**Solution Sheet**

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

For predicting the infection probability, I have used Gradient Boost Regressor as this model produced a smaller mean squared error & mean absolute error than other ML algorithms.

Gradient Boosting belongs to family of boosting ML algorithms. Ensemble methods uses multiple learning models to obtain a better predictive performance than any of its components.  Boosting is a method of converting weak learners into strong learners. In boosting, each new tree is a fit on a modified version of the original data set. Gradient Boosting trains many models in a gradual, additive and sequential manner.This model uses LS (Least Square) for its loss function.

My Gradient Boost Regressor has 100 Estimators and maximum Depth of a tree is 3. These parameters are carefully 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 ARIMA (Auto Regressive Moving Integrated Average) model.

First, for a given ‘ID’ Diuresis rate is plotted against time (in days), we can observe from the graph how Diuresis rate varies with time. For optimum performance I compared polynomial regression,Prophet and ARIMA model prediction by predicting 7th day Diuresis rate with the data of 6 days and measured the mean squared error produced by the models and finally chose ARIMA model.

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.