

# **Team GOAT! (Getting Our Act Together)**

## **Reproducing Paper: A cross-collection mixture model for comparative text mining**

### 1: Progress

As a group we have decided to choose a rather large dataset from the New York Times regarding the presidential election. Our paper that we are modeling also used the Expectation-Maximization (EM) algorithm to evaluate a model on two different text data sets one being a news article data set.

We are currently doing more research on the EM algorithm as this was a rather recent lesson in our class. Obviously, we have a GitHub repository to collaborate on for our coding portions. Also using Google Docs for the paper creation.

### 2: Remaining Tasks

We have to implement our algorithm and then use an experiment to verify our implementation.

### 3: Challenges/Issues Faced

Unfortunately, one of our group members lost a family member over the break, so we had to reallocate some of the workload recently. Other than that, we are having issues figuring out helpful libraries and strategies so that our EM algorithm can work on large datasets.