

## Analisis Sentimen Suporter terhadap Performa Tim Nasional Sepakbola Indonesia pada Turnamen Sea Games 2023 dengan Metode *Naive Bayes*

Al Diras Pradiptha<sup>1)</sup>, Farhan Siddiq Al Farisi<sup>2)</sup>, Muhammad Fahmi Pratama<sup>3)</sup>, Devi Aprianti Rimadhani Agustini<sup>4)</sup>, Munawir<sup>5)</sup>

<sup>1)</sup>Teknik Komputer, Kampus UPI di Cibiru, Universitas Pendidikan Indonesia  
Jl. Pendidikan No. 15, Cibiru, Bandung, Jawa Barat, 40625, Indonesia

<sup>1)</sup> aldiraspra@upi.edu

<sup>2)</sup> farhansiddiq@upi.edu

<sup>3)</sup> ppahme@upi.edu

<sup>4)</sup> deviaprianti@upi.edu

<sup>5)</sup> munawir@upi.edu

### Abstrak

Analisis pada sentimen suporter terhadap performa tim nasional sepakbola Indonesia menjadi topik yang menarik dikarenakan antusias pendukung yang begitu besar. Metode *Naive Bayes* dimanfaatkan dalam penelitian ini untuk menganalisis sentimen suporter berdasarkan data teks yang dikumpulkan dari media sosial melalui platform Twitter terkait pertandingan tim nasional sepakbola Indonesia. *Naive Bayes* dikenal dengan kemampuannya yang baik untuk mengklasifikasikan sebuah teks. Pengolahan data dilakukan dengan bantuan bahasa pemrograman Python dengan library yang telah disediakan. Analisis dilakukan dengan 265 tweet yang berkaitan dengan Tim Nasional Indonesia. Hasil penelitian ini membuktikan bahwa sentiment pengguna terklasifikasi sebagai puas sebanyak 70,94%, sementara tweet yang bersentimen negatif sebesar 29,06%. Nilai akurasi yang dihasilkan yaitu sebesar 81%.

**Kata kunci:** *naive bayes, analisis sentimen, crawling, suporter*

### Abstract

*Analysis of supporter sentiment towards the performance of the Indonesian national football team is an interesting topic due to the great enthusiasm of the supporters. The Naive Bayes method is used in this study to analyze supporter sentiment based on text data collected from social media via the Twitter platform related to the Indonesian national football team matches. Naive Bayes is known for its good ability to classify a text. Data processing is carried out with the help of the Python programming language with the library that has been provided. The analysis was carried out with 265 tweets related to the Indonesian National Team. The results of this study prove that user sentiment is classified as satisfied as much as 70.94%, while tweets with negative sentiment are 29.06%. The resulting accuracy value is 81%.*

**Keywords:** *naive bayes, sentiment analysis, crawling, supporter*

## 1. PENDAHULUAN

Salah satu acara olahraga terbesar di Asia Tenggara yang sangat digemari oleh penikmat sepakbola adalah SEA Games. Negara-negara anggota ASEAN bersaing untuk meraih kehormatan dan prestasi. Turnamen ini menjadi kesempatan penting bagi tim nasional sepakbola Indonesia untuk menunjukkan kemampuan dan potensi mereka di tingkat internasional. Performa dan keberhasilan timnas sepakbola Indonesia pada perhelatan turnamen SEA Games tahun 2023 akan menjadi perhatian besar bagi masyarakat umum dan para penggemar sepak bola di Indonesia.

Media sosial telah berkembang menjadi alat penting bagi masyarakat saat ini untuk berinteraksi, berbagi pendapat, dan mengungkapkan perasaan mereka tentang berbagai hal, salah satunya pada olahraga. Para penggemar sepak bola Indonesia menggunakan Twitter sebagai media populer untuk berbagi pendapat mereka tentang sepakbola, termasuk saling berdiskusi bagaimana tim nasional Indonesia di Turnamen SEA Games 2023. Dalam konteks ini, analisis sentimen bisa bermanfaat untuk memperluas wawasan dalam memahami perasaan dan sikap suporter terhadap performa timnas Indonesia. Dengan menggunakan metode crawling Twitter, kita dapat mengumpulkan data komentar-komentar yang diposting oleh suporter, menganalisisnya secara kualitatif dan kuantitatif, serta mengkategorikan komentar-komentar tersebut berdasarkan tingkat kepuasan suporter terhadap performa timnas.

Artikel ilmiah ini bertujuan untuk menyajikan analisis sentimen suporter Indonesia terhadap performa timnas sepakbola di turnamen SEA Games 2023 menggunakan metode crawling Twitter. Penelitian ini dilakukan agar pembaca dapat memperluas wawasan, khususnya mengenai persepsi, harapan, dan kekecewaan para suporter terhadap timnas Indonesia, serta memahami bagaimana komentar-komentar ini dapat mempengaruhi citra dan popularitas timnas. Hasil dari penelitian ini dapat memberikan masukan berharga bagi manajemen timnas Indonesia dan pihak terkait lainnya dalam meningkatkan kualitas dan performa timnas pada turnamen-turnamen berikutnya. Hasil penelitian juga dapat dimanfaatkan menjadi bahan kontribusi bagi bidang analisis sentimen dan pemrosesan bahasa alami dalam konteks olahraga.

Dalam artikel ini, kami akan mempresentasikan metodologi penelitian, analisis data, dan temuan utama yang dihasilkan dari penelitian ini. Analisis sentimen suporter Indonesia terhadap performa timnas sepakbola di turnamen SEA Games 2023 dengan metode crawling dan naive bayes ini dapat dijadikan bahan untuk memahami pandangan publik untuk prestasi sepakbola Indonesia serta mengarahkan langkah-langkah yang lebih baik dalam meningkatkan performa timnas di masa depan.

## 2. TINJAUAN PUSTAKA

### 2.1 Analisis Sistem

Analisis sentimen melibatkan proses identifikasi sentimen dan pengelompokan teks di dalam suatu dokumen dengan tujuan menentukan kategori dari sentiment di dalam dokumen atau kalimat tersebut, seperti positif, negatif, atau netral [1], [2]. Metode ini telah menjadi salah satu cabang penelitian yang signifikan dalam ilmu komputer, dengan penggunaan yang luas di jejaring sosial seperti Twitter untuk mengungkap persepsi publik [3], [4]. Analisis sentimen juga dikenal sebagai *opinion mining*, yang fokus pada pengklasifikasian pendapat yang bersifat positif atau negatif [5]. Dalam metode ini, *data mining* dilakukan dengan menganalisis, memproses, dan mengekstrak sebuah data yang terkait dengan sebuah identitas tertentu [6]. Proses ini melibatkan kalimat-kalimat pada forum *review*, post, ataupun blog. Metode ini juga melibatkan tahapan *preprocessing* data seperti tokenisasi, penghapusan stopwords, stemming data, *sentiment identification*, dan pengelompokan sentimen [7], [8], [9].

### 2.2 Twitter

Pada Maret 2006, Twitter diperkenalkan sebagai platform microblogging yang memungkinkan pengguna untuk mengirimkan pembaruan status dalam bentuk tweet dengan batasan 140 karakter. Sejak peluncurannya, Twitter telah berhasil mengumpulkan basis pengguna yang besar dan pada Juni 2011, jumlah pengguna mencapai lebih dari 300 juta. Dengan jumlah pengguna yang sangat banyak, Twitter telah memberikan pengaruh yang signifikan dalam masyarakat dengan memberikan wadah bagi individu untuk menyampaikan pendapat dan komentar terkait peristiwa yang sedang berlangsung atau telah terjadi.

### 2.3 Crawling

Crawling Twitter adalah metode pengumpulan data yang dilakukan dengan memanfaatkan API (*Application Programming Interface*) Twitter. Melalui API, pengguna dapat mengakses data publik yang diposting di platform Twitter. Metode crawling Twitter telah digunakan secara luas dalam

penelitian untuk mengumpulkan data sosial media, termasuk pendapat, pendapat, dan reaksi pengguna terhadap berbagai topik [10].

## 2.4 Preprocessing Text

*Preprocessing text* menjadi sebuah proses untuk mempersiapkan dan membersihkan data sebelum diolah lebih lanjut dalam tugas pemrosesan bahasa alami atau analisis teks. Tujuan utama dari preprocessing teks adalah untuk mengubah sebuah teks yang masih mentah menjadi bentuk yang jauh lebih baik, bersih, dan siap untuk digunakan dalam model pemrosesan bahasa alami atau aplikasi analisis teks lainnya.

## 2.5 Text Mining

Inti utama dari proses penambangan teks yaitu untuk mengidentifikasi pola, wawasan, dan pengetahuan baru dari teks yang dapat digunakan untuk pengambilan keputusan, pemahaman konten, dan penemuan informasi yang berharga. Text Mining, juga dikenal sebagai analisis teks atau penambangan teks, adalah proses ekstraksi informasi yang berarti dan berguna dari teks yang tidak terstruktur.

## 2.6 Metode Naïve Bayes

Metode Naïve Bayes, yang sering disebut juga sebagai teorema Bayes, merupakan suatu pendekatan yang menyediakan kerangka kerja untuk menghitung probabilitas suatu peristiwa berdasarkan informasi yang telah ada sebelumnya. Keunggulan dari Metode Naive Bayes terletak pada kecepatan komputasinya yang relatif tinggi serta kemampuannya dalam menangani jumlah fitur yang besar. Namun, metode ini memiliki kelemahan berupa asumsi independensi yang kuat antar fitur, yang dapat menyebabkan penurunan kinerja dalam beberapa situasi.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\sim A) \cdot P(\sim A)}$$

Gambar 1. Rumus teorema Naïve Bayes

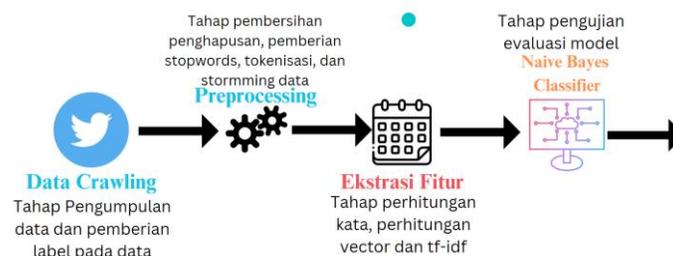
## 2.7 Implementasi

Penerapan processing dalam Naive Bayes sebagai penerapan atau implementasi metode tersebut dalam bahasa pemrograman yang dipakai. Dalam pengaplikasiannya menggunakan python dengan memakai salah satu tools, guna mempercepat mendapat kedua perbandingan dari data tersebut.

## 3. METODE PENELITIAN

### 3.1 Gambaran Umum Sistem

Studi ini menggunakan metode *crawling* Twitter untuk mengumpulkan data komentar-komentar dari supporter Timnas Sepakbola Indonesia terkait dengan performa Timnas di turnamen SEA Games 2023. Metode ini berfungsi untuk menganalisis data yang berasal dari publik yang diposting pada sebuah platform seperti Twitter. Gambar 2 adalah *step* yang dilalui dalam melakukan penelitian ini.



Gambar 2. Tahapan penelitian menggunakan Naïve Bayes

### 3.2 Pengumpulan Data Set

Data dikumpulkan dengan menggunakan metode crawling Twitter. Melalui API (*Application Programming Interface*) Twitter, data tweet yang terkait dengan performa timnas sepakbola Indonesia di turnamen SEA Games 2023 diambil secara real-time. Sampel tweet yang mencakup pendapat, komentar, dan sentimen supporter diambil untuk analisis lebih lanjut.

### 3.3 Data Preprocessing

Setelah data terkumpul, langkah berikutnya adalah melakukan pra-pemrosesan data untuk membersihkan dan menyiapkannya sebelum analisis sentimen dilakukan. *Preprocessing* mencakup penghilangan karakter khusus, konversi teks menjadi huruf kecil, penghapusan tautan URL, penghapusan *stopwords*, dan *stemming*.

### 3.4 Pengelompokkan Data

1. Puas: Komentar-komentar dalam kategori ini biasanya menyatakan bahwa timnas telah bermain dengan baik dan memenuhi harapan sebagian besar supporter. Mereka mungkin menyebut pencapaian timnas sebagai bagus, solid, atau layak diapresiasi. Kategori ini dapat mencakup komentar-komentar yang mengakui kemajuan timnas dan konsistensi yang baik dalam pertandingan.
2. Tidak Puas: Komentar-komentar dalam kategori ini biasanya sangat kritis terhadap timnas dan menggambarkan performa mereka sebagai buruk atau sangat mengecewakan. Mereka mungkin menyebut timnas sebagai gagal atau tidak layak mewakili negara. Kategori ini dapat mencakup komentar-komentar yang menyatakan bahwa timnas memiliki masalah serius dan perlu melakukan perubahan besar.

### 3.5 Pelatihan Model Naïve Bayes

Sebelum melakukan klasifikasi sentimen, model Naïve Bayes perlu dilatih menggunakan data latih yang telah dikategorikan secara manual. Data latih terdiri dari tweet-tweet dengan sentimen yang telah diketahui. Naïve Bayes mempelajari pola karakteristik dari data latih agar mampu mengklasifikasikan sentimen pada data uji dengan akurat.

### 3.6 Klasifikasi Sentimen

Setelah model Naïve Bayes dilatih, dilakukan proses klasifikasi sentimen pada data uji. Setiap tweet dalam data uji diklasifikasikan ke dalam salah satu kategori sentimen berdasarkan probabilitas yang dihasilkan oleh model. Klasifikasi sentimen ini memungkinkan untuk menentukan tingkat kepuasan supporter terhadap performa timnas di turnamen SEA Games 2023.

### 3.7 Evaluasi dan Interpretasi Hasil

Hasil klasifikasi sentimen dinilai menggunakan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Hasil analisis sentimen juga diinterpretasikan untuk memberikan pemahaman yang lebih mendalam mengenai preferensi dan persepsi pendukung terhadap performa tim nasional sepak bola Indonesia pada turnamen SEA Games 2023.

## 4. PEMBAHASAN

Pada penelitian ini digunakan beberapa kata kunci diantaranya ‘timnas indonesia’, ‘performa timnas’, ‘timnas sea games’, ‘timnas indonesia’, dan tagar #timnasday. Proses crawling data dilakukan dengan bahasa javascript, framework node js, dengan memanfaatkan token Twitter *API (Application Programming Interface)* dengan memakai tools *Command Prompt* untuk memperoleh informasi yang ingin didapatkan. Proses ini menghasilkan 265 *tweet* yang berkaitan dengan kata kunci yang telah ditentukan.

```

C:\Windows\system32\cmd.exe
Filling in keywords: performa timnas seagames

Got some tweets, saving to file...
No additional tweet, scrolling more...

Got some tweets, saving to file...
Your tweets saved to: C:\Users\rasxp\tweets-data\performa_timnas_seagames_22-05-2023_13-29-36.csv
Total tweets saved: 0

Got some tweets, saving to file...
Your tweets saved to: C:\Users\rasxp\tweets-data\performa_timnas_seagames_22-05-2023_13-29-36.csv
Total tweets saved: 21

Got some tweets, saving to file...
Your tweets saved to: C:\Users\rasxp\tweets-data\performa_timnas_seagames_22-05-2023_13-29-36.csv
Total tweets saved: 39

Got some tweets, saving to file...
Your tweets saved to: C:\Users\rasxp\tweets-data\performa_timnas_seagames_22-05-2023_13-29-36.csv
Total tweets saved: 57
Already got 57 tweets, done scrolling...
npm notice
npm notice New minor version of npm available! 9.5.1 -> 9.6.7
npm notice Changelog: https://github.com/npm/cli/releases/tag/v9.6.7
npm notice Run npm install -g npm@9.6.7 to update!
npm notice
C:\Users\rasxp>
    
```

Gambar 3. Proses crawling data

Gambar 3 memperlihatkan proses pengambilan data (crawling) menggunakan Node.js melalui *Command Prompt* dengan memanfaatkan kode akses token yang diperoleh dari Twitter API (*Application Programming Interface*). Hasil data yang diambil menggunakan Node.js ini ditampilkan pada Tabel 1, yang mencakup atribut seperti username dan teks.

Tabel 1. Data crawling

No	Tweet	Username
1	Performa permainan Timnas ini di Sea Games kali ini sangat baik dari segi attack, passing, defense, clearance, dan juga mentalitas.	@erwinyokoyama
2	@heightgakena @PSSI Benar, mulai menurun performa timnas semoga bisa segera diperbaiki dan tetap fokus sampai juara	@danangiffah

Proses pengelompokan data dilakukan untuk menentukan tweet tersebut termasuk ke dalam kategori “Puas” atau “Tidak puas”. Tweet yang masuk ke dalam kategori “Puas” adalah tweet yang menyebut pencapaian timnas sebagai bagus, solid, atau layak diapresiasi. Sedangkan kategori “Tidak puas” adalah tweet yang berisi keluhan, kritik, ataupun emosi negatif seperti amarah, kesal, dan kecewa.

Tabel 2. Pengelompokan data

No	Tweet	Kategori
1	Performa permainan Timnas ini di Sea Games kali ini sangat baik dari segi attack, passing, defense, clearance, dan juga mentalitas.	Puas
22	@heightgakena @PSSI Benar, mulai menurun performa timnas semoga bisa segera diperbaiki dan tetap fokus sampai juara	Tidak Puas

Untuk menghilangkan ketidakjelasan, proses konversi data dilakukan pada tahap pra-pemrosesan, dengan tujuan untuk menghapus atribut-atribut yang tidak relevan, seperti URL, mention, username, RT, atau tagar [5]. Tahap *preprocessing* dibutuhkan sebelum lanjut implementasi algoritma. Pada tahap ini digunakan bantuan *Python* untuk melakukan *preprocessing* data. Pada tahap ini dilakukan penghapusan URL untuk menghilangkan tautan yang terdapat dalam tweet, seperti yang ditunjukkan pada Tabel 3.

Tabel 3. Preprocessing - menghilangkan mention dan link

No	Sebelum	Sesudah
1	Enaknya Tidur Nyenyak Pemain Timnas Indonesia Usai Menang 5-2 <a href="https://t.co/2SihkNyTu1">https://t.co/2SihkNyTu1</a> lewat @YouTube #TimnasIndonesia #thailandelection2023	Enaknya Tidur Nyenyak Pemain Timnas Indonesia Usai Menang 5-2

Tweet telah melewati tahap *preprocessing* di mana langkah-langkah dilakukan untuk menghilangkan kata-kata mention (yang dimulai dengan '@') dan link yang ada dalam teks. Hasilnya, tweet telah diubah menjadi teks yang telah terbebas dari kata-kata mention dan link yang sebelumnya ada.

Tabel 4. Case folding

No	Sebelum	Sesudah
1	Performa permainan Timnas ini di Sea Games kali ini sangat baik dari segi attack, passing, defense, clearance, dan juga mentalitas.	performa permainan timnas ini di sea games kali ini sangat baik dari segi attack, passing, defense, clearance, dan juga mentalitas.

Tahap *case folding* dilakukan untuk mengubah semua teks dalam tweet menjadi huruf kecil. Hal tersebut akan mempermudah analisis teks, sehingga ketika semua huruf menjadi kecil maka akan mempermudah pencocokan teks menjadi lebih efisien.

Tabel 5. Tokenizing

No	Sebelum	Sesudah
1	performa permainan timnas ini di sea games kali ini sangat baik dari segi attack, passing, defense, clearance, dan juga mentalitas.	performa, permainan, timnas, ini, di, sea, games, kali, ini, sangat, baik, dari, segi, attack, passing, defense, clearance, dan, juga, mentalitas

Tokenisasi dilakukan untuk memecah teks menjadi unit-unit kecil yang disebut token. Proses ini penting dalam membangun kumpulan data yang terstruktur, memungkinkan teks diuraikan menjadi bagian-bagian yang dapat digunakan untuk melatih model, menghitung probabilitas kata, atau menganalisis pola linguistik.

Tabel 6. Filtering

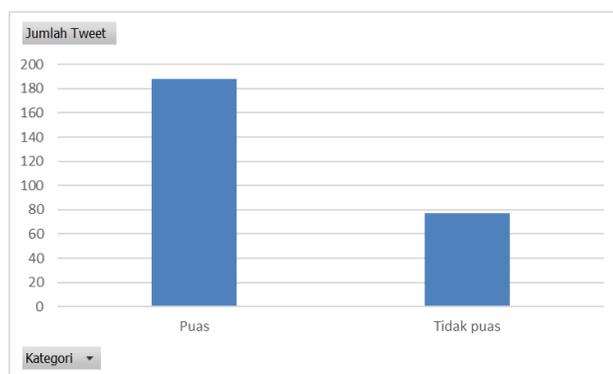
Tweet	Hasil
semangat	Puas
alhamdulillah	Puas
mantap	Puas
miris	Tidak puas

*Filtering* bertujuan untuk menyaring atau menghilangkan elemen-elemen yang tidak diinginkan atau tidak relevan dalam teks. Tujuan utama dari *filtering* dalam proses *preprocessing* adalah untuk membersihkan teks dari informasi yang tidak bermanfaat atau mengganggu, sehingga dapat meningkatkan kualitas analisis atau pemodelan teks yang dilakukan. Frekuensi kata divisualisasikan dalam bentuk *word cloud* untuk kategori "Puas" dan "Tidak Puas," seperti yang ditampilkan pada Gambar 4.



Gambar 4. *Wordcloud* puas dan tidak puas

Penelitian ini mengembangkan model analisis sentimen dengan memanfaatkan data latih yang telah diproses sebelumnya. Model tersebut diuji pada data pengujian untuk mengetahui sentimen terhadap performa Timnas Indonesia pada turnamen Sea Games 2023. Teks yang digunakan dalam analisis pada penelitian ini berupa tweet berbahasa Indonesia dengan kosakata yang sepenuhnya menggunakan bahasa Indonesia [2]. Hasil analisis menunjukkan bahwa sentimen positif sebesar 70,94%, sementara sentimen negatif sebesar 29,06%. Hasil pra-pemrosesan data kemudian dikelompokkan untuk menghitung frekuensi setiap kategori data atau tweet.



Gambar 5. Hasil pengukuran sentimen

Gambar 5 menunjukkan hasil pengukuran sentimen dengan 180 lebih tweet merupakan tweet yang menggambarkan kepuasan dari suporter dan 77 tweet merupakan tweet tidak puas terhadap performa Tim Nasional Indonesia. Hasil tersebut perlu diklasifikasikan ulang untuk memperoleh akurasi perhitungan yang lebih tepat.

Selanjutnya, agar data teks dapat diproses oleh algoritma klasifikasi, perlu mengubahnya menjadi bentuk numerik. Proses ini disebut *vectorizing* atau vektorisasi. Salah satu teknik vektorisasi yang banyak digunakan adalah pembobotan kata. Dalam penelitian ini, algoritma pembobotan kata yang dikenal sebagai TF-IDF digunakan. Di Python, algoritma ini diimplementasikan melalui perintah `TfidfVectorizer()`.

```
proses tf-idf

from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer

tf = TfidfVectorizer()
text_tf = tf.fit_transform(data_clean['Tweet'].astype('U'))
text_tf

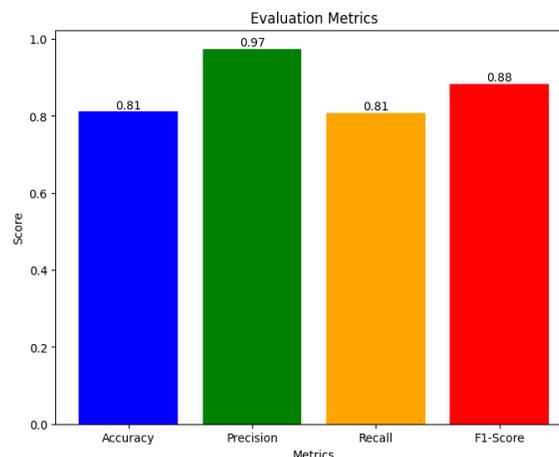
✓ 02s

<265x140 sparse matrix of type '<class 'numpy.float64''>'
with 267 stored elements in Compressed Sparse Row format>
```

Gambar 6. Proses TF-IDF

Langkah selanjutnya adalah membangun model klasifikasi menggunakan algoritma Multinomial Naïve Bayes yang diimplementasikan dengan library Scikit-learn (Sklearn) pada Python. Kinerja model kemudian dievaluasi untuk mengukur tingkat akurasi dalam mengklasifikasikan data. Evaluasi dilakukan menggunakan confusion matrix sebagai metode analisis dalam kasus klasifikasi. Parameter evaluasi yang digunakan meliputi akurasi (accuracy), presisi, recall, dan F1-score.

1. Pengujian Akurasi (Accuracy mengukur sejauh mana model klasifikasi dapat memberikan prediksi yang benar, sekaligus menunjukkan tingkat keakuratan keseluruhan dari model tersebut.
2. Pengujian Presisi (Precision) mengukur tingkat akurasi model dalam memprediksi kelas positif dibandingkan dengan total prediksi yang dilakukan untuk kelas positif. Presisi mencerminkan seberapa andal prediksi model terhadap kelas positif.
3. Pengujian Recall mengukur kemampuan model klasifikasi dalam mengenali kembali data bernilai positif, memberikan indikasi seberapa baik model mendeteksi informasi yang relevan.
4. Pengujian F1-Score menggabungkan presisi dan recall menggunakan harmonic mean untuk memberikan gambaran seimbang tentang performa model dalam kedua aspek tersebut. Nilai F1-Score yang tinggi (maksimum 1.0) menunjukkan bahwa model memiliki presisi dan recall yang seimbang serta berkinerja baik.



Gambar 7. Hasil kinerja model klasifikasi

Nilai akurasi sebesar 0,81 menunjukkan bahwa model klasifikasi mampu melakukan prediksi dengan tingkat keberhasilan sebesar 81%. Artinya, model ini dapat mengklasifikasikan sebagian besar data secara benar. Presisi yang mencapai 97% menandakan bahwa model memiliki tingkat

akurasi yang tinggi dalam memprediksi kelas positif (kepuasan supporter). Presisi mengukur seberapa tepat model dalam memprediksi kelas positif dibandingkan dengan total prediksi yang dilakukan untuk kelas tersebut. Nilai recall sebesar 81% mengindikasikan bahwa model berhasil dengan baik dalam menemukan kembali data yang termasuk dalam kategori positif (kepuasan supporter). Recall mengukur kemampuan model untuk mengenali informasi yang relevan dalam dataset. F1-score sebesar 88% merupakan rata-rata harmonis dari presisi dan recall, yang memberikan gambaran mengenai keseimbangan antara kedua metrik tersebut. Nilai F1-score yang tinggi menunjukkan bahwa model memiliki keseimbangan yang baik antara presisi dan recall secara keseluruhan.

Tabel 7. Confusion matrix

	Prediction Yes	Prediction No
True Yes	38	1
True No	9	5

Untuk memastikan bahwa nilai accuracy benar, dilakukan perhitungan manual dimana,  
 $Accuracy = (TP + TN) / (TP + FP + TN + FN) \times 100\%$   
 $= (38 + 5) / (38 + 1 + 9 + 5) \times 100\%$   
 $= 0,81$

## 5. KESIMPULAN

Penelitian ini menerapkan algoritma Naive Bayes untuk menguji klasifikasi sentimen. Data dikumpulkan melalui proses crawling Twitter dengan menggunakan kata kunci seperti 'timnas indonesia', 'performa timnas', 'timnas sea games', serta tagar #timnasday yang disaring berdasarkan kata kunci 'timnas' dan 'sea games' pada tweet dan komentar berbahasa Indonesia, dimulai sejak Mei 2023. Pengolahan data dilakukan menggunakan algoritma Naive Bayes dengan bantuan bahasa pemrograman Python. Hasil penelitian menunjukkan bahwa selama periode tersebut, sentimen positif mencapai 70,94%, sedangkan sentimen negatif sebesar 29,06%. Dari total 265 tweet, lebih dari 180 tweet menggambarkan kepuasan supporter dan 77 tweet menunjukkan ketidakpuasan mereka. Berdasarkan temuan ini, dapat disimpulkan bahwa model klasifikasi Naive Bayes menunjukkan performa yang baik dalam mengklasifikasikan sentimen kepuasan supporter terhadap performa tim nasional Indonesia. Meskipun terdapat perbedaan yang signifikan antara presisi dan recall, nilai F1-score yang tinggi menunjukkan bahwa model memiliki keseimbangan yang baik antara presisi dan recall dalam melakukan prediksi. Namun, penting untuk mempertimbangkan konteks dan karakteristik data yang digunakan saat menginterpretasi hasil evaluasi ini.

## DAFTAR PUSTAKA

- [1] Samsir *et al.*, "Naives Bayes Algorithm for Twitter Sentiment Analysis," *J Phys Conf Ser*, vol. 1933, no. 1, pp. 1–6, 2021, doi: 10.1088/1742-6596/1933/1/012019.
- [2] D. R. Lazuardi, T. A. Munandar, H. Harsiti, Z. Mutaqin, and R. N. Hays, "Sentiment Analysis of Public Opinions on the Welfare of Honorary Educators using Naive Bayes," *IOP Conf Ser Mater Sci Eng*, vol. 830, no. 3, pp. 1–6, 2020, doi: 10.1088/1757-899X/830/3/032018.
- [3] R. Ferryawan, Kusriani, and F. W. Wibowo, "Analisis Sentimen Wisata Jawa Tengah Menggunakan Naive Bayes," *Jurnal INFORMA Politeknik Indonusa Surakarta*, vol. 5, no. 3, pp. 55–60, 2019.
- [4] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *JURNAL*

- 
- MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 3, pp. 650–658, 2020, doi: 10.30865/mib.v4i3.2181.
- [5] C. B. Saputra, A. Muzakir, and D. Udariansyah, “Analisis Sentimen Masyarakat Terhadap #2019gantipresiden Berdasarkan Opini Dari Twitter Menggunakan Metode Naive Bayes Classifier,” in *Bina Darma Conference on Computer Science*, Palembang: Bina Darma University, 2019, pp. 403–413.
- [6] S. Bhatia, M. Sharma, and K. K. Bhatia, “Sentiment Analysis and Mining of Opinions,” in *Internet of Things and Big Data Analytics Toward Next-Generation Intelligence*, vol. 30, N. Dey, A. E. Hassanien, C. Bhatt, A. S. Ashour, and S. C. Satapathy, Eds., Cham: Springer, 2018, pp. 503–523. doi: 10.1007/978-3-319-60435-0\_20.
- [7] S. B. Bhonde and J. R. Prasad, “Sentiment Analysis-Methods, Applications & Challenges,” *International Journal of Electronics Communication and Computer Engineering*, vol. 6, no. 6, pp. 634–640, 2015.
- [8] R. Watrionthos, S. Suryadi, D. Irmayani, M. Nasution, and E. F. S. Simanjorang, “Sentiment Analysis Of Traveloka App Using Naïve Bayes Classifier Method,” *International Journal of Scientific & Technology Research*, vol. 8, no. 7, pp. 786–788, 2019.
- [9] R. Rasenda, H. Lubis, and R. Ridwan, “Implementasi K-NN Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 2, pp. 369–376, 2020, doi: 10.30865/mib.v4i2.2051.
- [10] F. Morstatter, J. Pfeffer, H. Liu, and K. Carley, “Is the Sample Good Enough? Comparing Data from Twitter’s Streaming API with Twitter’s Firehose,” in *Proceedings of the International AAAI Conference on Web and Social Media*, E. Kiciman, N. Ellison, B. Hogan, P. Resnick, and I. Soboroff, Eds., Cambridge: The AAAI Press, 2013, pp. 400–408. doi: 10.1609/icwsm.v7i1.14401.

#### **Biodata Penulis**

***Al Diras Pradiptha***, Lahir di Bandung, Jawa Barat, 05 Agustus 2002. Sedang menempuh pendidikan semester 4 di Universitas Pendidikan Indonesia, Jurusan Teknik Komputer.

***Farhan Siddiq Al Farisi***, Lahir di Kota Tasikmalaya, Jawa Barat, Indonesia, pada tanggal 02 September 2002, latar pendidikan SDN Sukasari 1 Tasikmalaya, SMPN 1 Tasikmalaya, SMAN 1 Tasikmalaya dan UPI Kamda Cibiru Teknik Komputer 2021- Sekarang.

***Muhammad Fahmi Pratama***, Lahir di Lebak, Banten, 11 Maret 2003. Mahasiswa Teknik Komputer tahun kedua di Universitas Indonesia Kampus Daerah di Cibiru.

***Devi Aprianti Rimadhani Agustini***, Dosen Program Studi Teknik Komputer Universitas Indonesia Kampus Daerah di Cibiru.

***Munawir***, Dosen Program Studi Teknik Komputer Universitas Indonesia Kampus Daerah di Cibiru.