

1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.
2. Which paper have you chosen?
3. Which programming language do you plan to use?
4. Can you obtain the datasets used in the paper for evaluation?
5. If you answer “no” to Question 4, can you obtain a similar dataset (e.g. a more recent version of the same dataset, or another dataset that is similar in nature)?
6. If you answer “no” to Questions 4 & 5, how are you going to demonstrate that you have successfully reproduced the method introduced in the paper?

The group captain of this project is Jonathan Kim (jck3) and the other two members of the group are Michael Xiang (mx9) and Tyler Ruckstaetter (thr2).

The paper we have chosen is “ A cross-collection mixture model for comparative text mining.”

The language that we will be using for this project is python.

Since the paper is from 2004, it seems like the exact datasets used in the paper will be unavailable for us. However, a similar dataset will be obtainable since the nature of the data is very publicly available. As of right now, we can still use reviews of laptops as used in the paper as well as more modern laptops. War news is available on many news sites, so while the exact documents from the paper may not be directly used, they will be very similar in nature.