

PROJECT HOUSING

Submitted by:

RAVI PRAKASH BAJPAI

**ACKNOWLEDGMENT**

IN this project data was provided by the flip robo ,I have also taken help from my mentor Mr. shweank mishra sir .i have also refered to google and many other websites.

**INTRODUCTION**

* Business Problem Framing

This project is related to prediction of price of the house nearly 80 features are given and price has to be predicted.

* Conceptual Background of the Domain Problem

For understanding this project some knowledge of real state should be there and in real state basically knowledge for housing is required.

* Motivation for the Problem Undertaken

From this project I have learn many things that are applicable in real world scenerio for prediction of price of any house.what are the factor that are necessary for prediction of price.

**Analytical Problem Framing**

* Data Sources and their formats

Data for the given project was provided by flip robo team.data basically consist of feature of both data type that is numeric and object.

Data Preprocessing Done

First all the null value has been filled through mean and mode method.

Than dummie encoding technique is used to convert object data type in numeric data type.

* Hardware and Software Requirements and Tools Used

Hardware required laptop i5 and mouse

Software required juyper notebook

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

For the given data set correlation has been used for finding the multicolinearity than data has been splited into train and test.

* Testing of Identified Approaches (Algorithms)

Ada boost regressor

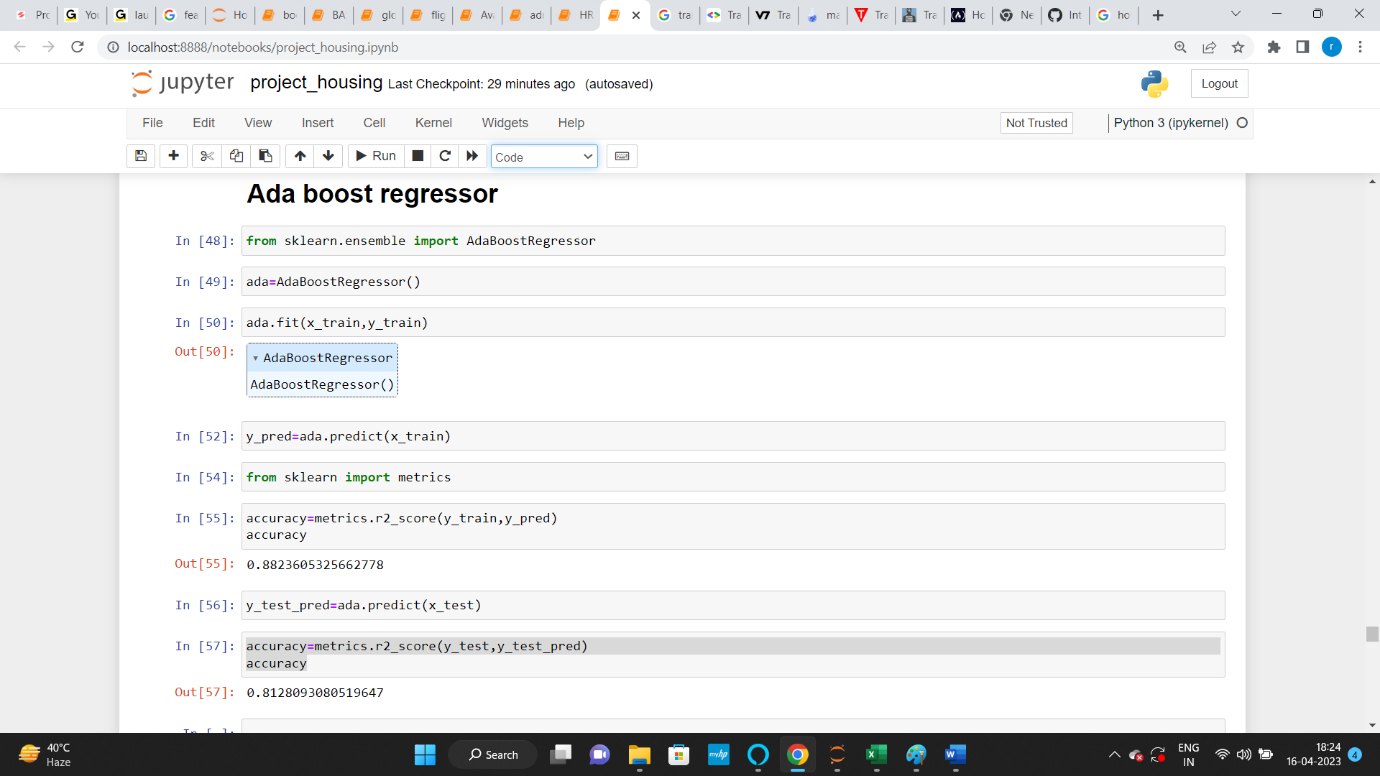
Random forest regressor

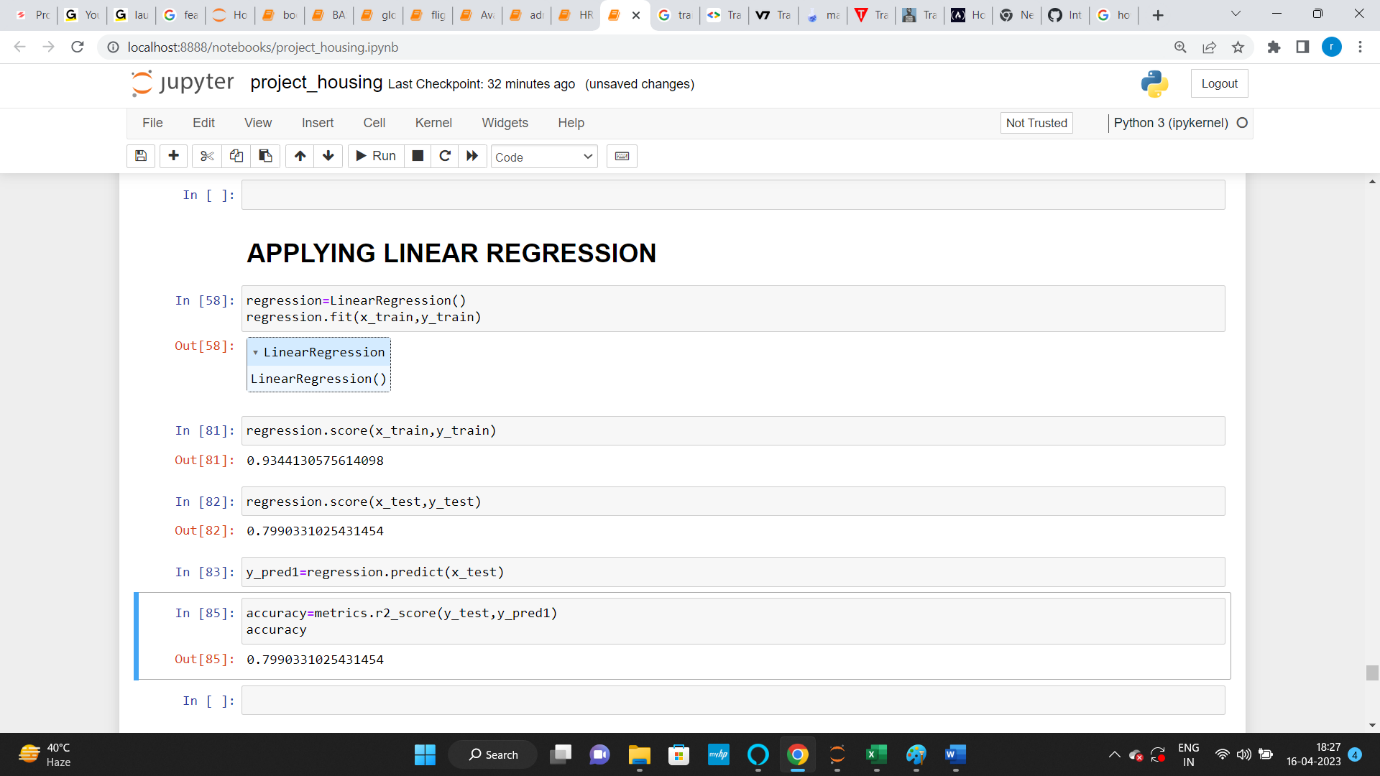
Linear regresssion

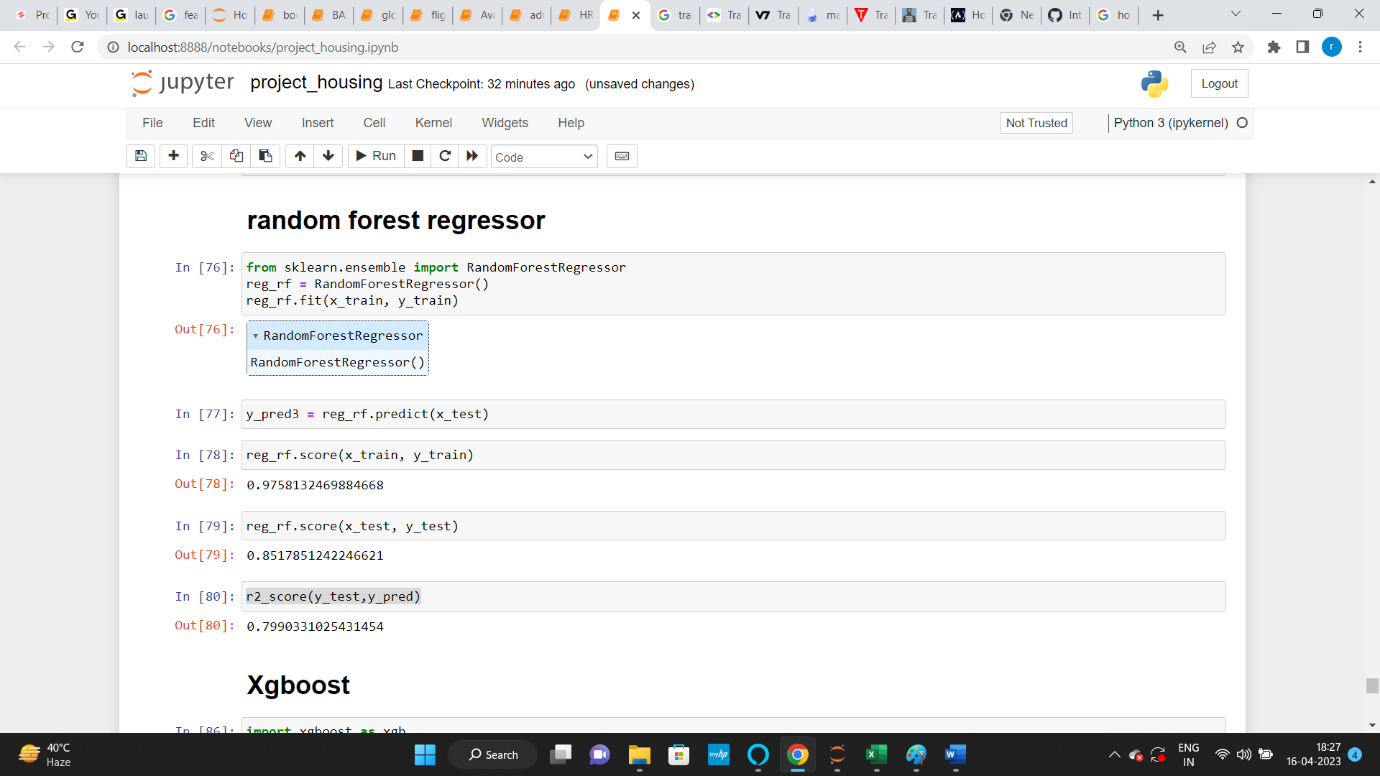
Xgb boost regressor

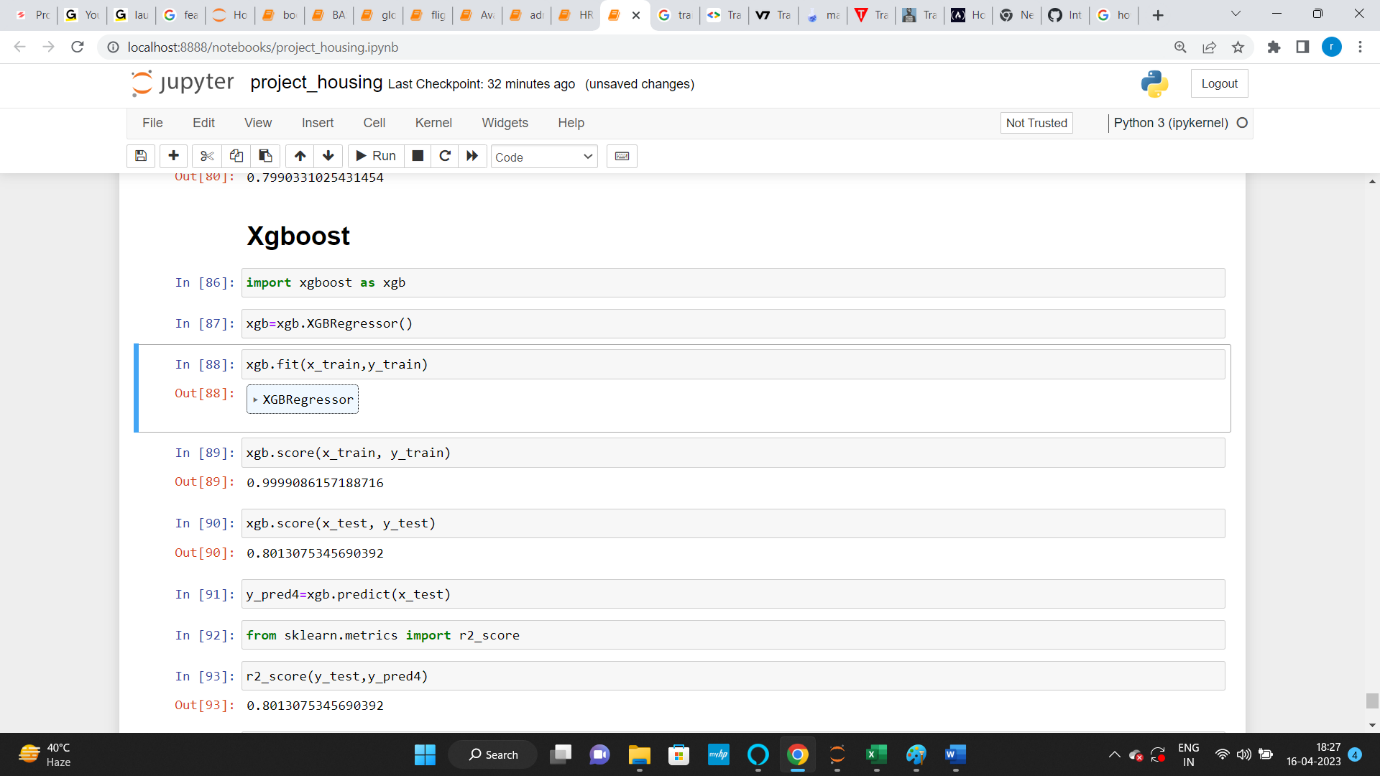
* Run and Evaluate selected modelsgb

Snap shot for the model performance has been attached









* Key Metrics for success in solving problem under consideration

R square has been used to measure the accuracy of the model

* Interpretation of the Results

From this model it can be interpreted that ada boost regressor is working best for the given data set ,as we can see from the above snapshot that all algorithm are nearly giving the same result there is only 1 to 2 % variation in the accuracy.

**CONCLUSION**

* Learning Outcomes of the Study in respect of Data Science

From the given model I have learn the data cleaning part in deeper and also I have made research for feature selection technique as there are many features in this project.

Mainly I have face challenges in dealing with too many features and secondly I have face challenge how to represent visulalization of the project because after applying get dummies feature were nearly 313.

Overall my experience for building this model was good because I have learn a lot,may be I have not applied all the things I have learned but I have over come with many thing.