

## ABSTRAK

### EVALUASI METODE *RANDOM FOREST*, *XGBOOST* DAN C5.0 DALAM KLASIFIKASI KUALITAS AIR BERSIH UNTUK Mendukung PENGELOLAAN SUMBER DAYA AIR

Oleh

MELAN CANIADI

Air bersih adalah kebutuhan dasar manusia yang penting untuk kehidupan sehari-hari dan kesehatan. Namun, banyak orang di seluruh dunia masih kekurangan akses air bersih dan sanitasi yang layak. Penelitian ini bertujuan untuk mengklasifikasikan kualitas air bersih menggunakan metode *Random Forest*, *XGBoost*, dan C5.0, guna mendukung pengelolaan sumber daya air. Data yang digunakan berasal dari *Kaggle*, mencakup 971 data kualitas air dari 62 titik lokasi sungai di Amerika Serikat dari tahun 1995 hingga 2014, pembagian data menggunakan metode *hold out* dan *stratified k-fold cross-validation*. Hasil penelitian menunjukkan bahwa metode *Random Forest* dengan menggunakan *hold out* memiliki akurasi tertinggi sebesar 0.979 dengan waktu eksekusi 429.806 ms, dibandingkan dengan *stratified k-fold cross-validation* dengan akurasi 0.977 dan waktu eksekusi 8584.102 ms. *XGBoost* dan C5.0 menunjukkan akurasi tertinggi sebesar 0.966 dengan *stratified k-fold cross-validation*, meskipun waktu eksekusi lebih lama dibandingkan dengan metode *hold out*. Akurasi *XGBoost* dengan *hold out* adalah 0.964 dengan waktu eksekusi 315.998 ms, sedangkan C5.0 memiliki akurasi 0.960 dengan waktu eksekusi 62.28 ms.

Kata Kunci: *Machine Learning*, *Random Forest*, *Extreme Gradien Boosting*, C5.0, Kualitas Air, Klasifikasi.

## **ABSTRACT**

### ***EVALUATION OF RANDOM FOREST, XGBOOST AND C5.0 METHODS IN CLEAN WATER QUALITY CLASSIFICATION TO SUPPORT WATER RESOURCES MANAGEMENT***

***By***

**MELAN CANIADI**

*Clean water is a basic human need that is important for daily life and health. However, many people around the world still lack access to clean water and proper sanitation. This research aims to classify clean water quality using the Random Forest, XGBoost, and C5.0 methods, to support water resource management. The data used comes from Kaggle, includes 971 water quality data from 62 river locations in the United States from 1995 to 2014, data division uses the hold out and stratified k-fold cross-validation methods. The research results show that the Random Forest method using hold out has the highest accuracy of 0.979 with an execution time of 429,806 ms, compared to stratified k-fold cross-validation with an accuracy of 0.977 and an execution time of 8584,102 ms. XGBoost and C5.0 show the highest accuracy of 0.966 with stratified k-fold cross-validation, although the execution time is longer compared to the hold out method. The accuracy of XGBoost with hold out is 0.964 with an execution time of 315.998 ms, while C5.0 has an accuracy of 0.960 with an execution time of 62.28 ms.*

*Keywords: Machine Learning, Random Forest, Extreme Gradient Boosting, C5.0, Water Quality, Classification.*