

▼ Fake news Detection

▼ Importing required library

Here I am going to importing some of the required library, if extra library is required to install It will be install later on.

```
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
```

▼ Inserting fake and real dataset

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

```
from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive",



```
df_fake.head(5)
```

	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
-	Trump Is So Obsessed He Even Has Obama's	On Christmas day. Donald Trump announced	..	December 29.

```
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
-	FBI Russia probe helped by Australian	WASHINGTON (Reuters) - Trump campaign	December 30.

Inserting a column called "class" for fake and real news dataset to categories fake and true news.

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

Removing last 10 rows from both the dataset, for manual testing

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

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

```
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))
```

Merging the manual testing dataframe in single dataset and save it in a csv file

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

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py: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:

```
"""Entry point for launching an IPython kernel.
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py: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:



```
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 ..	Middle-	January 19, 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
21412	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017	1
21413	LexisNexis withdrew two products from Chinese	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews	August 22, 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 the main fake and true dataframe

```
df_marge = pd.concat([df_fake, df_true], axis =0 )
df_marge.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 ...	News	December 23, 2017	0

```
df_marge.columns

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

- ▼ "title", "subject" and "date" columns is not required for detecting the fake news, so I am going to drop the columns.

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

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

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

- ▼ Randomly shuffling the dataframe

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

```
df.head()
```

	text	class	
14001	JOHANNESBURG (Reuters) - The Southern African ...	1	
13748	One short trip on a boat and this could be all...	0	
14173	Boom! Share this everywhere, as Hillary aligns...	0	
1817	Just when you thought you ve heard the ultimat...	0	
9562	WASHINGTON (Reuters) - Kentucky Senator Rand P...	1	

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

```
df.columns
```

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

```
df.head()
```

	text	class	
0	JOHANNESBURG (Reuters) - The Southern African ...	1	
1	One short trip on a boat and this could be all...	0	
2	Boom! Share this everywhere, as Hillary aligns...	0	
3	Just when you thought you ve heard the ultimat...	0	
4	WASHINGTON (Reuters) - Kentucky Senator Rand P...	1	

- ▼ Creating a function to convert the text in lowercase, remove the extra space, special chr., ulr and links.

```
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 variable as x and y

```

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

```

▼ Splitting the dataset into training set and testing set.

```
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)

```

▼ 1. Logistic Regression

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

```

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

```

```
LogisticRegression()
```

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

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

```
0.9836898395721925
```

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

	precision	recall	f1-score	support
0	0.99	0.98	0.98	5948
1	0.98	0.99	0.98	5272
accuracy			0.98	11220

macro avg	0.98	0.98	0.98	11220
weighted avg	0.98	0.98	0.98	11220

▼ 2. Decision Tree Classification

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

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

```
DecisionTreeClassifier()
```

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

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

```
0.9954545454545455
```

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

	precision	recall	f1-score	support
0	1.00	1.00	1.00	5948
1	1.00	0.99	1.00	5272
accuracy			1.00	11220
macro avg	1.00	1.00	1.00	11220
weighted avg	1.00	1.00	1.00	11220

▼ 3. Gradient Boosting Classifier

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

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

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

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

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

```
0.9940285204991087
```

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

	precision	recall	f1-score	support
0	1.00	0.99	0.99	5948
1	0.99	1.00	0.99	5272
accuracy			0.99	11220

macro avg	0.99	0.99	0.99	11220
weighted avg	0.99	0.99	0.99	11220

▼ Model Testing With Manual Entry

News

```
def output_lable(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)

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

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

☞ 21st Century Wire says This week, the historic international Iranian Nuclear Deal was punctuated by a two-w

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

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

SAO PAULO (Reuters) - Cesar Mata Pires, the owner and co-founder of Brazilian engineering conglomerate OAS

```
LR Prediction: Not A Fake News
DT Prediction: Not A Fake News
GBC Prediction: Not A Fake News
```

✓

15s

completed at 12:48 PM

●

✕