TopicModeling\_Adobe.ipynb

* Code file for computing coherence scores, performing topic modeling techniques, utilizing the filtered text containing the top 1000 important words.

category\_testing.ipynb

* Shows % overlap between new (NMF) categories and the original categories, the percent similarity via accuracy, and the percent similarity by category.

class\_important\_word\_comparisons.csv

* Shows important words for all categories, both in the original and new (NMF) category sets. The top 100 words as identified through regularized multinomial logistic regression are shown for each category. This should be used to gauge whether one category set is more valid than another.

class\_word\_importances.ipynb

* Code file used to produce class\_important\_word\_comparisions.csv.

nmf\_categories\_apr17.csv

* Output of the NMF topic modeling process. Column “type” indicates the text’s original category, while “topic” indicates the text’s NMF assigned category.