

ABSTRAK

Dalam *machine learning*, *clustering* adalah metode untuk menganalisis data statistik dalam *unsupervised learning*. K-Means adalah metode yang sering digunakan untuk mengelompokkan data dengan cepat dan sederhana dengan mencari kombinasi variabel dan atribut objek. Namun, metode ini akan mengalami masalah jika data memiliki *outlier*.

Dalam penelitian ini, metode *Local Outlier Factor* (LOF) digunakan untuk mengatasi masalah anomali *outlier* yang diangkat oleh K-Means dengan mendeteksi *outlier* berbasis kepadatan dengan menghitung *local variance* dari titik data tertentu. LOF akan diterapkan setelah *preprocessing* data dan sebelum masuk proses *clustering* data K-Means.

Data dari kuisioner performa akademik mahasiswa sebanyak 210 data yang terbagi dalam tiga *cluster*, digunakan untuk pengujian. Dengan terdeteksi sebanyak 38 data atau 18% data *outlier*. Penerapan LOF meningkatkan *silhouette score* sebesar 10,23%. *Elbow method* juga digunakan dengan *silhouette score* untuk mendapatkan nilai K optimal = 3.

Kata kunci: *K-Means*, *Local Outlier Factor*, Klasterisasi, *Outlier*

ABSTRACT

In machine learning, clustering is a method for analyzing statistical data in unsupervised learning. K-Means is a well-known method for quickly and simply clustering data by searching for combinations of variables and object attributes. The algorithm will have issues if the data has outliers.

In this study, the Local Outlier Factor (LOF) method was used to solve the outlier anomaly concerns raised by K-Means by detecting density-based outliers by calculating the local variance of a particular data. Before the K-Means data clustering process, LOF will be applied.

Data from the distribution of the student academic performance questionnaire, totaling 210 data that were separated into three clusters, were used for the testing. With up to 38 data detected, this represents for 18% outlier data. The LOF approach increased the silhouette score by 10.23%. The elbow method is used with the silhouette score to obtain an optimal K value = 3.

Keywords: K-Means, Local Outlier Factor, Clustering, Outlier