Abstract

The existence of internet technology has played an unprecedented role in increasing the number of social media users. Users are also accustomed to express their feelings with others using a text-based platform. This phenomenon produces the production and creation of a wide variety of data, which can be analyzed to assess sentiment, thereby benefiting both individuals and organizations. Sentiment analysis is a method for extracting text data for obtaining information about the personal views of individual or group on existing issues. To improve the performance of deep learning on sentiment analysis, a good text representation method is needed to be used as an embedding layer. This study analyzes deep learning using the Long Short-Term Memory (LSTM) method, which is compared with LSTM and *word embedding* for word feature expansion in sentiment classification. The data used comes from *tweet* data according to the specified topic with several searched keywords. The results of this research are the highest F1-score performance using the Long Short-Term Memory classification with labels by the system and the SMOTE data imbalance technique with TF-IDF feature extraction, feature selection, FastText feature expansion in the Wikipedia and tweet corpus to get an increase of 17.76% and F1-score 91.87%.

Keywords: deep learning, fasttext, feature expansion, long short-term memory, sentiment analysis, word2vec