Machine Learning Project -

Estimating 2nd hand Laptop Price

Business Objective

A C2B firm is in the process of creating an Online platform where the consumers can sell their used Laptops. This platform should host an automated mechanism that can suggest to a consumer the realistic price of their used laptop when the required details (different features of the laptop) are provided.

Problem Statement

You have been assigned the task to build for this C2B firm a Linear Regression Machine Learning model which will give as a prediction the used laptop's price. *For more details on the intermediate and final outputs expected, refer to the list of deliverables mentioned in the "Model Building" and "Model Validation" sections below.*

Data Description

The dataset provided for this activity consists of 12 features where 11 are independent features and 1 is a target variable. The features in the dataset are described below :

Variable Name	Description
Company	Laptop Manufacturer
Product	Brand and Model
TypeName	Type (Notebook, Ultrabook, Gaming, etc.)
Inches	Screen Size
Screen resolution	Screen Resolution
CPU	Central Processing Unit (CPU)
Ram	Laptop RAM
Memory	Hard Disk / SSD Memory
GPU	Graphics Processing Units (GPU)
Opsys	Operating System
Weight	Laptop Weight
Price euros	Price (Euro)

Model Building

- Show Bi-variate plots (scatter/ bar) of all meaningful variables with the dependent variable.

- Show correlation of all independent variables in a multicollinearity Matrix. Also, show the VIF values for the finally selected variables.

- Present your final model results (show what's applicable from the below list):
  - List of variables that came significant.
  - Final variables – their beta coefficients, standard error, and p-values.

- Summarize the steps followed to finalize your model - consisting of the below steps (as applicable)
  - Feature Engineering
  - Performance comparison between Train and Test
  - Use of Cross-validation

- While developing the model, you would have gotten a few candidate models which were not as good as the final model (in terms of performance, multicollinearity, statistical stability etc.). Show a few of these candidate models and explain their shortcomings

- Show what kind of feature engineering did you apply in your project and why (include in your results what's applicable from below)
  - Dummy variables
  - Label encoding
  - Any bin-based variable created -what was the significance/rationale of binning
  - Any new derived variables created using the raw variables – E.g., Ratio based, difference based, % difference based / Rate of change, etc.

- Demonstrate Live how your model scores a new data point?

- Provide your understanding of the next steps that the client/ end-user needs to follow to deploy your model at their end. Think about the below lines:
  - Any technical/infrastructure requirements that the client needs to meet?
  - What files do you need to provide them?
  - What kind of data cleaning and preprocessing would the client need to do before using the model?
  - How will the client use your model on new data?
  - How will the client know that the model is performing well on new data points?

Model Validation

- Show your model's performance on the below metrics (on both train and test samples).
  - ANOVA results
  - Model's R-squared and Adjusted R-squared values
  - Performance comparison on Train and Test Samples
  - Show the values for the following metrics. Pick anyone (RMSE/MSE/MAE/MAPE)
  - Show the histogram of residual deviance of the final model on the test data