

## **ABSTRACT**

### **ANALYSIS CLUSTER ROBUST USING K-MEDOIDS METHOD IN THE DATA CONTAINING OUTLIER**

**By**

**EPMI ANNISA NUR FAJRINA**

Cluster analysis is a multivariate analysis technique that is useful for grouping observational data or variables into clusters in such a way that each cluster is homogeneous according to the factors used for clustering. There are two assumptions that must be met in cluster analysis, namely a representative sample and non-multicollinearity between each variable. In fact, not all data meet the two assumptions above, especially for the assumption of a representative sample because of the problem of outliers. To overcome the problem of outliers, there is one method of non-hierarchical cluster analysis that is resistant to outliers, namely K-Medoids. K-Medoids method including the sealing technique (partition) which divides or separates objects to different groups so that objects with the same characteristics are in the same cluster and objects with different characteristics are in another clusters. This study aims to examine robust cluster analysis with the K-medoids method through simulation data with several percentages of outlier values and see the effectiveness of the K-medoids method to overcome outliers in terms of the sillhouette index and Dunn index validation values, and it is concluded that the analysis the cluster K-medoids method is an effective clustering method used on data that contains outliers. Based on the average value of the sillhouette index and Dunn index, the results show that the K-medoids method is effectively used for clustering data with a small number of objects, a small number of clusters, and a large proportion of outliers.

**Keywords:** Outliers, Cluster Analysis, K-Medoids, sillhouette index, Dunn index

## ABSTRAK

### ANALISIS *CLUSTER ROBUST* MENGGUNAKAN METODE *K-MEDOIDS* PADA DATA MENGANDUNG PENCILAN

Oleh

**EPMI ANNISA NUR FAJRINA**

Analisis kluster merupakan suatu teknik analisis multivariat yang berguna untuk mengelompokkan data observasi ataupun variabel-variabel ke dalam kluster sedemikian rupa sehingga masing-masing kluster bersifat homogen sesuai dengan faktor yang digunakan untuk melakukan pengklasteran. Ada dua asumsi yang harus dipenuhi pada analisis kluster, yaitu sampel representatif dan tidak ada multikolinearitas antara tiap variabel. Pada kenyataannya, tidak semua data memenuhi kedua asumsi di atas, khususnya untuk asumsi sampel representatif karena adanya masalah pencilan. Untuk mengatasi masalah pencilan tersebut, terdapat salah satu metode analisis kluster non hierarki yang tahan terhadap pencilan yaitu *K-Medoids*. Metode *K-Medoids* termasuk dalam teknik penyekatan (*partition*) yang membagi atau memisahkan objek ke-*k* kelompok yang berbeda sehingga objek yang memiliki karakteristik yang sama masuk ke dalam satu kluster yang sama dan objek yang mempunyai karakteristik berbeda masuk ke dalam kluster yang lain. Penelitian ini bertujuan untuk mengkaji analisis kluster *robust* dengan metode *K-medoids* melalui data simulasi dengan beberapa persentase nilai pencilan dan melihat keefektifan metode *K-medoids* untuk mengatasi pencilan ditinjau dari nilai validasi *sillhouette index* dan *Dunn index*, dan diperoleh kesimpulan bahwa analisis kluster metode *K-medoids* merupakan metode pengklasteran yang efektif digunakan pada data yang mengandung pencilan. Berdasarkan rata-rata nilai *sillhouette index* dan *Dunn index* diperoleh hasil bahwa metode *K-medoids* efektif digunakan untuk pengklasteran pada data dengan jumlah objek kecil, jumlah kluster kecil, dan proporsi pencilan yang besar.

**Kata kunci:** Pencilan, Analisis Kluster, *K-Medoids*, *sillhouette index*, *Dunn index*