PSTAT131Datamemo

2022-04-07

```
data <- read.csv('C:/Users/arthu/Dropbox/My PC (DESKTOP-9BV8I37)/Documents/PSTAT131Final/Cleaned_Laptop
```

—Information on Data—

My data set includes many of the specs of Laptops coming from some of the most popular and least brands. The data set compares many factors such as brand name processor, ram, ssd, os, and many others to look at the total price of the laptops. This data I have obtained from Kaggle where the user got it from flipkart.com. (https://www.kaggle.com/datasets/kuchhbhi/latest-laptop-price-list?resource=download&select=Laptop.csv) There are around 23 predictor variables used in this dataset with a total of 897 observations. I will mainly be looking at the processor, star rating, ram, and os to be able to compare and contrast the highest rated computers with the best specs on them, to be able to analyze what are some of the best computers with spec to price ratios as well as be able to predict a sample price based on the variables stated above. There are no missing data but there is a good amount of extra data that will be coded to become dummy variables.

I am interested in predicted the expected price of the laptop based on a few of the spec variables, excluding some of the categorical variables. I am interested in answering the question of which brand of laptops provides basically the best "bang for the buck", meaning which specs are the best for the price. In addition within the project I will define what is meant by "better specs" than others. These questions will best be answered using a classification approach. Again, the predictors that would be the most useful are the rating of the computers, as well as the other predictor variables that deal with the performance of the laptop, and to throw in Brand, to be able to compare and contrast which brand provides the best "bang for the buck". The goal of this model is inferential with a little bit of predictve as we will look at ratings that it was given to assume that future brands of the laptops should be able to perform at around the same capacity, but mainly will be looking at the facts and data to predict the outcome, as well as supposed price for a laptop with certain type of performance variables. I already have the data set loaded and am planning to work on this project on a weekly basis, starting with the third week of the quarter. Where I will begin using the dataset and simplying it to only the predictor models needed and start writing the more basic outline of specificng the variables and which are "good" vs " bad". My goal is too finish this project at the end of week 9. I believe the hardest aspects for me would be the coding part and being able to specify specific parts of the dataset to focus on and combine them with others. As well as looking which would be the best methods to analyze the results, which plots and analysis should be used.