Car Price Prediction

Submitted By

Anita Tarangi

**INTRODUCTION**

* Business Problem Framing

To find out the car prices of used cars in USA.

* Conceptual Background of the Domain Problem

Machine Learning helps buyers and sellers in predicting the value of used cars

* Review of Literature

The dataset had no NULL values but was imbalanced and treated , carried out different methods to train the model.

**Analytical Problem Framing**

* Identification of possible problem-solving approaches Linear Regression, Decision Tree, Random Regressor, XGB Regressor, Ridge, Lasso.
* Data Sources and their formats
* There were more than 10% null values in the dataset.
* The dataset is imbalanced.
* For some features, there may be values which might not be realistic. You may have to observe them and treat them with a suitable explanation.
* You might come across outliers in some features which you need to handle as per your understanding. Keep in mind that data is expensive, and we cannot lose more than 7-8% of the data.
* Data Preprocessing Done
* **Checking the size of the dataset**
* **Checking the summary statistics of the dataset**
* **Checking the data types and null values**
* **Dropping columns which aren’t important**
* **Converting categorical columns using dummy method.**
* Hardware and Software Requirements and Tools Used
* Python, Pandas, Seaborn, Matplotlib, sklearn, Model Selection, Linear Regression, Decision Tree, Random Regressor, XGB Regressor, Ridge, Lasso.

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)
* Linear Regression, Decision Tree, Random Regressor, XGB Regressor, Ridge, Lasso.
* Run and evaluate selected models
* Linear Regression, Decision Tree, Random Regressor, XGB Regressor, Ridge, Lasso.
* Visualizations

Seaborn, Matplotlib

* Interpretation of the Results
* Carried out Analysis as well as visualization of the Dataset, treated imbalanced dataand then, trained the model using Linear Regression, Decision Tree, Random Regressor, XGB Regressor, Ridge, Lasso, model predicted with different accuracy for every model, out of all the models, XGB has achieved with 83% accuracy rate.

**CONCLUSION**

* Data exploration, cleaning, visualization is the basic steps, in which imbalanced data was identified and treated. And then, trained the model using Linear Regression, Decision Tree, Random Regressor, XGB Regressor, Ridge, Lasso, model predicted with different accuracy for every model, out of all the models, Random Forest has achieved with 83% accuracy rate.