

# Komparasi Klasterisasi Data Historis Gempa Bumi Menggunakan DBSCAN, K-Means, dan Agglomerative Clustering

Eka Therina Lakeisyah, Ken Ditha Tania\*, Mira Afrina

Fakultas Ilmu Komputer, Sistem Informasi, Universitas Sriwijaya, Palembang, Indonesia

Email: <sup>1</sup>09031282227068@student.unsri.ac.id, <sup>2,\*</sup>kenya.tania@gmail.com, <sup>3</sup>miraafрина81@gmail.com

Email Penulis Korespondensi: kenya.tania@gmail.com

Submitted: 28/09/2025; Accepted: 06/12/2025; Published: 08/12/2025

**Abstrak**—Gempa bumi merupakan salah satu bencana alam yang rawan terjadi di Pulau Sumatera dan menjadi tantangan serius karena dapat menimbulkan dampak buruk bagi kehidupan manusia, seperti korban jiwa, kerugian materi, dan kerusakan lingkungan. Oleh karena itu, dibutuhkan pemetaan zona kerawanan gempa bumi untuk memberikan informasi mengenai potensi dan riwayat bencana serta alat penting upaya mitigasi bencana. Penelitian ini bertujuan untuk memetakan data historis gempa bumi di Pulau Sumatera dengan membandingkan tiga algoritma klasterisasi, yaitu *DBSCAN*, *K-Means*, dan *Agglomerative Clustering* berdasarkan data gempa bumi di Pulau Sumatera tahun 1973-2023 dengan atribut yang digunakan, yaitu latitude, longitude, depth, dan magnitude. Hal ini bertujuan untuk menemukan algoritma terbaik sehingga dapat memberikan rekomendasi strategi mitigasi resiko gempa bumi yang tepat. Hasil penelitian menunjukkan bahwa algoritma *K-Means* merupakan yang terbaik karena memperoleh nilai *Silhouette Coefficient* tertinggi, yaitu 0.3948 dengan jumlah klaster sebanyak 3. Diharapkan penelitian ini dapat meningkatkan pemahaman mengenai zona kerawanan gempa bumi di Pulau Sumatera serta memberikan kontribusi praktis berupa rekomendasi strategi mitigasi yang disesuaikan dengan karakteristik tiap klaster untuk mendukung pemanfaatan hasil penelitian ini menjadi lebih aplikatif bagi pemerintah dan masyarakat setempat.

**Kata Kunci:** Gempa Bumi; Klasterisasi; DBSCAN; K-Means; Agglomerative Clustering

**Abstract**—Earthquakes are one of the natural disasters that are prone to occur on the island of Sumatera and pose a serious challenge because they can have a devastating impact on human life, such as loss of life, material losses, and environmental damage. Therefore, earthquake hazard zone mapping is needed to provide information about the potential and history of disasters and is an important tool for disaster mitigation efforts. This study aims to map earthquake vulnerability in Sumatra by comparing three clustering algorithms, namely DBSCAN, K-Means, and Agglomerative Clustering, based on earthquake data in Sumatra from 1973 to 2023. This is to find the best algorithm so that it can provide recommendations for appropriate earthquake risk mitigation strategies. The results show that the K-Means algorithm is the best because it obtained the highest Silhouette Coefficient value, namely 0.3948 with a total of 3 clusters. It is hoped that this research can improve understanding of earthquake hazard zones on the island of Sumatra and provide practical contributions in the form of mitigation strategy recommendations tailored to the characteristics of each cluster to support the application of this research for the government and local communities.

**Keywords:** Earthquake; Clustering; DBSCAN; K-Means; Agglomerative Clustering

## 1. PENDAHULUAN

Gempa bumi merupakan bencana alam yang menjadi tantangan serius karena dapat menimbulkan penderitaan bagi kehidupan manusia seperti kehilangan nyawa, kerugian harta benda, dan kerusakan lingkungan [1]. Pulau Sumatera adalah salah satu daerah di Indonesia yang rentan terhadap gempa bumi. Hal tersebut dikarenakan posisi Pulau Sumatera yang dilalui oleh sesar aktif, jalur gunung berapi, dan zona subduksi. Menurut Badan Nasional Penanggulangan Bencana (BNPB), dalam seratus tahun terakhir (1921-2021) empat dari tujuh gempa terbesar yang pernah terjadi di Indonesia ada di Pulau Sumatera, yaitu gempa Aceh dengan kekuatan 9,3 M pada tahun 2004, gempa Pulau Nias dengan kekuatan 8,6 M pada tahun 2005, gempa Pulau Sumatera dengan kekuatan 8,5 M pada tahun 2012, dan gempa Pulau Mentawai dengan kekuatan 7,8 M pada tahun 2016 [2]. Gempa bumi merupakan bencana yang sulit dihindari tetapi dampaknya dapat diminimalisir. Oleh karena itu, dibutuhkan upaya mitigasi bencana yang efektif dan berkelanjutan. Salah satu caranya yaitu dengan melakukan pemetaan dan pemahaman mengenai zona kerawanan gempa bumi [3], [4].

Kerawanan (*susceptibility*) merupakan tahapan sebelum terjadi bencana yang menekankan pada kondisi fisik dari suatu wilayah yang membuatnya rentan terhadap suatu bencana tertentu [5]. Pemetaan kerawanan bencana berfungsi untuk menyediakan gambaran tentang potensi serta rekam jejak bencana pada suatu wilayah, yang dapat dimanfaatkan oleh para pengambil keputusan untuk menilai manfaat dan risiko sebelum menetapkan kebijakan [6]. Pemetaan kerawanan gempa bumi ini bertujuan untuk menggambarkan seberapa rentan suatu wilayah terhadap ancaman gempa bumi [7]. Salah satu upaya mitigasi gempa bumi ialah dengan mengetahui daerah rawan melalui pengelompokan data gempa bumi menggunakan metode *clustering*. Adanya *clustering* ini berguna untuk mengelompokkan gempa berdasarkan karakteristiknya, sehingga upaya mitigasi dapat disesuaikan dengan karakteristik unik pada tiap klaster [8], [9].

Beragam penelitian telah dilakukan mengenai klasterisasi gempa bumi di berbagai wilayah. Berbagai algoritma telah diterapkan termasuk algoritma DBSCAN, K-Means, dan Agglomerative Clustering. Penelitian oleh Arifin (2023) mengelompokkan titik kejadian gempa bumi di Asia Tenggara menggunakan metode Agglomerative Clustering. Penelitian ini menghasilkan nilai Silhouette Coefficient sebesar 0,434353 yang menandakan bahwa klaster memiliki struktur yang cukup baik [10]. Penelitian yang dilakukan oleh Bariklana dan Fauzan (2023)

mengelompokkan sebaran lokasi gempa di Jawa Barat selama tahun 2021 menggunakan metode DBSCAN yang menghasilkan 12 kluster. Hasil ini memperoleh nilai Silhouette Coefficient sebesar 0.713 yang dianggap menunjukkan struktur kluster yang kuat [11]. Penelitian oleh Arafat et al. (2023) mengelompokkan data gempa bumi di wilayah Regional VII (Jawa Tengah, Yogyakarta, dan Jawa Timur) dengan jumlah kluster yang terbentuk sebanyak 2 - 6 kluster tergantung periode data. Evaluasi kualitas kluster menggunakan Silhouette Coefficient menghasilkan nilai 0,270 - 0,499 yang berarti kualitas klusterisasi cukup baik namun bervariasi [12]. Sementara itu, penelitian oleh Sippan dan Setiyawati (2025) menggunakan metode K-Means dalam mengelompokkan daerah rawan bencana alam di Provinsi Sulawesi Tengah yang menghasilkan 5 daerah termasuk kategori rawan bencana dan 8 daerah termasuk kategori tidak rawan bencana. Hasil pengujian menunjukkan nilai Davies Bouldin Index adalah 0,6531711248 yang dianggap relatif baik [13]. Adapun penelitian oleh Wahyu dan Rushendra (2022) mengelompokkan tingkat keparahan dampak gempa bumi di Pulau Jawa menggunakan algoritma K-Means. Penelitian ini menghasilkan 4 kluster dengan nilai MAE, MSE, standar deviasi, dan variansi relatif kecil yang menunjukkan validitas model relatif baik sehingga penggunaan algoritma K-Means cukup efektif [14].

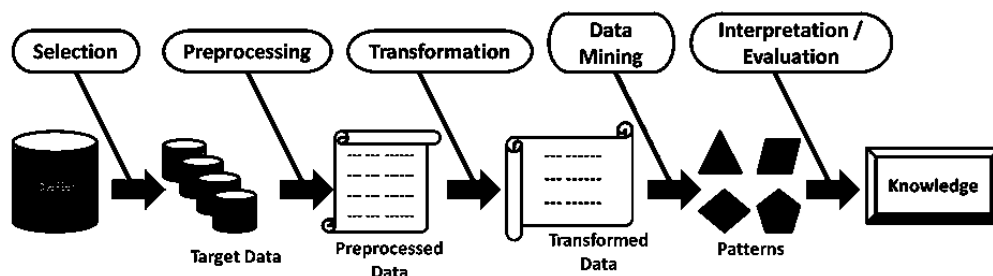
Berbeda dari penelitian sebelumnya, penelitian ini akan mencakup data gempa bumi yang terspesifikasi di pulau Sumatera dengan dataset dari tahun 1973 hingga 2023 menggunakan empat atribut utama, yaitu longitude, latitude, magnitude, dan depth. Selain itu, mengacu pada algoritma yang digunakan, pada kasus ini peneliti memilih untuk menerapkan algoritma DBSCAN, K-Means, dan Agglomerative Clustering yang masing-masing memiliki karakteristik yang unik serta diyakini layak untuk diuji pada data gempa bumi. Ketiga metode tersebut dipilih karena masing-masing mewakili paradigma klusterisasi yang berbeda, DBSCAN berbasis kepadatan, K-Means berbasis partisi, dan Agglomerative berbasis hierarki. Dengan karakteristik gempa bumi yang memiliki distribusi spasial tidak beraturan, rentang magnitudo yang beragam, serta kemungkinan adanya *noise*, penggunaan satu algoritma saja berpotensi menghasilkan gambaran yang kurang komprehensif. Pernyataan ini diperkuat dengan adanya penelitian oleh Jain et al. (2025) yang mengimplementasikan algoritma clustering dengan tiga paradigma yang berbeda, yaitu DBSCAN, K-Means, dan Agglomerative Clustering pada data geospasial dengan menekankan pentingnya pemilihan algoritma *clustering* yang tepat berdasarkan karakteristik dataset sehingga dilakukan evaluasi beberapa paradigma klusterisasi sebelum menentukan metode terbaik [15].

Penelitian ini bertujuan untuk memetakan data historis gempa bumi di Pulau Sumatera dengan membandingkan kinerja tiga algoritma yaitu DBSCAN, K-Means, dan Agglomerative Clustering untuk menemukan algoritma terbaik sehingga dapat merumuskan rekomendasi strategi mitigasi gempa berbasis data di Pulau Sumatera dengan tepat. Diharapkan penelitian ini dapat meningkatkan pemahaman mengenai zona kerawanan gempa bumi di wilayah tersebut, dan memberikan kontribusi kepada pemerintah dan masyarakat setempat terkait pengembangan strategi mitigasi risiko bencana gempa bumi di Pulau Sumatera.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian ini menerapkan metode Knowledge Discovery in Databases (KDD), yaitu sebuah proses multidisiplin yang bertujuan mengungkap pola yang valid, baru, bermanfaat, serta mudah dipahami dari suatu kumpulan data [16]. Pengolahan data dilakukan melalui Google Colab menggunakan bahasa pemrograman Python. Tahapan KDD disajikan pada gambar 1 yang diadopsi dari [17].



Gambar 1. Tahapan Knowledge Discovery in Databases (KDD)

Gambar 1 menunjukkan alur utama proses Knowledge Discovery in Databases (KDD). Proses ini dimulai dari selection, preprocessing, transformation, data mining, hingga interpretation/evaluation.

### 2.2 Selection

Penelitian ini menerapkan pendekatan kuantitatif dengan memanfaatkan data sekunder yang bersumber dari platform Kaggle melalui situs <https://www.kaggle.com/datasets/fikridwialpian/earthquake-in-sumatera>. Dataset ini berisi data gempa bumi di pulau Sumatera yang terjadi pada tahun 1973-2023. Pada tahap awal, dilakukan pemilihan atribut yang relevan dari kumpulan dataset yang ada [18]. Pada penelitian ini, dilakukan penyeleksian terhadap atribut-atribut data yang tidak relevan, sehingga diperoleh empat atribut utama, yaitu longitude, latitude, magnitude, dan depth.

### 2.3 Preprocessing

Tahap ini merupakan tahap evaluasi kualitas data. Data yang tidak memiliki nilai tidak akan dimasukkan ke dalam dataset [19]. Pada tahap ini akan dilakukan pembersihan dataset dari data yang bernilai *null* atau *missing*.

### 2.4 Transformation

Tahap transformasi merupakan proses mengonversi data ke dalam bentuk yang lebih cocok untuk diterapkan dalam data mining [20]. Data diproses dengan melakukan teknik normalisasi data untuk menyamakan skala pada setiap atribut data. Normalisasi data dilakukan menggunakan metode *Z-score* dengan *StandardScaler* terhadap seluruh atribut, yaitu longitude, latitude, depth, dan magnitude. Meskipun atribut latitude dan longitude merepresentasikan koordinat geografis, normalisasi tetap diterapkan untuk memastikan kesetaraan skala antar dimensi data. Metode *z-score normalization* yang ditunjukkan persamaan 1 dimana *zscore* adalah nilai hasil normalisasi,  $\mu_x$  adalah nilai rata-rata untuk variabel *x*, dan  $\sigma_x$  adalah simpangan baku variabel *x* [21].

$$zscore = \frac{x - \mu_x}{\sigma_x} \quad (1)$$

### 2.5 Data Mining

Tahapan data mining dilakukan menggunakan algoritma DBSCAN, K-Means, dan Agglomerative Clustering. Ketiga algoritma ini dibandingkan untuk mendapatkan hasil klusterisasi terbaik terhadap data historis gempa bumi di Pulau Sumatera.

#### 2.5.1 DBSCAN

Algoritma DBSCAN digunakan karena memiliki kemampuan dalam mengidentifikasi kluster dengan kepadatan yang bervariasi dan efektif menyaring noise [15]. Parameter utama dalam DBSCAN adalah Eps (epsilon) dan MinPts (*minimum points*). Penentuan parameter Eps dilakukan dengan menganalisis grafik *k-distance*, yaitu dengan memplot jarak ke *k*-tetangga terdekat dan memilih nilai ambang (Eps) pada titik siku (*elbow point*) dari grafik tersebut, sedangkan nilai MinPts ditentukan berdasarkan kepadatan data. Analisis *clustering* menggunakan algoritma DBSCAN mencakup beberapa langkah berikut [11]. DBSCAN membentuk kluster dengan memilih titik awal secara acak, lalu menggunakan parameter Eps dan MinPts untuk menentukan kepadatan di sekitar titik tersebut. Jika titik termasuk *core point*, maka akan membentuk kluster bersama titik-titik yang dapat dijangkau darinya. Proses terus diulangi hingga semua titik diproses dan hasil akhirnya tidak tergantung pada urutan pemrosesan.

#### 2.5.2 K-Means

Algoritma K-Means dipilih karena prosesnya yang sederhana dan mudah diadaptasi sehingga dapat diaplikasikan di berbagai bidang [22]. Parameter utama dalam algoritma ini adalah jumlah kluster (*k*). Pemilihan nilai *k* dilakukan dengan menerapkan metode elbow, di mana titik siku pada grafik digunakan sebagai acuan nilai *k* optimal. Algoritma K-Means melibatkan beberapa langkah berikut [23]. Pilih secara acak *k* dari *n* titik data untuk dijadikan centroid awal. Lalu, setiap titik data ke dalam kluster dengan centroid yang paling dekat. Setelah itu, perbarui posisi centroid dengan menghitung rata-rata dari seluruh data pada masing-masing kluster. Ulangi proses tersebut hingga posisi centroid stabil dan tidak mengalami perubahan lagi.

#### 2.5.3 Agglomerative Clustering

Algoritma Agglomerative Clustering dipilih karena fleksibilitasnya dalam membentuk struktur hierarkis tanpa memerlukan jumlah kluster di awal [24]. Parameter penting dalam algoritma ini meliputi metode linkage dan jumlah kluster hasil pemotongan dendrogram. Penelitian ini menggunakan metode Ward linkage, karena pendekatan ini meminimalkan jumlah variansi dalam kluster dan cocok untuk data numerik. Jumlah kluster ditentukan dengan mengamati dendrogram yang dihasilkan, kemudian memotongnya pada tingkat jarak yang menunjukkan pemisahan paling alami antar kelompok data. Algoritma Agglomerative Clustering juga melibatkan beberapa langkah penting [24]. Proses dimulai dengan masing-masing data dianggap sebagai kluster individu, kemudian algoritma secara berulang menggabungkan kelompok terdekat menjadi kelompok yang lebih besar. Ulangi langkah penggabungan dan pembaruan tersebut hingga hanya satu kelompok yang tersisa, membentuk struktur hierarki yang bergerak ke atas.

### 2.6 Evaluation

Evaluasi kualitas kluster dilakukan dengan menggunakan metode Silhouette Coefficient yang nilainya berkisar antara -1 sampai 1. Nilai Silhouette Coefficient yang semakin tinggi menunjukkan bahwa kualitas kluster yang terbentuk semakin baik [25]. Rumus untuk menghitung Silhouette Coefficient ditunjukkan pada persamaan 2 dimana *s(i)* adalah nilai Silhouette Coefficient data ke-*i*, *b(i)* adalah rata-rata jarak antara titik *i* ke seluruh titik dalam kluster yang berbeda, sedangkan *a(i)* adalah rata-rata jarak antara titik *i* ke seluruh titik dalam kluster yang sama [21].

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

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Selection

Pada tahap ini dilakukan penghapusan pada beberapa atribut data gempa bumi yang tidak relevan. Terdapat 13 atribut data dalam dataset dan hanya empat atribut yang digunakan, yaitu latitude, longitude, depth, dan magnitude. Hasil dari seleksi data disajikan dalam Tabel 1.

Tabel 1. Hasil Seleksi Data

Latitude	Longitude	Depth/km	Magnitude
-2,6871	101,2477	89,7	5,2
1,1532	101,2299	0	4,6
-4,389	102,9845	91,5	5,4
-1,7458	99,6507	57,7	5,4
-1,7261	99,6842	37	4,9

#### 3.2 Preprocessing

Pada tahap ini dilakukan pemeriksaan terhadap data yang hilang (*missing values*). Hasil pengecekan menunjukkan ada beberapa data bernilai 0 yang dianggap tidak valid dan diperlakukan sebagai *missing value*. Data tersebut akan dihapus sehingga dari 24082 data menjadi 19344 data. Hasil dari preprocessing data akan ditunjukkan pada Tabel 2.

Tabel 2. Hasil Preprocessing Data

Latitude	Longitude	Depth/km	Magnitude
-2,6871	101,2477	89,7	5,2
-4,389	102,9845	91,5	5,4
-1,7458	99,6507	57,7	5,4
-1,7261	99,6842	37	4,9
-1,8136	99,5909	34,4	5,2

#### 3.3 Transformation

Pada tahap ini dilakukan teknik normalisasi data dengan metode z-score untuk menyamakan skala yang diterapkan pada seluruh atribut. Hasil dari proses transformasi ditampilkan pada Tabel 3.

Tabel 3. Hasil Transformasi Data

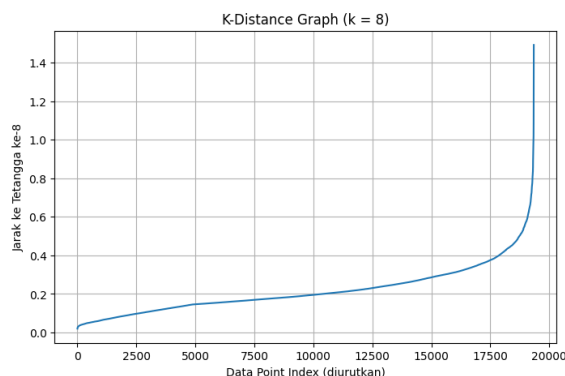
Latitude	Longitude	Depth/km	Magnitude
-0.710180	0.773092	0.730200	1.475429
-1.247832	1.449467	0.765433	1.757495
-0.412811	0.151161	0.103843	1.757495
-0.406587	0.164207	-0.301332	1.052332
-0.434229	0.127872	-0.352223	1.475429

#### 3.4 Data Mining

Pada tahap ini dilakukan pengujian tiga algoritma, yaitu DBSCAN, K-Means, dan Agglomerative Clustering.

##### 3.4.1 DBSCAN

Pengelompokan menggunakan algoritma DBSCAN dimulai dengan menentukan parameter epsilon ( $\epsilon$ ) dan MinPts. Pada penelitian ini, nilai  $\epsilon$  ditentukan menggunakan grafik k-distance.



Gambar 2. Grafik K-distance

Berdasarkan Gambar 2, nilai  $\epsilon$  yang optimal berada pada titik 0.4 dengan MinPts optimal sebesar 8. Tahap berikutnya, dilakukan proses klasterisasi menggunakan algoritma DBSCAN dengan nilai  $\epsilon$  dan MinPts yang telah didapatkan sebelumnya. Hasil pengolahan menggunakan algoritma DBSCAN dapat dilihat pada Gambar 3.

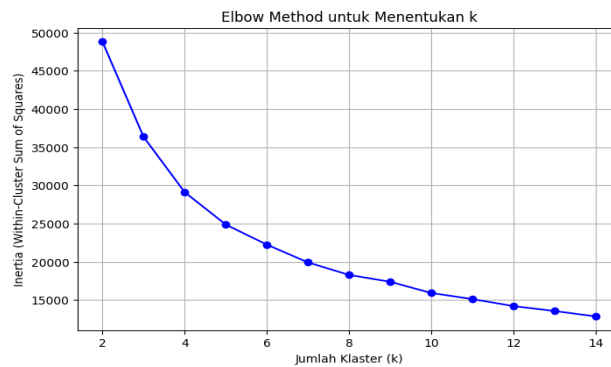


Gambar 3. Hasil Clustering DBSCAN

Gambar 3 menunjukkan hasil klasterisasi menggunakan algoritma DBSCAN dengan titik-titik di sekitar inti dianggap titik batas (*border point*), sedangkan titik yang tidak sesuai dengan kriteria akan dianggap sebagai *outlier* (label -1). Hasil DBSCAN menunjukkan terbentuknya 6 klaster dan 634 *outlier*.

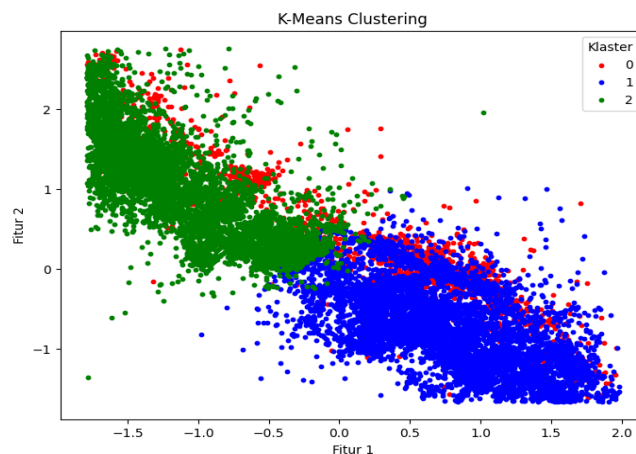
### 3.4.2 K-Means

Langkah pertama dalam pengelompokan algoritma K-Means yaitu dengan menentukan banyak klaster optimal ( $k$ ) dengan menggunakan elbow method yang ditunjukkan pada Gambar 4.



Gambar 4. Grafik Elbow Method

Dapat dilihat pada Gambar 4 bahwa hasil pengolahan elbow method adalah sebanyak 3 klaster. Selanjutnya dilakukan klasterisasi menggunakan algoritma K-Means dan hasilnya ditunjukkan pada Gambar 5 berikut.

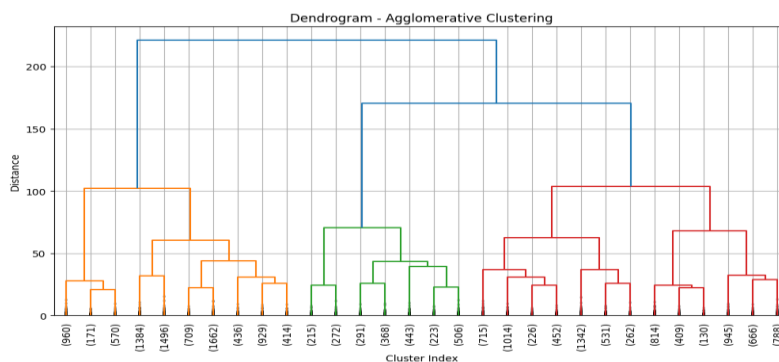


Gambar 5. Hasil Clustering K-Means

Gambar 5 menunjukkan hasil klusterisasi dengan algoritma K-Means dengan iterasi berhenti saat perubahan posisi centroid sudah sangat kecil atau model mencapai konvergensi sehingga membentuk 3 kluster.

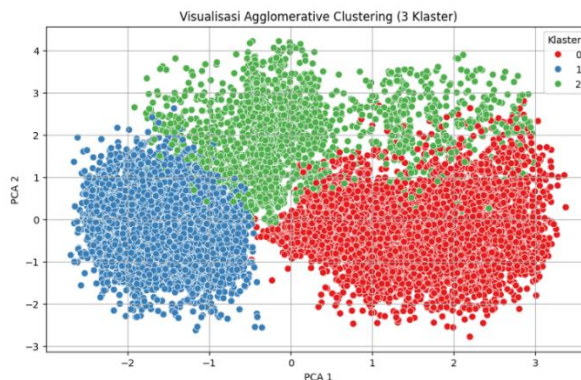
### 3.4.3 Agglomerative Clustering

Algoritma agglomerative digunakan dengan pendekatan *bottom-up*, dimana setiap titik awalnya menjadi kluster tunggal. Penggabungan kluster dilakukan secara berulang berdasarkan tingkat kemiripan tertinggi. Pada algoritma Agglomerative Clustering ini untuk menetapkan jumlah kluster yang ideal dilakukan menggunakan dendrogram yang dihasilkan melalui metode Ward linkage, yaitu metode penggabungan yang meminimalkan peningkatan varians dalam setiap kluster. Hasil pembentukan dendrogram ditunjukkan pada Gambar 6.



Gambar 6. Hasil Dendrogram

Pada Gambar 6 menunjukkan struktur kluster berdasarkan kedekatan karakteristiknya. Berdasarkan dendrogram tersebut, jumlah kluster optimal sebanyak 3 kluster. Hasil dendrogram tersebut selanjutnya digunakan untuk tahap klusterisasi menggunakan algoritma Agglomerative Clustering yang disajikan pada Gambar 7.



Gambar 7. Hasil Agglomerative Clustering

Gambar 7 menunjukkan hasil klusterisasi dengan algoritma Agglomerative Clustering yang menunjukkan bahwa pembagian menjadi 3 kluster paling representatif.

### 3.5 Evaluation

Setelah melakukan pengujian tiga algoritma di atas, berikut merupakan hasil evaluasi untuk mengetahui kualitas dan akurasi dari kluster tersebut. Evaluasi menggunakan Silhouette Coefficient yang disajikan Tabel 4 menunjukkan bahwa algoritma K-Means merupakan yang terbaik karena mendapatkan hasil Silhouette Coefficient tertinggi.

Tabel 4. Hasil Evaluasi

Algoritma	Silhouette Coefficient	Jumlah Kluster
DBSCAN	0.3864	6
K-Means	0.3948	3
Agglomerative Clustering	0.3638	3

Secara keseluruhan, silhouette coefficient yang relatif rendah pada ketiga algoritma ( $<0.5$ ) menunjukkan bahwa data gempa bumi di Sumatera bersifat kompleks, memiliki distribusi yang menyebar, dan tidak membentuk batas kluster yang tegas. Hal ini juga mencerminkan karakteristik fenomena geologi di mana aktivitas gempa mengikuti pola sesar dan zona subduksi yang memanjang, sehingga kluster tidak terkotak secara kompak sebagaimana idealnya dalam model kluster. Meskipun K-Means menghasilkan nilai Silhouette Coefficient tertinggi (0.3948), nilai DBSCAN (0.3864) dan Agglomerative Clustering (0.3638) yang lebih rendah perlu dianalisis. Pada DBSCAN, variasi kepadatan



titik gempa di sepanjang Sesar Sumatera dan zona subduksi menyebabkan algoritma ini membentuk kluster dengan kepadatan tidak seragam, sehingga menurunkan kompaksi kluster. Agglomerative Clustering pun mengalami kesulitan karena kedekatan spasial titik gempa yang tidak homogen, sehingga proses penggabungan tidak selalu menghasilkan batas kluster yang jelas. Dengan demikian, skor yang rendah mencerminkan bahwa struktur kluster pada data gempa tidak sepenuhnya terdefinisi dengan baik, terutama karena sifat data yang tidak membentuk kelompok yang jelas dan memiliki sebaran spasial yang kompleks. Namun, berdasarkan evaluasi kuantitatif, K-Means tetap memberikan pemisahan kluster yang paling stabil di antara ketiga metode yang diuji.

Menurut BMKG klasifikasi gempa berdasarkan kedalaman (depth) adalah sebagai berikut [26]. Gempa dangkal (< 60 km) adalah gempa yang paling merusak karena dekat dengan permukaan bumi, gempa menengah (60-300 km) adalah gempa yang menyebabkan getaran sedang dengan kerusakan tertentu, serta gempa dalam (>300 km) adalah gempa yang getarannya masih bisa dirasakan tetapi jarang menimbulkan kerusakan besar. Adapun klasifikasi gempa berdasarkan magnitudo menurut BMKG [26], yaitu gempa mag (2.5 – 5.4) adalah gempa yang sering dirasakan namun hanya menyebabkan kerusakan kecil, gempa mag (5.5 – 6.0) adalah gempa yang menimbulkan kerusakan ringan pada bangunan dan infrastruktur, gempa mag (6.1 – 6.9) adalah gempa yang menyebabkan banyak kerusakan pada pemukiman padat penduduk, gempa mag (7.0 – 7.9) adalah gempa besar yang menyebabkan banyak kerusakan fatal, serta gempa mag (≥ 8,0) adalah gempa hebat yang dapat menghancurkan komunitas di sekitar gempa. Berikut merupakan karakteristik parameter gempa per kluster yang disajikan pada Tabel 5.

**Tabel 5.** Karakteristik Parameter Gempa

Cluster	Depth/km		Magnitudo		Jumlah Data
	Range	Avg.	Range	Avg.	
0	67.00 - 205.40	144.87	2.50 - 6.10	3.84	2157
1	0.40 - 155.90	41.64	2.50 - 6.20	4.29	7350
2	0.30 - 120.60	32.93	2.50 - 6.20	4.11	9836

Berdasarkan klusterisasi menggunakan algoritma K-Means, berikut analisis setiap kluster dengan mempertimbangkan faktor-faktor gempa, seperti rata-rata kedalaman, rata-rata magnitudo, frekuensi kejadian, geografis, dan potensi penyebab gempa.

Kluster 0 memiliki rata-rata kedalaman menengah dengan rata-rata magnitudo paling rendah di antara kluster lainnya dan memiliki frekuensi kejadian yang tidak terlalu tinggi. Gempa pada kluster ini banyak terjadi di wilayah daratan dan dekat pusat pemukiman. Walaupun secara geofisik memiliki resiko rendah, getaran masih dapat dirasakan masyarakat karena lokasi yang dekat dengan pemukiman dan dapat merusak infrastruktur bangunan yang tidak dirancang dengan baik. Kluster ini berimpit dengan jalur Sesar Besar Sumatra (*Great Sumatran Fault*) yaitu sesar geser aktif yang memanjang di pulau Sumatera dan menjadi sumber utama gempa di daratan. Kluster 1 memiliki rata-rata kedalaman dangkal dengan rata-rata magnitudo tertinggi di antara kluster dan memiliki frekuensi kejadian yang tinggi. Gempa pada kluster ini berada di zona subduksi Sunda khususnya di segmen *megathrust* Mentawai yang menyimpan energi sangat besar. Lokasi yang terletak di lepas pantai dengan kombinasi magnitudo tinggi dan kedalaman dangkal dapat berpotensi memicu gempa *megathrust* dan risiko tsunami yang dahsyat. Kluster 2 memiliki rata-rata kedalaman paling dangkal dengan rata-rata magnitudo sedang dan memiliki frekuensi kejadian paling tinggi di antara kluster. Kluster ini terletak di zona subduksi Sunda bagian utara dimana adanya pertemuan lempeng Indo-Australia dan Eurasia yang aktif. Walaupun rata-rata magnitudonya tidak setinggi kluster 1, kluster ini rentan terhadap guncangan dan adanya risiko tsunami karena frekuensi kejadian yang sangat tinggi dan gempa yang dangkal.

Berikut merupakan klasifikasi tingkat kerawanan gempa untuk memberikan rekomendasi penyusunan strategi mitigasi bencana gempa bumi bagi masyarakat dan pemerintah di pulau Sumatera yang disajikan pada Tabel 6.

**Tabel 6.** Klasifikasi dan Mitigasi Kerawanan Gempa

Tingkat Kerawanan	Kluster	Warna	Rekomendasi Mitigasi
Sangat Rawan	1	Biru	<ol style="list-style-type: none"> <li>1. Tsunami dahsyat adalah ancaman utama, maka prioritasnya dengan memastikan keoptimalan sistem peringatan dini tsunami (InaTEWS) dan pemasangan sensor tekanan dasar laut (OBU) yang terhubung langsung dengan sirine di sepanjang pantai serta penyebaran hingga tingkat komunitas (seperti pengeras suara di masjid).</li> <li>2. Mengadakan latihan evakuasi secara berkala yang mencakup partisi dari semua kalangan, mulai dari anak-anak hingga dewasa.</li> <li>3. Mengedukasi masyarakat tentang <i>golden time</i> (waktu emas evakuasi).</li> <li>4. Memperbanyak dan merawat shelter evakuasi vertikal di area padat penduduk yang datar dan di zona pantai terdampak segmen Mentawai, serta memastikan jalur evakuasi horizontal (menuju perbukitan) selalu dalam kondisi baik, bebas hambatan, jelas dengan melengkapinya papan petunjuk yang mudah terlihat bahkan saat malam hari atau cuaca buruk, dan detail dengan peta evakuasi rute tercepat berbasis RT/RW.</li> </ol>

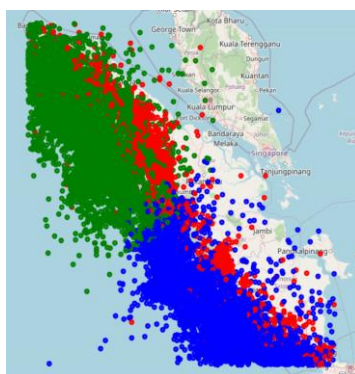
Tingkat Kerawanan	Klaster	Warna	Rekomendasi Mitigasi
			5. Menerapkan standar SNI gempa yang ketat untuk bangunan baru dan memprioritaskan program <i>retrofitting</i> (perkuatan struktur) untuk bangunan lama, terutama untuk fasilitas umum vital seperti sekolah dan rumah sakit.
Rawan	2	Hijau	<ol style="list-style-type: none"> <li>1. Dengan tingginya frekuensi gempa pada klaster ini, dibutuhkan struktur bangunan yang tahan terhadap guncangan gempa berulang dan melakukan audit ketahanan bangunan infrastruktur publik secara berkala karena seringnya guncangan dapat menyebabkan akumulasi kelelahan material.</li> <li>2. Mengintegrasikan materi kesiapsiagaan bencana gempa dan tsunami ke dalam kurikulum sekolah serta melakukan sosialisasi rutin dan pelatihan di tingkat komunitas mengenai teknik perlindungan diri dasar seperti <i>Drop, Cover, and Hold On</i> agar menjadi refleks otomatis individu saat gempa terjadi.</li> <li>3. Membentuk dan melatih Tim Reaksi Cepat bencana di tingkat desa/kelurahan yang mampu melakukan tindakan pertolongan pertama, penilaian kerusakan dini, dan komunikasi darurat untuk menciptakan kemandirian respons bencana sambil menunggu bantuan eksternal tiba.</li> <li>4. Menghimbau setiap keluarga menyiapkan Tas Siaga Bencana (TSB) dan menempatkannya di lokasi yang mudah dijangkau.</li> </ol>
Cukup Rawan	0	Merah	<ol style="list-style-type: none"> <li>1. Kerusakan struktural di darat akibat pusat gempa dekat dengan pemukiman merupakan ancaman utama, maka diperlukan inspeksi dan penguatan untuk infrastruktur kritis seperti jembatan, pipa, jaringan listrik yang berada di zona sesar untuk mencegah lumpuhnya layanan publik pasca-gempa.</li> <li>2. Memberikan bantuan teknis dan subsidi untuk penguatan struktur rumah warga (<i>non-engineered buildings</i>) yang berlokasi dalam radius paling rentan di sepanjang jalur sesar.</li> <li>3. Memberikan edukasi dan simulasi kesiapsiagaan yang terspesifikasi pada gempa darat seperti skenario evakuasi dari reruntuhan bangunan, evakuasi korban yang terisolasi, dan pemadaman kebakaran akibat korsleting listrik atau kebocoran gas, serta potensi bahaya sekunder, seperti tanah longsor di daerah berbukitan dan likuefaksi di daerah berpasir.</li> </ol>

Wilayah dalam klaster 1 menjadi prioritas utama dalam mitigasi bencana dengan daerah yang terkonsentrasi di lepas pantai sisi barat Sumtera yang sejajar dengan zona subduksi lempeng Samudra Hindia-Australia di bawah lempeng Eurasia. Daerah prioritas utamanya terdiri dari kepulauan Mentawai sebagai daerah terdekat dengan pusat potensi gempa *megathrust*, Pesisir Barat Provinsi Sumatra Barat, Pesisir Provinsi Bengkulu, daerah selatan Nias dan Pesisir Tapanuli Tengah.

Wilayah klaster 2 yang terkonsentrasi di lepas pantai utara sumatera yang memanjang ke wilayah utara terdiri atas daerah prioritas utama seperti Pesisir Provinsi Aceh yang memiliki kejadian gempa paling ekstrem, Pesisir Provinsi Sumatra Utara, dan Kepulauan Andaman.

Wilayah klaster 0 tersebar memanjang dari utara hingga selatan di daratan Pulau Sumatera mengikuti jalur Sesar Besar Sumatera. Daerah prioritas utama pada klaster ini terdiri dari Provinsi Sumatra Barat bagian darat yang dilintasi langsung oleh sesar, Kabupaten Kerinci dan Kota Sungai Penuh di Provinsi Jambi, daerah di sekitar zona sesar seperti Lebong, Kepahiang, dan dataran tinggi di perbatasan Bengkulu dan Sumatera Selatan, serta di sekitar segmen sesar Lampung Barat.

Berikut visualisasi peta sebaran titik gempa bumi di Sumatera menggunakan algoritma K-Means yang menghasilkan 3 klaster yang disajikan pada Gambar 8.



**Gambar 8.** Visualisasi Peta Gempa Sumatera

#### 4. KESIMPULAN

Penelitian ini membandingkan tiga algoritma yaitu DBSCAN, K-Means, dan Agglomerative Clustering dalam klusterisasi daerah gempa bumi di Pulau Sumatera menggunakan empat atribut utama gempa, yaitu latitude, longitude, magnitude, dan depth. Hasil pengujian menunjukkan bahwa algoritma K-Means adalah algoritma terbaik karena memperoleh Silhouette Score tertinggi, sebesar 0.3948 dengan jumlah kluster 3, yang menandakan algoritma ini memiliki struktur kluster yang lebih kompak dan terpisah dibandingkan algoritma lainnya. Sementara itu, algoritma DBSCAN memperoleh Silhouette Score 0.3864 dengan kluster yang terbentuk 6, lalu algoritma Agglomerative Clustering dengan perolehan Silhouette Score 0.3638 yang membentuk 3 kluster. Algoritma DBSCAN dan Agglomerative Clustering mencerminkan batas kluster yang kurang tegas dibandingkan K-Means. Penelitian ini tidak hanya berfokus pada pemetaan kluster risiko gempa, tetapi juga memberikan rekomendasi strategi mitigasi yang disesuaikan dengan karakteristik masing-masing tingkat kerawanan yang terdiri dari sangat rawan, rawan, dan cukup rawan. Dengan demikian, hasil penelitian menjadi lebih aplikatif dibandingkan penelitian terdahulu yang hanya terbatas pada analisis spasial. Penelitian ini dapat menjadi acuan bagi penelitian selanjutnya dengan menambahkan faktor non-seismik seperti kepadatan penduduk, kondisi infrastruktur, dan peta sesar aktif dari PUSGEN sehingga hasil penelitian menjadi lebih komprehensif. Selain itu, pengembangan *dashboard* interaktif berbasis klusterisasi juga akan mendukung pemanfaatan hasil studi ini agar lebih aplikatif bagi pemerintah dan masyarakat.

#### REFERENCES

- [1] A. Asman, E. Barlian, D. Hermon, I. Dewata, and I. Umar, "Mitigation and Adaptation of Community using AHP in Earthquake Disaster-Prone Areas in Pagar Alam City - Indonesia," *International Journal of Management and Humanities*, vol. 4, no. 9, pp. 34–38, May 2020, doi: 10.35940/ijmh.I0851.054920.
- [2] R. Fausta, "7 Gempa Terbesar Di Indonesia Dalam 100 Tahun Terakhir - Juara 2," Koleksi Sejarah Bencana (Portal Literasi Sejarah Bencana - Direktorat Sistem Penanggulangan Bencana, BNPB). Accessed: Sep. 16, 2025. [Online]. Available: <https://sejarah.dibi.bnpb.go.id/poster/7-Gempa-Terbesar-Di-Indonesia-Dalam-100-Tahun-Terakhir---Juara-2/6>
- [3] X. He, Z. Chen, Q. Yang, and C. Xu, "Advances in earthquake and cascading disasters," *Natural Hazards Research*, vol. 5, no. 2, pp. 421–431, 2025, doi: <https://doi.org/10.1016/j.nhres.2025.01.010>.
- [4] N. A. Akhundzadah, "Earthquake Risk Severity and Urgent Need for Disaster Management in Afghanistan," *GeoHazards*, vol. 6, no. 1, 2025, doi: 10.3390/geohazards6010009.
- [5] K. Thywissen, *Components of risk: a comparative glossary*. UNU-EHS, 2006.
- [6] A. L. Nugraha, M. Awaluddin, A. Sukmono, and N. Wakhidatus, "Pemetaan Dan Penilaian Kerentanan Bencana Alam Di Kabupaten Jepara Berbasis Sistem Informasi Geografis," *Geoid*, vol. 17, no. 2, p. 185, Apr. 2022, doi: 10.12962/j24423998.v17i2.9370.
- [7] S. H. Alavi, A. Bahrami, M. Mashayekhi, and M. Zolfaghari, "Optimizing Interpolation Methods and Point Distances for Accurate Earthquake Hazard Mapping," *Buildings*, vol. 14, no. 6, 2024, doi: 10.3390/buildings14061823.
- [8] B. I. Nasution, F. M. Saputra, R. Kurniawan, A. N. Ridwan, A. Fudholi, and B. Sumargo, "Urban vulnerability to floods investigation in Jakarta, Indonesia: A hybrid optimized fuzzy spatial clustering and news media analysis approach," *International Journal of Disaster Risk Reduction*, vol. 83, p. 103407, 2022, doi: <https://doi.org/10.1016/j.ijdr.2022.103407>.
- [9] A. Prasetyo, M. M. Effendi, and M. N. Dwi M, "Analisis Gempa Bumi Di Indonesia Dengan Metode Clustering," *Bulletin of Information Technology (BIT)*, vol. 4, no. 3, pp. 338–343, Sep. 2023, doi: 10.47065/bit.v4i3.820.
- [10] A. Arifin, "Clustering Earthquake Event Points in the Southeast Asia Region using Agglomerative Hierarchical Clustering," *SISTEMASI*, vol. 12, no. 3, p. 900, Sep. 2023, doi: 10.32520/stmsi.v12i3.3150.
- [11] M. Bariyklana and A. Fauzan, "Implementation of the dbscan method for cluster mapping of earthquake spread location," *Barekeng: Jurnal Ilmu Matematika dan Terapan*, vol. 17, no. 2, pp. 867–878, 2023, doi: 10.30598/barekengvol17iss2pp0869-0880.
- [12] I. B. F. Arafat, M. A. Hariyadi, I. B. Santoso, and C. Crysdiyan, "Clustering Gempabumi di Wilayah Regional VII Menggunakan Pendekatan DBSCAN," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 4, pp. 823–830, Aug. 2023, doi: 10.25126/jtiik.2024106918.
- [13] R. B. Sippan and N. Setiyawati, "Pemetaan dan Klusterisasi Daerah Rawan Bencana Alam di Provinsi Sulawesi Tengah Menggunakan K-Means," *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 2, pp. 1031–1045, Mar. 2025, doi: 10.29100/jipi.v10i2.6161.
- [14] A. Wahyu and R. Rushendra, "Klusterisasi Dampak Bencana Gempa Bumi Menggunakan Algoritma K-Means di Pulau Jawa," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 8, no. 1, p. 174, Apr. 2022, doi: 10.26418/jp.v8i1.52260.
- [15] A. Jain, K. Rathi, Y. Ganguly, A. Kumar, and Y. Bhale, "A Comparative Analysis of DBSCAN, K-Means and Agglomerative Clustering Algorithms for Geospatial Data," in *Proceedings of the International Conference on Recent Advancement and Modernization in Sustainable Intelligent Technologies & Applications (RAMSITA-2025)*, vol. 192, Springer Nature, 2025, pp. 212–221. doi: 10.2991/978-94-6463-716-8\_18.
- [16] X. Shu and Y. Ye, "Knowledge Discovery: Methods from data mining and machine learning," *Soc Sci Res*, vol. 110, p. 102817, 2023, doi: <https://doi.org/10.1016/j.ssrresearch.2022.102817>.
- [17] F. Gullo, "From Patterns in Data to Knowledge Discovery: What Data Mining Can Do," *Phys Procedia*, vol. 62, pp. 18–22, 2015, doi: 10.1016/j.phpro.2015.02.005.



- [18] D. Puspitasari, A. J. Aprian, E. D. Sikumbang, K. Ramanda, S. H. Sukmana, and Q. N. Azizah, “Heart Disease: Application of the K-Nearest Neighbor (KNN) Method,” *Ingénierie des systèmes d information*, vol. 29, no. 4, pp. 1275–1281, Aug. 2024, doi: 10.18280/isi.290403.
- [19] Nuraeni, Martanto, A. R. Dikananda, and A. Rifai, “Optimization of Kebaya Product Grouping Using K-Means Algorithm for Marketing Strategy of Rental Services at Gifaattire Store,” *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 4, no. 3, pp. 1549–1555, Jun. 2025, doi: 10.59934/jaiea.v4i3.776.
- [20] L. Q. Rosydah and T. Widiyaningtyas, “Perbandingan Cosine Similarity dan Mean Squared Difference dalam Rekomendasi Buku Fiksi berbasis Item,” *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 2, pp. 565–574, Dec. 2024, doi: 10.29408/edumatic.v8i2.27783.
- [21] R. R. A. Rahman and A. W. Wijayanto, “Pengelompokan Data Gempa Bumi Menggunakan Algoritma Dbscan,” *Jurnal Meteorologi dan Geofisika*, vol. 22, no. 1, p. 31, Oct. 2021, doi: 10.31172/jmg.v22i1.738.
- [22] V. R. Prasetyo, G. Erlangga, and D. A. Prima, “Analisis Sentimen untuk Identifikasi Bantuan Korban Bencana Alam berdasarkan Data di Twitter Menggunakan Metode K-Means dan Naive Bayes,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 5, pp. 1055–1062, Oct. 2023, doi: 10.25126/jtiik.2023107077.
- [23] N. Dwitiyanti, Siti Ayu Kumala, and Shinta Dwi Handayani, “Comparative Study of Earthquake Clustering in Indonesia Using K-Medoids, K-Means, DBSCAN, Fuzzy C-Means and K-AP Algorithms,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 6, pp. 768–778, Dec. 2024, doi: 10.29207/resti.v8i6.5514.
- [24] R. Kusumastuti, E. Bayunanda, A. M. Rifa’i, M. R. G. Asgar, F. I. Ilmawati, and K. Kusrini, “Clustering Titik Panas Menggunakan Algoritma Agglomerative Hierarchical Clustering (AHC),” *CogITO Smart Journal*, vol. 8, no. 2, pp. 501–513, Dec. 2022, doi: 10.31154/cogito.v8i2.438.501-513.
- [25] E. Sulistiyawan, A. Hapsery, and L. J. A. Arifahanum, “Perbandingan Metode Optimasi Untuk Pengelompokan Provinsi Berdasarkan Sektor Perikanan Di Indonesia (Studi Kasus Dinas Kelautan dan Perikanan Indonesia),” *Jurnal Gaussian*, vol. 10, no. 1, pp. 76–84, 2021, doi: <https://doi.org/10.14710/j.gauss.10.1.76-84>.
- [26] “BMKG, Skala Intensitas Gempabumi (SIG).” Accessed: Jul. 31, 2025. [Online]. Available: <https://kalbar.bmkg.go.id/?page=skala-intensitas-gempabumi>