**Project 1: Doing data Modeling using different machine Learning:**

**MLR: Multiple Linear Regression**

**SVM: Support Vector Machin**

**ANN: Artificial neural network**

You are required to complete this project, writing the codes that have not been implemented and run the program with three different machine learning tools. There are three files

* main\_datamining.py
* fitting\_scoring.py
* process\_input.py

In the fitting\_scoring.py you should implement

* the fitness function and
* calculateRMSE() function

In the process\_input.py you should implement

* sort\_descriptor\_matrix
* open\_descriptor\_matrix function
* open\_taget\_values function and
* Remove\_invalid\_rows function

Run the work with three machine learning model (MLR, SVM and ANN).

The fitness function is implemented as:

numerator = ((mt - n - 1 ) \* RMSEt2 )+ (mv\*RMSEv2)

denominator = mt - n -1 + mv

f = Sqrt (numerator / denominator)

* mt is the number of samples of training.
* mv is the number of samples of validation
* n is the total number of descriptors
* RMSEv is the Root mean square error of validation:
* RMSEt is the Root mean square error of training

**Test your program with some small data set first to see if you are getting correct result. Then test it with the given two data sets files, and take the snapshot of three outputs:**

* **Run it with MLR, get the snapshot of the results and place it in a file called Project-1-MLR**
* **Run it with SVM, get the snapshot of the results and place it in a file called Project-1-SVM**
* **Run it with ANN, get the snapshot of the results and place it in a file called Project-1-ANN**
* **Place all the files in a folder and call the folder based on your name and your partner’s name (ex:Jack-and-Nancy-Project-1)**
* **Zip the folder and only one person in the team is required to submit the work.**

**If don’t provide the wrong naming convention in your projects or labs you will lose 50% of the grade immediately.**