**PREDICTING USED CAR PRICES**

**-MACHINE LEARNING**

Problem Statement:

Deciding whether a used car is worth the posted price when you see listings online can be difficult. Several factors, including mileage, make, model, year, etc. can influence the actual worth of a car. From the perspective of a seller, it is also a dilemma to price a used car appropriately. Based on existing data, the aim is to use machine learning algorithms to develop models for predicting used car prices.

Solution:

* Deciding whether a used car is worth the posted price when you see listings online can be difficult.
* There is a need for a used car price prediction system to effectively determine the worthiness of the car using a variety of features.
* The focus of this project is developing machine learning models that can accurately predict the price of a used car based on its features, in order to make informed purchases.
* The Main Aim is to use machine Learning Algorithms to develop models for predicting used car prices

Machine Learning Pipeline:

Step 1:-Data Collection

-we are using the Dataset on used car sales from all over the United States from Truecar.com and we used true car listings data.

-The features available in this dataset are Mileage,Year,City,State,Vin,Make and Model .

-We need to Predict Price for the used car,So we can clearly concluded that it is a Regression Problem.

Step 2:-Data Preprocessing

In this step, we are going to gather and merge data into a single framework so it would be easy for us to work with it.

Step 3:- Data Cleaning

In this step we have used Box Plot, Distribution plot,Pie Chart,Scatter Plot and Heat map

to clean our data like if our data contains outliers,missing values,duplicate data,Categorical Encoding and other errors otherwise it will affect our prediction.

-Here,we found outliers for Price ,Mileage by Box plot and remove the outliers from price more than 45000 and for Mileage more than 150000 are removed and found using Z-Score.

Step 4:- Feature Engineering

Label Encoding: We converted the Make, Model, and City into labelled vectors.

Multicollinearity: To Check the Multicollinearity we used VIF and found Year is highly Multicollinear so we removed it.

Feature Scaling: Performed Standardisation to scale the data provided.

Step 5:- Model Selection

- In model selection, we performed different Machine Learning models like Linear Regression,Gradient Boosting,Random Forest Regressor,Linear Regression with KMeans ,XGBoost,Light gradient boosting.

- We had evaluate Performance of our model by accuracy and error.

- We had also evaluate features from the data that is which feature is important for the prediction or which feature is affecting our model performance by performing Feature Importances using Random Forest.

- We had Perform hyperparameter tuning for better performance of the model, then we had test our model on the test data.

-Conclusion:We can conclude that top three models are Random Forest Regressor with accuracy of 84.8% ,Extreme Gradient Boosting Regressor with accuracy of 78.2% and Light Gradient Boosting Regressor with accuracy of 77.9%.

Step 6:- Model Deployment

- For deployment we had pickle our model for uploading at cloud storage for production.

- We have deploy our model using Flask on Heroku.

Project Link : <https://used--car-price-prediction.herokuapp.com/>