Predicting Car Prices with Linear Regression

NAME - VEDANT KENKRE

CSEP PROJECT

BRANCH - CSD

ROLL NO - 23CD3050

This presentation will guide you through the process of building a linear regression model to predict actual car prices based on various features. We'll cover data preparation, feature engineering, model training, and evaluation, culminating in insights and next steps.

Data Preparation and Cleaning

Data Collection

We gathered a comprehensive dataset of car models, including details like make, model, year, mileage, and actual sales prices.

Data Cleaning

We identified and addressed any missing values, inconsistencies, or outliers in the data to ensure a high-quality dataset for modeling.

Exploratory Analysis

We conducted initial data exploration to uncover trends, relationships, and potential predictor variables for the linear regression model.

Feature Engineering

Feature Selection

1

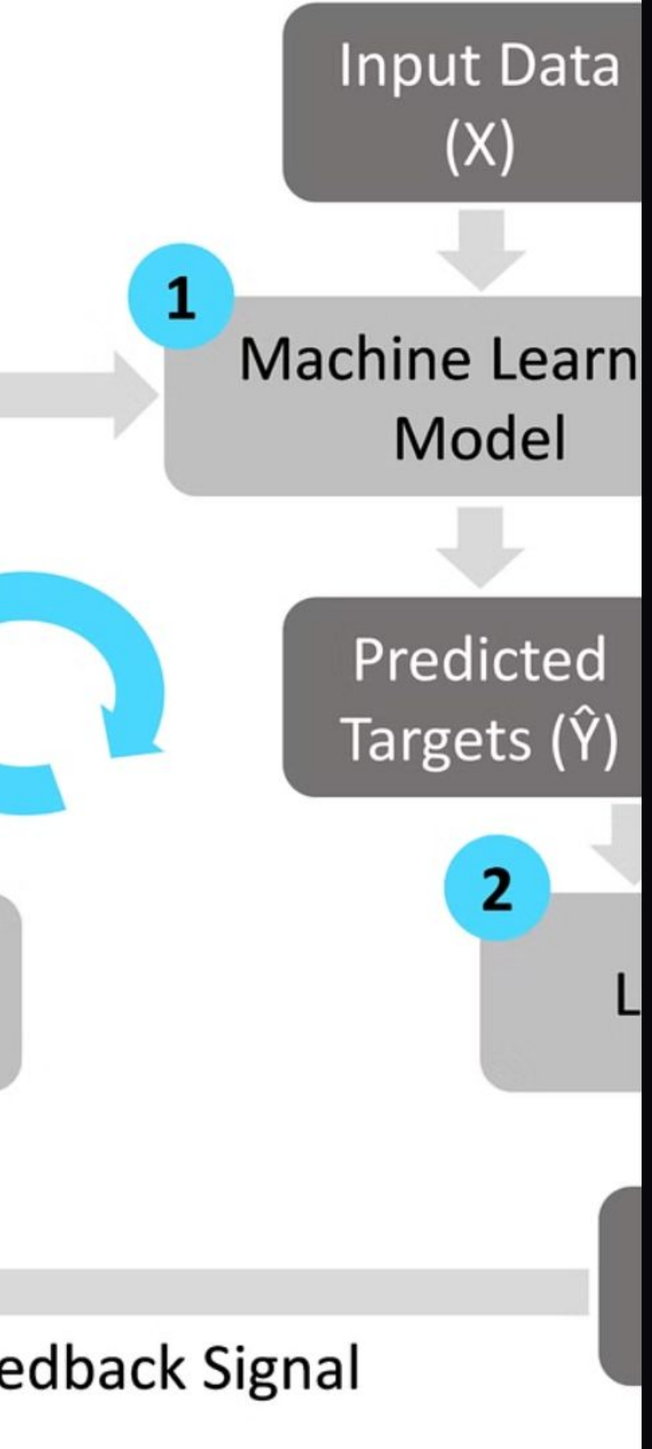
We carefully selected the most influential features, such as make, model, year, mileage, and engine size, to include in the model.

2 Feature Transformation

We transformed certain features, like mileage and engine size, to better capture their nonlinear relationships with the target variable.

3 Feature Encoding

We encoded categorical features, like make and model, to numerical values that the linear regression model can understand.



Model Training

1 **Algorithm Selection**

We chose linear regression as the appropriate algorithm for this predictive modeling task.

2 **Hyperparameter Tuning**

We experimented with different hyperparameter settings to optimize the model's performance.

3 **Model Fitting**

We trained the linear regression model on the prepared dataset, allowing it to learn the relationships between the features and the target variable.

Model Evaluation

R-Squared

We calculated the R-squared value to measure the proportion of the variance in the target variable that is predictable from the model.

Mean Absolute Error

We computed the mean absolute error to quantify the average absolute difference between the predicted and actual prices.

Cross-Validation

We performed cross-validation to ensure the model's robustness and generalizability to new, unseen data.

Actual Price vs Predicted Price Comparison

Scatter Plot

We created a scatter plot to visualize the relationship between the actual and predicted car prices, highlighting the model's accuracy.

Residual Analysis

We analyzed the residuals, or the differences between the actual and predicted values, to identify any patterns or assumptions violations.

Outlier Identification

We identified any outliers, or data points that deviate significantly from the overall pattern, to understand the model's limitations.

Insights and Findings

Affordability

The model can help identify cars that are underpriced or overpriced relative to their features, assisting customers in making informed purchasing decisions.

Profit Optimization

Dealers can use the model to price their inventory more effectively, maximizing profits while remaining competitive in the market.

Market Trends

By analyzing the model's predictions, we can uncover emerging trends in the car market, such as the impact of new technologies or consumer preferences.

Conclusion and Next Steps

1

Deployment

The next step is to deploy the linear regression model in a production environment, where it can be used to generate accurate price predictions for new car listings.

2

Continuous Improvement

We will continue to monitor the model's performance and refine it as new data becomes available, ensuring it remains up-to-date and accurate.

3

Expanding Capabilities

In the future, we may explore integrating the model with other data sources or incorporating additional features to further enhance its predictive power.