1. **Aim: Write a program to implement Naïve Bayesian classification**

**Step 1:** Download and Install Weka

          Link: <http://www.cs.waikato.ac.nz/ml/weka/downloading.html>

**Step 2:** Download sample data

          Link: [http://www.cs.cornell.edu/people/pabo/movie-review- data/mix20\_rand700\_tokens\_cleaned.zip](http://www.cs.cornell.edu/people/pabo/movie-review-%20%20%20%20%20data/mix20_rand700_tokens_cleaned.zip)

**Step 3:** Convert data into arff format

i) Run Weka

          ii) Choose “Simple CLI”

iii) Run the command in the command line:

java weka.core.converters.TextDirectoryLoader -dir "P:\tokens" > "P:\weka

data\sample.arff"

**Step 4:** Pre-process data and save processed data

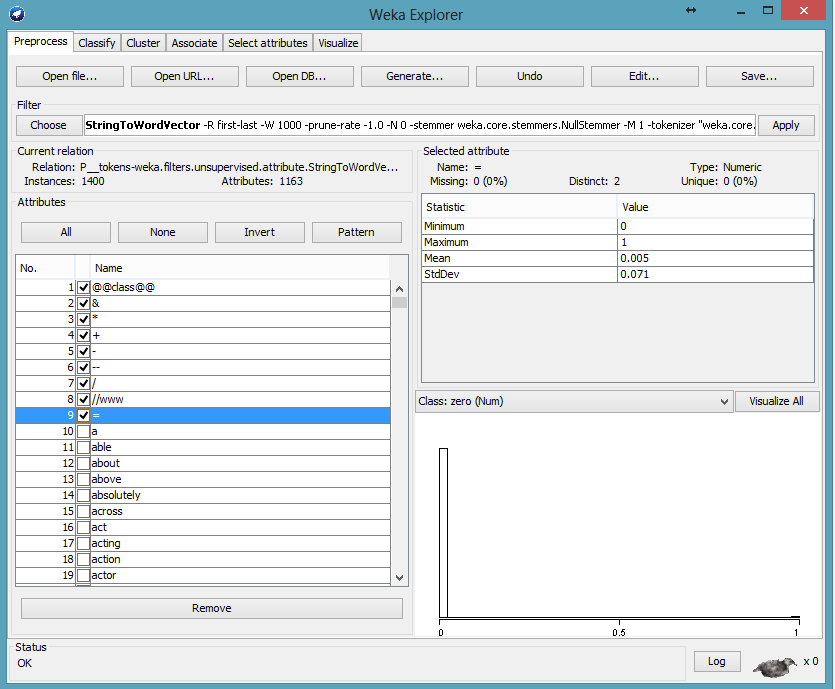
i) Choose “Explorer” from "Weka GUI Chooser"

ii) Load the arff file using “open file” option

iii) Choose a filter “StringToWordVector” and then "apply"

iv) Remove unexpected “words” using “pattern” and then “remove”

v) Remove other unexpected words manually

[](http://1.bp.blogspot.com/-smFWVXrNlak/VRJ_oUf9wJI/AAAAAAAAAn4/QbMy3rhv1iI/s1600/unexpected+words.png)

vi) Save the processed data using “save” option

**Step 5:** Run Classifier (as example : Naive Bayes)

          i) Select Classifier and method

   ii) Select evaluation option (click right to choose option)

**Step 6:** Result

i) Set output options using "more option" and run the classifier using "start"

ii) Save Result

