Progress Report

So far, we have researched the sources as mentioned in the paper. In order to keep this project as similar to the paper as possible, articles were pulled from the websites mentioned in the paper. This includes war news from CNN and BBC sites online and getting reviews from epinions.com. Since epinions no longer exists as a domain, an archive of the internet was used in order to extract data from the approximate time period the paper was published.

This led to a couple of complications: it was unclear in the paper what articles/reviews in particular were used despite mentioning the website they were taken from. While internet archives are available to extract data from that period of time, the archives are not comprehensive. Thus, we were not able to get all of the reviews mentioned in the paper (though we believe a sufficient number of them have been obtained to still run the model in a meaningful manner).

Additionally, since the specific articles/reviews were not released in the paper, it is unclear whether our extracted data matches the data used for the model in the paper. Regardless, we expect the content to be very similar, so this should not affect the results.

The only task that is left to be done is to implement the model itself. This involves two parts: creating the naive mixture model and creating the mixture model described in the paper. This will likely be in a format similar to that of MP3, so it should be a doable amount of time.