p–ISSN: 2723 – 6609 e-ISSN: 2745-5254
Vol. 6, No. 1, January 2025 http://jist.publikasiindonesia.id/

Analysis of Earthquake Intensity on Java Island Using K-Means Clustering and GeoMap

Galan Kristianto Anver[1*], Sri Yulianto J. Prasetyo[2]
Universitas Kristen Satya Wacana, Indonesia
Email: 672019316@student.uksw.edu[1*], sri.yulianto@uksw.edu[2]

*Correspondence

Keywords: earthquake, k-means clustering, geo map.

ABSTRACT

Indonesia is the country with the second highest earthquake intensity in the world because Indonesia is passed by the confluence of 3 tectonic plates, namely: the Indo-Australian Plate, the Eurasian Plate, and the Pacific Plate. Earthquakes occur due to the movement of tectonic plates that move freely and interact with each other. The impact of the earthquake includes: Tsunamis, collapsed buildings, fires, rock collapses, and ground cracks. The purpose of writing this study is to determine the intensity of earthquakes that occur on the island of Java using the K-Means Clustering method. The earthquake data used in this study uses magnitude, depth, and geographical location parameters obtained from existing data. The K-Means Clustering method is used to group earthquake data into several clusters based on their seismic characteristics. The result of this study is to identify seismic zones with similar earthquake intensity. The results of the cluster obtained, it was used to make data visualization of areas with high earthquake intensity using GeoMap.

Introduction

Indonesia is the country with the second highest earthquake intensity in the world because Indonesia is passed by the confluence of 3 tectonic plates, namely: the Indo-Australian Plate, the Eurasian Plate, and the Pacific Plate. Earthquakes occur due to the movement of tectonic plates that move freely and interact with each other. The impact of the earthquake includes: Tsunamis, collapsed buildings, fires, rock collapses, and ground cracks (Mardiani & Nugraha, 2023).

With the advancement of information and technology, anyone can access and get information quickly and correctly. Therefore, this knowledge is very useful for making decisions. Currently, information from various topics, including earthquake disasters, and although earthquakes are inevitable, the impact of earthquakes can be minimized by determining earthquake-prone areas using the K-Means Clustering method to map areas with high earthquake intensity (Rinjani et al., 2023).

The K-Means Clustering method is one of the effective methods in data management to group data based on similar data characteristics (Maruhawa & Mahdiana, 2024). In earthquake intensity analysis, this method can be used to group earthquake data based on earthquake intensity in the form of magnitude, depth, and location of the earthquake. This grouping makes it possible to identify earthquakes as well as zones with varying levels of intensity (Heraldi et al., 2019).

The K-Means Clustering method can be used well in grouping natural disasters. Wisnu Priyo Jatmoko in 2024 in a journal entitled "Application of the K-Means Clustering Method to Fire Disasters in Samarinda City" grouped fire data into 3 clusters using the Orange application, namely c1 the number of fire incidents, c2 the number of fire incidents in Samarinda city and c3 the number of fire incidents in sub-districts/complaints in Samarinda city (Jatmiko et al., 2024).

The impact of earthquakes will be divided based on the results of clustering where clusters with low intensity can be felt by many people but do not cause damage, earthquakes with moderate intensity cause impacts on non-structural parts of buildings experiencing minor damage such as cracks in walls, roofs shifting or falling, high-intensity earthquakes causing large cracks in buildings or collapsing, curved railroads, landslides, and tsunami (Hutabarat, 2023).

Thus, this study utilizes the K-Means Clustering method to see the spread of earthquakes on the island of Java as a whole by grouping the intensity of earthquakes based on the magnitude and impact produced and visualizing the results of clustering by mapping earthquake points on the island of Java. This is expected to be a reference for related parties in making decisions to overcome the impact of the earthquake.

Isna Hidayatur Rifa (2020) in a journal entitled "Clustering of Earthquake Risk In Indonesia Using K-Medoids And K-Means Algorithms" uses the K-Medoids and K-Means method in researching earthquake risk in Indonesia. This study organizes information about earthquakes based on how strong and how deep they occur using the clustering method. In his research, the data was divided into 5 clusters, namely, high-risk earthquake points, moderate, moderate, moderate, and low (Rifa et al., 2020).

Alicia Putri Widya Hadi (2023) in the journal "Clustering of Earthquake Data in Indonesia with K-Means and DBSCAN Algorithms" Where two algorithms are compared, namely K-Means and DBSCAN. In his research, K-Means produced 8 clusters, and the DBSCAN algorithm produced 14 clusters. In comparison, the K-Means silhouette value can cluster better than DBSCAN because it has a smaller cluster error (Prasetio & Effendi, 2023).

Muhammad Dwi Chandra (2021) in a journal entitled "Professional Application of K-Means Algorithm in Grouping Toddlers Who Suffer from Malnutrition by Province". In this study, data mining is used using the K-Means algorithm. In the results of the research from data from 34 provinces that had a vulnerable time in 2016-2018, 2 clusters were obtained, namely high and low clusters. The high cluster consists of 15 provinces and the low cluster consists of 19 provinces (Chandra et al., 2021).

Method

This research goes through several stages and flow of methods carried out to manage data and visualize data. First, by identifying problems and how to cluster earthquake data and visualizing the clustering results using GeoMap.

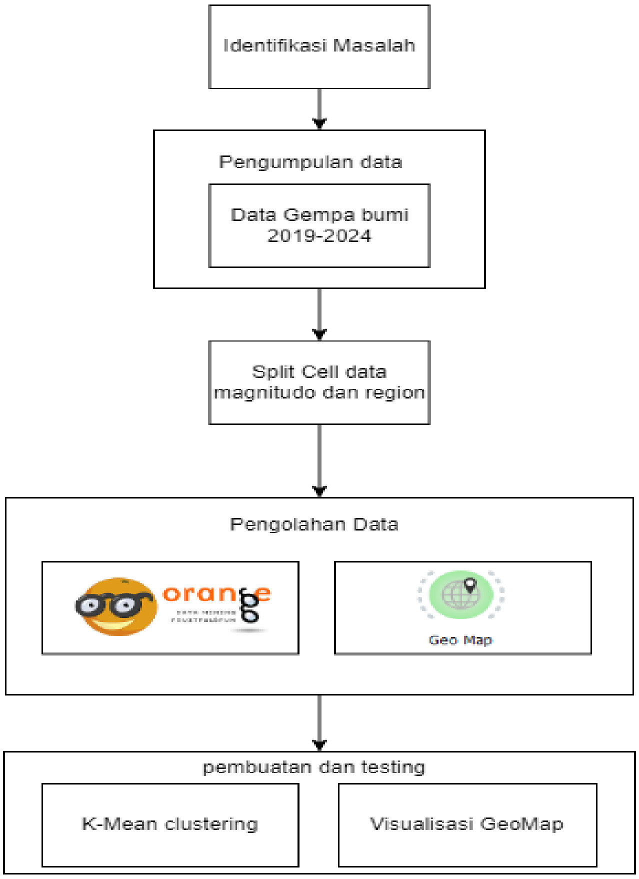

Figure 1 Research Methods

The first stage is to identify the problem and seek understanding, to understand the intensity of the earthquake on the island of Java. In this study, earthquake data will be grouped into 3 groups, namely high, medium, and low-intensity earthquakes. The process of collecting earthquake data was obtained through the EarthScope Consortium website which occurred on the island of Java in the vulnerable time of 2019-2024. The data obtained will then be split into cells so that the data can be processed using the Orange application. The next process is merging magnitude and region data in the orange application, after the merging process is successful, it can be processed by the Clustering method. After the results of Clustering are obtained, the data can be visualized using GeoMap which is an orange application tool. The last stage is the creation and testing with data from the results of K-Means Clustering. The K-Means Clustering process involves initializing where selecting a point as the initial cluster center (centroids), then assigning the cluster at each data point x_i to the cluster with the nearest Euclidean data center:

$$d(x_i,) =)c_j \sqrt{\sum_{m=1}^{n} \quad (x_{im} - \; c_{jm}}^{2}$$

Where is the ith data, is the jth cluster cent $x_i c_j$ d n is the number of features in the data (magnitude, depth, or location) the next stage is to update the cluster center. Calculate the new center for each *cluster* by taking the average of all the data points that have been assigned, which is the amount of data assigned to N_J*cluster* j.

$$c_j = \frac{1}{N_J} \sum_{i=1}^{n} \quad x_i$$

Then the last stage is convergence where the step is repeated 2-3 times until the center of the cluster does not change significantly. The visualization process uses data that has coordinates in the geographical location of Java Island in the form of latitude and longitude. Visualization is carried out by taking data that has been clustered. On the GeoMap, earthquake points with different colors obtained from the clustering results will be displayed.

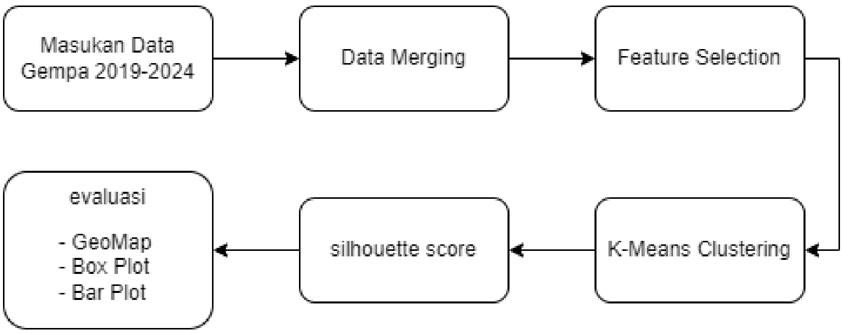

Figure 2 Data flow

In Figure 2, the data processing flow has gone through the split cell stage, then the dataset into the orange application and performs data merging or data merging to later be clustered. Next is to do feature selection to determine the column attributes of the dataset that will be used for the clustering process to get earthquake intensity results. Next is the clustering process using the K-Means algorithm, the data that has been successfully clustered is then evaluated using Silhouette Coefficient. The Silhouette Coefficient value is 0 to 1 where the evaluation is closer to 1, the more efficient the clustering obtained from the clustering results. Silhouette Coefficient value parameters based on Kaufman and Rousseeuw: 0.7 < 1 strong structure, 0.5 < 0.7 medium structure, 0.25 < 0.5 weak structure, <0.25 no structure (Yunistya et al., 2022).

The last process in conducting earthquake intensity analysis is the evaluation of the results using GeoMap displayed in the geographical image based on the data that has been processed to make it easier to understand the research results.

Results and Discussion

The data used in this study is data obtained from the EarthScope Consortium website, namely Java Island earthquake data in the 2019-2024 time frame (Pasaribu et al., 2021) The data obtained is raw data that has not gone through any process, as shown in Figure 3.

Year	Month	Day	Time	Lat	Lon	Depth	Mag	wilayah	Timestam
2024	5	2	06:20.0	-6.8386	110.7528	248.6	4.2	5 km SE of	1.71E+09
2024	5	1	06:11.0	-7.3772	107.4844	10	4.1	20 km SSE	1.71E+09
2024	3	24	34:05.0	-7.4873	108.6034	114.6	4.7	20 km W c	1.71E+09
2024	3	2	48:44.0	-7.0651	106.8453	62.9	4.9	18 km SSV	1.71E+09
2024	2	15	54:01.0	-8.3032	112.234	117.6	4.5	23 km SSE	1.71E+09
2024	2	14	01:11.0	-6.976	109.2674	212.8	4.1	12 km ESE	1.71E+09
2024	1	22	49:31.0	-7.5322	112.149	177.6	4.1	7 km NE o	1.71E+09
2024	1	1	46:46.0	-6.642	107.7736	10	4.5	25 km NE (	1.7E+09
2024	1	1	59:19.0	-6.8599	107.1224	178.1	4.1	4 km SSW	1.7E+09
2023	12	31	34:21.0	-6.7569	107.8204	12.4	4.8	14 km NW	1.7E+09
2023	12	31	35:34.0	-6.8551	108.001	10	4.1	9 km E of S	1.7E+09
2023	12	30	38:09.0	-7.9886	111.6142	154	4.2	14 km NW	1.7E+09
2023	12	15	42:56.0	-7.0736	108.8083	10.2	4	17 km SSV	1.7E+09
2023	12	15	24:33.0	-7.1302	108.8322	9.4	4.2	19 km NNI	1.7E+09
2023	12	13	35:10.0	-6.8597	106.5246	10.4	4.5	14 km NN'	1.7E+09
2023	12	7	00:46.0	-6.9698	106.534	9.9	3.9	2 km NW (	1.7E+09
2023	10	24	51:16.0	-6.7384	108.4228	237.6	4.2	3 km S of I	1.7E+09
2023	10	1	00:24.0	-6.9206	106.7799	112.9	5.3	15 km S of	1.7E+09
2023	5	8	21:08.0	-7.0763	107.2238	117.8	4.4	26 km WN	1.68E+09

Figure 3 Earthquake Data

The next data is data that has gone through the split cell stage. The data display consisting of magnitude data and region data is shown in Table 1.

Table 1
Magnitude Data and Region Data

Stomach	Region
4.2	5 km SE of Welahan, Indonesia
4.1	20 km SSE of Banjar, Indonesia
4.7	20 km W of Sidareja, Indonesia
4.9	18 km SSW of Sukabumi, Indonesia
4.5	23 km SSE of Blitar, Indonesia
4.1	12 km ESE of Tarub, Indonesia
4.1	4.5 miles from NE of Kertosono, Indonesia
4.5	25 km NE of Lembang, Indonesia
4.1	4 km SSW of Cianjur, Indonesia
4.8	14 km NW of North Sumedang, Indonesia
4.1	9 km E of North Sumedang, Indonesia
4.2	14 km NW of Trenggalek, Indonesia
4	17 km SSW of Ketanggungan, Indonesia
4.2	19 km NNE of Majenang, Indonesia
4.5	14 km NNW of Pelabuhanratu, Indonesia
3.9	2 km NW of Pelabuhanratu, Indonesia
4.2	3 km S of Palimanan, Indonesia
5.3	15 km S of Cicurug, Indonesia
4.4	26 km WNW of Banjar, Indonesia

Datasets - the dataset through column selection aims to determine the attributes that are used as parameters of a region. At this stage, all unnecessary data will be deleted, so that the processed data is more optimal. (Döhmen et al., 2016) The clustering process uses the K-Means algorithm that will use the orange application.

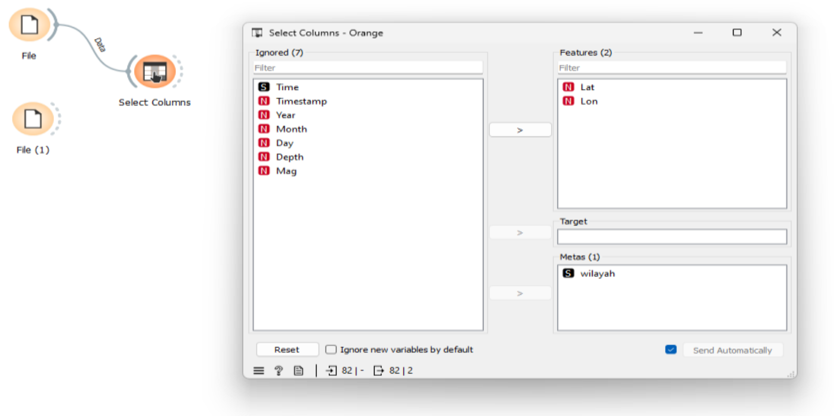

Figure 4 Selecting Attribute Columns

The first step to process the data in the orange application is to enter the dataset that has been obtained. Then the data attribute column will be selected for clustering, where the selected attributes are region, latitude, and longitude. This attribute is used to show earthquake points from the Java Island region.

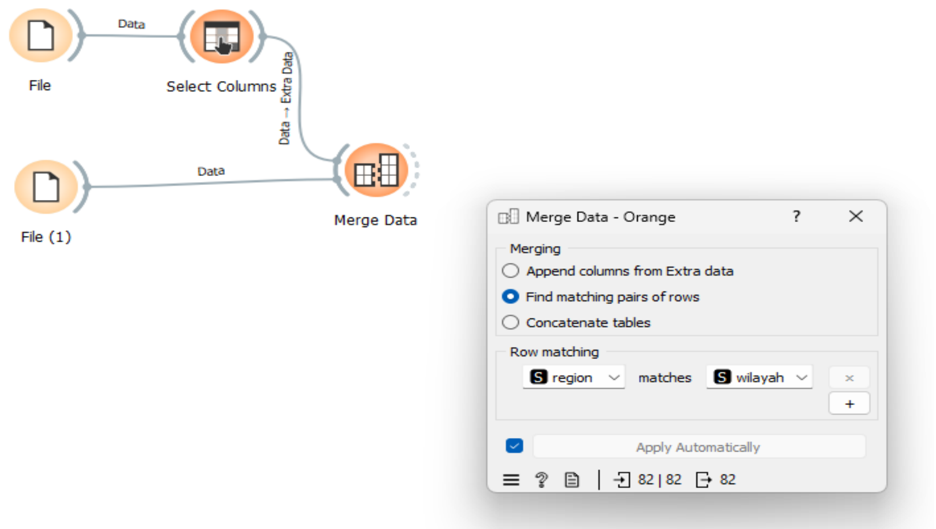

Figure 5 Merge Data

At this stage, the data from the two tables will be combined based on data that has similarities, namely region data and region data. To prevent the values in the clustering process from variously, standardization of data values using Distances from the data that has been obtained is carried out. The clustering results obtained are based on the magnitude and area of the earthquake, then the data will be divided into 3 clusters, namely: cluster 1 with high earthquake intensity, cluster 2 with medium earthquake intensity, and cluster 3 with low earthquake intensity. In the clustering results, it was found that cluster 1 had 31 data, cluster 2 had 19 data, and cluster 3 had 32 data.

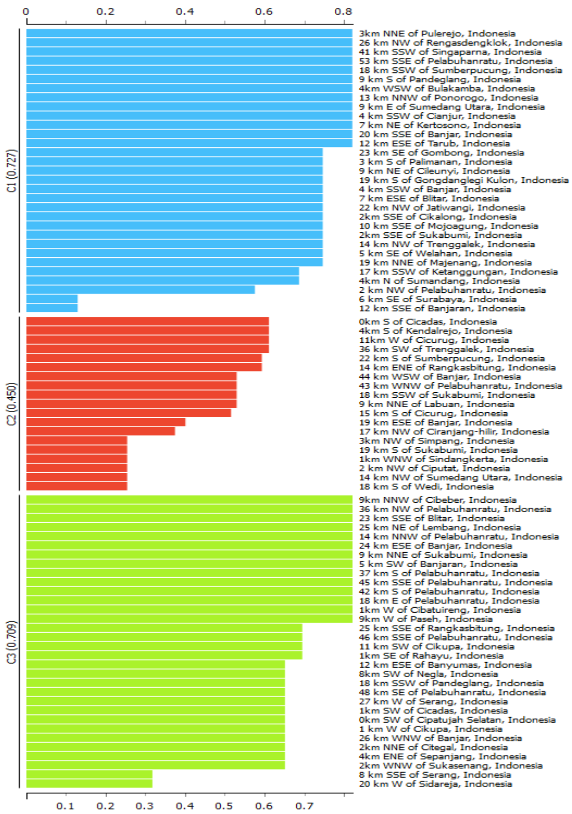

Figure 6 Clustering Results

Figure 6 shows the clustering results and the Silhouette Coefficient results used to see the evaluation of clustering results using the K-Means algorithm. In the evaluation of the Silhouette Coefficient, the Euclidean Distance was used as a calculation in two dimensions used to measure distances at latitude and longitude coordinates. The Silhouette Coefficient values obtained from each cluster are, Cluster 1 0.727 (Strong Structure), Cluster 2 0.450 (Weak Structure), and Cluster 3 0.709 (Strong Structure) (Shahapure & Nicholas, 2020).

Visualization of clustering results using GeoMap to make it easier to understand clustering results. The visualization process uses one of the tools in the orange application based on latitude and longitude data attributes as coordinate points in determining earthquake points and mapping in an area.

Figure 7
GeoMap Visualization Results

In the results of visualization and mapping, the earthquake points along with the intensity level of the earthquake based on their magnitude are displayed. The coordinate points have been differentiated in shape and coloring based on each cluster.

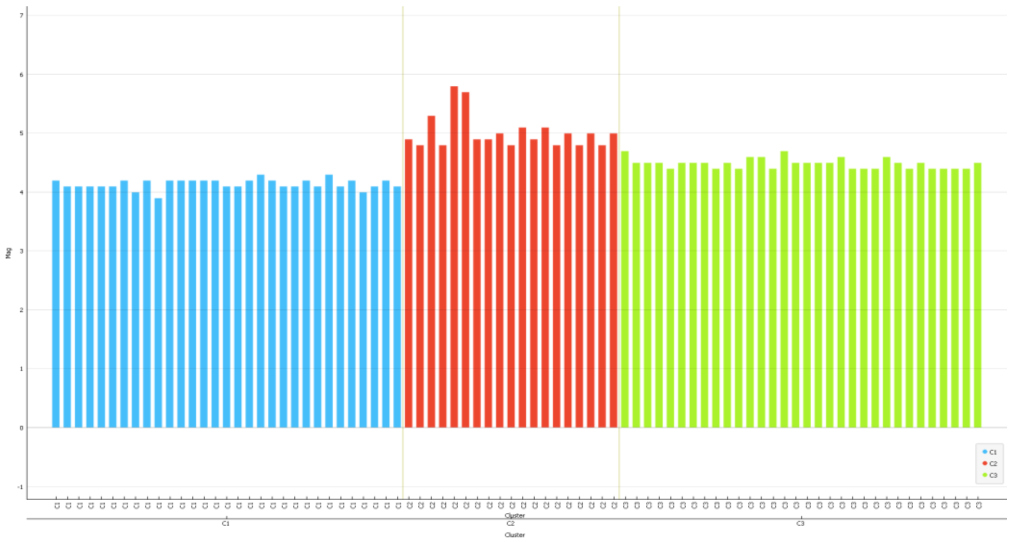

Figure 8 Plot Bar

In Figure 8, the results of data visualization with a Bar Plot can be seen that the intensity of the earthquake in Cluster 2 is higher than that of Cluster 1 and Cluster 3. Visualization of clustering results using Bar Plot using magnitude attribute as a parameter of earthquake intensity level. The intensity level of the earthquake can also be seen using the Boxplot in Figure 9.

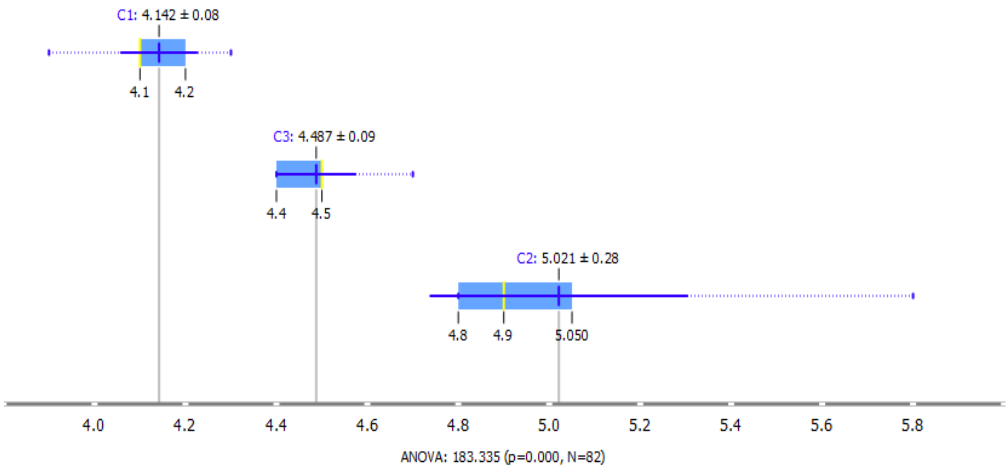

Figure 9 Box Plot

The results of using the Box Plot can show that cluster 1 has a low earthquake intensity with a data range of 4.1 to 4.2, cluster 2 has a moderate earthquake intensity with a data range of 4.4 to 4.5, and cluster 3 has a high earthquake intensity with a data range of 4.8 to 5.05 (Hofmann et al., 2017).

Conclusion

The results of this study using the K-Means clustering method and GeoMap visualization produced 3 clusters where cluster 1 had a low earthquake intensity level with an average data range of 4.1 to 4.2, cluster 3 had a medium earthquake intensity level with an average data range of 4.4 to 4.5, and a cluster 2 had an earthquake intensity level with an average data range of 4.8 to 5.05. The creation of clusters on earthquake intensity aims to group earthquake data based on its intensity so that it can provide structured information for related parties in reducing the impact of earthquakes. The analysis of earthquake intensity is also very important to be a reference in handling the impact of earthquakes, public safety, planning and development of earthquake-resistant infrastructure, reducing economic losses, and increasing public awareness and education about earthquakes.

Bibliography

Chandra, M. D., Irawan, E., Saragih, I. S., Windarto, A. P., & Suhendro, D. (2021). Penerapan Algoritma K-Means dalam Mengelompokkan Balita yang Mengalami Gizi Buruk Menurut Provinsi. *BIOS: Jurnal Teknologi Informasi Dan Rekayasa Komputer*, *2*(1), 30–38.

Döhmen, T., Bruntink, M., Ceolin, D., & Visser, J. (2016). Towards a benchmark for the maintainability evolution of industrial software systems. *2016 Joint Conference of the International Workshop on Software Measurement and the International Conference on Software Process and Product Measurement (IWSM-MENSURA)*, 11–21.

Heraldi, H. Y., Aprilia, N. C., & Pratiwi, H. (2019). Analisis Cluster Intensitas Kebencanaan di Indonesia Menggunakan Metode K-Means. *Indonesian Journal of Applied Statistics*, *2*(2), 137–144.

Hofmann, H., Wickham, H., & Kafadar, K. (2017). value plots: Boxplots for large data. *Journal of Computational and Graphical Statistics*, *26*(3), 469–477.

Hutabarat, L. E. (2023). Tinjauan Geologis Gempa Cianjur November 2022. *Jurnal Rekayasa Teknik Sipil Dan Lingkungan-CENTECH*, *4*(1), 46–53.

Jatmiko, W. P., Ramadhani, M. G., Romadhon, M. G., Adhmadani, G., Fardian, R., & Pranoto, W. J. (2024). Penerapan Metode K-Means Clustering Terhadap Bencana Kebakaran Di Kota Samarinda. *Jupiter: Publikasi Ilmu Keteknikan Industri, Teknik Elektro Dan Informatika*, *2*(1), 1–8.

Mardiani, L. Y., & Nugraha, S. (2023). Kesiapsiagaan Masyarakat Terhadap Megathrust Di Kelurahan Sidoharjo Dan Desa Sedeng Kecamatan Pacitan Kabupaten Pacitan. *Indonesian Journal of Environment and Disaster*, *2*(2), 178–186.

Maruhawa, L. S., & Mahdiana, D. (2024). Segmentasi Wilayah Banjir Daerah Khusus Jakarta Dengan Algoritma K-Medoids. *Prosiding Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, *3*(2), 1837.

Pasaribu, D. F., Damanik, I. S., Irawan, E., & Tambunan, H. S. (2021). Memanfaatkan Algoritma K-Means Dalam Memetakan Potensi Hasil Produksi Kelapa Sawit PTPN IV Marihat. *BIOS: Jurnal Teknologi Informasi Dan Rekayasa Komputer*, *2*(1), 11–20.

Prasetio, A., & Effendi, M. M. (2023). Analisis Gempa Bumi Di Indonesia Dengan Metode Clustering. *Bulletin of Information Technology (BIT)*, *4*(3), 338–343.

Rifa, I. H., Pratiwi, H., & Respatiwulan, R. (2020). Clustering of earthquake risk in Indonesia using k-medoids and k-means algorithms. *Media Statistika*, *13*(2), 194–205.

Rinjani, I., Anwar, S., & Herdiana, R. (2023). Pengelompokan Daerah Bencana Alam Menggunakan Algoritma K-Means Clustering. *Jurnal Ilmiah Sistem Informasi Dan Ilmu Komputer*, *3*(1), 35–51.

Shahapure, K. R., & Nicholas, C. (2020). Cluster quality analysis using silhouette score. *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, 747–748.

Yunistya, D. I., Goejantoro, R., & Amijaya, F. D. T. (2022). The Application Of K–Harmonic Means Method In District/City Grouping. *Jurnal Matematika, Statistika Dan Komputasi*, *19*(1), 51–64.