Fake News Detection Using ML

October 27, 2024

1 Fake News Detection Using ML

#Import Libraries

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

```python
from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

#Data Load

```python
[14]: data_fake = pd.read_csv('/content/drive/MyDrive/Data Analysis/Python Project/
      ↪Fake_News/Fake.csv') data_true =
      pd.read_csv('/content/drive/MyDrive/Data Analysis/Python Project/
      ↪Fake_News/True.csv')
```

2 Data preview

```python
[15]: data_fake.head()
```

```
[15]:                                                 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         News
         Americans …
      1  House Intelligence Committee Chairman Devin News
         Nu…
      2  On Friday, it was revealed that former      News
         Milwauk…
```

```
3  On Christmas day, Donald Trump announced     News
   that …
4  Pope Francis used his annual Christmas Day  News
   mes…
                  date
0     December 31,
2017
1  December 31, 2017
2  December 30, 2017
3  December 29, 2017
4  December 25, 2017
```

[16]: `data_true.tail()`

[16]:
```
                                        title \
21412 'Fully committed' NATO backs new U.S. approach…
21413 LexisNexis withdrew two products from Chinese …
21414 Minsk cultural hub becomes haven from authorities
21415 Vatican upbeat on possibility of Pope Francis …
21416 Indonesia to buy $1.14 billion worth of Russia…


                                          text    subject \
21412 BRUSSELS (Reuters) - NATO allies on        worldnews
         Tuesday we…
21413 LONDON (Reuters) - LexisNexis, a provider worldnews
         of l…
21414 MINSK (Reuters) - In the shadow of         worldnews
         disused Sov…
21415 MOSCOW (Reuters) - Vatican Secretary of   worldnews
         State …
21416 JAKARTA (Reuters) - Indonesia will buy 11 worldnews
         Sukh…
                  date
21412 August 22, 2017
21413 August 22, 2017
21414 August 22, 2017
21415 August 22, 2017
21416 August 22, 2017
```

#Adding Labels and Splitting the Data for Testing

[17]:
```
data_fake["class"] = 0
data_true['class'] = 1
```

- Fake news with class = 0
- True news with class = 1

[18]:
```
data_fake_manual_testing = data_fake.tail(10)
data_true_manual_testing = data_true.tail(10)
```

```
data_fake = data_fake.iloc[:-10]
data_true = data_true.iloc[:-10]
```

- Remove last 10 entries for manual testing later:

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

`[34]:`

```
                                                 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          News
   Americans …
1  House Intelligence Committee Chairman Devin News
   Nu…
2  On Friday, it was revealed that former       News
   Milwauk…
3  On Christmas day, Donald Trump announced     News
   that …
4  Pope Francis used his annual Christmas Day  News
   mes…

                date class
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
```

3 Combining and Shuffling the Data

`[19]:`
```
data_merge = pd.concat([data_fake, data_true], axis=0)
data_merge =
data_merge.sample(frac=1).reset_index(drop=True)
```

#Dropping Unnecessary Columns

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

#Preprocessing the Text Data

```
[21]: 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('\w*\d\w*', '', text)
          return text

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

4   Setting Up Independent and Dependent Variables

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

#Splitting Data into Training and Testing Sets

```
[23]: x_train, x_test, y_train, y_test = train_test_split(x, y,
test_size=0.25)
```

5   Text Vectorization

```
[24]: from sklearn.feature_extraction.text import TfidfVectorizer

      vectorization = TfidfVectorizer()
      xv_train =
      vectorization.fit_transform(x_train)
      xv_test =
      vectorization.transform(x_test)
```

#Model Training and Evaluation

Logistic Regression

```
[25]: from sklearn.linear_model import LogisticRegression

      LR = LogisticRegression()
      LR.fit(xv_train, y_train)
      pred_lr = LR.predict(xv_test)
      print("Logistic Regression:\n", classification_report(y_test, pred_lr))
```

```
      Logistic Regression:
                    precision  recall f1-score  support
```

```
            0       0.99       0.99       0.99       5927
            1       0.98       0.99       0.99       5293

     accuracy                             0.99      11220
    macro avg       0.99       0.99       0.99      11220
 weighted avg       0.99       0.99       0.99      11220
```

Decision Tree Classifier

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

      DT = DecisionTreeClassifier()
      DT.fit(xv_train, y_train)
      pred_dt = DT.predict(xv_test)
      print("Decision Tree:\n", classification_report(y_test, pred_dt))
```

```
Decision Tree:
              precision    recall  f1-score   support

           0       0.99       1.00       1.00       5927
           1       0.99       0.99       0.99       5293

    accuracy                             0.99      11220
   macro avg       0.99       0.99       0.99      11220
weighted avg       0.99       0.99       0.99      11220
```

Gradient Boosting Classifier

```python
[27]: from sklearn.ensemble import GradientBoostingClassifier

      GB = GradientBoostingClassifier(random_state=0)
      GB.fit(xv_train, y_train) pred_gb = GB.predict(xv_test)
      print("Gradient Boosting:\n", classification_report(y_test,
      pred_gb))
```

```
Gradient Boosting:
              precision    recall  f1-score   support

           0       1.00       0.99       0.99       5927
           1       0.99       1.00       0.99       5293

    accuracy                             0.99      11220
   macro avg       0.99       0.99       0.99      11220
weighted avg       0.99       0.99       0.99      11220
```

Random Forest Classifier

```python
[28]: from sklearn.ensemble import RandomForestClassifier

      RF = RandomForestClassifier(random_state=0)
```

```
RF.fit(xv_train, y_train) pred_rf = RF.predict(xv_test)
print("Random Forest:\n", classification_report(y_test,
pred_rf))
```

```
Random Forest:
              precision    recall f1-score   support
           0       0.99      0.99      0.99      5927
           1       0.99      0.99      0.99      5293

    accuracy                           0.99     11220
   macro avg       0.99      0.99      0.99     11220
weighted avg       0.99      0.99      0.99     11220
```
#Model Testing with Manual Entry

```python
[29]: def output_label(n):
          if n == 0:
              return "Fake News"
          elif n == 1:
              return "Not A Fake News "

      def manual_testing(news):
          testing_news = {"text":  news]}
          new_def_test = pd.DataFrame(testing_news)
          new_def_test['text'] = new_def_test['text'].apply(wordopt)
          new_x_test = new_def_test['text']
          new_xv_test = vectorization.transform(new_x_test)

          pred_LR = LR.predict(new_xv_test)
          pred_DT = DT.predict(new_xv_test)
          pred_GB = GB.predict(new_xv_test)
          pred_RF = RF.predict(new_xv_test)

          return print(f"\n\nLR Prediction: {output_label(pred_LR[0])} \n"
                       f"DT Prediction: {output_label(pred_DT[0])} \n"
                       f"GBC Prediction: {output_label(pred_GB[0])} \n"
                       f"RFC Prediction: {output_label(pred_RF[0])}")
```

• Define helper functions to label predictions as "Fake News" or "Not A Fake News" and test the models on new inputs:

6   Testing the Model with Manual Input

```python
[33]: news = input("Enter news text: ")
      manual_testing(news)
```

```
Enter news text: On Friday, it was revealed that former Milwaukee
Sheriff David Clarke, who was being considered for Homeland Security
Secretary in Donald Trump s administration, has an email scandal of
```

6

his own.In January, there was a brief run-in on a plane between Clarke and fellow passenger Dan Black, who he later had detained by the police for no reason whatsoever, except that maybe his feelings were hurt. Clarke messaged the police to stop Black after he deplaned, and now, a search warrant has been executed by the FBI to see the exchanges.Clarke is calling it fake news even though copies of the search warrant are on the Internet. I am UNINTIMIDATED by lib media attempts to smear and discredit me with their FAKE NEWS reports designed to silence me,   the former sheriff tweeted. I will continue to poke them in the eye with a sharp stick and bitch slap these scum bags til they get it. I have been attacked by better people than them #MAGA I am UNINTIMIDATED by lib media attempts to smear and discredit me with their FAKE NEWS reports designed to silence me. I will continue to poke them in the eye with a sharp stick and bitch slap these scum bags til they get it. I have been attacked by better people than them #MAGA pic.twitter.com/XtZW5PdU2b David A. Clarke, Jr. (@SheriffClarke) December 30, 2017He didn t stop there.BREAKING NEWS! When LYING LIB MEDIA makes up FAKE NEWS to smear me, the ANTIDOTE is go right at them. Punch them in the nose & MAKE THEM TASTE THEIR OWN BLOOD. Nothing gets a bully like LYING LIB MEDIA S attention better than to give them a taste of their own blood #neverbackdown pic.twitter.com/T2NY2psHCR   David A. Clarke, Jr. (@SheriffClarke) December 30, 2017The internet called him out.This is your local newspaper and that search warrant isn t fake, and just because the chose not to file charges at the time doesn t mean they won t! Especially if you continue to lie. Months after decision not to charge Clarke, email search warrant filed https://t.co/zcbyc4Wp5b KeithLeBlanc (@KeithLeBlanc63) December 30, 2017I just hope the rest of the Village People aren t implicated.   Kirk Ketchum (@kirkketchum) December 30, 2017Slaw, baked potatoes, or French fries? pic.twitter.com/fWfXsZupxyALT- Immigration (@ALT_uscis) December 30, 2017pic.twitter.com/ymsOBLjfxU Pendulum Swinger (@PendulumSwngr) December 30, 2017you called your police friends to stand up for you when someone made fun of your hat   Chris Jackson (@ChrisCJackson) December 30, 2017Is it me, with this masterful pshop of your hat, which I seem to never tire of. I think it s the steely resolve in your one visible eye pic.twitter.com/dWr5k8ZEZV   Chris Mohney (@chrismohney) December 30, 2017Are you indicating with your fingers how many people died in your jail? I think you re a few fingers short, dipshit Ike Barinholtz (@ikebarinholtz) December 30, 2017ROFL. Internet tough guy with fake flair. pic.twitter.com/ulCFddhkdy KellMeCrazy (@Kel_MoonFace) December 30, 2017You re so edgy, buddy. Mrs. SMH (@MRSSMH2) December 30, 2017Is his break over at Applebees?   Aaron (@feltrrr2) December 30, 2017Are you trying to earn your still relevant badge? CircusRebel (@CircusDrew) December 30, 2017make sure to hydrate, drink lots of water. It s rumored that prisoners can be denied water

by prison officials. Robert Klinc (@RobertKlinc1) December 30,
2017Terrill Thomas, the 38-year-old black man who died of thirst in
Clarke s Milwaukee County Jail cell this April, was a victim of
homicide. We just thought we should point that out. It can t be
repeated enough.Photo by Spencer Platt/Getty Images.

LR Prediction: Fake News
DT Prediction: Fake News
GBC Prediction: Fake News
RFC Prediction: Fake News

```
[32]: news = input("Enter news text: ")
      manual_testing(news)
```

Enter news text: NEW YORK/WASHINGTON (Reuters) - The new U.S. tax
code targets high-tax states and may be unconstitutional, New York
Governor Andrew Cuomo said on Thursday, saying that the bill may
violate New York residentsâ€ rights to due process and equal
protection.The sweeping Republican tax bill signed into law by U.S.
President Donald Trump on Friday introduces a cap, of $10,000,   on
deductions of state and local income and property taxes, known as
SALT. The tax overhaul was the partyâ€s first major legislative
victory since Trump took office in January. The SALT provision will
hit many taxpayers in states with high incomes, high property values
and high taxes, like New York, New Jersey and California. Those
states are generally Democratic leaning.   â€œIâ€m not even sure
what they did is legally constitutional and thatâ€s something weâ€re
looking at now,â€ Cuomo said in an interview with CNN. In an
interview with CNBC, Cuomo suggested why the bill may be
unconstitutional.   â€œPolitics does not trump the law,â€ Cuomo
said on CNBC. â€œYou have the constitution, you have the law, you
have due process, you have equal protection. You canâ€t use politics
just because the majority controls to override the law.â€ The Fifth
Amendment of the Constitution, better known for its protection
against selfincrimination, also protects individuals from seizure of
life, liberty or property without due process and has been
interpreted by the Supreme Court as guaranteeing equal protection by
the law.   Cuomo and California Governor Jerry Brown, both
Democrats, have previously said they were exploring legal challenges
to SALT deduction limits.   Law professors have said legal challenges
would likely rest on arguing that the provision interferes with the
protection of statesâ€ rights under the U.S. Constitutionâ€s 10th
Amendment. Tax attorneys said Cuomoâ€s legal argument against the tax
bill could be that it discriminates and places an unjust tax burden
on states that heavily voted for Democrats in the past - known as
â€œblue states.â€ â€œThe de facto effect of this legislation is to

discriminate against blue states and particularly from (Cuomoâ€s) perspective the state of New York,â€ said Joseph Callahan, an attorney with the law firm Mackay, Caswell & Callahan in New York.But some tax experts noted the U.S. Supreme Court has interpreted the 16th Amendment to give Congress broad latitude to tax as it sees fit. In a frequently cited 1934 decision, the Supreme Court called tax deductions a â€œlegislative graceâ€ rather than a vested right. â€œI donâ€t understand how they think they have a valid lawsuit here,â€ David Gamage, a professor of tax law at Indiana Universityâ€s Maurer School of Law, told Reuters last week, speaking generally about governors in blue states that could challenge the tax bill. Cuomo also said on Thursday that New York is proposing a restructuring of its tax code. He provided no details. A group of 13 law professors on Dec. 18 published a paper suggesting ways that high-tax states could minimize the effects of the SALT deduction cap. Their suggestions included shifting more of the tax burden onto businesses in the form of higher employer-side payroll taxes, since the federal tax billâ€s cap on SALT deductions only applies to individuals and not businesses. States also could raise taxes on pass-through entities, which the federal tax bill specifically benefits with a lower rate on a portion of their income. On Friday, Cuomo said he would allow state residents to make a partial or full pre-payment on their property tax bill before Jan. 1, allowing taxpapyers to deduct such payments for 2017 before the cap kicks in, prompting a wave of residents to pay early. However, the U.S. Internal Revenue Service on Wednesday advised homeowners that the pre-payment of 2018 property taxes may not be deductible.

LR Prediction: Not A Fake News
DT Prediction: Not A Fake News
GBC Prediction: Not A Fake News
RFC Prediction: Not A Fake News

Project Summary: Fake News Detection Using Machine Learning

The **Fake News Detection** project aims to create a reliable model for identifying fake news articles using machine learning techniques. The project employs a dataset of labeled news articles, with each article marked as either "True" or "Fake." By applying preprocessing techniques, text vectorization, and various machine learning algorithms, the model provides high accuracy in distinguishing between real and fake news.

Key Steps in the Project:

1. **Data Collection and Labeling**: The dataset comprises two CSV files, one for fake news and another for true news. Labels are added to differentiate the classes, where "0" indicates fake news, and "1" indicates real news.

2. **Data Preprocessing**:
   o **Data Cleaning**: Text data is cleaned by removing special characters, URLs, and unnecessary punctuation, transforming all text to lowercase to maintain uniformity.
   o **Feature Extraction**: Using TfidfVectorizer, the project converts text data into numerical format, making it usable for machine learning models.

3. **Model Selection and Training**: Several machine learning algorithms are trained and evaluated to classify news articles, including:
   o **Logistic Regression**: Achieved high accuracy with balanced precision, recall, and F1 scores.
   o **Decision Tree Classifier**: Showed robust performance with near-perfect precision and recall.
   o **Gradient Boosting Classifier**: Excelled in accuracy and minimized false positives.
   o **Random Forest Classifier**: Provided high accuracy with reliable predictions.

4. **Evaluation and Testing**:
   o The models were evaluated using metrics such as accuracy, precision, recall, and F1-score to ensure a well-rounded performance.
   o Manual testing was conducted with user-inputted news samples, allowing the models to predict real or fake news with high accuracy and consistency across different algorithms.

5. **Results**: All models demonstrated high accuracy (~99%), with Logistic Regression, Decision Tree, Gradient Boosting, and Random Forest classifiers all effectively distinguishing between fake and real news articles. The consistent performance across models indicates robustness and reliability in various real-world scenarios.

Conclusion:

The Fake News Detection model is highly accurate and reliable in identifying fake news articles. This project showcases the application of machine learning in addressing misinformation, highlighting the effectiveness of text processing, feature extraction, and classification algorithms in distinguishing between real and fake news. The tool can be further enhanced by training on more extensive and varied datasets to improve generalization across diverse topics and sources.

Shaun Mia | LinkedIn