```
# Import the Dependencies


import pandas as pd
import numpy as np
import re
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score

import nltk
nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

```
True
```

```
# Printing stopwords in English
print(stopwords.words('english'))
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you',
"you're", "you've", "you'll", "you'd", 'your', 'yours', 'yourself',
'yourselves', 'he', 'him', 'his', 'himself', 'she', "she's", 'her',
'hers', 'herself', 'it', "it's", 'its', 'itself', 'they', 'them',
'their', 'theirs', 'themselves', 'what', 'which', 'who', 'whom',
'this', 'that', "that'll", 'these', 'those', 'am', 'is', 'are', 'was',
'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do',
'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or',
'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with',
'about', 'against', 'between', 'into', 'through', 'during', 'before',
'after', 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out',
'on', 'off', 'over', 'under', 'again', 'further', 'then', 'once',
'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both',
'each', 'few', 'more', 'most', 'other', 'some', 'such', 'no', 'nor',
'not', 'only', 'own', 'same', 'so', 'than', 'too', 'very', 's', 't',
'can', 'will', 'just', 'don', "don't", 'should', "should've", 'now',
'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren', "aren't",
'couldn', "couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn',
"hadn't", 'hasn', "hasn't", 'haven', "haven't", 'isn', "isn't", 'ma',
'mightn', "mightn't", 'mustn', "mustn't", 'needn', "needn't", 'shan',
"shan't", 'shouldn', "shouldn't", 'wasn', "wasn't", 'weren',
"weren't", 'won', "won't", 'wouldn', "wouldn't"]
```

```
# Load the dataset to a pandas dataframe
news_dataset = pd.read_csv('train.csv')
```

```
# Check the number of rows and columns
news_dataset.shape
```

```
(20800, 5)
```

```
# Print the first 5 rows of this dateframe
news_dataset.head()
```


```
{"summary":"{\n  \"name\": \"news_dataset\",\n  \"rows\": 20800,\n
\"fields\": [\n    {\n      \"column\": \"id\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\": 6004,\n
\"min\": 0,\n        \"max\": 20799,\n        \"num_unique_values\":
20800,\n        \"samples\": [\n          14649,\n          9231,\n
6473\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"title\",\n      \"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 19803,\n        \"samples\": [\n
\"Florida Republicans Voting for Hillary Clinton \\u00b7 Guardian
Liberty Voice\",\n          \"\\u2018Weekend Warriors\\u2019 Show
Survival Benefits - The New York Times\",\n          \"Part 6,
Democratic Network Money: Commons Money that Works for Us, Instead of
Us Working for It\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"author\",\n      \"properties\": {\n        \"dtype\":
\"category\",\n        \"num_unique_values\": 4201,\n
\"samples\": [\n          \"AARGH63\",\n          \"Dave Kehr\",\n
\"Patrick Healy\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"text\",\n      \"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 20386,\n        \"samples\": [\n
\"Multan a un ultrasur por lanzar a un jugador del Bar\\u00e7a al
campo durante el partido Madrid-Legia EL PARTIDO SE CONSIDERABA DE
ALTO RIESGO PARA LOS JUGADORES DEL BARCELONA Real Madrid \\nUn jugador
barcelonista golpe\\u00f3 a varios jugadores durante el encuentro de
la Champions que tuvo lugar ayer entre el Real Madrid y el Legia
Varsovia despu\\u00e9s de ser arrojado desde la grada por un hincha
madridista. \\nEl portero del Legia, muy molesto, intent\\u00f3
devolver el jugador a las gradas de una patada y se encar\\u00f3 con
los Ultrasur, que ya preparaban a otro jugador para echarlo al campo y
entorpecer el encuentro. En esta ocasi\\u00f3n pretend\\u00edan
prenderlo con un mechero antes de lanzarlo al c\\u00e9sped. \\nLas
autoridades han vuelto a pedir a los aficionados que dejen de lanzar
cosas al campo como botellas, monedas y jugadores del Barcelona. El
dispositivo de seguridad no puedo hacer nada por evitar que algunos
hinchas introdujeran a jugadores barcelonistas a la grada, pues los
llevaban camuflados dentro de bocadillos de gran tama\\u00f1o. \\nEl
partido pudo ser reanudado cuando el \\u00e1rbitro apart\\u00f3 a un
rinc\\u00f3n del c\\u00e9sped al jugador barcelonista.\",\n
\"Written by Jacob G. Hornberger   As US officials continue to accuse
Russia of meddling with the US presidential election, an accusation
```

that they have provided no evidence whatsoever to support, let\\u2019s review some of the US government\\u2019s history of meddling with elections in others countries.1. In 1951, the democratically elected parliament of Iran elected a man named Mohammad Mossadegh to be Iran\\u2019s prime minister. Mossadegh angered British Empire officials by nationalizing British oil interests in the country.British officials then turned to the CIA for assistance. In 1953, the CIA secretly fomented a violent coup in Iran, which succeeded in ousting Mossadegh from power and making the Shah of Iran the supreme unelected dictator of the country.To fortify the Shah\\u2019s dictatorial hold on power, the CIA helped organize and train the his domestic police force, the Savak, which was essentially a combination of the CIA, the NSA, and the military. Part of the CIA\\u2019s training involved teaching Savak agents the art of torture.For the next 26 years, the Iranian people suffered under one of the most brutal and tyrannical dictatorships in the world, one that US officials fully supported and called an ally and friend of the United States.In 1979, Iranians successfully revolted against the Shah\\u2019s regime and ousted him from power. One result was not a restoration of the democratic system that had elected Mossadegh but rather another brutal dictatorship, this time a religious one. Another result is the bad relations between the Iran and US governments that continues to exist today.2. In 1951, the Guatemalan people democratically elected a man named Jacobo Arbenz to be their president. Arbenz, however, was not satisfactory to US officials, especially the national-security branch of the government, specifically the Pentagon and the CIA. The reason that US officials opposed Arbenz was that he was a socialist, and US officials considered a socialist president of Guatemala to be a threat to \\u201cnational security\\u201d here in the United States.In 1954 \\u2014 one year after the coup in Iran, the CIA fomented a violent military coup that succeeded in removing Arbenz from power and replacing him with one of the most brutal unelected military dictators in Latin American history, a man named Carlos Castillo Armas. The CIA had a kill list prepared for the coup, which Arbenz was able to escape by fleeing the country before Castillo was able to get him. The CIA\\u2019s destruction of Guatemala\\u2019s democratic system threw the nation into a 30-year civil war that ended up killing millions of Guatemalan people, especially many of the poor.3. In 1960 a man named Patrice Lamumba was elected Congo\\u2019s first prime minister after independence from Belgium. Lamumba spoke out against Western imperialism and refused to take sides in the Cold War, which caused the CIA to conclude that he was a threat to \\u201cnational security.\\u201d The CIA orchestrated the assassination of Lamumba, which ended up taking place on January 17, 1961, just three days before President Kennedy, who liked Lamumba and who would have ordered the CIA to stand down, was to be sworn into office.4. In 1970 a man named Salvador Allende received a plurality of votes in the presidential election in Chile. Pursuant to the Chilean constitution, the election was thrown into the national congress. President Richard

Nixon, his national-security team, the Pentagon, and the CIA concluded that because Allende believed in communism and socialism, he posed a grave threat to \\u201cnational security\\u201d here in the United States. The CIA attempted to bribe members of the congress to vote against Allende. It also orchestrated the kidnapping of the head of Chile\\u2019s armed forces, Gen Rene Schneider, who opposed a US military coup in his country, especially since a coup would violate the country\\u2019s constitution. The kidnapping attempt on Schneider left him dead.The CIA then fomented a coup that took place on 9/11 1973 that violently ousted Allende from power and left him dead. Replacing him was army Gen. Augusto Pinochet, one of the most brutal unelected military dictators in history. By the time Pinochet\\u2019s 17-year reign of military terror came to an end in 1990, he and his CIA-supported goons had incarcerated, raped, tortured, or killed tens of thousands of innocent people \\u2014 that is, people whose only \\ u201ccrime\\u201d was believing in socialism \\u2014 with the full support of the CIA, Pentagon, Nixon, and his \\u201cnational security\\u201d team.Of course, there are also the more recent support of regime-change operations that ousted democratically elected presidents that the US government disapproved of, such as in Ukraine and Egypt.And then there is the long list of countries where unelected dictators were targeted for regime change by the US national security state and, where successful, replaced with a brutal unelected pro-US dictator. Cuba, Indonesia, Iraq, Libya, and Syria all come to mind.US officials need to keep in mind that when they point their accusatory index finger at Russia for supposedly meddling in the US presidential election, US officials have, at the same time, three fingers pointing back at themselves. Reprinted with permission from the Future of Freedom Foundation . Related\",\n   \"When hearing of an establishment called the Black and Blue Steakhouse, one could reasonably surmise that some, or at the very least one, extremely tough person could be found somewhere in there. Well, the three terrorists who attacked London Bridge over the weekend found the Black and Blue Steakhouse, and they found the toughest man in the house. [Or, maybe more to the point, he found them. Roy Larner, 47, is a huge fan of the South East London soccer team Millwall F. C. The team is known as the Lions, and when terrorists entered the Black and Blue to kill innocent people, Larner let them know who is the true king of the jungle. After the terrorists entered, with obvious bad intentions, Larner sprang into action: \\u201cLike an idiot I shouted back at them. I thought, \\u2018I need to take the p*** out of these b*****s.\\u2019 \\u201cI took a few steps towards them and said, \\ u2018F*** you, I\\u2019m Millwall.\\u2019 So they started attacking me. \\u201cI stood in front of them trying to fight them off. Everyone else ran to the back. I was on my own against all three of them, that\\u2019s why I got hurt so much. It was just me, trying to grab them with my bare hands and hold on. I was swinging. \\u201cI got stabbed and sliced eight times. They got me in my head, chest and both hands. There was blood everywhere. \\u201cThey were saying, \\

```
u2018Islam, Islam!\\u2019 I said again, \\u2018F*** you, I\\u2019m
Millwall!\\u2019 \\u201cIt was the worst thing I could have done as
they carried on attacking me. \\u201cLuckily, none of the blows were
straight at me or I\\u2019d be dead. \\u201d A petition has begun,
with the aim of awarding Larner with the George Cross for bravery. The
award is the second highest commendation granted by Great Britain.
Looking at his scars, it\\u2019s really hard to imagine that he won\\
u2019t get it. If Millwall is looking for a new slogan, they may have
found it. \\u201cF*** you, I\\u2019m Millwall!\\u201d has a nice ring
to it. Follow Dylan Gwinn on Twitter: @themightygwinn\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"label\",\n      \"properties\": {\
n        \"dtype\": \"number\",\n        \"std\": 0,\n        \"min\":
0,\n        \"max\": 1,\n        \"num_unique_values\": 2,\n
\"samples\": [\n          0,\n          1\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe","variable_name":"news_dataset"}
```

```
# Checking for missing values
news_dataset.isnull().sum()
```

```
id           0
title      558
author    1957
text        39
label        0
dtype: int64
```

```
# Replacing null values with empty strings
news_dataset = news_dataset.fillna('')
```

```
# Merging the author name and news tite
news_dataset['content'] = news_dataset['author']+ ''+
news_dataset['title']
```

```
print(news_dataset['content'])
```

```
0        Darrell LucusHouse Dem Aide: We Didn’t Even Se...
1        Daniel J. FlynnFLYNN: Hillary Clinton, Big Wom...
2        Consortiumnews.comWhy the Truth Might Get You ...
3        Jessica Purkiss15 Civilians Killed In Single U...
4        Howard PortnoyIranian woman jailed for fiction...
                              ...
20795    Jerome HudsonRapper T.I.: Trump a ’Poster Chil...
20796    Benjamin HoffmanN.F.L. Playoffs: Schedule, Mat...
20797    Michael J. de la Merced and Rachel AbramsMacy’...
20798    Alex AnsaryNATO, Russia To Hold Parallel Exerc...
20799                David SwansonWhat Keeps the F-35 Alive
Name: content, Length: 20800, dtype: object
```

```
# Separating the data & label
X = news_dataset.drop(columns='label', axis=1)
Y = news_dataset['label']

print(X)
print(Y)
```

```
          id  ...                                            content
0          0  ...  Darrell LucusHouse Dem Aide: We Didn’t Even Se...
1          1  ...  Daniel J. FlynnFLYNN: Hillary Clinton, Big Wom...
2          2  ...  Consortiumnews.comWhy the Truth Might Get You ...
3          3  ...  Jessica Purkiss15 Civilians Killed In Single U...
4          4  ...  Howard PortnoyIranian woman jailed for fiction...
...      ...  ...                                                ...
20795  20795  ...  Jerome HudsonRapper T.I.: Trump a ’Poster Chil...
20796  20796  ...  Benjamin HoffmanN.F.L. Playoffs: Schedule, Mat...
20797  20797  ...  Michael J. de la Merced and Rachel AbramsMacy’...
20798  20798  ...  Alex AnsaryNATO, Russia To Hold Parallel Exerc...
20799  20799  ...            David SwansonWhat Keeps the F-35 Alive

[20800 rows x 5 columns]
0        1
1        0
2        1
3        1
4        1
        ..
20795    0
20796    0
20797    0
20798    1
20799    1
Name: label, Length: 20800, dtype: int64
```

```
# Stemming Procedure
port_stem = PorterStemmer()

def steming(content):
  stemmed_content = re.sub('[^a-zA-z]','', content)
  stemmed_content = stemmed_content.lower()
  stemmed_content = stemmed_content.split()
  stemmed_content = [port_stem.stem(word) for word in stemmed_content
if not word in stopwords.words('english')]
  stemmed_content = ''.join(stemmed_content)
  return stemmed_content

news_dataset['content'] = news_dataset['content'].apply(steming)

print(news_dataset['content'])
```

```
0        darrelllucushousedemaidewedidntevenseecomeysle...
1        danieljflynnflynnhillaryclintonbigwomanoncampu...
2               consortiumnewscomwhythetruthmightgetyoufir
3        jessicapurkisscivilianskilledinsingleusairstri...
4        howardportnoyiranianwomanjailedforfictionalunp...
                               ...
20795    jeromehudsonrappertitrumpaposterchildforwhites...
20796    benjaminhoffmannflplayoffsschedulematchupsando...
20797    michaeljdelamercedandrachelabramsmacysissaidto...
20798    alexansarynatorussiatoholdparallelexercisesinb...
20799                           davidswansonwhatkeepsthefal
Name: content, Length: 20800, dtype: object
```

```
# Create X and Y
X = news_dataset['content'].values
Y = news_dataset['label'].values

print(X)
print(Y)
```

```
['darrelllucushousedemaidewedidntevenseecomeysletteruntiljasonchaffetz
tweetedit'
 'danieljflynnflynnhillaryclintonbigwomanoncampusbreitbart'
 'consortiumnewscomwhythetruthmightgetyoufir' ...

'michaeljdelamercedandrachelabramsmacysissaidtoreceivetakeoverapproach
byhudsonsbaythenewyorktim'
 'alexansarynatorussiatoholdparallelexercisesinbalkan'
 'davidswansonwhatkeepsthefal']
[1 0 1 ... 0 1 1]
```

```
Y.shape
```

```
(20800,)
```

```
# Converting textual data to numerical data
vectorizer = TfidfVectorizer()
vectorizer.fit(X)

X = vectorizer.transform(X)

print(X)
```

```
  (0, 5210)      1.0
  (1, 4898)      1.0
  (2, 4551)      1.0
  (3, 9761)      1.0
  (4, 7983)      1.0
  (5, 4996)      1.0
  (6, 12013)     1.0
  (7, 1010)      1.0
```

```
  (8, 6809)      1.0
  (9, 12909)     1.0
  (10, 85) 1.0
  (11, 4203)     1.0
  (12, 1165)     0.72122843239707
  (12, 15651)    0.6926972991877945
  (13, 9115)     1.0
  (14, 1697)     1.0
  (15, 8745)     1.0
  (16, 13043)    1.0
  (17, 17480)    1.0
  (18, 18140)    1.0
  (19, 9289)     1.0
  (20, 14194)    1.0
  (21, 9671)     1.0
  (22, 15118)    1.0
  (23, 12718)    1.0
  :   :
  (20775, 1823)  1.0
  (20776, 8939)  1.0
  (20777, 409)   1.0
  (20778, 14005)1.0
  (20779, 18541)1.0
  (20780, 4167)  1.0
  (20781, 7757)  1.0
  (20782, 18172)1.0
  (20783, 3728)  1.0
  (20784, 5554)  1.0
  (20785, 1572)  1.0
  (20786, 7472)  1.0
  (20787, 8253)  1.0
  (20788, 9871)  1.0
  (20789, 19697)1.0
  (20790, 288)   1.0
  (20791, 4865)  1.0
  (20792, 10248)1.0
  (20793, 16444)1.0
  (20794, 11876)1.0
  (20795, 9686)  1.0
  (20796, 2366)  1.0
  (20797, 13154)1.0
  (20798, 890)   1.0
  (20799, 5476)  1.0
```

```
# Splitting the dataset training & testing data
X_train, X_test, Y_train, Y_test = train_test_split(X,Y,
test_size=0.2, stratify=Y, random_state=2)

# Training the model: Logistic regression model
model = LogisticRegression()
```

```
model.fit(X_train, Y_train)
```

```
LogisticRegression()
```

```
# Evaluation - accuracy score
# Accuracy_score on training data
X_train_prediction = model.predict(X_train)
training_data_accuracy = accuracy_score(X_train_prediction, Y_train)

print('Accuracy score of the training data:',training_data_accuracy)
```

```
Accuracy score of the training data: 0.9927283653846154
```

```
# Evaluation - Accuracy score on test data
X_test_prediction = model.predict(X_test)
test_data_accuracy = accuracy_score(X_test_prediction, Y_test)

print('Accurcay score of the testing data:', test_data_accuracy)
```

```
Accurcay score of the testing data: 0.5377403846153846
```

```
# Making a Prediction System
X_new = X_test[0]

prediction = model.predict(X_new)
print(prediction)

if (prediction[0]==0):
  print('The news is REAL')
else:
  print('The news is FAKE')
```

```
[0]
The news is REAL
```

```
print(Y_test[3])
```

```
0
```

```
# Plot the graph for fake news
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

# Convert X_test to a dense array & take first feature for plotting
X_test_dense =X_test.toarray()
X_test_plot = X_test_dense[:,0]

#Plotting the scatter plot
plt.scatter(X_test_plot,Y_test)
plt.xlabel(' First feature of X_test')
plt.ylabel('Y_test (Labels)')
```

```
plt.title('Scatter plot of fake news')
plt.show()
```