

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan pesat media sosial dan platform komunikasi daring telah secara signifikan meningkatkan kemudahan interaksi lintas batas bahasa dan budaya secara global. Namun, keterhubungan yang semakin meluas ini juga memicu peningkatan konten berbahaya, khususnya ujaran kebencian, yang menimbulkan tantangan sosial dan etika yang cukup serius dalam ranah digital [1]. Deteksi dan moderasi ujaran kebencian secara efektif menjadi hal yang sangat penting untuk memastikan lingkungan daring yang aman dan inklusif [2]. Namun demikian, tugas ini menjadi semakin kompleks ketika berhadapan dengan teks campuran bahasa, seperti dalam kasus alih kode (code-switching), di mana pengguna secara leluasa berganti bahasa atau menggabungkan istilah dari berbagai bahasa dalam satu percakapan [3].

Teks-teks multibahasa semacam ini kerap ditemui di kalangan individu dwibahasa maupun dalam komunitas yang menggunakan lebih dari satu bahasa, sementara sistem deteksi ujaran kebencian yang ada umumnya masih menunjukkan keterbatasan karena bergantung pada dataset dan algoritma monolingual [4]. Pendekatan konvensional dalam mendeteksi ujaran kebencian, yang biasanya menggunakan model pembelajaran mesin terawasi atau sistem berbasis aturan, memerlukan sumber daya yang bersifat spesifik terhadap suatu bahasa, seperti dataset beranotasi dan leksikon, agar dapat berfungsi secara optimal [2], [5–8].

Sebagai respons terhadap tantangan tersebut, penelitian ini bertujuan untuk memanfaatkan kemampuan multibahasa dari model BLOOM (BigScience Large Open-science Open-access Multilingual Language Model) guna meningkatkan deteksi ujaran kebencian pada teks campuran bahasa Indonesia dan Inggris. Tujuan utama dari studi ini adalah melakukan penyetelan hiperparameter (hyperparameter tuning) terhadap model BLOOM agar kinerjanya lebih optimal dalam konteks campuran bahasa. Penelitian ini akan mengevaluasi efektivitas model dalam membedakan antara ujaran kebencian dan bukan ujaran kebencian dalam konteks kode campuran, serta membandingkannya dengan performa pada teks monolingual.

Evaluasi ini mencakup analisis mendalam terhadap kemampuan BLOOM dalam memahami ekspresi linguistik yang kompleks dalam bahasa Indonesia dan

Inggris. Kinerja model akan diukur secara menyeluruh menggunakan metrik seperti akurasi, presisi, recall, F1-score, serta nilai loss pada proses validasi dan pelatihan. Pengukuran ini akan memberikan gambaran menyeluruh terkait keunggulan dan keterbatasan model dalam mengenali dan menginterpretasikan ujaran kebencian pada input beragam bahasa.

Pendekatan ini tidak hanya bertujuan untuk menyempurnakan proses deteksi konten berbahaya dalam situasi dengan batas bahasa yang kabur, tetapi juga memberikan wawasan mengenai kemampuan adaptasi dan efisiensi BLOOM dalam penerapan dunia nyata.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana implementasi dan performa model BLOOM dalam mendeteksi ujaran kebencian pada teks campuran Bahasa Indonesia-Inggris?
2. Bagaimana Akurasi dan F1-Score model BLOOM dalam mengenali ekspresi linguistik yang bersifat informal dan kontekstual dalam teks campuran?

1.3 Batasan Permasalahan

Penelitian ini memiliki batasan sebagai berikut:

- Penelitian hanya difokuskan pada deteksi ujaran kebencian dalam teks campuran Bahasa Indonesia-Inggris.
- Model yang digunakan adalah salah satu varian BLOOM, yaitu BLOOM-560M (varian terkecil yang tersedia secara *open source*).
- Dataset dibatasi pada 30.000 sampel yang telah dikategorikan ke dalam ujaran kebencian dan non-kebencian.
- Analisis hanya mencakup kalimat atau teks yang bersifat formal atau semi-formal, dan tidak mencakup bahasa gaul, slang, maupun singkatan tidak baku.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

- Mengimplementasikan model BLOOM untuk mendeteksi ujaran kebencian dalam teks campuran Bahasa Indonesia-Inggris.
- Mengevaluasi akurasi dan F1-Score model BLOOM dalam mengenali ekspresi linguistik yang bersifat informal dan kontekstual dalam teks campuran Bahasa Indonesia-Inggris.

1.5 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini antara lain:

- Memberikan kontribusi terhadap pengembangan sistem deteksi ujaran kebencian yang mampu menangani teks multibahasa.
- Menyediakan pendekatan berbasis model terbuka (open-access) yang dapat digunakan oleh peneliti atau pengembang sistem moderasi konten.
- Menawarkan analisis performa model BLOOM sebagai referensi bagi penelitian dan implementasi selanjutnya.

1.6 Sistematika Penulisan

Berikut adalah sistematika penulisan laporan penelitian ini:

- **Bab 1 PENDAHULUAN:** Berisi latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat, dan sistematika penulisan.
- **Bab 2 LANDASAN TEORI:** Mengulas penelitian-penelitian terkait dan studi literatur yang mendasari penelitian ini.
- **Bab 3 METODOLOGI PENELITIAN:** Menjelaskan secara rinci mengenai model BLOOM dan arsitekturnya.
- **Bab 4 HASIL DAN DISKUSI:** Menyajikan metodologi penelitian, meliputi pengumpulan data, praproses, pelatihan model, dan tuning hiperparameter.
- **Bab 5 KESIMPULAN DAN SARAN:** Membahas hasil eksperimen, kontribusi penelitian, keterbatasan, serta saran untuk penelitian selanjutnya.