

## CptS 315 Project Proposal

Derek Wright

3/21/2022

The data mining task I will pursue for my project is to use used car sales data to provide an estimated value of any given car based on the current listing of cars. The dataset I will use is from <https://www.kaggle.com/datasets/austinreese/craigslist-carstrucks-data>, which is a collection of used car listings on Craigslist that is collected every few months. The main task will be to determine the best way to compare the cars to find the most similar ones. Other things I will explore with this dataset will be how different aspects of a car, like condition, year, make, area, date of listing, number of exact matched cars listed at the same time, etc. affect the values of each car and how those could affect the recommended sales price. Most entries in the dataset do not contain complete data, so I will have to fill in the blanks most likely using item-item collaborative filtering, or possibly some other method if I find one to be more suitable. My reason for choosing this task is because my favorite hobby is working on cars. I spend a lot of my free time and money working on various repairs and projects on the three cars that I currently own. Looking at cars for sale is something I enjoy because its interesting to see the different values of not only the cars I own, but also other cars I like. It would be nice to have a better understanding of what affects those values and what other information can possibly be extracted from this. The final product of this will basically be a simpler version of tools like Kelly Blue Book, and as of now I plan to use results from Kelly Blue Book to measure the success of my project.