

Project Proposal: Predicting Sports Car Prices Using Machine Learning

Problem Statement & Background

Accurately determining the market price of sports cars is critical for buyers, sellers, dealerships, and auto marketplaces. Prices vary significantly based on features like make, model, mileage, engine size, horsepower, transmission type, and market conditions. Having an effective predictive model can help stakeholders quickly estimate fair market value, facilitating better-informed decisions and more transparent transactions.

Goals & Objectives

The primary goal of this project is to develop a machine learning model that accurately predicts the selling prices of sports cars based on car specifications and market data.

Objectives include:

- Conducting thorough exploratory data analysis (EDA) to uncover factors most influential on price.
- Cleaning, preprocessing, and engineering relevant features from the dataset.
- Evaluating multiple predictive modeling approaches (e.g., linear regression, random forest) to identify the most accurate model.
- Delivering an interpretable model with clear insights into how each feature impacts car prices.

Data Sources

Primary dataset:

- [Sports Car Prices Dataset](#)

This dataset contains detailed information about sports cars, including attributes such as price, make, model, year, mileage, transmission type, fuel type, horsepower, and more.

Proposed Methodology

1. **Data Wrangling and Cleaning:**

- Handle missing values, duplicates, and erroneous data.
- Standardize data formats and units for consistent analysis.

2. Exploratory Data Analysis (EDA):

- Analyze price distributions, correlations, and key price determinants.
- Visualize relationships and outliers.

3. Feature Engineering:

- Create relevant new features such as age of car, mileage per year, or categorical encoding of make and model.

4. Modeling:

- Split data into training and test sets.
- Test various models (Linear Regression, Random Forest).
- Tune hyperparameters using cross-validation.

5. Model Evaluation:

- Measure performance using metrics like RMSE, MAE, and R^2 .
- Interpret model coefficients or feature importances to highlight key predictors.

Expected Deliverables

- A well-documented GitHub repository containing:
 - Cleaned dataset ready for modeling.
 - Jupyter Notebooks clearly demonstrating each stage of the analysis.
 - Final model code and predictions.
- A brief report (PDF) summarizing findings and recommendations.

Potential Impact

This project will benefit stakeholders by providing quick, data-driven estimates of sports car values, reducing reliance on subjective judgments. Additionally, it could be adapted or extended by marketplaces or dealers, providing practical, real-world applicability beyond academic use.