

```
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('/content/Fake.csv')
data_true = pd.read_csv('/content/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	
-	Trump Is So Obsessed He Even Has	On Christmas dav. Donald Trump	..	December 29.	

```
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	WASHINGTON (Reuters) - Trump	December 30.	

```
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)
for i in range(23480, 23470,-1):
    data_fake.drop([i], axis = 0, inplace = True)
```

```
data_true_manual_testing = data_true.tail(10)
for i in range(21416, 21406,-1):
    data_true.drop([i], axis = 0, inplace = True)
```

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

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

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

```
<ipython-input-55-90008d39c97b>: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-
data_fake_manual_testing['class'] = 0
<ipython-input-55-90008d39c97b>: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-
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 ...	21st Century Wire says It s a familiar theme	Middle-east	January 16, 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	

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 Leaves Out A	Donald Trump spent a good portion of his day ...	News	December 23, 2017	0	

data_merge.columns

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

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

data.isnull().sum()

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

data = data.sample(frac = 1)

data.head()

	text	class	
13725	Sorry Bernie, but it s all about Obama s third...	0	
13600	Is there a law against waterboarding Republica...	0	
6801	Big news out of the Treasury Department: After...	0	
19326	In mid-October 2016, James O Keefe exposed man...	0	
15018	The more they hammer the GOP frontrunners, the...	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	Sorry Bernie, but it s all about Obama s third...	0	
1	Is there a law against waterboarding Republica...	0	
2	Big news out of the Treasury Department: After...	0	
3	In mid-October 2016, James O Keefe exposed man...	0	
4	The more they hammer the GOP frontrunners, the...	0	

```
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('\w*\d\w*', '', text)
    return text
```

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

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

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

```
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

LogisticRegression()

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

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

0.9871657754010695

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

           precision    recall  f1-score   support

0               0.99         0.99         0.99         5764
```

	1	0.99	0.99	0.99	5456
accuracy				0.99	11220
macro avg	0.99	0.99	0.99	0.99	11220
weighted avg	0.99	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)
```

```
0.996524064171123
```

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

	precision	recall	f1-score	support
0	0.99	1.00	1.00	5764
1	1.00	0.99	1.00	5456
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)
```

```
0.9953654188948307
```

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

	precision	recall	f1-score	support
0	1.00	0.99	1.00	5764
1	0.99	1.00	1.00	5456
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 RandomForestClassifier
```

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

```
RandomForestClassifier
```

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

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

```
0.9897504456327986
```

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

	precision	recall	f1-score	support
0	0.99	0.99	0.99	5764
1	0.99	0.99	0.99	5456
accuracy			0.99	11220
macro avg	0.99	0.99	0.99	11220
weighted avg	0.99	0.99	0.99	11220

[Colab paid products](#) - [Cancel contracts here](#)

✓ 0s completed at 10:42 PM

