

```
[1]: import pandas as pd
import numpy as nm
from sklearn.model_selection import train_test_split as train_test_split
from sklearn.metrics import classification_report
import re
import string
import matplotlib.pyplot as plt
```

```
[2]: data_true=pd.read_csv("/content/drive/MyDrive/True.csv")
data_fake=pd.read_csv("/content/drive/MyDrive/Fake.csv")
```

```
[3]: 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", force_remount=True).

```
[4]: data_true.head(5)
```

```
[4]:
```

	title \	text	subject \	date
0	As U.S. budget fight looms, Republicans flip t...			December 31, 2017
1	U.S. military to accept transgender recruits o...		politicsNews	December 29, 2017
2	Senior U.S. Republican senator: 'Let Mr. Muell...		politicsNews	December 31, 2017
3	FBI Russia probe helped by Australian diplomat...		politicsNews	December 30, 2017
4	Trump wants Postal Service to charge 'much mor...		politicsNews	

4 December 29, 2017

```
[5]: data_fake.head(5)
```

```
[5]:
```

	title \	text	subject \
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...		

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

```
[6]: data_true.shape,data_fake.shape
```

```
[6]: ((21417, 4), (23481, 4))
```

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

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

```
[9]: data_manual_testing = pd.  
    concat([data_fake_manual_testing,data_true_manual_testing])  
data_manual_testing.to_csv("manual_testing.csv")
```

```
[10]: data_merge = pd.concat([data_fake,data_true],axis=0)  
data_merge.head(10)#displaying first 10
```

```
[10]:
```

	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...
5	Racist Alabama Cops Brutalize Black Boy While...
6	Fresh Off The Golf Course, Trump Lashes Out A...
7	Trump Said Some INSANELY Racist Stuff Inside ...
8	Former CIA Director Slams Trump Over UN Bully...
9	WATCH: Brand-New Pro-Trump Ad Features So Muc...

	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
5	The number of cases of cops brutalizing and ki...	News
6	Donald Trump spent a good portion of his day a...	News
7	In the wake of yet another court decision that...	News
8	Many people have raised the alarm regarding th...	News
9	Just when you might have thought we d get a br...	News

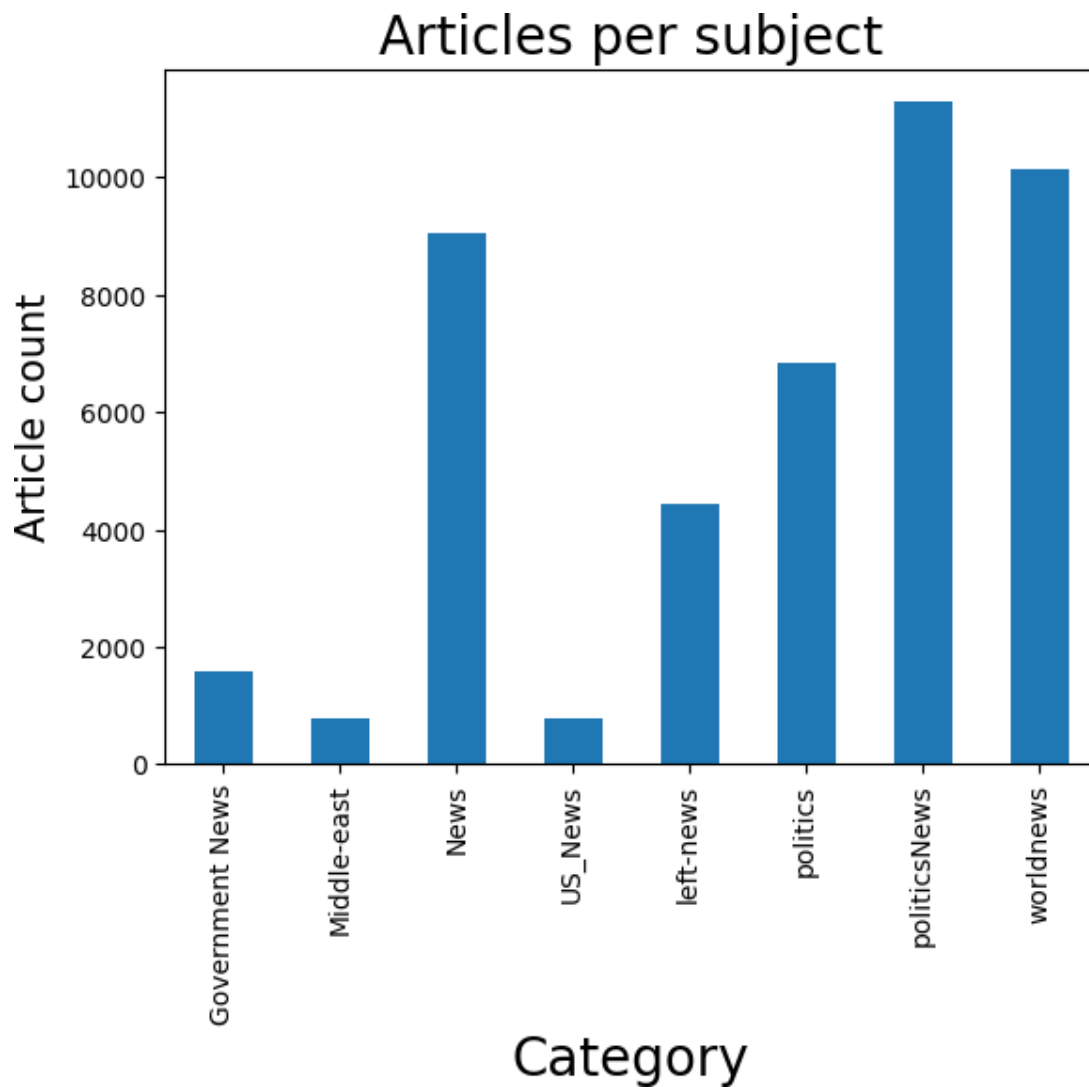
	date	class
0	December 31, 2017	0
1	December 31, 2017	0
2	December 30, 2017	0
3	December 29, 2017	0
4	December 25, 2017	0
5	December 25, 2017	0
6	December 23, 2017	0
7	December 23, 2017	0
8	December 22, 2017	0
9	December 21, 2017	0

```
[11]: print(data_merge.groupby(["subject"])[["text"]].count())
data_merge.groupby(["subject"])[["text"]].count().plot(kind="bar")
plt.title("Articles per subject",size=20)
plt.xlabel("Category",size=20)
plt.ylabel("Article count",size=15)
plt.show()
```

subject	
Government News	1570
Middle-east	778
News	9050

US_News	783
left-news	4449
politics	6841
politicsNews	11272
worldnews	10135

Name: text, dtype: int64



```
[12]: print(data_merge.groupby(["class"])["text"].count())
print("0 = Fake news\n1 = True news")
data_merge.groupby(["class"])["text"].count().plot(kind="pie")
plt.title("Fake news and true news", size=20)
plt.show
```

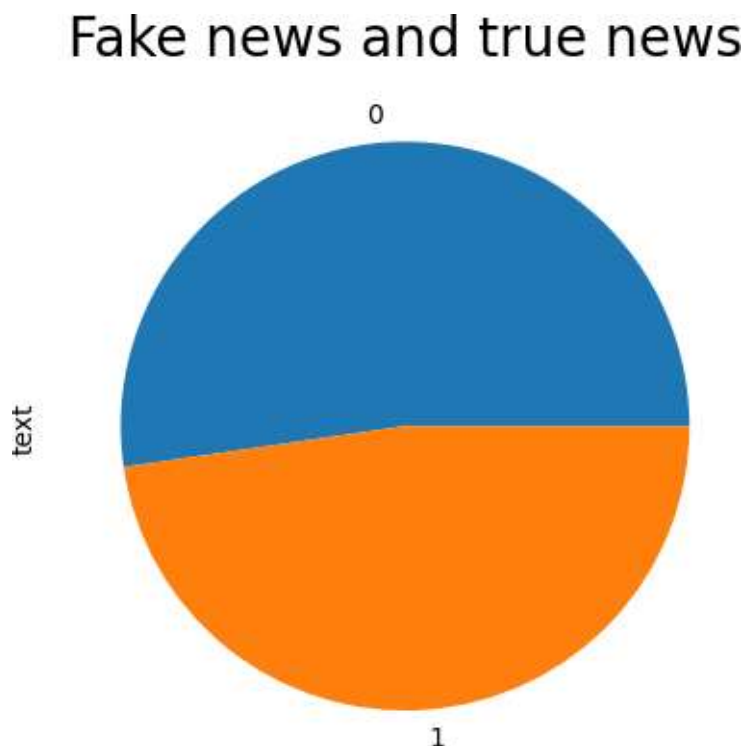
class

```

0    23471
1    21407
Name: text, dtype: int64
0 = Fake news
1 = True news

```

```
[12]: <function matplotlib.pyplot.show(close=None, block=None)>
```



```
[13]: data = data_merge.drop(["title", "subject", "date"], axis=1)
data.head(10)
```

```
[13]:
```

	text	class
0	Donald Trump just couldn t wish all Americans ...	0
1	House Intelligence Committee Chairman Devin Nu...	0
2	On Friday, it was revealed that former Milwauk...	0
3	On Christmas day, Donald Trump announced that ...	0
4	Pope Francis used his annual Christmas Day mes...	0
5	The number of cases of cops brutalizing and ki...	0
6	Donald Trump spent a good portion of his day a...	0
7	In the wake of yet another court decision that...	0
8	Many people have raised the alarm regarding th...	0
9	Just when you might have thought we d get a br...	0

```
[14]: data=data.sample(frac=1)
data.head(10)
```

```
[14]:
```

	text	class
20450	QUETTA, Pakistan (Reuters) - Gunmen in southwe...	1
16518	STRASBOURG (Reuters) - The European Parliament...	1
5316	(This story corrects paragraph 8 to clarify S...	1
21299	BEIJING (Reuters) - Venezuela s close ally Chi...	1
18852	BAMAKO (Reuters) - Mali s President Ibrahim Bo...	1
7277	This story is the perfect example of just how ...	0
9793	WASHINGTON (Reuters) - Democrats on a powerful...	1
20163	(Reuters) - Csx Corp said on Wednesday it has ...	1
9149	WASHINGTON (Reuters) - President Barack Obama ...	1
793	Trump s brand-new communications director, Ant...	0

```
[15]: data.isnull().sum()
```

```
[15]: text      0
class      0
dtype: int64
```

```
[16]: def filtering(data):
text=data.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
```

```
[39]: data["text"]= data["text"].apply(filtering)
data.head(5)
```

```
[39]:
```

	text	class
20450		1
16518		1
5316		1
21299	beijingreutersvenezuelascloseallychinasaidonmo...	1
18852		1

```
[18]: x=data["text"]
y=data["class"]
```

```
[19]: !pip install scikit-learn
from sklearn.model_selection import train_test_split
```

Requirement already satisfied: scikit-learn in /usr/local/lib/python3.10/dist-packages (1.2.2)
 Requirement already satisfied: numpy>=1.17.3 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.25.2)
 Requirement already satisfied: scipy>=1.3.2 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.11.4)
 Requirement already satisfied: joblib>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.4.2)
 Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (3.5.0)

```
[20]: x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25,
↳ random_state=0)
```

```
[21]: from sklearn.feature_extraction.text import TfidfVectorizer

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

```
[22]: from sklearn.linear_model import LogisticRegression

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

```
[22]: LogisticRegression()
```

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

LR.score(xv_test, y_test)

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

	precision	recall	f1-score	support
0	0.52	1.00	0.69	5867
1	1.00	0.01	0.01	5353
accuracy			0.53	11220
macro avg	0.76	0.50	0.35	11220
weighted avg	0.75	0.53	0.37	11220

```
[24]: from sklearn.tree import DecisionTreeClassifier

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

```
DecisionTreeClassifier()
DecisionTreeClassifier()
pred_dt = DT.predict(xv_test)
DT.score(xv_test, y_test)
```

[24]: 0.5260249554367201

[25]: `print (classification_report(y_test, pred_lr))`

	precision	recall	f1-score	support
0	0.52	1.00	0.69	5867
1	1.00	0.01	0.01	5353
accuracy			0.53	11220
macro avg	0.76	0.50	0.35	11220
weighted avg	0.75	0.53	0.37	11220

[26]: `def predict_news(text):`
`text_vectorized = vectorization.transform([filtering(text)])`
`prediction = DT.predict(text_vectorized)`
`if prediction == 1:`
`return "This news is likely true."`
`else:`
`return "This news is likely fake."`

[37]: `user_input = input("Enter news text: ")`
`result = predict_news(user_input)`
`print(result)`

Enter news text: beijingreutersvenezuelascloseallychinasaidonmo...
This news is fake.

[28]: `from sklearn.ensemble import RandomForestClassifier`

[29]: `RF = RandomForestClassifier(random_state = 0)`
`RF.fit(xv_train, y_train)`
`RandomForestClassifier(random_state=0)`
`RandomForestClassifier(random_state=0)`
`RandomForestClassifier(random_state=0)`
`pred_rf = RF.predict(xv_test)`
`RF.score(xv_test, y_test)`

[29]: 0.5260249554367201

[30]: `print (classification_report(y_test, pred_rf))`

	precision	recall	f1-score	support
0	0.52	1.00	0.69	5867
1	1.00	0.01	0.01	5353
accuracy			0.53	11220
macro avg	0.76	0.50	0.35	11220
weighted avg	0.75	0.53	0.37	11220

```
[34]: def predict_news(text):
      text_vectorized = vectorization.transform([filtering(text)])
      prediction = RF.predict(text_vectorized)
      if prediction == 1:
          return "This news is true."
      else:
          return "This news is fake."
```

```
[40]: user_input = input("Enter news text: ")
      result = predict_news(user_input)
      print(result)
```

Enter news text: beijingreutersvenezuelascloseallychinasaidonmo...
This news is fake.

```
[50]: def output_lable(n):
      if n==0:
          return "Fake News"
      elif n==1:
          return "Not A Fake News"
```

```
[58]: 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: {} \nGBC Prediction: {}
      \nRFC Prediction: {}".format(output_lable(pred_LR[0]),
                                     output_lable(pred_DT[0]),
```

```
↪ output_label(pred_GB[0]),  
↪ output_label(pred_RF[0]))
```

```
[59]: def wordopt(text):  
      text = text.lower()  
      text = re.sub('[.*?\\]', '', text)  
      # Define processed_text here with the result of your text processing  
      processed_text = text  
      return processed_text
```

```
[61]: from sklearn.ensemble import GradientBoostingClassifier  
      GB = GradientBoostingClassifier()  
      GB.fit(xv_train, y_train)  
  
      news = str(input())  
      manual_testing(news)
```

WASHINGTON (Reuters) – The head of a conservative Republican faction in the U.S.

LR Prediction: Fake News
DT Prediction: Fake News
GBC Prediction: Fake News
RFC Prediction: Fake News