

fake-news-detection

June 28, 2024

```
[ ]: 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/ML_Team12/1SV21CS003/True.csv")
data_fake=pd.read_csv("/content/drive/MyDrive/ML_Team12/1SV21CS003/Fake.csv")
```

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

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

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

```
[ ]: 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)
```

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

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

```
[ ]:                                     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

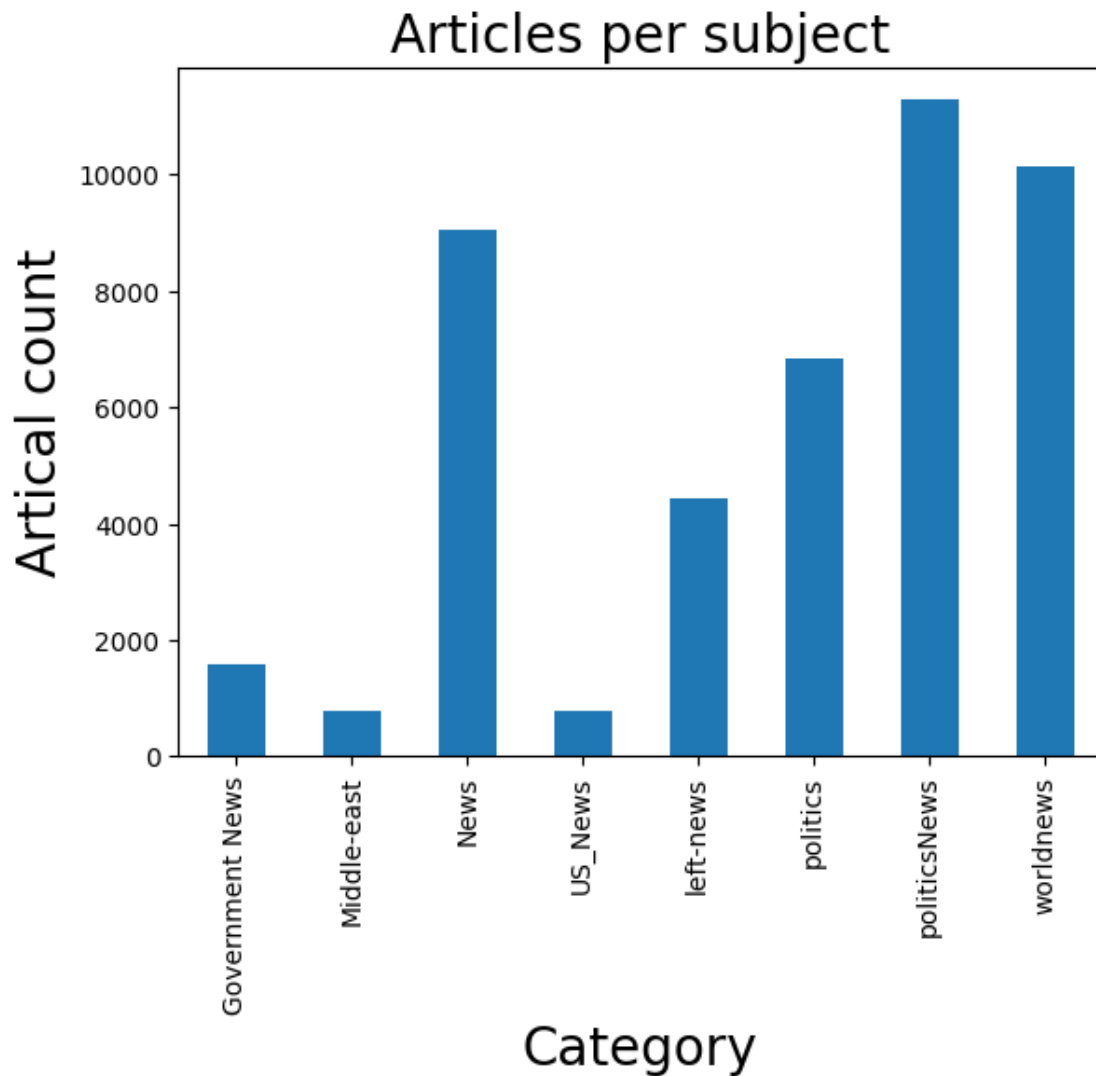
```

[ ]: 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

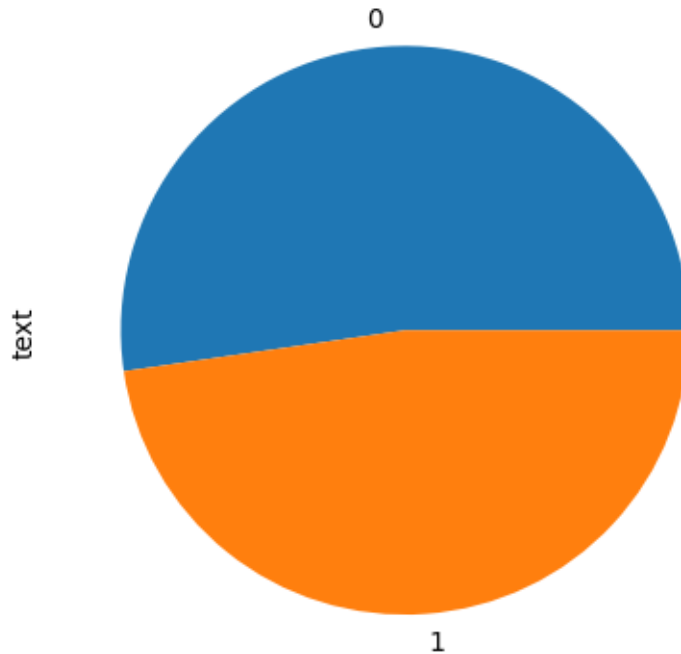
Name: text, dtype: int64



```
[ ]: 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



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

```
[ ]:
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
```

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

```
[ ]:
text  class
21837  Nothing but the best for the queen and the que...    0
13798  HARARE (Reuters) - New Zimbabwean President Em...    1
8201   WASHINGTON (Reuters) - Republican presidential...    1
```

1566	NEW YORK (Reuters) - Former U.S. Representativ...	1
13721	While researching for a wrap-up on the June 7 ...	0
16952	MANILA (Reuters) - Three Russian warships, inc...	1
12028	BERLIN (Reuters) - German Chancellor Angela Me...	1
9428	Trey Gowdy railed against the erosion of trust...	0
23224	21st Century Wire says For many viewers, this ...	0
5894	In the wake of the worst mass shooting in Amer...	0

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

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

```
[ ]: 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
```

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

```
[ ]:
           text      class
12837  nothing but the best for the queen and the que...      0
13798  harare reuters    new zimbabwean president em...      1
8201   washington reuters    republican presidential...      1
1566   new york reuters    former u s representativ...      1
13721  while researching for a wrap up on the june p...      0
16952  manila reuters    three russian warships inc...      1
12028  berlin reuters    german chancellor angela me...      1
9428   trey gowdy railed against the erosion of trust...      0
23224  century wire says for many viewers this came...      0
5894   in the wake of the worst mass shooting in amer...      0
```

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

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

```
[ ]: 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)

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

	precision	recall	f1-score	support
0	0.99	0.98	0.99	5928
1	0.98	0.99	0.98	5292
accuracy			0.99	11220
macro avg	0.98	0.99	0.99	11220
weighted avg	0.99	0.99	0.99	11220

```
[ ]: 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 died
This news is likely fake.
```

```
[ ]: 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)
```

```
[ ]: 0.9962566844919786
```

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

	precision	recall	f1-score	support
0	0.99	0.98	0.99	5928
1	0.98	0.99	0.98	5292
accuracy			0.99	11220
macro avg	0.98	0.99	0.99	11220
weighted avg	0.99	0.99	0.99	11220

```
[ ]: 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: modi is dies  
This news is likely fake.
```

```
[ ]: 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)
```

```
[ ]: 0.9883244206773618
```

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

	precision	recall	f1-score	support
0	0.99	0.99	0.99	5928
1	0.99	0.99	0.99	5292
accuracy			0.99	11220
macro avg	0.99	0.99	0.99	11220
weighted avg	0.99	0.99	0.99	11220

```
[ ]: 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

This news is likely fake.