

Mercedes-Benz Greener Manufacturing -Project

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).
- Check for null and unique values for test and train sets.
- Apply label encoder.
- Perform dimensionality reduction.
- Predict your test_df values using XGBoost.

Introduction:

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.

Data:

The dataset used for the project contained anonymized data on various features of the cars, such as the chassis, engine, and model type, as well as the time taken to test the car. The dataset contained 4209 rows and 378 columns. The data was preprocessed by removing any features with constant values, features with more than 95% of missing values, and features with a single unique value. Additionally, any rows with missing values were removed from the dataset. Perform dimensionality reduction was completed

Model:

The XGBoost algorithm was used to build the machine learning model for this project. XGBoost is a popular gradient boosting algorithm that is known for its speed and accuracy. The algorithm was trained on the preprocessed dataset, and the hyperparameters were optimized using grid search with cross-validation.

Conclusion:

The Mercedes-Benz Greener Manufacturing project was successful in building a machine learning model to predict the time taken to test a new car based on various features of the car. The XGBoost algorithm was found to be a suitable algorithm for this task, and the trained model had a reasonably good performance.