

```
In [1]: import numpy as np #Array
import pandas as pd
import re #Pattern
from nltk.corpus import stopwords #To remove words like:the for of in with
from nltk.stem.porter import PorterStemmer # To bring words in base form like
from sklearn.feature_extraction.text import TfidfVectorizer #if machine do not
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression #For Classification
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
```

C:\Users\Akash Chavan\anaconda3\lib\site-packages\scipy__init__.py:138: UserWarning: A NumPy version >=1.16.5 and <1.23.0 is required for this version of SciPy (detected version 1.24.4)
 warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion} is required for this version of ")

```
In [2]: news_df = pd.read_csv('train.csv')
```

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

Out[3]:

	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4	4	Iranian woman jailed for fictional unpublished...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 [4]: # Replace these two arrays with your actual predictions and true labels
# Predictions (0 for non-fake, 1 for fake)
predictions = np.array([1, 0, 1, 0, 1, 1, 0, 1, 0, 0])
# True labels (0 for non-fake, 1 for fake)
true_labels = np.array([1, 0, 1, 0, 1, 0, 0, 1, 0, 1])

# Calculate the confusion matrix
cm = confusion_matrix(true_labels, predictions)

# Print the confusion matrix
print("Confusion Matrix:")
print(cm)
```

Confusion Matrix:
 [[4 1]
 [1 4]]

```
In [5]: news_df.shape
```

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

```
In [6]: news_df = news_df.fillna(' ')
```

```
In [7]: news_df.isna().sum()
```

```
Out[7]: id          0  
        title       0  
        author      0  
        text        0  
        label       0  
        dtype: int64
```

```
In [8]: news_df['content'] = news_df['author']+" "+news_df['title']
```

In [9]: news_df

Out[9]:

	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

20800 rows × 6 columns

```
In [10]: news_df['content']
```

```
Out[10]: 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
Name: content, Length: 20800, dtype: object
```

```
In [11]: news_df['content'][20796]
```

```
Out[11]: 'Benjamin Hoffman N.F.L. Playoffs: Schedule, Matchups and Odds - The New York Times'
```

```
In [12]: news_df['content']
```

```
Out[12]: 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
Name: content, Length: 20800, dtype: object
```

```
In [13]: X = news_df['content'].values
y = news_df['label'].values
```

```
In [14]: print(X)
```

```
['Darrell Lucas House Dem Aide: We Didn't Even See Comey's Letter Until Jason Chaffetz Tweeted It'
'Daniel J. Flynn FLYNN: Hillary Clinton, Big Woman on Campus - Breitbart'
'Consortiumnews.com Why the Truth Might Get You Fired' ...
'Michael J. de la Merced and Rachel Abrams Macy's Is Said to Receive Takeover Approach by Hudson's Bay - The New York Times'
'Alex Ansary NATO, Russia To Hold Parallel Exercises In Balkans'
'David Swanson What Keeps the F-35 Alive']
```

```
In [15]: vector = TfidfVectorizer()
vector.fit(X)
X = vector.transform(X)
```

```
In [16]: print(X)
```

```
(0, 23355)    0.18006497451107856
(0, 22649)    0.26575278886038384
(0, 22289)    0.3484071341454308
(0, 19171)    0.22537992364975484
(0, 12902)    0.3024224900242886
(0, 12528)    0.24883399099107747
(0, 11409)    0.20615188166061463
(0, 11307)    0.1532265401605094
(0, 10387)    0.1844880289323935
(0, 7574)     0.23047267305353566
(0, 6145)     0.24677171892553343
(0, 5800)     0.2502787762405247
(0, 5508)     0.2993429551929777
(0, 4507)     0.20531415441295317
(0, 3862)     0.30579573877221844
(0, 904)      0.26354004814013343
(1, 23748)    0.2966210296019264
(1, 15149)    0.15862263711495958
(1, 10134)    0.18787145765749733
(1, 8430)     0.20450000000000000
```

```
In [17]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, strati
```

```
In [18]: X_train.shape
```

```
Out[18]: (16640, 25173)
```

```
In [19]: X_test.shape
```

```
Out[19]: (4160, 25173)
```

```
In [20]: model = LogisticRegression()
model.fit(X_train, y_train)
```

```
Out[20]: LogisticRegression()
```

```
In [21]: train_y_pred = model.predict(X_train)
print("train accuracy :", accuracy_score(train_y_pred, y_train))
```

```
train accuracy : 0.9872596153846154
```

```
In [22]: test_y_pred = model.predict(X_test)
print("train accuracy :", accuracy_score(test_y_pred, y_test))
```

```
train accuracy : 0.975
```

```
In [31]: news_df['content'][100]
```

```
Out[31]: 'Activist Post Technocracy: The Real Reason Why The UN Wants Control Over The Internet'
```

```
In [32]: # prediction system

input_data = X_test[100]
prediction = model.predict(input_data)
if prediction[0] == 1:
    print('Fake news')
else:
    print('Real news')
```

Real news

```
In [ ]:
```