

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Indeks Standar Pencemar Udara (ISPU) merupakan sistem indeks yang digunakan di Indonesia untuk menggambarkan kondisi mutu udara ambien berdasarkan dampaknya terhadap kesehatan manusia, nilai estetika, dan makhluk hidup lainnya. Melalui penyajian dalam bentuk indeks, informasi kualitas udara yang diperoleh dari hasil pengukuran berbagai parameter pencemar dapat disampaikan dalam bentuk yang lebih mudah dipahami oleh masyarakat maupun pemangku kebijakan [1]. Keberadaan informasi kualitas udara menjadi penting karena pencemaran udara masih berkontribusi terhadap berbagai permasalahan kesehatan di berbagai wilayah dunia, di mana laporan *State of Global Air 2024* menunjukkan bahwa pencemaran udara berkontribusi terhadap sekitar 8,1 juta kematian pada tahun 2021 dan menjadi faktor risiko kematian terbesar kedua secara global [2]. Temuan tersebut menunjukkan bahwa kualitas udara tidak hanya berkaitan dengan kondisi lingkungan, tetapi juga memiliki hubungan yang erat dengan kesehatan masyarakat. Penelitian lain menunjukkan bahwa paparan jangka pendek terhadap pencemaran udara diperkirakan berkontribusi terhadap sekitar 1 juta kematian prematur setiap tahun atau setara dengan 2,08% dari total kematian global [3]. Selain berdampak pada kesehatan, hasil pemantauan kualitas udara pada berbagai wilayah perkotaan juga menunjukkan bahwa konsentrasi pencemar udara pada banyak periode pengamatan masih melampaui nilai pedoman kualitas udara yang direkomendasikan sehingga memerlukan pemantauan yang dilakukan secara berkelanjutan [4]. Oleh karena itu, penyajian informasi kualitas udara melalui ISPU menjadi salah satu sarana yang digunakan untuk menggambarkan kondisi mutu udara secara lebih terstruktur berdasarkan hasil pemantauan yang dilakukan secara berkala [1].

DKI Jakarta merupakan kawasan metropolitan dengan tekanan pencemaran udara yang berkaitan erat dengan aktivitas transportasi, kepadatan penduduk, dan kegiatan perkotaan. Hasil pemantauan kualitas udara periode 2014–2018

menunjukkan bahwa konsentrasi PM<sub>10</sub> di DKI Jakarta berada pada rentang maksimum 158–206  $\mu\text{g}/\text{m}^3$  dan telah melewati nilai ambang batas 150  $\mu\text{g}/\text{m}^3$ , sedangkan O<sub>3</sub> berada pada rentang maksimum 249–456,10  $\mu\text{g}/\text{m}^3$  dan melewati nilai ambang batas 235  $\mu\text{g}/\text{m}^3$ . Kondisi kualitas udara juga terlihat dari kategori ISPU, di mana tahun 2018 tercatat sebagai periode dengan akumulasi 381 kejadian kategori tidak sehat dan 30 kejadian kategori sangat tidak sehat pada lima stasiun pemantauan kualitas udara di DKI Jakarta [5]. Tekanan tersebut diperkuat oleh kondisi transportasi, karena jumlah kendaraan di DKI Jakarta pada tahun 2020 mencapai 21.347.312 unit atau sekitar dua kali jumlah penduduknya, dengan kontribusi emisi kendaraan diperkirakan sebesar 60–70% terhadap pencemaran udara [6]. Penelitian lain juga menunjukkan bahwa sektor transportasi menyumbang sekitar 75% dari total emisi di Jakarta, sehingga aktivitas kendaraan menjadi salah satu faktor yang berperan besar terhadap penurunan kualitas udara perkotaan [7]. Dampak pencemaran udara di Jakarta tidak hanya berkaitan dengan lingkungan, tetapi juga dengan kesehatan masyarakat, karena polusi udara dikaitkan dengan lebih dari 10.000 kematian, lebih dari 5.000 rawat inap, dan biaya kesehatan sekitar USD 2,94 miliar per tahun [8]. Berdasarkan analisis data ISPU periode 2022–2025, kualitas udara Jakarta menunjukkan perbedaan pola antarstasiun dan antarperiode melalui tiga kelompok kualitas udara, yaitu baik, sedang, dan tidak sehat, yang menunjukkan bahwa kondisi kualitas udara berubah pada waktu dan lokasi pengamatan yang berbeda [9].

Indeks Standar Pencemar Udara (ISPU) digunakan sebagai indikator yang menggambarkan kondisi kualitas udara berdasarkan hasil pengukuran beberapa parameter pencemar udara yang telah ditetapkan dalam regulasi nasional. Parameter tersebut meliputi PM<sub>10</sub>, PM<sub>2.5</sub>, SO<sub>2</sub>, CO, O<sub>3</sub>, dan NO<sub>2</sub> yang dapat menunjukkan tingkat pencemaran udara dari sumber dan karakteristik yang berbeda sehingga menghasilkan nilai konsentrasi yang bervariasi pada setiap periode pengamatan [1]. Dalam sistem perhitungan ISPU, setiap parameter pencemar terlebih dahulu dikonversi ke dalam nilai indeks sesuai dengan ketentuan yang berlaku sehingga dapat dibandingkan dalam satu skala yang sama [10]. Nilai ISPU harian kemudian ditentukan berdasarkan parameter pencemar yang menghasilkan

nilai indeks tertinggi, sehingga perubahan konsentrasi pada parameter dominan akan menjadi faktor utama yang memengaruhi kategori kualitas udara yang ditampilkan pada hari tersebut [1]. Dengan mekanisme tersebut, berbagai informasi pencemar udara yang berasal dari beberapa parameter dapat direpresentasikan menjadi satu nilai indeks yang menggambarkan kondisi kualitas udara secara umum [11]. Pendekatan ini memungkinkan kualitas udara disajikan dalam bentuk informasi yang lebih sederhana tanpa menghilangkan representasi kondisi pencemaran yang terjadi pada periode pengamatan tertentu [10]. Oleh karena itu, penggunaan nilai ISPU sebagai representasi kondisi kualitas udara memberikan dasar yang terstruktur untuk menganalisis perubahan kualitas udara dari waktu ke waktu berdasarkan satu deret data yang konsisten [1].

Nilai ISPU yang diperoleh dari hasil pemantauan kualitas udara dicatat secara berkala pada setiap periode pengamatan sehingga membentuk rangkaian data berdasarkan urutan waktu [12]. Dalam praktiknya, nilai ISPU tidak selalu berada pada kondisi yang sama, melainkan dapat mengalami perubahan pada setiap periode pengamatan sesuai dengan kondisi lingkungan yang terjadi [13]. Penelitian mengenai kualitas udara menunjukkan bahwa data kualitas udara dapat memperlihatkan kecenderungan meningkat, menurun, maupun relatif stabil dalam rentang waktu tertentu yang mencerminkan adanya pola tren pada data pengamatan [12]. Selain itu, analisis kualitas udara juga menemukan adanya fluktuasi musiman yang menyebabkan konsentrasi pencemar dapat berubah pada periode-periode tertentu secara berulang [14]. Perubahan tersebut dipengaruhi oleh berbagai faktor, seperti aktivitas manusia, kondisi meteorologi, serta karakteristik lingkungan yang berbeda pada setiap waktu pengamatan [13]. Karakteristik data kualitas udara juga menunjukkan adanya keterkaitan antara nilai pada suatu periode dengan periode sebelumnya sehingga pola perubahan yang terjadi tidak bersifat acak sepenuhnya [12]. Oleh karena itu, ISPU dapat dipahami sebagai data deret waktu yang menyimpan informasi mengenai tren, fluktuasi, dan pola perubahan kualitas udara yang dapat dimanfaatkan untuk memahami kondisi kualitas udara pada periode berikutnya [14].

Informasi historis yang tersimpan pada data deret waktu dapat dimanfaatkan untuk memperkirakan kondisi pada periode mendatang melalui pendekatan *Forecasting* [15]. Pada data deret waktu, *Forecasting* dilakukan dengan membaca pola dari pengamatan sebelumnya sehingga hubungan antarperiode dapat digunakan sebagai dasar dalam proses prediksi [16]. Pendekatan ini banyak digunakan pada data yang memiliki pola historis karena mampu memanfaatkan informasi masa lalu sebagai dasar dalam proses prediksi [17]. Penerapan *Forecasting* pada data curah hujan menunjukkan bahwa pola historis dapat dimanfaatkan untuk mengenali perubahan kondisi pada periode berikutnya [15]. Pada bidang lain, pendekatan *time series regression* juga digunakan untuk memperkirakan kondisi masa depan berdasarkan pola historis yang tersedia pada data pengamatan [16]. Hasil penelitian lain menunjukkan bahwa model prediktif mampu memanfaatkan pola historis untuk memperkirakan kondisi yang belum diamati pada periode mendatang [17]. Berdasarkan hal tersebut, *forecasting* menjadi pendekatan yang relevan pada data deret waktu karena mampu memanfaatkan pola historis sebagai dasar dalam memperkirakan kondisi pada periode mendatang, termasuk pada data kualitas udara yang menunjukkan karakteristik temporal dan perubahan antarperiode.

Perkembangan penelitian prediksi kualitas udara menunjukkan bahwa berbagai metode telah digunakan untuk memanfaatkan pola historis dalam memperkirakan kondisi udara pada periode mendatang. Penelitian [18] membandingkan beberapa metode prediksi indeks kualitas udara, yaitu ARX-GARCH, AR *Logistic Regression*, SVM, dan NNARX, serta menunjukkan bahwa SVM memperoleh performa terbaik dengan tingkat akurasi sebesar 88,53%. Pendekatan yang lebih kompleks kemudian dikembangkan pada penelitian [19] melalui kombinasi CNN dan ARIMA dalam kerangka *Time Based Spatial Forecasting* yang menghasilkan nilai  $R^2$  sebesar 0,81 dengan RMSE sebesar 44,60 pada prediksi kualitas udara perkotaan. Penelitian [20] juga menunjukkan bahwa pemanfaatan algoritma *machine learning* mampu menghasilkan performa yang sangat baik, di mana CatBoost memperoleh nilai  $R^2$  sebesar 0,9998 dengan RMSE sebesar 0,76 pada prediksi kualitas udara. Perbedaan hasil yang diperoleh pada

masing-masing penelitian menunjukkan bahwa setiap metode memiliki kemampuan yang berbeda dalam menangkap pola temporal pada data kualitas udara. Dalam evaluasi model prediksi, nilai RMSE yang lebih kecil menunjukkan tingkat kesalahan prediksi yang lebih rendah, sedangkan nilai  $R^2$  yang semakin mendekati 1 menunjukkan kemampuan model yang lebih baik dalam menjelaskan variasi data aktual. Temuan tersebut mengindikasikan bahwa belum terdapat satu metode yang secara konsisten memberikan performa terbaik pada seluruh kondisi data kualitas udara.

Salah satu pendekatan yang dikembangkan untuk meningkatkan performa prediksi adalah *Stacking Ensemble*, yaitu metode yang menggabungkan beberapa model dalam satu arsitektur pembelajaran sehingga kelebihan dari masing-masing model dapat dimanfaatkan secara bersamaan. Penelitian [20] menunjukkan bahwa algoritma berbasis *ensemble* seperti *Random Forest*, XGBoost, dan CatBoost mampu menghasilkan performa yang baik pada prediksi kualitas udara, di mana CatBoost memperoleh nilai  $R^2$  tertinggi sebesar 0,9998 dengan RMSE sebesar 0,76, sementara *Random Forest* dan XGBoost juga menunjukkan performa prediksi yang kompetitif. Untuk memanfaatkan keunggulan beberapa model secara bersamaan, penelitian [21] mengembangkan arsitektur *Stacking Ensemble* yang mengombinasikan *Random Forest* dan XGBoost melalui *Ridge Regression* sebagai *meta learner*. Hasil penelitian tersebut menunjukkan bahwa model *Stacking* memperoleh akurasi sebesar 98,2% dengan nilai  $R^2$  sebesar 0,99, lebih baik dibandingkan *Random Forest* yang memperoleh akurasi 92,6% dan XGBoost sebesar 94,4%, sehingga menunjukkan bahwa penggabungan beberapa model mampu meningkatkan kualitas prediksi dibandingkan penggunaan model individual. Meskipun demikian, kualitas model *Stacking* tetap dipengaruhi oleh konfigurasi parameter yang digunakan pada setiap model penyusunnya sehingga optimasi *hyperparameter* menjadi salah satu aspek yang penting untuk diperhatikan. Penelitian [22] menunjukkan bahwa *Bayesian Optimization* mampu meningkatkan nilai  $R^2$  model dari 0,8747 menjadi 0,8861 serta menurunkan RMSE dari 0,0242 menjadi 0,0230, sedangkan penelitian [23] menunjukkan bahwa *Genetic Algorithm* mampu meningkatkan nilai  $R^2$  pengujian dari 0,9688 menjadi

0,9762 serta menurunkan RMSE dari 1,247 menjadi 1,089 dan MAE dari 0,7736 menjadi 0,7442. Berdasarkan temuan tersebut, penerapan *Bayesian Optimization* dan *Genetic Algorithm* pada arsitektur *Stacking Ensemble* berbasis *Random Forest* dan XGBoost masih memiliki potensi untuk dieksplorasi lebih lanjut pada prediksi ISPU Kota Jakarta sehingga menjadi fokus dalam penelitian ini.

Berdasarkan hasil kajian penelitian terdahulu, *Stacking Ensemble* terbukti mampu meningkatkan performa prediksi melalui penggabungan beberapa model, sedangkan *Bayesian Optimization* dan *Genetic Algorithm* menunjukkan potensi dalam meningkatkan kualitas model melalui proses optimasi *hyperparameter*. Meskipun demikian, penelitian yang mengintegrasikan *Random Forest* dan XGBoost dalam arsitektur *Stacking Ensemble* serta membandingkan model *Baseline*, *Bayesian Optimization*, dan *Genetic Algorithm* pada prediksi ISPU Kota Jakarta masih terbatas. Selain itu, kajian yang secara khusus mengevaluasi pengaruh kedua metode optimasi tersebut terhadap performa model *Stacking Ensemble* pada data kualitas udara di Indonesia juga masih relatif sedikit ditemukan. Oleh karena itu, penelitian ini mengusulkan model prediksi ISPU berbasis *Stacking Ensemble* dengan *Random Forest* dan XGBoost sebagai *base learner* serta *Ridge Regression* sebagai *meta learner* untuk memanfaatkan keunggulan masing-masing algoritma dalam menangkap pola kualitas udara. Penelitian ini tidak hanya berfokus pada pembangunan model prediksi, tetapi juga menganalisis pengaruh penerapan *Bayesian Optimization* dan *Genetic Algorithm* terhadap performa *Stacking Ensemble* sehingga dapat diketahui strategi optimasi yang memberikan hasil terbaik pada data ISPU Jakarta. Hasil perbandingan tersebut diharapkan dapat memberikan pemahaman yang lebih jelas mengenai kontribusi optimasi *hyperparameter* terhadap peningkatan akurasi prediksi kualitas udara. Selain itu, model terbaik yang diperoleh diimplementasikan ke dalam sistem prediksi berbasis *dashboard* yang memungkinkan proses *forecasting* dan penyajian informasi kualitas udara dilakukan secara lebih informatif, terstruktur, dan mudah dipahami oleh pengguna. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi pada aspek metodologis melalui evaluasi strategi optimasi

pada *Stacking Ensemble* maupun pada aspek praktis melalui pengembangan sistem prediksi kualitas udara.

### 1.2. Rumusan Masalah

Rumusan masalah yang menjadi fokus utama pada penelitian ini adalah sebagai berikut:

1. Bagaimana membangun model prediksi ISPU Kota Jakarta menggunakan arsitektur *Stacking Ensemble* yang mengombinasikan *Random Forest* dan *XGBoost* sebagai *base learner* serta *Ridge Regression* sebagai *meta learner*?
2. Bagaimana menerapkan optimasi *hyperparameter* menggunakan *Bayesian Optimization* dan *Genetic Algorithm* pada model *Stacking* untuk memperoleh konfigurasi model yang optimal?
3. Bagaimana mengimplementasikan model prediksi terbaik ke dalam sistem prediksi berbasis *dashboard* untuk menyajikan hasil *Forecasting* ISPU Kota Jakarta?

### 1.3. Batasan Masalah

Batasan yang ditetapkan pada masalah yang diamati pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan merupakan data kualitas udara harian di Kota Jakarta periode 2023–2025 yang diperoleh dari sumber resmi pemerintah daerah.
2. Penelitian ini menggunakan nilai maksimum ISPU harian sebagai variabel yang diprediksi, di mana nilai tersebut diperoleh dari parameter pencemar udara  $PM_{10}$ ,  $PM_{2.5}$ ,  $SO_2$ ,  $CO$ ,  $O_3$ , dan  $NO_2$ .
3. Model prediksi yang digunakan dibatasi pada arsitektur *Stacking* dengan *Random Forest* dan *XGBoost* sebagai *base learner* serta *Ridge Regression* sebagai *meta learner*.
4. Penelitian ini difokuskan pada optimasi model *Stacking* sehingga evaluasi penelitian tidak membahas perbandingan performa model tunggal secara terpisah.
5. Strategi optimasi *hyperparameter* yang digunakan dibatasi pada *Bayesian Optimization* dan *Genetic Algorithm*.

6. Evaluasi model difokuskan pada kemampuan prediksi nilai maksimum ISPU harian berdasarkan metrik *error* yang digunakan dalam penelitian, sehingga tidak mencakup analisis spasial antarwilayah atau antarstasiun pemantauan secara terpisah.

#### 1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk menjawab rumusan masalah serta memberikan arahan yang lebih jelas bagi penelitian ini. Adapun tujuan penelitian ini adalah sebagai berikut:

1. Membangun model prediksi ISPU Kota Jakarta menggunakan arsitektur *Stacking Ensemble* yang mengombinasikan *Random Forest* dan XGBoost sebagai *base learner* serta *Ridge Regression* sebagai *meta learner*.
2. Menerapkan optimasi *hyperparameter* menggunakan *Bayesian Optimization* dan *Genetic Algorithm* pada model *Stacking* untuk memperoleh konfigurasi parameter yang optimal.
3. Mengimplementasikan model prediksi terbaik ke dalam sistem prediksi berbasis *dashboard* untuk menyajikan hasil forecasting ISPU Kota Jakarta.

#### 1.5. Manfaat Penelitian

Manfaat penelitian yang dapat diperoleh dari hasil penelitian ini adalah sebagai berikut:

1. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan kajian pemodelan prediktif berbasis deret waktu, khususnya pada penerapan arsitektur *Stacking* dengan variasi strategi optimasi *hyperparameter*.
2. Penelitian ini diharapkan dapat memberikan gambaran mengenai penerapan model prediksi kategori kualitas udara harian secara terstruktur dan terukur. Hasil evaluasi komparatif yang diperoleh dapat menjadi pertimbangan dalam pemilihan strategi optimasi yang sesuai dengan karakteristik data kualitas udara.
3. Penelitian ini dapat menjadi bahan rujukan bagi penelitian selanjutnya yang mengkaji penerapan metode *ensemble* dan optimasi parameter pada domain data lingkungan maupun data deret waktu lainnya.