

SENTIMEN ANALISIS TIM SEPAK BOLA DI PIALA ASIA 2023 BERDASARKAN PENGGUNA TWITTER (X) MENGGUNAKAN ALGORITMA MULTINOMIAL NAÏVE BAYES

Muhammad Naufal¹⁾, Budi Arifitama²⁾, Silvester Dian Handy Permana³⁾

^{1,2,3)} Teknik Informatika Universitas Trilogi, Jalan TMP Kalibata 1, Jakarta Selatan, Indonesia
email : muhammad.naufal@trilogi.ac.id¹⁾

Abstrak

Perkembangan media sosial mengubah cara kita berkomunikasi dan berinteraksi, Dimana media sosial telah menyediakan sumber data yang kaya akan opini dan persepsi dari berbagai pengguna. Dalam konteks Olahraga, sentimen analisis terhadap Tim Sepak Bola Nasional Indonesia dapat memberikan wawasan yang baik tentang dukungan persepsi publik terhadap Tim Sepak Bola Nasional Indonesia dan juga menjadi bahan evaluasi tim nasional indonesia di masa yang akan datang. Studi ini akan menghasilkan opini publik positif dan negatif pada saat kondisi tim nasional indonesia mengalami kekalahan, kemenangan dan keseluruhan selama mengikuti ajang turnamen piala asia 2023, studi ini menggunakan pendekatan algoritma Multinomial Naïve Bayes Classification (MNB) untuk mengklasifikasikan dan training data, menggunakan library Valence Aware Dictionary and sEntiment Reasoner (VADER) untuk melakukan pelabelan otomatis pada data dan Term Frequency-Inverse Document Frequency (TF-IDF) untuk mengekstraksi fitur terhadap data tim sepak bola nasional Indonesia dalam turnamen Piala Asia 2023. Algoritma MNB digunakan untuk mengklasifikasikan teks yang sudah dilabeli secara otomatis menggunakan VADER . Data untuk studi ini diperoleh melalui proses pengambilan data dari sosial media X dengan kata kunci "timnas Indonesia Piala Asia 2023" dalam rentang tanggal pada saat turnamen piala asia 2023 yaitu 12 Januari 2024 hingga 10 Februari 2024. Hasil dari studi yang dilakukan oleh penulis pada studi berjudul "Sentimen Analisis Tim Sepak Bola di Piala Asia 2023 Berdasarkan Pengguna Twitter(X) Menggunakan Algoritma Multinomial Naïve Bayes" menghasilkan akurasi yang cukup baik yaitu 81%, 79% dan 79% pada kategori Keseluruhan, Kekalahan dan Kemenangan dengan menggunakan metode evaluasi Confussion Matrix.

Kata Kunci :

Algoritma Multinomial Naïve Bayes Classification (MNB), Crawling Data, Reasoner(VADER), Sentiment Analysis, TF-IDF, Valence Aware Dictionary and sEntiment Reasoner(VADER).

Abstract

The development of social media has transformed the way we communicate and interact, providing a rich source of data filled with opinions and perceptions from various users. In the context of sports, sentiment analysis of the Indonesian National Football Team can offer valuable insights into public support and perceptions, as well as serve as an evaluation tool for the team's future performance. This study aims to analyze public opinions, both positive and negative, during times of the Indonesian national team's defeats, victories, and overall performance throughout the 2023 Asian Cup tournament. The study employs the Multinomial Naïve Bayes Classification (MNB) algorithm to classify and train the data, using the Valence Aware Dictionary and sEntiment Reasoner (VADER) library for automatic labeling and Term Frequency-Inverse Document Frequency (TF-IDF) to extract features from data related to the Indonesian National Football Team during the 2023 Asian Cup tournament. The MNB algorithm is used to classify text that has been automatically labeled using VADER. Data for this study was collected from the social media platform X using the keyword "Indonesian national team Asian Cup 2023" within the date range of the Asian Cup tournament, from January 12, 2024, to February 10, 2024. The results of this study, titled "Sentiment Analysis of the Football Team in the 2023 Asian Cup Based on Twitter (X) Users Using the Multinomial Naïve Bayes Algorithm," demonstrate good accuracy, with 81%, 79%, and 79% in the Overall, Defeat, and Victory categories, respectively, using the Confusion Matrix evaluation method.

Keywords :

Algoritma Multinomial Naïve Bayes Classification (MNB), Crawling Data, Reasoner(VADER), Sentiment Analysis, TF-IDF, Valence Aware Dictionary and sEntiment Reasoner(VADER).

1. PENDAHULUAN

Dalam era digital saat ini, media sosial seperti *Twitter* (X) telah menjadi platform utama bagi pengguna untuk mengekspresikan pendapat, perasaan, dan dukungan mereka terhadap berbagai topik, termasuk olahraga dan khususnya tim sepak bola nasional di Indonesia. Menurut [1], *twitter* (X) adalah sosial media yang sangat memudahkan penggunaannya dalam berinteraksi dan berteman dengan pengguna lain dan memiliki berbagai fitur salah satunya “apa yang sedang populer” dimana dapat melihat kicauan yang sedang populer di kalangan pengguna *twitter* (X). Pada studi ini, penulis melakukan studi bagaimana respons dan tanggapan publik pengguna *twitter* (X) di Indonesia terhadap tim sepak bola nasional Indonesia pada gelaran Piala Asia 2023 berdasarkan pengguna *twitter* (X), Dimana studi ini menggunakan pendekatan algoritma *MNB* untuk melakukan klasifikasi data terhadap *cutan-cuitan twitter* (X), pada studi ini juga menggunakan *library Vader*, Menurut penelitian berjudul “*Comparison of the Multinomial Naive Bayes Algorithm and Decision Tree with the Application of AdaBoost in Sentiment Analysis Reviews PeduliLindungi Application*” [2] dimana penelitian ini mendapatkan akurasi sebesar 93,45% dimana pada penelitian ini menggunakan *labeling otomatis VADER* dan dikombinasikan juga menggunakan algoritma *Multinomial Naive Bayes* untuk melakukan pelabelan otomatis terhadap *dataframe* dan juga menggunakan *TF-IDF* untuk melakukan ekstraksi fitur supaya bisa memberikan *input* pada Algoritma *MNB*, Menurut penelitian berjudul “*The Impact of Feature Extraction to Naive Bayes Based Sentiment Analysis on Review Dataset of Indihome Services*” [3] mendapatkan akurasi yang cukup baik, yaitu sebesar 96% menggunakan ekstraksi fitur *TF-IDF*, Dimana dilakukan implementasi terhadap algoritma *Naive Bayes* pada dataset *Indihome Services* .

Studi sebelumnya berjudul "Analisis Sentimen Wacana Kenaikan Harga Tiket Candi Borobudur Menggunakan *Multinomial Naïve Bayes*" [4] menggunakan algoritma *Multinomial Naïve Bayes* untuk analisis sentimen terkait kenaikan harga tiket. Kelebihan studi ini adalah algoritma yang digunakan berhasil mencapai akurasi 81%. Namun, kelemahannya adalah hanya menganalisis sentimen terhadap tiket tanpa mempertimbangkan faktor lain seperti fasilitas atau kebersihan. Untuk studi selanjutnya, disarankan menambahkan variasi analisis sentimen untuk mendapatkan informasi yang lebih spesifik.

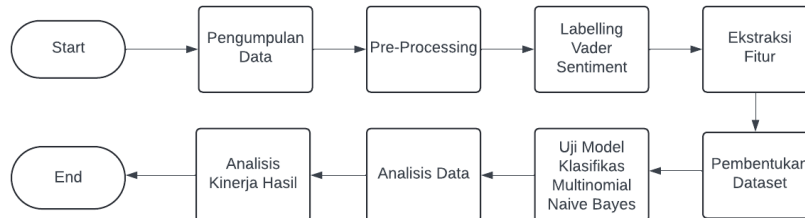
Lalu berdasarkan studi lain berjudul "Analysis of Public Sentiment Related to the Failure of Indonesia to Host U-20 Using *Multinomial Naïve Bayes Classifier*" [5] menggunakan algoritma *Multinomial Naïve Bayes* untuk analisis sentimen publik terkait kegagalan Indonesia menjadi tuan rumah U-20. Kelebihan studi ini adalah akurasi 97,26%, menunjukkan efektivitas algoritma untuk data olahraga, khususnya sepak bola. Kekurangannya adalah jumlah data yang kecil, hanya 250 data. Untuk studi selanjutnya, disarankan menggunakan jumlah data yang lebih besar.

Studi ini menganalisis sentimen masyarakat terhadap tim sepak bola Indonesia menggunakan metode *Multinomial Naïve Bayes*. Ini bertujuan membantu PSSI memahami sentimen publik dan mengukur kinerja organisasi sepak bola Indonesia. Tahapan studi meliputi pengumpulan data dengan *Tweet Harvest*, menurut penelitian berjudul “Analisis Sentimen Bakal Calon Presiden Indonesia 2024 Dengan Algoritma *Naïve Bayes*” [6] mendapatkan nilai akurasi sebesar 80% dimana pada penelitian ini menggunakan *library Tweet Harvest*, pre-processing (*cleaning, case folding, tokenization, stopword removal, stemming, dan translate*), *labeling otomatis* dengan *VADER*, ekstraksi fitur menggunakan *TF-IDF*, pembentukan dataset, klasifikasi, dan analisis hasil dengan *Confusion Matrix*. Hasil studi diharapkan memberi pandangan kepada PSSI tentang kinerja tim nasional dan menjadi dasar strategi pengembangan tim yang lebih efektif dan adaptif. Tujuan dilakukan studi ini adalah dengan

adanya sentimen analisis ini akan menjadi acuan terhadap kinerja dari organisasi sepak bola Indonesia untuk evaluasi pada turnamen-turnamen selanjutnya.

2. METODE / ALGORITMA

Pada studi ini metode yang digunakan terdiri dari beberapa tahapan dari proses penyimpulan data sampai evaluasi kinerja klasifikasi terhadap data yang dikumpulkan. Adapun metode studi dapat dilihat pada gambar berikut ini:



Gambar 1 Langkah Studi

Pengumpulan Data: Data yang terkait dengan timnas Indonesia dikumpulkan dari *Twitter* menggunakan kata kunci "timnas Indonesia Piala Asia 2023" dalam rentang tanggal 12 Januari 2024 hingga 10 Februari 2024. Data ini mencakup semua *tweet* yang menyebutkan tim nasional Indonesia selama periode tersebut. Tanggal 19 Januari 2024 ditetapkan sebagai hari kemenangan timnas, sementara periode kekalahan meliputi tanggal 12 hingga 18 Januari dan 20 hingga 28 Januari 2024. Pengumpulan data dibatasi hingga 3000 *tweet*.

Pre-Processing: *Preprocessing* data dalam studi ini melibatkan beberapa tahapan mendalam untuk mempersiapkan data *Twitter* yang dikumpulkan. Tahapan pertama adalah *Cleaning* data, di mana data mentah dari *Twitter* disaring untuk menghilangkan karakter yang tidak diinginkan seperti tautan atau simbol khusus. Kemudian, dilakukan *Case Folding* untuk menjaga konsistensi dalam teks dengan mengubah semua huruf menjadi huruf kecil. Proses selanjutnya adalah *Tokenization*, di mana teks dibagi menjadi *token* atau unit kata untuk analisis lebih lanjut. *Stopword Removal* dilakukan untuk menghapus kata-kata umum atau *stop words* yang tidak memberikan kontribusi signifikan dalam analisis sentimen. Terakhir, dilakukan *Translate* untuk menerjemahkan teks jika diperlukan, memungkinkan analisis yang lebih luas dan representatif terhadap data yang terkumpul.

Data Labelling: Pada studi ini *labelling* otomatis menggunakan *library VaderSentiment* dimana digunakan untuk menentukan skor sentimen setiap data *tweet* yang di dapatkan dari hasil *crawling* data sebelumnya, Untuk sentimen label yang digunakan adalah Positif dan Negatif dengan nilai skor yang dijadikan acuan pelabelan, jika ulasan mempunyai skor > 0 maka ulasan tersebut diberi label positif, jika ulasan mempunyai skor < 0 maka ulasan tersebut diberi label negatif.

Ekstraksi Fitur: Ekstraksi fitur adalah hasil ekstraksi dari data yang menghasilkan nilai numerik yang dapat menjadi *input* bagi model klasifikasi.

Pembentukan Dataset: Hasil ekstraksi fitur dari *TF-IDF* berupa matriks numerik akan digunakan sebagai fitur dalam pembentukan dataset untuk melatih dan menguji model klasifikasi. Ini akan membagi data *tweet* menjadi 3 pembagian, yaitu 70% dan 30%, 80% dan 20% dan 90% dan 10%. Pembagian ini dilakukan untuk mencegah *overfitting* dan memastikan bahwa model yang dilatih dapat menggeneralisasi dengan baik pada data yang belum pernah dilihat sebelumnya. Data pelatihan akan digunakan untuk melatih model, dan data pengujian akan digunakan untuk menguji kinerja model yang telah dilatih.

Implementasi Algoritma MNB: Hasil dari pembentukan dataset dimana pada studi ini menggunakan data latih sebanyak 80% yaitu 2400 dan 20% yaitu 600 untuk data testing, dimana hasil ekstraksi pada *TF-IDF* akan menjadi *input* untuk model klasifikasi *Multinomial Naïve Bayes* tepatnya pada tahapan perhitungan *Likelihood* (Probabilitas Kondisional) dan menggunakan library *MultinomialNB* dari *sklearn.naive_bayes*. Klasifikasi *Multinomial Naïve Bayes* adalah salah satu bentuk dari algoritma *Naïve Bayes* yang umumnya digunakan untuk klasifikasi teks. Metode ini bekerja berdasarkan asumsi bahwa setiap fitur (kata dalam teks) secara independen berkontribusi terhadap klasifikasi, dan nilai fitur tersebut dianggap terdistribusi *multinomial*. Advokasi: digunakan untuk kegiatan yang berupa pendampingan.

Analisa Data Hasil: Analisis data dilakukan untuk mengevaluasi kinerja model klasifikasi *Multinomial Naïve Bayes* dalam mengklasifikasikan *tweet* sentimen, pada tahap ini penulis menggunakan metode *Confussion Matrix*, pada metode *Confussion Matrix* memiliki empat nilai prediksi yang dihasilkan dari model klasifikasi, empat nilai prediksi dari hasil prediksi model klasifikasi adalah sebagai berikut:

1. *True Positive (TP)* adalah jumlah data positif yang di prediksi benar
2. *True Negative (TN)* adalah jumlah data negatif yang di prediksi benar
3. *False Positive (FP)* adalah jumlah data negatif yang di prediksi positif (salah)
4. *False Negative (FN)* adalah jumlah data positif yang di prediksi negatif (salah)

Pada *confusion matrix* akan melakukan perhitungan terhadap beberapa *matrix*, yaitu:

a. Akurasi (*Accuracy*)

Akurasi merupakan rasio prediksi yang benar (positif dan negatif) dengan total data yang diuji. *Matrix* ini memberikan gambaran umum tentang seberapa baik model dalam mengklasifikasikan sentiment, metode perhitungan dari akurasi adalah sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \dots \text{(Persamaan 1)}$$

b. Presisi (*Precision*)

Presisi mengukur proporsi data positif yang telah diprediksi dengan benar dari keseluruhan data yang diprediksi positif. *Matrix* ini menunjukkan seberapa tepat model dalam mengidentifikasi sentimen positif. Metode perhitungan dari presisi adalah sebagai berikut:

$$\text{Presisi} = \frac{TP}{TP+FP} \dots \text{(Persamaan 2)}$$

c. *Recall (Sensitivity)*

Recall, juga dikenal sebagai *sensitivity*, mengukur proporsi data positif yang diprediksi dengan benar dari semua data positif yang sebenarnya. *Matrix* ini menunjukkan seberapa baik model dalam mendeteksi sentimen positif. Metode perhitungan dari *recall* adalah sebagai berikut:

$$\text{Recall} = \frac{TP}{TP+FN} \dots \text{(Persamaan 3)}$$

d. *F1-Score*

F1-Score adalah rata-rata harmonik dari *presisi* dan *recall*. *Matrix* ini memberikan keseimbangan antara *presisi* dan *recall*, sehingga berguna untuk mengevaluasi performa model secara keseluruhan.

$$F1 - \text{Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \dots \text{(Persamaan 4)}$$

Kinerja Hasil: Dalam studi ini, kinerja model di analisis menggunakan berbagai *matrix* seperti *akurasi*, *presisi*, *recall* dan *F1-Score*. Hasil evaluasi ini memberikan gambaran tentang efektivitas model dalam melakukan klasifikasi pada *tweet* sentimen, pada bagian ini akan menghasilkan deskriptif analisis berdasarkan hasil evaluasi model berdasarkan beberapa *matrix* sebelumnya, dan akan melakukan visualisasi data terkait persentase sentimen positif dan negatif pada kategori kekalahan, kemenangan dan keseluruhan.

3. HASIL DAN PEMBAHASAN

Pengumpulan Data

Pengumpulan data adalah langkah awal yang dilakukan pada sentimen analisis, dalam proses pengumpulan data menggunakan *library Tweet Harvest*, dimana data diambil dari *tweet* user sosial media *Twitter(X)* menggunakan kata kunci “Timnas Indonesia Piala Asia 2023” dalam rentang waktu 12 Januari 2024 sampai 10 Februari 2024 dan data yang berhasil dikumpulkan sebanyak 3416 *tweet*.

index	created_at	full_text
0	Wed Feb 07 11:54:17 +0000 2024	#AksiBerani Yakob Sayuri selama gelaran Piala Asia Qatar 2023. Yakob bermain total 309 menit dalam empat pertandingan Timnas Indonesia di Piala Asia 2023. Dalam kurun waktu tersebut Yakob berhasil mengemas empat dibel sukses 46 operan sukses dua umpan kunci dan satu assist. https://t.co/EXbdJfPdB9
1	Wed Feb 07 10:28:00 +0000 2024	Masalah mulai mendatangi Shin Tae-yong jelang memimpin Timnas U-23 Indonesia di Piala Asia U-23 2024 diawali pemanggilan pemain dari Liga 1 2023 https://t.co/xvY8f5FzWg
2	Wed Feb 07 10:28:00 +0000 2024	Masalah mulai mendatangi Shin Tae-yong jelang memimpin Timnas U-23 Indonesia di Piala Asia U-23 2024 diawali pemanggilan pemain dari Liga 1 2023 https://t.co/Cocp5w9KP0
3	Wed Feb 07 10:05:00 +0000 2024	3 Hal yang Bisa Dipelajari Timnas Indonesia dari Kisah Cinderella Yordania di Piala Asia 2023 https://t.co/Q2grUogfKu
4	Wed Feb 07 09:29:47 +0000 2024	Baru-baru ini media Vietnam memberikan analisis terkait perjalanan wakil Asia Tenggara di Piala Asia 2023. Menurut mereka Timnas Indonesia merupakan tim Asia Tenggara yang paling mengejutkan https://t.co/aAsrB6lgDy
5	Wed Feb 07 09:29:38 +0000 2024	Baru-baru ini media Vietnam memberikan analisis terkait perjalanan wakil Asia Tenggara di Piala Asia 2023. Menurut mereka Timnas Indonesia merupakan tim Asia Tenggara yang paling mengejutkan https://t.co/ZZ7B1DO1GZ
6	Wed Feb 07 09:17:03 +0000 2024	Tembus Final Yordania Ternyata Segrup dengan Timnas Indonesia di Kualifikasi Piala Asia 2023 https://t.co/zyxBdiSPfB
7	Wed Feb 07 09:01:29 +0000 2024	#TauGakSih Indonesia berhasil menunjukkan kehebatannya lewat dunia esport. Termasuk kemenangan Timnas Indonesia yang berhasil menjadi juara esport Football di Piala Asia 2023. #kabarbak https://t.co/WWajE18l
8	Wed Feb 07 08:21:00 +0000 2024	Timnas Yordania menorehkan sejarah dengan lolos untuk pertama kali ke babak final Piala Asia pada edisi 2023. Namun Yordania lolos ke final dengan dugaan main sabun di fase grup yang merugikan Timnas Indonesia https://t.co/0opmWwVAh
9	Wed Feb 07 08:21:00 +0000 2024	Timnas Yordania menorehkan sejarah dengan lolos untuk pertama kali ke babak final Piala Asia pada edisi 2023. Namun Yordania lolos ke final dengan dugaan main sabun di fase grup yang merugikan Timnas Indonesia https://t.co/YPyBrd9eS
10	Wed Feb 07 07:52:10 +0000 2024	Timnas Indonesia Masih Akan Jalani 4 Liga Kualifikasi Piala Dunia 2026 Zona Asia. Jangan Kalah Lagi Minimal Rebut 10 Poin! Setelah terhenti di 16 besar Piala Asia 2023 Timnas Indonesia kini fokus untuk menjatani empat laga tersisa di putaran kedua Kualifikasi Piala Dunia 2026 https://t.co/D4G7dt5piQ
11	Wed Feb 07 07:12:48 +0000 2024	Yordania lolos ke final Piala Asia 2023 dengan dugaan main sabun di fase grup yang merugikan Timnas Indonesia. Ini menjadi sejarah bagi Yordania! https://t.co/0opmWwVAh
12	Wed Feb 07 07:07:05 +0000 2024	Yordania lolos ke final Piala Asia 2023 dengan dugaan main sabun di fase grup yang merugikan Timnas Indonesia. Ini menjadi sejarah bagi Yordania! https://t.co/YPyBrd9eS

Gambar 2 Tampilan Hasil *Crawling Data*

Pada gambar 2 diatas merupakan sebuah contoh tampilan hasil dari proses *crawling* data yang terdiri dari *tweet text* yaitu *full_text* dan *created_at* yaitu pada waktu kapan *tweet* tersebut dibuat

1. Pre – Processing

Setelah data dikumpulkan, kemudian dilakukan *pre-processing* atau melakukan persiapan data sebelum dilakukan implementasi pada model klasifikasi

2.1. *Cleaning Data*.

Untuk memproses data yang telah dikumpulkan, dilakukan proses pembersihan data untuk menghapus data yang berisi *URL*, *emoji*, *retweet*, simbol, tagar, dan spasi yang tidak diperlukan. Tabel 1 dibawah ini menunjukkan proses sebelum dan sesudah data masuk ke tahap pembersihan.

Tabel 1 *Cleaning Data*

Sebelum	Sesudah
Menurut Erick Thohir Timnas Indonesia akan melakoni laga yang tidak mudah melawan Jepang di Piala Asia 2023. Justin Hubner dan kawan-kawan tidak boleh loyo! IDJP #AsianCup2023 Selengkapny	Menurut Erick Thohir Timnas Indonesia akan melakoni laga yang tidak mudah melawan Jepang di Piala Asia Justin Hubner dan kawan-kawan tidak boleh loyo Asiancup Selengkapny
Terima Kasih Qatar Timnas Indonesia Selangkah Lagi Lolos ke 16 Besar Piala Asia 2023!	Terima Kasih Qatar Timnas Indonesia Selangkah Lagi Lolos ke Besar Piala Asia

2.2. *Case Folding*

Setelah melakukan *Cleaning Data*, selanjutnya dilakukan *Case Folding* dimana bertujuan untuk mengubah semua karakter menjadi huruf kecil.

Tabel 2 Case Folding

Sebelum	Sesudah
Menurut Erick Thohir Timnas Indonesia akan melakoni laga yang tidak mudah melawan Jepang di Piala Asia Justin Hubner dan kawankawan tidak boleh loyo Asiacup Selengkapnya	menurut erick thohir Timnas Indonesia akan melakoni laga yang tidak mudah melawan Jepang di Piala Asia Justin Hubner dan kawankawan tidak boleh loyo Asiacup selengkapnya
Terima Kasih Qatar Timnas Indonesia Selangkah Lagi Lolos ke Besar Piala Asia	terima kasih Qatar Timnas Indonesia selangkah lagi lolos ke besar Piala Asia

2.3. Tokenization

Setelah dilakukan *Case Folding* selanjutnya dilakukan *Tokenization*, yaitu memecahkan kata menjadi unit-unit kecil.

Tabel 3 Tokenization

Sebelum	Sesudah
menurut Erick Thohir Timnas Indonesia akan melakoni laga yang tidak mudah melawan Jepang di Piala Asia Justin Hubner dan kawankawan tidak boleh loyo Asiacup selengkapnya	[menurut, erick, thohir, Timnas, Indonesia, akan, melakoni, laga, yang, tidak, mudah, melawan, Jepang, di, Piala, Asia, Justin, Hubner, dan, kawankawan, tidak, boleh, loyo, Asiacup, selengkapnya]
terima kasih Qatar Timnas Indonesia selangkah lagi lolos ke besar Piala Asia	[terima, kasih, Qatar, Timnas, Indonesia, selangkah, lagi, lolos, ke, besar, Piala, Asia]

2.4. Stopword Removal

Setelah dilakukan *Tokenization* selanjutnya dilakukan *Stopword Removal* dimana di hapuskannya kalimat sambung dimana tidak memiliki makna spesifik pada *dataframe*

Tabel 4 Stopword Removal

Sebelum	Sesudah
[menurut, erick, thohir, Timnas, Indonesia, akan, melakoni, laga, yang, tidak, mudah, melawan, Jepang, di, Piala, Asia, Justin, Hubner, dan, kawankawan, tidak, boleh, loyo, Asiacup, selengkapnya]	[erick, thohir, Timnas, Indonesia, melakoni, laga, mudah, melawan, Jepang, Piala, Asia, Justin, Hubner, kawankawan, tidak, loyo, Asiacup]
[terima, kasih, Qatar, Timnas, Indonesia, selangkah, lagi, lolos, ke, besar, Piala, Asia]	[terima, kasih, Qatar, Timnas, Indonesia, selangkah, lolos, Piala, Asia]

2.5. Stemming

Setelah melakukan *Stopwords Removal*, langkah selanjutnya adalah tahap stemming adalah melakukan perubahan pada kalimat menjadi kata dasar.

Tabel 5 Stemming

Sebelum	Sesudah
[erick, thohir, Timnas, Indonesia, melakoni, laga, mudah, melawan, Jepang, Piala, Asia, Justin, Hubner, kawankawan, tidak, loyo, Asiacup]	[erick, thohir, Timnas, Indonesia, lakon, laga, mudah, lawan, Jepang, Piala, Asia, Justin, Hubner, kawankawan, tidak, loyo, Asiacup]
[terima, kasih, Qatar, Timnas, Indonesia,	[terima, kasih, Qatar, Timnas, Indonesia, selang,

selangkah, lolos, piala, asia]	lolos, piala, asia]
--------------------------------	---------------------

2.6. Translate

Setelah melakukan *stemming*, selanjutnya dilakukan translate data menggunakan *library googletrans* dimana dilakukan

Tabel 6 Translate

Sebelum	Sesudah
[erick, thohir, timnas, indonesia, lakon, laga, mudah, lawan, jepang, piala, asia, justin, hubner, kawankawan, tidak, loyo, asiancup]	[erick, thohir, timnas, indonesia, lakon, laga, mudah, lawan, jepang, piala, asia, justin, hubner, kawankawan, tidak, loyo, asiancup]
[terima, kasih, qatar, timnas, indonesia, selang, lolos, piala, asia]	[terima, kasih, qatar, timnas, indonesia, selang, lolos, piala, asia]

2. Pelabelan Data

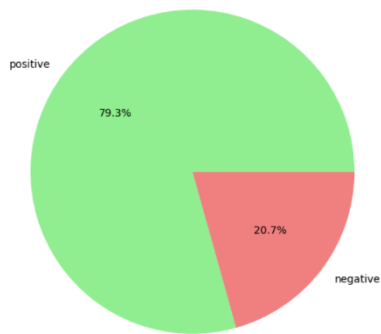
Setelah melakukan terjemahan data terhadap dataframe, langkah selanjutnya adalah memberikan label dengan membaginya ke dalam dua kategori yaitu Positif dan Negatif. Penentuan label ini dilakukan berdasarkan pada nilai *compound score* yang dihasilkan oleh model Vader Sentimen, dengan kondisi sebagai berikut:

- a) Dalam kasus nilai *compound score* yang lebih besar dari 0, maka ulasan akan diberi label "sentimen positif", menunjukkan adanya kecenderungan positif dalam ulasan tersebut.
- b) Jika nilai *compound score* kurang dari 0, maka ulasan akan diberi label "sentimen negatif", menunjukkan adanya kecenderungan negatif dalam ulasan tersebut.

full_text	tanggal	hari	jam	cleaned_text	case_folding	tokenization	stopwords	Tweet	Translate	Joined Translate	Translate_Stopwords	sentiment_score	label
4ksiBerani Yakob Sayuri selama gelaran Piala Asia Qatar Yakob bermain total menit dalam empat pertandingan Timnas Indonesia di Piala Asia 2023. Yakob rmain total 309 menit slam empat pertandingan Timnas donesia di Piala Asia 2023. Dalam kurun waktu rsebut Yakob berhasil engemas empat dribel ikses 46 operan sukses ja umpan kunci dan itu assist tps://t.co/EXbdJdP4B9	7/2/2024	Wednesday	11:54:17	Yakob Sayuri selama gelaran Piala Asia Qatar Yakob bermain total menit dalam empat pertandingan Timnas Indonesia di Piala Asia 2023. Dalam kurun waktu tersebut Yakob berhasil mengemas empat dribel sukses operan sukses dua umpan kunci dan satu assist	yakob sayuri selama gelaran piala asia qatar yakob bermain total menit dalam empat pertandingan timnas indonesia di piala asia 2023 dalam kurun waktu tersebut yakob berhasil mengemas empat dribel sukses operan sukses dua umpan kunci dan satu assist	[yakob, sayuri, selama, gelaran, piala, asia, qatar, yakob, bermain, total, menit, dalam, empat, pertandingan, timnas, indonesia, di, piala, asia, 2023, dalam, kurun, waktu, tersebut, yakob, berhasil, mengemas, empat, dribel, sukses, operan, sukses, dua, umpan, kunci, dan, satu, assist]	['yakob', 'sayuri', 'selama', 'gelaran', 'piala', 'asia', 'qatar', 'yakob', 'bermain', 'total', 'menit', 'dalam', 'empat', 'pertandingan', 'timnas', 'indonesia', 'di', 'piala', 'asia', '2023', 'dalam', 'kurun', 'waktu', 'tersebut', 'yakob', 'berhasil', 'mengemas', 'empat', 'dribel', 'sukses', 'operan', 'sukses', 'dua', 'umpan', 'kunci', 'dan', 'satu', 'assist']	[yakob, sayuri, gelaran, piala, asia, qatar, yakob, bermain, total, menit, pertandingan, timnas, indonesia, piala, asia, kurun, yakob, berhasil, mengemas, dribel, sukses, operan, sukses, umpan, kunci, assist]	[yakob, sayuri, holds, asian, cup, qatar, yakob, playing, total, minutes, indonesian, national, team, asian, cup, yakob, gold, results, dribel, success, success]	yakob sayuri holds asian cup qatar yakob playing total minutes indonesian national team asian cup yakob gold results dribel success success	0.8481	positive	
asalah mulai mendatangi Shin Taeyong jelang memimpin Timnas U-23 Indonesia di Piala Asia U-23 2024	7/2/2024	Wednesday	10:28:00	Masalah mulai mendatangi Shin Taeyong jelang memimpin Timnas U-23 Indonesia di Piala Asia U-23 2024	masalah mulai mendatangi shin taeyong jelang memimpin timnas u indonesia di piala asia u 23 2024	[masalah, mulai, mendatangi, shin, taeyong, jelang, memimpin, timnas, u, indonesia, di, piala, asia, u, 23, 2024]	['shin', 'taeyong', 'jelang', 'memimpin', 'timnas', 'u', 'indonesia', 'di', 'piala', 'asia', 'u', '23', '2024']	[shin, taeyong, jelang, memimpin, timnas, u, indonesia, piala, asia, u, 23, 2024]	[shin, taeyong, ahead, of, leading, the, indonesian, national, team, u, indonesia, u, initially]	shin taeyong ahead of leading the indonesian national team u indonesia u initially	0.2023	positive	

Gambar 3 Pelabelan Data

Pada gambar 3, menampilkan hasil pelabelan data seperti positif dan negatif yang telah diproses menggunakan *library Vader Sentiment* yang di sesuaikan dengan *sentiment_score*



Gambar 4 Diagram Pie Jumlah



Gambar 5 Wordcloud Analisis

Berdasarkan pada gambar 4 sentimen paling banyak adalah sentimen positif dengan persentase 79.3% sedangkan untuk sentimen positif memiliki persentase 20.7% , dan pada gambar 5 menunjukkan visualisasi *wordcloud* ulasan yang sering muncul dengan tampilan kata yang besar

3. Pembentukan Dataset

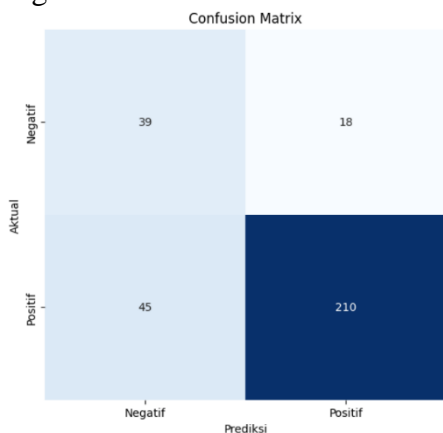
Sebelum implementasi terhadap model klasifikasi, data akan dilakukan pembagian data menjadi dua kategori, yaitu data pelatihan dan data testing, dimana di studi ini saya melakukan percobaan terhadap pembagian dengan rasio 80:20, setelah dilakukan pembagian data maka data pelatihan menjadi 2733 dan data testing menjadi 683.

4. Implementasi Model *MNB*

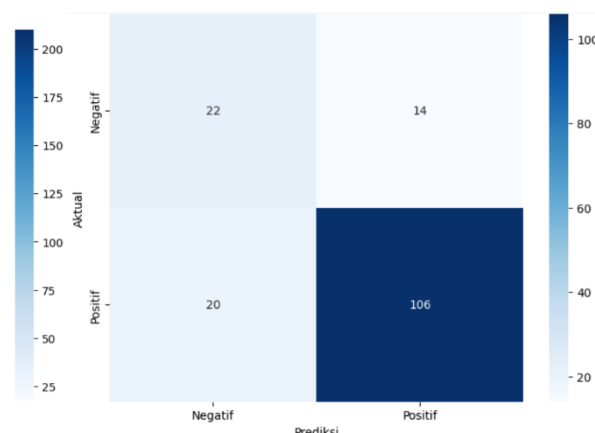
Setelah membagi data pelatihan dan data testing dengan rasio 80:20, 70:30, dan 90:10, penulis membagi data berdasarkan tanggal kemenangan, kekalahan, dan keseluruhan. Tanggal kemenangan adalah 19 Januari 2024, sementara kekalahan terjadi dari 12 Januari hingga 18 Januari 2024 dan dari 24 Januari 2024 hingga 12 Februari 2024. Selama implementasi Model *MNB*, penulis menghadapi masalah ketidakseimbangan distribusi kelas pada kolom label, di mana kategori TP (*True Positive*) dan FN (*False Negative*) menghasilkan nilai 0, menunjukkan adanya ketidakseimbangan data. Untuk mengatasi masalah ini, penulis menggunakan metode *Synthetic Minority Oversampling Technique (SMOTE)*, yang efektif meningkatkan representasi kelas minoritas sehingga hasil klasifikasi menjadi lebih seimbang.

Analisis Data

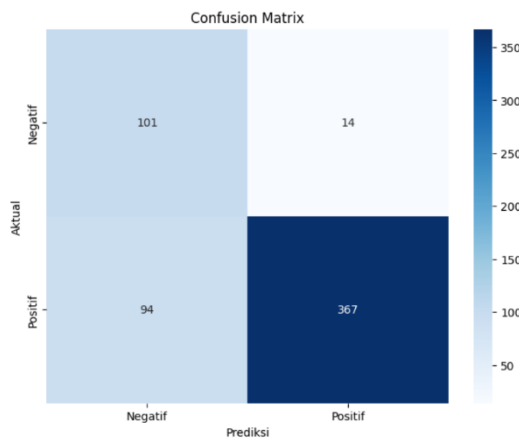
Kinerja model algoritma *Naïve Bayes* dievaluasi dengan *Confusion matrix* yang ditampilkan pada figur berikut.



Gambar 6 Confusion Matrix Kekalahan



Gambar 7 Confusion Matrix Kemenangan



Gambar 8 *Confusion Matrix* Keseluruhan

Pada gambar 6, 7, 8 menunjukkan gambar *Confusion Matrix* yang dapat menghasilkan beberapa parameter seperti *accuracy*, *precision*, *recall*, dan *f1-score* di ukuran. Hasil perhitungan parameter tersebut dideskripsikan pada tabel 7.

Tabel 7 Hasil Perhitungan Parameter *Confusion Matrix*

<i>Kategori</i>	<i>Akurasi</i>	<i>Presisi</i>	<i>Recall</i>	<i>F1-Score</i>
Kekalahan	79%	92%	82%	86%
Kemenangan	79%	83%	84%	86%
Keseluruhan	81%	96%	79%	81%

5. Analisis Kinerja Hasil

Hasil dari Tabel 7 menunjukkan evaluasi model klasifikasi sentimen menggunakan algoritma *Multinomial Naïve Bayes* dengan metrik Akurasi, Presisi, *Recall*, dan *F1-Score*. Untuk kategori "Kalah", model memiliki akurasi 79%, yang berarti model benar dalam mengklasifikasikan 79% dari data terkait kekalahan. Presisi untuk kategori ini adalah 92%, menunjukkan bahwa dari semua prediksi yang diklasifikasikan sebagai "kalah", 92% di antaranya benar. *Recall* sebesar 82% menunjukkan bahwa dari semua data yang benar-benar berlabel "kalah", model berhasil mengidentifikasi 82% di antaranya. *F1-Score* untuk kategori ini adalah 86%, menunjukkan keseimbangan antara presisi dan *recall*. Untuk kategori "Menang", model juga memiliki akurasi 79%. Presisi untuk kategori ini adalah 83%, dengan *recall* sebesar 84%, dan *F1-Score* 86%. Secara keseluruhan, model memiliki akurasi 81%, yang menunjukkan bahwa model ini benar dalam mengklasifikasikan 81% dari total data. Presisi keseluruhan adalah 96%, *recall* keseluruhan adalah 79%, dan *F1-Score* keseluruhan adalah 81%, menunjukkan keseimbangan yang baik antara presisi dan *recall* pada tingkat keseluruhan.

4. KESIMPULAN

Kesimpulan dari studi ini adalah Hasil akurasi tertinggi didapat pada waktu keseluruhan dibandingkan waktu kalah dan menang, dimana metode *multinomial naïve bayes* memiliki Tingkat akurasi yang cukup baik dimana mendapatkan akurasi sebesar 81% pada waktu keseluruhan, saran kedepannya terhadap studi ini adalah dapat menggunakan metode klasifikasi lain seperti *Support Vector Machine (SVM)*, dapat melakukan *crawling* data yang lebih banyak agar mendapatkan akurasi yang lebih tinggi dan memperluas cakupan dari sentimen analisis seperti pada waktu lolos 16 besar piala asia 2023 dan terakhir bisa menambahkan sentimen netral.

5. REFERENSI

- [1] Batu Bara, E. A., Nasution, K. A., Ginting, R. Z., & Kartini. (2022). Studi tentang *Twitter*. *Jurnal Universitas Muhammadiyah Enrekang*, 2715-2634.
- [2] C. Muhamad, S. Ramdani, A. N. Rachman, and R. Setiawan, "Comparison of the Multinomial Naive Bayes Algorithm and Decision Tree with the Application of AdaBoost in Sentiment Analysis Reviews PeduliLindungi Application," *Int. J. Inf. Syst. Technol. Akreditasi*, vol. 6, no. 158, pp. 419–430, 2022.
- [3] S. M. Permataning Tyas, B. S. Rintyarna, and W. Suharso, "The Impact of Feature Extraction to Naïve Bayes Based Sentiment Analysis on Review Dataset of Indihome Services," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 13, no. 1, pp. 1–10, 2022.
- [4] M. Rifki and I. Imelda, "Borobudur Menggunakan Multinomial Naïve Bayes Analysis of Discourse Sentiment of Borobudur Temple Ticket," *J. Inform. dan Komput.*, vol. 5, no. 2, pp. 156–163, 2022.
- [5] F. Zaini, J. W. Sari, and F. N. Hasan, "Analysis of Public Sentiment Related To the Failure of Indonesia To Host U-20 Using Multinomial Naïve Bayes Classifier," *J. Tek. Inform.*, vol. 4, no. 6, pp. 1409–1418, 2023.
- [6] M. F. HAIKAL, J. INDRA, and R. RAHMAT, "Analisis Sentimen Bakal Calon Presiden Indonesia 2024 Dengan Algoritma Naïve Bayes," *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 13, no. 1, p. 43, 2024.