```python
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
import re
import string
```

```python
import pandas as pd
df_fake = pd.read_csv("/content/fake_news_dataset 2.csv")
df_true = pd.read_csv("/content/fake_news_dataset 2.csv")
```

```python
df_fake.head()
```

| | title | text | date | source | author | category | label |
|---|---|---|---|---|---|---|---|
| 0 | Foreign Democrat final. | more tax development both store agreement lawy... | 2023-03-10 | NY Times | Paula George | Politics | real |
| 1 | To offer down resource great point. | probably guess western behind likely next inve... | 2022-05-25 | Fox News | Joseph Hill | Politics | fake |
| 2 | Himself church myself carry. | them identify forward present success risk sev... | 2022-09-01 | CNN | Julia Robinson | Business | fake |
| | | phone which item yard Republican safe where | 2023-02- | | Mr. David Foster | | |

```python
df_true.head(5)
```

| | title | text | date | source | author | category | label |
|---|---|---|---|---|---|---|---|
| 0 | Foreign Democrat final. | more tax development both store agreement lawy... | 2023-03-10 | NY Times | Paula George | Politics | real |
| 1 | To offer down resource great point. | probably guess western behind likely next inve... | 2022-05-25 | Fox News | Joseph Hill | Politics | fake |
| 2 | Himself church myself carry. | them identify forward present success risk sev... | 2022-09-01 | CNN | Julia Robinson | Business | fake |
| | | phone which item yard Republican safe where | 2023-02- | | Mr. David Foster | | |

```python
df_fake["class"] = 0
df_true["class"] = 1
df_fake.shape, df_true.shape
```

```
((20000, 8), (20000, 8))
```

```python
df_fake_manual_testing = df_fake.tail(10)
# Get the list of index labels to drop
indexes_to_drop = df_fake.tail(10).index.tolist()
# Drop the rows using the index labels
df_fake.drop(indexes_to_drop, axis=0, inplace=True)


df_true_manual_testing = df_true.tail(10)
# Get the list of index labels to drop
indexes_to_drop_true = df_true.tail(10).index.tolist()
# Drop the rows using the index labels
df_true.drop(indexes_to_drop_true, axis=0, inplace=True)


df_fake.shape, df_true.shape
```

```
((19990, 8), (19990, 8))
```

```python
df_fake_manual_testing["class"] = 0
df_true_manual_testing["class"] = 1
```

```
<ipython-input-13-3aaf8ec2aad1>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus
  df_fake_manual_testing["class"] = 0
<ipython-input-13-3aaf8ec2aad1>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus
  df_true_manual_testing["class"] = 1
```

```
df_fake_manual_testing.head(10)
```

| | title | text | date | source | author | category | label | class |
|---|---|---|---|---|---|---|---|---|
| **19990** | End item later outside. | quickly away item while goal company keep walk... | 2022-09-21 | The Guardian | Ian Perez | Politics | fake | 0 |
| **19991** | Beat generation individual describe. | size professor visit property worry fight six ... | 2024-09-17 | CNN | Norma Martin | Health | fake | 0 |
| **19992** | Rule participant catch her. | no sport decision dark level rate tax site qua... | 2023-12-03 | Global Times | Travis Jacobs | Business | real | 0 |
| **19993** | Money himself challenge once along. | new approach wide way provide western necessar... | 2024-10-21 | CNN | Alex Higgins | Politics | real | 0 |
| **19994** | Blue believe enjoy north return. | people think law represent arm apply governmen... | 2023-04-08 | Reuters | Sara Shaw | Health | real | 0 |
| **19995** | House party born. | hit and television I change very our happy doo... | 2024-12-04 | BBC | Gary Miles | Entertainment | fake | 0 |
| **19996** | Though nation people maybe price box. | fear most meet rock even sea value design stan... | 2024-05-26 | Daily News | Maria Mcbride | Entertainment | real | 0 |
| **19997** | Yet exist with experience unit. | activity loss very provide eye west create wha... | 2023-04-17 | BBC | Kristen Franklin | Entertainment | real | 0 |
| **19998** | School wide itself item. | term point general common training watch respo... | 2024-06-30 | Reuters | David Wise | Health | fake | 0 |
| **19999** | Offer chair cover senior born. | remain pressure glass me six senior though nor... | 2024-01-24 | Daily News | James Peterson | Health | fake | 0 |

```
df_true_manual_testing.head(10)
```

| | title | text | date | source | author | category | label | class |
|---|---|---|---|---|---|---|---|---|
| **19990** | End item later outside. | quickly away item while goal company keep walk... | 2022-09-21 | The Guardian | Ian Perez | Politics | fake | 1 |
| **19991** | Beat generation individual describe. | size professor visit property worry fight six ... | 2024-09-17 | CNN | Norma Martin | Health | fake | 1 |
| **19992** | Rule participant catch her. | no sport decision dark level rate tax site qua... | 2023-12-03 | Global Times | Travis Jacobs | Business | real | 1 |
| **19993** | Money himself challenge once along. | new approach wide way provide western necessar... | 2024-10-21 | CNN | Alex Higgins | Politics | real | 1 |
| **19994** | Blue believe enjoy north return. | people think law represent arm apply governmen... | 2023-04-08 | Reuters | Sara Shaw | Health | real | 1 |
| **19995** | House party born. | hit and television I change very our happy doo... | 2024-12-04 | BBC | Gary Miles | Entertainment | fake | 1 |
| **19996** | Though nation people maybe price box. | fear most meet rock even sea value design stan... | 2024-05-26 | Daily News | Maria Mcbride | Entertainment | real | 1 |
| **19997** | Yet exist with experience unit. | activity loss very provide eye west create wha... | 2023-04-17 | BBC | Kristen Franklin | Entertainment | real | 1 |
| **19998** | School wide itself item. | term point general common training watch respo... | 2024-06-30 | Reuters | David Wise | Health | fake | 1 |
| **19999** | Offer chair cover senior born. | remain pressure glass me six senior though nor... | 2024-01-24 | Daily News | James Peterson | Health | fake | 1 |

```
df_manual_testing = pd.concat([df_fake_manual_testing,df_true_manual_testing], axis = 0)
df_manual_testing.to_csv("manual_testing.csv")


df_merge = pd.concat([df_fake, df_true], axis =0 )
df_merge.head(10)
```

| | title | text | date | source | author | category | label | class |
|---|---|---|---|---|---|---|---|---|
| 0 | Foreign Democrat final. | more tax development both store agreement lawy... | 2023-03-10 | NY Times | Paula George | Politics | real | 0 |
| 1 | To offer down resource great point. | probably guess western behind likely next inve... | 2022-05-25 | Fox News | Joseph Hill | Politics | fake | 0 |
| 2 | Himself church myself carry. | them identify forward present success risk sev... | 2022-09-01 | CNN | Julia Robinson | Business | fake | 0 |
| 3 | You unit its should. | phone which item yard Republican safe where po... | 2023-02-07 | Reuters | Mr. David Foster DDS | Science | fake | 0 |
| 4 | Billion believe employee summer how. | wonder myself fact difficult course forget exa... | 2023-04-03 | CNN | Austin Walker | Technology | fake | 0 |
| 5 | Method purpose mission approach professor short. | affect too bill whether kind project turn offi... | 2023-12-02 | Reuters | Sherri Fry | Health | real | 0 |
| 6 | Laugh member step. | often along newspaper establish fall president... | 2022-06-01 | CNN | Alyssa Young | Business | fake | 0 |
| 7 | Center measure difference dark. | ready movement bed increase during or history ... | 2024-12-31 | NY Times | Tina Garrett | Sports | fake | 0 |
| 8 | Moment make those affect first difference. | officer mention dream fill later foot suffer d... | 2022-09-08 | Fox News | Heather Greene | Entertainment | fake | 0 |
| 9 | Reason physical contain total decision. | choose anything treat beyond political minute ... | 2022-11-08 | Daily News | Erin Hanson | Health | real | 0 |

```python
def wordopt(text):
    text = text.lower()
    text = re.sub('\[.*?\]', '', text)
    text = re.sub("\\W"," ",text)
    text = re.sub('https?://\S+|www\.\S+', '', text)
    text = re.sub('<.*?>+', '', text)
    text = re.sub('[%s]' % re.escape(string.punctuation), '', text)
    text = re.sub('\n', '', text)
    text = re.sub('\w*\d\w*', '', text)
    return text
```

```python
x = df_merge["text"] # Changed df to df_merge
y = df_merge["class"] # Changed df to df_merge


from sklearn.model_selection import train_test_split # importing train_test_split
import pandas as pd

# ... (your other code) ...

x = df_merge["text"] # Changed df to df_merge
y = df_merge["class"] # Changed df to df_merge

# Now you can use train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25)


from sklearn.feature_extraction.text import TfidfVectorizer

vectorization = TfidfVectorizer()
xv_train = vectorization.fit_transform(x_train)
xv_test = vectorization.transform(x_test)


from sklearn.linear_model import LogisticRegression

LR = LogisticRegression()
LR.fit(xv_train,y_train)
```

```
▾ LogisticRegression ⓘ �ⓘ
LogisticRegression()
```

```python
pred_lr=LR.predict(xv_test)
LR.score(xv_test, y_test)
```

```
0.39459729864932463
```

```python
from sklearn.metrics import classification_report # Importing the classification_report function

# ... your other code ...
```

```
pred_lr = LR.predict(xv_test)
LR.score(xv_test, y_test)

print(classification_report(y_test, pred_lr)) # Now classification_report should be recognized
```



```
               precision    recall  f1-score   support

           0       0.40      0.36      0.38      5100
           1       0.39      0.43      0.41      4895

    accuracy                           0.39      9995
   macro avg       0.39      0.40      0.39      9995
weighted avg       0.39      0.39      0.39      9995
```

```python
def manual_testing(news):
    testing_news = {"text":[news]}
    new_def_test = pd.DataFrame(testing_news)
    new_def_test["text"] = new_def_test["text"].apply(wordopt)
    new_x_test = new_def_test["text"]
    new_xv_test = vectorization.transform(new_x_test)
    pred_LR = LR.predict(new_xv_test)
    pred_DT = DT.predict(new_xv_test)
    pred_GBC = GBC.predict(new_xv_test)
def manual_testing(news):
    testing_news = {"text":[news]}
    new_def_test = pd.DataFrame(testing_news)
    new_def_test["text"] = new_def_test["text"].apply(wordopt)
    new_x_test = new_def_test["text"]
    new_xv_test = vectorization.transform(new_x_test)
    pred_LR = LR.predict(new_xv_test)
    pred_DT = DT.predict(new_xv_test)
    pred_GBC = GBC.predict(new_xv_test)
    pred_RFC = RFC.predict(new_xv_test)

    return print("\n\nLR Prediction: {} \nDT Prediction: {} \nGBC Prediction: {} \nRFC Prediction: {}".format(output_lable(pred_LR[0]),
                                                                                   output_lable(pred_GBC[0]),
                                                                                   output_lable(pred_RFC[0])

    testing_news = {"text":[news]}
    new_def_test = pd.DataFrame(testing_news)
    new_def_test["text"] = new_def_test["text"].apply(wordopt)
    new_x_test = new_def_test["text"]
    new_xv_test = vectorization.transform(new_x_test)
    pred_LR = LR.predict(new_xv_test)
    pred_DT = DT.predict(new_xv_test)
    pred_GBC = GBC.predict(new_xv_test)
    pred_RFC = RFC.predict(new_xv_test)

    return print("\n\nLR Prediction: {} \nDT Prediction: {} \nGBC Prediction: {} \nRFC Prediction: {}".format(output_lable(pred_LR[0]),
                                                                                   output_lable(pred_GBC[0]),
                                                                                   output_lable(pred_RFC[0])


def manual_testing(news):
    import pandas as pd # Importing pandas inside the function
    testing_news = {"text":[news]}
    new_def_test = pd.DataFrame(testing_news)
    new_def_test["text"] = new_def_test["text"].apply(wordopt)
    new_x_test = new_def_test["text"]
    new_xv_test = vectorization.transform(new_x_test)
    pred_LR = LR.predict(new_xv_test)
    pred_DT = DT.predict(new_xv_test)
    pred_GBC = GBC.predict(new_xv_test)
    pred_RFC = RFC.predict(new_xv_test)

    return print("\n\nLR Prediction: {} \nDT Prediction: {} \nGBC Prediction: {} \nRFC Prediction: {}".format(output_lable(pred_LR[0]),
                                                                                   output_lable(pred_GBC[0]),
                                                                                   output_lable(pred_RFC[0])
```



```
e function

ual_testing
```

```
ext)

tuation), '', text)
```

```
ly(wordopt)
```

```
:)
```

```
:ion: {} \nGBC Prediction: {} \nRFC Prediction: {}".format(output_lable(pred_LR[0]),
                                                    output_lable(pred_GBC[0]),
                                                    output_lable(pred_RFC[0])))
```