ABSTRAK

Seiring berjalannya waktu, jumlah data semakin bertambah banyak. Jumlah data yang terus bertambah dalam menyebabkan sulitnya data-data untuk dikelola, diatur, ataupun dicari. Dokumen teks adalah salah satu data yang jumlahnya bertambah dengan sangat pesat. Oleh karena itu, dibutuhkan suatu metode khusus untuk dapat mengelompokkan dokumen-dokumen teks, metode ini disebut *clustering*.

Teknik *clustering* yang paling umum dan banyak digunakan adalah *k-means clustering*. Selain *k-means clustering*, ada sebuah teknik *clustering* lain yaitu *spectral clustering*. Algoritma *spectral clustering* membuat graf terbobot, dan tidak berarah berdasarkan objek data, untuk menghasilkan hasil *clustering* yang optimal dengan memperhitungkan *eigenvalue* dan *eigenvector* yang terkait dengan graf. Penelitian ini akan mencoba menggunakan *spectral clustering* untuk *clustering* dokumen-dokumen teks dan membandingkan nilai *silhouette coefficient* dengan *k-means clustering*.

Tahapan proses dimulai dari pengolahan kata disebut *text mining*. Dalam *text mining* terdapat beberapa proses yaitu *stop-word removal*, *stemming*, pembobotan *tf-idf*, dan reduksi matriks menggunakan metode SVD. Tahap berikutnya dapat dihitung cosine similarity antara dokumen satu dengan yang lainnya, yang selanjutnya dapat dilakukan proses pengelompokan menggunakan *spectral clustering*. Setelah *cluster* terbentuk akan dievaluasi dengan menggunakan metode *Silhouette Coefficient*.

Dari penelitian yang dilakukan dari total data dokumen teks sebanyak 2225 file didapatkan *silhouette coefficient* optimal sebesar 0,201 untuk banyak *cluster k* 2 dan reduksi matriks *tf-idf* menjadi 1500 kolom.

Kata kunci : Dokumen teks, *spectral clustering*, *k-means clustering*

ABSTRACT

As time passes, the amount of the data keeps increasing. The amount of data that keeps increasing made it hard to organize, manage, and search. One of the data that increases significantly is the text document. Therefore, it needs a particular method to be able to group the text documents. This method named clustering.

The most common and widely used clustering technique is k-means clustering. Apart from k-means clustering, there is another clustering technique, namely spectral clustering. Spectral clustering algorithm made the undirected, weighted graph based on the object data to achieve the optimal clustering results by considering the eigenvalues and eigenvectors which are associated with the graph. The research uses spectral clustering to cluster document text and compares the silhouette coefficient with k-means clustering.

The process stage starts with the data processing called as text mining. In text mining, there are some processes. Text mining consist of stop-word removal, stemming, tf-idf weighting and SVD. The next stage calculates the cosine similarity between text documents, and next does the cluster using spectral clustering. After the clusters formed, it will be evaluated using Silhouette Coeficent method.

From the research that has been conducted, from a total 2225 text documents, the optimal silhouette coefficient is 0,201 for 2 cluster k and tf-idf matrix reductions to 1500 columns.

Keywords: Text documents, spectral clustering, k-means clustering