

Fake News Analysis

Importing Required Libraries

```
#modules
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
from sklearn.metrics import roc_curve, auc
import string
import re
```

Loading the Dataset

```
fake_data=pd.read_csv("Fake.csv")
true_data=pd.read_csv("True.csv")

#First five columns and last five columns of fake and true dataset
fake_data.head()
```

		title \
0	Donald Trump Sends Out Embarrassing New Year'...	
1	Drunk Bragging Trump Staffer Started Russian ...	
2	Sheriff David Clarke Becomes An Internet Joke...	
3	Trump Is So Obsessed He Even Has Obama's Name...	
4	Pope Francis Just Called Out Donald Trump Dur...	

		text	subject \
0	Donald Trump just couldn t wish all Americans ...		News
1	House Intelligence Committee Chairman Devin Nu...		News
2	On Friday, it was revealed that former Milwauk...		News
3	On Christmas day, Donald Trump announced that ...		News
4	Pope Francis used his annual Christmas Day mes...		News

	date
0	December 31, 2017
1	December 31, 2017
2	December 30, 2017
3	December 29, 2017
4	December 25, 2017

```
#last five rows of fake_data
```

```
fake_data.tail()
```

		title	\
23476	McPain: John McCain Furious That Iran Treated ...		
23477	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...		
23478	Sunnistan: US and Allied 'Safe Zone' Plan to T...		
23479	How to Blow \$700 Million: Al Jazeera America F...		
23480	10 U.S. Navy Sailors Held by Iranian Military ...		

		text	subject
	\		
23476	21st Century Wire says As 21WIRE reported earl...		Middle-east
23477	21st Century Wire says It s a familiar theme. ...		Middle-east
23478	Patrick Henningsen 21st Century WireRemember ...		Middle-east
23479	21st Century Wire says Al Jazeera America will...		Middle-east
23480	21st Century Wire says As 21WIRE predicted in ...		Middle-east

	date
23476	January 16, 2016
23477	January 16, 2016
23478	January 15, 2016
23479	January 14, 2016
23480	January 12, 2016

```
#first five rows of true data
```

```
true_data.head()
```

		title	\
0	As U.S. budget fight looms, Republicans flip t...		
1	U.S. military to accept transgender recruits o...		
2	Senior U.S. Republican senator: 'Let Mr. Muell...		
3	FBI Russia probe helped by Australian diplomat...		
4	Trump wants Postal Service to charge 'much mor...		

		text	subject	\
0	WASHINGTON (Reuters) - The head of a conservat...		politicsNews	
1	WASHINGTON (Reuters) - Transgender people will...		politicsNews	
2	WASHINGTON (Reuters) - The special counsel inv...		politicsNews	
3	WASHINGTON (Reuters) - Trump campaign adviser ...		politicsNews	
4	SEATTLE/WASHINGTON (Reuters) - President Donal...		politicsNews	

	date
0	December 31, 2017
1	December 29, 2017
2	December 31, 2017

```
3 December 30, 2017
4 December 29, 2017
```

```
#last five rows of true data
true_data.tail()
```

	title \
21412	'Fully committed' NATO backs new U.S. approach...
21413	LexisNexis withdrew two products from Chinese ...
21414	Minsk cultural hub becomes haven from authorities
21415	Vatican upbeat on possibility of Pope Francis ...
21416	Indonesia to buy \$1.14 billion worth of Russia...

	text	subject \
21412	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews
21413	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews
21414	MINSK (Reuters) - In the shadow of disused Sov...	worldnews
21415	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews
21416	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews

	date
21412	August 22, 2017
21413	August 22, 2017
21414	August 22, 2017
21415	August 22, 2017
21416	August 22, 2017

Labeling and Combining Data

```
# Add Label (fake=0, real=1)
fake_data["label"]=0
true_data["label"]=1
```

```
#shape of both
fake_data.shape
```

```
(23481, 5)
```

```
true_data.shape
```

```
(21417, 5)
```

```
#merge two data file and make one
df=pd.concat([fake_data,true_data], axis=0)

df.head()
```

	title \
0	Donald Trump Sends Out Embarrassing New Year'...
1	Drunk Bragging Trump Staffer Started Russian ...

```

2 Sheriff David Clarke Becomes An Internet Joke...
3 Trump Is So Obsessed He Even Has Obama's Name...
4 Pope Francis Just Called Out Donald Trump Dur...

                                text subject \
0 Donald Trump just couldn t wish all Americans ... News
1 House Intelligence Committee Chairman Devin Nu... News
2 On Friday, it was revealed that former Milwauk... News
3 On Christmas day, Donald Trump announced that ... News
4 Pope Francis used his annual Christmas Day mes... News

                                date label
0 December 31, 2017      0
1 December 31, 2017      0
2 December 30, 2017      0
3 December 29, 2017      0
4 December 25, 2017      0

df.shape
(44898, 5)

df.dtypes
title      object
text       object
subject     object
date       object
label      int64
dtype: object

#columns
df.columns
Index(['title', 'text', 'subject', 'date', 'label'], dtype='object')

```

Text Cleaning & Preprocessing

```

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

#Again shape
data.shape
(44898, 2)

print(df["title"])
0 Donald Trump Sends Out Embarrassing New Year'...
1 Drunk Bragging Trump Staffer Started Russian ...
2 Sheriff David Clarke Becomes An Internet Joke...

```

```

3      Trump Is So Obsessed He Even Has Obama's Name...
4      Pope Francis Just Called Out Donald Trump Dur...
...
21412  'Fully committed' NATO backs new U.S. approach...
21413  LexisNexis withdrew two products from Chinese ...
21414  Minsk cultural hub becomes haven from authorities
21415  Vatican upbeat on possibility of Pope Francis ...
21416  Indonesia to buy $1.14 billion worth of Russia...
Name: title, Length: 44898, dtype: object

```

```
print(df["label"])
```

```

0      0
1      0
2      0
3      0
4      0
...
21412  1
21413  1
21414  1
21415  1
21416  1
Name: label, Length: 44898, dtype: int64

```

```
#shuffle the data
```

```
data=data.sample(frac=1).reset_index(drop=True)
```

```
data.head()
```

	text	label
0	The Berkeley, California anti-Trump protesters...	0
1	So, if the climate scare-mongers can t fright...	0
2	By ramping up US troop levels in Afghanistan,...	0
3	J.R. Smith 21st Century WireFor the last two ...	0
4	Peter Certo Other WordsEven in an election yea...	0

```
data.head(50)
```

	text	label
0	The Berkeley, California anti-Trump protesters...	0
1	So, if the climate scare-mongers can t fright...	0
2	By ramping up US troop levels in Afghanistan,...	0
3	J.R. Smith 21st Century WireFor the last two ...	0
4	Peter Certo Other WordsEven in an election yea...	0
5	White House counselor Kellyanne Conway has pre...	0
6	Trump wins big in New Hampshire in what is an ...	0
7	Just about every journalist who was waiting to...	0
8	ISLAMABAD (Reuters) - A hardline Pakistani Isl...	1
9	Australian-born Creationist Ken Ham and his fu...	0
10	This story is from 2006 and is the first in a ...	0

11	WASHINGTON (Reuters) - With a hug and a handsh...	1
12	The Clinton Carpetbaggers have moved around do...	0
13	NEW YORK (Reuters) - A meeting in New York on ...	1
14	BANGKOK (Reuters) - An election could be a yea...	1
15	LONDON (Reuters) - British counter-terrorism p...	1
16	WASHINGTON (Reuters) - The U.S. military said ...	1
17	HARARE (Reuters) - Ousted Zimbabwean vice pres...	1
18	Vic Bishop Waking TimesOur reality is carefull...	0
19		0
20	WASHINGTON (Reuters) - The U.S. military said ...	1
21		0
22	Nothing like political cronyism to make your s...	0
23	WASHINGTON (Reuters) - The U.S. Senate cleared...	1
24	Sean Hannity and Tucker Carlson went after the...	0
25	TRIPOLI (Reuters) - Former Libyan Prime Minist...	1
26	NEW YORK/WASHINGTON (Reuters) - This year's pr...	1
27	LONDON (Reuters) - British Brexit minister Dav...	1
28	The irony of the Tweet put out on Mother s Day...	0
29	WASHINGTON (Reuters) - Republican presidential...	1
30	Dominican newspaper El Nacional had a bit of f...	0
31	Lawless, Crooked Hillary promises to continue ...	0
32	(This March 22 story was corrected to remove ...	1
33	MSNBC host Rachel Maddow found the perfect way...	0
34	ISLAMABAD/LAHORE, Pakistan (Reuters) - Pakista...	1
35	WASHINGTON/BEIJING (Reuters) - China's plan to...	1
36	Even Fox News understands that the Republican ...	0
37	Hell has a special spot reserved for this horr...	0
38	Now, you d think after calling Mexicans rapist...	0
39	NEW YORK/SAN FRANCISCO (Reuters) - Legal chall...	1
40	Director Michael Bay is perhaps best known for...	0
41	REINEH, Israel (Reuters) - A week after Daren...	1
42	WASHINGTON (Reuters) - New Jersey Governor Chr...	1
43	WASHINGTON (Reuters) - Paul Manafort, Presiden...	1
44	(Reuters) - Donald Trump knows he could stand ...	1
45	(This version of the November 8 story correct...	1
46	WASHINGTON (Reuters) - U.S. President Donald T...	1
47	Former reality show star Donald Trump has repe...	0
48	The saga surrounding the idiots who decided to...	0
49	POCHEON, South Korea (Reuters) - The screech o...	1

#Check null values

data.isnull().sum()

text 0

label 0

dtype: int64

data.duplicated().sum()

np.int64(6251)

```

data=data.drop_duplicates().reset_index(drop=True)

#again check
data.duplicated().sum()

np.int64(0)

data.shape

(38647, 2)

#text preprocessing
#This function will done to remove all punctuation, commas and also
headind,style,https and many mores
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

#apply on data
data["text"]=data["text"].apply(wordopt)

data.head(10)

```

Splitting Data into Train & Test

```

x=data["text"]
y=data["label"]

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

# Convert text into numerical ways that model learn out best
from sklearn.feature_extraction.text import TfidfVectorizer

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

```

Model Training (Logistic Regression, etc.)

```

from sklearn.linear_model import LogisticRegression
LR = LogisticRegression()

```

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

0.9828192920720348

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

	precision	recall	f1-score	support
0	0.99	0.98	0.98	4399
1	0.98	0.99	0.98	5263
accuracy			0.98	9662
macro avg	0.98	0.98	0.98	9662
weighted avg	0.98	0.98	0.98	9662

Decision tree

```
from sklearn.tree import DecisionTreeClassifier
DT = DecisionTreeClassifier()
DT.fit(xv_train, y_train)
pred_dt = DT.predict(xv_test)
DT.score(xv_test, y_test)
```

0.9941006002897951

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

	precision	recall	f1-score	support
0	0.99	0.99	0.99	4399
1	0.99	1.00	0.99	5263
accuracy			0.99	9662
macro avg	0.99	0.99	0.99	9662
weighted avg	0.99	0.99	0.99	9662

Gradient boosting classifier

```
from sklearn.ensemble import GradientBoostingClassifier
GB = GradientBoostingClassifier(random_state = 0)
GB.fit(xv_train, y_train)
pred_gb = GB.predict(xv_test)
GB.score(xv_test, y_test)
```

0.9958600703788035

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

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	1.00	0.99	1.00	4399
1	0.99	1.00	1.00	5263
accuracy			1.00	9662
macro avg	1.00	1.00	1.00	9662
weighted avg	1.00	1.00	1.00	9662

Random forest

```
from sklearn.ensemble import RandomForestClassifier
RF = RandomForestClassifier(random_state = 0)
RF.fit(xv_train, y_train)
pred_rf = RF.predict(xv_test)
RF.score(xv_test, y_test)
```

0.9835437797557441

```
def output_lable(n):
```

```
    if n ==0:
```

```
        return "Fake News"
```

```
    elif n==1:
```

```
        return "Not A Fake News"
```

```
def testing(news):
```

```
    testing_news = {"test":[news]}
```

```
    new_def_test = pd.DataFrame(testing_news)
```

```
    new_def_test["text"] = new_def_test["test"].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_GB = GB.predict(new_xv_test)
```

```
    pred_RF = RF.predict(new_xv_test)
```

```
    return print("\n\nLR Prediction: {} \nDT Prediction: {} \nGB
Prediction: {} \nRF Prediction: {}".format(output_lable(pred_LR[0]),
output_lable(pred_DT[0]), output_lable(pred_GB[0]),
output_lable(pred_RF[0])))
```

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

```
testing(news)
```

hello

LR Prediction: Fake News

DT Prediction: Fake News

GB Prediction: Fake News

RF Prediction: Fake News

