BRACT'S

Vishwakarma Institute of Information Technology Kondhwa (BK), Pune - 48

Department Of Computer Engineering Year 2021-2022



Data Science Mini Project On

**Car Price Prediction**

|  |  |  |
| --- | --- | --- |
| **Name** | **Roll Number** | **GR Number** |
| **Abdul Kadar Undia** | **321001** | **21910031** |
| **Sahil Agrawal** | **321003** | **21910562** |
| **Sourav Choudhari** | **321013** | **21910386** |
| **Yash Dandane** | **321014** | **21911106** |

**Under Guidance of Prof. Shweta Tiwaskar**

Project Problem Statement: Car Price Prediction

Dataset Used: Vehicle Dataset From cardekho

### Introduction of the problem

Vehicle price prediction especially when the vehicle is used and not coming direct from the factory, is both a critical and important task.

With increase in demand for used cars more and more vehicle buyers are finding alternatives of buying new cars.

There is a need of accurate price prediction mechanism for the used cars. Prediction techniques of machine learning can be helpful in this regard. It is common to lease a car in many countries rather than buying a new car.

#### Explanation of the dataset

This dataset contains information about used cars listed

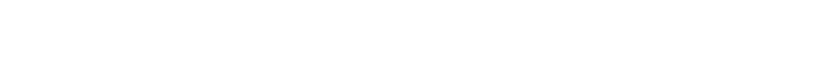
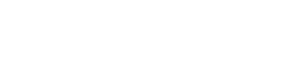
on [www.cardekho.com](http://www.cardekho.com/)

This data can be used for a lot of purposes such as price prediction to

Exemplify the use of linear regression in Machine Learning.

Source of the Dataset: Kaggle Number of the Attribute: 9

#### Name of the Attribute



*1. Car\_Name:* This column should be filled with the name of the car.

*2. Year:* This column should be filled with the year in which the car was bought.

*3.*

*Selling\_Price:*

the car at.

This column should be filled with the price the owner wants to sell

*4. Present\_Price:* This is the current ex-showroom price of the car.

*5. Kms\_Driven:* This is the distance completed by the car in km.

*6. Fuel\_Type:* Fuel type of the car i.e. Diesel or Petrol.

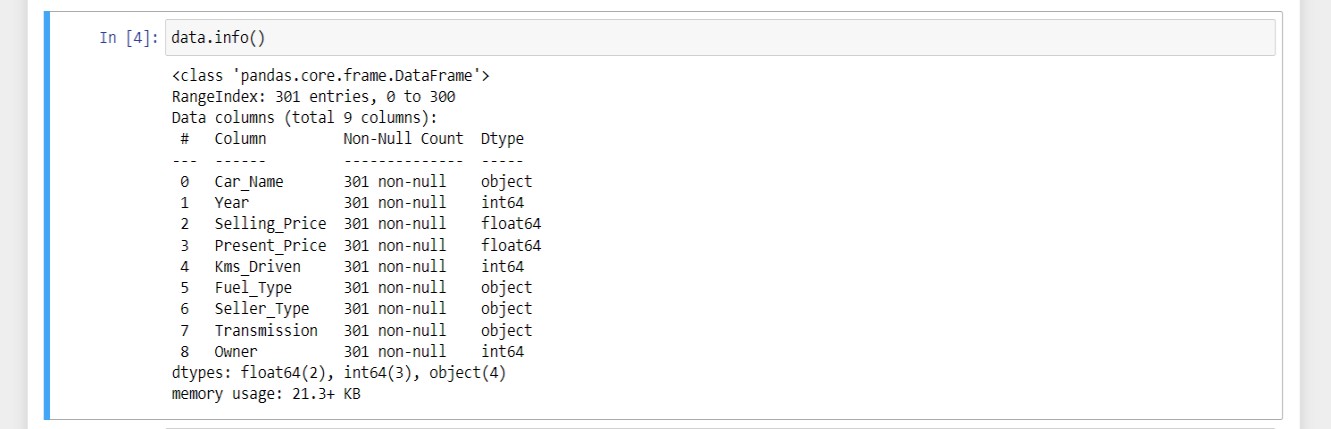
*7. Seller\_Type:* Defines whether the seller is a dealer or an individual.

*8. Transmission:* Defines whether the car is manual or automatic.

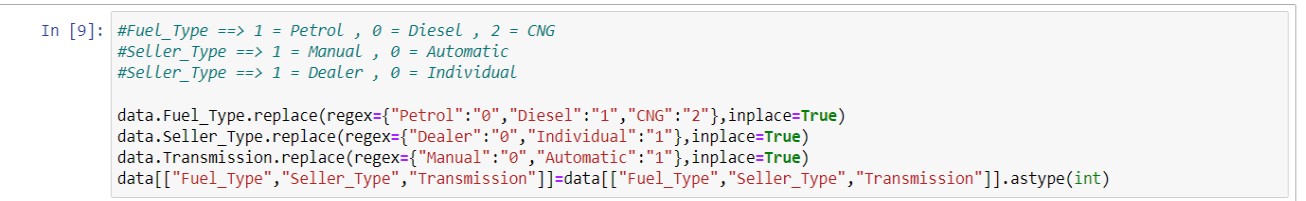
Type of the Attribute:

1. *Car\_Name : Discrete*
2. *Year : Discrete*
3. *Selling\_Price : Continuous*
4. *Present\_Price : Continuous*
5. *Kms\_Driven : Continuous*
6. *Fuel\_Type : Discrete*
7. *Seller\_Type : Discrete*
8. *Transmission : Discrete*
9. *Owner : Discrete*

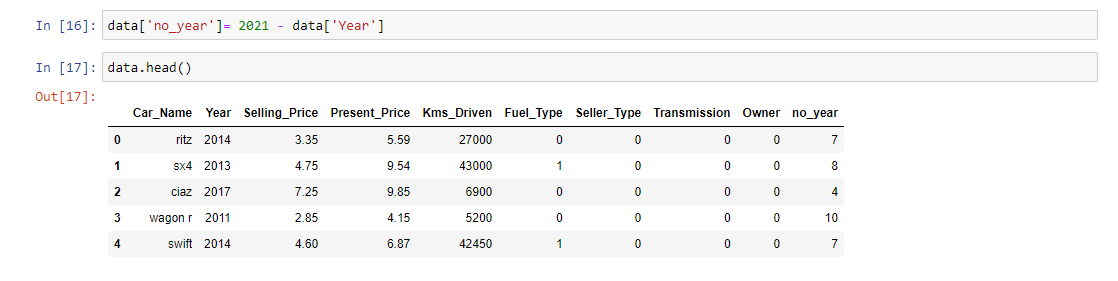
### Experimental Setup



Changing object values to numerical values to make it valid for regression models.

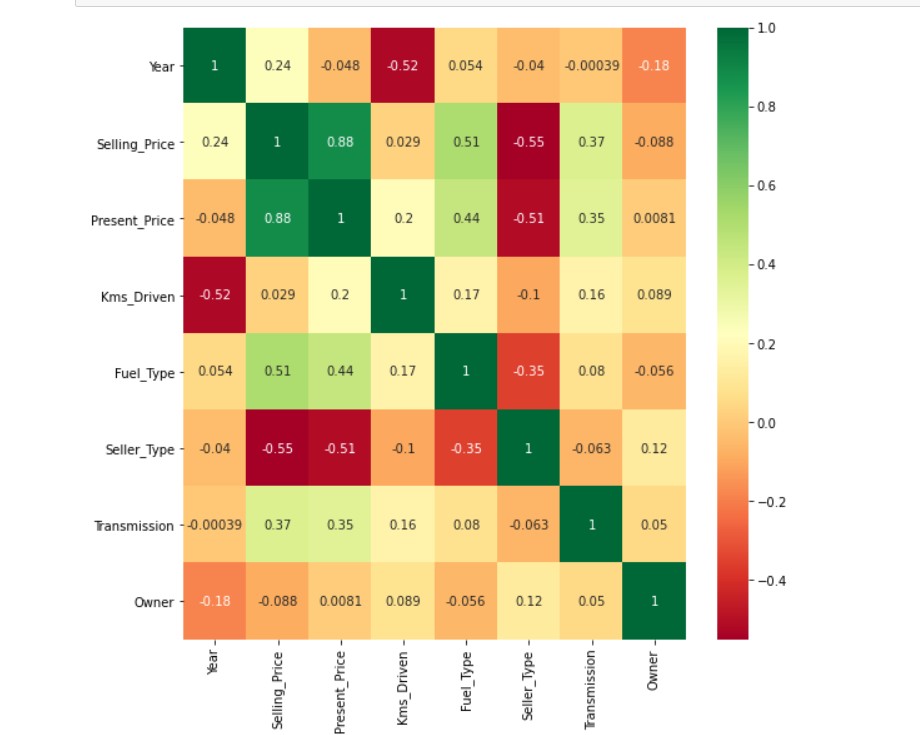


Adding a new column to the dataset for better Model building.



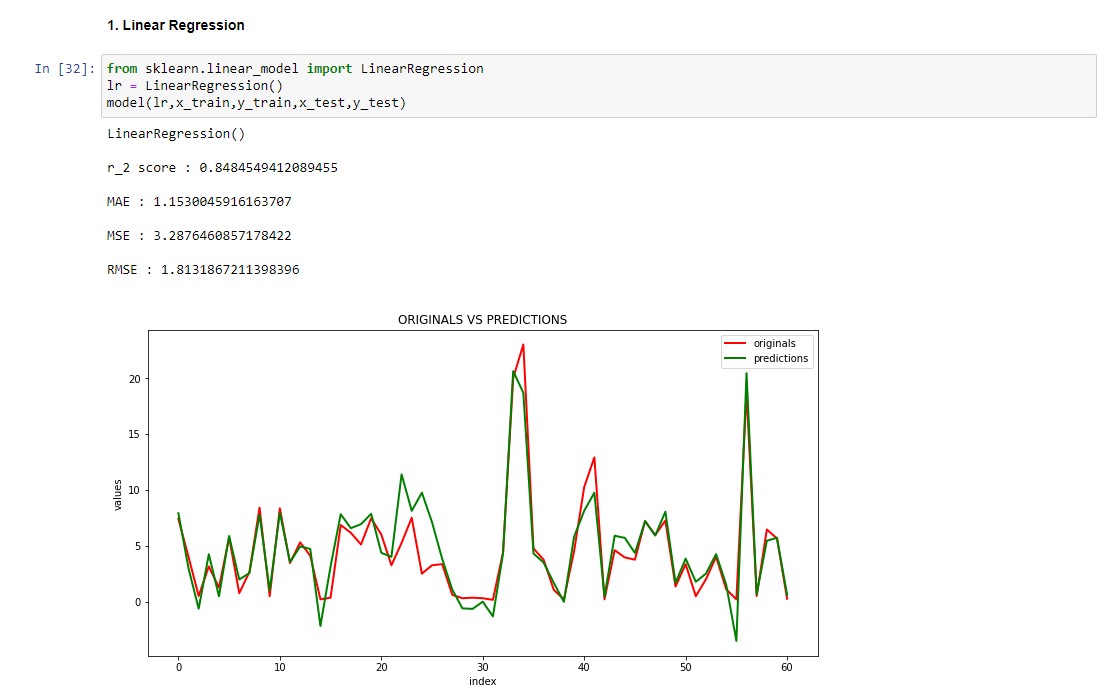
## Results

Correlation heat map

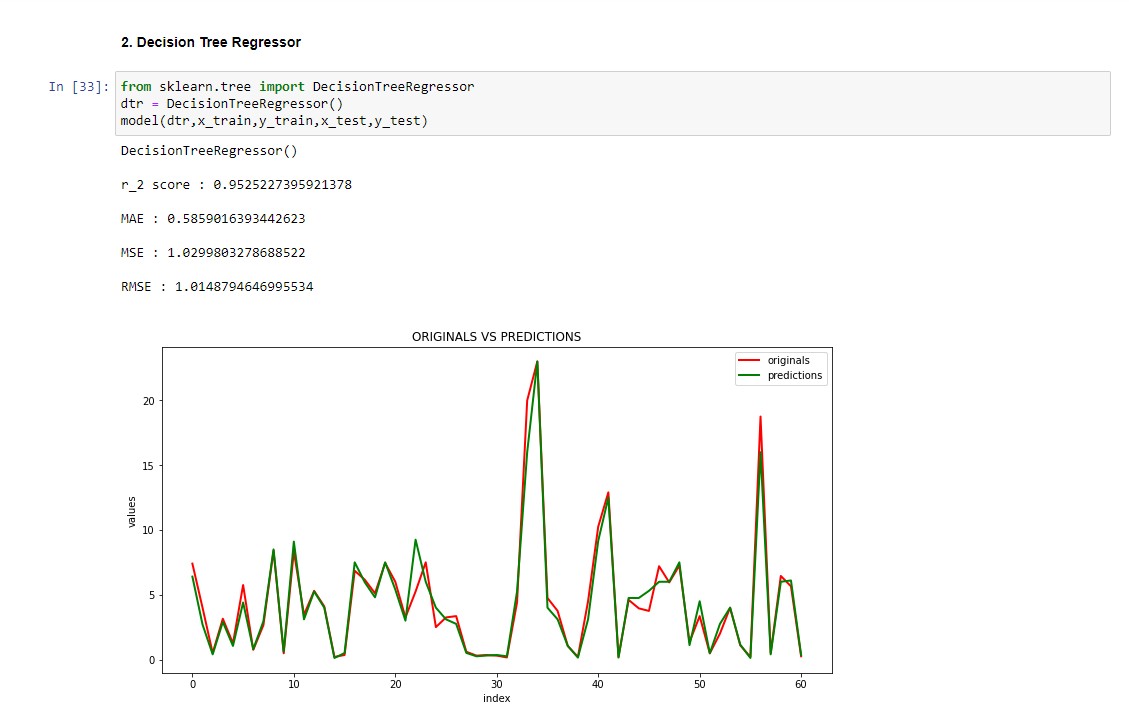


Present price of a car directly influences Selling Price prediction. Both are highly correlated and here directly proportional to each other.

## Linear Regression Model



Decision Tree Classification Model



Review

Car Price Prediction using Machine Learning Techniques (Enis Gegic, Becir Isakovic, Zerina Masetic)

International Berch University, Bosnia

* Abstract – A car price prediction has been a high interest research area, as it requires noticeable effort

and knowledge of the field expert. Considerable number of distinct attributes are examined for the reliable and accurate prediction. To build a model for

predicting the price of used cars .Respective performances of different algorithms were then compared to find one that best suits the available data set. The final prediction model

was integrated into Java application. Furthermore, the model was evaluated using test data and the accuracy of 87.38% was obtained. Predicting price of a used cars has been studied extensively in various researches.

* Richardson in his thesis work His theory was that car producers produce more durable cars. Richardson applied

multiple regression analysis and demonstrated that

hybrid cars retain their value for longer time than traditional cars.

This has roots in environmental concerns about the climate and it gives higher fuel efficiency

* In the related work shown a, authors

proposed prediction model based on the single machine learning algorithm. However, it is noticeable that single machine learning algorithm approach did not give remarkable prediction results and could be enhanced by assembling various machine learning methods in an ensemble.

* Car price prediction can be a challenging task due to the high number of attributes that should be considered for the accurate prediction. The major step in the prediction process is collection and preprocessing of the data. In this research, PHP scripts were built to normalize, standardize and clean data to avoid unnecessary noise for machine learning algorithms.
* Data cleaning is one of the processes that increases prediction performance, yet insufficient for the cases of complex data sets as the one in this research. Applying single machine algorithm on the data set accuracy was less than 50%. Therefore, the ensemble of multiple machine learning algorithms has been proposed and this combination of ML methods gains accuracy of 92.38%.
* This is significant improvement compared to single machine learning method approach. However, the drawback of the proposed system is that it consumes much more computational resources than single machine learning algorithm
* Although, this system has achieved astonishing performance in car price prediction problem our aim for the future research is to test this system to work successfully with various data sets.We will extend our test data with eBay and OLX used cars data sets and validate the proposed approach.

# Conclusion

* + As we can see in the correlation heat map, most correlating factor is Present\_Price(0.88), Seller\_Type(-0.55),

no\_year(-0.24).

* R2 higher the score, better the model. Therefore, decision tree regressor is better model than linear

regression.

* Mean absolute error - less the difference better the model. Therefore, decision tree regressor is better.
* Mean squared error - less the value better the model. Therefore, decision tree regressor is better.
* Root mean squared error - - less the value better the model.

Therefore, decision tree regressor is better.

* Hence, Decision tree Regressor model is good fit for prediction.

## Future Aspects

Keeping the current model as a baseline, we intend to use some advanced techniques like fuzzy logic and genetic

algorithms to predict car prices as our future work. We intend to develop a fully automatic, interactive system that contains a repository of used-cars with their prices. This enables a user to know the price of a similar car using are recommendation engine, which we would work in the future.

# Reference

#### <https://www.datarobot.com/wiki/prediction/>

* [https://machinelearningmastery.com/linear- regression-for-machine-learning/](https://machinelearningmastery.com/linear-regression-for-machine-learning/)
* [https://www.scribbr.com/frequently-asked- questions/error-in-a-linear-regression-model/](https://www.scribbr.com/frequently-asked-questions/error-in-a-linear-regression-model/)