CLASSIFICATION OF RIVER WATER QUALITY IN THE SPECIAL REGION OF YOGYAKARTA (DIY) USING MACHINE LEARNING ALGORITHM

Andika Ristianto
Data Science Study Program, Faculty of Science and Technology
University of Technology Yogyakarta
Jl. Ringroad Utara Jombor Sleman Yogyakarta
E-mail : andikaristianto66@gmail.com

ABSTRACT

Clean water is a basic need for human life and the global ecosystem. The availability of adequate clean water has a significant impact on public health, economic growth, and environmental sustainability. Data from the Central Statistics Agency of the Special Region of Yogyakarta in 2021 showed an increase in the number of clean water customers by 8.20%. Data on river water quality in Indonesia during 2020-2022 in the National Urban Water Supply Project (NUWSP) article showed serious pollution in almost all rivers. Therefore, effective monitoring and accurate water quality predictions are needed. Machine learning algorithms, especially random forests, are used to process water physics and chemistry data to produce real-time water quality predictions. This study developed a prediction model using the cross-validation method which showed an average accuracy of 100% on training data and 91.43% on test data. Although there were indications of overfitting, the model with the 4th index was selected and showed a prediction accuracy of 97.93% on new data. These results indicate that the model is able to classify river water quality in DIY well. The main constraints are the limited dataset and the removal of labels that meet standard quality due to too few samples, as well as the uneven distribution of labels even though using the last four years of data.

Keywords*: River Water Quality, Machine Learning, Python, Random Forest.*