

ANALISIS SENTIMEN KOMENTAR PADA KONTEN KUALIFIKASI PIALA DUNIA TIMNAS INDONESIA DI MEDIA SOSIAL TIKTOK MENGGUNAKAN ALGORITMA *DECISION TREE* DAN METODE *N-GRAM*

Ramadhanu, Taufik Hidayat, Doni Prastyo

Teknik Informatika, Universitas Islam Syekh Yusuf

Jl. Maulana Yusuf No.10, RT.001/RW.003, Babakan, Kec. Tangerang, Kota Tangerang, Banten 15118

Rhamadhanu2019@gmail.com

ABSTRAK

Perkembangan media sosial seperti TikTok telah menjadi platform utama bagi masyarakat untuk mengekspresikan opini, terutama pada topik populer seperti performa Tim Nasional (Timnas) Sepak Bola Indonesia. Tingginya antusiasme publik selama Kualifikasi Piala Dunia 2026 menghasilkan ribuan komentar yang tidak terstruktur, sehingga sulit untuk mengukur sentimen publik secara keseluruhan. Penelitian ini bertujuan untuk menganalisis dan mengklasifikasikan sentimen (positif, netral, dan negatif) dari komentar pengguna di akun TikTok resmi Timnas Indonesia (@timnasindonesia) serta menguji efektivitas algoritma Decision Tree dengan metode N-Gram. Proses penelitian meliputi pengumpulan 8419 data komentar menggunakan teknik crawling, preprocessing data teks, pelabelan sentimen menggunakan pendekatan berbasis lexicon, dan klasifikasi menggunakan algoritma Decision Tree. Hasil penelitian menunjukkan bahwa model klasifikasi mencapai tingkat akurasi sebesar 82%. Dari total data, sentimen negatif menjadi yang paling dominan dengan 4221 komentar, diikuti sentimen netral 812 komentar dan positif 630 komentar. Hasil visualisasi data melalui diagram batang dan WordCloud juga mengonfirmasi dominasi sentimen negatif ini, di mana kata-kata bernada kekecewaan seperti "kalah" dan "air mata" sering muncul. Hal ini mengindikasikan adanya ketidakpuasan dan kritik yang signifikan dari publik terhadap performa Timnas Indonesia selama periode kualifikasi tersebut.

Kata kunci : Analisis Sentimen, Decision Tree, N-Gram, Media Sosial, Timnas Indonesia, TikTok.

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah perubahan yang signifikan terhadap pola interaksi dan cara manusia berbagi informasi. Salah satu implikasi dari kemajuan ini adalah lahirnya berbagai platform media sosial yang memungkinkan pengguna untuk mengekspresikan opini, gagasan, dan perasaan mereka secara bebas di TikTok [1].

Sebagai salah satu platform media sosial yang berbasis video singkat, telah menjadi fenomena global [2] Dengan fitur-fitur interaktif yang memungkinkan pengguna berbagi konten kreatif, TikTok telah memengaruhi berbagai aspek kehidupan, termasuk hiburan, pendidikan, hingga diskusi sosial. Di Indonesia, TikTok menjadi platform dengan pengguna terbanyak, mencapai 112 juta pada tahun 2023 [3].

Salah satu topik yang selalu menarik perhatian publik Indonesia adalah sepak bola, khususnya performa tim nasional Indonesia Dalam ajang kualifikasi Piala Dunia, antusiasme masyarakat terlihat dari banyaknya diskusi yang terjadi di media sosial. Kualifikasi Piala Dunia adalah proses kompetisi yang diadakan FIFA untuk menentukan tim-tim yang berhak berlaga di putaran final Piala Dunia. Ajang ini tidak hanya menjadi simbol kebanggaan nasional, tetapi juga kesempatan bagi tim nasional untuk menunjukkan kemampuan mereka di tingkat dunia [4].

Dibalik tingginya antusiasme ini, terdapat tantangan dalam memahami opini publik secara keseluruhan dari komentar di TikTok. Ribuan komentar tersebar di berbagai konten video, sering kali menciptakan "noise" berupa data tidak terstruktur

yang menyulitkan upaya untuk merangkum pandangan umum terhadap tim nasional. Komentar-komentar tersebut umumnya bersifat singkat, tidak beraturan, singkatan, emoji.

Meskipun antusiasme masyarakat terhadap Timnas Indonesia sangat tinggi, tim ini masih menghadapi berbagai kendala dalam perjalanannya menuju Piala Dunia 2026. Beberapa permasalahan yang sering menjadi sorotan publik meliputi inkonsistensi performa tim, lemahnya lini pertahanan, kurangnya kreativitas di lini tengah, serta efektivitas strategi yang diterapkan oleh pelatih. persiapan tim yang kurang optimal, dan keputusan kontroversial dari PSSI juga menjadi perdebatan di kalangan suporter [5].

Tujuan dari penelitian ini adalah untuk menerapkan algoritma *Decision Tree* dan metode *N-Gram* dalam mengklasifikasikan sentimen (positif, negatif, dan netral) pada komentar pengguna TikTok yang berkaitan dengan performa Timnas Indonesia dalam ajang Kualifikasi Piala Dunia 2026. Selain itu, penelitian ini juga bertujuan untuk mengidentifikasi sentimen dominan yang diekspresikan oleh publik serta mengukur tingkat akurasi model klasifikasi yang diusulkan.

Untuk mencapai tujuan tersebut, penelitian ini akan dilakukan melalui beberapa tahapan utama. Dimulai dari pengumpulan data komentar dari akun TikTok resmi Timnas Indonesia, dilanjutkan dengan tahap *preprocessing* data untuk membersihkan dan menstandarisasi teks. Selanjutnya, dilakukan pelabelan sentimen secara otomatis menggunakan

pendekatan berbasis leksikon. Data yang telah bersih kemudian akan digunakan untuk melatih dan menguji model klasifikasi *Decision Tree* dengan ekstraksi fitur *N-Gram*. Terakhir, kinerja model akan dievaluasi menggunakan *confusion matrix* dan hasilnya divisualisasikan untuk interpretasi yang lebih mudah.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen, atau dikenal juga sebagai *opinion mining*, adalah proses mengidentifikasi dan mengkategorikan opini yang diekspresikan dalam sebuah teks untuk menentukan apakah sikap penulis terhadap suatu topik bersifat positif, negatif, atau netral. Dalam konteks media sosial, analisis sentimen menjadi alat yang sangat berguna untuk memahami persepsi publik secara otomatis terhadap berbagai isu, termasuk performa tim olahraga [6].

2.2. Text Preprocessing

Text preprocessing adalah tahapan krusial untuk membersihkan dan mempersiapkan data sebelum dianalisis. Proses ini bertujuan untuk menstandarisasi teks agar dapat diolah oleh model *machine learning*. Tahapan utamanya meliputi *emoji to teks* (konversi emoji to teks), *cleaning* (menghapus karakter tidak relevan), *case folding* (menyeragamkan huruf), *tokenizing* (memecah teks menjadi kata), *stopword removal* (menghapus kata umum), dan *stemming* (mengubah kata ke bentuk dasar) [7].

2.3. N-Gram

N-Gram adalah metode untuk merepresentasikan teks dengan membaginya menjadi urutan N kata yang berdekatan. Model ini digunakan sebagai teknik ekstraksi fitur untuk menangkap konteks lokal dalam kalimat. Jenis yang paling umum adalah *unigram* (N=1), *bigram* (N=2), dan *trigram* (N=3). Dalam penelitian ini, *unigram* digunakan untuk merepresentasikan setiap kata sebagai fitur independen, yang merupakan pendekatan dasar yang efektif untuk tugas klasifikasi teks [8] [9].

$$\text{Jumlah } N - \text{Gram} = (T - N + 1) \quad (1)$$

Di mana:

T = Jumlah kata dalam teks

N = panjang n-gram yang digunakan 1 untuk unigram, 2 untuk bigram, 3 untuk trigram.

2.4. Decision Tree

Decision Tree adalah salah satu algoritma supervised learning yang populer untuk tugas klasifikasi. Algoritma ini bekerja dengan membangun model prediksi dalam bentuk struktur pohon. Setiap internal node pada pohon merepresentasikan sebuah "tes" terhadap suatu atribut (fitur kata), setiap cabang merepresentasikan hasil dari tes tersebut, dan setiap leaf node (daun) merepresentasikan label kelas (sentimen) [10] [11].

$$\text{Entropy}(S) = \sum_{k=0}^n - p_i \cdot \log_2 p_i \quad (2)$$

Keterangan:

S = Himpunan kasus

n = Jumlah partisi S

p_i = proporsi S_i terhadap S

$$\text{Gain}(S, A) = \text{Entropy}(S) = \sum_{i=1}^n \frac{|S_i|}{S} * p_i \cdot \log_2 p_i \quad (3)$$

Keterangan:

S = Himpunan kasus

A = fitur

N = jumlah partisi atribut A

$|S_i|$ = proporsi terhadap S

$|S|$ = jumlah kasus dalam S

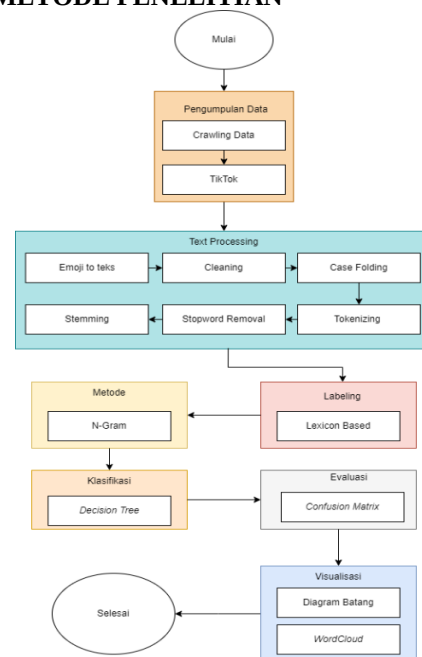
2.5. Penelitian Terdahulu

Penelitian sebelumnya telah mengeksplorasi analisis sentimen terkait sepak bola Indonesia di berbagai platform. Reda et al. [12]. menganalisis sentimen terhadap Timnas U-23 di media sosial X menggunakan metode *Naïve Bayes* dan memperoleh akurasi 73,94%. Serupa dengan itu, Prajamukti et al. [13].

Juga menggunakan *Naïve Bayes* di Twitter untuk menganalisis performa Timnas senior, U-23, dan U-19, dengan hasil akurasi 83%. Juniardi dan Sugianto [14] menerapkan *Naïve Bayes* untuk menganalisis sentimen pada Piala Dunia U-17 di Twitter dengan akurasi 78,34%.

Penelitian lain oleh Malik et al. [15]. menggunakan algoritma *Support Vector Machine* (SVM) untuk menganalisis komentar di platform YouTube terkait Piala Asia U-23 dengan akurasi 65,41%. Terakhir, Oktaviani et al. [16]. juga menggunakan *Naïve Bayes* untuk mengkaji sentimen pengguna Twitter terhadap Piala Dunia 2022 secara menunjukkan bahwa 770 di antaranya memiliki sentimen positif.

3. METODE PENELITIAN



Gambar 1. Alur Penelitian

Alur penelitian yang digunakan dalam studi ini disajikan pada Gambar 1, yang dirancang oleh peneliti untuk mempermudah pelaksanaan penelitian.

3.1. Text Preprocessing

Data mentah yang diperoleh mengandung banyak *noise* sehingga perlu melalui tahap pra-pemrosesan teks. Tahapan yang dilakukan meliputi:

- Emoji to Text:** Mengonversi simbol emoji menjadi representasi teksnya untuk menangkap makna emosional [17].
- Cleaning:** Menghapus karakter yang tidak relevan seperti tanda baca, baris kosong, data duplikat, angka, dan simbol khusus [18].
- Case Folding:** Mengubah seluruh huruf menjadi huruf kecil untuk standarisasi data [19].
- Tokenizing:** Memecah kalimat menjadi unit kata tunggal (*token*) [20].
- Stopword Removal:** Menghapus kata-kata umum yang tidak memiliki kontribusi signifikan terhadap analisis (misalnya "di", "dan", "yang") menggunakan pustaka Sastrawi [21].
- Stemming:** Mengubah kata berimbuhan menjadi kata dasarnya untuk mereduksi variasi kata [22].

3.2. Labelling

Setelah data melalui tahap pra-pemrosesan, dilakukan pelabelan sentimen untuk setiap komentar. Proses ini menggunakan metode berbasis leksikon (*Lexicon-Based*) dengan kamus yang memuat 10.251 kata berbobot. Setiap komentar diberi label sebagai positif, negatif, atau netral berdasarkan total skor sentimen dari kata-kata yang terkandung di dalamnya.

3.3. N-Gram

Untuk mengubah data teks menjadi format yang dapat diproses oleh model *machine learning*, dilakukan tahap ekstraksi fitur. Pada penelitian ini, metode *N-Gram*, khususnya *Unigram*, digunakan untuk merepresentasikan setiap kata unik dalam dataset sebagai sebuah fitur. Proses ini mengubah setiap komentar menjadi sebuah vektor numerik yang merepresentasikan frekuensi kemunculan kata.

3.4. Klasifikasi

Tahap klasifikasi bertujuan untuk membangun model prediksi. Dataset yang telah diubah menjadi vektor numerik dibagi menjadi dua bagian: 80% sebagai data latih (*training data*) dan 20% sebagai data uji (*testing data*). Selanjutnya, model klasifikasi dibangun menggunakan algoritma *Decision Tree*. Model ini dilatih menggunakan data latih untuk mempelajari pola-pola yang membedakan setiap kelas sentimen, yang kemudian digunakan untuk memprediksi label pada data uji.

3.5. Evaluasi

Kinerja model dievaluasi menggunakan *Confusion Matrix* untuk menghitung nilai metrik *accuracy*, *precision*, *recall*, dan *F1-score* [23].

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (6)$$

$$F1-Score = \frac{(2 \times Recall \times Precision)}{Recall + Precision} \times 100\% \quad (7)$$

3.6. Visualisasi

Untuk menyajikan hasil analisis secara lebih intuitif, digunakan dua teknik visualisasi. Pertama, diagram batang digunakan untuk menampilkan distribusi dan perbandingan jumlah komentar pada setiap kategori sentimen (positif, netral, dan negatif). Kedua, *WordCloud* digunakan untuk mengilustrasikan frekuensi kata-kata yang paling sering muncul dalam setiap kategori sentimen, di mana ukuran kata merepresentasikan tingkat kemunculannya.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Data komentar dikumpulkan dari akun *TikTok* resmi Timnas Indonesia (@timnasindonesia) selama periode Kualifikasi Piala Dunia 2026 ronde ketiga. Proses *crawling* data dilakukan menggunakan platform *Apify* yang diintegrasikan dengan *Google Colab*. Sebanyak 8.419 komentar berhasil dikumpulkan dari video konten terkait enam pertandingan Timnas. Adapun keenam pertandingan Timnas Indonesia yang dijadikan sumber pengambilan data adalah sebagai berikut:

Tabel 1. Enam Pertandingan Indonesia

1	5 September 2024	Arab Saudi vs Indonesia
2	10 September 2024	Indonesia vs Australia
3	10 Oktober 2024	Bahrain vs Indonesia
4	15 Oktober 2024	China vs Indonesia
5	14 November 2024	Indonesia vs Jepang
6	19 November 2024	Indonesia vs Arab Saudi

Tabel 2. Dataset Komentar Tiktok

User	createTimeISO	text
antromorphis	2024-09-05T20:11:08.000Z	Bang Paes Man Of The Match 🏆

4.2. Text Preprocessing

Data mentah yang diperoleh dari proses *crawling* umumnya tidak terstruktur dan mengandung banyak *noise* seperti penggunaan bahasa tidak baku, singkatan, dan simbol. Oleh karena itu, dilakukan serangkaian tahap *preprocessing* untuk membersihkan dan menstandarisasi data teks, sehingga meningkatkan kualitas input untuk model klasifikasi.

Tabel 3. Hasil Text Preprocessing

	Sebelum	Sesudah
Emoji to text	Terima Kasih Garuda Muda ID 🇮🇩 🦅 ❤️	Terima Kasih Garuda Muda : bendera_indonesia: :lengan_berotot_warna_kulit_cerah: :hati_merah:
Cleaning	Terima Kasih Garuda Muda : bendera_indonesia: :lengan_berotot_warna_kulit_cerah: :hati_merah:	Terima Kasih Garuda Muda bendera indonesia lengan berotot warna kulit cerah hati merah
Case Folding	Terima Kasih Garuda Muda bendera indonesia lengan berotot warna kulit cerah hati merah	terima kasih garuda muda bendera indonesia lengan berotot warna kulit cerah hati merah
Tokenizing	terima kasih garuda muda bendera indonesia lengan berotot warna kulit cerah hati merah	['terima', 'kasih', 'garuda', 'muda', 'bendera', 'indonesia', 'lengan', 'berotot', 'warna', 'kulit', 'cerah', 'hati', 'merah']
Stopword Removal	['terima', 'kasih', 'garuda', 'muda', 'bendera', 'indonesia', 'lengan', 'otot', 'warna', 'kulit', 'cerah', 'hati', 'merah']	['terima', 'kasih', 'garuda', 'muda', 'bendera', 'indonesia', 'lengan', 'berotot', 'warna', 'kulit', 'cerah', 'hati', 'merah']
Stemming	['terima', 'kasih', 'garuda', 'muda', 'bendera', 'indonesia', 'lengan', 'berotot', 'warna', 'kulit', 'cerah', 'hati', 'merah']	['terima', 'kasih', 'garuda', 'muda', 'bendera', 'indonesia', 'lengan', 'otot', 'warna', 'kulit', 'cerah', 'hati', 'merah']

4.3. Labelling Data

Pelabelan sentimen dilakukan menggunakan pendekatan berbasis kamus (*Lexicon-Based*). Setiap komentar yang telah dibersihkan diberi skor numerik berdasarkan kata-kata yang terkandung di dalamnya. Kamus *lexicon* yang digunakan berisi daftar kata positif dan negatif beserta bobotnya. Skor total dari sebuah komentar menentukan label sentimennya berdasarkan aturan: Positif jika skor > 0, Negatif jika skor < 0, dan Netral jika skor = 0.

Tabel 4. Hasil Labelling

Komentar	Skor	Kategori
good luck timnas angkat tangan angkat tangan	17	positif
tombol naturalisasi haland	0	netral
kurang stiker aja tumpul	-5	negatif
good	5	positif

Hasil dari proses labeling dapat diketahui komentar pengguna *tiktok* lebih banyak memberikan komentar sentimen negatif sebanyak 4221 dan netral sentimen sebanyak 812 dan positif sentimen sebanyak 630 komentar.

4.4. N-Gram

Untuk mempersiapkan data agar dapat diolah oleh algoritma *machine learning*, dilakukan ekstraksi fitur menggunakan metode *N-Gram*. Dalam penelitian ini, digunakan model *unigram* (N=1), yang merepresentasikan setiap komentar sebagai vektor fitur berdasarkan frekuensi kemunculan kata tunggal.

Tabel 5. Hasil N-Gram

Kata	Jumlah
wajah	3212
api	1455
hati	1421
senyum	1168
mata	947
menang	715
indonesia	684
gembira	665
air	571

4.5. Klasifikasi

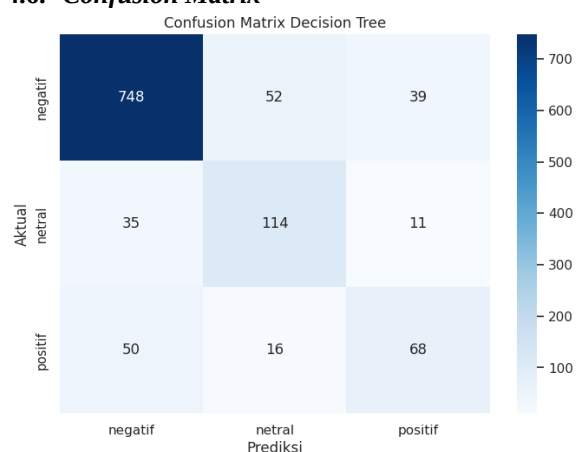
Untuk melakukan klasifikasi sentimen. Dataset yang telah dilabeli dan direpresentasikan dalam bentuk vektor fitur dibagi menjadi 80% data latih dan 20% data uji.

Tabel 6. Hasil Klasifikasi

Kelas	Precision	Recall	F1-score
Negatif	90%	89%	89%
Netral	63%	71%	67%
Positif	58%	51%	54%
Akurasi	82%		

Pada Tabel 5 terlihat bahwa model klasifikasi yang digunakan memberikan performa yang cukup baik pada kelas *negatif* dengan nilai precision, recall, dan f1-score masing-masing sebesar 90%, 89%, dan 89%. Untuk kelas *netral*, model memperoleh precision sebesar 63%, recall sebesar 71%, dan f1-score sebesar 67%. Sementara itu, performa model pada kelas *positif* masih lebih rendah dibandingkan kelas lainnya, dengan precision sebesar 58%, recall sebesar 51%, dan f1-score sebesar 54%. Secara keseluruhan, model klasifikasi memperoleh nilai akurasi sebesar 82%, yang menunjukkan bahwa model cukup baik dalam mengklasifikasikan sentimen terkait pertandingan timnas indonesia.

4.6. Confusion Matrix



Gambar 2. Confusion Matrix

Performa model klasifikasi dievaluasi menggunakan *Confusion Matrix*. Matriks ini memberikan gambaran rinci mengenai Prediksi yang dihasilkan diklasifikasikan ke dalam kategori prediksi benar (True Positive dan True Negative) maupun prediksi salah (False Positive dan False Negative).

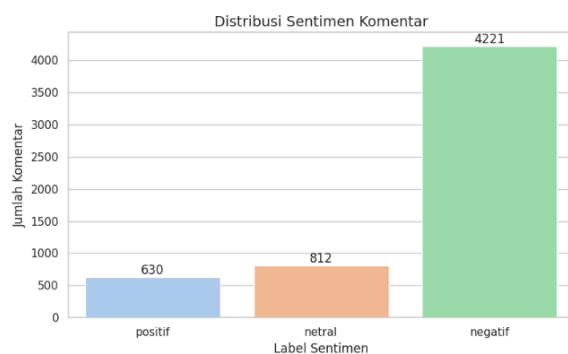
Berdasarkan dari Gambar 2 hasil *confusion matrix* menunjukkan klasifikasi model *Decision Tree* terhadap tiga kelas sentimen: negatif, netral, dan positif. Model mampu mengklasifikasikan komentar negatif dengan sangat baik (748 dari 839), cukup baik untuk komentar netral (114 dari 160), namun kurang akurat pada komentar positif, di mana hanya 68 dari 134 yang terklasifikasi dengan benar.

4.7. Visualisasi

Pada tahap di lakukan Visualisasi dengan Diagram batang dan Wordcloud.

4.8. Diagram Batang

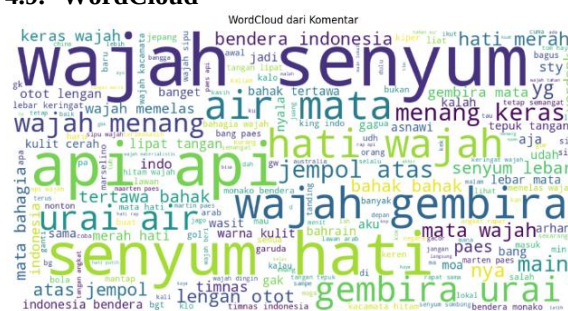
Diagram batang untuk menampilkan jumlah komentar berdasarkan klasifikasi sentimen, yaitu negatif, netral, dan positif.



Gambar 3. Distribusi Sentimen Komentar

Gambar 3 menunjukkan distribusi jumlah komentar berdasarkan label sentimen yang terbagi menjadi tiga kategori, yaitu positif, netral, dan negatif. Dari total data yang dianalisis, komentar dengan sentimen negatif mendominasi sebanyak 4.221 komentar, disusul oleh komentar netral sebanyak 812 komentar, dan komentar positif sebanyak 630 komentar.

4.9. WordCloud



Gambar 4. WordCloud Dataset

WordCloud dimanfaatkan untuk memvisualisasikan frekuensi kemunculan kata-kata yang paling sering muncul dalam komentar.

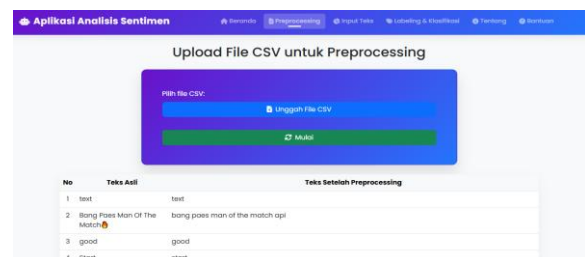
Visualisasi *WordCloud* pada Gambar 4 menunjukkan kata-kata yang paling sering muncul dalam komentar pengguna terkait konten kualifikasi Piala Dunia Timnas Indonesia di *TikTok*. Beberapa kata yang dominan antara lain “wajah”, “senyum”, “hati”, “air mata”, dan “api”. Ukuran kata yang lebih besar menandakan frekuensi kemunculan yang lebih tinggi. Hal ini mengindikasikan bahwa komentar-komentar banyak mengandung ekspresi emosi, baik positif seperti “senyum” dan “gembira”, maupun emosi mendalam seperti “air mata” dan “api”.

4.10. UI (User Interface) Aplikasi

Aplikasi sederhana dengan antarmuka pengguna (*User Interface*) yang intuitif. Aplikasi ini memfasilitasi pengguna untuk melakukan analisis sentimen, mulai dari pemrosesan data mentah hingga visualisasi hasil akhir. Berikut adalah penjelasan untuk setiap tampilan antarmuka aplikasi yang telah dikembangkan.

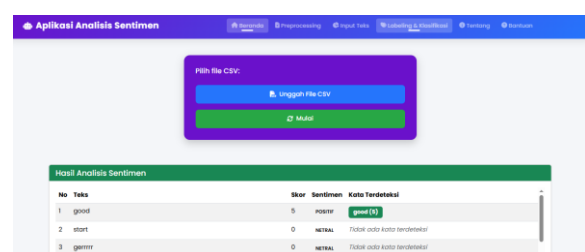


Gambar 5. Halaman Utama



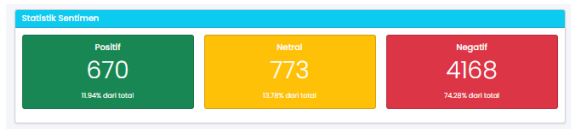
Gambar 6. Halaman *Preprocessing*

Pengguna dapat mengunggah *file* .CSV berisi data komentar mentah. Sistem kemudian secara otomatis menjalankan seluruh tahapan *preprocessing* dan menampilkan hasilnya dalam bentuk tabel perbandingan antara teks asli dan teks bersih.



Gambar 7. Halaman Labeling

Pada Gambar 7 setiap komentar disajikan dalam tabel bersama dengan skor sentimen yang dihitung menggunakan metode berbasis *lexicon*, label klasifikasi ("Positif", "Netral", atau "Negatif"), dan kata-kata spesifik yang terdeteksi dan mempengaruhi skor tersebut. Tampilan ini memungkinkan pengguna untuk memverifikasi hasil analisis pada tingkat individual komentar.



Gambar 8. Halaman Hasil Labeling

Untuk memberikan gambaran umum dari keseluruhan data, aplikasi menyajikan ringkasan statistik seperti yang ditunjukkan pada Gambar 8. Tampilan ini menyajikan jumlah total dan persentase untuk setiap kategori sentimen (Positif, Netral, dan Negatif) dalam format kartu visual yang mudah dibaca.



Gambar 8. Halaman Visualisasi Diagram batang

Pada gambar 8 Hasil analisis disajikan dalam bentuk visual. Diagram batang digunakan untuk membandingkan jumlah dari setiap sentimen.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan analisis yang telah dilakukan, dapat disimpulkan bahwa penerapan algoritma Decision Tree dengan metode ekstraksi fitur N-Gram (unigram) mampu mengklasifikasikan sentimen komentar pengguna TikTok terhadap Timnas Indonesia dengan tingkat akurasi keseluruhan sebesar 82%. Hasil ini menunjukkan bahwa model yang diusulkan cukup efektif untuk memahami respons publik di platform media sosial. Temuan utama dari penelitian ini adalah dominasi sentimen negatif yang mencapai 4.221 dari total 8.419 komentar. Hal ini mengindikasikan adanya persepsi publik yang cenderung kritis dan tidak puas terhadap performa Timnas Indonesia selama periode Kualifikasi Piala Dunia 2026 yang dianalisis. Meskipun demikian, model menunjukkan performa yang bervariasi antar kelas, dengan F1-Score tertinggi pada kelas negatif (89%) dan terendah pada kelas positif (54%), yang menyoroti adanya tantangan dalam mengklasifikasikan data yang tidak seimbang. Untuk penelitian selanjutnya, disarankan untuk melakukan perbandingan dengan model deep learning seperti

LSTM dan BERT yang lebih andal dalam memahami konteks kalimat yang kompleks, serta memperluas cakupan data dari platform media sosial lain untuk mendapatkan gambaran yang lebih komprehensif. Selain itu, pengembangan metode khusus untuk mendeteksi sarkasme dan bahasa informal juga dapat meningkatkan akurasi analisis di masa depan.

DAFTAR PUSTAKA

- [1] F. Fieryando and B. Kristianto, "Analisis Sentimen Terhadap TikTok Shop Dengan K-Nearest Neighbor, Decision Tree, dan Naive Bayes," *Jurnal Buana Informatika*, vol. 15, no. 01, pp. 21–29, Apr. 2024, doi: 10.24002/jbi.v15i1.8205.
- [2] J. Feng, "Review of Research on E-commerce Short Videos," *Frontiers in Humanities and Social Sciences*, vol. 4, no. Vol. 4 No. 4 (2024), p. 2024, Apr. 2024, doi: 10.54691/at122k44.
- [3] R. H. Pandjaitan, "The Social Media Marketing Mix Trends in Indonesia for 2024: Communication Perspective," *Jurnal Komunikasi Ikatan Sarjana Komunikasi Indonesia*, vol. 9, no. 1, pp. 251–269, Jun. 2024, doi: 10.25008/jkiski.v9i1.1005.
- [4] F. Hakim and Y. Astuti, "Analyzing Indonesian Football Sentiment Towards PSSI Performance Using Support Vector Machines," *Journal of Intelligent Software Systems*, vol. 3, no. 1, pp. 26–30, Jul. 2024, doi: 10.26798/jiss.v3i1.1330.
- [5] M. Try Andika, A. Habibi, and M. Ferari Bagaskoro, "Pengaruh Kepemimpinan Transformasional terhadap Perkembangan Sepak Bola Indonesia: Studi pada PSSI," *Fusion Multidisciplinary Journal Website*, vol. 2, no. 1, pp. 18–25, Jan. 2025, [Online]. Available: <https://yasyahikamatzu.com/index.php/FSN>
- [6] D. Rahmawati and T. Hidayat, "Analisis Kepuasan Mahasiswa Terhadap Sistem Informasi Akademik (Sina) Universitas Islam Syekh Yusuf Tangerang Berdasarkan Instrumen End User Computing Satisfaction (Eucs)," *Jurnal Teknik Informatika Unis*, vol. 9, no. 1, pp. 2252–5351, 2021.
- [7] D. Darwis, E. Shintya Pratiwi, A. Ferico, and O. Pasaribu, "PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA," *Jurnal Ilmiah Edutic*, vol. 7, no. 1, Nov. 2020.
- [8] H. Li, D. Cai, J. Xu, and T. Watanabe, "N-gram Is Back: Residual Learning of Neural Text Generation with n-gram Language Model," Oct. 2022, [Online]. Available: <http://arxiv.org/abs/2210.14431>
- [9] P. Pratama, E. Indarbansyah, and N. Rochmawati, "Penerapan N-Gram menggunakan Algoritma Random Forest dan Naïve Bayes Classifier pada Analisis Sentimen Kebijakan PPKM 2021," *Journal of Informatics and*

- Computer Science*, vol. 02, no. 4, pp. 235–244, 2021, doi: 10.26740/jinacs.v2n04.p235-244.
- [10] A. Tangkelayuk and E. Mailoa, “Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes Dan Decision Tree,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 2, pp. 1109–1119, Jun. 2022, doi: 10.35957/jatisi.v9i2.2048.
- [11] F. Fiddin, M. Yusuf Syahbarna, and M. Ridwan, “Penggunaan Supervised Learning untuk Prediksi Validitas Ulasan Negatif Aplikasi Tokopedia Berdasarkan Pengalaman Pengguna Ahli,” *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)*, vol. 23, no. 2, pp. 409–417, Aug. 2024, doi: 10.53513/jis.v23i2.10030.
- [12] R. H. Nufus *et al.*, “Analisis Sentimen Persepsi Masyarakat Terhadap Timnas Indonesia U-23 dalam AFC-23 Asian Cup 2024 Pada Media Sosial X Menggunakan Metode Naïve Bayes Classifier,” 2024. [Online]. Available: <https://journal.stmiki.ac.id>
- [13] R. Prajamukti and M. Mega Santoni, “KLASIFIKASI DAN ANALISIS SENTIMEN PADA DATA TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES (STUDI KASUS: TIMNAS INDONESIA SENIOR, U-23, DAN U-19),” (*SEINASI-KESI*), pp. 102–109, Oct. 2021, [Online]. Available: <https://t.co/SAbU6Prz>
- [14] T. Juniardi and C. A. Sugianto, “ANALISIS SENTIMEN TIM NASIONAL SEPAK BOLA INDONESIA DI TURNAMEN PIALA DUNIA U-17 INDONESIA PADA TWITTER (X) MENGGUNAKAN ALGORITMA NAÏVE BAYES,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3S1, Oct. 2024, doi: 10.23960/jitet.v12i3S1.5188.
- [15] M. Malik, D. Aji Pangestu, and M. Risky Pribadi, “Analisis Sentimen Hasil Pertandingan Sepakbola Timnas Indo-nesia di Piala Asia U-23 pada Platform Youtube menggunakan Algoritma Support Vector Machine (SVM),” 2024. [Online]. Available: <https://jurnal.politap.ac.id/index.php/aicoms>
- [16] S. Tetra Oktaviani, H. Baturohmah, P. Studi Sistem Informasi, U. Nusa Putra Jl Raya Cibolang No, C. Kaler, and J. Barat, “SISMATIK (Seminar Nasional Sistem Informasi dan Manajemen Informatika) Universitas Nusa Putra, 12 Agustus 2023 SENTIMEN ANALISIS PENGGUNA TWITTER INDONESIA TERHADAP PIALA DUNIA 2022 MENGGUNAKAN METODE NAÏVE BAYES CLASIFIER,” *SISMATIK (Seminar Nasional Sistem Informasi dan Manajemen Informatika)*, Aug. 2023.
- [17] M. Yunus, M. Husni, and M. M. Mufadhdhal, “Klasifikasi Sentimen Terhadap Badan Penyelenggara Jaminan Sosial (BPJS) Pada Media Sosial Twitter Menggunakan Naive Bayes,” *SMATIKA JURNAL*, vol. 11, no. 02, pp. 81–91, Dec. 2021, doi: 10.32664/smatika.v11i02.577.
- [18] O. I. Gifari, M. Adha, I. Rifky Hendrawan, F. Freddy, and S. Durrand, “Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine,” *JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY)*, vol. 2, no. 1, pp. 36–40, Mar. 2022, doi: 10.46229/jifotech.v2i1.330.
- [19] E. Suryati, A. Ari Aldino, N. Penulis Korespondensi, and E. Suryati Submitted, “Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM),” *JURNAL TEKNOLOGI DAN SISTEM INFORMASI*, vol. 4, no. 1, pp. 96–106, Mar. 2023, doi: 10.33365/jtsi.v4i1.2445.
- [20] Ramadhani, Ramadhanu, Amrullah fajri artha abdillah, and M. Ridwan, “Studi Komparatif Multinomial Naïve Bayes, Decision Tree, dan K-Nearest Neighbor dalam Klasifikasi Validasi Ulasan Clash of Clans oleh Pengguna Ahli,” *Jurnal sistem dan teknologi informasi*, vol. 12, no. 4, pp. 653–660, Oct. 2024, doi: 10.26418/justin.v12i4.81638.
- [21] D. Toresa, S. Rico, F. Sitorus, I. Muzdalifah, F. Wiza, and R. Syelly, “Analisis Sentimen Terhadap Ulasan Penggunaan Dompot Digital DANA Menggunakan Metode Klasifikasi Support Vector Machine Sentiment Analysis of Reviews on the Use of DANA’s Digital Wallet Using the Support Vector Machine Classification Method,” *Technologica*, vol. 3, no. 2, pp. 64–74, Jul. 2024, doi: 10.55043/technologica.v3i2.163.
- [22] M. U. Albab, Y. Karuniawati P, and M. N. Fawaiq, “Optimization of the Stemming Technique on Text preprocessing President 3 Periods Topic,” *Jurnal TRANSFORMATIKA*, vol. 20, no. 2, pp. 1–10, Jan. 2023, doi: 10.26623/transformatika.v20i2.5374.
- [23] T. Hidayat *et al.*, “Performance Prediction Using Cross Validation (GridSearchCV) for Stunting Prevalence,” in *2024 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*, IEEE, Feb. 2024, pp. 1–6. doi: 10.1109/AIMS61812.2024.10512657