|  |  |  |
| --- | --- | --- |
|  | **Use Case** | **Tasks** |
|  | -----------------------------------  ## **Problem Statement -**  #### **Daimler challenge - reduce the time that cars spend on the test bench.**  -----------------------------------  - The objective of the Mercedes-Benz Greener Manufacturing competition is to develop a machine learning model that can      - accurately `predict the time a car will spend on the test bench` based on the vehicle configuration.      - `reduce the total time spent testing vehicles by allowing cars with similar testing configurations to be run successively.  - The `vehicle configuration` is defined as the set of customization options and features selected for the particular vehicle.  - This problem is an example of a machine learning `regression` task because it requires predicting a continuous target variable (the duration of the test) based on one or more explanatory variables (the configuration of the vehicle).  - This problem is also a supervised task because the targets for the training data are known ahead of time and the model will learn based on labeled data.      2. Prepare the data to feed into the machine learning model.      3. Select an appropriate algorithm/method for the problem.      4. Optimize the model using the labeled training data. | * **Understanding of data** * **Nulls** * **Duplicates** * **Types of data columns** * **Analyze the cardinality** * **Histogram/distribution of y column** * **Apply Filter methods of Feature selection (var, corr)** * **Apply Inferential statistics to check the usefulness of categorical columns** * **List the columns to be dropped** * **On the cleaned data, perform KNN and Linear regression** * **Observe the regression performance metrics** * **Save the model/ load/predict** |
|  | location\_train = r"D:\AI-DATASETS\01-MISC\merc-train.csv" (5000 rows)  location\_test  = r"D:\AI-DATASETS\01-MISC\merc-test.csv" (ignored) |  |