

Assignment-2

Results

Name: Shaik Rumana Taj

Number: 16357409

a) Detecting and dealing with missing values:

First, we must find missing values in each column of the dataset.

We can substitute missing values in numerical columns (such as year, mileage, engine horsepower, and so on) with the column's mean or median.

For categorical columns (such as fuel type and transmission type), missing values can be replaced using the column's mode (most frequent value).

If the missing values are significant and cannot be properly imputed, we may consider removing those rows/columns.

b) Remove Units from Attributes:

To retain solely numerical values, we must eliminate the units (kmpl, CC, bhp, and lakh) from properties like as mileage, engine displacement, engine horsepower, and new price.

c) Changing Categorical Variables into Numerical One-Hot Encoded Values:

We'll transform the categorical variables "Fuel_Type" and "Transmission" into numerical one-hot encoded values, with each category assigned a binary column.

d) Introducing a new feature:

We will add a new column to the dataset to calculate the car's current age by subtracting the "Year" value from the current year.

e) Performing Different Operations on the Dataset:

We will utilize operations such as select (to choose certain columns), filter (to filter rows based on conditions), rename (to rename column names), modify (to create new columns), arrange (to sort rows), and summarize with group by procedures.

After these processes are done, the dataset will be preprocessed and available for additional analysis or modeling.

