

BAB V

PENUTUP

5.1 KESIMPULAN

Berdasarkan hasil penelitian mengenai perbandingan algoritma *K-Nearest Neighbor* (KNN) dan *Naïve Bayes* dalam klasifikasi kualitas udara, dapat disimpulkan sebagai berikut:

1. Informasi mengenai hasil performa dari *model K-Nearest Neighbors* (KNN) dan *Naïve Bayes* Berdasarkan hasil penelitian, algoritma KNN menunjukkan performa lebih baik dibandingkan dengan *Naïve Bayes*. KNN mencapai akurasi 83% dengan presisi tinggi di semua kelas. Model ini memiliki *nilai AUC (Area Under Curve)* yang baik untuk kelas mayoritas, yaitu 0,94 untuk kelas 0 dan 0,93 untuk kelas 1, menandakan kemampuannya yang kuat dalam memisahkan data. Namun, KNN memiliki performa yang lebih rendah untuk kelas minoritas, dengan AUC 0,78. Di sisi lain, *Naïve Bayes* lebih unggul dalam membedakan kelas minoritas, terbukti dari AUC sebesar 0,95, meski akurasi keseluruhannya hanya 55%. KNN memberikan presisi tertinggi di semua kelas, sedangkan *Naïve Bayes* memiliki recall sempurna (1,0) untuk kelas minoritas, tetapi dengan presisi yang rendah. Selain itu, nilai Log Loss *Naïve Bayes* lebih rendah (0,21) dibandingkan dengan KNN (0,8981), mencerminkan stabilitas probabilitas prediksinya yang lebih baik meskipun akurasi kurang memuaskan

2. Model terbaik dalam klasifikasi kualitas udara

Berdasarkan analisis keseluruhan, algoritma *K-Nearest Neighbors (KNN)* adalah model terbaik untuk klasifikasi kualitas udara dalam penelitian ini. Dengan akurasi yang lebih tinggi dan presisi yang konsisten di semua kelas, KNN lebih andal untuk aplikasi yang memerlukan prediksi akurat. Namun, untuk kasus yang memerlukan deteksi menyeluruh terhadap kelas minoritas, Naïve Bayes bisa menjadi pilihan alternatif karena recall yang tinggi meski mengorbankan akurasi dan presisi.

5.2 SARAN

1. Pemilihan Algoritma

Pemilihan algoritma disesuaikan dengan karakteristik data dan tujuan klasifikasi. KNN disarankan untuk diterapkan pada sistem klasifikasi kualitas udara yang memerlukan akurasi dan presisi tinggi dalam mengenali kelas mayoritas. Untuk aplikasi yang lebih berfokus pada deteksi data dari kelas minoritas, terutama jika data tidak seimbang, *Naive Bayes* dapat menjadi pilihan yang tepat..

2. Peningkatan Performa

Peningkatan performa dapat dilakukan melalui penyesuaian data dan parameter algoritma, seperti berikut:

1. Untuk meningkatkan kinerja *Naïve Bayes*, dapat dilakukan balancing dataset menggunakan metode oversampling atau undersampling agar lebih seimbang. Selain itu, penggabungan metode ensemble seperti bagging atau boosting dapat meningkatkan akurasi secara keseluruhan

2. *KNN* dapat dioptimalkan dengan pemilihan nilai k yang lebih sesuai serta menerapkan teknik seleksi fitur atau normalisasi data untuk meningkatkan hasil klasifikasi.

3. **Penggunaan Dataset yang Lebih Luas**

Penggunaan dataset yang lebih luas dapat menjadi langkah penting dalam pengembangan sistem klasifikasi. Penelitian lebih lanjut dapat menggunakan dataset yang lebih besar dan beragam untuk meningkatkan generalisasi model. Penambahan atribut lingkungan lainnya juga dapat memberikan wawasan yang lebih kaya dalam klasifikasi kualitas udara. Dengan memahami kelebihan dan kekurangan masing-masing algoritma, diharapkan penelitian ini dapat memberikan kontribusi signifikan dalam pengembangan sistem klasifikasi kualitas udara yang lebih efektif dan akurat.