

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Teknologi yang semakin maju menyebabkan masyarakat semakin mudah dalam berinteraksi melalui media sosial(1). Saat ini, media sosial menjadi salah satu sumber informasi yang sering digunakan oleh masyarakat umum. Berdasarkan data *Statista* pada tahun 2025, terdapat 143 juta pengguna media sosial aktif di Indonesia dengan rata-rata penggunaan media sosial 3 jam 8 menit per hari(2). Namun, dengan kemudahan tersebut muncul dampak negatif dari penggunaan media sosial, yaitu *cyberbullying* atau perundungan di media sosial. *Cyberbullying* merupakan tindakan yang disengaja oleh individu atau kelompok terhadap pihak lain melalui media sosial(3). Tindakan tersebut dapat berupa penghinaan, sarkasme, pelecehan, hingga pengancaman yang dilakukan secara berulang-ulang. Beberapa faktor yang menjadi pemicu *cyberbullying* antara lain rasa iri terhadap korban, pelaku pernah menjadi korban *cyberbullying*, serta pengaruh lingkungan yang negatif(4). *Cyberbullying* memiliki dampak psikologis yang serius terhadap korban, seperti penurunan rasa percaya diri, kecemasan, hingga dampak fisik berupa penurunan kesehatan yang diakibatkan oleh gangguan tidur(5).

YouTube menjadi salah satu media sosial yang paling banyak digunakan di Indonesia dengan pengguna sebanyak 139 juta per Oktober 2023(6). Hal ini menjadikan YouTube sebagai *platform* yang banyak digunakan dalam berinteraksi di media sosial, karena menyediakan fitur komentar yang berfungsi sebagai sarana interaksi antara *creator* (pembuat konten) dengan pengguna. Kolom komentar YouTube sering dimanfaatkan untuk memberikan kritik terhadap konten, tetapi kerap kali terdapat beberapa komentar yang mengarah ke arah negatif atau bahkan dapat dikategorikan sebagai *cyberbullying*(7). Jumlah komentar yang banyak membuat pengecekan komentar secara manual menjadi tidak efektif dan efisien. Selain itu, saat ini belum banyak sistem otomatis yang dikembangkan secara khusus mengidentifikasi *cyberbullying* pada *platform* YouTube. Oleh karena itu, dibutuhkan sebuah sistem yang mampu membantu dalam mengidentifikasi komentar yang terindikasi sebagai *cyberbullying*.

Identifikasi *cyberbullying* pada media sosial telah banyak dilakukan oleh peneliti sebelumnya dengan menggunakan beberapa metode seperti *K-Means*, *Naïve Bayes*, dan *Support Vector Machine* (SVM)(8–11). Metode-metode di atas masih memiliki keterbatasan dalam memahami konteks kalimat secara menyeluruh, serta kurang mampu menangkap variasi

penggunaan kosakata. Untuk mengatasi keterbatasan tersebut, penelitian ini menggunakan model *IndoBert* sebagai metode dalam mengidentifikasi *cyberbullying*. *IndoBert* merupakan model *Machine Learning* yang telah dilatih menggunakan *dataset* berbahasa Indonesia, sehingga mampu memahami konteks kalimat dalam bahasa Indonesia dengan lebih baik(12–14).

1.2 Rumusan Masalah

Adapun rumusan masalah penelitian ini adalah sebagai berikut:

1. Bagaimana cara mengidentifikasi indikasi *cyberbullying* pada kolom komentar konten video di Indonesia pada *platform* YouTube?
2. Bagaimana merancang dan mengimplementasikan model *IndoBert* untuk mengidentifikasi *cyberbullying* berdasarkan data yang diperoleh melalui *API* YouTube?
3. Bagaimana performa model *IndoBert* dalam mengidentifikasi komentar yang mengandung indikasi *cyberbullying*?

1.3 Tujuan Penelitian

Adapun tujuan dalam penelitian ini adalah sebagai berikut:

1. Mengidentifikasi pola komentar yang mengandung indikasi *cyberbullying* pada kolom komentar konten video YouTube di Indonesia.
2. Menerapkan model *IndoBert* untuk mengidentifikasi *cyberbullying* pada kolom komentar konten YouTube.
3. Mengembangkan aplikasi berbasis *web* untuk mengidentifikasi *cyberbullying* pada kolom komentar YouTube.
4. Mengevaluasi performa model menggunakan metrik evaluasi seperti *accuracy*, *precision*, *recall*, *confusion matrix*, *F1-score*, dan *AUC score*.

1.4 Manfaat Penelitian

Adapun manfaat dalam penelitian ini adalah sebagai berikut:

1. Memberikan kontribusi dalam penerapan model *IndoBert* untuk mengidentifikasi *cyberbullying* pada teks berbahasa Indonesia.
2. Menghasilkan sistem yang membantu dalam mengidentifikasi *cyberbullying* pada kolom komentar YouTube.

1.5 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya berfokus mengidentifikasi komentar pada konten YouTube.
2. Model yang digunakan pada penelitian ini adalah *IndoBert* dalam mengidentifikasi komentar pada *platform* YouTube.
3. *Dataset* yang digunakan diambil dari tahun 2019 hingga 2024 dan untuk *labelling* menggunakan *Google Spreadsheet*.

1.6 Keterbaruan

Adapun penelitian sebelumnya yang menjadi acuan dalam penelitian ini adalah sebagai berikut:

1. Fajri Koto (2021). "*IndoBertweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization*". Penelitian ini menunjukkan hasil model *IndoBERTweet* memiliki performa yang lebih unggul disbanding model *IndoBERT* pada berbagai tugas *Natural Language Processing (NLP)*, penelitian ini juga menekankan pentingnya penyesuaian kosakata pada *dataset* yang digunakan(15).
2. Hayel Khafajeh (2024). "*Cyberbullying Detection in Social Networks Using Deep Learning*". Penelitian ini menunjukkan bahwa model *BERT* memiliki performa terbaik dengannn akurasi 87,3%, diikuti *CNN-LSTM* sebesar 86,5% dalam mendeteksi *cyberbullying* pada kolom komentar. Namun penelitian ini masih terbatas pada bahasa Inggris(16).
3. Andika et al. (2023). "*Deteksi Komentar Cyberbullying Pada YouTube Dengan Metode CNN-LSTM*". Penelitian ini memiliki hasil yang menunjukkan bahwa metode *CNN-LSTM* mampu mendeteksi *cyberbullying* pada komentar YouTube berbahasa Indonesia dengan nilai *score F1* sebesar 0,841 serta telah diimplementasikan dalam bentuk *API* sederhana. Namun, metode yang digunakan masih memiliki keterbatasan dalam memahami konteks kalimat(17).
4. Muftie et al. (2024). "*Perbandingan Performa Deteksi Cyberbullying dengan Transformer, Deep Learning, dan Machine Learning*". Penelitian ini menunjukkan bahwa model *Transformer* memberikan performa paling unggul dengannn akurasi sebesar 98,49% pada *dataset* YouTube, diikuti oleh metode *SVM* dengannn *TF-IDF* sebesar 97,82%. Penelitian ini juga menunjukkan bahwa performa model sangat dipengaruhi oleh karakteristik *dataset* yang digunakan. Namun, penelitian tersebut masih menggunakan model *Transformer* umum

dan belum memanfaatkan model *pre-trained* seperti *BERT* atau *IndoBERT* yang mampu memahami konteks bahasa secara lebih mendalam(18).

5. Primanda et al. (2024). “*Implementation of IndoBert for Sentiment Analysis of Indonesian Presidential Candidates*”. Penelitian ini menunjukkan bahwa model *IndoBERT* mampu melakukan klasifikasi sentimen pada data twitter berbahasa Indonesia dengan akurasi sebesar 80%. Penelitian ini membuktikan bahwa *IndoBert* efektif dalam memahami konteks bahasa Indonesia pada media sosial(19).
6. Heri Santoso dkk (2023). “*Deteksi Komentar Cyberbullying pada Media Sosial Instagram Menggunakan Algoritma Random Forest*”. Penelitian ini menunjukkan bahwa penggunaan algoritma *Random Forest* dengan *FastText embedding* dalam mendeteksi komentar *cyberbullying* pada Instagram dan memperoleh akurasi sebesar 84%. Namun, pendekatan tersebut masih memiliki keterbatasan dalam memahami makna kalimat karena menggunakan *embedding* statis(20).

Berdasarkan penelitian di atas yang telah dilakukan oleh beberapa peneliti, dapat disimpulkan bahwa terdapat beberapa sistem yang telah dikembangkan dalam analisis teks pada media sosial. Berbagai metode *Machine Learning* telah diterapkan dalam penelitian sebelumnya dengan hasil yang bervariasi. Namun, masih terdapat sejumlah keterbatasan, di antaranya kurangnya variasi kosakata, penggunaan data berbahasa Inggris, serta cakupan *platform* yang masih terbatas pada Twitter dan Facebook. Jumlah dan karakteristik komentar yang berbeda menjadikan *YouTube* sebagai sumber data yang potensial dan berfokus pada satu konten. Oleh karena itu, penelitian ini menggunakan model *IndoBert* untuk mengidentifikasi *cyberbullying* pada komentar YouTube berbahasa Indonesia.