

ABSTRAK

Instagram merupakan salah satu media sosial populer. Seperti halnya media sosial lainnya, media sosial instagram merupakan media yang dapat digunakan untuk mencari informasi, berkomunikasi serta mengembangkan bisnis. Berkembangnya teknologi komunikasi serta informasi telah mengubah pola interaksi individu yang semula dilakukan secara konvensional menjadi modern. Kehadiran internet dan media sosial membuat masyarakat lebih mudah dalam berkomunikasi serta mendapatkan informasi di manapun dan kapanpun tanpa adanya batasan ruang, waktu dan jarak.

Terlepas sisi positif yang ditawarkan, kehadiran internet dan media sosial juga membawa dampak negatif, salah satunya berkembangnya fenomena *cyberbullying*. *Cyberbullying* dapat diartikan sebagai bentuk penggunaan teknologi untuk mengintimidasi, menjadikan korban, atau mengganggu individu atau sekelompok orang. Instagram merupakan salah satu media sosial dengan tingkat *cyberbullying* yang tinggi. Maka dari itu diperlukan sebuah sistem yang dapat mengidentifikasi komentar *cyberbullying* pada komentar instagram. *Sentimen analysis* dapat diterapkan untuk melakukan identifikasi *cyberbullying* pada komentar instagram. Algoritma yang digunakan untuk *sentimen analysis* adalah *Long Short Term Memory*. Algoritma ini akan melakukan proses analisis komentar dengan mengubah komentar menjadi data berbentuk *sequential*. Pada proses *embedding* memanfaatkan metode *word2vec* untuk menemukan makna semantik antar kata. Proses pada penelitian ini terdiri dari pengolahan data awal, *text preprocessing*, pembentukan model LSTM dan identifikasi. Identifikasi terdiri dari empat kelas yaitu *cyberbullying*, *irrelevant*, *neutral* dan *non cyberbullying*.

Pengujian dilakukan dengan skenario pengujian *k-fold cross validation* untuk menentukan tingkat akurasi, presisi, *recall* dan *F1-Score* dari model LSTM yang dibangun. Konfigurasi pelatihan meliputi 100 *epoch* dengan menambahkan *L1L2 regularizer* serta nilai *dropout* sebesar 0.3. Pelatihan dilakukan dengan membagi data latih sebesar 90% dan data uji sebesar 10% dengan membagi data menjadi 10 *subset fold*. Performa dari pengujian *10-fold* mendapatkan performa rata-rata hasil akurasi sebesar 68.1%, presisi 70.08%, *recall* 68.1% dan *F1-Score* 68.8%. Skenario lain membandingkan performa algoritma LSTM dengan *random forest classifier*. Hasil perbandingan menunjukkan keunggulan algoritma LSTM dibandingkan *random forest classifier* pada nilai akurasi 70% berbanding 62%, nilai presisi 69% berbanding 61%, nilai *recall* 70% berbanding 62% dan nilai *F1-Score* 69% berbanding 60%.

Kata kunci: *Cyberbullying, Sentimen analysis, Long Short Term Memory.*

ABSTRACT

Instagram is one of the popular social media. Like other social media, Instagram social media is a medium that can be used to find information, communicate, and develop business. The development of communication and information technology has changed the patterns of individual interaction from conventional to modern. The presence of the internet and social media makes it easier for people to communicate and get information anywhere and anytime without limitations on space, time, and distance.

Apart from the positive side offered, the presence of the internet and social media also has a negative impact, one of which is the development of the cyberbullying phenomenon. Cyberbullying can be defined as a form of using technology to intimidate, victimize, or disturb an individual or group of people. Instagram is one of the social media with a high level of cyberbullying. Therefore we need a system that can identify cyberbullying comments on Instagram comments. Sentiment analysis can be applied to identify cyberbullying on Instagram comments. The algorithm used for sentiment analysis is Long Short Term Memory. This algorithm will perform the comment analysis process by converting comments into sequential data. The embedding process uses the word2vec method to find semantic meanings between words. The process in this study consisted of initial data processing, text preprocessing, LSTM model formation, and identification. The identification consists of four classes, namely cyberbullying, irrelevant, neutral, and non cyberbullying.

The test was carried out with the k-fold cross-validation test scenario to determine the level of accuracy, precision, recall, and F1-Score of the LSTM model built. Training configuration includes 100 epochs by adding the activation of the L1L2 regularizer function and a dropout value of 0.3. The training is carried out by dividing the training data by 90% and the test data by 10% by dividing the data into 10 fold subsets. The performance of the 10-fold test gets an average performance of 68.1% accuracy, 70.08% precision, 68.1% recall, and 68.8% F1-Score. Another scenario compares the performance of the LSTM algorithm with a random forest classifier. The comparison results show the superiority of the LSTM algorithm over the random forest classifier at an accuracy value of 74% versus 63%, a precision value of 73% versus 61%, a recall value of 74% versus 63%, and an F1-score value of 73% versus 61%.

Keywords : *Cyberbullying, Sentimen analysis, Long Short Term Memory.*