ISSN: 1978-8126
e-ISSN: 2527-7340

Komparasi Algoritma K-Means dan DBSCAN dalam Klasterisasi Kualitas Udara di Wilayah Jabodetabek Berdasarkan Indeks Standar Pencemar Udara

Ahmad Fadhil Rizqi[1], Meiyin Monica Amilia Putri[2], Aulia Pinkasari[3], Fathoni[4]*

Sistem Informasi, Ilmu Komputer, Universitas Sriwijaya
Jl. Raya Palembang - Prabumulih No.KM.32, Indralaya Indah, Sumatera Selatan, Indonesia
09031282328058@student.unsri.ac.id[1], 09031182328097@student.unsri.ac.id[2],
09031182328011@student.unsri.ac.id[3] , fathoni@unsri.ac.id[4]

Submitted : **30/03/2026**; Reviewed :**02/04/2026**; Accepted :**26/04/2026**; Published : **30/04/2026**

Abstract

Air quality is a critical environmental issue because it directly impacts public health and the sustainability of urban environments. The Greater Jakarta area, as a hub of economic activity and population mobility, faces high levels of air pollution; however, a comprehensive understanding of air quality patterns remains limited due to the complexity of the data and the variation in pollutant parameters. This study aims to analyze and compare the performance of clustering algorithms in grouping air quality levels based on Air Pollutant Standard Index (ISPU) data in the Greater Jakarta area. The methods used were K-Means and DBSCAN with six input parameters PM2.5, PM10, SO_2, NO_2, O_3, and CO—which underwent data preprocessing, the clustering process, and evaluation using the Silhouette Score and the Davies-Bouldin Index (DBI). The results indicate that K-Means with three clusters yields the best performance, achieving a Silhouette Score of 0.324 and a DBI of 1.142, while DBSCAN achieves a Silhouette Score of 0.013 and a DBI of 1.240. These findings indicate that K-Means is more effective in forming cohesive groupings of air quality patterns and can be utilized as a basis for decision-making in the monitoring and control of urban air quality.

Keywords: DBSCAN, ISPU, clustering, k-means, air quality

Abstrak

Kualitas udara merupakan isu lingkungan yang krusial karena berdampak langsung terhadap kesehatan masyarakat dan keberlanjutan lingkungan perkotaan. Wilayah Jabodetabek sebagai pusat aktivitas ekonomi dan mobilitas penduduk menghadapi tingkat pencemaran udara yang tinggi, namun pemahaman terhadap pola kualitas udara secara menyeluruh masih terbatas akibat kompleksitas data dan variasi parameter pencemar. Penelitian ini bertujuan untuk menganalisis dan membandingkan performa algoritma *clustering* dalam mengelompokkan tingkat kualitas udara berdasarkan data Indeks Standar Pencemar Udara (ISPU) di wilayah Jabodetabek. Metode yang digunakan adalah K-Means dan DBSCAN dengan enam parameter masukan, yaitu PM2.5, PM10, SO_2, NO_2, O_3, dan CO, yang melalui tahapan pra-pemrosesan data, proses *clustering*, serta evaluasi menggunakan *Silhouette Score* dan *Davies-Bouldin Index* (DBI). Hasil penelitian menunjukkan bahwa K-Means dengan tiga klaster menghasilkan kinerja terbaik dengan nilai *Silhouette Score* sebesar 0,324 dan DBI sebesar 1,142, sedangkan DBSCAN memperoleh *Silhouette Score* sebesar 0,013 dan DBI sebesar 1,240. Temuan ini mengindikasikan bahwa K-Means lebih efektif dalam membentuk pengelompokan pola kualitas udara yang kohesif dan dapat dimanfaatkan sebagai dasar pengambilan keputusan dalam pemantauan serta pengendalian kualitas lingkungan udara perkotaan.

Kata kunci: DBSCAN*, ISPU,* klasterisas*i, k-means clustering,* kualitas udara

1. Pendahuluan

Indonesia menghadapi tantangan serius dalam pengendalian pencemaran udara, di mana berdasarkan data dari IQAir tahun 2024, Indonesia tercatat menempati peringkat ke-15 sebagai negara dengan tingkat polusi udara tertinggi secara global dengan konsentrasi $PM_{2.5}$ mencapai 35,54 µg/m³ [1]. Ketika berbicara tentang

polusi udara di Indonesia, semua mata tertuju pada Jakarta - ibu kota Indonesia yang besar dengan populasi perkotaan 35,4 juta [2].

Sebagai pusat pertumbuhan ekonomi dan mobilitas nasional, wilayah Jabodetabek memiliki konsentrasi emisi yang tinggi berasal dari segmen transportasi dan industri. Akumulasi polutan ini berkontribusi besar terhadap tingginya angka gangguan pernapasan di tengah masyarakat, di antaranya adalah Infeksi Saluran Pernapasan Akut (ISPA), serangan asma, dan komplikasi paru obstruktif kronis (PPOK) [3].

Ketika daerah metropolitan mengalami polusi udara yang signifikan, dampak terhadap kesehatan dan beban ekonomi berlipat ganda dan menjadi jauh lebih tinggi. Oleh karena itu, pemantauan kualitas udara di kota-kota besar telah meningkat pesat dalam beberapa tahun terakhir untuk menginformasikan kepada masyarakat tentang tindakan pencegahan yang diperlukan. Pemerintah Indonesia berdasarkan Peraturan Menteri Lingkungan Hidup dan Kehutanan Nomor 14 Tahun 2020 penggunaan Indeks Standar Pencemar Udara (ISPU) ditetapkan sebagai parameter standar pelaporan kualitas udara untuk memberikan informasi yang akurat tentang ancaman kesehatan tersebut.

Data ISPU yang dihasilkan dari stasiun pemantauan memiliki karakteristik volume yang besar dan parameter yang kompleks seperti PM_{10}, $PM_{2,5}$, SO_2, CO, O_3, dan NO_2. ISPU berfungsi sebagai indikator untuk memantau dan melaporkan kondisi kualitas udara secara terstruktur. Nilai yang dihasilkan merepresentasikan tingkat pencemaran udara serta kemungkinan dampaknya terhadap kesehatan masyarakat dalam bentuk angka tanpa satuan khusus [4].

Penelitian sebelumnya menitikberatkan pada pemilihan algoritma klasifikasi paling akurat untuk mengelompokkan kualitas udara menggunakan pendekatan *supervised learning* berdasarkan data berlabel yang tersedia [5]. Namun, studi tersebut hanya dilakukan di wilayah Jakarta, sehingga belum mampu merepresentasikan variasi spasial dan kompleksitas polutan di area yang lebih luas.

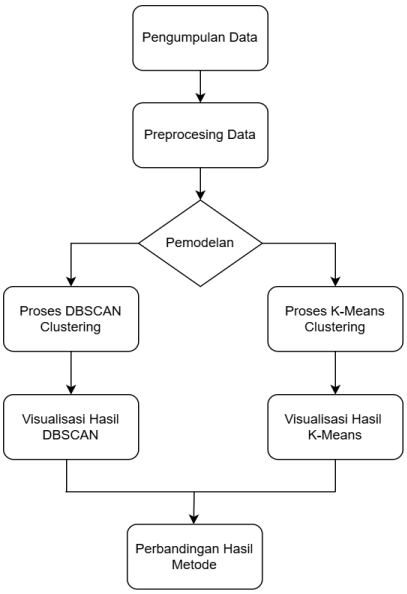
Sepanjang kurun waktu 2022 sampai 2025, dinamika tingkat polusi di ibukota Jakarta pernah dieksplorasi melalui teknik pengelompokan berbasis K-Means pada sebuah kajian relevan [6]. Namun, studi tersebut masih terbatas pada cakupan area tertentu dan belum melakukan perbandingan menyeluruh antara algoritma *clustering* berbasis *centroid* dan *density-based* dalam satu kerangka evaluasi terpadu. Selain itu, parameter penilaian seperti *Silhouette Score* dan *Davies-Bouldin Index*, dan K-Distance Graph belum digunakan secara komprehensif untuk menilai kualitas cluster.

Studi mengenai kualitas udara berbasis ISPU di wilayah Jakarta yang telah dilakukan sebelumnya, antara lain oleh Fajar et al, sebagian besar berfokus pada pemetaan kualitas udara dan komparasi beberapa algoritma klasterisasi (K-Means, GMM, DBSCAN) pada skala kota. Penelitian-penelitian tersebut umumnya mengkaji kinerja algoritma secara terpisah dan kurang menekankan penilaian komparatif yang mendalam antara K-Means dan DBSCAN dalam konteks sebaran spasial polusi udara di wilayah Jabodetabek. Hal ini menimbulkan kesenjangan penelitian (research gap) yang signifikan, mengingat masih terbatasnya studi yang secara eksplisit membandingkan kedua algoritma pada skala regional Jabodetabek dan mengintegrasikan hasil kluster dengan analisis pola spasial-temporal guna mendukung rekomendasi lokasi prioritas mitigasi polusi udara secara dinamis dan terukur [7].

2. Metodologi

2.1. Kerangka Penelitian

Tahapan Penelitian digunakan sebagai panduan untuk menjelaskan proses pelaksanaan penelitian secara runtut dan sistematis, sehingga proses penelitian dapat berlangsung sesuai dengan rencana serta tujuan yang telah ditetapkan.



Gambar 1. Kerangka Penelitian

2.1. Pengumpulan Data

Data dalam penelitian ini diperoleh melalui proses akuisisi data sekunder menggunakan antarmuka pemrograman aplikasi (Application Programming Interface/API) dari Open-Meteo Air Quality [8]. Proses pengumpulan data difokuskan pada wilayah metropolitan Jabodetabek dengan menetapkan sepuluh titik pemantauan strategis yang mewakili keragaman spasial di kawasan tersebut. Lokasi pengambilan sampel mencakup Jakarta Pusat, Jakarta Selatan, Jakarta Utara, Kota Bogor, Kabupaten Bogor, Kota Depok, Kota Tangerang, Tangerang Selatan, Kota Bekasi, dan Kabupaten Bekasi. Dari setiap lokasi tersebut, ditarik parameter polutan udara utama yang meliputi partikulat halus $PM_{2,5}$ dan PM_{10}serta gas pencemar lainnya seperti SO_2, CO, O_3, dan NO_2. Pemilihan Open-Meteo sebagai sumber data didasarkan pada kemampuannya menyediakan informasi polusi udara berbasis koordinat geografis yang presisi [8], yang kemudian dikonversi menjadi format numerik sesuai dengan parameter Indeks Standar Pencemar Udara (ISPU) yang berlaku di Indonesia.

2.2. *Preprocessing*

Pre-Processing adalah proses awal dalam mengelola data mentah agar bisa digunakan oleh model atau analisis lanjutan [9]. Tahapan pre-processing meliputi beberapa langkah penting, pertama adalah data selection, yaitu memilih data atau variabel yang relevan dan membuang data yang tidak relevan atau bahkan dapat mengganggu analisis jika digunakan. Kedua adalah data cleaning, yang berfungsi untuk mengurangi error pada data, mengatasi ketidakkonsistenan, serta membuang data yang tidak relevan. Terakhir, data transformation merupakan penyesuaian struktur dan representasi data agar siap diproses pada tahapan analisis lanjutan.

2.3. Algoritma

Tahapan pemetaan tingkat pencemaran udara di seputar Jabodetabek dieksekusi dengan memanfaatkan teknik *clustering*. Model analitik yang diujikan dalam penelitian ini mencakup algoritma berbasis *centroid* K-Means dan algoritma berbasis kepadatan spasial DBSCAN

2.4. Tahapan Clustering

Tahapan clustering pada penelitian ini dimulai dengan menentukan dasar pengelompokkan yaitu memilih parameter ISPU yang digunakan sebagai fitur analisis. Metode *clustering* bertujuan menempatkan objek-objek yang serupa dalam kelompok yang sama dan menjaga jarak antar kelompok agar tetap optimal [10].

2.5. Model K-Means Clustering

Algoritma *K-Means* adalah metode yang sering digunakan untuk mengelompokkan data berdasarkan kesamaan antar fitur. Proses ini dimulai dengan menentukan jumlah klaster (*k*) dan memilih titik pusat awal

(*centroid*). Selanjutnya, jarak setiap data dihitung terhadap masing-masing *centroid*, kemudian data ditempatkan ke klaster terdekat. Proses ini diulang terus-menerus hingga posisi *centroid* stabil atau Sebelum melakukan pengelompokan, data biasanya melalui tahap evaluasi menggunakan metode Elbow untuk menentukan jumlah klaster yang optimal [11].

Pendekatan ini menganalisis hubungan antara jumlah klaster dan nilai *Within-Cluster Sum of Squares* (WCSS). Titik optimal ditandai saat grafik posisi data mulai membentuk pola seperti siku (*elbow*) [11]. Setelah jumlah klaster ditetapkan, jarak antar titik data dan *centroid* dihitung menggunakan rumus jarak *Euclidean*, yaitu:

$$d(x_i, c_k) = \sqrt{\sum_{j=1}^{n} (x_{ij} - c_{kj})^2} \quad (1)$$

Berdasarkan persamaan jarak Euclidean tersebut, variabel x_i berperan sebagai representasi dari entitas data ke-i, sedangkan c_k merujuk pada titik pusat atau centroid untuk kelompok klaster ke-k . Simbol n mendefinisikan total fitur atau dimensi parameter yang digunakan dalam dataset, seperti konsentrasi $PM_{2,5}$, PM_{10}, CO, NO_2. Setelah prosedur klasifikasi data berakhir, titik centroid akan mengalami penyesuaian posisi secara iteratif dengan mengalkulasi nilai rata-rata dari keseluruhan objek data yang telah tergabung di dalam setiap klaster tersebut.

$$c_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i \quad (2)$$

Persamaan di atas mendefinisikan $c_k$sebagai pusat atau centroid dari klaster ke-k, yang nilainya dikalkulasi melalui penjumlahan seluruh objek data x_i yang tergabung dalam kelompok tersebut dibagi dengan N_k atau total jumlah data di dalamnya . Pemanfaatan algoritma K-Means memiliki keunggulan signifikan dalam hal efisiensi operasional, terutama saat memproses dataset dengan volume yang besar. Melalui pendekatan ini, berbagai wilayah dapat dikategorikan secara otomatis berdasarkan parameter pencemar udara seperti PM_{10}, CO, dan NO_2 guna memetakan area dengan tingkat risiko polusi yang tinggi. Wawasan berbasis data tersebut sangat krusial dalam memberikan landasan bagi perumusan kebijakan kesehatan lingkungan yang lebih presisi dan terintegrasi [12].

2.6 Model *DBSCAN*

Algoritma DBSCAN (Density-Based Spatial Clustering of Applications with Noise) dapat mengidentifikasi klaster dalam bentuk dan bentuk yang bervariasi serta efektif dalam menangani data yang mengandung noise [13]. Metode ini bekerja berdasarkan konsep kepadatan dan hubungan antar titik data, yang ditentukan oleh dua parameter utama, yaitu radius maksimum (epsilon) serta jumlah minimum titik dalam suatu klaster.

Dalam algoritma DBSCAN, terdapat dua parameter utama yang menentukan proses pembentukan klaster, yaitu radius maksimum (ε / epsilon) dan jumlah titik minimum dalam satu klaster (MinPts). Parameter epsilon merupakan batas jarak maksimum antara suatu titik dengan titik lain yang masih dianggap saling berdekatan, sedangkan MinPts adalah jumlah minimum titik yang harus berada dalam radius epsilon untuk dapat membentuk klaster yang valid. Konsep kepadatan inilah yang membedakan DBSCAN dengan metode klaster lain seperti K-Means, karena DBSCAN tidak bergantung pada bentuk klaster tertentu dan dapat mendeteksi noise secara eksplisit [14].

Objek dalam DBSCAN dibagi menjadi dua jenis, yaitu Objek Tepi dan Objek *Core*. Objek yang berada dekat dengan *core point* disebut sebagai objek tepi, sedangkan *core point* merupakan titik yang memiliki tetangga cukup dalam radius *epsilon* yang ditentukan. Titik yang tidak ada *core* maupun tepi dikategorikan sebagai anomaly atau *outlier*.

Tahapan algoritma DBSCAN dilakukan dengan beberapa langkah. Pertama, tentukan nilai *epsilon* dan MinPts, lalu pilih titik awal secara acak. Selanjutnya, lakukan proses pengulangan untuk semua titik sampai seluruh data diproses. Setelah itu, hitung ketercapaian kepadatan terhadap setiap titik menggunakan persamaan:

$$E(X,Y) = \sqrt{\sum_{i=0}^{n} (x_i - Y_i)^2} \quad (3)$$

Suatu titik akan dikategorikan sebagai core point apabila jumlah titik tetangga yang berada dalam radius epsilon sama dengan atau melebihi nilai MinPts. Jika suatu objek data teridentifikasi sebagai titik tepi (*border point*) dan tidak memiliki tetangga baru akibat kurangnya syarat densitas minimum, algoritma akan menghentikan ekspansi pada area tersebut dan beralih mengevaluasi objek data berikutnya.

2.7. Metode Evaluasi

Metode Evaluasi digunakan untuk menilai seberapa baik hasil pengelompokkan yang diperoleh dari proses *clustering.* Pada penelitian ini kualitas cluster diuji menggunakan tiga ukuran yaitu *Silhouette Score, k-Distance Graph* dan *Davies Bouldin.* Ketika metode tersebut membantu melihat apakah data dalam satu cluster sudah cukup mirip satu sama lain [15].

2.8. S*ilhouette Score*

Kualitas pengelompokan dari metode seperti K-Means dapat divalidasi melalui perhitungan *Silhouette Score*, yang secara khusus meninjau aspek jarak dan kerapatannya [16]. Hasil komputasi dari metrik ini mengindikasikan tingkat kesesuaian posisi suatu entitas di dalam kelompoknya saat ini relatif terhadap kelompok lain di sekitarnya. Adapun rumus yang digunakan adalah:

$$\mathrm{S(i)}\ \frac{b(i) - a(i)}{max(a(i),b(i))} \quad (4)$$

Nilai S(i) yang diperoleh dari setiap objek kemudian dirata-ratakan untuk menentukan *silhouette coefficient* global. Skala pengukuran pada pengujian ini dibatasi dari -1 sebagai titik terendah hingga 1 sebagai titik tertinggi. Perolehan angka di atas nol mengartikan bahwa sebuah *data point* telah menempati kelompok yang tepat, sedangkan hasil di bawah nol memberi sinyal bahwa data tersebut memiliki relevansi yang lebih kuat terhadap kelompok lain.

2.9. *K-Dist Graph*

Dalam penelitian ini, sebelum menerapkan algoritma *DBSCAN*, parameter *epsilon* (ε) ditentukan menggunakan *k-distance graph*. Grafik ini dibuat dengan menghitung jarak setiap titik data ke *k-nearest neighbor* terdekatnya, lalu jarak tersebut diurutkan dan diplot untuk membentuk kurva *k-distance*. Titik pada grafik yang menunjukkan perubahan tajam atau *elbow* dipilih sebagai nilai ε yang optimal, karena mencerminkan perbedaan kepadatan data antara area klaster dan area noise. Metode ini membantu mengidentifikasi radius lingkungan yang tepat di mana DBSCAN dapat membentuk klaster berdasarkan kepadatan data, sehingga meningkatkan kualitas hasil pengelompokan [17].

2.10. Davies Bouldin Index (DBI)

Davies-Bouldin Index (DBI) digunakan sebagai indikator untuk mengukur kualitas hasil pengelompokan sekaligus menilai performa algoritma clustering [18]. Nilai DBI mencerminkan hubungan yang searah dengan variasi dalam klaster (*within-class*) dan berbanding terbalik dengan jarak antar klaster (*between-class*). Pada penelitian ini, DBI dimanfaatkan sebagai alat evaluasi dalam menilai kualitas struktur klaster yang terbentuk. Secara umum, pendekatan validasi clustering terbagi menjadi dua, yaitu validasi internal dan validasi eksternal [19].

3. Hasil dan Pembahasan

Analisis perbandingan performa algoritma K-Means dan DBSCAN dalam pengelompokan kualitas udara di wilayah Jabodetabek melibatkan serangkaian tahapan komputasi yang dimulai dari akuisisi data melalui API *Open-Meteo*, pra-pemrosesan data untuk menjamin validitas input, hingga evaluasi performa model menggunakan metrik *Silhouette Score* dan *Davies Bouldin Index*. Implementasi algoritma K-Means dengan optimasi jumlah klaster melalui metode *Elbow* berhasil mengidentifikasi pola pengelompokan yang stabil, sementara DBSCAN memberikan wawasan tambahan mengenai karakteristik spasial polutan melalui deteksi *noise* atau pencilan data. Hasil evaluasi menunjukkan bahwa tahap pra-pemrosesan dan pemilihan parameter secara adaptif melalui K-Dist Graph berdampak signifikan terhadap kualitas klaster yang terbentuk, di mana K-Means cenderung menghasilkan kelompok yang lebih kompak dibandingkan DBSCAN pada dataset Indeks Standar Pencemar Udara (ISPU) ini.

3.1.1. Pengumpulan Data

Proses penelitian dimulai dengan tahap pengumpulan data sekunder yang dilakukan melalui antarmuka pemrograman aplikasi (*Application Programming Interface*/API) dari *Open-Meteo Air Quality*. Data yang diperoleh mencakup sepuluh titik wilayah strategis di Jabodetabek yang terdiri dari Jakarta Pusat, Jakarta Selatan, Jakarta Utara, Kota Bogor, Kabupaten Bogor, Kota Depok, Kota Tangerang, Tangerang Selatan, Kota Bekasi, dan Kabupaten Bekasi. Pengumpulan data ini difokuskan pada enam parameter utama pencemar udara sesuai dengan regulasi Indeks Standar Pencemar Udara (ISPU) di Indonesia. Karakteristik dari dataset yang telah dikumpulkan kemudian dianalisis secara statistik untuk memahami distribusi nilai polutan sebelum dilakukan proses klasterisasi, sebagaimana disajikan pada Tabel 1.

Tabel 1. Statistik Parameter Kualitas Udara Jabodetabek

Variabel	Min	Max	Mean	Standar Deviasi
PM_{10}	5.2	264.5	47.33	33.06
$PM_{2,5}$	4.8	260.8	46.56	33.05
CO	0.0	13489	1217.81	1179.81
NO_2	3.5	171	45.88	28.31
SO_2	6.1	135	43.50	18.20
O_3	0.0	289	39.42	43.17

Tabel 1 menyajikan ringkasan statistik dari dataset yang mencakup nilai minimum, maksimum, rata-rata (*mean*), dan standar deviasi dari setiap parameter polutan. Berdasarkan hasil pengolahan data tersebut, terlihat bahwa rata-rata konsentrasi $PM_{2,5}$ di wilayah Jabodetabek mencapai 46.56 µg/m³. Nilai ini secara signifikan lebih tinggi jika dibandingkan dengan data nasional yang dilaporkan oleh IQAir pada tahun 2024 yang sebesar 35,54 µg/m³. Tingginya nilai standar deviasi pada parameter Karbon Monoksida (CO) yang mencapai 1179.82 menunjukkan adanya fluktuasi konsentrasi yang sangat tajam antar lokasi pemantauan, yang mengindikasikan perbedaan kondisi lingkungan yang ekstrem antara pusat kota dan wilayah penyangga. Kondisi anomali ini memperkuat urgensi penggunaan metode klasterisasi yang mampu menangani variasi data untuk memetakan kualitas udara secara lebih presisi.

3.1.2. *Data Preprocessing*

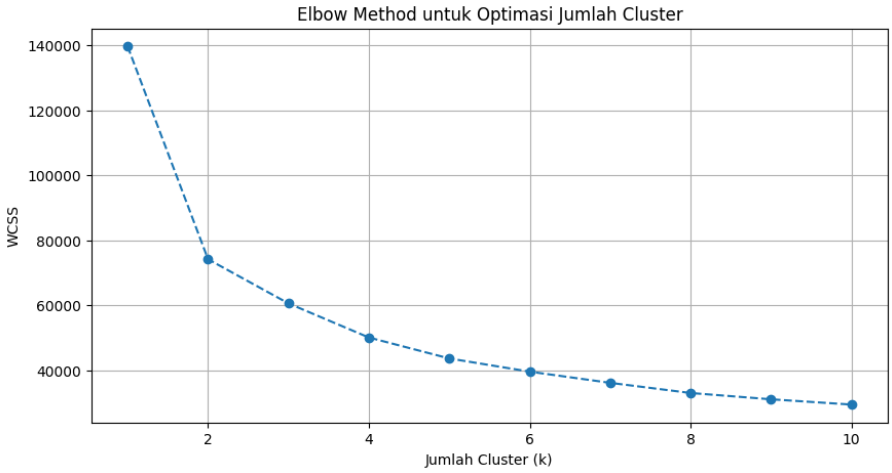
Implementasi pra-pemrosesan pada penelitian ini dimulai dengan mengintegrasikan data dari sepuluh wilayah Jabodetabek menjadi satu kesatuan dataset terpadu. Selama proses penggabungan, dilakukan penyelarasan nama kolom dan penambahan label lokasi untuk memastikan konteks spasial tetap terjaga. Pembersihan data dilakukan dengan menghapus baris yang memiliki nilai kosong pada parameter polutan utama, yang menghasilkan dataset bersih sebanyak (masukkan jumlah baris data bersih Anda) baris yang siap diolah lebih lanjut.

ISSN: 1978-8126
e-ISSN: 2527-7340

Tahap berikutnya adalah transformasi data melalui penghitungan nilai Indeks Standar Pencemar Udara (ISPU) menggunakan fungsi interpolasi linier yang telah diprogram. Melalui proses ini, konsentrasi mentah dari polutan seperti $PM_{2,5}$, PM_{10}, hingga CO berhasil dikonversi menjadi angka indeks yang seragam dalam rentang 0 hingga 500. Fitur baru berbasis ISPU ini memberikan representasi yang lebih akurat mengenai tingkat kualitas udara dibandingkan hanya menggunakan angka konsentrasi fisik. Tahap ini diakhiri dengan standardisasi fitur menggunakan StandardScaler untuk memastikan seluruh variabel memiliki bobot yang seimbang, sehingga fitur dengan nominal besar seperti Karbon Monoksida (CO) tidak mendominasi perhitungan jarak selama proses pembentukan klaster pada algoritma K-Means dan DBSCAN.

3.1.3. *Optimasi Parameter (Elbow Method)*

Tahap pertama yang dilakukan adalah menentukan jumlah cluster (K) dengan menggunakan *Elbow Method*. Proses ini melibatkan pemetaan nilai *inertia* pada sejumlah variasi K, diikuti dengan analisis grafik untuk menentukan *elbow point*, yaitu jumlah klaster yang paling sesuai. Visualisasi hasilnya tedapat pada Gambar 2.



Gambar 2. Hasil Elbow Method

Gambar 2 mendemonstrasikan visualisasi *Elbow Method* guna mengevaluasi parameter *cluster* terbaik untuk model K-Means. Pada grafik tersebut, patahan kurva (*elbow point*) terbentuk secara presisi di angka K=3. Hal ini mengindikasikan bahwa pembagian menjadi tiga kelompok merupakan skenario paling ideal, mengingat reduksi nilai WCSS sesudahnya cenderung melandai dan tidak lagi menunjukkan perubahan drastis.

3.1.4. Evaluasi Validitas K-Means

Tahap evaluasi dilakukan untuk menilai sejauh mana klaster yang dibentuk oleh algoritma *K-Means* menunjukkan kualitas pengelompokan. Penilaian ini memanfaatkan metrik validasi internal, yakni *Silhouette Score* dan *Davies-Bouldin Index (DBI)*, yang membantu menentukan tingkat kedekatan data dalam klaster (*cohesion*) serta pemisahan antar klaster (*separation*). Hasil ringkasan evaluasi tersebut ditampilkan pada Tabel 2.

Tabel 2. Tabel Hasil Evaluasi Validitas K-Means

Metrik Evaluasi	Skor
Silhouette Score	0.324464
Davies-Bouldin Index	1.142941

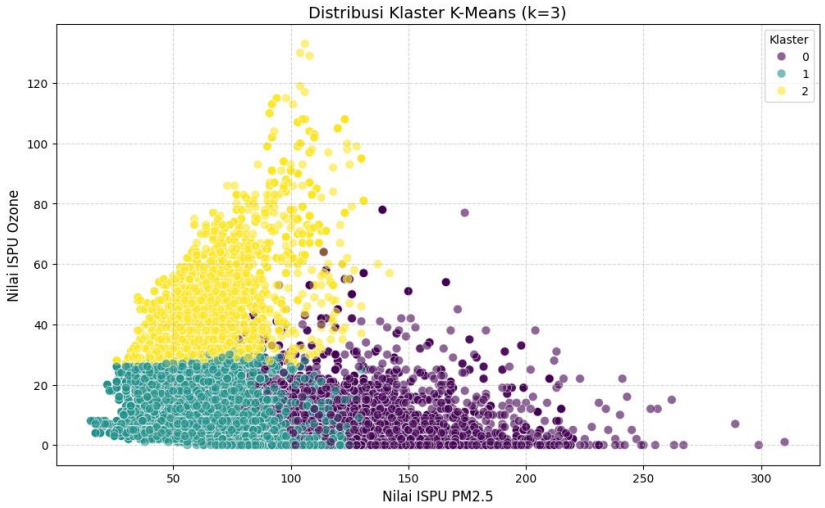
Hasil evaluasi validitas K-Means ditampilkan pada Tabel 2, menggunakan metrik Silhouette Score dan Davies-Bouldin Index (DBI). Nilai Silhouette Score sebesar 0,3245 mengindikasikan bahwa kualitas struktur klaster sudah cukup baik, walaupun pemisahan antar klaster masih berada pada tingkat menengah.

ISSN: 1978-8126
e-ISSN: 2527-7340

Sementara itu, nilai Davies-Bouldin Index (DBI) sebesar 1,1429 menunjukkan bahwa tingkat kekompakan dalam klaster serta jarak pemisahan antar klaster berada pada kondisi yang cukup baik.

3.1.5. Interpretasi Hasil Clustering K-Means

Pola kualitas udara di wilayah Jabodetabek berhasil dikelompokkan menjadi tiga kelompok utama melalui proses pengelompokan data melalui algoritma K-Means dengan nilai klaster optimal k = 3. Pembentukan klaster didasarkan pada enam parameter ISPU, sehingga setiap kelompok merepresentasikan karakteristik pencemaran udara yang berbeda. Visualisasi sebaran menunjukkan pemisahan yang jelas, terutama pada variabel $PM_{2.5}$ dan Ozone sebagai pemicu dominan variasi polusi. Distribusi sebaran klaster ditunjukkan pada gambar 3.



Gambar 3. Distribusi klaster Kmeans

Rangkuman statistik deskriptif nilai rata-rata ISPU pada tiap klaster hasil clustering K-Means disajikan pada Tabel 3, yang menjadi dasar interpretasi karakteristik masing-masing kelompok polutan.

Tabel 3. Rata-rata Nilai ISPU per Klaster (K-Means)

Klaster	PM_{10}	$PM_{2,5}$	SO_2	NO_2	CO	O_3	ISPU Final
Klaster 0	68.2	132.2	53.2	48.3	27.7	8	132.3
Klaster 1	28.4	64.3	34.4	20.6	10.0	10.1	64.4
Klaster 2	30.9	68.5	32.9	14.4	8.4	47.6	68.9

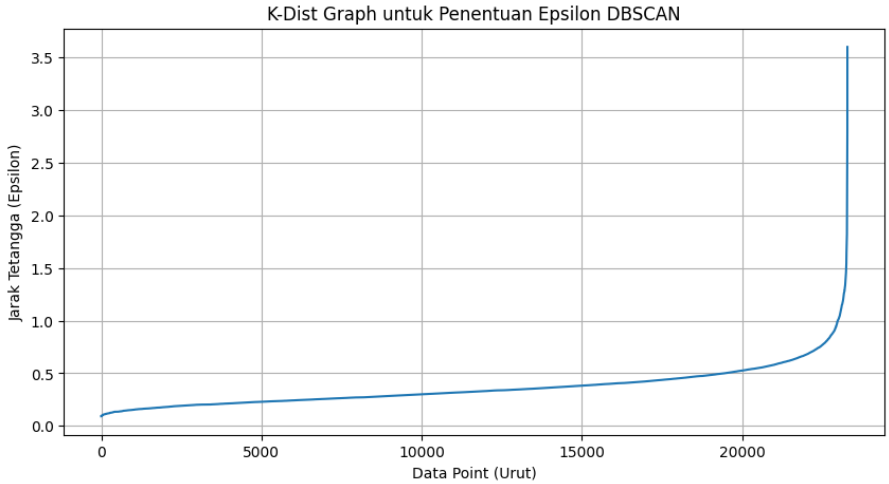
Berdasarkan nilai rata-rata ISPU pada tiap klaster, karakteristik kualitas udara pada masing-masing kelompok dapat dijelaskan sebagai berikut. Klaster 0 menggambarkan kondisi kualitas udara Tidak Sehat, dengan dominasi nilai PM2.5 tertinggi dibandingkan polutan lain yang berada pada tingkat menengah. Nilai ISPU final pada klaster ini merupakan yang paling tinggi, sehingga menunjukkan tekanan polusi partikulat yang signifikan, yang umumnya berasal dari emisi kendaraan, kawasan industri, atau pembakaran terbuka.

Klaster 1 menunjukkan kategori Sedang dengan nilai ISPU relatif rendah pada hampir semua polutan, termasuk Ozone yang memiliki tingkat terendah dibandingkan klaster lainnya, mencerminkan kondisi udara yang lebih stabil dan bersih serta merupakan kondisi yang paling sering muncul berdasarkan jumlah data. Sementara itu, klaster 2 juga berada pada kategori Sedang, namun memiliki karakteristik berbeda karena Ozone berada pada tingkat menengah, lebih tinggi daripada klaster 1, sedangkan polutan lain seperti PM10, PM2.5, dan NO_2 berada pada tingkatan sedang. Penetapan karakteristik klaster ini disusun dengan melihat dominasi parameter polutan tertentu pada tiap klaster sesuai praktik umum dalam penelitian kualitas udara menggunakan K-Means clustering untuk mengidentifikasi pola kategori kualitas udara yang berbeda [20].

ISSN: 1978-8126
e-ISSN: 2527-7340

3.1.6. Optimasi Parameter (K-Dist Graph)

Gambar 3 menunjukkan *K-Distance Graph* yang digunakan untuk menentukan nilai parameter epsilon (ε) pada algoritma DBSCAN. Pada grafik ini ditampilkan jarak ke tetangga terdekat ke-k secara berurutan dari yang terkecil hingga terbesar. Penentuan epsilon dilakukan dengan mengidentifikasi titik siku (*elbow point*), di mana terdapat lonjakan jarak yang jelas.



Gambar 4. *K-Dist Graph*

Berdasarkan grafik pada gambar 3 tersebut, terlihat adanya kenaikan yang cukup tajam pada bagian akhir kurva, yang mengindikasikan batas pemisahan antara *cluster* dan *noise*. Oleh karena itu, nilai epsilon dipilih pada titik sebelum kenaikan tajam tersebut, karena dianggap sebagai nilai optimal untuk membentuk cluster yang representatif.

3.1.7. Evaluasi Validitas DBSCAN

Tahap evaluasi ini bertujuan menilai kualitas klaster yang diperoleh dari algoritma DBSCAN. Penilaian dilaksanakan dengan menggunakan metrik validitas internal, seperti *Silhouette Score* dan *Davies-Bouldin Index* (DBI) digunakan sebagai metrik evaluasi untuk mengukur tingkat kekompakan dalam klaster (*cohesion*) serta jarak pemisah antar klaster (*separation*). Hasil perhitungan tersebut ditampilkan pada Tabel 4.

Tabel 4. Tabel Hasil Pengujian Validitas DBSCAN

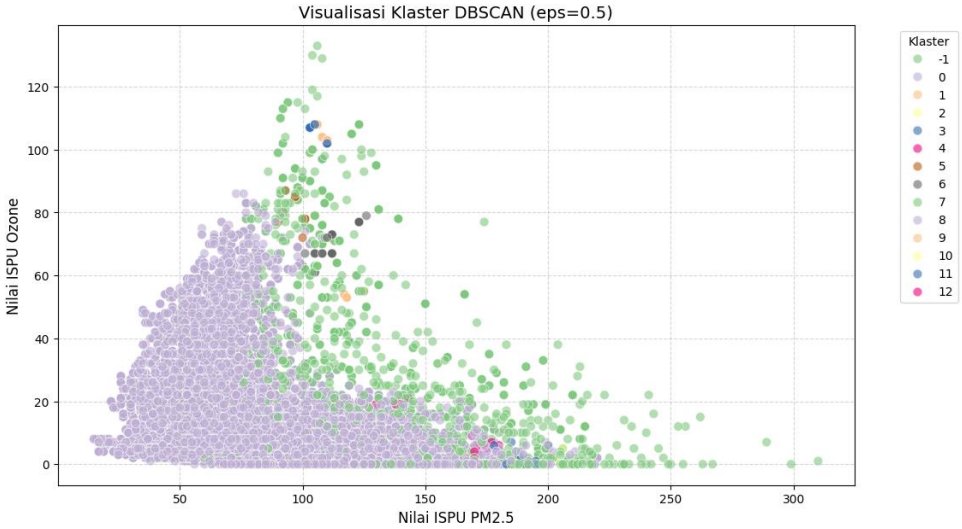
Metrik Evaluasi	Skor
Silhouette Score	0.013473
Davies-Bouldin Index	1.240546

Hasil penilaian validitas algoritma DBSCAN disajikan pada Tabel 4 dengan menggunakan metrik *Silhouette Score* dan *Davies-Bouldin Index*. Nilai *Silhouette Score* sebesar 0,013 mengidentifikasi bahwa struktur klaster yang dihasilkan belum optimal, ditandai dengan pemisahan antar klaster yang lemah serta adanya beberapa data yang berada di area batas klaster.

Nilai Davies-Bouldin Index sebesar 1,2405 mencerminkan bahwa baik pemisahan antar klaster maupun kekompakan internal klaster berada pada tingkat yang kurang baik. DBI yang lebih tinggi menandakan bahwa jarak antar klaster relatif dekat dan kohesi dalam klaster tidak terlalu kuat. Dengan demikian, hasil evaluasi ini mengindikasikan bahwa pemodelan menggunakan DBSCAN pada data ini belum menghasilkan struktur klaster yang optimal.

ISSN: 1978-8126
e-ISSN: 2527-7340

3.1.8. Interpretasi Hasil Clustering DBSCAN



Gambar 5. Grafik Hasil Clustering DBSCAN

Visualisasi hasil clustering DBSCAN dengan parameter ε = 0,5 ditampilkan pada Gambar 5. Pada scatter plot tersebut, sumbu horizontal menunjukkan nilai ISPU $PM_{2.5}$ sedangkan sumbu vertikal menunjukkan ISPU Ozone. Variasi warna masing-masing titik menunjukkan label klaster yang dihasilkan oleh algoritma DBSCAN, di mana label -1 menandakan titik data yang diidentifikasi sebagai *noise* atau *outlier* yang tidak termasuk ke dalam cluster mana pun. Fitur *noise* ini merupakan salah satu karakteristik khas metode DBSCAN yang membedakannya dari algoritma lain seperti K-Means, karena DBSCAN dapat secara eksplisit mempertahankan titik yang tidak memenuhi persyaratan kepadatan sebagai *noise*, bukan memaksanya masuk ke dalam kelompok manapun [21] .
Untuk memberikan dukungan numerik terhadap pola klaster, rata-rata nilai ISPU per klaster hasil DBSCAN disajikan pada Tabel 5 berikut. Tabel ini menyajikan ringkasan statistik deskriptif atas setiap klaster termasuk kelompok *noise*, sehingga pembaca dapat melihat bagaimana nilai polutan rata-rata berbeda antar kelompok hasil clustering.

Tabel 5. Rata-rata Nilai ISPU per Klaster (DBSCAN)

Klaster	PM_{10}	$PM_{2,5}$	SO_2	NO_2	CO	O_3	ISPU Final
Klaster -1	63.2	123.5	48.4	41.2	34	28.5	124.1
Klaster 0	38.7	81.9	39	26.8	13.3	15	82
Klaster 1	56.4	107.7	32.5	17.8	10.1	104.4	108.3
Klaster 2	106	208.6	53.8	70.6	27.7	0	208.6
Klaster 3	55.7	106.1	43.7	16.8	13.7	105.2	108.3
Klaster 4	70.7	138	47.7	57.3	48.6	20.3	138
Klaster 5	50.4	94.7	45.9	18.1	10.7	81.1	94.7
Klaster 6	58.8	111.8	53	12.6	8.8	70	111.8
Klaster 7	104	207.5	60.3	68.9	40	1.7	207.5
Klaster 8	110	218.3	56.6	82.8	24.1	1.8	218.3
Klaster 9	61.9	119.2	47.2	34.2	16.5	53.8	119.2
Klaster 10	102	202.8	62.6	72.5	21.7	1.8	202.8
Klaster 11	95.5	188.6	31.9	46.3	20.8	2.6	188.6
Klaster 12	87	170.5	24.9	33.6	7.6	3.8	170.5

Berdasarkan Tabel 5, istilah *noise* yang diberi label -1 mencerminkan data yang tidak memenuhi syarat kepadatan minimal untuk membentuk klaster, sehingga mewakili *anomali* atau pola polutan udara yang sangat ekstrem yang tidak terkelompok secara homogen bersama titik data lain dalam cluster. Kemampuan

DBSCAN untuk menangkap titik-titik semacam ini tanpa memaksanya masuk ke dalam cluster merupakan bagian dari kekuatan metode ini dalam menangani dataset nyata dengan variasi densitas yang tinggi [22] .

Selain *noise*, berbagai klaster yang terbentuk menunjukkan variasi dominasi polutan tertentu. Misalnya, beberapa klaster memiliki nilai rata-rata $PM_{2.5}$ dan PM_{10} yang tinggi, menunjukkan kelompok kondisi udara dengan konsentrasi partikulat yang lebih tinggi, sementara klaster lain menonjolkan nilai Ozone atau NO_2 sebagai karakter dominan. Berbeda dengan K-Means, DBSCAN secara otomatis menentukan struktur klaster berdasarkan kepadatan data sehingga pola kompleks dapat diidentifikasi secara adaptif. Namun, meskipun mampu mengenali pola kepadatan, tingkat segregasi antar-kelompok yang dihasilkan masih tergolong lemah. Hal ini selaras dengan hasil uji validitas sebelumnya, di mana capaian *Silhouette Score* yang rendah serta angka *Davies-Bouldin Index* yang tinggi mengindikasikan bahwa pemisahan antar-klaster belum terbentuk secara ideal.

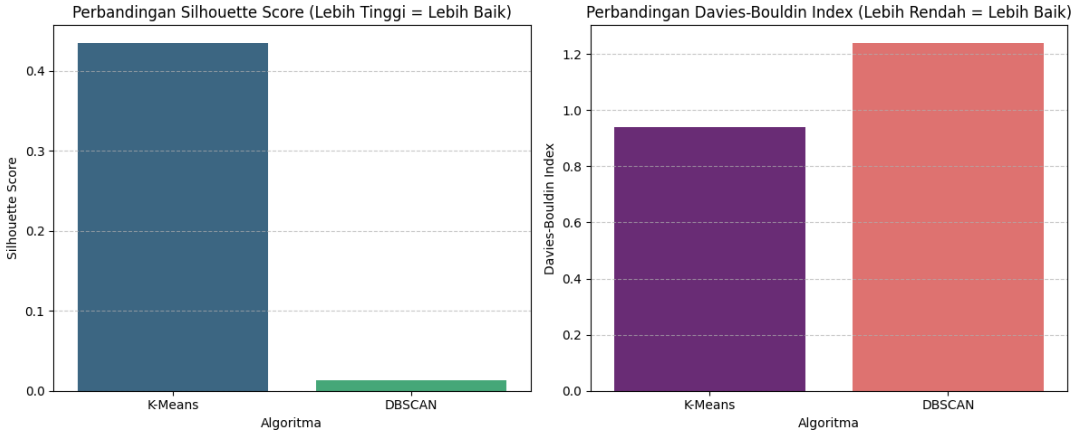
3.1.9. Analisis Perbandingan Performa Algoritma

Perbandingan performa algoritma dilakukan menggunakan matrik evaluasi yang sama untuk memastikan konsistensi analisis. Hasilnya ditampilkan pada Tabel 6.

Tabel 6. Tabel Analisis Perbandingan Performa Algoritma

Metrik Evaluasi	K-Means	DBSCAN
Silhouette Score	0.324464	0.013473
Davies-Bouldin Index	1.142941	1.240546

Tabel 6 menampilkan perbandingan performa algoritma K-Means dan DBSCAN berdasarkan metrik *Silhouette Score* dan *Davies-Bouldin Index* (DBI). Hasil pengujian menyatakan bahwa K-Means menghasilkan *Silhouette Score* sebesar 0,324464, yang lebih tinggi dibandingkan DBSCAN dengan nilai 0,013473. Selain itu, nilai DBI pada K-Means sebesar 1,142941 juga lebih rendah dibandingkan DBSCAN yang mencapai 1,240546. Hasil tersebut mengindikasikan bahwa K-Means dapat membentuk klaster yang lebih terstruktur dengan jarak antar klaster yang lebih jelas, sehingga memberikan performa yang lebih baik pada dataset yang digunakan dalam penelitian ini.



Gambar 6. Visualisasi Perbandigan K-Means dan DBSCAN

4. Kesimpulan

Penelitian ini membandingkan kinerja algoritma *K-Means* dan *DBSCAN* dalam mengelompokkan kualitas udara berdasarkan *Indeks Standar Pencemar Udara (ISPU)* di wilayah Jabodetabek. Hasil analisis menunjukkan bahwa penerapan *K-Means* dengan jumlah klaster optimal memberikan performa yang lebih baik dibandingkan *DBSCAN*. Berdasarkan evaluasi menggunakan metrik *Silhouette Score* dan *Davies-Bouldin Index (DBI)*, algoritma *K-Means* menghasilkan klaster dengan tingkat kerapatan yang lebih tinggi dan pemisahan antar klaster yang lebih jelas, dengan skor *Silhouette* sebesar 0,32446 dan DBI 1,142941. Sebaliknya, DBSCAN menampilkan performa yang kurang memadai, ditunjukkan oleh rendahnya nilai

Silhouette Score (0,013473) bersamaan dengan tingginya nilai DBI (1,240546), menandakan kohesi klaster yang lemah dan pemisahan antar klaster yang tidak optimal. Dengan demikian, K-Means menjadi pilihan yang lebih tepat untuk analisis kualitas udara di Jabodetabek, walaupun pengaturan parameter pada DBSCAN masih berpotensi ditingkatkan guna menghasilkan hasil yang lebih optimal.

Daftar Pustaka

[1] "Negara Paling Berpolusi di Dunia 2024 - Rangking PM2.5 | IQAir." Accessed: Mar. 17, 2026. [Online]. Available: https://www.iqair.com/id/world-most-polluted-countries

[2] Demographia, "Demographia World Urban Areas 19th Annual Edition (Built Up Urban Areas or World Agglomerations)," Demographia, Aug. 2023. [Online]. Available: https://www.demographia.com/db-worldua.pdf

[3] "Penentuan Model Algoritma Klasifikasi Terbaik Untuk Klasifikasi Kualitas Udara Di Jakarta 2023 | Jurnal Informatika dan Teknik Elektro Terapan." Accessed: Mar. 17, 2026. [Online]. Available: https://journal.eng.unila.ac.id/index.php/jitet/article/view/5664

[4] A. Amalia, A. Zaidiah, and I. N. Isnainiyah, "Prediksi Kualitas Udara Menggunakan Algoritma K-Nearest Neighbor," *JIPI J. Ilm. Penelit. Dan Pembelajaran Inform.*, vol. 7, no. 2, pp. 496–507, May 2022, doi: 10.29100/jipi.v7i2.2843.

[5] A. R. Kannajmi and D. Saputra, "Penentuan Model Algoritma Klasifikasi Terbaik Untuk Klasifikasi Kualitas Udara Di Jakarta 2023," *J. Inform. Dan Tek. Elektro Terap.*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5664.

[6] N. A. R. Rahmadenti, "Analisis Pola Dan Tren Kualitas Udara Dki Jakarta 2022–2025 Menggunakan K-Means Clustering," *J. Inform. Dan Tek. Elektro Terap.*, vol. 13, no. 3S1, Oct. 2025, doi: 10.23960/jitet.v13i3S1.8141.

[7] Mahendrasyah, A. Diana, Rusdah, and D. Mahdiana, "Penerapan Algoritma K-Means Untuk Klasterisasi Indeks Standar Pencemaran Udara," *Teknologi*, vol. 14, no. 2, pp. 146–156, Dec. 2024, doi: 10.26594/teknologi.v14i2.4088.

[8] I. G. S. Rahayuda and N. P. L. Santiari, "Sistem Monitoring Indek Kualitas Udara Berbasis Open Meteo Air Quality API | JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)", [Online]. Available: https://doi.org/10.24843/JLK.2023.v12.i03.p22

[9] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Glob. Transit. Proc.*, vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/j.gltp.2022.04.020.

[10] B. F. Azevedo, A. M. A. C. Rocha, and A. I. Pereira, "A multi-objective clustering approach based on different clustering measures combinations," *Comput. Appl. Math.*, vol. 44, no. 2, p. 59, Dec. 2024, doi: 10.1007/s40314-024-03004-x.

[11] N. A. Maori and E. Evanita, "Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering," *Simetris J. Tek. Mesin Elektro Dan Ilmu Komput.*, vol. 14, no. 2, pp. 277–288, Nov. 2023, doi: 10.24176/simet.v14i2.9630.

[12] Erlin Windia Ambarsari, Dedin Fathudin, and Gravita Alfiani, "Utilizing K-Means Clustering to Understanding Audience Interest in SEO-Optimized Media Content," *J. Comput. Inform. Res.*, vol. 3, no. 2, pp. 208–214, Mar. 2024, doi: 10.47065/comforch.v3i2.1207.

[13] K. D. Pertiwi, I. P. Lestari, and A. Afandi, "Analisis Risiko Kesehatan Lingkungan Pajanan Debu PM10 dan PM2.5 pada Relawan Lalu Lintas di Jalan Diponegoro Ungaran," *Health J. Ilm. Kesehat.*, vol. 6, no. 2, pp. 85–91, Jul. 2024.

[14] M. Ester, H.-P. Kriegel, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise".

[15] A. Chandra, "Perbandingan Algoritma Clustering K-Means, Gaussian Mixture Model, dan DBSCAN pada Data Indeks Standar Pencemar Udara (ISPU) Di Provinsi DKI Jakarta," *J. Komput. Dan Inform.*, vol. 19, no. 1, pp. 1–8, 2024.

[16] H. Rhomadhona, W. Kusrini, W. Aprianti, and J. Permadi, "Implementation of K-Means Clustering for Social Assistance Recipients with Silhouette Score Evaluation," *Brill. Res. Artif. Intell.*, vol. 5, no. 1, pp. 136–143, May 2025, doi: 10.47709/brilliance.v5i1.5900

[17] L. Yin, H. Hu, K. Li, G. Zheng, Y. Qu, and H. Chen, "Improvement of DBSCAN Algorithm Based on K-Dist Graph for Adaptive Determining Parameters," *Electronics*, vol. 12, no. 15, p. 3213, Jul. 2023, doi: 10.3390/electronics12153213.

ISSN: 1978-8126
e-ISSN: 2527-7340

[18] Y. Sopyan, A. D. Lesmana, and C. Juliane, "Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan," *Build. Inform. Technol. Sci. BITS*, vol. 4, no. 3, Dec. 2022, doi: 10.47065/bits.v4i3.2697.

[19] Y. Wijaya, D. Kurniadi, E. Setyanyo, T. Sanur, D. Rusmana, and R. Rahim, "Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities," *TEM J.*, vol. 10, no. 3, pp. 1099–1103, 2021.

[20] M. A. F. Razan, N. J. Alifah, Q. A'yuni, M. Wati, and Haviluddin -, "Application of K-Means Clustering Algorithm for Air Quality Pattern Analysis in Jakarta," *J. Teknol. DAN ILMU Komput. PRIMA JUTIKOMP*, vol. 8, no. 1, pp. 64–80, Apr. 2025, doi: 10.34012/jutikomp.v8i1.7028.

[21] A. Ramadhan, F. Achmad, I. Zulkarnain, and M. Aritsugi, "Evaluation of K-Means, DBSCAN, and Hierarchical Clustering for Strategic Segmentation of Tourism SMEs in Rembang, Indonesia," *J. Tek. Inform. Jutif*, vol. 6, no. 3, pp. 1605–1630, Jul. 2025, doi: 10.52436/1.jutif.2025.6.3.4602.

[22] D. Armiady, "Analisis Metode DBSCAN (Density-Based Spatial Clustering of Application with Noise) dalam Mendeteksi Data Outlier," *JURIKOM J. Ris. Komput.*, vol. 9, no. 6, p. 2158, Dec. 2022, doi: 10.30865/jurikom.v9i6.5080.