

INTISARI

PENGUNAAN METODE *K-MEANS* DENGAN *PROBABILITY SAMPLING* DAN PARAMETER DENSITAS UNTUK *CLUSTERING* JAWABAN URAIAN

Oleh

Nourma Jennie Aisyiyah

Evaluasi merupakan salah komponen penting yang digunakan untuk mengukur keberhasilan suatu proses pembelajaran. Salah satu cara yang dapat dilakukan dalam proses evaluasi adalah ujian tertulis menggunakan soal uraian. Soal uraian dianggap sebagai cara yang efektif karena melibatkan kemampuan siswa dalam pemecahan masalah dan penalaran. Akan tetapi, dengan banyaknya jawaban dari soal uraian, sulit bagi guru untuk mengetahui kemampuan dan variasi jawaban siswa secara keseluruhan. *Clustering* merupakan salah satu cara yang dapat digunakan untuk mengetahui variasi jawaban siswa secara keseluruhan. Dengan mengetahui variasi jawaban siswa, guru dapat mengetahui kemampuan siswa secara keseluruhan dan mengetahui kesalahan konsep yang mungkin terjadi sehingga dapat dilakukan evaluasi pada proses pembelajaran.

Pada penelitian ini digunakan algoritme *K-means* dengan metode *probability sampling* dan perhitungan parameter densitas pada proses inisialisasi pusat kluster. Dari hasil pengujian dengan menggunakan pengukuran nilai SSE, metode inisialisasi dengan parameter densitas menghasilkan kluster yang lebih baik dibandingkan dengan metode *probability sampling*. Namun dengan menggunakan pengukuran nilai *Silhouette*, metode *probability sampling* menghasilkan kluster yang lebih baik dibandingkan metode parameter densitas. Dengan karakteristik *dataset* yaitu adanya kemiripan antar jawaban, metode parameter densitas lebih tepat digunakan karena mengutamakan kedekatan data dengan data-data lain dalam membentuk kluster.

Kata Kunci: *Clustering* Teks, Algoritme *K-means*, *Clustering* Jawaban Uraian

ABSTRACT

APPLICATION OF PROBABILITY SAMPLING AND DENSITY PARAMETER IN K-MEANS FOR DESCRIPTION ANSWER CLUSTERING

by

Nourma Jennie Aisyiah

Evaluation is an important thing for education quality measurement. One of the evaluation methods is using descriptive or essay question. It is well known that there is greater for educational evaluation from essay question because, from this question, teachers could find out the ability of students to solve a problem and to give a reason of their answer. But it is hard for teachers to find out the insight from hundreds or thousands answers. One approach to addressing this is to automatically cluster the answers based on answer's similarity. From the clustering result, teachers could find out the insights of all answers. Also, teachers can find modes of understanding or misunderstanding concept so it can be used for teaching evaluation.

This research focuses on comparing probability sampling and density parameter for centroid initialization of K-means algorithm. Using SSE value measurement, clustering result of density parameter method is better than probability sampling method. But using Silhouette value measurement, probability sampling method is better than density parameter method. Based on answer similarity in a dataset, density parameter is an appropriate method because it considers the closeness of data to others when choosing initial centroid.

Keywords: Text clustering, K-means Algorithm, Essay Answer Clustering