

Project Summary Report

Manish Bhardwaj (2315510116)

Car Price Prediction Project

Building a Car Price Prediction Model – A Machine Learning Approach

Objective

The objective of this project was to build a predictive machine learning model that estimates the selling price of used cars based on various attributes, such as car brand, manufacturing year, mileage, fuel type, transmission type, owner type, and location.

Data Exploration and Understanding

- The dataset consists of **4,340 entries**, including columns like car name, year, km driven, fuel type, seller type, transmission, owner type, and selling price.
- Added new features:
 - **Mileage**: estimated based on car age ($(2025 - \text{year}) * 10,000$ km).
 - **Location**: randomly assigned among Delhi, Mumbai, Bangalore, Hyderabad, and Chennai.
- No major missing values were present.
- Numeric and categorical distributions were analyzed.

EDA Insights

- Newer cars (recent years) are generally priced higher.
- Cars with lower mileage (younger age) tend to have higher resale value.
- Diesel and petrol vehicles are more prevalent and show wider price ranges.
- Automatic transmission cars are priced higher than manual.
- Cars with fewer previous owners (especially first-owner) have higher selling prices.
- Cities like Mumbai and Delhi show slightly higher average car prices.

Data Preprocessing

- Categorical variables were encoded using **Label Encoding**.
- Numerical features were **standardized** using StandardScaler.
- Dataset was split into 80% training and 20% testing sets.

Model Development

- Linear Regression model was used to predict selling prices.
- Model trained on scaled features and evaluated on the test set.

Model Evaluation Metrics

Metric	Value
MAE	221,709.74 ₹
MSE	184,225,713,662 ₹²
RMSE	429,215.23 ₹
R² Score	0.396

The R² score of 0.396 suggests that the model explains about 40% of the variance in selling price.

Feature Importance (Coefficient Analysis)

- **Most positive impact:**
 - Year: +78,311 ₹
 - Location: +3,724 ₹
- **Most negative impact:**
 - Transmission: -275,773 ₹ (automatic cars generally priced higher; negative coefficient indicates opposite in encoded data)
 - Fuel: -144,528 ₹
 - Mileage: -78,311 ₹
 - km_driven: -47,345 ₹

Conclusion

The Linear Regression model was able to capture broad price trends in used cars, but its relatively moderate R² score suggests that further improvements are possible.

Future enhancement ideas:

- Remove or treat outliers in features like km driven and selling price.
 - Use advanced regression models (e.g., Ridge, Lasso, Random Forest Regressor).
 - Incorporate additional features such as service history, accident record, or car segment.
-