

BAB V PENUTUP

5.1 Kesimpulan

Penelitian ini membuktikan bahwa metode *Ensemble Stacking* mampu meningkatkan performa klasifikasi kualitas air secara menyeluruh pada seluruh metrik evaluasi dibandingkan model tunggal XGBoost, LightGBM, dan CatBoost. Pada tahap evaluasi model dasar, LightGBM menunjukkan performa terbaik dengan accuracy sebesar 96,69%, F1-Score 85,40%, dan AUC-ROC tertinggi sebesar 98,85%, sementara recall tertinggi sebesar 86,26% dicapai oleh CatBoost. Perbedaan performa ini menunjukkan bahwa tidak ada satu model tunggal yang unggul pada seluruh metrik.

Penerapan metode *Stacking* dengan Random Forest sebagai *meta-learner* menghasilkan peningkatan performa pada hampir seluruh metrik utama. Dibandingkan model tunggal terbaik (LightGBM), accuracy meningkat dari 96,69% menjadi 97,06% (naik 0,37 poin persentase), recall meningkat dari 85,16% menjadi 86,26% (naik 1,10 poin), serta F1-Score meningkat dari 85,40% menjadi 86,98% (naik 1,58 poin). Peningkatan ini menunjukkan bahwa *Stacking* mampu menyeimbangkan kemampuan deteksi kelas positif dan negatif secara lebih baik.

Optimalisasi lebih lanjut menggunakan mekanisme *passthrough* memberikan peningkatan yang lebih signifikan terutama pada aspek ketepatan prediksi. Model *Stacking Passthrough* menghasilkan accuracy sebesar 97,12%, meningkat 0,43 poin dibandingkan LightGBM, serta precision meningkat dari 85,64% menjadi 90,00% (naik 4,36 poin persentase). Pada saat yang sama, recall tetap tinggi sebesar 84,07% dan F1-Score meningkat menjadi 86,93%, menunjukkan bahwa peningkatan precision tidak mengorbankan keseimbangan performa model. Nilai AUC-ROC sebesar 97,63% juga menunjukkan kemampuan diskriminasi kelas yang tetap stabil.

Tahap *hyperparameter tuning* pada model *Stacking Passthrough* menghasilkan performa terbaik secara keseluruhan. Model ini mencapai accuracy

97,12%, precision 90,48%, recall 83,52%, F1-Score 86,86%, dan AUC-ROC 97,66%. Dibandingkan model tunggal terbaik, optimasi ini menghasilkan peningkatan accuracy sebesar 0,43 poin, precision sebesar 4,84 poin, F1-Score sebesar 1,46 poin, serta mempertahankan nilai AUC-ROC pada tingkat yang sangat tinggi.

5.2 Saran

Berdasarkan proses penelitian dan hasil yang diperoleh, terdapat beberapa hal yang dapat disarankan untuk penelitian selanjutnya:

1. Perluasan Dataset

Dataset yang digunakan dapat diperluas baik dari segi jumlah sampel, variasi kondisi lingkungan, maupun tambahan parameter kualitas air. Hal ini memungkinkan model menjadi lebih general dan adaptif terhadap kondisi nyata di berbagai wilayah.

2. Eksperimen Meta-Learner dan Base-Learner Lain

Penelitian selanjutnya dapat mempertimbangkan penggunaan *meta-learner* lain, seperti Logistic Regression atau Gradient Boosting, untuk mengevaluasi apakah terdapat kombinasi yang mampu menghasilkan performa lebih optimal dibandingkan Random Forest. Selain itu, juga dapat memperluas variasi *base learner* dengan menambahkan model-model tunggal lain seperti Support Vector Machine, Decision Tree, K-Nearest Neighbors, dan Naive Bayes. Penambahan keragaman model ini diharapkan dapat meningkatkan efektivitas Ensemble Stacking melalui kombinasi pola prediksi yang lebih beragam, sehingga potensi peningkatan akurasi dan stabilitas model dapat dieksplorasi lebih jauh.

3. Mengatasi Tantangan Saat Penelitian

Beberapa proses seperti penyeimbangan data (SMOTE), pemilihan parameter, hyperparameter tuning dan penyesuaian arsitektur stacking membutuhkan waktu komputasi tinggi. Penelitian berikutnya dapat menggunakan perangkat keras lebih kuat atau melakukan optimasi komputasi agar proses lebih efisien.