

ABSTRAK

Kualitas air merupakan aspek vital yang memengaruhi kesehatan masyarakat dan keberlanjutan ekosistem. Pemantauan status mutu air di Kota Yogyakarta menghadapi kendala dalam hal efisiensi waktu dan biaya analisis laboratorium, terutama ketika harus mengukur 24 parameter kualitas air secara rutin. Penelitian ini bertujuan mengembangkan model klasifikasi status mutu air yang optimal menggunakan algoritma *Random Forest* yang dioptimasi dengan *Particle Swarm Optimization* (PSO) untuk melakukan optimasi *hyperparameter* dan seleksi fitur secara simultan.

Penelitian ini menggunakan dataset kualitas air dari Dinas Lingkungan Hidup Kota Yogyakarta yang terdiri dari 260 sampel dengan 24 parameter kualitas air. Metodologi penelitian meliputi tahap *data preprocessing* (*data cleaning*, *label encoding*, *data splitting*, imputasi *missing value* menggunakan median, dan normalisasi), penanganan ketidakseimbangan kelas menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE), pengembangan model *baseline Random Forest*, serta optimasi menggunakan PSO dengan *Stratified K-Fold Cross Validation*. Tiga skenario diuji: Skenario 1 (*Baseline* tanpa optimasi), Skenario 2 (PSO untuk optimasi *hyperparameter* saja), dan Skenario 3 (PSO untuk optimasi *hyperparameter* dan seleksi fitur simultan). Evaluasi model menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

Hasil penelitian menunjukkan bahwa Skenario 1 dan Skenario 2 menghasilkan performa identik dengan akurasi 96,15%, mengindikasikan bahwa optimasi *hyperparameter* saja tidak cukup untuk meningkatkan performa model. Skenario 3 memberikan performa terbaik dengan *accuracy* 100%, *precision* 100%, *recall* 100%, dan *F1-score* 100% pada data uji, membuktikan bahwa kombinasi optimasi *hyperparameter* dan seleksi fitur sangat krusial dalam mengatasi tumpang tindih karakteristik antar kelas. Algoritma PSO berhasil mereduksi dimensi data dari 24 fitur menjadi 7 fitur optimal (reduksi 70,8%), yaitu: Total Coliform, Kromium Heksavalen, Sulfida, Fecal Coli, Amonia NH₃-N, Timbal terlarut (Pb), dan Tembaga (Cu) terlarut. Model yang dikembangkan telah diimplementasikan dalam sistem berbasis web menggunakan Streamlit dengan tingkat akurasi prediksi mencapai 100% pada pengujian batch 10 sampel. Penelitian ini membuktikan bahwa kombinasi PSO untuk seleksi fitur dan optimasi *hyperparameter* dapat meningkatkan performa model secara signifikan sekaligus memberikan efisiensi biaya dan waktu dalam pemantauan kualitas air.

Kata Kunci: Klasifikasi Mutu Air, *Random Forest*, *Particle Swarm Optimization*, Seleksi Fitur, Optimasi *Hyperparameter*

ABSTRACT

Water quality is a vital aspect that affects public health and ecosystem sustainability. Monitoring water quality status in Yogyakarta City faces challenges in terms of time efficiency and laboratory analysis costs, especially when measuring 24 water quality parameters routinely. This research aims to develop an optimal water quality status classification model using Random Forest algorithm optimized with Particle Swarm Optimization (PSO) to simultaneously perform hyperparameter optimization and feature selection.

This research utilized a water quality dataset from the Environmental Agency of Yogyakarta City consisting of 260 samples with 24 water quality parameters. The research methodology includes data preprocessing stages (data cleaning, label encoding, data splitting, missing value imputation using median, and normalization), handling class imbalance using Synthetic Minority Over-sampling Technique (SMOTE), baseline Random Forest model development, and optimization using PSO with Stratified K-Fold Cross Validation. Three scenarios were tested: Scenario 1 (Baseline without optimization), Scenario 2 (PSO for hyperparameter optimization only), and Scenario 3 (PSO for simultaneous hyperparameter optimization and feature selection). Model evaluation employed accuracy, precision, recall, and F1-score metrics.

The results demonstrate that Scenario 1 and Scenario 2 achieved identical performance with 96.15% accuracy, indicating that hyperparameter optimization alone is insufficient to improve model performance. Scenario 3 achieved the best performance with 100% accuracy, 100% precision, 100% recall, and 100% F1-score on test data, proving that the combination of hyperparameter optimization and feature selection is crucial in overcoming overlapping characteristics between classes. The PSO algorithm successfully reduced data dimensionality from 24 features to 7 optimal features (70.8% reduction): Total Coliform, Hexavalent Chromium, Sulfide, Fecal Coli, Ammonia NH₃-N, Dissolved Lead (Pb), and Dissolved Copper (Cu). The developed model has been implemented in a web-based system using Streamlit with prediction accuracy reaching 100% in batch testing of 10 samples. This research proves that the combination of PSO for feature selection and hyperparameter optimization can significantly improve model performance while providing cost and time efficiency in water quality monitoring.

Keywords: *Water Quality Classification, Random Forest, Particle Swarm Optimization, Feature Selection, Hyperparameter Optimization*