

Sentiment Analysis of Social Media Users On The 2024 Presidential Election Using the Naive Bayes Classifier Method

Ardhi Wahyu Nugroho¹, Rumini^{2*}, Tri Susanto³

^{1,2,3} Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta, Indonesia

Informasi Artikel

Diterima : 8 April 2025

Revisi : 3 Mei 2025

Publikasi : 20 Juni 2025

Kata Kunci:

Analisis Sentimen

Pemilihan Presiden 2024

Naive Bayes Classifier

Bag of Words (BoW)

Media Sosial

ABSTRAK

Dengan semakin berkembangnya penggunaan media sosial sebagai sarana komunikasi publik, banyak opini, komentar, dan sentimen yang disampaikan secara terbuka. Namun, volume data yang besar serta penggunaan bahasa informal di platform ini menimbulkan kesulitan dalam memahami sentimen secara efektif. Dampak dari permasalahan ini adalah kurangnya pemahaman yang akurat tentang preferensi politik masyarakat, yang dapat memengaruhi strategi kampanye dan kebijakan publik. Oleh karena itu, diperlukan sebuah metode untuk mengklasifikasikan sentimen secara lebih efektif dari data yang tidak terstruktur. Untuk menyelesaikan masalah tersebut, penelitian ini menggunakan algoritma *Naive Bayes Classifier* untuk mengklasifikasikan sentimen tweet menjadi tiga kategori: positif, netral, dan negatif. Dengan menggunakan dataset yang terdiri dari 1.945 data hasil crawling dan melalui representasi teks menggunakan *Bag of Words* (BoW) serta TF-IDF, penelitian ini menunjukkan bahwa algoritma Naive Bayes berhasil mengklasifikasikan sentimen positif, netral, dan negatif secara efektif, terutama pada ulasan negatif dan netral. Model Naive Bayes dengan BoW terbukti memiliki akurasi keseluruhan sebesar 90,15%, dengan keseimbangan yang lebih baik antara *precision* dan *recall* dibandingkan model dengan TF-IDF.

ABSTRACT

With the increasing use of social media as a means of public communication, many opinions, comments, and sentiments are expressed openly. However, the large volume of data and the use of informal language on this platform make it difficult to understand sentiment effectively. The impact of this problem is the lack of accurate understanding of people's political preferences, which can affect campaign strategies and public policies. Therefore, a method is needed to classify sentiment more effectively from unstructured data. To solve this problem, this study uses the Naive Bayes Classifier algorithm to classify tweet sentiment into three categories: positive, neutral, and negative. Using a dataset consisting of 1,945 crawled data and through text representation using Bag of Words (BoW) and TF-IDF, this study shows that the Naive Bayes algorithm successfully classifies positive, neutral, and negative sentiments effectively, especially in negative and neutral reviews. The Naive Bayes model with BoW is proven to have an overall accuracy of 90.15%, with a better balance between precision and recall than the model with TF-IDF.

This is an open-access article under the [CC BY-SA](#) license



*Penulis Koresponden

Email: rumini@amikom.ac.id

Cara sitasi IEEE::

A. W. Nugroho, Rumini, & T. Susanto, "Sentiment Analysis of Social Media Users On The 2024 Presidential Election Using the Naive Bayes Classifier Method" *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 5, no. 2, pp. 410-420, Juni 2025. doi: 10.30811/jaise.v5i2.6679

1. PENDAHULUAN

Jumlah pengguna media sosial terus meningkat seiring dengan kemajuan teknologi informasi yang pesat. Pengguna internet kini dapat dengan bebas berbagi informasi, yang kemudian membentuk opini publik yang beragam. Opini ini dianggap sebagai data yang dapat dipercaya karena berasal dari ekspresi emosi, pendapat, atau cerita yang sedang viral di media sosial, seperti Twitter, Instagram, dan Facebook, yang sering digunakan oleh pengguna internet untuk berbagi pengalaman. Selain pengguna individu, partai politik dan perusahaan juga dapat memanfaatkan data ini sebagai sumber informasi penting tentang respons masyarakat, mengukur popularitas, dan menilai kualitas layanan mereka. Namun, mengukur sentimen komentar di media sosial bukanlah hal yang mudah. Salah satu kendalanya adalah penggunaan bahasa informal yang seringkali sulit untuk diinterpretasikan dan digunakan dalam pengambilan keputusan. Media sosial saat ini mencakup berbagai platform online, seperti situs jejaring sosial, aplikasi jejaring sosial, game, dan media online lainnya. Beberapa contoh media sosial yang populer adalah Path, Line, Instagram, Whatsapp, Blackberry Messenger, Twitter, LinkedIn, Facebook, Wikipedia, Google Plus, dan YouTube. Konten yang disajikan dalam platform-platform tersebut meliputi berita politik, layanan sosial, ekonomi, olahraga, kesehatan, dan lain-lain. Namun, tidak semua informasi yang disampaikan bersifat positif, ada juga yang bersifat negatif[1].

Twitter saat ini merupakan salah satu alat komunikasi yang sangat populer di dunia maya dan sering digunakan oleh calon kandidat presiden terpilih sebagai sarana kampanye untuk memperkenalkan citra masing-masing kepada calon pemilih dan pendukungnya. Pada pemilihan umum di Asia pada tahun 2019, para kandidat menggunakan Twitter dan platform media sosial lainnya untuk membagikan slogan dan kebijakan mereka, yang mengguncang popularitas lawan-lawan mereka dan mendapatkan dukungan sebelum kampanye dimulai[2].

Semakin meluasnya penggunaan internet diperlihatkan oleh banyaknya platform yang menyediakan berbagai layanan untuk masyarakat mengaksesnya. Salah satu platform yang populer adalah media sosial, yang menjadi tempat komunikasi dan akses informasi yang cepat. Beragam informasi seperti komentar, kritik, dan opini dapat ditemukan di media sosial, termasuk di Twitter yang digunakan oleh banyak orang di dunia. Menurut Websitaring, jumlah pengguna Twitter di dunia mencapai 1.3 miliar akun Twitter, tetapi hanya 237.8 juta adalah pengguna aktif. Data ini menunjukkan bahwa Twitter adalah sumber sentimen masyarakat di dunia yang signifikan dan dapat dijadikan studi kasus[3]. Indonesia menjadi negara keempat terbesar sebagai pengguna X (Twitter) mencapai 24,69 juta pengguna Twitter pada awal tahun 2024[4].

Saat ini, Twitter banyak digunakan oleh masyarakat untuk berbagai keperluan, termasuk sebagai media sosial untuk mendapatkan informasi tentang pemerintahan, berdiskusi, dan menyampaikan pendapat. Indonesia, sebagai negara demokrasi, mengatur pemilihan umum dalam Undang-Undang Dasar Negara Republik Indonesia 1945 Nomor 23 Tahun 2003 Pasal 3, di mana pemilihan umum untuk anggota DPR, DPD, dan DPRD, serta presiden dan wakil presiden, diselenggarakan setiap lima tahun[5].

Pemilihan Presiden (Pilpres) yang akan berlangsung pada tahun 2024 sudah mulai dirasakan saat ini, terutama di media sosial, di mana orang-orang menyampaikan pandangan, sentimen, dan preferensi politik mereka. Salah satu topik hangat yang sedang dibahas adalah Pilpres 2024, yang banyak dibicarakan di Twitter melalui berbagai postingan opini dari masyarakat. Jumlah tweet yang sangat besar mengenai hal ini telah menimbulkan beragam pandangan, baik positif, netral maupun negatif. Data komentar pada sebuah tweet yang dikumpulkan tentang pembahasan ini bisa menjadi sumber data yang berharga untuk diolah atau dianalisis sesuai kebutuhan[6].

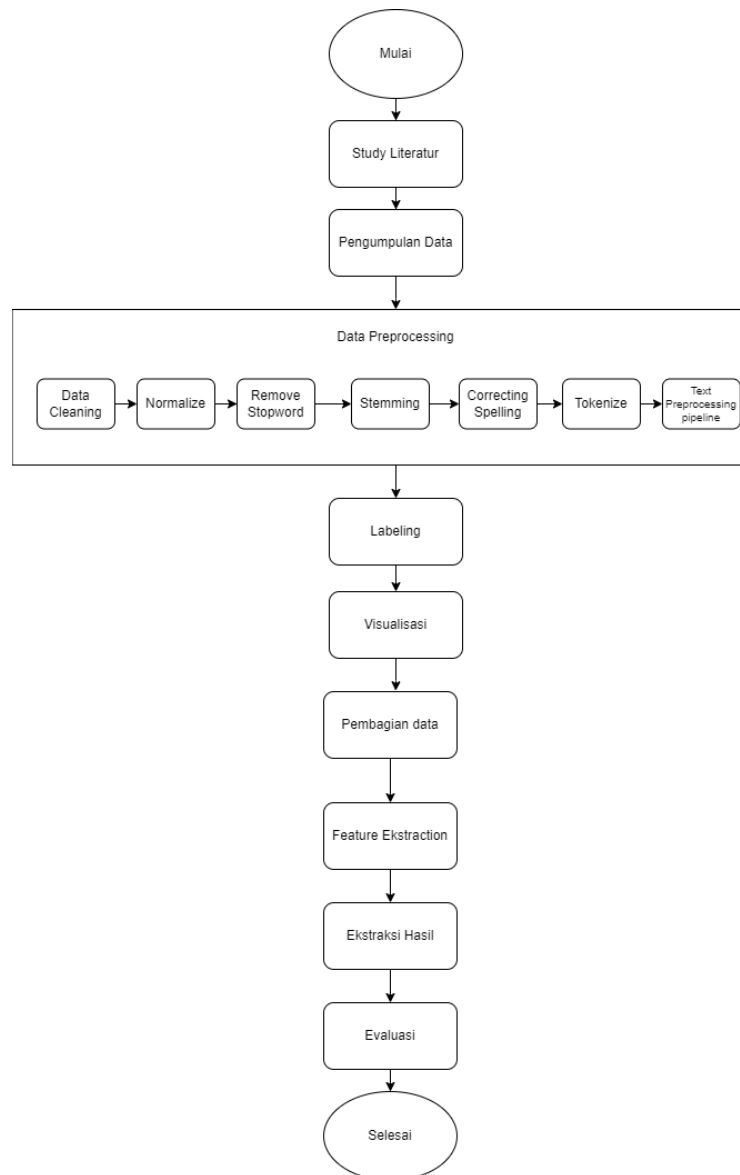
Penelitian yang berjudul Komparasi Algoritma Naive Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Pengguna Busway. Pada penelitian ini membahas Kemacetan di kota-kota besar di Indonesia semakin parah akibat peningkatan penggunaan kendaraan pribadi. Salah satu solusi yang diusulkan adalah penggunaan transportasi publik seperti busway atau TransJakarta. Analisis sentimen digunakan untuk mengidentifikasi dan mengklasifikasikan opini masyarakat tentang busway, dengan data yang dikumpulkan dari media sosial dan situs web. Penelitian ini menggunakan framework gataframework untuk preprocessing data, yang kemudian dianalisis dengan algoritma Naive Bayes dan optimasi fitur menggunakan algoritma

genetika. Kombinasi ini menghasilkan akurasi klasifikasi sebesar 88,55% dan nilai AUC sebesar 0,813%. Hasil ini menunjukkan bahwa metode ini efektif untuk analisis sentimen pengguna busway. Diharapkan, analisis sentimen ini dapat membantu meningkatkan fasilitas TransJakarta dan mendorong lebih banyak orang untuk beralih dari kendaraan pribadi ke transportasi umum, sehingga mengurangi kemacetan[7]. penelitian tersebut dilakukan analisis sentimen terhadap pengguna Busway menggunakan algoritma Naive Bayes berbasis Algoritma Genetika perbedaan penelitian adalah menggunakan algoritma naïve bayes berbasis metode *Bag of Words (Bow)* dan TF-IDF (*term frequency-inverse document frequency*).

Dalam analisis sentimen ini, digunakan algoritma Naive Bayes. Algoritma Naive Bayes adalah sebuah algoritma klasifikasi yang mencari nilai probabilitas tertinggi dengan menerapkan teorema Bayes. Algoritma ini memiliki keunggulan karena sederhana dalam komputasi dan mampu menangani data dalam jumlah besar[8]. Dalam penelitian ini, kelebihan algoritma Naive Bayes digunakan untuk mengklasifikasikan teks data tweet ke dalam kategori sentimen positif, netral, dan negatif. Hal ini bertujuan untuk mengetahui tingkat akurasi, presisi, dan recall dari algoritma tersebut dalam konteks Pemilihan Presiden 2024.

2. METODE

langkah-langkah penelitian yang divisualisasikan dengan gambar 1 Alur Penelitian.



Gambar 1. Alur Penelitian

Setiap tahapan penelitian pada alur penelitian adalah sebagai berikut

2.1 Studi literatur Tahap studi literatur

Melakukan pendalaman materi tentang Analisis sentimen, bagian dari pemrosesan bahasa alami (NLP), menentukan sikap atau opini melalui teks, dengan tahapan seperti pengumpulan data, pra-pemrosesan, representasi teks, pelabelan data, pelatihan model, evaluasi, dan prediksi sentimen. Twitter, sebagai platform yang sering digunakan untuk berbagi opini, memiliki karakteristik seperti panjang teks terbatas, penggunaan bahasa tidak baku, serta penggunaan hashtag dan mention. Naive Bayes Classifier, algoritma sederhana namun efektif, bekerja berdasarkan Teorema Bayes dengan asumsi kemandirian antara fitur. Algoritma ini cocok untuk klasifikasi teks dengan varian seperti Multinomial, Bernoulli, dan Gaussian Naive Bayes. Studi ini memberikan dasar teori dan metodologi untuk menghasilkan model yang efektif dalam mengklasifikasikan sentimen tweet terkait pemilihan presiden

2.2 Pengumpulan Data

mengumpulkan dan menganalisis data komentar Twitter terkait pemilihan presiden 2024. Kata kunci yang digunakan untuk pengumpulan data adalah "pemilihan presiden 2024". Data dikumpulkan menggunakan API Twitter berdasarkan kata kunci yang telah ditentukan. Pengambilan data dilakukan pada tanggal 22 Mei 2024 dan menghasilkan kurang lebih 2000 data komentar terkait pemilihan presiden 2024.

2.3 Data Processing

Berikut adalah proses preprocessing data yang dilakukan :

1. **Data Cleaning**
Proses pembersihan ini penting untuk menghilangkan elemen yang dapat mengganggu analisis lebih lanjut, menjadikan teks lebih konsisten dan siap untuk diproses
2. **Normalize**
Pada tahap normalisasi dalam preprocessing teks yang bertujuan untuk mengubah kata-kata informal atau slang menjadi bentuk standarnya.
3. **Remove Stopword**
Pada tahap ini, peneliti memanfaatkan library NLTK untuk menghilangkan stopwords dari teks dalam dataset.
4. **Stemming**
Pada tahap *stemming*, peneliti menggunakan library Sastrawi untuk mengubah kata-kata dalam teks menjadi bentuk dasarnya (*stem*).
5. **Correcting Spelling**
Pada tahap ini, peneliti menggunakan library SymSpell untuk melakukan koreksi ejaan pada teks.
6. **Tokenization**
Pada tahap ini, peneliti menggunakan Transformers dari library Hugging Face untuk melakukan analisis sentimen dengan model pre-trained khusus untuk bahasa Indonesia.
7. **Text Preprocessing Pipeline**
Text Preprocessing Pipeline memungkinkan penggabungan beberapa langkah menjadi satu objek terintegrasi, sehingga data yang telah melalui proses preprocessing dapat disimpan dalam file CSV.

2.4 Labeling

Proses penelitian labeling dimulai setelah proses *data preprocessing* selesai. Data kemudian akan dianalisis untuk diberi label berdasarkan klasifikasi positif, negatif, atau netral, sesuai dengan pendapat atau opini yang terkandung di dalamnya.

2.5 Visualisasi

Setelah proses labeling selesai, langkah selanjutnya dalam penelitian ini adalah visualisasi. Langkah pertama adalah mengimpor dua kelas penting dari pustaka wordcloud.

2.6 Pembagian Data,

Pada bagian ini, penulis membagi data (splitting data) menjadi set pelatihan dan set pengujian.

2.7 Feature Ekstraktion

Feature extraction adalah langkah penting dalam mempersiapkan data untuk analisis atau pemodelan machine learning. Dalam penelitian ini, metode Bag of Words (BoW) dan Term Frequency-Inverse Document Frequency (TF-IDF) digunakan untuk mengubah teks komentar terkait pemilihan presiden 2024 menjadi bentuk yang dapat digunakan oleh model machine learning.

2.8 Ekstraksi Hasil

Ekstraksi fitur menggunakan dua pendekatan utama, yaitu Bag of Words (BoW) dan Term Frequency-Inverse Document Frequency (TF-IDF).

2.9 Evaluasi

langkah evaluasi dilakukan untuk mengukur kinerja model dalam menganalisis sentimen terkait pemilihan presiden 2024.

representasi hasil dari proses klasifikasi, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN).

Akurasi (Accuracy):

Mengukur sejauh mana model dapat mengklasifikasikan semua kelas dengan benar

$$akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

presisi(Precision):

Precision untuk setiap kelas dihitung dengan rumus berikut:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall:

Recall untuk setiap kelas dihitung menggunakan

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1-Score:

F1-Score adalah harmoni dari precision dan recall, dihitung dengan rumus berikut

$$F1 - Score = 2 \frac{precision \cdot Recall}{Precision + Recall} \quad (4)$$

Keterangan :

TP: True Positive (jumlah prediksi benar untuk kelas positif)

TN: True Negative (jumlah prediksi benar untuk kelas negatif)

FP: False Positive (jumlah prediksi salah untuk kelas positif)

FN: False Negative (jumlah prediksi salah untuk kelas negatif)

3. HASIL DAN PEMBAHASAN

Data penelitian diambil dari Twitter melalui metode crawling pada tanggal 22 Mei 2024, dengan total dataset yang terkumpul sebanyak 2065. Dan hasil dari semua penelitian yang telah dilakukan adalah sebagai berikut :

3.1 Dataset Crawling

```

This script uses Chromium Browser to crawl data from Twitter with your Twitter auth token.
Please enter your Twitter auth token when prompted.

Note: Keep your access token secret! Don't share it with anyone else.
Note: This script only runs on your local device.

Opening twitter search page...

Created new directory: /content/tweets-data

Your tweets saved to: /content/tweets-data/datasetskripsil.csv
Total tweets saved: 10

-- Scrolling... (1) (2) (3)

Your tweets saved to: /content/tweets-data/datasetskripsil.csv
Total tweets saved: 19

```

Gambar 2. Proses Crawl data

Pada gambar 2 tersebut adalah hasil dari menjalankan sebuah skrip yang bertujuan untuk mendapatkan data dari twitter menggunakan *browser crome* menggunakan *google collab*.

3.2 Preprocessing Data

Pada tahap selanjutnya adalah melakukan preprocessing, preprocessing bertujuan untuk mengolah data menjadi efektif untuk digunakan.

```

0      kita dukung siapa et tergantung komando
1      pilih siapa
2      resmiiii gaess
3      bjirr rill cekk
4      mudryk onana ada ga sih eti
      ...
2060      pesimisti
2061      di repli pada marah marah hahahhaha ini perkat...
2062      betul tapi kalau negara diuru orang yg tepat p...
2063      bener lah orang penguasa juga rata rata berbis...
2064      setuju
Name: full_text, Length: 2065, dtype: object

```

Gambar 3. Preprocessing Data dan penghapusan kolom

Pada gambar 3 tersebut merupakan hasil dari menjalankan sebuah skrip yang bertujuan untuk melakukan praproses teks pada data. Praproses ini melibatkan beberapa langkah penting: pertama, teks dibersihkan dari URL, mention, dan karakter khusus menggunakan ekspresi reguler (re.sub). Selanjutnya, teks diubah menjadi huruf kecil untuk menghilangkan perbedaan kapitalisasi. Kemudian, teks dipecah menjadi token menggunakan fungsi tokenisasi dari NLTK (nltk.word_tokenize).

3.3 Labeling

Proses penelitian labeling dimulai setelah proses data preprocessing selesai. Data kemudian akan dianalisis untuk diberi label berdasarkan klasifikasi positif, negatif, atau netral, sesuai dengan pendapat atau opini yang terkandung di dalamnya. Proses labeling dilakukan menggunakan Indonesia-BERT-Sentiment-Classification untuk otomatisasi, di mana sistem ini akan menggunakan teknik pemrosesan bahasa alami untuk memahami dan mengklasifikasikan konten teks. Setelah diberi label oleh Indonesia-BERT-Sentiment-Classification, dilakukan pemeriksaan dan verifikasi ulang untuk memastikan konsistensi dan objektivitas hasil labeling. Pada gambar 4 adalah hasil dari proses labeling yang menggunakan '*mdhugol/indonesia-bert-sentiment-classification*', yaitu model ini tersedia di platform '*Hugging Face*', untuk '*indonesia-bert-sentiment-classification*' mengeluarkan label prediksi dalam bentuk kode numerik atau indeks label, yaitu LABEL_0, LABEL_1, dan LABEL_2. Namun, angka atau kode ini belum menyampaikan makna yang jelas bagi pengguna dibuatlah '*label_index*' yang memetakan setiap label model ke deskripsi sentimen yang lebih bermakna label_0 dipetakan menjadi positive, label_1 dipetakan menjadi neutral, label_2 dipetakan menjadi negative seperti yang terlihat pada hasil gambar 4.14 yang dimana dari 1952 data tweet yang dianalisis terdapat 561 data bersifat positive, 415 data bersifat neutral, 976 data bersifat negative.


```

# Preproses data
def preprocess_text(text):
    # Menghapus tanda baca, angka, dan karakter khusus
    text = re.sub(r'[^\w-zA-Z\s]', '', text)

    # Konversi ke huruf kecil
    text = text.lower()
    return text

# Preproses data
preprocessed_data = [preprocess_text(text) for text in my_df['cleaned_text']]
print("Sample preprocessed data:", preprocessed_data[:5]) # Print first 5 elements

# Split data into training and testing sets (70% training, 30% testing)
# X, X_test, y, y_test = train_test_split(preprocessed_data, my_df['sentiment'], test_size=0.3, random_state=42)

from sklearn.preprocessing import LabelEncoder

X = my_df['cleaned_text']
y = my_df['sentiment']

# Label Encode
encoder = LabelEncoder()

# Fitting encoder
y = encoder.fit_transform(y)
print(y)

# Splitting Data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

```

Gambar 6 hasil proses pembagian Dataset

Setelah proses visualisasi selesai, langkah selanjutnya dalam penelitian ini adalah membagi data (splitting data) menjadi set pelatihan dan set pengujian. Proses ini dimulai dengan memisahkan data yang telah diberi label menjadi fitur (X) dan label (y). Fitur (X) adalah teks komentar yang telah dibersihkan dan diproses, sedangkan label (y) adalah klasifikasi yang diberikan (positif, negatif, atau netral). Setelah pemisahan data, set pelatihan (X_train dan y_train) digunakan untuk melatih model, sementara set pengujian (X_test dan y_test) digunakan untuk mengevaluasi kinerja model. Pembagian data yang dilakukan secara acak namun terkontrol oleh parameter `random_state` memastikan bahwa eksperimen dapat diulang dengan hasil yang sama, sehingga menjaga validitas dan reliabilitas penelitian.

3.6 Feature Extraction

Feature extraction adalah langkah penting dalam mempersiapkan data untuk analisis atau pemodelan machine learning. Dalam penelitian ini, metode *Bag of Words* (BoW) dan *Term Frequency-Inverse Document Frequency* (TF-IDF) digunakan untuk mengubah teks komentar terkait pemilihan presiden 2024 menjadi bentuk yang dapat digunakan oleh model machine learning. Setelah data komentar terkait pemilihan presiden 2024 dikumpulkan dan diberi label menggunakan Indonesia-BERT-Sentiment-Classification, data tersebut dibagi menjadi set pelatihan dan set pengujian.

3.6.1 Bag of Words (BOW)

Pada gambar 7 adalah proses untuk mengubah teks menjadi representasi vektor numerik berdasarkan frekuensi kemunculan kata-kata dalam teks. Pertama, `CountVectorizer` diinisialisasi dengan tipe data integer, yang kemudian digunakan untuk membangun kosakata dari data teks dengan cara mempelajari semua kata unik di seluruh data. Setelah itu, fungsi `fit_transform()` mengubah setiap kalimat atau dokumen dalam data teks (X) menjadi vektor numerik di mana setiap nilai mewakili frekuensi kemunculan kata dalam dokumen tersebut. Hasilnya adalah matriks sparse, dengan baris mewakili dokumen dan kolom mewakili kata-kata unik dari kosakata. Matriks ini kemudian dapat digunakan sebagai input fitur untuk melatih model seperti Naive Bayes pada tahap selanjutnya.

```

# Bag of Words
vectorizer = CountVectorizer()
X_bow = vectorizer.fit_transform(X)

# Naive Bayes dengan BOW
vectorizer = CountVectorizer(dtype='int')
X_bow = vectorizer.fit_transform(X)

```

Gambar 7. proses Feature Engineering menggunakan Bag of Words(BOW)

3.6.2 TF-IDF

Pada gambar 8 adalah proses untuk mengubah teks menjadi representasi vektor numerik yang memperhitungkan frekuensi kemunculan kata dalam dokumen serta seberapa umum atau jarang kata tersebut di seluruh dokumen. TfidfVectorizer mempelajari kosakata dari data teks dan menghitung bobot TF-IDF untuk setiap kata. Hasilnya adalah matriks sparse yang berisi bobot TF-IDF untuk setiap dokumen, yang mencerminkan pentingnya kata tersebut dalam konteks dokumen tersebut. Matriks ini kemudian dapat digunakan sebagai input untuk melatih model Naive Bayes.

```
[ ] # TF-IDF
# tfidf_vectorizer = TfidfVectorizer()
# X_tfidf = tfidf_vectorizer.fit_transform(X)

# Naive Bayes Dengan TFIDF
tfidf_vectorizer = TfidfVectorizer()
X_tfidf = tfidf_vectorizer.fit_transform(X)
```

Gambar 8. proses Feature Engineering TF-IDF

3.7 Hasil

Pada gambar 9 adalah Hasil evaluasi dua model Naive Bayes yaitu BOW dan TF-IDF, dapat disimpulkan bahwa model dengan BoW menunjukkan performa yang lebih baik secara keseluruhan dengan akurasi 88,56%, dibandingkan dengan model TF-IDF yang memiliki akurasi 80,73%. Pada model BoW, kategori positif (kelas 0) memiliki precision 85% dan recall 97%, yang menunjukkan kemampuan model untuk mendeteksi dan memprediksi ulasan positif secara akurat. Sementara itu, pada kategori netral (kelas 1), precision mencapai 96% tetapi recall hanya 75%, menandakan bahwa meskipun model sangat akurat ketika memprediksi ulasan netral, ia cenderung melewatkan beberapa data yang seharusnya masuk kategori ini. Untuk kategori negatif (kelas 2), model juga tampil baik dengan precision 91% dan recall 84%. Sebaliknya, model dengan TF-IDF mengalami kesulitan terutama pada kelas netral, di mana precision-nya sempurna di 100%, tetapi recall sangat rendah pada 43%, menunjukkan bahwa model sering melewatkan ulasan netral. Secara keseluruhan, model dengan BoW lebih unggul dalam mengklasifikasikan ulasan positif, netral, dan negatif dibandingkan dengan model TF-IDF, terutama dalam hal keseimbangan antara precision dan recall.

Akurasi model Naive Bayes dengan Bag of Words: 0.8856259659969088					
	precision	recall	f1-score	support	
0	0.85	0.97	0.91	976	
1	0.96	0.75	0.84	415	
2	0.91	0.84	0.87	550	
accuracy			0.89	1941	
macro avg	0.91	0.85	0.87	1941	
weighted avg	0.89	0.89	0.88	1941	
Akurasi model Naive Bayes dengan TF-IDF: 0.8073158165893869					
	precision	recall	f1-score	support	
0	0.74	1.00	0.85	976	
1	1.00	0.43	0.60	415	
2	0.95	0.76	0.84	550	
accuracy			0.81	1941	
macro avg	0.89	0.73	0.76	1941	
weighted avg	0.85	0.81	0.79	1941	

Gambar 9 Hasil evaluasi dua model Naive Bayes yaitu BOW dan TF-IDF

Dengan akurasi keseluruhan sebesar 88,56%, model BoW mampu memberikan keseimbangan yang lebih baik antara precision dan recall pada ketiga kelas (positif, netral, dan negatif), terutama dalam mendeteksi ulasan netral dan negatif. Sebaliknya, meskipun model dengan TF-IDF memiliki precision yang lebih tinggi pada kelas netral, recall yang rendah (hanya 43%) menyebabkan banyak ulasan netral tidak terdeteksi dengan benar. Secara keseluruhan, BoW terbukti lebih efektif sebagai representasi teks dalam model Naive Bayes untuk tugas klasifikasi sentimen ini.

4. KESIMPULAN

Dalam penelitian ini, analisis sentimen media sosial twitter dilakukan dengan menggunakan dataset yang diambil dari tweet beberapa postingan yang diambil menggunakan Teknik crawling tweet-harvest versi

2.2.8, terdiri dari 2065 data. Setelah dilakukan pre-processing dan labeling secara otomatis ada 1952 data yang didapatkan, dari data tersebut terdapat 561 data positive, 415 data neutral, 976 data negative. Penggunaan Bag of Words dan TF-IDF, serta penerapan model klasifikasi Naive Bayes pada data training sebesar 70% dan data test sebesar 30%, hasil evaluasi menunjukkan kinerja yang sangat baik. model Naive Bayes dengan representasi Bag of Words (BoW) memiliki performa yang lebih baik dibandingkan model dengan TF-IDF dalam mengklasifikasikan ulasan berdasarkan sentimen. Dengan akurasi keseluruhan sebesar 88,56%, model BoW mampu memberikan keseimbangan yang lebih baik antara precision dan recall pada ketiga kelas (positif, netral, dan negatif), terutama dalam mendeteksi ulasan netral dan negatif. Sebaliknya, meskipun model dengan TF-IDF memiliki precision yang lebih tinggi pada kelas netral, recall yang rendah (hanya 43%) menyebabkan banyak ulasan netral tidak terdeteksi dengan benar. Secara keseluruhan, BoW terbukti lebih efektif sebagai representasi teks dalam model Naive Bayes untuk tugas klasifikasi sentimen ini. Hasil penelitian ini dapat memberikan kontribusi yang berarti dalam upaya meningkatkan keamanan dan kenyamanan dalam berinteraksi di media sosial, terutama dalam mendeteksi potensi polaritas opini menjelang pemilihan presiden kedepannya. Dengan demikian, algoritma Naive Bayes terbukti menjadi alat yang sangat efektif dalam analisis sentimen komentar di Twitter dan memiliki potensi untuk aplikasi lebih lanjut dalam pengembangan solusi terkait pemantauan dinamika opini publik. Hal ini diharapkan dapat membantu meningkatkan kesadaran akan pentingnya menjaga etika dan keamanan dalam bersosial media, terutama di platform seperti Twitter selama masa kampanye politik.

REFERENSI

- [1] Fauzi, A., Akbar, M. F., and Asmawan, Y. 2019. "Sentimen Analisis Berinternet Pada Media Sosial dengan Menggunakan Algoritma Bayes," *J. Inform.*, vol. 6, no. 1, pp. 77–83, doi: 10.31311/ji.v6i1.5437.
- [2] Kurniawan, I. and Susanto, A. 2019. "Implementasi Metode K-Means dan Naive Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019," *Eksplora Inform.*, vol. 9, no. 1, pp. 1–10, 2019, doi: 10.30864/eksplora.v9i1.237.
- [3] Ahlgren, Matt. 2024. "Statistik & Tren Twitter (X) [Pembaruan 2024]. <https://www.websiterating.com/id/research/twitter-statistics/>
- [4] S. Kemp., 2024. "Digital 2024: Indonesia," *datareportal*, <https://datareportal.com/reports/digital-2024-indonesia>
- [5] Hapsari, Y. D. P. and Saraswati, R., 2023. "Dampak Pelaksanaan Presidential Threshold pada Pemilu Serentak terhadap Demokrasi di Indonesia," *J. Pembang. Huk. Indones.*, vol. 5, no. 1, pp. 70–84, doi: 10.14710/jphi.v5i1.70-84.
- [6] Khatami, M. I. 2021. "The Existence Of ' Baliho 2024 ' In The Electricity Fight : Image Analysis Of Political Figures," *J. Ilmu Sos. dan Ilmu Polit. Univ. Jambi*, vol. 5, no. 2, pp. 14–24,
- [7] Aryanti, R. Saepudin, A. Eka, F. Permana, R. and Firmansyah, D. 2019. "Komparasi Algoritma Naive Bayes Dengan Algoritma genetika Pada Analisis Sentimen Pengguna Busway, *Jurnal Teknik Komputer Amik BSI Vol.5, No. 2.*
- [8] Ariansyah, A. and Kusmira, M. 2021. "Analisis Sentimen Pengaruh Pembelajaran Daring Terhadap Motivasi Belajar Di Masa Pandemi Menggunakan Naive Bayes Dan Svm," *Fakt. Exacta*, vol. 14, no. 3, p. 100, doi: 10.30998/faktorexacta.v14i3.10325.
- [9] Septianingrum, F. and Irawan, A. S. Y.. 2021. "Metode Seleksi Fitur Untuk Klasifikasi Sentimen Menggunakan Algoritma Naive Bayes: Sebuah Literature Review," *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 799, doi: 10.30865/mib.v5i3.2983.
- [10] A. Taufik et al., 2021. "Klasifikasi Tweet Influencer Nu Dengan Gnpf-Ulama Menggunakan Naive Bayes Dan Support Vector Machine Classification Of Nu Influencer Tweet With Gnpf-Ulama Populisme Islam belakangan ini merupakan salah satu topik sentral dalam kajian politik di Indonesia," vol. 5, pp. 258–276.
- [11] Normah, B. Rifai, Vambudi, S. and Maulana, R. 2022. "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, doi: 10.31294/jtk.v4i2.
- [12] Juanita, S. 2020. "Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 552, doi: 10.30865/mib.v4i3.2140.
- [13] Alfarizi, S. and Fitriani, E. 2023. "Analisis Sentimen Kendaraan Listrik Menggunakan Algoritma Naive Bayes dengan Seleksi Fitur Information Gain dan Particle Swarm Optimization," *Indones. J. Softw. Eng.*, vol. 9, no. 1, pp. 19–27, [Online]. Available: <http://ejournal.bsi.ac.id/ejournal/index.php/ijse>
- [14] Hayuningtyas, R. Y. and R. Sari, 2019. "Analisis Sentimen Opini Publik Bahasa Indonesia Terhadap Wisata Tmii Menggunakan Naive Bayes Dan Pso," *J. Techno Nusa Mandiri*, vol. 16, no. 1, pp. 37–42, doi: 10.33480/techno.v16i1.115.
- [15] M. U. Albab, Y. K. P, and M. N. Fawaiq, .2023. "Optimization of the Stemming Technique on Text preprocessing President 3 Periods Topic," vol. 20, no. 2, pp. 1–10,
- [16] Nardilasari, A. Hananto, S. Hilabi, S. Tukino, T. and Priyatna, B. 2023, "Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM Pada Media Sosial Twitter," *JOINTECS (Journal Inf. Technol. Comput. Sci.*, vol. 8, no. 1, p. 11, doi: 10.31328/jointecs.v8i1.4265.
- [17] Amrullah, A. Z. Sofyan Anas, A. and Hidayat, M. A. J. 2020, "Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square," *Jurnal*, vol. 2, no. 1, pp. 40–44, doi: 10.30812/bite.v2i1.804.
- [18] F. T. I. Uksw, "Implementasi Web Crawling Untuk Pencarian Harga Sparepart Pada PT Asuransi Sinar Mas," vol. 7, no. 3, pp. 416–428, 2020.
- [19] Firdaus A. and Firdaus, W. I. 2023. "Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan)," 2021.
- [20] Elza Putra, A. "Analisis Sentimen Tanggapan Masyarakat kepada Video Youtube mengenai Resesi 2023 menggunakan Metode Naive Bayes Dan K-Nearest Neighbor (KNN)," pp. 1–23.
- [21] Adnan, M. Wardhani, N. and Author, C. 2022, "Optimasi Normalisasi Kata Pada Data Twitter Untuk Meningkatkan Akurasi Analisis Sentimen (Studi Kasus Respon Masyarakat Terhadap Layanan Teman Bus)," *J. Fokus Elektroda Energi List. Telekomun. Komputer, Elektron. dan Kendali*, vol. 7, no. 4, pp. 237–243, [Online]. Available: <https://elektroda.uho.ac.id/index.php/journal/article/view/21>

-
- [22] Ardiansya, H. urhanuddin, A. Alike, S. D. 2023. "Analisis Sentimen Pendapat Masyarakat Terhadap PPKM DKI Jakarta Dengan Metode Naïve Bayes," vol. 10, no. 2, pp. 60–70.
- [23] Trisari, W. Putri, H. Hendrowati, R. 2020. "Penggalian Teks Dengan Model Bag Of Words Terhadap," vol. 2, no. 1, pp. 129–138.
- [24] I. Print, R. Kosasih, and A. Alberto, "InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan Analisis Sentimen Produk Permainan Menggunakan Metode TF-IDF Dan Algoritma K-Nearest Neighbor," vol. 1, 2021.
- [25] N. Bayes, 2019. "Analisis Sentimen Dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia," vol. 6, no. 1.
- [26] Duei Putri, D. Nama, G. F. and Sulistiono, W. E. 2022. "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, doi: 10.23960/jitet.v10i1.2262.