James Schwantes

Professor Locklair

CSC 419

12/10/21

Machine Learning Project Final Submission

For my machine learning project, I decided to use a dataset of cars that were registered under the state of Washington, and I wanted to see if I could find the Top 10 Best cars (with the cars’ Year, Make, and Model) that would be considered as a clean alternative. In other words, I wanted to find the top 10 best electric vehicles in the dataset. This goal was eventually modified to: Find the best electric range of an electric vehicle based on the Maximum Curb weight of a vehicle. Primarily because I was not able to find an appropriate algorithm, that was able to identify specific vehicles that would be used to accomplish my goal.

The process for creating this model, was that I choose two key features that I could use for my machine algorithm, which was the Maximum Curb Weight and the Electric Range of a vehicle, and I used a Simple Linear Regression algorithm to train and model the data. In addition to this modeling, I also did some research into the weight of each vehicle, in order to have some data to model, as this was not a category in the actual data set. After doing this, I modeled the data, and quickly realized that the score/accuracy off by a lot, and it did not model the data very well.

After I found out about this set back, I started to look at other algorithms to see if I could find one that would best fit the data. One of the algorithms that I did try, but did not have the best results, was the Plot Individual and Voting Regression Predictions Algorithm. I tried some other algorithms, but this algorithm was the most notable. Eventually, I choose the Plotting Cross-Validated Predictions algorithm, and this algorithm seemed to fit with the data well. Originally, I also had the Maximum Curb Weight of each vehicle in tons, but to have some more variation, I decided to have the Weight in lbs. If I had to redo the algorithm for this project, I probably would have done the Principal Component Regression vs Partial Least Squares Regression Algorithm or a similar algorithm, because I think it would be better suited for the data.