**Raw scale (₹)²**

**Model Validation Results Summary**

| **Model** | **MAE (↓)** | **MSE (↓)** | **R² Score (↑)** | **Verdict** |
| --- | --- | --- | --- | --- |
| Linear Regression | **2407.71** | **66,994,303.21** | 0.084 | Baseline, weak performance |
| Ridge Regression | 2404.66 | 67,006,417.40 | 0.084 | Same as Linear (adds L2 penalty) |
| Lasso Regression | 2403.19 | 67,006,848.19 | 0.084 | Same as Ridge, confirms linear limitations |
| Decision Tree | 1156.92 | **81,105,428.75** | -0.109 | Overfitting (low R² despite low MAE) |
| **Random Forest** | **1180.54** | **58,394,834.23** | **0.201** | **Best performance overall** |
| Gradient Boosting | 1295.72 | 59,886,781.41 | 0.181 | Second best, more stable than Decision Tree |

**Observations:**

* **Linear, Ridge, and Lasso** performed nearly identically, which is expected because they are all linear models and the relationships in your dataset are likely non-linear.
* **Decision Tree** has the lowest MAE but **negative R²**, which indicates it overfit the training data and failed to generalize.
* **Random Forest** gives the **best trade-off** — strong generalization and lowest MSE.
* **Gradient Boosting** is a close second, slightly less accurate but can be further optimized via tuning.

**Normalized Model Evaluation (All Metrics Scaled to 0–1)**

| **Model** | **MAE (Normalized)** | **MSE (Normalized)** | **R² Score** | **Verdict** |
| --- | --- | --- | --- | --- |
| **Random Forest** | **0.0117** | **0.0058** | **0.2184** | Best overall |
| Gradient Boosting | 0.0130 | 0.0061 | 0.1762 | 2nd best |
| Linear Regression | 0.0241 | 0.0067 | 0.0840 | Weak baseline |
| Ridge Regression | 0.0241 | 0.0067 | 0.0837 | Nearly same as Linear |
| Lasso Regression | 0.0184 | 0.0074 | -0.0004 | Not performing well |
| Decision Tree | 0.0118 | 0.0085 | -0.1495 | Overfit (R² < 0) |

**Interpretation:**

* **Random Forest** still performs the best:
  + Lowest normalized MAE and MSE
  + Best R² Score
  + Great generalization for non-linear data
* **Gradient Boosting** is close and more stable in noisy data, so a good alternative
* **Decision Tree** appears great by MAE but fails on R² → it's **overfitting** badly

**Best Model Selected: Random Forest Regressor**

**Why Random Forest?**

* **Lowest MAE** and **lowest MSE** among all models
* **Highest R² score**, indicating better ability to explain price variability
* Robust to overfitting and missing data (due to averaging multiple trees)
* Non-linear model — better suited for structured datasets like used cars