

# Assignment

## Text Analytics

**Question:** Build a Naïve Bayes Classifier on Newsgroup dataset.

Pick the top 5000 words as the dictionary. Learn  $P(w|c)$  for all words and classes.  
Compute the training and test set accuracy of the model.

**Dataset:** 20news-19997.tar (on LMS)