

Abstract

Twitter is a microblog-based social media site launched on July 13, 2006. In March 2020, 476.696 tweets about the government policy in COVID-19 spread on Twitter were captured by the Institute for Development of Economics and Finance (Indef). Government policy has a standard meaning, namely a decision systematically made by the government with specific goals and objectives relating to the public interest, whether carried out directly or indirectly. Sentiment analysis analyzes people's opinions, sentiments, evaluations, attitudes, and emotions from written language. In this decade, Sentiment Analysis is has become a trendy research area. The purpose of this paper is to focus how to implement word2vec using similarity word as a feature expansion for minimize the vocabulary mismatch in Twitter Sentiment Analysis using "word embeddings". This research contains 11.395 tweets for a dataset, where the dataset will be used in two classifications: Support Vector Machine Algorithm and Artificial Neural Network Algorithm. The output of Word2Vec will be used for feature expansion in this research, where the algorithm of expansion will check in each row in the corpus where has a similarity vector with that word and will replace the word with the similarity of this words if the value is 0. The dataset in Feature Expansion is using 142.545 articles from Indonesian media. The result of this research is ANN is better than SVM, where the ANN without feature expansion gets 68.89 % and using feature expansion gets 72.58 %. For SVM, the final accuracy without feature expansion is 63.95 %, and using feature expansion gets 68.56 %. This research proves that feature expansion can improve the final accuracy.

Keywords: Sentiment Analysis, SVM, ANN, Word2Vec, TF-IDF