

ABSTRAK

Salah satu indikator keberhasilan suatu sekolah adalah dengan mengukur nilai daya serap ujian akhir nasional dari sekolah tersebut karena daya serap adalah tolok ukur untuk mengetahui pemahaman peserta didik terhadap mata pelajaran yang dipelajari dan diajarkan oleh guru dalam suatu proses kegiatan belajar mengajar. Dalam mengkaji nilai daya serap ujian akhir nasional ini penulis akan menggunakan sebuah proses yang disebut penambangan data supaya ditemukan informasi mengenai kelompok-kelompok sekolah menengah atas di provinsi Daerah Istimewa Yogyakarta (DIY) berdasarkan nilai daya serap ujian nasional Bahasa Indonesia. Peneliti berfokus pada Bahasa Indonesia karena Bahasa Indonesia merupakan alat pemersatu bangsa, dan sarana komunikasi bagi Bangsa Indonesia yang memiliki banyak suku dan bahasa. Hasil dari penelitian ini dapat digunakan untuk membantu Departemen Pendidikan Provinsi dan juga penyelenggara pendidikan untuk melihat dan mengevaluasi sekolah-sekolah yang termasuk dalam kelompok nilai daya serap tinggi, sedang, dan rendah.

Teknik penambangan data yang digunakan dalam tugas akhir ini adalah *clustering* dengan algoritma *Agglomerative Hierarchical Clustering*. Algoritma ini akan mengelompokkan sekolah menengah atas di DIY berdasarkan nilai daya serap ujian nasional Bahasa Indonesia menjadi beberapa kelompok secara hirarki sehingga diperoleh informasi mengenai kelompok-kelompok sekolah di DIY berdasarkan nilai daya serap ujian nasional Bahasa Indonesia. Teknik untuk menghitung kedekatan dua *cluster* yang digunakan dalam penelitian ini adalah : *single linkage*, *complete linkage*, dan *average linkage*. Proses penambangan data yang dilakukan adalah proses KDD (*Knowledge Discovery in Database*) yaitu pembersihan data, integrasi data, pemilihan data, transformasi data, penambangan data, evaluasi pola, dan presentasi pengetahuan.

Penambangan data dan evaluasi dalam penelitian ini menggunakan 2 jenis data yaitu data daya serap semua butir kompetensi, dan data rata-rata seluruh butir kompetensi. Berdasarkan evaluasi menggunakan daya serap semua butir kompetensi dapat disimpulkan bahwa algoritma ini tidak tepat digunakan untuk analisis nilai daya serap ujian nasional Bahasa Indonesia. Menurut peraturan menteri pendidikan dan kebudayaan no.143 tahun 2014 tentang petunjuk teknis pelaksanaan jabatan fungsional pengawas sekolah dan angka kreditnya di mana satu pengawas sekolah mengawas minimal 7 sekolah, metode *single linkage*, *complete linkage*, dan *average linkage* tidak tepat untuk digunakan. Pengelompokan menggunakan metode *complete linkage* lebih baik dibandingkan menggunakan *single linkage* maupun *average linkage* meskipun masih ada cluster yang beranggotakan hanya 1 anggota. Berdasarkan evaluasi menggunakan rata-rata seluruh butir kompetensi dapat disimpulkan bahwa algoritma ini dapat mengelompokkan nilai rata-rata daya serap ke dalam 3 kelompok yaitu kelompok dengan nilai tinggi, sedang, dan rendah.

Kata Kunci : Penambangan Data, *Clustering*, *Agglomerative Hierarchical Clustering*, nilai daya serap

ABSTRACT

One of the indicators that indicates whether a school has been successful is by measuring the level of understanding (comprehension score) of the school national exam, because that value is a benchmark to determine the level of understanding of the students toward the subjects studied and taught by the teacher during the learning process. In assessing the level of understanding from the final national exam score, the author used data mining process in order to find information about clusters of high school in the Daerah Istimewa Yogyakarta (DIY) province based on the comprehension score of Indonesian national exam. The author sets the focus of this research on Indonesian language test because as a national language, Indonesian language is a means of unifying the nation, and means of communication for the nation of Indonesia which has many tribes and languages. The results of this research can be used to help provincial education departments and education providers to observe and evaluate the schools included in the cluster that has high score, medium score, and low score of understanding level.

Data mining that used in this research was clustering technique using Agglomerative Hierarchical clustering algorithm. This algorithm classified high school in the DIY province based on the comprehension score of the Indonesian national exam into several groups hierarchically in order to obtain information about the school groups in DIY based on the comprehension score of the Indonesian national exam. Techniques for calculating the proximity of the two clusters were used in this thesis are : single linkage, complete linkage and average linkage. The data mining process performed in this thesis is the KDD (Knowledge Discovery in Database) consists of data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation, and knowledge presentation.

Mining and evaluation was performed using two types of data, data of the comprehension score of all the competence items, and data of the average of all the competence items. Based on the evaluation using data of the comprehension score of all the competence items it can be concluded that this algorithm is not appropriate to be used for analysis of the comprehension score of Indonesian national exam. According to the regulation of the minister of education and culture rule no. 143 th 2014 about technical instruction of enforcement for functional position of school supervisor and credit numeral where one school supervisor at least be able to observed only 7 schools, unfortunately these method of single linkage, complete linkage and average linkage can not be use. Grouping using complete linkage method is better than using single linkage and average linkage although there still a cluster that consists of just one member. Based on the evaluation of the data using the average of all competence items it can be concluded that this algorithm is well suited to agglomerate the average value of comprehension score into 3 clusters with the following labels: high, medium and low score.

Keyword : Data Mining, Clustering, Agglomerative Hierarchical Clustering, comprehension score