Prosiding Seminar Nasional Teknologi dan Sistem Informasi (SITASI) 2026
Surabaya, 3 Juni 2026

KLASTERISASI WILAYAH RAWAN CUACA EKSTREM DI JAWA TIMUR BERBASIS K-MEANS

K-MEANS BASED CLUSTERING OF EXTREME WEATHER-PRONE AREAS IN EAST JAVA

Eka Asa Setyaning Pratiwi[1*], Alya Maytsa Ismawardi[1], Muhammad Alfian Handi Yudha[1] , Leilani Najma Rachmawati[1]
*E-mail: 23082010113@student.upnjatim.ac.id

[1]Sistem Informasi, Fakultas Ilmu Komputer, UPN "Veteran" Jawa Timur

Abstrak

Peningkatan frekuensi peristiwa cuaca ekstrem di Jawa Timur telah meningkatkan risiko bencana hidrometeorologi, sehingga diperlukan pemetaan kerawanan wilayah yang objektif untuk tujuan mitigasi. Penelitian ini bertujuan untuk mengklasifikasikan tingkat kerentanan terhadap cuaca ekstrem di 38 kabupaten dan kota di Jawa Timur melalui teknik *data mining* dengan algoritma *K-Means*. Data cuaca historis yang mencakup karakteristik curah hujan, kecepatan angin, suhu, dan kelembapan diperoleh langsung dari portal data Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) untuk periode Mei 2024 hingga Mei 2026. *Elbow Method* digunakan untuk menilai jumlah klaster yang optimal, sedangkan metrik *Silhouette Score* digunakan untuk memvalidasi kualitas struktur kluster. Hasil uji menunjukkan bahwa jumlah klaster optimal adalah tiga, didukung oleh validasi *Silhouette Score* sebesar 0,6477, yang menandakan struktur klaster yang tergolong kuat. Pendekatan pemodelan ini membagi wilayah menjadi 17 wilayah di Zona Aman dengan curah hujan terendah, 8 lokasi di Zona Waspada dengan karakteristik menengah, dan 13 area di Zona Rawan yang ditandai dengan curah hujan dan kelembapan tertinggi. Penelitian ini menyimpulkan bahwa algoritma *K-Means* secara efektif mentransformasi data meteorologi menjadi informasi zonasi risiko secara objektif. Data pemetaan yang dihasilkan dapat memberikan dasar analitis bagi pemerintah daerah untuk mengembangkan kebijakan dan distribusi sumber daya yang lebih terukur guna mitigasi bencana hidrometeorologi.

Kata kunci*: *k-means clustering, cuaca ekstrem, mitigasi bencana, jawa timur.

Abstract

The rising incidence of extreme weather events in East Java has heightened the risk of hydrometeorological disasters, necessitating objective regional vulnerability mapping for mitigation purposes. This study seeks to categorize the vulnerability levels to extreme weather in 38 regencies and cities in East Java via data mining techniques, namely the K-Means algorithm. Historical weather datasets encompassing rainfall, wind speed, temperature, and humidity characteristics were directly obtained from the Indonesian Meteorological, Climatological, and Geophysical Agency (BMKG) data online for the period spanning May 2024 to May 2026. The Elbow Method assessed the optimal cluster quantity, while the Silhouette Score metric validated the cluster structure's quality. The test findings indicated that the optimal number of clusters was three, supported by a validation Silhouette Score of 0.6477, signifying a robust cluster structure. The modeling approach divides the region into 17 regions in the Safe Zone with minimal rainfall, 8 places in the Caution Zone with moderate conditions, and 13 areas in the High-Risk Zone characterized by the highest amounts of rainfall and humidity. This study concludes that the K-Means algorithm effectively converts meteorological data into risk zoning information objectively. The resultant mapping data can provide an analytical basis for local governments to

develop more quantifiable policy and resource distributions for hydrometeorological catastrophe mitigation.

***Keywords**: k-means clustering, extreme weather, disaster mitigation, east java.*

1. PENDAHULUAN

Anomali cuaca yang dipicu oleh perubahan iklim telah secara signifikan meningkatkan intensitas cuaca ekstrem [3]. Di Indonesia, bencana hidrometeorologi, termasuk banjir, tanah longsor, dan angin puting beliung, terus mendominasi statistik bencana nasional setiap tahun [9]. Provinsi Jawa Timur, yang memiliki topografi yang kompleks yang mencakup dataran rendah dan perbukitan, termasuk di antara wilayah yang paling rentan terhadap dampak cuaca ekstrem [6]. Meningkatnya frekuensi dan intensitas bencana-bencana ini menuntut strategi mitigasi yang proaktif, sistematis, dan didasarkan pada bukti ilmiah.

Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Indonesia secara konsisten telah mendokumentasikan dan menyebarluaskan data historis mengenai parameter cuaca, seperti curah hujan, suhu, kelembapan, dan kecepatan angin. Namun, besarnya volume data observasi tersebut seringkali belum dimanfaatkan secara maksimal untuk mengekstraksi pola tersembunyi terkait kerawanan suatu wilayah [1]. Tanpa adanya pemetaan zonasi yang objektif, distribusi sumber daya dan pengembangan kebijakan manajemen bencana di tingkat kabupaten atau kota dapat menjadi tidak akurat. Oleh karena itu, teknik *data mining* untuk mengolah data meteorologi yang luas menjadi informasi analitis yang strategis sangat diperlukan.

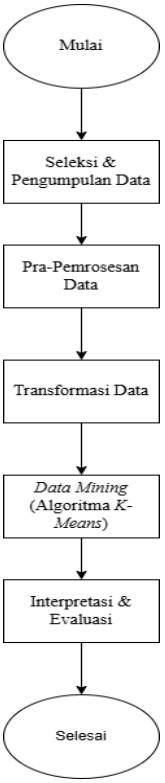
Algoritma K-Means merupakan salah satu algoritma pembelajaran mesin tanpa pengawasan (*unsupervised learning*) yang terbukti tangguh untuk mengidentifikasi kelompok data dengan karakteristik serupa [5]. Algoritma ini banyak digunakan dalam penelitian klimatologi karena efisiensi komputasinya dan kemampuannya untuk menghasilkan partisi kelompok secara objektif, bebas dari bias label sebelumnya [7]. Untuk menjamin integritas matematis dari hasil pengelompokan ini, implementasi K-Means harus dilengkapi dengan metode optimisasi dan validasi. *Elbow Method* diakui keefektifannya secara akademis dalam menentukan jumlah klaster (k) yang paling optimal, sementara metrik *Silhouette Score* secara kuantitatif menilai tingkat kohesi internal dan pemisahan antar klaster [4], [2].

Penelitian ini bertujuan untuk mengelompokkan wilayah-wilayah yang rentan terhadap cuaca ekstrem di Jawa Timur dengan menggunakan pendekatan K-Means, mengingat urgensi situasi dan keterbatasan dalam pemanfaatan data yang tersedia. Pendekatan ini memanfaatkan data yang terbaru dari BMKG dengan rentang waktu 2 tahun terakhir (Mei 2024 - Mei 2026) untuk mengklasifikasikan 38 kabupaten dan kota ke dalam zona risiko yang dapat diukur (aman, waspada, dan rentan). Hasil penelitian ini diharapkan dapat memberikan wawasan praktis sebagai landasan analitis bagi pemerintah daerah dan pemangku kepentingan terkait dalam mengembangkan program mitigasi bencana hidrometeorologi yang lebih efektif dan terarah.

2. METODOLOGI

Penelitian ini merupakan penelitian kuantitatif yang menggunakan metodologi deskriptif analitis melalui teknik *data mining*. Penelitian ini mengadopsi kerangka kerja dengan standar *Knowledge Discovery in Databases* (KDD) untuk memecahkan permasalahan pengelompokan wilayah rawan cuaca ekstrem secara sistematis. Penggunaan arsitektur KDD ini dipilih karena pendekatan yang lengkap dan sistematisnya dalam mengekstraksi informasi, pola, atau pengetahuan yang signifikan dari kumpulan data yang luas. KDD memastikan bahwa proses penemuan pola tidak hanya memprioritaskan implementasi algoritma, tetapi juga persiapan data mentah yang cermat, sehingga menghasilkan interpretasi hasil yang objektif. Prosedur ini dimulai dengan pemilihan dan pengumpulan data (1), dilanjutkan dengan pra-pemrosesan data untuk memperbaiki ketidakkonsistenan (2), dan diakhiri dengan transformasi data untuk memenuhi persyaratan

pemodelan (3). Setelah persiapan data selesai, tahap utama berupa proses *data mining* dilakukan dengan menggunakan algoritma *K-Means* (4). Siklus ini kemudian diakhiri dengan fase evaluasi dan interpretasi (5) untuk menilai kualitas pengelompokan dan mengubah data matematis menjadi informasi yang mudah dipahami mengenai kerentanan bencana. Seluruh alur penelitian ini direpresentasikan pada Gambar 1.



Gambar 1 Flowchart Penelitian

2.1 Seleksi dan Pengumpulan Data (*Data Selection*)

Tahap awal penelitian ini mencakup identifikasi sumber data yang relevan melalui portal resmi Data Online Pusat Database Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) [8]. Berdasarkan ketentuan operasional portal yang menyediakan aksesibilitas riwayat data untuk rentang waktu dua tahun terakhir, dataset yang diekstraksi dalam mencakup periode pengamatan mulai dari tanggal 06 Mei 2024 hingga 06 Mei 2026. Mengingat keterbatasan jumlah stasiun pengamatan yang tersedia di setiap wilayah administratif, penelitian ini menggunakan pendekatan Stasiun Referensi. Data diekstraksi dari tujuh stasiun utama di Jawa Timur, yaitu Stasiun Meteorologi Juanda, Stasiun Meteorologi Tuban, Stasiun Meteorologi Trunojoyo, Stasiun Meteorologi Banyuwangi, Stasiun Klimatologi Jawa Timur, Stasiun Geofisika Nganjuk dan Stasiun Geofisika Malang. Data dari ketujuh stasiun tersebut kemudian diasosiasikan untuk merepresentasikan 38 wilayah administratif kabupaten dan kota di sekitarnya berdasarkan kedekatan geografis.

Proses akuisisi data dilakukan secara bertahap melalui teknik *batch downloading*. Pada saat akuisisi data awal, parameter yang diunduh mencakup variasi nilai ekstrem harian, seperti Curah Hujan (RR), Kecepatan Angin Maksimum (FF_X), Kecepatan Angin Rata-Rata (FF_AVG), Kelembapan Rata-Rata (RH_AVG), Temperatur Maksimum (TX), Temperatur Minimum (TN), dan Temperatur Rata-Rata (TAVG). Namun, pada tahap seleksi fitur (*feature selection*), penelitian ini memutuskan untuk memfokuskan variabel masukan (*input variables*) pada nilai rata-rata (*average/mean*) serta akumulasi curah hujan.

2.2 Pra-Pemrosesan Data (Data Pra-Processing)

Tahap komputasi dieksekusi menggunakan pustaka Pandas dan NumPy pada Python. Mengingat format data yang terpisah dan adanya *missing values* (seperti kode anomali observasi BMKG), pra-pemrosesan dilakukan melalui dua tahap utama:

[16] Integrasi dan Pembersihan: Penggabungan data historis bulanan serta penangan nilai yang hilang melalui teknik imputasi menggunakan nilai rata-rata harian stasiun.

[17] Pemetaan dan Agregasi: Data dari tujuh stasiun diduplikasi untuk mengisi 38 entitas kabupaten/kota. Selanjutnya, dilakukan agregasi *time-series* menjadi satu nilai profil tunggal, dengan menggunakan rata-rata aritmetika untuk TAVG, RH_AVG, FF_AVG, dan akumulasi rata-rata untuk RR.

2.3 Transformasi Data

Tahap transformasi dilakukan untuk menstandarkan skala pengukuran semua variabel fitur yang membentuk matriks berdimensi 38 baris dan 4 kolom sebelum diproses oleh algoritma. Keempat parameter cuaca memiliki rentang nilai yang sangat bervariasi. Untuk mencegah dominasi variabel berskala besar, seperti curah hujan, terhadap variabel berskala kecil dalam perhitungan jarak matematis, diterapkan pendekatan normalisasi *Min-Max Scaler* untuk menstandarkan seluruh nilai data ke dalam rasio seragam dalam rentang 0 hingga 1. Persamaan untuk transformasi ini adalah persamaan (1).

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad \text{.....(1)}$$

Di mana X_{norm} adalah nilai hasil normalisasi, X adalah nilai atribut asli, X_{min} adalah nilai minimum dari atribut, dan X_{max} adalah nilai maksimum dari atribut terkait.

2.4 Data Mining (Algoritma K-Means)

Pemodelan klasterisasi dilakukan dengan menggunakan algoritma *K-Means*, yaitu sebuah metode *unsupervised learning* berbasis partisi k = 3 yang mengelompokkan wilayah berdasarkan kemiripan karakteristik cuacanya. Algoritma ini meminimalkan jarak antara titik-titik data suatu wilayah dengan pusat klaster (*centroid*). Pengukuran kedekatan tersebut dikalkulasi menggunakan formula *Euclidean Distance* dengan persamaan (2).

$$d(\chi, c) = \sqrt{\sum_{i=1}^{n} (\chi_i - c_i)^2} \quad \text{.....(2)}$$

Di mana $d(\chi, c)$ melambangkan jarak dari objek data χ ke titik pusat klaster c, n adalah jumlah variabel fitur cuaca, χ_i adalah nilai fitur ke-i pada data uji, dan c_i adalah nilai fitur ke-i pada pusat klaster. Proses pembaruan posisi *centroid* ini dilakukan secara iteratif hingga mencapai titik konvergen.

2.4 Interpretasi dan Evaluasi

Tahap terakhir dari kerangka kerja KDD adalah mengevaluasi kualitas model dan memberikan label klaster. Algoritma *K-Means* memerlukan inisialisasi jumlah kluster (k) di awal proses. Oleh karena itu, *Elbow Method* digunakan untuk menentukan nilai k yang optimal dengan mengidentifikasi penurunan paling drastis pada grafik inersia yang membentuk titik siku.

Setelah pembentukan klaster, kualitas struktur dievaluasi menggunakan *Silhouette Score* untuk menentukan tingkat kohesi internal di dalam klaster dan perbedaan antara klaster yang berbeda. Evaluasi ini dihitung melalui persamaan (3).

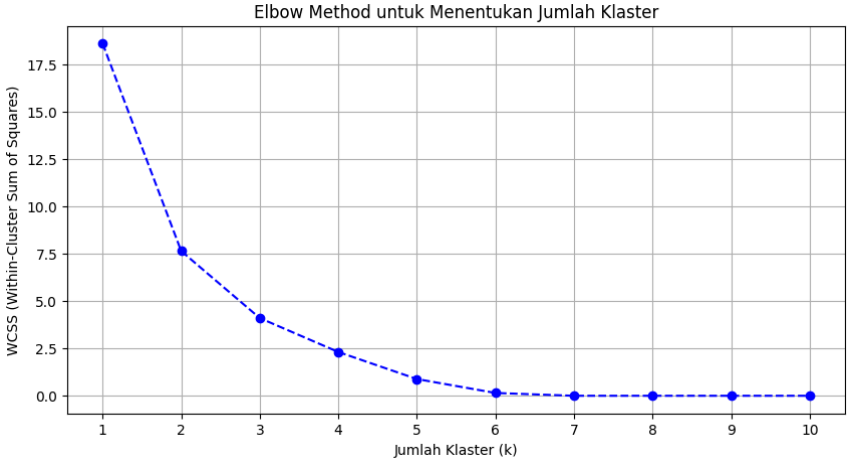
$$S(i) = \frac{b(i) - a(i)}{max(a(i), b(i))} \quad \text{.....(3)}$$

Di mana $a(i)$ adalah jarak rata-rata suatu objek ke objek lain dalam klaster yang sama, dan $b(i)$ adalah jarak rata-rata objek tersebut ke klaster tetangga terdekat. Hasil pengelompokan yang tervalidasi kemudian diinterpretasikan melalui *profiling* atribut *centroid* untuk menetapkan pelabelan zonasi tingkat kerawanan cuaca ekstrem secara spesifik, yakni Zona Rawan, Zona Waspada, dan Zona Aman.

3. HASIL DAN PEMBAHASAN

3.1 Penentuan Jumlah Klaster Optimal

Tahap awal dalam implementasi pemodelan *K-Means* adalah menentukan jumlah klaster (k) yang paling representatif untuk dataset cuaca historis yang telah diproses. Evaluasi komputasi dilakukan dengan menggunakan *Elbow Method* dengan rentang pengujian nilai k dalam rentang 1 hingga 10. Pengukuran ini diperoleh dari perhitungan nilai *Within-Cluster Sum of Squares* (WCSS) atau inersia pada setiap iterasi untuk mengevaluasi tingkat variansi di dalam klaster.



Gambar 2. Grafik *Elbow Method* untuk Penentuan Nilai k

Berdasarkan Gambar 2, terlihat adanya penurunan yang signifikan dan tajam pada nilai WCSS selama iterasi awal, yakni dari $k = 1$ hingga titik $k = 3$. setelah melewati titik $k = 3$, kurva menurun mulai melandai perlahan, membentuk pola siku (*elbow*). Berdasarkan visual pada grafik tersebut, jumlah klaster optimal yang ditetapkan untuk membagi 38 wilayah administratif di Jawa Timur adalah sebanyak 3 kelompok. Nilai ini sejalan dengan tujuan fungsional penelitian untuk mengklasifikasikan wilayah ke dalam tiga tingkatan zonasi kerawanan, yaitu Aman, Waspada, dan Rawan.

Grafik inersia WCSS terlihat konvergen dan menyentuh nilai nol (0) pada $k = 7$. Fenomena ini mengonfirmasi validitas prosedur persiapan data sebelumnya, di mana profil cuaca dari 38 kabupaten/kota tersebut pada dasarnya diperoleh dari pemetaan (duplikasi) 7 stasiun pengamatan BMKG referensi. Teknik *K-Means* secara efektif mengidentifikasi 7 entitas profil pengamatan yang berbeda dalam dataset, sehingga menghasilkan variansi nol di dalam klaster ketika jumlah klaster ditetapkan menjadi 7.

3.2 Validasi Klaster

Penelitian ini melakukan uji validasi internal dengan menggunakan metrik *Silhouette Score* untuk memastikan kualitas dan ketepatan pemisahan di antara ketiga klaster yang teridentifikasi. Pengujian ini ini mengevaluasi secara menyeluruh tingkat kohesi (kedekatan karakteristik antar objek di dalam satu klaster yang sama) dan tingkat pemisahan (jarak karakteristik objek tersebut terhadap klaster tetangga terdekatnya).

Berdasarkan hasil komputasi pada model dengan jumlah klaster $k = 3$, diperoleh nilai rata-rata koefisien *Silhouette Score* sebesar 0,6477. Nilai koefisien positif, yang melampaui kriteria 0,5 menunjukkan bahwa struktur partisi klaster yang dihasilkan sudah tergolong kuat (*strong structure*). Hal ini menandakan bahwa 38 wilayah kabupaten dan kota telah ditempatkan pada kategori kerawanan cuaca yang relevan, dengan batas pemisah (*decision boundary*) antar zona klaster yang sangat jelas dan tingkat tumpang tindih (*overlap*) yang minim.

3.3 Hasil Klasterisasi Wilayah

Berdasarkan pemrosesan algoritma *K-Means* dengan parameter optimal $k = 3$, 38 wilayah administrasi kabupaten dan kota di Provinsi Jawa Timur berhasil dipartisi ke dalam tiga kelompok kerawanan yang berbeda. Hasil pemetaan wilayah ke dalam masing-masing klaster ditentukan oleh kemiripan metrik cuaca, yaitu curah hujan, suhu, kelembapan, dan kecepatan angin. Rincian pembagian kabupaten dan kota untuk setiap klaster disajikan secara komprehensif pada Tabel 1.

Tabel 1. Distribusi Anggota Klaster Wilayah di Jawa Timur

Klaster	Jumlah Wilayah	Kabupaten/Kota
Klaster 0	8	Kota Malang, Kabupaten Malang, Kota Batu, Kota Pasuruan, Kabupaten Pasuruan, Kota Probolinggo, Kabupaten Probolinggo, Kabupaten Lumajang
Klaster 1	17	Kabupaten Banyuwangi, Kabupaten Jember, Kabupaten Situbondo, Kabupaten Bondowoso, Kota Surabaya, Kabupaten Sidoarjo, Kabupaten Gresik, Kota Mojokerto, Kabupaten Mojokerto, Kabupaten Jombang, Kabupaten Bangkalan, Kabupaten Sampang, Kabupaten Sumenep, Kabupaten Tuban, Kabupaten Bojonegoro, Kabupaten Lamongan
Klaster 2	13	Kota Kediri, Kabupaten Kediri, Kota Blitar, Kabupaten Blitar, Kabupaten Tulungagung, Kabupaten Trenggalek, Kabupaten Pacitan, Kabupaten Ponorogo, Kabupaten Nganjuk, Kota Madiun, Kabupaten Madiun, Kabupaten Ngawi, Kabupaten Magetan

Berdasarkan data dalam Tabel 1, terlihat bahwa distribusi kelompok cuaca di Jawa Timur tidak merata. Klaster 1 mendominasi klasifikasi dengan menaungi 17 wilayah (hampir separuh dari total wilayah observasi). Hal ini menunjukkan bahwa sebagian besar wilayah di Jawa Timur memiliki tren karakteristik tren karakteristik cuaca harian yang seragam selama periode pengamatan.
Di sisi lain, Klaster 0 terdiri dari 8 wilayah dan Klaster 2 terdiri dari 13 wilayah, keduanya memiliki jumlah anggota yang lebih sedikit. Keanggotaan yang lebih sedikit dalam kedua klaster ini menandakan kondisi meteorologi yang lebih spesifik, di mana wilayah-wilayah ini menunjukkan anomali atau penyimpangan cuaca yang seringkali lebih ekstrem, baik intensitas curah hujan yang jauh lebih baik maupun lebih rendah dibandingkan dengan wilayah lain di Jawa Timur. Klasifikasi ekstremitas untuk ketiga klaster ini akan dibahas lebih lanjut dalam profil karakteristik setiap pusat klaster.

3.4 Karakteristik dan Profiling Klaster

Setelah distribusi wilayah pada setiap klaster teridentifikasi, langkah penting berikutnya adalah melakukan analisis komprehensif terhadap karakteristik fisik dari masing-masing kelompok. Tahap

profiling ini dilakukan dengan menganalisis nilai pusat klaster (*centroid*) untuk setiap parameter fitur input pasca pemrosesan. Analisis ini bertujuan untuk mengidentifikasi tren meteorologi dominan yang membedakan satu klaster dengan klaster lainnya, sehingga memungkinkan transformasi objektif dari data matematis menjadi informasi zonasi kerawanan cuaca ekstrem. Ringkasan nilai *centroid* dan hasil pelabelan interpretasi zonasi untuk ketiga klaster disajikan pada Tabel 2.

Tabel 2. Nilai Pusat Klaster (Centroid) dan Interpretasi Zona

Klaster	Curah Hujan (RR)	Suhu (TAVG)	Kelembapan (RH)	Kecepatan Angin (FF)	Interpretasi Zona
1	1883,85	28,19	78,36	2,19	Zona Aman
0	2124,00	24,50	79,18	1,98	Zona Waspada
2	2993,95	25,41	81,95	1,37	Zona Rawan

Berdasarkan observasi terhadap titik pusat pada Tabel 2, karakteristik masing-masing zona dapat digambarkan sebagai berikut:

1. Zona Aman (Klaster 1)
   Kelompok ini ditandai oleh curah hujan akumulatif terendah (1.883,85 mm) dan kelembapan relatif terendah (78,36%). Meskipun memiliki kecepatan angin rata-rata tertinggi, ketidakstabilan karakteristik curah hujan menyebabkan kelompok ini memiliki probabilitas terendah untuk bencana hidrometeorologi dibandingkan dengan klaster lain.
2. Zona Waspada (Klaster 0)
   Kelompok ini menunjukkan karakteristik parameter menengah. Intensitas curah hujan sebesar 2.124,00 mm, disertai dengan suhu harian rata-rata terendah sebesar 24,50°C. Wilayah di dalam zona ini diklasifikasikan memiliki risiko menengah, sehingga memerlukan pemantauan terhadap potensi anomali cuaca yang fluktuatif.
3. Zona Rawan (Klaster 2)
   Kelompok ini mencakup wilayah-wilayah paling kritis, yang ditandai dengan nilai *centroid* curah hujan tertinggi sebesar 2.993,95 mm dan kelembapan relatif tertinggi sebesar 81,95%. Karakteristik ini menandakan kerentanan yang tinggi terhadap ancaman bencana hidrometeorologi, seperti banjir dan tanah longsor, sehingga memerlukan prioritas utama dalam manajemen risiko bencana.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Penelitian ini telah secara efektif menerapkan algoritma *K-Means* untuk mengklasifikasikan wilayah-wilayah yang rentan terhadap cuaca ekstrem di Provinsi Jawa Timur dalam kerangka *Knowledge Discovery in Databases* (KDD). berdasarkan analisis data iklim yang mencakup periode Mei 2024 hingga Mei 2026, dapat disimpulkan beberapa kesimpulan utama sebagai berikut:

12. Implementasi *Elbow Method* menunjukkan bahwa jumlah klaster yang optimal untuk mengelompokkan 38 kabupaten/kota di Jawa Timur adalah $k = 3$. Hasil ini didukung secara matematis melalui pengujian *Silhouette Score* yang menghasilkan nilai koefisien sebesar 0,6477, yang menandakan bahwa adanya kohesi dan pemisahan yang kuat dalam struktur klaster yang dihasilkan.
13. Proses klasterisasi menghasilkan tiga zona risiko cuaca ekstrem yang berbeda: Zona Aman (Klaster 1) mencakup 17 wilayah dengan curah hujan terendah, Zona Waspada (Klaster 0) yang terdiri dari 8 wilayah dengan karakteristik meteorologi menengah, serta Zona Rawan

(Klaster 2) mencakup 13 wilayah yang memiliki akumulasi curah hujan serta tingkat kelembapan udara tertinggi.

14. Penggunaan teknik *data mining* dengan menggunakan algoritma *K-Means* telah menunjukkan keefektifannya dalam mentransformasi data historis BMKG yang rumit menjadi informasi zona risiko yang objektif. Hasil pemetaan ini dapat berfungsi sebagai dasar analitis bagi pemerintah daerah dalam merumuskan rencana mitigasi bencana hidrometeorologi yang lebih terukur dan tepat sasaran.

4.2 Saran

Beberapa saran yang dapat diberikan untuk pengembangan penelitian selanjutnya antara lain:

1. Penambahan variabel meteorologi lain, seperti tekanan atmosfer atau durasi sinar matahari, untuk memperkaya dimensi profil karakteristik cuaca di setiap wilayah.
2. Perluasan rentang waktu dataset (*time-frame*) untuk menangkap pola anomali cuaca dan tren perubahan iklim jangka panjang secara lebih mendalam.
3. Eksperimentasi menggunakan algoritma klasterisasi lain atau penggabungan dengan teknik optimasi tertentu untuk mengevaluasi ketepatan dan keefektifan pengelompokan wilayah rawan bencana.

5. DAFTAR RUJUKAN

[1] A. Sofro, R. A. Riani, K. N. Khikmah, R. W. Romadhonia, dan D. Ariyanto, "ANALYSIS OF RAINFALL IN INDONESIA USING A TIME SERIES-BASED CLUSTERING APPROACH," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 18, no. 2, hlm. 0837–0848, Mei 2024, doi: 10.30598/barekengvol18iss2pp0837-0848.

[2] R. Sudrajat, A. I. Hadiana, dan M. Melina, "Evaluasi Kualitas Klaster Wilayah Rawan Bencana Menggunakan K-Means dengan Silhouette dan Elbow Method | Jurnal Algoritma", Diakses: 9 Mei 2026. [Daring]. Tersedia pada: https://jurnal.itg.ac.id/index.php/algoritma/article/view/2379

[3] B. Clarke, F. Otto, R. Stuart-Smith, dan L. Harrington, "Extreme weather impacts of climate change: an attribution perspective," *Environ. Res.: Climate*, vol. 1, no. 1, hlm. 012001, Sep 2022, doi: 10.1088/2752-5295/ac6e7d.

[4] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, dan B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 336, no. 1, hlm. 012017, Apr 2018, doi: 10.1088/1757-899X/336/1/012017.

[5] D. Lou *dkk.*, "K-Means and C4.5 Decision Tree Based Prediction of Long-Term Precipitation Variability in the Poyang Lake Basin, China," *Atmosphere*, vol. 12, no. 7, hlm. 834, Jul 2021, doi: 10.3390/atmos12070834.

[6] M. K. Adinata, A. Mahmudi, dan Y. A. Pranoto, "Klasterisasi Daerah Rawan Bencana Alam Menggunakan Algoritma K-Means," *Infotek: Jurnal Informatika dan Teknologi*, vol. 8, no. 1, hlm. 250–260, Jan 2025, doi: 10.29408/jit.v8i1.28196.

[7] F. Faisal, L. A. G. Giopani, M. F. Fitriah, Z. C. D. Dwynne, S. S. H. Helma, dan M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Suhu di Provinsi Riau : Comparison of K-Means and K-Medoids Algorithms for Temperature Grouping in Riau Province," *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, vol. 2, no. 2, hlm. 128–134, Sep 2022, doi: 10.57152/ijirse.v2i2.434.

[8] Badan Meteorologi, Klimatologi, dan Geofisika, "Data Online BMKG," 2026. [Online]. Available: https://dataonline.bmkg.go.id/ . [Accessed: 07-May-2026].

[9] Badan Nasional Penanggulangan Bencana, Buku Data Bencana Indonesia 2024, 2025. [Online]. Available: https://bnpb.go.id/storage/app/media/Buletin%20Info%20Bencana/Buku%20Data%20Bencana%202024/20250613_Buku%20Data%202024.pdf. [Accessed: 09-May-2026].