



ANALISIS SENTIMEN MASYARAKAT TERHADAP PON XII ACEH-SUMUT MENGGUNAKAN ALGORITMA NAÏVE BAYES

Maulian Saputra¹, Muhammad Iqbal²

¹maulian.saputra@gmail.com, ²muhammadiqbal@dosen.pancabudi.ac.id

^{1,2}Magister Teknik Informasi, Universitas Pembangunan Panca Budi

Abstrak

Penelitian ini menganalisis sentimen masyarakat terhadap Pekan Olahraga Nasional (PON) XII di Aceh-Sumatera Utara menggunakan algoritma *Multinomial Naive Bayes*. Data diperoleh melalui *crawling* media sosial X, kemudian diolah melalui tahap *pre-processing* yang meliputi pembersihan data, *case folding*, normalisasi kata, tokenisasi, penghapusan *stopword*, dan *stemming*. Sentimen pada data dilabeli menggunakan pendekatan *lexicon-based* untuk mengklasifikasikan teks menjadi positif dan negatif. Model klasifikasi dibangun menggunakan algoritma *Multinomial Naive Bayes*, dengan evaluasi kinerja melalui *confusion matrix*, akurasi, dan *classification report*. Hasil penelitian menunjukkan bahwa model mencapai akurasi 79,77% pada data uji, memberikan gambaran tentang pandangan masyarakat terhadap PON XII dan membuktikan efektivitas algoritma ini dalam analisis sentimen berbasis teks.

Kata kunci: Analisis Sentimen, Naïve Bayes, PON XII Aceh-Sumut

Abstract

This study analyzes public sentiment towards the 12th National Sports Week (PON XII) in Aceh-North Sumatra using the Multinomial Naive Bayes algorithm. Data was collected through crawling on the social media platform X and processed through pre-processing stages, including data cleaning, case folding, word normalization, tokenization, stopword removal, and stemming. Sentiment labeling was conducted using a lexicon-based approach to classify text into positive and negative categories. A sentiment classification model was developed using the Multinomial Naive Bayes algorithm, and its performance was evaluated using a confusion matrix, accuracy, and classification report. The results show that the model achieved an accuracy of 79.77% on the test data, providing insights into public perceptions of PON XII and demonstrating the effectiveness of this algorithm in text-based sentiment analysis.

Keywords: Sentiment analysis, Naïve Bayes, PON XII Aceh-Sumut

1. Pendahuluan

Pekan Olahraga Nasional (PON) merupakan acara olahraga terbesar di Indonesia yang diadakan setiap empat tahun sekali [1], dengan tujuan untuk mempererat persatuan antar Provinsi serta sebagai wadah untuk menemukan bakat-bakat olahraga terbaik dari seluruh penjuru Tanah Air. Tujuan olahraga nasional adalah untuk memelihara dan meningkatkan kesehatan dan kebugaran, prestasi, mutu hidup manusia, nilai moral dan akhlak mulia, sportivitas, mendisiplinkan, memantapkan dan memupuk persatuan dan kesatuan bangsa, memperkuat ketahanan nasional serta mengangkat harkat dan martabat bangsa [2]. PON tidak hanya berfungsi sebagai kompetisi olahraga, tetapi juga menjadi acara yang menyita perhatian masyarakat luas, baik dari sisi prestasi yang ditampilkan oleh atlet, maupun dari segi dampak sosial, ekonomi, dan politik yang ditimbulkan di daerah penyelenggara. PON XII yang diselenggarakan di Aceh-Sumut pada tahun 2024 menjadi salah satu momen penting dalam kalender olahraga nasional, karena selain menjadi tempat berkumpulnya atlet-atlet terbaik, ajang ini juga membawa perhatian besar dari publik, baik dalam hal pelaksanaan acara, hingga dampaknya terhadap masyarakat dan daerah. Namun, seiring dengan penyelenggaraan yang kompleks dan hasil kompetisi yang beragam, reaksi masyarakat terhadap acara ini turut beragam, mulai dari apresiasi hingga kritik yang disampaikan baik secara langsung maupun melalui platform digital.

Sebanyak 2.500 tweet dari pengguna media sosial X sepanjang tahun 2024, selama acara Pekan Olahraga Nasional (PON) berlangsung, mencurahkan pendapat mereka mengenai acara tersebut [3]. Media

sosial saat ini telah menjadi arena utama bagi publik untuk mengekspresikan opini mereka terhadap berbagai isu, termasuk peristiwa besar seperti PON. Platform seperti X, Facebook, Instagram, dan TikTok menjadi tempat dimana masyarakat dapat dengan bebas mengungkapkan pandangan mereka, baik itu berupa pujian terhadap prestasi atlet atau bahkan kekecewaan terhadap kekurangan dalam penyelenggaraan. Fenomena ini menimbulkan peluang untuk melakukan analisis terhadap sentimen publik, yaitu untuk mengukur apakah reaksi masyarakat cenderung positif atau negatif terhadap suatu peristiwa. Dalam hal ini, analisis sentimen dapat memberikan gambaran mengenai tingkat kepuasan atau ketidakpuasan masyarakat terhadap penyelenggaraan PON XII Aceh-Sumut, serta membantu memahami faktor-faktor yang mempengaruhi persepsi tersebut.

Media sosial merupakan media online yang digunakan sebagai kesempatan untuk berinteraksi dan mempresentasikan diri, baik secara langsung maupun tidak langsung dengan masyarakat yang mendorong nilai dari *user-generated* dan sudut pandang dengan orang lain [4]. Sentimen yang diungkapkan oleh publik di media sosial bisa memberikan wawasan yang sangat berharga, baik bagi pihak penyelenggara, pemerintah, maupun masyarakat itu sendiri. Tugas dasar dalam analisis sentimen adalah untuk mengklasifikasikan polaritas teks yang ada dalam dokumen, kalimat, atau pendapat. Polaritas berarti apakah teks yang ada dalam dokumen, kalimat, atau pendapat memiliki aspek positif atau negatif [5]. Misalnya, mengetahui sentimen negatif yang muncul terkait dengan logistik, fasilitas, atau bahkan isu-isu yang lebih sensitif seperti politik lokal, dapat menjadi bahan evaluasi untuk perbaikan pada acara-acara berikutnya. Sebaliknya, sentimen positif yang dominan bisa menjadi indikator keberhasilan penyelenggaraan PON dan sekaligus menjadi cerminan atas kontribusi positif yang diberikan oleh acara ini terhadap pembangunan daerah. Tujuan dari analisis sentimen tersebut ialah, sangat mungkin untuk menyelesaikan masalah yang muncul dari opini publik [6]. Oleh karena itu, penting untuk melakukan analisis mendalam terkait bagaimana masyarakat merespons ajang ini dan apa saja faktor yang membentuk sentimen mereka.

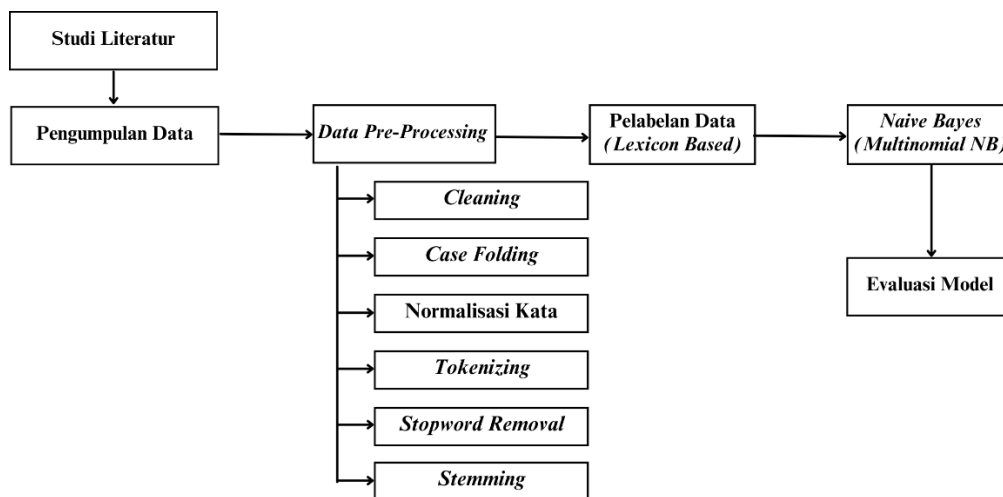
Untuk itu, penelitian ini bertujuan untuk menganalisis sentimen publik terhadap PON XII Aceh-Sumut dengan memanfaatkan data yang diperoleh dari platform digital, khususnya media sosial X. Sentimen publik ini akan dianalisis menggunakan teknik *machine learning*. *Machine learning* sendiri merupakan suatu bidang yang memberikan komputer kemampuan untuk belajar tanpa perlu diprogram secara eksplisit [7]. Selanjutnya, dengan algoritma *Naive Bayes* sebagai metode utama dalam klasifikasi teks. *Naive Bayes* merupakan algoritma klasifikasi yang berbasis probabilitas dalam statistik yang ditemukan oleh Thomas Bayes [8]. Algoritma *Naive Bayes* menggunakan metode *probabilistic* sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset [9]. Algoritma *Naive Bayes* dipilih karena kemampuannya yang efektif dalam mengklasifikasikan teks menjadi kategori-kategori tertentu, seperti positif dan negatif, yang sangat relevan dalam konteks analisis sentimen. Proses penelitian dimulai dengan pengumpulan data berupa komentar, *tweet*, atau postingan yang berkaitan dengan PON XII Aceh-Sumut. Data ini akan diproses terlebih dahulu dengan teknik *text mining* untuk membersihkan dan menyiapkan data untuk analisis lebih lanjut. Secara umum *Text mining* terdiri dari tiga langkah yaitu: *text preprocessing*, operasi penggalian teks, *postprocessing* [10]. Selanjutnya, algoritma *Naive Bayes* akan diterapkan untuk mengklasifikasikan sentimen berdasarkan teks yang terkumpul.

Metode penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam memahami bagaimana persepsi masyarakat terhadap penyelenggaraan PON, serta memberi wawasan tentang potensi keuntungan dan tantangan yang muncul dari acara besar semacam ini. Dengan analisis sentimen berbasis teks, diharapkan pihak penyelenggara dan pemangku kepentingan lainnya dapat memperoleh informasi yang lebih mendalam tentang apa yang dirasakan oleh masyarakat, serta dapat mengambil langkah-langkah perbaikan yang lebih tepat sasaran di masa mendatang. Selain itu, penelitian ini juga berpotensi untuk memperkaya khazanah riset dalam bidang analisis sentimen, terutama dalam konteks acara besar dan isu-isu sosial lainnya di Indonesia. Secara keseluruhan, penelitian ini diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai dinamika sentimen publik terhadap PON XII Aceh-Sumut dan mengarah pada pemahaman yang lebih baik tentang hubungan antara persepsi masyarakat dan keberhasilan sebuah acara olahraga nasional.

2. Metode

Pada penelitian ini, data dari media sosial X dikumpulkan menggunakan *library Python*, *tweet-harvest*. Data mentah yang diperoleh kemudian dibersihkan dan diproses melalui tahap *pra-processing*.

Sentimen dari setiap *tweet* kemudian dilabeli secara otomatis menggunakan pendekatan *lexicon-based*. Model klasifikasi *Multinomial Naive Bayes* dilatih menggunakan data yang telah diberi label. Kinerja model kemudian diuji pada data yang terpisah. Alur lengkap metodologi penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Alur Metodologi Penelitian

2.1. Studi Literatur

Melalui studi literatur, penelitian ini berusaha memetakan perkembangan penelitian terkait analisis sentimen, dengan fokus pada penerapan algoritma *Naive Bayes* dalam konteks olahraga. Kajian ini mencakup berbagai literatur, mulai dari konsep dasar analisis sentimen hingga studi kasus yang relevan dengan topik penelitian. Hasil kajian ini diharapkan dapat memberikan landasan yang kuat untuk merancang penelitian lebih lanjut.

2.2. Pengumpulan Data

Sebagai sumber data utama, penelitian ini memanfaatkan data sentimen yang diperoleh melalui proses *crawling* pada platform media sosial X. Data yang dikumpulkan berupa ungkapan sentimen masyarakat Indonesia terkait penyelenggaraan Pekan Olahraga Nasional (PON) XII di Aceh dan Sumatera Utara. Data ini kemudian dianalisis untuk memahami persepsi publik terhadap penyelenggaraan acara olahraga berskala nasional tersebut.

2.3. Data Pre-Processing

Praproses data merupakan tahap krusial dalam analisis sentimen. Tahap *Preprocessing* adalah komponen penting dalam persiapan data [11]. Tujuan utama dari tahap ini adalah untuk membersihkan dan mentransformasi data teks mentah menjadi bentuk yang lebih terstruktur dan relevan untuk proses analisis. Langkah-langkah praproses yang umum dilakukan meliputi:

- 1). *Cleaning*, tahap pembersihan data bertujuan untuk menghilangkan elemen-elemen yang tidak relevan dalam teks, seperti karakter khusus, tanda baca, angka, dan simbol. Proses ini penting untuk memastikan bahwa hanya kata-kata yang bermakna yang akan digunakan dalam analisis sentimen.
- 2). *Case folding*, adalah langkah persiapan data yang berupaya mengubah semua huruf kapital dalam kalimat menjadi karakter huruf kecil (*lowercase*) [12]. Tujuannya adalah untuk menyatukan bentuk kata yang berbeda namun memiliki makna yang sama, seperti "PON" dan "pon". Dengan demikian, variasi penulisan huruf besar dan kecil tidak akan memengaruhi hasil analisis sentimen.
- 3). Normalisasi kata, salah satu langkah penting dalam praproses teks adalah normalisasi kata. Proses ini bertujuan untuk mengubah kata-kata yang memiliki ejaan tidak baku atau singkatan

menjadi bentuk baku yang sesuai dengan kamus. Misalnya, singkatan "blm" akan diubah menjadi kata lengkap "belum" agar lebih mudah diproses oleh algoritma dan menghasilkan hasil analisis yang lebih akurat.

- 4). *Tokenizing*, merupakan proses awal dalam analisis teks yang bertujuan untuk membagi teks menjadi unit-unit terkecil yang disebut token atau kata. Dengan memecah teks menjadi kata-kata individu, kita dapat menganalisis kata per kata dan mengidentifikasi pola atau tren yang terkandung di dalamnya.
- 5). *Stopword removal*, Penghapusan kata-kata kosong (*stopword*) adalah proses menghilangkan kata-kata umum yang sering muncul dalam bahasa, seperti 'yang', 'dan', 'atau'. Kata-kata ini dianggap tidak membawa makna signifikan dalam analisis sentimen. Dengan menghapus *stopword*, model dapat lebih fokus pada kata-kata bermakna yang mampu menunjukkan sentimen positif atau negatif.
- 6). *Stemming*, proses pengubahan kata menjadi bentuk akarnya yang paling sederhana. Misalnya, kata 'membeli' akan diubah menjadi 'beli'. Tujuan dari proses ini adalah untuk mengurangi variasi bentuk kata yang memiliki makna dasar yang sama, sehingga algoritma dapat lebih mudah mengidentifikasi kata-kata yang terkait.

2.4. Pelabelan Data

Pelabelan sentimen *lexicon based* adalah metode yang digunakan untuk mengklasifikasikan teks menjadi sentimen positif atau negatif dengan cara membandingkan kata-kata dalam teks dengan kamus sentimen (*lexicon*). Setiap kata dalam leksikon telah diberi label sentimen tertentu. Dengan demikian, sentimen suatu teks dapat ditentukan berdasarkan jumlah kata positif dan negatif yang terkandung di dalamnya.

2.5. Naïve Bayes Multinomial

Data yang telah melalui proses pelabelan, dilakukan pembagian data menjadi 2, yaitu data latih dan data uji. Untuk masing-masing pembagian data data latih dan uji ini penulis membagi kembali menjadi 2 yaitu : - Pertama, data latih berjumlah lebih kecil dari data uji - Kedua, data latih berjumlah lebih banyak dari data uji [13].

Setelah data ulasan pengguna dibagi menjadi data latih dan data uji, algoritma *Naive Bayes* diterapkan pada data latih untuk membangun model klasifikasi sentimen. Model ini kemudian digunakan untuk memprediksi sentimen dari data uji, mengklasifikasikan setiap ulasan ke dalam kategori positif atau negatif.

2.6. Evaluasi Model

Kinerja model *Naive Bayes* dalam mengklasifikasikan sentimen masyarakat terhadap PON XII Aceh-SUMUT dievaluasi menggunakan tiga metrik utama: *confussion matrix*, *accuracy*, dan *clasificassion report*. Metrik-metrik ini memberikan gambaran yang komprehensif tentang kemampuan model dalam mengklasifikasikan sentimen dengan benar, termasuk kemampuannya dalam mengidentifikasi sentimen positif, negatif, dan kemungkinan kesalahan klasifikasi.

3. Hasil dan Pembahasan

Bagian ini memuat analisis hasil penelitian, termasuk evaluasi pelaksanaan, pengujian data, serta pembahasan topik yang relevan. Untuk mendukung penjelasan, disertakan pula visualisasi data berupa grafik, tabel, dan alat bantu visual lainnya guna memperjelas hasil yang diperoleh.

3.1. Pengumpulan data

Data dalam penelitian ini dikumpulkan menggunakan teknik *crawling* dari media sosial X. Proses pengambilan data dilakukan dengan memanfaatkan *library Tweet Harvest* pada *Python*, yang diimplementasikan melalui *Google Colab*. *Crawling* dilakukan dalam rentang waktu 10 hingga 14 September 2024, bertepatan dengan jadwal pelaksanaan PON XII Aceh-Sumut. Kata kunci yang digunakan adalah "pon aceh sumut", menghasilkan 2.500 *tweet* dalam pencarian awal, yang setelah proses filtrasi,

menyisakan 2.055 *tweet* untuk dijadikan dataset penelitian. Hasil dari proses ini disajikan pada Tabel 1 berikut.

Tabel 1. Hasil *Crawling Data*

<i>Username</i>	<i>Full Text</i>
MrNodava	Dukung PON XXI Aceh Sumut https://t.co/UVKzM667ME
david_irianto4	PON XXI 2024 Aceh Sumut emang fokus ke kesejahteraan atlet bikin mereka lebih termotivasi...
baktiadi_nababa	PON XXI 2024 Aceh Sumut ngebuat orang Aceh lebih merasa terhubung dan cinta tanah air...

Setelah data berhasil dikumpulkan dan melalui proses *pre-processing*, langkah berikutnya adalah menghapus data yang bersifat duplikat. Dari total 2.500 data hasil *crwaling*, tersisa 2.055 data yang unik setelah proses penghapusan duplikasi selesai dilakukan.

3.2. Data Pre-Processing

Tahap awal *pre-processing* dimulai dengan membersihkan data untuk menghapus elemen-elemen seperti *Uniform Resource Locator* (URL), *Hypertext Markup Language* (HTML), emoji, simbol, dan angka. Setelah proses pembersihan, dilakukan *Case Folding*, yaitu mengonversi semua huruf kapital menjadi huruf kecil guna menyamakan format teks. Hasil dari proses *Case Folding* disajikan pada Tabel 2.

Tabel 2. Hasil *Case Folding*

<i>Hasil Cleaning</i>	<i>Hasil Case Folding</i>
Dukung PON XXI Aceh Sumut	dukung pon xxi aceh sumut
PON XXI 2024 Aceh Sumut emang fokus...	pon xxi aceh sumut emang fokus ke kesejahteraa...
PON XXI 2024 Aceh Sumut ngebuat orang...	pon xxi aceh sumut ngebuat orang aceh lebih me...

Tahap selanjutnya adalah normalisasi kata, yaitu proses mengonversi kata-kata dalam ulasan menjadi bentuk baku sesuai dengan kaidah bahasa yang berlaku. Hasil dari proses normalisasi ini dapat dilihat pada Tabel 3.

Tabel 3. Hasil Normalisasi Kata

<i>Hasil Case Folding</i>	<i>Hasil Normalisasi Kata</i>
pon xxi aceh sumut bener bener jadi momen...	pon xxi aceh sumut benar benar jadi momen yang...
keren ya bisa liat talenta talenta muda di pon...	keren ya bisa lihat talenta talenta muda di pon...
gimana menurut kalian persiapan pon di ace...	bagaimana menurut kalian persiapan pon di aceh...

Setelah normalisasi kata selesai dilakukan, langkah berikutnya adalah *tokenizing*, yaitu memecah ulasan menjadi token atau kata-kata individu. Proses ini dilakukan menggunakan *library Natural Language Toolkit* (NLTK) untuk memastikan akurasi pemisahan. Hasil dari tahap *tokenizing* disajikan pada Tabel 4.

Tabel 4. Hasil *Tokenizing*

<i>Hasil Normalisasi Kata</i>	<i>Hasil Tokenizing</i>
dukung pon xxi aceh sumut	[dukung, pon, xxi, aceh, sumut]
pon xxi aceh sumut memang fokus ke kesej...	[pon, xxi, aceh, sumut, memang, fokus, ke, kes...
aceh dan sumut siap menyambut setiap peser...	[aceh, dan, sumut, siap, menyambut, setiap, pe..

Tahap berikutnya adalah penghapusan *stopwords*, yaitu menghilangkan kata-kata yang dianggap tidak memiliki makna signifikan dalam analisis, seperti "dan", "yang", atau "di". Proses ini dilakukan dengan bantuan *library NLTK* untuk memastikan penghapusan dilakukan secara efektif. Hasil dari penghapusan *stopwords* dapat dilihat pada Tabel 5.

Tabel 5. Hasil *Stopword Removal*

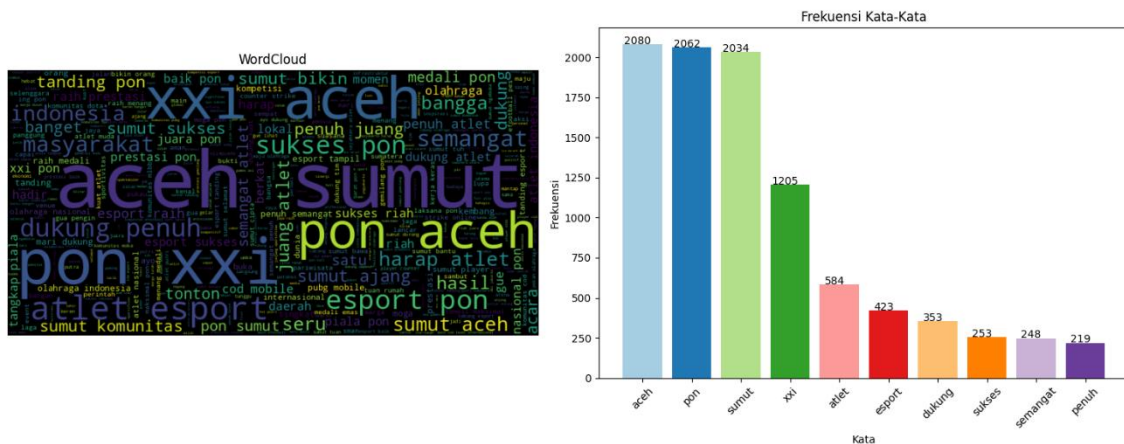
<i>Hasil Tokenizing</i>	<i>Hasil Stopword Removal</i>
[aceh, dan, sumut, siap, menyambut, setiap,p...	[aceh, sumut, menyambut, peserta, semangat, ke...
[atlet, atlet, yang, berpartisipasi, di, pon, ...	[atlet, atlet, berpartisipasi, pon, xxi, aceh,...
[pembangunan, di, aceh, dan, sumut, kelihat...	[pembangunan, aceh, sumut, maju, pon, xxi]

Tahap akhir dari *pre-processing* data adalah *stemming*, yaitu proses mengonversi kata berimbuhan menjadi bentuk dasar. Proses ini dilakukan dengan memanfaatkan *library Sastrawi* dan *Swifter* untuk memastikan efisiensi dan akurasi. Hasil dari proses *stemming* disajikan pada Tabel 6.

Tabel 6. Hasil *Stemming*

Hasil <i>Stopword Removal</i>	Hasil <i>Stemming</i>
[aceh, sumut, menyambut, peserta, semangat,...	aceh sumut sambut serta semangat sama pon xxi
[atlet, atlet, berpartisipasi, pon, xxi, aceh,...	atlet atlet partisipasi pon xxi aceh sumut ala...
[hujan, ringan, menambah, tantangan, caban...	hujan ringan tambah tantang cabang sepeda bmx ...

Setelah proses *stemming* selesai, dataset telah melalui seluruh tahap *pre-processing*, sehingga data *tweet* kini sudah dibersihkan dan siap untuk dianalisis lebih lanjut. Hasil dari *pre-processing* ini dapat dilihat dalam bentuk *wordcloud* dan frekuensi kata yang paling sering muncul ditampilkan pada Gambar 2.



Gambar 2. *Wordcloud* dan Frekuensi Kata

3.3. Pelabelan Data

Langkah berikutnya adalah menetapkan lapisan *output* untuk mengklasifikasikan sentimen menjadi positif dan negatif menggunakan metode berbasis kamus (*lexicon-based*). Kode program dari penerapan metode ini ditampilkan pada Gambar 3.

```

import pandas as pd

# Fungsi untuk menentukan sentimen
def determine_sentiment(text):
    positive_count = sum(1 for word in text.split() if word in positive_lexicon)
    negative_count = sum(1 for word in text.split() if word in negative_lexicon)
    if positive_count > negative_count:
        return "Positif"
    elif positive_count < negative_count:
        return "Negatif"

# Baca kamus leksikon positif dan negatif
positive_lexicon = set(pd.read_csv("positive.tsv", sep="\t", header=None)[0])
negative_lexicon = set(pd.read_csv("negative.tsv", sep="\t", header=None)[0])

def replace_none_sentiment(sentiments):
    replace_flag = "Positif"
    for i in range(len(sentiments)):
        if sentiments[i] is None:
            sentiments[i] = replace_flag
            replace_flag = "Negatif" if replace_flag == "Positif" else "Positif"
    return sentiments
    
```

Gambar 3. Kode Program Pelabelan Data Sentimen

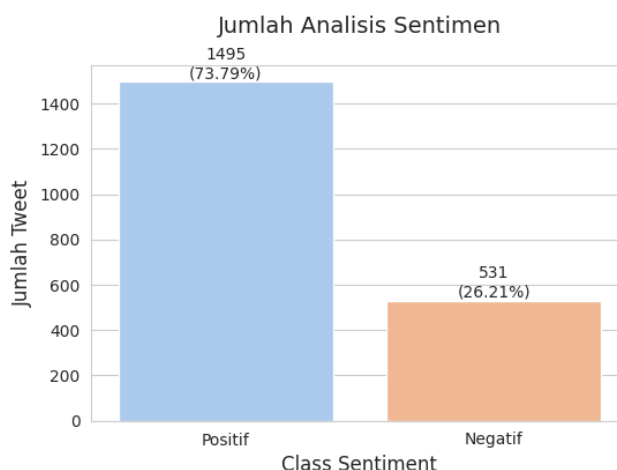
Metode *lexicon-based* dirancang untuk melakukan analisis sentimen secara efisien dan akurat pada teks, khususnya dalam konteks data berbasis media sosial. Metode ini memanfaatkan kamus kata yang telah dikategorikan ke dalam sentimen positif, negatif, atau netral untuk menilai dan mengidentifikasi pola

sentimen dalam teks. Pendekatan ini sangat relevan untuk digunakan pada data dari media sosial karena mampu mengolah data teks dengan volume besar dan mendeteksi sentimen yang terkandung di dalamnya secara cepat. Selain itu, metode ini tidak hanya mengklasifikasikan sentimen menjadi positif atau negatif tetapi juga mampu mengukur intensitas emosi dalam teks, sehingga memberikan wawasan lebih mendalam mengenai opini pengguna. Hasil pelabelan sentimen dapat dilihat pada Tabel 7.

Tabel 7. Pelabelan Sentimen

Hasil Pre-processing	Sentimen
bangun aceh sumut maju pon xxi	Positif
gua gabung giat volunteering pon xxi aceh sumut	Negatif
moga pon aceh sumut hasil prestasi	Positif

Jumlah total setiap kelas sentimen, baik positif maupun negatif, dalam dataset divisualisasikan untuk memberikan gambaran yang lebih jelas mengenai distribusi data. Visualisasi ini dilakukan menggunakan diagram batang yang dibuat dengan bantuan *library Seaborn*. Diagram batang dipilih karena mampu menyajikan perbandingan antar kelas sentimen secara sederhana dan efektif. Hasil visualisasi ini tidak hanya membantu memahami proporsi sentimen secara keseluruhan tetapi juga menjadi dasar dalam analisis lebih lanjut. Visualisasi kelas sentimen tersebut ditampilkan pada Gambar 4.

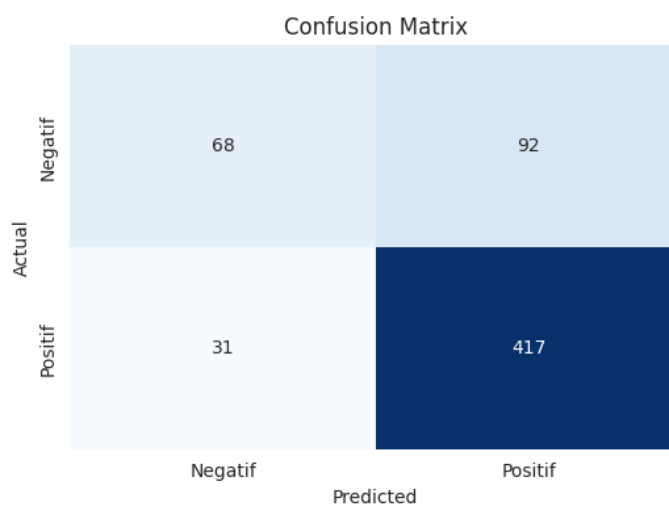


Gambar 4. Class Sentiment

3.4. Pengklasifikasian Menggunakan *Multinomial Naïve Bayes*

Tahap berikutnya dalam proses analisis adalah mengklasifikasikan ulasan pengguna dengan menerapkan algoritma *Multinomial Naive Bayes*, yang diimplementasikan menggunakan modul *Scikit-Learn*. Algoritma ini dipilih karena keandalannya dalam memproses data berbentuk teks dan memberikan hasil klasifikasi yang efisien. Dataset dibagi menjadi dua bagian, yaitu data latih dan data uji, dengan proporsi 70:30. Pembagian ini menghasilkan 1.418 data untuk pelatihan dan 608 data untuk pengujian, memastikan model memiliki cukup data untuk belajar dan diuji.

Proses klasifikasi menggunakan *kernel linier*, yang dirancang untuk menangani hubungan linier antar fitur dalam data. Teknik ini memungkinkan algoritma untuk mempelajari pola sentimen yang relevan dari dataset secara lebih efektif. Evaluasi kinerja model dilakukan dengan menyusun *Confusion Matrix*, yang memberikan gambaran mengenai jumlah prediksi benar dan salah untuk setiap kelas sentimen. Hasil visualisasi *Confusion Matrix* dari algoritma *Multinomial Naive Bayes* ditampilkan pada Gambar 5, memberikan pandangan lebih mendalam terhadap kemampuan model dalam mengklasifikasikan sentimen dengan akurasi yang tinggi.



Gambar 5. Confussion Matrix

Hasil *classification report* yang dihasilkan melalui penerapan algoritma *Multinomial Naive Bayes* memberikan wawasan rinci mengenai kinerja model dalam mengklasifikasikan ulasan pengguna. Laporan ini menyajikan berbagai metrik evaluasi penting, seperti *precision*, *recall*, dan *f1-score*, yang masing-masing menggambarkan kemampuan model dalam mengidentifikasi kelas sentimen secara akurat, cakupan prediksi yang benar, serta keseimbangan antara kedua metrik tersebut.

Selain itu, *classification report* ini juga mencakup metrik akurasi keseluruhan, memberikan gambaran umum tentang seberapa baik model bekerja pada dataset yang diberikan. Analisis lebih lanjut dari metrik evaluasi tersebut dapat membantu mengidentifikasi kekuatan dan kelemahan model dalam memprediksi sentimen positif maupun negatif.

Tabel 8 menyajikan hasil lengkap dari *classification report*, memberikan gambaran yang mendalam mengenai performa model.

Tabel 8. *Classification Report*

Akurasi Model	: 0.797697368421			
	: 79,77%			
<i>Classification Report</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Negatif	0.69	0.42	0.53	160
Positif	0.82	0.93	0.87	448
Akurasi			0.80	608
<i>Macro Avg</i>	0.75	0.68	0.70	608
<i>Weighted Avg</i>	0.78	0.80	0.78	608

Tabel 8 di atas menyajikan hasil evaluasi model *Multinomial Naive Bayes* yang diterapkan untuk analisis sentimen. Model ini berhasil mencapai tingkat akurasi sebesar 79,77% (0.7977), yang menunjukkan kinerjanya cukup baik dalam memprediksi sentimen pada dataset yang digunakan.

Laporan klasifikasi mengidentifikasi dua kategori sentimen, yaitu "Negatif" dan "Positif". Pada kategori "Negatif", model memiliki skor *precision* sebesar 0,69, *recall* 0,42, dan *f1-score* 0,53, dengan dukungan data sebanyak 160 sampel. Sementara itu, untuk kategori "Positif", model menunjukkan kinerja yang lebih tinggi dengan *precision* sebesar 0,82, *recall* 0,93, dan *f1-score* 0,87, berdasarkan total 448 sampel.

Hasil ini mencerminkan bahwa model lebih efektif dalam mengenali sentimen positif dibandingkan negatif. Perbedaan ini dapat disebabkan oleh distribusi data yang tidak seimbang atau karakteristik teks pada kategori tertentu, yang memberikan ruang untuk eksplorasi dan perbaikan model di masa mendatang.

4. Kesimpulan dan Saran

4.1. Kesimpulan

Hasil penelitian ini menunjukkan bahwa algoritma *Multinomial Naive Bayes* dapat diterapkan secara efektif untuk analisis sentimen pada ulasan masyarakat terkait PON XII Aceh-Sumut, dengan tingkat akurasi mencapai 79,77%. Evaluasi kinerja model melalui laporan klasifikasi mengindikasikan bahwa model lebih unggul dalam mengidentifikasi sentimen positif dibandingkan sentimen negatif. Pada kategori positif, model mencapai skor *precision* sebesar 0,82, *recall* 0,93, dan *f1-score* 0,87. Sementara itu, pada kategori negatif, *precision* tercatat sebesar 0,69, *recall* 0,42, dan *f1-score* 0,53. Ketimpangan kinerja antara kategori positif dan negatif menunjukkan bahwa model mengalami kesulitan dalam mengidentifikasi sentimen negatif, yang dipengaruhi oleh distribusi dataset yang tidak seimbang, dengan jumlah sampel negatif yang lebih kecil dibandingkan kategori positif. Selain itu, kompleksitas ekspresi sentimen negatif dalam ulasan juga turut memengaruhi hasil ini. Meskipun demikian, secara keseluruhan, model ini dapat memberikan gambaran yang cukup akurat tentang persepsi masyarakat dan dapat dimanfaatkan dalam analisis sentimen berbasis teks.

4.2. Saran

Berdasarkan temuan penelitian ini, disarankan agar dilakukan pengumpulan data yang lebih seimbang antara kategori positif dan negatif untuk meningkatkan kinerja model. Teknik data *augmentation* juga dapat dipertimbangkan untuk memperkaya variasi data pada kategori yang kurang terwakili, sehingga model dapat lebih optimal dalam mempelajari pola-pola sentimen negatif. Selain itu, eksperimen dengan algoritma lain seperti *Support Vector Machine* (SVM) atau *Random Forest* dapat dilakukan untuk memperoleh hasil yang lebih baik, terutama dalam menangani masalah ketidakseimbangan kelas. Pengaturan parameter lebih lanjut dan penerapan teknik *ensemble learning* juga dapat meningkatkan akurasi dan kinerja model secara keseluruhan. Selain itu, perhatian lebih pada *pre-processing* data, seperti penghapusan *stopwords*, *tokenizing*, dan *stemming*, sangat penting untuk memastikan kualitas data input yang tinggi dalam analisis sentimen.

Daftar Pustaka

- [1] U. Jember, "Analisis Sentimen Opini Publik Terhadap Pekan Olahraga Nasional Pada Instagram Menggunakan Metode Naïve Bayes Classifier Priza Pandunata, Caesarina Kurnia Ananta, Yanuar Nurdiansyah," 2022.
- [2] Muhammad Suhery, Gema Ramadhan, and Abdul Halim Hasugian, "Analisis Sentimen Masyarakat Kota Medan Terhadap Persiapan PON XXI Sumut-Aceh 2024 Menggunakan Metode Naive Bayes," *Polygon : Jurnal Ilmu Komputer dan Ilmu Pengetahuan Alam*, vol. 2, no. 5, pp. 10–18, Aug. 2024, doi: 10.62383/polygon.v2i5.207.
- [3] K. Jonathan Sitorus and A. Muliawati, "Klasifikasi dan Analisis Sentimen Pada Data Twitter Menggunakan Algoritma Naïve Bayes. (Studi Kasus: Pekan Olahraga Nasional XX 2021)," 2022. [Online]. Available: <https://t.co/dsH2pZdXvd>
- [4] I. Olive, D. Putra, K. Rega Prilianti, P. Lucky, and T. Irawan, "Implementasi Text Mining Untuk Analisis Opini Masyarakat Terhadap Kinerja Layanan Transportasi Online Dengan Analisis Faktor," vol. 8, no. 2, 2020.
- [5] R. Wahyudi *et al.*, "Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine," *Jurnal Informatika*, vol. 8, no. 2, 2021, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [6] M. S. Arrafiq and R. Kurniawan, "Analisis Sentimen Pengguna Terhadap Layanan Aplikasi Seabank Indonesia di Instagram Menggunakan Metode Support Vector Machine," *Journal of Information System Research*, vol. 5, no. 4, 2024, doi: 10.47065/josh.v5i4.5620.
- [7] F. Abdusyukur, "Penerapan Algoritma Support Vector Machine (Svm) Untuk Klasifikasi Pencemaran Nama Baik Di Media Sosial Twitter," *Komputa: Jurnal Ilmiah Komputer dan Informatika*, vol. 12, no. 1, 2023.
- [8] J. Homepage and A. Ardiansyah, "SATIN-Sains dan Teknologi Informasi Analisis Sentimen Aplikasi Quizizz pada Google Play Store Menggunakan Naïve Bayes", doi: 10.33372/stn.v9i2.1056.

- [9] R. N. Mauliza and Y. R. Sipayung, “Penerapan Text Mining Dalam Menganalisis Pendapat Masyarakat Terhadap Pemilu 2024 Pada Media Sosial X Menggunakan Metode Naive Bayes,” *Technomedia Journal*, vol. 9, no. 1, pp. 1–16, Feb. 2024, doi: 10.33050/tmj.v9i1.2212.
- [10] A. Firdaus and W. I. Firdaus, “Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan),” 2021.
- [11] N. Wijaya and E. Setiawan Panjaitan, “Analisis Sentimen Ulasan Aplikasi Instagram di Google Play Store: Pendekatan Multinomial Naive Bayes dan Berbasis Leksikon,” *Technology and Science (BITS)*, vol. 6, no. 2, 2024, doi: 10.47065/bits.v6i2.5615.
- [12] S. I. Nurhafida and F. Sembiring, “Analisis Sentimen Aplikasi Novel Online Di Google Play Store Menggunakan Algoritma Support Vector Machine (SVM),” 2022.
- [13] R. Apriani *et al.*, “Analisis Sentimen Dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia,” 2019.