

# Book Recommendation System



Jennifer Zhao

# Overview

---

This book recommendation system provides personalized recommendations based on user interests and preferences.

Additionally, exploratory data analysis will be utilized to draw insights about users' reading preferences and current trends in the book market.

# Process

---

Exploratory Data Analysis (EDA)

Data Analysis and Visualization

Modeling using KNN (K-Mean and Truncated  
SVD Latent Factors model)

Web Deployment

# Dataset

---

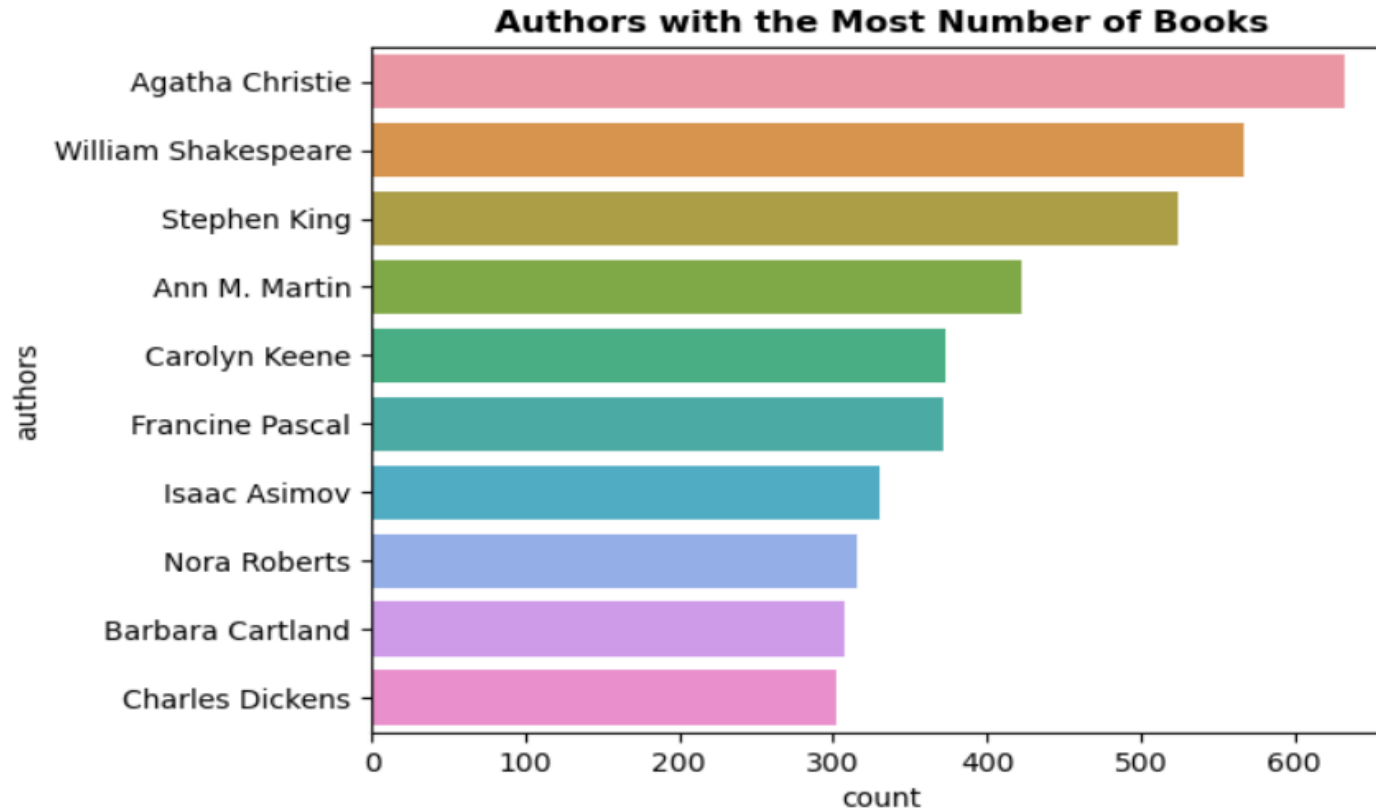
The dataset contains book information along with user ratings and is available from Kaggle.

- Books.csv: book information( book title, authors, publish year). It consists of 27K rows.
- Ratings.csv: book rating description( user\_id, book\_id, rating). It consists of 27K rows.
- Users.csv: user descriptions (user\_id, location, age). It consists of 160K rows.

*Data Sampling: Due to the large data size, a subset named "books\_with\_rating" was created by sampling 13,000 records from the original dataset.*

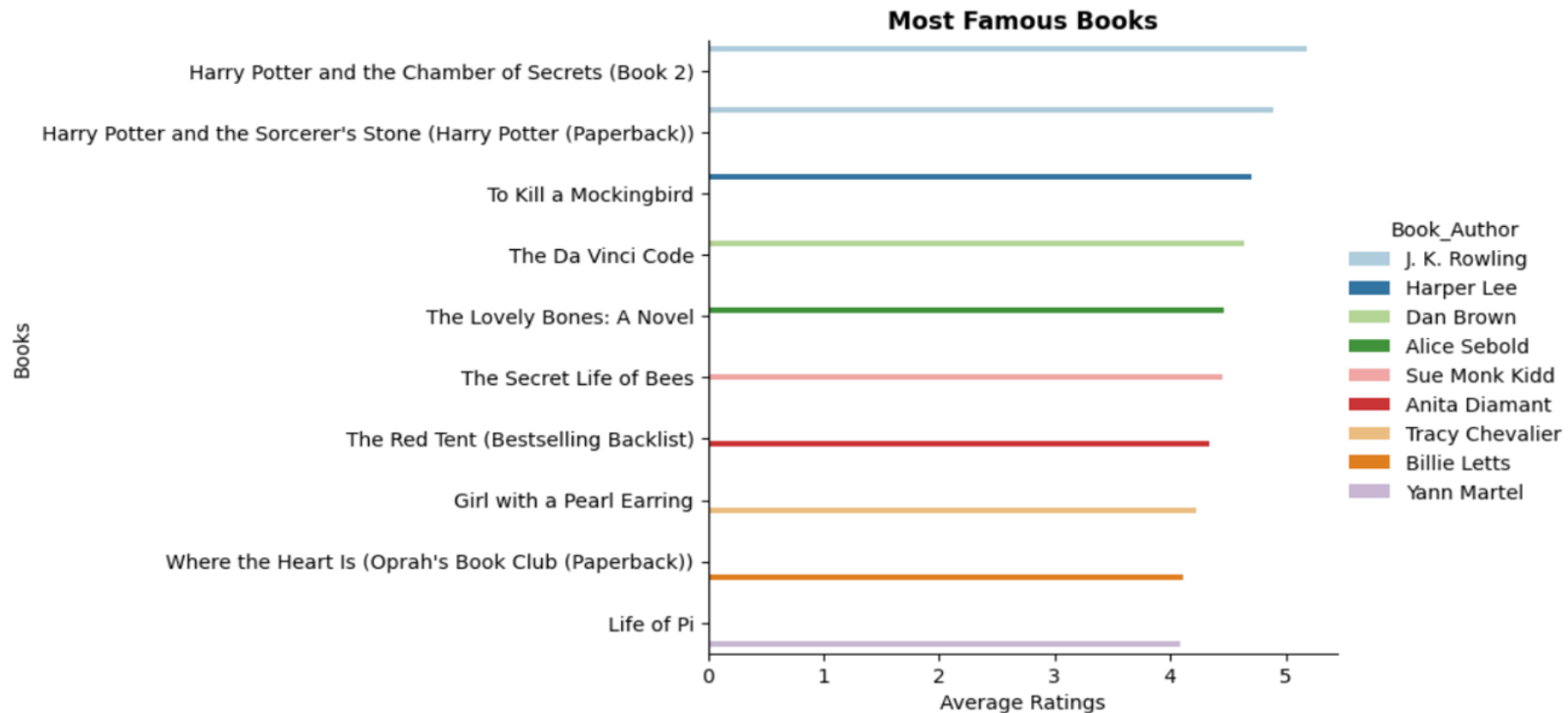
# Data Visualization

---



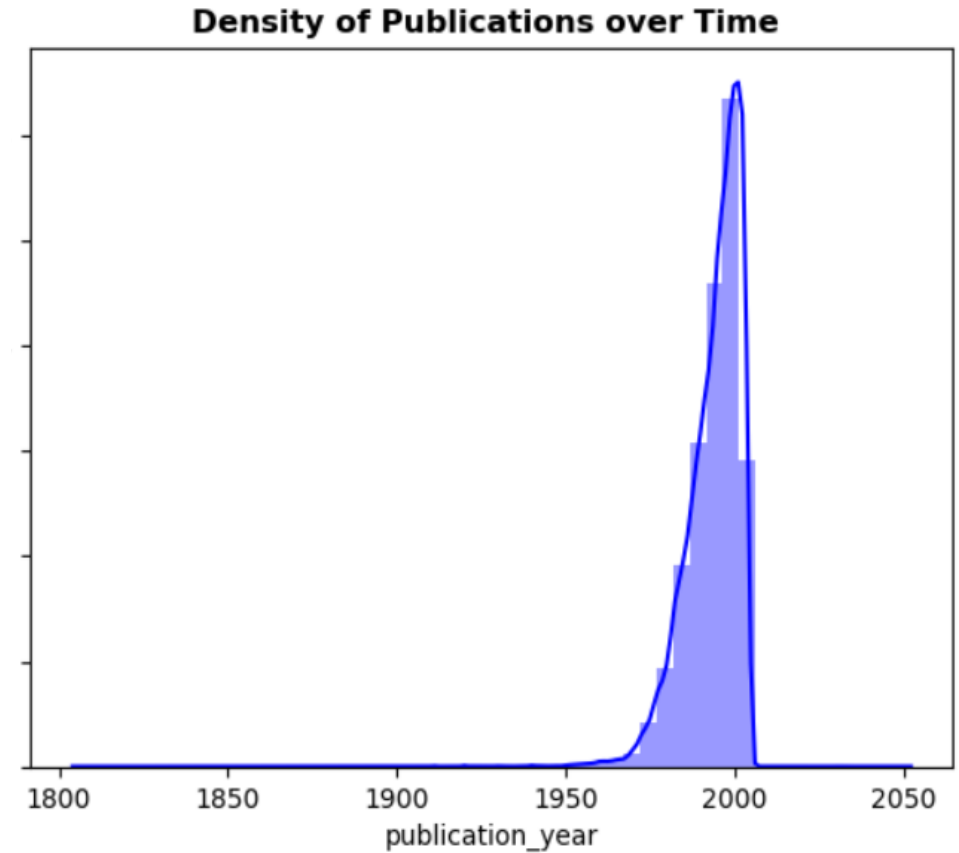
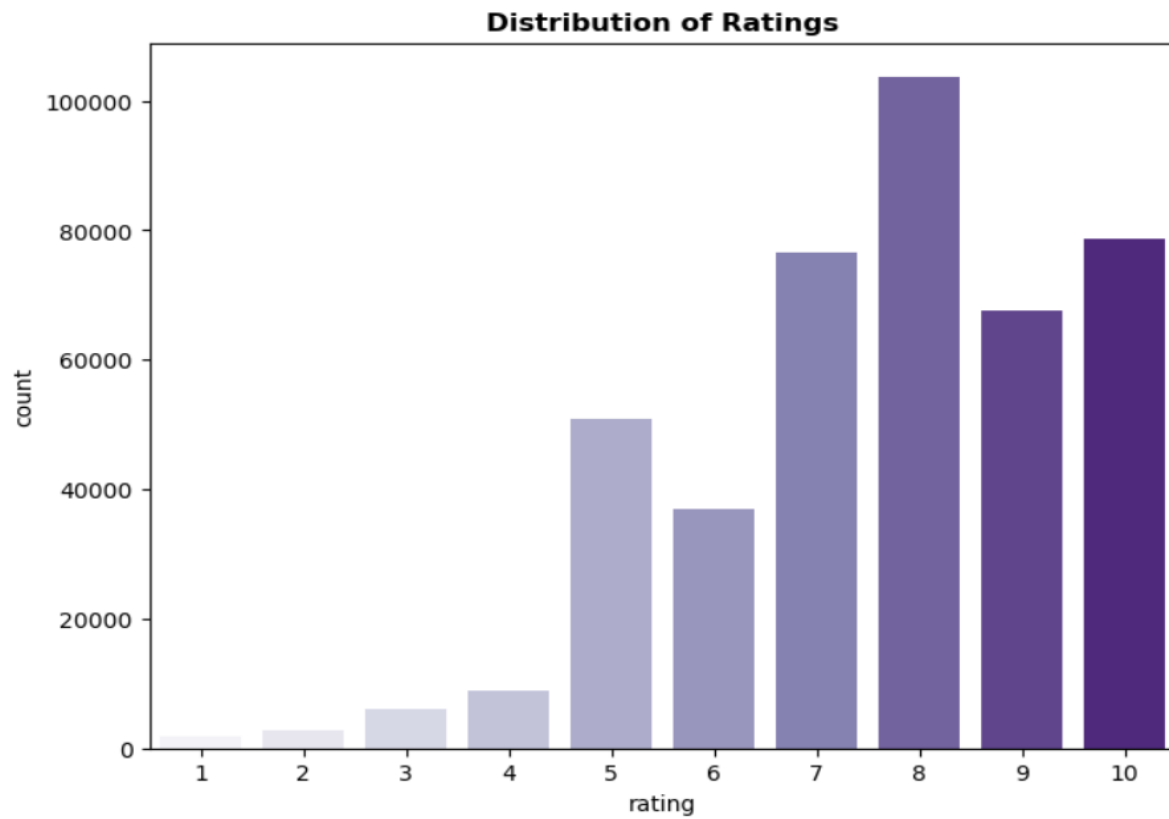
# Data Visualization

---



# Data Visualization

---



# Methodology

---



**Popularity-Based:** List the top 20 popular books for everyone.



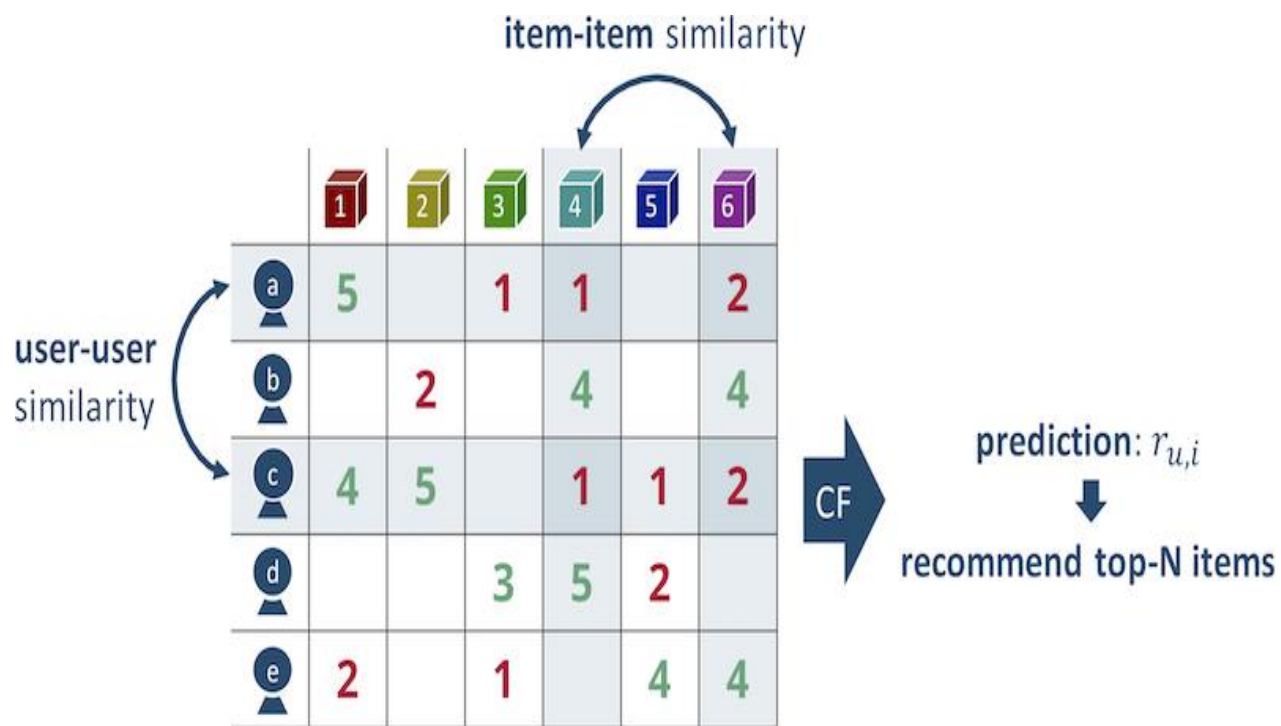
**Content-Based:** Calculate Cosine Similarity for book attributes (title, authors, etc.) to find similar books.



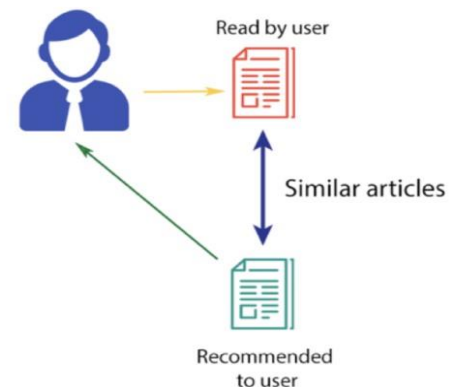
**Collaborative Recommendation using KNN and Truncated SVD Latent Factors model:** Utilizes user ratings and interactions to identify users with similar preferences.



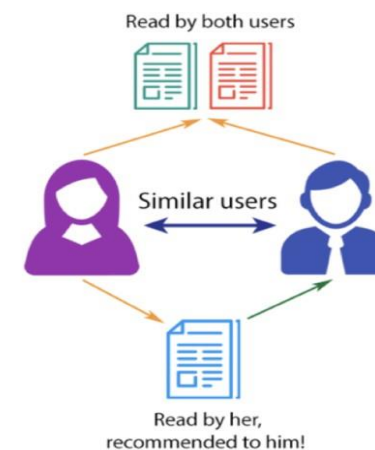
# Content-based vs. Collaborative Filtering



CONTENT-BASED FILTERING



COLLABORATIVE FILTERING



# TECH STACK

---

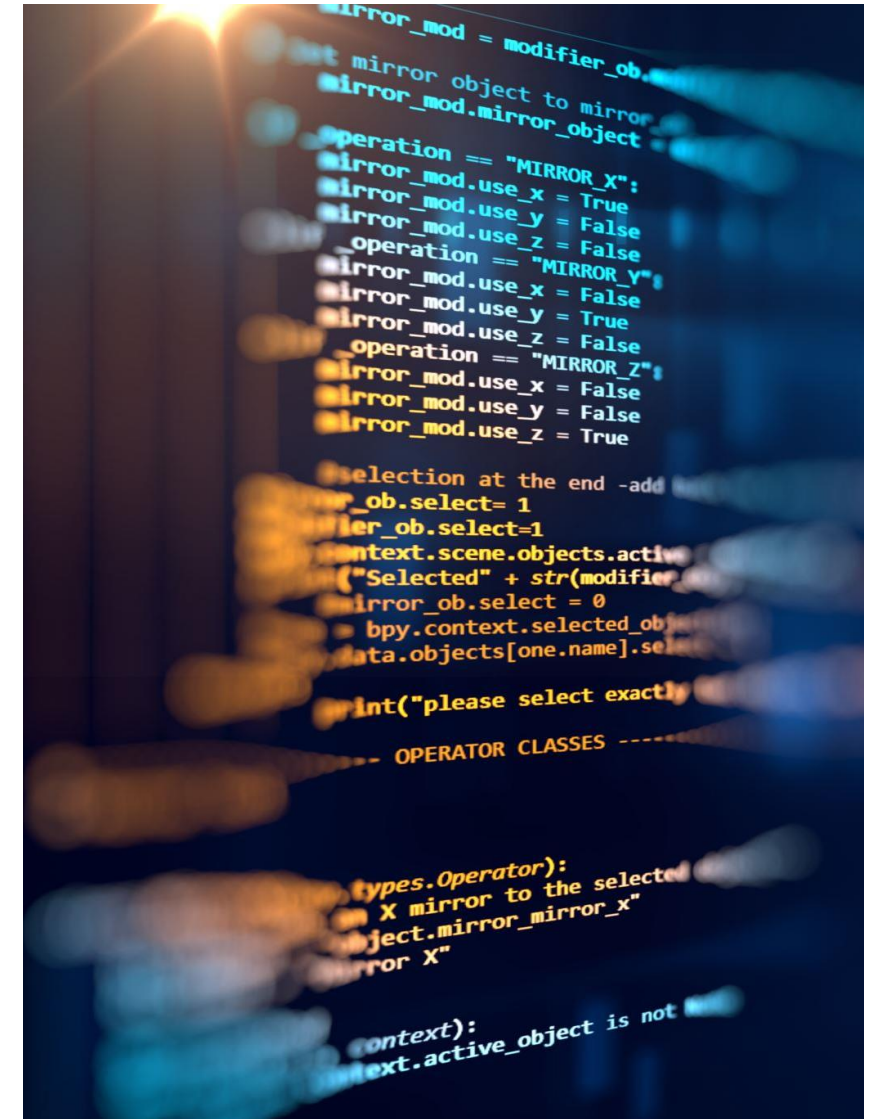
**Language:** Python

**Libraries:** Pandas, NumPy, Seaborn

**Sklearn Modules:** TfidfVectorizer, cosine\_similarity, TruncatedSVD, NearestNeighbors, Sparse\_matrix

**Web Tools:** Flask, HTML, Streamlit

**Models:** KNN (K-Mean) and Truncated SVD Latent Factors model



# Popularity based Recommendation

---

## THE TOP 5 POPULAR BOOKS

title	authors	publication_year	Cover_image
Harry Potter and the Chamber of Secrets Postcard Book	J. K. Rowling	2002	
Calvin and Hobbes	Bill Watterson	1987	
My Sister's Keeper : A Novel (Picoult, Jodi)	Jodi Picoult	2004	
Where the Sidewalk Ends : Poems and Drawings	Shel Silverstein	1974	
Johnny Got His Gun	Dalton Trumbo	1983	

# Content-Based Recommendation

---

```
search("Little Altars Everywhere", vectorizer)
```

title	authors	publication_year	Cover_image
Beautiful Necessity: The Art and Meaning of Women's Altars	Kay Turner	1999	
Altars in the Street: A Neighborhood Fights to Survive	Melody Ermachild Chavis	1997	
Little	David Treuer	1995	
Everywhere That Mary Went	Lisa Scottoline	1993	
In Search of Color Everywhere	E. Ethelbert Miller	1996	

# Collaborative Recommendation with Truncated SVD Latent Factors

---

`similar_user_ids`

`[ 500 2725 672 1240 2282]`

`Recommended books:`

`Raging Heart: The Intimate Story of the Tragic Marriage of O.J. and Nicole Brown Simpson`

`The All-american Cowboy Cookbook : Over 300 Recipes From the World's Greatest Cowboys`

`Grace Point`

`The Diaries of Adam and Eve`

`Fallout`

`The WAR AGAINST BOYS: How Misguided Feminism Is Harming Our Young Men`

---

# Future Work

---



Dataset: Seek a comprehensive dataset with more features (genres, descriptions) to improve the model's performance.



Building a more robust model by leveraging the PyTorch and TensorFlow deep learning libraries.

*Thank you!*

---