Abstract

Social media is a means to communicate and exchange information between people, and Twitter is one of them. But the information disseminated is not entirely true, but there is news that is not in accordance with the truth or often called hoaxes. There have been many cases of spreading hoaxes that cause concern and often harm a particular individual or group. So in this research, the authors build a system to identify hoax news on social media Twitter using the Decision Tree C4.5 classification method. What distinguishes this study from some previous studies is the existence of several test scenarios, classification only, classification using weighting features, and also classification using weighting and feature selection. The weighting method used is TF-IDF, while the feature selection uses Information Gain. The features used are also generated using n-grams consisting of unigram, bigram, and also trigrams. The final results show that the classification test that uses weighting and feature selection produces the best accuracy of 72.91% with a ratio of 90% training data and 10% test data (90:10) and the number of features used is 5000 in unigram features.

Keywords: twitter, hoax, Decision Tree C4.5, TF-IDF, Information Gain