# Self Evaluation Report

Anant Ashutosh Sharma
anantas2@illinois.edu

## Individual Project | Free Topic

Text and Tweet Classification using Machine Learning

## 1) Have you completed what you have planned ?

Yes, I have successfully completed what I planned and presented in the Project Proposal as well as the Progress Report. The following table mentions the tasks which have been successfully completed in this project along with the checklist. (These were the same tasks as mentioned in the Project Proposal)

| S. No. | Task | Completion |
|---|---|---|
| 1. | Import and Pre-process the input textual data<br>a. Cleaning of text<br>b. Stop Words removal<br>c. Stemming<br>d. Removal of non-alphabetic characters | Completed |
| 2. | Extract features from the pre-processed data<br>a. Count Vectors<br>b. TF-IDF Vectors<br>i. Bag of words (word level)<br>ii. Bag of n-grams (n-gram level)<br>iii. Character level | Completed |
| 3. | Training Model – Learning<br>a. Naïve Bayes<br>b. Linear Classification<br>c. SVM<br>d. Random Forest<br>e. Convolutional Neural Network | Completed |
| 1. | Classification of text using the trained model | Completed |
| 2. | Evaluation and Comparison of different models | Completed |
| 3. | Create new data set for classifying political tweets in India | Completed |
| 4. | Fetch tweets using different parameters and classify as Political / Non Political | Completed |

## 2) Have you got the expected outcome ?

Yes, I have got the expected outcome from my proposed project. All the machine learning models successfully classify the textual data from three different datasets into the respective categories with good accuracies and F1 scores. The purpose of this project was to present the different machine learning models which can be used to perform text classification instead of comparing and competing them against one another.

Further, as expected, the tweets are automatically streamed using the Twitter API and are classified into political and not political as expected. All the results and outcomes have been successfully presented in the GitHub repo (readme.md) as well as in the various Jupiter Notebooks.