

Car Price Predictor - Project Report

Objective

To develop a machine learning-based web application that predicts the resale value of a used car based on key features such as brand, model, manufacturing year, kilometers driven, and fuel type. The goal is to assist car buyers and sellers in making informed pricing decisions.

Dataset Used

- Source: Quikr.com
- File: quikr_car.csv
- Description: The dataset was scraped from online used car listings and includes various attributes such as car name, company, year, fuel type, price, and kilometers driven.
- Size: ~8,000 rows (after cleaning)
- Preprocessing:
 - Removed null and irrelevant entries
 - Extracted model and brand information
 - Converted prices and distances to numeric values
 - One-hot encoded categorical variables

Model Chosen

- Algorithm: Linear Regression
- Justification:
 - Simple and interpretable
 - Suitable for regression problems with continuous target variables
 - Performs well with properly preprocessed structured data

Performance Metrics

- R² Score: ~0.92 (on test data)
- Train/Test Split: 80/20
- Interpretation: The model is able to explain approximately 92% of the variance in car prices, indicating strong predictive performance given the cleaned dataset.

Car Price Predictor - Project Report

Challenges & Learnings

- Data Cleaning: The raw dataset was unstructured and required extensive cleaning (handling symbols, strings in numeric fields, duplicate entries).
- Feature Extraction: Extracting meaningful features from textual columns was critical (e.g., separating brand and model).
- Model Limitations: Initially used models without preprocessing pipelines, which caused versioning issues and deployment incompatibilities.
- Version Mismatch: Faced errors due to different scikit-learn versions during model loading. Learned the importance of version control and saving full pipelines.
- Deployment: Gained hands-on experience with Flask and deploying ML models in a lightweight web app.