

ABSTRAK

Ujian Nasional (UN) sebagai tolok ukur atau parameter akhir dari suatu proses pendidikan. Hasil Ujian Nasional digunakan sebagai dasar untuk pemetaan mutu program dan/atau satuan pendidikan di Indonesia. Setiap tahun UN diselenggarakan untuk mendapatkan sebuah informasi yang dapat bermanfaat untuk peningkatan mutu pendidikan. *Data mining* merupakan salah satu bidang ilmu yang dapat digunakan untuk mendapatkan informasi dari kumpulan data. Pada tugas akhir ini digunakan algoritma *K-means Clustering* yang akan menghasilkan sebuah sistem perangkat lunak yang dapat digunakan untuk mengelompokkan Sekolah Menengah Atas di DIY berdasarkan nilai Ujian Nasional. Sistem ini diuji dengan perbandingan pengujian manual dengan hasil sistem, pengujian *black box*, pengujian hasil pengelompokan menggunakan *Silhouette Coefficient*.

Dalam melakukan proses mengubah data mentah menjadi sebuah informasi yang bermanfaat, penulis menggunakan proses *Knowledge Discovery in Database* (KDD) yang terdiri dari pembersihan data, integrasi data, seleksi data, transformasi data, penambahan data, evaluasi data, dan presentasi pengetahuan. Pada tahap pembersihan data dilakukan secara manual, untuk itegrasi data dan transformasi data tidak dilakukan, sedangkan untuk seleksi data dan penambahan data penulis merancang perangkat lunak sebagai alat untuk melakukan tahap-tahap tersebut. Sedangkan untuk evaluasi pola dan presentasi pengetahuan, penulis melakukan evaluasi dari hasil penambahan data yang diperoleh dari hasil perangkat lunak dan menjelaskannya agar informasi tersebut dapat diterima oleh pihak-pihak yang membutuhkan. Perangkat lunak diujikan terhadap 8 *dataset* yang merupakan data Ujian Nasional SMA jurusan IPA dan IPS tahun 2014/2015.

Berdasarkan penelitian yang telah dilakukan, dapat diketahui bahwa algoritma *K-means* dapat digunakan untuk mengelompokkan data tersebut dengan nilai k yang diberikan. Pengujian hasil pengelompokan dengan menggunakan *Silhouette Coefficient*(*SC*) terhadap kedua set data yang digunakan yaitu IPA dan IPS tahun ajaran 2014/2015 dihasilkan nilai *SC* untuk data IPA sebesar 0.49 pada $k = 2$ yang dikategorikan sebagai *weak Classification* dan untuk data IPS sebesar 0.57 pada $k = 2$ dan 0.51 pada $k=3$ yang dikategorikan sebagai *good Classification*. Dengan demikian jumlah *cluster* yang disarankan untuk data nilai ujian jurusan IPA adalah 2 dan untuk IPS adalah 2 atau 3.

Kata Kunci : *Clustering, K-Means, data mining, nilai ujian nasional, Silhouette Coefficient*

ABSTRACT

As The National Exam is used as the final parameter of an educational process, the score resulted from National Exam is used as the basis to set the mapping of program quality and/or education unit in Indonesia. National Exam is conducted every year to get sufficient information in the purpose of upgrading the education quality. Data mining is one of disciplines aiming in gaining information among the data collection. In this undergraduate thesis, K-means Clustering algorithm is used to develop a software that can be used to classify Senior High School in Jogjakarta based on the the Final Score of National Exam. This system can be tested by compare the manual testing with result of system, black box testing, and result of clustering tested using Silhouette Coefficient.

In conducting the process of converting raw data into a useful information, the writer used the process of Knowledge Discovery in Database (KDD) consisting the data cleaning, data integration, data selection, data mining, data evaluation, and knowledge presentation. Data cleaning was done manually, data integration and data transformation were not conducted, while at the stage of data selection and data mining, the writer designed a software as the tool to succeed the whole stages. Meanwhile, for the pattern evaluation and knowledge presentation, the writer conducted an evaluation from the result of data mining that was obtained from the software result and the writer explained in order to have the information accepted by the people who required it. The software will be observed toward 8 datasets that were National Exam of Senior High School, both from Social and Exact Discipline –Year 2014/2015.

According to the research conducted, it is known that the K-means Algorithm can be used to classify the data by the given value of k . Result of clustering, tested using Silhouette Coefficient, toward both data set, that is Social and Exact Discipline year 2014/2015, resulting the value of SC for Exact Discipline 0.49 at $k = 2$, categorized as weak Classification and for Social Discipline 0.57 at $k = 2$ and 0.51 at $k=3$ which was categorized as good Classification. Therefore, the number of clusters recommended for the scoring of Exact Discipline is 2, and 2 or 3 for Social Discipline.

Keywords : Clustering, K-Means, data mining, National Exam Score, Silhouette Coefficient