Mandatory Assingment AI

August, Danijel

April 4, 2024

1 GitHub Repository

GitHub Repository

2 abstract

Throughout the project, we have worked with various regression models to predict the fuel efficiency (Miles Per Gallon) of cars with different attributes. We have analyzed the numbers and used tools such as matplotlib, numpy, pandas, and sklearn. We investigated the impact of factors such as weight, horsepower, and displacement on MPG. With the analysis, we can predict the fuel efficiency and calculate the accuracy based on our findings and trends.

3 Introduction

The purpose of this paper is to predict the fuel efficiency (In miles per gallon) of cars based on the data given. Our null-hypothesis says that the various features of our car within the dataset does not have any correlation with the miles per gallon. We will attempt to disprove this by predicting the MPG by looking at the dataset and finding relations between the data and MPG. Then using our findings to provide classification for a fuel-efficient car. Understanding the relationship between a car's attributes and its fuel efficiency is crucial for designing more fuel-efficient vehicles. This project aims to dive into the relationship between the data by applying regression analysis techniques to the datasets.

4 Methods

We will use various methods to analyze the datasets, some of which is employing multiple regression analysis methods, including OLS (Ordinary Least Squares) and linear regression. These methods will enable us to predict the target variable MPG, by using other variables such as weight, horsepower etc. We have also used models to aid in logistic regression, finding out whether a car is fuel-efficient or not if our previous predictions disprove the null-hypothesis.

5 Analysis

We checked the data for trends, then used regression to see how different factors affect MPG. We measured how well our models fit the data and looked at residuals to make sure they were good. This helped us learn more about car fuel efficiency, judging whether a car was fuel-efficient based off what factors affected MPG the most.

6 Findings

Our analysis revealed that weight, horsepower, and displacement all have a significant relationship with MPG. This could be seen on the plots and also on the summary data where they had a coefficient of displacement 0.0174, horsepower -0.0103 and weight -0.0068. The regression models demonstrate

moderate to high accuracy in predicting MPG, with slight variations depending on the chosen MPG data. The regression line indicates that our predictions closely match the actual data. While some points align perfectly with the line, others are spread around it, suggesting that while our predictions aren't flawless, they are generally close enough to reliably estimate the MPG (miles per gallon) to a certain extent.(see Appendix 1) When looking at the residuals, most points are clustered around the 0 point and well spread out, with some scattered further away, it suggests that our model's predictions are generally accurate.(see Appendix 2) When looking at the Q-Q plot we can see that the data closely matches a normal distribution. The other 2 Q-Q plots are somewhat similar to this one.(see Appendix 3)

In terms of fuel-efficiency, this means that we could determine whether a car was fuel-efficient or not based on weight, horsepower and displacements due to their significant relationships. We trained a model to find out the fuel-efficiency of cars based on the MPG median-threshold and weight, displacement, and horsepower as its relevant features. Model can bee seen in effect here (See Appendix 4) When looking at the matrix, we can see that the number of False-positive and False-negative cases are not zero, but minimal to acceptable degree. This is further supported by this ROC curve (see Appendix 5) which shows the trade-off between the true- and false positive rates at different thresholds.

7 Conclusion

In summary, our research reveals how various car attributes, like weight, horsepower, and displacement have a large impact on fuel efficiency. While weight seems to have the most impact, horsepower and displacement seems to also contribute to an extent. Using regression analysis, we've gained valuable insights into what drives fuel efficiency in cars, helping inform decisions about designing vehicles and creating policies in the automotive industry. We have now also defined a binary-classification for what determines if a car is fuel-efficient or not and trained a model to recognize it.

A Appendix

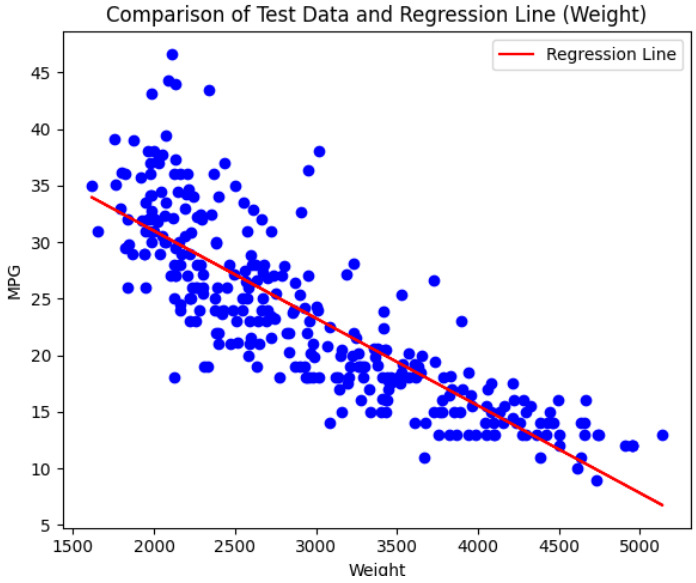

Figure 1: Regression Plot

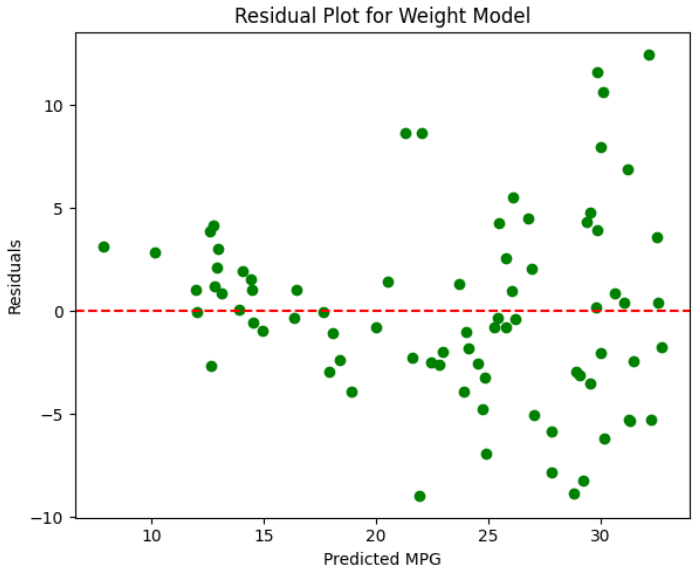

Figure 2: Residual Plot

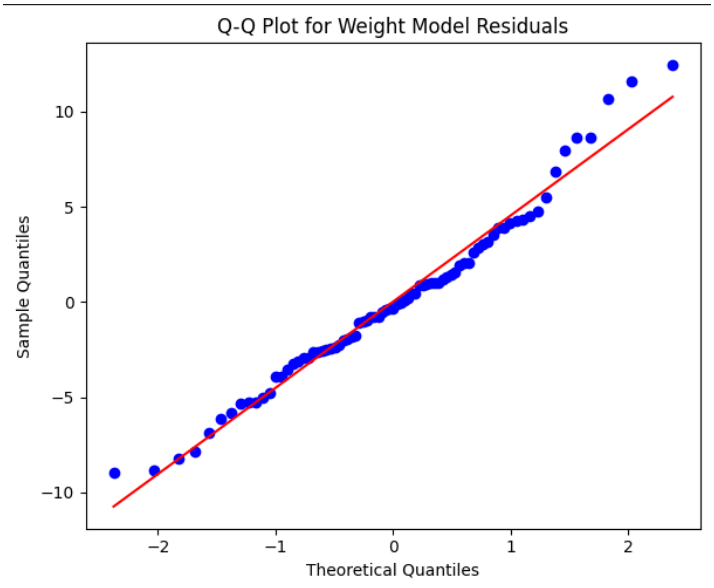

Figure 3: Q-Q Plot

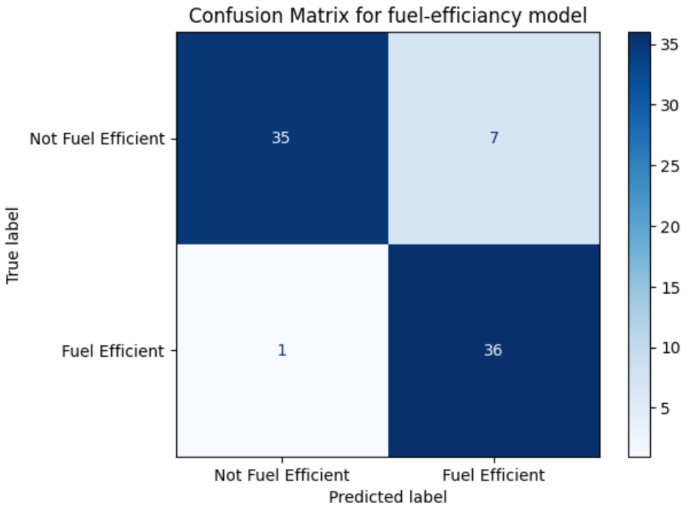

Figure 4: Confusion Matrix

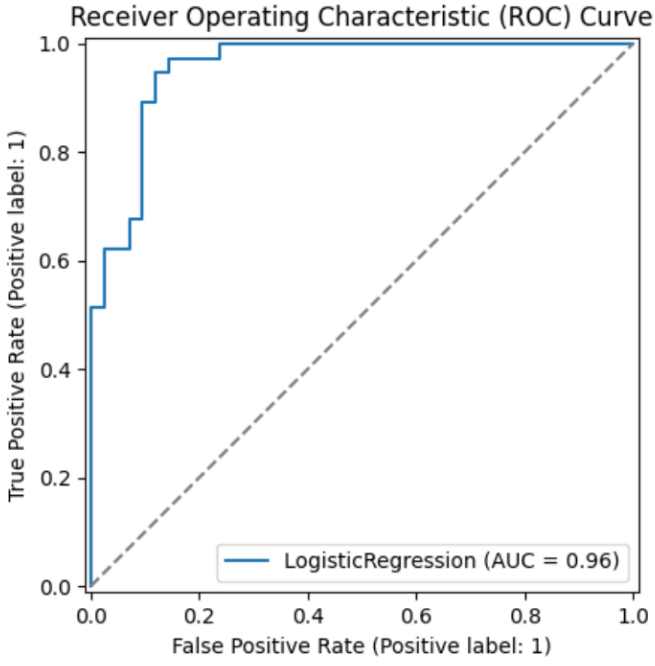

Figure 5: ROC curve