Justification for Method of Approach and Model Selection

# Approach Overview and Model Rationale

This document provides a detailed justification for the approach and model selection in the project. The chosen methodology and models are explained based on the dataset characteristics, performance metrics, and predictive capabilities.

## Method of Approach

The approach began with cleaning the dataset to ensure all features were relevant and without missing values. Exploratory Data Analysis (EDA) was conducted to identify trends and relationships within the data. Based on the insights gained, the following preprocessing steps were applied:  
• Handling missing data  
• Scaling numerical features  
• One-hot encoding for categorical variables  
• Label encoding for ordinal features

## Model Selection Process

After preprocessing the data, various regression models were considered to predict car prices accurately. The models selected included Linear Regression, Lasso Regression, Ridge Regression, Decision Tree, Random Forest, and Gradient Boosting.

## Justification for Model Selection

Gradient Boosting was selected for its excellent predictive performance, with a Test R² score of 0.9601, and low error rates. It proved to be robust to overfitting and offered the best generalization for this dataset.

## Model Comparison - Bar Plot

The bar plot below provides a comparison of MSE, MAE, and R² scores across different models:



## Conclusion

In conclusion, the Gradient Boosting model was selected for its superior performance in various metrics, making it the most suitable model for predicting car prices in this dataset.