Project Title: Opinion Miner using Decision Tree and Naive Bayes Classifier

Description: The aim is to develop a system for opinion mining that identifies the sentiments/opinions expressed in the online blogs, Tweets etc. The overall task is to recognize the opinions (positive, negative or neutral from the text) expressed in a Tweet (see **Semeval009**.pdf).

Training dataset: The datasets are in the folder "Train _data". Description of the datasets can be found in README_TRAIN.txt (***please note that it describes about two tasks. You have to focus only on Task B***).

Test dataset: Test-set.txt
Gold standard test dataset (requires for measuring the accuracy**):** Test-set-gold.txt

Evaluation script: Evaluation-script.pl

Compare the accuracies between naive Bayes and Decision tree. Please go through **README-evaluation.txt** to get the ideas about the test data formats and evaluating the systems.

Classification Algorithms:

1. Naive Bayes classifier: You have to implement Naive Bayes classifier yourself (in C/C++/Java/Python...). Use the multinomial model for implementation. **Note: Do not use any library.**
2. Decision tree: Install Weka data mining package (http://www.cs.waikato.ac.nz/ml/weka/) (See WekaManual-3-7-7.pdf for more details) and use Decision Tree algorithm (C4.5 / J48).

Features to implement

Please implement following set of features from the respective papers

A. Paper-Sentiment-I.pdf, Section 3.1
   a. Word & Char N-gram features
   b. Lexicons features
B. Paper-Sentiment-II.pdf, Section 2.2
   a. Emoticon features
   b. Encoding features

Evaluation:

Max Points for the Project-25

Resources/documents to be provided:

(i). Codes with proper documentation
(ii). Report of not more than 2-3 pages with Abstract, Introduction, Problem Description, Methods, Experiments, Conclusion and Bibliography.

Deadline for submission- November 04, 2016

Submit zipped folder at: ai@iitp.ac.in (mention your group number clearly in the subject as well as in the zipped file name.)