

ABSTRAK

Penambangan data (*data mining*) merupakan proses penemuan informasi otomatis dengan mengidentifikasi pola dari set data atau basis data besar. Proses penemuan informasi tersebut dapat dilakukan dengan metode pengelompokan data ke dalam beberapa kelompok dari sebuah set data yang dalam penambangan data disebut metode *clustering*. *Clustering* merupakan proses mempartisi data-set menjadi beberapa sub-set atau kelompok berdasarkan kesamaan karakteristik masing-masing data pada kelompok-kelompok yang ada. Salah satu algoritma *clustering* yaitu *Agglomerative Hierarchical Clustering (AHC)* di mana algoritma ini merupakan algoritma pengelompokan berbasis hirarki dengan pendekatan *bottom up*, yaitu proses pengelompokan dimulai dari masing-masing data sebagai satu buah *cluster*, kemudian secara rekursif mencari *cluster* terdekat sebagai pasangan untuk bergabung sebagai satu *cluster* yang lebih besar.

Pada tugas akhir ini diimplementasikan algoritma *Agglomerative Hierarchical Clustering* metode *single linkage*, *complete linkage*, dan *average linkage* untuk mengelompokkan Sekolah Menengah Atas di Provinsi DIY berdasarkan nilai Ujian Nasional tahun 2015. Hasil pengelompokan dapat digunakan untuk membantu penugasan pengawas sekolah untuk mendampingi sekolah. Dalam melakukan proses mengubah data mentah menjadi suatu informasi yang lebih bermanfaat, penulis menggunakan proses *Knowledge Discovery in Database (KDD)* yang terdiri dari pembersihan data, integrasi data, seleksi data, transformasi data, penambangan data, evaluasi pola, dan presentasi pengetahuan. Pada tahap awal KDD akan dilakukan pembersihan data dan integrasi data secara manual melalui aplikasi *spreadsheet*. Selanjutnya dilakukan seleksi data, transformasi data, dan penambangan data dengan menggunakan perangkat lunak yang dibuat. Tahap selanjutnya yaitu evaluasi pola dan presentasi pengetahuan.

Berdasarkan penelitian yang telah dilakukan, dapat diketahui bahwa algoritma *Agglomerative Hierarchical Clustering (AHC)* dapat digunakan untuk mengelompokkan Sekolah Menengah Atas di Provinsi DIY berdasarkan nilai Ujian Nasional. Hasil evaluasi *cluster* menunjukkan bahwa pengelompokan menggunakan algoritma AHC dengan metode *single linkage* memiliki struktur yang kuat. Sedangkan hasil pengelompokan menggunakan metode *complete linkage* dan *average linkage* memiliki struktur pengelompokan yang baik.

Berdasar peraturan Menteri Pendidikan dan Kebudayaan no.143 tahun 2014 tentang petunjuk teknis pelaksanaan jabatan fungsional pengawas sekolah dan angka kreditnya di mana pengawas bertugas untuk mendampingi minimal 7 sekolah, pada kasus ini tidak dapat menggunakan metode *single linkage* karena selalu menghasilkan kelompok yang memiliki 1 anggota. Metode *complete linkage* memberikan hasil yang efektif jika kelompok pendampingan dibagi menjadi 2 dan 3 kelompok untuk jurusan IPA, sedangkan untuk jurusan IPS dibagi menjadi 2 sampai 6 kelompok. Metode *average linkage* memberikan hasil yang efektif jika kelompok pendampingan dibagi menjadi 2 kelompok untuk jurusan IPA, sedangkan untuk jurusan IPS dibagi menjadi 2 sampai 4 kelompok.

Kata Kunci – Penambangan Data, Algoritma Agglomerative Hierarchical Clustering, Nilai Ujian Nasional

ABSTRACT

Data mining is a process to find information by identifying pattern from data set or big database automatically. The finding process can be done by clustering data to some clusters from data set in data mining called clustering method. Clustering is known as a partition process of data set into some sub-sets or clusters, based on the same characteristics of each data in clusters. One of clustering algorithm is Agglomerative Hierarchical Clustering (AHC), which is a clustering algorithm based on hierarchy with bottom up approach. Bottom up approach is a clustering process that start from each data as one cluster, then find the closest cluster in recursive to be merged into one bigger cluster.

This undergraduate thesis implemented Agglomerative Hierarchical Clustering algorithm by using single linkage method, complete linkage method, and average method to cluster Senior High Schools in Special Region of Yogyakarta Province based on the score of national examination in 2015. The clustering result can be used to help the assignment of school supervisors in assisting schools. In the process of converting raw data into useful information, the writer used Knowledge Discovery in Database (KDD) process which consist of data cleansing, data integration, data selection, data transformation, data mining, pattern of evaluation, and knowledge presentation. In the early stage of KDD process, the writer performed data cleansing and data integration manually by using spreadsheet application. The next processes are data selection, data transformation, and data mining, which were done by using the invented software. The last steps are pattern evaluation and knowledge presentation.

Based on the research that has been done, it can be concluded that Knowledge Discovery in Database (KDD) algorithm can be used for clustering Senior High School in Special Region of Yogyakarta Province based on the score of national examination. The result of cluster evaluation showing that the clustering process using AHC algorithm with single linkage method has a strong clustering structure, while complete linkage and average linkage have a good clustering structure.

Based on the regulation of Ministry of Education and Culture no. 143/2014 one school supervisor should observe seven schools at minimum. Therefore, the single linkage method is not appropriate because there always be a cluster that have one member. For natural science major, the complete linkage method gives effective result when the schools are divided into 2 and 3 clusters, while for social science major the schools are divided into 2 until 6 clusters. The average linkage method gives effective result when the schools are divided into 2 clusters for natural science major and 2 up to 4 clusters for social science major.

Keyword – Data Mining, Agglomerative Hierarchical Clustering Algorithm, Score of National Examination