Dynamic Time Warping (DTW) and to evaluate the clustering results using the Silhouette coefficient. The results of this study will be visualized through GeoJSON and a database. The results of rainfall clustering will identify the most optimal areas based on rainfall intensity levels and can be used as an estimate to determine which areas are prone to natural disasters based on rainfall intensity.

2. RESEARCH METHODS

2.1 Dynamic Time Warping

Dynamic Time Warping is a generalization of the classic algorithm for comparing discrete measures with sequences of continuous values. Data fluctuates and varies over time, so clustering methods can no longer use simple Euclidean distance to measure the proximity between objects. Euclidean distance is suitable for functions of data that do not change over time or static data. Meanwhile, in forming the proximity matrix between objects in time series data, it is also necessary to calculate the distance between time series that are dynamic in nature, so a different clustering method from static data is required. Therefore, an alternative measurement method that can be used is Dynamic Time Warping [12]. Dynamic Time Warping is used to measure distance by finding the optimal bending path between two time series data [8]. Suppose there are two time series of data, namely $x = x_1, x_2, \dots, x_n$ and $y = y_1, y_2, \dots, y_m$. The optimal bending path is determined using a cost matrix $\boldsymbol{C}$ of size $n \times m$ with each element (i, j) for $i = 1,2, \dots, n$ and $j = 1,2, \dots, m$. Matrix $\boldsymbol{C}$ is shown in Eq. (1).

$$\boldsymbol{C} = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,j} \\ \vdots & \vdots & \ddots & \vdots \\ c_{i,1} & c_{i,2} & \cdots & c_{i,j} \end{bmatrix}. \tag{1}$$

For the element $(1,1)$, the cost matrix $\boldsymbol{C}$ is obtained from the following Eq. (2).

$$c_{1,1} = w_{i,j} = w_{1,1}, \tag{2}$$

with $w_{i,j} = |X_i - Y_j|$.

However, the other elements (i, j) of the cost matrix $\boldsymbol{C}$ can be calculated using Eq. (3).

$$c_{i,j} = \begin{cases} w_{i,j} + c_{i,j-1}, & i = 1, \\ w_{i,j} + c_{i-1,j}, & j = 1, \\ w_{i,j} + min\{c_{i-1,j}, c_{i,j-1}, c_{i-1,j-1}\}, & others, \end{cases} \tag{3}$$

where $w_{i,j}$ is the difference between X_i and Y_j. In the cost matrix $\boldsymbol{C}$, the optimal warping path can be determined from the element (n, m) to element $(1,1)$ by considering the minimum values of $c_{i-1,j}, c_{i,j-1}, c_{i-1,j-1},$ as described in Eq. (2). The next step after obtaining the optimal bending path is to calculate the DTW distance, which is done using Eq. (4).

$$DTW(x, y) = c_{n,m}, \tag{4}$$

with:

$DTW(x, y)$: Dynamic Time Warping distance;

$c_{n,m}$: Element (n, m) of the cost matrix $\boldsymbol{C}$.

2.2 Multidimensional Scaling

Multidimensional scaling is a statistical method for mapping data from a high-dimensional space to a lower-dimensional space (such as 2D or 3D) with the aim of preserving as much distance or similarity between data points as possible [13]. The main objective of MDS is to represent data in a simpler form (e.g., from 100 dimensions to 2 or 3 dimensions) while maintaining the relationships between data points (distance or similarity). In the context of community detection in networks, the MDS method is used to map each node in the network to a lower-dimensional space to make the community detection process more efficient. MDS strives to maintain the distances between nodes as reflected in the distance matrix $\boldsymbol{D} \in \mathrm{R}^{n \times n}$ where each element d_{ij} indicates the distance between the nodes i and j. The steps for calculating Multidimensional Scaling are as follows: use the centering matrix $\boldsymbol{H}$, which is idempotent, to eliminate the average effect of the data so that the data has an average of zero [13]. The centering matrix $\boldsymbol{H}$ is defined by Eq. (5).

$$\boldsymbol{H} = \boldsymbol{I} - \frac{1}{n} 11^{\mathrm{T}}, \tag{5}$$

with:

$\boldsymbol{I}$: Identity matrix of size $n \times n$;

1 : an $n \times 1$ column vector with all elements equal to 1.

This H matrix serves to reduce the average influence and adjust the data so that the mapping process to a lower-dimensional space can be performed more effectively. The inner product matrix ($\boldsymbol{J}$) obtained from the distance matrix $\boldsymbol{D}$ can be calculated using Eq. (6).

$$\boldsymbol{J} = -\frac{1}{2}\boldsymbol{H}\,(\boldsymbol{D}\ o\ \boldsymbol{D})\boldsymbol{H}, \tag{6}$$

with:

o : Hadamard product (multiplication between matrix elements);
$\boldsymbol{H}$: The result of the centering matrix;
$\boldsymbol{D}$: Matrix of distance calculation results.

In this case, matrix $\boldsymbol{D}$ is an $n \times n$, distance matrix, where each element d_{ij} indicates the distance between the nodes i and j in the network. This matrix serves as the basis for calculations in the Multidimensional Scaling (MDS) method, as it reflects the degree of similarity or proximity between nodes. By maintaining this distance structure, MDS aims to produce a visual representation in a low-dimensional space without losing important information about the relationships between nodes. Furthermore, based on the Singular Value Decomposition (SVD) theorem for proof, matrix $\boldsymbol{D}$ can be decomposed using Eq. (7).

$$\boldsymbol{S} = \boldsymbol{B}^{-1}\,\boldsymbol{\Lambda}\boldsymbol{B}, \tag{7}$$

with:

$\boldsymbol{\Lambda}$: $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{D}$, sorted from largest to smallest, namely $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$;
$\boldsymbol{B}$: eigenvector matrix of $\boldsymbol{J}$.

The number of dimensional reductions produced depends on the number of positive eigenvalues found in the eigen decomposition calculation based on the selected order of eigenvalues. By selecting the top p vectors based on the positive eigenvalues of matrix $\boldsymbol{D}$ as the axes of the low-dimensional space, the data is then projected onto these axes, resulting in an approximate representation of the nodes in p-dimensional space, where $p < n$. In Multidimensional Scaling (MDS), this is used to transform the feature vectors into dimensions using Eq. (8) below.

$$\boldsymbol{X} = \boldsymbol{B}\boldsymbol{\Lambda}^{\frac{1}{2}} \tag{8}$$

With:

$\boldsymbol{X}$: Input matrix in the initial iteration process of Fuzzy C-Means;
$\boldsymbol{B}$: Eigenvector matrix of matrix $\boldsymbol{J}$;
Λ: Selected eigenvalue.

The metric used to measure the difference between the original distance dij and the projected distance d_{ij} is Kruskal's Stress, commonly known as Stress-1. This metric is used to evaluate the suitability of the distance structure in the original data with the representation of the projected data in low-dimensional space [14]. The general formula for Stress-1 is written in Eq. (9).

$$Stress = \sqrt{\frac{\sum_{i<j}\left(d_{i,j} - \hat{\delta}_{i,j}\right)^2}{\sum_{i<j} d_{i,j}{}^2}}. \tag{9}$$

With:

d_{ij} : actual distance between objects i and j;
d_{ij} : projected distance between objects i and j.

A smaller stress value indicates that the projection result is more accurate. In general, a stress value < 0.1 is considered good, and a stress value < 0.05 is considered very good.

2.3 Fuzzy C-Means

Fuzzy C-Means is a data clustering technique in which the presence of each data point in a cluster is determined by a specific value or degree of membership. This method uses fuzzy clustering, allowing each data point to be a member of several clusters with varying degrees of membership for each cluster. In fuzzy theory, the membership of a data point is not explicitly stated as 1 (member) or 0 (non-member), but rather as a membership degree ranging from 0 to 1 [15]. The steps of the FCM method are as follows:

1. Enter data in the form of an $(n \times m)$ where n denotes the number of observations and m denotes the number of variables for each observation. The element X_{ij} represents the i-th observation $(i = 1,2,\dots,n)$ for the j-th variable $(j = 1,2,\dots,m)$.
2. Determine the number of clusters to be formed $(c \geq 2)$, the exponent weight $(w > 1)$, the maximum iteration (MaxIter), the expected minimum error (ε), the initial objective function $P_0 = 0$, and the initial iteration $t = 1$.
3. Generate random numbers $(\mu_{ik}, , i = 1,2 \dots, n ; k = 1,2,\dots,c)$, as an element of the initial partition matrix $\boldsymbol{U}$ of size $n \times m$) shown in Eq. (10).

$$U = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1c} \\ \vdots & \ddots & & \vdots \\ \mu_{n1} & \mu_{n2} & \cdots & \mu_{nc} \end{bmatrix}. \tag{10}$$

4. Calculate cluster center k : V_{kj}, with $k = 1,2,\dots,c$, and $j = 1,2,\dots,m$ using Eq. (11).

$$V_{\mathrm{kj}} = \frac{\sum_{\mathrm{i}=1}^{n} (\mu_{i\mathrm{k}})^{\mathrm{w}} \cdot \chi_{i\mathrm{j}})}{\sum_{\mathrm{i}=1}^{n} (\mu_{i\mathrm{k}})^{\mathrm{w}}}. \tag{11}$$

5. Calculate the objective function at iteration t, P_t, using Eq. (12).

$$P_t = \sum_{i=1}^{n} \sum_{k=1}^{c} \left(\left[\sum_{j=1}^{\mathrm{m}} (\chi_{ij} - V_{\mathrm{kj}})^2 \right] (\mu_{i\mathrm{k}})^{\mathrm{w}} \right). \tag{12}$$

6. Calculating partition matrix changes following the equation (13)

$$\mu_{i\mathrm{k}} = \frac{\left[\sum_{\mathrm{j}=1}^{\mathrm{m}} (\chi_{ij} - V_{\mathrm{kj}})^2 \right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c} \left[\sum_{\mathrm{j}=1}^{\mathrm{p}} (\chi_{ij} - V_{\mathrm{kj}})^2 \right]^{\frac{-1}{w-1}}}. \tag{13}$$

7. Check the stopping condition if $(|P_t - P_{t-1}| < \xi)$ or (t < maximum iteration) then stop; otherwise, $t = t + 1$ and repeat step 4.

2.4 Silhouette Coefficient

The silhouette coefficient is a method for measuring the quality and strength of clusters. This method combines the concepts of cohesion, which measures the relationships between objects in a cluster, and separation, which measures the distance between different clusters [16]. The closer the silhouette coefficient is to 1, the higher the quality of the grouping in that cluster. Conversely, the closer the Silhouette Coefficient is to -1, the poorer the quality of clustering in that cluster [17]. The calculated Silhouette Coefficient can follow Eq. (14).

$$s(i) = \frac{b(i) - a(i)}{\max\big(a(i), b(i)\big)}. \tag{14}$$

3. RESULTS AND DISCUSSION

3.1 Data Collection

The data used in this study is secondary data with daily rainfall variables from districts/cities in Indonesia from 2021 to 2024, sourced from NASA Power on the website (https://power.larc.nasa.gov/data-access-viewer/). Data was obtained from NASA's Power Data Access Viewer (Prediction of Worldwide

Energy Resources – POWER) website, which provides satellite-based data on weather, climate, and energy. The data used covers 38 provinces and all regencies/cities in Indonesia. One millimeter of rainfall means that one square meter of flat surface area has accumulated one millimeter of water, or one liter of water [18]. There are three rainfall categories defined by Gatlin & Petersen, NASA: rainfall < 2.54 mm/day is classified as light rain, $2.54 \leq r < 7.62$ mm/day is classified as moderate rain, and ≥ 7.62 mm/day is classified as heavy rain. Topographic factors and regional weather systems are crucial in determining the amount and spatial patterns of rainfall in each area. Additionally, the distinct characteristics of different regions can also influence the intensity of rainfall across each province in Indonesia. The data covers 107 districts/cities on the island of Java. The data used in this study, along with its attributes, are shown in the following Table 1.

Table 1. Data Source

Date	Year	District/City	Province	Rainfall (mm/day)
01/01/2021	2021	Kabupaten Ciamis	Jawa Barat	19.36
02/01/2021	2021	Kabupaten Ciamis	Jawa Barat	11.42
03/01/2021	2021	Kabupaten Ciamis	Jawa Barat	8.33
⋮	⋮	⋮	⋮	⋮s
30/12/2024	2024	Ponorogo	Jawa Timur	5.74
31/12/2024	2024	Ponorogo	Jawa Timur	16.43

Data source: https://power.larc.nasa.gov/data-access-viewer/

After the data was collected, time series distance calculations were performed using the Dynamic Time Warping (DTW) method. This method is used to measure the similarity of patterns between two time series, in this case, the monthly rainfall time series from each district/city in East Java Province. The initial step in DTW calculation involves forming a cost matrix $\boldsymbol{E}$. For example, when comparing Malang Regency and Surabaya City, the time series from Malang Regency is represented as X, and the time series from Surabaya City as Y. Each element $E(i,j)$ in the matrix represents the distance between the i-th element of the time series X and the j-th element of the time series Y. After the cost matrix is formed, the process continues by forming a cumulative matrix, considering the minimum value of the three nearest neighboring elements. The result of the DTW process is obtained from the value in the last element of the cumulative matrix.

3.2 Dynamic Time Warping

After collecting daily rainfall data from all districts/cities in Java, the next step is to calculate the distance between time series. This distance measurement is crucial for determining the similarity of rainfall patterns between one region and another. At this stage, the distance between two time series is calculated for each pair of districts/cities using several distance measurement methods, one of which is Dynamic Time Warping (DTW). Each district/city is calculated for its distance relative to all other districts/cities, from the first to the 107th district/city. The results of this process form a distance matrix of size 107×107 for each distance measurement method used. An example of part of the DTW distance matrix between several districts/cities on the island of Java is shown in Table 2.

Table 2. Distance Matrix

	Karanganyar	Wonogiri	Kota Batu	…	Jakarta Barat
Karanganyar	0	2385.84	5386.45	…	6891.38
Wonogiri	2385.84	0	4224.68	…	6609.04
Kota Batu	5386.45	4224.68	0	…	5644.65
⋮	⋮	⋮	⋮	⋱	⋮
Jakarta Barat	6891.38	6609.04	5644.65	...	0

After obtaining the distance matrix between regions, dimensional reduction was performed using the Multidimensional Scaling (MDS) method. MDS aims to represent the DTW distance relationship in a two-dimensional visual form, so that the pattern of proximity between regions can be easily interpreted spatially.

3.3 Multidimensional Scaling

The calculation process of MDS involves eigen decomposition of the distance transformation matrix. The calculation results in 65 positive eigenvalues, which represent valid Euclidean space dimensions. The results of Multidimensional Scaling are shown in Table 3 below.

Table 3. The Results of Multidimensional Scaling

Dimension 1	Dimension 2	Dimension 3	...	Dimension 65
-61.487317	534.524915	39.846388	...	730.755166
-69.236864	292.547047	-79.313234	...	88.813491
234.577795	-159.531890	-848.503189	...	339.932690
⋮	⋮	⋮	⋮	⋮
-180.391856	-57.861875	409.803439	...	189.193007

After obtaining the coordinate representation of the area resulting from the Multidimensional Scaling process with 64 dimensions, the next step is to perform clustering with Fuzzy C-Means.

3.4 Fuzzy C-Means

Implementation of the Fuzzy C-Means algorithm requires initialization. Several initializations in the Fuzzy C-Means process involve data input. At this stage, data reduced from 65 dimensions from Multidimensional Scaling is used as input into the Fuzzy C-Means algorithm with a size of 107×65. There are several parameters with tuning hyperparameters that researchers need to determine first. These parameters are the number of clusters c (2-11), the weight exponent w (1.056), the maximum iteration (Maxiter) (100;1000), and the smallest error tolerance value ε (0.00001;0.000000001). After that, initialize the Fuzzy C-Means parameters as shown in Table 4.

Table 4. Fuzzy C-Means Parameters

Initiation	Score
Number of Clusters (n_cluster)	3
Fuzziness (m)	1.5
Initial objective function	0
Error Tolerance (error)	0.005
Maximum Iterations (maxiter)	1000

The next step is to generate the initial membership matrix. The following is the initial membership matrix that has been obtained.

$$\boldsymbol{U} = \begin{pmatrix} 0.136 & 0.206 & 0.658 \\ 0.224 & 0.414 & 0.362 \\ 0.202 & 0.634 & 0.163 \\ \vdots & \vdots & \vdots \\ 0.091 & 0.087 & 0.823 \end{pmatrix}$$

After calculating the cluster center, the final cluster center results are as below

$$V_{\mathrm{kj}} = \begin{pmatrix} 200.33 & -260.56 & 183.7 & \dots & -497.26 \\ -148.5 & 198.18 & 213.2 & \dots & 459.01 \\ -30.84 & 66.41 & -370.11 & \dots & 0.62 \end{pmatrix}.$$

The next step is to calculate the objective function to show the convergent decrease, and the process ends at the 19th iteration, as shown in Table 5.

Table 5. Objective Function Value

Iteration	Objective function value
1	1235714990
2	1116679980
3	1112703350
⋮	⋮
19	693628387

Based on the calculation of the objective function in Fig. 1, it can be seen that the objective function value decreases sharply in the initial iterations (1–4), indicating that the optimization process is running effectively from the start. After the sixth iteration, the decrease in value begins to slow down and approaches the convergence point. There is a smaller change in the objective function value in the subsequent iterations. From the 8th to the 10th iteration, the objective function value remains nearly unchanged, indicating that the clustering process has achieved optimal convergence. The iteration process stops at the 10th iteration, indicating that the maximum iteration limit has been reached or the algorithm has met the convergence criteria.

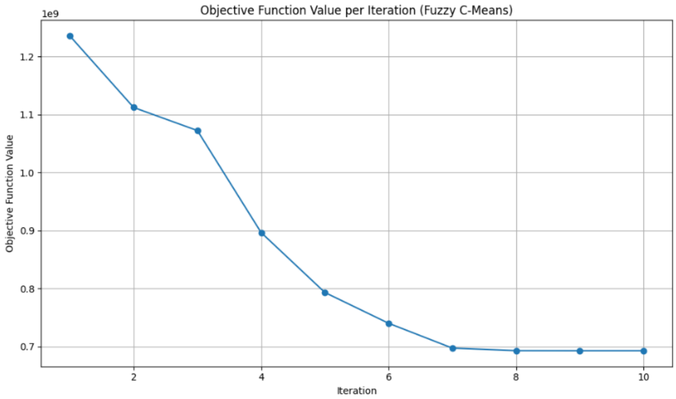

Figure 1. Objective Function Value

Based on the calculation results, the value of iteration 19 minus iteration 18 is -411, which is smaller than the specified error limit of 0.005. Thus, the convergence criteria have been met, so the iteration process is stopped at iteration 19. After that, check the final membership matrix results as a result of the Fuzzy C-Means process.

$$\mu_{ik} = \begin{pmatrix} 0.010608 & 0.983017 & 0.006375 \\ 0.021992 & 0.008667 & 0.969341 \\ 0.187467 & 0.068729 & 0.743803 \\ \vdots & \vdots & \vdots \\ 0.005354 & 0.992445 & 0.002201 \end{pmatrix}.$$

Each data point (region) is assigned to the cluster with the highest membership value. Table 6 shows the results of the clustering.

Table 6. Results of the Clustering

No	Cluster 0	Cluster 1	Cluster 2	Cluster	Regency/City
0	0.060326	0.874011	0.065663	1	Kabupaten Sumenep
1	0.004817	0.008530	0.986653	2	Kabupaten Majalengka
2	0.021281	0.962865	0.015854	1	Kota Mojokerto
3	0.724266	0.202587	0.073148	0	Kabupaten Tegal
4	0.007270	0.011641	0.981089	2	Kota Cimahi
⋮	⋮	⋮	⋮	⋮	⋮
114	0.002069	0.005228	0.992703	2	Kota Administrasi Jakarta Utara
115	0.001056	0.002745	0.996199	2	Kota Administrasi Jakarta Selatan
116	0.936797	0.051305	0.011898	0	Kota Yogyakarta
117	0.971568	0.022168	0.006264	0	Kabupaten Temanggung
118	0.001054	0.002744	0.996201	2	Kota Administrasi Jakarta Timur

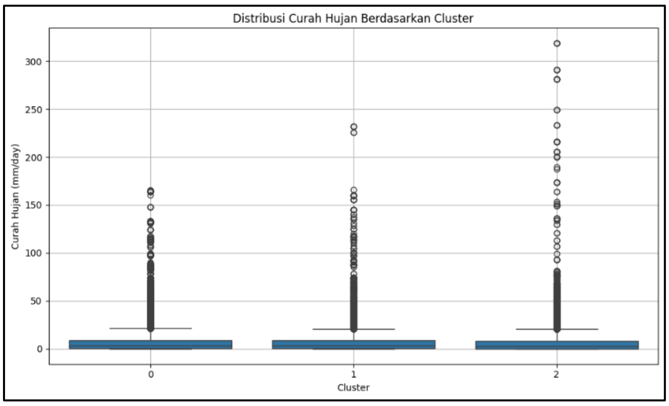

Figure 2. Boxplot Visualization

3.5 Characteristics of Clusters based on Dynamic Time Wrapping

Cluster 0 has a median rainfall in the low range, around 5 mm per day. The data distribution is relatively symmetrical and not too wide, but there are quite a few outliers above 50 mm per day, up to around 170 mm per day. This pattern shows that the areas in this cluster generally experience low to moderate rainfall, but occasionally experience fairly high-intensity rainfall.

Cluster 1 shows a median and data distribution that are almost identical to cluster 0. The difference lies in the higher distribution of outliers, with extreme rainfall values that can reach more than 200 mm per day. This indicates that the areas in cluster 1 have a rainfall pattern similar to cluster 0, but with a slightly greater potential for extreme rainfall under certain conditions.

Cluster 2 has a rainfall median nearly identical to the previous two clusters, but with outlier values that are far more extreme, exceeding 300 mm per day. The boxplot visualization in Fig. 2 shows that this cluster has a greater spread of extreme values than the other clusters. This condition leads to the interpretation that the areas in cluster 2 have the potential for very high or even extreme rainfall, even though on normal days the conditions may resemble those of other clusters.

Overall, the boxplot visualization in Fig. 2 shows that cluster 0 can be categorized as an area with low rainfall, cluster 1 as an area with moderate rainfall, and cluster 2 as an area with high rainfall. The most striking difference among the three clusters lies in the frequency and intensity of outliers, with cluster 2 standing out with a higher potential for extreme rainfall compared to clusters 0 and 1.

3.5 Evaluation with Silhouette Coefficient and Cluster Analysis with ANOVA

Evaluation of the clustering results using the Silhouette Score yielded a value of 0.413184. This value is quite high, indicating that the clusters formed are fairly compact and well separated. Next, cluster analysis will be performed using ANOVA.

Table 7. ANOVA test

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10,387.013	2	5,193.507	43.014	0.000
Within Groups	16,495,758.58	136,621	120.741		
Total	16,506,145.59	136,623			

The hypothesis tested in this study is H_0: There is no difference in rainfall variables in each cluster, and H_1: At least one cluster is different from the others. The test criterion used is to reject H_0, if the calculated F is greater than the table F. With a significance level of $\alpha = 0.05$, the table F value obtained is 2.995.

Based on the results of the ANOVA test in Table 7, the calculated F value is 43.014, which is greater than the table F value (43.014 > 2.995). Thus, H_0 is rejected. This means that there is a significant difference between the average rainfall in each cluster formed using the Fuzzy C-Means method with the Dynamic Time Warping (DTW) distance approach. These results indicate that DTW can effectively capture the similarity patterns of rainfall time series, thereby producing statistically distinct regional groupings.

3.6 Cluster Analysis with Turkey HSD Test

Table 8. Turkey HSD Test

(I) cluster	(J) cluster	Mean Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
0	1	0.08536	0.07876	0.524	-0.0992	0.2700
0	2	0.62129*	0.07994	0.000	0.4339	0.8086
1	0	-0.08536	0.07876	0.524	-0.2700	0.0992
1	2	0.53594*	0.06735	0.000	0.3781	0.6938
2	0	-0.62129*	0.07994	0.000	-0.8086	-0.4339
2	1	-0.53594*	0.06735	0.000	-0.6938	-0.3781

The research hypothesis is formulated as follows: H_0 states that there is no difference in average rainfall between clusters, while H_1 states that there is a difference in average rainfall between clusters. The decision criterion used is to reject H_0 if the significance value (Sig) is < 0.05. Thus, each comparison between clusters will be tested to determine whether the difference is significant or not.

In the comparison between Cluster 0 and Cluster 1, a Sig value of 0.524 was obtained. Since this value is greater than 0.05, the decision made is to fail to reject H_0. This means that there is no significant difference in the average rainfall between the two clusters. In other words, the rainfall characteristics in Cluster 0 and Cluster 1 can be said to be relatively similar.

Next, in the comparison between Cluster 0 and Cluster 2, the Sig value obtained is 0.000. This value is less than 0.05, so the decision is to reject H_0. This indicates that there is a significant difference in the average rainfall between Cluster 0 and Cluster 2. A similar condition is also found in the comparison between Cluster 1 and Cluster 2, where the Sig value of 0.000 indicates a significant difference in rainfall patterns.

Based on the results of the Tukey HSD test in Table 8, it can be concluded that Cluster 2 has a rainfall pattern that is significantly different from that of Cluster 0 and Cluster 1. Conversely, no significant differences were found between Cluster 0 and Cluster 1, so these two clusters can be classified as having relatively similar rainfall characteristics. This finding suggests that the main difference in rainfall data lies in the characteristics of Cluster 2. The results of rainfall clustering are visualized on Java Island using the Dynamic Time Warping (DTW) distance method with the Fuzzy C-Means algorithm, visualized using GeoJSON and a database in Fig. 3.

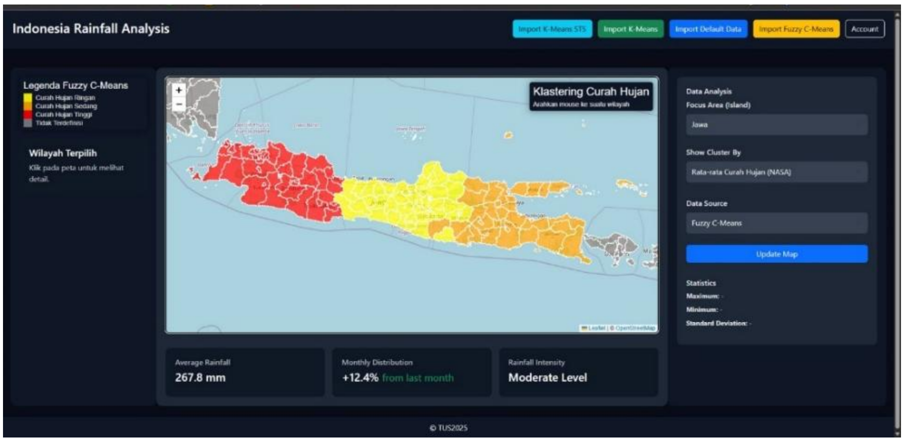

Figure 3. GeoJSON Display

Fig. 3 shows the interface of the Indonesia Rainfall Analysis application, which visualizes the results of rainfall clustering on the island of Java using the Fuzzy C-Means method. The interactive map in the center shows the distribution of areas based on rainfall levels, with a color legend representing the categories of light, moderate, and heavy rainfall. The data used is from NASA's average rainfall data, with an average rainfall value of 267.8 mm, a monthly distribution increase of +12.4% compared to the previous month, and rainfall intensity at a moderate level. Users can filter data based on focus areas and data sources and update the map according to selected parameters.

4. CONCLUSION

Based on the results of clustering analysis using the Fuzzy C-Means algorithm with the Dynamic Time Warping (DTW) distance approach, three groups of regions with different rainfall characteristics were obtained, namely low rainfall (Cluster 0), moderate rainfall (Cluster 1), and high rainfall (Cluster 2). Validation using the Silhouette Score yielded a value of 0.413184, indicating good clustering quality, with clusters that are sufficiently compact and distinct. The ANOVA test indicated significant differences in average rainfall between clusters, and the Tukey HSD post-hoc test reinforced this finding by showing that Cluster 2 was significantly different from the other two clusters, while Clusters 0 and 1 were not significantly different. These findings suggest that the Dynamic Time Warping method is effective in capturing temporal rainfall patterns to form statistically meaningful regional groups.

Author Contributions

Sri Hidayati: Conceptualization, Methodology, Writing-Original Draft, Formal Analysis, Validation. Regita Putri Permata: Data Curation, Resources, Draft Preparation, Writing Review-Editing. Fidi Wincoko Putro: Software, Visualization. All authors discussed the results and contributed to the final manuscript.

Funding Statement

This research was supported by funding from LPPM Telkom University, research scheme for non-external collaboration period 1 wave 1 fiscal year 2025 Telkom University No. 133/LIT06/PPM-LIT/2025.

Acknowledgement

The authors extend sincere thanks to Telkom University for its kind support and assistance that led to this article. Also, thanks to the reviewers and other parties who gave some constructive suggestions to make this article more significant and contribute to academic things.

Declarations

The authors declare no competing interests.

Declaration of Generative AI and AI-Assisted Technologies

The authors declare that no generative AI or AI-assisted technologies were used in the preparation of this manuscript, including writing, editing, data analysis, or the creation of tables and figures**.**

REFERENCES

[1] U. F. Kalude, J. Titaley, and A. Lapu Kalua, "ANALISIS CLUSTERING PERUBAHAN CUACA DAN IKLIM DI KOTA MANADO MENGGUNAKAN METODE K-MEANS ABSTRAKSI," 2022. [Online].

[2] A. Rahim, A. Sulfanita, and A. B. Didi, "AURELIA: JURNAL PENELITIAN DAN PENGABDIAN MASYARAKAT INDONESIA PENGOLAHAN AIR HUJAN SEBAGAI ALTERNATIF PEMENUHAN AIR BERSIH DI PERUMAHAN ANGING MAMMIRI KOTA MAKASSAR," *Aurelia: Jurnal Penelitian dan Pengabdian Masyarakat Indonesia*, vol. 3, 2024.

[3] N. Ananda and L. G. G.A. Monang, "ESTIMASI CURAH HUJAN BULANAN MENGGUNAKAN MULTI-LAYER PERCEPTRON DI KOTA TANGERANG SELATAN MONTHLY RAINFALL ESTIMATION USING MULTI-LAYER PERCEPTRON IN SOUTH TANGERANG CITY," *BULETIN METEOROLOGI, KLIMATOLOGI, DAN GEOFISIKA*, vol. 5, no. 4, pp. 17–26, 2024, [Online].

[4] W. A. Saputro, F. A. Harahap, I. Firdauzi, and R. Dewati, "THE RELATIONSHIP OF RAINFALL TO FOOD PRODUCTIVITY ON THE ISLAND OF JAVA," *Agrisaintifika: Jurnal Ilmu-Ilmu Pertanian*, vol. 9, no. 1, pp. 64–73, Feb. 2025. doi: https://doi.org/10.32585/ags.v9i1.6236

[5] A. Nurlatifah, R. B. Hatmaja, and A. A. Rakhman, "ANALISIS POTENSI KEJADIAN CURAH HUJAN EKSTREM DI MASA MENDATANG SEBAGAI DAMPAK DARI PERUBAHAN IKLIM DI PULAU JAWA BERBASIS MODEL IKLIM REGIONAL CCAM," *Jurnal Ilmu Lingkungan*, vol. 21, no. 4, pp. 980–986, Sep. 2023. doi: https://doi.org/10.14710/jil.21.4.980-986

[6] Ruliyansa, "DAMPAK PERUBAHAN IKLIM TERHADAP SEKTOR PERTANIAN DAN EKONOMI DI DESA SIDOMULYO," *Jurnal Aktual STIE Trisna Negara*, vol. 22, no. 2, pp. 75–80, 2024.

[7] M. N. Hayati, *et al.*, "PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN DATA JUMLAH KEJADIAN DAN DAMPAK BENCANA BANJIR MENGGUNAKAN METODE FUZZY C-MEANS," *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, vol. 6, no. 01, pp. 21–34, 2024,

[8] K. Haulia and N. Imro'ah, "PENERAPAN DYNAMIC TIME WARPING UNTUK PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN NILAI PDRB," 2024.

[9] W. S. Moola, W. Bijker, M. Belgiu, and M. Li, "VEGETABLE MAPPING USING FUZZY CLASSIFICATION OF DYNAMIC TIME WARPING DISTANCES FROM TIME SERIES OF SENTINEL-1A IMAGES," *International Journal of Applied Earth Observation and Geoinformation*, vol. 102, Oct. 2021. doi: https://doi.org/10.1016/j.jag.2021.102405

[10] I. A. Mahmudiati and R. Fajriyah, "GROUPING INDONESIAN PROVINCE FARMERS' TERM OF TRADE USING DYNAMIC TIME WARPING," *Indonesian Journal of Applied Statistics*, vol. 7, no. 2, p. 112, Dec. 2024. doi: https://doi.org/10.13057/ijas.v7i2.94456

[11] M. Rinaldi, A. Rmik, S. Hang, and T. Pekanbaru, "PENYEBARAN MAHASISWA BARU MENGGUNAKAN METODE FUZZY C-MEANS UNTUK MENCARI DAERAH PROMOSI YANG POTENSIAL DISTRIBUTION OF NEW STUDENTS USING THE FUZZY C-MEANS METHOD TO LOOK FOR POTENTIAL PROMOTIONAL AREAS,"

Journal of Information Technology and Computer Science (INTECOMS), vol. 3, no. 2, 2020. Doi: https://doi.org/10.31539/intecoms.v3i2.1524

[12] F. A. W. T. S. Wijaya, E. Prasetyo, and R. F. Tias, "DYNAMIC TIME WARPING PADA METODE K-MEANS UNTUK PENGELOMPOKAN DATA TREND PENJUALAN PRODUK," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 10, no. 2, pp. 100–109, Aug. 2024. doi: https://doi.org/10.25077/TEKNOSI.v10i2.2024.100-109

[13] W. Wang, D. Liu, X. Liu, and L. Pan, "FUZZY OVERLAPPING COMMUNITY DETECTION BASED ON LOCAL RANDOM WALK AND MULTIDIMENSIONAL SCALING," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 24, pp. 6578–6586, Dec. 2013. doi: https://doi.org/10.1016/j.physa.2013.08.028

[14] M. Walesiak and A. Dudek, "MDSOPT-SEARCHING FOR OPTIMAL MDS PROCEDURE FOR METRIC AND INTERVAL-VALUED DATA," 2023.

[15] S. Sandy Prasetyo and A. Rachman Hakim, "PENERAPAN FUZZY C-MEANS KLUSTER UNTUK SEGMENTASI PELANGGAN E-COMMERCE DENGAN METODE RECENCY FREQUENCY MONETARY (RFM)," *JURNAL GAUSSIAN*, vol. 9, no. 4, pp. 421–433, 2020, [Online]. Available: https://ejournal3.undip.ac.id/index.php/gaussian/. doi: https://doi.org/10.14710/j.gauss.v9i4.29445

[16] M. A. Haq, W. Purnomo, and N. Y. Setiawan, "ANALISIS CLUSTERING TOPIK SURVEY MENGGUNAKAN ALGORITME K-MEANS (STUDI KASUS: KUDATA)," Jul. 2023. [Online].

[17] T. Rahmawati, Y. Wilandari, and P. Kartikasari, "ANALISIS PERBANDINGAN SILHOUETTE COEFFICIENT DAN METODE ELBOW PADA PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN INDIKATOR IPM DENGAN K-MEDOIDS," *Jurnal Gaussian*, vol. 13, no. 1, pp. 13–24, Aug. 2024. doi: https://doi.org/10.14710/j.gauss.13.1.13-24

[18] P. N. Gatlin and W. A. Petersen, "A RAIN TAXONOMY FOR DEGRADED VISUAL ENVIRONMENT MITIGATION," 2018. [Online]. Available: http://www.sti.nasa.gov