Used Car Price Prediction

Abstract

Determining whether the listed price of a used car is a challenging task, due to the many factors that drive a used vehicle's price on the market. The focus of this project is developing machine learning models that can accurately predict the price of a used car based on its features, in order to make informed purchases. I implement and evaluate various learning methods on a dataset consisting of the vehicle trim, Dealer listing price of different makes and models.

Conclusion

By performing different models, it was aimed to get different perspectives and eventually compared their performance. With this study, it purpose was to predict Dealer Listing prices of used cars. With the help of the data visualizations and exploratory data analysis, the dataset was uncovered and features were explored deeply. The relation between features were examined. At the last stage, predictive models were applied to predict price of cars in an order: random forest, linear regression, ridge regression, lasso, Lasso, XGBoost.

Which variables are significant in predicting the price of a car?

newest car, vehicle with 1 owner, buyback and protection eligible, extended warranty, Vehicle exterior color with Bright White Clearcoat and vehicle with interior black.

- the Jeep manufacture has the higher price than Cadillac.
- the price of the newest car is higher than oldest car.
- the vehicle with 1 owner, buyback and protection eligible has more demand and the price of the vehicle is more than others.
- some states like IL, MD, IN, NC, PA, GA, MN and WI seller sell their vehicle at the highest price.
- The price of all vehicle with extended warranty are lower than other vehicles.
- the SUV is more demand than others style.
- the Vehicle exterior color with Bright White Clearcoat is mor demand and to be sold followed by other colors.
- Most vehicle with interior black is demand and sold more than other colors.

Future Work

- Keep the current model as a baseline, we intend to use some advanced techniques algorithms to predict used car prices as our future work.
- To improve model performance, I will use hyperparameter tuning like GridSearchCV and to improve prediction performance I will use ensemble method like voting classifier to get the better result.
- Adding more independent variables to a regression model tends to increase R-Squared.
- Cross Validation to improve the R Squared and use adjusted R Squared to compare fits.