

1. Pendahuluan

1.1. Latar Belakang

Media sosial adalah sebuah sarana komunikasi berbasis daring yang dapat mempermudah penggunaannya dalam berinteraksi antar sesama pengguna tanpa adanya batasan wilayah dan waktu. Indonesia yang memiliki angka pengguna sosial media tertinggi di dunia. Kementerian Komunikasi dan Informatika (Kemenkominfo) menyatakan ada sebanyak 95% dari 63 juta pengguna internet adalah pengguna sosial[1]. Salah satu media sosial yang sering digunakan para pengguna yaitu Twitter, *Country Industry Head Twitter* Indonesia mengklaim bahwa negara Indonesia merupakan negara dengan pertumbuhan aktif pengguna harian sosial media Twitter yang paling tinggi[1].

Namun tidak semua pengguna dapat menggunakan media sosial Twitter dengan bijak. Tak sedikit pengguna menggunakan media sosial Twitter untuk melakukan Tindakan yang dapat merugikan pengguna lain seperti penyebaran berita hoaks, penipuan, hingga pelecehan atau perundungan secara daring (*cyberbullying*) masalah ini timbul seiring dengan perkembangan media sosial akhir-akhir ini. *Cyberbullying* dapat diartikan sebagai penggunaan media sosial oleh individu atau sekelompok pengguna untuk melecehkan pengguna lain yang dapat mengakibatkan pengguna lain merasakan efek negatif terhadap korban. Satu studi yang dilakukan oleh badan amal anti-intimidasi nasional menunjukkan bahwa dua dari tiga anak berusia 13-22 tahun yang disurvei telah menjadi korban perundungan secara daring (*cyberbullying*)[2].

Oleh karena itu, penelitian ini bertujuan untuk mendeteksi perundungan secara daing (*cyberbullying*) pada *platform* Twitter. Yang menjadi tolak ukur dalam penelitian ini berdasarkan timbal balik serta interaksi cuitan pengguna yang ada di Twitter. Sebelumnya telah dilakukan penelitian yang dilakukan oleh Yinhan Liu mengenai perluasan arsitektur BERT yang telah dikembangkan dan diberi nama RoBERTa, penelitian tersebut dilakukan sebagai pengembangan arsitektur BERT yang menunjukkan hasil secara signifikan kurang terlatih dibandingkan dengan model setelahnya. Maka dari itu metode yang akan digunakan pada penelitian ini adalah RoBERTa yang dimana RoBERTa merupakan pelatihan ulang BERT dengan metodologi pelatihan yang ditingkatkan, memungkinkan menerima data yang lebih banyak, dan daya komputasi. Selain itu, metode ini menghilangkan *Next Sentence Prediction* (NSP) pada BERT dan menggunakan *Dynamic Masking*. Dengan beberapa kelebihan yang dimiliki oleh RoBERTa, membuat metode ini dapat diandalkan dengan peningkatan performa dibandingkan dengan metode BERT[3].

1.2. Topik dan Batasan

Penelitian ini membahas terkait deteksi tweet masyarakat Indonesia yang mengandung *cyberbullying* dengan kata kunci “gendut” , “makan”, “pemerintah”, “kece bong”, “anj*ng”, “tol*I” dan beberapa kata yang mengandung kata kasar serta menggunakan kata umpatan hewan melalui media sosial twitter dengan metode klasifikasi BERT yang telah dikembangkan yaitu RoBERTa.

1.3. Tujuan

Tujuan dilakukannya penelitian ini adalah untuk mengetahui performansi metode klasifikasi RoBERTa dengan mencari nilai akurasi terbaik terhadap cuitan masyarakat Indonesia pada media sosial twitter yang mengandung *cyberbullying*.

1.4. Organisasi Tulisan

Pada laporan penelitian ini, akan dibahas studi literatur yang terkait dan berhubungan yang menjadi acuan penelitian pada Bab 2. Kemudian sistem yang dibangun pada penelitian ini dapat dilihat pada Bab 3. Hasil dari penelitian yang telah dilakukan dapat dilihat pada Bab 4. Serta kesimpulan dari penelitian ini tertera pada Bab 5.