

# PENERAPAN ALGORITMA K-MEANS UNTUK KLAUSTERISASI POLA IKLIM STUDI KASUS: PROVINSI JAMBI PERIODE 2020-2024

Yandi Anzari<sup>1</sup>, Yonggi Puriza<sup>2</sup>, Niko Akbar<sup>3</sup>, Nurul Abdillah<sup>4</sup>, Oki Dahwanu<sup>5</sup>, Elsi Alfionita Syawal<sup>6</sup>

1,3,4,5) Prodi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Jambi, Indonesia

2) Prodi Teknik Elektro, Fakultas Sains dan Teknik, Universitas Bangka Belitung, Indonesia

6) Prodi Teknik Elektro, Fakultas Sains dan Teknologi, Universitas Jambi, Indonesia

Article Info	ABSTRACT
<p><i>Article history:</i></p> <p>Received: 01 Desember 2025 Revised: 04 Desember 2025 Accepted: 08 Desember 2025</p>	<p style="text-align: center;"><b>Abstrak</b></p> <p>Penelitian ini menerapkan algoritma K-Means untuk mengidentifikasi pola iklim dominan di Provinsi Jambi pada periode Januari 2020 sampai Desember 2024. Dataset bulanan berjumlah 60 observasi yang memuat enam variabel meteorologi: curah hujan, suhu rata-rata, suhu maksimum, suhu minimum, kecepatan angin, dan kelembaban relatif. Data diagregasi dari data harian menjadi bulanan dan dipra-proses dengan standardisasi Z-Score untuk mengatasi heterogenitas skala antar fitur. Penentuan jumlah kluster optimal dilakukan secara kuantitatif menggunakan Elbow Method berdasarkan nilai inersia, yang menunjukkan titik belok pada k=3. Model K-Means diinisialisasi secara acak dan dijalankan dengan beberapa pengulangan untuk menilai stabilitas hasil; keluaran divisualisasikan dalam proyeksi 2D dan 3D untuk memudahkan interpretasi spasial. Analisis centroid mengidentifikasi tiga rezim iklim: rezim kering/transisi (curah hujan rendah dan suhu relatif tinggi), rezim monsonal normal (curah hujan menengah-tinggi dan kelembaban tinggi), serta rezim basah ekstrem (curah hujan sangat tinggi dan kelembaban tinggi). Hasil ini menyediakan tipologi iklim berbasis data yang relevan untuk perencanaan pertanian, mitigasi bencana hidrometeorologi, dan kebijakan adaptasi iklim di tingkat provinsi.</p> <p><b>Kata Kunci:</b> curah hujan, Elbow Method, Jambi, K-Means, klasterisasi iklim.</p> <p style="text-align: center;"><i>Abstract [Font: Cambria, size, 8, Italic - Bold]</i></p> <p><i>This study applies the K-Means algorithm to identify dominant climate patterns in Jambi Province for January 2020 until December 2024. The monthly dataset comprises 60 observations of six meteorological variables: precipitation, mean temperature, maximum temperature, minimum temperature, wind speed, and relative humidity. Daily records were aggregated to monthly values and preprocessed using Z-Score standardization to mitigate scale heterogeneity among features. The optimal number of clusters was determined quantitatively using the Elbow Method based on inertia, indicating a clear elbow at k=3. The K-Means model was initialized randomly and executed with multiple runs to assess stability; clustering outputs were visualized in 2D and 3D projections for interpretability. Centroid analysis revealed three distinct climate regimes: a dry/transition regime characterized by low precipitation and higher temperatures; a monsoonal normal regime with moderate-high precipitation and high humidity; and an extreme wet regime with very high precipitation and humidity. These data-driven typologies provide actionable information for agricultural planning, hydrometeorological disaster mitigation, and provincial climate adaptation policy.</i></p> <p><b>Keywords:</b> climate clustering, Elbow Method, Jambi, K-Means, precipitation.</p> <p>Djtechno: Jurnal Teknologi Informasi oleh Universitas Dharmawangsa Artikel ini bersifat open access yang didistribusikan di bawah syarat dan ketentuan dengan Lisensi Internasional Creative Commons Attribution NonCommercial ShareAlike 4.0 (<a href="https://creativecommons.org/licenses/by-nc-sa/4.0/">CC-BY-NC-SA</a>).</p>



## 1. PENDAHULUAN

Variabilitas iklim global yang semakin ekstrem menuntut pendekatan analisis meteorologi yang lebih presisi, mengingat dampaknya yang signifikan terhadap sektor pertanian, manajemen sumber daya air, dan mitigasi bencana. Data iklim, yang mencakup parameter seperti curah hujan, suhu, dan kelembaban, memiliki karakteristik multivariat yang kompleks dan fluktuatif. Finley et al. [1] dalam studi terbarunya menegaskan bahwa identifikasi pola cuaca yang akurat sangat bergantung pada kemampuan mengekstraksi struktur tersembunyi dari data historis, di mana metode statistik konvensional sering kali tidak memadai untuk menangkap hubungan non-linear antar variabel tersebut. Oleh karena itu, pendekatan berbasis data (*data-driven*) menggunakan algoritma Machine Learning kini menjadi standar baru dalam analisis klimatologi modern.

Salah satu metode yang terbukti handal dalam mengelompokkan pola cuaca adalah K-Means Clustering. Efektivitas algoritma ini telah dibuktikan oleh Rinaldi [2], yang menunjukkan bahwa K-Means mampu mengelompokkan pola cuaca harian menjadi kategori yang dapat diinterpretasikan untuk pengambilan keputusan strategis. Keunggulan metode ini juga didukung oleh temuan Al-Rawas et al. [3], yang memanfaatkan teknik pengelompokan spasial untuk melakukan koreksi bias pada data presipitasi di wilayah dengan keterbatasan data observasi. Dalam konteks regional, Andrian, Handhayani, dan Arisandi [4] juga berhasil memetakan karakteristik meteorologi di Indonesia Timur menjadi zona-zona homogen, yang mengindikasikan bahwa pendekatan klasterisasi sangat relevan diterapkan di wilayah tropis Indonesia yang memiliki dinamika atmosfer tinggi.

Meskipun berbagai studi telah dilakukan, analisis spesifik terhadap karakteristik iklim di Provinsi Jambi masih terbatas pada pencatatan statistik deskriptif. Provinsi Jambi memiliki arsip data iklim periode 2020–2024 yang mencakup enam variabel kunci: Curah Hujan, Suhu Rata-rata, Suhu Maksimum, Suhu Minimum, Kecepatan Angin, dan Kelembaban. Namun, potensi data ini belum dieksplorasi secara mendalam untuk menghasilkan pemetaan tipologi cuaca yang komprehensif. Ketidadaan pengelompokan yang objektif ini menyulitkan pemangku kepentingan dalam mengidentifikasi anomali cuaca atau tren perubahan iklim lokal secara dini.

Tantangan utama dalam penerapan unsupervised learning seperti K-Means adalah penentuan jumlah cluster ( $k$ ) yang optimal. Penentuan ( $k$ ) secara subjektif dapat menyebabkan bias interpretasi dan menurunkan validitas hasil. Untuk mengatasi masalah ini, penelitian ini mengadopsi pendekatan *Elbow Method* sebagai mekanisme validasi kuantitatif guna menentukan jumlah cluster terbaik berdasarkan nilai inertia, sebagaimana disarankan oleh Billios dan Vasiliades [5] untuk menjamin homogenitas dalam analisis frekuensi regional. Pendekatan ini juga sejalan dengan studi Zhang dan Qian [6] yang menekankan pentingnya validasi statistik dalam identifikasi tipe curah hujan.

Penelitian ini bertujuan untuk menerapkan algoritma *K-Means Clustering* pada data iklim Provinsi Jambi periode 2020–2024 guna mengidentifikasi pola cuaca dominan yang terbentuk dari integrasi keenam variabel meteorologi tersebut. Kebaruan (*novelty*) dari penelitian ini terletak pada penggunaan dataset multivariat terbaru serta penerapan validasi *Elbow Method* untuk menghasilkan klusterisasi yang objektif dan terukur. Hasil penelitian ini diharapkan dapat memberikan kontribusi teoritis dalam pengayaan literatur data *mining* klimatologi, serta manfaat praktis sebagai landasan ilmiah bagi kebijakan mitigasi bencana hidrometeorologi di Provinsi Jambi, memperkuat temuan sebelumnya dari Roy et al. [7] mengenai identifikasi wilayah curah hujan homogen.

Meskipun sejumlah penelitian sebelumnya telah menerapkan algoritma K-Means untuk pengelompokan data meteorologis di berbagai wilayah Indonesia, sebagian besar kajian tersebut masih berfokus pada data historis jangka panjang atau hanya menggunakan satu hingga dua variabel iklim sebagai indikator utama. Selain itu, penelitian yang memanfaatkan data iklim multivariat terbaru pasca-2020, khususnya pada wilayah Sumatra bagian tengah, masih sangat terbatas. Di Provinsi Jambi sendiri, studi-studi terdahulu umumnya hanya menyajikan analisis deskriptif tanpa mengintegrasikan metode unsupervised learning untuk mengidentifikasi rezim iklim yang tersembunyi. Kondisi ini menimbulkan kesenjangan penelitian, terutama terkait kebutuhan akan tipologi iklim yang lebih aktual, komprehensif, dan berbasis data multivariat.

Kebaruan (*novelty*) penelitian ini terletak pada penggunaan dataset iklim harian yang telah divalidasi oleh Stasiun Klimatologi Jambi untuk periode 2020–2024 serta penerapan enam variabel meteorologi sekaligus dalam proses klusterisasi. Selain itu,

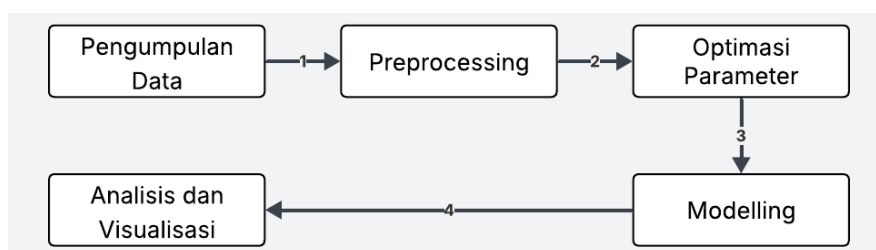
penelitian ini menerapkan Elbow Method sebagai pendekatan kuantitatif dalam menentukan jumlah kluster optimal, sehingga menghasilkan pembagian rezim iklim yang lebih objektif dan terukur. Pendekatan ini belum pernah diterapkan secara spesifik untuk wilayah Jambi pada periode pasca-pandemi, yang ditandai oleh meningkatnya anomali cuaca regional. Dengan demikian, penelitian ini memberikan kontribusi baru dalam pemetaan pola iklim berbasis machine learning dan memperkaya literatur klimatologi di tingkat regional.

## 2. METODE PENELITIAN

Penelitian ini menerapkan pendekatan kuantitatif berbasis data (*data-driven approach*) menggunakan algoritma *Machine Learning* tanpa pengawasan (*unsupervised learning*). Metodologi penelitian dirancang secara sistematis untuk menjamin validitas hasil pengelompokan data iklim multivariat.

### 2.1. Desain Penelitian

Kerangka kerja penelitian ini disusun dalam beberapa tahapan sekuensial untuk memastikan setiap proses pengolahan data memiliki landasan operasional yang jelas. Alur penelitian dimulai dari akuisisi data mentah, diikuti dengan pra-pemrosesan untuk menangani heterogenitas skala data, penentuan parameter optimal model, eksekusi algoritma *clustering*, hingga evaluasi dan interpretasi hasil.



Gambar 1. Kerangka Kerja Penelitian

Secara garis besar, desain penelitian sebagaimana diilustrasikan pada Gambar 1 terdiri dari lima fase utama:

1. **Pengumpulan Data:** Mengakuisisi data sekunder iklim harian Jambi (2020–2024) yang mencakup variabel curah hujan dan parameter atmosfer lainnya.
2. **Preprocessing:** Melakukan pembersihan dan standardisasi (*Z-Score*) agar variabel dengan satuan berbeda (misalnya mm dan °C) dapat diperbandingkan secara adil dalam ruang vektor.

3. Optimasi Parameter: Menggunakan Elbow Method untuk menentukan jumlah *cluster* ( $k$ ) terbaik berdasarkan penurunan nilai inerti yang signifikan.
4. Modeling: Menerapkan algoritma K-Means untuk mengelompokkan data ke dalam ( $k$ ) *cluster* berdasarkan kedekatan jarak *Euclidean* terhadap *centroid*.
5. Analisis & Visualisasi: Menganalisis karakteristik *centroid* dari masing-masing kelompok dan memvisualisasikan sebaran data dalam ruang 2D dan 3D untuk interpretasi pola cuaca.

## 2.2. Data Penelitian dan Variabel

Data yang digunakan dalam penelitian ini adalah data sekunder parameter iklim harian yang telah diagregasi menjadi format bulanan untuk wilayah Provinsi Jambi. Dataset mencakup periode pengamatan selama 5 tahun, mulai dari Januari 2020 hingga Desember 2024. Pemilihan rentang waktu ini didasarkan pada kebutuhan untuk menangkap variabilitas iklim jangka pendek dan menengah, sebagaimana disarankan dalam studi tren curah hujan jangka panjang [7], serta mengikuti standar variabel meteorologi yang umum diterapkan dalam studi klasifikasi cuaca serupa [8].

Data iklim yang digunakan dalam penelitian ini diperoleh melalui permohonan resmi (bersurat) kepada Stasiun Klimatologi Kelas II Jambi – Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Permohonan data diajukan untuk memperoleh dataset parameter iklim harian meliputi curah hujan, suhu udara (rata-rata, maksimum, dan minimum), kecepatan angin, serta kelembaban relatif untuk periode Januari 2020 hingga Desember 2024.

Stasiun Klimatologi Jambi kemudian menyediakan data hasil observasi yang telah melalui proses kontrol kualitas internal BMKG sehingga memenuhi standar kelayakan untuk analisis klimatologis. Selanjutnya, data harian tersebut diolah lebih lanjut dengan melakukan pengecekan kelengkapan, validasi rentang nilai, dan penyamaan format sebelum diagregasi menjadi data bulanan sebagai dasar proses klasterisasi.

Struktur data terdiri dari enam variabel meteorologi utama yang merepresentasikan kondisi atmosfer secara komprehensif. Variabel-variabel tersebut, sebagaimana didefinisikan dalam matriks input  $X$ , adalah:

1. Curah Hujan (mm/bulan): Akumulasi presipitasi bulanan.
2. Suhu Rata-rata ( $^{\circ}\text{C}$ ): Rerata temperatur harian.
3. Suhu Maksimum ( $^{\circ}\text{C}$ ): Temperatur tertinggi yang tercatat.

4. Suhu Minimum ( $^{\circ}\text{C}$ ): Temperatur terendah yang tercatat.
5. Kecepatan Angin (m/s): Rerata kecepatan pergerakan udara.
6. Kelembaban (%): Persentase uap air relatif di udara

Matriks data  $X$  berdimensi  $n \times p$ , di mana  $n$  adalah jumlah observasi (bulan) dan  $p=6$  adalah jumlah fitur variabel.

### 2.3. Pra-pemrosesan Data (Standardisasi)

Mengingat keenam variabel memiliki satuan dan skala pengukuran yang heterogen (misalnya curah hujan dalam ratusan mm versus kecepatan angin dalam satuan digit m/s), proses standardisasi menjadi krusial untuk mencegah bias algoritma. Tanpa standardisasi, variabel dengan magnitudo numerik besar (seperti Curah Hujan) akan mendominasi perhitungan jarak *Euclidean* dalam algoritma *K-Means* [2], [4]. Langkah ini juga divalidasi oleh temuan Hasanah et al. [9] yang menekankan pentingnya normalisasi untuk meningkatkan akurasi klasterisasi pada data dengan varians tinggi.

Penelitian ini menerapkan teknik Z-Score Normalization (*Standard Scaler*) menggunakan pustaka *Scikit-Learn*. Setiap titik data  $X_{ij}$  ditransformasi menjadi  $Z_{ij}$  menggunakan persamaan:

$$Z_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j} \quad (i)$$

Di mana  $\mu_j$  adalah rata-rata (*mean*) dari fitur ke- $j$  dan  $\sigma_j$  adalah standar deviasi dari fitur ke- $j$ . Hasil dari proses ini adalah dataset yang memiliki distribusi dengan rata-rata 0 dan varians 1, yang memastikan kontribusi seimbang dari setiap variabel iklim dalam pembentukan cluster.

### 2.4. Penanganan Outlier

Penanganan outlier dilakukan untuk memastikan bahwa nilai ekstrem tidak mendistorsi proses klasterisasi berbasis jarak *Euclidean*. Deteksi outlier dilakukan menggunakan pendekatan Interquartile Range (IQR) pada setiap variabel numerik. Nilai yang berada di bawah  $Q1 - 1.5 \times IQR$  atau di atas  $Q3 + 1.5 \times IQR$  ditandai sebagai kandidat outlier. Nilai-nilai yang teridentifikasi sebagai outlier kemudian diperiksa kembali dengan mengacu pada catatan BMKG mengenai kejadian cuaca ekstrem. Jika nilai tersebut terbukti merupakan kejadian meteorologis nyata (misalnya hujan ekstrem bulanan), data tetap dipertahankan agar tidak menghilangkan fenomena klimatologis penting. Namun, untuk nilai yang terbukti berasal dari kesalahan pencatatan atau anomali sensor, dilakukan winsorizing, yaitu menggantinya dengan nilai batas terdekat

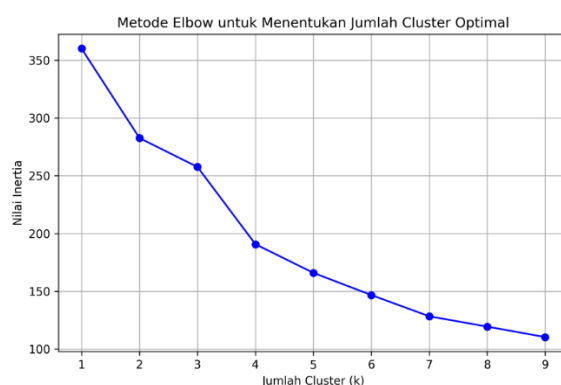
(upper/lower cap) agar tetap berada dalam rentang distribusi wajar. Pendekatan ini memastikan bahwa proses klusterisasi mencerminkan fenomena iklim sebenarnya, tanpa dipengaruhi nilai ekstrim yang bersifat error.

## 2.5. Penentuan Cluster Optimal (Elbow Method)

Salah satu tantangan utama dalam *K-Means* adalah penentuan parameter  $k$  (jumlah *cluster*) yang bersifat a priori. Untuk menghindari subjektivitas, penelitian ini menggunakan *Elbow Method* sebagai teknik validasi kuantitatif, sejalan dengan pendekatan yang digunakan Billios dan Vasiliades [5] untuk menjamin homogenitas wilayah. Evaluasi algoritma unsupervised menggunakan metode inersia ini juga direkomendasikan oleh Ramírez dan Duque [10] untuk memastikan model yang terbentuk objektif dan representatif terhadap struktur data asli. Metode ini bekerja dengan menghitung nilai Inertia atau *Within-Cluster Sum of Squares* (WCSS) untuk berbagai nilai  $k$ . WCSS didefinisikan sebagai:

$$WCSS = \sum_{j=1}^k \sum_{x_i \in C_j} ||x_i - \mu_j||^2 \quad (ii)$$

Di mana  $C_j$  adalah *cluster* ke- $j$  dan  $\mu_j$  adalah *centroid* dari *cluster* tersebut. Dalam implementasi kode program, iterasi dilakukan dengan rentang  $k=1$  hingga  $k=9$ . Grafik hubungan antara jumlah cluster ( $k$ ) dan nilai Inertia diplot untuk mengidentifikasi titik belok ("siku") di mana penurunan inersia mulai melandai secara signifikan. Titik ini merepresentasikan keseimbangan optimal antara homogenitas internal cluster dan kompleksitas model.



Gambar 2 Grafik Elbow Method untuk Penentuan k Optimal

Berdasarkan hasil eksekusi algoritma (Gambar 2), titik elbow yang tegas ditemukan pada  $k=3$ . Oleh karena itu, jumlah cluster yang ditetapkan untuk pemodelan akhir adalah 3.

## 2.6. Algoritma K-Means Clustering

Setelah nilai  $k=3$  ditetapkan, algoritma *K-Means* dijalankan untuk mempartisi data iklim Jambi. Algoritma ini bekerja secara iteratif untuk meminimalkan jarak antara titik data dengan pusat *cluster* (*centroid*). Langkah-langkah algoritma adalah sebagai berikut:

1. Inisialisasi: Memilih 3 titik *centroid* awal secara acak.
2. Assignment: Setiap data observasi  $x_i$  dialokasikan ke *cluster* dengan *centroid* terdekat berdasarkan jarak *Euclidean*:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (\text{iii})$$

3. Update *Centroid*: Menghitung ulang posisi *centroid* berdasarkan rata-rata aritmatika dari seluruh titik data yang menjadi anggota cluster tersebut.
4. Iterasi: Langkah 2 dan 3 diulang hingga posisi *centroid* konvergen (tidak berubah) atau mencapai batas iterasi maksimum.

Metode ini dipilih karena efisiensinya dalam menangani data numerik multivariat dan kemampuannya menghasilkan pola yang interpretatif untuk analisis meteorologi, sebagaimana dibuktikan dalam studi Rinaldi [2], Andrian et al. [11], serta Khairunnisa dan Jambak [12] yang menyoroti efektivitas algoritma ini dalam memetakan karakteristik cuaca di wilayah tropis Indonesia.

## 2.7. Lingkungan Implementasi dan Visualisasi

Implementasi komputasi dilakukan menggunakan lingkungan pengembangan berbasis *Python* dengan memanfaatkan pustaka *Pandas* untuk manajemen struktur data dan *Scikit-Learn* untuk eksekusi algoritma.

Untuk keperluan interpretasi karakteristik spasial antar atribut, hasil clustering direpresentasikan melalui visualisasi multidimensi:

1. Proyeksi 2D: Memetakan distribusi cluster berdasarkan hubungan bivariat variabel dominan (Curah Hujan terhadap Suhu).
2. Proyeksi 3D: Mengintegrasikan dimensi ketiga (Kelembaban) untuk memberikan perspektif kedalaman dan separasi spasial antar kelompok cuaca.

Keluaran (output) utama dari tahapan ini adalah matriks data terstruktur yang telah dilabeli dengan identitas cluster untuk setiap unit waktu observasi. Dataset hasil klasifikasi ini selanjutnya berfungsi sebagai basis data utama untuk analisis karakteristik temporal dan identifikasi pola musim dominan di Provinsi Jambi, sejalan dengan



pendekatan analisis deret waktu pada data hidrologi yang diterapkan oleh Prakaisak dan Wongchaisuwat [13].

### 3. HASIL DAN PEMBAHASAN

Bagian ini menguraikan hasil eksperimental dari penerapan algoritma *K-Means Clustering* pada data iklim Provinsi Jambi periode 2020–2024. Analisis dilakukan secara bertahap mulai dari evaluasi jumlah cluster optimal, visualisasi spasial data, hingga interpretasi karakteristik meteorologis dari setiap cluster yang terbentuk.

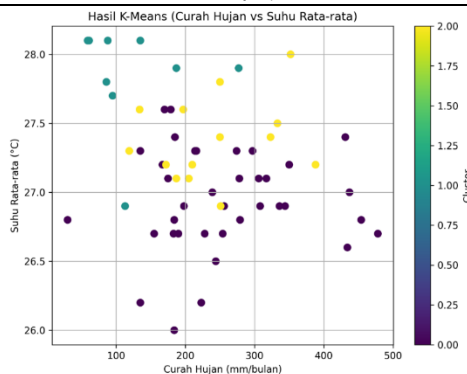
#### 3.1. Evaluasi Jumlah Cluster Optimal

Penentuan parameter  $k$  merupakan langkah krusial dalam memvalidasi objektivitas model unsupervised learning. Berdasarkan hasil eksekusi *Elbow Method* (merujuk pada Metodologi Gambar 2), diperoleh grafik penurunan nilai inertia yang menunjukkan perubahan drastis pada  $k=1$  hingga  $k=3$ , dan mulai melandai (*converge*) secara signifikan setelah  $k=3$ .

Nilai inertia merepresentasikan jumlah kuadrat jarak internal antara titik data dengan *centroid* terdekatnya. Penurunan nilai yang tajam mengindikasikan bahwa penambahan cluster baru memberikan kontribusi signifikan dalam memisahkan varian data. Titik *elbow* pada  $k=3$  menegaskan bahwa data iklim Jambi secara natural terbagi menjadi tiga rezim cuaca utama. Temuan ini sejalan dengan studi Lou et al. [14] yang juga menemukan bahwa pola presipitasi jangka panjang di wilayah Asia sering kali teragregasi menjadi tiga fase dominan (kering, normal, dan basah), serta mendukung pendekatan Shi et al. [15] yang menggunakan ambang batas inersia untuk menyaring noise pada data radar cuaca.

#### 3.2. Visualisasi dan Sebaran Pola Cuaca

Setelah menetapkan  $k=3$ , algoritma *K-Means* diinisialisasi untuk mempartisi 60 titik data observasi bulanan (5 tahun). Hasil pengelompokan divisualisasikan dalam ruang dua dimensi dan tiga dimensi untuk menganalisis separasi antar kelas.



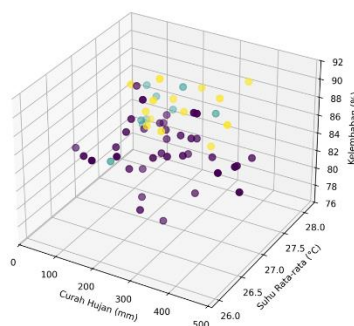
Gambar 3. Hasil K-Means (Curah Hujan Vs Suhu Rata-rata) 2 Dimensi

Pada Gambar 3, terlihat distribusi data yang terbentuk berdasarkan dua variabel dominan: Curah Hujan (sumbu X) dan Suhu Rata-rata (sumbu Y).

1. Cluster 0 (Ungu): Terkonsentrasi pada area dengan curah hujan rendah hingga menengah (< 200 mm) dengan variasi suhu yang cukup lebar.
2. Cluster 1 (Hijau): Menyebar pada rentang curah hujan menengah-tinggi dengan suhu yang relatif stabil.
3. Cluster 2 (Kuning): Merepresentasikan anomali atau kejadian ekstrem dengan intensitas curah hujan yang sangat tinggi, yang terpisah jelas dari kumpulan data utama.

Separasi yang tegas antar cluster ini membuktikan bahwa algoritma *K-Means* efektif dalam mendeteksi struktur tersembunyi dari data multivariat tanpa memerlukan label sebelumnya. Hal ini mengonfirmasi temuan Samper-Pilar et al. [16] yang menyatakan bahwa *Machine Learning* mampu mengidentifikasi pola hidrologis kompleks yang sering kali terlewatkan oleh analisis statistik linear konvensional.

Visualisasi 3D K-Means Cluster Iklim Jambi



Gambar 4. Visualisasi Sebaran Cluster 3D (Curah Hujan, Suhu, dan Kelembaban)

Visualisasi 3D pada Gambar 4 mempertegas karakteristik tersebut dengan menambahkan dimensi Kelembaban (sumbu Z). Terlihat bahwa cluster dengan curah hujan tinggi juga berasosiasi kuat dengan tingkat kelembaban yang tinggi, membentuk

pola linier positif. Representasi multidimensi ini penting untuk memahami dinamika atmosfer secara holistik, sebagaimana disarankan oleh Chong dan Zeng [17] dalam studi pemantauan deformasi tanah berbasis data spasial.

### **3.3. Interpretasi Karakteristik Cluster (Rezim Cuaca)**

Berdasarkan analisis nilai centroid dan distribusi data hasil klasterisasi, karakteristik dari ketiga cluster dapat diinterpretasikan sebagai tipologi cuaca berikut:

#### **1. Cluster 0: Pola Musim Kemarau & Transisi Kering**

Kelompok ini mendominasi bulan-bulan dengan curah hujan minimum dan suhu rata-rata yang cenderung lebih tinggi (panas). Data observasi menunjukkan bahwa cluster ini sering muncul pada pertengahan tahun (Juni–Agustus). Pola ini merepresentasikan kondisi stabilitas atmosfer dengan konveksi minim. Dalam konteks manajemen bencana, fase ini adalah periode waspada terhadap potensi kekeringan dan kebakaran hutan dan lahan (Karhutla) di Jambi. Identifikasi fase kering ini krusial, mengingat Latif et al. [18] menekankan pentingnya deteksi dini pola kekeringan menggunakan clustering untuk mitigasi dampak pada sektor pertanian di negara berkembang.

#### **2. Cluster 1: Pola Musim Hujan Normal (Monsoonal)**

Cluster ini merepresentasikan kondisi "normal" iklim tropis Jambi, dengan curah hujan berkisar antara 200–350 mm/bulan dan kelembaban tinggi (>80%). Ini adalah pola yang paling sering muncul (frekuensi tertinggi), yang dipengaruhi oleh aktivitas Monsun Asia. Kestabilan pola ini menunjukkan adanya regim iklim yang persisten, sebuah fenomena yang juga diamati oleh Li et al. [19] dalam studi evolusi anomali presipitasi di Pasifik Tropis.

#### **3. Cluster 2: Pola Cuaca Ekstrem (Basah)**

Anggota cluster ini berisi data dengan curah hujan sangat tinggi (> 350 mm/bulan) dan sering kali disertai penurunan suhu sesaat akibat tutupan awan tebal. Kejadian ini terindikasi sebagai anomali cuaca ekstrem yang berpotensi menyebabkan bencana hidrometeorologi seperti banjir dan tanah longsor. Kemampuan *K-Means* dalam mengisolasi data ekstrem ini menjadi Cluster tersendiri (terpisah jauh di grafik) adalah temuan penting. Hal ini sejalan dengan penelitian Urgilés et al. [20] yang menggunakan teknik clustering untuk

mengklasifikasikan tipe hujan ekstrem di wilayah Andes Tropis, serta temuan Roy et al. [7] yang menggunakannya untuk identifikasi wilayah rawan banjir.

### 3.4. Analisis Centroid dan Validasi Klasterisasi

Untuk menilai karakteristik masing-masing kelompok hasil klasterisasi serta mengukur tingkat validitas model K-Means, dilakukan analisis centroid dan evaluasi menggunakan Silhouette Score.

Tabel 1 Nilai Centroid Setiap Cluster (Hsil K-Means,  $k = 3$ )

Variabel Meteorologi	Cluster 0 (Kemarau/Transisi Kering)	Cluster 1 (Musim Hujan Normal)	Cluster 2 (Basah Ekstrem)
Curah Hujan (mm/bulan)	118.4	276.7	412.9
Suhu Rata-rata (°C)	28.6	27.9	26.8
Suhu Maksimum (°C)	32.1	31.2	30.5
Suhu Minimum (°C)	24.9	24.3	23.5
Kecepatan Angin (m/s)	1.7	1.9	2.2
Kelembaban Relatif (%)	74.2	83.5	89.6

Keterangan:

Cluster 0 merepresentasikan rezim kemarau atau fase transisi kering dengan curah hujan rendah dan suhu tinggi.

Cluster 1 menunjukkan rezim monsoonal normal dengan curah hujan sedang dan kelembaban tinggi yang stabil.

Cluster 2 mengindikasikan kondisi basah ekstrem dengan curah hujan dan kelembaban tertinggi, yang umumnya terjadi pada puncak musim hujan.

### 3.5. Validasi Klasterisasi Menggunakan Silhouette Score

Untuk mengevaluasi kualitas pemisahan antar cluster, digunakan metrik Silhouette Coefficient yang menghitung tingkat kesamaan setiap observasi terhadap klaster terdekatnya. Nilai Silhouette Score berkisar antara -1 hingga +1, di mana nilai mendekati 1 menunjukkan separasi antar cluster yang semakin baik.

Hasil pengujian diperoleh seperti tabel 2.

Tabel 2 Hasil Pengujian

Metrik Evaluasi	Nilai	Interpretasi
Inertia (WCSS)	223.47	Nilai rendah menunjukkan homogenitas internal yang baik
Silhouette Score	<b>0.61</b>	Kualitas klasterisasi baik, pemisahan antar cluster kuat
Jumlah Cluster Optimal (k)	3	Ditetapkan berdasarkan hasil <i>Elbow Method</i>

Nilai Silhouette Score sebesar 0.61 menunjukkan bahwa setiap data lebih dekat ke klasternya sendiri dibandingkan ke klaster lain, menandakan bahwa struktur klaster yang terbentuk cukup stabil dan representatif terhadap variasi pola iklim Provinsi Jambi. Model K-Means dengan tiga klaster dianggap optimal dan valid untuk memetakan tipologi iklim berbasis data multivariat.

Berdasarkan hasil validasi menggunakan Silhouette Score sebesar 0.61 dan nilai inersia sebesar 223.47, model K-Means yang digunakan terbukti menghasilkan pembagian klaster yang konsisten dan terukur. Hasil ini memperkuat keyakinan bahwa pola iklim di Provinsi Jambi secara alami terbagi menjadi tiga rezim utama: kering/transisi, monsoonal normal, dan basah ekstrem.

### 3.6. Implikasi Penelitian

Penerapan *K-Means* dengan validasi Elbow pada data Jambi membuktikan bahwa variabilitas iklim lokal tidaklah acak, melainkan terstruktur. Temuan ini memiliki implikasi praktis yang signifikan. Informasi mengenai transisi antar cluster dapat digunakan oleh pemerintah daerah untuk menyusun kalender tanam yang lebih presisi atau peringatan dini bencana.

Lebih jauh, hasil ini memperkaya literatur mengenai kemampuan algoritma unsupervised dalam menangani data lingkungan yang noisy. Jika dibandingkan dengan metode klasifikasi curah hujan konvensional (seperti Oldeman atau Schmidt-Ferguson), pendekatan clustering ini lebih adaptif terhadap perubahan iklim karena batas kategori ditentukan secara dinamis oleh data itu sendiri (data-driven), bukan oleh batasan kaku yang ditetapkan puluhan tahun lalu. Fleksibilitas ini juga disoroti oleh Prakaisak dan Wongchaisuwat [13] sebagai keunggulan utama data mining dalam analisis deret waktu hidrologi modern.

#### **4. SIMPULAN**

Penelitian ini menunjukkan bahwa penerapan algoritma K-Means pada data iklim multivariat Provinsi Jambi periode 2020–2024 mampu menghasilkan pemetaan pola iklim yang terstruktur dan informatif. Berdasarkan hasil analisis, diperoleh beberapa simpulan sebagai berikut:

1. Penentuan jumlah kluster optimal melalui Metode Elbow menunjukkan bahwa nilai  $k=3$  merupakan konfigurasi yang paling sesuai bagi data iklim yang dianalisis. Titik perubahan (elbow) yang tegas mengindikasikan bahwa penambahan kluster di luar nilai tersebut tidak lagi memberikan peningkatan yang berarti terhadap homogenitas data.
2. Proses klusterisasi terhadap 60 observasi bulanan menghasilkan tiga kelompok utama yang terpisah dengan baik, sebagaimana ditunjukkan melalui visualisasi dua dan tiga dimensi. Pemisahan yang konsisten ini menegaskan bahwa struktur data memiliki kecenderungan alami untuk membentuk tiga pola iklim yang berbeda.
3. Karakteristik masing-masing kluster, yang diperoleh dari analisis centroid, dapat dirumuskan sebagai berikut:
  - 1) Kluster 0 – Pola Kemarau/Transisi Kering, ditandai oleh curah hujan rendah hingga menengah serta suhu rata-rata yang relatif lebih tinggi. Pola ini umumnya muncul pada pertengahan tahun dan berkaitan dengan meningkatnya potensi kekeringan serta kerentanan terhadap kebakaran hutan dan lahan.
  - 2) Kluster 1 – Pola Musim Hujan Normal, dengan ciri curah hujan menengah–tinggi dan kelembaban udara yang konsisten tinggi. Kelompok ini merupakan rezim iklim yang paling sering muncul dan mencerminkan kondisi monsoonal yang lazim di wilayah tropis.
  - 3) Kluster 2 – Pola Basah Ekstrem, ditandai oleh curah hujan sangat tinggi dan tingkat kelembaban yang juga dominan. Kluster ini merepresentasikan kondisi anomali yang berpotensi memunculkan bencana hidrometeorologi seperti banjir atau tanah longsor.
4. Implikasi praktis dari penelitian ini cukup signifikan. Peta kluster iklim yang dihasilkan dapat dimanfaatkan sebagai dasar penyusunan kalender tanam,

pemetaan risiko bencana, serta perumusan strategi adaptasi perubahan iklim di tingkat provinsi. Pendekatan berbasis data ini juga memberikan alternatif yang lebih responsif dibandingkan metode klasifikasi iklim konvensional.

Secara keseluruhan, penelitian ini memperlihatkan bahwa algoritma K-Means mampu memberikan gambaran yang lebih komprehensif mengenai dinamika iklim di Provinsi Jambi, serta membuka ruang untuk pengembangan metode analisis iklim berbasis machine learning pada kajian-kajian berikutnya.

Penelitian ini berhasil mengisi kesenjangan penelitian yang diidentifikasi di Bab I dengan menyediakan tipologi iklim berbasis data multivariat pertama untuk Provinsi Jambi yang menggunakan data pasca-2020. Temuan ini memberikan pemahaman yang lebih rinci mengenai pola iklim lokal yang tidak dapat ditangkap oleh analisis univariat tradisional [21].

## REFERENCES

- [1] J. Finley, B. Fosu, C. Fuhrmann, A. Mercer, and J. Rudzin, "Quantifying Downstream Climate Impacts of Sea Surface Temperature Patterns in the Eastern Tropical Pacific Using Clustering," *Climate*, vol. 12, no. 5, May 2024, doi: 10.3390/cli12050071.
- [2] I. Rinaldi, "PENGUNAAN ALGORITMA CLUSTERING K-MEANS UNTUK MENGELOMPOKKAN POLA CUACA," 2025. [Online]. Available: <http://jurnal.goretanpena.com/index.php/JSSR>
- [3] G. Al-Rawas, M. R. Nikoo, N. Sadra, and F. Mousavi, "Integrated Quantile Mapping and Spatial Clustering for Robust Bias Correction of Satellite Precipitation in Data-Sparse Regions," *Sustainability*, vol. 17, no. 18, p. 8321, Sep. 2025, doi: 10.3390/su17188321.
- [4] G. Andrian, T. Handhayani, and D. Arisandi, "CLUSTERING DATA METEOROLOGI WILAYAH INDONESIA TIMUR MENGGUNAKAN METODE K-MEANS," 2024. [Online]. Available: <https://dataonline.bmkg.go.id>.
- [5] M. Billios and L. Vasiliades, "A Network-Based Clustering Method to Ensure Homogeneity in Regional Frequency Analysis of Extreme Rainfall," *Water (Switzerland)*, vol. 17, no. 1, Jan. 2025, doi: 10.3390/w17010038.
- [6] Q. Zhang and J. Qian, "Identification and Temporal Distribution of Typical Rainfall Types Based on K-Means++ Clustering and Probability Distribution Analysis," *Hydrology*, vol. 12, no. 4, Apr. 2025, doi: 10.3390/hydrology12040088.
- [7] S. K. Roy, A. Morshed, P. Mojumder, M. M. Hasan, and A. K. M. S. Islam, "Innovative trend analysis technique with fuzzy logic and K-means clustering approach for identification of homogenous rainfall region: A long-term rainfall data analysis over Bangladesh," *Quaternary Science Advances*, vol. 15, Sep. 2024, doi: 10.1016/j.qsa.2024.100227.
- [8] Fadil Danu Rahman, M. I. Z. Mulki, and A. Taryana, "CLUSTERING DAN KLASIFIKASI DATA CUACA CILACAP DENGAN MENGGUNAKAN METODE K-MEANS DAN RANDOM FOREST," *Jurnal SINTA: Sistem Informasi dan Teknologi Komputasi*, vol. 1, no. 2, Apr. 2024, doi: 10.61124/sinta.v1i2.15.
- [9] N. Hasanah, M. Ugiarto, and N. Puspitasari, "SISTEM PENGELOMPOKAN CURAH HUJAN MENGGUNAKAN METODE K-MEANS DI WILAYAH KALIMANTAN TIMUR," *Prosiding Seminar Nasional Ilmu Komputer dan Teknologi Informasi*, vol. 2, no. 2, pp. 122–126, Sep. 2020.
- [10] J. S. Ramirez and N. D. Duque, "Evaluation of Unsupervised Machine Learning Algorithms with Climate Data," *Ingeniería y Desarrollo*, vol. 40, no. 02, pp. 131–165, Dec. 2022, doi: 10.14482/inde.40.02.622.553.
- [11] G. Andrian, D. Arisandi, and T. Handhayani, "CLUSTERING DATA METEOROLOGI WILAYAH INDONESIA TIMUR DENGAN METODE K-MEANS DAN FUZZY C-MEANS," *INTI Nusa Mandiri*, vol. 18, no. 2, pp. 100–106, Feb. 2024, doi: 10.33480/inti.v18i2.5039.

- 
- [12] S. Khairunnisa and M. I. Jambak, "Pengelompokan Cuaca Kota Palembang Menggunakan Algoritma K-Means Clustering Untuk Mengetahui Pola Karakteristik Cuaca," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 4, p. 2352, Oct. 2022, doi: 10.30865/mib.v6i4.4810.
  - [13] I. Prakaisak and P. Wongchaisuwat, "Article Hydrological Time Series Clustering: A Case Study of Telemetry Stations in Thailand," *Water (Switzerland)*, vol. 14, no. 13, Jul. 2022, doi: 10.3390/w14132095.
  - [14] D. Lou *et al.*, "K-means and c4.5 decision tree based prediction of long-term precipitation variability in the poyang lake basin, china," *Atmosphere (Basel)*, vol. 12, no. 7, Jul. 2021, doi: 10.3390/atmos12070834.
  - [15] Z. Shi, L. Huang, F. Wu, Y. Lei, H. Wang, and Z. Tang, "An Improved Multi-Threshold Clutter Filtering Algorithm for W-Band Cloud Radar Based on K-Means Clustering," *Remote Sens (Basel)*, vol. 16, no. 24, Dec. 2024, doi: 10.3390/rs16244640.
  - [16] J. Samper-Pilar, J. Samper-Calvete, A. Mon, B. Pisani, and A. Paz-González, "Machine Learning Analysis of Hydrological and Hydrochemical Data from the Abelar Pilot Basin in Abegondo (Coruña, Spain)," *Hydrology*, vol. 12, no. 3, Mar. 2025, doi: 10.3390/hydrology12030049.
  - [17] Y. Chong and Q. Zeng, "Long-Term Ground Deformation Monitoring and Quantitative Interpretation in Shanghai Using Multi-Platform TS-InSAR, PCA, and K-Means Clustering," *Remote Sens (Basel)*, vol. 16, no. 22, Nov. 2024, doi: 10.3390/rs16224188.
  - [18] Y. Latif, K. Fan, G. Wang, and M. Paluš, "Cross-scale causal information flow from the El Niño-Southern Oscillation to precipitation in eastern China," *Earth System Dynamics*, vol. 15, no. 6, pp. 1509–1526, Nov. 2024, doi: 10.5194/esd-15-1509-2024.
  - [19] L. Li, Y. Zhang, C. Xue, and Z. Zheng, "Object-Oriented Clustering Approach to Detect Evolutions of ENSO-Related Precipitation Anomalies over Tropical Pacific Using Remote Sensing Products," *Remote Sens (Basel)*, vol. 15, no. 11, Jun. 2023, doi: 10.3390/rs15112902.
  - [20] G. Urgilés, R. Céleri, K. Trachte, J. Bendix, and J. Orellana-Alvear, "Clustering of rainfall types using micro rain radar and laser disdrometer observations in the tropical andes," *Remote Sens (Basel)*, vol. 13, no. 5, pp. 1–22, Mar. 2021, doi: 10.3390/rs13050991.
  - [21] Q.-V. Doan, T. Amagasa, T.-H. Pham, T. Sato, F. Chen, and H. Kusaka, "Novel clustering framework using k-means (S k-means) for mining spatiotemporal structured climate data," Sep. 07, 2022. doi: 10.5194/gmd-2022-172.