

day3-1

June 27, 2024

```
[1]: 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
```

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

```
[50]: #from google.colab import drive
#drive.mount('/content/drive')
```

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

```
[3]:                                     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
```

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

```
[4]:                                     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
```

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

```
[5]: ((21417, 4), (23481, 4))
```

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

```
[7]: 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)
```

```
[8]: data_manual_testing=pd.
     ↪concat([data_fake_manual_testing,data_true_manual_testing],axis=0)
     data_manual_testing.to_csv("manual_testing.csv")
```

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

```
[9]:                                     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

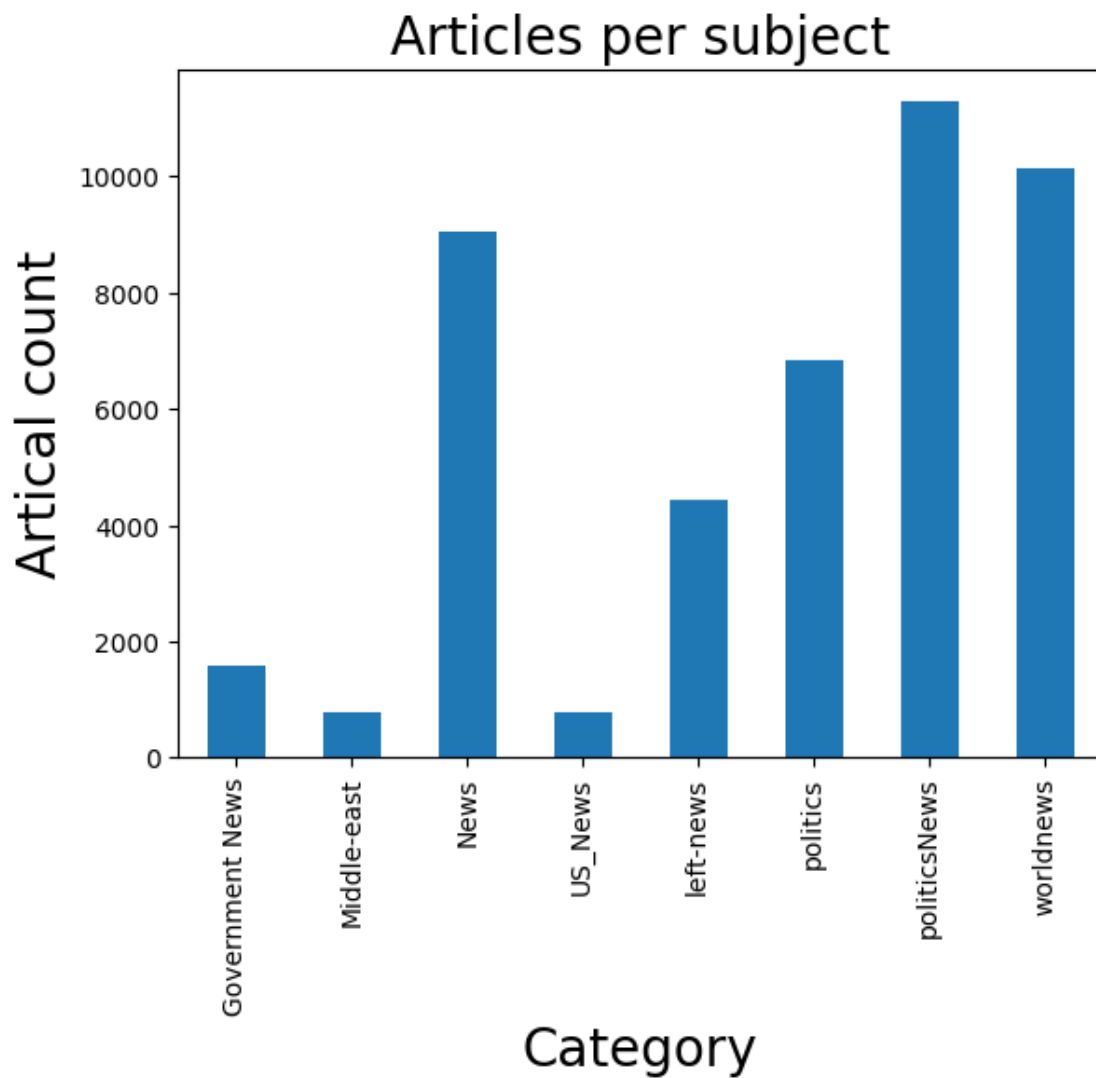
```

[10]: 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	778
News	9050
US_News	783
left-news	4449
politics	6841

```
politicsNews      11272
worldnews         10135
Name: text, dtype: int64
```

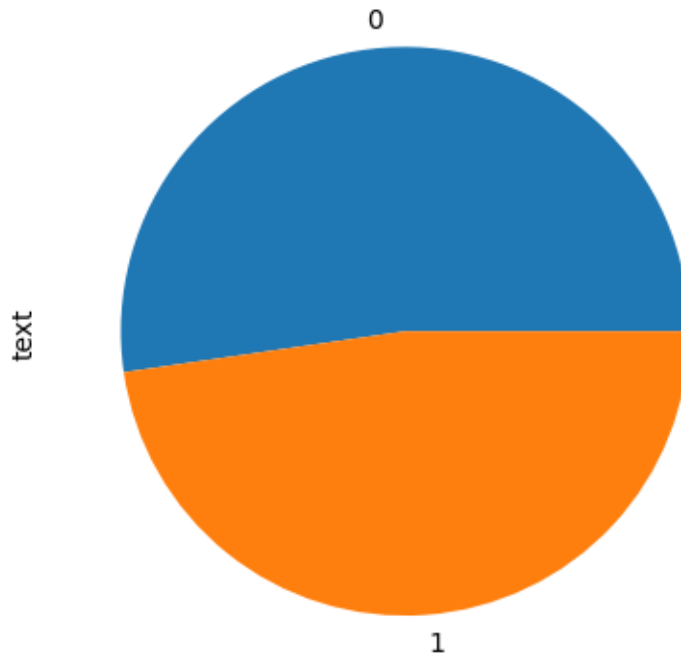


```
[11]: print(data_merge.groupby(['class'])['text'].count())
print("0 = Fake News\n1 = True News")
data_merge.groupby(['class'])['text'].count().plot(kind="pie")
plt.title("Fake News vs True News",size=20)
plt.show()
```

```
class
0    23471
1    21407
Name: text, dtype: int64
```

0 = Fake News
1 = True News

Fake News vs True News



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

```
[12]:
```

	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

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

```
[13]:
```

	text	class
13448	The tsunami has started President Obama s Keny...	0
16670	Let s face it, Obamacare had been a nightmare ...	0
6038	A 16-year-old is dead after a group of white t...	0
3152	House Republicans have long since made clear t...	0
7484	Watching people line up behind Donald Trump is...	0
13924	It s not shocking that Hillary s lying again, ...	0
17579	MOSCOW (Reuters) - If the United States withdr...	1
3852	Let s get something straight here Americans ...	0
12216	MOSCOW (Reuters) - Russia s FSB security servi...	1
5858	After the horrific terrorist attack in Orlando...	0

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

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

```
[15]: 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
```

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

```
[16]:
```

	text	class
13448	the tsunami has started president obama s kenya...	0
16670	let s face it obamacare had been a nightmare ...	0
6038	a year old is dead after a group of white thu...	0
3152	house republicans have long since made clear t...	0
7484	watching people line up behind donald trump is...	0
13924	it s not shocking that hillary s lying again ...	0
17579	moscow reuters if the united states withdr...	1
3852	let s get something straight here americans ...	0
12216	moscow reuters russia s fsb security servi...	1
5858	after the horrific terrorist attack in orlando...	0

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

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

```
[19]: 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)
```

```
[24]: LR.score(xv_test, y_test)
```

```
[24]: 0.9872549019607844
```

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

	precision	recall	f1-score	support
0	0.99	0.99	0.99	5892
1	0.99	0.99	0.99	5328
accuracy			0.99	11220
macro avg	0.99	0.99	0.99	11220
weighted avg	0.99	0.99	0.99	11220

```
[34]: 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: WASHINGTON (Reuters) - The head of a conservative Republican faction in the U.S.

This news is likely true.

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

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

```
[26]: DecisionTreeClassifier()
```

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

```
[28]: DT.score(xv_test, y_test)
```

```
[28]: 0.9955436720142602
```

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

	precision	recall	f1-score	support
0	0.99	0.99	0.99	5892
1	0.99	0.99	0.99	5328
accuracy			0.99	11220
macro avg	0.99	0.99	0.99	11220
weighted avg	0.99	0.99	0.99	11220

```
[ ]: from google.colab import drive  
drive.mount('/content/drive')
```

```
[35]: 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: WASHINGTON (Reuters) - The head of a conservative Republican faction in the U.S.

This news is likely true.

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

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



```
[37]: RandomForestClassifier(random_state=0)
```

```
[38]: pred_rf = RF.predict(xv_test)
```

```
[39]: RF.score(xv_test, y_test)
```

```
[39]: 0.9878787878787879
```

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

	precision	recall	f1-score	support
0	0.99	0.99	0.99	5892
1	0.99	0.98	0.99	5328
accuracy			0.99	11220
macro avg	0.99	0.99	0.99	11220
weighted avg	0.99	0.99	0.99	11220

```
[42]: 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: :WASHINGTON (Reuters) - The head of a conservative Republican faction in the U.S.

This news is likely true.