

**ANALYSIS CLUSTER FOR CATEGORICAL DATA USING K-MODES
METHOD
(STUDY CASE: COVID-19 PATIENTS DATA AT RSUD DR. H. ABDUL
MOELOEK LAMPUNG PROVINCE)**

By

SHABRINA NOVAINDAH DWIYAMTI

Cluster analysis is a multivariate technique with interdependence because there is no difference between independent variable and dependent variable. Cluster analysis consists of hierarchical and non hierarchical methods. K-Means is a non hierarchical cluster analysis method. However, the K-Means method is limited to numerical data. So we need a method to process the categorical data. One of the non hierarchical methods for categorical data often used is K-Modes. This study aims to apply K-Modes cluster analysis to COVID-19 patients data at RSUD Dr. H. Abdul Moeloek Lampung Province, totaling 560 patient data with variables of gender, age, admission, and their condition when leaving RSUD Dr. H. Abdul Moeloek Lampung Province. By using Davies-Bouldin Index (DBI) and Silhouette method, the optimal k value for the K-Modes cluster analysis is 8 clusters, consisting of 145 patients in the first cluster, 227 patients in the second cluster, 16 patients in the third cluster, 30 patients in the fourth cluster, 30 patients in the fifth cluster, 74 patients in the sixth cluster, 4 patients in the seventh cluster, dan 34 patients in the eighth cluster. Because there are more members of clusters 1 and 2 when compared to other clusters, more optimal handling is needed for clusters 1 and 2.

Keywords: Cluster Analysis, K-Modes, COVID-19, DBI, Silhouette Method.

**ANALISIS KLASTER UNTUK DATA KATEGORIK MENGGUNAKAN
METODE *K-MODES*
(STUDI KASUS: DATA PASIEN COVID-19 DI RSUD DR. H. ABDUL
MOELOEK PROVINSI LAMPUNG)**

Oleh

SHABRINA NOVAINDAH DWIYAMTI

Analisis klaster merupakan salah satu analisis multivariat metode interdependensi dikarenakan tidak ada perbedaan antara variabel bebas dan variabel tak bebas. Analisis klaster terdiri dari metode hierarki dan non hierarki. *K-Means* merupakan salah satu metode analisis klaster non hierarki. Namun, metode *K-Means* terbatas pada data numerik. Sehingga dibutuhkan metode untuk mengolah data kategorik. Salah satu metode non hierarki untuk data kategorik yang sering digunakan adalah *K-Modes*. Penelitian ini bertujuan untuk menerapkan analisis klaster *K-Modes* pada data pasien COVID-19 di RSUD Dr. H. Abdul Moeloek Provinsi Lampung yang berjumlah 560 data pasien dengan variabel jenis kelamin, usia, cara masuk, dan kondisi saat keluar dari RSUD Dr. H. Abdul Moeloek Provinsi Lampung. Dengan menggunakan Davies-Bouldin *Index* (DBI) dan metode *Silhouette*, diperoleh hasil nilai k optimal untuk analisis klaster *K-Modes* adalah sebanyak 8 klaster, yaitu terdiri dari 145 pasien klaster 1, 227 pasien klaster 2, 16 pasien klaster 3, 30 pasien klaster 4, 30 pasien klaster 5, 74 pasien klaster 6, 4 pasien klaster 7, dan 34 pasien klaster 8. Karena anggota klaster 1 dan 2 terbanyak jika dibandingkan dengan klaster lainnya, maka diperlukan penanganan yang lebih optimal untuk klaster 1 dan 2.

Kata Kunci: Analisis Klaster, *K-Modes*, COVID-19, DBI, Metode *Silhouette*.