

Perbandingan Algoritma Naïve Bayes dan TextBlob Untuk Mendapatkan Analisis Sentimen Masyarakat Pada Sosial Media

Giesta Rahguna Putri^{1*}, Muhammad Akbar Maulana², Samsul Bahri³

^{1,2,3}Program Studi Informatika, Institut Teknologi Tangerang Selatan, Tangerang Selatan
Email: ^{1*}giestarp12@gmail.com, ²makbar513@gmail.com, ³sambahrie@gmail.com

(Naskah masuk: 4 Mei 2024, direvisi: 24 Mei 2024, 9 Jun 2024, diterima: 10 Jun 2024)

Abstrak

Media sosial *Twitter* adalah platform yang populer di Indonesia untuk berkomunikasi dan mendapatkan informasi dengan cepat. Hal ini memungkinkan masyarakat dengan mudah mengungkapkan opini dan sentimen mereka. Penelitian ini berfokus pada perbandingan algoritma *TextBlob* dan *Naïve Bayes* dalam menganalisis sentimen masyarakat. Temuan menunjukkan bahwa *TextBlob* mengklasifikasikan sebagian besar *tweet* sebagai positif, sementara *Naïve Bayes* menunjukkan kecenderungan yang serupa dengan akurasi sebesar 78,18%. Dari analisis *TextBlob*, sekitar 50,98% komentar menunjukkan sentimen positif, 16,01% negatif, dan 33,33% netral. Dengan menggunakan kedua algoritma ini, penelitian berhasil mengidentifikasi sentimen masyarakat dengan akurasi yang baik, menunjukkan distribusi yang jelas antara sentimen positif, netral, dan negatif.

Kata Kunci: Analisis Sentimen, *Twitter*, *TextBlob*, *Naïve Bayes*.

Comparison of Naïve Bayes and TextBlob Algorithms for Analyzing Public Sentiment on Social Media

Abstract

Social media platform *Twitter* is popular in Indonesia for quick communication and information retrieval. It allows people to easily express their opinions and sentiments. This research focuses on comparing the *TextBlob* and *Naïve Bayes* algorithms in analyzing public sentiment. The findings show that *TextBlob* classifies most tweets as positive, while *Naïve Bayes* shows a similar trend with an accuracy of 78.18%. From the *TextBlob* analysis, about 50.98% of comments show positive sentiment, 16.01% negative, and 33.33% neutral. By using these two algorithms, the research successfully identified public sentiment with good accuracy, showing a clear distribution between positive, neutral, and negative sentiments.

Keywords: Analysis of Sentiment, *Twitter*, *TextBlob*, *Naïve Bayes*.

I. PENDAHULUAN

Dalam konteks politik, pemilihan presiden merupakan peristiwa yang sangat penting bagi sebuah negara. Seiring dengan perkembangan teknologi informasi dan media sosial, pendapat serta sentimen masyarakat terhadap calon presiden dapat dengan cepat tersebar dan berdampak besar pada proses pemilihan tersebut. Oleh karena itu, analisis sentimen masyarakat terhadap calon presiden menjadi suatu hal yang krusial untuk dipahami [1].

Dalam era digital, data teks dari berbagai platform media sosial seperti *Twitter*, *Facebook*, dan forum-forum *online* menjadi sumber informasi yang kaya akan pendapat dan sentimen masyarakat [2]. Melalui teknik analisis sentimen, kita dapat menggali dan memahami opini serta sikap masyarakat dengan lebih efisien dan tepat. Namun, masih terdapat kebutuhan akan pemahaman yang lebih dalam tentang cara masyarakat merespons berbagai isu atau topik tertentu di media sosial. Dalam penelitian ini opini masyarakat menjadi masalah dalam melihat sejauh mana sentimen masyarakat

terhadap isu atau topik tersebut, karena dengan banyaknya opini tersebut menjadi sulit untuk menentukan sentimen positif, negatif atau netral secara manual [3]. Dalam rangka menjawab kebutuhan tersebut, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap calon presiden Prabowo Subianto dalam pemilihan presiden 2024 dengan menggunakan data teks dari media sosial, khususnya *Twitter*.

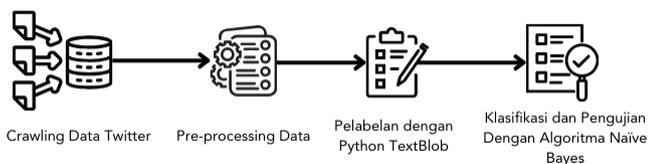
Dalam penelitian ini, akan dilakukan perbandingan performa antara dua pendekatan analisis sentimen yang umum digunakan, yaitu algoritma *TextBlob* dan *Naïve Bayes*. Algoritma *TextBlob* menggunakan pendekatan *rule-based* dan dapat melakukan analisis sentimen dalam bahasa natural dengan baik, sementara *Naïve Bayes* merupakan algoritma klasifikasi probabilistik yang sering digunakan dalam analisis sentimen [4].

Sentimen analisis atau opinion mining adalah metode pemrosesan bahasa alami (*natural language processing*) yang digunakan untuk mendeteksi *mood* atau perasaan masyarakat terhadap produk atau topik tertentu. Salah satu pendekatan yang efektif dalam analisis ini adalah menggunakan metode *machine learning*. *Python TextBlob* sering digunakan untuk pelabelan dalam analisis sentimen. Selanjutnya, performa metode ini dievaluasi melalui analisis dan pengujian akurasi menggunakan algoritma *Naïve Bayes*. Pada penelitian sebelumnya menunjukkan bahwa metode *Naïve Bayes* memiliki tingkat akurasi yang cukup baik dalam analisis sentimen. Penelitian menunjukkan bahwa setelah dilakukan pelabelan otomatis dengan *Python TextBlob*, hasil pengujian menggunakan algoritma *Naïve Bayes* menghasilkan nilai akurasi sebesar 71,98% [3].

Melalui perbandingan ini, diharapkan dapat diperoleh wawasan yang mendalam tentang bagaimana masyarakat merespons dan merasakan terhadap calon presiden, serta mengetahui kelebihan dan kekurangan dari setiap pendekatan analisis sentimen. Dengan pemahaman ini, penelitian ini diharapkan dapat memberikan sumbangan yang berarti dalam pengembangan teknik analisis sentimen yang lebih canggih dan akurat, terutama dalam konteks politik yang seringkali kompleks dan dinamis. Hal ini diharapkan dapat memperkuat pemahaman kita tentang opini [5] publik dan menjadi landasan bagi pengambilan keputusan yang lebih baik dalam proses pemilihan presiden.

II. METODOLOGI PENELITIAN

Ilustrasi langkah-langkah metodologi penelitian ini dapat dilihat dalam Gambar 1.



Gambar 1. Tahapan Penelitian

Dalam metodologi penelitian ini, tahapan dimulai dengan pengumpulan data dari platform *Twitter* melalui proses data *crawling*, yang melibatkan pengambilan *tweet-tweet* yang berkaitan dengan calon presiden 2024. Kemudian, setiap *tweet* akan dilabeli dengan sentimen yang sesuai (positif, negatif, atau netral) terhadap calon presiden. Proses *preprocessing* mencakup penggunaan metode TF-IDF untuk mengekstrak fitur-fitur penting dari setiap *tweet*, serta pembuatan visualisasi *Wordcloud* untuk menyoroti kata-kata yang paling sering muncul dalam setiap kategori sentimen. Dalam implementasi tahapan ini, alat pengembangan dan komputasi *Google Colab* digunakan untuk menjalankan proses analisis sentimen dengan lebih efisien dan efektif.

1. Auth Token

Dalam tahap awal proses *crawling* data *Twitter*, fokus utamanya adalah pada otentikasi token dari *Twitter* API [6]. Langkah ini mencakup memperoleh dan mengintegrasikan akses *token* [7] ke dalam kode program di *Google Colab*.

2. Data Crawling

Dalam penelitian ini, proses pengumpulan data dilakukan melalui metode *crawling* menggunakan *library tweet-harvest@latest* seperti terlihat pada Gambar 2. Proses dimulai dengan menentukan *file* penyimpanan dan pencarian *tweet* dengan kata kunci Prabowo. Batasan maksimum *tweet* yang diambil adalah 400, dan rentang waktu pengumpulan adalah 3 hari setelah tanggal pemilihan pada 14 Februari 2024.

```

# Crawl Data
filename = 'Prabowo.csv'
search_keyword = 'prabowo'
limit = 400

!pip --yes tweet-harvest -o "{filename}" -s "{search_keyword}" -l {limit} --token {twitter_auth_token}
  
```

Gambar 2. *Crawl Data Twitter*

3. Preprocessing Data

Setelah data berhasil diambil melalui *crawling process* kemudian disimpan dalam bentuk format CSV [8], proses selanjutnya adalah melakukan *preprocessing*. Hal ini diperlukan karena dataset yang diperoleh belum terstruktur secara optimal. Tujuan utama dari *preprocessing* data adalah untuk membersihkan dan mengatasi *noise* data sehingga analisis dapat dilakukan secara efektif. Tahapan dalam *preprocessing* data meliputi pembersihan data (*cleaning data*), penggunaan huruf kecil (*case folding*), normalisasi, penghapusan kata penghubung (*stopword*), tokenisasi, dan *stemming* [9].

3.1 Cleaning Data

Seperti terlihat pada Gambar 3, tahapan ini mencakup proses pembersihan data dari karakter atau informasi yang tidak relevan atau mengganggu, seperti tanda baca, URL, atau karakter khusus lainnya [10].

```
import re
import string

def cleaningText(text):
    text = re.sub(r'@[A-Za-z0-9]+', '', text) # remove mentions
    text = re.sub(r'[A-Za-z0-9]+', '', text) # remove hashtag
    text = re.sub(r'RT[^\s]', '', text) # remove RT
    text = re.sub(r'http\S+', '', text) # remove link
    text = re.sub(r'[0-9]+', '', text) # remove numbers

    text = text.replace('\n', ' ') # replace new line into space
    text = text.translate(str.maketrans('', '', string.punctuation)) # remove all punctuations
    text = text.strip(' ') # remove characters space from both left and right text
    return text

def emojiDel(text): #remove emoji
    regex_pattern = re.compile(pattern = "[
        u\u0001f600-\u0001f64f # emoticons
        u\u0001f300-\u0001f5ff # symbols & pictographs
        u\u0001f680-\u0001f6ff # transport & map symbols
        u\u0001f1e0-\u0001f1ff # flags (ios)
    ]+", flags = re.UNICODE)
    return regex_pattern.sub(r'',text)

# remove single char
def remove_single_char(text):
    return re.sub(r"[a-zA-Z]", "", text)
```

Gambar 3. Proses *Cleaning Data*

3.2 Case Folding

Case folding melibatkan konversi semua huruf dalam teks menjadi huruf kecil [11]. Ini dilakukan untuk memastikan konsistensi dalam teks, sehingga kata yang sama dengan huruf besar atau kecil akan dianggap sama.

3.3 Normalisasi

Normalisasi adalah proses mengubah kata-kata dengan variasi bentuk menjadi bentuk standar (Gambar 4). Misalnya, kata-kata yang memiliki makna sama tetapi ditulis dengan ejaan yang berbeda akan diubah menjadi bentuk yang konsisten. Hal ini dilakukan untuk mempermudah analisis dengan memastikan bahwa variasi ejaan tidak mengganggu konsistensi data.

```
import pandas as pd

def load_normalization_rules(file_path):
    df = pd.read_excel(file_path)
    return dict(zip(df['before'], df['after']))

def normalisasi(str_text, norm):
    for i in norm:
        str_text = str_text.replace(i, norm[i])
    return str_text

norm = load_normalization_rules('/content/drive/MyDrive/skripsi/normalisasi.xlsx')

df['full_text'] = df['full_text_clean'].apply(lambda x: normalisasi(x, norm))

df_processed.to_csv('/content/drive/MyDrive/skripsi/hasil_normalisasi Prabowo.csv', index=False)
```

Gambar 4. Proses Normalisasi

3.4 Stopword

Stopword merupakan kalimat umum yang sering muncul dalam teks dan tidak memberikan nilai tambah dalam analisis. Tahapan ini melibatkan penghapusan kata-kata tersebut dari teks untuk meningkatkan efisiensi analisis.

3.5 Tokenizing

Tokenizing proses membagi teks menjadi bagian-bagian kecil yang disebut token, seperti kata-kata atau frasa merupakan pengertian dari *tokenizing*, yang bertujuan untuk mempersiapkan teks agar dapat diproses lebih lanjut, misalnya untuk analisis teks atau untuk melatih model bahasa.

3.6 Stemming

Stemming adalah proses di mana awalan atau akhiran dari sebuah kata dihilangkan untuk mendapatkan bentuk dasarnya.

Tahap *stemming* dilakukan dengan menggunakan dukungan dari pustaka atau library dalam bahasa pemrograman *Python3* yang bernama *Sastrawi* [12]. *Stemming* membantu dalam mengurangi variasi kata yang memiliki akar kata yang sama, sehingga mempermudah analisis.

4. Pelabelan Dataset

Dalam penelitian ini, *TextBlob* digunakan untuk memberikan dua label pada data, yakni positif dan negatif. Proses pelabelan melibatkan terjemahan dari bahasa Indonesia ke bahasa Inggris [13]. Setelah tahap terjemahan, *TextBlob* melakukan pelabelan pada data *tweet* dengan menghasilkan nilai polaritas sebagai *output*. Jika nilai polaritas lebih besar dari 0, maka hasil pelabelan dianggap positif; sebaliknya, jika nilai polaritas kurang dari 0, hasil pelabelan dianggap negatif.

5. Visualisasi

Visualisasi ini berupa *wordcloud* [14], yang merupakan representasi visual dari kata-kata yang paling sering muncul dalam data. Kata-kata tersebut akan ditampilkan dengan ukuran yang berbeda-beda, di mana ukuran kata akan semakin besar seiring dengan frekuensi kemunculannya yang lebih tinggi dalam data.

6. Pembobotan TF-IDF

TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan metode pembobotan yang digunakan untuk analisis teks dalam menilai pentingnya sebuah kalimat di dalam sebuah dokumen terhadap kumpulan dokumen lainnya. Dengan TF-IDF, kalimat yang muncul secara sering dalam satu dokumen namun jarang dalam dokumen-dokumen lain akan memiliki bobot yang lebih tinggi [15]. Visualisasi ini akan menampilkan kata-kata dengan bobot TF-IDF tertinggi dalam dataset, memberikan gambaran yang jelas tentang kata-kata yang paling penting atau mendefinisikan dalam konteks analisis tersebut.

III. HASIL DAN PEMBAHASAN

A. Hasil *Crawling Data*

Terlihat pada Gambar 5, hasil *crawling data* pada tanggal 17 Februari 2024. Data yang berhasil di-crawl mengenai Prabowo sebanyak 503 *tweet*.

	full_text	username	created_at
0	@andre_rosiade eh botak, bilang ke capres lu d...	Yhafidzun15	Fri Dec 15 04:10:46 +0000 2023
1	Elektabilitas tembus 45 persen, AHY: Prabowo b...	Iza_Zot	Fri Dec 15 04:10:46 +0000 2023
2	Beri pujian ke AHY, Prabowo: Pemimpin yang heb...	Iza_Zot	Fri Dec 15 04:10:43 +0000 2023
3	@kurawa jokowi hebat...setuju tapi melihat pe...	romobing	Fri Dec 15 04:10:41 +0000 2023
4	Prabowo prediksi AHY akan jadi pemimpin di tin...	Iza_Zot	Fri Dec 15 04:10:38 +0000 2023
...
498	@kompascom Memang tidak mudah, sekarang masyar...	HAPOSANM	Fri Dec 15 03:56:02 +0000 2023
499	Visi Prabowo - Gibran membawa harapan baru bag...	istimatiara	Fri Dec 15 03:56:01 +0000 2023
500	@supersaiyaindo @ch_chotimah2 Pembeneran akan ...	DylanBr03197352	Fri Dec 15 03:56:01 +0000 2023
501	Prabowo Gibran berkomitmen untuk mengoptimalka...	clara144495	Fri Dec 15 03:56:00 +0000 2023
502	Program-program Prabowo-Gibran, jembatan menuj...	FalikaNata	Fri Dec 15 03:56:00 +0000 2023

Gambar 5. Hasil *Crawling* Dengan *Keyword* Prabowo

B. Hasil Text Pre-processing

Berikut pada Tabel 1 adalah hasil *pre-processing* dari satu contoh teks yang diambil. Proses ini dimulai dari data mentah yang diperoleh dari hasil *crawling*, dan melalui serangkaian tahapan. Hasil ini mencerminkan langkah-langkah yang diambil untuk mempersiapkan data sebelum dilakukan analisis sentimen lebih lanjut.

Tabel 1. Hasil *Text Pre-processing* Prabowo

Tahap Preprocessing	Hasil
Data Awal	Program Prabowo membawa perubahan positif bagi industri nasional, dengan peluang ekspor yang semakin terbuka. #02UntukIndonesiaMaju 17Program PrioritasPraGibs
Cleaning Data	program prabowo membawa perubahan positif bagi industri nasional dengan peluang ekspor yang semakin terbuka program prioritaspragibs
Tokenizing	['program', 'prabowo', 'membawa', 'perubahan', 'positif', 'industri', 'nasional', 'peluang', 'ekspor', 'semakin', 'terbuka', 'program', 'prioritaspragibs']
Hasil Pre-processing	program prabowo bawa ubah positif industri nasional peluang ekspor makin buka program prioritaspragibs
Hasil Translator	<i>prabowo program bring positive change to the national industry export opportunities are increasingly open priority programs pragibs</i>

C. Hasil Pembobotan TF-IDF

Setelah tahap *pre-processing*, teks masuk ke dalam tahap pembobotan TF-IDF. Proses ini memberikan bobot pada setiap kata atau frasa dalam teks berdasarkan seberapa pentingnya kata tersebut dalam dokumen dibandingkan dengan seluruh teks yang ada.

```

Scores :
[[0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 ...
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]]

Words head :
                term      rank
226 [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ... 12.064568
208 [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ... 6.853408
210 [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ... 4.544439
209 [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ... 4.544439
212 [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ... 4.359989
211 [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ... 4.359989
63  [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ... 3.392141
    
```

Gambar 6. Scores TF-IDF

Pada Gambar 6, menunjukkan bahwa kata atau frasa dengan indeks 226 memiliki bobot TF-IDF sebesar 12,064568. Bobot TF-IDF mengindikasikan seberapa pentingnya kata atau frasa tersebut dalam dokumen jika

dibandingkan dengan seluruh koleksi dokumen. Jadi, semakin tinggi bobot TF-IDF, semakin signifikan kata atau frasa tersebut dalam dokumen tertentu. Dengan kata lain, kata atau frasa dengan bobot TF-IDF yang tinggi cenderung lebih unik atau penting dalam konteks dokumen tersebut.

D. Wordcloud

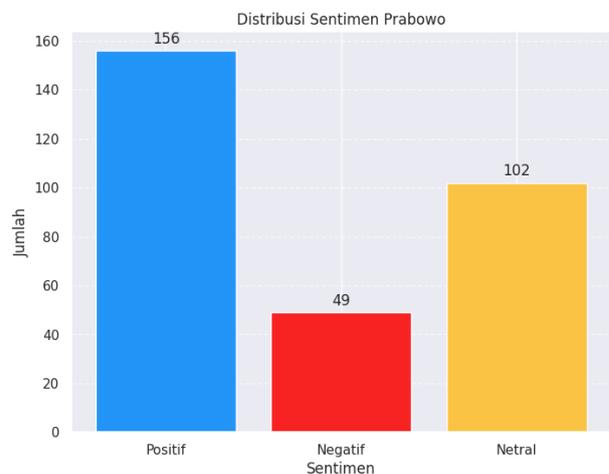
Wordcloud adalah representasi visual dari teks yang menyoroti kata-kata berdasarkan frekuensi kemunculannya [16]. Kata-kata yang paling sering muncul dalam teks akan ditampilkan dengan ukuran yang lebih besar, sementara kata-kata yang muncul lebih jarang akan ditampilkan dengan ukuran yang lebih kecil. Dalam kasus ini, hasil *wordcloud* Prabowo dapat dilihat pada Gambar 7.



Gambar 7. Hasil *Wordcloud* Prabowo

E. Pelatihan dan Pengujian Data dengan TextBlob

Dalam tahap pelatihan dan pengujian data menggunakan *TextBlob*, dilakukan analisis sentimen terhadap *tweet-tweet* yang telah dipreproses dalam bahasa Inggris. Hasil evaluasi menunjukkan bahwa dari total 306 data *tweet* yang dievaluasi, sebanyak 156 *tweet* memiliki sentimen positif, 102 *tweet* netral, dan 49 *tweet* memiliki sentimen negatif. Hal ini menggambarkan adanya perbedaan pendapat dan pandangan yang luas dalam masyarakat terkait Prabowo. Hasil bisa dilihat pada Gambar 8.



Gambar 8. Hasil Distribusi Sentimen Prabowo

Berdasarkan Gambar 9, hasil klasifikasi menggunakan *TextBlob* menunjukkan distribusi sentimen yang beragam, mencakup klasifikasi positif, netral, dan negatif. Hal ini mencerminkan keragaman pandangan dan perasaan yang ada dalam masyarakat terkait topik tersebut.

	full_text	tweet_english	klasifikasi
0	eh botak bilang capres lu dulu ngobrol ama ani...	uh bald said your presidential candidate used ...	Negatif
1	menhan prabowo subianto abdi korban prajurit t...	minister of defense prabowo subianto servant o...	Negatif
2	elektabilitas tembus 45 persen ahy prabowo men...	electability penetrates 45 percent ahy prabowo...	Positif
3	ahy demokrat siap kawal menang prabowo gibran ...	ahy democrat ready to control winning prabowo ...	Positif
4	kelola lingkungan lanjut komitmen prabowo gibran...	manage the environment further commitment prab...	Positif
...
302	sama pak mas silaturahmi kyai ulama amp santri...	with Mr. Kyai ulama & santri ponpes miftahul h...	Netral
303	sdhlah sdh lalu bukti makan ego sendiri pak pr...	sdhlah sdh then evidence of eating his own ego...	Positif
304	elektabilitas prabowo gibran selalu unggul ahy...	electability prabowo gibran always superior ah...	Positif

Gambar 9. Hasil Klasifikasi data Prabowo

Maka, persentase klasifikasi sentimen diperoleh dari hasil perhitungan menggunakan Persamaan 1 berikut:

$$\frac{\text{Jumlah komentar yang terklasifikasi}}{\text{Jumlah seluruh komentar}} \times 100\% \quad (1)$$

$$\text{Positif} = \frac{156}{306} \times 100\% = 50,98\%$$

$$\text{Negatif} = \frac{49}{306} \times 100\% = 16\%$$

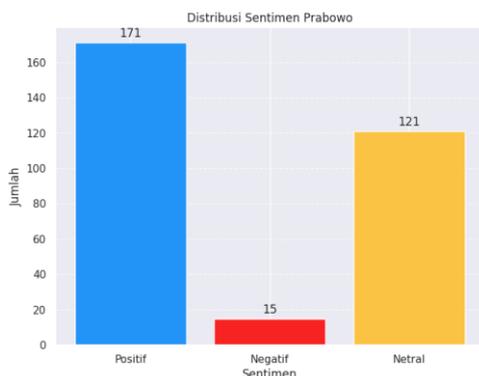
$$\text{Netral} = \frac{102}{306} \times 100\% = 33\%$$

F. Pelatihan dan Pengujian Data dengan Naïve Bayes

Pada penggunaan algoritma klasifikasi *Naïve Bayes*, diperoleh akurasi uji sebesar 78%, 0,3 data testing, serta 0,7 data training. Akurasi ini mencerminkan seberapa baik model *Naïve Bayes* dapat mengklasifikasikan data uji dengan benar berdasarkan pembelajaran dari data pelatihan. Berikut ini adalah hasil dari data testing *Naïve Bayes*.

Akurasi Tes: 0,7817589576547231

Hasil analisis menggunakan model klasifikasi *Naïve Bayes* menunjukkan distribusi sentimen dari total 307 data, dengan 171 diklasifikasikan sebagai positif, 121 sebagai netral, dan 15 sebagai negatif. Hasil dapat dilihat pada Gambar 10.



Gambar 10. Hasil Distribusi Sentimen Prabowo

Pada Gambar 11, dari hasil klasifikasi sentimen terhadap data yang berkaitan dengan Prabowo, terdapat perbedaan yang cukup signifikan antara penggunaan metode *TextBlob* dan *Naïve Bayes*. Dengan menggunakan *TextBlob*, jumlah *tweet* yang diklasifikasikan sebagai positif adalah 156, sedangkan yang diklasifikasikan sebagai negatif adalah 49, dan yang diklasifikasikan sebagai netral adalah 102. Di sisi lain, model *Naïve Bayes* menghasilkan jumlah yang berbeda, dengan 171 *tweet* yang diklasifikasikan sebagai positif, 15 *tweet* yang diklasifikasikan sebagai negatif, dan 121 *tweet* sebagai netral.

	full_text	tweet_english	klasifikasi	klasifikasi_Bayes
0	eh botak bilang capres lu dulu ngobrol ama ani...	uh bald said your presidential candidate used ...	Negatif	Netral
1	menhan prabowo subianto abdi korban prajurit t...	minister of defense prabowo subianto servant o...	Negatif	Negatif
2	elektabilitas tembus 45 persen ahy prabowo men...	electability penetrates 45 percent ahy prabowo...	Positif	Positif
3	ahy demokrat siap kawal menang prabowo gibran ...	ahy democrat ready to control winning prabowo ...	Positif	Positif
4	kelola lingkungan lanjut komitmen prabowo gibran...	manage the environment further commitment prab...	Positif	Netral
...
302	sama pak mas silaturahmi kyai ulama amp santri...	with Mr. Kyai ulama & santri ponpes miftahul h...	Netral	Netral
303	sdhlah sdh lalu bukti makan ego sendiri pak pr...	sdhlah sdh then evidence of eating his own ego...	Positif	Positif
304	elektabilitas prabowo gibran selalu unggul ahy...	electability prabowo gibran always superior ah...	Positif	Positif
305	jir striker prabowo	jir striker prabowo	Netral	Netral
306	2019 emank milu paling ngeri2 sedapebong kamp...	2019 emank milu most horrified 2 sedapebong k...	Positif	Positif

Gambar 11. Hasil Analisis Kedua Sentimen Data Prabowo

Dalam analisis sentimen menggunakan *TextBlob*, terlihat bahwa sekitar 50,98% dari total komentar diklasifikasikan sebagai sentimen positif, sementara 16,01% diklasifikasikan sebagai sentimen negatif, dan 33,33% diklasifikasikan sebagai sentimen netral. Sedangkan, hasil analisis menggunakan *Naïve Bayes* menunjukkan akurasi sebesar 78,18%. Selain itu, terdapat perbedaan dalam jumlah sentimen netral yang diklasifikasikan oleh kedua metode tersebut. Penelitian ini sejalan dengan [3] bahwa untuk mengetahui sentimen masyarakat terhadap isu atau topik yang menjadi trending di media sosial dapat menggunakan analisis sentimen. Menurut kajian teori, metode *Naïve Bayes* sering kali lebih efektif dalam mengklasifikasikan sentimen positif karena pendekatan probabilitiknya yang mampu mengenali pola-pola kata yang sering muncul dalam konteks positif. Sebaliknya, *TextBlob* menggunakan pendekatan berbasis aturan dan leksikon [17] yang memberikan penekanan lebih pada kata-kata negatif, sehingga cenderung menghasilkan klasifikasi negatif yang lebih tinggi.

IV. KESIMPULAN DAN SARAN

Kesimpulan yang dapat ditarik dari penelitian ini adalah analisis sentimen menggunakan algoritma *TextBlob* dan *Naïve Bayes* mampu mengklasifikasikan sentimen dengan baik. Hasil penelitian menunjukkan bahwa *Naïve Bayes* menghasilkan akurasi sebesar 78,18%. Yang mencerminkan kualitas model ini dalam klasifikasi yang optimal (optimal classification). Meskipun begitu, kelebihan *TextBlob* terletak pada kemudahannya dalam implementasi dan penggunaannya, serta kemampuannya untuk memberikan gambaran umum tentang sentimen masyarakat dengan cukup baik. Dengan begitu, penelitian ini memberikan solusi untuk menilai sentimen masyarakat secara otomatis dengan cepat dan akurat.

Berdasarkan hasil analisis, disarankan untuk penelitian selanjutnya dapat menggunakan metode serta algoritma yang

lebih beragam agar mendapatkan tingkat akurasi yang lebih baik lagi karena analisis sentimen pada saat ini hanya menggunakan dua algoritma saja yaitu *TextBlob* dan *Naïve Bayes*.

REFERENSI

- [1] S. Puad and A. Susilo Yuda Irawan, "Analisis Sentimen Masyarakat Pada Twitter Terhadap Pemilihan Umum 2024 Menggunakan Algoritma *Naïve Bayes*," 2023.
- [2] K. Zuhri, N. Adha, and O. Saputri, "Analisis Sentimen Masyarakat Terhadap Pilpres 2019 Berdasarkan Opini Dari Twitter Menggunakan Metode *Naïve Bayes Classifier*," 2020. [Online]. Available: <https://journal-computing.org/index.php/journal-cisa/index>
- [3] R. Azhar, A. Surahman, and C. Juliane, "Analisis Sentimen Terhadap *Cryptocurrency* Berbasis Python *TextBlob* Menggunakan Algoritma *Naïve Bayes*," 2022.
- [4] P. Al Muqsih Prasetyo and A. Hermawan, "Analisis Sentimen Twitter Terhadap Pemilihan Presiden Menggunakan Algoritma *Naïve Bayes*," *INFOTECH: Jurnal Informatika & Teknologi*, vol. 4, no. 2, pp. 224–233, Dec. 2023, doi: 10.37373/infotech.v4i2.863.
- [5] K. Verena, S. Toy, Y. A. Sari, and I. Cholissodin, "Analisis Sentimen Twitter menggunakan Metode *Naïve Bayes* dengan *Relevance Frequency Feature Selection* (Studi Kasus: Opini Masyarakat mengenai Kebijakan New Normal)," 2021. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [6] R. Vindua and A. U. Zailani, "Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python," *JURIKOM (Jurnal Riset Komputer)*, vol. 10, no. 2, p. 479, Apr. 2023, doi: 10.30865/jurikom.v10i2.5945.
- [7] M. Akbar Maulana, A. Setyanto, and M. P. Kurniawan, "Seminar Nasional Teknologi Informasi dan Multimedia," *UNIVERSITAS AMIKOM Yogyakarta*, 2018, [Online]. Available: <http://www.twitter.com>
- [8] F. A. Wenando, R. Hayami, and A. J. Anggrawan, "Analisis Sentimen Pada Pemerintahan Terpilih Pada Pilpres 2019 Ditwitter Menggunakan Algoritme *Naïve Bayes*," *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 7, no. 1, pp. 101–106, Dec. 2020, doi: 10.33330/jurteks.v7i1.851.
- [9] B. Indra Kusuma and A. Nugroho, "Cyberbullying Detection on Twitter Uses the Support Vector Machine Method," *Jurnal Teknik Informatika (JUTIF)*, vol. 5, no. 1, pp. 11–17, 2024, doi: 10.52436/1.jutif.2024.5.1.809.
- [10] D. Hazarika, G. Konwar, S. Deb, and D. J. Bora, "Sentiment Analysis on Twitter by Using TextBlob for Natural Language Processing," in *Proceedings of the International Conference on Research in Management & Technovation 2020*, PTI, Jan. 2020, pp. 63–67. doi: 10.15439/2020km20.
- [11] L. Aji Andika and P. Amalia Nur Azizah, "Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode *Naïve Bayes Classifier*," 2019.
- [12] P. Pasek, O. Mahawardana, G. A. Sasmita, P. Agus, and E. Pratama, "Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap 'Figure Pemimpin' Menggunakan Python," 2022.
- [13] M. H. Al-Areef and K. Saputra, "Nomor 2," *Agustus*, vol. 22, pp. 270–279, 2023, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jis/index>
- [14] D. Aby Vonega, A. Fadila, and D. Ely Kurniawan, "Analisis Sentimen Twitter Terhadap Opini Publik Atas Isu Pencalonan Puan Maharani dalam Pilpres 2024," 2022. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [15] Asno Azzawagama Firdaus, Anton Yudhana, and Imam Riadi, "Analisis Sentimen Pada Proyeksi Pemilihan Presiden 2024 Menggunakan Metode Support Vector Machine," *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 3, no. 2, pp. 236–245, Jun. 2023, doi: 10.51454/decode.v3i2.172.
- [16] Y. Akbar and T. Sugiharto, "Analisis Sentimen Pengguna Twitter di Indonesia Terhadap ChatGPT Menggunakan Algoritma C4.5 dan *Naïve Bayes* (Yuma Akbar 1*, Tri Sugiharto 2) Analisis Sentimen Pengguna Twitter di Indonesia Terhadap ChatGPT Menggunakan Algoritma C4.5 dan *Naïve Bayes*," *Jurnal Sains dan Teknologi*, vol. 5, no. 1, pp. 115–122, 2023, doi: 10.55338/saintek.v4i3.1368.
- [17] A. Robi Padri, "HCI dan Media Sosial: Studi Kasus Analisis Sentimen Pilpres 2024 di Twitter Menggunakan *Naïve Bayes Classifier*," *Jurnal SIMETRIS*, vol. 14, no. 2, 2023.