Abstract

An assessment of the content of a film review can be called sentiment analysis. Sentiment analysis on film reviews is divided into 2, namely in the form of positive reviews and negative reviews. One of the most frequently used data mining algorithms in research is Nave Bayes because it works quickly and efficiently as a text classification method but has a very sensitive lack of feature selection. In general, film review data contains very long content, so feature selection or feature trimming is needed to reduce dimensions during the classification process. In this study, using the Tf-Idf feature as a solution to simplify and speed up the search for appropriate information is to summarize the content. Tf-Idf (Term Frequency Inverse Document Frequency) is a weighting method in the form of integration between term frequency and inverse document frequency. The Tf-Idf method is used in this study to select features as a summary result, with its application to word weight feature selection. Before the classification process, preprocessing stages are carried out which include data cleaning and case folding, stop words removal, stemming, and tokenization. In this study, the accuracy value reached 86.48%. So that Naïve Bayes with the Tf-Idf feature on the problem of sentiment classification analysis for film reviews is proven to provide accurate accuracy.