ABSTRAK

Air merupakan sumber daya alam yang sangat penting bagi kehidupan manusia, salah satunya digunakan untuk keperluan minum. Namun sulitnya akses terhadap air bersih masih menjadi masalah serius di banyak daerah. Tujuan dari penelitian ini adalah untuk memprediksi apakah air tersebut layak atau tidak untuk dikonsumsi. Penelitian ini akan membandingkan kinerja dua algoritma *ensemble learning*, yaitu Extreme Gradient Boosting (XGBoost) dan Random Forest, dalam mengklasifikasikan kualitas air layak minum sebelum dan setelah proses *balancing* menggunakan metode ADASYN. Berdasarkan penelitian, algoritma XGBoost mendapat akurasi sebesar 62.60%, *precision* 61.6%, *recall* 62.6%, serta F1-*score* 61.5% pada K-*Fold* 10 dan n_*estimators* = 50 dengan waktu komputasi sebesar 0.0570 s. Sedangkan algoritma Random Forest menghasilkan akurasi sebesar 69.99% , *precision* 64%, *recall* 65%, serta F1-*score* 63% pada K-*Fold* 10 dan n_*estimators* = 200 dengan waktu komputasi sebesar 1.0121 s. Setelah *balancing* menggunakan metode ADASYN, algoritma XGBoost mendapat akurasi terbaiknya sebesar 67.22%, *precision* 67.4%, *recall* 67.2%, serta F1-*score* 67.2% pada K-*Fold* = 10 dan *n_estimator* = 50 dengan waktu komputasi sebesar 0.0720 s. Sedangkan algoritma Random Forest menghasilkan akurasi terbaiknya sebesar 70.29%, *precision* 70.5%, *recall* 70.3%, serta F1-*score* 70.2% pada K-Fold = 10 dan n_*estimators* = 100 dengan waktu komputasi sebesar 0.55469 s.

Kata kunci : Extreme Gradient Boosting, Random Forest, Kualitas Air Layak Konsumsi, ADASYN, *Balance, Imbalance*

ABSTRACT

Water is a crucial natural resource for human life, with one of its uses being drinking. However, access to clean water remains a serious issue in many areas. The aim of this research is to predict whether water is suitable for consumption. This study will compare the performance of two ensemble learning algorithms, namely Extreme Gradient Boosting (XGBoost) and Random Forest, in classifying potable water quality before and after balancing using the ADASYN method.According to the research, the XGBoost algorithm achieved an accuracy of 62.60%, precision of 61.6%, recall of 62.6%, and F1-score of 61.5% with 10-fold cross-validation and n_estimators = 50, with a computation time of 0.0570 seconds. On the other hand, the Random Forest algorithm achieved an accuracy of 69.99%, precision of 64%, recall of 65%, and F1-score of 63% with 10-fold cross-validation and n_estimators = 200, with a computation time of 1.0121 seconds. After balancing using the ADASYN method, the XGBoost algorithm achieved its best accuracy of 67.22%, precision of 67.4%, recall of 67.2%, and F1-score of 67.2% with 10-fold cross-validation and n_estimators = 50, with a computation time of 0.0720 seconds. Meanwhile, the Random Forest algorithm achieved its best accuracy of 70.29%, precision of 70.5%, recall of 70.3%, and F1-score of 70.2% with 10-fold cross-validation and n_estimators = 100, with a computation time of 0.55469 seconds.

Keywords : Extreme Gradient Boosting, Random Forest, *Potable Water Quality*, ADASYN, *Balance, Imbalance*