

ABSTRAK

Seiring dengan meningkatnya populasi dan perkembangan industri, pencemaran air telah menjadi isu serius yang menyebabkan penurunan kualitas air. Oleh karena itu, diperlukan metode yang efektif untuk memantau dan mengelompokkan kualitas air secara akurat. Penelitian ini bertujuan untuk membandingkan kinerja algoritma *K-Means* dan *K-Medoids* dalam pengelompokan kualitas air sungai di Yogyakarta berdasarkan parameter fisika, kimia, dan biologi. Dataset yang digunakan terdiri dari 852 data yang diperoleh dari Dinas Lingkungan Hidup Yogyakarta tahun 2020-2024. Tahapan penelitian meliputi *preprocessing* data dengan melakukan *Data Cleaning*, *Random Value Imputation (RVI)* dan seleksi fitur menggunakan ANOVA, dilanjutkan dengan penerapan algoritma *K-Means* dan *K-Medoids*. Evaluasi model dilakukan menggunakan *Silhouette Coefficient* dan waktu eksekusi (*running time*). Hasil penelitian menunjukkan bahwa *K-Means* menghasilkan nilai *Silhouette Score* tertinggi sebesar 0.5086 yang berada di $K=2$ dengan waktu eksekusi rata-rata lebih cepat yaitu 0.006s. Sementara itu, *K-Medoids* dengan $K=2$ menghasilkan nilai *Silhouette Score* 0.5073 dan memerlukan waktu rata-rata lebih lama yaitu 0.320s. Hasil *clustering* terbaik mengelompokkan data ke dalam dua label, yaitu C0 (Tidak Layak) yang terdiri dari 562 data dan C1 (Layak) yang terdiri dari 290 data.

Kata kunci : *K-Means, K-Medoids, Kualitas Air, Clustering, Silhouette Coefficient*

ABSTRACT

With the increase in population and industrial development, water pollution has become a serious issue that has led to a decline in water quality. Therefore, effective methods are needed to accurately monitor and classify water quality. This study aims to compare the performance of the K-Means and K-Medoids algorithms in classifying river water quality in Yogyakarta based on physical, chemical, and biological parameters. The dataset used consists of 852 data points obtained from the Yogyakarta Environmental Agency from 2020 to 2024. The research stages include data preprocessing through data cleaning, Random Value Imputation (RVI), and feature selection using ANOVA, followed by the application of the K-Means and K-Medoids algorithms. Model evaluation was conducted using the Silhouette Coefficient and execution time (running time). The results showed that K-Means produced the highest Silhouette Score, with a value of 0.5086 at $K=2$, and an average execution time of 0.006 seconds. Meanwhile, K-Medoids produced a Silhouette Score of 0.5073 and required a longer average time of 0.320s. The best clustering results grouped the data into two labels: C0 (Unsuitable) consisting of 562 data points and C1 (Suitable) consisting of 290 data points.

Keywords: K-Means, K-Medoids, Water Quality, Clustering, Silhouette Coefficient