

fake-news-detection

June 27, 2024

```
[ ]: import pandas as pd
import numpy as np
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
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression, LinearRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier
```

```
[ ]: data_true = pd.read_csv("/content/drive/MyDrive/ML_Team1/1sv21cs001/True.csv")
data_fake = pd.read_csv("/content/drive/MyDrive/ML_Team1/1sv21cs001/Fake.csv")
```

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

```
[ ]:                                     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  label
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

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

```
[ ]:                                     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 label
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
```

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

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

```
[ ]: data_true['label']=1
data_fake['label']=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_true_manual_testing,data_fake_manual_testing],axis=0)
data_manual_testing.to_csv('manual.csv')
```

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

```
[ ]:                                     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  label
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
```

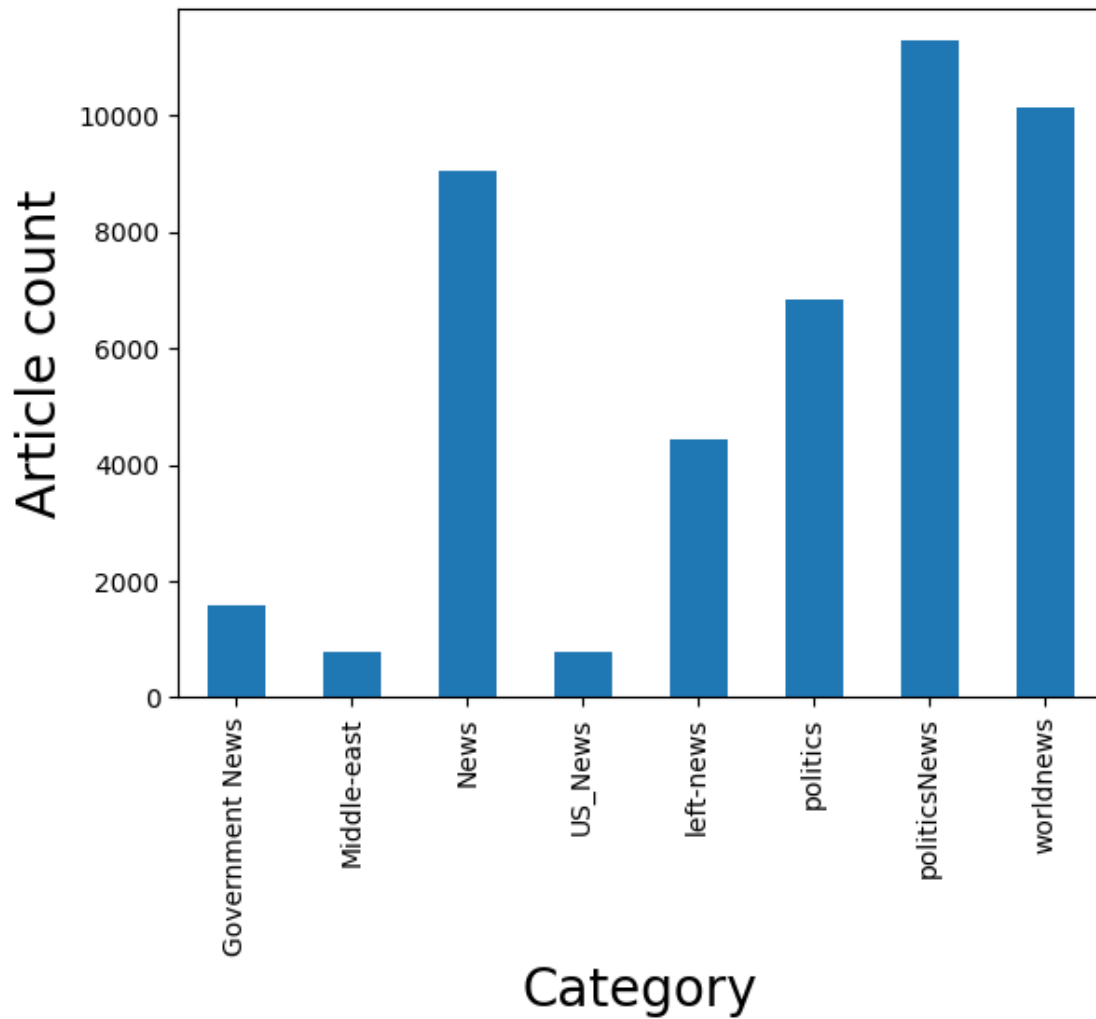
```
[ ]: print(data_merge.groupby(['subject'])['text'].count())
data_merge.groupby(['subject'])['text'].count().plot(kind='bar')
plt.xlabel("Category",size=20)
plt.ylabel("Article count",size=20)
plt.show()
```

```
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(['label'])['text'].count())
      print("0= Fake news\n1 = True news")
      data_merge.groupby(['label'])['text'].count().plot(kind='bar')
      plt.title("Fake news and True News",size=20)
      plt.show()

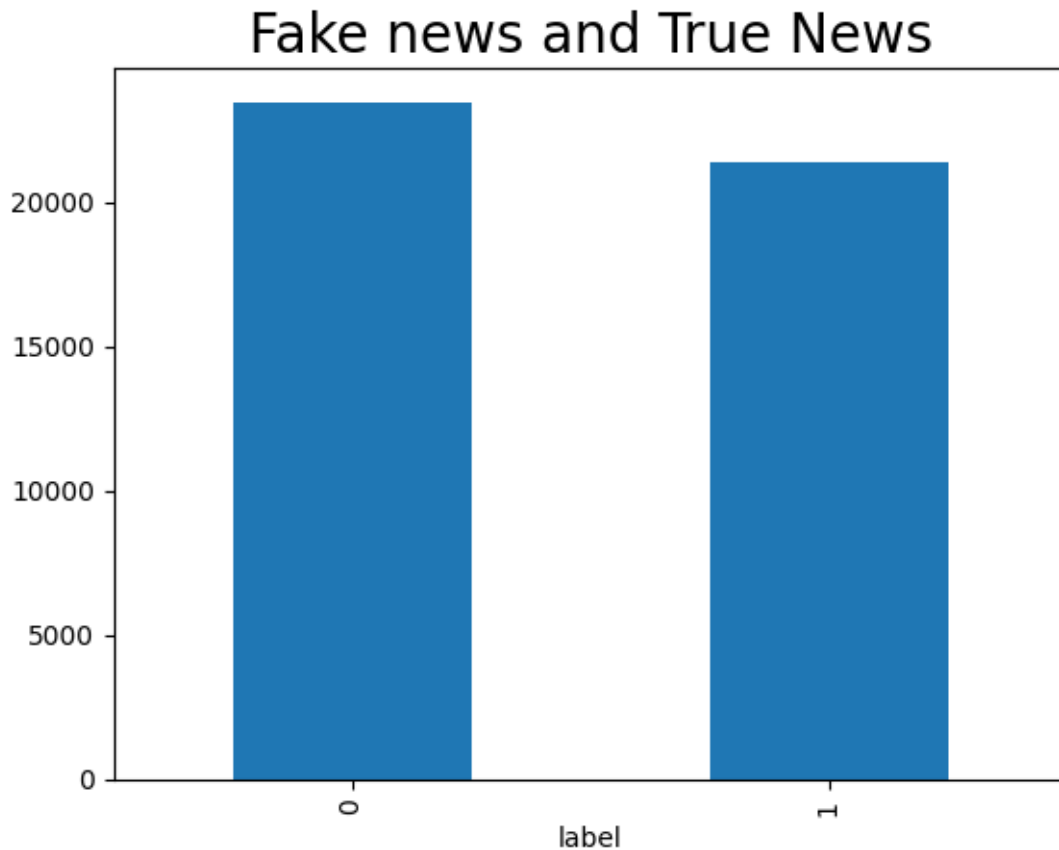
```

```

label
0    23471
1    21407
Name: text, dtype: int64

```

0= Fake news
1 = True news



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

```
[ ]:
      text  label
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  label
18641  A Fresno State University professor was recent...    0
2986   Conservatives are, once again, on the losing s...    0
22012  Our weekly documentary screening curated by th...    0
5300   Donald Trump is known for changing his mind, a...    0
15141  CAIRO (Reuters) - Suspected militants shot dea...    1
19808  Here s what we ve got with the new e-mails and...    0
20550  PARIS (Reuters) - The cost of Hurricane Irma, ...    1
12557  Another untold story of a random act of kindne...    0
12543  Please go to the BEST MULCH Facebook page and ...    0
8743   WASHINGTON (Reuters) - Saudi Arabia's foreign ...    1
```

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

```
[ ]: text      0
      label    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  label
18641  a fresno state university professor was recent...    0
2986   conservatives are  once again  on the losing s...    0
22012  our weekly documentary screening curated by th...    0
5300   donald trump is known for changing his mind  a...    0
15141  cairo reuters      suspected militants shot dea...    1
19808  here s what we ve got with the new e mails and...    0
20550  paris reuters      the cost of hurricane irma  ...    1
12557  another untold story of a random act of kindne...    0
12543  please go to the best mulch facebook page and ...    0
8743   washington reuters      saudi arabia s foreign ...    1
```

```
[ ]: from sklearn.feature_extraction.text import TfidfVectorizer
      from sklearn.linear_model import LogisticRegression
      vectorizer = TfidfVectorizer()
```

```
x = vectorizer.fit_transform(data['text'])
y = data['label']
```

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

```
[ ]: LR = LogisticRegression(max_iter=1000)
LR.fit(x_train, y_train)
```

```
[ ]: LogisticRegression(max_iter=1000)
```

```
[ ]: from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.ensemble import RandomForestClassifier
```

```
vectorizer = TfidfVectorizer()
x = vectorizer.fit_transform(data['text'])
y = data['label']
```

```
clf = RandomForestClassifier()
clf.fit(x, y)
```

```
[ ]: RandomForestClassifier()
```

```
[ ]: # Assuming x_train, y_train are your training data
RF = RandomForestClassifier()
RF.fit(x_train, y_train) # Fit the model to your training data

prediction_RF = RF.predict(user_input_transformed)
if prediction_RF == 1:
    print("Random Forest: This news article is real.")
else:
    print("Random Forest: This news article is fake.")
```

Random Forest: This news article is fake.

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

```
[ ]: # Assuming x_train, x_test, y_train, y_test are already defined from your code
DT = DecisionTreeClassifier()
DT.fit(x_train, y_train)
```

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

```
[ ]: y_pred = DT.predict(x_test)
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	5912
1	1.00	0.99	1.00	5308
accuracy			1.00	11220
macro avg	1.00	1.00	1.00	11220
weighted avg	1.00	1.00	1.00	11220

```
[ ]: from sklearn.neighbors import KNeighborsClassifier
```

```
[ ]: # Assuming x_train, x_test, y_train, y_test are already defined from your code
knn = KNeighborsClassifier(n_neighbors=5)
knn.fit(x_train, y_train)
```

```
[ ]: KNeighborsClassifier()
```

```
[ ]: # K-Nearest Neighbors prediction
prediction_KNN = knn.predict(user_input_transformed) # Use 'knn' instead of
↳ 'KNN'
if prediction_KNN == 1:
    print("K-Nearest Neighbors: This news article is real.")
else:
    print("K-Nearest Neighbors: This news article is fake.")
```

K-Nearest Neighbors: This news article is fake.

```
[ ]: LinR = LinearRegression()
LinR.fit(x_train, y_train)
```

```
[ ]: LinearRegression()
```

```
[ ]: user_input = input("Enter the news article text: ")
user_input_transformed = vectorizer.transform([user_input])
```

Enter the news article text: Spy chiefs pressure Congress to renew expiring surveillance law

```
[ ]: # Logistic Regression prediction
prediction_LR = LR.predict(user_input_transformed)
if prediction_LR == 1:
    print("Logistic Regression: This news article is real.")
else:
    print("Logistic Regression: This news article is fake.")

# Decision Tree prediction
prediction_DT = DT.predict(user_input_transformed)
```



```

if prediction_DT == 1:
    print("Decision Tree: This news article is real.")
else:
    print("Decision Tree: This news article is fake.")

# Random Forest prediction
prediction_RF = RF.predict(user_input_transformed)
if prediction_RF == 1:
    print("Random Forest: This news article is real.")
else:
    print("Random Forest: This news article is fake.")

# Linear Regression prediction
prediction_LinR = LinR.predict(user_input_transformed)
prediction_LinR = np.where(prediction_LinR >= 0.5, 1, 0) # Convert regression
↪ output to binary classification
if prediction_LinR == 1:
    print("Linear Regression: This news article is real.")
else:
    print("Linear Regression: This news article is fake.")

# K-Nearest Neighbors prediction
prediction_KNN = knn.predict(user_input_transformed) # Changed 'KNN' to 'knn'
if prediction_KNN == 1:
    print("K-Nearest Neighbors: This news article is real.")
else:
    print("K-Nearest Neighbors: This news article is fake.")

```

Logistic Regression: This news article is fake.
 Decision Tree: This news article is fake.
 Random Forest: This news article is fake.
 Linear Regression: This news article is fake.
 K-Nearest Neighbors: This news article is fake.