Fake News Detection Using NLP

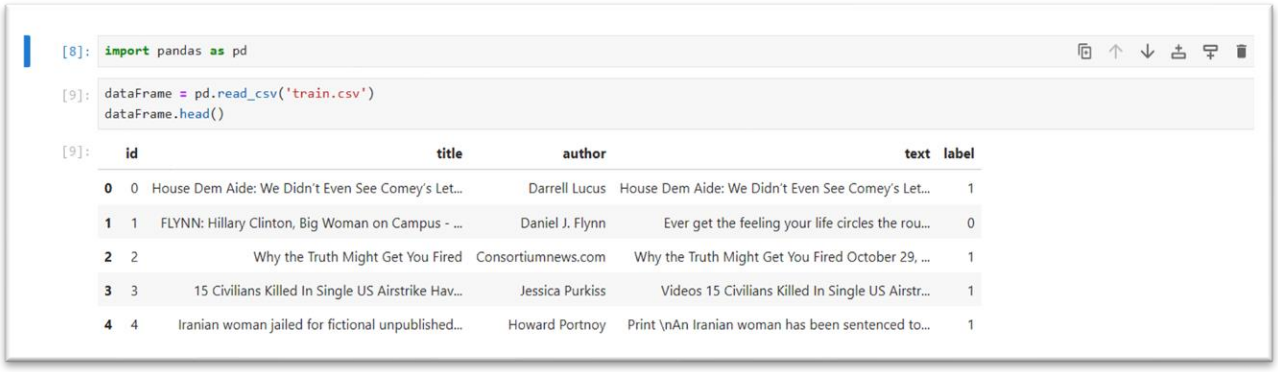
Phase – 3

We started building our project by loading the dataset, performing text preprocessing, and conducting analysis in Jupyter Notebook.

Loading the Dataset:

Downloaded the **train.csv** dataset from the Kaggle

We loaded a train.csv dataset using the **pandas** library.



```
[8]: import pandas as pd

[9]: dataFrame = pd.read_csv('train.csv')
     dataFrame.head()
```

[9]:

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4	4	Iranian woman jailed for fictional unpublished...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

Text Preprocessing:

Replacing null values in the dataset with empty string

```
[5]: dataFrame.isnull().sum()

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

[6]: #to remove the null values with an empty string
     dataFrame = dataFrame.fillna('')

[7]: dataFrame.isnull().sum()

[7]: id        0
     title     0
     author    0
     text      0
     label     0
```

Removing Unnecessary Columns in the dataFrame

```
[8]: #remove unnecessary columns
     dataFrame.drop(['id','title','author'],axis=1)
```

[8]:

	text	label
0	House Dem Aide: We Didn't Even See Comey's Let...	1
1	Ever get the feeling your life circles the rou...	0
2	Why the Truth Might Get You Fired October 29, ...	1
3	Videos 15 Civilians Killed In Single US Airstr...	1
4	Print \nAn Iranian woman has been sentenced to...	1
...	...	...

Using **Natural Language ToolKit** Library for Text Preprocessing

This code processes the 'text' column in the DataFrame by applying text preprocessing operations like stemming, lowercase conversion, and stopword removal to each text entry in that column.

```
[9]: import nltk
     nltk.download('stopwords')
     from nltk.corpus import stopwords
     from nltk.stem.porter import PorterStemmer

     import re
```

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

```
[10]: portStem = PorterStemmer()
```

```
[11]: def stemming(content):
          con = re.sub('[^a-zA-Z]',' ',content)
          con = con.lower()
          con = con.split()
          con = [portStem.stem(word) for word in con if not word in stopwords.words('english')]
          con = ''.join(con)
          return con
```

```
[*]: dataFrame['text'] = dataFrame['text'].apply(stemming)
```

Team:

MOHAMED ABDUL RAHMAN S (au951221104023)

YABESH JESILEN V (au951221104059)

SAM SELIN D (au951221104039)

RAHUL B (au951221104034)

NAGA MARI MUTHU A (au951221104028)