

ANALISIS SENTIMEN TIM NASIONAL SEPAK BOLA INDONESIA DI TURNAMEN PIALA DUNIA U-17 INDONESIA PADA TWITTER (X) MENGGUNAKAN ALGORITMA NAÏVE BAYES

Trianda Juniardi¹, Castaka Agus Sugianto²

^{1,2}Politeknik TEDC Bandung; Jl. Politeknik Jl. Pesantren No.2, Cibabat, Kec. Cimahi Utara, Kota Cimahi, Jawa Barat 40513; telp (022) 6645951

Received: 12 Agustus 2024
Accepted: 5 Oktober 2024
Published: 12 Oktober 2024

Keywords:

Analisis Sentimen;
Twitter;
Piala Dunia U-17;
Klasifikasi;
Algoritma Naïve Bayes.

Corespondent

Email:
triandajuniardi05@gmail.co
m

Abstrak. Piala Dunia U-17 adalah salah satu ajang bergengsi dalam sepak bola internasional yang menarik perhatian luas, termasuk di platform media sosial seperti Twitter(X). Penelitian ini bertujuan untuk menganalisis sentimen masyarakat Indonesia terhadap tim nasional sepak bola U-17 Indonesia selama ajang Piala Dunia U-17 menggunakan algoritma *Naïve Bayes*. Data dikumpulkan dari komentar-komentar Twitter(X) yang diklasifikasikan menjadi sentimen positif, negatif, dan netral. Metode ini melibatkan proses *preprocessing* untuk membersihkan dan mengelola data teks sebelum analisis. Hasil analisis menunjukkan mayoritas masyarakat mengekspresikan sentimen positif terhadap timnas U-17 Indonesia, dengan dukungan yang kuat terlihat dari kata-kata seperti "timnas," "Indonesia," dan "kebanggaan." Namun, terdapat juga kritik yang signifikan terhadap performa timnas, mencerminkan variasi opini dalam masyarakat. Metode *Naïve Bayes* berhasil mengklasifikasikan sentimen dengan akurasi 78.34%, presisi 82.96%, *recall* 78.34%, *f1-score* 74.95%. Penelitian ini memberikan wawasan penting tentang persepsi dan respons masyarakat terhadap event olahraga besar di Indonesia, serta relevansi dan kegunaan algoritma *Naïve Bayes* dalam menganalisis data sentimen media sosial secara efektif.

Abstract. The U-17 World Cup is one of the prestigious events in international soccer that attracts widespread attention, including on social media platforms such as Twitter(X). This study aims to analyze the sentiment of Indonesian people towards the Indonesian U-17 national football team during the U-17 World Cup event using the Naïve Bayes algorithm. Data was collected from Twitter(X) comments that were classified into positive, negative, and neutral sentiments. This method involves a preprocessing process to clean and organize the text data before analysis. The results of the analysis showed that the majority of people expressed positive sentiments towards the Indonesian U-17 national team, with strong support evident from words such as "national team," "Indonesia," and "pride." However, there was also significant criticism of the national team's performance, reflecting the variety of opinions within the community. The Naïve Bayes method successfully classified the sentiment with 78.34% accuracy, 82.96% precision, 78.34% recall, 74.95% f1-score. This research provides important insights into public perceptions and responses to major sporting events in Indonesia, as well as the relevance and usefulness of the Naïve Bayes algorithm in effectively analyzing social media sentiment data.

1. PENDAHULUAN

Piala Dunia U-17 adalah salah satu turnamen sepak bola internasional yang

mempertemukan tim-tim nasional sepak bola di bawah usia 17 tahun. Penyelenggaraan acara ini tidak hanya menciptakan antusiasme

di kalangan penggemar sepak bola, tetapi juga menjadi perhatian publik secara luas [1]. Seiring dengan perkembangan media sosial, *Twitter* telah menjadi *platform* yang sangat signifikan untuk berbagi opini, komentar, dan reaksi terkait berbagai peristiwa, seperti Piala Dunia U-17 [2].

Analisis sentimen opini publik di *Twitter* dapat memberikan wawasan yang berharga terkait persepsi dan tanggapan masyarakat terhadap suatu acara. Dalam konteks Piala Dunia U-17 di Indonesia, penting untuk memahami bagaimana *netizen* atau pengguna *Twitter* menilai terhadap tim nasional sepak bola (timnas) indonesia di turnamen ini . Hal ini dapat mencakup respons terhadap fasilitas, organisasi, kinerja tim, keamanan, dan performa tim nasional (timnas) indonesia yang memengaruhi pengalaman penonton dan citra penyelenggara. Agar kedepannya bisa menjadi evaluasi [3].

Selain itu, penggunaan media sosial sebagai sumber data memungkinkan analisis *real-time* yang dapat memberikan informasi cepat dan akurat terkait dengan perubahan sentimen selama penyelenggaraan Piala Dunia U-17. Analisis ini tidak hanya bermanfaat bagi tim nasional sepak bola (timnas) indonesia, tetapi juga dapat memberikan wawasan bagi pihak-pihak terkait untuk meningkatkan kualitas tim nasional sepak bola (timnas) Indonesia di masa depan dan memahami bagaimana peristiwa olahraga besar dapat memengaruhi opini publik secara keseluruhan [4].

Pada piala dunia U-17 yang di gelar 10 November performa tim nasional sepak bola Indonesia saat ini menjadi salah satu sorotan masyarakat Indonesia, menurunnya performa timnas Indonesia U-17 dalam ajang piala dunia U-17 membuat opini *tweet* negatif masyarakat sering bermunculan, saran dan argumentasi pun tidak luput mereka berikan untuk tim nasional sepak bola indonesia pentingnya analisis sentimen opini masyarakat Indonesia tentang performa tim nasional sepak bola Indonesia saat ini [5]. Berdasarkan penelitian terdahulu tedapat penelitian terkait dalam melakukan klasifikasi *tweet* untuk menentukan apakah sentimen dari *tweet* tersebut netral, negatif, atau positif menggunakan algoritma *Naïve Bayes*. Hasil analisis menunjukkan sentimen

dengan tingkat akurasi 57,22%. [6]. Ada juga penelitian lain menggunakan metode *Naïve Bayes* untuk menganalisis sentimen masyarakat terhadap SEA Games Kamboja 2023 di *Twitter*. Hasil analisis sentimen menunjukkan tingkat akurasi sebesar 92,70%[4]. Penelitian selanjutnya menggunakan metode *Support Vector Machine* (SVM) untuk menganalisis sentimen publik terhadap Tragedi Kanjuruhan Malang. Hasil penelitian menunjukkan akurasi 95,16%. [7]. Penelitian sebelumnya telah menunjukkan bahwa Algoritma *Naïve Bayes* efektif dalam berbagai situasi. Sebagai contoh, dalam sebuah studi tentang prediksi autisme, Algoritma *Naïve Bayes* mencapai akurasi sebesar 96,45%, jauh lebih tinggi dibandingkan dengan *Support Vector Machine* yang hanya mencapai 81,56% [8].

Jika dilihat dari penelitian terdahulu bahwa algoritma *Naïve Bayes* banyak digunakan dalam sentimen alalysis salah satunya. Algoritma *Naïve Bayes* dipilih sebagai metode analisis sentimen karena kemampuannya yang telah terbukti dalam klasifikasi teks. Metode ini memanfaatkan *teorema Bayes* untuk menghitung probabilitas suatu dokumen atau *tweet* tertentu termasuk dalam kategori sentimen positif, negatif, atau netral [9]. Dengan menggunakan Algoritma *Naïve Bayes* memiliki beberapa keunggulan utama, termasuk kecepatan komputasi dan keakuratan yang tinggi, kesederhanaan implementasi, dan tingkat akurasi yang memadai. Kecepatan, keakuratan kesederhanaan, dan akurasi yang baik menjadikannya populer dalam pembelajaran mesin dan analisis data [10]. *Naïve Bayes* sudah terbukti memiliki nilai akurasi yang tinggi dan kecepatan yang cepat jika diimplementasikan ke dalam dataset yang besar [11].

Dengan alasan-alasan di atas penelitian ini, akan dilakukan untuk menganalisis opini-opini masyarakat di *twitter* mengenai sentimen terhadap tim nasional sepak bola Indonesia di ajang piala dunia U-17 dengan menggunakan *python* dengan metode *naive bayes classifier*.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen merupakan kajian komputasional terhadap opini, sentimen, serta emosi yang diungkapkan dalam teks. Tugas utama dalam analisis sentimen adalah mengklasifikasikan polaritas dari teks yang terdapat dalam dokumen, kalimat, atau pernyataan[12]. Analisis sentimen bertujuan untuk mengelompokkan sentimen tersebut menjadi positif, negatif, atau netral.

2.2. Data Mining

Data mining diartikan sebagai serangkaian proses yang bertujuan untuk mengeksplorasi dan menemukan nilai berupa informasi serta relasi-relasi kompleks yang selama ini tersembunyi dalam suatu basis data[13]. Teknik data mining adalah proses mencari data yang ada untuk membangun sebuah model, kemudian menggunakan model tersebut untuk mengidentifikasi pola pada data lain yang belum ada dalam basis data yang tersimpan[14].

2.3. Algoritma Naïve Bayes

Naïve Bayes adalah metode klasifikasi berbasis probabilitas dan statistik yang diperkenalkan oleh ilmuwan Inggris, Thomas Bayes[15]. Naïve Bayes adalah algoritma klasifikasi dalam data mining yang menganggap bahwa setiap atribut bersifat independen dan tidak saling mempengaruhi satu sama lain[16].

3. METODE PENELITIAN

3.1. Menentukan Masalah

Publik *Twitter* memberikan tanggapan terkait kritik terhadap performa tim nasional sepak bola U-17 Indonesia yang dinilai kurang baik. Kritik ini datang dari berbagai pemain dan staf kepelatihan. Penulis menemukan kritik ini melalui komentar di Twitter terkait performa timnas U-17 yang berlaga di Indonesia. penelitian mengenai analisis sentimen opini publik di Twitter terhadap timnas U-17 Indonesia menggunakan algoritma *Naïve Bayes*.

3.2. Menentukan Tujuan dan Ruang Lingkup

Publik *Twitter* memberikan tanggapan terkait kritik terhadap performa tim nasional sepak bola U-17 Indonesia yang dinilai kurang baik. Kritik ini datang dari berbagai pemain dan staf kepelatihan. Penulis menemukan kritik ini melalui komentar di Twitter terkait performa timnas U-17 yang berlaga di Indonesia. Langkah-langkah untuk memecahkan masalah ini dijelaskan dalam latar belakang penelitian mengenai analisis sentimen opini publik di Twitter terhadap timnas U-17 Indonesia menggunakan algoritma *Naïve Bayes*.

3.3. Mencari Litelatur

Mencari litelature atau referensi untuk tujuan mendukung kerangka pemahaman yang berkaitan tentang data mining, analisis sentimen, algoritma *naïve baye*. Data yang dikumpulkan penulis melalui metode penelitian adalah.

1. Konsep *Data Mining*
2. Analisis Sentimen
3. Algoritma *Naïve Bayes*
4. Hasil Penelitian Terkait

3.4. Metode Pengumpulan Data

Berikut adalah metode pengumpulan data yang dilakukan penulis pada penelitian ini untuk mendapatkan data yang dibutuhkan.

3.4.1. Data Primer

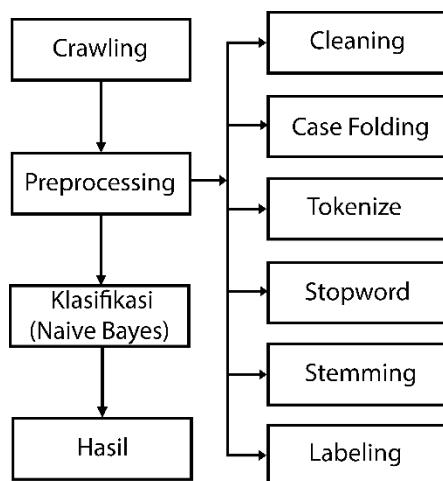
Data primer adalah data yang diperoleh langsung dari sumber terkait. Dalam penelitian ini, data primer berupa data dari komentar *Twitter* piala dunia u-17 yang dikumpulkan menggunakan *jupyter notebook* dalam format *Comma Separated Values* (CSV) selama 4 bulan, dari November 2023 hingga Februari 2024, dengan total sekitar 508 data.

3.4.2. Data Sekunder

Data sekunder adalah data yang diperoleh secara tidak langsung dari objeknya. Data ini dikumpulkan dari buku, jurnal, dan media internet yang relevan. Dalam penelitian ini, penulis mengumpulkan data sekunder melalui studi pustaka, yang mencakup pengetahuan tentang data mining dan teori-teori yang dibutuhkan.

3.5. Metode Yang Digunakan

Metode yang digunakan dalam pengelolaan data *mining* terdiri dari 4 tahapan utama yaitu *crawling*, *preprocessing*, ekstrasi *fitur* dan klasifikasi seperti pada Gambar 1.



Gambar 1. Metode yang digunakan

Gambar 1 menggambarkan 4 tahapan utama penelitian yang dilakukan oleh penulis. Pertama, metode *Crawling* digunakan untuk mengumpulkan data *tweet* dari *Twitter* yang berkaitan dengan Piala Dunia U-17. Selanjutnya, tahap *Preprocessing* meliputi berbagai proses yang bertujuan untuk mempersiapkan data ini untuk analisis lebih lanjut; proses-proses ini termasuk *Cleaning*, *Case Folding*, *Tokenize*, penghapusan *Stopword*, *Stemming*, dan *Labeling*. Tahap ketiga adalah *Klasifikasi* menggunakan algoritma *Naïve Bayes* untuk mengetahui hasil akurasi, presisi, *recall*, dan *f1-score* dari hasil sentimen positif, negatif, dan netral. Hasil dari klasifikasi tersebut dinilai untuk menentukan efektivitas model dalam menganalisis sentimen terhadap Piala Dunia U-17.

3.6. Proses dan Hasil Penelitian

Proses dalam penelitian ini adalah mencari nilai positif, negatif, dan netral dalam komentar *twitter* terkait piala dunia u-17 dengan menggunakan algoritma *naïve bayes*. Dalam penelitian ini, penulis melakukan analisis untuk mengidentifikasi sentimen positif, negatif, dan netral dari komentar *Twitter* terkait Piala Dunia U-17 menggunakan algoritma *Naïve Bayes*. Selain itu, penulis juga akan mengevaluasi akurasi, presisi, *recall*, dan *f1-score* dari hasil analisis tersebut.

3.7. Analisis

Hasil analisis mencakup identifikasi nilai positif, negatif, dan netral dari komentar-komentar Twitter terkait Piala Dunia U-17

dengan menggunakan algoritma *Naïve Bayes*. Selain itu, evaluasi dilakukan dengan mengukur akurasi, *presisi*, *recall*, dan *f1-score* untuk mengevaluasi kinerja dari model analisis sentimen yang dikembangkan.

3.8. Penarikan Kesimpulan

Penarikan kesimpulan dari analisis sentimen komentar *Twitter* terkait Piala Dunia U-17 menggunakan algoritma *Naïve Bayes* dengan bantuan *jupyter notebook*. Melalui evaluasi yang mencakup akurasi, *presisi*, *recall*, dan *f1-score*, penelitian ini akan mengidentifikasi dan membedakan sentimen positif, negatif, dan netral dari komentar-komentar tersebut. Hasil penelitian memberikan pemahaman yang lebih dalam tentang persepsi dan reaksi publik terhadap acara piala dunia.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Pada tahap pengumpulan data, penelitian ini memanfaatkan ulasan pengguna *Twitter/X* terkait Piala Dunia U-17. Metode yang digunakan untuk mengakses data tersebut adalah dengan memanfaatkan *clawling* data menggunakan *tools harvest* pada *jupyter notebook*. Total data yang berhasil dikumpulkan mencapai 508 ulasan *Twitter* terkait Piala Dunia U-17 tediri dari atribut *created_at*, *id_str*, *full_text*, *quote_count*, *reply_count*, *retweet_count*, *favorite_count*, *lang*, *user_id_str*, *conversation_id_str*, *username*, *tweet_url*. Dari 12 atribut yang penulis gunakan hanya atribut *full_text* yang isinya komentar masyarakat. Data yang diperoleh kemudian disimpan dalam format *csv* untuk memudahkan analisis sentimen pada tahap selanjutnya. Dapat dilihat pada gambar 2.

created_at	id_str	full_text	quote_count	reply_count	retweet_count	favorite_count	lang	user_id_str	conversation_id_str	username
Thu Dec 01 03:14 2022	1.740190e+18	Ala Uta Sis pungkiran serunya main di Liga TopSkor nih Skorerð~%o #skorlive #timnasu17 #ligatopskor https://t.co/LbA2a8Rjx3	0.0	0.0	0.0	1.0	in	8.382203e+18	1.740190e+18	alakonindonesia
Sun Dec 04 08:34 2022	1.738270e+18	Di balik karir yang sedang diperjuangkan pasti di belakangnya didukung oleh orang-orang tersayang seperti keluarga ya Skorerð~Š #skorlive #JiDaBin #LigaTopSkor #TimnasU17	0.0	0.0	0.0	1.0	in	8.382203e+18	1.738270e+18	alakonindonesia
Sat Dec 10 08:34 2022	1.733540e+18	Profil dan Ankatan Karir atau selaku bahegga kita kamu Yuk atau selaku bahegga kamu Yuk atau	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Sun Dec 11 22:20 2022	1.731440e+18	atau selaku bahegga kamu Yuk atau	0.0	0.0	1.0	2.0	in	1.194940e+18	1.731440e+18	Balitmasia88
Sat Dec 12 08:34 2022	1.731080e+18	atau selaku bahegga kamu Yuk atau	0.0	0.0	0.0	2.0	in	1.194940e+18	1.731080e+18	Balitmasia88
...
Fri Nov 10 02:12:36 2022	1.722820e+18	JADUL, SOLAR,HARD The Link Live Streaming Adm	0.0	0.0	0.0	1.0	in	1.857280e+18	1.722820e+18	ZeeMoonTV1
SUN 14 Nov 10 02:12:36 2022	1.722810e+18	Patra Dunia U-17 Indonesia Sampai	0.0	0.0	0.0	0.0	in	1.855130e+18	1.722810e+18	Hukupelipata
Fri Nov 10 02:12:36 2022	1.722810e+18	Patra Dunia U-17 Indonesia Sampai	0.0	0.0	0.0	0.0	in	1.855130e+18	1.722810e+18	Hukupelipata
Fri Nov 10 02:12:36 2022	1.722800e+18	#PES 2023 #LigaTopSkor #TimnasU17	1.0	0.0	0.0	0.0	in	1.843740e+18	1.722800e+18	LigaKembart77
Fri Nov 10 02:12:44 2022	1.722790e+18	Jadwal Patra Dunia U-17 Indonesia vs	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
SUN 17 Nov 10 02:12:44 2022	1.722780e+18	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Gambar 2. Data menta Piala Dunia hasil *scraping*

4.2. Preprocessing Data

Pada tahap ini, dilakukan serangkaian langkah untuk membersihkan data *cleaning*, termasuk menghilangkan tanda baca seperti *hashtag*, karakter, simbol dan mention. Selain itu, dilakukan pengubahan seluruh teks menjadi huruf kecil (*case folding*), menghapus kata-kata umum (*stopword*), dan mengubah kata-kata dengan imbuhan menjadi bentuk kata dasar (*stemming*) dan (*tokenize*) untuk memisahkan kata-kata agar dapat memudahkan pemrosesan.

4.2.1. Cleaning

Fungsi *cleaning* ini digunakan untuk membersihkan teks dari beberapa elemen yang umumnya tidak diinginkan dalam analisis teks. Pada penelitian ini atribut *created_at*, *id_str*, *quote_count*, *reply_count*, *retweet_count*, *favorite_count*, *lang*, *user_id_str*, *conversation_id_str*, *username*, *tweet_url* tidak dilakukan proses *cleaning*, hanya colom *full_text* saja dilakukan proses *cleaning*. Atribut *full_text* ini berisikan hasil komentar masyarakat tentang piala dunia. Berikut tabel 2 hasil pembersihan data yang telah diambil dari colom *full_text* dan sudah di *cleaning*. Seperti pada tabel 1.

Tabel 1. Cleaning

Sebelum	Sesudah
#Ji Da Bin berbagi pengalaman gimana serunya main di Liga TopSkor nih Skorerð~%o #skorlive #timnasu17 #ligatopskor https://t.co/LbA2a8Rjx3	Ji Da Bin berbagi pengalaman gimana serunya main di Liga TopSkor nih Skorer Skorlive timnasu ligatopskor

timnasu ligatopskor
Di balik karir yang sedang diperjuangkan pasti di belakangnya didukung oleh orang-orang tersayang seperti keluarga ya Skorerð~Š #skorlive #JiDaBin #LigaTopSkor #TimnasU17 https://t.co/7BC7KBBNvQ

Pada tabel 1. Proses ini melakukan penghapusan semua elemen yang tidak relevan untuk analisis teks dihapus seperti atribut yang ada pada data, atribut selain colom *full_text* dihilangkan, termasuk karakter ASCII yang tidak dikenali seperti emoji, angka-angka, mention (tag pengguna), *link* (URL), dan *hashtag*. Selain itu, *whitespace* yang berlebihan juga dihilangkan untuk memastikan teks yang dihasilkan lebih konsisten dan mudah diproses. Hasilnya adalah teks yang lebih bersih dan terfokus pada inti pesan, siap untuk digunakan dalam analisis atau pengolahan lebih lanjut, seperti dalam studi kasus analisis sentiment.

4.2.2. Case Folding

Case folding adalah proses mengubah semua huruf dalam teks menjadi huruf kecil. Seperti pada tabel 2.

Tabel 2. Case Folding

Sebelum	Sesudah
Ji Da Bin berbagi pengalaman gimana serunya main di Liga TopSkor nih Skorerð~%o #skorlive #timnasu17 #ligatopskor	ji da bin berbagi pengalaman gimana serunya main di liga topskor nih skorer skorlive timnasu ligatopskor
Di balik karir yang sedang diperjuangkan pasti di belakangnya didukung oleh orang	di balik karir yang sedang diperjuangkan pasti di belakangnya didukung oleh orang tersayang seperti

orang tersayang seperti keluarga ya Skorer skorlive JiDaBin LigaTopSkor Timnasu	keluarga ya skorer skorlive jidabin ligatopskor timnasu
--	---

Pada tabel 2, huruf kapital diubah menjadi huruf kecil untuk mempermudah identifikasi kata-kata berikutnya.

4.2.3. Tokenize

Tokenisasi sangat penting untuk mengenali dan mengolah setiap kata secara terpisah. Berikut tabel 3 hasil *Tokenize*.

Tabel 3. *Tokenize*

Sebelum	Sesudah
ji da bin berbagi pengalaman gimana serunya main di liga topskor nih skorer skorlive timnasu ligatopskor	['ji', 'da', 'bin', 'berbagi', 'pengalaman', 'gimana', 'serunya', 'main', 'di', 'liga', 'topskor', 'nih', 'skorer', 'skorlive', 'timnasu', 'ligatopskor']
di balik karir yang sedang diperjuangkan pasti di belakangnya didukung oleh orang orang tersayang seperti keluarga ya skorer skorlive jidabin ligatopskor timnasu	['di', 'balik', 'karir', 'yang', 'sedang', 'diperjuangkan', 'pasti', 'di', 'belakangnya', 'didukung', 'oleh', 'orang', 'orang', 'tersayang', 'seperti', 'keluarga', 'ya', 'skorer', 'skorlive', 'jidabin', 'ligatopskor', 'timnasu']

Pada tabel 3, setelah melakukan *case folding*, langkah berikutnya adalah tokenisasi. Proses ini membagi teks menjadi unit-unit seperti kata. Tujuannya adalah untuk menghasilkan token-token yang dapat dianalisis secara individu.

4.2.4. StopWord

Stopword adalah kata-kata umum dalam bahasa yang sering muncul namun biasanya tidak memberikan makna penting dalam analisis teks. Seperti pada tabel 4.

Tabel 4. *StopWord*

Sebelum	Sesudah
['ji', 'da', 'bin', 'berbagi', 'pengalaman', 'gimana', 'serunya', 'main', 'di', 'liga', 'topskor', 'nih', 'skorer', 'skorlive', 'timnasu', 'ligatopskor']	['ji', 'da', 'bin', 'berbagi', 'pengalaman', 'gimana', 'serunya', 'main', 'liga', 'topskor', 'nih', 'skorer', 'skorlive', 'timnasu', 'ligatopskor']
['di', 'balik', 'karir', 'yang', 'sedang', 'diperjuangkan', 'pasti', 'di', 'belakangnya', 'didukung', 'oleh', 'orang', 'orang', 'tersayang', 'seperti', 'keluarga', 'ya', 'skorer', 'skorlive', 'jidabin', 'ligatopskor', 'timnasu']	['karir', 'diperjuangkan', 'belakangnya', 'didukung', 'orang', 'orang', 'tersayang', 'keluarga', 'ya', 'skorer', 'skorlive', 'jidabin', 'ligatopskor', 'timnasu']

Pada tabel 4, kata-kata yang tidak relevan atau tidak mengandung opini telah dihilangkan.

4.2.5. Stemming

Stemming adalah proses menghapus imbuhan dari kata atau ulasan untuk mendapatkan bentuk dasar dari sebuah kata. Berikut tabel 5 hasil *stemming*.

Tabel 5. *Stemming*

Sebelum	Sesudah
['ji', 'da', 'bin', 'berbagi', 'pengalaman', 'gimana', 'serunya', 'main', 'liga', 'topskor', 'nih', 'skorer', 'skorlive', 'timnasu', 'ligatopskor']	['ji', 'da', 'bin', 'bagi', 'alam', 'gimana', 'seru', 'main', 'liga', 'topskor', 'nih', 'skorer', 'skorlive', 'timnasu', 'ligatopskor']
['karir', 'diperjuangkan', 'belakang', 'didukung', 'orang', 'orang', 'tersayang', 'sayang', 'keluarga', 'ya', 'skorer', 'skorlive',	['karir', 'juang', 'belakang', 'dukung', 'orang', 'orang', 'sayang', 'keluarga', 'ya', 'skorer', 'skorlive',

'skorer', 'skorlive', 'jidabin', 'ligatopskor', 'timnasu']	'jidabin', 'ligatopskor', 'timnasu']
---	---

Pada tabel 5, (imbuhan) dihapus dari kata-kata sehingga kata-kata dengan bentuk dasar yang sama dapat dikenali sebagai bentuk yang sama.

4.2.6. Pelabelan

Setelah tahap stemming dilakukan pelabelan data bertujuan untuk menunjukkan apakah suatu teks dianggap sentimen positif, negatif, atau netral. Untuk menentukan sentimen positif, negatif, atau netral ini, dilakukan dengan menggunakan "kampus lexicon", yang terdiri dari dua file: positiv.csv dan negatif.csv. Berikut tabel 6 data kamus lexicon positif.csv dan tabel 7 kamus lexicon neagtif.csv.

4.2.6.1 Kamus Lexicon Positif

Tabel 6. Kamus Lexicon Positif

No	Word	Weight
1.	Hai	3
2.	merekam	2
3.	Ekstensif	3
4.	paripurna	1
.	.	.
.	.	.
.	.	.
3610.	orisinal	3

Kata-kata yang termasuk dalam kategori sentimen positif diidentifikasi menggunakan *jupyter notebook* yang memuat file positiv.csv. File ini berisi daftar kata-kata yang memiliki konotasi positif dengan jumlah data 3610. Saat menganalisis komentar, setiap kata dalam komentar dibandingkan dengan daftar kata dalam positiv.csv. Jika kata tersebut ditemukan dalam daftar ini, maka kata tersebut diklasifikasikan sebagai kata dengan sentimen positif.

4.2.6.2 Kamus Lexicon Negatif

Tabel 7. Kamus Lexicon Negatif

No	Word	Weight
1.	putus tali gantung	-2
2.	gelebah	-2
3.	gobar hati	-2

4.	tersentuh (perasaan)	-1
.	.	.
.	.	.
.	.	.
6610	kacang botor	-3

Selanjutnya Sentimen negatif diidentifikasi menggunakan *jupyter notebook* yang memuat file negatif.csv. yang berisi daftar kata-kata dengan konotasi negatif dengan jumlah data 6610. Setiap kata dalam komentar yang dianalisis dibandingkan dengan daftar kata dalam negatif.csv. Jika ditemukan dalam daftar ini, maka kata tersebut diklasifikasikan sebagai kata dengan sentimen negatif.

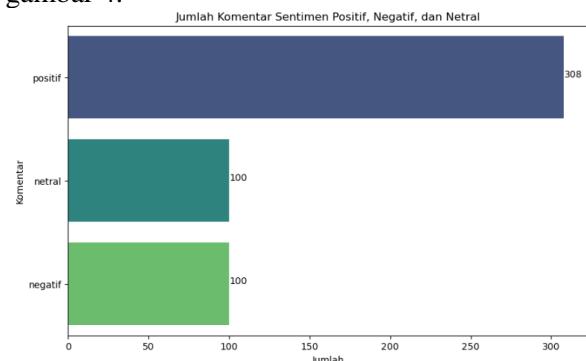
Sentimen netral Kata-kata yang tidak termasuk dalam positiv.csv maupun negatif.csv dianggap sebagai kata-kata dengan sentimen netral. Sentimen netral ditemukan dari kata-kata yang tidak ada dalam kedua file *lexicon* ini dikategorikan sebagai netral. Berikut gambar 3 hasil dari pelabelan.

stemming_string	label
0 ji da bin bagi alam gimana seru main liga tops...	positif
1 karir juang belakang dukung orang orang sayang...	positif
2 profil biodata arkhan kaka	netral
3 bahagia link daftar indonesia jakarta indonesia...	positif
4 langkah sukses link daftar indonesia jakarta i...	positif

Gambar 3. Hasil Pelabelan

Setelah data dibuat label, Kemudian data dibuatkan grafik hasil labeling seperti, positif, negatif, dan netral.

Berikut adalah grafik hasil dari proses pelabelan data yang menetapkan label sentimen positif, negatif, dan netral. Dapat dilihat pada gambar 4.



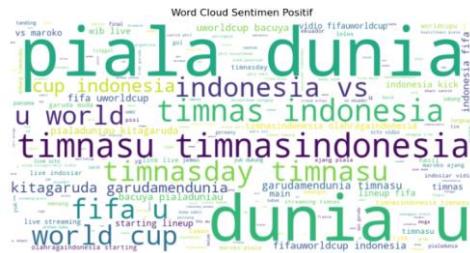
Gambar 3. Hasil Grafik Pelabelan Data

Dari gambar 4, diagram diatas menunjukkan hasil sentimen masyarakat tentang piala dunia u-17, dari data berjumlah 508, terlihat bahwa 308 orang-orang dari teks menunjukkan

sentimen positif, mengindikasikan mayoritas pengguna menyampaikan pandangan positif terkait Piala Dunia U-17 di Twitter. Sebanyak 100 teks memiliki sentimen negatif, ini mengidentifikasi pengguna yang menyampaikan pandangan negatif tentang Piala Dunia U-17. Untuk sentimen netral sama dengan sentimen negatif 100 teks, ini mengindikasikan sedikit.

Setelah proses pelabelan positif, negatif, dan netral selanjutnya dilakukan visualisasi *wordcloud* berfungsi untuk menampilkan kata-kata dari label positif, negatif, dan netral.

4.2.6.3 Word Cloud Sentimen Positif

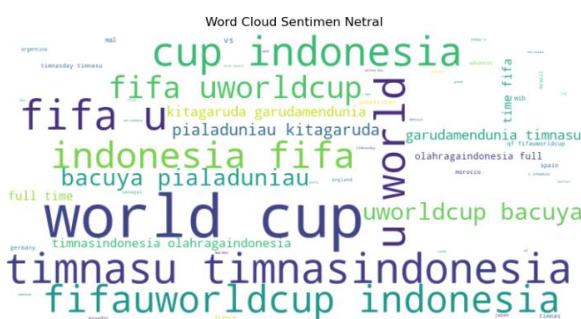


Gambar 5. Word Cloud Positif

Gambar 5 hasil visualisasi *wordcloud* sentimen positif.

4.2.6.4 Word Cloud Sentimen Netral

Gambar 6 hasil visualisasi *wordcloud* Analisis sentimen netral.



Gambar 6. Word Cloud Netral

4.2.6.5 Word Cloud Sentimen Negatif

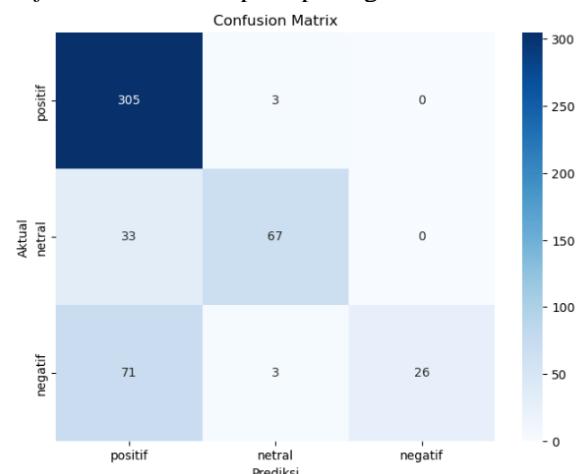


Gambar 7. Word Cloud Negatif

Gambar 7 hasil visualisasi *wordcloud* sentimen negatif.

4.3. Klasifikasi Naive Bayes

Dalam tahap klasifikasi menggunakan *Naive Bayes*, algoritma ini menghitung probabilitas kemunculan kata-kata spesifik dalam setiap kategori sentimen (positif, negatif, netral). Probabilitas ini kemudian digunakan untuk menentukan kategori sentimen dari ulasan pengguna. Dataset yang digunakan terdiri dari ulasan pengguna Twitter (x) dengan jumlah 508 ulasan terkait Piala Dunia U-17 di Indonesia, yang telah melalui proses prapemrosesan. Dari semua ulasan itu diujikan dan diimplementasikan dan menghasilkan *confusion matrix* seperti pada gambar 8.



Gambar 8. Confusion Matrik

Dengan menggunakan semua ulasan, pada tahap proses klasifikasi *naive bayes* menggunakan *jupyter notebook* dan bahasa pemrograman *python* untuk mengevaluasi model klasifikasi. Menghasilkan nilai akurasi, presisi, recal, *f1-score*. Dapat dilihat pada gambar 10.

Akurasi: 0.7834645669291339
Presisi: 0.8296515226516737
Recall: 0.7834645669291339
F1-Score: 0.7495329741580935

Gambar 19. Hasil Klasifikasi *Naïve Bayes* Pada Ulasan

4.4. Hasil

Berdasarkan hasil klasifikasi *Naïve Bayes* pada seluruh ulasan data Piala Dunia U-17, dapat disimpulkan bahwa model ini memiliki kinerja yang sangat baik dalam mengklasifikasikan sentimen pada semua ulasan.

Evaluasi model *Naïve Bayes* dalam klasifikasi sentimen pada ulasan Piala Dunia U-17 menunjukkan bahwa model ini mampu mengklasifikasikan sentimen dengan tingkat akurasi, presisi, *recall*, dan *F1-score* yang tinggi. Berikut tabel 8 adalah hasil lebih rinci dari evaluasi model.

Tabel 8. Hasil Evaluasi Model

Metrik	Nilai
Akurasi	78.34%
Presisi	82.96%
Recal	78.34%
F1-Score	74.95%

Dari hasil evaluasi model di atas, kita dapat melihat bahwa model memiliki akurasi sebesar 78.34%, Presisi sebesar 82.96%, Recall sebesar 78.34%, F1-Score 74.98%.

5. KESIMPULAN

Berdasarkan analisis sentimen terhadap komentar pengguna *Twitter(X)*, ditemukan bahwa mayoritas opini publik terkait tim nasional Indonesia U-17 bersifat positif, dari 508 ulasan terdapat sentimen positif sebanyak 308 ulasan, sentimen netral sebanyak 100 ulasan, dan sentimen negatif sebanyak 100 ulasan. Banyak pengguna *Twitter(X)* yang menunjukkan dukungan dan kebanggaan terhadap tim nasional meskipun terdapat beberapa kritik dan sentimen negatif. Sentimen netral sama dengan sentimen negatif, menunjukkan bahwa pengguna *Twitter(X)* memberikan komentar yang informatif atau tidak memiliki bias emosional tertentu terhadap tim nasional. Algoritma *Naïve Bayes* telah terbukti efektif dalam mengklasifikasikan sentimen dari komentar *Twitter(X)*. Hasil

analisis menunjukkan tingkat akurasi 78.34%, presisi 82.96%, recall 78.34%, dan F1-score sebesar 74.95%. dalam memprediksi sentimen positif, negatif, dan netral.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] Ajie Wicaksono and Maximianus Agus Prayudi, “Analisis Dampak Penyelenggaraan Fifa World Cup U-17 Pada Sektor Pariwisata Di Indonesia,” *EDUTOURISM J. Tour. Res.*, vol. 6, no. 01, pp. 90–101, 2024, doi: 10.53050/ejtr.v6i01.760.
- [2] A. M. Zuhdi, E. Utami, and S. Raharjo, “Analisis Sentiment Twitter Terhadap Capres Indonesia 2019 Dengan Metode K-NN,” vol. 5, pp. 1–7, 2019.
- [3] M. D. Alizah, A. Nugroho, U. Radiyah, and W. Gata, “Sentimen Analisis Terkait Lockdown pada Sosial Media Twitter,” *Indones. J. Softw. Eng.*, vol. 6, no. 2, pp. 223–229, 2020, doi: 10.31294/ijse.v6i2.8991.
- [4] Y. Akbar and A. N. Ihsan, “Analisis Sentimen Twitter Terhadap Opini Masyarakat Pada Sea Games Kamboja 2023 Menggunakan Algoritma Support Vector Machine,” *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 6, no. 2, pp. 814–821, 2023, doi: 10.31539/intecoms.v6i2.7670.
- [5] M. Kholilullah and U. Hayati, “Analisis Sentimen Pengguna Twitter (X) Tentang Piala Dunia Usia 17 Menggunakan Metode Naive,” vol. 8, no. 1, pp. 392–398, 2024.
- [6] F. B. Al Rouf, G. Gata, G. Triyono, and N. Nawindah, “Penerapan Algoritma Naïve Bayes Pada Twitter Untuk Analisis Sentimen Masyarakat Terhadap Ketua Umum Pssi Terpilih Periode 2023-2027 Bapak Erick Thohir,” *Pros. Semin. Nas. Mhs. Fak. Teknol. Inf.*, vol. 2, no. 2, pp. 1022–1031, 2023.
- [7] F. P. Herlambang and D. Avianto, “Analisis Sentimen Opini Pengguna Twitter Terhadap Tragedi Kanjuruhan Malang dengan Metode Support Vector Machine,” *J. Media Inform.* ..., vol. 7, pp. 1727–1739, 2023, doi: 10.30865/mib.v7i4.6332.
- [8] U. Fitriani and A. Wibowo, “Penerapan Algoritme Naïve Bayes Untuk Memprediksi Kelulusan Mahasiswa Universitas Budi Luhur Berbasis Web,” *3 rd Semin. Nas. Mhs. Fak. Teknol. Inf.*, vol. 2, no. 2, pp. 745–753, 2023.
- [9] N. Chatrina Siregar, R. Ruli, A. Siregar, ; M Yoga, and D. Sudirman, “Implementasi Metode Naïve Bayes Classifier (NBC) Pada Komentar Warga Sekolah Mengenai Pelaksanaan

- Pembelajaran Jarak Jauh (PJJ)," *J. Teknol. Aliansi Perguru. Tinggi BUMN*, vol. 3, no. 1, pp. 102–110, 2020.
- [10] H. F. Putro, R. T. Vulandari, and W. L. Y. Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," *J. Teknol. Inf. dan Komun.*, vol. 8, no. 2, pp. 20–24, 2020, doi: 10.30646/tikomsin.v8i2.500.
- [11] T. H. Apandi and C. A. Sugianto, "Analisis Komparasi Machine Learning Pada Data Spam Sms," *J. TEDC*, vol. 12, no. 1, pp. 58–62, 2018.
- [12] A. Nurian, T. N. Padilah, and Garno, "Analisis Sentimen Terhadap Pelayanan Disdukcapil Karawang Menggunakan Naive Bayes Classifier," *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 12, no. 2, 2024.
- [13] S. A. Faridzi, F. S. Azizah, F. Mustafa, A. N. Putri, G. Ramadhika, F. R. Aditya, R. S. Fadilah, Y. Habibi, M. Sutrisno, Jumail, R. D. Risanty, and N. Rosanti, "Pengolahan Data: Pemahaman Gempa Bumi di Indonesia Melalui Pendekatan Data Mining," *Jurnal Pengabdian Kolaborasi dan Inovasi IPTEKS*, vol. 2, no. 1, Feb. 2024.
- [14] N. Novianti, M. Zarlis, and P. Sihombing, "Penerapan Algoritma Adaboost Untuk Peningkatan Kinerja Klasifikasi Data Mining Pada Imbalance Dataset Diabetes," *Jurnal Media Informatika Budidarma*, vol. 6, no. 2, Apr. 2022.
- [15] A. Ramadhan and H. D. Bhakti, "Klasifikasi Penilaian Kinerja Karyawan Menggunakan Algoritme Naïve Bayes (Studi Kasus PT. As Sabar Sukses Berkah)," *Jurnal Multidisiplin Saintek*, vol. 4, no. 2, 2024.
- [16] S. F. Tahir and C. A. Sugianto, "Optimasi Naive Bayes Menggunakan Algoritma Genetika Pada Klasifikasi Komentar Cyberbullying Pada Media Sosial X," *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 12, no. 3, Aug. 2024.