FLIP ROBO

Car Price Prediction

Submitted by:

Sagnik Das

ACKNOWLEDGMENT

INTRODUCTION

- ## Business Problem Framing

  This assignment helps us in understanding the variables, which affect the price of used cars.

- ## Conceptual Background of the Domain Problem

  ### This problem has 4 steps:-

  a. Data Scrapping
  b. EDA Analysis, on train and test dataset.
  c. Hyperparameter optimization tuning, on train dataset.
  d. Predicting the value on test dataset

- ## Review of Literature

  Factors that have an effect in used car market.

- ## Motivation for the Problem Undertaken

  To find out the factors on which the value of a used car increase or decrease. Also, to find out the factors improve their marketing strategies of the websites that features used cars.

Analytical Problem Framing

- Mathematical/ Analytical Modeling of the Problem

  Linear Regresion

$$Y = a + bX$$

- Data Sources and their formats

  Data Provided by Flip Robo Technologies. Files provided in Microsoft Excel Comma Separated Values File (.csv) format. Websites (Olx, cardekho, Cars24)

- Data Preprocessing Done
  - Web Scrapping
  - Remove duplicate or irrelevant observations.
  - Fix structural errors
  - Filter unwanted outliers
  - Handle missing data (from both train and test dataset)
  - Changed strings into integers
  - Hyperparameter optimization using grid search cv
  - Validate

- Data Inputs- Logic- Output Relationships

  Input – Data from web scrapping

  Output – Predictions on the factors that have a positive or negative effect in the price of old cars.

- State the set of assumptions (if any) related to the problem under consideration

  Changed fuel type:-

Petrol – 1
Diesel – 2
CNG – 3

  Transmission :-

Manual – 1
Automatic – 2
Hybrid - 3

- Hardware and Software Requirements and Tools Used

Hardware – Laptop

Software –

import pandas as pd, import os, import csv, import sklearn, import numpy as np, import matplotlib.pyplot as plt, import seaborn as sns, %matplotlib inline, Hyperparameter optimization using grid search cv.

Model/s Development and Evaluation

- ## Identification of possible problem-solving approaches (methods)
  Extensive data cleaning, EDA Method, Linear Regression and prediction, Hyperparameter optimization using grid search cv.

- ## Testing of Identified Approaches (Algorithms)
  EDA Method, Hyperparameter optimization using grid search cv.

- ## Run and Evaluate selected models
  - EDA Algorithm, Linear Regression, prediction, Hyperparameter optimization using grid search cv.

- ## Key Metrics for success in solving problem under consideration
  What were the key metrics used along with justification for using it? You may also include statistical metrics used if any.

- ## Visualizations
  Contour, Heat map, pair plot, box plot, histogram, gridsearchcv.

- ## Interpretation of the Results

  This assignment gives us an overview of the factors on which the price of used cars vary.

CONCLUSION

- ## Key Findings and Conclusions of the Study

  Factors on which the price depends of a used car.

- ## Learning Outcomes of the Study in respect of Data Science

  Data cleaning helped in cleaning excess nan data and EDA algorithm helped it in putting it in visualised matter. Hyperparameter optimization using grid search cv helped to create a model and predict.

- ## Limitations of this work and Scope for Future Work

  More data would give better understanding of the factors that effect in the positivity or negativity of price of used cars.