Jamie Weiss

Dominic Holt

Brian Kilburn

Predicting the Price of a Vehicle based on Vehicle Features

Introduction:

With a constant stream of new automobiles being rolled out onto the market each year, a potential buyer may have trouble finding the best price range for what type of vehicle they want. This project aims to use a large dataset of vehicles, including their make, model, miles per gallon (MPG), price, and many other bits of information relative to a model to train the machine in determining the price of a vehicle based on several features presented. If successful, we hope that the model could be used to allow customers to input features of a vehicle and determine a relative price range.

Study Objectives

The current goals for this project are to train a machine using a dataset of vehicles to determine the price of said vehicle, or a set of features of a vehicle. This would ideally be done by setting price ranges of $5000 and having the vehicles grouped into these. From there, key features, such as MPG, body style, and vehicle brand would be examined to find correlation between them and cost. After testing the machine with the data provided, another test could be done by manually inputting features of an automobile and seeing if the machine could determine what price the automobile in question is. This test would occur after using a training set or K-fold Cross Validation. Additionally, if more data could be found, this would also aid in training the machine.

Project Activities

The activities involved within this project are: develop features that would allow the machine to categorize the vehicles by price range and examine the vehicle features provided in the dataset within each price range, train the program using these features using a training set or 5/10-fold Cross Validation, and develop a way for a user to input their own variables so that the machine may output a price range corresponding to the closest match. If possible and time permits, we also propose to implement image recognition within our program. Users can utilize the program via uploading an image of a vehicle of choice and the program will implement artificial neural networking to determine the year make and model of the vehicle. Once determined the system can use the data available that has been cross validated and tested with the initial implementation to determine the best marketable price for the consumer.

Benefits

The beneficial aspect of our project is to make the decision-making process for the consumer a lot easier than what current technology provides when purchasing a car. Currently, a consumer must do a lot of research about the vehicle market to know for certain if they are getting a good price, also the information that is found via research will likely be based upon suggestion and personal preference. Our software will remove the task of research involved for the consumer and the human factor when it comes to suggesting outcomes for a purchase. The system will pull the most accurate information possible based on facts calculated from the data and deliver to the consumer the best possible pricing.

Related Works

<https://www.ripublication.com/irph/ijict_spl/ijictv4n7spl_17.pdf>

<https://github.com/dkeske/CarPrice>

<https://towardsdatascience.com/build-develop-and-deploy-a-machine-learning-model-to-predict-cars-price-using-gradient-boosting-2d4d78fddf09>

References

Data set: <https://www.kaggle.com/ljanjughazyan/cars1>