Project Report

*HOUSING: PRICE PREDICTION*

-AYUSH RANJAN

210103028

1) Firstly, we have imported the data from the given file into the jupyter notebook.

2) Then, we handled the missing data by dropping the columns which have mostly Null values and imputing the object columns by its ‘mode’ and numerical columns by its ‘mean’.

3) After that we encoded the object columns by using “label encoder”.

4) After encoding we started EDA, first (i.e., univariate analysis) we plotted the distribution plot to check normal distribution and corrected it by using “power transformer”. Then we checked the correlation between variables and then removed outliers in the data preprocessing part.

5) We scaled the data after splitting it into x and y using “standard scalar”.

6) Finally, we trained the model using different type of classifiers and evaluation metrics as “mean absolute error” and “r2 score”.

7) “Random Forest Classifier” comes out as the best model and then we predicted values on test data and also done the hyperparameter tuning on “Random Forest”.