Predicting Latest Laptop Pricing Using Regression Methods: A Project Proposal into Predictive Regression

Kyle Brinker

University of Central Florida

STA 6714

Dr. Hsin-Hsiung Huang

January 31, 2024

Predicting Latest Laptop Pricing Using Regression Methods

In a world where the internet proves a necessity, the need for laptops has significantly increased. The price for these laptops depends on a variety of factors including specifications, brand name, and operating systems; however, these qualities may follow a pattern that will help create a model to predict the latest prices of laptops. The project seeks to determine if this pattern exists, and which model will predict the prices the best.
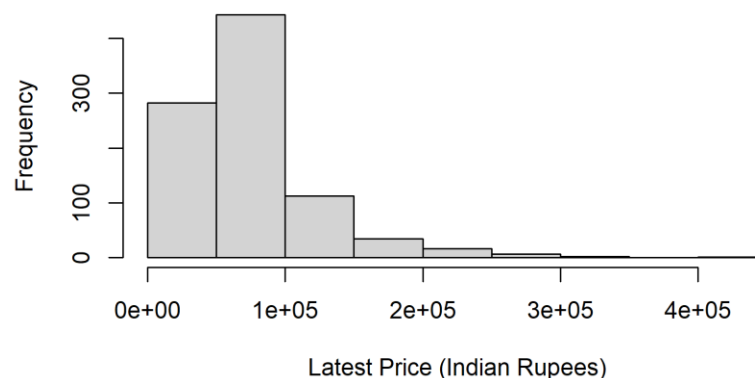
Variables and Target Problem

The target problem involves predicting the prices (in Indian Rupees) of various laptops based on specific attributes. A couple of key attributes to look at includes the make and model of the processor. Higher-end processors tend to affect the price of a computer, but examining the correlation between the price and processor may yield results on how one may affect the other for predictions. Some other key attributes that may influence laptop pricing involve Random Access Memory (RAM) and graphic memory. Some users need more RAM and graphical capabilities for gaming, data science, and engineering. Including these in the model might reveal insight into how these components influence the price. Some other features that will get included in the model includes screen size, touchscreen capability, storage options (hard drive or solid-state drive) as well as storage size, review rating, original price, the discount amount, and if Microsoft Office was installed.

Describing the Data

Looking at the known laptop prices and how it is distributed. Figure 1 shows the histogram of the latest laptop prices measured in Indian Rupees. This distribution has a skewed nature. This suggests that many of the population's needs regarding laptops may not require higher-end hardware found in more powerful machines such as gaming systems and workstations.

Histogram of Latest Laptop Prices

Further examination yields insights into the data and how the variables relate to each other. A heat map in Figure 2 of the dataset suggests that some of the independent variables may have collinearity considering the darker regions mixed in the plot; however, with so many predictor variables in the dataset will require more extensive examination to further determine. Collinearity would make sense in some cases as processor brand may be synonymous with a specific model.
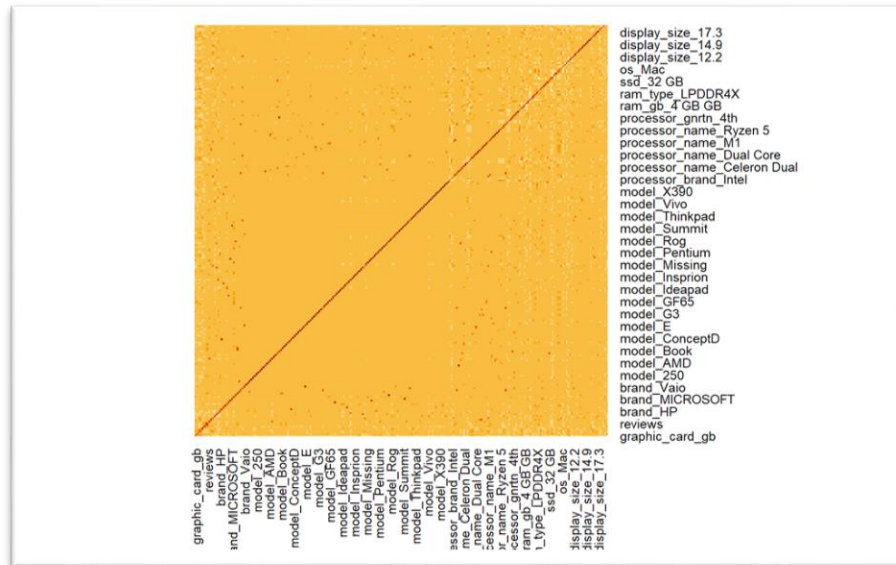
Figure 2: Heatmap for Laptop Data

Based on the results of the heatmap, a scatterplot shows the relationship between the variables with a high correlation. Note that the categorical data was transformed into dummy variables prior to the creation of the chart

Proposed Statistical Methods

Analyzing the dataset will involve the use of regression and random forest models to help choose the most accurate model. Regression models will determine if there exists a series of coefficients that can interpret the price of a laptop. The accuracy for these models will use testing and holdout data to create prediction prices before then determining the mean-square error. This same process will look at the accuracy of the random forest models as well. The model will then get re-evaluated using feature selection techniques and view the accuracy using the reduced model.

Application and Conclusion

Many aspects influence the prices of laptops; however, the project looks to predict the price of a laptop based on its specifications. These models would help influence consumer pricing for companies to produce competitive pricing.

References

Kumar, S. (2022, April 3). *Laptop specs and latest price*. Kaggle. https://www.kaggle.com/datasets/kuchhbhi/latest-laptop-price-list?rvi=1

Machine Learning for Business Analytics: Concepts, techniques, and applications in R, Second edition. (2023). . John Wiley and Sons Ltd.