**PROJECT REPORT**

**AUTOMOBILE ENGINE BENCH TEST PREDICTIONS**

**ABHILASH URMALA**

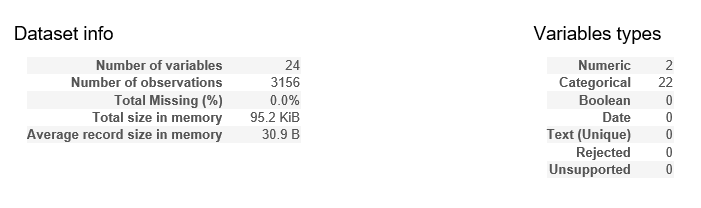
**BATCH-33 (1508)**

**The Problem:**

A leading car manufacturer is designing an automobile engine. It is required by law and regulations, that a given engine configuration goes through rigorous engine bench test before it is passed for production. Each bench test is an expensive, noisy and time-consuming process. Instead the manufacturer would like to use some previous data on various configurations tested and determine through analytics model if their new design will pass or not. This will help them narrow down only few configurations for further testing on physical bench test.

**Dataset:**

The given automobile dataset contains the various engine configurations as features and the target variable represents the outcome of the bench test.

****

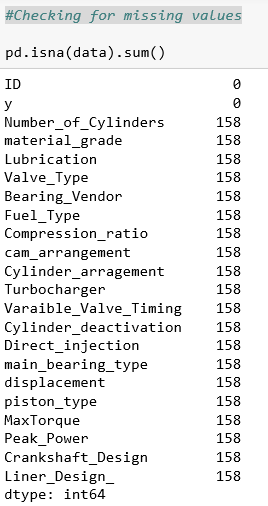
**Fig: Raw Data Characteristics**

**Data Preparation:**

Data preparations plays a vital role in building models. Before model building, the data preparation involved the following steps:

* Renaming columns
* Label Encoding (categorical variables)
* Adding additional data
* Handling missing values

**Checking for missing values :**

****

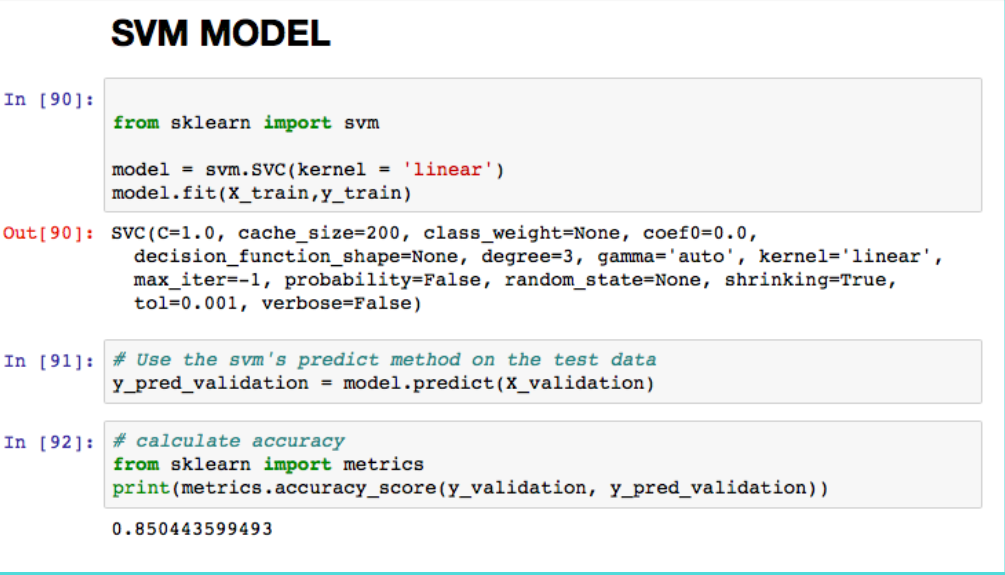
Except ID and the target variable, all the other variables contain missing values. Before imputing them, let’s check for the patterns in missing values. The following heatmap shows the correlation between missing values. One thing to note is that any columns that have no missing values will not show up in the heatmap. The remaining columns will appear with their correlation value between -1 and 1, and if the value rounds down to 0 (>-0.05 or < 0.05) then no value will be displayed. A value of -1 means that in all cases, when the first column is missing then the second column is not missing. A value of 1 means that in all cases, when the first column is missing the second column is missing also.

We observe that there are no patterns in the missing values.

**Imputation Methods Used:**

* fillna(method = ‘Mode’):
  + Propagates the previous value forward.
* KNN with k=3:
  + Considers the k nearest (in terms of characteristics) records for filling up missing values**.**

**Model Building:**

****

**Fig :** SVM Model in Python

* The predictions were 87% accurate.

**Conclusion:**

The problem of performing physical bench test can thus be solved by considering the past data about various engine configurations.