Machine Learning-I (CS/DS 706): Proposal

Books Recommendation - Goodreads Data

https://www.kaggle.com/zygmunt/goodbooks-10k

Chetna Jain (MT2016044)
Kanika Narang (MT2016069)
Tehreem Ansari (MT2016145)

DATA

The dataset contains ratings for ten thousand books. Generally, there are 100 reviews for each book, although some have less - fewer - ratings. Ratings go from one to five. Both book IDs and user IDs are contiguous. For books, they are 1-10000, for users, 1-53424. All users have made at least two ratings. Median number of ratings per user is 8.

There are two files in the datasets. They are as follows :

1. **ratings.csv file**
   It contains 3 columns book_id,user_id,rating and it looks like that:

```
book_id,user_id,rating
1,314,5
1,439,3
1,588,5
1,1169,4
1,1185,4
```

2. **Books.csv file**
   It provides title and goodreads IDs for each book. It contains four columns book_id, goodreads_title, goodreads_book_id, goodreads_work_id. It looks as follows:

```
book_id,goodreads_title,goodreads_book_id,goodreads_work_id
1,"The Hunger Games (The Hunger Games, #1)",2767052,2792775
2,"Harry Potter and the Sorcerer's Stone (Harry Potter, #1)",3,4640799
3,"Twilight (Twilight, #1)",41865,3212258
4,To Kill a Mockingbird (To Kill a Mockingbird #1),2657,3275794
5,The Great Gatsby,4671,245494
```

goodreads_title: Title of the book
goodreads_book_id: Points to the most popular edition of a given book.
goodreads_work_id: Refers to the book in the abstract sense.

Problem Statement : Recommend books to the user.

We need to recommend the books the user should read next.
We will be creating **Collaborative Recommender system**: Collaborative recommender systems aggregate ratings of books, recognize commonalities between the users on the basis of their ratings, and generate new recommendations based on inter-user comparisons.