

# Car Price Prediction Project

Submitted by :-

Mohit Meel




## Acknowledgement

The data used for the project is scraped from the website cars24.com which means it is primary data.

The data is related to the used cars.

The company has provided the problem statement of the client company as the client company works in buying and selling of used cars in the market.

The steps or guidelines are given by the company in order to help in completion of the project.




# Introduction

- The problem statement is related to the change in the supply and demand of the cars in the market after the covid 19. The cars which are in demand results rise in the prices and supply and vice-versa. The covid 19 has impact the market of India in all fields and therefore the company are now making those assets which are in demand due to the financial problems.
- The markets of India related to any field is facing the issues of the financial as the GDP has declined due to crises of covid 19, the sellers are only supplying those products which are in demand, the main reason is financial problem and this problem will take time to be solved.
- The researches done on the problem is related to prices of the used cars in the market on the basis of their properties. The study material is taken from the website cars24.com to exactly knowing about the market of used cars, which car is high demand or less demand and which car prices are high or low.
- The motive of making this project is analyzing the market of used cars and thus applying the predictive models for knowing the actual price as compared to prices set up by the market.

# Analytical Problem Framing

- The sources from where the data is collected is from the website cars24.com, the technique used for scraping the data is web scraping selenium. The data collected of used cars is not from one city it is from different cities.
- The data collected has the different properties of the used cars with their prices and the amount of data is 5575. The different properties related to used cars are brand, variant, manufacturing year, transmission, owner, driven, fuel type, down payment and prices. The data is in csv format and it comprises 5575 rows and 9 columns.
- The study of data is done by using different codes in the jupyter notebook as head, tail, dtypes, columns, info, is null, is null sum, shape, etc. After that the visualization of data is done for which I have used the heatmap in the codes of is null and knowing the correlation of features with target variable, used the hist and kde plot in checking skewness present in the dataset, used the boxplot for the checking of outliers present in the dataset. Label encoder is used to convert the object datatype into numerical datatype. Power transform is used in removing the skewness and for scaling the dataset, the standard scaler technique is used



## Model's Development and Evaluation

- For solving the problem I have used the different models for the machine to learn and predict the target variable. In the training and testing, the data is divided in such a way that 75% of data is for training and 25% of data is for testing. The random state used for achieving highest accuracy is 368.
- The problem is related to the predict the prices of the used cars to ascertain the actual value as compared to the prices of the market. The target variable which is to be predicted is continuous so to solve it the regression models will be used.
- The regression models used for predicting the target variable with their score achieved are linear regression-99%, lasso-99%, ridge-99%, support vector regressor with kernel list(rbf, poly, linear)-(62%, 61%, 33), random forest regressor-99% but the best model stands in all is the linear regression as the score is achieved is highest.
- The visualization used in the modeling phase is the bar plot in the lasso and ridge and the best fit line in the scatter plot is used with the linear regression and random forest regressor.
- The results achieved by the models is showing the predicted price(actual value by machine) is the somewhere little less in most cars than actual value(predicted value) of the market.

# Conclusion

- The findings or observations that I came across is that the prices that predicted by the machine as actual value is somewhere less than the prices of the market as predicted value. But the difference is very little. The market prices and the prices which are predicted are almost same. The prediction done for knowing the prices stands benefit for the client company for knowing the market and also ascertain the losses or profits.
- The learning which I got from making the project is that how the market is affected by the covid 19. The GDP of the country is declined during the session of covid which states that covid had affected everyone. The visualization done in the modeling phase has shown the best fit line in which every points have come on the line, thus this results that the model has predicted the best outcome.
- The limitation of the data is that it has very less feature columns which I think is not up to the mark of ascertaining the proper price of the car as there are so many things when a particular car price is fixed such as the wear and tear of the car, Tyres of the car, insurance of the car, how much year left for the car to be run on the roads, is there any claim taken by the owner, what is the actual claim amount sanctioned by the insurance company if any mishappening occurs, service records, etc. All these details to be present will help in achieving better results.