BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan komunikasi telah mendorong peningkatan penggunaan media sosial sebagai sarana penyebaran informasi dan penyampaian opini masyarakat. Media sosial telah menjadi sumber informasi yang sangat kaya untuk menggambarkan opini masyarakat terhadap suatu isu. Berbagai penelitian memanfaatkan komentar maupun unggahan pengguna media sosial sebagai dataset analisis sentimen karena mampu merepresentasikan persepsi publik secara aktual dan dinamis(Sayarizki and Nurrahmi 2024) Salah satu platform yang banyak digunakan adalah Instagram, yang menyediakan berbagai fitur interaksi, seperti unggahan foto, video, serta kolom komentar yang memungkinkan pengguna memberikan tanggapan terhadap berbagai isu, termasuk dalam bidang olahraga. Melalui fitur komentar tersebut, pengguna dapat menyampaikan pendapat, kritik, maupun dukungan terhadap performa seorang atlet secara langsung sehingga menciptakan interaksi yang berlangsung secara cepat dan real-time.(Alafwan, Siallagan, and Putro 2023)

Dalam dunia sepak bola Indonesia, Hokky Caraka merupakan salah satu pemain muda yang mendapatkan perhatian publik. Performa yang ditampilkan dalam berbagai pertandingan seringkali memicu beragam respons dari masyarakat, khususnya di media sosial Instagram. Penelitian mengenai analisis sentimen pada media sosial menunjukkan bahwa data yang berasal dari platform digital dapat dimanfaatkan untuk memahami kecenderungan opini masyarakat terhadap suatu objek maupun peristiwa. Oleh karena itu, komentar pengguna Instagram dipandang sebagai sumber data yang relevan untuk menggambarkan persepsi publik terhadap performa seorang atlet(Geni, Yulianti, and Sensuse 2023) Komentar-komentar tersebut mengandung opini yang beragam, baik positif, negatif, maupun netral, sehingga dapat dimanfaatkan sebagai sumber data untuk analisis sentimen guna memahami persepsi publik terhadap performa pemain.

Analisis sentimen merupakan salah satu bidang dalam Natural Language Processing (NLP) yang bertujuan untuk mengidentifikasi dan mengklasifikasikan opini dalam bentuk teks. Perkembangan model berbasis *transfer learning* juga semakin memperkuat kemampuan analisis teks berbahasa Indonesia. Melalui pendekatan tersebut, model IndoBERT mampu memanfaatkan pengetahuan hasil *pre-training* sehingga dapat digunakan secara efektif pada berbagai tugas Natural Language Processing, termasuk klasifikasi sentimen dan klasifikasi emo.(Shaw, LaCasse, and Champagne 2025) Teknik ini banyak digunakan untuk menganalisis

data dari media sosial karena mampu mengolah data dalam jumlah besar secara otomatis dan efisien (Mao, Liu, and Zhang 2024). Selain itu, media sosial dianggap sebagai sumber data yang relevan karena mencerminkan opini publik secara aktual dan dinamis.

Seiring dengan perkembangan teknologi, metode analisis sentimen telah mengalami evolusi dari pendekatan berbasis leksikon dan machine learning tradisional menuju metode berbasis deep learning. Berbagai penelitian terkini menunjukkan bahwa model berbasis Transformer, khususnya IndoBERT, mampu memberikan performa klasifikasi sentimen yang lebih baik pada data berbahasa Indonesia karena dapat memahami hubungan antarkata secara kontekstual. Selain digunakan pada analisis sentimen, IndoBERT juga telah berhasil diterapkan pada berbagai tugas klasifikasi teks lainnya, seperti deteksi berita palsu. Hal tersebut menunjukkan bahwa kemampuan representasi bahasa yang dimiliki IndoBERT cukup kuat untuk memahami konteks kalimat berbahasa Indonesia pada berbagai domain penelitian.(Isa, Nico, and Permana 2022) Kemampuan tersebut menjadikan IndoBERT sebagai salah satu model yang banyak digunakan dalam penelitian analisis sentimen pada berbagai domain, seperti ulasan hotel, media sosial, maupun layanan digital. (Singgalen 2025)Salah satu model yang banyak digunakan adalah BERT (Bidirectional Encoder Representations from Transformers), yang mampu memahami konteks bahasa secara lebih baik karena menggunakan pendekatan dua arah dalam pemrosesan teks . Model ini terbukti memberikan peningkatan performa yang signifikan dalam berbagai tugas NLP.(Devlin et al. n.d.)

Dalam konteks bahasa Indonesia, pengembangan model BERT menghasilkan IndoBERT, yaitu model yang telah dilatih menggunakan korpus berbahasa Indonesia. Selain arsitektur model yang mampu memahami konteks bahasa Indonesia, penelitian terbaru juga menunjukkan bahwa proses *fine-tuning* pada IndoBERT mampu meningkatkan performa model ketika diterapkan pada domain tertentu. Dengan melakukan penyesuaian terhadap karakteristik dataset penelitian, model dapat menghasilkan klasifikasi sentimen yang lebih akurat dibandingkan penggunaan model tanpa proses penyesuaian domain. (Perwira et al. 2025)IndoBERT dirancang untuk meningkatkan kinerja analisis teks dalam bahasa Indonesia dan telah terbukti efektif dalam berbagai penelitian NLP, termasuk analisis sentimen(Koto et al. n.d.)

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengimplementasikan model IndoBERT dalam melakukan analisis sentimen terhadap komentar Instagram terkait performa Hokky Caraka. Dengan memanfaatkan teknologi NLP berbasis deep learning, diharapkan penelitian ini dapat memberikan gambaran yang lebih objektif mengenai persepsi publik serta berkontribusi dalam pengembangan analisis sentimen

berbahasa Indonesia. Keberhasilan penerapan IndoBERT tidak hanya ditemukan pada data media sosial, tetapi juga pada analisis ulasan aplikasi layanan kesehatan. Hasil penelitian tersebut menunjukkan bahwa IndoBERT mampu mengklasifikasikan sentimen masyarakat secara efektif sehingga menjadi salah satu pendekatan yang sesuai untuk berbagai jenis data teks berbahasa Indonesia(Imaduddin, Yusfida A'la, and Nugroho n.d.)

1.2 Rumusan Masalah

Berdasarkan latar belakang penelitian, maka permasalahan yang akan dikaji dalam penelitian ini dirumuskan sebagai berikut:

1. Bagaimana proses implementasi model IndoBERT untuk menganalisis sentimen komentar pengguna Instagram terhadap performa Hokky Caraka?
2. Seberapa baik kinerja model IndoBERT dalam mengklasifikasikan sentimen komentar Instagram ke dalam kategori positif, negatif, dan netral berdasarkan nilai accuracy, precision, recall, dan F1-score?
3. Bagaimana gambaran persepsi publik terhadap performa Hokky Caraka berdasarkan hasil analisis sentimen yang diperoleh?

1.3 Tujuan

Sejalan dengan permasalahan penelitian yang telah dirumuskan, penelitian ini bertujuan untuk:

1. Mengolah dan menyiapkan data komentar Instagram yang berkaitan dengan performa Hokky Caraka agar dapat digunakan dalam proses analisis sentimen.
2. Menerapkan model IndoBERT untuk mengklasifikasikan sentimen komentar ke dalam kategori positif, negatif, dan netral.
3. Mengevaluasi hasil klasifikasi serta melihat kecenderungan opini publik terhadap performa Hokky Caraka berdasarkan analisis sentimen yang diperoleh.

1.4 Manfaat

Berdasarkan tujuan penelitian yang telah dirumuskan, penelitian ini diharapkan mampu memberikan kontribusi sebagai berikut:

1. Manfaat Teoritis Secara teoritis, penelitian ini diharapkan dapat menambah referensi dalam bidang Natural Language Processing (NLP), khususnya terkait penerapan model IndoBERT untuk analisis sentimen pada data berbahasa Indonesia yang berasal dari media sosial.
2. Manfaat Praktis Secara praktis, hasil penelitian ini dapat memberikan gambaran mengenai persepsi publik terhadap performa Hokky Caraka berdasarkan komentar di Instagram. Selain itu, penelitian ini juga dapat menjadi acuan bagi pihak-pihak terkait, seperti analis olahraga atau pengelola media sosial, dalam memahami opini masyarakat secara lebih sistematis.
3. Manfaat Akademis Dari sisi akademis, penelitian ini diharapkan dapat menjadi bahan rujukan bagi mahasiswa atau peneliti lain yang ingin mengembangkan penelitian serupa, khususnya yang berkaitan dengan analisis sentimen menggunakan model berbasis deep learning di Indonesia

1.5 Ruang Lingkup Masalah

Agar penelitian ini memiliki arah dan batasan yang jelas, maka ruang lingkup penelitian dibatasi sebagai berikut:

1. Data yang digunakan dalam penelitian ini berupa komentar pengguna Instagram yang berkaitan dengan performa Hokky Caraka.
2. Proses analisis difokuskan pada klasifikasi sentimen teks ke dalam tiga kategori, yaitu positif, negatif, dan netral.
3. Metode yang digunakan dalam penelitian ini adalah model IndoBERT sebagai pendekatan utama dalam analisis sentimen.
4. Tahapan pengolahan data meliputi case folding, cleaning, normalisasi kata, dan tokenisasi menggunakan tokenizer IndoBERT.
5. Evaluasi kinerja model dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score.
6. Penelitian ini tidak membahas faktor di luar teks komentar, seperti konteks pertandingan secara mendalam atau aspek non-teknis lainnya.

1.6 Kebaharuan

Penelitian ini memiliki nilai kebaruan pada fokus kajian yang mengangkat analisis sentimen terhadap performa pemain sepak bola muda Indonesia, yaitu Hokky Caraka, dengan memanfaatkan data komentar dari Instagram. Topik ini masih jarang dibahas secara spesifik, terutama dalam konteks analisis opini publik terhadap atlet sepak bola di Indonesia.

Selain itu, penelitian ini menggunakan model IndoBERT yang dirancang khusus untuk memahami bahasa Indonesia. Pendekatan ini diharapkan mampu memberikan hasil klasifikasi yang lebih baik dibandingkan metode sebelumnya yang belum sepenuhnya menyesuaikan karakteristik bahasa Indonesia, khususnya dalam konteks bahasa informal yang sering digunakan di media sosial.

Kebaruan lainnya terletak pada penggunaan data komentar yang bersifat aktual dan mencerminkan respons publik secara langsung. Dengan memanfaatkan data tersebut, penelitian ini tidak hanya berfokus pada aspek teknis klasifikasi sentimen, tetapi juga memberikan gambaran yang lebih nyata mengenai bagaimana opini masyarakat terbentuk terhadap performa seorang atlet di media sosial.