

Analisis sentimen publik di twitter terhadap pelantikan presiden Prabowo menggunakan algoritma *Naïve Bayes*

Analysis of public sentiment on twitter towards president Prabowo's inauguration using the Naïve Bayes algorithm

Adhika Pramita Widyassari¹⁾, Dea Salsabilla²⁾, M. Ali Amrozi³⁾

^{1), 2), 3)} Program Studi Informatika, Sekolah Tinggi Teknologi Ronggolawe

Jl. Kampus Ronggolawe No.1 Mentul, Kec. Cepu, Kabupaten Blora, Jawa Tengah

Email : dikasari9@gmail.com¹⁾, deasalsabilla06@gmail.com²⁾, maliamrozi8@gmail.com³⁾

Abstrak

Analisis sentimen di media sosial, khususnya Twitter, menjadi metode yang efektif untuk memahami opini publik terhadap peristiwa politik seperti pelantikan Presiden Prabowo. Penelitian ini bertujuan untuk mengidentifikasi sentimen masyarakat Indonesia terkait pelantikan Presiden Prabowo melalui analisis sentimen berbasis machine learning menggunakan algoritma Multinomial Naïve Bayes. Data dikumpulkan dari Twitter dengan kata kunci dan hashtag relevan, mencakup rentang waktu sebelum dan sesudah pelantikan untuk menangkap dinamika perubahan sentimen. Proses preprocessing dilakukan melalui pembersihan teks, penghapusan stop words, tokenisasi, dan stemming guna meningkatkan akurasi model. Hasil klasifikasi menunjukkan distribusi sentimen publik, dengan mayoritas bersifat netral (52,63%), diikuti sentimen positif (42,98%), dan negatif (4,39%). Model mencapai akurasi 75%, menunjukkan kinerja yang cukup baik untuk klasifikasi teks pendek. Kontribusi penelitian ini terletak pada penerapan analisis sentimen terhadap peristiwa pelantikan Presiden Indonesia yang spesifik, dengan fokus pada periode kritis sebelum dan sesudah pelantikan, yang belum banyak diteliti sebelumnya. Kebaruan dari penelitian ini adalah penggunaan data Twitter secara real-time terkait dengan peristiwa politik terkini (pelantikan Presiden Prabowo), serta penekanan pada sentimen netral yang memberikan dimensi lebih dalam terhadap pemahaman publik. Temuan ini diharapkan dapat menjadi dasar dalam merancang strategi komunikasi publik yang lebih efektif di media sosial.

Kata kunci: analisis sentimen, naïve bayes, pelantikan presiden prabowo, twitter

Abstract

Sentiment analysis on social media, especially Twitter, is an effective method to understand public opinion towards political events such as the inauguration of President Prabowo. This study aims to identify the sentiment of the Indonesian people regarding the inauguration of President Prabowo through machine learning-based sentiment analysis using the Multinomial Naïve Bayes algorithm. Data was collected from Twitter with relevant keywords and hashtags, covering the time span before and after the inauguration to capture the dynamics of sentiment changes. The preprocessing process was carried out through text cleaning, removing stop words, tokenization, and stemming to improve model accuracy. The classification results show the distribution of public sentiment, with the majority being neutral (52.63%), followed by positive sentiment (42.98%), and negative (4.39%). The model achieved an accuracy of 75%, showing quite good performance for short text classification. The contribution of this study lies in the application of sentiment analysis to the specific event of the inauguration of the President of Indonesia, with a focus on the critical period before and after the inauguration, which has not been widely studied before. The novelty of this study is the use of real-time Twitter data related to current political events (inauguration of President Prabowo), as well as the emphasis on neutral sentiment which provides a deeper dimension to public understanding. It is hoped that these findings can be the basis for designing more effective public communication strategies on social media.

Keywords: naïve bayes, prabowo presidential inauguration, twitter, sentiment analysis

1. PENDAHULUAN

Era digital saat ini, media sosial telah menjadi salah satu platform utama untuk mengekspresikan opini, terutama terkait isu-isu sosial dan politik. Twitter, sebagai salah satu media sosial yang paling banyak digunakan di Indonesia, menyediakan ruang bagi masyarakat untuk menyampaikan pandangan mereka dalam waktu nyata [1]. Dengan karakteristiknya yang memungkinkan penyebaran informasi secara cepat dan interaktif, Twitter menjadi salah satu sumber data yang sangat bernilai dalam memahami respons publik terhadap berbagai peristiwa,

termasuk peristiwa politik penting seperti pelantikan Presiden. Pelantikan Presiden Indonesia Prabowo Subianto yang dilaksanakan tanggal 20 Oktober 2024, merupakan momen politik yang dinanti dan berdampak luas terhadap berbagai lapisan masyarakat. Sebagai seorang tokoh yang memiliki pengikut serta kritikus dalam jumlah besar, pelantikannya diharapkan memicu reaksi publik yang beragam. Oleh karena itu, memahami opini publik terkait pelantikan ini menjadi penting, tidak hanya bagi pemerintah untuk memahami aspirasi rakyat, tetapi juga sebagai dasar dalam merancang strategi komunikasi dan kebijakan yang lebih efektif dan sesuai dengan harapan masyarakat. Dalam konteks ini, analisis sentimen di media sosial menjadi metode yang efektif untuk menangkap dan memetakan opini publik secara cepat dan sistematis. Analisis sentimen memungkinkan identifikasi pola emosional masyarakat (positif, negatif, atau netral) terhadap suatu peristiwa atau tokoh, sehingga dapat diinterpretasikan dalam kerangka sosial dan politik yang lebih luas [2]. Studi oleh [3] menunjukkan bahwa data dari media sosial dapat digunakan untuk memprediksi sikap politik masyarakat terhadap peristiwa politik besar seperti pemilu. Selanjutnya, penelitian [4] menemukan bahwa sentimen publik di media sosial selama periode pemilu berhubungan erat dengan tingkat kepercayaan terhadap pemerintah. Penelitian lain oleh [5] menyoroti bagaimana media sosial memperkuat polarisasi politik, yang tercermin dari pola sentimen selama kampanye politik.

Metode *machine learning* telah dikembangkan untuk analisis sentimen, masing-masing dengan keunggulan dan keterbatasannya. *Naïve Bayes* menjadi salah satu pendekatan yang banyak digunakan karena kesederhanaan model, efisiensi komputasi, serta performa yang baik dalam klasifikasi teks pendek seperti tweet [6][7]. Dalam konteks data berukuran besar dan *real-time* seperti Twitter, kecepatan proses klasifikasi *Naïve Bayes* menjadi nilai tambah penting [8]. Namun, beberapa algoritma lain seperti *Support Vector Machine* (SVM) dan *Random Forest* juga kerap digunakan dalam analisis sentimen karena keakuratannya yang lebih tinggi dalam beberapa kasus [9], [10], dan [11]. Misalnya, penelitian oleh [12] menunjukkan bahwa SVM memiliki keunggulan dalam menangani data sentimen yang seimbang dan menghasilkan klasifikasi yang lebih presisi dibandingkan *Naïve Bayes*. Sementara itu, model berbasis *Deep Learning* seperti LSTM (*Long Short-Term Memory*) menawarkan performa yang sangat baik dalam analisis sentimen dengan konteks kalimat yang lebih kompleks, meskipun membutuhkan sumber daya komputasi yang lebih besar [13], [14]. Dalam penelitian ini, *naïve bayes* dipilih karena mempertimbangkan efisiensi waktu, kebutuhan sumber daya komputasi yang rendah, serta karakteristik teks Twitter yang cenderung pendek dan eksplisit [15]. Pemilihan ini juga didukung oleh studi terbaru yang menyatakan bahwa untuk data teks pendek, *naïve bayes* tetap menjadi pilihan yang kompetitif dibandingkan model-model kompleks [16].

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat di Twitter terkait pelantikan Presiden Indonesia Prabowo dengan menerapkan algoritma *naïve bayes*. Proses penelitian melibatkan pengumpulan data melalui kata kunci dan tagar yang relevan dengan peristiwa pelantikan, serta tahap-tahap pra-pemrosesan data, seperti pembersihan teks, penghapusan *stop words*, *tokenisasi*, dan *stemming*, yang bertujuan untuk meningkatkan kualitas model analisis sentimen. Hasil akhir berupa klasifikasi sentimen (positif, negatif, atau netral) akan memberikan gambaran umum tentang persepsi masyarakat terhadap pelantikan Presiden Indonesia Prabowo, serta membantu pihak-pihak terkait dalam memahami respons publik terhadap pemerintahan yang baru.

Penelitian ini memiliki beberapa perbedaan signifikan dibandingkan dengan studi-studi sebelumnya yang melakukan analisis sentimen terhadap peristiwa politik menggunakan data media sosial, khususnya Twitter. Sebelumnya, penelitian oleh [2], [3], [4], dan [15] telah mengeksplorasi sentimen publik terhadap peristiwa besar seperti pemilu, di mana metode yang digunakan lebih fokus pada pemilihan presiden atau kampanye politik secara umum. Namun, penelitian ini lebih terfokus pada sentimen publik terhadap pelantikan Presiden Prabowo pada 2024, yang merupakan momen politik yang memiliki dinamika dan reaksi publik yang berbeda dibandingkan dengan proses pemilu atau kampanye. Selain itu, penelitian sebelumnya cenderung menggunakan berbagai algoritma seperti *Support Vector Machine* (SVM) atau *Random Forest*,

yang terkenal karena keakuratannya dalam menangani data sentimen yang seimbang. Namun, algoritma *naïve bayes* yang digunakan dalam penelitian ini dipilih karena kesederhanaannya, efisiensi waktu, kebutuhan sumber daya komputasi yang rendah dalam mengklasifikasikan teks pendek seperti tweet. Beberapa penelitian, seperti yang dilakukan oleh [6], juga menggunakan *naïve bayes*, namun dalam konteks yang lebih umum. Penelitian ini memberikan kontribusi dengan mengaplikasikan algoritma *naïve bayes* pada konteks sentimen politik dalam peristiwa pelantikan presiden, dengan memperhitungkan faktor waktu, yaitu sebelum dan sesudah pelantikan, untuk melihat perubahan respons publik. Penelitian ini memberikan gambaran yang lebih konkret tentang bagaimana masyarakat merespons pelantikan seorang presiden yang sangat dinantikan.

Hasil penelitian ini diharapkan tidak hanya memberikan wawasan tentang distribusi sentimen publik terkait pelantikan Presiden Prabowo tetapi juga menjadi masukan bagi para pemangku kepentingan dalam mengukur respons awal masyarakat terhadap pemerintahan baru. Dengan demikian, penelitian ini memiliki relevansi yang tinggi baik dari segi akademis maupun praktis dalam bidang ilmu politik, komunikasi, dan teknologi informasi.

2. DASAR TEORI

2.1 Sentimen Analisis

Sentimen analisis, juga dikenal sebagai *opinion mining*, adalah cabang dari *Natural Language Processing* (NLP) yang bertujuan untuk mengidentifikasi, mengekstrak, dan mengklasifikasikan opini, emosi, atau sikap yang terdapat dalam teks. Sentimen analisis biasanya digunakan untuk memahami persepsi publik terhadap suatu topik, produk, atau individu berdasarkan data tekstual yang berasal dari berbagai sumber, seperti media sosial, ulasan online, atau survei [17].

Komponen utama dalam sentimen analisis mencakup berbagai aspek penting. Pertama, teks sebagai input, di mana sentimen analisis memproses data tekstual berupa kalimat, paragraf, atau dokumen. Sumber data yang umum digunakan mencakup media sosial, ulasan pelanggan, artikel berita, atau komentar online. Kedua, sentimen dalam teks biasanya diklasifikasikan ke dalam tiga kategori utama: positif, negatif, dan netral. Sentimen positif menunjukkan opini atau emosi yang mendukung, menyukai, atau menghargai, sementara sentimen negatif mencerminkan kritik, ketidakpuasan, atau sikap negatif. Sentimen netral bersifat objektif dan tidak menunjukkan kecenderungan tertentu [18].

Tahapan sentimen analisis dimulai dengan *crawling* data, yaitu pengumpulan data dari berbagai platform menggunakan teknik *scraping* atau API, seperti Twitter API [19]. Selanjutnya, dilakukan *preprocessing* data, yang mencakup *filtering* untuk menghilangkan elemen tidak relevan, penghapusan *stopwords*, *tokenisasi* untuk memecah teks menjadi unit-unit kata, dan *stemming* untuk mengubah kata menjadi bentuk dasarnya. Setelah data diproses, tahapan klasifikasi sentimen dilakukan menggunakan algoritma tertentu, seperti *Naïve Bayes*, *Support Vector Machine* (SVM), atau *Deep Learning*, untuk menentukan kategori sentimen.

2.2 Naïve Bayes

Naïve Bayes adalah salah satu algoritma pembelajaran mesin (*machine learning*) berbasis teori probabilitas yang sering digunakan untuk tugas klasifikasi. Algoritma ini didasarkan pada *Bayes Theorem*, yang menggambarkan hubungan probabilistik antara variabel independen dan variabel dependen. Meskipun disebut “*naïve*” karena asumsi independensi antar fitur, algoritma ini telah terbukti efektif dan efisien untuk berbagai aplikasi klasifikasi berbasis teks, termasuk analisis sentimen, deteksi spam, dan pengenalan dokumen [20]. Berikut adalah tabel 1 yang merangkum kelebihan dan kekurangan *Naïve Bayes* [21]:

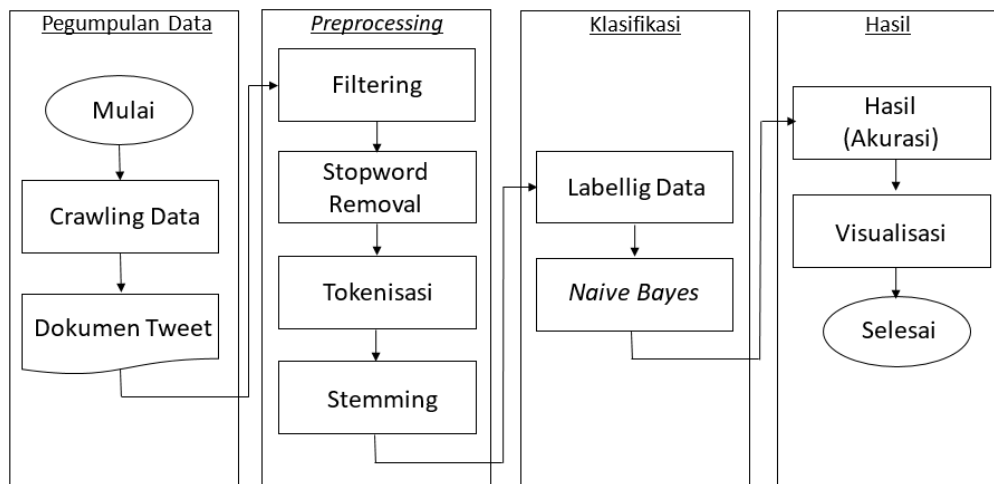
Tabel 1. *Tabel Kelebihan dan kekurangan Naïve Bayes*

Kelebihan	Kekurangan
Sederhana dan mudah diimplementasikan: algoritma ini mudah dipahami dan diimplementasikan.	Asumsi independensi fitur: mengasumsikan bahwa semua fitur independen, yang jarang terjadi dalam dunia nyata.
Cepat dan efisien: memiliki kecepatan komputasi tinggi, terutama untuk dataset besar.	Kesensitifan terhadap data latih: jika suatu kelas tidak terwakili dalam data latih, probabilitasnya menjadi nol.
Efektif untuk teks: sangat cocok untuk tugas klasifikasi berbasis teks, seperti analisis sentimen.	Tidak fleksibel untuk fitur berhubungan: tidak efektif jika terdapat hubungan antar fitur dalam data.
Mengatasi masalah overfitting: tahan terhadap overfitting pada dataset besar.	Kesulitan dengan data numerik kompleks: kurang cocok untuk data numerik yang tidak berdistribusi normal.
Mendukung klasifikasi multikelas: mampu mengklasifikasikan data ke lebih dari dua kategori.	Tidak dapat menangani data yang tidak terdistribusi merata: performa menurun pada dataset tidak seimbang.

Dalam berbagai penelitian, *Naïve Bayes* sering menjadi metode andalan untuk tugas-tugas klasifikasi berbasis teks. Meskipun sederhana, algoritma ini sering menunjukkan kinerja yang kompetitif jika dibandingkan dengan algoritma yang lebih kompleks. Kombinasi dengan teknik *preprocessing*, seperti *stopword removal*, tokenisasi, dan *stemming*, juga dapat meningkatkan akurasi *Naïve Bayes* dalam berbagai aplikasi. Dengan kemampuan untuk menangani data besar, efisiensi waktu komputasi, dan kemudahan implementasi, *Naïve Bayes* tetap menjadi algoritma yang relevan untuk diterapkan dalam berbagai konteks analisis data, terutama analisis berbasis teks.

3. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis sentimen berbasis *machine learning* yaitu *Naïve Bayes*. Metode ini dipilih untuk mengidentifikasi dan mengklasifikasikan sentimen publik secara objektif berdasarkan data teks dari media sosial, khususnya Twitter. Adapun tahapan penelitian ini seperti yang tertuang pada [gambar 1](#) meliputi: Pengumpulan data, *Preprocessing*, klasifikasi dan hasil.

**Gambar 1.** *Tahapan Penelitian*

3.1 Pengumpulan Data

Sumber data pada penelitian ini diambil dari twitter yang memuat opini publik terhadap pelantikan Presiden Indonesia 2024 Prabowo Subianto. Jenis data yang diambil berupa teks. Teknik Pengumpulan data dengan cara *crawling* data menggunakan API Twitter dengan bantuan tools “*tweet harvest*”. *Crawling* data dilakukan berdasarkan kata kunci atau tagar yang relevan, yaitu: #PelantikanPrabowo, #PresidenPrabowo, #Prabowo. Periode Pengumpulan Data: Data

dikumpulkan dalam rentang waktu tertentu, yaitu 1 hari sebelum dan 1 hari sesudah pelantikan. Hal ini bertujuan untuk menangkap perbedaan sentimen publik sebelum dan setelah pelantikan. Pemrograman yang digunakan untuk memproses tahap awal hingga akhir yaitu perograman phyton.

3.2 Preprocessing

Tahapan preprocessing teks merupakan langkah penting dalam analisis sentimen karena data teks yang diambil dari media sosial, seperti twitter, umumnya tidak terstruktur dan mengandung banyak “noise”. Dengan *preprocessing*, data dapat diubah menjadi bentuk yang lebih bersih dan relevan untuk pemrosesan lebih lanjut menggunakan algoritma *machine learning* [22]. Berikut adalah penjelasan dari tahapan-tahapan *preprocessing* yang umum digunakan dalam analisis sentiment [6][23]:

a) Filtering/Cleaning

Tahap pertama adalah membersihkan teks dari elemen-elemen yang tidak relevan. Pembersihan teks membantu mengurangi “noise” dalam data, sehingga algoritma dapat lebih fokus pada kata-kata yang relevan dengan sentimen [7]. Adapun elemen-elemen yang dimaksud seperti:

- 1] Karakter spesial (misalnya: @, #, !, ?).
- 2] Simbol dan angka yang tidak membawa makna sentimen, kecuali jika memiliki arti khusus.
- 3] URL atau tautan yang sering muncul dalam tweet dan tidak relevan untuk analisis sentimen.
- 4] Mentions (@username) yang mengacu pada pengguna lain, karena biasanya tidak berkontribusi pada isi sentimen.
- 5] Hashtags: Jika hashtag memiliki makna penting, dapat dipertimbangkan untuk diolah secara terpisah. Namun, biasanya hashtag dihapus kecuali jika digunakan sebagai kata kunci.

b) Stopword Removal

Stop words adalah kata-kata umum dalam bahasa, seperti “yang,” “dan,” “dari,” “ke,” yang tidak memiliki kontribusi penting terhadap sentimen atau makna utama teks. Menghapus stop words membantu mengurangi jumlah kata yang harus diproses, sehingga meningkatkan efisiensi algoritma. Ada daftar stop words khusus untuk setiap bahasa, dan dapat disesuaikan sesuai kebutuhan penelitian.

c) Tokenisasi

Tokenisasi adalah proses memecah teks menjadi unit-unit kata yang lebih kecil, yang disebut “token.” Misalnya, kalimat “Pelantikan Presiden sukses” akan diubah menjadi [“Pelantikan”, “Presiden”, “sukses”]. Tokenisasi membantu algoritma dalam memproses dan menganalisis teks per kata, sehingga setiap kata dapat diberi label sentimen secara individu.

d) Stemming

Stemming adalah proses mengubah kata menjadi bentuk dasarnya. Misalnya, kata “berlari,” “berlarian,” dan “lari” akan dikonversi menjadi “lari.” Stemming bertujuan untuk menghilangkan variasi bentuk kata yang berasal dari akhiran atau imbuhan, sehingga algoritma dapat memahami kata dasar yang sebenarnya. Stemming Bahasa Indonesia dilakukan menggunakan stemmer khusus karena morfologi bahasa Indonesia yang unik, seperti adanya awalan (misalnya, ber-, di-, ter-) dan akhiran (misalnya, -kan, -an).

3.3 Klasifikasi

Pada bagian ini dilakukan klasifikasi sentiment analisis, dimana pada penelitian ini ada dua tahap yaitu labelling data dan prediksi dengan algoritma *Naïve Bayes*.

a) Labelling data

Labelling atau pemberian label adalah tahap penting yang bertujuan untuk menentukan kategori sentimen setiap teks, misalnya, sebagai sentimen positif, negatif, atau netral [24]. Label ini akan digunakan oleh algoritma *machine learning* sebagai “target” atau “output” yang ingin diprediksi saat model dilatih. Dengan memberikan label pada setiap data teks, algoritma dapat belajar dari data berlabel ini untuk mendeteksi pola yang menunjukkan sentimen tertentu [25]. Ini penting untuk membuat model yang dapat memprediksi sentimen teks baru dengan baik berdasarkan pembelajaran sebelumnya.

b) *Naïve Bayes*

Algoritma *Naïve Bayes* merupakan salah satu metode machine learning yang sangat populer untuk analisis sentimen, terutama karena kesederhanaannya [26]. Algoritma ini didasarkan pada prinsip *Teorema Bayes*, yang menghitung probabilitas terjadinya suatu kejadian berdasarkan informasi atau data yang sudah ada [26].

Pada penelitian ini algoritma *Naïve Bayes* menghitung probabilitas bahwa teks (tweet, ulasan, atau kalimat) termasuk dalam kategori sentimen tertentu (misalnya, positif, negatif, atau netral) berdasarkan kata-kata yang ada dalam teks tersebut [27]. Disebut “*naïve*” (naif) karena algoritma ini mengasumsikan bahwa setiap kata dalam teks adalah independen atau tidak saling bergantung satu sama lain, meskipun dalam kenyataannya kata-kata sering kali memiliki korelasi [28], [29], [30].

Berdasarkan Teorema Bayes, algoritma menghitung probabilitas untuk setiap kelas sentimen dengan rumus 1 berikut [6], [26], [28]:

$$P(\text{Sentimen}|\text{Teks}) = \frac{P(\text{Teks}|\text{Sentimen}) \cdot P(\text{Sentimen})}{P(\text{Teks})} \quad \dots(1)$$

- a). $P(\text{Sentimen}|\text{Teks})$: Probabilitas suatu teks memiliki sentimen tertentu (target).
- b). $P(\text{Teks}|\text{Sentimen})$: Probabilitas munculnya teks tertentu dalam sentimen tertentu
- c). $P(\text{Sentimen})$: Probabilitas umum dari suatu sentimen (positif, negatif, atau netral) dalam keseluruhan data.
- d). $P(\text{Teks})$: Probabilitas umum dari teks dalam dataset (ini bersifat tetap untuk semua kelas, sehingga sering diabaikan dalam perhitungan komparatif).

3.4 Hasil

Hasil dari penelitian ini berupa: a) hasil akurasi dari metode *Naïve Bayes* dalam melakukan klasifikasi atau prediksi sentimen analisis melalui teks; dan b) Hasil berupa Frekuensi dari masing-masing sentimen (positif, negatif, netral) yang kemudian divisualisasikan dalam bentuk grafik atau diagram untuk memudahkan interpretasi dan pemahaman tentang sentimen publik secara keseluruhan.

4. PENGUJIAN DAN PEMBAHASAN

4.1 *Crawling Data*

Crawling data menggunakan API Twitter dengan tools “tweet harvest”, rentang waktu tanggal 19 Oktober 2024 sampai dengan 21 Oktober 2024 dengan tagar: #PelantikanPrabowo, #PresidenPrabowo, #Prabowo, diperoleh data sejumlah 114 tweet. [Gambar 2](#) berikut ini tampilan data teks tweet yang diperoleh atau data mentah hasil dari *crawling data*.

	conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_screen_name	lang	location	qu
0	1848152212570272087	Sun Oct 20 23:59:59 +0000 2024	3	Kabinet Merah Putih Prabowo-Gibran tergemuk se...	1848152212570272087	NaN	NaN	in	London - Jakarta	
1	1848142980169056259	Sun Oct 20 23:59:59 +0000 2024	0	@Vmbelink Keren @Russia	1848152212083704300	NaN	Vmbelink	hu	Cityam California	
2	1847966784483414323	Sun Oct 20 23:59:59 +0000 2024	0	@kinanthis @prabowo ak mnu mam gratis	1848152212058554516	NaN	kinanthis	in	Asgard	
3	1848152195046506908	Sun Oct 20 23:59:54 +0000 2024	1	Presiden RI terpilih Prabowo Subianto mengguna...	1848152195046506908	https://pbs.twimg.com/ext_tw_video_thumb/18481...	NaN	in	Indonesia	
4	1848146226291917243	Sun Oct 20 23:59:44 +0000 2024	0	@prabowo Selamat bertugas jenderal harapan	1848152150079390202	NaN	prabowo	in	Punwokerto, Indonesia	

Gambar 2. Data teks tweet / data mentah hasil dari crawling data

4.2 Preprocessing

Pada bagian ini dilakukan beberapa tahapan preprocessing, antara lain: filtering/cleaning, stopwords removal, tokenisasi, stemming.

a) Filtering/Cleaning

Berikut ini merupakan hasil *filtering/cleaning* yaitu menghapus elemen-elemen yang tidak relevan seperti yang tersaji pada [gambar 3](#) berikut.

```
def clean_twitter_text(text):
    text = re.sub(r'@[A-Za-z0-9_]+', '', text)
    text = re.sub(r'#\w+', '', text)
    text = re.sub(r'RT[\s]+', '', text)
    text = re.sub(r'https?://\S+', '', text)

    text = re.sub(r'[^A-Za-z0-9 ]', '', text)
    text = re.sub(r'\s+', ' ', text).strip()

    return text

df['full_text'] = df['full_text'].apply(clean_twitter_text)
```



	full_text	username	created_at
0	tinggal tunggu janji prabowo tuk memberantas k...	DimasS17827	Thu Oct 24 23:59:59 +0000 2024
1	pak ps pasti lebih faham apa yg harus dilakuka...	LilisSu27144014	Thu Oct 24 23:59:55 +0000 2024
2	gimana mau tangkap gibran ada disampingnya	DatuLiang	Thu Oct 24 23:59:52 +0000 2024
3	berikut adalah 150 komentar terkait masyarakat...	maudy_ayu20	Thu Oct 24 23:59:49 +0000 2024
4	makan bergizi gratis di sekolah cc pak bisa di...	delapandua10486	Thu Oct 24 23:59:46 +0000 2024
...	...	...	...
405	pertumbuhan ekonomi kita makin terasa di tanga...	akmal_januar7	Thu Oct 24 22:44:56 +0000 2024
406	kenapa ke pres prab katanya himbauankenapa ga...	sudsalesman	Thu Oct 24 22:44:43 +0000 2024
407	pinjol judul d tutup lahya ssh upetinya gedeee	chocopink799	Thu Oct 24 22:44:35 +0000 2024
408	mana makan gratisny	ucilfh	Thu Oct 24 22:44:33 +0000 2024
409	klo emg harus ada perwakilan papua di kabinet ...	Helmutputra48018	Thu Oct 24 22:44:29 +0000 2024

Gambar 3. Tampilan Data Teks setelah dilakukan Filtering/cleaning

b) Stopwod Removal

Berikut ini merupakan hasil penghapusan daftar *Stop words*, yang tidak memiliki kontribusi penting terhadap sentimen atau makna utama teks, seperti yang tersaji pada [gambar 4](#) berikut.

```
# Stopwords
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemo

more_stop_words = ["tidak"]

# Dapatkan stopwords dan tambahkan custom stopwords
stop_word_factory = StopWordRemoverFactory()
stop_words = stop_word_factory.get_stop_words()
stop_words.extend(more_stop_words)

# Buat stopwords remover dengan daftar stopwords terbaru
stopword_remover = stop_word_factory.create_stop_word_remover()

def stopword(str_text):
    str_text = stopword_remover.remove(str_text)
    return str_text

# Aplikasikan fungsi stopwords ke kolom 'full_text'
df['full_text'] = df['full_text'].apply(lambda x: stopword(x))
df.head()
```



	full_text	username	created_at
0	Sembah sujud fufufafa	niskala1214	Wed Oct 23 23:49:48 +0000 2024
1	Ini macet ato lancar	asulaut01	Wed Oct 23 23:49:47 +0000 2024
2	Wapresmu suruh ikut digembleng pak	EkoApan	Wed Oct 23 23:49:37 +0000 2024
3	Baru hari2 awal kerja udah banyak bener uh Yan...	_the_blue_sea_	Wed Oct 23 23:49:22 +0000 2024
4	ngelindas anak orang nyawa melayang cuma 5thun ...	raniaguan_	Wed Oct 23 23:49:20 +0000 2024

Gambar 4. Tampilan Data Teks setelah dilakukan Stopword Removalc) *Tokenisasi*

Berikut ini merupakan hasil tokenisasi yaitu memecah teks menjadi unit-unit kata yang lebih kecil atau token, seperti yang tersaji pada [gambar 5](#) berikut.

	full_text
0	[Sembah, sujud, fufufafa]
1	[Ini, macet, ato, lancar]
2	[Wapresmu, suruh, ikutdigembleng, pak]
3	[Baru, hari2, awal, kerja, udah, banyak, bener...
4	[ngelindas, anak, orang, nyawa, melayang, cuma...
...	...
109	[Gelar, S1, S2, Alumni, DPR, beberpa, periode...
110	[Lahh, dulu, pendukung, pak, Anies, mendukung...
111	[posisi, ketum, mungkin, takut, geser, ama, ba...

Gambar 5. Tampilan Data Teks setelah dilakukan Tokenisasid) *Stemming*

Berikut ini merupakan tampilan hasil stemming yaitu menghilangkan imbuhan atau mengubah kata menjadi bentuk dasarnya, seperti yang tersaji pada [gambar 6](#) berikut.

	full_text
0	sembah sujud fufufafa
1	ini macet ato lancar
2	wapresmu suruh ikutdigembleng pak
3	baru hari2 awal kerja udah banyak bener uh yan...
4	ngelindas anak orang nyawa layang cuma 5thun w...

Gambar 6. Tampilan Data Teks setelah dilakukan Stemming**4.3 Labelling Data**

Berikut ini merupakan tampilan hasil *labelling* yang bertujuan untuk menentukan kategori sentimen setiap teks: sentimen positif, negatif, atau netral, yang tersaji pada [gambar 7](#).

```
[22] print(f'Hasil Analisis Data:\nPositif = {total_positif}\nNetral = {total_netral}\nNegatif = {total_negatif}')
      print(f'\nTotal Data : {total}')
```

Hasil Analisis Data:
 Positif = 53
 Netral = 42
 Negatif = 19
 Total Data : 114

	full_text	tweet_english	klasifikasi
0	sembah sujud fufufafa	prostrate fufufafa	Netral
1	ini macet ato lancar	it's stuck ato smoothly	Positif
2	wapresmu suruh ikutdigembleng pak	your vice president told you to take part in d...	Netral
3	baru hari2 awal kerja udah banyak bener uh yan...	just the first days of work have been a lot ri...	Positif
4	ngelindas anak orang nyawa layang cuma 5thun w...	oppressing a child of a kite life is only 5thu...	Netral

Gambar 7. Tampilan Data Teks setelah dilakukan Labelling

4.4 Naïve Bayes

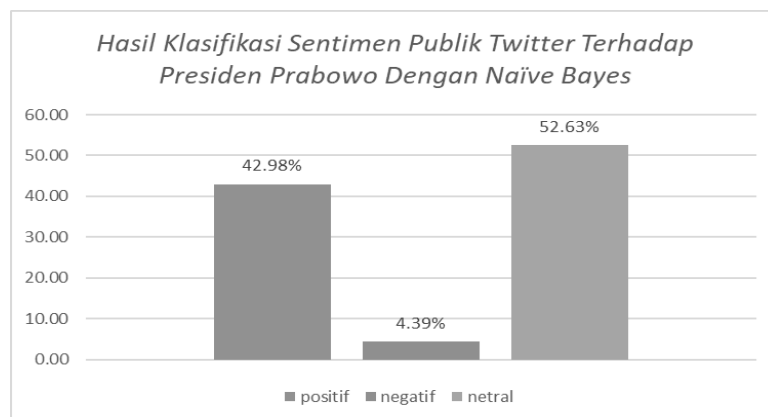
Berikut ini merupakan tampilan hasil klasifikasi teks komentar dengan menggunakan metode machine learning yaitu naïve bayes, untuk menentukan kategori sentimen setiap teks: sentimen positif, negatif, atau netral, yang tersaji pada [gambar 8](#).

	full_text	tweet_english	klasifikasi	klasifikasi_bayes
0	sembah sujud fufufafa	prostrate fufufafa	Netral	Netral
1	ini macet ato lancar	it's stuck ato smoothly	Positif	Netral
2	wapresmu suruh ikutdigembleng pak	your vice president told you to take part in d...	Netral	Netral
3	baru hari2 awal kerja udah banyak bener uh yan...	just the first days of work have been a lot ri...	Positif	Positif
4	ngelindas anak orang nyawa layang cuma 5thun w...	oppressing a child of a kite life is only 5thu...	Netral	Netral
...	...	...	...	...
109	gelar s1 s2 alumni dpr beberapa periode petingg...	degree s1 s2 alumni dpr beberapa period party o...	Netral	Netral
110	lahh dulu dukung pak anies dukung prabowo tapi...	lahh used to support pak anies support prabowo...	Negatif	Negatif
111	posisi tum mungkin takut geser ama bapak nya j...	the position of tum may be afraid to shift his...	Negatif	Positif
112	prabowo niat baik utk kurang gizi buruk lihat ...	prabowo good intentions for malnutrition see j...	Positif	Positif
113	rakyat sehat dulu baru mandiri makan gratis bu...	healthy people first independently eat for fre...	Positif	Positif
114	rows x 4 columns			

Hasil Analisis Data:
Positif = 49
Netral = 60
Negatif = 5

Gambar 8. Tampilan Data Teks setelah dilakukan klasifikasi dengan Naïve Bayes

Dari hasil pengujian klasifikasi sentimen berdasarkan data teks, dengan perbandingan 50% sebagai data latih dan 50% lagi sebagai data uji, menunjukkan bahwa metode *naïve bayes* mampu mengkalsifikasikan sentimen berdasarkan teks dengan cukup baik yaitu dengan akurasi sebesar 75%. Sedangkan hasil berupa frekuensi (jumlah) dari masing-masing sentimen (positif, negatif, netral) divisualisasikan dalam bentuk diagram seperti [gambar 9](#) berikut.



Gambar 9. Hasil klasifikasi sentimen publik twitter terhadap Presiden Prabowo dengan Naïve Bayes

Grafik diatas menunjukkan hasil klasifikasi sentimen publik Twitter terhadap Presiden Prabowo menggunakan metode *Naïve Bayes*. Tiga kategori sentimen yang dianalisis adalah positif, negatif, dan netral. Dari hasil yang ditampilkan, terlihat bahwa mayoritas sentimen publik cenderung netral, dengan 52.63% mengungkapkan sentimen netral. Sementara itu, sentimen positif memperoleh 42.98%, dan hanya 4.39% yang menunjukkan sentimen negatif terhadap Presiden Prabowo.

Dominasi sentimen netral dalam hasil klasifikasi publik Twitter terkait pelantikan Presiden Prabowo dapat dijelaskan melalui beberapa faktor. Pertama, ketidakpastian politik pasca-pelantikan sering kali membuat masyarakat lebih memilih untuk tidak memberikan opini yang tegas, baik positif maupun negatif, dan cenderung menunggu untuk melihat kebijakan-kebijakan yang akan diterapkan oleh pemerintahan baru. Ini mencerminkan sikap “*wait-and-see*” di kalangan masyarakat yang belum memutuskan untuk memberikan penilaian lebih lanjut.

Selain itu, faktor ketidakberpihakan juga berperan, di mana banyak pengguna Twitter yang memilih untuk tidak mengekspresikan afiliasi politik mereka secara terbuka, terutama dalam

peristiwa yang masih sangat baru. Hal ini menyebabkan sebagian besar tweet yang dihasilkan bersifat netral.

Selanjutnya, serta keberagaman pandangan di media sosial, turut menyebabkan banyak orang memilih untuk tetap netral daripada terlibat dalam perdebatan politik yang lebih polarisasi. Masyarakat yang merasa tidak memiliki informasi atau keyakinan yang cukup tentang pemerintahan baru juga cenderung memilih untuk tidak mengambil posisi yang kuat. Dengan demikian, pola dominasi sentimen netral ini menunjukkan bahwa banyak pengguna Twitter lebih memilih untuk menunggu dan melihat perkembangan lebih lanjut, yang mencerminkan ketidakpastian dan sikap hati-hati terhadap perubahan politik yang baru terjadi.

5. KESIMPULAN

Penelitian ini berhasil mengidentifikasi persepsi publik terhadap pelantikan Presiden Prabowo melalui analisis sentimen Twitter menggunakan algoritma *Naïve Bayes*, dengan akurasi klasifikasi mencapai 75%. Hasil analisis menunjukkan bahwa mayoritas sentimen bersifat netral (52,63%), diikuti oleh sentimen positif (42,98%) dan sentimen negatif (4,39%). Dominasi sentimen netral ini mencerminkan kecenderungan publik untuk bersikap hati-hati dan menunggu kinerja nyata Presiden terpilih sebelum memberikan penilaian lebih lanjut, sebuah pola umum dalam respons masyarakat terhadap peristiwa politik besar di Indonesia.

Kontribusi utama penelitian ini terletak pada penerapan pendekatan machine learning berbasis *Naïve Bayes* untuk menganalisis opini publik dalam konteks pelantikan presiden, yang masih jarang dikaji secara spesifik di Indonesia. Selain itu, penelitian ini memperkaya literatur dengan menyajikan gambaran sentimen masyarakat pada momen politik penting secara real-time.

Kedepan, penelitian ini dapat dikembangkan dengan memperluas cakupan data, memperpanjang periode pengumpulan data, serta membandingkan performa beberapa algoritma klasifikasi lain seperti SVM, *Random Forest*, atau metode *Deep Learning*. Penggunaan teknik NLP lanjutan untuk menangani variasi bahasa informal dan bahasa daerah juga diharapkan dapat meningkatkan akurasi dan representativitas hasil analisis.

DAFTAR PUSTAKA

- [1] M. O. Ibrohim and I. Budi, "A Dataset and Preliminaries Study for Abusive Language Detection in Indonesian Social Media," *Procedia Comput. Sci.*, vol. 135, pp. 222-229, 2018, doi: <https://doi.org/10.1016/j.procs.2018.08.169>
- [2] R. Luthfiansyah and B. Wasito, "Analisis Sentimen Terhadap Para Kandidat Presiden 2024 Berdasarkan Netizen Pengguna Twitter Dengan Metode Data Mining Dan Text Mining," *J. Inform. Dan Bisnis*, vol. 11, no. 2, pp. 85-104, Jul. 2023, doi: <https://doi.org/10.46806/jib.v11i2.994>
- [3] V. Pekar, H. Najafi, J. M. Binner, R. Swanson, C. Rickard, and J. Fry, "Voting intentions on social media and political opinion polls," *Gov. Inf. Q.*, vol. 39, no. 4, p. 101658, 2022, doi: <https://doi.org/10.1016/j.giq.2021.101658>
- [4] Mr. Swapnil P. Goje and Dr. Rupali H. Patil, "Sentiment Analysis of Political Parties on social media: A Machine Learning and Lexicon-Based Approach," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 10, no. 5, pp. 233-246, Nov. 2024, doi: <https://doi.org/10.32628/CSEIT24105103>
- [5] E. Kubin and C. Von Sikorski, "The role of (social) media in political polarization: a systematic review," *Ann. Int. Commun. Assoc.*, vol. 45, no. 3, pp. 188-206, Jul. 2021, doi: <https://doi.org/10.1080/23808985.2021.1976070>
- [6] C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J. H. Jeng, and J. G. Hsieh, "Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes," *Inf. Switz.*, vol. 12, no. 5, 2021, doi: <https://doi.org/10.3390/info12050204>
- [7] T. Arifqi, N. Suarna, and W. Prihartono, "Penggunaan Naïve Bayes Dalam Menganalisis Sentimen Ulasan Aplikasi Mcdonald'S Di Indonesia," *JATI J. Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1949-1956, 2024, doi: <https://doi.org/10.36040/jati.v8i2.8740>

- [8] Y. A. Singgalen, "Pemilihan Metode dan Algoritma dalam Analisis Sentimen di Media Sosial : Sistic Literature Review," J. Inf. Syst. Inform., vol. 3, no. 2, pp. 278-302, 2021, doi: <https://doi.org/10.33557/journalisi.v3i2.125>
- [9] I. Septiana and D. Alita, "Perbandingan Random Forest dan SVM dalam Analisis Sentimen Quick Count Pemilu 2024," J. Inform. J. Pengemb. IT, vol. 9, no. 3, pp. 224-233, Dec. 2024, doi: <https://doi.org/10.30591/jpit.v9i3.6640>
- [10] N. Sidauruk, N. Riza, and Rd. N. Siti Fatonah, "Penggunaan Metode Svm Dan Random Forest Untuk Analisis Sentimen Ulasan Pengguna Terhadap Kai Access Di Google Playstore," JATI J. Mhs. Tek. Inform., vol. 7, no. 3, pp. 1901-1906, Nov. 2023, doi: <https://doi.org/10.36040/jati.v7i3.6899>
- [11] A. Saepudin, A. Faqih, and G. Dwilestari, "Perbandingan Algoritma Klasifikasi Support Vector Machine, Random Forest dan Logistic Regression Pada Ulasan Shopee," J. Tekno Kompak, vol. 18, no. 1, p. 178, Feb. 2024, doi: <https://doi.org/10.33365/jtk.v18i1.3764>
- [12] L. Maada, K. Al Fararni, B. Aghoutane, M. Fattah, and Y. Farhaoui, "A comparative study of Sentiment Analysis Machine Learning Approaches," in 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), 2022, pp. 1-5. doi: <https://doi.org/10.1109/IRASET52964.2022.9738346>
- [13] S. Mutmatinah, "Metode Deep Learning LSTM dalam Analisis Sentimen Aplikasi PeduliLindungi," 2024. <https://doi.org/10.34304/scientific.v1i1.231>
- [14] W. Lin and L. Liao, "Lexicon-based prompt for financial dimensional sentiment analysis," Expert Syst. Appl., vol. 244, no. September 2022, p. 122936, 2024, doi: <https://doi.org/10.1016/j.eswa.2023.122936>
- [15] A. H. A. Zharifa and E. I. H. Ujianto, "Analisis Sentimen Publik di Twitter Pasca Debat Kelima Pilpres 2024 dengan Naïve Bayes," Edumatic J. Pendidik. Inform., vol. 8, no. 2, pp. 754-763, Dec. 2024, doi: <https://doi.org/10.29408/edumatic.v8i2.28048>
- [16] L. Abdillah Fudholi, N. Rahaningsih, and R. Danar Dana, "Sentimen Analisis Perilaku Penggemar Coldplay Di Media Sosial Twitter Menggunakan Metode Naïve Bayes," JATI J. Mhs. Tek. Inform., vol. 8, no. 3, pp. 4150-4159, Jun. 2024, doi: <https://doi.org/10.36040/jati.v8i3.9827>
- [17] V. R. Prasetyo et al., "Analisis Sentimen Dan Kategorisasi Review Pelanggan Pada Cafe Kopi Paste Dengan Metode Naïve Bayes Dan K-Nearest Neighbor Sentiment Analysis and Categorization of Customer Reviews on Kopi Paste Cafe Using Naïve Bayes and K-Nearest Neighbor Methods," J. Ilm. NERO, vol. 8, no. 1, p. 2023, 2023. <https://doi.org/10.21107/nero.v8i1.18465>
- [18] Q. A. Xu, V. Chang, and C. Jayne, "A systematic review of social media-based sentiment analysis: Emerging trends and challenges," Decis. Anal. J., vol. 3, no. May, p. 100073, 2022, doi: <https://doi.org/10.1016/j.dajour.2022.100073>
- [19] A. Cahya Kamilla, N. Priyani, R. Priskila, and V. Handrianus Pranatawijaya, "Analisis Sentimen Film Agak Laen Dengan Kecerdasan Buatan: Text Mining Metode Naïve Bayes Classifier," JATI J. Mhs. Tek. Inform., vol. 8, no. 3, pp. 2923-2928, 2024, doi: <https://doi.org/10.36040/jati.v8i3.9587>
- [20] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006. <https://doi.org/10.1007/978-1-4615-7566-5>
- [21] P.-N. Tan, A. Karpatne, M. Steinbach, and V. Kumar, *Introduction to Data Mining Global Edition*. Pearson Deutschland, 2019. [Online]. Available: <https://elibrary.pearson.de/book/99.150005/9780273775324>
- [22] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data," Front. Energy Res., vol. 9, no. March, pp. 1-17, 2021, doi: <https://doi.org/10.3389/fenrg.2021.652801>
- [23] A. E. Hassanien and A. Darwish, Machine Learning and Big Data Analytics Paradigms: Analysis, Applications and Challenges. 2021. doi: <https://doi.org/10.1007/978-3-030-59338-4>
- [24] J. R. Jim, M. A. R. Talukder, P. Malakar, M. M. Kabir, K. Nur, and M. F. Mridha, "Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art

- review,” *Nat. Lang. Process. J.*, vol. 6, no. February, p. 100059, 2024, doi: <https://doi.org/10.1016/j.nlp.2024.100059>
- [25] J. Tao and X. Fang, “Toward multi-label sentiment analysis: a transfer learning based approach,” *J. Big Data*, vol. 7, no. 1, pp. 1-26, 2020, doi: <https://doi.org/10.1186/s40537-019-0278-0>
- [26] D. Jurafsky and J. H. Martin, “Naïve Bayes, Text Classification, and Sentiment,” in *Speech and Language Processing*, 2024, ch. CHAPTER 4.
- [27] P. Nandwani and R. Verma, “A review on sentiment analysis and emotion detection from text,” *Soc. Netw. Anal. Min.*, vol. 11, no. 1, pp. 1-19, 2021, doi: <https://doi.org/10.1007/s13278-021-00776-6>
- [28] I. E. Syahputra, T. Tulus, and S. Efendi, “Indonesian Text Dataset for Determining Sentiment Classification Using Mechine Learning Approach,” *J. Inform. Telecommun. Eng.*, vol. 3, no. 2, pp. 192-201, 2020, doi: <https://doi.org/10.31289/jite.v3i2.3153>
- [29] S. Iyengar and S. J. Westwood, “Fear and Loathing across Party Lines: New Evidence on Group Polarization,” *Am. J. Polit. Sci.*, vol. 59, no. 3, pp. 690-707, 2015. <https://doi.org/10.1111/ajps.12152>
- [30] Y. Kim, “The contribution of social network sites to exposure to political difference: The relationships among SNSs, online political messaging, and exposure to cross-cutting perspectives,” *Comput. Hum. Behav.*, vol. 27, no. 2, pp. 971-977, 2011, doi: <https://doi.org/10.1016/j.chb.2010.12.001>