

AI_PHASE 3 IBM

TITLE: FAKE NEWS DETECTION USING NLP

Begin Building your Project by Loading and Preprocessing the Dataset.

TEAM MEMBERS:

REG NO

1. M. BHARANI DHARAN [LEADER]	731221104302
2. S. ARULMURUGAN.	731221104004
3. S. VASANTH KUMAR.	731221104038
4. S. SURYA.	731221104036
5. V. SRINIVAS.	731221104307



FAKE NEWS DETECTION USING NLP

INTRODUCTION

Whether you're a homeowner looking to estimate the value of your property, a real estate investor seeking profitable opportunities, or a data scientist aiming to build a predictive model, the foundation on this lies in loading and preprocessing the dataset.

Building a house price prediction model is a data-driven process that involves harnessing the power of deep learning to analyse the “FAKE NEWS DETECTION USING NLP”. This journey begins with the fundamental steps of data loading and preprocessing.

This introduction will guide you through the initial steps of the process. We'll explore how to import essential libraries, load the dataset, and perform critical preprocessing steps. Data preprocessing is crucial as it helps clean, format, and prepare the data for further analysis. This includes handling missing values, encoding categories variables, and ensuring that the data is appropriately scaled.

GIVEN DATASET:-

REAL DATASET

```
In [12]: real_data.head(23500)
```

Out[12]:

		title	text	subject	date	target
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017	1	
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017	1	
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017	1	
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017	1	
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017	1	
...	
21412	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017	1	
21413	LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of L...	worldnews	August 22, 2017	1	
21414	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disused Sov...	worldnews	August 22, 2017	1	
21415	Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017	1	
21416	Indonesia to buy \$1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	August 22, 2017	1	

21417 rows × 5 columns

FAKE DATASET

```
In [13]: fake_data.head(21400)
```

Out[13]:

		title	text	subject	date	target
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	
...	
21395	CHILLING: How America Looks After 8 Long Years...	The shocking truth about how close we are to b...	left-news	Oct 23, 2015	0	
21396	CANADA'S OBAMA? WATCH New Prime Minister Call ...	The global wussification of the male Identifyi...	left-news	Oct 23, 2015	0	
21397	THIS Is How Far The Left Will Go To Protect Hi...	Jeffrey Toobin chimes in to protect Hillary Cl...	left-news	Oct 23, 2015	0	
21398	NOT KIDDING! Obama's Education Department Want...	Barack Hussein Obama the most anti-American Pr...	left-news	Oct 22, 2015	0	
21399	SWEDEN HOUSES 600+ MUSLIM REFUGEES In Luxury S...	Nothing but the best for soldiers of the calip...	left-news	Oct 21, 2015	0	

21400 rows × 5 columns

Necessary step to follow:

1.Import Libraries:

Start by importing the necessary libraries:

PROGRAM:

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

2.Load the Dataset:

Load your dataset into a Pandas DataFrame. You can typically find Fake news and Real news datasets in CSV format, but you can adapt this code to other formats as needed.

PROGRAM:

```
real_data = pd.read_csv('True.csv')
fake_data = pd.read_csv('Fake.csv')
real_data.head(10)
fake_data.head(10)
```

3. Add Column and Merging

PROGRAM:

```
real_data['target'] = 1
fake_data['target'] = 0
data = pd.concat([real_data, fake_data], ignore_index=True,
sort=False) data.head()
```

4. Data cleaning

PROGRAM:

```
data.isnull().sum()
```

5. Train test Split and Train LSTM Model:

PROGRAM:

```
X_train, X_test, y_train, y_test = train_test_split(data['text'],
data['target'], random_state=0)
batch_size = 256 epochs = 10 embed_size = 100
model.summary()
```



Importance of loading and processing dataset:

Loading and preprocessing the dataset is an important first step in building any deep learning model. However, it is especially important for Fake news detection models, as fake news datasets are often complex and noisy.

By loading and preprocessing the dataset, we can ensure that the deep learning algorithm is able to learn from the data effectively and accurately.

How to overcome the challenges of loading and preprocessing a fake news dataset:

1. Use a data preprocessing library:

There are a number of libraries available that can help with data preprocessing tasks, such as handling missing values, encoding categorical variables, and scaling the features.

2. Carefully consider the specific needs of your model:

The best way to preprocess the data will depend on the specific deep learning algorithm that you are using. It is important to carefully consider the requirements of the algorithm and to preprocess the data in a way that is compatible with the algorithm.

1. Loading the Dataset:

Loading the dataset using deep learning is the process of bringing the data into the deep learning environment so that it can be used to train and evaluate a model.

LOADING DATASETS:

```
Dataset = pd.read_csv(r' C:\Users\surya\Downloads\archive(Real and Fake_csv))
```

Output:

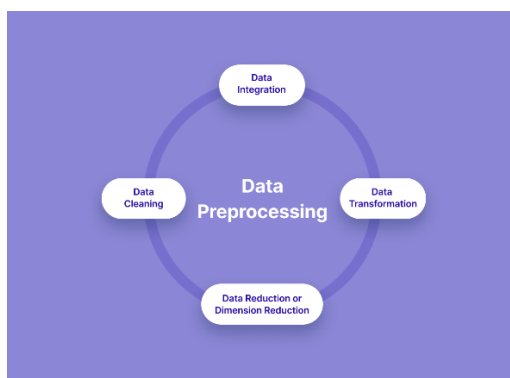
Out[20]:

	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
...
23476	McPain: John McCain Furious That Iran Treated ...	21st Century Wire says As 21WIRE reported earl...	Middle-east	January 16, 2016
23477	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...	21st Century Wire says It s a familiar theme. ...	Middle-east	January 16, 2016
23478	Sunnistan: US and Allied 'Safe Zone' Plan to T...	Patrick Henningsen 21st Century WireRemember ...	Middle-east	January 15, 2016
23479	How to Blow \$700 Million: Al Jazeera America F...	21st Century Wire says Al Jazeera America will...	Middle-east	January 14, 2016
23480	10 U.S. Navy Sailors Held by Iranian Military ...	21st Century Wire says As 21WIRE predicted in ...	Middle-east	January 12, 2016

23481 rows × 4 columns

```
In [221]: dataset = pd.read_csv('real and fake.csv')
```

2.Preprocessing the dataset:



- Data preprocessing is the process of cleaning, transforming, and integrating data in order to make it ready for analysis.
- This may involve removing errors and inconsistencies, handling missing values, transforming the data into a consistent format, and scaling the data to a suitable range.

VISUALISATION AND PREPROCESSING OF DATASET:

1. Count the fake and real news

```
print(data["target"].value_counts())

fig, ax = plt.subplots(1,2, figsize=(19, 5))

g1 = sns.countplot(data.target,ax=ax[0],palette="pastel");

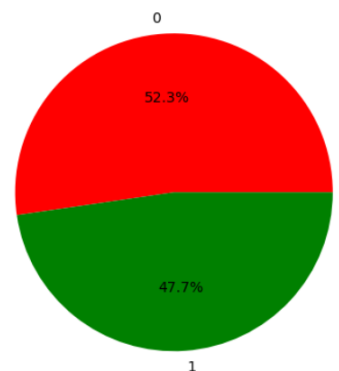
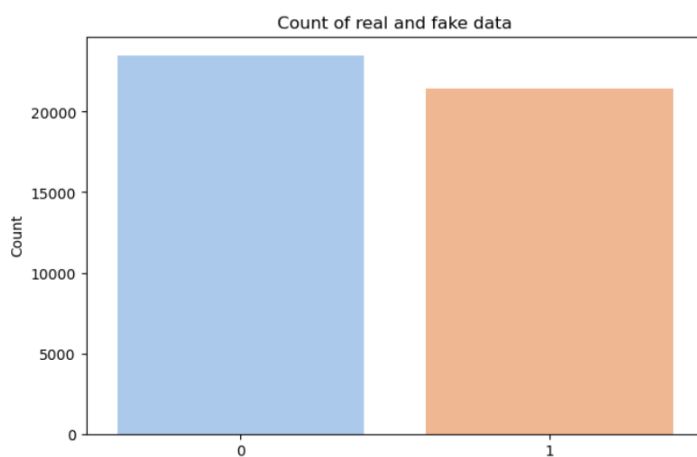
g1.set_title("Count of real and fake data")

g1.set_ylabel("Count")

g1.set_xlabel("Target")

g2=plt.pie(data["target"].value_counts().values,explode=[0,0],labels=data.
target.value_counts().index, ',colors=['red','green'])

fig.show()
```



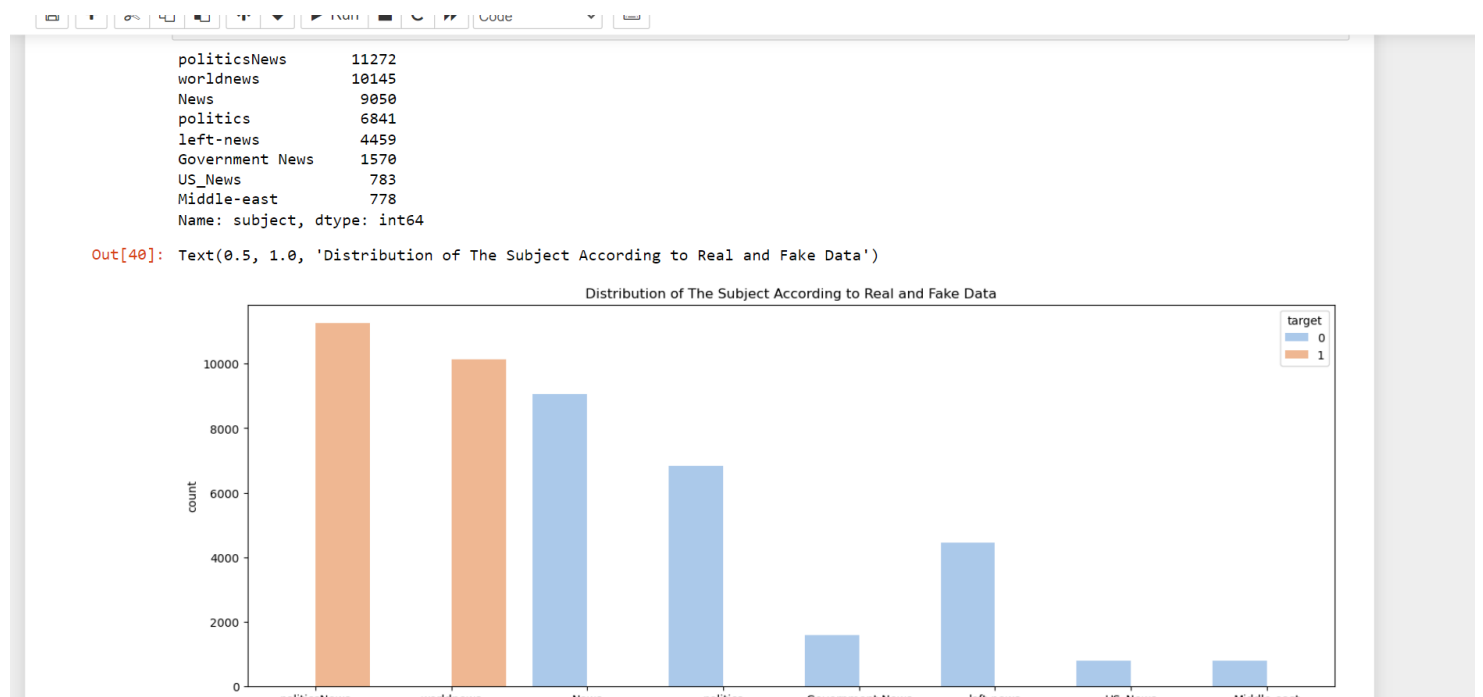
2. The Subject According to Real and Fake Data

```
print(data.subject.value_counts())
```

```
plt.figure(figsize=(17, 6))
```

```
ax = sns.countplot(x="subject", hue='target', data=data, palette="pastel")
```

```
plt.title("Distribution of The Subject According to Real and Fake Data")
```



3.Data cleaning:

```
data['text']= data['subject'] + " " + data['title'] + " " + data['text']
```

```
del data['title']
```

```
del data['subject']
```

```
del data['date']
```

```
data.head(10)
```

```
first_text = data.text[10]
```

```
first_text
```

Out[41]:

	text	target
0	politicsNews As U.S. budget fight looms, Repub...	1
1	politicsNews U.S. military to accept transgend...	1
2	politicsNews Senior U.S. Republican senator: '...	1
3	politicsNews FBI Russia probe helped by Austra...	1
4	politicsNews Trump wants Postal Service to cha...	1
5	politicsNews White House, Congress prepare for...	1
6	politicsNews Trump says Russia probe will be f...	1
7	politicsNews Factbox: Trump on Twitter (Dec 29...	1
8	politicsNews Trump on Twitter (Dec 28) - Globa...	1
9	politicsNews Alabama official to certify Senat...	1

9 politicsNews Alabama official to certify Senat... 1

In [48]: first_text = data.text[20]
first_text

Out[48]: "politicsNews Federal judge partially lifts Trump's latest refugee restrictions WASHINGTON (Reuters) - A federal judge in Seattle partially blocked U.S. President Donald Trump's newest restrictions on refugee admissions on Saturday, the latest legal defeat for his efforts to curtail immigration and travel to the United States. The decision by U.S. District Judge James Robart is the first judicial curb on rules the Trump administration put into place in late October that have contributed significantly to a precipitous drop in the number of refugees being admitted into the country. Refugees and groups that assist them argued in court that the administration's policies violated the Constitution and federal rulemaking procedures, among other claims. Department of Justice attorneys argued in part that U.S. law grants the executive branch the authority to limit refugee admissions in the way that it had done so. On Oct. 24, the Trump administration effectively paused refugee admissions from 11 countries mostly in the Middle East and Africa, pending a 90-day security review, which was set to expire in late January. The countries subject to the review are Egypt, Iran, Iraq, Libya, Mali, North Korea, Somalia, South Sudan, Sudan, Syria and Yemen. For each of the last three years, refugees from the 11 countries made up more than 40 percent of U.S. admissions. A Reuters review of State Department data showed that as the review went into effect, refugee admissions from the 11 countries plummeted. Robart ruled that the administration could carry out the security review, but that it could not stop processing or admitting refugees from the 11 countries in the meantime, as long as those refugees have a "bona fide" connection to the United States. As part of its new restrictions, the Trump administration had also paused a program that allowed for family reunification for refugees, pending further security screening procedures being put into place. Robart ordered the government to re-start the program, known as "follow-to-join". Approximately 2,000 refugees were admitted into the United States in fiscal year 2015 under the program, according to Department of Homeland Security data. Refugee advocacy groups praised Robart's decision. "This ruling brings relief to thousands of refugees in precarious situations in the Middle East and East Africa, as well as to refugees already in the U.S. who are trying to reunite with their spouses and children," said Mariko Hirose, litigation director for the International Refugee Assistance Project, one of the plaintiffs in the case. A Justice Department spokeswoman, Lauren Ehrsam, said the department disagrees with Robart's ruling and is "currently evaluating the next steps". Robart, who was appointed to the bench by Republican former President George W. Bush, emerged from relative obscurity in February, when he issued a temporary order to lift the first version of Trump's travel ban. On Twitter, Trump called him a "so-called judge" whose "ridiculous" opinion "essentially takes law-enforcement away from our country". Robart's ruling represented the second legal defeat in two days for the Trump administration. On Friday, a U.S. appeals court said Trump's travel ban targeting six Muslim-majority countries should not be applied to people with strong U.S. ties, but said its ruling would be put on hold pending a decision by the U.S. Supreme Court. "

In []:

Word Cloud for Real News:

PROGRAM:

```
from wordcloud import WordCloud, STOPWORDS
```

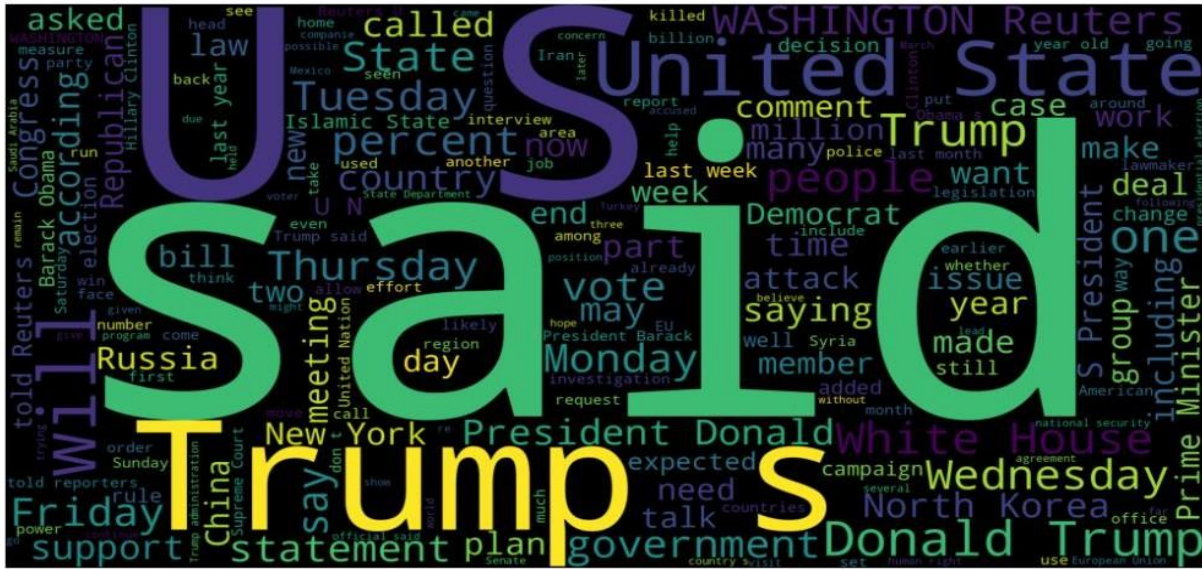
```
plt.figure(figsize = (15,15))
```

```
wc = WordCloud(max_words = 500 , width = 1000 ,
```

```
height = 500 stopwords = STOPWORDS)
```

```
.generate(" ".join(data[data.target == 1].text))
```

```
plt.imshow(wc , interpolation = 'bilinear')
```



```
CPU times: user 38.5 s, sys: 2.83 s, total: 41.3 s
Wall time: 41.4 s
```

Word Cloud for fake News:

```
plt.figure(figsize = (15,15))
```

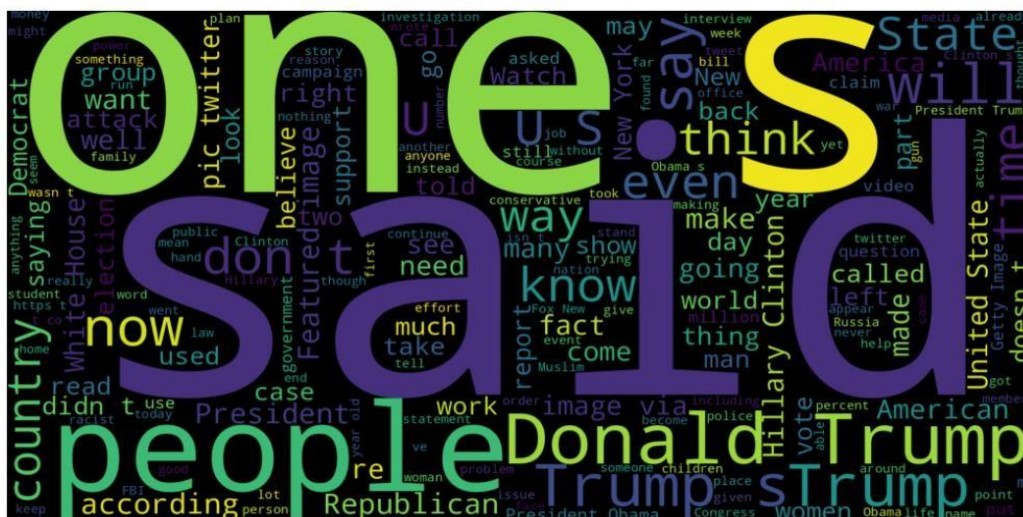
```
wc = WordCloud(max_words = 500, width = 1000,
```

height = 500 , stopwords =

```
STOPWORDS).generate(" ".join(data[data.target
```

```
== 0].text)plt.imshow(wc , interpolation = 'bilinear')
```

```
plt.show()
```



```
CPU times: user 43 s, sys: 3.33 s, total: 46.4 s
Wall time: 50.4 s
```

THANK YOU!