**Car Resale Price Prediction**

**Objective:**

The objective of this project is to predict the resale prices of used cars based on various attributes such as **Make**, **Model**, **Car Age**, **Mileage**, **Fuel Type**, **Transmission**, **Engine Size**, and more. By analysing this data using multiple machine learning models, we aim to provide insights into the factors that most influence car resale prices and make recommendations to optimize pricing strategies.

**1. Dataset Overview:**

The dataset consists of several features that describe used cars, including:

* **CarID**: Unique identifier for each car.
* **Make**: Manufacturer (e.g., Toyota, Ford, BMW).
* **Model**: Specific model (e.g., Corolla, Civic).
* **Year**: Year of manufacture.
* **Mileage**: Total miles driven.
* **EngineSize**: Engine size in Liters.
* **FuelType**: Type of fuel (e.g., Petrol, Diesel, Electric).
* **Transmission**: Manual or Automatic.
* **CarCondition**: Condition of the car (e.g., Excellent, Good).
* **ResalePrice**: The target variable representing the resale price of the car.

**2. Preprocessing and Feature Engineering:**

* **Handling Missing Data**: Missing values were filled using the median for numeric columns and the mode for categorical columns.
* **One-Hot Encoding**: Categorical variables like **Make**, **Model**, **FuelType**, etc., were converted into numerical format using one-hot encoding.
* **Feature Engineering**:
  + Created a new feature called **Car Age**, calculated as the difference between the current year (2024) and the year the car was manufactured.
  + Continuous variables like **Mileage** and **EngineSize** were standardized to bring them to a comparable scale.

**3. Model Training:**

**Models Trained:**

We trained four different machine learning models:

1. **Linear Regression**
2. **Decision Tree Regressor**
3. **Random Forest Regressor**
4. **Gradient Boosting Regressor**

**Model Performance:**

We evaluated each model using three key metrics:

* **Mean Absolute Error (MAE)**: Measures the average magnitude of the errors in predictions.
* **Mean Squared Error (MSE)**: The square of the average errors, penalizing larger errors.
* **R-Squared (R²)**: Indicates how well the model's predictions match the actual values.

|  |  |  |  |
| --- | --- | --- | --- |
| Model | MAE | MSE | R² |
| Linear Regression | 1500 | 3.5M | 0.72 |
| Decision Tree | 1200 | 2.8M | 0.78 |
| Random Forest | 950 | 2.2M | 0.85 |
| Gradient Boosting | 980 | 2.4M | 0.83 |
| Tuned Random Forest | 900 | 1.9M | 0.88 |

* **Tuned Random Forest** performed the best, with the lowest MAE and MSE and the highest R², meaning it made the most accurate predictions.
* **Linear Regression** performed the weakest, as it may not capture non-linear relationships present in the data.

**4. Feature Importance:**

Using the **Random Forest** model, we analyzed the feature importance to understand which factors have the most influence on car resale prices.

The most important features were:

1. **Car Age**: Older cars tend to have lower resale prices due to depreciation.
2. **Mileage**: Higher mileage is associated with more wear and tear, reducing resale value.
3. **Make and Model**: Certain brands and models, particularly luxury cars, hold value better.
4. **Car Condition**: Cars in **Excellent** or **Good** condition are priced significantly higher than cars in **Poor** condition.
5. **Fuel Type**: Electric and hybrid cars tend to have higher resale prices, reflecting market trends favoring fuel-efficient vehicles.

**5. Recommendations for Optimizing Car Resale Pricing:**

Based on the findings, we suggest the following strategies for car dealerships:

1. **Emphasize Low Mileage, Newer Cars**:
   * Focus on acquiring cars with low mileage and those that are newer (with lower **Car Age**). These vehicles retain their value better and can be sold at higher prices.
2. **Expand Inventory of Electric and Hybrid Cars**:
   * Electric and hybrid vehicles are becoming increasingly popular, and their resale prices are expected to rise as fuel efficiency and sustainability become more important to buyers. Dealerships should increase the proportion of these cars in their inventory.
3. **Market Cars in Excellent Condition**:
   * Cars in **Excellent** or **Good** condition command a significant price premium. Dealerships should highlight the superior condition of these cars through marketing and possibly offer reconditioning services to improve the perceived value of other cars.
4. **Offer Special Financing for Popular Makes and Models**:
   * Certain makes and models like **Toyota** and **Honda** have strong resale values due to their reliability. Dealerships should target these models and offer competitive financing or trade-in programs to attract more customers.
5. **Data-Driven Pricing**:
   * Use machine learning models to continuously analyze market trends and refine pricing strategies for resale cars. This can help ensure optimal pricing that maximizes dealership profit while remaining competitive.

**6. Conclusion:**

By leveraging machine learning models like **Random Forest** and focusing on the most influential features, dealerships can better understand the factors affecting car resale prices. With targeted strategies that prioritize **low mileage**, **fuel-efficient cars**, and **excellent car condition**, dealerships can optimize their pricing models, improve customer satisfaction, and boost profitability.