

Clustering of Disaster Risk in Indonesia Regions Using Self-Organizing Maps and K-Means

Hardianti Hafid^{1*}, Isma Muthahharah²

^{1,2} Department of Statistics, Universitas Negeri Makassar, 90224, Indonesia

Corresponding email*: hardiantihf@unm.ac.id

Abstract

Indonesia is one of the countries with a high level of vulnerability to natural disasters, making accurate risk mapping essential to support effective mitigation planning. This study aims to identify multi-disaster risk patterns and support region-based mitigation policies by clustering the 38 provinces of Indonesia based on disaster occurrence characteristics using a hybrid approach of Self-Organizing Maps (SOM) and K-Means. The data were obtained from the Indonesian National Disaster Management Agency (BNPB), covering the frequency and types of disasters such as floods, extreme weather, eruptions, abrasion, earthquakes, forest/land fires, droughts, and landslides. The SOM representation results were clustered using K-Means, with the optimal number of clusters determined through the evaluation of the Davies Bouldin index, Silhouette coefficient, and connectivity measure. The analysis revealed that two clusters provided the best separation: Cluster 1 includes most provinces with medium to low multi-hazard risk, while Cluster 2 consists of West Java, Central Java, and East Java, which have high hydrometeorological risk. The findings from Cluster 2 highlight the urgent need for the government to strengthen disaster mitigation infrastructure and early warning systems in Java Island, which serves as the country's economic and population center. This hybrid SOM and K-Means approach successfully identifies the spatial patterns of disaster risk and can serve as a reference for the government in formulating region-based mitigation strategies.

Keywords: disaster risk, k-means, self-organizing maps

Received :10-08-2025 Revised :08-10-2025 Accepted :14-10-2025 Published :17-10-2025

1. Introduction

Indonesia is one of the countries most vulnerable to natural disasters. Geographically, it lies along the Pacific Ring of Fire and is traversed by three active tectonic plates, placing it at high risk of earthquakes, tsunamis, and volcanic eruptions. Combined with its tropical climate and high rainfall intensity, this also makes Indonesia prone to disasters such as floods, landslides, droughts, and forest fires. [1]. According to [2] report, Indonesia is among the 35 countries with the highest disaster risk globally, underscoring the urgency of implementing spatial data based mitigation strategies as a foundation for decision-making.

To map disaster risk patterns spatially and multivariately, Self-Organizing Maps (SOM) is a commonly used method. SOM, also known as Kohonen maps, is a technique for reducing the dimensionality of multivariate data into two dimensions while preserving the original data's topological relationships. It is often used for exploring patterns and clustering large datasets due to its ability to maintain both the geographical proximity and the original characteristics of the data [3]. [1] Using Self-Organizing Maps (SOM) to cluster provinces in Indonesia based on landslide impacts successfully produced three main clusters, determined through internal cluster validation such as the Dunn index and Silhouette coefficient. Another study by [4] used SOM to cluster provinces based on earthquake impacts, with clusters formed according to the number of casualties and the extent of infrastructure damage.

However, SOM often produces a large number of clusters, equal to the number of neurons in the grid, making thematic interpretation difficult. To address this, a hybrid approach combining SOM and K-Means clustering is often used. SOM is applied for initial mapping and dimensionality reduction, followed by K-Means to simplify the clusters into a controlled number that is easier to interpret. [5] compared SOM and

K-Means for clustering village potential data and found that the combination of the two methods produced more stable results based on the Davies Bouldin index. Another study by [6] applied K-Means to divide post-earthquake hazard zones based on variables such as magnitude, casualty impact, and facility damage, and found this method effective in grouping disaster-prone areas at a regional scale. In addition, a study by [7] compared K-Means and K-Medoids for earthquake risk clustering in Indonesia, showing that K-Medoids yielded a higher silhouette score, but K-Means remains widely used due to its efficiency.

Despite these contributions, most existing studies still concentrate on a single type of disaster such as earthquakes, floods, or landslides without providing an integrated view of multi-hazard risks across regions. Furthermore, previous works have not explored the integration of SOM and K-Means specifically for multi-disaster clustering at the provincial or national scale. The combination of SOM ability for topological mapping with K-Means interpretability offers a promising framework to capture complex spatial disaster patterns. Therefore, this study aims to fill this research gap by applying a hybrid SOM and K-Means approach to classify Indonesian provinces based on multiple disaster types, supported by spatial visualization for regional mitigation planning. In doing so, this research is expected to make a significant contribution to the development of data-driven disaster risk analysis in Indonesia and provide a strategic information base for mitigation and resource allocation

2. Research Methods

This study employs a hybrid approach combining Self-Organizing Maps (SOM) and K-Means clustering to group provinces in Indonesia based on the frequency and types of natural disasters. This method was chosen because SOM excels at topologically mapping the structure of high-dimensional data, while K-Means serves to reduce the number of clusters to make them more thematically interpretable [8].

2.1 Data Sources

The primary data were obtained from the Indonesian National Disaster Management Agency (BNPB) through its official portal at <https://data.bnpb.go.id>, covering the summary of natural disaster occurrences across all provinces in Indonesia throughout 2024. The variables used in the analysis include: Province, Flood, Extreme Weather, Eruption, Coastal Abrasion, Earthquake, Forest/Land Fire, Drought, and Landslide.

Table 1. Disaster Frequency Data in Indonesia 2024

Province	Flood	Weather	Eruption	Abrasion	Earthquake	Forest/ Land Fire	Drought	Landslide
Aceh	47	14	0	0	0	35	1	2
Sumatera Utara	120	43	0	0	0	170	2	15
Sumatera Barat	64	10	3	3	0	7	1	10
Riau	53	2	0	2	0	10	0	1
Jambi	44	7	0	0	0	79	0	2
Sumatera Selatan	69	9	0	0	0	192	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Papua								
Pegunungan	0	0	0	0	0	0	0	3
Papua Barat Daya	4	0	0	0	0	0	0	0

2.2. Data Preprocessing

All preprocessing steps were performed using R. The data preprocessing steps include:

- Data Cleaning: Removing provinces with incomplete data or extreme outliers.
- Data Normalization: Standardizing each variable using the z-score method to ensure equal scaling and prevent any variable from dominating the clustering results.

- c. Transformation to a Numerical Matrix: Converting the dataset into a multivariate numerical matrix without the Province column for analysis purposes.

2.3. Self-Organizing Maps (SOM)

Self-Organizing Maps are used to map high-dimensional data into a two-dimensional structure in the form of a neuron grid. The implementation stages of SOM are as follows: [9]:

- a. Initialization of the SOM grid: Initialize the weights of each node. Before training begins, each node is initialized, typically with small random values.
- b. Model training is carried out over several iterations (epochs) using the batch training algorithm.
- c. For each selected input vector, compute the distance to the weights of all nodes and select the smallest distance as the Best Matching Unit (BMU). To determine the BMU, one common approach is to calculate the Euclidean distance between the weight of each node and the input vector.
- d. Visualization of SOM results in codebook vectors to observe the distribution patterns and proximity between units.

2.4. Clustering Using K-Means

Since the SOM contains a relatively large number of neurons (25 units), the next step is to apply the K-Means algorithm to simplify the clustering results into a more interpretable number of clusters. The stages include:

- a. Extracting the neuron weight vectors (codebook) from the SOM model.
- b. Applying the K-Means algorithm to the SOM codebook with a specified number of clusters. The optimal number of clusters is determined using :

- 1) Davies Bouldin Index value [10],

if we have k clusters, the DBI is often defined as:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{s_i + s_j}{d(c_i, c_j)} \right) \quad (1)$$

where

- s_i : the average distance between the members of cluster i and its centroid
- $d(c_i, c_j)$: the distance between the centroids c_i and c_j of clusters i and j , respectively

- 2) Silhouette

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2)$$

Where

- $a(i)$: the average distance from point i to all other points within the same cluster
- $b(i)$: the average distance from point i to all points in the nearest neighboring cluster
- $s(i)$: values closer to 1 indicate better cluster assignment (well-clustered)

- 3) Connectivity [11]

$$Conn(C) = \sum_{i=1}^N \sum_{j=1}^L X_{i, nni(j)}$$

Where

- $Conn(C)$: Connectivity index
- $nni(j)$: the j -th nearest neighbor observation of data point i and L

- N : total number of observations
- L : number of clusters
- c. Mapping provinces to the final clusters, by linking the K-Means clustering results to the SOM unit classification assigned to each province.

3. Results

3.1 Visualization of Disaster Composition per Neuron (SOM Codes Plot)

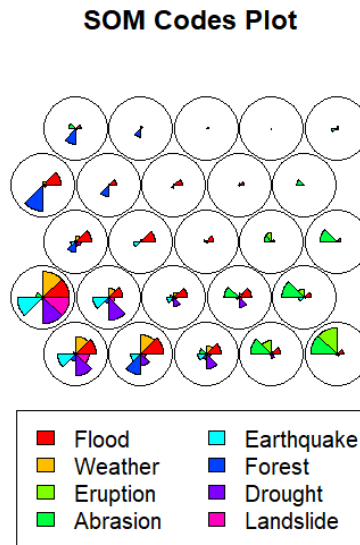


Figure 1. SOM Code Plot Disaster Composition per Neuron

Figure 1 presents the SOM Codes Plot, which represents the composition of eight major disaster types floods, extreme weather, eruptions, coastal abrasion, earthquakes, forest/land fires, droughts, and landslides across each neuron in the 5×5 grid. Each circle represents a single neuron, with slices in the form of a pie chart indicating the proportion of disaster variables within that unit.

For example, neurons located in the lower section (rows 4 and 5) are dominated by red (floods), purple (landslides), and light blue (forest/land fires), indicating a high concentration of hydrometeorological disaster combinations in the regions mapped to those neurons. Meanwhile, neurons in the upper right section tend to be dominated by light green (eruptions) and cyan (earthquakes), representing provinces with more prominent geological risks.

3.2 Clustering of Provinces Using SOM and K-Means

To simplify interpretation, K-Means clustering was applied to the trained SOM results. The 25 SOM neurons were reduced into 2 main clusters using the K-Means algorithm, as shown in Table 1. These results were then used to group provinces into the final clusters based on their positions in the SOM. An evaluation of the K-Means clustering results was conducted for cluster counts ranging from 2 to 6. The table below presents the values for each metric.

Table 2. Determination of the Optimal Number of Clusters

Number of Clusters	DBI	Silhouette	Connectivity
2	0.56875	0.36667	7.664
3	0.59653	0.29306	10.835
4	0.64444	0.13681	22.770
5	0.65	0.19722	25.772
6	0.61042	0.18889	28.081

Based on Table 2, the optimal clustering result for this study's data can be determined by considering three evaluation metrics: the Davies Bouldin Index (DBI), Silhouette, and Connectivity. The lowest DBI value is 0.56875 for $k = 2$. The closer the Silhouette value is to 1, the better, as it indicates more clearly separated clusters. The highest Silhouette value is also observed for $k = 2$, at 0.36667. Furthermore, the smaller the Connectivity value, the better, as it indicates that members within a cluster are appropriately close to each other and have minimal connections with other clusters. The lowest Connectivity value is found at $k = 2$, which is 7.664. These three metrics consistently indicate that $k = 2$ is the most optimal number of clusters for the data in this study..

Table 3. Cluster Distribution of Each Province

Cluster	Cluster Members
Cluster 1	Aceh, Sumatera Utara, Sumatera Selatan, Sulawesi Selatan, Riau, Jambi, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, Daerah Istimewa Yogyakarta, Banten, Bali, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Nusa Tenggara Timur, Nusa Tenggara Barat, Papua, Papua Barat, Papua Selatan, Papua Tengah, Papua Pegunungan, Papua Barat Daya Sumatera Barat,
Cluster 2	Jawa Barat, Jawa Tengah, Jawa Timur

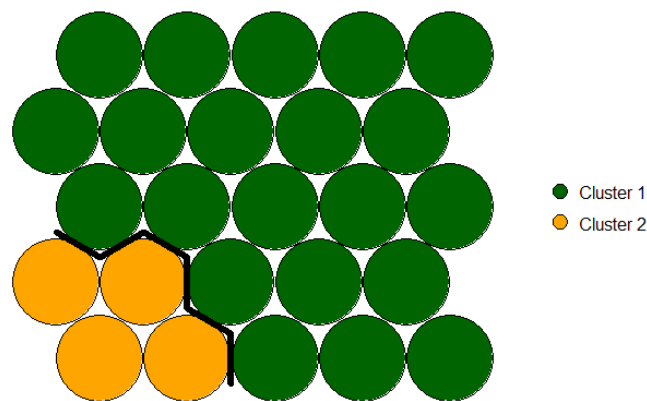


Figure 2. Cluster Result Plot

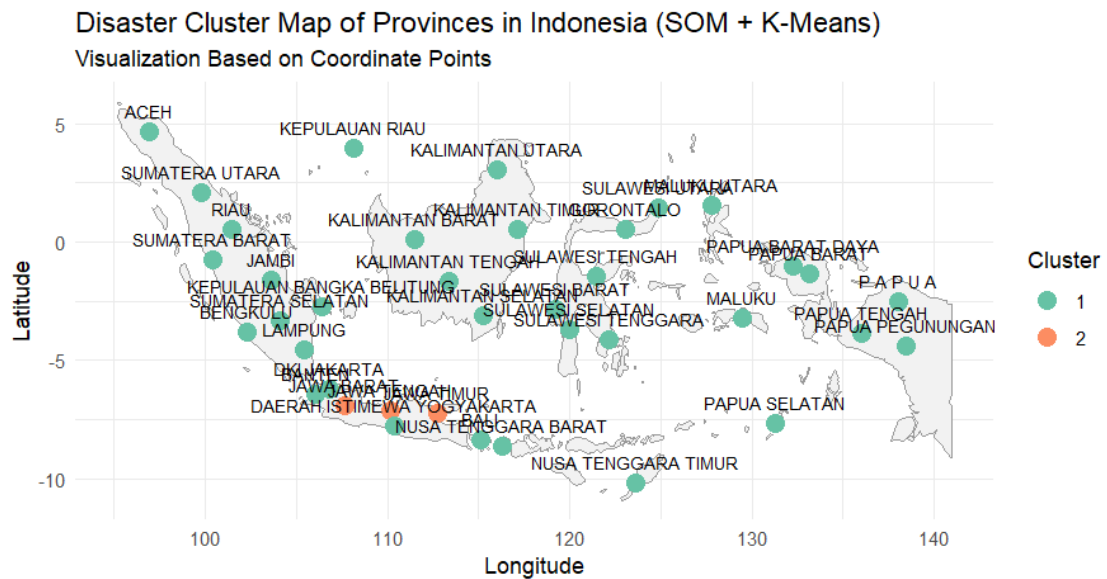


Figure 3. Map of Disaster Clustering Results for Indonesian Provinces Using the Hybrid Method (SOM–K-Means)

Figures 2 and 3 show the disaster cluster map of Indonesian provinces generated using the hybrid Self-Organizing Maps (SOM) and K-Means method, which divides the regions into two main clusters based on disaster occurrence characteristics. Cluster 1 covers most provinces in Indonesia, ranging from Sumatra, Kalimantan, most of Sulawesi, Maluku, to Papua. This cluster tends to represent provinces with relatively similar and widely distributed disaster occurrence levels. Meanwhile, Cluster 2 consists of several provinces in the southern part of Indonesia that have disaster characteristics distinct from those of the other cluster.

3.3 Characteristics of Provincial Clustering Results

a. Cluster 1 : Moderate–Low Disaster Risk with Varied Patterns

This cluster includes regions with generally moderate to low disaster risk levels but relatively high diversity in disaster types. In general, disasters in this cluster do not occur on a massive scale every year but may arise sporadically during certain periods.

- 1) Floods: The number of incidents varies between 0–120 cases/year. Most areas only experience seasonal flooding due to heavy rainfall, while provinces with higher case counts are typically located in coastal or lowland areas with poor drainage systems.
- 2) Extreme Weather: Occurs sporadically, usually with fewer than 50 incidents/year. These events typically include strong winds, heavy rain, or localized storms, with an uneven distribution of occurrences.
- 3) Volcanic Eruptions: Very rare. Only provinces with active volcanoes, such as West Sumatra, East Nusa Tenggara, and North Maluku, report incidents.
- 4) Coastal Abrasion: Occurs in certain coastal areas such as East Nusa Tenggara, Maluku, and Papua, mainly due to a combination of high ocean waves and the reduction of coastal protective vegetation.
- 5) Earthquakes: Low frequency and sporadic, generally occurring in Aceh, Bengkulu, and Sulawesi.
- 6) Forest and Land Fires: High in several provinces such as South Sumatra, Riau, and Kalimantan, but low in areas that are not major plantation centers or peat swamp forests.
- 7) Drought & Landslides: Generally low, although some areas in Sulawesi and Kalimantan experience increased occurrences during prolonged dry seasons or extreme rainfall.

General Pattern: Although the diversity of disaster types is relatively high, the average intensity for each disaster type is relatively low compared to Cluster 2. The risks that arise are usually influenced by

seasonal climatic factors and local environmental conditions rather than events occurring continuously throughout the year.

b. Cluster 2 : High Disaster Risk Dominated by Hydrometeorological Hazards

The regions in this cluster face high disaster risk, particularly in the hydrometeorological category disasters triggered by weather and climate phenomena. The annual frequency of events in this cluster is far above the national average, making it an area of very high vulnerability.

- 1) Floods: Very high, ranging from 78–140 incidents/year. Generally caused by prolonged heavy rainfall, land-use changes, and limited drainage capacity in urban areas.
- 2) Extreme Weather: Very high, reaching 71–193 incidents/year. Includes storms, tornadoes, extreme heavy rain, and high waves affecting fishing activities.
- 3) Earthquakes: Moderate frequency (1–3 times/year), but often with the potential to cause infrastructure damage in vulnerable areas such as the western coast of Sumatra and parts of Maluku.
- 4) Forest and Land Fires: High in East Java (145 incidents), moderate in West Java and Central Java. Mostly triggered by prolonged drought and land-clearing activities.
- 5) Drought: Occurs more frequently than in Cluster 1, ranging from 11–25 incidents/year, affecting agricultural production and clean water supply.
- 6) Landslides: Very high (14–77 incidents/year), often occurring in mountainous areas with steep slopes and unstable soil conditions.

General Pattern: This cluster is characterized by the dominance of hydrometeorological disasters with extremely high frequency, often exacerbated by high population density and vulnerable infrastructure. The socio-economic impacts in these areas tend to be greater due to the large number of assets and populations affected by each event. West Java, Central Java, and East Java fall into the high hydrometeorological disaster risk cluster, with the frequency of disasters such as floods, extreme weather, landslides, and droughts exceeding the national average. Frequent flooding in these regions is caused by high rainfall, land-use conversion, and limited drainage capacity, especially in urban areas [12] [13]. In addition, landslides with high frequency often occur in mountainous areas with steep slopes and fragile soil conditions [14]. The phenomenon of global climate change has also increased the intensity of these hydrometeorological disasters, thereby necessitating enhanced community-based mitigation and adaptation efforts, as well as improved spatial planning management to reduce the widespread social and economic impacts [15].

4. Conclusions

This study employs a hybrid approach that combines Self-Organizing Maps (SOM) and K-Means clustering to group the provinces of Indonesia based on the frequency and types of natural disasters. Internal validation results, including the Davies Bouldin Index, Silhouette, and Connectivity, indicate that forming two clusters is the most optimal solution. Nevertheless, the division into two main clusters already effectively reflects the spatial variation of national disaster risk.

Cluster 1 covers most regions of Indonesia including Sumatra, Kalimantan, Sulawesi, Maluku, Papua, and parts of Java which have moderate to low multi-hazard risk. This cluster is characterized by a diversity of disaster types, without the dominance of any particular type, making it important for cross-hazard mitigation planning. Cluster 2, consisting of West Java, Central Java, and East Java, exhibits high hydrometeorological disaster risk. This grouping is marked by far more intense occurrences of floods, extreme weather, landslides, and forest land fires compared to other provinces, indicating the need for priority mitigation interventions in these high-risk areas. Overall, the SOM and K-Means combination not only enables dimensionality reduction and topological visualization of the clustering results but also facilitates spatial interpretation through the mapping of provincial coordinate points. This approach

demonstrates strong potential as a data-driven disaster risk analysis tool that can support targeted mitigation strategies and resource planning at both national and regional levels.

References

- [1] S. Risky, D. Permana, and D. Fitria, "Self Organizing Maps Method for Grouping Provinces in Indonesia Based on the Landslide Impact," *UNP J. Stat. Data Sci.*, vol. 1, no. 1988, pp. 180–187, 2023.
- [2] W. Bank, "Climate risk country profile: Indonesia," *The World Bank Group*, 2021. <https://climateknowledgeportal.worldbank.org/country/indonesia/vulnerability>
- [3] D. Miljković, "Brief Review of Self-Organizing Maps," *Proc. MIPRO*, no. May, 2017, doi: 10.23919/MIPRO.2017.7973581.
- [4] I. Dermawan, A. Salma, Y. Kurniawati, and T. O. Mukhti, "Implementation of the Self Organizing Maps (SOM) Method for Grouping Provinces in Indonesia Based on the Earthquake Disaster Impact," *UNP J. Stat. Data Sci.*, vol. 1, no. 4, pp. 337–343, 2023, doi: <https://doi.org/10.24036/ujsds/vol1-iss4/83>.
- [5] L. R. Iyohu, I. Djakaria, and L. O. Nashar, "Perbandingan Metode K-Means Clustering dengan Self-Organizing Maps (SOM) untuk Pengelompokan Provinsi di Indonesia Berdasarkan Data Potensi Desa," *J. Stat. dan Apl.*, vol. 7, no. 2, pp. 195–206, 2023, doi: <https://doi.org/10.21009/JSA.07208>.
- [6] M. N. Bangun, O. Darnius, and Sutarman, "Optimization Model in Clustering The Hazard Zone After an Earthquake Disaster," *Sink. J. dan Penelit. Tek. Inform.*, vol. 6, no. 3, pp. 2089–2095, 2022, doi: <https://doi.org/10.33395/sinkron.v7i3.11598>.
- [7] I. H. Rifa, H. Pratiwi, N. Sciences, and U. S. Maret, "Clustering of earthquake risk in indonesia using k-medoids and k-means algorithms," *Media Stat.*, vol. 13, no. 2, pp. 194–205, 2020, doi: 10.14710/medstat.13.1.194-205.
- [8] T. Kohonen, *Self-Organizing Maps (3rd ed.)*. Berlin, Heidelberg: Springer, 2001. doi: <https://doi.org/10.1007/978-3-642-56927-2>.
- [9] N. P. N. Hendayanti, G. A. M. A. Putri, and M. Nurhidayati, "Ketepatan Klasifikasi Penerima Beasiswa STMIK STIKOM Bali dengan Hybrid Self Organizing Maps dan Algoritma K-Mean," *J. Varian*, vol. 2, no. 1, 2019, doi: 10.30812/varian.v2i1.316.
- [10] M. Kossakov, A. Mukasheva, G. Balbayev, S. Seidazimov, and D. Mukammejanova, "Quantitative Comparison of Machine Learning Clustering Methods for Tuberculosis Data Analysis †," *Eng. Proc. MDPI*, vol. 60, no. 20, 2024.
- [11] G. N. Brock, S. Datta, and S. Datta, "cIValid : An R Package for Cluster Validation," *J. Stat. Softw.*, vol. 25, no. 4, 2008, doi: 10.18637/jss.v025.i04.
- [12] BPBD Jawa Timur, "Dokumen Kajian Risiko Bencana Nasional Provinsi Jawa Timur 2022–2026," 2022. [Online]. Available: https://files.bpbd.jatimprov.go.id/DOKUMEN/DOKUMEN_KRB_NASIONAL_JAWA_TIMUR_2022-2026.pdf
- [13] B. J. Barat, "Dokumen Kajian Risiko Bencana Nasional Provinsi Jawa Barat 2022–2026," 2022. [Online]. Available: https://inarisk.bnppb.go.id/pdf/Jawa_Barat/Dokumen_KRB_Prov._Jawa_Barat_final_draft.pdf
- [14] I. G. Tejakusuma, E. H. Sittadewi, and R. Fitriani, "Hydrometeorological hazard detection and warning for risk reduction in West Java , Indonesia Hydrometeorological hazard detection and warning for risk reduction in West Java , Indonesia," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1192, no. 012043, 2023, doi: 10.1088/1755-1315/1192/1/012043.
- [15] A. Secretariat, "Changing disaster risk landscape due to climate change in ASEAN," Jakarta, 2025.



© **The Author(s) 2025**. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License. Editorial of Journal of Mathematics: Theory and Applications, Department of Mathematics, Universitas Sulawesi Barat, Jalan Prof. Dr. Baharuddin Lopa, S.H., Talumung, Majene 91412, Sulawesi Barat.