JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 168-177
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.945

Penerapan K-Means untuk Klasterisasi Pola Cuaca Spasial di Kawasan Sumatera Berbasis Data Reanalisis ERA5

Yehezkiel Haganta Tarigan[1], Sofia Zahra[2*], Christian Nicholas Sinaga[3]

[1,2,3]Fakultas Matematika dan Ilmu Pengetahuan Alam, Ilmu Komputer, Universitas Negeri Medan, Medan, Indonesia
Email: [1]yehezkielhaganta@gmail.com, [2*]sofiazahra2005@gmail.com, [3]christiannicholassinaga@gmail.com
(*Email Corresponding Author: sofiazahra2005@gmail.com)
Received: 20 Maret 2026 | Revision: 27 Maret 2026 | Accepted: 27 Maret 2026

Abstrak

Penelitian ini bertujuan untuk mengelompokkan pola cuaca spasial di wilayah Sumatera dengan memanfaatkan metode K-Means berbasis data reanalisis ERA5. Latar belakang penelitian ini didasari oleh kompleksitas dinamika cuaca yang tinggi serta keterbatasan data observasi yang tersebar tidak merata, sehingga diperlukan pendekatan berbasis data untuk memperoleh pola yang lebih jelas dan terstruktur. Proses penelitian dilakukan melalui beberapa tahapan, yaitu pembersihan data, normalisasi menggunakan metode Min-Max Scaling, penentuan jumlah cluster dengan metode Elbow, serta proses pengelompokan menggunakan algoritma K-Means. Variabel yang digunakan meliputi suhu udara, tekanan permukaan, dan kecepatan angin sebagai representasi kondisi atmosfer. Hasil penelitian menunjukkan bahwa pengelompokan yang dihasilkan mampu menggambarkan perbedaan karakteristik wilayah, seperti area perairan, pegunungan, dataran rendah, serta zona transisi pesisir. Selain itu, pola yang terbentuk juga mencerminkan kondisi geografis yang beragam di wilayah penelitian. Dengan demikian, metode K-Means dapat digunakan sebagai pendekatan yang efektif dalam mengidentifikasi pola cuaca spasial secara lebih sistematis.

Kata Kunci: K-Means, ERA5, Klasterisasi, Pola Cuaca, Data Mining

Abstract

This study aims to classify spatial weather patterns in the Sumatra region using the K-Means method based on ERA5 reanalysis data. The background of this research is driven by the complexity of weather dynamics and the uneven distribution of observational data, which makes it necessary to apply a data-driven approach to obtain clearer and more structured patterns. The research process consists of several stages, including data cleaning, normalization using the Min-Max Scaling method, determining the number of clusters with the Elbow method, and clustering using the K-Means algorithm. The variables used include air temperature, surface pressure, and wind speed as representations of atmospheric conditions. The results show that the clustering process is able to describe differences in regional characteristics, such as marine areas, mountainous regions, lowlands, and coastal transition zones. In addition, the resulting patterns reflect the diverse geographical conditions of the study area. Therefore, the K-Means method can be considered an effective approach for identifying spatial weather patterns in a more systematic manner.

***Keywords**: K-Means, ERA5, Clustering, Weather Patterns, Data Mining*

1. PENDAHULUAN

Wilayah Indonesia, khususnya kawasan Sumatera dan sekitarnya, memiliki karakteristik cuaca dan iklim yang sangat kompleks akibat interaksi antara faktor lokal seperti topografi dan kondisi geografis, serta faktor global seperti *El Niño–Southern Oscillation* (*ENSO*) dan *Madden–Julian Oscillation* (*MJO*) yang memengaruhi variabilitas atmosfer [1]. Kompleksitas ini menyebabkan pola curah hujan dan parameter meteorologi lainnya menunjukkan variabilitas tinggi baik secara spasial maupun temporal, sehingga menyulitkan proses identifikasi pola cuaca secara konvensional. Selain itu, keterbatasan distribusi stasiun pengamatan di Indonesia, terutama di wilayah terpencil, menyebabkan ketersediaan data observasi menjadi tidak merata dan seringkali tidak kontinu [2]. Kondisi tersebut menuntut adanya pendekatan berbasis data yang mampu mengolah informasi dalam skala besar untuk mengidentifikasi pola cuaca secara lebih efektif dan sistematis.

Dalam beberapa tahun terakhir, data *reanalysis* seperti ERA5 menjadi sumber data alternatif yang banyak digunakan dalam studi klimatologi karena menyediakan data atmosfer global yang konsisten dan berkelanjutan dengan resolusi temporal yang tinggi [3]. Data ini mampu merepresentasikan berbagai variabel meteorologi secara komprehensif dan telah dimanfaatkan dalam berbagai penelitian untuk analisis curah hujan, suhu, serta dinamika atmosfer [4]. Meskipun demikian, data ERA5 masih memiliki keterbatasan dalam resolusi spasial serta potensi bias pada kondisi ekstrem, sehingga diperlukan metode analisis tambahan untuk mengekstraksi informasi yang lebih representatif [5]. Di Indonesia, pemanfaatan data ERA5 terbukti mampu membantu mengatasi keterbatasan data observasi serta memberikan gambaran distribusi curah hujan yang lebih luas dan konsisten [6].

Seiring meningkatnya volume data iklim, pendekatan berbasis *data mining* dan *machine learning* menjadi solusi yang relevan dalam analisis pola cuaca. Salah satu teknik yang banyak digunakan adalah klasterisasi, yaitu metode pengelompokan data berdasarkan kemiripan karakteristik tanpa memerlukan label sebelumnya [7]. Algoritma *K-Means* merupakan salah satu metode klasterisasi yang paling populer karena kemudahannya dalam implementasi serta kemampuannya dalam

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 168-177
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.945

mengelompokkan data multidimensi secara efisien [8]. Metode ini telah banyak digunakan dalam berbagai bidang, termasuk klimatologi, untuk mengidentifikasi pola atmosfer, zona iklim homogen, serta karakteristik cuaca tertentu [9].

Berbagai penelitian terdahulu telah menunjukkan efektivitas penggunaan *K-Means* dalam analisis data iklim. Klasterisasi menggunakan *K-Means* mampu mengelompokkan wilayah berdasarkan kesamaan pola curah hujan dan suhu sehingga menghasilkan zona iklim yang lebih homogen [10]. Selain itu, pendekatan ini juga digunakan untuk menganalisis distribusi spasial data meteorologi berbasis grid yang kompleks [11]. Penelitian lain menunjukkan bahwa metode *K-Means* dapat digunakan untuk mengidentifikasi pola sirkulasi atmosfer dan rezim cuaca yang berulang [12]. Dalam konteks data ERA5, klasterisasi juga dimanfaatkan untuk menganalisis variabilitas curah hujan dan fenomena atmosfer ekstrem secara lebih terstruktur [13].

Di Indonesia, penerapan metode *machine learning* dalam analisis iklim juga mengalami perkembangan yang signifikan. Beberapa penelitian menunjukkan bahwa algoritma seperti *Random Forest*, *Support Vector Machine*, dan *Gradient Boosting* mampu meningkatkan akurasi prediksi curah hujan [14]. Selain itu, pendekatan berbasis *deep learning* seperti *Convolutional Neural Network* (*CNN*) dan arsitektur U-Net digunakan untuk meningkatkan resolusi spasial data curah hujan melalui proses *downscaling* [15]. Integrasi antara teknik klasterisasi dan metode prediktif juga telah dilakukan untuk meningkatkan pemahaman terhadap pola iklim dan mendukung sistem peringatan dini [16].

Dalam skala global, penggunaan teknik klasterisasi dalam studi klimatologi terus berkembang. Klasterisasi digunakan untuk mengidentifikasi pola curah hujan global, variabilitas suhu, serta distribusi fenomena atmosfer dalam jangka panjang [17]. Penelitian lain menunjukkan bahwa metode *K-Means* dapat digunakan untuk mengelompokkan kejadian cuaca ekstrem seperti hujan lebat dan badai berdasarkan karakteristik spasial dan temporalnya [18]. Selain itu, pendekatan klasterisasi juga digunakan untuk menganalisis hubungan antara variabel atmosfer dan fenomena iklim kompleks, termasuk interaksi antara angin, tekanan, dan kelembapan [19].

Penggunaan data ERA5 yang dikombinasikan dengan metode *machine learning* juga semakin berkembang dalam penelitian modern. Beberapa studi menunjukkan bahwa teknik berbasis kecerdasan buatan mampu meningkatkan resolusi dan akurasi data cuaca melalui proses *downscaling* serta pemodelan spasial-temporal [20]. Selain itu, pendekatan berbasis *pre-trained model* dan *self-supervised learning* mulai diterapkan untuk memahami pola cuaca secara lebih mendalam. Metode klasterisasi juga dikembangkan lebih lanjut melalui variasi algoritma seperti *structural K-Means* yang mempertimbangkan hubungan spasial dan temporal dalam data [21].

Meskipun berbagai penelitian telah dilakukan, masih terdapat keterbatasan dalam studi yang secara khusus mengkaji pengelompokan pola cuaca spasial di wilayah Indonesia, terutama Sumatera, menggunakan metode *K-Means* berbasis data ERA5. Sebagian besar penelitian lebih berfokus pada prediksi cuaca atau peningkatan resolusi data, sementara analisis pola spasial melalui pendekatan klasterisasi masih relatif terbatas [9]. Selain itu, beberapa penelitian menggunakan metode yang kompleks sehingga kurang praktis untuk diterapkan dalam analisis operasional [22]. Terdapat pula penelitian yang menggunakan klasterisasi, namun belum secara mendalam mengeksplorasi pola spasial cuaca dalam konteks regional Indonesia [23].

Berdasarkan kajian tersebut, dapat diidentifikasi adanya *research gap*, yaitu belum optimalnya pemanfaatan metode *K-Means* untuk mengelompokkan pola cuaca spasial berbasis data ERA5 secara khusus di wilayah Sumatera. Oleh karena itu, penelitian ini bertujuan untuk menerapkan metode *K-Means* dalam mengelompokkan pola cuaca spasial di kawasan Sumatera dan sekitarnya menggunakan data *reanalysis* ERA5. Penelitian ini diharapkan mampu menghasilkan klaster pola cuaca yang representatif serta memberikan pemahaman yang lebih mendalam mengenai distribusi dan variabilitas cuaca di wilayah tersebut. Selain itu, hasil penelitian ini diharapkan dapat mendukung pengembangan sistem analisis cuaca berbasis data serta memberikan kontribusi dalam pengambilan keputusan di bidang mitigasi bencana, pertanian, dan pengelolaan sumber daya air di Indonesia.

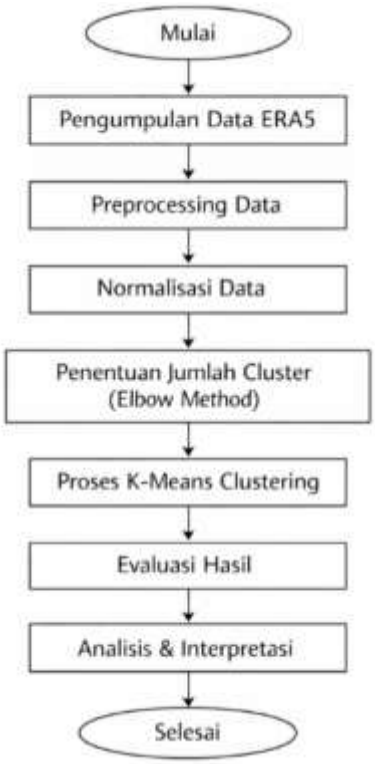
2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilakukan untuk mengelompokkan pola cuaca spasial di wilayah Sumatera menggunakan metode *K-Means clustering* berbasis data reanalisis ERA5. Tahapan penelitian disusun secara sistematis untuk memastikan bahwa proses analisis berjalan terstruktur dan hasil yang diperoleh sesuai dengan tujuan penelitian.

Secara umum, tahapan penelitian dimulai dari pengumpulan data, dilanjutkan dengan preprocessing, normalisasi, penentuan jumlah cluster, proses clustering menggunakan metode *K-Means*, serta evaluasi dan interpretasi hasil. Setiap tahapan memiliki peran penting dalam menghasilkan pengelompokan data yang optimal. Seluruh proses pengolahan data dilakukan menggunakan bahasa pemrograman Python dengan bantuan library seperti Pandas, NumPy, Matplotlib, dan Scikit-learn. Berikut diagram alur tahapan penelitian yang dituangkan pada gambar 1:

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 168-177
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.945



Gambar 1. Alur Tahapan Penelitian

Berdasarkan Gambar 1, tahapan pertama diawali dengan pengumpulan data dari *Copernicus Climate Change Service* (C3S) yang dikelola oleh *European Centre for Medium-Range Weather Forecasts* (ECMWF). Dataset yang diekstraksi adalah *ERA5 hourly data on single levels* untuk periode bulan Februari. Pengambilan data diotomatisasi menggunakan *Application Programming Interface* (API) cdsapi. Domain observasi dibatasi secara spasial pada kawasan Sumatera dan sekitarnya, dengan rentang koordinat 6° LU – 6° LS dan 95° BT – 106,5° BT. Untuk merepresentasikan dinamika atmosfer secara komprehensif, penelitian ini menggunakan empat variabel meteorologis utama, yaitu suhu udara (*temperature 2m* / t2m), tekanan permukaan (*surface pressure* / sp), serta komponen kecepatan angin arah zonal (u10) dan meridional (v10). Variabel ini dipilih karena mampu mendeskripsikan kondisi termal dan pergerakan massa udara secara presisi. Dataset ini membentuk dimensi yang sangat masif, mencapai 1.547.616 baris observasi, dengan spesifikasi lengkap yang dirangkum pada **Tabel 1**.

Tabel 1. Spesifikasi Dataset Penelitian

Komponen	Deskripsi
Sumber Data	Copernicus Climate Change Service (C3S), ECMWF
Nama Dataset	ERA5 hourly data on single levels from 1940 to present
Wilayah Penelitian	Pulau Sumatera dan sekitarnya
Batas Koordinat	6° LU – 6° LS dan 95° BT – 106,5° BT
Periode Data	Bulan Februari (data per jam)
Resolusi Temporal	Per jam (*hourly*)
Jumlah Data	1.547.616 baris × 6 kolom
Variabel	t2m, sp, u10, v10
Format Data	Data numerik berbasis grid
Jenis Data	Data spasial-temporal

Tahap selanjutnya adalah pra-pemrosesan data (*preprocessing*). Pada fase ini, data divalidasi dari anomali dan *missing values* untuk menjaga integritas informasi sebelum masuk ke tahap pemodelan. Dilakukan pula konversi satuan suhu atmosfer dari Kelvin menjadi Celsius agar hasil analisis akhir lebih intuitif untuk diinterpretasikan. Mengingat keempat variabel cuaca tersebut memiliki skala dan satuan yang sangat kontras (misalnya tekanan dalam orde ratusan ribu, sedangkan angin dalam satuan tunggal), maka tahap normalisasi menjadi prasyarat mutlak. Seluruh fitur ditransformasikan menggunakan *Z-score Standardization* (Standard Scaler) sehingga memiliki nilai rata-rata (*mu*) 0 dan standar deviasi (σ) 1. Persamaan matematis untuk standarisasi ini dirumuskan sebagai berikut:

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 168-177
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.945

$$Z = \frac{X - \mu}{\sigma} \quad (1)$$

Di mana X adalah nilai asli, μ adalah rata-rata, σ adalah standar deviasi dari masing-masing variabel.

Setelah ruang fitur berada pada skala yang seragam, pemodelan dimulai dengan penentuan jumlah klaster optimal (k) menggunakan metode *Elbow*. Metode ini mengevaluasi titik infleksi ("siku") terbaik dari penurunan metrik *Sum of Squared Errors* (SSE) seiring bertambahnya jumlah klaster. Nilai SSE dihitung berdasarkan jarak kuadrat setiap titik data terhadap pusat klasternya menggunakan Persamaan 2:

$$SSE = \sum_{i=1}^{k} \sum_{x \in C_i} (x - \mu_i)^2 \quad (2)$$

Dalam penelitian ini, jumlah cluster (k) ditentukan berdasarkan metode Elbow. Proses clustering dilakukan secara iteratif hingga mencapai kondisi konvergen dengan batas maksimum iterasi tertentu. Implementasi algoritma dilakukan menggunakan bahasa pemrograman Python dengan library *Scikit-learn*.

Evaluasi dilakukan untuk mengetahui kualitas cluster yang dihasilkan, salah satunya dengan melihat nilai SSE. Nilai SSE yang lebih kecil menunjukkan bahwa data dalam satu cluster memiliki tingkat kemiripan yang tinggi. Selain SSE, evaluasi clustering dapat menggunakan metrik lain seperti Silhouette Score untuk mengukur kualitas pemisahan antar cluster.

Tahap akhir adalah analisis dan interpretasi hasil clustering. Pada tahap ini, setiap cluster dianalisis untuk mengetahui karakteristiknya, seperti wilayah dengan suhu tinggi, tekanan rendah, atau kecepatan angin tertentu. Hasil ini kemudian digunakan untuk memahami pola cuaca yang terbentuk di wilayah Sumatera. Hasil clustering berupa label cluster pada setiap data yang menunjukkan pengelompokan pola cuaca berdasarkan karakteristik variabel yang digunakan.

Tabel 2. Variabel dataset

Nomor	Variabel	Keterangan
1	t2m	Suhu permukaan (*temperature 2 meter*)
2	sp	Tekanan udara permukaan (*surface pressure*)
3	u10	Komponen Kecepatan angin arah u
4	v10	Komponen kecepatan angin arah v

Tabel 2 menunjukkan variabel yang digunakan dalam penelitian ini. Variabel tersebut dipilih karena dapat merepresentasikan kondisi atmosfer secara umum. Suhu udara digunakan untuk menggambarkan kondisi termal, tekanan udara untuk menunjukkan dinamika atmosfer, serta komponen angin (u10 dan v10) untuk merepresentasikan pergerakan massa udara secara horizontal.

2.2 Metode K-Means Clustering

Metode *K-Means* merupakan salah satu algoritma *unsupervised learning* yang digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan karakteristik. Metode K-Means dipilih karena memiliki keunggulan dalam kesederhanaan, efisiensi komputasi, serta kemampuan dalam menangani dataset berukuran besar seperti data ERA5. Metode ini bekerja dengan menentukan pusat cluster (*centroid*) dan mengelompokkan data berdasarkan jarak terdekat terhadap centroid tersebut. Jarak Euclidean digunakan untuk mengukur kedekatan antara data dan centroid, di mana semakin kecil jarak maka data akan dimasukkan ke dalam cluster tersebut. Proses dalam metode *K-Means* dilakukan secara iteratif hingga posisi centroid stabil. Tahapan metode ini meliputi penentuan jumlah cluster, inisialisasi centroid, perhitungan jarak, pengelompokan data, serta pembaruan centroid hingga konvergen. Perhitungan jarak antara data dan centroid menggunakan jarak Euclidean yang dirumuskan sebagai berikut:

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^{n} (x_{ik} - c_{jk})^2} \quad (3)$$

Fungsi objektif dari metode *K-Means* adalah meminimalkan jumlah kuadrat jarak dalam cluster yang dirumuskan sebagai berikut:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i - c_j \right\|^2 \quad (4)$$

3. HASIL DAN PEMBAHASAN

Pada bagian ini, dipaparkan serangkaian hasil komputasi dan penemuan (*knowledge discovery*) dari penerapan algoritma *Unsupervised Learning* K-Means terhadap dataset cuaca reanalisis ERA5 di kawasan Sumatera dan sekitarnya. Pembahasan dalam bab ini disusun secara runut untuk menjawab tujuan penelitian, diawali dengan tahapan pra-pemrosesan dan eksplorasi data (*Exploratory Data Analysis*), penentuan parameter klaster optimal secara matematis, hingga visualisasi pemetaan spasial dari label yang terbentuk. Lebih lanjut, bab ini juga menyajikan validasi model menggunakan *Principal Component Analysis* (PCA) dan interpretasi mendalam terhadap profil karakteristik meteorologis (suhu, tekanan, dan angin) pada masing-masing

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 168-177
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.945

zona klaster. Keseluruhan tahapan ini dieksekusi untuk membuktikan sejauh mana algoritma *Machine Learning* mampu mengekstraksi pola laten batas-batas geografis murni dari variansi data numerik cuaca historis.

3.1 Eksplorasi Data Awal

Analisis dalam penelitian ini diawali dengan melakukan eksplorasi deskriptif terhadap dataset reanalisis atmosfer ERA5. Tahapan ini sangat krusial guna memahami karakteristik distribusi data serta memastikan validitas informasi sebelum diolah oleh algoritma *machine learning*. Berdasarkan hasil evaluasi statistik deskriptif, dipastikan bahwa seluruh dataset tidak memiliki nilai kosong (*missing values*). Ketiadaan data yang hilang ini merepresentasikan tingkat integritas dan kualitas data yang sangat tinggi, sehingga dataset dinyatakan layak untuk diproses lebih lanjut tanpa memerlukan tahapan imputasi data yang kompleks.

Dataset yang diolah dalam penelitian ini memiliki volume yang signifikan, yaitu mencakup 1.547.616 baris observasi yang terbagi ke dalam 6 kolom fitur utama. Besarnya jumlah sampel ini memberikan cakupan spasial dan temporal yang sangat rapat, yang memungkinkan model untuk menangkap dinamika cuaca di kawasan Sumatera secara mendetail. Representasi sebagian kecil dari dataset yang telah dibersihkan dan siap olah disajikan pada **Tabel 3.**

Tabel 3. Sebagian data yang digunakan pada penelitian ini

latitude	longitude	t2m	sp	u10	v10
6.0	95.00	300.81300	101168.44	-5.624283	-2.132980
6.0	95.25	300.86182	101186.44	-5.433853	-1.257980
6.0	95.50	300.78564	101233.44	-4.668228	-0.316574

Berikut merupakan rincian karakteristik dan parameter meteorologis dari dataset ERA5 yang digunakan dalam penelitian ini

a. **Latitude dan Longitude:**
Merupakan koordinat geografis yang menentukan lokasi titik *grid* di wilayah Sumatera. Data ini memiliki resolusi spasial sebesar $0{,}25^\circ \times 0{,}25^\circ$, yang berarti setiap titik mewakili luasan area sekitar 27-28 km. Akurasi koordinat ini sangat krusial untuk memastikan bahwa hasil klasterisasi mampu merepresentasikan kondisi topografi nyata di lapangan. b.

b. **u10 (10m u-component of wind):**
Merepresentasikan komponen horizontal angin pada ketinggian 10 meter di atas permukaan tanah yang bergerak dari arah Barat ke Timur (zona zonal). Parameter ini digunakan untuk menganalisis pengaruh massa udara yang datang dari Samudra Hindia menuju daratan Sumatera.

c. **v10 (10m v-component of wind):**
Merepresentasikan komponen vertikal angin pada ketinggian 10 meter yang bergerak dari arah Selatan ke Utara (zona meridional). Kombinasi antara fitur $u10$ dan $v10$ memungkinkan algoritma K-Means untuk mengidentifikasi pola sirkulasi angin regional di sepanjang jalur khatulistiwa.

d. **t2m (2m temperature):**
Merupakan suhu udara pada ketinggian 2 meter di atas permukaan tanah. Dalam dataset mentah, suhu tercatat dalam satuan Kelvin (K). Namun dalam penelitian ini, kami akan menggunakan satuan celsius (℃), Fitur ini memegang peranan vital dalam klasterisasi karena suhu memiliki gradien yang sangat tegas antara wilayah pesisir yang hangat dan wilayah pegunungan yang dingin.

e. **sp (Surface pressure):**
Menunjukkan tekanan udara tepat di permukaan tanah. Berbeda dengan tekanan permukaan laut (*mean sea level pressure*), parameter *sp* sangat sensitif terhadap ketinggian tempat (elevasi). Integrasi variabel ini memungkinkan model untuk mengenali struktur Pegunungan Bukit Barisan melalui perbedaan tekanan barometrik yang terekam secara spasial.

Tabel 4. Statistik Deskriptif Dataset Reanalisis ERA5

Statistik	latitude	longitude	t2m (∘C)	sp (Pa)	u10 (m/s)	v10 (m/s)
Total	1.547.616	1.547.616	1.547.616	1.547.616	1.547.616	1.547.616
Mean	0	100.75	26.58	100002.6	1.02	-1.88
Std	3.54	3.39	2.21	2559.98	3	2.44
Min	-6	95	11.82	82209.81	-10.79	-10.95
25%	-3	97.75	25.79	100540.9	-0.84	-3.43
50%	0	100.75	26.87	100881.9	0.69	-1.47

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 168-177
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.945

75%	3	103.75	27.86	101084.3	2.89	-0.19
Max	6	106.5	37.36	102683.9	11.58	10.06

Beberapa poin penting yang dapat disimpulkan dari data tersebut adalah sebagai berikut:

a. **Distribusi Spasial:**
Nilai rata-rata (*mean*) pada kolom *latitude* sebesar **0,00** mengonfirmasi bahwa domain penelitian berada tepat di garis khatulistiwa dan terdistribusi secara simetris antara belahan bumi utara dan selatan (6°LU hingga 6°LS).

b. **Variansi Suhu:**
Suhu udara (*t2m*) menunjukkan rentang yang cukup lebar, dengan nilai minimum mencapai 11,82℃ dan maksimum 37,36 ℃. Adanya nilai suhu yang sangat rendah ini menjadi indikator kuat keberadaan area dengan elevasi tinggi (pegunungan), mengingat suhu rata-rata di permukaan laut kawasan tropis biasanya berada di atas 25℃

c. **Tekanan Permukaan:**
Kolom *surface pressure* (*sp*) memiliki standar deviasi yang signifikan sebesar 2559,98 Pa. Hal ini menunjukkan variasi tekanan yang dipengaruhi oleh perbedaan ketinggian tempat, yang nantinya akan menjadi pembeda utama dalam penentuan klaster geografis.

Proses klasterisasi dalam penelitian ini difokuskan pada integrasi fitur-fitur meteorologis utama yang meliputi suhu udara (*t2m*), tekanan permukaan (*sp*), serta komponen kecepatan angin zonal (*u10*) dan meridional (*v10*). Setelah label klaster terbentuk, langkah selanjutnya adalah memproyeksikan hasil tersebut ke dalam koordinat geografis yang sesuai untuk menghasilkan pemetaan zona cuaca secara spasial.

3.2. Penentuan Jumlah Klaster dan Evaluasi Model

Tahap pemodelan merupakan inti dari penelitian ini, di mana algoritma K-Means diterapkan untuk mengidentifikasi pola tersembunyi dari data multivariat cuaca di Sumatera.

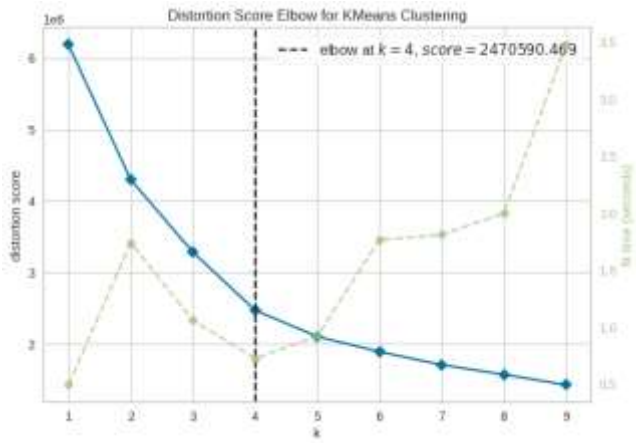
3.2.1 Standarisasi Data

Mengingat fitur-fitur meteorologis yang digunakan memiliki satuan dan rentang nilai yang sangat kontras—misalnya tekanan permukaan (*sp*) yang bernilai ratusan ribu Pascal dibandingkan dengan kecepatan angin (*u10* dan *v10*) yang hanya bernilai satuan—maka tahap standarisasi menjadi prasyarat mutlak. Seluruh fitur ditransformasikan menggunakan metode *Z-score Standardization*

Proses ini memastikan bahwa setiap variabel memiliki bobot yang setara dalam perhitungan jarak *Euclidean* pada algoritma K-Means, sehingga fitur dengan skala besar tidak mendominasi fitur lainnya.

3.2.2 Penentuan Jumlah Klaster (Elbow Method)

Setelah data berada pada skala yang seragam, langkah selanjutnya adalah menentukan jumlah klaster optimal (*k*) menggunakan metode *Elbow*. Evaluasi ini didasarkan pada nilai *Distortion Score* yang merepresentasikan total kuadrat jarak antara setiap titik data dengan pusat klasternya (*centroid*).



Gambar 2. Kurva Evaluasi Metode *Elbow* untuk Penentuan Jumlah Klaster Optimal.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 168-177
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.945

Berdasarkan hasil visualisasi pada **Gambar 2**, terlihat adanya penurunan nilai distorsi yang sangat signifikan dari *k=1* hingga *k=4*. Titik infleksi atau "siku" (*elbow*) terbentuk secara tegas pada nilai ***k=4***, di mana setelah titik tersebut, penurunan nilai distorsi mulai melandai secara bertahap (*plateau*). Hasil ini secara matematis mengonfirmasi bahwa pembagian wilayah Sumatera ke dalam empat zona klaster merupakan konfigurasi yang paling optimal untuk dataset ini.

3.2.3 Klasterisasi dan Evaluasi Model

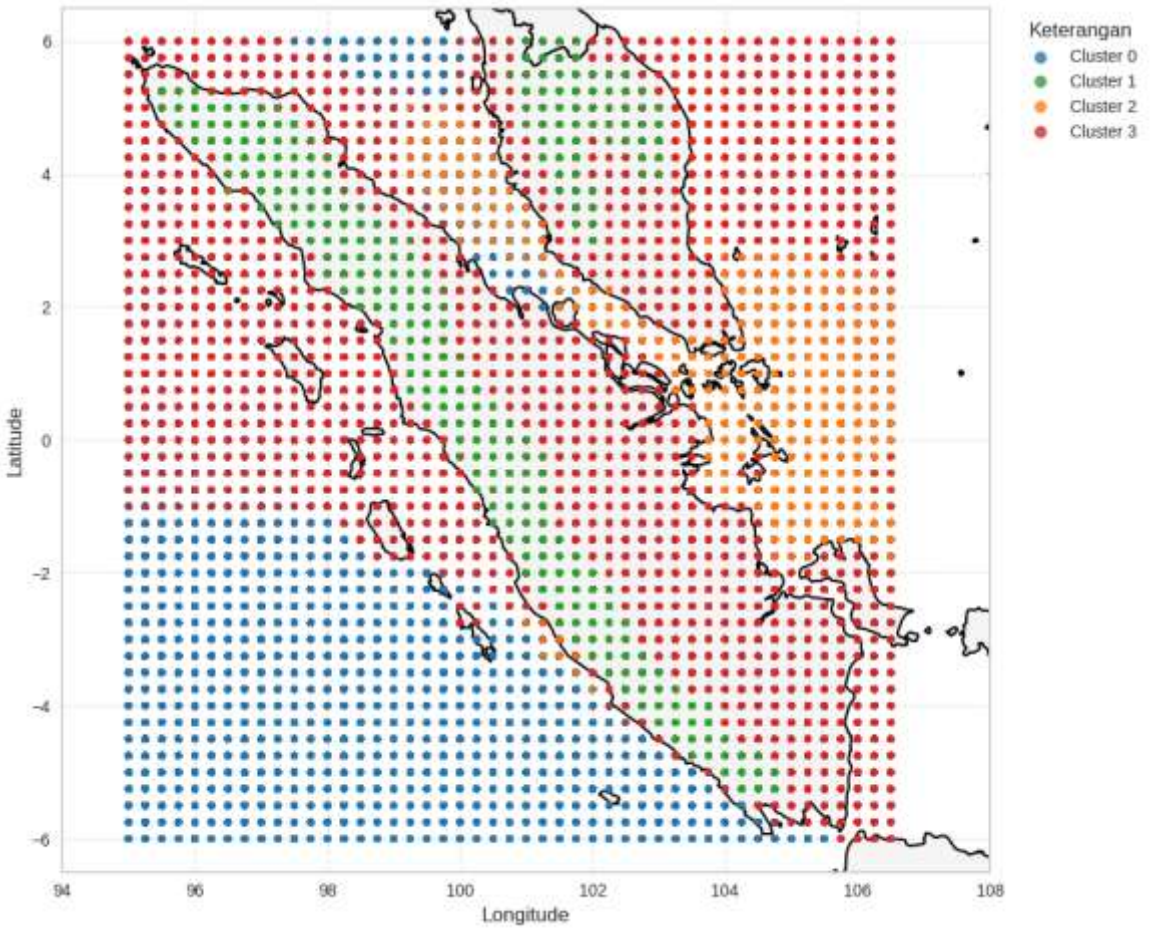
Berdasarkan parameter *k=4* yang telah ditetapkan melalui metode *Elbow*, algoritma K-Means dieksekusi untuk melabeli keseluruhan 1.547.616 baris data observasi. Setiap titik data dipetakan ke dalam salah satu dari empat klaster berdasarkan jarak terdekatnya dengan titik pusat (*centroid*).

Setelah proses pelabelan selesai, kualitas pemisahan dan kohesi antar klaster dievaluasi secara kuantitatif menggunakan metrik *Silhouette Coefficient*. Metrik ini mengukur seberapa serupa suatu objek data dengan klasternya sendiri (*cohesion*) dibandingkan dengan klaster lainnya (*separation*). Mengingat volume dataset yang sangat masif, komputasi *Silhouette Score* dilakukan dengan teknik *sampling* terhadap 20% sampel acak dari total populasi data dengan menetapkan parameter *random state* guna memastikan hasil yang dapat direproduksi (*reproducible*). Penggunaan porsi 20% ini secara statistik telah memenuhi representativitas distribusi data populasi sekaligus menjaga efisiensi memori komputasi.

Berdasarkan hasil pengujian pada sampel tersebut, model menghasilkan nilai *Silhouette Score* sebesar **0,3078**. Nilai positif ini mengonfirmasi bahwa algoritma K-Means telah berhasil membentuk struktur klaster yang valid secara empiris. Dalam konteks data atmosferik dan meteorologis yang sifat fluida-nya kontinu serta memiliki banyak irisan batas wilayah (*overlapping transition zones*), skor sebesar 0,3078 merupakan indikator yang kuat bahwa keempat zona cuaca di kawasan Sumatera terpisah menjadi kelompok-kelompok yang koheren dan bukan terbentuk secara acak.

3.3 Pemetaan Spasial Klaster Cuaca

Hasil prediksi label klaster kemudian dipetakan kembali ke dalam sistem koordinat geografis (Latitude dan Longitude) untuk melihat persebaran spasialnya. Pemetaan ini menghasilkan zonasi yang merepresentasikan karakteristik orografis dan geografis yang sangat tegas:



Gambar 3. Hasil Pemetaan Geografis Klaster Cuaca ERA5 di Wilayah Kajian.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 168-177
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.945

Berdasarkan pemetaan spasial pada Gambar X, algoritma K-Means terbukti mampu mendelineasi (memisahkan) wilayah kajian ke dalam empat zona iklim mikro yang sangat representatif terhadap topografi nyata di lapangan. Rincian karakteristik geografis dari masing-masing klaster adalah sebagai berikut:

a. **Klaster 0 (Zona Samudra Hindia / Perairan Terbuka):**
Secara spasial mendominasi wilayah perairan laut lepas di sebelah pesisir barat dan selatan Pulau Sumatera. Zona ini merepresentasikan karakteristik cuaca maritim murni dengan dinamika atmosfer yang sangat dipengaruhi oleh suhu permukaan laut samudra.
b. **Klaster 1 (Zona Orografis / Pegunungan):**
Membentuk pola klaster linier yang secara presisi membelah bagian tengah daratan Sumatera hingga menjangkau daratan Semenanjung Malaysia. Pola ini secara akurat merepresentasikan jajaran topografi tinggi, yakni Pegunungan Bukit Barisan dan Banjaran Titiwangsa, yang berhasil diidentifikasi oleh model berdasarkan gradien suhu yang lebih rendah dan anomali tekanan permukaan.
c. **Klaster 2 (Zona Dataran Rendah Timur dan Selat Malaka):**
Membentang secara ekstensif menutupi hamparan pesisir timur Sumatera, melintasi perairan dangkal Selat Malaka, hingga mencapai pesisir barat daratan Malaysia. Zona ini mewakili karakteristik meteorologis dataran rendah beriklim tropis yang memiliki sebaran suhu yang lebih homogen dan hangat.
d. **Klaster 3 (Zona Transisi Pesisir Barat):**
Terkonsentrasi di sepanjang garis pesisir barat Sumatera dan gugusan pulau-pulau kecil di sekitarnya (seperti Kepulauan Nias dan Mentawai). Zona ini bertindak sebagai sabuk batas (*buffer zone*) yang menangkap transisi dinamika cuaca antara massa udara lembap dari samudra lepas (Klaster 0) sebelum menabrak penghalang topografi pegunungan (Klaster 1).

Keberhasilan algoritma memetakan keempat zona geografis ini secara mandiri (*unsupervised*) mengonfirmasi bahwa parameter meteorologis seperti suhu dan tekanan memiliki korelasi spasial yang sangat identik dengan elevasi dan bentuk muka bumi.

3. 4 Analisis Profil Cuaca Tiap Klaster

Untuk memvalidasi partisi zona geografis yang telah divisualisasikan secara spasial pada tahap sebelumnya, dilakukan ekstraksi dan analisis terhadap nilai pusat (*centroid*) dari masing-masing klaster. Berbeda dengan proses pemodelan yang menggunakan data terstandarisasi, ekstraksi *centroid* pada tahap ini dikembalikan ke dalam skala nilai observasi aslinya (*original values*). Pengembalian nilai historis ini bertujuan untuk menginterpretasikan batas-batas klaster berdasarkan parameter fisika atmosfer yang nyata di lapangan. Rincian profil meteorologis dari setiap klaster disajikan pada Tabel X.

Tabel 5. Nilai *Centroid* Masing-masing Klaster pada Fitur Meteorologis

Klaster	Suhu (t2m)	Tekanan (sp)	Angin Zonal (u10)	Angin Meridional (v10)
0	27,46 °C	100.881,42 Pa	4,50 m/s	-2,72 m/s
1	21,94 °C	93.058,05 Pa	0,02 m/s	-0,21 m/s
2	26,35 °C	101.021,76 Pa	-1,97 m/s	-5,13 m/s
3	27,03 °C	100.416,29 Pa	0,14 m/s	-0,29 m/s

Berdasarkan metrik pada **Tabel 5**, profil meteorologis masing-masing zona dapat divalidasi dan diinterpretasikan secara fisis sebagai berikut:

a. **Klaster 1 (Zona Orografis/Pegunungan):**
Memiliki karakteristik ekstrem yang sangat mencolok dengan rata-rata suhu udara terendah mencapai 21,94℃ dan tekanan permukaan anjlok hingga 93.058,05 Pa. Profil ini secara mutlak mengonfirmasi Hukum Termodinamika Atmosfer lapse rate, di mana jajaran topografi tinggi seperti Pegunungan Bukit Barisan secara alami memiliki suhu dan tekanan barometrik yang jauh lebih rendah dibandingkan wilayah sekitarnya.
b. **Klaster 0 (Zona Samudra Hindia):**
Ditandai dengan intensitas kecepatan angin zonal (u10) dari arah barat yang sangat tinggi, yakni mencapai 4,50 m/s, serta suhu yang hangat (27,4℃). Angka kecepatan angin yang dominan ini secara akurat merepresentasikan dinamika lautan terbuka di pesisir barat yang menerima paparan sirkulasi monsun secara langsung tanpa adanya friksi atau hambatan dari topografi daratan.
c. **Klaster 2 (Zona Dataran Rendah Timur & Selat Malaka):**

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 168-177
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.945

Memiliki nilai tekanan permukaan tertinggi (101.021,76 Pa) yang mengindikasikan area dengan elevasi paling rendah atau sangat mendekati Mean Sea Level (MSL). Zona ini memiliki karakteristik pola pergerakan angin meridional (v10) yang bernilai negatif sangat kuat (-5,13 m/s), mencerminkan dominasi sirkulasi angin yang bertiup stabil dari arah utara di sepanjang pesisir timur Sumatera dan Selat Malaka.

d. **Klaster 3 (Zona Transisi Pesisir Barat):**

Menunjukkan profil angka yang paling ekuilibrium. Suhu terpantau normal di angka 27,03℃ dengan kondisi angin yang sangat tenang (*u10* sebesar 0,14 m/s dan *v10* sebesar -0,29 m/s). Karakteristik moderat ini sangat logis dan sesuai dengan letak geografisnya yang berhimpitan dengan garis pantai, menjadikannya zona penyangga (*buffer zone*) tempat bertemunya massa udara dari samudra terbuka sebelum menabrak penghalang topografi daratan.

4. KESIMPULAN

Penelitian ini berhasil membuktikan bahwa penggunaan algoritma *K-Means clustering* yang berbasis pada data reanalisis ERA5 merupakan metode yang sangat efektif untuk memetakan pola cuaca spasial di wilayah Sumatera secara sistematis. Melalui serangkaian tahapan ilmiah mulai dari pembersihan data hingga standarisasi menggunakan *Z-score*, algoritma ini mampu mengolah jutaan baris data meteorologi yang kompleks untuk menemukan pola tersembunyi tanpa memerlukan label data sebelumnya. Salah satu temuan paling krusial dalam studi ini adalah kuatnya korelasi antara suhu udara dan tekanan permukaan yang mencapai nilai 0,60, di mana sinyal ini menjadi kunci utama bagi model untuk mengenali struktur geografis seperti Pegunungan Bukit Barisan secara mandiri. Hal ini menunjukkan bahwa dinamika atmosfer yang terekam dalam data numerik memiliki keterkaitan fisik yang sangat erat dengan kondisi topografi nyata di lapangan. Hasil pengelompokan yang divalidasi melalui metode *Elbow* dan *Silhouette Score* menunjukkan bahwa pembagian wilayah ke dalam empat klaster utama adalah konfigurasi yang paling optimal. Keempat zona tersebut secara akurat merepresentasikan karakteristik wilayah yang berbeda, yakni Zona Samudra Hindia dengan angin zonal yang kuat, Zona Orografis atau pegunungan dengan suhu dan tekanan rendah yang ekstrem, Zona Dataran Rendah Timur yang hangat, serta Zona Transisi Pesisir Barat yang bertindak sebagai penyangga massa udara. Pemisahan yang tegas pada visualisasi *Principal Component Analysis* (PCA) semakin memperkuat bukti bahwa klaster yang terbentuk bukan sekadar hasil acak, melainkan representasi dari struktur data atmosfer yang koheren. Secara keseluruhan, penerapan *machine learning* ini memberikan pemahaman mendalam mengenai variabilitas cuaca regional yang sangat bermanfaat untuk mendukung pengambilan keputusan di berbagai sektor strategis seperti mitigasi bencana, pertanian, dan pengelolaan sumber daya air di Indonesia.

REFERENCES

[1] S. C. Peatman, J. Schwendike, C. E. Birch, J. H. Marsham, A. J. Matthews, and G. Yang, “A Local-to-Large Scale View of Maritime Continent Rainfall: Control by ENSO, MJO, and Equatorial Waves”, doi: 10.1175/JCLI-D-21.

[2] E. A. Reddy and K. S. Rajan, “Spatiotemporal Cluster Analysis of Gridded Temperature Data-A Comparison Between K-means and MiSTIC,” *International Journal of Scientific Research and Engineering Development*, vol. 6, [Online]. Available: www.ijsred.com

[3] H. Hersbach *et al.*, “The ERA5 global reanalysis,” *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, Jul. 2020, doi: 10.1002/qj.3803.

[4] M. Putra, M. S. Rosid, and D. Handoko, “High-Resolution Rainfall Estimation Using Ensemble Learning Techniques and Multisensor Data Integration,” *Sensors*, vol. 24, no. 15, Aug. 2024, doi: 10.3390/s24155030.

[5] H. Sitepu, D. Harisuseno, and J. S. Fidari, “Evaluasi Data Curah Hujan Satelit ERA-5 pada Berbagai Periode Data Hujan di Sub DAS Bodor Evaluation of ERA5 Satellite Rainfall Data at Various Rainfall Data Periods in Bodor Sub Watershed,” *Jurnal Teknologi dan Rekayasa Sumber Daya Air*, vol. 03, no. 02, pp. 626–636, 2023, doi: 10.21776/ub.jtresda.003.vol.no02.053.

[6] Uston Nawawi Christanto, Brina Miftahurrohmah, T. Bariyah, H. Kuswanto, and N. Faria, “CLUSTER-BASED MACHINE LEARNING APPROACHES FOR PREDICTING DAILY MAXIMUM TEMPERATURES IN INDONESIA UNDER CLIMATE CHANGE,” *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 11, no. 1, pp. 236–249, Aug. 2025, doi: 10.33480/jitk.v11i1.6749.

[7] Ayu Aprilia, Alka Budi Wahidin, Syafriadi Syafriadi, and Pulung Karo Karo, “Perbandingan Algoritma Machine Learning (Logistic Regression, SVM, KNN, Decision Tree, Random Forest, dan Gradient Boosting) dalam Prediksi Hujan Harian di Provinsi Lampung,” *Jurnal ilmiah Sistem Informasi dan Ilmu Komputer*, vol. 5, no. 3, pp. 755–764, Nov. 2025, doi: 10.55606/juisik.v5i3.1901.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 168-177
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.945

[8] N. P. Putri, A. H. Saputro, R. Prasetya, A. A. Soebroto, and P. Korespondensi, "Penerapan Model Arsitektur UNet untuk Peningkatan Resolusi Spasial Curah Hujan di Wilayah Pulau Jawa Berbasis Data MSWEP APPLICATION OF UNET ARCHITECTURE MODEL TO IMPROVE SPATIAL RESOLUTION OF RAINFALL IN JAVA ISLAND REGION BASED ON MSWEP DATA," vol. 13, no. 1, pp. 83–94, 2026, [Online]. Available: https://www.gloh2o.org/mswep/

[9] L. Glawion, J. Polz, H. Kunstmann, B. Fersch, and C. Chwala, "Global spatio-temporal ERA5 precipitation downscaling to km and sub-hourly scale using generative AI," *NPJ Clim. Atmos. Sci.*, vol. 8, no. 1, Dec. 2025, doi: 10.1038/s41612-025-01103-y.

[10] X. Man, C. Zhang, J. Feng, C. Li, and J. Shao, "W-MAE: Pre-trained weather model with masked autoencoder for multi-variable weather forecasting," Dec. 2023, [Online]. Available: http://arxiv.org/abs/2304.08754

[11] T. M. Ponjiger *et al.*, "Evaluation of Rainfall Erosivity in the Western Balkans by Mapping and Clustering ERA5 Reanalysis Data," *Atmosphere (Basel).*, vol. 14, no. 1, 2023, doi: 10.3390/atmos14010104.

[12] A. Boulin, E. Di Bernardino, T. Laloë, and G. Toulemonde, "Identifying regions of concomitant compound precipitation and wind speed extremes over Europe," Nov. 2023, [Online]. Available: http://arxiv.org/abs/2311.11292

[13] A. Dowdy, A. Brown, T. P. Lane, and M. Taszarek, "Climatological variability of a thunderstorm environment dataset in tropical and temperate regions." doi: DOI:10.1007/s00382-026-08076-5.

[14] O. Kisi, S. Heddam, K. S. Parmar, A. Petroselli, C. Külls, and M. Zounemat-Kermani, "Integration of Gaussian process regression and K means clustering for enhanced short term rainfall runoff modeling," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-91339-8.

[15] S. C. Hicks, R. Liu, Y. Ni, E. Purdom, and D. Risso, "mbkmeans: Fast clustering for single cell data using mini-batch k-means," *PLoS Comput. Biol.*, vol. 17, no. 1, p. e1008625, Jan. 2021, doi: 10.1371/journal.pcbi.1008625.

[16] Giarno *et al.*, "CLUSTERING-BASED EVALUATION OF SATELLITE RAINFALL PRODUCTS: A NOVEL PERSPECTIVE," *Bulletin of the Serbian Geographical Society*, vol. 105, no. 2, pp. 501–524, 2025, doi: 10.2298/GSGD2502501G.

[17] H. Firdausi, M. Ariska, S. Markos Siahaan, H. Akhsan, Y. Anwar, and I. Seprina, "Machine Learning untuk Memprediksi Perubahan Iklim Wilayah Pesisir Pantai Indonesia Machine Learning to Predict Climate Change in Coastal Areas of Indonesia." doi: https://doi.org/10.24843/BF.2026.v27.i01.p05.

[18] F. A. Sari, M. Y. N. Khakim, B. Setiawan, and P. P. Simanjuntak, "Pengaruh Variabilitas Iklim Terhadap Kesesuaian Lahan Lada (Piper nigrum L.) Berbasis Analisis Spasial di Kepulauan Bangka Belitung," *Jurnal Ilmu-Ilmu Pertanian Indonesia*, vol. 27, no. 2, pp. 148–155, Dec. 2025, doi: 10.31186/jipi.27.2.148-155.

[19] A. Chaqdid, A. Tuel, A. El Fatimy, and N. El Moçayd, "Toward Reducing Uncertainty in Simulating Temporal Clustering of Extreme Precipitation in Morocco: Insights from High-Resolution GCMs," *J. Clim.*, vol. 39, no. 6, pp. 1407–1431, Mar. 2026, doi: 10.1175/JCLI-D-25-0138.1.

[20] L. A. Pampuch, R. G. Negri, P. C. Loikith, and C. A. Bortolozo, "A Review on Clustering Methods for Climatology Analysis and Its Application over South America," *International Journal of Geosciences*, vol. 14, no. 09, pp. 877–894, 2023, doi: 10.4236/ijg.2023.149047.

[21] Q. Van Doan, T. Amagasa, T. H. Pham, T. Sato, F. Chen, and H. Kusaka, "Structural k-means (S k-means) and clustering uncertainty evaluation framework (CUEF) for mining climate data," *Geosci. Model Dev.*, vol. 16, no. 8, pp. 2215–2233, Apr. 2023, doi: 10.5194/gmd-16-2215-2023.

[22] E. A. Reddy and K. S. Rajan, "Spatiotemporal Cluster Analysis of Gridded Temperature Data-A Comparison Between K-means and MiSTIC," *International Journal of Scientific Research and Engineering Development*, vol. 6, [Online]. Available: www.ijsred.com

[23] A. Lojko, A. C. Winters, A. Oertel, C. Jablonowski, and A. E. Payne, "An ERA5 climatology of synoptic-scale negative potential vorticity–jet interactions over the western North Atlantic," *Weather and Climate Dynamics*, vol. 6, no. 2, pp. 387–411, Apr. 2025, doi: 10.5194/wcd-6-387-2025.