**Solution Sheet**

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

After lots of feature engineering with adaptive binning and binarizing I particularly used random forest regressor after checking other available options such as linear regression, elastic net, multi layer perceptron and xgboost.

For hyper parameter optimization, I used TPOT library which utilize genetic algorithm to find me best combination of hyper parameters. Further random search was also utilized for same tasks.

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

By looking at the data it was well understood that for most IDs diuresis values are in linear fashion. So, I applied log and sqrt transform before fitting linear regression and I found (based on R-squared scores ) that log transform worked slightly better than sqrt transform. Hence I used log transform, trained linear regression model, predicted and took integer value of exp value of predictions as the forecasts. Thereafter I used files from the prediction problem folder to make prediction on infection probability after replacing current diureses values with forecasted values.