Abstract

Nowadays news information is spread easily. Now people no longer need to buy print media to get information, because they have many kinds of news articles scattered on the internet. Various news portals can present various categories, instance criminal, political, business, sports and others. With the large number of articles, classification of categories is important because readers has a different favorite categories of articles. Sometimes each article that similiar to several different categories, this will be difficult in the selection of categories. With this system, it is expected to find the best category for an article. This research uses 1125 news article from kompas which are divided into five news categories. The classification that is done usually after preprocessing is the classification process. But before carrying out the classification process, preprocessing data can be sorted by a variety of feature selection methods. This methods that are expected to make improvements increase. In this Final Project will analyze several methods of feature selection. The methods used are the Frequency of Collection, Frequency of Documents, Frequency of Words, and Advantages of Information. By using the Naive Bayes classification method, the document frequency and word frequency produce the best performance, which is 88.6%. This feature increases by 8% compared to not using feature selection.

Keywords: feature selection, document classification, Indonesia news classification