Capstone Project Proposal (Car Evaluation)

Car Evaluation

Derived from simple hierarchical decision model, this database may be useful for testing constructive induction and structure discovery methods.

http://archive.ics.uci.edu/ml/datasets/Car+Evaluation

- What is the problem you want to solve?
    - I would like to be able to predict the value of a used car using its main characteristics as features .
    - *I would like to understand what metrics are involved in creating a valuation for a used car, and if any those metrics stand out in particular. I would also to observe if the same factors that drastically change the value in used cars, also affects new cars.*
- Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?
    - *My client in this situation would be a used car dealership, or any one who sells used cars or is looking to buy a second-hand car(Ability to find good prices, and avoid overpriced cars). Clients could also include the consumers of the used cars, as well as even new car manufacturers as they may be keen on seeing if those trends could correlate to new cars*
- What data are you using? How will you acquire the data?
    - *The data I will be using is included in the link above. If I am able to find data on new cars as well, then I may try to analyze that data set as well and compare the two.*
    - *The model evaluates car according to the following concept structure:*
        - *CAR car acceptability*
          . PRICE overall price
          . . buying buying price
          . . maint price of the maintenance
          . TECH technical characteristics
          . . COMFORT comfort

. . . doors *number of doors*
. . . persons *capacity in terms of persons to carry*
. . . lug_boot *the size of luggage boot*
. . safety *estimated safety of the car*

- *1728 Instances and 6 attributes*

- Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.

    - To begin this problem I will perform some exploratory analysis using some data visualization. I will attempt to solve this problem using a simple regression model, as well as various data visualization tools such as seaborn and matplotlib.

- What are your deliverables? Typically, this includes code, a paper, or a slide deck.

    - *My deliverables will definitely include code via Jupyter Notebook, as well a paper and/or slide deck*
    - *Ordered by priority:*
        - *Code: Jupyter notebook, with detailed description of the steps used to manipulate the data, goals on what I'm trying to achieve, a description of the data visualisation, cleaning, training and evaluation of the model*
        - *Slides: high-level description (few or no technical detail) of the problem, the most important plots to convey a message and a polished story telling.*
        - *Paper: detailed technical description of the data analysis and model training process*