```python
import pandas as pd
import numpy as np
import re
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import confusion_matrix
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import StratifiedKFold
from sklearn.cluster import KMeans
from sklearn.preprocessing import MinMaxScaler
from sklearn.naive_bayes import MultinomialNB
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn import tree
import nltk
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from textblob import TextBlob
import matplotlib.pyplot as plt
import seaborn as sn
from nltk.corpus import stopwords
from sklearn.ensemble import GradientBoostingClassifier
```

In [2]:
```python
df = pd.read_csv("complete dataset3.csv")
```

In [3]:
```python
df
```

Out[3]:

	Unnamed: 0.2	Unnamed: 0.1	Unnamed: 0	Time	Tweets	Location	Name	User ID	Language	Default Profile	defa Pro Ima
0	0	0	0	2022-09-13 04:01:53+00:00	i dont wanna be one of those people that vent ...	khi	GoodGuyAdeel	1057611001607131136	en	False	Fa
1	1	1	0	2022-09-12 01:40:22+00:00	hey do you wanna vent?	Kingdom in the Sky	Steel1901	1258203430281248768	en	True	Fa
2	2	2	1	2022-09-11 17:48:29+00:00	I wanna kill someone rn had to instantly chan...	Lahore, Pakistan	sleepblablublee	1341762820673777664	en	True	Fa
3	3	3	0	2022-09-18 19:19:44+00:00	Pak Estb is on suicide mission\n\nwhile India ...	Lahore, Pakistan	syedalisherazi5	151424609	en	False	Fa
4	4	4	1	2022-09-18 18:19:35+00:00	Trust in His Plans. Don't fall short of Life....	Srinagar, Jammu & Kashmir	sixsageuwu	1566387913054507010	en	True	Fa
...	...	...	...	...	...	...	...	...	...	...	...
5565	5565	2	2	2022-10-08 16:28:59+00:00	Late 2019 and the entirety of 2020 was horribl...	Karachi, Pakistan	Tattooed_Giant	445013621	en	False	Fa
5566	5566	0	0	2022-10-10 18:54:42+00:00	4 year old kid just wanted to become a dentist...	wano	IbsDaDon	753780898093490178	en	True	Fa
5567	5567	1	1	2022-10-10 10:30:35+00:00	😩 wanted to die 😩	Lahore	nauzu_billa	3383170039	en	True	Fa
5568	5568	2	2	2022-10-10 09:09:13+00:00	😩 I wanted to die	Lahore	nauzu_billa	3383170039	en	True	Fa

5569	5569	0	0	2022-10-04 11:32:05+00:00	He wanted her blood to have the throne, it's ...	Lyallpur, Pakistan	_autophilee__	1179371269394964482	en	True	Fa

5570 rows × 20 columns

In [4]:
```python
df = df.drop('Time', axis='columns')
df = df.drop('Location', axis='columns')
df = df.drop('Name', axis='columns')
df = df.drop('User ID', axis='columns')
df = df.drop('Language', axis='columns')
df = df.drop('Default Profile', axis='columns')
df = df.drop('default Profile Image', axis='columns')
df = df.drop('description', axis='columns')
df = df.drop('favourites Count', axis='columns')
df = df.drop('follow', axis='columns')
df = df.drop('followersCount', axis='columns')
df = df.drop('following', axis='columns')
df = df.drop('friends', axis='columns')
df = df.drop('profile Background', axis='columns')
df = df.drop('profile image url', axis='columns')
df = df.drop('staruses Count', axis='columns')
df = df.drop('Unnamed: 0.2', axis='columns')
df = df.drop('Unnamed: 0.1', axis='columns')
df = df.drop('Unnamed: 0', axis='columns')
```

In [5]:
```python
df
```

Out[5]:

	Tweets
0	i dont wanna be one of those people that vent ...
1	hey do you wanna vent?
2	I wanna kill someone rn had to instantly chan...
3	Pak Estb is on suicide mission\n\nwhile India ...
4	Trust in His Plans. Don't fall short of Life....
...	...
5565	Late 2019 and the entirety of 2020 was horribl...
5566	4 year old kid just wanted to become a dentist...
5567	😩 wanted to die 😩
5568	😩 I wanted to die
5569	He wanted her blood to have the throne, it's ...

5570 rows × 1 columns

In [6]:
```python
df = df.drop_duplicates(keep="last")
```

In [7]:
```python
df
```

Out[7]:

	Tweets
0	i dont wanna be one of those people that vent ...
1	hey do you wanna vent?
2	I wanna kill someone rn had to instantly chan...
24	A girl in Chandigarh University commited_suici...
39	A shocking news coming from Chandigarh Univers...
...	...
5565	Late 2019 and the entirety of 2020 was horribl...
5566	4 year old kid just wanted to become a dentist...
5567	😩 wanted to die 😩
5568	😩 I wanted to die
5569	He wanted her blood to have the throne, it's ...

4380 rows × 1 columns

In [8]:
```python
df= df.reset_index()
```

In [9]:
```python
df
```

Out[9]:

	index	Tweets
0	0	i dont wanna be one of those people that vent ...
1	1	hey do you wanna vent?
2	2	I wanna kill someone rn had to instantly chan...
3	24	A girl in Chandigarh University commited_suici...
4	39	A shocking news coming from Chandigarh Univers...
...	...	...
4375	5565	Late 2019 and the entirety of 2020 was horribl...
4376	5566	4 year old kid just wanted to become a dentist...
4377	5567	😩 wanted to die 😩
4378	5568	😩 I wanted to die
4379	5569	He wanted her blood to have the throne, it's ...

4380 rows × 2 columns

In [10]:
```python
df = df.drop('index', axis='columns')
```

In [11]:
```python
df
```

Out[11]:

	Tweets
0	i dont wanna be one of those people that vent ...
1	hey do you wanna vent?
2	I wanna kill someone rn had to instantly chan...
3	A girl in Chandigarh University commited_suici...
4	A shocking news coming from Chandigarh Univers...
...	...
4375	Late 2019 and the entirety of 2020 was horribl...
4376	4 year old kid just wanted to become a dentist...
4377	😩 wanted to die 😩
4378	😩 I wanted to die
4379	He wanted her blood to have the throne, it's ...

4380 rows × 1 columns

In [12]:
```python
#Data Cleaning
def cleanTxt(text):
    text = text.lower() #converting whole text into lower
    text = re.compile('rt @').sub('@', text) # removing tags
    text= re.sub('@[A-Za-z0-9]+','',text ) #removing mentions
    text= re.sub("#",'',text) #removing mentions
    text = re.sub('[()!?]', ' ', text) #removing special letters
    text = re.sub(r'http\S+', '', text) #removing links
    text = re.sub('\[.*?\]',' ', text) #removing special letters
    text = re.sub("[^a-z0-9]"," ", text) #removing special letters
    text = re.compile('https?:\/\/\S+').sub('@', text, count=1) #removing links
    return text #to get an output of text after cleaning it
```

In [13]:
```python
df['Tweets'] = df['Tweets'].apply(cleanTxt)
```

In [14]:
```python
df
```

Out[14]:

	Tweets
0	i dont wanna be one of those people that vent ...
1	hey do you wanna vent
2	i wanna kill someone rn had to instantly chan...
3	a girl in chandigarh university commited suici...
4	a shocking news coming from chandigarh univers...
...	...
4375	late 2019 and the entirety of 2020 was horribl...
4376	4 year old kid just wanted to become a dentist...
4377	wanted to die
4378	i wanted to die
4379	he wanted her blood to have the throne it s ...

4380 rows × 1 columns

In [15]:
```python
df = df.drop_duplicates(keep='last')
```

In [16]:
```python
df
```

Out[16]:

	Tweets
0	i dont wanna be one of those people that vent ...
1	hey do you wanna vent
2	i wanna kill someone rn had to instantly chan...
3	a girl in chandigarh university commited suici...
4	a shocking news coming from chandigarh univers...
...	...
4375	late 2019 and the entirety of 2020 was horribl...
4376	4 year old kid just wanted to become a dentist...
4377	wanted to die
4378	i wanted to die
4379	he wanted her blood to have the throne it s ...

4206 rows × 1 columns

In [17]:
```python
df = df.reset_index()
```

In [20]:
```python
df = df.drop('index', axis='columns')
```

In [21]:
```python
df
```

Out[21]:

	Tweets
0	i dont wanna be one of those people that vent ...
1	hey do you wanna vent
2	i wanna kill someone rn had to instantly chan...
3	a girl in chandigarh university commited suici...
4	a shocking news coming from chandigarh univers...
...	...
4201	late 2019 and the entirety of 2020 was horribl...
4202	4 year old kid just wanted to become a dentist...
4203	wanted to die
4204	i wanted to die
4205	he wanted her blood to have the throne it s ...

4206 rows × 1 columns

```python
In [22]: from nltk.corpus import stopwords
```

```python
In [27]: stop = set(stopwords.words('english'))
```

```python
In [28]: def remove_stopwords(word):
             return word not in words
```

```python
In [29]: df['Tweets'] = df['Tweets'].str.lower().str.split()
```

```python
In [32]: df['Tweets'] = df['Tweets'].apply(lambda x : [item for item in x if item not in stop])
```

```python
In [33]: df
```

Out[33]:

	Tweets
0	[dont, wanna, one, people, vent, twt, anything...
1	[hey, wanna, vent]
2	[wanna, kill, someone, rn, instantly, change, ...
3	[girl, chandigarh, university, commited, suici...
4	[shocking, news, coming, chandigarh, universit...
...	...
4201	[late, 2019, entirety, 2020, horrible, used, s...
4202	[4, year, old, kid, wanted, become, dentist, g...
4203	[wanted, die]
4204	[wanted, die]
4205	[wanted, blood, throne, kids, hear, larys, str...

4206 rows × 1 columns

```python
In [34]: df['Tweet'] = [','.join(map(str, l)) for l in df['Tweets']]
```

```python
In [35]: df
```

Out[35]:

	Tweets	Tweet
0	[dont, wanna, one, people, vent, twt, anything...	dont,wanna,one,people,vent,twt,anything,fuck,p...
1	[hey, wanna, vent]	hey,wanna,vent
2	[wanna, kill, someone, rn, instantly, change, ...	wanna,kill,someone,rn,instantly,change,vent,babar
3	[girl, chandigarh, university, commited, suici...	girl,chandigarh,university,commited,suicide,du...
4	[shocking, news, coming, chandigarh, universit...	shocking,news,coming,chandigarh,university,gir...
...	...	...
4201	[late, 2019, entirety, 2020, horrible, used, s...	late,2019,entirety,2020,horrible,used,sleep,pa...
4202	[4, year, old, kid, wanted, become, dentist, g...	4,year,old,kid,wanted,become,dentist,got,kidna...
4203	[wanted, die]	wanted,die
4204	[wanted, die]	wanted,die
4205	[wanted, blood, throne, kids, hear, larys, str...	wanted,blood,throne,kids,hear,larys,strong,quo...

4206 rows × 2 columns

```python
In [36]: df = df.drop('Tweets', axis = 'columns')
```

```python
In [37]: df
```

Out[37]:

	Tweet
0	dont,wanna,one,people,vent,twt,anything,fuck,p...

	Tweet
1	hey,wanna,vent
2	wanna,kill,someone,rn,instantly,change,vent,babar
3	girl,chandigarh,university,commited,suicide,du...
4	shocking,news,coming,chandigarh,university,gir...
...	...
4201	late,2019,entirety,2020,horrible,used,sleep,pa...
4202	4,year,old,kid,wanted,become,dentist,got,kidna...
4203	wanted,die
4204	wanted,die
4205	wanted,blood,throne,kids,hear,larys,strong,quo...

4206 rows × 1 columns

In [38]:
```python
df['Tweet'] = df['Tweet'].apply(cleanTxt)
```

In [39]:
```python
df
```

Out[39]:

	Tweet
0	dont wanna one people vent twt anything fuck p...
1	hey wanna vent
2	wanna kill someone rn instantly change vent babar
3	girl chandigarh university commited suicide du...
4	shocking news coming chandigarh university gir...
...	...
4201	late 2019 entirety 2020 horrible used sleep pa...
4202	4 year old kid wanted become dentist got kidna...
4203	wanted die
4204	wanted die
4205	wanted blood throne kids hear larys strong quo...

4206 rows × 1 columns

In [40]:
```python
def getSubjectivity(text):
    return TextBlob(text).sentiment.subjectivity

def getPolarity(text):
    return TextBlob(text).sentiment.polarity
df['Subjectivity'] = df['Tweet'].apply(getSubjectivity)
df['Polarity'] = df['Tweet'].apply(getPolarity)
```

In [44]:
```python
def getAnalysis(score):
    if score<-0.2:
        return '0'
    else:
        return '1'

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

In [46]:
```python
df = df.drop('Subjectivity', axis='columns')
df = df.drop('Polarity', axis='columns')
```

In [51]:
```python
df
```

Out[51]:

	Tweet	Analysis
0	dont wanna one people vent twt anything fuck p...	0
1	hey wanna vent	1
2	wanna kill someone rn instantly change vent babar	1
3	girl chandigarh university commited suicide du...	1
4	shocking news coming chandigarh university gir...	0

	Tweet	Analysis
...	...	...
4201	late 2019 entirety 2020 horrible used sleep pa...	0
4202	4 year old kid wanted become dentist got kidna...	1
4203	wanted die	1
4204	wanted die	1
4205	wanted blood throne kids hear larys strong quo...	1

4206 rows × 2 columns

In [52]:
```python
df2 = df[df.Analysis=='1']
```

In [53]:
```python
df3 = df[df.Analysis=='0']
```

In [54]:
```python
df3
```

Out[54]:

	Tweet	Analysis
0	dont wanna one people vent twt anything fuck p...	0
4	shocking news coming chandigarh university gir...	0
7	k22 cases like kidnapping suicide happened ear...	0
18	started little women thinking b depressy drama...	0
23	sick taliban brother calls fot anti nrf good t...	0
...	...	...
4124	story hazara wounded dehmzang lost little brot...	0
4138	stop unpatriotic abuzar wrong people die us ac...	0
4163	lahore css girl ne committed suicide due failu...	0
4193	justiceforsitarayaseen getting brutal threats ...	0
4201	late 2019 entirety 2020 horrible used sleep pa...	0

660 rows × 2 columns

In [55]:
```python
df2
```

Out[55]:

	Tweet	Analysis
1	hey wanna vent	1
2	wanna kill someone rn instantly change vent babar	1
3	girl chandigarh university commited suicide du...	1
5	justiceforcugirls chandigarh university multip...	1
6	kh bahiss yes head alqaida living guest house ...	1
...	...	...
4200	mazaq ke ilawa aaj kal rishta shadi scene se b...	1
4202	4 year old kid wanted become dentist got kidna...	1
4203	wanted die	1
4204	wanted die	1
4205	wanted blood throne kids hear larys strong quo...	1

3546 rows × 2 columns

In [56]:
```python
df.groupby('Analysis').describe()
```

Out[56]:

			Tweet	
	count	unique	top	freq
Analysis				
0	660	620	hate	9
1	3546	3391	either surrender fight freedom till death ask ...	16

```python
In [57]:   v = CountVectorizer()
           Tweet = v.fit_transform(df.Tweet.values)
```

```python
In [58]:   Tweet
```

```
Out[58]:   <4206x12426 sparse matrix of type '<class 'numpy.int64'>'
                   with 61737 stored elements in Compressed Sparse Row format>
```

```python
In [59]:   X_train, X_test, y_train, y_test = train_test_split(Tweet, df.Analysis, test_size=0.2)
```

```python
In [60]:   model = LogisticRegression()
```

```python
In [61]:   model.fit(X_train, y_train)
```

```
Out[61]:   LogisticRegression()
```

```python
In [62]:   model.score(X_test, y_test)
```

```
Out[62]:   0.9109263657957245
```

```python
In [63]:   from sklearn.metrics import classification_report, accuracy_score
           def classificationReport(y_true, y_pre):
               print(classification_report(y_true, y_pre))
               return accuracy_score(y_true, y_pre)
```

```python
In [64]:   y_pre = model.predict(X_test)
```

```python
In [65]:   print(classificationReport(y_test, y_pre))
```

```
                       precision    recall  f1-score   support

                   0        0.90      0.52      0.66       140
                   1        0.91      0.99      0.95       702

            accuracy                            0.91       842
           macro avg        0.91      0.76      0.80       842
        weighted avg        0.91      0.91      0.90       842

           0.9109263657957245
```

```python
In [78]:   model = RandomForestClassifier(n_estimators=140, criterion='entropy', max_depth=None, min_samples_split=2, min_sa
```

```python
In [79]:   model.fit(X_train, y_train)
```

```
Out[79]:   RandomForestClassifier(criterion='entropy', max_features='sqrt',
                                  n_estimators=140)
```

```python
In [80]:   model.score(X_test, y_test)
```

```
Out[80]:   0.9121140142517815
```

```python
In [81]:   y_pre = model.predict(X_test)
           print(classificationReport(y_test, y_pre))
```

```
                       precision    recall  f1-score   support
```

```
              0        0.92      0.51      0.66       140
              1        0.91      0.99      0.95       702

       accuracy                            0.91       842
      macro avg        0.92      0.75      0.81       842
   weighted avg        0.91      0.91      0.90       842
```

0.9121140142517815