Mercedes-Benz Greener Manufacturing

**Domain:** Automotive

**Objective:**

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.

**Analysis to be done:**

Reduce the time a Mercedes-Benz spends on the test bench.

1. If for any column(s), the variance is equal to zero, then you need to remove those variable(s).
2. Check for null and unique values for test and train sets.
3. Apply label encoder.
4. Perform dimensionality reduction.
5. Predict your test\_df values using XGBoost.

**Approach:**

1. If the column has only one value or cardinality=1 then we can drop it as it will not be of any use, and other values except 0 will be retained and updated to the same df.
2. This can be checked using the isnull() function.
3. This can be done while filtering the df to remove futile values and updating the df with usable columns
4. Principal component analysis (PCA) can be used. Linear dimensionality reduction using Singular Value Decomposition of the data to project it to a lower dimensional space.
5. Split data and feed in all the required parameters for xgboost and train() until r2 score improves in 50 rounds (early stop), we can predict() the test df values based on the trained model and export to a df or file.