**Model Comparison and Analysis Report**

**Project: Used Car Price Prediction - Model Evaluation**

**Objective:**

To train and evaluate 5 different regression models for predicting the price of used cars, and select the best-performing model based on standard evaluation metrics.

**Models Compared:**

| **Sl. No** | **Model** | **MAE** | **MSE** | **R² Score** |
| --- | --- | --- | --- | --- |
| 1 | Linear Regression | 0.036464 | 0.007758 | -0.054625 |
| 2 | Ridge Regression | 0.034139 | 0.007329 | 0.003747 |
| 3 | Decision Tree | 0.010563 | 0.007750 | -0.053447 |
| 4 | Random Forest | 0.011192 | 0.005563 | 0.243742 |
| **5** | **Gradient Boosting** | **0.013321** | **0.005476** | **0.255582** |

**Interpretation of Metrics:**

* **MAE (Mean Absolute Error):** Lower is better. Measures average magnitude of errors in predictions.
* **MSE (Mean Squared Error):** Lower is better. Punishes larger errors more severely.
* **R² Score:** Higher is better. Indicates how well predictions approximate actual values. (Closer to 1 = better)

**Model Performance Summary:**

* **Gradient Boosting** achieved the highest R² Score (0.255582) among all models.
* **Random Forest** was close but slightly behind.
* **Linear Regression**, **Ridge Regression**, and **Decision Tree** showed significantly lower performance.
* **Linear models** struggled because the dataset likely has **non-linear relationships** between features and car prices.

**Final Best Model Selected:**

* **Gradient Boosting Regressor**
* Based on:
  + Highest R² Score
  + Good balance of low MAE and low MSE
  + More stable and generalizable than Decision Trees or Random Forest alone

**Saved Artifacts:**

| **File** | **Purpose** |
| --- | --- |
| best\_model.pkl | Best trained Gradient Boosting model |
| imputer.pkl | Preprocessing imputer for numerical features |
| scaler.pkl | Preprocessing scaler for numerical features |
| price\_range.pkl | Saved min and max values for de-normalizing price predictions |

**Conclusion:**

Based on systematic evaluation of 5 regression models on the used car dataset,  
**Gradient Boosting** was found to be the most suitable model for deployment.  
It demonstrated better prediction accuracy, stability, and generalization across the dataset.

This model will now be integrated into the final Streamlit web application for real-time car price predictions.