**Solution Sheet**

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

The model used for the probability prediction is Random Forests. After analyzing the data given, the first thing to observe would be the total number of features present.

Probability requires it to be a Regression model and Random Forests fits best for this purpose as it works on the basis of ensemble learning.

It would be impossible to utilize other Regression Models such as linear Regression as there are too many features which have to be incorporated for the training purpose. After going through various combinations and logical thinking, some of the features can be excluded while training the model.

Random Forests not only can examine data with great number of features but also can provide us with a collective result of n trees. The number of trees that can be utilized is left upto the user(usually depends on memory availability). Since, the accuracy score of the model must be high, I’ve gone ahead and assigned 10000 trees in order to get a good prediction value. The code and output files are presented in the github link.

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

After carefully going through and analyzing the training set by plotting, it can be noted that the Diuresis value is increasing in a linear way.

Therefore, I concluded to train my model and predict value ‘N patients’ times.

What I’ve done is, train and predict the Diuresis value on 27th N times where N denotes total patients.

I implemented that using for loop. All the predicted values will be exported to initial.csv file.

Then the value for 27th is imported into the ML model used in the first part. i.e. Using random forests, the final probability is determined.

Then this value will update the original Diuresis value in the first model in order to get the infection probability.

Read the instructions file to get further understanding of the model