Car Price Prediction Project

Submitted by:
TARUN SAHU

ACKNOWLEDGMENT

Firstly, I would like to thank FlipRobo Technologies for giving me the opportunity to work on this project. Also, I would like to thank the DataTrained team, especially Deepika Sharma Ma'am for providing me the knowledge and guidance which helped me alot to work on this project.

INTRODUCTION

• Business Problem Framing

In this project, we have to make car price valuation model using new machine learning models from new data. Because with the change in market due to covid 19 impact, our client is facing problems with their previous car price valuation machine learning models.

• Conceptual Background of the Domain Problem

Firstly, we will prepare our own dataset using web scraping.

After that we will check whether the project is a regression type or a classification type.

We will also check whether our dataset is balanced or imbalanced. If it is an imbalanced one, we will apply sampling techniques to balance the dataset.

Then we will do model building and check its accuracy.

Our main motto is to build a model with good accuracy and for that we will also go for hyperparameter tuning.

• Review of Literature

I am summarizing my research done on the topic.

• I have created my own dataset using web scraping and imported important libraries for my project.

• I have created the dataframe.

• I have analysed my data by checking its shape, number of columns, presence of null values if any and checking the datatypes.

Then I have done some data cleaning steps,e.g Checking the value counts of the target variable, dropping some irrelevant columns from the dataset, checking correlation between the dependant and independent variables using heatmap, visualizing data using distribution plots, detecting and removing skewness in my data if any, outliers detection using boxplots and removing them, balancing dataset using Random overs sampler method, splitting the data into independent and dependant variables and finally scaling the data.

Then I have used 5 regressor models, out of which AdaBoost Regressor is giving a good accuracy score of 99% after hyperparameter tuning.

Analytical Problem Framing

- Mathematical/ Analytical Modeling of the Problem

If you look at data science, we are actually using mathematical models to model (and hopefully through the model to explain some of the things that we have seen) business circumstances, environment etc and through these model, we can get more insights such as the outcomes of our decision undertaken, what should we do next or how shall we do it to improve the odds. So mathematical models are important, selecting the right one to answer the business question can bring tremendous value to the organization.

Here I am using AdaBoost Regressor with accuracy 99% after hyper parameter tuning

- ## Data Sources and their formats

  Data Source: The read csv function of the pandas library is used to read the content of a CSV file into the python environment as a pandas DataFrame. The function can read the files from the OS by using proper path to the file.

  Data description: Pandas describe() is used to view some basic statistical details like percentile, mean, std etc. of a data frame or a series of numeric values.

- ## Data Preprocessing Done

  - I have checked for null values and there are some null values present. I have removed it using Simple Imputer method.
  - I have label encoded the object type columns in the dataset.