PROJECT DOCUMENTATION: FAKE NEWS DETECTION PHASE 5

Problem Statement

The problem is to develop a system that can automatically detect and classify fake news articles from real ones using Natural Language Processing (NLP) techniques. The goal is to create a model that can assist in identifying misinformation and disinformation in news articles, which is crucial for preventing the spread of false information and ensuring the accuracy of news sources.

Design thinking process:

Data Loading: The code begins by loading two datasets - 'True.csv' and 'Fake.csv'. These datasets presumably contain real news and fake news articles, respectively. This aligns with the problem of collecting labeled data for training a fake news detection model.

Labeling Data: It adds a 'label' column to each dataset to indicate whether an article is real (0) or fake (1). This is essential for supervised machine learning, as it provides ground truth labels for the model to learn from.

Concatenation and Shuffling: The code combines the two datasets into one and shuffles the data. This step ensures that the model is trained on a balanced and randomized mix of real and fake news, preventing it from learning any dataset-specific biases.

Data Cleaning and Preprocessing: Text preprocessing is a crucial aspect of NLP. The preprocess_text function lowers the text, splits it into words, and removes stopwords.

Text Vectorization: The code employs TF-IDF vectorization (Term Frequency-Inverse Document Frequency) to convert the text data into numerical features. This is a common feature extraction technique in NLP and aligns with the feature extraction phase.

Model Training: The code uses a Multinomial Naive Bayes classifier (MultinomialNB) for model training. Naive Bayes is a suitable choice for text classification tasks, as it's well-suited for working with text data**.**

Model Evaluation: The code evaluates the model's performance using the classification_report function, which provides metrics like precision, recall, and F1-score for both the real and fake news classes. This aligns with the phase of model evaluation.

Model Testing

Test the trained models using a separate dataset, assessing metrics like accuracy and precision. Conduct cross-validation for generalizability and analyze confusion matrices for error insights. Fine-tune hyperparameters based on testing results to optimize model performance for real-world application.

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
```

Importing dataset

```
df_fake = pd.read_csv("D:\\IBM AI NAN MUDHALVAN\\csvfakedata\\Fake.csv")

df_fake.head()
```

```
                                              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
0  December 31, 2017
1  December 31, 2017
2  December 30, 2017
3  December 29, 2017
4  December 25, 2017
```

```
df_true = pd.read_csv("D:\\IBM AI NAN MUDHALVAN\\csvtruedata\\True.csv")

df_true.head()
```

```
                                              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
0  December 31, 2017
1  December 29, 2017
2  December 31, 2017
3  December 30, 2017
4  December 29, 2017
```

```python
X = df_true["title"].tolist() + df_fake["title"].tolist()
y = [0] * len(df_true) + [1] * len(df_fake)
```

```python
from sklearn.model_selection import train_test_split
```

```python
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
```

```python
from simple_nlp_library import preprocessing, embeddings
```

```python
stop_words = preprocessing.stop_words()
vectors = embeddings.vectors()
```

```python
X_train_vec = [embeddings.tokens_vector(vectors, preprocessing.semantic_tokens(stop_words, x)) for x in X_train]
X_test_vec = [embeddings.tokens_vector(vectors, preprocessing.semantic_tokens(stop_words, x)) for x in X_test]
```

```python
from sklearn.neural_network import MLPClassifier
```

```python
clf = MLPClassifier(hidden_layer_sizes=(25), early_stopping=True)
clf.fit(X_train_vec, y_train)
```

```
MLPClassifier(early_stopping=True, hidden_layer_sizes=25)
```

```python
import matplotlib.pyplot as plt
```

```python
plt.plot(clf.validation_scores_)
```

```
[<matplotlib.lines.Line2D at 0x258a4597ac0>]
```

```python
df_fake.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 23481 entries, 0 to 23480
Data columns (total 4 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   title         23481 non-null  object
 1   text          23481 non-null  object
 2   subject  23481 non-null  object
 3   date          23481 non-null  object
dtypes: object(4)
memory usage: 733.9+ KB
```

df_true.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21417 entries, 0 to 21416
Data columns (total 4 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   title   21417 non-null  object
 1   text    21417 non-null  object
 2   subject  21417 non-null  object
 3   date    21417 non-null  object
dtypes: object(4)
memory usage: 669.4+ KB
```

df_fake.describe()

```
                               title   text subject  \
count                         23481  23481  23481
unique                        17903  17455         6
top     MEDIA IGNORES Time That Bill Clinton FIRED His...        News
freq                              6    626   9050

              date
count          23481
unique          1681
top     May 10, 2017
freq             46
```

df_true.describe()

```
                               title  \
count                         21417
unique                        20826
```

```
top      Factbox: Trump fills top jobs for his administ...
freq                                14


                                      text         subject  \
count                                21417           21417
unique                               21192            2
top      (Reuters) - Highlights for U.S. President Dona...  politicsNews
freq                                    8           11272


              date
count        21417
unique         716
top      December 20, 2017
freq           182
```

df_fake.columns

Index(['title', 'text', 'subject', 'date'], dtype='object')

df_true.columns

Index(['title', 'text', 'subject', 'date'], dtype='object')

Inserting a column "class" as target feature

```
df_fake["class"] = 0
df_true["class"] = 1
```

df_fake.shape, df_true.shape

((23481, 5), (21417, 5))

```
# Removing last 10 rows for manual testing
df_fake_manual_testing = df_fake.tail(10)
for i in range(23480,23470,-1):
    df_fake.drop([i], axis = 0, inplace = True)



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

df_fake.shape, df_true.shape

((23471, 5), (21407, 5))

```
df_fake_manual_testing["class"] = 0
df_true_manual_testing["class"] = 1
```

C:\Users\sasidharan\AppData\Local\Temp\ipykernel_11092\860779283.py:1:
SettingWithCopyWarning:
 A value is trying to be set on a copy of a slice from a DataFrame.
 Try using .loc[row_indexer,col_indexer] = value instead

 See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versu
s-a-copy
   df_fake_manual_testing["class"] = 0
 C:\Users\sasidharan\AppData\Local\Temp\ipykernel_11092\860779283.py:2:
SettingWithCopyWarning:
 A value is trying to be set on a copy of a slice from a DataFrame.
 Try using .loc[row_indexer,col_indexer] = value instead

 See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versu
s-a-copy
   df_true_manual_testing["class"] = 1

```
df_fake_manual_testing.head(10)
```

	title
23471	Seven Iranians freed in the prisoner swap have...
23472	#Hashtag Hell & The Fake Left
23473	Astroturfing: Journalist Reveals Brainwashing ...
23474	The New American Century: An Era of Fraud
23475	Hillary Clinton: 'Israel First' (and no peace ...
23476	McPain: John McCain Furious That Iran Treated ...
23477	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...
23478	Sunnistan: US and Allied 'Safe Zone' Plan to T...
23479	How to Blow $700 Million: Al Jazeera America F...
23480	10 U.S. Navy Sailors Held by Iranian Military ...

	text	subject
23471	21st Century Wire says This week, the historic...	Middle-east
23472	By Dady Chery and Gilbert MercierAll writers ...	Middle-east
23473	Vic Bishop Waking TimesOur reality is carefull...	Middle-east
23474	Paul Craig RobertsIn the last years of the 20t...	Middle-east
23475	Robert Fantina CounterpunchAlthough the United...	Middle-east
23476	21st Century Wire says As 21WIRE reported earl...	Middle-east
23477	21st Century Wire says It s a familiar theme. ...	Middle-east
23478	Patrick Henningsen 21st Century WireRemember ...	Middle-east

	title	
23479	21st Century Wire says Al Jazeera America will...	Middle-east
23480	21st Century Wire says As 21WIRE predicted in ...	Middle-east

	date	class
23471	January 20, 2016	0
23472	January 19, 2016	0
23473	January 19, 2016	0
23474	January 19, 2016	0
23475	January 18, 2016	0
23476	January 16, 2016	0
23477	January 16, 2016	0
23478	January 15, 2016	0
23479	January 14, 2016	0
23480	January 12, 2016	0

df_true_manual_testing.head(10)

	title \
21407	Mata Pires, owner of embattled Brazil builder ...
21408	U.S., North Korea clash at U.N. forum over nuc...
21409	U.S., North Korea clash at U.N. arms forum on ...
21410	Headless torso could belong to submarine journ...
21411	North Korea shipments to Syria chemical arms a...
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...

	textsubject \	
21407	SAO PAULO (Reuters) - Cesar Mata Pires, the ow...	worldnews
21408	GENEVA (Reuters) - North Korea and the United ...	worldnews
21409	GENEVA (Reuters) - North Korea and the United ...	worldnews
21410	COPENHAGEN (Reuters) - Danish police said on T...	worldnews
21411	UNITED NATIONS (Reuters) - Two North Korean sh...	worldnews
21412	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews
21413	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews
21414	MINSK (Reuters) - In the shadow of disused Sov...	worldnews
21415	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews
21416	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews

	date	class
21407	August 22, 2017	1
21408	August 22, 2017	1
21409	August 22, 2017	1

```
21410  August 22, 2017        1
21411  August 21, 2017        1
21412  August 22, 2017        1
21413  August 22, 2017        1
21414  August 22, 2017        1
21415  August 22, 2017        1
21416  August 22, 2017        1
```

```
df_manual_testing = pd.concat([df_fake_manual_testing,df_true_manual_testing], axis = 0)
df_manual_testing.to_csv("manual_testing.csv")
```

Merging True and Fake Dataframes

```
df_merge = pd.concat([df_fake, df_true], axis =0 )
df_merge.head(10)
```

```
                                    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...
5   Racist Alabama Cops Brutalize Black Boy While...
6   Fresh Off The Golf Course, Trump Lashes Out A...
7   Trump Said Some INSANELY Racist Stuff Inside ...
8   Former CIA Director Slams Trump Over UN Bully...
9   WATCH: Brand-New Pro-Trump Ad Features So Muc...
```

```
                                    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
5  The number of cases of cops brutalizing and ki...News
6  Donald Trump spent a good portion of his day a...      News
7  In the wake of yet another court decision that...    News
8  Many people have raised the alarm regarding th...      News
9  Just when you might have thought we d get a br...      News
```

```
               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
5  December 25, 2017        0
6  December 23, 2017        0
7  December 23, 2017        0
8  December 22, 2017        0
9  December 21, 2017        0
```

df_merge.columns

Index(['title', 'text', 'subject', 'date', 'class'], dtype='object')

Removing columns which are not required

df = df_merge.drop(["title", "subject","date"], axis = 1)

df.isnull().sum()

```
text    0
 class  0
 dtype: int64
```

Random Shuffling the dataframe

df = df.sample(frac = 1)

df.head()

```
                                    text  class
16707  The Obama s just added millions onto the taxpa...  0
19623  MEXICO CITY (Reuters) - At least 138 people we...        1
12212  ANGELA Merkel today completed an astonishing U...      0
20677  As a mother of three young girls, if anything ...      0
22001  IMAGE: Globalist scribe for the American branc...  0
```

df.reset_index(inplace = True)
 df.drop(["index"], axis = 1, inplace = True)

df.columns

Index(['text', 'class'], dtype='object')

df.head()

```
                                   text  class
0  The Obama s just added millions onto the taxpa...        0
1  MEXICO CITY (Reuters) - At least 138 people we...        1
2  ANGELA Merkel today completed an astonishing U...      0
```

3 As a mother of three young girls, if anything ...   0
4 IMAGE: Globalist scribe for the American branc...   0

Creating a function to process the texts

```
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

df["text"] = df["text"].apply(wordopt)
```

Defining dependent and independent variables

```
x = df["text"]
 y = df["class"]
```

Splitting Training and Testing

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

	precision	recall	f1-score	support
0	0.99	0.99	0.99	5928
1	0.98	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

	precision	recall	f1-score	support
accuracy			1.00	11220
macro avg	1.00	1.00	1.00	11220
weighted avg	1.00	1.00	1.00	11220

Model Testing

```python
def output_lable(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_GBC = GBC.predict(new_xv_test)
        pred_RFC = RFC.predict(new_xv_test)

        return print("\n\nLR Prediction: {} \nDT Prediction: {} \nGBC Prediction: {} \nRFC
Prediction: {}".format(output_lable(pred_LR[0]),
output_lable(pred_DT[0]),

output_lable(pred_GBC[0]),

output_lable(pred_RFC[0])))

news = str(input())
 manual_testing(news)
```

BRUSSELS (Reuters) - NATO allies on Tuesday welcomed President Donald Trump s decision to commit more forces to Afghanistan, as part of a new U.S. strategy he said would require more troops and funding from America s partners. Having run for the White House last year on a pledge to withdraw swiftly from Afghanistan, Trump reversed course on Monday and promised a stepped-up military campaign against Taliban insurgents, saying: Our troops will fight to win . U.S. officials said he had signed off on plans to send about 4,000 more U.S. troops to add to the roughly 8,400 now deployed in Afghanistan. But his speech did not define benchmarks for successfully ending the war that began with the U.S.-led invasion of Afghanistan in 2001, and which he acknowledged had required an   extraordinary sacrifice of blood and treasure . We will ask our NATO allies and global partners to support our new strategy, with additional troops and funding increases in line with our own. We are confident they will, Trump said. That comment signaled he would further increase pressure on U.S. partners who have already been jolted by his repeated demands to step up their contributions to NATO and his description of the alliance as obsolete - even though, since taking office, he has said this is no longer the case. NATO Secretary General Jens Stoltenberg said in a statement: NATO remains fully committed to Afghanistan and I am looking forward to discussing the way ahead with (Defense) Secretary (James) Mattis and our Allies and international partners. NATO has 12,000 troops in

Afghanistan, and 15 countries have pledged more, Stoltenberg said. Britain, a leading NATO member, called the U.S. commitment very welcome . In my call with Secretary Mattis yesterday we agreed that despite the challenges, we have to stay the course in Afghanistan to help build up its fragile democracy and reduce the terrorist threat to the West, Defence Secretary Michael Fallon said. Germany, which has borne the brunt of Trump s criticism over the scale of its defense spending, also welcomed the new U.S. plan. Our continued commitment is necessary on the path to stabilizing the country, a government spokeswoman said. In June, European allies had already pledged more troops but had not given details on numbers, waiting for the Trump administration to outline its strategy for the region.Nearly 16 years after the U.S.-led invasion - a response to the Sept. 11 attacks which were planned by al Qaeda leader Osama bin Laden from Afghanistan - the country is still struggling with weak central government and a Taliban insurgency. Trump said he shared the frustration of the American people who were weary of war without victory , but a hasty withdrawal would create a vacuum for groups like Islamic State and al Qaeda to fill.

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

```
news = str(input())
manual_testing(news)
```

Vic Bishop Waking TimesOur reality is carefully constructed by powerful corporate, political and special interest sources in order to covertly sway public opinion. Blatant lies are often televised regarding terrorism, food, war, health, etc. They are fashioned to sway public opinion and condition viewers to accept what have become destructive societal norms.The practice of manipulating and controlling public opinion with distorted media messages has become so common that there is a whole industry formed around this. The entire role of this brainwashing industry is to figure out how to spin information to journalists, similar to the lobbying of government. It is never really clear just how much truth the journalists receive because the news industry has become complacent. The messages that it presents are shaped by corporate powers who often spend millions on advertising with the six conglomerates that own 90% of the media:General Electric (GE), News-Corp, Disney, Viacom, Time Warner, and CBS. Yet, these corporations function under many different brands, such as FOX, ABC, CNN, Comcast, Wall Street Journal, etc, giving people the perception of choice   As Tavistock s researchers showed, it was important that the victims of mass brainwashing not be aware that their environment was being controlled; there should thus be a vast number of sources for information, whose messages could be varied slightly, so as to mask the sense of external control. ~ Specialist of mass brainwashing, L. WolfeNew Brainwashing Tactic Called AstroturfWith alternative media on the rise, the propaganda machine continues to expand. Below is a video of Sharyl Attkisson, investigative reporter with CBS, during which she explains how astroturf, or fake grassroots movements, are used to spin information not only to influence journalists but to sway public opinion. Astroturf is a perversion of grassroots. Astroturf is when political, corporate or other

special interests disguise themselves and publish blogs, start facebook and twitter accounts, publish ads, letters to the editor, or simply post comments online, to try to fool you into thinking an independent or grassroots movement is speaking. ~ Sharyl Attkisson, Investigative ReporterHow do you separate fact from fiction? Sharyl Attkisson finishes her talk with some insights on how to identify signs of propaganda and astroturfing These methods are used to give people the impression that there is widespread support for an agenda, when, in reality, one may not exist. Astroturf tactics are also used to discredit or criticize those that disagree with certain agendas, using stereotypical names such as conspiracy theorist or quack. When in fact when someone dares to reveal the truth or questions the official story, it should spark a deeper curiosity and encourage further scrutiny of the information.This article (Journalist Reveals Tactics Brainwashing Industry Uses to Manipulate the Public) was originally created and published by Waking Times and is published here under a Creative Commons license with attribution to Vic Bishop and WakingTimes.com. It may be re-posted freely with proper attribution, author bio, and this copyright statement. READ MORE MSM PROPAGANDA NEWS AT: 21st Century Wire MSM Watch Files

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

```
news = str(input())
manual_testing(news)
```

SAO PAULO (Reuters) - Cesar Mata Pires, the owner and co-founder of Brazilian engineering conglomerate OAS SA, one of the largest companies involved in Brazil s corruption scandal, died on Tuesday. He was 68. Mata Pires died of a heart attack while taking a morning walk in an upscale district of S o Paulo, where OAS is based, a person with direct knowledge of the matter said. Efforts to contact his family were unsuccessful. OAS declined to comment. The son of a wealthy cattle rancher in the northeastern state of Bahia, Mata Pires links to politicians were central to the expansion of OAS, which became Brazil s No. 4 builder earlier this decade, people familiar with his career told Reuters last year. His big break came when he befriended Antonio Carlos Magalh es, a popular politician who was Bahia governor several times, and eventually married his daughter Tereza. Brazilians joked that OAS stood for Obras Arranjadas pelo Sogro - or Work Arranged by the Father-In-Law.   After years of steady growth triggered by a flurry of massive government contracts, OAS was ensnared in Operation Car Wash which unearthed an illegal contracting ring between state firms and builders. The ensuing scandal helped topple former Brazilian President Dilma Rousseff last year. Trained as an engineer, Mata Pires founded OAS with two colleagues in 1976 to do sub-contracting work for larger rival Odebrecht SA - the biggest of the builders involved in the probe. Before the scandal, Forbes magazine estimated Mata Pires fortune at $1.6 billion. He dropped off the magazine s billionaire list in 2015, months after OAS sought bankruptcy protection after the Car Wash scandal. While Mata Pires was never accused of wrongdoing in the investigations, creditors demanded he and his family stay away from the builder s day-to-day operations, people directly involved in the

negotiations told Reuters at the time. He is survived by his wife and his two sons.

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

Conclusion

In conclusion, addressing the fake news challenge via Natural Language Processing (NLP) and design thinking is crucial. NLP aids in detecting deceptive information, while design thinking offers a user-centric approach. By combining linguistic analysis, scalability, and ethical considerations, we aim to empower users and platforms. The objective is to enhance information credibility, preserve public trust, and create a more resilient digital information ecosystem. This interdisciplinary approach stands as a vital step toward combating the detrimental effects of fake news in our digital age.