

✚ IMPORT LIBRARIES

```
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
```

✚ IMPORT DATASET

```
df_fake = pd.read_csv("/content/Fake.csv")
df_true = pd.read_csv("/content/True.csv")
```

```
df_fake.head()
```

| | title | text | subject | date | |
|---|--|---|---------|-------------------|--|
| 0 | Donald Trump Sends Out Embarrassing New Year'... | Donald Trump just couldn't wish all Americans ... | News | December 31, 2017 | |
| 1 | Drunk Bragging Trump Staffer Started Russian ... | House Intelligence Committee Chairman Devin Nu... | News | December 31, 2017 | |
| 2 | Sheriff David Clarke Becomes An Internet Joke... | On Friday, it was revealed that former Milwauk... | News | December 30, 2017 | |

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

| | title | text | subject | date | |
|---|---|---|--------------|-------------------|--|
| 0 | As U.S. budget fight looms, Republicans flip t... | WASHINGTON (Reuters) - The head of a conservat... | politicsNews | December 31, 2017 | |
| 1 | U.S. military to accept transgender recruits o... | WASHINGTON (Reuters) - Transgender people will... | politicsNews | December 29, 2017 | |
| 2 | Senior U.S. Republican senator: 'Let Mr. Muell... | WASHINGTON (Reuters) - The special counsel inv... | politicsNews | December 31, 2017 | |

✚ Inserting a column "class" as target feature

```
df_fake["class"] = 0
df_true["class"] = 1
```

```
# Removing last 10 rows for manual testing
df_fake_manual_testing = df_fake.tail(10)
for i in range(23480, 23470, -1):
    df_fake.drop([i], axis = 0, inplace = True)
```

```
df_true_manual_testing = df_true.tail(10)
for i in range(21416, 21406, -1):
    df_true.drop([i], axis = 0, inplace = True)
```

```
df_fake.shape, df_true.shape
```

```
((23471, 5), (21407, 5))
```

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

```
<ipython-input-10-3aaf8ec2aad1>: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-10-3aaf8ec2aad1>: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-a-copy

```
df_true_manual_testing["class"] = 1
```

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

| | title | text | subject | date | class | |
|-------|---|---|-------------|------------------|-------|-----|
| 23471 | Seven Iranians freed in the prisoner swap have... | 21st Century Wire says This week, the historic... | Middle-east | January 20, 2016 | 0 | |
| 23472 | #Hashtag Hell & The Fake Left | By Dady Chery and Gilbert MercierAll writers ... | Middle-east | January 19, 2016 | 0 | |
| 23473 | Astroturfing: Journalist Reveals Brainwashing ... | Vic Bishop Waking TimesOur reality is carefull... | Middle-east | January 19, 2016 | 0 | |
| 23474 | The New American Century: An Era of Fraud | Paul Craig RobertsIn the last years of the 20t... | Middle-east | January 19, 2016 | 0 | |
| 23475 | Hillary Clinton: 'Israel First' (and no peace ... | Robert Fantina CounterpunchAlthough the United... | Middle-east | January 18, 2016 | 0 | |
| ... | ... | ... | ... | ... | ... | ... |

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

| | title | text | subject | date | class | |
|-------|---|---|-----------|-----------------|-------|-----|
| 21407 | Mata Pires, owner of embattled Brazil builder ... | SAO PAULO (Reuters) - Cesar Mata Pires, the ow... | worldnews | August 22, 2017 | 1 | |
| 21408 | U.S., North Korea clash at U.N. forum over nuc... | GENEVA (Reuters) - North Korea and the United ... | worldnews | August 22, 2017 | 1 | |
| 21409 | U.S., North Korea clash at U.N. arms forum on ... | GENEVA (Reuters) - North Korea and the United ... | worldnews | August 22, 2017 | 1 | |
| 21410 | Headless torso could belong to submarine journ... | COPENHAGEN (Reuters) - Danish police said on T... | worldnews | August 22, 2017 | 1 | |
| 21411 | North Korea shipments to Syria chemical arms a... | UNITED NATIONS (Reuters) - Two North Korean sh... | worldnews | August 21, 2017 | 1 | |
| ... | ... | ... | ... | ... | ... | ... |

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

✓ Merging True and Fake Dataframes

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

| | title | text | subject | date | class | |
|-----|--|---|---------|-------------------|-------|-----|
| 0 | Donald Trump Sends Out Embarrassing New Year'... | Donald Trump just couldn't wish all Americans ... | News | December 31, 2017 | 0 | |
| 1 | Drunk Bragging Trump Staffer Started Russian ... | House Intelligence Committee Chairman Devin Nu... | News | December 31, 2017 | 0 | |
| 2 | Sheriff David Clarke Becomes An Internet Joke... | On Friday, it was revealed that former Milwauk... | News | December 30, 2017 | 0 | |
| 3 | Trump Is So Obsessed He Even Has Obama's Name... | On Christmas day, Donald Trump announced that ... | News | December 29, 2017 | 0 | |
| 4 | Pope Francis Just Called Out Donald Trump Dur... | Pope Francis used his annual Christmas Day mes... | News | December 25, 2017 | 0 | |
| ... | ... | ... | ... | ... | ... | ... |

```
df_merge.columns

Index(['title', 'text', 'subject', 'date', 'class'], dtype='object')
```

✓ Removing columns which are not required

```
df = df_merge.drop(["title", "subject", "date"], axis = 1)
```



```
df.isnull().sum()
```

```
text      0
class     0
dtype: int64
```

✓ Random Shuffling the dataframe

```
df = df.sample(frac = 1)
```

```
df.head()
```



| | text | class |  |
|-------|---|-------|---|
| 16082 | BELFAST (Reuters) - The British government on ... | 1 |  |
| 16970 | VALLETTA (Reuters) - Police believe a bomb tha... | 1 | |
| 13954 | Obama and HUD want to give one last freebie to... | 0 | |
| 10952 | The French citizens just voted for their demis... | 0 | |
| 1846 | Sean Penn is not exactly one to mince words, a... | 0 | |

```
df.reset_index(inplace = True)
df.drop(["index"], axis = 1, inplace = True)
```

```
df.columns
```

```
Index(['text', 'class'], dtype='object')
```

```
df.head()
```

| | text | class |  |
|---|---|-------|---|
| 0 | BELFAST (Reuters) - The British government on ... | 1 |  |
| 1 | VALLETTA (Reuters) - Police believe a bomb tha... | 1 | |
| 2 | Obama and HUD want to give one last freebie to... | 0 | |
| 3 | The French citizens just voted for their demis... | 0 | |
| 4 | Sean Penn is not exactly one to mince words, a... | 0 | |

✓ Creating a function to process the texts

```
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
```

```
df["text"] = df["text"].apply(wordopt)
```

✓ Defining dependent and independent variables

```
x = df["text"]
y = df["class"]
```

✓ Splitting Training and Testing

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25)
```

✓ Convert text to vectors

```
from sklearn.feature_extraction.text import TfidfVectorizer

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

✓ Logistic Regression

```
from sklearn.linear_model import LogisticRegression
```

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

```
▼ LogisticRegression
LogisticRegression()
```

```
pred_lr=LR.predict(xv_test)
```

```
LR.score(xv_test, y_test)
```

```
0.9874331550802139
```

```
print(classification_report(y_test, pred_lr))
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.99 | 0.99 | 0.99 | 5811 |
| 1 | 0.99 | 0.99 | 0.99 | 5409 |
| accuracy | | | 0.99 | 11220 |
| macro avg | 0.99 | 0.99 | 0.99 | 11220 |
| weighted avg | 0.99 | 0.99 | 0.99 | 11220 |

✓ Decision Tree Classification

```
from sklearn.tree import DecisionTreeClassifier
```

```
DT = DecisionTreeClassifier()
DT.fit(xv_train, y_train)
```

```
▼ DecisionTreeClassifier
DecisionTreeClassifier()
```

```
pred_dt = DT.predict(xv_test)
```

```
DT.score(xv_test, y_test)
```

```
0.996078431372549
```

```
print(classification_report(y_test, pred_dt))
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 1.00 | 1.00 | 1.00 | 5811 |
| 1 | 1.00 | 1.00 | 1.00 | 5409 |
| accuracy | | | 1.00 | 11220 |
| macro avg | 1.00 | 1.00 | 1.00 | 11220 |
| weighted avg | 1.00 | 1.00 | 1.00 | 11220 |

✓ Gradient Boosting Classifier

```
from sklearn.ensemble import GradientBoostingClassifier
```

```
GBC = GradientBoostingClassifier(random_state=0)
GBC.fit(xv_train, y_train)
```

```
▼ GradientBoostingClassifier
GradientBoostingClassifier(random_state=0)
```

```
pred_gbc = GBC.predict(xv_test)
```

```
GBC.score(xv_test, y_test)
```

```
0.995632798573975
```

```
print(classification_report(y_test, pred_gbc))
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 1.00 | 0.99 | 1.00 | 5811 |
| 1 | 0.99 | 1.00 | 1.00 | 5409 |
| accuracy | | | 1.00 | 11220 |
| macro avg | 1.00 | 1.00 | 1.00 | 11220 |
| weighted avg | 1.00 | 1.00 | 1.00 | 11220 |

✓ Random Forest Classifier

```
from sklearn.ensemble import RandomForestClassifier
```

```
RFC = RandomForestClassifier(random_state=0)
RFC.fit(xv_train, y_train)
```

```
▼ RandomForestClassifier
RandomForestClassifier(random_state=0)
```

```
pred_rfc = RFC.predict(xv_test)
```

```
RFC.score(xv_test, y_test)
```

```
0.9907308377896613
```

```
print(classification_report(y_test, pred_rfc))
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.99 | 0.99 | 0.99 | 5811 |
| 1 | 0.99 | 0.99 | 0.99 | 5409 |
| accuracy | | | 0.99 | 11220 |
| macro avg | 0.99 | 0.99 | 0.99 | 11220 |
| weighted avg | 0.99 | 0.99 | 0.99 | 11220 |

```
def output_label(n):
    if n == 0:
        return "Fake News"
    elif n == 1:
        return "Not A Fake News"

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_label(pred_LR[0]),
                                                                 output_label(pred_GBC[0]),
                                                                 output_label(pred_RFC[0]))
```

```
news = str(input())
manual_testing(news)
```

{euters) - Transgender people will be allowed for the first time to enlist in the U.S. military starting on Monday as ordered by fed

```
1: Not A Fake News
1: Not A Fake News
0n: Not A Fake News
0n: Not A Fake News
```



```
news = str(input())
manual_testing(news)
```

Donald Trump just couldn t wish all Americans a Happy New Year and leave it at that. Instead, he had to give a shout out to his ene

```
LR Prediction: Fake News
DT Prediction: Fake News
GBC Prediction: Fake News
RFC Prediction: Fake News
```

