Abstract

As the number of social media users rises, the probability of hate speech spread in social media also rises indirectly. Hate speech has become one of most common cases found on social media. The spread of hate speech can lead to a riot that might cause conflict, group extermination, and even human casualties. Some of the latest controversies in Indonesia related to hate speech was the hate speech uttered to the government that led to polemic and even demonstration in the country. Along with this, it is important to detect hate speech to avoid conflict to happen. As the spread of hate speech in social media increases, it requires significant human efforts and is costly to detect manually. Therefore, this experiment is built to detect hate speech detection in Indonesian twitter texts using several conventional machine learning and deep learning based, BiGRU, with various features. The machine learning approaches being used are SVM and RFDT, while deep learning based methods used are BiGRU and pre-trained IndoBERT with BiGRU. Several methods used are Word2vec and fastText. The experiment shows that BiGRU method with IndoBERT and no stop word removal achieves the best performance with 84.77% accuracy. BiGRU has advantages on storing important information from text, thus making a better result than conventional machine learning algorithm.

Keywords: hate speech, text classification, BiGRU, Word2vec, fastText