



INNOVATIVE: Journal Of Social Science Research

Volume 4 Nomor 5 Tahun 2024 Page 9033-9048

E-ISSN 2807-4238 and P-ISSN 2807-4246

Website: <https://j-innovative.org/index.php/Innovative>

## Analisis Sentimen Hasil Pemilihan Presiden Ri 2024 Pada *Twitter* Atau *X* Dengan Metode *Naive Bayes Classifier*

Ilham Lantasa Achmadi<sup>1✉</sup>, Haris Yuana<sup>2</sup>, Mohammad Faried Rahmat<sup>3</sup>

Universitas Islam Balitar

Email: [ilhamlantasaahmadi@gmail.com](mailto:ilhamlantasaahmadi@gmail.com)<sup>1✉</sup>

### Abstrak

Pemilihan Presiden Republik Indonesia tahun 2024 menjadi salah satu topik yang paling banyak dibicarakan di media sosial, khususnya di *Twitter* atau *X*. *Platform* ini digunakan oleh masyarakat untuk mengekspresikan opini, pandangan, dan perasaan mereka terkait hasil dari pemilihan presiden RI tahun 2024. Dalam penelitian ini, dilakukan analisis sentimen terhadap tweet yang berkaitan dengan hasil Pemilihan Presiden RI 2024 menggunakan metode *Naive Bayes Classifier*. Tujuan dari penelitian ini adalah untuk mengklasifikasikan sentimen tweet menjadi tiga kategori: positif, negatif, dan netral. Data dikumpulkan dan diproses melalui *crawling data*, *preprocessing data*, *labelling data*, dan implementasi algoritma *Naive Bayes*. Pengumpulan data menggunakan perangkat lunak *Google Colab* dengan bahasa pemrograman *python*. Hasil penelitian menunjukkan bahwa metode ini mampu mengklasifikasikan sentimen dengan hasil 68% dan 73%.. Temuan dari penelitian telah mendekati dengan hasil survei dari Lembaga Survei Indonesia yang menyatakan 71,2%. Dengan demikian, penelitian ini diharapkan dapat berkontribusi dalam memajukan studi analisis sentimen di Indonesia, khususnya dalam konteks pemilihan umum.

Kata Kunci: *Analisis Sentimen, Hasil Pemilihan Presiden 2024, Twitter atau X, Naive Bayes Classifier, Google Colab.*

## Abstract

The 2024 Presidential Election of the Republic of Indonesia has become one of the most talked about topics on social media, especially on Twitter or X. This platform is used by people to express their opinions, views and feelings regarding the outcome of the 2024 Presidential Election. This platform is used by the public to express their opinions, views, and feelings regarding the results of the 2024 Indonesian presidential election. In this study, a sentiment analysis of tweets related to the results of the 2024 RI Presidential Election using the Naive Bayes Classifier method was conducted. The purpose of this research is to classify the sentiment of tweets into three categories: positive, negative, and neutral. Data is collected and processed through data crawling, data preprocessing, data labelling, and Naive Bayes algorithm implementation. Bayes algorithm. Data collection uses Google Colab software with python programming language. The results show that this method is able to classify sentiment with 68% and 73% results. Findings from research have been close to the survey results from the Indonesian Survey Institute which states 71.2%. Thus, this research is expected to contribute to the advancement of sentiment analysis studies in Indonesia, especially in the in the context of general elections.

Keyword: *Sentiment Analysis, 2024 Presidential Election Results, Twitter or X, Naive Bayes Classifier, Google Colab.*

## PENDAHULUAN

Di dalam demokrasi yang ada di Indonesia, kedaulatan rakyat sangat dihargai. Kedaulatan dimiliki dan dilaksanakan oleh rakyat, seperti yang dinyatakan dalam Pasal 1 ayat (2) Undang-Undang Dasar Negara Republik Indonesia Tahun 1945. Salah satu definisi demokratis adalah kekuasaan yang sepenuhnya berasal dari rakyat. Oleh karena itu, rakyatlah yang menentukan, mengarahkan, dan mengelola kehidupan negara. Salah satu aspek terpenting demokrasi adalah kedaulatan rakyat melalui pelaksanaan Pemilihan Umum. Menurut Undang-Undang Nomor 7 Tahun 2017 tentang Pemilihan Umum, pemilihan umum adalah cara rakyat memilih presiden, wakil presiden, dan anggota Dewan Perwakilan Rakyat dan Dewan Perwakilan Rakyat Daerah. "Pemilihan umum dilaksanakan secara langsung, umum, bebas, rahasia, jujur, dan adil setiap lima tahun sekali," menurut Pasal 22 E Ayat 1 Undang-Undang Dasar NRI Tahun 1945. (Ristyawati, 2021).

Analisis sentimen adalah proses mengevaluasi pendapat publik tentang suatu subjek yang datanya diperoleh dari media sosial sosial (Kurniawan & Waluyo, 2022) seperti *Twitter* atau *X*. Analisis sentimen memungkinkan para peneliti dan praktisi politik untuk secara real-time memantau opini publik terhadap hasil pemilihan presiden yang ada pada *Twitter* atau *X*. Untuk melakukan analisis sentimen ini, penulis menggunakan tahapan ekstraksi fitur TF-IDF. Nilai bobot teks dapat dihitung dengan mudah dengan teknik TF-IDF, yang juga dapat

mengurangi ukuran teks untuk menghindari ruang fitur yang besar.

*Twitter* atau *X* adalah *platform* media sosial yang memungkinkan pengguna untuk berbagi pemikiran, berita, dan informasi dalam bentuk teks pendek yang disebut "*tweet*". Sejak *Twitter* diakuisisi oleh Elon Musk pada oktober 2022, Elon Musk merubah nama *Twitter* menjadi *X*. Perubahan ini bertujuan untuk merefleksikan visi *Twitter* atau *X* yang lebih luas dan menggambarkan evolusi perusahaan ke arah yang lebih inklusif dan beragam. *Twitter* atau *X* telah menjadi salah satu *platform* media sosial yang paling berpengaruh dalam konteks politik. Menurut laporan *We Are Social*, jumlah pengguna *Twitter* atau *X* di Indonesia ada sekitar 27,5 juta per Oktober 2023. Dengan jutaan pengguna aktif, *Twitter* atau *X* menjadi wadah penting bagi warga untuk menyampaikan pandangan, opini, dan dukungan mereka terhadap kandidat presiden. *Twitter* atau *X* telah menjadi saluran komunikasi yang sangat signifikan dalam proses politik kontemporer. Masyarakat semakin cenderung menggunakan *platform* media sosial untuk berdiskusi, berbagi informasi, dan menyatakan opini mereka tentang berbagai isu, termasuk pemilihan presiden.

Pemilihan presiden 2024 sudah selesai dilaksanakan, menandai babak baru dalam perjalanan demokrasi negara ini. Proses pemilihan kali ini diwarnai oleh semangat tinggi untuk berpartisipasi dari semua lapisan masyarakat, yang turut aktif dalam menentukan arah masa depan bangsa. Berdasarkan pengumuman dari Komisi Pemilihan Umum Republik Indonesia (KPU RI) pada hari Rabu, 20 Maret 2024, hasil Pemilihan Presiden 2024 telah ditetapkan melalui berita acara KPU nomor 218/PL.01.08-BA/05/2024. Hasil rekapitulasi suara menunjukkan bahwa pasangan nomor urut 2, Prabowo Subianto dan Gibran Rakabuming Raka, keluar sebagai pemenang dengan perolehan 96.214.691 suara dari total 164.227.475 suara sah secara nasional. Persentase suara yang diperoleh pasangan Prabowo-Gibran adalah sekitar 58% dari total suara sah tersebut. (Tim detikcom, 2024). Berdasarkan hasil Lembaga Survei Indonesia (LSI) menunjukkan 71,2% publik merasa puas dengan penyelenggaraan Pemilu 2024. Survei tersebut dilakukan pada 7 hingga 9 April 2024 dengan 1.213 responden (Sinambela, 2024). Dibutuhkannya analisis sentimen masyarakat untuk mengetahui apakah hasil dari pemilihan presiden 2024 pada *Twitter* atau *X* sesuai dengan hasil dari lembaga nasional.

Metode *Naive Bayes Classifier* adalah metode analisis data yang digunakan untuk memprediksi perasaan atau sentimen yang terkandung dalam teks, seperti yang terlihat dalam *tweet*. Dalam konteks pengisian latar belakang proposal analisis sentimen *Twitter* atau *X*, *Naive Bayes Classifier* digunakan untuk mengklasifikasikan *tweet* sebagai positif, negatif, atau netral berdasarkan kata-kata kunci dan pola yang terdapat dalam teks. Metode ini didasarkan pada Teorema *Bayes* dan menganggap bahwa setiap kata dalam teks

berkontribusi secara independen terhadap sentimen keseluruhan. Dengan menerapkan Metode *Naive Bayes Classifier* pada analisis sentimen *Twitter* atau *X*, proposal ini bertujuan untuk menyediakan landasan metodologi yang kuat untuk menggali dan memahami perasaan pengguna *Twitter* atau *X* terkait suatu topik atau peristiwa tertentu. Untuk memahami pandangan dan opini masyarakat, sangat penting untuk melakukan analisis sentimen hasil pemilihan presiden RI 2024 pada *Twitter* atau *X* menggunakan metode *Naive Bayes Classifier*.

## METODE PENELITIAN

Data yang di ambil setelah keputusan hasil pemilihan presiden telah di umumkan, yaitu dari 20 Maret 2024 hingga 20 April 2024 yang kemudian dilanjutkan sampai 31 Juli 2024. Data penelitian ini dikumpulkan di Laboratorium Komputer UNISBA Blitar. Penelitian ini termasuk dalam jenis penelitian kuantitatif karena menggunakan data numerik untuk menganalisis sentimen. Data berupa teks dari *Twitter* atau *X* akan dikonversi menjadi data numerik yang bisa dianalisis statistik. Pendekatan penelitian ini fokus pada analisis teks untuk menentukan sentimen (positif, negatif, atau netral) dari *tweet* atau postingan berkaitan dengan hasil dari pemilihan presiden RI 2024. Penelitian ini memanfaatkan teknik pengumpulan data yang dilakukan melalui *platform* media sosial *Twitter* atau *X*. Penelitian analisis sentimen hasil pemilihan Presiden RI 2024 pada *Twitter* atau *X* dengan metode *Naive Bayes Classifier* melibatkan beberapa tahap penting, yaitu pengumpulan data, *crawling tweet*, *preprocessing*, perancangan klasifikasi, implementasi algoritma (Darwis dkk., 2021).

## HASIL DAN PEMBAHASAN

### 1. Hasil *Crawl* Data

#### A. Sebelum *Crawl Data Twitter* atau *X*

Sebelum melakukan pengambilan data pada *Twitter* atau *X*, diperlukan menginstall *Node.js* dan *npm* sebagai penyedia lingkungan dan alat yang dibutuhkan untuk mengembangkan dan menjalankan aplikasi yang dapat mengakses dan memproses data dari API *Twitter* atau *X*. Berikut adalah *source code* penginstallannya.

---

```
!pip install pandas  
  
!sudo apt-get update  
!sudo apt-get install -y ca-certificates curl gnupg  
!sudo mkdir -p /etc/apt/keyrings
```

---

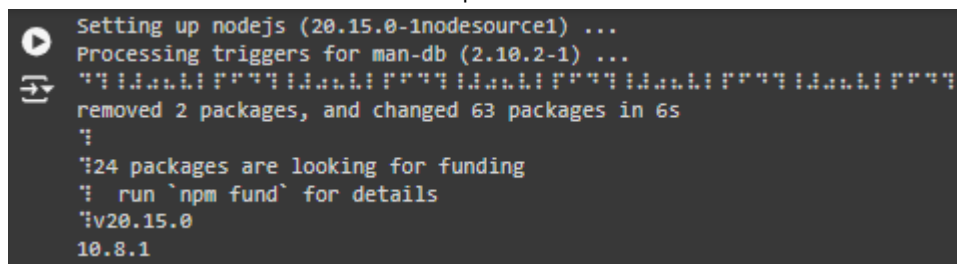
---

```
!curl -fsSL https://deb.nodesource.com/gpgkey/nodesource-
repo.gpg.key | sudo gpg --dearmor -o /etc/apt/keyrings/nodesource.gpg
!NODE_MAJOR=20 && echo "deb [signed-
by=/etc/apt/keyrings/nodesource.gpg]
https://deb.nodesource.com/node_${NODE_MAJOR}.x nodistro main" | sudo
tee /etc/apt/sources.list.d/nodesource.list
!sudo apt-get update
!sudo apt-get install nodejs -y

!npm install -g npm
# Cek versi node dan npm
!node -v
!npm -v
```

---

*!pip install pandas* adalah untuk menginstal paket *Python* bernama *pandas*, yang sering digunakan untuk analisis data dan manipulasi data.

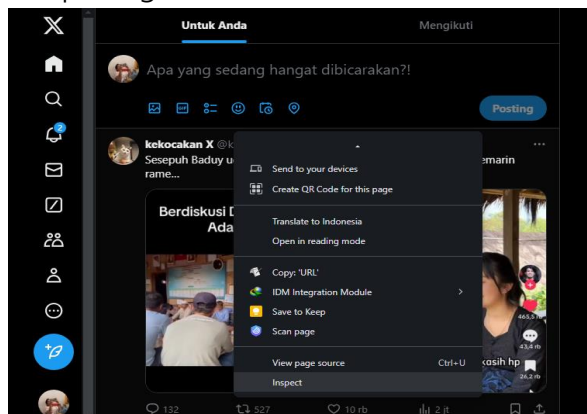


```
Setting up nodejs (20.15.0-1nodesource1) ...
Processing triggers for man-db (2.10.2-1) ...
removed 2 packages, and changed 63 packages in 6s
24 packages are looking for funding
run 'npm fund' for details
v20.15.0
10.8.1
```

Gambar 1 Hasil Instalasi *Node.js* dan *npm*

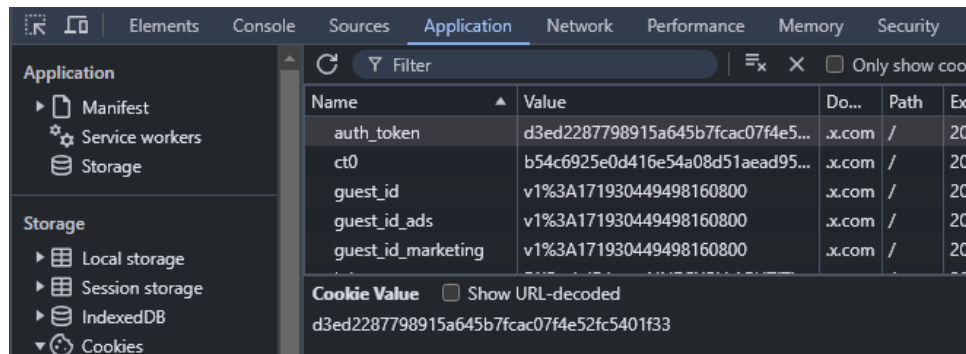
## B. *Twitter* atau *X Auth Token*

*Twitter* atau *X Auth Token* digunakan untuk memastikan bahwa hanya pengguna yang memiliki izin yang dapat mengakses dan mengumpulkan data. secara aman dan efisien selama proses *crawling*, dengan memberikan kontrol terhadap hak akses pengguna, mengelola sesi, serta mencegah penyalahgunaan dan akses tidak sah. Untuk mendapatkan *Auth Token* cukup buka *Twitter* atau *X*, pada halaman beranda tekan klik kanan pada *mouse*, kemudian pilih *Inspect* seperti pada gambar 4.2



Gambar 2 Halaman Sebelum *Inspect*

Akan muncul halaman *Inspect* seperti gambar 3 . Pada bagian *header* pilih yang bertuliskan *Application*, kemudian klik bagian *auth\_token* dan *copy value* nya untuk di *paste* kan pada *source code*.



Gambar 3 Halaman untuk Mendapatkan *Auth Token*

### C. *Crawling Data Twitter* atau *X*

Kembali lagi ke halaman *Google Colab* dan buat *source code* berikut.

```
import pandas as pd
```

```
twitter_auth_token = 'd3ed2287798915a645b7fcac07f4e52fc5401f33'
```

```
start = '2024-03-20'
```

```
end = '2024-04-21'
```

```
search_keyword = 'pilpres2024 until:"{end}" since:"{start}"
```

```
lang:id'.format(start=start, end=end)
```

```
filename = 'Hasil_Crawling_Data.csv'
```

```
limit = 2000
```

```
!npx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" --tab
```

```
"LATEST" -l {limit} --token {twitter_auth_token}
```

*import pandas as pd* digunakan untuk mengimpor *library pandas*, yang menyediakan alat untuk analisis dan manipulasi data, dan memberi alias "pd" agar dapat merujuk pada *pandas* dengan singkat dalam kode. *twitter\_auth\_token* berisi *value* yang di *copy* pada *Auth Token* dari *Twitter* atau *X*. *start* dan *end* merupakan jangka waktu untuk pengambilan data. *search\_keyword* digunakan untuk menentukan *keyword* pencarian data. *lang:id* berfungsi memastikan hasil pencarian hanya dalam bahasa Indonesia. *Filename* adalah penamaan file untuk menyimpan hasil *crawling data*. *limit* digunakan untuk membatasi jumlah pengambilan data. *npx -y tweet-harvest* merupakan sebuah alat yang digunakan untuk mengumpulkan tweet berdasarkan *keyword* pencarian.

```
..npm warn deprecated inflight@1.0.6: This module is not supported, and leaks memory. Do not use
..npm warn deprecated glob@7.2.3: Glob versions prior to v9 are no longer supported

Research by Helmi Satria
Use it for Educational Purposes only!

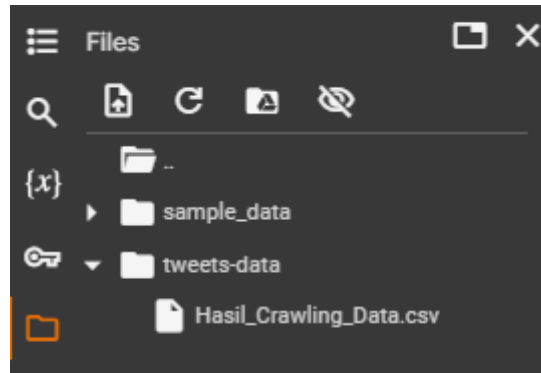
This script uses Chromium Browser to crawl data from Twitter with your Twitter auth token.
Please enter your Twitter auth token when prompted.

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...
```

Gambar 4 Proses *Crawling Data* Berjalan

Setelah proses selesai, hasil dari *crawling* akan tersimpan pada folder "tweets-data"



Gambar 5 File Hasil *Crawling Data*

Isi data yang ada pada "Hasil\_Crawling\_Data.csv" ditunjukkan pada gambar 4.6.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1734 entries, 0 to 1733
Data columns (total 15 columns):
#   Column                Non-Null Count  Dtype
---  -
0   conversation_id_str    1734 non-null   float64
1   created_at            1734 non-null   object
2   favorite_count        1734 non-null   int64
3   full_text             1734 non-null   object
4   id_str                1734 non-null   float64
5   image_url             1093 non-null   object
6   in_reply_to_screen_name 212 non-null   object
7   lang                  1734 non-null   object
8   location              893 non-null   object
9   quote_count           1734 non-null   int64
10  reply_count            1734 non-null   int64
11  retweet_count          1734 non-null   int64
12  tweet_url              1734 non-null   object
13  user_id_str            1734 non-null   float64
14  username               1734 non-null   object
dtypes: float64(3), int64(4), object(8)
memory usage: 203.3+ KB
```

Gambar 6 Isi File Hasil *Crawling Data*

Pada gambar 6 di bagian *Column* merupakan nama dari kolom dalam dataset. *Non-null count* adalah Jumlah baris dalam kolom tersebut yang tidak memiliki nilai null atau kosong. Ini menunjukkan seberapa banyak data yang *valid* terdapat dalam kolom tersebut. *Dtype* adalah Tipe dari data tersebut. *float64* adalah tipe data numerik yang digunakan untuk menyimpan angka desimal (*floating-point number*) dengan presisi 64-bit. *object* adalah tipe data yang dapat menyimpan berbagai jenis data, tetapi biasanya digunakan untuk teks (*string*) dalam konteks *pandas DataFrame*. *int64* adalah tipe data numerik yang digunakan untuk menyimpan angka bulat (*integer*) dengan presisi 64-bit. Ini berarti bahwa

tipe ini dapat menyimpan angka bulat yang sangat besar atau sangat kecil tanpa angka pecahan.

## 2. Term Frequency-Inverse Document Drequency (TF-IDF)

Selanjutnya merupakan pembobotan *Frequency-Inverse Document Drequency* (TF-IDF) pada setiap data. Semakin banyak kata yang berulang pada data, semakin tinggi nilai bobotnya. Berikut ini adalah *source code* untuk pembobotan.

---

```
!pip install pandas scikit-learn

import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
import numpy as np
from tabulate import tabulate

data = pd.read_csv("Hasil_Labelling.csv")
df = data[['username', 'stemming_data', 'compound_score', 'sentiment']]

tfidf_vectorizer = TfidfVectorizer()
tfidf_matrix = tfidf_vectorizer.fit_transform(df['stemming_data'])

terms = tfidf_vectorizer.get_feature_names_out()
idf = np.log(tfidf_matrix.shape[0] /
(np.count_nonzero(tfidf_matrix.toarray(), axis=0) + 1))

tfidf_df = pd.DataFrame({'term': terms, 'idf': idf})

for i, doc in enumerate(data['stemming_data']):
    tf = tfidf_matrix[i].toarray().flatten()
    tfidf_df[f'tf_{i}'] = tf
tfidf_df['tf'] = tfidf_df.iloc[:, 3:].sum(axis=1)
tfidf_df.drop(columns=tfidf_df.columns[3:-1], inplace=True)
# TF-IDF
tfidf_df['tf-idf'] = tfidf_df['tf'] * tfidf_df['idf']
# Print the updated DataFrame
tfidf_df2 = tfidf_df[['term', 'tf', 'idf', 'tf-idf']]
tfidf_df2.to_csv('Hasil_TF-IDF.csv', index=False)
```

---

*from sklearn.feature\_extraction.text import TfidfVectorizer*: Mengimpor kelas *TfidfVectorizer* dari paket *sklearn.feature\_extraction.text*. *TfidfVectorizer* digunakan untuk

mengonversi kumpulan dokumen teks menjadi matriks nilai TF-IDF. *from tabulate import tabulate*: Mengimpor fungsi *tabulate* dari paket *tabulate*. *tabulate* digunakan untuk menghasilkan tabel dari data dalam berbagai format. Hasil dari proses pembobotan TF-IDF ditunjukkan pada gambar 7.

	term	tf	idf	tf-idf
0	ab	0.361079	6.581332	2.376381
1	abah	0.920946	5.888185	5.422699
2	abababahnya	0.466929	6.581332	3.073012
3	abang	1.052345	5.888185	6.196405
4	abdul	0.218522	6.581332	1.438165
5	abdulaziz	0.291538	6.581332	1.918709
6	abdullah	0.214667	6.581332	1.412798
7	absah	0.330307	6.581332	2.173857
8	absurd	0.324283	6.581332	2.134216
9	abuse	0.315980	6.581332	2.079567
10	ac	0.400707	6.581332	2.637184

Gambar 7 Hasil Pembobotan TF-IDF

Gambar 7 merupakan hasil pembobotan TF-IDF. Untuk *term* dari data yang di peroleh berjumlah 3465 kata mulai dari abjad A sampai Z.

### 3. Implementasi Algoritma Naive Bayes Clasifier

Setelah tahap processing data sudah selesai, selanjutnya penerapan Algoritma Naive Bayes. Berikut ini adalah *source code* Nya.

---

```
!pip install scikit-learn # Instalasi Perpustakaan

# Import perpustakaan yang diperlukan
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, classification_report,
confusion_matrix

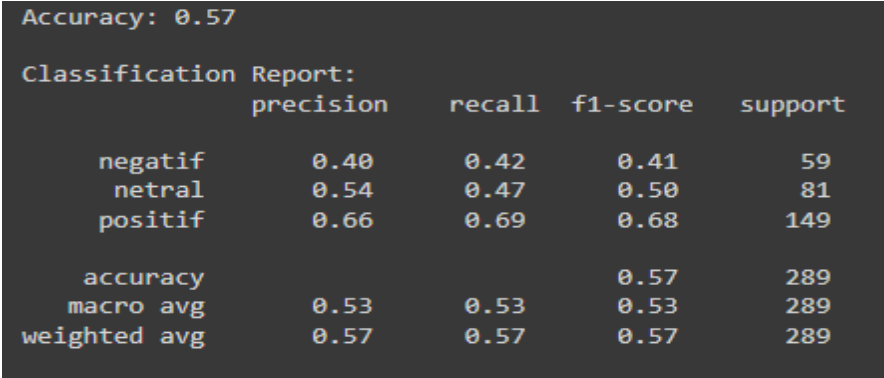
x_train, x_test, y_train, y_test =
train_test_split(df_cleaned['stemming_data'], df_cleaned['sentiment'],
test_size=0.2, random_state=42)
print(f'Jumlah data latih: {len(x_train)}')
print(f'Jumlah data uji: {len(x_test)}')
```

---

$x_{train}$ ,  $x_{test}$ ,  $y_{train}$ ,  $y_{test}$  digunakan untuk membagi data uji dan data latih. Dari 1443 data diambil 80% sebagai data latih menjadi 1154 data dan 20% sebagai data uji menjadi 289 data. Selanjutnya merupakan *source code* untuk latih model *Naive Bayes*.

```
model = MultinomialNB()
model.fit(x_train_vectorized, y_train)
import seaborn as sns
import matplotlib.pyplot as plt
predictions = model.predict(x_test_vectorized)
accuracy = accuracy_score(y_test, predictions)
print(f'Accuracy: {accuracy:.2f}')
print('\nClassification Report:\n', classification_report(y_test,
predictions))
print('\nConfusion Matrix:\n', confusion_matrix(y_test, predictions))
print('\n')
conf_matrix = confusion_matrix(y_test, predictions)
plt.figure(figsize=(8, 6))
sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues',
xticklabels=['Negatif', 'Netral', 'Positif'], yticklabels=['Negatif', 'Netral',
'Positif'])
plt.title('Confusion Matrix')
plt.xlabel('Prediksi')
plt.ylabel('Aktual')
plt.show()
```

*MultinomialNB()* merupakan salah satu kelas dari model *Naive Bayes* yang digunakan untuk klasifikasi teks. Pada gambar 8 merupakan hasil evaluasi model menggunakan data uji.



Accuracy: 0.57				
Classification Report:				
	precision	recall	f1-score	support
negatif	0.40	0.42	0.41	59
netral	0.54	0.47	0.50	81
positif	0.66	0.69	0.68	149
accuracy			0.57	289
macro avg	0.53	0.53	0.53	289
weighted avg	0.57	0.57	0.57	289

Gambar 8 Akurasi dan Klasifikasi dari Data Uji

Pada gambar 8 menunjukkan bahwa *Accuracy* model ini mengklasifikasikan 0.57 yang artinya memiliki akurasi 57% data uji dengan benar dari keseluruhan data uji. *Classification report* memberikan informasi lebih mendetail tentang kinerja model untuk setiap kelas.

*Precision* adalah rasio antara prediksi positif yang benar dengan total prediksi positif. Untuk kelas negatif 0.40 di antaranya benar-benar negatif. Untuk kelas netral 0.54 di antaranya benar-benar netral. Untuk kelas positif 0.66 di antaranya benar-benar positif.

*Recall* adalah rasio antara prediksi positif yang benar dengan total contoh positif sebenarnya. Untuk kelas negatif 0.42 dari semua contoh yang sebenarnya negatif. Untuk kelas netral 0.47 dari semua contoh yang sebenarnya netral. Untuk kelas positif 0.69 dari semua contoh yang sebenarnya positif.

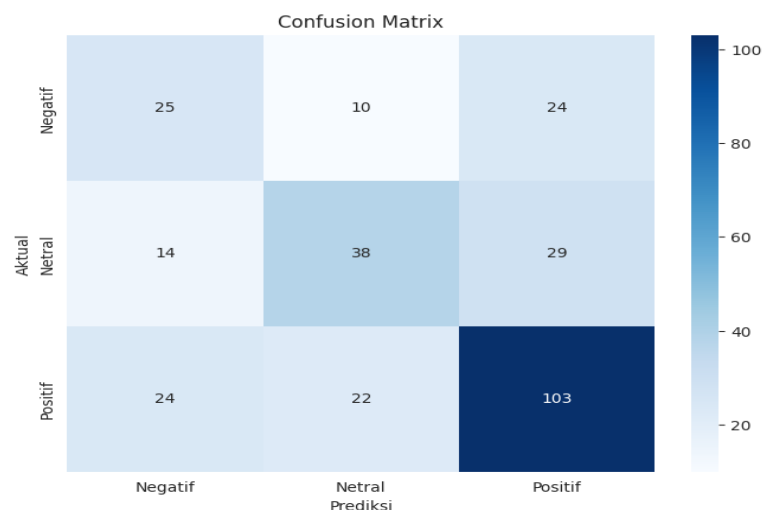
*F1-Score* adalah rata-rata harmonis dari *precision* dan *recall*. Ini memberikan gambaran seimbang dari kedua metrik tersebut. Untuk kelas negatif 0.41, kelas netral 0.50, dan kelas positif 0.68.

*Support* adalah jumlah contoh yang sebenarnya dari masing-masing kelas dalam data uji. Untuk kelas negatif 59 contoh, kelas netral 81 contoh, dan kelas positif 149 contoh.

*Macro avg* menghitung rata-rata dari masing-masing metrik (*precision*, *recall*, *f1-score*) tanpa mempertimbangkan jumlah contoh di setiap kelas. Hasil rata-rata untuk adalah 0.53 untuk *precision*, 0.53 untuk *recall*, dan 0.53 untuk *F1-score*.

*Weighted avg* menghitung rata-rata dari masing-masing metrik (*precision*, *recall*, *f1-score*) dengan mempertimbangkan jumlah contoh di setiap kelas. Untuk *precision* 0.57, *recall* 0.57, dan *F1-score* 0.57.

Untuk *Confusion matrix* ditunjukkan pada gambar 9. merupakan hasil evaluasi model menggunakan data uji.



Gambar 9 *Confusion Matrix* dari Data Uji

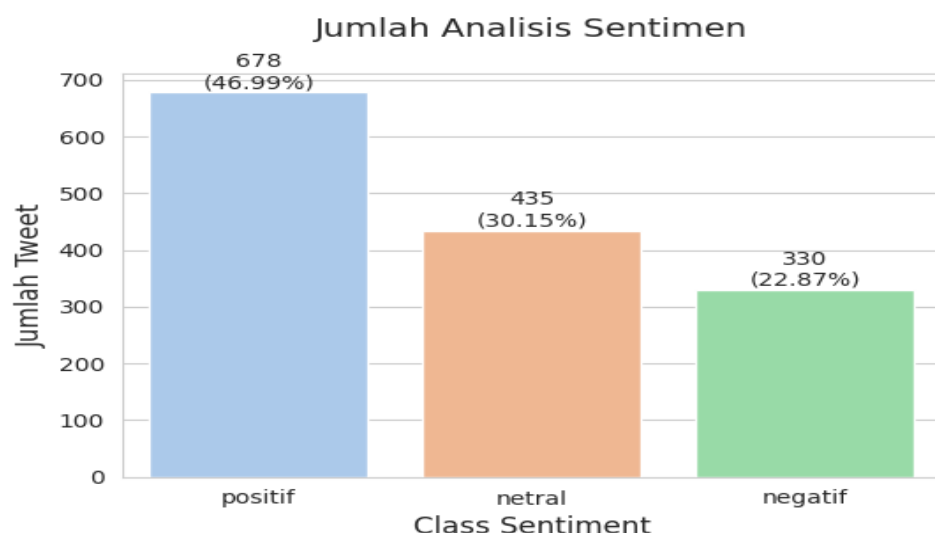
## Pembahasan

Pada bagian pembahasan akan membahas secara mendalam hasil yang diperoleh dari penerapan metode *Naive Bayes* dalam analisis sentimen Hasil Pemilihan Presiden RI 2024 di *Twitter* atau *X*. Proses pengambilan data (*crawling*) dari *Twitter* atau *X* berhasil mengumpulkan sejumlah 1734 data yang terkait dengan Hasil Pemilihan Presiden RI 2024.

Data setelah pengambilan data awal didapatkan 15 kolom dari dataset di antaranya *conversation\_id\_str*, *created\_at*, *favorite\_count*, *full\_text*, *id\_str*, *image\_url*, *in\_reply\_to\_screen\_name*, *lang*, *location*, *quote\_count*, *reply\_count*, *retweet\_count*, *tweet\_url*, *user\_id\_str*, dan *username*. Dari 15 kolom diambil hanya pada bagian *full\_text* karena merupakan isi dari *tweet* yang selanjutnya akan dilakukan tahap preprocessing data.

Tahap preprocessing data melibatkan beberapa langkah penting seperti *cleaning*, *case folding*, *tokenizing*, *stopword removal* dan *stemming*. Pada tahap *cleaning* dilakukan juga penghapusan data *tweet* yang *duplicate* bertujuan untuk menjadikan data lebih relevan. Data yang awalnya berjumlah 1734 menjadi 1443 data. Setelah melakukan tahap *case folding* yang seharusnya *tokenizing*, ada tahap tambahan yaitu normalisasi kata agar saat *preprocessing* mendapat hasil yang lebih siap untuk dianalisis lebih lanjut. Beberapa tantangan yang dihadapi termasuk menangani variasi bahasa dan slang yang sering muncul di *tweet*, serta memastikan bahwa proses *stemming* tidak mengurangi makna penting dari kata-kata yang ada.

Pada tahap *labelling data*, *tweet* dikategorikan ke dalam tiga kelas sentimen: positif, netral, dan negatif. Pada proses *labelling data* hasil dari *stemming data* dilakukan *translate* ke dalam bahasa *inggris*. Hal tersebut dilakukan agar hasil dari *compound score* dan *sentiment* lebih tepat. Hasil dari pelabelan menunjukkan distribusi sentimen yang beragam dengan proporsi yang signifikan pada ketiga kategori. Beberapa *tweet* mungkin sulit untuk dikategorikan dengan jelas karena ambiguitas dalam bahasa yang digunakan. Pada gambar 10 merupakan perbandingan sentimen positif, netral, dan negatif setelah proses *labelling data*.



Gambar 10 Perbandingan Jumlah Sentimen Positif, Netral , dan Negatif

Penerapan *Term Frequency-Inverse Document Frequency* (TF-IDF) pada data yang telah diproses memberikan bobot pada kata-kata berdasarkan frekuensi kemunculannya

dan kepentingannya dalam seluruh kumpulan data. Hasil TF-IDF membantu dalam mengidentifikasi kata-kata yang paling signifikan untuk analisis sentimen. Tantangan dalam tahap ini termasuk menangani kata-kata yang terlalu umum atau terlalu spesifik yang dapat mengganggu analisis.

Algoritma *Naive Bayes* diterapkan pada data yang telah ditransformasi dengan TF-IDF. Hasil implementasi menunjukkan bahwa algoritma ini mampu memberikan klasifikasi sentimen dengan tingkat akurasi yang memuaskan. Evaluasi kinerja algoritma dilakukan dengan menggunakan metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Gambar 11 dan gambar 12 merupakan evaluasi model menggunakan data latih.

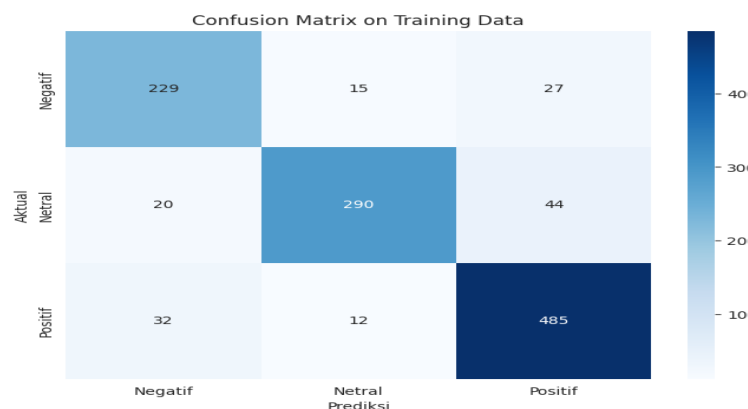
```
Accuracy on Training Data: 0.87

Classification Report on Training Data:
              precision    recall  f1-score   support

negatif      0.81      0.85      0.83      271
netral       0.91      0.82      0.86      354
positif      0.87      0.92      0.89      529

accuracy          0.87          0.87          0.87      1154
macro avg         0.87          0.86          0.86      1154
weighted avg     0.87          0.87          0.87      1154
```

Gambar 11 Akurasi dan Klasifikasi dari Data Latih

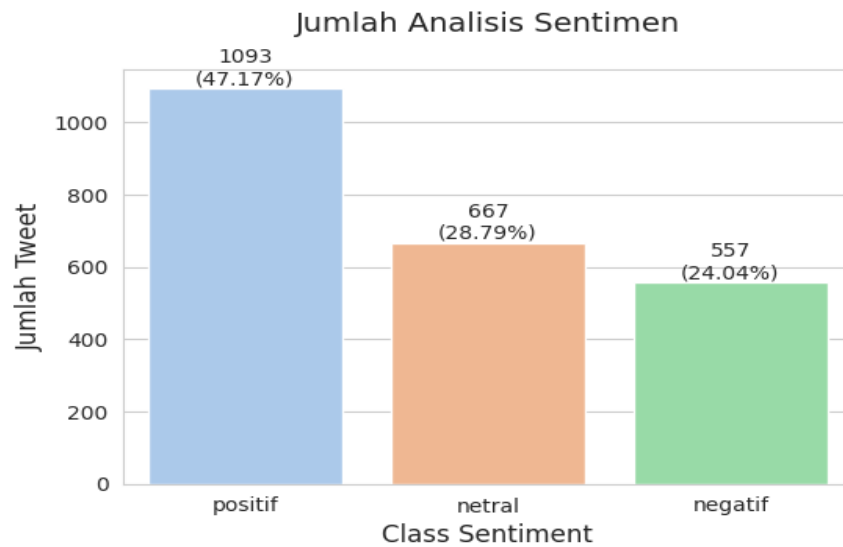


Gambar 12 *Confusion Matrix* dari Data Latih

Evaluasi model menggunakan data latih biasanya bukan praktik yang disarankan, karena ini dapat menyebabkan *overfitting*. *Overfitting* adalah sebuah fenomena dalam *machine learning* di mana sebuah model mempelajari data pelatihan terlalu baik, termasuk menangkap noise dan detail yang tidak relevan. *Overfitting* terjadi ketika model terlalu “menghafal” data latih dan tidak dapat menggeneralisasi dengan baik pada data baru. Oleh karena itu lebih disarankan menggunakan data uji untuk mengevaluasi kinerja model.

Untuk mengetahui apakah penelitian ini relevan, dilakukan pengambilan data kembali yang awalnya mengambil data selama sebulan mulai dari tanggal 20 Maret 2024 sampai 20 April 2024 menjadi 5 bulan, mulai tanggal 20 Maret 2024 sampai 31 Juli 2024. Dengan

pengambilan data selama 5 bulan data yang diperoleh sejumlah 3014 data yang kemudian dilakukan penghapusan data yang *duplicate* menjadi 2317. Pada gambar 13 merupakan perbandingan sentimen positif, netral, dan negatif setelah proses *labelling data* dari 2934 data.



Gambar 13 Perbandingan Jumlah Sentimen Positif, Netral , dan Negatif Pengambilan 5 bulan

Dari 2317 data diambil 80% sebagai data latih menjadi 1853 data dan 20% sebagai data uji menjadi 464 data. Pada gambar 14 merupakan hasil evaluasi model menggunakan data uji dari pengambilan data 5 bulan.

```

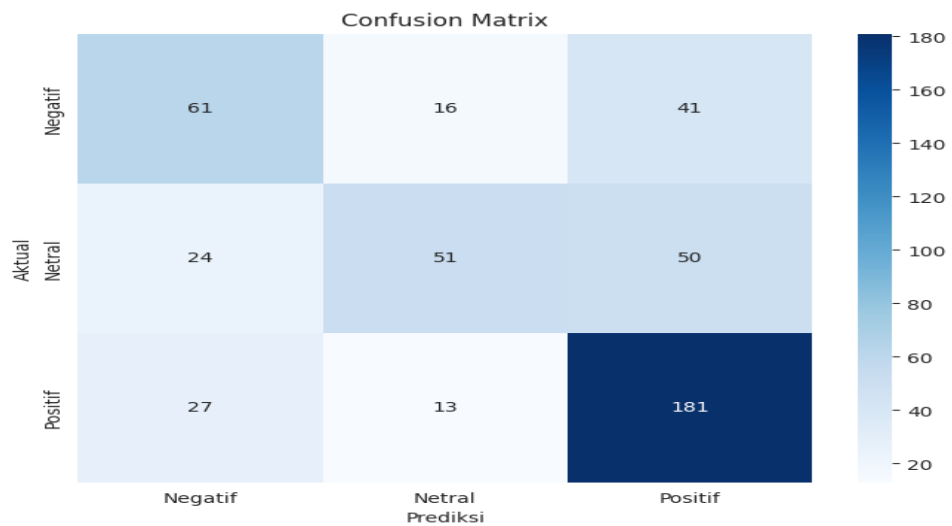
Accuracy: 0.63

Classification Report:
      precision    recall  f1-score   support

negatif    0.54    0.52    0.53     118
netral     0.64    0.41    0.50     125
positif    0.67    0.82    0.73     221

accuracy    0.63    0.63    0.63     464
macro avg   0.62    0.58    0.59     464
weighted avg 0.63    0.63    0.62     464
    
```

Gambar 4.17 Akurasi dan Klasifikasi dari Data Uji Pengambilan 5 Bulan



Gambar 15 *Confusion Matrix* dari Data Uji Pengambilan 5 Bulan

Pada gambar 15 dapat dilihat menunjukkan bahwa *f1-score* positif model ini mengklasifikasikan 0.73 yang artinya memiliki 73%. Hal tersebut membuktikan bahwa pengambilan data selama 1 bulan hampir sama dengan nilai pengambilan data selama 5 bulan.

## SIMPULAN

Dari penelitian yang telah dilakukan, Kesimpulan yang dapat di ambil penulis sebagai berikut :

1. Berdasarkan hasil penelitian ini, metode *Naive Bayes Classifier* terbukti efektif dalam mengklasifikasikan sentimen pada data *Twitter* atau *X* terkait hasil pemilihan presiden RI 2024.
2. Model ini mampu mengelompokkan tweet menjadi kategori positif, negatif, dan netral dengan evaluasi model pengambilan data selama 1 bulan didapatkan *accuracy* 57%, untuk kategori positif *precision* 66% *recall* 69% *f1-score* 68%, untuk kategori negatif *precision* 40% *recall* 42% *f1-score* 41% dan untuk kategori netral *precision* 54% *recall* 47% *f1-score* 50%. Untuk hasil evaluasi model pengambilan data selama 5 bulan didapatkan *accuracy* 63%, untuk kategori positif *precision* 67% *recall* 82% *f1-score* 73%, untuk kategori negatif *precision* 54% *recall* 52% *f1-score* 53% dan untuk kategori netral *precision* 64% *recall* 41% *f1-score* 50%. serta *confusion matrix* yang ditunjukkan di hasil dan pembahasan. Keberhasilan ini menegaskan bahwa Naive Bayes adalah salah satu algoritma yang handal untuk tugas klasifikasi teks dalam domain media sosial.

Analisis sentimen mengungkap bahwa sebagian besar *tweet* mengandung sentimen positif terhadap hasil pemilihan presiden RI 2024. Hal tersebut sudah sesuai dengan hasil dari Lembaga Survei Indonesia (LSI) yang menunjukkan presentase sebesar 71,2%. Namun,

terdapat juga sejumlah tweet dengan sentimen netral dan negatif, yang menunjukkan adanya spektrum opini yang beragam di kalangan pengguna *Twitter* atau *X*. Sentimen positif ini mungkin mencerminkan dukungan dan harapan masyarakat terhadap presiden terpilih, sementara sentimen netral dan negatif mencerminkan kekhawatiran atau ketidakpuasan sebagian masyarakat.

#### DAFTAR PUSTAKA

- Darwis, D., Siskawati, N., & Abidin, Z. (2021). *Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional*. 15(1), 131–145.
- Kurniawan, A., & Waluyo, S. (2022). Penerapan Algoritma Naive Bayes Dalam Analisis Sentimen Pemindahan Ibukota Pada Twitter Application Of Naive Bayes Algorithm In Capital Movement Sentiment Analysis On Twitter. *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) Jakarta-Indonesia, September*, 455–461.
- Ristyawati, A. (2021). *Analisis Wacana Pencalonan Presiden dan Wakil Presiden Non Partai Politik di Indonesia dalam Kerangka Negara Demokrasi*. 4(1), 117–125.
- Tim detikcom. (2024). *Lengkap! Hasil Pemilu 2024: Pemenang Pilpres hingga Daftar Parpol ke DPR*. News.Detik.Com. <https://news.detik.com/pemilu/d-7253356/lengkap-hasil-pemilu-2024-pemenang-pilpres-hingga-daftar-parpol-ke-dpr>