National University of Sciences and Technology (NUST)
School of Electrical Engineering and Computer Science (SEECS)
MSCS-07/MS IT–18, Information Retrieval
Semester Project Assignment, May 2018

Student Name: …………………….…………. Roll No: ………………..
Student Name: …………………….…………. Roll No: ………………..
Student Name: …………………….…………. Roll No: ………………..

Details:
This assignment is due on Monday 28th May 2017 at 3:00 pm. There will be a group comprising three students each. The evaluation of the project will be based on running code and demo in class. The student is required to submit an implementation of the following algorithms in any language of their choice (preferred Java) using machine learning toolkits / API (for java Weka is a recommended but can use any other library).

1. Algo – 1: Naive Bayes Classifer
2. Algo – 2: Rocciho Classification
3. Algo – 3: kNN Classification

Submission Procedure:
You shall create a repository on GitHub and submit your code using the GitHub. The project should have a readme.md file explaining the implementation details as well as detailed steps to run the system.

Note: Using the SSECS email address you can sign up for GitHub student account that will allow you to create private repositories. In case you have created a private repository, kindly make it public before submission of the assignment.

Datasets for Working:
Following links are for the datasets. The links also contain work details. You have to choose any three datasets to work.
1. *https://archive.ics.uci.edu/ml/datasets/Bag+of+Words*
2. *https://archive.ics.uci.edu/ml/datasets/DBWorld+e-mails*
3. *https://archive.ics.uci.edu/ml/datasets/Health+News+in+Twitter#*
4. *https://archive.ics.uci.edu/ml/datasets/NIPS+Conference+Papers+1987-2015*
5. *https://archive.ics.uci.edu/ml/datasets/Legal+Case+Reports*
6. *https://archive.ics.uci.edu/ml/datasets/Sentence+Classification*