Abstract

Social media is a communication tool that supports users to interact socially using technology. One of the most popular social media platforms is Twitter. However, Twitter has been considered by the virtual police as one of the main sources of spreading hate speech on social media. In this final project research, the authors conducted a study on the detection of hate speech in tweets on Twitter Indonesia. The method used in this research is multi-label classification by applying the LSTM and BiLSTM methods. The dataset used was 13,169 tweet data and there are 12 hate speech class labels from the dataset. The results showed that the LSTM and BiLSTM methods had good performance in classifying text data with 10 trials with an an average accuracy value of whole class of 78.67% for LSTM and 80.25% for BiLSTM. Based on the accuracy obtained, BiLSTM has higher accuracy than LSTM so it can be concluded that BiLSTM is superior to LSTM in identifying hate speech on social media.

Keywords: twitter, hate speech, social media, LSTM, BiLSTM