

ABSTRACT

PERFORMANCE ANALYSIS OF LATENT SEMANTIC ANALYSIS ALGORITHM AND K-MEANS CLUSTERING ALGORITHM TO CATEGORIZE DATA RESULT OF IT AUDIT EVIDENCE

by

Mohamad Andika Bhaskara

In an organization, the evaluation process becomes an important component supporting the running of the company's operations. Evaluation for this information system can be done with IT Audit. Supervision in the information system of a company is useful for monitoring the use of information systems, maintaining corporate integrity, up to the security of corporate data. However, the documents produced by IT audit for evidence can cause diverse manipulation. Documents that are not arranged accordingly may interrupt the Auditor and can cause Auditor to work inefficiently. To be able to overcome these problems a method is needed to be able to group documents.

There are 2 data sets used in this research, which are 36 documents in the form of .docx and 106 documents in the form of .docx and .pdf. These documents are going through a process of common words remover, reducing inflected words and feature extraction of documents. This research focused on comparing the results of the K-Means Clustering Algorithm for data categorization method using several reduced matrix inputs from matrix decomposition generated by the Singular Value Decomposition.

From the evaluation results, shows that the results of the Sum of Squared Error and Silhouette are not much different in determining the best number of clusters. The best number of clusters for 36 documents is 4 and the best number of cluster for 106 documents is 3.

Keywords: Text Clustering, K-means Algorithm, Singular Value Decomposition.

INTISARI

ANALISIS KINERJA ALGORITMA *LATENT SEMANTIC ANALYSIS* DAN ALGORITMA *K-MEANS CLUSTERING* UNTUK MENGELOMPOKKAN DATA HASIL BUKTI AUDIT TI

by

Mohamad Andika Bhaskara

Dalam suatu organisasi, proses evaluasi menjadi komponen penting yang mendukung jalannya operasi perusahaan. Evaluasi untuk sistem informasi ini dapat dilakukan dengan Audit TI. Pengawasan dalam sistem informasi perusahaan berguna untuk memantau penggunaan sistem informasi, menjaga integritas perusahaan, hingga keamanan data perusahaan. Namun, bukti dokumen yang dihasilkan oleh audit TI dapat menyebabkan manipulasi yang beragam. Dokumen yang tidak diatur dengan semestinya dapat mengganggu Auditor dan dapat menyebabkan Auditor bekerja tidak efisien. Untuk dapat mengatasi masalah ini diperlukan suatu metode untuk dapat mengelompokkan dokumen.

Terdapat 2 kumpulan data yang digunakan dalam penelitian ini yaitu 36 dokumen dalam bentuk .docx dan 106 dokumen dalam bentuk .docx dan .pdf. Dokumen tersebut akan melalui proses penghilangan kata umum, pemotongan kata dan ekstraksi fitur dari dokumen tersebut. Penelitian ini difokuskan untuk membandingkan hasil Algoritma *K-Means Clustering* untuk metode pengelompokan data menggunakan beberapa input matriks tereduksi dari dekomposisi matriks yang dihasilkan oleh *Singular Value Decomposition*.

Dari hasil evaluasi, menunjukkan bahwa hasil *Sum of Squared Error* dan *Silhouette* tidak jauh berbeda dalam menentukan jumlah cluster terbaik. Jumlah kluster terbaik untuk 36 dokumen adalah 4 dan jumlah kluster terbaik untuk 106 dokumen adalah 3.

Keywords: Text Clustering, K-means Algorithm, Singular Value Decomposition