

Intisari

Salah satu metode untuk memperoleh informasi dari sebuah data adalah dengan melakukan "*mining*" pada data tersebut. *Data mining* akan mengekstrak informasi baru dari sebuah data. Ada berbagai metode dalam *data mining*, satu diantaranya adalah *clustering*. *Clustering* mengelompokkan *instance* data ke dalam sejumlah kecil grup berdasarkan tingkat kemiripannya. Algoritme *Hierarchical Clustering* merupakan salah satu metode *clustering* yang sesuai untuk diterapkan pada dataset kecil. Algoritme *Agglomerative* adalah salah satu Algoritme *Hierarchical Clustering* yang mengeksekusi secara *bottom-up*.

Penelitian ini menerapkan Algoritme *Agglomerative* pada dataset *Breast Tissue*. Metode pendekatan yang digunakan dari Algoritme *Agglomerative* adalah *Single Linkage*, *Complete Linkage*, *Average Linkage* dan *Centroid Linkage*. Untuk menilai hasil *clustering* digunakan tiga metode yaitu *Silhouette Coefficient*, *Entropy* dan *Purity*. Penelitian dilakukan dengan menerapkan masing-masing metode berdasarkan *Euclidean Distance* dalam menentukan jarak antara *instance*. Kemudian dibandingkan dan diamati apakah terdapat *outlier* atau tidak. Ketika *outlier* terdeteksi, maka *outlier* tersebut dihilangkan kemudian proses diulangi dari awal.

Hasil dari penelitian ini menunjukkan bahwa metode pendekatan *Single Linkage*, *Average Linkage* dan *Centroid Linkage* menunjukkan hasil yang sama dalam pembagian anggota *cluster* maupun dalam nilai *Silhouette Coefficient*, *Entropy* dan *Purity* ketika *outlier* belum dihilangkan dengan nilai masing-masing secara berurutan 0,517, 1,644 dan 0,924.

Kata kunci : *Agglomerative Hierarchical Clustering, Silhouette Coefficient, Entropy, Purity*

Abstract

One method of obtaining information from a data is perform “mining” on the data. Data mining will extract new information from a data. There are various methods in data mining, one of them is clustering. Clustering classifying the instance data into a small number of groups based on the similarity. Algorithms Hierarchical Clustering is one method of clustering which appropriate for a small dataset. Agglomerative Algorithm is one of Hierarchical Clustering Algorithms that execute as bottom-up approach.

This research applied Agglomerative Algorithms in Breast Tissue dataset. The method used from Alogritme Agglomerative were Single Linkage, Complete Linkage, Average Linkage and Centroid Linkage. To measure the clustering results were used three methods Silhouette Coefficient, Entropy and Purity. The study was conducted by applying each method based on Euclidean Distance in determining the distance between instances. Then compared and observed whether there is an outlier or not. When outlier was detected, it was removed and then the process was repeated from the beginning.

The results of this study indicate that the approach Single Linkage, Average Linkage and Centroid Linkage showed the same results in the distribution of cluster members as well as the value Silhouette Coefficient, Entropy and Purity when the outlier is not removed with each value respectively 0.517, 1.644 and 0.924.

Keywords : *Agglomerative Hierachical Clustering, Silhouette Coefficient, Entropy, Purity*