# Sentimental Analysis

## Implementation :

1. In the given data, use all the words which represent adjectives (bag-of-words representation) from the training corpus and do the preprocessing of text like removing stop words , punctuations etc.
2. Compute the likelihood for all the features in a given class
   Determine s* of text T such that **s* =** argmax $_{si}$ p($s_i$ | T) for $s_i$ **ϵ** { neutral, positive, negative}
   $$p(s_i \mid T) = \frac{p(T|si)\,p(si)}{p(T)}$$
   where p($s_i$ | T) is posterior probability, p($s_i$) is prior probability and p(T|$s_i$) is likelihood.

3. Compute posteriors for all features and also ensure that you apply a smoothing Laplace feature to those words that are not available in the training corpus.
4. Predicting the sentiments based on map 3 values and 5 values scale.

**Performance Results :**

**F1 - score** is used to check the performance of this sentimental analysis.

| Map value | Dev | Train |
|---|---|---|
| 3 | 0.4651589043336535 | 0.5597503385205003 |
| 5 | 0.2135330136821063 | 0.32883495209517594 |

Other metrics such as accuracy, precision, recall etc can be used based on the business problem. Here F1 score is the harmonic mean of precision and recall. So we will choose f1 score to check the naive bayes classifier approach problem.

## Conclusions:

1. Naive Bayes works well in all such tasks.
2. We can improve the solution by considering verbs, intensifiers, negations and nouns to correctly predict the class such as verb - hate, dislike, intensifiers - very, much, a lot, negation - not, nouns - love , creativity.
3. We can also consider the algorithms such as SVM, Neural networks, no assumption of statistical independence of all the features in this. This can be complex but give better results.