In [ ]:
```python
pip install sklearn.feature_extraction
```

In [1]:
```python
import numpy as np
import pandas as pd
import re
from nltk.stem.porter import PorterStemmer
from nltk.corpus import stopwords
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
```

In [ ]:
```python
import nltk
nltk.download("stopwords")
```

In [3]:
```python
print(stopwords.words("english"))
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you'r
e", "you've", "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves',
'he', 'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'i
t', "it's", 'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselves',
'what', 'which', 'who', 'whom', 'this', 'that', "that'll", 'these', 'those',
'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'ha
d', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but',
'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'wit
h', 'about', 'against', 'between', 'into', 'through', 'during', 'before', 'af
ter', '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', 'th
an', 'too', 'very', 's', 't', 'can', 'will', 'just', 'don', "don't", 'shoul
d', "should've", 'now', 'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren',
"aren't", 'couldn', "couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn',
"hadn't", 'hasn', "hasn't", 'haven', "haven't", 'isn', "isn't", 'ma', 'might
n', "mightn't", 'mustn', "mustn't", 'needn', "needn't", 'shan', "shan't", 'sh
ouldn', "shouldn't", 'wasn', "wasn't", 'weren', "weren't", 'won', "won't", 'w
ouldn', "wouldn't"]
```

In [2]:
```python
news_datasets=pd.read_csv("News.csv")
```

In [3]: `news_datasets.head()`

Out[3]:

	id	title	author	text	label
0	0	House Dem Aide: We Didn't Even See Comey's Let...	Darrell Lucus	House Dem Aide: We Didn't Even See Comey's Let...	1
1	1	FLYNN: Hillary Clinton, Big Woman on Campus - ...	Daniel J. Flynn	Ever get the feeling your life circles the rou...	0
2	2	Why the Truth Might Get You Fired	Consortiumnews.com	Why the Truth Might Get You Fired October 29, ...	1
3	3	15 Civilians Killed In Single US Airstrike Hav...	Jessica Purkiss	Videos 15 Civilians Killed In Single US Airstr...	1
4	4	Iranian woman jailed for fictional unpublished...	Howard Portnoy	Print \nAn Iranian woman has been sentenced to...	1

In [4]: `news_datasets.shape`

Out[4]: `(20800, 5)`

In [5]: `news_datasets.describe()`

Out[5]:

	id	label
count	20800.000000	20800.000000
mean	10399.500000	0.500625
std	6004.587135	0.500012
min	0.000000	0.000000
25%	5199.750000	0.000000
50%	10399.500000	1.000000
75%	15599.250000	1.000000
max	20799.000000	1.000000

In [6]: `news_datasets.isnull().sum()`

Out[6]:
```
id          0
title     558
author   1957
text       39
label       0
dtype: int64
```

In [ ]:

In [7]: `news_datasets=news_datasets.fillna("")`

In [8]: `news_datasets['content']=news_datasets['author']+' '+news_datasets['text']`

In [9]: `news_datasets.head()`

Out[9]:

	id	title	author	text	label	content
0	0	House Dem Aide: We Didn't Even See Comey's Let...	Darrell Lucus	House Dem Aide: We Didn't Even See Comey's Let...	1	Darrell Lucus House Dem Aide: We Didn't Even S...
1	1	FLYNN: Hillary Clinton, Big Woman on Campus - ...	Daniel J. Flynn	Ever get the feeling your life circles the rou...	0	Daniel J. Flynn Ever get the feeling your life...
2	2	Why the Truth Might Get You Fired	Consortiumnews.com	Why the Truth Might Get You Fired October 29, ...	1	Consortiumnews.com Why the Truth Might Get You...
3	3	15 Civilians Killed In Single US Airstrike Hav...	Jessica Purkiss	Videos 15 Civilians Killed In Single US Airstr...	1	Jessica Purkiss Videos 15 Civilians Killed In ...
4	4	Iranian woman jailed for fictional unpublished...	Howard Portnoy	Print \nAn Iranian woman has been sentenced to...	1	Howard Portnoy Print \nAn Iranian woman has be...

In [10]: `x=news_datasets.drop(columns="label",axis=1)`

In [11]: `y=news_datasets["label"]`

In [12]:
```python
print(x)
print(y)
```

```
              id                                              title  \
0              0   House Dem Aide: We Didn't Even See Comey's Let...
1              1   FLYNN: Hillary Clinton, Big Woman on Campus - ...
2              2                    Why the Truth Might Get You Fired
3              3   15 Civilians Killed In Single US Airstrike Hav...
4              4   Iranian woman jailed for fictional unpublished...
...          ...                                                ...
20795      20795   Rapper T.I.: Trump a 'Poster Child For White S...
20796      20796   N.F.L. Playoffs: Schedule, Matchups and Odds -...
20797      20797   Macy's Is Said to Receive Takeover Approach by...
20798      20798   NATO, Russia To Hold Parallel Exercises In Bal...
20799      20799                            What Keeps the F-35 Alive

                                      author  \
0                              Darrell Lucus
1                            Daniel J. Flynn
2                          Consortiumnews.com
3                             Jessica Purkiss
4                              Howard Portnoy
...                                      ...
20795                           Jerome Hudson
20796                         Benjamin Hoffman
20797   Michael J. de la Merced and Rachel Abrams
20798                              Alex Ansary
20799                            David Swanson

                                               text  \
0       House Dem Aide: We Didn't Even See Comey's Let...
1       Ever get the feeling your life circles the rou...
2       Why the Truth Might Get You Fired October 29, ...
3       Videos 15 Civilians Killed In Single US Airstr...
4       Print \nAn Iranian woman has been sentenced to...
...                                               ...
20795   Rapper T. I. unloaded on black celebrities who...
20796   When the Green Bay Packers lost to the Washing...
20797   The Macy's of today grew from the union of sev...
20798   NATO, Russia To Hold Parallel Exercises In Bal...
20799     David Swanson is an author, activist, journa...

                                            content
0       Darrell Lucus House Dem Aide: We Didn't Even S...
1       Daniel J. Flynn Ever get the feeling your life...
2       Consortiumnews.com Why the Truth Might Get You...
3       Jessica Purkiss Videos 15 Civilians Killed In ...
4       Howard Portnoy Print \nAn Iranian woman has be...
...                                               ...
20795   Jerome Hudson Rapper T. I. unloaded on black c...
20796   Benjamin Hoffman When the Green Bay Packers lo...
20797   Michael J. de la Merced and Rachel Abrams The ...
20798   Alex Ansary NATO, Russia To Hold Parallel Exer...
20799   David Swanson   David Swanson is an author, ac...

[20800 rows x 5 columns]
0        1
1        0
2        1
3        1
```

```
4        1
         ..
20795    0
20796    0
20797    0
20798    1
20799    1
Name: label, Length: 20800, dtype: int64
```

In [13]:
```python
port_stem=PorterStemmer()
```

In [16]:
```python
def stemming(content):
    stemmed_content = re.sub('[^a-zA-Z]',' ',content)
    stemmed_content = stemmed_content.lower()
    stemmed_content = stemmed_content.split()
    stemmed_content = [port_stem.stem(word) for word in stemmed_content if not
    stemmed_content = ' '.join(stemmed_content)
    return stemmed_content
```

In [17]:
```python
print(news_datasets['content'])
```

```
0        Darrell Lucus House Dem Aide: We Didn't Even S...
1        Daniel J. Flynn Ever get the feeling your life...
2        Consortiumnews.com Why the Truth Might Get You...
3        Jessica Purkiss Videos 15 Civilians Killed In ...
4        Howard Portnoy Print \nAn Iranian woman has be...
                               ...
20795    Jerome Hudson Rapper T. I. unloaded on black c...
20796    Benjamin Hoffman When the Green Bay Packers lo...
20797    Michael J. de la Merced and Rachel Abrams The ...
20798    Alex Ansary NATO, Russia To Hold Parallel Exer...
20799    David Swanson   David Swanson is an author, ac...
Name: content, Length: 20800, dtype: object
```

In [18]:
```python
news_datasets['content']=news_datasets['content'].apply(stemming)
```

In [20]:
```python
print(news_datasets['content'])
```

```
0        darrel lucu hous dem aid even see comey letter...
1        daniel j flynn ever get feel life circl rounda...
2        consortiumnew com truth might get fire octob t...
3        jessica purkiss video civilian kill singl us a...
4        howard portnoy print iranian woman sentenc six...
                               ...
20795    jerom hudson rapper unload black celebr met do...
20796    benjamin hoffman green bay packer lost washing...
20797    michael j de la merc rachel abram maci today g...
20798    alex ansari nato russia hold parallel exercis ...
20799    david swanson david swanson author activist jo...
Name: content, Length: 20800, dtype: object
```

```
In [22]: x=news_datasets['content']
         y=news_datasets['label']
```

```
In [24]: vectorizer=TfidfVectorizer()
         vectorizer.fit(x)
         x=vectorizer.transform(x)
```

```
In [25]: print(x)
```

```
  (0, 109513)    0.05158710720776892
  (0, 109458)    0.020048695304140447
  (0, 108506)    0.04642287258829905
  (0, 108502)    0.0996559768298958
  (0, 108459)    0.03968664675416755
  (0, 108422)    0.011881983627763996
  (0, 107771)    0.018072875845931625
  (0, 106960)    0.017961156636243633
  (0, 106869)    0.0131648126444554
  (0, 106783)    0.030659734472833485
  (0, 106706)    0.01351546681491241
  (0, 106506)    0.012384173554702569
  (0, 105662)    0.027193531070699047
  (0, 105626)    0.032931097434687535
  (0, 104622)    0.022626272566892605
  (0, 103211)    0.06867906246194783
  (0, 102528)    0.03481803135793534
  (0, 102278)    0.017213348547683853
  (0, 101511)    0.03995296784256272
  (0, 100071)    0.01164076607705450
```

```
In [26]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,stratify=y,ra
```

```
In [27]: model=LogisticRegression()
         model.fit(x_train,y_train)
```

Out[27]: LogisticRegression()

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

```
In [28]: x_train_prediction=model.predict(x_train)
         train_data_accuracy=accuracy_score(x_train_prediction,y_train)
         print(train_data_accuracy)
```

```
0.9764423076923077
```

In [29]:
```python
x_test_prediction=model.predict(x_test)
test_data_accuracy=accuracy_score(x_test_prediction,y_test)
print(test_data_accuracy)
```

```
0.9454326923076923
```

In [30]:
```python
X_new = x_test[3]

prediction = model.predict(X_new)
print(prediction)

if (prediction[0]==0):
  print('The news is Real')
else:
  print('The news is Fake')
```

```
[1]
The news is Fake
```

In [35]: `print(x_test[3])`

```
  (0, 109458)     0.014322351073167887
  (0, 108642)     0.04848194891115942
  (0, 108527)     0.0070974457965219334
  (0, 108512)     0.028103714038680803
  (0, 108506)     0.016581744321717766
  (0, 108459)     0.009450425228960732
  (0, 108407)     0.012561928017893001
  (0, 108023)     0.02646876212239631
  (0, 107497)     0.015665419481082535
  (0, 107438)     0.013195154399467286
  (0, 107204)     0.025148260523565606
  (0, 106669)     0.015353538659602841
  (0, 106165)     0.02521978809460325
  (0, 105040)     0.027379566857922626
  (0, 104813)     0.012444476051077371
  (0, 104622)     0.016163716105421503
  (0, 104140)     0.034538501400213896
  (0, 103869)     0.03363764176514086
  (0, 103563)     0.015215396175071463
  (0, 103520)     0.021671364387764968
  (0, 103115)     0.008252483937015498
  (0, 103055)     0.027595204043219752
  (0, 102898)     0.01640649455323048
  (0, 102696)     0.030588592193605633
  (0, 102278)     0.024593682262587763
  :              :
  (0, 4893)       0.029641422689912312
  (0, 4867)       0.020276390630901177
  (0, 4865)       0.01684231290836362
  (0, 4577)       0.026554284966746437
  (0, 3823)       0.018676289097030455
  (0, 3655)       0.017892423762581957
  (0, 3266)       0.029510335588658655
  (0, 3037)       0.014262325254364546
  (0, 2974)       0.012825310684037469
  (0, 2895)       0.03202007609549731
  (0, 2777)       0.016057641586074943
  (0, 2655)       0.02183531561417725
  (0, 2601)       0.02762176648700157
  (0, 2258)       0.06695214117570211
  (0, 2149)       0.026931840185753626
  (0, 1932)       0.03957576514406355
  (0, 921)        0.012301938291419574
  (0, 902)        0.01636776636150318
  (0, 895)        0.013383454510221727
  (0, 886)        0.02519447701771436
  (0, 866)        0.012462953090469676
  (0, 604)        0.014566016237848058
  (0, 599)        0.043369382110471824
  (0, 559)        0.23035620667528905
  (0, 490)        0.023060140685291858
```