

## Solution Sheet

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

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean prediction of the individual trees.

I performed exploratory data analytics and overserved distribution of variables, upon performing I came to the conclusion that standard linear and logistic regression are not that efficient for the prediction as distribution was not normal, so I used random forest regression model.

Coming to the model part I used scikit-learn library for pre-processing the dataset. Functions used are imputer (to deal with the missing values) and Label encoder (to convert string values into numeric values). I also tuned the random forest model by changing its documentation's parameter based on loss and r2 score.