

INTISARI

METODE *K-MEDOIDS* PADA DATA DENGAN PENCILAN

Oleh

Annisa Larasati

13/348050/PA/15429

Analisis kluster adalah metode statistika multivariat yang bertujuan untuk mengelompokkan objek-objek yang memiliki kemiripan karakteristik ke dalam suatu kluster. *K-means* merupakan metode analisis kluster dengan menggunakan *mean* sebagai pusat klasternya. Namun, *mean* tidak *robust* terhadap adanya pencilan, sehingga algoritma *k-means* sangat sensitif terhadap data yang mengandung pencilan. Untuk mengatasi hal tersebut, dapat digunakan metode *k-medoids* untuk mengelompokkan data yang mengandung pencilan. *Medoid* merupakan objek yang letaknya terpusat di dalam suatu kluster, sehingga *robust* terhadap adanya pencilan. Metode *k-medoids* yang populer digunakan adalah *Partitioning Around Medoids* (PAM). Pada analisis kluster, objek-objek dikelompokkan berdasarkan kemiripannya. Untuk mengukur tingkat kemiripan tersebut digunakan ukuran jarak, yaitu jarak Euclidean dan jarak Manhattan. Selanjutnya, untuk mengetahui kualitas hasil analisis kluster dilakukan uji validasi dengan *silhouette width*.

Metode analisis kluster terbaik untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan produksi tanaman pangan pokok tahun 2015 adalah metode *k-medoids* dengan jarak Euclidean. Dapat diketahui pula bahwa metode *k-medoids* dengan jarak Euclidean lebih *robust* dibandingkan metode *k-means* dengan jarak Euclidean untuk mengelompokkan data dengan pencilan.

Kata kunci : *k-medoids*, *Partitioning Around Medoids*, pencilan, *silhouette width*

ABSTRACT

K-MEDOIDS METHOD FOR DATA WITH OUTLIERS

By

Annisa Larasati

13/348050/PA/15429

Cluster analysis is a multivariate statistical methods to classify objects that have similar characteristics into a cluster. K-means is a clustering method using mean as its cluster center. However, mean is not robust to the presence of outliers, so k-means algorithm is sensitive for data with outliers. To overcome this problem, k-medoids methods can be used to classify data with outliers. Medoid is the most centrally located object in a cluster, so it's robust to outliers. One of the popular methods for k-medoids is Partitioning Around Medoids (PAM). In cluster analysis, the objects are grouped by the similarity. To measure the similarity, it can be used distance measures, Euclidean distance and Manhattan distance. Then, to determine the quality of the clustering results can be used validity index with silhouette width.

The best clustering method to classify the provinces in Indonesia based on staple food production in 2015 is k-medoids with Euclidean distance. Furthermore, it can be concluded that k-medoids with Euclidean distance is more robust than k-means with Euclidean distance to classify data with outliers.

Keywords : k-medoids, Partitioning Around Medoids, outlier, silhouette width