JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 14 No. 1, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v14i1.8477

ANALISIS KLASIFIKASI INDEKS STANDAR PENCEMARAN UDARA JAKARTA TAHUN 2025 MENGGUNAKAN ALGORITMA RANDOM FOREST

Andhika Nur Fachmi[1], Anisa Permata Sari[2], Ferdinan Restu Ramadhan[3], Reihan Dwi Patria[4], Salma Pudjiati[5], Fuad Nur Hassan[6]

[1,2,3,4,5,6] Jurusan Informatika, Universitas Bina Sarana Informatika; Jl. Margonda No. 8, Pondok Cina, Kota Depok, Jawa Barat, 16424

Keywords:
Klasifikasi ISPU; Kualitas Udara; *Random Forest*; SMOTE; $PM_{2.5.}$

Corespondent Email:
andhikafachmi2545@gmail.com

Abstrak. Pencemaran udara di DKI Jakarta telah menjadi isu lingkungan kritis dengan fluktuasi konsentrasi polutan yang kompleks, menuntut metode pemantauan yang lebih presisi dibandingkan pendekatan konvensional. Penelitian ini bertujuan untuk mengklasifikasikan tiga kategori utama Indeks Standar Pencemaran Udara (ISPU) menggunakan algoritma *Machine Learning Random Forest*. Studi ini memanfaatkan dataset harian terbaru periode Januari hingga Agustus 2025 yang mencakup parameter PM_{10}, $PM_{2.5}$, SO_2, CO, O_3, dan NO_2. Guna mengatasi ketidakseimbangan distribusi kelas pada data kategori ISPU, diterapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) pada tahap pra-pemrosesan. Hasil evaluasi model menunjukkan kinerja yang sangat impresif dengan tingkat akurasi mencapai 99,50% pada data pengujian. Analisis *feature importance* mengidentifikasi bahwa $PM_{2.5}$ merupakan parameter paling dominan dengan kontribusi pengaruh sebesar 30,68% terhadap penentuan kualitas udara. Temuan ini memvalidasi efektivitas *Random Forest* sebagai instrumen sistem peringatan dini yang andal serta menekankan urgensi kebijakan pengendalian emisi partikulat di Jakarta.

Abstract. *Air pollution in DKI Jakarta has become a critical environmental issue with complex fluctuations in pollutant concentrations, demanding more precise monitoring methods compared to conventional approaches. This study aims to classify the three main categories of the Air Pollutant Standard Index (ISPU) categories using the Random Forest Machine Learning algorithm. The study utilizes the latest daily dataset from January to August 2025, covering PM_{10}, $PM_{2.5}$, SO_2, CO, O_3, and NO_2. To address the class distribution imbalance in the ISPU category data, the Synthetic Minority Over-sampling Technique (SMOTE) was applied during the pre-processing stage. Model evaluation results demonstrated impressive performance with an accuracy rate reaching 99.50% on testing data. Feature importance analysis identified $PM_{2.5.}$ as the most dominant parameter, contributing 30.68% to the determination of air quality. These findings validate the effectiveness of Random Forest as a reliable early warning system instrument and highlight the urgency of particulate emission control policies in Jakarta.*

1. PENDAHULUAN

Polusi udara di kota-kota seperti DKI Jakarta telah berkembang menjadi masalah lingkungan yang mendesak yang berdampak negatif pada kesehatan masyarakat. Penyakit pernapasan akut dan gangguan kardiovaskular telah terbukti berkorelasi linear dengan peningkatan konsentrasi polutan, terutama

Particulate Matter ($PM_{2.5}$) [1]. Sebagai tanggapan terhadap kondisi ini, pemerintah menggunakan Indeks Standar Pencemaran Udara (ISPU), yang memantau dan memberikan informasi tentang kondisi kualitas udara kepada masyarakat. ISPU mencakup parameter seperti PM_{10}, $PM_{2.5}$, SO_2, CO, O_3, dan NO_2 [2]. Namun, metode statistik konvensional seringkali sulit untuk memberikan klasifikasi status udara yang tepat secara real-time karena data polutan sangat berubah dan non-linear.

Mengatasi masalah ini dengan teknologi pembelajaran mesin telah banyak dipelajari. Metode clustering K-Means digunakan untuk memetakan pola kualitas udara di Jakarta dalam penelitian yang dilakukan oleh Rahmadenti [2025] [3].

Meskipun penelitian tersebut menemukan tren musiman, pendekatan mereka bersifat deskriptif (pembelajaran yang tidak diawasi), dan mereka tidak menghasilkan model prediktif yang dapat menentukan kategori risiko tertentu. Sebaliknya, Firdaus et al. (2024) berhasil menggunakan algoritma Random Forest yang sangat akurat untuk klasifikasi ISPU. [4]. Namun, penelitian tersebut menggunakan dataset historis dari tahun 2016 hingga 2021, sehingga tidak mewakili kondisi atmosfer Jakarta pasca pandemi maupun kondisi terbaru di tahun 2025.

Selain tantangan relevansi data tahunan, ketidakseimbangan kelas (Imbalanced Data) merupakan tantangan teknis signifikan lain dalam proses klasifikasi kualitas udara. Data ISPU yang historis menunjukkan bahwa jumlah sampel dalam kategori risiko tinggi seperti “Tidak Sehat” seringkali jauh lebih sedikit daripada kategori “Sedang” atau “Baik”. Kondisi data yang timpang ini berpotensi menyebabkan bias pada model Machine Learning, di mana model akan cenderung mengklasifikasikan data baru ke dalam kelas mayoritas. Oleh karena itu, penelitian ini mengimplementasikan teknik Synthetic Minority Over-sampling Technique (SMOTE) untuk menyeimbangkan distribusi kelas, memastikan model Random Forest yang dibangun mampu mengenali pola pada kelas minoritas dengan akurat [5],[6].

Penelitian ini bertujuan untuk membangun model klasifikasi ISPU di Jakarta dengan data terbaru dari Januari hingga Agustus 2025 menggunakan algoritma Random Forest yang dioptimasi menggunakan metode SMOTE, berdasarkan analisis kesenjangan. Penggunaan dataset terbaru serta analisis menyeluruh karakteristik penting untuk mengidentifikasi polutan yang paling dominan dalam pengaruh penurunan kualitas udara diberikan oleh penelitian ini. Hasil penelitian diharapkan dapat berfungsi sebagai referensi yang dapat diandalkan untuk mengembangkan sistem peringatan dini yang lebih akurat untuk kualitas udara.

2. TINJAUAN PUSTAKA

2.1. Kualitas Udara

Kualitas udara adalah istilah yang mengacu pada kondisi atmosfer yang diukur berdasarkan konsentrasi pencemar fisik dan kimia dalam jangka waktu tertentu. Faktor cuaca seperti curah hujan dan kecepatan angin memengaruhi dispersi polutan, dan ini memengaruhi penurunan kualitas udara di daerah metropolitan [7]. Paparan jangka panjang terhadap kualitas udara yang buruk, terutama yang mengandung partikel halus ($PM_{2.5}$), dikaitkan dengan risiko kesehatan masyarakat yang lebih tinggi di daerah urban yang padat aktivitas transportasi [8].

2.2. Indeks Standar Pencemaran Udara

Indeks Standar Pencemaran Udara (ISPU) dibuat oleh pemerintah Indonesia sebagai acuan resmi untuk kualitas udara ambien. Indeks ini dihitung berdasarkan kombinasi lima parameter polutan utama, yaitu PM10, $PM_{2.5}$, SO_2, CO, O_3, dan NO_2 [9]. Lalu kemudian digabungkan menjadi satu nilai untuk meningkatkan kesadaran publik tentang ancaman kesehatan yang terkait. Untuk menentukan kategori status mutu udara dari “Baik” hingga “Berbahaya”, data parameter ini dikonversi menjadi angka indeks tanpa satuan [10].

2.3. Algoritma Random Forest

Random Forest menggunakan pendekatan kelompok dalam kategori pembelajaran terawasi (*supervised learning*), yang menggabungkan kekuatan dari berbagai pohon keputusan *(decission tree*). Alih-alih bergantung pada satu model, algoritma ini

membangun sekumpulan pohon secara acak dan menggunakan mekanisme pemungutan suara mayoritas untuk menentukan hasil prediksi akhir. Terbukti bahwa pendekatan ini dapat mengurangi varians dan risiko *overfitting* yang sering terjadi pada model pohon keputusan tunggal [11].

Algoritma ini terbukti dapat menemukan variabel yang paling penting (variabel penting) dalam klasifikasi kualitas udara dan memberikan hasil prediksi yang lebih stabil dibandingkan metode lain [12]. *Random Forest* memiliki tingkat akurasi yang lebih tinggi daripada Support Vector Machine (SVM), menurut penelitian perbandingan [11]. Selain itu, lebih presisi daripada algoritma Naive Bayes dan C4.5 [12].

2.4. Teknik Resampling SMOTE

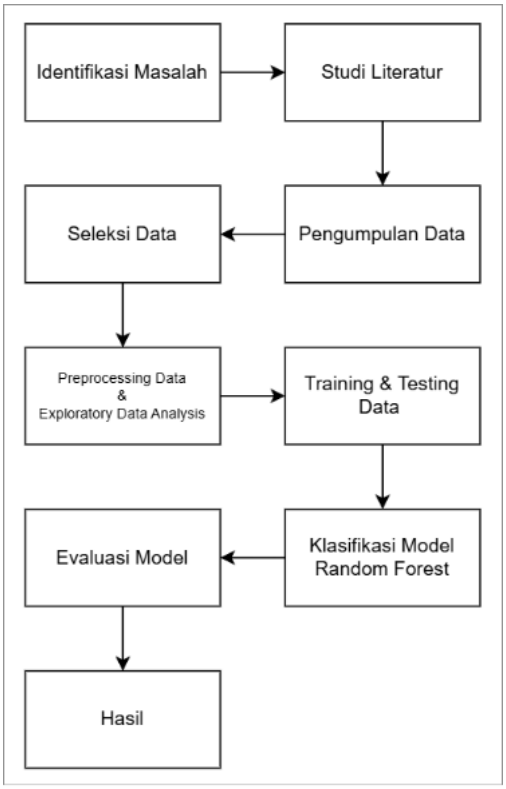
SMOTE menggunakan prinsip interpolasi linear, berbeda dengan metode oversampling konvensional yang hanya menduplikasi data yang ada. Data sintetis baru dibuat dari sampel kelas minoritas yang berdekatan dalam ruang fitur melalui metode ini. Oleh karena itu, SMOTE meningkatkan keragaman sampel pada kelas minoritas tanpa menyebabkan *overfitting*. Dengan demikian, model dapat mempelajari batas keputusan yang lebih representatif [13]. Untuk menjaga sensitivitas model dalam mendeteksi kelas minoritas (seperti kategori "Tidak Sehat"), penerapan teknik ini sangat penting.

2.5. Evaluasi Model

Kinerja model klasifikasi diukur Untuk mengetahui seberapa baik model klasifikasi bekerja menggunakan *Confusion Matrix*. Matriks ini membandingkan prediksi yang benar dengan yang salah. Akurasi (ketepatan global), Presisi, *Recall* (sensitivitas), dan *F1-Score* adalah metrik evaluasi standar yang dihitung berdasarkan matriks tersebut. Metode *K-Fold Cross Validation* memastikan pengujian yang objektif pada berbagai subset data dan menghasilkan kinerja yang kuat [10]**.**

3. METODE PENELITIAN

Penelitian ini menggunakan metodologi ilmu data yang terstruktur dan mencakup sembilan langkah utama.



Gambar 1. Kerangka Alur Tahapan Penelitian

3.1. Sumber dan Pengumpulan Data

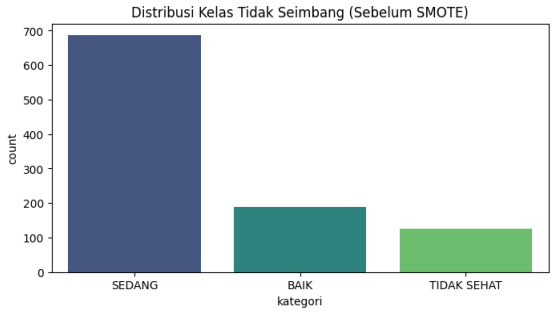
Data sekunder yang digunakan dalam penelitian ini diperoleh dari Dinas Lingkungan Hidup (DLH) Provinsi DKI Jakarta, yang dapat diakses melalui portal Satu Data Jakarta. Rekaman harian, enam parameter polutan (PM_{10}, $PM_{2.5}$, SO_2, CO, O_3, dan NO_2) diperoleh dari Januari hingga Agustus 2025.

3.2. Pra-Pemrosesan Data

Kualitas model sangat bergantung pada tahapan pra-pemrosesan. Proses ini termasuk:

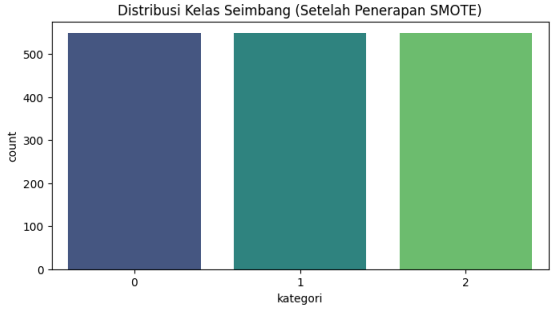
1. **Pembersihan Data,** adalah proses menangani nilai yang tidak ada dengan menggunakan imputasi median dan eliminasi *outliers* ekstrim menggunakan metode *Interquartile Range* (IQR). Pembersihan juga dilakukan pada kategori target "Sangat Tidak Sehat" yang merupakan outlier frekuensi dan "Tidak Ada Data" yang bukan label valid, guna meningkatkan konsistensi data dan kinerja model.
2. ***Feature Engineering,*** juga dikenal sebagai ekstraksi fitur, adalah prosedur temporal untuk mendapatkan informasi tentang hari dan penanda akhir pekan

3. **Transformasi Data,** Transformasi data, proses penerapan Label Encoding label untuk variabel target (kategori ISPU), dan *One-Hot Encoding* untuk fitur lokasi stasiun.
4. ***Scaling,*** Ini adalah prosedur untuk standarisasi fitur numerik dengan menggunakan StandardScaler agar semua variabel memiliki skala yang sama.
5. **Penanganan *Imbalance Data*,** Untuk menyeimbangkan distribusi kelas karena kategori "Sedang" mendominasi data asli, diterapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) pada data latih.



Gambar 2. Distribusi Kategori ISPU Sebelum Penerapan SMOTE

Analisis distribusi data awal menunjukkan ketidakseimbangan kelas yang signifikan, dengan dominasi kategori 'Sedang' yang berpotensi memicu bias model *Gambar 2*. Untuk mengatasi hal tersebut, teknik *Synthetic Minority Over-sampling Technique* (SMOTE) diterapkan guna menyeimbangkan proporsi sampel antar-kelas melalui pembentukan data sintetis.



Gambar 3. Distribusi Kategori ISPU Setelah Penerapan SMOTE

Hasilnya, distribusi data menjadi seimbang pada *Gambar 3*, memastikan model dapat mempelajari karakteristik seluruh kategori secara objektif tanpa mengabaikan kelas minoritas.

3.3. Pemodelan dan Evaluasi

Model diimplementasikan menggunakan bahasa pemrograman Python dan pustaka Scikit-Learn. Untuk memastikan distribusi kelas target konsisten pada kedua subset data, parameter *stratify* digunakan untuk membagi dataset menjadi proporsi 80:20 untuk data latih dan data uji. Dengan *random_state*=42, teknik SMOTE diterapkan pada data latih untuk mengatasi ketidakseimbangan data.

Untuk menjamin reprodusibilitas hasil eksperimen secara penuh pada seluruh tahapan stokastik, model klasifikasi dibangun menggunakan algoritma Random Forest. Konfigurasi parameter defaultnya adalah sebagai berikut (*n_estimators*=100, *criterion*='gini,' dan *random_state)* ditetapkan pada nilai 42.

Untuk menguji konsistensi, metode validasi *K-Fold Cross Validation* (*K=5*) digunakan untuk mengevaluasi kinerja secara menyeluruh. Pada data uji, metrik klasifikasi standar (Akurasi, Presisi, *Recall*, *F1-Score*), serta *Confusion Matrix* digunakan.

4. HASIL DAN PEMBAHASAN

4.1. Analisis Pola Data Polutan

Eksplorasi data mengungkapkan karakteristik menarik pada polusi Jakarta. Analisis korelasi menunjukkan bahwa hubungan linear terkuat terjadi antara konsentrasi PM_{10} dan $PM_{2.5}$ dengan koefisien korelasi sebesar 0,62. Tingginya korelasi positif ini mengindikasikan bahwa peningkatan partikel udara kasar (PM_{10}) cenderung diikuti oleh peningkatan partikel halus ($PM_{2.5}$), yang menunjukkan kemungkinan besar kedua polutan ini berasal dari sumber emisi yang sama atau memiliki pola penyebaran yang serupa. Selain itu, ditemukan anomali pola berdasarkan waktu di mana polutan hasil emisi kendaraan (CO dan NO_2) cenderung lebih tinggi pada hari kerja (*weekday*), sementara polutan partikulat (PM_{10}, $PM_{2.5}$) dan Ozon justru mengalami peningkatan rata-rata pada akhir pekan (*weekend*).

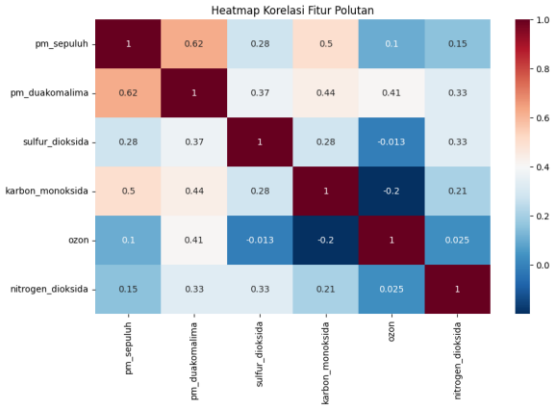
Tabel 1. Statistik Deskriptif Fitur Polutan

Statistik	pm_sepuluh	pm_duakomalima	sulfur_dioksida

count	1.001000e+03	1.001000e+03	1.001000e+03
mean	1.987532e-16	-5.678663e-17	1.135733e-16
std	1.000500e+00	1.000500e+00	1.000500e+00
min	-2.507857e+00	-2.353872e+00	-2.591238e+00
25%	-7.268465e-01	-6.306645e-01	-6.826811e-01
50%	7.153735e-02	1.052909e-02	-7.997896e-02
75%	6.856788e-01	6.917973e-01	6.231736e-01
max	2.589517e+00	2.735602e+00	2.531731e+00

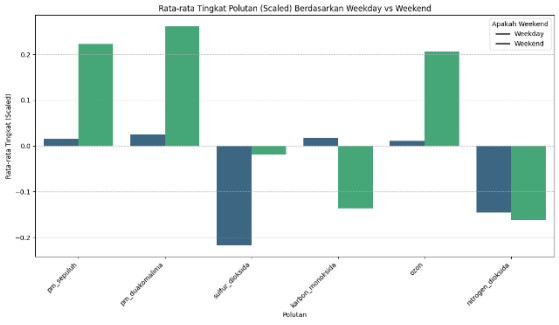
Statistik	karbon_monoksida	ozon	nitrogen_dioksida
count	**1.001000e+03**	1.001000e+03	1.001000e+03
mean	**4.258997e-17**	1.987532e-16	-1.419666e-16
std	**1.000500e+00**	1.000500e+00	1.000500e+00
min	**-2.359636e+00**	-2.288517e+00	-1.633742e+00
25%	**-7.019497e-01**	-7.836358e-01	-8.287661e-01
50%	**-8.031746e-02**	-1.566019e-01	-2.250341e-01
75%	**5.413148e-01**	7.212456e-01	6.470233e-01
max	**2.820633e+00**	2.978568e+00	3.061951e+00

Setelah data melalui tahap scaling, terlihat perbedaan karakteristik yang jelas pada distribusinya. Parameter partikulat PM_{10} dan $PM_{2.5}$ menunjukkan pola sebaran yang rapi dan simetris. Sebaliknya, polutan jenis gas justru memperlihatkan kecenderungan data yang miring ke kanan (*positive skew*). Perbedaan perilaku data ini kemudian dipetakan lebih lanjut untuk melihat interaksi antar-variabelnya, sebagaimana tersaji dalam matriks korelasi pada *Gambar 4.*



Gambar 4. Heatmap Matriks Korelasi Antar Variabel Polutan

Matriks korelasi memperlihatkan fakta menarik, di mana PM_{10} dan $PM_{2.5}$ memiliki hubungan linear yang paling menonjol dengan koefisien 0,62. Eratnya hubungan kedua jenis debu ini mengisyaratkan bahwa mereka berbagi sumber asal atau dipengaruhi kondisi atmosfer yang sama. Temuan ini memberikan indikasi penting saat dikaitkan dengan pola waktu pada *Gambar 5*, yang mengungkap adanya anomali unik. Meskipun gas emisi kendaraan CO dan NO_2 memuncak di hari kerja, polutan partikulat dan Ozon justru mencatat rata-rata tertinggi saat akhir pekan. Fenomena tersebut mengindikasikan bahwa masalah polusi di Jakarta sangat kompleks dan tidak bisa disederhanakan hanya sebagai akibat dari volume kendaraan semata.



Gambar 5. Rata-rata Polutan Weekday vs Weekend

4.2. Kinerja Model Klasifikasi Random Forest

```
Skor setiap fold: [1. 1. 1. 1. 1.]
Skor rata-rata: 100.00%
```

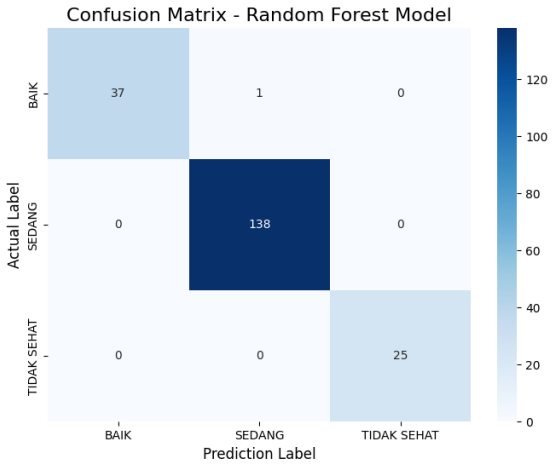
Gambar 6. Hasil Evaluasi K-Fold Cross Validation

Evaluasi kinerja model menunjukkan hasil yang sangat presisi. Validasi internal menggunakan *K-Fold Cross Validation* dengan 5 *fold* menghasilkan akurasi sempurna dengan skor 100% di setiap lipatan pengujian, menandakan stabilitas model yang tinggi. Pada pengujian terhadap data testing, model mencapai akurasi keseluruhan sebesar 99,50%. Rincian performa untuk setiap kelas disajikan dalam *Tabel* 2.

Tabel 2. Hasil Evaluasi Kinerja Model

	Precision	***Recall***	***F1-Score***	***Support***
0	1.0000	0.9737	0.9867	38
1	0.9928	1.0000	0.9964	138
2	1.0000	1.0000	1.0000	25
Accuracy			0.9950	201
Macro Avg	0.9976	0.9912	0.9944	201
Weighted Avg	0.9951	0.9950	0.9950	201

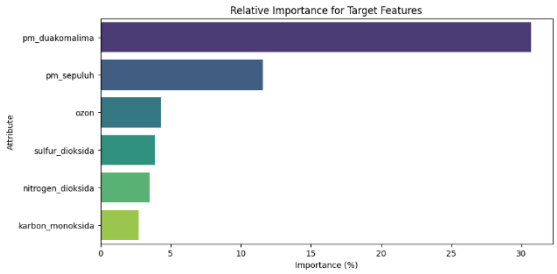
Kemampuan model membedakan antar-kelas divisualisasikan melalui *Confusion Matrix* pada Gambar 7. Dari total 201 data uji, model hanya melakukan satu kesalahan prediksi (misklasifikasi), membuktikan efektivitas algoritma dalam menangani klasifikasi multi-kelas pada data polusi udara yang kompleks



Gambar 7. Confusion Matrix Hasil Klasifikasi Model Random Forest

4.3. Identifikasi Polutan Paling Berpengaruh

Analisis *feature importance* dilakukan untuk mengkuantifikasi kontribusi setiap variabel polutan terhadap keputusan klasifikasi model.



Gambar 8. Visualisasi Feature Importance

Sebagaimana terlihat pada Gambar 8, terungkap bahwa $PM_{2.5}$ memiliki pengaruh terbesar dalam menyumbang lebih dari 30% terhadap total kontribusi. Temuan ini selaras dengan karakteristik data fisik dan menegaskan urgensi pengendalian emisi partikulat halus di Jakarta.

4.4. Pembahasan dan Implikasi

4.4.1. Interpretasi Performa Model dan Posisi Penelitian

Tingkat akurasi model Random Forest yang mencapai 99,50% pada data pengujian, didukung oleh skor sempurna pada validasi internal, memberikan bukti empiris yang kuat mengenai stabilitas metodologi yang diterapkan. Kinerja ini tidak hanya menegaskan keberhasilan teknik SMOTE dalam mengatasi ketimpangan data, tetapi juga menunjukkan keunggulan pendekatan

klasifikasi prediktif dibandingkan metode deskriptif murni seperti *clustering* yang dilakukan pada penelitian sebelumnya oleh Rahmadenti (2025) [3].

Secara komparatif, capaian akurasi model ini sebesar 99,50% menunjukkan daya saing yang tinggi jika disandingkan dengan studi serupa, seperti penelitian Firdaus et al. (2024) yang mencatatkan akurasi 99,95% [4]. Meskipun secara numerik terdapat selisih tipis, hasil penelitian ini memiliki nilai urgensi yang lebih relevan karena dilatih menggunakan dataset tahun 2025. Berbeda dengan data periode 2016-2021 yang digunakan pada studi terdahulu, data tahun 2025 merepresentasikan dinamika atmosfer pasca-pandemi yang memiliki variabilitas polutan lebih tinggi akibat pemulihan penuh aktivitas ekonomi dan mobilitas warga Jakarta. Model menunjukkan kemampuan dalam mempertahankan akurasi di atas 99% pada dataset yang lebih fluktuatif ini menegaskan bahwa arsitektur Random Forest yang diusulkan memiliki generalisasi yang sangat baik (*robust*), tidak sekadar menghafal pola data lama.

4.4.2. Dominasi $PM_{2.5}$ dan Implikasi Kebijakan

Tingginya kontribusi $PM_{2.5}$ (30,68%) dalam model klasifikasi juga selaras dengan temuan anomali pola waktu yang terdeteksi pada tahap analisis data. Sebagaimana dipaparkan pada bagian hasil, terdapat fenomena menarik di mana konsentrasi gas buang kendaraan (CO dan NO_2) menurun saat akhir pekan (*weekend*), namun konsentrasi partikulat ($PM_{2.5}$ dan PM_{10}) justru mengalami peningkatan rata-rata. Hal ini memberikan wawasan baru bahwa sumber polusi $PM_{2.5}$ di Jakarta tidak tunggal dan tidak sepenuhnya bergantung pada volume kendaraan bermotor harian. Dominasi $PM_{2.5}$ saat akhir pekan mengindikasikan adanya sumber emisi persisten lain. Seperti debu konstruksi, residu industri, atau transportasi logistik alat berat yang tetap beroperasi saat mobilitas warga berkurang. Oleh karena itu, keputusan model untuk menempatkan $PM_{2.5}$ sebagai fitur terpenting adalah langkah yang sangat logis secara ilmiah, mengingat parameter ini adalah polutan yang paling 'stabil' tingginya dan paling sulit turun meskipun aktivitas lalu lintas sedang lengang.

Temuan ini mengindikasikan bahwa strategi mitigasi polusi di masa depan harus bergeser dari pendekatan umum menjadi pengendalian yang terfokus secara agresif pada sumber-sumber emisi partikulat (seperti debu konstruksi atau pembakaran terbuka) untuk mencapai perbaikan status ISPU yang signifikan. Model klasifikasi yang dihasilkan dapat diintegrasikan sebagai dasar sistem peringatan dini yang proaktif.

5. KESIMPULAN

- Penerapan algoritma Random Forest dengan teknik *preprocessing* SMOTE terbukti sangat efektif untuk mengklasifikasikan kualitas udara Jakarta ke dalam tiga kategori utama (Baik, Sedang, Tidak Sehat). Model yang dikembangkan menunjukkan performa yang sangat *robust* dan stabil, dengan pencapaian akurasi pengujian sebesar 99,50%. Selain itu, analisis feature importance berhasil mengidentifikasi $PM_{2.5}$ sebagai parameter polutan paling dominan dengan kontribusi relatif sebesar 30,68%, menjadikannya indikator utama dalam penentuan status ISPU dibandingkan polutan lainnya.
- Keunggulan utama penelitian ini terletak pada penggunaan teknik penanganan *imbalance* data (SMOTE) yang membuat model yang objektif dan tidak bias terhadap kelas mayoritas. Model ini juga memiliki tingkat interpretabilitas yang baik melalui pemeringkatan fitur, memberikan wawasan yang lebih dalam dibandingkan metode klasifikasi konvensional.
- Keterbatasan penelitian ini terletak pada cakupan data yang hanya terbatas pada periode Januari hingga Agustus 2025, sehingga belum sepenuhnya menangkap variasi pola tahunan atau musiman jangka panjang. Selain itu, model ini hanya menggunakan variabel konsentrasi polutan dan belum mengintegrasikan faktor meteorologi eksternal.

- Untuk penelitian di masa depan, disarankan untuk menambahkan variabel data cuaca (seperti curah hujan, suhu, dan kecepatan angin) guna meningkatkan reliabilitas prediksi. Perluasan rentang waktu data dan komparasi dengan algoritma Deep Learning juga direkomendasikan untuk menguji batas performa model pada dataset yang lebih kompleks.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Bapak Fuad Nur Hasan, M.Kom (Universitas Bina Sarana Informatika) atas bimbingannya, Dinas Lingkungan Hidup DKI Jakarta atas penyediaan data ISPU melalui portal Satu Data Jakarta, serta seluruh rekan Kelompok 4 (empat) atas kerja sama dalam penyelesaian penelitian ini.

DAFTAR PUSTAKA

[1] Joko Sapto Pramono, Nuraini, Junardin Djamaluddin, Yoanita Hijriyati, and Yusriati, "The Effect of Air Pollution on the Health of Urban Residents (Case Study in Jakarta)," *Miracle Get Journal*, vol. 2, no. 2, pp. 34–43, May 2025, doi: 10.69855/mgj.v2i2.125.

[2] Z. Majidah, M. Ari Bianto, and B. Dwi Saputra, "Implementasi Fuzzy Logic Mamdani Untuk Monitoring Kualitas Udara Berbasis Iot," *Jurnal Pengembangan Teknologi Informasi dan Komunikasi (JUPTIK)*, vol. 2, no. 1, pp. 1–6, Jun. 2024, doi: 10.52060/juptik.v2i1.2091.

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

[4] R. Firdaus *et al.*, "Implementasi Algoritma Random Forest Untuk Klasifikasi Pencemaran Udara di Wilayah Jakarta Berdasarkan Jakarta Open Data," *JURNAL FASILKOM*, vol. 14, no. 2.

[5] H. Hairani, A. Anggrawan, and D. Priyanto, "Improvement Performance of the Random Forest Method on Unbalanced Diabetes Data Classification Using Smote-Tomek Link," *INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION*, vol. 7, no. 1, Mar. 2023, [Online]. Available: www.joiv.org/index.php/joiv

[6] A. J. Barid, Hadiyanto, and A. Wibowo, "Optimization of the algorithms use ensemble and synthetic minority oversampling technique for air quality classification," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, no. 3, pp. 1632–1640, Mar. 2024, doi: 10.11591/ijeecs.v33.i3.pp1632-1640.

[7] T. Handhayani, "An integrated analysis of air pollution and meteorological conditions in Jakarta," *Sci Rep*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-32817-9.

[8] A. Ghaida, F. M. Firdaus, K. M. Qatrunnada, D. Peters, B. Cardenas, and P. Lestari, "Spatial patterns of PM2.5 air pollution in Jakarta: insights from mobile monitoring," in *E3S Web of Conferences*, EDP Sciences, Feb. 2024. doi: 10.1051/e3sconf/202448506002.

[9] D. Septiyana, A. Sukmono, and M. A. Yusuf, "Pemantauan Kualitas Udara Ispu (Pm10, So2, No2) Menggunakan Citra Landsat 8 Dan 9 Untuk Kecamatan Mijen Selama Pandemi Covid-19," *Jurnal Geodesi Undip*, vol. 12, no. 3, Jul. 2023.

[10] D. Dzaky Daniswara, A. Terza Damaliana, I. Gede Susrama Mas Diyasa UPN, J. Timur JlRaya Rungkut Madya No, and G. Anyar, "Pengukuran Indeks Standar Pencemaran Udara Menggunakan Support Vector Machine," *JURNAL PENELITIAN Politeknik Penerbangan Surabaya*, vol. 9, no. 1, Apr. 2024.

[11] N. Adityo, A. Ibnu, and F. A. W. Yanuar, "Klasifikasi Tingkat Kualitas Udara DKI Jakarta Berdasarkan Open Government Data Menggunakan Algoritma Random Forest," *e-Proceeding of Engineering*, vol. 10, no. 2, 2023.

[12] A. R. Kannajmi and D. Saputra, "Penentuan Model Algoritma Klasifikasi Terbaik Untuk Klasifikasi Kualitas Udara Di Jakarta 2023," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5664.

[13] Anisa Ma'u Luthfi and Fatkhurokhman Fauzi, "Perbandingan Klasifikasi Random Forest, Support Vector Machines, dan LGBM Pada Klasifikasi Kualitas Udara di Jakarta," *JUSTINDO (Jurnal Sistem dan Teknologi Informasi Indonesia)*, vol. 9, no. 2, pp. 99–108, Aug. 2024, doi: 10.32528/justindo.v9i2.1912.

[14] D. P. Ramadhan and A. Triayudi, "Jakarta Air Quality Classification Based on Air Pollutant Standard Index Using C4.5 And Naïve Bayes Algorithms," *Journal of Technology and Information Systems*, vol. 2, no. 4, Nov. 2024, doi: 10.58905/SAGA.v2i4.395.

[15] M. Oumoulylte, A. El Allaoui, Y. Farhaoui, and A. A. Boughrous, "Efficient Air Quality Prediction Models Based on Supervised

Machine Learning Techniques," in *E3S Web of Conferences*, EDP Sciences, Jun. 2025. doi: 10.1051/e3sconf/202563202012.