

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi digital di Indonesia telah membawa perubahan besar dalam berbagai sektor, termasuk olahraga. Klub sepak bola saat ini tidak hanya mengandalkan pertandingan di lapangan, tetapi juga memanfaatkan platform digital untuk berinteraksi dengan pendukungnya. PERSIB tidak hanya dikenal sebagai salah satu klub terbesar di Indonesia, tetapi juga memiliki ekosistem digital yang sangat kuat. Berdasarkan studi terdahulu, PERSIB berhasil menempati urutan ke-22 di dunia sebagai klub dengan kehadiran digital terbanyak [1]. Dominasi ini berlanjut hingga April 2025, menempatkan PERSIB di urutan ketiga klub terpopuler Asia setelah Al-Nassr dan Al-Hilal berdasarkan aktivitas media sosial [2]. Tingginya aktivitas digital dari basis pendukung yang besar ini mendorong manajemen untuk menghadirkan aplikasi resmi yang berfungsi sebagai sarana penyampaian informasi, penjualan tiket, dan komunikasi dengan suporter. Namun, aplikasi juga memunculkan tantangan berupa ulasan beragam dari pengguna yang mencerminkan pengalaman positif maupun negatif.

Ulasan yang dipublikasikan di *Google Play Store* ditulis secara bebas dengan gaya bahasa yang bervariasi. Informasi tersebut dapat menjadi sumber berharga dalam mengevaluasi kualitas aplikasi, tetapi jumlahnya yang terus bertambah membuat analisis manual menjadi tidak efisien [3]. Selain itu, ulasan bersifat tidak terstruktur sehingga sulit diinterpretasikan oleh pengelola aplikasi. Kondisi ini juga terlihat pada aplikasi PERSIB yang memiliki ribuan ulasan dari pengguna. Oleh karena itu, dibutuhkan metode otomatis untuk mengklasifikasikan opini dengan cepat dan akurat, salah satunya melalui pendekatan analisis sentimen berbasis kecerdasan buatan.

Salah satu algoritma untuk analisis sentimen adalah *Bidirectional Encoder Representations from Transformers* (BERT). Model ini memahami konteks kata dalam kalimat melalui mekanisme *transformer*. Untuk bahasa Indonesia, telah

tersedia model IndoBERT yang dilatih menggunakan korpus teks berbahasa Indonesia [4]. Meskipun IndoBERT memiliki kemampuan mumpuni dalam memahami konteks kebahasaan, tantangan utama dalam data ulasan aplikasi nyata adalah ketidakseimbangan distribusi kelas (*imbalanced data*), dimana jumlah ulasan positif sering kali mendominasi dibandingkan ulasan negatif atau sebaliknya. Jika kondisi ini diabaikan, model cenderung mengalami bias dengan memprediksi kelas mayoritas secara terus-menerus dan mengabaikan kelas minoritas, sehingga akurasi yang dihasilkan menjadi semu dan tidak representatif.

Meskipun penelitian terdahulu telah membuktikan efektivitas IndoBERT pada domain *e-commerce* dan opini publik [5][6], penerapan metode ini pada aplikasi olahraga spesifik masih terbatas, dan belum banyak yang menyoroti secara khusus dampak ketidakseimbangan data terhadap performa model di domain ini. Oleh karena itu, penelitian ini berfokus pada analisis sentimen ulasan aplikasi PERSIB menggunakan algoritma IndoBERT dengan penekanan khusus pada teknik penanganan *imbalanced data*. Langkah ini diambil untuk mengisi celah penelitian guna memastikan model tidak hanya akurat secara keseluruhan, tetapi juga mampu mendeteksi sentimen pada kelas minoritas dengan presisi yang lebih baik.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka dapat dirumuskan permasalahan penelitian sebagai berikut:

1. Bagaimana implementasi metode *fine-tuning* pada model IndoBERT untuk klasifikasi sentimen ulasan aplikasi PERSIB?
2. Bagaimana pengaruh penerapan teknik penanganan *imbalanced data* terhadap performa model IndoBERT dibandingkan dengan penggunaan data asli?
3. Bagaimana kinerja model IndoBERT hasil *fine-tuning* dalam mengklasifikasikan sentimen ulasan aplikasi PERSIB berdasarkan metrik *accuracy*, *precision*, *recall*, dan *f1-score*?

1.3 Batasan Masalah

Untuk menjaga agar penelitian tetap fokus dan mendalam sesuai dengan tujuan yang telah ditetapkan, maka penelitian ini memiliki batasan masalah sebagai berikut:

1. Data yang digunakan adalah ulasan pengguna pada aplikasi PERSIB yang bersumber dari *Google Play Store*.
2. Data ulasan yang diambil dibatasi pada periode 1 Januari 2023 hingga 8 Januari 2026.
3. Penelitian berfokus pada ulasan berbahasa Indonesia, termasuk ragam bahasa tidak baku (*slang*) dan campuran bahasa daerah yang akan dinormalisasi ke dalam bahasa Indonesia baku.
4. Algoritma yang digunakan adalah *Pre-trained Model* IndoBERT-base-pl dengan menerapkan teknik *fine-tuning*.
5. Skenario pengujian dibatasi pada perbandingan kinerja model antara penggunaan data asli (*imbalanced*) dan data yang telah melalui proses penanganan ketidakseimbangan data (*handling imbalanced data*).
6. Analisis sentimen diklasifikasikan ke dalam dua kelas, yaitu positif dan negatif.

1.4 Tujuan Penelitian

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

1. Mengimplementasikan teknik *fine-tuning* pada model *pre-trained* IndoBERT untuk membangun model klasifikasi sentimen yang mampu mengkategorikan ulasan pengguna aplikasi PERSIB.
2. Menganalisis pengaruh penerapan teknik penanganan *imbalanced data* terhadap peningkatan performa model dibandingkan dengan penggunaan data asli tanpa penanganan.
3. Mengevaluasi kinerja model IndoBERT hasil *fine-tuning* berdasarkan metrik *accuracy*, *precision*, *recall*, dan *f1-score* untuk mengetahui tingkat

efektivitas model dalam mengklasifikasikan sentimen.

1.5 Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat memberikan manfaat, baik secara praktis maupun teoritis.

1. Manfaat Teoritis

Secara teoritis, penelitian ini berkontribusi pada pengembangan literatur *Natural Language Processing* (NLP), khususnya dalam pemrosesan teks berbahasa Indonesia yang bersifat tidak baku (*colloquial*). Penelitian ini menyajikan bukti empiris mengenai efektivitas penerapan teknik *fine-tuning* pada model IndoBERT serta pengaruh strategi penanganan *imbalanced data* terhadap performa model. Hasil evaluasi dari skenario pengujian ini dapat menjadi referensi ilmiah atau tolak ukur bagi peneliti selanjutnya dalam menentukan metode terbaik untuk mengatasi ketidakseimbangan kelas pada domain analisis sentimen ulasan aplikasi di Indonesia.

2. Manfaat Praktis

Secara praktis, penelitian ini memberikan kontribusi signifikan bagi PT Persib Bandung Bermartabat selaku pengembang aplikasi. Wawasan yang diperoleh dari model klasifikasi yang akurat, terutama dalam mendeteksi keluhan pengguna secara presisi, memungkinkan manajemen untuk memprioritaskan perbaikan *bug* maupun pengembangan fitur yang paling mendesak. Dengan strategi peningkatan produk yang lebih tepat sasaran, kualitas aplikasi akan meningkat, yang pada akhirnya berdampak positif pada pengalaman pengguna (*user experience*) yang lebih stabil, responsif, dan memuaskan.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun untuk memberikan gambaran yang jelas dan terstruktur mengenai tahapan penelitian. Uraian secara garis besar untuk setiap bab adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini memuat latar belakang, rumusan masalah, batasan, tujuan, dan manfaat penelitian. Bagian ini menjelaskan urgensi penelitian serta ruang lingkup permasalahan analisis sentimen pada ulasan aplikasi PERSIB.

BAB II TINJAUAN PUSTAKA

Bab ini membahas landasan teori mengenai Analisis Sentimen, NLP, teknik penanganan *imbalanced data*, dan arsitektur model IndoBERT. Bab ini juga menyajikan penelitian terdahulu yang relevan serta kerangka pemikiran penelitian.

BAB III METODE PENELITIAN

Bab ini menjelaskan alur sistematis penelitian mulai dari pengumpulan data, pra-pemrosesan, hingga perancangan skenario eksperimen (data asli vs data diseimbangkan). Dibahas pula tahapan *fine-tuning* model IndoBERT serta metrik evaluasi yang digunakan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi dan analisis komparatif performa model dari skenario pengujian yang dilakukan. Fokus utamanya adalah mengevaluasi efektivitas teknik penanganan *imbalanced data* terhadap akurasi klasifikasi sentimen.

BAB V PENUTUP

Bab ini berisi kesimpulan yang menjawab rumusan masalah berdasarkan hasil eksperimen, serta saran untuk pengembangan penelitian selanjutnya.