

Vehicle Price Prediction - Project Report

Objective:

Predict vehicle prices using machine learning based on vehicle specifications, make, and features.

Dataset:

Contains vehicle details and prices including columns like name, model, year, price, mileage, fuel, transmission, body, doors, drivetrain, and trim.

Data Preprocessing:

- Dropped unnecessary columns: name, description, engine
- Filled missing numeric with median, categorical with mode
- Hybrid encoding: OneHotEncoder for low-cardinality, TargetEncoder for high-cardinality features.

Modeling:

- Algorithm: Random Forest Regressor
- GridSearchCV hyperparameter tuning with 5-fold CV
- Best Params: max_depth=None, max_features=sqrt, min_samples_leaf=1, min_samples_split=2, n_estimators=200

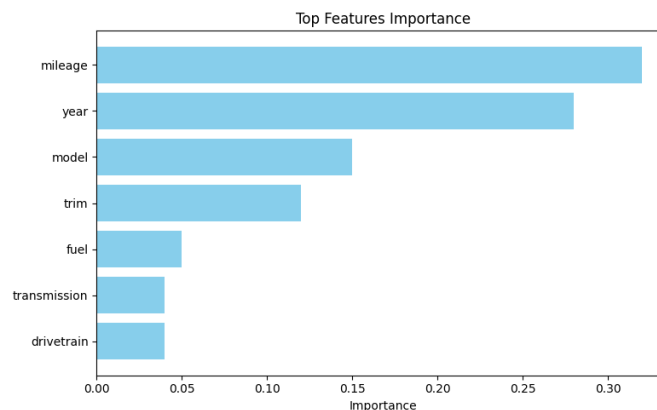
Evaluation Metrics:

R2 Score: 0.786

MAE: 4,061 USD

RMSE: 7,498 USD

Top Features (Importance):



Sample Prediction:

Vehicle - 2021 Camry LE, 15000 miles, 4 cylinders, 4 doors, Gasoline, Automatic, Sedan, FWD

Predicted Price: \$42,310.78

Conclusion:

- Successfully built a vehicle price prediction model using Random Forest
- Hybrid encoding prevents feature explosion
- Model can be extended with SHAP, XGBoost/LightGBM, and deployment for real-time predictions