Terbit online pada laman web jurnal: **https://jurnal.plb.ac.id/index.php/tematik/index**

TEMATIK

Jurnal **Teknologi Informasi Komunikasi (e-Journal)**

Vol. 12 No. 1 (2025) 59 - 71 ISSN Media Elektronik: 2443-3640

Analisis Klaster Daerah Rawan Gempa di Indonesia Menggunakan K-Means dan DBSCAN Berbasis Data Historis BMKG

Cluster Analysis of Earthquake Prone Areas in Indonesia Using K-Means and DBSCAN Based on BMKG Historical Data

Iqbal Prayoga Willyana[1*], Asep Id Hadiana[2], Ridwan Ilyas[3]
[1,2,3]Informatika, Fakultas Sains Dan Informatika, Universitas Jendral Achmad Yani
[1]Iqbalprayogaw20@if.unjani.ac.id[*], [2]asep.hadiana@lecture.unjani.ac.id, [3]ilyas@lecture.unjani.ac.id

Abstract

Indonesia is one of the countries with the highest levels of seismic activity in the world because it is located at the meeting point of three major plates. The high potential for earthquakes requires a data-based approach to map vulnerable areas more accurately. This study aims to group earthquake-prone areas in Indonesia using the K-Means and DBSCAN clustering algorithms. The dataset used includes spatial data (latitude, longitude) and seismic data (magnitude, depth, phasecount, azimuth_gap) obtained from the BMKG earthquake catalog for the period 2008–2025. The study begins with the data preprocessing stage, which includes data cleaning, type conversion, feature selection, missing value imputation, outlier detection and removal, and normalization. Furthermore, the clustering algorithm is applied in three main scenarios, namely spatial data, seismic data, and a combination of spatial and seismic data. Evaluation using Silhouette Score and Davies-Bouldin Index (DBI) metrics shows that the K-Means algorithm provides better cluster separation, with a DBI value of 1.2551 in the combined scenario, while the DBSCAN algorithm tends to form only one dominant cluster and is sensitive to the presence of outliers. The final result of this study produces a map of earthquake-prone areas in Indonesia, which are divided into several clusters with different risk characteristics. The cluster with the highest concentration includes areas such as western Sumatra, the southern coast of Java, and parts of Maluku and Papua, which have historically been recorded as having higher earthquake frequency and magnitude. Meanwhile, other clusters covering areas such as Kalimantan and parts of Sulawesi show lower seismic activity intensity.

Keywords: Clustering, K-Means, DBSCAN, Earthquake, Spatial, Seismic, Risk Mapping

Abstrak

Indonesia merupakan salah satu negara dengan tingkat aktivitas seismik tertinggi di dunia karena lokasinya berada di pertemuan tiga lempeng utama. Tingginya potensi gempa bumi menuntut adanya pendekatan berbasis data untuk memetakan wilayah rawan secara lebih akurat. Penelitian ini bertujuan untuk mengelompokkan daerah rawan gempa di Indonesia menggunakan algoritma klasterisasi K-Means dan DBSCAN. Dataset yang digunakan mencakup data spasial (latitude, longitude) dan data seismik (magnitude, depth, phasecount, azimuth_gap) yang diperoleh dari katalog gempa bumi BMKG periode 2008–2025. Penelitian diawali dengan tahapan praproses data, yang mencakup pembersihan data, konversi tipe, seleksi fitur, imputasi nilai hilang, deteksi dan penghapusan outlier, serta normalisasi. Selanjutnya, dilakukan penerapan algoritma klasterisasi dalam tiga skenario utama yaitu data spasial, data seismik, dan gabungan data spasial dan seismik. Evaluasi menggunakan metrik Silhouette Score dan Davies-Bouldin Index (DBI) menunjukkan bahwa algoritma K-Means memberikan pemisahan klaster yang lebih baik, dengan nilai DBI sebesar 1,2551 pada skenario gabungan, sementara algoritma DBSCAN cenderung hanya membentuk satu klaster dominan dan sensitif terhadap keberadaan outlier. Hasil akhir dari penelitian ini menghasilkan peta pengelompokkan daerah rawan gempa di Indonesia, yang terbagi ke dalam beberapa klaster dengan karakteristik risiko berbeda. Klaster dengan konsentrasi tertinggi mencakup wilayah-wilayah seperti Sumatra bagian barat, pantai selatan Jawa, serta sebagian wilayah Maluku dan Papua, yang secara historis tercatat memiliki frekuensi kejadian gempa dan magnitude yang lebih tinggi. Sementara itu, klaster lain yang mencakup wilayah seperti Kalimantan dan sebagian Sulawesi menunjukkan intensitas aktivitas seismik yang lebih rendah.

Kata Kunci: Klasterisasi, K-Means, DBSCAN, Gempa Bumi, Spasial, Seismik, Pemetaan Risiko

1. Pendahuluan

Indonesia merupakan negara kepulauan yang berada pada pertemuan tiga lempeng tektonik utama dunia, yaitu Lempeng Eurasia, Lempeng Indo-Australia, dan Lempeng Pasifik, serta didukung oleh aktivitas dari Lempeng Filipina sebagai lempeng minor. Kondisi geotektonik ini menyebabkan Indonesia memiliki tingkat kerawanan gempa bumi yang sangat tinggi [1].

Diterima Redaksi: 24-05-2025 | Selesai Revisi: 04-06-2025 | Diterbitkan Online: 25-06-2025

Berdasarkan laporan dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), tercatat lebih dari 11.000 kejadian gempa bumi dengan magnitudo di atas 5,0 skala moment terjadi di wilayah Indonesia dalam dua dekade terakhir. Fenomena tersebut berdampak besar tidak hanya terhadap kerusakan infrastruktur melainkan juga terhadap korban jiwa yang signifikan.

Beberapa kejadian gempa bumi dengan skala besar telah memberikan dampak yang luas, seperti gempa bumi di wilayah Aceh pada tahun 2004 yang disusul dengan tsunami yang menyebabkan lebih dari 230.000 korban jiwa. Kemudian gempa bumi di wilayah Yogyakarta pada tahun 2006 yang menyebabkan lebih dari 5.700 korban jiwa dan kerusakan infrastruktur secara masif. Kejadian gempa bumi lainnya seperti di wilayah Lombok dan Palu pada tahun 2018 juga menunjukkan betapa tingginya tingkat kerentanan wilayah Indonesia terhadap bencana gempa bumi [2]. Rangkaian peristiwa tersebut menegaskan pentingnya penerapan strategi mitigasi bencana terhadap zona rawan gempa bumi di Indonesia berbasis data dan teknologi untuk meminimalisasi dampak yang mungkin terjadi.

Identifikasi wilayah rawan gempa sangat penting sebagai dasar pengambilan keputusan dalam berbagai aspek kebijakan, seperti penguatan infrastruktur publik, perencanaan jalur evakuasi, serta penyusunan rencana tata ruang dan pembangunan berkelanjutan yang mempertimbangkan risiko bencana. Dengan mengetahui distribusi spasial dari zona-zona yang memiliki kerawanan tinggi, upaya mitigasi dapat dilakukan secara lebih terarah dan efektif [3].

Kemajuan teknologi informasi dan komputasi telah menghadirkan pendekatan baru dalam proses analisis data gempa bumi, salah satunya melalui penerapan algoritma *machine learning*. Salah satu teknik *machine learning* yang berkembang pesat adalah metode klasterisasi, yang memungkinkan analisis pola-pola kejadian gempa bumi berdasarkan karakteristik seismik yang serupa untuk identifikasi zona rawan gempa bumi [1]. Berbeda dengan pendekatan konvensional yang umumnya bersifat manual dan membutuhkan keahlian seismotektonik yang kurang unggul berdasarkan hal objektifitas, efisiensi waktu, dan kemampuan mengolah data dalam skala besar [4] [5].

Algoritma K-Means *clustering* merupakan salah satu metode populer dalam pengelompokan data berbasis partisi yang telah digunakan secara luas dalam berbagai kasus, termasuk untuk data spasial. Metode ini bekerja dengan menetapkan pusat klaster (*centroid*) dan mengelompokkan data berdasarkan jarak terdekat ke pusat klaster menggunakan proses iteratif [6]. Namun demikian, K-Means memiliki keterbatasan dalam mengidentifikasi klaster dengan bentuk yang tidak teratur serta ketidakmampuannya menangani *outlier* secara efektif. Untuk mengatasi hal tersebut, terdapat algoritma DBSCAN (Density-Based Spatial Clustering of Application with Noise) sebagai metode alternatif. Metode ini mengelompokkan data berdasarkan kepadatan dan memiliki keunggulan dalam mendeteksi *cluster* dengan bentuk tidak teratur hingga mengidentifikasi *noise* dan *outlier* secara alami [4]. Hasil dari kedua pendekatan ini akan memberikan hasil analisis spasial yang saling melengkapi dan komprehensif.

Beberapa penelitian terdahulu telah menunjukkan potensi teknik klasterisasi dalam bidang seismologi. Penelitian oleh Zamani et al. [7] di Iran menggunakan K-Means untuk mengelompokkan wilayah berdasarkan kedalaman dan magnitudo gempa, yang menunjukkan hasil sebaran zona seismik yang berkorelasi dengan struktur tektonik. Di Indonesia, Kristanto et al. [8] telah melakukan penelitian menggunakan metode K-Means untuk menganalisis intensitas gempa bumi di Pulau Jawa dan berhasil memvisualisasikan zona rawan gempa melalui geomap. Hasil penelitian ini menunjukkan bahwa wilayah selatan Pulau Jawa memiliki tingkat kerawanan gempa yang lebih tinggi dibandingkan dengan wilayah utara, yang sesuai dengan posisi tektonik wilayah tersebut yang berada di dekat zona subduksi. Penelitian lain yang dilakukan oleh Natawidjaja et al. [9] mengaplikasikan metode *clustering* untuk mengidentifikasi pola-pola seismisitas di sepanjang Sesar Sumatra. Dengan menggunakan data gempa historis yang komprehensif, penelitian ini berhasil mengidentifikasi segmen-segmen sesar dengan tingkat aktivitas seismik yang berbeda-beda yang memberikan informasi penting untuk penilaian risiko gempa bumi di wilayah tersebut.

Selanjutnya penelitian yang dilakukan oleh R. Indra Marihot et al. [5] membandingkan algoritma K-Means dan DBSCAN pada data gempa BKMG tahun 2022. Hasil yang diperoleh menunjukkan bahwa DBSCAN lebih adaptif dalam mendeteksi pola distribusi dengan kepadatan variatif. Sementara itu, M. Jalaludin Shofa et al. [10] menggunakan pendekatan serupa untuk data United States Geological Survey (USGS) 2004 – 2023 dan menyimpulkan bahwa DBSCAN lebih cocok untuk data dengan sebaran tidak seragam, sedangkan K-Means bekerja baik pada data homogen. Penelitian oleh S. Marzuki et al. [11] di Samosir menunjukkan bahwa kombinasi metode K-Means dan DBSCAN dapat memberikan wawasan lokal yang akurat terhadap wilayah seismik sempit.

Pendekatan lain dilakukan oleh A. Rohan et al. [12] yang membandingkan K-Means, CLARA, dan K-Medoids menggunakan dataset USGS global. Sementara itu, A. Hanifullah [13] menerapkan K-Medoids untuk wilayah Sumatera bagian utara, menunjukkan efektivitas metode tersebut dalam membentuk klaster spasial yang kuat. Penelitian lain oleh G. Sri Widyanti et al. [14] memperluas pendekatan

DOI: https://doi.org/10.38204/tematik.v12i1.2369
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

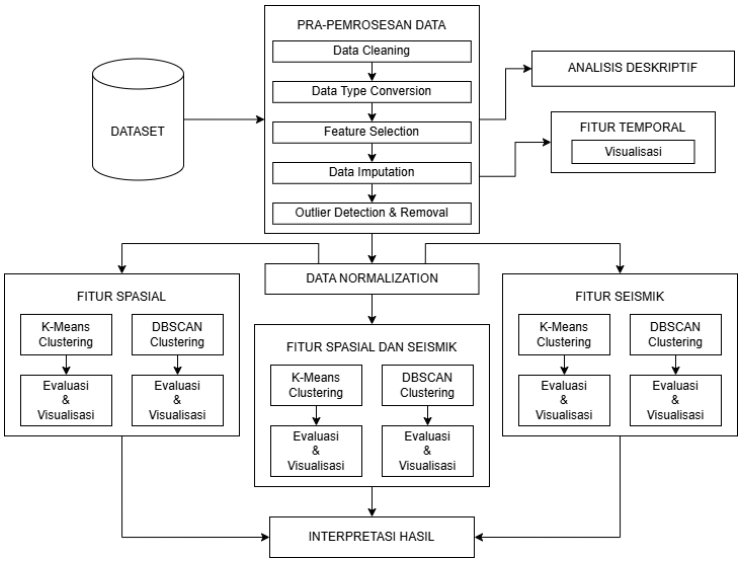
menggunakan K-Means untuk mengelompokkan kelompok penyintas trauma gempa, menunjukkan aplikasi lintas disiplin dari metode klasterisasi. Kemudian pendekatan prediktif dikaji oleh Wijaya et al. [15] yang memanfaatkan *time series* dalam memprediksi gempa bumi, adapun P. Hasih et al. [16] yang menggunakan *hierarchical clustering* untuk penilaian risiko gempa di Sumatera. Terakhir penelitian oleh W. Anton et al. [17] mengkombinasikan K-Means, DBSCAN, dan Fuzzy C-Means untuk data North Sumatra dan menghasilkan pemetaan risiko yang lebih fleksibel dan multikriteria.

Meskipun demikian, penelitian-penelitian sebelumnya masih memiliki keterbatasan yang dapat dijadikan sebagai celah penelitian. Umumnya, penelitian terdahulu hanya memanfaatkan satu jenis metode klasterisasi tanpa membandingkan efektivitas antar algoritma secara komprehensif. Selain itu, cakupan wilayah analisis sering kali terbatas pada area tertentu dan belum mencerminkan seluruh wilayah Indonesia secara spasial. Penelitian yang mengintegrasikan aspek geologi dan tektonik ke dalam proses klasterisasi data gempa masih realtif minim dan jarang ditemukan.

Oleh karena itu, penelitian ini dilakukan menggunakan algoritma K-Means dan DBSCAN untuk melakukan identifikasi dalam memetakan zona rawan gempa di seluruh wilayah Indonesia, berdasarkan data historis kejadian gempa bumi. Pendekatan ini diharapkan mampu menghasilkan pemetaan zona rawan gempa yang lebih akurat dan komprehensif yang dapat mendukung tenaga ahli dan masyarakat dalam melakukan upaya mitigasi bencana gempa bumi yang lebih efektif di Indonesia

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dan eksploratif yang saling berhubungan dalam proses analisis dan pemahaman data spasial dan seismik terkait aktivitas gempa bumi dalam pengelompokkan daerah rawan gempa. Pendekatan kuantitatif digunakan untuk melakukan analisis numerik terhadap data gempa menggunakan metrik evaluasi yang terukur, sedangkan pendekatan eksploratif digunakan untuk menggali pola tersembunyi dan struktur kelompok yang tidak diketahui sebelumnya. Klasterisasi dilakukan dengan algoritma K-Means dan DBSCAN yang termasuk dalam metode *unsupervised learning*, yaitu teknik pembelajaran tanpa label kelas.



Gambar 1. Desain Penelitian Klasterisasi Daerah Rawan Gempa

Proses penelitian ini dibagi ke dalam beberapa tahapan utama sebagaimana ditunjukkan dalam diagram alur pada Gambar 1.

2.1. Dataset

Data historis gempa bumi yang digunakan berasal dari Preliminary Earthquake Catalog dengan rentang waktu yang dimulai pada 1 November 2008 hingga 30 Maret 2025. Data ini dikelola oleh Lembaga BMKG (Badan Meteorologi Klimatologi dan Geofisika) di Indonesia (https://github.com/kekavigi/repo-gempa/tree/main/katalog_gempa_v2), memiliki 37 kolom atribut dengan 121.716 baris data yang merepresentasi data mekanisme fokus gempa yang disimpan menggunakan *.tsv sebagai ekstensi filenya.

Dari 37 atribut ini, yang akan dipilih hanyalah 6 atribut yaitu *latitude, longitude, magnitude, depth, phasecount, dan azimuth_gap* untuk proses *clustering* hingga mendapatkan hasil akhir dari *clustering*, serta 2 atribut tambahan yaitu *datetime* dan *location* untuk analisis geospasial dan temporal.

Pemilihan 6 atribut utama latitude, longitude, magnitude, depth, phasecount, dan azimuth_gap

DOI: https://doi.org/10.38204/tematik.v12i1.2369
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

dilandaskan pada pertimbangan seismologis yang esensial dalam analisis gempa bumi. Latitude dan longitude merepresentasikan lokasi episentrum yang krusial untuk mengidentifikasi distribusi spasial aktivitas seismik di wilayah tektonik aktif seperti Indonesia. Magnitude mencerminkan besar energi yang dilepaskan, menjadi indikator utama dalam membedakan dampak gempa dan potensinya. Depth berperan dalam menentukan jenis dan karakteristik sumber gempa, karena gempa dangkal lebih merusak dibanding gempa dalam. Phasecount menunjukkan jumlah fase seismik yang tercatat, yang berbanding lurus dengan keandalan penentuan lokasi gempa. Sedangkan azimuth_gap mengindikasikan cakupan sudut stasiun seismik terhadap episentrum; semakin kecil nilainya, semakin tinggi kualitas data lokasi gempa. Keenam atribut ini tidak hanya menyediakan dimensi spasial, energi, dan kedalaman gempa, tetapi juga mencerminkan kualitas observasi seismik, menjadikannya relevan secara ilmiah dalam pemodelan dan pengelompokan kejadian seismic.

2.2. Praprosess Data

Praproses data merupakan tahap penting yang harus dilakukan sebelum analisis lebih lanjut pada proses klasterisasi. Hal itu karena praproses data berperan sebagai fondasi utama dalam pembelajaran mesin yang umumnya menentukan efektivitas untuk proses selanjutnya [18]. Tahap ini umumya melakukan penyesuaian terhadap data dan meningkatkan kualitasnya agar hasil analisis menjadi lebih akurat dan representatif.

Ada enam tahap praproses yang dilakukan pada penelitian ini, yaitu pembersihan data (*data cleaning*), konversi tipe data (*data type conversion*), seleksi fitur (*feature selection*), imputasi data (*data imputation*), mendeteksi dan menghapus *outlier* (*outlier detection and removal*) dan normalisasi data (*data normalization*).

Data Cleaning merupakan tahap awal dalam praproses data yang bertujuan untuk meningkatkan kualitas dataset dengan cara menghilangkan, memperbaiki atau mengoreksi entri data yang tidak valid, duplikat, atau tidak konsisten. Dalam penelitian ini, tahap *data cleaning* dilakukan dengan pendekatan berbasis statistik deskriptif untuk mengidentifikasi atribut yang tidak sesuai. Tahap ini melalui dua proses utama, yaitu penghapusan atribut dengan persentase *missing values* melebihi ambang batas yang dipilih (pada penelitian ini 50%) serta eliminasi data duplikat.

Data Type Conversion merupakan langkah penting dalam tahap praproses data, terutama dalam konteks pembelajaran mesin dan analisis data. Dalam penelitian ini, proses konversi tipe data dilakukan dengan mengidentifikasi atribut yang memiliki tipe data tidak sesuai, seperti *string* yang seharusnya numerik, dan kemudian dilakukan konversi ke tipe data yang tepat. Hal ini dilakukan untuk mencegah kesalahan dalam proses analisis selanjutnya, seperti kesalahan dalam perhitungan jarak antar data atau kesalahan dalam penentuan pusat klister.

Data Imputation adalah teknik yang digunakan untuk menangani nilai hilang (*missing values*) dalam dataset agar tidak mengganggu proses analisis data dan pembelajaran mesin. Dalam konteks data gempa bumi, atribut seperti magnitudo, kedalaman, atau jumlah fase pengamatan dapat mengalami kehilangan data karena keterbatasan alat pencatatan atau faktor lingkungan. Oleh karena itu, metode imputasi dengan pendekatan median dipilih karena lebih *robust* terhadap *outlier* dibandingkan rata-rata (mean).

Outlier atau nilai pencilan merupakan observasi dalam dataset yang secara signifikan berbeda dari pola umum data lainnya. Salah satu metode statistik yang umum digunakan untuk mendeteksi *outlier* adalah Inter-Quartile Range (IQR). IQR mengukur rentang tengah dari distribusi data dan dihitung sebagai selisih antara kuartil ketiga (Q3) dan kuartil pertama (Q1) yang dapat dilihat pada Persamaan (1) berikut :

$$IQR = Q3 - Q1 \quad (1)$$

Dengan menggunakan IQR, batas bawah menggunakan Persamaan (5) dan batas atas menggunakan Persamaan (6) ditentukan untuk mengidentifikasi nilai-nilai yang dianggap *outlier*.

Data Normalization atau normalisasi data merupakan proses transformasi nilai-nilai fitur ke dalam skala tertentu yang seragam, sehingga semua atribut memiliki kontribusi yang setara dalam proses analisis. Salah satu teknik normalisasi yang umum digunakan adalah Min-Max Normalization yang mentransformasikan setiap nilai ke dalam rentang antara 0 hingga 1. Formula dari normalisasi Min-Max dapat dilihat pada Persamaan (2) sebagai berikut :

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2)$$

Dimana X adalah nilai asli dan X' adalah nilai yang telah dinormalisasi. Kemudian X_{min} dan juga X_{max} merupakan nilai minimum dan nilai maksimum dari atribut tersebut

2.3. Algoritma

Klasterisasi atau Clustering merupakan teknik dalam *data mining* yang bertujuan untuk mengelompokkan sekumpulan objek berdasarkan tingkat kesamaan diantara mereka. Salah satu studi memberikan tinjauan komprehensif terhadap algoritma K-Means, termasuk variannya dan penerapannya dalam era big data. Studi ini menyoroti tantangan dan solusi dalam menerapkan

DOI: https://doi.org/10.38204/tematik.v12i1.2369
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

K-Means pada dataset besar dan komplek [19]. Kemudian terdapat tinjauan statistik modern terhadap analisis klaster, membahas metode seperti klasterisasi *agglomeratif hierarkis*, K-Means, dan model campuran, serta topik terkait lainnya yang memberikan wawasan mendalam tentang perkembangan terkini dalam teknik klasterisasi dan aplikasinya dalam berbagai domain [20]. Pemilihan metode klasterisasi yang tepat sangat bergantung pada karakteristik data dan tujuan analisis.

K-Means adalah algoritma klasterisasi partisional yang bertujuan membagi data ke dalam k kelompok berdasarkan kemiripan fitur dengan meminimalkan variasi intra-klaster [21]. Algoritma ini bekerja secara iteratif untuk menetapkan pusat klaster (*centroid*) dan mengelompokkan data berdasarkan kedekatan terhadap pusat tersebut. Langkah - langkah utama dalam algoritma K-Means adalah sebagai berikut.

Inisialisasi, menentukan jumlah klaster k dan memilih pusat klaster (*centroid*) awal secara acak. Penugasan Klaster, menghitung jarak antara setiap data ke semua *centroid* menggunakan perhitungan Euclidean Distance seperti pada Persamaan (3) berikut :

$$d(x_i, \mu_j) = \sqrt{\sum_{n=1}^{N} (x_{in} - \mu_{jn})^2} \quad (3)$$

Di mana x_i adalah vector data ke-i, μ_j adalah *centroid* dari klaster ke-j dan N adalah jumlah fitur.

Persamaan (4) merupakan fungsi objektif utama yang diminimalkan dalam K-Means, sebagai berikut :

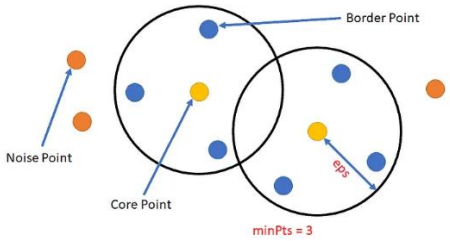
$$J = \sum_{i=1}^{K} \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \quad (4)$$

Di mana C_i adalah himpunan data dalam klaster ke-i, x_j adalah data dalam klaster tersebut dan μ_i adalah *centroid*.

Meskipun K-Means memiliki keunggulan dalam efisiensi komputasi dan kemudahan implementasi, algoritma ini sensitif terhadap inisialisasi *centroid* dan kesulitan menangani *outlier* atau klaster yang tidak berbentuk sferis.

DBSCAN merupakan algoritma klasterisasi berbasis kepadatan (*density-based clustering*) yang dirancang untuk mengelompokkan titik-titik data berdasarkan distribusi spasialnya tanpa memerlukan penentuan jumlah klaster di awal seperti pada K-Means. DBSCAN bekerja dengan mengidentifikasi area dengan kepadatan tinggi (*dense regions*) sebagai klaster dan memisahkan area dengan kepadatan rendah sebagai *outlier* atau *noise* [22].

Algoritma ini menggunakan dua parameter utama, yaitu epsilon (ε) sebagai radius pencarian tetangga dan $MinPts$ sebagai jumlah minimum titik dalam radius tersebut yang dibutuhkan untuk membentuk sebuah klaster. Berdasarkan kedua parameter tersebut, titik-titik data dikategorikan menjadi seperti pada Gambar 2 dibawah ini.



Gambar 2. Kategori Titik DBSCAN

(Sumber : https://medium.com/%40nitingour1203/dbscan-algorithm-f14e634abb02)

Core Point, merupakan titik yang memiliki setidaknya sejumlah $MinPts$ tetangga dalam radius ε. Border Point, merupakan titik yang berada dalam radius ε dari Core Point, tetapi memiliki tetangga kurang dari $MinPts$. Noise Point (Outlier), merupakan titik yang tidak termasuk dalam Core Point dan Border Point.

Dua titik p dan q dianggap *density-reachable* jika terdapat jalur dari p ke q melalui serangkaian Core Point sehingga setiap titik bertetangga dalam radius ε satu sama lain. Proses ini yang memungkinkan DBSCAN mendeteksi klaster dengan bentuk arbitrer serta *outlier* yang tidak dapat ditangani oleh pendekatan berbasis *centroid* seperti K-Mean [23].

Persamaan (5) merupakan persamaan untuk menghitung jumlah titik tetangga dalam radius epsilon (ε) dari titik x_i.

$$N_\varepsilon(x_i) = \{x_j \in D \mid dist(x_i, x_j) \leq \varepsilon\} \quad (5)$$

Di mana $dist$ dapat berupa metrik jarak seperti Euclidean Distance. Kemudian titik x_i dianggap sebagai *core point* jika :

$$|N_\varepsilon(x_i)| \geq MinPts$$

Meskipun DBSCAN unggul dalam menangani bentuk klaster yang kompleks dan mendeteksi *outlier* secara otomatis, algoritma ini memiliki kelemahan pada sensitivitas terhadap parameter epsilon (ε) dan $MinPts$ serta data dengan dimensi tinggi yang dapat menurunkan performa saat diterapkan.

2.4. Evaluasi

Pengukuran kinerja algoritma klasterisasi merupakan tahap penting dalam proses evaluasi untuk menilai sejauh mana hasil pengelompokan data sesuai dengan struktur alami di dalam data. Berbeda dengan

DOI: https://doi.org/10.38204/tematik.v12i1.2369
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

supervised learning, klasterisasi tidak memiliki label yang dapat dijadikan acuan evaluasi langsung, sehingga diperlukan metrik evaluasi internal yang berdasar pada karakteristik pembentukan klaster itu sendiri, seperti seberapa padat dan seberapa terpisah cluster yang terbentuk.

Silhouette Score adalah salah satu metrik evaluasi internal yang paling umum digunakan dalam klasterisasi untuk menilai seberapa baik objek berhasil cocok dengan klasternya sendiri dibandingkan dengan klaster lain. Metrik ini tidak memerlukan label data, sehingga sangat sesuai untuk evaluasi pada *unsupervised learning*. Nilai Silhouette mengukur kesamaan suatu titik data dengan klaster tempatnya berada dibandingkan dengan klaster terdekat lainnya. Secara matematis, Persamaan (6) merupakan persamaan yang digunakan untuk menghitung nilai Silhouette untuk setiap titik data i.

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \quad (6)$$

Dengan $a(i)$ adalah rata-rata jarak antara data i dengan seluruh data lain dalam klaster yang sama (intra-klaster), $b(i)$ adalah rata-rata jarak terendah antara data i dengan semua data dalam klaster lain (inter-klaster terdekat). Nilai Silhouette $s(i)$ berada dalam rentang $[-1, 1]$. $s(i) \approx 1$, berarti titik data sangat sesuai dengan klasternya dan jauh dari klaster lainnya, $s(i) \approx 0$, berarti titik data berada di perbatasan dua klister, $s(i) \approx -1$, berarti titik data lebih dekat ke klaster lain dibandingkan klasternya sendiri (keliru dalam penempatan klaster).

Davies-Bouldin Index (DBI) adalah salah satu metrik evaluasi internal yang digunakan untuk mengukur kualitas hasil klasterisasi tanpa memerlukan label data. DBI mengevaluasi seberapa baik klaster dipisahkan dan seberapa kompak anggota klaster tersebut. Nilai DBI yang lebih rendah mengindikasikan hasil klasterisasi yang lebih baik, karena menunjukkan bahwa klaster memiliki variansi internal yang rendah dan berada pada jarak yang cukup jauh dari klaster lainnya [24].

Secara matematis, indeks ini dihitung berdasarkan rasio antara *intra-cluster dispersion* (rata-rata jarak antar titik dalam satu klaster terhadap *centroid*-nya) dan *inter-cluster separation* (jarak antar centroid klaster). Rumus perhitungan DBI dapat dilihat pada Persamaan (7) berikut :

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{j \neq i}\left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)}\right) \quad (7)$$

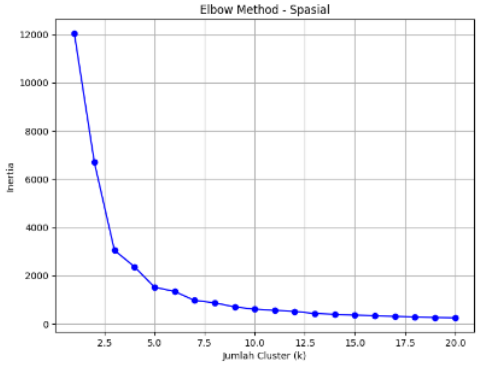
Dengan $k =$ jumlah klister, $\sigma_i =$ jarak rata-rata antar titik pada klaster i ke *centroid* nya, $c_i, c_j =$ *centroid* dari klaster i dan $jd(c_i, c_j) =$ jarak antar *centroid* klaster i dan $j, max_{j \neq i} =$ mencari klaster lain yang paling mirip dengan klaster i

3. Hasil dan Pembahasan

Proses klasterisasi berdasarkan fitur spasial latitude dan longitude yang telah dinormalisasi dilakukan terhadap 99.183 data gempa menggunakan algoritma K-Means dan DBSCAN untuk mengidentifikasi zona rawan gempa di Indonesia. Hasil klasterisasi K-Means menunjukkan pemisahan wilayah yang cukup selaras dengan kondisi tektonik Indonesia, di mana klaster terbentuk mengikuti zona subduksi di sepanjang pantai barat Sumatra dan Jawa, serta zona kompleks di Sulawesi, Maluku, dan Papua yang sesuai dengan interaksi antar lempeng aktif seperti Indo-Australia, Eurasia, dan Pasifik. Sebaliknya, DBSCAN gagal membentuk klaster yang bermakna karena tantangan inheren pada data spasial Indonesia, seperti ketidakteraturan kepadatan data antar wilayah, skala geografi yang luas, dan bentuk sebaran gempa yang memanjang linear mengikuti jalur tektonik—semua ini menyulitkan penentuan parameter eps dan min_samples yang optimal secara global. Akibatnya, DBSCAN mengklasifikasikan sebagian besar data sebagai noise atau hanya membentuk klaster di area sangat padat, yang tidak mencerminkan struktur geologis yang sebenarnya. Perbandingan nilai evaluasi antar metode—seperti silhouette score yang lebih tinggi pada K-Means dan persentase noise tinggi pada DBSCAN—dapat dirangkum dalam tabel untuk memperjelas efektivitas masing-masing pendekatan. Selain itu, visualisasi hasil klasterisasi dalam bentuk peta Indonesia akan sangat memperkaya interpretasi dan menjembatani pemahaman pembaca non-teknis terhadap keterkaitan spasial antara gempa dan kondisi geologi Indonesia.

3.1. Klasterisasi Spasial

Untuk menentukan jumlah klaster optimal pada algoritma K-Means, digunakan metode Elbow yang digambarkan pada Gambar 3 Hasil visualisasi menunjukkan bahwa titik siku (*elbow point*) berada pada $k = 3$, yang menandakan titik optimum di mana penurunan nilai inertia tidak lagi signifikan.



Gambar 3. Elbow Method Spasial

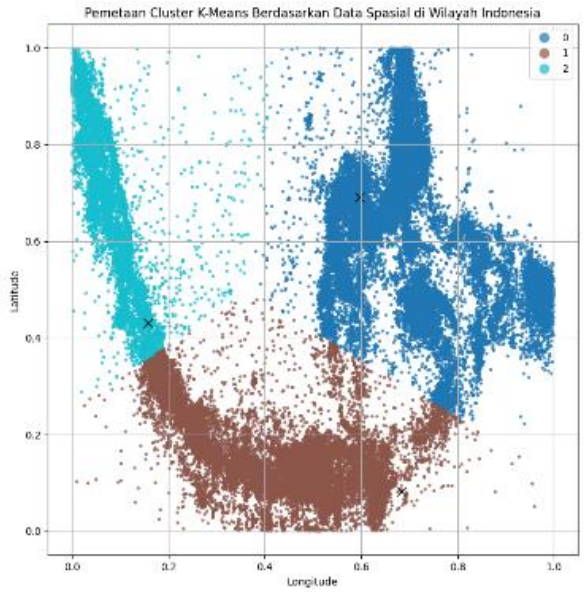
DOI: https://doi.org/10.38204/tematik.v12i1.2369
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Selanjutnya, untuk memperkuat hasil pemilihan nilai k, dilakukan evaluasi menggunakan metrik *Silhouette Score* pada rentang nilai $k = 2$ hingga $k = 5$. Hasil evaluasi ditunjukkan pada Tabel 1.

Tabel 1. Silhouette Score K-Means Spasial

Nilai k	Silhouette Score
2	0.4843
3	**0.5846**
4	0.5525
5	0.5202

Berdasarkan hasil tersebut, nilai k terbaik adalah 3 karena memiliki skor Silhouette tertinggi. Hasil klasterisasi ditampilkan pada Gambar 4, di mana tiga zona klaster teridentifikasi secara jelas. Setiap klaster memiliki sebaran lokasi yang menggambarkan segmentasi wilayah geografis berdasarkan kemiripan posisi.



Gambar 4. Visualisasi K-Means Clustering Data Spasial

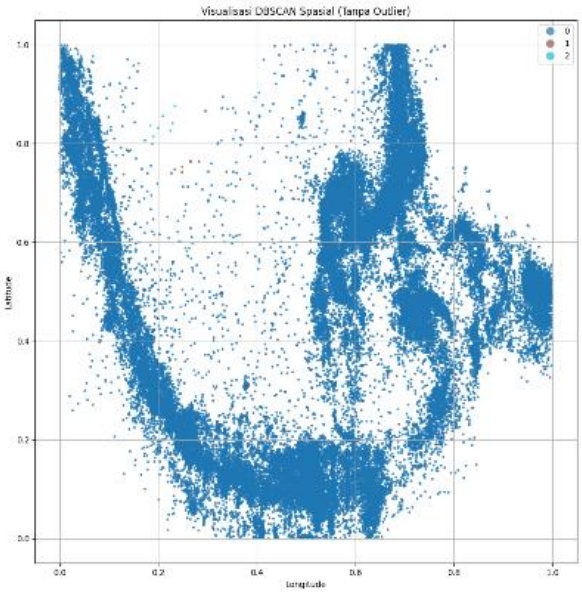
Evaluasi terhadap kualitas klaster dilakukan menggunakan Davies-Bouldin Index (DBI) yang menghasilkan nilai sebesar 0.5753. Nilai ini menunjukkan bahwa pemisahan antar klaster cukup baik dan memiliki kompaksi internal yang relatif kuat.

DBSCAN menggunakan pendekatan berbasis kerapatan sehingga diperlukan dua parameter utama : epsilon dan min_samples. Estimasi nilai epsilon dilakukan menggunakan K-Distance Graph pada berbagai nilai min_samples. Dari hasil estimasi yang dilakukan, kombinasi terbaik diperoleh pada nilai epsilon = 0.0354 dan min_samples = 6, yang menghasilkan skor Silhouette sebesar 0.0899.

Dalam algoritma DBSCAN, pemilihan parameter epsilon (ε) dan min_samples sangat penting karena keduanya menentukan batas jarak antar titik dan jumlah minimum tetangga yang diperlukan untuk membentuk suatu klaster. Dalam studi ini, nilai epsilon sebesar 0.0354 dan min_samples sebesar 6 dipilih berdasarkan analisis K-distance plot, yaitu grafik yang menunjukkan jarak tiap titik ke tetangga terdekatnya yang ke-6, kemudian diurutkan untuk menemukan titik tekukan (elbow) sebagai indikasi nilai epsilon optimal. Nilai min_samples = 6 dipilih karena sesuai dengan praktik umum pada data berdimensi dua, cukup sensitif untuk mendeteksi klaster lokal tanpa terlalu mudah menganggap titik sebagai noise. Kombinasi ini menghasilkan silhouette score sebesar 0.0899, yang tergolong rendah dan menunjukkan bahwa klaster yang terbentuk belum terpisah dengan baik. Hal ini disebabkan oleh karakteristik data gempa di Indonesia yang memiliki distribusi spasial sangat bervariasi, di mana beberapa wilayah sangat padat dan lainnya jarang, sehingga sulit bagi DBSCAN yang menggunakan satu nilai epsilon global untuk secara konsisten mengenali semua pola klaster.

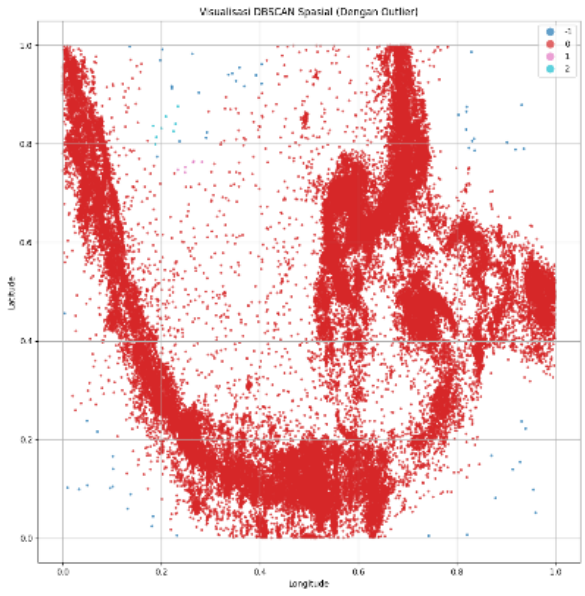
Hasil klasterisasi DBSCAN ditunjukkan dalam dua visualisasi. Berdasarkan Gambar 5 dan 6, bahwa DBSCAN mendeteksi 3 klaster dengan 57 data yang terdeteksi sebagai outlier. Evaluasi DBI terhadap hasil klaster tanpa *outlier* menunjukkan nilai sebesar 0.7958, yang relatif lebih tinggi dari K-Means, mengindikasikan bahwa struktur klaster kurang kompak dibandingkan K-Means.

Perbandingan antara metode K-Means dan DBSCAN menunjukkan bahwa K-Means lebih unggul dalam hal pembentukan klaster spasial dengan nilai Silhouette Score yang lebih tinggi (**0.5846** vs 0.0899) dan Davies-Bouldin Index yang lebih rendah (**0.5753** vs 0.7958).



Gambar 5. Visualisasi DBSCAN Spasial Tanpa Outlier

DOI: https://doi.org/10.38204/tematik.v12i1.2369
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)



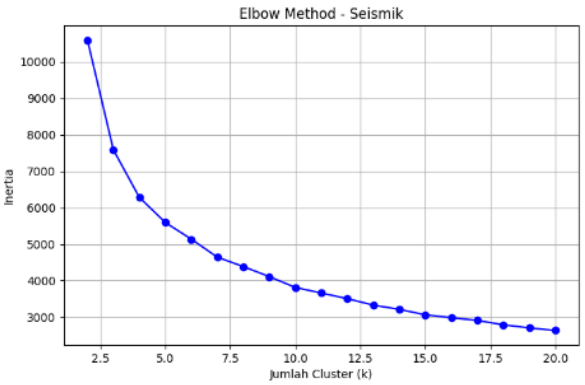
Gambar 6. Visualisasi DBSCAN Spasial Dengan Outlier

3.2. Klasterisasi Seismik

Pada bagian ini, dilakukan pengelompokkan data seismik gempa bumi berdasarkan atribut *magnitude*, *depth*, *phasecount*, dan *azimuth_gap*. Pemilihan fitur ini didasarkan pada relevansi karakteristik seismik yang secara langsung memengaruhi potensi risiko dan intensitas gempa. Atribut-atribut ini mewakili dimensi penting : kekuatan kejadian (*magnitude*), kedalaman pusat gempa (*depth*), jumlah fase gelombang yang tercatat (*phasecount*), dan distribusi arah gelombang (*azimuth_gap*), yang secara bersama-sama memberikan gambaran menyeluruh terkait perilaku gempa di wilayah Indonesia.

Analisis K-Means pada data seismik diawali dengan menentukan jumlah *cluster* optimal menggunakan metode Elbow yang dapat dilihat pada Gambar 7 dibawah ini. Metode ini memanfaatkan grafik hubungan antara jumlah *cluster* (k) dan nilai *inertia*, yaitu total jarak kuadrat antar titik data dengan pusat *cluster* terdekat. Berdasarkan kurva, terlihat bahwa seiring bertambahnya k, nilai *inertia* menurun tajam pada awalnya, tetapi mulai melandai setelah titik tertentu. Fenomena ini dikenal sebagai *elbow point*, yaitu titik optimal di mana penambahan jumlah cluster selanjutnya tidak lagi memberikan penurunan *inertia* yang signifikan.

Elbow Point teridentifikasi pada k = 4 sehingga dijadikan untuk pemilihan jumlah *cluster*. Penetapan k yang tepat penting karena secara langsung memengaruhi kualitas segmentasi data. Jika k terlalu kecil dapat menyebabkan *underfitting* (menggabungkan data berbeda ke dalam satu kelompok), sedangkan k terlalu besar berisiko *overfitting* (membuat *cluster* yang tidak bermakna secara nyata). Validasi dilakukan menggunakan Silhouette Score dengan hasil menunjukkan skor tertinggi sebesar 0.2654 pada k = 4 yang mengindikasikan pemisahan *cluster* sedang (moderat). Selain itu, perhitungan Davies-Bouldin Index (DBI) menghasilkan nilai 1.2017. Distribusi jumlah data per *cluster* ditunjukkan pada Tabel 2.

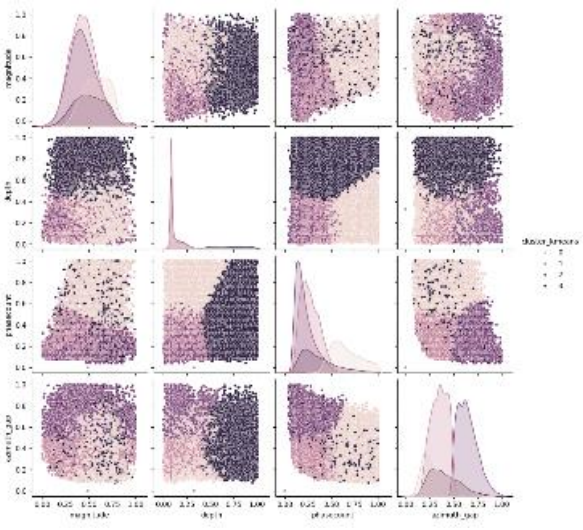


Gambar 7. Elbow Method Seismik

Tabel 2. Distribusi Jumlah Data Sseismik Setiap Cluster

Cluster ke-	Jumlah Data
0	18.418
1	35.849
2	31.798
3	11.136

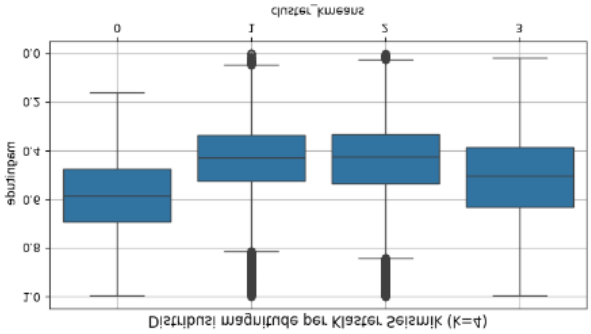
Karakteristik dari masing-masing *cluster* dilakukan analisis visualisasi pola distribusi dengan tujuan memperkaya informasi. Gambar Pairplot Seismik K-Means yang ditunjukkan pada Gambar 8, memperlihatkan secara umum bagaimana sebaran dan pola keterkaitan antar atribut di setiap *cluster*. Sementara itu, analisis yang lebih rinci setiap fitur atau atribut ditampilkan dalam bentuk visualisasi boxplot yang dapat dilihat pada Gambar 9 untuk fitur *magnitude*, Gambar 10 untuk fitur *depth*, Gambar 11 untuk fitur *phasecount*, dan Gambar 12 untuk fitur *azimuth_gap*.



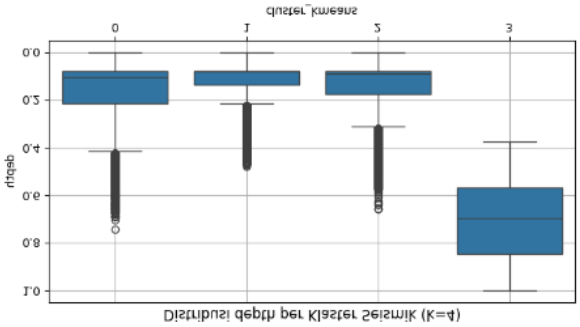
Gambar 8. Visualisasi Pairplot K-Means Data Seismik

DOI: https://doi.org/10.38204/tematik.v12i1.2369
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

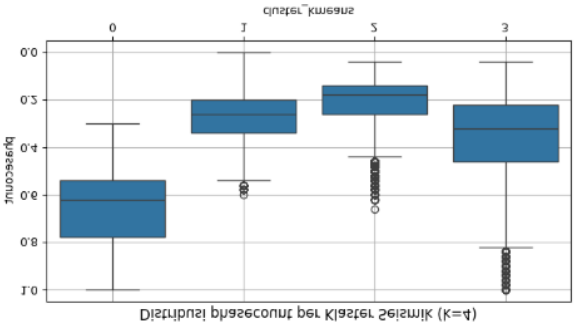
Hasil analisis menunjukkan bahwa masing-masing *cluster* memiliki pola dan karakteristik unik yang tercermin pada variasi nilai *magnitude*, *depth*, *phasecount*, dan *azimuth_gap*. Pola distribusi antar *cluster* terlihat jelas, baik dari sisi persebaran data maupun perbedaan nilai antar atribut. Visualisasi boxplot memperlihatkan bahwa *magnitude* memiliki penyebaran yang relatif merata, *depth* menunjukkan konsentrasi kedalaman yang lebih besar pada Cluster 3, *phasecount* mendominasi pada Cluster 0, sedangkan *azimuth_gap* memperlihatkan persebaran yang lebih luas namun tidak terlalu kontras antar *cluster*.



Gambar 9. Visualisasi Boxplot Untuk Atribut Magnitude Terhadap Masing-Masing Klaster

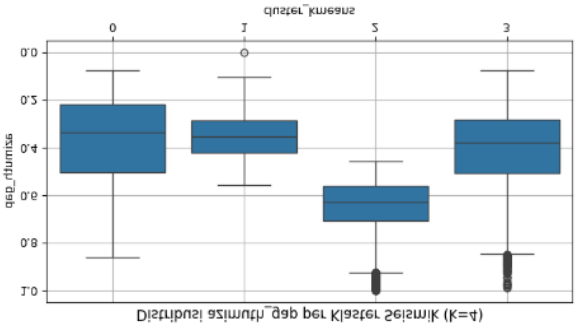


Gambar 10. Visualisasi Boxplot Untuk Atribut Depth Terhadap Masing-Masing Klaster



Gambar 11. Visualisasi Boxplot Untuk Atribut Phasecount Terhadap Masing-Masing Klaster

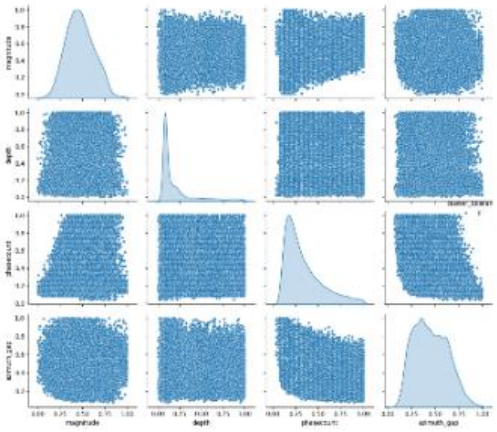
Analisis ini mendeskripsikan hasil pengelompokkan yang terbentuk pada tahap klasterisasi data seismik tanpa intervensi atau pembobotan tambahan, sehingga setiap cluster merepresentasikan segmentasi alami dari data yang digunakan. Hasil visualisasi dan distribusi angka memperlihatkan pola yang dapat diamati secara kuantitatif sesuai dengan keluaran algoritma.



Gambar 12. Visualisasi Boxplot Untuk Atribut Azimuth_Gap Terhadap Masing-Masing Klaster

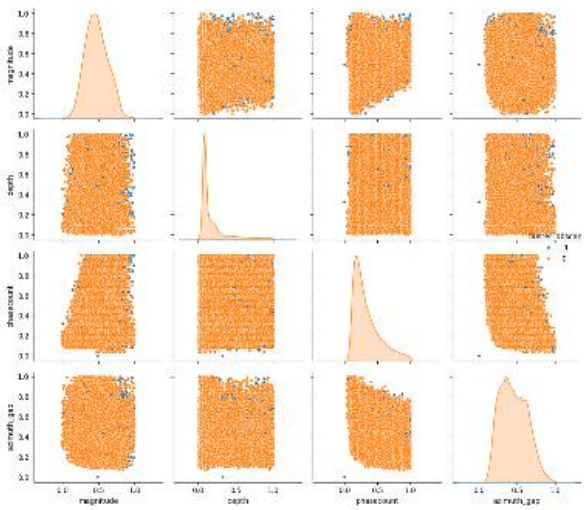
Analisis DBSCAN pada data seismik dilakukan dengan penentuan parameter optimal epsilon (ε) dan min_samples. Berdasarkan hasil estimasi menggunakan K - Distance Graph diperoleh rentang nilai *epsilon* optimal sekitar 0.0945 – 0.1300 untuk min_samples 5 – 12. Melalui proses *grid search* yang dilakukan, parameter terbaik diperoleh pada kombinasi $\varepsilon = 0.1072$ dan $min_samples = 8$ yang menghasilkan Silhouette Score sebesar 0.4043.

Implementasi DBSCAN dengan parameter tersebut hanya menghasilkan 1 *cluster* utama (Cluster 0) dengan 99.071 data, sementara 112 data lainnya teridentifikasi sebagai *outlier*. Hal ini menunjukkan bahwa sebagian besar data seismik memiliki pola yang sangat rapat sehingga hanya terbentuk satu klaster dominan, sedangkan data dengan pola berbeda atau jarang muncul diklasifikasikan sebagai *outlier*. Visualisasi pairplot pada Gambar 13, memperlihatkan persebaran atribut *magnitude*, *depth*, *phasecount*, dan *azimuth_gap* pada hasil klasterisasi. Jika *outlier* disertakan, terlihat pada Gambar 14 bahwa *outlier* tersebar pada nilai-nilai ekstrem atau jarang muncul, seperti pada kombinasi *magnitude* rendah namun *phasecount* tinggi.



Gambar 13. Pairplot Klasterisasi DBSCAN Data Seismik Tanpa Outlier

DOI: https://doi.org/10.38204/tematik.v12i1.2369
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)



Gambar 14. Pairplot Klasterisasi DBSCAN Data Seismik Dengan Outlier

Distribusi ini mengindikasikan bahwa DBSCAN cenderung sensitif terhadap kepadatan data. Mayoritas data seismik membentuk satu wilayah kepadatan tinggi yang dianggap satu klaster, sementara data dengan pola jarang langsung dikategorikan sebagai *outlier*.

3.3. Klasterisasi Gabungan

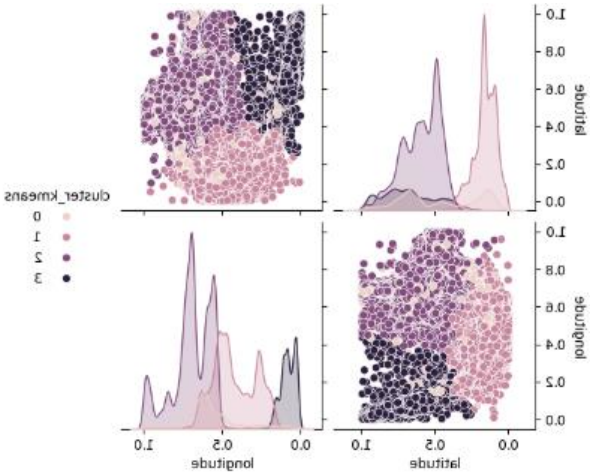
Pada tahap ini, analisis dilakukan dengan menggabungkan data spasial (*latitude*, *longitude*) dan data seismik (*magnitude*, *depth*, *phasecount*, *azimuth_gap*) dalam satu kesatuan dataset. Penggabungan ini dimaksudkan untuk mengeksplorasi pola klasterisasi yang terbentuk ketika informasi lokasi geografis dan karakteristik kegempaan dipertimbangkan secara bersamaan.

Pemilihan jumlah klaster optimal ditentukan menggunakan metode Elbow yang dapat dilihat pada Gambar 15, yang menunjukkan titik siku pada k = 4. Nilai ini kemudian digunakan sebagai jumlah klaster terbaik pada K-Means. Hasil Silhouette Score terbaik diperoleh pada k = 4 dengan skor 0,2904 yang menunjukkan pemisahan klaster yang cukup memadai. Jumlah data per klaster dapat dilihat pada Tabel 3 sebagai berikut.

Tabel 3. Distribusi Data Gabungan Setiap Klaster

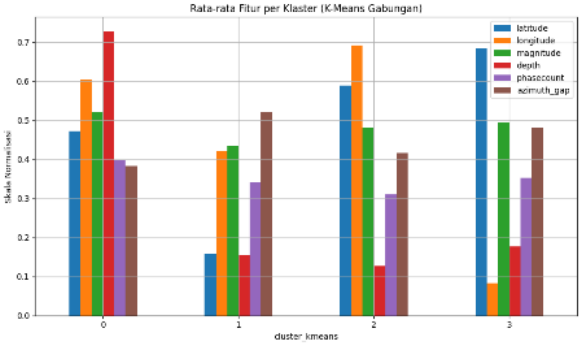
Cluster ke-	Jumlah Data
0	11.685
1	31.577
2	43.931
3	11.990

Nilai Davies-Bouldin Index (DBI) sebesar 1,2551 diperoleh sebagai ukuran validitas internal klasterisasi gabungan spasial dan seismik, di mana angka ini mengindikasikan pemisahan antar klaster yang masih moderat, belum optimal namun tetap menunjukkan pola terklaster.



Gambar 15. Visualisasi Pairplot Spasial Pada Data Gabungan

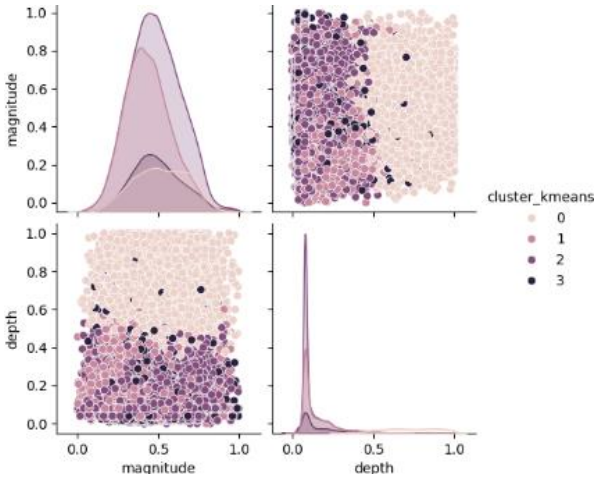
Analisis hasil visualisasi pairplot pada Gambar 14 memperlihatkan bahwa distribusi spasial (*latitude*, *longitude*) membentuk pemisahan klaster yang cukup jelas, terutama pada *cluster* 3 yang mendominasi area dengan rentang *latitude* tinggi. Sementara itu, *cluster* 0 dan 2 cenderung menyebar pada area *latitude* menengah hingga rendah mengindikasikan adanya perbedaan pola geografis antar kelompok. Berdasarkan distribusi atribut yang ditunjukkan pada Gambar 16, terlihat bahwa Cluster 0 memiliki rata-rata nilai *depth* dan *azimuth_gap* tertinggi yang menunjukkan bahwa area tersebut tidak hanya berada dalam lebih dalam secara representasi seismik, tetapi juga menunjukkan kesenjangan sudut yang signifikan dalam pola gelombang seismiknya. Cluster 1 menunjukkan nilai rata-rata *magnitude* dan *longitude* yang lebih tinggi yang mengindikasikan wilayah dengan gempa yang lebih kuat dan posisi geografis lebih ke arah timur. Cluster 2 memperlihatkan nilai *phasecount* yang relatif rendah yang berarti perlu ditinjau lebih lanjut apakah memang betul rendah atau terdapat kendala. Terakhir untuk Cluster 3 terlihat mendominasi pada area *latitude* tinggi dengan nilai *magnitude* yang sedang serta memiliki kedalaman lebih dangkal dibandingkan *cluster* lain yang mengarah pada pola gempa kerak dangkal di area tertentu.



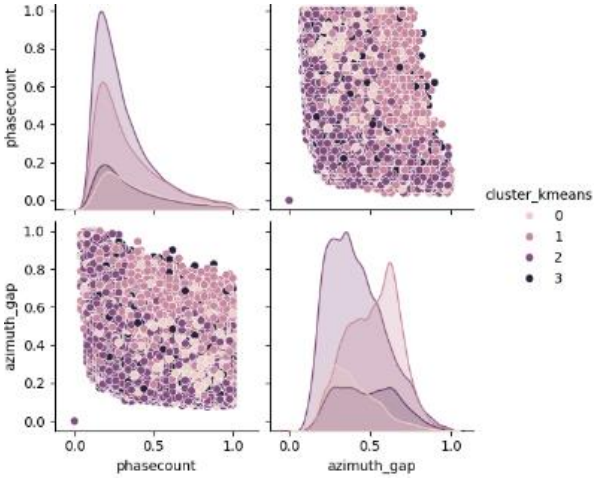
Gambar 16. Visualisasi Distribusi Fitur Pada Setiap Cluster

DOI: https://doi.org/10.38204/tematik.v12i1.2369
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Analisis lebih mendalam pada Pairplot Magnitude dan Depth yang dapat diihat pada Gambar 17, terlihat bahwa sebagian besar gempa di *cluster* 3 terjadi pada kedalaman kurang dari separuh skala maksimum, sementara *magnitude* relatif stabil pada kisaran menengah. Sebaliknya untuk *cluster* 0 memiliki persebaran kedalaman yang jauh lebih bervariasi termasuk beberapa kejadian gempa dalam yang signifikan.



Gambar 17. Visualisasi Pairplot Magnitude dan Depth Data Gabungan

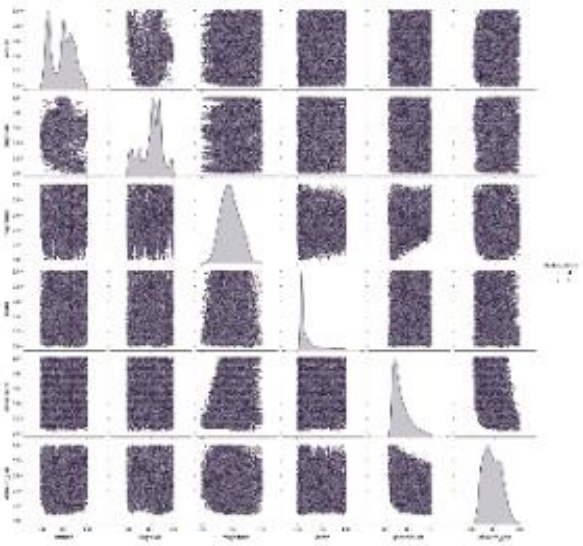


Gambar 18. Visualisasi Pairplot Phasecount dan Azimuth Gap Data Gabungan

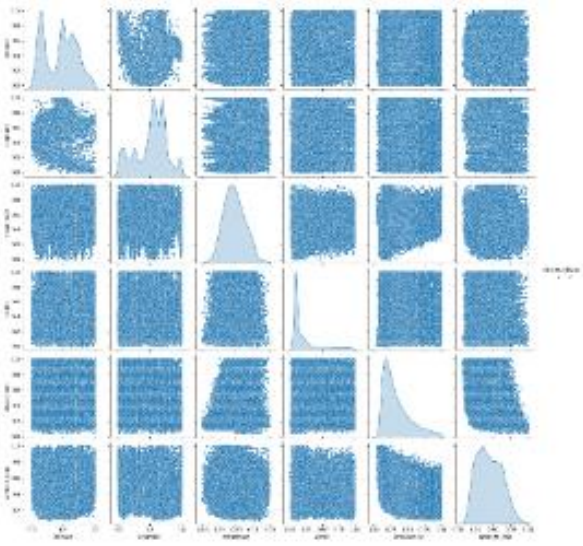
Sementara itu pada Pairplot Phasecount dan Azimuth Gap pada Gambar 18 menunjukkan bahwa *cluster* 0 memiliki variasi *azimuth gap* yang sangat lebar yang mengindikasikan potensi ketidakseragaman distribusi fase gempa yang terdeteksi, sedangkan *cluster* 2 dan 3 lebih terkonsentrasi pada rentang *phasecount* dan *azimuth gap* yang lebih sempet yang menandakan konsistensi pengamatan di wilayah tersebut.

Analisis DBSCAN pada data gabungan spasial dan seismik dilakukan dengan menentukan kombinasi parameter *epsilon* (*eps*) dan *min_samples* terbaik melalui Grid Seach. Berdasarkan hasil pencarian, diperoleh parameter optimal *eps* = 0,2030 dan *min_samples* = 10 dengan nilai Silhouette Score sebesar 0,2922. Meskipun nilai ini tidak tergolong tinggi namun tetap memberikan informasi terkait pemisahan struktur data pada gabungan atribut spasial dan seismik.

Hasil klasterisasi menunjukkan bahwa DBSCAN hanya membentuk satu klaster utama (Cluster 0) dengan total 99.054 data, sedangkan 129 data lainnya terdeteksi sebagai *outlier*. Situasi ini diperkuat oleh visualisasi distribusi pada Gambar Pairplot yang dapat dilihat pada Gambar 19 di mana seluruh titik data cenderung terkonsentrasi dalam satu kelompok besar.



Gambar 19. Visualisasi Pairplot DBSCAN Tanpa Outlier Data Gabungan



Gambar 20. Visualisasi Pairplot DBSCAN Dengan Outlier Data Gabungan

Ketika divisualisasikan tanpa memisahkan *outlier* yang dapat dilihat pada Gambar 20 persebaran data tetap tampak homogen, dengan *outlier* hanya muncul sebagai titik-titik minoritas di tepian distribusi. Analisis visual

DOI: https://doi.org/10.38204/tematik.v12i1.2369
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

menunjukkan bahwa tidak ada dominasi pola khusus yang mampu dipisahkan secara tegas oleh DBSCAN baik dari sisi atribut spasial (*latitude*, *longitude*) maupun atribut seismik (*magnitude*, *depth*, *phasecount*, *azimuth_gap*).

Jika dibandingkan dengan hasil K-Means pada data gabungan, DBSCAN pada konfigurasi ini menunjukkan keterbatasan dalam menangkap variasi klaster yang halus. Hal ini dapat terjadi karena sensitivitas DBSCAN terhadap parameter *epsilon*, di mana penyesuaian kecil dapat secara drastis mengubah hasil pemisahan. Selain itu, persebaran data yang cenderung rapat dan homogen pada ruang fitur gabungan juga menjadi tantangan, mengingat DBSCAN lebih efektif bekerja pada dataset dengan kepadatan berbeda yang kontras antar wilayah.

4. Kesimpulan

Penelitian ini telah berhasil melakukan pengelompokkan daerah rawan gempa di Indonesia dengan memanfaatkan metode klasterisasi K-Means dan DBSCAN pada tiga skenario dataset yaitu spasial (*latitude*, *longitude*), seismik (*magnitude*, *depth*, *phasecount*, *azimuth_gap*), serta gabungan spasial dan seismik. Hasil eksperimen menunjukkan bahwa pada data spasial, K-Means dengan 3 klaster optimal mampu memetakan wilayah Indonesia berdasarkan distribusi geografis dengan baik, sedangkan DBSCAN menghasilkan pemisahan klaster yang lebih terbatas. Pada data seismik, K-Means dengan 4 klaster berhasil membedakan pola berdasarkan kedalaman, kekuatan gempa, jumlah fase, dan azimuth gap, meskipun nilai evaluasi Davies-Bouldin Index (1,2017) mengindikasikan pemisahan antar klaster yang sedang. Pada data gabungan spasial dan seismik, K-Means kembali menunjukkan kinerja lebih baik dibandingkan DBSCAN, dengan nilai Davies-Bouldin Index sebesar 1,2551 dan distribusi antar klaster yang jelas terlihat pada visualisasi. Sebaliknya DBSCAN pada parameter optimal cenderung menggabungkan sebagian besar data ke satu klaster utama dengan sensitivitas tinggi terhadap *outlier*. Secara umum, penelitian ini menunjukkan bahwa metode K-Means lebih efektif dalam mendeteksi variasi klaster pada konteks pemetaan wilayah rawan gempa, sedangkan DBSCAN lebih cocok digunakan pada kasus data dengan pola kepadatan yang kompleks namun memerlukan penyesuaian parameter yang cermat. Temuan ini memberikan landasan penting untuk pengembangan model pemetaan risiko gempa yang lebih adaptif dan berbasis data spasial maupun seismic.

Daftar Rujukan

[1] F. F. Susanta, C. Pratama, T. Aditya, A. F. Khomaini, and H. W. K. Abdillah, "Geovisual Analytics of Spatio-Temporal Earthquake Data in Indonesia," *JGISE J. Geospatial Inf. Sci. Eng.*, vol. 2, no. 2, Dec. 2019, doi: 10.22146/jgise.51131.

[2] "This Final Report for the MDRID013 emergency appeal operation reflects the overall context of the disasters, the operational strategy, the overall achievements (in numbers), National Society Development and coordination components, the PMI branch construction projects, and findings from the CVA lessons learned and findings and recommendations from the Final Evaluation. Indonesia: 2018 Earthquakes and Tsunamis." [Online]. Available: http://www.ifrc.org/docs/appeals/Active/MDRID013.pdf

[3] "The National Medium-Term Development Plan For 2020-2024 Narration Republic Of Indonesia."

[4] M. Bariklana and A. Fauzan, "Implementation Of The DBSCAN Method For Cluster Mapping Of Earthquake Spread Location," *Barekeng J. Ilmu Mat. dan Terap.*, vol. 17, no. 2, pp. 0867–0878, Jun. 2023, doi: 10.30598/barekengvol17iss2pp0867-0878.

[5] I. M. Rajagukguk, B. K. Lukito, A. A. Santoso Gunawan, and M. E. Syahputra, "Optimizing Earthquake Forecast Through Clustering: K-Means and DBSCAN Approaches," in *Proceedings - 2024 International Conference on Information Technology and Computing, ICITCOM 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 224–229. doi: 10.1109/ICITCOM62788.2024.10762556.

[6] N. F. Saragih, Y. Y. Pratiwi, I. K. Jaya, I. M. Sarkis, H. H. Arrizal, and M. Sinambela, "Clustering Earthquakes in West Java Using Machine Learning Algorithm," in *2023 IEEE International Conference of Computer Science and Information Technology: The Role of Artificial Intelligence Technology in Human and Computer Interactions in the Industrial Era 5.0, ICOSNIKOM 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICoSNIKOM60230.2023.10364541.

[7] A. Zamani, M. R. Sorbi, and A. A. Safavi, "Application of neural network and ANFIS model for earthquake occurrence in Iran," *Earth Sci. Informatics*, vol. 6, no. 2, pp. 71–85, Jun. 2013, doi: 10.1007/s12145-013-0112-8.

[8] G. Kristianto Anver and Y. J. Prasetyo, "Analysis of Earthquake Intensity on Java Island Using K-Means Clustering and GeoMap," *Indones. J. Soc. Technol.*, vol. 6, no. 1, p. 469, 2025, [Online]. Available: http://jist.publikasiindonesia.id/

[9] D. H. Natawidjaja, "Updating active fault maps and sliprates along the Sumatran Fault Zone, Indonesia," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics Publishing, Feb. 2018. doi: 10.1088/1755-1315/118/1/012001.

[10] J. S. Mufti and A. Dhini, "Comparative Analysis of Centroid-Based and Density-Based Clustering for Indonesian Earthquake Data," in *International Conference on Computer, Control, Informatics and its Applications, IC3INA*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 464–467. doi: 10.1109/IC3INA64086.2024.10732579.

[11] M. Sinambela, E. Darnila, I. K. Jaya, I. M. Sarkis, and A. Rikki, "Cluster Analysis and Seismicity of The Samosir Using Machine Learning Approach," in *2023 8th International Conference on Informatics and Computing, ICIC 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICIC60109.2023.10381924.

[12] R. Alom, A. Mazumdar, R. Kr Prasad, G. Basumatary, and B. Baruah, "Analysis of Seismic Data Using Partition-Based Clustering Techniques," in *2022 IEEE Global Conference on Computing, Power and Communication Technologies, GlobConPT 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/GlobConPT57482.2022.9938362.

[13] H. H. Arrizal, M. Sinambela, Widodo, S. P. Adi, and H. T. Frianto, "Spatial Cluster Analysis of Land Seismicity in the Northem Part of Sumatra Using the K-Medoids Algorithm," in *Proceeding - International Conference on Information Technology and Computing 2023, ICITCOM 2023*, Institute of Electrical and Electronics Engineers Inc.,

DOI: https://doi.org/10.38204/tematik.v12i1.2369
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

2023, pp. 181–185. doi: 10.1109/ICITCOM60176.2023.10442535.

[14] S. W. Ginting, R. S. Hartati, M. Sudarma, and I. B. A. Swamardika, "Clustering of Earthquake and Volcanic Eruption Trauma Survivor Groups using K-Means Algorithm," in *ICSGTEIS 2021 - 2021 International Conference on Smart-Green Technology in Electrical and Information Systems: Advancing Smart and Green Technologies Toward Society 5.0, Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 69–73. doi: 10.1109/ICSGTEIS53426.2021.9650357.

[15] D. K. Wijaya, R. Widjaya, A. Kurniawan, and Diana, "Using Time Series Analysis to Predict Earthquake," in *2024 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics, ICoABCD 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 25–30. doi: 10.1109/ICoABCD63526.2024.10704296.

[16] H. Pratiwi, S. S. Handajani, I. Susanto, S. Sangadji, R. Meilawati, and I. S. Khairunnisa, "Hierarchical clustering algorithm for analyzing risk of earthquake on sumatra island," in *International Conference on Electrical, Computer, Communications and Mechatronics Engineering, ICECCME 2021*, Institute of Electrical and Electronics Engineers Inc., Oct. 2021. doi: 10.1109/ICECCME52200.2021.9590890.

[17] M. Sinambela, H. H. Arrizal, E. Darnila, Munawar, M. R. Nugraha, and A. Widodo, "Clustering Analysis of Earthquake Based on K-Means, DBSCAN, and Fuzzy C-Means in North Sumatra," in *Proceedings - 2024 International Conference on Information Technology and Computing, ICITCOM 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 242–247. doi: 10.1109/ICITCOM62788.2024.10762334.

[18] S. Gopal, K. Patro, and K. Kumar Sahu, "Normalization: A Preprocessing Stage." [Online]. Available: www.kiplinger.com,

[19] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci. (Ny).*, vol. 622, pp. 178–210, Apr. 2023, doi: 10.1016/j.ins.2022.11.139.

[20] A. Jaeger and D. Banks, "Cluster analysis: A modern statistical review," May 01, 2023, *John Wiley and Sons Inc*. doi: 10.1002/wics.1597.

[21] M. Suyal and S. Sharma, "A Review on Analysis of K-Means Clustering Machine Learning Algorithm based on Unsupervised Learning," *J. Artif. Intell. Syst.*, vol. 6, no. 1, pp. 85–95, 2024, doi: 10.33969/ais.2024060106.

[22] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," 1996. [Online]. Available: www.aaai.org

[23] J. ¨ O. Sander and M. Ester, "Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and Its Applications," 1998.

[24] A. E. Ezugwu *et al.*, "A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects," Apr. 01, 2022, *Elsevier Ltd*. doi: 10.1016/j.engappai.2022.104743.

DOI: https://doi.org/10.38204/tematik.v12i1.2369
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)