

Problem Description:

A Chinese automobile company aspires to enter the US market by setting up their manufacturing unit there and producing cars locally to give competition to their US and European counterparts. They have contracted an automobile consulting company to understand the factors on which the pricing of cars depends. Specifically, they want to understand the factors affecting the pricing of cars in the American market, since those may be very different from the Chinese market. Essentially, the company wants to know:

- Which variables are significant in predicting the price of a car
- How well those variables describe the price of a car

Based on various market surveys, the consulting firm has gathered a large dataset of different types of cars across the American market.

Business Goal:

You are required to model the price of cars with the available independent variables. It will be used by the management to understand how exactly the prices vary with the independent variables. They can accordingly manipulate the design of the cars, the business strategy etc. to meet certain price levels. Further, the model will be a good way for the management to understand the pricing dynamics of a new market.

Dataset:

https://drive.google.com/file/d/1FHmYNLs9v0Enc-UExEMpitOFGsWvB2dP/view?usp=drive_link

Key components to be fulfilled:

1. Loading and Preprocessing

- Load the dataset and perform necessary preprocessing steps.

2. Model Implementation

- Implement the following five regression algorithms:

- 1) Linear Regression
- 2) Decision Tree Regressor
- 3) Random Forest Regressor
- 4) Gradient Boosting Regressor
- 5) Support Vector Regressor

3. Model Evaluation

- Compare the performance of all the models based on R-squared, Mean Squared Error (MSE), and Mean Absolute Error (MAE).
- Identify the best performing model and justify why it is the best.

4. Feature Importance Analysis

Identify the significant variables affecting car prices (feature selection)

5. Hyperparameter Tuning

- Perform hyperparameter tuning and check whether the performance of the model has increased.