Abstract

This Final Project describes implementation of Divisive Information Theoretic-Feature Clustering Algorithm on data training to be used when the classification of text documents. Results of clustering are the words that will be used as the initial description text document classification data training using Naïve Bayes method. Tests on the process of clustering on data training to determine the accuracy of this algorithmis give an initial description of the classification of text documents. In addition, testing the influence of each input parameter on the results of classification accuracy. The quality of the results seen from the results of classification accuracy is seen from the value of precision, recall, and F-Measure. The results of the classification of text documents using Naïve Bayes method where data training performed clustering process using Divisive Information-Theoretic Feature Clustering Algorithm is influenced by the number of clusters desired number of documents used. Where the accuraty of the classification result with clustering first on data training perform better compared without clustering.

Keywords : Classification of test documents, Clustering, data training, Naïve Bayes, Divisive Information-Theoretic Feature Clustering Algorithm, precision, recall, F-Measure