

## 1. Pendahuluan

Penetrasi pengguna internet di Indonesia telah meningkat secara signifikan dari tahun ke tahun. Asosiasi Pengguna Jasa Internet Indonesia (APJII) melaporkan bahwa 210 juta orang menggunakan internet di Indonesia pada tahun 2022, atau 77,02 persen dari total populasi, dan 89,15 persen menggunakan internet untuk mengakses media sosial [1]. Twitter adalah salah satu platform media sosial yang paling menonjol di Indonesia [2]. Twitter adalah platform media sosial di mana pengguna berkomunikasi melalui tweet, yang merupakan pesan teks pendek. Pengguna Twitter dapat mengirim dan menerima tweet hingga 280 karakter dari pengguna lain [3]. Namun, tidak sedikit pengguna Twitter yang menyalahgunakannya untuk memposting hal-hal negatif, salah satunya adalah *Cyberbullying*.

*Cyberbullying* merupakan tindakan intimidasi dengan menggunakan teknologi digital, misalnya menakut-nakuti, membuat marah, atau mempermalukan seseorang [4]. Berbeda dengan Sentimen Analisis, Sentimen Analisis adalah teknik yang digunakan untuk mengidentifikasi, mengekstraksi, dan memahami sentimen atau emosi dari teks tertulis. Tujuan utamanya adalah untuk menentukan apakah pendapat dalam teks tersebut bersifat positif, negatif, atau netral [5]. dalam konteks media sosial, sementara *Cyberbullying* dapat dianggap sebagai suatu bentuk sentimen yang negatif, tidak semua sentimen negatif bisa dianggap sebagai *Cyberbullying*. Menurut data dari Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), 49 persen dari 5.900 pengguna internet pernah dirungung di media sosial [6] Oleh karena itu, perlu dikembangkan sebuah sistem pendeteksi *Cyberbullying* di Twitter.

Efektivitas model pembelajaran mesin tradisional dalam mendeteksi perundungan pada dunia maya yang terbatas karena sifat subjektif dari ekspresi yang melakukan perundungan [6]. Namun, penelitian terbaru menunjukkan bahwa pendekatan berbasis deep learning, seperti *Deep Neural Networks* seperti *Recurrent Neural Networks* (RNN), *Gated Recurrent Unit* (GRU) [8], *Long Short-term Memory* (LSTM) [9], *Bidirectional Long Short-term Memory* (Bi-LSTM) [10], *Convolutional Neural Networks* (CNN) [11], dan model serupa lainnya, mengungguli algoritma pembelajaran mesin konvensional dalam mengidentifikasi perundungan pada media sosial. Oleh karena itu, model-model pembelajaran mendalam ini dapat digunakan untuk mengatasi masalah ini.

Sebelumnya, banyak penelitian yang berfokus pada deteksi *cyberbullying* dalam Bahasa Indonesia [12]-[14]. Andriansyah et al [13] menggunakan *Support Vector Machine* (SVM) untuk mengkategorikan komentar pada akun Instagram selebgram. Nurrahmi et al [12] menggunakan delapan aturan umum berdasarkan Sarna [15] untuk mengekstraksi fitur, dan menggunakan *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (KNN) untuk mengklasifikasikan teks *cyberbullying* dari Twitter.

Pembatasan Twitter terhadap panjang tweet memaksa pengguna untuk menggunakan singkatan, frasa yang dipersingkat, dan bahasa gaul di internet. Penggunaan varian kata ini menyebabkan ketidaksesuaian kosa kata dan membuat tweet menjadi sulit dipahami. Oleh karena itu, untuk mengurangi kosakata yang tidak sesuai, fitur penyematan kata dapat digunakan [16]. GloVe, Word2Vec, dan FastText merupakan contoh dari fitur *word embedding* yang dapat digunakan sebagai fitur ekstraksi [16].

Kontribusi utama dalam penelitian ini sebagai berikut: i) Dalam penelitian ini diusulkan sebuah teknik untuk mendeteksi *cyberbullying* yang didasarkan pada *deep learning* dan C-RNN untuk *dataset* dalam bahasa Indonesia; ii) Pencarian pendekatan parameter terbaik untuk ekstraksi fitur untuk *dataset* dalam Bahasa Indonesia; iii) Menerapkan fitur penyematan kata menggunakan FastText yang dapat digunakan sebagai perluasan fitur. Sejauh pengetahuan kami, belum ada penelitian sebelumnya mengenai FastText untuk deteksi *cyberbullying* secara khusus dalam Bahasa Indonesia. Selain itu, penelitian tentang deteksi *cyberbullying* menggunakan FastText secara umum masih sangat sedikit jika dibandingkan dengan penggunaan metode hybrid deep learning secara luas.

Makalah ini disusun sebagai berikut: Pada Bagian 2, kami menguraikan komponen dan aspek utama dari metode yang diusulkan. Bagian 3 didedikasikan untuk menyajikan dan menganalisis hasil eksperimen yang diperoleh. Akhirnya, Bagian 4 menampilkan kesimpulan dari apa yang telah dilakukan sebelumnya pada bagian 3.

## 2. Sistem yang Dibangun

Dalam penelitian ini, teknik klasifikasi yang diterapkan pada tweet mencakup beberapa tahap yang berurutan. Tahap-tahapan ini meliputi pra-pemrosesan *dataset*, ekstraksi fitur, perluasan fitur, dan pada akhirnya, klasifikasi tweet. Representasi ilustratif dari proses-proses ini disajikan pada Gambar 1, untuk memberikan gambaran umum. Bagian selanjutnya akan membahas setiap langkah secara komprehensif, dengan memberikan pemeriksaan yang terperinci.