

Google Books Recommendations

COMBINING USER REVIEWS AND BOOK METADATA

Claudia Adam and Anastasiia Poverenova

AUGUST 25, 2025

Table of Contents

3	4	5	6
Project Goals	Dataset	Type of Recommender System	Models & Methods
7	8	9	10
Results	Evaluating Model Performance	Limitations & Future Improvements	Q&A

Project Goals

- 1 Build a personalized book recommender system
- 2 Address the limitations of single model recommenders by developing a hybrid model
- 3 Demonstrate practical value by providing accurate and diverse recommendations for users
- 4 Evaluate model performance with rating prediction metrics and Top-K ranking metrics
- 5 Explore scalability and applicability for large, sparse datasets

Dataset

Amazon Books Reviews

- Published on Kaggle
- Combines Amazon user reviews + Google books metadata
- Contains 420,000+ records covering 300,000+ books and 500,000+ users

Books Ratings File

- title
- description
- authors
- image
- previewLink
- publisher
- publishedDate
- infoLink
- categories
- ratingsCount

Books Metadata File

- id
- title
- price
- userId
- profileName
- helpfulness
- score
- time
- summary
- text

Type of Recommender System

1

Content-Based Filtering - recommends books with similar themes and language

2

Item-Based Collaborative Filtering - makes recommendations based on user co-preferences

3

Hybrid Model - combines normalized similarity scores from both models

Models & Methods

Content-Based Filtering

- TF-IDF vectorization with unigrams and english stopword removal
- Built sparse feature matrix of 62,000 books by 59,000 terms
- Used KNN with cosine similarity to compute item-item similarity scores

Collaborative Filtering

- Constructed a sparse user-item ratings matrix
- Applied mean-centering to each user to reduce rating bias
- Fit item-based KNN on the centered matrix

Hybrid Model

- Selected top rated item for each user
- Retrieved nearest neighbors from both models
- Normalized similarity scores to [0, 1] and applied weighted average parameter
- Tested the model at values of [0, 0.25, 0.5, 0.75, 1]
- Hybrid model performed the best at the weighted average value of 0.25

Interpretation of Results

- Our hybrid recommender model gives the Top-K ranked list of book suggestions for a given user
- If the weighted average is 0 → suggestions are semantically related books
- If weighted average is 1 → suggestions align with user co-preferences
- At the best performing weight average (0.25) → results balances thematic relevance and collaborative popularity trends
- Example below shows top recommendations that combines user interests in Jewish cooking and related Jewish holiday cookbooks

Hybrid recommendations for user 115202

- the art of jewish cooking (i=30081) score=0.750
- the mensch chef: or why delicious jewish food isn't an oxymoron (i=39312) score=0.662
- the complete american-jewish cookbook: in accordance with the jewish dietary laws (i=20867) score=0.598
- art of jewish cooking (i=50520) score=0.533
- joan nathan's jewish holiday cookbook (i=53311) score=0.471

Evaluation Metrics

Metric	Score	Interpretation
RMSE	7.12	on average, system's predictions differ from true ratings by 7.12
MAE	3.41	on average, predictions are 3.41 away from real rating
Precision	0.026	2.6% of predicted ratings were actually relevant or correct
Recall	0.26	the model found 26% of the actual positive instances in the dataset
F1 Score	0.047	model's overall performance is quite low at only 0.047
Hit Score	0.26	the algorithm makes a correct prediction 26% of the time out of all attempts
Coverage	0.013	the model recommends only 1.3% of all books in the catalog across users

Limitations

Data Sparsity

Scalability

New Users

Matrix
Factorization

Transformer
Based
Models

Implicit
Feedback

Future
Improvements

The End

THANK YOU FOR LISTENING

Claudia Adam and Anastasiia Poverenova

8/25/2025