Executive Report

Predicting the Selling Price of Cars

Group 4: Dailiang Pan (dailiang.pan@mail.utoronto.ca), Junwei Shen (junwei.shen@mail.utoronto.ca), Bryan Zhu (bryan.zhu@mail.utoronto.ca), & Che Zhu (alexfe.zhu@mail.utoronto.ca)

I. Introduction

Based on automotive industry trends, car sales in Canada appear to be on the rise. New car sales alone reached 1.63 million in 2023, an 11.8% rise from 2022[1]. Despite the per capita recession in Canada, February 2024 was the 16th consecutive month of year-over-year growth[1] in the automotive market. This gives automotive dealerships a prime opportunity to take advantage of this market. Our specific goal is to optimize our dealerships in doing so though providing a predictive model for pricing.

II. Business Objectives

The selling price of a vehicle can depend on multiple factors, which may give dealerships trouble in competitively pricing their vehicles. Fig. 1 shows the distribution of selling prices of cars from our dataset[2].

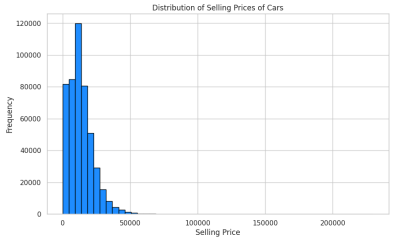

Fig. 1. The distribution of final selling prices of cars. The y-axis indicates the frequency of sales and the x-axis indicates price.

The distribution is right-skewed and also illustrates the wide range of different selling prices. Hence, we aim to be able to accurately predict the selling price of cars to give our dealerships an advantage. Our objectives are:

1) Predict the selling price of vehicles. This would give our dealerships a competitive advantage by pricing their vehicles according to what consumers will purchase.

2) Investigate which factors result in a vehicle being priced a certain amount. This allows our dealerships to make informed decisions on what to look for when acquiring inventory.

III. Results

A. Predict the selling price of vehicles.

Our machine learning model, which uses a random forest algorithm, can predict with high accuracy what the selling price of a car should be based on factors such as make, model, transmission type, place of registration, condition, odometer reading, interior/exterior colour, and vehicle age. The metrics of our random forest model include a mean absolute error (MAE) of 1,253 and a root mean squared error (RMSE) of 2,089. This means that when our model makes a prediction, on average it will correctly predict the selling price of a car within +/- $1,253 according to the MAE metric, or within +/- $2,089 according to the RMSE metric. Additionally, our model has an R^2 value of 0.952, meaning that 95.2% of the variance in selling price can be explained by our model. Hence, dealerships will be able to use this model to optimally price their cars, reducing time and fees spent on appraisal and being quicker to the market with more competitive prices.

B. Investigate which factors result in a vehicle being priced a certain amount.

We investigated which vehicle features would impact the final selling price of a car, as well as the correlation of some of those features together. In Fig. 2, we can see the correlation matrix of some features such as year, condition, odometer reading, along with the selling price. For instance, higher odometer reading is correlated with the age of the car, and is negatively correlated with selling price. Note that there is also a feature "mmr", which stands for Manheim Market Report. This is a third-party estimate of the market price of the vehicle, which we did not include as part of our model since we aim to be the stakeholders providing this value and it would thus be irrational to use this as a predictor.

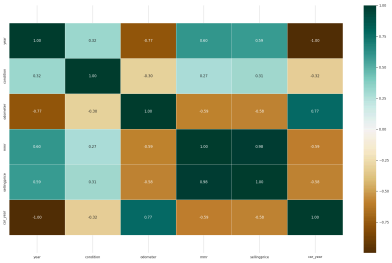

Fig. 2. Correlation matrix of car features. Values closer to +/-1.00 indicate a greater correlation. Variable "mmr" was dropped from our model.

In Fig. 3, we can see the feature importance graph. From this, we observed that odometer reading, make, body (e.g. SUV, sedan), and vehicle age are the four most relevant features for our model. This means that our dealerships should emphasize considering these four features when acquiring cars for inventory, as they are most crucial in determining price.

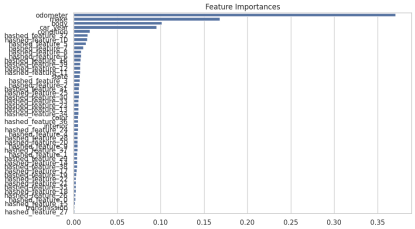

Fig. 3. Feature importance graph of car features. A greater bar length indicates a greater impact on the selling price.

[1] DesRosiers Automotive Consultants

[2] https://www.kaggle.com/datasets/syedanwarafridi/vehicle-sales-data/data