

KLASTERISASI INTENSITAS DAN TREN KEJADIAN BENCANA DI INDONESIA (2018–2024) MENGGUNAKAN ALGORITMA K-MEANS

Muhamad Fahrizal¹, Haidar Reyhan², Suhendra³, Muhammad Rizki Asyakur⁴, Muhamad Nurdin⁵

Universitas Pamulang

E-mail: fzal3005@gmail.com¹, reyhandavin0509@gmail.com²,
mandrasuhendra7@gmail.com³, asyakur354@gmail.com⁴,
nurdinrhk@gmail.com⁵

Abstrak

Penelitian ini bertujuan mengelompokkan provinsi di Indonesia berdasarkan intensitas dan tren bencana (Januari 2018–Mei 2024) menggunakan algoritma K-Means. Berdasarkan 28.773 data kejadian dari BNPB, evaluasi model (Elbow Method & Silhouette Score 0,81) menghasilkan 3 kluster optimal dengan struktur kuat. Hasil menunjukkan pola "Jawa-sentris" yang ekstrem: Jawa Barat menempati kluster intensitas "Tinggi", disusul Jawa Tengah dan Jawa Timur pada kluster "Sedang", sementara 35 provinsi lainnya berada di kluster "Rendah". Temuan ini menjadi landasan empiris perlunya pemusatan logistik dan prioritas mitigasi bencana yang lebih agresif di Pulau Jawa.

Kata Kunci — Klasterisasi Data, Bencana Alam, K-Means, Analisis Tren, Jawa Barat.

Abstract

This study clusters Indonesian provinces based on disaster intensity and trends (January 2018–May 2024) using K-Means. Analyzing 28,773 BNPB records, model evaluation (Elbow Method & Silhouette Score 0.81) identified 3 optimal clusters with a strong structure. Results reveal an extreme "Java-centric" pattern: West Java forms the "High" intensity cluster, followed by Central and East Java in the "Medium" cluster, while the other 35 provinces fall into the "Low" cluster. These findings provide an empirical basis for prioritizing centralized logistics and aggressive mitigation efforts on Java Island.

Keywords— Data Clustering, Natural Disasters, K-Means, Trend Analysis, West Java

1. PENDAHULUAN

Indonesia secara geografis terletak di pertemuan tiga lempeng tektonik utama dunia dan berada dalam kawasan Cincin Api Pasifik (Ring of Fire), yang menjadikannya salah satu negara dengan tingkat kerawanan bencana tertinggi di dunia. Ancaman bencana tidak hanya berasal dari aktivitas geologi seperti gempa bumi dan letusan gunung api, tetapi juga dari bencana hidrometeorologi seperti banjir, tanah longsor, dan cuaca ekstrem yang frekuensinya terus meningkat akibat dampak perubahan iklim global. Data historis menunjukkan bahwa sebaran kejadian bencana sangat bervariasi antarwilayah, dipengaruhi oleh kondisi geomorfologi, kepadatan penduduk, dan kapasitas infrastruktur daerah. Oleh karena itu, pemetaan wilayah rawan bencana yang akurat dan berbasis data menjadi kebutuhan mendesak untuk mendukung strategi mitigasi yang efektif dan tepat sasaran.

Pemanfaatan teknik data mining, khususnya algoritma klasterisasi, telah banyak diterapkan dalam penelitian terdahulu untuk memetakan risiko bencana. Penelitian oleh (Koagouw et al., 2024) menerapkan algoritma K-Means untuk memetakan daerah rawan banjir di Kota Manado, sementara (Rohman et al., 2024) melakukan klasterisasi bencana alam pada skala regional di Jawa Barat. Studi lain oleh (Hafid & Muthahharah, 2025) dan (Sudrajat et al., 2025) juga menggunakan pendekatan serupa untuk mengelompokkan wilayah berdasarkan indeks risiko. Meskipun memberikan kontribusi signifikan, mayoritas penelitian tersebut (State of the Art) cenderung berfokus pada wilayah administratif yang sempit (kabupaten/kota atau satu provinsi) dan umumnya hanya menggunakan akumulasi jumlah kejadian sebagai variabel tunggal tanpa memperhitungkan dinamika perubahan dari waktu ke waktu.

Keterbatasan utama dari studi-studi terdahulu adalah belum adanya integrasi analisis tren (trend analysis) sebagai fitur penentu dalam pembentukan kluster. Mengandalkan total kejadian semata dapat bias terhadap wilayah yang memiliki lonjakan kejadian di masa lalu namun trennya kini menurun, atau sebaliknya. Penelitian ini hadir untuk mengisi kesenjangan (gap) tersebut dengan menawarkan kebaruan (novelty) berupa penerapan Feature Engineering yang lebih kompleks pada skala nasional.

Penelitian ini menggunakan dataset terbaru dari Badan Nasional Penanggulangan Bencana (BNPB) periode Januari 2018 hingga Mei 2024. Kebaruan ilmiah utama dalam studi ini terletak pada penggunaan variabel kemiringan tren (slope) hasil regresi linear sederhana sebagai atribut klusterisasi, mendampingi atribut statistik standar (total, rata-rata, dan standar deviasi). Pendekatan ini memungkinkan identifikasi karakteristik wilayah yang lebih mendalam, membedakan antara provinsi dengan intensitas bencana yang statis dan provinsi yang mengalami tren peningkatan signifikan. Tujuan akhir dari penelitian ini adalah mengelompokkan 38 provinsi di Indonesia ke dalam level risiko yang objektif guna memberikan rekomendasi prioritas logistik dan kebijakan mitigasi yang lebih responsif terhadap anomali kejadian di Pulau Jawa maupun luar Jawa.

2. METODE PENELITIAN

Penelitian ini menerapkan pendekatan kuantitatif menggunakan teknik Data Mining untuk mengungkap pola tersembunyi pada data kejadian bencana. Sumber data utama yang digunakan adalah data sekunder dari Portal Data Terbuka Badan Nasional Penanggulangan Bencana (BNPB). Dataset mencakup rekapitulasi kejadian bencana alam harian di seluruh provinsi Indonesia dengan rentang waktu pengamatan dari Januari 2018 hingga Mei 2024. Total volume data yang diproses berjumlah 28.773 baris data mentah yang memuat atribut tanggal kejadian, jenis bencana, lokasi (provinsi/kabupaten), dan dampak yang ditimbulkan.

Proses analisis data dilakukan menggunakan bahasa pemrograman Python dengan pustaka utama Pandas untuk manipulasi data dan Scikit-Learn untuk pemodelan mesin pembelajaran (Pedregosa et al., 2011). Tahapan penelitian dimulai dengan pra-pemrosesan data (data preprocessing) yang meliputi pembersihan data dari nilai yang hilang (missing values) dan penyeragaman format nama provinsi. Selanjutnya, dilakukan tahapan rekayasa fitur (Feature Engineering) untuk mentransformasi data transaksional harian menjadi data agregat per provinsi. Empat variabel prediktor dibentuk untuk setiap provinsi, yaitu: (1) Total kejadian akumulatif; (2) Rata-rata kejadian bulanan (Mean); (3) Standar deviasi (Std Dev) untuk mengukur volatilitas atau ketidakpastian kejadian; dan (4) Nilai kemiringan tren (Slope) yang dihitung menggunakan regresi linear sederhana terhadap deret waktu kejadian untuk mengukur laju peningkatan atau penurunan bencana selama periode pengamatan.

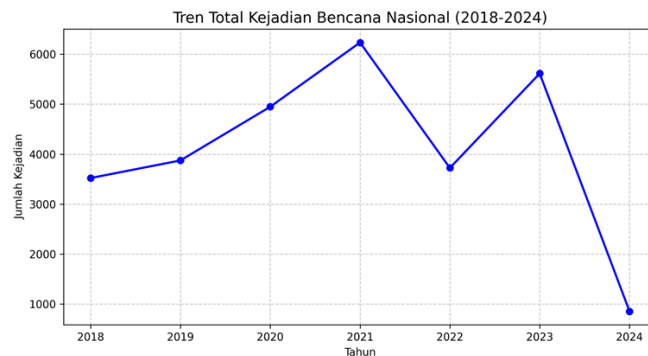
Sebelum proses klusterisasi, seluruh variabel dinormalisasi menggunakan teknik StandardScaler (Z-Score) untuk menyetarakan rentang nilai antar-variabel, sehingga mencegah dominasi fitur dengan satuan skala yang besar. Pengelompokan wilayah dilakukan menggunakan algoritma K-Means Clustering, sebuah metode unsupervised learning yang mempartisi data ke dalam k kelompok berdasarkan kedekatan jarak Euclidean terhadap titik pusat kluster (centroid) (Jain, 2010).

Penentuan jumlah kluster optimal dievaluasi menggunakan Elbow Method dengan mengamati titik perubahan drastis pada nilai Within-Cluster Sum of Squares (WCSS). Validasi kualitas hasil kluster diukur menggunakan metrik Silhouette Score, di mana nilai yang mendekati +1 mengindikasikan struktur kluster yang kuat dan terpisah dengan baik (Rousseeuw, 1987). Interpretasi hasil akhir dilakukan dengan memetakan label kluster ke dalam peta spasial dan menganalisis karakteristik statistik dari setiap kelompok untuk menentukan level risiko bencana (Rendah, Sedang, dan Tinggi).

3. HASIL DAN PEMBAHASAN

1. Analisis Tren dan Dinamika Data Kejadian Bencana

Sebelum dilakukan pengelompokan wilayah, analisis deskriptif dilakukan terhadap data deret waktu (time series) untuk memahami tren kejadian bencana secara nasional. Sebagaimana terlihat pada Gambar 1, fluktuasi kejadian bencana di Indonesia sangat dinamis sepanjang periode 2018 hingga 2023. Namun, pada tahun 2024 terlihat penurunan grafis yang tajam. Hal ini bukan disebabkan oleh penurunan risiko bencana secara drastis, melainkan karena keterbatasan data pengamatan yang baru terekam hingga bulan Mei 2024 (data parsial). Fenomena ini menjadi dasar pentingnya penggunaan variabel kemiringan tren (slope) dalam proses klusterisasi, agar wilayah yang mengalami penurunan semu akibat data parsial dapat dibedakan dengan wilayah yang memang memiliki intensitas rendah secara konsisten.

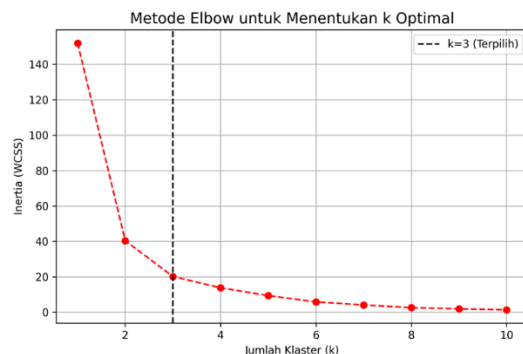


Gambar 1. Tren Fluktuasi Kejadian Bencana Nasional (Januari 2018 – Mei 2024) Sumber: Hasil Pengolahan Data Penulis (2025)

2. Evaluasi Penentuan Jumlah Kluster Optimal

Tahap krusial dalam algoritma K-Means adalah penentuan jumlah kluster (k) yang paling representatif untuk mempartisi data. Penelitian ini menggunakan pendekatan Elbow Method untuk mengidentifikasi titik optimal di mana penurunan nilai Within-Cluster Sum of Squares (WCSS) mulai melandai secara signifikan. Berdasarkan hasil pengujian terhadap rentang k=1 hingga k=10 yang ditampilkan pada Gambar 2, terlihat terbentuknya patahan sudut (elbow) yang sangat tajam pada titik k=3. Penambahan jumlah kluster setelah titik ini tidak memberikan penurunan inersia yang substansial, sehingga k=3 dipilih sebagai jumlah kluster terbaik.

Untuk memvalidasi kualitas struktur kluster yang terbentuk pada k=3, dilakukan perhitungan Silhouette Score. Metrik ini mengukur seberapa dekat setiap titik data dengan kluster tetangganya (separation) dibandingkan dengan kluster sendiri (cohesion). Hasil evaluasi menunjukkan nilai rata-rata Silhouette Score sebesar 0.81. Mengacu pada kriteria Rousseeuw (1987), nilai di atas 0.70 dikategorikan sebagai Strong Structure (Struktur Kuat). Hal ini mengindikasikan bahwa pengelompokan provinsi ke dalam tiga level risiko bencana (Rendah, Sedang, Tinggi) memiliki batas pemisah yang sangat tegas dan objektif secara statistik, bukan sekadar pengelompokan yang dipaksakan.



Gambar 2. Grafik Evaluasi Elbow Method Menunjukkan Titik Optimal pada k=3 Sumber: Hasil Pengolahan Data Penulis (2025)

3. Analisis Profil dan Karakteristik Klaster

Hasil klasterisasi menggunakan algoritma K-Means berhasil memetakan 38 provinsi di Indonesia ke dalam tiga strata risiko yang distingsif. Ringkasan karakteristik statistik dari pusat klaster (centroid) ditampilkan pada Tabel 1. Temuan paling menonjol dari analisis ini adalah terbentuknya pola distribusi yang sangat timpang dan terpusat di Pulau Jawa (Java-centric). Klaster 2 yang dilabeli sebagai level "Tinggi" hanya memiliki satu anggota tunggal, yaitu Provinsi Jawa Barat. Provinsi ini mencatat akumulasi kejadian bencana yang ekstrem sebesar 6.159 kejadian selama periode pengamatan, jauh melampaui rata-rata provinsi lainnya. Hal ini menjadikan Jawa Barat sebagai outlier utama dalam peta kebencanaan nasional, yang mengindikasikan bahwa provinsi ini menghadapi tekanan risiko hidrometeorologi yang jauh lebih besar dibandingkan wilayah lain.

Klaster 1 atau level "Sedang" ditempati oleh dua provinsi tetangga, yaitu Jawa Tengah dan Jawa Timur, dengan rata-rata total kejadian sebesar 3.444. Sementara itu, Klaster 0 atau level "Rendah" mencakup mayoritas wilayah Indonesia (35 provinsi) dengan rata-rata kejadian yang jauh lebih kecil, yakni sekitar 354 kejadian per provinsi. Ketimpangan jumlah anggota klaster ini (1 vs 2 vs 35) secara statistik mengonfirmasi bahwa beban bencana di Indonesia tidak terdistribusi secara normal, melainkan mengalami kemencengan (skewness) positif yang ekstrem akibat dominasi kejadian di Pulau Jawa.

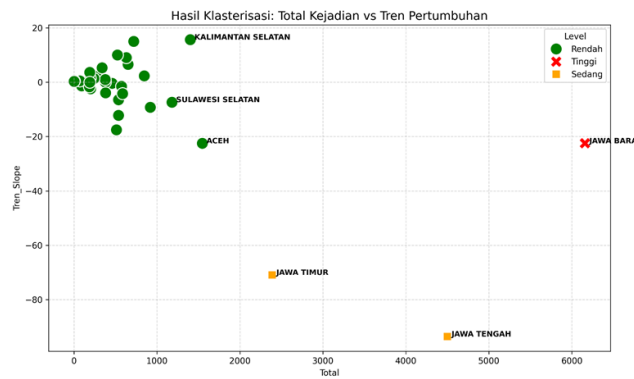
Tabel 1. Ringkasan Karakteristik Klaster Kejadian Bencana (2018–2024)

Level Kluster	Jumlah Provinsi	Rata-rata Total Kejadian	Rata-rata Tren (Slope)	Anggota Utama (Provinsi)
Tinggi	1	6.159	-22,46	Jawa Barat
Sedang	2	3.444	-82,21	Jawa Tengah, Jawa Timur
Rendah	35	354	-0,52	Aceh, Sumatera Utara, dll.

Analisis mendalam terhadap fitur Tren (Slope) pada Tabel 1 menunjukkan nilai negatif pada seluruh klaster, dengan penurunan paling tajam terlihat pada klaster Sedang (-82,21) dan Tinggi (-22,46). Nilai slope negatif ini perlu diinterpretasikan secara hati-hati (prudent). Secara teknis, nilai ini sangat dipengaruhi oleh titik data tahun 2024 yang baru terekam hingga bulan Mei (data parsial). Penurunan drastis jumlah kejadian dari titik puncak tahun 2023 ke data parsial 2024 menyebabkan garis regresi linear "tertarik" ke bawah secara signifikan. Meskipun demikian, besarnya magnitudo total kejadian pada klaster Tinggi dan Sedang menunjukkan bahwa meskipun tren statistik terlihat menurun akibat cutoff data, intensitas frekuensi bencana harian di wilayah tersebut tetap berada pada level yang mengkhawatirkan dan membutuhkan atensi prioritas.

Validasi visual terhadap karakteristik ini dapat dilihat pada Gambar 3 (Scatter Plot). Grafik tersebut memperlihatkan separasi spasial yang tegas, di mana titik data Jawa Barat (merah) terisolasi di kuadran kanan atas, terpisah jauh dari gerombolan provinsi lain. Jarak Euclidean yang lebar antara titik Jawa Barat dengan centroid klaster lainnya menegaskan

bahwa karakteristik risiko di Jawa Barat memiliki keunikan tersendiri yang tidak dapat disamakan dengan provinsi lain, bahkan dengan sesama provinsi di Pulau Jawa.



Gambar 3. Visualisasi Persebaran Provinsi Berdasarkan Klaster (Scatter Plot) Sumber: Hasil Pengolahan Data Penulis (2025)

4. Analisis Spasial dan Diskusi Wilayah Rawan Bencana

Interpretasi visual dari hasil klasterisasi ditampilkan melalui peta koroplet pada Gambar 4. Peta tersebut memperlihatkan polarisasi yang sangat kontras antara Pulau Jawa dan wilayah kepulauan lainnya. Warna merah menyala yang menandai Jawa Barat (Klaster Tinggi) dan warna oranye pada Jawa Tengah dan Jawa Timur (Klaster Sedang) mengonfirmasi bahwa Pulau Jawa merupakan "episentrum" bencana hidrometeorologi nasional. Sebaliknya, wilayah Sumatera, Kalimantan, Sulawesi, hingga Papua didominasi oleh warna hijau (Klaster Rendah).

Fenomena "Jawa-sentris" ini sejalan dengan teori kerentanan fisik dan sosial. Tingginya frekuensi kejadian di Jawa Barat tidak hanya dipicu oleh faktor alam, tetapi juga berbanding lurus dengan tekanan antropogenik. Sebagai provinsi dengan populasi terbesar di Indonesia, konversi lahan resapan menjadi area pemukiman dan industri di Jawa Barat memicu eskalasi kejadian banjir dan longsor yang lebih intensif dibandingkan provinsi lain dengan wilayah hutan yang masih luas seperti di Kalimantan atau Papua.

Temuan ini memperbarui studi sebelumnya oleh Rohman et al. (2024) yang hanya memetakan risiko internal Jawa Barat. Penelitian ini membuktikan bahwa dalam skala nasional, Jawa Barat bukan hanya "salah satu" daerah rawan, melainkan provinsi dengan anomali risiko tertinggi yang terpisah dari klaster lainnya. Selain itu, variabel tren (slope) negatif yang ditemukan pada klaster tinggi memberikan sinyal kewaspadaan; meskipun data parsial 2024 menunjukkan penurunan grafik, akumulasi historis membuktikan bahwa wilayah ini memiliki potensi bahaya laten yang sewaktu-waktu dapat melonjak kembali saat musim penghujan tiba.

Peta Sebaran Klaster Intensitas Bencana Indonesia (2018-2024)



Gambar 4. Peta Sebaran Tingkat Kerawanan Bencana Per Provinsi (2018–2024) Sumber: Hasil Pengolahan Data Penulis (2025)

4. KESIMPULAN

Penelitian ini berhasil menjawab tujuan utama untuk memetakan pola spasial dan mengelompokkan risiko bencana di Indonesia berdasarkan data intensitas serta tren kejadian periode 2018–2024. Melalui penerapan algoritma K-Means dengan optimasi feature engineering, disimpulkan bahwa distribusi bencana di Indonesia memiliki struktur ketimpangan yang ekstrem atau bersifat "Jawa-sentris". Hipotesis awal mengenai heterogenitas risiko terbukti dengan terbentuknya tiga kluster distingsif yang valid secara statistik (Silhouette Score = 0.81). Temuan ilmiah utama menunjukkan bahwa Provinsi Jawa Barat merupakan wilayah dengan anomali risiko tertinggi (Klaster Tinggi) yang terpisah secara signifikan dari provinsi lain, diikuti oleh Jawa Tengah dan Jawa Timur (Klaster Sedang).

Secara implikatif, hasil ini mengindikasikan bahwa manajemen penanggulangan bencana nasional tidak dapat lagi menggunakan pendekatan generalis "satu kebijakan untuk semua wilayah". Diperlukan strategi asimetris, di mana alokasi sumber daya, logistik, dan teknologi peringatan dini harus dipusatkan secara masif di Pulau Jawa, khususnya Jawa Barat, untuk mengantisipasi frekuensi kejadian yang tinggi. Untuk penelitian selanjutnya, disarankan untuk memperluas variabel analisis dengan mengintegrasikan data curah hujan, kepadatan penduduk, dan tutupan lahan (land cover) guna mendapatkan pemahaman kausalitas yang lebih komprehensif, serta memperbarui dataset hingga akhir tahun 2024 untuk menyempurnakan analisis tren.

DAFTAR PUSTAKA

- Badan Nasional Penanggulangan Bencana. (2024). Data Kejadian Bencana Indonesia: Portal Data Terbuka. Diakses dari <https://gis.bnpb.go.id/>
- Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2), 224–227.
- Hafid, H., & Muthahharah, I. (2025). Clustering of disaster risk in Indonesian regions using Self-Organizing Maps and K-Means. *Journal of Mathematics: Theory and Applications*, 7(2), 45-56.
- Haris, N., Furqan, A. C., Kahar, A., & Karim, F. (2023). Disaster risk index on disaster management budgeting: Indonesia's national data set. *Jambá: Journal of Disaster Risk Studies*, 15(1), 1–7.
- Hidayat, I., Darnila, E., & Afrillia, Y. (2023). Clustering zonasi daerah rawan bencana alam di Kabupaten Mandailing Natal menggunakan algoritma K-Means. *G-Tech: Jurnal Teknologi Terapan*, 7(3), 1218–1226.
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666.
- Januzaj, Y., Beqiri, E., & Luma, A. (2023). Determining the optimal number of clusters using Silhouette Score as a data mining technique. *International Journal of Online and Biomedical Engineering (iJOE)*, 19(4), 174–182.
- Koagouw, A., Paseru, D., & Kairupan, I. (2024). Geographical information system for mapping flood-prone areas in Manado City using the K-Means clustering method. *Journal of Information Technology and Its Utilization*, 7(1), 1–7.
- Kusnaldi, M. R., Gulo, T., & Aripin, S. (2022). Penerapan normalisasi data dalam mengelompokkan data mahasiswa dengan metode K-Means untuk menentukan prioritas bantuan uang kuliah tunggal. *Journal of Computer System and Informatics*, 3(4), 330–338.
- Nasliati, Bisri, M. B. F., Sasaki, D., & Oktari, R. S. (2025). Indonesia Disaster Education Repository House (In-Depth) as a knowledge management system platform for disaster education. *International Journal of Disaster Risk Reduction*, 123, 105502.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Rohman, D., Annisa, R., Efendi, D. I., & Solahudin, D. (2024). Clustering bencana alam menggunakan K-Means pada wilayah Jawa Barat. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 493–500.

- 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.
- Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-Means clustering algorithm. *IEEE Access*, 8, 80716–80727.
- Sudrajat, R., Hadiana, A. I., & Melina, M. (2025). Evaluasi kualitas klaster wilayah rawan bencana menggunakan K-Means dengan Silhouette dan Elbow Method. *Jurnal Algoritma*, 22(2), 127–139.
- Wibowo, S. A., Putra, E. P., Hanadila, B. L., & Sari, P. (2024). Natural disaster mapping and clustering in Indonesia. *Disaster Advances*, 17(3), 20–33.