

Clustering Daerah Rawan Gempa di Pulau Jawa Berbasis Metode Algoritma K-Means

Shinta Febriyani¹, Hafid Buroiroh², Narantaka Wahyu³, Nanda Tri Septiani⁴, Muhammad Faiz⁵

^{1,2,3,4,5}Sistem Informasi

Universitas Bina Sarana Informatika

Jl. Kramat Raya No.98, Jakarta Pusat 10450

e-mail: ¹shintafebriyani25@gmail.com, ²hafidburoiroh09@gmail.com, ³narantakawahyu450@gmail.com, ⁴trinandat8@gmail.com, ⁵thefaiz17@gmail.com

Artikel Info : Diterima : 05-12-2025 | Direvisi : 17-12-2025 | Disetujui : 18-12-2025

Abstrak - Peningkatan aktivitas gempa bumi dalam beberapa tahun terakhir menjadi perhatian serius, terlebih dengan maraknya pembahasan mengenai potensi gempa megathrust di wilayah selatan Jawa. Kondisi ini menunjukkan pentingnya pemetaan kerawanan yang lebih akurat, mengingat Pulau Jawa merupakan salah satu area dengan intensitas seismik tinggi. Meskipun data historis gempa tersedia cukup lengkap, pemetaan sebelumnya masih bersifat umum dan belum mampu menggambarkan pola spasial secara detail. Penelitian ini bertujuan mengidentifikasi tingkat kerawanan gempa di Pulau Jawa menggunakan algoritma *K-Means Clustering*. Data diambil dari platform Kaggle berupa katalog gempa bumi periode 2008–2023, namun analisis difokuskan pada Januari 2021 hingga Januari 2023 agar sesuai dengan kondisi seismik terbaru. Proses klusterisasi dilakukan menggunakan RapidMiner dengan *Euclidean Distance*, sedangkan jumlah kluster optimal ditentukan melalui *Davies-Bouldin Index* (DBI), yang menghasilkan nilai terbaik sebesar 0,349. Hasil klusterisasi menunjukkan tiga tingkat kerawanan, yaitu *Cluster 0* dengan 2.348 titik (kerawanan tinggi), *Cluster 2* dengan 276 titik (kerawanan sedang), dan *Cluster 1* dengan 12 titik (kerawanan rendah). Hasil tersebut memperlihatkan bahwa sebagian besar aktivitas gempa khususnya yang berada di zona subduksi selatan Jawa yang berpotensi memicu gempa megathrust terkonsentrasi pada *cluster* berkerawanan tinggi. Penelitian ini diharapkan dapat memberikan gambaran yang lebih presisi mengenai distribusi kerawanan gempa di Pulau Jawa, sehingga dapat menjadi dasar bagi BPBD dan BNPB dalam penyusunan strategi mitigasi dan sistem peringatan dini.

Kata Kunci : Gempa Bumi, K-Means, Pulau Jawa

Abstracts - The increase in earthquake activity in recent years has become a serious concern, especially with the growing discussion regarding the potential for a megathrust earthquake in the southern region of Java. This condition highlights the need for more accurate vulnerability mapping, considering that Java Island is one of the areas with high seismic intensity. Although historical earthquake data are widely available, previous mapping efforts have remained general and have not been able to illustrate spatial patterns in detail. This study aims to identify earthquake vulnerability levels on Java Island using the *K-Means Clustering* algorithm. The data were obtained from the Kaggle platform in the form of an earthquake catalog from 2008–2023, but the analysis focused on the period from January 2021 to January 2023 to better reflect recent seismic conditions. The clustering process was conducted using RapidMiner with *Euclidean Distance*, while the optimal number of clusters was determined using the *Davies–Bouldin Index* (DBI), which produced the best value of 0.349. The results show three levels of vulnerability, namely *Cluster 0* with 2,348 points (high vulnerability), *Cluster 2* with 276 points (medium vulnerability), and *Cluster 1* with 12 points (low vulnerability). These results indicate that most earthquake activity particularly along the southern Java subduction zone, which has the potential to trigger a megathrust event is concentrated within the high-vulnerability cluster. This study is expected to provide a more precise overview of earthquake-prone areas in Java Island, thereby serving as a reference for BPBD and BNPB in designing mitigation strategies and early warning systems.

Keywords : Earthquakes, Java Island, K-Means

PENDAHULUAN



Bencana alam merupakan peristiwa yang muncul akibat proses alamiah dan dapat menimbulkan kerugian yang signifikan bagi manusia maupun lingkungan. Salah satu bencana alam yang memiliki dampak besar adalah gempa bumi (Agustina & Mulyawan, 2023). Ditambah dengan kondisi letak wilayah Indonesia yang berada pada zona pertemuan empat lempeng utama yaitu Eurasia, Indo Australia, Filipina dan Pasifik (Halik & Septiana, 2022), membuat Indonesia memiliki risiko gempa bumi yang tinggi akibat seringnya aktivitas dan tumbukan antar lempeng tersebut (Rahmi et al., 2024). Gempa bumi sendiri memiliki definisi sebagai getaran pada permukaan bumi yang terjadi akibat pelepasan energi secara tiba-tiba dari patahan di dalam kerak bumi (Suara et al., 2023). Energi tersebut kemudian merambat sebagai gelombang seismik dan menimbulkan guncangan di permukaan tanah. Dalam beberapa kasus, gempa bumi dapat memicu bencana lanjutan seperti tsunami, yang berpotensi memperbesar kerusakan dan menambah jumlah korban jiwa (Maulana Kharyska Abadi et al., 2025).

Proses terjadinya gempa pada kenyataannya sangat sulit untuk diteliti secara langsung, karena melibatkan interaksi kompleks antara materi dan energi pada lapisan bumi. Hingga kini, belum ada metode atau lembaga yang mampu memprediksi waktu terjadinya gempa secara tepat (Cahyo et al., 2023). Di sisi lain, perkembangan teknologi saat ini telah mengalami kemajuan yang sangat pesat. Setiap orang dapat dengan mudah mengakses dan memperoleh informasi secara cepat serta akurat. Kondisi ini membuat masyarakat semakin mudah mengetahui berbagai informasi terkait bencana alam yang terjadi. Meskipun informasi yang disediakan bukan berupa prediksi waktu terjadinya gempa, namun berbagai platform seperti BMKG secara rutin memberikan informasi terkini mengenai aktivitas gempa yang terjadi di suatu wilayah. Informasi tersebut setidaknya dapat menjadi acuan bagi masyarakat untuk tetap waspada dan meningkatkan kesiapsiagaan terhadap potensi bencana gempa bumi (Agustina & Mulyawan, 2023).

Metode *K-Means Clustering* merupakan teknik dalam *data mining* yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok berdasarkan tingkat kemiripannya (Baldah et al., 2023). Dalam penelitian ini, metode tersebut diterapkan untuk mengidentifikasi wilayah yang berpotensi rawan gempa. Proses pengelompokannya dilakukan dengan menganalisis data gempa yang memiliki karakteristik serupa, sehingga setiap kelompok dapat menggambarkan tingkat kerawanan gempa pada suatu daerah. Beberapa penelitian terdahulu telah menunjukkan bahwa metode *K-Means* efektif digunakan dalam analisis berbasis spasial (Febrian et al., 2025). Namun, beberapa penelitian sebelumnya masih menerapkan pendekatan berbasis wilayah administratif dalam pemetaan risiko gempa, sehingga hasilnya kurang akurat dalam mempresentasikan distribusi spasial titik-titik kejadian gempa (Febrian et al., 2025).

Oleh karena itu, penelitian ini dilakukan dengan tujuan utama untuk mengidentifikasi dan memetakan area-area yang berpotensi rawan gempa bumi di Pulau Jawa menggunakan metode *K-Means Clustering*. Metode tersebut dipilih karena kemampuannya dalam mengelompokkan data berdasarkan kesamaan karakteristik, sehingga wilayah dengan pola aktivitas seismik yang mirip dapat terkumpul dalam satu kluster yang mencerminkan tingkat kerawannya. Adapun data yang digunakan dalam penelitian ini berasal dari sumber terbuka di platform Kaggle, yang menyediakan rekaman kejadian gempa bumi di Indonesia selama periode November 2008 hingga Januari 2023. Namun, data yang digunakan dalam analisis difokuskan hanya pada kejadian gempa mulai Januari 2021 hingga Januari 2023 agar hasil penelitian lebih relevan dengan kondisi seismik terbaru. Variabel utama yang dianalisis meliputi koordinat lintang, bujur, kedalaman, dan magnitudo gempa.

Melalui proses pengelompokan tersebut, penelitian ini berupaya menghasilkan visualisasi sebaran spasial wilayah rawan gempa di Pulau Jawa secara lebih presisi, sehingga mampu menggambarkan tingkat kerawanan dari kategori rendah, sedang, hingga tinggi. Hasil akhir dari penelitian ini diharapkan dapat memberikan informasi yang akurat dan terukur mengenai daerah-daerah dengan potensi risiko gempa tinggi, yang dapat dimanfaatkan oleh masyarakat maupun pihak berwenang sebagai dasar dalam meningkatkan kewaspadaan, kesiapsiagaan, serta melakukan langkah mitigasi dini guna meminimalkan dampak kerugian apabila gempa bumi terjadi di masa mendatang.

METODOLOGI PENELITIAN

Pada penelitian ini kami menggunakan beberapa tahapan seperti pengumpulan data, proses *KDD*, Penerapan *K-Means Clustering* menggunakan *Microsoft Excel* dan penerapan *K-Means Clustering* menggunakan *RapidMiner*.

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diambil dari platform Kaggle. Dataset tersebut tersedia pada tautan <https://www.kaggle.com/datasets/kekavigi/earthquakes-in-indonesia>. Sumber data tersebut dipilih karena memiliki cakupan yang luas dan disajikan secara terstruktur, sehingga sangat mendukung proses analisis serta pengelompokan wilayah rawan gempa. Dataset ini merupakan hasil kompilasi dari dua sumber utama, yaitu Katalog Gempa Bumi dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) serta *Earthquake Catalog* milik *U.S. Geological Survey (USGS)*. Data yang dihimpun mencakup kejadian gempa bumi di Indonesia dari tahun 2008 hingga 2023, dengan total 13 atribut. Atribut tersebut meliputi tanggal kejadian

(tgl), waktu kejadian (ot), latitude, longitude, kedalaman (*depth*), magnitudo (mag), lokasi kejadian (*remark*), serta informasi mekanisme fokus seperti strike, dip, dan rake.

	tgl	ot	lat	lon	depth	mag	remark	strike1	dip1	rake1	strike2	dip2	rake2
0	2008/11/01	21:02:43.058	-9.18	119.06	10	4.9	Sumba Region - Indonesia	NaN	NaN	NaN	NaN	NaN	NaN
1	2008/11/01	20:58:50.248	-6.55	129.64	10	4.6	Banda Sea	NaN	NaN	NaN	NaN	NaN	NaN
2	2008/11/01	17:43:12.941	-7.01	106.63	121	3.7	Java - Indonesia	NaN	NaN	NaN	NaN	NaN	NaN
3	2008/11/01	16:24:14.755	-3.30	127.85	10	3.2	Seram - Indonesia	NaN	NaN	NaN	NaN	NaN	NaN
4	2008/11/01	16:20:37.327	-6.41	129.54	70	4.3	Banda Sea	NaN	NaN	NaN	NaN	NaN	NaN

Sumber: Kaggle (Earthquakes in Indonesia Dataset, 2008–2023) (2025)

Gambar 1. Dataset Awal

2. Proses KDD (*Knowledge Discovery in Database*)

Tahapan selanjutnya adalah proses KDD (Lestari & Mulyawan, 2023). Adapun tahapan dalam proses KDD dapat dijelaskan sebagai berikut:

a. Seleksi Data (Data Selection)

Seleksi ini dilakukan untuk mendapatkan kelompok data berdasarkan atribut yang diperlukan. Pada penelitian ini kami menggunakan 5 dari 13 atribut yang tersedia, diantaranya adalah remark, latitude, longitude, depth, dan magnitudo. sebagai fokus pengerjaan penelitian. Pemilihan atribut ini didasarkan pada pertimbangan bahwa variabel tersebut bersifat numerik dan memiliki pengaruh langsung terhadap karakteristik spasial dan intensitas gempa, sehingga lebih sesuai untuk proses pengelompokan berbasis jarak. Atribut lainnya digunakan sebagai pelengkap informasi atau untuk keperluan identifikasi hasil kluster.

b. Preprocessing

Pada tahap preprocessing, dilakukan proses *cleaning* untuk memastikan bahwa hanya data gempa yang relevan dengan wilayah penelitian yang digunakan. Atribut remark, yang berisi informasi lokasi terjadinya gempa dimanfaatkan sebagai dasar untuk melakukan penyaringan data. Pada tahap ini, hanya data dengan remark yang menunjukkan lokasi di Pulau Jawa yang dipertahankan, sedangkan data dengan remark di luar wilayah tersebut dihapus agar analisis benar-benar berfokus pada area penelitian. Setelah proses penyaringan, dilakukan *reduction* untuk menyederhanakan atribut tertentu agar lebih sesuai dengan kebutuhan analisis, termasuk menghilangkan bagian remark yang tidak digunakan pada tahap klusterisasi. Remark tetap disimpan untuk proses identifikasi dan pengecekan data, namun tidak digunakan sebagai atribut perhitungan sehingga diatur sebagai "id" di RapidMiner.

c. Transformasi Data

Transformasi data merupakan proses mengubah format atau tipe data agar sesuai dengan kebutuhan analisis. Pada penelitian ini, tahap transformasi dilakukan dengan mengonversi atribut yang sebelumnya bersifat non-numerik menjadi nilai numerik sehingga dapat diproses dalam metode klusterisasi.

d. Data Mining

Tahap data mining merupakan bagian utama dalam proses *KDD*. Pada penelitian ini, teknik *data mining* yang digunakan adalah algoritma *K-Means Clustering* untuk mengelompokkan data gempa berdasarkan kemiripan karakteristiknya. Evaluasi performa model dilakukan menggunakan *Davies-Bouldin Index (DBI)* guna menentukan jumlah *cluster* yang paling optimal. Selain itu, penelitian ini menerapkan parameter *Numerical Measures* dengan menggunakan *Euclidean Distance* sebagai metode pengukuran kedekatan antar data. Proses iterasi kemudian dijalankan hingga nilai centroid dan keanggotaan setiap cluster mencapai kondisi stabil atau tidak mengalami perubahan lagi

e. Evaluasi

Pada tahap evaluasi dilakukan analisis terhadap hasil klusterisasi untuk menilai makna, pola, serta relevansi temuan yang diperoleh secara keseluruhan.

3. Penerapan K-Means Clustering pada Microsoft Excel

Tahap selanjutnya adalah penerapan *K-Means Clustering* pada *Microsoft Excel*. Pada tahap ini kami melakukan perhitungan jarak antar data menggunakan metode *Euclidean Distance* dengan terlebih dahulu menetapkan *centroid* awal yang dipilih secara acak. Perhitungan jarak tersebut digunakan untuk menentukan kedekatan setiap data terhadap masing-masing *centroid*, sehingga setiap titik dapat dikelompokkan ke dalam *cluster* yang paling sesuai (Karmanita & Hendrik, 2023). Setelah itu, nilai *centroid* diperbarui berdasarkan rata-rata anggota *cluster*, lalu proses perhitungan jarak diulangi secara iteratif hingga komposisi *cluster* dan posisi *centroid* tidak lagi berubah (Aria, 2024). Dengan demikian, proses klusterisasi dapat terbentuk secara sistematis berdasarkan kesamaan karakteristik data.

4. Penerapan K-Means Clustering pada RapidMiner

Pada tahapan yang terakhir dilakukan pengujian menggunakan *Davies Bouldin Index (DBI)* dengan bantuan

software RapidMiner. Tahap ini bertujuan untuk mengevaluasi kualitas hasil klasterisasi dan memastikan apakah metode yang diterapkan mampu menghasilkan kelompok data yang relevan dengan rumusan masalah penelitian. Nilai *DBI* digunakan sebagai indikator untuk menilai seberapa baik pemisahan dan kekompakan antar *cluster*. Semakin kecil nilai *DBI*, semakin baik kualitas klasterisasi yang dihasilkan. Algoritma *K-Means Clustering* dianggap efektif apabila pengujian ini menghasilkan nilai *DBI* berada pada kisaran rendah (Dwi Mauluda et al., 2025).

HASIL DAN PEMBAHASAN

1. Perhitungan dengan Algoritma K-Means

Data yang digunakan dalam penelitian ini adalah data katalog gempa bumi yang terjadi di Indonesia tahun 2008 hingga 2023, namun untuk mendapatkan hasil yang lebih relevan dalam penelitian ini data yang digunakan hanya data dari Januari 2021 hingga Januari 2023 dan hanya mencakup wilayah Pulau Jawa. Data yang akan diolah terdapat pada tabel 1 yang berisi 5 atribut yang terdiri dari remark, lat, lon, depth dan mag yang sebelumnya sudah dilakukan pembersihan.

Tabel 1. Data Gempa Bumi

No	remark	lat	lon	depth	mag
1	Java - Indonesia	-8.38	108.1	19	3
2	South of Java - Indonesia	-9.25	113.1	53	3.4
3	South of Java - Indonesia	-9.71	112.9	10	3.3
2635	Java - Indonesia	-7.98	110.5	10	3
2636	Java - Indonesia	-7.23	110.2	240	3.8

Sumber: Hasil Penelitian (2025)

a. Menentukan Jumlah Cluster

Dalam model ini, jumlah *cluster* yang ditentukan adalah sebanyak 3 *cluster*. Masing-masing *cluster* dipresentasikan sebagai tingkat rawan gempa tinggi yaitu *cluster* 1, tingkat rawan gempa sedang yaitu *cluster* 0 dan tingkat rawan gempa rendah yaitu *cluster* 2.

b. Menentukan Nilai Centroid

Pemilihan nilai *centroid* awal dilakukan dengan cara mengambil data secara acak. Berdasarkan data gempa bumi yang diambil sebagai *centroid* awal adalah data dari baris ke 1 (pertama) sebagai pusat *cluster* 0, data ke 1317 (tengah) sebagai pusat *cluster* 1 dan data ke 2636 (akhir) sebagai pusat *cluster* 2.

Tabel 2. Data Pusat *Centroid* Awal

No	remark	lat	lon	depth	mag
1	Java - Indonesia	-8.38	108.1	19	3
2	Java - Indonesia	-8.93	111.3	14	3.4
3	Java - Indonesia	-7.23	110.2	240	3.8

Sumber: Hasil Penelitian (2025)

c. Menentukan Jarak Cluster

Penentuan letak setiap *cluster* dilakukan dengan membandingkan nilai jarak dari masing-masing titik data terhadap ketiga *centroid*, di mana nilai jarak yang paling kecil menjadi penentu *cluster* yang dipilih. Proses perhitungan jarak ini menggunakan metode *Euclidean Distance* yang dihitung dengan bantuan *Microsoft Excel*. Perhitungan jarak dilakukan untuk mengetahui kedekatan setiap titik terhadap *centroid* masing-masing *cluster*, baik ke *centroid cluster* 0, *centroid cluster* 1, maupun *centroid cluster* 2. Selanjutnya, data dikelompokkan berdasarkan jarak yang paling dekat sehingga menghasilkan pemetaan *cluster* yang lebih akurat. Hasil perhitungan jarak *cluster* disajikan dalam tabel 3.

Tabel 3. Jarak *Cluster* Iterasi 1

No	remark	C0	C1	C2	Jarak Terdekat
1	Java - Indonesia	0	5,9538307	221,0142278	0
2	South of Java - Indonesia	34,37756536	39,04422236	187,0339769	34,37756536
3	South of Java - Indonesia	10,26286997	4,372047575	230,0292938	4,372047575

2635	Java - Indonesia	9,31034371	4,209346743	230,0027726	4,209346743
2636	Java - Indonesia	221,0142278	226,0093281	0	0

Sumber: Hasil Penelitian (2025)

d. Kelompokkan Objek Berdasarkan Jarak Minimum

Setelah menentukan posisi setiap data gempa berdasarkan jarak minimum terhadap masing-masing *centroid*, langkah selanjutnya adalah mengelompokkan data tersebut sesuai dengan *cluster* yang memiliki jarak terdekat. Hasil pengelompokan data gempa ke dalam cluster kemudian disajikan pada tabel 4.

Tabel 4. Hasil *Cluster* Iterasi 1

No	remark	C0	C1	C2	Jarak Terdekat	Cluster
1	Java - Indonesia	0	5,9538307	221,0142278	0	0
2	South of Java - Indonesia	34,37756536	39,04422236	187,0339769	34,37756536	0
3	South of Java - Indonesia	10,26286997	4,372047575	230,0292938	4,372047575	1
2635	Java - Indonesia	9,31034371	4,209346743	230,0027726	4,209346743	1
2636	Java - Indonesia	221,0142278	226,0093281	0	0	2

Sumber: Hasil Penelitian (2025)

e. Menentukan Nilai Centroid Baru

Setelah hasil pengelompokan diperoleh, langkah berikutnya adalah menentukan centroid baru dengan menghitung nilai rata-rata dari setiap cluster (Saputra & Nataliani, 2021). Setelah proses iterasi dilakukan sebanyak 13 kali dan tidak terjadi perubahan anggota cluster pada iterasi ke-12 dan ke-13, maka proses iterasi dinyatakan selesai dan dapat dilanjutkan ke tahap berikutnya. Begitu juga dengan nilai *SSE* (*Sum of Squared Errors*) menunjukkan penurunan yang cukup signifikan, yang artinya entitas yang masuk ke dalam jumlah *cluster* pada metode K-Means memiliki tingkat keseragaman atau kemiripan yang semakin tinggi di dalam setiap *cluster*nya (Maulana Kharyska Abadi et al., 2025). Adapun hasil dari setiap iterasi disajikan pada tabel 5.

Tabel 5. Tabel Iterasi 1 Sampai Iterasi 13

Iterasi	Jumlah Anggota <i>Cluster</i>			SSE
	C0	C1	C2	
1	1038	1510	88	3425448,685
2	631	1925	80	2615682,257
3	459	2126	51	2279505,558
4	385	2217	34	1724515,522
5	343	2265	28	1404976,609
6	318	2292	26	1319553,37
7	304	2309	23	1300729
8	298	2319	19	1263567
9	290	2330	16	1198293
10	289	2334	13	1132659
11	282	2342	12	1083613
12	279	2345	12	1077090,7
13	279	2345	12	1076998,51

Sumber: Hasil Penelitian (2025)

Jumlah anggota *cluster* yang paling mendekati hasil pada RapidMiner muncul pada iterasi ke-12, dan komposisinya tetap stabil pada iterasi berikutnya, yaitu 279 anggota C0, 2345 anggota C1 dan 12 anggota C2.

2. Pengujian Data dengan RapidMiner

Pemodelan proses analisis pada aplikasi RapidMiner dapat dilihat pada gambar 2, yang memperlihatkan

rangkaian operator serta alur kerja yang digunakan dalam tahap pengolahan data.



Sumber: Hasil Penelitian (2025)
Gambar 2. Model *Data Mining*

a. Read Excel

Operator pertama yang digunakan dalam proses *data mining* adalah *Read Excel*, yang berfungsi untuk memuat dataset dalam format *Excel*. Untuk mengimpor data, pengguna dapat membuka operator *Read Excel* dengan mengkliknya dua kali. Setelah itu, RapidMiner akan menampilkan jendela *Import Configuration Wizard* yang memungkinkan pengguna menelusuri *folder* penyimpanan dan memilih *file* dataset yang akan dimasukkan ke dalam RapidMiner. Selanjutnya, format data perlu disesuaikan dengan kebutuhan algoritma K-Means dengan menetapkan tipe data yang tepat untuk setiap atribut pada dataset.

b. Clustering

Data yang telah disimpan pada *local repository* RapidMiner selanjutnya diolah menggunakan metode *clustering* dengan terlebih dahulu menempatkan dataset ke dalam *workspace* RapidMiner. Pada penelitian ini digunakan jumlah *cluster* $k = 3$, sesuai dengan kebutuhan pengelompokan data gempa bumi berdasarkan karakteristik utama seperti kedalaman, magnitudo, dan koordinat lokasi gempa pada wilayah Pulau Jawa.

c. Performance

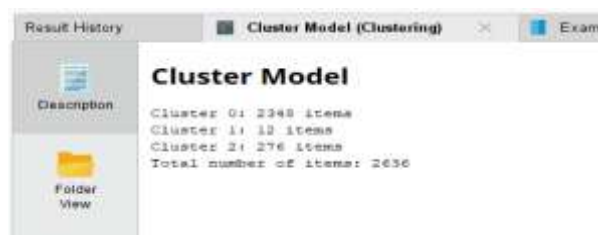
Cluster Distance Performance digunakan sebagai metode untuk mengukur kualitas hasil pengelompokan. Operator ini menghasilkan daftar nilai kinerja berdasarkan *centroid* dari setiap *cluster*. Mekanismenya adalah dengan mengambil model *centroid cluster* serta data yang telah dikelompokkan sebagai input, kemudian melakukan evaluasi untuk menghitung nilai *Davies-Bouldin Index* sebagai indikator pemisahan dan keseragaman *cluster*.



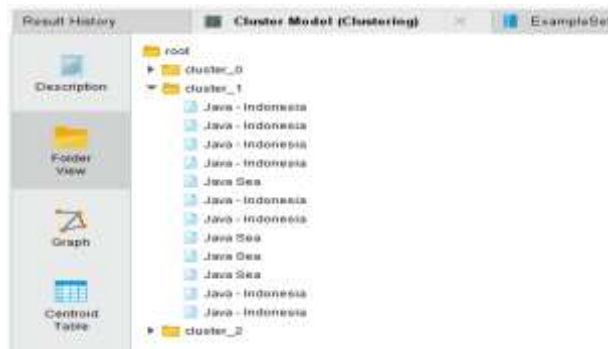
Sumber: Hasil Penelitian (2025)
Gambar 3. Nilai DBI Optimal pada *Cluster 3*

d. Cluster Model

Hasil cluster model dapat dilihat pada gambar 4 yang menunjukkan cluster beserta jumlah masing-masing anggotanya.



Sumber: Hasil Penelitian (2025)
Gambar 4. *Cluster Model*

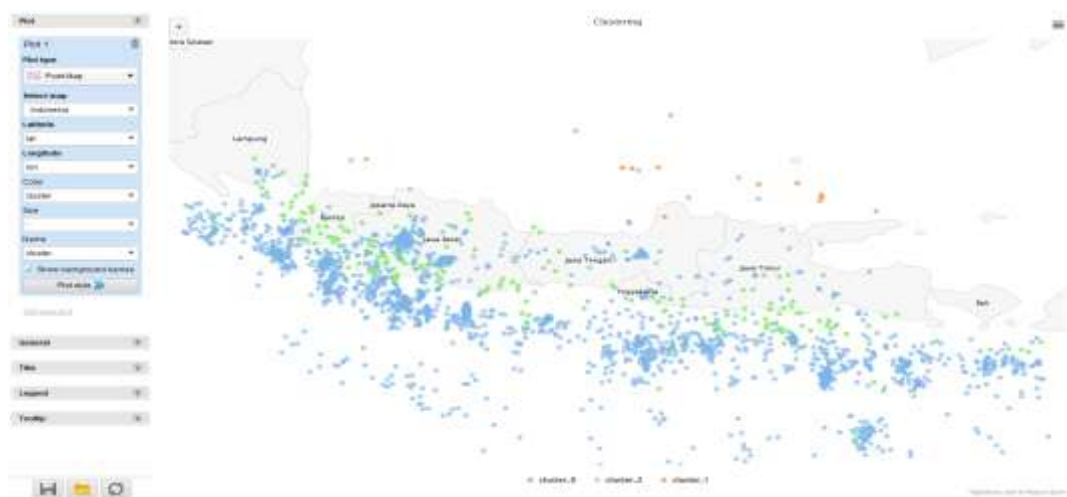


Sumber: Hasil Penelitian (2025)
Gambar 5. Anggota Cluster 1

Berdasarkan hasil yang ditunjukkan pada gambar 4 dan 5, diperoleh pengelompokan data gempa berdasarkan *cluster model* yaitu, *cluster 0* yang memiliki jumlah data terbesar yaitu sebanyak 2.348 titik menunjukkan tingginya frekuensi gempa. Selanjutnya, *cluster 1* berisi 12 titik, kelompok ini merupakan kategori kerawanan rendah. Sementara itu, *cluster 2* mencakup 276 titik tidak melebihi jumlah data pada *cluster 0* dan tidak lebih sedikit dari *cluster 1*, sehingga *cluster* ini masuk kedalam kategori kerawanan sedang.

e. Visualisasi

Gambar 6 memperlihatkan hasil pemetaan titik gempa pada wilayah Pulau Jawa berdasarkan *cluster* yang dihasilkan dari proses K-Means. *Cluster 0* ditandai dengan warna biru dan terlihat mendominasi hampir seluruh wilayah selatan Pulau Jawa, menunjukkan tingginya frekuensi gempa dangkal dengan magnitudo rendah. *Cluster 2* ditampilkan dengan warna hijau dan tersebar di beberapa titik, terutama di Jawa Barat, Jawa Tengah, dan sebagian wilayah Jawa Timur. Kelompok ini menggambarkan gempa dengan kedalaman dan magnitudo menengah. Sementara itu, *Cluster 1* yang ditandai dengan warna oranye memiliki jumlah titik paling sedikit dan tersebar secara terpisah, menunjukkan gempa yang terjadi pada kedalaman sangat besar dengan magnitudo lebih tinggi, namun frekuensinya rendah.



Sumber: Hasil Penelitian (2025)
Gambar 6. Visualisasi Anggota Cluster

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan mengenai pemetaan tingkat kerawanan gempa bumi di Pulau Jawa menggunakan algoritma K-Means Clustering, diperoleh pengelompokan data gempa menjadi tiga *cluster*. Pengelompokan ini dilakukan berdasarkan atribut lokasi (lat dan lon), kedalaman, dan magnitudo. Hasil akhir menunjukkan bahwa *Cluster 0* berisi 2.348 titik dan termasuk kategori kerawanan tinggi karena memiliki frekuensi kejadian terbesar meskipun kedalaman dan magnitudonya relatif rendah. *Cluster 2* berjumlah 276 titik dan dikategorikan sebagai kerawanan sedang, sedangkan *Cluster 1* berisi 12 titik dengan kedalaman dan magnitudo paling tinggi namun frekuensi kejadian rendah sehingga masuk kategori kerawanan rendah. Pemetaan ini

menunjukkan bahwa wilayah selatan Pulau Jawa merupakan daerah dengan aktivitas seismik yang paling intens dan perlu mendapatkan perhatian lebih dalam upaya mitigasi.

Penerapan algoritma K-Means menggunakan *Microsoft Excel* dan RapidMiner menghasilkan *output* yang berbeda, baik pada jumlah anggota *cluster* maupun penempatan label *clusternya*. Hal ini disebabkan oleh perbedaan proses kerja kedua platform. Pada *Microsoft Excel*, pemilihan centroid awal dilakukan secara manual atau acak berdasarkan data uji, serta terdapat proses iterasi awal yang dihitung secara bertahap. Sementara itu, RapidMiner telah menyediakan nilai centroid secara otomatis melalui centroid table dan langsung menampilkan hasil akhir tanpa menampilkan jumlah iterasi secara rinci. Selain itu, penomoran cluster pada RapidMiner bersifat otomatis (cluster 0, cluster 1, cluster 2), sehingga tidak selalu sama dengan penomoran *cluster* pada *Excel*. Perbedaan metode inisialisasi dan mekanisme iterasi inilah yang menyebabkan perbedaan jumlah anggota *cluster* serta posisi *cluster* pada kedua alat tersebut.

Secara keseluruhan, penelitian ini memberikan gambaran kerawanan gempa yang lebih terarah di Pulau Jawa dan dapat digunakan sebagai bahan pertimbangan bagi BPBD dan BNPB dalam penyusunan strategi mitigasi bencana dan sistem peringatan dini. Untuk penelitian selanjutnya, disarankan agar cakupan dataset diperluas serta mempertimbangkan variabel tambahan seperti intensitas getaran (MMI) atau mekanisme patahan agar hasil analisis dapat memberikan informasi yang lebih komprehensif bagi pemerintah maupun peneliti lain.

REFERENSI

- Agustina, M., & Mulyawan. (2023). *IMPLEMENTASI ALGORITMA K-MEANS PADA PERISTIWA GEMPA BUMI DI WILAYAH JAWA BARAT*. 2(2), 257–264.
- Aria, R. R. (2024). *KLASTERISASI ANGKATAN KERJA DI INDONESIA BERDASARKAN USIA MENGGUNAKAN METODE ALGORITMA K-MEANS*. *Inti Nusa Mandiri*, 18, 157–165. <https://doi.org/https://doi.org/10.33480/inti.v18i2.5056> VOL.
- Baldah, A., Duarisah, A. V., & Maulana, R. A. (2023). *Clustering Daerah Rawan Bencana Alam Di Indonesia Berdasarkan Provinsi Dengan Metode K-Means*. 14, 31–36.
- Cahyo, F. D., Ihsan, F., Roulita, R., Wijayanti, N., & Mirwanti, R. (2023). Kesiapsiagaan Bencana Gempa Bumi Dalam Keperawatan: Tinjauan Penelitian. *JPP (Jurnal Kesehatan Poltekkes Palembang)*, 18(1), 87–94. <https://doi.org/10.36086/jpp.v18i1.1525>
- Dwi Mauluda, A., Kuncoro, A. K., & Aningsih, S. (2025). Analisis Pengelompokan Gempa Bumi Di Sumatra Indonesia Menggunakan Metode K-Means, Dbscan, Agglomerative Clustering, Dan Mean Shift. *JoP*, 10(3), 75–86.
- Febrian, R., Suyatno, R., Salsabila, S., Budiman, F., & Sudaryanto, S. (2025). ANALISIS SPASIAL BERBASIS K-MEANS RAWAN GEMPA BERDASARKAN DATA HISTORIS. *RITECS (Integrative Research in Computer Science)*, 1(1), 108–123.
- Halik, M. F. Al, & Septiana, L. (2022). *Analisa Data Untuk Prediksi Daerah Rawan Bencana Alam Di Jawa Barat Menggunakan Algoritma K-Means Clustering*. 6(4), 856–870. <https://doi.org/10.52362/jisamar.v6i4.939>
- Karmanita, D., & Hendrik, B. (2023). *Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Peminatan Mata Kuliah*. 1(6).
- Lestari, P. D., & Mulyawan. (2023). *DATAMINING PADA PENJUALAN AIR BERSIH DI SPAM AKIDAH MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING*. 7(1), 412–416.
- Maulana Kharyska Abadi, M., Anjani Arifiyanti, A., & Satria Yudha Kartika, D. (2025). Klasterisasi Kerawanan Gempa Bumi Di Indonesia Menggunakan Algoritma Invasive Weed Optimization. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(4), 7158–7165. <https://doi.org/10.36040/jati.v9i4.14460>
- Rahmi, P. A., Muliawati, T., & Harbowo, D. G. (2024). Analisis Pola Aktivitas Gempa Bumi di Pulau Sumatera dengan Metode K-Means Clustering dan Rantai Markov. *Prosiding Seminar Nasional Sains Data*, 4(1), 592–601. <https://doi.org/10.33005/senada.v4i1.291>
- Saputra, E. A., & Nataliani, Y. (2021). *Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means*. 3(3), 424–439.
- Suara, M., Jati, B. L., Gufron, A., Hilmani, H., Helmalia, Fitri, N. L., & Andhia. (2023). *PENANGANAN DAMPAK H + 14 PASCA BENCANA GEMPA BUMI DENGAN MASALAH KESEHATAN DI DESA MANGUN KERTA KECAMATAN CUGENANG KABUPATEN CIANJUR*. 6(April), 1396–1411.