

Car Price Prediction Project

Submitted by:

**Amitava Sarkar**

**ACKNOWLEDGMENT**

**In preparation of this Project, I had to take the help and guidance of some video tutorials, article on medium, Kaggle and several google pages. As the completion of this Project gave me much pleasure. I would like to expand my gratitude to all those who have directly and indirectly who helped me to use wide thinking to explore here in dataset.**

**I want thank Flip Robo Technologies for providing such an opportunity to work on various types of projects which gradually improve my vision to apply for the datasets.**

**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. With the change in market due to covid 19 impact, used car sellers are facing problems with their previous car price valuation machine learning models. So, they are looking for new machine learning models from new data.

* Conceptual Background of the Domain Problem

Due to the pandemic situation all around, every sector is going through an unpredictable situation and the used car sellers are also not an exception of it. In this situation, sellers couldn’t determine what the customer want. The unpredictable increase and decrease of demand of different cars had created a confusion in valuation.

* Review of Literature

There are numerous times when seller couldn’t predict the price and due to that customer get dissatisfied. Customers always try to find things at low value and high benefit. So, it is quite obvious, that they would try to get a good conditioned used vehicle at reasonable rate. It is the seller, who must understand the need of the customer and satisfy them with their wants. Hence, if the seller can determine the market demand and can play their card properly, then definitely they will get benefited.

* Motivation for the Problem Undertaken

Here machine learning can be helpful to such situations. Machine can be trained in such a way that they can predict price based on the current scenario. With the market data collection, machine can predict the right price that can be usefully for both seller and buyer.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

Data collected, contained some numerical and some categorical data. Categorical data were treated with dummy variables. There were lot of duplicate values which were corrected. The skewness of the objects were treated with log transformation. And finally, data were scaled and models were created.

* Data Sources and their formats

The data were scaped using Selenium and finally after collection of required data, it was stored in csv format. Pre-processing were done on it

* Data Preprocessing Done

The data so collected were full of duplicate values that could overfit our model if could be passed to it. Hence duplicate values were replaced and dropped. The symbols and units were replaced by space and object data types were converted to float type.

Data Inputs- Logic- Output Relationships

**The input data for the processing and getting the output is converted in to the numerical forms or data words, in to the numerical form. It feed to the model in the form of series (one by one) which analyses by the model providing the certain score through the medium of performance metrics.**

* Hardware and Software Requirements and Tools Used
* **Hardware And Software required for this project:**
* **Laptop with I3 processor and 4 GB Ram**
* **Google Colab**
* **Python Pandas for processing**
* **Scikit learn library ..**
* **Matplotlib and Seaborn for visualization**

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

**These are the following approaches I have used here:**

**Importing and displaying insights of the data using pandas Dataframe.**

**Analysing the data and use proper pre-processing techniques to clean the it.**

**Applied log transformation to remove the skewness from the data.**

Lastly, scaled the data, after all pre-processing were completed.

* Testing of Identified Approaches (Algorithms)

Following are the list of algorithms used:

Bayesian Ridge

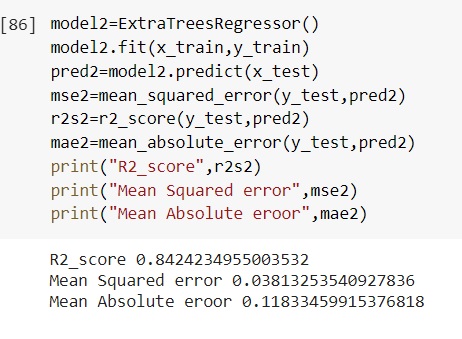
Ridge Regressor

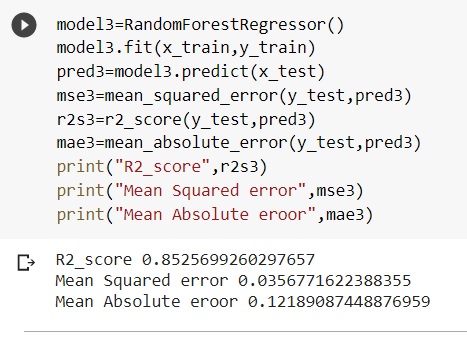
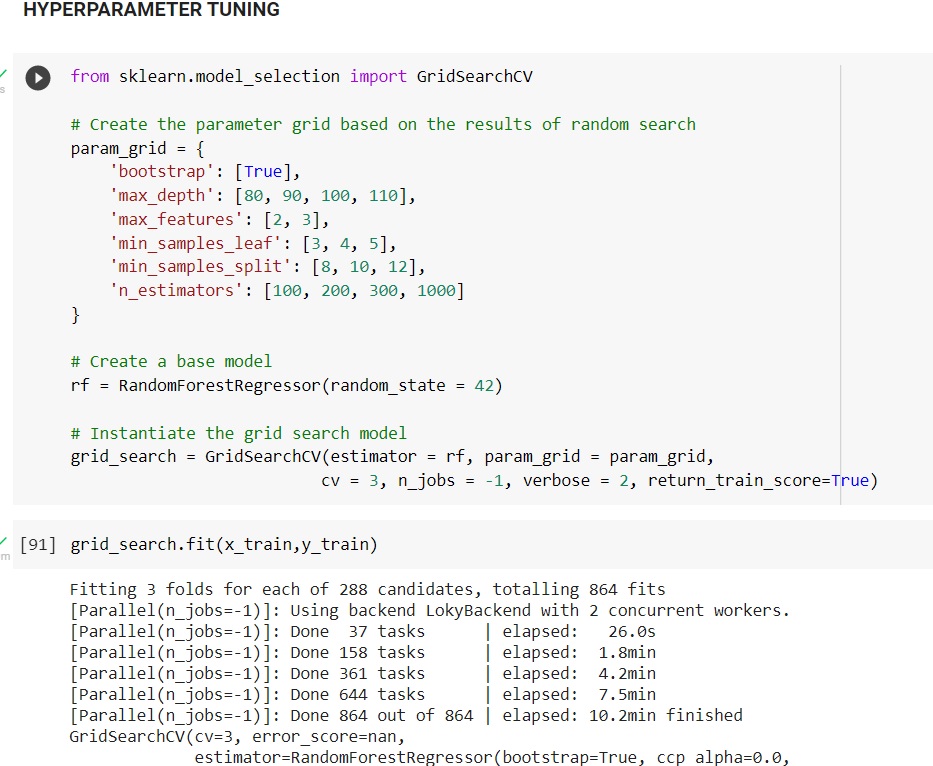
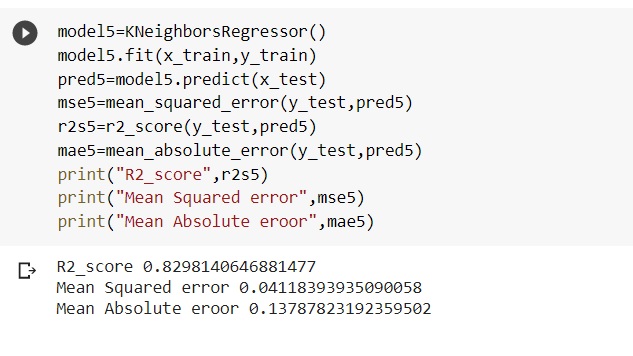
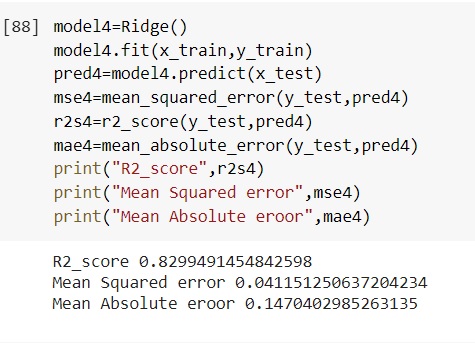
Random Forest Regressor

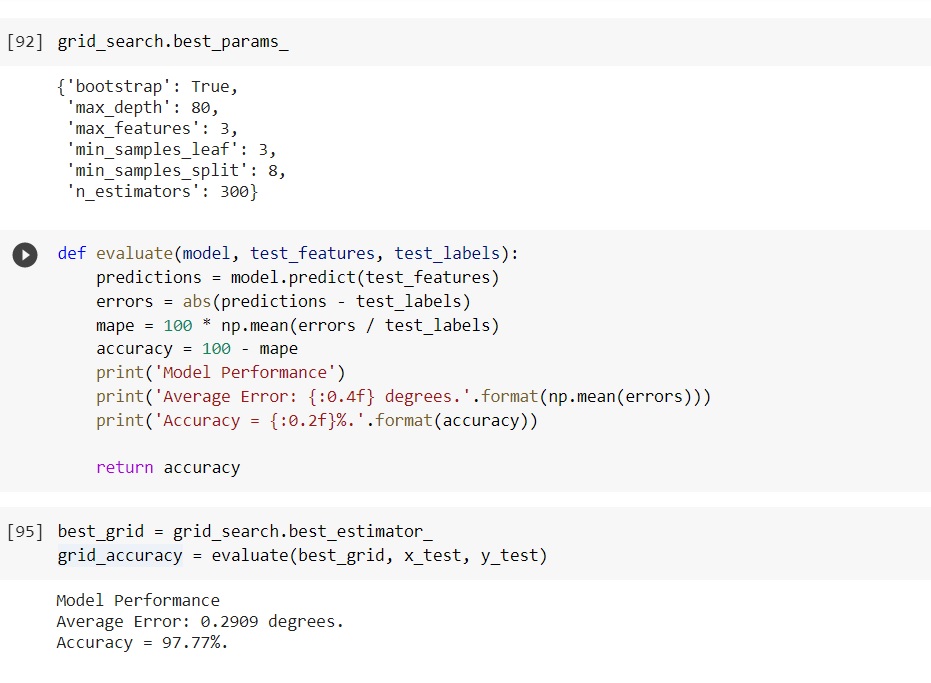
KNeighbor Regressor

ExtraTreeRegressor







* Key Metrics for success in solving problem under consideration

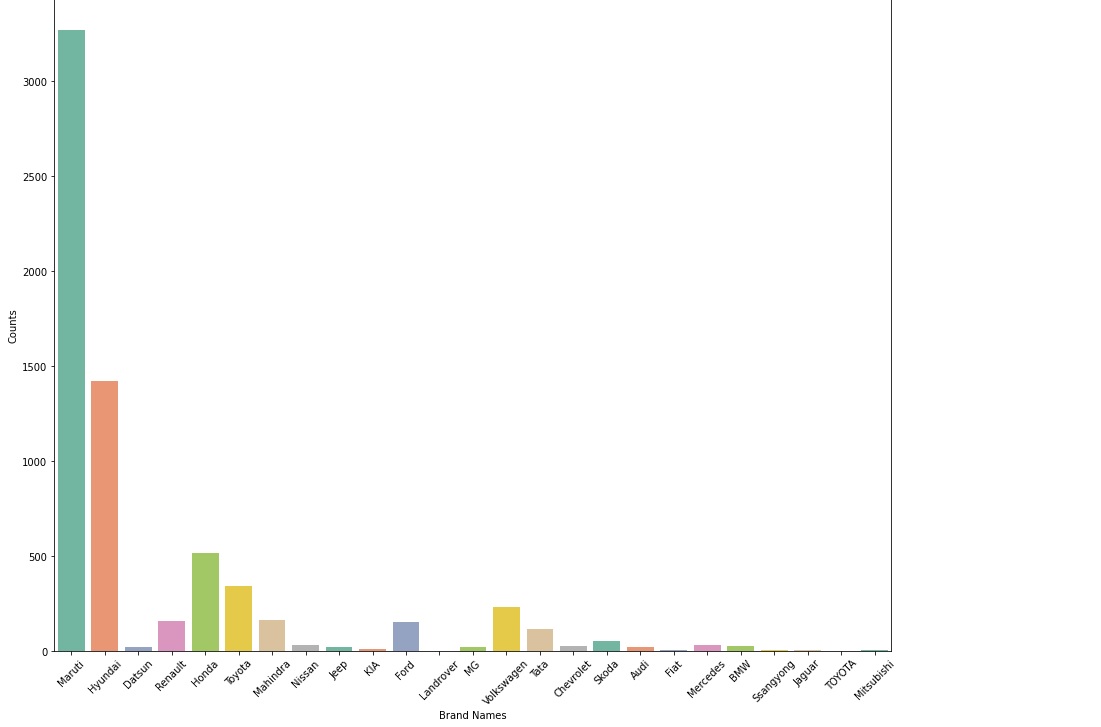
The Metrics used are:

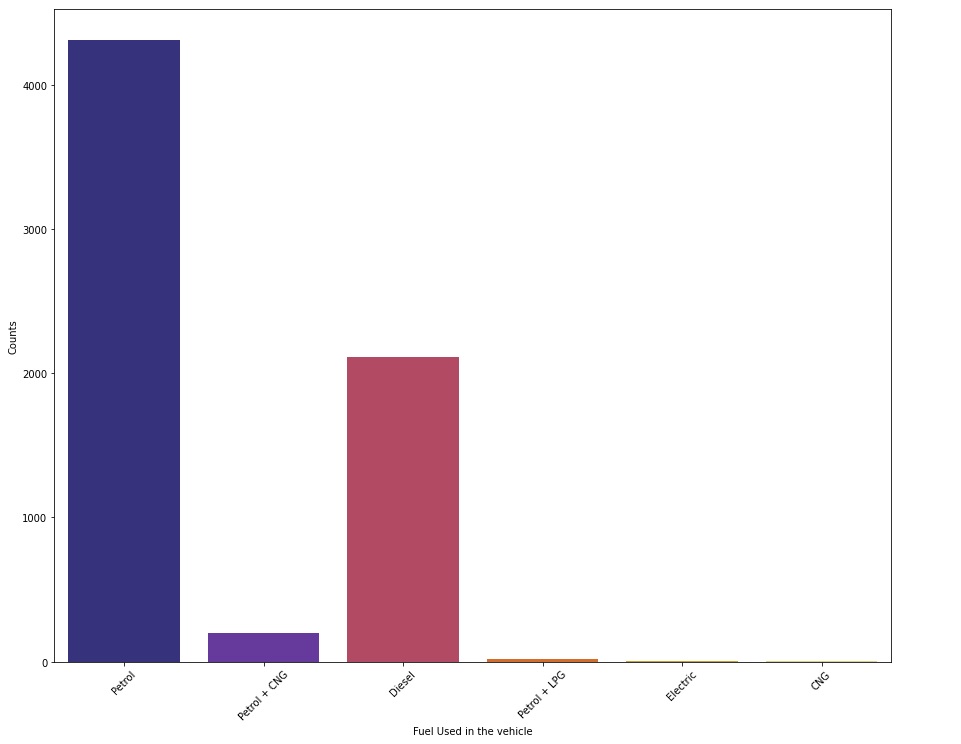
R2 Score

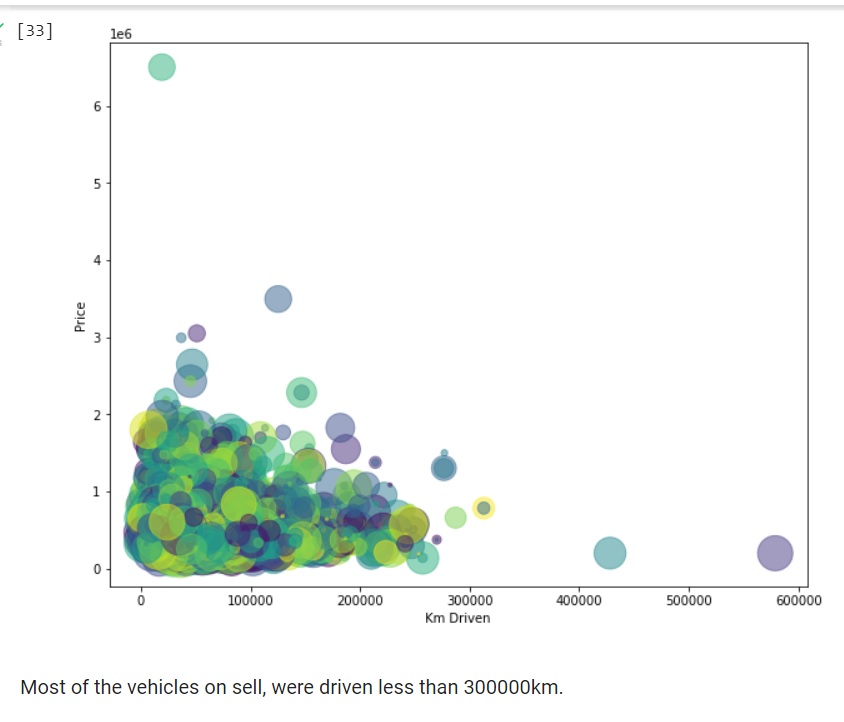
Mean Squared Error

Mean Absolute Error

* Visualizations







* Interpretation of the Results

1.Maruti cars are more for sell compare to other brands.

2. Most of the vehicle for sell, runs on petrol.

3. Most of the vehicles on sell, were driven less than 300000km.

**CONCLUSION**

* Key Findings and Conclusions of the Study

Most of the used cars on the market for resale are from Maruti and its models. The top 3 models for resale are Swift, Alto and WagonR. Most of the vehicles runs on petrol than any other fuel. The vehicles are mostly driven below 400000km.

* Learning Outcomes of the Study in respect of Data Science

Machine learning algorithms are very helpful in solving real world problems. If properly used they can predict lot of things which can be useful for our daily life. The visualization of the data are used to understand the insight of the dataset. We can create assumptions and conclusion with different types of visualizations.

Hence, we can conclude that, machine learning algorithms are a useful tool to create models which help us to overcome real world challenges and give an idea to make future planning according to the situations.

Thank You