



Analisis Sentimen X: Kegagalan Timnas ke Piala Dunia 2026 dengan Naive Bayes

Aditya Abdullah Masykur^{1*}, Rino Raihan Gumilang², Harun Al Rosyid³

^{1,2,3}Pendidikan Teknologi Informasi, Universitas Negeri Surabaya, Indonesia

Email author:

adityaabdulloh.22004@mhs.unesa.ac.id, rinoraihan.22025@mhs.unesa.ac.id, harusrosyid@unesa.ac.id

Article Info

Article history:

Received September 3, 2025

Revised Desember 3, 2025

Accepted Desember 26, 2025

Keywords:

Sentimen Analisis

Timnas Indonesia

Multinomial Naive Bayes

SMOTE

Platform X

ABSTRACT

The performance of the Indonesian National Team (Timnas) in the 2026 World Cup qualifications has triggered massive and diverse responses on social media, particularly on platform X. This study aims to identify and classify public sentiment regarding Timnas Indonesia's performance into positive, negative, and neutral categories using a data mining approach. Text data was processed through pre-processing stages, term weighting using TF-IDF, and the application of the Synthetic Minority Over-sampling Technique (SMOTE) to address significant class distribution imbalance. The classification algorithm employed was Multinomial Naive Bayes. Model performance evaluation was conducted by comparing two training-testing data split scenarios: 90:10 and 80:20 ratios. The results indicate that public opinion is dominated by negative sentiment at 73.2%, reflecting public disappointment. In terms of model performance, the 90:10 ratio scenario yielded the best accuracy of 80%, outperforming the 80:20 ratio which recorded an accuracy of 75%. These findings demonstrate that combining Multinomial Naive Bayes with the SMOTE technique is effective in handling imbalanced text data and is capable of accurately mapping public perception.

Corresponding Author:

Aditya Abdullah Masykur,

Universitas Negeri Surabaya

Jl. Ketintang, Ketintang, Kec. Gayungan, Kota Surabaya

Email: adityaabdulloh.22004@mhs.unesa.ac.id



1. INTRODUCTION

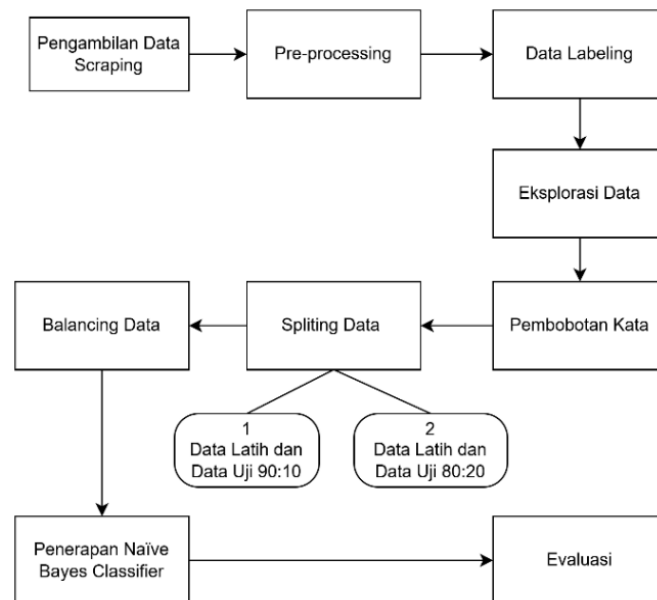
Sepak bola merupakan cabang olahraga yang memiliki popularitas dan pengaruh kultural yang masif di Indonesia. Dengan basis penggemar yang sangat besar, setiap perkembangan, terutama yang melibatkan Tim Nasional (Timnas), selalu menjadi topik perbincangan utama dan memicu reaksi publik yang kuat (Imaddudin et al., 2025). Perjalanan Timnas Indonesia dalam upaya kualifikasi menuju Piala Dunia FIFA 2026 merupakan momen penting yang menyita perhatian. Kegagalan untuk melaju ke tahap berikutnya dalam kualifikasi seringkali menimbulkan gelombang opini, kritik, dan dukungan di ruang publik, yang kini sebagian besar terwujud di platform media sosial (Apryani et al., 2025). Media sosial, khususnya X (sebelumnya Twitter), telah bertransformasi menjadi arena utama dalam ekspresi pandangan dan sentimen masyarakat. Platform ini menjadi sumber data yang kaya dan real-time untuk

menganalisis reaksi instan pengguna terhadap suatu peristiwa karena kecepatan penyebaran informasi dan tingginya volume data yang dihasilkan (Zuhrival, 2024). Respons masif di media sosial, khususnya pada platform X. Fenomena ini sejalan dengan konteks lain di mana platform digital digunakan untuk menilai keterlibatan dan persepsi publik melalui pendekatan analisis sentimen (Nabiem et al., 2024). Reaksi pengguna X terhadap kegagalan Timnas Indonesia penting untuk diteliti karena merefleksikan tingkat kepuasan, kekecewaan, hingga harapan publik terhadap kinerja tim dan federasi. Penelitian ini menggunakan pendekatan data mining. Pendekatan ini relevan sebagaimana diterapkan dalam studi pembelajaran daring, di mana efektivitas dianalisis melalui persepsi digital para pengguna (Yue, 2025). Analisis semacam ini memberikan wawasan terukur mengenai penerimaan publik terhadap prestasi olahraga nasional (Syahira & R, 2024).

Untuk memahami pandangan publik secara terstruktur, diperlukan pendekatan Analisis Sentimen (Sentiment Analysis). Dalam ranah data mining, terdapat metode yang dikenal sebagai analisis sentimen. Metode ini bertujuan untuk mengekstraksi pandangan atau emosi subjektif yang terkandung dalam data teks, yang kemudian dikelompokkan ke dalam tiga polaritas utama: positif, negatif, atau netral. (Zuhrival, 2024). Klasifikasi otomatis diperlukan untuk memproses volume besar data teks yang tidak terstruktur. Metode Machine Learning yang efektif dan efisien dibutuhkan untuk klasifikasi teks berskala besar. Dalam ranah pengolahan data, Naive Bayes Classifier tergolong sebagai algoritma andalan yang secara luas dimanfaatkan untuk keperluan pengklasifikasian teks karena kemudahannya, kecepatan komputasi, dan efektivitasnya. Algoritma ini berfungsi berdasarkan prinsip probabilitas Bayes dan telah terbukti efektif dalam studi kasus analisis sentimen yang spesifik terhadap kemampuan dan hasil Timnas senior Sepak Bola Indonesia di platform X (Syahira & R, 2024).

2. METHOD

Metode penelitian yang digunakan mengikuti beberapa tahapan sistematis yang umum diterapkan dalam analisis sentimen berbasis Machine Learning. Langkah – langkah yang dilakukan pada penelitian ini dimulai dengan pengumpulan data (*scraping*), diikuti dengan *pre-processing*, pelabelan sentimen, pembobotan fitur, pembagian data, serta penerapan algoritma klasifikasi Naive Bayes Classifier. Tahapan akhir meliputi evaluasi model menggunakan Confusion Matrix untuk mengukur Akurasi, Presisi, Recall, dan F1-Score.



Gambar 1. Alur Analisis Sentimen

2.1 Pengumpulan Data (*Data Scraping*)

Pengumpulan data (*Data Scraping*) adalah metode pengambilan data secara otomatis dari sebuah platform untuk memperoleh informasi dalam jumlah besar dan dapat diolah kembali. Data dikumpulkan dari platform X (sebelumnya Twitter), yang dikenal memiliki tingkat pembaruan yang tinggi,

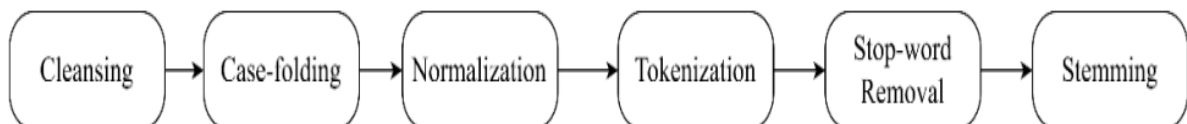
menjadikannya sumber efektif untuk analisis sentimen. Sebanyak 1.438 *tweet* dikumpulkan sebagai data utama untuk keperluan analisis pada penelitian ini yang diperoleh menggunakan *keyword* (“Timnas Indonesia”, “Gagal”, “Kualifikasi Piala Dunia 2026”, dll.) pada periode puncak perbincangan masyarakat. Pengambilan data umumnya dilaksanakan menggunakan bahasa pemrograman Python melalui API Twitter.

2.2 Pelabelan Data (*Data Labeling*)

Tahap pelabelan data dilaksanakan untuk mengelompokkan data ke dalam kategori sentimen yang akan diprediksi oleh model (*Supervised Learning*). Data yang sudah dikumpulkan dan dibersihkan diberi label berdasarkan polaritasnya, yaitu Positif, Negatif, atau Netral. Model RoBERTa (*Robustly optimized BERT approach*) dimanfaatkan dalam studi ini untuk membantu proses pemberian label secara semi-otomati yang telah dilatih sebelumnya (*pre-trained*). RoBERTa digunakan untuk menghasilkan prediksi label awal (*pseudo-labeling*) dengan memanfaatkan kemampuannya memahami konteks kalimat secara mendalam. Hasil prediksi tersebut kemudian divalidasi dan diverifikasi secara manual oleh anotator untuk memastikan kebenaran klasifikasi. Dalam proses validasi ini, data yang mengandung konteks kekecewaan, kritik, atau keluhan (misalnya 'gagal', 'kecewa', 'tidak lolos') dipastikan berlabel Negatif. Data yang memuat narasi dukungan, pujian, atau optimisme (misalnya 'semangat', 'dukung', 'bangga') dikonfirmasi sebagai Positif. Sementara itu, *tweet* yang dideteksi hanya berisi informasi objektif, fakta, atau pertanyaan (misalnya 'info skor', 'kapan main') ditetapkan sebagai label Netral

2.3 Pre-Processing

Preprocessing adalah serangkaian langkah yang dilakukan pada teks mentah untuk membersihkan, memformat, dan mengubah teks menjadi bentuk yang lebih terstruktur dan siap analisis. *Text Preprocessing* penting untuk menghilangkan *noise* dan meningkatkan kualitas data, sehingga model dapat bekerja lebih baik dan menghasilkan *output* yang lebih akurat, Sebagaimana diilustrasikan pada Gambar 2, penelitian ini menerapkan rangkaian prosedur pra-pemrosesan yang sistematis, meliputi tahapan cleansing, case-folding, normalization, tokenization, stop-word removal, dan stemming.



Gambar 2. Tahapan Pre-Processing data

- Cleansing:** Tahap awal ini yaitu untuk menghilangkan *noise* (derau) dan elemen pengganggu dalam *dataset*. Meliputi penghapusan tautan (URL), *mention* pengguna (@), *hashtag* (#), simbol *retweet*, angka/digit, karakter non-alfanumerik, tanda baca (koma, titik, tanda tanya), emotikon/emoji, dan tag HTML (Hakim & Sugiyono, 2024).
- Case Folding:** Fase ini berfungsi untuk menyeragamkan format teks dengan cara mengonversi seluruh karakter menjadi huruf kecil (*lowercase*). Hal ini digunakan untuk konsistensi data, sehingga kata dengan kapitalisasi berbeda (misalnya 'Indonesia' dan 'indonesia') diperlakukan sebagai entitas yang sama oleh sistem (Mubaroroh et al., 2022).
- Normalization** dilakukan dengan menyelaraskan kata-kata yang tidak baku, termasuk bentuk singkatan, kata gaul (*slang*), atau kata yang diperpanjang, agar kembali ke bentuk standar sesuai Kamus Besar Bahasa Indonesia (KBBI) (Fajriah, 2022).
- Tokenization:** Pada fase ini, Tahapan ini melibatkan penguraian struktur teks atau dokumen menjadi elemen-elemen kata yang berdiri sendiri, yang dikenal dengan istilah token (Prasetyo & Fitriani, 2024).

- e. Stop-word Removal: Langkah ini bertujuan untuk mereduksi dimensi data dengan menghapus kata-kata umum (*common words*) yang memiliki frekuensi kemunculan tinggi namun minim muatan informasi sentimen (misalnya kata hubung seperti 'dan', 'yang', 'di', 'adalah' (Misrun et al., 2024).
- f. Stemming: Tahap akhir pra-pemrosesan untuk mereduksi kata berimbuhan menjadi kata dasarnya (*root word*). Dalam tahap ini menghilangkan semua kata imbuhan, baik awalan kata, sisipan kata, serta akhiran kata, guna meminimalkan variasi morfologi kata (Surbakti et al., 2021).

Setelah seluruh tahapan *pre-processing* selesai dilakukan, hasil stemming digabungkan kembali sehingga membentuk dataset baru yang siap digunakan pada proses pemodelan atau analisis berikutnya. Contoh hasil pre-processing data disajikan pada Tabel dibawah ini.

Tabel 1. Tahap Pre-Processing

Proses	Hasil
Data awal	PSSI resmi mengakhiri kerja sama lebih awal dengan Patrick Kluivert sebagai pelatih timnas Indonesia pada Kamis, menyusul keagalannya
Cleansing	PSSI resmi mengakhiri kerja sama lebih awal dengan Patrick Kluivert sebagai pelatih timnas Indonesia pada Kamis menyusul keagalannya membawa Indonesia ke Piala Dunia meski sebelumnya Kluivert dikontrak selama dua tahun per Januari
Case-folding	pssi resmi mengakhiri kerja sama lebih awal dengan patrick kluivert sebagai pelatih timnas indonesia pada kamis menyusul keagalannya membawa indonesia ke piala dunia meski sebelumnya kluivert dikontrak selama dua tahun per januari
Normalization	pssi resmi mengakhiri kerja sama lebih awal patrick kluivert pelatih timnas indonesia kamis menyusul keagalannya membawa indonesia piala dunia meski sebelumnya kluivert dikontrak selama tahun per januari
Tokenization	['pssi', 'resmi', 'mengakhiri', 'kerja', 'sama', 'lebih', 'awal', 'dengan', 'patrick', 'kluivert', 'sebagai', 'pelatih', 'timnas', 'indonesia', 'pada', 'kamis', 'menyusul', 'keagalannya', 'membawa', 'indonesia', 'ke', 'piala', 'dunia', 'meski', 'sebelumnya', 'kluivert', 'dikontrak', 'selama', 'dua', 'tahun', 'per', 'januari']
Stop-word Removal	pssi resmi akhir kerja sama lebih awal patrick kluivert latih timnas indonesia kamis susul gagal bawa indonesia piala dunia meski belum kluivert kontrak lama tahun per januari
Stemming	pssi resmi akhir kerja sama lebih awal patrick kluivert latih timnas indonesia kamis susul gagal bawa indonesia piala dunia meski belum kluivert kontrak lama tahun per januari
Data akhir	pssi resmi akhir kerja sama lebih awal patrick kluivert latih timnas indonesia kamis susul gagal bawa indonesia piala dunia meski belum kluivert kontrak lama tahun per januari

2.4 Data Labeling

Dataset awal yang terdiri dari 1.438 data melalui proses pembersihan (cleaning) dan penyaringan ketat untuk menghapus konten spam, promosi judi, serta topik yang tidak relevan, dimana terdeteksi 400 data sampah sehingga menghasilkan 1.038 data bersih yang siap dianalisis. Pada tahap berikutnya dilakukan proses pelabelan data menggunakan pendekatan logika kata kunci (keyword labeling). Metode ini memberikan label sentimen pada setiap tweet berdasarkan skor kemunculan kata kunci tertentu, sehingga data dapat dikelompokkan menjadi 3 (tiga) kategori sentimen, yaitu negatif, netral, dan positif. Langkah awal yang dilakukan adalah menyusun daftar kata kunci yang merepresentasikan sentimen negatif dan positif sebagai dasar penentuan label. Sebagai contoh, kata-kata seperti 'gagal', 'kalah', 'kecewa', 'sedih', 'buruk', 'malu', 'pecat', dan 'hancur' digolongkan sebagai sentimen negatif. Sementara itu, kata seperti 'bangga', 'semangat', 'dukung', 'terima kasih', 'keren', 'optimis', dan 'apresiasi' termasuk dalam kategori sentimen positif. Adapun untuk kategori netral, label ini diberikan pada tweet yang

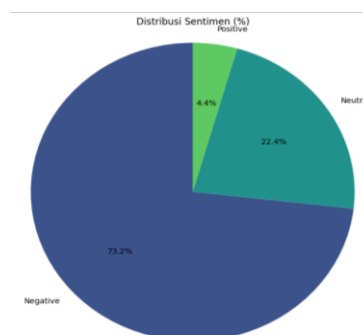
bersifat objektif atau tidak memuat dominasi kata kunci emosional, yang ditandai dengan kemunculan kata-kata seperti 'resmi', 'tanding', 'kualifikasi', 'main', 'poin', dan 'jadwal'. Mekanisme pelabelan dilakukan dengan menghitung skor; jika skor akhir lebih dari 0 maka tweet dilabeli positif, jika kurang dari 0 dilabeli negatif, dan jika bernilai 0 dikategorikan sebagai netral. Meskipun metode keyword labeling ini cepat dan mudah diterapkan, tetapi pendekatan ini memiliki kekurangan, dimana ketika menghadapi teks yang ambigu, seperti sarkasme atau sindiran halus, kata-kata yang positif dianggap menjadi teks yang bermakna negatif dan sebaliknya. Setelah proses labeling selesai, pemeriksaan ulang dilakukan secara manual dan hasil akhir distribusi jumlah data di setiap kelas sentimen yang sangat didominasi oleh respons negatif, dengan 759 data (73,2%) diberi label negatif, 233 data (22,4%) diberi label netral, dan 46 data (4,4%) diberi label positif.

Tabel 2. Pembagian Data Sentimen

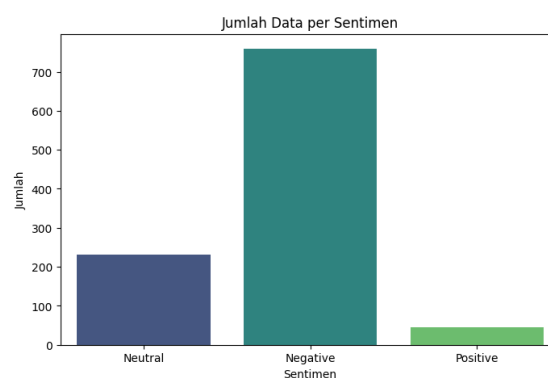
Kelas	Jumlah Data
Negatif	759
Positif	233
Netral	46

2.5 Data Exploration

Tahap eksplorasi data dilakukan untuk memperoleh pemahaman awal mengenai karakteristik *dataset* pada masing-masing kategori sentimen. Pada penelitian ini, eksplorasi dilakukan dengan menampilkan distribusi jumlah sentimen, visualisasi *Word Cloud*, serta analisis volume *tweet* harian. Berbagai visualisasi tersebut membantu mengidentifikasi pola penggunaan kata dan persebaran sentimen sehingga dapat menunjang proses analisis pada tahap berikutnya. Pada Gambar 3 dan 4 ditampilkan diagram batang atau *Bar Chart* dan diagram pie atau *Pie Chart* dengan komposisi dan total data pada setiap kelas sentimen.

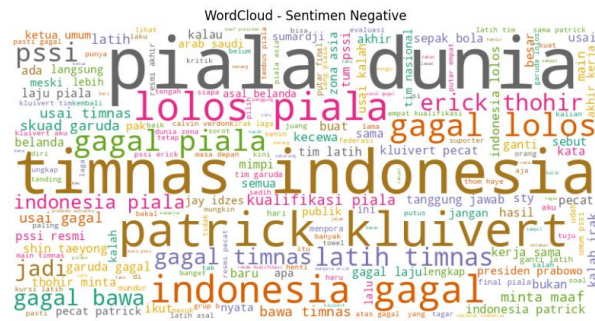


Gambar 3. Pie Chart Kelas Sentimen

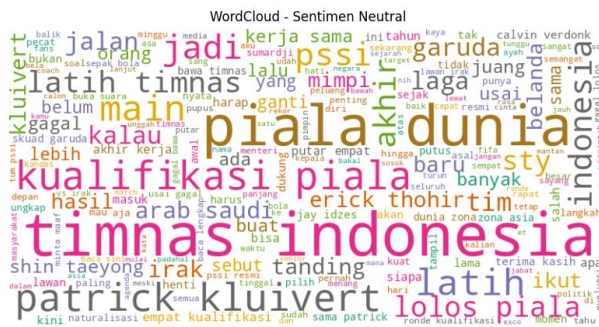


Gambar 4. Diagram Batang Data Sentimen

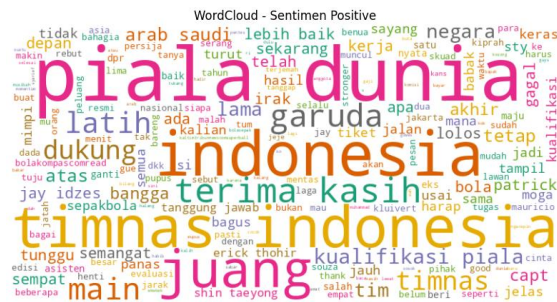
Gambar 4, 5, dan 6 menunjukkan *Word Cloud* untuk setiap kelas sentimen. Pada sentimen bermuatan negatif, kata-kata yang muncul mencerminkan respons pesimis dan kekecewaan publik terhadap pergerakan Timnas, didominasi oleh kata-kata kunci seperti GAGAL, TIMNAS, PIALA DUNIA, PATRICK KLUIVERT, dan LATIHAN/LATIH. Sebaliknya, sentimen netral banyak memuat kata informatif dan faktual, seperti TIMNAS, INDONESIA, PIALA DUNIA, KUALIFIKASI, LATIH, PSSI, dan ERIK THOHIR. Sementara itu, pada sentimen positif, kata-kata yang menonjol mencerminkan dukungan dan apresiasi, seperti PIALA DUNIA, DUKUNG, BAIK, JUANG/PERJUANGAN, dan TERIMA KASIH.



Gambar 5. Word Cloud Sentimen Negatif



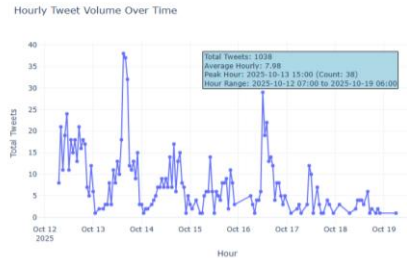
Gambar 6. Word Cloud Sentimen Netral



Gambar 7. Word Cloud Sentimen Positif

Gambar 8 *Hourly Tweet Volume Over Time* Grafik ini memperlihatkan total volume *tweet* yang berhasil dikumpulkan dalam rentang waktu 2025-10-12 07:00 hingga 2025-10-19 06:00 adalah sebanyak 1038 *tweet*. Rata-rata *tweet* yang muncul per jam adalah 7.98. Puncak pembicaraan (*Peak Hour*) terjadi pada tanggal 2025-10-13 pukul 15:00, dengan jumlah *tweet* mencapai 38. Peningkatan tajam volume *tweet*

ini mengindikasikan adanya peristiwa kunci atau pengumuman yang memicu reaksi instan dan masif dari publik pada momen tersebut.



Gambar 8 Hourly Tweet Volume Over Time

2.6 Pembobotan Kata

Metode *Term Frequency-Inverse Document Frequency* (TF-IDF) diterapkan untuk memetakan data teks menjadi format angka agar dapat diproses lebih lanjut. Pendekatan ini akan memberikan pembobotan pada tiap - tiap istilah menurut frekuensi kemunculannya dalam dokumen spesifik, yang kemudian dinormalisasi terhadap frekuensinya di seluruh korpus data. Mekanisme ini memungkinkan model untuk mengidentifikasi signifikansi setiap kata, di mana istilah yang relevan secara kontekstual akan memperoleh bobot lebih tinggi dibandingkan kata-kata umum yang muncul secara masif.

Formulasi matematis pembobotan ini ditunjukkan pada Persamaan 2:

$$W_{t,d} = tf_{t,d} \times \log \frac{N}{df_t} \quad (1)$$

Keterangan:

$w_{t,d}$ = bobot TF-IDF

$tf_{t,d}$ = banyaknya t dalam dokumen d

N = frekuensi dokumen

df_t = frekuensi dokumen yang mengandung kata t

Pada tahap perhitungan bobot, proses TF-IDF diimplementasikan menggunakan `TfidfVectorizer` dari *library scikit-learn*. Modul ini menyediakan berbagai pengaturan, seperti normalisasi dan pembatasan jumlah fitur, sehingga hasil pembobotan dapat disesuaikan dengan kebutuhan analisis. Karena komponen IDF menurunkan nilai kata yang terlalu sering muncul pada banyak dokumen, metode ini efektif untuk mengurangi dominasi kata-kata umum dan meningkatkan fokus pada kata yang lebih representatif. Hasil TF-IDF adalah sebagai berikut :

--- Hasil TF-IDF ---

	Metrik	Nilai
0	Shape matriks TF-IDF (Dokumen x Fitur)	(1037, 2724)
1	Jumlah elemen non-nol	18182
2	Rasio kepadatan (density ratio)	0.0064

Gambar 9. Hasil TF-IDF

2.7 Data Splitting

Dalam pengembangan model machine learning, data splitting adalah langkah fundamental untuk mencegah bias evaluasi. Jika sebuah model diuji menggunakan data yang sama dengan yang

dipelajarinya, model akan mengalami suatu kondisi di mana model hanya "menghafal" jawaban tanpa memahami pola sebenarnya (overfitting). Oleh karena itu, dataset harus dipisahkan menjadi training set (data latih) untuk pembelajaran dan testing set (data uji) untuk evaluasi performa. Pemilihan rasio pembagian data sangat krusial. Rasio data latih yang terlalu kecil membuat model kurang belajar, sedangkan rasio data uji yang terlalu kecil membuat evaluasi akurasi menjadi tidak representatif (Misrun et al., 2024).

2.8 Data Balancing

Salah satu tantangan terbesar dalam analisis sentimen publik, seperti kasus Timnas Indonesia dalam penelitian ini adalah ketidakseimbangan kelas (class imbalance). Berdasarkan visualisasi diagram lingkaran pada Gambar 3, sentimen negatif mendominasi secara ekstrem sebesar 73.2% (759 data), sementara sentimen positif hanya 4.4% (46 data). Ketimpangan ini sangat berbahaya karena algoritma akan cenderung mengabaikan kelas minoritas (positif) dan hanya memprediksi kelas mayoritas (negatif) demi mencapai akurasi semu yang tinggi. Model yang dilatih pada data tidak seimbang akan memiliki nilai recall yang sangat buruk pada kelas minoritas (Surbakti et al., 2021). Guna menangani kendala tersebut, metode *Synthetic Minority Over-sampling Technique* (SMOTE) diterapkan dengan memanfaatkan pustaka *imblearn*. Berbeda dengan metode random oversampling yang hanya menduplikasi data, SMOTE bekerja dengan cara mensintesis data baru. Algoritma ini memilih data dari kelas minoritas (Positif dan Netral), mencari tetangga terdekatnya (k-nearest neighbors) dalam ruang fitur, dan menciptakan sampel buatan di antara titik-titik data tersebut.

2.9 Penerapan Algoritma Naïve Bayes

Dengan berlandaskan pada Teorema Bayes dan asumsi bahwa setiap fitur bersifat saling bebas (independen), algoritma Naïve Bayes bekerja sebagai pengklasifikasi probabilistik yang efisien dalam mengestimasi peluang suatu data masuk ke dalam kategori tertentu. Dalam domain analisis sentimen, penelitian ini secara spesifik menerapkan varian Multinomial Naïve Bayes. Varian ini diadopsi karena keandalannya saat menangani fitur diskrit berupa bobot kata yang dihasilkan dari ekstraksi TF-IDF (Term Frequency-Inverse Document Frequency) (Trihardianingsih & Permatasari, 2025). Cara kerja dasar algoritma ini adalah menentukan label sentimen (c) yang memiliki probabilitas posterior tertinggi untuk sebuah dokumen (d), sebagaimana direpresentasikan secara matematis dalam Persamaan (3):

$$c_{map} = \operatorname{arg\,arg}[P(c) \cdot \prod_{t \in d} P(c)] \quad (2)$$

Untuk menyelesaikan persamaan di atas, algoritma perlu menghitung peluang kemunculan setiap fitur kata (t) terhadap masing-masing kelas sentimen. Mengingat karakteristik data teks yang sering kali memiliki nilai nol (*sparsity*), perhitungan ini melibatkan teknik *Laplace Smoothing* (α) guna menghindari probabilitas nol pada kata yang tidak muncul dalam data latih. Estimasi probabilitas kondisional (*likelihood*) tersebut dirumuskan dalam Persamaan (2):

$$P(c) = \frac{N_{tc} + \alpha}{N_c + \alpha \cdot |V|} \quad (3)$$

Meskipun algoritma ini andal, kinerja klasifikasi pada data riil sering kali terganggu oleh dominasi kelas mayoritas. Karena hal itu, penelitian yang akan dilakukan ini mengintegrasikan pendekatan SMOTE (*Synthetic Minority Over-sampling Technique*) untuk menyeimbangkan distribusi data sebelum proses perhitungan probabilitas di atas dilakukan, guna memastikan presisi yang optimal pada seluruh kategori sentimen.

2.10 Evaluasi

Pada tahap akhir, validasi performa model dilakukan melalui instrumen *confusion matrix*. Teknik ini berfungsi sebagai tabel evaluasi untuk menilai efektivitas algoritma klasifikasi (seperti Naive Bayes) dengan cara menyandingkan hasil prediksi sistem melawan data aktual (*ground truth*) (Rezzi, 2025). Dengan menggunakan tabel ini, peneliti tidak hanya mengetahui berapa banyak data yang diprediksi benar, tetapi juga dapat melihat secara detail jenis kesalahan yang dilakukan oleh model, apakah model cenderung salah memprediksi kelas positif sebagai negatif atau sebaliknya.

Untuk menghitung nilai evaluasi model, dapat menggunakan rumus perhitungan *accuracy*, *precision*, *recall* dan *f1-score* (Alfiyani et al., 2025).

Metrik pertama ialah Akurasi (Accuracy), yang berfungsi mengukur rasio total prediksi yang benar (baik positif maupun negatif) terhadap keseluruhan data. Rumus perhitungan akurasi adalah sebagai berikut :

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Selanjutnya, evaluasi difokuskan pada Presisi (Precision) untuk mengukur tingkat ketepatan prediksi pada kelas positif. Semakin tinggi nilai presisi, semakin sedikit kesalahan *False Positive* yang dihasilkan oleh model. Rumus perhitungan *precision* terdapat pada persamaan berikut :

$$precision = \frac{TP}{TP + FP} \quad (5)$$

Untuk melengkapi presisi, digunakan metrik *Recall* (Sensitivitas) yang mengukur keberhasilan model dalam menemukan kembali informasi positif yang sebenarnya ada dalam dataset (Akbar & Irfan, 2025). Tingginya nilai recall mengindikasikan minimnya kesalahan *False Negative*. Rumus perhitungan *recall* adalah sebagai berikut :

$$recall = \frac{TP}{TP + FN} \quad (6)$$

F1-Score, yang didefinisikan sebagai *harmonic mean* dari presisi dan *recall*, menjadi indikator evaluasi krusial saat menghadapi data yang tidak seimbang. Penggunaan metrik ini bertujuan untuk menjembatani celah antara presisi dan *recall* (Hartono et al., 2023). Perhitungan nilainya mengacu pada persamaan di bawah ini:

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

Penerapan serangkaian metrik evaluasi ini memungkinkan analisis kinerja model yang lebih komprehensif pada tiap kategori sentimen. Temuan yang diperoleh tidak hanya menyoroti keunggulan algoritma, tetapi juga mendeteksi kelemahan spesifik melalui pemetaan pola misklasifikasi yang terjadi.

3. RESULT DAN ANALISIS

Hasil eksperimen mengindikasikan bahwa model klasifikasi mampu mengenali pola sentimen pengguna mengenai topik Kegagalan Timnas Indonesia ke Piala Dunia 2026 dengan cukup efektif. Penerapan algoritma *Multinomial Naïve Bayes* yang didukung oleh metode *resampling* SMOTE terbukti efektif dalam memperbaiki distribusi data yang tidak seimbang dengan proporsi pembagian antara *training set* dan *testing set* diatur dalam dua variasi rasio, masing-masing sebesar 80:20 dan 90:10.

Tabel 3. Perbandingan Pengujian 2 Rasio Data

Rasio Data	Accuracy	Precision	Recall	F1-Score
80:20	0.75	0.77	0.75	0.76
90:10	0.79	0.82	0.80	0.81

Rincian performa klasifikasi untuk setiap kategori sentimen pada kedua skenario pembagian data ditampilkan dalam tabel 4 dan 5.

Tabel 4. Classification Report dengan rasio data 80 : 20

Classification Report	Precision	Recall	F1-Score	Support
Negatif	0.88	0.85	0.86	153
Netral	0.50	0.50	0.50	46
Positif	0.21	0.33	0.26	9
Accuracy			0.75	208
Macro avg	0.53	0.56	0.54	208
Weighted avg	0.77	0.75	0.76	208

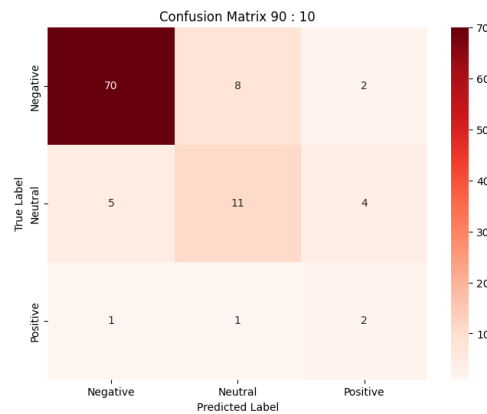
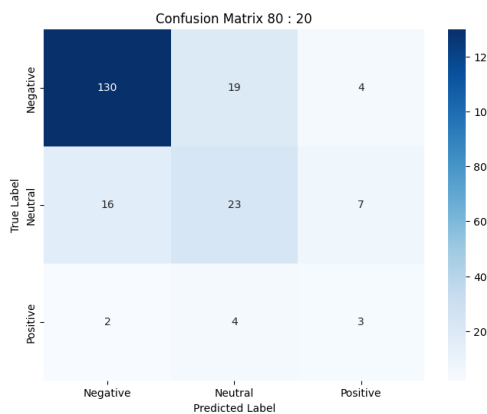
Tabel 5. Classification Report dengan rasio data 90 : 10

Classification Report	Precision	Recall	F1-Score	Support
Negatif	0.92	0.88	0.90	80
Netral	0.55	0.55	0.55	20
Positif	0.25	0.50	0.33	4
Accuracy			0.80	104
Macro avg	0.57	0.64	0.59	104
Weighted avg	0.82	0.80	0.81	104

Berdasarkan hasil evaluasi menggunakan algoritma *Naïve Bayes* dengan penerapan SMOTE di kedua skenario, rasio pembagian data 90:10 menunjukkan performa yang lebih unggul dibandingkan rasio 80:20. Model dengan rasio 90:10 mencatatkan akurasi sebesar 80%, lebih tinggi dari rasio 80:20 yang memperoleh akurasi 75%. Keunggulan ini juga tercermin pada kemampuan model dalam menangani kelas minoritas, di mana rasio 90:10 menghasilkan nilai *recall* yang lebih baik pada sentimen Positif (0.50) dan Netral (0.55) dibandingkan rasio 80:20 yang hanya mencapai 0.33 dan 0.50 untuk kelas yang sama, mengindikasikan bahwa proporsi data latih yang lebih besar (90%) memberikan informasi yang lebih kaya bagi algoritma untuk membentuk prediksi yang lebih presisi pada seluruh kategori kelas.

4. DISCUSSION/CONCLUSION

Penelitian ini berhasil mengidentifikasi dan mengklasifikasikan sentimen pengguna platform X terkait Kegagalan Timnas Indonesia ke Piala Dunia 2026 melalui pendekatan *data mining*. Masalah ketimpangan distribusi kelas yang parah pada data berhasil diredam secara efektif melalui integrasi antara algoritma *Multinomial Naïve Bayes* dan metode SMOTE. Konfigurasi rasio data latih dan uji sebesar 90:10 terbukti memberikan performa paling unggul dibandingkan skema lainnya, dengan capaian nilai akurasi 80%. Capaian ini lebih unggul daripada skenario rasio 80:20 yang menghasilkan akurasi sebesar 75%. Keunggulan performa pada rasio 90:10 mengindikasikan bahwasanya pembagian data latih yang lebih besar memberikan representasi fitur yang lebih komprehensif bagi model untuk mempelajari pola sentimen yang kompleks. Hal ini terlihat dari kemampuan model dalam mendeteksi kelas minoritas, di mana nilai *recall* untuk sentimen positif mencapai 0.50 dan sentimen netral mencapai 0.55, yang secara signifikan lebih baik dibandingkan hasil pada rasio 80:20. Temuan ini menegaskan bahwa dalam konteks dataset dengan variabilitas linguistik tinggi, maksimalisasi data pembelajaran memegang peranan krusial dalam meningkatkan presisi klasifikasi dan meminimalisir bias prediksi terhadap kelas mayoritas.

Gambar 10 . *Confusion Matrix* rasio data 90 : 10Gambar 11. *Confusion Matrix* rasio data 80 : 20

Analisis mendalam terhadap *confusion matrix* merefleksikan bahwasanya model *Multinomial Naïve Bayes* memiliki sensitivitas sangat tinggi untuk mendeteksi kelas mayoritas, yakni sentimen negatif, dengan tingkat keberhasilan prediksi (*recall*) yang konsisten di atas 85% pada kedua skenario pengujian. Meskipun demikian, pemetaan kesalahan klasifikasi (*misclassification*) masih teridentifikasi pada kelas minoritas, di mana model cenderung mengalami ambiguitas dalam membedakan sentimen Netral dan Positif, yang terlihat dari rendahnya nilai *recall* kelas Positif sebesar 33% pada rasio 80:20. Namun, peningkatan proporsi data latih pada rasio 90:10 terbukti mampu mereduksi "kebingungan" tersebut secara signifikan, meningkatkan kemampuan deteksi benar pada kelas Positif menjadi 50% dan Netral menjadi 55%, yang mengindikasikan bahwa ketersediaan data latih yang lebih masif memainkan peran vital dalam meminimalkan bias prediksi pada kategori dengan representasi data yang terbatas. Analisis distribusi sentimen mengungkapkan bahwa opini publik didominasi oleh sentimen negatif, yang merefleksikan tingginya tingkat kekecewaan dan kritik masyarakat terhadap kegagalan Timnas Indonesia. Dominasi ini ditandai dengan prevalensi kata kunci seperti "gagal", "kalah", dan "kecewa" dalam *word cloud*. Sebaliknya, proporsi sentimen netral dan positif yang lebih rendah menunjukkan bahwa diskursus yang bersifat informatif atau suportif cenderung tertutup oleh reaksi emosional yang reaktif. Secara keseluruhan, penelitian ini memberikan wawasan empiris mengenai dinamika persepsi publik terhadap sepak bola nasional, sekaligus membuktikan bahwa pendekatan *machine learning* dengan strategi penyeimbangan data yang tepat dapat menjadi instrumen yang andal untuk memetakan opini publik di media sosial.

REFERENCES

- Akbar, M. R., & Irfan, P. (2025). *Analisis Sentimen Ulasan Pengguna Aplikasi MyBCA dengan Naive Bayes, Random Forest, dan Decision Tree Sentiment Analysis of MyBCA Application User Reviews using Naive Bayes, Random Forest, and Decision Tree*. 14, 2464–2478.
- Alfiyani, W., Fatah, D. A., & Irhamni, F. (2025). *PENERAPAN ALGORITMA NAÏVE BAYES UNTUK ANALISIS SENTIMEN PADA MEDIA SOSIAL X TERHADAP PERFORMA TIM NASIONAL SEPAK BOLA INDONESIA DI ERA KEPEMIMPINAN SHIN TAE-YONG*. 9(3).
- 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.
- Fajriah, S. (2022). *Analisis sentimen terhadap body shaming pada twitter menggunakan metode Naive Bayes Classifier*. 3(2), 61–71.
- Hakim, Z. R., & Sugiyono. (2024). *Analisa Sentimen Terhadap Kereta Cepat Jakarta – Bandung*. 5(3), 939–945.
- Hartono, H., Hajjah, A., & Marlim, Y. N. (2023). *PENERAPAN METODE NAÏVE BAYES APPLICATION OF THE NAÏVE BAYES CLASSIFIER*. 12(1), 37–46.
- Imaddudin, S., Astuti, I., & Ruhama, S. (2025). *Studi Sentimen Masyarakat terhadap PSSI di Era Erick Thohir menggunakan Algoritma Support Vector Machine (SVM) pada Media Sosial X*. 1(8), 1003–1013.
- Misrun, C. A., Haerani, E., Fikry, M., & Budianita, E. (2024). *Jurnal Computer Science and Information Technology (CoSciTech) naive bayes classifier method*. 4(1), 207–215.
- Mubaroroh, H. H., Yasin, H., & Rsugiyono, A. (2022). *Analisis sentimen data ulasan aplikasi ruangguru pada situs google play menggunakan algoritma naive bayes classifier dengan normalisasi kata levenshtein distance*. 11(1), 257–266.
- Nabiem, C., Stephen, A., Sunday, A., Hilary, A., & Sidney, O. (2024). *Impact of online learning on student ' s performance and engagement: a systematic review*. *Discover Education*. <https://doi.org/10.1007/s44217-024-00253-0>
- Prasetyo, H., & Fitriani, A. S. (2024). *Procedia of Engineering and Life Science Vol . 4 June 2023 Seminar Nasional & Call Paper Fakultas Sains dan Teknologi (SENASAINS 6 th) Universitas Muhammadiyah Sidoarjo Procedia of Engineering and Life Science Vol . 4 June 2023 Seminar Nasional & Call P. 4(June 2023)*.
- Rezzi, F. (2025). *Penerapan Algoritma Multinomial Naive Bayes dengan Penyeimbangan Data SMOTE pada Kl Asifikasi Sentimen Pengguna Shopee terhadap Produk Facial Wash Kahf Farendika Rezzi Analisis Sentimen pengolahan data opini atau ulasan untuk menentukan kecenderungan sent. September*.
- Surbakti, A. Q., Hayami, R., & Al-amien, J. (2021). *Analisa tanggapan terhadap PSBB di indonesia dengan algoritma decision tree pada twitter*. 2(2), 91–97.
- Syahira, M. A., & R, R. K. (2024). *Analisis Sentimen Cyberbullying Pada Media Sosial X Menggunakan Metode Support Vector Machine*. 8, 1724–1733. <https://doi.org/10.30865/mib.v8i3.7926>
- Trihardianingsih, L., & Permatasari, H. (2025). *PENGARUH OPTIMIZER TERHADAP AKURASI KLASIFIKASI PISTACHIO MENGGUNAKAN MOBILENETV2*. 644–649.
- Yue, Q. (2025). *Online vs . face-to-face : a long-term study on the effectiveness and essence of learning*. *Cogent Education*, 12(1). <https://doi.org/10.1080/2331186X.2025.2554314>
- Zuhrival, G. E. (2024). *Analisis Sentimen Performa Tim Nasional Sepak Bola Indonesia Era Kepemimpinan Shin Tae-yong pada X Menggunakan Algoritma Naive Bayes Analisis Sentimen Performa Tim Nasional Sepak Bola Indonesia Era Kepemimpinan Shin Tae-yong pada X Menggunakan Algoritma N*.