



ANALISIS SPASIAL BERBASIS K-MEANS CLUSTERING UNTUK IDENTIFIKASI DAERAH RAWAN GEMPA BERDASARKAN DATA HISTORIS

**Randy Febrian, Revalina Suyatno, Shafa Salsabila,
Fikri Budiman*, Sudaryanto Sudaryanto**

Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Indonesia

*Penulis Korespondensi: fikri.budiman@dsn.dinus.ac.id

Abstrak: Indonesia merupakan salah satu negara dengan tingkat risiko bencana alam yang tinggi karena letak geografisnya berada pada pertemuan tiga lempeng tektonik besar, serta dipengaruhi oleh kondisi topografi dan iklim tropis yang beragam. Bencana seperti gempa bumi, tsunami, banjir, dan tanah longsor sering terjadi, sehingga menimbulkan kerugian signifikan baik secara ekonomi maupun infrastruktur. Penelitian ini berfokus pada analisis pola historis kejadian gempa bumi di Indonesia dengan menggunakan algoritma K-Means Clustering untuk mengidentifikasi daerah rawan gempa berdasarkan tingkat intensitas, yaitu tinggi, sedang, dan rendah. Data yang digunakan diperoleh dari platform Kaggle dengan sumber utama dari BMKG. Metode K-Means diterapkan untuk mengelompokkan wilayah berdasarkan kesamaan karakteristik kejadian gempa. Hasil penelitian menunjukkan terbentuknya tiga kluster wilayah dengan kategori intensitas gempa: tinggi (0), sedang (1), dan rendah (2). Temuan ini diharapkan dapat menjadi referensi bagi pemerintah dalam merancang strategi mitigasi bencana yang lebih terarah dan efisien, sehingga mampu meminimalkan dampak gempa serta meningkatkan kesiapsiagaan masyarakat.

Kata Kunci: Analisis Spasial, K-Means Clustering, Data Historis, Identifikasi Daerah Rawan Gempa, Mitigasi Bencana.

1. PENDAHULUAN

Indonesia merupakan salah satu negara dengan tingkat risiko bencana alam tertinggi di dunia. Letaknya yang berada di pertemuan tiga lempeng tektonik utama Lempeng Eurasia, Indo Australia, dan Pasifik menjadikan wilayah ini sangat rawan terhadap gempa bumi dan tsunami [1-3]. Aktivitas seismik yang tinggi menyebabkan Indonesia sering mengalami gempa dengan berbagai magnitudo, baik di darat maupun di laut [4]. Di samping itu, kondisi geografis yang kompleks dan iklim tropis yang lembab turut meningkatkan potensi terjadinya bencana hidrometeorologi seperti banjir, tanah longsor, dan badai tropis [5]. Dampak dari bencana-bencana tersebut tidak hanya mengakibatkan kerugian ekonomi yang besar, tetapi juga mengancam

Cara Mengutip:

Febrian, R., Suyatno, R., Salsabila, S., Budiman, F., & Sudaryanto, S. (2025). Analisis spasial berbasis k-means clustering untuk identifikasi daerah rawan gempa berdasarkan data. IRCS: Integrative Research in Computer Science, 1(1), 108-123.

keselamatan jiwa, merusak infrastruktur, dan menimbulkan ketidakstabilan sosial. Oleh karena itu, pemetaan wilayah rawan bencana berbasis data menjadi kebutuhan mendesak dalam perencanaan mitigasi yang lebih terstruktur dan tepat sasaran [6].

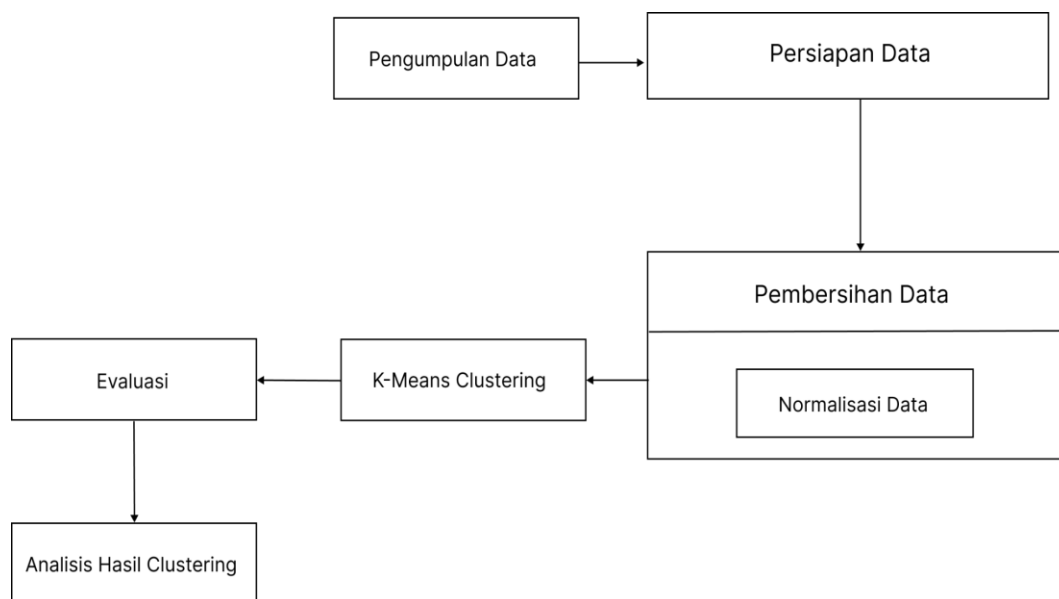
Perkembangan teknologi informasi telah membuka peluang besar dalam pengelolaan risiko bencana melalui pendekatan berbasis data. Salah satu metode yang umum digunakan dalam analisis spasial adalah K-Means Clustering, sebuah teknik pengelompokan data berdasarkan kemiripan karakteristik [7-8]. Metode ini memungkinkan klasifikasi wilayah ke dalam kategori rawan tinggi, sedang, dan rendah berdasarkan pola kejadian gempa historis. Clustering sendiri merupakan proses pengelompokan yang bertujuan membentuk himpunan data dengan kesamaan internal yang tinggi dibandingkan kelompok lainnya [9-11]. Meskipun K-Means banyak digunakan karena kesederhanaannya dan efisiensi komputasinya, metode ini memiliki beberapa keterbatasan. Hasil pengelompokan sangat dipengaruhi oleh pemilihan titik pusat awal (initial centroid) [12-13] dan hanya optimal jika kluster memiliki bentuk dan ukuran yang seragam. Selain itu, K-Means tidak secara otomatis menentukan jumlah cluster yang ideal, sehingga diperlukan pendekatan tambahan seperti metode Elbow atau Silhouette Score untuk validasi [14-15]. Sensitivitas terhadap outlier juga menjadi tantangan tersendiri yang dapat mempengaruhi hasil clustering.

Sejumlah penelitian sebelumnya telah membuktikan efektivitas metode K-Means dalam konteks spasial. [7] mengelompokkan wilayah rawan bencana di Indonesia berdasarkan data provinsi dan menunjukkan bahwa metode ini mampu mengungkap pola kerawanan yang signifikan [16-17]. Sementara itu, [18] menggunakan K-Means untuk memetakan kerentanan bencana di wilayah Jawa Barat dan membuktikan efektivitasnya dalam klasifikasi wilayah berdasarkan karakteristik historis. Lebih lanjut, [19] juga menerapkan K-Means untuk menganalisis distribusi gempa, namun masih terbatas pada analisis deskriptif tanpa adanya metrik validasi kuantitatif seperti Silhouette Score atau Davies-Bouldin Index. Berdasarkan tinjauan pustaka yang telah dilakukan, terdapat beberapa kesenjangan dalam penelitian sebelumnya. Sebagian besar studi sebelumnya masih menggunakan pendekatan administratif dalam memetakan risiko gempa, sehingga kurang presisi dalam menggambarkan sebaran spasial titik kejadian. Selain itu, pendekatan yang digunakan cenderung bersifat supervised, sedangkan penelitian ini mengusulkan pendekatan unsupervised dengan K-Means Clustering untuk segmentasi wilayah [20-21]. Penelitian sebelumnya juga jarang melakukan evaluasi kuantitatif terhadap kualitas cluster yang dihasilkan. Oleh karena itu, penelitian ini hadir untuk mengisi kesenjangan tersebut dengan menggabungkan data historis gempa berbasis spasial, metode clustering yang terukur, dan visualisasi spasial sebagai alat analisis risiko bencana yang lebih tepat.

Berdasarkan kondisi tersebut, penelitian ini berupaya mengeksplorasi bagaimana penerapan algoritma K-Means Clustering terhadap data historis kejadian gempa bumi di Indonesia dapat digunakan untuk mengidentifikasi wilayah-wilayah dengan tingkat kerawanan yang berbeda secara spasial. Dengan menganalisis data gempa dari tahun 2015 hingga 2025 yang diperoleh dari platform Kaggle dan bersumber dari BMKG, diharapkan hasil pengelompokan dapat memberikan gambaran objektif mengenai zona resiko gempa di Indonesia. Informasi ini penting untuk mendukung pengambilan keputusan dalam penyusunan strategi mitigasi bencana yang lebih efektif dan adaptif terhadap dinamika risiko bencana yang terus berkembang.

2. METODOLOGI PENELITIAN

Penelitian ini dilakukan menggunakan lima tahap: pengumpulan data, persiapan data, pembersihan data, pemrograman K-Means Clustering, dan evaluasi. Gambar 1 memperlihatkan diagram tahapan penelitian. Proses metodologi ini dirancang secara sistematis agar mampu menghasilkan hasil analisis yang terukur dan dapat dipertanggungjawabkan secara ilmiah. Setiap tahapan saling berkaitan, sehingga kualitas hasil akhir sangat bergantung pada ketelitian di setiap tahap sebelumnya. Oleh karena itu, metodologi ini disusun dengan memperhatikan prinsip transparansi, replikasi, dan validitas data.



Gambar 1. Diagram Tahapan Penelitian

Penelitian ini juga menekankan pada penggunaan pendekatan kuantitatif berbasis data spasial, sehingga hasil yang diperoleh dapat digunakan dalam pengambilan keputusan berbasis bukti (evidence-based decision making). Dengan demikian, metodologi ini tidak hanya bermanfaat bagi

pengembangan akademik, tetapi juga memiliki relevansi praktis dalam perencanaan mitigasi bencana gempa bumi di Indonesia.

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari platform Kaggle, yaitu dataset berisi informasi historis mengenai kejadian bencana gempa di Indonesia. Dataset mencakup rentang waktu dari tahun 2008 hingga 2025. Namun, untuk menjaga relevansi analisis dengan kondisi terkini, hanya data pada periode 2015 hingga 2025 yang digunakan dalam penelitian ini. Pemilihan periode ini bertujuan agar pola kejadian bencana yang dianalisis tetap aktual dan mampu merepresentasikan dinamika risiko bencana saat ini. Dataset berisikan data bencana gempa dari sumber Kaggle yang diambil dari repositori Gempa Bumi yang dikelola oleh BMKG (lembaga pemerintah non-departemen Indonesia). Dataset yang digunakan adalah data gempa dari 1 November 2008 sampai 30 Maret 2025, yang mencakup 92.887 entri. Namun penelitian ini hanya berfokus pada data gempa dari tahun 2020 - Maret 2025. Gambar 2 menunjukkan contoh lima baris teratas data yang diperoleh sebelum diproses lebih lanjut.

	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

Gambar 2. Dataset awal

Selain itu, dataset yang digunakan mencakup atribut-atribut penting seperti waktu kejadian gempa, lokasi geografis (latitude dan longitude), kedalaman (depth), magnitudo (mag), dan keterangan daerah (remark). Atribut-atribut ini dipilih karena dianggap paling relevan dalam menentukan pola spasial dan intensitas kejadian gempa. Proses pengumpulan data dilakukan dengan memastikan keabsahan sumber, sehingga data yang diperoleh dapat dipertanggungjawabkan. Dengan menggunakan dataset dari BMKG melalui Kaggle, penelitian ini memiliki dasar data yang kuat dan kredibel. Pemanfaatan data sekunder dari repositori publik juga memungkinkan penelitian ini direplikasi oleh peneliti lain, sehingga meningkatkan aspek keterbukaan ilmiah.

2.2 Persiapan Data

Tahap persiapan data dilakukan dengan seleksi atribut dan penyesuaian rentang waktu. Seleksi atribut bertujuan untuk memilih kolom-kolom data yang relevan. Sementara itu, penyesuaian rentang waktu dilakukan dengan menyaring data sehingga hanya mencakup kejadian

dari tahun 2015 hingga 2025. Data yang berada di luar rentang ini dikeluarkan dari dataset, sehingga analisis lebih fokus dan tidak terpengaruh oleh pola historis yang sudah terlalu jauh dari kondisi terkini. Proses persiapan data ini penting untuk memastikan bahwa hasil clustering benar-benar mencerminkan karakteristik kejadian bencana dalam kurun waktu yang masih relevan untuk kebutuhan mitigasi. Persiapan data dilakukan untuk memastikan data yang digunakan relevan dan bersih. Langkah pertama adalah melakukan filterisasi berdasarkan tahun, yaitu hanya memilih data kejadian gempa bumi dalam rentang tahun tertentu yang sesuai dengan kebutuhan penelitian. Data yang berada di luar periode tersebut diabaikan untuk menjaga konsistensi dan relevansi analisis terhadap kondisi terkini. Gambar 3 memperlihatkan contoh hasil filterisasi pada data.

	tgl	ot	lat	lon	depth	mag	remark	strike1	dip1	rake1	strike2	dip2	rake2
61458	2020-01-01	23:50:16.765	-6.54	132.81	10	4.4	Tanimbar Islands Reg. - Indonesia	NaN	NaN	NaN	NaN	NaN	NaN
61459	2020-01-01	22:41:30.691	-6.27	104.13	17	3.1	Sunda Strait - Indonesia	NaN	NaN	NaN	NaN	NaN	NaN
61460	2020-01-01	22:18:50.864	-2.13	127.93	72	3.9	Ceram Sea	NaN	NaN	NaN	NaN	NaN	NaN
61461	2020-01-01	19:50:47.534	1.48	126.42	10	3.5	Northern Molucca Sea	NaN	NaN	NaN	NaN	NaN	NaN
61462	2020-01-01	19:07:31.971	-9.75	112.98	10	3.9	South of Java - Indonesia	NaN	NaN	NaN	NaN	NaN	NaN

Gambar 3. Dataset setelah filterisasi tahun

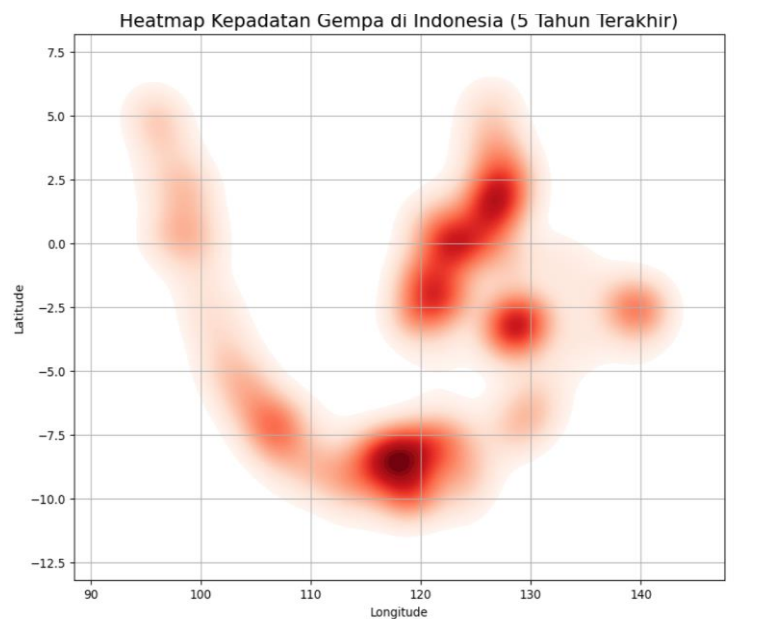
Setelah proses filterisasi berdasarkan tahun, data juga diperiksa kembali untuk mengidentifikasi kemungkinan adanya anomali seperti duplikasi entri atau pencatatan yang tidak wajar. Tahap ini diikuti dengan pemilihan atribut utama, yaitu tanggal, koordinat geografis, kedalaman, dan magnitudo. Dengan melakukan reduksi atribut, dataset menjadi lebih ringkas dan efisien untuk diproses oleh algoritma K-Means. Selain itu, dilakukan pula pengecekan konsistensi satuan pada setiap atribut, misalnya memastikan kedalaman menggunakan meter dan magnitudo sesuai standar skala Richter. Keseluruhan langkah ini bertujuan untuk memastikan bahwa data yang masuk ke tahap analisis benar-benar berkualitas tinggi.

Setelah itu, dilakukan filterisasi kolom dengan menghapus kolom-kolom yang tidak diperlukan, seperti strike1, dip1, rake1, strike2, dip2, dan rake2. Kolom-kolom ini dihilangkan karena sebagian besar isinya kosong (NaN) dan tidak berkontribusi langsung terhadap analisis spasial dan clustering yang akan dilakukan. Dengan langkah ini, dataset yang dihasilkan menjadi lebih ringkas, hanya berisi atribut-atribut utama seperti tanggal kejadian (tgl), lokasi (lat, lon), kedalaman (depth), magnitudo (mag), dan keterangan daerah (remark). Persiapan data ini penting dilakukan untuk meningkatkan kualitas analisis pada tahap selanjutnya. Gambar 4 memperlihatkan contoh hasil menghapus kolom dalam proses filterisasi.

	tgl	ot	lat	lon	depth	mag	remark
61458	2020-01-01	23:50:16.765	-6.54	132.81	10	4.4	Tanimbar Islands Reg. - Indonesia
61459	2020-01-01	22:41:30.691	-6.27	104.13	17	3.1	Sunda Strait - Indonesia
61460	2020-01-01	22:18:50.864	-2.13	127.93	72	3.9	Ceram Sea
61461	2020-01-01	19:50:47.534	1.48	126.42	10	3.5	Northern Molucca Sea
61462	2020-01-01	19:07:31.971	-9.75	112.98	10	3.9	South of Java - Indonesia

Gambar 4. Dataset setelah filterisasi kolom kosong

Heatmap yang ditampilkan pada Gambar 5 menggambarkan kepadatan kejadian gempa bumi di Indonesia selama lima tahun terakhir. Pada visualisasi ini, daerah dengan tingkat kejadian gempa yang lebih tinggi ditunjukkan dengan warna merah yang lebih pekat, sedangkan area dengan intensitas gempa yang lebih rendah ditampilkan dengan warna yang lebih terang. Sumbu horizontal dan vertikal merepresentasikan posisi geografis berdasarkan longitude dan latitude. Heatmap ini memperjelas area-area yang menjadi pusat aktivitas seismik terbesar di Indonesia, sehingga memudahkan dalam mengidentifikasi zona rawan gempa secara visual. Dengan melihat konsentrasi warna, dapat disimpulkan bahwa beberapa daerah seperti Sulawesi, selatan Jawa, dan kawasan timur Indonesia mengalami kepadatan gempa yang lebih tinggi dibandingkan daerah lainnya.



Gambar 5. Heatmap kepadatan gempa

2.3 Pembersihan Data

Tahap pembersihan data bertujuan untuk meningkatkan kualitas dataset yang akan dianalisis. Proses ini dilakukan dengan normalisasi data, yaitu mengubah skala nilai atribut menjadi

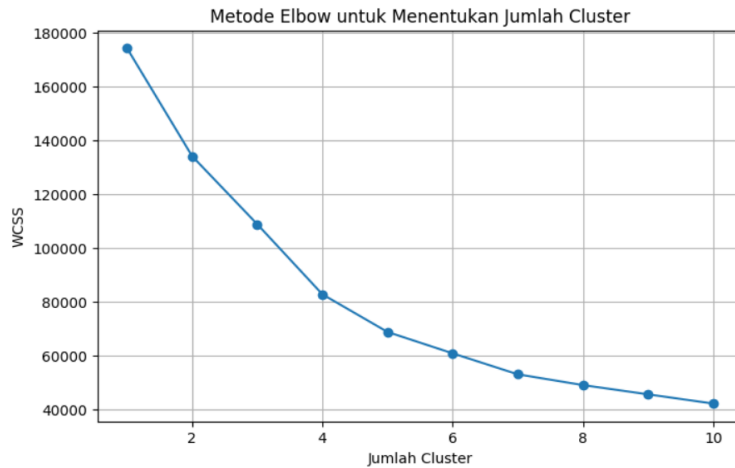
seragam. Normalisasi penting untuk mencegah atribut tertentu mendominasi hasil pengelompokan saat menggunakan algoritma K-Means. Dengan normalisasi, setiap variabel memiliki pengaruh yang seimbang dalam proses clustering, sehingga hasil analisis menjadi lebih akurat dan konsisten.

Pada tahap pembersihan data, dilakukan proses pengecekan terhadap keberadaan data yang hilang (missing value) pada setiap kolom dalam dataset. Berdasarkan hasil pemeriksaan, seluruh atribut utama seperti tgl, ot, lat, lon, depth, mag, dan remark menunjukkan nilai nol pada jumlah data hilang, yang berarti tidak terdapat nilai kosong pada kolom-kolom tersebut. Kondisi ini mengindikasikan bahwa dataset telah bersih dan tidak memerlukan proses penanganan data hilang, seperti imputasi atau penghapusan data. Dengan data yang sepenuhnya lengkap, analisis lebih lanjut dapat dilakukan secara optimal tanpa resiko bias atau kesalahan akibat ketidaklengkapan data, sehingga meningkatkan validitas dan reliabilitas hasil penelitian.

2.4 K-Means Clustering

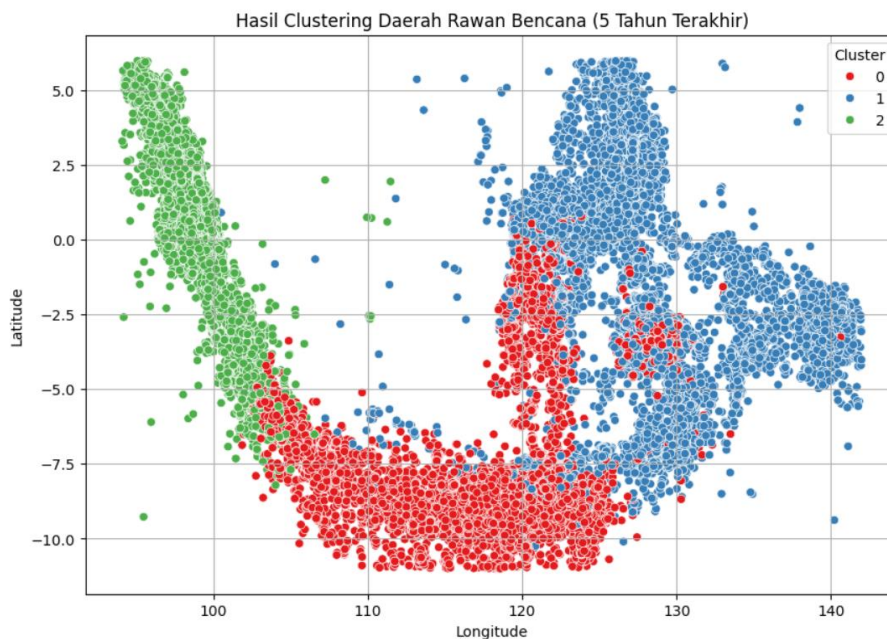
K-Means Clustering adalah metode yang digunakan untuk mengelompokkan data ke dalam beberapa grup atau klaster. Proses pengelompokkan ini memanfaatkan kemiripan data yang masuk ke dalam beberapa klaster. Sederhananya, data yang memiliki karakteristik yang mirip akan masuk ke dalam klaster yang sama, sedangkan karakteristik data lain akan masuk ke klaster yang berbeda. Ada delapan tahap yang dilalui dalam proses K-Means. Tahap ini dimulai dengan menentukan jumlah cluster optimal dengan menggunakan metode Elbow. Metode ini memplot nilai inertia (jumlah kuadrat jarak antara data dan centroid dalam cluster) terhadap berbagai jumlah cluster untuk menentukan jumlah cluster yang optimal. Setelah ditemukan jumlah cluster optimal, model K-Means diterapkan pada data dengan jumlah cluster yang sesuai. Hasil clustering kemudian ditambahkan ke dalam dataset sebagai kolom baru yang menunjukkan label cluster untuk masing-masing daerah. Langkah ini memungkinkan analisis lebih lanjut untuk memahami karakteristik dan pola dari setiap kelompok yang dihasilkan.

Pada tahap penentuan jumlah cluster optimal, digunakan metode Elbow untuk mengetahui titik terbaik dalam pembentukan kelompok data. Metode ini menghitung nilai Within-Cluster Sum of Square (WCSS) untuk berbagai jumlah cluster. Grafik Elbow (Gambar 6) menunjukkan bahwa jumlah cluster optimal tercapai saat penurunan nilai WCSS mulai melambat secara signifikan, membentuk sudut siku-siku pada jumlah cluster tiga. Penurunan WCSS sebelum titik ini menunjukkan pembagian data yang semakin efisien, sedangkan setelah tiga cluster, penurunannya cenderung melandai. Oleh karena itu, tiga cluster dipilih sebagai konfigurasi terbaik untuk mendapatkan hasil clustering yang efektif dan efisien.



Gambar 6. Jumlah cluster berdsarkan metode Elbow

Hasil clustering menggunakan metode K-Means, di mana data dikelompokkan ke dalam tiga cluster berdasarkan lokasi kejadian gempa bumi berdasarkan nilai latitude dan longitude. Setiap titik pada grafik merepresentasikan suatu kejadian gempa, dengan warna berbeda untuk setiap cluster. Cluster 0 (Merah) mencakup daerah dengan tingkat kejadian gempa bumi yang paling tinggi. Titik-titik merah tersebar cukup padat di area yang menunjukkan aktivitas seismik yang intens. Daerah dalam cluster ini memerlukan perhatian prioritas dalam upaya mitigasi bencana, karena potensi dampak kerusakan yang besar akibat frekuensi kejadian gempa yang tinggi. Gambar 7 menunjukkan diagram scatter plot hasil clustering.



Gambar 7. Scatter plot K-Means Clustering

Cluster 1 (Biru) terdiri dari daerah dengan tingkat kejadian gempa bumi yang sedang. Kejadian gempa di area ini masih cukup sering terjadi, namun intensitas dan frekuensinya tidak setinggi cluster 0. Daerah dalam cluster ini perlu mendapatkan langkah mitigasi sistematis untuk mengantisipasi potensi peningkatan risiko ke depan. Sedangkan Cluster 2 (Hijau) mencakup daerah dengan tingkat kejadian gempa bumi yang rendah. Sebagian besar titik-titik hijau tersebar lebih jarang, menunjukkan bahwa daerah dalam cluster ini relatif lebih aman dari aktivitas gempa bumi dalam lima tahun terakhir. Meskipun demikian, pemantauan berkala tetap diperlukan untuk memastikan kondisi tetap stabil dan mengantisipasi perubahan risiko.

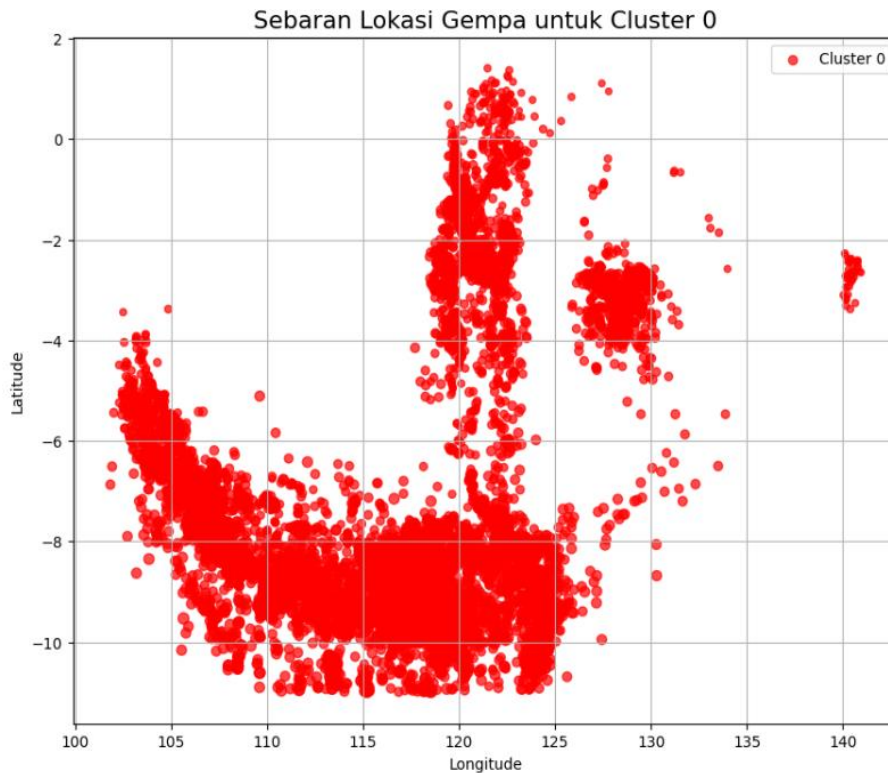
2.5 Evaluasi Hasil Clustering

Hasil clustering dievaluasi untuk memastikan pembagian cluster yang baik. Evaluasi dilakukan menggunakan Silhouette Score untuk mengukur kualitas cluster, di mana nilai yang tinggi menunjukkan pemisahan cluster yang baik. Selain itu, digunakan juga Davies-Bouldin Index untuk menilai validitas cluster, dengan nilai yang lebih rendah menunjukkan hasil clustering yang lebih optimal. Hasil clustering divisualisasikan menggunakan scatter plot untuk menunjukkan distribusi data berdasarkan cluster. Selanjutnya, dilakukan analisis statistik deskriptif untuk setiap cluster, termasuk perhitungan rata-rata, rentang nilai, dan jumlah data per cluster, guna memahami karakteristik masing-masing kelompok.

Evaluasi hasil clustering bertujuan untuk menilai sejauh mana kualitas pengelompokan data yang telah dilakukan menggunakan metode K-Means Clustering [19-20]. Dalam penelitian ini, evaluasi dilakukan menggunakan dua metrik utama: silhouette score dan Davies-Bouldin index. Silhouette Score mengukur seberapa mirip sebuah data dengan cluster-nya sendiri dibandingkan dengan cluster lain. Nilainya berada di rentang -1 hingga 1: nilai mendekati 1 menunjukkan bahwa data berada di cluster yang benar dan sangat sesuai, nilai mendekati 0 menunjukkan bahwa data berada di batas antara dua cluster, dan nilai negatif menunjukkan bahwa data mungkin salah pengelompokan. Berdasarkan hasil evaluasi, diperoleh Silhouette Score sebesar 0.2871, yang berarti bahwa hasil clustering tergolong cukup, meskipun masih ada kemungkinan perbaikan (umumnya nilai > 0.5 dianggap clustering sangat baik). Davies-Bouldin Index (DBI) mengukur rata-rata kesamaan setiap cluster dengan cluster yang paling mirip. Dalam metrik ini, semakin kecil nilainya, semakin baik hasil clustering, di mana nilai DBI rendah menunjukkan bahwa cluster yang terbentuk lebih kompak dan lebih terpisah satu sama lain. Pada evaluasi ini, diperoleh Davies-Bouldin Index sebesar 1.2852. Nilai ini menunjukkan bahwa meskipun cluster cukup terbentuk, masih ada ruang untuk meningkatkan separasi antar cluster agar lebih optimal.

Gambar 8 memperlihatkan sebaran lokasi gempa bumi yang tergolong dalam Cluster 0 berdasarkan hasil K-Means Clustering. Titik-titik merah menunjukkan bahwa gempa banyak terjadi di daerah

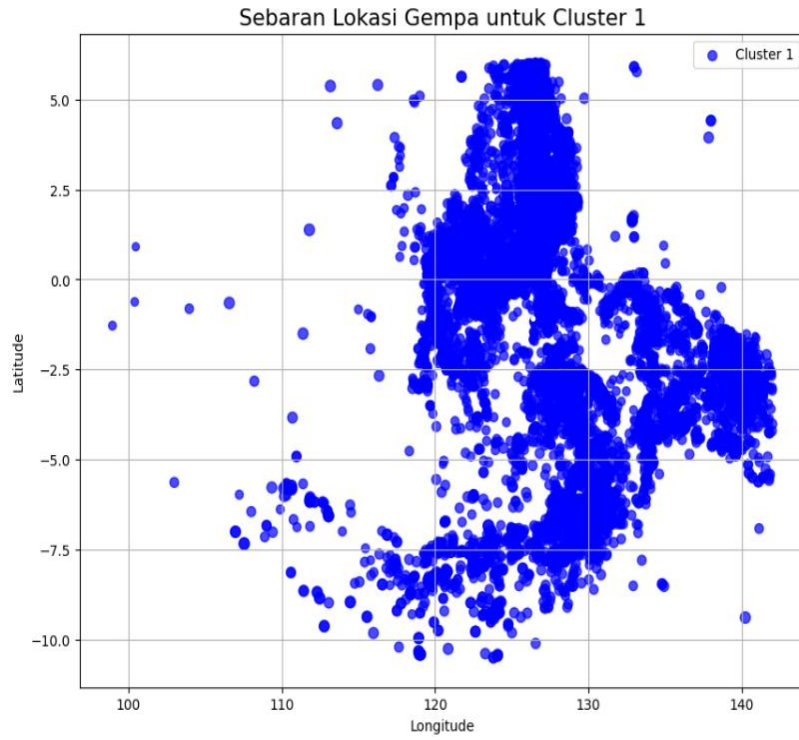
selatan Indonesia, terutama di sepanjang pantai selatan Jawa, Bali, Nusa Tenggara Barat, dan Nusa Tenggara Timur. Sebaran yang padat ini mengindikasikan bahwa Cluster 0 mencakup daerah dengan tingkat aktivitas seismik tinggi. Oleh karena itu, daerah ini perlu menjadi prioritas dalam upaya mitigasi bencana gempa bumi.



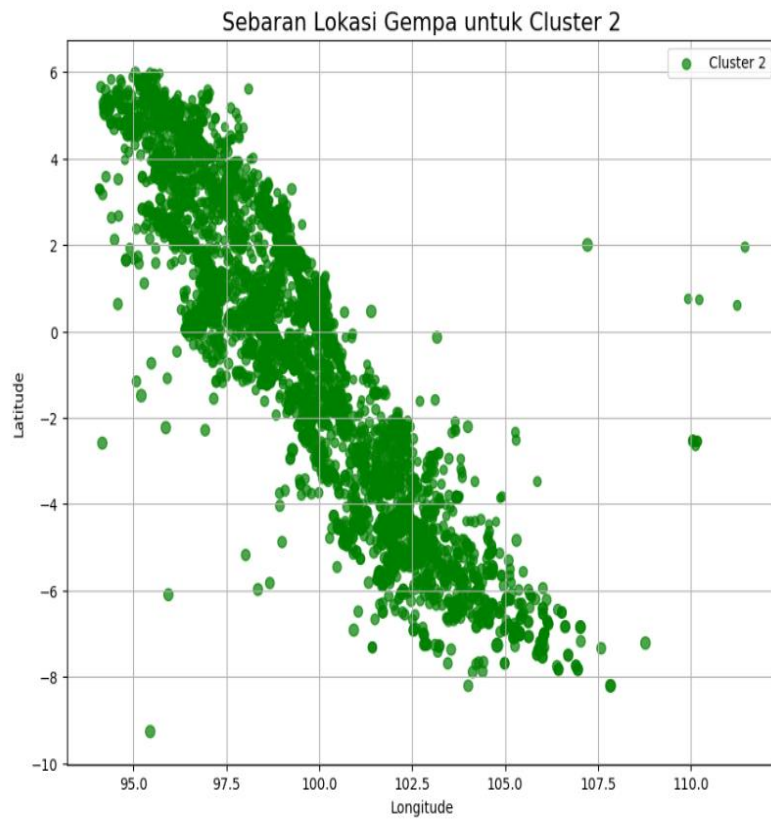
Gambar 8. Scatter Plot untuk Cluster 0

Gambar 9 menunjukkan sebaran lokasi gempa untuk Cluster 1 yang ditandai dengan titik-titik berwarna biru. Sebaran gempa di cluster ini cukup merata, terutama di daerah tengah Indonesia seperti Sulawesi, sebagian Kalimantan bagian timur, dan Maluku. Titik-titiknya tidak sepadat Cluster 0, menunjukkan bahwa daerah ini memiliki frekuensi gempa yang sedang. Oleh karena itu, Cluster 1 dikategorikan sebagai daerah dengan tingkat aktivitas gempa menengah, yang tetap membutuhkan perhatian dalam upaya mitigasi bencana.

Gambar 10 memperlihatkan sebaran lokasi gempa untuk Cluster 2 yang digambarkan dengan titik-titik hijau. Sebaran gempa di cluster ini terlihat memanjang dari barat ke timur, dominan di sepanjang Pulau Sumatera. Pola ini mengikuti jalur subduksi di sepanjang zona Megathrust Sumatera. Meskipun titik-titiknya lebih jarang dibandingkan Cluster 0 dan Cluster 1, daerah ini tetap berpotensi mengalami gempa besar, sehingga tetap perlu dipertimbangkan dalam strategi mitigasi bencana.



Gambar 9. Scatter Plot untuk Cluster 1

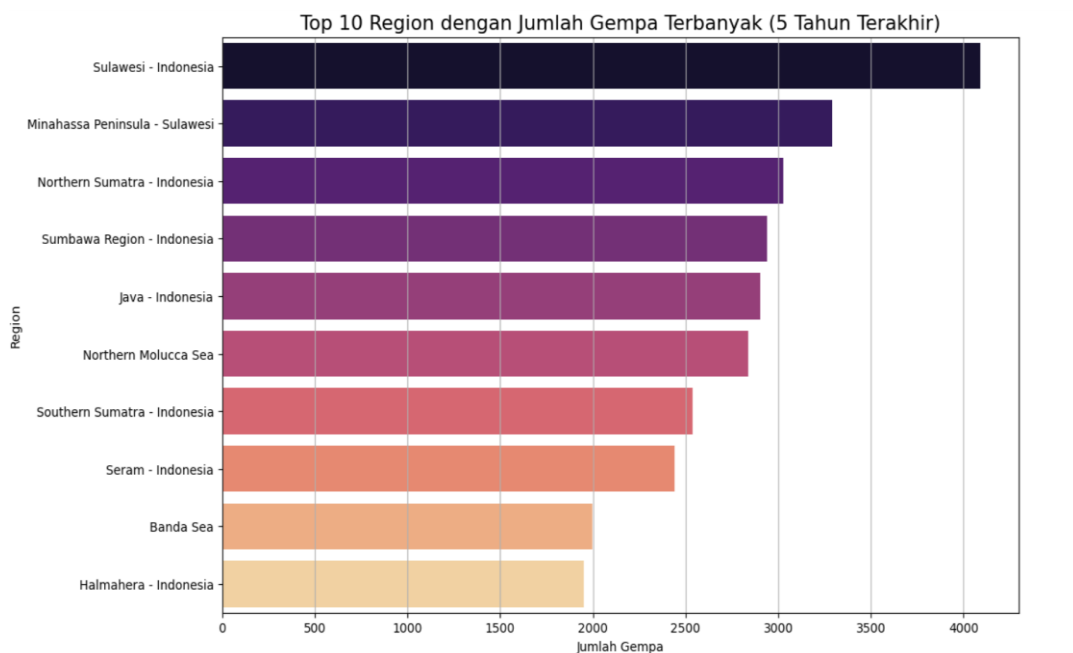


Gambar 10. Scatter Plot untuk Cluster 2

3. HASIL DAN PEMBAHASAN

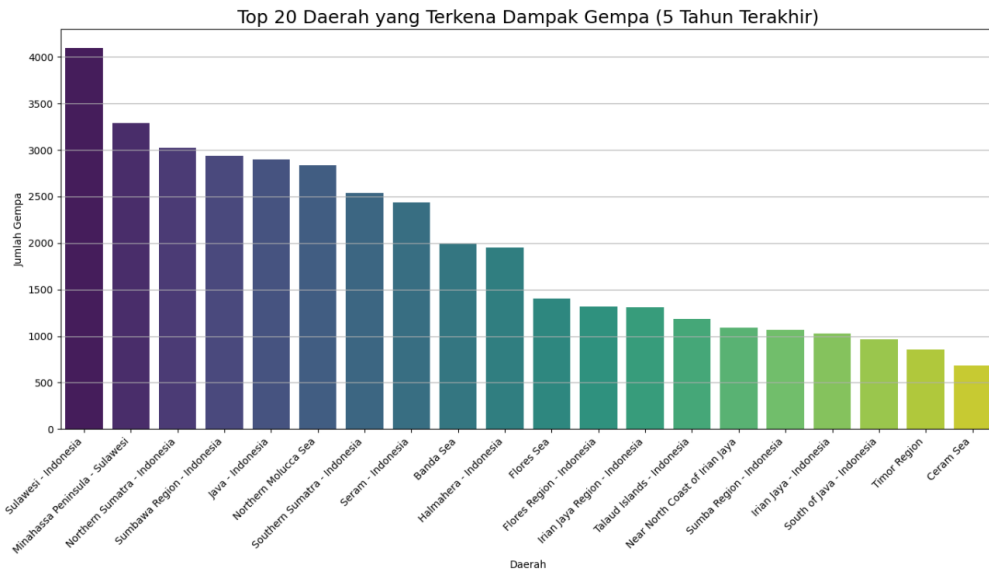
Hasil clustering dianalisis untuk memahami karakteristik setiap cluster. Analisis hasil clustering memberikan wawasan mengenai karakteristik daerah berdasarkan faktor-faktor bencana gempa. Setiap cluster yang terbentuk mencerminkan pola tertentu dari faktor-faktor yang dianalisis. Analisis statistik dilakukan untuk memahami distribusi data dalam setiap cluster, seperti jumlah daerah, rata-rata nilai faktor, serta rentang nilai pada masing-masing variabel.

Berdasarkan Gambar 11, daftar Top 10 daerah dengan jumlah gempa terbanyak dalam lima tahun terakhir menunjukkan bahwa daerah Sulawesi menjadi daerah dengan frekuensi kejadian gempa paling tinggi di Indonesia. Disusul oleh Semenanjung Minahasa (Sulawesi), Sumatera bagian utara, dan daerah Sumbawa. Jawa juga termasuk dalam daftar ini, meskipun dengan jumlah kejadian yang sedikit lebih rendah dibandingkan daerah Indonesia tengah dan timur. Data ini mengindikasikan bahwa aktivitas seismik terbesar terjadi di kawasan Sulawesi, Sumatera Utara, dan Nusa Tenggara.



Gambar 11. Bar plot top 10 daerah jumlah gempa terbanyak

Sementara itu, Gambar 12 menampilkan Top 20 daerah yang paling terdampak gempa dalam lima tahun terakhir. Daerah Sulawesi tetap menduduki posisi pertama sebagai daerah paling terdampak, diikuti oleh Semenanjung Minahasa dan Sumatera Utara. Selain itu, beberapa daerah perairan seperti Laut Maluku Utara, Laut Banda, dan Laut Flores juga tercatat cukup terdampak, menunjukkan bahwa gempa di kawasan ini tidak hanya berpengaruh di daratan, tetapi juga di daerah laut yang berdekatan dengan pusat-pusat aktivitas seismik.



Gambar 12. Bar plot top 20 daerah terkena dampak gempa

Temuan ini menegaskan bahwa daerah Indonesia bagian tengah dan barat, khususnya Sulawesi dan Sumatera, perlu mendapatkan perhatian lebih besar dalam upaya mitigasi risiko bencana gempa, baik melalui pembangunan infrastruktur tahan gempa, peningkatan sistem peringatan dini, maupun edukasi kesiapsiagaan masyarakat.

Hasil proses clustering menggunakan algoritma K-Means menghasilkan tiga kluster utama wilayah rawan gempa di Indonesia. Kluster 0 terdiri dari 31 wilayah dengan konsentrasi yang cukup tinggi di kawasan timur dan tengah Indonesia. Beberapa di antaranya meliputi Seram, Flores Region, Sulawesi, Sumbawa, Bali, Jawa, dan Sumatra bagian selatan. Selain itu, wilayah laut seperti Laut Banda, Laut Seram, Laut Savu, Laut Bali, Laut Jawa, Laut Flores, serta Laut Maluku Selatan juga termasuk dalam kluster ini. Kluster ini juga mencakup wilayah sekitar Irian Jaya, Kepulauan Aru, Kepulauan Tanimbar, dan Semenanjung Minahassa. Sebaran wilayah ini menunjukkan bahwa gempa sering terjadi di zona tektonik aktif yang dipengaruhi oleh tumbukan antar lempeng di kawasan Indonesia timur dan tengah.

Sementara itu, Kluster 1 mencakup 43 wilayah yang lebih luas dibandingkan kluster lainnya, tersebar dari kawasan barat hingga timur Indonesia, bahkan menjangkau beberapa wilayah di luar perairan Indonesia. Wilayah yang masuk kluster ini antara lain Halmahera, Talud Islands, Sulawesi, Sumba, Sumbawa, Seram, Timor, Bali, Jawa, serta Sumatra bagian utara dan selatan. Sejumlah laut di kawasan ini juga termasuk, seperti Laut Banda, Laut Seram, Laut Celebes, Laut Savu, Laut Arafura, Laut Flores, Laut Jawa, dan Laut Cina Selatan. Selain itu, terdapat pula wilayah di luar Indonesia seperti Papua Nugini, Filipina (Mindanao dan Philippine Islands Region), serta Kepulauan Caroline di Mikronesia. Jumlah wilayah yang lebih banyak menunjukkan bahwa kluster

ini merepresentasikan zona rawan gempa dengan intensitas menengah yang tersebar luas di kawasan Asia Tenggara dan sekitarnya.

Adapun Klaster 2 terdiri dari 9 wilayah dengan sebaran lebih terbatas dibandingkan dua klaster sebelumnya. Wilayah yang termasuk dalam klaster ini adalah Selat Sunda, Sumatra bagian utara dan selatan, serta Jawa. Selain itu, terdapat pula wilayah lepas pantai seperti Off West Coast of North Sumatra, Southwest of Sumatra, Laut Cina Selatan, Borneo, dan Samudra Hindia bagian selatan. Sebaran yang relatif sedikit mengindikasikan bahwa klaster ini merepresentasikan wilayah dengan intensitas gempa rendah, meskipun tetap berada pada zona tektonik aktif yang rawan terhadap aktivitas seismik.

4. KESIMPULAN

Penelitian ini berhasil mengelompokkan daerah di Indonesia berdasarkan tingkat intensitas kejadian gempa menggunakan algoritma K-Means Clustering. Hasil analisis menghasilkan tiga klaster utama, yaitu klaster dengan risiko tinggi, risiko sedang, dan risiko rendah. Daerah-daerah seperti Sulawesi, Semenanjung Minahasa, dan Sumatera Utara teridentifikasi sebagai wilayah dengan intensitas gempa tinggi, sehingga memerlukan perhatian lebih besar dalam perencanaan mitigasi. Sementara itu, wilayah dengan intensitas sedang tetap menunjukkan frekuensi gempa yang signifikan meskipun tidak sepadat klaster risiko tinggi, sedangkan wilayah pada klaster rendah relatif jarang mengalami gempa, namun tidak berarti sepenuhnya bebas risiko.

Evaluasi hasil clustering menggunakan Silhouette Score dan Davies-Bouldin Index menunjukkan bahwa kualitas pengelompokan tergolong cukup baik dan dapat digunakan sebagai acuan awal dalam penyusunan strategi mitigasi bencana gempa. Meskipun nilai Silhouette Score belum mencapai kategori optimal, hasil ini tetap memberikan gambaran pola spasial kejadian gempa yang dapat dipertanggungjawabkan secara ilmiah. Dengan demikian, penelitian ini dapat menjadi dasar penting bagi pengembangan kajian lebih lanjut yang menggunakan metode analisis spasial maupun algoritma clustering yang lebih kompleks untuk memperoleh hasil yang lebih akurat.

Berdasarkan temuan ini, diperlukan penguatan upaya mitigasi di daerah dengan klaster risiko tinggi, misalnya melalui pembangunan dan perbaikan infrastruktur tahan gempa, penerapan standar konstruksi yang ketat, serta pengembangan sistem peringatan dini yang lebih andal. Upaya ini tidak hanya bertujuan untuk mengurangi potensi kerusakan fisik, tetapi juga untuk melindungi keselamatan jiwa masyarakat yang tinggal di daerah rawan. Di sisi lain, wilayah dalam klaster risiko sedang perlu dipantau secara konsisten karena meskipun intensitas gempa relatif lebih rendah, potensi bencana besar tetap dapat terjadi sewaktu-waktu.

Selain itu, pemerintah bersama lembaga terkait perlu memperbaharui data historis gempa secara berkala agar analisis berbasis data dapat mencerminkan kondisi terbaru. Pendekatan berbasis data (data-driven decision making) menjadi sangat penting untuk memastikan strategi mitigasi yang disusun benar-benar relevan dan efektif. Data terbaru juga dapat membantu dalam pengembangan model prediksi yang lebih akurat sehingga tindakan preventif dapat dilakukan lebih awal.

Tidak kalah penting, sosialisasi dan pelatihan kesiapsiagaan bencana kepada masyarakat di daerah rawan gempa perlu ditingkatkan. Edukasi mengenai prosedur evakuasi, penggunaan sistem peringatan dini, dan simulasi bencana dapat membantu masyarakat lebih siap menghadapi risiko. Keterlibatan masyarakat menjadi kunci karena strategi mitigasi tidak hanya bergantung pada infrastruktur, tetapi juga pada tingkat kesadaran dan kesiapan penduduk.

Dengan demikian, hasil penelitian ini memberikan kontribusi nyata dalam memahami distribusi spasial risiko gempa di Indonesia. Penelitian ini diharapkan dapat menjadi rujukan awal bagi perumusan kebijakan mitigasi bencana, pengembangan sistem monitoring, serta penelitian lanjutan yang lebih mendalam. Pada akhirnya, tujuan utama dari pemetaan risiko ini adalah untuk meminimalkan dampak gempa bumi terhadap kehidupan masyarakat, infrastruktur, dan perekonomian nasional melalui strategi mitigasi yang lebih terarah, komprehensif, dan berkelanjutan.

REFERENSI

- [1] Angraeni, A., Lubis, L. H., Sugeng, S., & Ginting, M. H. (2023). Analisis Kerentanan Seismik Gempa Bumi berdasarkan Nilai PGA Menggunakan Metode Esteva pada Wilayah Kepulauan Nias. *Geo-Image Journal*, 12(2), 141-147.
- [2] Yudha, I. P. P. W. S., & Sinambela, M. (2024). Mitigasi Bencana Gempa Bumi dengan Integrasi Analisis Geofisika dan Data Mining. *Geosfera: Jurnal Penelitian Geografi*, 3(2), 121-129.
- [3] Metrikasari, R., & Choiruddin, A. (2021). Pemodelan risiko gempa bumi di pulau Sumatera menggunakan model inhomogeneous neyman-scott cox process. *Jurnal Sains dan Seni ITS*, 9(2), 487941.
- [4] Munir, R., Nurahmah, L., Friesky, V., Imran, M., & Haryono, A. (2024). Kajian Nilai B-Value untuk Menganalisis Aktivitas Seismik di Wilayah Pulau Jawa, Bali dan Nusa Tenggara Barat. *GEOSAINS KUTAI BASIN*, 7(2), 57-63.
- [5] Latif, M., Andriani, A., & Hakam, A. (2023). Analisis Tingkat dan Sebaran Bencana Tanah Longsor di Kabupaten Bengkulu Tengah. *Bentang: Jurnal Teoritis Dan Terapan Bidang Rekayasa Sipil*, 11 (2), 217–226.
- [6] Alamsyah, A., & Wahyudi, E. (2024). Transformasi Digital untuk Mitigasi Banjir: Optimalisasi Sistem Informasi di Jawa Barat. *Jurnal Perlindungan Masyarakat: Bestuur Praesidium*, 1(2), 50-62.
- [7] Baldah, A., Duarisah, A. V., & Maulana, R. A. (2023). Clustering Daerah Rawan Bencana Alam Di Indonesia Berdasarkan Provinsi Dengan Metode K-Means. *Jurnal Ilmiah Informatika Global*, 14(2), 31-36.
- [8] Wijaya, N. S., Jajuli, M., & Dermawan, B. A. (2024). Penerapan algoritma K-Means clustering dalam menentukan daerah prioritas penanganan kemiskinan di wilayah Jawa Timur. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(4), 7579-7584.

-
- [9] Faradilla, Z., & Veritawati, I. (2022). Clustering Daftar Harga Rumah di Jakarta Dengan Algoritma K-Means. *Journal of Informatics and Advanced Computing (JIAC)*, 3(2), 155-160.
- [10] Widyadhan, D., Hastuti, R. B., Kharisudin, I., & Fauzi, F. (2021, February). Perbandingan analisis kluster k-means dan average linkage untuk pengklasteran kemiskinan di provinsi jawa tengah. In *PRISMA, Prosiding Seminar Nasional Matematika* (Vol. 4, pp. 584-594).
- [11] Sari, R. M., Rizka, A., Putri, N. A., & Efriana, A. (2024). Perhitungan Metode Clustering. *Serasi Media Teknologi*.
- [12] Anjarwati, D. (2025). Clustering penindakan kasus korupsi di indonesia menggunakan k-means: analisis data kpk tahun 2004-2024. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 12(2).
- [13] Kinanti, R. (2024). Penerapan Metode Clustering K-Means Untuk Menentukan Prioritas Penerima Bantuan Program Beras Untuk Rakyat Miskin (Raskin) Studi Kasus: Kecamatan Siulak. *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)*, 4(2), 1135-1146.
- [14] Maori, N. A., & Evanita, E. (2023). Metode elbow dalam optimasi jumlah cluster pada k-means clustering. *Jurnal Simetris*, 14(2), 1-11.
- [15] Kiat, A. B. H., Azhar, Y., & Rahmayanti, V. (2020). Penerapan Metode K-Means Dengan Metode Elbow Untuk Segmentasi Pelanggan Menggunakan Model RFM (Recency, Frequency & Monetary). *Jurnal Repositor*, 2(7).
- [16] Baldah, A., Duarisah, A. V., & Maulana, R. A. (2023). Clustering Daerah Rawan Bencana Alam Di Indonesia Berdasarkan Provinsi Dengan Metode K-Means. *Jurnal Ilmiah Informatika Global*, 14(2), 31-36.
- [17] Hidayat, R. S., Muttaqin, M. R., & Irmayanti, D. (2024). Pengelompokan Daerah Rawan Bencana Di Jawa Tengah Menggunakan Algoritma K-Means Clustering. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(5), 10035-10042.
- [18] Al Halik, M. F., & Septiana, L. (2022). Analisa Data Untuk Prediksi Daerah Rawan Bencana Alam Di Jawa Barat Menggunakan Algoritma K-Means Clustering. *Journal of Information System, Applied, Management, Accounting and Research*, 6(4), 856-870.
- [19] Prasetio, A., & Effendi, M. M. (2023). Analisis Gempa Bumi Di Indonesia Dengan Metode Clustering. *Bulletin of Information Technology (BIT)*, 4(3), 338-343.
- [20] Massinai, M. F. I., Zarkasi, A., Maharani, A. A., & Desiani, R. (2022). Klasifikasi Kedalaman Kejadian Gempa Menggunakan Algoritma K-Means Clustering: Studi Kasus Kejadian Gempa Di Sulawesi. *JFT: Jurnal Fisika dan Terapannya*, 9(2), 79-88.
- [21] Wibowo, A., & Gunawan, W. (2024). Pemanfaatan Algoritma K-Means dalam Klasterisasi Gempa Sulawesi. *Faktor Exacta*, 17(3), 228-240.