

CAR PRICE PREDICTION PROJECT

Submitted by:

HARSHIL GORADIA

**ACKNOWLEDGMENT**

The data of cars that are for sale was extracted from ‘OLX.COM’. The data extracted from the website contains car data from various cities, different number of owners and few other features that will be useful for predicting car price.

**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.

* Motivation for the Problem Undertaken

The motive for taking down this project is to know the car price demand for used cars and use machine learning models to make predictions of the new trend

**Analytical Problem Framing**

* Data Sources and their formats

I extracted the data from olx.com using web scraping techniques and converted to into dataframe for further process. I scraped brand, fuel type, number of owners, transmission, location, variant, kilometer travelled and price of the car.

* Data Preprocessing Done

For cleaning the data, null values were replaced by mode value of each column assuming that replacing them won’t affect the model.

* Hardware and Software Requirements and Tools Used

Jupyter Notebook - The whole project was done in jupyter notebook which provides python environment.

Selenium – Selenium was used for web scarping to extra data from olx.com. This data was later used for analysis.

Pandas and Numpy - These libraries which are most important for an predictive analysis project. These libraries provide a function of shaping the data, importing the data and viewing the data.

Matplotlib and Seaborn – These libraries were used for visualising the data in the entire project. All the visualization done in the project were done using either one of these libraries.

Sklearn – This library was the most used library as compared to other libraries. This library has been used in pre-processing for imputing numerical null values and for scaling the dataset. Later it was widely used for importing algorithms for machine learning and predicting.

**Model/s Development and Evaluation**

* Testing of Identified Approaches (Algorithms)

1. LinearRegression
2. RandomForestRegressor
3. DecisionTreeRegressor
4. SVR
5. GradientBoostingRegressor
6. AdaBoostRegressor

* Run and Evaluate selected models

Graphical user interface, text, application, email

Description automatically generated

A computer screen capture

Description automatically generated with medium confidence

Graphical user interface, text

Description automatically generated

Graphical user interface, text

Description automatically generated

Graphical user interface, text

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text

Description automatically generated

* Visualizations

Chart, bar chart

Description automatically generated Chart

Description automatically generated

Chart, histogram

Description automatically generated

Chart, bar chart

Description automatically generated A screenshot of a computer

Description automatically generated with medium confidence

* Interpretation of the Results

After visualizing I found that there is no co-relation between different variables. Post modelling, I interpreted that RandomForestRegressor was the best performing model.

**CONCLUSION**

* Learning Outcomes of the Study in respect of Data Science

Visualization helped to conclude that most of the cars that are to be sold are of Maruti’s. Most the car sellers are 1st owners, and the cars are mostly diesel cars or petrol cars. The best model for predicting the price was RandomForestRegressor with parameters (max\_depth = 100, max\_features = 3, n\_estimators = 100).

* Limitations of this work and Scope for Future Work

The data used for this project is collected from few states only out of 29 states of India. So, the model built can only help predict prices from these states, this model cannot predict prices for all over India.