

ANALISIS DAN VISUALISASI POLA PARAMETER CUACA HARIAN DAN BULANAN JAKARTA TIMUR DENGAN K-MEANS CLUSTERING

Lola¹, Aji Nurrohman², Sigit Wibisono³, Rachmat Setiabudi⁴, Atang Supena⁵

Institut Teknologi Budi Utomo, Jakarta Timur

e-mail: ¹lola@itbu.ac.id, ²ajinurrohman@itbu.ac.id, ³sigitwibisono@itbu.ac.id,

⁴rachmatsetiabudi@itbu.ac.id, ⁵atang@itbu.ac.id

Abstract: *Weather observation data produced by the Meteorology, Climatology and Geophysics Agency (BMKG) has multivariate characteristics, so it requires an analysis method that is able to identify patterns based on similarities in data characteristics. This research aims to analyze and visualize daily and monthly weather parameter patterns in the East Jakarta area using the K-Means Clustering algorithm. The dataset used is daily observation data from the Halim Perdanakusuma Meteorological Station for the period January–December 2025, totaling 365 observations with six meteorological parameters, namely TN, TX, TAVG, RH_AVG, RR, and FF_AVG. The research stages include merging data, handling missing values using median imputation, data normalization using StandardScaler, data exploration, determining the optimal number of clusters using the Elbow Method and Silhouette Score, followed by the clustering process using K-Means. The evaluation results show that there are three optimal clusters (K=3), which represent variations in the daily weather characteristics of East Jakarta. The clustering results identified different characteristics, namely Cluster 0 (Hot Dry Weather), Cluster 1 (Cool Cloudy Weather), and Cluster 2 (Heavy Rainy Weather). Visualization in the form of correlation heatmaps, time series graphs, scatter plots, box plots and bar charts provide informative representations of weather parameter distribution patterns, making it easier to interpret analysis results. This research proves that K-Means Clustering is used as an effective exploratory method for identifying weather parameter patterns and visualization of meteorological analysis.*

Keyword: *BMKG; K-Means Clustering; Weather Analysis; Data Visualization; Meteorological Parameters.*

Abstrak: Data observasi cuaca yang dihasilkan Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) memiliki karakteristik multivariat, sehingga memerlukan metode analisis yang mampu mengidentifikasi pola berdasarkan kemiripan karakteristik data. Penelitian ini bertujuan menganalisis dan memvisualisasikan pola parameter cuaca harian dan bulanan wilayah Jakarta Timur menggunakan algoritma K-Means Clustering. Dataset yang digunakan berupa data observasi harian Stasiun Meteorologi Halim Perdanakusuma periode Januari–Desember 2025 sebanyak 365 observasi dengan enam parameter meteorologi, yaitu TN, TX, TAVG, RH_AVG, RR, dan FF_AVG. Tahapan penelitian meliputi penggabungan data, penanganan missing value menggunakan imputasi median, normalisasi data dengan StandardScaler, eksplorasi data, penentuan jumlah klaster optimal menggunakan Elbow Method dan Silhouette Score, dilanjutkan proses klasterisasi menggunakan K-Means. Hasil evaluasi menunjukkan ada tiga klaster optimal (K=3), yang merepresentasikan variasi karakteristik cuaca harian Jakarta Timur. Hasil klasterisasi mengidentifikasi perbedaan karakteristik yaitu Klaster 0 (Cuaca Panas Kering), Klaster 1 (Cuaca Sejuk Berawan), dan Klaster 2 (Cuaca Hujan Lebat). Visualisasi dalam bentuk heatmap korelasi, grafik time series, scatter plot, box plot, dan bar chart memberikan representasi informatif pola distribusi parameter cuaca, sehingga memudahkan proses interpretasi hasil analisis. Penelitian ini membuktikan K-Means Clustering digunakan sebagai metode eksploratif efektif untuk mengidentifikasi pola

parameter cuaca dan visualisasi analisis meteorologi.

Kata kunci: BMKG; K-Means Clustering; Analisis Cuaca; Visualisasi Data; Parameter Meteorologi.

PENDAHULUAN

Menurut Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), kondisi iklim perkotaan di wilayah DKI Jakarta pada tahun 2024 mengalami dinamika yang signifikan. Kompleksitas anomali cuaca tersebut menjadi landasan penting untuk memetakan pola parameter cuaca pada tahun berikutnya, khususnya di wilayah Jakarta Timur tahun 2025 menggunakan pendekatan komputasional.

Parameter cuaca merupakan variabel atmosfer yang sangat menentukan kenyamanan dan keselamatan aktivitas masyarakat di kota besar. Wilayah Jakarta Timur memiliki karakteristik antropogenik yang kompleks, sehingga fluktuasi parameter cuaca sepanjang tahun 2025 yang terekam dalam data historis meteorologi telah dianalisis secara mendalam dalam penelitian ini untuk memetakan pola iklim mikro di Jakarta Timur.

Fenomena *Urban Heat Island* (UHI) menjadi salah satu tantangan lingkungan utama, di mana kawasan perkotaan mencatat suhu udara yang lebih tinggi dibandingkan wilayah penyangganya. Fenomena ini memicu fluktuasi parameter cuaca harian yang dinamis, meliputi perubahan suhu minimum, suhu maksimum, tingkat kelembapan udara, hingga intensitas curah hujan, yang berdampak pada kenyamanan termal dan aktivitas mobilitas masyarakat. Tingkat kenyamanan ini tidak hanya dipengaruhi oleh satu faktor saja, melainkan interaksi kompleks antara berbagai parameter cuaca yang perlu dipahami polanya secara mendalam (BMKG, 2023).

Penelitian ini berfokus pada pemanfaatan algoritma K-Means Clustering sebagai alat bantu analisis utama. Algoritma ini digunakan untuk mengelompokkan data cuaca harian dan

bulanan berdasarkan kemiripan fiturnya, sehingga memungkinkan identifikasi pola cuaca secara objektif. Penelitian ini memberikan penekanan utama pada hasil analisis data dan penyajian dalam bentuk visualisasi grafik. Dengan teknik visualisasi yang tepat, hasil pengelompokan yang rumit dari algoritma K-Means dapat diubah menjadi grafik yang jelas dan mudah dipahami, sehingga pola cuaca di Jakarta Timur tahun 2025 dapat dipetakan dengan baik.

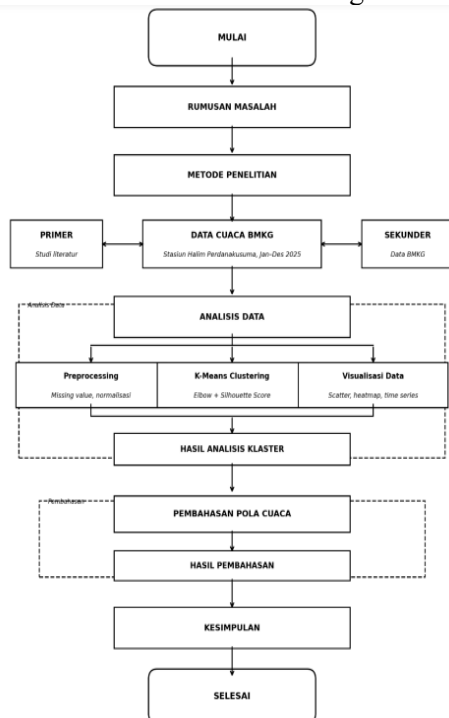
Berdasarkan penelitian terdahulu, sebagian besar penelitian lebih berfokus pada prediksi cuaca, kualitas udara, maupun analisis statistik konvensional. Penelitian yang secara khusus mengidentifikasi pola parameter cuaca harian dan bulanan menggunakan algoritma K-Means Clustering pada wilayah Jakarta Timur masih sangat terbatas. Oleh karena itu penelitian ini dilakukan untuk mengisi kekosongan tersebut.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis data berbasis *machine learning unsupervised learning* menggunakan algoritma K-Means Clustering. Statistik deskriptif digunakan sebagai tahap eksplorasi awal untuk memahami karakteristik data sebelum proses clustering dilakukan. Penelitian ini tidak membangun sistem prediksi maupun melakukan analisis tren jangka panjang, melainkan berfokus pada historical data analysis dengan menerapkan metode machine learning K-Means Clustering untuk menemukan pola pengelompokan (klasterisasi) yang tersembunyi di dalam data cuaca. Hasil analisis disajikan dalam bentuk visualisasi grafik menggunakan

bahasa pemrograman Python di lingkungan *Google Colaboratory*.

Penelitian dimulai dari identifikasi permasalahan yaitu sulitnya memahami pola cuaca Jakarta Timur dari data mentah yang bersifat numerik. Solusi yang ditawarkan adalah penerapan algoritma K-Means Clustering untuk mengelompokkan data cuaca secara otomatis berdasarkan kemiripan fitur, yang kemudian hasilnya disajikan dalam bentuk visualisasi grafik yang komunikatif dan mudah dipahami. Secara garis besar, kerangka pemikiran penelitian ini dapat digambarkan sebagai berikut: Data cuaca dari BMKG Jakarta Timur periode Januari–Desember 2025 → Preprocessing data (pembersihan, normalisasi) → Penentuan jumlah kluster optimal (Metode Elbow dan 26 Silhouette Score) → Penerapan K-Means Clustering → Analisis dan Interpretasi Kluster → Visualisasi Hasil dalam bentuk grafik.

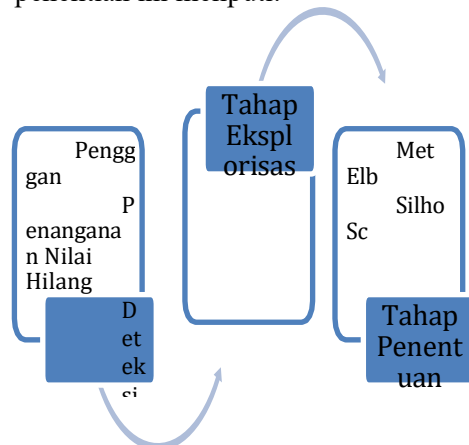


Gambar 1 Alur Penelitian
Sumber: Diolah oleh Peneliti (2026)

Metode pengumpulan data yang digunakan dalam penelitian ini adalah metode Dokumen, yaitu pengambilan data dalam bentuk dokumen elektronik dari institusi resmi pemilik data. Data yang

digunakan merupakan Data Sekunder berupa data pengamatan cuaca harian yang diperoleh dari Stasiun Meteorologi BMKG Halim Perdanakusuma, Jakarta Timur (ID WMO: 96747, Lintang: -6.27036°, Bujur: 106.88926°, Elevasi: 28 meter). Data tersebut mencakup periode Januari hingga Desember 2025, yang diunduh dalam format Excel (.xlsx) melalui sistem basis data *online* BMKG.

Analisis data dalam penelitian ini dilakukan dengan pendekatan *machine learning unsupervised learning* menggunakan algoritma K-Means Clustering. Analisis statistik deskriptif digunakan sebagai *exploratory data analysis* (EDA) untuk mengetahui karakteristik awal dataset. Keseluruhan proses analisis data dilakukan menggunakan bahasa pemrograman Python di lingkungan *Google Colaboratory*. Tahapan analisis data dalam penelitian ini meliputi:



Gambar 2 Alur Metode Analisis Data

Adapun spesifikasi data yang digunakan dalam penelitian ini adalah sebagai berikut:

Tabel 1 Parameter Cuaca yang Dianalisis

No.		Satuan	Kode BMKG
1	Suhu Minimum	°C	TN
2	Suhu Maksimum	°C	TX
3	Suhu Rata-rata	°C	TAVG
4	Kelembaban Rata-rata	%	RH_AVG

5	Curah Hujan	mm	RR
6	Kecepatan Angin	m/s	FF_AV G

Berdasarkan hasil pembacaan keseluruhan file data, diperoleh karakteristik dataset sebagai berikut:

Tabel 2 Karakteristik Dataset

Keterangan	Nilai	Catatan
Stasiun	Halim Perdanakusuma Jakarta Timur	ID WMO: 96747
Periode	Januari – Desember 2025	12 bulan lengkap
Total Hari	365 hari observasi	Data lengkap satu tahun
Jumlah Parameter	6 parameter cuaca	TN, TX, TAVG, RH_AVG, RR, FF_AVG
Format File	Microsoft Excel (.xlsx)	12 file per bulan (Januari– Desember 2025)
Kode Data Hilang	8888 = tidak terukur; 9999 = tidak ada data	Standar kode BMKG

HASIL DAN PEMBAHASAN

Klaster cuaca pada cuaca harian dan bulanan di Jakarta Timur terbagi atas tiga klaster, yaitu :

1. Klaster 0 (Cuaca Panas Kering). Berdasarkan nilai centroid, klaster ini memiliki suhu rata-rata dan suhu maksimum tertinggi, kelembapan relatif terendah, curah hujan yang rendah, serta kecepatan angin rata-rata tertinggi dibandingkan klaster lainnya.
2. Klaster 1 (Cuaca Sejuk Berawan). Kombinasi suhu rendah dan kelembapan tinggi mencerminkan kondisi cuaca yang relatif lebih

lembap dengan dominasi tutupan awan. Klaster 1 mendominasi Januari dan Februari, yang menunjukkan karakteristik kondisi udara lebih lembap dan suhu lebih rendah dibandingkan periode lainnya. Kelembapan tinggi (83,06%) namun curah hujan relatif rendah (7,95 mm) menunjukkan kondisi udara lembap dengan awan tebal namun hujan tidak intens.

3. Klaster 2 (Cuaca Hujan Lebat), klaster ini secara sah diklasifikasikan sebagai Klaster Cuaca Hujan Lebat karena memiliki nilai rata-rata pusat (centroid) curah hujan harian (RR) sebesar 86,15 mm (berada dalam rentang 50,0 – 100,0 mm/hari.

Sebelum proses klusterisasi dilakukan, data mentah hasil unduhan BMKG diproses melalui tahapan-tahapan berikut:

1. Penggabungan Data (*Data Merging*). Sebanyak 12 file Excel per bulan digabungkan menjadi satu DataFrame tunggal menggunakan pustaka Pandas. Proses pembacaan dilakukan dengan mengidentifikasi baris header (baris yang mengandung kata "TANGGAL") secara otomatis karena setiap file memiliki metadata stasiun di baris atas.
2. Penanganan Nilai Hilang (*Missing Values*). Data yang memiliki kode 8888 (data tidak terukur) dan 9999 (tidak ada data) sesuai standar BMKG dikonversi menjadi nilai NaN. Berdasarkan hasil pemeriksaan dataset, terdapat *missing values* pada beberapa parameter: TN (6,3%), TX (6,0%), TAVG (6,3%), RH_AVG (6,3%), RR (11,8%). Di sisi lain, parameter \$FF_{AVG}\$ memiliki data yang lengkap 100% tanpa adanya data hilang. Penanganan *missing values* pada parameter lainnya dilakukan dengan metode imputasi menggunakan nilai median per bulan agar tidak terdistorsi oleh *outlier*.
3. Deteksi dan Penanganan *Outlier*

Deteksi outlier dilakukan menggunakan metode IQR (Interquartile Range). Data yang berada di luar batas $Q1 - 1,5 \times IQR$ dan $Q3 + 1,5 \times IQR$ diperiksa kewajaran fisiknya. Nilai yang tidak wajar secara meteorologis akan diimputasi, sementara nilai ekstrem yang wajar (seperti curah hujan tinggi pada musim hujan) dipertahankan.

4. **Normalisasi Data (*Feature Scaling*)**
Sebelum dilakukan proses normalisasi, dilakukan penentuan parameter yang digunakan sebagai variabel input dalam proses *clustering*. Berdasarkan karakteristik data cuaca yang tersedia, penelitian ini menggunakan enam parameter utama yaitu TN (suhu minimum), TX (suhu maksimum), TAVG (suhu rata-rata), RH_AVG (kelembapan rata-rata), RR (curah hujan), dan FF_AVG (kecepatan angin rata-rata).

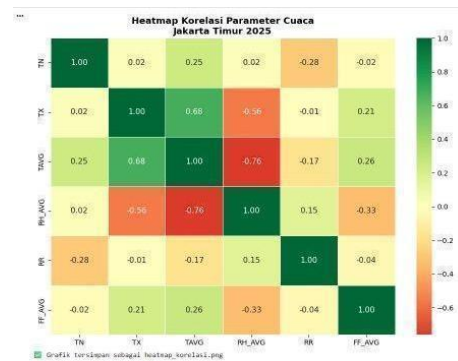
Parameter tersebut dipilih karena memiliki keterkaitan langsung dalam menggambarkan kondisi cuaca harian dan seluruh variabel memiliki bentuk numerik yang sesuai untuk proses perhitungan jarak pada algoritma K-Means Clustering.

Exploratory Data Analysis (EDA) dilakukan untuk memahami karakteristik distribusi dan hubungan antar parameter cuaca sebelum proses klusterisasi. Analisis EDA meliputi:

1. Statistik deskriptif (mean, median, std, min, max) untuk setiap parameter cuaca.
2. Visualisasi distribusi data menggunakan histogram dan *box plot*.
3. Analisis korelasi antar parameter menggunakan *Pearson Correlation Coefficient* yang divisualisasikan dalam bentuk heatmap.
4. Visualisasi time series untuk melihat tren dan pola musiman setiap parameter sepanjang tahun 2025.

Salah satu hasil EDA yang penting adalah analisis korelasi antar parameter yang divisualisasikan dalam bentuk heatmap pada Gambar 3.

Gambar 3 Heatmap Korelasi Antar



Parameter Cuaca Jakarta Timur 2025

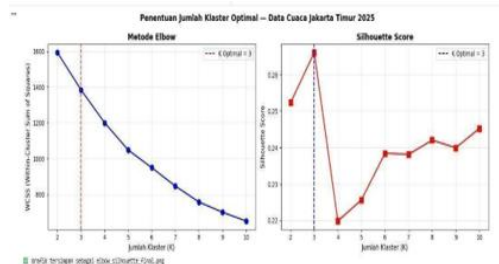
Penentuan nilai K (jumlah kluster) yang optimal dilakukan menggunakan dua metode evaluasi:

1. **Metode *Elbow***
Metode *Elbow* digunakan untuk menentukan jumlah kluster optimal dengan cara menghitung nilai *Within-Cluster Sum of Squares (WCSS)* untuk berbagai nilai K (K = 2 hingga K = 10). Nilai WCSS yang menurun tajam kemudian mulai mendatar membentuk "siku" menunjukkan jumlah kluster yang optimal. Grafik *Elbow* divisualisasikan menggunakan Matplotlib.
2. ***Silhouette Score***
Silhouette Score digunakan sebagai metode konfirmasi untuk memvalidasi hasil Metode *Elbow*. Nilai *Silhouette Score* berkisar antara -1 hingga +1, di mana nilai mendekati +1 menunjukkan kluster yang padat dan terpisah dengan baik. Nilai *Silhouette Score* dihitung untuk setiap nilai K menggunakan fungsi *silhouette_score* dari pustaka *scikit-learn*.

Tabel 3 Hasil Perhitungan WCSS dan Silhouette Score untuk K=2 hingga K=10

Jumlah Klaster (K)	WCSS (Elbow)	Silhouette Score
2	1.593,5	0,2524
3	1.384,6	0,2661 ← Tertinggi ✓
4	1.200,7	0,2199

5	1.048,3	0,2257
6	949,2	0,2385
7	846,6	0,2382
8	757,4	0,2420
9	700,2	0,2400
10	651,0	0,2453



Gambar 4 Grafik Metode Elbow dan Silhouette Score untuk Penentuan K Optimal

Berdasarkan hasil evaluasi menggunakan metode *Elbow* dan *Silhouette Score* pada tabel 3 dan gambar 4, jumlah kluster yang dipilih dalam penelitian ini adalah $K = 3$. Pemilihan tersebut didasarkan pada dua pertimbangan. Pertama, grafik *Elbow* menunjukkan penurunan nilai *Within kluster Sum of Squares (WCSS)* yang mulai melandai setelah $K = 3$, sehingga penambahan jumlah kluster setelah titik tersebut tidak memberikan peningkatan kualitas pengelompokan yang signifikan. Kedua, nilai *Silhouette Score* tertinggi diperoleh pada $K = 3$ sebesar 0,2661, yang menunjukkan bahwa pemisahan antar kluster pada jumlah tersebut lebih baik dibandingkan jumlah kluster lainnya. Oleh karena itu, penggunaan tiga kluster dinilai paling mampu merepresentasikan variasi karakteristik cuaca harian di wilayah Jakarta Timur selama tahun 2025.

Meskipun nilai *Silhouette Score* sebesar 0,2661, menunjukkan bahwa pemisahan antar kluster belum sepenuhnya tegas, namun masih dapat diterima karena karakteristik data cuaca bersifat kontinu, sehingga batas antar kelompok tidak terbentuk secara jelas. Hasil pengelompokan $K=3$ ini tetap dinilai valid dan dapat diterima secara fisis untuk data meteorologi. Karakteristik data cuaca harian bersifat kontinu dan dinamis, di

mana transisi fluktuasi parameter dari hari ke hari terjadi secara gradual tanpa adanya pembatas nilai yang ekstrem atau kaku di alam nyata. Sebagaimana dikonfirmasi dalam penelitian klimatologi serupa oleh Prasetyo (2023), pengklasteran data atmosferik perkotaan sering kali menghasilkan *Silhouette Score* di bawah 0.30 karena tingginya variabilitas harian, namun secara fisis mampu merepresentasikan diferensiasi kondisi cuaca (panas, normal, hujan) dengan sangat logis.

Proses *clustering* dilakukan menggunakan data harian sebanyak 365 observasi dengan enam parameter cuaca yaitu TN, TX, TAVG, RH_AVG, RR, dan FF_AVG sebagai variabel masukan (*input*).

Tabel 4 Distribusi Hari per Kluster Hasil K-Means Clustering

Kluster	Nama Kluster	Jumlah Hari	Proporsi (%)
0	Cuaca Panas Kering	201	55,1%
1	Cuaca Sejuk Berawan	140	38,4%
2	Cuaca Hujan Lebat	24	6,6%
Total	-	365	100%

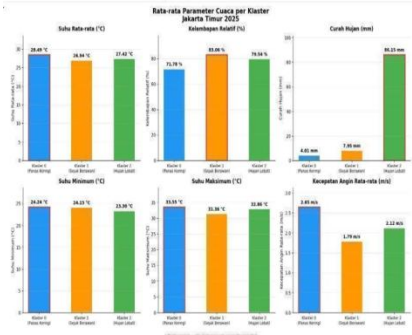
Tabel 4 menunjukkan bahwa hasil klusterisasi terhadap 365 data harian menghasilkan tiga kluster dengan jumlah anggota yang berbeda. Perbedaan jumlah anggota pada setiap kluster menunjukkan bahwa karakteristik cuaca di Jakarta Timur sepanjang tahun 2025 tidak tersebar secara merata, melainkan lebih sering memiliki karakteristik suhu relatif tinggi, curah hujan rendah, dan kelembapan yang lebih rendah sebagaimana direpresentasikan oleh Kluster 0.

Penamaan kluster dilakukan berdasarkan kombinasi nilai centroid seluruh parameter, bukan berdasarkan satu parameter tertentu. Oleh karena itu, interpretasi nama kluster tidak ditentukan berdasarkan bulan terjadinya, melainkan

berdasarkan kombinasi karakteristik parameter cuaca yang dimiliki masing-masing kluster. Interpretasi nama kluster dilakukan setelah membandingkan seluruh nilai centroid masing-masing parameter sehingga penamaan mencerminkan karakteristik dominan setiap kelompok, bukan hasil subjektif peneliti.

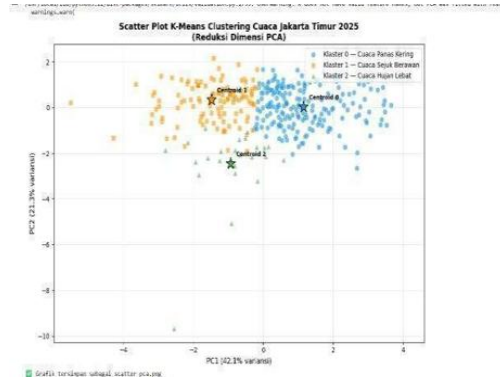
Berdasarkan nilai centroid, diketahui bahwa setiap kluster memiliki karakteristik parameter cuaca yang berbeda sehingga proses pengelompokan yang dilakukan algoritma K-Means mampu membedakan kondisi cuaca harian berdasarkan kombinasi suhu, kelembapan, curah hujan, dan kecepatan angin. Perbedaan karakteristik tersebut menunjukkan bahwa algoritma K-Means berhasil mengelompokkan data berdasarkan kemiripan pola parameter cuaca, sehingga setiap kluster merepresentasikan kondisi cuaca harian yang berbeda di wilayah Jakarta Timur.

Untuk menampilkan visualisasi hasil *clustering* digunakan grafik. Pada gambar 5 ditampilkan rata-rata enam parameter cuaca untuk masing-masing kluster secara terpisah, sehingga perbedaan karakteristik antar kluster dapat terlihat dengan jelas.



Gambar 5 Rata-rata Parameter Cuaca per Kluster di Jakarta Timur tahun 2025

Pada gambar 6, hasil K-Means Clustering ditampilkan dalam *scatter plot* dengan reduksi dimensi PCA untuk memvisualisasikan separasi antar kluster dalam ruang dua dimensi.



Gambar 6 Scatter Plot K-Means Clustering dengan Reduksi Dimensi PCA (PC1=42,1%, PC2=21,3%)

Visualisasi PCA digunakan untuk membantu melihat persebaran hasil clustering dalam ruang dua dimensi. Proses pembentukan kluster tetap dilakukan menggunakan algoritma K-Means terhadap enam parameter cuaca, sedangkan PCA hanya digunakan sebagai metode visualisasi untuk menggambarkan pola pemisahan antar kluster.

Hal tersebut menunjukkan bahwa sebagian observasi memiliki karakteristik cuaca yang saling berdekatan sehingga batas antar kluster tidak sepenuhnya terpisah, sejalan dengan nilai *Silhouette Score* yang relatif rendah.

SIMPULAN

Berdasarkan hasil penelitian analisis dan visualisasi pola parameter cuaca harian dan bulanan Jakarta Timur dengan K-Means Clustering, maka dapat disimpulkan dari hasil analisis eksploratif (Exploratory Data Analysis/EDA) dan analisis korelasi, diketahui setiap parameter cuaca memiliki karakteristik hubungan yang berbeda.

Suhu maksimum (TX) dan suhu rata-rata (TAVG) memiliki korelasi positif yang kuat, sedangkan suhu minimum (TN) hanya berkorelasi lemah terhadap TX dan TAVG, dan suhu rata-rata (TAVG) memiliki korelasi negatif yang kuat terhadap kelembapan relatif (RH_AVG). Meskipun terdapat korelasi antarparameter suhu, seluruh parameter tetap digunakan

dalam proses clustering karena masing-masing memiliki makna meteorologis yang berbeda serta berkontribusi dalam menggambarkan karakteristik kondisi cuaca secara menyeluruh.

Penerapan algoritma K-Means Clustering dipadukan dengan teknik visualisasi grafik (termasuk antarmuka interaktif Streamlit), terbukti efektif memetakan serta menyajikan pola parameter cuaca harian dan bulanan Jakarta Timur secara informatif. Hasil penelitian ini dapat dimanfaatkan sebagai bahan rujukan eksploratif bagi analisis klimatologi lokal maupun pengembangan model pembelajaran mesin selanjutnya

Hasil visualisasi menunjukkan distribusi kondisi cuaca di Jakarta Timur selama tahun 2025 didominasi oleh Klaster 0 (Cuaca Panas Kering), sedangkan Klaster 2 (Cuaca Hujan Lebat) muncul dengan frekuensi yang lebih sedikit dan lebih banyak ditemukan pada periode akhir tahun.

Visualisasi menggunakan grafik distribusi klaster serta reduksi dimensi Principal Component Analysis (PCA) menunjukkan bahwa meskipun terdapat sedikit tumpang tindih antar klaster, proses pengelompokan mampu membedakan karakteristik cuaca berdasarkan kombinasi parameter suhu, kelembapan, curah hujan, dan kecepatan angin.

Secara keseluruhan, penerapan algoritma K-Means Clustering mampu mengidentifikasi pola karakteristik cuaca harian berdasarkan data historis BMKG Jakarta Timur tahun 2025. Hasil pengelompokan yang diperoleh memberikan informasi mengenai variasi kondisi cuaca yang dapat digunakan sebagai dasar dalam analisis pola cuaca serta mendukung pengambilan keputusan pada bidang meteorologi dan pengelolaan lingkungan.

DAFTAR PUSTAKA

- Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). (2024), Buku Iklim DKI Jakarta: Analisis Kondisi Cuaca dan Iklim Tahun 2020–2024. Jakarta: BBMKG Wilayah II.
- Damayanti, S., Lestari, P., & Sofyan, A. (2023) 'Urban Heat Island Intensity and Its Relationship with Land Surface Temperature in Jakarta Metropolitan Area', IOP Conference Series: Earth and Environmental Science, vol. 944, no. 1, p. 012023, doi: 10.1088/1755-1315/944/1/012023.
- Fitriana, H. S., Sumantyo, J. T. S., & Kuze, H. (2022) 'Multi-temporal Analysis of Land Use/Land Cover Changes and Urban Heat Island Dynamics in Jakarta Using Remote Sensing and GIS', Remote Sensing Applications: Society and Environment, vol. 26, p. 100724, doi: 10.1016/j.rsase.2022.100724.
- Handyka. (2022) 'Pemanfaatan Python dan Google Colab Dalam Pembelajaran Statistika Deskriptif Handika STKIP Muhammadiyah Pagaralam', pp. 379–389.
- Hidayat, R. & Nugroho, S. (2024) 'Spatial and Temporal Variability of Extreme Rainfall in Jakarta and Its Suburbs: A Data-Driven Approach Using Clustering and Geostatistics', J. Hydrology: Regional Studies, vol. 51, p. 101644, doi: 10.1016/j.ejrh.2024.101644.
- J. Enterprise, J. (2023) 'Python untuk Analisis dan Visualisasi Data'.
- Kurniawan, R., Suratno, A., & Hidayati, R. (2022) 'Variabilitas Curah Hujan di Jakarta dan Hubungannya dengan Indeks Iklim Pasifik', J. Ilmu dan Inovasi Fisika, vol. 5, no. 1, pp. 1–10, doi: 10.24198/jiif.v5i1.30459.
- Maori, N. A., and Evanita, E. (2023) 'Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering', Simetris J. Tek. Mesin, Elektro dan Ilmu Komputer, 14(2), pp. 277–288, doi: 10.24176/simet.v14i2.9630.
- Prasetyo, R., Wijayanto, A. K., & Sukmara, R. H. (2023) 'Penerapan Machine Learning untuk Analisis Data Lingkungan: Studi Kasus Cuaca Perkotaan', J. Teknologi Informasi dan Ilmu Komputer (JTIK), vol. 9,

-
- no. 4, pp. 721–730, doi: 10.25126/jtiik.2022x4821.
- Rahmadenti, N. A. R. (2025) 'Analisis Pola dan Tren Kualitas Udara DKI Jakarta 2022-2025 Menggunakan K-Means Clustering', *J. Inform. dan Tek. Elektro Terap.*, 13(Oct), 3S1, doi: 10.23960/jitet.v13i3S1.8141.
- Saputra, D. M. & Novianti, P. (2021) 'Optimasi Klasterisasi Data Iklim Kota Bandung Menggunakan Algoritma K-Means dengan Reduksi Dimensi PCA', *J. RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 714–722, doi: 10.29207/resti.v5i4.3187.
- Surbakti, N. M. *et al.* (2024) 'Penggunaan Bahasa Pemrograman Python dalam Pembelajaran Kalkulus Fungsi Dua Variabel', *Algoritma. J. Mat. Ilmu Pengetah. Alam, Kebumihan dan Angkasa*, 2(3), pp. 98–107, doi: 10.62383/algoritma.v2i3.67.
- Wibowo, D. H., Santoso, P. B., & Yunianto, M. (2022) 'Perbandingan Metode K-Means dan Fuzzy C-Means dalam Klasterisasi Data Curah Hujan Musiman di Daerah Istimewa Yogyakarta', *J. Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 11,