DOCUMENTATION FOR DATA SCIENCE TASK IMPLEMENTATION

CAR PRICE PREDICTION PROJECT

Task Description

Task: Regression
Target Variable : Selling Price
Features: Various features were used in predicting the car price such as; name, year, kilometers driven, fuel type, seller type, transmission, owner, mileage, engine, max power, torque and seats.

Data Preprocessing/Data Cleaning

Before using the data for prediction, various preprocessing steps are applied including;

1. Handling missing values : Dropping the null values.
2. Duplicates: Removing Duplicates.
3. Encoding categorical features / name column.

Feature Engineering

Extractions (stripping of the string character) were done in the name, fuel, owner, transmission, mileage, torque columns using a function.

Chart

Chart Used: Bar chart, Histogram, Heat map, Scatter plot
Reason: Bar chart was used to visualize the count of categorical features. The Histogram for each feature was plotted to visualize the distribution (it reveals if there is either a normal distribution or not). The heat map reveals the correlation matrix to understand the relationship between features.

Algorithm

Gradient Boosting Regressor: it is well suited in predicting the selling price of cars. It is an ensemble method that can handle outliers. It has higher predictive accuracy.

Model Evaluation

The trained model is evaluated using the metrics such as R2 score and Max error to assess its performance in predicting the selling prices of cars accurately.