

Analisis Sentimen Berbasis Aspek dan Deteksi Emosi dari Ulasan Timnas Indonesia di Instagram

I Made Bayu Artha Pratama¹⁾, Viktor Handrianus Pranatawijaya²⁾, Putu Bagus Adidyana Anugrah Putra³⁾

¹⁾²⁾³⁾ Teknik Informatika, Fakultas Teknik, Universitas Palangka Raya Kampus UPR Tunjung Nyaho Jl. Yos Sudarso, Palangka Raya

¹⁾ madebayu26@mhs.eng.upr.ac.id

²⁾ viktorhp@it.upr.ac.id

³⁾ putubagus@it.upr.ac.id

Abstrak

Olahraga, khususnya sepak bola ini memiliki peran penting dalam memperkuat citra nasional. dalam analisis sentimen dan emosi penggemar. terhadap Timnas Indonesia di piala Asia 2023 melalui komentar Instagram. Sebanyak 96.485 komentar dikumpulkan dan di proses menggunakan teknik *Latent Dirichlet Allocation (LDA)* untuk indentifikasi topik serta cosine similarity untuk menghubungkan aspek-aspek relevan. Emosi dilabeli menggunakan kamus *NRC lexicon*, sementara sentimen dianalisis secara manual dan dengan model Indobert, yang menunjukkan akurasi tertinggi, yaitu 97,58%.

Hasil penelitian menunjukkan dominasi sentimen positif (74,9%) dengan emosi utama trust dan joy, meski terdapat kritik signifikan terkait hasil pertandingan. Emosi negatif seperti anger dan fear muncul, mencerminkan ketidakpuasan sebagai penggemar.

Meskipun mayoritas penggemar merasa puas, terdapat kelompok yang sangat tidak puas dengan kinerja tim. Temuan ini memberikan wawasan bagi pihak terkait untuk meningkatkan performa dan kualitas interaksi dengan penggemar.

Kata kunci: Sentimen, Emosi, Timnas Indonesia, Media Sosial, algoritma

Abstract

Sports, especially football, play an important role in strengthening national image. In the sentiment and emotion analysis of fans towards the Indonesian National Team at the 2023 Asian Cup through Instagram comments, a total of 96,485 comments were collected and processed using Latent Dirichlet Allocation (LDA) for topic identification and cosine similarity to connect relevant aspects. Emotions were labeled using the NRC lexicon, while sentiment was analyzed both manually and with the Indobert model, which showed the highest accuracy of 97.58%.

The results of the study showed a dominance of positive sentiment (74.9%) with the main emotions being trust and joy, although there was significant criticism regarding the match results. Negative emotions such as anger and fear emerged, reflecting fan dissatisfaction.

Although the majority of fans felt satisfied, there was a group that was very dissatisfied with the team's performance. These findings provide insights for the relevant parties to improve performance and the quality of interaction with fans.

Keywords: Sentiment, Emotion, Indonesia Nasional Team, Social Media, Algorithm

1. PENDAHULUAN

Sepak bola tidak hanya berperan dalam kebugaran fisik, tetapi juga menjadi sarana penting dalam membangun citra positif suatu bangsa di kancah internasional. Tim Nasional Indonesia (Timnas Indonesia) kembali menjadi sorotan setelah berpartisipasi dalam Piala Asia 2023, menandai kembalinya mereka ke ajang besar setelah 16 tahun absen. Momen ini memicu euforia

serta beragam tanggapan dari penggemar, terutama di Instagram, melalui akun @timnas.indonesia. Media sosial menjadi platform utama bagi penggemar untuk menyampaikan opini, baik dukungan maupun kritik, terutama setelah hasil pertandingan. Komentar tersebut mencakup kekecewaan terhadap performa pemain, kritik terhadap strategi pelatih, hingga harapan untuk pertandingan berikutnya.

Di era digital, penting bagi pelatih, manajemen tim, dan pemain untuk memahami persepsi publik terhadap kinerja Timnas Indonesia. Salah satu cara untuk melakukannya adalah dengan analisis sentimen berbasis aspek dan deteksi emosi dari ulasan di media sosial seperti Instagram. Teknologi *Machine Learning*, khususnya *Natural Language Processing (NLP)*, dapat membantu komputer memahami bahasa manusia dan memberikan respons yang sesuai[1]. Untuk memahami pandangan pengguna, digunakan teknik *Aspect Based Sentiment Analysis (ABSA)*, yang mengekstraksi aspek dan menentukan polaritas sentimen dari setiap aspek dalam sebuah kalimat [2].

Penelitian Penelitian terbaru mengeksplorasi deteksi emosi berbasis aspek dan korelasi untuk memahami opini publik [3] Deteksi Emosi Berbasis Aspek dan Korelasi pada Tweet untuk Memahami Opini Publik tentang Covid-19 Mereka Labeling data menjadi sentimen dan emosi menggunakan korelasi Pearson. Setelah Labeling, data mengalami pemrosesan ulang, diikuti oleh klasifikasi model tunggal menggunakan metode pembelajaran mesin (*Support Vector Machine*, *Random Forest*, *Naïve Bayes*) dan metode pembelajaran mendalam (*Bidirectional Encoder Representation from Transformers*). Proses klasifikasi difokuskan pada akurasi dan evaluasi skor F1. Tiga skenario digunakan untuk menentukan sentimen dan emosi: faktor berbasis aspek dan korelasi, tanpa faktor-faktor ini, dan sentimen berbasis aspek saja. Skenario yang menggabungkan kedua faktor mencapai akurasi 97%, sedangkan skenario tanpa mereka memperoleh akurasi 96%.

Penelitian Dalam penelitian ini, data dikumpulkan dari ulasan Instagram @timnas.indonesia menggunakan teknik seperti *Latent Dirichlet Allocation (LDA)* untuk mengidentifikasi topik, Cosine Similarity untuk pelabelan aspek, dan kamus NRC Lexicon untuk deteksi emosi. Model pembelajaran mesin seperti *Decision Tree*, *Gradient Boosting*, dan *IndoBERT* digunakan untuk mengevaluasi sentimen, dengan tujuan utama memberikan wawasan tentang pola sentimen penggemar dan mengidentifikasi area yang memerlukan peningkatan dalam kinerja Timnas Indonesia.

2. TINJAUAN PUSTAKA

2.1 Analisis Sentimen Berbasis Aspek

Analisis Sentimen berbasis aspek (*Aspect-based Sentiment Analysis/ABSA*) adalah pendekatan yang digunakan untuk mengidentifikasi dan mengekstrak sentimen terkait aspek-aspek spesifik dari suatu objek atau entitas. Dalam konteks ulasan, ABSA membantu evaluasi pengguna terhadap berbagai aspek produk, layanan, atau destinasi[4] ABSA bekerja dengan memeriksa sentimen terhadap berbagai aspek secara individual, sering kali menggunakan teknik seperti POS Taggers untuk mengekstraksi aspek dan opini sebelum mengklasifikasikannya [5]

2.2 Deteksi Emosi

Deteksi emosi adalah proses mengenali emosi yang diekspresikan oleh pengguna melalui teks biasanya menggunakan algoritma dan model pembelajaran mesin.[6]

2.3 Web Scrapping

Web Scraping adalah proses mengambil data dari halaman web yang ditulis dalam bahasa markup seperti HTML atau XHTML, lalu mengolahnya untuk mengekstraksi informasi yang diperlukan.) [7]

Web Scraping Sering menggunakan framework Scrapy untuk mengumpulkan dan mengekstraksi data, serta Splash untuk merender halaman berbasis JavaScript sebelum data

diekstraksi. Hasil scraping biasanya disimpan dalam database seperti MySQL untuk analisis lebih lanjut.[8].

2.4 Text Preprocessing

Tahapan preprocessing merupakan langkah penting sebelum data digunakan untuk membuat model, bertujuan untuk membersihkan dan menyiapkan data[9]

2.4.1 Data Cleansing

Data Cleansing tahap menghapus tanda baca dan angka.[10]

2.4.2 Case Folding

Case Folding mengkonversi seluruh teks ke dalam bentuk huruf kecil/low-case[10]

2.4.3 Normalisasi

Normalization atau normalisasi, Proses normalisasi bertujuan untuk mengubah kata-kata tidak baku menjadi baku yang sesuai dengan kamus Besar Bahasa Indonesia (KBBI). [11]

2.4.4 Stopword

Stopwords yaitu proses pemilihan kata-kata penting yang mempunyai arti dan tidak, sehingga kata yang tidak mempunyai arti akan dibuang, contohnya “ini”, “itu”, “yang”, “di” dan lain-lain.[12]

2.4.5 Tokenizing

Tokenizing adalah proses memecah urutan karakter menjadi beberapa bagian (kata/frasa) yang disebut token.[13]

2.4.6 Stemming

Stemming menghilangkan semua kata imbuhan baik sufiks dan prefiks. Berfungsi untuk mengurangi variasi pada kata yang memiliki kata dasar semua.[10]

2.4.7 Bag Of Word

Bag Of Words (Bow) adalah cara untuk merepresentasikan kata-kata yang unit dari seluruh teks dalam sebuah korpus, sering digunakan dalam bidang Pemrosesan Bahasa Alami (NLP) dan Pemulihan Informasi[14]

2.4.8 Topic modelling Latent Dirichlet Allocation (LDA)

Topic Modelling merupakan proses analisis data teks berdasarkan topik-topik tertentu. Ini melibatkan pengelompokan dokumen-dokumen berdasarkan kemiripannya, sering kali dengan menggunakan metode clustering[15]

2.4.9 Cosine Similarity

Cosine Similarity adalah ukuran yang digunakan untuk menilai seberapa mirip dua dokumen, tanpa memperhatikan[16]

2.4.10 NRC Lexicon

NRC Lexicon adalah kumpulan kata-kata yang dikaitkan dengan emosi[17] Lexicon ini mencakup berbagai kata dalam bahasa Inggris yang dikaitkan dengan kategori emosi atau dimensi efek seperti kegembiraan, ketakutan, valensi, dan rangsangan. Model Emosi *Plutchik* mencakup delapan emosi utama: *anticipation*, *joy*, *trust*, *fear*, *surprise*, *sadness*, *disgust*, dan *anger*, yang dianggap sebagai emosi dasar universal[18]

2.4.11 Lexicon Vader

Lexicon Vader Adalah pendekatan Leksikal yang digunakan untuk analisis sentimen teks pada media sosial. Pendekatan ini menggabungkan kosakata umum dengan ekspresi microblogging[19]

2.4.12 Confusion Matrix

Confusion matrix adalah alat yang digunakan untuk mengevaluasi kinerja algoritma pembelajaran mesin[20]

2.4.13 Pembobotan TF-IDF

Pembobotan *Term Frequency-Inverse Document Frequency (TF-IDF)* adalah algoritma yang digunakan untuk menilai seberapa sering kata-kata muncul dalam dokumen[21]

2.4.14 Oversampling SMOTE

Oversampling SMOTE ((*Synthetic Minority Over-sampling Technique*)) adalah sebuah metode yang digunakan untuk menyeimbangkan dataset yang tidak seimbang dengan cara menghasilkan sampel sintesis dari kelas minoritas daripada hanya menggandakan sampel yang sudah ada[22]

2.4.15 Desicion tree

Decision Tree adalah pendekatan yang digunakan untuk menggambarkan dan menggali informasi dari data dengan cara menggambarkan struktur pohon keputusan.[23]

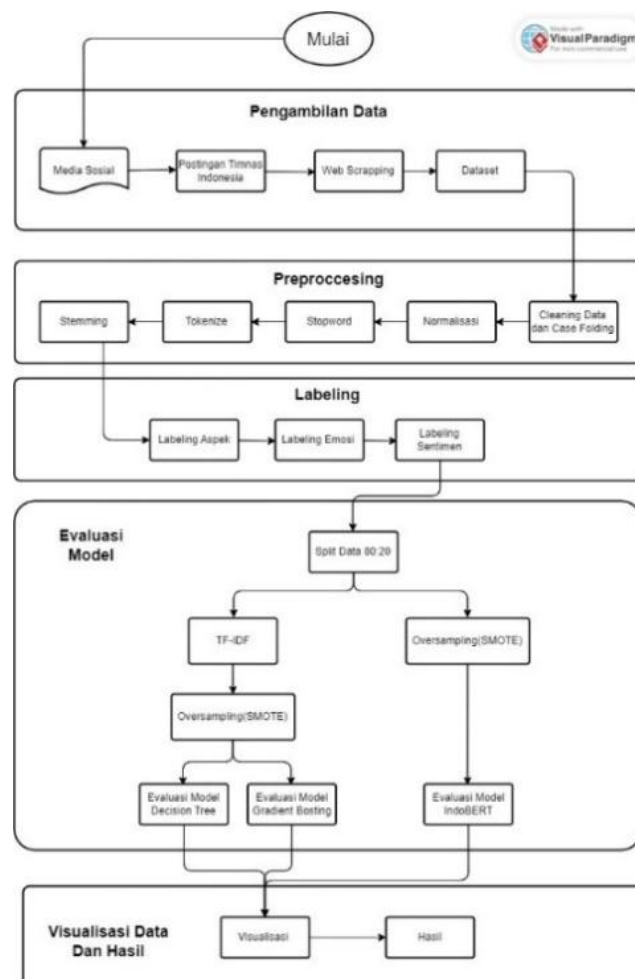
2.4.16 Gradient Boosting

Gradient Boosting adalah metode *machine learning* yang menggabungkan beberapa model prediksi lemah(biasanya pohon keputusan) untuk membentuk model yang lebih kuat[24]

2.4.17 IndoBERT

Model ini adalah model pralatih yang didasarkan pada arsitektur BERT ((*Bidirectional Encoder Representations from Transformers*)). Model ini secara khusus dilatih menggunakan dataset bahasa indonesia yang bernama Indo4B, yang mencakup berbagai sumber teks dalam bahasa indonesia, seperti berita online, media sosial, wikipedia, artikel online, subtitle, dan dataset paralel[25]

3. METODE PENELITIAN



Gambar 1. Alur Metode Penelitian

Pada gambar di atas, terdapat visualisasi alur metode penelitian yang digunakan penulis dalam studi ini. Alur metode penelitian tersebut disusun untuk mengilustrasikan hubungan antara metode-metode dan variabel yang digunakan dalam penelitian. Terdapat lima tahapan dalam alur metode tersebut, yaitu pengambilan data, *preprocessing*, *Labeling*, visualisasi data dan hasil Pada bagian berikutnya, akan dijelaskan lebih lanjut mengenai metode-metode yang tercantum dalam gambar sebelumnya.

4. PEMBAHASAN

4.1.1 Scraping Data

Data ulasan dikumpulkan dari Instagram menggunakan ekstensi Super Comment Exporter Proses pengambilan data dilakukan dari tanggal 14 hingga 28 Januari 2024, dengan total 98.000 ulasan yang disimpan dalam format CSV. Data ini mencakup informasi seperti nama pengguna dan komentar. Hasil scraping ini akan digunakan sebagai dataset utama dalam penelitian untuk dianalisis lebih lanjut. Contoh hasil scrapping data ulasan dapat dilihat pada Tabel 1.

Tabel 1. Hasil Scrapping Data ulasan Media Sosial *Instagram*

Username	Comment
nijitoursandtravel	Sty memang top percaya akan proses.. wajib perpanjang kontrak 5 thn ke depan.. tp klo tiba tiba gus hiddink menawarkan diri jadi pelatih gimana ya guys..
riska_fnj	Pakai AI mah gak perlu ide, kocak nih manusia satuðŸ™, bukannya support artist lokal malah pakai AI yang hobi comot karya orang. Mana fans nya mau aja dibegoin begini
afganr_	Kalo memang coach Shin Tea Yong mau mundur kita hargai saja keputusan Coach Shin,,mungkin dia udah gak tahan di Indonesia,,
ansori_ar	Tetap semangat Garuda... Lupakan kegagalan saat ini.. kalian sudah luar biasa dan dapat banyak pengalaman...
rbu_englishcourse	Pemain irak sudah bermain bersama lbh dr 10th..jd kita harus bersabar.. Ada pemain naturalisasi yg baru gabung bwberapa hari, bulan atau tahun.. Jd ada nama nya proses..... go Indonesia ðŸ†®ðŸ†©ðŸ†®ðŸ†©ðŸ†®ðŸ†©

4.1.2 Preprocessing Data

Setelah mengumpulkan dataset, dilakukan tahap preprocessing untuk membersihkan dan memproses teks dalam data. Tahapan Preprocessing dapat dilihat pada tabel 2.

Tabel 2. Hasil Preprocessing Data

No	Pemrosesan Awal	Data Mentah	Pembersihan Data
1	Data Cleansing	Tetap semangat Garuda... Lupakan kegagalan saat ini.. kalian sudah luar biasa dan dapat banyak pengalaman...	Tetap semangat Garuda Lupakan kegagalan saat ini kalian sudah luar biasa dan dapat banyak pengalaman
2	Normalisasi	Tetap semangat Garuda... Lupakan kegagalan saat ini.. kalian sudah luar biasa dan dapat banyak pengalaman...	Tetap semangat Garuda Lupakan kegagalan saat ini kalian sudah luar biasa dan dapat banyak pengalaman Fokus untuk kualifikasi piala dunia lagi Semoga kita lolos piala dunia untuk pertama kali nya
3	Stopword Removal	Tetap semangat Garuda... Lupakan kegagalan saat ini.. kalian sudah luar biasa dan dapat banyak pengalaman...	semangat Garuda Lupakan kegagalan pengalaman Fokus kualifikasi piala dunia Semoga lolos piala dunia kali
4	Tokenisasi	Tetap semangat Garuda... Lupakan kegagalan saat ini.. kalian sudah luar biasa dan dapat banyak pengalaman...	['semangat', 'Garuda', 'Lupakan', 'kegagalan', 'pengalaman', 'Fokus', 'kualifikasi', 'piala', 'dunia', 'Semoga', 'lolos', 'piala', 'dunia', 'kali']

5	Stemming Data	Tetap semangat Garuda... Lupakan kegagalan saat ini.. kalian sudah luar biasa dan dapat banyak pengalaman...	['semangat', 'garuda', 'lupa', 'gagal', 'alam', 'fokus', 'kualifikasi', 'piala', 'dunia', 'moga', 'lolos', 'piala', 'dunia', 'kali']
---	---------------	--	--

4.1.3 Labeling Aspek

Algoritma *Latent Dirichlet Allocation (LDA)* untuk analisis topik dan aspek pada teks komentar melibatkan beberapa langkah utama. Pertama, komentar dibersihkan dan diubah menjadi representasi bag-of-words. Setelah itu, model LDA diterapkan dengan berbagai jumlah topik. Model yang menghasilkan nilai coherence tertinggi dipilih sebagai model optimal. Dalam penelitian ini, model LDA dieksplorasi dengan batas maksimal lima topik, minimal dua topik, dan penambahan jumlah topik dengan langkah satu. Hasil ini kemudian digunakan untuk menilai koherensi dari setiap topik yang diidentifikasi.

Setelah mendapatkan nilai koherensi, langkah berikutnya adalah menampilkan grafik untuk melihat topik dengan nilai tertinggi, yaitu 3 topik. Kata kunci dari setiap topik diekstraksi untuk mengidentifikasi aspek dominan. *Cosine similarity* dihitung antara vektor TF-IDF dari komentar dan topik untuk menentukan relevansinya. Hasil akhirnya menunjukkan distribusi topik, membantu mengungkap tema utama dalam komentar, seperti diskusi tentang "Tim Nasional Indonesia", "Pemain", dan "Hasil Pertandingan". Detailnya dapat dilihat dalam tabel 3

Tabel 3. Sebaran Kata Tiap Jumlah Topik

Topik	Aspek	Kata per Topik
1	Hasil Pertandingan	jepang, menang, lawan, vietnam, nguyen, bang, poin, min, pas, tonton, allah, king, adam, imbang, pulang, seri, habis, gua, japan, deg, langsung, mode, alis, darah, an, hujat, minimal, bismillah, parah, irak, makassar, kuat, tuhan, argentina, tombol, asa, mati, amin, info, ku, bro, tenang, jam, merah, wkwkwk, iraq, madrid, kapai, putih, var
2	Pemain	main, hokky, banget, keren, ernando, sandy, shin, bagus, menit, kalo, tae-yong, kayak, hoki, kelas, bola, hubner, tahan, yakob, walsh, peluang, jordi, mantap, lihat, sananta, nih, kiper, hoky, gue, sayur, masuk, match, gol, bikin, latih, cepat, gacor, tim, tinggal, greget, striker, coach, tuh, motm, nando, full, turun, caraka, save, lini, coba
3	Tim Nasional Indonesia	indonesia, alhamdulillah, timnas, garuda, next, kalah, moga, kasih, malam, jantung, babak, semangat, baik, selamat, juang, terima, terimakasih, asia, tanding, gila, lu, lolos, piala, kali, bangga, ayo, point, besok, senam, aff, muda, bantai, fokus, harap, tunggu, hasil, game, apresiasi, aman, maju, the, apa, laga, optimis, tidur, kerja, doa, hati, dingin, hebat

4.1.4 Labeling Emosi

Algoritma untuk melabeli emosi pada teks ulasan menggunakan *NRC Lexicon* dalam Bahasa Indonesia mencakup delapan emosi: marah, antisipasi, jijik, takut, senang, sedih, kejutan, dan kepercayaan. Proses dimulai dengan membaca dataset, menghapus data yang hilang, dan memuat kamus *EmoLex NRC*. Setiap kata dalam ulasan diperiksa kecocokannya dengan lexicon, dan emosi yang sesuai diberi bobot berdasarkan frekuensi kata. Emosi dominan ditentukan dari skor tertinggi, sementara ulasan tanpa emosi diberi label "nan". Contoh komentar dari setiap emosi Tabel 5.

Tabel 4. Hasil Labeling Deteksi Emosi

Komentar_Clean	fear	Anger	Anticipation	Trust	Surprise	Sadness	Disgust	Joy	emotion
shin tae-yong top percaya proses wajib panjang kontrak kalo gus hiddink tawar latih guys	0	0	0.0714286	0.1428571	0.0714286	0	0	0.0714286	trust

4.1.5 Labeling Sentimen

Penelitian ini menggunakan data komentar dari file CSV yang berisi kolom '*Komentar_Clean*', '*Aspek*', dan '*Emotion*'. Langkah pertama adalah membaca data dan mengambil kolom relevan. Kamus *emotion_mapping* digunakan untuk mengaitkan emosi yang terdeteksi dengan sentimen dan tingkat kepuasan pengguna. Misalnya, emosi '*Joy*' dikategorikan sebagai sentimen positif yang menunjukkan kepuasan, sementara '*Anger*' sebagai sentimen negatif menunjukkan ketidakpuasan. Nama emosi distandarisasi dengan huruf kapital di awal, dan hasil pemetaan emosi ini ditampilkan dalam tabel yang menunjukkan sentimen dan tingkat kepuasan. berikut adalah sentimen dan satisfied pada tabel 6.

Tabel 6. Hasil Sentimen dan Satisfied

Komentar	Aspek	Emotion	Sentimen	User Satisfaction
latih butuh siangkat main naturalisasi latih shin tae-yong panjang kontrak segi main timnas indonesia bagus kalo shin tae-yong out progesnya buyar	Pemain	Trust	Positive	Satisfied

4.1.6 Split Data

Pada dataset Ulasan Timnas Indonesia Di Instagram yang telah memiliki label sentimen entitas dan sentimen. Selanjutnya dibagi menjadi data latih dan data uji. Detailnya dapat dilihat dalam Tabel 7.

Tabel 7. Split Data Model

Metode	Klasifikasi	Data Latih (80%)	Data Uji (20%)
Decision Tree, Gradient Boosting Dan BERT	Positive	161816	40454
	Negative	65264	133
	Jumlah	5389	16316

4.1.7 Evaluasi Sentimen

Model *Decision Tree* dilatih menggunakan parameter seperti `max_depth=None`, `min_samples_split=2`, `min_samples_leaf=1`, dan `max_features=None`, serta didukung teknik SMOTE untuk menyeimbangkan dataset. Hasil evaluasi kinerja menunjukkan kemampuan model ini dalam mengklasifikasikan berbagai emosi dan aspek dengan akurasi, presisi, recall, dan F1 Score yang mencerminkan performanya dalam berbagai konteks. Model *Gradient Boosting* dilatih

dengan parameter serupa dan model *BERT* yang telah disesuaikan untuk tugas ini berhasil memberikan hasil evaluasi yang komprehensif. Evaluasi model terdapat pada Tabel 7

Tabel 8. Evaluasi Sentimen

Dataset	Pelabelan	Sentimen			
		Accuracy	Precision	Recall	F1-Score
Ulasan Timnas Indonesia Di Instagram	Decision Tree	92.16%	92.37%	92.16%	92.23%
	Gradient Boosting	92.55%	92.79%	92.55%	92.63%
	IndoBERT	97.58%	97.03%	98.10%	97.56%

4.1.8 Evaluasi Emosi

Pada evaluasi emosi *Decision Tree*, *Gradient Boosting* dan *IndoBERT* dataset yang digunakan terdiri dari ulasan yang telah diberi label emosi seperti *joy*, *sadness*, *fear*, *disgust*, *anger*, *anticipation*, *trust*, dan *surprise*. Pada proses ini dihasilkan data hasil evaluasi untuk emosi dari dataset "ulasan Timnas Indonesia Di Instagram". Berikut adalah data evaluasi terdapat pada tabel 9.

Tabel 9. Evaluasi Emosi

Dataset	Pelabelan	Emosi			
		Accuracy	Precision	Recall	F1-Score
Ulasan Timnas Indonesia Di Instagram	Decision Tree	85.54%	86.16%	85.54%	85.71%
	Gradient Boosting	85.54%	86.16%	85.54%	85.71%
	IndoBERT	94.06%	94.04%	96.06%	93.99%

4.1.9 Evaluasi Aspek

Pada Evaluasi aspek menggunakan model *Decision Tree*, *Gradient Boosting* dan *IndoBERT* Pada proses ini dihasilkan data hasil evaluasi untuk emosi dari dataset "ulasan Timnas Indonesia Di Instagram". Berikut adalah data evaluasi terdapat pada tabel 10, Tabel 11 dan Tabel 12.

Tabel 10. Evaluasi Aspek *Decision Tree*

Dataset	Model	Aspek	Accuracy	Precision	Recall	F1-Score
Ulasan Timnas Indonesia Di Instagram	<i>Decision Tree</i>	Hasil Pertandingan	90.32%	90.44%	90.32%	90.35%
		Pemain	92.69%	92.83%	92.69%	92.74%
		Tim Nasional Indonesia	91.92%	91.93%	91.93%	91.93%

Tabel 11. Evaluasi Aspek *Gradient Boosting*

Dataset	Model	Aspek	Accuracy	Precision	Recall	F1-Score
Ulasan Timnas Indonesia Di Instagram	<i>Gradient Boosting</i>	Hasil Pertandingan	90.99%	91.07%	90.99%	91.02%
		Pemain	92.89%	93.04%	92.89%	92.95%

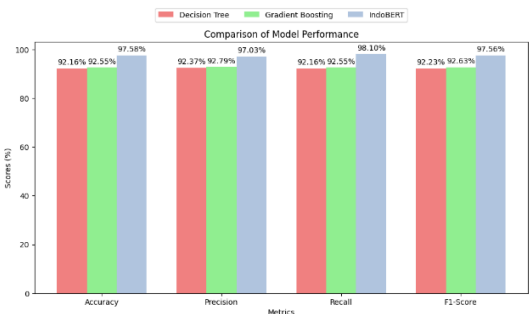
	Tim Nasional Indonesia	91.55%	91.66%	91.55%	91.56%
--	---------------------------	--------	--------	--------	--------

Tabel 12. Evaluasi Aspek IndoBERT

Dataset	Model	Aspek	Accuracy	Precision	Recall	F1-Score
Ulasan Timnas Indonesia Di Instagram	IndoBERT	Hasil Pertandingan	91.37%	93.51%	88.83%	91.11%
		Pemain	95.88%	93.60%	98.49%	95.98%
		Tim Nasional Indonesia	97.51%	96.65%	98.61%	97.61%

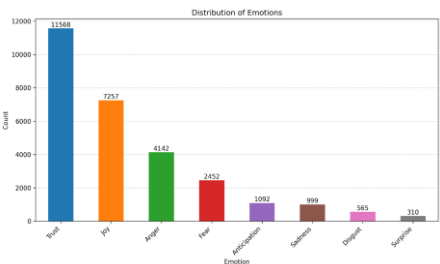
4.1.10 Analisis dan Visualisasi Data

Hasil evaluasi sentiment pada dataset Ulasan Timnas Indonesia Di Instagram dengan menggunakan model *Decision Tree*, *Gradient Boosting* dan *IndoBERT*. Hasil yang didapatkan di visualisasikan dengan menggunakan chart seperti pada gambar 2.



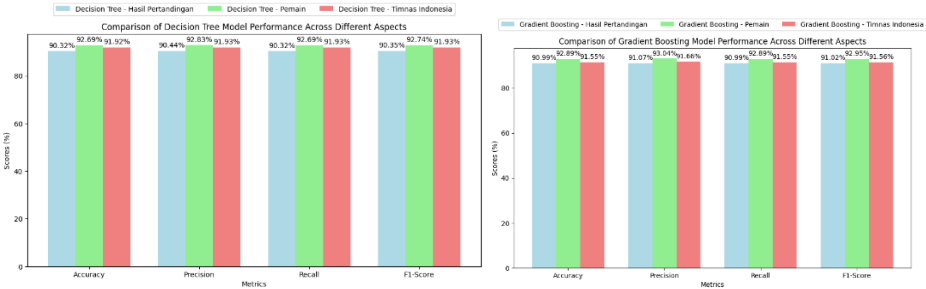
Gambar 3. Evaluasi Sentimen Model Decision Tree, Gradient Boosting dan IndoBERT

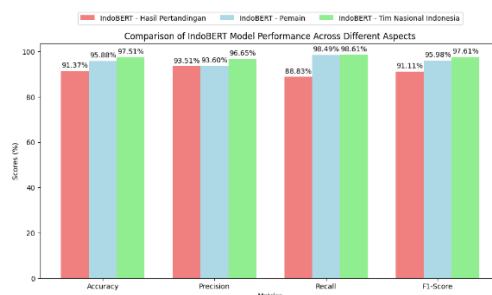
Selain itu, dilakukan juga perhitungan frekuensi kemunculan emosi dari hasil deteksi emosi awal, di mana emosi Trust(Kepercayaan) muncul paling sering, diikuti oleh *Joy*, *Anger* dan *Fear*.



Gambar 4. Distribusi Emosi

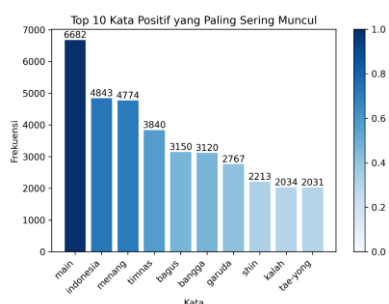
Hasil evaluasi berbagai aspek menggunakan model *Decision Tree*, *Gradient Boosting* dan *IndoBERT* divisualisasikan dalam bentuk diagram seperti pada gambar 5.



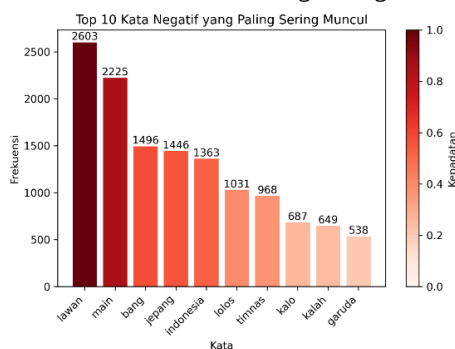


Gambar 5. Evaluasi Aspek Model *Decision Tree*, *Gradient Boosting* dan *IndoBERT*

Selanjutnya pada analisis sentimen berbasis aspek ini didapat daftar kata-kata yang sering muncul pada ulasan merepresentasikan kata positif dan negatif. Dari ulasan tersebut dianalisis 10 kata positif dan negatif yang sering muncul dapat dilihat pada gambar dibawah 6 dan 7.



Gambar 5. Kata Positif Yang Sering Muncul



Gambar 6 Kata Negatif Yang Sering Muncul

5. KESIMPULAN

Penelitian ini menganalisis 96.485 komentar dari postingan Tim Nasional Indonesia di Instagram selama AFC 2023 di Qatar, yang disaring menjadi 59.941 komentar setelah preprocessing. Fokus penelitian adalah mengidentifikasi aspek utama, emosi, dan sentimen publik terhadap tim. Aspek-aspek seperti Tim Nasional Indonesia, Pemain, dan Hasil Pertandingan diidentifikasi menggunakan topic modeling LDA, dan proses pelabelan menggunakan Cosine Similarity menghasilkan 48.641 komentar terlabel. Sebagian besar komentar terkait dengan aspek Tim Nasional Indonesia, diikuti oleh aspek Pemain dan Hasil Pertandingan.

Pelabelan emosi menggunakan *NRC Emolex* menunjukkan distribusi emosi seperti trust (11.568 komentar), joy (7.257 komentar), anger (4.142 komentar), dan fear (2.452 komentar). Emosi trust dan joy mendominasi, mencerminkan sentimen positif yang kuat, meskipun terdapat

juga emosi negatif seperti *anger* dan *fear*, yang mencerminkan kritik dan kekhawatiran penggemar. Analisis sentimen menunjukkan bahwa 20.227 komentar bernada positif, sementara 8.158 komentar negatif. Tingkat kepuasan pengguna juga dianalisis, dengan kategori terbesar adalah *Satisfied* dan *Very Satisfied*, meskipun ada yang merasa sangat tidak puas, terutama terkait dengan hasil pertandingan.

Evaluasi model dilakukan dengan menggunakan *Decision Tree*, *Gradient Boosting*, dan *IndoBERT*, di mana *IndoBERT* menunjukkan performa terbaik dengan akurasi 97,58% dalam analisis sentimen. Model *Gradient Boosting* dan *Decision Tree* juga menunjukkan hasil yang baik, namun tidak seakurat *IndoBERT*. Secara keseluruhan, penelitian ini memberikan wawasan mengenai persepsi publik terhadap Tim Nasional Indonesia selama Piala Asia 2023, termasuk emosi, sentimen, dan aspek utama yang menjadi perhatian. Temuan ini dapat digunakan sebagai dasar strategi komunikasi dan peningkatan performa tim untuk meningkatkan kepuasan penggemar.

DAFTAR PUSTAKA

- [1] A. Y. Chandra, D. Kurniawan, and R. Musa, "Perancangan Chatbot Menggunakan Dialogflow Natural Language Processing (Studi Kasus: Sistem Pemesanan pada Coffee Shop)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 1, p. 208, Jan. 2020, doi: 10.30865/mib.v4i1.1505.
- [2] S. A. P. Perdana, T. B. Aji, and R. Ferdiana, "Aspect Category Classification dengan Pendekatan Machine Learning Menggunakan Dataset Bahasa Indonesia," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi* | Vol, vol. 10, no. 3, 2021.
- [3] S. M. P. Tyas, Y. Romadhona, and D. Purwitasari, "Aspect-based Sentiment and Correlation-based Emotion Detection on Tweets for Understanding Public Opinion of Covid-19.," *Journal of Information Systems Engineering & Business Intelligence*, vol. 9, no. 1, 2023.
- [4] C. A. Bahri and L. H. Suadaa, "Aspect-based sentiment analysis in bromo tengger semeru national park indonesia based on google maps user reviews," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 17, no. 1, pp. 79–90, 2023.
- [5] S. Cahyaningtyas, D. H. Fudholi, and A. F. Hidayatullah, "Deep learning for aspect-based sentiment analysis on Indonesian hotels reviews," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 2021.
- [6] M. A. Riza and N. Charibaldi, "Emotion detection in Twitter social media using long short-term memory (LSTM) and fast text," *Int. J. Artif. Intell. Robot*, vol. 3, no. 1, pp. 15–26, 2021.
- [7] D. D. A. Yani, H. S. Pratiwi, and H. Muhandi, "Implementasi web scraping untuk pengambilan data pada situs marketplace," *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, vol. 7, no. 4, pp. 257–262, 2019.
- [8] I. N. Husada, E. H. Fernando, H. Sagala, A. E. Budiman, and H. Toba, "Ekstraksi dan Analisis Produk di Marketplace Secara Otomatis dengan Memanfaatkan Teknologi Web Crawling," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 5, no. 3, 2019.
- [9] A. Hermawan, I. Jowensen, and J. Junaedi, "Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine," *Jurnal Sains dan Teknologi*, vol. 12, no. 1, pp. 129–137, 2023.
- [10] S. A. Azzahra and A. Wibowo, "Analisis sentimen multi-aspek berbasis konversi ikon emosi dengan algoritme naïve bayes untuk ulasan wisata kuliner pada web tripadvisor," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 4, pp. 737–744, 2020.
- [11] V. K. S. Que, A. Iriani, and H. D. Purnomo, "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization," *Jurnal*

-
- Nasional Teknik Elektro Dan Teknologi Informasi*. <https://doi.org/10.22146/jnteti.v9i2>, vol. 102, 2020.
- [12] N. L. P. M. Putu and A. Z. Amrullah, "Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, pp. 123–131, 2021.
- [13] M. Lestandy, A. Abdurrahim, and L. Syafa'ah, "Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes," *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, vol. 5, no. 4, pp. 802–808, 2021.
- [14] M. P. Geetha and D. K. Renuka, "Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model," *International Journal of Intelligent Networks*, vol. 2, pp. 64–69, 2021.
- [15] W. A. N. Sari and H. D. Purnomo, "Topic Modeling Using the Latent Dirichlet Allocation Method on Wikipedia Pandemic Covid-19 Data in Indonesia," *Jurnal Teknik Informatika (Jutif)*, vol. 3, no. 5, pp. 1223–1230, 2022.
- [16] J. M. Alyza, F. S. Utomo, Y. Purwati, B. A. Kusuma, and M. S. Azmi, "MUSIC RECOMMENDATION SYSTEM BASED ON COSINE SIMILARITY AND SUPERVISED GENRE CLASSIFICATION," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 9, no. 1, pp. 77–80, 2023.
- [17] S. M. Mohammad, "Practical and ethical considerations in the effective use of emotion and sentiment lexicons," *arXiv preprint arXiv:2011.03492*, 2020.
- [18] S. Imron, E. I. Setiawan, J. Santoso, and M. H. Purnomo, "Aspect Based Sentiment Analysis Marketplace Product Reviews Using BERT, LSTM, and CNN," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 3, pp. 586–591, 2023.
- [19] Y. Asri, W. N. Suliyanti, D. Kuswardani, and M. Fajri, "Analisis Sentimen Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile: Analisis Sentimen.," 2022.
- [20] M. C. Wijaya, "Automatic Short Answer Grading System in Indonesian Language Using BERT Machine Learning," *Revue d'Intelligence Artificielle*, vol. 35, no. 6, pp. 503–509, Dec. 2021, doi: 10.18280/ria.350609.
- [21] S. Rizkia, E. B. Setiawan, and D. Puspadari, "Analisis Sentimen Kepuasan Pelanggan Terhadap Internet Provider Indihome di Twitter Menggunakan Metode Decision Tree dan Pembobotan TF-IDF," *eProceedings of Engineering*, vol. 6, no. 2, 2019.
- [22] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.
- [23] P. A. Harrison *et al.*, "Selecting methods for ecosystem service assessment: A decision tree approach," *Ecosyst Serv*, vol. 29, pp. 481–498, 2018.
- [24] W. M. Baihaqi, I. R. Yunita, A. S. T. Damayanti, and L. Akhaerunnisa, "Analysis of Classification Algorithm Performance on User Review Sentiment of the Muamalat DIN Application: Analisis Performa Algoritma Klasifikasi pada Sentimen Ulasan Pengguna terhadap Aplikasi Muamalat DIN," *CogITO Smart Journal*, vol. 9, no. 2, pp. 241–251, 2023.
- [25] H. Jayadianti, W. Kaswidjanti, A. T. Utomo, S. Saifullah, F. A. Dwiyanto, and R. Drezewski, "Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN," *ILKOM Jurnal Ilmiah*, vol. 14, no. 3, pp. 348–354, Dec. 2022, doi: 10.33096/ilkom.v14i3.1505.348-354.
-