1. Dataset:

IMDB MOVIE REVIEWS DATASET

<https://ieee-dataport.org/open-access/imdb-movie-reviews-dataset>

This dataset contains nearly 1 Million unique movie reviews from 1150 different IMDb movies spread across 17 IMDb genres - Action, Adventure, Animation, Biography, Comedy, Crime, Drama, Fantasy, History, Horror, Music, Mystery, Romance, Sci-Fi, Sport, Thriller and War. The dataset also contains movie metadata such as date of release of the movie, run length, IMDb rating, movie rating (PG-13, R, etc), number of IMDb raters, and number of reviews per movie.

Analysis:

1. Avg genres rating
2. length of movie v.s. timeline
3. users’ preference
4. Movie numbers per genres
5. data distribution -> Normalization
6. Movie rating timeline
7. Tasks:
   1. Movie per user prediction
8. Would watch
   1. Baseline
      1. Jaccard Similarity
      2. BPR
   2. Improved
      1. NN
9. rating prediction
   1. Baseline
      1. Latent Factor without ru, ri
      2. Latent Factor
   2. Improved
   3. Movie rating prediction

Comparison:

1. Normalization

Features:

1. Users relationship
2. Movies relationship
3. Relations between genres
4. Number of rates
5. Number of reviews
6. Review length
7. Movie release date
8. Riview commented date
9. Helpful/Not helpful
10. (Review text length)