**Solution Sheet**

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

## Initially I converted the output variable in the training set from probability to 0&1 using a threshold of 50%

## In the pre-processing step I used Knearest Neighbours to fill in the missing values instead of using mean and mode

## Also I didn't standardize the data because Gradient boosting algorithm doesn't assume the data to be in guassian distribution unlike knn algorithm.

## I used Gradient boosting algorithm for training the classifier on 70% of training data and tested on the remaining 30% to get an accuracy of 90.4 percent(accuracy\_score)

## A challenge I faced was to include the regions that the train set never saw but are present in the test set

## Now I trained on the 100% training data to predict probabilites for the given test data and exported them to output.csv

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

## I didn’t complete the part-2

## I would have used arima to predict the diuresis value on 27th march and use the gradient boosting like in part-2 to predict the probabilities.