BAB II

TINJAUAN PUSTAKA

2.1 Penelitian yang Relevan

Sebagai bahan perbandingan dengan penelitian yang akan dilakukan, bagian ini akan mencantumkan beberapa penelitian yang telah dilakukan oleh peneliti sebelumnya dengan topik dan penggunaan metode yang sama dan hampir mirip serta pengembangan teori yang berkaitan pada penelitian ini.

2.1.1 Matriks Penelitian

Tabel 2. 1 Matriks Penelitian Terdahulu

No	Nama Penelitian	Koto, F., et al.
1	Judul Jurnal	IndoBERT: A Pre-trained Language Model for Indonesian NLP
	Jenis Data	Data teks berbahasa Indonesia dari berbagai domain NLP
	Sumber Data	Korpus besar bahasa Indonesia yang dikumpulkan dari berbagai sumber digital
	Metode	Penelitian ini menggunakan model IndoBERT berbasis arsitektur Transformer yang telah melalui proses pre-training pada dataset skala besar. Model kemudian diuji pada beberapa tugas NLP termasuk analisis sentimen.
	Hasil	Hasil penelitian menunjukkan bahwa IndoBERT mampu meningkatkan performa secara signifikan dibandingkan metode tradisional. Model ini sangat efektif dalam memahami konteks bahasa Indonesia dan menghasilkan akurasi tinggi pada berbagai tugas klasifikasi teks
	Model Evaluasi	Evaluasi dilakukan menggunakan accuracy, precision, recall, dan F1-score
	Tahun	2020
2	Nama Penelitian	Salsabila, A., et al.
	Judul Jurnal	Sentiment Analysis of Indonesian Social Media Using IndoBERT
	Jenis Data	Data teks dari media sosial berbahasa Indonesia
	Sumber Data	Twitter dengan jumlah data ribuan tweet
	Metode	Penelitian ini menerapkan model IndoBERT yang telah melalui proses fine-tuning untuk klasifikasi sentimen. Tahapan

		meliputi preprocessing, tokenisasi, pelatihan model, dan evaluasi hasil.
	Hasil	Model IndoBERT mampu mengklasifikasikan sentimen positif, negatif, dan netral dengan akurasi di atas 85%. Model menunjukkan performa yang stabil terutama pada data dengan struktur bahasa informal.
	Model Evaluasi	Confusion Matrix, accuracy, dan F1-score
	Tahun	2021
3	Nama Penelitian	Maulidi, R.
	Judul Jurnal	Analisis Sentimen Instagram Menggunakan BERT
	Jenis Data	Komentar pengguna Instagram
	Sumber Data	Data hasil scraping dari beberapa akun Instagram
	Metode	Penelitian ini menggunakan metode BERT untuk melakukan klasifikasi sentimen. Proses dimulai dari pengumpulan data, preprocessing, pelatihan model, hingga evaluasi performa model.
	Hasil	Hasil penelitian menunjukkan bahwa model BERT mampu memahami konteks komentar Instagram dengan cukup baik. Model berhasil mengklasifikasikan sentimen dengan tingkat akurasi yang tinggi, meskipun masih terdapat kesalahan pada komentar ambigu.
	Model Evaluasi	Accuracy, precision, recall, dan F1-score
	Tahun	2022
4	Nama Penelitian	Prasetyo, E.
	Judul Jurnal	Sentiment Analysis on Social Media Using Deep Learning
	Jenis Data	Data teks dari media sosial
	Sumber Data	Twitter dan Instagram
	Metode	Penelitian ini membandingkan metode LSTM dan BERT dalam melakukan analisis sentimen. Proses meliputi preprocessing, pelatihan dua model, dan evaluasi perbandingan performa.
	Hasil	Hasil menunjukkan bahwa model BERT memiliki performa yang lebih baik dibandingkan LSTM dalam memahami konteks kalimat yang kompleks. BERT menghasilkan akurasi yang lebih tinggi dan lebih stabil.
	Model Evaluasi	Accuracy dan F1-score
	Tahun	2021
5	Nama Penelitian	Kurniawan, A., et al.
	Judul Jurnal	Aspect-Based Sentiment Analysis for Indonesian Language
	Jenis Data	Data teks berbahasa Indonesia yang mengandung opini

		pengguna
	Sumber Data	Dataset ulasan pengguna yang dikumpulkan dari platform digital
	Metode	Penelitian ini menggunakan pendekatan *aspect-based sentiment analysis* dengan memanfaatkan teknik deep learning untuk mengidentifikasi aspek tertentu dalam teks serta menentukan polaritas sentimen. Proses dilakukan melalui tahapan preprocessing, ekstraksi fitur, serta pelatihan model berbasis neural network
	Hasil	Hasil penelitian menunjukkan bahwa model mampu mengidentifikasi aspek dan sentimen secara cukup akurat pada data berbahasa Indonesia. Pendekatan ini memberikan hasil yang lebih detail dibandingkan klasifikasi sentimen umum karena mampu menganalisis opini berdasarkan aspek tertentu dalam teks.
	Model Evaluasi	Evaluasi dilakukan menggunakan accuracy, precision, recall, dan F1-score
	Tahun	2019
6	Nama Penelitian	Minaee, S., et al.
	Judul Jurnal	Deep Learning Based Text Classification: A Comprehensive Review
	Jenis Data	Data teks dari berbagai domain seperti media sosial, berita, dan ulasan pengguna
	Sumber Data	Dataset publik yang digunakan dalam berbagai penelitian klasifikasi teks
	Metode	Penelitian ini merupakan studi literatur yang membahas berbagai metode deep learning untuk klasifikasi teks, termasuk CNN, RNN, LSTM, dan Transformer. Setiap metode dianalisis berdasarkan cara kerja serta kemampuannya dalam memahami konteks teks.
	Hasil	Hasil penelitian menunjukkan bahwa metode deep learning memiliki performa yang lebih baik dibandingkan metode tradisional dalam klasifikasi teks. Model berbasis Transformer menunjukkan keunggulan dalam memahami konteks bahasa secara lebih mendalam.
	Model Evaluasi	Accuracy, precision, recall, dan F1-score dari berbagai penelitian yang dianalisis
	Tahun	2021
7	Nama Penelitian	Putri, D.
	Judul Jurnal	Sentiment Analysis on Instagram Comments Using Deep Learning

	Jenis Data	Komentar Instagram
	Sumber Data	Instagram API
	Metode	Penelitian ini menggunakan model CNN untuk melakukan klasifikasi sentimen. Proses mencakup preprocessing data, embedding, pelatihan model CNN, dan evaluasi.
	Hasil	Model CNN mampu mengklasifikasikan sentimen dengan cukup baik. Namun, performa masih di bawah model berbasis Transformer dalam memahami konteks yang kompleks.
	Model Evaluasi	Accuracy, precision, recall, dan F1-score
	Tahun	2021
8	Nama Penelitian	Zhang, L., et al.
	Judul Jurnal	A Survey on Sentiment Analysis using Deep Learning
	Jenis Data	Data teks dari berbagai domain seperti media sosial, ulasan produk, dan berita
	Sumber Data	Berbagai dataset penelitian yang digunakan dalam studi sebelumnya
	Metode	Penelitian ini merupakan studi literatur yang mengkaji perkembangan metode deep learning dalam analisis sentimen, termasuk penggunaan CNN, RNN, serta model berbasis Transformer. Analisis dilakukan dengan membandingkan performa masing-masing metode pada berbagai jenis data.
	Hasil	Hasil penelitian menunjukkan bahwa model berbasis deep learning, khususnya yang menggunakan arsitektur modern seperti Transformer, memiliki kinerja yang lebih unggul dalam memahami konteks dan menghasilkan klasifikasi sentimen yang lebih akurat.
	Model Evaluasi	Accuracy, precision, recall, dan F1-score
	Tahun	2018
9	Nama Penelitian	Saputra, R.
	Penulis	R. Saputra, D. Pratama, dan A. Nugraha
	Judul Jurnal	Analisis Sentimen Bahasa Indonesia Menggunakan Transformer
	Jenis Data	Data ulasan pengguna
	Sumber Data	Google Play Store
	Metode	Penelitian ini menggunakan model berbasis Transformer untuk klasifikasi sentimen. Proses meliputi data cleaning, tokenisasi, pelatihan model, dan evaluasi.
	Hasil	Model Transformer mampu menghasilkan akurasi tinggi terutama pada dataset berukuran besar. Model juga mampu menangkap hubungan konteks antar kata dengan baik.
	Model Evaluasi	Confusion Matrix dan accuracy

	Tahun	2023
10	Nama Penelitian	Nugroho, F.
	Penulis	F. Nugroho dan M. Rizky
	Judul Jurnal	Sentiment Analysis pada Komentar Sepak Bola di Media Sosial
	Jenis Data	Komentar terkait sepak bola
	Sumber Data	Twitter dan Instagram
	Metode	Penelitian ini menggunakan metode LSTM untuk menganalisis sentimen komentar terhadap pemain sepak bola.
	Hasil	Hasil menunjukkan bahwa model mampu menangkap opini publik terhadap performa pemain. Namun, masih terdapat kesalahan pada komentar yang
	Model Evaluasi	Accuracy dan F1-score
	Tahun	2022
11	Nama Penelitian	Firmansyah, M.
	Penulis	M. Firmansyah, A. Setiawan, dan R. Hidayat
	Judul Jurnal	Analisis Sentimen Pemain Sepak Bola Menggunakan NLP
	Jenis Data	Komentar netizen
	Sumber Data	Instagram
	Metode	Penelitian ini menggunakan IndoBERT untuk klasifikasi sentimen terhadap pemain sepak bola. Proses meliputi preprocessing, pelabelan data, pelatihan model, dan evaluasi.
	Hasil	Model IndoBERT mampu mengklasifikasikan sentimen dengan baik pada tiga kelas. Model menunjukkan performa tinggi dalam memahami bahasa informal yang sering digunakan di media sosial.
	Model Evaluasi	Accuracy, precision, dan recall
	Tahun	2023

Penelitian mengenai analisis sentimen pada media sosial terus berkembang seiring meningkatnya penggunaan platform digital sebagai sarana penyampaian opini publik. Dalam konteks bahasa Indonesia, penggunaan model IndoBERT dinilai mampu memberikan hasil yang lebih optimal karena telah dilatih menggunakan korpus berbahasa Indonesia(Wilie et al. n.d.) . Beberapa penelitian menunjukkan bahwa penerapan IndoBERT dalam analisis sentimen mampu menghasilkan tingkat akurasi yang tinggi, terutama pada data media sosial yang memiliki karakteristik bahasa informal Oleh karena itu, penelitian ini berfokus pada penerapan model IndoBERT untuk menganalisis sentimen komentar pengguna Instagram terhadap performa Hokky Caraka. Melalui pendekatan ini, diharapkan dapat diperoleh hasil klasifikasi sentimen yang lebih akurat serta mampu memberikan gambaran mengenai persepsi publik

terhadap performa pemain tersebut di media sosial.

2.1.2 Research Gap

Berdasarkan hasil telaah terhadap penelitian terdahulu, ditemukan beberapa kesenjangan penelitian (research gap) yang masih perlu dikaji lebih lanjut.

Pertama, sebagian besar penelitian analisis sentimen masih menggunakan metode machine learning dan deep learning konvensional, seperti Naïve Bayes, Support Vector Machine (SVM), Convolutional Neural Network (CNN), dan Long Short-Term Memory (LSTM). Meskipun metode-metode tersebut mampu menghasilkan kinerja yang cukup baik pada berbagai tugas klasifikasi teks, pendekatan tersebut memiliki keterbatasan dalam menangkap konteks semantik secara menyeluruh, terutama pada teks media sosial yang bersifat informal, mengandung ambiguitas, serta memiliki variasi bahasa yang tinggi. Keterbatasan tersebut mendorong berkembangnya model berbasis Transformer, seperti BERT, yang mampu memahami hubungan kontekstual antarkata secara lebih efektif melalui mekanisme representasi dua arah.(Tan, Lee, and Lim 2023)

Kedua, penelitian yang memanfaatkan model berbasis Transformer, khususnya IndoBERT, telah banyak diterapkan pada berbagai domain, seperti data Twitter, ulasan produk, ulasan aplikasi, maupun komentar Instagram. Namun, sebagian besar penelitian tersebut berfokus pada analisis sentimen secara umum atau pada objek seperti organisasi, layanan publik, dan Tim Nasional Indonesia. Penelitian yang secara khusus menganalisis sentimen komentar Instagram terhadap performa pemain sepak bola Indonesia secara individual masih relatif terbatas, sehingga masih terdapat peluang untuk mengembangkan penelitian pada konteks tersebut.(Abidin and Pamungkas 2025)

Ketiga, berdasarkan telaah literatur yang telah dilakukan, penelitian yang secara khusus menerapkan model IndoBERT untuk menganalisis sentimen komentar Instagram terhadap performa Hokky Caraka masih relatif terbatas. Sebagian besar penelitian sebelumnya lebih banyak berfokus pada analisis sentimen terhadap objek umum, organisasi, atau tim olahraga, sedangkan kajian yang menjadikan pemain sepak bola Indonesia secara individual sebagai objek penelitian masih jarang dilakukan. Kondisi tersebut menunjukkan adanya peluang untuk mengimplementasikan model IndoBERT dalam menganalisis persepsi publik terhadap performa Hokky Caraka melalui komentar Instagram.(Koto et al. n.d.)

Berdasarkan kesenjangan tersebut, penelitian ini mengusulkan penggunaan model IndoBERT untuk menganalisis sentimen komentar Instagram terhadap performa Hokky Caraka. Penggunaan IndoBERT diharapkan mampu menghasilkan klasifikasi sentimen yang

lebih akurat karena model ini telah dilatih menggunakan korpus berbahasa Indonesia serta memiliki kemampuan memahami konteks bahasa secara lebih baik melalui representasi kontekstual dua arah (*bidirectional contextual representation*) dibandingkan metode klasifikasi teks konvensional Dengan demikian, model IndoBERT diharapkan mampu memberikan hasil analisis sentimen yang lebih akurat terhadap komentar pengguna Instagram mengenai performa Hokky Caraka(Wilie et al. n.d.)

2.1.3 Alasan pemilihan Indobert

Pemilihan model IndoBERT dalam penelitian ini didasarkan pada beberapa pertimbangan. IndoBERT merupakan model bahasa berbasis Transformer yang telah dilatih menggunakan korpus bahasa Indonesia dalam jumlah besar sehingga mampu memahami karakteristik bahasa Indonesia secara lebih baik dibandingkan model umum yang dilatih menggunakan bahasa multibahasa.

Berbagai penelitian menunjukkan bahwa IndoBERT mampu menghasilkan performa yang lebih tinggi dibandingkan metode machine learning tradisional maupun beberapa model deep learning sebelumnya dalam berbagai tugas Natural Language Processing (NLP), termasuk analisis sentimen. Kemampuan mekanisme self-attention pada arsitektur Transformer memungkinkan IndoBERT memahami hubungan antar kata secara kontekstual sehingga lebih efektif dalam menangani kalimat yang kompleks.(Devlin et al. n.d.)

Selain itu, komentar Instagram memiliki karakteristik berupa bahasa informal, singkatan, bahasa gaul, serta penggunaan kata yang kontekstual. Karakteristik tersebut memerlukan model yang mampu memahami makna kalimat secara menyeluruh. Oleh karena itu, IndoBERT dipilih karena dinilai lebih sesuai untuk mengolah data komentar Instagram berbahasa Indonesia dibandingkan metode konvensional lainnya.

Tabel 2. 2 Perbandingan Metode Klasifikasi Teks

Metode	Kelebihan	Kekurangan	Referensi
Naive Bayes	Memiliki proses klasifikasi yang sederhana, cepat, dan dapat bekerja dengan baik pada data teks berukuran kecil karena menggunakan pendekatan probabilitas.	Memiliki keterbatasan karena menganggap setiap kata bersifat independen sehingga kurang mampu memahami hubungan konteks	(Medhat, Hassan, and Korashy 2014)

		antar kata dalam kalimat.	
Support Vector Machine (SVM)	Mampu menghasilkan performa klasifikasi yang baik pada data teks dan efektif dalam menemukan batas pemisah antar kelas.	Memerlukan proses ekstraksi fitur manual seperti TF-IDF dan belum mampu memahami makna kalimat secara kontekstual.	(Brasoveanu, Moodie, and Agrawal 2020a)
Convolutional Neural Network (CNN)	Mampu melakukan ekstraksi fitur secara otomatis dan mengenali pola penting pada teks melalui proses convolution.	Lebih fokus terhadap pola lokal sehingga memiliki keterbatasan dalam memahami hubungan antar kata dengan jarak yang panjang.	(Kim 2014)
Long Short-Term Memory (LSTM)	Mampu memproses data teks secara berurutan dan menangkap informasi dari urutan kata melalui mekanisme memori.	Membutuhkan waktu pelatihan lebih lama dan memiliki keterbatasan dalam memahami konteks dua arah dibandingkan model Transformer.	(Brasoveanu, Moodie, and Agrawal 2020a)
IndoBERT	Memanfaatkan arsitektur Transformer dengan mekanisme self-attention sehingga mampu memahami konteks kata secara dua arah. Model ini juga telah dilatih menggunakan korpus bahasa Indonesia sehingga sesuai untuk analisis teks bahasa Indonesia.	Membutuhkan sumber daya komputasi lebih besar dibandingkan metode machine learning tradisional.	(Wilie et al. n.d.)

Berdasarkan perbandingan beberapa metode analisis sentimen pada Tabel 2.2, setiap metode memiliki karakteristik serta kemampuan yang berbeda dalam melakukan klasifikasi teks. Metode machine learning tradisional seperti Naive Bayes dan Support Vector Machine (SVM) memiliki keunggulan dalam proses komputasi yang sederhana, namun masih bergantung pada ekstraksi fitur manual.

Metode deep learning seperti CNN dan LSTM mampu melakukan pembelajaran fitur secara otomatis sehingga dapat meningkatkan performa klasifikasi teks. Namun, kedua metode tersebut masih memiliki keterbatasan dalam memahami hubungan konteks kata secara menyeluruh pada kalimat yang kompleks.

Model berbasis Transformer seperti IndoBERT mampu mengatasi keterbatasan tersebut melalui mekanisme self-attention yang memungkinkan model memahami hubungan antar kata secara dua arah. Selain itu, IndoBERT telah melalui proses pre-training menggunakan korpus bahasa Indonesia sehingga lebih sesuai digunakan untuk pengolahan teks berbahasa Indonesia, terutama pada komentar media sosial yang mengandung bahasa informal dan variasi penulisan.

Oleh karena itu, penelitian ini menggunakan IndoBERT sebagai metode utama dalam analisis sentimen komentar Instagram terhadap performa Hokky Caraka karena memiliki kemampuan pemahaman konteks yang lebih baik dibandingkan metode sebelumnya.

2.2 Landasan Teori

Landasan teori merupakan kerangka konseptual yang menjadi dasar ilmiah penelitian dalam mengidentifikasi dan menyusun konsep serta variabel secara sistematis. Kerangka ini mengintegrasikan teori, metode, dan temuan penelitian sehingga proses analisis berlangsung secara logis, konsisten, dan dapat dipertanggungjawabkan secara akademik

2.2.1 IndoBERT

IndoBERT merupakan model *pre-trained* dalam bidang *Natural Language Processing* (NLP) yang dikembangkan khusus untuk bahasa Indonesia dengan memanfaatkan arsitektur Transformer. Model ini dilatih menggunakan korpus teks berbahasa Indonesia dalam jumlah besar sehingga mampu memahami struktur bahasa, konteks kalimat, serta variasi penggunaan kata dalam bahasa Indonesia secara lebih baik (Koto et al. n.d.). Kemampuan tersebut menjadikan IndoBERT banyak digunakan dalam berbagai tugas NLP, termasuk analisis sentimen, klasifikasi teks.

Dalam konteks analisis sentimen pada komentar Instagram, IndoBERT memiliki keunggulan dalam memahami karakteristik bahasa yang digunakan oleh pengguna, yang umumnya bersifat informal, mengandung singkatan, serta penggunaan kata yang tidak baku. Kemampuan tersebut didukung oleh proses *pre-training* menggunakan korpus berbahasa Indonesia yang beragam sehingga model mampu mempelajari karakteristik linguistik bahasa Indonesia secara lebih komprehensif (Wilie et al. n.d.)Selain itu, mekanisme *self-attention* pada arsitektur Transformer memungkinkan IndoBERT memahami hubungan antarkata secara kontekstual, sehingga mampu menangkap makna suatu kalimat dengan lebih akurat dibandingkan pendekatan yang hanya mengandalkan representasi kata secara terpisah. Berbagai penelitian menunjukkan bahwa IndoBERT memiliki kemampuan representasi bahasa yang baik karena dibangun menggunakan arsitektur Transformer serta telah melalui proses pre-training pada korpus bahasa Indonesia berskala besar. Dengan kemampuan tersebut, IndoBERT mampu menangkap hubungan kontekstual antarkata sehingga menghasilkan performa yang tinggi pada berbagai tugas klasifikasi teks, termasuk analisis sentimen, klasifikasi dokumen, maupun klasifikasi multilabel.(Singgalen 2025; Wilie et al. n.d.)

Oleh karena itu, proses klasifikasi sentimen ke dalam kategori positif, negatif, dan netral dapat dilakukan secara lebih efektif pada data komentar media sosial berbahasa Indonesia. Secara arsitektural, IndoBERT dibangun berdasarkan komponen *encoder* dari arsitektur Transformer yang memungkinkan pemrosesan teks dilakukan secara paralel. Mekanisme *self-attention* yang diterapkan memungkinkan setiap kata dalam suatu kalimat berinteraksi dengan seluruh kata lainnya sehingga model mampu memahami hubungan kontekstual secara menyeluruh dalam satu proses komputasi (Devlin et al. n.d.)Selain itu, penggunaan *positional encoding* memungkinkan model mengenali informasi urutan kata, sehingga hubungan antarposisi dalam kalimat tetap dipertahankan selama proses pembelajaran meskipun pemrosesan dilakukan secara paralel (Vaswani et al. 2023)

Pada tahap pelatihan, IndoBERT memanfaatkan proses pembelajaran berbasis konteks dari data dalam jumlah besar, sehingga mampu menghasilkan representasi bahasa yang kaya dan mendalam. Model ini kemudian dapat disesuaikan (*fine-tuning*) dengan data spesifik sesuai kebutuhan penelitian. Dalam penelitian ini, IndoBERT digunakan untuk menganalisis sentimen komentar pengguna Instagram terhadap performa Hokky Caraka, sehingga diharapkan dapat memberikan hasil klasifikasi yang akurat serta menggambarkan persepsi publik secara lebih komprehensif.Beberapa penelitian terdahulu menunjukkan bahwa IndoBERT mampu memberikan performa yang lebih baik dibandingkan metode klasifikasi teks lainnya. (Koto et al. n.d.) menunjukkan bahwa IndoBERT berhasil memperoleh hasil yang

kompetitif pada berbagai benchmark NLP bahasa Indonesia. Selain itu, penelitian (Wilie et al. n.d.) juga menunjukkan bahwa model berbasis IndoBERT menghasilkan performa yang unggul pada berbagai tugas klasifikasi teks berbahasa Indonesia. Temuan tersebut menunjukkan bahwa IndoBERT merupakan salah satu model yang relevan untuk diterapkan pada penelitian analisis sentimen komentar Instagram terhadap performa Hokky Caraka. Selain digunakan pada analisis sentimen, IndoBERT juga telah diterapkan pada berbagai tugas Natural Language Processing lainnya, seperti klasifikasi multilabel, deteksi berita palsu, klasifikasi soal ujian, hingga klasifikasi emosi. Hal tersebut menunjukkan bahwa IndoBERT memiliki kemampuan generalisasi yang baik pada berbagai karakteristik data teks berbahasa Indonesia.(Nabiilah et al. 2024)

2.2.1.1 Performa Hokky Caraka

Dalam penelitian ini, performa Hokky Caraka merujuk pada penilaian publik terhadap penampilan dan kontribusinya sebagai pemain sepak bola yang tercermin melalui komentar pengguna Instagram. Penilaian tersebut dapat berkaitan dengan kemampuan bermain, kontribusi terhadap tim, efektivitas dalam pertandingan, konsistensi penampilan, kemampuan mencetak gol, serta respons pengguna terhadap keputusan dan aksi yang dilakukan selama pertandingan. Oleh karena itu, performa dalam penelitian ini tidak diukur secara langsung menggunakan statistik pertandingan, tetapi dianalisis berdasarkan kecenderungan sentimen yang terdapat dalam komentar Instagram.

Selain aspek teknis di lapangan, persepsi pengguna dalam dataset juga menunjukkan adanya penilaian terhadap aspek non-teknis, khususnya sikap, attitude, perilaku, dan cara pemain menunjukkan respons dalam situasi tertentu. Aspek tersebut digunakan sebagai bagian dari konteks persepsi publik terhadap Hokky Caraka karena komentar pengguna tidak hanya membahas kemampuan bermain, tetapi juga memberikan penilaian terhadap sikap yang ditampilkan oleh pemain. Beberapa komentar dalam dataset secara langsung menggunakan istilah seperti “attitude”, “respect”, dan “sombong” untuk memberikan penilaian terhadap perilaku Hokky. Salah satu komentar menyatakan bahwa “kalau selebrasi jangan tengil, attitude itu penting”, sedangkan komentar lainnya menyebutkan “BAD ATTITUDE” dan “Tak punya attitude sebagai pemain kaga punya respec sama tim lawan”. Terdapat pula komentar yang menyebutkan “Ketok sombonge” sebagai bentuk penilaian negatif terhadap sikap yang dipersepsikan oleh pengguna.

Persepsi terhadap aspek di luar lapangan juga muncul dalam komentar yang berkaitan dengan kehidupan personal yang terlihat melalui media sosial. Salah satu komentar dalam

dataset menyebutkan bahwa pengguna melihat unggahan atau story Hokky dan kemudian memberikan penilaian terhadap hubungan personalnya. Hal tersebut menunjukkan bahwa sebagian pengguna dapat mengaitkan citra pemain di media sosial dengan penilaian mereka terhadap dirinya. Namun, komentar tersebut tidak dapat digunakan sebagai bukti objektif mengenai kepribadian atau perilaku Hokky di luar lapangan. Dalam penelitian ini, komentar tersebut hanya dipandang sebagai bentuk persepsi pengguna yang muncul dalam data media sosial.

Dengan demikian, aspek performa Hokky Caraka dalam penelitian ini mencakup dua konteks, yaitu performa dalam pertandingan dan persepsi terhadap aspek non-teknis yang ditampilkan melalui media sosial. Performa dalam pertandingan mencakup kemampuan bermain, kontribusi terhadap tim, konsistensi, dan efektivitas permainan. Sementara itu, aspek non-teknis mencakup persepsi pengguna terhadap attitude, sikap, respect, perilaku, serta citra personal yang ditampilkan melalui media sosial. Kedua aspek tersebut dapat memengaruhi kecenderungan sentimen pengguna yang kemudian diklasifikasikan ke dalam sentimen positif, negatif, dan netral.

Analisis sentimen digunakan untuk mengidentifikasi kecenderungan persepsi pengguna terhadap performa Hokky Caraka ke dalam sentimen positif, negatif, dan netral. Sentimen positif menunjukkan komentar yang memberikan penilaian atau tanggapan yang mendukung terhadap performa maupun sikap yang dipersepsikan positif. Sentimen negatif menunjukkan kritik atau penilaian kurang baik terhadap kemampuan bermain, tindakan dalam pertandingan, attitude, atau perilaku yang dipersepsikan oleh pengguna. Sementara itu, sentimen netral menunjukkan komentar yang tidak memberikan kecenderungan penilaian positif maupun negatif. Dengan pendekatan tersebut, hasil analisis dapat memberikan gambaran mengenai bagaimana pengguna Instagram mempersepsikan Hokky Caraka, baik dari aspek performa di lapangan maupun aspek non-teknis yang muncul dalam interaksi dan pembahasan di media sosial.

2.2.2 Deep Learning

Deep Learning merupakan salah satu cabang dari machine learning yang memanfaatkan jaringan saraf tiruan berlapis (*deep neural networks*) untuk mengolah data dalam jumlah besar dan kompleks secara lebih efektif. Pendekatan ini mampu mempelajari representasi data secara otomatis tanpa memerlukan ekstraksi fitur secara manual, sehingga sangat cocok digunakan dalam berbagai tugas seperti pengolahan bahasa alami (*Natural Language Processing*), termasuk analisis sentimen pada media sosial. Dalam konteks penelitian ini, Deep Learning

digunakan untuk mengolah data komentar Instagram yang memiliki karakteristik bahasa yang beragam, tidak baku, dan kontekstual. Perkembangan model berbasis Transformer menjadi salah satu tonggak penting dalam evolusi deep learning pada bidang Natural Language Processing. Dibandingkan pendekatan CNN maupun LSTM, Transformer mampu memahami hubungan antarkata secara paralel melalui mekanisme self-attention sehingga menghasilkan representasi bahasa yang lebih kaya.(Vaswani et al. 2023)

Berdasarkan arsitektur yang digunakan, Deep Learning memiliki beberapa model yang umum digunakan, seperti *Convolutional Neural Network* (CNN), *Recurrent Neural Network* (RNN), serta model berbasis Transformer. Model Transformer, seperti IndoBERT, menjadi salah satu pendekatan yang banyak digunakan karena kemampuannya dalam memahami konteks kalimat secara lebih mendalam melalui mekanisme *self-attention*. Dengan kemampuan tersebut, model ini dinilai lebih unggul dalam melakukan klasifikasi teks dibandingkan metode sebelumnya (Vaswani et al. 2023)

Perkembangan Deep Learning tidak terlepas dari meningkatnya ketersediaan data digital serta kemajuan teknologi komputasi. Deep Learning kembali berkembang pesat setelah diperkenalkan oleh Hinton et al. (2006), yang menjadi salah satu tonggak penting dalam pengembangan model pembelajaran berbasis jaringan saraf berlapis. Perkembangan tersebut mendorong kemajuan yang signifikan dalam berbagai bidang, termasuk *Natural Language Processing* (NLP), di mana model *Deep Learning* telah banyak dimanfaatkan untuk berbagai tugas, seperti klasifikasi teks, analisis sentimen, *question answering*, dan *machine translation* ((Hinton, Osindero, and Teh n.d.). Saat ini, pendekatan *Deep Learning* juga banyak diterapkan dalam analisis media sosial untuk mengidentifikasi dan memahami opini publik yang disampaikan melalui komentar pengguna terhadap berbagai isu, tokoh, maupun peristiwa (Zhang, Wang, and Liu n.d.). Oleh karena itu, penelitian ini mengimplementasikan model IndoBERT sebagai pendekatan *Deep Learning* untuk menganalisis sentimen komentar Instagram terhadap performa Hokky Caraka.

2.2.3 Analisis Sentimen

Analisis sentimen merupakan salah satu teknik dalam *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi serta mengklasifikasikan opini atau persepsi seseorang terhadap suatu topik tertentu berdasarkan data teks. Secara umum, analisis sentimen bertujuan untuk menentukan polaritas suatu teks ke dalam kategori positif, negatif, atau netral. Teknik ini banyak dimanfaatkan dalam berbagai bidang, terutama untuk memahami opini publik yang tersebar luas di media sosial, karena platform tersebut menjadi sarana utama bagi

masyarakat dalam menyampaikan pendapat secara terbuka dan spontan (Mao, Liu, and Zhang 2024)

Dalam penerapannya, analisis sentimen pada media sosial memiliki tantangan tersendiri. Data yang diperoleh biasanya bersifat tidak terstruktur, mengandung bahasa informal, singkatan, serta variasi ekspresi yang sangat beragam. Selain itu, penggunaan emotikon, kata tidak baku, hingga adanya sarkasme juga dapat mempengaruhi hasil analisis. Oleh karena itu, diperlukan pendekatan yang mampu memahami konteks bahasa secara mendalam agar proses klasifikasi sentimen dapat dilakukan dengan lebih akurat. Pendekatan berbasis deep learning, seperti IndoBERT, dinilai mampu mengatasi permasalahan tersebut karena memiliki kemampuan dalam memahami hubungan antar kata dalam suatu kalimat secara kontekstual (Koto et al. n.d.)

Dalam penelitian ini, analisis sentimen digunakan untuk mengolah komentar pengguna Instagram yang berkaitan dengan performa Hokky Caraka. Komentar-komentar tersebut kemudian diklasifikasikan ke dalam tiga kategori utama, yaitu sentimen positif, negatif, dan netral. Hasil dari proses ini diharapkan dapat memberikan gambaran yang lebih jelas mengenai persepsi publik terhadap performa pemain tersebut

2.2.4 Media Sosial (Instagram)

Media sosial merupakan platform digital yang memungkinkan pengguna untuk berinteraksi, berbagi informasi, serta menyampaikan opini secara terbuka tanpa dibatasi oleh ruang dan waktu. Perkembangan media sosial yang semakin pesat menjadikan platform ini sebagai salah satu sumber utama dalam memahami persepsi publik terhadap suatu topik tertentu. Melalui media sosial, pengguna dapat dengan mudah mengekspresikan pendapat, memberikan tanggapan, serta berpartisipasi dalam berbagai diskusi yang bersifat publik (Kaplan and Haenlein 2010)

Salah satu platform media sosial yang memiliki tingkat interaksi tinggi adalah Instagram. Platform ini tidak hanya digunakan untuk berbagi foto dan video, tetapi juga menyediakan fitur komentar yang memungkinkan pengguna untuk memberikan respon terhadap suatu konten. Komentar yang dihasilkan umumnya mencerminkan opini, emosi, serta persepsi pengguna terhadap topik yang sedang dibahas. Dalam konteks ini, komentar pada Instagram dapat dimanfaatkan sebagai sumber data untuk analisis sentimen karena mengandung berbagai bentuk ekspresi yang merepresentasikan pandangan masyarakat secara langsung.

Namun demikian, data yang berasal dari Instagram memiliki karakteristik yang cukup kompleks. Pengguna sering menggunakan bahasa informal, singkatan, serta variasi penulisan

yang tidak baku. Selain itu, penggunaan emotikon dan gaya bahasa yang beragam juga menjadi tantangan dalam proses analisis. Oleh karena itu, diperlukan metode yang mampu memahami konteks bahasa secara lebih mendalam agar hasil analisis dapat dilakukan secara lebih akurat (Mao, Liu, and Zhang 2024)

Dalam penelitian ini, Instagram digunakan sebagai sumber data utama untuk mengumpulkan komentar pengguna yang berkaitan dengan performa Hokky Caraka. Data tersebut kemudian dianalisis menggunakan pendekatan analisis sentimen untuk mengetahui kecenderungan opini publik, sehingga dapat memberikan gambaran yang lebih jelas mengenai persepsi masyarakat terhadap performa pemain tersebut di media sosial.

2.2.5 Preprocessing Teks

Preprocessing teks merupakan tahapan awal dalam pengolahan data teks yang bertujuan untuk membersihkan dan menyiapkan data agar dapat digunakan pada proses analisis selanjutnya. Data mentah yang diperoleh dari media sosial umumnya mengandung berbagai elemen yang tidak relevan, seperti tanda baca, simbol, angka, *hyperlink*, emoji, maupun kata-kata yang tidak memiliki makna penting. Oleh karena itu, diperlukan proses *preprocessing* untuk meningkatkan kualitas data sehingga mampu menghasilkan representasi teks yang lebih baik dan mendukung peningkatan kinerja model klasifikasi (Koto et al. n.d.) Tahapan preprocessing memiliki peranan penting dalam meningkatkan kualitas data sebelum proses pelatihan model dilakukan. Pembersihan data, normalisasi kata, serta tokenisasi yang sesuai terbukti mampu meningkatkan kualitas representasi teks sehingga proses fine-tuning menghasilkan performa klasifikasi yang lebih optimal.(Sathyanarayanan 2024; Singgalen 2025)

Dalam penelitian ini, tahapan preprocessing yang digunakan meliputi case folding, cleaning, normalisasi kata, dan tokenisasi IndoBERT. Tahapan tersebut bertujuan untuk membersihkan teks komentar Instagram serta menyesuaikan format data agar dapat diproses oleh model IndoBERT.Dalam penelitian ini, preprocessing dilakukan pada komentar Instagram sebelum data tersebut diproses menggunakan model IndoBERT. Dengan melalui tahapan ini, diharapkan data yang digunakan menjadi lebih bersih dan terstruktur sehingga dapat meningkatkan akurasi dalam proses klasifikasi sentimen.

2.2.6 Klasifikasi Sentimen

Klasifikasi sentimen merupakan proses pengelompokan data teks berdasarkan polaritas opini yang terkandung di dalamnya. Secara umum, klasifikasi sentimen membagi opini ke

dalam tiga kategori utama, yaitu positif, negatif, dan netral. Proses klasifikasi dilakukan dengan memanfaatkan model pembelajaran yang telah dilatih untuk mengenali pola dan karakteristik tertentu pada data teks sehingga mampu menentukan kategori sentimen secara otomatis (Aggarwal and Zhai 2012)

Dalam penelitian ini, proses klasifikasi sentimen dilakukan terhadap komentar pengguna Instagram dengan menggunakan model IndoBERT. Model ini memiliki kemampuan dalam memahami konteks bahasa Indonesia secara lebih mendalam, sehingga dapat mengklasifikasikan sentimen dengan tingkat akurasi yang lebih baik dibandingkan metode konvensional. Setiap komentar yang telah melalui tahap preprocessing akan dianalisis dan dikategorikan sesuai dengan polaritasnya.

Hasil klasifikasi sentimen ini kemudian digunakan untuk melihat kecenderungan opini publik terhadap performa Hokky Caraka. Dengan demikian, penelitian ini tidak hanya berfokus pada proses klasifikasi, tetapi juga pada interpretasi hasil untuk memahami persepsi masyarakat secara lebih luas.

2.2.7 Evaluasi Model

Evaluasi model merupakan tahap untuk mengukur kinerja model dalam mengklasifikasikan data menggunakan data uji (*testing data*) yang tidak digunakan selama proses pelatihan. Tahap ini bertujuan untuk mengetahui kemampuan model dalam melakukan prediksi secara akurat terhadap data baru. Pada penelitian ini, evaluasi dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score, serta Confusion Matrix untuk menganalisis hasil klasifikasi secara lebih rinci. Penggunaan metrik tersebut memberikan gambaran mengenai tingkat ketepatan prediksi, kemampuan model dalam mengenali setiap kelas sentimen, serta pola kesalahan klasifikasi yang dihasilkan oleh model. Selanjutnya, penjelasan mengenai masing-masing metrik evaluasi disajikan pada subbab berikutnya

2.2.8 CRISP-DM

CRISP-DM (*Cross Industry Standard Process for Data Mining*) merupakan kerangka kerja standar yang banyak digunakan dalam pelaksanaan proyek *data mining* di berbagai sektor industri Model ini menyediakan pendekatan yang sistematis dan terstruktur melalui enam tahapan utama, sehingga proses analisis data dapat dilaksanakan secara terarah, terintegrasi, dan sesuai dengan tujuan penelitian.

Tahap pertama yaitu *business understanding*, yang berfokus pada pemahaman terhadap tujuan, kebutuhan, serta permasalahan yang mendasari penelitian. Pada tahap ini dilakukan

perumusan masalah, penetapan tujuan, serta penyusunan rencana penelitian yang sesuai dengan kebutuhan analisis. (Wirth and Hipp n.d.) menjelaskan bahwa tahap ini berfokus pada penerjemahan tujuan dan kebutuhan ke dalam permasalahan *data mining* serta perancangan rencana penelitian.

Setelah permasalahan ditentukan, proses dilanjutkan ke tahap *data understanding*. Tahap ini meliputi pengumpulan data awal, eksplorasi karakteristik dan struktur data, identifikasi kualitas data, serta analisis awal untuk memperoleh pemahaman terhadap informasi yang terkandung dalam data. Pemahaman terhadap karakteristik data diperlukan sebelum data diproses lebih lanjut.

Selanjutnya, tahap *data preparation* mencakup proses pembersihan, transformasi, dan seleksi data agar berada dalam kondisi yang sesuai untuk pemodelan. Pada tahap ini juga dilakukan pelabelan data, yaitu pemberian kelas atau kategori pada data sesuai dengan tujuan klasifikasi. Setelah data memiliki label, dataset dapat dipersiapkan untuk proses pemodelan melalui tahapan preprocessing dan transformasi data.

Tahap berikutnya adalah *modeling*, yaitu proses membangun model menggunakan algoritma yang sesuai dengan tujuan penelitian. Pada tahap ini, data yang telah dipersiapkan dan diberi label dibagi menjadi data pelatihan **(***training data***)** sebesar 80% dan data pengujian **(***testing data***)** sebesar 20%**.** Data pelatihan digunakan untuk melatih model, sedangkan data pengujian digunakan untuk mengukur kemampuan model dalam melakukan prediksi terhadap data yang belum digunakan selama proses pelatihan. Selanjutnya, model dilatih dan disesuaikan untuk memperoleh performa yang optimal.

Setelah model terbentuk, dilakukan tahap *evaluation* untuk menilai kinerja model secara menyeluruh dan memastikan hasil yang diperoleh sesuai dengan tujuan penelitian. Evaluasi dilakukan menggunakan metrik yang sesuai dengan permasalahan klasifikasi, seperti Accuracy, Precision, Recall, dan F1-Score**,** serta dapat didukung dengan *confusion matrix* untuk melihat distribusi hasil prediksi pada setiap kelas. Tahap ini juga digunakan untuk mengetahui kesalahan klasifikasi yang dihasilkan oleh model.

Tahap terakhir adalah *deployment*, yaitu proses penerapan model yang telah selesai dilatih dan dievaluasi agar dapat digunakan pada data baru. Pada tahap ini, model dapat disimpan bersama komponen pendukungnya sehingga dapat digunakan kembali tanpa melakukan proses pelatihan dari awal. Dalam penelitian analisis sentimen, deployment memungkinkan model digunakan untuk melakukan klasifikasi sentimen terhadap komentar baru sesuai dengan kategori yang telah dipelajari selama proses pemodelan.

2.2.9 Flowchart

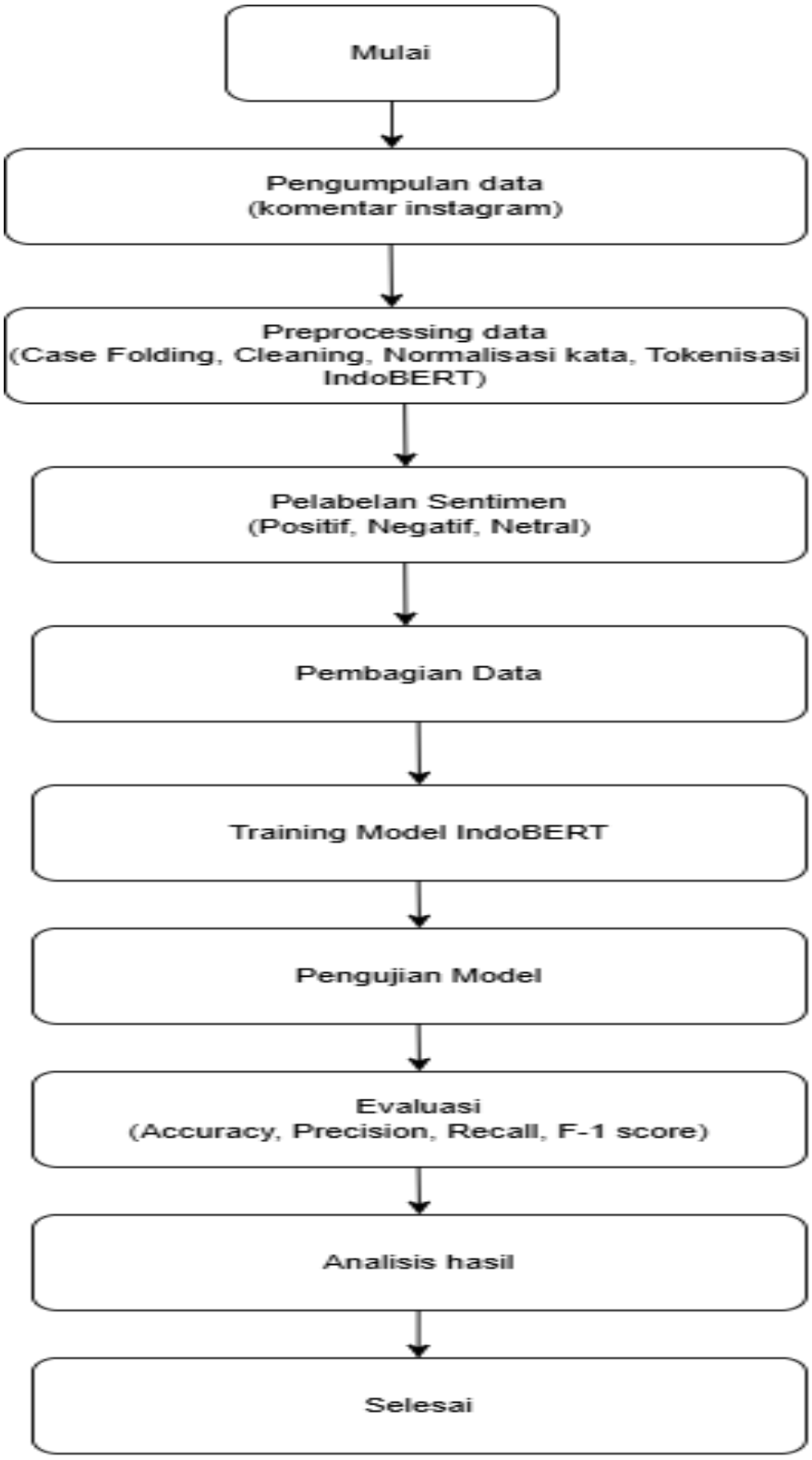
Flowchart penelitian ini menggambarkan alur sistematis dalam melakukan analisis sentimen terhadap komentar Instagram terkait performa Hokky Caraka dengan memanfaatkan model IndoBERT. Tahapan penelitian diawali dengan proses pengumpulan data, yaitu mengambil komentar dari platform Instagram yang relevan dengan objek penelitian. Pengumpulan data dari media sosial merupakan salah satu sumber utama dalam analisis sentimen karena mengandung opini publik yang bersifat real-time dan beragam

Setelah data terkumpul, dilakukan tahap *preprocessing* yang mencakup pembersihan teks (*cleaning*), penghapusan simbol, angka, tautan (*URL*), maupun karakter yang tidak diperlukan, normalisasi kata, serta tokenisasi. Tahapan ini bertujuan untuk meningkatkan kualitas data, mengurangi *noise*, dan menghasilkan representasi teks yang lebih baik sehingga dapat meningkatkan kinerja model pada proses klasifikasi sentimen (Brasoveanu, Moodie, and Agrawal 2020b)

Selanjutnya, data yang telah melalui tahap *preprocessing* diberi label sentimen ke dalam tiga kategori, yaitu positif, negatif, dan netral. Proses pelabelan ini berperan penting dalam membentuk dataset berlabel yang digunakan sebagai dasar pelatihan model klasifikasi sentimen. Setelah proses pelabelan selesai, dataset dibagi menjadi data latih (*training data*) dan data uji (*testing data*) untuk melatih model serta mengevaluasi kemampuannya dalam mengklasifikasikan data yang belum pernah dilihat sebelumnya (Koto et al. n.d.)

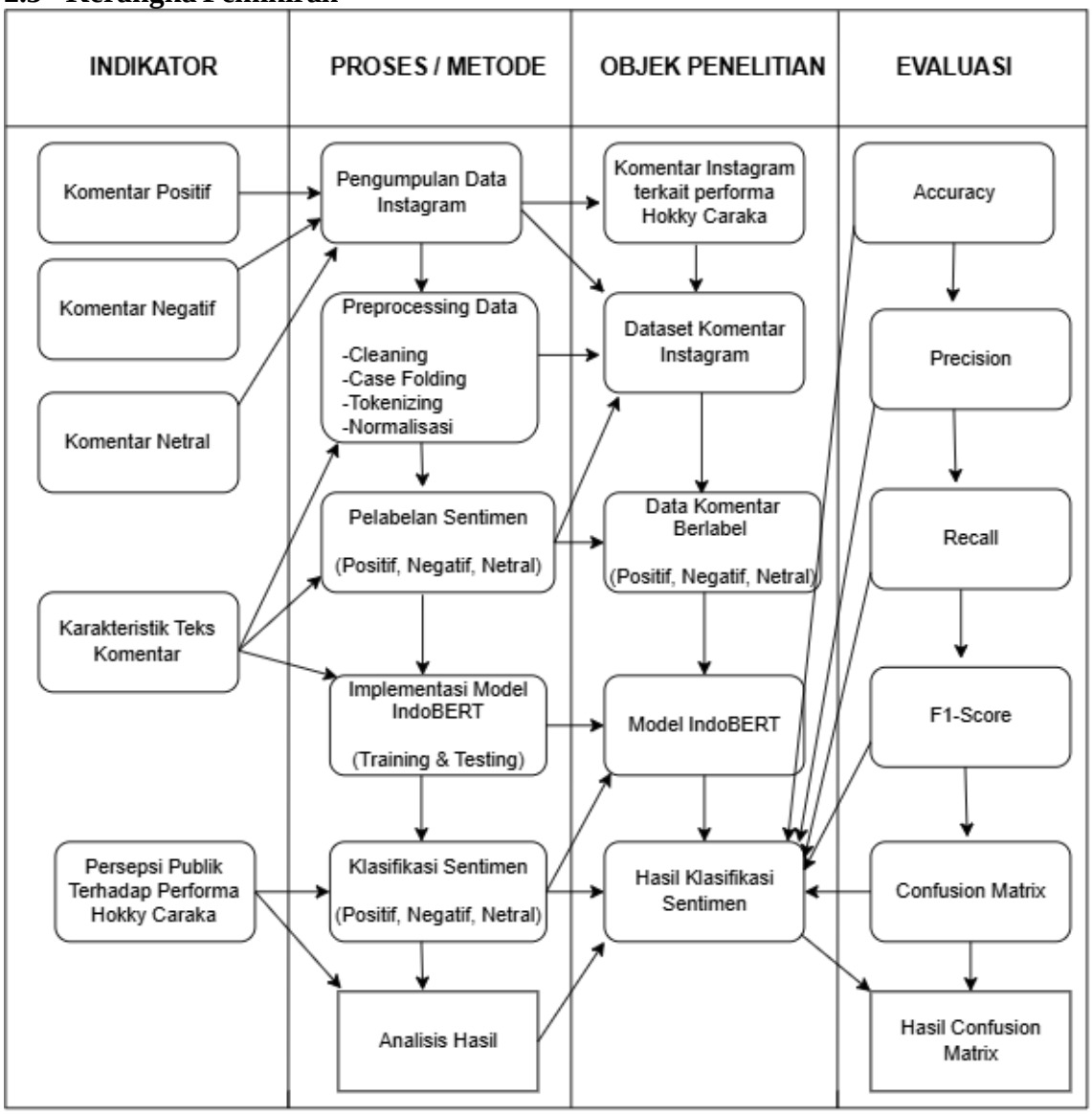
Pada tahap berikutnya, model IndoBERT diimplementasikan dan dilatih menggunakan data latih. IndoBERT merupakan model berbasis *Bidirectional Encoder Representations from Transformers* (BERT) yang telah diadaptasi untuk bahasa Indonesia, sehingga mampu memahami konteks kalimat secara lebih mendalam dibandingkan metode tradisional (Koto et al. n.d.)

Setelah proses pelatihan selesai, model diuji menggunakan data uji (*testing data*) untuk mengukur performanya dalam mengklasifikasikan sentimen. Evaluasi dilakukan menggunakan metrik *accuracy, precision, recall*, dan *F1-score*, yang merupakan metrik evaluasi umum pada tugas klasifikasi teks dan analisis sentimen. Metrik-metrik tersebut digunakan untuk mengukur tingkat ketepatan, kelengkapan, serta keseimbangan kinerja model dalam mengklasifikasikan setiap kategori sentimen Tahap terakhir adalah analisis hasil klasifikasi sentimen serta penarikan kesimpulan untuk memahami kecenderungan opini publik terhadap performa Hokky Caraka berdasarkan data yang telah diolah. (Brasoveanu, Moodie, and Agrawal 2020b)



Gambar 2. 1 Alur Penelitian Analisis Sentimen Menggunakan IndoBERT

2.3 Kerangka Pemikiran



Gambar 2. 2 Kerangka Pemikiran Penelitian Analisis Sentimen IndoBERT

Kerangka pemikiran pada penelitian ini menggambarkan alur yang digunakan untuk menganalisis sentimen komentar Instagram mengenai performa Hokky Caraka dengan memanfaatkan model IndoBERT sebagai metode klasifikasi. Proses penelitian dimulai dari pengumpulan komentar pengguna Instagram yang berkaitan dengan performa Hokky Caraka. Komentar tersebut menjadi sumber data utama yang akan digunakan dalam proses analisis sentimen.

Setelah data berhasil dikumpulkan, tahap berikutnya adalah preprocessing untuk meningkatkan kualitas data sebelum diproses oleh model. Tahapan ini meliputi proses cleaning untuk menghilangkan karakter yang tidak diperlukan, case folding untuk menyeragamkan huruf menjadi huruf kecil, normalisasi untuk memperbaiki kata yang tidak

baku, tokenizing untuk memisahkan kalimat menjadi token-token, serta stemming untuk mengubah kata menjadi bentuk dasarnya. Proses tersebut bertujuan agar data memiliki format yang lebih bersih dan konsisten sehingga dapat diproses secara optimal oleh model IndoBERT.

Data yang telah melalui tahap preprocessing kemudian diberikan label sentimen yang terdiri atas tiga kategori, yaitu positif, negatif, dan netral. Pelabelan dilakukan sebagai dasar dalam proses pembelajaran model sehingga setiap komentar memiliki kelas yang sesuai dengan makna yang terkandung di dalamnya.

Selanjutnya, dataset yang telah diberi label dibagi menjadi data pelatihan (training) dan data pengujian (testing). Data pelatihan digunakan untuk melakukan proses fine-tuning pada model IndoBERT sehingga model mampu mempelajari karakteristik bahasa yang terdapat pada komentar Instagram. Setelah proses pelatihan selesai, model digunakan untuk mengklasifikasikan sentimen pada data pengujian ke dalam kategori positif, negatif, atau netral.

Tahap terakhir adalah evaluasi terhadap performa model menggunakan beberapa metrik pengukuran, yaitu accuracy, precision, recall, F1-score, dan confusion matrix. Accuracy digunakan untuk mengetahui tingkat ketepatan prediksi secara keseluruhan. Precision mengukur ketepatan model dalam memberikan prediksi pada setiap kelas. Recall menunjukkan kemampuan model dalam mengenali seluruh data yang termasuk pada masing-masing kategori sentimen. F1-score digunakan untuk melihat keseimbangan antara precision dan recall.

Melalui rangkaian tahapan tersebut diharapkan model IndoBERT mampu mengklasifikasikan sentimen komentar Instagram terhadap performa Hokky Caraka secara akurat. Hasil klasifikasi kemudian digunakan sebagai dasar dalam menganalisis kecenderungan persepsi publik sehingga dapat memberikan gambaran mengenai respons pengguna Instagram terhadap performa pemain tersebut