

# 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(r"C:\Users\Vaish\Desktop\NLP(AD)\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4	4	Iranian woman jailed for fictional unpublished...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...

## DATA PREPROCESSING

```
In [12]: data['content'] = data['content'].apply(lambda x: " ".join(x.lower() for x in x.split()))
```

```
In [13]: data['content'] = data['content'].str.replace('[^\w\s]','')
```

```
<>:1: SyntaxWarning: invalid escape sequence '\w'
<>:1: SyntaxWarning: invalid escape sequence '\w'
C:\Users\Vaish\AppData\Local\Temp\ipykernel_9388\3643324700.py:1: SyntaxWarning: invalid escape sequence '\w'
data['content'] = data['content'].str.replace('[^\w\s]','')
```

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

```
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\Vaish\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords 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
```

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

data['content'].head()
```

```
Out[16]: 0    darrell lucus house dem aide: didn't even see ...
1    daniel j. flynn flynn: hillary clinton, big wo...
2    consortiumnews.com 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]: # splitting into training and testing data

X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.3, random_state=45,
```

```
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 [21]: from sklearn.feature_extraction.text import TfidfVectorizer
```

```
In [22]: tfidf_vect = TfidfVectorizer(analyzer='word', token_pattern=r'\w{1,}', max_features

tfidf_vect.fit(data['content'])

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

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

## MODEL BUILDING

```
In [24]: #passive Aggressive Classifier:supervised Learning for Large data
```

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

```
[[2998  118]
 [ 119 3005]]
```

## MLP CLASSIFIER

```
In [28]: 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	0.96	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 [29]: print(metrics.confusion_matrix(y_test, predictions))
```

```
[[3000  116]
 [ 125 2999]]
```

```
In [30]: import pickle
```

```
# Save trained model to file
```

```
pickle.dump(mlpclf, open("fakenews1.pkl", "wb"))
```

```
In [79]: def fake_news_det(news):  
        input_data = [news]  
        vectorized_input_data = tfidf_vect.transform(input_data)  
        prediction = pclf.predict(vectorized_input_data)  
        print(prediction)
```

```
In [81]: fake_news_det('U.S. Secretary of State John F. Kerry said Monday that he will stop  
[1]
```

```
In [83]: fake_news_det(""" President Barack Obama has been campaigning hard for the  
        woman who is supposedly going to extend his legacy four more years.  
        The only problem with stumping for Hillary Clinton, however, is she is not  
        exactly a candidate easy to get too enthused about. """)  
[1]
```

```
In [86]: !pip install Flask
```

```
Requirement already satisfied: Flask in c:\users\vaish\anaconda3\lib\site-packages  
(3.0.3)  
Requirement already satisfied: Werkzeug>=3.0.0 in c:\users\vaish\anaconda3\lib\site-  
packages (from Flask) (3.0.3)  
Requirement already satisfied: Jinja2>=3.1.2 in c:\users\vaish\anaconda3\lib\site-pa  
ckages (from Flask) (3.1.4)  
Requirement already satisfied: itsdangerous>=2.1.2 in c:\users\vaish\anaconda3\lib\s  
ite-packages (from Flask) (2.2.0)  
Requirement already satisfied: click>=8.1.3 in c:\users\vaish\anaconda3\lib\site-pac  
kages (from Flask) (8.1.7)  
Requirement already satisfied: blinker>=1.6.2 in c:\users\vaish\anaconda3\lib\site-p  
ackages (from Flask) (1.6.2)  
Requirement already satisfied: colorama in c:\users\vaish\anaconda3\lib\site-package  
s (from click>=8.1.3->Flask) (0.4.6)  
Requirement already satisfied: MarkupSafe>=2.0 in c:\users\vaish\anaconda3\lib\site-  
packages (from Jinja2>=3.1.2->Flask) (2.1.3)
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

