Hello Folks Myself Nirbhay Tiwari (Data Scientist) Lets Begin Our Sentiment Analysis NLP Project (Date: 1/2/2024)

```
#!pip install nltk

#!pip install youtube-comment-scraper-python

#Importing Modules & Required Libraries To Work With

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import os

#import functions for data preprocessing & data prepration

from sklearn.preprocessing import LabelEncoder
from sklearn.utils import resample
from sklearn.feature_extraction.text import CountVectorizer
from nltk.sentiment.vader import SentimentIntensityAnalyzer


from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
from nltk.stem.snowball import SnowballStemmer
from nltk.corpus import stopwords
from nltk.corpus import wordnet
import string
from string import punctuation
import nltk
import re

#Data Extraction

#from youtube_comment_scraper_python import *
#import pandas as pd

#link = input("Youtube Link: https://youtu.be/X0tOpBuYasI?
si=nud0kijxn9ESltiT")
#saved = input("Output name: BlackAdam")
#youtube.open(link)

#response = youtube.video_comments()

#all_data = []

#for i in range(0, 20): # it will scroll 10 times

#    response = youtube.video_comments()
```

```
#    data = response['body']
#    all_data.extend(data)
#    df.to_csv(saved)

#Reading Black Adam Trailer Extracted Comments Stored in Csv File

data = pd.read_csv('C:/Users/info4/OneDrive/Desktop/comments.csv')
data.columns
```

```
Index(['Unnamed: 0', 'Comment', 'Likes', 'Time', 'user', 'UserLink'],
dtype='object')
```

```
data1=data.drop(['Unnamed:
0','Likes','Time','user','UserLink'],axis=1)
data1
```

```
                                                Comment
0    Love how Dr. Fate's design looks and how cool ...
1    I can’t get over how good everything looks. Dr...
2    Really hoping that this can save DC's movie un...
3    U cant deny how good this looks.Now if they ca...
4    From this trailer, I have a feeling that this ...
..                                                 ...
275  I want to see this. It may be one of his most ...
276       wow thats very amazing. I can't wait to see.
277                    Doctor Fate is why i'm watching
278  This looks fire. DC looks like they stepping t...
279  Shazam : "I don't want fight you Black Adam."B...

[280 rows x 1 columns]
```

```
#Data Lablleing

nltk.download('vader_lexicon')
sentiments = SentimentIntensityAnalyzer()
data1["Positive"] = [sentiments.polarity_scores(i)["pos"] for i in
data1["Comment"]]
data1["Negative"] = [sentiments.polarity_scores(i)["neg"] for i in
data1["Comment"]]
data1["Neutral"] = [sentiments.polarity_scores(i)["neu"] for i in
data1["Comment"]]
data1['Compound'] = [sentiments.polarity_scores(i)["compound"] for i
in data1["Comment"]]
score = data1["Compound"].values
sentiment = []
for i in score:
    if i >= 0.05 :
        sentiment.append('Positive')
    elif i <= -0.05 :
        sentiment.append('Negative')
    else:
```

```
        sentiment.append('Neutral')
data1["Sentiment"] = sentiment
data1.head(20)
```

```
[nltk_data] Downloading package vader_lexicon to
[nltk_data]     C:\Users\info4\AppData\Roaming\nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!
```

	Comment	Positive	Negative	\
0	Love how Dr. Fate's design looks and how cool ...	0.384	0.000	
1	I can’t get over how good everything looks. Dr...	0.153	0.000	
2	Really hoping that this can save DC's movie un...	0.375	0.000	
3	U cant deny how good this looks.Now if they ca...	0.302	0.049	
4	From this trailer, I have a feeling that this ...	0.131	0.000	
5	I am just so happy that a fan cast has actuall...	0.497	0.000	
6	I have never been more hyped for a comic book ...	0.000	0.247	
7	I like a lot of the imagery and action, and it...	0.122	0.058	
8	Every time I watch this trailer, I get chills....	0.039	0.000	
9	This movie feels like a step in the right dire...	0.247	0.039	
10	Anguish and Rage are the two of most dangerous...	0.095	0.193	
11	I hope we get to see more of the Justice Socie...	0.328	0.000	
12	This looks AMAZING! All of the marvel installm...	0.182	0.000	
13	2:12 anyone else love this scene where you can...	0.197	0.000	
14	Whoever's idea to cast Pierce Brosnan as Dr Fa...	0.163	0.064	
15	This movie looks absolutely amazing I absolute...	0.239	0.000	
16	This is going to be SO good. There had to have...	0.073	0.000	
17	So looks like it’s going to be Black Adam Vs t...	0.125	0.175	
18	Honestly, it's refreshing to see Dwayne playin...	0.271	0.060	
19	Será o melhor filme de quadrinhos do ano. DC»»...	0.000	0.000	

	Neutral	Compound	Sentiment
0	0.616	0.8910	Positive
1	0.847	0.6801	Positive
2	0.625	0.9216	Positive
3	0.649	0.9262	Positive
4	0.869	0.4416	Positive
5	0.503	0.9501	Positive
6	0.753	-0.6002	Negative
7	0.820	0.3071	Positive
8	0.961	0.2714	Positive
9	0.714	0.9020	Positive
10	0.712	-0.8860	Negative
11	0.672	0.9730	Positive
12	0.818	0.7103	Positive
13	0.803	0.7319	Positive
14	0.773	0.6573	Positive
15	0.761	0.6240	Positive
16	0.927	0.6028	Positive
17	0.700	-0.2782	Negative
18	0.669	0.8999	Positive
19	1.000	0.0000	Neutral

```
data2=data1.drop(['Positive','Negative','Neutral','Compound'],axis=1)
data2.head(30)
```

	Comment	Sentiment
0	Love how Dr. Fate's design looks and how cool ...	Positive
1	I can’t get over how good everything looks. Dr...	Positive
2	Really hoping that this can save DC's movie un...	Positive
3	U cant deny how good this looks.Now if they ca...	Positive
4	From this trailer, I have a feeling that this ...	Positive
5	I am just so happy that a fan cast has actuall...	Positive
6	I have never been more hyped for a comic book ...	Negative
7	I like a lot of the imagery and action, and it...	Positive
8	Every time I watch this trailer, I get chills....	Positive
9	This movie feels like a step in the right dire...	Positive
10	Anguish and Rage are the two of most dangerous...	Negative
11	I hope we get to see more of the Justice Socie...	Positive
12	This looks AMAZING! All of the marvel installm...	Positive
13	2:12 anyone else love this scene where you can...	Positive
14	Whoever's idea to cast Pierce Brosnan as Dr Fa...	Positive
15	This movie looks absolutely amazing I absolute...	Positive
16	This is going to be SO good. There had to have...	Positive
17	So looks like it’s going to be Black Adam Vs t...	Negative
18	Honestly, it's refreshing to see Dwayne playin...	Positive
19	Será o melhor filme de quadrinhos do ano. DC»»...	Neutral
20	Can't wait to see more villain redemption. I w...	Negative
21	Super stoked for this! All the costumes look a...	Positive
22	Finally something that is pure badass and no j...	Negative

```
23  Damn!!!This looks freaking awesome!!\nCan't wa...  Positive
24  Finally some more of the other characters. I h...  Positive
25  Finally! The first superhero movie I've wanted...  Positive
26   I’ve never been excited for a DC film until now.  Negative
27  The effects look very good. They make this mov...  Positive
28  This looks insane. Love the way they are portr...  Positive
29  I feel like DC is finally starting to embrace ...  Positive
```

```
nltk.download('stopwords')
from nltk.stem import PorterStemmer
from nltk.stem import LancasterStemmer
stop_words = stopwords.words('english')
porter_stemmer = PorterStemmer()
lancaster_stemmer = LancasterStemmer()
snowball_stemer = SnowballStemmer(language="english")
lzr = WordNetLemmatizer()
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\info4\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

```
def text_processing(text):
    # convert text into lowercase
    text = text.lower()

    # remove new line characters in text
    text = re.sub(r'\n',' ', text)

    # remove punctuations from text
    text = re.sub('[%s]' % re.escape(punctuation), "", text)

    # remove references and hashtags from text
    text = re.sub("^a-zA-Z0-9$,.", "", text)

    # remove multiple spaces from text
    text = re.sub(r'\s+', ' ', text, flags=re.I)

    # remove special characters from text
    text = re.sub(r'\W', ' ', text)

    text = ' '.join([word for word in word_tokenize(text) if word not
in stop_words])

    # stemming using porter stemmer from nltk package - msh a7sn 7aga
- momken: lancaster, snowball
    # text=' '.join([porter_stemmer.stem(word) for word in
word_tokenize(text)])
    # text=' '.join([lancaster_stemmer.stem(word) for word in
word_tokenize(text)])
    # text=' '.join([snowball_stemer.stem(word) for word in
```

```
word_tokenize(text)])

    # lemmatizer using WordNetLemmatizer from nltk package
    text=' '.join([lzr.lemmatize(word) for word in
word_tokenize(text)])

    return text

nltk.download('omw-1.4')
nltk.download('punkt')
nltk.download('wordnet')
data_copy = data2.copy()
data_copy.Comment = data_copy.Comment.apply(lambda text:
text_processing(text))
```

```
[nltk_data] Downloading package omw-1.4 to
[nltk_data]     C:\Users\info4\AppData\Roaming\nltk_data...
[nltk_data]   Package omw-1.4 is already up-to-date!
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\info4\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package wordnet to
[nltk_data]     C:\Users\info4\AppData\Roaming\nltk_data...
```

```
le = LabelEncoder()
data_copy['Sentiment'] = le.fit_transform(data_copy['Sentiment'])

processed_data = {
    'Sentence':data_copy.Comment,
    'Sentiment':data_copy['Sentiment']
}

processed_data = pd.DataFrame(processed_data)
processed_data.head(20)

#2 Stands for Positive Score and 1 Stands for Negative Score and 0
Stands for Neutral Score
```

	Sentence	Sentiment
0	love dr fate design look cool scene look power...	2
1	get good everything look dr fate magic cyclone...	2
2	really hoping save dc movie universe looking n...	2
3	u cant deny good looksnow follow rest movie go...	2
4	trailer feeling movie going one movie would ne...	2
5	happy fan cast actually come true design good ...	2
6	never hyped comic book movie ever doesnt matte...	0
7	like lot imagery action great see jsa especial...	2
8	every time watch trailer get chill im used see...	2
9	movie feel like step right direction dc univer...	2
10	anguish rage two dangerous emotion human black...	0
11	hope get see justice society movie im fine cam...	2

```
12  look amazing marvel installment year good some...          2
13  212 anyone else love scene see atom smasher sp...          2
14  whoevers idea cast pierce brosnan dr fate dese...          2
15  movie look absolutely amazing absolutely wait ...          2
16  going good least four time trailer put hand he...          2
17  look like going black adam v jsa first two act...          0
18  honestly refreshing see dwayne playing rather ...          2
19      será melhor filme de quadrinhos ano dc marvel          1
```

```
processed_data['Sentiment'].value_counts()
```

```
2    205
1     39
0     36
Name: Sentiment, dtype: int64
```

```
df_neutral = processed_data[(processed_data['Sentiment']==1)]
df_negative = processed_data[(processed_data['Sentiment']==0)]
df_positive = processed_data[(processed_data['Sentiment']==2)]

# upsample minority classes
df_negative_upsampled = resample(df_negative,
                                 replace=True,
                                 n_samples= 205,
                                 random_state=42)

df_neutral_upsampled = resample(df_neutral,
                                 replace=True,
                                 n_samples= 205,
                                 random_state=42)


# Concatenate the upsampled dataframes with the neutral dataframe
final_data =
pd.concat([df_negative_upsampled,df_neutral_upsampled,df_positive])
```

```
final_data['Sentiment'].value_counts()
```

```
0    205
1    205
2    205
Name: Sentiment, dtype: int64
```

```
corpus = []
for sentence in final_data['Sentence']:
    corpus.append(sentence)
corpus[0:5]
```

```
['trailer look sick im definitely watching movie',
 'actually look like villain trailer',
 'movie going push dc top comic book movie disaster early dceu new
```

```
msheu mess two awesome projekts behind suicide squad peacemaker yeah
going rock pun kinda intended',
 'damn sure im gon na watchdc seems going right track',
 'okay look absolutely incredible dc making look foolish ever even
skeptical film definitely seeing opening weekend']

from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer(max_features=1500)
X = cv.fit_transform(corpus).toarray()
y = final_data.iloc[:, -1].values

# Mapping Data to Machine learning model

from sklearn.naive_bayes import GaussianNB
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3, random_state=0)
classifier = GaussianNB()
classifier.fit(X_train, y_train)

GaussianNB()

#Generating Confusion Matrix for checking accuracy score of model

from sklearn.metrics import confusion_matrix, accuracy_score
y_pred = classifier.predict(X_test)
cm = confusion_matrix(y_test, y_pred)
cm

array([[58,  0,  0],
       [ 0, 70,  0],
       [11,  1, 45]], dtype=int64)

nb_score = accuracy_score(y_test, y_pred)
print('accuracy',nb_score)

accuracy 0.9351351351351351
```