
Abstrak

Cyberbullying adalah tindakan mengancam atau membahayakan orang lain dengan memposting teks atau gambar yang memermalukan atau melecehkan seseorang melalui internet atau perangkat komunikasi lainnya, seperti *cell phone* dan *computer*. Menurut survei yang telah dilakukan pada tahun 2019 oleh Polling Indonesia dan Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), dari 5.900 peserta survei, didapatkan 49% peserta menyatakan pernah di bully. Oleh karena itu, dilakukan penelitian mengenai cyberbullying detection, sebagai langkah awal mengurangi cyberbullying. Telah dikumpulkan data dari Twitter berdasarkan kata kunci Trending Twitter yang berkorelasi dengan peristiwa cyberbully. Kemudian data tersebut digabungkan dengan data dari penelitian sebelumnya. Sehingga diperoleh total 1425 tweet, terdiri dari 393 data berlabel cyberbully dan 1032 data berlabel non-cyberbully. Lalu digunakan model Doc2Vec untuk ekstraksi fitur, dan model klasifikasi menggunakan *baseline classification method* (SVM dan RF) dan CNN untuk mendeteksi teks cyberbully. Hasil penelitian menunjukkan bahwa model klasifikasi menggunakan CNN dan Doc2vec memiliki skor F1 tertinggi yaitu 65.08%.

Kata kunci : cyberbullying, doc2vec, text classification, convolutional neural network, twitter