

Abstract

The widespread use of SMS makes some people take advantage of this condition, by sending SMS to many people for his own benefit. This SMS is also called SMS Spam. For some people, it is very disturbing. Therefore, SMS Spam Filter will built for filter SMS. In this final project, conducted on SMS Spam Filter research, would classify SMS into Spam and Ham(Not Spam). SMS contents in the form of text and tend to be irregular, making the data to be used for classification must be done with preprocessing. Preprocessing will be used in this final project is Slang Handling, Stopword Elimination, Stemming, and Tokenization. After all the data is done with preprocessing, algorithm will classify its SMS is Multinomial Naïve Bayes Algorithm. This final project will use two sources of data, from The SMS Spam SMS Collection v.1 and British. Which will both be composed in accordance testing. Tests were conducted comparing several tset scenarios based on the used of preprocessing. And have obtained the best results with an accuracy reaching 98.15% with preprocessing selection Stopword Elimination with Stemming.

Keywords : Classification, SMS Spam Filter, Slang, Stopword Elimination, Stemming, Multinomial Naïve Bayes.