

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kualitas udara merupakan salah satu aspek dalam lingkungan yang memiliki pengaruh signifikan terhadap kesehatan masyarakat, khususnya di wilayah perkotaan dengan tingkat aktivitas industri dan transportasi yang tinggi. Sebagai ibu kota Negara Kesatuan Republik Indonesia, DKI Jakarta memiliki kepadatan penduduk dan mobilitas yang tinggi, sehingga berpotensi mengalami pencemaran udara yang signifikan. Kondisi kualitas udara yang buruk dapat menimbulkan berbagai dampak negatif, seperti gangguan pernapasan, penurunan kualitas hidup, dan peningkatan risiko penyakit kronis (WHO, 2023).

Untuk mengukur tingkat pencemaran udara, Pemerintah Republik Indonesia telah menggunakan Indeks Standar Pencemaran Udara (ISPU) sebagai indikator resmi dalam menilai tingkat kualitas udara. ISPU dihitung berdasarkan konsentrasi beberapa parameter polutan utama, yaitu PM_{2.5}, PM₁₀, SO₂, CO, O₃, dan NO₂. Setelah dilakukan perhitungan, nilai ISPU kemudian diklasifikasikan ke dalam beberapa kategori kualitas udara, seperti Baik, Sedang, Tidak Sehat, Sangat Tidak Sehat, dan Berbahaya.

Informasi kualitas udara yang tersedia saat ini umumnya masih bersifat monitoring, yaitu hanya menyajikan data hasil pengukuran tanpa analisis lanjutan untuk mengelompokkan kondisi kualitas udara secara otomatis. Padahal, data kualitas udara historis yang tersedia dari instansi pemerintah memiliki potensi besar untuk dianalisis lebih lanjut guna mengklasifikasikan kondisi kualitas udara berdasarkan pola data polutan.

Perkembangan teknologi komputasi memungkinkan penerapan metode machine learning dalam proses klasifikasi data secara lebih sistematis dan objektif. Algoritma Decision Tree dan Random Forest termasuk metode klasifikasi yang banyak digunakan karena mampu mengolah data multivariat serta menghasilkan performa klasifikasi yang cukup baik. Decision Tree memiliki kelebihan dalam menghasilkan aturan klasifikasi yang mudah dipahami, sedangkan Random

Forest dapat meningkatkan tingkat akurasi dan stabilitas model melalui pendekatan ensemble.

Berdasarkan latar belakang permasalahan tersebut, Studi ini difokuskan pada analisis perbandingan performa algoritma Decision Tree dan Random Forest dalam proses klasifikasi kualitas udara di Jakarta berdasarkan Indeks Standar Pencemaran Udara (ISPU). Penelitian ini bertujuan untuk menganalisis serta membandingkan kinerja kedua algoritma menggunakan berbagai metrik evaluasi sehingga menentukan model yang lebih optimal dalam konteks klasifikasi kualitas udara. Implementasi berbasis web dikembangkan sebagai media pendukung untuk menampilkan hasil klasifikasi, namun bukan menjadi fokus utama penelitian.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang telah dijabarkan sebelumnya, maka permasalahan yang dikaji dalam penelitian ini dirumuskan sebagai berikut:

1. Bagaimana tahapan pengolahan data mentah kualitas udara Jakarta hingga data tersebut siap digunakan dalam proses klasifikasi kualitas udara?
2. Bagaimana kinerja algoritma Decision Tree dalam melakukan klasifikasi kualitas udara?
3. Bagaimana kinerja algoritma Random Forest dalam proses klasifikasi kualitas udara?
4. Algoritma manakah yang memiliki performa paling optimal berdasarkan hasil evaluasi menggunakan metrik klasifikasi?

1.3 Tujuan Penelitian

Penelitian ini memiliki beberapa tujuan yang dirumuskan sebagai berikut:

1. Melakukan proses pengolahan data mentah kualitas udara Jakarta yang diperoleh dari instansi pemerintah sehingga menghasilkan dataset yang siap digunakan dalam proses klasifikasi.
2. Menganalisis kinerja algoritma Decision Tree dalam melakukan klasifikasi kualitas udara.

3. Menganalisis performa algoritma Random Forest dalam mengklasifikasikan kualitas udara.
4. Melakukan analisis perbandingan kinerja algoritma Decision Tree dan Random Forest dalam proses klasifikasi kualitas udara guna menentukan model yang memiliki performa paling optimal.

1.4 Batasan Masalah

Agar penelitian tetap berada dalam ruang lingkup yang jelas, beberapa batasan penelitian ditetapkan sebagai berikut:

1. Data yang digunakan dalam penelitian ini berupa data mentah Indeks Standar Pencemaran Udara (ISPU) Tahun 2021 yang diperoleh melalui Open Data DKI Jakarta.
2. Penelitian ini dibatasi pada data kualitas udara yang berasal dari lima stasiun pemantauan kualitas udara yang tersedia pada dataset.
3. Parameter polutan yang dianalisis dalam penelitian ini meliputi PM10, SO₂, CO, O₃, dan NO₂.
4. Proses klasifikasi bertujuan untuk menentukan kategori kualitas udara terdiri dari Baik, Sedang, Tidak Sehat.
5. Penilaian performa model dilakukan dengan menggunakan beberapa metrik evaluasi, yaitu Accuracy, Precision, Recall, dan F1-score.
6. Implementasi berbasis web hanya digunakan sebagai media visualisasi hasil klasifikasi dan bukan menjadi fokus utama penelitian.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan beberapa manfaat sebagai berikut:

1. Manfaat Teoritis
Penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang data mining dan machine learning, khususnya terkait analisis perbandingan algoritma klasifikasi Decision Tree dan Random Forest pada data kualitas udara.
2. Manfaat Praktis

- a. Menyediakan informasi mengenai kinerja algoritma Decision Tree dan Random Forest dalam proses klasifikasi kualitas udara.
 - b. Menjadi referensi dalam pemilihan algoritma klasifikasi yang lebih optimal untuk data ISPU.
 - c. Menjadi bahan pertimbangan bagi penelitian selanjutnya dalam pengembangan model klasifikasi kualitas udara.
3. Manfaat Akademis
- Penelitian ini juga diharapkan dapat menjadi referensi bagi mahasiswa maupun penelitian yang tertarik melakukan penelitian sejenis dalam bidang klasifikasi data lingkungan dengan menggunakan metode machine learning,

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun untuk memberikan gambaran mengenai alur pembahasan penelitian yang disajikan dalam beberapa bab sebagai berikut:

- **BAB I Pendahuluan** memaparkan latar belakang penelitian, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, serta sistematika penulisan skripsi.
- **BAB II Tinjauan Pustaka** menguraikan penelitian terdahulu, landasan teori, serta berbagai konsep yang relevan dengan topik klasifikasi kualitas udara dan penerapan machine learning.
- **BAB III Metodologi Penelitian** menjelaskan pendekatan dan metode penelitian yang digunakan, sumber data penelitian, tahapan pengolahan data, serta algoritma klasifikasi yang diterapkan dalam penelitian ini.
- **BAB IV Hasil dan Pembahasan** menyajikan hasil proses klasifikasi, evaluasi performa model, analisis perbandingan antara algoritma Decision Tree dan Random Forest, serta pembahasan terhadap hasil penelitian yang diperoleh.
- **BAB V Penutup** memuat kesimpulan yang diperoleh dari keseluruhan hasil penelitian serta saran yang dapat digunakan sebagai bahan pengembangan penelitian selanjutnya.