**Solution Sheet**

1. Which model have you used for Covid Cases prediction? Explain your model.

Random Forests Model

The data provided is skewed with variance varying according to the different features present. Since the data is of covid cases of a real population, it cannot be normalized/log transformed. It will show a false representation of population.

In cases like these, Ensemble methods, ANN and SVMs can be used.

After having tried all the above methods, including ridge, lasso and polynomial regression techniques,

Random forests Regressor gives Maximum accuracy. Score used=r2.

With best selected features (using recursive feature elimination) the accuracy given by random forests is approx 77% and using all the available features present, the accuracy is 83.6%.

Hence, when all features (except city names, population 2001) are used, model is nicely generalization.

Random Forest Model, hyper parameters tuned and generalized.

1. Which model have you used for Foreign Visitors Time series prediction? Explain your model.

Stacked LSTM (Long Short Term Memory)

Foreign visitors is a time dependent feature and the same for September and October is to be predicted. Since 500 cities are present it is a multi variate, multi step LSTM.

LSTMs best suit time series prediction because here, the input in sequential data. The sequence of data was day-wise put-call ratios.

LSTMs have a memory in them that can maintain memory of sequential,time-series data.