

**Module 2: Building the care of the future**

**ALY 6020**

Name: Kavitha Palthayvath

Instructor’s name: Justin Grosz

Due Date: 07/25/2021

**Abstract**

This report summarizes the methods and the process we have followed to improve the sales posture of a car manufacturing company. To make this happen we are focusing on making fuel efficiency the parameter for increase in sales.

**Introduction**

Fuel efficiency attributed for the best car, and our goal is to design the perfect car, to do so we are going to find the variables that can help us define what makes the most fuel-efficient car. We are trying to find a linear equation of the form y=m1x1 + m2x2 + m3x3 + m4x4 + m5x5 + m6x6 + m7x7 + c. Where y is mpg and x1, x2, …. x7 is ('Cylinders', 'Displacement', 'Horsepower', 'Weight', 'Acceleration', 'Model Year', 'US Made').

**Analysis**

**Data Cleaning**

While going over all variables in the dataset we found outliers in Horsepower of the car, in some of the records (6 in number) the horsepower of the car was not known hence we got rid of these entries as there were only 6 entries out of 398. Having these entries out of the way the Horsepower variable was good to be used as a variable. We also say that there are some outliers in Acceleration of cars, but we would like to consider these outliers as it could account for distribution of data.

**Designing Model**

On considering all the variables we built a model (model\_1) which will be able to predict the fuel efficiency. On following forward selection method for feature selection (Luhaniwal, 2020), we observed that Cylinders, Horsepower, and Acceleration are not so important variables to predict the higher gas mileage hence we removed these variables from the model and retrained the model (model\_2). On retraining we found some increase in the performance of the model but still found another variable that we can get rid of another variable(‘Displacement’). After removing this variable, we saw an increase in the error made by the model(model\_3). Hence model\_2 was the perfect model for predicting the fuel efficiency of the car with the following variables (Weight, Model Year, Us Made, Displacement). Out of the following variables we found that “US Made” is the most variable as it affects the prediction the most negatively and the “Model Year” is the variable which affects the model the most positively.

**Conclusion**

On thorough analysis we found the linear equation for predicting the MPG of a car,

*mpg = 0.01\*Displacement - 0.0068\*Weight + 0.835\*Model\_Year - 2.2109\*US\_Made - 20.39.* Hence, we can plug in the respective parameters and have the perfect car we want. In order to check the accuracy of the model we calculated the MSE value of the model after each model training the MSE value of each model, model\_1, model\_2, model\_3 was 11.816, 11.530, and 11.566, this is the reason why we chose the model\_2 as the best model. The most important thing to observe here is that if the car is US Made it affect the mpg negatively with a factor of 2.29 but another more important conclusion, we draw is that a newer car can increase the mpg by a factor of 0.835 which can also up the sales of the car manufacturing company.

**References**

1. Luhaniwal, V. (2020, October 24). *A comprehensive guide to Feature Selection using Wrapper methods in Python*. Retrieved from Analytics Vidya: https://www.analyticsvidhya.com/blog/2020/10/a-comprehensive-guide-to-feature-selection-using-wrapper-methods-in-python/