

BAB I

PENDAHULUAN

1.1. Latar Belakang

Pencemaran udara telah menjadi ancaman lingkungan global yang sangat berbahaya di masa kini. Laporan kualitas udara dunia memposisikan Indonesia sebagai negara ke-17 dengan polusi udara terparah pada tahun 2025, dengan rata-rata konsentrasi $PM_{2.5}$ (partikel kecil dengan diameter 2,5 mikron) sebesar $30,0 \mu g/m^3$ (IQAir, 2026). Angka ini jauh melebihi hingga 7 kali lebih besar dibandingkan dengan standar kualitas udara global yang telah ditentukan oleh *World Health Organization* (WHO), yaitu sebesar $5 \mu g/m^3$ (IQAir, 2026). Tingginya tingkat polusi ini didorong oleh berbagai faktor, di antaranya yaitu industrialisasi, emisi dari sektor transportasi, ketergantungan pada bahan bakar fosil, dan pembakaran pada limbah dan sampah (Candrasari et al., 2023).

Pencemaran udara yang terjadi memberikan dampak buruk yang signifikan pada kesehatan masyarakat dan bersifat merusak. Paparan polusi ini bisa menyebabkan berbagai gangguan pernapasan, baik yang bersifat akut maupun kronis seperti asma, bronkitis, Infeksi Saluran Pernafasan Akut (ISPA), hingga penyakit jantung, kanker paru-paru, dan kematian dini (Candrasari et al., 2023). Menurut Sudaryanto et al. (2022), dampak polusi udara secara global sangat serius karena setiap tahunnya menyebabkan sedikitnya 3,7 juta kematian dini di seluruh dunia akibat paparan udara tercemar.

Sebagai pusat pemerintahan dan ekonomi Indonesia, Provinsi DKI Jakarta menghadapi tantangan pencemaran udara yang krusial. Data pemantauan kualitas udara secara konsisten menempatkan Jakarta sebagai salah satu kota dengan Indeks Standar Pencemaran Udara (ISPU) tertinggi di Indonesia. Pada September 2025, nilai ISPU Jakarta tercatat mencapai 113, sebuah angka yang masuk dalam kategori "Tidak Sehat" (Fadhlurrahman, 2026). Berdasarkan laporan dari Kementerian Kesehatan Republik Indonesia (2024) diperkirakan bahwa lebih dari 57% kasus ISPA di Jakarta berhubungan langsung dengan paparan polutan udara ambien, terutama dari $PM_{2.5}$ (partikel halus 2.5 mikron) yang mampu menembus sistem pernapasan bagian dalam.

Indeks Standar Pencemaran Udara (ISPU) menjadi salah satu indikator penting yang digunakan pemerintah untuk mengukur tingkat kualitas udara berdasarkan parameter utama seperti PM_{10} (partikel halus 10 mikron), $PM_{2.5}$ (partikel halus 2.5 mikron), SO_2 (Sulfur Dioksida), CO (Karbon Monoksida), NO_2 (Nitrogen Dioksida), dan O_3 (Ozon). Nilai parameter tertinggi akan menjadi penentu kategori ISPU pada waktu tertentu yang terdiri dari kategori “baik”, “sedang”, “tidak sehat”, “sangat tidak sehat”, dan “berbahaya” (KLHK Republik Indonesia, 2020).

Sistem pemantauan kualitas udara yang ada saat ini sudah cukup baik dalam menyajikan data secara deskriptif, namun belum didukung oleh model analitis yang kuat untuk memahami pola kompleks di dalam data tersebut (Fitri & Saryoko, 2024). Penentuan kategori ISPU yang didasarkan pada nilai parameter pencemar tertinggi belum sepenuhnya mampu merepresentasikan kontribusi relatif antar parameter pencemar secara komprehensif, terutama pada data historis yang bersifat besar dan kompleks (KLHK Republik Indonesia, 2020). Kondisi ini menyebabkan upaya mitigasi pencemaran udara masih cenderung bersifat responsif, sehingga langkah pencegahan dini terhadap potensi memburuknya kualitas udara sulit untuk dilakukan secara optimal (Candrasari et al., 2023).

Pendekatan berbasis *machine learning* menawarkan kemampuan untuk mempelajari hubungan non linier antar variabel, melakukan klasifikasi suatu kategori secara lebih akurat, serta menghasilkan model yang dapat digunakan sebagai alat pendukung pengambilan keputusan (Fitri & Saryoko, 2024). Oleh karena itu, penerapan algoritma klasifikasi berbasis *machine learning* dapat menjadi solusi untuk meningkatkan pemanfaatan data ISPU yang tidak hanya sebagai alat pemantauan, tetapi juga sebagai sarana analisis prediktif kualitas udara.

Sejumlah penelitian terdahulu telah membuktikan efektivitas pemanfaatan *machine learning* dalam menganalisis dan mengklasifikasikan kualitas udara berdasarkan nilai ISPU seperti yang dilakukan oleh Fitri & Saryoko (2024) dalam penelitiannya yang berjudul *Determining Air Quality Influential Parameters Using Machine Learning Techniques*. Hasil penelitian tersebut menunjukkan bahwa pendekatan berbasis *machine learning* mampu menangkap hubungan non linier

antar parameter pencemar seperti $PM_{2.5}$, CO, dan NO_2 secara lebih akurat dibanding metode konvensional statistik.

Dalam kasus yang sama, Luthfi & Fauzi (2024) dalam *Jurnal Sistem dan Teknologi Informasi Indonesia (JUSTINDO)* melakukan perbandingan kinerja beberapa algoritma klasifikasi dalam mengklasifikasikan kualitas udara di Jakarta. Hasilnya menunjukkan bahwa algoritma *Random Forest* memiliki tingkat akurasi tertinggi sebesar 95,67%, mengungguli algoritma pembanding lainnya yaitu SVM dengan 90,12% dan LGBM dengan 93,40%. Tingginya tingkat akurasi yang dihasilkan algoritma *Random Forest* juga dibuktikan oleh Manik et al. (2026) yang melakukan optimasi pada algoritma *Random Forest* dalam klasifikasi data ISPU di beberapa wilayah kota Indonesia dan menghasilkan tingkat akurasi hingga 96,02%.

Penelitian yang dilakukan oleh Sinaga et al. (2025) dalam *Jurnal Teknik Informatika (JATI) ITN Malang* mencoba mengombinasikan algoritma *Random Forest* dengan *Synthetic Minority Oversampling Technique (SMOTE)* untuk menangani ketidakseimbangan data dalam klasifikasi penyakit jantung dan berhasil meningkatkan akurasi model dari 92% menjadi 96%, disertai peningkatan signifikan pada nilai *recall* kelas minoritas. Hasil ini menunjukkan bahwa penerapan *SMOTE* mampu menyeimbangkan distribusi data dan meningkatkan sensitivitas model terhadap kelas dengan jumlah sampel rendah.

Dari hasil-hasil penelitian terdahulu tersebut menunjukkan bahwa algoritma *Random Forest* memiliki sejumlah keunggulan dibanding algoritma lain, antara lain kemampuannya untuk menangani hubungan non linier antar variabel, kestabilannya terhadap *noise*, kemampuannya dalam memproses data berdimensi tinggi, serta mampu mengurangi risiko *overfitting* (Krisbiantoro et al., 2024). Selain itu, *Random Forest* memberikan nilai penting dalam interpretabilitas fitur melalui pengukuran *feature importance*, yang membantu mengidentifikasi parameter pencemar udara paling signifikan dalam menentukan kategori ISPU.

Meskipun beberapa penelitian telah menunjukkan tingginya akurasi yang dihasilkan oleh algoritma *Random Forest* dalam melakukan klasifikasi terhadap data ISPU, namun evaluasi tersebut belum mempertimbangkan adanya distribusi kelas yang tidak seimbang (*imbalanced*). Fokus evaluasi yang hanya berpatokan pada akurasi sering kali mengabaikan tingkat sensitivitas model terhadap kelas

minoritas, yang mana kategori yang memiliki urgensi kesehatan tinggi seperti kategori 'Tidak Sehat' justru memiliki jumlah sampel yang jauh lebih sedikit dibandingkan kategori lainnya.

Oleh karena itu, penelitian ini mengadopsi pendekatan algoritma *Random Forest* dalam klasifikasi Indeks Standar Pencemaran Udara (ISPU) di Provinsi DKI Jakarta dengan mengombinasikan teknik *SMOTE* sebagai upaya mengatasi ketidakseimbangan data guna memaksimalkan tingkat akurasi yang dihasilkan. Pendekatan ini diharapkan dapat menghasilkan model klasifikasi yang lebih akurat, sehingga tidak hanya berfungsi sebagai sistem pemantauan kualitas udara yang deskriptif, tetapi juga sebagai sistem analisis prediktif yang dapat mendukung pengambilan keputusan berbasis data dalam upaya mitigasi pencemaran udara.

1.2. Identifikasi Masalah

Dari latar belakang yang sudah diuraikan, maka diidentifikasi beberapa permasalahan yang menjadi fokus dalam penelitian ini, di antaranya sebagai berikut:

1. Sistem klasifikasi ISPU yang ada masih bersifat deskriptif dan belum mampu menangkap pola hubungan kompleks antar parameter pencemar udara.
2. Belum adanya penerapan model prediktif berbasis *machine learning* yang secara efektif mempertimbangkan distribusi data yang tidak seimbang dalam dataset kualitas udara di Provinsi DKI Jakarta.
3. Keterbatasan penelitian terdahulu yang belum secara spesifik mendalami penggunaan algoritma *Random Forest* untuk meningkatkan akurasi dan sensitivitas klasifikasi ISPU.
4. Kebutuhan akan model prediktif yang lebih akurat dan andal guna mendukung kebijakan mitigasi pencemaran udara di Provinsi DKI Jakarta berbasis analisis data historis yang kompleks.

1.3. Batasan Masalah

Berdasarkan permasalahan yang telah diidentifikasi, ditetapkan batasan masalah demi menjaga fokus penelitian agar tetap terarah dan tidak meluas dari tujuan utamanya, di antaranya yaitu:

1. Penelitian ini hanya berfokus pada penerapan algoritma *Random Forest* sebagai metode utama klasifikasi ISPU. Teknik *SMOTE (Synthetic Minority Oversampling Technique)* digunakan sebagai strategi penyeimbang distribusi kelas untuk mengatasi masalah *imbalanced* pada dataset.
2. Penelitian ini terbatas pada wilayah administratif Provinsi DKI Jakarta dengan memanfaatkan data historis ISPU Provinsi DKI Jakarta periode bulan Januari 2024 hingga November 2025.
3. Variabel penelitian terbatas pada parameter-parameter pencemar udara yang digunakan dalam perhitungan ISPU yaitu *Particulate Matter 10 Microns (PM₁₀)*, *Particulate Matter 2.5 Microns (PM_{2.5})*, Sulfur Dioksida (SO₂), Karbon Monoksida (CO), Nitrogen Dioksida (NO₂), dan Ozon (O₃).
4. Analisis penelitian ini terbatas pada tahapan pengolahan data, penerapan model klasifikasi, dan evaluasi performa model berdasarkan metrik yang dihasilkan dari *Confusion Matrix* seperti *Accuracy*, *Precision*, *Recall*, dan *F1-score*.

1.4. Rumusan Masalah

Adapun rumusan masalah dari penelitian ini yaitu: Bagaimana hasil performa model *Random Forest* dalam mengklasifikasikan kategori Indeks Standar Pencemaran Udara (ISPU) di Provinsi DKI Jakarta?

1.5. Tujuan Penelitian

Adapun tujuan yang ingin dicapai dari penelitian ini berdasarkan rumusan masalah yang dipaparkan yaitu menghasilkan analisis hasil performa model *Random Forest* dalam mengklasifikasikan kategori Indeks Standar Pencemaran Udara (ISPU) di Provinsi DKI Jakarta.

1.6. Manfaat Penelitian

Hasil dari penelitian ini diharapkan mampu memberikan manfaat, baik itu dari sisi teoritis maupun dari sisi praktis.

1.6.1. Manfaat Teoritis

Secara teoritis, penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang *machine learning*, khususnya dalam penerapan algoritma *Random Forest* untuk klasifikasi data lingkungan yang bersifat kompleks dan tidak seimbang.
2. Memperluas pemahaman mengenai efektivitas penerapan teknik *Synthetic Minority Oversampling Technique (SMOTE)* dalam meningkatkan akurasi dan sensitivitas model klasifikasi pada data kualitas udara.
3. Menjadi referensi akademik bagi penelitian lanjutan dalam penggunaan model prediktif untuk data lingkungan perkotaan seperti di Provinsi DKI Jakarta.

1.6.2. Manfaat Praktis

Secara praktis, hasil dari penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan dukungan analitis bagi instansi seperti Kementerian Lingkungan Hidup dan Kehutanan (KLHK) serta Dinas Lingkungan Hidup (DLH) Provinsi DKI Jakarta dalam melakukan evaluasi dan prediksi kualitas udara secara lebih akurat.
2. Menyediakan dasar bagi pengambilan keputusan berbasis data dalam mitigasi pencemaran udara serta perumusan kebijakan pengendalian emisi di wilayah perkotaan.
3. Menjadi acuan pengembangan sistem pemantauan udara cerdas berbasis model prediktif untuk mendukung pengawasan lingkungan di masa depan.