

CLUSTERING POLUSI UDARA BERDASARKAN FACTOR METEOROLOGI BERBASIS PRINCIPAL COMPONENT ANALYSIS

Maimuna Syariefatussholehah, Agung Teguh Wibowo Almais*, Ririen Kusumawati,
M. Amin Hariyadi, M. Faisal, Ahmad Zarkoni, Mohammad Mamba'ul Ulum
Magister Informatika, Universitas Islam Negeri Maulana Malik Ibrahim Malang
Jalan Gajayana No. 50, Kelurahan Jatimulyo, Kecamatan Lowokwaru, Kota Malang, Jawa Timur, Indonesia
agung.twa@ti.uin-malang.ac.id

ABSTRAK

Polusi udara di wilayah urban, khususnya Kota Malang, telah mencapai tingkat yang mengkhawatirkan dengan konsentrasi $PM_{2.5}$ yang sering kali melebihi ambang batas aman WHO. Kompleksitas interaksi antara polutan dan faktor meteorologi lokal menuntut adanya pendekatan analisis yang lebih cerdas dan holistik dibandingkan metode konvensional. Penelitian ini bertujuan untuk mengidentifikasi pola pengelompokan polusi udara di Kota Malang melalui pengembangan model pelabelan otomatis secara *unsupervised*. Metode yang diimplementasikan adalah integrasi *Principal Component Analysis* (PCA) dan teknik *clustering* untuk mereduksi dimensi data meteorologi multidimensi menjadi komponen utama yang esensial. Data historis $PM_{2.5}$ dan enam variabel cuaca (suhu, kelembapan, curah hujan, penyinaran matahari, dan kecepatan angin) periode 2024 dari stasiun BMKG diproses melalui tahapan normalisasi *StandardScaler*, ekstraksi nilai eigen, hingga penentuan klaster berdasarkan ambang batas PC1. Hasil penelitian menunjukkan bahwa penggunaan dua komponen utama (PC1 dan PC2) mampu menjelaskan 66,76% variansi data asli dan berhasil membagi kondisi kualitas udara ke dalam dua kategori, yaitu "Kondisi Normal" dan "Kondisi Ekstrem". Validasi menggunakan *Silhouette Score* sebesar 0,87 menunjukkan bahwa klaster yang terbentuk memiliki kepadatan dan pemisahan yang sangat baik. Model ini terbukti efektif sebagai instrumen pelabelan otomatis untuk mendukung sistem mitigasi risiko kesehatan masyarakat dan kebijakan pengendalian kualitas udara yang lebih terukur.

Kata kunci : Polusi Udara, $PM_{2.5}$, PCA-Clustering, *Unsupervised Labeling*, Meteorologi, Kota Malang.

1. PENDAHULUAN

Polusi udara telah menjadi krisis global yang mengancam kesehatan manusia dan keseimbangan ekosistem secara masif [1] Organisasi Kesehatan Dunia (WHO) melaporkan bahwa 99% populasi dunia saat ini terpapar udara dengan konsentrasi $PM_{2.5}$ dan PM_{10} yang melampaui ambang batas aman, di mana paparan kronis ini berkontribusi langsung pada 7 juta kematian prematur setiap tahunnya akibat penyakit pernapasan dan kardiovaskular [2].

Di Indonesia, wilayah urban seperti Kota Malang menghadapi tantangan serius dengan rata-rata konsentrasi $PM_{2.5}$ mencapai $45 \mu g/m^3$, angka yang tercatat tujuh kali lipat lebih tinggi dari standar keselamatan yang direkomendasikan (Badan Pusat Statistik Kota Malang, 2023)[3].

Pertumbuhan kendaraan bermotor sebesar 9% per tahun menyumbang 58% emisi harian, sementara sektor industri, kebakaran lahan, dan fenomena alam seperti inversi termal memperburuk kondisi kualitas udara di lapisan atmosfer bawah [4].

Dinamika polusi udara ini sangat dipengaruhi oleh interaksi parameter meteorologi yang kompleks, seperti kecepatan angin, curah hujan, suhu, dan kelembapan [5]

Studi oleh Shafi & Waheed (2020) [6] menunjukkan bahwa konsentrasi partikulat dapat melonjak drastis hingga $242 \mu g/m^3$ dalam waktu singkat akibat perubahan kondisi atmosfer, sementara struktur morfologi kota dan kanopi urban terbukti

dapat menjebak polutan sehingga menciptakan *hotspot* polusi [7].

Pemantauan yang lebih cerdas sangat diperlukan melalui pendekatan analitik yang mampu mereduksi dimensi data multidimensi untuk menonjolkan fitur meteorologi paling berpengaruh [8][9].

Metode *Principal Component Analysis* (PCA) terbukti efektif dalam menyederhanakan kompleksitas variabel cuaca tanpa kehilangan informasi esensial [10].

Penelitian terbaru oleh Syahreni dkk. (2025) [11] memperkuat efektivitas integrasi PCA-Clustering dalam pembentukan label otomatis (*unsupervised labeling*) pada data klimatologi dengan hasil pemisahan klaster yang stabil dan valid. Namun, penerapan metode ini untuk segmentasi $PM_{2.5}$ di wilayah urban tropis Indonesia masih terbatas, sehingga pengembangan model ini menjadi krusial untuk mendukung sistem pemantauan yang lebih responsif [12][13].

Berdasarkan latar belakang tersebut, Penelitian ini bertujuan untuk mengkaji sejauh mana pengaruh parameter meteorologi seperti suhu, kelembapan, curah hujan, penyinaran matahari, dan kecepatan angin, secara simultan terhadap distribusi $PM_{2.5}$ di Kota Malang. Selain itu, penelitian ini juga mengevaluasi efektivitas integrasi metode PCA-Clustering dalam mengidentifikasi pola pengelompokan polusi udara berdasarkan interaksi kompleks antara polutan dan kondisi cuaca harian, serta menelaah implikasi hasil segmentasi tersebut

sebagai dasar pelabelan otomatis untuk mendukung kebijakan pengendalian kualitas udara dan perlindungan kesehatan masyarakat. Penelitian ini bertujuan untuk menganalisis korelasi serta pengaruh parameter meteorologi terhadap variasi polusi udara harian di Kota Malang. Selain itu, penelitian ini juga mengembangkan model PCA-Clustering yang mampu mengelompokkan kondisi kualitas udara ke dalam klaster “Normal” dan “Ekstrem” secara objektif melalui pendekatan *unsupervised* [10].

Hasil penelitian ini diharapkan dapat memberikan landasan ilmiah berbasis data bagi pemangku kebijakan dalam merancang sistem mitigasi risiko kesehatan masyarakat yang lebih terukur. Penelitian ini dilakukan melalui beberapa tahapan teknis utama. Tahap awal meliputi normalisasi data dan reduksi dimensi menggunakan PCA pada dataset historis polusi udara PM_{2.5} dan parameter meteorologi harian periode 2024 yang diperoleh dari stasiun pemantauan BMKG untuk meminimalkan redundansi variable [10][11].

Selanjutnya, dikembangkan model PCA-Clustering untuk memisahkan data berdasarkan nilai eigenvalue dominan (PC1 dan PC2) yang merepresentasikan variansi terbesar [10].

Tahap akhir mencakup pelabelan otomatis berdasarkan ambang batas koordinat PC1 yang divalidasi menggunakan *Silhouette Score* guna memastikan kualitas klaster yang optimal [10][11].

Hasil penelitian kemudian dihubungkan dengan standar kualitas udara BMKG dan Indeks Standar Pencemar Udara (ISPU) untuk mendukung tujuan SDGs poin ke-11 dan ke-13 (The Global Goals, 2024).

2. TINJAUAN PUSTAKA

2.1. Analisis Klaster Polusi Udara

Analisis kualitas udara di wilayah perkotaan telah berkembang pesat dengan memanfaatkan pendekatan *machine learning*, khususnya dalam mengidentifikasi pola polusi dan faktor-faktor yang memengaruhinya. Metode klasterisasi seperti K-Means banyak digunakan untuk mengelompokkan data polusi udara berdasarkan kemiripan karakteristik, serta efektif dalam mendeteksi perubahan konsentrasi polutan secara temporal [6][5].

Selain itu, integrasi antara data polusi dan parameter meteorologi juga menjadi fokus utama dalam berbagai penelitian. Variabel seperti suhu, kelembaban, kecepatan angin, dan curah hujan terbukti memiliki pengaruh signifikan terhadap variasi konsentrasi PM_{2.5} [14][8].

Di sisi lain, faktor lingkungan seperti tata guna lahan juga berperan dalam memediasi distribusi polutan, sehingga memperkuat kompleksitas analisis kualitas udara di wilayah urban [9].

Seiring dengan meningkatnya kompleksitas data, penggunaan teknik reduksi dimensi seperti Principal Component Analysis (PCA) mulai banyak diterapkan untuk menyederhanakan variabel tanpa menghilangkan informasi utama. Integrasi PCA

dengan metode clustering terbukti mampu meningkatkan stabilitas pengelompokan data multidimensi serta menghasilkan klasifikasi yang lebih objektif [10].

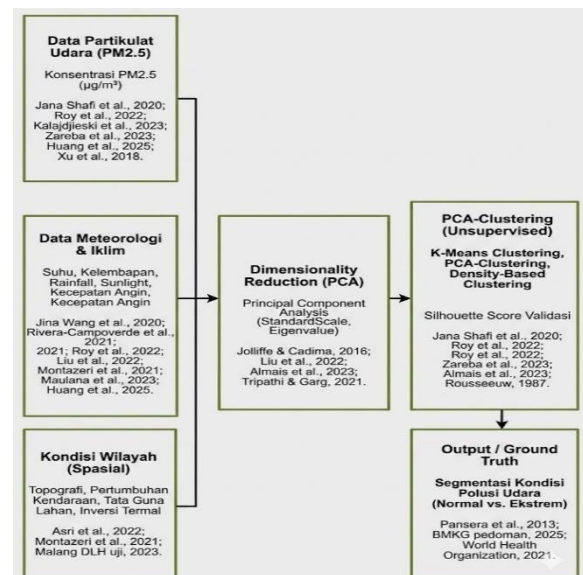
Dengan demikian, pendekatan berbasis PCA-Clustering menjadi salah satu metode yang potensial dalam menganalisis dan mengelompokkan kualitas udara secara efektif, khususnya dalam menghadapi kompleksitas data meteorologi dan polusi udara di wilayah perkotaan.

2.2. Kerangka Teori

Penelitian ini menggunakan pendekatan *Principal Component Analysis* (PCA) untuk mereduksi dimensi data polusi udara dan parameter meteorologi, sehingga pola antar variabel dapat diidentifikasi secara lebih efisien. Data masukan berupa parameter seperti PM_{2.5}, suhu, kelembaban, curah hujan, penyinaran matahari, dan kecepatan angin diolah untuk menghasilkan komponen utama (*principal components*) yang merepresentasikan variasi dominan dalam data.

Hasil transformasi tersebut kemudian digunakan sebagai dasar dalam proses *clustering* untuk mengelompokkan kondisi polusi udara berdasarkan karakteristik cuaca tertentu. Pendekatan ini bertujuan untuk mendukung pemahaman pola kualitas udara di wilayah urban secara lebih komprehensif.

Berdasarkan alur tersebut, kerangka teori penelitian ini disusun untuk menggambarkan hubungan antara data masukan, proses analisis, dan hasil pengelompokan. Kerangka teori tersebut disajikan pada Gambar 2.1.



Gambar 1 Kerangka Teori

Kerangka teori pada Gambar 2.1 menunjukkan bahwa proses analisis diawali dengan input data polusi udara dan parameter meteorologi. Data tersebut kemudian diproses melalui tahap pra-proses dan reduksi dimensi untuk memperoleh komponen utama

yang merepresentasikan variansi terbesar dalam data. Selanjutnya, komponen utama digunakan sebagai input dalam proses *clustering* untuk mengelompokkan data ke dalam beberapa kluster berdasarkan kemiripan karakteristik. Output dari proses ini berupa kluster kualitas udara yang dapat digunakan untuk memahami pola distribusi polusi serta mendukung pengambilan keputusan dalam pengelolaan lingkungan perkotaan.

2.3. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) merupakan teknik statistik multivariat yang digunakan untuk mereduksi dimensi data dengan mengidentifikasi komponen utama yang paling signifikan dalam menjelaskan variabilitas dataset polusi udara. Proses ini sangat efektif untuk menyederhanakan kompleksitas data berdimensi tinggi tanpa kehilangan informasi yang relevan. Implementasi PCA dalam penelitian ini melibatkan beberapa tahapan teknis sebagai berikut:

2.4. Normalisasi Data

Normalisasi dilakukan agar seluruh fitur meteorologi memiliki skala yang seragam sehingga kontribusi setiap variabel menjadi seimbang dalam pembentukan komponen utama. Penelitian ini menggunakan skala standar (*StandardScaler*) dengan persamaan (1) berikut:

$$X_{standar} = \frac{x - \mu}{\sigma} \quad (1)$$

Keterangan: μ adalah nilai rata-rata (*mean*) dan σ adalah standar deviasi dari himpunan data. Standar deviasi tersebut dihitung menggunakan persamaan (2) berikut:

$$\bar{X} = \sum_{i=1}^n Xi / n \quad (2)$$

2.5. Penentuan Nilai Eigen dan Rasio Varians

Tahap ini bertujuan untuk menentukan besarnya penyebaran data yang mampu ditangkap oleh setiap komponen. Rasio varians dihitung menggunakan persamaan (3) untuk mengetahui seberapa besar informasi yang terwakili pada setiap dimensi baru:

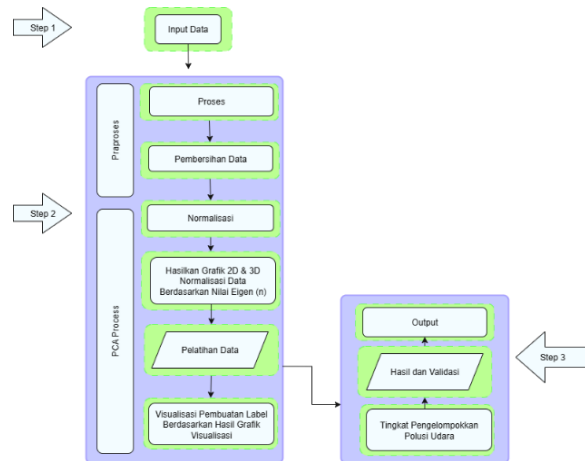
$$S^2 = \sum_{i=1}^n (Xi - \bar{X})^2 / (n - 1) \quad (3)$$

2.6. Pemilihan Komponen Utama (PC)

Komponen utama dipilih berdasarkan *eigenvalue* terbesar yang mencerminkan arah dengan variansi data tertinggi. Dalam penelitian ini, PC1 dan PC2 digunakan sebagai koordinat baru guna merepresentasikan variasi dominan antara konsentrasi polutan dan parameter cuaca harian. Hasil reduksi ini kemudian divisualisasikan dalam bentuk grafik guna mengidentifikasi pola pengelompokan alami sebelum dilanjutkan ke tahap klusterisasi.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan strategi eksploratif untuk mengelompokkan polusi udara di Kota Malang berdasarkan faktor meteorologi. Tahapan penelitian dirancang secara sistematis yang meliputi pengumpulan data, praproses, analisis reduksi dimensi menggunakan PCA, hingga validasi hasil kluster. Alur kerja sistem secara lengkap disajikan pada Gambar 2.



Gambar 2. Desain Sistem

Dari Gambar 2, dapat dilihat bahwa proses diawali dengan pengumpulan data partikulat PM_{2.5} dan parameter cuaca (suhu, kelembapan, curah hujan, penyinaran matahari, dan kecepatan angin) periode tahun 2024 dari stasiun pemantauan BMKG. Tahap selanjutnya adalah praproses data yang meliputi pembersihan data untuk menangani nilai yang hilang (*missing values*) serta normalisasi data menggunakan metode *StandardScaler* agar seluruh variabel berada pada skala yang seragam.

4. HASIL DAN PEMBAHASAN

Bagian ini menjelaskan tahapan implementasi metode PCA-Clustering untuk segmentasi polusi udara di Kota Malang, mulai dari normalisasi data hingga validasi hasil pengelompokan.

4.1. Normalisasi Data

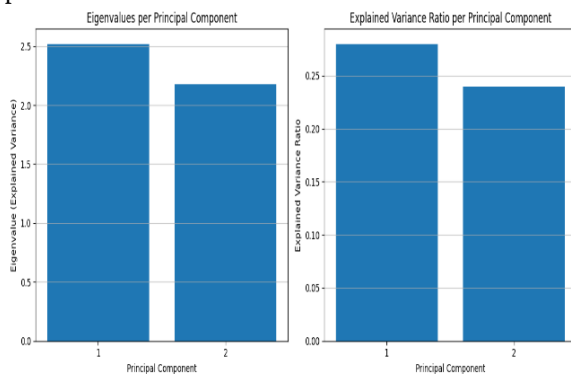
Proses standarisasi data dilakukan menggunakan metode *StandardScaler* untuk menyamakan skala fitur meteorologi yang beragam (PM_{2.5}, suhu, kelembapan, curah hujan, penyinaran matahari, dan kecepatan angin). Langkah ini krusial agar variabel dengan rentang nilai besar tidak mendominasi pembentukan komponen utama dalam perhitungan varians. Hasil normalisasi mentransformasikan setiap fitur sehingga memiliki nilai rata-rata (*mean*) nol dan standar deviasi satu sebelum masuk ke tahap reduksi dimensi. Ampel hasil proses normalisasi data disajikan pada Tabel 1.

Tabel 1. Hasil Normalisasi Data

Partikulat matter_2.5	temp_avg	humidity_avg	rainfall	sunlight	wind_avg
-1.153998388	0.035726014	0.248900674	0.20996807	-0.123126018	0.486312401
0.07520506	0.035726014	0.170899435	-0.416889083	-0.13641731	-1.152640995
0.894674026	-0.037381053	0.326901913	-0.416889083	-0.118695588	-1.316536335
0.940200079	-0.068712654	0.170899435	-0.416889083	-0.202873766	-0.988745656
0.257309275	-0.162707455	0.638906868	-0.416889083	-0.180721614	0.158521722

4.2. Penentuan Nilai Eigen dan Rasio Varians

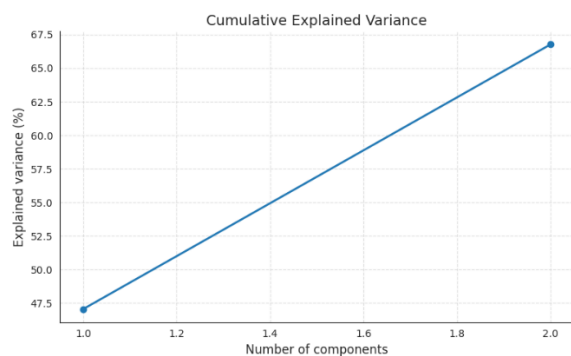
Penentuan jumlah *Principal Component* (PC) didasarkan pada analisis nilai *eigen* dan rasio varians. Parameter ini menggambarkan proporsi informasi yang berhasil ditangkap oleh setiap komponen. Berdasarkan hasil ekstraksi, PC1 memiliki nilai *eigen* sebesar 2,61 dengan rasio varians 43,57%, sedangkan PC2 memiliki nilai *eigen* 1,39 dengan rasio varians 23,19%. Secara kumulatif, penggunaan dua komponen utama ini mampu menjelaskan 66,76% variansi data asli, sehingga cukup representatif untuk digunakan dalam tahap analisis selanjutnya. Hubungan jumlah komponen utama dengan variansi kumulatif disajikan pada Gambar 3



Gambar 3. Nilai Eigen dan Rasio Varians

4.3. Proses Pengelompokan Data (Clustering)

Penentuan jumlah *principal component* (PC) dilakukan dengan mengevaluasi variansi kumulatif. Penggunaan PC1 dan PC2 mampu menjelaskan 66,76% variansi data asli, sehingga cukup representatif untuk analisis selanjutnya. Grafik kumulatif rasio varians disajikan pada Gambar 4.

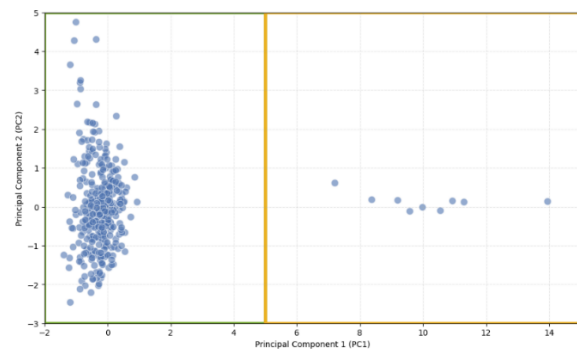


Gambar 4. Grafik Kumulatif Rasio Varians

Penentuan jumlah *Principal Component* (PC) dilakukan dengan mengevaluasi *variansi* kumulatif

untuk memastikan representasi data yang optimal. Berdasarkan Gambar 3, komponen utama pertama (PC1) mampu menjelaskan 47,06% variasi, sementara penambahan komponen utama kedua (PC2) meningkatkan total variasi kumulatif menjadi 66,76%. Hal ini menunjukkan bahwa dua komponen utama sudah memadai untuk digunakan sebagai koordinat baru dalam proses klasterisasi tanpa kehilangan informasi esensial.

Selanjutnya, proses pengelompokan dilakukan menggunakan metode *thresholding* pada sumbu PC1 yang mewakili dimensi dengan variansi terbesar. Distribusi titik data hasil transformasi PCA tersebut dipisahkan ke dalam dua kelompok homogen berdasarkan rentang nilai koordinatnya sebagaimana disajikan pada Gambar 5.



Gambar 5 Coordinate Points Data Distribution

Berdasarkan Gambar 4, kelompok Kondisi Normal diidentifikasi pada rentang nilai PC1 < 5,0 (kotak hijau) dengan sebaran titik yang padat, mencerminkan karakteristik polusi udara harian yang stabil. Sementara itu, kelompok Kondisi Ekstrem (kotak oranye) mencakup titik data dengan nilai PC1 > 5,0 yang tersebar secara jarang dan menjauh dari pusat massa data, mengindikasikan adanya anomali kualitas udara yang dipengaruhi secara signifikan oleh dinamika faktor meteorologi tertentu. Pemisahan yang kontras pada ruang *komponen* utama ini memvalidasi bahwa model mampu membentuk segmentasi kualitas udara secara otomatis melalui pendekatan *unsupervised learning*.

4.4. Validasi Hasil Clustering

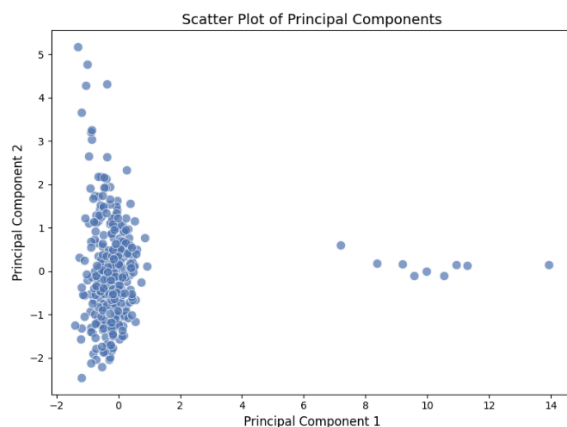
Validasi dilakukan untuk mengukur kualitas dan konsistensi pengelompokan yang dihasilkan oleh metode PCA-Clustering. Pengujian ini mencakup evaluasi kuantitatif menggunakan metrik statistik serta validasi visual melalui representasi dimensi data.

4.5. Evaluasi Kuantitatif dengan Silhouette Score

Evaluasi kualitas kluster dilakukan dengan menggunakan *Silhouette Score* untuk mengukur seberapa baik setiap titik data ditempatkan ke dalam kelompoknya. Berdasarkan hasil pengujian, diperoleh nilai *Silhouette Score* sebesar 0,87. Nilai yang mendekati 1 ini memberikan bukti statistik yang kuat bahwa kluster memiliki tingkat *kepadatan* intra-kluster yang tinggi dan jarak pemisahan antar-kluster yang jelas. Hasil ini memvalidasi secara matematis bahwa pembagian kelompok berdasarkan *threshold* PC1 pada angka 5,0 telah menghasilkan pemisahan yang valid dan konsisten dengan karakteristik data asli.

4.6. Visualisasi Hasil Clustering 2D dan 3D

Untuk memperkuat hasil analisis, sebaran data hasil transformasi PCA divisualisasikan ke dalam bentuk grafik 2 dimensi (2D) dan 3 dimensi (3D). Visualisasi ini bertujuan untuk mempresentasikan rentang nilai hasil reduksi secara grafis sehingga pola pengelompokan alami data polusi udara dan meteorologi dapat diidentifikasi dengan lebih mudah.

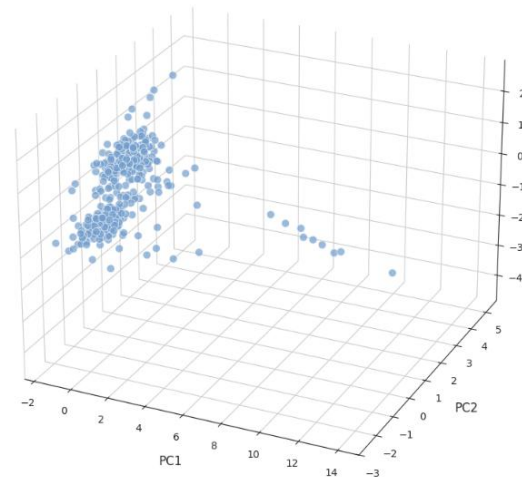


Gambar 6. Visualisasi 2D

Berdasarkan Gambar 6, terlihat sebaran titik data terkonsentrasi secara padat di area nilai negatif sumbu PC1 yang merepresentasikan Kondisi Normal, serta beberapa titik yang tersebar secara jauh di area nilai positif yang merepresentasikan Kondisi Ekstrem. Meskipun visualisasi 2D telah memberikan informasi yang jelas mengenai pemisahan kelompok, perspektif tambahan diperlukan untuk melihat kedalaman variansi data yang mungkin tidak tertangkap pada dua sumbu saja. Oleh karena itu, penelitian ini juga menyertakan visualisasi 3 dimensi (3D) dengan menambahkan komponen utama ketiga (PC3) guna memberikan representasi ruang data yang lebih komprehensif dan meminimalkan risiko kehilangan informasi akibat reduksi dimensi.

Dalam upaya memberikan pemahaman yang lebih komprehensif mengenai struktur data, penelitian ini juga menghadirkan visualisasi *dalam* dimensi yang lebih tinggi. Meskipun grafik 2D telah mampu memperlihatkan pemisahan kluster di sepanjang sumbu PC1, representasi 3 dimensi (3D) diperlukan

untuk mengamati kedalaman variansi data yang melibatkan komponen utama ketiga (PC3). Visualisasi 3D ini memungkinkan identifikasi sebaran titik data dari berbagai sudut pandang ruang, sehingga *meminimalkan* risiko kehilangan informasi penting yang terjadi pada saat reduksi dimensi. Sebaran data polusi udara dalam ruang koordinat PC1, PC2, dan PC3 disajikan pada Gambar 7.



Gambar 7. Visualisasi 3D

Berdasarkan Gambar 6, terlihat bahwa penambahan sumbu PC3 mempertegas jarak antar-kelompok data yang sebelumnya terlihat berhimpitan pada proyeksi 2D. Kelompok Kondisi Normal membentuk massa data yang padat dan terkonsentrasi di tengah, sementara kelompok Kondisi Ekstrem membentuk lintasan titik yang menjauh secara linier mengikuti arah komponen utama pertama. Pola sebaran yang konsisten pada ruang 3D ini memberikan bukti visual tambahan bahwa metode *unsupervised PCA-Clustering* telah *berhasil* mengidentifikasi struktur alami dari interaksi faktor meteorologi dan polusi udara di Kota Malang secara akurat.

5. KESIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa integrasi metode *Principal Component Analysis* (PCA) dan teknik *clustering* terbukti efektif dalam melakukan segmentasi pola polusi udara di Kota Malang secara *unsupervised*. Melalui reduksi dimensi, enam variabel meteorologi berhasil disederhanakan menjadi dua komponen utama (PC1 dan PC2) yang mampu menjelaskan 66,76% variansi data asli tanpa kehilangan informasi esensial. Proses pengelompokan berdasarkan ambang batas PC1 berhasil memisahkan data ke dalam dua kategori utama, yaitu kondisi normal yang menunjukkan stabilitas kualitas udara harian dan kondisi ekstrem yang merepresentasikan anomali polusi akibat pengaruh faktor meteorologi tertentu. Validasi menggunakan *Silhouette Score* sebesar 0,87 serta visualisasi 3D mempertegas bahwa

klaster yang terbentuk memiliki kepadatan dan pemisahan yang sangat baik, sehingga layak digunakan sebagai dasar pelabelan otomatis. Untuk pengembangan selanjutnya, disarankan agar penelitian dapat memperluas cakupan data spasial dengan melibatkan lebih banyak stasiun pemantauan di Jawa Timur serta mengintegrasikan model prediksi berbasis *machine learning* untuk analisis pola polusi udara yang lebih dinamis di masa depan.

DAFTAR PUSTAKA

- [1] H. Xu *et al.*, "Identification of the pollution sources and hidden clustering patterns for potentially toxic elements in typical peri-urban agricultural soils in southern China," *Environ. Pollut.*, vol. 370, no. January, p. 125904, 2025, doi: 10.1016/j.envpol.2025.125904.
- [2] World Health Organization, "Ambient (outdoor) air pollution," World Health Organization.
- [3] Badan Pusat Statistik Kota Malang, "Pengamatan Kelembaban Udara di Kota Malang (Persen (%)), 2023," BPS Kota Malang. [Online]. Available: <https://malangkota.bps.go.id>
- [4] L. Nor Asri, K. Eka Sari, and Christia Meidiana, "Emisi CO Kendaraan Bermotor Pada Ruas Jalan Dengan Tingkat Pelayanan Rendah Di Kota Malang," *Plan. Urban Reg. Environ.*, vol. 11, no. 1, pp. 31–38, 2022.
- [5] M. Roy, S. Balachandran, and C. Brokamp, "Clustering and Regression-Based Analysis of PM2.5 Sensitivity to Meteorology in Cincinnati, Ohio," *Atmosphere (Basel)*, vol. 13, no. 4, pp. 1–20, 2022, doi: 10.3390/atmos13040545.
- [6] J. Shafi and A. Waheed, "K-Means Clustering Analysing Abrupt Changes in Air Quality," *Proc. 4th Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2020*, pp. 26–30, 2020, doi: 10.1109/ICECA49313.2020.9297493.
- [7] Y. Huang *et al.*, "The Impact of Meteorological Factors and Canopy Structure on PM2.5 Dynamics Under Different Urban Functional Zones in a Subtropical City," *Forests*, vol. 16, no. 3, 2025, doi: 10.3390/f16030479.
- [8] W. Liu, F. Chen, and Y. Chen, "PM2.5 Concentration Prediction Based on Pollutant Pattern Recognition Using PCA-clustering Method and CS Algorithm Optimized SVR," *Nat. Environ. Pollut. Technol.*, vol. 21, no. 1, pp. 393–403, 2022, doi: 10.46488/NEPT.2022.v21i01.047.
- [9] A. Montazeri, A. J. Lilienthal, and J. D. Albertson, "A spatial land use clustering framework for investigating the role of land use in mediating the effect of meteorology on urban air quality," *Atmos. Environ. X*, vol. 12, no. August, p. 100126, 2021, doi: 10.1016/j.aeaoa.2021.100126.
- [10] A. T. W. Almais *et al.*, "Principal Component Analysis-Based Data Clustering for Labeling of Level Damage Sector in Post-Natural Disasters," *IEEE Access*, vol. 11, no. May, pp. 74590–74601, 2023, doi: 10.1109/ACCESS.2023.3275852.
- [11] S. Syahreni and A. Zarkoni, "Principal Component Analysis-Based Climatology Data for Rainfall Clustering".
- [12] J. Kalajdzieski, K. Trivodaliev, G. Mirceva, S. Kalajdziski, and S. Gievska, "A Complete Air Pollution Monitoring and Prediction Framework," *IEEE Access*, vol. 11, no. January, pp. 88730–88744, 2023, doi: 10.1109/ACCESS.2023.3251346.
- [13] M. Zareba, H. Dlugosz, T. Danek, and E. Weglinska, "Big-Data-Driven Machine Learning for Enhancing Spatiotemporal Air Pollution Pattern Analysis," *Atmosphere (Basel)*, vol. 14, no. 4, 2023, doi: 10.3390/atmos14040760.
- [14] H. Zhou *et al.*, "Characteristics of air pollution and their relationship with meteorological parameters: Northern versus southern cities of China," *Atmosphere (Basel)*, vol. 11, no. 3, pp. 1–15, 2020, doi: 10.3390/atmos11030253.