

INTISARI

ANALISIS KLASTER MENGGUNAKAN ALGORITMA I-CLARANS UNTUK DATASET BESAR DENGAN PENCILAN

Oleh

Kadek Dewi Yulianingsih

15/383361/PA/17021

Analisis kluster merupakan metode statistika multivariat yang bertujuan untuk mengelompokkan objek-objek yang memiliki kesamaan karakteristik ke dalam suatu kluster. *K-medoids* adalah metode analisis kluster dengan *medoid* sebagai pusat klusternya, di mana *medoid* merupakan objek yang letaknya terpusat di dalam suatu kluster, sehingga *robust* terhadap pencilan. Algoritma *k-medoids* yang digunakan dalam penelitian ini adalah *Improved Clustering Large Applications based on Randomized Search* (I-CLARANS), di mana I-CLARANS merupakan pengembangan dari algoritma *Clustering Large Applications based on Randomized Search* (CLARANS) dalam meningkatkan kualitas klustering. Metode I-CLARANS menggunakan dua batasan, yaitu *numlocal*, untuk membatasi iterasi dan *maxneighbour*, untuk membatasi *neighbour* pada suatu *node*. Pengelompokan didasarkan pada ukuran jarak Euclidean dan Manhattan. Selanjutnya, untuk mengetahui tingkat validasi digunakan *silhouette width*.

Metode analisis kluster terbaik untuk mengelompokkan beton-beton berdasarkan komponen yang digunakan adalah metode I-CLARANS dengan jarak Manhattan, $k = 6$, $numlocal = 2$, dan $maxneighbour = 80$. Pada studi kasus, dapat diketahui bahwa berdasarkan nilai *overall average silhouette width* metode I-CLARANS lebih baik dibandingkan metode CLARANS.

Kata kunci : *k-medoids*, *Improved Clustering Large Applications based on Randomized Search*, pencilan, dataset besar, *silhouette width*, *Clustering Large Applications based on Randomized Search*

ABSTRACT

CLUSTER ANALYSIS USING I-CLARANS ALGORITHM FOR LARGE DATASET WITH OUTLIERS

By

Kadek Dewi Yulianingsih

15/383361/PA/17021

Cluster analysis is a multivariate statistical method to classify objects that have similar characteristics into a cluster. K-medoids is a clustering method with the medoid as its center cluster, where medoid is the most centrally located object in a cluster, which is robust to outliers. The k-medoids algorithm that used in this study is Improved Clustering Large Applications based on Randomized Search (I-CLARANS), where I-CLARANS is development of Clustering Large Applications based on Randomized Search (CLARANS) algorithm in improving quality of clusters. I-CLARANS method uses two parameters, named as numlocal, to limit the iteration and maxneighbour, to limit neighbours 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 concretes based on components used is I-CLARANS method with Manhattan distance, $k = 6$, numlocal = 2, and maxneighbour = 80. In the case study, can be known based on the overall average silhouette width value that I-CLARANS method is better than CLARANS method.

Keywords: k-medoids, Improved Clustering Large Applications based on Randomized Search, outliers, large dataset, silhouette width, Clustering Large Applications based on Randomized Search