

Abstract

During the development of technology, there's more information provided in the form of digital text documents. The document text has a lot of type of information, so to ease in retrieving information that match with the one we want, there's need for grouping of document. Clustering is a process for classifying data into a cluster, so the objects in a cluster has a very large similarity with other objects in the same clusters, but has very little similarity to the object on the other clusters. Clustering that performed on the documents referred as document clustering.

In this final task a clustering algorithm is implemented, that is Cure Algorithm. Cure algorithm is an algorithm that works by measuring the distance between documents with points representative list of the cluster that has been previously selected.

Testing is done by counting the number of cluster and calculate the value of cohesion and cluster separation that produce using Cure Algorithm. Based on the value of cohesion that is produced in this test in forming the right cluster with the category that is from dataset, it is obtained that the quality that have been produced is good enough around 0.0855. As based the value of separation that have been produced in this test is 0.927039 although forming clusters that do not fit with dataset category. But the cluster quality that have been produced is good enough, because the quality of clustering is good if the smaller cohesion value and the bigger separation value.

Keyword : cure algorithm, clustering, document, cohesion, separation.