

INTISARI

Algoritma *K-modes* dengan Tambahan Informasi Antar Klaster

oleh

Farrah Gita Nurmala

15/383356/PA/17016

K-modes merupakan teknik analisis klaster untuk mengelompokkan data kategorik. Performa dari hasil akhir analisis bergantung pada pusat klaster awal. Pada umumnya pusat klaster awal dipilih secara random dan hanya akan baik jika pusat klaster yang dipilih dekat dengan nilai klaster yang sesungguhnya. Maka diusulkan inisialisasi pusat klaster awal dengan memperhatikan jarak dan densitas dari data. Dalam algoritma *k-modes*, pada umumnya hanya memperhatikan informasi dalam klaster dan mengabaikan informasi antar klaster yang mungkin akan mengakibatkan lemahnya pemisahan antar klaster. Penelitian ini akan memperhatikan informasi antar klaster dalam algoritma *k-modes*. Metode yang diusulkan akan diaplikasikan pada dataset *Soybean (Small)*. Analisis klaster yang menggunakan inisialisasi pusat klaster awal dan memperhatikan informasi antar klaster memiliki hasil yang lebih baik dan lebih akurat.

Kata kunci : *k-modes*, inisialisasi pusat klaster awal, analisis klaster, data kategorik

ABSTRACT

K-modes Clustering Algorithm Plus Between-Cluster Information

by

Farrah Gita Nurmala

15/383356/PA/17016

K-modes algorithm is a cluster analysis technique used to cluster categorical data. The performance of the algorithm depends on initial cluster center. Usually, initial cluster center is chosen randomly from the data and it will work well only if that initial choice is close to a good solution. This research aims to initialize initial cluster center which the distance and the density between the object considered. In the k-modes algorithm, generally computed based on within-cluster information only and the between-cluster information is not considered, which may result in the clustering result with weak separation among different clusters. This research aims to add the between-cluster information to the *k*-modes algorithm. The proposed method will be applied to the Soybean (Small) dataset. Cluster analysis using initial cluster center initialization and between-cluster information has better and more accurate results.

Keywords : k-modes, initial cluster center, cluster analysis, categorical data