

USED CAR PRICE PREDICTOR

BUSINESS PROBLEM

Curacel is in the business of second-hand vehicle sales. It helps owners and renters predict the price range to value a vehicle. It helps shoppers plan for their car needs. Currently, there are easy ways to predict the price range to value a car and help shoppers plan for their car needs. Owners basically browse through their onlistings to determine what to value their car.

As a Software Engineer - ML, I was tasked to build an automated service (web) to constantly update a prediction model with happenings in the marketplace so that, at every point in time, users can get semi-accurate prediction of the prices of proposed cars or advice them to make price adjustments for existing ones.

Also, one could predict what a car might have been sold at four(4) months ago because the app keeps track of the model for each time period. The marketplace is the available car listing services, providing authenticated api to access the service.

BUSINESS UNDERSTANDING

The problem is a Machine learning problem. In solving this, I built a regression model that predicts price based on the predictor variables supplied to the model.

After building the model that predicts price, I built an automated service (web) - REST api that calls our target variable after supplying parameters for predictor variables.

DATA PREPARATION

The used car dataset used for the regression model was obtained from Kaggle. 'Train-data.csv'. Multiple Linear Regression Model (MLRM) was used because our target variable is a continuous value. The analysis was done using Python Pandas, numpy, matplotlib, and pylab libraries. The variables used for our analysis are:

Target/Dependent Variable - 'Price'

Predictor/Independent Variables - 'Kilometers Driven',

'Year of Purchase of Car'

'Fuel Type'

'Transmission'

Price

'Price' in the dataset has numerical values. The unit of Price in the dataset is lakh. It's named 'Price' in the dataset.

Kilometers Driven

The values of Kilometers driven in the dataset are numerical and continuous. The variable is named 'Kilometers_Driven' in the dataset.

Year of Purchase of Car

I could not get a dataset that had ages of used cars in years and months. The dataset I got only had values in years. If provided with a dataset that has values in years and months, I'll be able to build a model that will predict prices of used cars based on years and months since purchase of car.

The variable is named 'Year' in the dataset.

Fuel Type

The fuel type variable in the dataset is a categorical variable with three classes. Diesel, Petrol and CNG. It's named 'Fuel_Type' in the dataset.

Transmission

The transmission variable in the dataset is categorical with two classes. Manual and Automatic. It's named 'Transmission' in the dataset.

MODELLING

Multiple Linear Regression Model (MLRM) was used for our model. The independent variables which had categories were converted into continuous variables by encoding the categories to make our regression model work.

The encodings are: Fuel Type: 0=CNG, 1=Diesel, 2=Petrol
Transmission: 0=Manual, 1=Automatic

The main dataset gotten from kaggle was cleaned to include only the variables for our analysis.

The cleaned data was then split into train and test sets. Train set was used to train our model and the test set was used to test our model and make predictions.

The regression model was saved and model columns of the independent variables were also saved.

MODEL EVALUATION

The model was evaluated using our test dataset.

MODEL DEPLOYMENT

The model was first deployed to the local host on Postman App to test.

The Model was deployed on a REST-API hosted on Heroku.

On Postman App, Basic json passed to POST METHOD as

```
[  
  {"Kilometers_Driven": 87000, "Year": 2012, "FuelType": 1, "TransmissionR": 0},  
  {"Kilometers_Driven": 25000, "Year": 2017, "FuelType": 1, "TransmissionR": 0}  
]
```

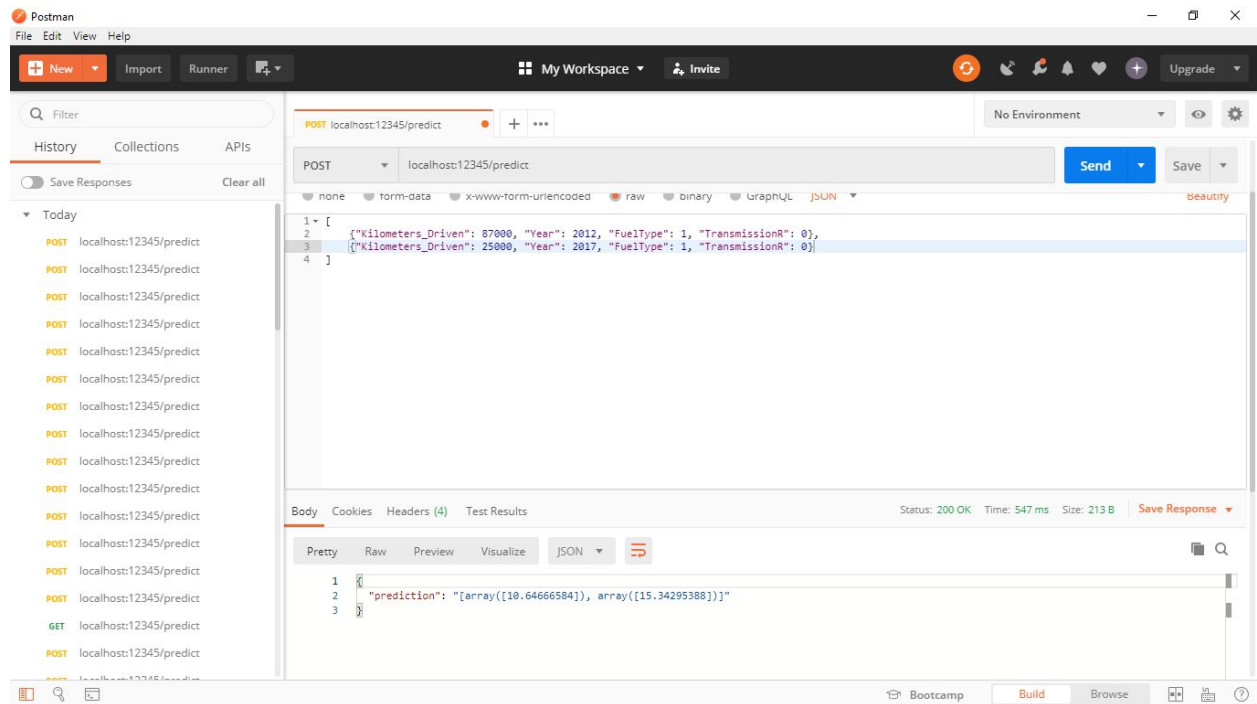
Returns predicted Used Car Price Value as

```
{  
  "prediction": "[array([10.64666584]), array([15.34295388])]"  
}
```

Where, **10.64 lakhs** is the Predicted Used Car price when Kilometers driven is 87,000, Year of Purchase of Car is 2012, Fuel Type is Diesel, and Transmission is Manual.

And **15.34 lakhs** is the predicted used car price value when Kilometers driven is 25,000, Year of Purchase of Car is 2017, Fuel Type is Diesel, and Transmission is Manual.

Here's a Screenshot of Postman.



Here's the link to the repository containing the notebooks, models, app.py, and dataset used for the project. <https://github.com/aoluwatooni/Usedcarpricepredictor>

Here's the link to the automated web service REST api -
<https://oadebiyi-usedcarpricepredictor.herokuapp.com/>

Thank you,

Oluwatooni Adebisi