Goals : sentiment analysis of twets , finding top keywords , finding top users

Procedure : we have found the sentiment of the tweet using decision tree classifier , where the classifier was trained on matlab using treebagger with a single tree and the decision boundaries were obtained , we did not bother to put the classifier in the code as it is not an online training. For a particular tweet we have given a score with the help of AFINN database , which is a database containing words rated between -5 and +5 based on their use manually . We resorted to this mode of classification as it is intuitive compared to having a dictionary and obtaining a feature vector as that vector contains mostly 0 and on performing dimensionality reduction with PCA / NMF shows 0 feature vector everywhere. Most importantly the feature vectors are gigantic and difficult for the classifier to train on , so we resorted to the other approach of obtaining feature vector by giving a score on the basis of a database. After all this is the technique that is found in most of the text existing . AFINN is a very small database though a larger database would have been way far better, but were not freely available and easy to use as AFINN. Decision tree classifier is a weak classifier so the accuracy of this classifier is less which could have been experimented upon.

Input : the programme takes it automatically

Output : The top keywords and users along with a sentiment analysis of the available tweets(ie. We shall show what overall sentiment is on analysing tweets for a particular amount of time )