

BAB I

PENDAHULUAN

1.1. Latar Belakang

Udara merupakan gas-gas di permukaan bumi dan mengelilingi bumi serta tidak memiliki bentuk, wujud, maupun rasa. Gas tersebut salah satunya adalah oksigen, dimana gas tersebut sangat diperlukan oleh makhluk hidup dalam proses respirasi. Seiring bertambahnya populasi manusia, aktivitas manusia menjadi tekanan signifikan yang menyebabkan turunnya kualitas udara. Kualitas udara merupakan acuan kondisi kebersihan udara pada suatu wilayah yang diukur menggunakan Indeks Kualitas Udara (AQI)[1].

Penurunan kualitas udara ini merupakan salah satu masalah global yang mengalami peningkatan dari tahun 2008. Di Indonesia, Kualitas udara saat ini menjadi isu yang serius pada kota-kota besar seperti Jakarta. Penurunan kualitas udara akibat adanya polutan di atmosfer dapat membahayakan kesehatan manusia. Beberapa polutan udara yang berdampak pada kesehatan manusia meliputi karbon monoksida, nitrogen dioksida, sulfur dioksida, partikel partikulat, serta ozon[2]. Bahan pencemar tersebut dihasilkan dari aktivitas manusia seperti polusi hasil kendaraan, aktivitas industrinya, pembakaran sampah, kebakaran hutan atau lahan dan lain sebagainya. Penurunan kualitas udara juga diperparah dengan kurangnya intensitas hujan akibat musim kemarau[3].

Berdasarkan *Centre for Research on Energy and Clean Air (CREA)* polusi PLTU batu bara juga menjadi kontributor utama pencemaran udara di Jakarta dan menyebabkan 1.470 kematian setiap tahun serta menimbulkan kerugian kesehatan hingga 14,2 triliun[4]. Dimana, aktivitasnya merupakan penyumbang polutan paling tinggi yang berasal dari industri pembangkit listrik dan manufaktur. Ditemukan bahwa selama musim hujan (November hingga Mei) angin bergerak dari arah timur laut dan tenggara sehingga membawa emisi dari sumber polutan di Sumatera Selatan, Banten, dan Jawa Barat ke Jakarta. Dan pada musim kemarau (Juni hingga Oktober) angin dari Jawa Barat membawa sumber emisi ke wilayah timur dan tenggara Jakarta.

Berdasarkan IMHE, *Global Burden of Disease* tahun 2019, di Indonesia terdapat lebih dari 123.000 kematian setiap tahunnya akibat polusi udara[5]. Menurut data BPJS Kesehatan pada tahun 2022, total penyakit pernapasan mencapai biaya sekitar 10 triliun dengan rincian beban biaya pada penyakit pneumonia sekitar 5,8 triliun, TBC 1,3 triliun, infeksi saluran pernafasan atas (ISPA) 1,1 triliun dan asma 602 juta. Dimana, TBC di Indonesia merupakan tertinggi ke-2 di dunia. Penyakit pernafasan juga meningkat dalam satu tahun terakhir. Dan diperkirakan akan ada sekitar 1.090.000 kasus baru untuk TBC di tahun 2024. Berdasarkan data dari Nafas Indonesia, kasus ISPA di DKI Jakarta juga meningkat dari 50.000 pasien pada Januari 2021 menjadi sekitar 150.000 pasien di Juni 2023.

Berdasarkan situs pemantauan kualitas udara IQAir sering menunjukkan bahwa udara di Jakarta adalah kota terburuk di dunia. Pada tanggal 16 Agustus 2024, Jakarta kembali menempati posisi pertama dengan indeks kualitas udara (AQI) 156, dengan $PM_{2.5}$ sebesar 62.7, 13 kali di atas nilai panduan Organisasi Kesehatan Dunia (WHO). Menurut Kementerian Lingkungan Hidup dan Kehutanan (KLHK), penurunan kualitas udara di Jakarta memang sudah menjadi pola tahunan sejak tahun 1998. Penurunan kualitas udara tersebut dimulai dari bulan Juni, Juli, Agustus, dan September. Pamantauan ini dilakukan dengan menggunakan alat *air quality monitoring* milik Kementerian Lingkungan Hidup dan Kehutanan (KLHK) yang ditempatkan pada 16 titik di Jabodetabek. Berdasarkan IQAir, polusi udara di Jakarta 2024 ini diperkirakan telah menyebabkan 8400 kematian, dan merugikan sekitar \$2,200,000 USD atau sekitar 34 miliar rupiah.

Kualitas udara di Jakarta dapat memberikan gambaran tentang dinamika lingkungan di kawasan perkotaan. Sebagai kota metropolitan Indonesia, Jakarta mencerminkan kompleksitas tantangan yang dihadapi kota-kota besar lainnya, terutama terkait polusi udara yang disebabkan oleh aktivitas industri, transportasi, dan pembangkit listrik. Memahami kualitas udara di Jakarta tidak hanya penting untuk kesehatan masyarakat, tetapi juga untuk menunjang keberlanjutan pembangunan ekonomi dan peningkatan kualitas hidup di wilayah perkotaan. Sehingga sangat penting untuk memprediksi kualitas udara DKI Jakarta dan

memahami bagaimana konsentrasi polutan udara berubah serta kontribusi polutan udara terhadap kualitas udara di Jakarta guna menentukan upaya yang lebih efektif dalam mengatasi polusi udara di Jakarta. Prediksi ini dilakukan dengan menggunakan data historis Indeks Kualitas Udara DKI Jakarta dari Januari 2021 hingga November 2024.

Fokus utama dari penelitian ini adalah untuk meramalkan konsentrasi $PM_{2.5}$, yang merupakan salah satu polutan udara paling berbahaya bagi kesehatan manusia dan menjadi indikator utama dalam pengukuran kualitas udara. $PM_{2.5}$ merupakan komponen utama dalam indeks kualitas udara karena kemampuannya menembus sistem pernapasan dan dampaknya yang luas terhadap kesehatan, sebagaimana dinyatakan oleh United States Environmental Protection Agency[6]. Menurut World Health Organization (WHO), paparan jangka panjang terhadap $PM_{2.5}$ dapat meningkatkan risiko penyakit jantung iskemik, stroke, kanker paru-paru, dan infeksi saluran pernapasan bawah akut, terutama pada anak-anak dan lansia[7]. Oleh karena itu, memahami pola dan dinamika konsentrasi $PM_{2.5}$ sangat penting untuk mitigasi kesehatan dan perencanaan kebijakan lingkungan yang efektif.

Dalam meramalkan kualitas udara, penggunaan model *ensemble XGBoost* sangat dianjurkan karena kompleksitas data yang terlibat. Serta model *XGBoost* ini memiliki regularisasi untuk mengurangi risiko *overfitting*, terutama pada dataset besar. *XGBoost* merupakan model ansambel efektif dalam membuat model prediktif yang unggul [8][9]. *XGBoost* memiliki kemampuan untuk menangani berbagai jenis tugas pembelajaran, seperti regresi dan klasifikasi. *Extreme Gradient Boosting (XGBoost)* adalah salah satu metode yang menarik perhatian karena kemampuannya untuk memprediksi fenomena yang kompleks dengan tepat [10][11]. Dengan kemampuan pengolahan volume data yang besar dengan berbagai fitur dan hasil prediksi yang cepat dan akurat, *XGBoost* menjadi pilihan yang populer di kompetisi *data science*[12].

Namun, meskipun *XGBoost* menjadi pilihan utama, penting untuk membandingkan kinerjanya dengan metode *ensemble* lain seperti *Random Forest*. Dimana, *Random Forest* dikenal lebih stabil dan cenderung lebih mudah diterapkan, terutama karena tidak terlalu sensitif terhadap pengaturan *hyperparameter*. Perbandingan kedua metode ini dilakukan untuk memberikan

gambaran yang lebih jelas tentang keunggulan metode *boosting* dalam memprediksi data *non-linear* yang beragam seperti kualitas udara DKI Jakarta.

Meskipun pembelajaran mesin telah menunjukkan hasil yang menjanjikan dalam memprediksi kualitas udara, kurangnya interpretabilitas dalam banyak model tetap menjadi tantangan utama, terutama di bidang lingkungan di mana transparansi dan pemahaman sangat dibutuhkan. Model pembelajaran mesin yang mudah dijelaskan, seperti yang menggunakan *Shapley Additive Explanations (SHAP)*, menawarkan tingkat transparansi untuk memahami bagaimana model membuat keputusan. Transparansi ini tidak hanya mendukung proses pengambilan keputusan tetapi juga meningkatkan kepercayaan terhadap penerapan AI di sektor lingkungan.

Pada penelitian yang berjudul '*Application of XGBoost algorithm in hourly PM_{2.5} concentraton prediction*' dimana penelitian ini memprediksi konsentrasi PM_{2.5} di Tianjin, China. Penelitian ini menggunakan model *XGBoost*. Dengan menggunakan data setiap jam PM_{2.5}. Dimana, penelitian ini membandingkan *XGBoost*, *Random Forest*, *SVM regression*, *Multiple Linear Regression*, dan *Decision Tree Regression*. Didapatkan bahwa *XGBoost* mendapatkan RMSE sebesar 17,298, MAE sebesar 11,774, dan R² sebesar 0,952.

Pada penelitian lain yang memprediksi PM_{2.5} menggunakan data harian dari Januari 2014 hingga Desember 2019 dengan jumlah data sebanyak 2044 data. Penelitian ini juga membandingkan *Linear Regression*, *Random Forest*, KNN, *Ridge and Lasso Regression*, *XGBoost*, dan *AdaBoost*. Didapatkan *XGBoost* merupakan model terbaik dengan MAE sebesar 8.27 dan RMSE 13.85[13]. Penelitian ini juga memperkirakan kategori PM_{2.5} yang telah diprediksi tersebut, dengan standar PM_{2.5} paling rendah 0 dan paling tinggi 30 sehingga termasuk pada kategori baik dan berpengaruh baik terhadap kesehatan.

Sedangkan pada penelitian yang memprediksi konsentrasi polutan akhir dengan menggunakan detektor kualitas udara mikro. Penelitian ini menggunakan 4200 sampel data yang mencatat konsentrasi polutan per jam dari 14 November 2018 hingga 11 Juni 2019. Penelitian ini membandingkan model *Ridge*, *XGBoost*, *RR-XGBoost*, *Random Forest Regression*, *SVR*, dan *MLP*. Didapatkan bahwa dengan penggabungan model *Ridge Regression* dan *Extreme Gradient Boosting*

(RR-*XGBoost*) mampu mendapatkan *MAPE* dengan rata-rata 0,0733 serta R^2 rata-rata 0.98[14].

Berdasarkan penelitian terdahulu didapatkan bahwa model *XGBoost* mampu melakukan prediksi kualitas udara. Model *XGBoost* juga mampu menangani data yang tidak terstruktur dan mampu mencegah *overfitting*. Oleh karena itu, penelitian ini menggunakan model *XGBoost* dalam melakukan prediksi kualitas udara DKI Jakarta. Akan tetapi, *XGBoost* merupakan model *black-box* sehingga memiliki kesulitan dalam memahami pengaruh tiap variabel fitur terhadap hasil prediksi. Sehingga kombinasi antara *XGBoost* dan *SHAP* dalam memprediksi parameter $PM_{2.5}$ akan memberikan pemahaman yang lebih baik tentang dinamika kualitas udara. Diharapkan penelitian ini akan memberikan masyarakat wawasan tentang kualitas udara serta memberikan referensi untuk pemerintah dalam merumuskan kebijakan atau langkah preventif yang tepat.

1.2. Rumusan Masalah

Berdasarkan latar belakang, berikut merupakan permasalahan yang menjadi fokus utama pembahasan:

1. Bagaimana metode *XGBoost* dalam meramalkan $PM_{2.5}$ di Jakarta?
2. Bagaimana perbandingan antara *XGBoost* dan *Random Forest* dalam meramalkan $PM_{2.5}$ di Jakarta?
3. Bagaimana *SHAP* dalam memahami model prediksi dalam meramalkan $PM_{2.5}$ di Jakarta?
4. Bagaimana model prediksi terbaik dan *SHAP* diterapkan dalam aplikasi berbasis *website*?

1.3. Batasan Masalah

Berikut merupakan batasan masalah pada penelitian ini yang membantu penelitian ini agar tidak terlalu luas dan fokus dengan permasalahan penelitian:

1. Penelitian ini menggunakan data sekunder yakni data harian Indeks Standar Pencemar Udara (ISPU) DKI Jakarta dari Januari 2021 hingga November 2024.

2. Peneliti menggunakan metode *XGBoost* dan *Random Forest* dalam meramalkan $PM_{2.5}$ DKI Jakarta. Dan menggunakan *SHAP* dalam menginterpretasikan hasil prediksi.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah, berikut merupakan tujuan yang ingin dicapai dalam penelitian ini:

1. Membuat sebuah model prediksi $PM_{2.5}$ di Jakarta menggunakan metode *XGBoost* dan *Random Forest*.
2. Membandingkan hasil peramalan dari *XGBoost* dan *Random Forest* untuk menilai performa masing-masing model dalam meramalkan $PM_{2.5}$ di Jakarta.
3. Menganalisis hasil peramalan $PM_{2.5}$ di Jakarta yang dihasilkan oleh model prediksi dengan metode *SHAP* untuk memahami kontribusi masing-masing variabel terhadap hasil prediksi.
4. Menerapkan hasil peramalan dari model prediksi terbaik untuk *deployment* aplikasi berbasis *website* prediksi $PM_{2.5}$ di Jakarta.

1.5. Manfaat Penelitian

Berikut merupakan manfaat yang dapat diperoleh dari hasil penelitian ini:

1. Bidang Keilmuan:
Meningkatkan pemahaman tentang faktor-faktor yang mempengaruhi $PM_{2.5}$. Mengetahui pengaplikasian dari metode *XGBoost* dan interpretasi *SHAP* di bidang lingkungan dan statistik.
2. Bagi Masyarakat:
Meningkatkan kesadaran masyarakat akan dampak aktivitas sehari-hari terhadap kualitas udara dan mendorong partisipasi aktif dalam program pelestarian lingkungan.
3. Bagi Pemerintah:
Menyediakan informasi pendukung kebijakan dan regulasi yang lebih efektif dalam pengelolaan polutan udara.