

INTISARI

Analisis Klaster Hierarki untuk Data Kategorik dengan Algoritma *Divisive Hierarchical Clustering of Categorical Data* (DHCC) (Studi Kasus: Pengelompokan Anggota UKM Mapagama)

Oleh

Muhammad Burhan Aziz

17/412744/PA/18063

Analisis klaster adalah analisis yang bertujuan untuk mengelompokkan objek-objek yang memiliki tingkat kemiripan tinggi ke dalam suatu kelompok. Analisis klaster pada data kategorik menjadi lebih kompleks karena tidak ada ukuran kemiripan yang memiliki makna berarti diantara objek kategorik. Pada tugas akhir ini, akan dipaparkan suatu metode analisis klaster untuk data kategorik, yaitu algoritma DHCC. Algoritma ini berjalan dengan pendekatan atas-bawah. Pemisahan klaster berdasarkan analisis multikorespondensi pada matriks indikator yang menunjukkan kemunculan nilai kategorik pada suatu objek. Total inersia atau rata-rata jarak *Chi-square* pada perhitungan analisis multikorespondensi yang berupa jumlah kuadrat dari elemen matriks residual standar dapat dipandang sebagai ukuran heterogenitas suatu klaster dari prespektif analisis klaster. Dimensi pertama dari ruang yang ditransformasi berdasarkan MCA menjelaskan variansi terbesar dari rata-rata jarak *Chi-square* sehingga pemisahan klaster berdasarkan dimensi pertama akan menurunkan rata-rata jarak *Chi-Square*. Unsur optimisasi pada algoritma DHCC berjalan dengan memperbaiki pemisahan hasil dari analisis multikorespondensi. Metode ini diaplikasikan untuk melakukan segmentasi anggota UKM Mapagama periode 2020/2021. Diperoleh 8 segmen dari anggota UKM Mapagama dengan masing-masing karakteristiknya. Optimisasi yang

terdapat pada algoritma DHCC terbukti meningkatkan kualitas hasil analisis klaster yang lebih baik.

Kata kunci : Analisis klaster hierarki, Data kategorik, Analisis multikorespondensi, Algoritma DHCC, Segmentasi.

ABSTRACT

Hierarchical Clustering Analysis for Categorical Data Using Divisive Hierarchical Clustering of Categorical Data (DHCC) Algorithm (Case Study: Grouping Members of UKM Mapagama)

By

Muhammad Burhan Aziz

17/412744/PA/18063

Cluster analysis is an analysis that aims to group objects that have a high level of similarity into a group. Cluster analysis on categorical data becomes more complex because there is no significant similarity measure between categorical objects. In this final project, a cluster analysis method for categorical data will be presented, namely the DHCC algorithm. This algorithm runs with a top-down approach. Cluster separation is based on multiple correspondence analysis on an indicator matrix that shows the occurrence of categorical values in an object. The total inertia or the average Chi-square distance in the calculation of multiple correspondence analysis is the sum squares of elements from the standard residual matrix elements can be viewed as a measure of the heterogeneity of a cluster from the perspective of cluster analysis. The first dimension of the transformed space based on MCA explains the largest variance of the average Chi-square distance so that the cluster separation based on the first dimension will decrease the average Chi-Square distance. The optimization element in the DHCC algorithm works by improving the separation of results from multiple correspondence analysis. This method was applied to segment Mapagama UKM members for the 2020/2021 period. Obtained 8 segments from members of UKM Mapagama with their respective characteristics. The optimization contained in the DHCC algorithm is proven to improve the quality of the results of better cluster analysis.

Keywords : Hierarchical cluster analysis, Categorical data, Multiple correspondence analysis, DHCC algorithm, Segmentation.