Identification of Social, Economic and Building Density Vulnerability To Earthquake and Tsunami Hazards In Bantul District

Yohana Noradika Maharani[1]*, Yody Rizkianto[2], Ikhsan Ikhsan[3]
*Corresponding author email: yohananm@upnyk.ac.id
[1]Disaster Management, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia
[2]Geology Engineering, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia
[3]Indonesian Agency for Meteorological Climatological and Geophysics, Yogyakarta, Indonesia

Abstract. Bantul Regency is one of the regions in Indonesia with a high risk of earthquake and tsunami disasters due to the presence of an active fault (Opak fault) and proximity to the plate subduction zone. Furthermore, as Yogyakarta's most popular tourist destination, population density and new development areas are increasing. The increase in density and development has an effect on the level of vulnerability (vulnerability), which is increasing, particularly in the absence of risk management, which is the main cause of increased casualties, infrastructure damage, and poverty during disasters. This study aims to identify social, economic, and building density vulnerability in the study area by clustering the level of vulnerability and identifying the dominant variables in the study area using the Self Organizing Maps (SOM) method. As a result, the research area has been divided into three clusters, with Bantul, Imogiri, Banguntapan, and Sewon sub-districts being the most vulnerable. Research on vulnerability in a given area can serve as a starting point for stakeholders and policymakers to use in supporting contingency planning.

Keywords : earthquake and tsunami hazards, vulnerabilities, clusters, som, bantul regency

1 Introduction

Bantul Regency is among the regions in Indonesia with a high risk of earthquakes and tsunamis, because to the subduction zone, where the Indo-Australian Plate is subducting beneath the Eurasian Plate, and active land faults like the Opak Fault. There is a lot of seismic activity in the southern Javan subduction zone, which can result in tsunamis. The Bantul coastal region has an earthquake potential of up to 8.7 SR, according to the 2017 Megathrust National Earthquake Map segmentation update [1].

Building more infrastructure is undoubtedly essential to boosting the economy. But such progress ought to consider the possibility of catastrophe. The aging of assets and greater growth in the absence of risk management are the primary factors contributing to increased damage during disasters [2][3]. The cost of constructing a huge bridge is five to eight times the damage and losses resulting from earthquakes in Indonesia, indicating that the impact of this natural disaster undoubtedly slows down growth [4].

B. Sobirov et al. (eds.), *Proceedings of the 2nd International Conference on Advanced Research in Social and Economic Science (ICARSE 2023)*, Advances in Social Science, Education and Humanities Research 842,
https://doi.org/10.2991/978-2-38476-247-7_61

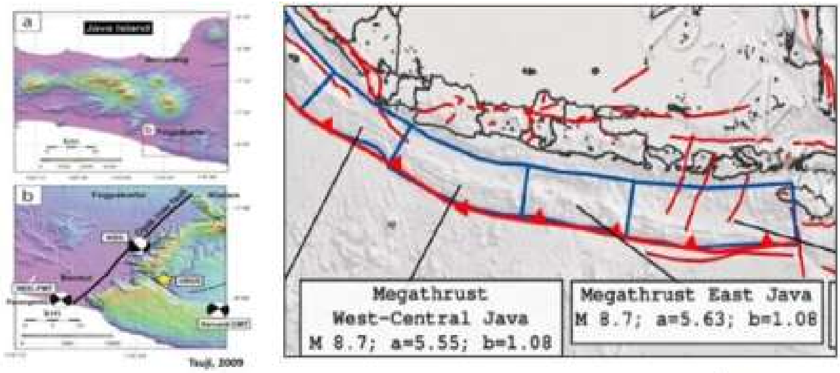

Fig.1. Potential for a Large Earthquake and Tsunami South of Java [1]

When disasters strike a region, they can result in poverty, infrastructure destruction, and fatalities. Poor households typically exhibit lower levels of resilience and experience more challenges in coping with and recovering from the effects of disasters. Vulnerability, though, is not limited to poverty. A large number of people struggle to manage everyday life as a result of disparities in income, skill level, etc. Social vulnerabilities may not always result in disasters; rather, social inequality and its underlying causes lead to unequal social conditions generally. These disparities also have an impact on people's ability to react to and recover from dangerous situations. Figure 2 show of poverty and other multifaceted causes and drivers which are linked to specific groups such as women, children, the elderly, people with disabilities, migrants, etc. are frequently, but not always, what make individuals vulnerable to the effects of catastrophes. Over time, these weaknesses may increase people's vulnerability to the effects of hazards, leading to higher losses in terms of lives and property. Consequently, social vulnerability studies must unquestionably be included in order to lessen the impacts that at-risk groups will experience in the near future and to determine what steps will strengthen their resistance to the effects of hazards.

Fig.2.Conceptual model of how vulnerability can lead to differences in response [5]

As part of disaster mitigation, research on the vulnerability to disasters resulting from social, economic, tourism, and development density activities must be carried out. Disaster mitigation is a strategy that can be used to lessen the effects of natural disasters, such as the number of casualties and longer-term, higher economic losses. This strategy can also help prevent the destruction of infrastructure, facilities, and other buildings that could cause trauma and harm to other natural resources, which could upset the economy's wheels. Complete mitigation of a disaster is necessary, engaging all relevant parties including the public, private, and community sectors. This calls for thorough planning and preparation. The assessment of a region's vulnerability to impending disasters is a crucial component of disaster mitigation activities that should be examined as a preventive step to boost awareness and readiness.

The objective of this study is to pinpoint the study area's vulnerabilities related to social, economic, and building density. Utilizing statistical data, the Self Organizing Maps (SOM) technique was applied to group the study area's degree of vulnerability. Studies on an area's vulnerability can be utilized as a starting point for practical information that helps stakeholders and policy makers assist catastrophe mitigation and emergency preparedness.

2 Data and Methods

Bantul Regency, which includes 17 sub-districts (Srandakan, Sanden, Kretek, Pundong, Bambanglipuro, Pandak, Bantul, Imogiri, Jetis, Dlingo, Plereti, Piyungan, Banguntapan, Sewon, Kasihan, Pajangan, and Sedayu) is the location of the case study for this research.

Table 1. Vulnerability variable used in the research

Label	Vulnerability Variable	Vulnerability Concept	Vulnerability Rationalization
FEMALE	Total female population	Gender	Correlated with lack of resources
MALE	Total male population	Gender	Correlated with lack of resources
TODDLER	Number of baby and toddler	Age	Needs more assistance during disaster phases
ELDRY	Number of seniors (over 60 years old)	Age	Requires more assistance during disasters
DISBLD	Persons with disabilities (blind, deaf, physically disabled, sick)	Disability	Need additional support and assistance in coping with the impact of hazards
HOUSE	Number of residential buildings	infrastructure	Correlated with building vulnerability level
EDUCTN	Number of permanent buildings educational facilities	infrastructure	Correlated with building vulnerability level

HEALTHFAC	Number of permanent buildings health facilities	infrastructure	Correlated with building vulnerability level
WRSSHP	Number of permanent buildings infrastructure worship facilities	Infrastructure	Correlated with building vulnerability level
HLTHWORK	Number of health workers	Health services	Required during emergency and recovery phases
HOTEL	Number of hotels	Economy	Correlates with economic level and recovery
TOURISM	Number of tourisms	Economy	Correlated with economic level and recovery
RESTO	Number of restaurants	Number of restaurant businesses	Correlated with economic level and recovery

2.1. Data Acquisition

The statistical data utilized in the research of social vulnerability instances came from the "Angka 2022" publication on the Bantul Regency website, BMKG and additional statistics and information on earthquake disasters in the study area. These resources are excellent for getting detailed data on the population, building count, and economic conditions. because representative data for multiple variables in a more constrained area the district was needed for the study. The collected data was combined and adjusted according to Bantul's subdistrict population.

2.2. Methods Self Organizing Maps (SOM)

Neural network models' unsupervised methods are used in this technique. The process analysis was carried out using the SOM tool that was integrated into the MATLAB computer environment. SOM was a useful tool for identifying case study locations that were comparable to one another and for identifying critical characteristics that characterize each cluster's social vulnerability. SOM and correlation coefficients were utilized in this work for the reduction variables (selection variables). Redundancy (correlated variables) and noise (variance), two undesired aspects of data, are reduced through the application of dimensionality reduction techniques.

Reduction in this study aims to get rid of redundant or strongly associated variables. The chosen variables are used to designate regions in the following procedure. SOM may locate clusters using participatory clustering and hierarchical agglomerative clustering using k-means and similarity matrices. Within these cluster groupings, a dendrogram, or cluster tree, can be made. By dividing the dendrogram to a specific level, the visualization of the dendrogram can be used to comprehend the data structure and calculate the number of clusters.

Self-Organizing Maps (SOM), generated using an unsupervised artificial neural network modeling approach, were used in this study's data processing [7]. Process analysis was conducted using MATLAB's SOM toolset. [7] state that the SOM algorithm can be explained as follows.

$$X = [x_1, x_2, \ldots, x_M]^T \quad (1)$$

There are M neurons in the input layer. The output neurons, which are typically arranged in a 2D grid, are found in the output layer. The vector of weights for each neuron whose dimension matches that of the input pattern represents the weights from the input layer neurons to the output layer neurons. denoted by

$$W_j = [w_{1j}, w_{2j}, \ldots, w_{Mj}]^T, \quad j = 1, 2, \ldots, N \quad (2)$$

Initialization of the analysis process with all weights is set to a lower random integer. For every neuron, SOM determines the similarity distance between the weight vector and the input vector X. The Euclidean distance, which is frequently employed as a similarity metric, is the Euclidean distance between the weight vector and the input vector X.

$$d_j = \|x - w_j\| = \sqrt{\sum_{i=1}^{M} (x_i - w_{ij})^2} \quad (3)$$

A two-dimensional map with n inputs and n connection weights (w(i)=(w0(i),...,wn(i)) that are all connected topologically makes up a SOM. The topological borders of the input space determine how the parser arranges this map. As a result, two close-ups at the entry will activate two units close to the SOM, creating a description between the input space and the network space. With the use of the SOM, the ideal room arrangement is discovered from the input data. The locations of the weight vectors and the direct neighbor relationships between cells can be visually depicted if the input space is smaller than three. As a result, a close look at the map yields excellent insights regarding its design and architectural decisions. As a result of updating their prototype vectors, winning neurons become more responsive to that kind of input when it is presented again. This makes it possible to train various cells with various kinds of data. Neighbors of the winning neuron can also match their prototype vectors to the input vectors to accomplish topological mapping, albeit the degree of matching will vary based on the distance from the winner. The distribution of samples on the SOM topological map is the direct result of the SOM method, as seen in Figure 4. From a conceptual standpoint, the dark lines represent neurons that are trained on multidimensional data, which causes them to cluster. Currently, each unit cell's cluster and SOM learning method is a clustering method that sequentially organizes samples into cells. The so-called U-Matrix [8] can be used to illustrate the density or cluster structure of the data. SOM offers a map view with a lattice-organized arrangement of the map units. The SOM is enhanced by the U-matrix. In the SOM view, this could make the data density—that is, the clusters—appear gray. By labeling places on the map that are comparable and have comparable variable/component values as neighbors, SOM is used to classify samples.

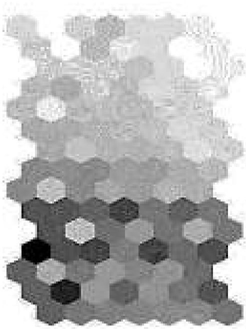

Fig.3. The U-Matrix shows the SOM dataset in shades of gray, showing the topological coordinates of the hexagons themselves and the difference between the two hexagons.

2.3. Map Size

The map is first formed during the training process, just like in other artificial neural networks. In this work, a two-dimensional hexagonal SOM structure with hexagonal neighbors was used for all of the simulations. Heuristics can be used to determine the default number of SOM rows and columns, where n is the number of data samples. We assess the SOM algorithm's training performance using two quality index evaluation techniques: quantization error *(Qe)* and topography error *(Te)*. The Best Matching Unit (BMU) and each data vector are separated by an average distance.

The algorithm performs better when the *Qe* value is low. That is, it indicates how well the SOM preserves the learned material's topology [9]. A lower *Te* value is preferable. Although increasing the size of the data set guarantees lower errors (better quantification), this is not always desirable because redundant or insignificant neurons grow at the same time [10].

2.4. Sample Size Selection

The goal of sample size reduction (dimensionality) is to obtain a simplified representation of the original data while retaining as much information as possible. When nm, reduction is defined as a decrease in the number of features that represent data objects from m original data to n items of reduced data. The results show that reducing the sample size has an effect on the resolution of the sensitivity map as well as the cluster accuracy. A two-way solution is proposed in this study to select an effective sample size. The first option is to represent the correlation coefficient and component level visually. The second point is the significance of the SOM pie chart.

2.5. Clustering Analysis

Similar units should be grouped using clustering methods to facilitate quantitative analysis of maps and data [11]. The goal of clustering is to find similar subsets (groups) of objects in a dataset. However, there are two main data clustering methods for achieving unique clustering and meaningful data partitioning, namely hierarchical agglomerative clustering and positive clustering using k-means. A clustering tree (dendrogram) can be constructed using agglomerative clustering. By slicing the dendrogram to a specific level, visualization of the dendrogram can be used to interpret the data structure and determine the number of groups. *K*-means seeks spherical groups such as those labeled

$$E = \sum_{k=1}^{c} \sum_{x \in Q_k} \|x - c_k\|^2 \tag{4}$$

Where *C* is the number and the group center is *k*. To minimize the best clustering, the Davies-Bouldin clustering index is used for both in-group distance and inter-group distance, so that the best clustering is minimized by:

$$\frac{1}{C} \sum_{k=1}^{c} \max_{l \neq k} \left\{ \frac{S_c(Q_k) + S_c(Q_l)}{d_{ce}(Q_k, Q_l)} \right\} \tag{5}$$

The Davies-Bouldin index can be used to evaluate k-means partitioning because a low value indicates good clustering results.

3 Results and Discussion

3.1. NJOP Saptosari District and Panggang District

All simulations in SOM are performed to determine the Quantification error *(Qe)* and topographic error *(Te)* in a two-dimensional hexagonal map with a structure in a hexagonal neighborhood during the training process to form an artificial neural network. Quantification error *(Qe)* and topographic error *(Te)* are two quantities used to estimate the quality index used to evaluate the SOM algorithm's performance during the training process. A larger data set guarantees smaller errors (better quantization), but not necessarily the best or desired results, because the presence of redundant or insignificant neurons causes a parallel increase (Lin & Chen, 2005). As shown in Figure 2, a series of experiments were carried out to determine Quantification error *(Qe)* and topographic error *(Te)*. Figure 3 shows the results of finding a reasonably good representation of the results using the map size [8 3] with *Qe* 0.354 and *Te* 0.000 and variable visualization.

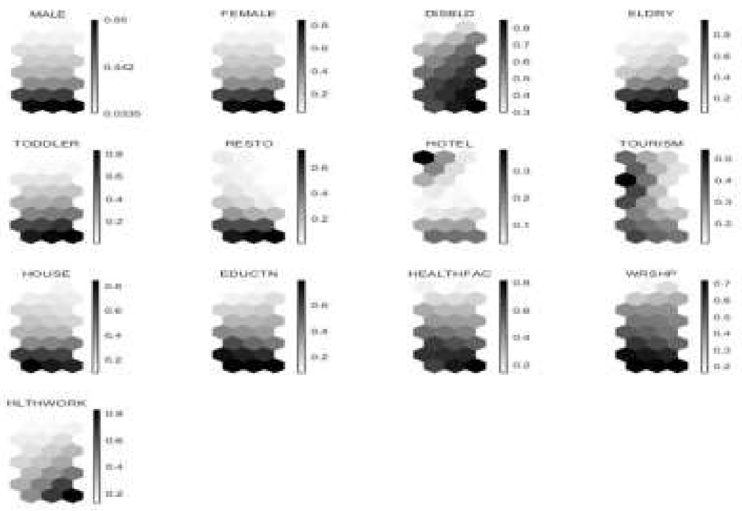

Fig. 4. Visualization of the components of each social vulnerability variable to determine their contribution in building the Hexagonal map

The thirteen variables chosen from the SOM relative importance pie chart were processed in the same way as in the previous analysis. 'Hits' were added to the U-Matrix to show the actual distribution of decisions on the map, as shown in Figure 5a. As shown in Figure 5c, it is possible to detect spatial differences in the distribution of the eta dataset using hits, and the variables that influence the differences are clearly described. An investigation into the influence of individual variables was conducted based on a border selection criterion to determine which variables predominantly fund social vulnerability, as shown in Figure 5b (solid red box), which shows pie charts for thirteen variables in map units. The size and color of the various graphs within this map unit describe the relative importance of each variable. These are then sorted in order of importance (Figure 5c). The number of health workers (WRSHP) 10.11% is the most important variable influencing the increase in social, economic, and building vulnerability. Health workers are required at times in the effort to overcome resilience, particularly during the post-disaster recovery period. The number of educational facilities (EDUCTN) is 9.69%, and the number of health facilities (HEALTHFAC) is 9.05%. Disability (DISBLD) ranks fourth with 8,80%. Finally, the number of seniors (EDLRY) is 8.23%.

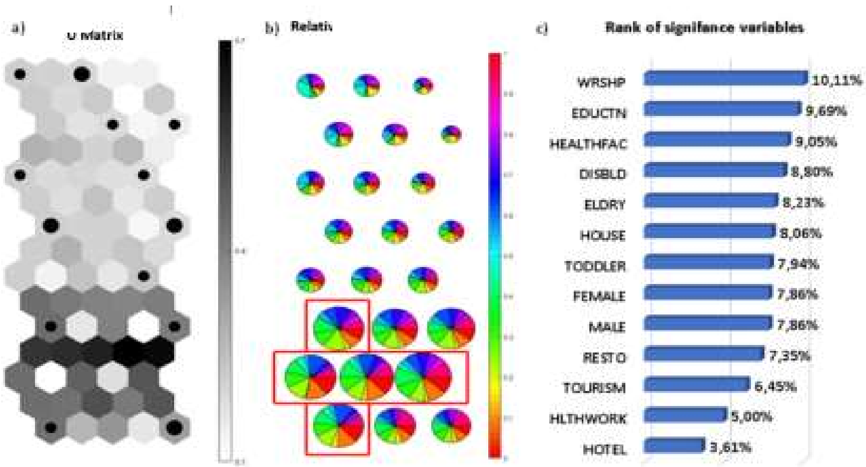

Fig.5. Spatial variation of social vulnerability variables: a) hit histogram showing data density indicated by the size of the circles in the U-matrix; b) relative importance of variables indicated by each pie chart of different size and color; and c) order of importance of variables from high to low.

3.2. Clustering and Dominant Variables

SOM was used to classify the samples so that similar sites with similar/component variable values are arranged as neighbors on the map, as illustrated in Figure 6, which employs a hexagonal map structure with distributed labels (subdistrict names). According to the findings of the SOM analysis, the social, economic, and building density vulnerabilities in Bantul regency are classified into three groups. The highest vulnerability is in Cluster 3, which includes the Bantul, Imogiri, Banguntapan, and Sewon sub-regions. The high number of elderly and female residents contribute to this area's high level of social vulnerability. The number of permanent buildings such as worship facilities, health facilities, and educational facilities is a construction aspect.

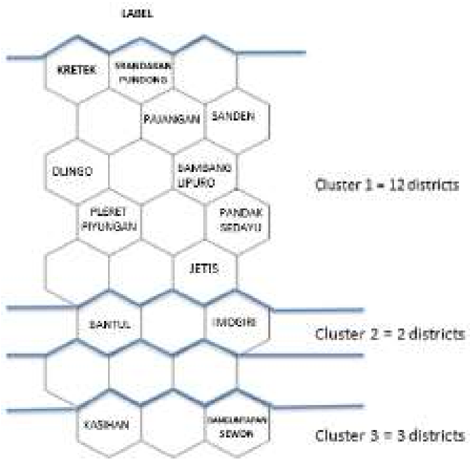

Fig. 6. Classification of 17 sub-districts with 13 selected variables through SOM relative importance-classification of sample sub-districts based on similarity with close variable/component values based on the clusters obtained, the ranking of significant variables in each cluster will be determined, indicating the most dominant variables that cause vulnerability, as shown in Figure 7.

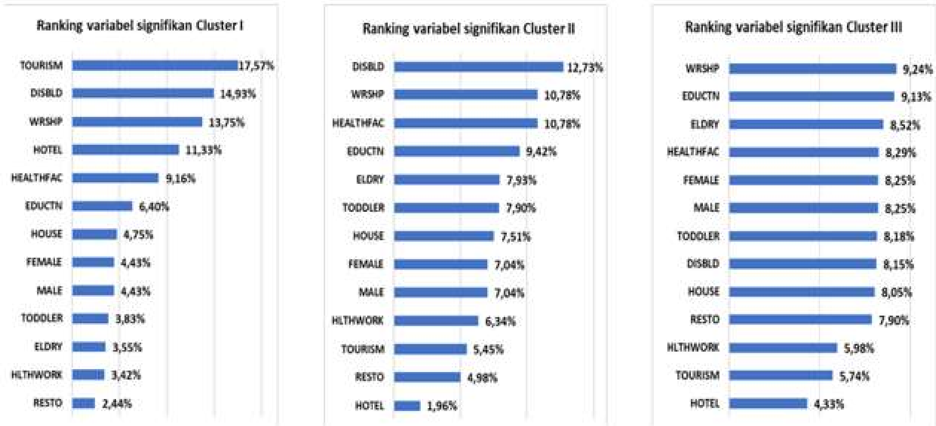

Fig. 7. Ranking of significant variables in each cluster

4 Conclusions

The SOM is able to explicate the effects of removing redundant variables on social vulnerability and clearly separates the data set based on the similarity and dissimilarity of each variable; it also offers unique clustering and logical partitioning of the data. Three clusters of social, economic, and building density vulnerabilities have been identified in Bantul regency based on the findings of the SOM analysis. Policymakers and other decision-makers can find this to be a helpful approach when planning and prioritizing tasks based on time and resource constraints related to emergencies and hazards. In relation to natural hazards, it can help to ensure that development plans and budgets account for the sensitivity of the populations and their reduced ability to respond to hazards can help identify which areas may require special attention during the immediate response and long-term recovery following a hazard event, as well as more comprehensive framework and the identification of communities that are most in need of social services.

Acknowledgments

This research is supported by a grant from the Institute for Research and Community Service (LPPM), Universitas Pembangunan Nasional Veteran Yogyakarta through the "UPN Veteran Yogyakarta Basic Research Grant Year 2023".

References

[1] Pusgen, "Peta Sumber dan Bahaya Gempa Indonesia Tahun 2017," Badan Penelitian dan Pengembangan Kementrian Pekerjaan Umum dan Perumahan Rakyat, ISBN978- 602-5489-01-3, 2017.

[2] R. Davidson, "AN Urban Earthquake Disaster Risk Index," Department of Civil and Environmental Engineering, Stanford University, 1997.

[3] World Bank, "The Sendai Report: Managing Disaster Risks for a Resilient Future," 2012.

[4] S. P. Nugroho, "Pusat Data dan Informasi BNPB. Dampak Bencana Terhadap Ekonomi Indonesia," http://www.majalahglobalreview.com/ opini/8-opini/25-dampak-bencana-terhadap-ekonomi-indonesia.html. Di unduh tanggal 5 September 2013.

[5] Y. N. Maharani, A. R. B. Nugroho, D. F. Adiba, dan I. Sulistiyowati, "Pengaruh Kerentanan Sosial Terhadap Ketangguhan Masyarakat dalam Menghadapi Bencana Erupsi Gunung Merapi di Kabupaten Sleman," *Jurnal Dialog Penanggulangan Bencana*, vol. 11, no. 1, hal. 307-313, 2020.

[6] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, no. 1, hal. 59-69, 1982.

[7] G. F. Lin dan L. H. Chen, "Identification of homogenous regions for regional frequency analysis using the self-organizing map," *Journal of Hydrology*, vol. 324, hal. 1-9, 2005.

[8] A. Ultsch, "Self-organizing neural networks for visualization and classification," 1993.

[9] K. Kiviluoto, "Topology preservation in Self-Organizing Map," dalam *International Conference on Neural Networks*, hal. 294-299, 1996.

[10] S. J. Ki, J. H. Kang, S. W. Lee, Y. S. Lee, K. H. Cho, K. G. An, dan J. H. Kim, "Advancing assessment and design of stormwater monitoring programs using a self-organizing map: Characterization of trace metal concentration profiles in stormwater runoff," *Water Research*, vol. 45, hal. 4183-4197, 2011.

[11] J. Vesanto dan E. Alhoniemi, "Clustering of the Self-Organizing Map," *IEEE Transaction on Neural Networks*, vol. 11, no. 3, hal. 586-600, 2000.

[12] BMKG Earthquake Catalog: http://repogempa.bmkg.go.id/repo_new/, 2021.