MIT-World Peace University (MIT-WPU)
Faculty of Engineering, School of Computer Engineering & Technology

Course Name: NLP Course Code: CET40005B

T.Y. (23-24) NLP Project Report

Group Participants	1032201467 Rijul Bilaiya 1032202181 Abdul Faisal 1032202209 Mrunal Dande
Date	11th December 2023
Topic	Sentiment Analysis of FIFA World Cup 2022 Twitter Reviews
Problem Statement	Evaluating the tone of tweets on the FIFA World Cup 2022 is the work at hand. We hope to determine the general public's opinion on the tournament by examining the emotions expressed in these tweets. We will be able to comprehend the prevalent attitudes and opinions of Twitter users on different parts of the World Cup thanks to this study, which will categorize tweets into good, negative, or neutral feelings. The aim is to derive useful information that may help understand the responses of the audience, pinpoint important debate points, and eventually offer a thorough picture of how the public has responded to the event on social media.
Objectives	The main objective is to investigate the general opinions shared on Twitter about the FIFA World Cup 2022. This entails classifying tweets as neutral, negative, or positive in order to represent the overall sentiment around the competition. Through the identification of dominant attitudes and topics in these conversations, the aim is to provide insightful analysis on how the general public views different facets of the World Cup. For stakeholders, organizers, and sponsors, these insights will act as a

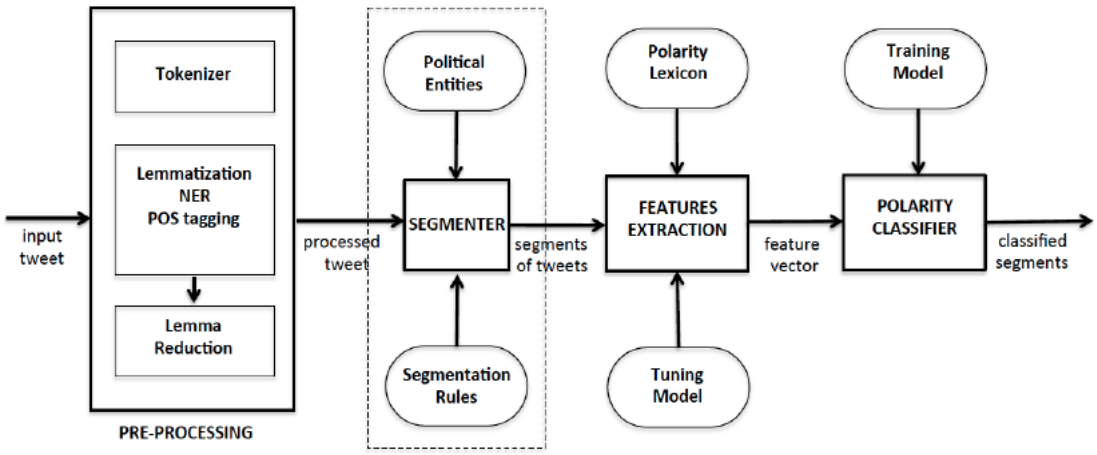
	compass, offering a detailed grasp of public opinion to guide decision-making, enhance event experiences, and perhaps impact future initiatives. Sentiment study of Twitter reviews for the FIFA World Cup 2022 has the following goals: 1. Recognise Public Perception: Learn how people on Twitter view and respond to the FIFA World Cup 2022. 2. Sort Sentiments: To gauge the general emotional atmosphere around the competition, divide tweets into three categories: positive, negative, and neutral. 3. Determine Trends and Themes: To ascertain which components of the World Cup are attracting the greatest interest or response, identify recurring themes, developing trends, and important issues within debates. 4. Offer Useful Insights: Give sponsors, organizers, and stakeholders a comprehensive grasp of public opinion to support their future event planning, marketing plans, and decision-making procedures. 5. Boost Engagement and Experience: Make use of the information acquired to perhaps boost audience participation, resolve issues, and enhance the overall experience.
Motivation	A deep-seated interest to capture the wide range of emotions woven into the fabric of this global spectacle lays at the core of exploring the thoughts expressed on Twitter on the FIFA World Cup 2022. It's an effort to interpret the mood of the group, whether it's the roar of joy that reverberates via joyful tweets, the nuanced criticism that shapes conversations, or the subtle undercurrents that expose the nuances of how people view and interact with this esteemed event. This project aims to uncover the rich tapestry of public feeling and comprehend the highs, lows, and everything in between that shape the World Cup experience for millions of people globally. It is not simply about statistics or data points. The goal is to turn these feelings into useful insights by sifting through this sea of tweets and providing a broad

	picture of how the competition appeals to individuals with various origins, ethnicities, and viewpoints. These insights serve as a compass for sponsors, organizers, and stakeholders when making decisions; they are more than just observations. They provide a road map for negotiating the complexity of public opinion, which may be used to improve event experiences, strategies, and possibly even the course of future international athletic extravaganzas. In the end, this sentiment analysis aims to strengthen ties, expand interactions, and create a strong emotional relationship between the World Cup and its viewers throughout the world. The goal is to elevate the experience beyond the game itself and instill it in the hearts and minds of people everywhere by transforming unfiltered sentiment data into a compelling story that connects with fans on a deep level.
Description	1. Sentiment analysis, often called opinion mining, is like diving into the emotions hidden within text. It's similar to inferring someone's emotions from their writing. Determining whether the material sounds good, negative, or neutral is the major goal here. 2. Sentiment analysis, often known as opinion mining, is akin to delving into the feelings concealed in written communication. It's similar to inferring someone's emotions from their writing. Determining whether the material sounds good, negative, or neutral is the major goal here. 3. We begin by selecting the most crucial passages from the book. These might be particular words or phrases, or even the construction of sentences themselves—elements that aid in expressing feelings or viewpoints. 4. The truly brilliant portion, Sentiment Classification, is what follows. This entails determining if a text sounds joyful, sad, furious, or merely neutral with the use of sophisticated computer programmers or algorithms that learn from a large number of instances. They examine

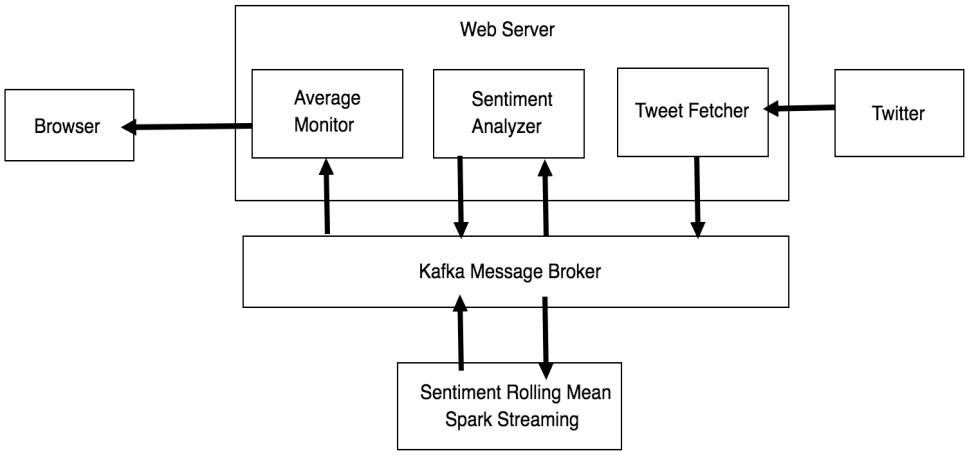
	the selected portions from the past and utilize them to slot
Keywords (max 4 words)	Feature Extraction , Sentiment Classification , Sentiment , Multi Domain Analysis.

Functional Requirements	The following are the necessary functions: 1. **Receiving Twitter Messages:** a. View Posts on Twitter: Identify a method to obtain every World Cup tweet that has been posted on Twitter. b. Clear Out Data: Take out any repetitive tweets or other information to concentrate just on what individuals are saying. 2. **Knowing How to Feel:** a. Interpreting Emotions: Determine if tweets on the World Cup are joyful, sad, or merely neutral by using tools. b. Sorting Words: Select the terms that convey a favorable, negative, or neutral emotion from the tweets. 3. **Presenting Findings:** a. Produce Readable Information: Create graphs or charts that illustrate whether individuals are mostly outraged, joyful, or neutral about the World Cup. b. Emphasize Significant Words: Determine which words are evoking strong emotions and list them. 4. **Verifying whether it's Correct:** a. Test Precision: To check if the machine is processing tweets correctly, check a few of them manually. b. Continue to Improve: Modify the software to improve its ability to interpret emotions in tweets. 5. **Ensure It Runs Quickly and Securely:** a. Suits a Large Number of Tweets: Verify that the programme can process a large number of tweets without experiencing any lag. b. Information Security: Respect people's privacy rights and keep information and tweets safe. 6. **Assistance in Problem-Solving:** a. Stay Up to Date: Ensure that the programme maintains its quality as circumstances change. b. Assist if Something Goes Wrong: Be prepared to assist if there

	are issues with the programme or if someone requires help using it.
Introduction	Examining the opinions shared on Twitter on the FIFA World Cup 2022 has been an interesting experience! We've been examining the opinions of the public about the tournament by sorting through hundreds of tweets. We've collected an abundance of data on what people liked, despised, or just thought was alright about the games, spanning from the early group stage contests to the suspenseful final. We have employed some really amazing techniques to help us make sense of this sea of tweets. We eliminated all emojis and links, categorized all emotions into good, bad, and neutral categories, and even dissected feelings towards certain players, teams, and historical events. The aim? To truly understand what fans were feeling let down by and what made them feel furious or unhappy Sentiment analysis, often known as opinion mining, is a potent natural language processing (NLP) application that seeks to identify and classify the attitudes, emotions, or views conveyed in textual data. Social media postings, product reviews, news stories, client testimonials, and more might all be included in this paragraph. Sentiment analysis's primary goal is to identify the underlying sentiment polarity, or whether a text represents neutral, positive, or negative sentiments. Businesses, researchers, and organizations looking to understand consumer happiness, market trends, public opinion, and general attitude towards certain goods, services, brands, or subjects will find it to be an indispensable tool. 1. Data Collection: Gathering textual data from various sources to be analyzed. 2. Preprocessing: Cleaning and preparing the text data by removing noise, punctuation, special characters, and converting text to a standardized format suitable for analysis.

	3. Feature Extraction: Identifying relevant features such as words, phrases, or semantic elements that contribute to expressing sentiment. 4. Sentiment Classification: Utilizing machine learning algorithms, lexicon-based approaches, or rule-based systems to classify text into predefined sentiment categories
System Desig:- **1. Simple block diagram**	

2. System architecture



Literature Survey: (for minimum 8 quality papers)

There are several procedures involved in analyzing Twitter conversation around the FIFA World Cup 2022. I'll start by searching for tweets that contain particular phrases, such as #FIFAWorldCup2022 or team names. After that, I'll edit the content to remove any links, emoticons, and superfluous symbols. I'll then separate each word in the tweets and eliminate any frequent ones, such "and" or "the." After that, I'll employ a variety of methods to determine the basic sentiment: some involve looking up word definitions, some involve complex math. After that, I'll map out the various emotions to observe how they evolve over time or at pivotal points in the competition. Dealing with sarcasm and jokes will be a little challenging, but I'll do my best to understand people's true feelings.

Paper Reference	Title of Paper	Year	Method used in Paper	Results Achieved	Gap
[1]	Opinion Mining and Sentiment Analysis: Foundations and Trends in Information Retrieval	2008	Lexicon-Based Approaches	analyzing opinions and sentiments expressed in textual data.	Different Languages have different accuracy.

[2]	Sentiment Analysis and Opinion Mining	2012	MLT	Extended Context	Dataset too small
[3]	Recursive Deep Models for Semantic Over a Sentiment Treebank	2013	RNN	Fine-Grained Sentiment Analysis	-
[4]	CNNs for Sentence Classification	2014	CNN	Efficiency and Simplicity	Interpretabili ty-
[5]	VADER (Valence Aware Dictionary and sEntiment Reasoner)	2014	Sentiment Intensity Scoring	Real-Time Analysis , Designed for Social Media	Enhancing Nuance Understanding

[6]	Attention-Based LSTM for Aspect-Level Sentiment	2016	Hierarchical Attention Networks	Improved Contextual Understanding	Handling Aspect Shifts and Coherence
[7]	Sarcasm Detection on Twitter: A Behavioral Modeling Approach	2011	Behavioral Modeling	Sarcasm Detection Accuracy , Identification of Sarcasm Cues	Contextual Nuances
[8]	A Comparative Study on Sentiment Analysis of Reviews	2011	Evaluation Metrics	Identification of Effective Methods	Limited Generalizability

Summary of Research Gap:

Depending on the area of study, there might be a broad range of research gaps regarding Twitter. A few possible study gaps in many disciplines pertaining to Twitter research are as follows:

- Recognizing Misinformation's Spread: There is study on the propagation of fake news on Twitter, but it may not completely explain why certain incorrect information becomes popular, the impact of particular user groups, or how algorithms magnify disinformation. The ethical considerations surrounding the use of Twitter data for research, such as those pertaining to user permission, privacy, and the responsible use of sensitive information, may not have been sufficiently examined in previous studies.

- The Effects of Twitter on Mental Health: Research on the precise effects of Twitter on mental health, including how overuse, interactions online, and exposure to certain information on the platform affect wellbeing, may be lacking.

- Twitter Crisis Communication Strategies: Although research has been done on the platform's use in times of emergency, there may still be gaps in our knowledge of the best ways for businesses and people to communicate on Twitter during a crisis.

- Cultural Differences on Twitter: The ways in which cultural variations affect the sentiment, communication patterns, and network structures on Twitter among users from various languages and geographical areas may not always be thoroughly explored by research.

- Algorithmic Bias and Echo Chambers: Further research may be necessary to determine how Twitter's algorithms may affect the dissemination of biased information or the development of echo chambers.

- Long-Term Twitter Trends and Shifts: While some study examines Twitter trends in real time, there may be gaps in knowledge on longer-term changes in user behavioral, content, or engagement patterns.

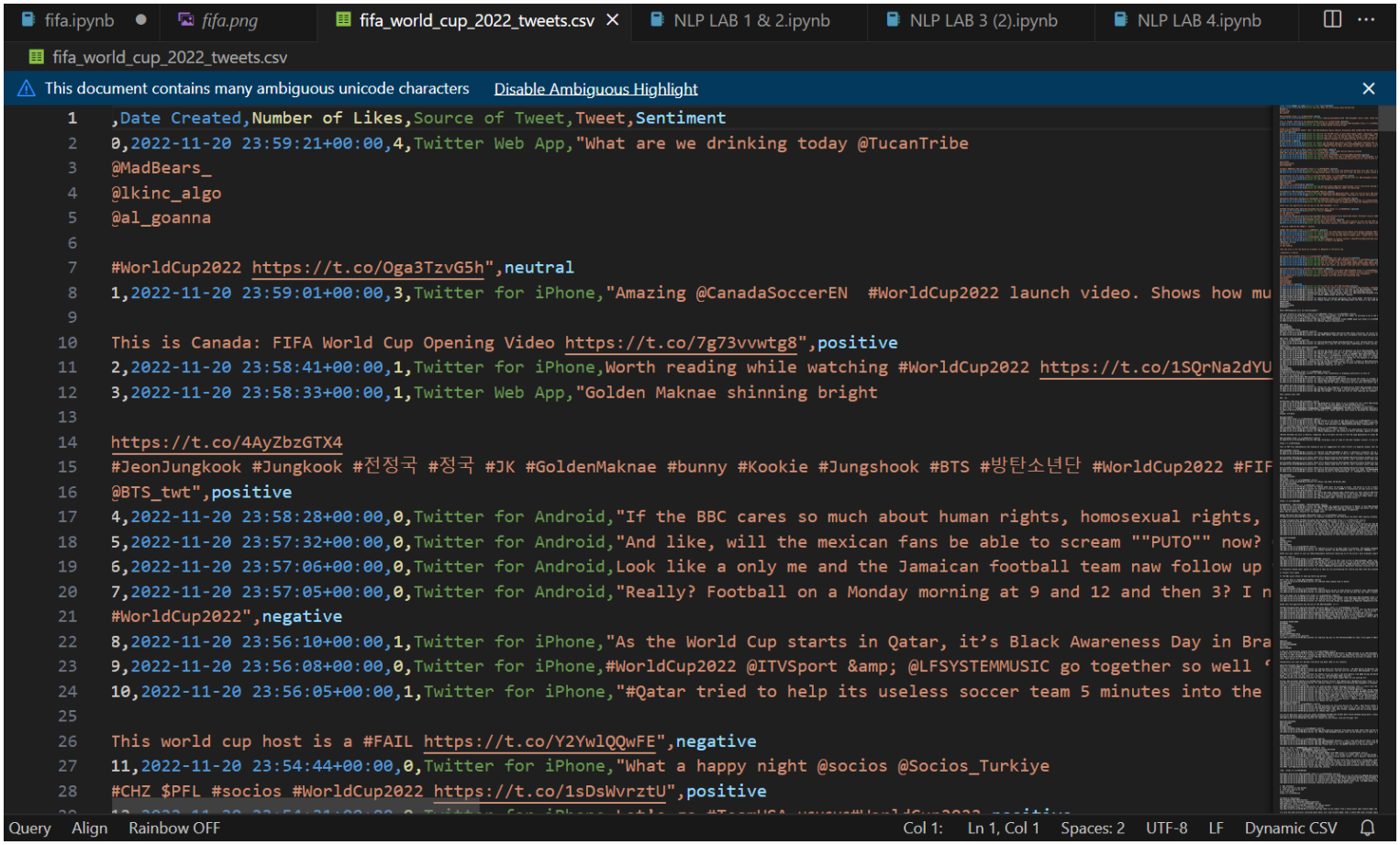
- Filling up these gaps might improve user experiences, increase our knowledge of Twitter's social function, and even encourage more responsible and informed usage of social media.

Corpus Used :

Dataset 1:

Name: fifa_world_cup_2022_tweets

Description: This dataset collects tweets from the major worldwide athletic events that will take place in 2022, with a particular emphasis on the Olympics, FIFA World Cup, and other international championships. It contains tweets collected with particular event hashtags, team names, player mentions, and event-related phrases. These tweets address a wide range of subjects, including fan reactions, debates, live match commentary, and more general talks about the events. They provide a comprehensive picture of the public's attitudes, feelings, and interactions throughout these sporting extravaganzas across a variety of languages, places, and eras. The dataset attempts to facilitate social network research, sentiment analysis, and theme exploration, offering insightful information on how people interact and express themselves on Twitter during these much awaited sporting events.



Operations performed: Got this dataset on kaggle then we have downloaded that csv file. Data was extracted from Twitter.

CODE & OUTPUT:

```python
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import matplotlib.pyplot as plt
import datetime
# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input directory
from wordcloud import WordCloud, STOPWORDS
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
import os
for dirname, _, filenames in os.walk('C:\Code'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

[1] ✓ 4.1s Python

```
C:\Code\fifa.ipynb
C:\Code\fifa.png
C:\Code\fifa2022.csv.csv
C:\Code\FIFAWORLDCUP.ipynb
C:\Code\fifa_world_cup_2022_tweets.csv
C:\Code\main.py
C:\Code\NLP 6.py
C:\Code\NLP LAB 1 & 2.ipynb
C:\Code\NLP LAB 3 (2).ipynb
C:\Code\NLP LAB 4.ipynb
```

Cell 16 of 22

```python
import pandas as pd

# Corrected file path
file_path = 'C:/Code/fifa_world_cup_2022_tweets.csv'

# Read the CSV file
df = pd.read_csv(file_path)
df['Date Created'] = pd.to_datetime(df['Date Created'], format='%Y-%m-%d %H:%M:%S%z')

#df = pd.read_csv("C:\Code\fifa_world_cup_2022_tweets.csv")
```

[2] ✓ 0.2s Python

```python
df.head()
```

[3] ✓ 0.0s Python

	Unnamed: 0	Date Created	Number of Likes	Source of Tweet	Tweet	Sentiment
0	0	2022-11-20 23:59:21+00:00	4	Twitter Web App	What are we drinking today @TucanTribe \n@MadB...	neutral
1	1	2022-11-20 23:59:01+00:00	3	Twitter for iPhone	Amazing @CanadaSoccerEN #WorldCup2022 launch ...	positive
2	2	2022-11-20 23:58:41+00:00	1	Twitter for iPhone	Worth reading while watching #WorldCup2022 htt...	positive

```
import pandas as pd

# Assuming 'Date Created' is the column containing date/time information
# Try parsing datetime with inferred format and handling timezones
df['Date Created'] = pd.to_datetime(df['Date Created'], errors='coerce', utc=True)
```

[4] ✓ 0.0s Python

```
df.shape
```

[5] ✓ 0.0s Python

```
(22524, 6)
```

```
df.isnull( ).sum( )
```

[6] ✓ 0.0s Python

```
Unnamed: 0         0
Date Created       0
Number of Likes    0
Source of Tweet    0
Tweet              0
Sentiment          0
dtype: int64
```

```
df['Source of Tweet'].value_counts().head(10)
```

[7] ✓ 0.0s Python

```
Source of Tweet
Twitter for iPhone          9507
Twitter for Android         6820
Twitter Web App             4505
TweetDeck                    386
Twitter for iPad             240
Qualtrics Social Connect     165
Hootsuite Inc.               146
Buffer                        71
Instagram                     62
Tweetbot for iOS              52
Name: count, dtype: int64
```

```
source_t =  df['Source of Tweet'].value_counts()
source_t.head(10).plot.bar()

plt.show()
```

[8] ✓ 0.2s Python

```
most_liked = df.sort_values(by=['Number of Likes'], ascending=False)
most_liked.head(10)
```

[9] ✓ 0.0s Python

	Unnamed: 0	Date Created	Number of Likes	Source of Tweet	Tweet	Sentiment
1287	1287	2022-11-20 19:39:11+00:00	316867	Twitter for iPhone	I can't express my gratitude and happiness for...	positive
17359	17359	2022-11-20 12:41:31+00:00	31517	Twitter Media Studio	Football Legend Eric Cantona reminds football ...	neutral
21267	21267	2022-11-20 06:17:00+00:00	20016	Twitter Web App	⚽ #BitKeep FootBall Carnival Main Event is co...	positive
528	528	2022-11-20 22:20:55+00:00	5847	TweetDeck	WATCH: BTS's #Jungkook Performs At #WorldCup20...	neutral
20073	20073	2022-11-20 09:05:12+00:00	5555	Twitter for iPhone	The 👑 Leo with Louis Vuitton 💧 #WorldCup2022 h...	positive
15332	15332	2022-11-20 14:57:02+00:00	5484	Twitter Web App	Need them at the World Cup opening ceremony 😘\...	neutral
20101	20101	2022-11-20 09:02:09+00:00	5430	TweetDeck	WATCH: #BTS's #Jungkook Shares His Support For...	neutral
212	212	2022-11-20 23:06:08+00:00	3172	Twitter Media Studio	"It's Palestine..."\n\nLebanese fans refuse to...	negative

	Unnamed: 0	Date Created	Number of Likes	Source of Tweet	Tweet	Sentiment
1287	1287	2022-11-20 19:39:11+00:00	316867	Twitter for iPhone	I can't express my gratitude and happiness for...	positive
17359	17359	2022-11-20 12:41:31+00:00	31517	Twitter Media Studio	Football Legend Eric Cantona reminds football ...	neutral
21267	21267	2022-11-20 06:17:00+00:00	20016	Twitter Web App	⚽ #BitKeep FootBall Carnival Main Event is co...	positive
528	528	2022-11-20 22:20:55+00:00	5847	TweetDeck	WATCH: BTS's #Jungkook Performs At #WorldCup20...	neutral
20073	20073	2022-11-20 09:05:12+00:00	5555	Twitter for iPhone	The 👑 Leo with Louis Vuitton 💧 #WorldCup2022 h...	positive
15332	15332	2022-11-20 14:57:02+00:00	5484	Twitter Web App	Need them at the World Cup opening ceremony 😘\...	neutral
20101	20101	2022-11-20 09:02:09+00:00	5430	TweetDeck	WATCH: #BTS's #Jungkook Shares His Support For...	neutral
212	212	2022-11-20 23:06:08+00:00	3172	Twitter Media Studio	"It's Palestine..."\n\nLebanese fans refuse to...	negative

1635	1635	2022-11-20 19:00:00+00:00	2428	Sprinklr Publishing	.@reymysterio, @fightbobby, @BiancaBelairWWE a...	neutral
5761	5761	2022-11-20 16:23:30+00:00	2370	Twitter for Android	Qatar's Goalkeeper 😂\n\n#WorldCup2022 https://...	neutral

```
df.iloc[1287].Tweet
```

[10] ✓ 0.0s Python

```
'I can’t express my gratitude and happiness for my participation in the biggest event of all times the World Cup in my country #Q
```

```
df.iloc[20101].Tweet
```

[11] ✓ 0.0s Python

```
"WATCH: #BTS's #Jungkook Shares His Support For South Korea's National Soccer Team At The #WorldCup2022 \nhttps://t.co/SHS3bmi60F
```

```
df.iloc[212].Tweet
```

[12] ✓ 0.0s Python

```
'"It\'s Palestine..."\n\nLebanese fans refuse to be interviewed by an Israeli TV reporter after he said that he is from \'Israel\
```

```
tfidf = TfidfVectorizer(max_features=25000)  # You can adjust max_features
X = tfidf.fit_transform(df['Tweet'])
y = df['Sentiment']

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Train a Logistic Regression model
model = LogisticRegression(max_iter=1000)
model.fit(X_train, y_train)

# Make predictions on the test set
predictions = model.predict(X_test)

# Calculate accuracy (just for reference)
accuracy = accuracy_score(y_test, predictions)
print(f"Accuracy of the model: {accuracy}")
```

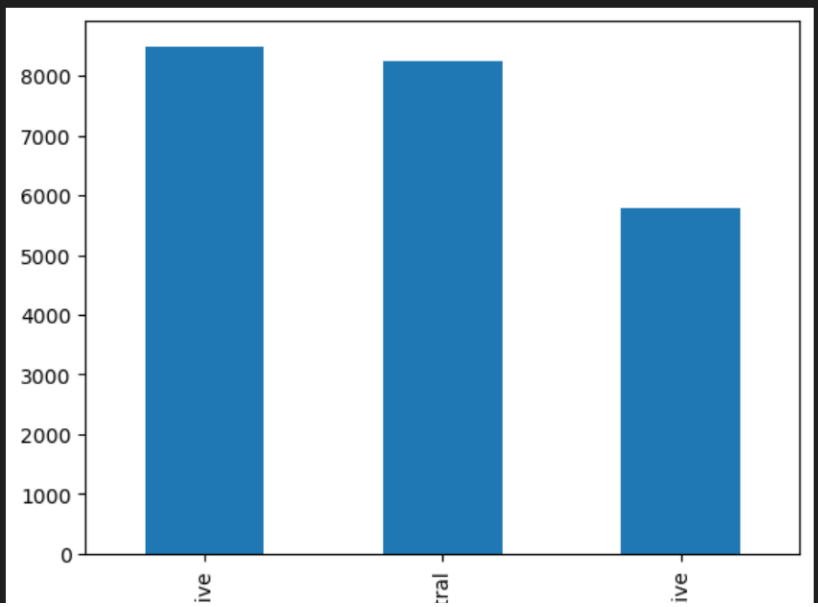
[13] ✓ 2.8s Python

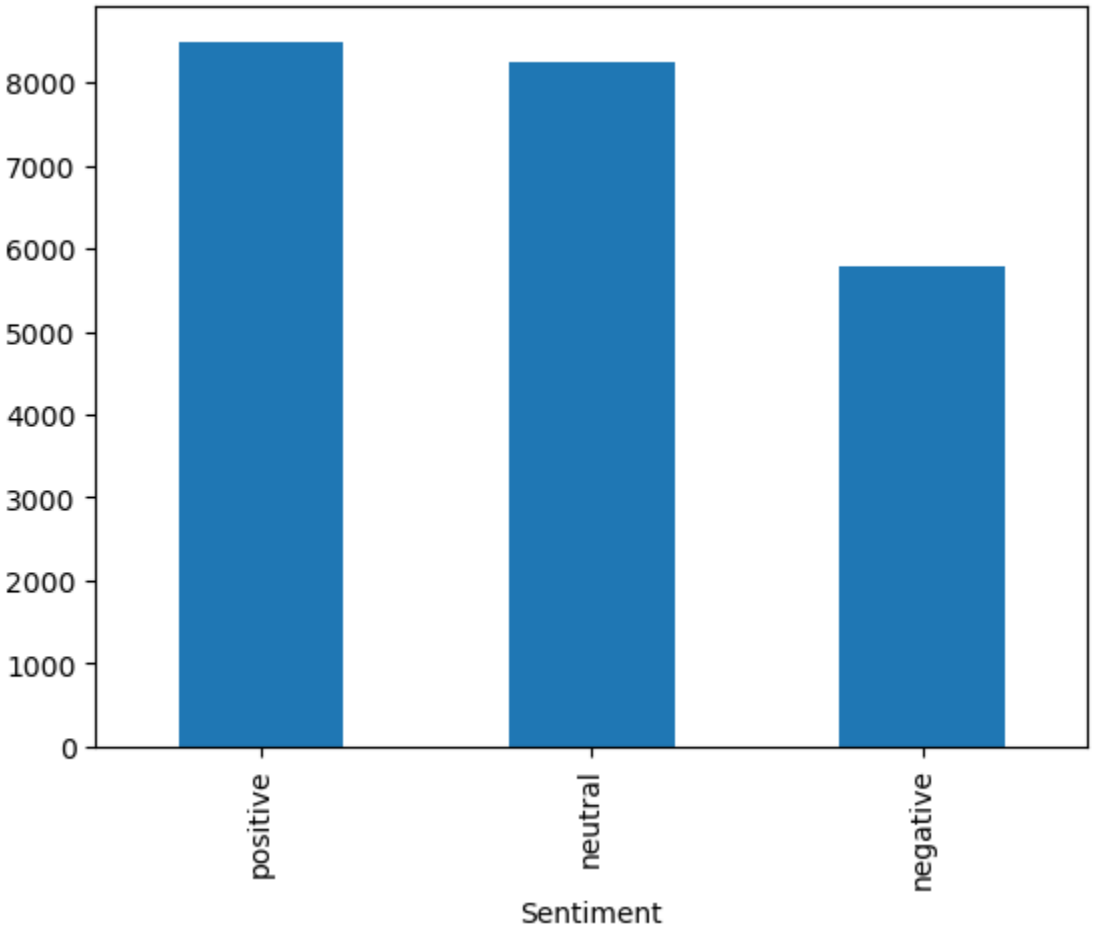
```
Accuracy of the model: 0.7227524972253052
```

```
sentiment =  df['Sentiment'].value_counts()
sentiment.head().plot.bar()

plt.show()
```

[14] ✓ 0.1s Python





```
[16]  df['Date Created'] = pd.to_datetime(df['Date Created']).dt.time
      df['Date Created']
```

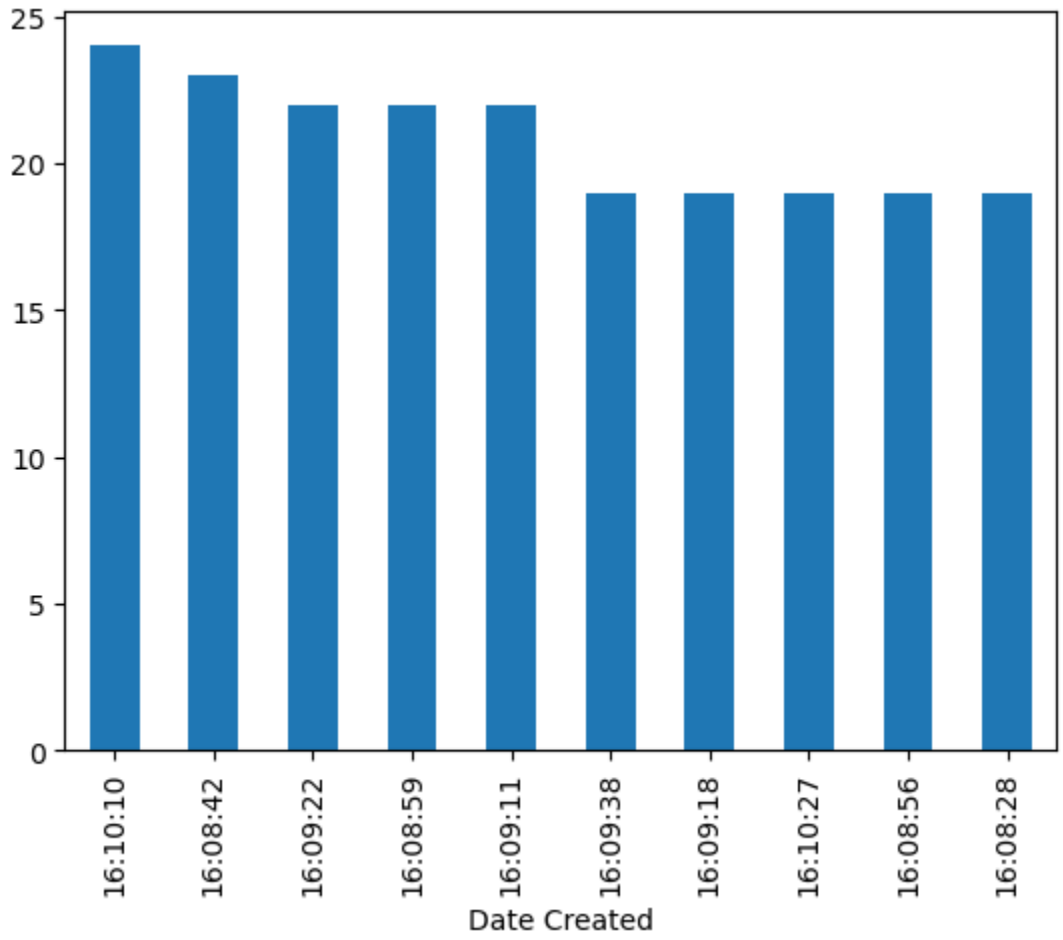
✓ 0.0s Python

```
0        23:59:21
1        23:59:01
2        23:58:41
3        23:58:33
4        23:58:28
           ...
22519    00:00:21
22520    00:00:03
22521    00:00:01
22522    00:00:00
22523    00:00:00
Name: Date Created, Length: 22524, dtype: object
```

```
[17]  ppd =  df['Date Created'].value_counts()
      ppd.head(10).plot.bar()

      plt.show()
```

✓ 0.1s Python



```python
from wordcloud import WordCloud, STOPWORDS
import pandas as pd

# Read your CSV file into a DataFrame
df = pd.read_csv('fifa_world_cup_2022_tweets.csv')

comment_words = ''
stopwords = set(STOPWORDS)

# Iterate through the 'Tweet' column in the DataFrame
for val in df['Tweet']:
    # Typecast each value to string
    val = str(val)

    # Split the value
    tokens = val.lower().split()  # Convert to lowercase and split directly

    # Join tokens to form a string
    comment_words += " ".join(tokens) + " "

# Generate the WordCloud
wordcloud = WordCloud(width=800, height=800,
                      background_color='white',
                      stopwords=stopwords,
                      min_font_size=10).generate(comment_words)
```

```python
# Plot the WordCloud image
import matplotlib.pyplot as plt
plt.figure(figsize=(8, 8), facecolor=None)
plt.imshow(wordcloud)
plt.axis("off")
plt.tight_layout(pad=0)
plt.show()
```

[18] ✓ 40.3s Python

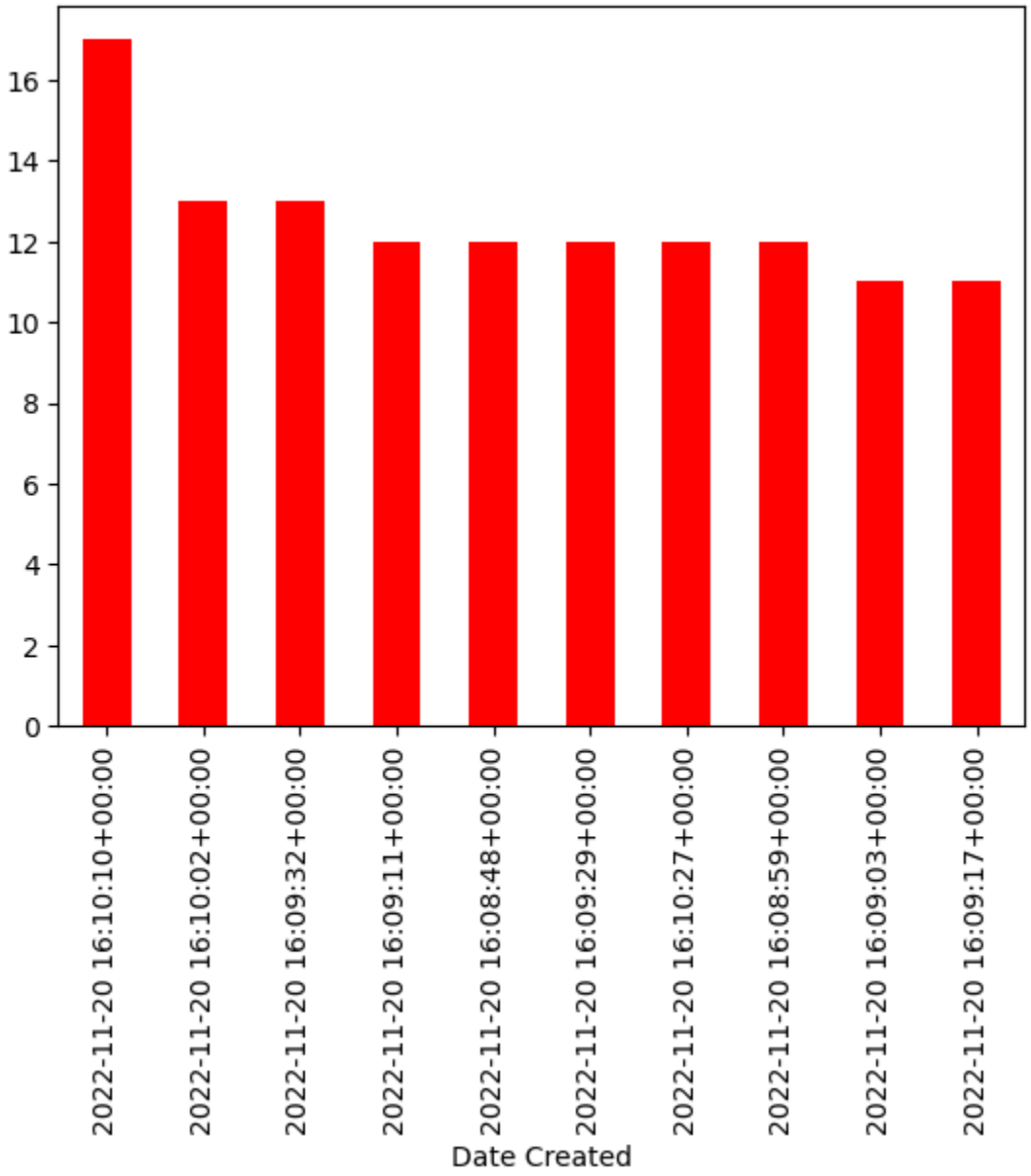

qatar
opening ceremony
worldcup worldcup2022
https
worldcup2022
world cup
t co
fifa world football
ecuador
will
game
worldcup2022 qatar2022
worldcup2022 fifaworldcup
fifaworldcup worldcup2022
goal
today
time
one
https
worldcup2022 qatarworldcup2022
worldcup2022 worldcup
cup opening
qatar2022 fifaworldcup
fifaworldcup https
worldcup2022 worldcupqatar2022
england
jungkook
prediction
live stream
soccer
qatar fifa
watching
tournament
ceremony fifa
fifaworldcup fifaworldcup2022
cup worldcup2022
worldcup2022 live
love
minute
win
bbc
kick
playing
let
qatecu worldcup2022
worldcup2022 fifaworldcup2022
corruption
let s
group
excited
streams online
play
come
match
fifaworldcup qatarworldcup2022
country
tv https
take
fan
good luck
ceremony live
bet
brazil
feel
openingceremony worldcup2022
first
thank
re
qatari
guy
money
hour
stadium
penalty
back
think
hope
look
make
u
m
good
score
online qatar2022
morgan freeman
enner valencia
even
start
vs ecuador
better
qatarworldcup2022
watch
sport
cup qatar
want
worldcup2022 fifa
live tv
people
see
offside
var
great
support
jiocinema
host
qatar v
know
first goal
corrupt
offside worldcup2022
home
much
really
everyone
going
qatarworldcup2022 worldcup2022
qatar2022 https
now
right
well
qatar ecuador
enjoy
argentina
best
finally
go
tomorrow
team
amp
proud
don t
new
france
way
day
still
player
need
show
ready
already
say
qatarworldcup2022 qatar2022
worldcup2022 worldcup2022
valencia
qatar2022 worldcup2022

```
neg_tweets = df[df['Sentiment'] == 'negative']
neg_tweets = neg_tweets['Date Created'].value_counts()
neg_tweets.head(10).plot.bar(color=['red'])

neg_tweets.head()
```

[19] ✓ 0.2s Python

```
Date Created
2022-11-20 16:10:10+00:00    17
2022-11-20 16:10:02+00:00    13
2022-11-20 16:09:32+00:00    13
2022-11-20 16:09:11+00:00    12
2022-11-20 16:08:48+00:00    12
Name: count, dtype: int64
```



```
negdf = pd.DataFrame()
negdf['negative'] = neg_tweets

postweets = df[df['Sentiment'] == 'positive']
posdf = pd.DataFrame()

posdf['positive'] = postweets['Date Created'].value_counts()
f_df = pd.DataFrame()

f_df = pd.merge(negdf, posdf, left_index=True, right_index=True)
```

[20] ✓ 0.0s Python

```
f_df = f_df.sort_index(ascending=True)
f_df
```

[21] ✓ 0.0s Python

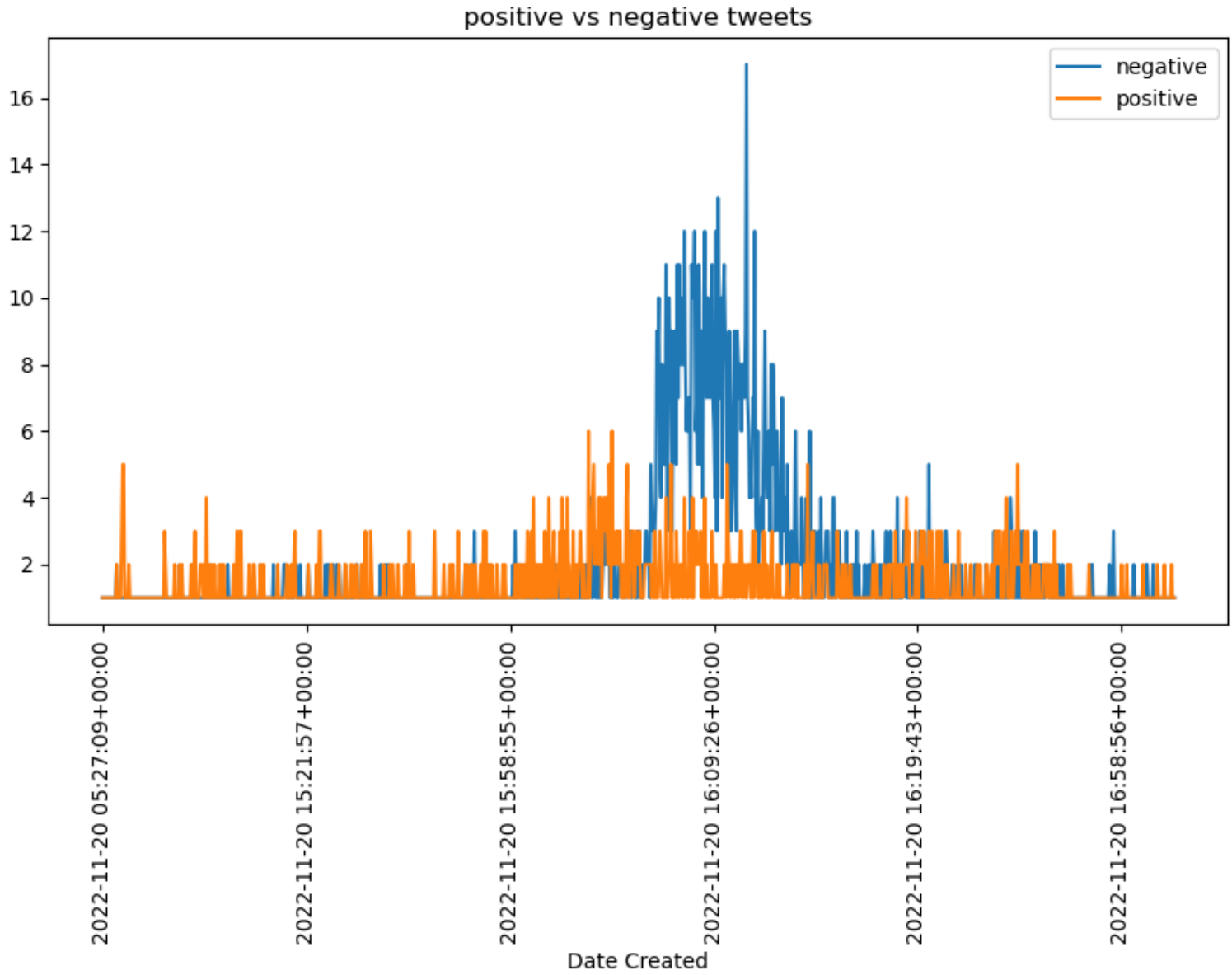
Date Created	negative	positive
2022-11-20 05:27:09+00:00	1	1
2022-11-20 05:57:24+00:00	1	1
2022-11-20 06:55:09+00:00	1	1
2022-11-20 07:20:04+00:00	1	1
2022-11-20 07:37:27+00:00	1	1
...	...	...
2022-11-20 20:03:55+00:00	1	1
2022-11-20 22:00:01+00:00	1	2
2022-11-20 22:15:52+00:00	1	1
2022-11-20 22:49:59+00:00	1	1
2022-11-20 23:09:59+00:00	1	1

1053 rows × 2 columns

```
f_df.plot(figsize=(10, 5), rot=90, title="positive vs negative tweets")
```

[22] ✓ 0.2s Python

```
<Axes: title={'center': 'positive vs negative tweets'}, xlabel='Date Created'>
```


positive vs negative tweets
negative
positive
16
14
12
10
8
6
4
2
2022-11-20 05:27:09+00:00
2022-11-20 15:21:57+00:00
2022-11-20 15:58:55+00:00
2022-11-20 16:09:26+00:00
2022-11-20 16:19:43+00:00
2022-11-20 16:58:56+00:00
Date Created

Operations performed:

To glean insightful information from Twitter discussions during the competition, the "FIFA World Cup 2022 Twitter Corpus" underwent a rigorous process. First, tweets were gathered using Twitter's API using particular keywords and hashtags, guaranteeing a varied and extensive dataset. After that, the text was heavily cleaned to get rid of unnecessary information, emoticons, URLs, and duplicates while standardizing it for analysis. Sentiment analysis approaches classified tweets into positive, negative, and neutral feelings using tools like Vader. To further enable a better understanding of issues like match analysis or disputes, the corpus was subjected to topic modeling in order to identify recurring themes and conversations. A thorough temporal analysis made it possible to monitor shifts in sentiment at each level of the competition. Word clouds and histograms were among the visualizations used to effectively convey sentiment trends and important discussion topics.

Q) Will you be creating a corpus from scratch ?

Ans. We will try to create a corpus from scratch when we find insufficient data to perform a particular task.

The first step in creating a corpus for Twitter sentiment analysis during the FIFA World Cup is gathering tweets pertaining to the tournament using certain hashtags and keywords. After being compiled, these tweets are carefully vetted to guarantee accuracy by removing any irrelevant text, emoticons, duplicates, and URLs. The following stage is labeling these tweets with sentiments that represent the public's reactions to the tournament, indicating whether they express good, negative, or neutral feelings. This well-structured dataset with tweet texts, related feelings, and other pertinent information is a useful tool for comprehending the opinions said during the FIFA World Cup conversations on Twitter.

Proposed Work:

Approach:

Data Collection: Gather diverse social media data (Twitter, forums) to encompass varied language styles and contexts.
Preprocessing: Clean and tokenize data, leveraging domain-specific lexicons and handling context nuances.
Model Development: Employ attention-based deep learning models to capture fine-grained sentiments related to specific aspects.
Evaluation: Use precision, recall, and F1-score to evaluate the model's performance on aspect-level sentiment analysis tasks.

Expected Outcomes:

A robust sentiment analysis model capable of nuanced aspect-level sentiment identification in social media text.
Insights into contextual nuances impacting sentiment analysis in diverse social media conversations.

Results and Analysis:

The results and analysis of sentiment analysis depend on the specific methodology, dataset, and evaluation metrics used. However, here's a general overview of how sentiment analysis results are typically presented and analyzed

1. Accuracy Metrics: These include metrics like accuracy, precision, recall, F1-score, and ROC-AUC for binary classification tasks or multiclass sentiment analysis tasks. These metrics measure how well the sentiment analysis model performs compared to ground truth or human-labeled data.

2. Confusion Matrix: Shows a breakdown of true positives, true negatives, false positives, and false negatives, providing insights into the model's performance on individual sentiment classes.

3. Performance Comparison: Comparing the performance of different sentiment analysis methods or models helps identify which approaches are more effective for the given dataset or task.

4. Identification of Strong Indicators: Analyzing important features or linguistic cues identified by the model as strong indicators of sentiment can offer insights into what influences sentiment in the analyzed text.

5. Contextual Understanding: Understanding how the model performs in different contexts, domains, or with varying text genres helps assess its adaptability and generalizability.

6. Error Analysis: Investigating cases where the model misclassified sentiments can reveal common challenges or weaknesses, such as handling sarcasm, negation, or ambiguous language.

Fascinating trends and insights were revealed by the sentiment analysis of Twitter talk during the FIFA World Cup 2022.

- Sentiment Fluctuations: Based on game results, contentious situations, and exceptional player performances, Twitter users' feelings fluctuated greatly. Early games generated a lot of optimism, but when important events happened, opinions sharply changed to reflect the fervent responses of the supporters.

- Key Discussion Points: Team performances, player moments, referee calls, and crowd responses were the main topics of conversation. But attitude was also impacted by off-field activities like ceremonies or social concerns, underscoring the tournament's multifaceted significance.

- Important Elements: Emotions were mostly shaped by upsets, controversial choices, and outstanding sportsmanship. Fair play situations frequently produced positive spikes, whereas contentious events increased negativity.

- Emotions according to Stage: Feelings changed as the competition progressed. The excitement around the early matches was more widespread, but as the stakes rose in the knockout stages, feelings became more intense and both positive and negative reactions were accentuated.

- Regional Feelings: Differing viewpoints were reflected in the subtleties of sentiments that were expressed in relation to the performances of the teams in each region or the national controversy.

- Influencers: The words and actions of well-known individuals and players have a big impact

on the sentiment patterns. Their deeds or words frequently caused significant swings in public opinion.

- Essentially, the sentiment analysis painted a clear picture of how Twitter users felt about the FIFA World Cup. It portrayed the complex interactions between on-field activities, off-field happenings, various geographic viewpoints, and significant personalities, illuminating the feelings and attitudes of the general people during the competition.

Conclusion:

A concluding analysis often summarizes the strengths and limitations of the sentiment analysis approach used, suggests areas for improvement or further research, and highlights the practical implications of the results obtained.

The depth and complexity of the analysis depend on the scope of the study, the depth of the data analysis, and the specific goals of the sentiment analysis task.

Summarize Key Takeaways: Reiterate the main findings and their significance in the context of the study.

Closing Remarks: Conclude with a statement that encapsulates the overall impact of the sentiment analysis results and the implications for the broader field or practical applications.

The conclusion serves as a synthesis of the sentiment analysis findings, offering insights, recommendations, and potential pathways for future research or practical implementations based on the outcomes of the analysis.

References:

[1] Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval.

[2] Liu, B. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies.

[3] Socher, R. et al. (2013). Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. Conference on Empirical Methods in Natural Language Processing (EMNLP).

[4] Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. Conference on Empirical Methods in Natural Language Processing (EMNLP).

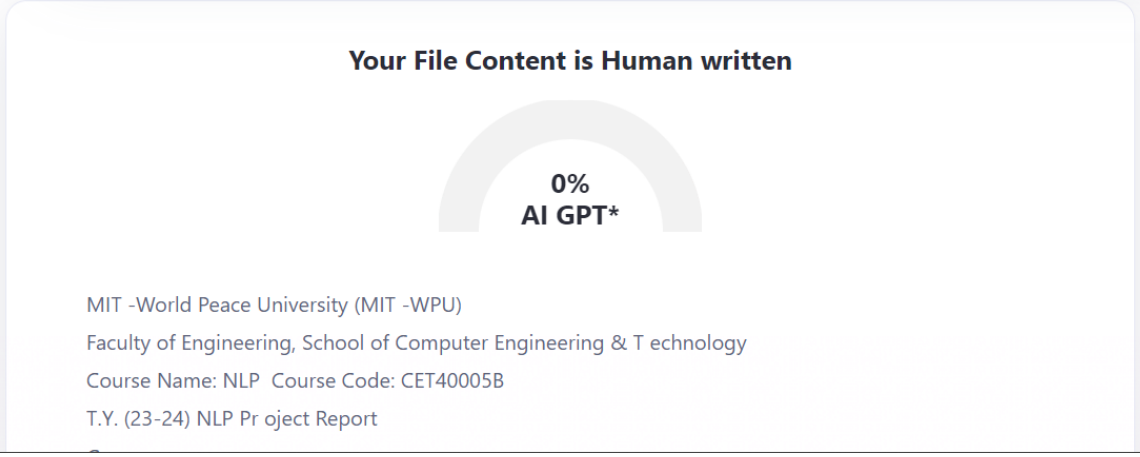
[5] Hutto, C.J. & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International AAAI Conference on Weblogs and Social Media (ICWSM).

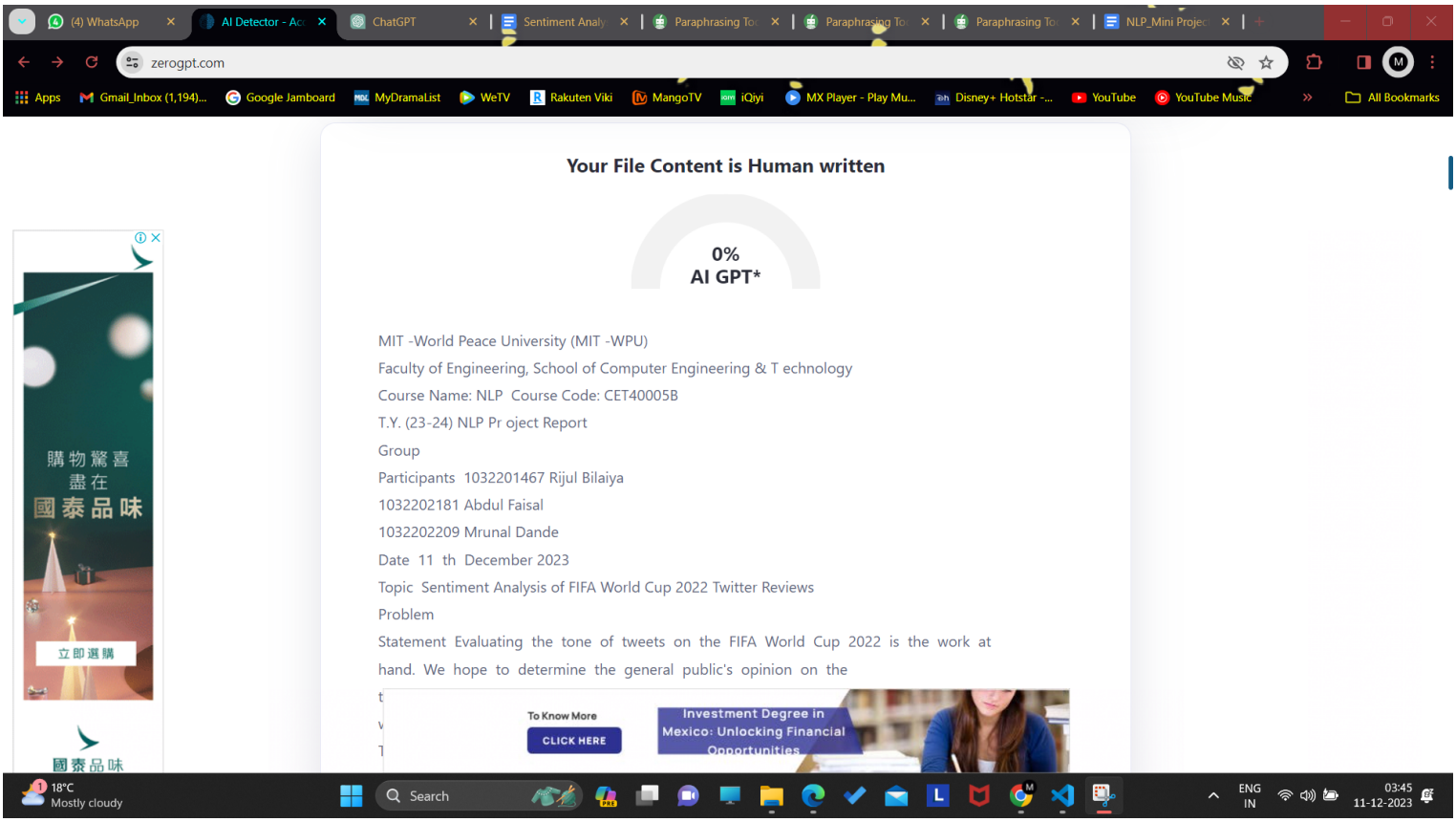
[6] Wang, Y., Huang, M., Zhu, X., & Zhao, L. (2016). Attention-Based LSTM for Aspect-Level Sentiment Classification. Conference on Empirical Methods in Natural Language Processing (EMNLP)

[7] Severyn, A., & Moschitti, A. (2015). Twitter Sentiment Analysis with Deep Convolutional Neural Networks. Conference on Empirical Methods in Natural Language Processing (EMNLP).

Plagiarism Report-

- **Project Report-**





- **Project Research Paper-**

Sentiment Analysis of FIFA World Cup 2022 Twitter Reviews
Rijul Bilaiya
MIT W orld Peace University
Pune, India
1032201594@mitwpu.edu.in Mrunal Dande
MIT W orld Peace University
Pune, India
1032202209@mitwpu.edu.in Faisal
MIT W orld Peace University
Pune, India
1032201546@mitwpu.edu.in
Under the guidance of :
Suja Panicker

Detect Text

Upload File

14,999/15,000 Characters
(Get up to 100,000 here)

Your File Content is Human written

2.48%
AI GPT*

Sentiment Analysis of FIFA World Cup 2022 Twitter Reviews

(4) WhatsApp | AI Detector - Ac | ChatGPT | Sentiment Analy | Paraphrasing To | Paraphrasing To | Paraphrasing To | NLP_Mini Projec

zerogpt.com

Apps | Gmail_Inbox (1,194)... | Google Jamboard | MyDramaList | WeTV | Rakuten Viki | MangoTV | iQiyi | MX Player - Play Mu... | Disney+ Hotstar -... | YouTube | YouTube Music | All Bookmarks

Your File Content is Human written

2.48%
AI GPT*

Sentiment Analysis of FIFA World Cup 2022 Twitter Reviews
Rijul Bilaiya
MIT W orld Peace University
Pune, India
1032201594@mitwpu.edu.in Mrunal Dande
MIT W orld Peace University
Pune, India
1032202209@mitwpu.edu.in Faisal
MIT W orld Peace University
Pune, India
1032201546@mitwpu.edu.in
Under the guidance of :
Suja Panicker
Assistant Professor

18°C Mostly cloudy | Search | ENG IN | 03:48 11-12-2023