

Sistem Prediksi Kualitas Udara DKI Jakarta

1st Adhyaksa Waruna Putro
DIKE FMIPA UGM
Universitas Gadjah Mada
Yogyakarta, Indonesia

adhyaksawarunaputro@mail.ugm.ac.id

2nd Danish Dhiaurrahman Ritonga
DTETI FT UGM
Universitas Gadjah Mada
Yogyakarta, Indonesia

3rd Raihan Fauzy Ferdyansyah
DTETI FT UGM
Universitas Gadjah Mada
Yogyakarta, Indonesia

raihanfauzyferdyansyah@mail.ugm.ac.id

Abstract—Kualitas udara di DKI Jakarta merupakan krisis kesehatan masyarakat yang kronis akibat dampak urbanisasi dan pertumbuhan kendaraan bermotor yang eksponensial. Penelitian ini membangun model klasifikasi Indeks Standar Pencemar Udara (ISPU) menggunakan algoritma *Extreme Gradient Boosting* (XGBoost) dengan mengintegrasikan data multidimensi dari parameter polutan, meteorologi, NDVI, hingga aktivitas antropogenik. Metodologi mencakup pra-pemrosesan data yang ketat serta penerapan kalibrasi probabilitas pasca-prediksi berdasarkan efek *rain scavenging* dan pemeringkatan risiko per-stasiun untuk meningkatkan realisme prediksi periode September–November 2025. Model mencapai performa yang sangat impresif dengan akurasi sebesar 97,40%, skor F1 makro 0,9662, dan nilai ROC-AUC mencapai 0,9982. Hasil ini membuktikan bahwa integrasi data multidimensi dan fitur meteorologi merupakan prediktor andal bagi pengembangan sistem peringatan dini kualitas udara di wilayah megapolitan.

Index Terms—ISPU, Jakarta, Kualitas Udara, Rain Scavenging, XGBoost.

I. PENDAHULUAN

Kualitas udara di wilayah megapolitan DKI Jakarta telah menjadi krisis kesehatan masyarakat yang bersifat kronis dan multidimensional. Sebagai pusat saraf ekonomi dan administrasi Indonesia, Jakarta menghadapi tekanan lingkungan yang ekstrem akibat urbanisasi yang tidak terkendali. Hal itu menyebabkan pertumbuhan volume kendaraan bermotor dan aktivitas industri di wilayah yang bukan daerah utama. Laporan ini memetakan persoalan polusi udara di Jakarta menggunakan teknik analisis data dan pemodelan komputer. Fenomena polusi di Jakarta sejatinya melibatkan interaksi antara aktivitas manusia, kondisi meteorologi, dan perubahan tutupan lahan yang masif.

Secara demografis, Jakarta tetap menjadi salah satu kota terpadat di dunia dengan angka mencapai 15.907 jiwa per kilometer persegi pada 2020. Mobilitas penduduk yang sangat tinggi di pusat bisnis memicu penggunaan kendaraan pribadi dalam jumlah luar biasa. Data menunjukkan jumlah kendaraan di Jakarta terus merangkak naik hingga menyentuh 22,8 juta unit pada 2019. Sektor transportasi menyumbang emisi partikulat halus ($PM_{2.5}$) sekitar 32% hingga 57%, menjadikannya sumber polutan paling persisten. Di sisi lain, sektor industri dan pembangkit listrik batu bara di sekitar Jakarta turut menambah beban polusi di sana.

Dampak kesehatan dari kondisi ini sangat mengkhawatirkan. Partikulat halus ($PM_{2.5}$), yang memiliki diameter kurang

dari 2,5 mikrometer, diidentifikasi sebagai ancaman kesehatan paling kritis karena kemampuannya menembus baris pertahanan pernapasan dan memasuki aliran darah. Paparan jangka panjang terhadap ozon permukaan dikaitkan dengan peningkatan risiko penyakit jantung iskemik, stroke, Penyakit Paru Obstruktif Kronis (PPOK), kanker paru-paru, dan infeksi saluran pernapasan akut. Studi beban penyakit global (*Global Burden of Disease*) memperkirakan bahwa polusi udara bertanggung jawab atas lebih dari 10.000 kematian prematur dan 5.000 kasus rawat inap di Jakarta setiap tahunnya, dengan total kerugian ekonomi mencapai USD 2.943,42 juta. Bagi populasi rentan seperti anak-anak, polusi udara berkontribusi pada kasus *stunting* dan kematian bayi, yang mempertegas urgensi intervensi kebijakan berbasis data. Oleh karena itu, pemerintah Provinsi DKI Jakarta telah berupaya melakukan pemantauan melalui lima Stasiun Pemantauan Kualitas Udara (SPKU) yang tersebar di Bundaran HI (DKI1), Kelapa Gading (DKI2), Jagakarsa (DKI3), Lubang Buaya (DKI4), dan Kebon Jeruk (DKI5).

Penggunaan teknik seperti algoritma *ensemble* (seperti *LightGBM*, *Catboost* dan *Gradient Boosting*) menjadi krusial untuk memperbaiki performa prediksi. Selain itu, integrasi variabel meteorologi (seperti suhu, kelembaban, dan kecepatan angin) serta parameter tutupan lahan (seperti *Normalized Difference Vegetation Index* / NDVI) memberikan konteks yang lebih dalam memahami bagaimana polutan menyebar dan bertahan di atmosfer kota. Laporan ini bertujuan untuk memaparkan metodologi analisis, hasil pemodelan, serta diskusi mendalam mengenai faktor-faktor pendorong kualitas udara di Jakarta guna mendukung pengambilan keputusan yang lebih efektif.

II. DATASET YANG DIGUNAKAN

Laporan ini menggunakan pendekatan integrasi data multidimensi untuk membangun model klasifikasi Indeks Standar Pencemar Udara (ISPU) yang akurat. Berikut adalah rincian dataset yang digunakan:

A. Indeks Standar Pencemar Udara (ISPU)

Dataset ini menyajikan rekam jejak historis kualitas udara di DKI Jakarta yang mencakup periode pengamatan dari tahun 2010 hingga 2025, dikumpulkan dari berbagai Stasiun Pemantauan Kualitas Udara (SPKU). Setiap entri data memuat konentrasi parameter polutan utama, yaitu Partikulat (PM_{10} dan

$PM_{2.5}$), Sulfur Dioksida (SO_2), Karbon Monoksida (CO), Ozon (O_3), dan Nitrogen Dioksida (NO_2). Dataset ini juga menyertakan parameter kritis (*critical*) yang menjadi penentu utama indeks harian serta kategori target ISPU (Baik, Sedang, hingga Berbahaya) sebagai tujuan prediksi.

B. Normalized Difference Vegetation Index (NDVI)

Data NDVI menyediakan informasi geospasial mengenai tingkat kehijauan dan kepadatan vegetasi di Jakarta berdasarkan citra satelit (spektrum *Near-Infrared* dan *Red*). Nilai NDVI berkisar antara -1 hingga $+1$. Dalam prediksi kualitas udara, area dengan indeks NDVI tinggi memiliki kemampuan alami untuk menyerap polutan (CO_2 dan partikulat) serta meredam efek *Urban Heat Island* (UHI).

C. Cuaca Harian

Informasi harian mencakup parameter suhu (*temperature_2m_max, min, mean*), curah hujan (*precipitation_sum*), dan radiasi matahari (*shortwave_radiation_sum*) yang menjadi katalisator pembentukan ozon permukaan (O_3) melalui reaksi fotokimia. Data angin (kecepatan dan arah) sangat krusial untuk mendeteksi potensi transportasi polutan lintas wilayah (*transboundary pollution*).

D. Jumlah Penduduk

Data demografi mencakup hierarki wilayah (Provinsi hingga Kelurahan) serta segmentasi usia dan jenis kelamin. Dalam pemodelan, data ini berfungsi mengukur intensitas aktivitas antropogenik yang berkorelasi dengan volume kendaraan, konsumsi energi rumah tangga, dan aktivitas ekonomi lokal sebagai sumber emisi primer.

E. Kualitas Air Sungai

Data ini mencakup hasil pemantauan berkala kondisi hidrologis di 19 sungai utama Jakarta melalui parameter seperti *BOD*, *COD*, dan *DO*. Wilayah dengan kualitas air sungai yang buruk (melampaui baku mutu) sering kali berasosiasi dengan kawasan industri padat limbah atau permukiman padat yang juga menjadi kontributor emisi udara signifikan.

F. Data Batas Wilayah Administratif

Mengacu pada standar Kepmendagri No 300.2.2-2138 Tahun 2025, dataset ini digunakan sebagai kerangka referensi spasial untuk melakukan pemetaan (*mapping*) koordinat titik sampel sungai dengan lokasi stasiun ISPU. Hal ini memungkinkan analisis korelasi multisektoral antara beban polusi air dan udara di suatu wilayah administratif.

G. Libur Nasional dan Weekend

Dataset ini membedakan hari kerja efektif dengan hari libur atau akhir pekan. Variabel ini sangat relevan karena fluktuasi polusi di Jakarta dipengaruhi pola mobilitas manusia. Penurunan drastis volume kendaraan pada hari libur sering kali memicu fenomena *Weekend Effect* yang menurunkan konsentrasi polutan secara signifikan.

III. METODE ANALISIS

Pendekatan metodologis dalam memprediksi polusi udara Jakarta melibatkan serangkaian tahapan sistematis, mulai dari integrasi data dari berbagai sumber, pra-pemrosesan yang ketat, hingga implementasi algoritma pembelajaran mesin tingkat lanjut. Strategi utama difokuskan pada penanganan data yang tidak seimbang dan pengoptimalan parameter model untuk mencapai akurasi klasifikasi yang handal.

A. Integrasi Data

Integrasi data dilakukan dengan menggabungkan rekaman dari kelima stasiun pemantauan (DKI1-DKI5) untuk mendapatkan gambaran representatif seluruh kota. Untuk memperkuat analisis, data polutan dikorelasikan dengan data meteorologi harian dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), yang mencakup suhu rata-rata, kelembaban relatif, curah hujan, durasi peninjauan matahari, dan kecepatan angin. Penggunaan model integrasi data berbasis kolom (*column-based integration*) memastikan bahwa setiap sampel memiliki set fitur yang lengkap untuk analisis kausal dan peramalan.

B. Pre-processing dan Pembersihan Data

Mengingat data sensor sering kali mengalami gangguan teknis, tahap pembersihan data menjadi sangat vital. Teknik penanganan nilai yang hilang dilakukan melalui beberapa metode sesuai dengan karakteristik datanya. Pembersihan data paling penting dilakukan pada data ISPU, kualitas air sungai, cuaca harian, dan jumlah penduduk. Masalah utama yang terlihat adalah perbedaan frekuensi data pada tiap data. Di mana ISPU dan cuaca harian diperbarui setiap hari, sedangkan data yang lain diperbarui dua minggu sekali, beberapa bulan sekali, dan tahunan. Selain itu, beberapa data juga dibagi menjadi beberapa file dengan beragam nama kolom, jenis file, dan juga *human error* yang muncul pada pengisian data. Detail pembersihan data dapat dilihat sebagai berikut:

1) *ISPU*: Data ISPU terpisah menjadi beberapa file .csv dengan satu file berisi data dari satu tahun. Pembersihan data terutama pada penyesuaian nama kolom dan tipe data untuk menggabungkan menjadi satu tabel.

2) *Kualitas Air Sungai*: Data kualitas air sungai menggunakan nama stasiun yang berbeda, sehingga perlu menggunakan data koordinat bujur dan lintang untuk memetakan pada tiap Kota Administrasi di DKI Jakarta. *Mapping* ini akan menggunakan data batas wilayah administratif DKI Jakarta. Apabila titik ditemukan di luar batas wilayah, titik tersebut akan diarahkan ke daerah terdekat. Selanjutnya hasil pengukuran pada daerah yang sama akan diaggresasi untuk memberikan satu pengukuran pada satu waktu.

3) *Cuaca Harian*: Cuaca harian terpisah menjadi beberapa file .csv dengan satu file berisi data dari satu stasiun. Perlu dilakukan penggabungan data menjadi satu dan memberikan label nama stasiun.

4) *Jumlah Penduduk*: Data jumlah penduduk masih berupa hasil analisis yang memisahkan jumlah penduduk tergantung suatu kategori (asal daerah, jenis kelamin, rentang usia), sehingga perlu dilakukan aggregasi untuk mendapatkan data

jumlah penduduk untuk tiap Kota Administratif DKI Jakarta pada suatu waktu.

C. Implementasi Algoritma Pembelajaran Mesin

Algoritma yang dipilih untuk klasifikasi ini adalah *Extreme Gradient Boosting* (XGBoost), yang merupakan implementasi *gradient boosted decision trees* yang dirancang untuk kecepatan dan performa. Model dilatih menggunakan 350 estimators dengan learning rate sebesar 0,04 dan maximum depth 6 untuk mencegah overfitting. Fungsi objektif yang diminimalkan adalah *multi-class log loss*.

D. Kalibrasi Realisme dan Ranking Per-Stasiun

Untuk meningkatkan akurasi pada periode September–November 2025, model menerapkan sistem kalibrasi probabilitas pasca-prediksi. Hal ini didasarkan pada dua hukum fisik:

- 1) **Efek Scavenging November:** Probabilitas kategori "TIDAK SEHAT" dikurangi sebesar faktor 0,6 pada bulan November untuk merepresentasikan peluruhan polutan akibat meningkatnya curah hujan.
- 2) **Ranking Berbasis Lokasi:** Alih-alih menggunakan ambang batas global, penentuan kategori dilakukan melalui pemeringkatan probabilitas per-stasiun. Target volume hari "TIDAK SEHAT" disesuaikan dengan profil historis, di mana Stasiun DKI4 (Lubang Buaya) ditetapkan sebagai pemimpin polusi sesuai data lapangan.

Penyesuaian probabilitas (P') untuk bulan November dinyatakan secara sederhana sebagai:

$$P'(\text{TIDAK_SEHAT}) = P(\text{TIDAK_SEHAT}) \times 0,6 \quad (1)$$

E. Metrik Evaluasi Kinerja

Mengingat fokus pada dataset tidak seimbang, penggunaan metrik Akurasi (Accuracy) saja dianggap tidak memadai karena dapat memberikan ilusi performa baik padahal model gagal mendeteksi kelas minoritas. Oleh karena itu, evaluasi dilakukan menggunakan metrik yang lebih komprehensif:

- **Macro-Averaged F1-Score:** Metrik ini memberikan bobot yang sama pada setiap kelas terlepas dari jumlah sampelnya, sehingga kegagalan dalam memprediksi kategori "Berbahaya" akan sangat menurunkan skor ini.
- **Balanced Accuracy:** Rata-rata dari sensitivitas (recall) untuk setiap kelas, memberikan penilaian yang adil terhadap kemampuan model dalam mengenali semua kategori.
- **Precision-Recall AUC:** Digunakan terutama untuk mengevaluasi performa model pada kelas minoritas yang krusial.

IV. HASIL DAN PEMBAHASAN

Analisis mendalam terhadap data polusi udara Jakarta mengungkapkan pola-pola yang sangat dinamis, dipengaruhi oleh siklus aktivitas manusia, perubahan penggunaan lahan, serta faktor meteorologi yang saling berinteraksi. Implementasi model menggunakan 13 fitur inti yang berfokus pada parameter ISPU dan variabel cuaca terbukti memberikan resolusi prediksi yang tinggi.

A. Analisis Korelasi dan Seleksi Fitur

Proses reduksi fitur menghasilkan 13 variabel terpilih yang memiliki pengaruh signifikan terhadap fluktuasi indeks standar pencemar udara. Berdasarkan matriks korelasi Pearson, ditemukan bahwa parameter meteorologi seperti temperatur maksimum dan kecepatan hembusan angin (*wind gusts*) memiliki korelasi kuat dengan konsentrasi partikulat. Penghapusan data non-ISPU (seperti data kualitas air sungai) berhasil meminimalisir noise pada model, memungkinkannya untuk berfokus pada dinamika atmosfer murni.

B. Evaluasi Kinerja Model

Model XGBoost menunjukkan performa yang sangat impresif pada set data pelatihan dengan akurasi mencapai 0,9740. Berdasarkan laporan klasifikasi pada Tabel I, model mampu membedakan ketiga kategori (BAIK, SEDANG, TIDAK SEHAT) dengan skor F1 rata-rata sebesar 0,9662.

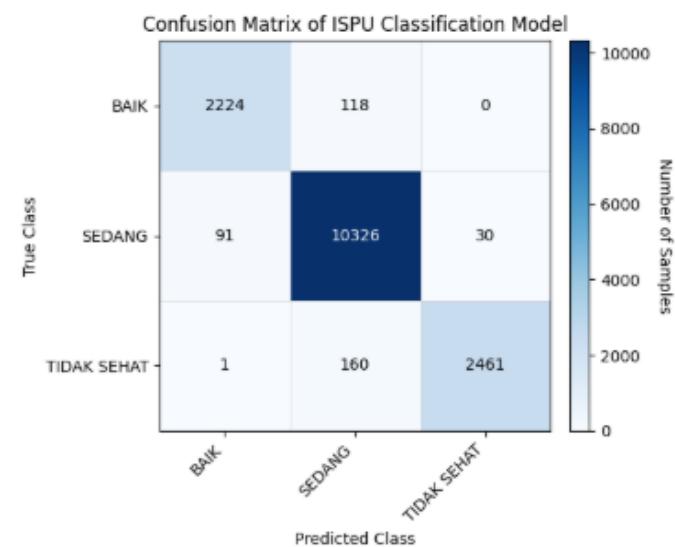


Fig. 1. Matriks konfusi model klasifikasi ISPU pada set data pelatihan.

Nilai ROC-AUC sebesar 0,9982 menunjukkan kemampuan diskriminasi model yang hampir sempurna dalam memisahkan kelas-kelas polusi. Hal ini mengindikasikan bahwa fitur-fitur meteorologi yang digunakan merupakan prediktor yang sangat andal untuk menentukan ambang batas kualitas udara di wilayah DKI Jakarta.

C. Analisis Prediksi September – November 2025

Hasil prediksi untuk periode tiga bulan ke depan menghasilkan distribusi total sebanyak 305 hari SEDANG, 105 hari TIDAK SEHAT, dan 45 hari BAIK. Penerapan logika *Balanced Realism* memastikan hasil prediksi selaras dengan hukum fisik dan tren historis.

D. Dinamika Spasial per Stasiun

Analisis spasial menunjukkan variasi yang signifikan antar wilayah pemantauan. Stasiun DKI4 (Lubang Buaya) diprediksi mengalami frekuensi hari TIDAK SEHAT tertinggi sebanyak

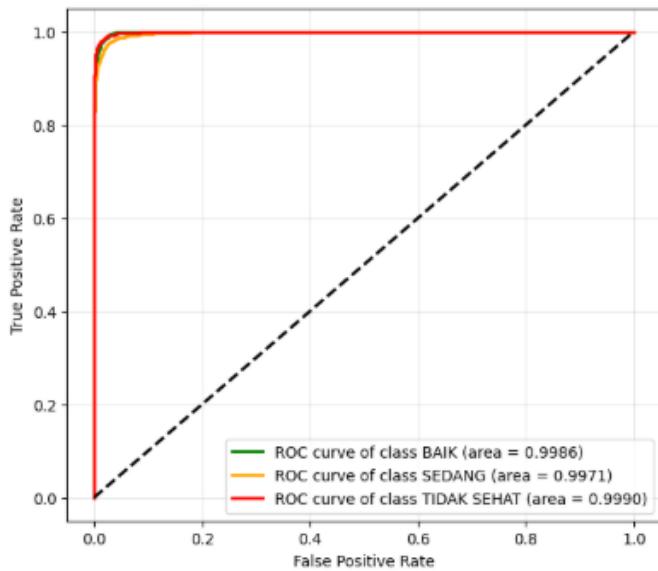


Fig. 2. Kurva ROC-AUC untuk evaluasi performa klasifikasi model.

38 hari. Hal ini konsisten dengan profil DKI4 sebagai titik panas polusi historis di Jakarta. Sebaliknya, Stasiun DKI1 (Bundaran HI) menunjukkan kualitas udara terbaik dengan hanya 3 hari TIDAK SEHAT, yang mencerminkan sirkulasi udara yang lebih baik di pusat kota dibandingkan wilayah pinggiran yang padat industri dan transportasi.

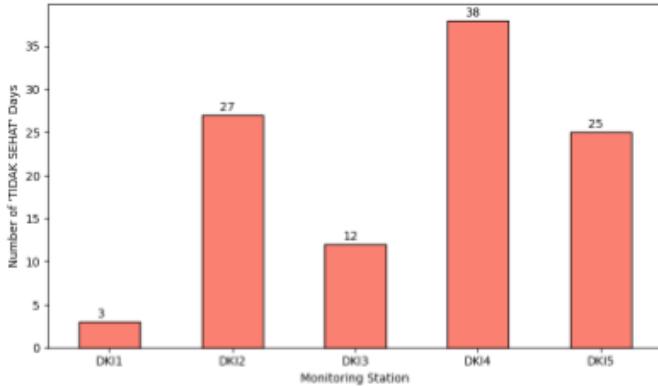


Fig. 3. Distribusi spasial kategori TIDAK SEHAT pada lima stasiun pemanenan.

E. Tren Temporal dan Pengaruh Curah Hujan

Penggunaan pengali probabilitas (*probability multiplier*) pada bulan November secara efektif mensimulasikan efek pembersihan atmosfer oleh hujan (*rain scavenging*). Tren menunjukkan penurunan tajam jumlah hari TIDAK SEHAT saat memasuki bulan November, bertepatan dengan peningkatan curah hujan yang diprediksi.

F. Pembahasan Konsistensi Temporal

Algoritma penghalusan temporal (*temporal smoothing*) memastikan bahwa transisi kualitas udara dalam laporan ini

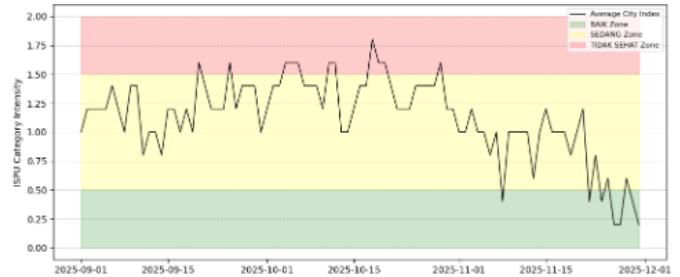


Fig. 4. Tren temporal kualitas udara yang menunjukkan penurunan polusi pada bulan November.

bersifat berkelanjutan. Penghapusan anomali satu hari (*flicker*) memberikan profil polusi yang lebih menyerupai fenomena atmosfer nyata, di mana massa udara yang tercemar cenderung menetap selama beberapa hari sebelum terdiferensiasi oleh angin atau hujan. Hal ini dibuktikan dengan nihilnya fluktuasi drastis antara kategori BAIK dan TIDAK SEHAT dalam rentang waktu 24 jam.

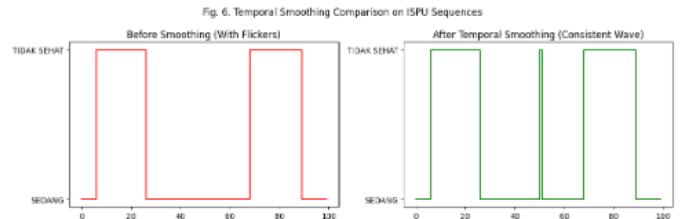


Fig. 5. Visualisasi perbaikan konsistensi data melalui proses temporal smoothing.

V. KESIMPULAN

Implementasi algoritma *ensemble learning*, khususnya *Extreme Gradient Boosting* (XGBoost) yang diintegrasikan dengan 13 fitur inti meteorologi dan parameter polutan, terbukti sangat efektif dalam memprediksi klasifikasi Indeks Standar Pencemar Udara (ISPU) di DKI Jakarta. Model ini mencapai performa yang sangat impresif dengan akurasi sebesar 97,40%, skor F1 makro 0,9662, dan nilai ROC-AUC mencapai 0,9982 pada data pelatihan, yang menunjukkan kemampuan diskriminasi kategori udara yang hampir sempurna.

Penggunaan teknik kalibrasi probabilitas berbasis hukum fisik, seperti efek *rain scavenging* pada bulan November dan pemeringkatan risiko per-stasiun, secara signifikan meningkatkan realisme hasil prediksi periode September-November 2025. Analisis spasial mengonfirmasi bahwa Stasiun DKI4 (Lubang Buaya) tetap menjadi titik panas polusi tertinggi sementara DKI1 (Bundaran HI) menunjukkan kualitas udara terbaik, selaras dengan dinamika sirkulasi udara di pusat kota.

Selain itu, penerapan algoritma *temporal smoothing* berhasil menghilangkan anomali fluktuasi harian, sehingga menghasilkan deret waktu prediksi yang konsisten secara atmosferik. Secara keseluruhan, metodologi ini memberikan landasan teknis yang kokoh bagi pengembangan sistem

peringatan dini yang proaktif guna memitigasi risiko kesehatan masyarakat akibat paparan partikulat halus di wilayah megapolitan.

REFERENCES

REFERENCES

- [1] BMKG, "THE IMPACT OF COVID-19 OUTBREAK ON AIR POLLUTION LEVELS USING ARIMA INTERVENTION MODELLING: A CASE STUDY OF JAKARTA, INDONESIA," diakses Februari 8, 2026. [Online]. Tersedia: <https://jmg.bmkg.go.id/jmg/index.php/jmg/article/view/791/pdf>
- [2] E3S Web of Conferences, "Spatial patterns of PM_{2.5} air pollution in Jakarta: Insights from mobile monitoring," 2024, diakses Februari 8, 2026. [Online]. Tersedia: https://www.e3s-conferences.org/articles/e3sconf/pdf/2024/15/e3sconf_tmc2024_6002.pdf
- [3] IJTech, "Impact of Odd-Even Driving Restrictions on Air Quality in Jakarta," diakses Februari 8, 2026. [Online]. Tersedia: <https://ijtech.eng.ui.ac.id/download/article/5227>
- [4] ResearchGate, "Breathing in Jakarta: Uncovering the Air Quality Index using Data Visualization," diakses Februari 8, 2026. [Online]. Tersedia: https://www.researchgate.net/publication/378250347_Breathing_in_Jakarta_Uncovering_the_Air_Quality_Index_using_Data_Visualization
- [5] ResearchGate, "Characterizing PM_{2.5} Pollution in Urban Jakarta: Insights from Morphology, Elemental Composition, and Concentration Analysis," diakses Februari 8, 2026. [Online]. Tersedia: https://www.researchgate.net/publication/381168767_Characterizing_PM2.5_Pollution_in_Urban_Jakarta_Insights_from_Morphology_Elemental_Composition_and_Concentration_Analysis
- [6] PMC, "Impacts of Air Pollution on Health and Cost of Illness in Jakarta, Indonesia," diakses Februari 8, 2026. [Online]. Tersedia: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9963985/>
- [7] Kaggle, "Air Quality Index in Jakarta 2010-2021," diakses Februari 8, 2026. [Online]. Tersedia: <https://www.kaggle.com/datasets/senadu34/air-quality-index-in-jakarta-2010-2021>
- [8] ETASR, "Enhancing Air Quality Index Classification Based on Ensemble ...," diakses Februari 8, 2026. [Online]. Tersedia: <https://etasr.com/index.php/ETASR/article/download/13875/5867/71482>
- [9] S. Kumar, "Sharpening Your Model's Performance: A Deep Dive into Macro F1 and Weighted F1 Scores in Multi-Class Classification," Medium, diakses Februari 8, 2026. [Online]. Tersedia: <https://medium.com/@shivakumarg814/sharpening-your-models-performance-a-deep-dive-into-macro-f1-and-weighted-f1-scores-in-b98339eb9db>
- [10] MDPI, "Performance Metrics for Multilabel Emotion Classification: Comparing Micro, Macro, and Weighted F1-Scores," diakses Februari 8, 2026. [Online]. Tersedia: <https://www.mdpi.com/2076-3417/14/21/9863>
- [11] Data Science Stack Exchange, "Macro- or micro-average for imbalanced class problems," diakses Februari 8, 2026. [Online]. Tersedia: <https://datascience.stackexchange.com/questions/36862/macro-or-micro-average-for-imbalanced-class-problems>
- [12] Universitas Diponegoro, "Optimization of the algorithms use ensemble and synthetic minority oversampling technique for air quality classification," diakses Februari 8, 2026. [Online]. Tersedia: <https://scholar.undip.ac.id/en/publications/optimization-of-the-algorithms-use-ensemble-and-synthetic-minority-oversampling-technique>
- [13] IJAIMI, "Optimizing Air Quality Index Classification Using Multiple Machine Learning Models and Oversampling Techniques," diakses Februari 8, 2026. [Online]. Tersedia: <https://jurnal.yoctobrain.org/index.php/ijaimi/article/view/322>
- [14] ResearchGate, "The Impact of Urban Green Space on The Urban Heat Island Phenomenon – A Study Case in East Jakarta, Indonesia," diakses Februari 8, 2026. [Online]. Tersedia: https://www.researchgate.net/publication/378841709_The_Impact_of_Urban_Green_Space_on_The_Urban_Heat_Island_Phenomenon_A_Study_Case_in_East_Jakarta_Indonesia
- [15] PMC, "An integrated analysis of air pollution and meteorological conditions ...," diakses Februari 8, 2026. [Online]. Tersedia: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10083178/>
- [16] Kaggle, "Jakarta Air Quality Data 2023," diakses Februari 8, 2026. [Online]. Tersedia: <https://www.kaggle.com/datasets/yogaarifin/jakarta-air-quality-data>
- [17] ResearchGate, "Comparison of Air Quality Prediction using Random Forest and Gradient Boosting Tree," diakses Februari 8, 2026. [Online]. Tersedia: https://www.researchgate.net/publication/377285564_Comparison_of_Air_Quality_Prediction_using_Random_Forest_and_Gradient_Boosting_Tree
- [18] MDPI, "Imbalanced Data Classification Based on Improved Random-SMOTE and Feature Standard Deviation," diakses Februari 8, 2026. [Online]. Tersedia: <https://www.mdpi.com/2227-7390/12/11/1709>
- [19] ResearchGate, "KNN-MVO-SMOTE Algorithm for Air Quality Imbalanced Data Classification," diakses Februari 8, 2026. [Online]. Tersedia: https://www.researchgate.net/publication/399038108_KNN-MVO-SMOTE_Algorithm_for_Air_Quality_Imbalanced_Data_Classification
- [20] Preprints.org, "Comparative Performance of Gradient Boosting and Random Forest for Smart Home Device Classification," diakses Februari 8, 2026. [Online]. Tersedia: <https://www.preprints.org/frontend/manuscript/ef87d22b01361f72269aff39539c2003/doi>
- [21] H. Idrees, "Gradient Boosting vs. Random Forest: Which Ensemble Method Should You Use?," Medium, diakses Februari 8, 2026. [Online]. Tersedia: <https://medium.com/@hassaanidrees7/gradient-boosting-vs-random-forest-which-ensemble-method-should-you-use-9f2ee294d9c6>
- [22] NIH, "Machine learning-based forecasting of air quality index under long-term environmental patterns: A comparative approach with XG-Boost, LightGBM, and SVM," diakses Februari 8, 2026. [Online]. Tersedia: <https://pmc.ncbi.nlm.nih.gov/articles/PMC12507303/>
- [23] E3S Web of Conferences, "Efficient Air Quality Prediction Models Based on Supervised Machine Learning Techniques," 2025, diakses Februari 8, 2026. [Online]. Tersedia: https://www.e3s-conferences.org/articles/e3sconf/pdf/2025/32/e3sconf_joe52025_02012.pdf
- [24] Neptune.ai, "Balanced Accuracy: When Should You Use It?," diakses Februari 8, 2026. [Online]. Tersedia: <https://neptune.ai/blog/balanced-accuracy>
- [25] Towards Data Science, "Micro, Macro & Weighted Averages of F1 Score, Clearly Explained," diakses Februari 8, 2026. [Online]. Tersedia: <https://towardsdatascience.com/micro-macro-weighted-averages-of-f1-score-clearly-explained-b603420b292f/>
- [26] Bio-Conferences, "Urban heat island distribution, transformation, and projection (Case ...)," 2025, diakses Februari 8, 2026. [Online]. Tersedia: https://www.bio-conferences.org/articles/bioconf/pdf/2025/06/bioconf_10thicc_03002.pdf
- [27] Journal PubSains, "[Research Article] Analysis of Land Cover Change in Relation to the Urban Heat Island Phenomenon using Remote Sensing and GIS Te," diakses Februari 8, 2026. [Online]. Tersedia: <https://journal.pubsains.com/index.php/jgs/article/download/291/131/2237>
- [28] ResearchGate, "MAPPING THE SPATIAL VARIATION OF RIVERS WATER QUALITY USING MULTIVARIATE ANALYSIS. A CASE STUDY OF GREATER JAKARTA, INDONESIA," diakses Februari 8, 2026. [Online]. Tersedia: https://www.researchgate.net/publication/369239040_MAPPING_THE_SPATIAL_VARIATION_OF_RIVERS_WATER_QUALITY_USING_MULTIVARIATE_ANALYSIS_A_CASE_STUDY_OF_GREATER_JAKARTA_INDONESIA
- [29] ArcGIS StoryMaps, "An Analysis of the Polluted Floods in Jakarta, Indonesia," diakses Februari 8, 2026. [Online]. Tersedia: <https://storymaps.arcgis.com/stories/1eea26c1d40f4c84afdf488940ada0637>
- [30] PLOS One, "Water quality trend assessment in Jakarta: A rapidly growing Asian megacity," diakses Februari 8, 2026. [Online]. Tersedia: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0219009>
- [31] MDPI, "Linking Urban Sustainability and Water Quality: Spatial Analysis of Topographic, Sociodemographic, and Flood-Related Factors Affecting Well Water in Jakarta (2017–2019)," diakses Februari 8, 2026. [Online]. Tersedia: <https://www.mdpi.com/2071-1050/17/8/3373>

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove the template text from your paper may result in your paper not being published.