

## SARI

Cyberbullying menjadi sebuah masalah yang harus mendapat perhatian serius oleh semua pihak. Di samping tindakan ini merupakan kebiasaan yang buruk, *cyberbullying* juga memberikan dampak yang mengerikan, mulai dari gangguan psikis korban, hingga berujung pada kasus bunuh diri.

Tujuan dari penelitian ini adalah membuat sebuah aplikasi untuk mengidentifikasi konten yang mengandung makna perundungan secara daring (*Cyberbullying*) pada media sosial. Media sosial yang dipilih adalah *Twitter*. Setidaknya, ada 3900 baris data yang telah dikumpulkan. Data-data tersebut berisi dua jenis cuitan baik cuitan yang memiliki kecenderungan *Cyberbullying* dan yang tidak.

Untuk mencapai tujuan penelitian, peneliti menggunakan lima langkah penelitian, yaitu pengumpulan data, preprocessing, ekstraksi fitur, klasifikasi, dan evaluasi. Algoritma *Machine Learning* yang diimplementasikan dalam penelitian ini adalah *Multinomial Naïve Bayes*. Setelah model terbentuk, kemudian dibuat sebuah aplikasi berbasis *website* menggunakan *Django Framework*.

Klasifikasi menggunakan algoritma *Multinomial Naïve Bayes* terbukti memiliki performa yang baik dengan akurasi sebesar 81%. Model tersebut menjadi inti dari aplikasi Deteksi *Cyberbullying* yang diimplementasikan dalam sebuah aplikasi berbasis *website*.

**Kata kunci:** *Twitter*, *cyberbullying*, *Machine Learning*, klasifikasi