BOOK RECOMMENDATION SYSTEM

Shayan Abdul Karim Khan

07/08/2023

Book of The Quarter

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
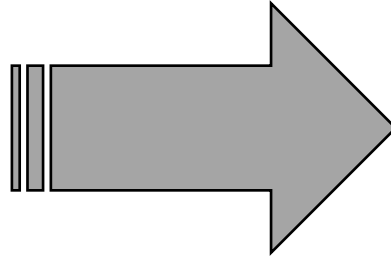QUARTERLY 5 BOOKS AVAILABLE

GOAL

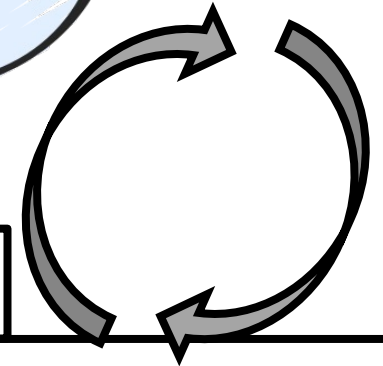**TOP 5**

RECOMMENDATION MODEL
5 BOOK RECOMMENDATIONS

ANALYSIS PROCESS

BOOK-CROSSING DATA CROSSED WITH AMAZON DATA FROM KAGGLE

ITERATIVE DATA CLEANING & PROCESSING

MODELLING

KNN Algorithm

www.educba.com

SVD

NMF

EXPLORATORY ANALYSIS
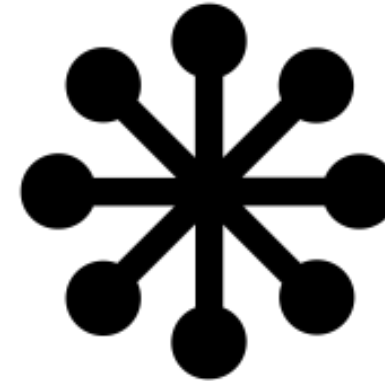
TARGETTED NO OF
REVIEWS > 19

AVG RATING 7.9/10

TARGETTED AGES
16 – 60

USA DOMNIATED
80% OF THE DATA

60% RATINGS
UNKNOWN

60% OF COMPLETE DATA
USED FOR TRAINING

MODELING RESULTS

RMSE → measures accuracy of predicted ratings

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
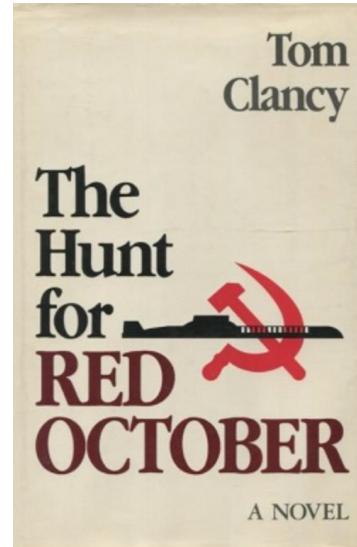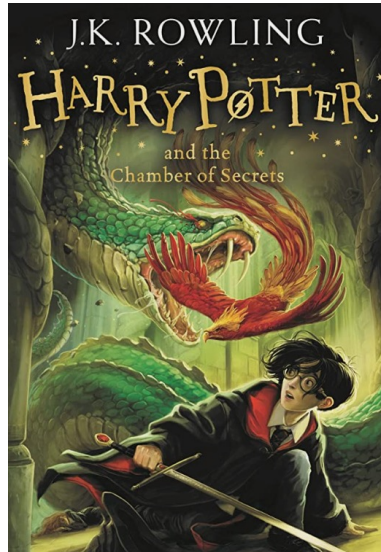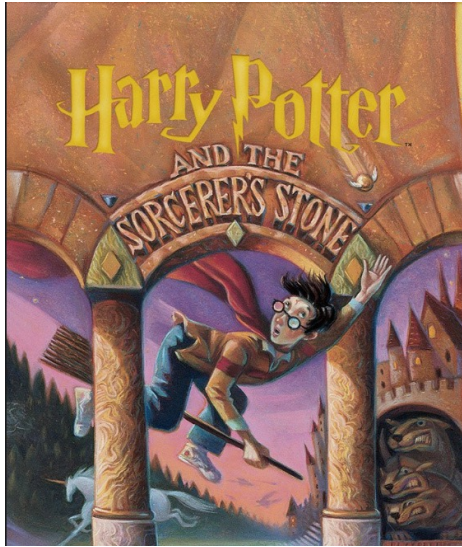
SVD used in conjunction with **collaborative filtering**

SVD still can be **off by ~1.3 rating points** for predictions

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

MODELS	RMSE
SVD	1.37
KNN BASELINE	1.40
KNN WITH MEANS	1.50
KNN WITH Z-SCORE	1.51
NMF	1.58
KNN BASIC	1.62

RECOMMENDATION SYSTEM RESULTS
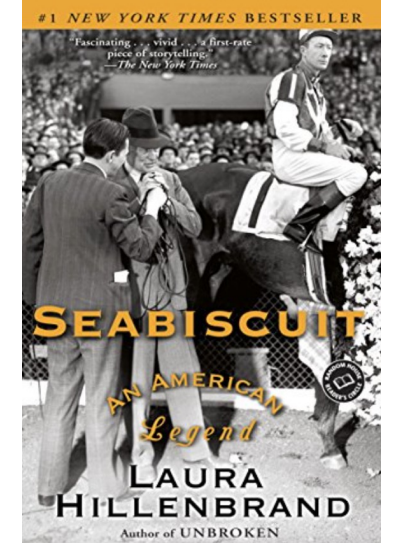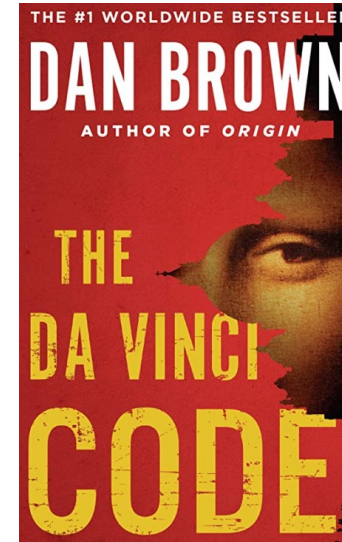
- Original liking for Self-Help, Fantasy, Fiction, Romance

- Fantasy, Thriller, Fiction, and Adventure genres recommended

- Fiction Recommendations line up with liking for Hobbit and Lord of the Rings

- Popular titles recommended in similar genres

- There is a shift in genres and themes

- Popular titles not already reviews might be because of either missing data or book-based movies

- Eases into exploration of other genres and themes

- Identified widely recognized books that have the potential to engage and captivate readers

CONCLUSION

- Ratings are subjective to user preferences

- Divergence of genres and themes can be good or bad

- Not enough information available to determine user personas

- Accuracy range of 1.3 rating points is still high

- No way to predict how may books a user will keep

- No solution for a new user with no data

- Use as baseline model for MVP product to gather more data

- Use user feedback and recommendations to improve for production

NEXT STEPS & IMPROVEMENTS

- **Higher Granularity** and more features for user and product personas

- Further modelling to bring **accuracy close to 0.5** rating points

- Improve Sampling Methods to **increase amount and diversity of data**

- Incorporate **solution for new users**

- Incorporate information on whether user has **owned a book previously**

- Incorporate information on **book-based movies / tv shows**

QUESTIONS?

BOOK

RECOMMENDATION

SYSTEM

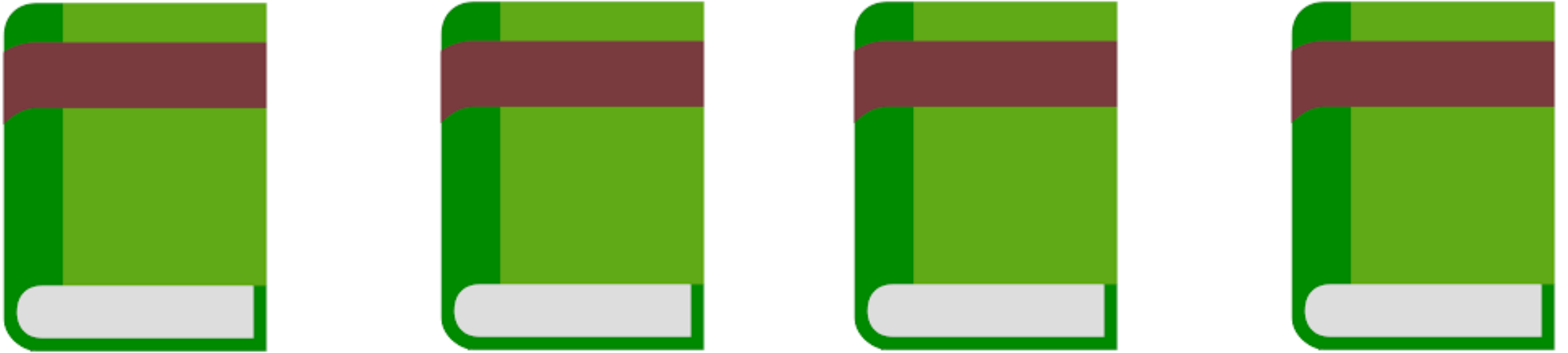Book of The Quarter

Shayan Abdul Karim Khan

07/08/2023

BACKUP

OVERVIEW

AVERAGE 4 BOOKS READ PER YEAR PER PERSON

365 DAYS

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**