

Cluster Ensemble pada Data Campuran dalam Pengelompokan Sekolah Menengah Pertama di Provinsi Jawa Barat = Cluster Ensemble Based Mixed Data Clustering of Junior High School in West Java Province

Bayu Permata Negara, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20509668&lokasi=lokal>

Abstrak

Analisis kelompok adalah metode multivariat yang bertujuan mengelompokkan pengamatan berdasarkan karakteristiknya. Salah satu metode analisis pengelompokan adalah metode cluster ensemble dengan pengelompokan dilakukan dengan satu metode berulang kali hingga diperoleh hasil yang lebih baik dibandingkan jika dilakukan satu kali. Penelitian ini mencoba menggunakan Cluster Ensemble Based Mixed Data Clustering (CEBMDC), yaitu metode pengelompokan yang biasa dilakukan untuk data dengan variabel campuran yaitu numerik dan kategorik. Tahap awal dalam metode ini yaitu membagi data awal menjadi data dengan hanya variabel-variabel numerik dan data dengan hanya variabel-variabel kategorik. Data yang telah dipisahkan berdasarkan jenis variabelnya kemudian dikelompokkan menggunakan metode yang sesuai secara simultan. Hasil pengelompokan ini menjadi data baru dengan dua variabel kategorik yaitu hasil pengelompokan dengan variabel numerik dan hasil pengelompokan dengan variabel kategorik. Data baru dengan dua variabel kategorik ini kemudian dilakukan proses pengelompokan. Metode pengelompokan untuk data dengan variabel numerik adalah metode Hierarchical Agglomerative Clustering. Metode clustering untuk data kategorik adalah ROCK (RObust Clustering using linKs) dan K-medoids/PAM (Partition Around Medoids). Penelitian ini membandingkan hasil pengelompokan ROCK dan K-medoids. Pengelompokan dilakukan pada data mengenai sarana dan prasarana sekolah yang diambil dari 5.094 SMP yang ada di Jawa barat. Metode pengelompokan dengan kinerja terbaik pada penelitian ini adalah Ensemble K-medoids berdasarkan rasio antara simpangan baku di dalam kelompok ($\hat{\sigma}^2$ -SW) dan simpangan baku antar kelompok (SB) terkecil. Penelitian ini menghasilkan 3 kelompok yang mencerminkan kondisi sekolah-sekolah pada jenjang SMP di Jawa Barat.

Clustering analysis is a multivariate method that aims to classify observations based on their characteristics. One method of clustering analysis is the ensemble clustering method in which the grouping is done using a method repeatedly until better results are obtained than if it is done once. This study uses the Cluster Ensemble Based Mixed Data Clustering (CEBMDC), which is a grouping method that commonly used for data with numerical and categorical variables. The first step in this method is to divide the initial data into two parts, that is data with only numerical variables and data with categorical variables. After data has been separated based on the types of variables, and then clustering using the appropriate method is conducted simultaneously. The results of these two clustering method become a new data with two categorical variables, namely the results of clustering with numeric variables and the results of clustering with categorical variables. The new data with two categorical variables are then carried out the clustering process. The clustering method for data with numerical variables is the Hierarchical Agglomerative Clustering method. Clustering methods for categorical data are ROCK (RObust Clustering using linKs) and K-medoids / PAM (Partition Around Medoids). This study compares the results of ROCK and K-medoids clustering. The study was conducted on data of school facilities and infrastructure taken from 5094 junior high schools in West Java. The best performance grouping method in this study is the Ensemble K-medoids based on the ratio between the standard deviation in the group (SW)

and the smallest standard inter-group (SB) deviation. This study produced 3 groups that reflect the condition junior high schools in West Java.