

# Fake News Detection

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [2]: data = pd.read_csv("fake_news.csv")
data.head()
```

Out[2]:

	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4	4	Iranian woman jailed for fictional unpublished...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 [3]: data.shape
```

Out[3]: (20800, 5)

```
In [4]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20800 entries, 0 to 20799
Data columns (total 5 columns):
#   Column  Non-Null Count  Dtype
---  -
0    id      20800 non-null    int64
1   title   20242 non-null    object
2   author  18843 non-null    object
3   text    20761 non-null    object
4   label   20800 non-null    int64
dtypes: int64(2), object(3)
memory usage: 812.6+ KB
```

```
In [5]: data.isna().sum()
```

```
Out[5]: id      0
title    558
author   1957
text     39
label     0
dtype: int64
```

```
In [6]: data = data.drop(['id'], axis=1)
```

```
In [7]: data = data.fillna('')
```

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

```
In [9]: data = data.drop(['title', 'author', 'text'], axis=1)
```

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

Out[10]:

	label	content
0	1	Darrell Lucas House Dem Aide: We Didn't Even S...
1	0	Daniel J. Flynn FLYNN: Hillary Clinton, Big Wo...
2	1	Consortiumnews.com Why the Truth Might Get You...
3	1	Jessica Purkiss 15 Civilians Killed In Single ...
4	1	Howard Portnoy Iranian woman jailed for fictio...

```
In [11]: data['content'] = data['content'].apply(lambda x: " ".join(x.lower() for x in x.split()))
```

```
In [12]: data['content'] = data['content'].str.replace('[^\w\s]','')
```

C:\Users\Pooja Reddy\AppData\Local\Temp\ipykernel\_24476\3643324700.py:1: FutureWarning: The default value of regex will change from True to False in a future version.

```
data['content'] = data['content'].str.replace('[^\w\s]','')
```

```
In [13]: import nltk
nltk.download('stopwords')
```

[nltk\_data] Downloading package stopwords to C:\Users\Pooja Reddy\AppData\Roaming\nltk\_data...  
[nltk\_data] Package stopwords is already up-to-date!

Out[13]: True

```
In [14]: import nltk
nltk.download('wordnet')
```

[nltk\_data] Downloading package wordnet to C:\Users\Pooja Reddy\AppData\Roaming\nltk\_data...  
[nltk\_data] Package wordnet is already up-to-date!

Out[14]: True

```
In [15]: from nltk.corpus import stopwords
stop = stopwords.words('english')
data['content'] = data['content'].apply(lambda x: " ".join(x for x in x.split() if x not in stop))
```

```
In [16]: from nltk.stem import WordNetLemmatizer
from textblob import Word
data['content'] = data['content'].apply(lambda x: " ".join([Word(word).lemm
data['content'].head()
```

```
Out[16]: 0    darrell lucus house dem aide didnt even see co...
1    daniel j flynn flynn hillary clinton big woman...
2    consortiumnewscom truth might get fired truth ...
3    jessica purkiss 15 civilian killed single u ai...
4    howard portnoy iranian woman jailed fictional ...
Name: content, dtype: object
```

```
In [17]: X = data[['content']]
y = data['label']
```

```
In [18]: from sklearn.model_selection import train_test_split
```

```
In [19]: X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.3, random_st
```

```
In [20]: print (X_train.shape)
print (y_train.shape)
print (X_test.shape)
print (y_test.shape)
```

```
(14560, 1)
(14560,)
(6240, 1)
(6240,)
```

```
In [22]: from sklearn.feature_extraction.text import TfidfVectorizer
```

```
In [23]: tfidf_vect = TfidfVectorizer(analyzer='word', token_pattern=r'\w{1,}', max_
tfidf_vect.fit(data['content'])
```

```
xtrain_tfidf = tfidf_vect.transform(X_train['content'])
xtest_tfidf = tfidf_vect.transform(X_test['content'])
```

```
In [25]: from sklearn.linear_model import PassiveAggressiveClassifier
from sklearn import metrics
pclf = PassiveAggressiveClassifier()
pclf.fit(xtrain_tfidf, y_train)
predictions = pclf.predict(xtest_tfidf)
print(metrics.classification_report(y_test, predictions))
```

	precision	recall	f1-score	support
0	0.96	0.96	0.96	3116
1	0.96	0.96	0.96	3124
accuracy			0.96	6240
macro avg	0.96	0.96	0.96	6240
weighted avg	0.96	0.96	0.96	6240

In [27]: `print(metrics.confusion_matrix(y_test,predictions))`

```
[[2977  139]
 [ 118 3006]]
```

In [30]: `from sklearn.neural_network import MLPClassifier`  
`mlpclf = MLPClassifier(hidden_layer_sizes=(256,64,16),`  
 `activation = 'relu',`  
 `solver = 'adam')`  
`mlpclf.fit(xtrain_tfidf, y_train)`  
`predictions = mlpclf.predict(xtest_tfidf)`  
`print(metrics.classification_report(y_test, predictions))`

	precision	recall	f1-score	support
0	0.96	0.95	0.95	3116
1	0.95	0.96	0.95	3124
accuracy			0.95	6240
macro avg	0.95	0.95	0.95	6240
weighted avg	0.95	0.95	0.95	6240

In [31]: `print(metrics.confusion_matrix(y_test,predictions))`

```
[[2969  147]
 [ 135 2989]]
```

In [32]: `import pickle`  
`pickle.dump(mlpclf, open("fakenews1.pkl", "wb"))`

In [ ]: