
PREDICTING CAR PRICES

Abstract

The purchase of a car can take a considerable amount of time. Deciding whether a car is worth the posted price when you see listings online can be difficult. It is possible to determine the actual worth of a car based on many factors, including the mileage, make, model, and year. The aim is to develop models based on our data to predict car prices using machine learning algorithms. We would like to know.

Design

Web scraping using Selenium was used to obtain the data. Exploratory data analysis was carried out next. Following the exploration of the data and finding the most relevant features that affect the price, we can use machine learning to predict car prices using selected features, thus allowing companies to offer reasonable car prices depending on the given feature.

Data

There are 205 observations and 26 features, 10 of which are categorical. There are several features that are categorical, such as the CarName, fuel type, and aspiration. Dummy variables were used to convert the non-numeric data to binary.

Algorithms

- 1.Outliers cleaning using z-score
- 2.Fixing miss spelled brands names
- 3.Took the log for price
- 4.Converting categorical features to binary dummy variables
- 5.Combining particular dummies and ranges of numeric features to highlight strong relationship with price
- 6.Apply scaler() to all the columns except the 'dummy' variables
- 7.Dividing into X and Y sets for the model building

Tools

- Jupyter Notebook will be used to create and document live code and visualization
- Pandas for data manipulation
- Matplotlib and Seaborn for plotting