

Implementasi Metode *ST-DBSCAN* untuk Pengelompokan Pola Gempa Bumi di Pulau Jawa

Amanda Devanti Prameswari *, Suliadi

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Islam Bandung, Indonesia.

amandadevantiprameswari@gmail.com, suliadi@unisba.ac.id

Abstract. The aim of this study is to cluster earthquake occurrence patterns on Java Island using the Spatio-Temporal DBSCAN (ST-DBSCAN) algorithm, which simultaneously processes spatial and temporal data. Data were obtained from BMKG (Meteorology, Climatology, and Geophysics Agency) comprising of 230 earthquake events from January to July 2025. The variables analyzed in this research are latitude, longitude, event time, depth, and magnitude. We conducted 27 parameter combinations of ϵ_1 (spatial radius), ϵ_2 (temporal radius), and MinPts (minimum number of points in a cluster). The results were evaluated using the Density-Based Clustering Validation (DBCV) metric to assess clustering quality. The best parameter combination was $\epsilon_1 = 50$ km, $\epsilon_2 = 12$ hours, and MinPts = 2, producing 32 clusters with a DBCV score of 0.8421. Visualization revealed that earthquake clusters were concentrated in the southern region of Java Island, which is part of an active subduction zone, while noise points were randomly distributed. The findings demonstrate that ST-DBSCAN is effective in identifying spatial-temporal earthquake patterns and can serve as a valuable tool for seismic risk mapping and disaster mitigation planning in vulnerable areas.

Keywords: *ST-DBSCAN, Spatio-Temporal Clustering, Earthquake.*

Abstrak. Penelitian ini bertujuan untuk mengelompokkan pola kejadian gempa bumi di Pulau Jawa menggunakan algoritma *Spatio-Temporal DBSCAN* (ST-DBSCAN) yang mampu mengolah data spasial dan temporal secara bersamaan. Data sekunder dari BMKG sebanyak 230 kejadian gempa periode Januari–Juli 2025 digunakan sebagai objek penelitian. Variabel utama yang dianalisis mencakup lintang, bujur, waktu kejadian, kedalaman, dan magnitudo. Proses klasterisasi dilakukan dengan menguji 27 kombinasi parameter ϵ_1 (radius spasial), ϵ_2 (radius temporal), dan MinPts (jumlah minimum titik dalam klaster). Evaluasi hasil dilakukan menggunakan metrik *Density-Based Clustering Validation* (DBCV) untuk menilai kualitas klaster. Kombinasi parameter terbaik diperoleh pada $\epsilon_1 = 50$ km, $\epsilon_2 = 12$ jam, dan MinPts = 2, menghasilkan 32 klaster dengan nilai DBCV sebesar 0,8421. Visualisasi menunjukkan bahwa klaster gempa terkonsentrasi di wilayah selatan Pulau Jawa yang merupakan zona subduksi aktif, sedangkan titik-titik *noise* tersebar acak. Hasil penelitian menunjukkan bahwa ST-DBSCAN efektif dalam mendeteksi pola gempa yang saling berkaitan secara spasial-temporal dan dapat menjadi alat bantu dalam pemetaan risiko serta upaya mitigasi bencana gempa bumi di wilayah rawan.

Kata Kunci: *ST-DBSCAN, Pengelompokan Spasial-Temporal, Gempa Bumi.*

A. Pendahuluan

Pengelompokan data (*clustering*) merupakan teknik penting dalam ilmu data untuk mengidentifikasi pola berdasarkan kemiripan antar data. Salah satu algoritma yang banyak digunakan adalah DBSCAN (Density-Based Spatial Clustering of Applications with Noise), yang mampu mengidentifikasi kluster dengan bentuk tak beraturan dan menangani data berisi noise tanpa perlu menentukan jumlah kluster di awal (Xu & Tian, 2015; Zhou dkk., 2000). DBSCAN bekerja berdasarkan dua parameter utama: epsilon (ϵ), yang menentukan jarak maksimum antar titik dalam kluster, dan MinPts, jumlah minimum titik untuk membentuk kluster. Meski efektif, akurasi algoritma ini sangat bergantung pada pemilihan parameter yang tepat (Shen dkk., 2016).

Untuk mengatasi keterbatasan DBSCAN dalam menangani data spasial dan temporal secara bersamaan, dikembangkan algoritma ST-DBSCAN (Spatial-Temporal DBSCAN). Algoritma ini menggabungkan informasi spasial (lokasi) dan temporal (waktu kejadian) dalam satu proses klustering, menjadikannya sangat cocok untuk menganalisis data dinamis seperti pergerakan kendaraan, pola cuaca, serta kejadian gempa bumi (Birant & Kut, 2007). Meski demikian, ST-DBSCAN juga memiliki kelemahan serupa, yaitu sensitivitas terhadap parameter spasial dan temporal yang dapat menyebabkan kesalahan dalam pembentukan kluster (Nicolis dkk., 2023).

Gempa bumi adalah fenomena geologis yang kompleks dan dapat menimbulkan kerusakan besar, baik materiil maupun immateriil. Indonesia, khususnya Pulau Jawa, merupakan kawasan dengan tingkat aktivitas seismik yang tinggi karena berada di pertemuan tiga lempeng utama dunia: Indo-Australia, Eurasia, dan Pasifik (Widiyantoro dkk., 2020). Sepanjang tahun 2023, Indonesia tercatat mengalami lebih dari 2.200 gempa bermagnitudo ≥ 4 (Annur, 2024). Kondisi ini menjadikan pemetaan pola gempa sangat penting untuk upaya mitigasi risiko. Namun, sebagian besar masyarakat masih belum memiliki kesiapan dan pengetahuan memadai dalam menghadapi bencana alam (Prasetyo dkk., 2023). Lima parameter utama yang digunakan adalah waktu kejadian (tanggal), lintang (*latitude*), bujur (*longitude*), kedalaman, dan magnitudo. Kelima parameter tersebut digunakan untuk membedakan gempa utama (*mainshock*) dan susulan (*aftershock*), serta mengidentifikasi sebaran spasial dan temporal dari kejadian seismik. Dengan menggunakan algoritma ini, diharapkan pola tersembunyi dalam distribusi gempa dapat terungkap lebih akurat dan komprehensif.

Pulau Jawa dipilih sebagai wilayah studi karena letaknya di zona subduksi aktif dan memiliki karakteristik tektonik yang mirip dengan Chile, negara dengan tingkat seismisitas tinggi yang juga menjadi lokasi uji coba ST-DBSCAN dalam penelitian sebelumnya (Cesca dkk., 2016). Seperti di Chile, subduksi di Jawa terjadi akibat pertemuan antara lempeng samudra dan benua yang menyebabkan gempa dengan pola distribusi kompleks. Hasil penelitian ini diharapkan dapat menjadi dasar ilmiah untuk mitigasi bencana berbasis data dan memberi kontribusi pada pengembangan model klustering spasial-temporal dalam bidang geofisika dan kebencanaan.

Tujuan dari penelitian ini adalah untuk mengelompokkan kejadian gempa bumi yang terjadi di Pulau Jawa dengan menggunakan algoritma ST-DBSCAN (*Spatial-Temporal DBSCAN*). Pendekatan ini digunakan untuk mengidentifikasi pola spasial dan temporal dari aktivitas seismik dengan mempertimbangkan lokasi dan waktu kejadian gempa secara simultan, sehingga diharapkan mampu mengungkap kluster gempa yang signifikan.

Manfaat dari penelitian ini antara lain adalah memberikan kontribusi terhadap pengembangan metode pengelompokan berbasis densitas dengan penerapan algoritma ST-DBSCAN, khususnya dalam konteks data seismik yang memiliki distribusi tidak merata dan bersifat dinamis. Penelitian ini juga bertujuan untuk mengidentifikasi wilayah-wilayah di Pulau Jawa yang rentan terhadap gempa bumi, guna mendukung upaya mitigasi bencana secara lebih terarah. Selain itu, hasil penelitian ini diharapkan dapat menjadi referensi bagi studi lanjutan dalam bidang analisis spasial-temporal dan penerapan algoritma ST-DBSCAN pada data geofisika lainnya.

B. Metode

Penelitian ini menggunakan data sekunder kejadian gempa bumi yang terjadi di Pulau Jawa dalam rentang waktu Januari hingga Juli 2025. Data tersebut diperoleh dari situs resmi BMKG (Badan Meteorologi Klimatologi dan Geofisika) dan mencakup 230 peristiwa gempa <https://dataonline.bmkg.go.id/data-gempabumi>. Terdapat lima variabel utama yang dianalisis, yaitu waktu kejadian (*date*), koordinat geografis (*latitude* dan *longitude*), kedalaman (*depth*), dan

magnitudo. Dalam proses klusterisasi menggunakan metode ST-DBSCAN, variabel koordinat geografis (*latitude* dan *longitude*) (X_1) dan waktu kejadian (*date*) (X_2) digunakan sebagai input utama, sedangkan variabel *depth* dan magnitudo berfungsi sebagai informasi pendukung yang menjelaskan waktu dan lokasi masing-masing kejadian gempa. Data yang digunakan dapat dilihat pada link berikut ini <https://bit.ly/DataGempaBumiBMKGJansdJuli2025>

Spatial-Temporal DBSCAN (ST-DBSCAN)

Spatio-Temporal DBSCAN (ST-DBSCAN) merupakan pengembangan dari algoritma DBSCAN yang dirancang untuk mendeteksi kluster dalam data spasial-temporal dengan mempertimbangkan tidak hanya dimensi spasial (seperti *latitude* dan *longitude*), tetapi juga dimensi temporal (waktu kejadian). Algoritma ini menggunakan tiga parameter utama, yaitu epsilon spasial (ϵ_1), epsilon temporal (ϵ_2), dan MinPts sebagai jumlah minimum titik untuk membentuk kluster (Birant & Kut, 2007). Keunggulan ST-DBSCAN dibandingkan DBSCAN konvensional terletak pada kemampuannya dalam menangani variasi densitas lokal antar wilayah kluster, sehingga lebih efektif dalam mendeteksi noise dan mengelompokkan titik-titik yang berada di perbatasan antar kluster. Dengan cara membandingkan kepadatan lokal suatu titik terhadap rata-rata kepadatan kluster terdekat, algoritma ini dapat memasukkan titik tersebut ke dalam kluster yang paling sesuai, sehingga meningkatkan akurasi terutama pada area yang tumpang tindih antar kluster (Birant & Kut, 2007).

Tahapan algoritma Spatio-Temporal DBSCAN (ST-DBSCAN) diawali dengan menentukan tiga parameter utama: epsilon spasial ($\epsilon_{spasial}$), epsilon temporal ($\epsilon_{temporal}$), dan MinPts sebagai jumlah minimum tetangga yang dibutuhkan untuk membentuk suatu kluster. Setelah parameter ditentukan, langkah berikutnya adalah menghitung jarak *Haversine* antar setiap pasangan objek dalam dataset berdasarkan atribut spasial (*latitude* dan *longitude*) dan atribut temporal (waktu). Kemudian, dibentuk matriks jarak yang mencakup seluruh kombinasi objek terhadap atribut spasial dan temporal tersebut (Birant & Kut, 2007).

Proses pembentukan kluster dimulai dengan memilih titik acak dalam dataset sebagai titik awal (q). Dua himpunan dibentuk berdasarkan jarak dari titik q : A untuk titik-titik dalam radius $\epsilon_{spasial}$, dan B untuk titik-titik dalam radius $\epsilon_{temporal}$. Irisan dari kedua himpunan ini ($A \cap B$) merupakan titik-titik yang memenuhi syarat spasial dan temporal sekaligus. Jika jumlah titik dalam $A \cap B$ kurang dari MinPts, maka titik q dianggap sebagai noise. Namun, jika jumlahnya sama dengan atau melebihi MinPts, maka titik q adalah core point dan membentuk kluster awal (C_1). Proses ini kemudian dilanjutkan dengan memperluas kluster tersebut menggunakan titik-titik yang juga memenuhi kriteria spasial-temporal, sedangkan titik yang tidak bisa memperluas kluster namun berada di batasannya akan diberi label sebagai border point. Langkah-langkah ini diulang hingga seluruh titik dalam dataset telah diperiksa dan diidentifikasi sebagai bagian dari kluster tertentu atau sebagai noise (Birant & Kut, 2007).

Matriks Jarak

Dalam algoritma ST-DBSCAN, digunakan dua parameter epsilon (ϵ) yang berfungsi untuk mengukur kedekatan spasial dan temporal antar titik. Parameter ϵ_1 digunakan untuk menentukan jarak spasial, yaitu seberapa dekat dua objek secara geografis agar dapat dimasukkan ke dalam kluster yang sama. Sementara itu, ϵ_2 mengukur kedekatan temporal, yaitu kesamaan waktu kejadian suatu data. Penggunaan jarak Euclidean dalam perhitungan jarak spasial hanya cocok pada wilayah kecil atau dalam sistem koordinat datar seperti UTM (Universal Transverse Mercator). Ketika data spasial dinyatakan dalam sistem koordinat geografis (*latitude* dan *longitude*) pada permukaan bumi yang melengkung, penggunaan rumus Euclidean tidak lagi akurat. Untuk itu, rumus Haversine lebih tepat digunakan karena mempertimbangkan kelengkungan bumi dalam perhitungan jarak antar dua titik. Haversine memberikan estimasi yang lebih realistis pada skala global atau area yang luas, dan karenanya lebih cocok dalam konteks analisis spasial dengan data koordinat geografis (Esenbuga, 2016). Di mana, jarak spasial dihitung dengan rumus jarak Haversine untuk n -dimensi dalam Persamaan berikut:

$$d_1(i, j) = d_{spasial}(i, j) = 2r \cdot \arcsin \left(\sqrt{\sin^2 \left(\frac{\phi_i - \phi_j}{2} \right) + \cos(\phi_i) \cdot \cos(\phi_j) \cdot \sin^2 \left(\frac{\lambda_i - \lambda_j}{2} \right)} \right)$$

Parameter ϕ_i dan ϕ_j masing-masing merepresentasikan nilai *latitude* dari titik ke- i dan ke- j , sedangkan λ_i dan λ_j menyatakan nilai *longitude* dari kedua titik tersebut. Seluruh koordinat tersebut dinyatakan dalam satuan radian agar sesuai dengan perhitungan dalam fungsi trigonometri. Nilai r merupakan jari-jari bumi yang secara umum diasumsikan sebesar 6.371 kilometer dalam perhitungan jarak geografis. Hasil perhitungan dari rumus *Haversine* akan memberikan nilai $d_{spasial}(i, j)$, yaitu jarak antara dua titik berdasarkan atribut spasial berupa koordinat geografis yang digunakan dalam algoritma ST-DBSCAN.

Sebelum menerapkan rumus jarak *Haversine*, nilai koordinat latitude dan longitude dalam satuan derajat harus terlebih dahulu dikonversi ke satuan radian, karena rumus jarak *Haversine* bekerja dengan satuan radian. Konversi ini dilakukan dengan rumus berikut ini $\phi = \text{radian}_{latitude} = \text{derajat}(\circ) \times \frac{\pi}{180}$ dan $\lambda = \text{radian}_{longitude} = \text{derajat}(\circ) \times \frac{\pi}{180}$. Di mana, $\pi = 3,14$ (Akbar & Gautama, 2023):

Nilai ε_2 digunakan sebagai ambang batas untuk mengevaluasi jarak antar titik berdasarkan atribut temporal. Jarak temporal ini dihitung menggunakan rumus Manhattan Distance, yang hanya mempertimbangkan satu dimensi, yaitu perbedaan waktu antar dua titik. Perhitungannya dinyatakan dalam Persamaan berikut: $d_2(i, j) = d_{temporal}(i, j) = |X_i - X_j|$, di mana X_i dan X_j masing-masing merepresentasikan nilai waktu dari titik ke- i dan ke- j . Hasil dari rumus ini memberikan jarak absolut antara dua peristiwa dalam dimensi waktu, sehingga memungkinkan algoritma ST-DBSCAN untuk mengevaluasi kesamaan temporal secara efisien (Han dkk., 2011).

Pemilihan Parameter

Pemilihan parameter dalam algoritma ST-DBSCAN pada studi kasus gempa bumi di Pulau Jawa melibatkan dua komponen utama, yaitu parameter spasial (ε_1) dan parameter temporal (ε_2). Penentuan nilai kedua parameter ini harus disesuaikan dengan karakteristik data seismik dan kondisi geografis wilayah penelitian. Dalam aspek temporal, ε_2 digunakan untuk mengukur kedekatan waktu antar kejadian gempa. Beberapa studi menunjukkan bahwa nilai ε_2 dalam satuan jam efektif untuk mendeteksi hubungan antar gempa yang terjadi secara berurutan. Misalnya, Samsudin dkk. (2024) menyarankan penggunaan rentang waktu 12 hingga 24 jam untuk menangkap pola gempa susulan di zona subduksi Jawa Selatan. Sementara itu, Iswari (2022) menekankan bahwa ambang waktu hingga 48 jam diperlukan untuk menganalisis rangkaian gempa secara mikro dalam satu sistem patahan aktif. Oleh karena itu, penggunaan ε_2 sebesar 12, 24, dan 48 jam dianggap cukup representatif untuk mengidentifikasi dinamika temporal dalam kluster gempa bumi di Pulau Jawa.

Selain itu, dalam dimensi spasial, parameter ε_1 ditentukan berdasarkan jarak maksimum antar titik episentrum yang dianggap masih berkaitan secara tektonik. Penentuan nilai ε_1 mempertimbangkan pola sebaran gempa di Pulau Jawa serta karakteristik propagasi gelombang seismik. Wilayah ini dikenal memiliki risiko gempa tinggi karena terletak di zona subduksi aktif. Samsudin dkk. (2024), mengusulkan penggunaan variasi ε_1 antara 50 km hingga 150 km untuk mencerminkan jangkauan seismik antara gempa utama dan gempa susulan.

Metrik Validasi Kluster

Validasi kluster merupakan tantangan utama dalam analisis data, terutama bagi algoritma berbasis densitas seperti DBSCAN, OPTICS, dan HDBSCAN, yang menghasilkan kluster dengan bentuk tidak beraturan dan mengandung noise. Menurut Moulavi dkk. (2014), sebagian besar metode validasi tradisional hanya cocok untuk kluster berbentuk cembung (globular). Untuk mengatasi hal tersebut, dikembangkanlah DBCV (*Density-Based Clustering Validation*), yaitu indeks validasi relatif yang dirancang khusus untuk menilai kualitas kluster berbasis densitas dengan mempertimbangkan kepadatan, bentuk kluster, dan keberadaan *noise*.

Tahapan awal dalam DBCV adalah menghitung nilai *aptscoredist* untuk setiap objek dalam kluster, yang mengukur seberapa padat suatu titik terhadap tetangganya menggunakan gabungan jarak spasial dan temporal. Jarak ini dihitung menggunakan rumus gabungan spasial-temporal, yaitu

$$dist_{ST}(i, j) = \sqrt{\alpha \cdot (d_{spasial}(i, j))^2 + (1-\alpha) \cdot (d_{temporal}(i, j))^2}$$

Dengan α sebagai koefisien penyeimbang (biasanya 0,5). Nilai *aptscoredist* ini digunakan dalam perhitungan *Mutual Reachability Distance* (MRD) antar objek, yaitu jarak maksimum antara

dua titik berdasarkan jarak langsung dan kepadatan lokal keduanya. Hasil MRD ini kemudian digunakan untuk membangun *Minimum Spanning Tree of Mutual Reachability Distance* (MSTM RD) untuk setiap klaster.

Selanjutnya, dua ukuran penting dihitung, yaitu *Density Sparseness* (DSC) dan *Density Separation* (DSPC). DSC mengukur seberapa jarang titik-titik terluar dalam satu klaster, sedangkan DSPC mengukur jarak densitas minimum antar klaster. Kedua nilai ini digunakan dalam perhitungan Validitas Klaster (VC) untuk setiap klaster menggunakan rasio antara DSPC dan DSC. Nilai VC akan menunjukkan seberapa kuat pemisahan antar klaster dibandingkan dengan kepadatannya sendiri. Semakin tinggi nilainya, semakin baik kualitas klaster tersebut.

$$DBCV(C) = \sum_{i=1}^l \frac{|C_i|}{|O|} \cdot VC(C_i)$$

Terakhir, nilai DBCV dihitung sebagai rata-rata tertimbang dari nilai VC setiap klaster, berdasarkan ukuran masing-masing klaster. Nilai DBCV berada pada rentang -1 hingga 1 , dengan interpretasi sebagai berikut: nilai di atas $0,5$ menunjukkan klaster sangat baik dan terpisah jelas, nilai $0,2$ hingga $0,5$ menunjukkan klaster cukup baik, sedangkan nilai di bawah 0 menunjukkan struktur klaster yang lemah atau buruk. Dengan demikian, DBCV menjadi metode validasi yang lebih akurat dan sesuai untuk model berbasis densitas seperti ST-DBSCAN.

C. Hasil Penelitian dan Pembahasan

Analisis Pengelompokan ST-DBSCAN

Langkah awal dalam analisis pengelompokan ST-DBSCAN adalah transformasi data. Dalam penelitian ini, transformasi data mencakup pengolahan variabel waktu (*Date*) menjadi format numerik *Timestamp* yang menggabungkan tanggal dan waktu dalam satu fitur. Transformasi ini bertujuan untuk mempermudah analisis berbasis waktu dan dilakukan menggunakan software *Python*. Selain itu, koordinat geografis *Latitude* dan *Longitude* dikonversi dari derajat ke radian agar sesuai dengan kebutuhan perhitungan spasial menggunakan rumus *Haversine*.

Selanjutnya, lakukan perhitungan jarak spasial dan temporal dilakukan sebagai bagian awal dari metode ST-DBSCAN untuk membentuk klaster. Jarak spasial dihitung menggunakan rumus *Haversine* berdasarkan nilai *Latitude* dan *Longitude*, sedangkan jarak temporal dihitung dengan rumus *Manhattan* dari nilai *Timestamp*. Hasilnya adalah dua matriks jarak berukuran 230×230 yang masing-masing merepresentasikan kedekatan antar kejadian gempa secara spasial dan temporal.

Setelah melakukan transformasi data dan menghitung jarak spasial-temporal, selanjutnya lakukan analisis klaster ST-DBSCAN dari 27 kombinasi parameter, yaitu epsilon spasial (ϵ_1), epsilon temporal (ϵ_2), dan MinPts. Proses ini menggunakan matriks jarak yang telah dihitung sebelumnya dan menghasilkan sejumlah klaster berdasarkan kedekatan spasial-temporal antar gempa.

Setelah analisis klaster dilakukan menggunakan ST-DBSCAN dengan 27 kombinasi parameter, tahap berikutnya adalah evaluasi kualitas klaster menggunakan metode *Density-Based Clustering Validation* (DBCV). Evaluasi dimulai dengan menghitung jarak spasial dan temporal antar titik klaster (dengan mengecualikan *noise*), kemudian hasilnya dinormalisasi menggunakan *Min-Max Scaler* untuk menyamakan skala dimensi. Proses ini menghasilkan data dalam rentang $[0,1]$ yang kemudian digunakan untuk menghitung jarak gabungan. Setiap kombinasi parameter menghasilkan matriks jarak spasial dan temporal yang berbeda.

Tahapan akhir adalah perhitungan skor DBCV untuk menilai kualitas masing-masing model klaster. Perhitungan meliputi nilai jarak gabungan, apscore, MSTM RD, hingga validitas klaster (VC) dan skor DBCV total. Hasilnya akan disajikan pada tabel dibawah ini yaitu 5 dari 27 kombinasi parameter yang menghasilkan nilai DBCV tertinggi.

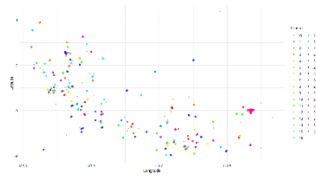
Tabel 1. Skor DBCV 5 Kombinasi Parameter Terbaik.

Kombinas i	epsilon spasial (ϵ_1)	epsilon temporal (ϵ_2)	MinPts	Hasil		
				Jumlah Klaster	Nois e	DBCV Score
1	50	12	2	32	55	0,8421*
2	50	24	2	25	21	0,8015

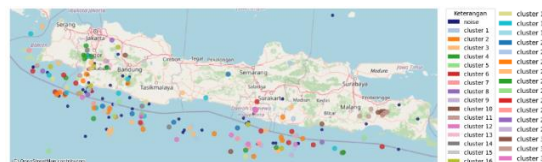
3	150	12	3	24	85	0,8014
4	150	48	2	15	6	0,7528
5	100	48	3	14	9	0,7491

Tabel 2 yang menunjukkan bahwa dari 5 kombinasi parameter yang memiliki nilai DBCV tertinggi, bahwa kombinasi parameter terbaik adalah pada kombinasi ke-1 dengan $\epsilon_1 = 50$ km, $\epsilon_2 = 12$ jam, dan MinPts = 2 menghasilkan skor DBCV sebesar 0,8421 yang membentuk 32 kluster dengan jumlah *noise* sejumlah 55. Nilai ini menunjukkan bahwa model tersebut menghasilkan kluster yang padat dan terpisah jelas, sehingga dapat disimpulkan sebagai parameter paling optimal dalam penelitian ini.

Hasil pengelompokan gempa di Pulau Jawa menggunakan ST-DBSCAN dengan menggunakan parameter terbaik yakni $\epsilon_1 = 50$ km, $\epsilon_2 = 12$ jam, dan MinPts = 2 dapat divisualisasikan pada Gambar 2 berikut ini.



Gambar 1. Hasil Pengelompokan ST-DBSCAN dengan Parameter Terbaik



Gambar 2. Hasil Pengelompokan ST-DBSCAN dengan Parameter Terbaik Dilihat dari Peta Pulau Jawa

Gambar 1 memperlihatkan hasil klusterisasi spasial-temporal menggunakan algoritma ST-DBSCAN, berdasarkan koordinat geografis lintang (*latitude*) dan bujur (*longitude*). Masing-masing titik pada grafik ini mewakili suatu kejadian spasial gempa bumi, dan diberikan warna sesuai dengan kluster yang dihasilkan. Secara keseluruhan terdapat 32 kluster teridentifikasi, ditandai dengan variasi warna, serta sejumlah titik berwarna abu-abu yang dikategorikan sebagai *noise*, yaitu kejadian yang tidak cukup padat secara spasial-temporal untuk membentuk kluster. Kluster-kluster yang terbentuk memperlihatkan distribusi spasial yang tidak merata, di mana beberapa kluster tampak sangat padat dan terkonsentrasi di area tertentu, sedangkan lainnya lebih menyebar atau hanya terdiri dari sedikit titik. Keberadaan kluster yang padat kemungkinan merefleksikan adanya hubungan spasial dan temporal yang kuat antar kejadian, seperti pola *foreshock*, *mainshock*, dan *aftershock* yang sering ditemukan pada zona seismik aktif. Sementara itu, titik-titik *noise* yang tidak masuk dalam kluster dapat menunjukkan adanya kejadian gempa yang berdiri sendiri tanpa keterkaitan dengan kejadian lainnya, baik secara lokasi maupun waktu. Dengan demikian, plot ini menggambarkan bagaimana ST-DBSCAN dapat memisahkan kejadian-kejadian yang berkorelasi spasial-temporal dari kejadian-kejadian tunggal yang bersifat lebih acak.

Gambar 2 menyajikan hasil yang sama namun dalam konteks spasial yang lebih realistis, yaitu dengan peta Pulau Jawa. Dalam visualisasi ini, distribusi titik-titik kluster dapat diamati lebih jelas dalam hubungannya dengan lokasi geografis sebenarnya. Kluster-kluster tampak banyak terkonsentrasi di wilayah pesisir selatan Pulau Jawa, meliputi daerah Banten bagian selatan, Sukabumi, Yogyakarta, Pacitan, Trenggalek, hingga Banyuwangi. Wilayah-wilayah ini merupakan bagian dari zona tektonik aktif yang dipengaruhi oleh zona subduksi Lempeng Indo-Australia dan Eurasia, serta keberadaan sesar-sesar aktif seperti Sesar Cimandiri, Sesar Opak, dan Sesar Grindulu. Sementara itu, titik-titik *noise* (berwarna biru tua) tampak tersebar di berbagai bagian pulau, termasuk wilayah tengah dan utara, menandakan adanya kejadian-kejadian yang lebih terisolasi secara spasial-temporal.

Secara umum, kedua visualisasi ini memberikan pemahaman komprehensif terhadap struktur spasial-temporal dari aktivitas seismik di Pulau Jawa. Hasil klasterisasi menunjukkan bahwa aktivitas gempa sangat terkonsentrasi di sepanjang selatan Jawa yang dikenal sebagai zona rawan gempa akibat tumbukan lempeng. Distribusi klaster yang jelas dalam wilayah tertentu memperlihatkan bahwa metode ST-DBSCAN efektif dalam mengidentifikasi pola kejadian yang berkaitan erat satu sama lain, baik dari segi lokasi maupun waktu. Di sisi lain, keberadaan *noise* juga menjadi informasi penting, karena mengindikasikan banyaknya kejadian tunggal yang tidak berulang atau tidak terkait secara spasial dan temporal. Interpretasi ini sangat berguna untuk mendukung kajian risiko gempa dan pengembangan strategi mitigasi bencana geologi secara lebih tepat sasaran.

D. Kesimpulan

Penelitian ini berhasil mengimplementasikan algoritma *Spatio-Temporal* DBSCAN (ST-DBSCAN) untuk mengelompokkan kejadian gempa bumi di Pulau Jawa berdasarkan dimensi spasial dan temporal. Dengan menggunakan data sekunder dari BMKG yang mencakup 230 kejadian gempa selama Januari hingga Juli 2025, algoritma ST-DBSCAN mampu membentuk klaster secara adaptif tanpa perlu menentukan jumlah klaster di awal. Melalui pengujian 27 kombinasi parameter, ditemukan bahwa kombinasi $\varepsilon_1 = 50$ km, $\varepsilon_2 = 12$ jam, dan MinPts = 2 merupakan kombinasi parameter terbaik yang menghasilkan 32 klaster dengan skor validitas DBCV tertinggi sebesar 0,8421. Ini menunjukkan struktur klaster yang padat dan terpisah secara signifikan dari noise.

Hasil klasterisasi menunjukkan bahwa gempa-gempa yang terjadi di Pulau Jawa cenderung membentuk pola spasial-temporal tertentu, terutama terkonsentrasi di wilayah selatan Jawa yang berdekatan dengan zona subduksi dan sesar aktif. Klaster-klaster padat yang terbentuk mencerminkan kemungkinan hubungan antara gempa awal, utama, dan susulan (*foreshock-mainshock-aftershock*), sedangkan sejumlah besar titik *noise* menunjukkan adanya kejadian gempa yang terisolasi secara spasial dan temporal. Visualisasi hasil klasterisasi, baik dalam plot koordinat maupun peta spasial, memberikan gambaran yang jelas mengenai persebaran dan intensitas aktivitas seismik di kawasan tersebut.

Ucapan Terimakasih

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) atas ketersediaan data gempa bumi yang menjadi sumber utama dalam penelitian ini.

Daftar Pustaka

- Dewi, E. F., & Hajarisman, N. (2023). Penanganan Data Hilang pada Pemodelan Persamaan Terstruktur melalui Metode Full Information Maximum Likelihood (FIML). *DataMath: Journal of Statistics and Mathematics*, 1(1), 11–18. <https://doi.org/10.29313/datamath.v1i1.10>
- Muhammad Rizq Nafisyah Alam, & Aceng Komarudin Mutaqin. (2023). Pemodelan Distribusi Poisson-Sujatha pada Data Frekuensi Klaim Asuransi Kendaraan Bermotor di Indonesia. *Jurnal Riset Statistika*, 71–78. <https://doi.org/10.29313/jrs.v3i1.1944>
- Akbar, I., & Gautama, E. (2023). Locating Identification the Nearest Social Security Organizing Agency for Health Recipient Hospital Using the Haversine Formula and Black Box Method. *International Journal of Advances in Scientific Research and Engineering*, 09(03), 25–36. <https://doi.org/10.31695/IJASRE.2023.9.3.4>
- Birant, D., & Kut, A. (2007). ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data & Knowledge Engineering*, 60(1), 208–221. <https://doi.org/10.1016/j.datak.2006.01.013>

- Cesca, S., Grigoli, F., Heimann, S., Dahm, T., Kriegerowski, M., Sobiesiak, M., Tassara, C., & Olcay, M. (2016). The M_w 8.1 2014 Iquique, Chile, seismic sequence: A tale of foreshocks and aftershocks. *Geophysical Journal International*, 204(3), 1766–1780. <https://doi.org/10.1093/gji/ggv544>
- Esenbuga, O. G. (2016, Juni 28). *COMPARISON OF PRINCIPAL GEODETIC DISTANCE CALCULATION METHODS FOR AUTOMATED PROVINCE ASSIGNMENT IN TURKEY*. 16th International Multidisciplinary Scientific GeoConference SGEM2016. <https://doi.org/10.5593/SGEM2016/B22/S09.019>
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3 ed.). Elsevier.
- Moulavi, D., Jaskowiak, P. A., Campello, R. J. G. B., Zimek, A., & Sander, J. (2014). Density-Based Clustering Validation. *Proceedings of the 2014 SIAM International Conference on Data Mining*, 839–847. <https://doi.org/10.1137/1.9781611973440.96>
- Nicolis, O., Delgado, L., Peralta, B., Diaz, M., & Chiodi, M. (2023). *Space-time clustering of seismic events in Chile using ST-DBSCAN-EV algorithm*. In Review. <https://doi.org/10.21203/rs.3.rs-3469745/v1>
- Prasetio, A., Effendi, M. M., & Dwi M, M. N. (2023). Analisis Gempa Bumi Di Indonesia Dengan Metode Clustering. *Bulletin of Information Technology (BIT)*, 4(3), 338–343. <https://doi.org/10.47065/bit.v4i3.820>
- Samsudin, A. R., Fudholi, D. H., & Iswari, L. (2024). TEMPORAL SPATIAL PROPERTY PROFILING AND IDENTIFICATION OF EARTHQUAKE PRONE AREAS USING ST-DBSCAN AND K-MEANS CLUSTERING. *Jurnal Teknik Informatika (Jutif)*, 5(3), 917–929. <https://doi.org/10.52436/1.jutif.2024.5.3.1293>
- Shen, J., Hao, X., Liang, Z., Liu, Y., Wang, W., & Shao, L. (2016). Real-Time Superpixel Segmentation by DBSCAN Clustering Algorithm. *IEEE Transactions on Image Processing*, 25(12), 5933–5942. *IEEE Transactions on Image Processing*. <https://doi.org/10.1109/TIP.2016.2616302>
- Widiyanoro, S., Gunawan, E., Muhari, A., Rawlinson, N., Mori, J., Hanifa, N. R., Susilo, S., Supendi, P., Shiddiqi, H. A., Nugraha, A. D., & Putra, H. E. (2020). Implications for megathrust earthquakes and tsunamis from seismic gaps south of Java Indonesia. *Scientific Reports*, 10(1), 15274. <https://doi.org/10.1038/s41598-020-72142-z>
- 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>
- Zhou, A., Zhou, S., Cao, J., Fan, Y., & Hu, Y. (2000). Approaches for scaling DBSCAN algorithm to large spatial databases. *Journal of Computer Science and Technology*, 15(6), 509–526. <https://doi.org/10.1007/BF02948834>
- Raisa Filmi Suryahadikusumah, & Sutawanir Darwis. (2024). Penentuan Premi Asuransi Gempa Berdasarkan Declustering Katalog Jawa Barat. *Jurnal Riset Statistika*, 127–136. <https://doi.org/10.29313/jrs.v4i2.5029>