

INTISARI

METODE K-MEDOIDS DENGAN ALGORITME CLARANS PADA DATASET BESAR DENGAN PENCILAN

Oleh

Anastasya Viviana Gunawan

14/363747/PA/15856

K-medoids merupakan metode analisis kluster yang digunakan untuk mengelompokkan objek-objek ke dalam beberapa kelompok berdasarkan kesamaan karakteristik dengan *medoid* sebagai pusat klasternya, di mana *medoid* merupakan objek yang memiliki jarak paling minimum, sehingga tangguh terhadap pencilan. Algoritme *k-medoids* yang digunakan dalam penelitian ini adalah *Clustering Large Applications based on RANdomized Search* (CLARANS), di mana CLARANS merupakan perbaikan dari algoritme *Partitioning Around Medoid* (PAM) dalam menangani dataset besar dan perbaikan dari algoritme *Clustering Large Applications* (CLARA) dalam meningkatkan kualitas klustering. Metode CLARANS menggunakan dua batasan yaitu *numlocal*, untuk membatasi iterasi dan *maxneighbor*, untuk membatasi *neighbor* pada suatu *node*. Pengelompokkan didasarkan pada ukuran jarak Euclidean dan Manhattan. Selanjutnya untuk mengetahui tingkat validasi digunakan *silhouette width*.

Metode analisis kluster terbaik untuk mengelompokkan nilai ujian nasional SMA di Provinsi Jawa Barat, Jawa Tengah dan Jawa Timur pada tahun 2017 adalah metode CLARANS dengan jarak Manhattan, $k = 3$, *numlocal* = 2 dan *maxneighbor* = 44. Pada studi kasus, dapat diketahui dengan nilai *overall average silhouette width* bahwa metode CLARANS lebih baik dari metode PAM dan CLARA.

Kata kunci : *k-medoids*, *Clustering Large Applications based on RANdomized Search*, pencilan, dataset besar, *silhouette width*, *Partitioning Around Medoid*, *Clustering Large Applications*

ABSTRACT

K-MEDOIDS METHOD WITH CLARANS ALGORITHM FOR LARGE DATASET WITH OUTLIERS

By

Anastasya Viviana Gunawan

14/363747/PA/15856

K-medoids is a clustering method which used to cluster objects into several groups based on their characteristics similarity with the medoid as its center cluster, where medoid is an object that has the smallest minimum distance, which is robust to outliers. The k-medoids algorithm that used in this study is Clustering Large Applications based on RANdomized Search (CLARANS), where CLARANS is an improvement of the Partitioning Around Medoid (PAM) algorithm in handling large datasets and improvement of the Clustering Large Applications (CLARA) algorithm in improving quality of clusters. CLARANS method uses two parameters, named as *numlocal*, to limit the iteration and *maxneighbor*, to limit neighbors to a node. Clustering is based on Euclidean distance and Manhattan distance. Then, to determine the validation level used silhouette width as evaluation method.

The best clustering method for classifying high school national exam scores in West Java, Central Java and East Java in 2017 is CLARANS method with Manhattan distance, $k = 3$, *numlocal* = 2 and *maxneighbor* = 44. In the case study, can be known with the overall average silhouette width value that CLARANS method is better than PAM and CLARA methods.

Keywords: k-medoids, Large Clustering Applications based on RANdomized Search, outliers, large dataset, silhouette width, Partitioning Around Medoid, Clustering Large Applications