# Price Prediction of Used Cars

Group 9: Giriraj Heda (giriraj.heda@mail.utoronto.ca), Seray Pulluk (seray.pulluk@mail.utoronto.ca), Si Cheng (cs.cheng@mail.utoronto.ca)

## I. Introduction

Platinum Cars is an established used car dealership since 2004.[1] We deliver quality and unbeatable cost savings through integrity pricing.[2]

## II. Business Objectives

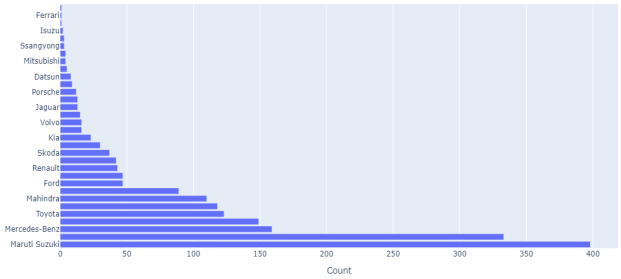


**Figure 1:** An overview of the number of cars in our inventory by their brand.

The assumption is that pricing each used car by evaluating various conditions such as damage level, engine-related components, color, etc., would be very time-consuming. We desire to provide better service by increasing work efficiency, therefore, we aim to automate the pricing process. Our objective is as follows: *Predicting a used car's worth based on the original configuration and the usage.*

We envision that by utilizing our model, we will be able to provide precise and competitive prices for our inventory.

## III. Results

Our machine learning model which uses the techniques of random forest can precisely estimate the value of used cars in the market, with a test accuracy of 80.28%. By adopting this model, we can provide the best offer that can tailor to our customer's needs and also maximize our profit. If we are given several features of a car, such as the production year, size, fuel capacity, engine capacity, and the number of previous owners our estimated selling price from the model would align with the current market price 80.28% of the time. Eventually, we can automate this pricing business process so that our employees can focus on more meaningful work, such as providing better customer service.

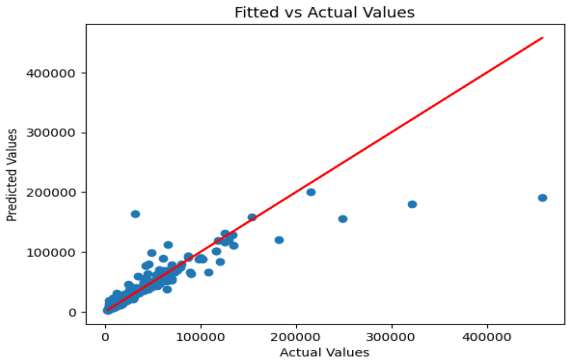


**Figure 2:** Scatterplot of the actual and fitted values for car prices

Fig. 2 displays a scatterplot of actual sale prices versus the predicted prices generated by the model. The scatterplot demonstrates a relatively good fit for prices under $200,000, as the majority of the data points lie near the linear pattern that follows the actual line. However, as we examine data points with prices above $200,000, the model struggles to accurately predict these values, as these data points are further from the actual line. This suggests that there are outliers in the dataset, and the model has failed to account for them. This is particularly evident for high-end cars like Ferraris and Rolls Royces, which are priced much higher than other cars in the dataset and are scarce. As a result, the model is unable to learn from these points and struggles to accurately predict their sale prices.

The length, width, year, and engine are the most important features for our model to predict the pricing. As such, these must be collected for any given car in the inventory.

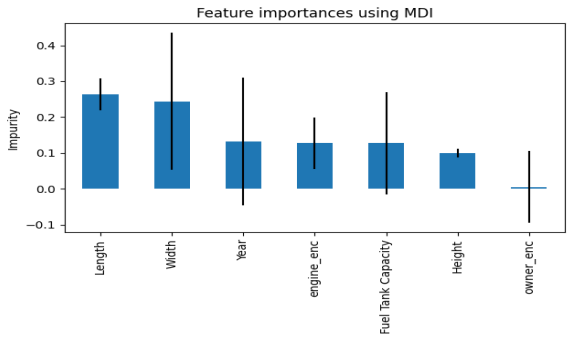


**Figure 3:** Feature importance plot for Random Forest.

[1] https://www.platinumcars.ca/usp/

[2] https://www.jpmorgan.com/insights/research/when-will-car-prices-drop