**AML-3204 Social Media Analytics**

**Assignment-1\_Part-2 (10%)**

**Topic Modeling Using Twitter and LDA**

In this assignment, you will write Python scripts to download Tweets for any Twitter account OR News using any News API. Then, on the downloaded tweets, you need to apply the following:

1. Cleaning of the tweets or News
2. Apply LDA for topic modeling; you can run LDA for 10-15 topics.
3. Based on the Topic\term strength, you need to create a word cloud
4. When a word is clicked, then it should display the related tweets for that word.

The output of your program should behave in the same way as it is done in the following link [ignore, if it does not work now]:

<https://tweettopicexplorer.neoformix.com/#n=NYTimes>

FYI, the tool in the given URL is just calculating the frequent words, but here we are not calculating the frequent words. Here, the collection of words will be selected based on the LDA model. So, make sure you are using the LDA model properly.

Here is a tutorial on LDA:

<https://towardsdatascience.com/topic-modeling-and-latent-dirichlet-allocation-in-python-9bf156893c24>

Submission:

To submit, please do the following:

1. Write a short report on the assignment. The report should have the following:
   1. Data collection procedure
   2. Data cleansing steps
   3. A Bar Diagram showing the top-5 topics for any Twitter handler of your choice
2. Please upload the report and the Python files

Presentation:

During the class, you will also demonstrate your Python solution. You will share your screen to describe your Python program. You can prepare slides to present remaining stuff.