In [ ]:
```
FAKE NEWS DETECTION USING NLP

PHASE 03:LOADING AND PREPROCESSING THE DATASET

The dataset for Fake News Detection using NLP is:

https://www.kaggle.com/code/therealsampat/fake-news-detection#Fake-News-Detection

        In this phase, we will building a project by loading and preprocessing the dataset. Visit the Kaggle link provided above and download the dataset to t

With the explosion of online fake news and disinformation, it is increasingly difficult to discern fact from fiction. And as machine learning and natural langu

Google Cloud Natural Language API is a great platform to use for this project. Simply upload a dataset, train the model, and use it to predict new articles.

But before we download a Kaggle dataset and get cracking on Google Cloud, it's in our best interest to pre-process the dataset.

PREPROCESSING:

      To preprocess your text simply means to bring your text into a form that is predictable and analyzable for your task.

The goal of pre-processing is to remove noise. By removing unnecessary features from our text, we can reduce complexity and increase predictability (i.e. our m

APPROACH:

There are many types of text pre-processing and their approaches varied. We will cover the following:

      Removing columns

      Convert text to vectors

      Data Analysis

      Text Analysis

      Modeling

      Evaluation
```

In [ ]:
```
IMPORTING LIBRARIES
```

In [17]:
```
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
```

In [ ]:
```
IMPORTING DATASET
```

In [18]:
```
df_fake = pd.read_csv("Fake.csv")
df_true = pd.read_csv("True.csv")
```

In [6]:
```
df_fake.head()
```

Out[6]:

	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

In [19]:
```
df_true.head(5)
```

Out[19]:

	title	text	subject	date
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservati...	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

In [ ]:
```
DATA ANALYSIS
```

In [20]:
```
df_fake["class"] = 0
df_true["class"] = 1
```

In [21]:
```
df_fake.shape, df_true.shape
```

Out[21]:
```
((23481, 5), (21417, 5))
```

In [27]:
```
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)
```

In [28]:
```
df_fake.shape, df_true.shape
```

Out[28]:
```
((23471, 5), (21407, 5))
```

In [ ]:
```
TEST ANALYSIS
```

In [29]:
```
df_fake_manual_testing["class"] = 0
df_true_manual_testing["class"] = 1
```

```
C:\Users\CSE_BAY44\AppData\Local\Temp\ipykernel_5940\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-versus-a-copy
  df_fake_manual_testing["class"] = 0
C:\Users\CSE_BAY44\AppData\Local\Temp\ipykernel_5940\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-versus-a-copy
  df_true_manual_testing["class"] = 1
```

In [30]:
```
df_true_manual_testing.head(10)
```

Out[30]:

	title	text	subject	date	class
21407	Mata Pires, owner of embattled Brazil builder ...	SAO PAULO (Reuters) - Cesar Mata Pires, the ow...	worldnews	August 22, 2017	1
21408	U.S., North Korea clash at U.N. forum over nuc...	GENEVA (Reuters) - North Korea and the United ...	worldnews	August 22, 2017	1
21409	U.S., North Korea clash at U.N. arms forum on ...	GENEVA (Reuters) - North Korea and the United ...	worldnews	August 22, 2017	1
21410	Headless torso could belong to submarine journ...	COPENHAGEN (Reuters) - Danish police said on T...	worldnews	August 22, 2017	1
21411	North Korea shipments to Syria chemical arms a...	UNITED NATIONS (Reuters) - Two North Korean sh...	worldnews	August 21, 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

In [31]:
```
df_manual_testing = pd.concat([df_fake_manual_testing,df_true_manual_testing], axis = 0)
df_manual_testing.to_csv("manual_testing.csv")
```

In [32]:
```
df_merge = pd.concat([df_fake, df_true], axis =0 )
df_merge.head(10)
```

Out[32]:

	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

In [33]:
```
df_merge.columns
```

Out[33]:
```
Index(['title', 'text', 'subject', 'date', 'class'], dtype='object')
```

In [34]:
```
df = df_merge.drop(["title", "subject","date"], axis = 1)
```

In [35]:
```
df.isnull().sum()
```

Out[35]:
```
text     0
class    0
dtype: int64
```

RANDOM SHUFFLING THE DATAFRAMES

In [36]:
```
df = df.sample(frac = 1)
```

In [37]:
```
df.head()
```

Out[37]:

	text	class
13677	HARARE (Reuters) - Robert Mugabe s 37-year rul...	1
1839	WASHINGTON (Reuters) - Treasury Secretary Stev...	1
6913	(Reuters) - U.S. Republican President-elect Do...	1
13968	Says the guy who was a Jr. Senator and Communi...	0
1000	GENEVA (Reuters) - President Bashar al-Assad a...	1

In [38]:
```
df.reset_index(inplace = True)
df.drop(["index"], axis = 1, inplace = True)
```

In [39]:
```
df.columns
```

Out[39]:
```
Index(['text', 'class'], dtype='object')
```

In [44]:
```
df.head()
```

Out[44]:

	text	class
0	HARARE (Reuters) - Robert Mugabe s 37-year rul...	1
1	WASHINGTON (Reuters) - Treasury Secretary Stev...	1
2	(Reuters) - U.S. Republican President-elect Do...	1
3	Says the guy who was a Jr. Senator and Communi...	0
4	GENEVA (Reuters) - President Bashar al-Assad a...	1

In [46]:
```
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
```

In [47]:
```
df["text"] = df["text"].apply(wordopt)
```

In [48]:
```
x = df["text"]
y = df["class"]
```

In [49]:
```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25)
```

In [50]:
```
from sklearn.feature_extraction.text import TfidfVectorizer

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

In [51]:
```
from sklearn.linear_model import LogisticRegression

LR = LogisticRegression()
LR.fit(xv_train,y_train)
```

Out[51]:
```
LogisticRegression()
```

In [52]:
```
pred_lr=LR.predict(xv_test)
```

In [53]:
```
LR.score(xv_test, y_test)
```

Out[53]:
```
0.9829119254583851
```

In [54]:
```
print(classification_report(y_test, pred_lr))
```

```
              precision    recall  f1-score   support

           0       0.99      0.97      0.98      4443
           1       0.98      0.99      0.98      5217

    accuracy                           0.98      9660
   macro avg       0.98      0.98      0.98      9660
weighted avg       0.98      0.98      0.98      9660
```

In [57]:
```
from sklearn.tree import DecisionTreeClassifier

DT = DecisionTreeClassifier()
DT.fit(xv_train, y_train)
```

Out[57]:
```
DecisionTreeClassifier()
```

In [58]:
```
pred_dt = DT.predict(xv_test)
```

In [59]:
```
DT.score(xv_test, y_test)
```

Out[59]:
```
0.9943064182194618
```

In [60]:
```
print(classification_report(y_test, pred_dt))
```

```
              precision    recall  f1-score   support

           0       1.00      0.99      0.99      4443
           1       0.99      1.00      0.99      5217

    accuracy                           0.99      9660
   macro avg       0.99      0.99      0.99      9660
weighted avg       0.99      0.99      0.99      9660
```

In [ ]:
```
from sklearn.ensemble import GradientBoostingClassifier

GBC = GradientBoostingClassifier(random_state=0)
GBC.fit(xv_train, y_train)
```

In [ ]:
```
pred_gbc = GBC.predict(xv_test)
```

In [ ]:
```
GBC.score(xv_test, y_test)
```

In [ ]:
```
print(classification_report(y_test, pred_gbc))
```

In [ ]:
```
from sklearn.ensemble import RandomForestClassifier

RFC = RandomForestClassifier(random_state=0)
RFC.fit(xv_train, y_train)
```

In [ ]:
```
pred_rfc = RFC.predict(xv_test)
```

In [ ]:
```
RFC.score(xv_test, y_test)
```

In [ ]:
```
print(classification_report(y_test, pred_rfc))
```

In [ ]:
```

```

In [ ]: