**Solution Sheet**

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

For predicting the infection probability, I have used ‘Random Forest Regressor’ as this model performed better than other conventional ML models and Neural Networks.

Random Forest is an ensemble ML algorithm. Ensemble methods uses multiple learning models to obtain a better predictive performance than any of its components. Random Forest uses Decision Trees as it’s components. In a Random Forest Regressor, the mean of output of every component Decision Tree is taken. Each Decision Tree is built by random sampling from the train dataset. This model uses MSE for its loss function.

My Random Forest Regressor has 250 Decision Trees and maximum Depth of a tree is 8. These params are chosen after doing hyper-parameter tuning of the model.

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

For Diuresis Time series prediction, I have used both Polynomial and Linear Regression, these models are chosen based on the visualizations.

First, for a given ‘ID’ Diuresis rate is plotted against time (in days), we can observe from the graph that Diuresis rate varies parabolically with time. Hence, for each ‘ID’, Diuresis rate on March 27th is found out by doing polynomial regression with time.

Then, Diuresis rate on March 20th is plotted against Diuresis rate on March 27th. Here, we can observe that they vary linearly with each other. Hence, Diuresis rate of the people in test dataset is fed to a Linear Regressor and then transformed. Then, the new dataset is used to predict the probability of infection on 27th March.