

MERCEDES-BENZ GREENER MANUFACTURING

DESCRIPTION

REDUCE THE TIME A MERCEDES-BENZ SPENDS ON THE TEST BENCH.

PROBLEM STATEMENT SCENARIO:

You are required to reduce the time that cars spend on the test bench. Others will work with a dataset representing different permutations of features in a Mercedes-Benz car to predict the time it takes to pass testing. Optimal algorithms will contribute to faster testing, resulting in lower carbon dioxide emissions without reducing Mercedes-Benz's standards.

THE GOAL OF THE PROJECT WAS TO BUILD A MACHINE LEARNING MODEL TO PREDICT THE TIME A MERCEDES-BENZ SPENDS ON THE TEST BENCH. THE PROJECT WAS PERFORMED IN SEVERAL STEPS AS FOLLOWS:

Step 1: Importing the necessary libraries and reading in the data from the CSV file using pandas.

Step 2: Exploring the data by checking for missing values, data types, and descriptive statistics.

Step 3: Visualizing the distribution of the target variable in the training data.

Step 4: Encoding the categorical features using one-hot encoding.

Step 5: Splitting the data into training and testing sets.

Step 6: Building a baseline linear regression model and evaluating its performance using mean squared error and R-squared score.

Step 7: Building a more complex model using XGBoost and evaluating its performance using mean squared error and R-squared score.

Step 8: Tuning the XGBoost model using GridSearchCV to find the optimal hyperparameters.

Step 9: Building an ensemble of XGBoost models with different random seeds and taking the average of their predictions to improve the performance.

Step 10: Making predictions on the test data using the best-performing model and submitting the predictions in CSV format.

Finally, the model's predicted values were analysed to draw conclusions about the time a Mercedes-Benz spends on the test bench, and it was emphasized that the accuracy of the model's predictions should be validated before making any decisions based on them.