

BAB I

PENDAHULUAN

1.1 Latar Belakang

TikTok saat ini menjadi salah satu media sosial yang berkembang pesat dan digunakan oleh berbagai kalangan, mulai dari anak-anak hingga orang dewasa, sehingga interaksi pengguna di dalamnya semakin tinggi. Kondisi ini menjadikan TikTok tidak hanya dimanfaatkan sebagai media hiburan, tetapi juga sebagai ruang penyampaian respons dan persepsi pengguna terhadap berbagai konten yang beredar, terutama melalui kolom komentar yang dapat berisi beragam tanggapan terhadap isu tertentu[1]. Melalui fitur komentar, pengguna dapat berinteraksi secara langsung dengan menyampaikan pandangan, kritik, maupun dukungan terhadap isu yang disajikan. Kondisi ini menjadikan TikTok relevan sebagai sumber data opini publik. Hal ini diperkuat oleh data GoodStats Indonesia yang menunjukkan bahwa Indonesia termasuk dalam 10 negara dengan jumlah pengguna TikTok terbesar sebagai media informasi[2]. Selain itu, interaksi dan penyebaran konten di TikTok juga dipengaruhi oleh mekanisme algoritma platform, yang memungkinkan suatu konten menjangkau audiens dalam skala luas dengan relatif cepat dibandingkan media lainnya[3].

Namun, data komentar TikTok sebagai bentuk opini publik tidak mudah di analisis secara langsung, karena umumnya berupa teks bebas yang tidak terstruktur, di tulis secara informal, serta memuat variasi penulisan dan noise seperti simbol, emoji, tautan, pengulangan karakter, hingga kata tidak baku. Kondisi tersebut membuat kualitas data mentah masih kurang layak digunakan langsung untuk pemodelan. Selain itu, ketersediaan opini publik dalam jumlah besar membuat analisis manual menjadi tidak efisien dan berpotensi subjektif[4]. Sementara metode konvensional juga sulit menangani skala data yang besar tanpa dukungan pendekatan komputasional yang tepat[5].

Analisis sentimen digunakan untuk mengelompokkan opini publik berdasarkan kecenderungan sikap terhadap suatu isu yang diperoleh dari jejaring

sosial. Dengan analisis sentimen, kecenderungan opini publik terhadap suatu isu dapat dipetakan secara lebih terstruktur dan berbasis data[6]. Selain itu, pendekatan berbasis pembelajaran mesin juga membantu meningkatkan konsistensi dan objektivitas hasil analisis dibandingkan pengamatan secara manual[4].

Namun, penerapan analisis sentimen pada komentar TikTok menghadirkan tantangan lanjutan, terutama pada kemampuan model memahami konteks Bahasa Indonesia yang bervariasi dan sering tersirat. Keakuratan hasil analisis sangat dipengaruhi oleh kemampuan model menangkap makna secara kontekstual, sementara teks komentar di media sosial umumnya singkat dan beragam. Pendekatan berbasis transformer dinilai lebih efektif dalam menangani kompleksitas Bahasa Indonesia dibandingkan pendekatan secara konvensional [7]. Oleh karena itu, evaluasi kinerja model menggunakan metrik seperti akurasi, presisi, recall, dan F1-Score menjadi sangat penting untuk menilai keunggulan model klasifikasi[8].

Dalam penelitian ini, komentar TikTok pada isu yang diteliti terlebih dahulu diproses melalui tahap preprocessing agar teks lebih bersih dan konsisten sebelum masuk ke tahap pelabelan dan pemodelan. Pelabelan otomatis berbasis leksikon menghasilkan tiga kelas sentimen, yaitu negatif, netral, dan positif, dengan distribusi yang cenderung didominasi oleh kelas netral. Kondisi ini berpotensi memengaruhi hasil klasifikasi karena komentar netral sering bersifat samar, singkat, dan memiliki kemiripan makna dengan komentar negatif maupun positif, sehingga peluang tertukarnya prediksi antarkelas menjadi lebih besar.

Tahap pemodelan dilakukan menggunakan IndoBERT karena IndoBERT merupakan model bahasa prelatih yang dirilis khusus untuk Bahasa Indonesia dan dievaluasi pada benchmark IndoLEM, sehingga relevan untuk mengolah komentar berbahasa Indonesia yang beragam di media sosial[9]. Proses pelatihan menggunakan pembagian data latih, validasi, dan uji dengan stratified split agar proporsi kelas tetap terjaga, lalu dipantau menggunakan validasi serta mekanisme early stopping untuk membantu memilih model yang stabil. Evaluasi pada data uji dilakukan menggunakan metrik akurasi, macro F1, serta analisis per kelas dan confusion matrix. Hasil evaluasi menunjukkan performa model yang baik, dengan

kemampuan terbaik pada kelas negatif, disusul positif, dan tantangan terbesar tetap berada pada kelas netral karena karakteristiknya yang cenderung ambigu dan mudah bergeser ke kelas lain.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, permasalahan utama dalam penelitian ini tidak hanya terletak pada ketersediaan data opini publik, tetapi sejauh mana model klasifikasi mampu menginterpretasikan opini tersebut secara akurat. Oleh karena itu, rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana hasil pelabelan otomatis menggunakan leksikon InSet membentuk tiga kelas sentimen negatif, netral, dan positif, serta bagaimana karakteristik distribusi kelas yang dihasilkan memengaruhi kestapan dataset untuk proses klasifikasi?
2. Bagaimana kinerja pemodelan klasifikasi sentimen menggunakan IndoBERT pada komentar TikTok melalui tahapan preprocessing, pembagian data dengan stratified split, serta pemilihan konfigurasi pelatihan terbaik dari proses tuning untuk menghasilkan model dengan kemampuan generalisasi yang baik?
3. Bagaimana performa model IndoBERT dievaluasi menggunakan metrik akurasi, macro F1, F1-score per kelas, dan confusion matrix, serta bagaimana pola kesalahan klasifikasi menunjukkan bahwa kelas netral cenderung lebih sulit dibedakan dari kelas negatif dan positif?

1.3 Batasan Masalah

Agar penelitian ini berfokus dan sesuai dengan tujuan yang telah ditetapkan. Maka batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Data yang dianalisis berupa teks komentar pengguna media sosial Tiktok yang berkaitan dengan isu tertentu, sehingga hasil penelitian tidak mempresentasikan opini pengguna dari platform media sosial lainnya.

2. Penelitian ini dibatasi pada analisis sentimen berbasis teks, sehingga tidak mencakup analisis topik, emosi, maupun aspek non-teks lainnya.
3. Model yang digunakan dalam penelitian ini dibatasi pada model IndoBERT, tanpa melakukan perbandingan dengan model klasifikasi sentimen lain diluar lingkup penelitian.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Mendeskripsikan keluaran pelabelan otomatis berbasis leksikon InSet pada tiga kelas sentimen (negatif, netral, positif) dan menganalisis distribusi kelas yang terbentuk, khususnya dominasi kelas netral, sebagai dasar kesiapan dataset untuk klasifikasi
2. Menguji kinerja pemodelan sentimen menggunakan IndoBERT dengan stratified split dan konfigurasi tuning terbaik berdasarkan hasil pencarian hyperparameter untuk memperoleh model dengan generalisasi yang baik
3. Mengevaluasi performa model IndoBERT menggunakan akurasi, macro F1, F1-score per kelas, dan confusion matrix, serta mengidentifikasi sumber kesalahan utama pada kelas netral sebagai kelas yang paling sulit diklasifikasikan.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut :

1.5.1 Manfaat Teoretis

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan kajian analisis sentimen berbasis Natural Language Processing, khususnya dalam pemanfaatan model IndoBERT untuk klasifikasi sentimen opini publik pada data media sosial.

1.5.2 Manfaat praktis

Hasil penelitian ini dapat dimanfaatkan sebagai referensi dalam memahami kecenderungan opini publik berdasarkan data media sosial secara

objektid. Selain itu, hasil evaluasi kinerja model dapat menjadi acuan bagi pengembang sistem analisis sentimen berbasis teks di berbagai konteks isu publik.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun dalam lima bab dengan uraian sebagai berikut:

- Bab I :** Bab ini menguraikan latar belakang penelitian yang menjelaskan konteks dan urgensi dilakukannya analisis sentimen opini publik pada media sosial TikTok. Selain itu, bab ini memuat rumusan masalah yang menjadi fokus penelitian, batasan masalah untuk memperjelas ruang lingkup kajian, tujuan penelitian yang ingin dicapai, serta manfaat penelitian baik secara teoretis maupun praktis. Bab ini diakhiri dengan sistematika penulisan yang memberikan gambaran umum mengenai struktur skripsi secara keseluruhan.
- Bab II :** Bab ini menyajikan kajian literatur dan landasan teori yang relevan sebagai dasar konseptual penelitian. Pembahasan dalam bab ini meliputi konsep opini publik di media sosial, analisis sentimen, pemrosesan bahasa alami (Natural Language Processing), serta model bahasa berbasis transformer khususnya IndoBERT. Selain itu, bab ini juga meninjau penelitian-penelitian terdahulu yang berkaitan dengan analisis sentimen untuk menunjukkan posisi dan keaslian penelitian ini dibandingkan studi sebelumnya.
- Bab III :** Bab ini menjelaskan metode penelitian yang digunakan secara sistematis dan terstruktur. Pembahasan meliputi objek dan sumber data penelitian, teknik pengumpulan data komentar pengguna TikTok, tahapan pengolahan dan pembersihan data teks, proses pelabelan sentimen, serta perancangan dan pelatihan model klasifikasi sentimen menggunakan IndoBERT. Bab ini juga menguraikan skema pembagian data serta metrik evaluasi yang digunakan untuk menilai kinerja model.
- Bab IV :** Bab ini memaparkan hasil penelitian yang diperoleh dari proses

pelatihan dan pengujian model klasifikasi sentimen. Pembahasan mencakup karakteristik dataset setelah melalui tahap pengolahan data, hasil evaluasi kinerja model berdasarkan metrik yang digunakan, serta analisis confusion matrix untuk melihat pola kesalahan klasifikasi. Selain itu, bab ini juga membahas hasil klasifikasi sentimen secara keseluruhan untuk memberikan gambaran kecenderungan sentimen berdasarkan output model.

Bab V : Bab ini berisi kesimpulan yang dirumuskan berdasarkan hasil dan pembahasan penelitian. Kesimpulan disusun untuk menjawab rumusan masalah yang telah ditetapkan pada Bab I. Selain itu, bab ini juga memuat saran yang dapat digunakan sebagai bahan pertimbangan bagi penelitian selanjutnya, baik dalam pengembangan metode, penggunaan data, maupun pendekatan analisis sentimen.

