**CS5542 Big Data Apps and Analytics**

**In Class Programming –3**

**11th September 2020**

**Submit ICP Feedback in Class. :** [**Lnik to Feed back Form**](https://docs.google.com/forms/d/e/1FAIpQLSesllFh5_STnj7RbHyQainRG_2EIKw1csp8ObP5FWjpVnGVOg/viewform)

**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). (70 points)
2. Data Visualization and analysis for critical steps (WordCloud, Bar plots, etc) (10 points)
3. Model building and successfully executing the model to make prediction. (10 points )
4. Code quality, Pdf Report quality, video explanation (10 points)

Submission Guidelines:

Same as ICP 2.