

Analisis Sentimen Kemenangan Timnas U-23 Menggunakan Naive Bayes

Diterima:

10 Juni 2024

Revisi:

10 Juli 2024

Terbit:

1 Agustus 2024

^{1*}**Anwar Muzaki**, ²**Risky Aswi Ramadhani**, ³**Patmi Kasih**
¹⁻³*Prodi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer,*
Universitas Nusantara PGRI Kediri
^{1*}*zackcorporation2020@gmail.com*, ²*risky_aswi@unpkediri.ac.id*,
³*fatkasih@gmail.com*

Abstrak—Media sosial telah menjadi bagian integral dalam kehidupan masyarakat modern, memfasilitasi berbagi informasi dan interaksi secara real-time. Twitter, sebagai platform yang sangat populer, memungkinkan pengguna untuk menyampaikan pendapat dan reaksi mereka secara langsung. Dalam konteks kemenangan Tim Nasional Indonesia pada semifinal Piala Asia U-23 AFC 2024, reaksi masyarakat terhadap peristiwa ini dieksplorasi melalui analisis sentimen menggunakan algoritma Naive Bayes. Tujuan penelitian adalah untuk mengeksplorasi tingkat dukungan dan loyalitas masyarakat Indonesia, khususnya pengguna Twitter, terhadap kemenangan tersebut. Hasil penelitian menunjukkan bahwa penggunaan Inset Lexicon dalam proses pelabelan cukup efektif karena berbasis bahasa Indonesia. Algoritma Naive Bayes terbukti memberikan akurasi yang baik dalam klasifikasi sentimen. Visualisasi hasil menunjukkan bahwa mayoritas sentimen masyarakat Indonesia terhadap kemenangan Tim Nasional Indonesia sangat positif, dengan 64,2% sentimen positif dan hanya 35,8% sentimen negatif. Ini menggambarkan tingkat dukungan yang tinggi dan loyalitas terhadap tim nasional dalam peristiwa tersebut.

Kata Kunci—Analisis Sentimen; Naive Bayes; Timnas; Inset

Abstract— *Social media has become an integral part of modern society, facilitating real-time information sharing and interaction. Twitter, as a highly popular platform, allows users to express their opinions and reactions in real time. In the context of the Indonesian National Team's victory in the semifinals of the AFC U-23 Asian Cup 2024, people's reactions to this event were explored through sentiment analysis using the Naive Bayes algorithm. The purpose of the study was to explore the level of support and loyalty of the Indonesian public, particularly Twitter users, towards the victory. The results show that the use of Inset Lexicon in the labeling process is quite effective because it is based on Indonesian language. The Naive Bayes algorithm proved to provide good accuracy in sentiment classification. Visualization of the results shows that the majority of Indonesian sentiment towards the Indonesian National Team's victory is very positive, with 64.2% positive sentiment and only 35.8% negative sentiment. This illustrates the high level of support and loyalty towards the national team in the event.*

Keywords—*Sentiment Analysis; Naive Bayes; Timnas; Inset*

This is an open access article under the CC BY-SA License.



Penulis Korespondensi:

Anwar Muzaki,
Fakultas Teknik dan Ilmu Komputer,
Universitas Nusantara PGRI Kediri,

Email: zackcorporation2020@gmail.com
ID Orcid: [<https://orcid.org/register>]
Handphone: 085366258996

I. PENDAHULUAN

Media sosial telah menjadi bagian integral dari kehidupan masyarakat modern, terutama dalam era digital saat ini. Dalam beberapa tahun terakhir, media sosial telah berkembang dengan sangat cepat[1], memungkinkan orang-orang untuk berbagi informasi, berkomunikasi, dan berinteraksi dengan mudah. Twitter sebagai media sosial yang sangat populer dan digunakan oleh masyarakat, memungkinkan pengguna untuk berbagi pendapat atau opininya secara real-time[2].

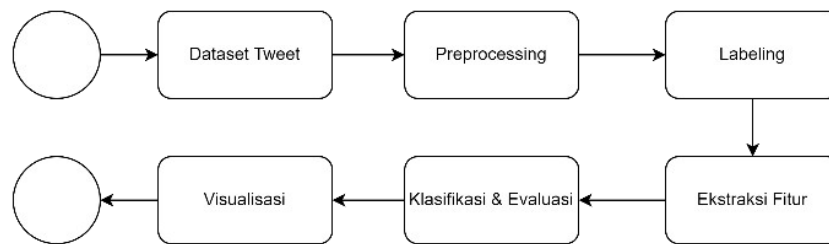
Banyak opini dan reaksi yang disampaikan oleh masyarakat melalui Twitter, opini tersebut tidak lepas dari penilaian positif maupun penilaian negatif. Seperti pada pertandingan semifinal piala Asia U-23 AFC, Indonesia menang dengan skor 11-10 pada babak penalti. Kemenangan tersebut tentu memunculkan banyak reaksi dari masyarakat khususnya pendukung setia Timnas di Twitter. Pemanfaatan *machine learning* untuk pengolahan data besar termasuk pada tweet yang diungkapkan oleh masyarakat melalui media sosial sangat diperlukan. Dengan menggunakan *machine learning*, pengolahan tweet melalui analisis sentimen menjadi mudah dan efisien.

Seperti pada penelitian yang dilakukan oleh[3][4][5][6], penggunaan algoritma *machine learning* yaitu Naive Bayes untuk analisis sentimen dengan menggunakan studi kasus yang berbeda mendapatkan hasil akurasi yang baik. Pada penelitian ini, peneliti mencoba mengolah data Twitter untuk dilakukan analisis sentimen mengenai kemenangan Tim Nasional Indonesia pada semifinal Piala Asia U-23 AFC 2024 menggunakan algoritma *naive bayes*. Algoritma *naive bayes* dipilih karena algoritma tersebut tergolong sederhana[7] dan mempunyai akurasi yang baik pada hasil klasifikasi[8].

Tujuan dari penelitian ini adalah untuk mengetahui dukungan masyarakat Indonesia khususnya pengguna media sosial Twitter terhadap kemenangan Tim Nasional Indonesia pada pertandingan semifinal Piala Asia U-23 AFC 2024 melalui analisis sentimen.

II. METODE

Penelitian ini menggunakan pendekatan kuantitatif, dimana pendekatan kuantitatif merupakan proses pengumpulan data, pengolahan data, proses analisis, serta penyajian data menggunakan perhitungan statistik, matematika dan komputasi cerdas[9][10]. Prosedur pada penelitian menggunakan metode *waterfall* yang merupakan tahapan penyelesaian secara berurutan agar tidak terjadi kesalahan pada setiap fase[10]. Gambaran proses dari sistem dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alir Analisis Sentimen

- 1.
- 2.

2.1. Dataset

Data yang digunakan pada penelitian ini berupa komentar / tweet dari media sosial Twitter. Data yang diambil merupakan tweet ataupun komentar pengguna media sosial Twitter terkait kemenangan Tim Nasional Indonesia pada semifinal Piala Asia U-23 AFC 2024.

2.2. Preprocessing

Preprocessing merupakan suatu proses pengubahan bentuk data yang belum terstruktur menjadi data yang terstruktur sesuai dengan kebutuhan, untuk proses mining yang lebih lanjut, seperti analisis sentimen, peringkatan, clustering dokumen, dan lain-lain. Proses ini melibatkan beberapa langkah, seperti pembersihan data, tokenisasi, menghapus stopword, dan stemming, untuk menghasilkan sebuah *set term index* yang dapat mewakili dokumen dan meningkatkan kinerja sistem[11].

2.3. Labelling

Tahapan *labelling* atau pelabelan pada dataset yang telah di *preprocessing* yaitu tahap menghitung nilai polaritas dari kata / kalimat untuk diberi label. Perhitungan ini menggunakan Inset (Indonesia Sentimen Lexicon) yang merupakan *lexicon* kata berbahasa Indonesia[12]. Tujuan dari pelabelan ini nantinya dijadikan sebagai fitur target pada proses klasifikasi Naive Bayes.

2.4. Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan salah satu metode ekstraksi fitur yang banyak digunakan dalam melakukan vektorisasi pada data, cara kerja pada metode ini yaitu melakukan pembobotan data *term* pada suatu *document*[13]. Tahapan dari ekstraksi fitur menggunakan TF-IDF adalah sebagai berikut:

- 1) Menghitung nilai *Term Frequency* (TF): yaitu perhitungan total dari kemunculan *term* pada suatu *document*.

- 2) Menghitung nilai *Document Frequency* (DF): yaitu perhitungan *term* muncul pada semua *document*.
- 3) Menghitung *Inverse Document Frequency* (IDF): yaitu perhitungan dari hasil kebalikan nilai DF. Perhitungan ini menggunakan persamaan (1). N merupakan total keseluruhan *document*.

$$IDF_t = \log_2 \left(\frac{N}{df_t} \right) \quad (1)$$

- 4) Menghitung TF-IDF: perkalian hasil TF dan juga IDF menggunakan persamaan (2).

$$Tf - IDF = tf \times \log_2 \left(\frac{N}{df_t} \right) \quad (2)$$

2.5. Naive Bayes

Naive Bayes adalah sebuah algoritma yang dapat digunakan untuk pengklasifikasian data binary dan multiclass. Klasifikasi akan menggunakan tipe Multinomial Naive Bayes. Perhitungan probabilitas dalam Naive Bayes secara umum dapat dirumuskan pada persamaan 3[14].

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (3)$$

Keterangan:

X = Data class yang belum diketahui

H = Hipotesis X Kelas spesifik

P(H|X) = Probabilitas Hipotesis H berdasarkan X

P(X|H) = Probabilitas X berdasar kondisi

P(H) = Probabilitas H

P(X) = Probabilitas X

2.6. Evaluasi

Confusion Matrix adalah sebuah teknik yang digunakan dalam *data mining* dan *machine learning* untuk menghitung seberapa baik sebuah model dapat memprediksi label dari sebuah data. Teknik ini sering digunakan dalam evaluasi model *classification* yang mana model harus memprediksi label dari sebuah data[15]. Pada confusion matrix di hitung akurasi, *precision*, dan *recall*.

III. HASIL DAN PEMBAHASAN

3.

3.1. Dataset Tweet

Pengumpulan dataset menggunakan bantuan bahasa Python dan API tweet-harvest[16]. Pengumpulan dataset tersebut menghasilkan dataset berjumlah 2000 data yang kemudian di ekspor dalam bentuk CSV.

3.2. Preprocessing

Dari dataset tweet tersebut dimasukkan pada sistem *preprocessing*. Bentuk dataset dari proses preprocessing ini menghasilkan dataset bersih. Berikut merupakan contoh dari dataset yang telah di *preprocessing* disajikan pada Tabel 1.

Tabel 1. Hasil Preprocessing

<i>Inde x</i>	<i>Tweet</i>	<i>Preprocessing</i>
0	Eh nyata nathan jadi kapten #TimnasDay #AFCU23AsianCup #AFCU23	['eh', 'nyata', 'nathan', 'kapten']
1	NGAKAK BANGET LIAT BIJINYA DITENDANG indonesia vs uzbekistan Afc u23 #TikTok #AFCU23AsianCup #Indonesia #viral #meme https://t.co/eQc4iac5E9	['tertawa', 'banget', 'lihat', 'biji', 'tendang', 'indonesia', 'vs', 'uzbekistan', 'afc']
2	Mendukung tim nasional bukan hanya saat mereka menang. Sebagai penggemar dukungan berkelanjutan kita dalam suka maupun duka adalah yang paling berharga. #TimnasDay #AFCU23AsianCup #AsianCupU23	['dukung', 'tim', 'nasional', 'menang', 'gemar', 'dukung', 'lanjut', 'suka', 'duka', 'harga']
3	Setelah tadi malam emosi naik turun lihat keputusan keputusan #wasit #AFCU23AsianCup terhadap ðŸ†@ðŸ†© #TimnasIndonesia kayanya nanti harus pelampiasan dengan rewatch #CivilWarMovie di bioskop deh https://t.co/THK2H6FzY3	['malam', 'emosi', 'turun', 'lihat', 'putus', 'putus', 'kaya', 'lampias', 'rewatch', 'bioskop', 'deh']
4	ðŸ†@ðŸ†© Apakah sepakbola hanya permainan? Di Indonesia ini adalah panggung drama sosial politik dan emosi yang kaya. Lanjut... #TimnasDay #AFCU23AsianCup #AsianCupU23	['sepakbola', 'main', 'indonesia', 'panggung', 'drama', 'sosial', 'politik', 'emosi', 'kaya']

3.3. Labelling

Tahapan selanjutnya yaitu pelabelan dengan menghitung jumlah polaritas dari setiap komentar / tweet. Jika polaritas lebih dari sama dengan 0 maka diberi label positif, jika polaritas bernilai kurang dari 0 maka diberi label negatif. Hasil *labelling* dari 2000 dataset menghasilkan 1134 komentar positif dan 866 komentar negatif. Berikut merupakan sajian data hasil pelabelan ditunjukkan pada Tabel 2.

Tabel 2. Hasil Pelabelan

<i>Inde x</i>	<i>Preprocessing</i>	<i>Skor Polaritas</i>	<i>Label</i>
0	['eh', 'nyata', 'nathan', 'kapten']	2	Positif
1	['tertawa', 'banget', 'lihat', 'biji', 'tendang', 'indonesia', 'vs', 'uzbekistan', 'afc']	5	Positif

2	['dukung', 'tim', 'nasional', 'menang', 'gemar', 'dukung', 'lanjut', 'suka', 'duka', 'harga']	8	Positif
3	['malam', 'emosi', 'turun', 'lihat', 'putus', 'putus', 'kaya', 'lampias', 'rewatch', 'bioskop', 'deh']	-5	Negatif
4	['sepakbola', 'main', 'indonesia', 'panggung', 'drama', 'sosial', 'politik', 'emosi', 'kaya']	-2	Negatif

3.4. Ekstraksi Fitur

Setelah data selesai pada proses *preprocessing* dan *labelling*, kemudian diproses ekstraksi fitur, ekstraksi fitur pada penelitian ini menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF).

3.5. Klasifikasi dan Evaluasi

Label / sentimen pada dataset dilakukan *encoding* menjadi data numerik agar mudah dalam proses klasifikasi. Kemudian model di *train* pada data tweet / komentar yang telah dibentuk vektor pada proses ekstraksi fitur menggunakan algoritma TF-IDF. Hasil evaluasi menggunakan data *testing* yang diklasifikasi menggunakan model yang telah di *training* untuk menghasilkan akurasi prediksi.

Pada proses klasifikasi, peneliti membagi dataset dengan proporsi 80:20. Berikut merupakan sajian hasil klasifikasi dari masing-masing pengujian dapat dilihat pada gambar 2.

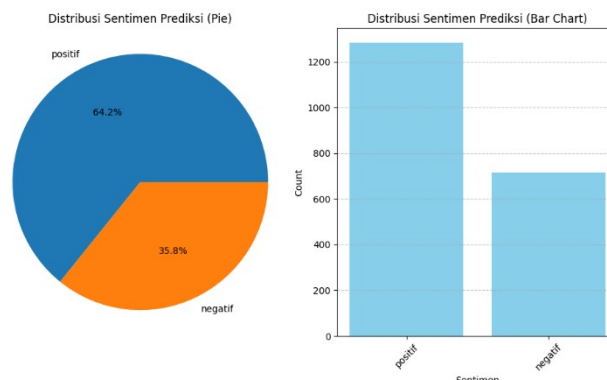
Classification Report Data Testing

precision	recall	f1-score	support	
	0.89	0.59	0.71	180
negatif				
	0.74	0.94	0.83	220
positif				
accuracy			0.79	400
macro avg	0.82	0.77	0.77	400
weighted avg	0.81	0.79	0.78	400

Gambar 2. Hasil Klasifikasi

3.6. Visualisasi

Setelah proses klasifikasi menggunakan algoritma *Naive Bayes*, kemudian dilakukan proses visualisasi dengan tujuan untuk melihat berapa banyak sentimen positif ataupun sentimen negatif dari hasil proses klasifikasi. Sajian visualisasi terdapat pada gambar 3.



Gambar 3. Visualisasi Sentimen Prediksi

IV. KESIMPULAN

Dari hasil penelitian analisis sentimen tentang kemenangan Tim Nasional Indonesia pada Piala Asia U-23 AFC 2024 ini, dapat disimpulkan bahwa:

1. Penggunaan Inset Lexicon sudah cukup baik dalam proses pelabelan karena Inset berbasis bahasa Indonesia.
2. Algoritma *Naive Bayes* mempunyai akurasi yang baik dalam proses klasifikasi.
3. Hasil visualisasi menunjukkan bahwa sentimen mayoritas masyarakat Indonesia terhadap kemenangan Tim Nasional Indonesia sangat baik, 64,2% sentimen positif dan 35,8% sentimen negatif.

DAFTAR PUSTAKA

- [1] S. Fernando, A. Voutama, dan A. Andri Hendriadi, "KLASIFIKASI BERITA HOAKS KAMPANYE PEMILIHAN UMUM (PEMILU) 2024 MENGGUNAKAN ALGORITMA NAÏVE BAYES," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, hlm. 2112–2115, Apr 2024, doi: 10.36040/jati.v8i2.9400.
- [2] S. H. Putri dan L. O. Maharani, "The role of @Txdaripemerintah Twitter account as a public space for the community in the perspective of communication on social media," *ArtComm : Jurnal Komunikasi dan Desain*, vol. 4, no. 01, hlm. 37–47, Apr 2021, doi: 10.37278/artcomm.v4i01.401.
- [3] N. P. G. Naraswati, D. Cindy Rosmilda, D. Desinta, F. Khairi, R. Damaiyanti, dan R. Nooraeni, "SISTEMASI: Jurnal Sistem Informasi Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification," 2021. [Daring]. Tersedia pada: <http://sistemasi.ftik.unisi.ac.id>
- [4] N. S. Marga, "SENTIMEN ANALISIS TENTANG KEBIJAKAN PEMERINTAH TERHADAP KASUS CORONA MENGGUNAKAN METODE NAIVE BAYES," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 2, no. 4, hlm. 453–463, Feb 2022, doi: 10.33365/jatika.v2i4.1602.
- [5] S. Tetra Oktaviani, H. Baturohmah, P. Studi Sistem Informasi, U. Nusa Putra Jl Raya Cibolang No, C. Kaler, dan J. Barat, "SENTIMEN ANALISIS PENGGUNA TWITTER INDONESIA TERHADAP PIALA DUNIA 2022 MENGGUNAKAN METODE NAÏVE BAYES CLASIFIER," 2023.

- [6] P. Samotana Zalukhu, T. Handhayani, dan M. Sitorus, “ANALISIS SENTIMEN TERHADAP KENAIKAN BBM DI INDONESIA PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE NAÏVE BAYES,” *Jurnal Sistem Informasi dan Ilmu Komputer (SIMTEK)*, vol. 8, no. 1, 2023.
- [7] D. F. Zhafira, B. Rahayudi, dan I. Indriati, “Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes dan Pembobotan TF-IDF Berdasarkan Komentar pada Youtube,” *Jurnal Sistem Informasi, Teknologi Informasi, dan Edukasi Sistem Informasi*, vol. 2, no. 1, Agu 2021, doi: 10.25126/justsi.v2i1.24.
- [8] N. A. Widiastuti, A. K. Zyen, dan N. Safik, “PREDIKSI PENENTUAN PEMOHON KREDIT SEPEDA MOTOR MENGGUNAKAN ALGORITMA NAIVE BAYES,” *JURNAL DISPROTEK*, vol. 10, no. 2, hlm. 130–134, Feb 2023, doi: 10.34001/jdpt.v10i2.4689.
- [9] S. P. M. M. Dr. Muhammad Ramdhan, *Metode Penelitian*. Cipta Media Nusantara. [Daring]. Tersedia pada: https://books.google.co.id/books?id=Ntw_EAAAQBAJ
- [10] D. S. Wijaya, A. Sanjaya, dan W. C. Utomo, “Analisis Sentimen Masyarakat Tentang Naiknya Harga BBM Dengan Metode Fasttext dan Naïve Bayes,” *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, vol. 7, no. 1, hlm. 27–35, Jul 2023.
- [11] B. Hakim, “Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning,” *JBASE - Journal of Business and Audit Information Systems*, vol. 4, no. 2, Agu 2021, doi: 10.30813/jbase.v4i2.3000.
- [12] F. Koto dan G. Y. Rahmaningtyas, “Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs,” dalam *2017 International Conference on Asian Language Processing (IALP)*, IEEE, Des 2017, hlm. 391–394. doi: 10.1109/IALP.2017.8300625.
- [13] Regina, T. H. Saragih, dan D. Kartini, “ANALISIS SENTIMEN BRAND AMBASSADOR BTS TERHADAP TOKOPEDIA MENGGUNAKAN KLASIFIKASI BAYESIAN NETWORK DENGAN EKSTRAKSI FITUR TF-IDF,” *Jurnal Informatika Polinema*, vol. 9, no. 4, hlm. 383–390, Agu 2023, doi: 10.33795/jip.v9i4.1333.
- [14] B. Haya Pangestu, “Data Mining Menggunakan Algoritma Naïve Bayes Classifier Untuk Evaluasi Kinerja Karyawan,” *Jurnal Riset Matematika*, hlm. 177–184, Des 2023, doi: 10.29313/jrm.v3i2.2837.
- [15] R. Nurhidayat dan K. E. Dewi, “KOMPUTA : Jurnal Ilmiah Komputer dan Informatika PENERAPAN ALGORITMA K-NEAREST NEIGHBOR DAN FITUR EKSTRAKSI N-GRAM DALAM ANALISIS SENTIMEN BERBASIS ASPEK,” vol. 12, no. 1, 2023, [Daring]. Tersedia pada: <https://www.kaggle.com/datasets/hafidahmusthaanah/skincare-review?select=00.+Review.csv>.
- [16] H. Satria, “Cara Mendapatkan Data (Crawl) Twitter X - Maret 2024.” Diakses: 9 Juni 2024. [Daring]. Tersedia pada: <https://helimisatria.com/blog/updated-crawl-data-twitter-x-maret-2024>