

task2

January 29, 2024

```
[1]: # 1) Data Exploration
```

```
import pandas as pd
column_names = ['target', 'ids', 'date', 'flag', 'user', 'text']
# Loading the dataset
df = pd.read_csv("training.1600000.processed.noemoticon.csv", names = _
    ↪column_names, encoding='ISO-8859-1')
df
```

```
[1]:
```

	target	ids	date	flag	\
0	0	1467810369	Mon Apr 06 22:19:45 PDT 2009	NO_QUERY	
1	0	1467810672	Mon Apr 06 22:19:49 PDT 2009	NO_QUERY	
2	0	1467810917	Mon Apr 06 22:19:53 PDT 2009	NO_QUERY	
3	0	1467811184	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	
4	0	1467811193	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	
...	
1599995	4	2193601966	Tue Jun 16 08:40:49 PDT 2009	NO_QUERY	
1599996	4	2193601969	Tue Jun 16 08:40:49 PDT 2009	NO_QUERY	
1599997	4	2193601991	Tue Jun 16 08:40:49 PDT 2009	NO_QUERY	
1599998	4	2193602064	Tue Jun 16 08:40:49 PDT 2009	NO_QUERY	
1599999	4	2193602129	Tue Jun 16 08:40:50 PDT 2009	NO_QUERY	
		user			text
0		_TheSpecialOne_	@switchfoot	http://twitpic.com/2y1zl	- Awww, t...
1		scotthamilton	is upset that he can't	update his	Facebook by ...
2		mattycus	@Kenichan	I dived many times	for the ball. Man...
3		ElleCTF	my whole body feels	itchy and like its	on fire
4		Karoli	@nationwideclass	no, it's not	behaving at all...
...	
1599995		AmandaMarie1028	Just woke up. Having	no school is the	best fee...
1599996		TheWDBboards	TheWDB.com - Very	cool to hear old	Walt interv...
1599997		bpbabe	Are you ready for	your MoJo	Makeover? Ask me f...
1599998		tinydiamondz	Happy 38th	Birthday to my	boo of alll time!!! ...
1599999		RyanTrevMorris	happy #charitytuesday	@theNSPCC	@SparksCharity...

```
[1600000 rows x 6 columns]
```

```
[2]: # Displaying the basic information about the dataset
print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1600000 entries, 0 to 1599999
Data columns (total 6 columns):
#   Column  Non-Null Count  Dtype
---  -
0   target  1600000 non-null    int64
1   ids     1600000 non-null    int64
2   date    1600000 non-null    object
3   flag    1600000 non-null    object
4   user    1600000 non-null    object
5   text    1600000 non-null    object
dtypes: int64(2), object(4)
memory usage: 73.2+ MB
None
```

```
[3]: # Displaying the first few rows of the dataset
print(df.head())
```

	target	ids	date	flag	\
0	0	1467810369	Mon Apr 06 22:19:45 PDT 2009	NO_QUERY	
1	0	1467810672	Mon Apr 06 22:19:49 PDT 2009	NO_QUERY	
2	0	1467810917	Mon Apr 06 22:19:53 PDT 2009	NO_QUERY	
3	0	1467811184	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	
4	0	1467811193	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	

	user	text
0	_TheSpecialOne_	@switchfoot http://twitpic.com/2y1z1l - Awww, t...
1	scotthamilton	is upset that he can't update his Facebook by ...
2	mattycus	@Kenichan I dived many times for the ball. Man...
3	ElleCTF	my whole body feels itchy and like its on fire
4	Karoli	@nationwideclass no, it's not behaving at all...

```
[4]: # 2) Data Cleaning

# (i) Dropping the irrelevant columns
# Drop unnecessary columns
# df = df[['target', 'text']]

# (i) Handling the missing values
df.dropna(inplace=True)

# (ii) Dropping the duplicate entries
df.drop_duplicates(inplace=True)
```

```
# (iii) Displaying the cleaned dataset
print(df.head())
```

	target	ids	date	flag	\
0	0	1467810369	Mon Apr 06 22:19:45 PDT 2009	NO_QUERY	
1	0	1467810672	Mon Apr 06 22:19:49 PDT 2009	NO_QUERY	
2	0	1467810917	Mon Apr 06 22:19:53 PDT 2009	NO_QUERY	
3	0	1467811184	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	
4	0	1467811193	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	

	user	text
0	_TheSpecialOne_	@switchfoot http://twitpic.com/2ylzl - Awww, t...
1	scotthamilton	is upset that he can't update his Facebook by ...
2	mattycus	@Kenichan I dived many times for the ball. Man...
3	ElleCTF	my whole body feels itchy and like its on fire
4	Karoli	@nationwideclass no, it's not behaving at all...

```
[5]: pip install seaborn
```

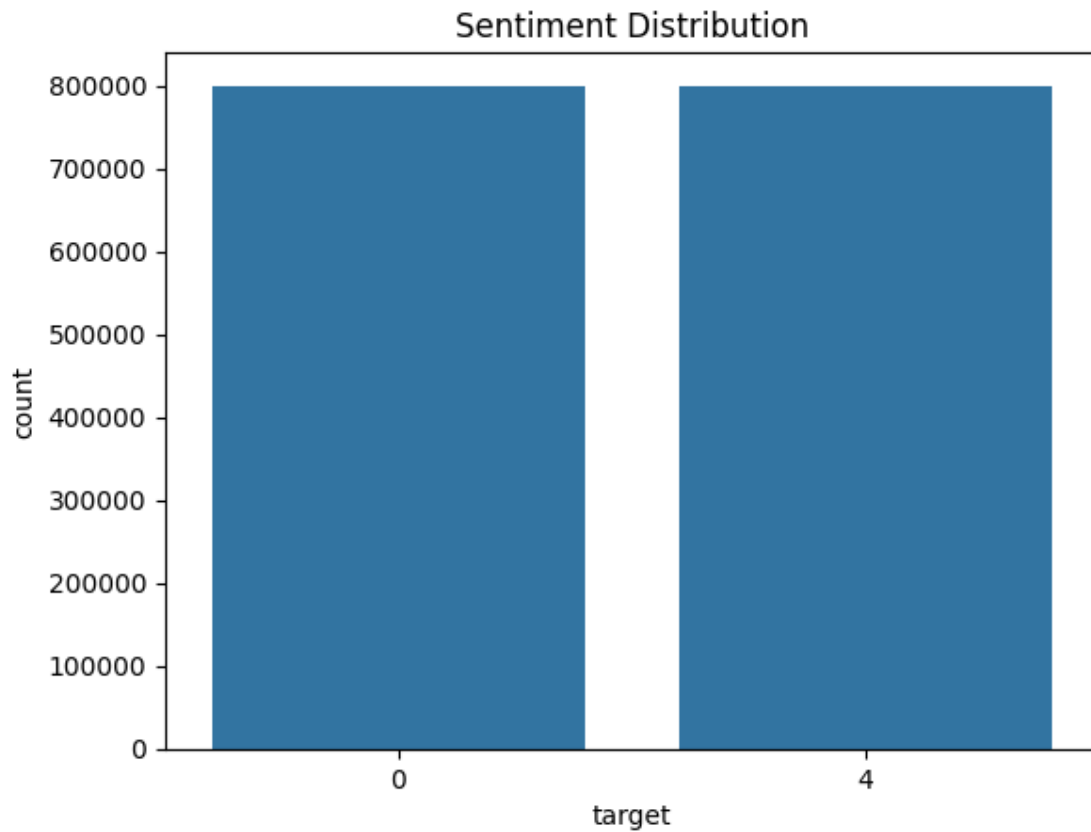
```
Requirement already satisfied: seaborn in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(0.13.1)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from seaborn) (1.26.2)
Requirement already satisfied: pandas>=1.2 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from seaborn) (2.1.4)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from seaborn) (3.8.2)
Requirement already satisfied: contourpy>=1.0.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from matplotlib!=3.6.1,>=3.4->seaborn) (1.2.0)
Requirement already satisfied: cycler>=0.10 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from matplotlib!=3.6.1,>=3.4->seaborn) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from matplotlib!=3.6.1,>=3.4->seaborn) (4.47.2)
Requirement already satisfied: kiwisolver>=1.3.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from matplotlib!=3.6.1,>=3.4->seaborn) (1.4.5)
Requirement already satisfied: packaging>=20.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from matplotlib!=3.6.1,>=3.4->seaborn) (23.2)
Requirement already satisfied: pillow>=8 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
```

```
(from matplotlib!=3.6.1,>=3.4->seaborn) (10.2.0)
Requirement already satisfied: pyparsing>=2.3.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from matplotlib!=3.6.1,>=3.4->seaborn) (3.1.1)
Requirement already satisfied: python-dateutil>=2.7 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from matplotlib!=3.6.1,>=3.4->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from pandas>=1.2->seaborn) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from pandas>=1.2->seaborn) (2023.3)
Requirement already satisfied: six>=1.5 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

```
[6]: # 3) Exploratory Data Analysis (EDA)
import matplotlib.pyplot as plt
import seaborn as sns

# Conducting EDA to gain initial insights
# (e.g., summary statistics, distribution of tweet lengths)

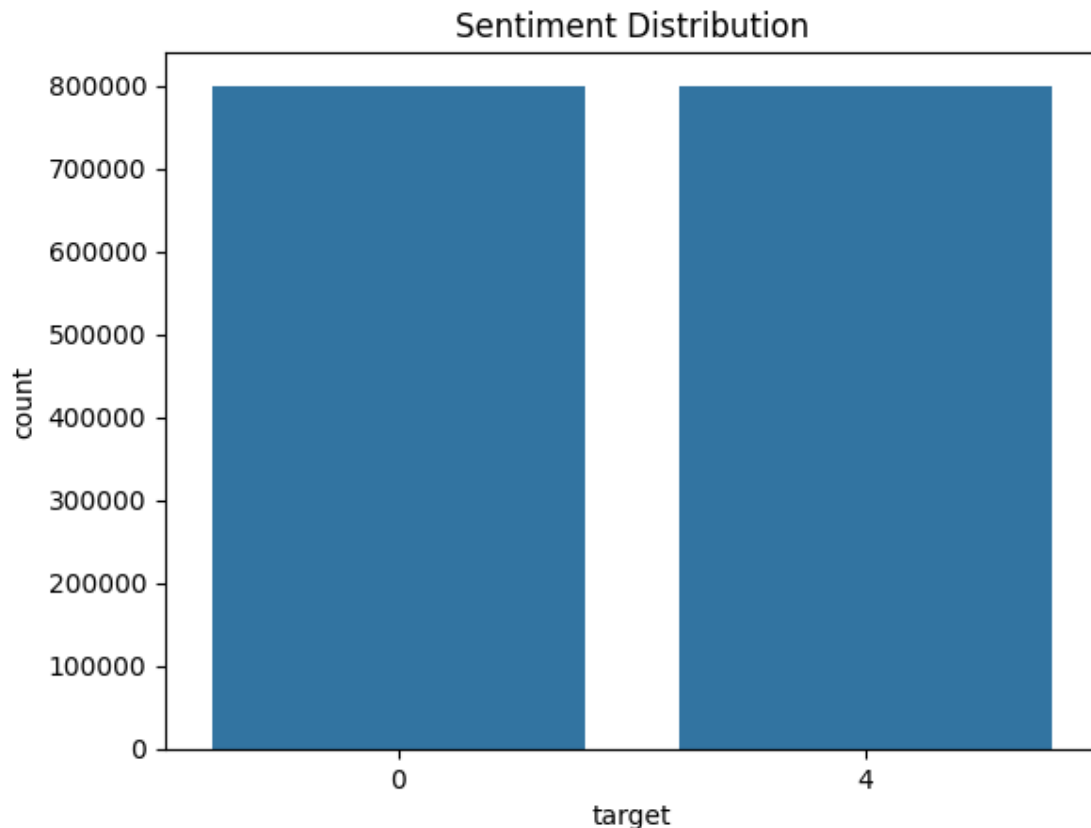
# Visualizing sentiment distribution
sns.countplot(x='target', data=df)
plt.title('Sentiment Distribution')
plt.show()
```



```
[7]: # 4) Sentiment Distribution

# Visualizing the distribution of sentiment labels
sns.countplot(x='target', data=df)
plt.title('Sentiment Distribution')
plt.show()

# Analyzing the balance of sentiment classes
sentiment_counts = df['target'].value_counts()
print(sentiment_counts)
```



```
target
0      800000
4      800000
Name: count, dtype: int64
```

```
[8]: pip install wordcloud
```

```
Requirement already satisfied: wordcloud in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(1.9.3)
Requirement already satisfied: numpy>=1.6.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from wordcloud) (1.26.2)
Requirement already satisfied: pillow in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from wordcloud) (10.2.0)
Requirement already satisfied: matplotlib in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from wordcloud) (3.8.2)
Requirement already satisfied: contourpy>=1.0.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
```

```
(from matplotlib->wordcloud) (1.2.0)
Requirement already satisfied: cyclcr>=0.10 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from matplotlib->wordcloud) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from matplotlib->wordcloud) (4.47.2)
Requirement already satisfied: kiwisolver>=1.3.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from matplotlib->wordcloud) (1.4.5)
Requirement already satisfied: packaging>=20.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from matplotlib->wordcloud) (23.2)
Requirement already satisfied: pyparsing>=2.3.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from matplotlib->wordcloud) (3.1.1)
Requirement already satisfied: python-dateutil>=2.7 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: six>=1.5 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from python-dateutil>=2.7->matplotlib->wordcloud) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

[9]: *# 5) Word Frequency Analysis*

```
from wordcloud import WordCloud

# Analyzing the word frequency in tweets
positive_tweets = df[df['target'] == 4]['text']
negative_tweets = df[df['target'] == 0]['text']

# Creating the word clouds for positive and negative sentiments
positive_wordcloud = WordCloud().generate(' '.join(positive_tweets))
negative_wordcloud = WordCloud().generate(' '.join(negative_tweets))

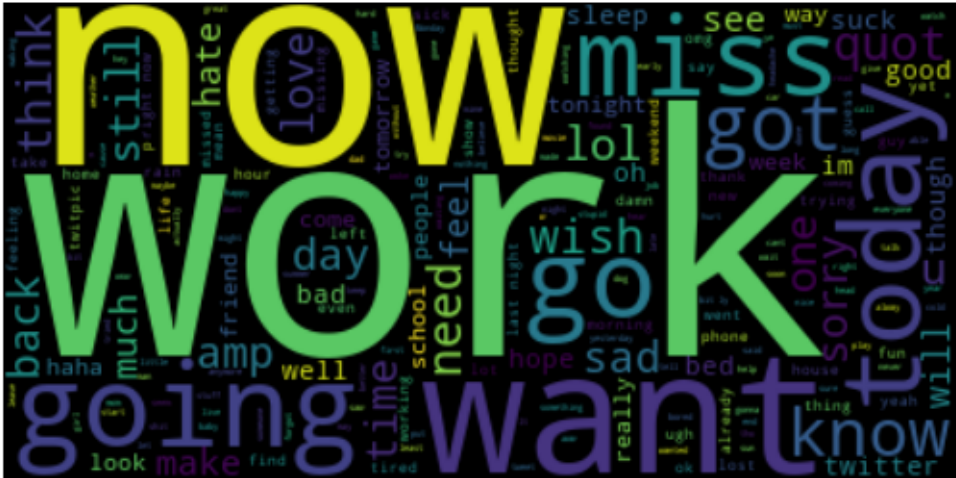
# Displaying the word clouds
plt.imshow(positive_wordcloud, interpolation='bilinear')
plt.title('Positive Word Cloud')
plt.axis('off')
plt.show()

plt.imshow(negative_wordcloud, interpolation='bilinear')
plt.title('Negative Word Cloud')
plt.axis('off')
plt.show()
```

Positive Word Cloud



Negative Word Cloud



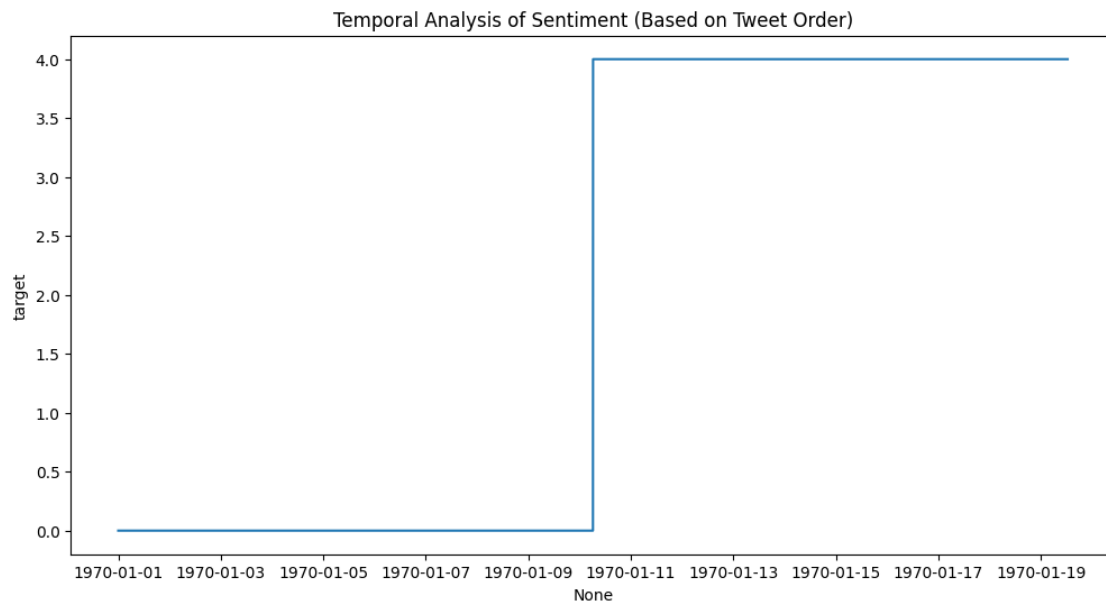
```
[10]: # 6) Temporal Analysis
```

```
# Assuming the index represents the order of tweets in the dataset
# Converting the index to datetime (assuming it's a numerical index)
df.index = pd.to_datetime(df.index, unit='s') # 's' assumes the index is in
↪seconds, adjust if needed

# Exploring how sentiment varies over the "pseudo-time" index
plt.figure(figsize=(12, 6))
sns.lineplot(x=df.index, y='target', data=df)
```



```
plt.title('Temporal Analysis of Sentiment (Based on Tweet Order)')
plt.show()
```



```
[11]: pip install nltk
```

```
Requirement already satisfied: nltk in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(3.8.1)
Requirement already satisfied: click in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from nltk) (8.1.7)
Requirement already satisfied: joblib in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from nltk) (1.3.2)
Requirement already satisfied: regex>=2021.8.3 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from nltk) (2023.12.25)
Requirement already satisfied: tqdm in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from nltk) (4.66.1)
Requirement already satisfied: colorama in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from click->nltk) (0.4.6)
Note: you may need to restart the kernel to use updated packages.
```

```
[12]: import nltk
nltk.download('punkt')
```

```
nltk.download('stopwords')
nltk.download('wordnet')
```

```
[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\Vishakha\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\Vishakha\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to
[nltk_data] C:\Users\Vishakha\AppData\Roaming\nltk_data...
[nltk_data] Package wordnet is already up-to-date!
```

[12]: True

[13]:

```
pip install pandas nltk scikit-learn matplotlib wordcloud
```

```
Requirement already satisfied: pandas in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(2.1.4)
Requirement already satisfied: nltk in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(3.8.1)
Requirement already satisfied: scikit-learn in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(1.4.0)
Requirement already satisfied: matplotlib in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(3.8.2)
Requirement already satisfied: wordcloud in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(1.9.3)
Requirement already satisfied: numpy<2,>=1.22.4 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from pandas) (1.26.2)
Requirement already satisfied: python-dateutil>=2.8.2 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from pandas) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from pandas) (2023.3)
Requirement already satisfied: click in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from nltk) (8.1.7)
Requirement already satisfied: joblib in
```

```

c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from nltk) (1.3.2)
Requirement already satisfied: regex>=2021.8.3 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from nltk) (2023.12.25)
Requirement already satisfied: tqdm in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from nltk) (4.66.1)
Requirement already satisfied: scipy>=1.6.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from scikit-learn) (1.12.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from scikit-learn) (3.2.0)
Requirement already satisfied: contourpy>=1.0.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from matplotlib) (1.2.0)
Requirement already satisfied: cycycler>=0.10 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from matplotlib) (4.47.2)
Requirement already satisfied: kiwisolver>=1.3.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from matplotlib) (1.4.5)
Requirement already satisfied: packaging>=20.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from matplotlib) (23.2)
Requirement already satisfied: pillow>=8 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from matplotlib) (10.2.0)
Requirement already satisfied: pyparsing>=2.3.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from matplotlib) (3.1.1)
Requirement already satisfied: six>=1.5 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from python-dateutil>=2.8.2->pandas) (1.16.0)
Requirement already satisfied: colorama in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from click->nltk) (0.4.6)
Note: you may need to restart the kernel to use updated packages.

```

```

[14]: import re
      from nltk.corpus import stopwords
      from nltk.tokenize import word_tokenize
      from nltk.stem import WordNetLemmatizer

```

```

from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, f1_score
import matplotlib.pyplot as plt

```

```

[15]: # 7) Text Preprocessing
stop_words = set(stopwords.words('english'))
lemmatizer = WordNetLemmatizer()

def preprocess_text(text):
    text = re.sub(r'http\S+', '', text) # Remove URLs
    text = re.sub(r'[^a-zA-Z\s]', '', text) # Remove special characters and
    ↪ numbers
    tokens = word_tokenize(text)
    tokens = [lemmatizer.lemmatize(token.lower()) for token in tokens if token.
    ↪ lower() not in stop_words]
    return ' '.join(tokens)
df2 = df.head(15000)
df2['processed_text'] = df2['text'].apply(preprocess_text)
df2
print("After Text Processing:- ")
df2['processed_text']

```

After Text Processing:-

C:\Users\Vishakha\AppData\Local\Temp\ipykernel_8832\397820761.py:12:

SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
df2['processed_text'] = df2['text'].apply(preprocess_text)
```

```

[15]: 1970-01-01 00:00:00    switchfoot awww thats bumner shoulda got david...
      1970-01-01 00:00:01    upset cant update facebook texting might cry r...
      1970-01-01 00:00:02    kenichan dived many time ball managed save res...
      1970-01-01 00:00:03                                whole body feel itchy like fire
      1970-01-01 00:00:04    nationwideclass behaving im mad cant see
      ...
      1970-01-01 04:09:55                                aubreyoday hey aubrey ya ya dk split
      1970-01-01 04:09:56    twaiting wife come stair knock tivo leaf tomor...
      1970-01-01 04:09:57                                yer hgistory presetnation
      1970-01-01 04:09:58                                really miss talking missy
      1970-01-01 04:09:59    dinner ok mussel appetizer followed chicken st...
      Name: processed_text, Length: 15000, dtype: object

```

```
[16]: # 8) Sentiment Prediction Model

X_train, X_test, y_train, y_test = train_test_split(df2['processed_text'],
    ↪df2['target'], test_size=0.2, random_state=42)

vectorizer = TfidfVectorizer(max_features=5000)
X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)

model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train_tfidf, y_train)

# Evaluate the model
y_pred = model.predict(X_test_tfidf)
accuracy = accuracy_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred, average='weighted')
print(f'Accuracy: {accuracy:.2f}')
print(f'F1 Score: {f1:.2f}')
```

Accuracy: 1.00

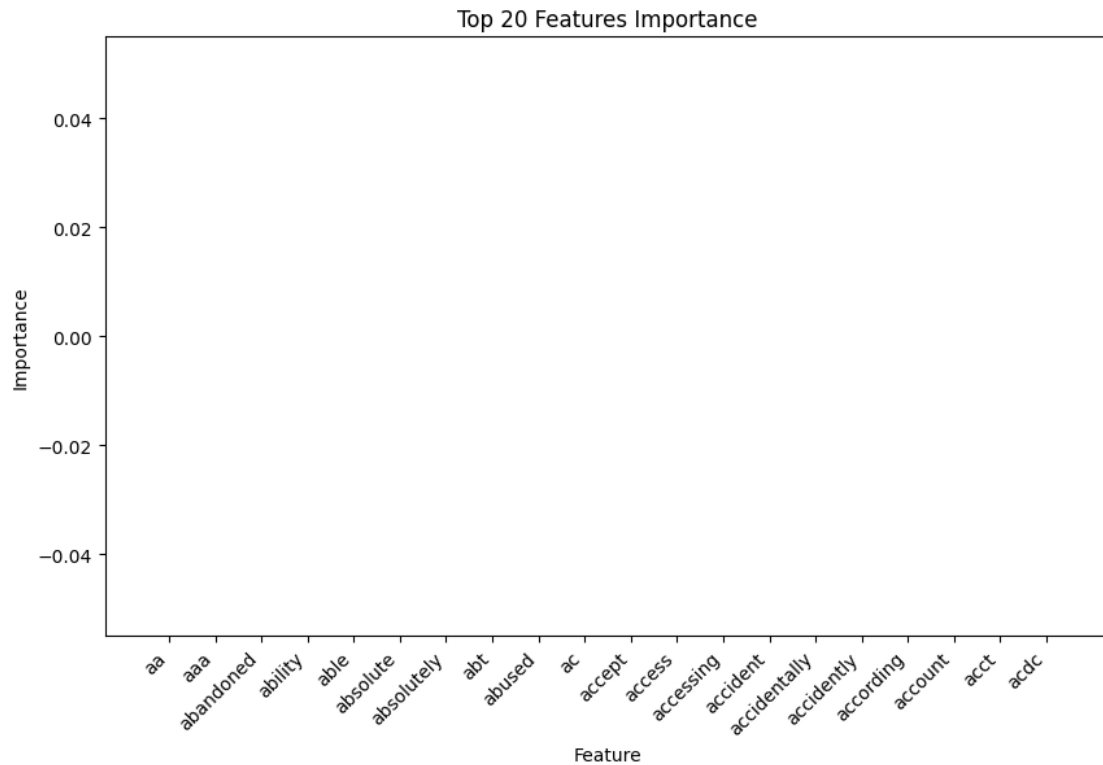
F1 Score: 1.00

```
[17]: # 9) Feature Importance

feature_names = vectorizer.get_feature_names_out()
feature_importance = model.feature_importances_

# Creating a DataFrame to store feature names and their importance scores
feature_df = pd.DataFrame({'Feature': feature_names, 'Importance':
    ↪feature_importance})
top_features = feature_df.nlargest(20, 'Importance')

# Visualizing feature importance
plt.figure(figsize=(10, 6))
plt.bar(top_features['Feature'], top_features['Importance'])
plt.xticks(rotation=45, ha='right')
plt.xlabel('Feature')
plt.ylabel('Importance')
plt.title('Top 20 Features Importance')
plt.show()
```



```
[18]: pip install streamlit
```

```
Requirement already satisfied: streamlit in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(1.30.0)
Requirement already satisfied: altair<6,>=4.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (5.2.0)
Requirement already satisfied: blinker<2,>=1.0.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (1.7.0)
Requirement already satisfied: cachetools<6,>=4.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (5.3.2)
Requirement already satisfied: click<9,>=7.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (8.1.7)
Requirement already satisfied: importlib-metadata<8,>=1.4 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (7.0.1)
Requirement already satisfied: numpy<2,>=1.19.3 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (1.26.2)
```

Requirement already satisfied: packaging<24,>=16.8 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (23.2)

Requirement already satisfied: pandas<3,>=1.3.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (2.1.4)

Requirement already satisfied: pillow<11,>=7.1.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (10.2.0)

Requirement already satisfied: protobuf<5,>=3.20 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (4.25.2)

Requirement already satisfied: pyarrow>=6.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (15.0.0)

Requirement already satisfied: python-dateutil<3,>=2.7.3 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (2.8.2)

Requirement already satisfied: requests<3,>=2.27 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (2.31.0)

Requirement already satisfied: rich<14,>=10.14.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (13.7.0)

Requirement already satisfied: tenacity<9,>=8.1.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (8.2.3)

Requirement already satisfied: toml<2,>=0.10.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (0.10.2)

Requirement already satisfied: typing-extensions<5,>=4.3.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (4.9.0)

Requirement already satisfied: tzlocal<6,>=1.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (5.2)

Requirement already satisfied: validators<1,>=0.2 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (0.22.0)

Requirement already satisfied: gitpython!=3.1.19,<4,>=3.0.7 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (3.1.41)

Requirement already satisfied: pydeck<1,>=0.8.0b4 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (0.8.1b0)

Requirement already satisfied: tornado<7,>=6.0.3 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (6.4)

Requirement already satisfied: watchdog>=2.1.5 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from streamlit) (3.0.0)

Requirement already satisfied: jinja2 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from altair<6,>=4.0->streamlit) (3.1.2)

Requirement already satisfied: jsonschema>=3.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from altair<6,>=4.0->streamlit) (4.20.0)

Requirement already satisfied: toolz in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from altair<6,>=4.0->streamlit) (0.12.0)

Requirement already satisfied: colorama in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from click<9,>=7.0->streamlit) (0.4.6)

Requirement already satisfied: gitdb<5,>=4.0.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from gitpython!=3.1.19,<4,>=3.0.7->streamlit) (4.0.11)

Requirement already satisfied: zipp>=0.5 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from importlib-metadata<8,>=1.4->streamlit) (3.17.0)

Requirement already satisfied: pytz>=2020.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from pandas<3,>=1.3.0->streamlit) (2023.3.post1)

Requirement already satisfied: tzdata>=2022.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from pandas<3,>=1.3.0->streamlit) (2023.3)

Requirement already satisfied: six>=1.5 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from python-dateutil<3,>=2.7.3->streamlit) (1.16.0)

Requirement already satisfied: charset-normalizer<4,>=2 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from requests<3,>=2.27->streamlit) (3.3.2)

Requirement already satisfied: idna<4,>=2.5 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from requests<3,>=2.27->streamlit) (3.6)

Requirement already satisfied: urllib3<3,>=1.21.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from requests<3,>=2.27->streamlit) (2.1.0)

Requirement already satisfied: certifi>=2017.4.17 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from requests<3,>=2.27->streamlit) (2023.11.17)

Requirement already satisfied: markdown-it-py>=2.2.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from rich<14,>=10.14.0->streamlit) (3.0.0)

Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from rich<14,>=10.14.0->streamlit) (2.17.2)


```

Requirement already satisfied: smmap<6,>=3.0.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from gitdb<5,>=4.0.1->gitpython!=3.1.19,<4,>=3.0.7->streamlit) (5.0.1)
Requirement already satisfied: MarkupSafe>=2.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from jinja2->altair<6,>=4.0->streamlit) (2.1.3)
Requirement already satisfied: attrs>=22.2.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from jsonschema>=3.0->altair<6,>=4.0->streamlit) (23.1.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from jsonschema>=3.0->altair<6,>=4.0->streamlit) (2023.11.2)
Requirement already satisfied: referencing>=0.28.4 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from jsonschema>=3.0->altair<6,>=4.0->streamlit) (0.32.0)
Requirement already satisfied: rpds-py>=0.7.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from jsonschema>=3.0->altair<6,>=4.0->streamlit) (0.15.2)
Requirement already satisfied: mdurl~=0.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from markdown-it-py>=2.2.0->rich<14,>=10.14.0->streamlit) (0.1.2)
Note: you may need to restart the kernel to use updated packages.

```

[19]: `!pip install ipywidgets`

```

Requirement already satisfied: ipywidgets in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(8.1.1)
Requirement already satisfied: comm>=0.1.3 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from ipywidgets) (0.2.0)
Requirement already satisfied: ipython>=6.1.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from ipywidgets) (8.18.1)
Requirement already satisfied: traitlets>=4.3.1 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from ipywidgets) (5.14.0)
Requirement already satisfied: widgetsnbextension~=4.0.9 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from ipywidgets) (4.0.9)
Requirement already satisfied: jupyterlab-widgets~=3.0.9 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from ipywidgets) (3.0.9)
Requirement already satisfied: decorator in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from ipython>=6.1.0->ipywidgets) (5.1.1)
Requirement already satisfied: jedi>=0.16 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages

```

```
(from ipython>=6.1.0->ipywidgets) (0.19.1)
Requirement already satisfied: matplotlib-inline in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from ipython>=6.1.0->ipywidgets) (0.1.6)
Requirement already satisfied: prompt-toolkit<3.1.0,>=3.0.41 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from ipython>=6.1.0->ipywidgets) (3.0.43)
Requirement already satisfied: pygments>=2.4.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from ipython>=6.1.0->ipywidgets) (2.17.2)
Requirement already satisfied: stack-data in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from ipython>=6.1.0->ipywidgets) (0.6.3)
Requirement already satisfied: exceptiongroup in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from ipython>=6.1.0->ipywidgets) (1.2.0)
Requirement already satisfied: colorama in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from ipython>=6.1.0->ipywidgets) (0.4.6)
Requirement already satisfied: parso<0.9.0,>=0.8.3 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from jedi>=0.16->ipython>=6.1.0->ipywidgets) (0.8.3)
Requirement already satisfied: wcwidth in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from prompt-toolkit<3.1.0,>=3.0.41->ipython>=6.1.0->ipywidgets) (0.2.12)
Requirement already satisfied: executing>=1.2.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from stack-data->ipython>=6.1.0->ipywidgets) (2.0.1)
Requirement already satisfied: asttokens>=2.1.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from stack-data->ipython>=6.1.0->ipywidgets) (2.4.1)
Requirement already satisfied: pure-eval in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from stack-data->ipython>=6.1.0->ipywidgets) (0.2.2)
Requirement already satisfied: six>=1.12.0 in
c:\users\vishakha\appdata\local\programs\python\python310\lib\site-packages
(from asttokens>=2.1.0->stack-data->ipython>=6.1.0->ipywidgets) (1.16.0)
```

```
[20]: # 10. User Interface:
import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.naive_bayes import MultinomialNB
from ipywidgets import widgets, interact
from IPython.display import display

# Load the dataset
```

```

df = pd.read_csv('training.1600000.processed.noemoticon.csv', header=None,
    ↪encoding='ISO-8859-1')
df.columns = ['target', 'id', 'date', 'query', 'user', 'text']

# Remove unnecessary columns
df = df[['target', 'text']]

# Map target values to sentiment labels
df['sentiment'] = df['target'].map({0: 'negative', 2: 'neutral', 4: 'positive'})

# Clean the text data (remove URLs, special characters, etc.)
def clean_text(text):
    # Your text cleaning code here
    # For simplicity, let's assume we only want to convert text to lowercase
    cleaned_text = text.lower()
    return cleaned_text

df['cleaned_text'] = df['text'].apply(clean_text)

# Convert text data to numerical features using CountVectorizer
vectorizer = CountVectorizer()
X_vectorized = vectorizer.fit_transform(df['cleaned_text'])
y = df['sentiment']

# Train a Naive Bayes classifier
model = MultinomialNB()
model.fit(X_vectorized, y)

# Function to predict sentiment for input text
def predict_sentiment(text):
    cleaned_text = clean_text(text)
    vectorized_text = vectorizer.transform([cleaned_text])
    prediction = model.predict(vectorized_text)
    return prediction[0]

# User Interface
def analyze_sentiment(input_text):
    result = predict_sentiment(input_text)
    print(f"Sentiment Prediction: {result}")

# Create interactive widget
text_input = widgets.Text(description="Enter text:", value="Lyx is cool")
interact(analyze_sentiment, input_text=text_input)

```

```

interactive(children=(Text(value='Lyx is cool', description='Enter text:'),
    ↪Output()), _dom_classes=('widget-i...

```

```
[20]: <function __main__.analyze_sentiment(input_text)>
```