# TM18004 - Privacy and Security on Online Social Media
# Assignment 1

### Instructions :

- Languages allowed : Python/Java
- You are free to use API documentation but if referring to any other sources, please cite them
- Please write your own code. All codes will be tested for Plagiarism and if found, institute policy for plagiarism will be followed.
- Document your code properly.
- You can use any database for storing the data but it will be tested at the time of demo.
- Write all the analysis along with graphs, charts etc in analysis.pdf
- Make a readme.txt file with the instructions how to run the code. All libraries, sources etc used should be properly mentioned in it.
- Do the Assignment individually
- Zip all your code files along with analysis and readme file in RollNo_Assignment1.zip format. Example 201402230_Assignment1.zip

Collect tweets using keywords: @iiit_hyderabad and @iiscbangalore separately (should be case insensitive). Use twitter search api (REST API) for this. Do the following analysis on both handles separately. *(10 + 15 + 10 + 10 + 25 = 70 marks)*

a) Make separate histograms of top 20 most frequent words in tweets. There will be two histograms one for IIIT & another for IISC. Feel free to creatively present the results.
b) Remove the stop words, and create a word cloud.
c) Make a time series graph from your data in which x axis is the time and y axis is number of tweets tweeted at that time. You can take range of values for x in terms of week or month or day. Feel free to creatively present the results.
d) Analyse the data in terms of WHAT these tweet contains. You can present this analysis in form of pie charts. Example x number of tweets contained images, y number contained urls etc.
e) Do the sentiment analysis on collected tweets of handles separately using the keyword 'Research'. Find out the percentage of polarity of sentiment of both handles. (Use either of NLTK or TextBlob libraries of python)
f) As we did in class, get the top 10 most popular tweets of both the accounts.

Try to draw inferences from this instead of just reporting graphs and numbers. Compare the analysis you have drawn from both handles.

Note : Make sure you collect only english language tweets. Do not include stop words in your analysis. Filter out all repetitive tweets.

**Bonus Question**

Given a twitter ID, print the very first tweet (you are free to explore and try out different methods but the response should be in real time from your script and printed on console)

(20 marks)