
Abstract

Stemming is an important part of a preprocessing process in sentiment analysis. In previous studies, the problem of stemming for overstemming and understemming cannot yet be solved using the Indonesian stemming algorithm. This research investigates and measures the accuracy of the influence of Indonesian stemming on twitter sentiment analysis regarding online transportation user satisfaction. This research compares the accuracy among three conditions: without using stemming in preprocessing, using Nazief Adriani Algorithm stemming, and using Sastrawi stemming, as well as knowing the handling of overstemming, understemming, and word sense disambiguation. This research uses the data from the crawling process from twitter of gojek customers. This research collects 2500 tweets from July until December with crawling Gojek keyword, then we process the tweets and give label manually before doing preprocessing process and compare among three conditions mentioned above, and after that the three conditions the three conditions will be processed by feature extraction using unigram, bigram and unigram bigram, then classified using the SVM algorithm. The result are tested using the confusion matrix of the data grouping. The results in this study showed that Sastrawi's has a higher accuracy as much as 0.604% compare to Nazi-ef Adriani, and remembering compared to not using stemming nor using Nazief Adriani stemming. And Sastrawi's stemming can handling of overstemming also understemming, but can not overcome word sense disambiguation problem.

Keywords: Sentiment Analysis, Stemming, Feature Extraction, SVM
