Sunbeam Institute of Information Technology

Machine Learning Case Study : Used Car Price Prediction

Develop a flask application to perform Used Car Price Prediction using machine learning

The price of a new car in the industry is fixed by the manufacturer with some additional costs incurred by the Government in the form of taxes. So, customers buying a new car can be assured of the money they invest to be worthy. But, due to the increased prices of new cars and the financial incapability of the customers to buy them, Used Car sales are on a global increase. With the covid 19 impact in the market, we have seen lot of changes in the car market. Now some cars are in demand hence making them costly and some are not in demand hence cheaper.

Therefore, there is an urgent need for a Used Car Price Prediction system which effectively determines the worthiness of the car using a variety of features. Existing System includes a process where a seller decides a price randomly and buyer has no idea about the car and its value in the present-day scenario.. In fact, seller also has no idea about the car's existing value or the price he should be selling the car at. To overcome this problem there is need to developed a model which will be highly effective to predict the actual price a car rather than the price range of a car.

This application is having two parts

1) Data Cleaning EDA and Model Building

Data Cleaning -:

- The dataset is large and it may contain some data error. In order to reach clean, error free data some pre-processing Require. Read "**vehicles.csv**" datasets and perform data cleansing, convert datatypes in appropriate datatypes,remove duplicates, remove missing values,remove extra column keep only require column change posting date to date time format and get the day and month. Visualize missing data using bar plot and heat map. Remove outlier and save the cleaned data
- Observe the variables and their distribution do EDA and save the data . Classify manufacturers by their countries, classify car type and car color in large groups.

Data Modeling (Price Prediction) -:

- Handle the categorical column data, Prepare categorical features for correlation matrix .Use label encoding technique, LabelEncoder to convert categorical values to assign numerical values or the model.

Model Building -:

Once the data is pre-processed split the data in training and testing build three machine learning models .

1. Linear Regression model
2. XGBoost Regressor model
3. Random Forest Regressor model

Evaluate the performance of above models and identify the best performing model for price prediction. Save that model as "car_price_prediction.pkl".

2) Flask and User interface

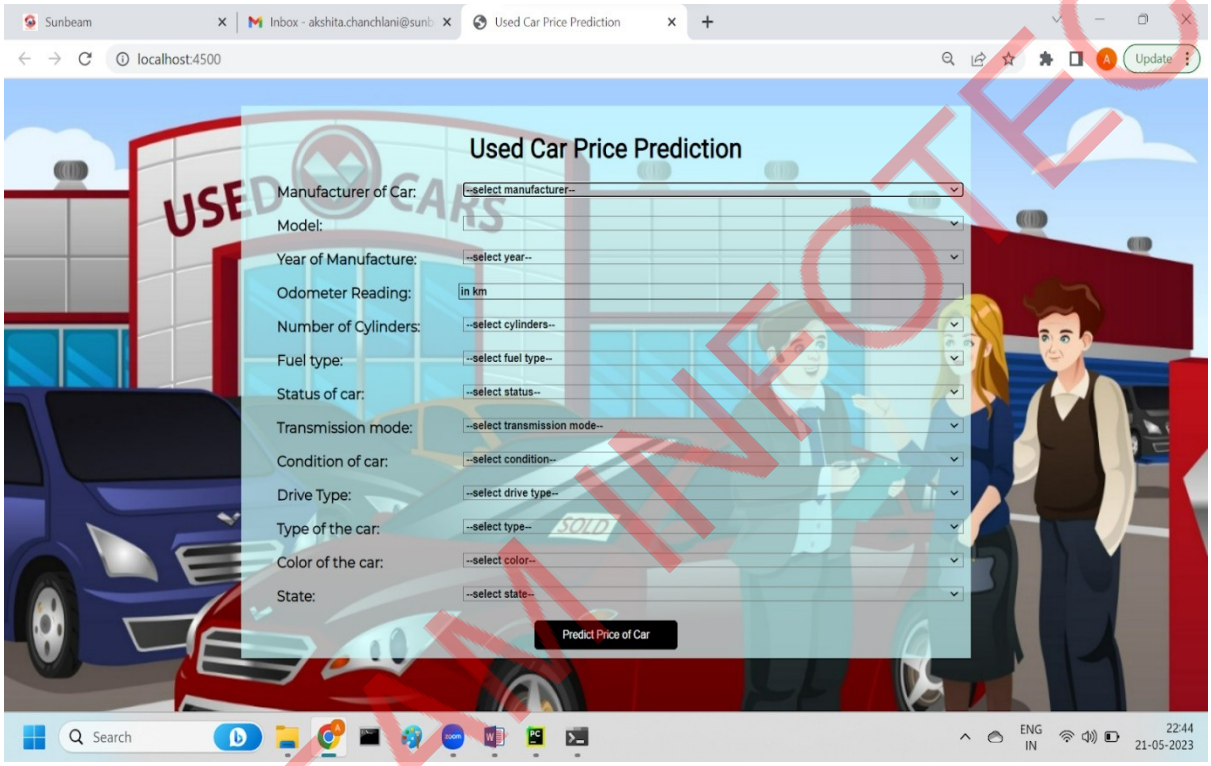
- Load the trained model car_price_prediction.pkl

Create an API with (“/”,method = “GET”)

It should display “car_price_prediction.html”

Create another API "/predict"

It should take input from user and do the prediction on price



Sample html form “/”