

```
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
```

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

```
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

Next steps: [Generate code with data_fake](#) [View recommended plots](#) [New interactive sheet](#)

```
data_true.head()
```

</

Next steps: [Generate code with data_true](#) [View recommended plots](#) [New interactive sheet](#)

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

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

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

```
data_fake_manual_testing = data_fake.tail(10)
data_fake = data_fake[:-10]
```

```
data_true_manual_testing = data_true.tail(10)
data_true = data_true[:-10]
```


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

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

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

```
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-ac...	21st Century Wire says It s a familiar theme. ...	Middle-east	January 16, 2016	0	
23478	Sunnistan: US and Allied 'Safe Zone' Plan to T...	Patrick Henningsen 21st Century WireRemember ...	Middle-east	January 15, 2016	0	
23479	How to Blow \$700 Million: Al Jazeera America F...	21st Century Wire says Al Jazeera America will...	Middle-east	January 14, 2016	0	
23480	10 U.S. Navy Sailors Held by Iranian Military ...	21st Century Wire says As 21WIRE predicted in ...	Middle-east	January 12, 2016	0	



```
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	
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	
21414	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disused Sov...	worldnews	August 22, 2017	1	
21415	Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017	1	
21416	Indonesia to buy \$1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	August 22, 2017	1	

```
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	
7	Trump Said Some INSANELY Racist Stuff Inside ...	In the wake of yet another court decision that...	News	December 23, 2017	0	
8	Former CIA Director Slams Trump Over UN Bully...	Many people have raised the alarm regarding th...	News	December 22, 2017	0	
9	WATCH: Brand-New Pro-Trump Ad Features So Muc...	Just when you might have thought we d get a br...	News	December 21, 2017	0	

Next steps:

[Generate code with data_merge](#)

[View recommended plots](#)

[New interactive sheet](#)

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

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

data.isnull().sum()
```

```
0
text 0
class 0
```

data = data.sample(frac = 1)

data.head()

```

text  class
6007  PHILADELPHIA (Reuters) - U.S. Republican leade...  1
10229  The Washington Post Nearly a third of territ...  0
12581  (Reuters) - Reuters photographers witnessed th...  1
5172   Senator Kelly Ayotte (R-NH) is fighting hard t...  0
18419  BAGHDAD (Reuters) - Iraq stopped selling dolla...  1
```

Next steps: [Generate code with data](#) [View recommended plots](#) [New interactive sheet](#)

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

data.columns

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

data.head()

```

text  class
0  PHILADELPHIA (Reuters) - U.S. Republican leade...  1
1  The Washington Post Nearly a third of territ...  0
2  (Reuters) - Reuters photographers witnessed th...  1
3  Senator Kelly Ayotte (R-NH) is fighting hard t...  0
4  BAGHDAD (Reuters) - Iraq stopped selling dolla...  1
```

Next steps: [Generate code with data](#) [View recommended plots](#) [New interactive sheet](#)

```
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
```

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

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

```
x_train, x_text, 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)
```

```
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)
```

```
0.985650623885918
```

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

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

	precision	recall	f1-score	support
0	0.99	0.98	0.99	5984
1	0.98	0.99	0.98	5236
accuracy			0.99	11220
macro avg	0.99	0.99	0.99	11220
weighted avg	0.99	0.99	0.99	11220

```
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
```

	precision	recall	f1-score	support
0	0.99	1.00	1.00	5984
1	1.00	0.99	1.00	5236
accuracy			1.00	11220
macro avg	1.00	1.00	1.00	11220
weighted avg	1.00	1.00	1.00	11220

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

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

```
GradientBoostingClassifier
```

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

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

0.9950089126559715

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

	precision	recall	f1-score	support
0	1.00	0.99	1.00	5984
1	0.99	1.00	0.99	5236
accuracy			1.00	11220
macro avg	0.99	1.00	0.99	11220
weighted avg	1.00	1.00	1.00	11220

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

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

RandomForestClassifier

RandomForestClassifier(random_state=0)

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

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

0.9905525846702318

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

	precision	recall	f1-score	support
0	0.99	0.99	0.99	5984
1	0.99	0.99	0.99	5236
accuracy			0.99	11220
macro avg	0.99	0.99	0.99	11220
weighted avg	0.99	0.99	0.99	11220

```
#Model Testing
```

```
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_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())
manual_testing(news)
```

NEW DELHI, July 14 (Reuters) - A preliminary investigation into the crash of an Air India passenger jet last month that killed 260 people

LR Prediction: Not A Fake News
 DT Prediction: Not A Fake News
 GB Prediction: Not A Fake News

RF Prediction: Not A Fake News

Could not connect to the reCAPTCHA service. Please check your internet connection and reload to get a reCAPTCHA challenge.