Abstract

Abusive language is an expression used by a person with insulting delivery of any person's aspect. In the modern era, the use of harsh words is often found on the internet, one of them is in the comment section of online news articles which contains harassment, insult, or a curse. An abusive language detection system is important to prevent the negative effect of such comments. Detecting abusive language in the online comment section is a challenge since abusive languages can be expressed in various words. Moreover, only a few studies have been conducted in Indonesian language. In this paper, we present an Indonesian abusive language detection system by tackling this problem as a classification task and solving it using the following classifiers: Naive Bayes, SVM, and KNN. We also performed feature selection procedure based on Mutual Information value between words. The experimental results show that SVM is the best classifier for detecting the abusive language in news comment with an accuracy score of 90,19% and the use of Mutual Information able to improve the classification accuracy by 1.63%. Mutual Information can increase the accuracy performance of the classifier.

Keywords: Abusive Language, Indonesian, Detection, SVM, Mutual Information