

CAR PRICE PREDICTION

Submitted by:

MITESH VERMA

**ACKNOWLEDGMENT**

I would like to acknowledge Mr. Shwetank Mishra (FlipRobo) for his vital cooperation and help in ensuring the successful completion of my assignment. He deserves the utmost credit for the assignment’s outcome.

Finally, I would want to convey my sincere thanks Datatrained Academy and their guidance without them, the task would not have been accomplished.

The website that I referred are:

<https://learning.datatrained.com>

<https://www.w3schools.com>

<https://www.freecodecamp.org>

<https://github.com>

<https://www.geeksforgeeks.org>

https://www.carwale.com

https://www.cardekho.com

https://www.cartrade.com

https://stackoverflow.com

https://www.kaggle.com

**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 car price valuation model.

* Conceptual Background of the Domain Problem

This project is about predicting the price of used cars in India, using the data of some websites. There are two phases in this project:

* + Data Collection Phase.
  + Model Building Phase.
* Review of Literature

1. First phase is Data Scraping using Selenium.

Scraped data from:

* + www.cardekho.com
  + www.carwale.com
  + www.cartrade.com

Features:

* + Brand (Brand of the car)
  + Name (Name of the car)
  + Year (Year of manufactured)
  + Fuel (Fuel used in the car)
  + Driven(km) (How much the car is driven in km)
  + City (In which city is car available to sell)

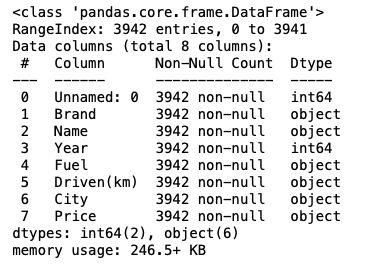
Target:

* + Price (Price of the used car)
* Motivation for the Problem Undertaken

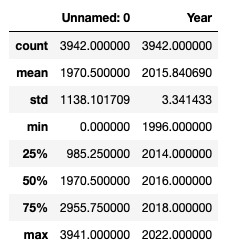
This project is on the data scraping, data science and machine learning model, build the model to predict the used car price based on some features.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem
  + **Information of the dataset:**



* + **Description of the dataset:**



* Data Sources and their formats
  + Data Collection Phase.
    1. Collected the data from different websites such as carwale.com, cardekho.com & cartrade.com.
    2. Collected data like name, year, km, fuel, city etc.
    3. Saved the dataset as a csv file.
    4. Data cleaning from excel and through python.
  + Model Building Phase.
    1. Data Cleaning.
    2. EDA
    3. Data Pre-processing
    4. Model Building
    5. Model Evalution
    6. Selecting the best model
    7. Hyperparameter tuning
* Data Preprocessing Done
* **EDA**
* **Description**
* **No null present**
* **Data cleaning**
* **Visualization**
* **Encoding**
* Hardware and Software Requirements and Tools Used

Anaconda-navigator

jupyter notebook

matplotlib-inline==0.1.6

numpy==1.23.2

packaging==21.3

pickleshare==0.7.5

platformdirs==2.5.2

prompt-toolkit==3.0.30

pyparsing==3.0.9

python-dateutil==2.8.2

scikit-learn==1.1.2

scipy==1.9.0

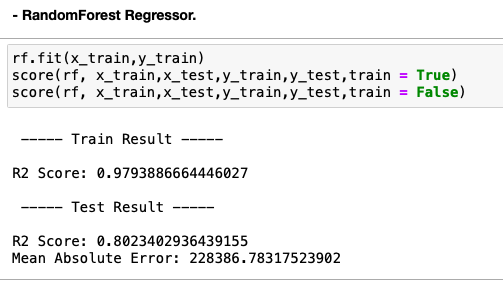
sklearn==0.05

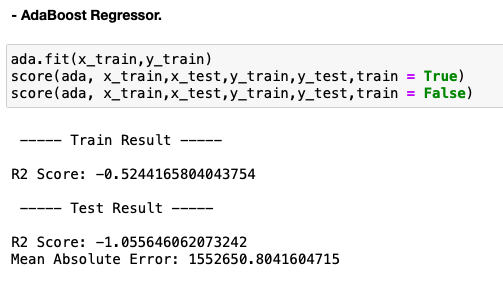
**Model/s Development and Evaluation**

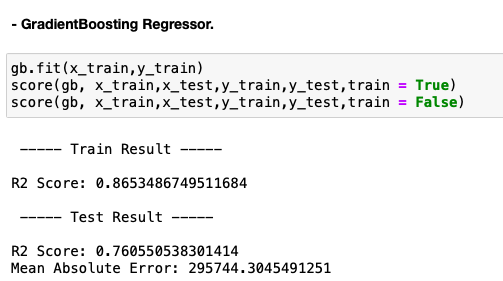
* Identification of possible problem-solving approaches (methods)
* **EDA**
* **Description**
* **No null present**
* **Data cleaning**
* **Visualization**
* **Encoding**
* **Model Building**
* **Select the best model**
* **Hyperparameter tuning**
* Testing of Identified Approaches (Algorithms)

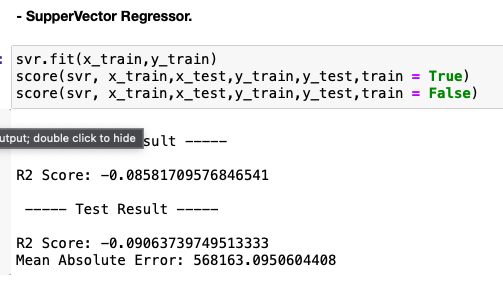
Algorithms used for the training and testing:

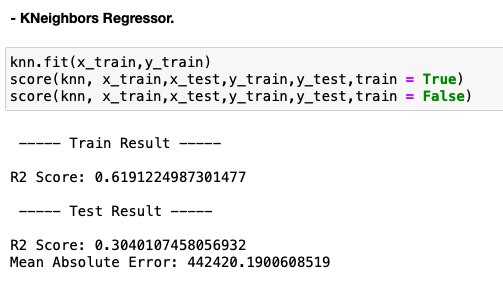
* + **RandomForest Regressor.**
  + AdaBoost Regressor.
  + GradientBoosting Regressor.
  + Super Vector Regressor.
  + Kneighbors Regressor.
* Run and Evaluate selected models









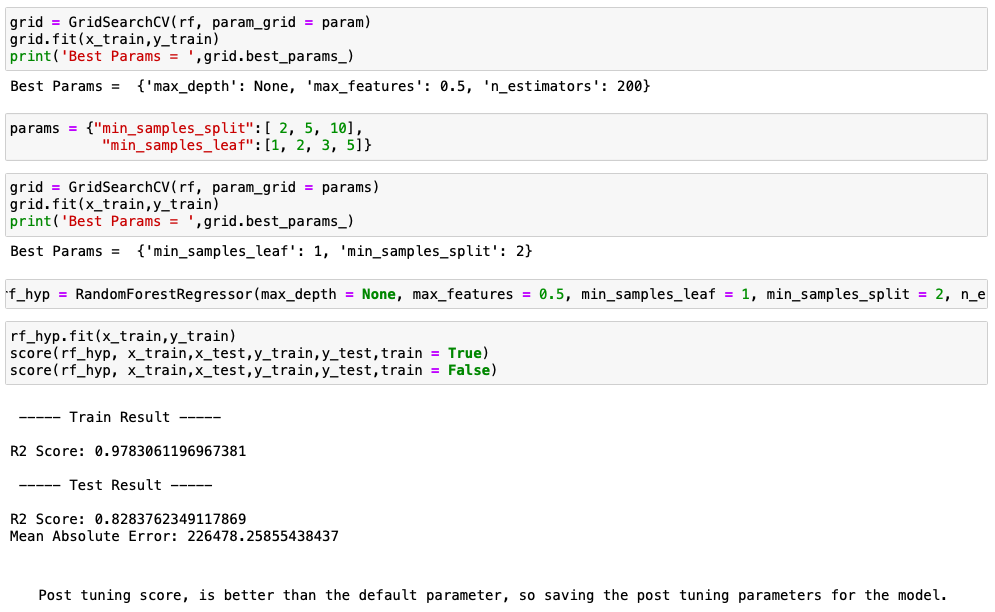


* Interpretation of the Results

RandomForest Regressor is giving the best score.

**CONCLUSION**

* Key Findings and Conclusions of the Study



Post Tuning is giving the better score & found the best parameter for RandomForest Regressor.