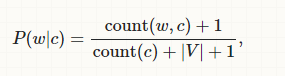
**Report**

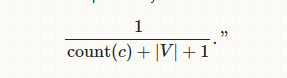
**On**

**Bayesian learning for classifying Netnews text articles**

1. Packages used NLTK , Python 2.7 .
2. Steps:
   1. Remove Stopwords and special Character, stemming the words
   2. Created Global Dictonary for each Class and stored total number of words and total count in each class
   3. Calculated Unique Words in all Class
   4. For Testing Set, Remove stopwords and special characters, stemming the words
   5. Created a list of words in that Document
   6. Calculated the probability using below formula.



* 1. Calculated new feature using laplacian smoothing formula:



* 1. Took the overall math.log of the probability and return the max value among the probabilities for each class
  2. And Calculated the accuracy using the true predication / Total number of inputs

1. Accuracy:
   1. Training set = 500 and Testing \_set = 500
      1. Accuracy: 78.97%
   2. Training set = 700 and Testing \_set = 300
      1. Accuracy: 80.81%
2. To run this Program:
   1. Update the “training\_set” variable with your new path training set
   2. Update the “testing\_set” variable with your new path testing set

Note Folder Structure should be 20\_newsgroups/20\_newsgroups/ Class\_Name/Files