

README

The cars dataset was originally scraped from thecarconnction.com with full specs (dimensions, fuel economy, performance specs, safety features, warranty, etc.) and pricing information.

Data wrangling and exploratory analysis was conducted subsequently. Car price distribution, buying guide, and price evolution over the years were illustrated.

After imputation, one-hot encoding of the dataset, machine learning was implemented using algorithms including linear regression, ridge regression, lasso regression, decision tree, random forests. A r^2 score close to 0.99 was achieved by Random Forest. Feature importance analysis revealed the torque spec and displacement are the most important factors determining the car prices.