Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9627

Clustering Algorithm Comparison and Seasonal Validation for Daily Weather Patterns in Bima Regency

Wulandari[1*], Rumini[2]
[1,2]Informatics, Faculty of Computer Science, Universitas Amikom Yogyakarta, Indonesia
[1*]wulandari@students.amikom.ac.id, [2]rumini@amikom.ac.id

open access

***Corresponding Author**

Article History:
Submitted: 05-09-2026
Accepted: 12-09-2026
Published: 01-11-2026

Keywords:
clustering; daily weather patterns; external validation; K-Means; NASA POWER.

Brilliance: Research of Artificial Intelligence is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

ABSTRACT

Bima Regency in West Nusa Tenggara lies in a monsoonal region with sharp seasonal contrast, yet its daily weather typology has never been characterized quantitatively and its surface observation network is sparse. In comparable studies, algorithm choice usually rests on internal validity indices alone, so the physical meaning of the clusters is rarely demonstrated. This study identifies the daily weather typology of Bima Regency and compares four clustering algorithms using a framework combining internal validation, external validation, and a stability test. The data comprise 4,383 daily observations of six meteorological variables from NASA POWER for 2014 to 2025. Preprocessing covered a quality audit, a log transformation of rainfall, and standardization. The number of clusters was set to three from a composite score of validity indices. Four algorithms were tested: K-Means, Gaussian Mixture Model, Self-Organizing Map, and Hierarchical Ward. K-Means achieved the best internal performance with a Silhouette of 0.2748, a Calinski-Harabasz of 1825.39, and a Davies-Bouldin of 1.1993. Hierarchical Ward reached the highest seasonal agreement with a Cramer's V of 0.6324, only marginally above the 0.6321 of K-Means. Based on the weighted composite score, K-Means was selected as the main algorithm and proved highly stable, with a bootstrap Adjusted Rand Index of 0.9800. The typologies are wet monsoon at 45.40 percent of days, windy dry season at 31.46 percent, and hot-dry high-radiation conditions at 23.13 percent. The clusters therefore carry meteorological meaning and can support agricultural planning and drought preparedness.

INTRODUCTION

Indonesia lies where the Asian and Australian monsoon systems meet. This produces a firm seasonal cycle and large interannual variability. The monsoonal character is not uniform across the country. The Nusa Tenggara Islands, including Bima Regency, show a far sharper seasonal contrast than western Indonesia because their dry season is longer and their rainfall is lower. Interannual variability in Indonesian monsoon rainfall is driven by several external forcings acting on different time scales (Mulsandi et al., 2024), The 2023 El Nino event further showed that its effect on the Indonesian Maritime Continent varies widely in space (Zehri et al., 2025).

Characterizing daily weather at the regency level remains difficult. Surface station density in eastern Indonesia is still limited. Available information is usually a monthly average, which hides day-to-day variety. Yet operational decisions such as setting a planting window or anticipating drought depend on daily conditions. Agricultural drought in Nusa Tenggara can be diagnosed by combining surface information with satellite precipitation (Krisnawan et al., 2025), Multivariate daily data are therefore available, but they have not been turned into a typology that practitioners can use.

Previous work has followed two largely separate paths. The first path tests the reliability of reanalysis data. NASA POWER was judged adequate for weather monitoring in West Sumbawa (Darman et al., 2024) and as an alternative rainfall data source in Bali (I Gusti Ngurah Putu Dharmayasa et al., 2022), It also performed acceptably in a Mediterranean climate (Rodrigues & Braga, 2021) and over a long analysis span in Oman (Marzouk, 2021). The second path applies clustering to Indonesian data. (Andrian et al., 2024) compared K-Means and Fuzzy C-Means on eastern Indonesian meteorological data; (Risaldi et al., 2025) combined DBSCAN with a Gaussian Mixture Model for daily weather patterns in Makassar; (Uston Nawawi Christanto et al., 2025) used a cluster-based approach to predict maximum temperature; while (Mujahidin & Hasanah, 2025) and (Andayani et al., 2025) compared several algorithms on province-level aggregate data. On the methodological side, (Slaoui & Dafir, 2021) developed a K-Means variant that requires no initial parameter setting, while (Stryhal & Plavcová, 2023) and (Yoon et al., 2025) showed that the Self-Organizing Map is well established for atmospheric circulation patterns but is also prone to misinterpretation when its results are not verified.

Four gaps emerge from this synthesis. First, algorithm comparisons are usually limited to two or three methods from overlapping families, so differences in underlying assumptions are never truly tested. Second, cluster quality is almost always judged by internal indices alone, even though (Chicco et al., 2025) and (Ikotun et al., 2025) show that

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9627

these indices carry structural bias and can rank the same data differently. Third, the stability of results across data subsets is rarely tested. Fourth, most studies work on province-level or monthly aggregates rather than multivariate daily series, and the semi-arid region of West Nusa Tenggara still has no data-driven daily weather typology despite its high drought vulnerability.

This study compares four algorithms chosen deliberately for their differing assumptions. K-Means assumes spherical clusters with hard membership. The Gaussian Mixture Model assumes Gaussian components with probabilistic membership. The Self-Organizing Map preserves topological relations. Hierarchical Ward builds a nested structure. All four were applied to identical data: 4,383 daily observations of six meteorological variables for Bima Regency from 2014 to 2025. The novelty lies in the evaluation framework. Performance is judged not only by three internal indices but also by cluster agreement with the calendar season as an independent external reference, by between-algorithm agreement, and by a bootstrap stability test.

The study has three objectives: to identify the daily weather typology of Bima Regency, to compare the performance of four algorithms using a combined evaluation framework, and to characterize the resulting clusters within a monsoonal meteorological frame. Its contributions are an algorithm selection framework that does not rely on internal indices alone, a validated daily weather typology for Bima Regency, and empirical evidence that algorithm rankings by internal metrics do not always match meteorological meaningfulness.

LITERATURE REVIEW

Sparse station networks encourage the use of reanalysis products as an alternative data source. NASA POWER has been tested in various regions with results that vary by variable and by local climate. It was judged suitable as an alternative for weather monitoring in West Sumbawa (Darman et al., 2024) and adequate for hydrological studies in Bali (I Gusti Ngurah Putu Dharmayasa et al., 2022), Testing in a Mediterranean climate showed agreement that differs between variables (Rodrigues & Braga, 2021) and a 39-year analysis in Al Buraimi showed reliability against meteorological measurements (Marzouk, 2021). The consensus is that NASA POWER suits the analysis of patterns and trends better than the measurement of absolute values at a single point. This study uses it in that sense.

Interpreting weather clusters in Indonesia requires a monsoonal frame. Interannual variability in Indonesian monsoon rainfall is shaped by several external forcings acting together (Mulsandi et al., 2024), The response of Maritime Continent weather to El Nino is not uniform across regions (Zehri et al., 2025). At the regional scale of Nusa Tenggara, agricultural drought can be diagnosed by combining surface information with satellite precipitation (Krisnawan et al., 2025). These findings support the use of calendar seasons as an external validation reference here. They also explain why that reference is only approximate, since actual season boundaries shift from year to year.

The four algorithms tested rest on different structural assumptions. K-Means minimizes the sum of squared distances to cluster centers, so it tends to produce spherical clusters of relatively balanced size, and a variant without initial parameter setting has been developed to reduce dependence on initialization (Slaoui & Dafir, 2021). The Gaussian Mixture Model treats data as a mixture of Gaussian distributions and yields membership probabilities, which suits gradually changing phenomena such as weather; hybrid approaches involving it have been applied to daily weather patterns in Makassar (Risaldi et al., 2025) and to grouping provinces by welfare indicators (Andayani et al., 2025). The Self-Organizing Map maps data onto a grid of nodes while preserving topological relations and has long been used for atmospheric circulation patterns (Stryhal & Plavcová, 2023), Its interpretation nonetheless demands care, because topological mapping can produce groupings that appear meaningful but are no (Yoon et al., 2025). Hierarchical Ward minimizes the increase in variance at each merge, so it needs no initial cluster centers.

Cross-algorithm comparison in the Indonesian context has been carried out on eastern Indonesian meteorological data (Andrian et al., 2024) and on province-level rice production data (Mujahidin & Hasanah, 2025). Both report K-Means as competitive, but both compare algorithms from the centroid-based family, so differences in structural assumption are never truly tested. (Uston Nawawi Christanto et al., 2025) used clustering as an intermediate step for predicting daily maximum temperature, so cluster structure was not the object of analysis. Cross-family studies such as (Andayani et al., 2025) which set DBSCAN against the Gaussian Mixture Model, were carried out on cross-sectional data rather than on daily weather series.

On the evaluation side, internal validity indices measure cluster cohesion and separation without a reference label. (Chicco et al., 2025) show that the Silhouette coefficient and the Davies-Bouldin index give more reliable information than the Dunn index, Calinski-Harabasz index, Shannon entropy, and the Gap statistic for two convex clusters. (Ikotun et al., 2025) report that different validity indices can rank the same data differently, so no single index is universally superior. Judging clusters by one index alone therefore risks being misleading. Most of the meteorological clustering studies reviewed above stop at internal indices. They do not test whether the cluster structure holds across data subsets, nor whether the clusters agree with an independent meteorological reference. That gap is the entry point for this study, which uses three internal indices together, adds the calendar season as an external reference, and tests stability by bootstrap.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9627

METHOD

This is a quantitative study using an unsupervised learning approach. The research flow consists of seven sequential stages, as shown in Figure 1. The first stage is daily data collection from NASA POWER. It is followed by data cleaning and a quality audit, preprocessing, determining the number of clusters, applying four clustering algorithms, internal and external evaluation together with a stability test, and finally cluster characterization and interpretation.

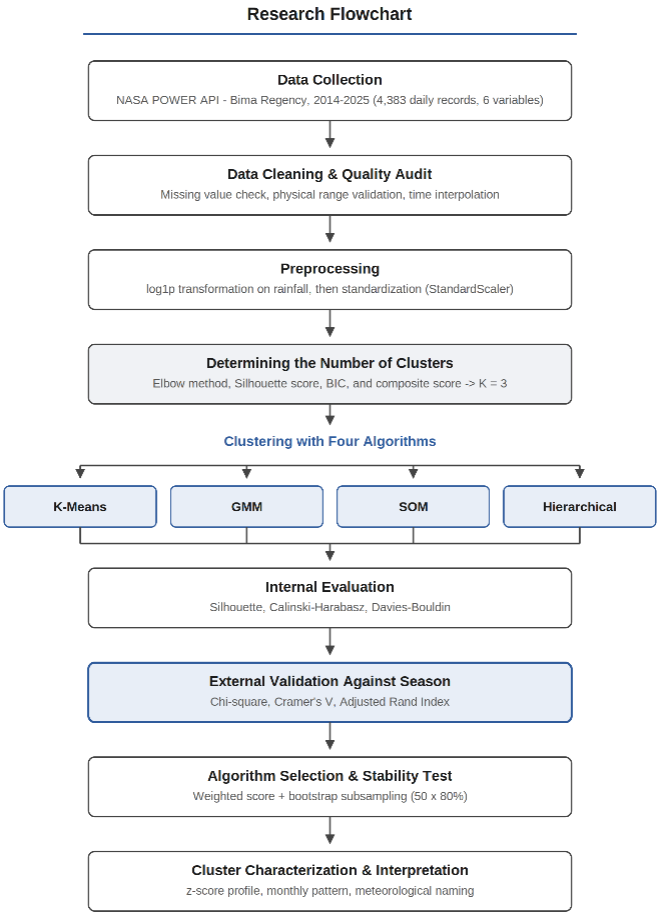


Figure 1. Research flow diagram

Data Collection

Data were obtained from the NASA POWER service through the temporal/daily/point endpoint of the AG community for the grid point at latitude -8.47 and longitude 118.75, representing Bima Regency, West Nusa Tenggara. The collection period runs from 1 January 2014 to 31 December 2025, a full twelve calendar years, yielding 4,383 daily observations. Six parameters were retrieved: air temperature at 2 meres (T2M, degrees Celsius), relative humidity at 2 meters (RH2M, percent), corrected precipitation (PRECTOTCORR, millimeters), wind speed at 2 meters (WS2M, meters per second), surface pressure (PS, kilopascals), and all-sky surface shortwave downward irradiance (ALLSKY_SFC_SW_DWN, megajoules per square meter). Retrieval included a retry mechanism of up to three attempts and local caching so that the results can be reproduced identically.

Data Cleaning and Quality Audit

The quality audit ran in five steps. The NASA POWER missing-value marker of -999 was replaced with a null value. Duplicate dates were checked and removed where found. Date-series completeness was checked against the full date range, and missing dates were inserted as empty rows. Each variable was validated against a plausible physical range: temperature from -10 to 55 degrees Celsius, humidity from 0 to 100 percent, rainfall and wind speed non-negative, pressure from 50 to 110 kilopascals, and radiation non-negative. Values outside these ranges were treated as missing. Remaining nulls were filled by time-based linear interpolation. Meteorological extremes such as very wet days were not removed, since they are valid physical events; the effect of their skewed distribution is handled by the transformation in the next stage.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, Number 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9627

Preprocessing

Redundancy between variables was checked with a Pearson correlation matrix using a monitoring threshold of 0.85 and a removal threshold of 0.95. Variables with absolute skewness above 1.0 and a non-negative minimum received a log1p transformation, $x' = \ln(1 + x)$. All variables were then standardized with StandardScaler to zero mean and unit standard deviation, since Euclidean distance is sensitive to differences in unit. To test whether the transformation actually helps, two preprocessing scenarios were compared explicitly. Scenario A applied standardization without transformation. Scenario B applied the log1p transformation followed by standardization. Both were evaluated with K-Means over K from 2 to 8 and selected by the highest Silhouette value, with a tolerance of -0.01 favoring the transformed scenario because it better satisfies the Euclidean distance assumption.

Determining the Number of Clusters

Four criteria were used. The Elbow method was applied quantitatively by finding the point on the inertia curve with the maximum perpendicular distance to the line joining the first and last points after both axes were normalized, so the elbow does not depend on visual judgement. The second criterion is the maximum Silhouette value and the third is the minimum Bayesian Information Criterion of the Gaussian Mixture Model. The fourth is a composite score combining the three internal indices after min-max normalization: one third of the sum of the normalized Silhouette, the normalized Calinski-Harabasz, and one minus the normalized Davies-Bouldin. The Davies-Bouldin index is inverted because lower values are better.

Applying the Four Clustering Algorithms

All four algorithms were applied to the same standardized data matrix with the same number of clusters. K-Means was run with n_init of 10 and random_state of 42 as the baseline, following the standard formulation (Ikotun et al., 2025). The Gaussian Mixture Model used a full covariance matrix estimated by the expectation-maximization algorithm with n_init of 10. It was chosen because it yields membership probabilities, so transitional days can be identified explicitly. The Self-Organizing Map was applied in two levels, with grid size following the rule of roughly five times the square root of the number of observations, which for 4,383 observations gives a 19 by 19 grid of 361 nodes. Training ran for 12,000 iterations with an initial learning rate of 0.5 decaying exponentially and a Gaussian neighborhood function. The node weight vectors were then grouped by K-Means into three clusters, because the number of nodes far exceeds the number of clusters sought. Hierarchical Clustering was run agglomeratively with Ward linkage.

Evaluation

Internal validation uses three indices. The Silhouette coefficient for observation i is the difference between b(i) and a(i) divided by the larger of the two, where a(i) is the mean distance from observation i to members of its own cluster and b(i) is the smallest mean distance to members of another cluster. It ranges from -1 to 1 and higher is better. The Calinski-Harabasz index is the ratio of between-cluster to within-cluster dispersion, where higher is better. The Davies-Bouldin index is the mean maximum similarity ratio between cluster pairs, where lower is better. The three indices are used together because (Chicco et al., 2025) and (Ikotun et al., 2025) show that no single index is universally superior.

External validation compares cluster labels against calendar seasons, which never entered the training process. The seasons are defined as the wet season from December to March, the first transition from April to May, the dry season from June to September, and the second transition from October to November. Association was measured by a Chi-square test on the contingency table, by Cramer's V, which is the square root of the Chi-square to sample size ratio divided by the smaller of the number of rows minus one and the number of columns minus one, and by the Adjusted Rand Index and Normalized Mutual Information. Between-algorithm agreement was computed with the same two indices pairwise.

The main algorithm was selected by a weighted composite score giving 0.4 to the mean of the three normalized internal indices and 0.6 to the normalized Cramer's V. The larger weight goes to external validation because the aim is clusters that are meteorologically meaningful rather than merely optimal in geometric terms. The stability test used bootstrap subsampling of 50 iterations at 80 percent without replacement, comparing each iteration's labels against the full-model labels with the Adjusted Rand Index. Cluster characterization used the mean z-score profile of each variable per cluster against the overall mean, with an interpretation threshold of plus or minus 0.5, together with monthly and annual distributions, a between-cluster difference test, and a Dunn post-hoc test with Bonferroni correction. The difference test was chosen automatically: ANOVA when normality and homogeneity of variance held at the 0.05 level, and Kruskal-Wallis otherwise, with effect size reported as eta-squared or epsilon-squared.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, Number 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9627

Experimental Environment

All analyze ran on Python 3.13.15 under Linux with scikit-learn 1.6.1, SciPy 1.16.3, NumPy 2.1.3, and pandas 2.2.3. The Self-Organizing Map was implemented directly in NumPy without an additional library. The random seed was fixed at 42 for every stochastic step so that the results can be reproduced.

RESULT

Data Characteristics and Quality Audit

The audit shows the data are intact. No -999 markers, duplicate dates, missing dates, or values outside the physical range were found in any of the six variables. Completeness therefore reached 100 percent and interpolation was not applied to a single value.

Descriptive statistics for the six variables appear in Table 1. Mean air temperature is 27.235 degrees Celsius with a coefficient of variation of only 4.196 percent, reflecting the temperature stability typical of the tropics. Rainfall shows the most extreme spread, with a mean of 3.444 millimeters, a coefficient of variation of 171.243 percent, skewness of 3.414, and kurtosis of 16.688. It is the only variable with absolute skewness above 1.0. The correlation check found no variable pair reaching an absolute correlation of 0.85, so no variable was dropped and all six were retained as clustering features.

Table 1. Descriptive statistics of six meteorological variables (n = 4,383)

Variable	Mean	Std. deviation	Min	Median	Max	Skewness	CV (%)
Temperature (C)	27.235	1.143	24.42	27.05	31.31	0.651	4.196
Humidity (%)	74.537	8.634	51.27	76.40	91.33	-0.467	11.584
Rainfall (mm)	3.444	5.897	0.00	0.98	62.94	3.414	171.243
Wind (m/s)	2.806	1.091	0.47	2.76	6.95	0.259	38.880
Pressure (kPa)	98.689	0.172	98.11	98.70	99.23	-0.183	0.175
Radiation (MJ/m2)	19.991	3.503	3.22	20.35	27.77	-0.837	17.525

Preprocessing and Determining the Number of Clusters

The log1p transformation was applied to rainfall and reduced skewness from 3.414 to 0.802, which on the standardized scale lowered its maximum from 10.09 to 3.39. The transformed scenario produced a higher Silhouette and a lower Davies-Bouldin for K from 2 to 4, with the best Silhouette of 0.2748 at K equal to 3 against 0.2695 at K equal to 4 for the untransformed scenario. The transformed scenario was therefore chosen for all subsequent analysis.

Results for the number of clusters appear in Table 2 and Figure 2. The four criteria do not fully converge. The quantitative Elbow method points to K equal to 4, the maximum Silhouette points to K equal to 3 with a value of 0.2748, and the minimum Bayesian Information Criterion points to K equal to 8 with a value of 54464.1470. The composite score peaks at 0.9754 for K equal to 3, followed by 0.9149 for K equal to 4 and 0.6963 for K equal to 2. Based on the composite score, the number of clusters was set to three.

Table 2. Cluster number evaluation under the selected preprocessing scenario

K	Inertia	Silhouette	Calinski-Harabasz	Davies-Bouldin	BIC GMM	Composite score
2	18423.1232	0.2712	1872.6489	1.4474	58261.0104	0.6963
3	14342.9969	0.2748	1825.3865	1.1993	56216.7166	0.9754
4	11720.6587	0.2727	1815.4336	1.2401	55619.1229	0.9149
5	10787.1492	0.2234	1573.8011	1.3351	54826.5937	0.4953
6	9981.5174	0.2009	1430.9907	1.4841	54749.9799	0.1632
7	9369.5440	0.1858	1317.7320	1.4835	54521.7447	0.0491

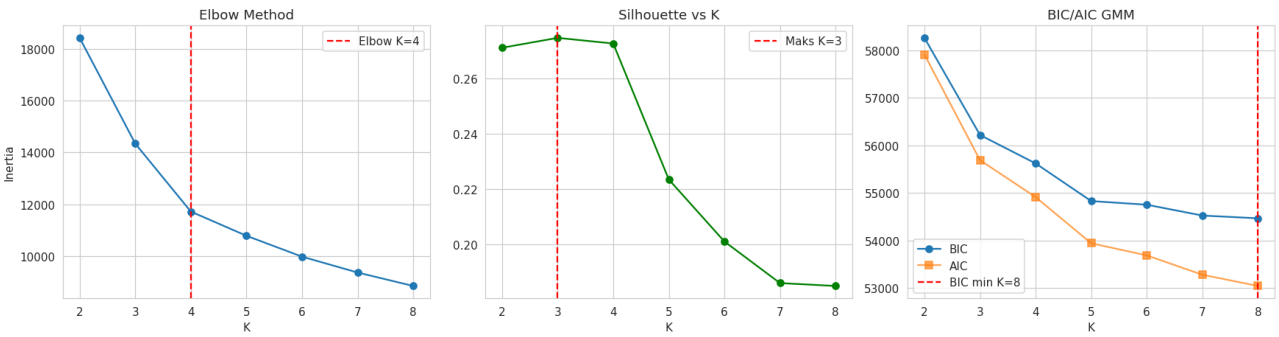


Figure 2. Elbow, Silhouette, and BIC/AIC curves for determining the number of clusters

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9627

Figure 2 presents these four criteria as three curves plotted against the number of clusters (K) from 2 to 8. The left panel shows the Elbow curve, where inertia falls steeply from K = 2 to K = 4 and then flattens, so the marked bend at K = 4 is the point beyond which adding clusters gives diminishing reduction in within-cluster variance. The middle panel shows the Silhouette curve, which rises slightly and reaches a visible peak at K = 3 before dropping sharply from K = 4 onward, indicating that cohesion and separation are best balanced at three clusters and worsen quickly beyond that point. The right panel shows the BIC and AIC curves of the Gaussian Mixture Model, both of which decline across the whole range without a clear inflection and reach their lowest value at K = 8, a pattern typical of likelihood-based criteria that keep favoring more components unless a strong complexity penalty is applied. Read together, the three panels show visually why the four criteria disagree: two of them (Elbow, BIC/AIC) keep improving toward higher K while only the Silhouette curve shows a distinct peak, which is why the composite score in Table 2 is used to reconcile them and supports setting K = 3 as the number of clusters carried into the algorithm comparison.

Performance Comparison of the Four Algorithms

Results for the four algorithms at K equal to 3 appear in Table 3 and Figure 3. K-Means obtained the best value on all three internal indices: a Silhouette of 0.2748, a Calinski-Harabasz of 1825.3865, and a Davies-Bouldin of 1.1993. SOM ranked second with 0.2576, 1711.4018, and 1.2274, followed by Hierarchical Ward with 0.2231, 1485.6875, and 1.4093. The Gaussian Mixture Model obtained the lowest value on every internal index: 0.1708, 1020.6849, and 2.4080. The smallest cluster share ranges from 19.60 to 24.57 percent, so no algorithm produced severely unbalanced clusters.

Table 3. Performance comparison of four clustering algorithms at K = 3

Algorithm	Silhouette	Calinski-Harabasz	Davies-Bouldin	Smallest cluster (%)
K-Means	0.2748	1825.3865	1.1993	23.13
GMM	0.1708	1020.6849	2.4080	19.60
SOM	0.2576	1711.4018	1.2274	24.57
Hierarchical	0.2231	1485.6875	1.4093	24.09

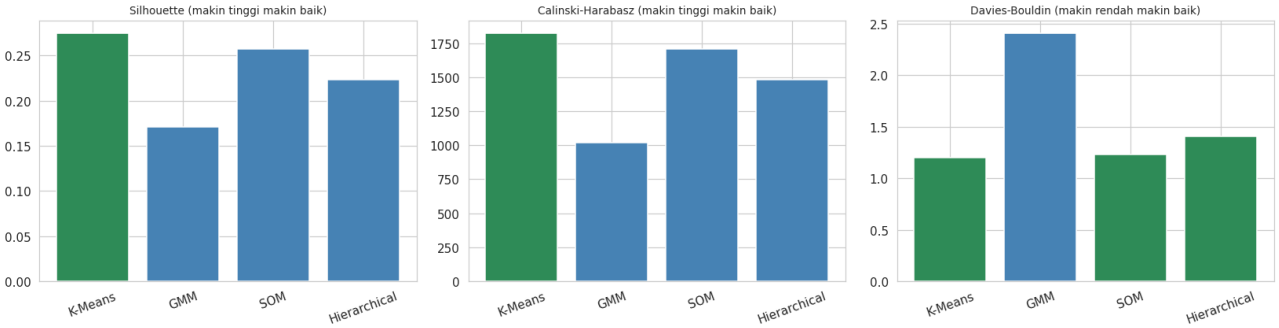


Figure 3. Comparison of three internal validity indices across algorithms

Between-algorithm agreement measured by the Adjusted Rand Index averages 0.5543 with a range from 0.4665 to 0.7629, showing that the four algorithms find a similar but not identical structure. Cluster membership also differs: 1,379, 1,990, and 1,014 days for K-Means; 1,286, 859, and 2,238 days for GMM; 1,875, 1,077, and 1,431 days for SOM; and 1,880, 1,447, and 1,056 days for Hierarchical. The SOM mapping yielded a quantization error of 0.7900 with all 361 nodes used. For GMM, the mean highest membership probability reached 0.9426, and 114 days or 2.60 percent had a highest probability below 0.60, most often in May at 7.5 percent.

External Validation, Algorithm Selection, and Stability Test

Validation against calendar seasons appears in Table 4. The Chi-square test returned p values below machine precision for all four algorithms with 6 degrees of freedom. Hierarchical Ward obtained the highest Cramer's V at 0.6324, followed by K-Means at 0.6321, SOM at 0.6270, and GMM at 0.5721, with only 0.0003 separating the top two. The Adjusted Rand Index against season ranges from 0.3302 to 0.3692 and Normalized Mutual Information from 0.3200 to 0.3800, both highest for Hierarchical.

Table 4. External validation of clusters against calendar seasons

Algorithm	Chi-square	df	Cramer's V	ARI vs season	NMI vs season
K-Means	3502.81	6	0.6321	0.3612	0.3563
GMM	2868.75	6	0.5721	0.3302	0.3200
SOM	3446.43	6	0.6270	0.3641	0.3418
Hierarchical	3506.07	6	0.6324	0.3692	0.3800

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9627

The weighted composite score puts K-Means highest at 0.9970, followed by SOM at 0.9022, Hierarchical at 0.8543, and GMM at 0.0000. K-Means reached the maximum normalized value on all three internal components and 0.9950 on the external component. The bootstrap stability test on K-Means with 50 iterations at 80 percent subsampling gave a mean Adjusted Rand Index of 0.9800 with a standard deviation of 0.0167 and a range from 0.9269 to 0.9972, and a mean subsample Silhouette of 0.2749 with a standard deviation of 0.0025, almost identical to the full-model value of 0.2748.

Cluster Characterization

Mean and z-score profiles of the three K-Means clusters appear in Table 5 and Figure 4. Cluster 0 is marked by relatively high wind speed and surface pressure with relatively low rainfall. Cluster 1, the largest, is marked by relatively high humidity and rainfall with relatively low pressure. Cluster 2 is marked by the highest temperature and radiation with the lowest humidity.

Table 5. Characteristics of three daily weather clusters in Bima Regency

Characteristic	Cluster 0	Cluster 1	Cluster 2
Name	Windy dry season (eastern monsoon)	Wet monsoon	Hot-dry high radiation (pre-rain)
Number of days	1,379 (31.46%)	1,990 (45.40%)	1,014 (23.13%)
Temperature (C)	26.538 (z = -0.61)	26.920 (z = -0.28)	28.803 (z = +1.37)
Humidity (%)	71.782 (z = -0.32)	81.344 (z = +0.79)	64.924 (z = -1.11)
Rainfall (mm)	0.290 (z = -0.53)	6.813 (z = +0.57)	1.120 (z = -0.39)
Wind (m/s)	3.667 (z = +0.79)	2.266 (z = -0.50)	2.695 (z = -0.10)
Pressure (kPa)	98.830 (z = +0.82)	98.587 (z = -0.59)	98.699 (z = +0.06)
Radiation (MJ/m2)	19.302 (z = -0.20)	18.661 (z = -0.38)	23.536 (z = +1.01)
Dominant months	Jul, Aug, Jun	Feb, Mar, Jan	Oct, Sep, Nov
Dominant season	Dry season	Wet season	Second transition

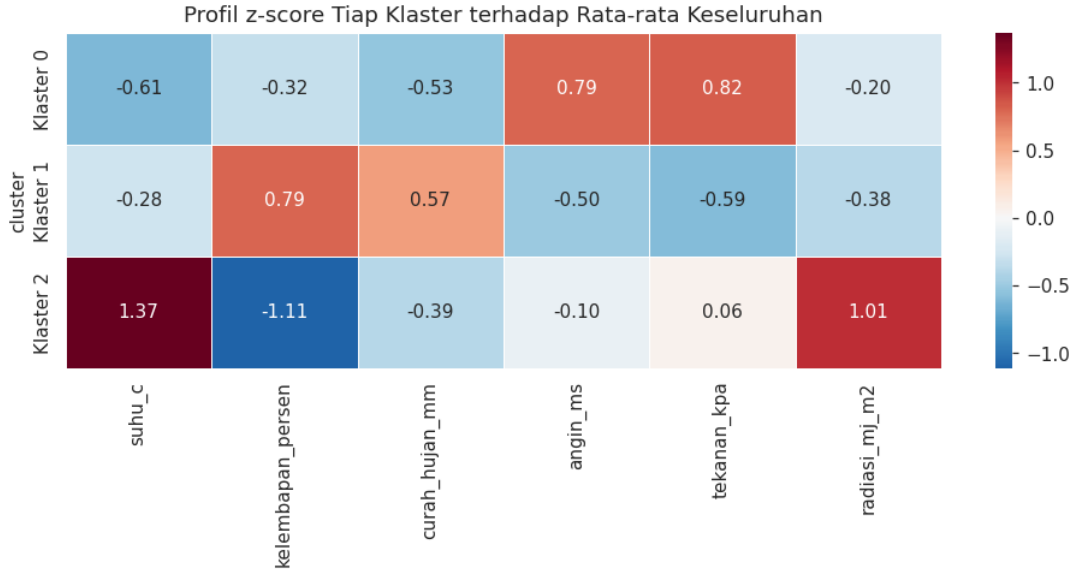


Figure 4. Z-score profile of the three clusters against the overall mean

Figure 4 renders the same z-scores from Table 5 as a diverging color-coded matrix, with the three clusters as rows and the six standardized meteorological variables as columns; warm (red) cells mark values above the twelve-year mean and cool (blue) cells mark values below it, so the dominant color of each row summarizes that cluster's meteorological signature at a glance. Cluster 0's row is warm for wind speed (z = +0.79) and pressure (z = +0.82) and cool for temperature, humidity, and rainfall, visually confirming a dry, windy profile. Cluster 1's row is warm only for humidity (z = +0.79) and rainfall (z = +0.57) and cool for wind and pressure, the near mirror image of Cluster 0, consistent with its identification as the wet monsoon type. Cluster 2's row stands out with the deepest red for temperature (z = +1.37) and radiation (z = +1.01) alongside the deepest blue for humidity (z = -1.11), the most extreme contrast in the matrix, marking it as the hot, dry, high-radiation type. This near mirror-symmetry between the Cluster 0 and Cluster 1 rows, together with the extreme values isolated in Cluster 2, shows that the three clusters are separated mainly along a wet-dry axis and a temperature-radiation axis, consistent with the two leading components of the principal component projection reported later in this section.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9627

The monthly distribution of clusters appears in Figure 5. Cluster 1 dominates January to March with more than 300 days per month across twelve years. Cluster 0 dominates June to August, peaking at 350 days in July. Cluster 2 concentrates in September to November, peaking at 309 days in October. Cluster 0 barely appears from January to March, while Cluster 2 is absent in June and July. Across years, the share of Cluster 1 ranges from 29.86 percent in 2019 to 65.75 percent in 2022, and Cluster 2 from 12.88 percent in 2021 to 29.51 percent in 2024.

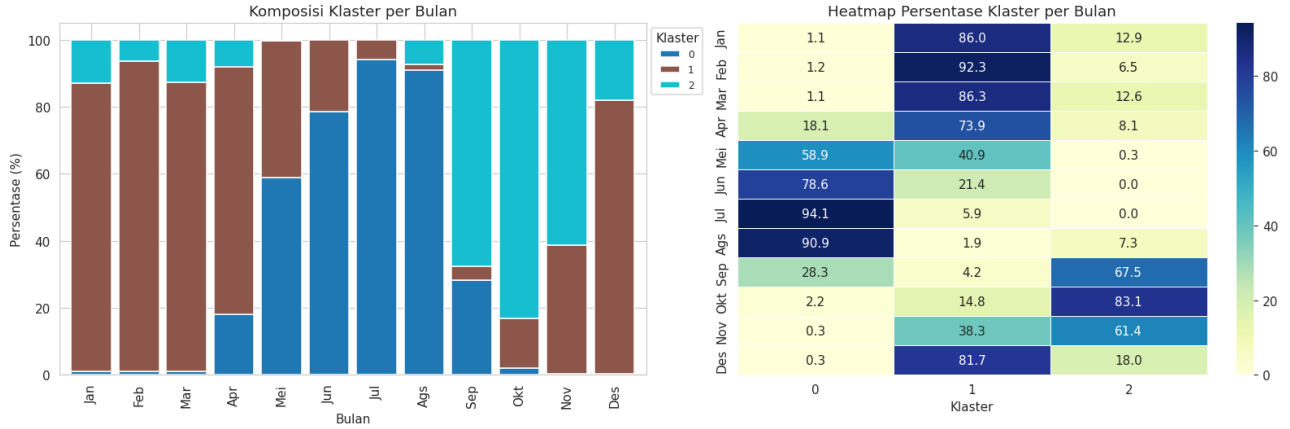


Figure 5. Monthly cluster composition

The between-cluster difference test used Kruskal-Wallis, because normality and homogeneity of variance did not hold for any variable. All variables show a large effect size. The highest epsilon-squared is humidity at 0.6109, followed by rainfall at 0.5843 and temperature at 0.5077. The Dunn post-hoc test with Bonferroni correction shows that Cluster 0 against Cluster 2 and Cluster 1 against Cluster 2 differ on all six variables, while Cluster 0 against Cluster 1 differs on five of six. This test is descriptive and measures the degree of separation. It is not a confirmatory hypothesis test, since it is applied to the same variables that formed the clusters.

The principal component projection, used for visualization only, shows that the first component explains 43.14 percent of the variance and the second 29.12 percent, for a total of 72.27 percent. The first component is dominated by humidity at 0.567 and rainfall at 0.481 on the positive side and by radiation at -0.434 and pressure at -0.347 on the negative side, which is physically a wet-dry axis. The second component is dominated by wind at 0.570 and pressure at 0.416 on the positive side and by temperature at -0.533 on the negative side, which is a windy eastern monsoon axis set against calm hot conditions.

DISCUSSION

K-Means obtained the best value on all three internal indices. The explanation that best fits the data is that the preprocessed data structure closely matches its assumptions. After rainfall was transformed and all variables standardized, the three clusters are not severely unbalanced, ranging from 23.13 to 45.40 percent, and no single variable dominates the distance space because no correlation redundancy was found. These conditions favor an algorithm that assumes compact clusters of balanced size, consistent with (Andrian et al., 2024) and (Mujahidin & Hasanah, 2025).

A Silhouette of 0.2748 is moderate and needs careful reading. It shows that daily weather clusters are not sharply separated, which is physically reasonable because weather changes gradually and produces a continuous gradation between weather types. A moderate Silhouette is therefore not a sign of algorithm failure but a reflection of the phenomenon being clustered, consistent with the reminder from (Chicco et al., 2025) that internal validity indices measure cluster geometry and do not directly measure meaningfulness.

The weaker performance of the Gaussian Mixture Model can be explained by a difference in objective function. The internal indices used here are all based on Euclidean distance and reward compact, spherical clusters. GMM with a full covariance matrix optimizes likelihood fit and can produce elongated or overlapping components, which automatically lowers the Silhouette and raises the Davies-Bouldin. The same pattern appears in the Bayesian Information Criterion, which keeps falling up to K equal to 8, indicating that GMM tends to split the data into more than three components.

Even so, GMM provides information no other algorithm can. A mean highest membership probability of 0.9426 shows that most days can be assigned firmly, while 114 days or 2.60 percent have ambiguous membership, most often in May at 7.5 percent, coinciding with the transition from wet to dry season. This finding gives quantitative support to the idea that transitional periods are genuinely hard to classify, and it adds nuance to the application of GMM to daily weather patterns reported by daily weather patterns (Risaldi et al., 2025). In practice, GMM is a poor choice as the main algorithm when the evaluation criteria are geometric, but it is useful as a complementary tool for identifying transitional days.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

E-ISSN : 2807-9035
Volume 6, Number 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9627

The first three criteria give different answers on the number of clusters: Elbow at K equal to 4, Silhouette at K equal to 3, and the Bayesian Information Criterion at K equal to 8. This is not an anomaly but a consequence of what each criterion measures. Elbow measures the rate of inertia decline, Silhouette measures the balance of cohesion and separation, and the Bayesian Information Criterion measures probabilistic model fit with a complexity penalty. (Ikotun et al., 2025) and the Bayesian Information Criterion measures probabilistic model fit with a complexity penalty. (Ikotun et al., 2025) report that different validity indices can indeed rank the same data differently. Using a composite score is therefore a reasonable middle path, and the choice of three clusters is reinforced by two further pieces of evidence from outside the internal-index frame: strong agreement with calendar seasons and high bootstrap stability. Even so, setting three clusters still involves researcher judgement, which is stated here openly.

The most important methodological finding is that the algorithm ranking by internal indices does not match the ranking by seasonal agreement. K-Means leads on every internal index, yet the highest Cramer's V goes to Hierarchical Ward at 0.6324 against 0.6321. That margin is too small to claim a meaningful advantage, and this study makes no claim that either is superior on the external dimension. What matters is the mismatch in order on other dimensions: Hierarchical ranks third on internal indices yet leads on the Adjusted Rand Index and Normalized Mutual Information against season, at 0.3692 and 0.3800. This strengthens the argument that clustering evaluation resting on internal indices alone can mislead, as warned by (Chicco et al., 2025) and (Ikotun et al., 2025), and it underlines the interpretive caution raised by (Yoon et al., 2025) in the SOM context.

A Cramer's V above 0.57 for all four algorithms is substantively meaningful, because time information never entered as a feature. The algorithms received only six daily meteorological values with no knowledge of date or month. A strong association with season therefore shows that the clusters reflect physical atmospheric structure rather than a purely mathematical partition. The Adjusted Rand Index against season, at only 0.33 to 0.37, is far lower than Cramer's V. This gap follows from the unequal number of categories, three clusters against four seasons, so perfect agreement is impossible by definition.

The three clusters can be read consistently within a monsoonal frame. Cluster 0, with high surface pressure and wind speed and very low rainfall, peaks in July and August and matches the eastern monsoon that brings dry air masses from Australia. Cluster 1, with high humidity and rainfall and low pressure, peaks from January to March and matches the western monsoon. Cluster 2, with the highest temperature and radiation and the lowest humidity, peaks from September to November and matches the pre-rain period when maximum radiation is not yet offset by increased cloud cover. This structure is consistent with the picture of Indonesian monsoon variability (Mulsandi et al., 2024) and with the dry character of the Nusa Tenggara Islands (Krisnawan et al., 2025).

Interannual variation in cluster share, for example 65.75 percent in 2022 against 29.86 percent in 2019 for Cluster 1, indicates that the typology is sensitive to interannual variability. This study did not cross-analyze against ENSO or Indian Ocean Dipole indices so that link can only be raised as a possibility requiring further testing.

A bootstrap Adjusted Rand Index of 0.9800 with a standard deviation of 0.0167 shows that the K-Means cluster structure is highly robust to variation in data subsets, reinforced by a subsample Silhouette almost identical to the full model. A mean between-algorithm Adjusted Rand Index of 0.5543 shows that the four algorithms find a similar underlying structure even though they place cluster boundaries differently. The three-cluster structure is therefore present in the data and is not an artefact of one particular algorithm. Stability tests of this kind are rarely reported in the Indonesian meteorological clustering studies reviewed here.

In practical terms, the three-cluster typology offers a simple frame for turning daily weather data into operational information. The dominance of Cluster 1 from January to March and Cluster 0 from June to August can guide planting windows and water supply planning, while Cluster 2, concentrated from September to November, marks a period of high evaporative demand and is therefore relevant to drought preparedness. Methodologically, adding an external reference and a stability test costs little computation and should become standard practice in meteorological clustering studies.

This study has several limitations. Each day is treated as an independent observation even though weather on consecutive days is strongly correlated, so the effective sample is smaller than 4,383 and statistical tests tend to reach significance too readily. The between-cluster difference test is circular because it is applied to the same variables that formed the clusters, so it serves a descriptive role only. NASA POWER represents the mean condition of a grid cell rather than a point measurement, so small-scale phenomena and topographic effects are not fully captured. A twelve-year span is adequate for recognizing seasonal patterns but falls below the thirty-year standard for climate normals. External validation uses season boundaries based on calendar months, whereas actual season onset varies from year to year. The criteria for the number of clusters do not fully converge, and the results have not been cross-verified against data from the nearest observation station.

CONCLUSION

This study identified the daily weather typology of Bima Regency using 4,383 daily observations of six meteorological variables from NASA POWER for 2014 to 2025, and compared four clustering algorithms under a combined evaluation framework. Three typologies were identified: wet monsoon with 1,990 days or 45.40 percent,

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9627

peaking from January to March; windy dry season with 1,379 days or 31.46 percent, peaking from June to August; and hot-dry high-radiation conditions with 1,014 days or 23.13 percent, peaking from September to November.

K-Means was selected as the main algorithm with a weighted composite score of 0.9970. It leads on all three internal indices with a Silhouette of 0.2748, a Calinski-Harabasz of 1825.39, and a Davies-Bouldin of 1.1993, and it proved highly stable with a bootstrap Adjusted Rand Index of 0.9800. An important accompanying finding is that the ranking by internal indices does not match the ranking by seasonal agreement. Hierarchical Ward, third on internal indices, leads on the Adjusted Rand Index and Normalized Mutual Information against season, although its Cramer's V exceeds that of K-Means by only 0.0003, which is too small to be called a meaningful advantage. All four algorithms produced a Cramer's V above 0.57 without ever receiving time information, showing that the clusters reflect real meteorological structure. The Gaussian Mixture Model, lowest on internal indices, still contributed complementary information by identifying 114 days of ambiguous membership concentrated in May.

The contributions of this study are an algorithm selection framework combining internal validation, external validation, and a stability test; a validated daily weather typology for Bima Regency; and empirical evidence on the limits of judging clusters by internal indices alone. The main limitations are the assumption of independence between days, the grid resolution of reanalysis data, the twelve-year span, and the absence of verification against station data. Further work should cross-verify against data from the nearest observation station, use season boundaries that vary by year, analyze links with ENSO and Indian Ocean Dipole indices, apply time-series clustering methods that account for temporal autocorrelation, and extend the study area to several regencies in Nusa Tenggara to test whether this typology is local or regional.

REFERENCES

Andayani, S., Retnani, N., Yusri, T. A. S., & Marwoto, B. S. H. (2025). A COMPARATIVE ANALYSIS OF DBSCAN AND GAUSSIAN MIXTURE MODEL FOR CLUSTERING INDONESIAN PROVINCES BASED ON SOCIOECONOMIC WELFARE INDICATORS. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *19*(3), 2039–2056. https://doi.org/10.30598/barekengvol19iss3pp2039-2056

Andrian, G., Arisandi, D., & Handhayani, T. (2024). CLUSTERING DATA METEOROLOGI WILAYAH INDONESIA TIMUR DENGAN METODE K-MEANS DAN FUZZY C-MEANS. *INTI Nusa Mandiri*, *18*(2), 100–106. https://doi.org/10.33480/inti.v18i2.5039

Chicco, D., Campagner, A., Spagnolo, A., Ciucci, D., & Jurman, G. (2025). The Silhouette coefficient and the Davies-Bouldin index are more informative than Dunn index, Calinski-Harabasz index, Shannon entropy, and Gap statistic for unsupervised clustering internal evaluation of two convex clusters. *PeerJ Computer Science*, *11*, e3309. https://doi.org/10.7717/peerj-cs.3309

Darman, L. P., Januhariadi, Yudha, M. P., & Aslan. (2024). Assessment of NASA POWER reanalysis products as data resources alternative for weather monitoring in West Sumbawa, Indonesia. *E3S Web of Conferences*, *485*. https://doi.org/10.1051/e3sconf/202448506006

I Gusti Ngurah Putu Dharmayasa, Cathleen Ariella Simatupang, & Doni Marisi Sinaga. (2022). NASA Power's: an alternative rainfall data resources for hydrology research and planning activities in Bali Island, Indonesia. *Journal of Infrastructure Planning and Engineering (JIPE)*, *1*(1), 1–7. https://doi.org/10.22225/jipe.1.1.2022.1-7

Ikotun, A. M., Habyarimana, F., & Ezugwu, A. E. (2025). Benchmarking validity indices for evolutionary K-means clustering performance. *Scientific Reports*, *15*(1), 21842. https://doi.org/10.1038/s41598-025-08473-6

Krisnawan, G. D., Chang, Y.-L., Tsai, F., Tseng, K.-H., & Lin, T.-H. (2025). Meteorological Drivers and Agricultural Drought Diagnosis Based on Surface Information and Precipitation from Satellite Observations in Nusa Tenggara Islands, Indonesia. *Remote Sensing*, *17*(14), 2460. https://doi.org/10.3390/rs17142460

Marzouk, O. A. (2021). Assessment of global warming in Al Buraimi, sultanate of Oman based on statistical analysis of NASA POWER data over 39 years, and testing the reliability of NASA POWER against meteorological measurements. *Heliyon*, *7*(3), e06625. https://doi.org/10.1016/j.heliyon.2021.e06625

Mujahidin, I., & Hasanah, S. H. (2025). COMPARATIVE ANALYSIS OF K-MEANS, K-MEDOIDS, AND FUZZY C-MEANS FOR CLUSTERING PROVINCES IN INDONESIA BASED ON RICE PRODUCTION IN 2024. *Jurnal Gaussian*, *14*(2), 356–365. https://doi.org/10.14710/j.gauss.14.2.356-365

Mulsandi, A., Koesmaryono, Y., Hidayat, R., Faqih, A., & Sopaheluwakan, A. (2024). On The Interannual Variability of Indonesian Monsoon Rainfall (IMR): A Literature Review of The Role of its External Forcing. *Jurnal Meteorologi Dan Geofisika*, *24*(2), 115–127. https://doi.org/10.31172/jmg.v24i2.1049

Risaldi, M., Safitri, A., Nur Risal, A. A., Surianto, D. F., Andayani, D. D., Edy, M. R., Firdaus, F., & Parenreng, J. M. (2025). Clustering and Modeling of Daily Weather Pattern Distribution in Makassar City Using Hybrid DBSCAN-Gaussian Mixture Model. *Jurnal Teknik Informatika (Jutif)*, *6*(5), 5237–5250. https://doi.org/10.52436/1.jutif.2025.6.5.5254

Rodrigues, G. C., & Braga, R. P. (2021). Evaluation of NASA POWER Reanalysis Products to Estimate Daily Weather Variables in a Hot Summer Mediterranean Climate. *Agronomy*, *11*(6), 1207.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9627

https://doi.org/10.3390/agronomy11061207

Slaoui, S., & Dafir, Z. (2021). A Parameter-free Clustering Algorithm based K-means. *International Journal of Advanced Computer Science and Applications*, *12*(3). https://doi.org/10.14569/IJACSA.2021.0120372

Stryhal, J., & Plavcová, E. (2023). On using self-organizing maps and discretized Sammon maps to study links between atmospheric circulation and weather extremes. *International Journal of Climatology*, *43*(6), 2678–2698. https://doi.org/10.1002/joc.7996

Uston Nawawi Christanto, Brina Miftahurrohmah, Bariyah, T., Kuswanto, H., & Faria, N. (2025). CLUSTER-BASED MACHINE LEARNING APPROACHES FOR PREDICTING DAILY MAXIMUM TEMPERATURES IN INDONESIA UNDER CLIMATE CHANGE. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, *11*(1), 236–249. https://doi.org/10.33480/jitk.v11i1.6749

Yoon, J.-S., Park, S., & Chung, I.-U. (2025). Interpretational Pitfalls in SOM-Based Clustering: A Case Study of Extreme Cold Events in South Korea. *Atmosphere*, *17*(1), 44. https://doi.org/10.3390/atmos17010044

Zehri, S., Yulihastin, E., Marpaung, F., Adiputra, A., Mushoddik, Purwadani, N. N., & Gammamerdianti. (2025). Diverse impact of 2023 El Niño on weather patterns over the Indonesian Maritime Continent. *Journal of Southern Hemisphere Earth Systems Science*, *75*(2). https://doi.org/10.1071/ES25005

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.