



Vehicle Price Prediction

Capstone Project Executive Summary

By: Robby Khoutsaysana

Email: Robbykh87@hotmail.com



EXECUTIVE SUMMARY

Introduction

One of the most important practices in the automotive industry is the ability to predict the price of a vehicle. This is especially important for the automotive industry as it is an extremely competitive market and the ability to predict the price of a vehicle is a key factor in the success of a dealership. The ability to predict the price of a vehicle is also important for the consumer as it allows them to make an informed decision on the purchase of a vehicle. This is a perfect area to apply data science techniques as the data is readily available and the problem is well defined.

Problem Statement

The sale of vehicles is a complex process. There are many factors that influence the price which are not always obvious. This could mean that a vehicle is not sold at the price it is worth, or that a vehicle is sold at a price that is too high. This equates to a loss of revenue for the dealership and a loss of value for the customer. To maximize the revenue of a dealership, it is important to understand the factors that influence the sale of a vehicle.

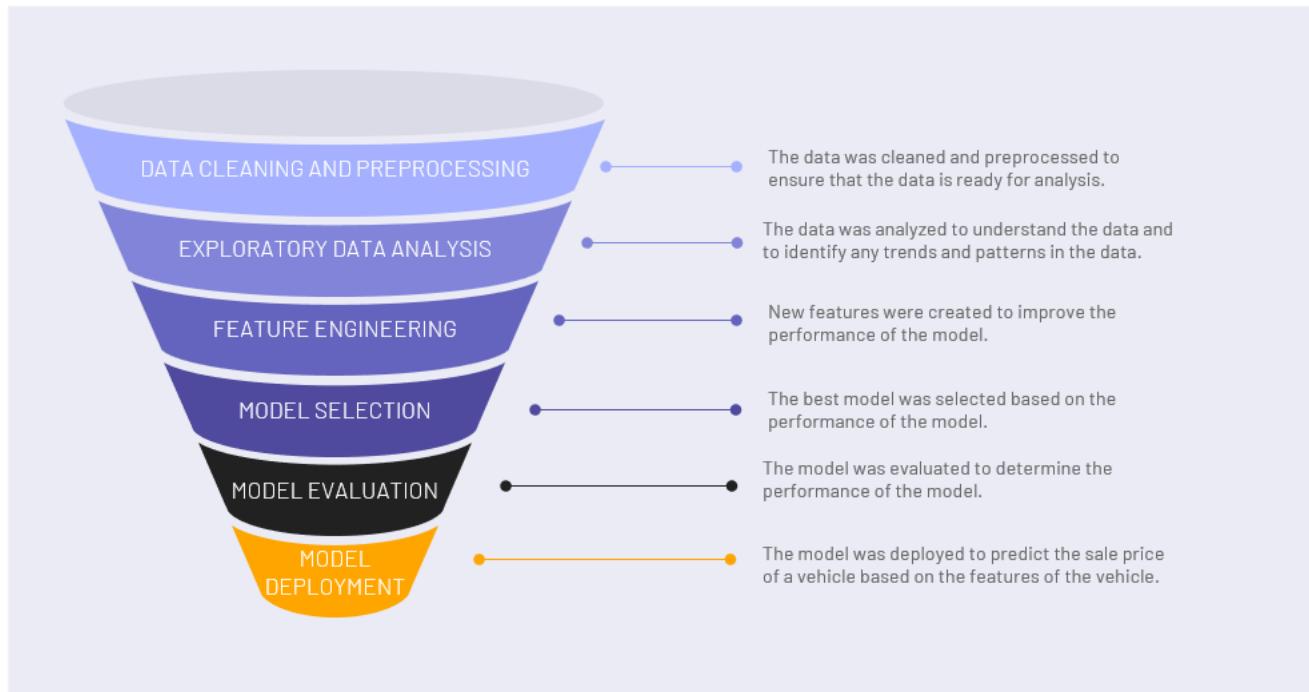
Business Task

The key task is to find out what features have a significant impact on the sale price of a vehicle and build a predictive model that can be used to predict the sale price based on its features. The insights gathered here can be used by the car dealers to price their vehicles competitively and by the customers to get a fair idea of the price of the vehicle they are interested in.

The Data

This data was obtained using web scraping techniques on cargurus.com. This data is for academic, research and individual experimentation only and is not intended for commercial purposes. This dataset contains 3 million new and used vehicle listings in the United States. The data set can be downloaded on kaggle.com. This data dictionary can be downloaded/reviewed [here](#).

Methodology



Data Cleaning and Preprocessing

In preparation for analysis and modeling, the data was cleaned and preprocessed using traditional data cleaning techniques. Standard data cleaning techniques such as removing duplicates, removing, or imputing missing values, removing outliers, and removing unnecessary columns were performed. Then the data was transformed using feature engineering techniques that were specific to the data.

Exploratory Data Analysis

Through the EDA, we have learned that the features of the vehicle are important to predicting the price of a vehicle. The features of the vehicle include the year, make, model, engine type, transmission type, mileage, speed, size, and if the vehicle has an incident. These features are important to predicting the price of a vehicle because they have a high correlation with the target variable. The higher the correlation, the more important the feature is to predict the price of a vehicle.

Correlation Matrix

	price
price	1.000000
horsepower	0.611184
torque	0.560698
size_of_vehicle	0.508744
year	0.456368
has_incidents_False	0.438806
major_options_count	0.366454
body_type_Pickup Truck	0.332278
engine_type_V8	0.321653
legroom	0.317011
maximum_seating	0.205582
engine_type_V6	0.190623

Insights

- The best vehicles to sell are SUV / Crossover, sedan, or a pickup truck. These are likely to sell fast.
- The average price of a vehicle is \$29933.37. This means that vehicles that are priced above this value will be harder to sell. Consider selling it for \$30,000 or less.
- There is a high importance in fuel economy and savings. It would be wise to purchase vehicles that emphasizes in these areas. For example, buying smaller vehicles with smaller sized engines that have great fuel economy.

Model Selection

Six different models were evaluated to determine which model would be the best fit for the data. The outcome resulted in XGBoost being the best model for predicting the price of a vehicle.

Model	R2 Score	MSE	RMSE	MAE
Linear Regression	0.82	0.05	0.22	0.15
KNN	0.94	0.02	0.13	0.09
Random Forest	0.96	0.01	0.13	0.09
LightGBM	0.94	0.02	0.12	0.09
XGBoost	0.96	0.01	0.1	0.07
Neural Network	0.95	0.01	0.11	NA

Conclusion

In recent years, it has been noted that many vehicles are being sold online. This is because it is more convenient for the buyer and the seller. The buyer can shop for a vehicle from the comfort of their home, and the seller can sell their vehicle without having to deal with the hassle of selling it in person. Therefore, it is important to understand the factors that affect the price of a vehicle. As a result, we have created a model that can predict the price of a vehicle based on the features of the vehicle. This model can be used by sellers to determine the price of their vehicle, and by buyers to determine if the price of the vehicle is fair. In the future, we can expand this model to include more features of the vehicle, and to include more data. This will allow us to create a more accurate model that can predict the price of a vehicle with more accuracy.