Understanding Problem Statement

Machine Learning has become a tool used in almost every task that requires estimation. Companies like Cars24 and Cardekho.com use Regression analysis to estimate the used car prices. So we need to build a model to estimate the price of used cars. The model should take car-related parameters and output a selling price. The selling price of a used car depends on certain features as mentioned below:

1.Fuel Type
2.Manufacturing Year
3.Miles Driven
4.Number of Historical Owners
5.Maintenance Record

This is a supervised learning problem and can be solved using regression techniques. We need to predict the selling price of a car based on the given car's features. Supervised Regression problems require labelled data

where our target or dependent variable is the selling price of a car. All other features are independent

variables.

Following are some regression algorithms that can be used for predicting the selling price.

1.Linear Regression
2.Decision Tree Regressor
3.Support Vector Regressor
4.KNN Regressor
5.Random Forest Regressor

Linear Models are relatively less complex and explainable, but linear models perform poorly on data containing the outliers. Linear models fail to perform well on non-linear datasets. In such cases, non-linear regression algorithms random forest regression and xgboost Regressor perform better in fitting the nonlinear data.

In this tutorial, we will use Random Forest Regressor for predicting the selling price of cars. Our data contains some outliers, and treating them is entirely possible, but the performance of nonlinear regression models is insensitive to outliers.