**COVID19 PREDICTION REPORT**

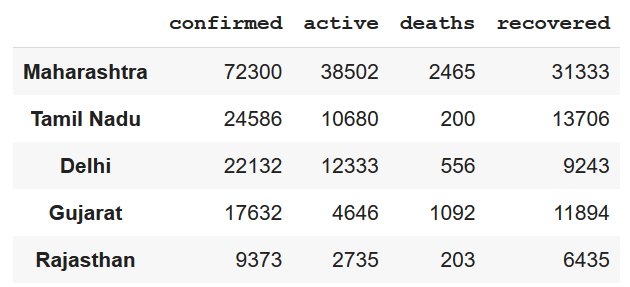
By Subham Kumar Sahoo

**Problem statement:**

COVID19 has become a global pandemic. And the daily total cases of affected people are increasing day by day. So, there is a need to predict the number of affected cases in our country we will be seeing in the future if it increases at same rate.

**Data:**

We got the required state-wise data from covid19india.org and selected the features we will be working on. So, it looks like this:



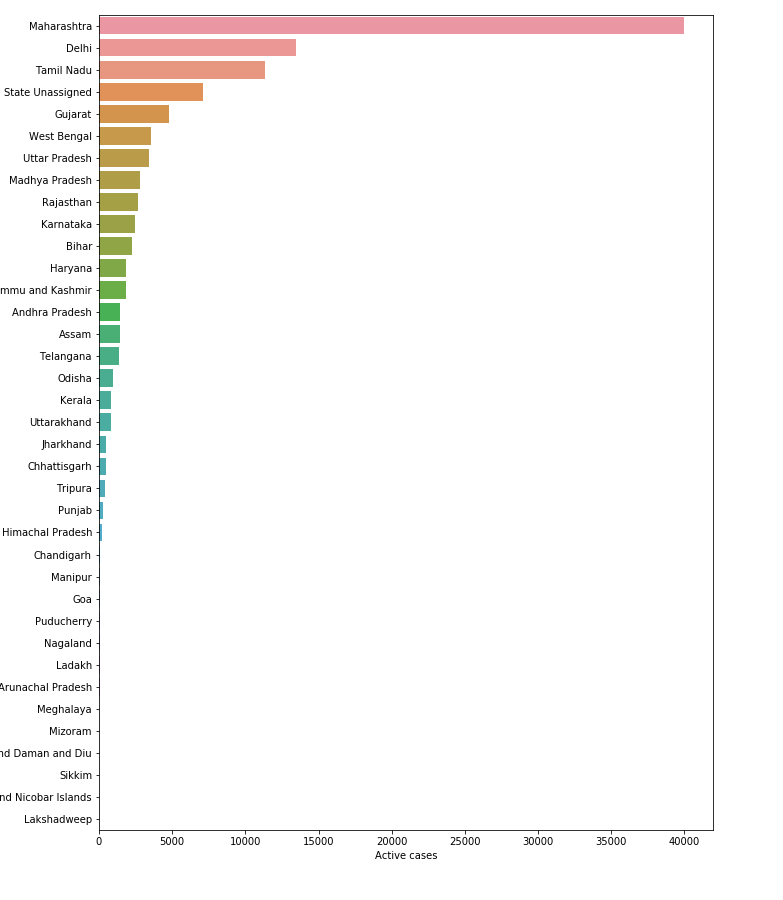
The time-series data looks like this:



We will be using this for future cases prediction.

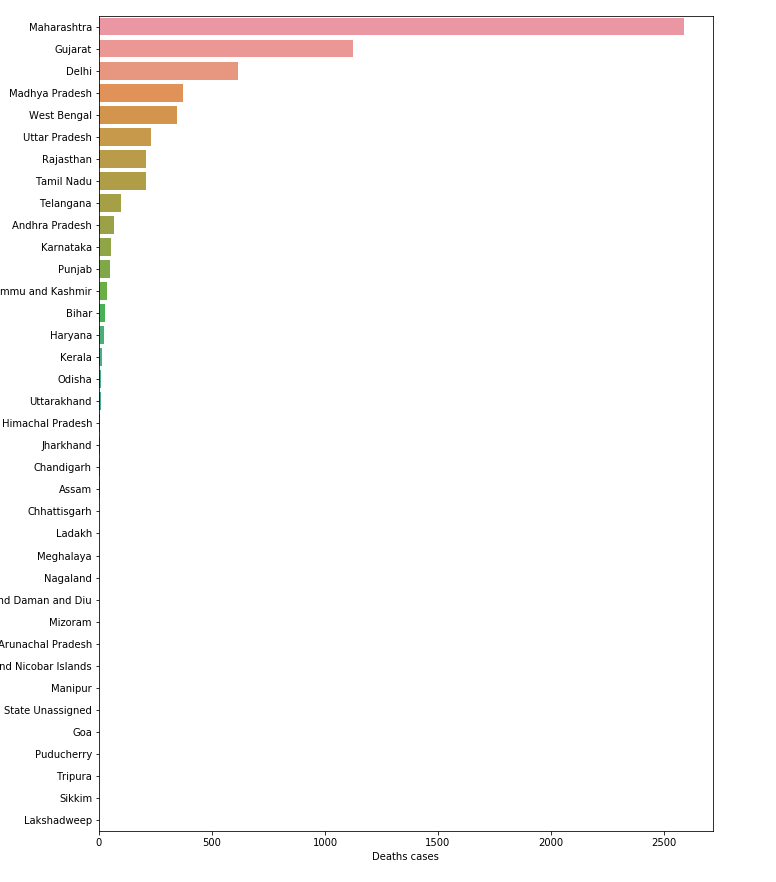
**Analysing the data:**

Active cases per state



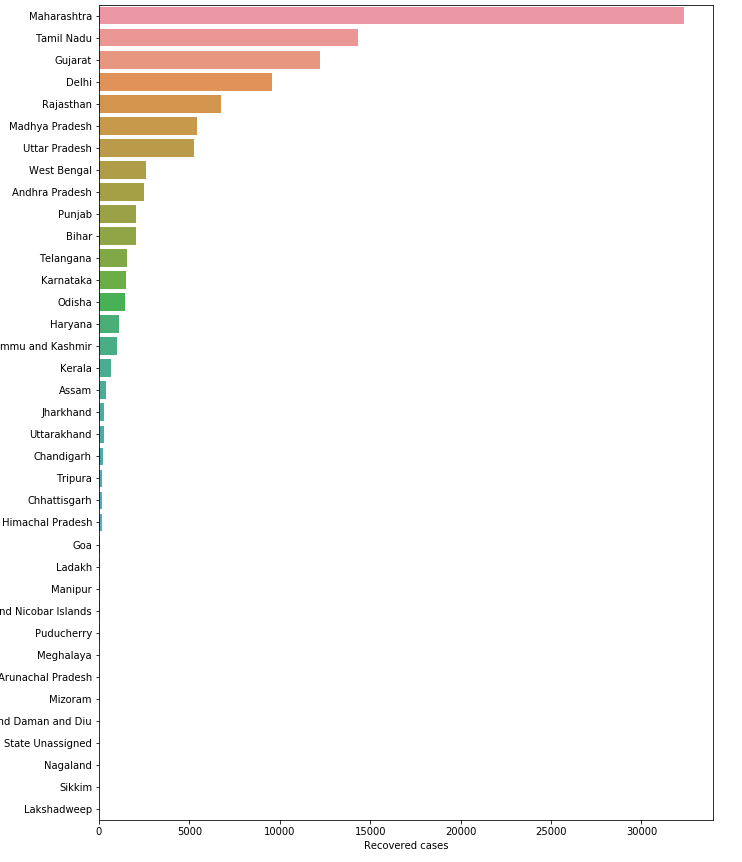
We can see Maharashtra has the most active cases followed by Delhi.

Death cases per state



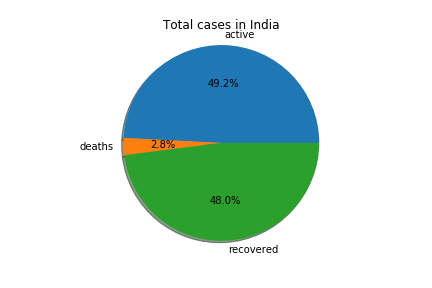
Again Maharashtra is having most registered deaths. But Gujarat, which was on 5th position in most active cases list is now on 2nd position here.

Recovered cases per state



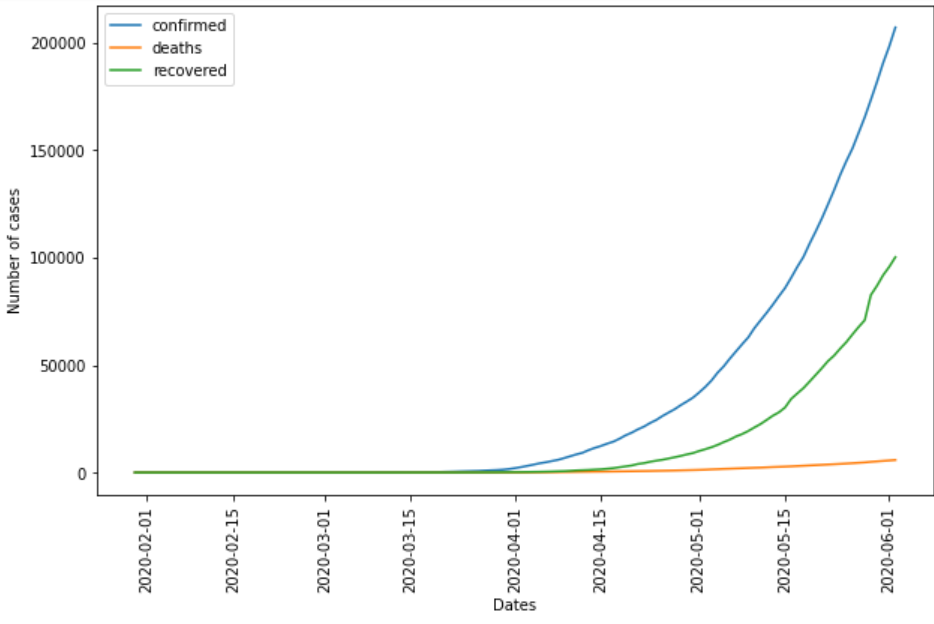
Maharashtra has also most number of recovered cases followed by Tamil Nadu.

Total cases in India:



We can see the number of active and recovered cases are nearly same while number of deaths is quite low.

Date-wise trend in number of cases (till now):

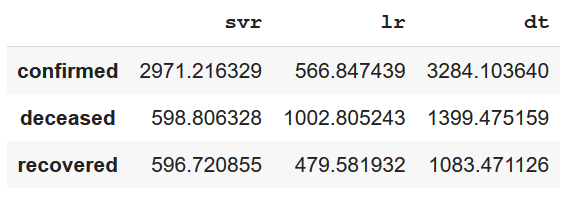


**Building ML model:**

Here we have used Support Vector regression (SVR), linear regression, and Decision tree regression to predict the future cases. So, here we can see the trends are like polynomial curves. Therefore we used polynomial features with linear regression and polynomial kernel with SVR.

**Evaluation metric:** RMSE (Root mean square error) and R2 score

So, the RMSE score of each model for confirmed, recovered and deceased cases are as follows:



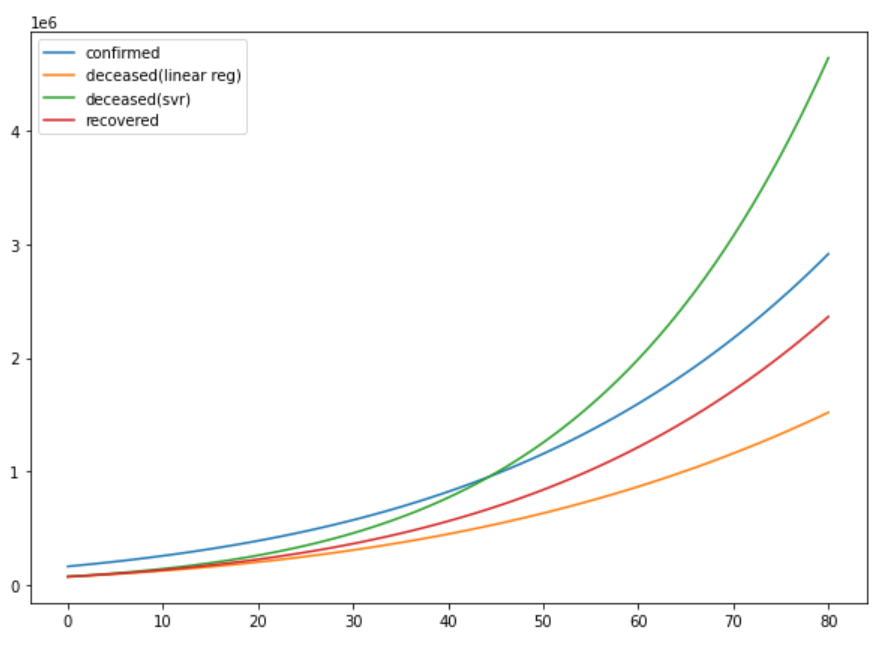
svr: Support vector regression

lr: Linear regression

dt: Decision tree regression

For **confirmed** and **recovered** cases, linear regression model has best score. For **deceased** cases SVR has best score.

Let’s view the predicted values of around 75 days in future.

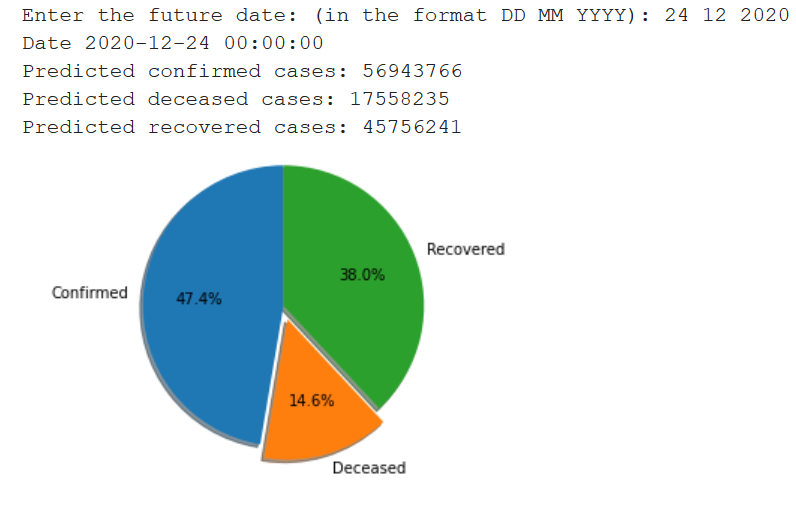


But by comparing predicted confirmed cases with predicted deceased cases (predicted by both Linear regression and SVR) for next 80 days, we can see that deceased cases predicted by SVR is higher than number of confirmed cases which is not the trend according to the historical data we have where deceased cases are quite low compared to confirmed cases.

Along with that for linear regression we got both training and testing R2 score as 0.99 which indicates that the model has captured the variations in the data quite well and there is neither overfitting nor underfitting.

Therefore we are choosing linear regression for predicting all three cases.

The prediction for 12th December 2020 is:



**Conclusion:**

So, we can conclude that the predicted results are quite satisfactory and the model has performed well. As we can see from the visualizations that the Covid19 cases are quite high is India and it is growing exponentially, we need to take proper precautions to control its spread.

***“Stay home, stay safe”***