Personalized Book Recommendation System  
**Team BARtS**: **B**ibhabasu Das, **A**bhranil Chakrabarti, **R**achi**t** Jain, **S**hreya Gupta

**Description**

The objective of this project is to build a “Book Recommendation System” that relies on user information on past books reviewed by them and book meta-data to give recommendations to users on the next book to read, both at an overall and genre level.

**Data Used** Kaggle Data**:** 271k books, 279k Users, 1.1MReviews

* **Book Data Labels**: ISBN, Book-Title, Book-Author, Publisher, Year-of-Publication, Image-URL
* **Rating Data Labels**: UserID, ISBN, Rating; **User Data**: User-ID, Location, Age

Since there isn’t a lot of book meta-data in this dataset, we extract information on ‘Book Description’, ‘Also Bought’ and ‘Also Viewed’ using data from Amazon.

**Approach Used**

**Key assumption:** If a user hasn’t rated a book, he/she hasn’t read the book

**Predictive Methods:** Use regression to predict the ratings of all books by a user and recommend the books with best ratings. Methods used: Linear Regression, Holistic Regression, XGBoost

**Recommendation Methods:**

* **Popularity-Based:** Recommend the most popular book in a genre/by an author to a user
* **Content based:** Use text analyticson the description, title, author and other field of the books rated by the user to find similar books and recommend the top-rated book
* **Identifying Archetypal User:** We classify our users into *X* archetypes and based on the archetype into which a new user falls, we recommend the books that have been rated the best by this archetype and which hasn’t been rated by the user yet
* **User-Item Collaborative filtering (SVD)**: We find the users most similar in behavior to current user and recommend the top-rated books by these users
* **KNN Based:** Create compressed sparse user-book matrix (for each user and each book) and apply KNN to find recommendation for given user
* **Hybrid:** Combine collaborative and content-based approaches to find recommendations using collective matrix hybridization

**Results**

We consider the RMSE for the difference between generated rating and observed ratings for a given user-book combination as a measure of performance. XGBoost gives an RMSE of 0.78 and user-item collaborative filtering gives an RMSE of 0.81.

We use optimization to identify the best approach that minimizes the prediction error and generate final recommendations based on this selection.

**Impact**

The recommendation system can be used by a marketplace as a part of its push strategy to understand its users, recommend them books while also deciding price point of these books dynamically (based on trending prices in the book category and user willingness to pay historically).

The system can also be used by libraries to support search for relevant books. This system is agnostic to product category and can be embedded as a part of customer experience improvement for any platform, thereby empowering the customer with reliable and personalized information.