



Books Recommender System



By: Gary Chang

Table of contents

01. Introduction

02. EDA

03. Recommender
Models

04. Conclusion

01

Introduction



Recommendation System

Filtering systems that seek to predict and suggest relevant items to users

Impact on market:

- 35% of revenue on Amazon
- 70% of watch time on YouTube
- \$1B savings/year on Netflix

The Amazon logo, featuring the word "amazon" in a dark blue, lowercase sans-serif font, with a curved orange arrow underneath it pointing from the 'a' to the 'z'.The YouTube logo, consisting of a red rounded rectangle with a white play button icon inside, followed by the word "YouTube" in a bold, black, sans-serif font.The Netflix logo, featuring the word "NETFLIX" in a bold, red, sans-serif font.

Problem Statement

How can we implement a recommendation system to predict books that a reader likes?

Using the goodbooks 10k extended dataset, last updated in 2021

- `books_enriched.csv` – 10K entries of book details
- `ratings.csv` – 5.9M ratings provided by 53K unique users



02

Exploratory Data Analysis



Top rated books

- Ratings alone is not sufficient for recommendation

title	authors	average_rating
The Complete Calvin and Hobbes	[Bill Watterson]	4.82
Words of Radiance (The Stormlight Archive, #2)	[Brandon Sanderson]	4.77
Harry Potter Boxed Set, Books 1-5 (Harry Potte...	[J.K. Rowling, Mary GrandPré]	4.77
ESV Study Bible	[Anonymous, Lane T. Dennis, Wayne A. Grudem]	4.76
Mark of the Lion Trilogy	[Francine Rivers]	4.76
It's a Magical World: A Calvin and Hobbes Coll...	[Bill Watterson]	4.75
There's Treasure Everywhere: A Calvin and Hobb...	[Bill Watterson]	4.74
Harry Potter Boxset (Harry Potter, #1-7)	[J.K. Rowling]	4.74
Harry Potter Collection (Harry Potter, #1-6)	[J.K. Rowling]	4.73
The Indispensable Calvin and Hobbes	[Bill Watterson]	4.73

Most popular books

- Books with high rating counts can contain low ratings

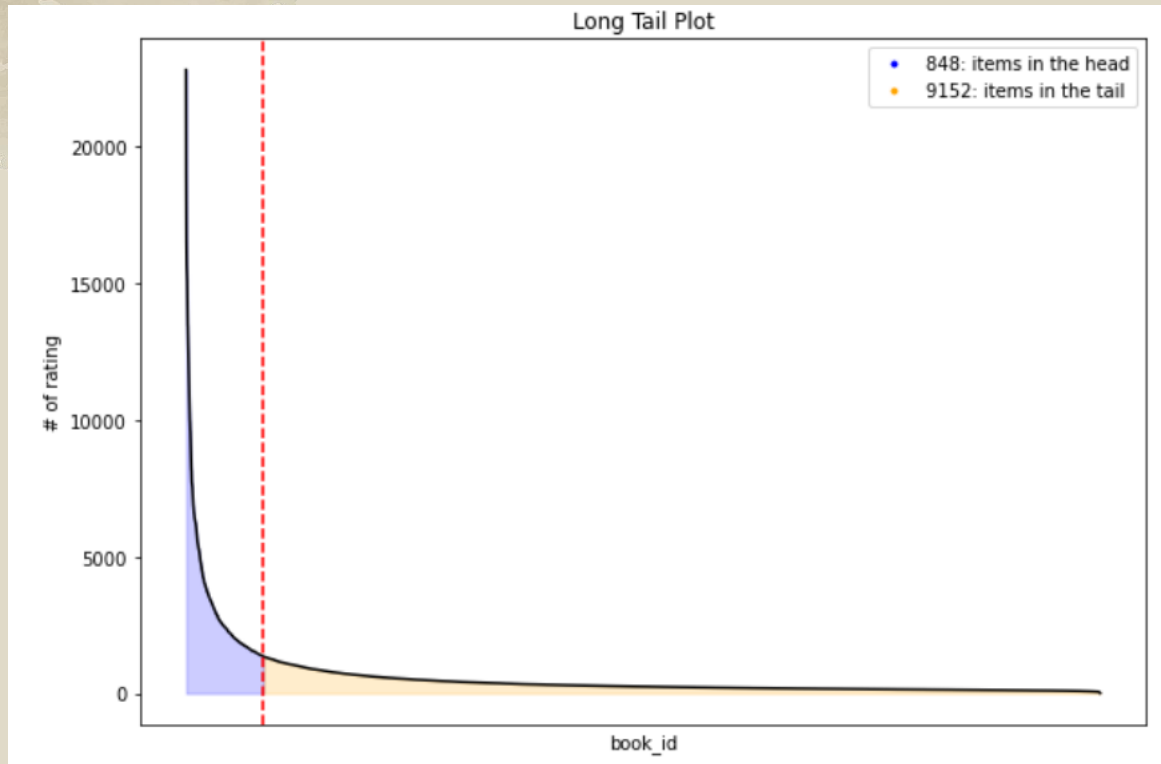
title	authors	average_rating	ratings_count
The Hunger Games (The Hunger Games, #1)	[Suzanne Collins]	4.34	4780653
Harry Potter and the Sorcerer's Stone (Harry P...	[J.K. Rowling, Mary GrandPré]	4.44	4602479
Twilight (Twilight, #1)	[Stephenie Meyer]	3.57	3866839
To Kill a Mockingbird	[Harper Lee]	4.25	3198671
The Great Gatsby	[F. Scott Fitzgerald]	3.89	2683664
The Fault in Our Stars	[John Green]	4.26	2346404
The Hobbit	[J.R.R. Tolkien]	4.25	2071616
The Catcher in the Rye	[J.D. Salinger]	3.79	2044241
Pride and Prejudice	[Jane Austen]	4.24	2035490
Angels & Demons (Robert Langdon, #1)	[Dan Brown]	3.85	2001311

Weighted rating / book score

- How much weightage to set?

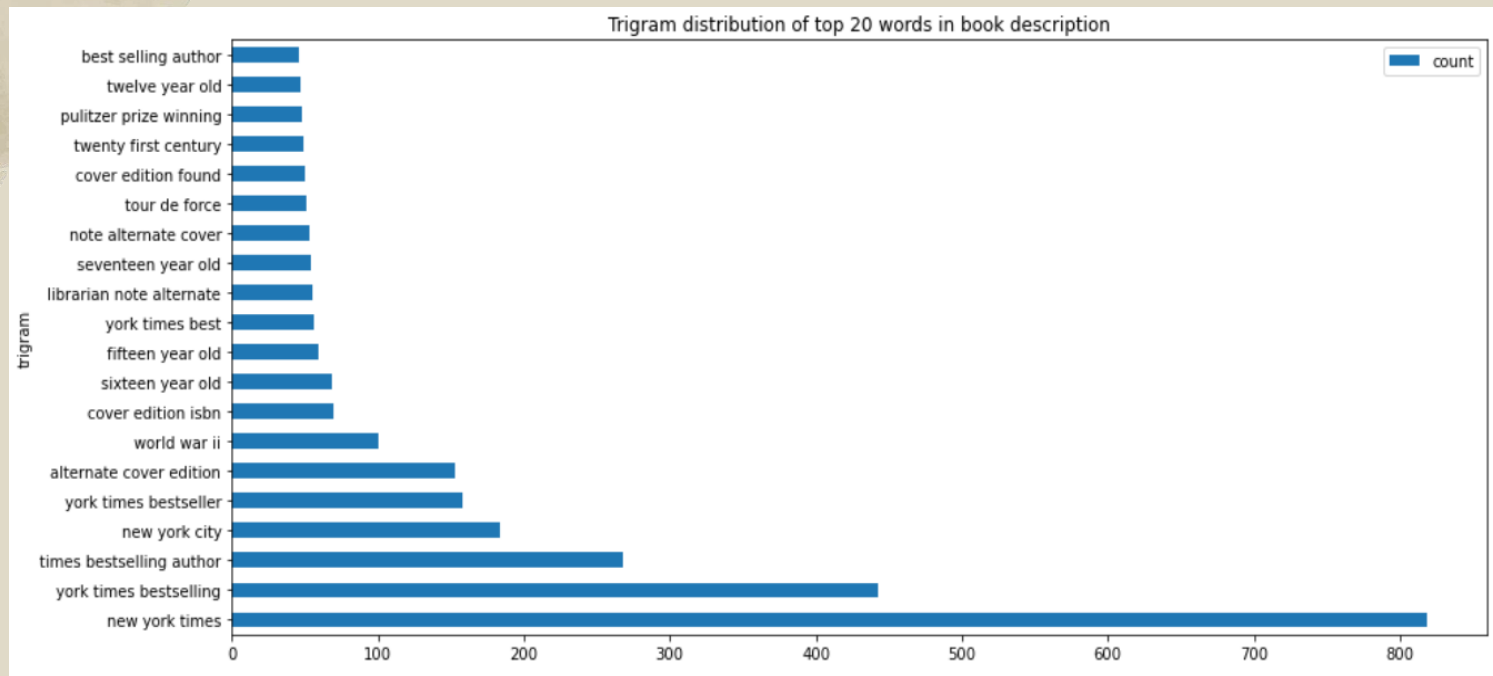
title	authors	average_rating	ratings_count	book_score
Harry Potter and the Sorcerer's Stone (Harry P...	[J.K. Rowling, Mary GrandPré]	0.838298	0.962709	0.900503
The Hunger Games (The Hunger Games, #1)	[Suzanne Collins]	0.795745	1.000000	0.897872
To Kill a Mockingbird	[Harper Lee]	0.757447	0.668895	0.713171
Twilight (Twilight, #1)	[Stephenie Meyer]	0.468085	0.808741	0.638413
Harry Potter and the Deathly Hallows (Harry Po...	[J.K. Rowling, Mary GrandPré]	0.910638	0.364974	0.637806
Harry Potter and the Prisoner of Azkaban (Harr...	[J.K. Rowling, Mary GrandPré, Rufus Beck]	0.876596	0.383026	0.629811
The Fault in Our Stars	[John Green]	0.761702	0.490517	0.626110
Harry Potter and the Goblet of Fire (Harry Pot...	[J.K. Rowling, Mary GrandPré]	0.876596	0.366328	0.621462
Harry Potter and the Half-Blood Prince (Harry ...	[J.K. Rowling, Mary GrandPré]	0.880851	0.350794	0.615822
Harry Potter and the Order of the Phoenix (Har...	[J.K. Rowling, Mary GrandPré]	0.846809	0.362628	0.604718

Long Tail Plot



- Only small % of books have high interactions
- Aim to provide relevant and useful recommendations for users

Book description



03

Recommender Models



Data Pre-processing

books.csv

Use text features for modelling :

- Authors
- Titles
- Genres
- Book Description

ratings.csv

Filtered books with at least 1000 ratings

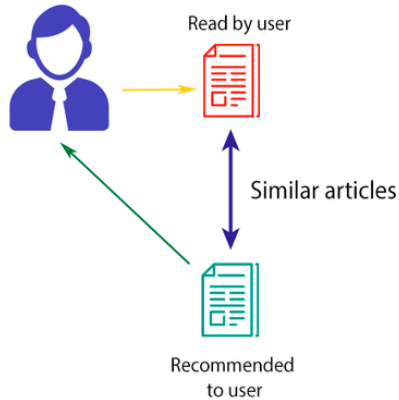
Users with at least 135 ratings given

The original data frame shape: (5976479, 3)

The new data frame shape: (885453, 3)

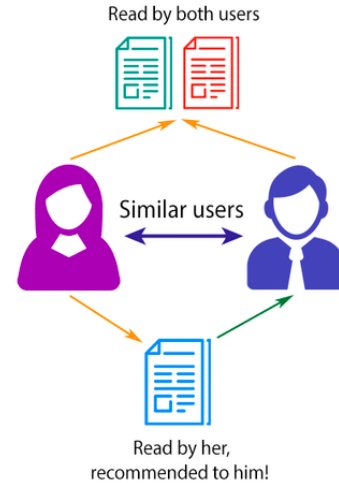
Recommender System

CONTENT-BASED FILTERING



Finding books with similar features

COLLABORATIVE FILTERING



Finding users with similar preferences

Content-Based Filtering

Combine all text
features, **apply TF-
IDF Vectorizer**

1

Use **cosine similarity** to
compute vector distances







2

Calculate **pairwise
similarity scores**







3

Content-Based Filtering

get_recommendations("Angels & Demons (Robert Langdon, #1)")

	title	authors	cover_image	goodreads_link	average_rating	ratings_count
3775	Angels and Demons / The Da Vinci Code (Robert Langdon, #1-2)	['Dan Brown']		Goodreads	4.190000	18711
158	The Lost Symbol (Robert Langdon, #3)	['Dan Brown']		Goodreads	3.660000	369428
21	The Da Vinci Code (Robert Langdon, #2)	['Dan Brown']		Goodreads	3.790000	1447148
192	Inferno (Robert Langdon, #4)	['Dan Brown']		Goodreads	3.800000	287533
9077	The Gemini Contenders	['Robert Ludlum']		Goodreads	3.870000	10428
5180	Temple	['Matthew Reilly']		Goodreads	4.040000	14961

get_recommendations("The Five People You Meet in Heaven")

	title	authors	cover_image	goodreads_link	average_rating	ratings_count
1917	The First Phone Call from Heaven	['Mitch Albom']		Goodreads	3.730000	38957
662	For One More Day	['Mitch Albom']		Goodreads	4.090000	102193
7197	The Boy Who Came Back from Heaven: A Remarkable Account of Miracles, Angels, and Life beyond This World	['Kevin Malarkey', 'Alex Malarkey']		Goodreads	3.890000	8945
242	Heaven is for Real: A Little Boy's Astounding Story of His Trip to Heaven and Back	['Todd Burpo', 'Lynn Vincent']		Goodreads	4.010000	228371
90	Tuesdays with Morrie	['Mitch Albom', 'Saulius Dagsys']		Goodreads	4.060000	556518
5054	Heaven (Casteel, #1)	['V.C. Andrews']		Goodreads	4.010000	17289

Collaborative Filtering

- Using algorithms from *Surprise* library to build model using explicit ratings
- SVD chosen as baseline model for hyperparameters tuning

	test_rmse	fit_time	test_time
Algorithm			
SVDpp	0.843055	566.634514	27.828063
SVD	0.859417	35.636158	4.998294
KNNBaseline	0.864164	64.243537	367.939671
SlopeOne	0.869239	3.018741	24.117684
KNNWithZScore	0.870423	64.629724	390.459211
KNNWithMeans	0.871343	58.662309	339.339279
BaselineOnly	0.876753	2.033532	4.390604
CoClustering	0.883880	21.376527	3.852897
NMF	0.885807	44.309575	4.446064
KNNBasic	0.894614	81.125520	380.896954

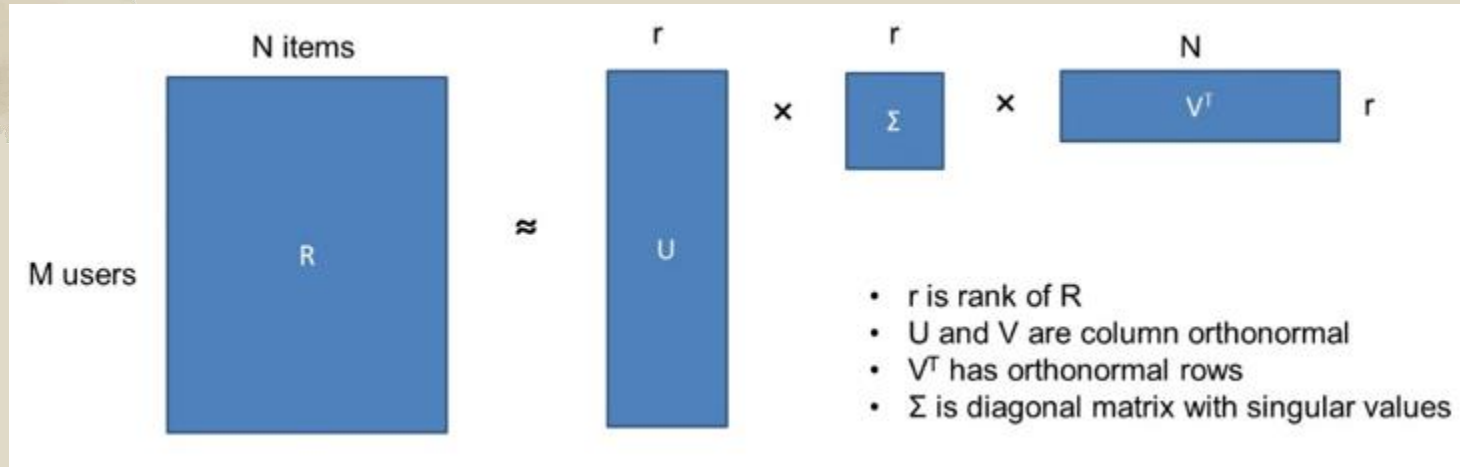
Hyperparameters tuning

Model	test_rmse	Difference	fit-time
Base SVD	0.8594	-	-
Tuned SVD	0.8288	0.0306	16 mins
Tuned SVDpp	0.8281	0.0007	5.5 hrs

```
# Performing GridSearchCV on SVD
param_grid = {"n_factors": [200],      #[70, 80, 100, 120, 150, 180, 200]
              "n_epochs": [40],        #[20, 25, 30, 35, 45, 50]
              "lr_all": [0.015],       #[0.005, 0.010, 0.03, 0.018]
              "reg_all": [0.08],       #[0.02, 0.05, 0.1, 0.075, 0.09]
              "random_state": [23],
              }

gs = GridSearchCV(SVD, param_grid, measures=['RMSE'], cv=3)
```

Singular Value Decomposition (SVD)









Factorize the rating matrix into a product of matrices, provide a condensed representation

Decompose into user feature and item feature matrix / reveal latent factors

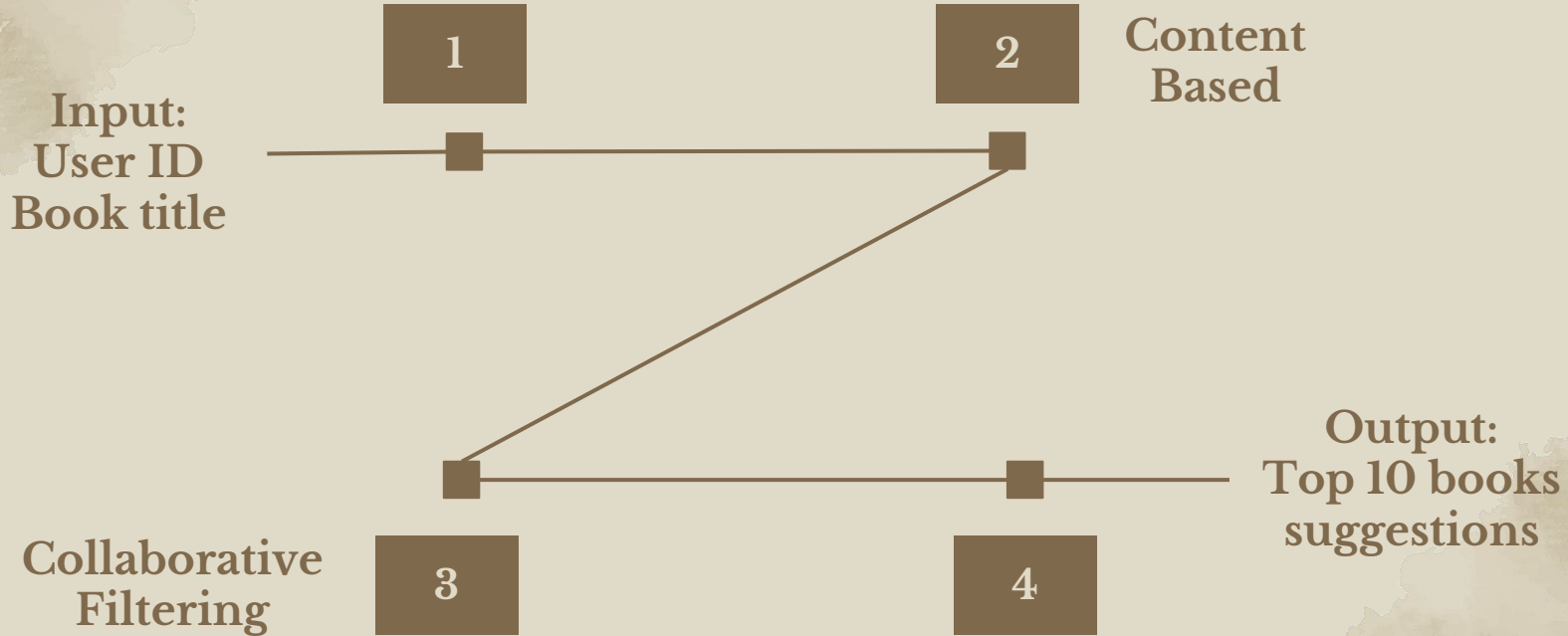
Collaborative Filtering

- Obtain highest predicted ratings for books

```
get_user_pred(6335)
```







	user_id	book_id	title	authors	cover_image	goodreads_link	average_rating	ratings_count
0	6335	1010	The Essential Calvin and Hobbes: A Calvin and Hobbes Treasury	['Bill Watterson']		Goodreads	4.650000	93001
1	6335	780	Calvin and Hobbes	['Bill Watterson', 'G.B. Trudeau']		Goodreads	4.610000	117788
2	6335	422	Harry Potter Boxset (Harry Potter, #1-7)	['J.K. Rowling']		Goodreads	4.740000	190050
3	6335	4	To Kill a Mockingbird	['Harper Lee']		Goodreads	4.250000	3198671
4	6335	490	Maus I: A Survivor's Tale: My Father Bleeds History (Maus, #1)	['Art Spiegelman']		Goodreads	4.350000	184007
5	6335	862	Words of Radiance (The Stormlight Archive, #2)	['Brandon Sanderson']		Goodreads	4.770000	73572

Hybrid Recommender









Hybrid Recommender

hybrid(202, 'The Five People You Meet in Heaven')

	title	authors	cover_image	goodreads_link	average_rating	ratings_count	est
869	Lord of Chaos (Wheel of Time, #6)	['Robert Jordan']		Goodreads	4.100000	91046	4.016919
787	The Alchemyst (The Secrets of the Immortal Nicholas Flamel, #1)	['Michael Scott']		Goodreads	3.840000	58396	4.003085
1917	The First Phone Call from Heaven	['Mitch Albom']		Goodreads	3.730000	38957	3.995743
7734	The First Fifteen Lives of Harry August	['Claire North']		Goodreads	4.040000	22327	3.995743
4530	The Rose Garden	['Susanna Kearsley']		Goodreads	4.010000	18553	3.995743
4364	Abandon (Abandon, #1)	['Meg Cabot']		Goodreads	3.710000	30934	3.995743

hybrid(6335, 'The Five People You Meet in Heaven')

	title	authors	cover_image	goodreads_link	average_rating	ratings_count	est
17	The Lovely Bones	['Alice Sebold']		Goodreads	3.770000	1605173	4.056850
869	Lord of Chaos (Wheel of Time, #6)	['Robert Jordan']		Goodreads	4.100000	91046	4.017087
208	How to Win Friends and Influence People	['Dale Carnegie']		Goodreads	4.130000	282623	3.987156
1917	The First Phone Call from Heaven	['Mitch Albom']		Goodreads	3.730000	38957	3.979531
7735	The Guns of August	['Claire North']		Goodreads	4.180000	35147	3.979531
4364	Abandon (Abandon, #1)	['Meg Cabot']		Goodreads	3.710000	30934	3.979531

04

Conclusion



Summary

1. Content-based Recommender:

Using book titles, genres, authors and description as features to find similar books

2. Collaborative Filtering Recommender

Used SVD to obtain estimated ratings for a given user

3. Hybrid Recommender

Combined both content and collaborative filtering models, create book suggestions after inputting user ID and book title



Future Works

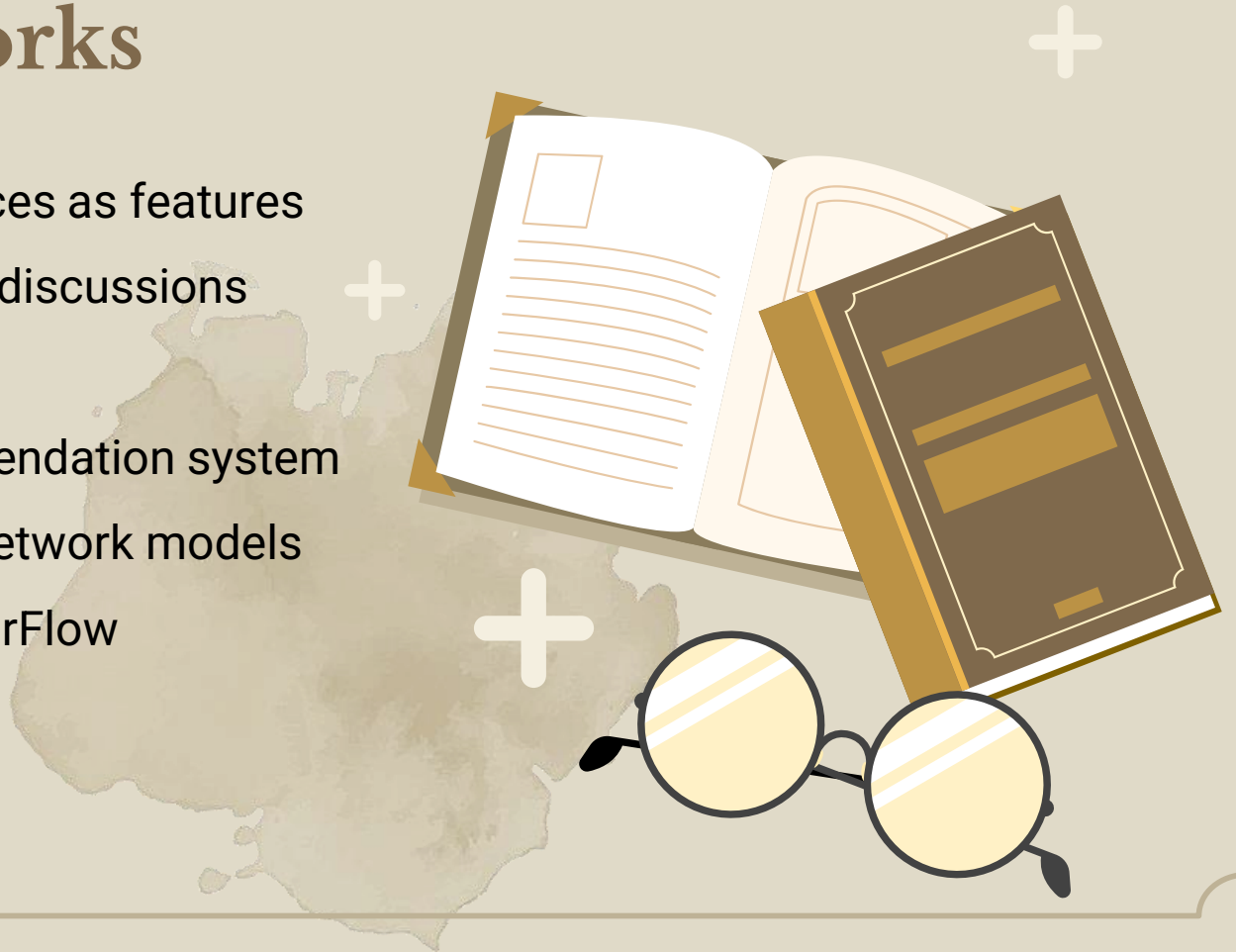
1. Include other sources as features

- e.g. text reviews, discussions

2. Use other recommendation system libraries and neural network models

- e.g. LightFM, TensorFlow

Recommenders





Thank You

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

Please keep this slide for attribution

