

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Sepak bola merupakan olahraga paling populer di Indonesia dan telah menjadi bagian integral dari identitas nasional [1]. Tim Nasional Indonesia (Timnas) selalu menjadi sorotan utama publik, khususnya ketika berlaga di kompetisi internasional bergengsi. Antusiasme masyarakat mencapai puncaknya pada Kualifikasi Piala Dunia 2026 zona Asia, di mana Timnas Indonesia menunjukkan perkembangan yang sangat pesat di bawah arahan pelatih baru. Setiap hasil pertandingan, mulai dari kemenangan bersejarah hingga kekalahan yang menyakitkan, secara langsung memantik gelombang respons emosional yang masif dari jutaan pendukung di seluruh Indonesia.

Platform Twitter/X telah menjadi salah satu arena utama di mana opini, kekhawatiran, dan reaksi spontan masyarakat terhadap Timnas Indonesia terekam secara digital. [2] menegaskan bahwa Twitter/X merupakan sumber data yang berharga untuk memahami sentimen publik terhadap tim sepak bola nasional karena bersifat publik, real-time, dan kaya akan ekspresi emosi yang otentik. Temuan tersebut juga didukung oleh penelitian Marpaung dkk. [3] yang menganalisis sentimen masyarakat terhadap Timnas Indonesia pada Kualifikasi Piala Dunia 2026 menggunakan data Twitter/X, sehingga diperoleh gambaran mengenai dinamika opini publik pada domain sepak bola nasional.

Periode Kualifikasi Piala Dunia 2026 menjadi momentum penting karena setiap hasil pertandingan Timnas Indonesia memengaruhi persepsi dan respons masyarakat secara cepat. Sentimen yang muncul di media sosial tidak hanya mencerminkan tingkat dukungan maupun kritik terhadap performa tim, tetapi juga menggambarkan ekspektasi publik terhadap perkembangan Timnas Indonesia. Oleh karena itu, analisis sentimen diperlukan untuk mengidentifikasi kecenderungan opini masyarakat secara sistematis sehingga dapat memberikan informasi berbasis data mengenai respons publik terhadap performa Timnas Indonesia. Informasi tersebut dapat dimanfaatkan oleh berbagai pemangku kepentingan, seperti PSSI, pelatih, media, maupun sponsor, sebagai salah satu bahan pertimbangan dalam mengevaluasi komunikasi publik dan memahami persepsi masyarakat terhadap perkembangan tim.

Beberapa penelitian sebelumnya telah mengkaji analisis sentimen pada domain sepak bola dan Timnas Indonesia melalui berbagai pendekatan. [4] menggunakan FastText dan LSTM untuk menganalisis sentimen pembatalan tuan rumah Indonesia di Piala Dunia U-20 dan hanya memperoleh akurasi 74,92%. Pada domain yang lebih relevan, [2] menganalisis sentimen suporter di Piala Asia 2023 menggunakan Multinomial Naive Bayes dengan akurasi 81%. Sementara itu, [5] menganalisis sentimen komentar Instagram terhadap Timnas di Piala Asia 2023 menggunakan model Transformer berbahasa Indonesia, di mana IndoBERT memberikan hasil terbaik dengan akurasi 88,97% dan F1-score 88,59%. Adapun [3] membandingkan IndoBERT dan Indonesian RoBERTa Base pada tweet Timnas di Kualifikasi Piala Dunia 2026 dengan IndoBERT mencapai akurasi 79,7%.

Penelitian-penelitian di atas menunjukkan bahwa metode klasik seperti Naive Bayes memiliki keterbatasan dalam menangkap konteks, ironi, dan bahasa gaul khas tweet berbahasa Indonesia. Penelitian yang dilakukan oleh [6] menunjukkan bahwa model berbasis *deep learning* seperti BERT mampu memberikan performa yang lebih baik dibandingkan metode klasik seperti Random Forest pada tugas analisis sentimen teks berbahasa Indonesia. Hal ini diperkuat oleh [7] yang membuktikan bahwa IndoBERT secara signifikan mengungguli SVM dan Naive Bayes dengan F1-score 84% untuk analisis sentimen teks berbahasa Indonesia di Twitter, menjadikan pendekatan berbasis Transformer sebagai pilihan yang lebih tepat untuk menangani kompleksitas bahasa informal media sosial.

Selain keterbatasan pada arsitektur pemodelan, ditemukan pula keterbatasan pada tahap pelabelan data. Sebagian besar penelitian terdahulu yang menggunakan pendekatan pelabelan otomatis berbasis leksikon umum, seperti InSet (*Indonesia Sentiment Lexicon*), tidak dirancang untuk memahami konteks domain olahraga. Leksikon umum semacam ini cenderung salah menilai polaritas kosakata khas sepak bola, sebagai contoh kata “*menang*” atau “*keren*” yang seharusnya berkonotasi sangat positif dalam konteks performa tim, namun tidak dapat dinilai secara tepat oleh leksikon umum yang tidak disusun untuk domain spesifik ini. Hal ini menunjukkan adanya kebutuhan akan pendekatan pelabelan otomatis yang disesuaikan secara khusus dengan kosakata dan ekspresi yang umum digunakan suporter sepak bola Indonesia di media sosial.

Dalam konteks pengembangan model, [8] telah membuktikan bahwa kombinasi hybrid IndoBERT-LSTM mampu menghasilkan F1-score sebesar 98,90% pada klasifikasi multiclass tweet berbahasa Indonesia, mengungguli penggunaan IndoBERT maupun Word2Vec-LSTM secara terpisah. Lebih lanjut,

[9] menunjukkan bahwa arsitektur hybrid yang menggabungkan IndoBERT dengan *Bidirectional* LSTM (BiLSTM) secara konsisten memberikan peningkatan kinerja pada tugas analisis sentimen dan klasifikasi emosi berbahasa Indonesia dibandingkan IndoBERT yang berdiri sendiri. Arsitektur hybrid ini memanfaatkan kemampuan IndoBERT dalam menghasilkan representasi kontekstual yang kaya dari setiap token teks, yang kemudian diproses lebih lanjut oleh lapisan LSTM untuk menangkap pola dependensi sekuensial jangka panjang antar kata dalam kalimat, dengan varian *bidirectional* yang memproses urutan token dari dua arah sekaligus untuk menangkap konteks yang lebih kaya.

Meskipun penelitian-penelitian tersebut telah memberikan kontribusi yang signifikan, masih terdapat beberapa celah penelitian (*research gap*) yang perlu dikaji lebih lanjut. Berdasarkan hasil telaah literatur yang dilakukan pada penelitian ini, belum ditemukan penelitian yang secara spesifik menerapkan model *hybrid* IndoBERT-LSTM untuk analisis sentimen *tweet* berbahasa Indonesia terkait performa Timnas Indonesia pada Kualifikasi Piala Dunia 2026. Selain itu, sebagian besar penelitian sebelumnya masih menggunakan pendekatan pelabelan berbasis leksikon umum sehingga belum mampu menangkap karakteristik bahasa yang digunakan oleh penggemar sepak bola Indonesia, seperti istilah teknis pertandingan, singkatan, nama pemain, julukan tim, maupun ungkapan khas yang sering muncul pada media sosial.

Penelitian terdahulu juga umumnya hanya berfokus pada pengembangan model klasifikasi sentimen tanpa melakukan optimasi hiperparameter secara sistematis. Padahal, proses optimasi hiperparameter berperan penting dalam meningkatkan kemampuan generalisasi model dan memperoleh konfigurasi pelatihan yang optimal. Oleh karena itu, penelitian ini memanfaatkan algoritma Optuna untuk mengoptimalkan *learning rate* pada komponen IndoBERT dan LSTM sehingga diperoleh kombinasi hiperparameter terbaik berdasarkan nilai *F1-macro* pada data validasi.

Berdasarkan celah penelitian tersebut, penelitian ini mengusulkan model *hybrid* IndoBERT-LSTM untuk analisis sentimen *tweet* terhadap performa Timnas Indonesia pada Kualifikasi Piala Dunia 2026. Model ini memanfaatkan IndoBERT sebagai *encoder* yang dikembangkan khusus untuk Bahasa Indonesia dan BiLSTM untuk menangkap ketergantungan kontekstual antarkata. Selain itu, penelitian ini menerapkan pelabelan otomatis menggunakan kamus sentimen domain sepak bola yang dilengkapi dengan penanganan negasi serta pemberian bobot intensitas pada kata maupun frasa sentimen. Untuk meningkatkan

kinerja model, dilakukan optimasi hiperparameter menggunakan Optuna serta penanganan ketidakseimbangan kelas melalui *class weighting* pada fungsi *loss*. Pendekatan tersebut diharapkan mampu menghasilkan model klasifikasi sentimen yang lebih akurat sekaligus memberikan gambaran mengenai kecenderungan sentimen masyarakat terhadap performa Timnas Indonesia pada Kualifikasi Piala Dunia 2026.

1.2 Rumusan Masalah

Merujuk pada uraian latar belakang yang telah disampaikan sebelumnya, maka dapat dirumuskan permasalahan penelitian sebagai berikut.

1. Bagaimana karakteristik dan distribusi sentimen publik terhadap performa Timnas Indonesia pada Kualifikasi Piala Dunia 2026 berdasarkan *tweet* yang diperoleh dari platform X (Twitter)?
2. Bagaimana kinerja model *Hybrid* IndoBERT-LSTM dalam mengklasifikasikan sentimen positif, netral, dan negatif pada *tweet* berbahasa Indonesia yang membahas performa Timnas Indonesia pada Kualifikasi Piala Dunia 2026?

1.3 Batasan Permasalahan

Untuk memastikan ruang lingkup penelitian tetap terfokus dan tidak melebar, maka batasan permasalahan dalam penelitian ini ditetapkan sebagai berikut.

1. Data yang digunakan pada penelitian ini hanya berasal dari platform X (Twitter) dan berkaitan dengan performa Timnas Indonesia pada Kualifikasi Piala Dunia 2026.
2. Klasifikasi sentimen dibatasi menjadi tiga kelas, yaitu positif, netral, dan negatif.
3. Model yang digunakan adalah model *Hybrid* IndoBERT-BiLSTM untuk mengklasifikasikan *tweet* berbahasa Indonesia.
4. Pelabelan data dilakukan secara otomatis menggunakan kamus sentimen domain sepak bola yang disusun oleh peneliti khusus untuk konteks Timnas

Indonesia. Kamus tersebut dilengkapi dengan penanganan negasi dan pembobotan intensitas pada kata maupun frasa sentimen tanpa melibatkan anotasi manual oleh pakar ataupun penggunaan leksikon sentimen umum berbahasa Indonesia.

5. Penelitian ini tidak mencakup prediksi hasil pertandingan, analisis taktik permainan, maupun analisis sentimen yang berasal dari platform media sosial lain, seperti Instagram, Facebook, YouTube, dan platform sejenis.

1.4 Tujuan Penelitian

Selaras dengan permasalahan yang telah dirumuskan sebelumnya, penelitian ini bertujuan untuk:

1. Membangun dan mengevaluasi model *Hybrid* IndoBERT-LSTM dalam mengklasifikasikan sentimen positif, netral, dan negatif pada *tweet* berbahasa Indonesia yang membahas performa Timnas Indonesia pada Kualifikasi Piala Dunia 2026.
2. Mengetahui karakteristik dan distribusi sentimen publik terhadap performa Timnas Indonesia pada Kualifikasi Piala Dunia 2026 berdasarkan *tweet* yang diperoleh dari platform X (Twitter).

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut.

1. Memberikan kontribusi terhadap pengembangan metode analisis sentimen pada domain olahraga, khususnya melalui penerapan model *Hybrid* IndoBERT-LSTM yang dipadukan dengan pelabelan sentimen otomatis menggunakan kamus sentimen domain sepak bola, penanganan negasi, dan pembobotan intensitas sentimen pada *tweet* berbahasa Indonesia yang membahas performa Timnas Indonesia.
2. Menambah referensi akademik mengenai efektivitas model *Hybrid* berbasis Transformer yang dikombinasikan dengan LSTM serta didukung oleh optimasi hiperparameter menggunakan Optuna dalam tugas klasifikasi sentimen multikelas pada *tweet* berbahasa Indonesia.

3. Memberikan informasi berbasis data kepada pemangku kepentingan, seperti PSSI, pelatih, media, dan sponsor, mengenai dinamika sentimen publik terhadap performa Timnas Indonesia pada Kualifikasi Piala Dunia 2026 sehingga dapat menjadi salah satu bahan pertimbangan dalam memahami respons masyarakat terhadap performa tim.

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut.

- Bab 1 PENDAHULUAN

Pada bab ini diuraikan gambaran umum penelitian yang menjadi dasar dilakukannya penelitian ini. Bagian ini mencakup latar belakang masalah yang diperkuat dengan data dari berbagai penelitian terdahulu mengenai analisis sentimen sepak bola Indonesia dan perkembangan model *Hybrid IndoBERT-LSTM*, yang kemudian mengerucut pada perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, hingga sistematika penulisan laporan secara keseluruhan.

- Bab 2 LANDASAN TEORI

Bab ini membahas berbagai teori dan konsep relevan yang berfungsi sebagai fondasi teoretis untuk penelitian. Pembahasan mencakup tinjauan penelitian terdahulu, konsep analisis sentimen, Twitter/X sebagai sumber data, *Natural Language Processing* (NLP), *deep learning*, metodologi CRISP-DM, serta arsitektur model yang digunakan mencakup LSTM dan IndoBERT beserta arsitektur *hybrid*-nya. Bab ini juga menyajikan metode evaluasi yang digunakan untuk mengukur kinerja model.

- Bab 3 METODOLOGI PENELITIAN

Bab ini menjabarkan langkah-langkah penelitian secara teknis, mulai dari pengumpulan data, pelabelan otomatis berbasis kamus domain-spesifik, pra-proses teks, tokenisasi, pembagian dataset, perancangan arsitektur model *Hybrid IndoBERT-LSTM*, pencarian hiperparameter otomatis menggunakan Optuna, hingga konfigurasi pelatihan, evaluasi model, dan pengujian inferensi pada data baru.

- Bab 4 HASIL DAN DISKUSI

Bab ini memaparkan hasil dari setiap tahapan penelitian yang telah

dilaksanakan, mencakup hasil pengumpulan dan pelabelan data, hasil pencarian hiperparameter, hasil pelatihan model, hingga evaluasi akhir pada data uji dan pengujian inferensi pada data baru, disertai pembahasan dan perbandingan kinerja model dengan penelitian-penelitian terdahulu yang relevan.

- **Bab 5 SIMPULAN DAN SARAN**

Bab ini berisikan simpulan dari keseluruhan hasil penelitian yang telah dilaksanakan, beserta saran untuk pengembangan penelitian selanjutnya.

