Car Resale Value Prediction

INTRODUCTION: Nowadays, Machine Learning and Artificial Intelligence are applicable in almost every sector. Companies are adopting smart AI solutions in their product to eliminate manual interventions. Let's keep ourselves confined to cars, and we will see how it has changed the driving experiences. Explores the potential of forecasting methods to support decision making in the automotive industry. More specifically, we concentrate on the second-hand market and develop empirical models to forecast resale prices. In the second-hand car market, demand depends substantially on the difference between a car's residual value and its offer price. Thus, to decide on offer prices, sellers need to estimate residual values and resale prices, respectively. Such forecasts are also important in the new car business where leasing has become a major sales channel. Leasing companies set prices based on expected residual values If the actual resale price of a car falls below expectations, the company faces a loss. Consequently, the accuracy of resale price forecasts is linked directly to the profitability of car leasing. More generally, given the impact of pricing policy on firm performance and the dependence of pricing decisions on residual value estimates, we argue that resale price forecasting is indirectly connected to the profitability of car selling.

PROBLEM STATEMENT:

With difficult economic conditions, it is likely that sales of second-hand imported (reconditioned) cars and used cars will increase. In many developed countries, it is common to lease a car rather than buying it outright. After the lease period is over, the buyer has the possibility to buy the car at its residual value, i.e., its expected resale value. Thus, it is of commercial interest to sellers/financers to be able to predict the salvage value (residual value) of cars with accuracy.

To predict the resale value of the car, we proposed an intelligent, flexible, and effective system that is based on using regression algorithms. Considering the main factors which would affect the resale value of a vehicle a regression model is to be built that would give the nearest resale value of the vehicle. We will be using various regression algorithms and algorithm with the best accuracy will be taken as a solution, then it will be integrated to the web-based application where the user is notified with the status of his product.

TECHNICAL FEATURES:

The main idea of making a car resale value prediction system is to get hands-on practice for python using Data Science. Car resale value prediction is the system to predict the amount of resale value based on the parameters provided by the user. User enters the details of the car into the form given and accordingly the car resale value is predicted.

- Fuel Type
- Manufacturing Year
- Miles Driven
- Number of Historical Owners

This is a supervised learning problem and can be solved using regression techniques. We need to predict the selling price of a car based on the given car's features. Supervised Regression problems require labelled data

where our target or dependent variable is the selling price of a car. All other features are independent variables. Maintenance Record Following are some regression algorithms that can be used for predicting the selling price.

- Linear Regression
- Decision Tree Regressor
- Support Vector Regressor
- KNN Regressor
- Random Forest Regressor
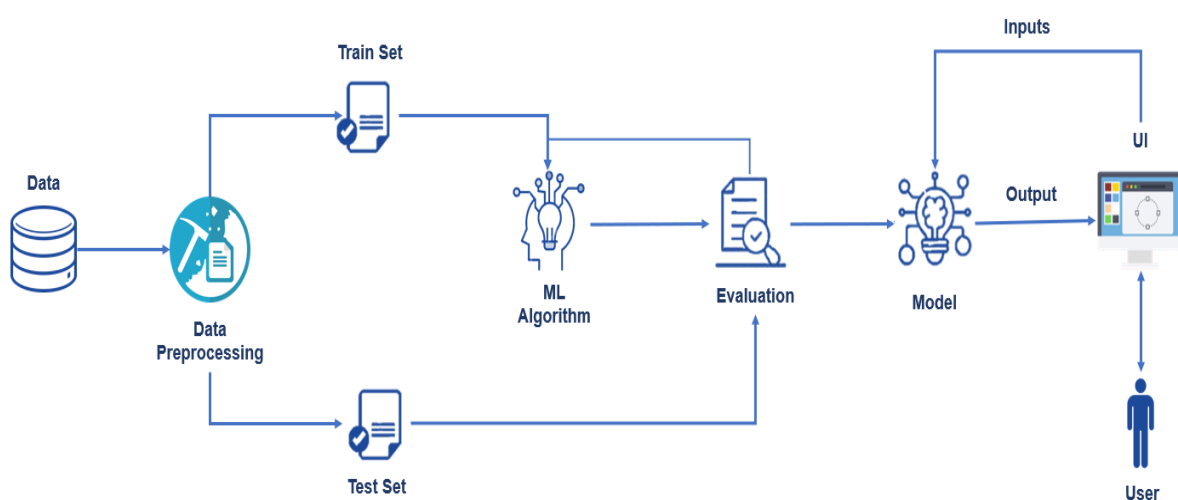
- DATA ANALYSIS
- PRICE DISTRIBUTION FOR DIFFERENT FUEL TYPE CARS
- CORRELATION MATRIX
- PAIR PLOT
- FITTING THE MODEL

CONCLUSION:

The prediction error rate of all the models was well under the accepted 5% of error. But, on further analysis, the mean error of the regression tree model was found to be more than the mean error rate of the multiple regression and lasso regression models. Even though for some seeds the regression tree has better accuracy, its error rates are higher for the rest. This has been confirmed by performing an ANOVA. Also, the post-hoc test revealed that the error rates in multiple regression models and lasso regression models aren't significantly different from each other. To get even more accurate models, we can also choose more advanced machine learning algorithms such as random forests, an ensemble learning algorithm which creates multiple decision/regression trees, which brings down overfitting massively or Boosting, which tries to bias the overall model by weighing in the favour of good performers. More data from newer websites and different countries can also be scraped and this data can be used to retrain these models to check for reproducibility