Twitter Sentiment Analysis

(Akshaj Gupta,Maddela Sai Karthik)

In the project we need to classify the tweets into different orientation of their sentiment. A tweet can be classified into 3 different categories namely, positive , negative and neutral.
Positive tweet - A tweet has a positive sentiment regarding the subject.
Negative tweet - A tweet has a negative sentiment regarding the subject.
Neutral - The tweet has no polarity at all regarding the subject.

Data Collection and Preprocessing - Twitter has its own API that allows to download the tweets in the customized way. After downloading the data it need to be preprocessed before training and testing. Data preprocessing include the separation of the words in a single hash tag, making sensible words out of creative language e.g. lol = laughing out loud, replacing the urls with proper single words, replacing the emoticons with the best describing labels e.g. :) = positive, and removing unnecessary words called as stop words.

Training and Testing - To train the data we used 2 different approaches . In our first approach we used a CNN model to train the data. We ran it for 5 different epochs and the average accuracy was 76.94% on a training sample size of 32000 and testing size of 8000. Another method included the use of TF-IDF, for that the model was trained for the Support Vector Machine and Naive Bayes classifier for multivariate Bernoulli models, the reported accuracies were 77% and 75.38% respectively.