

fake-data

June 27, 2024

```
[2]: import pandas as pd
import numpy as nm
from sklearn.model_selection import train_test_split as ttp
from sklearn.metrics import classification_report
import re
import string
import matplotlib.pyplot as plt
data_true=pd.read_csv("/content/drive/MyDrive/True.csv")
data_fake=pd.read_csv("/content/drive/MyDrive/Fake.csv")
data_true.head(5)
```

```
[2]:                                     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
```

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

```
[3]:                                     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

```

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

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

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

```
[15]: data_true_manual_testing=data_true.tail(10)
      # Drop the last 10 indices from data_true
      data_true.drop(data_true.tail(10).index, inplace=True)
      data_fake_manual_testing=data_fake.tail(10) # Removed extra space at_
      ↪the_beginning of this line
      # Drop the last 10 indices from data_fake
      data_fake.drop(data_fake.tail(10).index, inplace=True)
      data_manual_testing=pd.
      ↪concat([data_fake_manual_testing,data_true_manual_testing],axis=0)
      data_manual_testing.to_csv("manual_testing.csv") # Removed extra space at_
      ↪the_beginning of this line
      data_merge=pd.concat([data_true,data_fake],axis=0)
      data_merge.head(10)
```

```
[15]:
                                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...
5 White House, Congress prepare for talks on spe...
6 Trump says Russia probe will be fair, but time...
```

```

7 Factbox: Trump on Twitter (Dec 29) - Approval ...
8 Trump on Twitter (Dec 28) - Global Warming
9 Alabama official to certify Senator-elect Jone...

```

```

                                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
5 WEST PALM BEACH, Fla./WASHINGTON (Reuters) - T... politicsNews
6 WEST PALM BEACH, Fla (Reuters) - President Don... politicsNews
7 The following statements were posted to the ve... politicsNews
8 The following statements were posted to the ve... politicsNews
9 WASHINGTON (Reuters) - Alabama Secretary of St... politicsNews

```

```

                                date  class
0 December 31, 2017                1
1 December 29, 2017                1
2 December 31, 2017                1
3 December 30, 2017                1
4 December 29, 2017                1
5 December 29, 2017                1
6 December 29, 2017                1
7 December 29, 2017                1
8 December 29, 2017                1
9 December 28, 2017                1

```

```

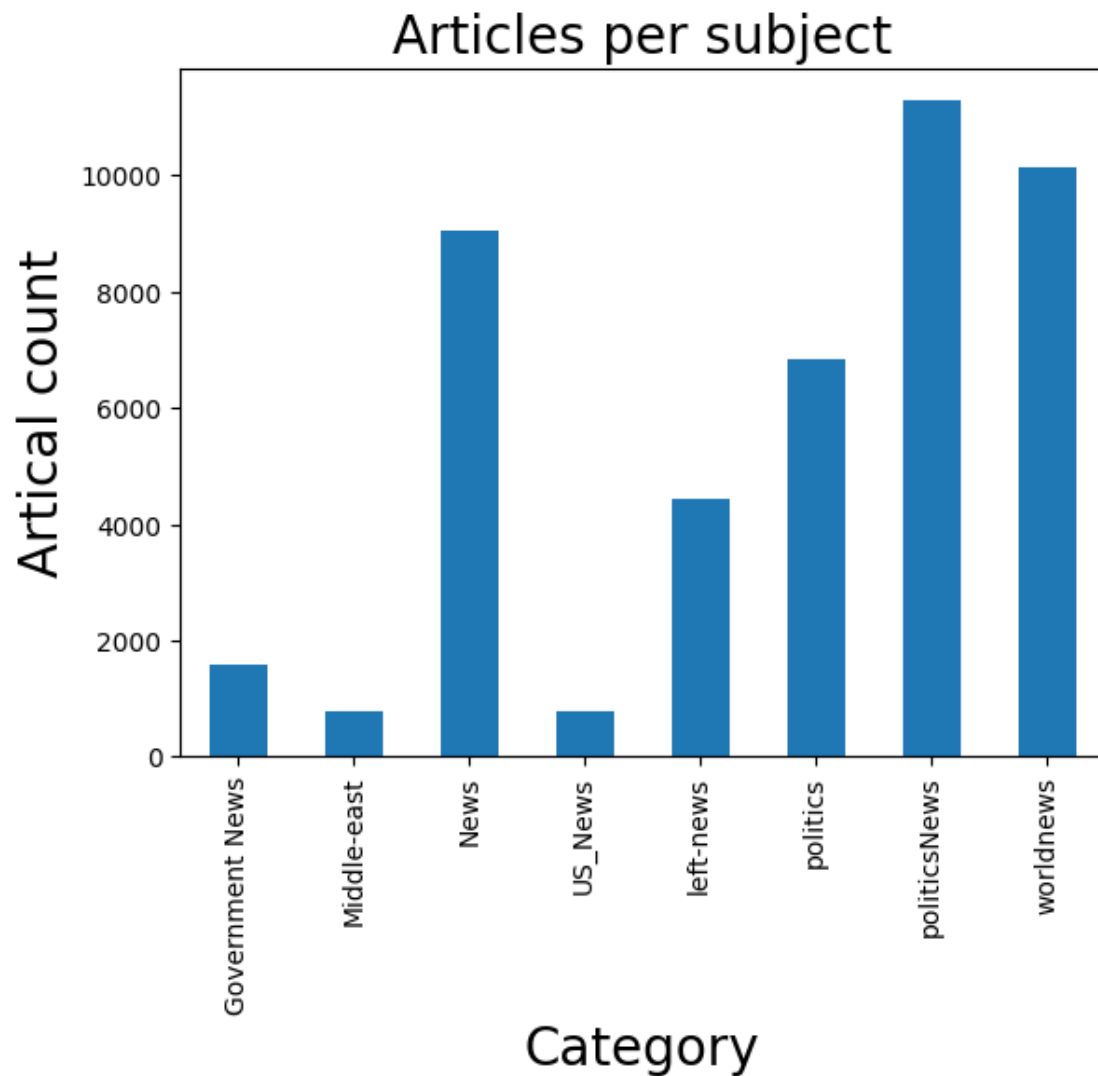
[16]: print(data_merge.groupby(['subject'])['text'].count())
      data_merge.groupby(['subject'])['text'].count().plot(kind="bar")
      plt.title("Articles per subject",size=20)
      plt.xlabel("Category",size=20)
      plt.ylabel("Artical count",size=20)
      plt.show()

```

```

subject
Government News    1570
Middle-east        768
News               9050
US_News            783
left-news          4449
politics           6841
politicsNews       11272
worldnews          10125
Name: text, dtype: int64

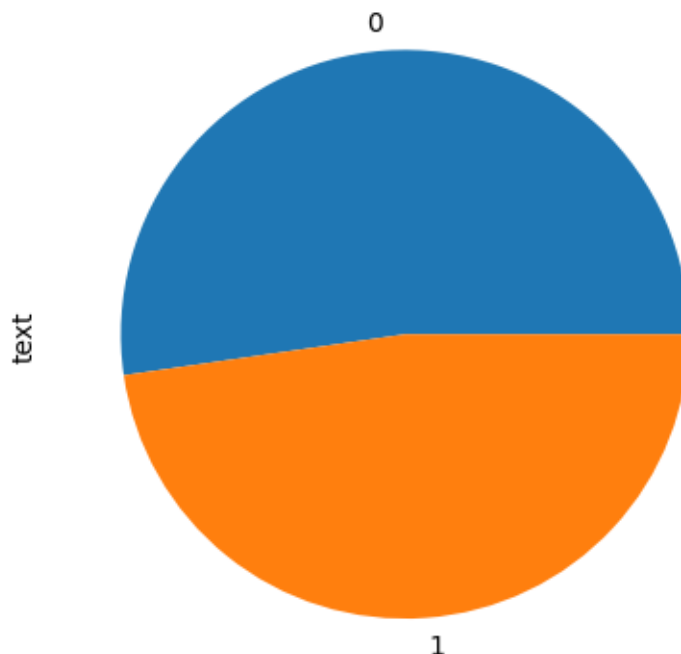
```



```
[17]: 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 vs True News",size=20)
plt.show()
```

```
class
0    23461
1    21397
Name: text, dtype: int64
0 = Fake News
1 = True News
```

Fake News vs True News



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

```
[18]:
```

	text	class
0	WASHINGTON (Reuters) - The head of a conservat...	1
1	WASHINGTON (Reuters) - Transgender people will...	1
2	WASHINGTON (Reuters) - The special counsel inv...	1
3	WASHINGTON (Reuters) - Trump campaign adviser ...	1
4	SEATTLE/WASHINGTON (Reuters) - President Donal...	1
5	WEST PALM BEACH, Fla./WASHINGTON (Reuters) - T...	1
6	WEST PALM BEACH, Fla (Reuters) - President Don...	1
7	The following statements were posted to the ve...	1
8	The following statements were posted to the ve...	1
9	WASHINGTON (Reuters) - Alabama Secretary of St...	1

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

```
[19]:
```

	text	class
3877	MILAN (Reuters) - Former U.S. President Barack...	1
16720	Everyone needs to read this letter (click HERE...	0
5226	WASHINGTON (Reuters) - Billionaire investor Wi...	1

20646	SINGAPORE (Reuters) - Singapore should end har...	1
10650	WASHINGTON (Reuters) - U.S. Senator John McCai...	1
710	WASHINGTON (Reuters) - The National Federation...	1
14463	HARARE (Reuters) - Crowds of Zimbabweans booed...	1
5818	Donald Trump s presumptive nomination just cos...	0
11049	DES MOINES, Iowa (Reuters) - Donald Trump shun...	1
3548	With Donald Trump winning the election, albeit...	0

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

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

```
[22]: 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('\n', '', text)
      text=re.sub('\w*\d\w*', '', text)
      return text
```

```
[23]: data['text']=data['text'].apply(filtering)
      data.head(10)
```

```
[23]:
```

	text	class
3877	milan reuters former u s president barack...	1
16720	everyone needs to read this letter click here...	0
5226	washington reuters billionaire investor wi...	1
20646	singapore reuters singapore should end har...	1
10650	washington reuters u s senator john mccai...	1
710	washington reuters the national federation...	1
14463	harare reuters crowds of zimbabweans booed...	1
5818	donald trump s presumptive nomination just cos...	0
11049	des moines iowa reuters donald trump shun...	1
3548	with donald trump winning the election albeit...	0

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

```
[25]: x_train,x_test,y_train,y_test=ttp(x,y,test_size=0.25,random_state=0)
```

```
[26]: from sklearn.feature_extraction.text import TfidfVectorizer
      vectorization = TfidfVectorizer()
```

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

```
[27]: from sklearn.linear_model import LogisticRegression
LR = LogisticRegression()
LR.fit(xv_train,y_train)
```

```
[27]: LogisticRegression()
```

```
[30]: def predict_news(text):
        text_vectorized = vectorization.transform([filtering(text)])
        prediction = LR.predict(text_vectorized)
        if prediction == 1:
            return "This news is likely true."
        else:
            return "This news is likely fake."
user_input = input("Enter news text: ")
result = predict_news(user_input)
print(result)
```

Enter news text: Modi is president

This news is likely fake.

```
[31]: LogisticRegression()
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	0.99	0.98	0.99	5844
1	0.98	0.99	0.98	5371
accuracy			0.99	11215
macro avg	0.99	0.99	0.99	11215
weighted avg	0.99	0.99	0.99	11215

```
[32]: 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)
0.996524064171123
```

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

	precision	recall	f1-score	support
0	0.99	0.98	0.99	5844
1	0.98	0.99	0.98	5371
accuracy			0.99	11215
macro avg	0.99	0.99	0.99	11215
weighted avg	0.99	0.99	0.99	11215

```
[34]: 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."
user_input = input("Enter news text: ")
result = predict_news(user_input)
print(result)
```

Enter news text: doll is intelligent
This news is likely fake.

```
[35]: from sklearn.ensemble import RandomForestClassifier
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)
0.989572192513369
print (classification_report(y_test, pred_rf))
```

	precision	recall	f1-score	support
0	0.99	0.99	0.99	5844
1	0.99	0.99	0.99	5371
accuracy			0.99	11215
macro avg	0.99	0.99	0.99	11215
weighted avg	0.99	0.99	0.99	11215


```
[37]: def predict_news(text):  
        text_vectorized = vectorization.transform([filtering(text)])  
        prediction = RF.predict(text_vectorized)  
        if prediction == 1:  
            return "This news is likely true."  
        else:  
            return "This news is likely fake."  
user_input = input("Enter news text: ")  
result = predict_news(user_input)  
print(result)
```

Enter news text: humans are not intelligent
This news is likely fake.