|  |
| --- |
| Toronto Metropolitan university |
| References |
|  |
| **Kyle Cheng** |
| **500627248** |
| **5/15/2023** |

A problem car manufacturers must face is the ability to balance desirability and profits. It is easy to assume that the more features a car has, the higher the likelihood a consumer will find that vehicle appealing. As the cost to produce increases with more features, determining what is a worthwhile feature is an important task to consider in the balancing of cost versus add-ons. For example, we know that Safety in a vehicle is important to a driver and has been proven that a potential buyer is willing to pay more money for a safer vehicle (Koppel et al., 2008). Determining what other automobile features can create a similar effect is a worthwhile venture for car manufacturers.

The theme of this task will explore classification and regression. These themes are both machine learning techniques that can be used to predict categorical or continuous target variables. Classification and regression can be used in a wide range of applications and can utilize the power of decision trees and logistic regressions, linear regressions and polynomial regressions.

The problem we wish to solve is to determine what feature in a car is most important for consumers and creates the biggest pull in deciding on a car. Below are samples of several questions to consider:

* “How much is a customer willing to pay based on its given features?”
* “Can we predict the price range of a car based on its features?”
* “How does the maintenance cost of a car relate to its overall rating?”
* “What size of car has the highest level of desirability?”
* “How much trunk space do consumers want?”
* “What is the most important factor in determining a car's acceptability rating?”

By answering these statements, we will aim to explore the relationship between various features of an automobile and the acceptability of that vehicle. By creating a predictive model, we can isolate the most important factor in determining a car’s desirability and draw conclusions from that.

In order to answer the questions posed above, the dataset to be analyzed will be the Car Evaluation dataset available in the UCI Machine Learning Repository: <http://archive.ics.uci.edu/ml/datasets/Car+Evaluation>. Entries have six attributes that consist of the buying price, maintenance price, number of doors, seating capacity, size of the trunk and safety rating. These attributes are then classified in an acceptability level that is low, medium or high. With over 1700 instances, the dataset provides a robust and diverse set of data that can be used to analyze the factors that contribute to a car's acceptability level.

Some techniques and tools we are proposing to solve the stated problem include a correlation analysis, predictive models such as decision trees or the Naïve Bayes test and principal component analysis.

References

Dua, D. and Graff, C. (2019). UCI Machine Learning Repository. http://archive.ics.uci.edu/ml.

Koppel, S., Charlton, J. L., Fildes, B. N., & Fitzharris, M. (2008). How important is vehicle safety in the new

vehicle purchase process? PubMed. https://www.researchgate.net/publication/5393358\_How\_

important\_is\_vehicle\_safety\_in\_the\_new\_vehicle\_purchase\_process