

ABSTRAK

Analisis kluster merupakan salah satu metode multivariat untuk menggabungkan objek ke dalam beberapa kelompok atau kluster, di mana anggota kluster itu tidak diketahui sebelumnya. Dengan kata lain, analisis kluster merupakan analisis statistik yang digunakan untuk mengelompokkan n observasi ke dalam k buah kluster dengan setiap objek dalam kluster tersebut memiliki kemiripan yang tinggi dibandingkan objek antar kluster. Tidak dapat dipungkiri terkadang data yang akan diolah ke dalam beberapa kluster adalah data yang besar dan mengandung pencilan serta multikolinearitas. Untuk mengolah data tersebut, metode yang digunakan pada kajian ini adalah *Clustering Large Applications* (CLARA) yang sudah tahan terhadap pencilan, berkerja dengan baik pada himpunan data yang besar serta proses pengklusterannya lebih cepat. Proses pengambilan sampelnya secara acak dan pengolahan untuk menemukan medoid terbaiknya dengan PAM (*Partitioning Around Medoids*). Untuk validasinya dapat dilihat dari nilai rata-rata *silhouette width* atau biasa disebut nilai *silhouette coefficient*nya. *Silhouette coefficient* menggabungkan ide *cohesion* dan *separation* di mana *cohesion* digunakan untuk mengukur seberapa dekat hubungan objek-objek pada kluster yang sama sedangkan *separation* digunakan untuk mengukur seberapa berbeda atau terpisahnya sebuah kluster dengan kluster lainnya. Semakin tinggi nilai *silhouette coefficient* pada suatu analisis kluster maka semakin akurat pengelompokan objek tersebut dan kriteria pengujian yang dilakukan berdasarkan interpretasi Kaufmann dan Rousseeuw.

Kata Kunci : Analisis Kluster, PAM, CLARA, *Silhouette Coefficient*

ABSTRACT

Cluster analysis is one of the multivariate methods for incorporating objects into groups or clusters, where cluster members are not previously known. In other words, cluster analysis is a statistical analysis used to group n observations into k clusters with each object in the cluster having a high resemblance compared to clustered objects. It can not be denied that sometimes the data to be processed into multiple clusters is large data and contains both outlier and multicollinearity. To process the data, the method used in this study is Clustering Large Applications (CLARA) that has been resistant to outlier, work well on large data sets and clustering process faster. The process of random sampling and processing to find the best medoid with PAM (Partitioning Around Medoids). For validation can be seen from the average value of silhouette width or so-called silhouette coefficientnya value. Silhouette coefficient combines cohesion and separation ideas in which cohesion is used to measure how closely the relationships of objects in the same cluster whereas separation is used to measure how different or clustered a cluster is with another cluster. The higher the value of silhouette coefficient in a cluster analysis the more accurate the grouping of the object and the criteria of the test based on the Kaufmann and Rousseeuw interpretations.

Keywords: *Cluster Analysis, PAM, CLARA, Silhouette Coefficient*