LEVEL 2 : Price Prediction

Data Description

Source: Kaggle

This is a dataset picked from websites selling used cars and contains information about used cars that were sold.

There are 6702 records with 13 columns (*name,year,selling_price,km_driven,fuel,seller_type,transmission,owner,mileage,engine,max_power,torque,seats*).

Problem Statement

To predict the selling_price of the car using the given features of used car.

R-squared or MSE or MAE can be used for evaluating the model's predictions.

Features -
name,year,km_driven,fuel,seller_type,transmission,owner,mileage,engine,max_power,torque,seats.

Target column – ***selling_price.***

Hints

Steps that can be followed to achieve decent predictions:

1) Normalize the data : As the data contains columns in different scales, bringing all the columns to a common scale will boost performance of model building and gets less influenced by outliers.
2) Feature Engineering : Perform feature engineering on columns **mileage**, **engine**, **max_power** and **torque**. To make better use of the integer values in them.
3) Feature Selection : Perform feature selection and look for those features which will affect model adversely and eliminate them.
4) Strings to Integer conversion: Convert all the string column values to integers using One-hot encoding or Label Encoding. (This is a compulsory step)
5) Model Building : Use the processed data to build a ML model using any algorithm of your choice. (This is a compulsory step)

Note : The above-mentioned steps are not exhaustive, they are just for guidance, one can use additional ML techniques and algorithms to achieve better predictions.

References

https://towardsdatascience.com/used-car-price-prediction-using-machine-learning-e3be02d977b2