

Progress Report

Captitan: Mack Teng (mact2)
Member: Hsin Hsien Chung (hhc3)

Progress:

- Finished parsing and filtering the data sets
- Finished building the matrices
- Implementing cross-collection mixture model EM algorithm

Challenges:

- Originally, we were planning to crawl data from the web, but it took longer than expected. We used existing data sets from Kaggle.com instead.
- Some data are duplicated
- For the mobile phone review data set, some of the comparison is not valid. Because some mobile phone models are outdated, but we do not have starting sales time or review time to distinguish them, we need to come up with a way to make a valid set for comparison. Therefore, we filtered the data with price range.

Remaining Tasks:

- Might need to compare with filtered data sets or unfiltered data sets
- Add documentations
- Create presentation