e-ISSN 2620 – 8369
VOLUME 9 (ISSUE 2) DECEMBER 2025: 46-54
https://doi.org/10.21009/JSA.09205

APPLICATION OF THE GUSTAFSON–KESSEL ALGORITHM FOR IDENTIFYING SPATIAL PATTERNS OF NATURAL DISASTERS IN EAST NUSA TENGGARA

Mitha Rabiyatul Nufus[1*], Chandrawati[2], Erlyne Nadhilah Widyaningrum[3]

[1]Kupang State Agricultural Polytechnic
Jl. Prof. Dr. Herman Johanes, Kupang 85391, Indonesia.
[2]Hamzanwadi University
Jl. Gajah Mada No. 101, East Lombok, West Nusa Tenggara 83612, Indonesia.
[3]Mulawarman University
Jl. Kuaro, Samarinda, East Kalimantan 75123, Indonesia.

Corresponding author's e-mail: ****mhytha.nufus88@gmail.com****

ABSTRACT

Article History:
Received: 19, November 2025
Revised: 30, November 2025
Accepted: 29, December 2025
Published: 31, December 2025
Available online.

Keywords:
Clustering; East Nusa Tenggara; Gustafson-Kessel; Natural Disasters

This study examines spatial patterns of disaster vulnerability across districts and cities in East Nusa Tenggara Province, one of Indonesia's most disaster-prone regions. Although previous studies have highlighted the province's exposure to multiple hazards, limited attention has been given to clustering methods capable of capturing non-homogeneous and elliptical data structures. This research aims to classify regional disaster vulnerability based on the characteristics of disaster occurrences and to provide empirical support for more targeted mitigation strategies. Secondary data on floods, forest fires, hurricanes, and landslides recorded in 2023 were analyzed using the adaptive Gustafson–Kessel clustering algorithm. The optimal number of clusters was determined using the Silhouette validity index. The results identify three distinct vulnerability groups: regions highly prone to multiple types of disasters, regions predominantly affected by a single hazard, and regions with relatively low disaster risk. The resulting spatial patterns reveal clear differences in disaster intensity and complexity among regions, emphasizing the need for location-specific disaster management policies. This study contributes to disaster risk analysis by demonstrating the applicability of the Gustafson–Kessel algorithm in capturing complex spatial vulnerability patterns that are often overlooked by conventional clustering approaches.

How to cite this article:
M. R. Nufus, Chandrawati, E. N. Widyaningrum, "APPLICATION OF THE GUSTAFSON–KESSEL ALGORITHM FOR IDENTIFYING SPATIAL PATTERNS OF NATURAL DISASTERS IN EAST NUSA TENGGARA", Jurnal Statistika dan Aplikasinya, vol. 9, iss. 2, pp. 46 – 54, December 2025

Journal e-mail: *jsa@unj.ac.id*
Research Article · Open Access

1. INTRODUCTION

Indonesia is a country with a high level of vulnerability to natural disasters due to its complex geographical, geological, hydrological, and demographic conditions [1]. The geographical position of Indonesia places it at the convergence zone of the Asian, Australian, Indian Ocean, and Pacific tectonic plates. This tectonic setting makes Indonesia highly susceptible to various types of natural disasters [2]. East Nusa Tenggara (NTT), a province in Indonesia, is recognized as one of the country's most disaster-prone regions. NTT's location in an archipelagic area with active geological structures makes it prone to disasters such as droughts, floods, hurricanes, landslides, and forest fires. These conditions not only cause material losses but also pose serious threats to the livelihoods of communities and regional development.

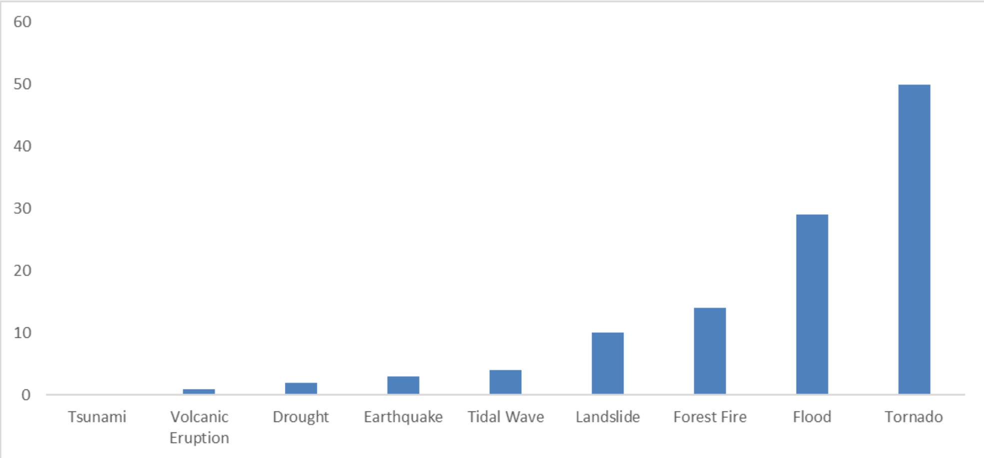

Figure 1. Number of Natural Disasters in East Nusa Tenggara Province (2023)

Figure 1. illustrates that the majority of natural disasters occurring in East Nusa Tenggara Province are hurricanes, floods, forest fires, and landslides. This suggests that the region is more frequently affected by disasters associated with extreme weather and climatic conditions, which aligns with the characteristics of its archipelagic geography and prolonged dry seasons, both of which are influenced by its geographic location and global atmospheric dynamics [3]. According to Statistics Indonesia as cited in [4], the number of people affected and displaced by extreme weather events in East Nusa Tenggara reached 5,376 individuals, representing a 41.4% increase from the previous year. To support effective disaster mitigation and response efforts, a comprehensive understanding of the patterns and characteristics of disaster occurrences in each region is essential. One approach that can be employed is clustering of disaster data, which enables more systematic identification of spatial and temporal patterns. Accurate clustering results can help policymakers develop more targeted, data-driven disaster management strategies tailored to the specific needs of affected areas.

This study aims to cluster the regions within East Nusa Tenggara Province based on the characteristics of natural disaster events using the Gustafson-Kessel Clustering method. Traditional clustering approaches such as K-Means and Fuzzy C-Means (FCM) have limitations in representing non-homogeneous data distributions, particularly in terms of cluster shapes and orientations. In the context of spatial disaster data, which often exhibit elliptical or anisotropic distributions, the Gustafson-Kessel (GK) clustering method offers a more relevant advantage. GK Clustering is an extension of FCM that incorporates a local covariance matrix for each cluster, enabling it to capture more complex data structures and produce more accurate clustering outcomes for multidimensional data, such as natural disaster occurrences. The results of this clustering analysis are expected to provide valuable spatial insights for disaster mitigation planning, resource allocation, and enhancing community preparedness in disaster-prone areas.

2. METHODS

Material and Data

Secondary data for this study were derived from the East Nusa Tenggara Provincial Statistics Agency, detailing natural disaster occurrences in all districts and cities throughout the province in 2023. The unit of observation is each district/ cities in East Nusa Tenggara Province, with the sample including all administrative units, totaling 22 districts and cities. All disaster-related variables are measured on a ratio scale and represent the total number of occurrences recorded in each district/ cities during the 2023 observation period. Specifically, X_1denotes the annual frequency of flood events, X_2represents the number of forest fire occurrences, X_3indicates the number of tornado events, and X_4corresponds to the number of landslide occurrences in each district/municipality.

Gustafson-Kessel Clustering

Gustafson-Kessel Clustering is a method developed from Fuzzy C-Means clustering by utilizing an adaptive distance measure, allowing it to accommodate various geometric shapes of membership functions that are suitable for a given dataset [5]. The algorithm is as follows [6].

1. Inputting the data to be clustered.
2. Determining the Number of Clusters to be Formed $(1 < c < N)$, weighting exponent value $(m > 1)$, maximum number of iterations (MaxIter), the minimum expected error threshold $(\varepsilon > 0)$, the initial value of the objective function= 0 and the initial iteration $(t = 1)$.
3. Generate random numbers μ_{ik} where $i = 1,2,3, \ldots, c$ and $1 \leq k \leq N$ as the elements of the initial partition matrix $\boldsymbol{U_0}$,

$$\sum_{i=1}^{c} u_{ik} = 1 \tag{1}$$

4. Calculating the centroid of the kth cluster using the formula,

$$v = \frac{\sum_{k=1}^{N} u_{ik}^{m} x_k}{\sum_{k=1}^{N} u_{ik}^{m}} \tag{2}$$

5. Calculating the covariance matrix for each group using the formula,

$$F_i = \frac{\sum_{k=1}^{N} u_{ik}^{m} (x_k - v_i)(x_k - v_i)^T}{\sum_{k=1}^{N} u_{ik}^{m}} \tag{3}$$

6. Calculating distance using a formula,

$$D_{ikA_i}^2 = (x_k - v_i)^T [\rho_i \det(F_i)]^{1/h F_i^{-1}(x_k - v_i)} \tag{4}$$

7. Calculating the objective function at the t-th iteration using the formula,

$$J_{GK}(X; U, V, A) = \sum_{i=1}^{c} \sum_{k=1}^{N} (u_{ik})^m D_{ikA_i}^2 \tag{5}$$

8. Calculating the latest value of the membership function U_{t+1}

$$u_{ik} = \left[\sum_{j=1}^{c} \left(\frac{D_{ikA_i}}{D_{jkA_i}} \right)^{\frac{2}{m-1}} \right]^{-1} \tag{6}$$

Comparing membership values in the U matrix. If $||U_{t+1} - U_t|| < \varepsilon$ or (t > MaxIter) thus, the process is considered to have converged. The threshold value (ε) is a very small value approaching zero. If $||U_{t+1} - U_t|| \geq \varepsilon$ thus, the process returns to Step 3 for the next iteration.

Preprocessing and Data Preparation

Before performing clustering analysis, a data preprocessing stage was conducted to ensure the quality and comparability of the variables. This stage included handling missing values, identifying potential outliers, and standardizing the data to prevent variables with larger scales from dominating the clustering process. These steps are essential to obtain meaningful and reliable clustering results.

1. Handling Missing Values
   The first step involved examining the dataset for missing values across all variables. When missing values were identified in a relatively small proportion, mean imputation was applied to preserve the number of observations while maintaining the overall data structure [7]. The imputed mean for variable j is computed as,
   $$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \tag{7}$$
2. Outlier Detection
   Outlier detection is critical for ensuring that extreme values do not distort distance-based clustering. A standard approach uses z-score transformation, where observations with absolute z-scores greater than a threshold (e.g., 3) are flagged for further review [8]. Z-score formula:
   $$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \tag{8}$$
3. Data Standardization
   Given the varying scales of measurements across variables, standardization ensures that each feature contributes equally to the clustering process. Standardization transforms each variable to have zero mean and unit variance, a common practice for distance-based algorithms. With standardization formula:
   $$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \tag{9}$$
4. Distance Measure
   After standardization, a distance measure is selected to quantify similarity between observations. For numerical data, Euclidean distance is widely used due to its straightforward interpretation and compatibility with many clustering algorithms [9].

Validity Index

Validity index is a measure used to determine the most optimal number of clusters by testing several variations of cluster counts. In this study, the validity index used is the Silhouette index. The Silhouette method evaluates the quality of the resulting clusters by combining the concepts of cohesion (similarity within a cluster) and separation (difference between clusters). This index calculates the average dissimilarity between points within the same cluster and those in different clusters to assess how well the clusters are formed [10].

$$S = \frac{b_{(i)} - a_{(i)}}{max\{a_{(i)}, b_{(i)}\}} \tag{10}$$

Here, $a_{(i)}$ refers to the average intra-cluster distance for point i, whereas $b_{(i)}$ corresponds to the lowest average inter-cluster distance between point i and the points belonging to another cluster. In general, Silhouette values close to 1 indicate well-separated and compact clusters, while values near 0 suggest overlapping clusters. Several studies suggest that Silhouette scores above 0.5 reflect reasonably well-structured clusters, and values exceeding 0.7 indicate a strong and reliable clustering structure [11]. Therefore, the number of clusters with the highest Silhouette value and a score exceeding these thresholds was considered the most appropriate solution in this study.

Step of Analysis

The steps carried out in this study are as follows:

1. Collecting and performing preprocessing on natural disaster data in East Nusa Tenggara Province.
2. Conducting clustering assumption tests, including multivariate normality testing, data adequacy testing, and multicollinearity testing.
3. Conducting clustering analysis with the Gustafson-Kessel approach to establish the appropriate number of clusters.
4. Visualizing and interpreting the clustering results using the Gustafson-Kessel method.
5. Comparing the clustering results using the Silhouette validity index for 2, 3, and 4 clusters.
6. Drawing conclusions and providing recommendations.

3. RESULTS

Data Characteristics

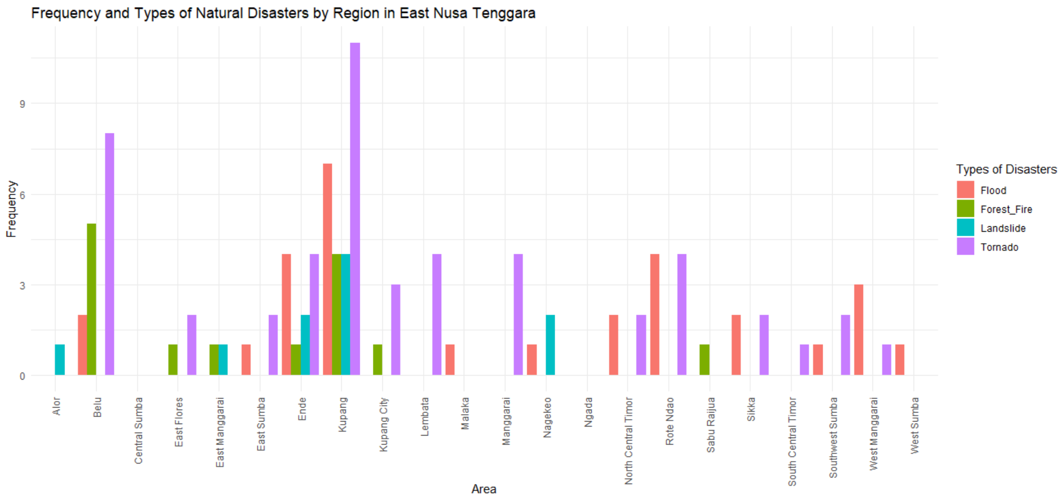

Figure 2. Number of Natural Disasters in Each Regency/City (2023)

Figure 2 illustrates the number and types of natural disasters in various regions of East Nusa Tenggara (NTT), including hurricanes, floods, fires, and landslides. In general, hurricanes occur most frequently, especially in Kupang Regency, Kupang City, and Alor Regency. Floods are also dominant in Kupang Regency, East Flores Regency, and Ngada Regency, even comparable to tornado occurrences in Ngada and Rote Ndao Regencies. Fires are most prevalent in Kupang City, Alor Regency, and Ngada Regency, whereas landslides are the least frequent, appearing only in a few areas such as Ende and Nagekeo Regencies. Several regions, including the West Manggarai, Malaka, and Central Sumba Regencies, experienced only a few disaster events. The province of NTT is highly vulnerable to hurricanes and floods, particularly in and around Kupang Regency. The distribution of disasters varies significantly across regions, with some areas experiencing more than one type of disaster. This highlights the need for disaster mitigation strategies that are tailored to the specific characteristics of each region.

Clustering Using Gustafson-Kessel

The Gustafson-Kessel method will be used to determine the optimal number of clusters in the 2023 natural disaster data in Indonesia by taking into account the following cluster validity indices.

Table 3. Comparison of Validity Indices for Each Cluster

Number of Cluster	Iteration	Time	*Silhouette*
2	169	0.36	0.0038

3	**20**	**0.07**	**0.4861**
4	25	0.13	0.2034

Referring to Table 3 above, the Gustafson-Kessel method produced an optimal number of three clusters for the 2023 natural disaster data in East Nusa Tenggara Province. The clustering process using the Gustafson–Kessel method produced a Silhouette Score of 0.4861. In general, the Silhouette index ranges from −1 to 1, where values closer to 1 indicate well-defined and clearly separated clusters, while values approaching 0 suggest increasing overlap between clusters. Based on commonly accepted interpretive thresholds, Silhouette values between 0.25 and 0.50 are typically associated with a moderate level of cluster structure. Accordingly, the clustering results in this study demonstrate that the underlying cluster patterns are reasonably established, although the separation among clusters cannot be considered strong. This outcome indicates the presence of overlapping data characteristics, which is understandable given the complexity and heterogeneity of the dataset. Nevertheless, the Gustafson–Kessel method remains effective in capturing clusters with flexible and non-spherical shapes, supporting the relevance of the obtained results for exploratory analytical purposes. This result is based on the highest validity index value indicated by the Silhouette score for three clusters. The optimal number of clusters was determined solely based on the highest Silhouette index value, which was obtained for three clusters. The computation time and iteration count are reported only to describe the computational efficiency of the algorithm. Furthermore, by setting the number of clusters to three, the allocation of data points to each cluster is provided in the following summary.

Table 4. Number of Members in Each Cluster

Cluster	Members
1	3
2	3
3	16

Based on Table 4, three clusters of regional classifications were identified according to the characteristics of natural disasters occurring in each area. The first cluster consists of multi-disaster-prone areas, which are regions that exhibit high frequencies of more than one type of natural disaster. Examples include areas frequently affected by combinations of disasters such as floods and hurricanes or landslides and forest fires. The second cluster represents single-disaster-prone areas, characterized by the dominance of one particular type of disaster—such as regions that frequently experience floods or forest fires, while other types of disasters rarely or never occur. The third cluster includes low-disaster-risk areas, which are regions with low frequencies of all types of disasters or areas that have almost no record of disasters at all.

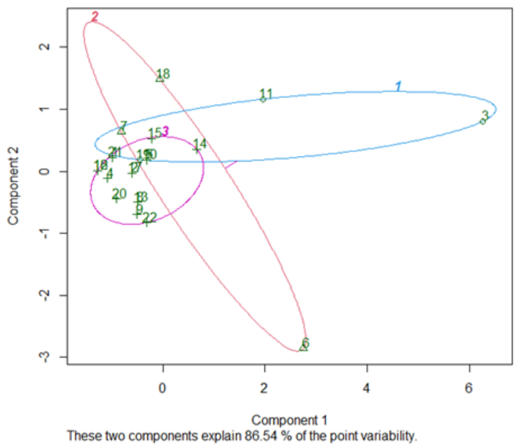

Figure 4. Cluster Plot of Natural Disaster Data in East Nusa Tenggara Province

To clarify Figure 4 above, the members of each cluster are presented in the form of a map by region using the Gustafson-Kessel method, as follows.

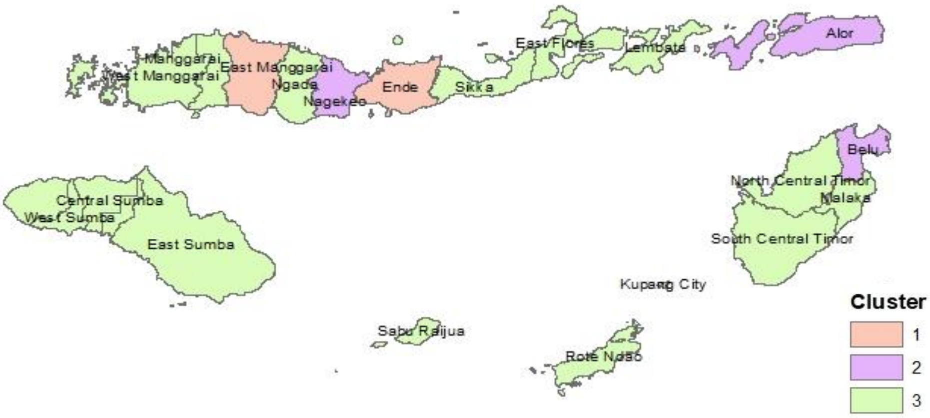

Figure 5. Regional Map Based on Clusters

Based on Figure 5, it is observed that Kupang Regency, Ende Regency, and East Manggarai Regency are classified under Cluster 1, representing areas with high vulnerability to multi-hazard disasters. Kupang Regency, being a hub of human activities and a major transportation route, is particularly prone to hurricanes, flash floods, and extreme droughts. Ende Regency, located near the volcanic area of Mount Kelimutu, is highly susceptible to volcanic eruptions, landslides, and earthquakes. Meanwhile, East Manggarai Regency, characterized by mountainous terrain, is also vulnerable to landslides and tectonic earthquakes. Cluster 2 includes regions that are vulnerable to specific types of disasters, namely Belu Regency, Alor Regency, and Nagekeo Regency. Belu Regency, which borders Timor-Leste, lies in a semi-arid zone and frequently experiences droughts and forest fires. Alor Regency, an archipelagic area, is vulnerable to earthquakes and tsunamis. Nagekeo Regency, with its diverse topography of mountains and lowlands, is prone to landslides and seasonal droughts. The final group, Cluster 3, consists of 15 other regencies and 1 city not previously mentioned. These areas are considered to have low disaster vulnerability. Regions in this cluster are generally geologically stable or experience disasters that are more predictable and manageable. For example, although Sumba Regency has a dry climate, it rarely experiences major geological disasters. Sabu Raijua, Rote Ndao, and the City of Kupang have relatively flat topography and are not located in active seismic zones. Moreover, while the Lembata and East Flores Regencies do have active volcanoes, their volcanic activity has been relatively low in recent years.

To strengthen the interpretation of the clustering results, a summary of the average values of each hazard-related variable for all clusters is presented in Table 5.

Table 5. Mean values of each hazard-related variable

Variable	Cluster 1	Cluster 2	Cluster 3
Flood (X1)	3.667	1.000	0.938
Forest Fire (X2)	2.000	1.667	0.188
Tornado (X3)	5.000	2.667	1.688
Landslide (X4)	2.333	1.000	0.000

The table shows that Cluster 1 consistently exhibits higher mean values across most hazard indicators, particularly for drought, flood, and geological hazards. These numerical characteristics confirm that areas grouped in Cluster 1 experience a relatively higher level of multi-hazard exposure compared to other clusters. The spatial distribution shown in Figure 5 further supports this result, where Kupang Regency, Ende Regency, and East Manggarai Regency are assigned to Cluster 1. The high cluster mean values align with the physical and environmental conditions of these regions, such as intensive human activities, proximity to active volcanic zones, and complex topography. Therefore, the classification of Cluster 1 as a high multi-hazard vulnerability zone is supported both quantitatively and spatially.

4. DISCUSSIONS

The clustering results generated by the Gustafson–Kessel method indicate that disaster occurrences in East Nusa Tenggara form clear spatial patterns rather than emerging randomly across regions. The separation into three clusters reflects differences in environmental conditions, topography, and local climatic factors that shape each region's exposure to specific hazards. Areas grouped in the first cluster face multiple hazards because they lie in zones where geological activity, extreme weather, and land-use pressures occur simultaneously. Meanwhile, regions in the second cluster tend to experience one dominant disaster type due to more uniform ecological characteristics. The third cluster comprises areas with relatively stable environmental conditions, resulting in lower disaster frequencies.

These findings align with Rahmatika et al. (2015), who noted that the GK method performs well when data distributions are irregular or elliptical. The results also support the view presented by Azizah et al. (2022) that adaptive clustering techniques are needed when handling ecological data with complex variation. However, unlike Azzahra and Koesyanto (2023), who emphasized institutional factors in disaster management outcomes, this study highlights the influence of physical geographic conditions. Together, these perspectives show that disaster vulnerability is shaped by both environmental and governance-related factors.

5. CONCLUSION

This study applied the Gustafson–Kessel clustering method to identify regional patterns of disaster characteristics. Several clustering scenarios were evaluated to determine the most appropriate partition. Based on the Silhouette index, the three-cluster solution demonstrated the highest validity value, indicating a clearer and more coherent grouping structure compared to alternative cluster numbers. The final clustering results categorized the regions into three distinct groups. The first cluster represents areas exposed to multiple types of disasters, reflecting complex and diverse risk profiles. The second cluster consists of regions dominated by a single recurring hazard, suggesting more specific and concentrated disaster patterns. The third cluster includes regions with relatively low disaster frequency, indicating comparatively lower risk levels. Overall, the clustering results provide meaningful insights into the spatial variation of disaster characteristics and offer a useful basis for supporting region-specific disaster risk management and mitigation planning.

6. REFERENCES

[1] Q. A. Azzahra and H. Koesyanto, "Efektivitas tim reaksi cepat," *Higeia Journal of Public Health Research and Development*, vol. 7, no. 1, pp. 1–11, 2023, doi: 10.15294/higeia.v7i1.58841.

[2] M. Mughron, "Efektivitas Badan Penanggulangan Bencana Daerah (BPBD) dalam pelaksanaan program kelurahan siaga bencana di Kota Semarang," *Jurnal Administrasi Publik*, 2019.

[3] A. Agustina, M. Ariska, and H. Akhsan, "Dinamika temperatur ekstrem di Nusa Tenggara," *Jurnal Online Universitas Jambi*, pp. 104–109, 2024.

[4] Badan Pusat Statistik Provinsi Nusa Tenggara Timur, "Jumlah korban yang diakibatkan bencana alam menurut kabupaten/kota di Provinsi Nusa Tenggara Timur," 2024. [Online]. Available: https://ntt.bps.go.id/id/statistics-table/3/T0VGNVR6UkVaakVyUmxWTU1YTlhSQzkzYW1vclFUMDkjMw==/jumlah-korban-yang-diakibatkan-bencana-alam-menurut-kabupaten-kota-di-provinsi-nusa-tenggara-timur--2020.html?year=2023

[5] S. Singh and P. Singh, "Speaker specific feature based clustering and its application in language independent forensic speaker recognition," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 4, pp. 3508–3518, 2020, doi: 10.11591/ijece.v10i4.pp3508-3518.

[6] L. Rahmatika, Suparti, and D. Safitri, "Analisis kelompok dengan algoritma C-means dan Gustafson–Kessel clustering pada indeks LQ45," *Jurnal Gaussian*, vol. 4, no. 3, pp. 543–551, 2015, doi: 10.14710/j.gauss.4.3.543-551.

[7] R. Aschenbruck, G. Szepannek, and A. F. Wilhelm, "Imputation strategies for clustering mixed-type data with missing values," *Journal of Classification*, vol. 40, pp. 2–24, 2023, doi: 10.1007/s00357-022-09422-y.

[8] N. Kholila, M. Mujiono, and D. Wahyudi, "Pemetaan kondisi lingkungan tanam menggunakan K-means clustering," *Jurnal Sistem Informasi dan Teknologi Informasi Komputer (JSITIK)*, vol. 1, no. 2, pp. 37–47, 2023, doi: 10.53624/jsitik.v1i2.182.

[9] M. R. Mubarok, A. T. Harjanto, and R. Renaldy, "Peningkatan performa DBSCAN dengan reduksi dimensi principal component analysis (PCA) dalam klasterisasi tingkat kemiskinan di Indonesia," *JINTEKS*, vol. 7, no. 3, pp. 1176–1184, 2025, doi: 10.51401/jinteks.v7i3.6129.

[10] A. N. Azizah, T. Widiharih, and A. R. Hakim, "Kernel K-means clustering untuk pengelompokan sungai di Kota Semarang berdasarkan faktor pencemaran air," *Jurnal Gaussian*, vol. 11, no. 2, pp. 232–243, 2022, doi: 10.14710/j.gauss.v11i2.35470.

[11] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.