Oblig ML DAT158 HT2024

Members: Leonard Heldal (669778), Severin Johannessen (669799), Iver Thoresen Malme(669820)

#Gruppe: 21

Project Name: Car Price Prediction

Introduction

The purpose of this project is to develop a car price estimation service that provides users with a reliable estimate of a car’s current market value, requiring no login or sharing of sensitive information. The target users are individuals and car dealers.

Business Objectives

1. * **Optimize Pricing**: Give a realistic price estimate to dealers and sellers.
2. * **Support Investment Decisions**: Help car dealers make good decisions when buying or selling a car.
3. * **Simplify the Buying and Selling Process**: A reliable price estimation.

Business Impact

- * **Streamlined Decision-Making**: Accurate price estimates offer dealers a market reflected pricing, which then speeds up the process of buying a car..
- * **Enhanced Market Positioning**: Accurate price analysis.

Comparison with Existing Solutions

The solutions we have today often require a long registration process and seem to only provide a general estimate. Our service will deliver price estimates based on car attributes.

Manual Method for Price Estimation:

Machine learning helps us with price estimation which would normally require manually comparing cars with historical sale prices, which then could be very time consuming.

Machine Learning and Software Metrics

1. * **Root Mean Squared Error (RMSE)**: Root of the average of squared differences between predicted and actual car prices.
2. * **Latency**:The response time from user input to the return of estimate.
3. * **Throughput**: Number of price estimates per minute.

Stakeholders

1. * **Customers**: Individuals and car dealers.
2. * **Car Dealers**: Car dealers who need a service for a more accurate price.

Resources

- * **Personell:**
  Developers.
- * **Computational Resources**:
  Computers for development and cloud resources if needed.
- * **Data Resources**:
  Historical car prices and market trends.

Data

This project uses training and test datasets. To ensure consistency and data quality, cross-validation is applied.

Data Preprocessing:

- * **Missing Data Handling**: Filled missing values in fuel_type, accident, and clean_title with default values. For numeric columns, iterative and simple imputers were used.
- * **Feature Engineering**: Horsepower, displacement, engine_type, and cylinders from the enginecolumn and encoded categorical variables, such as brand and fuel_type, used Label Encoding.
- * **Scaling**: Standardized continuous variables for better performance.

Modeling

Exploratory Data Analysis:
A heatmap and price distribution plots were used to get a better view of feature relationships. Correlations between model year, milage, and variables like accident and price indicated connections between vehicle age, mileage, and price.

Model Selection and Optimization:
Random Forest Regressor was chosen as the primary model due to its performance with structured data. The optimized parameters achieved an RMSE of 66,035.47.

Label Encoding:
Random Forest models are effective with label encoded data, which simplified the data representation without slowing our performance.

Deployment

The model will be deployed via Gradio, providing a user-friendly interface. The goal is to maintain a fast response time.

Summary

This car price estimation project utilized Random Forest Regressor as the model, achieving an RMSE of 66,035.47. Label Encoding was preferred over One-Hot Encoding for easier readability. The car price estimation service offers reliable price estimates for private users and dealers. The final deployment through Gradio offers a user-friendly experience with minimal latency.

References

We used GitHub, Scikit-learn, gradio, Pandas and python to build the program.