# PROJECT: 2

# Vehicle Price Prediction

**Name:** Mani P

**Email:** maniprabu991@gmail.com

**Program:** Unified Mentor ML Internship Program

## 1. Problem Statement

Determining a fair market price for vehicles is challenging because prices depend on multiple factors such as brand, model year, mileage, fuel type, transmission, and drivetrain. Manual pricing methods are often inaccurate and inconsistent. This project aims to solve this problem by developing a machine learning model that predicts vehicle prices using historical data and vehicle specifications.

## 2. Objective

The main objective of this project is to build a reliable machine learning system that can accurately predict vehicle prices based on vehicle features such as make, model, year, mileage, fuel type, transmission, body type, and drivetrain.

## 3. Tech Stack Used

- **Programming Language:** Python
- **Libraries:** Pandas, NumPy, Scikit-learn
- **Machine Learning Algorithm:** Random Forest Regressor
- **Model Optimization:** GridSearchCV
- **Feature Engineering:**
  - One-Hot Encoding (low-cardinality features)
  - Target Encoding (high-cardinality features)

## 4. Project Architecture / Workflow

1. Data collection from vehicle dataset
2. Data cleaning and preprocessing
3. Handling missing values
4. Feature encoding
5. Model training
6. Hyperparameter tuning
7. Model evaluation
8. Vehicle price prediction

# 5. Implementation Details

- Removed irrelevant columns such as **name**, **description**, and **engine**
- Filled missing numerical values using **median**
- Filled missing categorical values using **mode**
- Applied **hybrid encoding** to prevent feature explosion
- Trained a **Random Forest Regressor** using optimized hyperparameters from GridSearchCV

**Best Model Parameters:**

- n_estimators: 200
- max_depth: None
- max_features: sqrt
- min_samples_split: 2
- min_samples_leaf: 1

# 6. Output / Results

**Evaluation Metrics:**

- $R^2$ Score: **0.786**
- Mean Absolute Error (MAE): **$4,061**
- Root Mean Squared Error (RMSE): **$7,498**

**Sample Prediction:**
Vehicle: *2021 Toyota Camry LE*
Mileage: 15,000 miles
Fuel Type: Gasoline
Transmission: Automatic
Body Type: Sedan
Drivetrain: FWD

**Predicted Price:** $42,310.78

# 7. Challenges Faced

- Handling high-cardinality categorical features

- Preventing dimensionality explosion during encoding
- Selecting optimal hyperparameters for better accuracy

## 8. Future Enhancements

- Use advanced models like **XGBoost** or **LightGBM**
- Add **SHAP** for feature importance and explainability
- Deploy the model using **Streamlit** or **Flask API**
- Integrate real-time vehicle data

## 9. Conclusion

This project successfully demonstrates how machine learning can be applied to predict vehicle prices effectively. The use of Random Forest and hybrid encoding improved accuracy and scalability. The system can be extended further for real-time and production-level deployment.