

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.

```
In [1]: 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 plotly.graph_objects as go
```

Inserting fake and real dataset

```
In [2]: df_fake = pd.read_csv("Fake.csv")
df_true = pd.read_csv("True.csv")

In [3]: df_fake.head(10)
```

	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 Milwauc...	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
5	Racist Alabama Cops Brutalize Black Boy While...	The number of cases of cops brutalizing and ki...	News	December 25, 2017
6	Fresh Off The Golf Course, Trump Lashes Out A...	Donald Trump spent a good portion of his day a...	News	December 23, 2017
7	Trump Said Some INSANELY Racist Stuff Inside...	In the wake of yet another court decision tha...	News	December 23, 2017
8	Former CIA Director Slams Trump Over UN Bully...	Many people have raised the alarm regarding th...	News	December 22, 2017
9	WATCH: Brand-New Pro-Trump Ad Features So Mu...	Just when you might have thought we'd get a br...	News	December 21, 2017

```
In [4]: df_true.head(10)
```

	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 u...	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
3	FBI Russia probe helped by Australian diplom...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017
4	Trump wants Postal Service to change 'much mor...	SEATTLEWASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017
5	White House: Congress 'prone to taking on sp...	WEST PALM BEACH, Fla. WASHINGTON (Reuters) - T...	politicsNews	December 29, 2017
6	Trump says Russia probe will be fair, but ene...	WEST PALM BEACH, Fla. (Reuters) - President Don...	politicsNews	December 29, 2017
7	Factio: Trump on Twitter (Dec 28) - Approval ...	The following statements were posted to the ve...	politicsNews	December 29, 2017
8	Trump on Twitter (Dec 28) - Global Warning	The following statements were posted to the ve...	politicsNews	December 29, 2017
9	Alabama official to certify Senator-elect Jone...	WASHINGTON (Reuters) - Alabama Secretary of St...	politicsNews	December 28, 2017

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

```
In [5]: df_fake["type"] = "FAKE"
df_true["type"] = "REAL"

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

```
In [6]: df_fake.shape, df_true.shape
((23481, 5), (23417, 5))

In [7]: 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(23416, 23406, -1):
    df_true.drop(i, axis=0, inplace = True)
```

```
In [8]: df_fake.shape, df_true.shape
((23471, 5), (23407, 5))

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

```
In [9]: df_fake_manual_testing.head(10)
```

	title	text	subject	date	type
23471	Seven Iranians freed in the prisoner swap have...	21st Century Wire says This week, the historic...	Middle-east	January 20, 2016	FAKE
23472	#Hashtag Hell & The Fake Left	By Dady Cherry and Gilbert MercierAll writers ...	Middle-east	January 19, 2016	FAKE
23473	Astronauting: Journalism Reveals Brainwashing...	Vic Bishop Waking TimesOur reality is careful...	Middle-east	January 19, 2016	FAKE
23474	The New American Century: An Era of Fraud	Paul Craig RobertsIn the last years of the 20...	Middle-east	January 19, 2016	FAKE
23475	Hillary Clinton: 'Israel First' (and no peace...	Robert FarinaIn CounterpunchAlthough the United...	Middle-east	January 18, 2016	FAKE
23476	McPain: John McCain Furious That Iran Treated...	21st Century Wire says As 21WIRE reported earl...	Middle-east	January 16, 2016	FAKE
23477	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...	21st Century Wire says It's a familiar theme...	Middle-east	January 16, 2016	FAKE
23478	Sunistan: US and Allied 'Safe Zone' Plan to T...	Patrick Henningsen 21st Century WireRemember ...	Middle-east	January 15, 2016	FAKE
23479	How to Blow \$700 Million: Al Jazeera America F...	21st Century Wire says Al Jazeera America will...	Middle-east	January 14, 2016	FAKE
23480	10 U.S. Navy Sailors Held by Iranian Military ...	21st Century Wire says As 21WIRE predicted in ...	Middle-east	January 12, 2016	FAKE

```
In [10]: df_true_manual_testing.head(10)
```

	title	text	subject	date	type
21407	Mata Pires, owner of embattled Brazil builder...	SAO PAULO (Reuters) - Cesar Mata Pires, the ow...	worldnews	August 22, 2017	REAL
21408	U.S., North Korea clash at U.N. forum over nuc...	GENEVA (Reuters) - North Korea and the United ...	worldnews	August 22, 2017	REAL
21409	U.S., North Korea clash at U.N. arms forum on ...	GENEVA (Reuters) - North Korea and the United ...	worldnews	August 22, 2017	REAL
21410	Headless torso could belong to submarine journ...	COPENHAGEN (Reuters) - Danish police said on T...	worldnews	August 22, 2017	REAL
21411	North Korea shipments to Syria chemical arms a...	UNITED NATIONS (Reuters) - Two North Korean sh...	worldnews	August 21, 2017	REAL
21412	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017	REAL
21413	LexisNexis withdraws two products from Chinese...	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews	August 22, 2017	REAL
21414	Minsk cultural hub becomes haven from authorit...	MINSK (Reuters) - In the shadow of disused Sov...	worldnews	August 22, 2017	REAL
21415	Vatican appeal on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017	REAL
21416	Indonesia to buy \$1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	August 22, 2017	REAL

```
In [11]: 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
```

```
In [12]: df_marge = pd.concat([df_fake, df_true], axis =0 )
df_marge.head(10)
```

	title	text	subject	date	type
0	Donald Trump Sends Out Embarrassing New Year...	Donald Trump just couldn't wish all Americans ...	News	December 31, 2017	FAKE
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	FAKE
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauc...	News	December 30, 2017	FAKE
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that...	News	December 29, 2017	FAKE
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	FAKE
5	Racist Alabama Cops Brutalize Black Boy While...	The number of cases of cops brutalizing and ki...	News	December 25, 2017	FAKE
6	Fresh Off The Golf Course, Trump Lashes Out A...	Donald Trump spent a good portion of his day a...	News	December 23, 2017	FAKE
7	Trump Said Some INSANELY Racist Stuff Inside...	In the wake of yet another court decision tha...	News	December 23, 2017	FAKE
8	Former CIA Director Slams Trump Over UN Bully...	Many people have raised the alarm regarding th...	News	December 22, 2017	FAKE
9	WATCH: Brand-New Pro-Trump Ad Features So Mu...	Just when you might have thought we'd get a br...	News	December 21, 2017	FAKE

```
In [13]: df_marge.columns
Out[13]: Index(['title', 'text', 'subject', 'date', 'type'], dtype='object')
```

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

```
In [14]: df = df_marge.drop(["title", "subject", "date"], axis = 1)
```

Randomly shuffling the dataframe

```
In [15]: df = df.sample(frac = 1)

In [16]: df.head(10)
```

	text	type
22176	21st Century Wire says While the US media conti...	FAKE
980	WASHINGTON (Reuters) - The U.S. House intell...	REAL
4176	Donald Trump, while pretending to reach out to...	FAKE
19056	Barack Obama is turning his new home in the po...	FAKE
19999	LONDON (Reuters) - British Prime Minister Ther...	REAL
14741	ZURICH (Reuters) - Switzerland will stop seiz...	REAL
7351	CHICAGO (Reuters) - A former U.S. representat...	REAL
19945	Watch Thugs Protest Cop Killing By Committing...	FAKE
4453	Fox News needs to take Andrea Tantaros sexual...	FAKE
7593	HUNTINGTON, WV (Reuters) - When Carissa Sell...	REAL

```
In [17]: df.type.value_counts()
Out[17]:
FAKE    23471
REAL    21407
Name: type, dtype: int64
```

```
In [18]: !df.type.value_counts()

In [19]: fig=go.Figure(data=[go.Bar(x=['Real', 'Fake'], y=[1, 2], text=['', ''], textposition='auto')])

In [20]: fig.show()
```



```
In [21]: df.isnull().sum()
Out[21]:
text      0
type      0
dtype: int64

Creating a function to convert the text in lowercase, remove the extra space, special chr, url and links.
```

```
In [22]: def wordopt(text):
    text = text.lower()
    text = re.sub('[\r\n]', '', text)
    text = re.sub("[\\W]", "", text)
    text = re.sub("https?://\S+|www.\S+", '', text)
    text = re.sub("http://\S+", '', text)
    text = re.sub("[\s]", % re.escape(string.punctuation), '')
    text = re.sub('[\n]', '', text)
    text = re.sub("[a-z\dW+]", '', text)
    return text
```

```
In [23]: df["text"] = df["text"].apply(wordopt)
df.head(10)
```

	text	type
22176	century wire says while the us media continue...	FAKE
980	washington reuters the u s house intell...	REAL
4176	donald trump while pretending to reach out to...	FAKE
19056	barack obama is turning his new home in the po...	FAKE
19999	london reuters british prime minister ther...	REAL
14741	zurich reuters switzerland will stop seiz...	REAL
7351	chicago reuters a former u s representat...	REAL
19945	watch thugs protest cop killing by committing...	FAKE
4453	fox news needs to take andrea tantaros sexual...	FAKE
7593	huntington wv reuters when carissa sell...	REAL

Defining dependent and independent variable as x and y

```
In [24]: x = df["text"]
y = df["type"]
```

Splitting the dataset into training set and testing set.

```
In [25]: x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25)

In [26]: y_test
```

	type
1017	REAL
12975	FAKE
12780	REAL
12521	REAL
8693	REAL
...	
22047	FAKE
12634	REAL
16824	FAKE
9641	REAL
5467	FAKE
Name:	type, Length: 11226, dtype: object

```
In [27]: x_train
```

	text
23922	century wire says in a press release from the...
21116	if ga residents are paying attention these sl...
12404	tampa fla reuters democratic presidential...
8255	washington dc foreigners aiming for tempor...
11486	virginia s voter id law is being challenged i...
...	
6514	washington reuters president barack obama ...
7632	after they were crushed in the election repu...
14941	as it turns out police brutally really isn't...
5119	berkeley calif banning mich reuters s...
3598	former vermont governor howard dean has alread...
Name:	text, Length: 33658, dtype: object

Convert text to vectors

```
In [28]: from sklearn.feature_extraction.text import TfidfVectorizer

In [29]: vectorization = TfidfVectorizer()
xv_train = vectorization.fit_transform(x_train)
xv_test = vectorization.transform(x_test)
```

1. Logistic Regression

```
In [30]: from sklearn.linear_model import LogisticRegression

In [31]: LR = LogisticRegression()
LR.fit(xv_train, y_train)
```

```
Out[31]: LogisticRegression()

In [32]: LR.score(xv_test, y_test)
print("Accuracy", LR.score(xv_test, y_test)*100, "%")
Accuracy= 98.53832442967736 %
```

```
In [33]: pred_lr=LR.predict(xv_test)

In [34]: print(classification_report(y_test, pred_lr))
```

	precision	recall	f1-score	support
FAKE	0.99	0.99	0.99	5886
REAL	0.98	0.99	0.98	5414
accuracy			0.99	11220
macro avg	0.99	0.99	0.99	11220
weighted avg	0.99	0.99	0.99	11220

2. Decision Tree Classification

```
In [35]: from sklearn.tree import DecisionTreeClassifier

In [36]: DT = DecisionTreeClassifier()
DT.fit(xv_train, y_train)
```

```
Out[36]: DecisionTreeClassifier()

In [37]: DT.score(xv_test, y_test)
print("Accuracy", DT.score(xv_test, y_test)*100, "%")
Accuracy= 99.6978431372549 %
```

```
In [38]: pred_dt = DT.predict(xv_test)

In [39]: print(classification_report(y_test, pred_dt))
```

	precision	recall	f1-score	support
FAKE	1.00	1.00	1.00	5886
REAL	1.00	1.00	1.00	5414
accuracy			1.00	11220
macro avg	1.00	1.00	1.00	11220
weighted avg	1.00	1.00	1.00	11220

Model Testing With Manual Entry

News

```
In [40]: def output_label(n):
    if n == "FAKE":
        return "Fake News!!!!!!!"
    elif n == "REAL":
        return "Not A Fake News(Verified 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"]
    pred_lr = LR.predict(new_x_test)
    pred_DT = DT.predict(new_x_test)
    # pred_GBC = GBC.predict(new_x_test)
    # pred_RFC = RFC.predict(new_x_test)

    return print("\n\nt*****RESULTS*****\n\nLogistic Regression Prediction: {} \nDecision Tree Prediction: {}".format(output_label(pred_lr[0]), output_label(pred_DT[0])))
    #
    #
    output_label(pred_GBC[0]),
    output_label(pred_RFC[0]))
```

```
In [41]: 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
```

*****RESULTS*****:

Logistic Regression Prediction: Fake News!!!!!!

Decision Tree Prediction: Fake News!!!!!!

```
In [43]: news = str(input())
manual_testing(news)

GENEVA (Reuters) - North Korea and the United States clashed at a U.N. forum on Tuesday over their military intentions towards one another, with Pyongyang's envoy declaring it would never put its nuclear deterrent on the negotiating table. Japan, well within reach of North Korea's missiles, said the world must maintain pressure on the exclusive country to rein in its nuclear and missile programs and now was not the time for a resumption of multi-party talks. North Korea has pursued its weapons programs in defiance of U.N. Security Council sanctions and ignored all calls, including from major ally China, to stop, prompting a bellicose exchange of rhetoric between the North and the United States. North Korea justifies its weapons programs, including its recent threat to throw missiles towards the U.S. Pacific territory of Guam, by pointing to perceived U.S. hostility, such as military exercises with South Korea this week. U.S. disarmament ambassador Robert Wood told a U.N.-sponsored Conference on Disarmament in Geneva U.S. President Donald Trump's top priority was to protect the United States and its allies against the growing threat from North Korea and America was ready to use the full range of capabilities at its disposal. The path to dialogue still remained an option for Pyongyang and it had the choice between poverty and belligerence on the one hand and prosperity and acceptance. North Korea's envoy told the same forum the North's nuclear deterrent would never be up for negotiation, echoing Pyongyang's regular denunciation of U.S. aggression. The measures taken by the DPRK to strengthen its nuclear deterrence and develop inter-continental rockets is justifiable and a legitimate option for self-defense in the face of such apparent and real threats, diplomat Ju Yong Chol told the forum, referring to constant nuclear threats by the United States. DPRK stands for the North's official name, the Democratic People's Republic of Korea. Reg arding joint U.S.-South Korean military exercises that began on Monday, he warned: The ongoing military adventure would certainly add gasoline to the fire, driving the current tense situation to further deterioration. Japanese Foreign Minister Taro Kono said pressure must be maintained until the North demonstrated it would give up its nuclear program. It's not the time to discuss (the resumption of) peace talks, Kono said, referring to stalled negotiations involving both Korea, the United States, Russia, China and Japan. It's time to exert pressure, he told reporters. The head of the U.S. military's Pacific Command said diplomacy was key. Admiral Harry Harris was in South Korea to observe annual joint military drills with the South Korean military, which the North called a step towards nuclear conflict averted by war maniacs. So we hope and we work for diplomatic solutions to the challenge presented by Kim Jong Un, Harris told reporters at a U.S. air base in Osan, about an hour from Seoul, referring to the North Korean leader. He said diplomacy was the most important starting point in response to the North's threat, when asked what actions by North Korea might trigger a preemptive U.S. strike against it. As far as a timeline, it would be crazy for me to share with you those tripwires in advance. If we did that, it would hardly be a military strategy, he said. The United States and South Korea began the long-planned exercises on Monday, called the Ulchi Freedom Guardian, which the allies have said are purely defensive. The drills involve tens of thousands of troops as well as computer simulations designed to prepare for war with a nuclear-capable North Korea. The United States and South Korea are technically still at war with the North because their 1950-53 conflict ended in a truce, not a peace treaty. Delegations from about 20 countries spoke at the four-hour U.N. session, including Britain, France, Australia and South Korea, all of which criticized North Korea. I would like to repeat the appeal to the DPRK to listen to the fact that there is no alternative to stopping the different provocations and to return to dialogue, South Korean ambassador Kwon Inchul said. We have never threatened the DPRK with any attacks and we have never promoted the use of force.
```

*****RESULTS*****:

Logistic Regression Prediction: Not A Fake News(Verified News)

Decision Tree Prediction: Not A Fake News(Verified News)

```
In [44]: news = str(input())
manual_testing(news)

21st Century Wire says This week, the historic international Iranian Nuclear Deal was punctuated by a two-way prisoner swap between Washington and Tehran, but on the U.S. side all 7 of the Iranians held in U.S. prisons DID NOT show up to their flight to Geneva for the prisoner exchange with at least 3 electing to stay in the U.S. TEHRAN SIDE: In Iran, 5 U.S. prisoners were released, with 4 of them making their way to Germany via Switzerland.Will Robinson Daily MailNone of the Iranians freed in the prisoner swap have returned home and could still be in the United States, it has been reported. The seven former inmates, who were released as part of a deal with the Islamic republic, did not show up to get a flight to Geneva, Switzerland, where the exchange was set to take place on Sunday.Three of the Iranians have decided to stay in the United States, ABC reported, with some moving in with their families. However it is not known where the other four are. Three of the Americans who had been detained in Iran - Washington Post journalist Jason Rezaian, former U.S. Marine Amir Hekmati and Christian pastor Scott Abbeduto - left Tehran at around 7am the same day, but weren't met by their counterparts in Switzerland Continue this story at the Mail OnlineREAD MORE IRAN NEWS AT: 21st Century Wire Iran Files
```

*****RESULTS*****:

Logistic Regression Prediction: Fake News!!!!!!

Decision Tree Prediction: Fake News!!!!!!

```
In [45]: news = str(input())
manual_testing(news)

LONDON (Reuters) - LexisNexis, a provider of legal, regulatory and business information, said on Tuesday it had withdrawn two products from the Chinese market in March this year after it was asked to remove some content. The issue of academic freedom in China hit the headlines this week after the leading British academic publisher, Cambridge University Press, said it had complied with a request to block online access to some scholarly articles in a statement. It later reversed its position. Earlier this year LexisNexis Business Insight Solutions in China was asked to remove some content from its database. LexisNexis said in a statement. In March 2017, the company withdrew two products (Nexis and LexisNexis Academic) from the Chinese market. LexisNexis is owned by information group Reix.
```

*****RESULTS*****:

Logistic Regression Prediction: Not A Fake News(Verified News)

Decision Tree Prediction: Not A Fake News(Verified News)

```
In [46]: news = str(input())
manual_testing(news)

21st Century Wire says It's a familiar theme. Whenever there is a dispute or a change of law, and two tribes go to war, there is normally only one real winner after the tribulation the lawyers. Ars Technica late 2013, Yahoo was hit with six lawsuits over its practice of using automated scans of e-mail to produce targeted ads. The cases, which were consolidated in federal court, all argued that the privacy rights of non-Yahoo users, who did not consent to Yahoo's interception and scanning of their emails, were being violated by a multi-billion dollar company.Now, lawyers representing the plaintiffs are getting a different tune. Last week, they asked US District Judge Lucy Koh to accept a proposed settlement (PTC). Under the proposal, the massive class of non-Yahoo users won't get any payment, but the class lawyers at Girard Gibbs and Kaplan Fox intend to ask for up to $4 million in fees. The ultimate amount of fees will be up to the judge, but Yahoo has agreed not to oppose any fee request up to $4 million.While users won't get any payment, Yahoo will change how it handles user e-mails but it isn't the change that the plaintiffs attorneys were originally asking for. Yahoo won't stop scanning e-mails. Instead, the company has agreed to make a technical change to when it scans e-mails. In the settlement (PDF), Yahoo has agreed that e-mail content will be only sent to servers for advertising purposes after a technical audit. Yahoo had been asked to remove some content from its database. LexisNexis said in a statement. In March 2017, the company withdrew two products (Nexis and LexisNexis Academic) from the Chinese market. LexisNexis is owned by information group Reix.
```

*****RESULTS*****:

Logistic Regression Prediction: Fake News!!!!!!

Decision Tree Prediction: Fake News!!!!!!

```
In [ ]: 
```