

## ABSTRAK

Ujian Nasional merupakan salah satu upaya pemerintah dalam rangka memacu peningkatan mutu pendidikan. Berbagai upaya pasti dilakukan setiap sekolah agar para peserta didiknya mampu memberikan hasil yang memuaskan. Untuk mendapatkan hasil ujian nasional yang memuaskan juga diperlukan upaya para guru dalam peningkatan kualitas belajar siswa. Kemampuan setiap siswa dalam memahami materi dapat diukur salah satunya dengan daya serap setiap pelajaran. Daya serap diukur dengan cara menghitung setiap kompetensi dari soal pada suatu mata pelajaran. Pada tugas akhir ini menggunakan metode pengelompokan *K-means*, metode ini dapat digunakan untuk mengelompokkan Sekolah Menengah Atas di Provinsi Daerah Istimewa Yogyakarta berdasarkan data nilai Daya Serap ujian nasional mata pelajaran Bahasa Indonesia. Dalam tugas akhir ini penulis akan melakukan penelitian dengan mencari jumlah cluster terbaik dari 4 dataset dengan jumlah data yang berbeda-beda (34, 68, 102, dan 137). Terdapat banyak cara dalam menentukan hal tersebut, salah satunya dengan metode *Elbow*. Penentuan metode ini dilihat dari grafik SSE (*Sum Square Error*) dari beberapa jumlah *cluster*. Kemudian dalam melakukan proses mengubah data mentah menjadi suatu informasi yang bermanfaat, penulis menggunakan proses *Knowledge Discovery in Database* (KDD) yang terdiri dari pembersihan data, seleksi data, panambahan data, evaluasi yang terbentuk dan presentasi pengetahuan. Hasil dari penelitian ini akan dijadikan dasar untuk penentuan jumlah *cluster* terbaik dari data nilai Daya Serap ujian nasional mata pelajaran Bahasa Indonesia di Provinsi Daerah Istimewa Yogyakarta.

Kata Kunci : *K-means*, metode *Elbow*, nilai daya serap, SSE (*Sum Square Error*)

## ABSTRACT

National Examination is one of the government's efforts in improving the educational quality. Multiple efforts have certainly been pursued by each school in order that the students will be able to attain satisfying results. In order to attain the satisfying results in the National Examination, teacher should also improve the students' learning quality. The students' capacity in understanding the materials can be measured and one of the measures is the absorption capacity for each subject. The absorption capacity was measured by calculating the score of each competency in a subject. In relation to this final assignment, the researcher has implemented the K-Means clustering method; this method can be implemented in order to cluster the Senior High Schools in the Province of Yogyakarta Special Region based on the Absorption Capacity Score from the Bahasa Indonesia National Examination. In this final assignment, the researcher will conduct a study by looking for the best cluster of 4 dataset with different amount of data (34, 68, 102, and 137). There are many methods for clustering the data and one of the methods is the Elbow method. The selection of this method is seen from the SSE (Sum Square Error) method from several numbers of clusters. Then, in turning the raw data into beneficial information, the researcher has made use of Knowledge in Discovery Database (KDD) that consists of data cleaning, data selection, data mining, evaluation that has been formed, and knowledge presentation. The results of this study will be the basis for determining the best cluster from the data of Absorption Capacity Test Score in the Bahasa Indonesia National Examination conducted in the Province of Yogyakarta Special Region.

Keyword: K-means, Elbow method, absorption capacity score, SSE (Sum Square Error)