1. Pendahuluan

Popularitas media sosial yang semakin meningkat tidak lepas dari fenomena *cyberbullying*. Facebook, twitter, ask.fm dan instagram terdaftar dalam 5 platform media sosial yang memiliki persentase tertinggi pada peristiwa *cyberbullying*, dengan Instagram menduduki peringkat pertama[1]. *Cyberbullying* adalah suatu tindakan penyalahgunaan teknologi melalui media sosial untuk mempermalukan atau mengancam orang lain[2].

Pengguna instagram dapat menulis komentar dalam kolom komentar pada unggahan foto atau video seseorang. Komentar yang ditulis berupa kalimat dijadikan sebagai input dengan output berupa identifikasi kalimat yang mengandung *cyberbullying* dan non-*cyberbullying*.

Berdasar beberapa penelitian yang menyelesaikan persoalan identifikasi *cyberbullying* dengan pendekatan pembelajaran mesin, metode yang mempunyai kinerja paling baik adalah SVM, dibanding metode Naïve Bayes dan KNN[3]. Pada penelitian [3] terdapat persoalan terkait akurasi pada klasifikasi karena jumlah data yang digunakan sedikit yaitu 700 komentar. Peningkatan akurasi dapat dilakukan antara lain dengan menambahkan informasi semantik seperti yang dilakukan pada penelitian klasifikasi pada topik ulasan produk dengan kenaikan akurasi dari 88.2% menjadi 89.9%[4].

Berdasar hasil penelitian [3][4] identifikasi *cyberbullying* pada Tugas Akhir ini menggunakan metode SVM sebagai metode klasifikasi pada data latih dengan menambahkan informasi semantik pada data uji. Informasi semantik yang akan digunakan berasal dari basis pengetahuan korpus, yaitu berupa nilai kemiripan antar kata, sesuai dengan yang digunakan pada penelitian[4]. Penambahan kata-kata yang mempunyai nilai kemiripan tinggi dengan kata-kata pada data pengujian dapat membantu meningkatkan akurasi klasifikasi, ketika terdapat kondisi kata pada data pengujian yang tidak tercakup dalam data pelatihan[4]. Penelitian [4] menggunakan basis pengetahuan wordnet dalam topik ulasan produk. Namun, penerapan *Word2vec* lebih dipilih dalam penelitian ini karena komentar instagram mengandung bahasa terkini yang tidak terdapat dalam WordNet. Selain itu, *Word2vec* bekerja mirip dengan pikiran manusia menggunakan asosiasi kata untuk membantu komputer mencari kemungkinan kombinasi kata dari basis pengetahuan korpus[5].

Topik dan Batasannya

Topik yang dibahas pada tugas akhir ini adalah identifikasi kalimat komentar pada kolom komentar instagram yang mengandung *cyberbullying* ataupun non-*cyberbullying*. Metode klasifikasi yang dipilih dalam penelitian ini adalah Support Vector Machine (SVM) dengan menambahkan informasi semantik saat pengujian menggunakan *word2vec*. Pada penelitian ini terdapat beberapa batasan yakni dataset yang dianalisis berupa kalimat komentar berbahasa inggris dengan jumlah data latih sebanyak 595 komentar, data uji 99 komentar, dan data untuk membangun kamus wor2vec 62088 komentar.

Tujuan

Tujuan penelitian ini adalah menerapkan penambahan kata pada data uji klasifikasi *cyberbullying* dengan SVM berdasar kemiripan semantik yang informasinya diambil dari basis pengetahuan *word2vec* dan menganalisis pengaruh penggunaan informasi semantik pada akurasi klasifikasi kalimat yang bersifat *cyberbullying*.

Organisasi Tulisan

Struktur penulisan tugas akhir tersusun atas pendahuluan yang menjelaskan permasalahan identifikasi *cyberbullying*, dilanjuttkan dengan bagian studi terkait yang memaparkan studi literatur terkait mencakup istilah umum *cyberbullying*, kategori *cyberbullying*, dan penelitian sebelumnya yang membahas identifikasi *cyberbullying*. Bagian selanjutnya merupakan sistem yang dibangun membahas tahapan yang dilakukan untuk menyelesaikan tugas akhir dengan penjelasan rinci dan singkat yang diawali dari pengumpulan data hingga kesimpulan. Hasil pengujian program dan analisis hasil pengujian terdapat pada bagian evaluasi berisi analisa sistem yang dibangun sesuai dengan tujuan penelitian. Bagian kesimpulan dan saran dipaparkan secara singkat hasil pengujian dan apa saja yang perlu ditambahkan untuk penelitian selanjutnya. Bagian terakhir adalah daftar pustaka yang merupakan daftar dari setiap referensi yang digunakan dalam penyelesaian tugas akhir.