

Problem description:

1. Find 100 CNN news articles online (try to find them in different categories, e.g., sports and finance). You need to find the new article by yourself. Pls ignore picture or other non-text data in the new article.
2. Convert them to data matrix (each row is an article and each column is a unique term).
3. Run kNN and Decision Tree with 5-folder cross validation. (label each cnn new article with the category)
4. Evaluate classification results with accuracy, F-measure

Deadline: please submit your results by 02/20 . You may compress your 100 news articles, data matrix, classification results and your own discussion into a zip file to submit.