**Car Price Prediction**

**Project Members:**

PARA HITESH BT17GCS154

EAGALA RAKESH BT17GCS025

CHENNUPATI SAI REVANTH BT17GCS152

**Objective:**

Our Goal is to design a working model which can predict the prices of second-hand cars based on user inputs.

**Motivation:**

Lots of people in market ask for personal details when we want to check the second-hand price of our car. So, our Model can predict their prices without compromising on the User’s Privacy and Data Integrity

**Summary of Mid Semester-1:**

* We read three research papers(Shalini Goyal, Panwar Abish Ali, Encs Gokce).
* According to Literature review we found that regression techniques like XGBoost, Linear regression and Random forest regression will give better accuracy and least errors.
* We integrated the model with Flask and HTML.

**Methodologies:**

**A. Dataset:**

We took dataset from famous second-hand car selling website “Car Dekho” to use in our project. And it has feature like Present price, selling price, Odometer readings etc.

**B. Data Pre-processing:**

We exploring the data and filling the null values and performing some data proposing techniques like one hot encoding technique to convert categorical variables to numerical.

**C. Algorithms we used**:

The algorithms we are going to use are

1. Random Forest
2. Linear Regression

**Random Forest:**

Random Forest is an ensemble machine learning technique capable of performing both regression and classification tasks using multiple decision trees and a statistical technique called bagging.A RF instead of just averaging the prediction of trees it uses two key concepts:

1. Random sampling of training observations when building trees
2. Random subsets of features for splitting nodes

In other words, Random forest builds multiple decision trees and merge their predictions together more accurately and stable prediction rather than relying on individual decision trees.

In our model we use sklearn library to implement Random Forest Regressor and we use Randomized Search CV with 3 fold cross validation for hyper parameter tuning.

**Linear Regression:**

Linear regression is a simple yet powerful supervised learning technique. The aim of linear regression is to identify how the input variable(explanatory variable) influences the output variable(response variable). The core components in a simple linear regression are:

1. Continuous input variable
2. Continuous response variable.
3. The assumptions of linear regression being meet.

* The assumptions of linear regression are,

1. Linear association between input and output variable
2. Normally distributed errors and
3. Independence of error term with input.

In our model we use stats models to create linear regression model and it also tells about all the statical parameter values like and adjusted value and P values. And we choose the best model from the two.

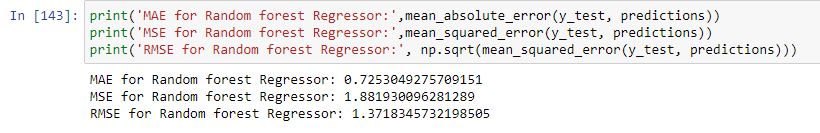
**D. Integrating with flask:**

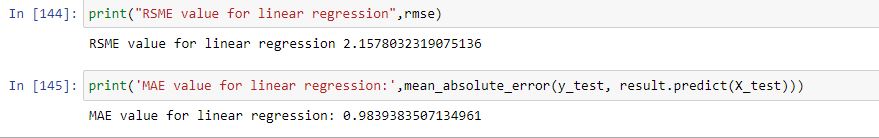
After successfully creating the model we saved this in pickle format and we integrated this model in flask and we use HTML form for user input.

**E. Deployment:**

We also deployed this model in Heroku.

**Results:**

Random Forest Regression

Linear Regression

**Result Analysis:**

1.We calculated the value of some regression model evaluations metric like RSME and MAE.

2.We clearly observed that the error rates of Random forest Regressor is very low when compared to Linear regression.

3.So we concluded that the Random forest Regressor will give better results for out data set.

**Future Scope:**

1. Now we implemented only two regression algorithms.We could also perform other complex regression algorithms and can find the best algorithm for car priceprediction.
2. We can develop very useful model with more accuracy when our dataset contains a greater number of features.