

Abstract

Nowadays, the presence of Short Message Service (SMS) that indicate fraud acts is rising and very disturbing for SMS users. SMS is known as SMS Spam. Therefore, detection or filtering spam SMS automatically indispensable. This research will build a system that classifies Filtering SMS between SMS Spam with not Spam (Ham). The method used in Multinomial Naïve Bayes classification used in combination with the feature weighting Term Frequency - Inverse Document Frequency (TF-IDF). Before the classification of the data will be in the preprocessing in advance using techniques tokenisasi, hose handling, stopword, and stemming. The distribution of the data to obtain training data and data testing in the testing process is done by using cross validation. Tests were conducted comparing several test scenarios based on the use preprocessingnya. The best results were obtained after the testing process scored 94.44% accuracy with preprocessing election Slang Handling and Stemming which will be implemented on the mobile Android. Android mobile implementation rule is added to the form if the sender is not in the contact list, the user of incoming SMS will perform the classification process. The results obtained in Android mobile application can detect and classify SMS Spam with a value of 94.74% accuracy.

Keywords: *SMS Spam Filter, Classification, multinomial naïve bayes, term frequency - inverse document frequency*