# Assignment 1

Deadline: 4th March 2018
Note: Delay in the late submission will automatically deduct the 10% of the marks scored per day.
Plagiarism will be thoroughly checked and plagiarism found in any task would result in ZERO marks for the whole assignment.

## Task 1 (30 Marks)

Consider the regional schools' record file: Tehsil Schools.xlsx – Read that file in your program and load all the data in a list or numpy array

1. Print the names of all schools for which the number of students passed in 10th Class Exam is ZERO
2. Find the percentage of large-sized schools w.r.t. student enrollment, assuming a school to be large if it had more than 50 Students in 9th (2012) as per Registration.
3. Among all large sized schools, print the name of the school with the highest % dropout

## Task 2 (70 Marks)

A number of datasets have been shared with you for task 2. You have to see the **last two digits** of your registration number and calculate it **modulo with 21**. The value you get is the dataset number you will have to work on.

> Eg: Registration number: 1723**44**
>
> Last two digits of the registration number: **44**
> Modulo with 21: 44%21 = **2**

So, the student with registration number 1723**44** will be working on dataset number **2**. You will always get a value between 0 and 20.

The dataset contains the attributes along with the labels. These datasets were pre-processed and converted to vectorized form. Now you have to download the dataset and divide it into two sets, called training and testing. You can set a division of your dataset as 80% training and 20% testing. Implement the **Naive Bayes Classifier** for your dataset and analyse its performance on the train and test corpus.

## LMS Upload:

A zip file containing the source code for both tasks and a one page document reporting your findings for both tasks.

----------GOOD LUCK----------