

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kualitas udara merupakan salah satu faktor lingkungan yang memiliki pengaruh signifikan terhadap kesehatan dan kesejahteraan manusia. Udara yang bersih berperan penting dalam mengurangi risiko penyakit pernapasan, gangguan kardiovaskular, hingga kanker paru-paru yang disebabkan oleh paparan polutan udara [1]. Selain itu, lingkungan dengan kualitas udara yang baik juga mendukung peningkatan produktivitas, kenyamanan dalam beraktivitas, serta kualitas hidup secara menyeluruh [2].

Namun, kondisi udara di berbagai kota besar menunjukkan tren yang memburuk, termasuk di Jakarta [3]. Sebagai ibu kota sekaligus pusat kegiatan ekonomi dan transportasi, Jakarta menghadapi tekanan besar terhadap kualitas udaranya [3]. Berdasarkan laporan IQAir [4] pada Agustus 2023, Jakarta menempati posisi kedua sebagai kota dengan kualitas udara terburuk di dunia dengan nilai US Air Quality Index (AQI) sebesar 154 yang tergolong dalam kategori “tidak sehat”. Kondisi tersebut berdampak signifikan terhadap kesehatan masyarakat, di antaranya peningkatan angka gangguan pernapasan, iritasi mata, serta risiko penyakit kronis, khususnya pada sistem pernapasan dan kardiovaskular [5].

Menurut Abdurrohman, dkk, lebih dari 10.000 kematian setiap tahun di Jakarta berkaitan langsung dengan paparan polusi udara dan sekitar 5.000 kasus rawat inap terjadi akibat gangguan pernapasan akut . Polusi udara di Jakarta sebagian besar bersumber dari emisi kendaraan bermotor, aktivitas industri, pembakaran sampah yang tidak terkendali, serta kebakaran hutan. Partikel halus seperti PM2.5 dan PM10 menjadi penyebab utama yang paling berbahaya karena mampu masuk ke dalam saluran pernapasan hingga mencapai aliran darah [7]. Penelitian [7] menunjukkan bahwa konsentrasi PM2.5 di Jakarta secara konsisten melebihi ambang batas aman yang ditetapkan oleh WHO, sehingga menempatkan masyarakat pada risiko tinggi terhadap berbagai penyakit kronis.

Dampak polusi udara juga dirasakan secara lebih signifikan oleh kelompok rentan, seperti anak-anak dan lansia. Paparan jangka panjang terhadap udara berpolusi dapat menyebabkan penurunan fungsi paru-paru secara permanen pada anak-anak serta memperburuk kondisi kesehatan kronis pada lansia [8].

Untuk mengukur kualitas udara, Indonesia menggunakan Indeks Standar Pencemar Udara (ISPU) yang membagi kualitas udara ke dalam beberapa kategori, yaitu baik, sedang, tidak sehat, sangat tidak sehat, dan berbahaya [9]. Ketika nilai ISPU melebihi ambang batas aman, maka risiko terhadap kesehatan masyarakat pun meningkat secara signifikan. Oleh karena itu, pemantauan kualitas udara yang akurat dan berkelanjutan menjadi hal yang sangat penting dalam upaya mitigasi dampak polusi udara [10].

Dalam upaya mewujudkan sistem pemantauan yang lebih efektif, pemanfaatan teknologi analisis data menjadi suatu kebutuhan. Salah satu pendekatan yang relevan adalah penerapan data mining, yaitu proses penggalian informasi dan pola tersembunyi dari kumpulan data berskala besar. Di dalam data mining, terdapat berbagai teknik analisis, salah satunya adalah klasifikasi. Klasifikasi merupakan metode untuk membangun model yang mampu memetakan data ke dalam kategori tertentu berdasarkan pola dari data historis. Dalam kualitas udara, teknik klasifikasi dapat digunakan untuk memprediksi kategori ISPU berdasarkan data parameter pencemar yang tersedia. Dengan kata lain, klasifikasi memungkinkan sistem untuk secara otomatis menentukan tingkat kualitas udara, sehingga mendukung pemantauan yang lebih cepat, akurat, dan berbasis data.

Untuk mengimplementasikan teknik klasifikasi tersebut, salah satu pendekatan yang umum digunakan adalah melalui algoritma machine learning. Machine learning memungkinkan sistem untuk belajar dari data historis dan membangun model prediktif yang dapat mengklasifikasikan data secara otomatis berdasarkan pola yang telah dipelajari. Berdasarkan hal tersebut, penelitian ini menggunakan tiga algoritma machine learning, yaitu Support Vector Machine (SVM), K-Nearest Neighbors (KNN), dan Random Forest (RF) untuk mengklasifikasikan kualitas udara di Jakarta. Pemilihan ketiga algoritma ini didasarkan pada perbedaan karakteristik dan pendekatan masing-masing model dalam proses klasifikasi. SVM merupakan algoritma margin-based yang dikenal

efektif untuk data berdimensi tinggi dan mampu menangani klasifikasi non-linear melalui penggunaan fungsi kernel. KNN adalah algoritma lazy learning yang sederhana namun cukup kuat, dengan prinsip kerja berdasarkan kedekatan antar data. Sementara itu, Random Forest merupakan metode ensemble yang terdiri dari banyak pohon keputusan dan memiliki keunggulan dalam menangani variabel kompleks serta mengurangi risiko overfitting.

Ketiga algoritma tersebut dipilih untuk dibandingkan karena mewakili pendekatan yang berbeda dalam machine learning, sehingga hasil evaluasi dapat memberikan gambaran yang lebih menyeluruh terhadap performa model klasifikasi pada data kualitas udara. Dengan melakukan perbandingan, dapat diketahui model mana yang paling efektif dan akurat untuk diterapkan dalam sistem pemantauan kualitas udara. Selain itu, pemilihan model yang optimal juga penting dalam mendukung pengambilan keputusan oleh instansi terkait secara lebih tepat, efisien, dan berbasis data. Oleh karena itu, penelitian ini difokuskan pada perbandingan performa tiga algoritma klasifikasi, yaitu KNN, SVM, dan Random Forest untuk mengetahui model yang paling baik dalam mengklasifikasikan kualitas udara di Jakarta.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, permasalahan utama dalam penelitian ini terletak pada bagaimana membangun model klasifikasi yang efektif dan akurat untuk mengidentifikasi kategori kualitas udara di Jakarta berdasarkan data parameter pencemar udara. Untuk menjawab permasalahan tersebut, penelitian ini difokuskan pada penerapan dan perbandingan tiga algoritma klasifikasi, yaitu Support Vector Machine (SVM), K-Nearest Neighbors (KNN), dan Random Forest (RF). Ketiga algoritma tersebut dipilih karena memiliki karakteristik dan pendekatan yang berbeda dalam menangani proses klasifikasi, sehingga penting untuk dianalisis kinerjanya dalam memodelkan data kualitas udara serta kemampuannya dalam mengelompokkan data ke dalam kategori ISPU yang sesuai. Penelitian ini juga bertujuan untuk mengevaluasi performa masing-masing model berdasarkan metrik seperti akurasi, precision, recall, dan F1-score,

guna mengidentifikasi algoritma yang paling optimal sebagai dasar pengembangan sistem pemantauan kualitas udara berbasis machine learning.

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah membandingkan performa model Support Vector Machine (SVM), K-Nearest Neighbors (KNN), dan Random Forest (RF) berdasarkan nilai evaluasi, seperti akurasi, presisi, recall, dan F1-score.

1.4 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut.

1. Menyediakan dasar bagi pengembangan sistem monitoring kualitas udara berbasis machine learning untuk deteksi dini terhadap peningkatan kadar polusi.
2. Menjadi referensi untuk penelitian selanjutnya yang berkaitan dengan pengolahan dan analisis data kualitas udara menggunakan pendekatan machine learning.