

ABSTRAK

Penambangan data (*data mining*) merupakan proses penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah besar. Salah satu metode dalam *data mining* adalah *clustering*. *Clustering* adalah proses mengelompokkan sejumlah data/obyek ke dalam klaster (*group*) sehingga dalam setiap klaster akan berisi data yang semirip mungkin. Salah satu algoritma *clustering* adalah *Partitioning Around Medoids* (PAM). PAM adalah metode *clustering* menggunakan metode partisi untuk mengelompokkan sekumpulan n obyek menjadi sejumlah k *cluster*. Algoritma PAM menggunakan obyek pada kumpulan obyek untuk mewakili sebuah *cluster*. Obyek yang terpilih untuk mewakili sebuah *cluster* disebut dengan *medoids*.

Pada tugas akhir ini akan diimplementasikan algoritma PAM untuk mengelompokkan data Sekolah Menengah Atas di DIY berdasarkan nilai daya serap ujian nasional mata pelajaran matematika tahun ajaran 2014/2015 menjadi beberapa kelompok sehingga diperoleh informasi mengenai kelompok-kelompok sekolah di DIY. Proses penambangan data yang dilakukan adalah proses KDD (*Knowledge Discovery in Database*) yaitu pembersihan data, integrasi data, seleksi data, transformasi data, penambangan data, evaluasi pola, dan presentasi pengetahuan.

Penambangan data dan evaluasi dalam penelitian menggunakan dua jenis data yaitu data nilai daya serap 29 kompetensi dan data rata-rata dari nilai daya serap 29 kompetensi. Berdasarkan evaluasi menggunakan nilai daya serap 29 kompetensi dapat disimpulkan bahwa algoritma ini dapat digunakan untuk mengelompokkan data sekolah dengan nilai k yang diberikan. Berdasarkan evaluasi menggunakan rata-rata dari nilai daya serap 29 kompetensi dapat disimpulkan bahwa algoritma ini dapat mengelompokkan nilai rata-rata daya serap ke dalam tiga kelompok yaitu kelompok dengan standar deviasi tinggi, sedang, dan rendah.

Kata kunci : penambangan data, *clustering*, *Partitioning Around Medoids* (PAM), *medoids*, KDD, nilai daya serap ujian nasional.

ABSTRACT

Data mining is a mining process or extracting information by getting pattern or specific rules from a large amount of data. One of the data mining method is clustering. Clustering is a process of grouping data/objects into a cluster (group) so in that every cluster contains data as closely as possible. One of clustering algorithm is Partitioning Around Medoids (PAM). PAM is a clustering method that used partitioning method for cluster some n object to k cluster. PAM algorithm used object on the set of objects for representing a cluster. The chosen object for representing a cluster called as medoids.

This thesis will implement of PAM algorithm for clustering high school data in DIY based on the value of absorption data of national exam score mathematics subject in 2014/2015 become some of group so that obtained the information about high school groups in DIY. Process of data mining which is conducted is KDD (Knowledge Discovery in Database) process that is data cleaning, data integration, selection data, transformation data, data mining, pattern evaluation, and knowledge presentation.

Data mining and evaluation in research using two types of data that is value of absorption data 29 comprehension, and data of the average of value of absorption data 29 comprehension. Based on the evaluation using value of absorption data 29 comprehension can be concluded that the algorithm can be used to classify the data of high school by giving the value of k . Based on the evaluation using the average of value of absorption data 29 comprehension can be concluded that this algorithm can be used for grouping the value of the average of absorption data into three groups which is group with high, medium, and low standard deviation.

Keyword : Data mining, *Clustering*, *Partitioning Around Medoids* (PAM), *Medoids*, KDD, value of absorption data of national exam score.