

Problem 2

- Text classification:
 - This data set contains 1000 text articles posted to each of 20 online newsgroups, for a total of 20,000 articles. For documentation and download, see <http://www-2.cs.cmu.edu/afs/cs/project/theo-11/www/naive-bayes.html>.
 - The "label" of each article is which of the 20 newsgroups it belongs to. The newsgroups (labels) are hierarchically organized (e.g., "sports", "hockey").
 - You should provide model evaluation results and discuss the reasons of the results.