BAB 5
KESIMPULAN

5.1 Kesimpulan

Berdasarkan analisis hasil terhadap confusion matrix dan hasil pengujian dan evluasi dalam membandingkan kedua algoritma, kesimpulan perbandingan antara KNN dan Random Forest dapat disimpulkan sebagai berikut:

1. Kedua algoritma mampu mengklasifikan data ISPU di jakarta namun algoritma Random Forest menghasilkan performa yang lebih stabil dan akurat dibandingkan KNN. Random Forest memiliki akurasi yang jauh lebih baik dibandingkan KNN karena menghasilkan sedikit kesalahan dibandingkan KNN, bila dilihat perbandingan nilai klasifikasi KNN dengan akurasi 96% dan Random Forest 99% hal ini menunjukkan selisih yang tidak terlampau jauh namun diungguli Random Forest.
2. Perbandingan kinerja secara keseluruhan, Random Forest lebih efektif karena memiliki keunggulan dalam akurasi, presisi, dan recall dibandingkan dengan KNN. KNN memiliki lebih banyak kesalahan dalam membedakan antara BAIK dan SEDANG. Random Forest lebih konsisten dalam memisahkan ketiga kategori. Random Forest juga lebih efektif dalam mengidentifikasi fitur penting dalam klasifikasi, sedangkan KNN lebih sederhana namun kurang fleksibel untuk data skala besar. Nilai metrik Presisi, recall, dan F1-score juga menunjukkan bahwa Random Forest lebih konsisten dalam menangani data tidak seimbang pada beberapa target.
3. Klasifikasi yang sudah dilakukan terhadap ISPU dengan KNN dan Random Forest menghasilkan sebuah dashboard interaktif yang memuat informasi penting dalam beberapa grafik terkait sebaran polutan dalam rentang tahun 2022 – 2023 pada setiap wilayah/stasiun di jakarta selatan, pusat, timur, barat dan utara. Visualisasi bertujuan memberikan informasi yang mudah dipahami serta dapat bermanfaat bagi pembaca sebagai memberikan informasi baru sehingga kedepannya dapat meningkatkan

kesadaran masyarakat terhadap kondisi udara, serta menjadi bahan pertimbangan dalam mengambil langkah preventif untuk menjaga kualitas udara yang baik kedepan.

5.2 Saran

1. Penelitian ini menggunakan dataset yang diperoleh dari pihak luar yakni DLH DKI Jakarta, guna memaksimalkan hasil penelitian selanjutnya diharapkan dapat menggunakan bantuan alat berupa sensor yang ditempatkan pada wilayah tertentu agar data dapat lebih beragam. Sehingga dapat memperluas cakupan data dengan periode waktu yang lebih panjang dan variasi lokasi yang lebih beragam agar model dapat diadaptasi untuk berbagai kondisi lingkungan, sekaligus dapat diintegrasikan ke dalam sebuah sistem monitoring kualitas udara secara real-time untuk membantu pengambilan keputusan yang lebih cepat dan akurat dalam manajemen lingkungan.
2. Kedepannya diperlukan teknik pra-pemrosesan data , dikarenakan data ISPU yang unik berisi polutan yang memiliki nilai outlier yang tinggi serta dalam hal balancing data karena data kategori dalam rentang yang jauh / beragam sehingga diperlukan balancing ekstrim.
3. Penggunaan algoritma lain yang lebih kompleks dapat digunakan seperti deep learning, dan time series analysis dikarenakan data ispu sendiri memiliki runtutan tanggal per tahun untuk diolah.
4. Pada penelitian ini polutan yang paling dominan adalah PM2.5, kedepan dapat diasjikan analisis lebih mendalam terkait faktor apa saja yang mempengaruhi polutan tersebut sehingga dapat memberikan solusi pada hasil klasifikasi kualitas udara.