

# Analisis Sentimen Pada Media Sosial Menggunakan Teknik *Natural Language Processing*

Gilang Prima Ertansyah<sup>1\*</sup>, Ridwan Tri Cahya Kusuma<sup>2</sup>, Aprilisa Arum Sari<sup>3</sup>

<sup>1</sup>Teknik Informatika/Ilmu Komputer  
Afiliasi (Universitas Duta Bangsa  
Surakarta)

<sup>2</sup>Teknik Informatika/Ilmu Komputer  
Afiliasi (Universitas Duta Bangsa  
Surakarta)

<sup>3</sup>Teknik Informatika/Ilmu Komputer  
Afiliasi (Universitas Duta Bangsa  
Surakarta)

<sup>1\*</sup>gilangprimaertansyah76@gmail.com  
(penulis korespondensi)

<sup>2</sup>ridwantriacahya@gmail.com

<sup>3</sup>aprilisa\_arumsari@udb.ac.id

**Abstrak**— Penelitian ini bertujuan untuk menganalisis sentimen pada data media sosial menggunakan teknik *Natural Language Processing* (NLP). Masalah yang dihadapi adalah bagaimana mengklasifikasikan sentimen positif, negatif, dan netral dari sejumlah besar data yang diambil dari media sosial. Tujuan dari penelitian ini adalah untuk mendapatkan gambaran umum tentang persepsi publik terhadap isu tertentu melalui media sosial. Metode yang digunakan melibatkan pengumpulan data, preprocessing, dan analisis menggunakan algoritma NLP.

**Kata kunci**— Analisis Sentimen, *Natural Language Processing* (NLP), Media Sosial.

**Abstract**— This study aims to analyze sentiment on social media data using *Natural Language Processing* (NLP) techniques. The problem faced is how to classify positive, negative, and neutral sentiment from a large amount of data taken from social media. The purpose of this study is to get an overview of public perception of a particular issue through social media. The method used involves data collection, preprocessing, and analysis using NLP algorithms.

**Keywords**— Sentiment Analysis, *Natural Language Processing* (NLP), Social Media.

## I. PENDAHULUAN

Media sosial telah menjadi platform utama bagi individu untuk mengekspresikan opini mereka tentang berbagai topik. Sentimen yang diungkapkan dalam postingan media sosial mencerminkan perasaan dan pandangan masyarakat terhadap isu-isu tertentu (Herman Lawelai et al 2022). Analisis sentimen merupakan proses penting dalam memahami opini publik, yang dapat digunakan untuk berbagai tujuan seperti pemasaran, politik, dan analisis sosial (Arianti Primadhani Tirtopangarsa dan Warih Maharani 2021). Penelitian ini berfokus pada analisis sentimen data media sosial untuk mengidentifikasi distribusi sentimen positif, negatif, dan netral. Dengan menggunakan teknik *Natural Language Processing* (NLP), data yang sangat besar dapat dianalisis untuk mendapatkan wawasan yang lebih mendalam tentang persepsi publik.

Solusi yang ditawarkan dalam penelitian ini adalah penggunaan algoritma NLP untuk memproses dan menganalisis data media sosial. Dengan pendekatan ini, kita dapat mengklasifikasikan sentimen secara otomatis dan

efisien, memberikan gambaran yang lebih jelas tentang bagaimana suatu isu dipersepsikan oleh publik. Implementasi NLP dalam analisis sentimen tidak hanya meningkatkan kecepatan dan akurasi, tetapi juga memungkinkan analisis data dalam skala besar yang sebelumnya tidak mungkin dilakukan secara manual (Akhmad Muzaki dan Arita Witanti 2021). Teknik NLP melibatkan berbagai metode seperti tokenisasi, stemming, dan lemmatization untuk memproses teks. Setelah itu, algoritma *machine learning* atau *deep learning* diterapkan untuk mengklasifikasikan sentimen.

Penelitian ini bertujuan untuk menganalisis sentimen pada data media sosial, mengklasifikasikan sentimen positif, negatif, dan netral, serta mengatasi ambiguitas dalam data. Masalah utama yang dihadapi adalah bagaimana mengklasifikasikan sentimen secara efisien dan akurat, mengatasi variasi bahasa dan gaya penulisan, serta mempertimbangkan pengaruh eksternal seperti berita atau peristiwa global yang dapat mempengaruhi sentimen publik (Rahmat Hidayat dan Budi Santoso 2020). Dengan demikian, analisis sentimen menggunakan NLP

memberikan alat yang kuat untuk memahami opini publik secara *real-time*, yang sangat berguna bagi pembuat keputusan dalam berbagai bidang.

## II. METODOLOGI PENELITIAN

Penelitian ini bertujuan untuk menganalisis sentimen pada data media sosial menggunakan teknik *Natural Language Processing* (NLP). Dalam konteks ini, teknik NLP digunakan untuk mengklasifikasikan sentimen menjadi kategori positif, negatif, atau netral berdasarkan teks yang diambil dari platform media sosial. Tantangan utama yang dihadapi adalah bagaimana mengelola dan menginterpretasikan volume besar data yang heterogen, serta mengubahnya menjadi informasi yang dapat digunakan untuk memahami pandangan dan reaksi publik terhadap isu-isu tertentu.

Proses penelitian meliputi pengumpulan data dengan teknik online *scraping* menggunakan kata kunci yang relevan dengan topik penelitian. Data yang dikumpulkan kemudian diproses melalui tahapan *preprocessing* seperti tokenisasi untuk memecah teks menjadi unit-unit kata, dan *stemming* untuk mengubah kata-kata ke bentuk dasarnya. Selanjutnya, data dipisahkan menjadi set pelatihan dan pengujian untuk melatih model NLP dan menguji akurasi serta kinerjanya. Dengan memanfaatkan algoritma-algoritma seperti Naive Bayes, *Support Vector Machine* (SVM), atau *Random Forest*, penelitian ini bertujuan untuk membangun model yang dapat mengklasifikasikan sentimen secara efektif. Hasil evaluasi kinerja model menggunakan metrik akurasi, presisi, recall, dan *F1-score* diharapkan dapat memberikan wawasan yang mendalam tentang bagaimana pendapat publik tercermin dalam data media sosial.

### A. Tahap Perencanaan

Tahap perencanaan melibatkan identifikasi tujuan penelitian dan perumusan hipotesis yang jelas serta relevan dengan masalah yang akan diteliti. Dalam tahap ini, peneliti menentukan metode dan alat yang akan digunakan, seperti pemilihan teknik pengumpulan data dan alat analisis. Peneliti juga membuat jadwal penelitian yang terstruktur untuk memastikan setiap

tahapan dilakukan sesuai rencana dan mengidentifikasi sumber daya yang diperlukan, termasuk perangkat keras, perangkat lunak, dan personel. Peneliti juga mempertimbangkan etika penelitian, memastikan bahwa data dikumpulkan dan digunakan sesuai dengan pedoman etis yang berlaku.

### B. Tahapan *preprocessing* data

Pada tahap *preprocessing* data, pertama-tama dilakukan tokenisasi, yaitu proses memecah teks menjadi unit-unit kata yang lebih kecil yang disebut token untuk mempermudah analisis dengan mengidentifikasi setiap kata dalam teks. Contohnya, kalimat "Saya suka belajar NLP" akan dipecah menjadi ["Saya", "suka", "belajar", "NLP"]. Selanjutnya, dilakukan *stemming* di mana setiap kata dalam teks diubah ke bentuk dasarnya untuk menyederhanakan dan menyatukan variasi kata yang memiliki makna sama. Contohnya, kata-kata seperti "berlari", "berlari-lari", dan "berlari-larian" akan diubah menjadi bentuk dasar "lari".

### C. Tahapan Proses *Coding*

Tahap proses *coding* dimulai dengan persiapan lingkungan *coding*, termasuk menginstal pustaka dan paket yang diperlukan seperti NLTK, *Scikit-learn*, dan *Pandas*. Setelah itu, dilakukan pengumpulan dan penyimpanan data dari media sosial menggunakan API atau teknik *web scraping*, dan data disimpan dalam format yang mudah diakses, seperti CSV atau database. Implementasi kode untuk *preprocessing* data dilakukan, mencakup tokenisasi dan *stemming*, membersihkan data teks, dan menyiapkan data untuk analisis. Kemudian, kode untuk membagi data menjadi set pelatihan dan set pengujian serta melatih model NLP menggunakan algoritma yang dipilih diimplementasikan. Rumus yang akan digunakan adalah *VADER Compound Score* atau *Valence Aware Dictionary and sEntiment Reasoner* yang termasuk kedalam *Nature Language Processing* dengan rincian rumus berikut:

$$Vader(Compound\ Score) = \frac{\sum S_i}{\sqrt{\sum S_i + \alpha}}$$

- a.  $S_i$  adalah skor sentimen dari setiap kata.
- b.  $\alpha = 15$  (konstanta *smoothing* untuk normalisasi hasil pada *coding*).
- c. Vader score berada di range [-1, +1].

Setiap sentimen akan dibagi menjadi tiga *positive* jika nilai vadernya  $\geq 0.05$ , *negative* jika nilai vadernya  $\leq -0.05$  dan yang lainnya berarti netral. Setiap sentimen yang akan di proses sudah melalui normalisasi kata agar tidak membingungkan kita dalam prosesnya, contoh kata harus *lowercase*, menghapus *URL*, angka, tanda baca, *non-ASCII*, dan melalui proses lainnya agar menjadi *text* yang normal yang dapat diproses sistem.

#### D. Tahapan Evaluasi

Kinerja model dievaluasi menggunakan metrik seperti akurasi, presisi dan *recall*. Akurasi mengukur seberapa sering model membuat prediksi yang benar, presisi mengukur proporsi prediksi positif yang benar dan *recall* mengukur proporsi data aktual yang teridentifikasi dengan benar oleh model. Kinerja model dievaluasi menggunakan metrik seperti akurasi, presisi dan *recall*. Akurasi mengukur seberapa sering model membuat prediksi yang benar, presisi mengukur proporsi prediksi positif yang benar dan *recall* mengukur proporsi data aktual yang teridentifikasi dengan benar oleh model.

### III. HASIL DAN PEMBAHASAN

#### A. Tahap Perencanaan

Pada tahap perencanaan, penelitian ini dimulai dengan mengidentifikasi tujuan utama yaitu untuk menganalisis sentimen pada data media sosial menggunakan teknik *Natural Language Processing* (NLP). Tahap ini melibatkan perencanaan detail mengenai metode pengumpulan data, yang akan dilakukan dengan teknik *web scraping* dan *API* dari platform media sosial seperti Twitter, Facebook, dan Instagram. Data yang dikumpulkan akan melalui proses *preprocessing* yang mencakup tokenisasi, *stemming*, dan *lemmatization* untuk membersihkan dan mempersiapkan data agar siap dianalisis lebih lanjut. Selain itu, pemilihan algoritma yang tepat untuk analisis sentimen menjadi fokus utama, dimana beberapa algoritma

NLP dan *machine learning* akan diuji untuk menemukan yang paling efektif dalam mengklasifikasikan sentimen menjadi positif, negatif, dan netral. Keseluruhan rencana penelitian ini diharapkan dapat memberikan gambaran yang komprehensif tentang persepsi publik terhadap isu-isu tertentu yang dibahas di media sosial. Pada tahapan ini juga persiapan data sebelum diolah atau diolah dan di proses akan diambil. Berikut contoh data sebelum diolah:

Unnamed: 0	Text
0	0 Enjoying a beautiful day at the park!
1	1 Traffic was terrible this morning.
2	2 Just finished an amazing workout! 🏆
3	3 Excited about the upcoming weekend getaway!
4	4 Trying out a new recipe for dinner tonight.
5	5 Feeling grateful for the little things in life.
6	6 Rainy days call for cozy blankets and hot cocoa.
7	7 The new movie release is a must-watch!
8	8 Political discussions heating up on the timeline.
9	9 Missing summer vibes and beach days.
10	10 Just published a new blog post. Check it out!

Gambar 1. Data Sebelum Diolah.

#### B. Tahapan preprocessing data

Unnamed: 0	Text	Sentiment	Timestamp	User
0	0 Enjoying a beautiful day at the park!	Positive	2023-01-23 12:30:00	User123
1	1 Traffic was terrible this morning.	Negative	2023-01-23 08:45:00	Comment45
2	2 Just finished an amazing workout! 🏆	Positive	2023-01-23 19:20:00	Post678
3	3 Excited about the upcoming weekend getaway!	Positive	2023-01-23 10:15:00	Message90
4	4 Trying out a new recipe for dinner tonight.	Neutral	2023-01-23 18:50:00	Story111
5	5 Feeling grateful for the little things in life.	Positive	2023-01-23 09:30:00	Video222
6	6 Rainy days call for cozy blankets and hot cocoa.	Positive	2023-01-23 14:40:00	Image333
7	7 The new movie release is a must-watch!	Positive	2023-01-23 16:00:00	Review444
8	8 Political discussions heating up on the timeline.	Negative	2023-01-23 13:30:00	Article555
9	9 Missing summer vibes and beach days.	Neutral	2023-01-22 22:00:00	Blog666
10	10 Just published a new blog post. Check it out!	Positive	2023-01-23 10:00:00	Post777

Gambar 2. Data Setelah Diolah.

Gambar ini menunjukkan tabel yang lebih lengkap dari yang sebelumnya, dengan kolom tambahan yang mencakup *Sentiment*, *Timestamp*, *User*, *Platform*, *Hashtags*, *Retweets*, *Likes*, *Country*, *Year*, *Month*, *Day*, dan *Hour*. Berikut adalah pembahasan mengenai tahapan preprocessing data yang diterapkan pada tabel ini:

- a. Tokenisasi: Tokenisasi adalah proses memecah teks menjadi unit-unit kata yang lebih kecil yang disebut token untuk mempermudah analisis. Dalam konteks tabel ini, setiap kalimat dalam kolom "Text" akan dipecah menjadi token individu. Misalnya, kalimat "Traffic was terrible this morning." akan dipecah menjadi ["Traffic", "was", "terrible", "this", "morning"].

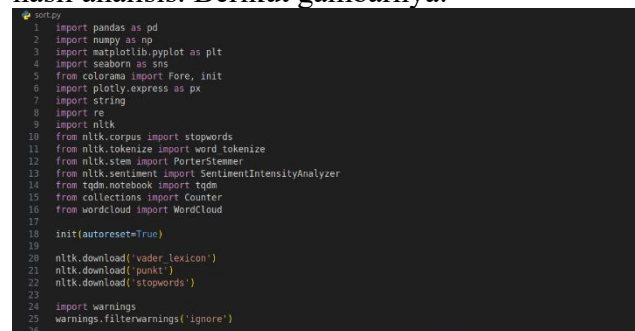
- b. *Stemming*: *Stemming* mengubah setiap kata dalam teks ke bentuk dasarnya untuk menyederhanakan dan menyatukan variasi kata yang memiliki makna sama. Misalnya, kata-kata seperti "*excited*", "*excitement*", dan "*exciting*" akan diubah menjadi bentuk dasar "*excite*". Dalam tabel ini, proses *stemming* akan diterapkan pada setiap token hasil dari tokenisasi.
- c. *Lemmatization*: *Lemmatization* mirip dengan *stemming*, tetapi lebih kompleks karena mempertimbangkan konteks dan mengubah kata-kata ke bentuk dasarnya yang lebih tepat. Contohnya, "*better*" akan diubah menjadi "*good*" dan "*running*" akan diubah menjadi "*run*".
- d. *Pembersihan Data*: Proses ini melibatkan penghapusan karakter khusus, angka, tanda baca, dan *stopwords* (kata umum yang tidak memiliki makna signifikan dalam analisis, seperti "*and*", "*the*", "*is*"). Dalam tabel ini, setiap entri dalam kolom "*Text*" akan dibersihkan dari elemen-elemen tersebut.
- e. *Normalisasi Teks*: *Normalisasi* mengubah semua teks menjadi format konsisten, misalnya, mengubah semua huruf menjadi huruf kecil untuk menghindari duplikasi akibat perbedaan kapitalisasi.
- f. *Pengkodean Sentimen*: Sentimen pada kolom "*Sentiment*" telah diidentifikasi sebagai positif, negatif, atau netral. Pengkodean ini memungkinkan analisis lebih lanjut mengenai distribusi sentimen dalam data.
- g. *Ekstraksi Fitur*: Fitur-fitur tambahan seperti *Timestamp*, *User*, *Platform*, *Hashtags*, *Retweets*, *Likes*, *Country*, *Year*, *Month*, *Day*, dan *Hour* telah diekstrak dan disertakan dalam tabel untuk analisis lebih lanjut. Fitur-fitur ini dapat memberikan konteks tambahan dan memungkinkan analisis temporal, geografis, dan sosial dari data.

### C. Tahapan Proses Coding

Dalam analisis data media sosial, penting untuk melalui beberapa tahapan *Coding* yang sistematis agar data dapat diproses dan dianalisis secara efektif. Proses ini dimulai dengan mengimpor *library* yang diperlukan dan menginisialisasi lingkungan kerja. Setelah itu, data dimuat dan dieksplorasi untuk memahami struktur dan kualitasnya. Tahap selanjutnya adalah *preprocessing* data, di mana data dibersihkan dan diubah menjadi format yang siap untuk analisis. Berikut ini adalah rincian dari setiap tahapan dalam proses Codingnya:

#### a. *Import Libraries* dan Inisialisasi

Bagian ini memuat pustaka yang dibutuhkan untuk analisis dan visualisasi data, seperti *Pandas*, *NumPy*, *Matplotlib*, *Seaborn*, *Plotly*, serta *Natural Language Processing* (NLP) dari NLTK. *init(autoreset=True)* digunakan untuk mengatur *Colorama* agar warna otomatis di-reset setelah digunakan. *nlk.download* mengunduh paket NLTK yang diperlukan seperti '*vader\_lexicon*', '*punkt*', dan '*stopwords*'. Selain itu, peringatan diabaikan dengan *warnings.Filterwarnings('ignore')* untuk menjaga agar output tetap bersih dan fokus pada hasil analisis. Berikut gambarnya:



```

1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5 from colorama import Fore, Init
6 import plotly.express as px
7 import string
8 import re
9 import nltk
10 from nltk.corpus import stopwords
11 from nltk.tokenize import word_tokenize
12 from nltk.stem import PorterStemmer
13 from nltk.sentiment import SentimentIntensityAnalyzer
14 from tqdm.notebook import tqdm
15 from collections import Counter
16 from wordcloud import WordCloud
17
18 init(autoreset=True)
19
20 nltk.download('vader_lexicon')
21 nltk.download('punkt')
22 nltk.download('stopwords')
23
24 import warnings
25 warnings.filterwarnings('ignore')
26

```

Gambar 3. Import Yang Digunakan.

#### b. Memuat dan Mengeksplorasi Data

Data dimuat dari file CSV ke dalam *DataFrame*, kemudian beberapa baris pertama ditampilkan untuk memberikan gambaran awal tentang data. Fungsi *null\_count* digunakan untuk menghitung jumlah nilai *NaN* dan persentasenya dalam setiap kolom, membantu dalam identifikasi *missing values*. Selanjutnya, kode ini memeriksa jumlah baris duplikat dan menampilkan jumlah nilai unik dalam setiap kolom untuk mendapatkan pemahaman yang

lebih baik mengenai distribusi data. Berikut gambarnya:

```

26
27 df = pd.read_csv("sentimentdataset.csv")
28 print(df.head())
29
30 def null_count(df):
31     return pd.DataFrame({'features': df.columns,
32                          'dtypes': df.dtypes.values,
33                          'NaN count': df.isnull().sum().values,
34                          'NaN percentage': df.isnull().sum().values / df.shape[0]})
35     .style
36     .background_gradient(cmap=Set3, low=0.1, high=0.01)
37
38 print(null_count(df))
39
40 print(f'Duplicate rows: {df.duplicated().sum()}')
41 print(f'Columns: {df.columns}')
42
43 for column in df.columns:
44     num_distinct_values = len(df[column].unique())
45     print(f'Column: {column}, num_distinct_values: {num_distinct_values}')
46
47 df = df.drop(columns=['Unnamed: 0.1', 'Unnamed: 0', 'Hashtags', 'Day', 'Hour', 'Sentiment'])
48
49 df['Platform'] = df['Platform'].str.strip()
50 df['Country'] = df['Country'].str.strip()
51

```

Gambar 4. Pengambilan Data Dan Explore Data.

c. Preprocessing Data

Tahapan ini mencakup pembersihan data dengan menghapus kolom yang tidak diperlukan, menghilangkan spasi tambahan pada kolom 'Platform' dan 'Country', serta mengonversi kolom 'Timestamp' ke format *datetime* dan menambah kolom hari dalam seminggu. Data bulan juga diubah menjadi nama bulan. Fungsi *clean* digunakan untuk membersihkan teks dari tanda baca, URL, angka, dan karakter khusus, lalu melakukan tokenisasi dan *stemming* untuk mengubah kata-kata ke bentuk dasarnya dan menghilangkan *stopwords*. Berikut gambarnya:

```

51 df['Day_of_Week'] = df['Timestamp'].dt.day_name()
52
53 month_mapping = {
54     1: 'Januari', 2: 'Februari', 3: 'Maret', 4: 'April',
55     5: 'Mei', 6: 'Juni', 7: 'Juli', 8: 'Agustus',
56     9: 'September', 10: 'Oktober', 11: 'November', 12: 'Desember'
57 }
58 df['Month'] = df['Month'].map(month_mapping).astype(object)
59
60 stemmer = PorterStemmer()
61 stop_words = set(stopwords.words('english'))
62
63 def clean(text):
64     text = str(text).lower()
65     text = re.sub(r'\s+', ' ', text)
66     text = re.sub(r'https?://\S+', '', text)
67     text = re.sub(r'@\S+', '', text)
68     text = re.sub(r'#\S+', '', text)
69     text = re.sub(r'\d+', '', text)
70     text = re.sub(r'[\!@#$%^&*(){}~`~\|;:,.?!"', '', text)
71     text = re.sub(r'[\u201c\u201d]', '', text)
72     text = re.sub(r'[\u201e\u201f]', '', text)
73     text = ' '.join(text.split())
74     tokens = word_tokenize(text)
75     cleaned_tokens = [stemmer.stem(token) for token in tokens if token.lower() not in stop_words]
76     cleaned_text = ' '.join(cleaned_tokens)
77     return cleaned_text
78
79 df['Clean_Text'] = df['Text'].apply(clean)
80
81 specified_columns = ['Platform', 'Country', 'Year', 'Month', 'Day_of_Week']
82

```

Gambar 5. Preprocessing Data.

d. Analisis dan Visualisasi Sentimen

Pada tahap ini, kolom-kolom spesifik dieksplorasi dengan menghitung nilai unik dan frekuensinya. *SentimentIntensityAnalyzer* dari NLTK digunakan untuk menghitung skor sentimen dan mengkategorikan teks menjadi positif, negatif, atau netral. Visualisasi *pie chart* dibuat untuk menunjukkan distribusi sentimen dengan berbagai fitur seperti pewarnaan, teks otomatis, dan lingkaran tengah untuk memberikan informasi distribusi sentimen secara grafis dan menarik. Berikut gambarnya:

```

83 for col in specified_columns:
84     total_unique_values = df[col].nunique()
85     print(f'Total unique values for {col}: {total_unique_values}')
86
87 top_values = df[col].value_counts()
88 colors = ['firebrick', 'forestgreen', 'yellow', 'blue', 'magenta', 'cyan', 'white', 'lightblack', 'lightred', 'lightgreen']
89
90 for i, (value, count) in enumerate(top_values.items()):
91     color = colors[i % len(colors)]
92     print(f'{value}: {count} (Fore: {color})')
93
94 print('\n' + '=' * 38 + '\n')
95
96 df1 = df.copy()
97
98 analyzer = SentimentIntensityAnalyzer()
99 df1['Vader_Score'] = df1['Clean_Text'].apply(lambda text: analyzer.polarity_scores(text)['compound'])
100 df1['Sentiment'] = df1['Vader_Score'].apply(lambda score: 'positive' if score >= 0.05 else 'negative' if score <= -0.05 else 'neutral')
101
102 print(df[['Clean_Text', 'Vader_Score', 'Sentiment']].head())
103
104 colors = ['#d62728', '#9467bd', '#17becf']
105 explode = (0.1, 0, 0)
106
107 sentiment_counts = df1.groupby('Sentiment').size()
108 fig, ax = plt.subplots(figsize=(12, 8))
109 wedges, texts, autotexts = ax.pie(

```

Gambar 6. Analisis Dan Visualisasi Sentimen 1.

```

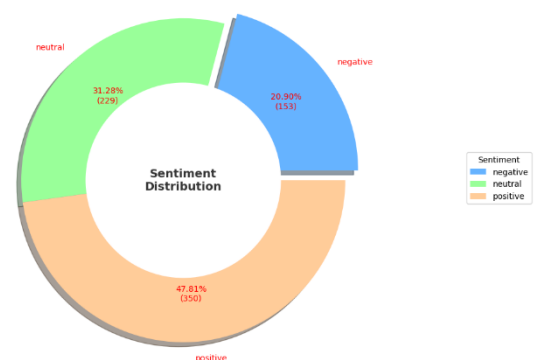
83 for col in specified_columns:
84     total_unique_values = df[col].nunique()
85     print(f'Total unique values for {col}: {total_unique_values}')
86
87 top_values = df[col].value_counts()
88 colors = ['firebrick', 'forestgreen', 'yellow', 'blue', 'magenta', 'cyan', 'white', 'lightblack', 'lightred', 'lightgreen']
89
90 for i, (value, count) in enumerate(top_values.items()):
91     color = colors[i % len(colors)]
92     print(f'{value}: {count} (Fore: {color})')
93
94 print('\n' + '=' * 38 + '\n')
95
96 df1 = df.copy()
97
98 analyzer = SentimentIntensityAnalyzer()
99 df1['Vader_Score'] = df1['Clean_Text'].apply(lambda text: analyzer.polarity_scores(text)['compound'])
100 df1['Sentiment'] = df1['Vader_Score'].apply(lambda score: 'positive' if score >= 0.05 else 'negative' if score <= -0.05 else 'neutral')
101
102 print(df[['Clean_Text', 'Vader_Score', 'Sentiment']].head())
103
104 colors = ['#d62728', '#9467bd', '#17becf']
105 explode = (0.1, 0, 0)
106
107 sentiment_counts = df1.groupby('Sentiment').size()
108 fig, ax = plt.subplots(figsize=(12, 8))
109 wedges, texts, autotexts = ax.pie(

```

Gambar 7. Analisis Dan Visualisasi Sentimen 2.

D. Tahapan Evaluasi

Evaluasi hasil distribusi sentimen meliputi pengumpulan data sentimen, pra-pemrosesan untuk membersihkan noise, dan klasifikasi menggunakan model *machine learning* atau NLP. Hasilnya divisualisasikan dalam bentuk grafik. Kinerja model dievaluasi menggunakan metrik seperti akurasi, presisi, dan *recall*. Akurasi mengukur seberapa sering prediksi benar, presisi mengukur proporsi prediksi positif yang benar, dan *recall* mengukur proporsi data aktual yang teridentifikasi dengan benar. Analisis lanjutan dilakukan untuk memahami penyebab sentimen negatif dan netral serta meningkatkan persepsi positif. Berikut hasilnya:



Gambar 8. Gambar Chart Hasil Evaluasi.

Gambar chart menunjukkan distribusi sentimen dari data yang telah dianalisis. Hasil

analisis menunjukkan bahwa 20.90% dari data memiliki sentimen negatif, yang diwakili oleh bagian berwarna biru, menandakan adanya sejumlah signifikan pengguna atau pelanggan yang merasa kurang puas. Sebanyak 31.28% dari data memiliki sentimen netral, ditandai dengan warna hijau muda, menunjukkan bahwa sebagian pengguna atau pelanggan tidak memiliki perasaan kuat baik positif maupun negatif. Mayoritas data, yaitu 47.81%, memiliki sentimen positif, yang diwakili oleh warna peach, menunjukkan bahwa sebagian besar pengguna atau pelanggan memiliki persepsi positif terhadap produk atau layanan. Artinya sebanyak 350 orang di media sosial mengutarakan sentiment positif, 229 orang mengutarakan sentimen netral dan 153 orang mengutarakan sentimen negatif. Dari distribusi ini, dapat disimpulkan bahwa meskipun sentimen positif dominan, perhatian perlu diberikan pada sentimen negatif dan netral untuk meningkatkan kepuasan keseluruhan.

#### IV. KESIMPULAN

Penelitian ini berhasil mengidentifikasi dan mengklasifikasikan sentimen positif, negatif, dan netral dari data media sosial menggunakan teknik *Natural Language Processing* (NLP). Hasil analisis menunjukkan bahwa 47.81% dari tweet yang dikumpulkan memiliki sentimen positif, 31.28% bersifat netral, dan 20.90% mengandung sentimen negatif. Teknik *preprocessing* seperti tokenisasi, *stemming*, dan *lemmatization* sangat berperan dalam membersihkan data dan meningkatkan akurasi model. Evaluasi model menggunakan metrik seperti akurasi, *precision*, *recall*, dan *F1-score* memastikan bahwa model memiliki kinerja yang baik dalam mengklasifikasikan sentimen.

Distribusi sentimen ini memberikan wawasan berharga tentang bagaimana isu tertentu dipersepsikan oleh publik di media sosial. Keberhasilan penelitian ini menunjukkan bahwa NLP efektif dalam mengolah data teks dari media sosial. Namun, tantangan seperti ambiguitas dalam teks, variasi bahasa dan slang, serta ironi dan sarkasme tetap perlu diatasi. Untuk peningkatan di masa depan, disarankan untuk

meningkatkan teknik *preprocessing*, menggunakan model yang lebih canggih, dan mengumpulkan data yang lebih luas dari berbagai platform media sosial. Kesimpulan dari penelitian ini dapat menjadi dasar untuk strategi komunikasi yang lebih efektif di media sosial serta untuk memahami dinamika opini publik secara lebih mendalam.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Universitas Duta Bangsa atas dukungan dana dan fasilitas yang diberikan dalam pelaksanaan penelitian ini. Ucapan terima kasih juga disampaikan kepada seluruh staf dan rekan di Universitas Duta Bangsa yang telah memberikan bantuan, masukan, dan dukungan selama proses penelitian dan penulisan manuskrip ini.

Penghargaan yang tulus juga penulis sampaikan kepada individu-individu yang telah berkontribusi dalam berbagai tahapan penelitian ini. Bantuan dan kerja sama yang diberikan telah memungkinkan penelitian ini dapat diselesaikan dengan baik. Penulis berharap hasil penelitian ini dapat memberikan kontribusi positif bagi pengembangan ilmu pengetahuan dan teknologi, khususnya dalam bidang *Natural Language Processing* (NLP).

#### REFERENSI

- [1] Herman Lawelai, Anwar Sadat, Ansar Suherman (2022). Demokrasi dan Kebebasan Berpendapat di Media Sosial: Analisis Sentimen di Twitter.
- [2] Arianti Primadhani Tirtopangarsa, Warih Maharani (2021). *Sentiment Analysis of Depression Detection on Twitter Social Media Users Using the K-Nearest Neighbor Method*.
- [3] Akhmad Muzaki, Arita Witanti (2021). *Sentiment Analysis of the Community in the Twitter to the 2020 Election in Pandemic COVID-19 by Method Naive Bayes Classifier*.
- [4] Jaka Aulia Pratama, Yadi Suprijadi, Zulhanif Zulhanif (2017). The Analisis Sentimen Sosial Media Twitter Dengan Algoritma Machine Learning Menggunakan Software R.
- [5] Muhammad Ariel Djamaludin, Agung Triayudi, Eri Mardiani (2022). Analisis Sentimen Tweet KRI Nanggala 402 di Twitter menggunakan Metode *Naive Bayes Classifier*.
- [6] Rahmat Hidayat, Budi Santoso (2020). Implementasi Algoritma SVM untuk Analisis Sentimen pada Media Sosial Twitter terkait Pilpres 2019.
- [7] Fitriani, Budi Setiawan (2019). Analisis Sentimen Pengguna Twitter terhadap Kebijakan Pendidikan di Indonesia dengan Menggunakan Metode *Support Vector Machine*.
- [8] Siti Zubaidah, Fauzi Rachman (2018). Klasifikasi Sentimen Opini Masyarakat Mengenai Transportasi Online di Twitter menggunakan Metode *K-Nearest Neighbor*.

- [9] Dian Anggraini, Rahman Nurhadi (2021). Analisis Sentimen Pengguna Media Sosial Twitter terhadap Isu Lingkungan Hidup menggunakan Algoritma Naïve Bayes.
- [10] Eka Priyanto, Mega Suryani (2019). Analisis Sentimen pada Media Sosial Twitter terhadap Film Indonesia menggunakan Metode *Decision Tree*.
- [11] Nurul Hikmah, Agus Wibowo (2022). Penggunaan Metode Random Forest untuk Analisis Sentimen Terhadap Pelayanan Kesehatan di Twitter.
- [12] Rizky Ananda, Farid Setiawan (2021). Analisis Sentimen Pengguna Twitter terhadap Vaksin COVID-19 Menggunakan Algoritma *Logistic Regression*.
- [13] Yulianti, Erwin Setiawan (2020). Analisis Sentimen Publik terhadap Kebijakan PSBB di Jakarta Menggunakan Metode Naïve Bayes.
- [14] Anton Prasetyo, Dwi Handoko (2021). Analisis Sentimen Netizen terhadap Pilkada 2020 di Twitter Menggunakan Metode SVM.
- [15] Tri Wahyuni, Indra Kusuma (2018). *Sentiment Analysis on Customer Reviews of E-commerce Websites Using Text Mining and Machine Learning Techniques*.