## Abstract

Today, amount of information, such as news articles, available on the web are growing fastly. Large number of information can causes the user get into trouble in finding important information. One of *Data Mining* task, *Text Categorization*, which is the task of assigning documents to pre-specified classes (categories) of documents can be used as the solution to organizing news documents.

One of *Text Categorization* method is *Centroid Based Classifier*. *Centroid Based* method represent documents as the vector. This method create the *centroid vector* for each set of documents belonging to the same class. That centroid vector will be used as model to classify documents using *cosinus similarity*.

In this final project, performance of *Centroid Based Classifier* method is compared with performance of k-NN and Naïve Bayes. Accuracy and f- measure are used to compare the performance of those methods. Beside of that, the final project also analyze outlier detection for increasing the accuracy of *Centroid Based*. The experiments show that *Centroid Based Classifier* give better performance than k-NN and Naïve Bayes, and outlier detection experiments show that outlier detection can improve the accuracy of *Centroid Based*.

**Keywords** : *text* categorization, centroid based classifier, centroid vector, akurasi, *f*-measure.