

ABSTRAK

Spectral clustering merupakan salah satu metode pengelompokan yang cukup sederhana diimplementasikan dan mempunyai performa yang menjanjikan dibanyak kasus pengelompokan berbasis grafik. *spectral clustering* bekerja dengan cara membuat grafik terbobot dan tidak berarah berdasarkan objek data, untuk menghasilkan hasil pengelompokan yang optimal dengan menghitung *eigenvalue* dan *eigenvector* yang terkait dengan grafik. Penelitian ini akan menganalisis unjuk kerja dari metode *spectral clustering* dengan menggunakan himpunan data *convex* dan *nonconvex*, serta membandingkan hasil unjuk kerja tersebut dengan metode *kmeans clustering*. Proses penelitian diawali dengan melakukan preprocessing data. Dalam *preprocessing* terdapat 2 proses yaitu standardisasi dan *cleaning*. Tahap berikutnya data akan dikelompokkan menggunakan *spectral clustering* dengan afinitas *nearest neighbor* dan *radial basis functions*. Cluster yang terbentuk akan divisualisasikan dan dievaluasi dengan menggunakan metode *silhouette coefficient*, *davies bouldin index*, *calinski harabasz index*, dan *confusion matrix*.

Hasil penelitian menunjukkan dari total 5 himpunan data yang terdiri dari 1 data *convex* dan 4 data *nonconvex*, *spectral clustering* memberikan performa yang baik pada data *convex* dan 3 variasi data *nonconvex*. Pada salah satu data *nonconvex* yang bersifat tandensi *spectral clustering* masih gagal mengelompokkan data tersebut. Peneliti juga mendapati metode *silhouette coefficient*, *calinski harabasz index* dan *davies-bouldin index* tidak cocok digunakan untuk mengevaluasi himpunan data *nonconvex* karena memberikan hasil yang kontradiksi dengan hasil visualisasi.

Kata kunci: *spectral clustering*, *convex*, *non-convex*.

ABSTRACT

Spectral clustering is one of the clustering methods that is quite simple to implement and has promising performance in many cases of graph-based grouping. Spectral clustering works by creating weighted and non-directional graphs based on data objects, to produce optimal grouping results by calculating the eigenvalue and eigenvector associated with the graph. The thesis will analyze the performance of spectral clustering methods using convex and nonconvex data sets and compare the results of these performances with kmeans clustering methods. The process begins by preprocessing data. In preprocessing there are 2 processes, namely standardization and cleaning. The next stage of data will be clustering by spectral clustering using nearest neighbor and radial basis functions affinity. The cluster formed will be visualized and evaluated using silhouette coefficient methods, davies bouldin index, calinski harabasz index, and confusion matrix.

The results showed that from a total of 5 data sets consisting of 1 convex data and 4 nonconvex data, spectral clustering provides good performance on convex data and 3 variations of nonconvex data. In one of the nonconvex data that is counter-spectral clustering still fails to group the data. Researchers also found that silhouette coefficient, calinski harabasz index and davies-bouldin index methods were not suitable for evaluating nonconvex data sets because they provided results that contradicted the results of visualization.

Keyword: *spectral clustering, convex, non-convex.*