

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan media sosial di Indonesia menunjukkan tren yang signifikan. Menurut laporan Digital 2025: Indonesia dari DataReportal, pada Januari 2025 terdapat 143 juta pengguna aktif media sosial di Indonesia, yang setara dengan 50,2% dari total populasi negara tersebut. Angka ini mencerminkan peran penting media sosial sebagai platform utama bagi masyarakat Indonesia dalam berkomunikasi, berbagi informasi, dan mengekspresikan opini terhadap berbagai isu sosial dan politik. di mana platform X (sebelumnya Twitter) menjadi salah satu kanal utama penyebaran opini publik secara cepat dan luas [1].

Salah satu bentuk ekspresi digital tersebut adalah penggunaan tagar (#) sebagai bentuk kritik sosial. Tagar #KaburAjaDulu, misalnya, sempat viral di X sebagai reaksi terhadap situasi politik tertentu dan menyuarakan keresahan masyarakat terhadap pemerintah. Tagar ini mencerminkan bentuk protes digital yang direspons dengan berbagai macam sentimen, mulai dari kritik tajam hingga candaan sarkastik. Namun, persepsi publik terhadap fenomena ini masih belum banyak diteliti secara sistematis, padahal analisis sentimen dapat menjadi alat penting untuk memahami opini masyarakat secara objektif dan terukur.

Namun, menganalisis sentimen secara manual pada data media sosial memiliki tantangan besar karena jumlah data yang masif dan sifatnya yang dinamis. Pendekatan tradisional seperti survei atau analisis konten manual tidak lagi efisien untuk menangani kompleksitas bahasa informal, singkatan, slang, dan konteks kultural yang khas di platform seperti X[2]. Di sinilah teknologi pemrosesan bahasa alami (NLP) seperti model transformer berbasis Bahasa Indonesia, IndoBERT, dapat menjadi solusi. IndoBERT telah terbukti efektif dalam tugas klasifikasi teks berbahasa Indonesia karena kemampuannya menangkap konteks semantic dan linguistic yang kompleks[3].

Penelitian sebelumnya tentang analisis sentimen di media sosial banyak menggunakan lexicon-based approach atau model machine learning konvensional seperti SVM dan Naive Bayes. Namun, metode ini seringkali kurang akurat dalam menangkap nuansa bahasa, terutama untuk kasus-kasus ironi, sarkasme, atau konteks spesifik seperti isu politik. Keunggulan IndoBERT terletak pada arsitektur pretrained-nya yang mampu memahami hubungan kata dalam kalimat secara mendalam, sehingga cocok untuk diaplikasikan pada analisis sentimen tagar #KaburaJadDulu yang mungkin mengandung ekspresi linguistik yang beragam.

Selain itu, minimnya penelitian yang mengkaji sentimen publik terhadap tagar politik di X dengan pendekatan berbasis transformer menjadi celah akademik yang perlu diisi. Studi serupa lebih banyak fokus pada bahasa Inggris[4]. Padahal, karakteristik unik Bahasa Indonesia, termasuk campuran bahasa daerah dan slang internet, memerlukan model yang spesifik. IndoBERT, yang dilatih pada korpus Bahasa Indonesia, menawarkan solusi tepat untuk tantangan ini.

Dari segi sosial, analisis sentimen terhadap tagar #KaburaJadDulu dapat memberikan gambaran tentang tren opini publik terhadap isu tertentu, yang bermanfaat bagi berbagai pemangku kepentingan. Misalnya, pemerintah dapat mengevaluasi respons masyarakat terhadap kebijakan, sementara akademisi dapat memetakan pola komunikasi digital di Indonesia. Dengan demikian, penelitian ini tidak hanya relevan secara teknis tetapi juga memiliki nilai praktis.

Berdasarkan gap penelitian dan peluang tersebut, studi ini bertujuan mengimplementasikan IndoBERT untuk menganalisis sentimen dalam cuitan bermuatan tagar #KaburaJadDulu di X. Hasilnya diharapkan memberikan dua kontribusi utama: validasi kinerja IndoBERT dalam menangani kasus sentimen politik berbahasa Indonesia, dan pemetaan objektif atas distribusi sentimen publik (positif, negatif, netral) terkait isu yang diangkat.

Oleh karena itu, penelitian ini dirancang untuk menjawab pertanyaan: Bagaimana efektivitas IndoBERT dalam mengklasifikasikan sentimen pada tagar #KaburajadDulu di X, dan apa pola sentimen yang terbentuk? Dengan

menggabungkan pendekatan komputasi NLP dan analisis sosial, proposal ini menawarkan kerangka metodologis yang robust untuk mengeksplorasi opini digital di Indonesia.

1.2 Rumusan Masalah

Perumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana implementasi model IndoBERT dalam analisis sentimen terhadap tweet dengan tagar #KaburAjaDulu di media sosial X?
2. Bagaimana akurasi model IndoBERT dalam mengklasifikasikan sentimen (positif, negatif, netral) pada tweet bertagar #KaburAjaDulu?

1.3 Tujuan Penelitian

Adapun tujuan dari pembuatan sistem ini adalah sebagai berikut:

1. Menerapkan model IndoBERT untuk melakukan analisis dan klasifikasi sentimen terhadap tweet yang mengandung tagar #KaburAjaDulu di media sosial X.
2. Mengevaluasi tingkat akurasi model IndoBERT dalam mengklasifikasikan sentimen (positif, negatif, netral).

1.4 Manfaat Penelitian

Manfaat yang didapatkan dari pengguna adalah sebagai berikut:

1. Memberikan kontribusi terhadap pengembangan riset di bidang Natural Language Processing (NLP), khususnya dalam penerapan model transformer-based seperti IndoBERT untuk analisis sentimen bahasa Indonesia.
2. Menjadi acuan untuk penelitian sejenis yang berfokus pada analisis media sosial, terutama dalam memahami persepsi publik terhadap isu sosial yang sedang viral seperti tagar #KABURAJADULU.
3. Menyediakan hasil analisis sentimen yang dapat digunakan oleh akademisi, media, atau pembuat kebijakan untuk melihat pola sentimen masyarakat secara objektif dan terstruktur.

4. Menghadirkan pemahaman mendalam tentang bagaimana model berbasis BERT dapat menangani kompleksitas bahasa informal di media sosial, khususnya dalam konteks Indonesia.

1.5 Batasan Masalah

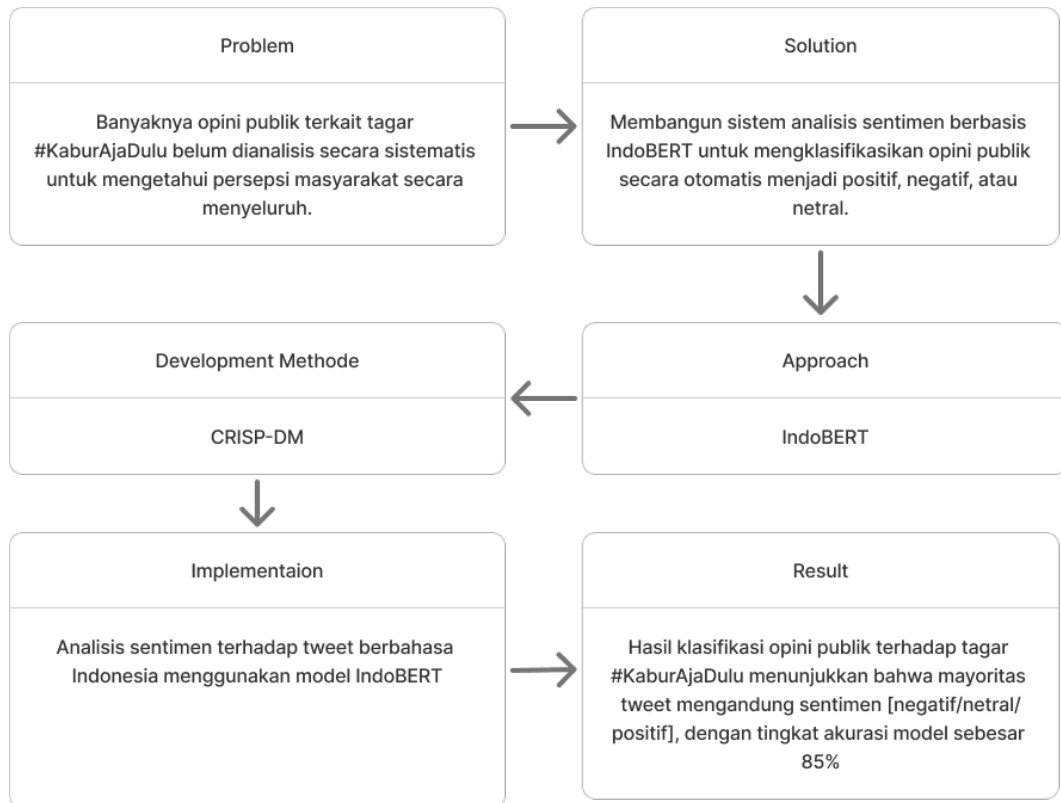
Agar penelitian ini lebih terarah dan sesuai dengan tujuan yang diinginkan maka berikut adalah batasan masalah yang ditetapkan:

1. Data yang dianalisis terbatas pada tweet berbahasa Indonesia yang mengandung tagar #KABURAJADULU, dikumpulkan dari media sosial X (Twitter).
2. Klasifikasi sentimen dibatasi pada tiga label: positif, negatif, dan netral.
3. Model yang digunakan dalam penelitian ini adalah IndoBERT.
4. Penelitian berfokus pada pengolahan data, fine-tuning, dan evaluasi performa model IndoBERT, tanpa membangun aplikasi antarmuka pengguna (UI) atau sistem produksi.
5. Evaluasi performa model dilakukan menggunakan metrik evaluasi umum dalam klasifikasi seperti accuracy, precision, recall, dan F1-score.
6. Bahasa pemrograman yang digunakan adalah Python.

1.6 Kerangka Pemikiran

Agar penelitian ini lebih terarah dan sesuai dengan tujuan yang diinginkan maka berikut adalah batasan masalah yang ditetapkan:

Pada Gambar 1.1 dibawah, akan menjelaskan kerangka pemikiran.



Gambar 1.1 Kerangka Pemikiran