Abstract

In this era, the number of digital text documents grows rapidly and in large number of variation. To facilitate the easiest way to retrieve information from those digital text documents, then classifying those documents are surely needed according to the topic. The classifying can be conducted by using data mining technique that called clustering. Clustering is an unsupervised learning technique which have meaning that the data is not being labeled, but then the clustering algorithm will classify documents based on their similarity. To make the document clustering, 'committee clustering' algorithm is used, which is working by building a cluster center with uniformly averaged value of feature vector of a subset of cluster members that called the committee. This committee will be responsible for determining whether a document is a part of a cluster or not. By selecting the committee members carefully, the feature of the center cluster will tend to lead to the target class [8].

In this final project, the number of cluster which formed by the committee clustering algorithm is adjusted with number of categories of data that used. That number of clusters produced the average value of silhouette coefficient 0.2296. Thus the quality of the resulting clusters are no structure.

Keywords: *algorithm, committee clustering, clustering, document, feature, silhouette coefficient.*