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.

Ardel Kardi (pnkardi2) - Captain
Putra Firmansyah (pdf2)

2. What topic have you chosen? Why is it a problem? How does it relate to the theme and to the class?

We have chosen intelligent browsing and specifically, we will make an intelligent bookmark for movies that can be appended by many different other users (in real life would be like a friend group) and recommend movies based on the choices. The recommendation will be based on similarity between synopsis of movies and recommend a movie. It is a problem because currently, there are not many well-known technologies that can be used by many users on a common platform, in this case on chrome, to pick their favorite movies and recommend movies that they can add to the bookmark and, say, be watched on a movie night. It relates to the theme as it personalizes browsing for people so that they can find more relevant movies to watch. It relates to the class as we are implementing a recommendation system which is a push retrieval system which is a big section of the class.

3. Briefly describe any datasets, algorithms or techniques you plan to use

The technique that we plan to use is content-based filtering. In order to do this, we plan to find similar movies based on the list of movies that we have. Specifically, we will use the movie synopsis which will act as our dataset that can be taken from IMDB and find similar movies that will be recommended to the bookmark. We plan to use a mixture model estimation with

4. How will you demonstrate that your approach will work as expected?

We will generate offline metrics such as F-score or MAP from relevance judgements sessions where the movies returned as similar will be scored by the judges (the two of us) as either relevant or not-relevant.

5. Which programming language do you plan to use?

We plan to use Python as that is what we are most familiar with and also how we have done all our MPs in.

6. Please justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.
   Total: 48 hours

- Create data scraper from all movies on IMDB: 2 hours
- NLP and text representation: 10 hours
- Word association mining and analysis: 8 hours
- Topic mining and analysis: 3 hours
- Opinion mining and sentiment analysis: 5 hours
- Recommender system: 10 hours
- Frontend: 10 hours