ADVANCED BOOK RECOMMENDATION SYSTEM

Shayan Abdul Karim Khan

07/21/2023

OVERVIEW

INCREASE THE NUMBER OF BOOKS READ BY PEOPLE

MAKE IT SIMPLE AND EASY TO FIND NEW GOOD BOOKS

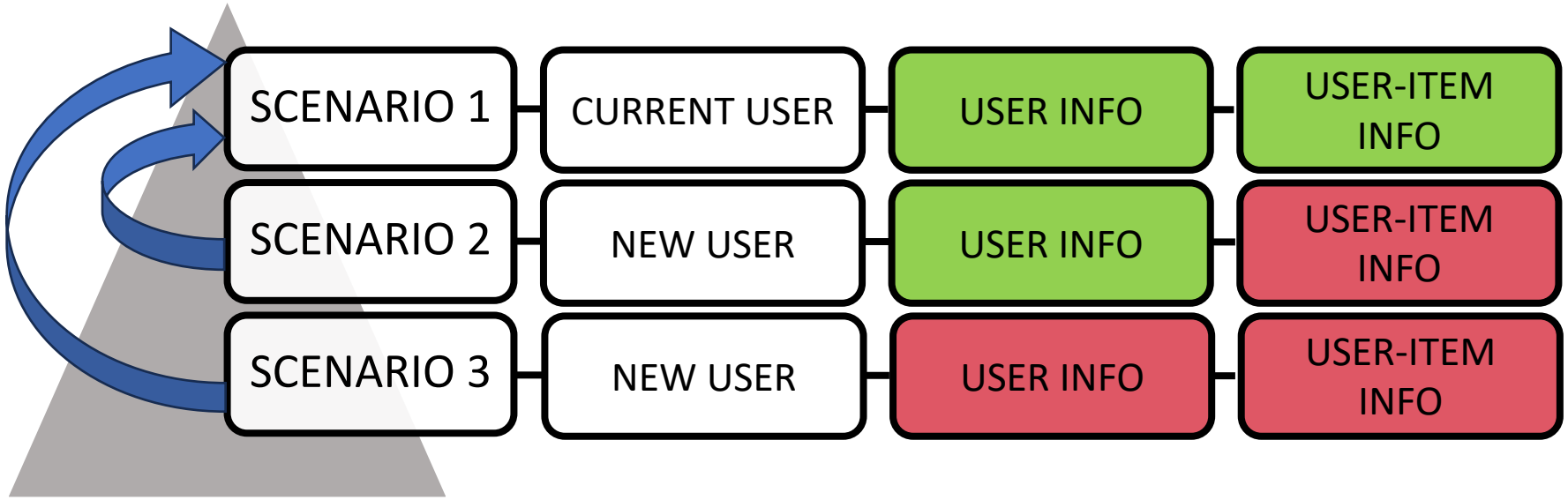
MAKE IT SIMPLE AND EASY TO EXPLORE NEW GENRES

SOLUTION

RECOMMENDATION SYSTEM
NO BOOK STORE RUNS
NO RESEARCHING BOOKS
QUARTERLY 5 BOOKS AVAILABLE

GOAL

RECOMMENDATION MODEL
5 BOOK RECOMMENDATIONS



ANALYSIS PROCESS

BOOK-CROSSING & GOOGLE BOOK API DATA

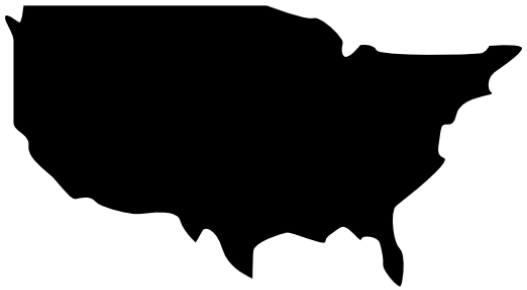
ITERATIVE DATA CLEANING & PROCESSING

EXPLORATORY ANALYSIS

TARGETTED NO OF REVIEWS > 19

AVG RATING 3.8/5

TARGETTED AGES 16 – 60



USA DOMINATED THE DATA

60% RATINGS UNKNOWN

70% OF COMPLETE DATA USED FOR TRAINING

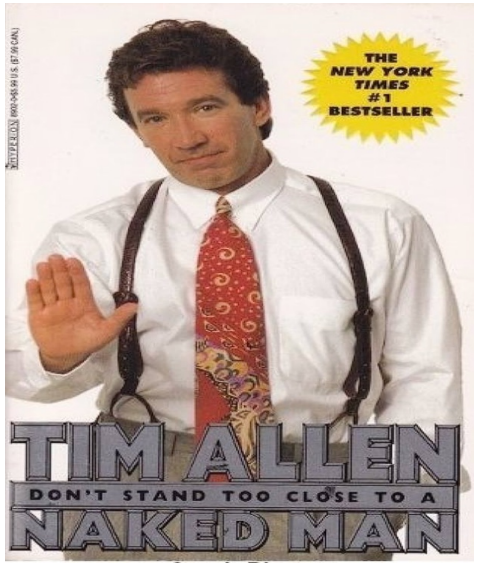
MODELING RESULTS

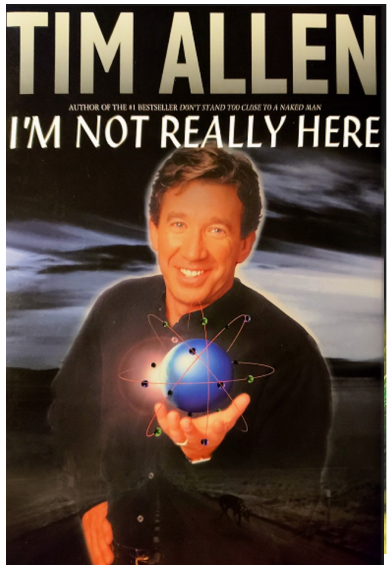
Accuracy main evaluation criteria for models

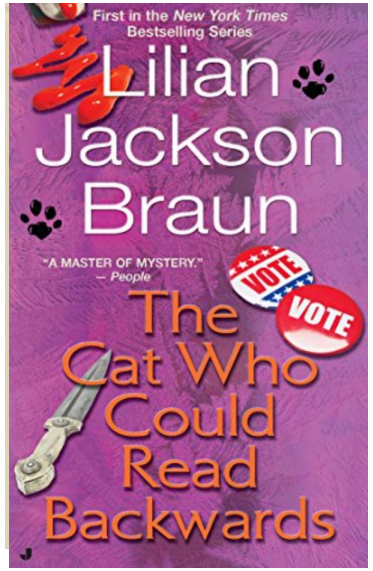
SVD Best Performing Model
Accuracy range: 0.7

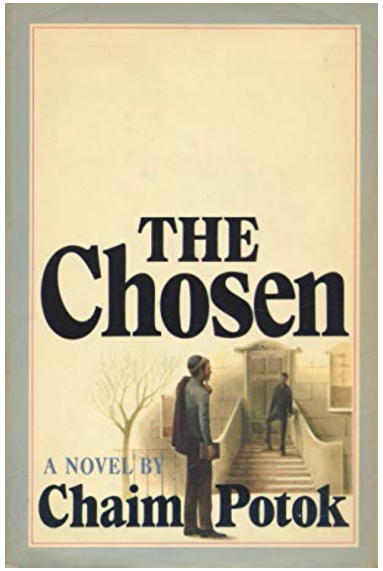
Example: Predicted Rating: 4.0 → Actual Rating: 3.3 – 4.7

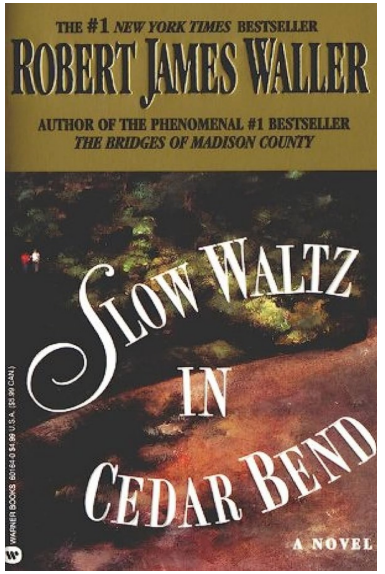
RECOMMENDATION SYSTEM RESULTS











- Original liking for **Humor, Mystery, Fiction, Storytelling**
- **Humor, Mystery, Entertainment, Fiction** genres recommended
- Themes and genres area very similar to original likings
- Popular titles recommended in similar genres

- **Eases into exploration** of other genres and themes
- **4/5 recs** spot on with original likings
- Popular titles not already reviewed might be because of either missing data or book-based movies
- Identified widely recognized books that have the potential to engage and captivate readers

CONCLUSION

- Need more data on users and user-item interactions
- **MVP for initial customer acquisition and data gathering**
- **User feedback and recommendations to improve for production**

NEXT STEPS & IMPROVEMENTS

- **Higher Granularity**
- **Increase amount and diversity of data**
- **Pipelines and User-Feedback loop Updates**
- **User Purchase History**
- **Book-based movies / tv shows**

QUESTIONS?

ADVANCED BOOK RECOMMENDATION SYSTEM

Shayan Abdul Karim Khan

07/21/2023

BACKUP

MODELING RESULTS

RMSE → measures accuracy of predicted ratings

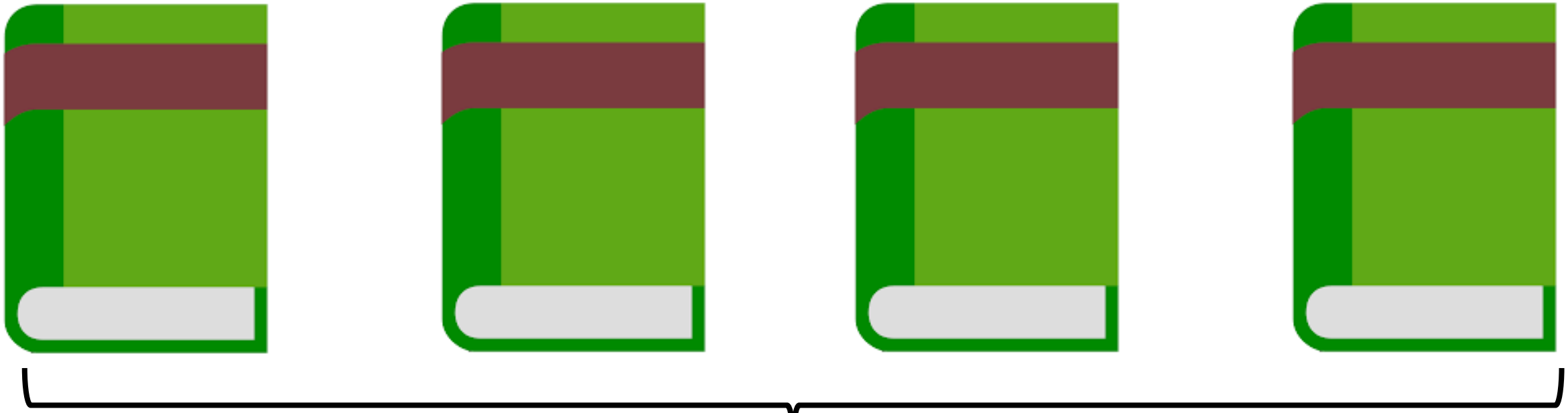
SVD was the best performing algorithm for RMSE

SVD still can be **off by ~0.7 rating points** for predictions

0.7 ratings points can push a below average book into average or above average category

Example: Predicted Rating: 4.0 → Actual Rating: 3.3 – 4.7

OVERVIEW



AVERAGE 4 BOOKS READ PER YEAR PER PERSON

MODELING RESULTS

MODELS	RMSE	MAE	MSE	FCP
SVD	1.37	1.10	1.88	0.54
KNN BASELINE	1.40	1.13	1.96	0.55
KNN WITH MEANS	1.50	1.16	2.26	0.71
KNN WITH Z-SCORE	1.51	1.16	2.27	0.71
NMF	1.58	1.26	2.50	0.57
KNN BASIC	1.62	1.27	2.62	0.59

RMSE → measures accuracy of predicted ratings

MAE → measures average magnitude of prediction errors

MSE → measures average squared difference b/w predicted and actual ratings

FCP → measures the system's ranking accuracy

RMSE → used for selecting the final model

Ratings accuracy important for predicting unknow ratings

SVD was the best performing algorithm for RMSE, MAE and MSE

SVD used in conjunction with **collaborative filtering**

NEXT STEPS & IMPROVEMENTS

- **Higher Granularity** and more features for user and product personas
- Further modelling to bring **accuracy close to 0.25** rating points
- Improve Sampling Methods to **increase amount and diversity of data**
- Incorporate **Pipelines and User-Feedback loop Updates**
- Incorporate information on whether user has **owned a book previously**
- Incorporate information on **book-based movies / tv shows**