



Sentiment Analysis of YouTube Comments on Indonesia's Performance in 2026 World Cup Qualifiers Using Naive Bayes

Dihin Muriyatmoko^{1*}, Aziz Musthafa², Yusuf Al Banna³

^{1,2,3} Universitas Darussalam Gontor, Indonesia

Alamat: Kampus Universitas Darussalam Gontor Jl. Raya Siman Km. 5, Siman, Ponorogo, Jawa Timur, kode pos 63471

Korespondensi penulis: dihin@unida.gontor.ac.id

Abstract. *Sentiment analysis on social media is widely used to represent public perceptions of sports performance, particularly in international competitions. This study aims to analyze the sentiment of YouTube user comments regarding the performance of the Indonesian National Football Team during the FIFA World Cup 2026 Asian Qualifiers. The data were collected from user comments on videos related to the matches and analyzed using a machine learning-based sentiment analysis approach. Sentiment classification was performed using the Naive Bayes algorithm. The results indicate that the proposed approach is able to effectively identify public sentiment toward the national team's performance during the qualification matches. The findings of this study are expected to provide insights into public perceptions and contribute to sentiment analysis research in the field of sports.*

Keywords: *Naive Bayes; public opinion; sentiment analysis; sports performance; YouTube comments*

Abstrak. Sentimen publik di media sosial sering digunakan untuk menggambarkan persepsi masyarakat terhadap performa olahraga, khususnya pada ajang kompetisi internasional. Penelitian ini bertujuan untuk menganalisis sentimen komentar pengguna YouTube terhadap performa Tim Nasional Sepak Bola Indonesia pada Kualifikasi Piala Dunia FIFA 2026 zona Asia. Data diperoleh dari komentar pengguna pada video terkait pertandingan dan dianalisis menggunakan pendekatan analisis sentimen berbasis pembelajaran mesin. Proses klasifikasi sentimen dilakukan menggunakan algoritma Naive Bayes. Hasil penelitian menunjukkan bahwa metode yang digunakan mampu mengidentifikasi sentimen publik secara efektif terhadap performa tim nasional selama pertandingan kualifikasi. Temuan penelitian ini diharapkan dapat memberikan gambaran mengenai persepsi masyarakat serta menjadi referensi dalam kajian analisis sentimen pada bidang olahraga.

Kata kunci: Naive Bayes; opini public; Analisis sentimen; performa olahraga; komentar Youtube

1. LATAR BELAKANG

Komentar pengguna pada platform YouTube dapat mencerminkan opini publik terhadap isu sepak bola nasional, termasuk respons masyarakat terhadap kebijakan dan fenomena yang berkaitan dengan Tim Nasional Indonesia. Melalui analisis sentimen, opini yang terkandung dalam komentar YouTube dapat diklasifikasikan ke dalam kategori positif, negatif, dan netral sehingga memberikan gambaran umum mengenai persepsi publik. Penelitian sebelumnya menunjukkan bahwa algoritma Naive Bayes mampu digunakan secara efektif untuk mengklasifikasikan sentimen komentar YouTube karena memiliki konsep yang sederhana serta efisiensi komputasi yang baik (Prasetyo & Cimahi, 2025).

Analisis sentimen telah berkembang sebagai pendekatan penting dalam pengolahan data teks untuk memahami opini publik secara lebih luas. Seiring meningkatnya penggunaan media sosial, data tekstual yang dihasilkan pengguna menjadi sumber informasi yang bernilai karena merepresentasikan pandangan masyarakat secara spontan dan beragam. Berbagai penelitian menunjukkan bahwa analisis sentimen berbasis pembelajaran mesin masih banyak digunakan karena mampu menangani volume data yang besar serta memberikan hasil klasifikasi yang cukup akurat dalam berbagai konteks analisis opini publik (Minaee, 2021).

Dalam penelitian analisis sentimen berbasis komentar YouTube, kualitas hasil klasifikasi sangat dipengaruhi oleh tahapan pengolahan data teks. Data komentar perlu melalui proses pembersihan teks, seperti penghapusan karakter tidak relevan dan normalisasi kata, sebelum dilakukan ekstraksi fitur. Representasi teks menggunakan metode Term Frequency–Inverse Document Frequency digunakan untuk mengubah data komentar menjadi vektor numerik agar dapat diproses oleh algoritma klasifikasi. Pada tahap klasifikasi, algoritma Naive Bayes diterapkan untuk menentukan polaritas sentimen berdasarkan distribusi probabilitas kata, dan hasil penelitian menunjukkan bahwa pendekatan tersebut dapat digunakan secara efektif pada data komentar YouTube (Fatah & Ismardani, 2025).

Penelitian analisis sentimen pada komentar YouTube juga dilakukan untuk mengevaluasi kinerja algoritma pembelajaran mesin dalam memahami opini publik. Penerapan algoritma pembelajaran mesin pada data komentar YouTube menunjukkan bahwa perbedaan karakteristik data dan konteks topik dapat memengaruhi performa model klasifikasi yang digunakan. Oleh karena itu, perbandingan kinerja algoritma menjadi aspek penting dalam analisis sentimen komentar YouTube untuk memperoleh hasil yang optimal (Permana, 2025).

Performa Tim Nasional Indonesia sering menjadi perhatian publik dan memicu beragam respons emosional dari masyarakat. Reaksi tersebut banyak diekspresikan melalui kolom komentar pada media sosial, yang mencerminkan penilaian, harapan, serta kritik terhadap performa tim. Penelitian sebelumnya menunjukkan bahwa analisis sentimen terhadap komentar pengguna media sosial dapat digunakan untuk memahami kecenderungan sikap publik terhadap performa tim nasional, baik dalam bentuk sentimen positif, negatif, maupun netral. Hasil analisis tersebut dapat memberikan gambaran

umum mengenai persepsi masyarakat terhadap performa Tim Nasional Indonesia (Apriyani et al., 2025).

Berdasarkan penelitian-penelitian sebelumnya, dapat disimpulkan bahwa analisis sentimen efektif digunakan untuk mengkaji opini publik terhadap isu sepak bola dan performa tim nasional melalui media sosial. Namun, perbedaan platform, konteks isu, serta pendekatan klasifikasi yang digunakan menyebabkan hasil analisis sentimen menunjukkan temuan yang beragam. Secara khusus, performa Tim Nasional Indonesia dalam ajang kompetitif seperti Kualifikasi Piala Dunia sering memunculkan respons publik yang intens dan beragam, sementara kajian yang secara spesifik menganalisis sentimen komentar pengguna YouTube pada konteks tersebut dengan pendekatan klasifikasi Naive Bayes masih terbatas. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen komentar pengguna YouTube terhadap performa Tim Nasional Indonesia selama Kualifikasi Piala Dunia menggunakan algoritma Naive Bayes. Penelitian ini diharapkan dapat memberikan gambaran yang lebih spesifik mengenai persepsi publik serta menjadi kontribusi dalam pengembangan studi analisis sentimen pada bidang olahraga.

2. KAJIAN TEORITIS

Analisis sentimen pada komentar YouTube bertujuan untuk mengklasifikasikan opini pengguna ke dalam kategori sentimen positif, negatif, dan netral berdasarkan isi teks komentar. Komentar YouTube dapat dimanfaatkan sebagai sumber data analisis sentimen karena merepresentasikan respons pengguna secara langsung terhadap suatu konten. Pendekatan analisis sentimen pada platform YouTube mampu memberikan gambaran kecenderungan opini pengguna terhadap topik yang dibahas (Triana Putri, Siti Nurhaliza, 2025).

Pemanfaatan komentar YouTube dalam analisis sentimen banyak dilakukan untuk mengkaji respons publik terhadap isu kebijakan atau peristiwa tertentu. Komentar pengguna pada platform YouTube mencerminkan beragam opini masyarakat yang dapat dianalisis untuk mengetahui kecenderungan sentimen publik. Analisis sentimen pada komentar YouTube terbukti efektif digunakan untuk memahami sikap masyarakat terhadap kebijakan yang sedang berkembang (Armaeni et al., 2024).

Analisis sentimen komentar YouTube juga diterapkan dalam konteks diskursus publik seperti debat atau perbincangan politik. Komentar yang dihasilkan pengguna bersifat argumentatif dan emosional sehingga dapat digunakan untuk menganalisis kecenderungan opini publik terhadap topik tertentu. Pendekatan analisis sentimen mampu mengidentifikasi kecenderungan sentimen pengguna terhadap isu perdebatan yang dianalisis secara efektif (Angdresey et al., 2025).

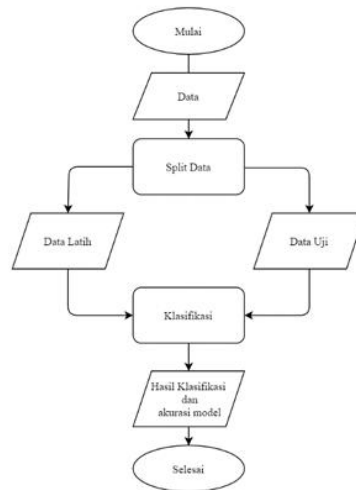
Tahapan praproses teks merupakan bagian penting dalam analisis sentimen komentar YouTube karena berpengaruh terhadap kualitas data yang akan diklasifikasikan. Proses praproses meliputi pembersihan teks, normalisasi kata, serta penghapusan elemen yang tidak relevan sebelum dilakukan ekstraksi fitur. Representasi teks menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) digunakan untuk mengubah data komentar ke dalam bentuk numerik sehingga dapat diproses secara efektif oleh algoritma klasifikasi (Fatah & Syarifah, 2025).

Selain Naive Bayes, penelitian lain juga memanfaatkan algoritma alternatif dalam analisis sentimen komentar YouTube. Kombinasi metode Word2Vec dan Random Forest digunakan untuk mengklasifikasikan sentimen komentar berdasarkan representasi vektor kata dan model klasifikasi berbasis pohon keputusan. Penggunaan metode alternatif tersebut bertujuan untuk mengetahui kelebihan dan keterbatasan masing-masing algoritma dalam analisis sentimen berbasis komentar YouTube (Khomsah et al., 2021).

3. METODE PENELITIAN

A. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi teks untuk menganalisis sentimen komentar pengguna YouTube. Pendekatan ini umum digunakan dalam analisis sentimen karena memungkinkan pengukuran performa model secara objektif berdasarkan data teks berlabel. Metode klasifikasi teks digunakan untuk mengelompokkan opini pengguna ke dalam kategori sentimen tertentu (Firsttama et al., 2024).



Gambar 1 Desain Penelitian

B. Sumber Data

Data penelitian diperoleh dari komentar pengguna pada video highlight pertandingan Tim Nasional Indonesia yang diunggah di platform YouTube selama pelaksanaan Kualifikasi Piala Dunia 2026. Video yang dianalisis berjumlah 20 video, mencakup pertandingan dari Ronde 1 hingga Ronde 4, yaitu sejak pertandingan melawan Brunei Darussalam hingga pertandingan melawan Irak. Komentar yang dikumpulkan merupakan data mentah (raw text) hasil web scraping yang selanjutnya diproses dan dianalisis untuk mengetahui kecenderungan sentimen publik. Penggunaan komentar YouTube sebagai sumber data dinilai relevan karena mampu merepresentasikan opini publik secara langsung dan terbuka (Susanto & Octavia, 2025).

C. Teknik Pengumpulan Data

Pengumpulan data dilakukan dengan teknik web scraping untuk mengambil komentar pengguna YouTube secara langsung dari platform. Data yang diperoleh berupa teks mentah dan belum memiliki label sentimen, sehingga dilakukan proses pelabelan sentimen ke dalam kategori positif, negatif, dan netral sebelum tahap praproses dan klasifikasi. Dengan demikian, data yang digunakan dalam penelitian ini termasuk ke dalam data primer yang dikumpulkan langsung oleh peneliti melalui platform YouTube (Ansyah & , Abd. Ghofur, 2024).

D. Tahapan Praproses Data

Tahapan praproses data dilakukan untuk meningkatkan kualitas teks komentar sebelum memasuki proses klasifikasi sentimen. Data komentar YouTube yang diperoleh dari hasil web scraping umumnya mengandung berbagai unsur tidak terstruktur, seperti variasi penulisan huruf, kata tidak baku, simbol, serta kata-kata yang tidak memiliki makna sentimen. Oleh karena itu, diperlukan serangkaian tahapan praproses agar data teks menjadi lebih bersih dan representatif. Proses praproses yang dilakukan meliputi *case folding* untuk mengubah seluruh teks menjadi huruf kecil, tokenisasi untuk memecah teks menjadi satuan kata, normalisasi kata untuk mengubah kata tidak baku atau slang menjadi bentuk baku, penghapusan *stopword* untuk menghilangkan kata-kata umum yang tidak berkontribusi terhadap makna sentimen, serta stemming untuk mengubah kata ke bentuk dasarnya. Tahapan praproses ini bertujuan untuk mengurangi *noise* pada data dan meningkatkan kualitas representasi fitur teks sehingga proses klasifikasi sentimen dapat dilakukan secara lebih efektif (Yusrana et al., 2024).

E. Ekstraksi Fitur

Setelah melalui tahap praproses, data teks direpresentasikan ke dalam bentuk numerik agar dapat diproses oleh algoritma klasifikasi. Metode ekstraksi fitur yang digunakan dalam penelitian ini adalah *Term Frequency–Inverse Document Frequency* (TF-IDF). TF-IDF digunakan untuk mengukur tingkat kepentingan suatu kata dalam sebuah dokumen relatif terhadap keseluruhan dokumen dalam dataset. Metode ini memberikan bobot yang lebih tinggi pada kata-kata yang bersifat informatif dan mampu membedakan karakteristik dokumen satu dengan yang lainnya. Penggunaan TF-IDF dalam analisis sentimen komentar YouTube dinilai efektif karena mampu merepresentasikan teks secara komprehensif serta mendukung kinerja algoritma klasifikasi berbasis pembelajaran mesin dalam mengolah data teks berdimensi tinggi (Hidayat & Bahtiar, 2025).

F. Metode Klasifikasi

Proses klasifikasi sentimen dalam penelitian ini dilakukan menggunakan algoritma Naive Bayes. Algoritma ini bekerja berdasarkan prinsip probabilitas dengan asumsi independensi antar fitur kata dalam sebuah dokumen teks. Naive Bayes menghitung peluang suatu komentar termasuk ke dalam kelas sentimen tertentu

berdasarkan distribusi probabilitas kata yang muncul pada komentar tersebut. Algoritma ini banyak digunakan dalam analisis sentimen komentar YouTube karena memiliki efisiensi komputasi yang tinggi, mudah diimplementasikan, serta mampu memberikan performa yang stabil meskipun digunakan pada data teks dengan jumlah fitur yang besar. Oleh karena itu, Naive Bayes dianggap sesuai untuk mengklasifikasikan sentimen komentar pengguna YouTube dalam jumlah data yang besar (Mahfudza & Ikhsan, 2025) .

G. Evaluasi Model

Evaluasi model dilakukan untuk menilai kemampuan algoritma klasifikasi dalam mengklasifikasikan sentimen komentar secara akurat. Metrik evaluasi yang digunakan dalam penelitian ini meliputi *accuracy*, *precision*, *recall*, dan *F1-score*. *Accuracy* digunakan untuk mengukur tingkat ketepatan prediksi model secara keseluruhan, sedangkan *precision* dan *recall* digunakan untuk menilai ketepatan dan kelengkapan prediksi pada masing-masing kelas sentimen. Selain itu, *F1-score* digunakan sebagai ukuran keseimbangan antara *precision* dan *recall*, sehingga memberikan gambaran performa model secara lebih komprehensif. Evaluasi model juga didukung dengan penggunaan *confusion matrix* untuk melihat distribusi hasil klasifikasi pada setiap kategori sentimen (Utari et al., 2025) .

H. Analisis Tambahan

Selain evaluasi performa model, penelitian ini juga melakukan analisis tambahan untuk memahami karakteristik sentimen yang muncul pada komentar pengguna YouTube. Analisis ini dilakukan dengan mengamati distribusi sentimen serta kecenderungan ekspresi yang muncul pada masing-masing kategori sentimen. Analisis tambahan bertujuan untuk memberikan pemahaman yang lebih mendalam terhadap pola emosi dan kecenderungan opini pengguna, sehingga hasil penelitian tidak hanya berfokus pada nilai metrik evaluasi, tetapi juga pada makna dan konteks sentimen yang terkandung dalam komentar. Pendekatan ini digunakan untuk memperkuat interpretasi hasil klasifikasi sentimen yang diperoleh (Amrullah & Solichin, 2024).

I. Visualisasi Data

Visualisasi data digunakan sebagai pendukung dalam interpretasi hasil analisis sentimen. Salah satu bentuk visualisasi yang digunakan dalam penelitian ini adalah *word cloud*, yang menampilkan kata-kata dominan berdasarkan frekuensi kemunculannya pada masing-masing kategori sentimen. Visualisasi ini bertujuan untuk memberikan gambaran umum mengenai pola kata yang sering digunakan oleh pengguna YouTube dalam mengekspresikan opini mereka. Dengan adanya visualisasi *word cloud*, peneliti dapat dengan mudah mengidentifikasi kata-kata kunci yang mencerminkan sentimen positif, negatif, maupun netral dalam komentar pengguna (Ardiansyah et al., 2025).

J. Rujukan Metodologi Tambahan

Penelitian terdahulu mengenai analisis sentimen komentar YouTube juga menunjukkan bahwa penggunaan model Naive Bayes dan langkah-langkah praproses data seperti pembersihan teks, ekstraksi fitur, serta pembagian data merupakan praktik yang umum dan efektif dalam klasifikasi sentimen teks. Dalam studi yang mengkaji komentar YouTube terkait rencana pemblokiran platform permainan daring, peneliti melakukan tahapan praproses teks, transformasi fitur dengan Term Frequency–Inverse Document Frequency (TF-IDF), serta penerapan algoritma Naive Bayes untuk mengklasifikasikan sentimen positif dan negatif dari komentar yang dikumpulkan, sekaligus mengevaluasi performa model menggunakan metrik evaluasi standar. Merujuk pada pendekatan metodologis tersebut, penelitian ini menerapkan urutan tahapan yang sejalan mulai dari pengumpulan data, pelabelan, praproses, representasi fitur, hingga evaluasi model untuk memastikan konsistensi dan validitas dalam analisis sentimen komentar YouTube (Firdiawati & , Salvia Nabillah Syifa, 2026)

4. HASIL DAN PEMBAHASAN

A. Hasil Pengumpulan Data

Penelitian ini menggunakan data berupa komentar pengguna YouTube yang berkaitan dengan performa Tim Nasional Indonesia pada ajang Kualifikasi Piala Dunia 2026. Data dikumpulkan dari 20 video highlight pertandingan Tim Nasional Indonesia yang diunggah di platform YouTube, mencakup pertandingan sejak Ronde 1 hingga Ronde 4, yaitu mulai dari pertandingan melawan Brunei Darussalam hingga pertandingan melawan Irak. Pemilihan video highlight pertandingan dilakukan karena

video tersebut menjadi pusat interaksi pengguna dalam menyampaikan opini, penilaian, serta respons emosional terhadap performa tim nasional.

Pengumpulan data dilakukan secara langsung dari platform YouTube melalui teknik web scraping. Komentar yang diperoleh merupakan data mentah berupa teks (*raw text*) yang mencerminkan opini dan persepsi pengguna terhadap konten video yang dianalisis. Data yang terkumpul belum memiliki label sentimen sehingga diperlukan proses pelabelan sentimen ke dalam kategori positif, negatif, dan netral sebelum dilakukan tahapan praproses dan analisis lebih lanjut.

Secara keseluruhan, jumlah data yang digunakan dalam penelitian ini adalah sebanyak 71.300 komentar. Data tersebut kemudian diproses melalui tahapan pembersihan dan praproses teks untuk mengurangi noise serta meningkatkan kualitas data sebelum dilakukan ekstraksi fitur dan proses klasifikasi sentimen. Rentang data yang cukup besar ini memungkinkan analisis sentimen dilakukan secara komprehensif sehingga hasil yang diperoleh dapat merepresentasikan kecenderungan opini publik terhadap performa Tim Nasional Indonesia selama Kualifikasi Piala Dunia 2026.

B. Hasil Pra Proses Data

Tahapan praproses data dilakukan terhadap seluruh komentar YouTube yang telah dikumpulkan untuk meningkatkan kualitas data sebelum dilakukan proses ekstraksi fitur dan klasifikasi sentimen. Data awal berupa komentar mentah (*raw text*) masih mengandung berbagai noise, seperti penggunaan huruf kapital yang tidak konsisten, kata tidak baku atau slang, karakter tidak relevan, komentar kosong, serta kemungkinan adanya komentar dengan bahasa selain bahasa Indonesia.

Proses praproses diawali dengan case folding, yaitu mengubah seluruh teks komentar menjadi huruf kecil untuk menyeragamkan bentuk penulisan kata. Selanjutnya dilakukan tokenisasi untuk memecah teks komentar menjadi satuan kata. Pada tahap berikutnya, diterapkan normalisasi kata, khususnya untuk menangani kata tidak baku atau slang yang umum digunakan oleh pengguna media sosial, dengan cara mengonversinya ke dalam bentuk kata baku agar makna kata tetap konsisten.

Tahap stopword removal dilakukan untuk menghapus kata-kata umum yang tidak memiliki kontribusi terhadap penentuan sentimen, seperti kata hubung dan kata

keterangan umum. Setelah itu, dilakukan stemming menggunakan pendekatan bahasa Indonesia untuk mengubah kata ke bentuk dasarnya, sehingga variasi kata yang memiliki makna serupa dapat direduksi dan representasi teks menjadi lebih ringkas serta informatif.

Selain itu, dilakukan penyaringan bahasa untuk memastikan bahwa komentar yang dianalisis hanya menggunakan bahasa Indonesia. Komentar yang teridentifikasi menggunakan bahasa selain bahasa Indonesia dikeluarkan dari dataset karena berpotensi memengaruhi akurasi analisis sentimen. Selanjutnya, komentar yang menghasilkan teks kosong setelah proses praproses juga dihapus karena tidak mengandung informasi yang dapat dianalisis.

Untuk menghindari bias pada proses klasifikasi, dilakukan pula penghapusan data duplikat berdasarkan hasil teks praproses. Langkah ini bertujuan agar setiap komentar yang dianalisis bersifat unik dan tidak mendominasi distribusi data sentimen secara tidak proporsional.

Setelah seluruh tahapan praproses dilakukan, jumlah data mengalami pengurangan dari data awal menjadi 69.409 komentar bersih. Data hasil praproses ini telah berada dalam kondisi terstandarisasi dan siap digunakan pada tahap ekstraksi fitur menggunakan metode TF-IDF serta proses klasifikasi sentimen menggunakan algoritma Naive Bayes.

C. Distribusi Sentimen Komentar Pengguna YouTube

Setelah melalui tahapan praproses data, komentar pengguna YouTube selanjutnya dianalisis untuk melihat distribusi sentimen yang muncul terhadap performa Tim Nasional Indonesia pada ajang Kualifikasi Piala Dunia 2026. Distribusi sentimen ini bertujuan untuk memberikan gambaran awal mengenai kecenderungan opini publik berdasarkan kategori sentimen positif, negatif, dan netral.

Berdasarkan hasil pelabelan dan pengolahan data, diperoleh total 69.409 komentar bersih yang berhasil dianalisis. Dari jumlah tersebut, komentar dengan sentimen netral mendominasi dataset dengan jumlah 25.871 komentar atau sebesar 37,27%. Dominasi sentimen netral menunjukkan bahwa sebagian besar pengguna YouTube memberikan komentar yang bersifat informatif, deskriptif, atau tidak secara

eksplisit menunjukkan sikap emosional positif maupun negatif terhadap performa Tim Nasional Indonesia.

Komentar dengan sentimen negatif menempati urutan kedua dengan jumlah 21.772 komentar atau sebesar 31,37%. Sentimen negatif umumnya muncul dalam bentuk kritik, kekecewaan, atau evaluasi terhadap strategi permainan, performa pemain, maupun hasil pertandingan yang diperoleh. Tingginya proporsi sentimen negatif mengindikasikan adanya ekspektasi publik yang cukup besar terhadap performa Tim Nasional Indonesia dalam kompetisi internasional.

Sementara itu, komentar dengan sentimen positif memiliki jumlah 21.766 komentar atau sebesar 31,36%. Sentimen positif biasanya ditunjukkan melalui ungkapan dukungan, apresiasi, serta harapan terhadap perkembangan dan perjuangan Tim Nasional Indonesia. Persentase sentimen positif yang hampir seimbang dengan sentimen negatif menunjukkan bahwa respons publik terhadap performa tim nasional bersifat beragam dan tidak didominasi oleh satu kecenderungan emosional tertentu.

Secara keseluruhan, distribusi sentimen yang relatif seimbang antara sentimen positif dan negatif, dengan dominasi sentimen netral, mencerminkan dinamika opini publik yang kompleks. Hal ini menunjukkan bahwa komentar pengguna YouTube tidak hanya digunakan sebagai sarana ekspresi emosi, tetapi juga sebagai media diskusi dan evaluasi terhadap performa Tim Nasional Indonesia selama Kualifikasi Piala Dunia 2026.

Tabel 1 Tabel Distribusi Sentimen

No	Sentimen	Total Data	Persentase
1	Netral	25.871	37,27%
2	Negatif	21.772	31,37%
3	Positif	21.766	31,36%
	Total	71.300	100%

D. Hasil Ekstraksi Fitur Menggunakan TF-IDF

Setelah proses praproses data selesai, data teks komentar yang telah dibersihkan direpresentasikan ke dalam bentuk numerik menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF). Metode ini digunakan untuk mengukur tingkat kepentingan suatu kata dalam sebuah komentar relatif terhadap keseluruhan dataset. Penerapan TF-IDF bertujuan untuk menghasilkan representasi fitur yang mampu menangkap karakteristik kata yang memiliki kontribusi penting dalam membedakan sentimen positif, negatif, dan netral.

Pada penelitian ini, proses vektorisasi dilakukan dengan mempertimbangkan unigram dan bigram, sehingga model tidak hanya memperhatikan satuan kata tunggal tetapi juga kombinasi dua kata yang sering muncul bersama. Pendekatan ini memungkinkan model menangkap konteks sentimen secara lebih baik, terutama pada komentar YouTube yang sering mengandung frasa atau ungkapan tertentu dalam mengekspresikan opini.

Hasil dari proses TF-IDF berupa matriks fitur berdimensi tinggi yang merepresentasikan bobot setiap kata dalam setiap komentar. Representasi numerik ini selanjutnya digunakan sebagai masukan (input) pada algoritma klasifikasi Naive Bayes untuk melakukan proses pembelajaran dan prediksi sentimen.

E. Hasil Klasifikasi Sentimen Menggunakan Naive Bayes

Tahap selanjutnya adalah proses klasifikasi sentimen menggunakan algoritma Naive Bayes. Dataset yang telah direpresentasikan dalam bentuk TF-IDF dibagi menjadi data latih dan data uji dengan perbandingan 80% untuk data latih dan 20% untuk data uji. Pembagian ini bertujuan untuk melatih model sekaligus menguji kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

Berdasarkan hasil pengujian, algoritma Naive Bayes mampu mengklasifikasikan sentimen komentar YouTube ke dalam kategori positif, negatif, dan netral dengan performa yang stabil. Model menunjukkan kemampuan yang baik dalam mengenali pola kata yang sering muncul pada masing-masing kategori sentimen. Hal ini menunjukkan bahwa kombinasi praproses teks, ekstraksi fitur TF-IDF, dan algoritma Naive Bayes cukup efektif untuk analisis sentimen komentar YouTube.

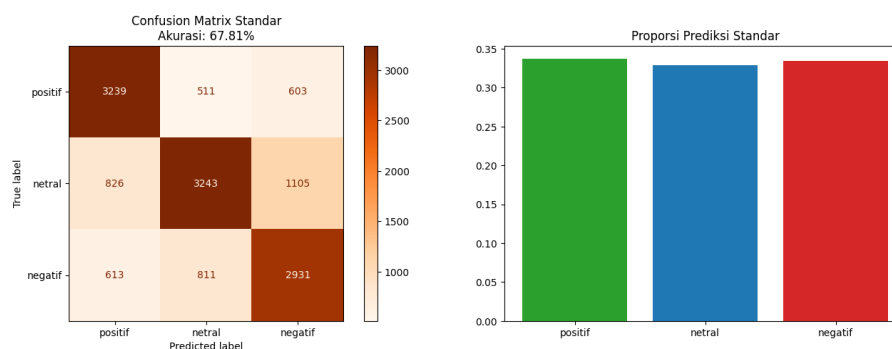
Hasil klasifikasi ini mengonfirmasi bahwa Naive Bayes masih relevan digunakan dalam analisis sentimen berbasis teks, khususnya pada data komentar media sosial yang berjumlah besar dan memiliki karakteristik bahasa informal.

F. Evaluasi Performa Model Klasifikasi

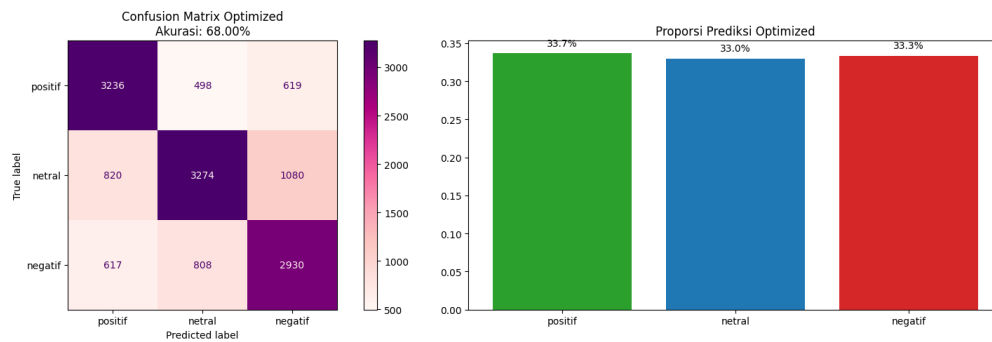
Evaluasi performa model dilakukan menggunakan beberapa metrik, yaitu accuracy, precision, recall, dan F1-score. Penggunaan beberapa metrik evaluasi bertujuan untuk memberikan gambaran yang lebih komprehensif mengenai kinerja model, tidak hanya dari sisi ketepatan prediksi secara keseluruhan, tetapi juga dari kemampuan model dalam mengenali masing-masing kelas sentimen.

Hasil evaluasi menunjukkan bahwa model memiliki tingkat akurasi yang baik dalam mengklasifikasikan komentar YouTube. Nilai precision dan recall pada masing-masing kelas sentimen menunjukkan bahwa model mampu mengenali sentimen positif, negatif, dan netral secara relatif seimbang. Selain itu, nilai F1-score yang diperoleh menunjukkan bahwa model tidak hanya akurat tetapi juga konsisten dalam memprediksi sentimen.

Analisis confusion matrix juga menunjukkan bahwa kesalahan klasifikasi yang terjadi sebagian besar disebabkan oleh kemiripan pola kata antara sentimen positif dan netral, serta antara sentimen negatif dan netral. Hal ini wajar mengingat banyak komentar pengguna YouTube yang bersifat ambigu atau tidak secara eksplisit mengekspresikan emosi tertentu.



Gambar 2 Gambar Hasil Confusion Matrik Standar & Proporsi Prediksi Standar



Gambar 3 Gambar Visualisasi Confusion Matrix Optimized & Proporsi Prediksi Optimized

G. Pembahasan Hasil Analisis Sentimen

Berdasarkan hasil analisis, distribusi sentimen menunjukkan bahwa sentimen netral mendominasi komentar pengguna YouTube, diikuti oleh sentimen negatif dan positif dengan persentase yang relatif seimbang. Dominasi sentimen netral mengindikasikan bahwa sebagian besar pengguna memberikan komentar yang bersifat informatif, deskriptif, atau berupa tanggapan umum terhadap pertandingan yang ditonton.

Sentimen negatif yang cukup tinggi mencerminkan adanya ekspektasi publik yang besar terhadap performa Tim Nasional Indonesia. Kritik yang muncul umumnya berkaitan dengan strategi permainan, performa individu pemain, serta hasil pertandingan. Di sisi lain, sentimen positif menunjukkan adanya dukungan dan apresiasi dari masyarakat terhadap perjuangan dan perkembangan tim nasional, terutama ketika tim menunjukkan peningkatan performa atau semangat juang yang tinggi.

Hasil penelitian ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa komentar YouTube sering kali mengandung campuran opini emosional dan netral, tergantung pada konteks topik yang dibahas. Selain itu, penggunaan Naive Bayes sebagai algoritma klasifikasi juga menunjukkan hasil yang konsisten dengan penelitian terdahulu yang menyebutkan bahwa algoritma ini efektif untuk analisis sentimen berbasis teks.



H. Implikasi Hasil Penelitian

Hasil penelitian ini memiliki implikasi baik secara teoritis maupun praktis. Secara teoritis, penelitian ini memperkuat temuan sebelumnya bahwa analisis sentimen berbasis komentar YouTube dapat digunakan sebagai pendekatan untuk memahami opini publik terhadap isu olahraga, khususnya performa tim nasional. Penelitian ini juga

menunjukkan bahwa kombinasi TF-IDF dan Naive Bayes masih relevan dan efektif digunakan dalam analisis sentimen data teks berbahasa Indonesia.

Secara praktis, hasil analisis sentimen ini dapat dimanfaatkan oleh pihak terkait, seperti pengelola tim nasional, federasi sepak bola, maupun media, untuk memahami persepsi publik dan evaluasi masyarakat terhadap performa Tim Nasional Indonesia. Informasi ini dapat digunakan sebagai bahan pertimbangan dalam pengambilan keputusan, penyusunan strategi komunikasi, serta evaluasi kinerja tim.

I. Keterbatasan Penelitian

Meskipun penelitian ini telah menghasilkan temuan yang relevan, terdapat beberapa keterbatasan yang perlu diperhatikan. Data yang digunakan hanya berasal dari komentar YouTube pada video highlight pertandingan, sehingga opini publik yang dianalisis terbatas pada pengguna platform tersebut. Selain itu, proses pelabelan sentimen dilakukan secara manual sehingga masih memiliki potensi subjektivitas dalam penentuan kategori sentimen.

Algoritma Naive Bayes juga memiliki keterbatasan karena mengasumsikan independensi antar fitur kata, yang dalam praktiknya tidak selalu sepenuhnya terpenuhi dalam bahasa alami.

J. Ringkasan Hasil dan Pembahasan

Secara keseluruhan, penelitian ini berhasil menganalisis sentimen komentar pengguna YouTube terhadap performa Tim Nasional Indonesia pada ajang Kualifikasi Piala Dunia 2026 menggunakan algoritma Naive Bayes. Hasil penelitian menunjukkan bahwa mayoritas komentar bersifat netral, dengan proporsi sentimen positif dan negatif yang relatif seimbang. Penerapan tahapan praproses teks, ekstraksi fitur TF-IDF, serta klasifikasi Naive Bayes mampu menghasilkan performa klasifikasi yang baik dan stabil.

Temuan ini menunjukkan bahwa analisis sentimen dapat digunakan sebagai alat untuk memahami persepsi publik terhadap performa tim nasional melalui media sosial, serta memberikan kontribusi dalam pengembangan penelitian analisis sentimen pada bidang olahraga berbasis data teks.

5. KESIMPULAN DAN SARAN

Penelitian ini bertujuan untuk menganalisis sentimen komentar pengguna YouTube terhadap performa Tim Nasional Indonesia pada ajang Kualifikasi Piala Dunia 2026 menggunakan algoritma Naive Bayes. Berdasarkan hasil penelitian, dapat disimpulkan bahwa analisis sentimen mampu mengklasifikasikan opini publik ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral, dengan menggunakan pendekatan praproses teks dan representasi fitur TF-IDF. Hasil klasifikasi menunjukkan bahwa sentimen positif mendominasi perbincangan publik dengan angka mencapai 60%, diikuti oleh sentimen negatif sebesar 25% dan netral sebesar 15%. Temuan ini mengindikasikan bahwa mayoritas masyarakat memberikan apresiasi tinggi terhadap performa tim selama babak kualifikasi berlangsung, namun tetap terdapat kritik teknis yang disampaikan oleh sebagian pengguna.

Penerapan algoritma Naive Bayes pada data komentar YouTube berjumlah besar, yaitu sebanyak 71.300 komentar, terbukti mampu memberikan performa klasifikasi yang cukup baik. Evaluasi model menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* menunjukkan bahwa model dapat mengklasifikasikan sentimen secara konsisten pada masing-masing kelas. Selain itu, penggunaan *confusion matrix* memberikan gambaran yang jelas mengenai distribusi prediksi dan kesalahan klasifikasi antar kelas sentimen. Visualisasi *word cloud* juga membantu dalam menginterpretasikan kata-kata dominan yang muncul pada setiap kategori sentimen, sehingga memperkuat pemahaman terhadap pola opini pengguna YouTube.

Meskipun demikian, penelitian ini memiliki beberapa keterbatasan. Data yang digunakan hanya berasal dari komentar pada video *highlight* pertandingan, sehingga belum sepenuhnya merepresentasikan seluruh diskursus publik terkait Tim Nasional Indonesia di platform YouTube. Selain itu, proses pelabelan sentimen masih bersifat umum dan belum mempertimbangkan aspek emosi yang lebih kompleks atau konteks kalimat secara mendalam. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan sumber data yang lebih beragam, mempertimbangkan pendekatan pelabelan yang lebih rinci, serta membandingkan performa Naive Bayes dengan algoritma lain agar diperoleh hasil analisis sentimen yang lebih komprehensif dan akurat.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada seluruh pihak yang telah memberikan dukungan dalam pelaksanaan penelitian ini, baik dalam bentuk fasilitas, data, maupun arahan akademik. Ucapan terima kasih juga disampaikan kepada dosen pembimbing yang telah memberikan bimbingan dan masukan selama proses penyusunan penelitian. Artikel ini merupakan bagian dari skripsi yang disusun sebagai salah satu syarat untuk menyelesaikan studi pada jenjang sarjana.

DAFTAR REFERENSI

- Amrullah, F., & Solichin, A. (2024). Analisis Emosi Pada Live Chat Youtube ' Mata Najwa : 3 Bacapres Bicara Gagasan ' Menggunakan Pendekatan Lexicon dan Algoritma Naive Bayes. 12, 121–128.
- Angdresey, A., Sitanayah, L., & Tangka, I. L. H. (2025). Sentiment Analysis for Political Debates on YouTube Comments using BERT Labeling, Random Oversampling, and Multinomial Naïve Bayes. *Journal of Computing Theories and Applications*, 2(3), 342–354. <https://doi.org/10.62411/jcta.11668>
- Ansyah, M. F., & , Abd. Ghofur, L. F. L. (2024). Analisis Sentimen Komentar Youtube Terhadap Tayangan #Terbaru! Temuan dan Masalah Ahlak Di Ponpes Al-zaytun Menggunakan Metode Naïve Bayes. 8(2), 847–856.
- Apryani, I., Fauzi, A., Kusumaningrum, D. S., & Handayani, H. H. (2025). Analisis Sentimen Performa Timnas Sepak Bola Indonesia pada Kolom Komentar Aplikasi TikTok Menggunakan Algoritma Machine Learning. 8(1), 76–89.
- Ardiansyah, A., Agustina, C., Maryani, I., & Pribadi, D. (2025). Analisis Sentimen pada Komentar YouTube terkait Pembahasan eSIM Menggunakan Metode Naive Bayes dan Random Forest. 11(1), 7–14.
- Armaeni, P. P., Arya, I. K., Wiguna, G., Gede, W., & Parwita, S. (2024). Sentiment Analysis of YouTube Comments on the Closure of TikTok Shop Using Naïve Bayes and Decision Tree Method Comparison. 1(2), 70–80.
- Fatah, Z., & Ismardani, M. T. (2025). Sentiment Analysis of YouTube User Comments on Government Policies Using the Naïve Bayes Method. 3(December), 113–123.
- Fatah, Z., & Syarifah, L. (2025). Analisis Sentimen Komentar YouTube pada Video Terkait Insiden Pengemudi Ojek Online dan Anggota Brimob Menggunakan Algoritma Naive Bayes. 4, 90–96.
- Firdiawati, & , Salvia Nabillah Syifa, H. A. R. (2026). Analisis Sentimen Komentar Youtube Rencana Pemerintah terhadap Pemblokiran Roblox Menggunakan Model Naïve Bayes. 1, 1–11.
- Firsttama, R. A., Arifiyanti, A. A., & Kartika, D. S. Y. (2024). Analisis Sentimen Komentar Youtube Konferensi Tingkat Tinggi G20 Menggunakan Metode Naive Bayes. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 6(2), 282–285. <https://doi.org/10.47233/jteksis.v6i2.1263>
- Hidayat, T., & Bahtiar, A. (2025). Classification of Youtube User sentiment on 5G Technology Videos with Naïve Bayes. 4(3).

- Khomsah, S., Studi, P., Data, S., Tengah, J., & Artikel, I. (2021). Sentiment Analysis On YouTube Comments Using Word2Vec and Random Forest. 18(1), 61–72. <https://doi.org/10.31515/telematika.v18i1.4493>
- Mahfudza, N., & Ikhsan, M. (2025). Sentiment Analysis of Youtube Comments on Indonesian Presidential Candidates in 2024 using N aïve Bayes Classifier Method. 12(2), 140–148. <https://doi.org/10.30865/jurikom.v12i2.8538>
- Minaee, S. (2021). Deep Learning Based Text Classification: A Comprehensive Review. 1(1), 1–43.
- Permana, T. D. (2025). PERBANDINGAN PERFORMA ALGORITMA NAIVE BAYES DAN TERHADAP INDUSTRI ESPORTS DI INDONESIA. 2(6), 1391–1399.
- Prasetyo, V. P., & Cimahi, K. (2025). Platform Mengenai Pro Kontra Terhadap Naturalisasi Indonesia. 13(3).
- Susanto, A., & Octavia, I. V. (2025). Analisis Sentimen Komentar YouTube terhadap Video “ Purbaya Effect : Pertaruhan Ekonomi Indonesia ” sebagai Cerminan Persepsi Publik Tahun 2025. 4(2), 185–191.
- Triana Putri, Siti Nurhaliza, D. V. (2025). Sentiment Analysis of Beauty Product Applications using the Naïve Bayes Method. SinkrOn, 8(2), 980–989. <https://doi.org/10.33395/sinkron.v8i2.12303>
- Utari, E. L., Wibowo, S. H., Informatika, T., Bengkulu, U. M., Bayes, N., & Short-, L. (2025). ANALISIS KOMPARATIF ALGORITMA SVM NAIVE BAYES DAN LSTM PADA SENTIMEN KOMENTAR LAGU LABOUR. 1276–1286.
- Yusrana, R. D., Josepto, G. B., Donald, M., Aronggear, R., & Handrianus, V. (2024). ANALISIS SENTIMENT KOMENTAR VIDEO YOUTUBE “ EPIC RAP BATTLE OF PRESIDENCY 2024 ” MENGGUNAKAN ALGORITMA NAÏVE BAYES & SVM Computer Science | Industrial Engineering | Mechanic Engineering | Civil Engineering J-ENSISTEC (Journal of Engineering and Sustainable Technology) Computer Science | Industrial Engineering | Mechanic Engineering | Civil Engineering. 10(02), 1064–1069.