News Classification using Natural Language Processing

ABSTRACT:

Internet is one of the most important inventions and a large number of people are its users. And there are various platforms which posts the news without checking whether it is genuine or fake. A normal human being is unable to detect the fake news. So, News Classification is developed using Natural Language Processing. This is used to determine whether the upcoming news or the news displayed on the feed is genuine or fake. Using this NLP, we can differentiate based on the language used in different set of data of fake and genuine news.

OBJECTIVE:

The goal of this project is to find out whether the news in an online platform is genuine or fake using Natural Language Processing. It is done based on the analysis of a defined set of genuine and fake news. It is done in order to make the world a better place.

INTRODUCTION:

There are millions of online platforms in this world. They will be both boon and bane for the human race. Anyone can easily spread a fake news in order to destroy the reputation of the person or the organization. Natural Language Processing is a part of the artificial intelligence which learns from the previous data. A variety of algorithms are available that include the supervised, unsupervised, reinforcement . The algorithms first have to be trained with a data set called train data set. After the training, these algorithms can be used to perform different tasks' is used in different sectors to perform different tasks.

Online platforms are helpful for the users because they can easily access a news. Many of the cybercriminals use the fake news as a weapon in order to make a war or demand for ransom. Readers read the news and start believing it without its verification. So once NLP is trained with a set of data then it becomes more vulnerable to detect fake news.

METHODOLOGY:

Various concepts involved in this project. Those are:

TOKENIZATION:

This process divides a large piece of continuous text into distinct units or tokens basically this process is often known as tokenization. NLTK provides the word tokenize() for splitting strings into tokens(nominally words).It splits tokens based on while space and punctuation.

STEMMING:

This is the idea of removing the suffix of a word and reducing different forms of a word to a core root. Some of the packages in stemmer are:

1)Snowball

2)Porter

3)Lancaster

Word	Porter	Lancaster	Lemmatiser
wrote	wrote	wrot	write
thinking	think	think	think
remembered	rememb	rememb	remember
relies	reli	rely	rely
ate	ate	at	eat
gone	gone	gon	go
won	won	won	win
ran	ran	ran	run
swimming	swim	swim	swim
mistreated	mistreat	mist	mistreat

STOPWORDS REMOVAL:

A stop word is a commonly used word that a search engine has been programmed to ignore. Typically, articles and pronouns are generally classified as stopwords. By removing these words, we remove the low-level information from our text in order to give more focus to the important information. The removal of stop words is highly dependent on the task we are performing and the goal we want to achieve.

VECTORIZATION:

The scikit-learn library offers easy-to-use tools to perform feature extraction of your text data is called as vectorization is a technique used to convert textual data to numerical format. Using vectorization, a matrix is created where each column represents a feature and each row represents an individual review.

Term Frequency is defined as how frequently the word appear in the document.

$$\mathrm{tf}(t,d) = \frac{f_{t,d}}{\text{number of words in d}} = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}}$$

TF-IDF stands for *term frequency-inverse document frequency* and it is a measure, used in the fields of information retrieval (IR) and machine learning, that can quantify the importance or relevance of string representations (words, phrases, lemmas, etc) in a document amongst a collection of documents (also known as a corpus).

W(x,y)=weight which signifies how important a word is for individual text message.

$$w_{x,y} = tf_{x,y} \times \log\left(\frac{N}{df_x}\right)$$

TF-IDF

Term x within document y

$tf_{x,y}$ = frequency of x in y

df_x = number of documents containing x

N = total number of documents

PASSIVE-AGGRESSIVE CLASSIFIER:

PASSIVE:

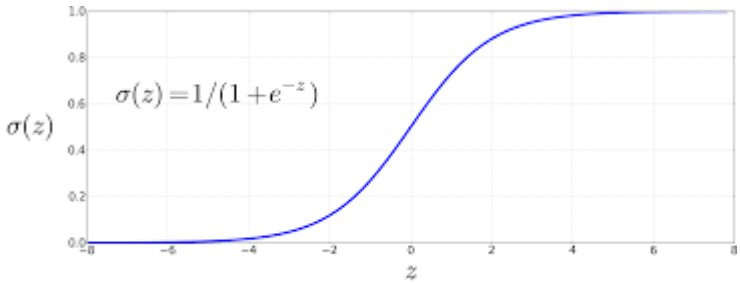
If the prediction is correct, keep the model and do not make any changes.i.e.,the data is not enough to cause any change in the model.

AGGRESSIVE:

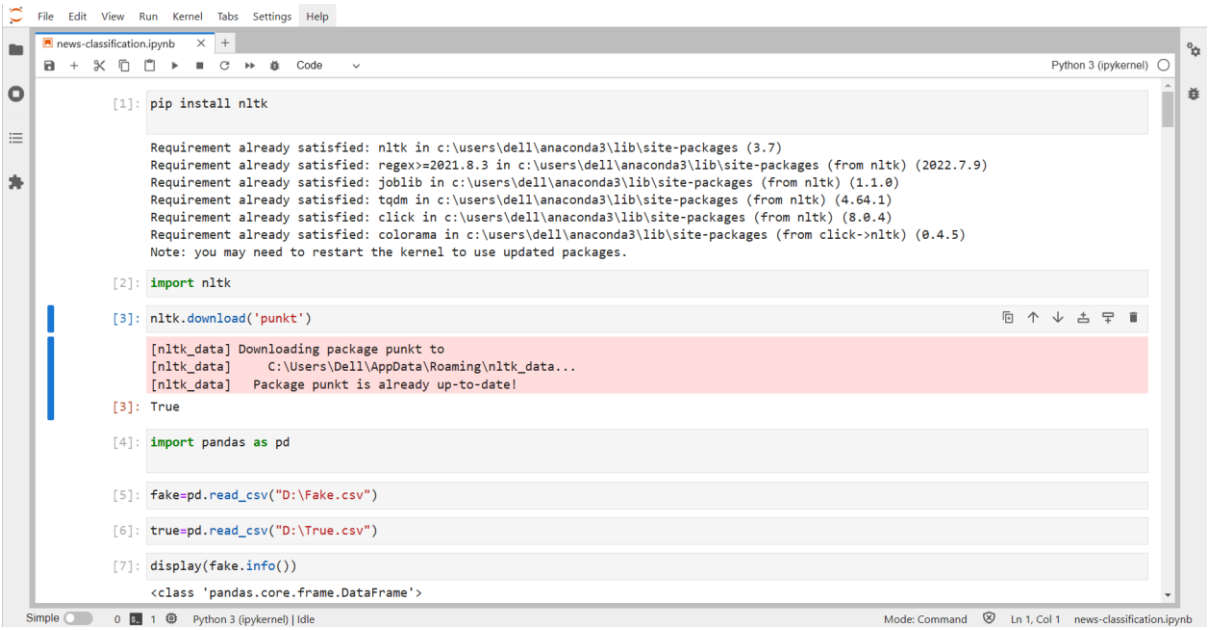
If the prediction is incorrect, make changes to the model.i.e.,some changes to the model may correct it.

LOGISTIC REGRESSION:

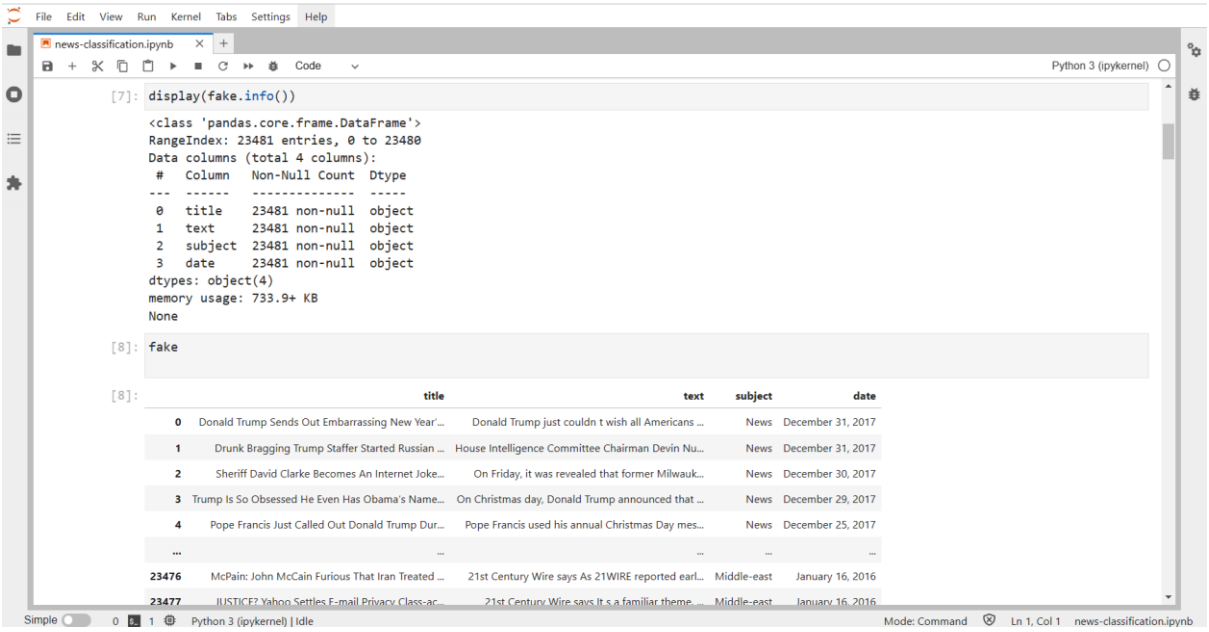
Logistic Regression is a machine learning algorithm which is used for the classification problems, it is a predictive analysis algorithm and based on the concept of probability.



CODE:



```
[1]: pip install nltk

Requirement already satisfied: nltk in c:\users\dell\anaconda3\lib\site-packages (3.7)
Requirement already satisfied: regex>=2021.8.3 in c:\users\dell\anaconda3\lib\site-packages (from nltk) (2022.7.9)
Requirement already satisfied: joblib in c:\users\dell\anaconda3\lib\site-packages (from nltk) (1.1.0)
Requirement already satisfied: tqdm in c:\users\dell\anaconda3\lib\site-packages (from nltk) (4.64.1)
Requirement already satisfied: click in c:\users\dell\anaconda3\lib\site-packages (from nltk) (8.0.4)
Requirement already satisfied: colorama in c:\users\dell\anaconda3\lib\site-packages (from click->nltk) (0.4.5)
Note: you may need to restart the kernel to use updated packages.

[2]: import nltk

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

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

[3]: True

[4]: import pandas as pd

[5]: fake=pd.read_csv("D:\Fake.csv")

[6]: true=pd.read_csv("D:\True.csv")

[7]: display(fake.info())

<class 'pandas.core.frame.DataFrame'>
```



```
[7]: display(fake.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

[8]: fake
```

[8]:

	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
...	...	...	...	...
23476	McPain: John McCain Furious That Iran Treated ...	21st Century Wire says As 21WIRE reported earl...	Middle-east	January 16, 2016
23477	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...	21st Century Wire says It s a familiar theme. ...	Middle-east	January 16, 2016

File Edit View Run Kernel Tabs Settings Help

news-classification.ipynb

Code

Python 3 (ipykernel)

[8]:
```
fake
```

[8]:

	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
...	...	...	...	...
23476	McPain: John McCain Furious That Iran Treated ...	21st Century Wire says As 21WIRE reported earl...	Middle-east	January 16, 2016
23477	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...	21st Century Wire says It s a familiar theme. ...	Middle-east	January 16, 2016
23478	Sunnistan: US and Allied 'Safe Zone' Plan to T...	Patrick Henningsen 21st Century WireRemember ...	Middle-east	January 15, 2016
23479	How to Blow $700 Million: Al Jazeera America F...	21st Century Wire says Al Jazeera America will...	Middle-east	January 14, 2016
23480	10 U.S. Navy Sailors Held by Iranian Military ...	21st Century Wire says As 21WIRE predicted in ...	Middle-east	January 12, 2016

23481 rows × 4 columns

[9]:
```
display(true.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21417 entries, 0 to 21416
Data columns (total 4 columns):
 #   Column   Non-Null Count  Dtype
```

Simple 0 1 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 news-classification.ipynb

File Edit View Run Kernel Tabs Settings Help

news-classification.ipynb

Code

Python 3 (ipykernel)

[9]:
```
display(true.info())
```

```
<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
```

[10]:
```
display(true.head(10))
```

	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
5	White House, Congress prepare for talks on spe...	WEST PALM BEACH, Fla./WASHINGTON (Reuters) - T...	politicsNews	December 29, 2017
6	Trump says Russia probe will be fair, but time...	WEST PALM BEACH, Fla (Reuters) - President Don...	politicsNews	December 29, 2017
7	Factbox: Trump on Twitter (Dec 29) - Approval ...	The following statements were posted to the ve...	politicsNews	December 29, 2017

Simple 0 1 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 news-classification.ipynb

File Edit View Run Kernel Tabs Settings Help

news-classification.ipynb

Code

Python 3 (ipykernel)

[10]:
```
display(true.head(10))
```

	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
5	White House, Congress prepare for talks on spe...	WEST PALM BEACH, Fla./WASHINGTON (Reuters) - T...	politicsNews	December 29, 2017
6	Trump says Russia probe will be fair, but time...	WEST PALM BEACH, Fla (Reuters) - President Don...	politicsNews	December 29, 2017
7	Factbox: Trump on Twitter (Dec 29) - Approval ...	The following statements were posted to the ve...	politicsNews	December 29, 2017
8	Trump on Twitter (Dec 28) - Global Warming	The following statements were posted to the ve...	politicsNews	December 29, 2017
9	Alabama official to certify Senator-elect Jone...	WASHINGTON (Reuters) - Alabama Secretary of St...	politicsNews	December 28, 2017

[11]:
```
display(fake.subject.value_counts())
```

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

[12]:
```
fake['target']=0
```

Simple 0 1 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 news-classification.ipynb

File Edit View Run Kernel Tabs Settings Help

news-classification.ipynb

Code

Python 3 (ipykernel)

[12]:
```
fake['target']=0
true['target']=1
```

[13]:
```
display(true.head(10))
```

	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
5	White House, Congress prepare for talks on spe...	WEST PALM BEACH, Fla./WASHINGTON (Reuters) - T...	politicsNews	December 29, 2017	1
6	Trump says Russia probe will be fair, but time...	WEST PALM BEACH, Fla (Reuters) - President Don...	politicsNews	December 29, 2017	1
7	Factbox: Trump on Twitter (Dec 29) - Approval ...	The following statements were posted to the ve...	politicsNews	December 29, 2017	1
8	Trump on Twitter (Dec 28) - Global Warming	The following statements were posted to the ve...	politicsNews	December 29, 2017	1
9	Alabama official to certify Senator-elect Jone...	WASHINGTON (Reuters) - Alabama Secretary of St...	politicsNews	December 28, 2017	1

[14]:
```
display(fake.head(10))
```

	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

Simple 0 1 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 news-classification.ipynb

File Edit View Run Kernel Tabs Settings Help

news-classification.ipynb

Code | Python 3 (ipykernel)

[14]:
```
display(fake.head(10))
```

	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
5	Racist Alabama Cops Brutalize Black Boy While...	The number of cases of cops brutalizing and ki...	News	December 25, 2017	0
6	Fresh Off The Golf Course, Trump Lashes Out A...	Donald Trump spent a good portion of his day a...	News	December 23, 2017	0
7	Trump Said Some INSANELY Racist Stuff Inside ...	In the wake of yet another court decision that...	News	December 23, 2017	0
8	Former CIA Director Slams Trump Over UN Bully...	Many people have raised the alarm regarding th...	News	December 22, 2017	0
9	WATCH: Brand-New Pro-Trump Ad Features So Muc...	Just when you might have thought we d get a br...	News	December 21, 2017	0

[15]:
```
data=pd.concat([fake,true],axis=0)
```

[16]:
```
date=data.reset_index(drop=True)
```

[17]:
```
data=data.drop(['subject','date','title'],axis=1)
```

[18]:
```
data
```

[18]:

	text	target

Simple 0 1 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 news-classification.ipynb

File Edit View Run Kernel Tabs Settings Help

news-classification.ipynb

Code | Python 3 (ipykernel)

[17]:
```
data=data.drop(['subject','date','title'],axis=1)
```

[18]:
```
data
```

[18]:

	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
...	...	...
21412	BRUSSELS (Reuters) - NATO allies on Tuesday we...	1
21413	LONDON (Reuters) - LexisNexis, a provider of l...	1
21414	MINSK (Reuters) - In the shadow of disused Sov...	1
21415	MOSCOW (Reuters) - Vatican Secretary of State ...	1
21416	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	1

44898 rows × 2 columns

[19]:
```
print(data.columns)
```

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

Simple 0 1 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 news-classification.ipynb

Tokenization

[20]:
```
from nltk.tokenize import word_tokenize
```

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

Stemming

[24]:
```
from nltk.stem.snowball import SnowballStemmer
porter=SnowballStemmer("english",ignore_stopwords=False)
```

[25]:
```
def stem_it(text):
    return[porter.stem(word) for word in text]
```

[26]:
```
data['text']=data['text'].apply(stem_it)
```

STOPWORD REMOVAL

[27]:
```
from nltk.corpus import stopwords
nltk.download('stopwords')
print(stopwords.words('english'))
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your', 'yours', 'yoursel
f', 'yourselves', 'he', 'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its', 'itself', 'they', 'the
m', 'their', 'theirs', 'themselves', 'what', 'which', 'who', 'whom', 'this', 'that', "that'll", 'these', 'those', 'am', 'is', 'are', 'w
as', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if',
'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through', 'during',
```


STOPWORD REMOVAL

[27]:
```
from nltk.corpus import stopwords
nltk.download('stopwords')
print(stopwords.words('english'))
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your', 'yours', 'yoursel
f', 'yourselves', 'he', 'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its', 'itself', 'they', 'the
m', 'their', 'theirs', 'themselves', 'what', 'which', 'who', 'whom', 'this', 'that', "that'll", 'these', 'those', 'am', 'is', 'are', 'w
as', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if',
'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through', 'during',
'before', 'after', 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'off', 'over', 'under', 'again', 'further', 'then',
'once', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both', 'each', 'few', 'more', 'most', 'other', 'some', 'such',
'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'than', 'too', 'very', 's', 't', 'can', 'will', 'just', 'don', "don't", 'should', "sho
uld've", 'now', 'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren', "aren't", 'couldn', "couldn't", 'didn', "didn't", 'doesn', "does
n't", 'hadn', "hadn't", 'hasn', "hasn't", 'haven', "haven't", 'isn', "isn't", 'ma', 'mightn', "mightn't", 'mustn', "mustn't", 'needn',
"needn't", 'shan', "shan't", 'shouldn', "shouldn't", 'wasn', "wasn't", 'weren', "weren't", 'won', "won't", 'wouldn', "wouldn't"]
```


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

[28]:
```
def stop_it(t):
    dt=[word for word in t if len(word)>2]
    return dt
```

[29]:
```
data['text']=data['text'].apply(stop_it)
```

[30]:
```
data['text']=data['text'].apply(' '.join)
```

[31]:
```
data
```

```
    return dt
```

[29]:
```
data['text']=data['text'].apply(stop_it)
```

[30]:
```
data['text']=data['text'].apply(' '.join)
```

[31]:
```
data
```

[31]:

	text	target
0	donald trump just couldn wish all american hap...	0
1	hous intellig committe chairman devin nune hav...	0
2	friday was reveal that former milwauke sheriff...	0
3	christma day donald trump announc that would b...	0
4	pope franci use his annual christma day messag...	0
...	...	...
21412	brussel reuter nato alli tuesday welcom presid...	1
21413	london reuter lexisnexi provid legal regulator...	1
21414	minsk reuter the shadow disus soviet-era facto...	1
21415	moscow reuter vatican secretari state cardin p...	1
21416	jakarta reuter indonesia will buy sukhoi fight...	1

44898 rows × 2 columns

Splitting

[32]:
```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(data['text'],data['target'],test_size=0.25)
display(X_train.head())
print('\n')
display(y_train.head())
```

```
3216     case you are unawar becaus donald trump the pr...
23215    shawn helton 21st centuri wireyesterday wave d...
22105    tune the altern current radio network acr for ...
9903     st. loui former st. loui polic offic jason sto...
16139    texa feder judg appoint obama has again reject...
Name: text, dtype: object


3216     0
23215    0
22105    0
9903     0
16139    0
Name: target, dtype: int64
```

Vectorization

[33]:
```
from sklearn.feature_extraction.text import TfidfVectorizer
my_tfidf=TfidfVectorizer(max_df=0.7)

tfidf_train=my_tfidf.fit_transform(X_train)
```

Vectorization

[33]:
```python
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)
```

[35]:
```python
print(tfidf_train)
```

```
  (0, 73147)    0.035820446257037966
  (0, 60505)    0.04412895366071528
  (0, 73065)    0.03189011963026803
  (0, 34721)    0.022625363117862146
  (0, 86753)    0.01562949248406545
  (0, 40999)    0.03322920101378302
  (0, 31375)    0.017520129772859785
  (0, 82719)    0.043800019486398854
  (0, 82375)    0.02083819270424446
  (0, 69539)    0.022516458431202407
  (0, 55639)    0.011767674758056191
  (0, 30377)    0.019847548065475992
  (0, 85589)    0.02180043207989916
  (0, 77025)    0.03420384097173025
  (0, 71336)    0.022341637636171537
  (0, 79238)    0.027454833744283296
  (0, 30389)    0.04040146077496972
  (0, 77636)    0.011407654134414023
  (0, 42831)    0.01715892566089102
  (0, 25962)    0.026416906301292297
```

LogisticRegression

[36]:
```python
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
```

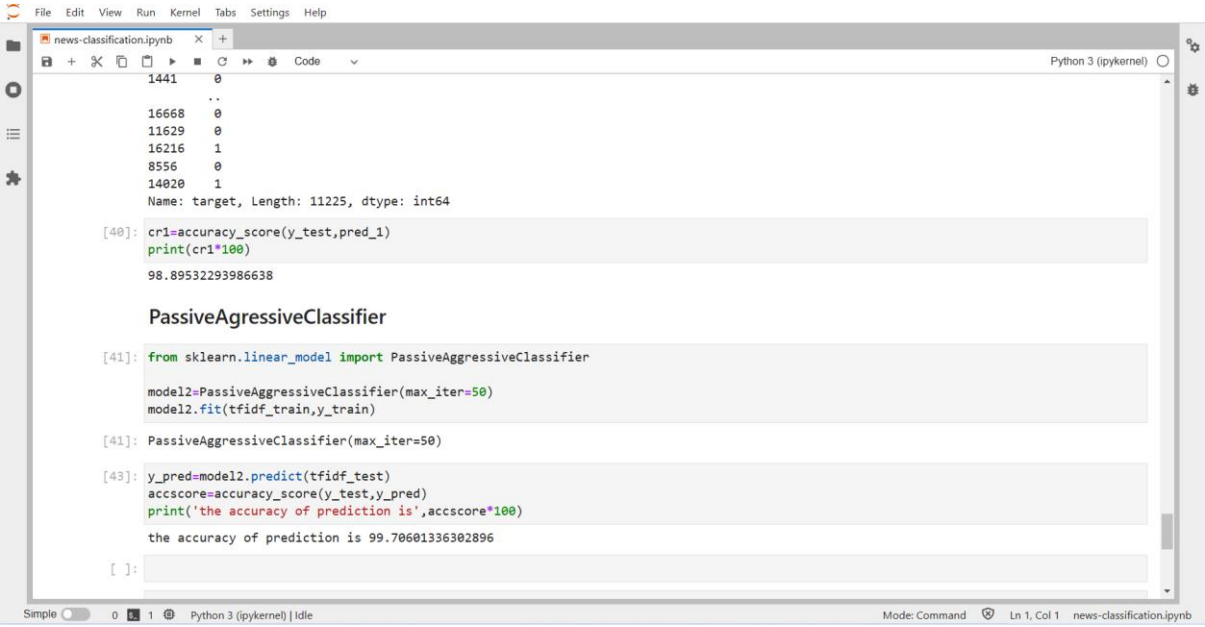
[37]:
```python
model_1=LogisticRegression(max_iter=900)
model_1.fit(tfidf_train,y_train)
pred_1=model_1.predict(tfidf_test)
```

[38]:
```python
pred_1
```

[38]:
```
array([1, 0, 1, ..., 1, 0, 1], dtype=int64)
```

[39]:
```python
y_test
```

[39]:
```
18718    1
16404    0
7385     0
8044     0
1441     0
        ..
16668    0
11629    0
16216    1
8556     0
14020    1
Name: target, Length: 11225, dtype: int64
```



```
1441     0
         ..
16668    0
11629    0
16216    1
8556     0
14020    1
Name: target, Length: 11225, dtype: int64
```

[40]:
```
cr1=accuracy_score(y_test,pred_1)
print(cr1*100)
```

```
98.89532293986638
```

PassiveAgressiveClassifier

[41]:
```
from sklearn.linear_model import PassiveAggressiveClassifier

model2=PassiveAggressiveClassifier(max_iter=50)
model2.fit(tfidf_train,y_train)
```

[41]:
```
PassiveAggressiveClassifier(max_iter=50)
```

[43]:
```
y_pred=model2.predict(tfidf_test)
accscore=accuracy_score(y_test,y_pred)
print('the accuracy of prediction is',accscore*100)
```

```
the accuracy of prediction is 99.70601336302896
```

CONCLUSION:

Due to increasing use of internet, it is now easy to spread fake news. A huge number of persons are regularly connected with internet and social media platforms. There is no any restriction while posting any news on these platforms. So, some of the people takes the advantage of these platforms and start spreading fake news against the individuals or organizations. This can destroy the repute of an individual or can affect a business. Through fake news, the opinions of the people can also be changed for a political party. There is a need for a way to detect these fake news. Natural Language Processing is used for different purposes and these can also be used for detecting the fake news. The classifiers are first trained with a data set called training data set. After that, these classifiers can automatically detect fake news.

The supervised machine learning classifiers are discussed that requires the labelled data for training. Labelled data is not easily available that can be used for training the classifiers for detecting the fake news. In future research can be on the use of the unsupervised machine learning classifiers for the detection of fake news.