

PENGGUNAAN ALGORITMA CLUSTERING K-MEANS UNTUK MENGELOMPOKKAN POLA CUACA

Ilham Rinaldi¹, Harmayani²

Universitas Asahan

e-mail: ¹ilhamrinaldi57@gmail.com , ²mayong3010@gmail.com

Abstract: *Weather pattern analysis is crucial for enhancing disaster preparedness and mitigation in Indonesia, particularly in the face of climate change and extreme weather conditions. This study applies the K-Means algorithm to cluster daily weather data obtained from the Meteorology, Climatology, and Geophysical Agency (BMKG). Using average temperature (Tavg) and average humidity (RH_avg) as variables, the data were analyzed to identify two primary weather patterns: "rainy weather" and "cool humid weather." The clustering results are visualized in a web-based application designed to facilitate interpretation for the general public. The implementation of the K-Means algorithm in this application provides clearer insights into daily weather variations and has the potential to support planning and disaster mitigation in Indonesia.*

Keywords: *Weather, K-Means, weather patterns, disaster mitigation, web-based application.*

Abstrak: Analisis pola cuaca sangat penting untuk mendukung kesiapsiagaan dan mitigasi bencana di Indonesia, terutama dalam menghadapi perubahan iklim dan cuaca ekstrem. Penelitian ini menerapkan algoritma K-means untuk mengelompokkan data cuaca harian yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Dengan menggunakan variabel suhu rata-rata (Tavg) dan kelembapan rata-rata (RH_avg), data dianalisis untuk mengidentifikasi dua pola cuaca utama: ‘cuaca hujan’ dan ‘cuaca dingin lembab’. Hasil pengelompokan ini divisualisasikan dalam aplikasi berbasis web yang dirancang untuk mempermudah interpretasi bagi masyarakat umum. Implementasi algoritma K-means dalam aplikasi ini memberikan gambaran yang lebih jelas mengenai cuaca harian, serta berpotensi mendukung perencanaan dan mitigasi bencana di Indonesia.

Kata kunci: Cuaca, K-means, pola cuaca, mitigasi bencana, aplikasi berbasis web.

PENDAHULUAN

Perubahan iklim dan variabilitas cuaca menjadi isu yang semakin penting dalam beberapa tahun terakhir. Data cuaca yang akurat dan analisis yang mendalam sangat diperlukan untuk memahami pola cuaca dan dampaknya terhadap kehidupan manusia. Di Indonesia, yang merupakan negara dengan keragaman iklim dan cuaca, pemantauan dan analisis data cuaca dilakukan oleh Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Data yang diperoleh dari BMKG mencakup berbagai parameter cuaca,

seperti suhu, kelembapan, curah hujan, dan kecepatan angin, yang sangat penting untuk perencanaan dan mitigasi bencana alam.

Pentingnya analisis data cuaca juga tercermin dalam dampaknya terhadap sektor ekonomi. Misalnya, sektor pertanian sangat bergantung pada pola cuaca yang stabil. Perubahan yang tiba-tiba dalam suhu atau curah hujan dapat memengaruhi hasil panen, yang pada gilirannya berdampak pada ketahanan pangan. Penelitian menunjukkan bahwa pengelolaan risiko cuaca yang baik dapat meningkatkan produktivitas pertanian dan mengurangi kerugian ekonomi (Nurina

Sari et al., 2023). Oleh karena itu, pemahaman yang lebih baik tentang pola cuaca melalui data analisis sangat diperlukan.

Penggunaan algoritma clustering, seperti K-Means, dalam analisis data cuaca dapat membantu dalam mengidentifikasi pola dan tren yang tidak terlihat dengan jelas. Menurut (Supardi & Kanedi, 2020) *K-Means Clustering* adalah metode yang digunakan untuk menemukan dan mengelompokkan data berdasarkan kesamaan karakteristik (*similarity*) antara satu data dengan data lainnya. Dengan menggunakan K-Means, data cuaca dapat dikelompokkan untuk menemukan pola yang serupa di berbagai lokasi atau waktu yang berbeda.

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah: Apa saja pola cuaca yang dapat diidentifikasi dari data harian BMKG menggunakan algoritma K-Means? Bagaimana efektivitas algoritma K-Means dalam mengelompokkan data cuaca dibandingkan dengan metode analisis lainnya? Apa implikasi hasil pengelompokan pola cuaca terhadap perencanaan dan mitigasi bencana di Indonesia?.

Penelitian ini memiliki beberapa batasan, antara lain: Penelitian ini dibatasi pada data cuaca dari periode Maret 2024 hingga April 2024. Data cuaca historis sebelum atau setelah periode tersebut tidak termasuk dalam analisis. Penelitian ini terbatas pada penggunaan algoritma K-Means untuk melakukan *clustering*, tanpa membandingkan dengan algoritma *machine learning* lainnya, meskipun evaluasi dasar terhadap efektivitasnya akan dilakukan. Analisis akan difokuskan pada satu lokasi stasiun cuaca dari BMKG, yaitu Balai Besar Meteorologi Klimatologi dan Geofisika Wilayah I, untuk mengidentifikasi pola cuaca yang spesifik dan relevan bagi daerah tersebut.

Adapun tujuan penelitian ini adalah sebagai berikut : mengidentifikasi pola cuaca yang ada dari data harian BMKG menggunakan algoritma K-Means, sehingga dapat diketahui karakteristik

cua di wilayah tersebut. Penelitian ini bertujuan untuk mengevaluasi efektivitas algoritma K-Means dalam mengelompokkan data cuaca dibandingkan dengan metode analisis yang umum digunakan, untuk menentukan kehandalan dan akurasi hasilnya. Penelitian ini bertujuan untuk memberikan rekomendasi berdasarkan hasil pengelompokan pola cuaca yang dapat digunakan untuk perencanaan dan mitigasi bencana di Indonesia, dengan fokus pada peningkatan kesiapsiagaan masyarakat terhadap perubahan cuaca ekstrem.

Penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan ilmu meteorologi dan klimatologi, khususnya dalam penerapan teknik *machine learning* untuk analisis data cuaca, serta memperkaya referensi akademis di bidang tersebut. Hasil penelitian dapat digunakan oleh pemerintah dan instansi terkait untuk meningkatkan kesiapsiagaan dan respons terhadap bencana cuaca, dengan cara memahami pola cuaca yang berpotensi memengaruhi masyarakat. Penelitian ini dapat memberikan informasi yang berguna bagi sektor pertanian dan ekonomi, sehingga para petani dan pelaku bisnis dapat mengambil keputusan yang lebih baik berdasarkan prediksi pola cuaca yang dihasilkan dari analisis data.

METODE

Cuaca adalah kondisi atmosfer yang terjadi dalam periode waktu pendek di suatu tempat, mencakup parameter seperti suhu, kelembapan, curah hujan, dan tekanan udara. Sebaliknya, iklim adalah pola cuaca yang teramati dalam jangka waktu yang lebih panjang, biasanya selama 30 tahun atau lebih. Pemahaman yang mendalam tentang perbedaan ini sangat penting, terutama dalam konteks perubahan iklim yang memengaruhi pola cuaca di seluruh dunia (Suhartono, 2021). Perubahan iklim juga berakibat pada fenomena cuaca ekstrem,

seperti peningkatan frekuensi dan intensitas hujan, yang dapat menyebabkan banjir. Oleh karena itu, penting untuk memahami hubungan antara cuaca dan iklim dalam konteks perencanaan dan adaptasi terhadap perubahan iklim. Penelitian yang dilakukan di berbagai wilayah di Indonesia menunjukkan bahwa pemodelan cuaca dan iklim yang tepat dapat membantu dalam meramalkan kejadian ekstrem dan mengurangi dampaknya pada Masyarakat.

Menurut (Ikhwan & Aslami, 2020), Data mining merupakan sebuah proses pengumpulan serta analisis data historis guna mengidentifikasi keteraturan, pola, dan hubungan di dalam kumpulan data yang besar. Proses ini melibatkan penggunaan teknik statistik, algoritma pembelajaran mesin, dan analisis data untuk mengekstrak informasi yang berharga. Dalam era digital saat ini, data mining berperan penting dalam berbagai bidang, termasuk pemasaran, kesehatan, dan analisis sosial, di mana pengambilan keputusan berbasis data menjadi semakin umum.

Dalam proses *clustering*, data yang serupa akan dikelompokkan bersama, sedangkan data yang berbeda akan ditempatkan di kelompok yang terpisah. Menurut (Nugroho & Wahyuni, 2022), ada berbagai algoritma *clustering* yang dapat digunakan, termasuk K-Means, *Hierarchical Clustering*, dan DBSCAN. Setiap algoritma memiliki pendekatan dan keunggulan tersendiri. Misalnya, K-Means adalah salah satu algoritma yang paling populer karena kesederhanaan dan efisiensinya dalam mengelompokkan data besar. Algoritma ini bekerja dengan menentukan jumlah *cluster* yang diinginkan dan mengelompokkan data berdasarkan jarak terdekat ke pusat *cluster* (*centroid*).

Website: Situs web akan digunakan sebagai platform untuk menampilkan hasil pengelompokan pola cuaca dan informasi terkait. Menurut Ariyani et al. (2023) dan Satya Saputra et al. (2023), website merupakan media yang efektif untuk menyajikan konten interaktif dan

multimedia yang dapat diakses melalui internet. Dalam penelitian ini, website dinamis akan dikembangkan untuk memfasilitasi pembaruan data cuaca secara berkala dan menampilkan analisis hasil pengelompokan cuaca secara real-time.

PHP: Dalam pengembangan situs web ini, PHP digunakan sebagai bahasa pemrograman *server-side* untuk memproses data cuaca yang diolah menggunakan algoritma K-Means. Menurut Hidayat et al. (2019), PHP merupakan bahasa yang dirancang untuk mengelola aplikasi web dinamis, memungkinkan komunikasi antara server dan pengguna secara real-time. PHP akan memungkinkan sistem untuk memperbarui dan menampilkan hasil pengelompokan pola cuaca berdasarkan data cuaca terbaru secara otomatis di halaman web.

Pengumpulan Data

Langkah 1: Mengambil data cuaca harian dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) yang mencakup suhu rata-rata dan kelembapan harian.

Langkah 2: Memastikan data yang dikumpulkan mencakup periode waktu yang relevan (seperti Maret 2024) untuk analisis pola cuaca.

Pra-pemrosesan Data

Langkah 1: Membersihkan dataset dengan menghapus data duplikat atau data yang tidak relevan.

Langkah 2: Normalisasi data untuk memastikan bahwa semua variabel memiliki skala yang sama, seperti mengonversi suhu dan kelembapan ke skala standar (z-score atau min-max scaling).

Langkah 3: Menangani nilai-nilai kosong atau hilang (*missing values*) dengan metode interpolasi atau penghapusan jika diperlukan, agar data siap untuk dianalisis.

Penentuan Fitur Utama

Langkah 1: Mengidentifikasi variabel utama yang relevan, seperti suhu rata-rata

(Tavg) dan kelembapan rata-rata (RH avg), yang memengaruhi pola cuaca.
 Langkah 2: Memilih fitur-fitur yang signifikan agar model K-means dapat mengelompokkan data dengan baik.

Penerapan Algoritma K-means

Langkah 1: Menentukan jumlah cluster optimal menggunakan metode Elbow atau Silhouette untuk mengidentifikasi titik optimal untuk jumlah cluster.

Langkah 2: Melakukan pengelompokan data dengan algoritma K-means untuk menemukan pola cuaca berdasarkan kemiripan suhu dan kelembapan.

Langkah 3: Mengelompokkan data ke dalam dua pola utama yaitu: ‘cuaca dingin lembab’ dan ‘cuaca hujan’ yang memudahkan interpretasi hasil untuk mitigasi cuaca ekstrem.

Evaluasi Kualitas Clustering

Langkah 1: Menggunakan Elbow Method untuk mengevaluasi jumlah cluster yang optimal, serta memastikan bahwa pembagian cluster menghasilkan nilai inerti 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.

HASIL DAN PEMBAHASAN

Pengumpulan Data

Pengelompokkan data cuaca dilakukan dengan algoritma K-means yang membagi data menjadi dua pola utama yaitu ‘cuaca dingin lembab’ dan ‘cuaca hujan’. Proses ini bertujuan untuk membantu pemahaman terkait kondisi cuaca yang dapat mendukung Upaya mitigasi bencana. Dalam penelitian ini, data cuaca yang digunakan adalah suhu rata-rata (Tavg) dan kelembapan rata-rata (RH_avg) dari enam sample data harian.

Data Cuaca Sample untuk Clustering
Tabel 1. Data Cuaca Sample untuk Clustering

sample data	Tavg	RH_avg
1	28.4	83
2	27.9	83
3	28.6	79
4	29.0	79
5	28.0	84
6	28.6	82

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	28.4	83
K2	27.9	83

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

$$d = \sqrt{(x1 - x2)^2 + (y1 - y2)^2} \quad (1)$$

Iterasi pertama

Pada iterasi pertama, dua *centroid* awal di pilih.

Jarak antara sample pertama (28.4,83) ke centroid cluster 1 dan cluster 2 :

- o Jarak ke cluster 1: 0
- o Jarak ke cluster 2 : 0.5

1. Pengelompokkan Data Berdasarkan Iterasi Pertama

Tabel 2. Pengelompokkan Data Iterasi Pertama

N o	x	y	Jarak 1	Jarak 2	kelompok cluster
1	28.4	83	0	0.5	1
2	27.9	83	0.5	0	2
3	28.6	79	4.01	4.06	1
4	29.0	79	4.04	4.20	1
5	28.0	84	1.08	1.0	2
6	28.6	82	1.02	1.22	1

Iterasi kedua

Pada iterasi kedua, jarak setiap sampel dihitung kembali berdasarkan *centroid* baru.

2. Perhitungan Data Berdasarkan Iterasi Kedua

Pada iterasi kedua, jarak setiap sampel dihitung kembali berdasarkan *centroid* baru:

Tabel 3. Pengelompokan Data Iterasi Kedua

No	x	y	Jarak 1	Jarak 2	Kelompok cluster
1	28.4	83	2.26	0.67	2
2	27.9	83	2.37	0.5	2
3	28.6	79	1.75	4.55	1
4	29.0	79	1.78	4.65	1
5	28.0	84	3.32	0.5	2
6	28.6	82	1.25	1.63	1

Pada iterasi kedua, hasil pengelompokan tetap stabil, menunjukkan konvergensi model. Pengelompokan akhir menunjukkan bahwa data terbagi menjadi dua pola yang stabil

- o Cluster 1 (cuaca hujan) : data sample 3, 4, dan 6
- o Cluster 2 (cuaca dingin lembab) : data sample 1, 2, dan 5

SIMPULAN

Pada penelitian ini berhasil menerapkan teknik data mining *K-means clustering* untuk mengelompokkan pola cuaca berdasarkan data harian BMKG. Hasil yang diperoleh menunjukkan adanya pola cuaca yang terbagi menjadi dua kluster utama yaitu: cuaca hujan dan cuaca dingin lembab meskipun terdapat tantangan dalam pemrosesan data cuaca yang beragam, pendekatan yang digunakan telah memberikan gambaran yang jelas mengenai karakteristik cuaca yang mendominasi di wilayah studi. Temuan ini menegaskan bahwa algoritma *K-means* dapat menjadi alat yang efektif

dalam mendukung analisis pola cuaca untuk kepentingan mitigasi bencana dan perencanaan sumber daya.

DAFTAR PUSTAKA

- Ariyani, M., Surahman, A., & Wantoro, A. (2023). Implementasi metode aida dalam pengembangan website sebagai peningkatan promosi produk makanan umkm puding hayu. 4(3), 250–261.
<https://doi.org/10.33365/jtsi>
- Aufa, A., Putri, A., & Rahmah, S. A. (2024). Implementasi data mining dengan algoritma k-means clustering untuk analisis bisnis pada perusahaan asuransi. Jurnal Teknologi Informasi, 5(1).
<https://doi.org/10.46576/djtechno>
- Dwitri, N., Tampubolon, J. A., Prayoga, S., Fikrul, P. P. P. A. N. W., Zer, I. R. H., Studi, P., Informatika, T., Tunas, S., Pematangsiantar, B., Jend, J., Sudirman Blok, N., Pematangsiantar, K., & Utara, S. (2020). Penerapan algoritma k-means dalam menentukan tingkat penyebaran pandemi covid-19 di indonesia. Jurnal Teknologi Informasi, 4(1).
- Hany, S. Samsugi, & Ari Sulistiyawati. (2023). Rancang bangun sistem informasi pendataan calon penerima bantuan sosial dan desa berbasis web (studi kasus : desa cilimus).
- Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa.
<https://doi.org/10.58602/jima-ilkom.v3i1.26>
- Hidayat, A., Yani, A., Studi Sistem Informasi, P., & Mahakarya, S. (2019). Membangun website sma PGRI gunung raya ranau menggunakan php dan mysql (Vol. 2, Issue 2).
- Ikhwan, A., & Aslami, N. (2020). Implementasi data mining untuk manajemen bantuan sosial menggunakan algoritma k-means.

- Jurnal Teknologi Informasi, 4(2).
- Inggi, R., Alam, H. P., Studi, P., Informasi, S., Bina, S., & Kendari, B. (2023). Analisis forensik web browser pada perangkat android. 8(1).
- Maori, N. A. (2023). Metode elbow dalam optimasi jumlah cluster pada k-means clustering. *Jurnal SIMETRIS*, 14.
- Nugroho, A., & Wahyuni, S. (2022). Algoritma Clustering dan Penerapannya. *Jurnal Sistem Informasi*, 8(2), 100–110.
- Nurina Sari, Betha, & Muhammad Yamin Nurzaman. (2023). Implementasi K-Means Clustering Dalam Pengelompokan Banyaknya Jumlah Petani Berdasarkan Kecamatan Di Provinsi Jawa Barat. *Jurnal Teknik Informatika Dan Sistem Informasi*, 10(3). <http://jurnal.mdp.ac.id>
- Romzi, M., & Kurniawan, B. (2020). Pembelajaran pemrograman python dengan pendekatan logika algoritma (Issue 2).
- Rusdi Oktapalisa, M., Murti, W., & Informatika dan Komputer *Jurnal Informatika dan Komputer*, J. (2022). Membuat Aplikasi Penjualan Pada CV. Sumber Bakti Mandiri Berbasis Website Menggunakan PHP dan MYSQL. In *JIK* (Vol. 13, Issue 2).
- Santos, A. Dos, Pereira, G. S., Syuhada, R. A., Malays, E., & Sakti, S. (2024). Uji Coba Keamanan Database Website Menggunakan Python Dan Sqlmap Melalui Command Prompt Pada Sistem Operasi Windows. <https://doi.org/10.37817/tekinfo.v25i1>
- Satya Saputra, P., Aditya Pratama, P., & Putu Ary Sri Tjahyanti, L. (2023). Perancangan dan komparasi web server nginx dengan web server apache serta pemanfaatan reverse proxy server pada nginx. *Jurnal Komputer Dan Teknologi Sains (KOMTEKS)*, 2(1), 16–21.
- Suhartono. (2021). Climate Change and Extreme Weather Events in Indonesia: Impacts and Adaptation Strategies. *Journal of Environmental Management*. <https://doi.org/10.1016/j.jenvman.2021.112048>
- Supardi, R., & Kanedi, I. (2020). Implementasi metode algoritma k-means clustering pada toko eidelweis. *Jurnal Teknologi Informasi*, 4(2).