

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

Requirement already satisfied: nltk in c:\users\mdnayaz\anaconda3\lib\site-packages (3.5)Note: you may need to restart the kernel to use updated packages.

Requirement already satisfied: click in c:\users\mdnayaz\anaconda3\lib\site-packages (from nltk) (7.1.2)

Requirement already satisfied: joblib in c:\users\mdnayaz\anaconda3\lib\site-packages (from nltk) (0.16.0)

Requirement already satisfied: tqdm in c:\users\mdnayaz\anaconda3\lib\site-packages (from nltk) (4.47.0)

Requirement already satisfied: regex in c:\users\mdnayaz\anaconda3\lib\site-packages (from nltk) (2020.6.8)

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

```
In [3]: nltk.download('punkt')
```

[nltk_data] Downloading package punkt to

[nltk_data] C:\Users\MDNayaz\AppData\Roaming\nltk_data...

[nltk_data] Unzipping tokenizers\punkt.zip.

```
Out[3]: True
```

```
In [4]: import pandas as pd
```

```
In [6]: fake = pd.read_csv("Fake.csv")
genuine = pd.read_csv("True.csv")
```

```
In [7]: display(fake.info())
display(genuine.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 23481 entries, 0 to 23480
Data columns (total 4 columns):
#   Column   Non-Null Count  Dtype
---  ---
0    title   23481 non-null  object
1    text    23481 non-null  object
2    subject 23481 non-null  object
3    date    23481 non-null  object
dtypes: object(4)
memory usage: 733.9+ KB
```

None

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21417 entries, 0 to 21416
Data columns (total 4 columns):
#   Column   Non-Null Count  Dtype
---  ---
0    title   21417 non-null  object
1    text    21417 non-null  object
2    subject 21417 non-null  object
3    date    21417 non-null  object
dtypes: object(4)
memory usage: 669.4+ KB
```

None

```
In [8]: display(fake.head())
display(genuine.head())
```

	title	text	subject	date
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017

	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

```
In [9]: display(fake.subject.value_counts())
print('\n')
display(genuine.subject.value_counts())
```

```
News          9050
politics      6841
left-news     4459
Government News 1570
US_News       783
Middle-east   778
Name: subject, dtype: int64
```

```
politicsNews  11272
worldnews     10145
Name: subject, dtype: int64
```

```
In [10]: fake['target'] = 0
genuine['target'] = 1
```

```
In [11]: display(fake.head())
```

	title	text	subject	date	target
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

```
In [12]: display(genuine.head())
```

	title	text	subject	date	target
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

```
In [13]: data = pd.concat([fake, genuine], axis=0)
```

```
In [14]: data = data.reset_index(drop=True)
```

```
In [15]: data=data.drop(['subject','date','title'], axis=1)
```

```
In [16]: print(data.columns)
print(data)
```

Index(['text', 'target'], dtype='object')

	text	target
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 disused Sov...	1
44896	MOSCOW (Reuters) - Vatican Secretary of State ...	1
44897	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	1

[44898 rows x 2 columns]

TOKENIZATION

```
In [17]: from nltk.tokenize import word_tokenize
data['text']=data['text'].apply(word_tokenize)
```

```
In [18]: print(data.head(10))
```

	text	target
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
5	[The, number, of, cases, of, cops, brutalizing...	0
6	[Donald, Trump, spent, a, good, portion, of, h...	0
7	[In, the, wake, of, yet, another, court, decis...	0
8	[Many, people, have, raised, the, alarm, regar...	0
9	[Just, when, you, might, have, thought, we, d,...	0

STEMMING

```
In [19]: from nltk.stem.snowball import SnowballStemmer
porter = SnowballStemmer("english", ignore_stopwords=False)
```

```
In [20]: def stem_it(text):
return [porter.stem(word) for word in text]
```

```
In [21]: data['text']=data['text'].apply(stem_it)
```

```
In [22]: print(data.head(10))
```

	text	target
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
5	[the, number, of, case, of, cop, brutal, and, ...	0
6	[donald, trump, spent, a, good, portion, of, h...	0
7	[in, the, wake, of, yet, anoth, court, decis, ...	0
8	[mani, peopl, have, rais, the, alarm, regard, ...	0
9	[just, when, you, might, have, thought, we, d,...	0

Stopword removal

```
In [23]: #import nltk
#from nltk.corpus import stopwords
#nltk.download('stopwords')
#print(stopwords.words('english'))
```

```
In [24]: def stop_it(t):
dt = [word for word in t if len(word)>2]
return dt
```

```
In [25]: data['text']=data['text'].apply(stop_it)
```

```
In [26]: print(data['text'].head(10))
```

```
0    [donald, trump, just, couldn, wish, all, ameri...
1    [hous, intellig, committe, chairman, devin, nu...
2    [friday, was, reveal, that, former, milwauke, ...
3    [christma, day, donald, trump, announc, that, ...
4    [pope, franci, use, his, annual, christma, day...
5    [the, number, case, cop, brutal, and, kill, pe...
6    [donald, trump, spent, good, portion, his, day...
7    [the, wake, yet, anoth, court, decis, that, de...
8    [mani, peopl, have, rais, the, alarm, regard, ...
9    [just, when, you, might, have, thought, get, b...
Name: text, dtype: object
```

```
In [27]: data['text']=data['text'].apply(' '.join)
```

Splitting up of data

```
In [28]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(data['text'], data['target'], tes
t_size=0.25)
display(X_train.head())
print('\n')
display(y_train.head())
```

```
37214    brasilia reuter the unit nation has ask brazil...
3911     donald trump twitter account litter with insul...
18945    menac donald trump gaze down phoenix grand ave...
27438    tallahassee fla. reuter republican democrat and...
2699     you haven notic now donald trump has been sign...
Name: text, dtype: object
```

```
37214    1
3911     0
18945    0
27438    1
2699     0
Name: target, dtype: int64
```

Vectorization

```
In [29]: from sklearn.feature_extraction.text import TfidfVectorizer
my_tfidf = TfidfVectorizer( max_df=0.7)

tfidf_train = my_tfidf.fit_transform(X_train)
tfidf_test = my_tfidf.transform(X_test)
```

```
In [30]: print(tfidf_train)
```

```
(0, 69727)    0.03332777546594283
(0, 84939)    0.021895638919709
(0, 88608)    0.015968372548811262
(0, 89812)    0.04928639380224759
(0, 8899)     0.016201934964951956
(0, 9036)     0.03233232079164979
(0, 28222)    0.052217232938539325
(0, 59911)    0.03875431075734922
(0, 64593)    0.02076985522657517
(0, 10421)    0.03435157054329103
(0, 75501)    0.020980772927501612
(0, 70641)    0.030850874147623202
(0, 25836)    0.034437952979640724
(0, 33120)    0.04026983040028881
(0, 17677)    0.01699639134018499
(0, 88219)    0.021201379604791366
(0, 28212)    0.031051990797366452
(0, 55747)    0.019964719877644102
(0, 69495)    0.03959411279801234
(0, 86032)    0.020121594642351356
(0, 66194)    0.0673956429123436
(0, 30232)    0.025886404489956036
(0, 81374)    0.016976808713143998
(0, 26912)    0.02212534260869897
(0, 89847)    0.014456887356608295
:
(33672, 89847)    0.1031826749067512
(33672, 81090)    0.015937550449049175
(33672, 77734)    0.013283404614141275
(33672, 30533)    0.044443062479030127
(33672, 22275)    0.01650270369060576
(33672, 81463)    0.012482835937566553
(33672, 68252)    0.047475927301042734
(33672, 61270)    0.05052502134656695
(33672, 46363)    0.028288029465554768
(33672, 81621)    0.03817960047511823
(33672, 24586)    0.034073179153938295
(33672, 46772)    0.09774647457522485
(33672, 33243)    0.054058991574478577
(33672, 69387)    0.02720300108523269
(33672, 32521)    0.021186615473584344
(33672, 79138)    0.017027678466854542
(33672, 13405)    0.03958951320053995
(33672, 88970)    0.0693442538360088
(33672, 56524)    0.3399626191005392
(33672, 61287)    0.04236457829488147
(33672, 87124)    0.02851373356714735
(33672, 88828)    0.019082771204478606
(33672, 81546)    0.022290893003339866
(33672, 37705)    0.08558564150393264
(33672, 85323)    0.017298265069416317
```

LogisticRegression

```
In [31]: from sklearn.linear_model import LogisticRegression
         from sklearn.metrics import accuracy_score
```

```
In [32]: model_1 = LogisticRegression(max_iter=900)
model_1.fit(tfidf_train, y_train)
pred_1 = model_1.predict(tfidf_test)
cr1     = accuracy_score(y_test, pred_1)
print(cr1*100)
```

98.76169265033408

PassiveAggressiveClassifier

```
In [33]: from sklearn.linear_model import PassiveAggressiveClassifier

model = PassiveAggressiveClassifier(max_iter=50)
model.fit(tfidf_train, y_train)
```

Out[33]: PassiveAggressiveClassifier(max_iter=50)

```
In [34]: y_pred = model.predict(tfidf_test)
accscore = accuracy_score(y_test, y_pred)
print('The accuracy of prediction is ', accscore*100)
```

The accuracy of prediction is 99.61692650334075

In []: