

BAB I

PENDAHULUAN

1.1 Latar Belakang

Korupsi merupakan permasalahan global yang dialami oleh hampir setiap bangsa. Di Indonesia, meskipun memiliki kekayaan alam melimpah, pengelolaannya sering kali tidak transparan dan adil [1]. Salah satu kasus korupsi yang menyita perhatian publik secara masif adalah dugaan korupsi di lingkungan PT. Pertamina (Persero). Kasus ini mencakup dari dugaan "oplosan BBM" yang merugikan konsumen dan negara, hingga korupsi tata kelola minyak tanah dan produk kilang periode 2018-2023 dengan dugaan manipulasi proses pengadaan [1][2].

Skala kerugian negara dalam kasus ini sangat signifikan, dengan perkiraan mencapai Rp 193,7 triliun [2][3]. Besarnya nilai kerugian ini tidak hanya berdampak pada stabilitas ekonomi, tetapi juga secara fundamental merusak kepercayaan publik terhadap institusi negara dan perusahaan milik negara (BUMN) [1][3]. Dalam kasus Pertamina, ribuan komentar dan konten video memicu opini publik yang masif. Platform seperti X (sebelumnya Twitter) dan TikTok menjadi ekosistem data kaya, tempat publik menumpahkan emosi serta kritik [3].

Namun, data opini publik di media sosial ini memiliki karakteristik yang unik, yaitu tidak terstruktur, menggunakan bahasa informal (slang), penuh dengan singkatan, serta mengandung konteks ambigu dan nuansa emosional yang kental [3]. Untuk memahaminya, diperlukan pendekatan melalui Analisis Sentimen untuk mengolahnya. Proses ini bekerja dengan mengolah data opini mentah untuk menghasilkan polaritas yang terukur, yaitu mengklasifikasikan pandangan masyarakat ke dalam kategori sentimen positif, negatif, atau netral [2].

Meskipun banyak penelitian analisis sentimen, terdapat celah penelitian di mana belum banyak studi spesifik membedah sentimen kasus korupsi PT.

Pertamina menggunakan data gabungan X dan TikTok dengan algoritma terkini [4]. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen publik terhadap kasus korupsi PT. Pertamina. Penelitian ini diharapkan dapat memetakan persepsi masyarakat di media sosial, sekaligus menguji efektivitas model dalam menangani data komentar di media sosial yang kompleks.

1.2 Rumusan Masalah

Berdasar latar belakang di atas, rumusan masalah yang menjadi faktor dalam penelitian ini adalah :

1. Bagaimana pengaruh penggunaan bahasa tidak baku (*slang*) pada komentar media sosial terhadap performa model IndoBERT, dan seberapa efektif teknik normalisasi dalam mengatasi tantangan tersebut?
2. Bagaimana performa model IndoBERT yang telah dioptimalkan, diukur berdasarkan metrik akurasi, presisi, *recall*, dan *F1-score*, terutama dalam kemampuannya mengklasifikasikan sentimen kelas minoritas?
3. Berapa persentase distribusi sentimen (positif, negatif, dan netral) masyarakat terhadap berita Pertamina di platform TikTok dan X, berdasarkan hasil klasifikasi dari model yang telah divalidasi?

1.3 Batasan Masalah

Agar permasalahan yang dibahas tidak terlalu luas, maka perlu adanya batasan-batasan pada ruang lingkup penelitian yang akan dilakukan. Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah komentar berbahasa Indonesia yang berasal dari komentar pada akun resmi TikTok dan X (sebelumnya Twitter) yang berkaitan secara langsung dengan isu-isu Pertamina.
2. Data yang dianalisis adalah komentar yang dipublikasikan dalam rentang waktu 1 Februari hingga 1 April 2025, mencakup periode sejak munculnya isu korupsi hingga isu-isu relevan lainnya dalam kurun waktu tersebut.
3. Analisis sentimen dilakukan menggunakan algoritma IndoBERT varian base-pl untuk mengklasifikasikan komentar ke dalam tiga kategori sentimen, yaitu

positif, netral, dan negatif.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Menerapkan strategi penanganan ketidakseimbangan data dan optimisasi *hyperparameter* pada algoritma IndoBERT untuk membangun model klasifikasi sentimen yang andal.
2. Mengevaluasi performa model IndoBERT yang telah dioptimalkan berdasarkan metrik akurasi, presisi, *recall*, dan *F1-score* untuk memvalidasi efektivitasnya.
3. Menganalisis hasil klasifikasi untuk mengetahui distribusi persentase sentimen (positif, negatif, dan netral) masyarakat terhadap berita Pertamina, berdasarkan model yang telah divalidasi.

1.5 Manfaat Penelitian

Temuan dari penelitian ini dapat bermanfaat untuk:

1. Memperkuat validitas penggunaan algoritma IndoBERT dalam mengklasifikasikan opini publik di media sosial.
2. Menambah referensi ilmiah dalam studi komunikasi digital, opini publik, dan citra perusahaan di era media sosial.
3. Memberikan gambaran bagi Pertamina dalam memahami persepsi masyarakat secara kuantitatif untuk mendukung strategi komunikasi dan pengelolaan citra perusahaan.
4. Menjadi acuan bagi humas atau pengambil kebijakan dalam merespons isu-isu publik yang berkembang di media sosial secara lebih tepat.
5. Menjadi referensi awal bagi peneliti selanjutnya yang ingin mengembangkan studi terkait analisis sentimen, media sosial, dan NLP.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi digunakan agar penyusunan skripsi lebih terarah dan mudah dipahami. Sistematika penulisan pada skripsi ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini membahas latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini berisi tinjauan pustaka, dan dasar-dasar teori yang digunakan dalam penyusunan skripsi ini.

BAB III METODE PENELITIAN

Bab ini menjelaskan tentang objek penelitian, alur penelitian, alat, dan bahan yang dalam penelitian ini.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan mengenai hasil tentang setimen analisis yang ditemukan dalam penelitian ini.

BAB V PENUTUP

Bab ini berisi kesimpulan dari keseluruhan penelitian dan saran yang membangun untuk menambah kesempurnaan penelitian ini.