

Pengelompokan Daerah Rawan Gempa Bumi di Pulau Jawa Menggunakan Metode *Density Based Spatial Clustering of Application with Noise* (DBSCAN)

Naufalia Alfiryal*, Ilham Faishal Mahdy

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

*alfiryaln@gmail.com, ilham.faishal@unisba.ac.id

Abstract. In cluster analysis, there are often several problems, such as arbitrary cluster shapes and different cluster sizes and densities. One method that can handle these problems is Density Based Spatial Clustering of Application with Noise (DBSCAN). Java Island is one of the areas in Indonesia that often experiences earthquakes, due to the collision zone of the Indo-Australian plate colliding with the Eurasian plate and the presence of 34 active volcanoes. The earthquakes that occur also vary in terms of magnitude and depth. The greater the magnitude and the shallower the depth, the greater the damage. Java is one of the most populous islands in Indonesia, so it is important to know the earthquake-prone areas in Java to minimize the risk through cluster analysis. The data used in this study is earthquake event data in the January 1, 2021 - January 1, 2024 date range obtained from the United States Geological Survey (USGS). From the cluster analysis or grouping carried out, 2 clusters and 1 noise were obtained with the optimal parameter value $\epsilon = 0.2857$ and $\text{MinPts} = 4$. Cluster 1 consists of 167 areas prone to medium earthquakes and has an average magnitude of 4.42 mb. Cluster 2 consists of 8 areas prone to deep earthquakes and have an average magnitude of 4.46 mb. Meanwhile, the noise regions are areas prone to deep and high magnitude earthquakes. The Silhouette coefficient of 0.6788 means that the cluster structure formed is already in the good structure category.

Keywords: *Cluster Analysis, DBSCAN, Earthquake.*

Abstrak. Dalam analisis kluster seringkali terjadi beberapa permasalahan, seperti bentuk kluster yang tidak beraturan (*arbitrary*) serta ukuran dan kepadatan kluster yang berbeda. Salah satu metode yang mampu menangani permasalahan tersebut adalah *Density Based Spatial Clustering of Application with Noise* (DBSCAN). Pulau Jawa merupakan salah satu daerah di Indonesia yang sering mengalami gempa bumi. Hal ini dikarenakan adanya zona tumbukan lempeng Indo-Australia yang bertabrakan dengan lempeng Eurasia serta adanya 34 gunung berapi aktif. Gempa bumi yang terjadi juga bervariasi dari segi magnitudo maupun kedalamannya (*depth*). Semakin besar magnitudo dan semakin dangkal kedalamannya, maka akan semakin besar pula kerusakan yang terjadi. Pulau Jawa merupakan salah satu pulau terpadat di Indonesia sehingga penting untuk mengetahui daerah rawan gempa bumi di Pulau Jawa untuk meminimalisir risiko melalui analisis kluster. Data yang digunakan pada penelitian ini merupakan data kejadian gempa bumi pada rentang tanggal 1 Januari 2021-1 Januari 2024 yang diperoleh dari *United States Geological Survey* (USGS). Dari analisis kluster atau pengelompokan yang dilakukan diperoleh 2 kluster dan 1 noise dengan nilai parameter optimal $\epsilon = 0.2857$ dan $\text{MinPts} = 4$. Kluster 1 terdiri dari 167 daerah yang rawan akan gempa bumi menengah dan memiliki rata-rata magnitudo 4.42 mb. Kluster 2 terdiri dari 8 daerah yang rawan akan gempa bumi dalam dan memiliki rata-rata magnitudo 4.46 mb. Sedangkan, daerah yang menjadi noise merupakan daerah yang rawan gempa bumi dalam dan magnitudo tinggi. *Silhouette coefficient* sebesar 0.6788 mengartikan bahwa struktur kluster yang terbentuk sudah dalam kategori struktur yang baik.

Kata Kunci: *Analisis Klaster, DBSCAN, Gempa Bumi.*

A. Pendahuluan

Data mining merupakan merupakan sebuah proses menemukan pola yang menarik dari sejumlah besar data. Salah satu teknik yang termasuk dalam data mining adalah *machine learning*. *Unsupervised learning* (pembelajaran tanpa pengawasan) merupakan salah satu jenis *machine learning* yang pada dasarnya merupakan sinonim dari pengelompokan atau analisis kluster dimana dengan metode ini dapat ditemukan pola-pola atau kelas dalam data yang sebelumnya belum diketahui [7].

Analisis kluster atau pengelompokan merupakan suatu proses membagi sekelompok objek pada data menjadi beberapa subset. Masing-masing subset merupakan sebuah kluster, maka objek yang memiliki karakteristik serupa satu sama lain akan berada dalam satu kluster yang sama. Sebaliknya, objek yang tidak memiliki karakteristik serupa akan berada dalam kluster yang berbeda (Meidianingsih et al., n.d.). Dalam melakukan analisis kluster ini seringkali muncul permasalahan, seperti bentuk kluster yang tidak beraturan (*arbitrary*) serta ukuran dan kepadatan kluster yang berbeda. Permasalahan ini dapat diselesaikan dengan menggunakan analisis kluster berbasis kepadatan, salah satunya seperti *Density Based Spatial Clustering of Application with Noise* (DBSCAN) [7]. DBSCAN dapat diterapkan dalam berbagai hal, seperti untuk mengelompokkan daerah rawan kejadian gempa bumi.

Gempa bumi merupakan bencana alam yang terjadi dalam waktu yang relatif singkat dan terjadi secara tiba-tiba sehingga dapat menghancurkan apapun baik harta, benda, manusia, dan lainnya yang ada di muka bumi. Negara Indonesia merupakan salah satu negara yang rawan terjadi gempa bumi (Karimuse et al., n.d.). Hal tersebut dikarenakan adanya jalur pertemuan antara 3 buah lempeng tektonik, yakni lempeng Indo-Australia, lempeng Pasifik, dan lempeng Eurasia [8]. Kejadian gempa bumi yang biasa terjadi juga bervariasi berdasarkan parameter yang mengikutinya, seperti dari segi magnitudo maupun kedalamannya (*depth*). Menurut kedalamannya, gempa bumi dapat diklasifikasikan menjadi tiga, yakni gempa dangkal, gempa intermediet, dan gempa dalam. Dengan begitu, semakin besar magnitudo maka akan semakin besar pula kerusakan yang terjadi, sedangkan semakin rendah kedalaman kejadian gempa bumi, maka akan semakin besar kerusakan yang akan ditimbulkannya [8].

Salah satu daerah di Indonesia yang rawan akan terjadinya gempa bumi adalah Pulau Jawa. Hal ini berhubungan dengan adanya zona tumbukan lempeng Indo-Australia di bagian selatan Pulau Jawa yang bertabrakan dengan lempeng Eurasia di bagian utaranya [12] dan 34 gunung berapi aktif [3]. Kemudian, dilansir pada laman dataindonesia.id, pada tahun 2021 *World Economic Forum* (WEF) mencatat bahwa kepadatan penduduk di Pulau Jawa adalah sebesar 1.015,9 jiwa per kilometer persegi (km²). Maka dari itu, penting untuk mengetahui daerah mana saja di Pulau Jawa yang rawan terjadi gempa bumi dengan memperhatikan magnitudo dan kedalaman kejadian gempa bumi untuk meminimalisir risiko yang dapat terjadi dengan pengelompokan (Samudera Kutaraja Banda Aceh Rizwan et al., 2023). DBSCAN dapat mengelompokkan kejadian gempa bumi yang tidak memiliki pola tertentu atau bervariasi dari segi kedalaman dan magnitudonya sehingga nantinya dapat terbentuk kluster-kluster dengan karakteristik kedalaman dan magnitudo gempa bumi yang dominan untuk setiap anggota dalam satu kluster. Melalui metode DBSCAN juga dapat diketahui kejadian gempa bumi yang memiliki karakteristik berbeda dengan kluster-kluster utama yang telah terbentuk melalui keunggulannya dalam mendeteksi *noise* atau *outliers*.

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, adapun tujuan dari penelitian ini sebagai berikut:

1. Mengetahui hasil pengelompokan daerah rawan gempa bumi berdasarkan magnitudo dan kedalaman kejadian gempa bumi di Pulau Jawa menggunakan metode *Density Based Spatial Clustering of Application with Noise* (DBSCAN)?
2. Mengetahui hasil kinerja metode *Density Based Spatial Clustering of Application with Noise* (DBSCAN) untuk pengelompokan daerah rawan gempa bumi berdasarkan magnitudo dan kedalaman kejadian gempa bumi di Pulau Jawa?

B. Metodologi Penelitian

Pada penelitian ini, data yang digunakan merupakan data sekunder kejadian gempa bumi di

Pulau Jawa dalam rentang waktu 1 Januari 2021 sampai dengan 1 Januari 2024 yang diperoleh melalui website resmi *United States Geological Survey* (USGS) dengan 176 kejadian serta 5 variabel *time*, *place*, *depth*, dan magnitudo. Variabel *depth* (X_1) dan magnitudo (X_2) akan digunakan dalam proses pengelompokan menggunakan metode DBSCAN sedangkan variabel *time* dan *place* akan digunakan sebagai keterangan waktu dan tempat dari setiap kejadian gempa bumi. Berikut merupakan ringkasan data yang digunakan.

Tabel 1. Ringkasan Data Penelitian

<i>Time</i>	<i>Place</i>	<i>Depth</i>	<i>Magnitudo</i>
2024-01-01T13:46:45.922Z	25 km Timur laut dari Lembang, Indonesia	10	4,5
2024-01-01T00:59:19.656Z	4 km Selatan barat daya dari Cianjur, Indonesia	178,051	4,1
⋮	⋮	⋮	⋮
2021-01-15T14:17:02.903Z	95 km Selatan dari Sidareja, Indonesia	101,89	4,1
2021-01-14T01:51:17.879Z	175 km Barat daya dari Pelabuhanratu, Indonesia	10	4,3

Density Based Spatial Clustering of Application with Noise (DBSCAN)

DBSCAN merupakan sebuah algoritma pengelompokan yang berfokus pada kepadatan (*density*) data. Aspek spasial pada algoritma ini terkait dengan posisi dan kedekatan titik dalam data ketika membuat suatu klaster, dalam hal ini tergantung pada kepadatan. Kepadatan yang dimaksud dalam DBSCAN merujuk pada pembentukan klasternya, daerah yang padat atau memiliki densitas tinggi dianggap sebagai klaster. Sebaliknya, daerah yang jarang atau memiliki densitas rendah dianggap sebagai *noise* [7]. Suatu daerah dapat dikatakan padat atau tidak dilihat dari tingkat kerapatan atau kedekatan jarak antar objek pada sebuah data [6].

Untuk mengestimasi kepadatan tersebut, algoritma ini perlu mengetahui terlebih dahulu nilai untuk parameter epsilon (ϵ) dan *MinPts* (*Minimal Points*). Parameter ϵ merupakan sebuah nilai untuk radius antar titik yang menjadi dasar pembentukan *neighborhood* dari sebuah titik pada data [4]. Sedangkan *MinPts* merupakan minimal titik dalam sebuah klaster [7]. Dalam DBSCAN dikenal beberapa istilah penting, yakni [4]:

1. *Core Point*, merupakan sebuah titik yang berada dalam suatu klaster dan memiliki setidaknya *MinPts* dalam radius ϵ . *Core Point* dapat didefinisikan dalam persamaan berikut:

$$\text{core point} = |N_{\epsilon}(p)| \geq \text{MinPts} \quad (1)$$

dengan, $N_{\epsilon}(p)$: titik yang berada dalam radius ϵ titik p , *MinPts* : minimal titik dalam sebuah klaster, dan p : pemisalan untuk sebuah titik dalam dataset.

2. *Border Point*, merupakan sebuah titik yang berada pada perbatasan suatu klaster.

$$\text{Border point} = \{|N_{\epsilon}(p)| < \text{MinPts} \mid q \in N_{\epsilon}(p)\} \quad (2)$$

dengan, $N_{\epsilon}(p)$: titik yang berada dalam radius ϵ titik p , *MinPts* : minimal titik dalam sebuah klaster, p : pemisalan untuk sebuah titik dalam dataset, dan q : *core point*.

3. *Noise*, merupakan sebuah titik yang termasuk dalam sebuah dataset, tetapi tidak termasuk dalam anggota klaster manapun.

$$\text{noise} = \{p \in D \mid \forall_i: p \notin C_i\} \quad (3)$$

dengan, D : dataset, C_i : klaster ke- i , dan p : pemisalan untuk sebuah titik dalam dataset.

4. *Directly Density-Reachable*, suatu titik p dapat dijangkau secara langsung dari titik q jika memenuhi syarat berikut:

- 1) $p \in N_{\epsilon}(q)$

$$2) |N_{eps}(q)| \geq MinPts$$

dengan, $N_{eps}(q)$: titik yang berada dalam radius ϵ titik q , $MinPts$: minimal titik dalam sebuah kluster, p : *border point*, dan q : *core point*.

5. *Density-Reachable*, suatu titik dikatakan terjangkau secara densitas apabila titik tersebut tidak terjangkau secara langsung dari titik inti (*core point*). Titik p dapat dikatakan dijangkau secara densitas dari titik q apabila terdapat rantai titik p_1, \dots, p_n sedemikian sehingga $p_1 = q$, $p_n = p$, dan p_{i+1} secara langsung dapat dijangkau dari p_i dengan memperhatikan ϵ dan $MinPts$. Hal ini dapat didefinisikan pada persamaan berikut:

$$\begin{aligned} \text{density - reachable} &= \exists \{p_1, p_2, \dots, p_n\} \mid p_1 = q, p_n = p, \\ \forall 1 \leq i < n, p_{i+1} &\text{ adalah directly density - reachable dari } p_i \end{aligned} \quad (4)$$

dengan, p : titik dalam dataset yang dapat berupa *border point* dan *core point*, q : *core point*, n : panjang rantai titik, dan i : indeks titik dalam rantai titik.

6. *Density-Connected*, suatu titik p dikatakan terhubung secara densitas ke sebuah titik q dalam radius ϵ apabila terdapat titik o (*core point*) sehingga kedua titik tersebut juga dapat dijangkau secara densitas dari o .

$$\begin{aligned} \text{density - connected} &= \exists o \mid p \text{ adalah density -} \\ &\text{reachable dari } o \text{ dan } q \text{ adalah density - reachable dari } o \end{aligned} \quad (5)$$

dengan, p : titik dalam dataset yang ingin diperiksa apakah terhubung secara kepadatan dengan titik q dan dapat berupa *border point*, q : titik dalam dataset yang ingin diperiksa apakah terhubung secara kepadatan dengan titik p dan dapat berupa *border point*, dan o : titik perantara yang merupakan *core point* yang terjangkau secara kepadatan baik dari p maupun q .

Adapun tahapan analisis kluster DBSCAN sebagai berikut:

1. Masukkan D (dataset), ϵ , dan $MinPts$.
2. Tandai semua titik pada D sebagai belum dikunjungi.
3. Ulangi langkah berikut hingga semua titik pada D sudah dikunjungi semua.
 - a. Tentukan titik p secara acak. Lalu, tandai p sebagai sudah dikunjungi.
 - b. Hitung jarak dari titik p ke titik lainnya dengan menggunakan perhitungan jarak *Euclidean*.
 - c. Periksa apakah terdapat sebanyak $MinPts$ titik di dalam radius ϵ dari p . Apabila p tidak termasuk dalam suatu kluster, maka kluster baru dibuat dan titik p serta titik-titik dalam radius ϵ dari p ditambahkan dalam kluster tersebut. Namun, apabila p termasuk dalam suatu kluster, maka titik-titik dalam radius ϵ dari titik p ditambahkan dalam kluster tersebut.
Apabila banyak titik dalam radius ϵ dari p kurang dari $MinPts$ dan p tidak termasuk dalam kluster manapun, maka p ditandai sebagai *noise*.
 - d. Kemudian, ulangi langkah 3a sampai 3c pada setiap titik yang tersisa hingga tidak ada titik yang ditandai sebagai belum dikunjungi.

Berdasarkan algoritma DBSCAN tersebut, DBSCAN mampu mendeteksi *noise* melalui algoritma pengelompokannya. Selain itu, DBSCAN juga masih memiliki keuntungan lainnya, yakni [10]:

1. DBSCAN tidak mengharuskan untuk mengetahui jumlah kluster di awal.
2. DBSCAN mampu untuk menemukan kluster dengan bentuk sembarang.
3. DBSCAN hanya membutuhkan dua parameter, yakni ϵ dan $MinPts$ yang sebagian besar tidak sensitif terhadap urutan titik dalam dataset.

K-Nearest Neighbour (KNN)

KNN merupakan sebuah algoritma yang didasarkan pada pembelajaran berbasis analogi, yakni dengan membandingkan data uji yang diberikan dengan data latih yang memiliki kemiripan. Di mana suatu data latih dikatakan mirip dengan data ujinya ketika memiliki jarak yang dekat [7]. Algoritma ini memiliki satu parameter, yakni K (jumlah tetangga terdekat). Algoritma KNN ini akan digunakan untuk menentukan nilai ϵ yang optimal dengan cara melihat grafik *k-distance*. Setelah memperoleh nilai ϵ yang optimal, maka dilanjutkan dengan menentukan nilai $MinPts$.

Seperti yang telah disebutkan sebelumnya, nilai $MinPts$ diperoleh dari nilai $K+1$ [5]. Adapun tahapannya sebagai berikut:

1. Tentukan parameter K . Dalam hal ini, nilai K diperoleh dari $((2 \times \text{dimensi}) - 1)$.
2. Hitunglah jarak antar objek atau data menggunakan rumus jarak *Euclidean* pada Persamaan (2.2).
3. Ambil K tetangga terdekat atau objek yang memiliki jarak paling kecil.
4. Kemudian, ambil nilai jarak dari tetangga terdekat ke- K untuk masing-masing data.
5. Susun nilai jarak dari tetangga terdekat ke- K yang telah diperoleh sebelumnya dari yang terbesar menuju ke terkecil ke dalam grafik *k-distance* sebagai sumbu y dan diikuti indeks objek sebagai sumbu x .
6. Amati grafik *k-distance* dan lihat *threshold point* yang merupakan radius pertama pada lembah pertama.
7. Hasil dari *threshold point* itulah yang dipilih sebagai nilai ϵ .

Kemudian, untuk nilai $MinPts$ diperoleh dari nilai K yang telah ditentukan sebelumnya dan ditambahkan dengan nilai 1.

Silhouette Coefficient

Silhouette coefficient merupakan hasil penggabungan dari dua metode, yakni metode *cohesion* dan metode *separation*. Metode *cohesion* digunakan untuk mengukur tingkat kedekatan antara objek dalam satu kluster, sedangkan metode *separation* digunakan untuk mengukur seberapa jauh terpisahnya satu kluster dengan kluster lainnya [1]. Dalam proses perhitungan *silhouette coefficient*, hal pertama yang perlu dilakukan adalah mencari jarak rata-rata data ke- i dengan seluruh data di dalam kluster yang sama, dalam hal ini mari asumsikan data ke- i berada dalam kluster A. Rumus dari $a_{(i)}$ (*cohesion*) dapat ditulis sebagai berikut [11]:

$$a_{(i)} = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j) \quad (6)$$

dengan, $a_{(i)}$ = rata-rata jarak data ke- i terhadap semua data dalam satu kluster, A = banyaknya data di kluster A, dan $d(i, j)$ = jarak antar objek i dan objek j .

Kemudian, untuk $b_{(i)}$ (*separation*) merupakan nilai minimum dari jarak rata-rata data ke- i dengan seluruh data dalam kluster yang berbeda, misalnya C diasumsikan sebagai kluster yang berbeda selain A. Untuk perhitungannya dapat digunakan persamaan berikut:

$$b_{(i)} = \min \frac{1}{|C|} \sum_{j \in C} d(i, j) \quad (7)$$

dengan, $b_{(i)}$ = rata-rata jarak antara titik i dengan keseluruhan titik pada kluster lainnya,

C = jumlah data pada kluster C dan $d(i, j)$ = jarak antar objek i dan objek j pada kluster C. Setelah nilai $a_{(i)}$ dan $b_{(i)}$ diketahui, maka selanjutnya menghitung nilai *silhouette score* untuk masing-masing sampel dengan menggunakan persamaan berikut:

$$s_{(i)} = \frac{(b_{(i)} - a_{(i)})}{\max(a_{(i)}, b_{(i)})} \quad (8)$$

Nilai *silhouette score* ini berada pada rentang -1 sampai 1 yang berarti struktur kluster yang baik akan memiliki nilai *silhouette score* mendekati 1. Apabila nilai $s_{(i)} \approx 1$, maka data ke- i diklasifikasikan dengan baik (di dalam kluster A); nilai $s_{(i)} \approx 0$, maka data ke- i berada di tengah-tengah antara kedua kluster (A dan C); dan nilai $s_{(i)} \approx -1$, maka data ke- i diklasifikasikan lemah atau lebih dekat ke kluster C daripada A [11].

Kemudian, untuk memperoleh nilai *silhouette coefficient*, maka diperlukan rata-rata dari keseluruhan $s(i)$ dengan persamaan sebagai berikut [2]:

$$S = \frac{\sum_{i=1}^N \frac{(b_{(i)} - a_{(i)})}{\max(a_{(i)}, b_{(i)})}}{N} = \frac{\sum_{i=1}^N s_{(i)}}{N} \quad (9)$$

dengan S merupakan *silhouette coefficient* dan N merupakan jumlah keseluruhan data. Adapun kriteria pengukuran *silhouette coefficient* sebagai berikut [9].

Tabel 1. Kriteria Pengukuran *Silhouette Coefficient*

Nilai <i>Silhouette Coefficient</i>	Kriteria Struktur Klaster
0.71 – 1.00	Struktur Klaster Kuat
0.51 – 0.70	Struktur Klaster Baik
0.26 – 0.50	Struktur Klaster Lemah
≤0.25	Struktur Klaster Buruk

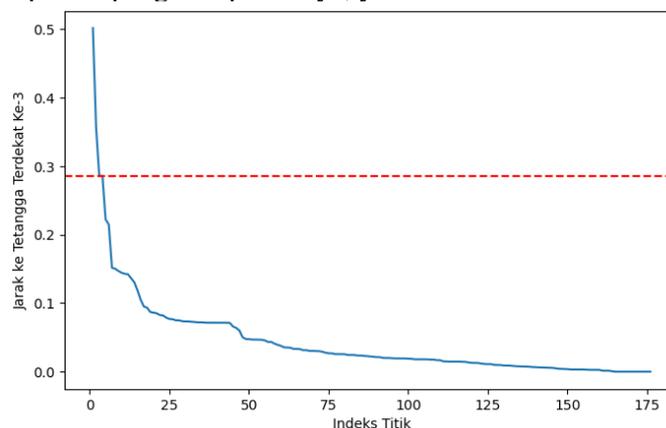
C. Hasil Penelitian dan Pembahasan

Data Pre-Processing

Sebelum melakukan analisis data, maka terlebih dahulu perlu dilakukan *data pre-processing*. Beberapa tahapan yang perlu dilakukan sebelum analisis klaster adalah *data cleaning* dan normalisasi data. Pada tahapan *data cleaning*, tidak ditemukan adanya *missing value* dan data duplikat. Selanjutnya, dilakukan normalisasi data untuk variabel *depth* dan magnitudo karena terdapat ketidaksamaan satuan pada kedua variabel tersebut. Metode normalisasi yang digunakan adalah metode normalisasi *min-max* yang mana dapat mengubah skala data agar berada pada rentang yang sama, yakni 0-1.

Penentuan Parameter ϵ dan *MinPts*

Dalam pengelompokan menggunakan metode DBSCAN terdapat 2 parameter penting yang mempengaruhi proses pengelompokannya, yakni ϵ dan *MinPts*.

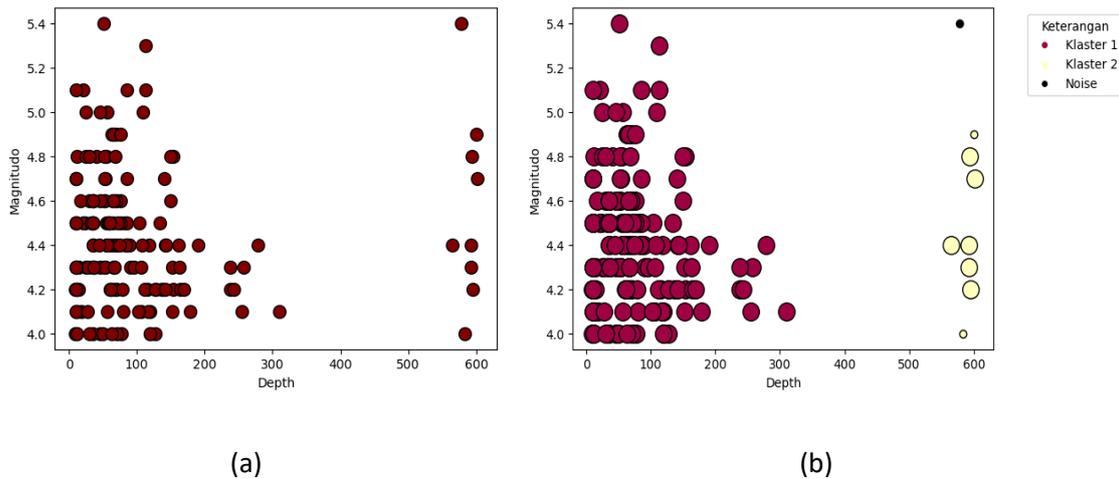


Gambar 1. Grafik *K-distance*

Berdasarkan grafik *k-distance* dapat diketahui bahwa nilai ϵ yang optimal berada pada radius 0,2857 melalui pengamatan *threshold point* yang ditandai dengan garis merah putus-putus. *Threshold point* atau nilai ambang batas merupakan sebuah radius pertama yang ada pada lembah pertama. Dalam hal ini yang dimaksud sebagai lembah adalah bagian grafik di mana penurunan berakhir dan mulai datar atau menurun berada pada radius 0,2857. Sedangkan, untuk nilai *MinPts* yang optimal diperoleh sebesar 4.

Analisis Klaster

Setelah diperoleh nilai ϵ dan *MinPts* yang optimal, maka dilakukan analisis klaster. Hasil pengelompokan atau analisis klaster untuk daerah rawan gempa bumi di Pulau Jawa pada rentang tanggal 1 Januari 2021-1 Januari 2024 berdasarkan kedalaman (*depth*) dan magnitudo kejadian gempa bumi dapat dilihat pada Gambar 2b.



Gambar 2. Perbandingan Sebaran Data Sebelum Analisis Klaster (a) dan Setelah Analisis Klaster (b)

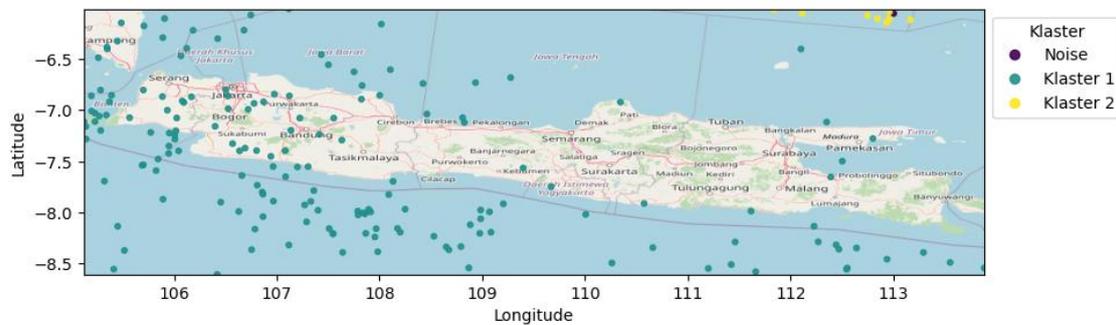
Gambar 2a menunjukkan bahwa sebaran data sebelum analisis klaster dilakukan berdasarkan variabel *depth* dan maganitudo memang tampak sudah terpisah dengan baik dan grafik sebaran data setelah analisis klaster memvalidasi hal tersebut serta mengidentifikasi adanya *noise* yang tidak terlalu tampak jelas pada Gambar 2a. Lebih lanjut pada Gambar 2b dapat diketahui mana yang merupakan *core point* dan *border point*. Perbedaannya terlihat dari ukurannya, *core point* memiliki ukuran yang jauh lebih besar dibandingkan dengan *border point* yang tampak lebih kecil dibandingkan dengan titik lainnya. Secara jumlahnya pun *border point* hanya ada 2 buah dan menjadi bagian dari anggota klaster 2. Sedangkan, pada klaster 1 tidak terdapat *border point* dan hanya terdiri dari *core point*. Hal tersebut juga disebabkan karena jarak antar titiknya yang berdekatan dan padat pada klaster 1 sehingga seluruh titiknya merupakan *core point*. Selain dapat melihat *core point* dan *border point*, hal lain yang dapat diketahui dari Gambar 2b adalah berapa banyak klaster yang terbentuk dan apakah terdapat *noise* atau tidak. Pada Gambar 2b terlihat bahwa terbentuk 2 klaster dengan 1 *noise*. Klaster 1 ditandai dengan warna ungu dan terdiri dari 167 anggota. Sedangkan, klaster 2 ditandai dengan warna cream dan terdiri atas 8 anggota.

Interpretasi Klaster

Tabel 2. Rata-Rata Klaster 1 dan Klaster 2

Nama Klaster	Variabel	Rata-Rata	
Klaster 1	<i>Depth</i>	74,0306	
	Magnitudo	4,42	
Klaster 2	<i>Depth</i>	590.6364	
	Magnitudo	4.46	

Berdasarkan Tabel 2 dapat dilihat bahwa kluster 1 memiliki kejadian gempa bumi dengan rata-rata kedalamannya sebesar 74,0306 km sehingga mengacu pada rata-rata tersebut dapat dikatakan bahwa gempa bumi pada daerah kluster 1 termasuk ke dalam jenis gempa bumi menengah. Sedangkan, menurut magnitudonya, kluster 1 memiliki kejadian gempa bumi dengan rata-rata kekuatan sebesar 4,42 mb. Selanjutnya, kluster 2 memiliki kejadian gempa bumi dengan rata-rata kedalamannya yang lebih besar dibandingkan dengan kluster 1, yakni sebesar 590,6364 km. Berbeda dengan kluster 1, kluster 2 memiliki kejadian gempa bumi dengan jenis gempa bumi dalam. Sedangkan, menurut magnitudonya, kluster 2 memiliki kejadian gempa bumi dengan rata-rata kekuatan yang juga lebih besar dibandingkan dengan kluster 1, yakni sebesar 4,46 mb. Apabila dibandingkan dengan rata-rata dari kedalaman dan kekuatan gempa bumi keseluruhan di Pulau Jawa pada tanggal 1 Januari 2021-1 Januari 2024, kluster 2 memiliki nilai rata-rata yang jauh lebih besar untuk kedua variabel tersebut. Adapun daerah yang menjadi *noise* memiliki kejadian gempa bumi dengan kedalaman sebesar 578,0500 km yang termasuk ke dalam jenis gempa bumi dalam. Namun, besaran magnitudonya, yakni sebesar 5,4 mb yang mana jauh lebih besar dibandingkan dengan rata-rata magnitudo pada kejadian gempa bumi yang terjadi pada daerah anggota kluster 1 maupun kluster 2. Untuk sebaran daerahnya dapat dilihat pada peta di bawah ini.

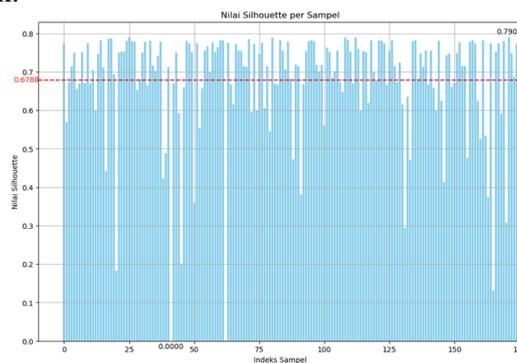


Gambar 3. Peta Sebaran Daerah Kejadian Gempa Bumi

Berdasarkan Gambar 3, dapat dilihat pada kluster 1 titik-titik kejadian gempa bumi menyebar di berbagai bagian Pulau Jawa, tetapi cenderung bertumpuk di daerah Jawa Barat. Hal tersebut kemungkinan dikarenakan daerah Jawa Barat memiliki 15 gunung berapi aktif atau setengahnya dari jumlah gunung berapi aktif di Pulau Jawa [3]. Sedangkan, kluster 2 dan *noise* titik-titik kejadian gempa buminya bertumpuk di Laut Jawa. Namun, secara keseluruhan titik-titik kejadian gempa bumi ini banyak berpusat di laut dibandingkan di daratan.

Validasi Kluster

Analisis dan interpretasi kluster telah dilakukan, maka selanjutnya yang perlu dilakukan adalah validasi struktur kluster yang telah terbentuk menggunakan *silhouette coefficient*. Untuk memperoleh nilai *silhouette coefficient*, maka harus diketahui terlebih dahulu *silhouette score* untuk masing-masing titik.



Gambar 4. Grafik Hasil Perhitungan Nilai *Silhouette* untuk Setiap Titik

Pada Gambar 4 menunjukkan hasil perhitungan nilai *silhouette score* pada setiap titik yang telah dilakukan pada *software Python* dengan nilai paling tinggi sebesar 0,7902 dan nilai paling kecilnya sebesar 0. Garis merah putus-putus pada grafik menunjukkan nilai *silhouette coefficient* itu sendiri, yakni sebesar 0,6778. Sebagaimana kriteria pengukuran dari *silhouette coefficient* pada Tabel 1, maka struktur klaster dari hasil analisis klaster pada daerah rawan gempa bumi di Pulau Jawa pada rentang tanggal 1 Januari 2021-1 Januari 2024 dapat dikategorikan sebagai struktur klaster yang baik.

D. Kesimpulan

Berdasarkan hasil analisis dan pembahasan dalam penelitian ini, maka dapat disimpulkan beberapa hal sebagai berikut:

1. Pengelompokan daerah rawan gempa bumi di Pulau Jawa pada rentang tanggal 1 Januari 2021-1 Januari 2024 berdasarkan variabel *depth* (kedalaman) dan magnitudo kejadian gempa bumi dengan metode analisis klaster DBSCAN menghasilkan 2 klaster dengan 1 *noise*. Hasil tersebut diperoleh dengan menggunakan parameter ϵ sebesar 0.2857 dan *MinPts* sebesar 4. Klaster 1 terdiri dari 167 daerah dan klaster 2 terdiri dari 8 daerah yang rawan dengan jenis gempa bumi menengah dan rata-rata magnitudo sebesar 4,42 mb. Sedangkan, klaster 2 merupakan daerah yang rawan dengan jenis gempa bumi dalam dengan rata-rata magnitudo sebesar 4,46 mb. Di antara kedua klaster tersebut, terdapat juga satu daerah sebagai *noise* yang dengan karakteristik rawan akan jenis gempa bumi dalam dan diikuti dengan magnitudo yang besar, yakni sebesar 5,4 mb.
2. Pengelompokan daerah rawan gempa bumi di Pulau Jawa pada rentang tanggal 1 Januari 2021-1 Januari 2024 berdasarkan variabel *depth* (kedalaman) dan magnitudo kejadian gempa bumi menghasilkan struktur klaster yang baik. Hal tersebut dilihat dari nilai *silhouette coefficient* nya, yakni sebesar 0,6788 dalam artian setiap klasternya mampu terpisahkan dengan cukup baik dengan menggunakan metode DBSCAN.

Acknowledge

Penelitian ini tidak akan selesai tanpa bantuan serta dukungan dari banyak pihak. Penulis ingin mengucapkan terima kasih yang tulus kepada semua pihak yang telah ikut membantu dalam penyelesaian penelitian ini. Terutama kepada Bapak Ilham Faishal Mahdy, S.Stat., M. Stat., sebagai dosen pembimbing yang telah memberikan ilmu, saran, dan waktu yang berharga bagi penulis. Tidak lupa juga kepada orang tua serta teman-teman saya yang telah memberikan dukungan moral sampai akhir.

Daftar Pustaka

- [1] Azuri, D. F., Zulhanif, & Pontoh, R. S. (2016). Pengelompokan Kabupaten/Kota Di Pulau Jawa Berdasarkan Pembangunan Manusia Berbasis Gender Menggunakan Bisecting K-Means. *Prosiding Seminar Nasional MIPA 2016*, 78–83.
- [2] Anggarwai, C. C., & Reddy, C. K. (2013). *Data Clustering: Algorithms and Applications*. CRC Press.
- [3] Bariklana, M., & Fauzan, A. (2023). Implementation of The DBSCAN Method for Cluster Mapping of Earthquake Spread Location. *Journal of Mathematics and Its Applications*, 17(2), 867-878.
- [4] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *KDD-96 Proceedings*, 96(34), 226-231.
- [5] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1998). Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and Its Applications. *KDD-96 Proceedings*, 96(34), 169-194.
- [6] Furqon, M. T., & Muflikhah. L. (2016). Clustering The Potential Risk of Tsunami Using Density-Based Spatial Clustering of Application with Noise. *Journal of Environmental Engineering & Sustainable Technology*, 3(1), 1-8.

- [7] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques*. (Third Edition). Waltham: Elsevier.
- [8] Hidayat, N., & Santoso, E. W. (1997). Gempa Bumi dan Mekanismenya. *Alami: Jurnal Teknologi Reduksi Risiko Bencana*, 2(3), 50-52.
- [9] Khurin'in, A. I. (2021). Pengelompokan Kabupaten/Kota di Provinsi Jawa Barat Berdasarkan Tingkat Sebaran Pengangguran Menggunakan Metode Density-Based Spatial Clustering Algorithm with Noise (DBSCAN). Surabaya. Program Studi Matematika. Universitas Islam Negeri Sunan Ampel Surabaya.
- [10] Mumtaz, K., & Duraiswamy, K. (2010). An Analysis on Density Based Clustering of Multi dimensional Spatial Data. *Indian Journal of Computer Science and Engineering*, 1(1), 8-12.
- [11] Rousseeuw, P. J. (1987). Silhouettes: A Graphical Aid to The Interpretation and Validation of Cluster Analysis. *Journal of computational and applied mathematics*, 20, 53-65.
- [12] Zakaria, Z., Ismawan., & Haryanto, I. (2011). Identifikasi dan Mitigasi pada Zona Rawan Gempa Bumi di Jawa Barat. *Bulletin of Scientific Contribution*, 9(1), 35-41
- [13] Karimuse, W. Y., Nohe, D. A., & Siringoringo, M. (n.d.). Pendekatan Regresi Nonparametrik Kernel pada Data IHSG Periode Januari 2020-Desember 2021 (Vol. 23, Issue 1). <https://www.idx.co.id/>
- [14] Meidianingsih, Q., Wardani, D. E., Salsabila, E., Nafisah, L., & Mutia, A. N. (n.d.). Perbandingan Performa Metode Berbasis Support Vector Machine untuk Penanganan Klasifikasi Multi Kelas Tidak Seimbang (Vol. 23, Issue 1).
- [15] Samudera Kutaraja Banda Aceh Rizwan, P., Saputra, E., Shiddiq, M., Studi Pemanfaatan Sumberdaya Perikanan, P., Syiah Kuala, U., & Studi Statistika, P. (2023). Analisis Jumlah Kedatangan Kapal terhadap Hasil Tangkapan Ikan di Pelabuhan (Vol. 23, Issue 1).