ABSTRACT

The need for text mining analysis is needed in handling the unstructured text. One of the important activities in text mining is the classification or categorization of texts. Text mining analysis is done in order to facilitate us in taking information or managing so much information from the world of the Internet or digital, one of them by classifying the data already available. Text categorization has various ways to approach, among others, probabilistic approach, support vector machine, artificial neural network, or decision tree classification. In statistical learning. Support Vector Machine is chosen because this method has advantages in the field of classification with the help of the kernel. In this final project support vector machine will group news based on topic into 3 part or class that is: government, economy and sport. The kernel of Support Vector Mechine will be combined with stopword, tokenisasi, tf-idf, chi-square is expected to make it easier to recognize the news rolled into the class of topics that should be.

With the kernel trick and the help of weighting method, Document Frequency, Chi square is expected to help classify the text with both non-linear and can improve accuracy, thus classification with support vector machine method can be the highest accuracy with the combination of stopword, tokenizing, term frequency & chi- square 47.43%

Keywords: text mining, support vector machine, tf-idf, chi square, stopword, tokenisasi.