



Analisis Sentimen Publik Terhadap Peluang Timnas Indonesia Lolos ke Piala Dunia 2026 Dengan Algoritma Naïve Bayes dan Random Forest

Faizura Zadri¹

¹Program Studi Teknik Informatika S-2, Universitas Pamulang

Email: ¹faizurazadri00@gmail.com

ABSTRACT

Football is a sport that is highly anticipated by the Indonesian people in 2025. This enthusiasm increases with the opportunity for the Indonesian national team to compete in the biggest football event in the world, namely the 2026 World Cup which will be held in Canada, Mexico and the United States. This study aims to analyze public sentiment towards the chances of the Indonesian national team qualifying for the 2026 World Cup. The data used is sourced from Youtube social media, with a total of 5910 comments. The data is divided into two parts, namely 80% for training and 20% for testing. The results of this study indicate that the Random Forest algorithm produces the best performance with an accuracy of 73%, a precision of 73%, and a recall of 73%. This value is obtained based on the results of model evaluation using a confusion matrix. Thus, the Random Forest algorithm is considered more suitable than Naïve Bayes in classifying public sentiment towards the chances of the Indonesian national team qualifying for the 2026 World Cup.

Keywords: *Sentiment Analysis; Indonesian National Team; Random Forest; Naïve Bayes; Public Opinion; World Cup*

ABSTRAK

Sepakbola menjadi olahraga yang sangat dinantikan oleh masyarakat Indonesia pada tahun 2025. Antusiasme ini meningkat dengan seiring adanya peluang timnas Indonesia untuk berlaga di ajang sepak bola terbesar didunia, yaitu Piala Dunia 2026 yang akan diselenggarakan di Kanada, Meksiko dan Amerika Serikat. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap peluang tim nasional Indonesia lolos ke Piala Dunia 2026. Data yang digunakan bersumber dari media sosial Youtube, dengan jumlah data sebanyak 5910 komentar. Data dibagi menjadi dua bagian, yaitu 80% untuk pelatihan (*training*) dan 20% untuk pengujian (*testing*). Hasil penelitian ini menunjukkan bahwa algoritma Random Forest menghasilkan performa terbaik dengan akurasi sebesar 73%, *precision* sebesar 73%, dan *recall* sebesar 73%. Nilai ini diperoleh berdasarkan hasil evaluasi model menggunakan *confusion matrix*. Dengan demikian, algoritma Random Forest dinilai lebih cocok dibandingkan Naïve Bayes dalam mengklasifikasikan sentimen masyarakat terhadap peluang tim nasional Indonesia lolos ke Piala Dunia 2026.

Kata kunci: Sentimen Analisis; Timnas Indonesia; Random Forest; Naïve Bayes; Opini Publik; Piala Dunia.

1. PENDAHULUAN

Pada saat sekarang ini media sosial menjadi tempat bagi orang-orang untuk berinteraksi satu sama lain dari jarak jauh. Platform seperti Youtube, Instagram, tiktok serta twitter dijadikan sebagai tempat untuk berekspresi dan menyampaikan pendapat terhadap isu-isu maupun hal yang banyak dibicarakan saat ini. Opini yang disampaikan

pada media sosial bisa mempengaruhi cara berpikir, berperilaku maupun dalam mengambil keputusan[1]. Pada Analisis sentimen terdapat suatu proses pemahaman, ekstraksi data, serta melakukan proses pengolahan data secara otomatis untuk mendapatkan informasi tertentu[2]. Dari analisis sentimen ini menghasilkan informasi terkait opini pengguna baik itu bersifat positif, netral, maupun negatif.

Salah satu penerapan penggunaan sentiment analisis yaitu pada sektor olahraga tepatnya pada sepak bola. Sepak bola menjadi salah satu olahraga terpopuler di dunia yang mana memiliki penggemar dari segala usia[3]. Analisis sentimen dalam sepak bola dapat kita gunakan sebagai sarana untuk memahami bagaimana perasaan atau opini masyarakat terhadap suatu pertandingan, tim, pemain serta hal-hal yang terjadi seputar sepak bola. Dengan adanya analisis ini dapat membantu pihak-pihak dalam dunia sepak bola untuk memahami opini publik serta dapat membuat keputusan yang lebih baik[4].

Di Indonesia, sepak bola menjadi olahraga yang banyak digemari oleh masyarakatnya, sebanyak 90,8% dari populasi Indonesia mengenal atau mengetahui apa itu olahraga sepak bola. Dari kelompok yang mengenal tersebut, ada sekitar 46,7% memiliki minat aktif, misalnya seperti mereka yang menonton pertandingan, mengikuti perkembangan liga, mendukung tim maupun bermain sepak bola[5]. Ditahun 2025, pertandingan sepak bola menjadi tontonan yang banyak dinanti oleh masyarakat Indonesia. Kondisi ini sejalan dengan adanya peluang bagi Timnas Indonesia untuk melangkah ke level yang lebih tinggi, yaitu tampil di ajang Piala Dunia 2026 yang akan diselenggarakan di Kanada, Meksiko dan Amerika Serikat. Jika berhasil lolos ke Piala Dunia 2026, Indonesia akan mencetak sejarah baru, karena untuk pertama kalinya Timnas akan tampil di perhelatan sepak bola terbesar di dunia.

Dalam analisis sentimen kita bisa menggunakan algoritma seperti Naïve Bayes atau bisa juga menggunakan Random Forest. Naïve Bayes merupakan algoritma yang sederhana serta memiliki akurasi yang tinggi. Sedangkan Random Forest memiliki kemampuan untuk menghasilkan kesalahan yang relatif rendah, klasifikasi yang baik, serta memiliki kemampuan untuk menangani data pelatihan yang besar secara efisien[6]. Kedua algoritma ini banyak digunakan oleh peneliti dalam analisis sentiment, seperti pada penelitian yang dilakukan oleh Irma Apriyani, mengenai Analisis Sentimen Performa Timnas Sepak Bola Indonesia pada Kolom Komentar Aplikasi Tiktok, yang

menghasilkan akurasi sebesar 97.30%[7]. Selain itu penelitian yang dilakukan oleh Wulan Alfiyani, mengenai Penerapan Algoritma Naïve Bayes Untuk Analisis Sentimen pada Pada Media Sosial X, menghasilkan akurasi prediksi sebesar 98,73% [4]

Pada penelitian ini, penulis akan menggunakan kedua metode tersebut untuk menganalisis sentimen masyarakat terhadap peluang Timnas Indonesia lolos ke Piala Dunia 2026. Dari hasil analisis, akan diperoleh satu metode yang memiliki tingkat akurasi paling tinggi.

2. METODE

2.1. Tempat dan Waktu Penelitian

Penelitian ini dilaksanakan secara daring menggunakan platform Google Colab sebagai tempat untuk pengambilan data yang ada pada video Youtube. Sedangkan untuk *training,testing* serta prediksi model menggunakan perangkat lunak Orange. Proses pengumpulan data dari *platform* youtube dilakukan dua dalam dua periode, yaitu pada tanggal 18 Mei 2025 dan 13 juni 2025.

2.2. Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan text mining dan machine learning untuk mengklasifikasi sentiment publik. Penelitian ini bertujuan untuk mengidentifikasi sentimen publik terhadap peluang timnas Indonesia lolos ke Piala Dunia 2026 melalui komentar pengguna Youtube.

2.3. Metode Pengumpulan Data

Dataset diperoleh dari komentar publik pada platform Youtube yang membahas terkait peluang Timnas Indonesia untuk lolos ke Piala Dunia 2026. Pengambilan Data dilakukan dua periode, yaitu pada tanggal 18 Mei 2025 sebanyak 12 video, dan pada tanggal 13 Juni 2025 sebanyak 10 vidio.

Data dikumpulkan dari komentar publik melalui platform Youtube yang membahas terkait Peluang timnas Indonesia lolos ke piala dunia 2026. Pengumpulan data diambil berdasarkan kata kunci yang telah ditentukan.

Berikut adalah kata kunci yang digunakan :

Tabel 2. Kata Kunci Pengumpulan Data

| No | Kata Kunci | Jumlah Video |
|----|---|--------------|
| 1 | Indonesia lolos piala dunia 2026 | 4 |
| 2 | Peluang indonesia ke piala dunia 2026 | 5 |
| 3 | Timnas indonesia ke piala dunia 2026 | 3 |
| 4 | Peluang timnas Indonesia lolos piala dunia 2026 | 4 |
| 5 | Timnas Indonesia lolos piala dunia 2026 | 2 |
| 6 | Timnas indonesia | 3 |
| 7 | Peluang timnas indonesia | 1 |

Komentar dari masing-masing video dikumpulkan menggunakan Teknik *web scrapping* dengan menggunakan bahasa pemrograman python yang dijalankan pada layanan Google Colab. Secara keseluruhan, jumlah komentar yang berhasil dikumpulkan sebanyak 9291 komentar. Namun, untuk keperluan pelatihan (*training*) dan pengujian (*testing*) model, digunakan sebanyak 5910 komentar. Selain itu, sebanyak 101 komentar digunakan untuk proses prediksi sentimen terhadap data baru. Berikut pada table 2.1 adalah informasi terkait data yang digunakan.

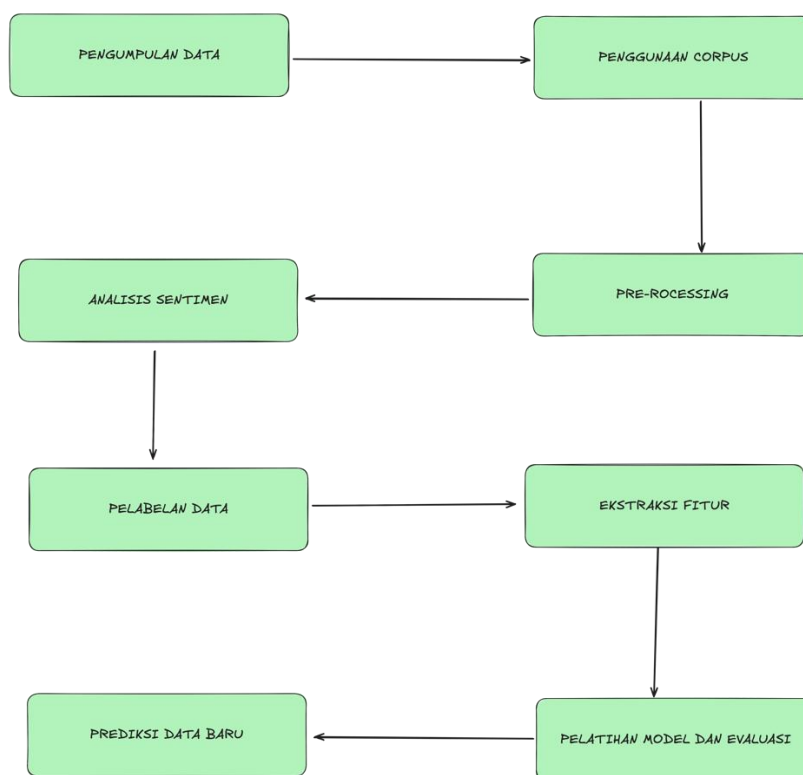
Tabel 2.1 Detail Dataset

| Author | comment |
|-----------------|---|
| @kidsdoballoon | ♡ Garuda |
| @Home_Workout34 | Rata rata saya liat komenan org ² nya pesimis,,tapi setelah pertandingan melawan Bahrain apakah kalian masih pesimis???????????? |
| @RizaldiB-gq3cw | Semoga do'a Rakyat indonesia terkabul , Selalu Menang terhadap lawan Tandingan nya |

| | |
|-----------------------|--|
| @FrancelinoMonteiro | Kalaupun lolos ke pdpun nanti jdi lumbung gol tim2 raksasa 😡😡😡 |
| @jatnuriseptiawan9257 | Gw ykn masih bisa masuk piala dunia. Klo ga jadi runer up minimal klasemen di peringkat 3 atau 4 |

2.4. Tahapan Penelitian

Pada Gambar 2 merupakan alur atau workflow dari proses penelitian analisis sentimen publik terhadap peluang timnas Indonesia lolos ke Piala Dunia 2026.




Gambar 2. Perancangan Penelitian

Dalam melakukan penelitian ini dibagi menjadi beberapa tahapan, diantaranya :

2.4.1 Pengumpulan data

Data komentar masyarakat yang diambil dari media sosial Youtube diambil menggunakan metode web scraping. *Web scraping* merupakan proses pengambilan data dari halaman web, yang mana dengan ini kita bisa mengumpulkan dan ekstraksi data[8]. Pada penelitian ini dilakukan menggunakan layanan dari GoogleColab dengan memanfaatkan library dari *Youtube comment downloader*.



```
import csv
import time
from youtube_comment_downloader import YoutubeCommentDownloader

# Inisialisasi downloader
downloader = YoutubeCommentDownloader()

# Ganti dengan ID video YouTube yang diinginkan
video_id = 'NEqMQEBTII'

# Mendownload komentar
comments = downloader.get_comments(video_id)

# Generate nama file dengan timestamp
timestamp = int(time.time()) # Mendapatkan timestamp saat ini sebagai angka
file_name = f'youtube_comments_{timestamp}.csv'

# Membuka file CSV untuk menulis
with open(file_name, mode='w', newline='', encoding='utf-8') as file:
    writer = csv.writer(file)
    # Menulis header
    writer.writerow(['Author', 'Comment'])

    # Menuliskan setiap komentar ke file CSV
    for comment in comments:
        writer.writerow([comment['author'], comment['text']])

print(f"Komentar berhasil disimpan ke {file_name}")
```

Komentar berhasil disimpan ke youtube_comments_1749825608.csv

Gambar 2.1 Kode python untuk pengumpulan data

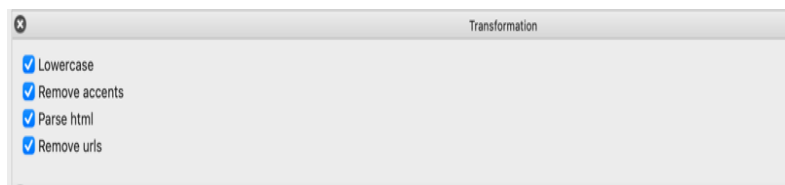
Gambar 2.1 menunjukkan implementasi kode python yang digunakan untuk ekstraksi komentar dari vidio youtube yang membahas terkait peluang tim nasional Indonesia lolos ke piala dunia 2026. Proses dimulai dengan mengimport *library youtube comment downloader* yang berfungsi sebagai alat bantu untuk mengakses komentar publik dari vidio tersebut.

2.4.2 Pre-processing data

Pada tahap ini akan dilakukan pembersihan data, yang awalnya tidak terstruktur menjadi terstruktur sehingga siap digunakan untuk merancang model sentimen analisis[7]. Selain itu *preprocessing* dilakukan untuk membersihkan data dari kata-kata yang dapat mempengaruhi akurasi seperti simbol “@” dan “#” . hal ini bisa dilakukan dengan memanfaatkan berbagai fitur yang ada pada *processing*, seperti *Tokenization, Transformation, filtering* dan yang lainnya. Berikut adalah beberapa proses yang dilakukan[1].

a. Transformation

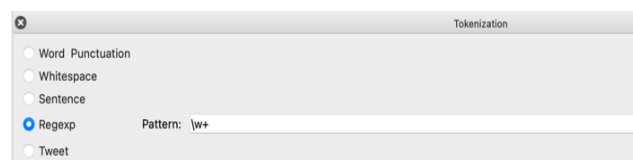
Pada transformasi akan dilakukan proses perubahan data menjadi kecil, menghapus *accents, parse* html serta menghapus url.



Gambar 2.2 Transformation

b. Tokenization

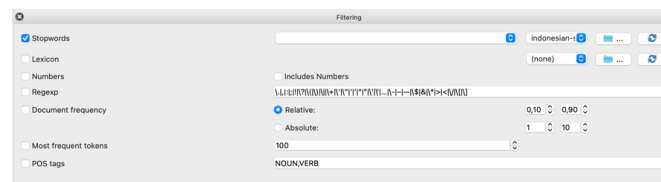
Tokenization digunakan untuk memecah teks menjadi unit-unit kata agar bisa dianalisis lebih lanjut, seperti sentimen maupun klasifikasi. *Tokenization* yang digunakan yaitu *Regular Expression*, yang mana menggunakan pola khusus.



Gambar 2.3 Tokenization

c. Filtering

Pada *filtering* bagian yang digunakan yaitu StopWords. Fitur ini digunakan untuk menghilangkan kata yang tidak memiliki makna penting dalam teks.



Gambar 2.4 Filtering

2.4.3 Pelabelan Data

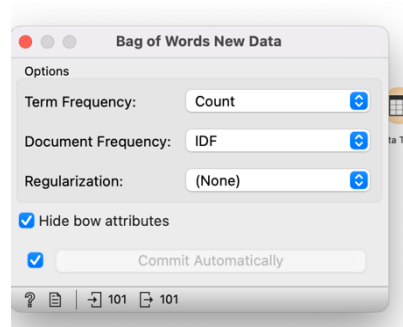
Setelah proses preprocessing dan didapat hasil sentimen dari widget sentimen analisis yang ada pada perangkat lunak orange, selanjutnya data disimpan dalam format excel lalu dilakukan pelabelan secara manual. Pelabelan merupakan proses pengelompokan data teks menjadi beberapa bagian. Kata-kata yang ada dalam teks akan dievaluasi dan diberi skor sesuai dengan kecenderungan emosionalnya[9]. Pelabelan dilakukan pada microsoft excel, berikut adalah datanya :

Tabel 2.2 Pelabelan Data

| Author | Comment | score | sentiment |
|---------------------|---|----------|-----------|
| @AMiRA-ud8he | Bahkan sampai kiamat pun aku rasa gak akan pernah bisa masuk piala dunia | 0 | NETRAL |
| @RizaldiB-gq3cw | Semoga do'a Rakyat indonesia terkabul , Selalu Menang terhadap lawan Tandingan nya | 0 | NETRAL |
| @kidsdoballoon | ♡ Garuda | 0 | NETRAL |
| @septiekowibowo6848 | Mau gak mau harus menang ni sama Japan n China wkwwkwkw | 142.857 | POSITIF |
| @DedenJunaidi-x5o | Klo menang ama jepang dan naga china ya lolos lah min min jngn ngawurr | 833.333 | POSITIF |
| @Home_Workout34 | Rata rata saya liat komenan org ² nya pesimis,,tapi setelah pertandingan melawan Bahrain apakah kalian masih pesimis???????????? | -111.111 | NEGATIF |
| @tutorialjoe773 | Seandainya waktu lawan bahrain di kandang bahrain indonesia menang ga di curangin indonesia yg menemani jepang lolos ke piala dunia | 0 | NETRAL |
| @jesika4410 | Kko indo menang lwan jpan bisa gk ya posisi kedua | 12.5 | POSITIF |
| @asnawipuryani5494 | Negara miskin | -50 | NEGATIF |

2.4.4 Ekstraksi Fitur (Bag of Words)

Metode ini merupakan cara sederhana dalam memproses data teks dengan mengubahnya menjadi *vector* angka agar dapat diproses oleh computer[10]. Berikut adalah konfigurasi dari *widget bag of words* :



Gambar 2.5 Ekstraksi Fitur

Pada Gambar 2.5 merupakan tampilan pengaturan dari *widget Bag of Wors(BoW)*. Pada bagian *Document Frequency* yang digunakan, yaitu IDF. Dengan IDF akan mengurangi bobot kata yang muncul di banyak dokumen. Kata yang jarang muncul di dokumen lain akan memiliki bobot lebih tinggi. Semakin kecil nilai IDF maka akan dianggap semakin tidak penting kata tersebut. Berikut ini adalah persamaan dari TF-IDF [11]

$$TF_i = \frac{n_i}{\sum n} \quad (1)$$

$$IDF_i = \log\left(\frac{N}{df_i}\right) \quad (2)$$

$$TF - IDF_i = TF_i \times IDF_i \quad (3)$$

Keterangan :

TF (*Term Frequency*) : Frekuensi suku kata

IDF (*Inverse Document Frequency*) : Frekuensi dokumen inversi

n_i : Jumlah kemunculan kata ke-i dalam satu dokumen

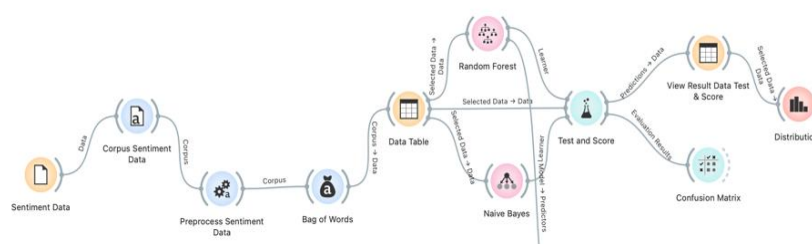
df : Banyaknya dokumen yang memiliki term tertentu

$\sum n$: Total jumlah kata

N : Banyak term dokumen

2.4.5 Pelatihan dan Pengujian Model

Pelatihan dan pengujian model menggunakan algoritma Naïve Bayes dan Random Forest. Dimana Random Forest akan mengkombinasikan banyak decision tree melalui *system voting* untuk hasil yang lebih statbil dan tahan terhadap *overvitting*[9]. Sedangkan untuk Naïve Bayes merupakan algoritma yang memanfaatkan distribusi frekuensi dari atribut untuk menentukan probabilitas suatu data termasuk ke dalam kelas tertentu[12]. Kedua Model dilatih menggunakan 80% dari total data serta 20% akan digunakan untuk pengujian model. Berikut adalah alur untuk melakukan training dan testing pada perangkat lunak orange:



Gambar 2.6 Workflow Pelatihan dan Pengujian

Pada Gambar 2.6 merupakan *workflow* analisis sentimen menggunakan Orange. Data komentar yang telah dilabeli dimuat dan diubah menjadi format corpus, yang kemudian diproses melalui tahapan pra-pemrosesan seperti tokenisasi, penggunaan *stopword*. Selanjutnya, fitur teks diekstraksi menggunakan metode *Bag of Words* dengan *term frequency* berbasis *count* dan pembobotan IDF untuk menghasilkan representasi TF-IDF. Data yang telah diproses sebelumnya digunakan untuk melatih dua model klasifikasi, yaitu *Random Forest* dan *Naive Bayes*. Setelah proses *training model* dilakukan, selanjutnya akan dievaluasi menggunakan *confusion matrix*. Pada bagian akhir akan dilihat bagaimana hasil prediksi menggunakan *widget distributions*, sehingga dengan ini dapat diketahui model mana yang memberikan performa paling efektif dalam mengklasifikasikan sentimen masyarakat secara akurat[13].

2.4.6 Evaluasi Model

Untuk mengevaluasi model yang telah melalui proses *training* dan *testing* kita bisa menggunakan *Confusion Matrix*. *Confusion matrix* merupakan *matriks* yang bisa digunakan untuk mengevaluasi kinerja model dalam mengklasifikasikan data dengan memberikan informasi berapa jumlah data uji yang benar dan salah[14]. Berikut ini adalah tabel dari *Confusion Matrix* :

Tabel 2.3 Tabel Confusion Matrix

| Actual | Predicted | |
|----------|----------------|---------------|
| | Positive | Negative |
| Positive | True Positive | False Negatif |
| Negative | False Positive | True Negatif |

Berdasarkan keterangan dari tabel *confusion matrix* yang ada pada tabel 2.3, maka dilakukan perhitungan terhadap metrik evaluasi untuk mengukur performa model, yaitu : [15]

- Accuracy*, memberikan informasi seberapa akurat model dapat memprediksi dengan benar. Berikut ini persamaan dari *accuracy*:

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \times 100 \quad (1)$$

- Precision*, memberikan informasi dari prediksi sebagai kelas positif yang sebenarnya positif. Berikut ini persamaan dari *precision* :

$$Precision = \frac{TP}{(TP+FP)} \times 100 \quad (2)$$

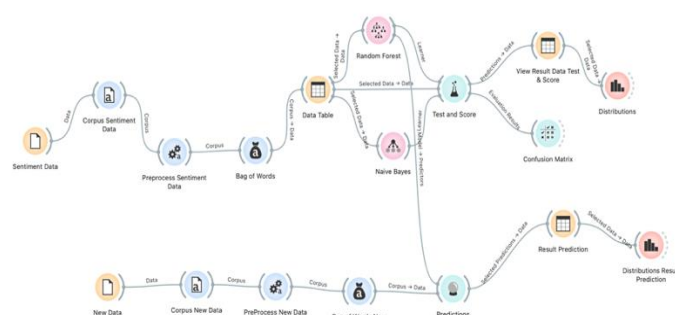
c. *Recall*, memberikan informasi dari prediksi kelas positif yang di prediksi negatif.

Berikut ini persamaan dari *Recall* :

$$Recall = \frac{TP}{(TP+FN)} \times 100 \quad (3)$$

2.4.7 Prediksi Data Baru

Dalam melakukan prediksi sentimen terhadap komentar baru masyarakat terhadap peluang timnas Indonesia lolos Piala Dunia 2026 dilakukan seperti gambar dibawah ini.



Gambar 2.7 Workflow Prediksi Data Baru

Pada gambar diatas merupakan design pada *canvas software orange*. Prediksi menggunakan data baru sebanyak 101 komentar dimana sebelum melakukan prediksi dilakukan *preprocessing* hingga penggunaan *bag of words*. Model yang digunakan untuk prediksi komentar baru adalah algoritma Random Forest

3. HASIL DAN PEMBAHASAN

3.1. Analisis Model Naïve Bayes

Dari hasil pengujian yang telah dilakukan menggunakan model Naïve Bayes didapat akurasi sebesar 79.6%.

| | | Predicted | | | Σ |
|--------|---------|-----------|--------|---------|------|
| | | NEGATIF | NETRAL | POSITIF | |
| Actual | NEGATIF | 630 | 627 | 78 | 1335 |
| | NETRAL | 303 | 2428 | 214 | 2945 |
| | POSITIF | 167 | 820 | 643 | 1630 |
| Σ | | 1100 | 3875 | 935 | 5910 |

Gambar 3. Confusion Matrix Model Naïve Bayes

Pada gambar diatas terlihat dari total 5910 komentar, model dapat memprediksi 2428 dari 2945 komentar dapat diprediksi dengan akurat. Namun, terdapat 627 komentar yang harusnya negatif malah diprediksi Netral.Serta, 820 komentar Positif malah diprediksi Netral.Hal ini, menunjukkan model cenderung memprediksi ke kelas Netral,dan mengalami kesulitan dalam membedakan komentar Negatif dan Positif.

3.2. Analisis Model Random Forest

Berikut ini adalah hasil evaluasi dari modelnya menggunakan *confusion matrix*:

| | | Predicted | | | Σ |
|--------|---------|-----------|--------|---------|------|
| | | NEGATIF | NETRAL | POSITIF | |
| Actual | NEGATIF | 798 | 460 | 77 | 1335 |
| | NETRAL | 140 | 2537 | 268 | 2945 |
| | POSITIF | 116 | 507 | 1007 | 1630 |
| Σ | | 1054 | 3504 | 1352 | 5910 |

Gambar 3.1 Confusion Matrix Model Random Forest

Pada gambar *confusion matrix* diatas menunjukkan bagaimana kinerja model dalam mengklasifikasikan sentimen menjadi 3 kelas yaitu Negatif, Netral, dan Positif. Model mampu mengklasifikasi data komentar netral dengan cukup baik(2428 dari 2945 komentar), namun sering kali keliru membedakan antara kelas Negatif dan Netral, serta antara Positif dan Netral.

| | | Test and Score | | | | | |
|---------------|-------|---|-------|-------|--------|-------|--|
| | | Evaluation results for target (None, show average over classes) | | | | | |
| Model | AUC | CA | F1 | Prec | Recall | MCC | |
| Naive Bayes | 0.796 | 0.626 | 0.610 | 0.631 | 0.626 | 0.379 | |
| Random Forest | 0.874 | 0.737 | 0.731 | 0.739 | 0.737 | 0.569 | |

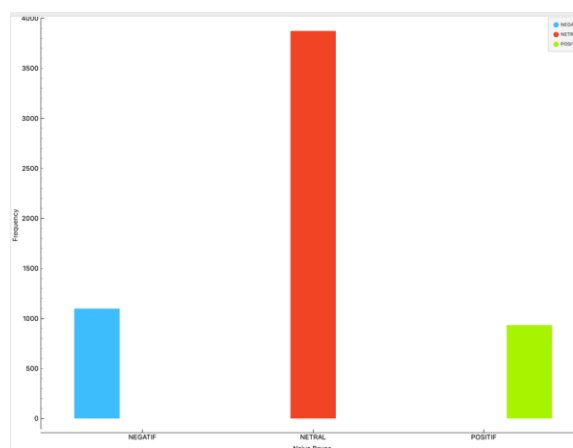
Gambar 3.2 Evaluasi Performa Model Naïve Bayes dan Random Forest

Selain menggunakan *confusion matrix*, model juga dievaluasi menggunakan AUC(Area Under The Receiver Operating Characteristic Curve). Hasil evaluasi menunjukkan bahwa model Random Forest menghasilkan skor AUC sebesar 87.4%, yang mana ini lebih tinggi jika dibandingkan dengan model Naïve bayes dengan AUC sebesar 79.6%. Selanjutnya, jika dilihat dari segi akuras (CA – *Classification Accuracy*), Random Forest memperoleh nilai 0.737 sedangkan Naïve Bayes memperoleh 0.626. Hal ini

didukung juga oleh *F1-Score*, *Precision*, dan *Recall* yang lebih tinggi pada Random Forest (masing-masing 0.731, 0.739, dan 0.737) dibandingkan dengan Naïve bayes (masing-masing 0.610, 0.631, dan 0.626). Nilai MCC (*Matthews Correlation Coefficient*), yang merupakan metrik yang mempertimbangkan keseimbangan antar kelas, juga menunjukkan performa Random Forest yang lebih baik (0.569) dibandingkan Naïve Bayes (0.379).

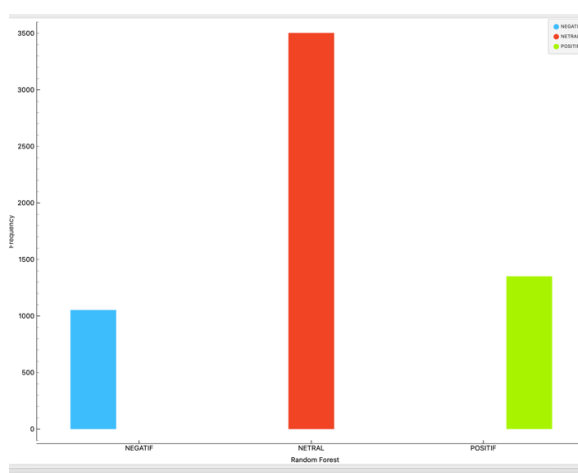
3.3. Hasil dan Visualisasi Sentimen

Data 80% total data, dimana 20% untuk pengujian diperoleh sentiment Netral sebanyak 65.57% , positif sebanya 15.82%, serta Negatif sebanyak 18.61%. Berikut adalah hasil visualiasi sentimen menggunakan model Naïve Bayes.



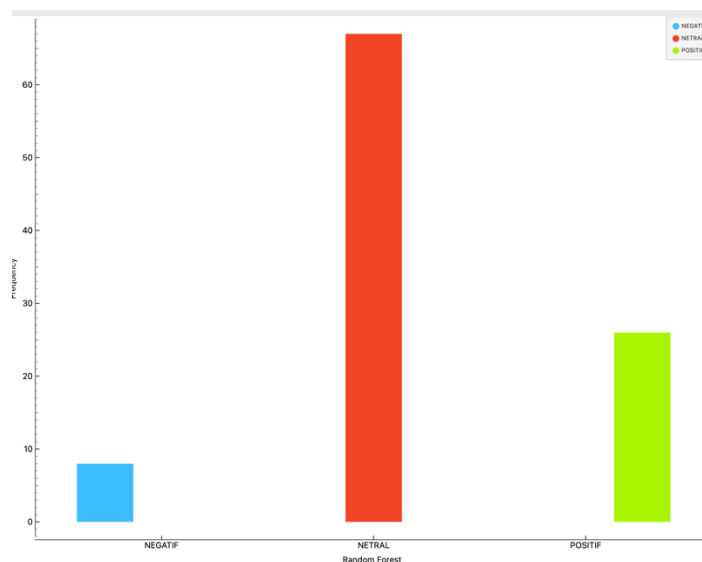
Gambar 3.3 Visualisasi Distribusi Sentimen Hasil Prediksi Model Naïve Bayes

Untuk model Random Forest dari hasil prediksi diperoleh sentiment netral sebanyak 59.29%, positif sebanyak 22.88% dan negative sebanyak 17.83%. berikut adalah visualiasi sentimentnya.



Gambar 3.4 Visualisasi Distribusi Sentimen Hasil Prediksi Model Random Forest

Prediksi juga dilakukan menggunakan data baru dengan model Random Forest, dimana diperoleh hasil prediksi untuk sentimen Netral sebesar 66.34%, Positif sebesar 25.74% serta Negatif sebesar 7.92%. berikut adalah Visualisasi distribusi sentimen dari prediksi menggunakan data baru pada model Random Forest.



Gambar 3.5 Visualisasi Distribusi Sentimen Hasil Prediksi Data Baru

4. KESIMPULAN

Dari total 5910 komentar yang digunakan pada proses *training* dan *testing*. Dimana diambil 80% untuk *training* serta 20% untuk *testing*. Dari kedua model diperoleh skor evaluasi AUC 0.874, CA 0.737, F-Score 0.731, Precision 0.739, Recall 0.737, MCC 0.569 untuk algoritma Random serta Naïve Bayes dengan skor evaluasi AUC 0.796, CA 0.626, F-Score 0.731, Precision 0.631, Recall 0.626, MCC 0.379. Sehingga dapat diperoleh algoritma yang memiliki akurasi tinggi pada analisis sentimen publik terhadap peluang timnas Indonesia lolos ke piala dunia 2026 yaitu Random Forest dengan akurasi 73%. Selain itu dari hasil prediksi menggunakan algoritma Random Forest pada data baru menghasilkan persentase untuk sentimen Positif sebesar 25.74%, Netral sebesar 66.34% serta sentimen Negatif sebesar 7.92%. Untuk penelitian selanjutnya bisa dilakukan dengan menggunakan data yang lebih bervariasi lagi, seperti menggabungkan data yang diperoleh dari media sosial Youtube, Twitter maupun Instagram.

5. UCAPAN TERIMA KASIH

Rasa terima kasih yang sebesar-besarnya kepada Bapak Dr. Abu Khalid Rivai, S.Si., M.Eng., dosen pengampu mata kuliah Metode Penelitian, atas bimbingan dan arahan beliau yang berharga selama proses penulisan artikel ilmiah ini. Terima kasih juga disampaikan kepada Program Studi PascaSarjana Teknik Informatika, Universitas Pamulang atas fasilitas dan lingkungan akademik yang mendukung penyelesaian penelitian ini.

6. DAFTAR PUSTAKA

- [1] M. Muthmainnah, R. Rudiman, and F. Yulianto, "Sentiment Analysis of the Public on the Deployment of Smart Robots in Indonesia Using the Naïve Bayes Method," *JSE Journal of Science and Engineering*, vol. 3, no. 2, pp. 85–89, Jan. 2025, doi: 10.30650/jse.v3i2.3887.
- [2] W. Widyastuti, D. Dasari, and S. H. Noer, "‘‘Apa yang Siswa Pikirkan Tentang Matematika?’ Analisis Sentimen pada Siswa Kelas Tinggi Sekolah Dasar,” *Jurnal Pendidikan Matematika Universitas Lampung*, vol. 12, no. 4, pp. 252–265, Dec. 2024, doi: 10.23960/mtk/v12i4.pp252-265.
- [3] A. Erlangga, Y. P. Astuti, E. Kartikadarma, S. Rakasiwi, and E. R. Subhiyakto, "Penggunaan Algoritma Naïve Bayes dengan Polarity Textblob untuk Analisis Sentimen pada Acara ASEAN CUP 2024 U-16 di Media Sosial Twitter,” *Jurnal Sains dan Teknologi Informasi*, vol. 3, no. 1, pp. 177–189, doi: 10.62951/switch.v3i1.357.
- [4] W. Alfiyani, D. Abdul Fatah, and F. Irhamni, "PENERAPAN ALGORITMA NAÏVE BAYES UNTUK ANALISIS SENTIMEN PADA MEDIA SOSIAL X TERHADAP PERFORMA TIM NASIONAL SEPAK BOLA INDONESIA DI ERA KEPEMIMPINAN SHIN TAE-YONG,” *Jurnal Mahasiswa Teknik Informatika*, vol. 9, no. 3, Jun. 2025.
- [5] M. Kholilullah and U. Hayati, "ANALISIS SENTIMEN PENGGUNA TWITTER(X) TENTANG PIALA DUNIA USIA 17 MENGGUNAKAN METODE NAIVE BAYES,” *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 1, Feb. 2024, [Online]. Available: <https://apify.com/quacker/twitter-scraper>
- [6] T. Dwi, P. 1✉, and D. Oktafiani, "Klasifikasi Sentimen Postingan Sosial Media Menggunakan Machine Learning Random Forest dan Naïve Bayes,”

- INNOVATIVE: Journal Of Social Science Research*, vol. 5, no. 1, pp. 2338–2347, 2025.
- [7] 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,” *Jurnal Accounting Information System (AIMS)*, vol. 8, no. 1, pp. 76–89, Mar. 2025, doi: 10.32627.
- [8] I. Made *et al.*, “Analisis Sentimen Berbasis Aspek dan Deteksi Emosi dari Ulasan Timnas Indonesia di Instagram,” *JOINTECOMS (Journal of Information Technology and Computer Science)* p-ISSN: 2798-284X, vol. 4, no. 4, pp. 2798–3862, Dec. 2024, doi: 10.47111/jointecoms.v4i4.
- [9] G. Gregorius, F. Djema, and O. Suria, “PUBLIC SENTIMENT ANALYSIS OF DANANTARA POLICY THROUGH SOCIAL MEDIA X USING SVM AND RANDOM FOREST ANALISIS SENTIMEN PUBLIK TERHADAP KEBIJAKAN DANANTARA MELALUI MEDIA SOSIAL X DENGAN METODE SVM DAN RANDOM FOREST,” *JURNAL INOVTEK POLBENG - SERI INFORMATIKA*, vol. 10, no. 2, p. 2025, Jul. 2025.
- [10] A. A. Firdaus, A. Id Hadiana, and A. K. Ningsih, “Klasifikasi Sentimen pada Aplikasi Shopee Menggunakan Fitur Bag of Word dan Algoritma Random Forest,” *R2J*, vol. 6, no. 5, pp. 1678–1683, Jul. 2024, doi: 10.38035/rrj.v6i5.
- [11] 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,” *Jurnal Accounting Information System (AIMS)*, vol. 8, no. 1, pp. 76–89, Mar. 2025, doi: 10.32627.
- [12] F. M. Fadillah, Y. Cahyana, and A. Fauzi, “BULLETIN OF COMPUTER SCIENCE RESEARCH Analisis Sentimen Masyarakat Terhadap Pembatasan BBM Pertalite Menggunakan Random Forest dan K-Nearest Neighbor,” *BULLETIN OF COMPUTER SCIENCE RESEARCH*, vol. 5, no. 4, pp. 340–352, Jun. 2025, doi: 10.47065/bulletincsr.v5i4.547.
- [13] K. Tri Putra, M. Amin Hariyadi, and C. Crysdiyan, “PERBANDINGAN FEATURE EXTRACTION TF-IDF DAN BOW UNTUK ANALISIS SENTIMEN BERBASIS SVM,” *JURNAL CAHAYA MANDALIKA*, vol. 3, no. 2, pp. 1449–1463, Nov. 2023.

- [14] C. Shifa Khoirunnisa, S. Anggai, T. Informatika, U. Pamulang, and K. Tangerang Selatan, “Analisis Sentimen Opini Masyarakat Terhadap Pemilu 2024 Melalui Media Sosial X Dengan Menggunakan Naive Bayes, K-Nearest Neighbor Dan Decision Tree,” *JURNAL ILMU KOMPUTER*, vol. 2, no. 2, pp. 157–168, Feb. 2024.
- [15] A. Yusupa and V. Tarigan, “PERBANDINGAN ALGORITMA MACHING LEARNING DALAM ANALISIS SENTIMEN MOBIL LISTRIK DI INDONESIA PADA MEDIA SOSIAL TWITTER/X,” *JIP (Jurnal Informatika Polinema)*, vol. 10, no. 4, pp. 479–489, Aug. 2024.