

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian mengenai klasifikasi kualitas udara di Jakarta berdasarkan Indeks Standar Pencemaran Udara (ISPU) dengan menggunakan algoritma *Decision Tree* dan *Random Forest*, beberapa kesimpulan yang dapat diperoleh adalah sebagai berikut:

1. Proses Pengolahan Data Berhasil Menghasilkan Dataset Siap Klasifikasi

Data mentah ISPU Tahun 2021 yang diperoleh dari portal Open Data DKI Jakarta telah melalui serangkaian tahapan pengolahan data, yaitu *data understanding*, *data cleaning*, *data transformation*, *data selection*, dan *data splitting*. Tahapan *preprocessing* tersebut berhasil menangani permasalahan seperti *missing value* dan data tidak valid sehingga menghasilkan dataset yang bersih serta layak digunakan dalam proses pemodelan klasifikasi.

2. Algoritma *Decision Tree* dan *Random Forest* Dapat Digunakan untuk Klasifikasi Kualitas Udara

Penerapan kedua algoritma *machine learning* tersebut mampu melakukan klasifikasi kualitas udara ke dalam kategori Baik, Sedang, dan Tidak Sehat berdasarkan parameter polutan PM10, SO₂, CO, O₃, NO₂. Model yang dibangun mampu mempelajari hubungan antara konsentrasi polutan dengan kategori kualitas udara sehingga dapat digunakan untuk melakukan prediksi klasifikasi.

3. *Random Forest* Menunjukkan Performa Lebih Baik Dibandingkan *Decision Tree*

Berdasarkan hasil evaluasi menggunakan confusion matrix, accuracy, precision, recall, dan F1-score, diperoleh hasil bahwa:

1. *Decision Tree* menghasilkan tingkat akurasi sekitar 80,75%
2. *Random Forest* mencapai tingkat akurasi sekitar 85,17%

Perbedaan nilai akurasi tersebut menunjukkan bahwa Random Forest memiliki performa yang lebih baik dibandingkan Decision Tree. Hal ini disebabkan karena Random Forest menggunakan pendekatan ensemble learning yang menggabungkan beberapa pohon keputusan melalui mekanisme majority voting, sehingga mampu mengurangi risiko overfitting dan meningkatkan kemampuan generalisasi model.

4. **Evaluasi Multi-Metrik Memberikan Gambaran Performa yang Lebih Komprehensif**

Analisis menggunakan metrik precision, recall, dan F1-score menunjukkan bahwa kelas dengan jumlah data terbesar (Sedang) memiliki performa terbaik, sedangkan kelas minoritas memiliki tingkat kesalahan klasifikasi lebih tinggi. Oleh karena itu, evaluasi performa model tidak cukup hanya menggunakan akurasi, tetapi juga perlu mempertimbangkan metrik lain untuk mendapatkan gambaran performa yang lebih objektif.

5. **Integrasi Model ke Sistem Informasi Berbasis Web Berhasil Dilakukan**

Hasil klasifikasi berhasil diintegrasikan ke dalam sistem informasi berbasis web sebagai media pendukung penyampaian informasi kualitas udara. Sistem ini mampu menampilkan parameter polutan dan kategori kualitas udara secara informatif, sehingga memberikan nilai aplikatif terhadap hasil penelitian.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa algoritma *Random Forest* lebih efektif digunakan dalam klasifikasi kualitas udara Jakarta berdasarkan ISPU dibandingkan dengan model *Decision Tree* tunggal.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, beberapa saran yang dapat dipertimbangkan untuk pengembangan penelitian selanjutnya adalah sebagai berikut:

a. **Penambahan Dataset dan Rentang Waktu**

Penelitian selanjutnya dapat menggunakan data dengan rentang waktu lebih panjang (lebih dari satu tahun) untuk meningkatkan variasi data dan memperkuat kemampuan generalisasi.

b. Penanganan Ketidakseimbangan Data (Imbalanced Data)

Untuk meningkatkan performa model pada kelas dengan jumlah data yang lebih sedikit, penelitian berikutnya dapat menerapkan teknik penyeimbangan data seperti SMOTE (*Synthetic Minority Oversampling Technique*) atau metode *resampling* lainnya.

c. Penggunaan Algoritma Lain sebagai Pembanding

Penelitian berikutnya dapat mempertimbangkan penggunaan algoritma lain seperti Support Vector Machine (SVM), Gradient Boosting, atau XGBoost agar diperoleh perbandingan performa model yang lebih beragam.

d. Optimasi Hyperparameter

Model dalam penelitian ini menggunakan parameter default. Penggunaan teknik *hyperparameter tuning* seperti *Grid Search* atau *Random Search* berpotensi meningkatkan performa model secara signifikan.

e. Pengembangan Sistem Informasi yang Lebih Interaktif

Sistem berbasis web dapat dikembangkan lebih lanjut dengan fitur visualisasi grafik tren kualitas udara, peta interaktif, atau sistem peringatan dini berbasis notifikasi.

f. Integrasi dengan Data Real-Time

Untuk meningkatkan manfaat praktis, sistem dapat diintegrasikan dengan data kualitas udara real-time sehingga mampu memberikan informasi aktual kepada masyarakat.