PYTHON PROJECT

PROJECT NAME: FAKE NEWS RECOGNITION

SUBMITTING TO:

Ms. Mallika Srivastava
Trainer,Programming & Algorithms,
Eisystems Services

SUBMITTING BY:

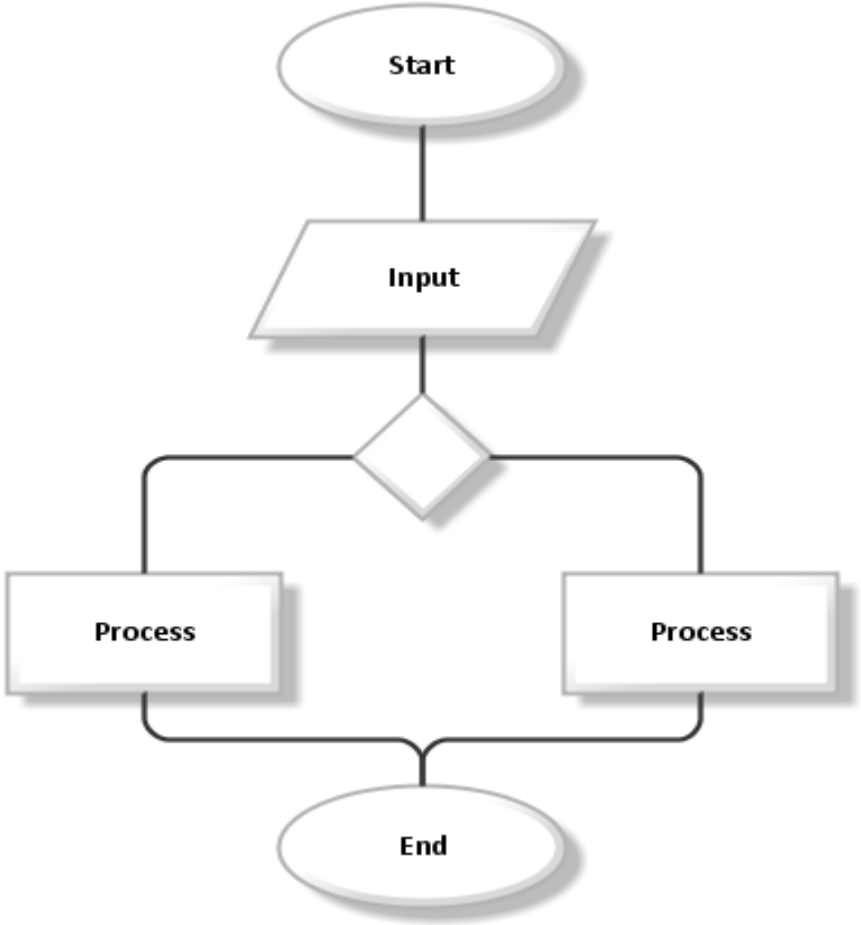
Aakanksha Mourya
Btech.CSE
Bundelkhand University, Jhansi

10	**Input output dataset**	
11	**Text Code**	
12	**References**	

1. Details of project Developed:

   Project is based on Machine Learning.Machine learning is a branch of AI and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

2. Data Flow Diagram:



Abstract Of Project

This paper presents the development of an intelligent certificate verification system for fraud detection using machine learning technique. The research was
embarked upon after noticing the rate of document forgery in the Nigerian society. Thus an exhaustive review of literatures was made which identified the
challenges public and private institutions encounter due to lack of automated means to verify any legal document. The flaws in the conventional verification

system such as delay time, cumbersome, high cost, lack of intelligence and above all, not being reliable; have been exploited over the years by fraudsters to
fabricate fake documents such as certificates mostly and commit fraud. This research seeks to address the problems via the development of a machine learning
based verification system and localizing it for the verification of certificate at the Nnamdi Azikiwe University (Unizik), Awka, Nigeria. To achieve this, the
methods of data collection, data acquisition, data processing, feature extraction, artificial neural network, training, and classification were used. Self defining
equations and modeling diagrams were used to develop the artificial neural network model and then train with 1180 authorized data collection of Unizik
certificates from 2016 to 2020, to generate the reference verification model which was used to develop the expert system for verification of documents. The
system was implemented using image acquisition toolbox, image processing toolbox, statistical and feature extraction toolbox, neural network toolbox, Matlab
and then tested for evaluation. The result recorded however, achieved a Mean Square Error (MSE) performance of 0.000100Mu and Regression value of R=
0.99373 which is very good, with implication that the new system is very reliable.

Project Summary

The study has successfully presented an intelligent system for the verification of documents using artificial intelligence technique. This was done to

combat the increased fraud and the rate at which documents are fabricated and manipulated all over the world today for many selfish reasons, in order
to gain wealth, fame, employment, among others. This study collected data from the bursary department, Unizik; and then developed an intelligent
document verification algorithm and deployed as an expert system using Matlab. The system was tested and the result showed hi gh rate of verification
regression and MSE performance. The implication is that when deployed, it was able to recognize and verify results accurately.

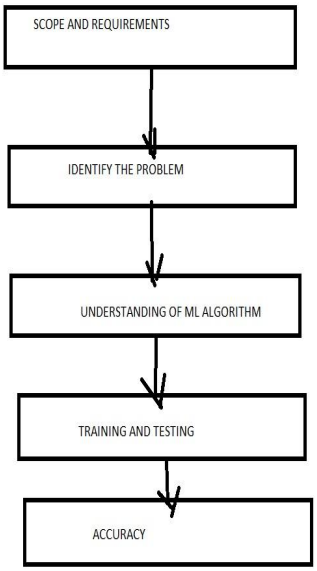
Objectives Of Project

1. Digital certificate that adopts digital signature technology provides the authority to validate the user himself in the digital fields used to validate the identity of a user and the authorisation to access the network resources.
2. This provides employers with clarity to check workers' educational credentials during the recruiting process and saves time for the review of educational documents.

Details Of Project Developed

The software that we implement first scanned the QR-code of the document and the sign, stamp and logo of the document

using Image processing techniques in deep learning.The Image Processing Module basically includes of two parts:Error Level Analysis and NeuralNetwork.These parts in combination help to detect whether the document image is manipulated by any means ornot.Deployment phase of the system is the main part that is how the system is to be used in the real life.



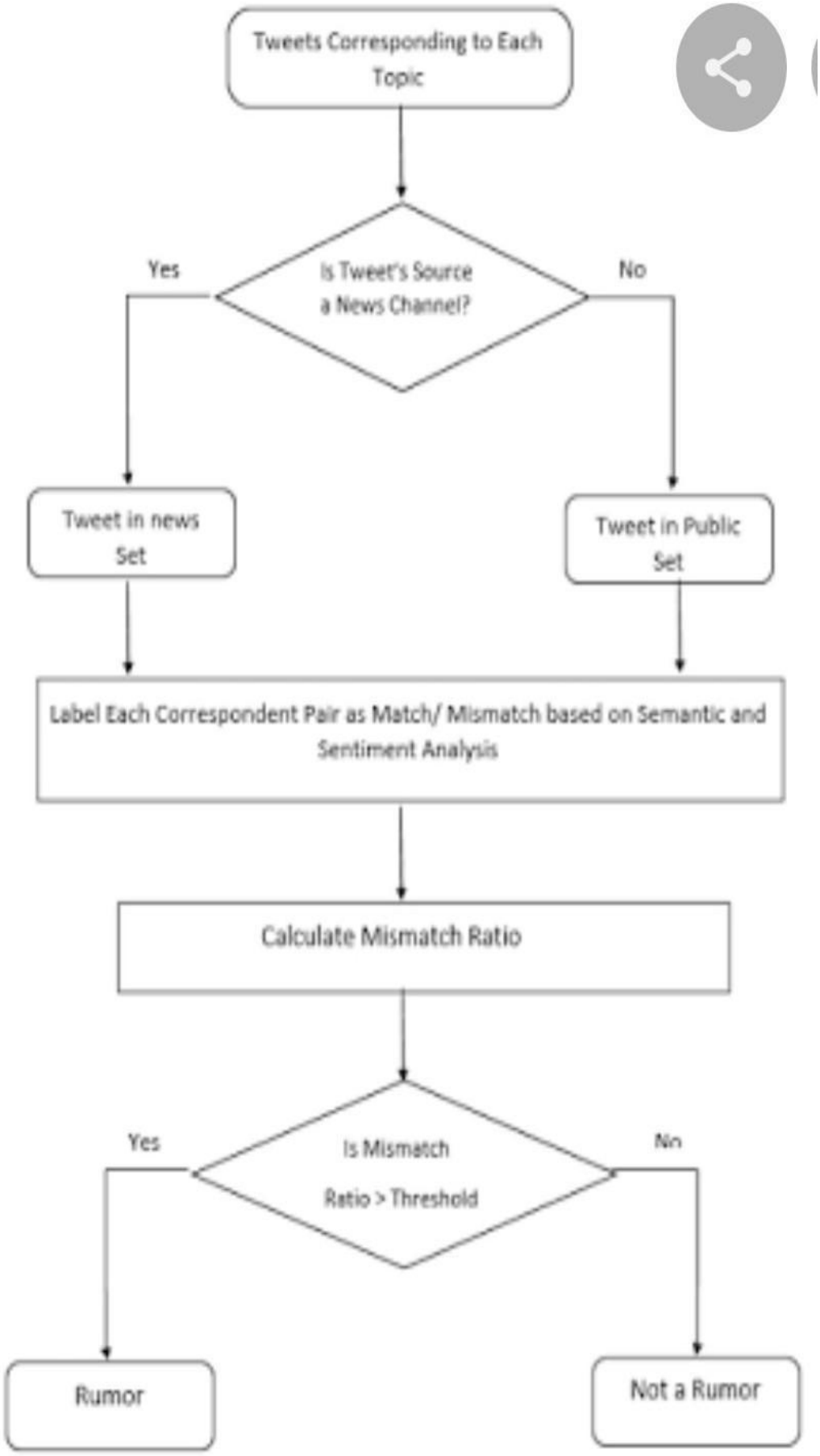
Details Of Project Developed

SYSTEM REQUIREMENT USED:

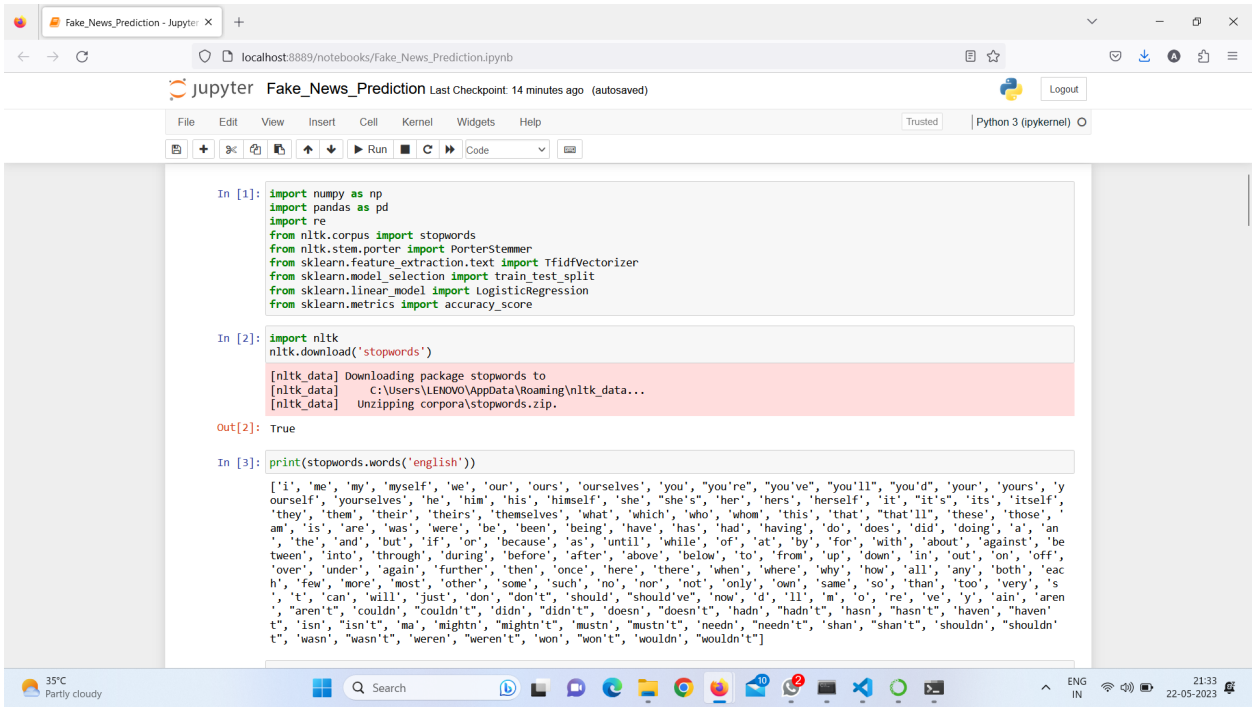
1.WINDOWS 11

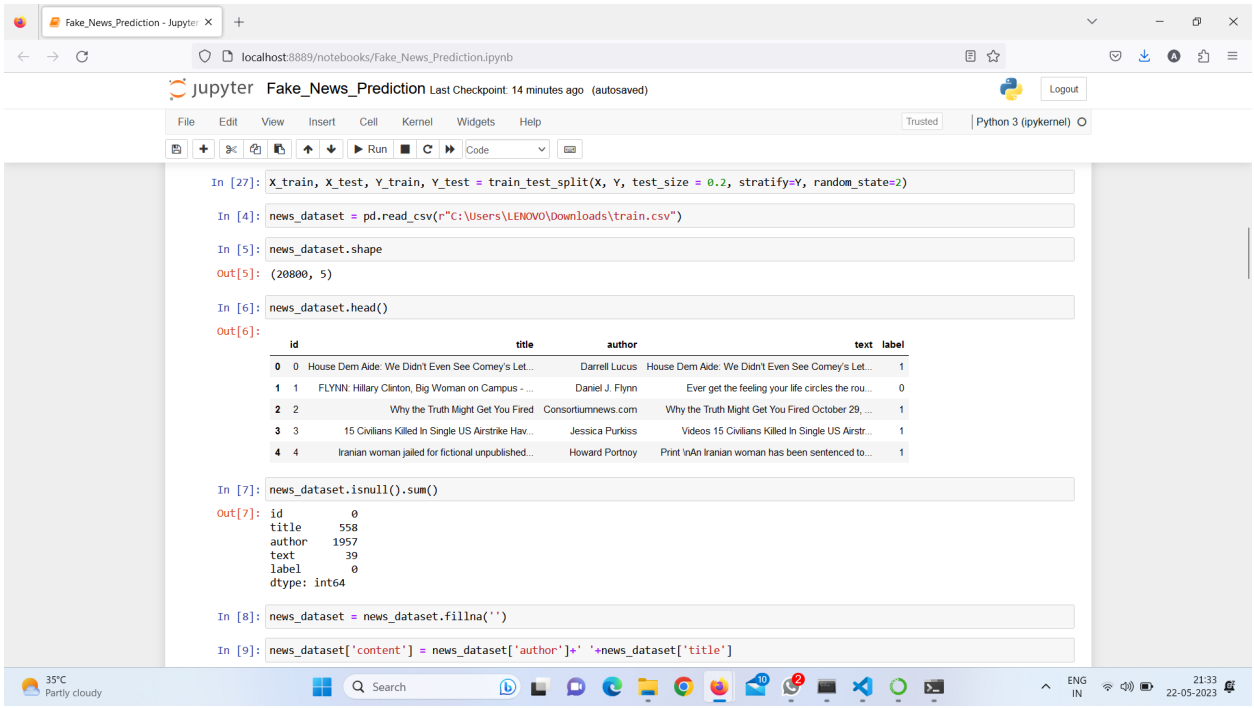
2.JUPYTER NOTEBOOK
3. PYTHON3

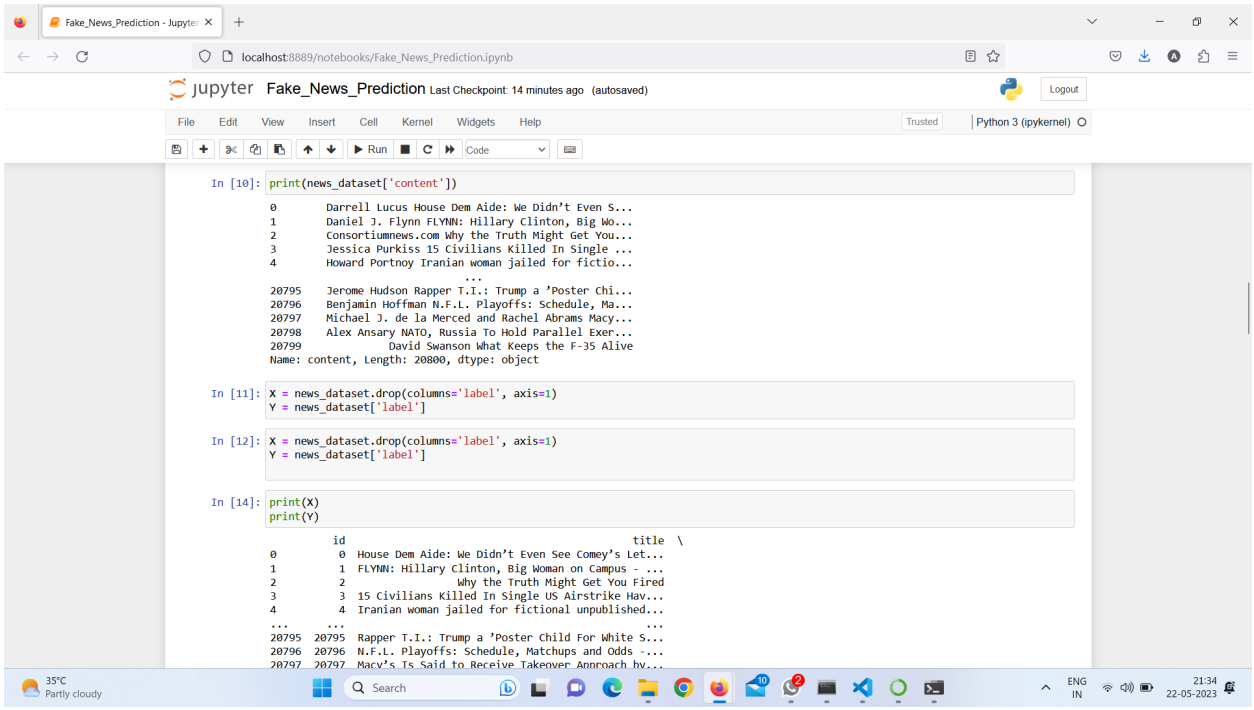
Data Flow Diagram Algorithm

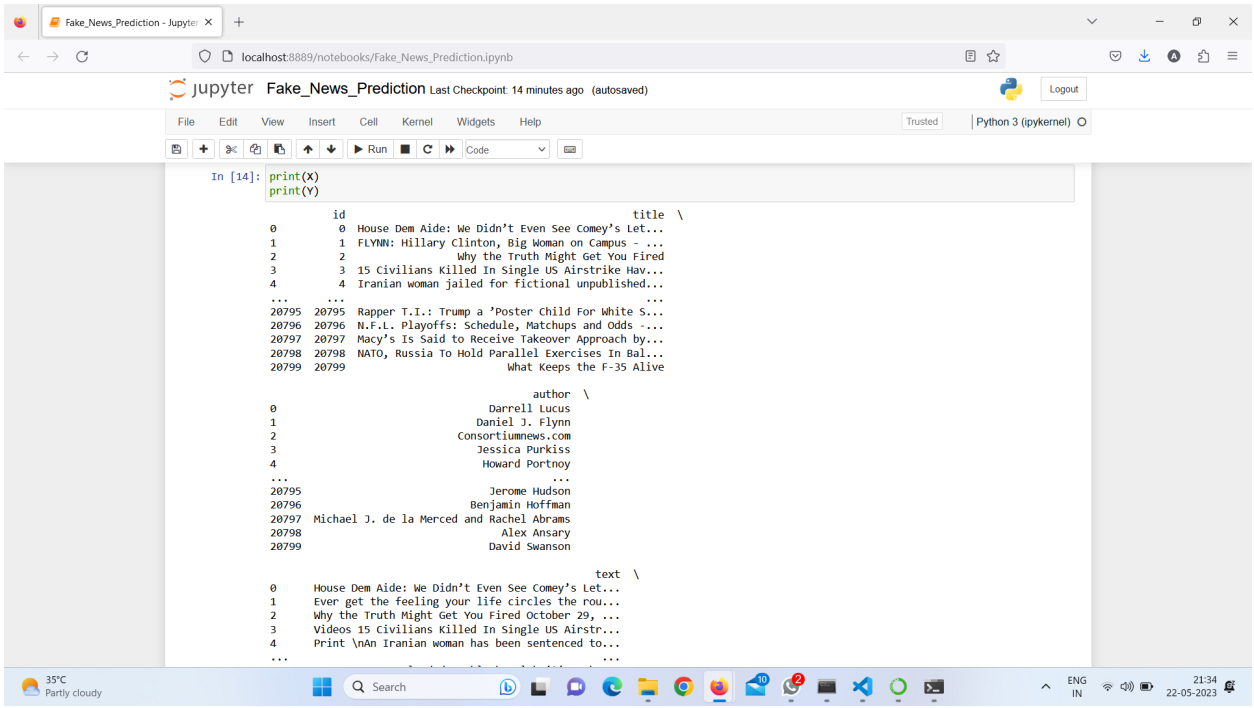

Tweets Corresponding to Each Topic
Is Tweet's Source a News Channel?
Yes
No
Tweet in news Set
Tweet in Public Set
Label Each Correspondent Pair as Match/ Mismatch based on Semantic and Sentiment Analysis
Calculate Mismatch Ratio
Is Mismatch Ratio > Threshold
Yes
No
Rumor
Not a Rumor

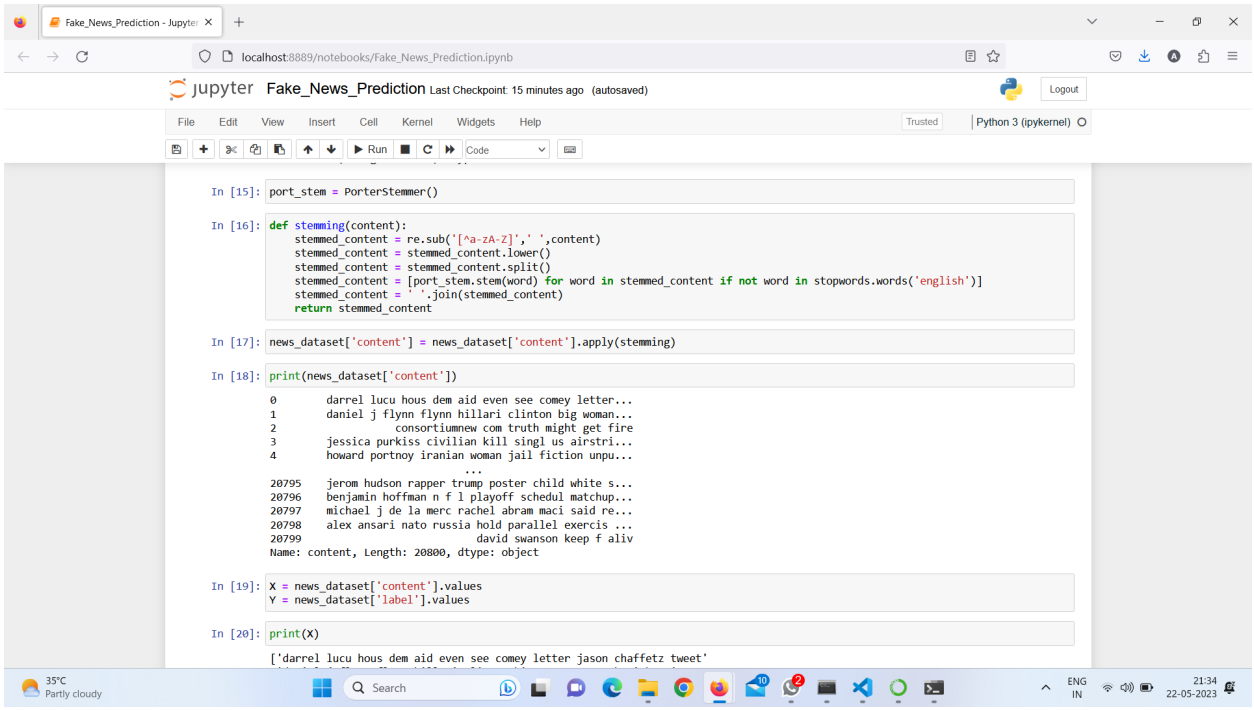
Text Code:

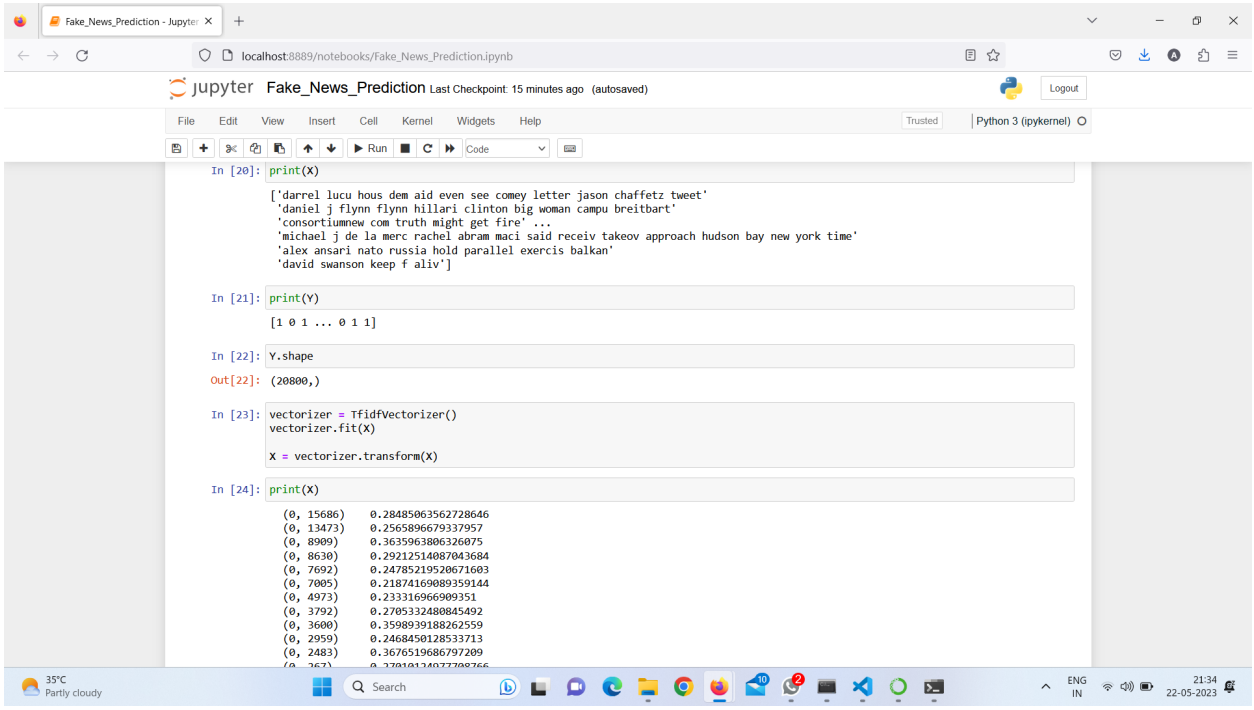

Fake_News_Prediction - Jupyter
localhost:8889/notebooks/Fake_News_Prediction.ipynb
jupyter Fake_News_Prediction Last Checkpoint: 14 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [1]:
import numpy as np
import pandas as pd
import 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
In [2]:
import nltk
nltk.download('stopwords')
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\LENOVO\AppData\Roaming\nltk_data...
[nltk_data] Unzipping corpora\stopwords.zip.
Out[2]:
True
In [3]:
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"]
35°C
Partly cloudy
Search
ENG
IN
21:33
22-05-2023

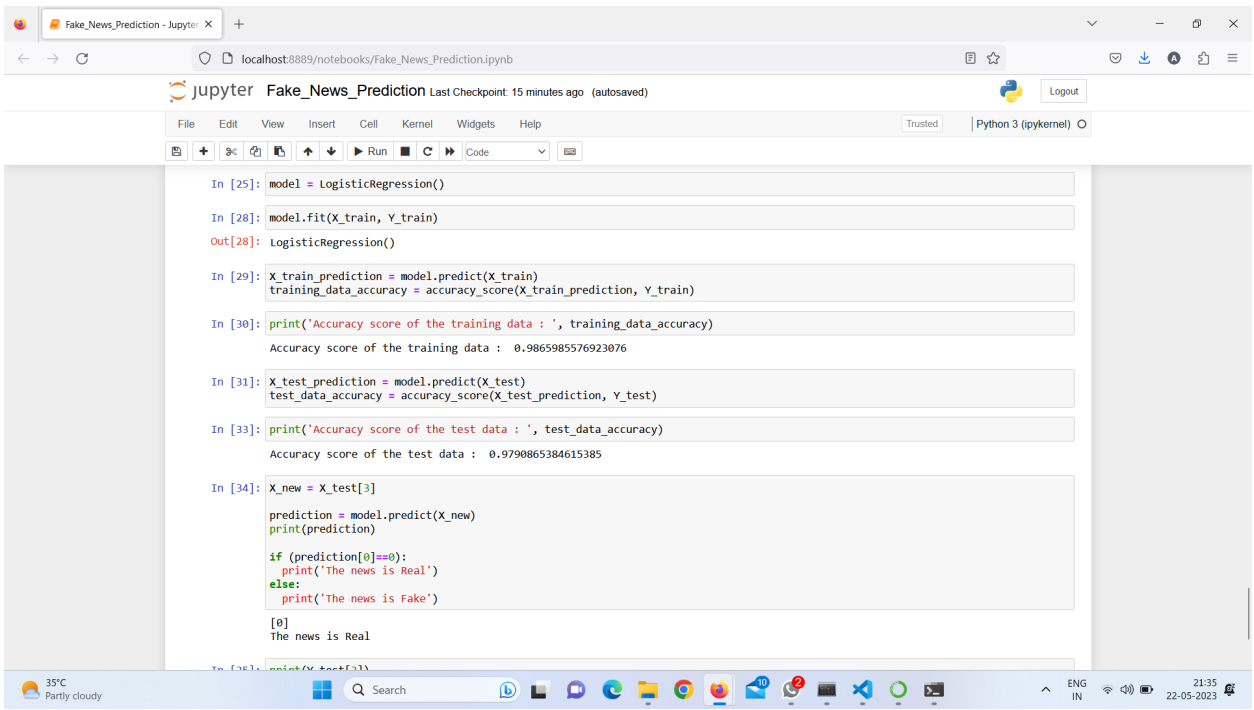

Fake_News_Prediction - Jupyter
localhost:8889/notebooks/Fake_News_Prediction.ipynb
jupyter Fake_News_Prediction Last Checkpoint: 14 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [27]: X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size = 0.2, stratify=Y, random_state=2)
In [4]: news_dataset = pd.read_csv(r"C:\Users\LENOVO\Downloads\train.csv")
In [5]: news_dataset.shape
Out[5]: (20800, 5)
In [6]: news_dataset.head()
Out[6]:
id title author text label
0 0 House Dem Aide: We Didn't Even See Comey's Let... Darrell Lucus House Dem Aide: We Didn't Even See Comey's Let... 1
1 1 FLYNN: Hillary Clinton, Big Woman on Campus - ... Daniel J. Flynn Ever get the feeling your life circles the rou... 0
2 2 Why the Truth Might Get You Fired Consortiumnews.com Why the Truth Might Get You Fired October 29, ... 1
3 3 15 Civilians Killed In Single US Airstrike Hav... Jessica Purkiss Videos 15 Civilians Killed In Single US Airstr... 1
4 4 Iranian woman jailed for fictional unpublished... Howard Portnoy Print \nAn Iranian woman has been sentenced to... 1
In [7]: news_dataset.isnull().sum()
Out[7]: id 0
title 558
author 1957
text 39
label 0
dtype: int64
In [8]: news_dataset = news_dataset.fillna('')
In [9]: news_dataset['content'] = news_dataset['author']+' '+news_dataset['title']
35°C
Partly cloudy
Search
ENG
IN
21:33
22-05-2023

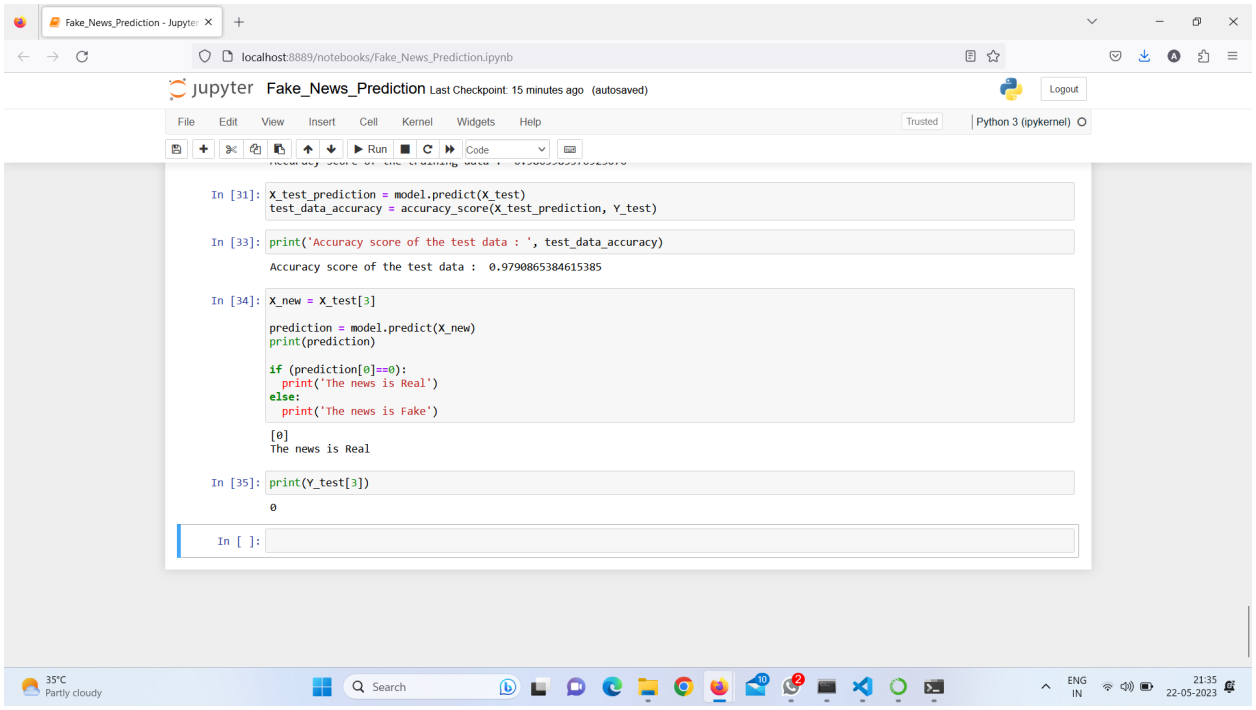

Fake_News_Prediction - Jupyter
localhost:8889/notebooks/Fake_News_Prediction.ipynb
jupyter Fake_News_Prediction Last Checkpoint: 14 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [10]: print(news_dataset['content'])
0 Darrell Lucus House Dem Aide: We Didn’t Even S...
1 Daniel J. Flynn FLYNN: Hillary Clinton, Big Wo...
2 Consortiumnews.com Why the Truth Might Get You...
3 Jessica Purkiss 15 Civilians Killed In Single ...
4 Howard Portnoy Iranian woman jailed for fictio...
...
20795 Jerome Hudson Rapper T.I.: Trump a ’Poster Chi...
20796 Benjamin Hoffman N.F.L. Playoffs: Schedule, Ma...
20797 Michael J. de la Merced and Rachel Abrams Macy...
20798 Alex Ansary NATO, Russia To Hold Parallel Exer...
20799 David Swanson What Keeps the F-35 Alive
Name: content, Length: 20800, dtype: object
In [11]: X = news_dataset.drop(columns='label', axis=1)
Y = news_dataset['label']
In [12]: X = news_dataset.drop(columns='label', axis=1)
Y = news_dataset['label']
In [14]: print(X)
print(Y)
id title \
0 0 House Dem Aide: We Didn’t Even See Comey’s Let...
1 1 FLYNN: Hillary Clinton, Big Woman on Campus - ...
2 2 Why the Truth Might Get You Fired
3 3 15 Civilians Killed In Single US Airstrike Hav...
4 4 Iranian woman jailed for fictional unpublished...
... ...
20795 20795 Rapper T.I.: Trump a ’Poster Child For White S...
20796 20796 N.F.L. Playoffs: Schedule, Matchups and Odds -...
20797 20797 Macy’s Is Said to Receive Takeover Approach by...
35°C
Partly cloudy
Search
ENG
IN
21:34
22-05-2023


Fake_News_Prediction - Jupyter
localhost:8889/notebooks/Fake_News_Prediction.ipynb
jupyter Fake_News_Prediction Last Checkpoint: 14 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [14]:
print(X)
print(Y)
id title \
0 0 House Dem Aide: We Didn’t Even See Comey’s Let...
1 1 FLYNN: Hillary Clinton, Big Woman on Campus - ...
2 2 Why the Truth Might Get You Fired
3 3 15 Civilians Killed In Single US Airstrike Hav...
4 4 Iranian woman jailed for fictional unpublished...
... ...
20795 20795 Rapper T.I.: Trump a ’Poster Child For White S...
20796 20796 N.F.L. Playoffs: Schedule, Matchups and Odds -...
20797 20797 Macy’s Is Said to Receive Takeover Approach by...
20798 20798 NATO, Russia To Hold Parallel Exercises In Bal...
20799 20799 What Keeps the F-35 Alive
author \
0 Darrell Lucus
1 Daniel J. Flynn
2 Consortiumnews.com
3 Jessica Purkiss
4 Howard Portnoy
... ...
20795 Jerome Hudson
20796 Benjamin Hoffman
20797 Michael J. de la Merced and Rachel Abrams
20798 Alex Ansary
20799 David Swanson
text \
0 House Dem Aide: We Didn’t Even See Comey’s Let...
1 Ever get the feeling your life circles the rou...
2 Why the Truth Might Get You Fired October 29, ...
3 Videos 15 Civilians Killed In Single US Airstr...
4 Print \nAn Iranian woman has been sentenced to...
...
35°C
Partly cloudy
Search
ENG
IN
21:34
22-05-2023


Fake_News_Prediction - Jupyter
localhost:8889/notebooks/Fake_News_Prediction.ipynb
jupyter Fake_News_Prediction Last Checkpoint: 15 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [15]:
port_stem = PorterStemmer()
In [16]:
def stemming(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
In [17]:
news_dataset['content'] = news_dataset['content'].apply(stemming)
In [18]:
print(news_dataset['content'])
0 darrel lucu hous dem aid even see comey letter...
1 daniel j flynn flynn hillari clinton big woman...
2 consortiumnew com truth might get fire
3 jessica purkiss civilian kill singl us airstri...
4 howard portnoy iranian woman jail fiction unpu...
...
20795 jerom hudson rapper trump poster child white s...
20796 benjamin hoffman n f l playoff schedul matchup...
20797 michael j de la merc rachel abram maci said re...
20798 alex ansari nato russia hold parallel exercis ...
20799 david swanson keep f aliv
Name: content, Length: 20800, dtype: object
In [19]:
X = news_dataset['content'].values
Y = news_dataset['label'].values
In [20]:
print(X)
['darrel lucu hous dem aid even see comey letter jason chaffetz tweet'
35°C
Partly cloudy
Search
ENG
IN
21:34
22-05-2023


Fake_News_Prediction - Jupyter
localhost:8889/notebooks/Fake_News_Prediction.ipynb
jupyter Fake_News_Prediction Last Checkpoint: 15 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [20]: print(X)
['darrel lucu hous dem aid even see comey letter jason chaffetz tweet'
'daniel j flynn flynn hillari clinton big woman campu breitbart'
'consortiumnew com truth might get fire' ...
'michael j de la merc rachel abram maci said receiv takeov approach hudson bay new york time'
'alex ansari nato russia hold parallel exercis balkan'
'david swanson keep f aliv']
In [21]: print(Y)
[1 0 1 ... 0 1 1]
In [22]: Y.shape
Out[22]: (20800,)
In [23]: vectorizer = TfidfVectorizer()
vectorizer.fit(X)
X = vectorizer.transform(X)
In [24]: print(X)
(0, 15686) 0.28485063562728646
(0, 13473) 0.2565896679337957
(0, 8909) 0.3635963806326075
(0, 8630) 0.29212514087043684
(0, 7692) 0.24785219520671603
(0, 7005) 0.21874169089359144
(0, 4973) 0.233316966909351
(0, 3792) 0.27053324808454 92
(0, 3600) 0.3598939188262559
(0, 2959) 0.2468450128533713
(0, 2483) 0.36765196867972 09
35°C
Partly cloudy
Search
ENG
IN
21:34
22-05-2023


Fake_News_Prediction - Jupyter
localhost:8889/notebooks/Fake_News_Prediction.ipynb
jupyter Fake_News_Prediction Last Checkpoint: 15 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [25]: model = LogisticRegression()
In [28]: model.fit(X_train, Y_train)
Out[28]: LogisticRegression()
In [29]: X_train_prediction = model.predict(X_train)
training_data_accuracy = accuracy_score(X_train_prediction, Y_train)
In [30]: print('Accuracy score of the training data : ', training_data_accuracy)
Accuracy score of the training data : 0.9865985576923076
In [31]: X_test_prediction = model.predict(X_test)
test_data_accuracy = accuracy_score(X_test_prediction, Y_test)
In [33]: print('Accuracy score of the test data : ', test_data_accuracy)
Accuracy score of the test data : 0.9790865384615385
In [34]: X_new = X_test[3]
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
35°C
Partly cloudy
Search
ENG
IN
21:35
22-05-2023



References:

- Kaggle
- Medium
- wikipedia