# PROJECT REPORT ON

# *Fake News Prediction*

## Submitted in fulfilment of the degree of

Department of Information Technology

***BY***


***NAME:-***

**Sirshendhu Ghosh (11900220001)**

**Dipesh Das (11900220012)**

**Atul Krishna Gupta (11900220019)**

**Susanta Dey (11900220024)**

**Tamosa Mondal (11900220033)**

**Soumili Rooj (11900220034)**

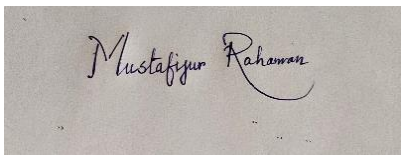
**Rohan Barick (11900220037)**

**Mustafijur Rahaman (11900220038)**

# DEPARTMENT OF IT

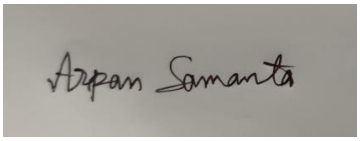
I hereby forward the documentation prepared under my supervision by SIRSHENDHU GHOSH, DIPESH DAS, ATUL KRISHNA GUPTA, SUSANTA DEY, TAMOSA MONDAL, SOUMILI ROOJ, ROHAN BARICK(TL), MUSTAFIJUR RAHAMAN entitled SILIGURI INSTITUTE OF TECHNOLOGY be accepted as fulfilment of the requirement for the Degree of B. TECH in INFORMATION TECHNOLOGY from SILIGURI INSTITUTE OF TECHNOLOGY affiliated to Maulana Abul Kalam Azad University of Technology (MAKAUT).

By




Mustafijur Rahaman

……..




Mr. ARPAN SHAMANTA

Project Guide

Ltd: Sikharthy Infotech pvt.ltd

SILIGURI INSTITUTE OF TECHNOLOGY

DEPT OF INFORMATION TECHNOLOGY(3RD YEAR)

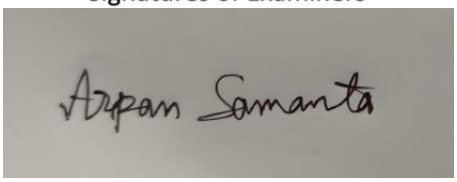
# Certificate of Approval

The foregoing project is hereby approved as a creditable study for the ***B.TECH IN I.T*** and presented in a manner of satisfactory to warrant its acceptance as a prerequisite to the degree for which it has been submitted. It is understood that by this approval the undersigned do not necessarily endorsed or approved any statement made, opinion express or conclusion therein but approve this project only for the purpose for which it is submitted.

Evaluation of the

---

Signatures of Examiners



Arpan Samanta
Sikharthy Infotech Pvt.Ltd

# ABSTRACT

The purpose of the project entitled as “Community” is to stop the fake rumours spreading application which is user friendly simple, fast, and cost effective. It is a android application for chatting. Chatting could be that one-to one or one-to-many. In one-to-many chatting system (group chat, formally known as ‘Community’) you can add or remove one or more user in your group and also leave that group. At the same time, it is completely secure. Anyone can take this advantage of this as a user by Sign up or Login. There is a “Help” section for any problems. User can also contact with us through “Contact Us”.

# TABLE OF CONTENTS

# ACKNOWLEDGEMENT

It is a great pleasure for me to acknowledge the assistance and participation of a large number of individuals to this attempt. My project report has been structured under the valued suggestion, support and guidance of Mr. Arpan Samantha. Under his guidance I have accomplished the challenging task in a very short time. Finally, I express my sincere thankfulness to my family members for inspiring me all throughout and always encouraging me.

# INTRODUCTION

**DEFINITION OF MACHINE LEARNING:-** Machine Learning (ML) is a sub-category of artificial intelligence that refers to the process by which computers develop pattern recognition, or the ability to continuously learn from and make predictions based on data, then make adjustments without being specifically programmed to do so.

Whether or not you're excited by the idea of artificial neural networks one day growing sophisticated enough to replicate human consciousness, there are undeniable practical advantages to machine learning, namely:

- **Intelligent big data management** – The sheer volume and variety of data being generated as humans and other environmental forces interact with technology would be impossible to process and draw insights from without the speed and sophistication of machine learning.
- **Smart devices** – From wearable devices that track health and fitness goals to self-driving cars to "smart cities" with infrastructure that can automatically reduce wasted time and energy, the Internet of Things (IoT) holds great promise, and machine learning can help make sense of this significant increase in data.
- **Rich consumer experiences** – Machine learning enables search engines, web apps and other technology to customise results and recommendations to match user preferences, creating delightfully personalised experiences for consumers.

# How does machine learning work?

Machine learning is incredibly complex and how it works varies depending on the task and the algorithm used to accomplish it. However, at its core, a machine learning model is a computer looking at data and identifying patterns, and then using those insights to complete its assigned task more effectively. Any task that relies

upon a set of data points or rules can be automated using machine learning, even those more complex tasks such as responding to customer service calls and reviewing CVs.

Depending on the situation, machine learning algorithms function using more or less human intervention/reinforcement. The four major machine learning models are supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning.

With **supervised learning**, the computer is provided with a labelled set of data that enables it to learn how to do a human task. This is the least complex model, as it attempts to replicate human learning.

With **unsupervised learning**, the computer is provided with unlabelled data and extracts previously unknown patterns/insights from it. There are many different ways that machine learning algorithms do this, including:

- Clustering, in which the computer finds similar data points within a data set and groups them accordingly (creating “clusters”).
- Density estimation, in which the computer discovers insights by looking at how a data set is distributed.
- Anomaly detection, in which the computer identifies data points within a data set that are significantly different from the rest of the data.
- Principal component analysis (PCA), in which the computer analyses a data set and summarises it so that it can be used to make accurate predictions.

With **semi-supervised learning**, the computer is provided with a set of partially labelled data and performs its task using the labelled data to understand the parameters for interpreting the unlabelled data.

With **reinforcement learning**, the computer observes its environment and uses that data to identify the ideal behaviour that will minimise risk and/or maximise reward**.**

# STEPS OF MACHINE LEARNING

## 1. Collecting Data:

As you know, machines initially learn from the data that you give them. It is of the utmost importance to collect reliable data so that your machine learning model can find the correct patterns. The quality of the data that you feed to the machine will determine how accurate your model is. If you have incorrect or outdated data, you will have wrong outcomes or predictions which are not relevant.

Make sure you use data from a reliable source, as it will directly affect the outcome of your model. Good data is relevant, contains very few missing and repeated values, and has a good representation of the various subcategories/classes present.

## 2. Preparing the Data:

After you have your data, you have to prepare it. You can do this by :

- Putting together all the data you have and randomizing it. This helps make sure that data is evenly distributed, and the ordering does not affect the learning process.
- Cleaning the data to remove unwanted data, missing values, rows, and columns, duplicate values, data type conversion, etc. You might even have to restructure the dataset and change the rows and columns or index of rows and columns.
- Visualize the data to understand how it is structured and understand the relationship between various variables and classes present.
- Splitting the cleaned data into two sets - a training set and a testing set. The training set is the set your model learns from. A testing set is used to check the accuracy of your model after training.

## 3. Choosing a Model:

A machine learning model determines the output you get after running a machine learning algorithm on the collected data. It is important to choose a model which is relevant to the task at hand. Over the years, scientists and engineers developed various models suited for different tasks like speech recognition, image recognition,

prediction, etc. Apart from this, you also have to see if your model is suited for numerical or categorical data and choose accordingly.

## 4. Training the Model:

Training is the most important step in machine learning. In training, you pass the prepared data to your machine learning model to find patterns and make predictions. It results in the model learning from the data so that it can accomplish the task set. Over time, with training, the model gets better at predicting.

## 5. Evaluating the Model:

After training your model, you have to check to see how it's performing. This is done by testing the performance of the model on previously unseen data. The unseen data used is the testing set that you split our data into earlier. If testing was done on the same data which is used for training, you will not get an accurate measure, as the model is already used to the data, and finds the same patterns in it, as it previously did. This will give you disproportionately high accuracy.

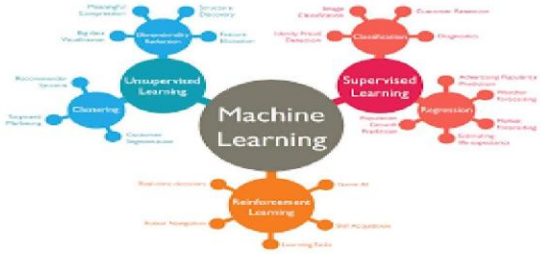
When used on testing data, you get an accurate measure of how your model will perform and its speed.

## 6. Parameter Tuning:

Once you have created and evaluated your model, see if its accuracy can be improved in any way. This is done by tuning the parameters present in your model. Parameters are the variables in the model that the programmer generally decides. At a particular value of your parameter, the accuracy will be the maximum. Parameter tuning refers to finding these values.

## 7. Making Predictions

In the end, you can use your model on unseen data to make predictions accurately.



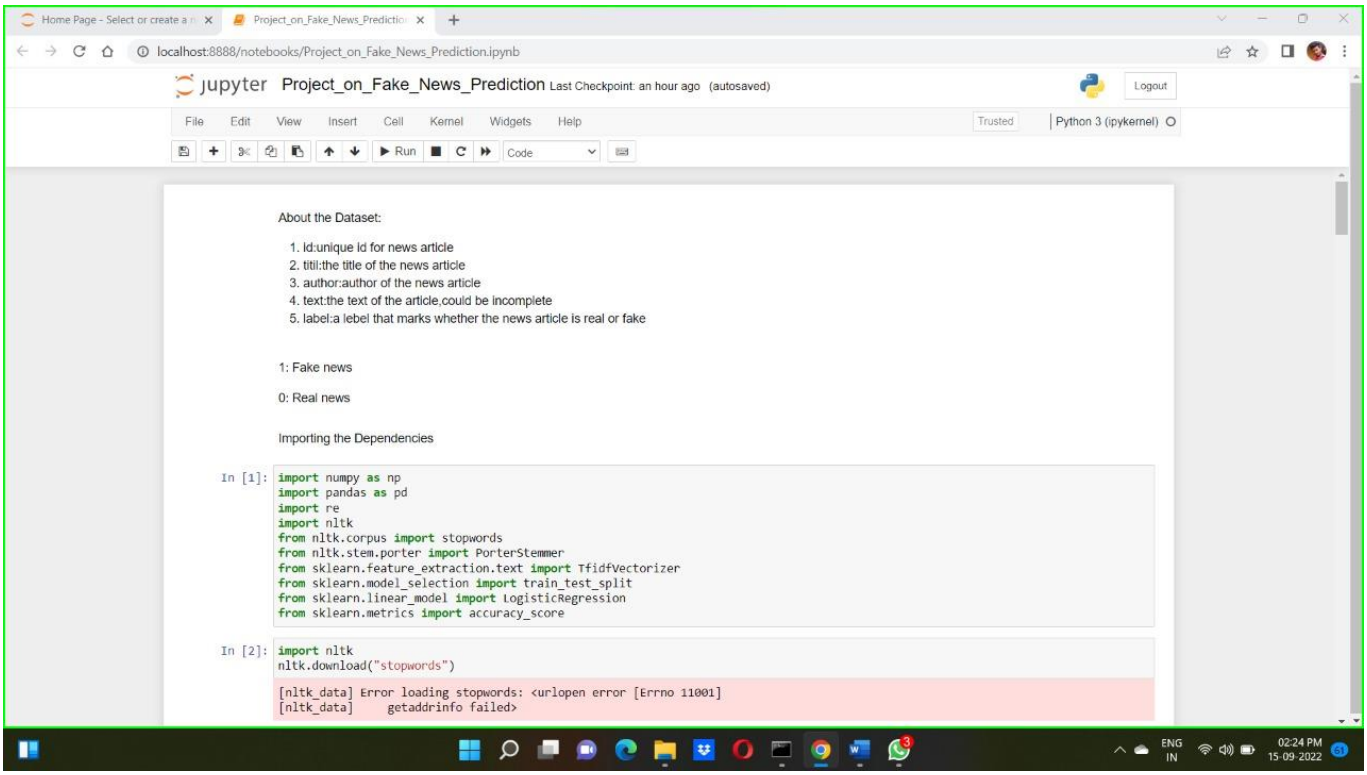

## How to Implement Machine Learning Steps in Python?

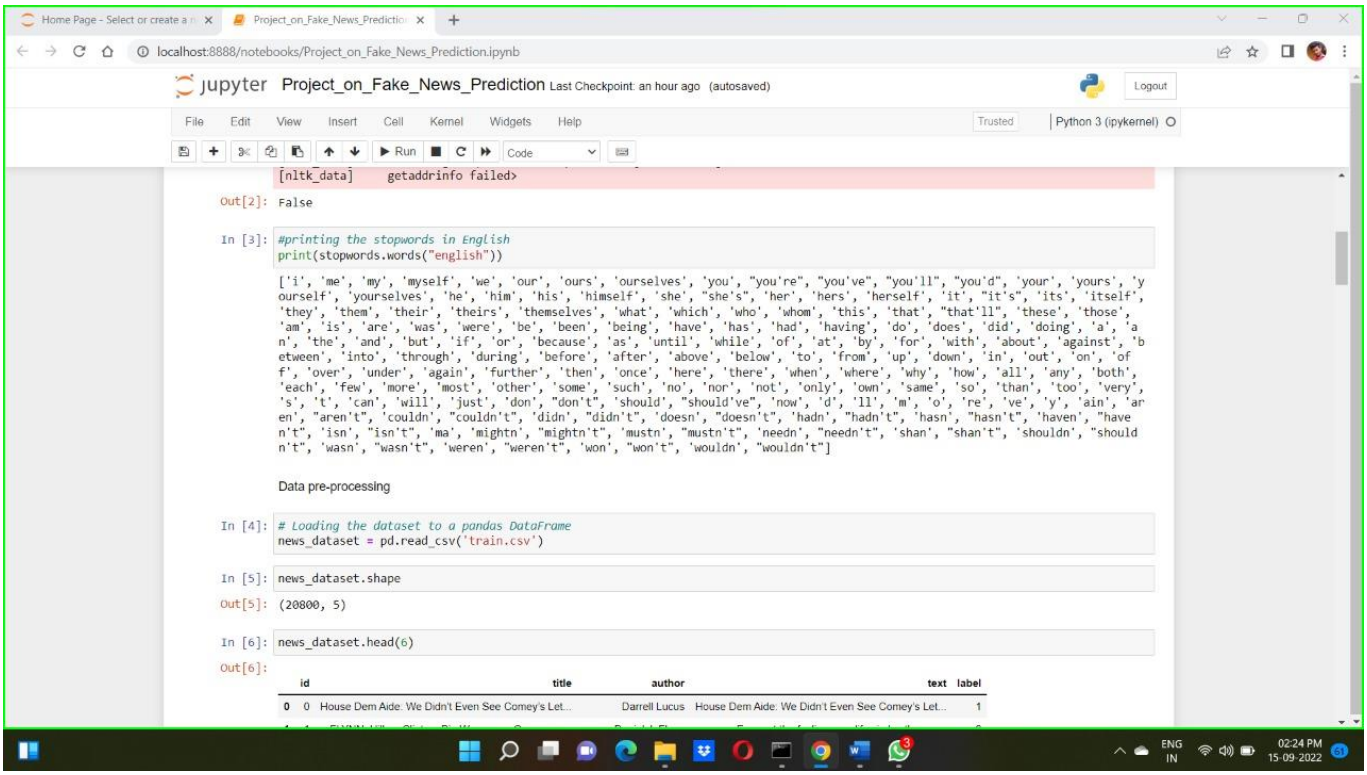
You will now see how to implement a machine learning model using Python.

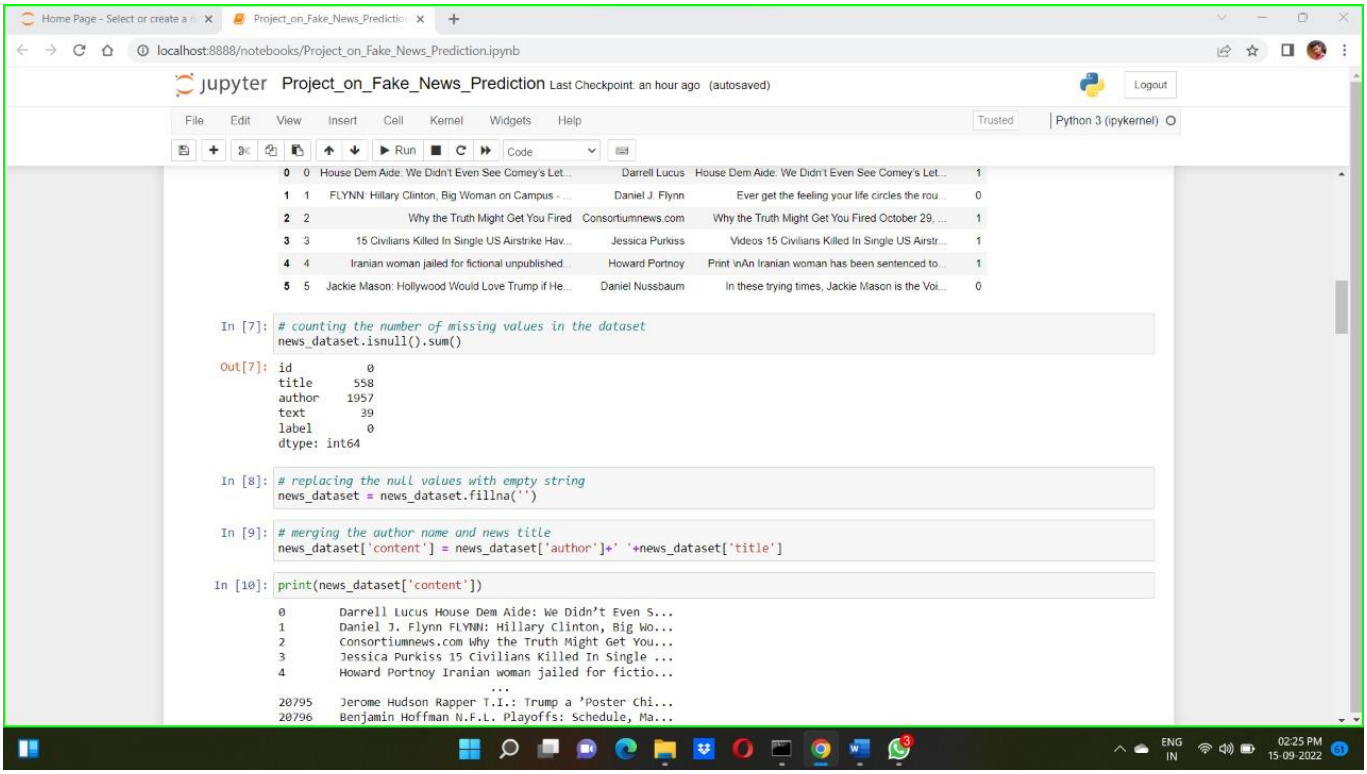
In this example, data collected is from an insurance company, which tells you the variables that come into play when an insurance amount is set. Using this, you will have to predict the insurance amount for a person. This data was collected from Kaggle.com, which has many reliable datasets.

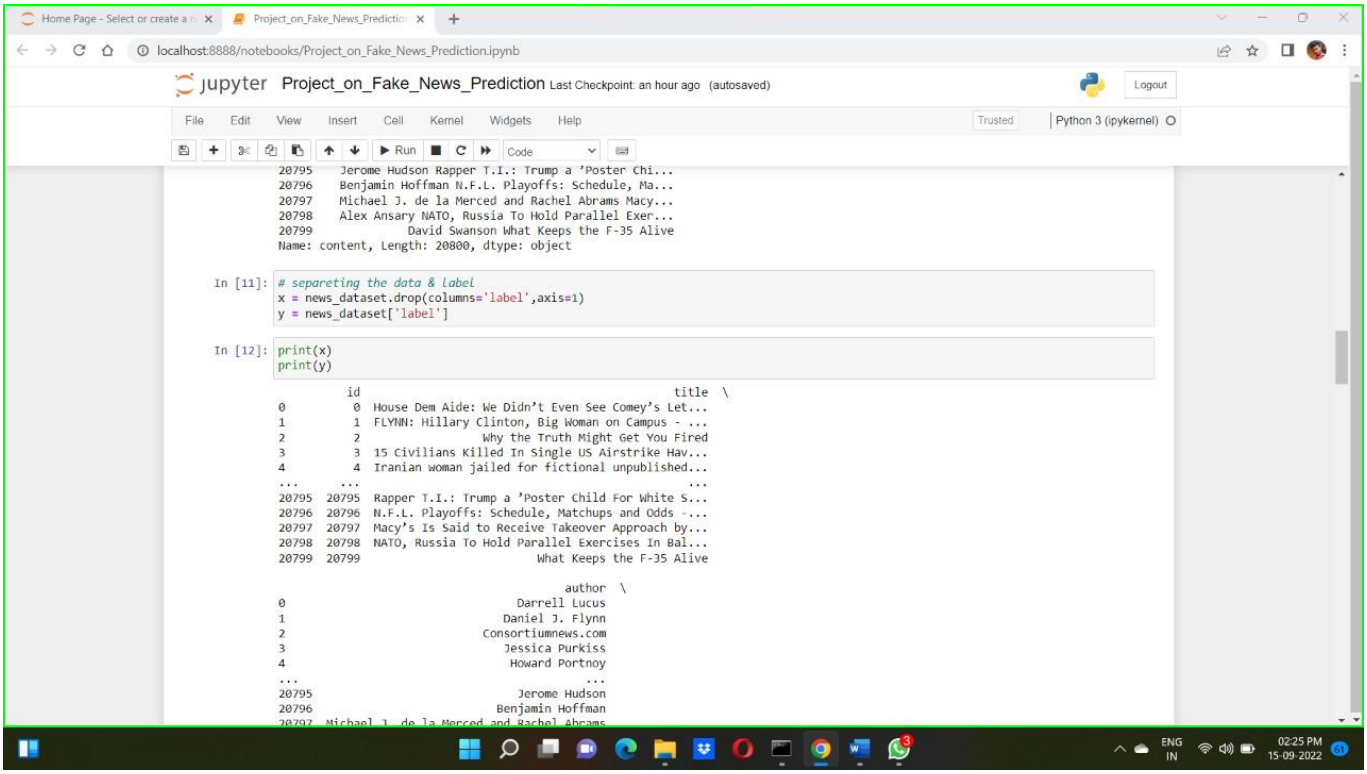
You need to start by importing any necessary modules, as shown.

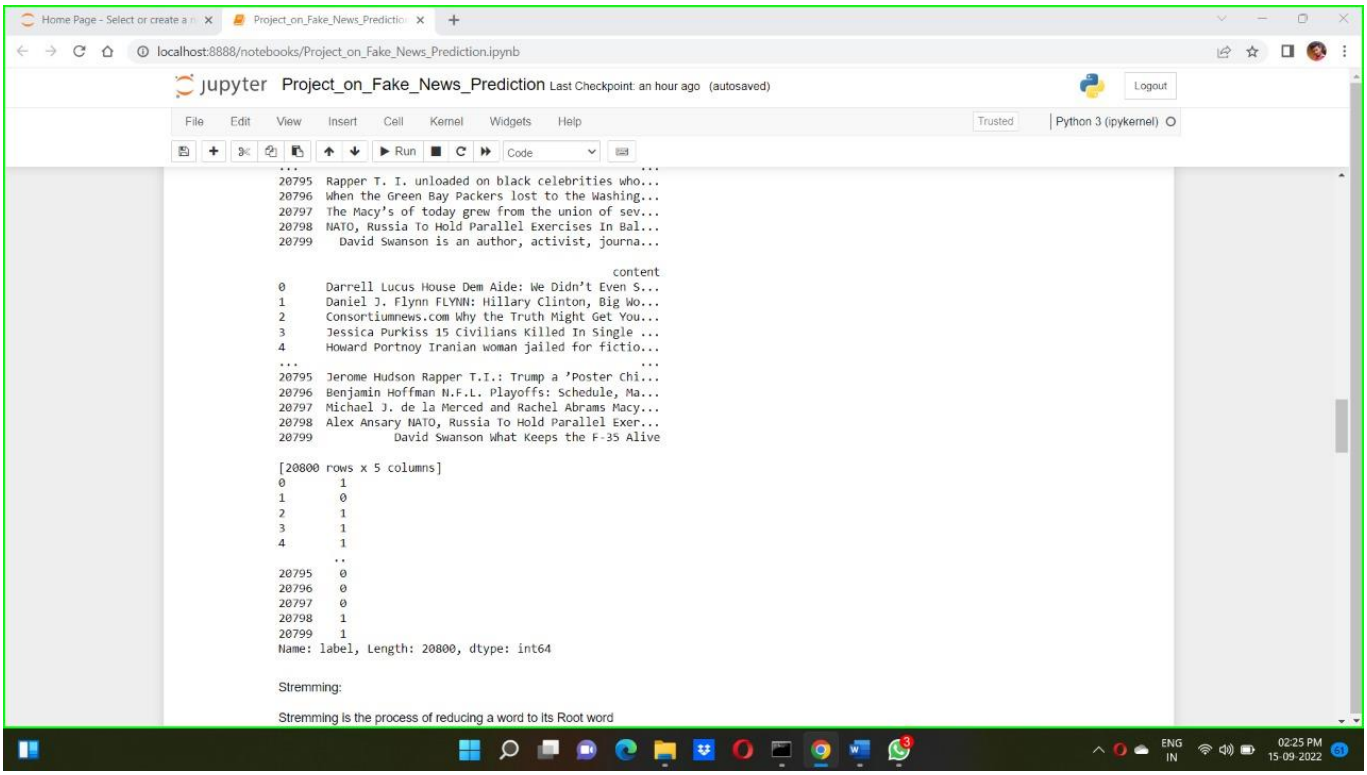
```
import pandas as pd
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import accuracy_score
```

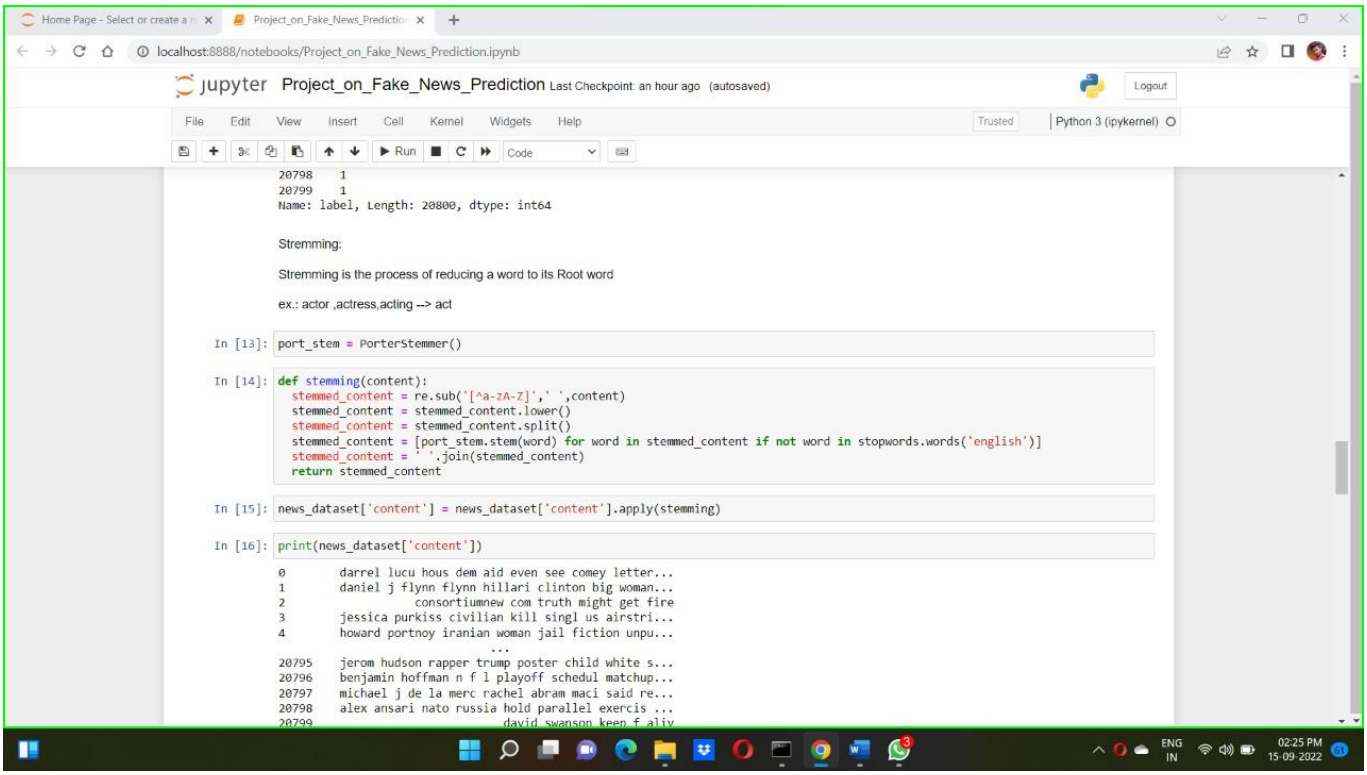

Home Page - Select or create a n
Project_on_Fake_News_Prediction
localhost:8888/notebooks/Project_on_Fake_News_Prediction.ipynb
jupyter Project_on_Fake_News_Prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
About the Dataset:
1. id:unique id for news article
2. titil:the title of the news article
3. author:author of the news article
4. text:the text of the article,could be incomplete
5. label:a lebel that marks whether the news article is real or fake
1: Fake news
0: Real news
Importing the Dependencies
In [1]:
import numpy as np
import pandas as pd
import re
import nltk
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] Error loading stopwords: <urlopen error [Errno 11001]
[nltk_data] getaddrinfo failed>
ENG
IN
02:24 PM
15-09-2022

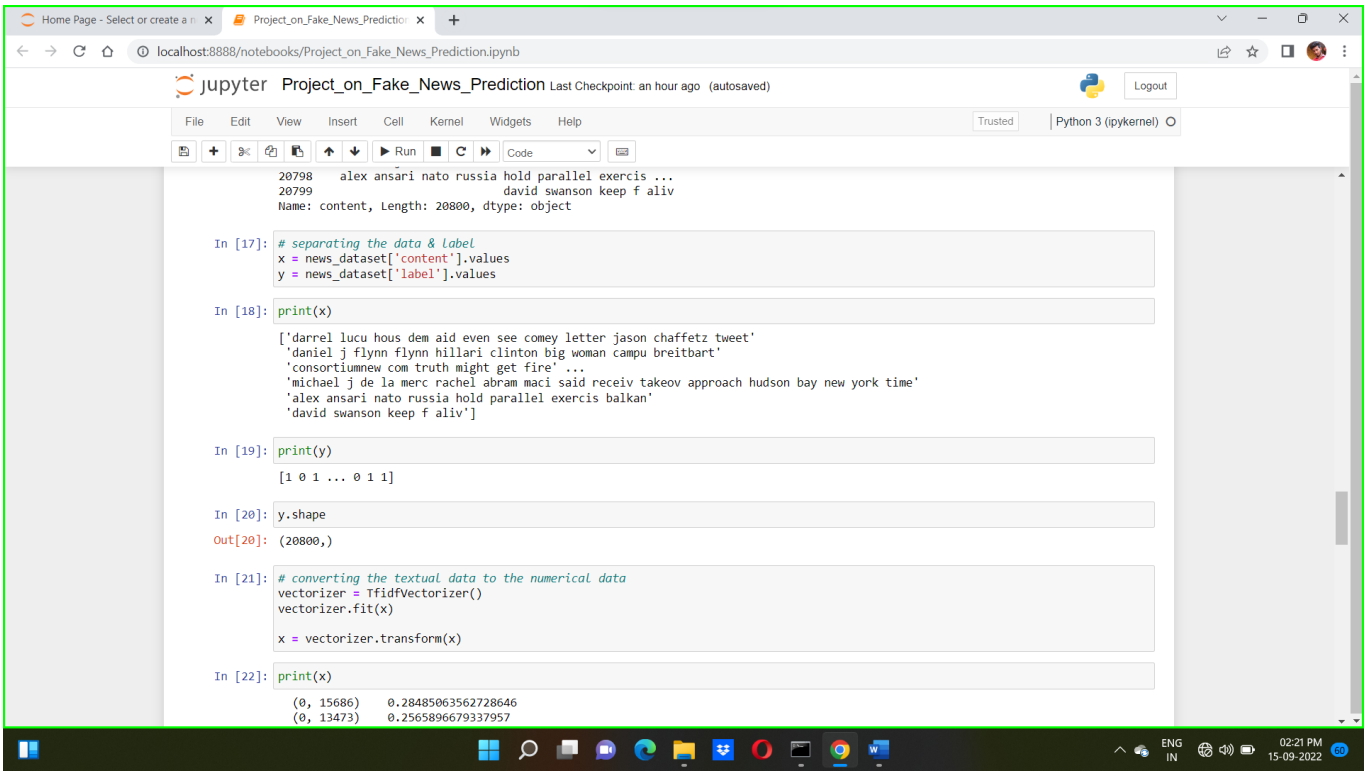

Home Page - Select or create a n
Project_on_Fake_News_Prediction
localhost:8888/notebooks/Project_on_Fake_News_Prediction.ipynb
jupyter Project_on_Fake_News_Prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
[nltk_data] getaddrinfo failed>
Out[2]: False
In [3]: #printing the 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"]
Data pre-processing
In [4]: # Loading the dataset to a pandas DataFrame
news_dataset = pd.read_csv('train.csv')
In [5]: news_dataset.shape
Out[5]: (20800, 5)
In [6]: news_dataset.head(6)
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
ENG IN
02:24 PM 15-09-2022

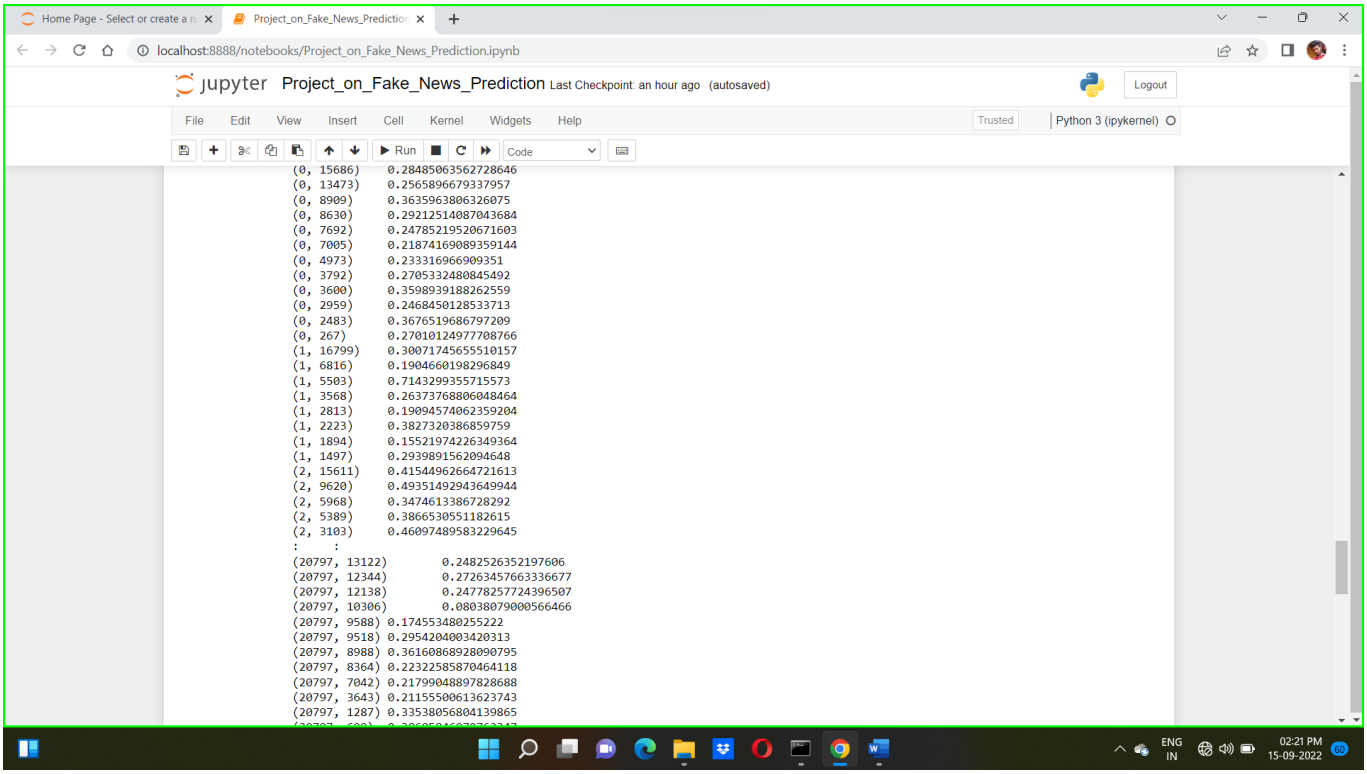

Project_on_Fake_News_Prediction
localhost:8888/notebooks/Project_on_Fake_News_Prediction.ipynb
jupyter Project_on_Fake_News_Prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
0 0 House Dem Aide: We Didn't Even 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
5 5 Jackie Mason: Hollywood Would Love Trump if He... Daniel Nussbaum In these trying times, Jackie Mason is the Voi... 0
In [7]:
# counting the number of missing values in the dataset
news_dataset.isnull().sum()
Out[7]:
id 0
title 558
author 1957
text 39
label 0
dtype: int64
In [8]:
# replacing the null values with empty string
news_dataset = news_dataset.fillna('')
In [9]:
# merging the author name and news title
news_dataset['content'] = news_dataset['author']+' '+news_dataset['title']
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...
ENG IN
02:25 PM 15-09-2022

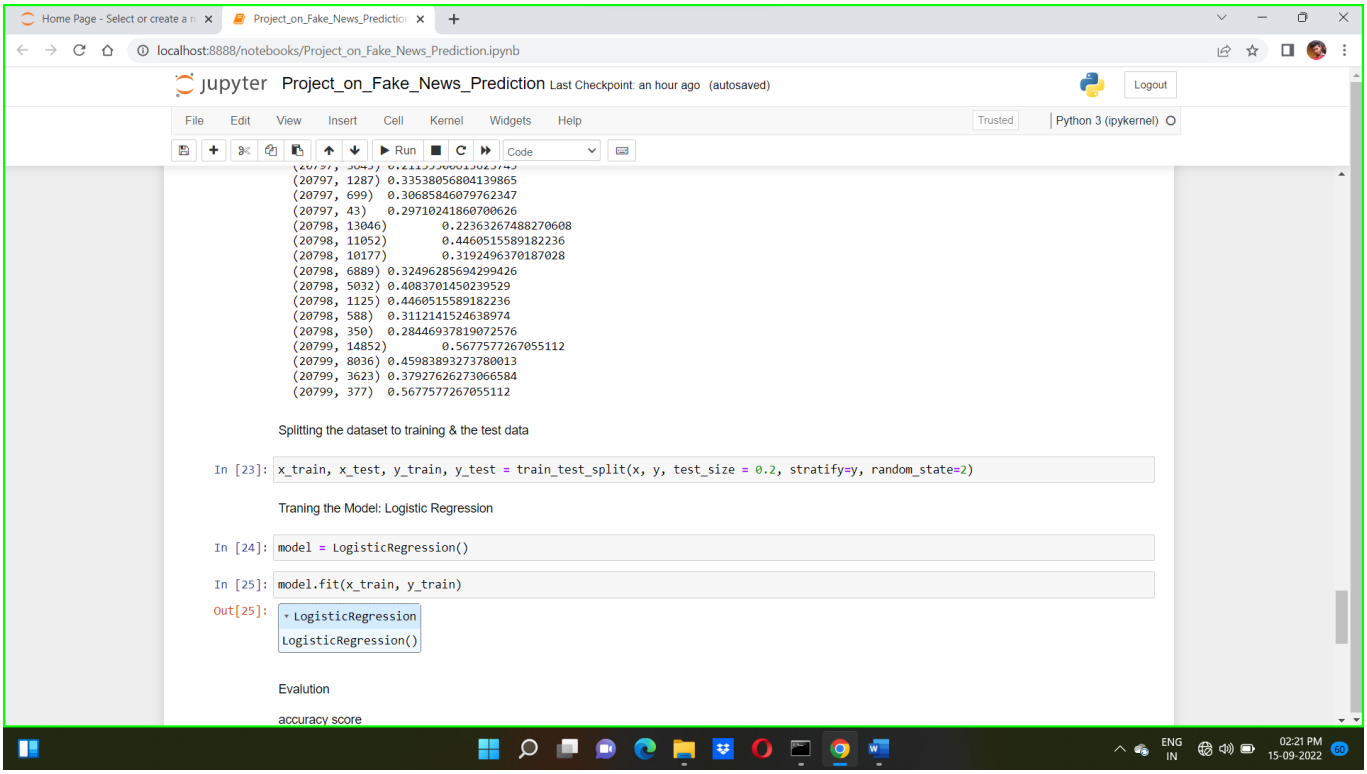

Home Page - Select or create a n
Project_on_Fake_News_Prediction
localhost:8888/notebooks/Project_on_Fake_News_Prediction.ipynb
jupyter Project_on_Fake_News_Prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
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]:
# separeting the data & label
x = news_dataset.drop(columns='label',axis=1)
y = news_dataset['label']
In [12]:
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
ENG
IN
02:25 PM
15-09-2022

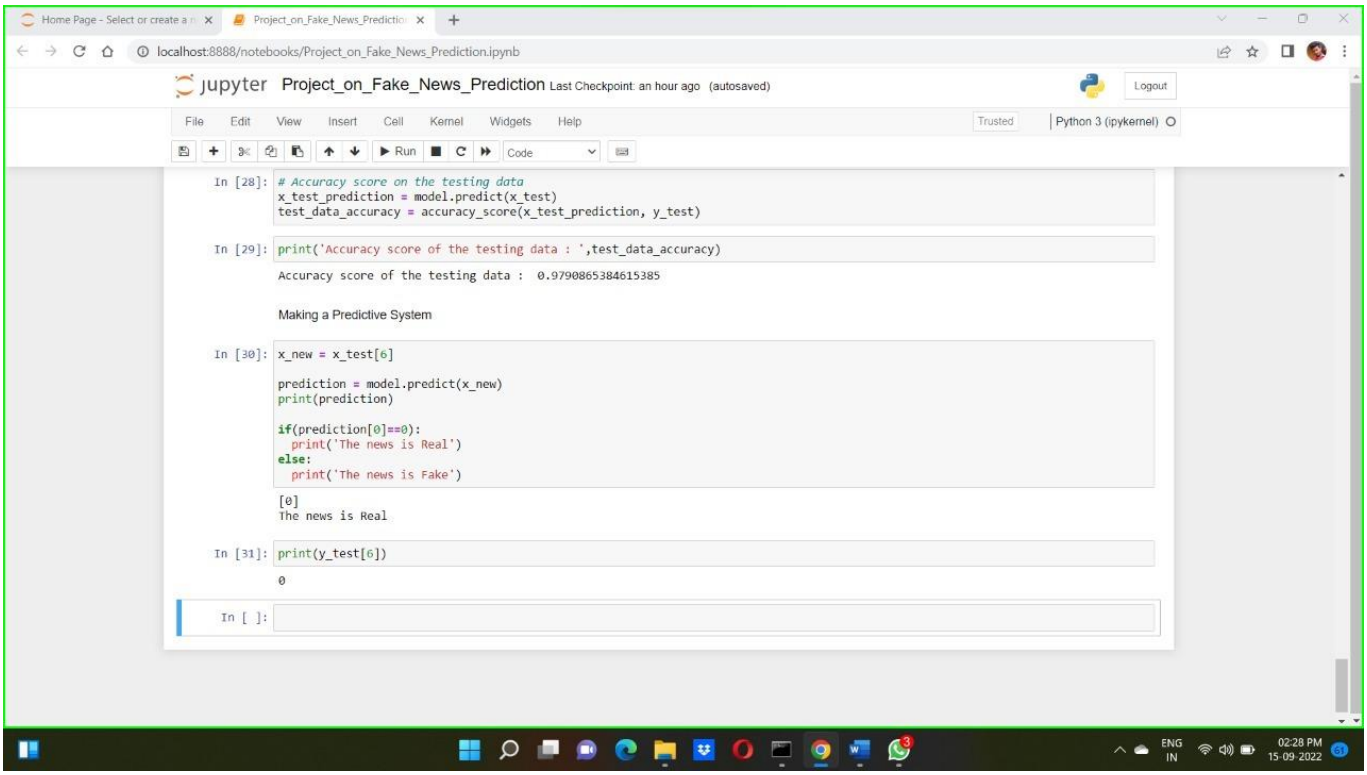

Home Page - Select or create a n
Project_on_Fake_News_Predictio
localhost:8888/notebooks/Project_on_Fake_News_Prediction.ipynb
jupyter Project_on_Fake_News_Prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
20795 Rapper T. I. unloaded on black celebrities who...
20796 When the Green Bay Packers lost to the Washing...
20797 The Macy's of today grew from the union of sev...
20798 NATO, Russia To Hold Parallel Exercises In Bal...
20799 David Swanson is an author, activist, journa...
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
[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
Stremming:
Stremming is the process of reducing a word to its Root word
ENG
IN
02:25 PM
15-09-2022

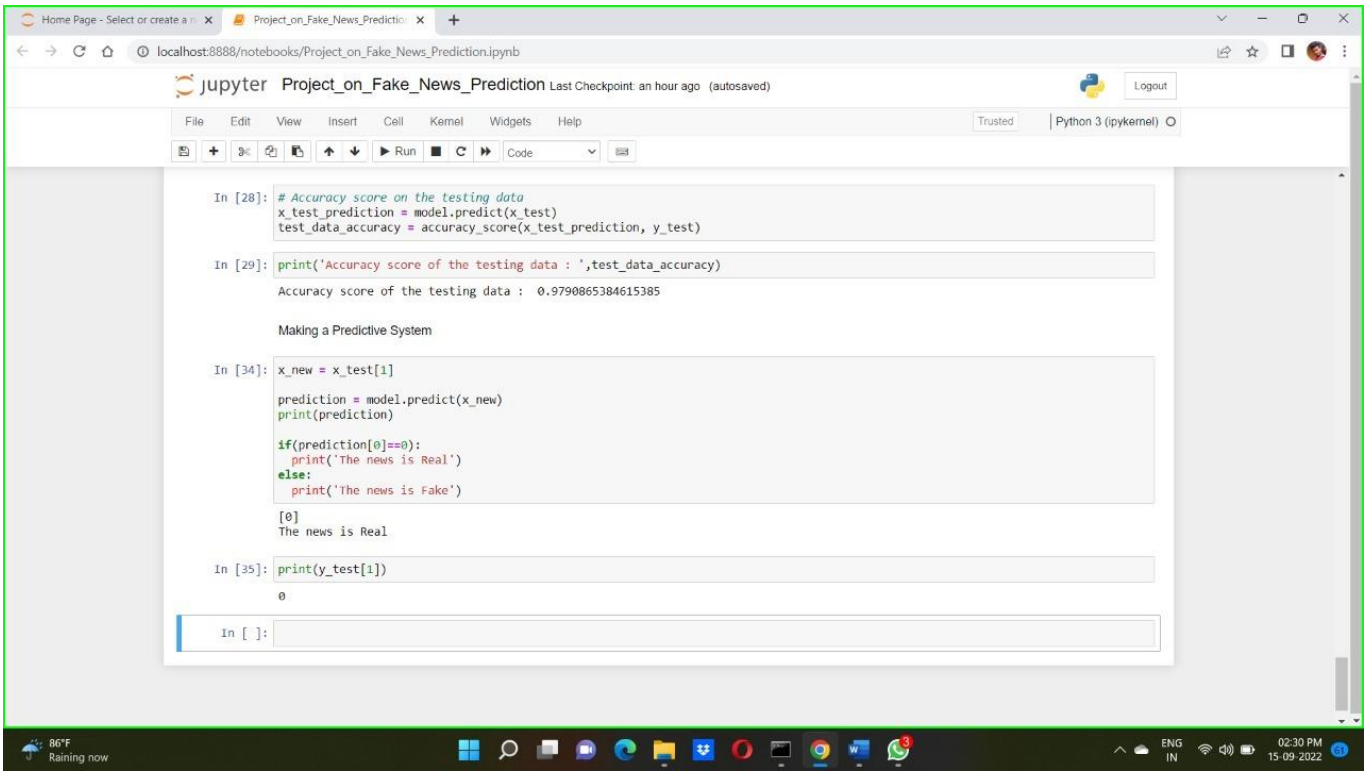

Project_on_Fake_News_Prediction
localhost:8888/notebooks/Project_on_Fake_News_Prediction.ipynb
jupyter Project_on_Fake_News_Prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
20798 1
20799 1
Name: label, Length: 20800, dtype: int64
Stremming:
Stremming is the process of reducing a word to its Root word
ex.: actor ,actress,acting --> act
In [13]: port_stem = PorterStemmer()
In [14]: 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 [15]: news_dataset['content'] = news_dataset['content'].apply(stemming)
In [16]: 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
ENG IN
02:25 PM 15-09-2022

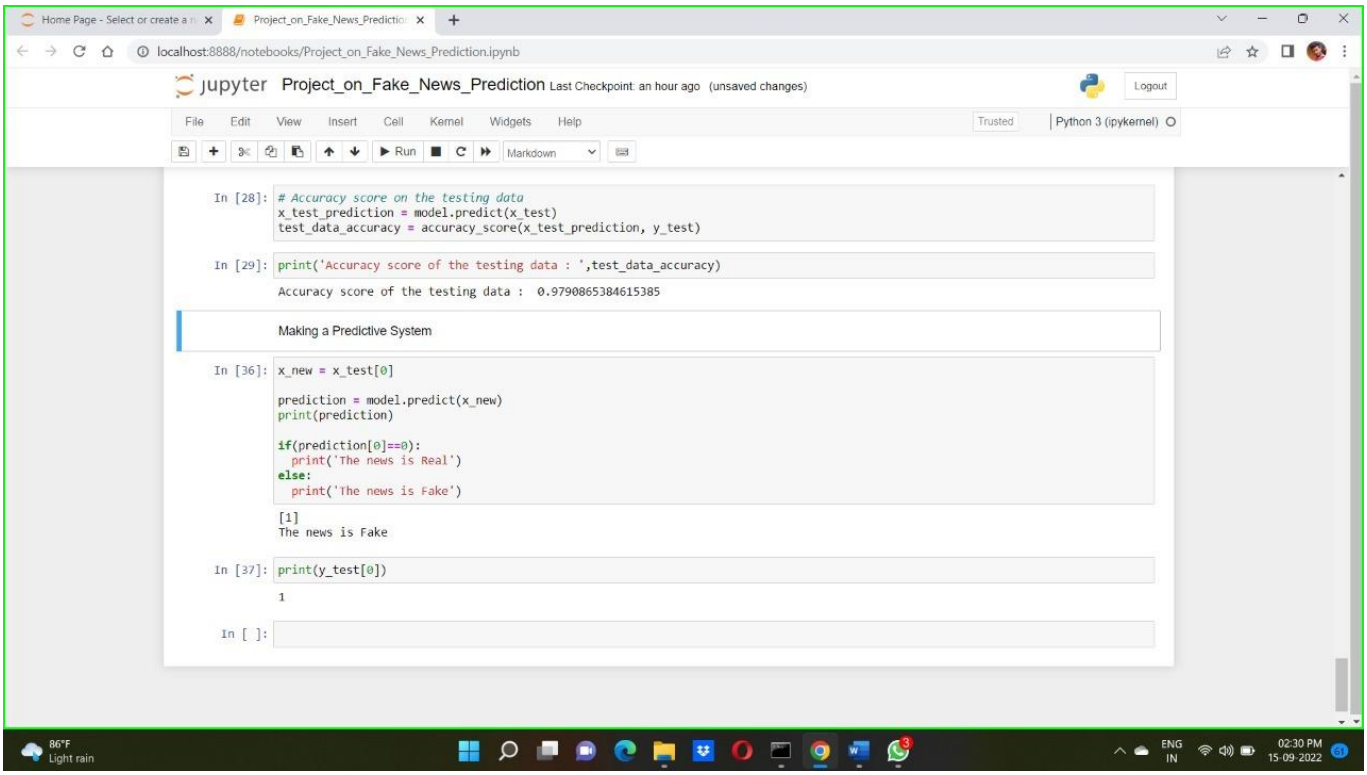

Home Page - Select or create a n
Project_on_Fake_News_Prediction
localhost:8888/notebooks/Project_on_Fake_News_Prediction.ipynb
jupyter Project_on_Fake_News_Prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
20798 alex ansari nato russia hold parallel exercis ...
20799 david swanson keep f aliv
Name: content, Length: 20800, dtype: object
In [17]:
# separating the data & label
x = news_dataset['content'].values
y = news_dataset['label'].values
In [18]:
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 [19]:
print(y)
[1 0 1 ... 0 1 1]
In [20]:
y.shape
Out[20]:
(20800,)
In [21]:
# converting the textual data to the numerical data
vectorizer = TfidfVectorizer()
vectorizer.fit(x)
x = vectorizer.transform(x)
In [22]:
print(x)
(0, 15686) 0.28485063562728646
(0, 13473) 0.2565896679337957
ENG
IN
02:21 PM
15-09-2022


Home Page - Select or create a n
Project_on_Fake_News_Prediction
localhost:8888/notebooks/Project_on_Fake_News_Prediction.ipynb
jupyter Project_on_Fake_News_Prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
(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) 0.2705332480845492
(0, 3600) 0.3598939188262559
(0, 2959) 0.2468450128533713
(0, 2483) 0.3676519686797209
(0, 267) 0.27010124977708766
(1, 16799) 0.30071745655510157
(1, 6816) 0.1904660198296849
(1, 5503) 0.7143299355715573
(1, 3568) 0.26373768806048464
(1, 2813) 0.19094574062359204
(1, 2223) 0.3827320386859759
(1, 1894) 0.15521974226349364
(1, 1497) 0.2939891562094648
(2, 15611) 0.41544962664721613
(2, 9620) 0.49351492943649944
(2, 5968) 0.3474613386728292
(2, 5389) 0.3866530551182615
(2, 3103) 0.46097489583229645
(20797, 13122) 0.2482526352197606
(20797, 12344) 0.27263457663336677
(20797, 12138) 0.24778257724396507
(20797, 10306) 0.08038079000566466
(20797, 9588) 0.174553480255222
(20797, 9518) 0.2954204003420313
(20797, 8988) 0.36160868928090795
(20797, 8364) 0.22322585870464118
(20797, 7042) 0.21799048897828688
(20797, 3643) 0.21155500613623743
(20797, 1287) 0.33538056804139865
ENG IN
02:21 PM 15-09-2022


Home Page - Select or create a n
Project_on_Fake_News_Prediction
localhost:8888/notebooks/Project_on_Fake_News_Prediction.ipynb
jupyter Project_on_Fake_News_Prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
(20797, 1287) 0.33538056804139865
(20797, 699) 0.30685846079762347
(20797, 43) 0.29710241860700626
(20798, 13046) 0.22363267488270608
(20798, 11052) 0.4460515589182236
(20798, 10177) 0.3192496370187028
(20798, 6889) 0.32496285694299426
(20798, 5032) 0.4083701450239529
(20798, 1125) 0.4460515589182236
(20798, 588) 0.3112141524638974
(20798, 350) 0.28446937819072576
(20799, 14852) 0.5677577267055112
(20799, 8036) 0.45983893273780013
(20799, 3623) 0.37927626273066584
(20799, 377) 0.5677577267055112
Splitting the dataset to training & the test data
In [23]: x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, stratify=y, random_state=2)
Traning the Model: Logistic Regression
In [24]: model = LogisticRegression()
In [25]: model.fit(x_train, y_train)
Out[25]: LogisticRegression
LogisticRegression()
Evalution
accuracy score
ENG
IN
02:21 PM
15-09-2022


Home Page - Select or create a n
Project_on_Fake_News_Prediction
localhost:8888/notebooks/Project_on_Fake_News_Prediction.ipynb
jupyter Project_on_Fake_News_Prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [28]:
# Accuracy score on the testing data
x_test_prediction = model.predict(x_test)
test_data_accuracy = accuracy_score(x_test_prediction, y_test)
In [29]:
print('Accuracy score of the testing data : ',test_data_accuracy)
Accuracy score of the testing data : 0.9790865384615385
Making a Predictive System
In [30]:
x_new = x_test[6]
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
In [31]:
print(y_test[6])
0
In [ ]:
ENG
IN
02:28 PM
15-09-2022


Home Page - Select or create a n
Project_on_Fake_News_Prediction
localhost:8888/notebooks/Project_on_Fake_News_Prediction.ipynb
jupyter Project_on_Fake_News_Prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [28]:
# Accuracy score on the testing data
x_test_prediction = model.predict(x_test)
test_data_accuracy = accuracy_score(x_test_prediction, y_test)
In [29]:
print('Accuracy score of the testing data : ',test_data_accuracy)
Accuracy score of the testing data : 0.9790865384615385
Making a Predictive System
In [34]:
x_new = x_test[1]
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
In [35]:
print(y_test[1])
0
In [ ]:
86°F
Raining now
ENG
IN
02:30 PM
15-09-2022


Home Page - Select or create a n
Project_on_Fake_News_Predictio
localhost:8888/notebooks/Project_on_Fake_News_Prediction.ipynb
jupyter Project_on_Fake_News_Prediction Last Checkpoint: an hour ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Markdown
In [28]:
# Accuracy score on the testing data
x_test_prediction = model.predict(x_test)
test_data_accuracy = accuracy_score(x_test_prediction, y_test)
In [29]:
print('Accuracy score of the testing data : ',test_data_accuracy)
Accuracy score of the testing data : 0.9790865384615385
Making a Predictive System
In [36]:
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')
[1]
The news is Fake
In [37]:
print(y_test[0])
1
In [ ]:
86°F
Light rain
ENG
IN
02:30 PM
15-09-2022

## CONCLUSION

Today, we learned to detect fake news with Python. We took a Fake and True News dataset, implemented a Text cleaning function, TfidfVectorizer, initialized Multinomial Naive Bayes Classifier, and fit our model. We ended up obtaining an accuracy of 97.90% in magnitude.

# REFERENCE


1. https//youtube.com
2. https://medium.com/swlh/fake-news-detection-using-machine-learning-69ff9050351f#:~:text=Today%2C%20we%20learned%20to%20detect,accuracy%20of%2095.31%25%20in%20magnitude