

Application of the Gustafson–Kessel Algorithm for Identifying Spatial Patterns of Natural Disasters in East Nusa Tenggara

Mitha Rabiyyatul Nufus^{1*}, Chandrawati², Erlyne Nadhilah Widyaningrum³

¹Kupang State Agricultural Polytechnic, Jl. Prof. Dr. Herman Johanes, Kupang 85391, Indonesia.

²Hamzanwadi University, Jl. Gajah Mada No. 101, East Lombok, West Nusa Tenggara 83612, Indonesia.

³Mulawarman 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 addresses the persistent challenge of identifying spatial patterns of natural disasters in East Nusa Tenggara, one of Indonesia's most hazard-prone provinces. While previous research has documented the region's exposure to multi-hazard events, limited attention has been given to clustering approaches capable of capturing the non-homogeneous and elliptical structure of disaster data. This study therefore examines how regional disaster vulnerability can be classified more accurately using an adaptive clustering method. The purpose of this research is to develop a spatial classification of districts and cities in East Nusa Tenggara based on the characteristics of disaster occurrences and to provide empirical evidence that supports more targeted mitigation policies. Secondary data on floods, forest fires, hurricanes, and landslides for the year 2023 were analyzed. The study applied the Gustafson–Kessel clustering algorithm, preceded by multivariate normality testing, data adequacy assessment, and multicollinearity evaluation to ensure methodological suitability. The optimal number of clusters was determined using the Silhouette validity index. Clustering results were then interpreted spatially to identify distinct vulnerability profiles across the province. The findings show that the Gustafson–Kessel method yields three meaningful clusters: (1) multi-hazard-prone regions such as Kupang, Ende, and East Manggarai; (2) regions dominated by a single type of disaster, including Belu, Alor, and Nagekeo; and (3) low-risk areas comprising the remaining districts. These classifications reveal clear spatial patterns and highlight substantial differences in hazard intensities across the province, offering practical input for localized disaster management strategies. The study is limited by its reliance on one-year cross-sectional data and the absence of socio-economic variables that may influence vulnerability. The novelty of this work lies in the application of the Gustafson–Kessel algorithm, which has not been widely used in Indonesian disaster studies, to capture complex spatial patterns that conventional clustering methods often overlook.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License.

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 – 56, December 2025

Copyright © 2025 Author(s)

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.

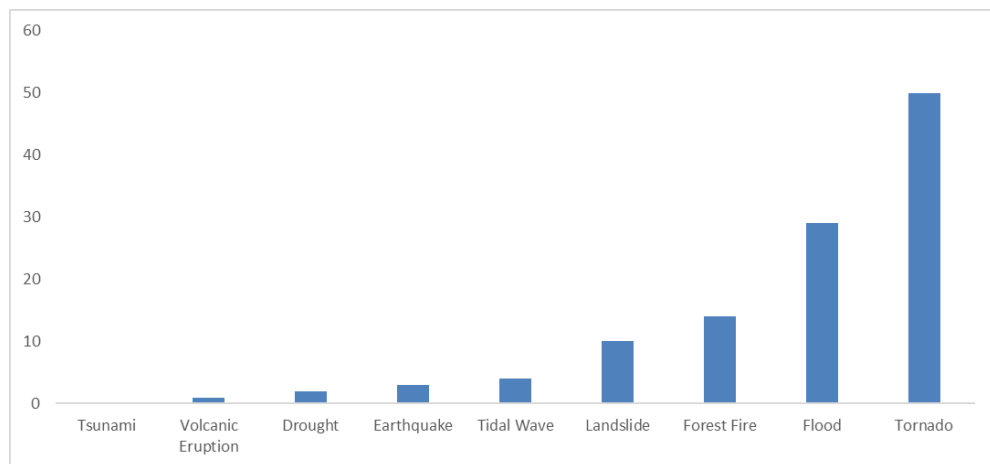


Figure 1. Number of Natural Disasters in East Nusa Tenggara Province (2023)

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 dataset covers the number of flood occurrences (X1), forest fire occurrences (X2), tornado occurrences (X3), and landslide occurrences (X4).

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 $i = 1, 2, 3, \dots, c$ and $1 \leq k \leq N$ as the elements of the initial partition matrix $\mathbf{U}_0$,

$$\sum_{i=1}^c u_{ik} = 1 \quad (1)$$

4. Calculating the centroid of the k^{th} cluster using the formula,

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

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

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

6. Calculating distance using a formula,

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

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

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

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

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

Comparing membership values in the U matrix. If $||U_{t+1} - U_t|| < \varepsilon$ or ($t > \text{MaxIter}$) 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.

Assumption Testing

Before conducting the analysis using a clustering method, it is necessary to perform assumption testing to ensure that the clustering method is appropriate and applicable to the data used.

Multivariate Normality Assumption

The multivariate normality assumption is evaluated to determine whether the data distribution is normal, which is assessed by examining the plot of d_j^2 against the chi-square quantile $\left(\frac{j-0,5}{n}\right)$. This assumption is tested using the following hypotheses [7].

H_0 : The data does not follow a multivariate normal distribution

H_1 : The data follows a multivariate normal distribution

The assessment of multivariate normality was carried out through the following procedures,

- Computing $d_j^2 = (\mathbf{X}_j - \bar{\mathbf{X}})' \mathbf{S}^{-1} (\mathbf{X}_j - \bar{\mathbf{X}})$; $j = 1, 2, 3, \dots, n$ (7)

Where d_j^2 is the distance value of the j^{th} data, $\mathbf{X}_j$ is the vector of the j^{th} variable, $\bar{\mathbf{X}}$ is the mean vector value, and $\mathbf{S}^{-1}$ is the inverse of the variance-covariance matrix

- Sorting the d_j^2 values of all observations obtained so that $d_1^2 \leq d_2^2 \leq \dots \leq d_j^2$
- Constructing a Q-Q Plot or Chi-Square Plot with d_j^2 values on the X-axis and upper quantile values on the Y-axis.

$$\left(d_j^2, \chi_p^2 \left(\frac{j-0,5}{n} \right) \right) \quad (8)$$

If the plot tends to form a straight line above 0.5 and the value of $d_j^2 \leq \chi_{p;0,05}^2$ or the p-value > 0.05 , then the data are concluded to follow a multivariate normal distribution [8].

Data Adequacy Assumption

The adequacy of the data was tested using the Kaiser-Meyer-Olkin (KMO) formula, with the hypothesis for the data adequacy test stated as follows [9].

H_0 : The data is sufficient for analysis

H_1 : The data is insufficient for analysis

Statistics test:

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} \quad (9)$$

where $i = 1, 2, 3, \dots, p$ and $j = 1, 2, 3, \dots, p$. And r_{ij} is the correlation coefficient between variable i and variable j ; and a_{ij} is the partial coefficient between variable i and variable j .

If the KMO calculation result is less than 0.5, the decision is to reject H_0 , which means the data is insufficient for analysis.

Multicollinearity Assumption

The multicollinearity test can be conducted using Bartlett's test, which aims to assess the correlation between two or more independent variables by evaluating the correlation values among those variables [8].

H_0 : $\rho = 1$ (there is a correlation)

H_1 : $\rho \neq 1$ (there is no correlation)

Statistics test:

$$X^2 = - \left(C - 1 - \frac{2N + 5}{6} \right) \ln |R| \quad (10)$$

with $|R|$ is the determinant value of the correlation matrix, N is the number of variables and C is the number of observations.

H_0 will be rejected if $X^2 > X^2_{\alpha, \frac{1}{2}N(N-1)}$ p-value < 0.05. This indicates that there is no multivariate correlation or multicollinearity among the variables studied.

Validity Index

The 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 indexing method. The Silhouette indexing 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].

$$S = \frac{b_{(i)} - a_{(i)}}{\max\{a_{(i)}, b_{(i)}\}} \quad (11)$$

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. The highest Silhouette value indicates the most appropriate number of clusters to select.

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

This section presents the outcomes of applying the Gustafson–Kessel clustering approach to the dataset that has been processed in the earlier stage of the study. The analysis aims to observe how the data objects are grouped based on their similarity patterns, enabling a deeper understanding of the characteristics of each resulting cluster. The discussion does not only highlight the allocation of objects into specific groups but also evaluates the separation quality and variability in the shape of the detected clusters.

Data Characteristics

This section presents an initial overview of the data used in the study. The collected data are shown using charts to display the distribution of each variable. Through these visualizations, we can see general patterns, differences between samples, and the possibility of unusual values in the dataset.

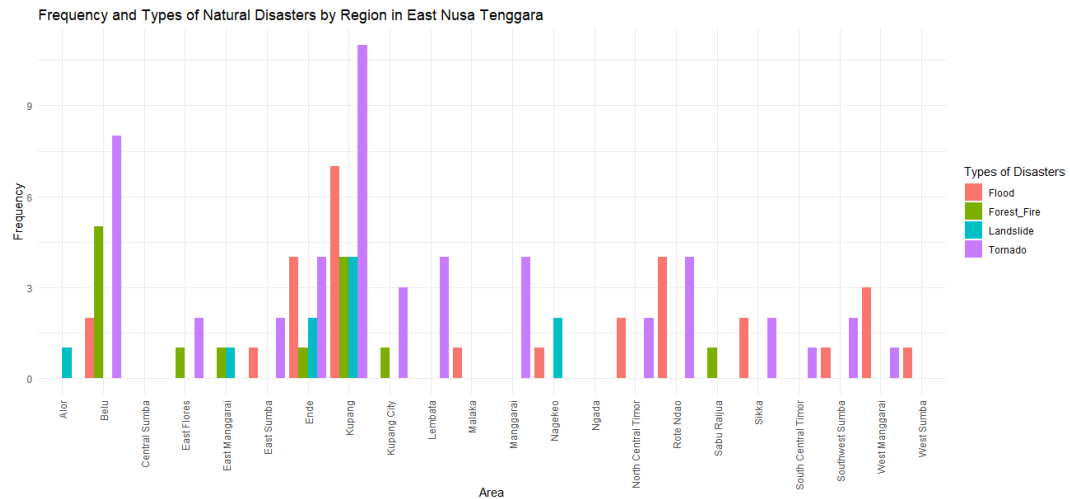


Figure 2. Number of Natural Disasters in Each Regency/City (2023)

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 Assumption Testing

Multivariate Normality Assumption

Before performing clustering analysis on natural disaster data using Mahalanobis distance, the primary requirement is that the assumption of multivariate normality in the data must be satisfied.

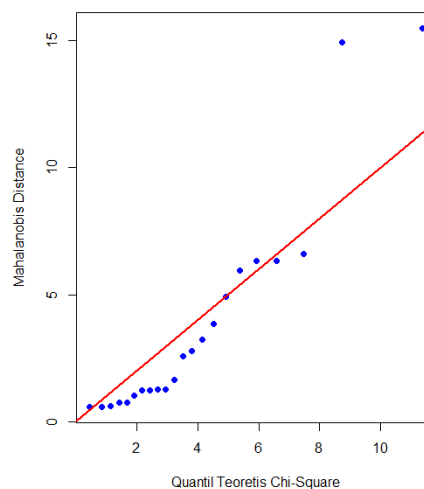


Figure 3. Q-Q Plot of Mahalanobis Distance

Figure 3 does not yet allow for a visual conclusion regarding whether the data follow a multivariate normal distribution. Although the displayed Q-Q plot forms a straight line and the data points are spread

around it, the presence of outliers prevents a definitive visual assessment. Therefore, further testing is necessary by examining the correlation value of the Mahalanobis distance, as follows:

Table 1. Mahalanobis Distance Correlation Results

		Mahalanobis Distance	qi
Mahalanobis Distance	Pearson Correlation	1	0.933
	P-Value		0.001
	N	22	22

As presented in Table 1, the correlation coefficient of 0.933 demonstrates a robust relationship, while the corresponding p-value of 0.001 ($p < 0.05$) affirms the significance of this correlation. Hence, it is reasonable to infer that the data are drawn from a sample that adheres to a multivariate normal distribution.

Data Adequacy Assumption

The adequacy of the data in this study was tested using the KMO test, with the following results.

Table 2. The result of the KMO test

X1 MSA	X2 MSA	X3 MSA	X4 MSA	Overall MSA
0.63	0.62	0.61	0.72	0.64

To meet the assumption of sample adequacy, the KMO value must be at least 0.6. Based on Table 2, all four variables in the study have KMO values greater than 0.6, with an overall MSA (Measure of Sampling Adequacy) value of 0.64. This indicates that all the variables meet the data adequacy requirements for further analysis.

Multicollinearity Assumption

The multicollinearity assumption test was conducted using Bartlett's Test of Sphericity with the following hypothesis testing,

$H_0 : \rho = 0$ (no correlation among variables)

$H_1 : \rho \neq 0$ (a correlation exists among variables)

The test results show a p-value of 2.09×10^{-7} , which is below the 0.05 significance level. Therefore, H_0 is rejected, indicating that the variables are correlated. This shows that the dataset contains a relationship structure among variables. The presence of correlation suggests that multicollinearity may exist, thus further evaluation such as VIF is required to confirm the severity of multicollinearity.

Table 3. VIF Result

Variable	VIF	Tolerance (1/VIF)
X2	2.580	0.387
X3	2.714	0.368
X4	1.336	0.748

From Table 3, we know that the multicollinearity assessment was carried out using the Variance Inflation Factor (VIF) for the predictors included in the regression model. The results show that the VIF

values range from 1.336 to 2.714, which are well below the commonly accepted threshold of 5 or even the more conservative value of 10. This indicates that there is no evidence of harmful multicollinearity among the predictors. Furthermore, tolerance values above 0.1 confirm that each predictor still contributes unique information to the model and is not highly dependent on the others. Therefore, all predictors can be retained for further analysis, as they do not pose multicollinearity concerns in the regression model.

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 4. Comparison of Validity Indices for Each Cluster

Number of Cluster	Iteration	Time	<i>Silhouette</i>
2	169	0.36	0.0038
3	20	0.07	0.4861
4	25	0.13	0.2034

Referring to Table 4 above, the Gustafson-Kessel method produced an optimal number of three clusters for the 2023 natural disaster data in East Nusa Tenggara Province. This result is based on the highest validity index value indicated by the Silhouette score for three clusters, achieved with the shortest time and fewest iterations. Furthermore, by setting the number of clusters to four, The allocation of data points to each cluster is provided in the following summary.

Table 5. Number of Members in Each Cluster

Cluster	Members
1	3
2	3
3	16

Based on Table 5, 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.

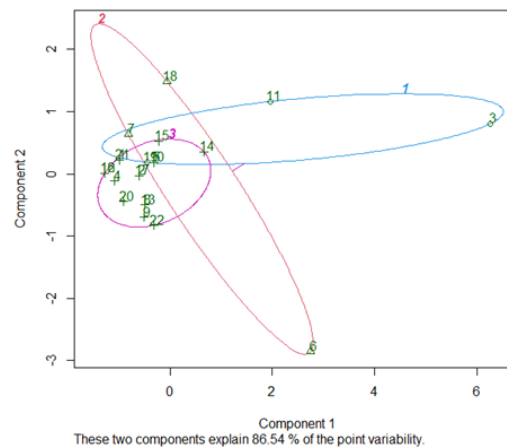


Figure 4. Cluster Plot of Natural Disaster Data in East Nusa Tenggara Province

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.

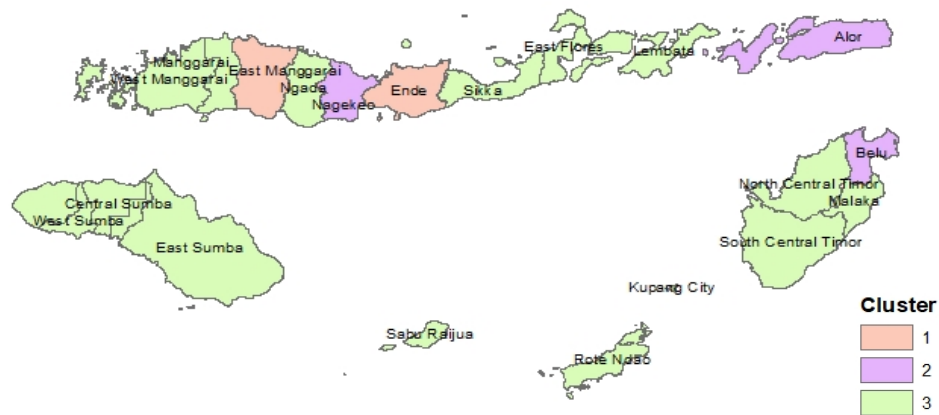


Figure 5. Regional Map Based on Clusters

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.

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

Prior to performing the clustering procedure, preliminary diagnostics were conducted to examine the structure and quality of the data. These checks were intended to ensure that the variables used did not contain redundancy, extreme distortions, or measurement scales that could bias the clustering results. Through this step, the dataset was confirmed to be appropriate for grouping analysis and allowed the clustering process to be carried out with more reliable outcomes. The multivariate normality check showed a strong and significant correlation between the Mahalanobis distance and the chi-square quantiles, indicating that the data followed a multivariate normal pattern. The KMO value of 0.64 confirmed that the variables were adequate for analysis. While Bartlett's test indicated significant correlations among variables, while Bartlett's test showed that the correlations among variables were statistically significant, this result only confirms that the dataset contains a sufficient relationship structure to proceed with multivariate analysis. However, the presence of correlation does not necessarily indicate desirable multicollinearity. Therefore, a further assessment with VIF was conducted to ensure that multicollinearity does not become a problem. Using the Gustafson–Kessel method, several cluster scenarios were tested. The Silhouette index demonstrated that three clusters provided the most coherent structure, with the highest validity score compared to two or four clusters. The final clustering grouped the regions into three categories: the first one is areas exposed to multiple types of disasters, the second one is areas dominated by a single recurring hazard, and the last one is regions with relatively low disaster frequency. These groupings were consistent with the spatial and environmental characteristics of each district, reinforcing the suitability of the Gustafson–Kessel approach for datasets with non-homogeneous patterns.

6. REFERENCES

- [1] Q. A. Azzahra and H. Koesyanto. (2023). "Efektivitas Tim Reaksi Cepat". *Higeia Journal of Public Health Research and Development*, 1-11. DOI: <https://doi.org/10.15294/higeia.v7i1.58841>.
- [2] M. Mughron. (2019). "Efektifitas Badan Penanggulangan Bencana Daerah (BPBD) Dalam Pelaksanaan Program Kelurahan Siaga Bencana Di Kota Semarang". *Jurnal Administrasi Publik*.
- [3] A. Agustina, M. Ariska and H. Akhsan. (2024). "Dinamika Temperatur Ekstrem di Nusa Tenggara". *Jurnal Online Universitas Jambi*, 104-109.
- [4] Badan Pusat Statistik. (2024, Februari 22). Jumlah Korban yang Diakibatkan Bencana Alam Menurut Kabupaten/Kota di Provinsi Nusa Tenggara Timur. Retrieved from Badan Pusat Statistik Provinsi Nusa Tenggara Timur: <https://ntt.bps.go.id/id/statistics->

table/3/T0VGNVR6UkVaakVyUmxWTU1YTlhSQzkzYW1vcIFUMDkjMw==/jumlah-korban-yang-diakibatkan-bencana-alam-menurut-kabupaten-kota-di-provinsi-nusa-tenggara-timur--2020.html?year=2023

- [5] S. Singh and P. Singh. (2020). “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. DOI: <http://doi.org/10.11591/ijece.v10i4.pp3508-3518>
- [6] L. Rahmatika, Suparti and D. Safitri. (2015). “Analisis Kelompok dengan Algoritma C-Means dan Gustafson Kessel Clustering pada Indeks LQ45”. *Jurnal Gaussian*, 4(3). DOI: <https://doi.org/10.14710/j.gauss.4.3.543-551>.
- [7] Rusli. *Pengujian Normal Multivariat dan Homoskedastisitas Matriks Varians-Kovarians Pada Prestasi Belajar dan Kredit Lulus Mahasiswa UNM Angkatan 2014*. Malang: Ilmu Pendidikan UNM, 2014.
- [8] R. Johnson and D. Wichern. *Applied Multivariate Statistical Analysis 6th Edition*. New Jersey: Prentice Hall International, 2007.
- [9] J. Hair, W. Black, B. Babin and R. Anderson. *Multivariate Data Analysis Seventh Edition*. Pearson, 2009.
- [10] A. N. Azizah, T. Widiharah and A. R. Hakim. (2022). “Kernel K-Means Clustering Untuk Pengelompokkan Sungai di Kota Semarang Berdasarkan Faktor Pencemaran Air”. *Jurnal Gaussian*, 232, 11(2). DOI: <https://doi.org/10.14710/j.gauss.v11i2.35470>.