**CS5542 Big Data Apps and Analytics**

**In Class Programming –2**

**7th February 2022(11:59 pm CST)**

**NLP:**

**Use the same data (that we obtained by in source code** Data = pd.read\_csv('https://raw.githubusercontent.com/dD2405/Twitter\_Sentiment\_Analysis/master/train.csv')**) and perform the sentiment analysis task on this data using one of the scikit learn classifier for text.**

ICP Requirements:

1. Data cleaning and preprocessing (at minimum have the following: Removing unnecessary columns or data, Removing Twitter Handles( @user ), Removing punctuation, numbers, special characters, removing stop words, Tokenization, and Stemming, TFIDF vectors, POS tagging, checking for missing values , train/test split of data).
2. Data Visualization and analysis for critical steps (WordCloud, Bar plots, etc)
3. Model building and successfully executing the model to make prediction.
4. Code quality, Wiki Report quality, video explanation

Submission Guidelines:

1. Sign in into your github account
2. Click on this link : <https://classroom.github.com/a/oZ2Rr7OL>
3. Accept the ICP-2 (Assignment 2)
4. Complete your ICP and create your wiki report (Pdf or Word doc).
5. Folders for ICP:
   1. Create two folders for source code and documentation
      1. Source code folder contains only the code and output file (if applicable)
      2. Documentation folder contains wiki report and the images of your results
6. The wiki report should have at least the following:
   1. What you learned in the ICP
   2. ICP description what was the task you were performing
   3. Challenges that you faced
   4. Screen shots that shows the successful execution of each required step of your code
   5. Out put file link if applicable
   6. Video link (YouTube or any other publicly available video platform)
   7. Any inside about the data or the ICP in general
7. Upload your ICP folders to your assignment GitHub repository
8. Click on the add a readme file on the next page and write done your and your partner’s name and email

You are all Done!!!!!!!