**COVID – 19 : PREDICTION , ANALYSIS AND TRENDS**

A PROJECT REPORT

BY

S.Srihari17BCE1308

Shyam Sundar 17BCE1309



**WEB MINING**

**CSE3024**

**SUBMITTED TO**

**DR Shridevi**

**Table of Contents**

**Title Page No**

**Introduction 3**

**Modules with explanation 4-5**

**Results and Performance Analysis6-7**

**Technical Specifications 8**

**Conclusion and Future Scope 9**

**Appendix**

**Code 10-19**

**Introduction**

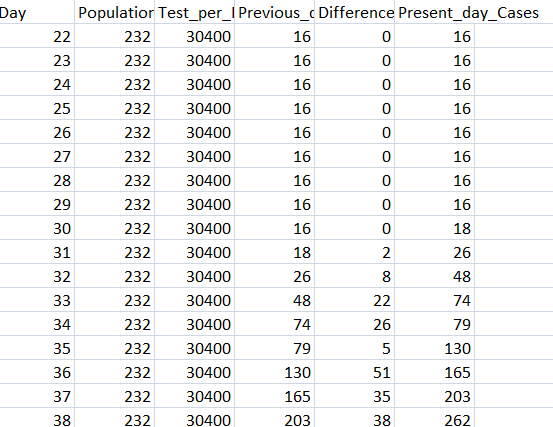
The present COVID – 19 pandemic has changed the world upside down. It has wrecked an havoc across continents irrespective of race , religion , culture among other things . Ever since the crisis came into existence governments across the world have imposed lockdown to curb the effect of the killer virus and they have been aiming at the phenomena what is called “flattening the curve”. Flattening the curve means the no of cases which has been emanating across consecutive days have been very negligible/zero . This effectively means that the spread of the virus in a particular locality has been curbed. News and developments across the world has been very useful in determining the trends across various parts of the world . The aim of the system is to predict the curve of the pandemic and show various related news articles about a country when a country is searched upon.

**Modules**

**Training the ML model for curve prediction**

**1.Dataset creation**

First we have taken a list of countries who have been successful in flattening the curve . then we take the data of cumulative cases per day from the first day of reporting COVID-19 in the country . Then we take the data of test per million people and population density of each country whose data we are taking for training the dataset



The dataset was created in the above manner . The day , Population Density,Test per Million Cases , Previous day Count and the difference b/w consecutive days is taken as the imput values and the present day cases are predicted . The dataset was trained using both linear regression and polynomial regression . As linear regression offered higher accuracy it was chosen above polynomial regression .

**2 . Tracing and predicting the curve**

We take day by day cumulative cases of the country for which we need to predict the curve . Also we take the population density and test per million cases for the country whose curve is to predicted . Based on the trained model and the available data we recursively predict the data(cases) for every next day based on the data predicted for the previous day(cases) . This process is continued until the difference in the no of cases are negligible/close to 0. Then the no of case is plotted versus no of days.

**Search process for searching the news articles related to the country searched**

**1. Scraping , Parsing and Collecting News data**

We take a repository of BBC links which contains news about COVID-19 . We scrape the web pages linked to web links and parse the news content and headlines from the scraped content .

**2. Indexing,TF – IDF Vectorization and save related data in files**

We do Indexing and TF\_IDF Vectorization on the parsed news content and save the data after these processes in files.

For instance we save the lengths of each document in a file, Term/Count/Document data in a file,idf values in one file and other such data

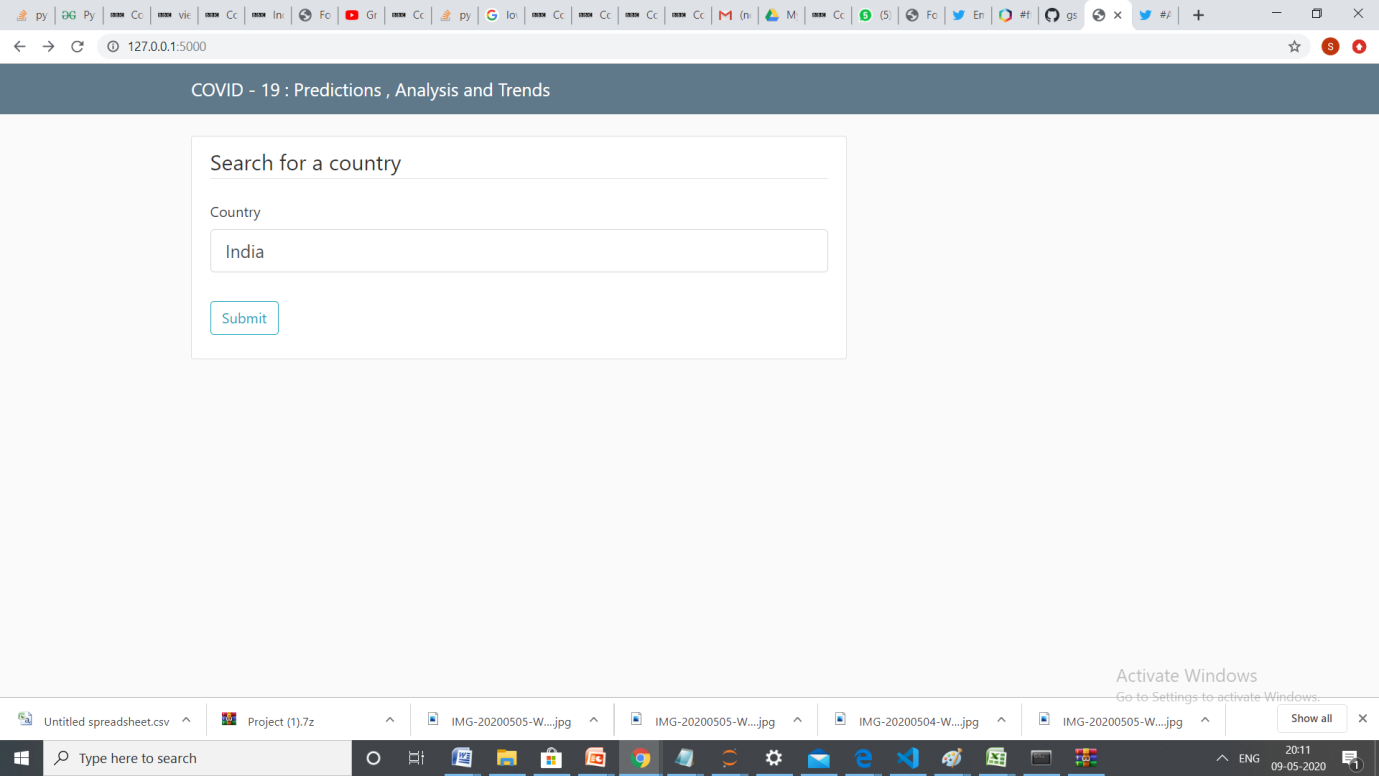
**3.Using cosine similarity for retrieving the most related articles**

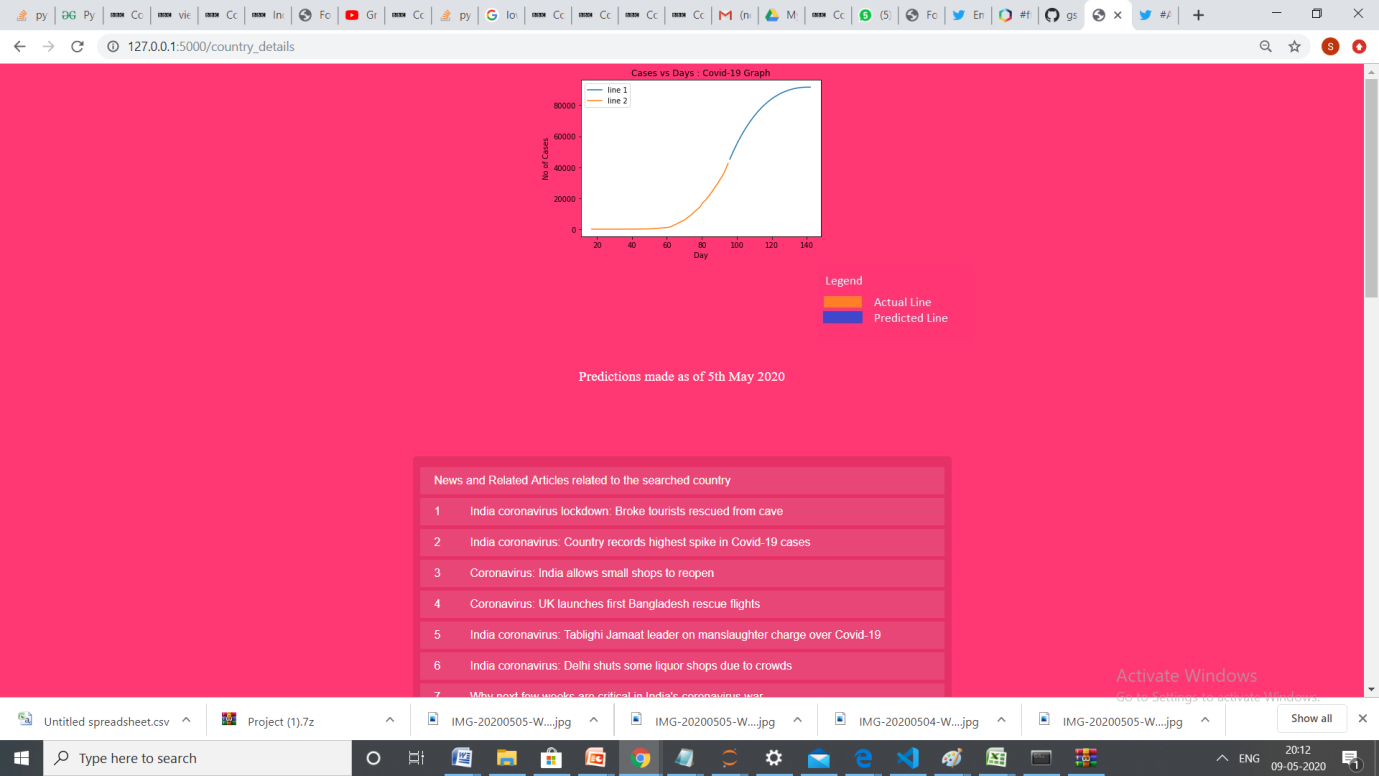
Using the saved data based on the search query we use cosine similarity to sort the links from highest to lowest based on relevancy [similarity].

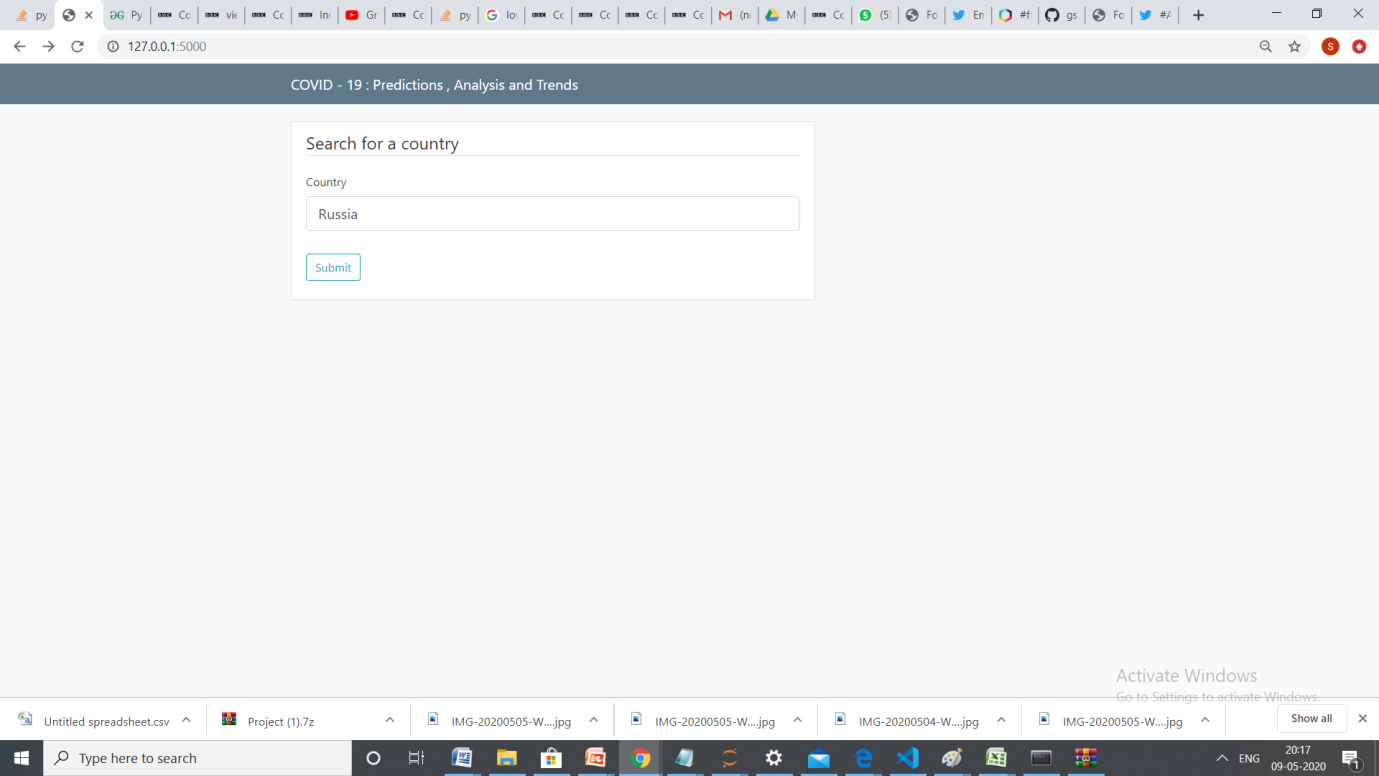
**Working of the website**

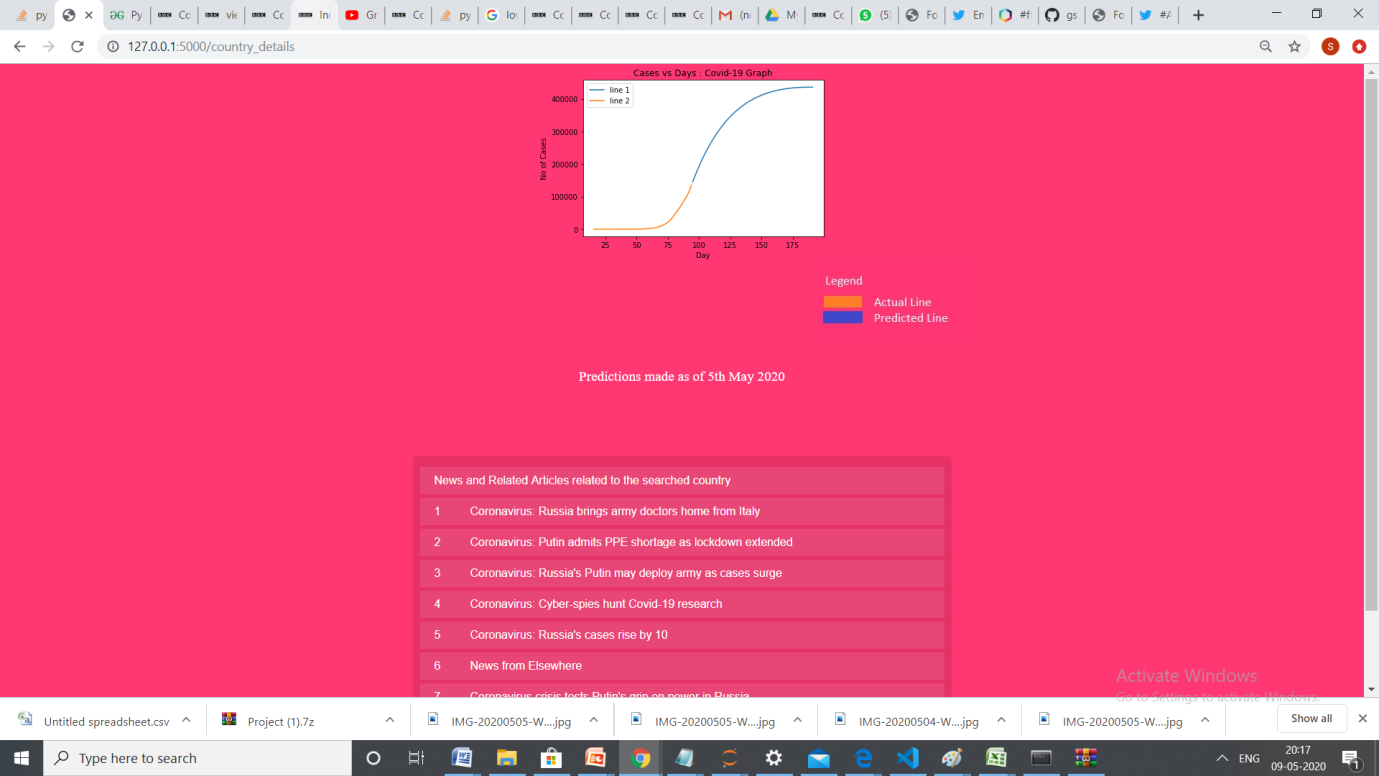
Users can search data about countries and the COVID graph and related web news articles for the particular country

**Result and Performance Analysis**





****

****

**Technical Specifications**

**Software Used**

Flask Framework [Python]

Jupyter Notebook

HTML

CSS

Scikit – learn

Beautiful Soup

And other related libraries

**How to run ?**

python abc.py

**Conclusion and Future Scope**

A successful system was developed wherein based on the search query the COVID-19 predictions followed by related by various news articles . This helps to understand the various trends across various parts of the world followed up with predictions of intensity of COVID – 19 on a particular day. This also gives us an estimate of how long and how much the situation will persist.

There is still scope for improvement though. The project could be extended make a prediction of no of casualities and have a much greater insight on growth rate etc.