

## **Tutorial-5**

### **(Vaccination Coverage, Daily cases, Mobility in 2021)**

You are given 3 csv files (Vaccination Coverage, Daily Cases, Mobility in 2021). You have to build a nonlinear regression model to estimate the vaccination coverage in **Delhi** in 2021 using different predictors/regressors and test the model and also compute RMSE error and  $R^2$  value (measure of fit of the model). Vaccination Coverage is to be taken from the day it started till the end (31st October 2021). You are required to proceed in the following manner :

- As it will be very tedious and exhausting to find the right set of regressors for the nonlinear regression model, you can take **Sites/sq.km, Confirmed cases%, death%, recovered% and pharmaceutical/grocery mobility** as the regressors for Delhi. However, you are encouraged to try a different set of regressors or maybe something else which you think could give you a better model with lesser error and better goodness of fit for estimating Vaccination Coverage.
- Find out the **correlation** of the output with each predictor.
- You have to take care of the size of lists that you are going to consider. The size should be same for the output as well as inputs. If the size of one predictor is less than the other, reduce the size of the larger sized predictor accordingly.
- Build non linear regression models taking different values of  $p$ . Compute **RMSE** and  $R^2$  (you can read more about calculating goodness of fit and how to do it in python or R) for each model and gauge which model is the best for analysis.
- Now the estimated list of vaccination coverage is expected to contain some spikes so to estimate vaccination coverage by the end of 31st October in Delhi take the maximum of that list. That's your answer.
- Plot the Actual Vaccination Coverage VS Predicted Vaccination Coverage.
- After you are done with all of this, vary the sites/sq.km to see how it affected vaccination coverage. (You can do anything, like increasing it by some percent on a daily basis or increasing every element by a constant etc.). What can you observe from this ?

Hint: All the input and output lists should be taken from 16th January 2021 to 31st October 2021. Do not waste your time in deciding this. Moreover you can try normalizing and standardizing these lists to build the model though it is not needed.