Clustering of the 2018 Lombok Earthquakes using an Agglomerative Hierarchical Clustering Algorithm

Cipta Ramadhani
Dept. of Electrical Engineering
University of Mataram
Mataram, Indonesia
cipta.ramadhani@unram.ac.id

Abdullah Zainuddin
Dept. of Electrical Engineering
University of Mataram
Mataram, Indonesia
abdullahzain@unram.ac.id

Abdul Natsir
Dept. of Electrical Engineering
University of Mataram
Mataram, Indonesia
natsir.amin@unram.ac.id

Rosmaliati
Dept. of Electrical Engineering
University of Mataram
Mataram, Indonesia
rosmaliati@unram.ac.id

Teti Zubaidah
Dept. of Electrical Engineering
University of Mataram
Mataram, Indonesia
tetizubaidah@unram.ac.id

***Abstract*—Lombok Island is located between the Indo-Australian subduction zone and the Flores back-arc thrust, makes it vulnerable to earthquake. During July to August 2018, significant earthquakes occurred in Lombok Island region and led to severe damages in the northern area. In this study, we propose hierarchical clustering of historical earthquakes that occurred in Lombok region during July 2018 to December 2018 and model the clustering of earthquakes. We used spatial and non-spatial attributes in three different conditions. The results show that the most homogenous seismic sources were achieved by using spatial attributes.**

Keywords—earthquakes, clustering, agglomerative, spatial attributes

I. Introduction

Lombok is an island, located in West Nusa Tenggara province, Indonesia. It is a part of the Lesser Sunda Island region, lies between Bali island in the West and Sumbawa island in the East. In 2018, Indonesian Meteorology, Climatology and Geophysical Agency (Badan Meteorologi Klimatologi dan Geofisika/BMKG) reported that several significant earthquakes shook Lombok [1]. This earthquake sequence began with an Mw 6.4 earthquake on 28 July 2018, followed by earthquakes on 5 August (Mw 6.8), 9 August (Mw 5.8) and 19 August (Mw 6.9). The detailed parameters of earthquakes are listed in Table I, and locations of earthquakes are depicted in Fig. 1. The earthquakes that occurred from 5 to 19 August 2018 in the West-, East-, and North Lombok reached the maximum intensity of VIII-IX and were strongly felt by all residents in the whole of Lombok [2].

The Seismicity of Lombok island mainly controlled by two major geological structures, i.e. the Flores back-arc thrust in the North and the Indo-Australian subduction zone to the South [3], [4]. A unique feature of geomagnetic anomaly pattern in Lombok consists of contiguous negative–positive anomalies are closely related to this subduction. A strong dipolar magnetic anomaly was found in the southern region of Lombok could be associated with a large magnetic body or a discontinuity in the geological structure (e.g. potentially local fault) [5]. Another stronger one was found in the northern region, which was formerly identified as formations of young lavas of Mount Rinjani, but later on it could also be identified related with a new mature subduction along the Flores thrust [6].

Earthquakes and effort to reduce the hazard have attracted the attention of many researchers. Mitigation plans of seismic hazard and risk play an important role particularly for decision makers to prepare an earthquake mitigation in an optimal way [7], [8]. One of step in any seismic hazard analysis is the ability of modeling the earthquake source [9], [10]. A potential single earthquake that occurred in one place is supposed to be uniform. The chance of one earthquake to be occurred with certain magnitude is the same through the source, which may be linear or areal [10].

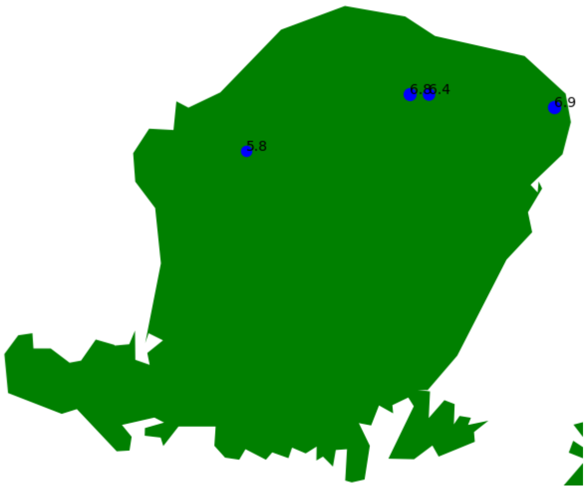

Fig. 1. The relatives Location of The 2018 Lombok Earthquakes.

TABLE I. PARAMETER FOR THE MAINSHOCKS OF LOMBOK 2018 EARTHQUAKES

Date	Time (UTC)	Latitude (°N)	Longitude (°E)	Magnitude (Mw)	Depth (km)
July 28th	22:47:38.491	8.35	116.50	6.4	13
August 5th	11:46:37.363	8.35	116.47	6.8	32
August 9th	05:25:32.601	8.44	116.21	5.8	14
August 19th	14:56:27.086	8.37	116.70	6.9	18

Modeling of earthquake sources is a fundamental in creating hazard maps and estimating the probability of upcoming earthquake that will be occurred with different magnitude in the future [11]. Using database collection of historical earthquakes has made it possible to compose these sources more efficient. However, seismologists construct the boundary of seismic sources manually based on tectonic features and historical earthquakes without using standard method [12]-[14]. Furthermore, since the size of historical earthquake grows, the manual depiction of source boundaries becomes less accurate and complicated.

Several studies have been conducted to cluster earthquakes sources. A study for earthquake preparedness in the Istanbul city has used K-means clustering method to create the training dataset for earthquakes vulnerability analysis [15]. Another study in the Sea of Marmara region has used high-resolution seismicity catalog and the nearest-neighbor earthquake cluster approach for identification of clusters seismicity [16]. Using this method, they could identified whether the events of earthquake are foreshocks, mainshocks and aftershocks. More recent study has been conducted in the area of Corinth Gulf, Greece to identify evolution of mainshock–aftershock sequences and swarms, along with periods of seismic quiescence using Markovian Arrival Process (MAP) [17]. The study could identify selected seismic sequences and the hidden states and found their close relationships with mainshock–aftershock and swarm-like sequences.

Agglomerative Hierarchical Clustering (AHC) algorithm has been proposed to overcome clustering problem. Authors [18] proposed AHC algorithm to cluster Linear Ordinal Ranking (LOR) information and used it to illustrate online financial product recommendation. Distance measure and the aggregation method have been proposed under the framework of AHC. Meanwhile authors [19] used AHC as a data mining procedure to provide a simplified optimization of economic, environmental and energy (3E) in assessing of building retrofit on a macro-scale. Based on set performance target, an innovative framework for a facile and holistic assessment can be offered to investors with a broader range of retrofit alternatives.

AHC has been applied also in the power transmission network investment [20]. Demand patterns are extracted from hourly demand data based on the Elbow's rule and a linkage criterion. Three different categories (i.e. seasonal, monthly, and weekly) are used to test the representative demand curves, thereby a 24 hours demand pattern can be provided.

II. DATA AND METHOD

The work of this paper is focused to generate clusters of earthquakes based on seismic sources from historical earthquakes. In order to achieve this task, we used an agglomerative clustering algorithm to generate clustering of the 2018 earthquakes in Lombok Island. We interpreted that AHC algorithm can cluster the earthquake without predetermined the number of clustering as an input parameter.

Spatial (latitude and longitude) and non-spatial attributes (depth, magnitude, and occurrence date) are used in this work. Table II shows the description of each attribute used in the dataset.

TABLE II. AVAILABLE ATTRIBUTES FOR EARTHQUAKES.

Variable	Description
Latitude	Decimal degrees Latitude
Longitude	Decimal degrees Longitude
Depth	Depth of the earthquake in kilometers.
Magnitude	The magnitude for the event
DOY	Day of years

A. Data

Earthquake data with magnitude from M_W 1.7 to M_W 6.9 in Lombok Island from July to December 2018 were collected from the official BMKG website for earthquake repository (http://repogempa.bmkg.go.id/). The Data set contains 1400 earthquakes.

The occurrence day of earthquake is used based on Day of Years which means that we count the day from 1 to 365 days in one year.

B. Method

Hierarchical clustering algorithm can be either divisive or agglomerative. In this case, the AHC algorithm does not require us to predetermine the number of clusters. Bottom-up algorithms initially consider each piece of data as a singleton cluster before subsequently combining pairs of clusters into a single cluster that contains all the data.

Hierarchical clustering creates a (usually binary) tree through the data. The leaves are individual data objects, but the root is a single cluster that contains all of the data. Individual pieces of data make up the leaves, but the root is a single cluster that contains all of the data. Intermediate clusters that contain parts of the data are placed between the root and the leaves. The main goal of hierarchical clustering is to build upward-moving "clusters of clusters." To make a tree, there are primarily two conceptual methods. AHC merges groups together from the bottom up, starting with each datum in its own singleton cluster. Divisive clustering begins with all of the data in a single large group and then separates it into singletons for each piece of data. In this paper, we consider to use AHC as our method.

Fig. 2 represents the AHC technique. The clusters with just one element are labeled as leaves. The internal nodes depict the division of parent node (division), or the union of their two children (agglomeration). The root is a single cluster that contains the entire element in the collection. The AHC in agglomerative clustering is defined by a hierarchy structure for the data, which are usually deterministic. The advantage of hierarchical clustering is that the number of clusters is not required as an input variable. In comparison with partitional algorithm, it provides more detail information than using the unorganized set.

In order to cluster the earthquake sources, we used three different approaches:

- Clustering based on spatial variables.

  In this approach, we clustered the earthquakes based on their location (latitude and longitude). Because the distance of both variables is compatible and compensate each other, the variables do not need to be normalized for this clustering.

- Clustering based on non-spatial variables.

  In non-spatial variables, the earthquakes are grouped based on occurrence date, magnitude and depth. We hypothesize that the earthquakes which a similar depth, magnitude and occurrence date are likely to be related to the same cluster. Because the range of occurrence date is larger than the range of both depth and magnitude so that all non-spatial data is normalized to ensure that data is similar across all records.

- Clustering based on all variables.

  For the last clustering approach, we used all variable (latitude, longitude, depth, magnitude, and occurrence date) to see their dependency to cluster the earthquakes.

In order to create a cluster among the earthquake sources, first we measured the distance between the earthquakes using the Euclidean technique. This technique is used to calculate the distance between two objects. Euclidean distance is determined between both the center of the source object and the centers of all surrounding objects. The Euclidean distance can be computed using (1).

$$D(x,y) = \sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\cdots+(x_n-y_n)^2} \quad (1)$$

where:

D(x,y)	:	The distance between the earthquake x and y
x_n	:	number variable from the earthquake x
y_n	:	number variable from the earthquake y

Next step, we used agglomerative clustering with ward's method with pandas and Sklearn library in Python programming. The goal of Ward's Method is to reduce variance within a single cluster of objects. Errors Sum of square (ESS) between the two clusters is used to calculate the distance between the two clusters created using Ward's approach. Only when there are multiple element objects in the cluster can Ward's be determined [21]. The Ward's method of agglomerative clustering algorithm can be measured using (2).

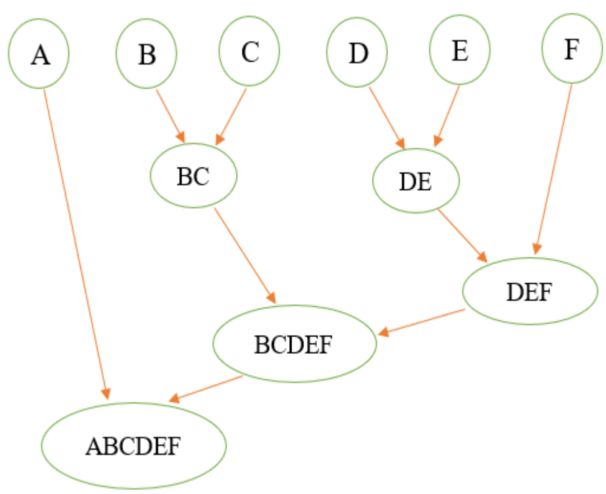

Fig. 2. Agglomerative Hierarchical Clustering Technique.

$$ESS = \sum_{j=1}^{P}\left(\sum_{i=1}^{n} X_{ij}^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_{ij}\right)^2\right) \quad (2)$$

where:

X_{ij}	:	value for the object to-i in the cluster to-j
P	:	number of parameter calculated
N	:	the number of objects made into the the cluster

The goal of Ward's Method is to select the subsequent clustering steps with the least amount of ESS increase possible. The sum of the squares in the two clusters for each variable is the distance between the two clusters. This strategy uses a variance analysis approach to determine the distance between clusters [22].

III. Result and Discussion

In this section, we present the results of earthquakes clustering that occurred in Lombok Island from July to December 2018 using agglomerative Hierarchy algorithm. As mentioned in the previous section, the result will be presented in three different approaches. First, we cluster the earthquakes based on spatial variables. The second clustered is modeled based on non-spatial variables. The last model, we used all variable to see their dependences in clustering model.

Fig. 3 represents the size of cluster based on spatial attribute. Meanwhile, Fig. 4 and Fig. 5 represent the size of cluster based on non-spatial attribute and all variable in our dataset, respectively. Based on observation, earthquakes that occurred in Lombok Island are distributed uniformly in a few clusters when spatial attribute are used as variables. Although, in terms of depth, magnitude and occurrence date, the earthquakes are mostly in two classes and one small class.

By referring to Fig. 3, it was found that the earthquakes are clustered around three homogenous seismic sources ranging from 448 to 497 numbers of earthquakes. This result is in-line with three significant earthquakes that shook the Lombok island in 2018. On the other hand, both Fig. 4 and Fig. 5 are opposite. In Fig. 4, the sizes of each clustering value are 162, 639, and 616, respectively. Furthermore in Fig. 5, the clustering pattern is similar; with size of the three clustering models are of 166, 676, and 575, respectively. Based on these results, we interpret that clustering might work only on spatial attributes rather than non-spatial attributes.

Fig. 6 shows the number of clustering that are colored based on spatial attribute (longitude and latitude). In Fig. 7, the number of clustering is depicted based on non-spatial attribute (depth, magnitude, and occurrence date). Meanwhile, Fig. 8 depicts the number of clusters based on all attributes that were already mentioned in the previous section. All of these figures depict earthquakes that occurred at least 1400 times on Lombok island from June to December 2018.

When earthquakes are clustered according to spatial attributes (see Fig. 6), the geographical distribution of clusters seems uniforms and their geographic distribution appears uniform and follow the pattern of mainshocks (see Table I). In other words, each cluster is located near by the three significant earthquakes that shook Lombok Island in 2018. This cluster gave similar result which was conducted by [23]. On the other hand, when earthquakes are clustered according to non-spatial attributes (see Fig. 7), their geographic location grouping appears random and does not correspond to the locations of main earthquakes. It means that the earthquakes of one cluster potentially occur in different parts of the region. In the last approach, when both non-spatial and spatial attributes of earthquakes are considered as clustering process (see Fig. 8), these additional attributes does not affect the geographical distribution of clusters. In other words, non-spatial attributes seem not be appropriate for use as attributes for the clustering process.

Fig. 3. The Size of cluster based on spatial variables.

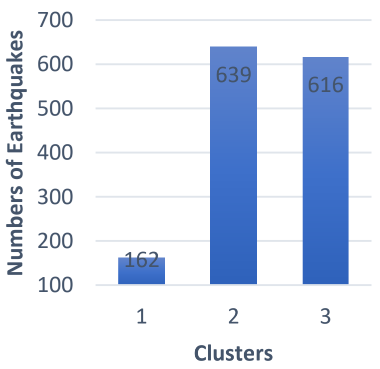

Fig. 4. The Size of cluster based on non-spatial variables.

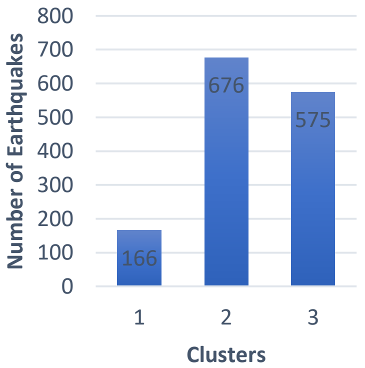

Fig. 5. The Size of cluster based on all variables.

IV. Conclusion

This research clustered earthquakes based on spatial and non-spatial attributes that occurred in Lombok Island from July to December 2018. The clustering pattern can be seen when the earthquakes are clustered based on their location. Three clustering patterns of earthquakes represented three significant earthquakes that shook Lombok Island in 2018. However, the result of clustering pattern using non-spatial attributes is opposite. The distribution of clustering in certain locations seems random and does not follow the pattern of the significant earthquakes. This observation implies that clustering earthquakes is feasible only for spatial attributes rather than non-spatial attributes. However, this hypothesis needs further investigation and is a subject for future research.

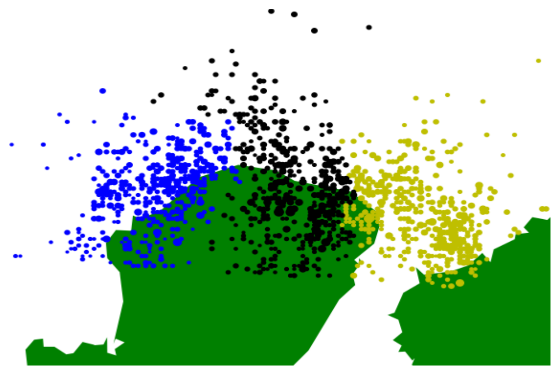

Fig. 6. Earthquakes clustering based on spatial attributes

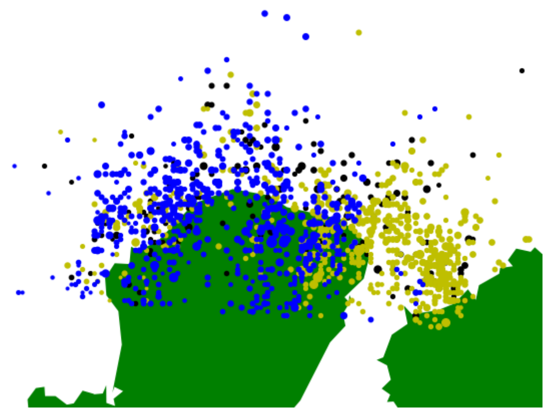

Fig. 7. Earthquakes clustering based on non-spatial attributes

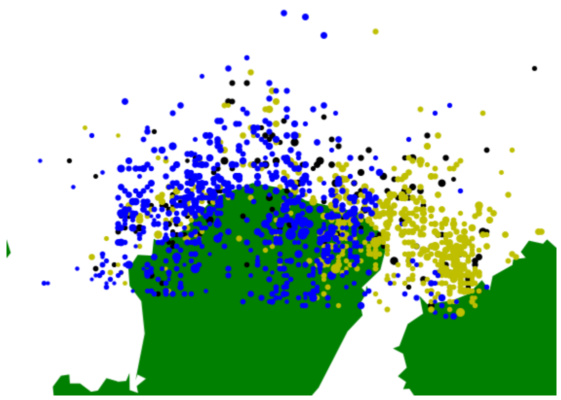

Fig. 8. Earthquakes clustering based on spatial and non-spatial attributes

ACKNOWLEDGMENT

This research is funded by Universitas Mataram under Hibah Dana PNBP 2022 program entitled “Pengelompokan (Clustering) Kejadian Gempa Pulau Lombok Zona Backarc Thrust untuk Mitigasi Bencana”. The data that has been presented in this paper is collected from Indonesian Meteorology, Climatology, and Geophysical Agency (BMKG). Our appreciation is presented to BMKG as their effort to promote high standard earthquakes repository (http://repogempa.bmkg.go.id/).

REFERENCES

[1] U. Sutiyono et al., Katalog Gempa Bumi Signifikan dan Merusak Tahun 1821–2018. Jakarta: Pusat Gempa Bumi dan Tsunami Kedeputian Bidang Geofisika Badan Meterologi Klimatologi dan Geofisika, 2019.

[2] R. Robiana, A. Minarno, S. Hidayati, S. Supartoyo, and F. Nurfalah, *Gempa Lombok Tanggal 29 Juli 2018 dan Dampaknya*. Kajian Rangkaian Gempa Lombok Provinsi Nusa Tenggara Barat, Pusat Penelitian dan Pengembangan Perumahan dan Pemukiman, Badan Penelitian dan Pengembangan, Kementerian Pekerjaan Umum dan Perumahan Rakyat, 2018.

[3] W. Hamilton, “Subduction in the Indonesion region,” 1977, pp. 15–31. doi: 10.1029/ME001p0015.

[4] W. B. Hamilton, “Tectonics of the Indonesian region,” Washington, D.C, 1979.

[5] T. Zubaidah, M. Korte, M. Mandea, Y. Quesnel, B. Kanata, “Geomagnetic field anomalies over the Lombok Island region: an attempt to understand the local tectonic changes,” Int. J Earth Sci., Vol. 99, pp. 1123-1132, July 2010, doi: 10.1007/s00531-009-0450-4.

[6] T. Zubaidah, M. Korte, M. Mandea, M. Hamoudi, “New insights into regional tectonics of the Sunda‑Banda Arcs region from integrated magnetic and gravity modelling,” J. Asian Earth Sci., Vol. 80, pp. 172‑184, 2014. http://dx.doi.org/10.1016/j.jseaes.2013.11.013.

[7] C. H. Scholz, “Large Earthquake Triggering, Clustering, and the Synchronization of Faults,” Bulletin of the Seismological Society of America, vol. 100, no. 3, pp. 901–909, Jun. 2010, doi: 10.1785/0120090309.

[8] J. G. Anderson and K. Nanjo, “Distribution of Earthquake Cluster Sizes in the Western United States and in Japan,” Bulletin of the Seismological Society of America, vol. 103, no. 1, pp. 412–423, Feb. 2013, doi: 10.1785/0120100212.

[9] C. A. Cornell, “Engineering seismic risk analysis,” Bulletin of the Seismological Society of America, vol. 58, no. 5, pp. 1583–1606, Oct. 1968, doi: 10.1785/BSSA0580051583.

[10] Reiter. L, Earthquake hazard analysis : issues and insights. New York: Columbia University Press, 1990.

[11] T. Anagnos and A. S. Kiremidjian, “A review of earthquake occurrence models for seismic hazard analysis,” Probabilistic Engineering Mechanics, vol. 3, no. 1, pp. 3–11, Mar. 1988, doi: 10.1016/0266-8920(88)90002-1.

[12] M. Hashemi, A. A. Alesheikh, and M. R. Zolfaghari, “A spatio-temporal model for probabilistic seismic hazard zonation of Tehran,” Comput Geosci, vol. 58, pp. 8–18, Aug. 2013, doi: 10.1016/j.cageo.2013.04.005.

[13] M. Erdik, Y. A. Biro, T. Onur, K. Sesetyan, and G. Birgoren, “Assessment of earthquake hazard in Turkey and neighboring,” Annals of Geophysics, vol. 42, no. 6, Dec. 1999, doi: 10.4401/ag-3773.

[14] M. Erdik, M. Demircioglu, K. Sesetyan, E. Durukal, and B. Siyahi, “Earthquake hazard in Marmara Region, Turkey,” Soil Dynamics and Earthquake Engineering, vol. 24, no. 8, pp. 605–631, Sep. 2004, doi: 10.1016/j.soildyn.2004.04.003.

[15] M. Shafapourtehrany, P. Yariyan, H. Özener, B. Pradhan, and F. Shabani, “Evaluating the application of K-mean clustering in Earthquake vulnerability mapping of Istanbul, Turkey,” International Journal of Disaster Risk Reduction, vol. 79, p. 103154, Sep. 2022, doi: 10.1016/j.ijdrr.2022.103154.

[16] P. Martínez-Garzón, Y. Ben-Zion, I. Zaliapin, and M. Bohnhoff, “Seismic clustering in the Sea of Marmara: Implications for monitoring earthquake processes,” Tectonophysics, vol. 768, p. 228176, Oct. 2019, doi: 10.1016/j.tecto.2019.228176.

[17] P. Bountzis, E. Papadimitriou, and G. Tsaklidis, “Earthquake clusters identification through a Markovian Arrival Process (MAP): Application in Corinth Gulf (Greece),” Physica A: Statistical Mechanics and its Applications, vol. 545, p. 123655, May 2020, doi: 10.1016/j.physa.2019.123655.

[18] N. Liu, Z. Xu, X.-J. Zeng, and P. Ren, “An agglomerative hierarchical clustering algorithm for linear ordinal rankings,” Inf Sci (N Y), vol. 557, pp. 170–193, May 2021, doi: 10.1016/j.ins.2020.12.056.

[19] Y. Hong, C. I. Ezeh, H. Zhao, W. Deng, S.-H. Hong, and Y. Tang, “A target-driven decision-making multi-layered approach for optimal building retrofits via agglomerative hierarchical clustering: A case study in China,” Build Environ, vol. 197, p. 107849, Jun. 2021, doi: 10.1016/j.buildenv.2021.107849.

[20] N. González-Cabrera, J. Ortiz-Bejar, A. Zamora-Mendez, and M. R. Arrieta Paternina, “On the Improvement of representative demand curves via a hierarchical agglomerative clustering for power transmission network investment,” Energy, vol. 222, p. 119989, May 2021, doi: 10.1016/j.energy.2021.119989.

[21] A. N. Fathia, R. Rahmawati, and T. Tarno, “Analisis Klaster Kecamatan Di Kabupaten Semarang Berdasarkan Potensi Desa Menggunakan Metode Ward Dan Single Linkage,” Jurnal Gaussian, vol. 5, no. 4, pp. 801–810, Oct. 2016.

[22] C. Suhaeni, A. Kurnia, and R. Ristiyanti, “Perbandingan Hasil Pengelompokan menggunakan Analisis Cluster Berhirarki, K-Means Cluster, dan Cluster Ensemble (Studi Kasus Data Indikator Pelayanan Kesehatan Ibu Hamil),” JURNAL MEDIA INFOTAMA, vol. 14, no. 1, Feb. 2018, doi: 10.37676/jmi.v14i1.469.

[23] A. T. Sasmi et al., “Hypocenter and Magnitude Analysis of Aftershocks of the 2018 Lombok, Indonesia, Earthquakes Using Local Seismographic Networks,” Seismological Research Letters, vol. 91, no. 4, pp. 2152–2162, Jul. 2020, doi: 10.1785/0220190348.