

# Analisis Sentimen Timnas Indonesia pada Data Tidak Seimbang Menggunakan Perbandingan Naïve Bayes dan IndoBERT

Maharani Navila Salsa Bela\*, Putry Wahyu Setyaningsih

Fakultas Teknologi Informasi, Program Studi Sistem Informasi, Universitas Mercu Buana Yogyakarta, Yogyakarta, Indonesia

Email: <sup>1,\*</sup>maharaninavilasalsa@gmail.com, <sup>2</sup>putryws@mercubuana-yogya.ac.id

Email Penulis Korespondensi: maharaninavilasalsa@gmail.com

Submitted: 07/04/2026; Accepted: 26/05/2026; Published: 26/05/2026

**Abstrak**—Platform media sosial X banyak dimanfaatkan masyarakat untuk menyampaikan opini terhadap performa Tim Nasional Indonesia, khususnya pada putaran keempat Kualifikasi Piala Dunia 2026. Pada fase ini, Timnas Indonesia mengalami dua kekalahan berturut, yaitu 2–3 dari Arab Saudi dan 0–1 dari Irak, yang memicu meningkatnya respons emosional dan kritik publik di media sosial. Kondisi ini menjadikan analisis sentimen penting untuk memahami persepsi masyarakat secara lebih objektif. Penelitian ini bertujuan untuk menganalisis sentimen pengguna media sosial X serta membandingkan kinerja model Naïve Bayes dan IndoBERT pada kondisi data tidak seimbang. Data penelitian berjumlah 1.268 tweet yang diproses melalui tahap pra-pemrosesan, kemudian dilabeli secara otomatis menggunakan pendekatan lexicon-based sebagai pelabelan awal menjadi dua kelas, yaitu positif dan negatif. Dataset dibagi menjadi data latih dan data uji dengan rasio 70:30. Distribusi data menunjukkan dominasi sentimen negatif sebesar 84,1% dan positif sebesar 15,9%. Klasifikasi dilakukan menggunakan Naïve Bayes berbasis TF-IDF dan IndoBERT-base-p1, dengan penanganan ketidakseimbangan data berupa random oversampling dan class weighting. Hasil menunjukkan bahwa Naïve Bayes tanpa penanganan memperoleh akurasi 84% namun gagal mengenali kelas positif. Setelah oversampling, recall kelas positif meningkat menjadi 45%. IndoBERT menghasilkan akurasi 85%, dengan peningkatan recall positif dari 35% menjadi 43% serta F1-score kelas positif sebesar 47% setelah penerapan class weighting. Meskipun akurasi relatif tinggi, evaluasi menunjukkan pentingnya mempertimbangkan performa pada kelas minoritas. Secara keseluruhan, IndoBERT dengan class weighting memberikan hasil yang lebih seimbang. Namun, penggunaan pelabelan otomatis berbasis leksikon menjadi keterbatasan penelitian ini.

**Kata Kunci:** Analisis Sentimen; Media Sosial X; Naïve Bayes; IndoBERT; Timnas Indonesia; Ketidakseimbangan Data

**Abstract**—Social media platform X is widely used by the public to express opinions on the performance of the Indonesian National Team, especially in the fourth round of the 2026 World Cup Qualifiers. In this phase, the Indonesian National Team suffered two consecutive defeats, namely 2–3 to Saudi Arabia and 0–1 to Iraq, which triggered an increase in emotional responses and public criticism on social media. This condition makes sentiment analysis important to understand public perception more objectively. This study aims to analyze the sentiment of social media users X and compare the performance of the Naïve Bayes and IndoBERT models in imbalanced data conditions. The research data amounted to 1,268 tweets that were processed through a pre-processing stage, then automatically labeled using a lexicon-based approach as an initial labeling into two classes, namely positive and negative. The dataset was divided into training data and test data with a ratio of 70:30. The data distribution shows the dominance of negative sentiment at 84.1% and positive at 15.9%. Classification was performed using TF-IDF-based Naïve Bayes and IndoBERT-base-p1, with data imbalance management using random oversampling and class weighting. The results show that Naïve Bayes without treatment achieved 84% accuracy but failed to recognize the positive class. After oversampling, the positive class recall increased to 45%. IndoBERT achieved 85% accuracy, with positive recall increasing from 35% to 43% and the positive class F1-score increasing by 47% after applying class weighting. Despite the relatively high accuracy, the evaluation shows the importance of considering performance on minority classes. Overall, IndoBERT with class weighting provided more balanced results. However, the use of lexicon-based automatic labeling is a limitation of this study.

**Keywords:** Sentimen Analysis; Social Media X; Naïve Bayes; IndoBERT; Indonesian National Team; Data Imbalance

## 1. PENDAHULUAN

Sepak bola adalah satu dari disiplin olahraga yang memiliki daya tarik besar dan mendapat perhatian publik di seluruh dunia, termasuk di Indonesia. Performa tim nasional dalam ajang internasional sering kali menjadi fokus utama yang mencerminkan tidak hanya prestasi olahraga tetapi juga kebanggaan dan emosi publik [1]. Dalam konteks ini, keberhasilan Tim Nasional Indonesia melaju ke putaran keempat Kualifikasi Piala Dunia 2026 menjadi pencapaian yang cukup bersejarah setelah penantian panjang selama 16 tahun. Pencapaian tersebut memicu antusiasme tinggi dan mendorong masyarakat untuk secara aktif menyampaikan pandangan, dukungan, maupun kritik melalui media sosial [2].

Di antara sosial media yang kerap digunakan sebagai media untuk berinteraksi dan bertukar opini ialah Twitter atau X, yang popularitasnya konsisten mengalami peningkatan di setiap tahunnya seiring dengan melonjaknya jumlah pengguna. Di Indonesia, jumlah pengguna Twitter tercatat mencapai sekitar 23,76 juta pada tahun 2025, sehingga menghasilkan volume data yang besar dan dinamis dalam bentuk opini publik [3]. Aktivitas pengguna yang intens, khususnya dalam menanggapi hasil pertandingan Timnas Indonesia seperti kekalahan dengan selisih yang sedikit 2–3 dari Arab Saudi serta kekalahan 0–1 dari Irak pada putaran keempat, serta performa yang lebih baik pada putaran sebelumnya, menghasilkan beragam sentimen yang berkembang

secara dinamis. Kondisi ini membuka peluang untuk dilakukan analisis sentimen secara sistematis guna memahami persepsi publik terhadap performa Timnas Indonesia secara lebih objektif [4].

Analisis sentimen dengan pendekatan machine learning sudah sering dilakukan menggunakan algoritma klasik seperti Naïve Bayes [5]. Penelitian Minardi dkk. menganalisis sentimen warganet atas tragedi kanjuruhan malang dengan Naïve Bayes dengan akurasi 77,67% [6]. Reda Hayati dkk. menunjukan nilai akurasi sebesar 73,94% dengan dominasi sentimen positif terhadap Timnas Indonesia dalam AFC-23 Asian Cup menggunakan Naïve Bayes [7]. Dalam beberapa tahun terakhir, bidang NLP berkembang pesat berkat model Transformer seperti BERT yang bisa memahami konteks kata secara dua arah dengan sangat baik [8]. Teknologi inilah yang kemudian diadaptasi ke dalam IndoBERT, sebuah model yang dikembangkan khusus untuk memahami karakteristik bahasa Indonesia dengan memanfaatkan himpunan data teks lokal yang sangat luas. Hasilnya, IndoBERT terbukti jauh lebih efektif dalam menangani berbagai tugas bahasa, mulai dari memilah kategori tulisan hingga mendeteksi kecenderungan opini dalam ulasan [9]. Dalam risetnya, Ahmad Fauzi dkk. menunjukkan bahwa IndoBERT memberikan hasil yang lebih presisi dibandingkan metode konvensional untuk data ulasan e-commerce. Dengan tingkat akurasi mencapai 94,1%, IndoBERT sukses mengungguli metode SVM dan Naïve Bayes yang masing-masing hanya meraih skor 89,5% dan 84,2% [10]. Muhammad Irfan dkk. membandingkan tiga model transformer IndoBERT, IndoRoBERTa, dan DistilBERT Multilingual dalam analisis sentimen Timnas Indonesia di Piala Asia 2023 dengan akurasi terbaik 0.8897 yang dihasilkan oleh model IndoBERT mengungguli model lainnya [11]. Penelitian oleh Chandra dkk. membandingkan model CNN dan IndoBERT dalam analisis berita politik Indonesia, IndoBERT menunjukkan performa terbaik dengan tingkat akurasi sebesar 92,93%, khususnya dalam mengklasifikasikan sentimen netral. Sementara itu, CNN memperoleh akurasi 89,13%, tetapi masih mengalami kendala dalam membedakan kelas sentimen netral [12]. Dalam konteks analisis sentimen pada domain sepak bola, penggunaan model seperti IndoBERT menjadi relevan karena mampu memahami konteks bahasa Indonesia yang bersifat dinamis, termasuk istilah khas pertandingan dan ekspresi emosional pengguna. Dibandingkan dengan BERT yang bersifat umum, IndoBERT lebih sesuai karena dilatih pada korpus bahasa Indonesia, sementara model seperti IndoBERTtweet yang berfokus pada data media sosial masih memiliki keterbatasan dalam penerapan yang luas. Meskipun demikian, sebagian besar penelitian tersebut lebih berfokus pada perbandingan akurasi model, sementara aspek distribusi kelas data belum menjadi perhatian utama.

Sentimen terhadap performa Tim Nasional Indonesia pada putaran keempat Kualifikasi Piala Dunia 2026 menunjukkan kecenderungan dominasi opini negatif yang dipengaruhi oleh hasil pertandingan yang kurang memuaskan, termasuk kekalahan beruntun dari Arab Saudi dan Irak. Kondisi ini tidak hanya mencerminkan reaksi emosional publik, tetapi juga membentuk persepsi kolektif terhadap performa tim nasional. Oleh karena itu, analisis sentimen pada konteks ini menjadi penting untuk memberikan gambaran yang lebih objektif mengenai opini publik, yang dapat menjadi masukan dalam evaluasi performa tim maupun pengambilan keputusan strategis di bidang olahraga. Meskipun penelitian terkait analisis sentimen dan klasifikasi teks tidak seimbang telah banyak dikembangkan dalam bidang *Natural Language Processing* (NLP), sebagian besar studi masih berfokus pada peningkatan akurasi model tanpa mempertimbangkan kompleksitas data media sosial yang bersifat informal, dinamis, dan kaya akan penggunaan bahasa gaul. Selain itu, pendekatan yang digunakan umumnya belum secara spesifik mengkaji bagaimana ketidakseimbangan distribusi kelas memengaruhi kemampuan model dalam memahami konteks bahasa Indonesia pada domain tertentu, seperti sepak bola. Pada platform X, karakteristik teks yang singkat, tidak baku, dan penuh ekspresi emosional dapat menurunkan performa model klasifikasi, terutama dalam mendeteksi kelas minoritas.

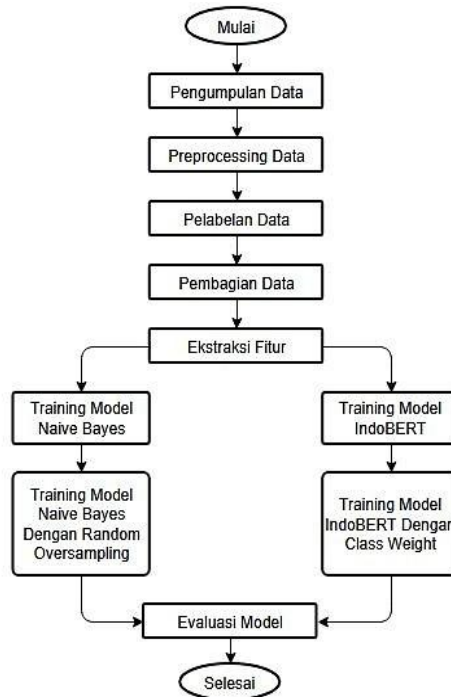
Berdasarkan permasalahan tersebut, riset ini bertujuan untuk menganalisis dan membandingkan kinerja Naïve Bayes dan IndoBERT dalam mengklasifikasikan sentimen terhadap performa Tim Nasional Indonesia pada Kualifikasi Piala Dunia 2026 dengan mempertimbangkan kondisi data yang tidak seimbang. Selain itu, riset ini juga mengevaluasi pengaruh penerapan teknik penanganan ketidakseimbangan data terhadap kemampuan model dalam mengidentifikasi sentimen positif sebagai kelas minoritas. Dengan demikian, riset ini diharapkan dapat memberikan kontribusi dalam pengembangan analisis sentimen berbasis media sosial, khususnya pada konteks bahasa informal dengan distribusi data yang tidak seimbang.

## **2. METODOLOGI PENELITIAN**

### **2.1 Tahapan Penelitian**

Riset berikut dilaksanakan melewati sejumlah prosedur yang disusun secara sistematis untuk menganalisis sentimen pengguna media sosial X terhadap performa Timnas Indonesia pada putaran keempat Kualifikasi Piala Dunia 2026. Tahapan penelitian disusun agar setiap proses pengolahan data, pemodelan, dan evaluasi dapat dilakukan secara terarah dan sistematis. Gambar 1 menunjukkan alur penelitian yang dimulai dari pengumpulan data, kemudian dilanjutkan dengan tahap preprocessing dan pelabelan sentimen. Data yang telah dilabeli selanjutnya dibagi menjadi data latih dan data uji, kemudian dilakukan ekstraksi fitur. Proses pemodelan dilakukan menggunakan dua pendekatan, yaitu Naïve Bayes dan IndoBERT, masing-masing dengan dan tanpa

penanganan ketidakseimbangan data berupa random oversampling dan class weighting. Tahap akhir adalah evaluasi model menggunakan metrik klasifikasi untuk menilai kinerja masing-masing metode.



Gambar 1. Alur Penelitian

## 2.2 Pengumpulan Data

Data penelitian dihimpun melalui platform media sosial X dengan mengekstraksi tweet berdasarkan *key word* tertentu serta rentang waktu yang telah ditentukan. Proses pengambilan data dilakukan pada dua periode, yaitu 8–9 Oktober 2025 yang bertepatan dengan pertandingan Indonesia melawan Arab Saudi, serta 11–12 Oktober 2025 saat pertandingan Indonesia menghadapi Irak. Kata kunci yang dipakai dalam tahap mengumpulkan data yaitu “*timnas*”, “*arab saudi*”, dan “*irak*”.

## 2.3 Preprocessing Data

Tahap preprocessing merupakan proses permulaan yang memiliki tujuan untuk membersihkan serta menyiapkan data teks supaya berada dalam bentuk yang lebih terstruktur sehingga dapat digunakan pada tahap analisis selanjutnya [13]. Pada penelitian ini, beberapa prosedur preprocessing yang dilaksanakan meliputi:

a. Cleaning

Pada tahap berikut, tweet yang kosong atau berulang dibuang terlebih dahulu. Setelah itu, teks dibersihkan dengan menghapus bagian-bagian yang tidak dibutuhkan, semisal URL, mention, hashtag, retweet, emoji, angka, dan punctuation yang tidak perlu.

b. Case Folding

Dalam fase case folding, seluruh huruf dalam teks yang telah melalui proses pembersihan dikonversi menjadi huruf kecil untuk menyederhanakan proses pengolahan data [14].

c. Normalization

Pada tahap normalisasi, dilakukan standarisasi ejaan kata melalui perubahan bahasa gaul, singkatan, maupun bahasa tidak baku menjadi bentuk baku sesuai kaidah bahasa Indonesia. Proses ini dilakukan dengan menggunakan kamus slang yang disusun berdasarkan referensi penelitian sebelumnya serta disesuaikan dengan karakteristik data media sosial X [15].

d. Tokenizing

Tahapan tokenisasi dilakukan dengan memecah sebuah kalimat menjadi unit-unit kecil, sehingga setiap kata dapat diproses sebagai bagian yang terpisah [16].

e. Stopword Removal

Tahapan berikut memiliki tujuan untuk mengeliminasi kosakata lazim yang sering muncul namun tidak memberi informasi penting pada makna kalimat, sehingga dapat meningkatkan efektivitas proses analisis sentimen.

f. Stemming

Stemming bertujuan mengubah kata berimbuhan menjadi bentuk dasarnya. Proses ini menyatukan berbagai variasi atau turunan kata ke dalam satu bentuk dasar yang sama.

## 2.4 Pelabelan Data

Penelitian ini menerapkan skema pelabelan sentimen otomatis dengan bantuan *Indonesia Sentiment Lexicon* (InSet), yaitu kumpulan kosakata berbahasa Indonesia yang telah diberi bobot sentimen positif dan negatif [17]. Setiap tweet dianalisis berdasarkan kemunculan kata-kata dalam leksikon tersebut untuk menentukan kecenderungan sentimennya.

Riset berikut membatasi kategori sentimen ke dalam dua klasifikasi utama, yakni positif serta negatif, tanpa menyertakan kelas netral. Hal ini didasarkan pada pertimbangan bahwa sentimen netral biasanya memiliki sifat informatif ataupun ambigu, alhasil tidak dengan jelas merepresentasikan pandangan pengguna akan topik yang dibahas [18]. Ketidakjelasan tersebut berpotensi menimbulkan inkonsistensi dalam proses pelabelan dan memengaruhi kualitas data. Maka dari itu, dengan memfokuskan analisis terhadap dua kelas sentimen utama, tahap klasifikasi dapat dilakukan secara lebih efektif dan memberikan hasil berupa penafsiran output yang lebih jelas dan anda.

## 2.5 Pembagian Data dan Ekstraksi Fitur

Data tweet yang sudah melewati prosedur pre-processing serta pelabelan kemudian diklasifikasi pada dua bagian, yakni data latih serta uji dengan rasio 70:30. Pembagian ini dilakukan secara konsisten pada seluruh model yang digunakan dalam penelitian agar proses pelatihan dan evaluasi dapat berlangsung secara seimbang. Dari total 1.268 data tweet yang diperoleh, sebanyak 887 tweet berfungsi menjadi data latih, adapun 381 tweet difungsikan menjadi data uji.

Pada metode Naïve Bayes, teks diubah menjadi representasi numerik menggunakan teknik *Term Frequency–Inverse Document Frequency* (TF-IDF). Teknik berikut memiliki fungsi untuk mengukur skala prioritas sebuah kata pada suatu dokumen relatif terhadap berkas secara komprehensif dalam korpus. Hasil TF-IDF yang dihasilkan kemudian dimanfaatkan sebagai fitur numerik dalam proses klasifikasi [19]. Berbeda dengan pendekatan tersebut, pada model IndoBERT proses ekstraksi fitur dilakukan melalui tahap tokenisasi menggunakan tokenizer berbasis IndoBERT, yaitu *indobert-base-pl*. Melalui pendekatan *WordPiece*, tokenizer milik BERT berfungsi membagi input teks ke dalam potongan sub-kata. Tujuannya agar setiap kalimat dapat dipetakan ke dalam fragmen yang lebih sederhana sehingga model lebih mudah memahami struktur bahasanya. Setiap tweet selanjutnya diubah menjadi token yang sejalan dengan kosakata model IndoBERT, serta dilengkapi dengan token spesifik seperti [CLS] serta [SEP] [20]. Untuk menjaga keseragaman panjang input, tweet dengan jumlah token lebih sedikit akan menambahkan *padding*, sedangkan tweet yang melebihi batas maksimum akan dipotong (*truncation*). Selain itu, *attention mask* dipakai untuk memilah antara token yang mengandung informasi dan token hasil *padding* selama proses pemodelan.

## 2.6 Penanganan Ketidakseimbangan Data

Ketidakseimbangan data (*imbalanced data*) merupakan kondisi saat total sampel pada satu kelas melebihi kelas yang lain. Pada konteks analisis sentimen, ketidakseimbangan sering terjadi ketika salah satu sentimen, seperti negatif atau positif, mendominasi dataset. Di antara pendekatan untuk menangani problematika tersebut adalah *oversampling*, yakni teknik penyeimbangan data melalui penambahan jumlah sampel pada kelas minoritas. Pada model Naïve Bayes, *oversampling* data dilakukan dengan menerapkan *random oversampling* pada data latih. *Random oversampling* dilakukan dengan menambahkan sampel dari kelas minoritas melalui proses *sampling* dengan pengembalian (*sampling with replacement*). Jumlah sampel tambahan dihitung berdasarkan selisih antara jumlah data kelas mayoritas dan kelas minoritas. Pendekatan berikut memiliki tujuan untuk memberi peluang yang lebih pada model dalam mempelajari karakteristik kelas minoritas. Dengan distribusi yang lebih seimbang, model seperti Naïve Bayes dapat mengestimasi probabilitas posterior secara lebih proporsional dan tidak terlalu bias terhadap kelas mayoritas. Tanpa dilakukannya penyesuaian data, model berpotensi lebih sering memprediksi kelas mayoritas karena jumlah datanya lebih dominan, sehingga menyebabkan rendahnya kapabilitas model untuk mengidentifikasi kelas minoritas dan menurunkan nilai recall pada kelas tersebut.

Selain pendekatan *oversampling*, penanganan ketidakseimbangan data dalam penelitian ini juga dilakukan menggunakan teknik *class weighting*, khususnya pada model IndoBERT. Berbeda dengan *oversampling* yang memodifikasi distribusi data dengan menggandakan sampel kelas minoritas, *class weighting* tidak mengubah jumlah data, melainkan memberikan bobot yang berbeda dalam tiap-tiap kelas selama tahapan pelatihan model. Pada model berbasis Transformer seperti IndoBERT, *class weighting* diterapkan pada fungsi *loss* (umumnya *Cross-Entropy Loss*). Bobot kelas dimasukkan sebagai parameter dalam perhitungan *loss*, sehingga selama proses backpropagation, gradien untuk kelas minoritas menjadi lebih signifikan. Hal ini memungkinkan model untuk menyesuaikan parameter secara lebih sensitif terhadap kesalahan pada kelas minoritas tanpa perlu melakukan duplikasi data. Dengan demikian, model tidak sekadar berfokus kepada kelas mayoritas, namun juga lebih memperhatikan kesalahan prediksi pada kelas minoritas, sehingga performa klasifikasi menjadi lebih seimbang.

## 2.7 Pemodelan dan Klasifikasi

Proses pemodelan dan klasifikasi dalam riset berikut dilakukan memakai dua pendekatan, yakni algoritma Naïve Bayes dan model berbasis Transformer IndoBERT. Pada model Naïve Bayes, proses klasifikasi dilakukan dengan memanfaatkan fitur teks hasil ekstraksi TF-IDF. Model yang digunakan adalah Multinomial Naïve Bayes, dengan pengaturan *hyperparameter* menggunakan nilai  $\alpha$  (alpha) untuk mengantisipasi probabilitas nol dalam kata yang jarang muncul. Nilai alpha digunakan untuk pengaturan standar yang umum diterapkan pada klasifikasi teks, sehingga model dapat menangani distribusi fitur yang bersifat jarang (*sparse*).

Sementara itu, pemodelan menggunakan IndoBERT dilakukan dengan memanfaatkan model pretrained *IndoBERT-base-p1* yang disesuaikan untuk tugas klasifikasi sentimen. Proses pelatihan dilakukan dengan beberapa pengaturan hiperparameter utama, antara lain jumlah *epoch* sebanyak lima, *learning rate* sebesar  $2^{-5}$ , *batch size* sebesar 16, serta penerapan *weight decay* untuk mengurangi risiko *overfitting*. Pada skenario penanganan ketidakseimbangan data, pembobotan kelas (*class weighting*) diterapkan pada fungsi *loss* guna memberikan penalti yang lebih besar terhadap kesalahan prediksi pada kelas minoritas. Seluruh proses pemodelan dilakukan secara terpisah untuk masing-masing skenario, baik dengan maupun tanpa penanganan ketidakseimbangan data, sehingga memungkinkan perbandingan performa model secara objektif. Rincian hyperparameter yang digunakan pada masing-masing model ditunjukkan pada Tabel 1.

**Tabel 1.** Hyperparameter

Model	Hyperparameter	Nilai
Naïve Bayes	Alpha	1.0
	Learning rate	$2e-5$
IndoBERT	Batch size	16
	Weight decay	0.01
	Epochs	5

## 2.8 Evaluasi Model

Evaluasi berikut dikerjakan untuk menganalisis kinerja algoritma Naïve Bayes dan IndoBERT dalam mengkategorisasi sentimen pengguna media sosial X terhadap performa Tim Nasional Indonesia dalam Putaran Keempat Kualifikasi Piala Dunia 2026. Proses evaluasi memiliki tujuan untuk membandingkan performa kedua model, baik sebelum maupun sesudah penerapan penanganan ketidakseimbangan data. Evaluasi dilaksanakan memakai data pengujian yang sudah dipisahkan sebelumnya agar output pengukuran mencerminkan kapabilitas generalisasi model terhadap data yang tidak pernah dilihat selama proses pelatihan. Penilaian dilaksanakan memakai empat metrik esensial, yakni *accuracy*, *precision*, *recall*, serta *F1-score*, yang umum dipakai dalam klasifikasi multi-kelas. *Accuracy* menghitung rasio prediksi yang tepat terhadap keseluruhan data pengujian, serta dirumuskan di bawah ini:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Dengan TP (*true positive*), TN (*true negative*), FP (*false positive*), dan FN (*false negative*) masing-masing menggambarkan kualitas prediksi model terhadap setiap kelas yang diuji. *Precision* digunakan untuk mengukur ketepatan prediksi suatu kelas, yaitu sebesar apa bagian dari prediksi positif yang tercakup dalam kelas yang dimaksud. Rumus *precision* ditunjukkan pada Persamaan berikut:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Sementara itu, *recall* digunakan untuk menilai sensitivitas model, yaitu seberapa banyak data aktual pada sebuah kelas yang berhasil dikenali dengan benar. Rumus *recall* adalah sebagai berikut:

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

*F1-score* digunakan sebagai metrik gabungan antara *precision* serta *recall* untuk memberi asesmen yang lebih seimbang, terutama ketika terdapat distribusi kelas yang tidak merata. Rumus *F1-score* ditunjukkan pada Persamaan berikut:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Penggunaan keempat metrik ini memberikan gambaran kuantitatif yang komprehensif mengenai performa model dalam membedakan sentimen negatif, dan positif. Selain itu, evaluasi juga dilengkapi dengan analisis confusion matrix untuk melihat pola kesalahan model dalam mengklasifikasikan antar kelas. Confusion matrix membantu mengidentifikasi bias model, misalnya kecenderungan salah mengklasifikasikan sentimen negatif sebagai positif atau sebaliknya, sehingga memberikan pemahaman tambahan mengenai kekuatan dan kelemahan kedua model.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Hasil Pengumpulan Data

Penghimpunan data dalam riset berikut dilaksanakan menggunakan platform media sosial X memakai tools Tweet Harvest pada Google Colab. Proses pengumpulan data dilaksanakan dengan menerapkan *key word* “timnas”, “arab”, “irak”, yang berhubungan dengan Kualifikasi Piala Dunia 2026. Pengambilan data dilaksanakan pada 8–9 Oktober 2025 untuk pertandingan melawan Arab Saudi, serta 11–12 Oktober 2025 untuk pertandingan melawan Irak, mencakup unggahan sebelum dan sesudah pertandingan. Melalui proses ini diperoleh total 1.304 komentar, data itu disimpan dalam format CSV.

#### 3.2 Hasil Preprocessing

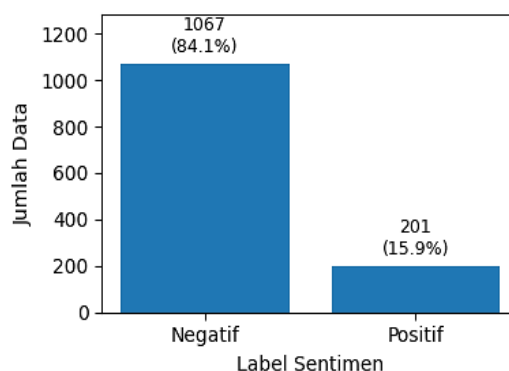
Proses preprocessing bertujuan untuk memperbaiki mutu data melalui penyaringan komentar yang tak sesuai dan menyamakan bentuk tulisan sebelum dilaksanakan analisis sentimen. Dari 1.304 diperoleh sebanyak 1.268 komentar yang dinilai layak untuk digunakan pada proses klasifikasi sentimen. Setiap tahapan preprocessing mengubah teks secara bertahap dari bentuk asli menjadi lebih terstruktur melalui proses cleaning, case folding, normalisasi, tokenizing, dan stopword removal, sehingga menghasilkan teks yang siap digunakan untuk analisis sentimen. Tahapan tersebut dapat dilihat dalam Tabel 2.

**Tabel 2.** Hasil Preprocessing

Alur Proses	Hasil
Teks Asli	Tembok itu bernama Maarten Paes! Dari tembakan tepat sasaran Arab Saudi semalam <u>di</u> antaranya berhasil digagalkan oleh beliau ini. Timnas Indonesia pun terhindar dari skor kekalahan yang lebih besar. <a href="https://t.co/SE4j9VnRKt">https://t.co/SE4j9VnRKt</a>
Cleaning	Tembok itu bernama Maarten Paes Dari tembakan tepat sasaran Arab Saudi semalam di antaranya berhasil digagalkan oleh beliau ini Timnas Indonesia pun terhindar dari skor kekalahan yang lebih besar
Case Folding	tembok itu bernama maarten paes dari tembakan tepat sasaran arab saudi semalam di antaranya berhasil digagalkan oleh beliau ini timnas indonesia pun terhindar dari skor kekalahan yang lebih besar
Normalisasi	tembok itu bernama maarten paes dari tembakan tepat sasaran arab saudi semalam di antaranya berhasil digagalkan oleh beliau ini timnas indonesia pun terhindar dari skor kekalahan yang lebih besar
Tokenizing	['tembok', 'itu', 'bernama', 'maarten', 'paes', 'dari', 'tembakan', 'tepat', 'sasaran', 'arab', 'saudi', 'semalam', 'di', 'antaranya', 'berhasil', 'digagalkan', 'oleh', 'beliau', 'ini', 'timnas', 'indonesia', 'pun', 'terhindar', 'dari', 'skor', 'kekalahan', 'yang', 'lebih', 'besar']
Stopword Removal	['tembok', 'bernama', 'maarten', 'paes', 'tembakan', 'tepat', 'sasaran', 'arab', 'saudi', 'semalam', 'antaranya', 'berhasil', 'digagalkan', 'beliau', 'timnas', 'indonesia', 'terhindar', 'skor', 'kekalahan', 'lebih', 'besar']

#### 3.3 Hasil Pelabelan Data

Setelah tahap pra-pemrosesan, tiap tweet diberikan poin sentimen memakai pendekatan leksikon melalui penjumlahan porsi kata dalam kamus sentimen. Tweet diklasifikasikan sebagai positif apabila poinnya  $\geq 0$  serta negatif apabila poinnya  $< 0$ . Output pelabelan ditunjukkan dalam Gambar 2.



**Gambar 2.** Hasil Pelabelan Data

Gambar 2 memperlihatkan bahwasanya sentimen publik terhadap performa Tim Nasional Indonesia pada putaran keempat Kualifikasi Piala Dunia 2026 didominasi oleh sentimen negatif. Dari total data yang dianalisis, sebanyak 84,1% (1.067 tweet) tergolong sebagai sentimen negatif, sementara 15,9% (201 tweet) termasuk dalam

sentimen positif. Dominasi sentimen negatif ini mengindikasikan bahwa mayoritas pengguna media sosial X mengekspresikan ketidakpuasan atau pandangan kritis terhadap performa Timnas Indonesia, yang sejalan dengan hasil pertandingan yang kurang optimal pada fase tersebut. Pada Tabel 3, ditampilkan contoh hasil pelabelan data yang menunjukkan bahwa setiap teks telah diklasifikasikan ke dalam kategori sentimen positif dan negatif sesuai dengan hasil perhitungan skor menggunakan pendekatan lexicon-based.

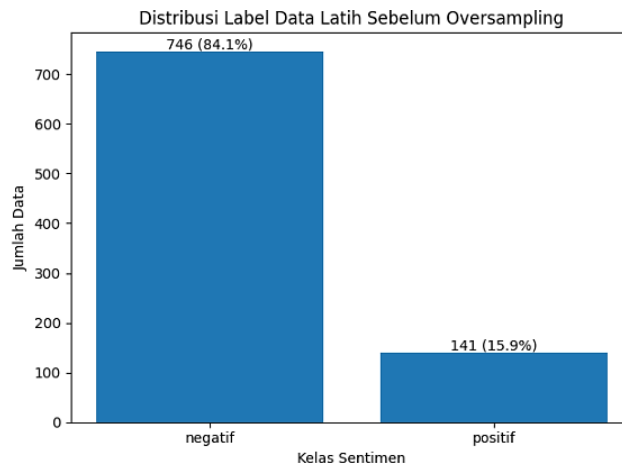
**Tabel 3.** Hasil Pelabelan Data

Teks	Label
goal arab semuanya berasal dari kesalahan yakob wkwk	Negatif
indonesia kalah tipis dari arab saudi	Negatif
babak pertama laga Indonesia vs irak belum ada gol tercipta Nasib akan ditentukan dalam menit ke depan	Positif

### 3.4 Hasil Penanganan Ketidakseimbangan Data

#### 3.4.1 Naïve Bayes Tanpa Oversampling

Hasil distribusi label Naïve Bayes tanpa oversampling tertera pada Gambar 3 menunjukkan ketidakseimbangan yang signifikan antara kelas negatif dan positif.

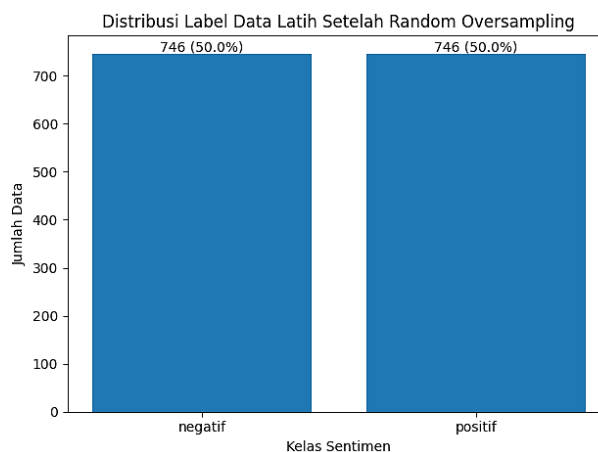


**Gambar 2.** Hasil Pelabelan Tanpa Oversampling

Berdasarkan Gambar 3, distribusi data memperlihatkan bahwasanya kelas negatif mendominasi dengan jumlah 746 data atau sebesar 84,1%, sedangkan kelas positif hanya berjumlah 141 data atau 15,9%. Perbedaan jumlah yang cukup signifikan ini menunjukkan adanya ketidakseimbangan distribusi antar kelas. Kondisi tersebut berpotensi memengaruhi proses pembelajaran model, karena model cenderung lebih sering mengidentifikasi corak di dalam kelas mayoritas dibanding kelas minoritas. Akibatnya, model dapat menjadi bias dan lebih banyak memprediksi sentimen negatif selama proses klasifikasi.

#### 3.4.2 Naïve Bayes Dengan Oversampling

Distribusi label Naïve Bayes setelah diterapkan oversampling bisa ditemukan di dalam Gambar 4 yang menunjukkan distribusi data menjadi seimbang antara kelas negatif dan positif.



**Gambar 3.** Hasil Pelabelan Dengan Oversampling

Berlاندaskan pada Gambar 4. setelah dilakukan penerapan Random Oversampling, distribusi data menjadi seimbang antara kelas negatif dan positif. Masing-masing kelas memiliki jumlah 746 data atau sebesar 50%. Penyeimbangan ini dilakukan untuk mengurangi bias model terhadap kelas mayoritas, sehingga proses pelatihan dapat mempelajari pola dari kedua kelas secara lebih proporsional. Dengan distribusi yang seimbang, diharapkan performa model terhadap kelas minoritas dapat meningkat.

### 3.4.3 IndoBERT Tanpa Class Weight

Proses *fine-tuning* dilakukan selama beberapa tahap penelitian (*training steps*). Tabel 4. Nilai *training loss* tanpa teknik *class weight*.

**Tabel 4.** Hasil Training Loss Tanpa CW

Step	Training Loss
50	0.455600
100	0.377100
150	0.244400
200	0.120500
250	0.058000

Berdasarkan Tabel 4. Skenario tanpa penanganan ketidakseimbangan data, model IndoBERT dilatih menggunakan distribusi data latih asli tanpa penerapan *class weighting*. Proses *fine-tuning* menunjukkan penurunan *training loss* yang konsisten, dari 0,4556 pada step ke-50 menjadi 0,0580 pada step ke-250. Penurunan tersebut membuktikan bahwasanya model dapat menelaah corak pola dalam data latih. Tetapi, karena distribusi data tidak seimbang dan didominasi oleh kelas negatif, model berpotensi lebih cenderung memprediksi kelas mayoritas dibandingkan kelas minoritas.

### 3.4.4 IndoBERT Dengan Class Weight

Proses *fine-tuning* setelah diterapkan *class weight* dapat terlihat pada Tabel 5. Dengan penurunan nilai *training loss* yang konsisten.

**Tabel 5.** Hasil Training Loss CW

Step	Training Loss
50	0.697500
100	0.528800
150	0.340100
200	0.174500
250	0.072100

Berdasarkan Tabel 5. Skenario dengan penerapan *class weighting*, proses *fine-tuning* IndoBERT menunjukkan penurunan *training loss* yang konsisten dari 0,6975 pada step ke-50 menjadi 0,0721 pada step ke-250. Nilai awal *loss* yang lebih tinggi dibandingkan skenario tanpa *class weight* menunjukkan bahwa model memberikan penalti yang lebih besar terhadap kesalahan prediksi, khususnya pada kelas minoritas. Seiring bertambahnya langkah pelatihan, *loss* yang semakin menurun mengindikasikan bahwa model mulai mampu menyesuaikan diri dengan distribusi bobot kelas yang diterapkan, sehingga pembelajaran menjadi lebih seimbang antar kelas.

## 3.5 Evaluasi Perbandingan Model

Berikut Tabel 6, ringkasan Confusion Matrix untuk seluruh model Naïve Bayes dan IndoBERT dalam bentuk TP, TN, FP, FN.

**Tabel 6.** Hasil Perbandingan Confusion Matrix Model

Model	TN	TP	FN	FP
NB Tanpa RO	321	0	60	0
NB RO	272	27	33	49
IndoBERT Tanpa CW	303	21	39	18
IndoBERT CW	296	26	34	25

Berdasarkan Tabel 6, terlihat perbandingan performa model Naïve Bayes dan IndoBERT dalam kondisi tanpa dan dengan penanganan ketidakseimbangan data. Pada Naïve Bayes tanpa *oversampling*, model menunjukkan kecenderungan kuat memprediksi kelas negatif dengan nilai *True Negative* sebesar 321 dan tidak menghasilkan prediksi positif sama sekali (*True Positive* = 0, *False Negative* = 0). Hal ini menunjukkan adanya bias terhadap kelas mayoritas. Setelah diterapkan *Random Oversampling*, kemampuan Naïve Bayes dalam mengenali kelas positif meningkat, ditunjukkan dengan diperolehnya 27 data terklasifikasi benar sebagai *True*





tidak seimbang. Selain itu, pengembangan kamus atau model yang lebih sensitif terhadap variasi bahasa informal dan slang pada media sosial juga menjadi aspek penting untuk meningkatkan akurasi dan generalisasi model dalam konteks NLP bahasa Indonesia.

## REFERENCES

- [1] I. Apriyani, A. Fauzi, D. S. Kusumaningrum, and H. H. Handayani, "Analisis Sentimen Performa Timnas Sepak Bola Indonesia pada Kolom Komentar Aplikasi TikTok Menggunakan Algoritma Machine Learning," *J. Account. Inf. Syst.*, vol. 8, no. 1, pp. 76–89, 2025, doi: 10.32627/aims.v8i1.1160.
- [2] A. D. Pradiptha, F. Siddiq, A. Farisi, M. F. Pratama, and D. Aprianti, "Analisis Sentimen Suporter terhadap Performa Tim Nasional Sepakbola Indonesia pada Turnamen Sea Games 2023 dengan Metode Naive Bayes," *J. Sains dan Inform.*, vol. 11, pp. 30–39, 2025, doi: 10.34128/jsi.v11i1.775.
- [3] Nouvan, "Negara Dengan Pengguna X (twitter) terbanyak Juli 2025, Indonesia urutan keempat," *Dataloka.id*, <https://dataloka.id/humaniora/4627/negara-dengan-pengguna-x-twitter-terbanyak-juli-2025-indonesia-urutan-keempat/> (accessed Apr. 22, 2026).
- [4] W. Alfiyani, D. Abdul Fatah, and F. Irahmani, "Penerapan algoritma naïve Bayes Untuk Analisis Sentimen Pada media sosial X Terhadap Performa tim nasional sepak bola indonesia di era Kepemimpinan Shin Tae-yong," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 3, pp. 3969–3977, 2025, doi:10.36040/jati.v9i3.13451.
- [5] F. F. Wati, Suleman, and A. E. Widodo, "Analisis Sentimen Ulasan Pengguna Aplikasi Deepseek Menggunakan Algoritma Random Forest Dan Naive Bayes," *CONTEN Comput. Netw. Technol.*, vol. 5, no. 1, pp. 8–15, 2025, doi: 10.31294/hqpha267.
- [6] R. Lasepa, S. Riyadi, S. Ramadhan, and D. D. Saputra, "Sentiment Analysis Terhadap Perspektif Warganet Atas Tragedi Kanjuruhan Malang di Twitter Menggunakan Naïve Bayes Classifier," *J. Inform.*, vol. 10, no. 1, pp. 45–53, 2023, doi: 10.31294/inf.v10i1.14546.
- [7] R. H. Nufus and U. Surapati, "Analisis Sentimen Persepsi Masyarakat Terhadap Timnas Indonesia U-23 dalam AFC-23 Asian Cup 2024 Pada Media Sosial X Menggunakan Metode Naïve Bayes Classifier," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 3, pp. 2647–2657, 2024, doi: 10.35870/jimik.v5i3.964.
- [8] H. S. Anggraheni *et al.*, "Deteksi Spam Berbahasa Indonesia Berbasis Teks Menggunakan Model Bert," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 6, pp. 1291–1301, 2024, doi: 10.25126/jtiik.2024118121.
- [9] C. J. Tobing, IGN Lanang Wijayakusuma, and Luh Putu Ida Harini, "Perbandingan Kinerja Indobert dan Mbert Untuk Deteksi Berita Hoaks Politik Dalam Bahasa Indonesia," *JST (Jurnal Sains dan Teknologi)*, vol. 14, no. 1, pp. 114–123, 2025, doi:10.23887/jstundiksha.v14i1.92126.
- [10] W. Widyandana, Maskur, and A. Fauzi, "Machine Learning and Transformer-based Model for Sentiment Analysis of Indonesian E-Commerce Reviews," *Indones. J. Comput. Sci.*, vol. 14, no. 4, pp. 6262–6271, 2025, doi: 10.33022/ijcs.v14i4.4980.
- [11] M. I. Abidin and E. W. Pamungkas, "Analisis Sentimen terhadap Tim Nasional Indonesia pada Piala Asia 2023 Menggunakan Model Transformer Bahasa Indonesia," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 482–496, 2025, doi:10.36341/rabit.v10i2.6142.
- [12] C. Ramadhan, V. Atina, and H. Permatasari, "Analisis Perbandingan Model CNN dan IndoBERT Dalam Sentimen Berita Politik Indonesia," *Pros. Semin. Nas. Teknol. Inf. dan Bisnis*, pp. 110–118, 2025, doi: 10.47701/v1r9ka69.
- [13] A. Agustin, S. Andrea, S. Susanti, R. Rahmiati, and H. Hamdani, "Review Aplikasi Kredivo Menggunakan Analisis Sentimen dengan Algoritma Support Vector Machine," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 9, no. 1, pp. 39–49, 2023, doi: 10.36341/rabit.v9i1.4107.
- [14] N. N. Aini Aryanti and O. Suria, "Analisis Sentimen terhadap Pemutusan Hubungan Kerja di Indonesia: Komparasi IndoBERT dengan SVM, Random Forest, dan Decision Tree dengan Optimasi TF-IDF," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 1158–1176, 2025, doi:10.36341/rabit.v10i2.6364.
- [15] F. Alvin, N. Anisa, and S. Winarsih, "Perbandingan Kinerja Model IndoBERT , IndoBERTweet , dan Algoritma Klasik pada Analisis Sentimen Isu Indonesia Gelap," *Build. Informatics, Technol. Sci.*, vol. 7, no. 3, pp. 1601–1613, 2025, doi: 10.47065/bits.v7i3.8636.
- [16] A. A. Rohman and G. A. Trisnapradika, "Perbandingan Algoritma NBC , SVM , Logistic Regression untuk Analisis Sentimen Terhadap Wacana KaburAjaDulu di Media Sosial X," *Build. Informatics, Technol. Sci.*, vol. 7, no. 1, pp. 169–178, 2025, doi: 10.47065/bits.v7i1.7261.
- [17] E. A. Winanto, S. M. Z. Ali Difyah, P. A. Jusia, and Sharipuddin, "Analisis Sentimen Terhadap Tagar Kabur Aja Dulu Di Twitter Menggunakan Metode Lexicon-Based," *Jurnal PROCESSOR*, vol. 20, no. 2, 2025, doi:10.33998/processor.2025.20.2.2542.
- [18] A. A. Qolbu, N. Fitriyati, and N. Inayah, "Performa Naïve Bayes , SVM , dan IndoBERT pada Analisis Sentimen Twitter IndiHome dengan Strategi Penanganan Data Tidak Seimbang," *J. FOURIER*, vol. 814, no. 1, pp. 29–44, 2025, doi: 10.14421/fourier.2025.141.29-44.
- [19] A. D. M. Putri, N. Sulistianingsih, and R. Rismayati, "Pengaruh Teknik Representasi Teks Bag-of-Words dan TF-IDF terhadap Akurasi Klasifikasi Sentimen Teks Multi-Domain," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 7, no. 4, pp. 675–688, 2025, doi: 10.35746/jtim.v7i4.756.
- [20] N. M. Damayanti, I. D. Ariningtyas, M. Izuddin, and A. Icham, "Analisis Sentimen Tagar #BTSComeback di Platform X Menggunakan IndoBERTweet," *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, vol. 13, no. 3, 2025, doi: 10.23960/jitet.v13i3.7176.
- [21] T. Juniardi and C. A. Sugiarto, "Analisis Sentimen Terhadap Tim Nasional Sepak Bola Indonesia pada Piala Dunia U-17 Menggunakan Algoritma Naïve Bayes di Platform X," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3S1, 2024, doi:10.23960/jitet.v12i3s1.5188.