

ABSTRAK

Pada era digital ini, laju perkembangan teknologi semakin pesat. Kemajuan ini membawa perubahan bagi kehidupan masyarakat. Hampir seluruh kegiatan sehari-hari sudah menggunakan teknologi terbaru saat ini. Teknologi yang sudah tidak asing lagi adalah internet. Penggunaan internet sudah meliputi hampir seluruh aspek kehidupan tidak terkecuali dunia hiburan. Salah satu situs di internet yang menawarkan hiburan bagi orang yang mengaksesnya adalah Youtube. Dibalik banyaknya pengguna aktif Youtube, terdapat beberapa kelemahan diantaranya adalah banyaknya komentar negatif baik yang berisikan spam, ujaran kebencian, hingga perundungan atau yang sering disebut dengan cyberbullying. Contoh video yang banyak terdapat komentar negatif adalah video – video yang terdapat pada kanal Lutfi Agizal dan rahmawati kekeyi putri cantikka. Beberapa video yang terdapat pada kanal tersebut sempat berada pada trending video di Youtube. Konten video yang rata – rata berisi musik yang dinilai plagiat membuat video tersebut banyak dihujat.

Dalam melakukan *text classification*, salah satu metode yang digunakan adalah metode *Support vector Machine* (SVM). Metode tersebut memiliki tingkat akurasi yang baik dalam melakukan analisis sentimen. Adapun dua tahapan yang dilakukan dalam melakukan klasifikasi *cyberbullying* ini yaitu klasifikasi komentar *cyberbullying* atau *non-cyberbullying* kemudian dilanjutkan dengan klasifikasi *cyberbullying*. Keduanya menggunakan metode SVM. Proses klasifikasi diawali dengan scraping data, kemudian dilakukan *text preprocessing* diantaranya *case folding*, *tokenization*, *stopword removal* atau *filtering*, dan *stemming*. Proses selanjutnya adalah pembobotan TF-IDF dan melakukan analisis menggunakan metode SVM. Pengujian menggunakan tabel *Confusion Matrix* dan *K-Fold Cross Validation* dengan iterasi sebanyak 5 kali.

Hasil pengujian menunjukkan bahwa nilai *accuracy* tertinggi yang dihasilkan oleh metode SVM dalam pencarian kelas *cyberbullying* pada komentar Youtube ada 66% dengan nilai *precision* sebesar 77% dan nilai *recall* sebesar 55% sedangkan nilai *accuracy* tertinggi untuk pencarian kategori *cyberbullying* adalah 59% dengan nilai *precision* sebesar 66% dan nilai *recall* sebesar 44%.