

ABSTRACT

Twitter is a social media service whose users can create, post, update, and read short messages called tweets. However, Twitter is often abused by its users by posting negative messages (tweets) that contain cyberbullying. Cyberbullying has a devastating impact on the mental health of victims, with some cases tragically leading to suicide due to the intense stress. Therefore, it is necessary to take preventive measures, including creating a cyberbullying detection system on Twitter. This research proposes a hybrid deep learning approach, feature expansion using Word2Vec, and feature extraction using TF-IDF in building a cyberbullying detection system on Twitter in the Indonesian language. A series of test scenarios were carried out on a system built using 29,085 Indonesian language tweet data. The final results of this study show that the best accuracy was obtained for the CNN-LSTM hybrid with an accuracy of 79.26% and for the LSTM-CNN hybrid with an accuracy of 79.48%. These results prove that combining hybrid models, Word2Vec as feature extension, and TF-IDF as feature extraction provides the best accuracy score compared to other deep learning models. Thus, this study has succeeded in detecting cyberbullying on Twitter so that it can be used to create a healthier social media environment for its users.

Keywords: cyberbullying detection, hybrid deep learning, feature expansion, word2vec