Solution Sheet

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

   I have used Random Forest Regressor for probability prediction. As probability ranges from 0-100%, with any real value between them, I used regression to predict the probability and then normalised the predictions so that the probability falls between 0-100.

   I have preprocessed the training data by taking care of numerical and categorical variables separately and encapsulating all of this in a data processing pipeline.
   Filling all the empty values, encoding the categorical variables were some of the tasks involved in preprocessing.

   For training the data, I used Random Forest Regressor as it is an ensemble machine learning

model involving multiple decision trees and would work better on large datasets.

I used k-fold cross validation and hyperparameter tuning to improve my model.

Lastly I normalised the predictions so that every prediction falls under 0-100 range.

The Code for this model is present in Flipr_ML.ipynb file.

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

For Time series prediction I used a neural network model consisting of Convolutional Neural Network(CNN) and Long Short Term Memory (LSTM) together wrapped in a model.

The code for this model is present in
Time_Series.ipynb

This data did not require any preprocessing as it was properly labeled already.

LSTM was used so that it could get the relation between terms in a sequence and CNN was used so that it could capture relations between subsequences within a sequence.

I opted to use my own built network instead of pre trained models as this data is different and I trained my model on the data and got the diuresis value for 27-March 2020.

I used the predicted values of Diuresis I got from time-series prediction and retrained the training dataset using the updated values of diuresis and got the results for Part-2.