

Abstract

Sometimes, when given a document collection, will come the need to classify documents into clusters based-similarity level of the existing document content. For a small number of documents, grouping process manually is still possible to do. But, to classify documents with large amounts, manually clustering process will take a long time. Process automation into a single solution that can be used to reduce the time required to perform clustering. Therefore, we need a method of clustering which produces clusters of documents in accordance with existing topics.

In this final project implemented Non-negative Matrix Factorization (NMF) method to perform the document clustering process. This method is doing the factorization on term-document matrix that is the result of preprocessing the documents into two-dimensional matrix in which each row represents a term vector and each column represents the weight associated with the term on the line. Test on the process of clustering process is performed to determine the accuracy of the quality of clusters produced by Non-negative Matrix Factorization (NMF) method when compared with the initial cluster of datasets. It also tested the influenced of each input parameter on the results of cluster quality. The result of cluster quality as indicated by an Accuracy.

The result of clustering by Non-negative Matrix Factorization (NMF) method, indicates that the value of Accuracy cluster after the clustering process using Non-negative Matrix Factorization (NMF), influenced by the desired number of clusters and the number of documents used. Where the value of accuracy will not always rise in the number of different clusters with the same number of documents.

Keywords : Document Clustering, *Non-negative Matrix Factorization (NMF)*, *Accuracy*