

# FAKE NEWS DETECTION

```
In [9]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [11]: data = pd.read_csv(r"C:\Users\Vaish\Desktop\NLP(AD)\fake_news.csv")
data.head()
```

```
Out[11]:
```

	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4	4	Iranian woman jailed for fictional unpublished...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 [15]: data.shape
```

```
Out[15]: (20800, 5)
```

```
In [19]: 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 [21]: data.isna().sum()
```

```
Out[21]: id          0
         title      558
         author     1957
         text       39
         label      0
         dtype: int64
```

```
In [23]: data = data.drop(['id'],axis=1)
```

```
In [25]: data=data.fillna('')
```

```
In [27]: data['content'] = data['author']+' '+data['title']+' '+data['text']
```

```
In [31]: data = data.drop(['title','author','text'],axis=1)
```

```
In [33]: data.head()
```

```
Out[33]:
```

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

```
In [38]: 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_11076\3643324700.py:1: SyntaxWarning: in
valid escape sequence '\w'
    data['content'] = data['content'].str.replace('[^\w\s]','')
```

```
In [40]: 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[40]: True
```

```
In [42]: 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 [46]: from nltk.stem import WordNetLemmatizer
from textblob import Word
data['content'] = data['content'].apply(lambda x: " ".join([Word(word).lemmatize()

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

```
Out[46]: 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 [48]: X = data[['content']]
y = data['label']
```

```
In [50]: from sklearn.model_selection import train_test_split
```

```
In [54]: # 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 [56]: print(X_train.shape)
print(y_train.shape)
print(X_test.shape)
print(y_test.shape)
```

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

```
In [58]: from sklearn.feature_extraction.text import TfidfVectorizer
```

```
In [ ]: 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'])
```

```
In [ ]:
```

```
In [ ]:
```