

## ABSTRAK

Dalam *machine learning*, *clustering* adalah teknik yang digunakan untuk mengelompokkan data secara *unsupervised*. Salah satu algoritma yang umum digunakan untuk *clustering* adalah *K-Means*, yang bekerja dengan mencari kombinasi variabel dan atribut objek untuk mengelompokkan data dengan cepat dan sederhana. Namun, *K-Means* dapat mengalami masalah ketika terdapat *outlier* dalam data. Dalam penelitian ini, metode *Local Outlier Factor* (LOF) digunakan untuk mengatasi masalah *outlier* yang dihadapi oleh *K-Means*. LOF bekerja dengan mendeteksi *outlier* melalui perhitungan kepadatan titik data tertentu. Proses LOF diimplementasikan setelah proses *preprocessing* data dan sebelum proses *clustering* menggunakan *K-Means*. Data yang digunakan dalam penelitian ini adalah data UKT mahasiswa dari kampus, dengan total 317 data yang terbagi menjadi 8 golongan. Percobaan dengan menggunakan *K-Means* tanpa LOF menghasilkan akurasi sebesar 34,90%. Namun, setelah *K-Means* dipadukan dengan LOF, akurasi meningkat signifikan menjadi 98,30%. Dalam proses deteksi *outlier*, 30 data diidentifikasi sebagai *outlier*, yang menyebabkan peningkatan nilai *silhouette score* menjadi 0.69499 dalam pengujian model. Dengan demikian, penggunaan metode LOF sebelum *clustering* dengan *K-Means* terbukti efektif dalam meningkatkan akurasi dan kualitas model *cluster*.

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

## ABSTRACT

*In machine learning, clustering is a technique used to group data in an unsupervised manner. One commonly used algorithm for clustering is K-Means, which works by finding a combination of variables and object attributes to group data quickly and simply. However, K-Means can encounter problems when there are outliers in the data. In this research, the Local Outlier Factor (LOF) method is used to address the outlier problem faced by K-Means. LOF works by detecting outliers through the calculation of the density of certain data points. The LOF process is implemented after data preprocessing and before the clustering process using K-Means. The data used in this research is student UKT (tuition fee) data from a university, with a total of 317 data points divided into 8 categories. Experiments using K-Means without LOF resulted in an accuracy of 34.90%. However, after combining K-Means with LOF, the accuracy significantly increased to 98.30%. In the outlier detection process, 30 data points were identified as outliers, which led to an increase in the silhouette score to 0.69499 in model testing. Thus, using the LOF method before clustering with K-Means has proven effective in improving the accuracy and quality of the clustering model.*

*Keywords: Outlier, Local Outlier Factor, Clustering, K-means*