Used Car Price Prediction

Introduction

Lack of knowledge regarding the features put customers in loss either someone purchasing or selling a used car. Knowing features well help customers to make informed decision and prevent them making good negotiation during car purchase. We will find out, what is the price of a used car? What are the features one should look into before buying a car based on given features and factors such as brand or company, manufacturing year, mileage, Accident, Engine Type etc. in order to buy more used cars.

Background

- Objective: To predict the price of used cars based on muliple features.
- Importance: Based on car's features, we predicted the price of used cars.
- Target variable: Price

Methodology

Pre-Processing

- .Label encoder used to change categorical data so that we can improve performance and our data fit and predict accurately using each of the model algorithms.
- Missing data checked and removed. Removed unwanted data to avoid over fitting such as clean_title, accident, fuel type

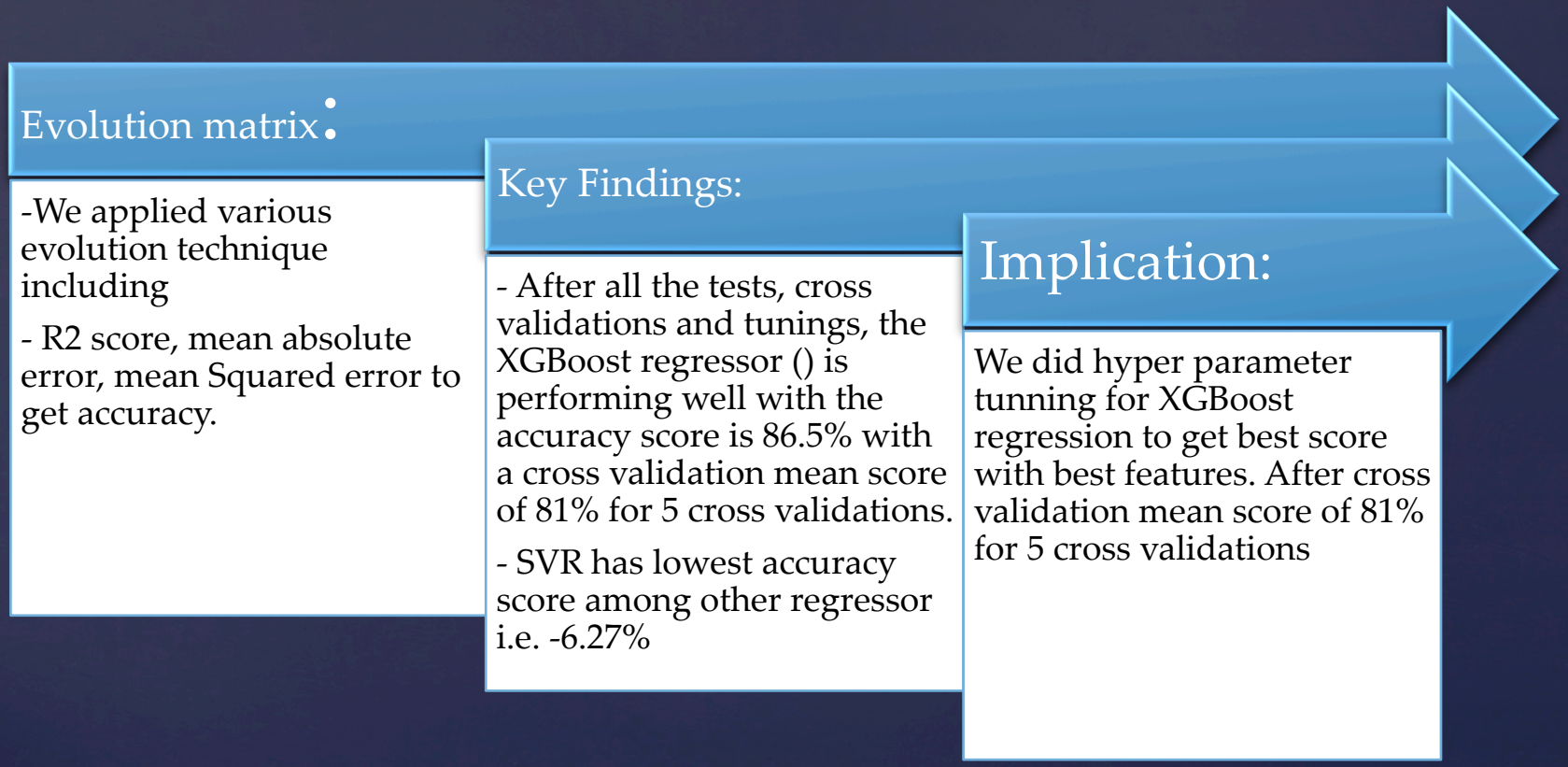
Feature Engineering

- Outliers removed, Model year, mileages, price have some outliers. We have removed them so that our data will fit properly when proceeding next step of modelling.
- Kept all highly correlated data such as price, mileage, engine .

Machine learning

- I have used several regression classifier model to predict the price of used car using different features with Cross validation of 5 split to avoid over fitting. .
- After all the tests, cross validations and tunings, the xgb regressor () is performing well with the accuracy score is 86.5% with a cross validation mean score of 81% for 5 cross validations. On the other hand SVR has lowest accuracy score among other regressor i.e. -6.27% .

Result and Analysis



Findings

- Mileage, engine, model year are top features for predicting used car prices.
- accident, clean title, fuel type are not recorded as they are not used in modeling .
- Most of the cars manufactured in 2022. Last Spike of car manufacture was in 2008 and gradually increased 2015 to 2022. Cars with the model year 2022 have the highest count and lowest in 1974 and dropped after 2023 specially in 2024 .

Future Improvements

For future improvements I would like to set

Different n_estimators, max_depth, max_features, **and** criterion as I have selected best params.

I would like to focus on refining my model's features by addressing multicollinearity.

Conclusion

- In this presentation, we have predicted the price of used car on the bases of features of car. We have seen mileage, engine, model year are the top three features people see to buy a car. On the other hand Accident, clean title, fuel type are also important factors to predict the price although we have missed data in our dataset which we have not considered to predict the car price.
- XGBoost gave 81% accuracy even after the hyper parameter tunning and cross validation. To compare better result in future we can modify n_estimator, max_depth to see if there is any changes in accuracy scores.

Overall, model works well and help us to predict good scores.