

## Milestone 3

### 1. Dictionary (AFINN)

Use  $\text{sum}(\text{tf-idf} * \text{score}_t)$ , where  $\text{score}_t$  is the score of term  $t$  when it appears in the tweet, to compute scores for all the tweets and classify them accordingly.

Highest score is 15 and lowest score is -15; threshold is set from 0 to 9, and precision, recall, and f-measure are computed for positive, negative and neutral, respectively.

### 2. SentiWordNet

Leverage the approach introduced in the SentiFul paper to classify terms and compute score and weight for each term.

- a) Use  $\text{sum}(\text{sentiment} * \text{weight}_t * \text{score}_t)$  for all term  $t$  in one tweet to compute the score for each tweet.

Three thresholds are set:

obj\_weight\_threshold – the objective weight to shortcut the classification as neutral

pos\_score\_threshold – the threshold to classify one tweet as positive

neg\_score\_threshold – the threshold to classify one tweet as negative

Tested for obj\_weight\_threshold from 0.01 to 0.2, pos\_score\_threshold from 0.08 to 0.2, and neg\_score\_threshold from -0.01 to -0.2.

- b) Use  $\text{sum}(\text{tf-idf} * \text{score}_t)$  to compute scores for all the tweets and classify them.

### 3. More Live Feeds From Twitter

Download more tweets with smileys, such as :), :D, and =(, as key words, and pre-classify them according to the facial notation in each tweet. Apply naïve Bayes classifier with certain features to test against those tweets. This achieves the best performance as we showed during the demo.