

Task 4

Task-04

Task: Analyze and visualize sentiment patterns in social media data to understand public opinion and attitudes towards specific topics or brands.

DataSet Link :

<https://www.kaggle.com/datasets/jp797498e/twitter-entity-sentiment-analysis>

In [205...]

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
import numpy as np
```

In [249...]

```
col=['Id' , 'Entity' , 'Sentiment' , 'Content']
df_train=pd.read_csv("C:\\\\Users\\\\abhis\\\\Downloads\\\\twitter_training.csv" , names=col)
df_train
```

Out[249]:

	Id	Entity	Sentiment	Content
0	2401	Borderlands	Positive	im getting on borderlands and i will murder yo...
1	2401	Borderlands	Positive	I am coming to the borders and I will kill you...
2	2401	Borderlands	Positive	im getting on borderlands and i will kill you ...
3	2401	Borderlands	Positive	im coming on borderlands and i will murder you...
4	2401	Borderlands	Positive	im getting on borderlands 2 and i will murder ...
...
74677	9200	Nvidia	Positive	Just realized that the Windows partition of my...
74678	9200	Nvidia	Positive	Just realized that my Mac window partition is ...
74679	9200	Nvidia	Positive	Just realized the windows partition of my Mac ...
74680	9200	Nvidia	Positive	Just realized between the windows partition of...
74681	9200	Nvidia	Positive	Just like the windows partition of my Mac is l...

74682 rows × 4 columns

In [251...]

```
df_test=pd.read_csv("C:\\\\Users\\\\abhis\\\\Downloads\\\\twitter_validation (2).csv" , names=col)
df_test
```

Out[251]:

	Id	Entity	Sentiment	Content
0	3364	Facebook	Irrelevant	I mentioned on Facebook that I was struggling ...
1	352	Amazon	Neutral	BBC News - Amazon boss Jeff Bezos rejects clai...
2	8312	Microsoft	Negative	@Microsoft Why do I pay for WORD when it funct...
3	4371	CS-GO	Negative	CSGO matchmaking is so full of closet hacking,...
4	4433	Google	Neutral	Now the President is slapping Americans in the...
...
995	4891	GrandTheftAuto(GTA)	Irrelevant	⭐ Toronto is the arts and culture capital of ...
996	4359	CS-GO	Irrelevant	tHIS IS ACTUALLY A GOOD MOVE TOT BRING MORE VI...
997	2652	Borderlands	Positive	Today sucked so it's time to drink wine n play...
998	8069	Microsoft	Positive	Bought a fraction of Microsoft today. Small wins.
999	6960	johnson&johnson	Neutral	Johnson & Johnson to stop selling talc baby po...

1000 rows × 4 columns

```
In [252...]: print(df_train.shape)
(74682, 4)

In [253...]: print(df_train.columns)
Index(['Id', 'Entity', 'Sentiment', 'Content'], dtype='object')

In [254...]: print(df_train.dtypes)
Id          int64
Entity      object
Sentiment    object
Content     object
dtype: object

In [255...]: df_train.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 74682 entries, 0 to 74681
Data columns (total 4 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   Id          74682 non-null   int64  
 1   Entity      74682 non-null   object  
 2   Sentiment    74682 non-null   object  
 3   Content     73996 non-null   object  
dtypes: int64(1), object(3)
memory usage: 2.3+ MB

In [256...]: df_train.describe()
```

```
Out[256]:
```

	Id
count	74682.000000
mean	6432.586165
std	3740.427870
min	1.000000
25%	3195.000000
50%	6422.000000
75%	9601.000000
max	13200.000000

```
In [257...:
```

```
df_train.isnull().sum()
```

```
Out[257]:
```

```
Id          0  
Entity      0  
Sentiment    0  
Content     686  
dtype: int64
```

```
In [258...:
```

```
col=['Id' , 'Entity' , 'Sentiment' , 'Content']  
df_train
```

```
Out[258]:
```

	Id	Entity	Sentiment	Content
0	2401	Borderlands	Positive	im getting on borderlands and i will murder yo...
1	2401	Borderlands	Positive	I am coming to the borders and I will kill you...
2	2401	Borderlands	Positive	im getting on borderlands and i will kill you ...
3	2401	Borderlands	Positive	im coming on borderlands and i will murder you...
4	2401	Borderlands	Positive	im getting on borderlands 2 and i will murder ...
...
74677	9200	Nvidia	Positive	Just realized that the Windows partition of my...
74678	9200	Nvidia	Positive	Just realized that my Mac window partition is ...
74679	9200	Nvidia	Positive	Just realized the windows partition of my Mac ...
74680	9200	Nvidia	Positive	Just realized between the windows partition of...
74681	9200	Nvidia	Positive	Just like the windows partition of my Mac is I...

74682 rows × 4 columns

```
In [259...:
```

```
df_train.dropna(subset=['Content'] , inplace=True)  
df_train
```

```
Out[259]:
```

	Id	Entity	Sentiment	Content
0	2401	Borderlands	Positive	im getting on borderlands and i will murder yo...
1	2401	Borderlands	Positive	I am coming to the borders and I will kill you...
2	2401	Borderlands	Positive	im getting on borderlands and i will kill you ...
3	2401	Borderlands	Positive	im coming on borderlands and i will murder you...
4	2401	Borderlands	Positive	im getting on borderlands 2 and i will murder ...
...
74677	9200	Nvidia	Positive	Just realized that the Windows partition of my...
74678	9200	Nvidia	Positive	Just realized that my Mac window partition is ...
74679	9200	Nvidia	Positive	Just realized the windows partition of my Mac ...
74680	9200	Nvidia	Positive	Just realized between the windows partition of...
74681	9200	Nvidia	Positive	Just like the windows partition of my Mac is l...

73996 rows × 4 columns

```
In [260...:
```

```
(df_train.shape)
```

```
Out[260]:
```

```
(73996, 4)
```

```
In [261...:
```

```
df_train.Sentiment.unique()
```

```
Out[261]:
```

```
array(['Positive', 'Neutral', 'Negative', 'Irrelevant'], dtype=object)
```

```
In [262...:
```

```
df_train.Sentiment=df_train.Sentiment.replace('Irrelevant' , 'Neutral')
```

```
In [263...:
```

```
df_test.columns = ['Id' , 'Entity' , 'Sentiment' , 'Content']  
df_test
```

Out[263]:

	Id	Entity	Sentiment	Content
0	3364	Facebook	Irrelevant	I mentioned on Facebook that I was struggling ...
1	352	Amazon	Neutral	BBC News - Amazon boss Jeff Bezos rejects clai...
2	8312	Microsoft	Negative	@Microsoft Why do I pay for WORD when it funct...
3	4371	CS-GO	Negative	CSGO matchmaking is so full of closet hacking,...
4	4433	Google	Neutral	Now the President is slapping Americans in the...
...
995	4891	GrandTheftAuto(GTA)	Irrelevant	⭐ Toronto is the arts and culture capital of ...
996	4359	CS-GO	Irrelevant	tHIS IS ACTUALLY A GOOD MOVE TOT BRING MORE VI...
997	2652	Borderlands	Positive	Today sucked so it's time to drink wine n play...
998	8069	Microsoft	Positive	Bought a fraction of Microsoft today. Small wins.
999	6960	johnson&johnson	Neutral	Johnson & Johnson to stop selling talc baby po...

1000 rows × 4 columns

In [264...]

```
df_test.Sentiment=df_train.Sentiment.replace('Irrelevant' , 'Neutral')
```

In [265...]

```
df_train.Sentiment.unique()
```

Out[265]:

```
array(['Positive', 'Neutral', 'Negative'], dtype=object)
```

In [266...]

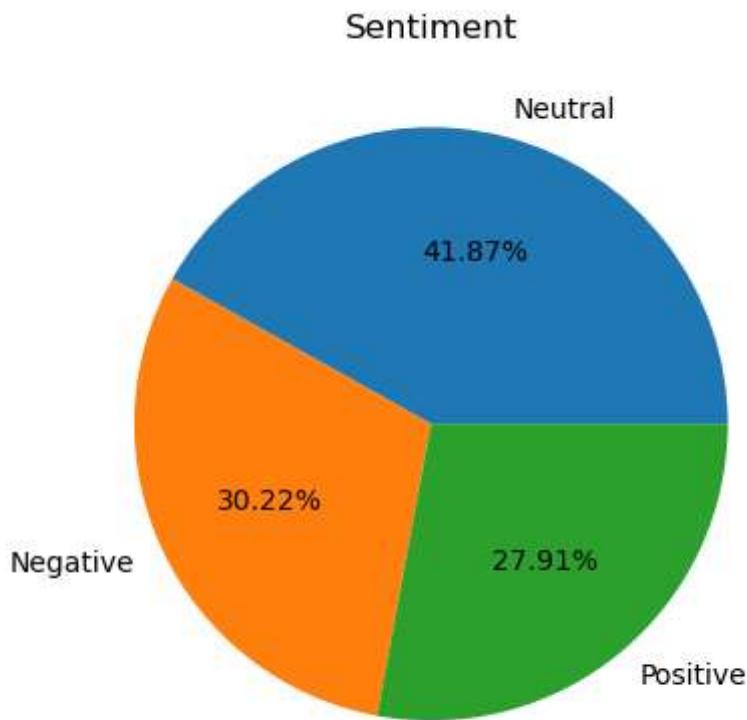
```
sentiment_count=df_train.Sentiment.value_counts()  
sentiment_count
```

Out[266]:

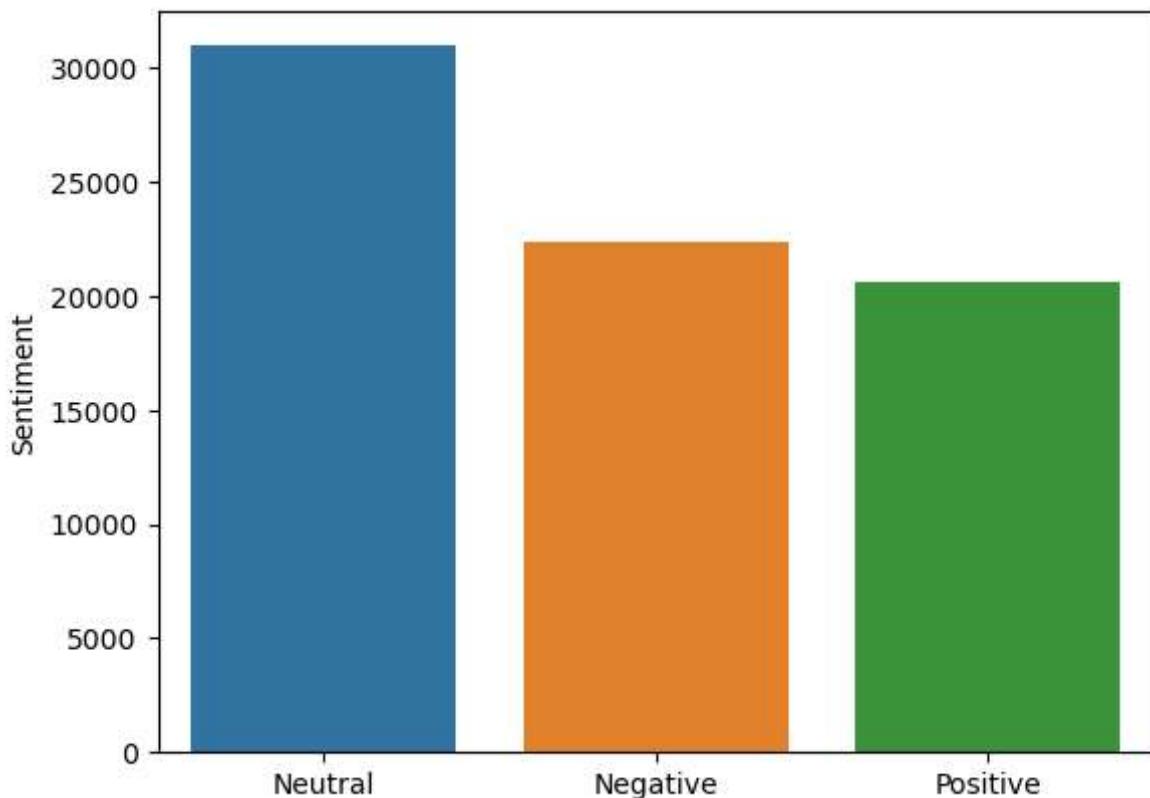
```
Neutral      30983  
Negative    22358  
Positive     20655  
Name: Sentiment, dtype: int64
```

In [267...]

```
y=['Neutral','Negative','Positive']  
plt.pie(sentiment_count,labels=y , autopct='%.2f%%')  
plt.title("Sentiment")  
plt.show()
```



```
In [270]: sb.barplot(y=sentiment_count,x=['Neutral','Negative','Positive'],)  
plt.show()
```



```
In [269]: df_train.Entity.unique()
```

```
Out[269]: array(['Borderlands', 'CallOfDutyBlackopsColdWar', 'Amazon', 'Overwatch',
   'Xbox(Xseries)', 'NBA2K', 'Dota2', 'PlayStation5(PS5)',
   'WorldOfCraft', 'CS-GO', 'Google', 'AssassinsCreed', 'ApexLegends',
   'LeagueOfLegends', 'Fortnite', 'Microsoft', 'Hearthstone',
   'Battlefield', 'PlayerUnknownsBattlegrounds(PUBG)', 'Verizon',
   'HomeDepot', 'FIFA', 'RedDeadRedemption(RDR)', 'CallOfDuty',
   'TomClancysRainbowSix', 'Facebook', 'GrandTheftAuto(GTA)',
   'MaddenNFL', 'johnson&johnson', 'Cyberpunk2077',
   'TomClancysGhostRecon', 'Nvidia'], dtype=object)
```

```
In [271... df_train.Entity.value_counts()
```

```
Out[271]:
```

MaddenNFL	2377
LeagueOfLegends	2377
CallOfDuty	2376
Verizon	2365
TomClancysRainbowSix	2364
Facebook	2362
Microsoft	2361
Dota2	2359
WorldOfCraft	2357
ApexLegends	2353
NBA2K	2343
CallOfDutyBlackopsColdWar	2343
FIFA	2324
johnson&johnson	2324
TomClancysGhostRecon	2321
Battlefield	2316
Overwatch	2316
GrandTheftAuto(GTA)	2293
HomeDepot	2292
PlayStation5(PS5)	2291
Hearthstone	2286
CS-GO	2284
Xbox(Xseries)	2283
Borderlands	2280
Amazon	2276
Google	2274
Nvidia	2271
Cyberpunk2077	2262
RedDeadRedemption(RDR)	2249
Fortnite	2249
PlayerUnknownsBattlegrounds(PUBG)	2234
AssassinsCreed	2234

```
Name: Entity, dtype: int64
```

```
In [272... y=df_train.Entity.value_counts().sort_values(ascending=False)
```

```
In [273... y
```

```
Out[273]: MaddenNFL          2377
LeagueOfLegends      2377
CallOfDuty           2376
Verizon              2365
TomClancysRainbowSix 2364
Facebook             2362
Microsoft            2361
Dota2                2359
WorldOfCraft         2357
ApexLegends          2353
NBA2K                2343
CallOfDutyBlackopsColdWar 2343
FIFA                 2324
johnson&johnson     2324
TomClancysGhostRecon 2321
Overwatch             2316
Battlefield           2316
GrandTheftAuto(GTA)   2293
HomeDepot             2292
PlayStation5(PS5)     2291
Hearthstone          2286
CS-GO                2284
Xbox(Xseries)        2283
Borderlands           2280
Amazon               2276
Google               2274
Nvidia               2271
Cyberpunk2077        2262
RedDeadRedemption(RDR) 2249
Fortnite              2249
PlayerUnknownsBattlegrounds(PUBG) 2234
AssassinsCreed       2234
Name: Entity, dtype: int64
```

```
In [274...]: entity_head=y.head(10)
entity_head
```

```
Out[274]: MaddenNFL          2377
LeagueOfLegends      2377
CallOfDuty           2376
Verizon              2365
TomClancysRainbowSix 2364
Facebook             2362
Microsoft            2361
Dota2                2359
WorldOfCraft         2357
ApexLegends          2353
Name: Entity, dtype: int64
```

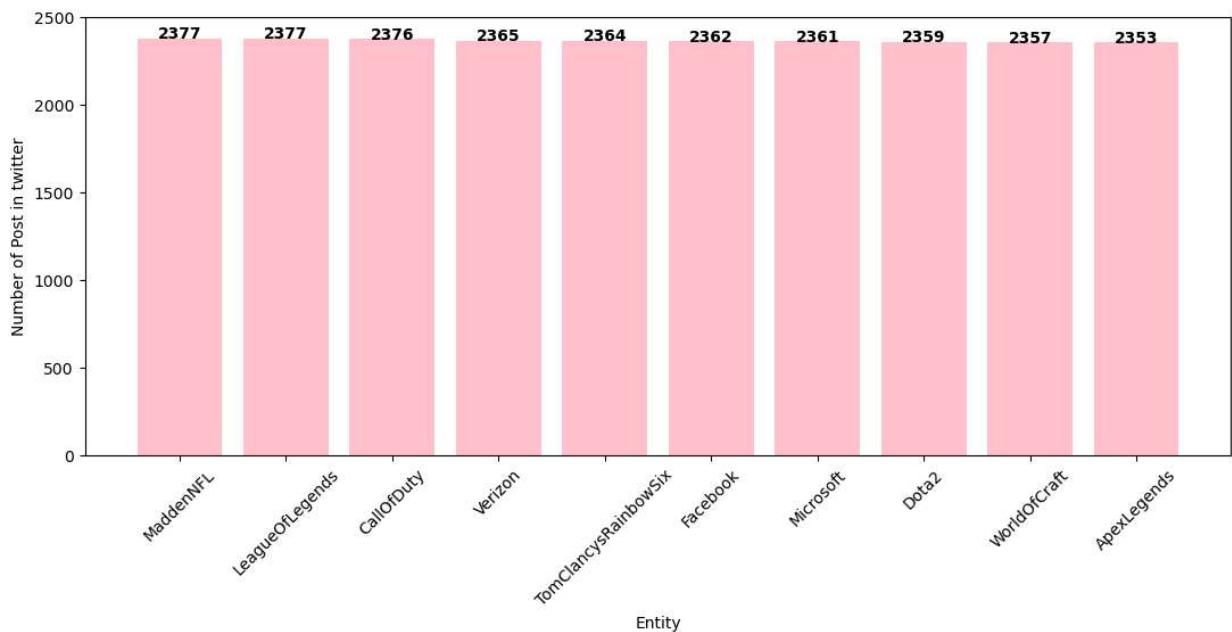
```
In [295...]: x=(entity_head.index)
x
```

```
Out[295]: Index(['MaddenNFL', 'LeagueOfLegends', 'CallOfDuty', 'Verizon',
                  'TomClancysRainbowSix', 'Facebook', 'Microsoft', 'Dota2',
                  'WorldOfCraft', 'ApexLegends'],
                  dtype='object')
```

```
In [294...]: y=entity_head.values
y
```

```
Out[294]: array([2377, 2377, 2376, 2365, 2364, 2362, 2361, 2359, 2357, 2353],  
                 dtype=int64)
```

```
In [296... plt.figure(figsize=(13,5))  
plt.bar(x,y, color='pink')  
plt.xticks(rotation=45)  
for i,v in enumerate(y):  
    plt.text(i,v,str(v),ha='center',weight='bold' )  
plt.yticks(range(0,3000,500))  
plt.xlabel('Entity')  
plt.ylabel('Number of Post in twitter')  
plt.show()
```



```
In [ ]: entity_top3_df=entity_head.head(3)  
entity_top3_df
```

```
In [234... entity_top3=entity_top3_df.index.tolist()  
entity_top3
```

```
Out[234]: ['MaddenNFL', 'LeagueOfLegends', 'CallOfDuty']
```

```
In [235... sentiment_by_entity=df_train.loc[df_train['Entity'].isin(entity_top3)].groupby('Entity')  
sentiment_by_entity
```

```
Out[235]: Entity      Sentiment  
CallOfDuty    Negative      883  
                  Neutral     1047  
                  Positive     446  
LeagueOfLegends Negative     632  
                      Neutral    1130  
                      Positive    615  
MaddenNFL     Negative    1694  
                  Neutral     287  
                  Positive    396  
Name: Sentiment, dtype: int64
```

```
In [277... plt.figure(figsize=(10,5))  
  
y=['Neutral' , 'Negative' , 'Positive']
```

```

color=['#9C6383' , '#839C63' , '#63839C']

plt.subplot(1,3,1)
plt.pie(sentiment_by_entity[:3] , labels=y , autopct='%.1f%%' , textprops={'fontsize':10, 'color':color[0]})

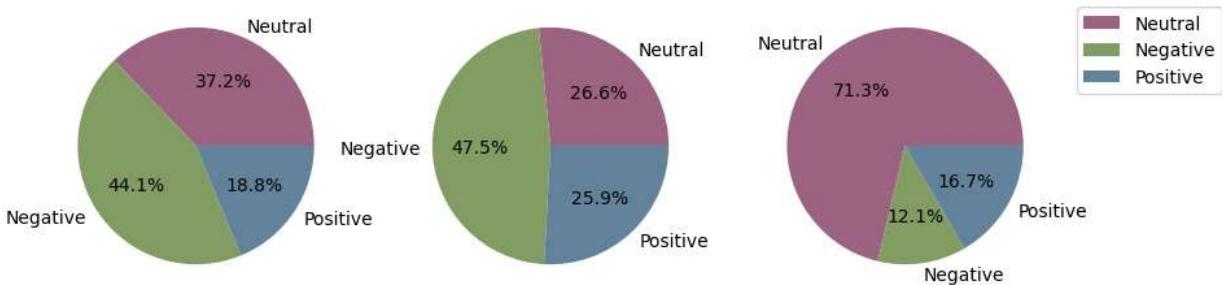
plt.subplot(1,3,2)
plt.pie(sentiment_by_entity[3:6] , labels=y , autopct='%.1f%%' , textprops={'fontsize':10, 'color':color[1]})

plt.subplot(1,3,3)
plt.pie(sentiment_by_entity[6:] , labels=y , autopct='%.1f%%' , textprops={'fontsize':10, 'color':color[2]})

plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left' , fontsize="10")

```

Out[277]: <matplotlib.legend.Legend at 0x28e807645d0>



In [278]: df_train

	Id	Entity	Sentiment	Content
0	2401	Borderlands	Positive	im getting on borderlands and i will murder yo...
1	2401	Borderlands	Positive	I am coming to the borders and I will kill you...
2	2401	Borderlands	Positive	im getting on borderlands and i will kill you ...
3	2401	Borderlands	Positive	im coming on borderlands and i will murder you...
4	2401	Borderlands	Positive	im getting on borderlands 2 and i will murder ...
...
74677	9200	Nvidia	Positive	Just realized that the Windows partition of my...
74678	9200	Nvidia	Positive	Just realized that my Mac window partition is ...
74679	9200	Nvidia	Positive	Just realized the windows partition of my Mac ...
74680	9200	Nvidia	Positive	Just realized between the windows partition of...
74681	9200	Nvidia	Positive	Just like the windows partition of my Mac is l...

73996 rows × 4 columns

In [279]: df_train.drop(['Id'] , axis=1 , inplace=True)
df_test.drop(['Id'] , axis=1 , inplace=True)

In [280]: #train test split
X_train=df_train.drop(['Sentiment'] , axis=1)
X_test=df_test.drop(['Sentiment'] , axis=1)
y_train=df_train['Sentiment']
y_test=df_test['Sentiment']

```
In [281... df_train.Sentiment.unique())
Out[281]: array(['Positive', 'Neutral', 'Negative'], dtype=object)

In [282... #count the no of words in a sentence
from sklearn.feature_extraction.text import CountVectorizer

In [283... v=CountVectorizer()
X_train_count=v.fit_transform(X_train.Content)

In [284... #Label Encoding
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
y_train=le.fit_transform(y_train)
y_test=le.fit_transform(y_test)

In [285... y_train
Out[285]: array([2, 2, 2, ..., 2, 2, 2])

In [286... X_train.drop(['Entity'],axis=1,inplace=True)
X_test.drop(['Entity'],axis=1,inplace=True)

In [287... #model
from sklearn.naive_bayes import MultinomialNB
model=MultinomialNB()
model.fit(X_train_count,y_train)

Out[287]: ▾ MultinomialNB
           MultinomialNB()

In [288... comment=[
    'I am coming to the borders and I will kill you.'
]
comment_count=v.transform(comment)
model.predict(comment_count)

Out[288]: array([2])

In [289... X_test_count=v.transform(X_test.Content)
X_test_count.toarray()

Out[289]: array([[0, 0, 0, ..., 0, 0, 0],
   [0, 0, 0, ..., 0, 0, 0],
   [0, 0, 0, ..., 0, 0, 0],
   ...,
   [0, 0, 0, ..., 0, 0, 0],
   [0, 0, 0, ..., 0, 0, 0],
   [0, 0, 0, ..., 0, 0, 0]], dtype=int64)

In [290... X_test_count.shape
Out[290]: (1000, 31062)
```

```
In [291... #score
```

```
model.score(X_test_count,y_test)
```

```
Out[291]: 0.358
```

```
In [292... X_test_count
```

```
Out[292]: <1000x31062 sparse matrix of type '<class 'numpy.int64'>'  
with 18593 stored elements in Compressed Sparse Row format>
```

```
In [ ]:
```