

Used car price prediction

Problem statement

In this dataset we will predict the price of used cars from various sites.



understanding

- This is a car price prediction dataset in which data is collected through web scrapping from various sites.
- Data is collected in a dataframes from various sites.
- All the dataframes are then concatenated to become a single dataframe .

understanding

- Because of covid there are lot of changes in the cars price .so people are preferring used cars .
- Peoples are facing problems in the valuation of used cars .Lots of features are to be considered for the price of car.
- New machine learning models are to be needed from new data for the valuation of cars.

EDA

- EDA is exploratory data analysis
- Which includes the following:-
- Checking the shape (.shape) gives the number of rows and columns.
- Checking the description(.describe()) gives the count,mean,standard deviation,max and min value and quartiles.
- See the datatypes of columns(.dtypes)and if they are object change them into integer and float.

EDA

- Checking for nulls (using `.isnull().sum()`) and treat them using imputers.
- Checking for zeros and treat for it using log, mean and mode.
- Checking for whitespaces and replace them with nulls and then treat them

VISUALIZATION

- This includes visualizing the data as we can better understand through charts and graphs.
- For that importing necessary libraries.
- Checking for outliers using boxplot.
- There are many techniques to deal with outliers such as z score, inter quartile range.
- After dealing with the outliers lets split the data into features and target

MODEL BUILDING

- Train, test and split the data.
- After preprocessing model initializing and model building .
- Check for accuracy.
- Cross validation and hyper parameter tuning is necessary.
- Try to improve the accuracy using different parameters.
- Thank you