



Pengelompokkan Daerah Rawan Bencana di Indonesia Menggunakan Metode Clustering K-Means

Seprina Aulia Putri *

Universitas Asahan, Indonesia

Email: seprinaauliaputri02@gmail.com

Alamat: Jl. Ahmad Yani No. 1, Sumatera Utara 21214, Indonesia

Korespondensi penulis : seprinaauliaputri02@gmail.com*

Abstract. Indonesia is a country with a high level of disaster vulnerability, influenced by tectonic activity and its tropical climate. This study uses the K-Means clustering method to identify and group disaster-prone areas based on the level of vulnerability. The data used included average temperature (Tavg) and rainfall (RR) which were processed using Python. The analysis process includes data collection, pre-processing, determination of key features, and evaluation of clustering quality using the Elbow and Silhouette Score methods. The results of the grouping show two main patterns, namely flood-prone areas and drought-prone areas. These findings are expected to support the government in more effective and data-based disaster mitigation planning.

Keywords: K-Means, clustering, disaster-prone, Python, disaster mitigation

Abstrak. Indonesia merupakan negara dengan tingkat kerawanan bencana yang tinggi, dipengaruhi oleh aktivitas tektonik serta iklim tropisnya. Penelitian ini menggunakan metode clustering K-Means untuk mengidentifikasi dan mengelompokkan daerah rawan bencana berdasarkan tingkat kerawanan. Data yang digunakan meliputi suhu rata-rata (Tavg) dan curah hujan (RR) yang diolah menggunakan Python. Proses analisis mencakup pengumpulan data, pra-pemrosesan, penentuan fitur utama, serta evaluasi kualitas clustering menggunakan metode Elbow dan Silhouette Score. Hasil pengelompokan menunjukkan dua pola utama, yaitu daerah rawan banjir dan daerah rawan kekeringan. Temuan ini diharapkan dapat mendukung pemerintah dalam perencanaan mitigasi bencana yang lebih efektif dan berbasis data.

Kata kunci: K-Means, clustering, rawan bencana, Python, mitigasi bencana

1. LATAR BELAKANG

Indonesia berada di kawasan Cincin Asia Pasifik, yang dikenal memiliki aktivitas tektonik yang sangat tinggi. Kondisi ini membuat Indonesia sangat rentan terhadap bencana alam, seperti gempa bumi dan erupsi gunung berapi (Wicaksono dan Susetyo, 2023). Selain itu, Indonesia memiliki dua musim utama, yaitu musim hujan dan musim kemarau, yang turut berkontribusi pada potensi terjadinya bencana. Selama musim kemarau, risiko bencana seperti kebakaran hutan dan kekeringan cenderung meningkat, sedangkan di musim hujan, ancaman banjir, tanah longsor, dan angin puting beliung menjadi lebih tinggi (Ramadhani et al., 2022).

Bencana alam memiliki dampak signifikan bagi masyarakat, infrastruktur, dan perekonomian. Contohnya, bencana banjir dapat merusak rumah-rumah warga, mengganggu aktivitas bisnis, serta menyebabkan korban jiwa. Banjir juga dapat merusak infrastruktur publik seperti pentingnya analisis data cuaca jalan, jembatan, dan fasilitas umum, sehingga memperlambat rekonstruksi dan pembangunan ulang.

Identifikasi serta pengelompokan daerah rawan bencana di Indonesia telah dilakukan melalui berbagai metode, seperti analisis manual menggunakan data historis atau pemetaan berbasis geografis. Namun, pendekatan-pendekatan tersebut seringkali kurang efisien dalam memberikan hasil yang cepat dan akurat, terutama dalam mengelompokkan wilayah berdasarkan tingkat kerawanan yang kompleks. Hingga kini, penelitian yang memanfaatkan algoritma berbasis data, seperti metode clustering, untuk mengidentifikasi daerah rawan bencana di Indonesia masih terbatas.

Oleh karena itu penelitian ini bertujuan untuk mengatasi kesenjangan tersebut dengan menggunakan metode clustering *K-Means* untuk menganalisis dan memetakan daerah rawan bencana. Metode ini memungkinkan pengelompokan data secara otomatis berdasarkan tingkat kerawanan bencana, sehingga dapat memberikan gambaran yang lebih jelas tentang daerah-daerah prioritas. Melalui metode ini, penelitian ini tidak hanya mendukung perencanaan mitigasi berbasis data, tetapi juga membantu pemerintah dan pembuat kebijakan dalam merancang strategi penanggulangan bencana yang lebih efektif dan efisien.

2. KAJIAN TEORITIS

Konsep Bencana Alam

Bencana alam adalah peristiwa yang mengancam dan mengganggu kehidupan manusia. Bencana ini dapat dipicu oleh faktor alam, non-alam, maupun aktivitas manusia, yang mengakibatkan kehilangan nyawa, kerusakan lingkungan, kerugian harta benda, serta tekanan psikologis. Selain dampak langsungnya, bencana alam juga memengaruhi kestabilan sosial, ekonomi, dan politik, terutama di wilayah yang rentan terhadap bencana. Contohnya, gempa bumi, banjir, tanah longsor, dan letusan gunung berapi tidak hanya menimbulkan kerugian langsung, tetapi juga dapat memicu masalah lain seperti kelaparan, wabah penyakit, dan konflik sosial. Oleh karena itu, penting bagi masyarakat dan pemerintah untuk mengenali potensi risiko bencana, meningkatkan kesadaran, serta menerapkan strategi mitigasi yang efektif untuk meminimalkan dampaknya.

Data Mining

Data mining adalah proses pengumpulan dan analisis data historis untuk mengungkap pola, tren dan hubungan dalam kumpulan data besar (Ikhwan & Aslami, 2020). Prosedur ini mengambil informasi penting melalui penggunaan pendekatan statistik, algoritma pembelajaran mesin, dan analisis data. Di era digital saat ini, data mining sangat penting daldi

banyak industri, termasuk pemasaran, kesehatan, dan analisis sosial, di mana pengambilan keputusan berbasis data menjadi lebih populer.

Algoritma Clustering

Selama proses clustering, data yang serupa akan disatukan, sementara data yang berbeda akan dipisahkan. Menurut (Nugroho & Wahyuni, 2022), ada beberapa algoritma clustering yang tersedia, antara lain K-Means, Hierarchical Clustering, dan DBSCAN. Setiap algoritma memiliki pendekatan dan keunggulan yang unik. Misalnya, K-Means adalah salah satu algoritma yang umum untuk mengelompokkan data dalam jumlah besar karena sederhana dan efisien. Algoritma ini bekerja dengan menentukan jumlah cluster yang diperlukan dan mengelompokkan data berdasarkan jarak terdekat ke pusat cluster (centroid).

Python

Python adalah bahasa pemrograman tingkat tinggi. Menurut (Kadarsih & Andrianto, 2022), perusahaan dan pengembang signifikan menggunakan Python untuk membuat berbagai aplikasi desktop, online, dan seluler. Python juga merupakan bahasa pemrograman yang mudah beradaptasi dan mudah dipelajari. Python menawarkan modul standar yang menyediakan berbagai macam metode dan algoritme untuk memenuhi tugas-tugas seperti memproses data teks, memanipulasi dan menemukan file pada disk, membaca dan menulis file terkompresi, dan menerima data dari server web.

3. METODE PENELITIAN

Teknik Analisis Data

1. Pengumpulan Data

- **Langkah 1:** Mengumpulkan data dari sumber terpercaya, yaitu Badan Meteorologi, Klimatologi, dan Geofisika (BMKG).
- **Langkah 2:** Data yang dikumpulkan mencakup informasi penting seperti suhu rata-rata (Tavg), kelembapan rata-rata (RR), dan curah hujan harian dari berbagai wilayah di Indonesia.
- **Langkah 3:** Memastikan data yang digunakan bersifat konsisten dan mencakup periode waktu yang relevan untuk penelitian ini, yaitu hingga 5 Oktober 2024.

2. Pra-pemrosesan Data

- **Langkah 1:** Membersihkan data mentah dengan menghapus informasi duplikat atau data yang tidak relevan untuk analisis.

- **Langkah 2:** Melakukan proses normalisasi data agar semua variabel memiliki skala yang seragam untuk menghindari bias dalam perhitungan.
- **Langkah 3:** Mengonversi data kategorikal, seperti jenis bencana, menjadi format numerik agar dapat diproses oleh algoritma K-Means.

3. Penentuan Fitur Utama

- **Langkah 1:** Menentukan variabel utama yang relevan untuk analisis, yaitu suhu rata-rata (T_{avg}) dan curah hujan (RR).
- **Langkah 2:** Memilih fitur-fitur yang signifikan agar model K-means dapat mengelompokkan data dengan baik.

4. Penerapan Algoritma K-means

- **Langkah 1:** Menentukan jumlah cluster yang optimal menggunakan metode seperti Elbow atau Silhouette untuk menemukan titik terbaik jumlah cluster.
- **Langkah 2:** Mengelompokkan data menggunakan algoritma K-Means berdasarkan variabel yang telah dipilih.
- **Langkah 3:** Mengelompokkan data ke dalam dua pola utama yaitu : ‘daerah rawan banjir, dan ‘daerah rawan kekeringan’ untuk mitigasi bencana alam.

5. Evaluasi Kualitas Clustering

- **Langkah 1:** Menggunakan Elbow Method untuk mengevaluasi jumlah cluster yang optimal, serta memastikan bahwa pembagian cluster menghasilkan nilai inertia yang stabil.
- **Langkah 2:** Menghitung Silhouette Score untuk mengukur sejauh mana data terkelompok dengan benar dalam cluster yang sesuai.
- **Langkah 3:** Melakukan validasi hasil pengelompokan dengan membandingkan pola yang terbentuk dengan data cuaca historis atau tren cuaca yang sudah ada.

4. HASIL DAN PEMBAHASAN

Pengumpulan Data

Pengelompokan data daerah rawan bencana dilakukan dengan algoritma K-means yang mengkategorikan data menjadi dua pola utama, yaitu ‘daerah rawan banjir, dan ‘daerah rawan kekeringan’. Proses ini bertujuan untuk membantu pemahaman yang lebih baik mengenai kondisi cuaca, sehingga dapat mendukung upaya mitigasi bencana. Dalam penelitian ini, data cuaca yang dianalisis meliputi suhu rata-rata dan curah hujan berdasarkan enam sample data harian.

1. Data Bencana Sample untuk Clustering

Tabel 1. Data bencana sample unruk clustering

sample data	X	Y
1	34.5	21.0
2	55.7	22.4
3	64.2	25.2
4	89.8	25.6
5	194.3	24.6
6	32.1	29.1

Keterangan:

- Sample data 1-6 adalah urutan nama stasiun.
- X adalah nilai curah hujan dan Y adalah nilai suhu rata-rata.

Analisis Metode K- means

Kita akan mengaplikasikan k-means clustering untuk data diatas menjadi dua cluster,pertama kita akan hitung centroid.

Cluster	X	Y
K1	34.5	21.0
K2	55.7	22.4

Perhitungan jarak menggunakan persamaan Euclidean Distance dilakukan antara setiap data sampel dengan dua centroid ini:

$$d = \sqrt{(x1 - x2)^2 + (y1 - y2)^2}$$

Iterasi pertama

Pada iterasi pertama , dua centroid awal di pilih

Jarak antara sample pertama (34.5,21.0) ke centroid cluster 1 dan cluster 2 :

- o Jarak ke cluster 1: 0
- o Jarak ke cluster 2 : 21.24

1. Pengelompokan data berdasarkan iterasi pertama

Tabel 2. Pengelompokan data iterasi pertama

No	X	y	Jarak 1	Jarak 2	kelompok cluster
1	34.5	21.0	0	21.24	1
2	55.7	22.4	21.24	0	2
3	64.2	25.2	29.99	8.94	2
4	89.8	25.6	55.49	34.24	2
5	194.3	24.6	159.84	138.61	2
6	32.1	29.1	8.44	24.53	1

Iterasi kedua

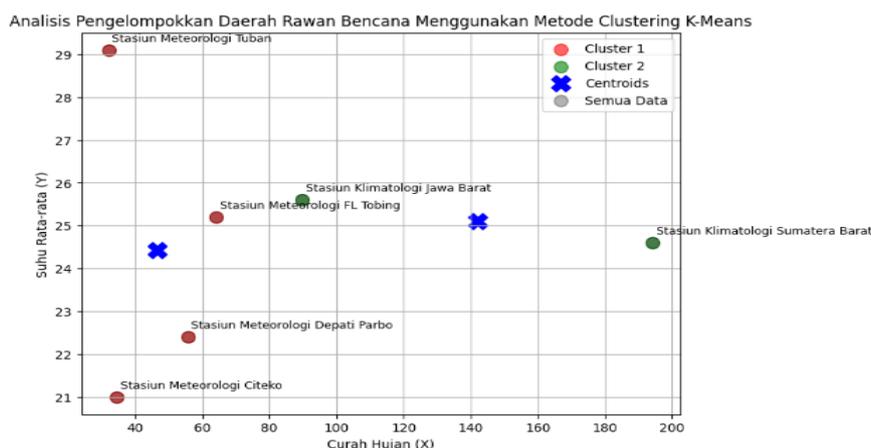
Pada iterasi kedua, jarak setiap sampel dihitung kembali berdasarkan centroid baru setelah langkah 1

Cluster 1 (33.3,25.05)

Cluster 2 (101.025,24.45)

Tabel 3. Pengelompokan data iterasi kedua

No	x	y	Jarak 1	Jarak 2	kelompok cluster
1	34.5	21.0	4.22	66.61	1
2	55.7	22.4	22.55	45.37	1
3	64.2	25.2	30.90	36.83	1
4	89.8	25.6	56.50	11.28	2
5	194.3	24.6	161.00	93.27	2
6	32.1	29.1	4.22	69.08	1



Grafik 1. Pengelompokan Daerah Rawan Bencana

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengimplementasikan metode clustering K-Means untuk mengelompokkan daerah rawan bencana di Indonesia. Berdasarkan hasil analisis, dua pola utama berhasil diidentifikasi, yaitu *daerah rawan banjir* dan *daerah rawan kekeringan*. Data yang digunakan mencakup suhu rata-rata (T_{avg}) dan kelembapan rata-rata (RR), yang diolah secara efektif dengan algoritma K-Means menggunakan Python. Evaluasi kualitas clustering melalui metode Elbow dan Silhouette Score menunjukkan hasil yang optimal dalam pengelompokan data. Dengan pendekatan berbasis data ini, hasil penelitian diharapkan dapat menjadi panduan bagi pemerintah dan pemangku kebijakan dalam merancang strategi mitigasi bencana yang lebih terarah dan efisien. Selain itu, metode ini juga dapat diterapkan pada data cuaca lainnya untuk pengelompokan serupa, memberikan peluang pengembangan lebih lanjut dalam analisis risiko bencana di masa mendatang.

DAFTAR REFERENSI

- Al Halik, M. F., & Septiana, L. (2022). Analisa Data Untuk Prediksi Daerah Rawan Bencana Alam Di Jawa Barat Menggunakan Algoritma K-Means Clustering. *JISAMAR (Journal of Information System, Applied, Management, Accounting and Research)*, 6(4), 856-870.
- Anam, M. K., Anto, A., & Wulandari, C. (2023). A PENERAPAN METODE CLUSTERING K-MEANS DALAM PENGELOMPOKAN PENERIMA KWH METER GRATIS DENGAN BAHASA PEMROGRAMAN PYTHON. *JTIM: Jurnal Teknik Informatika Mahakarya*, 6(1), 31-35.
- Asyuti, S., & Setyawan, A. A. (2023). Data Mining Dalam Penggunaan Presensi Karyawan Denga Cluster Means. *Jurnal Ilmiah Sains Teknologi dan Informasi*, 1(1), 1-10.

- Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., & Jajuli, M. (2022). Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi dan Informasi*, 12(1), 64-77.
- Hasanah, M., Defit, S., & Nurcahyo, G. W. (2019). Implementasi Algoritma K-Means untuk Klasterisasi Peserta Olimpiade Sains Nasional Tingkat SMA. *Jurnal Sistim Informasi dan Teknologi*, 30-35.
- Ikhwan, A., & Aslami, N. (2020). Implementasi Data Mining untuk Manajemen Bantuan Sosial Menggunakan Algoritma K-Means. (*JurTI*) *Jurnal Teknologi Informasi*, 4(2), 208-217.
- Lubis, N. J. A., Harahap, S. Z., & Ritonga, I. (2024). Analisis K-Means dan Naive Bayes Untuk Pengelompokan Rawan Bencana di Daerah Kabupaten Labuhanbatu. *INFORMATIKA*, 12(1), 81-92.
- Otniel, M. V., & Prasetyo, S. Y. J. (2024). Pemanfaatan K-Means Clustering Untuk Pengelompokan Dan Pemetaan Bencana Alam Di Indonesia. *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*, 5(4), 1670-1678.
- Pratama, Y., Hendrawan, H., Rasywir, E., Carenina, B. T., & Anggraini, D. R. (2022). Penerapan Algoritma K-Means clustering Untuk Mengelompokkan Provinsi Berdasarkan Banyaknya Desa/Kelurahan Dengan Upaya Antisipasi/Mitigasi Bencana Alam. *Building of Informatics, Technology and Science (BITS)*, 4(3), 1232-1240.
- Rakuasa, H., & Latue, P. C. (2023). Analisis Spasial Daerah Rawan Banjir Di Das Wae Heru, Kota Ambon. *Jurnal Tanah Dan Sumberdaya Lahan*, 10(1), 75-82
- Ramadhan, A., Mustakim, M., & Handinata, R. (2019). Implementasi Algoritma Fuzzy C-Means untuk Pengelompokan Wilayah Bencana Banjir. In *Seminar Nasional Teknologi Informasi Komunikasi dan Industri* (pp. 171-177).
- Romzi, M., & Kurniawan, B. (2020). Pembelajaran Pemrograman Python Dengan Pendekatan Logika Algoritma. *JTIM: Jurnal Teknik Informatika Mahakarya*, 3(2), 37-44.
- Sadewo, M. G., Eriza, A., Windarto, A. P., & Hartama, D. (2019, February). Algoritma K-Means Dalam Mengelompokkan Desa/Kelurahan Menurut Keberadaan Keluarga Pengguna Listrik dan Sumber Penerangan Jalan Utama Berdasarkan Provinsi. In *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)* (Vol. 1, No. 1).
- Wati, M., Hatta, H. R., Bahtiar, A. A., Ervan, M. G. K., Gading, F. R. N., Nuzulan, A., & Sembiring, W. H. S. (2023). Pelatihan Pemrograman Dasar Python Sebagai Upaya Meningkatkan Minat Siswa di Bidang Informatika. *Inovasi Teknologi Masyarakat (INTEKMAS)*, 1(2), 46-51.
- Yulianto, T., Rahmah, A. F., Faisol, F., & Amalia, R. (2023). Clustering Daerah Bencana Alam di Indonesia Menggunakan Metode Fuzzy C-Means. *UJMC (Unisda Journal of Mathematics and Computer science)*, 9(2), 29-39.