Used cars

Abstract

This project provides the best possible price of a used car by taking attributes such as Manufactured Year, Number of Cylinders, Condition, and many more as inputs. Out of all models tested

Design

This project is one of the T5-Batch 3 Data Science BootCamp requirements. Data provided by Kaggle "used cars " A web application built with Flask which scrapes used vehicle listings weekly from all regional Craigslist sites in the United States and allows users to filter them by criteria such as (but not limited to) city, price, manufacturer, and odometer.

Data

The dataset is provided in .csv format. It contains 426880, each flight has 26 features. The most relevant feature to this project is the Price . The date has 26 features, as instance:**(id , url , region , region_url , price , year , manufacturer , model , condition , cylinders, fuel, odometer, title_status , transmission , VIN , drive , size, type , paint_color, image_url, description, county, state, lat, long, posting_date)**

Algorithm

-Split data into (train,validation,test)

-EDA train part

-Visualization train data

-Ecode train part by hot encode(categories value)

-Clean validation part and encode the categories value

-bulid linear registration model and R2 by validation (the score in negative)

-try to bulid polynomial model

TOOLS

- Technologies: Python using Jupyter Notebook
- Libraries : Pandas , Linear Regression , Ridge ,Lasso ,Standard Scaler , Polynomial Features ,Matplotlib and Seaborn

Communication