

Analisis Spasial Klaster Atmosfer di Sumatera Utara Menggunakan *K-Means*

Jhon Gabriel Simarmata¹, Natasha Patricia Nainggolan^{2*}, Elga Sari Tanjung³

^{1,2,3}Fakultas Matematika dan Ilmu Pengetahuan Alam, Program Studi Ilmu Komputer, Universitas Negeri Medan, Medan, Indonesia

Email: ¹rielsimarmata06@gmail.com, ^{2*}nattadsah@gmail.com, ³elgasaritanjung123@gmail.com

(*Email Corresponding Author: nattadsah@gmail.com)

Received: 17 Maret 2026 | Revision: 20 Juni 2026 | Accepted: 25 Juni 2026

Abstrak

Variabilitas atmosfer jangka pendek di Sumatera Utara pada periode puncak monsun belum dipetakan secara kuantitatif dan spasial. Penelitian ini bertujuan mengidentifikasi dan mengklasifikasikan pola atmosfer pada 28 Desember 2025–1 Januari 2026 menggunakan metode *K-Means clustering* sebagai solusi analitis berbasis data. Data yang digunakan berasal dari *Modern-Era Retrospective Analysis for Research and Applications* Version 2 (MERRA-2) dengan variabel suhu udara permukaan, kelembapan spesifik, tekanan permukaan, dan kecepatan angin. Penelitian menggunakan pendekatan kuantitatif deskriptif dengan tahapan pra-proses data, normalisasi, optimasi jumlah cluster menggunakan Elbow Method, serta evaluasi menggunakan *Silhouette Score*. Jumlah cluster optimal adalah tiga dengan nilai *Silhouette Score* global sebesar 0,4089 yang menunjukkan struktur cluster moderat. Hasil penelitian mengidentifikasi tiga pola atmosfer dominan, yaitu Panas Tropis & Sangat Lembap (43,44%), Hangat & Sangat Lembap (36,93%), dan Sejuk & Lembap Sedang (19,63%). Secara spasial, wilayah dataran tinggi berkorelasi dengan kondisi lebih sejuk, sedangkan wilayah pesisir dan kepulauan didominasi kondisi panas dan sangat lembap. Temuan ini menunjukkan bahwa metode *K-Means* efektif dalam memetakan variabilitas atmosfer spasial jangka pendek secara kuantitatif.

Kata Kunci: *K-Means clustering*, Variabilitas Atmosfer, MERRA-2, Analisis Spasial, Sumatera Utara

Abstract

Short-term atmospheric variability in North Sumatra during peak monsoon periods has not been quantitatively mapped in a spatially explicit manner. This study aims to identify and classify spatial atmospheric patterns during 28 December 2025–1 January 2026 using K-Means clustering as a data-driven solution. The analysis utilizes the Modern-Era Retrospective Analysis for Research and Applications Version 2 (MERRA-2) Dataset, incorporating surface air temperature, specific humidity, surface pressure, and wind speed variables. A quantitative descriptive approach with unsupervised Machine learning was applied, including data preprocessing, normalization, Elbow Method optimization, and Silhouette Score evaluation. The optimal number of clusters was three, with a global Silhouette Score of 0.4089 indicating moderate cluster structure. The results reveal three dominant atmospheric patterns: Hot Tropical & Very Humid (43.44%), Warm & Very Humid (36.93%), and Cool & Moderately Humid (19.63%). Spatially, highland areas correspond to cooler clusters, while coastal and island regions are dominated by hot and very humid conditions. These findings demonstrate that K-Means clustering effectively captures short-term spatial atmospheric variability and provides a quantitative basis for Regional atmospheric characterization.

Keywords: *K-Means clustering*, Atmospheric Variability, MERRA-2, Spatial Analysis, North Sumatra

1. PENDAHULUAN

Provinsi Sumatera Utara merupakan salah satu wilayah strategis di Indonesia yang memiliki karakteristik atmosfer sangat dinamis karena posisi geografisnya yang terletak tepat di kawasan ekuatorial. Secara fisik, wilayah ini diapit oleh dua massa perairan besar yang sangat berpengaruh terhadap sistem sirkulasi udara *Regional*, yaitu Samudra Hindia di sisi barat dan Selat Malaka di sisi timur. Kondisi posisi geografis yang unik ini menjadikan Sumatera Utara sangat sensitif terhadap berbagai fenomena meteorologi ekstrem, terutama pada periode transisi musim yang sering kali ditandai dengan fluktuasi parameter cuaca yang tajam dan tidak menentu. Dinamika atmosfer di wilayah ini dipengaruhi oleh interaksi sirkulasi monsun skala global dan aktivitas konveksi tropis yang membentuk pola cuaca dengan variabilitas spasial tinggi. Keragaman topografi, mulai dari dataran rendah pesisir hingga pegunungan Bukit Barisan, menciptakan profil kondisi atmosfer yang berbeda secara signifikan antar sub-wilayah. Pemahaman mendalam mengenai karakteristik cuaca lokal ini sangat krusial guna mendukung sektor strategis seperti ketahanan pangan, khususnya di wilayah produktif seperti Kabupaten Asahan dan Karo yang sangat bergantung pada stabilitas iklim *Regional* [1]. Selain itu, analisis pola cuaca di tingkat provinsi menjadi landasan penting dalam mengidentifikasi anomali atmosfer yang dapat memicu risiko lingkungan [2].

Fenomena perubahan pola atmosfer jangka pendek sering kali berdampak langsung pada peningkatan intensitas bencana hidrometeorologi. Kota Medan, sebagai pusat populasi terbesar di Sumatera Utara dengan luas wilayah mencapai 26.510 hektar, telah diklasifikasikan sebagai daerah rawan banjir yang memerlukan perhatian serius melalui

analisis zonasi daerah terdampak secara mendetail [3]. Kejadian banjir di wilayah perkotaan ini secara teknis dipengaruhi oleh kondisi hidrologis hilir sungai Deli dan Belawan, di mana frekuensi kejadian genangan dilaporkan dapat terjadi secara berulang dalam kurun waktu satu tahun [3]. Permasalahan serupa juga teramati di pusat populasi lainnya di Indonesia, seperti Kota Bengkulu, di mana identifikasi daerah rawan banjir dilakukan melalui analisis pengelompokan data meteorologi guna mendukung sistem mitigasi bencana yang lebih terukur [4]. Ketersediaan informasi pola atmosfer yang akurat pada periode kritis, seperti saat pergantian tahun antara 28 Desember 2025 hingga 1 Januari 2026, menjadi kebutuhan mendesak untuk mendukung perencanaan keselamatan publik dan strategi pengurangan risiko bencana secara efektif bagi masyarakat luas [3], [4].

Sebagai solusi atas kompleksitas data meteorologi tersebut, pendekatan *data mining* berbasis klusterisasi melalui algoritma *K-Means clustering* diterapkan dalam penelitian ini. Penggunaan algoritma ini didasarkan pada efektivitasnya sebagai metode pembelajaran mesin tanpa supervisi (*Unsupervised learning*) untuk memproses *Dataset* berdimensi besar tanpa memerlukan pelabelan data awal [5]. Proses pembelajaran mesin memberikan kemampuan bagi sistem untuk mengenali pola-pola tersembunyi dari pengalaman data dan meningkatkan akurasi analisis secara berkelanjutan tanpa memerlukan pengkodean manual yang rumit [5]. Selain pemilihan algoritma, penggunaan sumber data yang presisi juga menjadi faktor penentu kualitas analisis fisis. Penelitian ini memanfaatkan data reanalisis atmosfer *Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2)* yang disediakan oleh NASA *Goddard Earth Sciences Data and Information Services Center* [6]. Penggunaan data reanalisis atmosfer dinilai sangat andal karena mampu menyediakan variabel meteorologi permukaan secara kontinu di seluruh cakupan wilayah studi dengan resolusi temporal satu jam, yang melampaui keterbatasan spasial dari jaringan stasiun pengamatan konvensional [6].

Efektivitas metodologi klusterisasi dalam bidang meteorologi telah dibuktikan oleh berbagai penelitian terdahulu yang mengolah *Dataset* dengan karakteristik serupa. Identifikasi profil cuaca di Provinsi Sumatera Utara berhasil dilakukan melalui pengelompokan ke dalam tiga pola utama, yaitu kondisi cerah, lembap, dan hujan, berdasarkan parameter fisis seperti suhu dan kelembapan [2]. Selain itu, analisis terhadap pola kondisi atmosfer berbasis data historis jangka panjang juga menunjukkan konsistensi dalam pembagian karakteristik fisis atmosfer yang bermakna bagi interpretasi klimatologis *Regional* [7]. Implementasi klusterisasi ini bahkan telah diperluas ke berbagai wilayah lainnya di Indonesia, termasuk pengelompokan data meteorologi harian di wilayah Indonesia Timur untuk mendeteksi variasi cuaca *Regional* melalui analisis variabel suhu dan kecepatan angin [8]. Di wilayah pesisir Jawa Tengah, kombinasi algoritma klusterisasi dan metode klasifikasi lainnya terbukti efektif dalam memproses data stasiun guna memetakan variabilitas cuaca lokal secara lebih spesifik [9]. Bahkan, teknik klusterisasi juga telah diterapkan pada skala yang lebih mikro, seperti pengelompokan suhu di dalam gedung komersial guna mendukung manajemen efisiensi energi berbasis data deret waktu [10].

Keabsahan hasil pengelompokan dalam riset *data mining* sangat bergantung pada tahap evaluasi kualitas kluster guna memastikan akurasi interpretasi fisis. Penggunaan metrik internal seperti *Silhouette Score* dan *Davies-Bouldin Index (DBI)* merupakan langkah standar dalam analisis tanpa supervisi untuk mengukur seberapa baik setiap titik data ditempatkan dalam kelompoknya sendiri dibandingkan dengan kelompok lainnya [11]. Metrik *Silhouette Score* memberikan gambaran objektif mengenai kepadatan dan pemisahan antar kluster, di mana nilai yang optimal menunjukkan struktur kelompok yang kuat dan terpisah dengan baik secara statistik [11]. Selain evaluasi kluster, analisis dinamika atmosfer juga dapat didukung oleh teknik pemrosesan sinyal digital, seperti aplikasi *Fast Fourier Transform (FFT)* yang digunakan untuk menganalisis korelasi periodisitas curah hujan guna mengungkap pola siklus cuaca yang kompleks [12]. Integrasi antara teknik klusterisasi dan evaluasi metrik yang tepat diharapkan mampu meminimalisir kesalahan interpretasi pada *Dataset* meteorologi multidimensi yang bersifat kontinu [11], [5].

Dinamika kondisi atmosfer juga dapat dipahami lebih dalam melalui analisis studi kasus terhadap berbagai fenomena konvektif yang terjadi di wilayah maritim Indonesia. Analisis terhadap dinamika atmosfer saat kejadian hujan lebat menunjukkan adanya pengaruh signifikan dari parameter atmosfer lokal yang dipantau melalui metode citra satelit *Red Green Blue (RGB)* dan *Cloud Convective Overlays (CCO)* [13]. Selain fenomena hujan lebat, kejadian cuaca ekstrem berupa angin puting beliung di wilayah pesisir seperti Tanjung Balai Karimun juga memberikan gambaran mengenai pentingnya pemantauan profil tekanan dan konvergensi udara secara aktual guna deteksi dini [14]. Faktor-faktor fisis seperti fluktuasi suhu muka laut dan tekanan udara permukaan terbukti berhubungan erat dengan aktivitas pertumbuhan awan *Cumulonimbus* yang menjadi pemicu utama fenomena atmosfer intens di kawasan tropis [15]. Keandalan data reanalisis dalam memantau parameter tersebut juga telah dibuktikan dalam analisis spasial dan temporal *Aerosol* di wilayah Pulau Borneo menggunakan data MERRA-2 [16]. Seluruh studi kasus tersebut memperkuat urgensi pemetaan parameter atmosfer *Regional* guna memperdalam pemahaman mengenai mekanisme cuaca jangka pendek selama periode pergantian tahun di Sumatera Utara.

Berdasarkan kajian terhadap literatur-literatur terkait tersebut, ditemukan beberapa celah penelitian (*Gap analysis*) yang menjadi dasar urgensi penelitian ini. Mayoritas penelitian berbasis klusterisasi sebelumnya di Sumatera Utara umumnya hanya menggunakan data observasi stasiun permukaan dengan sebaran spasial yang sangat terbatas [7], [2]. Selain itu, penelitian atmosfer sebelumnya cenderung difokuskan pada rentang data historis yang sangat panjang untuk tujuan klimatologis, sehingga dinamika harian aktual pada periode spesifik transisi tahun baru belum terungkap secara mendetail dan komprehensif. Nilai kebaruan (*Novelty*) penelitian ini terletak pada integrasi data reanalisis spasial

penyempurnaan MERRA-2 untuk menganalisis distribusi spasial tiga kluster atmosfer di enam sub-wilayah Sumatera Utara secara kontinu selama periode transisi lima hari yang krusial.

Tujuan utama dari penelitian ini adalah untuk menganalisis dan mengidentifikasi pola spasial kluster kondisi atmosfer jangka pendek di wilayah Sumatera Utara pada periode 28 Desember 2025 hingga 1 Januari 2026 menggunakan algoritma *K-Means clustering* terhadap data MERRA-2. Hasil penelitian ini diharapkan dapat memberikan gambaran komprehensif mengenai profil atmosfer *Regional* pada periode pergantian tahun, mengungkap keterkaitan antara pola kluster dengan karakteristik topografi wilayah, serta memberikan kontribusi terhadap pemahaman dinamika atmosfer jangka pendek di kawasan Sumatera Utara di masa mendatang.

2. METODOLOGI PENELITIAN

2.1 Metode Pengumpulan dan Instrumen Penelitian

Data yang digunakan dalam penelitian ini bersumber dari *Modern-Era Retrospective Analysis for Research and Applications Version 2* (MERRA-2) yang dikembangkan oleh NASA *Global Modeling and Assimilation Office* (GMAO). *Dataset* MERRA-2 merupakan produk reanalisis atmosfer global yang mengintegrasikan data observasi satelit, radiosonde, dan berbagai instrumen meteorologi dengan model sirkulasi atmosfer melalui teknik asimilasi data.

Produk data yang digunakan adalah *tavg1_2d_slv_Nx* yang menyediakan variabel meteorologi permukaan dengan resolusi spasial sebesar $0,5^\circ \times 0,625^\circ$ (lintang \times bujur) serta resolusi temporal satu jam. Data yang dianalisis mencakup periode 28 Desember 2025 hingga 1 Januari 2026 atau selama lima hari pengamatan.

Wilayah penelitian difokuskan pada Provinsi Sumatera Utara yang secara geografis berada pada kisaran koordinat sekitar 1°LU hingga 4°LU dan 98°BT hingga 100°BT di bagian utara Pulau Sumatera. Untuk memastikan seluruh wilayah administratif Sumatera Utara tercakup secara lengkap, termasuk Kepulauan Nias yang terletak lebih ke arah barat dan utara, batas wilayah kajian diperluas dengan menambahkan buffer spasial sehingga mencakup koordinat $0,5^\circ\text{LU}$ hingga $5,0^\circ\text{LU}$ dan $95,0^\circ\text{BT}$ hingga $100,5^\circ\text{BT}$.

Setelah proses ekstraksi wilayah dilakukan, total data yang diperoleh berjumlah 10.800 observasi yang merupakan kombinasi dari 120 langkah waktu pengamatan, 10 titik grid lintang, serta 9 titik grid bujur. Variabel meteorologi yang digunakan dalam penelitian ini meliputi:

Tabel 1. Jenis-jenis variabel

Variabel	Deskripsi	Satuan
T2M	Suhu udara pada ketinggian 2 meter	Kelvin (K)
QV2M	Kelembapan spesifik pada ketinggian 2 meter	kg/kg
PS	Tekanan Permukaan	Pascal (Pa)
U10M	Komponen angin zonal pada ketinggian 10 meter	m/s
V10M	Komponen angin meridional pada ketinggian 10 meter	m/s

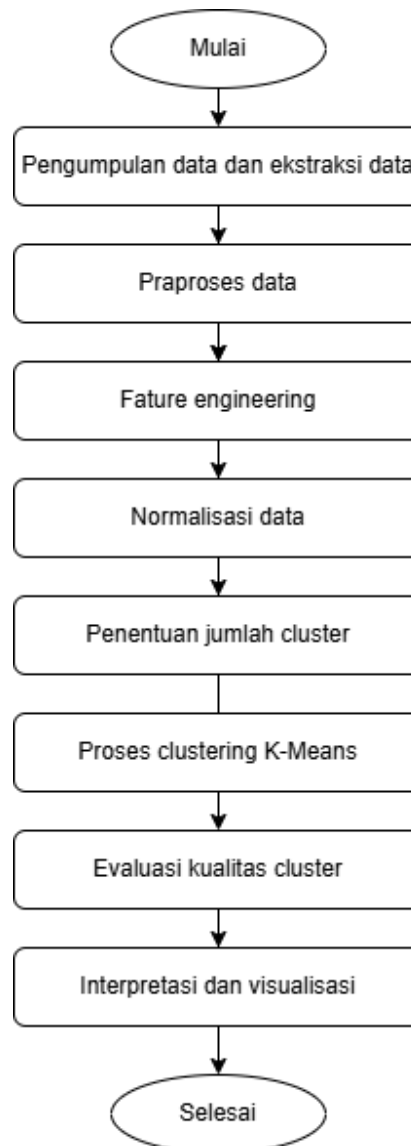
Selain variabel tersebut, penelitian ini juga menghitung kecepatan angin (WS10M) sebagai fitur tambahan yang diperoleh melalui kombinasi komponen angin zonal dan meridional menggunakan persamaan berikut:

$$WS10M = \sqrt{U10M^2 + V10M^2} \quad (1)$$

Seluruh proses analisis dilakukan menggunakan bahasa pemrograman Python 3.11 dengan pustaka utama: *xarray* untuk pembacaan data *NetCDF*, *pandas* dan *numpy* untuk pengolahan data, *scikit-learn* untuk proses normalisasi dan *clustering*, serta *matplotlib*, *seaborn*, dan *cartopy* untuk visualisasi.

2.2 Tahapan Penelitian

Tahapan penelitian dilakukan secara sistematis untuk memastikan validitas hasil pengelompokan. Urutan langkah-langkah penelitian dituangkan dalam diagram alur yang ditunjukkan pada Gambar 1:



Gambar 1. Alur Tahapan Penelitian

Penjelasan mengenai alur tahapan penelitian pada Gambar 1 adalah sebagai berikut:

a. Pengumpulan dan Ekstraksi Data

Data MERRA-2 dalam format NetCDF diunduh dari portal NASA Earthdata untuk periode penelitian yang telah ditentukan. Data global kemudian digabungkan berdasarkan koordinat spasial dan temporal, kemudian dilakukan pemotongan wilayah (*spatial subsetting*) sesuai batas koordinat wilayah Sumatera Utara.

b. Praproses Data

Tahapan praproses dilakukan untuk memastikan kualitas data yang digunakan dalam analisis. Proses ini meliputi penghapusan nilai kosong (*missing values*) serta penyaringan nilai yang tidak realistis secara fisik. Nilai suhu udara difilter pada rentang 200 K hingga 330 K untuk menghindari anomali data yang tidak valid.

c. Feature Engineering

Pada tahap ini dilakukan perhitungan parameter tambahan berupa kecepatan angin (WS10M) yang diperoleh dari kombinasi komponen angin zonal (U10M) dan meridional (V10M). Parameter ini digunakan sebagai salah satu fitur dalam proses analisis *clustering*.

d. Normalisasi Data

Seluruh variabel meteorologi dinormalisasi menggunakan metode *StandardScaler* untuk memastikan bahwa setiap variabel memiliki skala yang sebanding. Proses ini penting karena algoritma *K-Means* menggunakan perhitungan jarak *Euclidean* yang sensitif terhadap perbedaan skala antar variabel.

e. Penentuan Jumlah Cluster Optimal

Jumlah cluster optimal ditentukan menggunakan metode *Elbow Method* dengan mengevaluasi nilai *Within-Cluster Sum of Squares (WCSS)* pada beberapa nilai *K*. Titik perubahan kemiringan pada grafik elbow digunakan sebagai indikator jumlah cluster yang paling representatif untuk menggambarkan pola data.

f. Proses Clustering

Proses pengelompokan data dilakukan menggunakan algoritma *K-Means* dengan jumlah cluster yang telah ditentukan pada tahap sebelumnya. Algoritma ini bekerja dengan meminimalkan nilai *Within-Cluster Sum of Squares (WCSS)* secara iteratif, di mana setiap observasi ditetapkan ke cluster dengan centroid terdekat berdasarkan jarak *Euclidean*. Parameter *random_state=42* ditetapkan untuk memastikan reproduktibilitas hasil analisis.

g. Evaluasi Kualitas Cluster

Kualitas hasil pengelompokan dievaluasi menggunakan *Silhouette Score* yang mengukur tingkat kesamaan suatu data dengan clusternya sendiri dibandingkan dengan cluster lain. Nilai *Silhouette Score* berkisar antara -1 hingga 1, di mana nilai yang lebih tinggi menunjukkan kualitas pemisahan cluster yang lebih baik. Selain *Silhouette Score global*, distribusi nilai *silhouette* per observasi juga divisualisasikan untuk melihat konsistensi kualitas cluster secara keseluruhan.

h. Interpretasi dan Visualisasi Hasil

Setiap cluster diinterpretasikan berdasarkan nilai centroid yang ditransformasi kembali ke satuan aslinya menggunakan *inverse transform*. Klasifikasi kondisi atmosfer dilakukan berdasarkan kategori suhu dan kelembapan yang telah ditetapkan. Hasil *clustering* kemudian dipetakan secara spasial menggunakan koordinat lintang dan bujur untuk mengidentifikasi distribusi kondisi atmosfer di enam sub-wilayah Sumatera Utara, yaitu Kepulauan Nias, Pantai Barat, Kawasan Danau Toba, Dataran Tinggi Karo, Pegunungan Bukit Barisan, dan Pantai Timur Sumatera Utara.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil analisis pengelompokan kondisi atmosfer di wilayah Sumatera Utara menggunakan metode *K-Means clustering*. Analisis dilakukan pada data reanalisis atmosfer yang diperoleh dari produk *MERRA-2 atmospheric reanalysis Dataset* dengan resolusi spasial $0,5^\circ \times 0,625^\circ$ dan resolusi temporal 1 jam selama periode 28 Desember 2025 hingga 1 Januari 2026.

3.1 Hasil Ekstraksi dan Praproses Data

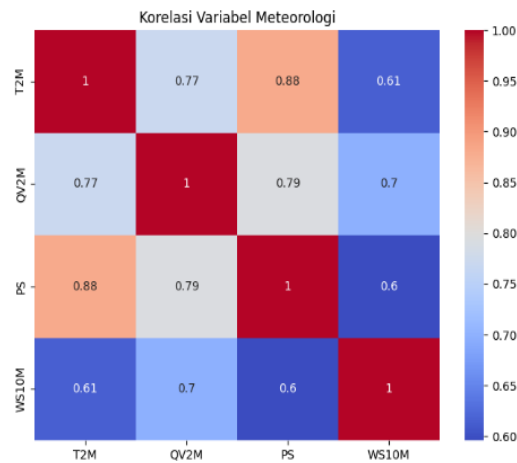
Proses ekstraksi data *MERRA-2* menghasilkan *Dataset* dengan dimensi 120 langkah waktu, 10 titik grid lintang, dan 9 titik grid bujur, sehingga total observasi yang diperoleh adalah 10.800 data. Statistik deskriptif keempat variabel fitur ditunjukkan pada Tabel 2.

Tabel 2. Statistik Deskriptif Variabel Meteorologi Wilayah Sumatera Utara

Variabel	Minimum	Maksimum	Rata-rata	Std.Deviasi	Satuan
T2M	289,91	303,47	299,61	2,62	K
QV2M	0,01321	0,02110	0,01862	0,00124	Kg/kg
PS	86.866,63	101.231,14	98.711,70	3.509,01	Pa
WS10M	0,019	9,007	2,855	1,717	m/s

Berdasarkan Tabel 2, variabel suhu udara (T2M) berkisar antara 289,91 K hingga 303,47 K dengan rata-rata 299,61 K. Nilai standar deviasi sebesar 2,62 menunjukkan adanya variasi suhu yang cukup signifikan antara wilayah dataran tinggi dan dataran rendah di Sumatera Utara. Kelembapan spesifik (QV2M) memiliki nilai rata-rata 0,01862 kg/kg, yang menunjukkan karakteristik wilayah tropis lembap yang konsisten. Tekanan permukaan (PS) menunjukkan rentang yang sangat lebar, di mana nilai minimum yang rendah (86.866,63 Pa) mencerminkan kondisi di wilayah dengan elevasi tinggi seperti Pegunungan Bukit Barisan. Kecepatan angin (WS10M) menunjukkan dinamika yang beragam dengan nilai maksimum mencapai 9,007 m/s, mengindikasikan adanya pergerakan massa udara yang aktif selama periode monsun.

Hubungan antar variabel meteorologi dianalisis melalui matriks korelasi yang ditunjukkan pada Gambar 2.

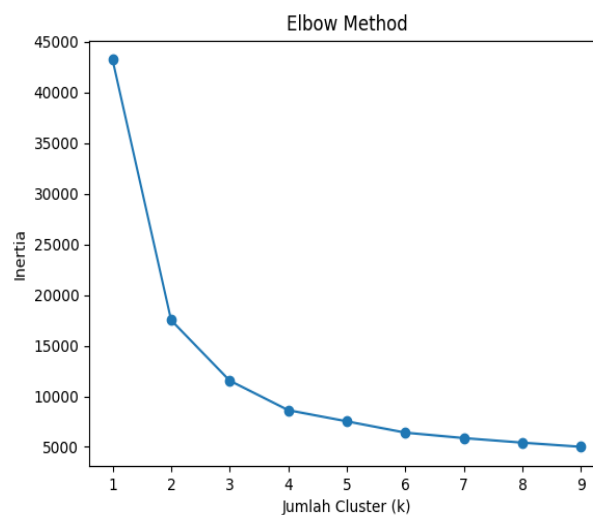


Gambar 2. Matriks Korelasi Variabel Meteorologi Wilayah Sumatera Utara

Berdasarkan Gambar 2, seluruh pasangan variabel menunjukkan korelasi positif. T2M dan PS memiliki korelasi tertinggi sebesar 0,88. Hal ini menunjukkan bahwa wilayah dengan tekanan permukaan lebih tinggi cenderung memiliki suhu yang lebih tinggi, yang merupakan karakteristik umum dataran rendah pesisir jika dibandingkan dengan dataran tinggi. Korelasi antara PS dan QV2M sebesar 0,79 juga menunjukkan bahwa tekanan udara memiliki peran penting dalam mendistribusikan uap air di lapisan permukaan.

3.2. Penentuan Jumlah Cluster Optimal

Tahap krusial dalam algoritma *K-Means* adalah menentukan jumlah kluster optimal (k) agar pembagian data tidak terlalu umum namun juga tidak terlalu kompleks. Penentuan dilakukan menggunakan *Elbow Method* dengan mengevaluasi nilai inersia untuk $K=1$ hingga $K=9$, sebagaimana ditunjukkan pada Gambar 3.

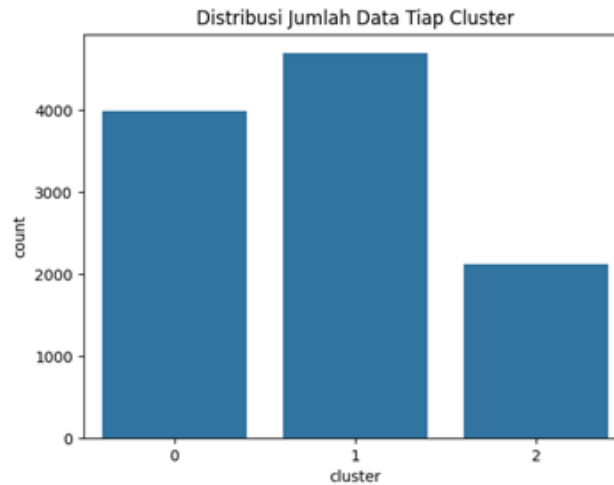


Gambar 3. Grafik Elbow Method Penentuan Jumlah Cluster Optimal

Kurva inersia menunjukkan penurunan tajam dari $K=1$ hingga $K=2$. Titik infleksi atau "siku" mulai terbentuk secara jelas pada $K=3$. Setelah titik tersebut, penurunan inersia melandai secara signifikan, sehingga penambahan kluster lebih lanjut tidak memberikan peningkatan kualitas pemisahan yang efisien. Oleh karena itu, jumlah kluster optimal ditetapkan sebanyak tiga kelompok.

3.3. Hasil *K-Means clustering*

Penerapan algoritma *K-Means* dengan $K = 3$ menghasilkan tiga cluster dengan distribusi jumlah data sebagaimana ditunjukkan pada Gambar 4 dan Tabel 3.



Gambar 4. Distribusi Jumlah Data per Cluster

Berdasarkan Gambar 3, distribusi data per cluster ditunjukkan secara lebih rinci pada Tabel 3.

Tabel 3. Distribusi Jumlah Data per Cluster

Cluster	Jumlah Data	Persentase
Cluster 0	3.988	36,93%
Cluster 1	4.692	43,44%
Cluster 2	2.120	19,63%
Total	10.800	100%

Berdasarkan Tabel 3, Cluster 1 merupakan cluster dengan jumlah data terbesar (43,44%), diikuti Cluster 0 (36,93%), dan Cluster 2 (19,63%). Distribusi yang tidak merata ini mencerminkan perbedaan luas wilayah dan frekuensi kemunculan masing-masing kondisi atmosfer di Sumatera Utara selama periode pengamatan.

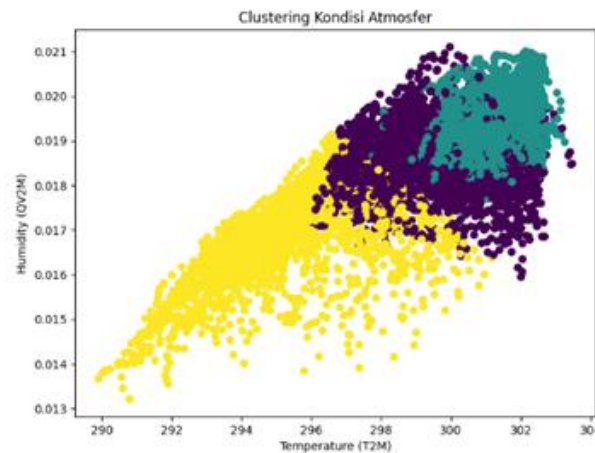
Nilai centroid masing-masing cluster setelah ditransformasi kembali ke satuan aslinya ditunjukkan pada Tabel 4.

Tabel 4. Nilai Centroid Setiap Cluster

Cluster	T2M (K)	T2M (°C)	QV2M (kg/kg)	Ps (Pa)	WS10M (m/s)	Kondisi
Cluster 0	300,04	26,89	0,01856	99.531,20	1,90	Hangat & Sangat Lembap
Cluster 1	301,22	28,07	0,01953	100.776,53	4,51	Panas Tropis & Sangat Lembap
Cluster 2	295,20	22,05	0,01674	92.583,31	0,98	Sejuk & Lembap Sedang

Berdasarkan Tabel 3, ketiga cluster menunjukkan karakteristik atmosfer yang berbeda secara jelas. Cluster 0 dicirikan oleh suhu hangat (26,89°C), kelembapan sangat lembap (0,01856 kg/kg), tekanan permukaan sedang (99.531,20 Pa), dan kecepatan angin rendah (1,90 m/s), merepresentasikan kondisi Hangat & Sangat Lembap yang umum dijumpai di wilayah dataran rendah dan pesisir Sumatera Utara. Cluster 1 memiliki suhu tertinggi (28,07°C) dengan kelembapan paling tinggi (0,01953 kg/kg), tekanan permukaan tertinggi (100.776,53 Pa), dan kecepatan angin terkuat (4,51 m/s), merepresentasikan kondisi Panas Tropis & Sangat Lembap yang dominan di wilayah kepulauan dan pesisir yang terpapar langsung angin laut. Cluster 2 menunjukkan karakteristik yang berbeda signifikan dengan suhu terendah (22,05°C), kelembapan paling rendah (0,01674 kg/kg), tekanan permukaan terendah (92.583,31 Pa), dan kecepatan angin paling tenang (0,98 m/s), merepresentasikan kondisi Sejuk & Lembap Sedang khas wilayah dataran tinggi.

Visualisasi sebaran cluster pada ruang fitur T2M dan QV2M ditunjukkan pada Gambar 5.

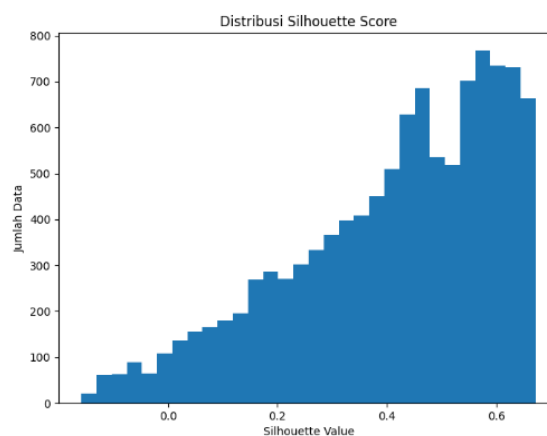


Gambar 5. Visualisasi *Clustering* Kondisi Atmosfer pada Ruang Fitur T2M dan QV2M

Berdasarkan Gambar 5, ketiga cluster memiliki separasi yang cukup jelas pada ruang fitur suhu dan kelembapan. Cluster 2 (kuning) terpisah jauh di sisi kiri dengan rentang suhu sekitar 289–296 K dan kelembapan relatif lebih rendah, sedangkan Cluster 0 (ungu) dan Cluster 1 (hijau toska) berada pada suhu lebih tinggi (296–304 K) dengan perbedaan utama pada kecepatan angin dan tekanan permukaan yang tidak terlihat langsung pada scatter plot dua dimensi ini.

3.4. Evaluasi Kualitas Cluster

Kualitas hasil *clustering* dievaluasi menggunakan *Silhouette Score*. Distribusi nilai *silhouette* per observasi ditunjukkan pada Gambar 6.

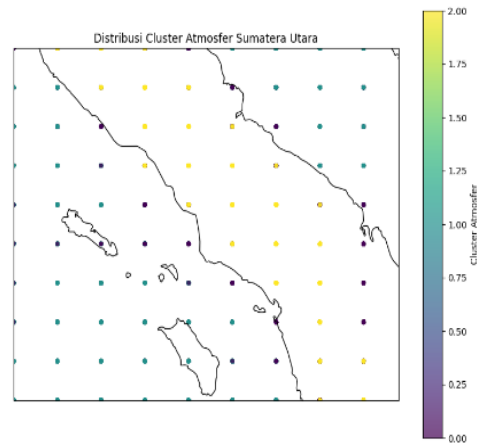


Gambar 6. Distribusi *Silhouette Score* per Observasi

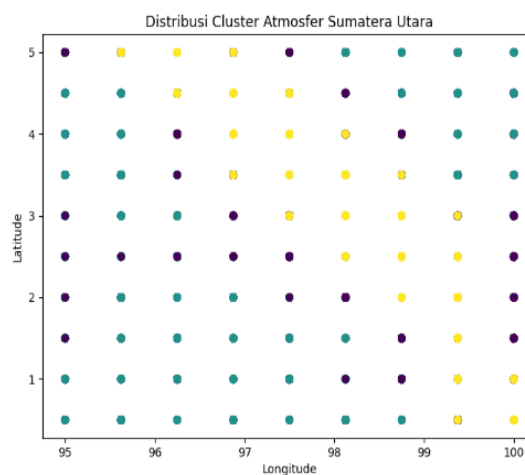
Berdasarkan Gambar 6, distribusi nilai *silhouette* menunjukkan bahwa sebagian besar data memiliki nilai *silhouette* positif dengan konsentrasi terbesar pada rentang 0,5 – 0,65, yang mengindikasikan bahwa mayoritas observasi telah ditempatkan pada cluster yang tepat. Terdapat sebagian kecil data dengan nilai *silhouette* negatif yang mencerminkan adanya tumpang tindih (*overlap*) pada batas antar cluster, hal ini wajar terjadi mengingat variabel atmosfer bersifat kontinu tanpa pemisah yang tegas. Nilai *Silhouette Score global* yang diperoleh adalah 0,4089, berada pada rentang 0,25–0,50 yang mengindikasikan struktur cluster moderat (*reasonable structure*) dan masih dapat diterima untuk data geofisika yang memiliki gradien spasial kontinu.

3.5. Distribusi Spasial Cluster

Distribusi spasial hasil *clustering* dipetakan berdasarkan koordinat lintang dan bujur sebagaimana ditunjukkan pada Gambar 7 dan Gambar 8.



Gambar 7. Distribusi Spasial Cluster Atmosfer Sumatera Utara (Tampilan Grid)



Gambar 8. Peta Distribusi Spasial Cluster Atmosfer Wilayah Sumatera Utara

Berdasarkan Gambar 7 dan Gambar 8, distribusi cluster memiliki pola spasial yang konsisten dengan karakteristik topografi Sumatera Utara. Cluster 2 (kuning — Sejuk & Lembap Sedang) teridentifikasi pada titik-titik grid yang berkorespondensi dengan wilayah dataran tinggi di bagian tengah Sumatera Utara. Cluster 1 (hijau toska — Panas Tropis & Sangat Lembap) mendominasi wilayah bagian barat dan utara yang mencakup Kepulauan Nias dan pesisir yang terpapar langsung massa udara laut. Cluster 0 (ungu — Hangat & Sangat Lembap) tersebar merata di wilayah peralihan antara dataran tinggi dan pesisir.

Ringkasan distribusi kondisi atmosfer per sub-wilayah berdasarkan klasifikasi koordinat ditunjukkan pada Tabel 5.

Tabel 5. Distribusi Kondisi Atmosfer per Sub-Wilayah Sumatera Utara

Sub-Wilayah	Hangat & Sangat Lembap	Panas Tropis & Sangat Lembap	Sejuk & Lembap Sedang	Total
Kepulauan Nias	1.865	3.286	849	6.000
Pantai Barat	448	423	329	1.200
Kawasan Danau Toba	-	-	360	360
Dataran Tinggi Karo	226	252	2	480
Pegunungan Bukit Barisan	227	133	-	360
Pantai Timur	1.222	598	580	2.400
Total	3.988	4.692	2.120	10.800

Pembagian sub-wilayah dilakukan berdasarkan klasifikasi koordinat geografis menggunakan fungsi wilayah `sumut()` dengan batas: Kepulauan Nias ($lon < 97,8^{\circ}BT$), Pantai Barat ($97,8^{\circ}BT \leq lon < 98,4^{\circ}BT$), Kawasan Danau Toba ($98,4^{\circ}BT \leq lon \leq 99,0^{\circ}BT$, $2,0^{\circ}LU \leq lat \leq 3,0^{\circ}LU$), Dataran Tinggi Karo ($98,4^{\circ}BT \leq lon \leq 99,0^{\circ}BT$, $lat > 3,0^{\circ}LU$), Pegunungan Bukit Barisan ($98,4^{\circ}BT \leq lon \leq 99,0^{\circ}BT$, $lat < 2,0^{\circ}LU$), dan Pantai Timur ($lon > 99,0^{\circ}BT$). Berdasarkan Tabel 5, terdapat beberapa temuan penting yang dapat diuraikan sebagai berikut.

Kepulauan Nias didominasi oleh kondisi Panas Tropis & Sangat Lembap (3.286 data atau 54,8% dari total Nias), yang konsisten dengan posisi geografisnya sebagai wilayah kepulauan di perairan terbuka Samudera Hindia yang menerima pengaruh langsung massa udara lembap dari laut selama periode monsun.

Kawasan Danau Toba ($98,4^{\circ}BT-99,0^{\circ}BT$, $2,0^{\circ}LU-3,0^{\circ}LU$) seluruh titik gridnya menunjukkan kondisi Sejuk & Lembap Sedang (360 data). Hal ini konsisten dengan karakteristik kawasan Danau Toba sebagai dataran tinggi berketinggian sekitar 900 mdpl yang secara alami memiliki suhu udara lebih rendah dibandingkan wilayah dataran rendah sekitarnya.

Dataran Tinggi Karo ($98,4^{\circ}BT-99,0^{\circ}BT$, $lat > 3,0^{\circ}LU$) yang secara geografis berada pada ketinggian lebih tinggi (~1.400 mdpl) hanya menghasilkan 2 data kondisi Sejuk dari total 480 data. Hal ini disebabkan oleh keterbatasan resolusi spasial MERRA-2 sebesar $0,5^{\circ} \times 0,625^{\circ}$ yang tidak mampu merepresentasikan variasi topografi mikro secara detail, sehingga nilai rata-rata per grid mencakup kombinasi wilayah dataran tinggi dan lereng.

Pantai Timur Sumatera Utara menunjukkan distribusi yang lebih merata antara ketiga kondisi atmosfer (Hangat: 1.222, Panas Tropis: 598, Sejuk: 580), yang mencerminkan variabilitas kondisi meteorologi yang lebih dinamis di wilayah transisi antara daratan dan perairan Selat Malaka.

Pegunungan Bukit Barisan tidak menunjukkan kemunculan kondisi Sejuk dalam periode pengamatan, serupa dengan Dataran Tinggi Karo, yang mengindikasikan keterbatasan resolusi spasial MERRA-2 dalam merepresentasikan variasi elevasi di sepanjang rangkaian pegunungan tersebut.

3.6. Perbandingan dengan Penelitian Sebelumnya

Temuan dalam studi ini menunjukkan relevansi dan konsistensi yang signifikan apabila dibandingkan dengan beberapa literatur terdahulu, meskipun tetap mempertahankan nilai kebaruan yang spesifik. Komparasi utama dilakukan terhadap hasil penelitian yang disusun oleh Manalu dkk. [2], yang juga menitikberatkan pada analisis pola cuaca di wilayah Provinsi Sumatera Utara melalui implementasi algoritma *K-Means clustering*. Persamaan fundamental ditemukan pada identifikasi karakteristik wilayah berbasis aspek topografi, di mana zona dataran tinggi dan kawasan pesisir secara konsisten menghasilkan pembentukan kluster atmosfer yang kontras. Namun, perbedaan mendasar ditemukan pada dimensi temporal; penelitian ini merupakan studi kasus jangka pendek selama lima hari yang mampu menangkap dinamika atmosfer harian secara lebih mendetail dibandingkan dengan pola klimatologis tahunan yang bersifat lebih umum.

Implementasi *Dataset* reanalisis MERRA-2 dalam kajian ini memiliki keselarasan metodologis dengan prosedur yang diterapkan oleh Sentian dkk. [16] di wilayah Borneo. Kedua studi tersebut memberikan validasi bahwa produk reanalisis dari NASA memiliki keandalan tinggi dalam memetakan parameter atmosfer di kawasan tropis Asia Tenggara dengan resolusi temporal yang rapat. Selain itu, teknik pengelompokan pola atmosfer ini juga menunjukkan kemiripan dengan pendekatan yang digunakan oleh Prayigo dan Marwati [7] yang mengidentifikasi karakteristik cuaca berdasarkan data historis melalui metode serupa. Keunggulan penelitian ini terletak pada fokus spesifik yang mengunci periode transisi tahun baru guna mengamati profil atmosfer *Regional* pada momen pergantian tahun secara aktual.

Keberhasilan dalam pemisahan kluster pada wilayah Karo dan kawasan dataran tinggi lainnya dalam analisis ini memperkuat temuan yang dilaporkan oleh Clario dkk. [1] mengenai karakteristik tren iklim lokal di Sumatera Utara yang sangat dipengaruhi oleh faktor elevasi. Meskipun pembatasan data hanya dilakukan dalam rentang waktu lima hari, integrasi antara analisis spasial dan pembelajaran mesin (*Machine learning*) terbukti berhasil mencapai objektif penelitian dalam mengidentifikasi profil atmosfer *Regional* secara representatif. Hal ini sekaligus membuktikan bahwa pendekatan studi kasus harian, sebagaimana juga diimplementasikan dalam penelitian dinamika atmosfer oleh Sawal dkk. [14] dan Esti Vidia dkk. [13], merupakan metodologi yang valid untuk memahami fenomena meteorologi spesifik meskipun tidak mencakup rentang waktu pengamatan klimatologis yang panjang.

4. KESIMPULAN

Penelitian ini bertujuan untuk mengidentifikasi dan memetakan pola spasial kondisi atmosfer di wilayah Sumatera Utara selama periode 28 Desember 2025 hingga 1 Januari 2026 menggunakan pendekatan *K-Means clustering* berbasis data reanalisis MERRA-2. Berdasarkan hasil analisis, permasalahan penelitian terkait bagaimana karakteristik dan distribusi spasial kondisi atmosfer pada periode puncak monsun dapat terjawab melalui terbentuknya tiga kluster atmosfer yang memiliki perbedaan karakteristik fisik yang jelas. Kluster yang dihasilkan menunjukkan bahwa variasi suhu udara, kelembapan spesifik, tekanan permukaan, dan kecepatan angin membentuk zonasi atmosfer yang konsisten dengan kondisi topografi dan pengaruh maritim *Regional*. Wilayah dataran tinggi cenderung membentuk pola atmosfer sejuk dengan kelembapan sedang, sedangkan wilayah pesisir dan kepulauan menunjukkan karakter panas tropis dengan kelembapan sangat tinggi serta dinamika angin yang lebih kuat. Temuan ini menegaskan

bahwa pendekatan unsupervised *Machine learning* mampu mengungkap struktur spasial atmosfer jangka pendek secara objektif tanpa memerlukan klasifikasi awal. Nilai evaluasi *clustering* yang berada pada kategori moderat menunjukkan bahwa meskipun variabel atmosfer bersifat kontinu, pola pengelompokan tetap dapat diidentifikasi secara representatif. Secara ilmiah, penelitian ini memberikan kontribusi pada pemanfaatan data reanalisis global untuk kajian dinamika atmosfer *Regional* berbasis studi kasus harian. Ke depan, penelitian serupa dapat dikembangkan dengan cakupan waktu yang lebih panjang untuk analisis musiman atau klimatologis, serta integrasi dengan data observasi permukaan guna meningkatkan validitas spasial hasil klusterisasi.

REFERENCES

- [1] S. R. Clario, T. Anggraeni, T. C. Putri, and J. Y. Ariantono, "Tren Iklim dan Implikasinya terhadap Ketahanan Pangan di Wilayah Asahan dan Karo, Sumatera Utara," no. 2024, 2025, doi: <https://doi.org/10.56211/tabela.v3i2.1112>.
- [2] J. Ilmu, K. Dan, V. N. J. Hal, E. Manalu, E. Surbakti, and S. Pardingotan, "Analisis Pola Cuaca di Provinsi Sumatera Utara Menggunakan Metode Clustering K-Means," vol. 2, no. 3, pp. 81–89, 2026, [Online]. Available: <https://jurnal.globalscients.com/index.php/jiki/article/view/988%0A>
- [3] F. A. Zurfani, M. R. Syahputra, F. Matematika, P. Alam, and U. S. Utara, "Analisis Metode Clustering K-Means pada Zonasi Daerah Terdampak Banjir di Kota Medan dengan Evaluasi Silhouette Coefficient bahkan menjadi tradisi tahunan. Telah banyak upaya dilakukan untuk menanggulangnya," no. 6, 2024, doi: <https://doi.org/10.62383/algorithm.v2i6.270>.
- [4] B. Hadi, R. G. G. Alam, U. Juhardi, and A. K. Hidayah, "IDENTIFIKASI DAERAH RAWAN BANJIR DI KOTA BENGKULU MENGGUNAKAN K-MEANS CLUSTERING DENGAN PYTHON," vol. 9, no. 4, pp. 5978–5983, 2025, doi: <https://doi.org/10.36040/jati.v9i4.13976>.
- [5] M. Suyal and S. Sharma, "A Review on Analysis of K-Means Clustering Machine Learning Algorithm based on Unsupervised Learning," pp. 85–95, 2024, doi: 10.33969/AIS.2024060106.
- [6] Global Modeling and Assimilation Office (GMAO), "MERRA-2 tavg1_2d_slv_Nx: 2d, 1-Hourly, Time-Averaged, Single-Level, Assimilation, Single-Level Diagnostics V5.12.4," NASA Goddard Earth Sciences Data and Information Services Center (GES DISC).
- [7] Y. P. Permana, F. Marwati, S. Informasi, U. Pamulang, E. Method, and S. Score, "Analisis Pola Kondisi Atmosfer Berdasarkan Data Cuaca Historis Menggunakan Metode K-Means Clustering," vol. 5, no. 2, pp. 226–234, 2025, doi: 10.59395/zb8w0x66.
- [8] D. Metode and K. D. A. N. Fuzzy, "INTI NUSA MANDIRI CLUSTERING DATA METEOROLOGI WILAYAH INDONESIA TIMUR," vol. 18, no. 2, pp. 100–106, 2024, doi: <https://doi.org/10.33480/techno.v16i2.642>.
- [9] R. Forest and T. P. Terdahulu, "CLUSTERING DAN KLASIFIKASI DATA CUACA KOTA CILACAP MENGGUNAKAN K-MEANS DAN RANDOM FOREST," vol. 1, no. April, pp. 90–97, 2024, doi: 10.61124/sinta.v1i2.15.
- [10] A. Wickramasinghe, S. Muthukumarana, D. Loewen, and M. Schaubroeck, "Temperature clusters in commercial buildings using k - means and time series clustering," *Energy Informatics*, pp. 1–14, 2022, doi: 10.1186/s42162-022-00186-8.
- [11] K. Dbscan and Y. Hasan, "Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster," vol. 06, no. 01, pp. 60–74, 2024, doi: <http://dx.doi.org/10.23960/jitet.v12i3S1.5001>.
- [12] A. Of *et al.*, "APLIKASI FAST FOURIER TRANSFORM DALAM ANALISIS KORELASI PERIODISITAS CURAH HUJAN DI SUMATERA UTARA DENGAN FENOMENA VARIABILITAS IKLIM GLOBAL APPLICATION OF FAST FOURIER TRANSFORM IN ANALYSING THE CORRELATION OF RAINFALL PERIODICITY IN NORTH SUMATRA WITH T," vol. 15, no. 1, pp. 29–36, 2024, doi: <https://doi.org/10.46824/megasains.v15i1.148>.
- [13] T. E. Vidia, Y. Darmawan, and W. Widodo, "ANALISIS DINAMIKA ATMOSFER KEJADIAN HUJAN LEBAT DENGAN METODE RED GREEN BLUE (RGB) DAN CLOUD CONVECTIVE OVERLAYS (CCO) DI BALIKPAPAN," vol. 06, no. 02, pp. 112–122, 2025, doi: <https://doi.org/10.30587/indexia.v6i2.9374>.
- [14] T. Balai, K. Case, and S. May, "ANALISIS DINAMIKA ATMOSFER KEJADIAN ANGIN PUTING ANALYSIS OF ATMOSPHERIC DYNAMICS OF THE TORNADO EVENT IN," vol. 15, no. 2, pp. 38–48, 2024, doi: 10.46824/megasains.v15i2.237.
- [15] O. Wulandari, A. A. Rofiah, N. J. Putri, and P. P. Sani, "ANALISIS SUHU MUKA LAUT, SUHU BOLA KERING, DAN SELISIH TEKANAN UDARA UNTUK MEMPREDIKSI PERTUMBUHAN AWAN CUMULONIMBUS DI STASIUN METEOROLOGI MARITIM TANJUNG PERAK SURABAYA JANUARI-AGUSTUS 2024," *J. Inov. Fis. Indones.*, vol. 14, pp. 92–102, 2025, doi: <https://doi.org/10.26740/ifi.v14n1.p92-102>.
- [16] J. Sentian *et al.*, "Recent Spatial and Temporal Analysis of Aerosol Optical Depth From MERRA - 2 Over Borneo Island," *Aerosol Air Qual. Res.*, vol. 26, no. 1, pp. 1–18, 2026, doi: 10.1007/s44408-025-00081-3.