In [1]:
```
pip install nltk
```

```
Requirement already satisfied: nltk in c:\users\reshma_koduri\anaconda3\lib\site-pack
ages (3.6.5)
Requirement already satisfied: click in c:\users\reshma_koduri\anaconda3\lib\site-pac
kages (from nltk) (8.0.3)
Requirement already satisfied: joblib in c:\users\reshma_koduri\anaconda3\lib\site-pa
ckages (from nltk) (1.1.0)
Requirement already satisfied: regex>=2021.8.3 in c:\users\reshma_koduri\anaconda3\li
b\site-packages (from nltk) (2021.8.3)
Requirement already satisfied: tqdm in c:\users\reshma_koduri\anaconda3\lib\site-pack
ages (from nltk) (4.62.3)
Requirement already satisfied: colorama in c:\users\reshma_koduri\anaconda3\lib\site-
packages (from click->nltk) (0.4.4)
Note: you may need to restart the kernel to use updated packages.
```

In [2]:
```python
import nltk
```

In [3]:
```python
import pandas as pd
```

In [4]:
```python
nltk.download ('punkt')
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\reshma_koduri\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```
Out[4]:
```
True
```

In [5]:
```python
fake=pd.read_csv("C:\\Users\\reshma_koduri\\Downloads\\Fake-210604-161841.csv")
genuine=pd.read_csv("C:\\Users\\reshma_koduri\\Downloads\\True-210604-161650.csv")
```

In [6]:
```python
genuine
```

Out[6]:

	title	text	subject	date
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017
...	...	...	...	...
21412	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017
21413	LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews	August 22, 2017

	title	text	subject	date
21414	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disused Sov...	worldnews	August 22, 2017
21415	Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017
21416	Indonesia to buy $1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	August 22, 2017

21417 rows × 4 columns

In [7]:
```python
fake['genuiness']=0
genuine['genuiness']=1
```

In [8]:
```python
genuine
```

Out[8]:

	title	text	subject	date	genuiness
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017	1
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017	1
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017	1
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017	1
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017	1
...	...	...	...	...	...
21412	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017	1
21413	LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews	August 22, 2017	1
21414	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disused Sov...	worldnews	August 22, 2017	1
21415	Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017	1
21416	Indonesia to buy $1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	August 22, 2017	1

21417 rows × 5 columns

In [9]:
```python
dataset = pd.concat([fake,genuine])
```

In [10]:
```python
dataset
```

Out[10]:

	title	text	subject	date	genuiness
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017	0
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	0
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	0
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	0
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	0
...	...	...	...	...	...
21412	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017	1
21413	LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews	August 22, 2017	1
21414	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disued Sov...	worldnews	August 22, 2017	1
21415	Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017	1
21416	Indonesia to buy $1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	August 22, 2017	1

44898 rows × 5 columns

In [11]:
```python
dataset=dataset.reset_index(drop=True)
```

In [12]:
```python
dataset
```

Out[12]:

	title	text	subject	date	genuiness
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017	0
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	0
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	0
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	0

	title	text	subject	date	genuiness
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	0
...	...	...	...	...	...
44893	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017	1
44894	LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews	August 22, 2017	1
44895	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disued Sov...	worldnews	August 22, 2017	1
44896	Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017	1
44897	Indonesia to buy $1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	August 22, 2017	1

44898 rows × 5 columns

```
In [13]:  dataset= dataset.drop(['date','title','subject'], axis=1)
```

```
In [14]:  dataset
```

Out[14]:

	text	genuiness
0	Donald Trump just couldn t wish all Americans ...	0
1	House Intelligence Committee Chairman Devin Nu...	0
2	On Friday, it was revealed that former Milwauk...	0
3	On Christmas day, Donald Trump announced that ...	0
4	Pope Francis used his annual Christmas Day mes...	0
...	...	...
44893	BRUSSELS (Reuters) - NATO allies on Tuesday we...	1
44894	LONDON (Reuters) - LexisNexis, a provider of l...	1
44895	MINSK (Reuters) - In the shadow of disued Sov...	1
44896	MOSCOW (Reuters) - Vatican Secretary of State ...	1
44897	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	1

44898 rows × 2 columns

```
In [15]:  from nltk.tokenize import word_tokenize
          dataset['text'] = dataset['text'].apply(word_tokenize)
```

```
In [16]:  dataset
```

Out[16]:

	text	genuiness
0	[Donald, Trump, just, couldn, t, wish, all, Am...	0
1	[House, Intelligence, Committee, Chairman, Dev...	0
2	[On, Friday, ,, it, was, revealed, that, forme...	0
3	[On, Christmas, day, ,, Donald, Trump, announc...	0
4	[Pope, Francis, used, his, annual, Christmas, ...	0
...	...	...
44893	[BRUSSELS, (, Reuters,), -, NATO, allies, on,...	1
44894	[LONDON, (, Reuters,), -, LexisNexis, ,, a, p...	1
44895	[MINSK, (, Reuters,), -, In, the, shadow, of,...	1
44896	[MOSCOW, (, Reuters,), -, Vatican, Secretary,...	1
44897	[JAKARTA, (, Reuters,), -, Indonesia, will, b...	1

44898 rows × 2 columns

In [17]:
```python
from nltk.stem.snowball import SnowballStemmer
```

In [18]:
```python
sb =SnowballStemmer('english')
```

In [19]:
```python
def stem_it(text):
    return[sb.stem(word) for word in text]
```

In [20]:
```python
dataset['text']=dataset['text'].apply(stem_it)
```

In [21]:
```python
dataset
```

Out[21]:

	text	genuiness
0	[donald, trump, just, couldn, t, wish, all, am...	0
1	[hous, intellig, committe, chairman, devin, nu...	0
2	[on, friday, ,, it, was, reveal, that, former,...	0
3	[on, christma, day, ,, donald, trump, announc,...	0
4	[pope, franci, use, his, annual, christma, day...	0
...	...	...
44893	[brussel, (, reuter,), -, nato, alli, on, tue...	1
44894	[london, (, reuter,), -, lexisnexi, ,, a, pro...	1
44895	[minsk, (, reuter,), -, in, the, shadow, of, ...	1
44896	[moscow, (, reuter,), -, vatican, secretari, ...	1
44897	[jakarta, (, reuter,), -, indonesia, will, bu...	1

44898 rows × 2 columns

stopword removal

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

In [23]:
```python
def stopword_removal(text):
    dt=[word for word in text if len(word)>>2]
    return dt
```

In [24]:
```python
dataset['text']=dataset['text'].apply(stopword_removal)
```

In [25]:
```python
dataset
```

Out[25]:

	text	genuiness
0	[donald, trump, just, couldn, wish, american, ...	0
1	[hous, intellig, committe, chairman, devin, nu...	0
2	[friday, reveal, that, former, milwauke, sheri...	0
3	[christma, donald, trump, announc, that, would...	0
4	[pope, franci, annual, christma, messag, rebuk...	0
...	...	...
44893	[brussel, reuter, nato, alli, tuesday, welcom,...	1
44894	[london, reuter, lexisnexi, provid, legal, reg...	1
44895	[minsk, reuter, shadow, disus, soviet-era, fac...	1
44896	[moscow, reuter, vatican, secretari, state, ca...	1
44897	[jakarta, reuter, indonesia, will, sukhoi, fig...	1

44898 rows × 2 columns

In [26]:
```python
dataset['text']=dataset['text'].apply(''.join)
```

In [27]:
```python
dataset
```

Out[27]:

	text	genuiness
0	donaldtrumpjustcouldnwishamericanhappiyearleav...	0
1	housintelligcommittechairmandevinnunehavebeenu...	0
2	fridayrevealthatformermilwaukesheriffdavidclar...	0
3	christmadonaldtrumpannouncthatwouldbackworkfol...	0
4	popefranciannualchristmamessagrebukdonaldtrump...	0
...	...	...

	text	genuiness
44893	brusselreuternatoallituesdaywelcompresiddonald...	1
44894	londonreuterlexisnexiprovidlegalregulatoribusi...	1
44895	minskreutershadowdisussoviet-erafactoriminskst...	1
44896	moscowreutervaticansecretaristatecardinpietrop...	1
44897	jakartareuterindonesiawillsukhoifighterworth1....	1

44898 rows × 2 columns

In [28]:
```python
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(dataset['text'],dataset['genuiness'],
```

In [29]:
```python
x_train
```

Out[29]:
```
17348
38243    colomboreuterbritishwoman800,000rupe5,200compe...
7022     muchdonaldtrumpdoesnwantadmitabsolutbombpollwi...
22769    dailishooter21stcenturiwirejustroutdeirezzorsy...
21640    donaldtrumpannouncwillpresidrepublican2016here...
                        ...
29624    berlinreutergermanimustbraceitselfturbultimeun...
30070    washingtonreuterpresident-electdonaldtrumpstat...
7070     donaldtrumpentirimagbasearoundideathatwinnerwh...
35130    cairoreuterfoodsecurmiddleastnorthafricaquickd...
42278    istanbulreuterturkeysaidthursdaystoptrainpeshm...
Name: text, Length: 33673, dtype: object
```

In [30]:
```python
from sklearn.feature_extraction.text import TfidfVectorizer
tfidfv=TfidfVectorizer(max_df = 0.7)
```

In [31]:
```python
tfidfv_train=tfidfv.fit_transform(x_train)
tfidfv_test=tfidfv.transform(x_test)
```

In [32]:
```python
print(tfidfv_train)
```

```
  (1, 138202)    0.4180281451798651
  (1, 154457)    0.35533413737974595
  (1, 5224)      0.4180281451798651
  (1, 11055)     0.4180281451798651
  (1, 5222)      0.4180281451798651
  (1, 61774)     0.4180281451798651
  (2, 260856)    0.3779644730092272
  (2, 265770)    0.3779644730092272
  (2, 244820)    0.3779644730092272
  (2, 32960)     0.3779644730092272
  (2, 261026)    0.3779644730092272
  (2, 29516)     0.3779644730092272
  (2, 151459)    0.3779644730092272
  (3, 242898)    0.08200799177694976
  (3, 194061)    0.16826487435065487
  (3, 46145)     0.15228180205398417
  (3, 181591)    0.16826487435065487
  (3, 211636)    0.16826487435065487
  (3, 10705)     0.16826487435065487
```

```
(3, 225885)    0.16826487435065487
(3, 118384)    0.16826487435065487
(3, 187634)    0.16826487435065487
(3, 22506)     0.16826487435065487
(3, 35073)     0.16826487435065487
(3, 109731)    0.16826487435065487
  :      :
(33668, 248822)        0.3162277660168379
(33668, 177131)        0.3162277660168379
(33668, 45534)         0.3162277660168379
(33669, 84826)         0.6382886173859965
(33669, 82794)         0.6382886173859965
(33669, 251035)        0.4303199761002805
(33670, 52615)         0.36960514132709316
(33670, 74419)         0.36960514132709316
(33670, 63183)         0.36960514132709316
(33670, 241760)        0.36960514132709316
(33670, 108494)        0.36960514132709316
(33670, 189552)        0.36960514132709316
(33670, 76790)         0.36960514132709316
(33670, 91138)         0.20915132448076074
(33671, 99584)         0.38059680191442596
(33671, 109476)        0.38059680191442596
(33671, 184064)        0.38059680191442596
(33671, 40007)         0.38059680191442596
(33671, 94638)         0.38059680191442596
(33671, 53279)         0.38059680191442596
(33671, 28335)         0.3617684981242388
(33672, 132075)        0.5
(33672, 70892)         0.5
(33672, 145725)        0.5
(33672, 123792)        0.5
```

In [33]:
```python
from sklearn.linear_model import LogisticRegression
```

In [34]:
```python
lr=LogisticRegression(max_iter=1000)
lr.fit(tfidfv_train,y_train)
```

Out[34]: LogisticRegression(max_iter=1000)

In [35]:
```python
y_pred=lr.predict(tfidfv_test)
```

In [36]:
```python
from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred)
```

Out[36]: 0.8547884187082405

In [37]:
```python
from sklearn.linear_model import PassiveAggressiveClassifier
pac=PassiveAggressiveClassifier(max_iter=100)
pac.fit(tfidfv_train,y_train)
```

Out[37]: PassiveAggressiveClassifier(max_iter=100)

In [38]:
```python
y_pred_pac=pac.predict(tfidfv_test)
accuracy_score(y_test,y_pred_pac)
```

Out[38]:   0.8653897550111359

In [ ]: