

PENERAPAN METODE K-MEANS PADA KLASTERISASI WILAYAH RAWAN GEMPA DI INDONESIA

Implementation of K-Means Method in Classterization of Earthquake Prone Areas in Indonesia

Nurfidah Dwitiyanti¹, Siti Ayu Kumala², Shinta Dwi Handayani³

^{1,2,3} Teknik Informatika, Universitas Indraprasta PGRI, Jakarta

Corresponding author : nurfidah.pulungan@gmail.com

Abstrak

Indonesia terletak pada pertemuan antara tiga lempeng tektonik utama bumi sehingga mengakibatkan wilayah di Indonesia rawan terjadi gempa bumi. Perlu dilakukan upaya untuk pengelompokan wilayah di Indonesia yang sering terjadi gempa bumi agar dapat memudahkan proses mitigasi bencana. Metode clustering menjadi salah satu metode yang dapat digunakan untuk mengelompokkan wilayah rawan gempa bumi. Pada penelitian ini, metode clustering yang digunakan adalah metode K-Means. Tahapan penelitian yang dilakukan terdiri dari pengambilan data dari BMKG, preprocessing data, penerapan metode K-Means, visualisasi dan interpretasi hasil klaster. Hasil dari penelitian ini diperoleh untuk jumlah klaster yang optimal adalah $k = 6$, dengan melihat nilai shilhoutte index sebesar 0,3245. Jumlah kejadian gempa paling banyak terjadi sebesar 5.125 untuk sebuah klaster dan paling sedikit sebesar 632 untuk klaster lainnya pada jumlah klaster $k = 6$.

Kata Kunci : Data Gempa, Metode K-Means, Shilhoutte Index, Metode Clustering

Abstract

Indonesia is located at the meeting point of three of the Earth's main tectonic plates, making areas in Indonesia prone to earthquakes. Efforts need to be made to group areas in Indonesia where earthquakes frequently occur in order to facilitate the disaster mitigation process. The clustering method is one method that can be used to group earthquake-prone areas. In this research, the clustering method used is the K-Means method. The research stages carried out consisted of collecting data from BMKG, preprocessing data, applying the K-Means method, and visualizing and interpreting cluster results. The results of this research were obtained for the optimal number of clusters, $k=6$, by looking at the Silhouette index value of 0.3245. The maximum number of earthquake events was 5,125 in one cluster, and the minimum was 632 in another cluster, with the number of clusters $k=6$.

Keywords : Earthquake Data, K-Means Method, Shilhoutte Index, Clustering Method.

PENDAHULUAN

Berdasarkan letak astronomis, Indonesia berada antara 6° LU- 11° LS dan 95° BT- 141° BT.(Senduk & Nhita, 2019) Secara geografis Indonesia terletak antara Samudera Hindia/Samudera Indonesia dan antara Benua Asia dan Benua Australia. Secara regional Indonesia juga berada antara tiga lempeng tektonik utama bumi, yaitu lempeng Eurasia di Utara dan lempeng Indo-Australia di Selatan, dan lempeng pasifik di sebelah Timur.(Yana et al., 2018) Di daerah pertemuan antara dua lempeng tektonik banyak



menimbulkan gempa bumi (Hadi et al., 2023) Dengan banyaknya kejadian gempa bumi yang terjadi di Indonesia, maka berpotensi menyebabkan jatuhnya korban, baik korban jiwa termasuk luka-luka hingga meninggal dunia maupun kerusakan bangunan yang ada. (Setiawan et al., 2022) Bahaya bencana alam juga diperparah dengan waktu kejadiannya yang tak terduga atau datang secara tiba-tiba. Untuk itu diperlukan upaya mitigasi bencana yang dilakukan oleh semua pihak, baik dari pemerintah maupun peneliti dan akademisi sesuai dengan bidang keahliannya masing-masing. Dalam kaitannya dengan upaya mitigasi tersebut, teknik *clustering* pada *data mining* sangat bermanfaat untuk mengelompokkan data gempa bumi berdasarkan karakteristik yang sama sehingga dapat dijadikan dasar untuk memprediksi peristiwa gempa bumi kedepannya. (Murdiaty et al., 2020).

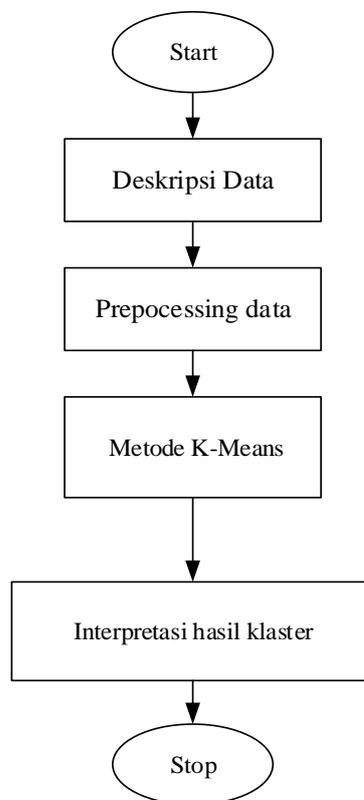
Data mining adalah suatu teknik yang digunakan untuk mengeksplorasi data dan mengekstraksi informasi penting dan diperlukan dari *database* besar (Shah & Jivani, 2013), Teknik *clustering* adalah salah satu dari Teknik data mining. Teknik *clustering* adalah teknik mengelompokkan data-data berdasarkan karakteristik yang dimilikinya sehingga dapat ditarik informasi yang bermanfaat dan diperlukan. (Prihandoko et al., 2017) Prinsip kerja dari teknik *clustering* adalah memaksimalkan kesamaan antar anggota dalam satu kelompok dan meminimumkan kesamaan antar kelompok (Artatia & Hakim, 2015) Ada banyak algoritma yang digunakan dalam analisis *clustering*. Penelitian ini menggunakan algoritma K-Means untuk mengelompokkan data gempa bumi di Indonesia dimana hasil penelitian ini diharapkan dapat digunakan untuk tindakan mitigasi bencana gempa bumi pada wilayah rawan gempa di Indonesia. K-Means merupakan salah satu metode pengelompokan data nonhierarki yang akan mempartisi data yang ada menjadi dua kelompok data atau lebih. Metode ini mempartisi data ke dalam kelompok sehingga data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain. Algoritma K-Means sederhana untuk diimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi, umumnya digunakan dalam praktek. (Iin et al., 2018). Parameter yang diinput pada saat menggunakan algoritma K-Means adalah nilai K. Nilai K yang digunakan didasarkan pada informasi terdahulu atau sebelumnya tentang berapa kelompok yang biasanya muncul.

Metode K-Means ini sudah banyak digunakan dalam mengolah data gempa bumi, seperti penelitian yang dilakukan oleh Kamath (2017), Suksmadana & Ramadhani (2022), dan Reviantika dkk (2020). Menurut penelitian yang dilakukan oleh Jufriansah dkk (2021) dan Hadi dkk (2023) dimana membandingkan metode *clustering* K=Means dan DBSCAN, hasil kedua penelitian tersebut disimpulkan bahwa metode K-Means dinilai lebih baik dalam melakukan clustering dibandingkan dengan hasil kluster dari metode DBSCAN, dilihat dari nilai Shilhouette Indeks. Masing-masing penelitian menggunakan data gempa di Indonesia yang bersumber dari Badan Meteorologi, Klimatologi dan Geofisika (BMKG) pada tahun 2019 dan United States Geological Survey (USGS) pada tahun 2012-2021. Pada penelitian ini, digunakan metode K-Means dengan menggunakan kebaruan data gempa bumi di Indonesia dengan rentang tahun

mulai dari Januari 2017 sampai dengan Januari 2023. Tujuan penelitian ini adalah untuk melihat hasil dari klasterisasi wilayah rentan gempa bumi di Indonesia menggunakan metode K-Means dengan keterbaruan data yang ada.

METODE

Pada penelitian ini dilakukan metode *clustering* dengan K-Means. Metode ini dipilih dengan alasan sudah umum dan banyak digunakan pada penelitian terkait klasterisasi gempa bumi. Kemudian, hasil klaster terbaik ditentukan dari perbandingan nilai *Silhouette Index*. Adapun, gambaran tahapan penelitian ini ditunjukkan pada gambar 1.



Gambar 1. Tahapan Penelitian

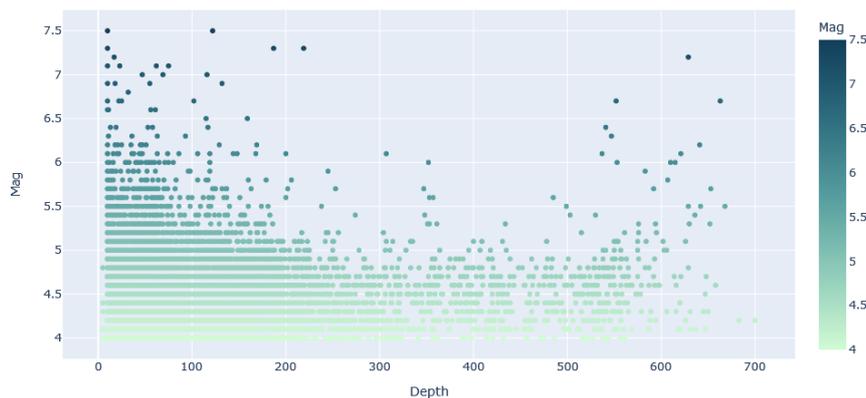
Pada tahap pertama data gempa bumi di Indonesia pada rentang tahun Januari 2017-Januari 2023 dikumpulkan berdasarkan magnitude ≥ 3 SR. Data bersumber dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Variabel yang digunakan meliputi tanggal (*date*), waktu (*time*), *latitude*, *longitude*, kekuatan gempa (*magnitude*), dan kedalaman (*depth*). Selanjutnya, dilakukan *preprocessing data* supaya data bersih dan tidak mengandung *missing value*. Tahap dilanjutkan dengan klasterisasi menggunakan metode *clustering* yaitu K-Means. Setelah itu, hasil klasterisasi dievaluasi dengan melihat nilai *Silhouette Index* sehingga diperoleh kesimpulan hasil klaster yang terbaik dari metode tersebut.

HASIL DAN PEMBAHASAN

Pembahasan akan diuraikan berdasarkan tahapan penelitian meliputi tahap deskripsi data, *preprocessing* data, penerapan metode K-Means, dan kemudian evaluasi dari hasil kluster dengan melihat nilai *Shilhouette Index* serta interpretasi dari hasil kluster Metode K-Means.

Deskripsi Data

Pada penelitian ini, data yang digunakan berupa data gempa bumi di Indonesia sepanjang tahun 2017-2023. Data bersumber dari BMKG. Total data gempa yang terekam sebanyak 13.860 dengan kekuatan gempa (*magnitudo*) ≥ 4 SR dan kedalaman (*depth*) 0 -700 km yang terjadi di Indonesia sepanjang tahun bulan Januari 2017 sampai dengan bulan Januari 2023, dimana Indonesia terbentang pada 6° LU - 11° LS dan 95° W - 141° BT (Senduk & Nhita, 2019). Dataset terdiri dari 7 atribut diantaranya tanggal, waktu, wilayah, *latitude*, *longitude*, kedalaman (*depth*) dan kekuatan (*magnitudo*). Adapun distribusi untuk kekuatan (*mag*) dan kedalaman (*depth*) data gempa dapat dilihat pada gambar 2.

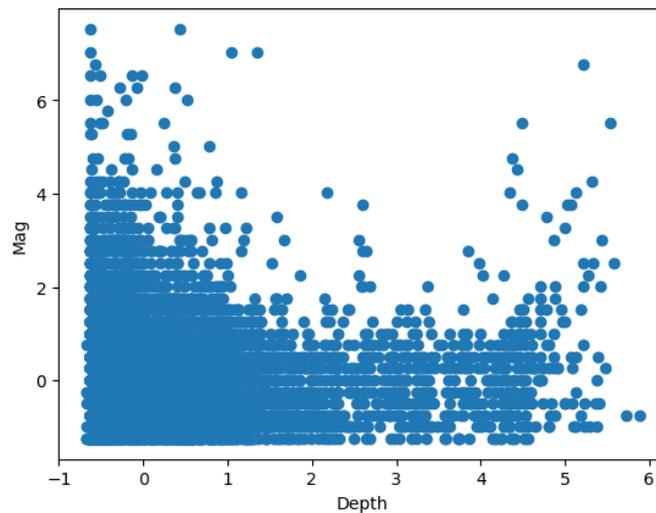


Gambar 2. Distribusi data gempa pada tahun 2017-2023

Pada gambar 2 dapat dilihat bahwa rentang kekuatan gempa terjadi di Indonesia sepanjang tahun 2017-2023 berkisar antara 4 – 7,5 SR. Sedangkan untuk rentang kedalaman berkisar antara 5 km sampai dengan 700 km.

Preprocessing data

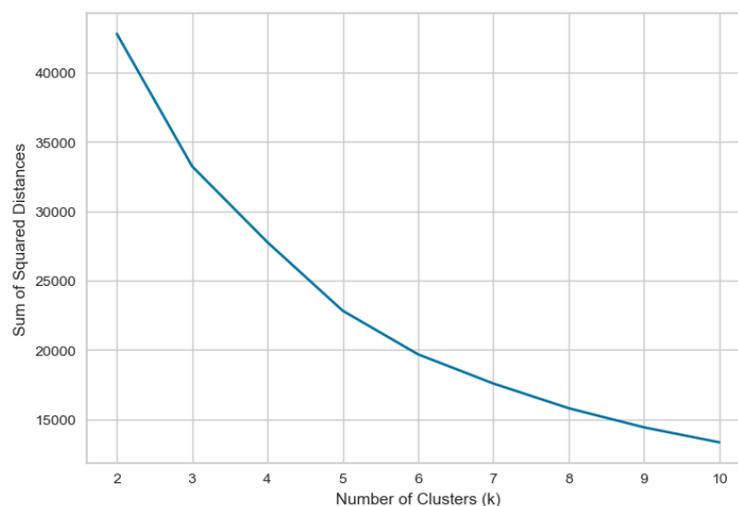
Pada tahapan *preprocessing* data dilakukan standarisasi data. Tujuan dilakukannya standarisasi terhadap data untuk memusatkan nilai *feature* pada mean 0 dan standar deviasi 1.



Gambar 3. Standardisasi data Magnitudo dan Depth

Metode K-Means Clustering

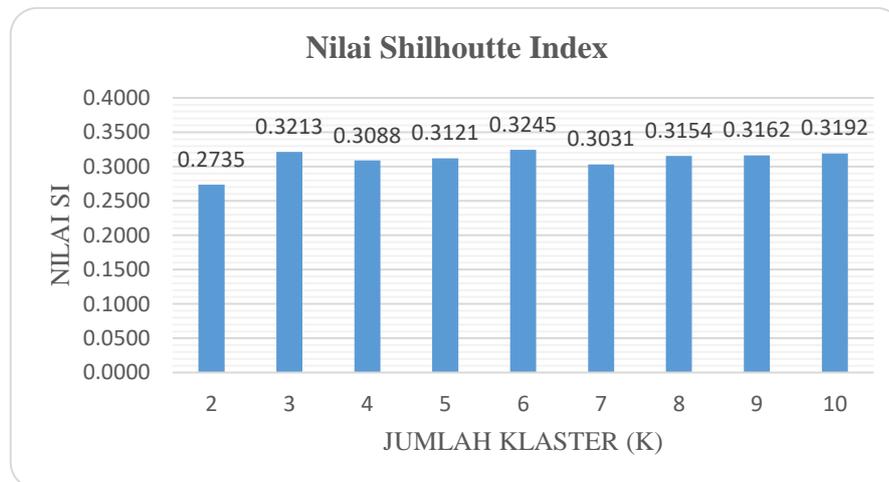
Dalam memanfaatkan metode *clustering* K-means, langkah awal yang dilakukan adalah penentuan jumlah kluster yang optimal. Hal ini sering dicapai dengan mempergunakan metode Elbow. Teknik ini sering digunakan dalam penentuan jumlah kluster terbaik dengan memeriksa sudut yang terbentuk pada titik tertentu ketika membandingkan jumlah kluster. (Novianta et al., 2023) Hasil dari metode Elbow, seperti ditunjukkan pada Gambar 4, menampilkan adanya lengkungan garis yang nyata pada $k = 6$. Hal ini memperlihatkan bahwa jumlah kluster yang ideal, sesuai dengan metode Elbow, adalah enam. Namun, untuk membandingkan jumlah kluster, perhitungan juga dilakukan untuk nilai k dari 2, 3, 4, 5, 7, 8, 9 dan 10.



Gambar 4. Grafik dari Metode Elbow

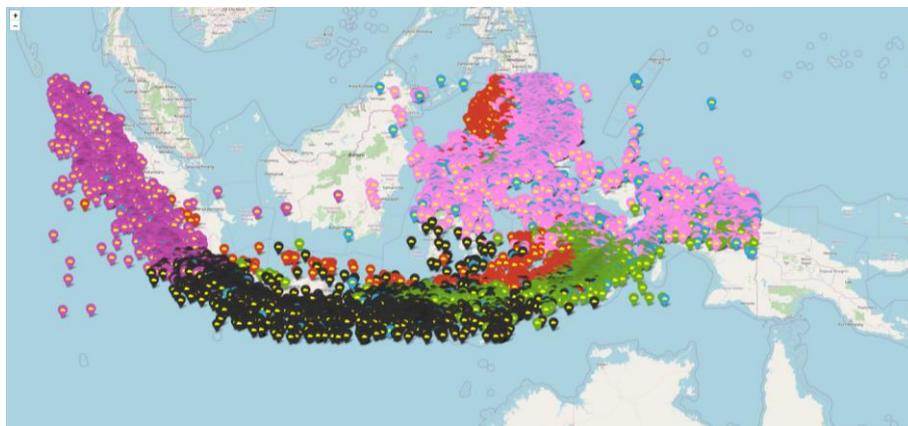
Selain itu, validitas kluster juga dilakukan untuk mengetahui hasil kluster yang diperoleh valid untuk digunakan atau tidak. Metode yang digunakan adalah dengan melihat *Silhouette Index* (SI). (Adha et al., 2021) Berdasarkan perbandingan, *Silhouette*

Index (SI) terbesar sebesar 0,3245 dengan jumlah kluster 6. Visualisasi hasil perbandingan ditunjukkan pada Gambar 5.



Gambar 5. Nilai *Shilhouette Index* untuk Metode K-Means

Pada penelitian ini dilakukan analisis dengan metode K-Means, dengan memilih jumlah kluster yang optimal adalah $k = 6$. Gambar 6. menunjukkan hasil pemetaan kluster menggunakan K-Means dengan jumlah kluster sebanyak 6 sebagai kluster dengan rata-rata *Silhouette Index* (SI) terbesar.



Gambar 6. Hasil Pemetaan Kluster Menggunakan Metode K-Means dengan $k = 6$

Setelah mendapatkan jumlah kluster yang optimal, selanjutnya dilakukan pembuatan profil pada $k=6$. Profiling ini bertujuan untuk mengetahui gambaran dan karakteristik masing-masing variabel pada setiap kluster. (Novianta et al., 2023) Pembuatan profil dilakukan berdasarkan nilai rata-rata. Tabel 1 menunjukkan profiling setiap variabel di setiap kluster dengan $k=6$.

Tabel 1. Profil Hasil Klaster dengan K-Means ($k = 6$)

Jumlah klaster (k)	Latitude	Longitude	Kekuatan Gempa (Magnitude)	Kedalaman (Depth)	Banyak Kejadian
1	-6,73	128,77	4,51	152,17	2164
2	-0,21	126,87	5,16	49,59	1682
3	-1,03	99,47	4,51	42,00	1633
4	-8,66	114,56	4,40	28,91	2624
5	0,38	127,79	4,32	42,93	5125
6	-3,69	123,69	4,53	451,62	632

Hasil klaster dengan $k = 6$ dapat dilihat pada tabel 1, pada tabel ini tampak kejadian gempa dikelompokkan menjadi 6 klaster. Klaster 1 sebanyak 2.164, klaster 2 sebanyak 1.682, klaster 3 sebanyak 1.633, klaster 4 sebanyak 2.624, klaster 5 sebanyak 5.125, dan klaster 6 sebanyak 623. Pembentukan klaster hampir sebagian besar terjadi pada wilayah di Indonesia meliputi Laut Banda, Sumatra Selatan, Sumbawa, Halmahera, Minahassa Peninsula, Laut Maluku Utara di Sulawesi, dan Pulau Talaud.

KESIMPULAN

Berdasarkan hasil klasterisasi metode K-Means yang digunakan pada penelitian ini, diperoleh jumlah klaster yang optimal terletak pada $k = 6$ dengan nilai Shilhouette Index sebesar 0,3245, dimana dihasilkan 6 klaster dengan Jumlah kejadian gempa paling banyak terjadi sebesar 5.125 untuk sebuah klaster dan paling sedikit sebesar 632 untuk klaster lainnya. Profil dari hasil klaster penelitian ini diharapkan dapat dijadikan sebagai acuan dalam menggambarkan model clustering daerah rawan gempa bumi di Indonesia. Kedepannya, hasil penelitian ini penulis pergunakan dalam penelitian lainnya yang berjudul “*Determining Clusters of Earthquake Prone Areas in Indonesia Using K-Medoids Algorithm*”. Pada penelitian ini, dilakukan perbandingan dengan metode clustering lainnya, meliputi metode K-Medoids, DBSCAN, Fuzzy C Means dan K-AP dengan menggunakan rentang tahun data yang sama.

UCAPAN TERIMA KASIH

Terima kasih diucapkan oleh penulis atas pendanaan melalui program PDP 2023 dari Direktorat Riset, Teknologi, dan Pengabdian kepada Masyarakat, serta Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi di bawah Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.

DAFTAR PUSTAKA

- Adha, R., Nurhaliza, N., Soleha, U., Studi, P., Informasi, S., & Sains, F. (2021). *Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia*. 18(2), 206–211.
- Artatia, H., & Hakim, R. F. (2015). Pengelompokan dampak gempa bumi dari segi kerusakan fasilitas pada provinsi yang berpotensi gempa di Indonesia menggunakan k-means-clustering. *Seminar Nasional Matematika Dan Pendidikan Matematika UMS 2015*, 742–748.
- Hadi, A. P. W., Pratiwi, H., & Slamet, I. (2023). Pengelompokan Data Gempa Bumi di Indonesia dengan Algoritma K-Means dan DBSCAN. *Seminar Nasional Pendidikan Matematika Ahmad Dahlan 2023*, 52–60.
- Iin, P., Windarto, A. P., Wanto, A., & Lubis, M. R. (2018). Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Assessment Center Untuk Clustering Program SDP. *Journal of Computer Engineering System and Science*, 3(1), 87–93.
- Jufriansah, A., Pramudya, Y., Khusnani, A., & Saputra, S. (2021). Analysis of Earthquake Activity in Indonesia by Clustering Method. *Journal of Physics: Theories and Applications*, 5(2), 92–103. <https://doi.org/10.20961/jphys theor-appl.v5i2.59133>
- Kamath, R. S. (2017). Earthquake Cluster Analysis : K-Means Approach Earthquake Cluster Analysis : K-Means Approach. *Journal of Chemical and Pharmaceutical Sciences*, 10(January), 250–253.
- Murdiaty, Angela, & Sylvia, C. (2020). Pengelompokan Data Bencana Alam Berdasarkan Wilayah , Waktu , Jumlah Korban dan Kerusakan Fasilitas Dengan Algoritma K-Means. *Jurnal Media Informatika Budidarma*, 4(3), 744–752. <https://doi.org/10.30865/mib.v4i3.2213>
- Novianta, M. A., Syafrudin, S., & Warsito, B. (2023). *K-Means Clustering for Grouping Rivers in DIY based on Water Quality Parameters*. 11(1), 155–163.
- Prihandoko, Bertalya, & Ramadhan, M. I. (2017). An Analysis of Natural Disaster Data by Using K-Means and K-Medoids Algorithm of Data Mining Techniques. *Intl. Conf. QiR: Intl. Symp. Elec. and Com. Eng*, 221–225.
- Reviantika, F., Harahap, C. N., Azhar, Y., Informatika, J. T., Malang, U. M., Selatan, L., & Data, P. (2020). *Analisis Gempa Bumi Pada Pulau Jawa Menggunakan Clustering Algoritma K-Means Analysis Earthquake in Java Island Using Clustering K-Means Algorithm*. 9(1), 51–60.
- Senduk, F. R., & Nhita, F. (2019). *Clustering of Earthquake Prone Areas in Indonesia Using K-Medoids Algorithm*. 4(2016), 65–76.



<https://doi.org/10.21108/indojc.2019.4.3.359>

- Setiawan, I., Krismawati, D., Pramana, S., & Tanur, E. (2022). Klasterisasi Wilayah Rentan Bencana Alam Berupa Gerakan Tanah dan Gempa Bumi di Indonesia. *Seminar Nasional Official Statistics*, 669–676.
- Shah, C., & Jivani, A. G. (2013). Comparison of Data Mining Classification Algorithms for Breast Cancer Prediction. *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, 1–4. <https://doi.org/10.1109/ICCCNT.2013.6726477>
- Suksmadana, I. M. B., & Ramadhani, C. (2022). Pengklasteran Kejadian Gempa Wilayah Indonesia Menggunakan Algoritma *k-Means*. 8(2), 62–67.
- Yana, M. S., Setiawan, L., Ulfa, E. M., & Rusyana, A. (2018). Penerapan Metode *K-Means* dalam Pengelompokan Wilayah Menurut Intensitas Kejadian Bencana Alam di Indonesia Tahun 2013-2018. *Journal of Data Analysis*, 1(2), 93–102.