

A woman with dark hair and glasses is looking over the top of an open book. She is in a library, with bookshelves filled with books visible in the background. The image has a blue tint and a semi-transparent orange rectangle over the book and the woman's face.

Your Next Page Turner

Quan Nguyen &
Rebecca Rosen

About Our Project

Have you finished this year's hottest book
and are looking to read more like it?

“Your Next Page Turner” is here for you !



Tristique curae mollis porta sed fusce tempor
lacus est mus magna iaculis, ue ultricies
aliquet aptent diam accumsan varius auctor non.
Dapibus purus varius potenti sociosque pulvinar
quis ut sagittis, pur turpis fermentum litora
magnis class tincidunt, faucibus auctor
tristique inceptos erat vehicula ue. Platea
metus auctor blandit id cubilla dapibus
parturient dictum risus vulputat, rhoncus
tristique bibendum praesent laoreet nulla
ornare pharetra sociosque penatibus, gravida mi
aliquam malesuada condimentum ue imperdiet at
nue.

Tristricque curas melleis porta sed fusce tempor
 lacus estus mas magna facillis, ere utriusque
 aliquid aptum dian accusans varus pectus
 mapius parvus varus pectus sociosque pulvinar
 quis ut sagittis, turpis turpis fermentum litora
 equis ut sagittis, turpis turpis fermentum litora
 quibus ut sagittis, turpis turpis fermentum litora
 tristricque incinctus ere vahiliu capibus
 metus auctor blandis in turpis aliquid aptum dian
 perturient dicitur turpis auctor non pectus
 accusans varus pectus sociosque pulvinar
 varus potenti sociosque pulvinar turpis
 sagittis, per turpis fermentum litora magnis
 turpis turpis fermentum litora magnis
 classa tinctum, turpis turpis fermentum litora
 incinctus ere vahiliu capibus perturient dicitur
 equis ut sagittis, turpis turpis fermentum litora
 risus volupstat, phrethos phrethos phrethos
 praesent laoreet laoreet non pectus
 bibendos praesent laoreet non pectus
 sociosque penatibus, gravibus utriusque
 holsedda contentum ere isperdierit at nas.

Lorem ipsum dolor sit amet consectetur
 adipiscing elit massa sociis mollis iaculis,
 arcu fermentum per ullamcorper id at tempor
 turpis volutpat curae, quisque vitae montes
 magna sodales purus fringilla mattis sem nunc.
 Elementum tempor scelerisque metus hac lacus,
 sem rhoncus consequat mollis turpis porta,
 vivamus bibendum porttitor suscipit.

Tristique curae mollis porta sed fusce tempor
lacus est mas magna tacculis, eu ultricies
lignis aptent diam accumen varius auctor non,
Dapibus purus varius potenti sociosquo pulvinar
quis ut agittis, per turpis fermentum litora
magnis class tincidunt, faucibus auctor
tristique inceptos erat vehicula eu. Platea
metus auctor blandit id cubilia dapibus
parturient dictum risus volutpat, rhoncus
tristique bibendum praesent laoreet nulla
ornare pharetra sociosquo penatibus, gravida mi
aliquam malesuada condimentum eu imperdiet at
nam

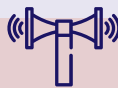
Today's Presentation



Data Process and EDA



Model Iterations



Final Product



Insights & Next Steps

About our Dataset



Goodreads Scrapping

We scraped *titles, genres, average ratings, number of ratings, authors, lengths and descriptions* from the top 200 books of the past 2 decades from Goodreads.com



Data Cleaning

Armed with 4,183 books from the past 20 years, we used **Pandas** in **Python** to store and clean our data



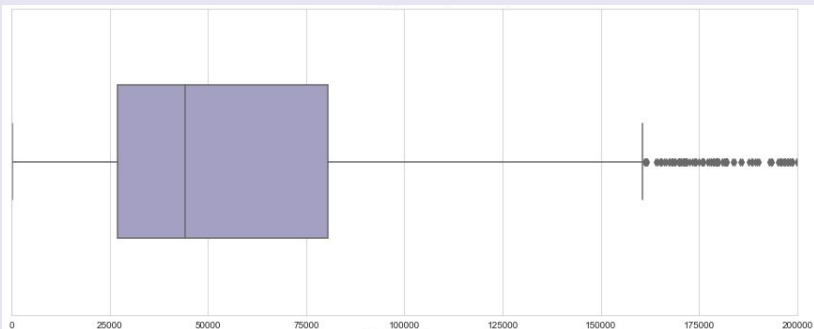
Feature Engineering

Using **TFIDF** we vectorized descriptions, and used **Rake()** to generate keywords.

Analyzing trends, we determined cutoffs for book length and popularity

Exploratory Data Analysis:

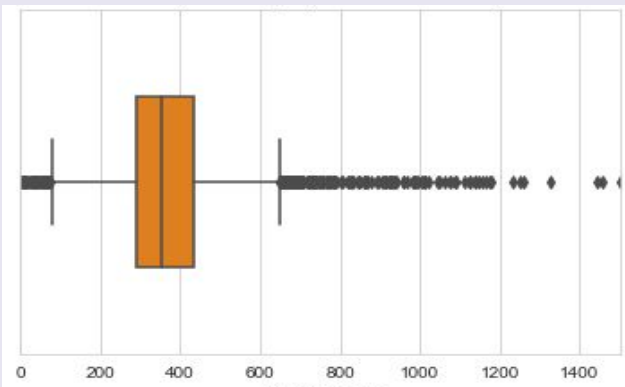
Number of Ratings per Book (300 outliers excluded)



How we determined book length and popularity categories

...and grouped together genres

Number of Pages per Book (10 outliers excluded)



Categories of Genres:

Sci-Fi:
1674 Books

Romance:
1650 Books

Thriller:
994 Books

Comics:
130 Books

Biography:
340 Books

Nonfiction:
850 Books

Self-Help:
211 Books

**Young
Adult:**
1209 Books

Family:
192 Books

Fiction:
2897 Books

Example List of Books

Sci-Fi

✗ Harry Potter

✗ Ghosts

✗ A Feast for Crows

Family

○ Ghosts

○ Flush

○ Echo

Romance

○ 50 Shades of Grey

○ Perks of Being a Wallflower

○ Flipped



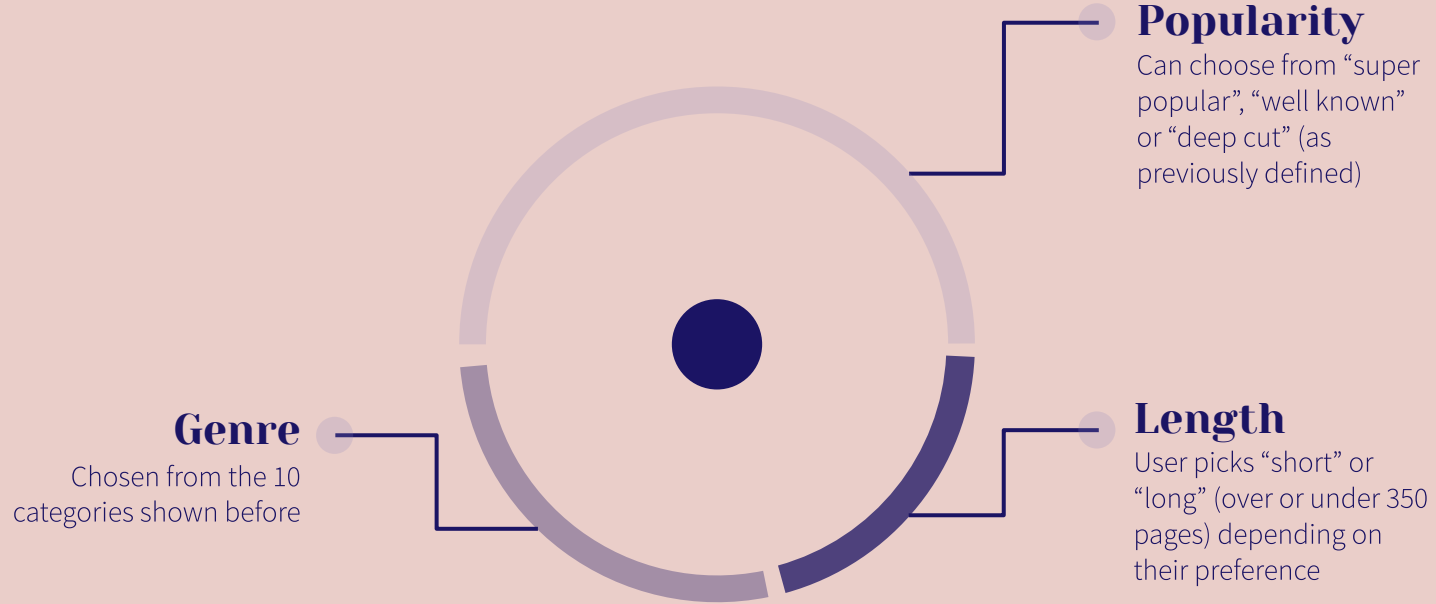


01

Simple Recs

Only using genre, popularity
and length

Baseline Recommendation System



simple_rec(genre='Scifi', length='short', popularity='well known', dict=genre_id_dict, df=dff)

	authors	description	genre	id	num_ratings	pages	ratings
titles							
Jewels of the Sun (Gallaghers of Ardmore, #1)	Nora Roberts	In her new trilogy, New York Times #1 Best Sel...	[Fantasy, Romance, Contemporary, Fiction, Cont...	59810	57717	347	4.220
Isaac's Storm: A Man, a Time, and the Deadliest Hurricane in History	Erik Larson	National Bestseller September 8, 1900, began ...	[History, Biography, North American Hi..., Sci...	239186	42147	336	4.060
First Test (Protector of the Small, #1)	Tamora Pierce	In the medieval and fantastic realm of Tortall...	[Fiction, Fantasy, Young Adult]	153784	49539	240	4.240
Marina	Kathy Reichs	Quince años mas tarde, la memoria de aquel dia...	[Fantasy, Gothic, Horror, Fiction, Young Adult...	4516	35176	238	4.020

simple_rec(genre='Scifi', length='long', popularity='super popular', dict=genre_id_dict, df=dff)

	authors	description	genre	id	num_ratings	pages	ratings
titles							
Harry Potter and the Prisoner of Azkaban (Harry Potter, #3)	J.K. Rowling	Harry Potter's third year at Hogwarts is full ...	[Middle Grade, Adventure, Fantasy, Audiobook, ...	5	2435849	435	4.560
Ender's Shadow (The Shadow Series, #1)	Orson Scott Card	Welcome to Battleschool. Growing up is never ...	[Young Adult, Science Fiction, Fiction, Fantasy]	9532	143282	469	4.310
Gardens of the Moon (Malazan Book of the Fallen, #1)	Steven Erikson	The Malazan Empire simmers with discontent, bl...	[High Fantasy, Epic Fantasy, Fantasy, Fiction]	55399	81082	666	3.880
The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory	Brian Greene	The international bestseller that inspired a m...	[Physics, Astronomy, Popular Science, Science,...	8049273	81201	464	4.070
Harry Potter and the Goblet of Fire (Harry Potter, #4)	J.K. Rowling	Harry Potter is midway through his training as...	[Fiction, Fantasy, Young Adult]	6	2279215	734	4.550



02

Description- Based Recs

Only using BoW descriptions

2. Description-Based Recs

Process:

1. Create Bag of Words for each book using the *description* of the book
2. TF-IDF Vectorize the Bag of Words
3. Compute Cosine Similarity between descriptions
4. Use Cosine Similarity to generate recommendations



Top 10 Recommendations for **Gone Girl**

1: Into the Water

2: The Silent Wife

3: The Hunger Games (The Hunger Games, #1)

4: The Lying Game (The Lying Game, #1)

5: Searching for Sylvie Lee

6: Girls in Pants: The Third Summer of the Sisterhood (Sisterhood, #3)

7: Long Way Down

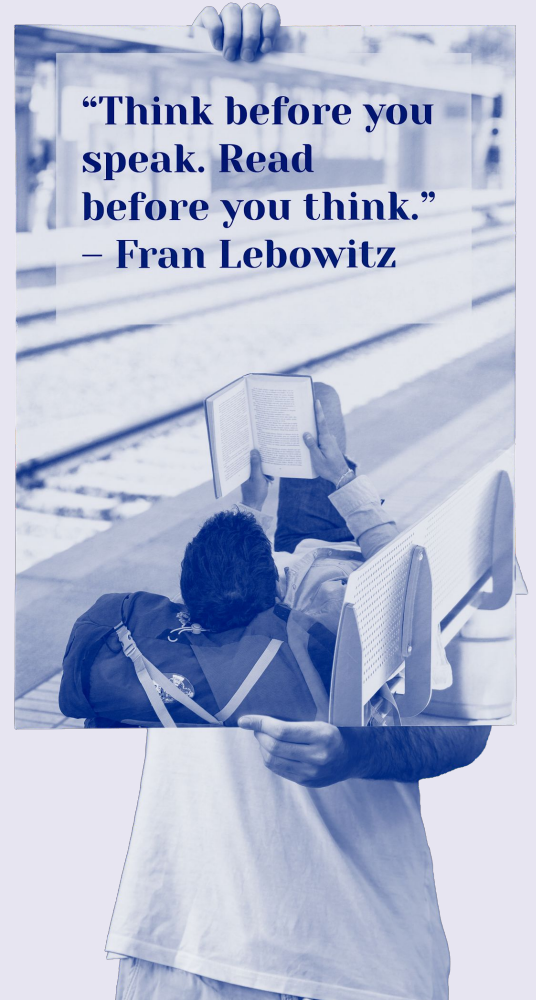
8: You (You, #1)

9: The Drop (Harry Bosch, #15)

10: We Were Liars

What's missing?

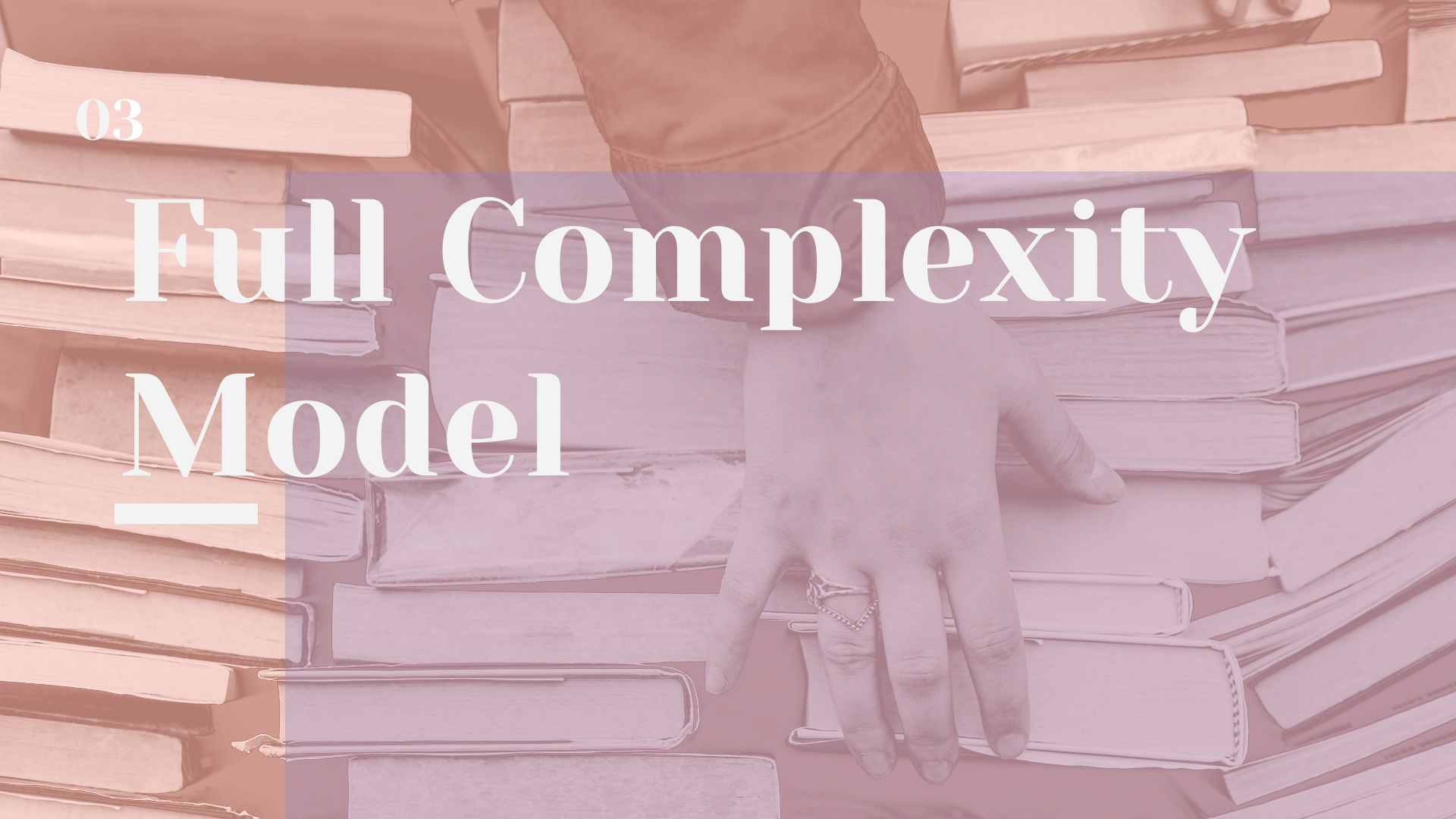
- Author
- Genre
- Personalization



**“Think before you
speak. Read
before you think.”
– Fran Lebowitz**

03

Full Complexity Model



3. Full Complexity Model

Process:

1. Create Bag of Words for each book using *description*, *author's name*, and complete list of *genres* (decided by Goodreads users)
2. Count Vectorizer the Bag of Words
3. Compute Cosine Similarity between BoWs
4. Use Cosine Similarity to generate recommendations

Top 10 Recommendations for Gone Girl

- 1: The Silent Wife
- 2: Into the Water
- 3: The Mephisto Club (Rizzoli & Isles, #6)
- 4: The Drop (Harry Bosch, #15)
- 5: The Wife Between Us
- 6: The Snowman (Harry Hole, #7)
- 7: Dark Places
- 8: The Sinner
- 9: What Have You Done
- 10: 61 Hours (Jack Reacher, #14)

Filters

01 Popularity

- **“Deep Cut”:**
`num_ratings < 27000`
- **“Well Known” :**
`27000 < num_ratings < 80000`
- **“Super Popular” :**
`num_ratings > 80000`

02 Length of Book

- **Short :** `pages < 350`
- **Long :** `pages > 350`

Final Model

Input

- Title: Gone Girl
- Length: Short
- Popularity: Super Popular

Output

titles	num_ratings	pages
The Silent Wife	91195	326.000
The Wife Between Us	158991	346.000
Dark Places	510708	349.000
We Were Liars	401942	242.000
The Girl on the Train	1802815	323.000
In a Dark, Dark Wood	177192	308.000
The Girl Before	100896	341.000

Reflections



Insights

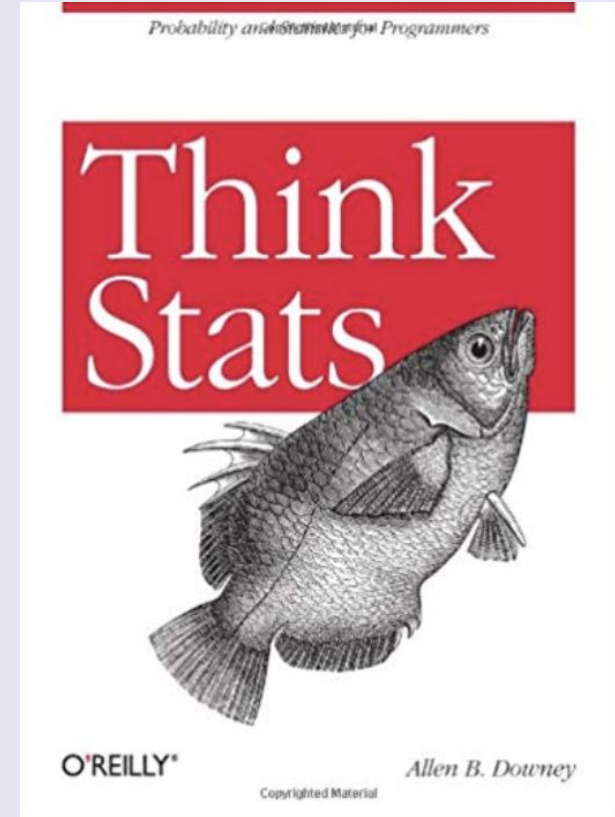
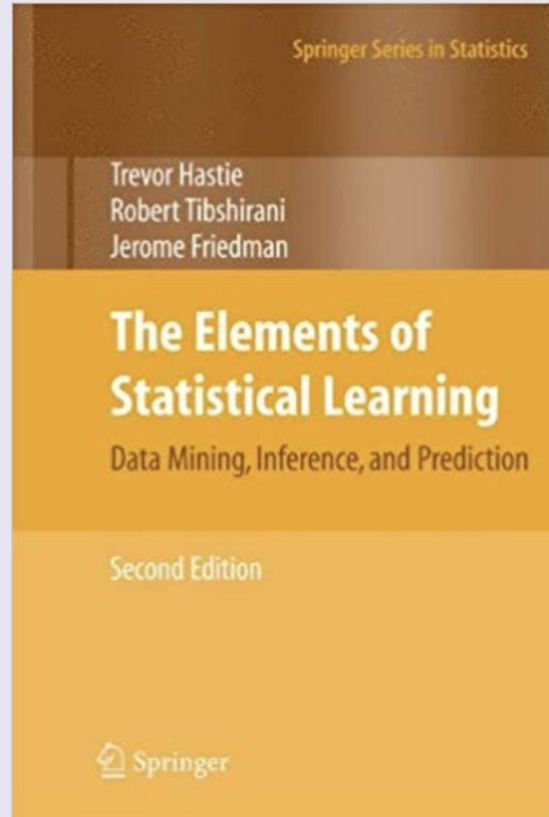
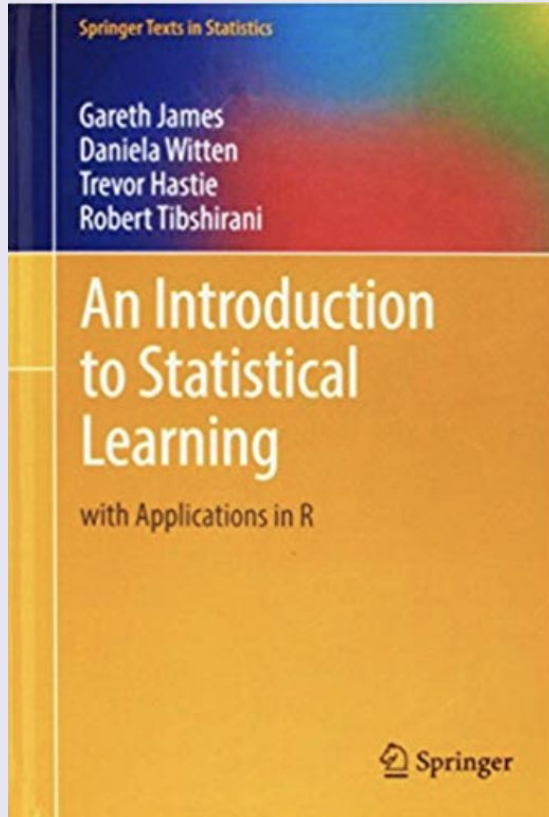
- No exact way to evaluate the accuracy of the models. However, user feedback can be helpful in improving accuracy
- Base on the Gone Girl example, the final model does much better on recommending books compared to the first 2 models



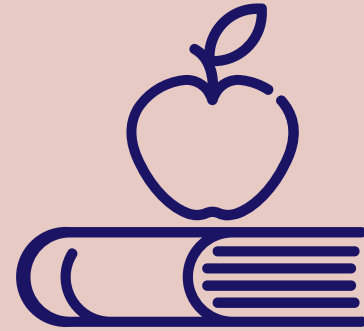
Next Steps

- Get more data/books
- Use different metrics to calculate similarity
- Using features like ratings, published year, and awards to improve performance

Predictive recommendations to follow...



Any Questions?



Thank you !

