

Fake news detection using nlp

Fake news:

- The fundamental definition of fake news is information that leads people wrong. Nowadays, fake news spreads like wildfire, and people share it without confirming it. This is frequently done to advance or enforce specific beliefs and is frequently accomplished through political agendas.
-
- The ability to draw users to media organizations' websites is required to create online advertising revenue. As a result, it is vital to recognize fake news.

How to Create a Fake News Detection System?

- To create a Fake news detection system and to make the system functional, python provides a bunch of libraries. To understand how to create a system using python and make it functional for the Fake News detection system, stay tuned till the end of the article.

Step 1: Importing Libraries.

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
import re
import string
```

Step 2: Importing the Dataset:

```
data_fake = pd.read_csv('Fake.csv')  
data_true = pd.read_csv('True.csv')  
data_fake.head()  
data_true.head()
```

Result:

1: Fake news data

2: True news data

	title	text	subject	date
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn't wish all Americans ...	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017

	title	text	subject	date
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017

Step 3: Assigning Classes to the Dataset:

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


Step 4: Checking Number of Rows and Columns in the Dataset

```
data_fake.shape, data_true.shape
```

Step 5: Manual Testing for Both the Dataset

```
data_fake_manual_testing = data_fake.tail(10)
for i in range(23480, 23470, -1):
    data_fake.drop([i], axis = 0, inplace = True)

data_true_manual_testing = data_true.tail(10)
for i in range(21416, 21406, -1):
    data_true.drop([i], axis = 0, inplace = True)
```

Step 6: Assigning Classes to the Dataset

```
data_fake_manual_testing['class'] = 0  
data_true_manual_testing['class'] = 1
```

Step 7: Merging Both the Dataset

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

Result:

	title	text	subject	date	class
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn't wish all Americans ...	News	December 31, 2017	0
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	0
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	0
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	0
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	0
5	Racist Alabama Cops Brutalize Black Boy While...	The number of cases of cops brutalizing and ki...	News	December 25, 2017	0
6	Fresh Off The Golf Course, Trump Lashes Out A...	Donald Trump spent a good portion of his day a...	News	December 23, 2017	0
7	Trump Said Some INSANELY Racist Stuff Inside ...	In the wake of yet another court decision that...	News	December 23, 2017	0
8	Former CIA Director Slams Trump Over UN Bully...	Many people have raised the alarm regarding th...	News	December 22, 2017	0
9	WATCH: Brand-New Pro-Trump Ad Features So Muc...	Just when you might have thought we'd get a br...	News	December 21, 2017	0

Step 8: Dropping Unwanted Columns

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

Step 9: Create a Function to Clean Text

```
def wordopt(text):  
    text = text.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
```

Step 10: Applying Function to Text Column and Assigning X and Y

```
data['text'] = data['text'].apply(wordopt)
```

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


Step 11: Defining Training and Testing Data and Splitting Them Into 75 -25 Percent Ratio.

```
x_train, x_test, y_train, y_test = train_test_split(x,y, test_size= 0.25)
```

Step 12: Converting Raw Data Into Matrix for Further Process.

```
from sklearn.feature_extraction.text import TfidfVectorizer  
  
vectorization = TfidfVectorizer()  
xv_train = vectorization.fit_transform(x_train)  
xv_test = vectorization.transform(x_test)
```

Step 13: Creating First Model.

```
from sklearn.linear_model import LogisticRegression  
  
LR = LogisticRegression()  
LR.fit(xv_train, y_train)
```

Step 14: Checking the Model Accuracy and Classification Report

```
pred_lr = LR.predict(xv_test)
LR.score(xv_test, y_test)
print(classification_report(y_test, pred_lr))
```

Result:

Logistic Regression Model Accuracy:

0.9876114081996435

Classification Report:

	precision	recall	f1-score	support
0	0.99	0.99	0.99	5819
1	0.99	0.99	0.99	5401
accuracy			0.99	11220
macro avg	0.99	0.99	0.99	11220
weighted avg	0.99	0.99	0.99	11220



Step 15: Creating a Second Model.

```
from sklearn.tree import DecisionTreeClassifier  
  
DT = DecisionTreeClassifier()  
DT.fit(xv_train, y_train)
```

Step 16: Checking the Model Accuracy and Classification Report

```
pred_dt = DT.predict(xv_test)
DT.score(xv_test, y_test)
print(classification_report(y_test, pred_dt))
```

Result:

DecisionTreeClassifier Model Accuracy:

0.9953654188948307

Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	5819
1	1.00	0.99	1.00	5401
accuracy			1.00	11220
macro avg	1.00	1.00	1.00	11220
weighted avg	1.00	1.00	1.00	11220

Step 17: Checking Fake News

Here you have to give input of random news to check whether it's fake or not

Example:

Pro-Russian users have often repeated the Kremlin's original position that the invasion of Ukraine is a "special military operation" to "denazify" and "demilitarise" a "Neo-Nazi state". Many have downplayed allegations of Russian war crimes or even claimed that the war is a supposed "hoax". In one widely shared video, a news reporter could be seen standing in front of lines of body bags, one of which was moving. However, the footage did not show invented war casualties in Ukraine, but a "Fridays for Future" climate change protest in Vienna in February, three weeks before the invasion began. Days later, another viral video of a mannequin claimed to be proof that Ukrainian authorities had "staged" the mass killing of civilians in the town of Bucha. The misleading clip showed a prosthetic doll being dressed and prepared by two men. Nadezhda, an assistant director for a Russian television programme, confirmed to Euronews that the video showed their film set near St. Petersburg and not Ukrainian military personnel. "The information being given [to Russian citizens] is one-sided, it has nothing to do with reality, it is as badly done as any fake," she told Euronews. Euronews The mannequin was being prepared for a television scene in Vsevolozhsk in Russia's Leningrad region. Euronews Other examples of Ukraine war misinformation have centred on "crisis actors" – people who are supposedly hired to act out the role of terrified or deceased war victims. One false claim suggested that a well-known beauty blogger had "acted" as the pregnant victim of a deadly attack on a maternity hospital in the city of Mariupol on 9 March. Russia has shifted its stance on the attack, accusing Ukrainian Azov nationalists of staging a "hoax" bombing at a "non-operational" hospital. Its unfounded claims were later removed by both Facebook and Twitter.

LR Prediction: Fake News
DT Prediction: Fake News

Conclusion:

- The manual classification of false political news requires for a deeper understanding of the field. The problem of predicting and categorizing data in the fake news detection issue needs to be confirmed using training data. Reducing the amount of these features could increase the accuracy of the fake news detection algorithm because the majority of fake news datasets have many attributes, many of which are redundant and useless. As a result, this research suggests a technique for dimensionality reduction-based fake news detection. The dimension-reduced dataset is constructed using the final set of features.