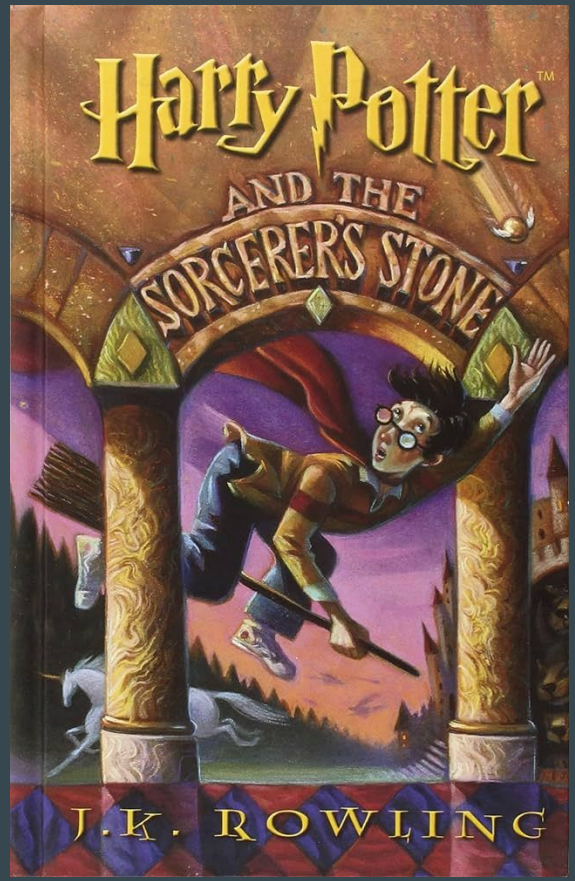
Book Recommendations

Using Singular Value Decomposition to Create A Recommendation System

Problem Identification

Cold Start Problem

- Difficult to personalize recommendations for new users
- Most popular books become default recommendation
  - Harry Potter and Hunger Games
- Lack of item data like genre for item-based or hybrid model
- Ask users for some help to get started



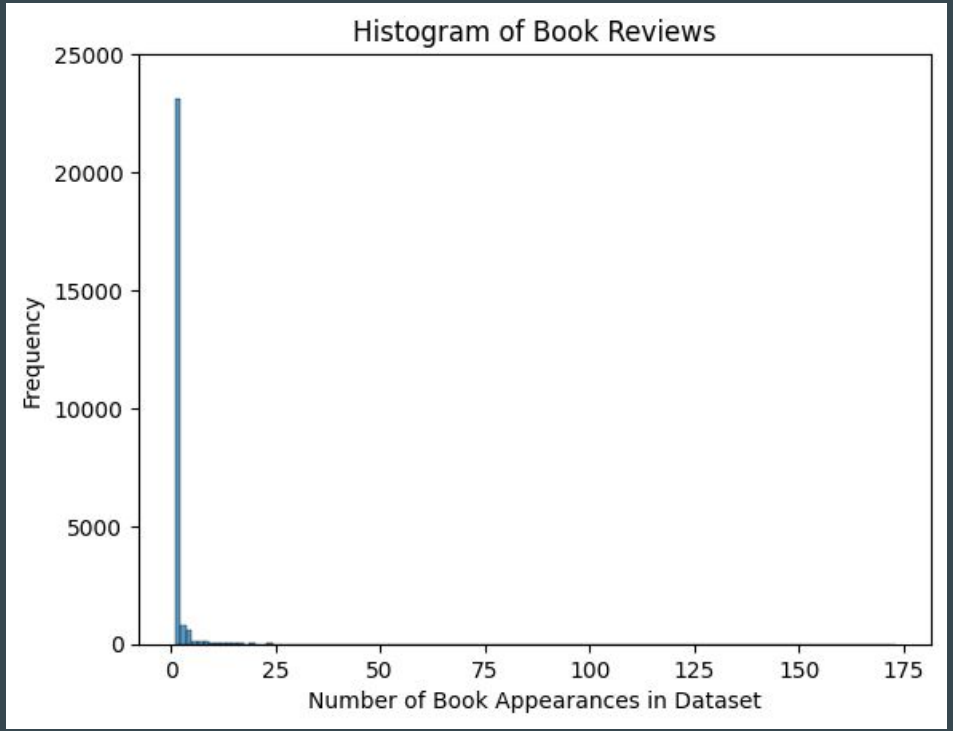
How do I provide users with personalized book recommendations in under 30 seconds from a list of their 5 favorite books?

Problem Statement

Data Collection & Exploration

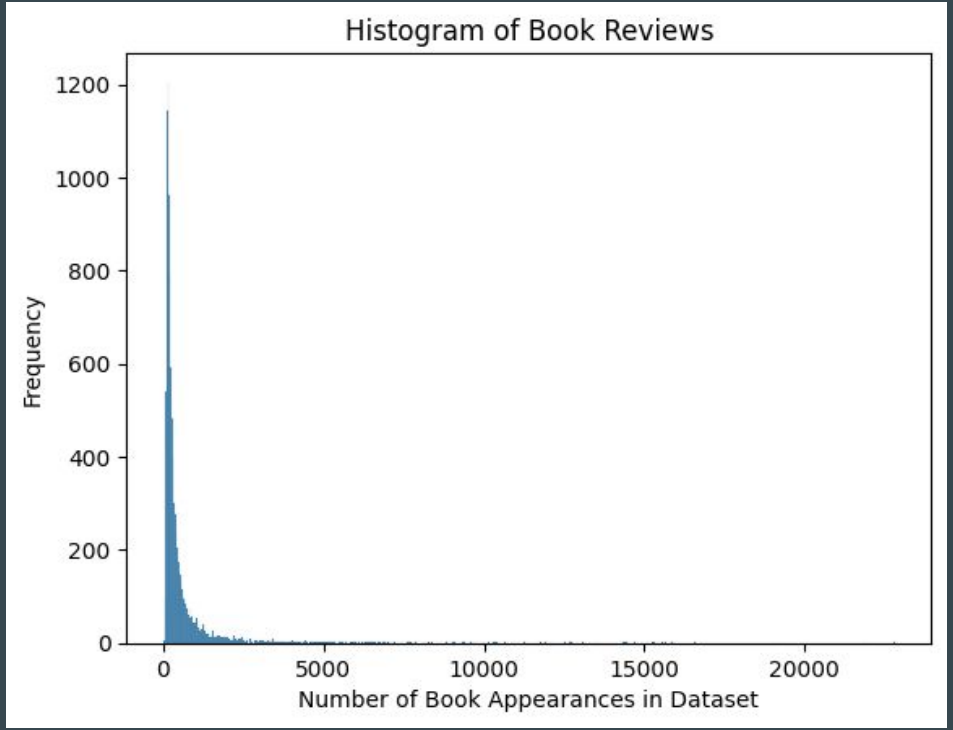
Goodreads Scraping

- Time consuming (~20s per user)
- 300 users, 25,000 books, 55,000 ratings
- Only 116 books with more than ten reviews
- Abandoned due to poor user diversity



Goodbooks Dataset

- GitHub download
- 53,000 users, 10,000 books, 6 million ratings
- 1901 books with more than one review
- Performed well with diverse users



Modeling

Model Selection

- SVD outperforms other algorithms on test dataset
- Hyperparameter tuning yields optimal parameters
- Impossible to evaluate performance on new users, which makes this a unique problem

Algorithm	test_rmse	test_mae	fit_time	test_time
SVD	1.070495	0.753834	0.489696	0.052653
BaselineOnly	1.070785	0.762052	0.092678	0.034303
SVDpp	1.076974	0.753623	7.935296	0.932344
KNNBaseline	1.102457	0.774174	0.154247	0.382105
KNNWithZScore	1.206980	0.853376	0.060582	0.267725
KNNWithMeans	1.207775	0.857880	0.041189	0.288606
SlopeOne	1.221169	0.869391	8.487853	0.789540
CoClustering	1.243652	0.890011	2.079022	0.063248
NMF	1.258026	0.919447	1.361211	0.052810
KNNBasic	1.277513	0.902087	0.038896	0.274676
NormalPredictor	1.648319	1.243276	0.039295	0.034303

Prediction Strategy

Overcoming the Cold Start Problem

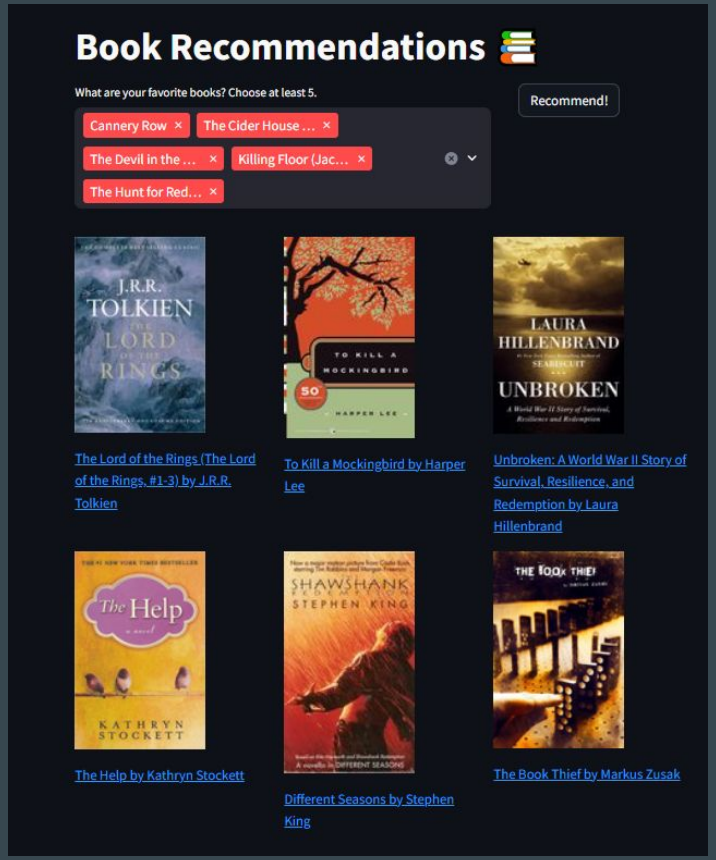
- Filter dataset for all users with at least 3 books in common with new user
- Add new user, rate all of their favorite books 5 stars
- Fit SVD model to full dataset, predict new user's top 6 books

Storing and Accessing Data

- Create local SQLite database
- Create tables for Ratings & Books
- Look up book ID from title, look up ratings from book ID
- Return title, author, image, and URL from book ID

Deployment

- Streamlit hosts the webapp
- GitHub Large File Storage hosts database
- Simple web interface with auto-complete book titles
- Links to recommended books



Further Improvements

- Error handling
  - Detect when not enough similar users, prompt for more books
- Book series filtering
  - Don’t recommend books from a series the user has read
- Living database development
  - Add users and their favorite books, let them rate recommendations
- Enhance UI
  - Improve visual balance, better color schemes