SVD-Based Book Recommender Using ISBN

Aatif Ahmad (B22AI002), Anushk Gupta (B22AI007),
Yashraj Chaturvedi (B22AI059), Omprakash Nain (B22AI062)
Indian Institute of Technology Jodhpur

April 2025

Abstract

We developed an ISBN-based book recommender system using collaborative filtering with Singular Value Decomposition (SVD). By analyzing a large dataset of user ratings and book metadata, the system predicts top 5 similar books given a target ISBN. The system is deployed as an interactive Streamlit web app at: `https://svd-book-recommender.streamlit.app/`

1. Introduction

Book recommendation systems help readers discover new books based on their interests. Our project focuses on collaborative filtering using matrix factorization with SVD to find book similarities based on reader ratings.

2. Dataset Overview

We used the **Book-Crossing Dataset** from Kaggle, containing:

- `books.csv`: 271,379 entries
- `ratings.csv`: 1,149,780 entries
- `users.csv`: 278,859 users

2.1 Missing Values

- **Books:** 2 missing authors
- **Ratings:** No missing values
- **Users:** 110,232 missing ages

3. Preprocessing

- Retained users with ≥50 ratings and books with ≥20 ratings.
- Removed entries with missing book titles/authors.
- Created a dense user-item rating matrix.

Post-filtering:

- Books: 4,493 (**1.66%** of original)
- Ratings: 201,726
- Users: 3,312 (**1.19%** of original)
- Rating Matrix: 3312×4493
- Missing values in rating matrix: 0

4. Methodology

We used matrix factorization with SVD:

$$R \approx U\Sigma V^T$$

Where:

- R: user-book rating matrix
- U: user latent matrix (3312×50)
- Σ: diagonal singular values (50×50)
- V^T: book latent matrix (50×4493)

We used the top 50 latent factors to capture key user preferences.

5. Recommendation Strategy

1. Get latent vector of input ISBN from V^T
2. Compute cosine similarity with all other book vectors
3. Recommend top 5 most similar books (excluding input)

6. Results

6.1 Sample ISBNs

ISBN	Title	Author
000649840X	Angelas Ashes	Frank McCourt
0007154615	Unless: A Novel	Carol Shields
0020198906	Joshua	Joseph F Girzone
0020199600	The Great Gatsby (Reissue)	F. Scott Fitzgerald
002026478X	Age of Innocence (Movie Tie-In)	Edith Wharton

6.2 Recommendations for ISBN 000649840X (Angela's Ashes)

Title	Author	Similarity
Shipping News	E Annie Proulx	0.825
Jitterbug Perfume	Tom Robbins	0.821
Kentucky Rich	Fern Michaels	0.749
I Heard the Owl Call My Name	Margaret Craven	0.742
Sabriel	Garth Nix	0.733

6.3 Recommendations for ISBN 0020199600 (The Great Gatsby)

Title	Author	Similarity
Spy Hook	Len Deighton	0.912
Out of Africa / Shadows on the Grass	Isak Dinesen	0.911
Love and War	John Jakes	0.911
The Two Mrs. Grenvilles	Dominick Dunne	0.909
A Woman of Substance	Barbara Taylor Bradford	0.904

7. Streamlit App

The recommender system is deployed at:

`https://svd-book-recommender.streamlit.app/`

Features:

- Enter ISBN to get book info and 5 similar titles.
- Suggestions displayed with author and similarity.

- Responsive design with sidebar sample ISBNs.

The source code for the project is available on GitHub at:

`https://github.com/aatifahmad123/svd_book_recommender`

8. Contributions

- **Aatif Ahmad**: SVD implementation, recommender logic, Streamlit frontend/backend integration.
- **Anushk Gupta**: Data filtering, matrix construction, evaluation.
- **Yashraj Chaturvedi**: UI/UX for app, ISBN validation, deployment.
- **Omprakash Nain**: Testing, visualizations, and debugging.

9. Conclusion

The SVD-based recommender performs well in identifying similar books using user rating patterns. Despite filtering down to just 1–2% of the original data, the system provides meaningful results and is accessible via a web app.

10. References

- Book-Crossing Dataset: `https://www.kaggle.com/datasets/somnambwl/bookcrossing-datas`
- Funk, S. (2006). Netflix Update: Try This at Home.
- Ricci, F., Rokach, L., Shapira, B. (2011). Recommender Systems Handbook.