

INTISARI

PENGUNAAN INDEKS DUNN UNTUK PENENTUAN JUMLAH KLASTER OPTIMAL PADA METODE CLARA

Oleh
Raissa Nuzul Muzaffar
(19/439126/PA/19039)

Analisis kluster merupakan salah satu metode yang terdapat pada statistika multivariat yang bertujuan untuk mengelompokkan objek-objek yang tidak diketahui kelompoknya menjadi beberapa kelompok atau kluster dengan memperhatikan karakteristik dari data yang dimiliki sehingga objek-objek dalam suatu kelompok memiliki homogenitas yang tinggi, namun akan bersifat heterogen untuk anggota kelompok lainnya. Analisis kluster terbagi menjadi *hierarchical clustering* dan *partitional clustering*. Metode *partitional* memiliki kelebihan dibandingkan metode hirarki yaitu mampu menganalisis sampel yang berukuran besar dan dapat menangani penciran pada data. Salah satu metode *partitional* yaitu *Clustering Large Application* yang dikenalkan oleh Kaufman dan Rousseeuw pada tahun 1990 dan merupakan pengembangan dari metode *Partition Around Medoids*. Pada analisis dengan metode *partitional*, salah satu hal yang penting dilakukan yaitu menentukan jumlah kluster terlebih dahulu sebelum melakukan analisis. *Dunn Index* merupakan metode validasi yang diperkenalkan oleh J. C. Dunn pada tahun 1994 yang mana metode ini memberikan skor terbaik dalam analisis kluster untuk menghasilkan kluster dengan variansi yang kecil pada anggota dalam kluster, namun menghasilkan variansi yang besar untuk anggota kluster lainnya. Setelah dilakukan analisis terhadap data nasabah kartu kredit, hasil kluster dengan metode *Clustering Large Application* ukuran jarak Manhattan diperoleh jumlah kluster sebanyak tiga kluster yaitu kluster nasabah pemakaian kartu kredit rendah, sedang dan tinggi. Kluster tersebut diperoleh dari metode *Dunn Index* yang merupakan hasil kluster yang optimal dengan nilai validasi tertinggi sebesar 0,7221353.

Kata kunci: Analisis Kluster, Jumlah Kluster, Dunn Index, Clustering Large Application

ABSTRACT

APPLICATION OF DUNN INDEX FOR DETERMINING THE OPTIMAL NUMBER OF CLUSTERS USING THE CLARA METHOD

By

Raissa Nuzul Muzaffar
(19/439126/PA/19039)

Cluster analysis is one of method in multivariate statistics which aims to grouping objects whose groups are unknown into several groups or clusters by observe to the characteristics of the data so the objects in a group have high homogeneity, but will be heterogeneous with other group members. Cluster analysis is divided into hierarchical clustering and partitional clustering. Partitional method has advantages over the hierarchical method, which is capable to analyze large samples and handle outliers in the data. One of partitional method is Clustering Large Application which was introduced by Kaufman and Rousseeuw in 1990 and it is a development of the Partition Around Medoids method. In analysis using the partitional method, one of the important things to do is determine the number of clusters first before carrying out the analysis. The Dunn Index is a validation method introduced by J. C. Dunn in 1994, where this method provides the best score in cluster analysis to produce clusters with small variances for members in the cluster, but produces large variances for other cluster members. After analyzing the Credit Card Customer data, the cluster results using the Clustering Large Application method measuring the Manhattan distance resulted in three clusters, namely low, medium and high credit card usage customer clusters. This cluster was obtained from the Dunn Index method which is an optimal cluster result with the highest validation value of 0.7221353.

Keywords: Cluster Analysis, Number of Clusters, Dunn Index, Clustering Large Application