**Solution Sheet**

1. Which model have you used for probability prediction? Explain your model.
2. I have deleted extra columns and selectively used only one of the two among of the various highly related attribute/variable pairs.
3. Then I have handled missing data by replacing them with mean/most\_frequent values .
4. I have evaluated various models which I thought ,might give required results such as :-
5. Random Forest Regressor
6. SVR
7. ANN
8. Linear Regression
9. I concluded that the best mean\_squared\_error results were given by **Random Forest Regressor** .

[['Random Forest', 81.41187910796387],

['SVR', 87.70487367605601],

['ANN', 98.29827112299039],

['Multiple Linear Regression', 87.43791803735338],

['Random Forest 5000', 81.29019905898481],

['Random Forest(2000), min\_leaf(5)', 81.29806944624573],

['Random Forest(2000), min\_leaf(5) min\_samples\_split(5)', 81.29806944624573]]

1. Then I tried to optimize the model by changing some parametes , So far I got the best results for

n\_estimator = 2000 , min\_samples\_leaf=5

1. My **final mean\_squared** value for test\_split of 20% was equal to **81.298**.
2. Which model have you used for Diuresis Time series prediction? Explain your model.
3. After analyzing the time-series data provided , I concluded that we can predict the diuresis values for next day using the diuresis values one day before and applying a regression model. No need of a typical time-series model because there is a direct relation between these values.
4. So I converted the data from 7days attribute to only containing values of 2days (Today and tomorrow).
5. Then I trained Random Forest Reression model to predict the values of Tomorrow using the values of Today.
6. I got an amazing result with mean\_squared\_error for 20% test data split equal to 1.980 only .
7. I used this model to predict the diuresis values on 27th March for every people\_id in Test\_dataset.
8. In Test\_dataset ,The Diuresis values on 20th March was replaced by Diuresis values on 27th March(Predicted by Time Series Model).
9. Now Regressor model from previous model was utilized to predict infect\_prob for each people\_Id in the Test\_dataset.

Thank you for organizing such a nice Hackathon.