# Car Price Prediction

This project predicts the selling price of used cars using Machine Learning.

## Project Overview

This project is to estimate a car's Selling Price based on features such as brand, fuel type, transmission, owner type, and car age. Both Linear Regression and Random Forest Regressor models are trained, evaluated, and saved for later use.

## Project Structure

- car_data.csv → Original dataset
- car_data_cleaned.csv → Cleaned dataset (Year converted to Car_Age)
- car_data_encoded.csv → Encoded dataset for ML models
- car_data_new.csv → New dataset for price prediction (no Selling_Price column)
- predicted_prices.csv → Final predictions generated by saved models
- testing_and_training_data.ipynb → Model training and evaluation notebook
- save_model.py → Saves trained models using Pickle
- use_model.py → Loads saved models and predicts prices on new data
- linear_model.pkl → Saved Linear Regression model
- rf_model.pkl → Saved Random Forest model

## Workflow

1. Data Understanding – Explore and analyze the dataset.
2. Data Preprocessing – Clean data, handle missing values, encode categorical features, and create new features like Car_Age.
3. Train-Test Split – Divide the data into 80% training and 20% testing.
4. Model Training – Train two models: Linear Regression and Random Forest Regressor.
5. Model Evaluation – Evaluate performance using $R^2$ Score and MAE (Mean Absolute Error).
6. Model Saving – Save trained models with Pickle for future use without retraining.
7. Model Loading & Prediction – Load saved models and predict selling prices for new car data.

## Results

| Metric | Random Forest | Linear Regression |
|---|---|---|
| $R^2$ Score / MAE | 0.9582 / 0.6414 | 0.8489 / — |

## Model Saving and Usage

The models were saved using Pickle files (linear_model.pkl and rf_model.pkl). These saved models can be directly loaded in any environment to predict prices on unseen data. Predictions are generated and saved in the predicted_prices.csv file for reference.

## Conclusion

This project successfully demonstrates a complete end-to-end ML workflow. Random Forest outperformed Linear Regression by capturing non-linear patterns in the data. The use of Pickle makes the models reusable and efficient, enabling predictions without retraining.

---