Mercedes-Benz Greener Manufacturing

Machine Learning Course-End Project Write-up

Introduction:

Mercedes-Benz, a pioneer in automotive innovation, continually strives to enhance its manufacturing processes. The objective of this project is to reduce the time spent on the test bench for Mercedes-Benz cars using a robust machine learning approach.

1. Dataset Overview:

The dataset consists of two CSV files - `train.csv` and `test.csv`, representing the training and testing datasets, respectively. The training set contains various features related to the manufacturing process, and the target variable, 'y,' represents the time a Mercedes-Benz spends on the test bench.

2. Data Preprocessing:

- Handling Non-Numeric Values:
  - Identified non-numeric columns using `select_dtypes` and `LabelEncoder`.
  - Applied label encoding to convert categorical variables into numeric format.
- Variance Analysis:
  - Calculated the variance for each column to identify low-variance features.
  - Removed columns with variance below a threshold (0.2) to reduce dimensionality.
- Correlation Analysis:
  - Computed the correlation matrix for the remaining features.
  - Detected highly correlated features (> 0.9 correlation) and retained only one of them.
- Handling Missing Values:
  - Checked for missing values in the dataset.

- Imputed missing values with the mean for numeric columns.

3. Dimensionality Reduction:

- Applied Principal Component Analysis (PCA) to further reduce the number of features.
- Selected the top principal components that capture the most variance.

4. Model Training:

- Utilized the XGBoost algorithm, a powerful gradient boosting technique, for regression.
- Split the preprocessed training data into features (X) and the target variable (y).
- Trained the XGBoost model on the training data.

5. Testing and Predictions:

- Loaded the test dataset and applied the same preprocessing steps as the training set.
- Made predictions using the trained XGBoost model on the preprocessed test data.

6. Results:

- Evaluated the model using Mean Squared Error (MSE) on a validation set (not explicitly shown).
- Applied the model to the test dataset and obtained predictions for the target variable.

Conclusion:

By employing a combination of feature engineering, variance analysis, correlation analysis, and machine learning techniques, this project aims to optimize the testing time for Mercedes-Benz cars, contributing to a greener manufacturing process.

Future Steps:

For future improvements, consider hyper parameter tuning, exploring other dimensionality reduction techniques, and refining the feature engineering process. Additionally, assess the model's generalization performance on unseen data and monitor its impact on reducing testing time.

This machine learning project provides a comprehensive overview, from data preprocessing to model training and testing, aiming to reduce the testing time for Mercedes-Benz cars while maintaining high standards of safety and reliability.