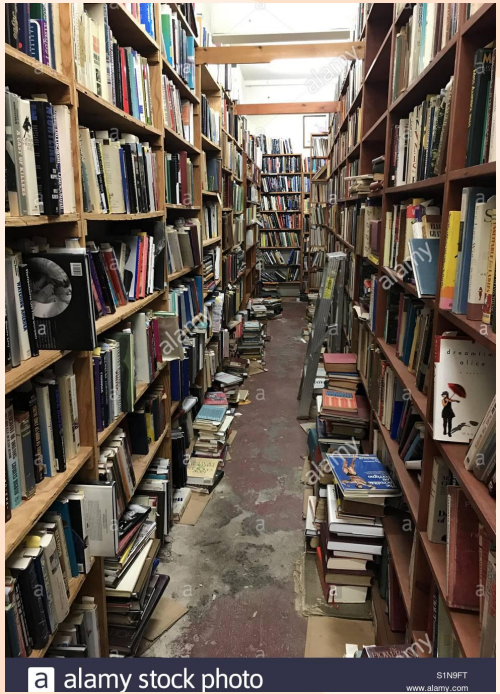
Book Recommender Systems

“If you don’t like to read, you haven’t found the right book.” – J.K. Rowling

Liew Chin Xia (DSI 21)

Where to begin?



https://www.alamy.com/stock-photo-the-old-bookstore-310869772.html
https://www.123rf.com/photo_29271029_3d-rendering-of-thinking-frastrated-man-with-red-question-mark-3d-white-people-man-character.html

Reading books?

Available books globally

> 130 million books (2010)

Estimated books a person reads per year

12-120 books

Number of books published in a year

2.2 million books

Process of selecting a book

- Social media
- Online browsing
- Book referrals

10 Best Book Recommendation Sites You Need To Know

Royale Scuderi

A creative strategist, consultant and writer who specializes in cultivating human potential for happiness, health and fulfillment. Read full profile

Share | Pin it | Tweet | Share | Email

Looking for a great summer read? Want to read more on a topic you're interested in or see what's new or trending in the book world? There are plenty of places to turn for book recommendations on the internet. Most are simple and free and best of all, they'll help prevent that feeling of reader's remorse, when you get home from the bookstore or library, start reading and realize that you've just wasted your time and possibly your hard earned money on a lousy book in which you have no interest in reading.

https://www.lifehack.org/articles/technology/10-best-book-recommendation-sites-you-need-know.html

GatesNotes THE BLOG OF BILL GATES LOG IN SIGN UP

GREAT ESCAPES

5 summer books and other things to do at home

Whether you're looking for a distraction or just spending a lot more time at home, you can't beat reading a book.

By Bill Gates | May 18, 2020 • 6 minute read

https://www.gatesnotes.com/About-Bill-Gates/Summer-Books-2020

Problem Statement

Readers faced with vast availability of book choices, and it takes huge amount of time, effort and energy to select a suitable book.

As a data scientist in the online bookstore, we aim to build a book recommendation system in our online bookstore website to help readers save up time to search for the books based on their preference.

The recommender system's success will be evaluated based on the root mean square error (RMSE) between the actual and the predicted rating.

Overview

Data Collection and cleaning

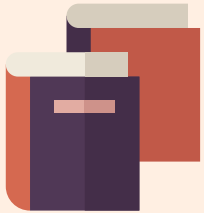
Recommender System

Exploratory Data Analysis

Conclusion and Recommendation

Datasets

(Source: UCSD Book Graph)



Metadata of books

- 6 datasets

  (books, best book edition, genre, author, book series, book id map)
- 2.3 million books
- 829 thousand authors
- 10 genres

User-book interactions

- 228 million interactions
- 876 thousand users
- 2.36 million books
- Ratings (1 to 5)

Exploratory Data Analysis

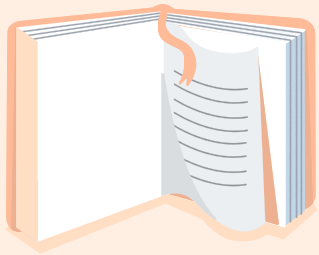
Majority of the Books Characteristics in the Dataset

150-350 pages
(Range: 0-300 000 pages)
Number of books pages

Fiction
Genre

<10 book editions
(Range: 1-3676 books)
Number of book editions

~3 books
(Range: 0-738 books)
Number of books in a series



Undefined, smashwords edition, harlequin and createspace
Book publisher

Paperback, undefined, hardcover, ebook
Format book types

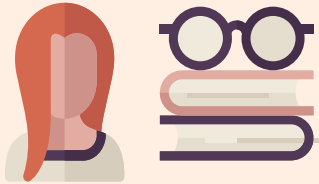
Undefined, english
Language

Increasing trend from 1950 to 21st century
(Range: Prior BC to beyond 2017)
Publishing year

User-book Interactions in the Dataset

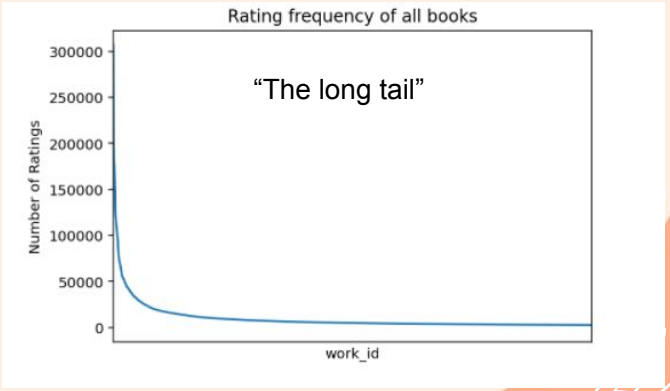
Majority of books read and rated are not reviewed

Average rating: ~4.0 (Range: 1.0–5.0)



Majority of the users rated < 150 books
(Range: 0–17000 books)

Majority of the books rated < 20 users
(Range: 1–307,000 users)



Challenges

Memory Error!

Computational Limitation

Threshold on the user-book interactions were set for

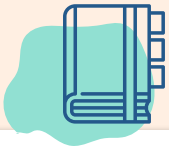
- Ratings of a book required
- Number of books rated by a user

Acquire a subset fraction of the full dataset for further analysis.

Recommender System

- Popularity-based
- Content-based

Non-personalized Recommendation



Collaborative Filtering
- User-based
- Item-based

Personalized Recommendation

Model-based System

- Non-negative Matrix Factorization
- Singular Value Decomposition (SVD)
- SVD ++

Personalized Recommendation

01

Recommender System

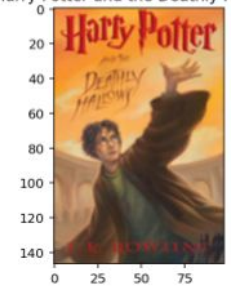
Non-personalized Recommendation

- Popularity-based
- Content-based

Popularity-based Recommender System

Top 10 recommended books based on popularity-based recommender system

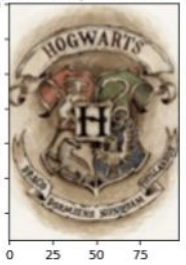
1
J.K. Rowling
Harry Potter and the Deathly Hallows



2
J.K. Rowling
Harry Potter and the Prisoner of Azkaban

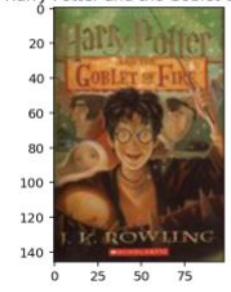
3
Sarah J. Maas
A Court of Mist and Fury

4
J.K. Rowling
Complete Harry Potter Boxed Set



5
J.K. Rowling
Harry Potter and the Half-Blood Prince

6
J.K. Rowling
Harry Potter and the Goblet of Fire

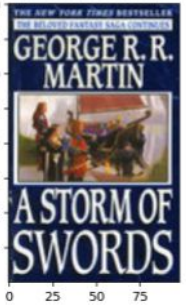


7
J.K. Rowling
Harry Potter and the Philosopher's Stone

8
Cassandra Clare
Clockwork Princess

9
J.K. Rowling
Harry Potter and the Order of the Phoenix

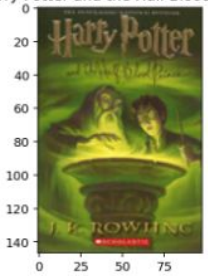
10
George R.R. Martin
A Storm of Swords



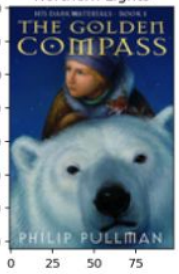
Content-based Recommender System

Top 10 recommended books based on content-based recommender system for J.K. Rowling - Harry Potter and the Deathly Hallows

1
J.K. Rowling
Harry Potter and the Half-Blood Prince

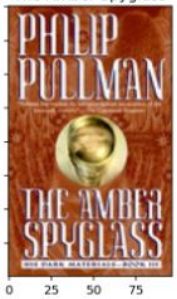


2
Philip Pullman
Northern Lights



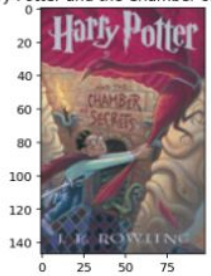
3
J.K. Rowling
Complete Harry Potter Boxed Set

4
Philip Pullman
The Amber Spyglass



5
J.K. Rowling
Harry Potter and the Prisoner of Azkaban

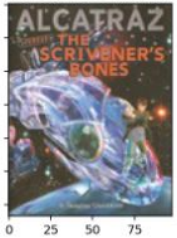
6
J.K. Rowling
Harry Potter and the Chamber of Secrets



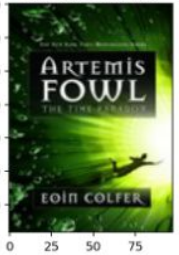
7
China Mieville
Un Lun Dun

8
J.K. Rowling
Harry Potter and the Philosopher's Stone

9
Brandon Sanderson
Alcatraz Versus the Scrivener's Bones



10
Eoin Colfer
The Time Paradox



02

Recommender System

Personalized Recommendation

- Collaborative Filtering
- Model-based Systems

Collaborative Filtering Recommender System

Dataset

User-based

Collaborative Filtering (Cosine similarity)

Item-based

Predicted-rating

	RMSE	Coverage
Null (Random)	1.9265	100.0%
User-based	1.2720	7.0%
Item-based	1.2663	8.4%

K-means-Collaborative Filtering Recommender System

Dataset

K-means Clustering by Genres

Fiction

Romance

User-based

Collaborative Filtering (Cosine similarity)

Item-based

Predicted-rating

	RMSE	Coverage
Null (Random)	1.9265	100.0%
K-means User-based	1.0886	55.0%
K-means Item-based	1.0454	58.4%

Model-based Recommender System

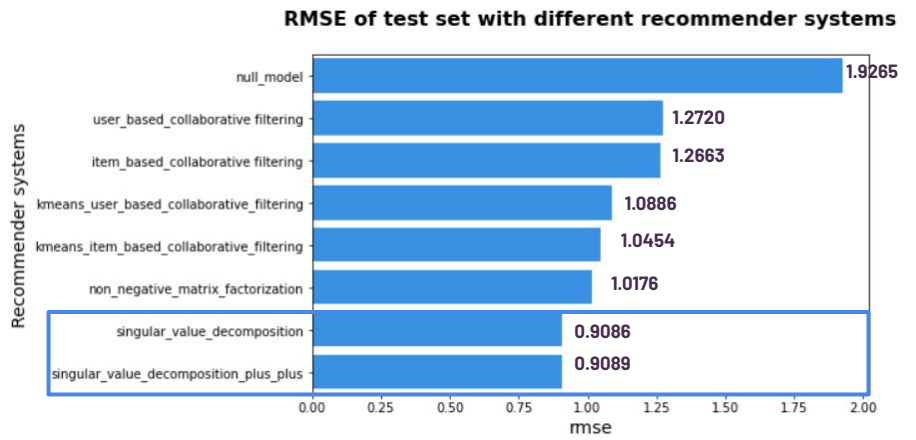
Dataset

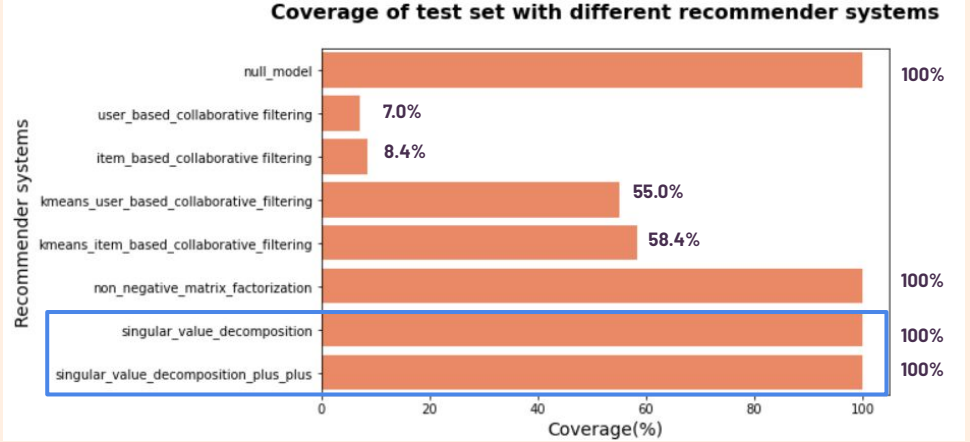
Model-based Systems

Predicted-rating

	RMSE	Coverage
Null (Random)	1.9265	100.0%
Non-negative Matrix Factorization	1.0178	100.0%
Singular Value Decomposition	0.9086	100.0%
Singular Value Decomposition ++	0.9089	100.0%

Evaluation





Evaluation: Root Mean Square Error (RMSE)

Chosen Model: Singular Value Decomposition

Conclusion & Recommendations

Exploratory Data Analysis

- Discovered majority of the books characteristics in the dataset
- The user book interactions illustrates the long tail scenario

Computational Limitation

- Applied threshold and obtained a subset fraction of the dataset

Non-personalized Recommendation

- Popularity-based
- Content-based

Personalized recommendation

- Collaborative filtering
- Model-based

Selected Model

- Singular value decomposition algorithm (Lowest RMSE)
- Improvement of 53% from the null model

Recommend Books

- Predict ratings more accurately
- Enhanced reader's experience

Future Work

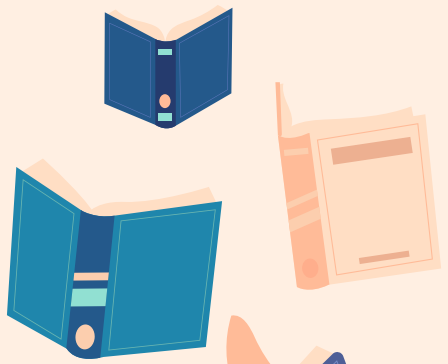
- Explore more algorithms:
  - Further improve the RMSE score.
  - Handle greater user-book-interactions with good scalability and faster processing speed.

- Implicit data includes user purchased history and more details on the users such age, geographical location may also be explored to provide better readers' experience.

Thank you!

“This is one of the things I love about reading. Each book opens up new avenues of knowledge to explore”

– Bill Gates



Thanks!

youremail@freepik.com | +00 240 576 092 | yourcompany.com

CREDITS: This presentation template was created by **Slidesgo,** including icons from **Flaticon,** infographics & images by **Freepik.**

Please keep this slide for attribution.