

Abstract

The many platforms that are equipped with review features make it easy for people to convey anything. Product reviews are judgments that are opinions from consumers about the products they have purchased. These reviews can provide benefits for both producers and consumers. Reviews from consumers can contain ratings that cover aspects of the product and reviews can run into hundreds or even thousands. The large number of reviews makes it difficult in the sentiment analysis process. Therefore we need a model that can analyze sentiment based on aspects of the product. Sentiment analysis was performed using the naive Bayes algorithm, feature extraction with TF-IDF, and feature selection with chi square. The application of stopwords removal or stemming processes in preprocessing and the use of n-grams in feature extraction can affect the resulting performance. In addition, the application of feature selection to the built model has an important role because it can improve classification performance. From the research results obtained the best accuracy of 80,18%, recall of 72,49%, precision of 77,25% and f1-score of 74,73%.

Keywords: sentiment analysis, TF-IDF, n-gram, chi square, naive bayes.