

## BAB VI

### KESIMPULAN

Bab ini akan memaparkan hasil dari penelitian yang telah dilakukan serta pembahasan dari hasil tersebut. Uraian dari hasil penelitian ini terdiri dari dua bagian yaitu kesimpulan dan saran.

#### 6.1. Kesimpulan

Pemodelan kualitas air menggunakan algoritma *random forest* dan *naïve bayes* telah dilakukan. Masing-masing *machine learning* yang menggunakan algoritma berbeda telah memiliki hasil performanya. Perbandingan akurasi, presisi, recall, dan F1-score bertujuan untuk mengetahui algoritma yang lebih efektif dalam kasus kualitas air. Hasil dari penelitian ini dapat dijadikan acuan dalam pembangunan sistem klasifikasi air bersih.

Dari pemodelan yang telah dilakukan menggunakan data sekunder berasal dari data.jakarta.go.id. Merupakan data dari 4 periode yakni data tahun 2018 periode 1, data tahun 2018 periode 2, data tahun 2019 periode 1, dan data tahun 2019 periode 2. Fitur yang digunakan di antaranya adalah Kekeruhan, Air Raksa, Besi (Fe), Fluorida, Kesadahan, Krom Heksavalen, Mangan (Mn), Nitrat, Nitrit, Seng (Zn), Sulfat, Detergen, KMnO<sub>4</sub>, Total Coliform, E.Coli. Pemodelan *machine learning* dilakukan menggunakan bahasa pemrograman Python. Model dievaluasi menggunakan *confusion matrix* untuk mengetahui model yang memiliki kinerja paling baik.

Pembagian data latih dan data uji algoritma *random forest* menggunakan rasio 75:25. *Random forest* dibangun dengan 100 pohon dan menggunakan gini indeks sebagai kriteria pembagiannya. Klasifikasi kualitas air dengan *random forest* menghasilkan nilai akurasi sebesar 85.23%, presisi 88.11%, recall 79.89%, dan F1-score 83.35. Sementara itu pembagian data latih dan data uji algoritma *naïve bayes* menggunakan rasio 80:20. *Naïve bayes* dibangun menggunakan Gaussian yakni salah satu jenis dari *naïve bayes* yang disediakan oleh *library* sklearn.

Menghasilkan nilai akurasi 72.51%, presisi 76.19%, recall 74.51%, dan F1-score 73.08%.

Berdasarkan penelitian yang dilakukan algoritma *random forest* didapatkan hasil performa dengan pengujian menggunakan *confussion matrix*. Hasil tersebut menunjukkan bahwa algoritma *random forest* memiliki performa lebih baik dibandingkan dengan algoritma *naïve bayes* di setiap parameternya.

## 6.2. Saran

Penelitian dengan studi kasus kualitas air yang telah dilakukan masih dapat ditingkatkan kualitasnya dan dikembangkan kembali. Berikut merupakan beberapa saran yang dapat dijadikan sebagai referensi dan pertimbangan untuk penelitian selanjutnya.

1. Menggunakan dataset yang lebih besar untuk pemodelan *machine learning*.
2. Membangun sistem berbasis *website* atau *mobile* sehingga dapat digunakan oleh badan tertentu maupun masyarakat umum untuk mengetahui kualitas air.