

CORONA VIRUS PREDICTION APPLICATION

PROBLEM STATEMENT:

Due to the spread and impact of the novel coronavirus, which has taken the world by storm with its rapid growth, and we need a strong model that predicts how the virus could spread across different countries and regions. The main aim of this application is to build a model that predicts the spread of the virus in the next coming days (provided by user but range from 0 to 20 days). In this project, both linear regression model and Support Vector Machine (SVM) are used which compares that SVM is better than the linear regression.

There are four stages defined according to the level of the spread of COVID-19 such as Import, Contact, Community transmission and Prediction stage. Coronavirus is a chain reaction influencing the most important aspects of our daily life. The growth of this pandemic virus is exponential. It grows extremely fast once it reaches a significant point. After a sharp rise, growth will slow down and such a curve is called sigmoid which means a logistic growth. Each point on sigmoid curve tells the number of cases on a given day and the slope at any given point will show the number of new cases on a daily basis.

PROBLEM ANALYSIS:

The overall process of this application can be broken down into the following four subtasks:

1. Pulling the GitHub raw formatted data into .csv file.
2. Extracting the data using pandas.
3. Cleaning the data like removal of outliers.
4. Visualizing the data using different ways such as bar graph, pie chart etcetera.

This application is a machine learning model used to analyze the impacts of coronavirus, so far and the outbreak of COVID 19 across various regions. The visualization of data is done in

the form of graphs and charts, and ultimately predicting the number of upcoming confirmed cases using two algorithms.

SPECIFICATIONS:

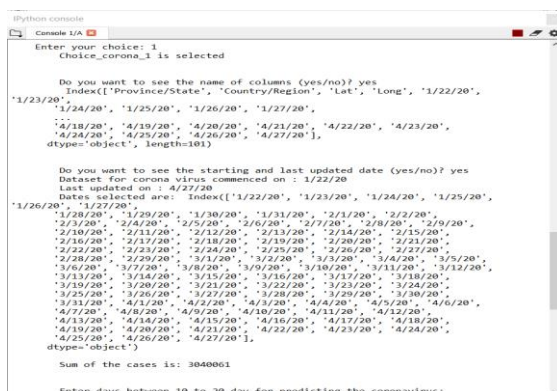
In this model, the raw data is fetched from GitHub and perform operations to clean the data like by replacing missing values with 'Nan'. There are around 4 main variables and the rest for each day. The rest of the variables except last one are used for training the model and last one [04/27/2020] is for estimated value.

The dataset variables are as below:

- Province/State: listed the names of the provinces
- Country/Region: listed the names of the nations
- Lat: Latitudinal value of the area recorded
- Long: Longitudinal value of the region recorded
- Dates till now [Target value]

DESIGN:

The program will start by fetching the data from github and loading it with the help of panda's libraries. It will start by asking the user for choice and then load the data for confirm global, deaths global and recover global.



```

Python console
Console 1/A
Enter your choice: 1
Choice_corona_1 is selected

Do you want to see the name of columns (yes/no)? yes
Index(['Province/State', 'Country/Region', 'Lat', 'Long', '1/22/20',
      '1/23/20', '1/24/20', '1/25/20', '1/26/20', '1/27/20',
      '4/18/20', '4/19/20', '4/20/20', '4/21/20', '4/22/20', '4/23/20',
      '4/24/20', '4/25/20', '4/26/20', '4/27/20'],
      dtype='object', length=18)

Do you want to see the starting and last updated date (yes/no)? yes
Dataset for corona virus commenced on : 1/22/20
Last updated on : 4/27/20
Dates selected are: Index(['1/22/20', '1/23/20', '1/24/20', '1/25/20',
      '1/26/20', '1/27/20', '1/28/20', '1/29/20', '1/30/20', '1/31/20', '2/1/20', '2/2/20',
      '2/3/20', '2/4/20', '2/5/20', '2/6/20', '2/7/20', '2/8/20', '2/9/20',
      '2/10/20', '2/11/20', '2/12/20', '2/13/20', '2/14/20', '2/15/20',
      '2/16/20', '2/17/20', '2/18/20', '2/19/20', '2/20/20', '2/21/20',
      '2/22/20', '2/23/20', '2/24/20', '2/25/20', '2/26/20', '2/27/20',
      '2/28/20', '2/29/20', '3/1/20', '3/2/20', '3/3/20', '3/4/20', '3/5/20',
      '3/6/20', '3/7/20', '3/8/20', '3/9/20', '3/10/20', '3/11/20', '3/12/20',
      '3/13/20', '3/14/20', '3/15/20', '3/16/20', '3/17/20', '3/18/20',
      '3/19/20', '3/20/20', '3/21/20', '3/22/20', '3/23/20', '3/24/20',
      '3/25/20', '3/26/20', '3/27/20', '3/28/20', '3/29/20', '3/30/20',
      '3/31/20', '4/1/20', '4/2/20', '4/3/20', '4/4/20', '4/5/20', '4/6/20',
      '4/7/20', '4/8/20', '4/9/20', '4/10/20', '4/11/20', '4/12/20',
      '4/13/20', '4/14/20', '4/15/20', '4/16/20', '4/17/20', '4/18/20',
      '4/19/20', '4/20/20', '4/21/20', '4/22/20', '4/23/20', '4/24/20',
      '4/25/20', '4/26/20', '4/27/20'],
      dtype='object')

Sum of the cases is: 3040061

Enter days between 10 to 20 day for predicting the coronavirus:

```

Fig. 1- Dates selected from the column

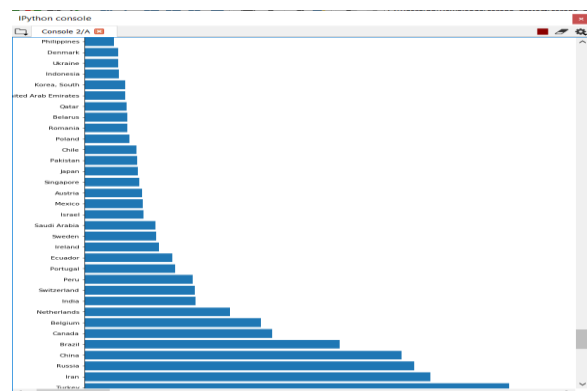


Fig. 2- Bar graph showing the number of

table after entering 'yes'

cases in various nations

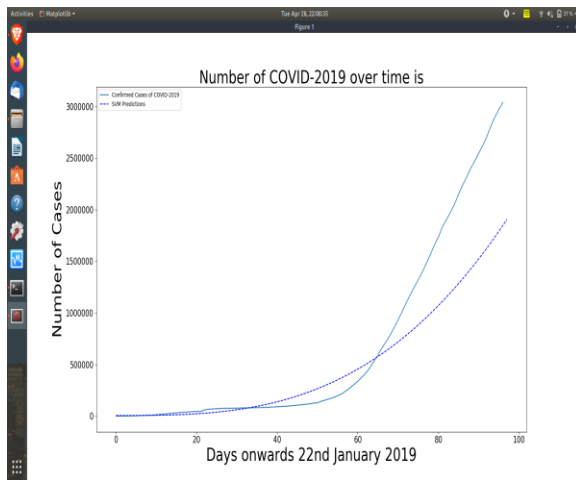


Fig. 3- SVM prediction

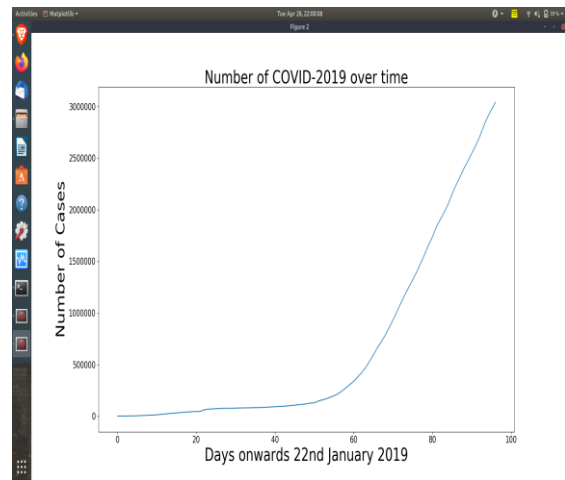


Fig. 4- Increasing slope with respect to
escalating number of cases

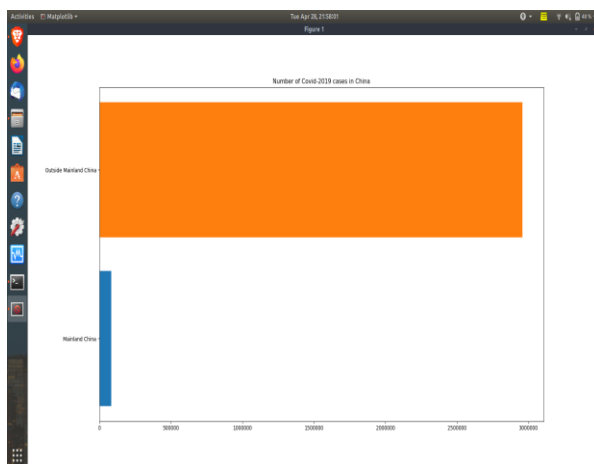


Fig. 5- Data of China v/s Mainland China

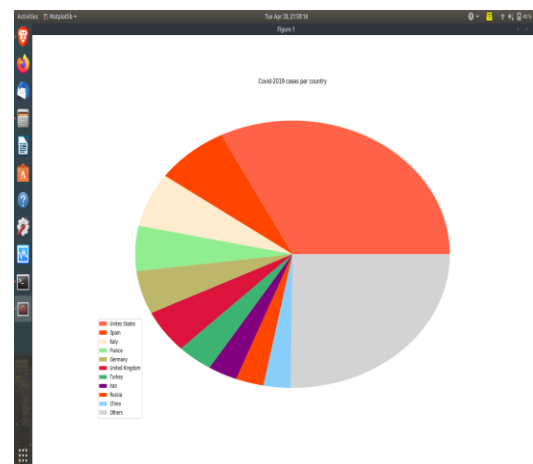


Fig. 6- Top 10 countries shown in pie chart

```
Fitting 4 folds for each of 17 candidates, totalling 68 fits
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 8 concurrent workers.
[Parallel(n_jobs=-1)]: Done 34 tasks | elapsed: 1.3s
[Parallel(n_jobs=-1)]: Done 68 out of 68 | elapsed: 13.0s finished
```

Fig. 7- Steps involved for executing the SVM predictions with 4 folds

TESTING:

Table 1 – List of test cases

S. No	Description	Expected Result	Actual Result	Conclusion
1	Entering “no” when asked for different inputs	It will move forward	Moving to next task of asking another question for different visualization technique	PASS
2	Top 10 countries situation shown with the help of bar graph	Graph should show only top 10 nations affected at that time	Data represented in the form of bar graph is displayed on the screen	PASS
3	Selection of choice other than 1,2 and 3	It should ask you to enter valid choice	Statement showing “invalid character” or “invalid choice” and again have to enter another choice	PASS
4	Range should be within 0 to 20 days	Other than the number from 0 to 20, it should not work and some statement should be shown about out of range	Message displayed showing “invalid character” or “invalid choice” and again ask for valid range	PASS

```

Enter your choice: 1
Choice 1 and corona_1 method is selected

Do you want to see the name of columns (yes/no)? no
Did not enter 'yes', so moving forward

Do you want to see the starting and last updated date (yes/no)? no
Executing preceding code because not written 'yes'

Sum of the cases is: 3040061

Enter days between 0 to 20 day for predicting the coronavirus: 15

Do you want to see the list of countries affected by Corona (yes/no)? no
Did not enter 'yes', so moving forward
Confirmed Positive cases by Country/Regions are:

United States: 988197 cases
Spain: 228422 cases
Italy: 199414 cases
France: 165963 cases
Germany: 158758 cases
United Kingdom: 158348 cases
Turkey: 112261 cases
Iran: 91472 cases
Russia: 87147 cases
China: 83918 cases
Brazil: 67646 cases
Canada: 49616 cases
Belgium: 46687 cases
Netherlands: 38440 cases
India: 29451 cases
Switzerland: 28164 cases
Peru: 28699 cases
Portugal: 24827 cases
Ecuador: 23248 cases
Ireland: 19648 cases
Sweden: 18926 cases
Saudi Arabia: 18811 cases
Israel: 15555 cases
Mexico: 15529 cases

```

Fig. 8- Test case 1

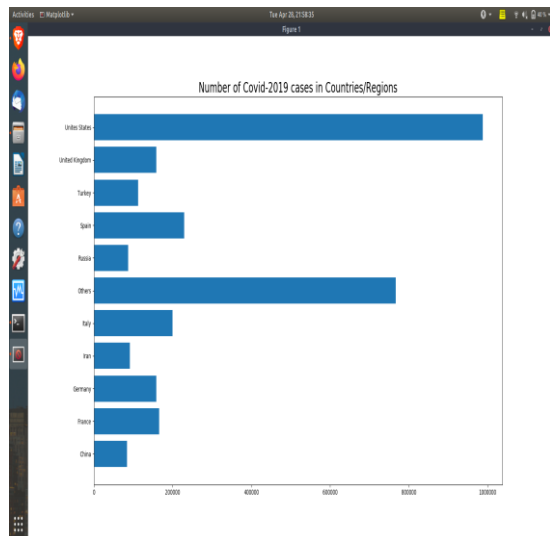


Fig. 9- Test case 2

```

puneet@puneet:~/python_project
puneet@puneet:~/python_project$ python3 main_corona.py
Press 1 to see data for confirmed global cases
Press 2 to see data for global deaths cases
Press 3 to see data for recovered global cases

Enter your choice: 6
Incorrect choice
Enter your choice: 7
Incorrect choice
Enter your choice: er
Invalid character
Enter your choice: 3
Choice 3 and corona_3 method is selected

Do you want to see the name of columns (yes/no)?

```

Fig. 10- Test case 3

```

puneet@puneet:~/python_project$ python3 main_corona.py
Press 1 to see data for confirmed global cases
Press 2 to see data for global deaths cases
Press 3 to see data for recovered global cases

Enter your choice: 6
Incorrect choice
Enter your choice: 7
Incorrect choice
Enter your choice: er
Invalid character
Enter your choice: 3
Choice 3 and corona_3 method is selected

Do you want to see the name of columns (yes/no)? no
Did not enter 'yes', so moving forward

Do you want to see the starting and last updated date (yes/no)? no
Executing preceding code because not written 'yes'

Sum of the cases is: 893747

Enter days between 0 to 20 day for predicting the coronavirus: 22
Incorrect choice

Enter days between 0 to 20 day for predicting the coronavirus: re
Invalid character

Enter days between 0 to 20 day for predicting the coronavirus: 2
<class 'int'>

Do you want to see the list of countries affected by Corona (yes/no)?

```

Fig. 11- Test case 4

IMPACT:

By executing this machine learning model and discussing the outbreak, a person can get to know about the disastrous impact on various countries. This model predicts the number of persons in different regions that can be affected by the spread of this virus in next week.

The implementation of this project gives analytical and visual understanding of the coronavirus prediction model as well as tells the number of confirmed cases in various

nations. The future project will be target on the areas within the provinces or cities where it is surging up and help the government to deploy or allocate the resources to that particular city.

REFERENCES:

- <https://towardsdatascience.com/using-kalman-filter-to-predict-corona-virus-spread-72d91b74cc8>
- <https://medium.com/analytics-vidhya/coronavirus-updated-prediction-using-kalman-filter-3ef8b7a72409>
- <https://towardsdatascience.com/machine-learning-the-coronavirus-9cb8352e1b36>
- https://github.com/CSSEGISandData/COVID19/tree/master/csse_covid_19_data/csse_covid_19_time_series