

ABSTRAK

Cyberbullying sama seperti perundungan tetapi dilakukan melalui teknologi digital sebagai medianya. Tindakan *cyberbullying* sudah marak terjadi seiring dengan semakin pesatnya perkembangan teknologi dan media sosial di masyarakat. Media dan pengguna perlu untuk menyaring komentar-komentar yang bersifat *bully* tersebut karena secara tidak langsung dapat memengaruhi mental psikologi yang membacanya terutama langsung ditujukan kepada orang tersebut. Tetapi, untuk melakukan *filtering* teks komentar secara manual, tidak mungkin dilakukan oleh manusia setiap saat. Diperlukan proses klasifikasi yang dilakukan oleh teknologi itu sendiri. Dengan memanfaatkan penggalian informasi, sistem diharapkan mampu mengklasifikasikan informasi yang beredar di masyarakat. Beberapa teknik klasifikasi yang dapat diterapkan pada klasifikasi berbasis teks yaitu Naïve Bayes, Support Vector Machine (SVM) dan K-Nearest Neighbor. Ketiga algoritma tersebut termasuk baik dalam melakukan proses klasifikasi. Dalam penelitian ini, telah dilakukan simulasi uji keakuratan ketiga algoritma tersebut pada 1000 dataset komentar. Data dikelompokkan dahulu secara manual ke label “bully” dan “tidak bully” kemudian data dibagi menjadi data latih dan data uji. Untuk menguji kemampuan sistem, data yang sudah diklasifikasikan oleh masing-masing algoritma kemudian dianalisis menggunakan *confusion matrix*. Hasil penelitian menunjukkan Algoritma Naïve Bayes mendapatkan tingkat akurasi paling tinggi yakni 87,775% kemudian diikuti Algoritma SVM sebesar 87,75% dan terakhir Algoritma KNN sebesar 80%. Dari segi kecepatan menyelesaikan sistem, Algoritma Naïve Bayes juga memiliki tingkat kecepatan yang paling baik diantara kedua algoritma lainnya dengan waktu penyelesaian selama 0,033 detik.

Kata Kunci: *Cyberbullying*, Algoritma Naïve Bayes, Algoritma K-Nearest Neighbor (KNN), Algoritma Support Vector Machine (SVM), Confusion Matrix, Google Colab, Python.