**Assignment No: 3 Matriculation Number: 00387718, 00413018**

**Data Extraction: Psychological Disorder Profiling through Social Media Mining**

**Overview:**

For collecting the data from twitter, a twitter developer account is required which enables to extract the data by providing the access credentials. Using “Tweepy” library, python can be linked to the twitter server to stream the live tweets using python script. This live streamed data is a printed out in raw JSON format. Processing the JSON format data becomes a tedious task and increases the code complexity as well. Thus, we tried an alternative approach to extract the data in CSV format. TAGS (Twitter Archive Google Sheets) – Plugin was used to extract the twitter data on to the Google Spreadsheet.

**Data extraction methodology:**

**Failed approach:**

* Twitter API
* Access Credentials
* Connect to Python using “Tweepy”
* Data set in JSON format

**Success approach:**

**Using TAGS-Plugins**

* Setup Twitter Access.
* Enter term - #depression, #ocd, #bipolar, #anxiety.
* Tweet count – 3000 per category.
* TAGS output – CSV format.

**Future Work:**

* Cleaning collected Data Set
* Implementing Machine Learning Algorithms
* Perform Sentiment Analysis
* Psychological Profiling