|  |  |
| --- | --- |
| **R Logistic Regression + DT + KNN + RF + SVM Class Assessment** | **[Time: 4 hrs] [Total Marks: 100]** |

**Chemical Classification**

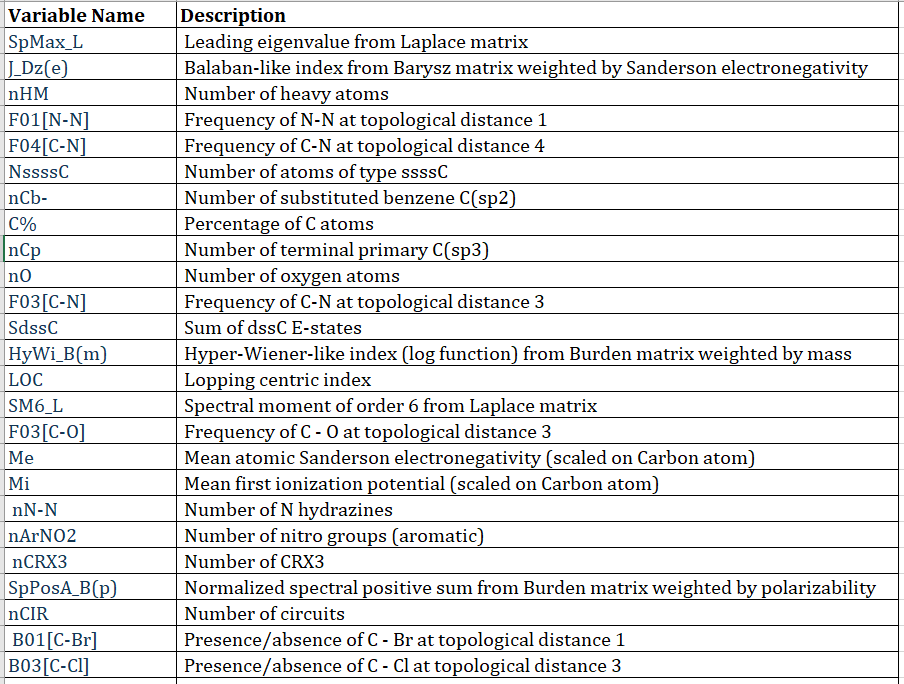
## Problem Statement:

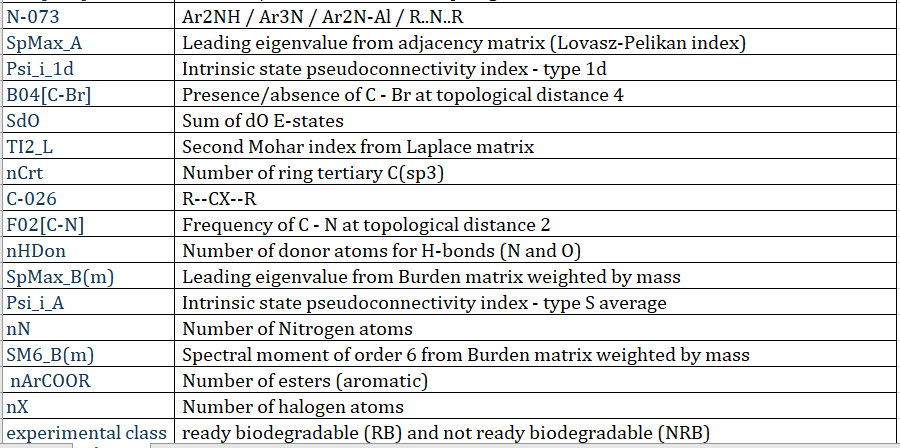
To predict whether a given chemical is biodegradable or not based on the given chemical structural information.

## Dataset Information:

Data set containing values for 41 attributes (molecular descriptors) used to classify 1055 chemicals into 2 classes (ready and not ready biodegradable).

# Variable Description:





**Scope:**

* Data Pre-processing
* Training Logistic Regression, Decision tree, SVM classifier, Random Forest and KNN to predict whether a given chemical compound is biodegradable or not
* Assess the correctness in classifying data with respect to efficiency and effectiveness of all the classifier in terms of accuracy, precision, sensitivity, specificity and AUC ROC
* Tuning the hyperparameters of the models and identifying the optimum model