Canisha Barron

7/8/2021

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.

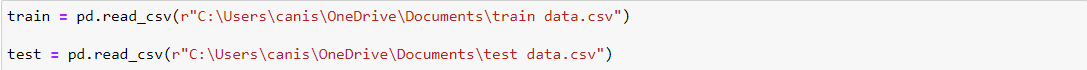
In order to reduce the time a Mercedes-Benz spends on the test bench; the following will be completed:

* 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.

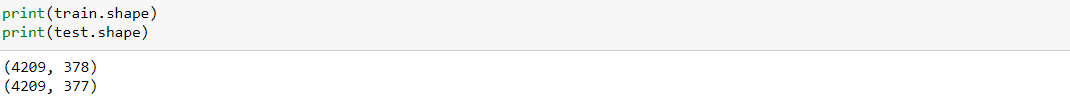
#Import libraries and loading datasets

A picture containing shape

Description automatically generated



Check the shape for train and test dataset and view dataset.



A screenshot of a computer

Description automatically generated with medium confidence

Drop the ID column in both train and test dataset. Remove variables where the variance is equal to zero and check for columns with null and unique values.

Table

Description automatically generated with low confidence

A picture containing graphical user interface

Description automatically generated

Graphical user interface, application

Description automatically generated

A picture containing table

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

Label Encoding:

Check for columns having dtype: Object. Perform label encoder on those columns.

Graphical user interface, text, application

Description automatically generated

Dimensionality Reduction:

Using PCA components with 98% variance will be retained. Separate X and y from the dataset and perform test train split to get the train and validation data. Apply n\_components and experienced\_variance\_ratio on the PCA. Transform the following using pca for xgboost:

X\_train 🡪 training data (pca\_fit\_transform\_X\_train)

X\_val 🡪 validation data(y\_train)

test\_df🡪 test data(pca\_transform\_X\_test)

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application, Teams

Description automatically generated

A picture containing graphical user interface

Description automatically generated

Chart

Description automatically generated

Graphical user interface, application

Description automatically generated with medium confidence

Xgboost:

Build an xgboost model for linear regression and fit X\_train and y\_train in the model. Predict on pca\_transform\_X\_test. Check RMSE for the model.

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

A helpful task would be to create a different model and compare RMSE as highlighted below:

Graphical user interface, text, application, email

Description automatically generated