ARTIFICIAL INTELLIGENCE PROJECT

FAKE NEWS DETECTION ON SOCIAL MEDIA

SUNNY WAHANE

ROLL NO. 18

REG NO. 11812594

SECTION : K18KK

SUBMITTED TO : DR. SAGAR PANDE

SCHOOL OF COMPUTER SCIENCE

LOVELY PROFESSIONAL UNIVERSITY

ABSTRACT

This Project is about fake news detection on social media. Implemented with the help of scikit library in python.Nowadays Fake news is spreading at a rapid pace .Detecting Fake news is a great challenge of Artificial Intelligence.Machine learning is used to detect whether a news is fake or not.It's a binary classification of Machine learning where the output is either 0 or 1 representing REAL or Fake.

This Project is a very basic project which decides whether the given news is fake or not based on the occurrence of some specific words in the news which are present in any other irrelevant words.

In this project I could have you logistic regression but Passive Aggressive algorithms are online learning algorithms. Such an algorithm remains passive for a correct classification outcome, and turns aggressive in the event of a miscalculation, updating and adjusting. Unlike most other algorithms, it does not converge. Its purpose is to make updates that correct the loss, causing very little change in the norm of the weight vector.And hence for data which is changing Passive Aggressive algorithms works better than logistic regression .

RELATED WORK

The major problem was getting a dataset. Since for building a classifier or any other machine learning project well formatted data is required.This problem was solved by taking a dataset from kaggle (link at the end of the report).

It was hard to find the right dataset according to the need .

Lot of help was taken from “stack overflow” regarding errors and syntax
while I was trying to implement the algorithm in python.

IMPLEMENTATION

1. GUI :

- For Gui I use the tkinter library of python .
- This was used for the purpose of getting the news from the user and displaying the result whether it is fake or not .
- It consist of main window which ask for the text from the news
- After that the model will take the text and outputs in a new window whether its fake or real

2.Feature Extraction :

- To convert raw text into features I have used TfidfVectorizer from scikit learn
- TfidfVectorizer Converts a collection of raw documents to a matrix of TF-IDF features.

1. TF(Term Frequency) :
   The number of times a word appears in a document is its Term Frequency. A higher value means a term appears more often than others, and so, the document is a good match when the term is part of the search terms.
2. IDF (Inverse Document Frequency):
   Words that occur many times in a document, but also occur many times in many others, may be irrelevant. IDF is a measure of how significant a term is in the entire corpus.

3. Training the Model

For this I have use Passive Aggressive Classifier
Passive Aggressive algorithms are online learning algorithms. Such an algorithm remains passive for a correct classification outcome, and turns aggressive in the event of a miscalculation, updating and adjusting. Unlike most other algorithms, it does not converge. Its purpose is to make updates that correct the loss, causing very little change in the norm of the weight vector.

About the Dataset :

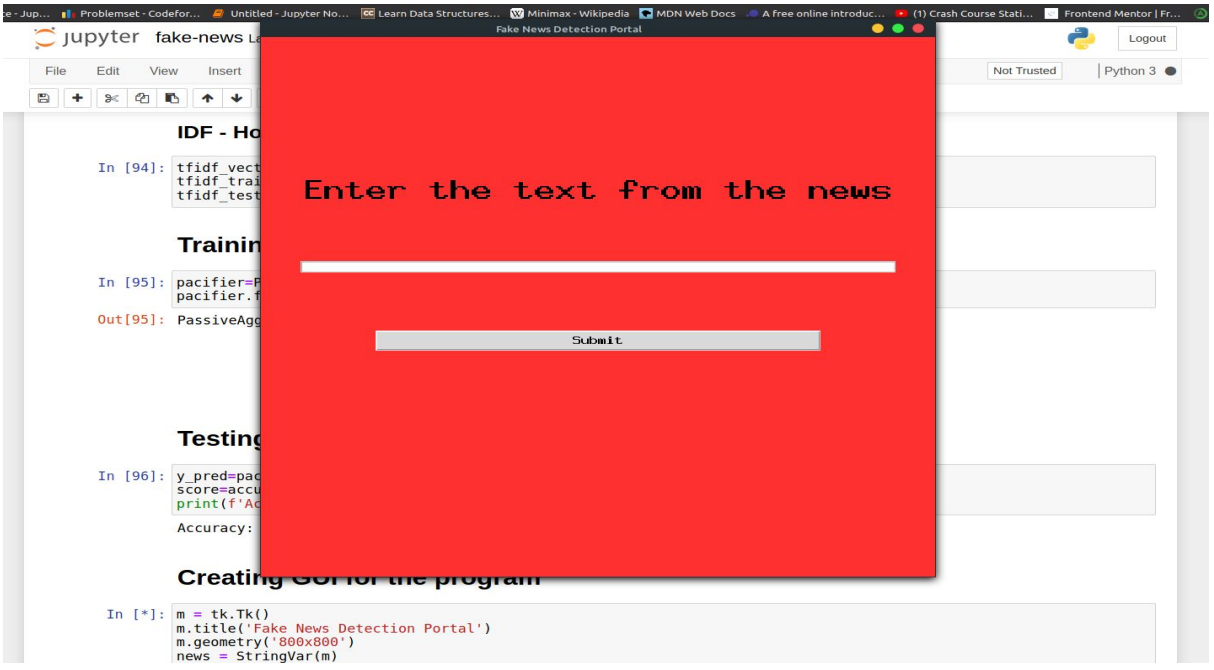
train.csv: A full training dataset with the following attributes:

- **id**: unique id for a news article
- **title**: the title of a news article
- **author**: author of the news article
- **text**: the text of the article; could be incomplete
- **label**: a label that marks the article as potentially unreliable
    - 1: unreliable
    - 0: reliable

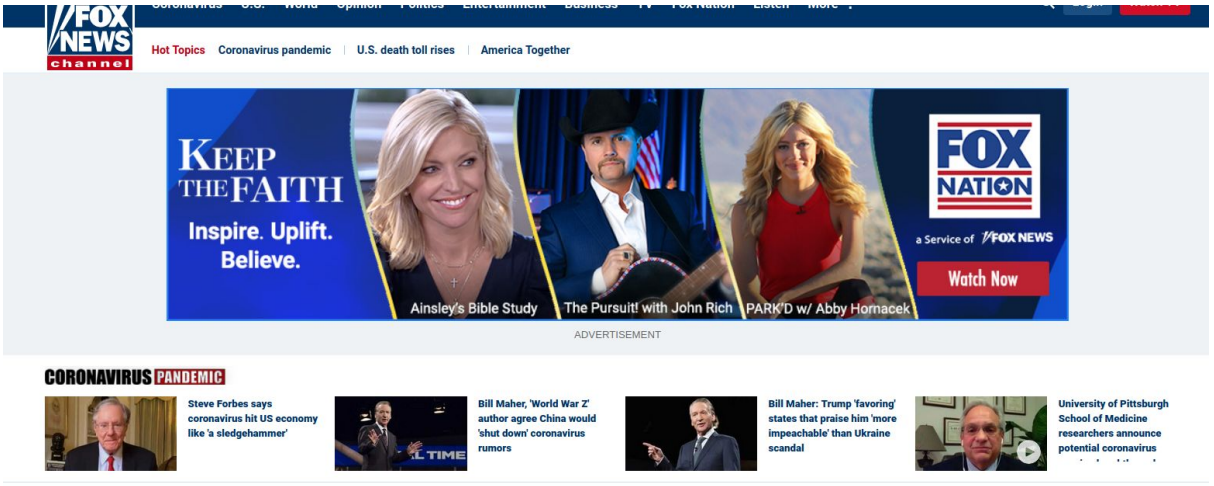
Screenshot/Result :

Test Result for Fox News (Link:https://www.foxnews.com/politics/biden-campaign-says-ex-vp-backs-trump-move-limit-travel-china):

Input Prompt:



Tested News (from Fox News) :



CORONAVIRUS · **Published** 5 hours ago · **Last Update** 4 hours ago

Biden campaign says former VP backs Trump move to limit travel from China

The outbreak first emerged in China's city of Wuhan and the virus rapidly spread – eventually becoming a global pandemic.

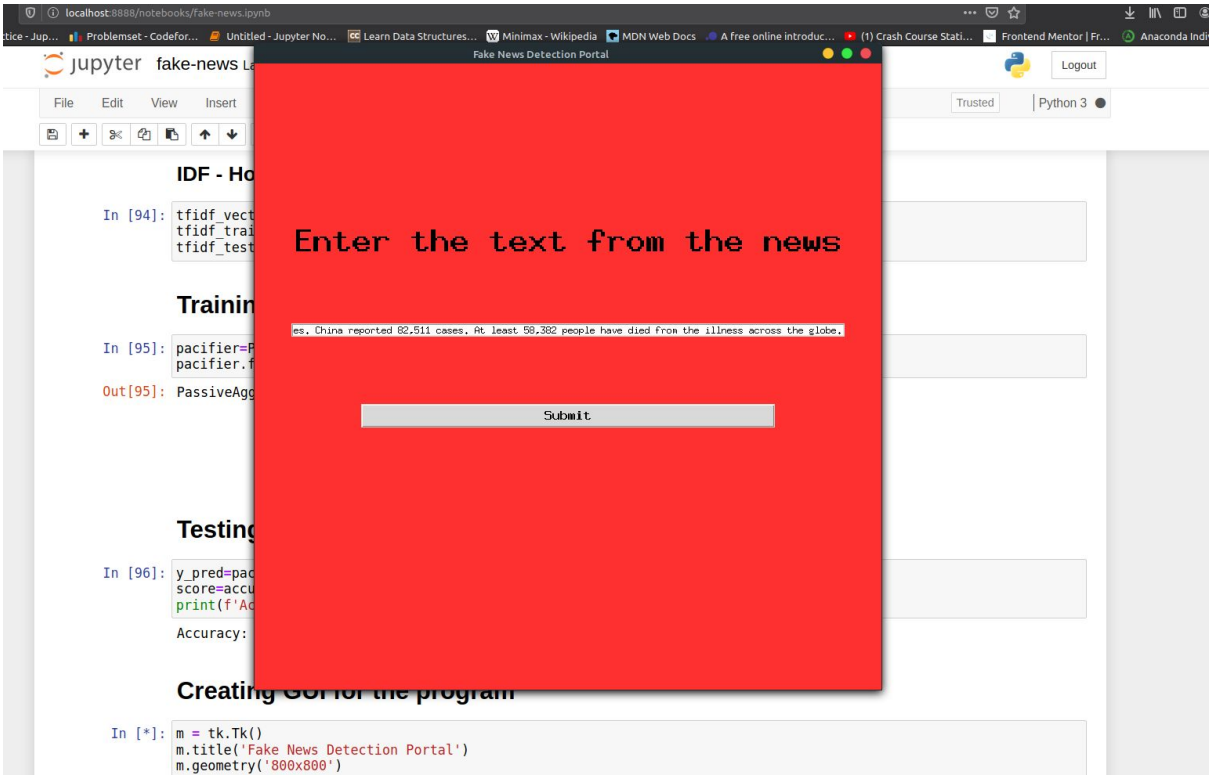
China has enacted restrictions of its own as officials there say the crisis there is winding down. Among them was limiting the New England Patriots' private jet to just three hours on the ground when it flew there this week for an emergency shipment of over a million N95 masks. The crew was not allowed to get off the plane.

President Trump has said that he could not confirm the veracity of the numbers of coronavirus cases and deaths being reported by the government in China, as he urged Americans to adhere to guidelines on social distancing and other health precautions as the outbreak continues here.

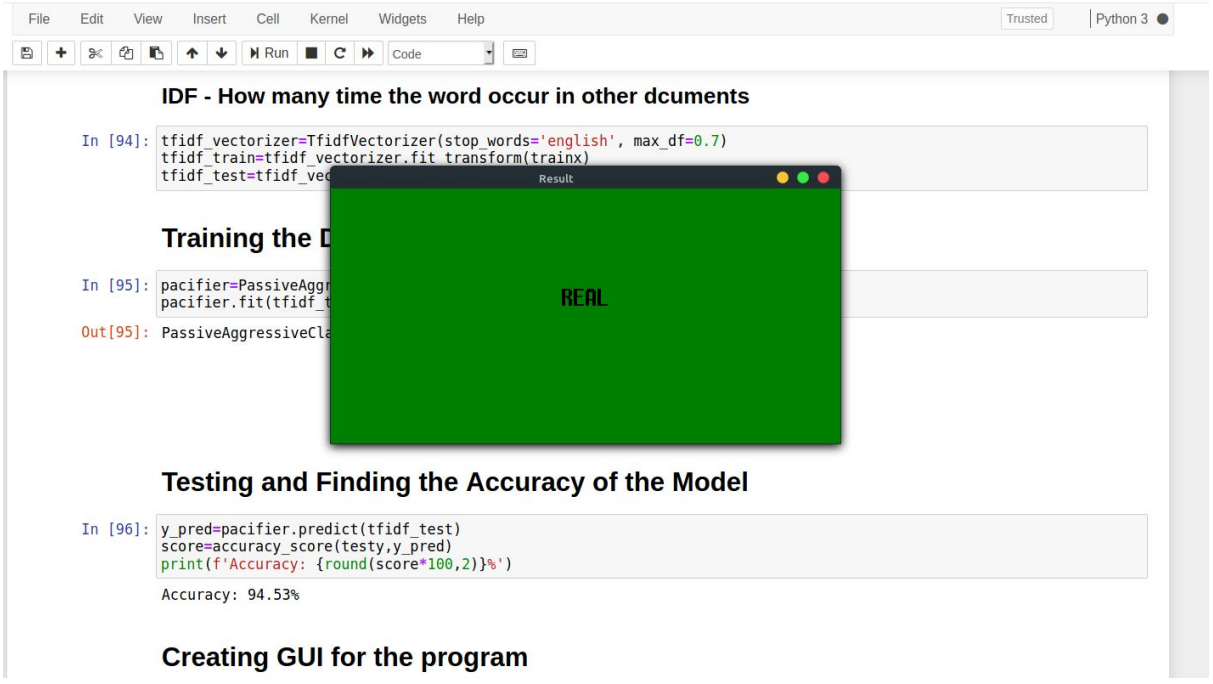
There were at least 1,088,878 confirmed coronavirus cases worldwide Friday night, with 270,473 of them in the United States. China reported 82,511 cases. At least 58,382 people have died from the illness across the globe.

Fox News' Gregg Re contributed to this report.

Copying Raw Text from the news into the Entry area:



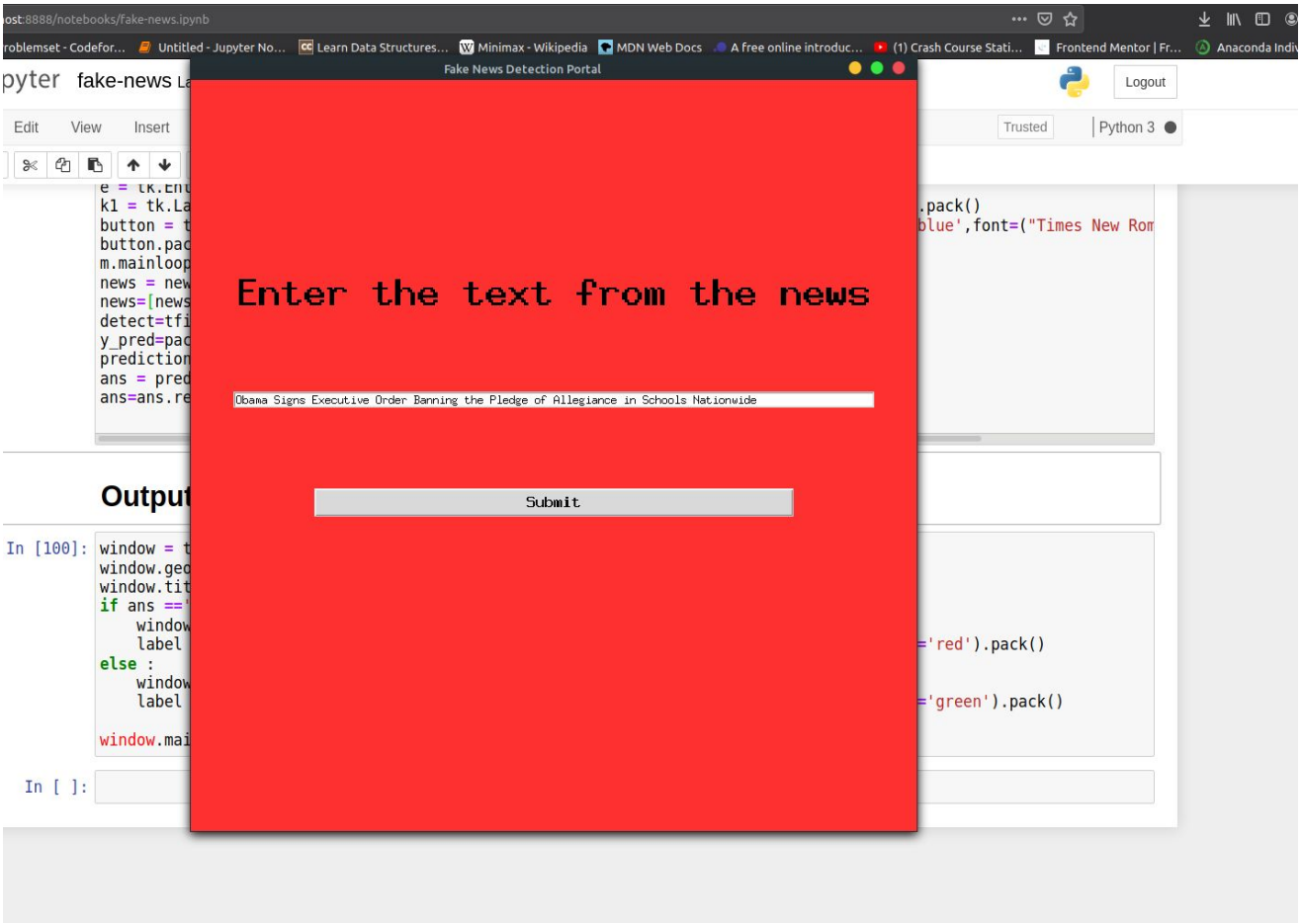
Result :



Testing for Fake news from verified sources :

> ## 4. Obama Signs Executive Order Banning the Pledge of Allegiance in Schools Nationwide
>
> This whopper had over two million shares, comments, and reactions on Facebook (FB). It was published on a site made to look like ABC News. When faced with a website that looks like reputable news, but has sensational headlines, I recommend you take a good look at the URL. In this case, the extra ".co" at the end should be a clear giveaway that this is a fake news site: *ABCNews.com.co* = fake news.

Copying Raw Text from the news into the Entry area:



Result :

LIBRARIES USED :

1. Pandas
2. Numpy
3. Scikit learn
4. Tkinter

REFERENCES :

1. Dataset for fake news detection : Kaggle
   Link : https://www.kaggle.com/c/fake-news/data
2. TfidfVectorizer
   Link: sklearn.feature_extraction.text.TfidfVectorizer – scikit-learn 0.22.2 documentation
3. Passive Aggressive Classifier :
   Link:https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.PassiveAggressiveClassifier.html