Car Price Prediction - A Machine Learning Approach

Objective:

To build a Linear Regression model that accurately estimates the selling price of used cars using features such as brand, manufacturing year, mileage, fuel type, transmission, and owner type.

1. Dataset Summary:

The dataset contains 4340 rows and 8 columns, with no missing values.

Features:

- name: Car brand and model
- year: Year of manufacture
- selling_price: Target variable
- km_driven: Total kilometers driven
- fuel: Fuel type (Petrol/Diesel/CNG/LPG/Electric)
- seller_type: Individual or Dealer
- transmission: Manual or Automatic
- owner: Ownership status (First, Second, etc.)

2. Exploratory Data Analysis (EDA):

- Heatmap: Shows correlations between year, km_driven, and selling_price.
- Histogram: Selling price distribution is right-skewed; most cars are priced under Rs.10 lakhs.
- Boxplot: Fuel type and transmission affect the price range significantly.

3. Data Preprocessing:

- Label Encoding applied to categorical variables.
- StandardScaler used for feature scaling.

- Data split: 80% training and 20% testing.

4. Model Development:

- Linear Regression model trained using sklearn.

- Target: Predicting car selling price based on available features.

5. Model Evaluation (Actual Results):

- MAE: Rs.2,21,821

- MSE: Rs.1.84×10^{11}

- RMSE: Rs.4,29,339

- R^2 Score: 0.396 (moderate accuracy)

6. Model Interpretation:

- Most impactful features: year, km_driven, fuel, transmission.

- Model shows slight underfitting - complex models like XGBoost or Random Forest may improve accuracy.

7. Conclusion:

The Linear Regression model provides a simple yet interpretable prediction of car prices.