Abstract

Feature extraction method commonly used for sentiment analysis is weighting the TF-IDF word. However, the TF-IDF method has a deficiency in determining weighting for sentiment analysis because it cannot give different weight to positive or negative words (terms). Delta TF-IDF is an improvised method of weighting the TF-IDF method which can give different weight to positive and negative words, because this method gives word weight based on the existence of words on the corpus that have been labeled positive and negative. In this study, a sentiment analysis system was made by weighting using Delta TF-IDF, the weighting method was compared to its performance with TF-IDF weighting, linear SVM was used as a classifier algorithm for sentiment analysis system in this study. Performance parameters analyzed in this study are Precision, Recall, and F1score. The system with the weight of Delta TF-IDF has an F1score value of 80% while the one using TF-IDF is 78%. It can be concluded that the weighting of Delta TF-IDF has better performance than the TF-IDF weighting for sentiment analysis systems. In this study also analyzed the correlation between the use of the number of datasets on system performance, correlation was calculated using Pearson's correlation, both systems have a dataset correlation value of F1Score above 0.80, which means that the correlation is positive and strong so it can be concluded that the more many datasets used will improve system performance. This study has also identified weaknesses that can occur in the Delta TF-IDF method.

Keywords: term weighting, delta TF-IDF,TF-IDF,SVM,sentiment analysis