

Project Proposal: Vehicle Price Prediction

Student: Naume Karamitreski (ID: 89231338)

January 2026

1 Problem Description

The objective of this project is to develop and evaluate machine learning models for predicting the selling price of vehicles based on their characteristics. This is a supervised learning regression problem, where the target variable is the vehicle price. The project focuses on applying standard data mining and machine learning techniques to a real-world tabular dataset and comparing multiple models based on their predictive performance.

2 Dataset

I will use the Australian Vehicle Prices dataset available on Kaggle.

- Link: <https://www.kaggle.com/code/rewidashabaanmohamed/machine-learning/input>
- Description: The dataset contains vehicle listings with attributes such as brand, model, year of manufacture, mileage (kilometers driven), fuel type, transmission, body type, and price.
- Target Variable: Vehicle selling price (continuous numerical variable).

3 Modeling Algorithms

I intend to implement and compare multiple regression approaches:

- **Baseline Model:** Linear Regression (and/or Ridge Regression) to establish a simple baseline for prediction performance.
- **Advanced Model:** Random Forest Regressor to capture non-linear relationships between vehicle features and price. Optionally, Gradient Boosting methods may be explored if time permits.

4 Expected Outcomes

- **Quantitative Evaluation:** Models will be evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2).
- **Model Comparison:** Tree-based ensemble models are expected to outperform linear models due to their ability to model complex interactions among features.
- **Interpretation:** Feature importance analysis will be used to identify the most influential factors affecting vehicle prices, such as vehicle age, mileage, and manufacturer.