

Mercedes-Benz Greener Manufacturing

Problem Statement Scenario:

Since the first automobile, the Benz Patent Motor Car in 1886, Mercedes-Benz has stood for important automotive innovations. These include the passenger safety cell with a crumple zone, the airbag, and intelligent assistance systems. Mercedes-Benz applies for nearly 2000 patents per year, making the brand the European leader among premium carmakers. Mercedes-Benz is the leader in the premium car industry. With a huge selection of features and options, customers can choose the customized Mercedes-Benz of their dreams.

To ensure the safety and reliability of every unique car configuration before they hit the road, the company's engineers have developed a robust testing system. As one of the world's biggest manufacturers of premium cars, safety and efficiency are paramount on Mercedes-Benz's production lines. However, optimizing the speed of their testing system for many possible feature combinations is complex and time-consuming without a powerful algorithmic approach.

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.

Following actions should be performed:

- If for any column(s), the variance is equal to zero, then you need to remove those variable(s).

Both the training and the test data set are combined and the variance of each column is checked. The column with zero variance is removed.

- Check for null and unique values for test and train sets.

Both the training and the test data set are combined and the each column is checked for null values. The column with null value is removed.

- Apply label encoder.

Both the training and the test data set are combined and the categorical columns are label encoded.

- Perform dimensionality reduction.

Dimensionality reduction is performed on train and test data.

- Predict your test_df values using XGBoost.

XGBoost is trained and evaluated on the training data set and then tested on the testing data set.