CAR PRICE PREDICTION PROJECT

Submitted by:

Rajesh Kamatham

ACKNOWLEDGMENT

I like to thank FlipRobo Technology to give a project like this and thanks to Rashi Mathur who though that we will do this project.This project helps me understand more about web scraping and machine learning got to improve my skill. I have taken lots of efforts to complete this project, but without help of internet and team leader could not possible complete this project.

INTRODUCTION

BUSINESS PROBLEM FRAMING

- ✓ *With the covid 19 impact in the market, we have seen lot of changes in the car market. Now some cars are in demand hence making them costly and some are not in demand hence cheaper. One of our clients works with small traders, who sell used cars. With the change in market due to covid 19 impact, our client is facing problems with their previous car price valuation machine learning models. So, they are looking for new machine learning models from new data. We have to make carprice valuation model.*

✓ Conceptual Background of the Domain Problem

- ✓ *So we scrap the data from internet and build the machine learning model that helps us to understand which car is costlier and which car is cheaper.I have perform exploratory data analysis and visualization to give brief idea of what is condition of car market.*

✓ Review of Literature

- ✓ *As seen in the data the manual transmission car has more demand in market any others. The people are going for high brand car like Bmwx5.The price has increased and demand for suv car has increased in market due to covid19.The industries are has lost profit in normal cars. People are taking Good mileage car and manual transmission car.*

✓ Motivation for the Problem Undertaken

- ✓ *Every projects begins with idea that are futher developed and inspired variety of situation and circumstances.The client want prediction of used car in markets.With the data I found people are prefering higher brand and manual transmission car.Also the prices are increased compared to older one.My motivation behind this project is upgrade my skill and to learn new thinks from internet by doing this project.So after doing all the data analysis and programming to give perfect solution for a client.*

Analytical Problem Framing

- ✓ Mathematical/ Analytical Modeling of the Problem
    - ✓ *First I check the information of dataset that its shape,info by appliying pandas it gives all the information about columns and dtypes.*
    - ✓ *After that I check the null values ,there were no null values was present,if it was I have removed it because of our data is text and numeric data.*
    - ✓ *After that I performed summary statistic of dataset.This part tell us the stastistic of our dataset and give the information about mean and median.*
    - ✓ *To visualize the missing value I used matplotlib and seaborn to see the graphical representation.*
    - ✓ *I also checked our data is balance or not.*
    - ✓ *I also performed hyperparameter tuning to find best score.*

- ✓ Data Sources and their formats
    - ✓ *The data I collected by webscraping than I collected data from internet which was in csv format(CommaSeparatedValue)*

- ✓ *The dataset contains 4340 rows and 8 columns.In datasets there were 8 features that is brand,Year,Price of the car,km_driven,fuel,seller_type,transmission,owner.*
- ✓ *The dependent variable was price of the car.*

- Data Preprocessing Done
  - ✓ *First I check the information of dataset and shape of dataset it gives the information about number of columns and rows and type of data present like int,float and object.*
  - ✓ *After that I check statistic summary and correlation with target variable.Also I have done visualization of the data.*
  - ✓ *After doing some data was in object type that converted into integers by imputing dummies values.*
  - ✓ *Than I performed Machine learning and hyperparameter tuning by splitting the data into independent variable and dependent that is x and y.*

- Data Inputs- Logic- Output Relationships
  - ✓ *In this data I focused on visualization and performing hyperparameter tuning. Than i*

conducted the set of learning experiment to build the accurate prediction of price of the car to see the how the input feature affecting the target variable.

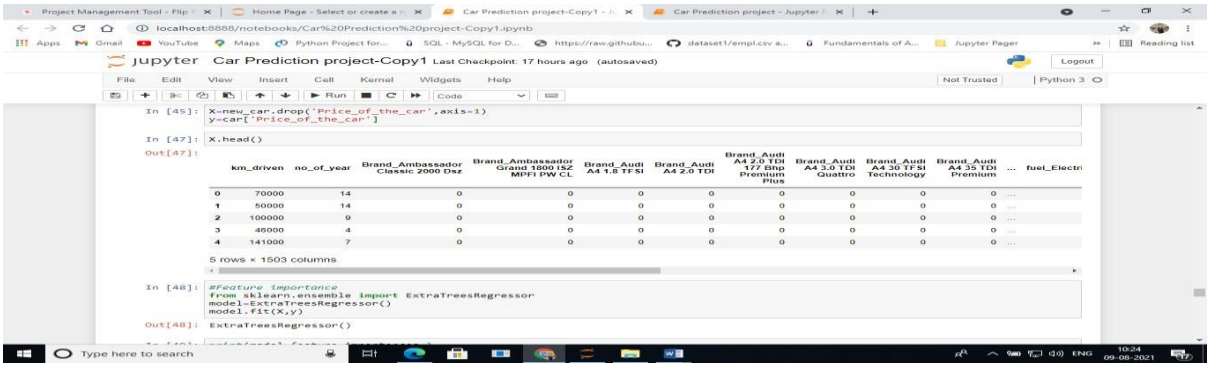
- Hardware and Software Requirements and Tools Used
  - ✓ Processor-intel®core™ i5 CPU M480 @2.67GHz 2.66.
  - ✓ *Installed RAM:4.00GB(3.80 GB Usable)*
  - ✓ *System type:64 bit operating system*
  - ✓ *Software: window 10 pro*
  - ✓ *I used python package because it is powerful and programming language and works faster.*
  - ✓ *Numpy: It is math library operates for numpy array and multidimensional array, it is used for highly mathematical function and scientific computing.it also used less storage.*
  - ✓ *Pandas: It is high level python library and it is easy to use to make dataframe, manipulation of data, processing the data, importing the data and data analysis.*
  - ✓ *Matplotlib: it is provided for 2d and 3d plotting.*
  - ✓ *Seaborn: Seaborn is a python visulization library based on matplotlib, It provides high level interface for plotting attractive and graphical design.*
  - ✓ *(Scikit learn(Sklearn):It is collection of tool and algorithm which is used for machine learning.*

It works with with numpy and scipy it is easy to implement machine learning model.

Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods)

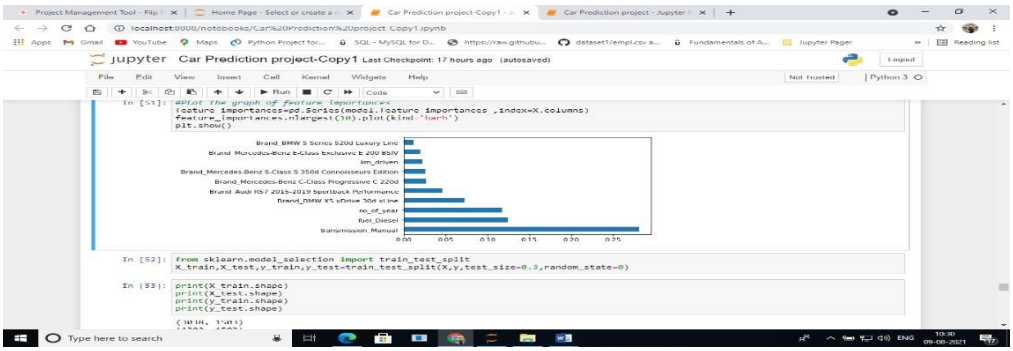
✓ *Before making the model I converted the text into One-Hot- encoder*.



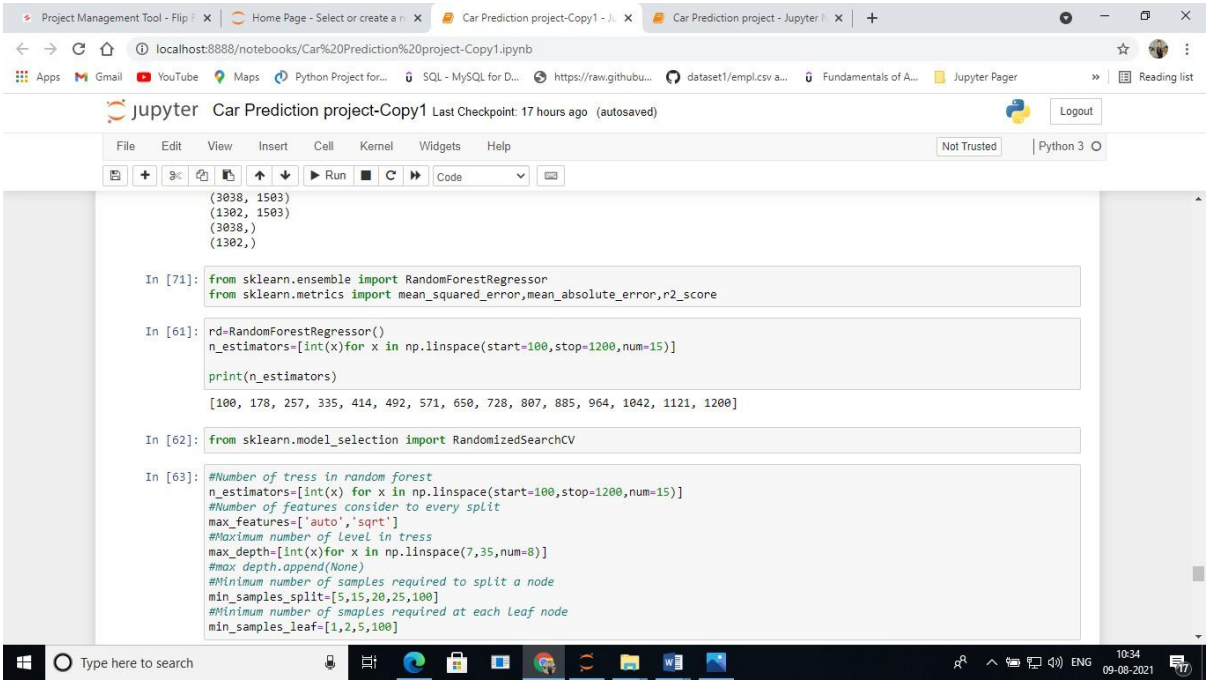
- Testing of Identified Approaches (Algorithms)

✓ *As we know the data need predict price so I gone randomforestregressor and perform hyperparameter tuning.*

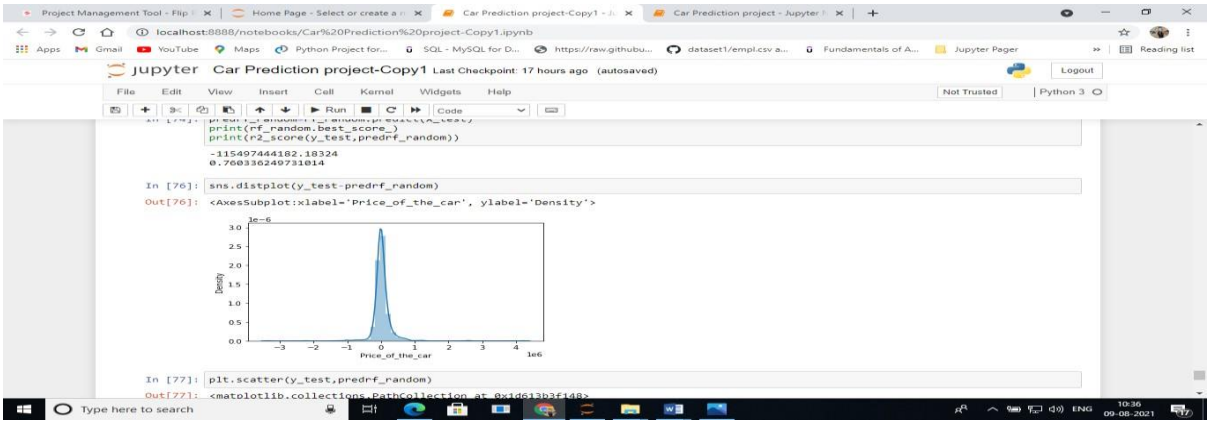
✓ Run and Evaluate selected models



RandomForestRegressor

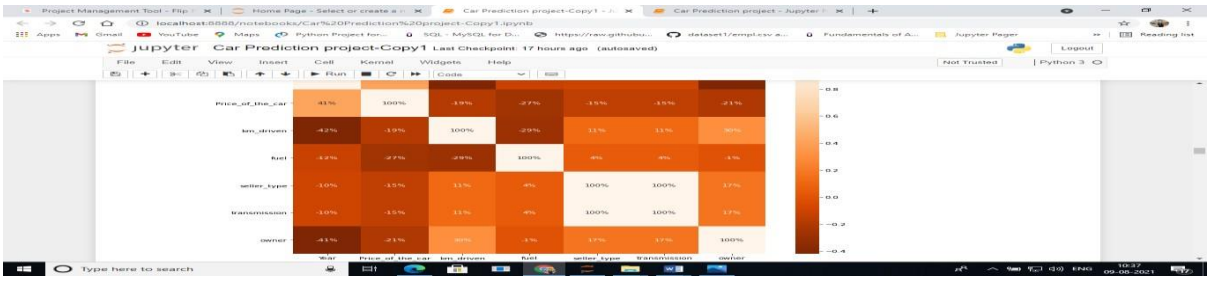


1. Key Metrics for success in solving problem under consideration

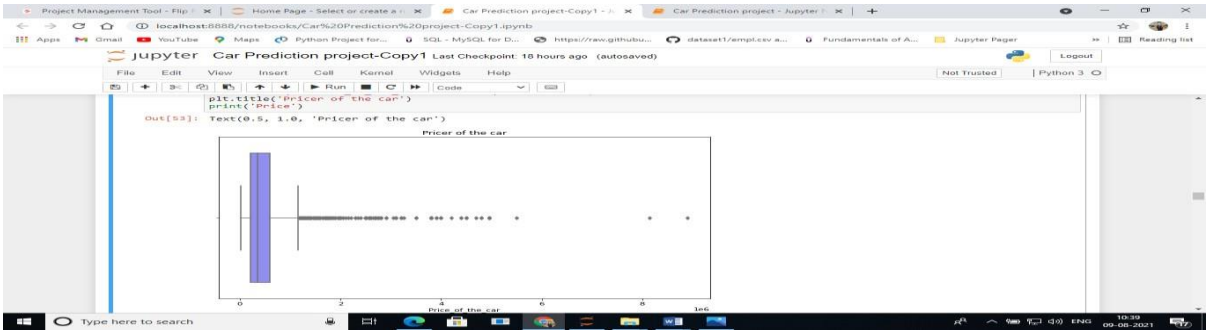


- Visualizations

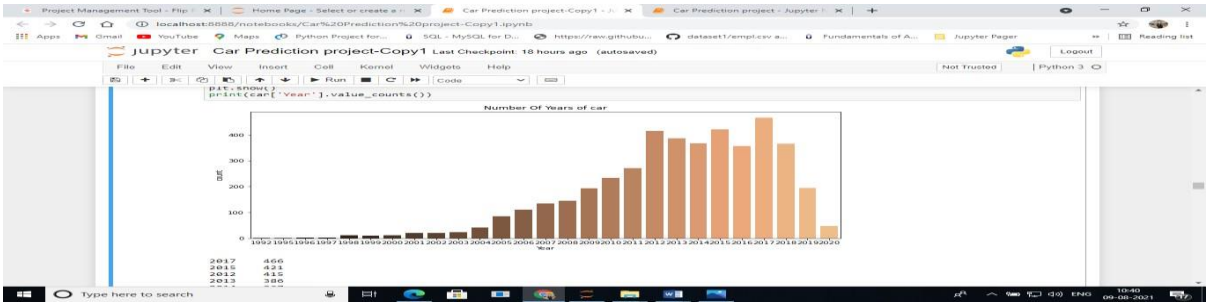
✓ ***Correlation***



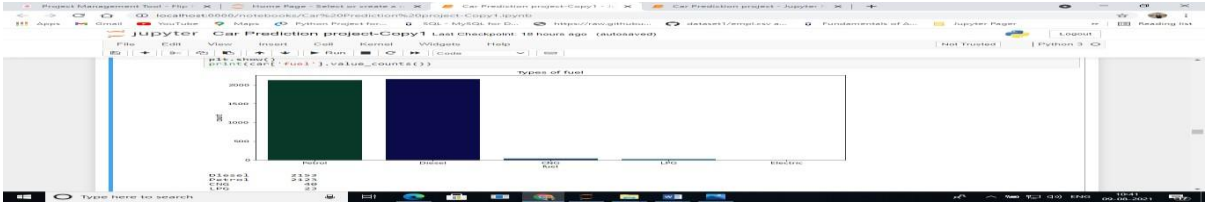
✓ ***Plotting of price of the car.***



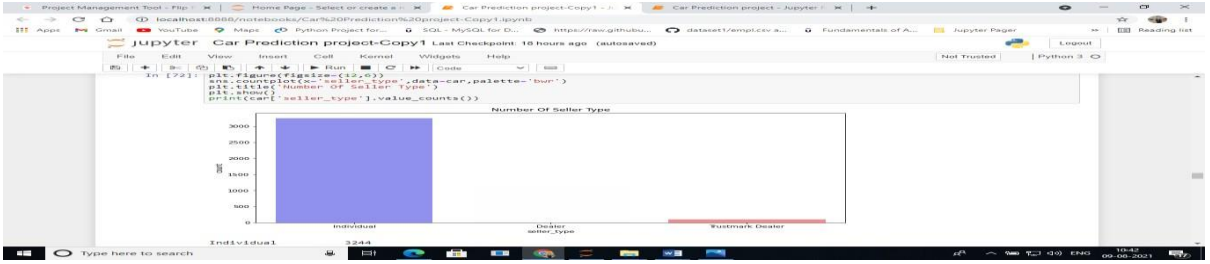
✓ ***Year***



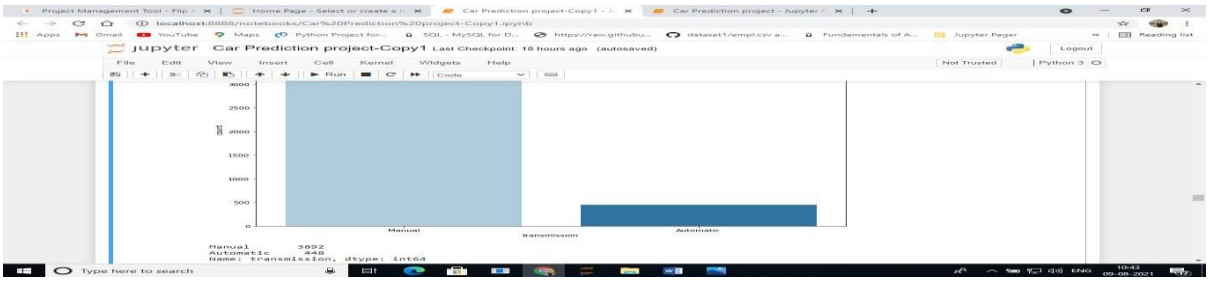
Fuel



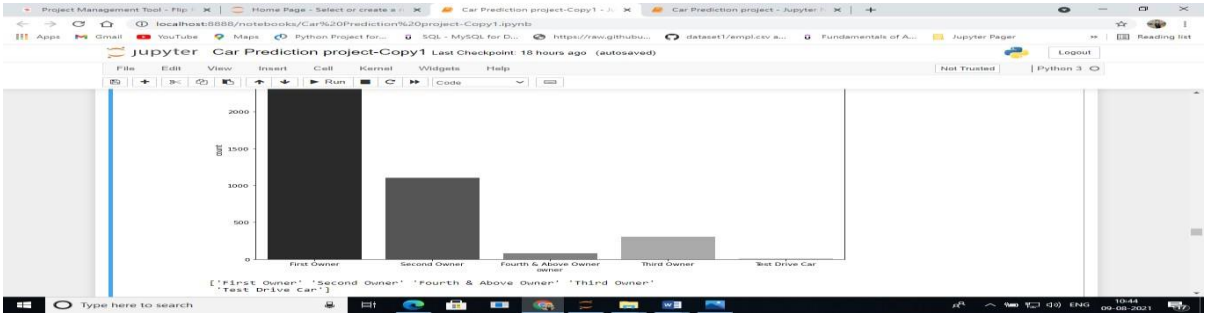
Seller type



Transmission



Number of owners

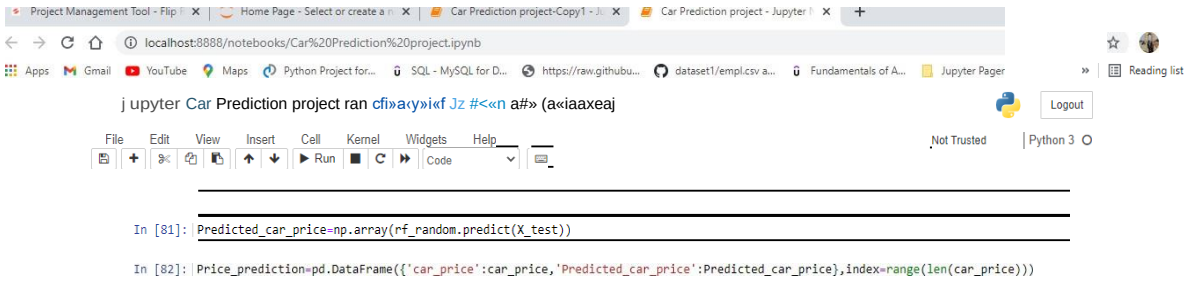


- Interpretation of the Results

✓ *From the above intereption I came to know that this is regression problem so I have build Regression Machine Learning Model.*

✓ *The Goal of my machine Learning model to predict the price of used car and giving a good result to client.*

CONCLUSION



In [81]:
```
Predicted_car_price=np.array(rf_random.predict(X_test))
```

In [82]:
```
Price_prediction=pd.DataFrame({'car_price':car_price,'Predicted_car_price':Predicted_car_price},index=range(len(car_price)))
```