**Solution Sheet**

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

I have used RandomForestRegressor model which is an ensemble model from sklearn.

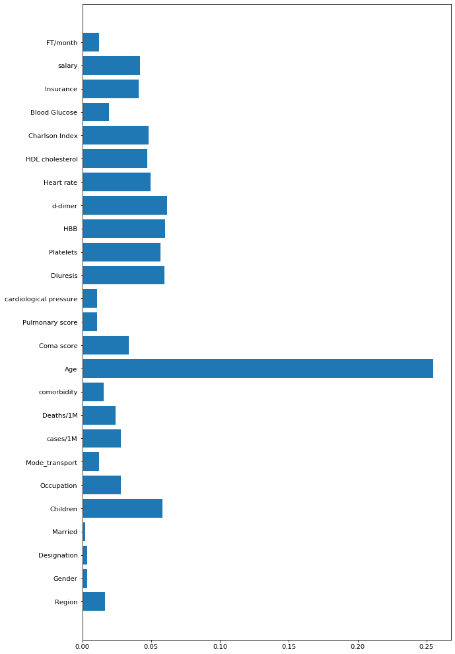
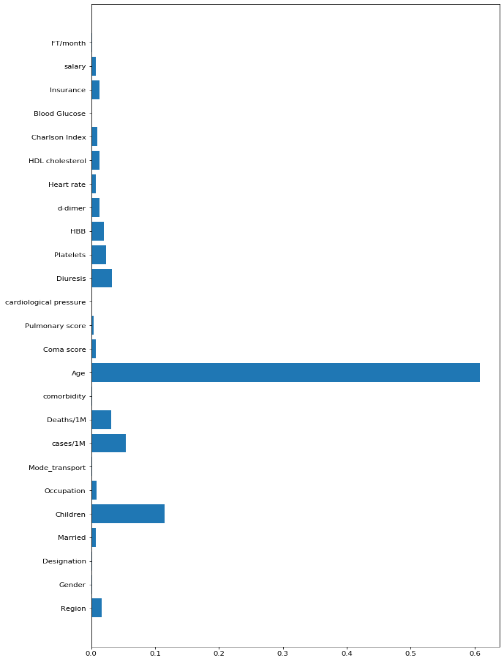
A random forest is a meta estimator that fits a number of classifying decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.

For the project I considered many models such as LinearRegression , GradientBoostingRegressor , CatboostRegressor , XGBoostRegressor but when compared, their score was significantly less than RandomForestReggressor

Then I tried OneHotEncoder instead of category conversion to make a better prediction but it made the whole training process a lot slower without making any positive changes in the prediction score.

After all this analysis I had to make a close comparision between RandomForestRegressor and the GradientBoostingRegressor.

For that I compared their scores after imposing RandomSearchCV and then analyzing their feature performances.

**F.I. for RandomForestRegr. F.I. for GradientBoostingRegre.**

So considering these things I preffered RandomForest over GradientBoosting.

And this is how I made the predictions i.e. Prediction\_20th\_March\_2020 : for Problem 1 and Prediction\_27th\_March\_2020 : for Problem 2 attached along with this solution sheet.

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

The model used for the Diuresis Time Series Prediction is AutoRegression model from statsmodels.

Since the data was limited to only 7 days per person so this was the best model which we could use to get the desired result

For analyzing the model initially I considered the data for the first 6 days as the training data and the data for the 7th day as the test data and achieved the r2\_score of 0.98.

After that I made the diuresis prediction and exported it in Diuresis\_Predictions.csv that were used to predict the Infect\_Prob for Problem 2