Car Price Prediction Project

1. Introduction

Project Objective: I developed a project called 'Car Price Predictor' using Linear Regression to determine the significant factors affecting car prices in the US market and to predict car prices based on those factors. This project was based on a dataset provided by an automobile consulting company.

Motivation: The insights from this model can help car manufacturers adjust designs and strategies to compete effectively in the US market.

2. Dataset

Source and Description: I worked with a dataset containing details of various cars in the American market, including variables such as car dimensions, engine specifications, and fuel efficiency. The target variable was 'price'.

Data Exploration: I explored the dataset to understand its structure and performed preprocessing like handling categorical variables, encoding them, and replacing textual data with numerical values for seamless modeling.

3. Data Preprocessing

Cleaning and Transformation: I checked for missing values, duplicates, and outliers. While outliers were present in the price column, I chose not to remove them as they might represent premium or luxury cars.

Feature Encoding: Categorical features like 'fuel type' and 'drive wheel' were label-encoded to convert them into numerical format.

4. Feature Selection

Variable Selection: Using scatter plots and a heatmap, I identified 12 key variables significantly correlated with car price, such as 'drivewheel', 'wheelbase', 'carwidth', and 'horsepower'. Features

with strong correlations were retained for the model.

Rationale: This step ensured the model focused on impactful variables, improving prediction accuracy.

5. Model Building

Choice of Algorithm: I used Linear Regression because the objective was to model the relationship between multiple independent variables and the target variable, 'price'.

Splitting Data: The dataset was split into independent variables (like engine size and curb weight) and the dependent variable (price).

6. Results and Evaluation

Model Training and Prediction: After training the model, I tested it by predicting the price of a car with given specifications. For example, I input values like engine size and city mileage, and the model predicted a price close to the actual value.

Accuracy: The model had a good score, indicating its reliability in predicting car prices based on the provided features.

7. Impact and Future Scope

Business Utility: The model provides actionable insights for car manufacturers to strategize better pricing and design in the competitive US market.

Improvements: In the future, I plan to enhance the model by incorporating more features, using advanced algorithms like Random Forests or XGBoost, and performing hyperparameter tuning for better accuracy.

8. Technical Tools Used

I used Python along with libraries like Pandas and NumPy for data preprocessing, Seaborn and Matplotlib for visualization, and Scikit-learn for building and evaluating the model.

Closing

This project showcased my ability to preprocess data, perform exploratory data analysis, build and evaluate predictive models, and draw actionable insights, all of which are crucial for real-world machine learning applications.