**Covid-19 visualization, prediction and forecasting**

**1. Abstract**

The goal of the project is to predict coronavirus outbreak using python code. I analyse the outbreak of Coronavirus across various regions, visualize them using charts and graphs, and predict the number of upcoming confirmed cases using the SVM and the Linear Regression model in Python.

It requires precise forecasting the spread of affirmed cases even as investigation of the number of deaths and recoveries. Forecasting, in any case, requires abundant authentic information. Simultaneously, no prediction is bound because the future infrequently rehashes itself similarly because the past. Besides, gauges am impacted by the dependability of the data, personal stakes, and what factors am being anticipated.

Additionally, mental components assume an enormous job in how individuals see and predict the danger from the infection and the dread that it would influence them. Expecting that the data utilized is solid which the longer term will continue following the past example of the illness, our conjectures recommend a proceeding with increment within the affirmed COVID-19 cases with sizable related vulnerability.

**2. Project Design**

I canter around the combined day by day figures collected globally of the three main variables of interest: confirmed cases, deaths and recoveries. The data allude to day-by-day aggregate cases and spread the period from January 22, 2020 until June 22, 2020. While each of the three data designs shows an exponential increment, the patterns of both the confirmed cases and the deaths Ire diminished in the mid of February. Simultaneously, the quantity of recovered cases is consistently expanding.

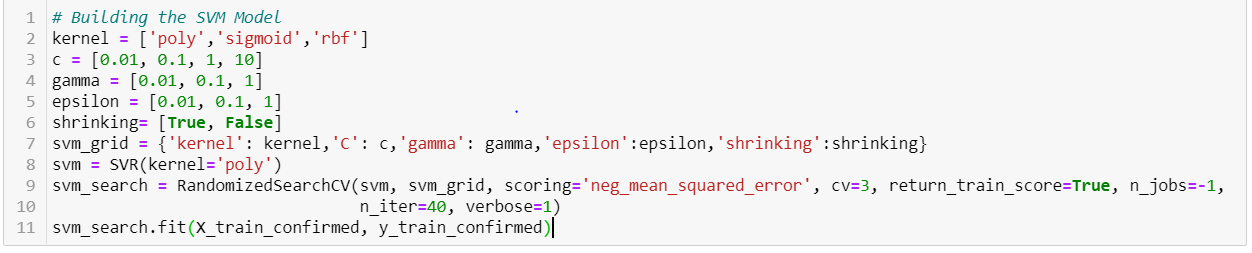
I am using Support vector machine (SVM) and Linear Regression to build our model and predict the total number of confirmed cases that may come us in the next 10 days. The accuracy of traditional forecasting largely depends on the availability of data to base its predictions and estimates of uncertainty.

Building the SVM model and show the confirmed cases vs SVM predicted cases for analysing the result then I will do predictions for the next 10 days using SVM. Next, I am also using the Linear Regression model to make predictions for next 10 days and analyse the outbreak of Coronavirus across various regions, visualize them using charts and graphs.

* For this project I am using python programming language.
* For this project I am using Jupyter Notebook IDE.

2.1 Building the SVM model

SVM uses different parameters to build a model. These parameters am kernel, c, gamma, epsilon, shrinking and svm\_grid.

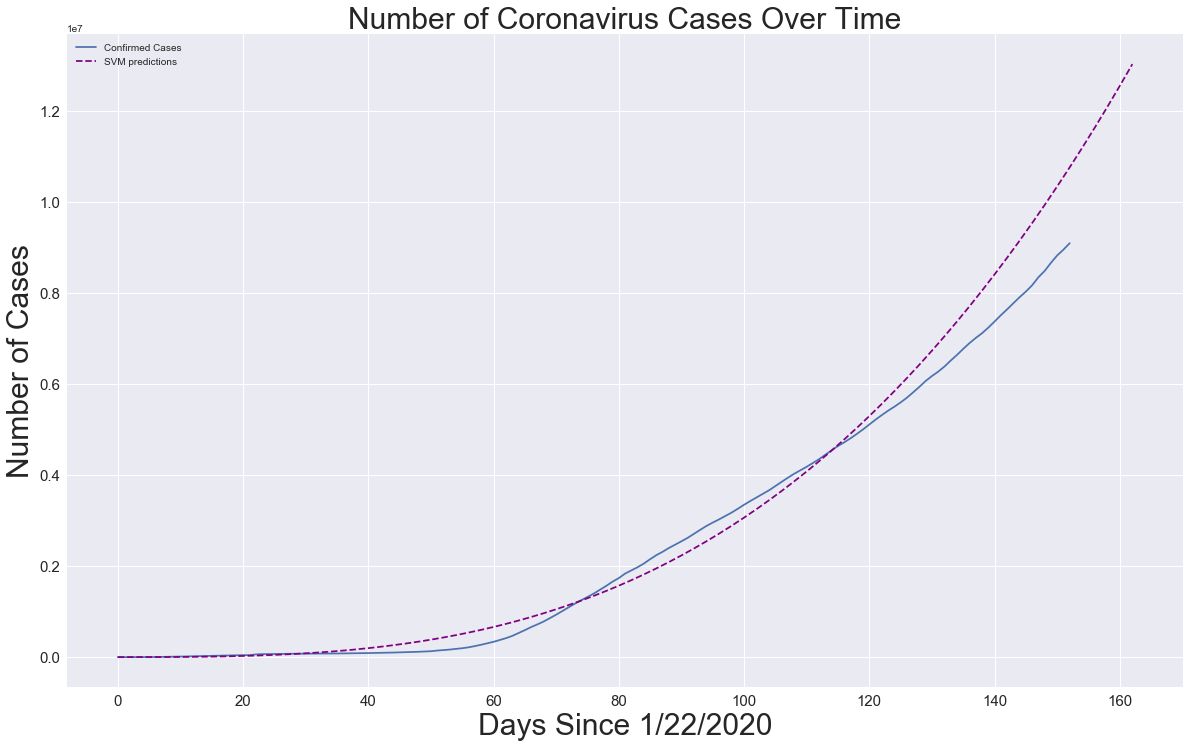


1. Kernel specifies the kernel type to be used in the algorithm
2. C is a regularization parameter
3. Gamma is the kernel coefficient of poly, sigmoid, rbf
4. Epsilon specifies the epsilon tuple within which no penalities associated in the training loss function
5. Shrinking takes Boolean values i.e., true or false
6. Svm grid has all the values passed as parameters to it

With the help of Randomized Searchcv function I am building the function passing necessary parameters

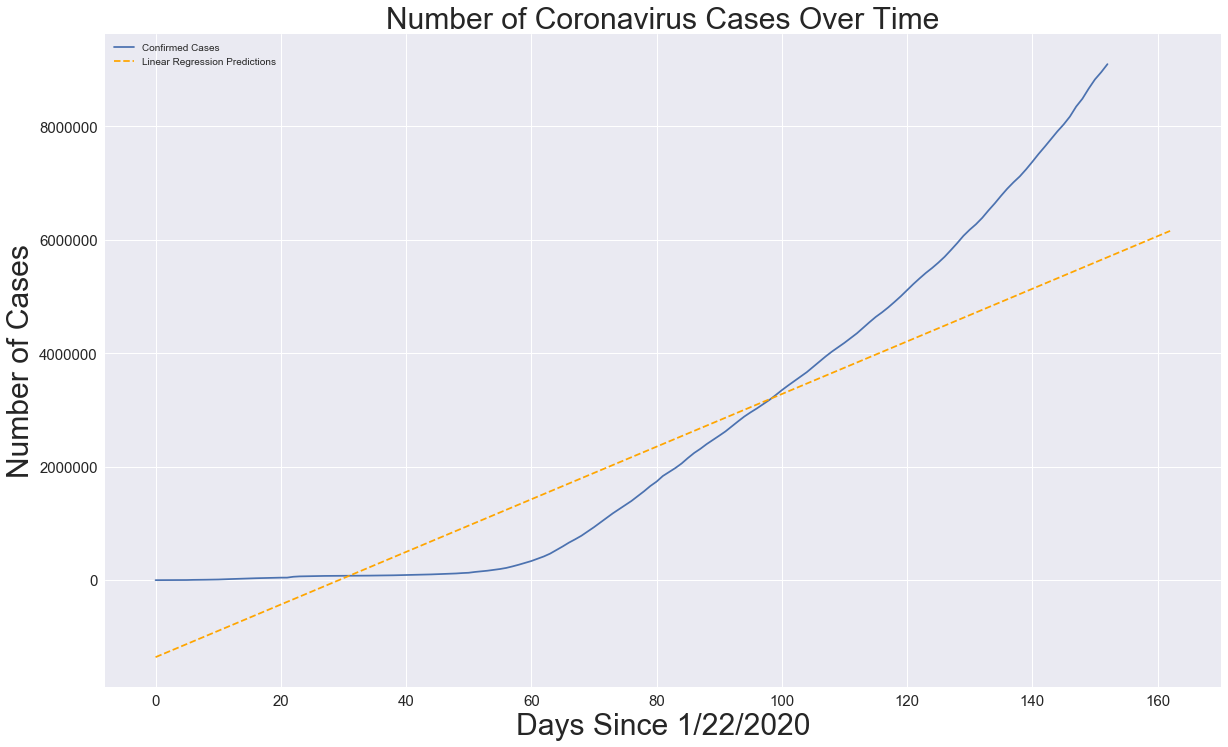
2.2 Confirmed vs SVM Predicted Cases for Number of Coronavirus Cases Over Time

The following is the confirmed number of coronavirus cases over time vs SVM predicted number of coronavirus cases over time from January 22, 2020 until June 22, 2020



2.3 Confirmed vs Linear Regression Predicted Cases for Number of Coronavirus Cases Over Time

The following is the confirmed number of coronavirus cases over time vs Linear Regression Predicted number of coronavirus cases over time from January 22, 2020 until June 22, 2020



2.4 Results of SVM future Predictions and Linear regression future predictions for Next 10 days

|  |  |  |
| --- | --- | --- |
| Date | SVM future predictions | Linear regression future predictions |
| '06/23/2020' | 10979931 | 5740468 |
| '06/24/2020' | 11196596 | 5786857 |
| '06/25/2020' | 11416093 | 5833246 |
| '06/26/2020' | 11638441 | 5879635 |
| '06/27/2020' | 11863658 | 5926024 |
| '06/28/2020' | 12091762 | 5972413 |
| '06/29/2020' | 12322772 | 6018802 |
| '06/30/2020' | 12556706 | 6065191 |
| '07/01/2020' | 12793583 | 6111580 |
| '07/02/2020' | 13033420 | 6157969 |

2.5 Error values of VM and Linear regression model

|  |  |  |
| --- | --- | --- |
| Error | SVM prediction | Linear regression predictions |
| MAE (mean absolute error) | 1090973 | 2376440 |
| MSE (mean squamd error) | 1300741123700 | 5979711806506 |

Here I can see that SVM model performs better than Linear regression model.

**3. Milestones:**

1. Extract, convert and calculate the data for our use.

2. Building the SVM model and showing confirmed cases vs SVM predicted cases.

3. Predictions for the next 10 days using SVM.

4. Use Linear Regression model to make predictions for next 10 days

5. Analyse the outbreak of Coronavirus and visualize them using charts and graphs.

**4. Reference material and Related projects:**

[www.geeksforgeeks.com](http://www.geeksforgeeks.com)

[www.kaggle.com](https://www.kaggle.com/)

I looked up for basic functions, keywords and datasets in these websites.

[www.kaggle.com/therealcyberlord/coronavirus-covid-19-visualization-prediction](https://www.kaggle.com/therealcyberlord/coronavirus-covid-19-visualization-prediction)

[www.kaggle.com/ushashwat/coronavirus-case-study/comments](https://www.kaggle.com/ushashwat/coronavirus-case-study/comments)

Here they have worked on analysing and visualizing various coronavirus outbreaks.