

ABSTRAK

Media sosial memungkinkan pengguna untuk saling terhubung satu sama lain di belahan dunia manapun tak terbatas ruang dan waktu. Dengan media sosial, suatu informasi dapat menyebar dengan cepat dan mudah. Namun media sosial juga memberikan dampak negatif bila terjadi penyalahgunaan, misalnya *cyberbullying*. Berdasarkan data statistik yang diperoleh, 41-51 % remaja Indonesia yang berusia 13-15 tahun pernah menjadi korban *cyberbullying*. Padahal dampak yang ditimbulkan oleh tindakan *cyberbullying* ini lebih menyakitkan daripada *bullying* secara fisik. Instagram merupakan salah satu media sosial terpopuler untuk berbagi gambar, foto, dan video yang saat ini masih belum menyediakan fitur untuk mendeteksi *cyberbullying*. Penelitian ini bertujuan untuk membuat sistem klasifikasi komentar pengguna Instagram ke dalam kelas *cyberbullying* dan bukan *cyberbullying*. Algoritma yang digunakan untuk klasifikasi adalah SVM dan NBC dengan ekstraksi fitur TF dan TF-IDF. Dengan penggunaan *stemming* terbukti mampu meningkatkan akurasi namun berdampak pada penurunan performansi. Dari hasil pengujian dengan 500 *dataset* diketahui bahwa akurasi yang dihasilkan SVM lebih baik daripada NBC. Akurasi terbaik pada SVM diperoleh pada kombinasi SVM dan TF sebesar 85.2 %. Sedangkan akurasi terbaik pada NBC diperoleh sebesar 83.2% dengan kombinasi NBC dan TF-IDF. Pada penelitian selanjutnya dapat dilakukan dengan menambahkan jumlah *dataset* yang lebih banyak, mencoba ekstraksi fitur lainnya seperti *unigram* atau *convert negation*, dan menggunakan algoritma klasifikasi lainnya seperti K-NN, *decision tree*, atau *neural network* untuk mendapatkan hasil klasifikasi terbaik.

Kata kunci : *cyberbullying*, instagram, klasifikasi, nbc, svm.