

MODEL *ENSEMBLE LEARNING* UNTUK PREDIKSI PENCEMARAN AIR LAUT DALAM RANGKA MENDUKUNG KONSERVASI BIOTA TELUK JAKARTA

Lily Fitri Hasanah, 202231001

Dibimbing oleh Dr. Ir. Abdul Haris, S.Kom., M.Kom., IPM.

ABSTRAK

Peningkatan tekanan pencemaran di perairan pesisir Teluk Jakarta menimbulkan ancaman serius terhadap keberlanjutan ekosistem dan kehidupan biota laut. Kompleksitas karakteristik kualitas air yang bersifat multivariat serta keterbatasan metode evaluasi konvensional menuntut pendekatan yang lebih adaptif dan akurat dalam memodelkan tingkat pencemaran perairan laut. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model prediksi kualitas air laut berbasis *machine learning* menggunakan algoritma *Extreme Gradient Boosting* (XGBoost) guna mendukung pengelolaan dan perlindungan ekosistem biota laut secara berkelanjutan.

Dataset yang digunakan terdiri atas 3.180 data pengamatan dari 53 lokasi sampling pada periode Maret dan Agustus 2024, mencakup 28 parameter kualitas air yang mengacu pada baku mutu biota laut berdasarkan Keputusan Menteri Lingkungan Hidup Nomor 51 Tahun 2004. Pelabelan kelas mutu air dilakukan melalui perhitungan *Pollution Index* (PI) sesuai Keputusan Menteri Lingkungan Hidup Nomor 115 Tahun 2003, yang menghasilkan empat kategori mutu air, yaitu Baik, Cemar Ringan, Cemar Sedang, dan Cemar Berat. Model XGBoost dikembangkan melalui proses *hyperparameter tuning* menggunakan metode *grid search* terhadap 13.122 kombinasi parameter untuk memperoleh konfigurasi model yang optimal.

Hasil evaluasi menunjukkan tingkat kesesuaian prediksi PI yang sangat tinggi dengan nilai R^2 sebesar 0,9958, RMSE 0,2728, MAE 0,1942, dan MSE 0,0744 pada Periode I dan R^2 sebesar 0,9819, RMSE 0,4708, MAE 0,3702, dan MSE 0,2217 pada Periode II, serta menghasilkan akurasi klasifikasi sebesar 100% pada kedua periode pengamatan. Analisis *feature importance* menunjukkan bahwa Nitrat dan Salinitas menjadi parameter dominan pada Periode I, sedangkan Kecerahan dan Nitrat mendominasi pada Periode II. Dominasi parameter nutrisi mengindikasikan adanya tekanan eutrofikasi yang signifikan terhadap habitat biota laut. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa XGBoost efektif sebagai alat pendukung pengambilan keputusan dalam pengelolaan kualitas air laut yang berorientasi pada perlindungan ekosistem biota laut.

Kata kunci: Kualitas air laut, Indeks Pencemaran, XGBoost, klasifikasi multi-kelas, biota laut, Teluk Jakarta

ENSEMBLE LEARNING MODELS FOR PREDICTING SEA WATER POLLUTION TO SUPPORT THE CONSERVATION OF JAKARTA BAY BIOTA

Lily Fitri Hasanah, 202231001

Supervised by Dr. Ir. Abdul Haris, S.Kom., M.Kom., IPM.

ABSTRACT

Increasing pollution pressure in the coastal waters of Jakarta Bay poses a serious threat to ecosystem sustainability and marine biota. The multivariate nature of water quality characteristics, combined with the limitations of conventional evaluation methods, necessitates a more adaptive and accurate approach to modeling marine water pollution levels. Therefore, this study aims to develop a machine learning–based marine water quality prediction model using the Extreme Gradient Boosting (XGBoost) algorithm to support sustainable marine ecosystem management and protection.

The dataset comprises 3,180 observations collected from 53 sampling locations during March and August 2024, covering 28 water quality parameters in accordance with the marine biota water quality standards stipulated in Decree of the Minister of Environment No. 51 of 2004. Water quality classes were labeled using the Pollution Index (PI) calculated based on Decree of the Minister of Environment No. 115 of 2003, resulting in four water quality categories: Good, Lightly Polluted, Moderately Polluted, and Heavily Polluted. The XGBoost model was developed through hyperparameter tuning using a grid search approach over 13,122 parameter combinations to obtain an optimal model configuration.

The results indicate a very high level of agreement between predicted and observed PI values, with R^2 of 0.9958, RMSE of 0.2728, MAE of 0.1942, and MSE of 0.0744 in Period I, and R^2 of 0.9819, RMSE of 0.4708, MAE of 0.3702, and MSE of 0.2217 in Period II, while achieving a classification accuracy of 100% in both observation periods. Feature importance analysis reveals that nitrate and salinity are the dominant parameters in Period I, whereas water transparency and nitrate dominate in Period II. The predominance of nutrient-related parameters indicates significant eutrophication pressure on marine biota habitats. Overall, this study demonstrates that XGBoost is an effective decision-support tool for marine water quality management oriented toward the protection of marine ecosystems.

Keywords: *Seawater quality, Pollution Index, XGBoost, multi-class classification, marine biota, Jakarta Bay*