▾ Twitter Sentiment Analysis

Uses Naive Bayes Classifier. TextBlob library- with inbuilt Naive Bayes Classifier

```
# Minor Project
# AYUSH SHARMA
# 2K19/SE/027


# Description: Sentiment analysis on tweets parsed directly from live Twitter feed
```

▾ Import Libraries and Define Functions

```python
import pandas as pd
import tweepy
from textblob import TextBlob
from wordcloud import WordCloud
import numpy as np
import re
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
import os


clearConsole = lambda: os.system('cls' if os.name in ('nt', 'dos') else 'clear')


# Function to clean the tweets text
# Uses Regular Expressions (re) library

def cleanText(text):
  text= re.sub(r'@[A-Za-z0-9]+', '', text)  #Removes @ mentions
  text=re.sub(r'#', '', text)               #Removes hash tags
  text=re.sub(r'RT[\s]+', '', text)         #Remove Re tweet symbol
  text=re.sub(r'https?:\/\/\S+', '', text)  #Removes any hyperlinks from the tweet

  return text


#Function to calculate subjectivity of a tweet
#Using TextBlob

def getSubjectivity(text):
  return TextBlob(text).sentiment.subjectivity


#Function to calculate polarity of a tweet
#using TextBlob
```

```python
def getPolarity(text):
  return TextBlob(text).sentiment.polarity

#Analyse

def getAnalysis(score):
  if score<0:
    return 'Negative'
  elif score==0:
    return 'Neutral'
  else:
    return 'Positive'


# Rumour Detection
def fact_opinion(score, t1=0.25, t2=0.75):
  if score<t1:
    return 'Fact'
  elif score<t2:
    return 'Neutral'
  else:
    return 'Opinion'
```

▾ Twiiter Authentication

```python
# Twitter API credentials
# Which are secret so I can't reveal them
# Hence I will upload them from csv file I made
cred= pd.read_csv('credentials.csv')


ck = cred['consumer_key'][0]
cs = cred['consumer_secret'][0]
ak = cred['access_key'][0]
as_ = cred['access_secret'][0]


# Authenticate
auth = tweepy.OAuthHandler(ck,cs)

# Access
auth.set_access_token(ak, as_)

#API
api= tweepy.API(auth, wait_on_rate_limit=True)
```

▾ Gather Tweets

```python
#Initialise the posts array
num_posts=500    #Number of tweets you want to extract
```

```python
posts=[]

# Listener Object

class Listener (tweepy.StreamListener):

  def on_status(self, status):
    global num_posts
    if num_posts>0:
      posts.append(status.text)
      num_posts-=1

    else:
      print("Collected desired number of posts")
      return False

  def on_error(self, status_code):
    if status_code == 420:
      return False



listener_= Listener()

myStream = tweepy.Stream(auth = api.auth, listener=listener_)


myStream.filter(track=['Indian Cricket Team', 'Cricket', 'World Cup', 'T20', 'Virat Kohli'
                       'Jasprit Bumrah', 'Ind vs NZ', 'Ind vs Pak', 'India vs New Zealand'
```

```
    Collected desired number of posts
```

```python
len(posts)
```

```
    500
```

```python
for tweet in posts[0:5]:
  print(tweet+'\n')
```

```
    Only one makes it through to The Final... who ya got?

    Adelaide, you won't wanna miss this! Get your tickets now ✏… https://t.co/U20UM43M5/

    RT @WasimJaffer14: Warne won 1 World Cup.
    Warner won 2 World Cups.
    Looking forward to seeing an Aussie kid named Warnest win 3 World Cups.…

    RT @man4_cricket: Rohit Sharma moves to Number 13 position in the latest ICC T20I Bat

    16 is the number of teams that qualified for the finals of the FIFA World Cup from 19

    A special coaching event organised by @GlamCricket and @CricketWales held at Rydal Pe
```

```
df = pd.DataFrame([tweet for tweet in posts], columns=['Tweets'])
df.head()
```

	Tweets
0	Only one makes it through to The Final... who ...
1	RT @WasimJaffer14: Warne won 1 World Cup.\nWar...
2	RT @man4_cricket: Rohit Sharma moves to Number...
3	16 is the number of teams that qualified for t...
4	A special coaching event organised by @GlamCri...

```
df.to_csv('tweets_posts.csv')
```

▾ Clean Tweets

```
#Clean the Tweets text
```

```
df['Tweets']=df['Tweets'].apply(cleanText)
```

```
df
```

	Tweets
0	Only one makes it through to The Final... who ...
1	: Warne won 1 World Cup.\nWarner won 2 World C...
2	_cricket: Rohit Sharma moves to Number 13 posi...
3	16 is the number of teams that qualified for t...
4	A special coaching event organised by and he...
...	...
495	I am not impressed by your cricket.\nIdiot ....
496	Autumn series provides glorious glimpse of wha...
497	We are awaiting any recommendations from a...
498	: I played competitive football, rugby & c...
499	Why is nobody talking about Junior Hockey Worl...

500 rows × 1 columns

```
df.to_csv("tweets_clean.csv")
```

▾ Sentiment Analysis

```
df= pd.read_csv('tweets_clean.csv')


# Get the subjectivity and polarity of tweets

# Subjectivity- [0 to 1]. More subjective means the tweet is more opinionated than fact
# Polarity- [-1 to 1]. Positive means good, negative means bad

df['Subjectivity']= df['Tweets'].apply(getSubjectivity)
df['Polarity']=df['Tweets'].apply(getPolarity)

#Display
df
```
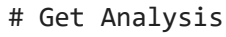
	Tweets	Subjectivity	Polarity
0	Only one makes it through to The Final... who ...	1.000000	0.000000
1	: Warne won 1 World Cup.\nWarner won 2 World C...	0.400000	0.800000
2	_cricket: Rohit Sharma moves to Number 13 posi...	0.900000	0.500000
3	16 is the number of teams that qualified for t...	0.000000	0.000000
4	A special coaching event organised by and he...	0.571429	0.357143
...	...	...	...
495	I am not impressed by your cricket.\nIdiot	0.900000	-0.650000
496	Autumn series provides glorious glimpse of wha...	0.000000	0.000000
497	We are awaiting any recommendations from a...	0.125000	0.000000
498	: I played competitive football, rugby & c...	0.500000	0.500000
499	Why is nobody talking about Junior Hockey Worl...	0.000000	0.000000

500 rows × 3 columns

```
# Plot Word Cloud

allWords= ' '.join( [twts for twts in df['Tweets']] )
wordCloud= WordCloud(width=500, height=300, random_state=21, max_font_size=119).generate(a

plt.imshow(wordCloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```

Get Analysis

```
df['Analysis']=df['Polarity'].apply(getAnalysis)
```

```
#Display
df
```

	Tweets	Subjectivity	Polarity	Analysis
0	Only one makes it through to The Final... who ...	1.000000	0.000000	Neutral
1	: Warne won 1 World Cup.\nWarner won 2 World C...	0.400000	0.800000	Positive
2	_cricket: Rohit Sharma moves to Number 13 posi...	0.900000	0.500000	Positive
3	16 is the number of teams that qualified for t...	0.000000	0.000000	Neutral
4	A special coaching event organised by and he...	0.571429	0.357143	Positive
...	...	...	...	...
495	I am not impressed by your cricket.\nIdiot	0.900000	-0.650000	Negative
496	Autumn series provides glorious glimpse of wha...	0.000000	0.000000	Neutral
497	We are awaiting any recommendations from a...	0.125000	0.000000	Neutral
498	: I played competitive football, rugby & c...	0.500000	0.500000	Positive
499	Why is nobody talking about Junior Hockey Worl...	0.000000	0.000000	Neutral

500 rows × 4 columns

See whether the tweets are fact based or opinionated

```
df['Facts']= df['Subjectivity'].apply(fact_opinion)
```

```
df
```

	Tweets	Subjectivity	Polarity	Analysis	Facts
0	Only one makes it through to The Final... who ...	1.000000	0.000000	Neutral	Opinion
1	: Warne won 1 World Cup.\nWarner won 2 World C...	0.400000	0.800000	Positive	Neutral
2	_cricket: Rohit Sharma moves to Number 13 posi...	0.900000	0.500000	Positive	Opinion
3	16 is the number of teams that qualified for t...	0.000000	0.000000	Neutral	Fact
	A special coaching event organised by and				

```
# Display all the Positive Tweets

ptweets= df[df.Analysis=='Positive']
ptweets= ptweets['Tweets']
ptweets
```

```
1      : Warne won 1 World Cup.\nWarner won 2 World C...
2      _cricket: Rohit Sharma moves to Number 13 posi...
4      A special coaching event organised by  and  he...
6      : OTD in 1976.  3-0 . A team full of 1974 Worl...
7      Abu Dhabi T10 is in full-swing as the greatest...
                             ...
490    : IN THE GABBA VICTORY\n[ A THREAD ]\n\nAustra...
491    _Odisha: Wish the Indian junior hockey team al...
493    : OnThisDay in 2009\n\nUmar Akmal, playing his...
494    : Getting match ready 👌 👌\n\nTeamIndia get int...
498    : I played competitive football, rugby &amp; c...
Name: Tweets, Length: 248, dtype: object
```

```
# Display all the Negative Tweets

ntweets= df[df.Analysis=='Negative']
ntweets= ntweets['Tweets']
ntweets
```

```
9      _Mystical:  I stopped watching these grown chi...
27     _don_: Rohit Sharma moves to 13 rank in icc Me...
31     : Highest % of team runs for India in Tests at...
36     _venkat: Rohit Sharma to be blamed for this fo...
39     Thank God you knocked 20 balls off the short f...
42     : Expected, was never leaving the European Cha...
45     : Highest Test Average at Home:-\n[min. 1500 r...
58     AjinkyaRahane was not amused with questions re...
60     Windies is very near to break the Indian recor...
72     There'll be a few of the  family in attendance...
75     _se_hai: Dear    either ensure transparency in...
80      _robin I couldn't agree more Nick and hate to...
87     : Highest Test Average at Home:-\n[min. 1500 r...
95     : NL Mathieu van der Poel and BE Wout van Aert...
117    _49: I think his HS of 202 is against Australi...
121    : Fifa football conference today for clubs and...
122    ***1977 Albums World Cup Tournament****\n\nGRO...
132    Quote of the day: Don't ask difficult question...
```

```
142    : Virat Kohli in 2021:\n\n- Out of top 5 Tes...
171    _khen _freak_9  Beeter you go and play oh sorr...
185    Hafeez stepping down for Pak series in which h...
197    __za:  Bafana didn't fail to qualify they lost...
219    : Kya Drama laga rakha hai Cricket Australia n...
224    : There'll be a few of the  family in attendan...
232    _18: A Random Thread on Virat Kohli Performanc...
239    : Highest % of team runs for India in Tests at...
244    : Former cricketer and commentator Ian Smith i...
254    Ind vs NZ first Test | Spinners hold the key f...
255    : Virat Kohli in 2021:\n\n- Out of top 5 Tes...
264                    Ind vs NZ first Test tough to crack
271    : Hindi Cricket Commentary is shit. Mostly the...
293    _45: Highest % of team runs for India in Tests...
296     Muyise akabbo oba it's too late we can't make it
299    : Highest % of team runs for India in Tests at...
301    _45: Highest % of team runs for India in Tests...
315    : Dr Hilary say hardly affects anyone with sid...
318      _morikawa  Bloody loved you taking about da...
337     y this sudden Virat Kohli hating fake behavio...
343     Blatters corruption led to this "World Cup" h...
368    __: Michael Vaughan will not have role in BBC'...
370    : Highest % of team runs for India in Tests at...
384    : Funnily enough if you add this up you could ...
386    __: Michael Vaughan will not have role in BBC'...
403    6 years it took, sextortion is complicated!\nT...
423    _ash1: Kitni flat pitches hoti thi uss time pe...
467     As champions, it seems right that England hos...
484    some reason 😂 something stupid like the Bradfo...
485                    Ind vs NZ first Test tough to crack
492     Reason why non test match bowlers shouldn't b...
495      I am not impressed by your cricket.\nIdiot ....
Name: Tweets, dtype: object
```
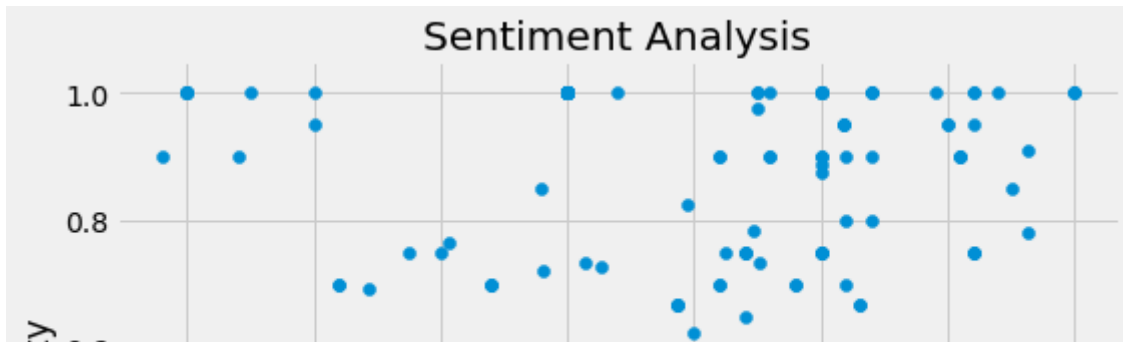
```python
# Plot the sentiment analysis

fig = plt.figure(figsize=(8,6))
plt.scatter(df['Polarity'],df['Subjectivity'])
plt.title("Sentiment Analysis")
plt.xlabel("Polarity")
plt.ylabel("Subjectivity")
plt.show()
```

```
# Visualise the percentage of positive and negative tweets

pos = (ptweets.shape[0]/df.shape[0])*100
neg = (ntweets.shape[0]/df.shape[0])*100

print("Positive: "+ str(pos)+ "%\nNegative: "+ str(neg))
```
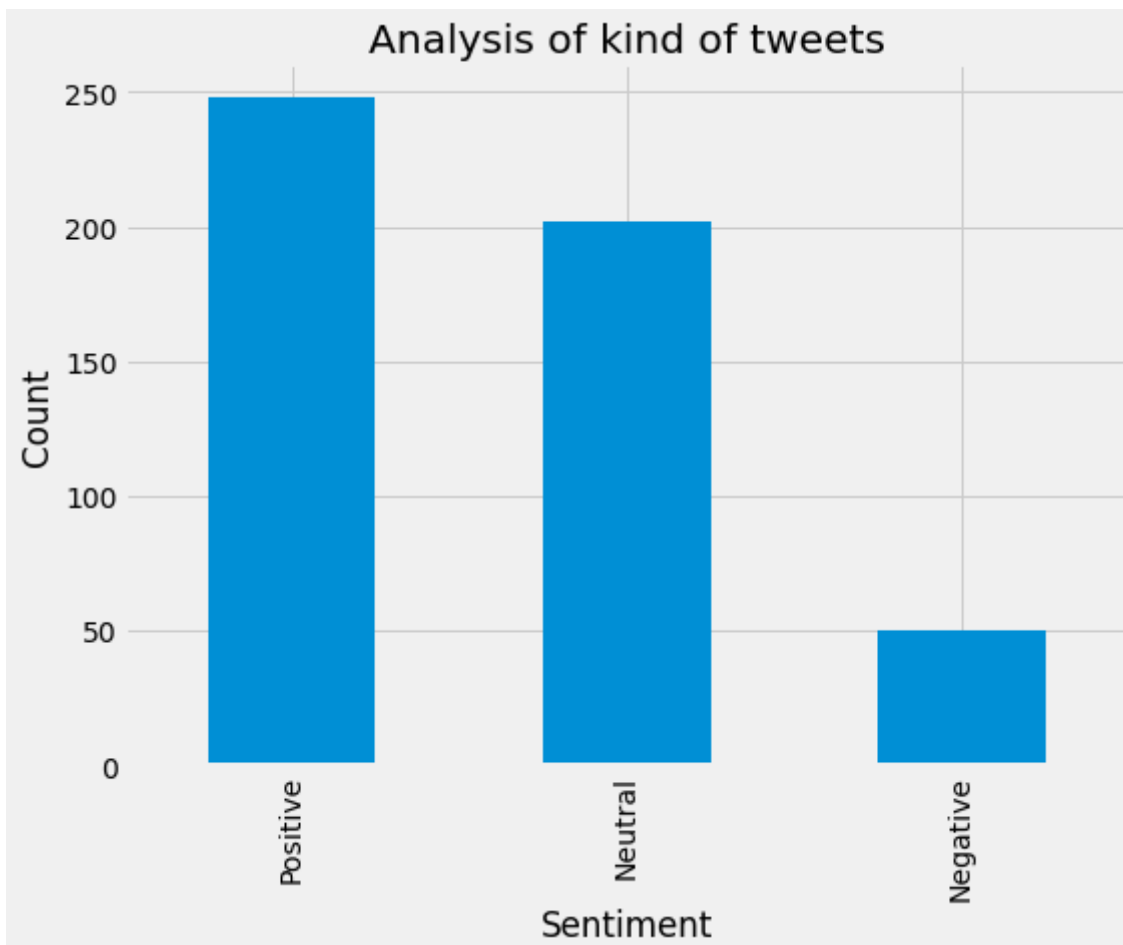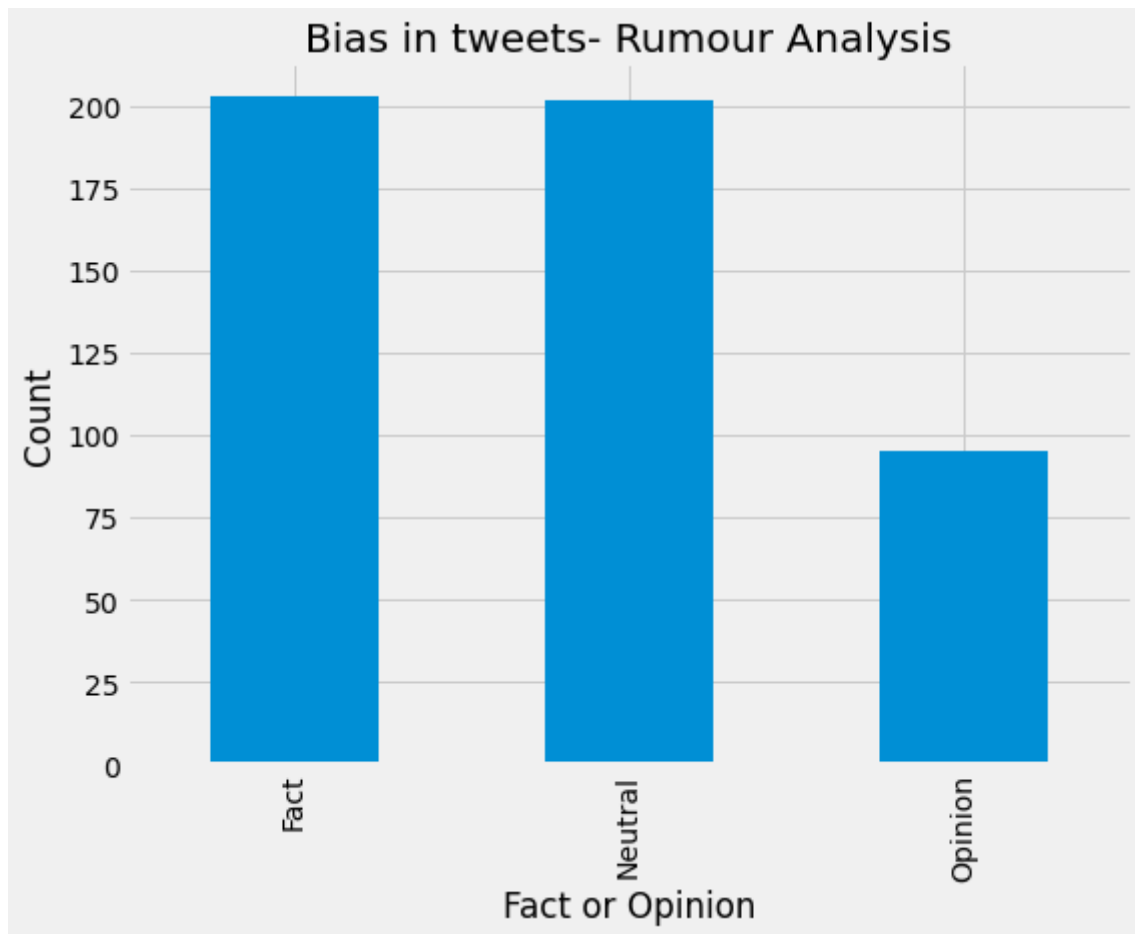
```
Positive: 49.6%
Negative: 10.0
```



```
fig= plt.figure(figsize=(8,6))
df['Analysis'].value_counts().plot(kind='bar')
plt.title("Analysis of kind of tweets")
plt.xlabel("Sentiment")
plt.ylabel("Count")
plt.show()
```



```
# Visualise the number of tweets that are factual vs the tweets that are biased
```

```python
fig= plt.figure(figsize=(8,6))
df['Facts'].value_counts().plot(kind='bar')
plt.title("Bias in tweets- Rumour Analysis")
plt.xlabel("Fact or Opinion")
plt.ylabel("Count")
plt.show()
```



```python
df.to_csv('tweets_data.csv')
```

✓ 0s   completed at 5:40 PM   ● ✕