IJASST
International Journal of
Applied Sciences and Smart Technologies

Volume 08
Issue 01

p-ISSN 2655-8564, e-ISSN 2685-943, pp. 293-320, 2026

Grouping Weekly Weather Based on Weather Elements in Pagaralam by Using *K*-Means Clustering Analysis

Sri Indra Maiyanti[1], Irmeilyana[1*], Putri Nilam Cayo[1],
Dinny Indah Angelia, Angelina[1]

Faculty of Mathematics and Natural Sciences, University of Sriwijaya,
Jln. Lintas Timur Palembang-Prabumulih KM. 32,
Indralaya, 30129, Indonesia
[*]Corresponding Author: irmeilyana@unsri.ac.id

(Received 12-12-2025; Revised 01-02-2026; Accepted 02-03-2026)

Abstract

Weather is the result of the interaction and combination of various atmospheric elements that occurs in a relatively small place or region over a short period of time. This study aims to analyze weekly weather characteristics in Pagaralam using the *K*-Means Clustering method. The data used was weekly weather in 2022 and 2023 with 15 variables, namely Maximum temperature, Minimum temperature, Temperature, Dew, Humidity, Precipitation, Precipitation cover, Wind gust, Wind speed, Wind direction, Sea pressure, Cloud cover, Solar radiation, UV Index, and Moon phase. The analysis process began with data standardization, followed by clustering using *K*-Means Clustering with several *K* values to observe variations in cluster structure. The optimal number of clusters was determined using the elbow and silhouette methods. The best optimal *K* value for each of the 2022 and 2023 data was *K*=3. A small number of weeks in both years had high temperatures and solar radiation and accompanied by lower dew, humidity, precipitation, and cloud cover than other weeks. A small number of weeks in 2022 had low minimum temperature and were also accompanied by lower cloud cover, dew, and humidity, but they had higher wind gusts and wind speeds than other weeks. Meanwhile, a small number of weeks in 2023 had lower temperatures and accompanied by higher cloud cover, wind gusts, and wind speeds than other weeks.

Keywords: Elbow, *K*-Means Clustering, Pagaralam, Silhouette, Weather elements

1 Introduction

Weather is an atmospheric condition that occurs in a relatively small place or region over a short period of time. Weather is the result of the interaction and combination of various atmospheric elements such as temperature, rainfall, humidity, wind speed, and duration of sunlight, where changes in one element will affect other elements [1]. Weather is a condition resulting from differences in temperature and humidity in various regions [2-3]. Changing weather patterns on a global scale, lasting for a very long time (slow pace) and progressing slowly (slow onset) indicate the phenomenon of climate change [4]. Weather changes are characterized by an increase in the Earth's surface temperature and an increase in extreme weather events such as extreme rainfall, which are the main cause of disasters [5]. Global warming arises from increasingly intensive human activities, such as the burning of fossil fuels, increased industrial production, and large-scale deforestation [6], which can disrupt the balance of the Earth's meteorological and geological cycles [3]. Global climate change is affecting the decline in coffee productivity [7-10], including in Indonesia.

The main elements that shape weather include temperature, air pressure, humidity, wind, and rainfall. Sunlight, clouds, lightning, and rainbows also influence the formation of weather in a region [11]. Weather changes also occur in tropical regions like Indonesia, which are highly sensitive to changes in weather elements. In recent years, Indonesia has been impacted by the La Nina phenomenon, which is known to increase rainfall and change wind direction. These changes will continue with varying intensity and affect seasonal patterns and weather stability in various regions [12]. Detailed weather information is highly beneficial because it impacts various sectors of life, such as agriculture, health, tourism, and the economy [13-14].

Pagaralam Municipality, South Sumatra, is one of the areas heavily affected by weather variability due to its location at an altitude of 500-3,000 meters above sea level and its unique microclimate. Weather instability in this region can hamper the coffee cultivation and post-harvest processes, while in the tourism sector [15-17], so understanding the weather characteristics in Pagaralam is important to maintain the sustainability of these sectors [18].

Assessment of the drought threat index in Dempo Tengah District, Pagaralam, using GIS shows that most areas are included in the high threat zone [19]. However, in this study, the weather element data used only rainfall within 1 month and is less visually descriptive. Based on interviews conducted in [20-26], coffee farmers in Pagaralam stated that the decline in coffee production was mostly due to the influence of weather.

Information on time groups dominated by high or low values of certain weather elements can represent the characteristics of general weather conditions for 1 harvest period. Based on the mean difference test in [27], the variables of temperature, maximum temperature, minimum temperature, dew point, rainfall coverage, and cloudiness in Dempo Tengah District in 2021 were significantly different than in 2022. Likewise, the test results for 2 time periods from 2022 to 2024 showed significant differences for temperature, dew, humidity, wind conditions, cloud cover, and sea level pressure [28].

Pagaralam's geographical conditions, which are sensitive to increasingly complex weather changes, require analytical methods capable of explaining the interrelationships between weather elements. Biplot analysis can map the interrelationships between weather elements, simplifying the interpretation of complex weather data into more concise and clear information [29]. Weather data in Pagaralam changes daily, necessitating analytical methods to simplify and display the main patterns. The simplification focuses on weekly weather patterns to make inter week variations more visible. Clustering, an unsupervised data mining method, is used to group data and consists of hierarchical and non-hierarchical clustering. Clustering is a data mining method that can be used to extract valuable information by grouping data into clusters with similar characteristics. Examples of non-hierarchical clustering methods are *K*-means clustering [30] and *K*-medoids clustering [31].

Combining data using *K*-Means Clustering can improve data quality [32]. An example of the application of *K*-Means Clustering was to assess the effectiveness of language learning techniques on students by grouping pre-test and post-test data [33], but the analysis results were not supported by cluster plot visualization. *K*-Means Clustering was also used to group commodity priorities for food policy [34]. However, the determination of *K* did not use statistical justification methods, such as the Elbow method or the Silhouette Index and DBI (Davies-Bouldin Index) value.

Applications of *K*-Means Clustering include Pramoedya, et al. [35] who grouped weather conditions in the Kediri region and obtained three optimal clusters based on the DBI value. Cahaya, et al. [36] grouped rice potential in 35 sub-districts in Pagaralam in 2020 to 2022 period into three rice potential clusters: low, medium, and high. Another study by Rahman, et al. [37] combined *K*-Means Clustering and Random Forest to group and classify annual weather data in Cilacap City using the silhouette and elbow methods. However, in these studies, weather grouping generally still focuses on regions (spatially) or uses annual data. Even though changes in weather conditions often occur rapidly from week to week.

The purpose of this study is to analyze weekly weather clustering in Pagaralam Municipality 2022 and 2023 based on weather elements using the K-Means Clustering method. The results obtained by identifying the characteristics of each weather cluster can provide basic information that can be used in practical activity planning, such as developing agricultural cultivation calendars and other activities that depend on weather conditions.

2 Material and Methods

The data used in this study is secondary data obtained from the website visualcrossing.com. The data contains daily weather information for Pagaralam Municipality, South Sumatra, in 2022 and 2023. The objects in the data matrix are weekly periods, from January 1 to December 31 in a year, resulting in 53 weeks, with the 53rd week consisting of only one or two days. Variables and their notations and units include: Maximum Temperature (Tmax; X_1; in ^oC), Minimum Temperature (Tmin; X_2; in ^oC), Mean temperature (Temp; X_3; in ^oC), Dew point (Dew; X_4; in ^oC), Relative humidity (Hum; X_5; in %), Precipitation (Precip; X_6; in mm), Precipitation Cover (Pcover; X_7; in mm), Wind Gust (Wgust; X_8; in km/h), Wind Speed (WSpeed; X_9; in km/h), Wind Direction (WDir; X_{10}; in $degrees\ from\ North$), Sea level pressure (Seapress; X_{11}; in mb), Cloud cover (CCover; X_{12}; in %), Solar radiation (SolRad; X_{13}; in W/m^2), UV Index (UVI; X_{14}; $between\ 0\ and\ 10$), Moonphase (Moophase; X_{15}; $between\ 0\ and\ 10$). Two data matrices used were data on 15 weather elements in Pagaralam in 2022 and 2023.

The following were the steps taken in this research:

1. Describe the data for each variable.

2. Construct a data matrix for 2022 and 2023. For example, a data matrix consisting of n objects (i.e. the number of weeks) and p variables (i.e. the number of weather elements) is denoted as $\boldsymbol{X} = \left(x_{ij}*\right)$; where $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, p$.
3. Standardize the data using Equations (1), (2), and (3):

$$z_{ij} = \frac{x_{ij} * - \bar{x}_j}{s_j} \quad (1)$$

$$\bar{x}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n} \quad (2)$$

$$s_j = \sqrt{\frac{\sum_{i=1}^{n}\left(x_{ij} * - \bar{x}_j\right)^2}{n-1}} \quad (3)$$

where

z_{ij} : Standardized value of the $i-$th object on the $j-$th variable;

$i = 1, 2, \dots, n; j = 1, 2, \dots, p$

$x_{ij}*$: Initial value of the $i-$th on the $j-$th variable

$\bar{x}_j$: Average value of the $j-$th variable

s_j : Standard deviation value of the $j-$th variable

4. Perform *K*-Means Clustering analysis [38-40]
   a. Determine the number of initial clusters *K* to be used in the analysis process.
   b. Determine the initial centroid $\boldsymbol{c_k}$.
   c. Calculate the distance of each object i to centroid $\boldsymbol{c_k}$ using Equation (4):

$$d(\boldsymbol{x_i}, \boldsymbol{c_k}) = \sqrt{\sum_{j=1}^{p}\left(x_{ij} - c_{kj}\right)^2} \quad (4)$$

   where

   $d(\boldsymbol{x_i}, \boldsymbol{c_k})$: Euclidean distance of the $i-$th object to the $k-$th cluster centroid;

   $k = 1, 2, \dots, K$

   x_{ij} : Standardized value of the $i-$th object to the $k-$th variable

   c_{kj} : Centroid value of the $k-$th for $j-$th variable

   d. Group each object i into the $k-$th cluster with the closest distance.

e. Recalculate the centroid position $\boldsymbol{c_k}$ based on the average value of the members in cluster k using Equation (5).

$$c_{kj} = \frac{\sum_{i=1}^{n_k} x_{ij}}{n_k} \tag{5}$$

where

c_{kj} : Centroid value of the $k-$th for the $j-$th variable

x_{ij} : Standardized value for the $i-$th object for the $j-$th variable in the $k-$th cluster; $i = 1, 2, 3, \dots, n_k$

n_k : Number of objects in the $k-$th cluster

f. Repeat the calculation steps until the centroid position $\boldsymbol{c_k}$ remains unchanged (converges).

5. Determine the optimal number of clusters, K, based on the SSE (Sum of Squared Errors in elbow method) and silhouette values. The elbow method is a method for determining the best number of clusters in a model by looking at the optimal point from the comparison between the number of clusters and the data variation value. The silhouette coefficient method combines two metrics, namely cohesion which assesses the proximity between objects within a cluster, and separation which measures the distance between one cluster and another [40]. The silhouette coefficient ranges from -1 to 1, with higher values indicating better cluster formation [41]. The SSE and silhouette values are calculated by using Equation (6) and Equation (7), respectively.

$$SSE = \sum_{k=1}^{K} \sum_{x_i \in c_k} \sum_{j=1}^{p} \left(x_{ij} - c_{kj}\right)^2 \tag{6}$$

$$s(i) = \frac{b(i) - a(i)}{\max\left(a(i) - b(i)\right)} \tag{7}$$

Where:

$s(i)$: Silhouette value for the $i-$th object

$a(i)$: The average value of the distance between the $i-$th object and all other objects in the same cluster

$b(i)$: The average value of the distance between the $i-$th object and objects in other nearby clusters

$\max\left(a(i)-b(i)\right)$: Maximum value between $a(i)$ and $b(i)$, used as a divisor so that the result is in the range -1 to 1

6. Interpretation of results.

Data processing and analysis using software Minitab 19, SPSS 24, and R Studio version 4.3.1.

3 Results and Discussions

The data aggregation process was carried out to convert daily weather data into weekly data. Construct the 2022 and 2023 data matrices, respectively, as $\boldsymbol{X} = \left(x_{ij}*\right)$ with $i = 1, 2, \dots, 53$ and $j = 1, 2, \dots, 15$. The next step is to conduct a descriptive analysis to provide an initial overview of the data distribution. This analysis includes calculating the minimum, maximum, median, mean, and standard deviation values. Descriptive statistics for variables in 2022 as shown in Table 1.

Data standardization on Table 2 was performed using Equation (1) based on average and standard deviation values in Table 1. The standardized data used in a *K*-Means Clustering analysis to group weekly weather conditions based on their weather elements.

Table 1. Descriptive statistics for variables in 2022

Variable	Average	Minimum	Maximum	Median	Standard Deviation
X_1	78,8518	76.57	81.51	78.71	1.2118
X_2	66.1369	63.80	67.93	66.13	0.7801
X_3	71.3493	70.29	73.57	71.30	0.8060
X_4	66.9453	64.46	69.01	67.03	0.9845
X_5	86.8577	79.03	91.80	87.14	2.9053
X_6	0.3794	0.07	0.99	0.33	0.2206
X_7	51.5837	12.50	92.26	48.81	16.6234
X_8	18.5264	12.77	36.00	16.46	4.9401

Variable	**Average**	**Minimum**	**Maximum**	**Median**	**Standard Deviation**
X_9	4.1057	2.63	8.50	3.79	1.2158
X_{10}	210.3827	107.60	298.20	225.60	49.7460
X_{11}	1012.3191	1010.00	1015.00	1012.00	1.0222
X_{12}	90.9844	56.83	99.64	93.31	8.8680
X_{13}	197.9790	135.90	276.90	192.30	27.0349
X_{14}	7.5148	5.86	9.43	7.29	0.7597
X_{15}	0.4772	0.10	0.87	0.45	0.2202

Table 2. Standardized data in 2022

Week-i	X_1	X_2	X_3	X_4	...	X_{14}	X_{15}
1	0.3935	0.7218	0.5060	0.6940	...	-0.1135	-0.6620
2	-1.0094	0.5203	-1.0183	0.5344	...	-0.4896	-0.9020
3	-0.9386	-0.2854	-1.3196	-0.1766	...	0.4506	0.1361
4	0.7825	0.3189	0.4351	0.0556	...	0.4506	1.2522
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
53	-0.8679	-2.9955	-0.9297	-1.7727	...	0.6386	-0.9410

The number of clusters tested was $K = 2$. Some of the calculation examples were performed manually using Equation (4), while the overall clustering process was completed using RStudio software version 4.3.1. The steps of the *K*-Means Clustering analysis for the 2022 data matrix are explained below.

1. Determine the initial centroid by randomly selecting several data points in Table 3. The centroid for cluster 1 is denoted as $\boldsymbol{c_1}$, derived from the weather data from week 13. Similarly, the centroids for clusters 2 is denoted as $\boldsymbol{c_2}$ derived from weather data from week 45.
2. Perform iteration 1 by calculating the object's distance to the initial centroid using the Euclidean distance in Equation (4). For example, the distance from week 2 to the centroid in cluster 1 is

$$d(\boldsymbol{x_2}, \boldsymbol{c_1}) = \sqrt{(-1.0094 - 2.1971)^2 + (0.5203 + 0.2854)^2 + \cdots + \big(-0.9020 - (1.2262)\big)^2}$$
$$= 9.8346$$

3. The results of the Euclidean distance calculation and cluster grouping are shown in Table 4.
4. Iteration 2

   The centroid is calculated by taking the average value of each variable in each cluster. Based on Table 4, the weekly weather for 2022 consists of 18 weeks in cluster 1 and 35 weeks in cluster 2. The new centroid coordinates c_{kj} are determined using the standardized initial data and Equation (5), with the number of cluster members being $n_1 = 18$ and $n_2 = 35$.

   For example, to calculate the new centroid for variable X_1 in cluster 1 is based on the X_1 values of 18 weeks in cluster 1. These values are based on Table 2.

$$c_{1,1} = \frac{0.7825 + 0.6882 + \cdots + (-1.0919) + (-0.8679)}{18} = 0.7759$$

   New centroid for variable X_1 in cluster 2 is

$$c_{2,1} = \frac{0.3935 + (-1.0094) + \cdots + (-0.9033) + (-0.7264)}{35} = -0.3991$$

5. By the same calculation, new centroid values for each cluster from the 2022 data are shown in Table 5. The Euclidean distance and clustering results from the second iteration can be seen in Table 6.

Table 3. Initial centroid values for 2022

Centroid	X_1	X_2	X_3	$\cdots$	X_{14}	X_{15}
$\boldsymbol{c_1}$	2.1971	-0.2854	2.2784	$\cdots$	2.5190	1.2262
$\boldsymbol{c_2}$	-1.8817	0.1358	-1.1601	$\cdots$	-1.9939	0.0583

Table 4. Euclidean distance and clustering from iteration 1 on 2022 data

Week-i	$\boldsymbol{c_1}$	$\boldsymbol{c_2}$	Minimum distance	Cluster-*k*

1	7.8420	**5.4629**	5.4629	2
2	9.8346	**3.3025**	3.3025	2
3	8.5152	**5.2021**	5.2021	2
4	**6.4918**	6.8082	6.4918	1
⋮	⋮	⋮	⋮	⋮
53	**9.3431**	9.9322	9.3431	1

Note: Bold numbers indicate the minimum distance from week-i to the centroid $\boldsymbol{c_k}$

Table 5. New centroid values for the second iteration from the 2022 data

Centroid	X_1	X_2	X_3	$\cdots$	X_{14}	X_{15}
$\boldsymbol{c_1}$	0.7759	-0.3556	0.6872	$\cdots$	0.9833	-0.0005
$\boldsymbol{c_2}$	-0.3991	0.1829	-0.3534	$\cdots$	-0.5057	0.0002

Table 6. Euclidean distance and clustering in the second iteration of 2022 data

Week-i	$\boldsymbol{c_1}$	$\boldsymbol{c_2}$	Minimum distance	Cluster-k
1	3.5435	**2.6195**	2.6195	2
2	5.3713	**2.3735**	2.3735	2
3	3.9657	**2.3234**	2.3234	2
4	**2.8835**	3.5707	2.8835	1
⋮	⋮	⋮	⋮	⋮
53	**6.8134**	7.9778	6.8134	1

Table 6 still shows differences in cluster members from the previous calculation. Therefore, the iteration continues using the same process until the cluster members are the same as those calculated in the previous iteration (convergence). In this case, the calculation for $K = 2$ on the 2022 data matrix reaches iteration 6. New centroid values for each cluster from the 2022 data are shown in Table 7. The Euclidean distance and clustering results from the second iteration can be seen in Table 8.

Table 7. New centroid values for the sixth iteration from the 2022 data

Centroid	X_1	X_2	X_3	$\cdots$	X_{14}	X_{15}
$\boldsymbol{c_1}$	0.5167	-0.3978	0.4391	$\cdots$	0.7925	-0.0851
$\boldsymbol{c_2}$	-0.3667	0.2823	-0.3116	$\cdots$	-0.5624	0.0604

Table 8. Euclidean distance and clustering in the sixth iteration of 2022 data

Week-i	$\boldsymbol{c_1}$	$\boldsymbol{c_2}$	Minimum distance	Cluster-*k*
1	3.4187	**2.6863**	2.6863	2
2	5.0844	**2.2889**	2.2889	2
3	3.6431	**2.5060**	2.5060	2
4	**2.7653**	3.7774	2.7653	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
53	**6.3493**	8.4069	6.3493	1

Furthermore, the *K*-Means Clustering visualization for *K* = 2 on the 2022 and 2023 data matrices can be seen in Figs. 1 and 2. Based on Fig. 1, the *K*-Means clustering results show two clusters distributed in different positions, although some objects in both clusters are relatively close together. In general, the distribution of each cluster indicates that each group has distinct characteristics.

The number of clusters is evaluated to determine the optimal *K* value using the elbow and silhouette methods. The stage of evaluating the number of clusters using the elbow method is to determine the SSE value for week i in cluster *k* using Equation (6):

$$SSE_{k,i} = \sum_{k=1}^{K} \sum_{x_i \in c_k} \sum_{j=1}^{p} \left(x_{ij} - c_{kj}\right)^2$$

For example, to calculate the SSE value for week 4 in cluster 1 (notated as $SSE_{1,4}$) is based on data of week 4 in Table 2 and centroid of cluster 1 ($\boldsymbol{c_1}$) in Table 7.

$$SSE_{1,4} = \Big((0.7825 - 0.5167)^2 + \big(0.3189 - (-0.3978)\big)^2 + \cdots + (0.4506 - 0.7945)^2 + \big(1.2522 - (-0.0850)\big)^2\Big)$$

Calculation of the SSE value for week 53 in cluster 1 (notated as $SSE_{1,53}$) is based on data of week 53 in Table 2 and centroid of cluster 1 ($\boldsymbol{c_1}$) in Table 7.

$$SSE_{1,53} = \left((-0.8679 - 0.5167)^2 + \left(-2.9955 - (0.3978)\right)^2 + \cdots + (0.6386 - 0.7945)^2 + \left(-0.9410 - (-0.0850)\right)^2\right) = 40.3131$$

So, the SSE for cluster 1 is obtained as follows

$$SSE_1 = SSE_{1,4} + SSE_{1,6} + \cdots + SSE_{1,53}$$
$$= 7.6470 + 14.6696 + \cdots + 40.3131 = 308.0786$$

Figure 1. K-Means clustering for $K = 2$ in 2022

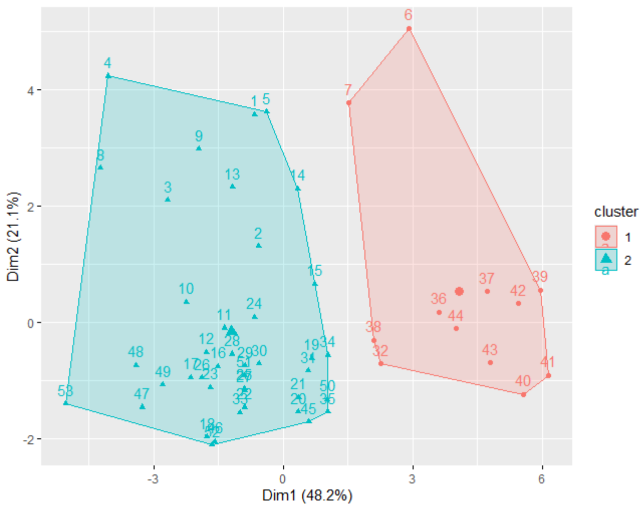

Figure 2. K-Means clustering for $K = 2$ in 2023

Based on manual calculations, the SSE value for cluster 1 was 308.0786. The calculation process was repeated until the optimal clustering result was achieved. For example, at K=2, the SSE value for cluster 1 was combined with the SSE value for cluster 2. Using the same steps, the SSE for cluster 2 was 270, resulting in a total SSE of 577.9574 at K=2. The overall SSE calculation results can be seen in Table 9.

Based on Table 9, the SSE value in the 2022 data matrix experienced a very sharp decline at $K = 4$. At $K > 4$, the percentage change in the SSE value was less than 22. This indicates that increasing the number of clusters at $K > 4$ did not provide optimal clustering. Thus, the elbow method resulted an optimal number of clusters of $K = 4$.

The SSE value in the 2023 data matrix shows the largest decreasing at $K = 3$. At $K > 3$, the percentage change in the SSE value becomes smaller than 23, so increasing the number of clusters to K >3 does not provide optimal results. Thus, the elbow method results the optimal number of clusters at $K = 3$.

Next, the silhouette method was used to measure the proximity between objects within a cluster. The calculation example was performed on objects in week 4 based on the initial data and using Equation (7).

Table 9. SSE calculation results for K-clusters in 2022 and 2023

Number of Clusters (K)	2022 Data Matrix			2023 Data Matrix		
	SSE Value	Change in SSE Value	%age Change	SSE Value	Change in SSE Value	%age Change
2	577.9574	0	0	513.6831	0	0
3	506.0911	71.8663	12	393.4374	**120.2457**	**23**
4	395.7869	**110.3042**	**22**	353.1861	40.2513	10
5	370.5297	25.2572	6	330.4316	22.7545	6
6	340.0241	30.5056	8	273.2691	57.1625	17
7	300.0259	39.9982	12	257.5825	15.6866	6
8	285.2293	14.7966	5	227.3601	30.2224	12
9	259.3365	25.8928	9	210.2731	17.0870	8
10	241.3491	17.9874	7	197.4074	12.8657	6

Note: Bold numbers indicate a significant decrease in SSE.

Based on the results of previous calculations, the week 4 is included in cluster 1. The first step taken is to calculate the distance between objects in the same cluster.

$$d(4,6) = \sqrt{\left(0.7825 - (-1.1626)\right)^2 + \cdots + \left(1.2522 - (-1.1291)\right)^2} = 5.0471$$

$$d(4,53) = \sqrt{\left(0.7825 - (-1.8679)\right)^2 + \cdots + \left(1.2522 - (-1.9410)\right)^2} = 7.6325$$

The average distance between the 4th week and all objects in the same cluster is

$$a(4) = \frac{5.0471 + 3.1511 + 3.1187 + \cdots + 6.4894 + 7.6325}{22} = 4.3903$$

The next step is to calculate the distance between week 4 and other objects in different clusters.

$$d(4,1) = \sqrt{\left(0.7825 - (0.3935)\right)^2 + \cdots + \left(1.2522 - (-0.6620)\right)^2} = 2.9701$$

$$d(4,49) = \sqrt{\left(0.7825 - (-0.9033)\right)^2 + \cdots + \left(1.2522 - (-0.2078)\right)^2} = 4.5553$$

The average distance between the 4th week and all objects in different clusters.

$$b(4) = \frac{2.9701 + 4.7830 + 4.2221 + \cdots + 3.9250 + 4.5553}{31} = 4.6742$$

The silhouette coefficient values for week 4 are as follows.

$$s(4) = \frac{4.6742 - 4.3903}{4.6742} = 0.0607$$

Based on manual calculations, the silhouette coefficient value for week 4 was 0.0607. The calculation process was repeated for all objects to obtain the average silhouette coefficient value at *K*=2. The overall calculation results are shown in Table 10.

According to Table 10, the highest silhouette coefficient value for the 2022 data matrix at *K*=3 was 0.2722. This value is the largest compared to other cluster size, indicating a good separation distance between clusters. Therefore, the silhouette method indicates *K*=3 as the optimal number of clusters.

Table 10. Silhouette coefficient calculation results in 2022 and 2023

Cluster (K)	Silhouette Coefficient Value (s_K)	
	2022 Data Matrix	2023 Data Matrix
2	0.2198	**0.3503**

3	**0.2722**	0.3101
4	0.2130	0.2957
5	0.1719	0.2425
6	0.1683	0.1797
7	0.1650	0.1843
8	0.1598	0.2193
9	0.1563	0.1735
10	0.1688	0.1663

The highest silhouette coefficient value for the 2023 data matrix was at K=2, namely 0.3503. This indicates the largest distance between clusters, so the silhouette method indicates K=2 as the optimal number of clusters. Next, a comparison was conducted to examine the characteristics of each cluster formed. The results of the comparison of cluster members in 2022 and 2023 can be seen in Table 11.

Table 11. Comparison of cluster members in 2022 and 2023

Month	Week	2022 Data Matrix		2023 Data Matrix	
		Silhouette 3 Clusters	Elbow 4 Clusters	Silhouette 2 Clusters	Elbow 3 Clusters
January	Week 1	2	2	2	3
	Week 2	2	4	2	3
	Week 3	2	4	2	3
	Week 4	2	2	2	3
January/February	Week 5	2	4	2	3
February	Week 6	1	3	1	3
	Week 7	2	2	1	3
	Week 8	3	1	2	3
February/March	Week 9	3	1	2	3
March	Week 10	2	2	2	2
	Week 11	2	4	2	2
	Week 12	3	1	2	2

Month	Week	2022 Data Matrix		2023 Data Matrix	
		Silhouette 3 Clusters	**Elbow 4 Clusters**	**Silhouette 2 Clusters**	**Elbow 3 Clusters**
	Week 13	3	1	2	3
March/April	Week 14	3	1	2	3
April	Week 15	2	2	2	2
	Week 16	2	2	2	2
	Week 17	2	2	2	2
April/May	Week 18	2	2	2	2
May	Week 19	2	2	2	2
	Week 20	2	2	2	2
	Week 21	2	2	2	2
May/Jun	Week 22	2	2	2	2
Jun	Week 23	2	4	2	2
	Week 24	2	4	2	2
	Week 25	2	4	2	2
	Week 26	2	4	2	2
Jun/July	Week 27	2	2	2	2
July	Week 28	2	2	2	2
	Week 29	2	2	2	2
	Week 30	2	4	2	2
July/August	Week 31	2	4	2	2
August	Week 32	3	1	1	1
	Week 33	2	2	2	2
	Week 34	2	4	2	2
	Week 35	2	4	2	2
August/Sept.	Week 36	2	4	1	1
September	Week 37	3	1	1	1
	Week 38	2	2	1	1

Month	Week	2022 Data Matrix		2023 Data Matrix	
		Silhouette 3 Clusters	Elbow 4 Clusters	Silhouett e 2 Clusters	Elbow 3 Clusters
	Week 39	2	2	1	1
September/Oct.	Week 40	2	4	1	1
October	Week 41	2	4	1	1
	Week 42	2	4	1	1
	Week 43	2	4	1	1
October/Nov.	Week 44	2	2	1	1
November	Week 45	2	4	2	2
	Week 46	2	4	2	2
	Week 47	1	3	2	2
	Week 48	2	4	2	2
November/Dec.	Week 49	2	4	2	2
December	Week 50	1	3	2	2
	Week 51	2	2	2	2
	Week 52	1	3	2	2
	Week 53	1	3	2	2

Based on Table 11, the 2022 data matrix shows differences in cluster members between the elbow and silhouette methods, indicating that each cluster has a varied data pattern. At K =3, many weekly weather patterns cluster in the same group, so the variation pattern remains quite general. At K =4, the data is more clearly separated, but 2 of 4 clusters have same characteristics. Meanwhile, the other two clusters have the same members and characteristics as the 2 clusters in the $K = 3$ grouping results. This shows that the clustering for $K = 3$ is quite appropriate in distinguishing weekly weather characteristics more specifically. Comparison of cluster centroids can be seen in Table 12.

In Table 12, the results of the K=3 clustering of the 2022 show that each cluster has distinct weather characteristics, as follows:

(i) Cluster 1 consists of 5 weeks, predominantly characterized by Tmin, Dew, WGust, Wspeed, and WDir. This condition illustrates that these weeks have stronger winds, with minimum temperatures and dew points tending to be lower than other weeks.

(ii) Cluster 2 consists of 41 weeks, indicating that most weeks in 2022 tended to be less dominated by the more "extreme" weather elements.

(iii) Cluster 3 consists of 7 weeks, predominantly characterized by the Tmax, Temp, SolRad, and UVI variables, as well as the Dew, Hum, Precip, PCover, and CCover variables. This condition illustrates that these weeks experience warmer temperatures, but lower precipitation, dew, and humidity than other weeks.

Table 12. Comparison of cluster centroids in 2022 data

Variable	$K = 3$ Cluster Centroid			$K = 4$ Cluster Centroid			
	1	2	3	1	2	3	4
Tmax	-0.7547	-0.1310	**1.3063**	**1.3063**	0.5650	-0.7547	-0.7938
Tmin	**-0.9373**	0.2497	-0.7929	-0.7929	0.5075	**-0.9373**	0.0041
Temp	-0.6461	-0.0811	**0.9364**	**0.9364**	0.6850	-0.6461	-0.8107
Dew	**-1.4941**	0.3929	**-1.2338**	**-1.2338**	0.4691	**-1.4941**	0.3202
Hum	**-0.9511**	0.4095	**-1.7190**	**-1.7190**	-0.1034	**-0.9511**	0.8980
Precip	-0.6921	0.2607	**-1.0325**	**-1.0325**	-0.3843	-0.6921	0.8750
PCover	-0.4963	0.3030	**-1.4201**	**-1.4201**	-0.2958	-0.4963	0.8732
WGust	**2.4024**	-0.2918	-0.0066	-0.0066	-0.2592	**2.4024**	-0.3229
WSpeed	**2.4723**	-0.3311	0.1733	0.1733	-0.3883	**2.4723**	-0.2766
WDir	**1.0280**	-0.1863	0.3568	0.3568	-0.1618	**1.0280**	-0.2096
Seapress	0.5962	-0.1632	0.5303	0.5303	-0.4373	0.5962	0.0977
CCover	0.6790	0.1456	**-1.3376**	**-1.3376**	-0.1819	0.6790	0.4574
SolRad	0.2682	-0.3217	**1.6926**	**1.6926**	0.1978	0.2682	-0.8164
UVI	-0.0759	-0.2878	**1.7400**	**1.7400**	0.1403	-0.0759	-0.6956
MooPhase	-0.3933	-0.0091	0.3345	0.3345	0.0485	-0.3933	-0.0641

Number of objects	5	41	7	7	20	5	21

Note: Bold numbers indicate the dominant variables characterizing the cluster.

Furthermore, the results of the K=4 clustering of the 2022 show that:

(i) Cluster 1, consisting of 7 weeks, tends to have the same characteristics as cluster 3 at K=3, namely higher Tmax, Temp, SolRad, and UVI values, but lower Dew, Hum, Precip, PCover, and PCover values compared to other weeks.

(ii) Cluster 2, consisting of 20 weeks, and cluster 4 consisting of 21 weeks, each tend not to be dominated by more "extreme" weather elements. These two clusters represent a separation from cluster 2 at K=3.

(iii) Cluster 3, consisting of 5 weeks, has the same characteristics as cluster 1 at K=3, namely predominantly characterized by Tmin, Dew, WGust, Wspeed, and WDir.

Based on a comparison of the clustering results at the two K values in 2022 data, the optimal number of clusters was set at K =3. Although the elbow method results show the maximum SSE change at K=4, the resulting cluster characteristics tend to be similar to those obtained with K=3. The silhouette coefficient value suggests K=3 as the best choice. Based on the K=3 clustering results in Table 11, there are only 12 weeks with slightly more extreme weather than the others. These weeks, from late February to March, early August, and early September, are dominated by higher temperatures, solar radiation, and UVI, but those are as opposed with dew, humidity, and precipitation. Meanwhile, in early February, mid-November, and December, the weather is dominated by wind, dew and humidity.

Based on Table 11, the clustering results for the 2023 data matrix K=2 and K=3 do not result the same membership pattern. All members of one cluster at K=2 are separated into two clusters. Thus, the three resulting clusters have distinct characteristics. This indicates that some weeks exhibit inconsistent patterns that change as the number of clusters increases. Visualization for K = 2 on 2023 data matrix can be seen in Fig. 2.

Table 13 shows the results of clustering in the K=2, indicating that each cluster has distinct weather characteristics, as follows:

(i) Cluster 1, consisting of 12 weeks, tends to be characterized by the Tmax, Temp, WSpeed, SolRad, and UVI variables, as well as the Dew, Hum, Precip, PCover, and CCover variables. These two variable groups have opposite values. These conditions indicate that the weeks in this cluster are characterized by higher temperature, wind speed, solar radiation, and UVI, but also lower Dew, humidity, precipitation, and cover precipitation compared to other weeks.

(ii) Cluster 2 consists of 41 weeks, so that most of the weeks in 2023 are likely not to be dominated by more 'extreme' weather elements.

Table 13. Comparison of cluster centroids in 2023

Variable	$K = 2$		$K = 3$		
	Cluster Centroid		Cluster Centroid		
	1	2	1	2	3
Tmax	**1.2543**	-0.3671	**1.5238**	-0.1538	**-0.9378**
Tmin	-0.6742	0.1973	-0.5317	0.5291	**-1.0557**
Temp	**0.9317**	-0.2727	**1.2325**	0.0347	**-1.2214**
Dew	**-1.5566**	0.4556	**-1.6116**	0.6248	-0.3526
Hum	**-1.5863**	0.4643	**-1.7610**	0.4321	0.3438
Precip	**-0.8962**	0.2623	**-0.8739**	0.2439	0.0848
PCover	**-1.2357**	0.3617	**-1.2383**	0.3025	0.2458
WGust	0.6768	-0.1981	0.3088	-0.6008	**1.4672**
WSpeed	**1.0407**	-0.3046	0.6955	-0.6193	**1.1694**
WDir	-0.5482	0.1605	**-0.9465**	-0.1803	**1.3850**
Seapress	0.1622	-0.0475	0.4084	0.0006	-0.3729
CCover	**-1.0987**	0.3216	**-1.4463**	0.1049	**1.0096**
SolRad	**1.3950**	-0.4083	**1.5013**	-0.3635	-0.3073
UVI	**1.3662**	-0.3999	**1.4703**	-0.3901	-0.2017
MooPhase	0.3390	-0.0992	0.2151	-0.0180	-0.1431
Number of Objects	12	41	10	32	11

Based on Table 13, the results of the $K = 3$ clustering show that each cluster has distinct weather characteristics, as follows:

(i) Cluster 1 which consists of 10 weeks has the same characteristics as Cluster 1 at $K = 2$.

(ii) Cluster 2, consisting of 32 weeks, has the same characteristics as Cluster 2 at $K = 2$. Most weeks in 2023 tend not to be dominated by more extreme weather elements.

(iii) Cluster 3, consisting of 11 weeks, is predominantly characterized by the Tmax, Tmin, and Temp variables, as well as the Wgust, WSpeed, WDir, and CCover variables. These two groups of weather elements are inversely related. These two groups of weather elements are inversely related. It can be said that a small part of the weeks in 2023, low temperature conditions are characterized by high cloud cover, wind speed, and wind gust.

In cluster 1 and cluster 3, the wind direction is opposite to the North.

In the 2023 data, the optimal number of clusters using the silhouette coefficient method was $K = 2$, but the difference between $K = 2$ and $K = 3$ was not significant. Meanwhile, the optimal number of clusters using the elbow method was $K = 3$. Based on the $K = 3$ clustering results, the weeks between January and early March and from late March to early April were characterized by low temperatures accompanied by higher cloud cover, wind speed, and wind gusts than other weeks. Meanwhile, in early August, between September and early November, weather conditions tend to have characteristics of high temperature and solar radiation but accompanied by dew, humidity, precipitation, and cloud cover that tend to be lower than in other weeks.

The differences in weather elements that dominate a weekly time span within a 1-year period indicate weather variability in Pagaralam. Differences in weekly clustering results over two consecutive years also indicate changes in weather elements, which can affect the flowering process from cultivated plants to harvest, the harvest period, and crop production. Solar intensity, temperature, humidity, rainfall, and wind speed significantly influence fruit formation. If the weather changes are too extreme and the plants are vulnerable to the weather, then it can lead to pest and disease attacks, which can reduce fruit production and even crop failure.

For example, in Pagaralam's robusta coffee plants, the flowering process, which typically occurs throughout the year, can shift to specific months, thus altering the harvest period. The timing of plant maintenance strategies, such as weed control, pest and disease management, and pruning, can also shift and be adjusted to weather conditions. This agricultural calendar planning strategy represents an adaptive and mitigation management strategy to minimize the risk of climate impacts on crop production. Coffee farmers can also implement strategies such as creating rorak for drainage, proper pruning techniques, planting shade tree species that are more effective in controlling the microclimate, and also more intensively implementing GAP (Good Agricultural Practices) in sustainable agriculture.

4 Conclusions

For both 2022 and 2023 data, the elbow and silhouette methods resulted different optimal K values. However, considering the characteristics of each cluster generated at these two different K values, the best optimal K value for each of the 2022 and 2023 data was K=3. Most weeks in 2022 and 2023 in Pagaralam did not have dominant weather elements, especially from April to July. Furthermore, a small number of weeks in both years had high temperatures and solar radiation and they were accompanied by lower dew, humidity, precipitation, and cloud cover than other weeks.

A small number of weeks in 2022 had low minimum temperature and were also accompanied by lower cloud cover, dew, and humidity, but they had higher wind gusts and wind speeds than other weeks. Meanwhile, a small number of weeks in 2023 had lower temperatures and accompanied by higher cloud cover, wind gusts, and wind speeds than other weeks.

The average weekly weather conditions in Pagaralam during certain periods throughout 2022 and 2023 varied considerably. To overcome the impact of variability in weather elements on the harvest period of cultivated crops, especially coffee, farmers should implement a schedule planning strategy and also appropriate plant maintenance management, so that coffee production can be stable and even increase.

This study used a non-hierarchical clustering method, allowing for further comparison with other clustering methods to determine the optimal number of clusters on a larger data range, such as hierarchical clustering, including single linkage, complete linkage, and average linkage methods. Another example of a clustering method used for further research is Density-Based Spatial Clustering with Noise (DBSCAN), a density-based clustering algorithm that groups adjacent data points (high density) and separates noise (sparse areas) to find irregularly shaped clusters. Gaussian Mixture Models (GMM), a probabilistic unsupervised learnng model, can also be used, assuming that the data is generated from a combination of several different Gaussian distributions.

Acknowledgements

The research of this article was funded Universitas Sriwijaya 2025, in accordance with the Rector's Decree Number: 0027/UN9/LPPM.PT/2025, on September 17, 2025.

References

[1] N. Sunarmi, E. N. Kumailia, N. Nurfaiza, A. K. Nikmah, H. N. Aisyah, I. Sriwahyuni, and S. N. Lailly, "Analisis Faktor Unsur Cuaca Terhadap Perubahan Iklim di Kabupaten Pasuruan pada Tahun 2021 dengan Metode Principal Component Analysis." *Newton-Maxwell Journal of Physics*, vol. 3, no. 2, pp. 56–64, 2022.

[2] S. Marlina, *Dampak Perubahan Iklim pada Kesehatan Masyarakat*. Palangka Raya: NEM, 2022.

[3] T. Mulyaningsih, Y. Purwaningsih, Barokatuminalloh, W. Perwithosuci, and I. A. Sasanti, *Adaptasi Perubahan Iklim dan Ketahanan Pangan*, 1st ed. Yogyakarta: Jejak Pustaka, 2022.

[4] E. Aldrian, M. Karmini, and Budiman, *Adaptasi dan Mitigasi Perubahan Iklim di Indonesia*. Jakarta Pusat: Pusat Perubahan Iklim dan Kualitas Udara BMKG, 2011.

[5] S. Y. Andarini and Sudarti. "Analisis Efek global warming terhadap perubahan iklim,". *Jurnal Phi*, vol. 4, no. 2, pp. 31–38, 2023.

[6] O. R. Pinontoan, O. J. Sumampouw, and J. E. Nelwan, *Perubahan Iklim dan Pemanasan Global*. Yogyakarta: Deepublish, 2022.

[7] S. G. Mbwambo, S. K. Mourice, and A. J. P. Tarimo, “The Impacts of Current Climate Variability on Coffee Production in the Northern and Southern Highlands of Tanzania,” *J. Agric. Sci.*, vol. 14, no. 3, pp. 78–99, 2022.

[8] W. Merga and D. Alemayehu, “Effects of Climate Change on Global Arabica Coffee (Coffea arabica L) Production,” *Greener J. Plant Breed. Crop Sci.*, vol. 7, no. 1, pp. 23–30, 2019, doi: https://doi.org/10.15580/GJPBCS.2019.1.072319143.

[9] E. Rahn, P. Vaastd, P. Läderachb, P. van Astenc, L. Jassognec, and J. Ghazoul, “Exploring adaptation strategies of coffee production to climate change using a process-based model,” *Ecol. Modell.*, vol. 371, pp. 76–89, 2018.

[10] T. R. Moreira, S. F. da Silva, N. B. da Silva, G. M. A. D. A. dos Santos, and A. R. dos Santos, “Global warming and the effects of climate change on coffee production,” in *Food Engineering Series: Quality Determinants In Coffee Production*, 1st ed., T. R. M. Lucas Louzada Pereira, Ed. Switzerland: Springer Cham, 2021, pp. 65–100.

[11] B. Susilo, *Mengenal Iklim dan Cuaca di Indonesia*. Yogyakarta: Diva Press, 2021.

[12] W. N. Harahap, B. Yuniasih, and S. Gunawan, “Dampak La Nina 2021-2022 terhadap Peningkatan Curah Hujan,” *Jurnal Phi: Jurnal Pendidikan Fisika dan Fisika Terapan*, vol. 7, no. 1, pp. 26–32, 2023.

[13] A. A. Lestari, “Dampak Perubahan Iklim Terhadap Sistem Pertanian Berkelanjutan di Wilayah Tropis,” *Literacy Notes*, vol. 2, no. 4, pp. 1–8, 2024.

[14] S. Febriosa, W. S. Pratama, Z. Mahdalena, and Ikhwan, "Analisis Dampak Perubahan Iklim Terhadap Kualitas Lingkungan Hidup dan Kehidupan Sosial Masyarakat," *J. MUDABBIR J. Res. Educ. Stud.*, vol. 5, no. 2, pp. 2211–2214, 2025.

[15] S. Maleachi and F. Christianus, "Analisis Pengelolaan Panen dan Pasca Panen Kopi Sebagai Faktor yang Mempengaruhi Kualitas Biji Kopi Robusta di Kota Pagar Alam," *J. Glob. Ilm.*, vol. 1, no. 7, pp. 482–488, 2024.

[16] D. Pratiyudha, M. Permatasari, and P. K. Karo, "Analisis Penataan Kamar Home Stay di Kota Pagar Alam, Sumatera Selatan," *J. Bisnis Hosp.*, vol. 12, no. 1, pp. 39–55, 2023.

[17] R. A. Qusyairi, M. T. Hendratono, and A. W. Sarwono, "Penerapan Konsep Pariwisata Hijau di Gunung Dempo Kota Pagar Alam Sumatera Selatan," *J-CEKI : Jurnal Cendekia Ilmiah*, vol. 2, no. 4, pp. 385–395, 2023.

[18] J. L. Rindika, S. Subiyanto, and F. J. Amarrohman, "Analisis Geospasial Perkembangan Nilai Ekonomi Kawasan Wisata Kota Pagar Alam Menggunakan Sistem Informasi Geografis," *Jurnal Geodesi Undip*, vol. 9, no. 4, pp. 22-31, Oct. 2020.

[19] K. Pasmah, F. Dhiniati, and B. Azizah, "Penilaian Indeks Ancaman Kekeringan di Kecamatan Dempo Tengah Kota Pagaralam Menggunakan Gis," *Bering's*, vol. 09, no. 01, pp. 23–30, Mar. 2022.

[20] Ngudiantoro, Irmeilyana, M. Nirwan Samsuri, "Binary Logistic Regression Modeling on Net Income of Pagar Alam Coffee Farmers," *International Journal of Applied Sciences and Smart Technologies*, vol. 2, no. 2, pp. 137–156, 2020.

[21] Irmeilyana, Ngudiantoro, M. N. Samsuri, and B. Suprihatin, "Logistic regression model on land productivity of Pagar Alam coffee farming," in *J. Phys.: Conf. Ser. 1943*, 2021, pp. 1–12, doi: 10.1088/1742-6596/1943/1/012135.

[22] Irmeilyana, N. Ngudiantoro, and D. Rodiah, "Correspondence Analysis pada Hubungan Faktor-faktor yang Mempengaruhi Pendapatan Petani Kopi Pagaralam," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 15, no. 1, pp. 179–192, Mar. 2021.

[23] Irmeilyana, Ngudiantoro, S. I. Maiyanti, and I. Febrianti, "Correspondence Analysis to Know Factors Related to the Use of Reducant Herbicide on Pagaralam Coffee Farmers," *Pattimura Int. J. Math.*, vol. 1, no. 2, pp. 69–80, 2022, doi: DOI https://doi.org/10.30598/pijmathvol1iss2pp69-80.

[24] Irmeilyana, N. Ngudiantoro, and S. I. Maiyanti, "Regression Model on Pagaralam Coffee Farmers' Income with the Influence of The Use of Herbicide Reductant Variable," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 16, no. 2, pp. 409–420, Jun. 2022, doi: 10.30598/barekengvol16iss2pp409-420.

[25] Irmeilyana, Ngudiantoro, and S. I. Maiyanti, "Hypothesis testing in the study of the characteristics of Pagaralam coffee farmers as herbicide reductant users," in *AIP Conf. Proc. 2913, 040005*, 2023, pp. 1–13, doi: https://doi.org/10.1063/5.0172023.

[26] Irmeilyana, B. Suprihatin, A. Desiani, O. Dwipurwani, and N. N. Aisyah, "Comparison of Graphical Representation of Simple and Multiple Correspondence Analysis Results on the Category of Factors Related to Farmers Status in Using Reductant Herbicide," in *Proceedings of the 3rd Sriwijaya International Conference on Basic and Applied Sciences, SICBAS 2023*, European Alliance for Innovation, Aug. 2024, pp. 1–12, doi: http://dx.doi.org/10.4108/eai.3-11-2023.2347775.

[27] Irmeilyana, Ngudiantoro, and S. I. Maiyanti, “Application of Two Groups Analysis and Cluster Analysis on comparison of Characteristics of Pagaralam Coffee Farmers Categories,” *Barekeng*, vol. 17, no. 1, pp. 101–112, 2023, doi: https://doi.org/10.30598/barekengvol17iss1year2023.

[28] Irmeilyana, B. Suratama, S. I. Maiyanti, N. Ngudiantoro, B. Suprihatin, and A. Desiani, “Deskripsi dan uji hipotesis data elemen-elemen cuaca menggunakan uji selisih mean dan uji rasio variansi pada Tahun 2021-2022 di dua kecamatan penghasil kopi,” *Jurnal Penelitian Sains*, vol. 27, no. 2, pp. 1–11, Jun. 2025.

[29] H. Venelia, K. Nisa, R. A. Wibowo, M. A. Muda, “Robust Biplot Analysis of Natural Disasters in Indonesia from 2019 to 2021”. *Jurnal Aplikasi Statistika & Komputasi Statistik*, vol. 13, no. 2, pp. 61–68, 2021.

[30] R. Supardi and I. Kanedi, “Implementasi Metode Algoritma K-Means Clustering pada Toko Eidelweis,” *J. Teknol. Inf.*, vol. 4, no. 2, pp. 270–277, 2020.

[31] E. H. S. Atmaja, “Implementation of k-Medoids Clustering Algorithm to Cluster Crime Patterns in Yogyakarta,” *International Journal of Applied Sciences and Smart Technologies*, vol. 1, no. 1, pp. 33–44, 2019.

[32] I. A. Arshella, I W. Mustika, and P. Nugroho, “Improving the Accuracy of Prediction of Dissolved Oxygen and Nitrate Level Using LSTM with K-Means Clustering and Spearman Analysis,” *International Journal of Applied Sciences and Smart Technologies*, vol. 7, no. 2, pp. 277-298, 2025.

[33] N. H. Wulandari and V. Purwayoga, “Cluster Change Analysis to Assess the Effectiveness of Speaking Skill Techniques Using Machine Learning,” *International Journal of Applied Sciences and Smart Technologies*, vol. 7, no. 1, pp. 1–14, 2025.

[34] P. Pangestu, S. Maarip, Y. N. Addinsyah, and V. Purwayoga, “Clustering and Trend Analysis of Priority Commodities in The Archipelago Capital Region (IKN) Using A Data Mining Approach,” *International Journal of Applied Sciences and Smart Technologies*, vol. 6, no. 1, pp. 169-182, 2024.

[35] M. F. Pramoedya, R. A. Ramadhani, and B. Setiawan, “Pemanfaatan K-Means Clustering dalam Pengelompokan Cuaca di Wilayah Kediri,” *Inotek*, vol. 9, pp. 2155–2164, 2025.

[36] D. Tri Cahaya, D. Puspita, and R. Syahri, “Penerapan Metode K-Means Clustering untuk Pengelompokan Potensi Padi di Kota Pagar Alam,” *JATI J. Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 2187–2193, 2024.

[37] F. D. Rahman, M. I. Zulfa, and A. Taryana, “Clustering dan Klasifikasi Data Cuaca Kota Cilacap dengan Menggunakan Metode K-Means dan Random Forest,” *J. SINTA Sist. Inf. dan Teknol. Komputasi*, vol. 1, no. 2, pp. 90–97, 2024.

[38] W. Härdle and L. Simar, *Applied Multivariate Statistical Analysis*. Hiedelberg: Springer-Verlag, 2015.

[39] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. New Jersey: Prentice-Hall, Inc., 2007.

[40] M. Sholeh and K. Aeni, “Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means,” *STRING Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 8, no. 1, pp. 56–59, 2023.

[41] G. Andrian, D. Arisandi, and T. Handhayani, “Clustering Data Meteorologi Wilayah Indonesia Timur dengan Metode K-Means dan Fuzzy C-Means,” *Inti Nusa Mandiri*, vol. 18, no. 2, pp. 100–106, 2024.