Abstract

Twitter is a popular social media platform that gives users the ability to send text messages with a maximum length of 280 characters which causes a lot of use of word variations that cause vocabulary writing errors and nowadays more and more tweets are spread and because of the very rapid spread it causes information overload. From the problems raised, it requires the ability to recognize words that have errors in writing and categorize tweets into certain categories. Therefore, this study aims to build a topic classification system on tweets that can study writing errors in a word and feature expansion using pretrained word vectors from FastText can be used to recognize writing errors in a word because the process of building word vectors from FastText can study the structure internal of a word that will be used in the *Support Vector Machine* classification model. The best results from this study get an accuracy of 76.88% with the application of feature expansion on top-1 but application of feature expansion using Indonesian pretrained word vectors on top-2 to top-10 reduces the performance of the *Support Vector Machine* classification.

Keywords: Topic Classification, Feature Expansion, FastText, *Support Vector Machine*