

```
In [34]: import numpy as np
import re
import pandas as pd
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
```

```
Requirement already satisfied: gensim in c:\users\dell\.conda\acc\lib\site-packages (4.3.0)
Requirement already satisfied: numpy>=1.18.5 in c:\users\dell\.conda\acc\lib\site-packages (from gensim) (1.23.5)
Requirement already satisfied: scipy>=1.7.0 in c:\users\dell\.conda\acc\lib\site-packages (from gensim) (1.10.0)
Requirement already satisfied: FuzzyTM>=0.4.0 in c:\users\dell\.conda\acc\lib\site-packages (from gensim) (2.0.9)
Requirement already satisfied: smart-open>=1.8.1 in c:\users\dell\.conda\acc\lib\site-packages (from gensim) (5.2.1)
Requirement already satisfied: pyfume in c:\users\dell\.conda\acc\lib\site-packages (from FuzzyTM>=0.4.0->gensim) (0.3.1)
Requirement already satisfied: pandas in c:\users\dell\.conda\acc\lib\site-packages (from FuzzyTM>=0.4.0->gensim) (1.5.3)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\dell\.conda\acc\lib\site-packages (from pandas->FuzzyTM>=0.4.0->gensim) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\dell\.conda\acc\lib\site-packages (from pandas->FuzzyTM>=0.4.0->gensim) (2022.7)
Requirement already satisfied: typing-extensions in c:\users\dell\.conda\acc\lib\site-packages (from pyfume->FuzzyTM>=0.4.0->gensim) (4.4.0)
Requirement already satisfied: simpful in c:\users\dell\.conda\acc\lib\site-packages (from pyfume->FuzzyTM>=0.4.0->gensim) (2.12.0)
Requirement already satisfied: fst-pso in c:\users\dell\.conda\acc\lib\site-packages (from pyfume->FuzzyTM>=0.4.0->gensim) (1.8.1)
Requirement already satisfied: six>=1.5 in c:\users\dell\.conda\acc\lib\site-packages (from python-dateutil>=2.8.1->pandas->FuzzyTM>=0.4.0->gensim) (1.16.0)
Requirement already satisfied: miniful in c:\users\dell\.conda\acc\lib\site-packages (from fst-pso->pyfume->FuzzyTM>=0.4.0->gensim) (0.0.6)
Requirement already satisfied: lightgbm in c:\users\dell\.conda\acc\lib\site-packages (4.3.0)
Requirement already satisfied: numpy in c:\users\dell\.conda\acc\lib\site-packages (from lightgbm) (1.23.5)
Requirement already satisfied: scipy in c:\users\dell\.conda\acc\lib\site-packages (from lightgbm) (1.10.0)
```

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

```
In [35]: !rm *.*
!ls
```

```
'rm' is not recognized as an internal or external command,
operable program or batch file.
'ls' is not recognized as an internal or external command,
operable program or batch file.
```

```
In [36]: df = pd.read_csv('train[1].csv')
```

In [37]: df.head()

Out[37]:

	id	title	author	text	label
0	0	House Dem Aide: We Didn't Even See Comey's Let...	Darrell Lucas	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 Aistr...	1
4	4	Iranian woman jailed for fictional unpublished...	Howard Portnoy	Print \nAn Iranian woman has been sentenced to...	1

In [38]: df.isnull().sum()

Out[38]:

```
id          0
title      558
author    1957
text        39
label        0
dtype: int64
```

In [39]: df.shape

Out[39]: (20800, 5)

In [40]: df = df.fillna(' ')

In [41]: df.isnull().sum()

Out[41]:

```
id          0
title        0
author        0
text          0
label         0
dtype: int64
```

In [42]: df['content'] = df['author']+' '+df['title']

In [43]: df

Out[43]:

	id	title	author	text	label	content
0	0	House Dem Aide: We Didn't Even See Comey's Let...	Darrell Lucas	House Dem Aide: We Didn't Even See Comey's Let...	1	Darrell Lucas 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 FLYNN: Hillary Clinton, Big Wo...
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 Aistr...	1	Jessica Purkiss 15 Civilians Killed In Single ...
4	4	Iranian woman jailed for fictional unpublished...	Howard Portnoy	Print \nAn Iranian woman has been sentenced to...	1	Howard Portnoy Iranian woman jailed for fictio...
...
20795	20795	Rapper T.I.: Trump a 'Poster Child For White S...	Jerome Hudson	Rapper T. I. unloaded on black celebrities who...	0	Jerome Hudson Rapper T.I.: Trump a 'Poster Chi...
20796	20796	N.F.L. Playoffs: Schedule, Matchups and Odds -...	Benjamin Hoffman	When the Green Bay Packers lost to the Washing...	0	Benjamin Hoffman N.F.L. Playoffs: Schedule, Ma...
20797	20797	Macy's Is Said to Receive Takeover Approach by...	Michael J. de la Merced and Rachel Abrams	The Macy's of today grew from the union of sev...	0	Michael J. de la Merced and Rachel Abrams Macy...
20798	20798	NATO, Russia To Hold Parallel Exercises In Bal...	Alex Ansary	NATO, Russia To Hold Parallel Exercises In Bal...	1	Alex Ansary NATO, Russia To Hold Parallel Exer...
20799	20799	What Keeps the F-35 Alive	David Swanson	David Swanson is an author, activist, journa...	1	David Swanson What Keeps the F-35 Alive

```
In [44]: X = df.drop('label',axis=1)
y = df['label']
```

```
In [45]: print(X)
```

	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 Lucas
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 Aistr...
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 Lucas House Dem Aide: We Didn't Even S...
1	Daniel J. Flynn FLYNN: Hillary Clinton, Big Wo...
2	Consortiumnews.com Why the Truth Might Get You...
3	Jessica Purkiss 15 Civilians Killed In Single ...
4	Howard Portnoy Iranian woman jailed for fictio...
...	...
20795	Jerome Hudson Rapper T.I.: Trump a 'Poster Chi...
20796	Benjamin Hoffman N.F.L. Playoffs: Schedule, Ma...
20797	Michael J. de la Merced and Rachel Abrams Macy...
20798	Alex Ansary NATO, Russia To Hold Parallel Exer...
20799	David Swanson What Keeps the F-35 Alive

[20800 rows x 5 columns]

```
In [46]: ps = PorterStemmer()
def stemming(content):
    stemmed_content = re.sub('[^a-zA-Z]', ' ', content)
    stemmed_content = stemmed_content.lower()
    stemmed_content = stemmed_content.split()
    stemmed_content = [ps.stem(word) for word in stemmed_content if not word in stopwords]
    stemmed_content = ' '.join(stemmed_content)
    return stemmed_content
```

```
In [47]: df['content'] = df['content'].apply(stemming)
```

```
In [48]: df['content']
```

```
Out[48]: 0      darrel lucu hous dem aid even see come letter...
1      daniel j flynn flynn hillari clinton big woman...
2      consortiumnew com truth might get fire
3      jessica purkiss civilian kill singl us airstri...
4      howard portnoy iranian woman jail fiction unpu...
...
20795   jerom hudson rapper trump poster child white s...
20796   benjamin hoffman n f l playoff schedul matchup...
20797   michael j de la merc rachel abram maci said re...
20798   alex ansari nato russia hold parallel exercis ...
20799   david swanson keep f aliv
Name: content, Length: 20800, dtype: object
```

```
In [49]: X = df['content'].values
y = df['label'].values
```

```
In [50]: vector = TfidfVectorizer()
vector.fit(X)
X = vector.transform(X)
```

```
In [51]: print(X)
```

```

(0, 15686)    0.28485063562728646
(0, 13473)    0.2565896679337957
(0, 8909)     0.3635963806326075
(0, 8630)     0.29212514087043684
(0, 7692)     0.24785219520671603
(0, 7005)     0.21874169089359144
(0, 4973)     0.233316966909351
(0, 3792)     0.2705332480845492
(0, 3600)     0.3598939188262559
(0, 2959)     0.2468450128533713
(0, 2483)     0.3676519686797209
(0, 267)      0.27010124977708766
(1, 16799)    0.30071745655510157
(1, 6816)     0.1904660198296849
(1, 5503)     0.7143299355715573
(1, 3568)     0.26373768806048464
(1, 2813)     0.19094574062359204
(1, 2223)     0.3827320386859759
(1, 1894)     0.15521974226349364
(1, 1497)     0.2939891562094648
(2, 15611)    0.41544962664721613
(2, 9620)     0.49351492943649944
(2, 5968)     0.3474613386728292
(2, 5389)     0.3866530551182615
(2, 3103)     0.46097489583229645
:             :
(20797, 13122) 0.2482526352197606
(20797, 12344) 0.27263457663336677
(20797, 12138) 0.24778257724396507
(20797, 10306) 0.08038079000566466
(20797, 9588)  0.174553480255222
(20797, 9518)  0.2954204003420313
(20797, 8988)  0.36160868928090795
(20797, 8364)  0.22322585870464118
(20797, 7042)  0.21799048897828688
(20797, 3643)  0.21155500613623743
(20797, 1287)  0.33538056804139865
(20797, 699)   0.30685846079762347
(20797, 43)    0.29710241860700626
(20798, 13046) 0.22363267488270608
(20798, 11052) 0.4460515589182236
(20798, 10177) 0.3192496370187028
(20798, 6889)  0.32496285694299426
(20798, 5032)  0.4083701450239529
(20798, 1125)  0.4460515589182236
(20798, 588)   0.3112141524638974
(20798, 350)   0.28446937819072576
(20799, 14852) 0.5677577267055112
(20799, 8036)  0.45983893273780013
(20799, 3623)  0.37927626273066584
(20799, 377)   0.5677577267055112

```

```

In [52]: from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, Y_test = train_test_split(X, y, test_size = 0.2, stratify=y,

```

```

In [53]: X_train.shape

```

```

Out[53]: (16640, 17128)

```

```
In [54]: model = LogisticRegression()  
         model.fit(X_train,Y_train)
```

```
Out[54]: ▾ LogisticRegression  
         LogisticRegression()
```

```
In [55]: train_y_pred = model.predict(X_train)  
         print(accuracy_score(train_y_pred,Y_train))  
  
0.9865985576923076
```

```
In [56]: testing_y_pred = model.predict(X_test)  
         print(accuracy_score(testing_y_pred,Y_test))  
  
0.9790865384615385
```

```
In [57]: input_data = X_test[10]  
         prediction = model.predict(input_data)
```

```
In [58]: if prediction[0] == 0:  
         print('The News Is Real')  
         else:  
         print('The News is Fake')  
  
The News Is Real
```

```
In [59]: df['content'][2]
```

```
Out[59]: 'consortiumnew com truth might get fire'
```

```
In [ ]:
```