```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
from google.colab import files
uploaded=files.upload()
```

Choose Files 2 files
- **True.csv**(text/csv) - 53582940 bytes, last modified: 12/26/2023 - 100% done
- **Fake.csv**(text/csv) - 62789876 bytes, last modified: 12/26/2023 - 100% done
Saving True.csv to True (1).csv
Saving Fake.csv to Fake (1).csv

```python
data_fake = pd.read_csv("Fake.csv")
data_true = pd.read_csv("True.csv")
```

```python
data_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
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017

```python
data_true.head()
```

	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
	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsN	December 30,

```
data_fake.tail()
```

	title	text	subject	date
23476	McPain: John McCain Furious That Iran Treated ...	21st Century Wire says As 21WIRE reported earl...	Middle-east	January 16, 2016
23477	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...	21st Century Wire says It s a familiar theme. ...	Middle-east	January 16, 2016
23478	Sunnistan: US and Allied 'Safe Zone' Plan to T...	Patrick Henningsen 21st Century WireRemember ...	Middle-east	January 15, 2016
23479	How to Blow $700 Million: Al Jazeera America F...	21st Century Wire says Al Jazeera America will...	Middle-east	January 14, 2016
23480	10 U.S. Navy Sailors Held by Iranian Military ...	21st Century Wire says As 21WIRE predicted in ...	Middle-east	January 12, 2016

```
data_true.tail()
```

	title	text	subject	date
21412	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017
21413	LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews	August 22, 2017
21414	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disused Sov...	worldnews	August 22, 2017
21415	Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017
21416	Indonesia to buy $1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	August 22, 2017

```
data_fake["class"]=0
data_true["class"]=1
```

```
data_fake.shape, data_true.shape
```

```
((23481, 5), (21417, 5))
```

```
#Dropping the last 10 columns for manual testing
```

```
data_fake_manual_testing = data_fake.tail(10)
for i in range(23480, 23469, -1):
    data_fake.drop([i], axis=0, inplace=True)
```

```
data_true_manual_testing = data_true.tail(10)
for i in range(21416, 21405, -1):
    data_true.drop([i], axis=0, inplace=True)
```

```
data_fake.shape, data_true.shape
```

```
((23470, 5), (21406, 5))
```

```
data_fake_manual_testing["class"] = 0
data_true_manual_testing["class"] = 1
```

```
    <ipython-input-52-a26436eb6887>: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-a-copy
      data_fake_manual_testing["class"] = 0
    <ipython-input-52-a26436eb6887>: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
      data_true_manual_testing["class"] = 1
```

```
#Data head for fake news
data_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	
23476	McPain: John McCain Furious That Iran Treated ...	21st Century Wire says As 21WIRE reported earl...	Middle-east	January 16, 2016	0	
23477	JUSTICE? Yahoo Settles E-mail Privacy Class-...	21st Century Wire says It s a familiar theme. ...	Middle-east	January 16, 2016	0	

```
#Data head for True news
data_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	
	U.S., North Korea clash at U.N. arms forum on	GENEVA (Reuters) - North Korea and the United		August 22,	

```
#merging the two data heads

data_merge = pd.concat([data_fake, data_true], axis = 0)
data_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	
5	Racist Alabama Cops Brutalize Black Boy While...	The number of cases of cops brutalizing and ki...	News	December 25, 2017	0	
6	Fresh Off The Golf Course, Trump Lashes Out A...	Donald Trump spent a good portion of his day a...	News	December 23, 2017	0	

```
data_merge.columns
```

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

```
#Dropping Columns
data = data_merge.drop(['title','subject','date'], axis =1)
```

```
#checking null values
data.isnull().sum()
```

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

```
#Random shuffling of the dataframes
data = data.sample(frac = 1)
```

```
data.head()
```

	text	class
20853	Hillary is failing miserably with the young vo...	0
16956	TOKYO (Reuters) - Typhoon Lan with heavy rains...	1
17661	Before Keurig, and other major companies commi...	0
19436	A new report from Harvard Kennedy School s Sho...	0
17318	It s no accident that President Obama named Va...	0

```
data.reset_index(inplace = True)
data.drop(['index'], axis = 1, inplace = True)
```

```
data.columns
```

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

```
data.head()
```

	text	class
0	Hillary is failing miserably with the young vo...	0
1	TOKYO (Reuters) - Typhoon Lan with heavy rains...	1
2	Before Keurig, and other major companies commi...	0
3	A new report from Harvard Kennedy School s Sho...	0
4	It s no accident that President Obama named Va...	0

```
#Removing Special characters

def word(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

data['text'] = data['text'].apply(word)
```

```python
#Defining dependent and independent variables

x = data['text']
y = data['class']
```

```python
#splitting training and testing data

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

```python
#converting to vectors

from sklearn.feature_extraction.text import TfidfVectorizer
vectorization = TfidfVectorizer()
xv_train = vectorization.fit_transform(x_train)
xv_test = vectorization.transform(x_test)
```

```python
from sklearn.linear_model import LogisticRegression
LR = LogisticRegression()
LR.fit(xv_train, y_train)
```

```
▾ LogisticRegression
LogisticRegression()
```

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

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

```
0.9877885729565915
```

Hence, we can see that the accuracy is 98.7% which is good.

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

```
              precision    recall  f1-score   support

           0       0.99      0.99      0.99      5901
           1       0.99      0.99      0.99      5318

    accuracy                           0.99     11219
   macro avg       0.99      0.99      0.99     11219
weighted avg       0.99      0.99      0.99     11219
```

```python
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)
```

```
0.9958998128175417
```

Here, we can see that the accuracy is 99.5% which is better than the Linear Regression

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

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      5901
           1       1.00      1.00      1.00      5318

    accuracy                           1.00     11219
   macro avg       1.00      1.00      1.00     11219
weighted avg       1.00      1.00      1.00     11219
```

```
from sklearn.ensemble import GradientBoostingClassifier
GB = GradientBoostingClassifier(random_state = 0)
GB.fit(xv_train, y_train)
```

```
▾        GradientBoostingClassifier
GradientBoostingClassifier(random_state=0)
```

```
pred_gb = GB.predict(xv_test)
```

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

```
0.9950976022818433
```

```
print(classification_report(y_test,pred_gb))
```

```
              precision    recall  f1-score   support

           0       1.00      0.99      1.00      5901
           1       0.99      1.00      0.99      5318

    accuracy                           1.00     11219
   macro avg       1.00      1.00      1.00     11219
weighted avg       1.00      1.00      1.00     11219
```

Hence, we can conclude that the accuracy is not good than the DecisionTreeClassifier but almost near to it.

```python
from sklearn.ensemble import RandomForestClassifier

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

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

```python
pred_rf = RF.predict(xv_test)
```

```python
RF.score(xv_test, y_test)
```

```
    0.9907300115874855
```

```python
print(classification_report(y_test,pred_rf))
```

```
                  precision    recall  f1-score   support

               0       0.99      0.99      0.99      5901
               1       0.99      0.99      0.99      5318

        accuracy                           0.99     11219
       macro avg       0.99      0.99      0.99     11219
    weighted avg       0.99      0.99      0.99     11219
```

Here, we can conclude that the score is 99% which is a good score again.

```python
#Code for Model Testing:

def output_label(n):
  if n == 0:
    return "Its a Fake News"
  elif n == 1:
    return "It's a True 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(word)
  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_GB = GB.predict(new_xv_test)
```

```
    pred_RF = RF.predict(new_xv_test)
    print("\n\nLR Prediction: {} \nDT Prediction: {} \nGB Prediction: {} \nRF Prediction: {}".format(
        output_label(pred_LR[0]),
        output_label(pred_GB[0]),
        output_label(pred_RF[0]),
        output_label(pred_DT[0])
))


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

```
St. Paul says it fixed every pothole in the city


    LR Prediction: Its a Fake News
    DT Prediction: Its a Fake News
    GB Prediction: Its a Fake News
    RF Prediction: Its a Fake News
```