Abstract

Social media users especially on Twitter, can freely express opinions or other information in the form of tweets about anything, including responding to a public policy. In a written tweet, there is a limit of 280 characters per tweet and this allows for problems such as vocabulary mismatches. Therefore, in this study, the feature expansion Word2vec method was applied to overcome when the vocabulary mismatches occur. This study develops and compares the Twitter sentiment analysis system using the feature expansion Word2vec method with the Logistic Regression (LR) and Support Vector Machine (SVM) classification algorithms and the system without the feature expansion Word2Vec method. The results of this study, the feature expansion Word2Vec method on the SVM classification algorithm succeeded in increasing the system accuracy up to 0.99% with an accuracy value of 78.99% and succeeded in increasing the F1-Score up to 1.23% with a value of 0.7884.

Keywords: sentiment analysis, public policy, feature expansion, word2vec.