detecting fake news

December 6, 2021

0.0.1 InfoPillar Solution Pvt Ltd

0.0.2 Task-2 Fake News Detection Project

Submitted by- Shraddha Gadale

[1]:
```
# importing all the libraries

import numpy as np
import pandas as pd
import itertools
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import PassiveAggressiveClassifier
from sklearn.metrics import accuracy_score
```

[2]:
```
df=pd.read_csv("news.csv")
df.shape
```

[2]:
```
(6335, 4)
```

[3]:
```
df.head()
```

[3]:
```
   Unnamed: 0                                              title  \
0        8476                       You Can Smell Hillary's Fear
1       10294  Watch The Exact Moment Paul Ryan Committed Pol…
2        3608        Kerry to go to Paris in gesture of sympathy
3       10142  Bernie supporters on Twitter erupt in anger ag…
4         875   The Battle of New York: Why This Primary Matters

                                                text label
0  Daniel Greenfield, a Shillman Journalism Fello…  FAKE
1  Google Pinterest Digg Linkedin Reddit Stumbleu…  FAKE
2  U.S. Secretary of State John F. Kerry said Mon…  REAL
3  - Kaydee King (@KaydeeKing) November 9, 2016 T…  FAKE
4  It's primary day in New York and front-runners…  REAL
```

[4]:
```
#get the labels
labels=df.label
```

```
labels.head()
```

[4]:
```
0    FAKE
1    FAKE
2    REAL
3    FAKE
4    REAL
Name: label, dtype: object
```

[5]:
```
#splitting into training and testing sets
x_train, x_test, y_train, y_test =␣
 ↪train_test_split(df['text'],labels,test_size=0.2)
```

0.0.3 Fit and transform the vectorizer on the train set, and transform the vectorizer on the test set

[6]:
```
#Initialize a TfidfVectorizer
tfidf_vec=TfidfVectorizer(stop_words='english',max_df=0.7)

#Fit and transform train set, transform test set
tfidf_train=tfidf_vec.fit_transform(x_train)
tfidf_test=tfidf_vec.transform(x_test)
```

[7]:
```
pac=PassiveAggressiveClassifier(max_iter=50)
pac.fit(tfidf_train,y_train)

#accuracy
y_pred=pac.predict(tfidf_test)
score=accuracy_score(y_test,y_pred)
print("Accuracy:",round(score*100),"%")
```

```
Accuracy: 94 %
```