

Predicting the Price of Laptops in Amazon

Abdullah Al-Huwaishel

Abdalla Al-Nujaidy

Abstract

The goal of this project was to use Regression models to predict the Price of Laptops in Amazon to help People to have an idea How the features can change the price. We worked with data scrapped from Amazon by us, after refining a model, we predict a price of our own dataset to visualize and communicate my results using regplot.

Data

The dataset contains around 1759 with 12 features for each, 5 of which are categorical. A few feature highlights include display size and Processor, Ram, Hard Drive, Battery life, Price, Weight, Processor Cores, Brand, Color, Processor Brand. Some of the features were dropped and other grouped into more general categories, and an in-depth analysis of the rest of them was undertaken to inform baseline models and feature engineering.

The data was scrapped from Amazon.com because it is the most famous website and provide many different brands. The focus was targeted on price.

The last predictions of laptops prices were drawn from multiple features such as brand, processor, RAM, Hard drive.

Algorithms

Pre-processing

- 1- Cleaned the data and impute the null, checked for duplicate, and drop outliers.
- 2- Created plots to visualize the relationships between the target and predictors, and between the variables themselves.
- 3- Decided which features to keep and which to drop.

Feature Engineering

- 1- The categorical data were turned to numerical using dummy variables.
- 2- Combined the dummies to reduce the number of columns.
- 3- Turned price columns to price numerical.
- 4- Split data to train and test data.

Models

- 1- Several linear regressions were developed, and we reduced the number of features in each one based on the coefficient of each model.
- 2- and choose the best model with the right number of features to avoid the complexity and error.

Tools

- Numpy and Pandas for data manipulation
- Scikit-learn for modeling
- Matplotlib and Seaborn for plotting
- Tableau for interactive visualizations