



# Predicting Used Car Prices Using Machine Learning.

By  
Bhargava rathod  
SaiPrathyusha veguru  
Sheshidhar reddy shaga  
Likith Chowdary  
Meghana eravelli

# Abstract

- The core objective of this project is to recommend the selling price of a used car. As many people want to sell their used car but most of them are not sure how much is the value of their car. In such cases our project will be more suitable for the consumers.
- So, we have trained a regression model that will predict the selling price. We have applied few machine learning techniques for the price of cars with the available independent variables.

# Data Specification

- We have used the dataset which was downloaded from the Kaggle repository (<https://www.kaggle.com/nehalbirla/vehicle-dataset-from-cardekho>). The dataset contains 7 independent variables and 1 dependent variable. The below are the details.
- Where car name, year, kms driven, Fuel, seller\_type, Transmission, Owner, Present Price are the independent variable. And Selling price is the dependent variable.
- The data set has 301 data points and 9 columns.
- The dataset is divided into 80% training and 20% testing.

## Design and Milestones

- Our whole project is divided into 2 phases.
  - 1. Training the Regression model.
  - 2. Deployment of project using Heroku.

# Training the Regression model

- For this phase we have used Jupyter notebooks to train the model.
- Before Training the model, we have performed EDA, Pre-processing, Vectorization, and feature engineering on the taken dataset.
- In EDA phase we came to know the dataset has zero null values, there are 4 columns that are string type means categorical.
- In pre-processing, We have converted all the car names into lowercase where we have replaced the space between the car names with underscore. We have performed standard scaling for the kms driven.
- In vectorization and Feature engineering all the categorical columns except the car name are converted using label encoding.
- For the car name we have used one hot encoding.
- Since the dataset is from 2020, we have created a new feature i.e., **num years**.

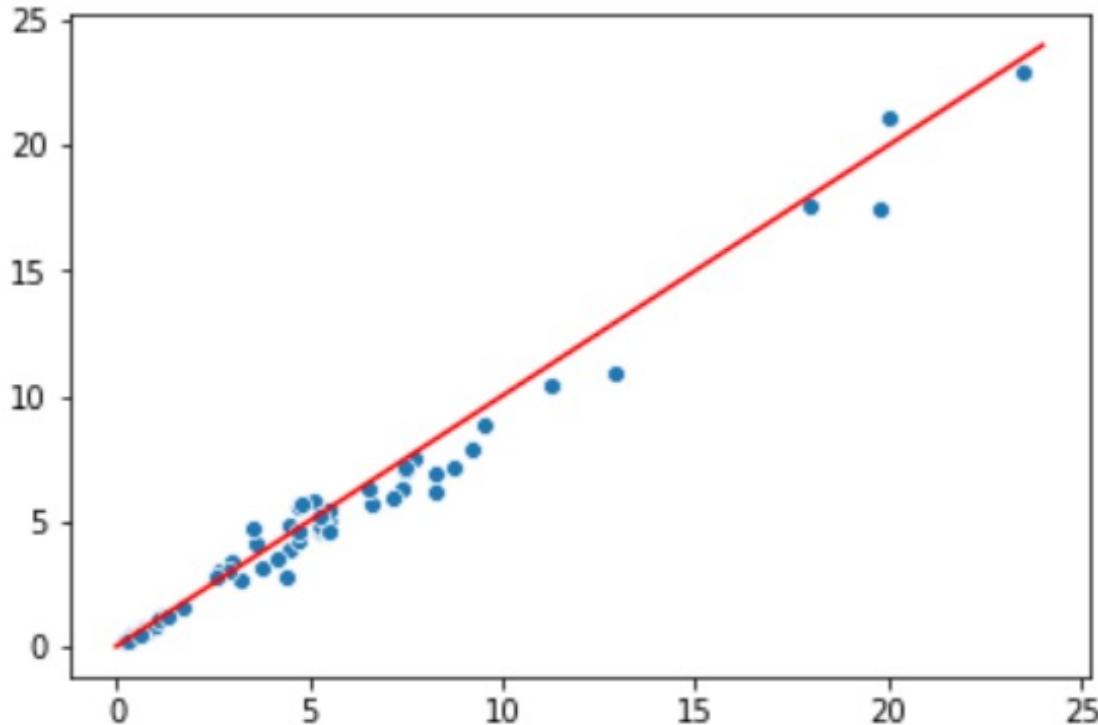
## Modelling

- We have trained 7 different models with different regression algorithms. And measured the mean square error for all of them.
- Out of 7 we have achieved the lowest mean square error for ExtraTreesRegressor model. So, we have picked this model for predictions.
- Below is the table for all the mean square achieved with respective models.

<u>Model</u>	<u>Mean Squared Error</u>
Linear Regression	1.23
Lasso	3.33
Ridge	1.52
KNeighborsRegressor	1.15
DecisionTreeRegressor	2.53
RandomForestRegressor	1.69
ExtraTreesRegressor	0.59

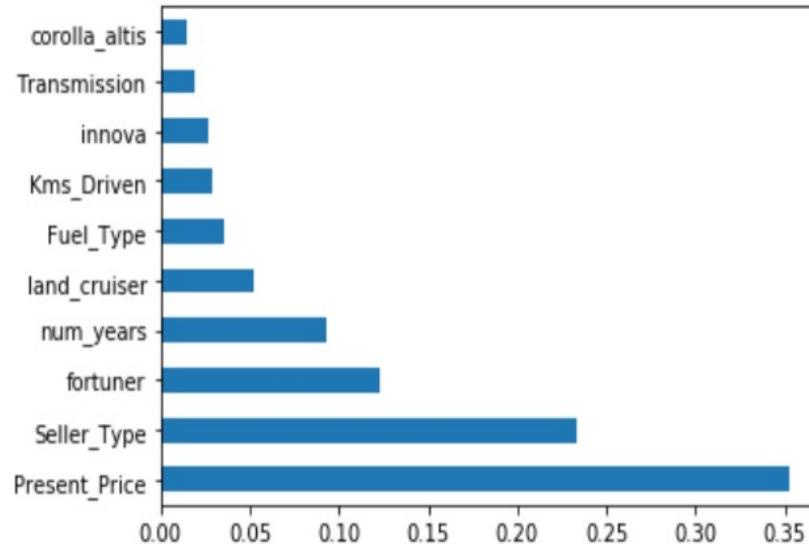
# Scatter Plot

```
: sns.scatterplot(y_test.reshape(-1),predicted)  
plt.plot(line,color = 'r')  
plt.show()
```



# Top important features.

```
#plot graph of feature importances for better visualization  
feat_importances = pd.Series(ExtraTrees_model.feature_importances_, index=columns)  
feat_importances.nlargest(10).plot(kind='barh')  
plt.show()
```



## Saving the models

- After training process is done, by using the best estimator technique we have saved the model.pkl for predictions.
- Along with the trained model we have also saved the vectorizer.pkl and scaler.pkl.

# Deployment using Heroku

- We have used flask API where we have integrated the front-end programmed using HTML, CSS and back-end using python.
- Then we have deployed the whole code into GitHub repository. And from the Heroku app we have imported the code from GitHub directly and deployed the model into production in just few steps.

# User Interface

Google | Music Recommendation by Fai... | bhargavarathod/Predicting-us... | Document | k\_ Vehicle dataset | Kaggle | +

used-car-price-estimate.herokuapp.com/?

Car Name: ignis

Present Showroom Price(In lakhs): 571000

Car Registration Year: 2017

Total KiloMetres Driven: 2400

Number of Owners: 0

What is the Fuel Type?

- Petrol
- Diesel
- CNG

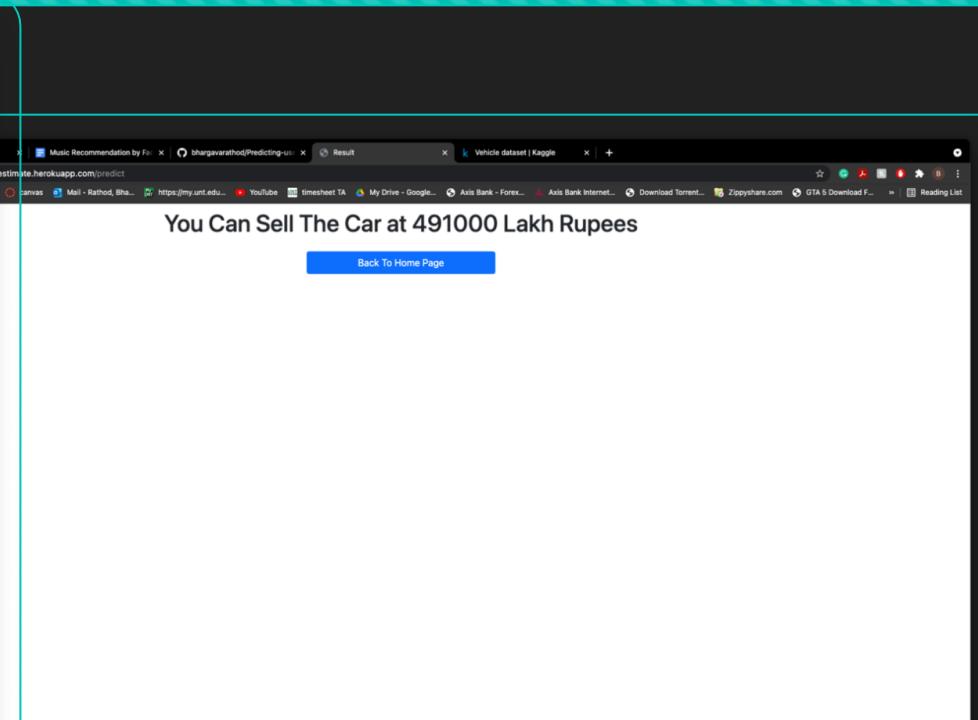
What is the Seller Type?

- Dealer
- Individual

What is the Transmission Type?

- Manual
- Automatic

[Calculate the Selling Price](#)



## Results and Analysis

- After deployment and testing with different input variables. The results were very optimal. The project came out as good as we planned.
- Furthermore, we tried to get a bigger dataset with Present price column in it.
- On performing 7 types of multiple regression model the best mean square error was achieved by using extratreeregressor model.

# LINKS

Github repository:

- <https://github.com/bhargavarathod/Predicting-used-car-prices-using-machine-learning.git>

Link to access the User interface to get predictions:

- <https://used-car-prediction-machine.herokuapp.com/>