

INTISARI

ANALISIS KLASTER MENGGUNAKAN METODE CLARA PADA DATA YANG MENGANDUNG PENCILAN

Oleh:

Ahmad Bukhari Muslim

14/364256/PA/15968

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. Salah satu metode *k-medoids* yang populer digunakan adalah *Clustering Large Application* (Clara). Metode Clara juga baik digunakan pada data dalam jumlah besar. 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 film-film berdasarkan popularitas pada tahun 2009 sampai 2016 adalah metode Clara dengan jarak Manhattan. Dapat diketahui pula bahwa metode Clara dengan jarak Manhattan lebih *robust* dibandingkan metode *K-Means* dengan jarak Euclidean dan metode Pam dengan jarak Manhattan untuk mengelompokkan data dengan pencilan dan data dalam jumlah besar.

Kata kunci: *k-medoids, Clustering Large Application, pencilan, silhouette width*

ABSTRACT

CLUSTER ANALYSIS USING CLARA METHOD FOR DATA WITH OUTLIERS

By:

Ahmad Bukhari Muslim

14/364256/PA/15968

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 Clustering Large Application (Clara). Clara method is also good to use for data with large quantities. 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 film based on popularity in 2009 until 2016 is Clara with Manhattan distance. Furthermore, it can be concluded that Clara with Manhattan distance is more robust than K-Means with Manhattan distance and Pam method with Manhattan distance to classify data with outliers and data with large quantities.

Keywords: *k-medoids, Clustering Large Application, pencilan, silhouette width*