Abstract

Online Social Networks have an essential role as a source of information, especially during an emergency. One of them is Twitter, a service that allows users to send and read messages but is limited in character. Thus, tweets that are written are very short and do not always use the correct grammar, and use many variations of words. Using word variations can increase the likelihood of vocabulary mismatches and make tweets difficult to understand. One solution to overcome this problem is to expand the features of the tweet. The feature expansion on Twitter is a semantic addition to the process of multiplying the original text to make it look like large text. In this study, word2vec will be used with the Gradient Boosted Decision Tree Method to classify it. The expected result of this research is to reduce words or sentences in the classification of Twitter topics which are evaluated using the accuracy value, F1-Measure. The highest accuracy value in the application of feature expansion using Word2Vec with the Gradient Boosted Decision Tree classification method is 85.44%.

Keywords: *classification, word2vec, tweet, twitter, expansion, Gradient Boosted Decision Tree*