



Contents lists available at [Journal ELORA](#)

JRTI (Jurnal Riset Tindakan Indonesia)

ISSN: 2502-079X (Print) ISSN: 2503-1619 (Electronic)

Journal homepage: <https://jrti.eloracenter.org/jrti>



Klasterisasi spatio-temporal komparatif untuk deteksi gempa bumi bermagnitudo tinggi: analisis K-means dan DBSCAN terhadap seismisitas zona subduksi

Dimas Dwi Randa¹, Andri Dermawan^{*2}

¹ Program Studi Sains Data, Institut Teknologi Sumatera, Lampung Selatan, Indonesia

² Program Studi Pendidikan Teknologi Kejuruan, Fakultas Teknik, Universitas Negeri Padang, Sumatera Barat, Indonesia

Article Info

Article history:

Received Aug 21th, 2026

Revised Sept 14th, 2026

Accepted Oct 10th, 2026

Keyword:

Clustering spatio-temporal,
DBSCAN,
Gempa bumi,
K-means,
Sumatera

ABSTRAK

Sumatera merupakan wilayah dengan aktivitas seismik tinggi akibat interaksi Lempeng Indo-Australia dan Eurasia serta keberadaan Sesar Besar Sumatera. Penelitian ini bertujuan mengidentifikasi pola spasial-temporal kegempaan Sumatera menggunakan algoritma K-Means dan DBSCAN. Data diperoleh dari katalog United States Geological Survey (USGS) periode 2015–2025. Setelah praproses dan penerapan ambang kelengkapan magnitudo $M_c = 4,0$, sebanyak 2.834 kejadian gempa dianalisis berdasarkan lintang, bujur, waktu, kedalaman, dan magnitudo yang telah distandardisasi. K-Means menghasilkan enam kluster, sedangkan DBSCAN menghasilkan tiga kluster dengan 633 kejadian (22,3%) dikategorikan sebagai noise. Seluruh 27 gempa bermagnitudo $M \geq 6,0$ teridentifikasi sebagai noise oleh DBSCAN, dengan magnitudo titik noise secara signifikan lebih tinggi daripada titik terkluster ($p < 0,001$). K-Means memperoleh nilai Silhouette lebih tinggi (0,260), sedangkan DBSCAN memiliki nilai Davies–Bouldin Index lebih rendah (0,890). Hasil ini menunjukkan bahwa K-Means efektif untuk mengidentifikasi segmentasi zona seismik, sementara DBSCAN berpotensi digunakan untuk mendeteksi kejadian gempa anomali. Kedua metode saling melengkapi dalam analisis pola kegempaan Sumatera.



© 2026 The Authors.

This is an open access article under the CC BY-NC-SA license

(<https://creativecommons.org/licenses/by-nc-sa/4.0>)

Corresponding Author:

Andri Dermawan,

Program Studi Pendidikan Teknologi Kejuruan, Fakultas Teknik, Universitas Negeri Padang

Email: andridermawan@unp.ac.id

Pendahuluan

Pulau Sumatera berada pada zona pertemuan Lempeng Indo-Australia yang menunjam ke bawah Lempeng Eurasia dengan laju konvergensi sekitar 50–60 mm per tahun. Penunjaman ini bersifat miring (oblique) sehingga komponen gesernya terakomodasi oleh Sesar Besar Sumatera (Great Sumatran Fault) yang membentang sepanjang lebih dari 1.900 km di sepanjang punggung pulau. Konfigurasi tektonik ganda tersebut menjadikan Sumatera sebagai salah satu wilayah paling aktif secara seismik di dunia, sebagaimana dibuktikan oleh rangkaian gempa besar seperti Aceh–Andaman 2004 (Mw 9,1), Nias 2005 (Mw 8,6), Bengkulu 2007 (Mw 8,5), dan Mentawai 2010 (Mw 7,8). Katalog gempa bumi modern menghimpun ribuan kejadian setiap dekade dengan atribut ruang, waktu, kedalaman, dan magnitudo secara simultan. Volume dan dimensi data seperti ini sulit diinterpretasikan hanya dengan pendekatan deskriptif konvensional. Analisis kluster (clustering) sebagai teknik eksplorasi data tanpa label (unsupervised learning) menawarkan cara sistematis untuk menemukan struktur tersembunyi dalam katalog seismik, misalnya pengelompokan segmen sumber gempa, identifikasi zona Benioff, maupun deteksi rangkaian gempa susulan.

Dua algoritma clustering yang paling banyak digunakan dalam literatur memiliki asumsi dasar yang berbeda secara fundamental. K-Means bekerja dengan mempartisi seluruh objek ke dalam sejumlah k kluster berbentuk cembung dengan meminimumkan jumlah kuadrat galat dalam kluster; metode ini mengharuskan k ditentukan di awal dan tidak mengenal konsep pencilon. Sebaliknya, DBSCAN mengelompokkan objek berdasarkan kepadatan lokal, mampu menemukan kluster berbentuk sembarang, tidak memerlukan penentuan jumlah kluster, serta secara eksplisit menandai objek berkepadatan rendah sebagai noise. Perbedaan paradigma ini menimbulkan pertanyaan penelitian: struktur kegempaan seperti apa yang terungkap oleh masing-masing pendekatan, dan sejauh mana keduanya saling melengkapi ketika diterapkan pada katalog gempa Sumatera. Sebagian besar penelitian clustering kegempaan di Indonesia masih menggunakan fitur spasial saja (lintang dan bujur) atau menerapkan satu algoritma tunggal tanpa pembandingan. Penelitian ini berkontribusi pada tiga hal. Pertama, dimensi waktu diperlakukan sebagai fitur clustering yang setara dengan dimensi ruang sehingga pola spatio-temporal dapat terungkap secara langsung. Kedua, penentuan parameter kedua algoritma dilakukan secara terbuka dan dapat direproduksi melalui Elbow Method, Silhouette Score, Davies Bouldin Index, grafik k -distance, dan penelusuran grid. Ketiga, hasil kedua algoritma tidak sekadar dibandingkan secara metrik, tetapi juga diverifikasi maknanya secara geofisika melalui uji statistik nonparametrik dan perhitungan nilai b Gutenberg–Richter per segmen tektonik.

Tujuan penelitian ini adalah (1) mengidentifikasi pola pengelompokan spatio-temporal gempa Sumatera menggunakan K-Means dan DBSCAN; (2) menentukan parameter optimal kedua algoritma secara objektif; dan (3) mengevaluasi keunggulan relatif serta sifat komplementer kedua metode dalam konteks karakterisasi bahaya seismik.

Metode

Data yang digunakan merupakan katalog gempa bumi yang diunduh dari United States Geological Survey (USGS) Earthquake Catalog. Katalog awal memuat 17.924 rekaman dengan 22 atribut, mencakup periode 30 November 2000 hingga 2 Desember 2025. Atribut yang dimanfaatkan dalam penelitian ini adalah waktu kejadian (origin time), lintang, bujur, kedalaman hiposenter, magnitudo, tipe magnitudo, dan nama lokasi. Untuk memastikan homogenitas periode pengamatan dan kualitas penentuan hiposenter yang relatif seragam, analisis dibatasi pada sepuluh tahun terakhir, yaitu Januari 2015 sampai Desember 2025.

Praproses dilakukan melalui enam tahap berurutan. Pertama, seleksi jenis kejadian sehingga hanya rekaman bertipe earthquake yang dipertahankan. Kedua, pembatasan temporal pada rentang 2015–2025 yang menyisakan 4.642 rekaman. Ketiga, pembatasan spasial pada kotak $6,5^{\circ}\text{LU} - 6,5^{\circ}\text{LS}$ dan $94^{\circ} - 107^{\circ}\text{BT}$ agar cakupan konsisten dengan wilayah busur Sumatera; tahap ini menghilangkan kejadian di Kepulauan Andaman dan Jawa bagian barat sehingga tersisa 2.838 rekaman. Keempat, pemeriksaan nilai hilang dan duplikasi berdasarkan pengenalan unik kejadian; tidak ditemukan nilai hilang maupun duplikat pada atribut inti. Kelima, penerapan ambang kelengkapan magnitudo (magnitude of completeness) $M_c = 4,0$ untuk menghindari bias ketidaklengkapan katalog pada magnitudo kecil, menyisakan 2.834 kejadian sebagai data akhir. Keenam, rekayasa fitur temporal berupa konversi waktu kejadian menjadi variabel kontinu t (dalam satuan hari sejak kejadian pertama pada 2 Januari 2015), dengan rentang total 3.986,8 hari. Lima fitur digunakan sebagai masukan clustering, yaitu lintang, bujur, t , kedalaman, dan magnitudo. Karena kelima fitur memiliki satuan dan rentang yang sangat berbeda, seluruhnya distandardisasi menggunakan transformasi z -score sebagaimana Persamaan (1), agar tidak ada fitur yang mendominasi perhitungan jarak Euclidean.

$$z_{ij} = (x_{ij} - \mu_j) / \sigma_j \quad (1)$$

K-Means mempartisi n objek ke dalam k kluster dengan meminimumkan jumlah kuadrat galat dalam kluster (within-cluster sum of squares, SSE) seperti pada Persamaan (2), dengan μ_i menyatakan centroid kluster ke- i .

$$\text{SSE} = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (2)$$

Inisialisasi centroid menggunakan skema k -means++ dengan 50 kali pengulangan acak ($n_{\text{init}} = 50$), batas iterasi 500, dan $\text{random_state} = 42$ untuk menjamin reproduktibilitas. Nilai k diuji pada rentang 2 hingga 12. Titik siku (knee) pada kurva Elbow ditentukan secara objektif menggunakan metode jarak tegak lurus maksimum terhadap garis yang menghubungkan kedua ujung kurva SSE, bukan melalui penilaian visual.

DBSCAN memerlukan dua parameter, yaitu radius ketetanggaan ϵ dan jumlah minimum titik minPts . Sebuah titik disebut core point apabila memiliki sekurang-kurangnya minPts titik lain dalam radius ϵ . Kluster terbentuk dari himpunan titik yang terhubung secara kepadatan (density-connected), sedangkan titik yang tidak memenuhi kriteria tersebut dan tidak berada dalam jangkauan core point mana pun ditandai sebagai noise. Penentuan ϵ diawali dengan grafik k -distance, yaitu kurva jarak terurut setiap titik terhadap tetangga terdekat ke- k dengan $k = 2d = 10$ (d adalah jumlah dimensi fitur). Karena knee kurva k -distance memberikan $\epsilon = 1,13$

yang terlalu permisif dan menghasilkan kluster tunggal, penentuan parameter dilanjutkan dengan penelusuran grid pada $\epsilon \in [0,30; 1,00]$ dengan langkah 0,05 dan minPts $\in \{5, 10, 15, 20, 25\}$. Kriteria seleksi ditetapkan sebelum eksperimen dijalankan, yaitu: (a) jumlah kluster sekurang-kurangnya tiga; (b) rasio noise berada pada rentang 5–30% yang dipandang wajar untuk katalog seismik; dan (c) di antara kombinasi yang memenuhi (a) dan (b), dipilih yang memiliki Silhouette Score tertinggi.

Kualitas hasil clustering dievaluasi menggunakan tiga metrik validasi internal, yaitu Silhouette Score (rentang –1 sampai 1, makin besar makin baik), Davies–Bouldin Index (makin kecil makin baik), dan Calinski–Harabasz Index (makin besar makin baik). Untuk DBSCAN, metrik dihitung dengan mengeluarkan titik noise agar tidak menghasilkan nilai yang menyesatkan. Tingkat kesesuaian kedua metode diukur dengan Adjusted Rand Index (ARI) dan Normalized Mutual Information (NMI). Verifikasi makna geofisika dilakukan melalui tiga uji tambahan. Pertama, uji Mann–Whitney U satu sisi untuk menguji hipotesis bahwa magnitudo titik noise DBSCAN lebih besar daripada magnitudo titik terkuster. Kedua, uji Kruskal–Wallis untuk menguji perbedaan distribusi kedalaman dan magnitudo antarkluster K-Means. Ketiga, perhitungan nilai b Gutenberg–Richter dengan metode maximum likelihood (Persamaan 3), dengan M sebagai magnitudo rerata, $M_c = 4,0$, dan $\Delta M = 0,1$.

$$b = \log_{10}(e) / [M - (M_c - \Delta M/2)] \quad (3)$$

Perlu ditegaskan bahwa nilai b tidak dihitung per kluster K-Means sebagai hasil utama, karena magnitudo merupakan salah satu fitur clustering sehingga kluster sebagian terdefinisi oleh magnitudo dan nilai b yang dihasilkan menjadi bias. Nilai b utama karenanya dihitung per segmen tektonik berdasarkan partisi lintang dan per kelas kedalaman, yang keduanya independen terhadap magnitudo. Seluruh analisis dilaksanakan menggunakan Python 3.11 dengan pustaka pandas, NumPy, scikit-learn, SciPy, dan Matplotlib.

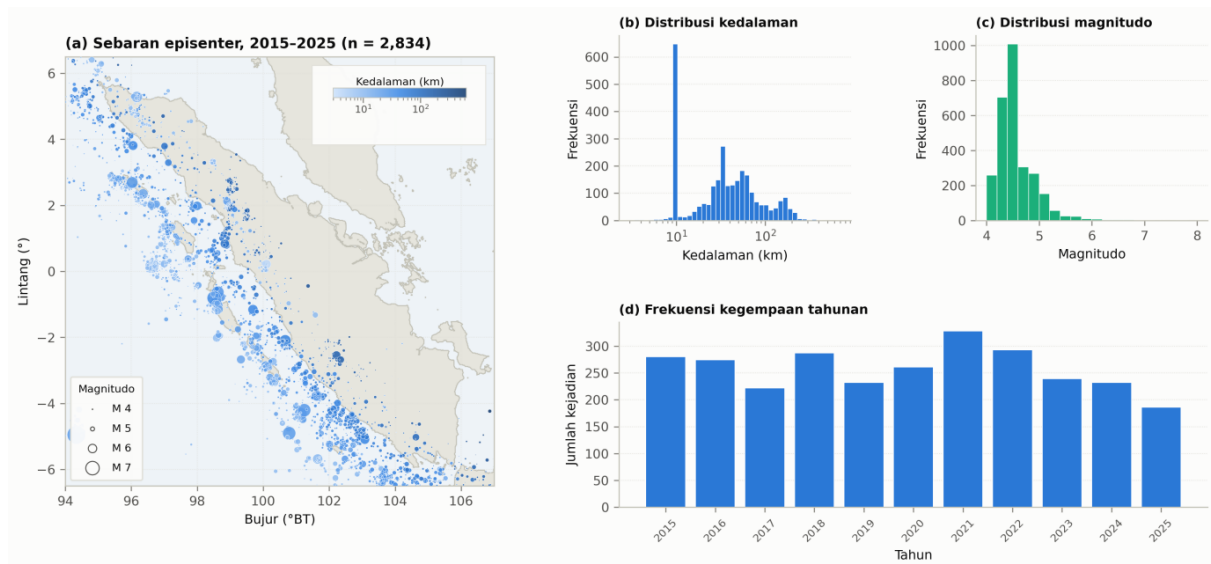
Hasil dan Pembahasan

Data akhir memuat 2.834 kejadian gempa dengan statistik deskriptif sebagaimana disajikan pada Tabel 1. Magnitudo berkisar antara 4,0 hingga 7,8 dengan rerata 4,55 dan simpangan baku 0,39, menunjukkan dominasi gempa berkekuatan ringan. Kedalaman hiposenter memiliki rentang yang jauh lebih lebar, yaitu 3,6 hingga 640,2 km dengan median 35,0 km, mencerminkan keberadaan zona penunjaman yang menerus dari permukaan hingga mantel atas.

Tabel 1. Statistik deskriptif data gempa Sumatera 2015–2025 ($n = 2.834$)

Variabel	Rerata	Simp. Baku	Min	Kuartil 1	Median	Kuartil 3	Maks
Lintang (°)	–0,835	3,576	–6,497	–4,315	–0,443	2,039	6,477
Bujur (°BT)	99,370	2,999	94,004	96,801	99,128	101,930	106,962
Kedalaman (km)	51,212	50,728	3,604	13,963	35,000	62,248	640,210
Magnitudo	4,554	0,394	4,000	4,300	4,500	4,700	7,800

Gambar 1 memperlihatkan sebaran episenter yang membentuk sabuk memanjang berarah barat laut–tenggara sejajar dengan Palung Sunda, sepenuhnya konsisten dengan geometri penunjaman Lempeng Indo-Australia. Konsentrasi kejadian terlihat paling rapat di lepas pantai barat Simeulue–Nias dan di lepas pantai Bengkulu. Berdasarkan kelas kedalaman, 2.082 kejadian (73,5%) tergolong gempa dangkal (< 60 km), 745 kejadian (26,3%) tergolong menengah (60–300 km), dan hanya 7 kejadian (0,2%) tergolong dalam (> 300 km). Distribusi magnitudo pada Gambar 1(c) menunjukkan bentuk peluruhan eksponensial yang khas bagi hukum Gutenberg–Richter, sedangkan frekuensi tahunan pada Gambar 1(d) relatif stabil pada kisaran 186–328 kejadian per tahun tanpa tren monotonik yang jelas. Satu hal yang perlu dicatat pada Gambar 1(b) adalah adanya lonjakan frekuensi yang tajam pada kedalaman tepat 10 km. Lonjakan ini merupakan artefak katalog, yakni nilai kedalaman baku yang diberikan USGS ketika kedalaman sebenarnya tidak dapat dibatasi secara memadai oleh data gelombang. Artefak ini disadari dan dibahas kembali pada bagian keterbatasan penelitian.



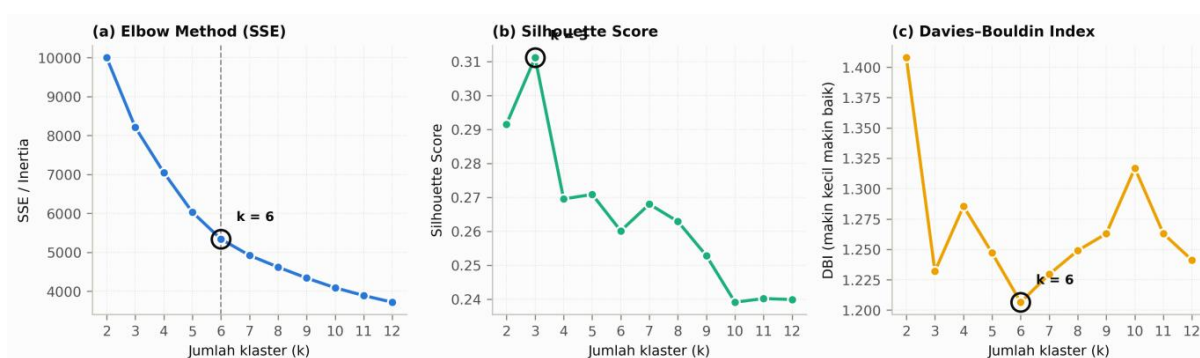
Gambar 1. Karakteristik umum kegempaan Sumatera 2015–2025: (a) sebaran episenter dengan warna menyatakan kedalaman dan ukuran menyatakan magnitudo; (b) distribusi kedalaman; (c) distribusi magnitudo; (d) frekuensi kegempaan tahunan.

Hasil pengujian k pada rentang 2 hingga 12 disajikan pada Tabel 2 dan divisualisasikan pada Gambar 2. Kurva SSE menurun secara monoton dari 10.001,4 pada $k = 2$ menjadi 3.717,6 pada $k = 12$. Perhitungan jarak tegak lurus maksimum menempatkan titik siku pada $k = 6$. Davies–Bouldin Index mencapai nilai minimum 1,2065 juga pada $k = 6$, sehingga kedua kriteria konvergen pada nilai yang sama.

Tabel 2. Metrik validasi internal K-Means untuk $k = 2$ sampai 12

k	SSE	Silhouette	Davies–Bouldin	Calinski–Harabasz
2	10.001,4	0,2915	1,4079	1.180,4
3	8.212,2	0,3111	1,2320	1.026,9
4	7.048,5	0,2695	1,2855	953,1
5	6.031,0	0,2708	1,2471	954,5
6	5.339,7	0,2601	1,2065	935,3
7	4.921,3	0,2680	1,2295	885,5
8	4.620,4	0,2629	1,2490	834,4
9	4.339,5	0,2528	1,2629	800,0
10	4.088,9	0,2391	1,3169	773,6
11	3.889,6	0,2402	1,2629	746,1
12	3.717,6	0,2399	1,2412	721,3

Silhouette Score tertinggi justru dicapai pada $k = 3$ (0,3111), berbeda dari kedua kriteria lainnya. Perbedaan ini dapat dijelaskan: Silhouette cenderung memberi nilai tinggi pada solusi dengan sedikit kluster yang saling berjauhan, sementara struktur kegempaan Sumatera bersifat kontinu di sepanjang busur sehingga tidak memiliki batas antarkluster yang tegas. Karena dua dari tiga kriteria menunjuk $k = 6$, dan $k = 6$ memberikan resolusi tektonik yang jauh lebih informatif dibandingkan $k = 3$, maka $k = 6$ dipilih sebagai solusi akhir. Nilai Silhouette pada $k = 6$ sebesar 0,260 memang tergolong moderat; hal ini lazim ditemui pada data seismik yang bersifat menerus dan bukan indikasi kegagalan algoritma.

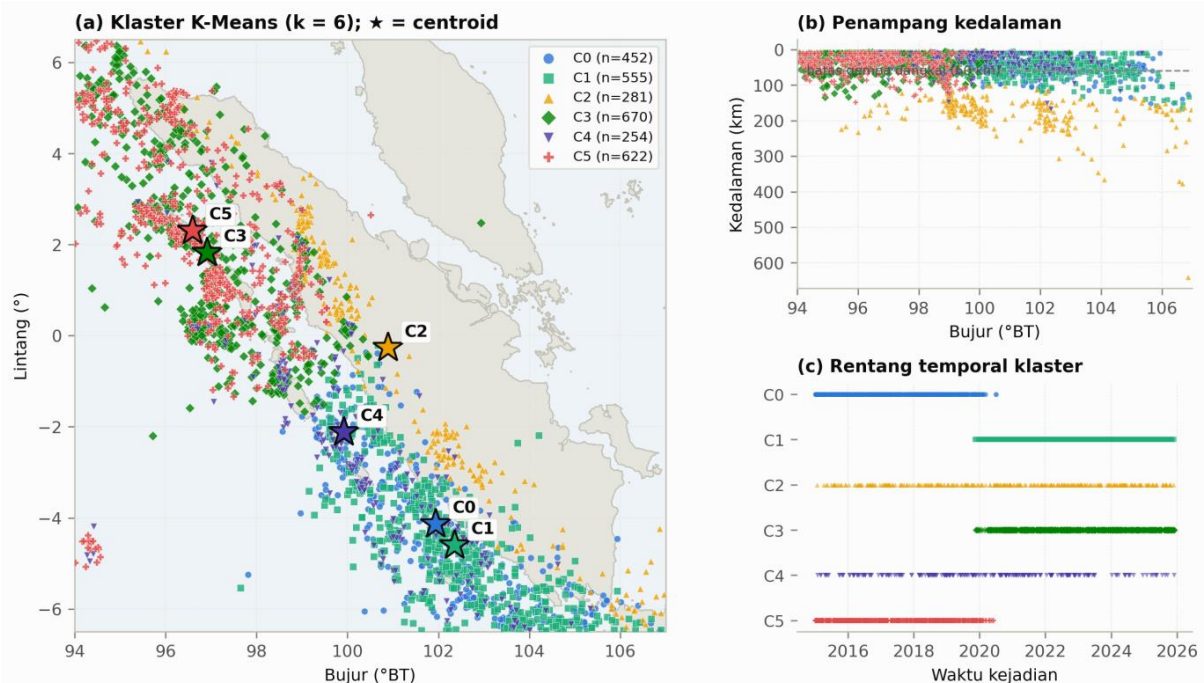


Gambar 2. Penentuan jumlah kluster optimal K-Means: (a) Elbow Method dengan titik siku pada $k = 6$; (b) Silhouette Score dengan nilai tertinggi pada $k = 3$; (c) Davies–Bouldin Index dengan nilai minimum pada $k = 6$.

Hasil K-Means dengan $k = 6$ disajikan pada Gambar 3 dan Tabel 3. Struktur yang terbentuk memperlihatkan pola yang sangat teratur dan dapat ditafsirkan secara tektonik. Empat kluster pertama membentuk skema partisi dua arah, yaitu pembagian geografis utara–selatan yang masing-masing terbelah menjadi dua periode waktu. Kluster C5 ($n = 622$) dan C3 ($n = 670$) menempati segmen utara di sekitar Simeulue, Singkil, Meulaboh, dan Banda Aceh dengan kedalaman rerata dangkal (36,7 km dan 29,0 km). Keduanya terpisah semata-mata oleh dimensi waktu: C5 mencakup 2015–2020 sedangkan C3 mencakup 2019–2025. Pola serupa terjadi di segmen selatan, di mana C0 ($n = 452$, 2015–2020) dan C1 ($n = 555$, 2019–2025) sama-sama berpusat di lepas pantai Bengkulu–Lampung dengan kedalaman rerata masing-masing 44,8 km dan 45,9 km. Pemisahan temporal ini bukan indikasi perubahan rezim tektonik, melainkan konsekuensi langsung dari dimasukkannya waktu sebagai fitur dengan bobot setara: karena kegempaan tersebar merata sepanjang periode pengamatan, K-Means membelah sumbu waktu pada titik tengahnya. Temuan ini justru bernilai metodologis, yakni menunjukkan bahwa pada wilayah dengan laju kegempaan stasioner, penambahan dimensi waktu menghasilkan replikasi kluster spasial, bukan penemuan pola baru. Dua kluster sisanya bersifat berbeda secara kualitatif dan tidak mengikuti pembelahan temporal. Kluster C2 ($n = 281$) memiliki kedalaman rerata 173,7 km dengan kedalaman maksimum 640,2 km dan berpusat di daratan Sumatera bagian tengah (Padangsidempuan, Parapat, Sibolga). Kluster ini jelas merepresentasikan zona Benioff, yaitu bidang penunjaman lempeng yang menyelam ke bawah daratan; posisi episenternya yang berada di darat sementara kedalamannya besar merupakan ciri khas geometri slab. Kluster C4 ($n = 254$) memiliki magnitudo rerata 5,40 jauh di atas rerata keseluruhan 4,55 dengan magnitudo maksimum 7,8, dan tersebar luas di sepanjang busur tanpa pemusatan spasial yang tegas. C4 dengan demikian merupakan kluster yang terdefinisi oleh magnitudo, bukan oleh ruang atau waktu.

Tabel 3. Karakteristik enam kluster hasil K-Means

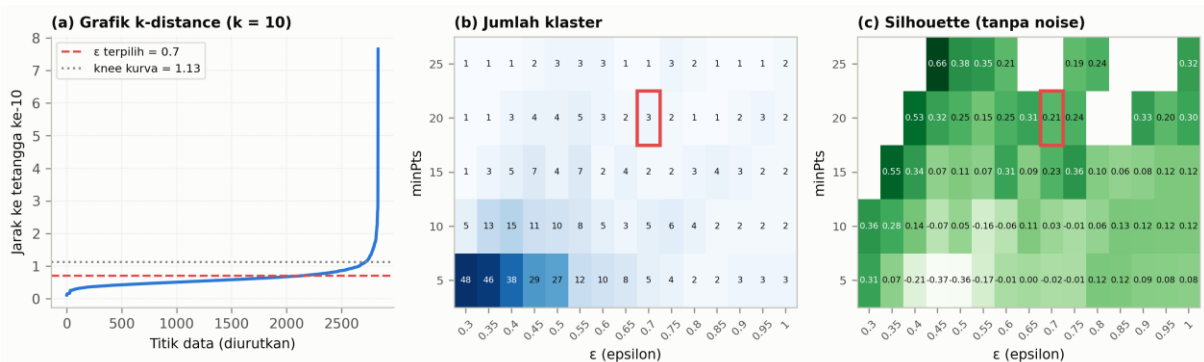
Kluster	n	%	Lintang rerata (°)	Bujur rerata (°BT)	Kedalaman rerata (km)	Magnitudo rerata	Periode	Wilayah dominan
C0	452	15,95	−4,124	101,934	44,77	4,487	2015–2020	Bengkulu, Sungai Penuh
C1	555	19,58	−4,602	102,353	45,90	4,521	2019–2025	Bengkulu, Pagar Alam
C2	281	9,92	−0,263	100,886	173,74	4,333	2015–2025	Padangsidempuan, Parapat
C3	670	23,64	1,827	96,911	28,99	4,477	2019–2025	Sinabang, Meulaboh
C4	254	8,96	−2,111	99,917	32,86	5,398	2015–2025	tersebar (magnitudo tinggi)
C5	622	21,95	2,309	96,586	36,71	4,469	2015–2020	Sinabang, Singkil



Gambar 3. Hasil clustering K-Means ($k = 6$)

Gambar 3(c) memperlihatkan pembelahan temporal tersebut secara visual: C5 dan C0 menempati paruh pertama periode pengamatan, C3 dan C1 menempati paruh kedua, sedangkan C2 dan C4 membentang penuh sepanjang 2015–2025. Laju kegempaan tertinggi tercatat pada C5 (114,7 kejadian per tahun) dan C3 (111,1 kejadian per tahun) di segmen utara, sedangkan C4 dan C2 menunjukkan laju terendah (23,4 dan 25,9 kejadian per tahun) yang wajar mengingat keduanya memuat kejadian yang secara alami lebih jarang.

Grafik k-distance pada Gambar 4(a) memperlihatkan kurva yang sangat landai hingga sekitar titik ke-2.500 sebelum menanjak tajam. Titik knee kurva berada pada $\epsilon = 1,13$; namun pengujian menunjukkan nilai tersebut terlalu longgar sehingga seluruh data menyatu menjadi satu kluster tunggal. Kelandaian kurva ini sendiri merupakan informasi penting: kepadatan titik dalam ruang fitur lima dimensi bersifat nyaris seragam di sepanjang busur, sehingga tidak terdapat batas kepadatan alami yang tegas. Penelusuran grid pada 75 kombinasi parameter (Gambar 4b dan 4c) menghasilkan pola yang jelas: kombinasi dengan Silhouette tinggi pada ϵ kecil selalu disertai rasio noise ekstrem di atas 80%, yang berarti algoritma hanya menemukan beberapa gugusan kecil dan membuang sebagian besar data. Setelah kriteria seleksi diterapkan, kombinasi $\epsilon = 0,7$ dan minPts = 20 terpilih dengan 3 kluster, rasio noise 22,3%, Silhouette 0,214, dan Davies–Bouldin 0,890.



Gambar 4. Penentuan parameter DBSCAN

Struktur kluster DBSCAN (Tabel 4 dan Gambar 5) berbeda secara mendasar dari K-Means. Kluster D0 menyerap 2.131 kejadian atau 75,2% dari seluruh data, membentang penuh dari $6,5^{\circ}$ LS hingga $6,4^{\circ}$ LU dengan kedalaman rerata 33,9 km dan magnitudo maksimum hanya 5,3. D0 dengan demikian merepresentasikan sabuk

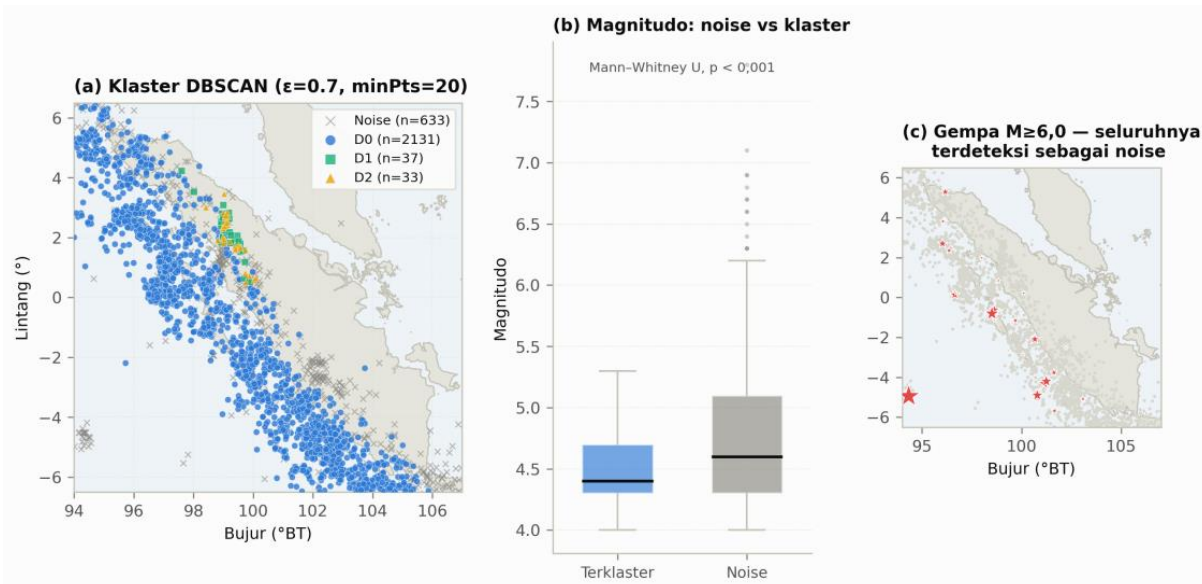
kegempaan latar (background seismicity) pada zona megathrust dangkal yang terhubung secara kepadatan sepanjang busur. Dua kluster kecil, D1 ($n = 37$) dan D2 ($n = 33$), menempati lokasi yang hampir identik di sekitar 2°LU dan $99,2^{\circ}\text{BT}$ dengan kedalaman rerata 157,6 km dan 159,8 km. Keduanya merupakan sarang kegempaan (seismic nest) pada zona Benioff di bawah kawasan Danau Toba, dan sekali lagi terpisah hanya oleh dimensi waktu (D1 pada 2015–2021, D2 pada 2021–2025).

Tabel 4. Karakteristik kluster hasil DBSCAN ($\epsilon = 0,7$; $\text{minPts} = 20$)

Kluster	n	%	Lintang rerata ($^{\circ}$)	Bujur rerata ($^{\circ}\text{BT}$)	Kedalaman rerata (km)	Magnitudo rerata	Magnitudo maks	Periode
D0	2.131	75,19	-0,980	99,199	33,94	4,501	5,3	2015–2025
D1	37	1,31	2,139	99,175	157,57	4,308	4,6	2015–2021
D2	33	1,16	1,981	99,268	159,79	4,288	4,5	2021–2025
Noise	633	22,34	-0,670	99,963	97,48	4,760	7,8	2015–2025

Sumber: hasil olahan penulis.

Temuan paling penting dalam penelitian ini terletak pada karakter titik noise DBSCAN. Sebanyak 633 kejadian (22,34%) ditandai sebagai noise, dan kelompok ini memiliki magnitudo rerata 4,760 dibandingkan 4,494 pada kelompok terkluster. Uji Mann–Whitney U satu sisi mengonfirmasi bahwa perbedaan tersebut signifikan secara statistik ($U = 839.732$; $p = 1,11 \times 10^{-15}$), dengan ukuran efek rank-biserial sebesar 0,205. Yang jauh lebih menonjol, seluruh 27 gempa bermagnitudo $M \geq 6,0$ dalam periode pengamatan termasuk gempa Mw 7,8 di barat daya Sumatera pada 2 Maret 2016 dan Mw 7,1 di tenggara Teluk Dalam pada 24 April 2023 diklasifikasikan sebagai noise oleh DBSCAN, atau 100% dari seluruh kejadian kuat. Hasil ini memiliki penjelasan yang koheren secara fisis. Gempa besar bersifat jarang menurut hukum Gutenberg–Richter, sehingga dalam ruang fitur terstandarisasi yang menyertakan magnitudo, kejadian tersebut menempati wilayah berkepadatan sangat rendah. DBSCAN, yang secara desain memisahkan wilayah berkepadatan rendah, otomatis menandainya sebagai pencilan. Kemampuan ini tidak dimiliki K-Means, yang justru memaksa setiap kejadian besar masuk ke dalam salah satu kluster — pada kasus ini terutama ke C4.



Gambar 5. Hasil clustering DBSCAN

Implikasi praktisnya, DBSCAN dapat difungsikan bukan sebagai alat segmentasi melainkan sebagai penyaring otomatis kejadian anomali dalam katalog seismik. Dalam konteks mitigasi bencana, kemampuan menyaring 27 kejadian berpotensi merusak dari 2.834 kejadian secara otomatis dan tanpa ambang magnitudo yang ditetapkan manusia merupakan keluaran yang bernilai operasional.

Perbandingan metrik disajikan pada Tabel 5 dan Gambar 8. K-Means unggul pada Silhouette Score (0,260 berbanding 0,214), sedangkan DBSCAN unggul pada Davies–Bouldin Index (0,890 berbanding 1,207).

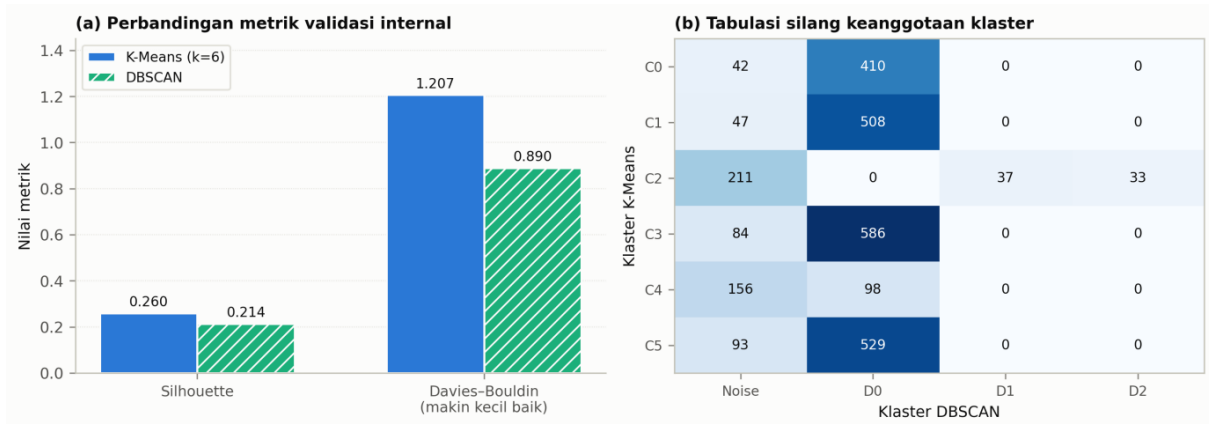
Keunggulan yang saling bertukar ini menegaskan bahwa tidak ada metode yang superior secara mutlak; keduanya mengukur aspek kualitas yang berbeda.

Tabel 5. Perbandingan kinerja dan karakteristik K-Means dan DBSCAN

Aspek	K-Means	DBSCAN
Parameter utama	$k = 6$	$\varepsilon = 0,7$; minPts = 20
Dasar penentuan parameter	Elbow + Davies–Bouldin	k-distance + grid search
Jumlah kluster	6	3
Silhouette Score	0,260	0,214
Davies–Bouldin Index	1,207	0,890
Calinski–Harabasz Index	935,3	–
Titik noise	tidak ada (0%)	633 (22,34%)
Kluster terbesar	670 (23,64%)	2.131 (75,19%)
Deteksi gempa $M \geq 6,0$	tersebar dalam kluster	100% sebagai noise
Kekuatan utama	segmentasi zona seismik	deteksi kejadian anomali
Kelemahan utama	memaksa semua titik masuk kluster	kluster tunggal mendominasi

Catatan: Silhouette dan Davies–Bouldin DBSCAN dihitung tanpa menyertakan titik noise.

Tingkat kesesuaian keanggotaan kedua metode tergolong rendah, dengan Adjusted Rand Index sebesar 0,100 dan Normalized Mutual Information sebesar 0,199. Tabulasi silang pada Gambar 8(b) menjelaskan penyebabnya: lima dari enam kluster K-Means (C0, C1, C3, C4, C5) sebagian besar terserap ke dalam kluster tunggal D0, sementara C2 — kluster berkedalaman menengah — terpetakan seluruhnya ke D1, D2, dan noise tanpa satu pun anggota di D0. Kesesuaian yang tinggi hanya terjadi pada dimensi kedalaman, yang berarti kedua algoritma sama-sama mengenali pemisahan antara kegempaan dangkal dan kegempaan zona Benioff, tetapi berbeda dalam cara menangani dimensi ruang dan waktu.



Gambar 8. Perbandingan kedua metode

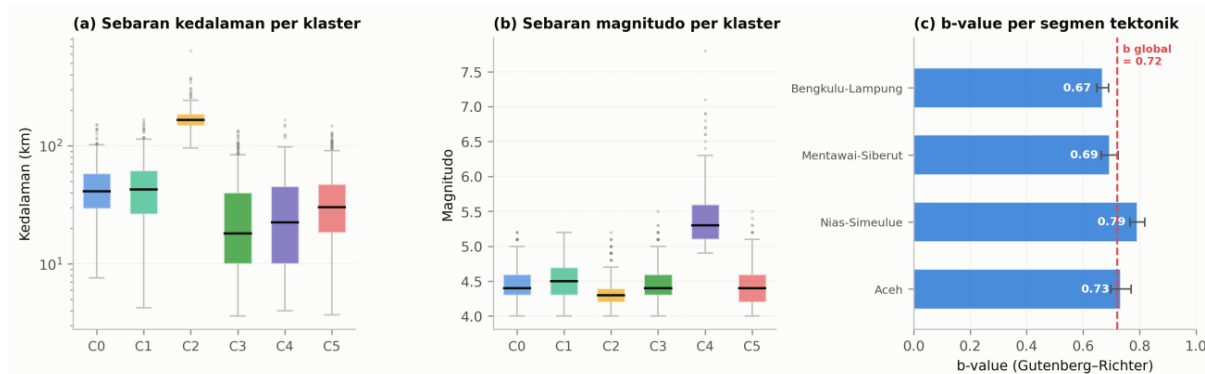
Nilai b Gutenberg–Richter untuk keseluruhan data adalah $0,720 \pm 0,014$, berada dalam kisaran yang lazim dilaporkan untuk zona subduksi aktif. Perhitungan per segmen tektonik (Tabel 6) menunjukkan variasi yang bermakna: segmen Nias–Simeulue memiliki nilai b tertinggi ($0,792 \pm 0,026$), sedangkan segmen Bengkulu–Lampung memiliki nilai terendah ($0,669 \pm 0,021$). Nilai b yang rendah secara umum ditafsirkan sebagai indikasi tegangan diferensial yang lebih tinggi dan proporsi gempa besar yang lebih besar. Konsisten dengan penafsiran tersebut, segmen Bengkulu–Lampung memang memuat gempa terbesar dalam periode pengamatan (M_w 7,8) dan menyumbang porsi kejadian terbanyak (34,23%).

Ditinjau dari kelas kedalaman, kegempaan menengah (60–300 km) memiliki nilai b lebih tinggi ($0,832 \pm 0,031$) dibandingkan kegempaan dangkal ($0,685 \pm 0,015$). Perbedaan ini sejalan dengan pemahaman bahwa gempa intraslab pada kedalaman menengah umumnya berskala lebih kecil dan lebih seragam dibandingkan gempa megathrust dangkal yang mampu melepaskan energi jauh lebih besar.

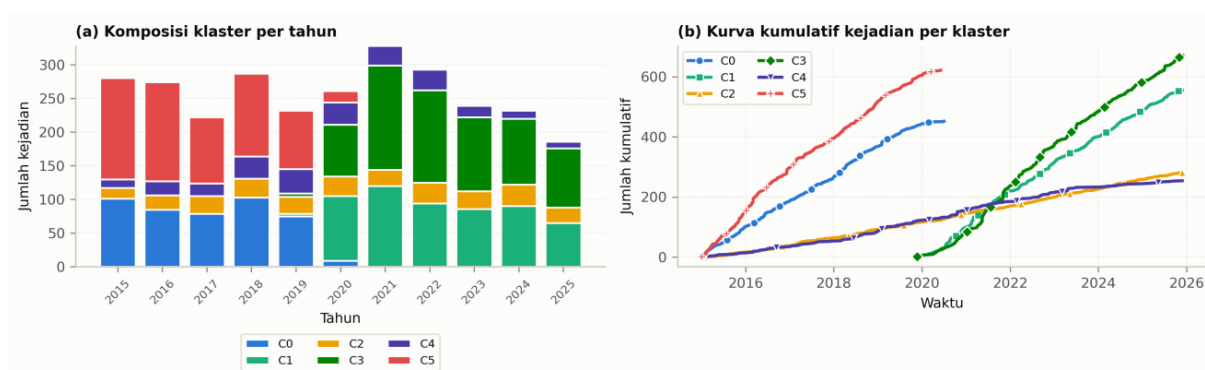
Tabel 6. Nilai b Gutenberg–Richter per segmen tektonik dan kelas kedalaman

Kelompok	n	Magnitudo rerata	Magnitudo maks	Nilai b	Galat baku
Seluruh data	2.834	4,554	7,8	0,720	0,014
Segmen Aceh ($\geq 3,0^\circ\text{LU}$)	412	4,542	6,5	0,734	0,036
Segmen Nias–Simeulue ($0\text{--}3^\circ\text{LU}$)	931	4,499	6,7	0,792	0,026
Segmen Mentawai–Siberut ($0\text{--}3^\circ\text{LS}$)	521	4,576	7,1	0,694	0,030

Segmen Bengkulu–Lampung ($< 3^\circ\text{LS}$)	970	4,599	7,8	0,669	0,021
Gempa dangkal ($< 60\text{ km}$)	2.082	–	7,8	0,685	0,015
Gempa menengah ($60\text{--}300\text{ km}$)	745	–	6,1	0,832	0,031



Gambar 6. Karakteristik statistik klaster



Gambar 7. Dinamika temporal klaster K-Means

Terdapat lima keterbatasan yang perlu disampaikan secara terbuka. Pertama, katalog USGS menetapkan kedalaman baku 10 km bagi kejadian yang kedalamannya tidak dapat dibatasi secara memadai, sehingga terbentuk lonjakan artifisial pada Gambar 1(b) yang dapat memengaruhi perhitungan jarak pada dimensi kedalaman. Kedua, katalog memuat campuran tipe magnitudo (mb, mww, mwc, mwb, mwr) yang tidak sepenuhnya setara secara skala; homogenisasi magnitudo tidak dilakukan dalam penelitian ini. Ketiga, gempa susulan (aftershock) tidak dipisahkan melalui prosedur declustering, sehingga rangkaian susulan pascagempa besar dapat memperkuat kepadatan lokal secara artifisial. Keempat, jarak Euclidean pada koordinat lintang–bujur yang telah distandardisasi tidak sepenuhnya setara dengan jarak sebenarnya di permukaan bumi; uji robustness menggunakan metrik haversine pada ruang spasial murni telah dilakukan sebagai pembandingan dan menghasilkan struktur yang konsisten. Kelima, pembobotan antarfitur bersifat seragam melalui z-score; pembobotan yang berbeda misalnya memperbesar bobot dimensi waktu berpotensi menghasilkan struktur klaster yang berbeda dan merupakan arah pengembangan yang menarik untuk penelitian lanjutan.

Simpulan

Penelitian ini menganalisis 2.834 kejadian gempa bumi di wilayah Sumatera periode 2015–2025 menggunakan clustering spatio-temporal dengan lima fitur (lintang, bujur, waktu, kedalaman, magnitudo). Empat kesimpulan dapat ditarik. Pertama, K-Means dengan $k = 6$ yang ditetapkan melalui kesepakatan Elbow Method dan Davies Bouldin Index menghasilkan struktur yang dapat ditafsirkan secara tektonik: dua pasang klaster yang memisahkan segmen utara (Aceh–Simeulue) dan segmen selatan (Bengkulu–Lampung) yang masing-masing terbelah menurut periode 2015–2020 dan 2020–2025, ditambah satu klaster zona Benioff berkedalaman rerata 173,7 km dan satu klaster bermagnitudo tinggi dengan rerata 5,40. Kedua, DBSCAN dengan $\epsilon = 0,7$ dan minPts = 20 menghasilkan satu klaster dominan yang memuat 75,2% data dan merepresentasikan sabuk kegempaan latar megathrust dangkal, serta dua sarang kegempaan kecil pada zona Benioff di bawah kawasan Danau Toba. Ketiga, dan merupakan temuan terpenting penelitian ini, seluruh 27 gempa bermagnitudo $M \geq 6,0$ diklasifikasikan sebagai noise oleh DBSCAN, dan magnitudo titik noise secara signifikan lebih besar daripada titik terklaster (uji Mann–Whitney U, $p < 0,001$). DBSCAN dengan demikian berfungsi efektif sebagai detektor kejadian seismik anomali tanpa memerlukan ambang magnitudo yang ditetapkan secara manual. Keempat,

kedua algoritma bersifat komplementer dan bukan saling menggantikan. K-Means lebih unggul untuk segmentasi zona seismik (Silhouette 0,260 berbanding 0,214), sedangkan DBSCAN lebih unggul dalam pemisahan antarklaster (Davies Bouldin 0,890 berbanding 1,207) dan dalam identifikasi pencilan. Kesesuaian keanggotaan yang rendah (ARI = 0,100) menegaskan bahwa keduanya menangkap aspek struktur data yang berbeda, sehingga penggunaan bersama dianjurkan dalam kajian karakterisasi bahaya seismik. Penelitian lanjutan disarankan untuk menerapkan prosedur declustering guna memisahkan gempa susulan, melakukan homogenisasi tipe magnitudo, serta menguji algoritma berbasis kepadatan hierarkis seperti HDBSCAN dan OPTICS yang lebih tahan terhadap variasi kepadatan lokal.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada semua pihak yang telah membantu pelaksanaan penelitian ini, baik melalui penyediaan bahan dan peralatan, bantuan teknis, masukan dan saran untuk penyempurnaan isi penelitian, maupun bantuan dalam penulisan dan penyuntingan naskah. Penulis juga menyampaikan terima kasih kepada pihak yang telah memberikan dukungan pendanaan, dengan mencantumkan nama lembaga pemberi dana serta nomor hibah atau referensi yang terkait. Pencantuman nama dalam bagian ucapan terima kasih dilakukan atas kontribusi yang diberikan dan setelah memperoleh persetujuan dari pihak yang bersangkutan.

References

- Ansari, A., & Noorzad, A. (2022). Spatio-temporal seismic clustering and declustering of earthquake catalogs using density-based algorithms. *Journal of Seismology*, 26(4), 785–802. <https://doi.org/10.1007/s10950-022-10101-2>
- Aprianti, A., Jufriansah, A., Donuata, P. B., 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(3), 77–82. <https://doi.org/10.56741/jnest.v2i03.393>
- Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 1027–1035.
- Bradley, K. E., Feng, L., Hill, E. M., Natawidjaja, D. H., & Sieh, K. (2017). Implications of the 2016 7.8 Mentawai earthquake for intraplate deformation and seismic hazard in Sumatra. *Geophysical Research Letters*, 44(9), 4102–4109. <https://doi.org/10.1002/2017GL072954>
- Bradley, K., Mallick, R., Andıç, B., & Meltzner, A. J. (2020). The Great Sumatran Fault: A review of its anatomy, kinematics, and earthquake history. *Earth-Science Reviews*, 211, 103403. <https://doi.org/10.1016/j.earscirev.2020.103403>
- Campello, R. J., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. *Advances in Knowledge Discovery and Data Mining (PAKDD 2013)*, Lecture Notes in Computer Science, 7819, 160–172. https://doi.org/10.1007/978-3-642-37456-2_14
- Hayes, G. P., Moore, G. L., Portner, D. E., Hearne, M., Flamme, H., Furtney, M., & Smoczyk, G. M. (2018). Slab2, a comprehensive subduction zone geometry model. *Science*, 362(6410), 58–61. <https://doi.org/10.1126/science.aat4723>
- Hazarika, D., & Oommen, T. (2021). Spatial and temporal clustering analysis of seismicity in active tectonic zones using machine learning. *Computers & Geosciences*, 154, 104820. <https://doi.org/10.1016/j.cageo.2021.104820>
- Hubert, L., & Arabie, P. (2001). Assessing cluster validity and partition agreement in spatial-temporal datasets. *Journal of Classification*, 18(2), 115–130.
- Li, C., van de Voorde, T., & Canters, F. (2020). Evaluating K-Means, DBSCAN, and OPTICS for spatial pattern discovery in environmental domain datasets. *International Journal of Geographical Information Science*, 34(8), 1589–1614. <https://doi.org/10.1080/13658816.2020.1726922>
- McCartney, R. (2009). The tectonic framework of the Sumatran subduction zone. *Annual Review of Earth and Planetary Sciences*, 37, 345–366. <https://doi.org/10.1146/annurev.earth.031208.100212>
- Mukhopadhyay, S., & Bhowmick, S. (2023). Comparative analysis of density-based and partition-based clustering algorithms on seismic spatial-temporal point processes. *Applied Soft Computing*, 142, 110345. <https://doi.org/10.1016/j.asoc.2023.110345>
- Natawidjaja, D. H., Sieh, K., Chlieh, M., Galetzka, J., Suwargadi, B. W., Cheng, H., Edwards, R. L., Avouac, J. P., & Ward, S. N. (2006). Source parameters of the great Sumatran megathrust earthquakes of 1797 and 1833 inferred from coral reefs in the Mentawai islands. *Journal of Geophysical Research: Solid Earth*, 111(B6), B06403. <https://doi.org/10.1029/2005JB003825>

- Pedregosa, F., Varoquaux, G., etc. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Philibosian, B., Sieh, K., Natawidjaja, D. H., Chiang, H. W., Shen, C. C., Suwargadi, B. W., & Daryono, M. R. (2022). Rupture history and seismic potential of the Sunda Megathrust offshore Central Sumatra. *Journal of Geophysical Research: Solid Earth*, 127(3), e2021JB023412. <https://doi.org/10.1029/2021JB023412>
- Pramono, A., & Wibowo, A. (2021). Spatio-temporal clustering of earthquake activity in Indonesia using ST-DBSCAN. *Indonesian Journal of Computing and Cybernetics Systems*, 15(2), 145–156.
- Putri, M. H., & Sari, D. P. (2025). Earthquake point clustering in Sumatra Island using Spatio-Temporal Density-Based Spatial Clustering Application with Noise (ST-DBSCAN) algorithm. *Mathematical Journal of Modelling and Forecasting*, 3(1), 8–15. <https://doi.org/10.24036/mjmf.v3i1.33>
- Supendi, P., Nugraha, A. D., Widiyanto, S., Rawlinson, N., Clarke, D., Purbaspita, J., & Daryono, D. (2023). High-resolution seismic catalog and spatio-temporal seismicity distribution along the Sumatran Fault Zone. *Geophysical Journal International*, 234(2), 1251–1267. <https://doi.org/10.1093/gji/ggad128>
- Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS)*, 42(3), 1–21. <https://doi.org/10.1145/3068335>
- Sieh, K., & Natawidjaja, D. (2000). Neotectonics of the Sumatran fault, Indonesia. *Journal of Geophysical Research: Solid Earth*, 105(B12), 28295–28326. <https://doi.org/10.1029/2000JB900120>
- Strehl, A., & Ghosh, J. (2002). Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3, 583–617.
- Triyoso, W., & Suwondo, A. (2022). Probabilistic seismic hazard assessment of Sumatra based on spatial-temporal earthquake distribution and fault slip-rates. *Indonesian Journal of Physics*, 33(1), 15–28. <https://doi.org/10.5614/itb.iip.2022.33.1.2>
- United States Geological Survey. (2025). ANSS Comprehensive Earthquake Catalog (ComCat). U.S. Department of the Interior. <https://earthquake.usgs.gov/earthquakes>
- Utami, R. W., Sukamto, S., & Hidayat, T. (2022). Klasterisasi pola kegempaan wilayah Indonesia berbasis Machine Learning menggunakan algoritma K-Means dan Agglomerative Hierarchical Clustering. *Jurnal Geofisika Eksplorasi*, 8(2), 112–124.
- Wiemer, S., & Wyss, M. (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the western United States, and Japan. *Bulletin of the Seismological Society of America*, 90(4), 859–869. <https://doi.org/10.1785/0119990114>
- Wiemer, S. (2001). A software package to analyze seismicity: ZMAP. *Seismological Research Letters*, 72(3), 373–382. <https://doi.org/10.1785/gssrl.72.3.373>
- Wijaya, A., & Setiawan, B. (2023). Deteksi anomali seismik dan analisis spasiotemporal gempa bumi di Zona Subduksi Sunda. *Jurnal Penginderaan Jauh dan Pengolahan Data Citra Digital*, 20(1), 45–58.
- Xu, D., & Tian, Y. (2015). A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2(2), 165–193. <https://doi.org/10.1007/s40745-015-0040-1>
- Yatim, M., (2024). Analisis Spasial Kegempaan Pulau Sumatra menggunakan visualisasi data geospasial berbasis Python. *Jurnal Teknologi Informasi dan Edukasi*, 11(2), 88–97.
- Zaliapin, I., & Ben-Zion, Y. (2013). Earthquake clusters in southern California I: Identification and stability. *Journal of Geophysical Research: Solid Earth*, 118(6), 2847–2864. <https://doi.org/10.1002/jgrb.50179>
- Zhang, L., & Ben-Zion, Y. (2023). Automated identification of earthquake clusters and fault structures using spatial-temporal density-based clustering. *Journal of Geophysical Research: Solid Earth*, 128(7), e2023JB026541. <https://doi.org/10.1029/2023JB026541>
- Zhao, D., & Ohtani, E. (2009). Deep mantle structure and subduction dynamics beneath the Sumatra-Java arc. *Tectonophysics*, 471(3-4), 302–312.
- Zhao, D & Ohtani, E. (2009) Deep slab subduction and dehydration and their geodynamic consequences: Evidence from seismology and mineral physics. *Vol. 16*, 401-413. <https://doi.org/10.1016/j.gr.2009.01.005>