أكاديمية سدايا
SDAIA Academy

Predicting car price using linear regression

Overview:

An automobile company called Cars.com sells new and used cars and we are focusing in cars that have been used as customers aspire to understand the factors affecting the prices of each used car based on some factors/variables provided, we are going to build a linear regression model that predicts the prices of used cars based on these variables, using the data from the website by web scrapping them with a comprehensive analysis with all data cleaning, exploration, visualization, feature selection, model building, and evaluation.

Dataset Description:

Using Cars.com data that have 7 features:

FEATURE	DESCRIPTION
DESCRIPTION	Column contains the car name and the model of the car
DEALAR	Name of the automobile dealer.
MILES _DRIVEN	The kilometers driven for the used cars.
RATE	Rate of the dealer out of 5
NO_OF_REVIEWS	Number of reviews for each cumulative rate
PRICE	Price of the car.

Feature Engineering Columns:

FEATURES	DESCRIPTION
MODEL	Model year of each car.
CAR_NAME	The name of the car

Tools:

- Pandas and NumPy packages to manipulate data.
- Matplotlib library for visualizing data.
- BeautifulSoup library for pulling data out of the Cars.com website.
- LinearRegression from the sklearn.linear_model class of the sklearn module, to perform the linear regression then predicted the price.
- mean_squared_error from the sklearn.metrics module to measures the average squared difference between the estimated values and actual value.
- mean_absolute_error from the sklearn.metrics module to measure the accuracy of the model.
- Jupyter notebook to execute the code.

Conclusion:

Predicting the price of used cars is both an important and interesting problem. In this project, we want to use linear regression to predict the price of used cars using various features that will help determine the expected price for the car and sell it.