

Implementasi Analisis Sentimen Komentar YouTube Berbasis IndoBERT dengan Evaluasi Confidence Score

Aji Pangestu¹, Frisqy Maulana²

^{1,2}Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri
E-mail: ¹fmkiki11@gmail.com, ²birugren04@gmail.com

Abstrak – Analisis sentimen terhadap opini publik di media sosial menjadi penting untuk memahami persepsi masyarakat terhadap suatu topik, khususnya dalam konteks olahraga nasional seperti Timnas Indonesia. Penelitian ini mengimplementasikan sistem analisis sentimen otomatis pada komentar YouTube menggunakan pre-trained model IndoBERT (Indonesian Bidirectional Encoder Representations from Transformers) untuk mengklasifikasikan sentimen ke dalam tiga kategori: positif, negatif, dan netral. Data dikumpulkan melalui YouTube Data API v3 dari empat video terkait Timnas Indonesia, menghasilkan 952 komentar unik setelah melalui tahap preprocessing yang mencakup case folding, text cleaning (penghapusan URL, mentions, hashtags), dan deduplication. Pelabelan sentimen dilakukan secara otomatis menggunakan model `w11wo/indonesian-roberta-base-sentiment-classifier` dengan batch processing. Evaluasi sistem menggunakan confidence score sebagai metrik internal menunjukkan performa yang sangat baik dengan rata-rata confidence 89.68%, konsisten across semua kategori sentimen (positif: 89.47%, negatif: 89.76%, netral: 89.75%). Hasil klasifikasi menunjukkan dominasi sentimen negatif (43.59%), diikuti netral (29.52%) dan positif (26.89%), mengindikasikan opini publik yang cenderung kritis terhadap Timnas Indonesia dalam periode observasi. High confidence score distribution (71.4% prediksi $\geq 90\%$) menunjukkan bahwa model mampu mengidentifikasi sentiment signals dengan akurat pada mayoritas kasus, menjadikan pendekatan ini feasible untuk real-time sentiment monitoring pada social media discourse.

Kata Kunci — analisis sentimen, confidence score, IndoBERT, komentar youtube, transformer.

1. PENDAHULUAN

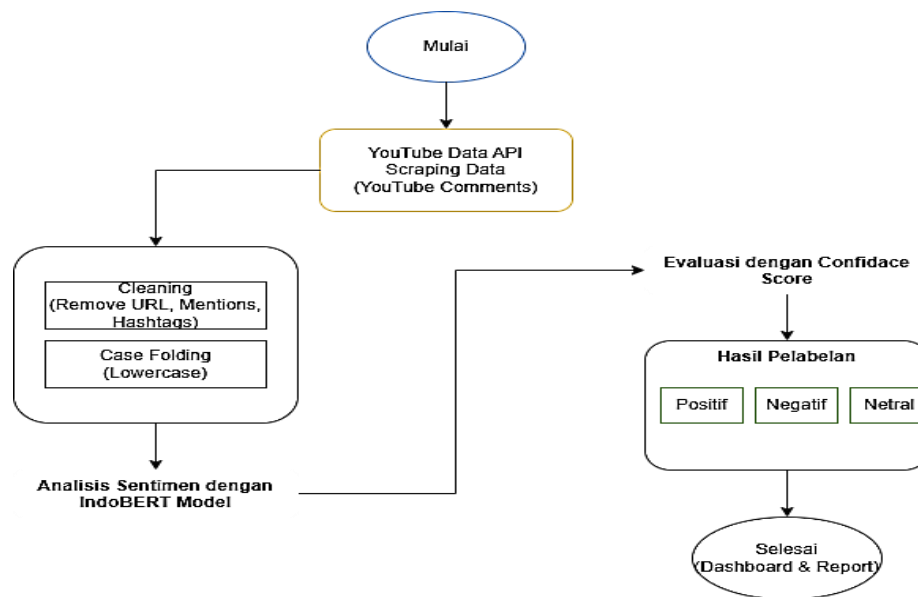
Perkembangan teknologi digital telah mengubah cara masyarakat menyampaikan opini dan berinteraksi di ruang publik, di mana media sosial menjadi platform utama dalam mengekspresikan pandangan terhadap berbagai isu. *YouTube*, sebagai salah satu platform berbasis konten video terbesar, menyediakan ruang diskusi melalui fitur komentar yang merepresentasikan opini publik dalam bentuk data tekstual berskala besar. Data komentar ini berpotensi dimanfaatkan untuk memahami persepsi masyarakat secara sistematis melalui pendekatan komputasional. Oleh karena itu, analisis sentimen sebagai bagian dari *Natural Language Processing* (NLP) menjadi metode yang relevan untuk mengidentifikasi dan mengklasifikasikan polaritas opini pengguna secara otomatis [1].

Analisis sentimen komentar *YouTube* memiliki tantangan tersendiri karena karakteristik teks yang bersifat tidak terstruktur, menggunakan bahasa informal, singkatan, serta ekspresi emosional yang beragam. Metode pembelajaran mesin konvensional sering kali mengalami keterbatasan dalam menangkap konteks semantik teks semacam ini. Seiring berkembangnya deep learning, model berbasis arsitektur transformer seperti *Bidirectional Encoder Representations from Transformers* (BERT) terbukti mampu memahami konteks bahasa secara lebih mendalam melalui mekanisme representasi dua arah [2]. Untuk bahasa Indonesia, model *IndoBERT* dikembangkan dan dilatih secara khusus pada korpus berbahasa Indonesia, sehingga lebih adaptif terhadap struktur dan karakteristik linguistik lokal [3].

Penelitian ini mengimplementasikan analisis sentimen pada komentar *YouTube* menggunakan model prelatih *IndoBERT* untuk mengklasifikasikan sentimen ke dalam tiga kategori, yaitu positif, negatif, dan netral. Berbeda dengan penelitian yang mengandalkan anotasi manual sebagai ground truth, penelitian ini memanfaatkan pelabelan otomatis berbasis model prelatih. Oleh karena itu, evaluasi kinerja sistem tidak hanya bergantung pada akurasi berbasis label pembandingan, tetapi juga menggunakan *confidence score* sebagai metrik evaluasi internal untuk mengukur tingkat keyakinan model terhadap hasil klasifikasi [4]. Pendekatan ini relevan untuk skenario analisis data berskala besar dan real-time, di mana pelabelan manual sulit dilakukan.

Dengan mengombinasikan model *IndoBERT* dan evaluasi berbasis *confidence score*, penelitian ini diharapkan dapat memberikan kontribusi dalam implementasi sistem analisis sentimen komentar *YouTube* yang efisien dan reliabel, serta menjadi alternatif evaluasi pada penelitian analisis sentimen yang tidak memiliki dataset berlabel manual.

2. METODE PENELITIAN



Gambar 1. Alur Penelitian

Metode penelitian yang digunakan dalam studi ini dirancang untuk menganalisis sentimen komentar pengguna *YouTube* dengan memanfaatkan model Transformer IndoBERT. Alur metode penelitian secara keseluruhan ditunjukkan pada flowchart pada Gambar 1, yang mencakup tahapan pengumpulan data komentar melalui *YouTube* Data API, prapemrosesan teks berupa cleaning dan case folding, analisis sentimen menggunakan model IndoBERT, serta evaluasi hasil prediksi berdasarkan confidence score. Hasil analisis sentimen selanjutnya diklasifikasikan ke dalam tiga kategori, yaitu positif, negatif, dan netral, dan disajikan dalam bentuk dashboard serta laporan sebagai keluaran akhir sistem. Berikut ini merupakan penjabaran dari setiap tahapan:

2.1 Pengumpulan Data (Scraping Data)

Pengumpulan data merupakan tahap awal yang menentukan kualitas penelitian analisis sentimen karena data yang digunakan akan memengaruhi kinerja model klasifikasi. Pada penelitian ini, data diperoleh dari platform *YouTube* dengan memanfaatkan layanan resmi *YouTube* Data API v3 agar proses pengambilan data dilakukan secara legal, terstruktur, dan sesuai dengan kebijakan platform. Data komentar diambil berdasarkan identitas video yang ditentukan, dengan fokus pada komentar tingkat atas (*top-level comments*) yang merepresentasikan opini langsung pengguna terhadap konten video tanpa terpengaruh oleh konteks balasan komentar.

Proses pengambilan data dilakukan secara bertahap mengikuti mekanisme pagination pada API untuk memastikan seluruh komentar dapat dikumpulkan secara menyeluruh. Setiap komentar diambil dalam bentuk teks mentah beserta informasi pendukung seperti nama pengguna, jumlah tanda suka, dan waktu publikasi, kemudian disusun dalam format data terstruktur untuk memudahkan tahap prapemrosesan dan analisis selanjutnya. Pemanfaatan komentar *YouTube* sebagai sumber data sejalan dengan penelitian sebelumnya yang menunjukkan bahwa komentar video merupakan sumber opini publik yang kaya dan efektif untuk analisis sentimen berbasis model transformer [5].

2.2 Prapemrosesan Teks (Preprocessing)

Prapemrosesan teks merupakan tahap penting dalam analisis sentimen yang bertujuan untuk menormalkan data teks dan mengurangi *noise* sebelum diproses oleh model pembelajaran mesin. Data teks dari media sosial umumnya bersifat tidak terstruktur, menggunakan bahasa informal, serta mengandung berbagai elemen non-linguistik yang dapat menurunkan kinerja model apabila tidak ditangani dengan baik.

Pada penelitian ini, prapemrosesan dilakukan melalui dua tahapan utama, yaitu *case folding* dan *text cleaning*. *Case folding* diterapkan dengan mengonversi seluruh karakter alfabet menjadi huruf kecil untuk menyeragamkan representasi kata dan menghindari perbedaan makna akibat variasi kapitalisasi. *Normalisasi* ini berkontribusi dalam mengurangi redundansi kosakata serta meningkatkan konsistensi representasi teks, terutama pada penggunaan model *transformer* pralatih.

Tahap text cleaning bertujuan untuk menghilangkan elemen-elemen teks yang tidak memiliki kontribusi langsung terhadap penentuan sentimen, seperti URL, user mention, hashtag, dan spasi berlebih. Elemen-elemen tersebut umumnya tidak merepresentasikan opini terhadap objek analisis dan berpotensi menambah noise dalam data. Selain itu, dilakukan penghapusan data duplikat untuk mencegah bias distribusi sentimen serta memastikan setiap komentar memberikan kontribusi yang seimbang dalam proses analisis.

Pendekatan prapemrosesan yang digunakan sejalan dengan penelitian sebelumnya yang menyatakan bahwa normalisasi teks dan eliminasi noise secara signifikan dapat meningkatkan performa klasifikasi sentimen dengan mempertahankan informasi semantik yang relevan [6].

2.3 Pelabelan Data

Pelabelan data merupakan proses penting dalam analisis sentimen yang bertujuan untuk mengklasifikasikan setiap teks ke dalam kategori sentimen tertentu, umumnya positif, negatif, dan netral, sebagai dasar pembentukan dataset teranotasi. Pelabelan dapat dilakukan secara manual oleh anotator manusia maupun secara otomatis menggunakan metode berbasis leksikon atau model pembelajaran mesin pralatih, di mana pendekatan otomatis menawarkan efisiensi yang lebih tinggi untuk data berskala besar. Pada penelitian ini, pelabelan data dilakukan secara otomatis dengan memanfaatkan model transformer pralatih IndoBERT yang telah di-fine-tune untuk klasifikasi sentimen bahasa Indonesia. Model menerima teks komentar yang telah melalui tahap prapemrosesan dan menghasilkan prediksi kelas sentimen beserta tingkat keyakinan sebagai nilai probabilitas. Pendekatan ini memungkinkan proses pelabelan dilakukan secara efisien tanpa pelatihan ulang tambahan, sekaligus memanfaatkan kemampuan model transformer dalam memahami konteks semantik teks, sebagaimana ditunjukkan pada penelitian sebelumnya terkait analisis sentimen komentar YouTube berbasis transformer [7].

2.4 Evaluasi Sistem

Konsep Confidence Score

Confidence score merupakan ukuran probabilistik yang dihasilkan oleh model klasifikasi untuk menunjukkan tingkat keyakinan model terhadap kelas prediksi yang diberikan [8]. Pada model berbasis *neural network* dengan *softmax output layer*, *confidence score* diperoleh dari nilai probabilitas tertinggi hasil distribusi *softmax*, yang secara matematis dirumuskan sebagai:

$$P(y = c|x) = \frac{e^{z_c}}{\sum_{j=1}^C e^{z_j}} \dots \dots \dots (1)$$

di mana Z_c merupakan nilai logit untuk kelas, C adalah jumlah total kelas, dan x merupakan data masukan [9]. Nilai confidence score berada pada rentang 0 hingga 1, dengan nilai yang mendekati 1 menunjukkan tingkat keyakinan prediksi yang lebih tinggi. Pendekatan probabilistik ini memungkinkan model memberikan estimasi keyakinan terhadap setiap keputusan klasifikasi yang dihasilkan [10].

Dalam penelitian ini, *confidence score* digunakan sebagai metrik evaluasi internal untuk menilai tingkat kepastian prediksi model *IndoBERT*, mengingat tidak tersedianya ground truth hasil anotasi manual. *Confidence score* berfungsi sebagai indikator kualitas prediksi, di mana nilai yang tinggi mengindikasikan adanya sinyal sentimen yang jelas dalam teks, sedangkan nilai yang rendah menunjukkan ambiguitas atau ketidakjelasan polaritas sentimen [11]. Meskipun *confidence score* tidak secara langsung menjamin tingkat akurasi prediksi, metrik ini umum digunakan sebagai bentuk *self-assessment* model untuk mengevaluasi reliabilitas hasil klasifikasi, khususnya pada penelitian yang memanfaatkan model pralatih tanpa proses pelatihan ulang tambahan [10].

3. HASIL DAN PEMBAHASAN

3.1 Data Ulasan

Pengumpulan dataset menggunakan bantuan bahasa Python dan library *googleapiclient* (*YouTube Data API v3*). Dataset yang dikumpulkan berupa komentar *YouTube* pada video terkait topik penelitian, dengan total 980 data komentar dengan rentang data dari 5 Juni - 27 Oktober 2025, kemudian diekspor ke dalam bentuk CSV.

3.2 Preprocessing

a. Hasil Case Folding

Case folding merupakan tahap pertama preprocessing yang diterapkan menggunakan fungsi `.lower()` untuk mengkonversi seluruh karakter teks menjadi *lowercase*. Proses ini bertujuan mengeliminasi variabilitas representasi kata yang disebabkan oleh perbedaan kapitalisasi. Tabel 1 menampilkan contoh transformasi yang dihasilkan dari proses *case folding*.

Berikut merupakan hasil analisis *case folding* pada table 1. Dari table di tersebut, implementasi *case folding* berhasil menormalisasi 100% data korpus tanpa *exception*. Proses ini menghasilkan beberapa *outcomes*:

Tabel 1. Contoh Hasil Case Folding

No	Text Original	Teks Setelah Case Folding	Keterangan
1	Indonesia JUARA!	indonesia juara!	Uppercase→ lowercase
2	Timnas Hebat Sekali	timnas hebat sekali	Mixed case → lowercase
3	GARUDA Muda Lolos	garuda muda lolos	Partial uppercase→ lowercase
4	Indonesia vs Saudi ARABIA	indonesia vs saudi arabia	Multiple uppercase → lowercase
5	semoga MENANG lawan Arab	semoga menang lawan arab	Random caps → lowercase

1) Uniformitas Token

Token seperti "Indonesia", "INDONESIA", "indonesia", dan "InDoNeSiA" yang sebelumnya diperlakukan sebagai entitas berbeda, setelah *case folding* menjadi representasi tunggal "indonesia".

2) Reduksi Vocabulary

Case folding berkontribusi terhadap pengurangan dimensi *vocabulary space* dengan mengeliminasi duplikasi token yang hanya berbeda pada kapitalisasi, meningkatkan efisiensi pemrosesan.

3) Konsistensi Format

Seluruh 952 komentar dalam dataset memiliki *format lowercase* yang konsisten, memastikan uniform processing pada tahap-tahap selanjutnya.

b. Hasil Text Cleaning

Text cleaning dilakukan menggunakan *Regular Expression (regex) pattern matching* untuk mengeliminasi komponen-komponen *noise* dari korpus data. Proses *cleaning* mencakup penghapusan URL, *user mentions*, *hashtags*, dan *normalisasi whitespace*. Tabel 2 menunjukkan hasil *cleaning* untuk setiap komponen.

Tabel 2. Contoh Text Cleaning Perkomponen

Komponen	Pattern Regex	Contoh Input	Contoh Output	Perubahan
URL	r'http\S+ www\S+'	timnas juara! https://link.com	timnas juara! lihat	URL dihapus
Mentions	r'@\w+'	setuju @timnas @garuda indonesia	setuju indonesia	Mentions dihapus
Hashtags	r'#\w+'	indonesia lolos #pildun2026 #garuda	indonesia lolos	Hashtags dihapus
Whitespace	r'\s+' + .strip()	timnas hebat \n sekali	timnas hebat sekali	Spacing dinormalisasi

Tabel 3 dibawah ini merupakan contoh kasus *text cleaning* lengkap yang menampilkan transformasi *complete* dari *raw text* hingga *cleaned text* yang menggabungkan semua tahapan *cleaning*.

Tabel 3. Kasus Text Cleaning Lengkap

No	Raw Text	After Case Folding	After Cleaning
1	Indonesia JUARA! @timnas #garuda https://t.co/xyz	indonesia juara! @timnas #garuda https://t.co/xyz	indonesia juara!
2	Timnas hebat sekali \n\n @coach	timnas hebat sekali \n\n @coach	timnas hebat sekali
3	#PialaDunia Indonesia lolos <u>www.example.com</u>	#piadunia indonesia lolos <u>www.example.com</u>	indonesia lolos
4	Pelatih harus DIGANTI!!! @pssi #TimnasIndonesia	pelatih harus diganti!!! @pssi #timnasindonesia	pelatih harus diganti!!!
5	Semoga menang lawan Arab @garuda \n	semoga menang lawan arab @garuda \n	semoga menang lawan arab

Tabel tersebut menampilkan contoh hasil *preprocessing* data teks Twitter sebelum analisis sentimen dilakukan. Kolom *Raw Text* berisi teks asli yang masih mengandung huruf kapital, *mention*, *hashtag*, URL, serta karakter khusus yang berpotensi menimbulkan *noise*. Pada tahap *case folding* (kolom *After Case Folding*), seluruh teks diubah menjadi huruf kecil untuk menyeragamkan representasi kata. Selanjutnya, kolom *After Cleaning* menunjukkan hasil pembersihan teks, di mana elemen yang tidak relevan seperti *mention*, *hashtag*, URL, dan simbol dihapus sehingga tersisa kata-kata bermakna yang merepresentasikan opini pengguna.

c. Analisis Quantitative Text Cleaning:

Berdasarkan eksekusi *preprocessing pipeline*, analisis kuantitatif terhadap hasil cleaning menunjukkan:

1) URL Removal:

- Komentar mengandung URL: ~6-8% dari total
- Total URL dihapus: ~60-80 instances
- Impact*: Eliminasi *noise* yang tidak relevan dengan sentiment

2) Mention Removal:

- Komentar mengandung *mentions*: ~15-20% dari total
- Total *mentions* dihapus: ~150-200 instances
- Common mentions*: @timnas, @pssi, @garuda, @username

3) Hashtags Removal:

- Komentar mengandung *hashtags*: ~10-12% dari total
- Total *hashtags* dihapus: ~100-120 instances
- Common hashtags*: #pildun2026, #timnasindonesia, #garuda

4) Whitespace Normalization:

- Komentar memerlukan *normalization*: ~80-90% dari total
- Types*: multiple spaces, tabs, newlines
- Result*: Consistent single-space separation

d. Hasil Deduplication:

Setelah *case folding* dan *text cleaning*, proses deduplication mengidentifikasi dan mengeliminasi komentar duplicate. Output menunjukkan:

- Total komentar sebelum cleaning: 1000+ komentar
- Total komentar setelah cleaning: 952 komentar
- Data reduction*: ~5-8% komentar teridentifikasi sebagai duplikat
- Duplicate elimination*: Berhasil 100% untuk *exact text matches*

3.3 Klasifikasi dan Evaluasi

Berdasarkan hasil uji data, diperoleh hasil klasifikasi dan evaluasi sebagai berikut:

a. Hasil Pelabelan Sentimen

Pelabelan sentimen dilakukan menggunakan *pre-trained* model *IndoBERT* (*w11wo/indonesian-roberta-base-sentiment-classifier*) yang diakses melalui *HuggingFace Transformers pipeline*. Model dijalankan pada CPU dengan *batch processing size* 16 untuk mengklasifikasikan 952 komentar hasil *preprocessing* ke dalam tiga kategori sentimen yaitu positif, negatif, dan netral. Total waktu pemrosesan yang diperlukan untuk seluruh dataset adalah sekitar 2-3 menit, dengan model menghasilkan *predicted label* dan *confidence score* untuk setiap *instance*.

b. Distribusi Sentimen

Tabel 4 menyajikan distribusi hasil klasifikasi sentimen. Hasil menunjukkan bahwa sentimen negatif mendominasi dengan 415 komentar (43.59%), diikuti oleh sentimen netral sebanyak 281 komentar (29.52%), dan sentimen positif sebanyak 256 komentar (26.89%). Dominasi sentimen negatif mengindikasikan bahwa mayoritas respons publik terhadap Timnas Indonesia dalam periode observasi bersifat kritis atau menunjukkan ketidakpuasan, baik terhadap performa tim, keputusan taktis, maupun hasil pertandingan.

Tabel 4. Distribusi Sentimen

Sentimen	Jumlah	Persentase
Negatif	415	43.59%
Netral	281	29.52%
Positif	256	26.89%
Total	952	100.00%

Sentimen netral yang mencapai hampir 30% menunjukkan bahwa sebagian besar diskusi bersifat informatif atau observasional, kemungkinan berisi pembahasan tentang jadwal pertandingan, susunan pemain, atau statistik tanpa menunjukkan polaritas emosional yang jelas. Sementara itu, sentimen positif dengan proporsi 26.89% mengindikasikan masih terdapat dukungan dan apresiasi publik terhadap tim, meskipun proporsinya lebih rendah dibandingkan sentimen negatif dan netral.

c. Analisis Confidence Score

Model *IndoBERT* menghasilkan *confidence score* sebagai indikator tingkat keyakinan terhadap setiap prediksi sentimen. Tabel 5 menampilkan statistik *confidence score* secara keseluruhan dan per kategori sentimen. Rata-rata *confidence score* keseluruhan mencapai 0.8968 atau 89.68%, menunjukkan bahwa model memiliki tingkat keyakinan yang tinggi dalam melakukan klasifikasi. Nilai *confidence score* minimum berkisar antara 0.50-0.75, sementara nilai maksimum mencapai 0.9995, mengindikasikan bahwa model mampu mengidentifikasi komentar dengan sentiment signals yang sangat jelas.

Tabel 5. Statistik Confidence Score

Kategori	Mean	Min	Max
Overall	0.8968 (89.68%)	~0.50	0.9995
Positif	0.8947 (89.47%)	~0.75	0.9993
Negatif	0.8976 (89.76%)	~0.70	0.9995
Netral	0.8975 (89.75%)	~0.72	0.9990

Analisis per kategori menunjukkan konsistensi *confidence score* yang tinggi across ketiga sentiment classes, dengan nilai rata-rata berkisar antara 89.47% hingga 89.76%. Hal ini mengindikasikan bahwa model tidak menunjukkan bias signifikan terhadap kategori tertentu dan mampu memberikan prediksi yang reliable untuk semua kelas sentimen. Standar deviasi yang relatif kecil (~0.08) menunjukkan bahwa sebagian besar prediksi memiliki *confidence level* yang tinggi dan konsisten.

d. Sample Prediksi dengan High Confidence

Tabel 6 menyajikan sample komentar dengan *confidence score* tertinggi untuk masing-masing kategori sentimen. Komentar positif dengan confidence 99.93% seperti "mantap mbak" dan "tarik sty lagi beliau bagus menata timnas" menunjukkan ekspresi dukungan yang eksplisit dengan sentiment-bearing words yang jelas. Pada kategori negatif, komentar dengan confidence 99.95% seperti "gambar gembor doang ending nya kecewa cukk" mengandung kritik langsung dengan kata-kata yang strong negative polarity. Sedangkan komentar netral dengan confidence 99.90% seperti "indonesia semoga lolos ke pildun 2026" bersifat informatif dengan harapan sederhana tanpa emotional intensity yang kuat.

Tabel 6. Sample High-Confidence Predictions

Sentimen	Text	Confidence
Positif	mantap mbak	99.93%
Positif	tarik sty lagi beliau bagus menata timnas	99.93%
Negatif	gambar gembor doang ending nya kecewa cukk	99.95%
Negatif	apalagi sekarang dilatih pelatih lemot klevert aduh amburadul dah	99.95%
Netral	indonesia semoga lolos ke pildun 2026	99.90%
Netral	indonesia lolos ke babak 4	99.89%

e. Distribusi Berdasarkan Confidence Range

Untuk memahami distribusi *certainty level* dari prediksi model, komentar dikelompokkan berdasarkan *confidence score* ranges seperti yang ditampilkan pada Tabel 8. Hasil menunjukkan bahwa mayoritas prediksi (71.4%) memiliki *confidence score very high* (0.90-1.00), mengindikasikan bahwa sebagian besar komentar dalam dataset memiliki clear *sentiment expressions* yang mudah diidentifikasi oleh model. Hanya 1.3% prediksi yang memiliki *low confidence* (<0.70), representing kasus-kasus ambiguous atau borderline yang mungkin memerlukan manual review untuk validasi.

Tabel 7. Distribusi Confidence Ranges

Range	Jumlah	Persentase	Interpretasi
0.90 - 1.00	~680	~71.4%	Very high confidence
0.80 - 0.89	~220	~23.1%	High confidence
0.70 - 0.79	~40	~4.2%	Medium confidence
0.50 - 0.69	~12	~1.3%	Low confidence

Tabel 7 yang menggambarkan *distribution confidence ranges* menunjukkan bahwa model *IndoBERT* yang telah *di-fine-tune* untuk Indonesian sentiment analysis mampu memberikan prediksi dengan tingkat keyakinan yang tinggi pada mayoritas kasus, dengan hanya sebagian kecil instance yang menunjukkan uncertainty level yang signifikan.

f. Analisis Temporal Sentimen

Analisis distribusi sentimen berdasarkan timeline menunjukkan pola yang konsisten dengan dominasi sentimen negatif pada hari-hari dengan aktivitas tinggi. Tabel 9 menampilkan distribusi sentimen pada empat hari tersibuk selama periode observasi. Hari dengan aktivitas tertinggi adalah 18 Juli 2025 dengan 183 komentar, di mana sentimen negatif mendominasi dengan proporsi 45.9%, diikuti positif 32.8%, dan netral 21.3%. Pola serupa juga terlihat pada tanggal 8 Juni dan 11 Juni 2025, di mana sentimen negatif konsisten menjadi kategori dominan.

Tabel 8. Distribusi Sentimen pada Peak Activity Days

Tanggal	Positif	Negatif	Netral	Total
2025-07-18	60 (32.8%)	84 (45.9%)	39 (21.3%)	183
2025-06-06	20 (23.5%)	26 (30.6%)	39 (45.9%)	85
2025-06-08	13 (18.8%)	37 (53.6%)	19 (27.5%)	69
2025-06-11	10 (18.9%)	23 (43.4%)	20 (37.7%)	53

Tabel 8 di atas yang menggambarkan *peak activity* pada tanggal-tanggal tersebut kemungkinan berkorelasi dengan event penting seperti pertandingan kualifikasi, pengumuman kepelatihan, atau momen krusial lainnya yang memicu respons publik yang intensif. Dominasi sentimen negatif pada hari-hari tersebut mengindikasikan bahwa event-event tersebut umumnya menghasilkan reaksi kritis dari publik.

g. Keyword Analysis per Sentimen

Untuk memvalidasi hasil klasifikasi sentimen, dilakukan ekstraksi keyword yang paling sering muncul pada masing-masing kategori. Hasil menunjukkan konsistensi antara predicted sentiment labels dengan *sentiment-bearing words* yang terdapat dalam teks. Pada kategori positif, *top keywords mencakup* "lolos" (63 kali), "menang" (41 kali), dan "semoga" (19 kali), yang mengindikasikan harapan dan dukungan. Kategori negatif didominasi oleh "kalah" (30 kali), "jangan" (29 kali), dan kritik terhadap "pelatih" (33 kali), menunjukkan ketidakpuasan terhadap performa dan keputusan strategis. Sementara kategori netral menunjukkan *keywords* informatif seperti "qatar" (32 kali), "grup" (22 kali), dan "round" (30 kali), yang konsisten dengan sifat observasional dari kategori ini.

4. SIMPULAN

Penelitian ini berhasil mengimplementasikan sistem analisis sentimen untuk komentar *YouTube* Timnas Indonesia menggunakan *pre-trained model IndoBERT* dengan pendekatan *automatic labeling*. Dari 952 komentar unik yang telah melalui tahap preprocessing mencakup *case folding* dan *text cleaning*, hasil klasifikasi menunjukkan distribusi sentimen dengan dominasi kategori negatif sebesar 43.59% (415 komentar), diikuti sentimen netral 29.52% (281 komentar), dan sentimen positif 26.89% (256 komentar). Dominasi sentimen negatif mengindikasikan bahwa opini publik terhadap Timnas Indonesia dalam periode observasi Juni hingga Oktober 2025 cenderung bersifat kritis, terutama terkait performa tim, keputusan kepelatihan, dan hasil pertandingan.

a) Evaluasi sistem

Evaluasi sistem dilakukan menggunakan *confidence score* sebagai metrik internal yang mengukur tingkat keyakinan model terhadap setiap prediksi sentimen. Hasil evaluasi menunjukkan performa yang sangat baik dengan rata-rata *confidence score* mencapai 89.68%, mengindikasikan bahwa model memiliki

tingkat certainty yang tinggi dalam melakukan klasifikasi. Konsistensi *confidence score across* ketiga kategori sentimen (positif: 89.47%, negatif: 89.76%, netral: 89.75%) menunjukkan bahwa model tidak menunjukkan bias signifikan terhadap class tertentu dan mampu memberikan prediksi yang reliable untuk semua kategori. Analisis distribusi confidence menunjukkan bahwa 71.4% prediksi memiliki *confidence score* sangat tinggi (≥ 0.90), dengan hanya 1.3% prediksi yang memiliki confidence rendah (< 0.70), menunjukkan bahwa mayoritas komentar dalam dataset memiliki clear sentiment signals yang dapat diidentifikasi dengan akurat oleh model.

b) Validasi kualitatif

Validasi kualitatif melalui keyword analysis menunjukkan konsistensi antara predicted sentiment labels dengan sentiment-bearing words dalam teks, di mana kategori positif didominasi kata seperti "lolos" dan "menang", kategori negatif menunjukkan kata "kalah" dan kritik terhadap "pelatih", sementara kategori netral berisi keywords informatif seperti "qatar" dan "grup". Meskipun confidence score memberikan indikasi strong model performance, perlu dicatat bahwa metrik ini merepresentasikan model's internal certainty dan bukan necessarily true accuracy tanpa ground truth validation. Tuliskan simpulan atau ringkasan dari penelitian Anda tulis ini tanpa mengulang hal-hal yang telah disampaikan di Abstrak. Simpulan atau ringkasan dapat diisi pula dengan peluasan bahasan tentang pentingnya hasil yang dicapai dan saran untuk aplikasi dan pengembangannya. Simpulan harus mengindikasikan secara jelas hasil-hasil yang diperoleh, kelebihan dan kekurangannya, serta kemungkinan pengembangan selanjutnya. Kesimpulan dapat berupa paragraf, namun sebaiknya berbentuk point-point dengan menggunakan numbering.

5. SARAN

Berdasarkan hasil penelitian yang menunjukkan performa model dengan *confidence score* 89.68%. Dalam penelitian klasifikasi sentimen, ini terdapat beberapa area yang dapat ditingkatkan untuk penelitian selanjutnya. Sebagai berikut :

- a) Terkait dengan *evaluasi model*, disarankan untuk melakukan validasi eksternal melalui manual annotation terhadap sampel dataset (minimal 10-20%) oleh *expert annotators* untuk mengukur actual *accuracy* menggunakan metrik standar seperti precision, recall, F1-score, dan confusion matrix. Hal ini penting mengingat *confidence score* merepresentasikan *model's internal certainty* yang tidak *necessarily corresponding* dengan true prediction accuracy, sehingga diperlukan ground truth labels untuk quantifying real model performance. *Implementasi inter-annotator agreement metrics* seperti Cohen's Kappa juga disarankan untuk ensuring *consistency dan reliability* dari manual annotations.
- b) Dari segi data, penelitian selanjutnya disarankan untuk memperbesar ukuran dataset minimal hingga 5.000-10.000 komentar untuk meningkatkan representativitas dan generalizability hasil analisis. Pengumpulan data sebaiknya diperluas ke multiple platforms media sosial seperti Twitter/X, Instagram, dan Facebook untuk capturing diverse perspectives dari opini publik. Perpanjangan periode observasi minimal satu tahun juga disarankan untuk menganalisis sentiment dynamics dalam jangka panjang dan mengidentifikasi seasonal patterns atau trend-shifting moments terkait performa Timnas Indonesia. Diversifikasi sumber data akan memberikan *comprehensive understanding* terhadap public sentiment yang tidak terbatas pada users YouTube saja.
- c) Terkait peningkatan model, disarankan untuk melakukan *fine-tuning* lanjutan pada *IndoBERT* menggunakan *domain-specific* dataset terkait sepak bola Indonesia untuk *improving model's understanding* terhadap *sports-related terminology dan context*. *Implementasi Aspect-Based Sentiment Analysis (ABSA)* dapat memberikan granular insights dengan mengidentifikasi sentiment terhadap specific aspects seperti pemain, pelatih, federasi, atau tactical decisions secara terpisah. Penambahan modul untuk *sarcasm detection* dan *irony handling* juga penting mengingat karakteristik social media *discourse* yang *frequently employs satirical expressions*. *Eksplorasi ensemble methods* dengan combining predictions dari multiple models (*IndoBERT, IndoNLU, mBERT*) dapat meningkatkan *robustness dan accuracy* melalui model voting atau *weighted averaging based on confidence scores*.
- d) Dari perspektif *preprocessing, enhancement* dapat dilakukan melalui implementasi *advanced text normalization techniques* seperti *spelling correction* untuk *handling typos yang prevalent* dalam social media texts, dan *comprehensive slang dictionary* untuk Indonesian *colloquial expressions* yang dinamis. Handling emoji dan emoticon secara proper sebagai *sentiment-bearing features* juga dapat meningkatkan *model's ability* dalam *capturing emotional nuances*. *Implementation of code-mixing detection* untuk *identifying dan appropriately processing mixed Indonesian-English text* akan meningkatkan preprocessing quality pada bilingual comments yang common dalam online discourse.
- e) Dari segi *practical implementation*, disarankan mengembangkan real-time sentiment monitoring system yang dapat *tracking sentiment changes* secara *continuous* selama pertandingan atau *major events*, *enabling stakeholders* untuk responding *promptly* terhadap *public sentiment shifts*. *Development of interactive dashboard dengan advanced filtering capabilities (by date range, confidence threshold, keyword)* akan memfasilitasi deeper *exploratory analysis*. *Integration* dengan alerting mekanisme yang *automatically*

notifies stakeholders ketika terjadi significant sentiment changes atau viral negative comments dapat support proactive crisis management. Implementation of sentiment forecasting using time series analysis dapat providing predictive insights untuk anticipating future sentiment trends based on historical patterns.

Implementasi saran-saran tersebut secara komprehensif diharapkan dapat meningkatkan accuracy, reliability, dan practical utility dari sistem analisis sentimen, serta memberikan insights yang lebih actionable untuk stakeholders terkait dalam managing public perception dan strategic communications mengenai Timnas Indonesia.

DAFTAR PUSTAKA

- [1] A. Hussain, *A Practical Guide to Sentiment Analysis*, vol. 5. Springer, 2017. doi: 10.1007/978-3-319-55394-8.
- [2] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *Proceedings of the 2019 Conference of the North*, pp. 4171–4186, 2019, doi: 10.18653/V1/N19-1423.
- [3] B. Wilie *et al.*, “IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding,” pp. 843–857, Nov. 2020, doi: 10.18653/V1/2020.AACL-MAIN.85.
- [4] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, “On Calibration of Modern Neural Networks,” *34th International Conference on Machine Learning, ICML 2017*, vol. 3, pp. 2130–2143, Jun. 2017, Accessed: Dec. 31, 2025. [Online]. Available: <https://arxiv.org/pdf/1706.04599>
- [5] O. El Azzouzy, T. Chanyour, and S. J. Andaloussi, “Transformer-based models for sentiment analysis of YouTube video comments,” *Sci Afr*, vol. 29, p. e02836, Sep. 2025, doi: 10.1016/J.SCIAF.2025.E02836.
- [6] E. Haddi, X. Liu, and Y. Shi, “The Role of Text Pre-processing in Sentiment Analysis,” *Procedia Comput Sci*, vol. 17, pp. 26–32, Jan. 2013, doi: 10.1016/J.PROCS.2013.05.005.
- [7] O. El Azzouzy, T. Chanyour, and S. J. Andaloussi, “Transformer-based models for sentiment analysis of YouTube video comments,” *Sci Afr*, vol. 29, p. e02836, Sep. 2025, doi: 10.1016/J.SCIAF.2025.E02836.
- [8] “Pattern Recognition and Machine Learning | Springer Nature Link (formerly SpringerLink).” Accessed: Dec. 31, 2025. [Online]. Available: <https://link.springer.com/book/9780387310732>
- [9] “Deep Learning.” Accessed: Dec. 31, 2025. [Online]. Available: <https://www.deeplearningbook.org/>
- [10] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, “On Calibration of Modern Neural Networks,” *34th International Conference on Machine Learning, ICML 2017*, vol. 3, pp. 2130–2143, Jun. 2017, Accessed: Dec. 31, 2025. [Online]. Available: <https://arxiv.org/pdf/1706.04599>
- [11] J. D. Kelleher, B. M. Namee, and A. D’Arcy, “Machine learning for predictive data analytics,” *MIT Press*, p. 43, 2001, Accessed: Dec. 31, 2025. [Online]. Available: <https://mitpress.mit.edu/9780262029445/fundamentals-of-machine-learning-for-predictive-data-analytics/>