ABSTRACT

Drinking water is a very important thing to know, considering that water is easily contaminated and can harm the human body in the long run. There are several manual methods used to determine the feasibility of drinking water including STORET and Pollution Index. However, these two methods are not efficient enough because they take a long time, thus requiring a lot of human resources and high costs to be able to carry out testing. Therefore, a technology is needed that is able to classify the feasibility of drinking water quickly and accurately so that the classification of drinking water is more efficient.

To handle the problems described above, a classification using Machine Learning technology is proposed. Machine learning is chosen because machine learning is able to make decisions based on data patterns that have been studied or given before, so that in the case of drinking water eligibility machine learning can classify the eligibility of drinking water based on the data given to be studied by the machine. In this capstone project, three classification algorithms are applied, namely Decision Tree, K-Nearest Neighbors, and Extreme Learning Machine. In addition, the SMOTE technique is used to balance the data in the target class. The dataset used for the learning process comes from various sources, one of which is PDAM.

This research proves that the use of machine learning models with SMOTE resampling techniques is effective in classifying water quality. The Decision Tree and K-Nearest Neighbors (KNN) models showed excellent performance with accuracy and AUC values of 1.00. Meanwhile, the Extreme Learning Machine (ELM) model achieved a high accuracy of 98.6% with an AUC value of 1.00. The implementation of the model into the website successfully reduced the time and cost previously required for water quality evaluation.

Keywords : Classification Algorithm, Drinking Water Feasibility, Machine Learning, Model Accuracy, SMOTE.