

# ANALISIS PENGELOMPOKAN GEMPA BUMI DI SUMATRA INDONESIA MENGUNAKAN METODE K-MEANS, DBSCAN, AGGLOMERATIVE CLUSTERING, DAN MEAN SHIFT

**Alfikri Dwi Mauluda<sup>1\*</sup>, Alvina Kusumadewi Kuncoro<sup>2</sup>, Sri Aningsih<sup>3</sup>**

<sup>1</sup> Program Studi Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Kebangsaan Republik Indonesia, Bandung, Kota, 40263, Indonesia

<sup>2</sup> Program Studi Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Bandung, Bandung, 40132, Indonesia

<sup>3</sup> Program Studi Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Medan, Medan, 20221, Indonesia

email: alfikridwimauluda@mipa.ukri.ac.id

## ABSTRAK

Indonesia merupakan salah satu negara dengan tingkat aktivitas seismik tertinggi di dunia, yang dipengaruhi oleh keberadaan tiga lempeng tektonik aktif. Pulau Sumatra khususnya berada di atas zona subduksi dan patahan aktif, sehingga memiliki kerentanan tinggi terhadap bencana gempa bumi. Untuk memahami pola distribusi spasial gempa bumi di wilayah ini, diperlukan analisis berbasis data yang mampu mengidentifikasi zona-zona rawan secara lebih akurat. Penelitian ini bertujuan untuk menganalisis pola sebaran gempa bumi di Sumatra, Indonesia dengan menerapkan empat metode pengelompokan (clustering), yaitu K-Means, DBSCAN, Agglomerative Clustering, dan Mean Shift. Data gempa bumi diperoleh dari United States Geological Survey (USGS) untuk periode tahun 2019 hingga 2024, dengan parameter yang digunakan meliputi lintang, bujur, kedalaman, dan magnitudo. Seluruh data distandarisasi dan dianalisis menggunakan pendekatan berbasis unsupervised learning. Hasil penelitian menunjukkan bahwa metode K-Means dan DBSCAN mampu mengelompokkan data gempa secara efektif berdasarkan parameter spasial dan seismik, dengan kluster yang cenderung padat dan tersebar sesuai struktur geologis. Sementara itu, metode Agglomerative Clustering dan Mean Shift menghasilkan pola yang lebih kompleks dan menyiratkan potensi adanya konsentrasi aktivitas seismik di area tertentu. Beberapa kluster berhasil mengidentifikasi zona seismik penting seperti patahan Sumatra dan zona subduksi Indo-Australia. Temuan ini dapat dimanfaatkan untuk mendukung strategi mitigasi bencana dan perencanaan pembangunan berbasis risiko di wilayah rawan gempa di Sumatra, Indonesia.

**Kata Kunci:** Pengelompokan; Gempa bumi; K-means; DBSCAN; Mean Shift

## ABSTRACT

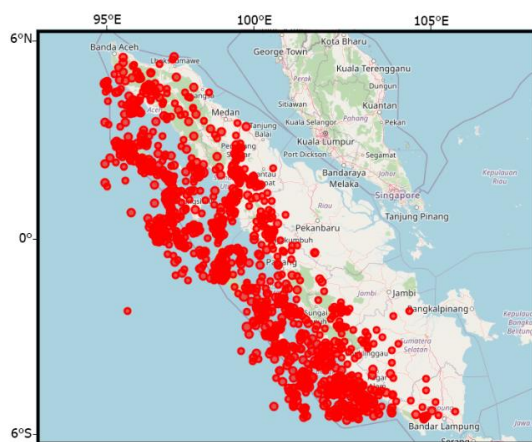
**[Title: Analysis Of Earthquake Clustering In Sumatra, Indonesia Using K-Means, DBSCAN, Agglomerative Clustering, and Mean Shift Methods]** Indonesia is one of the countries with the highest seismic activity in the world, influenced by the interaction of three active tectonic plates. Sumatra Island, in particular, lies above a subduction zone and active fault lines, making it highly vulnerable to earthquakes. To better understand the spatial distribution patterns of earthquakes in this region, data-driven analysis is essential to accurately identify seismically hazardous zones. This study aims to analyze earthquake distribution patterns in Sumatra, Indonesia by applying four clustering methods: K-Means, DBSCAN, Agglomerative Clustering, and Mean Shift. Earthquake data were obtained from the United States Geological Survey (USGS) for the period from 2019 to 2024, using parameters such as latitude, longitude, depth, and magnitude. All data were standardized and analyzed using an unsupervised learning approach. The results show that K-Means and DBSCAN effectively grouped the earthquake data based on spatial and seismic attributes, producing dense and distinct clusters that align with geological structures. Meanwhile, Agglomerative Clustering and Mean Shift produced more complex patterns, suggesting potential concentrations of seismic activity in specific areas. Several clusters successfully identified key seismic zones such as the Sumatra Fault and the Indo-Australian subduction zone. These findings can be utilized to support disaster mitigation strategies and risk-informed spatial planning in earthquake-prone areas across Sumatra, Indonesia.

**Keywords:** Clustering; Earthquake; K-means; DBSCAN; Mean Shift.

## PENDAHULUAN

Indonesia terletak di jalur cincin api dengan aktivitas kegempaan yang sangat tinggi di dunia, salah satunya area subduksi Sumatra (Gambar 1) (Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), 2018). Tingginya frekuensi gempa bumi di Indonesia mengakibatkan dibutuhkan analisis yang mendalam terhadap pola sebaran dan karakteristik gempa untuk meningkatkan upaya mitigasi bencana. Salah satu tantangan utama dalam analisis tersebut adalah mengelola dan memahami data gempa yang besar dan kompleks, yang terus bertambah setiap tahunnya seiring dengan kemajuan teknologi pencatatan seismik. Menurut data dari United States Geological Survey (USGS), Indonesia mengalami lebih dari 7.000 gempa bumi setiap tahun, dan sekitar 500 di antaranya memiliki magnitudo lebih besar dari 4.0 Mw, menjadikannya salah satu negara dengan frekuensi gempa tinggi secara global. Badan Geologi (PVMBG) juga mencatat bahwa sebagian besar gempa signifikan di Indonesia berasosiasi dengan sistem subduksi dan patahan aktif, seperti Patahan Sumatra yang memanjang sepanjang pulau.

Namun demikian, salah satu tantangan utama dalam analisis gempa modern adalah mengelola dan menafsirkan data gempa yang besar dan kompleks, yang terus bertambah seiring kemajuan teknologi pencatatan seismik. Oleh karena itu, dibutuhkan pendekatan komputasi dan metode analitik seperti *clustering* untuk mengidentifikasi pola tersembunyi dan mendukung pengambilan keputusan yang lebih akurat dalam mitigasi risiko gempa.



**Gambar 1.** Peta persebaran gempa di Sumatra, Indonesia tahun 2019-2024.

Dalam dekade terakhir, perkembangan metode komputasi telah memungkinkan pengolahan

data skala besar secara lebih efisien. Salah satu pendekatan populer adalah teknik pengelompokan (*clustering*), yang banyak digunakan untuk mengidentifikasi pola tersembunyi dalam data spasial. Metode pengelompokan seperti *K-Means*, *Density-Based Spatial Clustering of Applications with Noise (DBSCAN)*, *Agglomerative Clustering*, dan *Mean Shift* menawarkan berbagai keunggulan dalam mengelompokkan data gempa berdasarkan parameter geografis seperti lintang, bujur, kedalaman, dan magnitudo. Setiap metode memiliki keunggulan unik: *K-Means* efektif untuk pola sederhana dengan jumlah kluster yang telah ditentukan sebelumnya (Ramadhani, 2020). Metode ini bertujuan untuk menemukan pola dalam kumpulan data besar melalui proses pengelompokan yang sistematis (Han dkk., 2011).

Referensi dari berbagai penelitian sebelumnya mendukung pemilihan algoritma tersebut. Ramadhani (2020) dan Maulana dkk. (2021) menunjukkan bahwa *K-Means* efektif untuk pengelompokan data gempa dengan parameter geografis. Sementara itu, penelitian Dewi & Anggraini (2020) serta Priyono (2019) menyoroti keunggulan *DBSCAN* dalam mengidentifikasi pusat kluster dan kluster kompleks. Penelitian lain juga mengindikasikan keberhasilan *K-Means* dalam analisis zonasi daerah rawan bencana alam (Wicaksono & Susetyo, 2023) dan klasifikasi wilayah rawan gempa (Dwitiyanti dkk., 2023). Penggunaan *Agglomerative* dan *Mean Shift* untuk analisis pola gempa bumi juga telah dilaporkan dalam beberapa studi (Setiawan dkk, 2022; Wati dkk., 2024; Ramadhani, 2020; Aprianti dkk. (2023)).

Oleh karena itu, dengan adanya kebutuhan untuk memahami pola distribusi gempa bumi di Sumatra, Indonesia serta memanfaatkan keunggulan masing-masing metode pengelompokan, penelitian ini bertujuan untuk membandingkan efektivitas metode *K-Means*, *DBSCAN*, *Agglomerative Clustering*, dan *Mean Shift*.

## METODE

Penelitian ini menggunakan empat algoritma pengelompokan untuk menganalisis pola gempa bumi di Sumatra, Indonesia: *K-Means*, *Density-Based Spatial Clustering of Applications with Noise (DBSCAN)*, *pengklasteran Agglomerative*, dan *Mean Shift*. Setiap metode memiliki pendekatan unik: *K-Means* membagi data berdasarkan pusat kluster, *DBSCAN* menggunakan kepadatan data untuk mendeteksi kluster dan pencilan, *Agglomerative Clustering* membentuk tingkatan kluster berdasarkan kedekatan antar data, dan *Mean Shift* menemukan kluster secara dinamis dengan

estimasi kepadatan kernel. Implementasi dan hasil masing-masing metode akan dibahas untuk mengevaluasi keunggulan dan keterbatasannya dalam analisis gempa bumi.

Pada penelitian ini data yang digunakan berasal dari United States Geological Survey (USGS) untuk periode 2019 hingga 2024. Hasil dari studi ini diharapkan dapat memberikan wawasan baru yang bermanfaat dalam upaya mitigasi risiko gempa bumi di Indonesia.

**Metode K-Means**

Metode K-Means adalah salah satu algoritma pengelompokan yang paling umum digunakan karena kesederhanaannya dalam membagi data ke dalam kluster berdasarkan kedekatan pusat kluster (Ester dkk., 1996). Namun, metode ini memiliki kelemahan dalam mendeteksi kluster dengan bentuk tidak bulat atau jika terdapat pencilan.

K-Means merupakan pengelompokan data (dalam hal ini, data gempa bumi) yang dipartisi dengan masing-masing kelompok dihubungkan oleh sebuah pusat kluster atau titik pusat dan setiap titik ditempatkan ke dalam kelompok yang mempunyai pusat kluster terdekat (Rifa dkk., 2020). Langkah-langkah yang digunakan pada algoritma K-Means antara lain:

1. Menentukan kelompok k
2. Memilih pusat massa benda yang akan dikelompokkan sebanyak k
3. Menentukan jarak setiap benda terhadap setiap pusat massa dengan cara menghitung jarak menggunakan jarak Euclidean atau metrik lainnya.
4. Mengalokasikan setiap objek ke pusat kluster terdekat dan menentukan pusat kluster baru dengan menghitung rata-rata data pada setiap kelompok menggunakan rumus sebagai berikut:

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \tag{1}$$

dengan  $\mu_k$  adalah pusat kluster yang baru pada kelompok ke-k,  $N_k$  adalah jumlah data pada kelompok ke-k, dan  $x_q$  adalah data (titik gempa) ke-q pada kluster ke-k, dan

5. Ulangi langkah 3 sampai 4, hingga tidak ada perubahan pada pusat kluster.

**Density-Based Spatial Clustering of Applications with Noise (DBSCAN)**

DBSCAN mengelompokkan data berdasarkan kepadatan. Tidak seperti K-Means, DBSCAN tidak memerlukan jumlah kluster yang ditentukan sebelumnya. DBSCAN memiliki dua parameter utama: epsilon ( $\epsilon$ ) dan minPts (jumlah titik minimum dalam jangkauan  $\epsilon$ ).

Dalam konteks data **gempa bumi**, metode DBSCAN sangat berguna karena distribusi gempa di permukaan bumi tidak selalu membentuk pola geometris sederhana. Sebaliknya, gempa sering terkonsentrasi di sepanjang **patahan aktif**, yang dapat membentuk pola kluster **memanjang, tidak beraturan, atau berkepadatan tidak seragam**. Oleh karena itu, algoritma berbasis kepadatan seperti DBSCAN lebih tepat digunakan untuk mendeteksi pola spasial semacam ini.

Persamaan utama yang digunakan dalam DBSCAN adalah untuk menentukan apakah sebuah titik adalah *core point*, *border point*, atau *noise*. Dua kondisi berikut harus dipenuhi untuk sebuah titik untuk menjadi core point:

1. Jarak antar titik: Jika jarak antara dua titik (gempa)  $x_i$  dan  $x_j$  (2)

lebih kecil dari  $\epsilon$ , maka titik tersebut dikatakan saling terhubung. Dengan kata lain,

$$\|x_i - x_j\| \leq \epsilon. \tag{3}$$

2. Jumlah titik dalam radius  $\epsilon$ : Sebuah titik dianggap sebagai *core point* jika ada setidaknya minPts titik dalam jangkauan radius  $\epsilon$  (termasuk titik itu sendiri). Untuk data gempa di Sumatra, jika koordinat distandardisasi, nilai epsilon bisa berkisar antara **0.1 hingga 0.5** (dalam skala data standar), dan *minPts* biasanya antara **4–10** tergantung kepadatan data.

Setelah *core points* ditemukan, kluster terbentuk dengan menghubungkan titik-titik yang terhubung langsung atau secara transitif (titik yang terhubung dengan *core points* lainnya). Titik yang tidak memenuhi kriteria ini disebut noise.

**Pengklasteran Agglomerative**

Pengklasteran Agglomerative adalah metode hierarkis yang menggabungkan kluster berdasarkan

jarak antar kluster. Dalam metode ini, jarak antar kluster dapat dihitung dengan beberapa cara, seperti single linkage, complete linkage, atau average linkage (Kononenko & Kukar, 2007). Berikut adalah rumus yang digunakan untuk menghitung jarak antar kluster:

1. Single Linkage (jarak minimum):

$$d(C_i, C_j) = \min \|x_i - x_j\| \quad (4)$$

2. Complete Linkage (jarak maksimum):

$$d(C_i, C_j) = \max \|x_i - x_j\| \quad (5)$$

3. Average Linkage (jarak rata-rata):

$$d(C_i, C_j) = \frac{1}{|C_i||C_j|} \sum_{i \in C_i, x_j \in C_j} \|x_i - x_j\| \quad (6)$$

Dimana:

$C_i$  dan  $C_j$  adalah dua kluster yang berbeda.

$x_i$  dan  $x_j$  adalah titik data dalam kluster  $C_i$  dan  $C_j$ .

$\|x_i - x_j\|$  adalah jarak antara dua titik data.

Proses Pengklasteran Agglomerative menggabungkan kluster berdasarkan jarak yang terpendek atau terdekat (tergantung metode yang digunakan), hingga mencapai jumlah kluster yang diinginkan atau semua data tergabung dalam satu kluster besar.

### Pengklasteran Mean Shift

Pada Mean Shift, proses utamanya adalah memindahkan titik data menuju titik pusat massa di sekitar titik tersebut. Langkah iteratif ini dilakukan hingga titik tersebut mencapai posisi pusat kepadatan data yang lebih tinggi (Comaniciu & Meer, 2002) Proses ini dapat dijelaskan dengan persamaan berikut:

Iterasi Perpindahan Mean:

$$x_i^{(t)} = \frac{\sum_{x_j \in S_i} K(x_j - x_i^{(t)}) x_j}{\sum_{x_j \in S_i} K(x_j - x_i^{(t)})} \quad (7)$$

Dimana:

$x_i^{(t)}$  adalah posisi titik data  $x_i$  pada iterasi t.

$S_i$  adalah set titik data dalam radius tertentu dari  $x_i$ .

$K(x_j - x_i^{(t)})$  adalah fungsi kernel (Gaussian) yang menghitung bobot berdasarkan jarak antara titik  $x_i$  dan titik  $x_j^{(t)}$ .

Bentuk matematis dari fungsi kernel Gaussian adalah sebagai berikut:

$$K(x) = \exp\left(-\frac{\|x\|^2}{2h^2}\right) \quad (8)$$

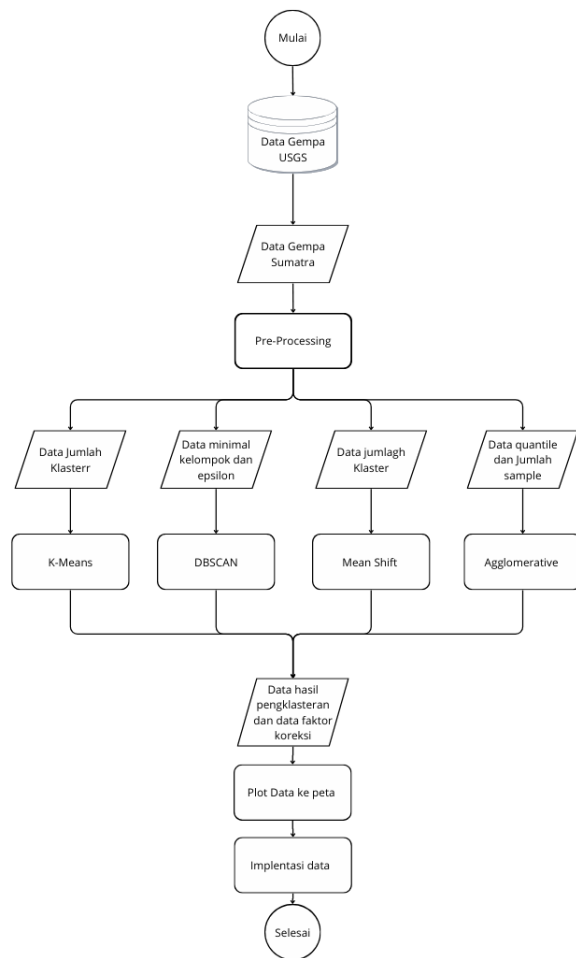
dengan:

- x: selisih vektor antara titik tetangga dan titik yang sedang diproses
- h: parameter bandwidth (radius pencarian)

Pusat Massa: Proses perpindahan dilakukan untuk setiap titik data, dengan menghitung pusat massa (mean) dari titik-titik di sekitar titik tersebut, menggunakan kernel sebagai pembobot.

Proses ini berlanjut hingga titik-titik tidak lagi bergerak atau hanya bergerak sedikit, yang menandakan bahwa mereka telah mencapai pusat kepadatan yang lebih tinggi.

Untuk melakukan pengolahan data, data diambil dari USGS dengan parameter lokasi geografis, kedalaman, dan magnitudo dengan tahapan pelaksanaan ditunjukkan pada Gambar 2.



**Gambar 2.** Digram alur pengolahan data gempa Sumatra 2019-2024.

1. Pengumpulan Data:

Data gempa diperoleh dari USGS Earthquake Catalog dengan format GeoJSON, mencakup periode 2019–2024. Untuk memastikan fokus pada wilayah Pulau Sumatra, data disaring berdasarkan batas koordinat geografis dengan rentang lintang – 11° hingga 7° LU dan bujur 90° hingga 110° BT. Rentang ini dipilih agar mencakup keseluruhan wilayah tektonik aktif di Sumatra, termasuk zona subduksi di sebelah barat pulau dan patahan besar Sumatra yang membentang sepanjang pulau dari utara ke selatan. Data tersebut disaring berdasarkan batas koordinat geografis Sumatra, Indonesia.

2. Preprocessing Data:

Data yang tidak relevan, seperti gempa dengan magnitudo kurang dari 4.0, tidak tersedia di USGS. Parameter lintang, bujur, kedalaman, dan

magnitudo digunakan untuk proses pengelompokan. Data distandardisasi menggunakan metode StandardScaler dari scikit-learn.

3. Evaluasi dan Perbandingan Hasil.

Hasil pengklasteran dari masing-masing metode dievaluasi berdasarkan keberhasilan dalam mengidentifikasi pola distribusi gempa dan zona seismik utama di Sumatra. Selain itu, dilakukan analisis terhadap jumlah klaster, pemisahan antar klaster, serta deteksi pusat klaster yang dapat memberikan wawasan lebih dalam mengenai pola gempa bumi di wilayah tersebut.

4. Implementasi Pengklasteran:

**K-Means:** Menentukan pusat klaster awal secara acak untuk membagi data menjadi tiga klaster. **DBSCAN:** Menentukan nilai epsilon ( $\epsilon$ ) menggunakan metode *nearest neighbor* untuk menemukan jarak optimal. Dalam pengolahan data ini kami menggunakan nilai epsilon ( $\epsilon$ ) = 2.4. **Pengklasteran Agglomerative:** Menggunakan pendekatan *bottom-up* untuk membentuk tiga klaster berdasarkan kedekatan jarak. **Mean Shift:** Menggunakan estimasi kepadatan kernel untuk menentukan jumlah klaster secara otomatis.

Selain itu, data gempa juga dianalisis berdasarkan kedalaman hiposentrum, yang dalam seismologi digunakan untuk mengklasifikasikan tipe gempa bumi. Berdasarkan klasifikasi standar internasional, gempa bumi dibagi menjadi tiga kategori: gempa dangkal (kedalaman < 70 km), gempa menengah (kedalaman antara 70–300 km), dan gempa dalam (kedalaman > 300 km). Pengelompokan berdasarkan kedalaman ini penting untuk memahami karakteristik sumber gempa, mekanisme pelepasan energi, serta potensi dampaknya di permukaan, karena gempa dangkal umumnya lebih merusak dibandingkan gempa dalam. Dengan mempertimbangkan baik dimensi spasial (lokasi geografis) maupun dimensi vertikal (kedalaman), hasil analisis klaster dalam penelitian ini menjadi lebih komprehensif dan relevan dalam konteks zonasi risiko seismik di wilayah Sumatra.

Parameter yang digunakan dalam proses pengelompokan meliputi lintang, bujur, kedalaman, dan magnitudo. Untuk mempertimbangkan skala yang berbeda antar parameter (misalnya, kedalaman dalam kilometer dan magnitudo dalam skala logaritmik), dilakukan proses standarisasi data menggunakan metode StandardScaler dari pustaka *scikit-learn*. *StandardScaler* bekerja dengan cara

mengubah setiap fitur (parameter) sehingga memiliki rata-rata 0 dan standar deviasi 1, menggunakan rumus berikut:

$$z = \frac{x - \mu}{\sigma} \tag{9}$$

di mana:

x: nilai asli

μ: rata-rata dari fitur

σ: standar deviasi dari fitur

Proses ini penting agar setiap parameter memiliki kontribusi yang setara dalam penghitungan jarak antar titik, khususnya dalam algoritma pengelompokan seperti *K-Means* dan *DBSCAN* yang sensitif terhadap skala data.

Silhouette Score mengukur seberapa baik suatu titik cocok dengan klasternya dibandingkan dengan klaster lain. Nilainya berkisar dari -1 hingga 1: nilai mendekati +1 menunjukkan pengelompokan yang baik, mendekati 0 menunjukkan batas antar klaster kurang jelas, dan nilai negatif menandakan kemungkinan data salah klaster. Semakin tinggi nilainya, semakin baik kualitas pengelompokan.

Untuk mengevaluasi kualitas hasil pengelompokan, digunakan Davies-Bouldin Index (DBI), yaitu metrik evaluasi internal yang mengukur seberapa baik klaster dipisahkan dan seberapa kompak anggota klaster. Davies-Bouldin Index didefinisikan sebagai:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left( \frac{S_i + S_j}{M_{ij}} \right) \tag{9}$$

dengan:

k: jumlah klaster

S<sub>i</sub>: rata-rata jarak antara tiap titik dalam klaster ke-*i* dan pusat klaster-nya (mengukur **kompaksi/dispersion**)

M<sub>ij</sub>: jarak antara pusat klaster *i* dan *j* (**jarak antar-klaster**)

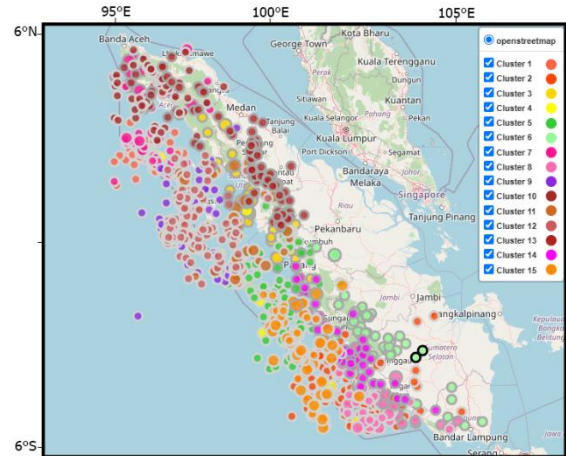
$\frac{S_i + S_j}{M_{ij}}$ : rasio antara kompaksi dan separasi dua klaster

Nilai DBI yang **lebih rendah** menunjukkan performa klaster yang lebih baik karena mencerminkan klaster yang **lebih kompak dan terpisah dengan jelas** dari klaster lain.

## HASIL DAN PEMBAHASAN

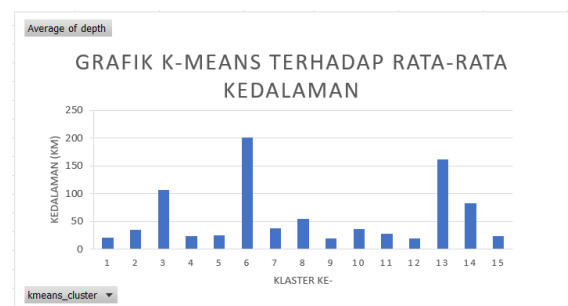
Visualisasi hasil pengelompokan menunjukkan bahwa setiap metode memiliki kemampuan berbeda dalam menangkap pola distribusi gempa bumi:

K-Means menghasilkan kelompok berbentuk bulat yang merata dan bergantung pada pusat klaster awal yang dipilih.



**Gambar 3.** Peta persebaran kelompok gempa metode K-Means di Sumatra, Indonesia tahun 2019-2024.

Metode ini efektif untuk dataset yang memiliki distribusi yang relatif rapat (homogen), tetapi kurang mampu menangani bentuk kelompok yang kompleks. Dari 15 klaster yang terbentuk 14 data distribusinya lebih terpusat, sedangkan satu data lebih menyebar (klaster 15) (Gambar 3).

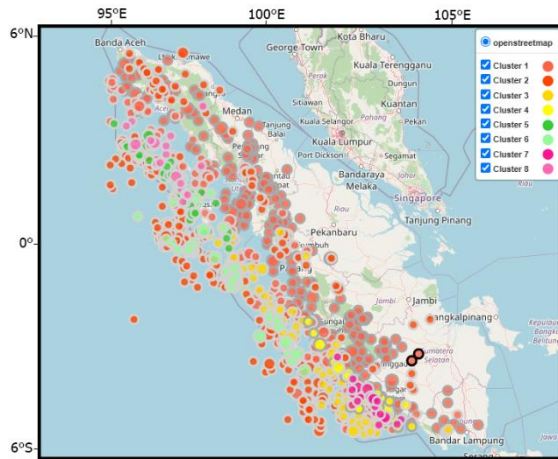


**Gambar 4.** Grafik K-Means terhadap rata-rata kedalaman

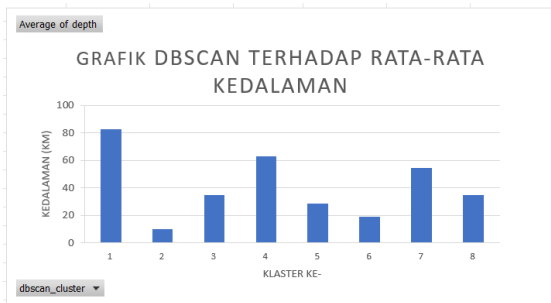
Hasil pengelompokan menggunakan metode **K-Means** menunjukkan bahwa kedalaman rata-rata gempa bervariasi antar klaster, dengan nilai terdangkal pada **Cluster 9** (18,78 km) dan terdalam pada **Cluster 6** (201,37 km). Rata-rata keseluruhan kedalaman gempa sebesar **48,81 km** mengindikasikan dominasi gempa dangkal hingga

menengah. Variasi ini mencerminkan efektivitas K-Means dalam memisahkan kejadian gempa berdasarkan kedalaman hiposentrum di wilayah Sumatra.

DBSCAN mengidentifikasi gangguan dan pusat klaster dengan lebih baik, terutama pada daerah dengan kepadatan rendah.



**Gambar 5.** Peta persebaran kelompok gempa metode DBSCAN di Sumatra, Indonesia tahun 2019-2024.

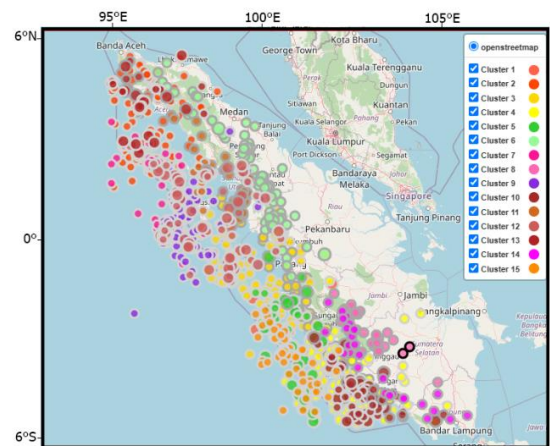


**Gambar 6.** Grafik DBSCAN terhadap rata-rata kedalaman

Berdasarkan hasil pengelompokan menggunakan metode DBSCAN, kedalaman rata-rata gempa untuk setiap klaster menunjukkan variasi yang cukup jelas. Klaster 2 memiliki rata-rata kedalaman paling dangkal, yaitu 9,99 km, disusul oleh Klaster 6 (19,07 km) dan Klaster 5 (28,44 km), yang mencerminkan dominasi gempa-gempa dangkal. Di sisi lain, Klaster 1 memiliki kedalaman rata-rata tertinggi, yakni 82,82 km, menunjukkan konsentrasi gempa menengah hingga dalam. Rata-rata keseluruhan dari seluruh klaster adalah 48,81 km, yang secara umum mengarah pada kategori gempa dangkal-menengah. Hasil ini menunjukkan bahwa DBSCAN mampu mengelompokkan kejadian gempa berdasarkan kedekatan spasial dan densitas

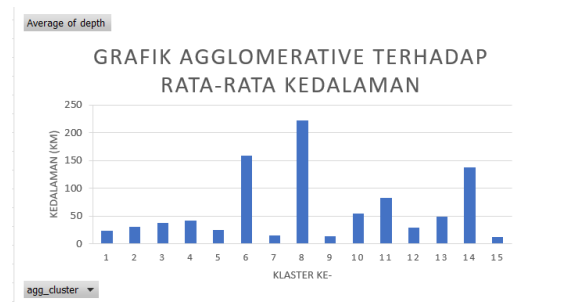
data, sekaligus memisahkan karakteristik kedalamannya secara efektif.

Metode ini mampu mendeteksi kelompok dengan bentuk yang tidak beraturan, namun hasilnya sangat bergantung pada pemilihan parameter epsilon ( $\epsilon$ ) dan jumlah minimum titik dalam kelompok. Klaster yang di dapatkan cenderung tersebar melintang pada arah barat laut-tenggara dari lokasi, dan dikelompokkan berdasarkan data kedalaman (Gambar 7). Pengklasteran Agglomerative memberikan struktur hierarkis yang memfasilitasi analisis multi-level.



**Gambar 7.** Peta persebaran kelompok gempa metode Pengklasteran Agglomerative di Sumatra, Indonesia tahun 2019-2024.

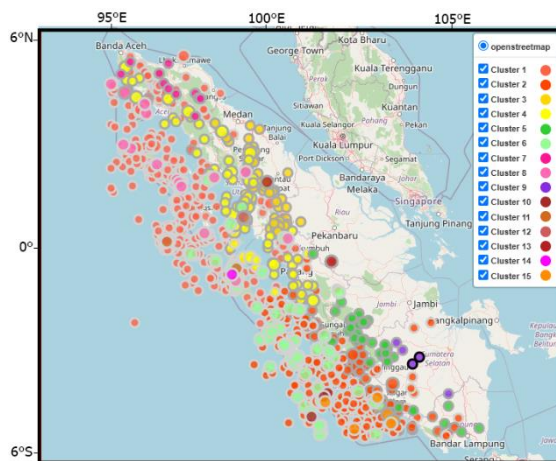
Hasil dari pengelompokan Agglomerative hampir mirip dengan K-means, namun hasil pengelompokan dari Agglomerative berdasarkan tetangga dan tidak terpusat. Sehingga data klaster yang dihasilkan lebih menyebar dibandingkan K-means. Hal ini membantu melakukan interpretasi yang lebih fleksibel terkait hubungan antar kelompok. (Gambar 5).



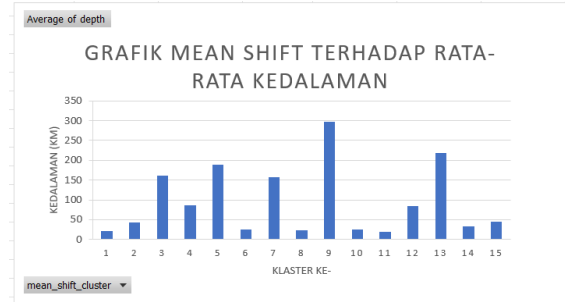
**Gambar 8.** Grafik Agglomerative terhadap rata-rata kedalaman

Berdasarkan hasil pengelompokan menggunakan metode Agglomerative Clustering, terlihat bahwa rata-rata kedalaman gempa bervariasi cukup signifikan antar kluster. Kluster dengan kedalaman rata-rata paling dangkal adalah Cluster 15 (12,16 km), Cluster 9 (13,55 km), dan Cluster 7 (15,67 km), yang mengindikasikan dominasi gempa dangkal pada kluster tersebut. Sementara itu, Cluster 8 mencatatkan rata-rata kedalaman tertinggi, yaitu 222,68 km, diikuti oleh Cluster 6 (158,53 km) dan Cluster 14 (137,88 km), yang menunjukkan konsentrasi gempa dalam. Nilai rata-rata keseluruhan kedalaman untuk seluruh kluster adalah 48,81 km, yang termasuk dalam kategori gempa dangkal hingga menengah. Pola ini menguatkan bahwa metode agglomerative berhasil mengelompokkan gempa berdasarkan kedalaman hiposentrum dengan cukup baik, yang penting untuk mendukung interpretasi zona-zona seismik di wilayah Sumatra.

*Mean Shift* secara adaptif membentuk kelompok dengan pusat kepadatan tertinggi. Kluster yang dihasilkan dari metode ini cenderung mengelompokkan hasil berdasarkan besarnya nilai parameter (magnitudo), tidak terlalu bergantung dengan kerapatan data yang dikelompokkan (Gambar 9). Metode ini tidak memerlukan asumsi jumlah kelompok sebelumnya, tetapi dapat menjadi lambat untuk dataset besar dengan distribusi yang kompleks.



**Gambar 9.** Peta persebaran kelompok gempa metode Mean Shift di Sumatra, Indonesia tahun 2019-2024.



**Gambar 10.** Grafik *Means Shift* terhadap rata-rata kedalaman

Berdasarkan hasil perhitungan rata-rata kedalaman gempa pada setiap kluster (*Means Shift*), diketahui bahwa kedalaman gempa sangat bervariasi antar kluster. Kluster dengan kedalaman rata-rata terdangkal terdapat pada Cluster 11 (19,5 km), Cluster 1 (21,49 km), dan Cluster 10 (24,33 km), yang kemungkinan besar merepresentasikan gempa dangkal. Sementara itu, Cluster 9 menunjukkan kedalaman rata-rata tertinggi, yaitu 297,84 km, diikuti oleh Cluster 13 (217,63 km) dan Cluster 5 (189,18 km), yang mengindikasikan dominasi gempa dalam. Perbedaan ini menunjukkan bahwa klusterisasi berhasil memisahkan kejadian gempa berdasarkan kedalaman hiposentrum secara signifikan, yang dapat membantu dalam identifikasi zona seismik aktif baik di permukaan maupun di kedalaman litosfer Sumatra. Rata-rata keseluruhan kedalaman gempa dari seluruh kluster adalah sekitar 48,81 km, yang termasuk kategori gempa dangkal-menengah.

**Analisis Kinerja Metode Pengklusteran.**

Untuk mengevaluasi kinerja setiap metode pengelompokan, digunakan dua indikator utama: *Silhouette Score* dan *Davies-Bouldin Index*. Berikut adalah hasil evaluasi untuk masing-masing metode:

**Tabel 1.** Tabel evaluasi kinerja setiap metode-metode pengelompokan untuk data gempa Sumatra, Indonesia.

Metode	SILHOUETTE SCORE	Davies-Bouldin Index
K-Means	0.262427	1.158211
DBSCAN	N/A	N/A
Agglomerative	0.258579	1.108437
Mean Shift	0.311806	0.863011

*Silhouette Score* mengukur seberapa baik objek di dalam sebuah kelompok dibandingkan dengan objek

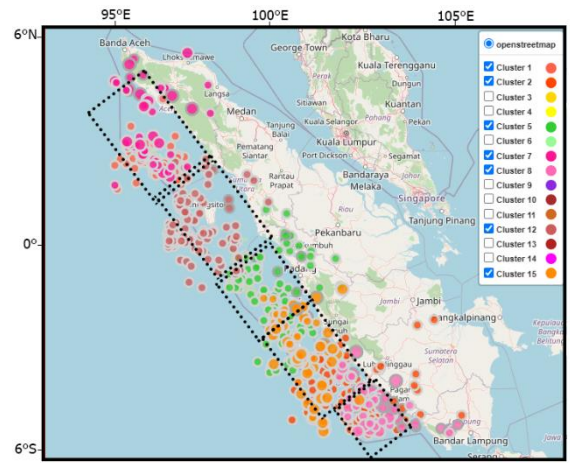
di kelompok lain. Nilai *Silhouette Score* yang lebih tinggi menunjukkan hasil pengelompokan yang lebih baik. Berdasarkan hasil, *Mean Shift* memiliki nilai *Silhouette Score* tertinggi (0.311806) (tabel 1). Hal ini menunjukkan bahwa metode *Mean Shift* memberikan pengelompokan yang paling baik di antara metode yang digunakan. K-Means dan Pengklasteran Agglomerative memiliki nilai *Silhouette Score* yang cukup dekat (0.262427 dan 0.258579), namun lebih rendah dibandingkan *Mean Shift*. DBSCAN tidak dapat dihitung karena metode ini mengidentifikasi beberapa data sebagai gangguan, yang tidak membentuk kelompok yang jelas untuk evaluasi menggunakan *Silhouette Score*.

*Davies-Bouldin Index* mengukur kualitas pengelompokan dengan membandingkan rasio antara jarak rata-rata antar kelompok dengan ukuran kelompok. Nilai yang lebih rendah menunjukkan performa pengelompokan yang lebih baik. Dari tabel, terlihat bahwa *Mean Shift* memiliki *Davies-Bouldin Index* terendah (0.863011). Hal ini menunjukkan bahwa pengelompokan yang dihasilkan lebih baik dibandingkan dengan metode lainnya. Pengklasteran Agglomerative dan K-Means memiliki nilai *Davies-Bouldin Index* yang lebih tinggi (1.108437 dan 1.158211), menunjukkan bahwa meskipun metode ini memberikan hasil yang dapat diterima, mereka kurang optimal dalam memisahkan kelompok dengan jarak yang jelas.

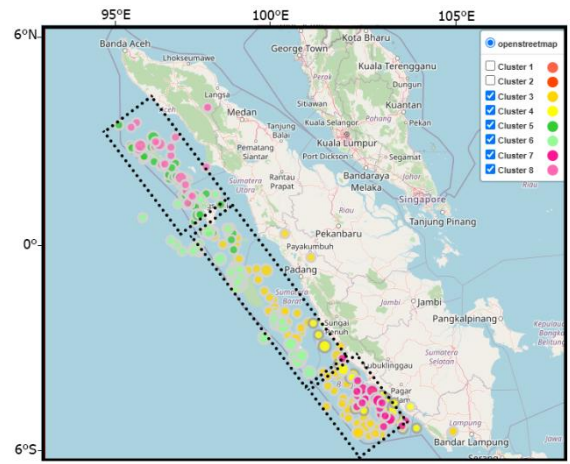
**Distribusi Spasial dan Pola Seismik**

Hasil dari analisis pengelompokan menunjukkan adanya pola sebaran gempa bumi yang terkait dengan segmentasi zona seismik di Sumatra, Indonesia. Setiap metode pengelompokan memberikan informasi yang berbeda mengenai pola distribusi spasial gempa bumi. Klaster yang dihasilkan mendekati pola segmentasi area kegempaan di zona subduksi Sumatra dari Tim Pusat Studi Gempa Nasional (2017). Berdasarkan hasil pengelompokan, teridentifikasi beberapa daerah segmentasi yang mempengaruhi distribusi gempa bumi di wilayah Sumatra, Indonesia.

Zona subduksi Sumatra merupakan wilayah dengan aktivitas seismik paling tinggi dan terlihat jelas dalam hasil pengelompokan. Sebagian besar gempa dengan kedalaman dangkal hingga menengah dan magnitudo besar terpusat di sepanjang zona subduksi ini. Klaster gempa yang dihasilkan oleh metode K-Means terdiri dari setidaknya 5 segmentasi di area sisi barat pulau Sumatra.



(a)



(b)

**Gambar 11.** Beberapa pola klaster dari metode (a) K-Means dan (b) DBSCAN yang mendekati pola segmentasi seismik di zona subduksi Sumatra.

Klaster 1 dan 11 (Gambar 11a) berada di area segmentasi gempa Aceh 2004, klaster 12 di area gempa Nias 2005, dan klaster 5 berada di area kepulauan Siberut. Klaster 15 mengelompokkan data gempa di area pulau Pagai dimana gempa pernah terjadi di tahun 2007 dan 2010. Data gempa lainnya terkumpul dalam klaster 8 di sisi barat pulau Sumatra paling selatan atau area pulau Enggano dimana gempa pernah terjadi pada tahun 2000. Untuk di area pulau (darat), pengelompokan distribusi gempa yang dihasilkan terdiri dari 2 klaster yaitu klaster 6 dan 13. Lokasi klaster ini bertepatan dengan area Sesar Sumatra yang melintang dari ujung utara hingga ujung selatan pulau Sumatra dimana gempa pernah terjadi di Padang pada tahun 2007 dan 2022 dengan magnitudo 6.4 dan 6.1, serta Aceh pada tahun 2013 dan 2016 dengan magnitudo 6.1 dan 6.5 (USGS).

Di sisi lain, metode DBSCAN menghasilkan 3 area segmentasi (sisi barat pulau Sumatra) dan cenderung menghasilkan pola pengelompokan berbentuk memanjang dan melintang pada arah Barat Laut-Tenggara (Gambar 7b). Dalam penggunaan metode DBSCAN, diperlukan pemilihan parameter yang lebih baik untuk mencapai hasil yang optimal. Pengelompokan berdasarkan posisi,

terdistribusi di wilayah samudra dan pulau (Gambar 4).

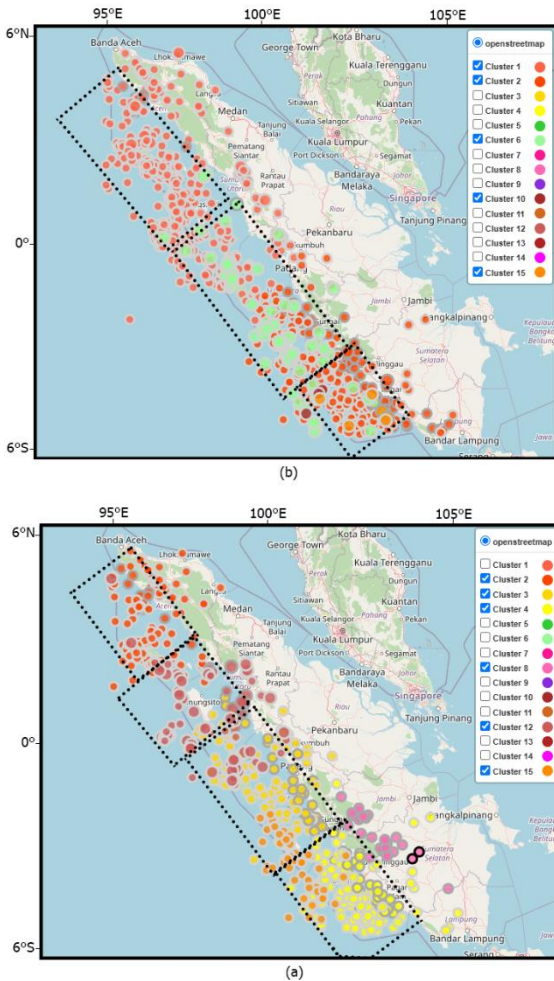
Hasil pola segmentasi dari metode Agglomerative menghasilkan pola kluster yang mirip dengan metode K-Means. Akan tetapi, pola segmentasi terdiri dari 4 area yaitu segmen Aceh, Nias, Pulau Siberut dan Pagai (Kepulauan Mentawai), dan Enggano (Gambar 8). Untuk area Sesar Sumatra, dihasilkan 4 kluster: kluster 6, 8, 13, dan 14 (Gambar 8).

Selain itu, hasil pengelompokan menggunakan metode Agglomerative Clustering di area daratan mendukung temuan Widyantoro & Natawidjaja (2007) yang menyatakan bahwa segmen-segmen Patahan Sumatra memiliki pola seismik yang kompleks dengan frekuensi gempa kecil yang tinggi.

Di zona subduksi Sumatra, pola distribusi gempa yang dihasilkan oleh Metode *Mean Shift* lebih tersebar (Gambar 8b) dan menghasilkan 3 area segmentasi. Metode ini juga mampu membentuk mengelompokkan gempa dengan magnitudo besar dengan adanya kluster 10, 11, 12, 13, 14, dan 15 (Gambar 8) walaupun dengan jumlah data yang sedikit dimana umumnya pengelompokan dilakukan berdasarkan distribusi data yang besar dan padat. Selain itu, kelompok di area Sesar Sumatra yang sebagian besar terdiri dari gempa dengan magnitudo rendah dapat teridentifikasi di dalam kluster 3, 5, 7, 9, dan 13 (Gambar 8). Wilayah ini cenderung memiliki kepadatan gempa yang lebih rendah dibandingkan dengan zona subduksi dan patahan. Wilayah ini cenderung memiliki kepadatan gempa yang lebih rendah dibandingkan dengan zona subduksi dan patahan utama.

Keterbatasan data dari USGS yang hanya mencatat gempa dengan magnitudo di atas 4,0 menyebabkan jumlah gempa bermagnitudo kecil yang teridentifikasi relatif sedikit. Meskipun demikian, gempa dengan magnitudo kecil tetap memiliki potensi mempengaruhi stabilitas struktural dan geologi di wilayah tersebut, terutama jika terjadi secara berulang atau dalam waktu yang berdekatan. Selain itu, gempa-gempa dengan magnitudo rendah di area darat berkaitan dengan aktivitas tektonik akibat patahan regional (Natawidjaja, 2007).

Secara keseluruhan, hasil pengelompokan ini menunjukkan bahwa masing-masing metode memiliki keunggulannya dalam mengungkapkan pola-pola seismik yang berbeda, sesuai dengan karakteristik masing-masing zona seismik (Setiawan



**Gambar 12.** Beberapa pola kluster dari metode (a) Agglomerative dan (b) Mean Shift yang mendekati pola segmentasi seismik di zona subduksi Sumatra.

kedalaman, dan magnitudo gempa yang besar di zona ini berhubungan dengan arah perbatasan zona subduksi Sumatra, dimana lempeng Indo-Australia tersubduksi ke bawah lempeng Eurasia dari arah Barat Daya ke Timur Laut. Akan tetapi, DBSCAN tidak mampu melakukan pengelompokan segmen kegempaan di area daratan (Sesar Sumatra). Gempa dalam kluster 1 yang dihasilkan oleh metode ini

& Hadi, 2021). Pemetaan pola spasial gempa bumi ini memberikan pemahaman yang lebih mendalam tentang dinamika seismik di Sumatra dan sangat penting untuk merancang strategi mitigasi bencana yang lebih efektif, terutama di zona yang rawan terhadap gempa besar dan sering terjadi.

### KESIMPULAN DAN SARAN

Penelitian ini telah berhasil menerapkan empat metode pengelompokan untuk menganalisis pola sebaran gempa bumi di Sumatra, Indonesia. Berdasarkan hasil analisis, dapat disimpulkan bahwa masing-masing metode pengelompokan memberikan temuan yang signifikan dalam memahami pola seismik di wilayah tersebut.

K-Means dan DBSCAN efektif dalam mengelompokkan data gempa berdasarkan parameter geografis, kedalaman, dan magnitudo, Metode K-Means ini efektif untuk dataset yang memiliki distribusi yang relatif rapat (homogen), tetapi kurang mampu menangani bentuk kelompok yang kompleks. Dari 15 kluster yang terbentuk 14 data distribusinya lebih terpusat, sedangkan satu data lebih menyebar.

Metode ini mampu mendeteksi kelompok dengan bentuk yang tidak beraturan, namun hasilnya sangat bergantung pada pemilihan parameter epsilon ( $\epsilon$ ) dan jumlah minimum titik dalam kelompok. Kluster kegempaan yang didapatkan cenderung tersebar melintang pada arah barat laut-tenggara dari lokasi, dan dikelompokkan berdasarkan data kedalaman gempa.

Hasil dari pengelompokan Agglomerative hampir mirip dengan K-means, namun hasil pengelompokan dari Agglomerative berdasarkan tetangga dan tidak terpusat. Sehingga data kluster yang dihasilkan lebih menyebar dibandingkan K-means. Hal ini membantu melakukan interpretasi yang lebih fleksibel terkait hubungan antar kelompok. Namun, metode ini memiliki kelemahan dalam komputasi jika dataset sangat besar.

*Mean Shift* memberikan hasil yang adaptif Kluster yang dihasilkan dari metode ini cenderung mengelompokkan hasil berdasarkan besarnya nilai parameter (magnitudo), tidak terlalu bergantung dengan kerapatan data yang dikelompokkan. Metode ini tidak memerlukan asumsi jumlah kelompok sebelumnya, tetapi dapat menjadi lambat untuk dataset besar dengan distribusi yang kompleks.

Hasil pengelompokan ini berhasil mengidentifikasi beberapa zona seismik utama di

Sumatra, termasuk di area zona subduksi dan area Sesar Sumatra. Masing-masing metode menghasilkan kluster yang mendekati pola segmentasi kegempaan dengan jumlah kluster yang berbeda-beda. Temuan ini berpotensi menjadi dasar dalam perencanaan strategi mitigasi bencana, khususnya di daerah rawan gempa besar dan kecil. Secara keseluruhan, analisis pengelompokan ini memberikan kontribusi penting dalam memahami dinamika seismik di Sumatra, Indonesia dan dapat mendukung kebijakan mitigasi bencana yang lebih tepat sasaran.

### DAFTAR PUSTAKA

- Aprianti, A., Jufriansah, A., Donuata, P., Khusnani, A., & Ayuba, J. (2023). Comparison of K-Means Algorithm and DBSCAN on Aftershock Activity in the Flores Sea: Seismic Activity 2019-2022. *Journal of Novel Engineering Science and Technology*, 2, 77–82. <https://doi.org/10.56741/jnest.v2i03.393>
- Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). (2018). Memahami Potensi, Mitigasi, dan Sistem Informasi Gempabumi Tsunami Di Indonesia, BMKG KMS, 2 Februari 2018.
- Badan Meteorologi, Klimatologi, dan Geofisika. (n.d.). *Data gempabumi*. <https://data.bmkg.go.id/gempabumi/>
- Comaniciu, D., & Meer, P. (2002). *Mean shift: A robust approach toward feature space analysis*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 24(5), 603–619.
- Dewi, S., & Anggraini, S. (2020). Analisis clustering data gempa menggunakan DBSCAN. *Jurnal Geomatika*, 5.
- Dwitiyanti, N., Ayu Kumala, S., & Dwi Handayani, S. (2023). *PENERAPAN METODE K-MEANS PADA KLASERISASI WILAYAH RAWAN GEMPA DI INDONESIA Implementation of K-Means Method in Clasterization of Earthquake Prone Areas in Indonesia*.
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *In Proceedings of the 2nd International*

- Conference on Knowledge Discovery and Data Mining*. [www.aaai.org](http://www.aaai.org)
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques 3rd Edition Solution Manual*. Morgan Kaufmann.
- Jain, A. (2010). Data Clustering: 50 Years Beyond K-Means. *Pattern Recognition Letters*, 31, 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- Kononenko, I., & Kukar, M. (2007). Machine learning and data mining. *Machine Learning and Data Mining*. <https://doi.org/10.1533/9780857099440>
- Lestari, N. I. D., & Madlazim. (2022). Analisis seismisitas dan potensi bahaya bencana seismik Pulau Sumatera berdasarkan data gempa 1970–2020. *Jurnal Inovasi Fisika Indonesia (IFI)*, 11(2), 12–19. <https://doi.org/10.26740/ifi.v11n2.p12-19>
- Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 129–137. <https://doi.org/10.1109/TIT.1982.1056489>
- Maulana, D., Yuniarto, M., & Kusuma, T. (2021). Implementasi DBSCAN dalam identifikasi outlier data gempa bumi. *Jurnal Sistem Informasi*, 18(4).
- Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.
- Natawidjaja, D. H. (2007). The Sumatran Fault Zone—from Source to Hazard. *Journal of Earthquake and Tsunami*.
- Priyono, P. (2019). Hierarchical clustering untuk data spasial gempa. *Jurnal Statistika*, 7(2).
- Ramadhani, R. . (2020). Penerapan K-Means untuk analisis gempa di wilayah Sumatera Barat. *Jurnal Informatika*, 14(2).
- Rifa, I. H., Pratiwi, H., & Respatiwan, R. (2020). Clustering of earthquake risk in Indonesia using K-medoids and K-means algorithms. *Media Statistika*, .
- Setiawan, B., & Hadi, L. (2021). Clustering techniques for earthquake analysis in Indonesia: A case study of Sumatra region. . *Geophysics Journal International*.
- Setiawan, B., & Hadi, L. (2022). Penerapan Mean Shift untuk analisis kepadatan data gempa bumi. . *Jurnal Teknologi Informasi dan Komputer*, 9(1).
- Tim Pusat Studi Gempa Nasional. (2017). *PETA SUMBER DAN BAHAYA GEMPA INDONESIA TAHUN 2017*.
- Wati, R. K., Pratiwi, H., & Winita Sulandari. (2024). Perbandingan Algoritma Density-Based Spatial Clustering Algorithm with Noise (DBSCAN) dan Self-Organizing Map (SOM) untuk Clustering Data Gempa Bumi. *Statistika*, 24(2). <https://doi.org/10.29313/statistika.v24i2.3645>
- Wicaksono, D., & Susetyo, Y. (2023). CLUSTERING ZONASI DAERAH RAWAN BENCANA ALAM DI PROVINSI SUMATERA BARAT MENGGUNAKAN ALGORITMA K-MEANS DAN LIBRARY GEOPANDAS. *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, 4, 426–438. <https://doi.org/10.35870/jimik.v4i2.225>
- Widyantoro, S., & Natawidjaja, D. H. (2007). Subduction zone and earthquake characteristics in Sumatra, Indonesia: Evidence from seismological studies. *Journal of Geophysical Research*.