Used Car Price Prediction

1. Objectives

The primary goal of this project was to build a predictive model that can accurately estimate the price of a used car based on several features. The specific objectives included:

- **Understanding the factors** that influence the resale value of used cars.
- **Collecting and preprocessing data** related to used cars, including attributes such as brand, model, year, mileage, fuel type, transmission, and ownership.
- **Developing and evaluating** multiple machine learning models to identify the best performer in predicting car prices.
- **Deploying the model** in a usable format for potential users such as car dealerships, resale platforms, and individual buyers/sellers.

---

2. Key Learnings

a) Data Analysis & Feature Engineering

- **Mileage, age of the car, and brand** were found to be major influencers on price.
- **Luxury brands and newer models** tend to retain more value.
- Cars with **manual transmission or multiple owners** often showed lower resale prices.
- **Data preprocessing** steps such as handling missing values, encoding categorical variables, and normalizing numerical values were crucial for model performance.

b) Modeling Techniques

- Several models were tested, including:
    - **Linear Regression** – Simple, interpretable, but underperformed on non-linear relationships.
    - **Decision Tree & Random Forest** – Captured complex patterns and non-linearity better.
    - **Gradient Boosting (XGBoost)** – Provided the most accurate predictions with good generalization.
- **Model evaluation** was done using metrics such as RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), and R^2 Score.

c) Challenges Faced

- **Data imbalance** with certain car brands dominating the dataset.

- **Overfitting** in complex models like Random Forest and XGBoost was mitigated using hyperparameter tuning and cross-validation.

---

3. Conclusion

The project successfully demonstrated the feasibility of predicting used car prices using machine learning techniques. The most effective model, **XGBoost**, achieved a high level of accuracy and generalization. The project highlighted the importance of quality data, thoughtful feature selection, and model tuning.

Implications:

- Car resale platforms can use the model to provide fair price suggestions.
- Buyers and sellers can make informed decisions based on objective data-driven estimates.
- Dealerships can optimize their inventory pricing strategy.
