



Car Price Prediction

Project Title: *Building a Car Price Prediction Model – A Machine Learning Approach*

Overview

In the dynamic world of automobile sales, predicting the right price of a used car is both a business necessity and a data-driven challenge. In this project, you will **build a Car Price Prediction model using Linear Regression**, aimed at forecasting car prices based on various real-world attributes such as brand, age, mileage, fuel type, and more.

This project simulates a real industry use-case and is designed to help you apply machine learning techniques in a practical and meaningful way—making your learning portfolio strong and interview-ready.

Objective

The goal is to develop a machine learning model that can accurately estimate the selling price of used cars. You will handle data preprocessing, perform exploratory analysis, train a predictive model, and evaluate its accuracy using statistical metrics.

Dataset

A dataset containing information about used cars will be **shared in the group**. It includes:

- Car Brand & Model
- Year of Manufacture
- Mileage
- Fuel Type
- Transmission
- Owner Type

- Location
 - Selling Price (Target Variable)
-

Key Tasks

1. Data Understanding and Exploration

- Load and inspect the dataset.
- Handle missing values and check data types.
- Identify correlations and data distributions.

2. Exploratory Data Analysis (EDA)

- Visualize how features affect the car price.
- Plot heatmaps, histograms, and boxplots to gain insights.

3. Data Preprocessing

- Encode categorical variables using Label or OneHot Encoding.
- Normalize or scale numerical features.
- Split data into training and testing sets (e.g., 80:20).

4. Model Development

- Train a **Linear Regression** model on the training data.
- Predict car prices on the test data.

5. Model Evaluation

Assess the model using:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- R^2 Score (Goodness of Fit)

6. Model Interpretation

- Identify the most impactful features.
- Discuss the model's strengths and limitations.
- Explain potential underfitting or overfitting issues.

What to Submit

1. **Source Code:** Jupyter Notebook or Python script with proper comments and outputs.
2. **Short Report (PDF):** Summary of your work (1–2 pages) including EDA insights, model results, and conclusion.

Bonus (Optional)

- Perform feature selection or remove outliers to improve model performance.

