James Schwantes

Professor Locklair

CSC 419

9/23/21

Machine Learning Project Proposal: Clean Alternative Car Ranking.

For my Machine Learning Project, I will be determining the top 10 best clean alternative cars for a healthier environment. As I have mentioned in my topic idea, my plan is to narrow down my data set to the top 40 cars, out of a range of roughly 80,000 or less, depending on how big the range is. Once I have this range, my new plan is to narrow down the cars to the top 10, and then the top car in my model. If this gets two complicated, I can scale down the data set to meet one of my goals instead of three.

My outputs for this project would be the Electric Range, along with any information pertaining to the car, this would include the: Make, Model, and Year that the car was made. If this information pertaining to the Make, Model, and Year of the car becomes overwhelming, I may be able to combine columns to simply my model. The ideal dataset would contain all of these features listed above with one extra feature; this feature would be a Clean Alternative Category. As far as variables, I would want it to contain a numeric Electric Range, and a numeric year.

This Project as a whole would need a supervised learning algorithm, and I would potentially like to use one of the following Algorithms: Nearest Neighbors Algorithm (specifically the NeighborhoodComponentAnalysis, along with the KNeighborsClassifier), Cross detection, Feature Selection, and maybe a Neural network model, but I mean lean away from this algorithm. For sure, I would use the first 3 algorithms I listed for this project.