Predicting Old Car Prices
A Machine Learning Approach

By – Pranita das

Capstone project – sprint 1

Brain station

Project Overview

Introduction to the problem statement of predicting old car prices using machine learning techniques.

Opportunity Identified: Highlight the significance of accurate price prediction for both buyers and sellers in the used car market.

Vision for Tackling the Problem

Proposed Approach:

Utilizing Supervised Learning: planning to employ supervised learning techniques since the dataset contains labeled data, with 'selling_price' being the target variable.

Feature Engineering: conducting feature engineering to extract relevant information from the existing features and create new features, such as calculating the age of the car from the 'year' feature.

Handling Categorical Variables: Since the dataset includes categorical variables like 'fuel,' 'seller_type,' 'transmission,' and 'owner,' encode them appropriately using techniques like one-hot encoding or label encoding.

Introduction to the Dataset

- Exploring the Old Car Dataset – My dataset name is car_data.csv

Size of Dataset: The dataset comprises 4340 entries and 8 attributes.

Conclusion: With 4340 rows and 8 columns, the dataset provides a comprehensive foundation for analyzing old car prices and exploring various factors influencing selling prices.

Overview of Dataset

Name

Year

Selling Price

Kilometers Driven

Fuel

Seller Type

Transmission

Owner

Conclusion: This table provides an overview of the attributes present in the dataset, facilitating a better understanding of the information available for analysis.

Potential Impact of Predicting Old Car Prices

- Impact on the Automotive Market

Buyers: Empowered decision-making, better negotiation, and confidence in investment.

Sellers: Competitive pricing, reduced time to sell, and maximized returns on sales.

Market Transparency: Enhanced efficiency, trust-building, and smoother transactions.

Moving Forward with Model Development

Data Preprocessing: Check each column for missing values using methods like isnull().

Feature Engineering: Calculate the age of the car by subtracting the 'year' feature from the current year.

Visualization of Data Distribution: Plot histograms or density plots to visualize the distribution of numerical features such as 'year', 'selling_price', and 'km_driven'.

Used Seaborn or Matplotlib for plotting.

Visualize Distribution of Categorical Features: Plot count plots to visualize the distribution of categorical features such as 'fuel', 'seller_type', 'transmission', and 'owner'.

Baseline Modeling: This will include code snippets demonstrating the creation of baseline models, splitting the data, training the models, and evaluating their performance.

In conclusion, these data preprocessing, feature engineering, baseline modeling, and visualization techniques provide a solid foundation for the machine learning project on predicting old car prices. By carefully preparing the dataset, extracting relevant features, and building baseline models, I have laid the groundwork for further analysis and model refinement.

THANK YOU