

ABSTRAK

Klasifikasi status pencemaran air sungai memerlukan pendekatan analisis yang mampu menangani ketidakseimbangan data. Penelitian ini bertujuan untuk mengetahui kinerja algoritma *machine learning* dalam mengklasifikasikan status pencemaran air sungai di Kabupaten Banyumas. Variabel penelitian terdiri dari delapan parameter kualitas air, yaitu pH, BOD, COD, TSS, DO, nitrat, fosfat, dan fecal coliform, dengan variabel target berupa status mutu air yang diklasifikasikan ke dalam tiga kelas, yaitu memenuhi baku mutu, cemar ringan, dan cemar sedang. Metode analisis yang digunakan meliputi algoritma *Decision tree*, *Random Forest*, *Support Vector Machine* (SVM), dan *K-Nearest Neighbors* (KNN) dengan menggunakan teknik augmentasi data berbasis *Conditional Generative Adversarial Networks* (CGAN). Hasil penelitian menunjukkan bahwa penerapan augmentasi data CGAN memberikan pengaruh yang berbeda terhadap kinerja masing-masing algoritma. Algoritma *Random Forest* mengalami peningkatan akurasi dari 87,80% menjadi 93,98% dan menunjukkan performa terbaik dibandingkan algoritma lainnya. Peningkatan akurasi juga terjadi pada algoritma *Support Vector Machine* (SVM) dari 78,05% menjadi 85,54%, serta pada algoritma *K-Nearest Neighbors* (KNN) dari 75,61% menjadi 85,54%. Sebaliknya, algoritma *Decision tree* mengalami penurunan akurasi dari 95,12% menjadi 90,36% setelah diterapkan augmentasi CGAN. Penelitian ini menunjukkan bahwa efektivitas augmentasi data berbasis *Conditional Generative Adversarial Network* (CGAN) dalam klasifikasi pencemaran air sungai dipengaruhi oleh karakteristik algoritma yang digunakan, sehingga pemilihan algoritma menjadi faktor penting, dan pendekatan ini berpotensi dikembangkan lebih lanjut untuk meningkatkan keandalan hasil klasifikasi.

Kata kunci: pencemaran air, *machine learning*, CGAN, augmentasi data, klasifikasi

ABSTRACT

Classifying river water pollution status requires an analytical approach that can handle data imbalances. This study aims to determine the performance of machine learning algorithms in classifying river water pollution status in Banyumas Regency. The research variables consist of eight water quality parameters, namely pH, BOD, COD, TSS, DO, nitrate, phosphate, and fecal coliform, with the target variable being water quality status classified into three classes, namely meeting quality standards, mild pollution, and moderate pollution. The analysis methods used include Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) algorithms using Conditional Generative Adversarial Networks (CGAN)-based data augmentation techniques. The results show that the application of CGAN data augmentation has different effects on the performance of each algorithm. The Random Forest algorithm experienced an increase in accuracy from 87.80% to 93.98% and showed the best performance compared to other algorithms. Accuracy also increased in the Support Vector Machine (SVM) algorithm from 78.05% to 85.54%, as well as in the K-Nearest Neighbors (KNN) algorithm from 75.61% to 85.54%. Conversely, the Decision Tree algorithm experienced a decrease in accuracy from 95.12% to 90.36% after applying CGAN augmentation. This study shows that the effectiveness of Conditional Generative Adversarial Network (CGAN)-based data augmentation in river water pollution classification is influenced by the characteristics of the algorithm used, making algorithm selection an important factor. This approach has the potential to be further developed to improve the reliability of classification results.

Keywords: water pollution, machine learning, CGAN, data augmentation, classification