Anmol Kashyap
2511AI44

Car Price Prediction Using UCI Automobile Dataset

1. Introduction

This project aims to predict car prices using machine learning models trained on the UCI Automobile dataset. Predicting car prices is important for buyers, sellers, and dealers to make informed decisions.

2. Dataset Description

The dataset is sourced from the UCI Machine Learning Repository. It includes technical specifications such as engine size, horsepower, fuel type, body style, and other attributes. The target variable is car price.

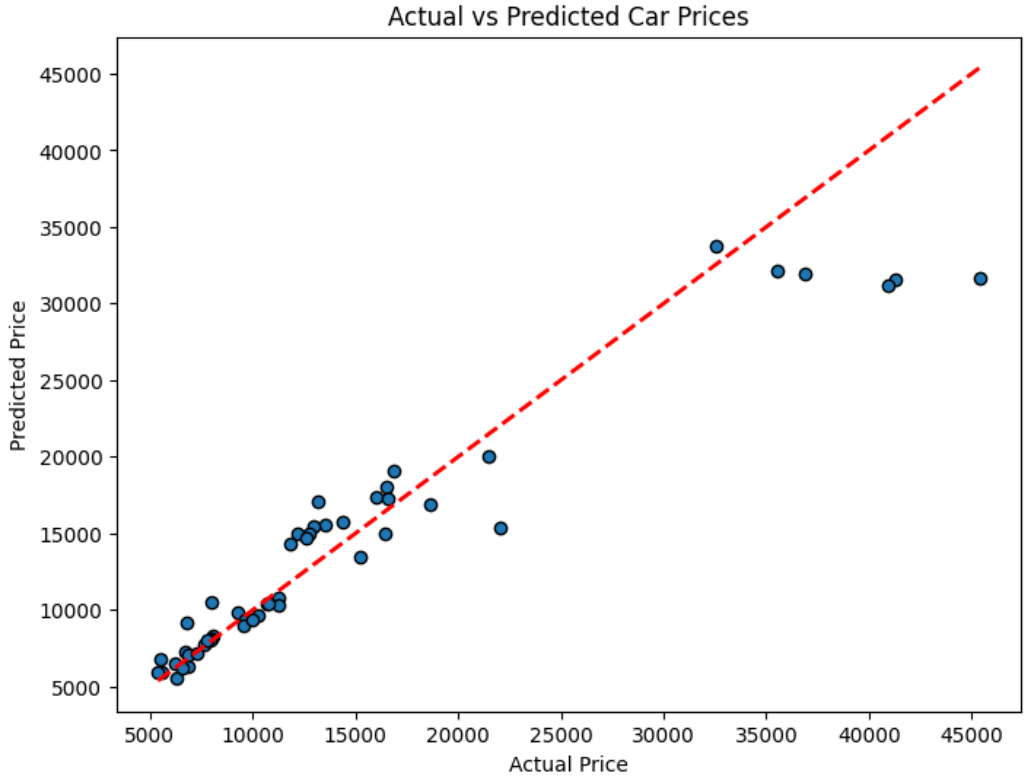
3. Methodology

Data preprocessing steps include handling missing values, converting numeric features, and encoding categorical variables. A Random Forest Regressor is used as the prediction model, combined with OneHotEncoding and StandardScaler in a pipeline. The dataset is split into training and testing sets in a 75/25 ratio.

4. Results

The model was evaluated using R^2 score and Mean Absolute Error (MAE). The scatter plot below shows actual vs predicted prices.

```
R2: 0.8894726660622242
MAE: 1963.9643027210886
```



5. Conclusion

The Random Forest model performs reasonably well in predicting car prices. Future improvements can include hyperparameter tuning and testing other regression models like Gradient Boosting or Neural Networks.