

Abstract

Internet use in Indonesia continues to increase, because the more people who can access the internet, the more abuse in the internet itself, especially social media. It is necessary to be able to detect *hoaxes* on social media such as Twitter. Twitter as one of the social media has a significant role in the dissemination of information, everyone can respond to the information obtained. In the content of a tweet allows for vocabulary incompatibility. Therefore, in this study, the application of the Word2Vec Feature Expansion method to overcome vocabulary incompatibility. This research conducted the development and comparison of Twitter's hoax classification system using the Word2Vec Feature Expansion method with Logistic Regression, Support Vector Machine (SVM), Random Forest and system without the Word2Vec feature expansion method. As a result of this study, Word2Vec's Feature Expansion method on the RandomForest classification algorithm managed to increase system accuracy by 1.46% with an accuracy value of 89.53%.

Keywords: *Word2Vec, Random Forest, Feature Expansion, Twitter, Hoax Classification*