Abstract

Twitter is one of the popular social network sites, that was first launched in 2006. This service allows users to spread real-time information. However, the information obtained is not always based on facts and sometimes deliberately used to spread rumors that cause fear to the public. So detection efforts are needed to overcome and prevent the spread of rumors on Twitter. Much research regarding the detection of rumors but is limited to English and Chinese. In this reserach, the authors built a system to detect Indonesian-language rumors based on the implementation of the SVM classification and feature selection using the TF-IDF weighting. Data collection was conducted in November 2019 to February 2020 using crawling methods by keywords and manual labeling process. Research data used topics around government and trending with 47,449 records and features combination based on users and tweets. Stages of research include the process of collecting data on the Twitter social networking site which is then carried out preprocessing consists of case-folding, URL removal, normalization, stopwords removal, and stemming. The next stage is feature selection, N-Gram modeling, classification, and evaluation using a confusion matrix. Based on the results of the study, the system gets good performance in the test scenario using 10% of testing data and unigram features with the highest accuracy value of 78.71%. As for features twitter that affected the detection of rumors covering the number of following, the number of like and mention.

Keywords : SVM, TF-IDF, Crawling¸ twitter, rumor