DATA ANALYTICS

BOOK RECOMMENDATION SYSTEM

PRESENTED BY:
ANSHIKA PAL - PES1UG20CS062
ROHAN - PES1UG20CS346
CHANNABASAVANNA HATTI - PES1UG20CS110

ABSTRACT!

Need and significance of the problem

ABSTRACT

The main objective is to create a book recommendation system for users. Recommender systems are really critical in some industries as they can generate a huge amount of income when they are efficient or also be a way to stand out significantly from competitors. Collaborative Filltering, Content-Based Filtering is used. And in Collaborative filtering, user-Based and Item-Based Collaborative filtering is used.

A recommendation system broadly recommends items to the user best suited to their tastes and traits. It uses the user's previous data and other user's data to give new recommendations.
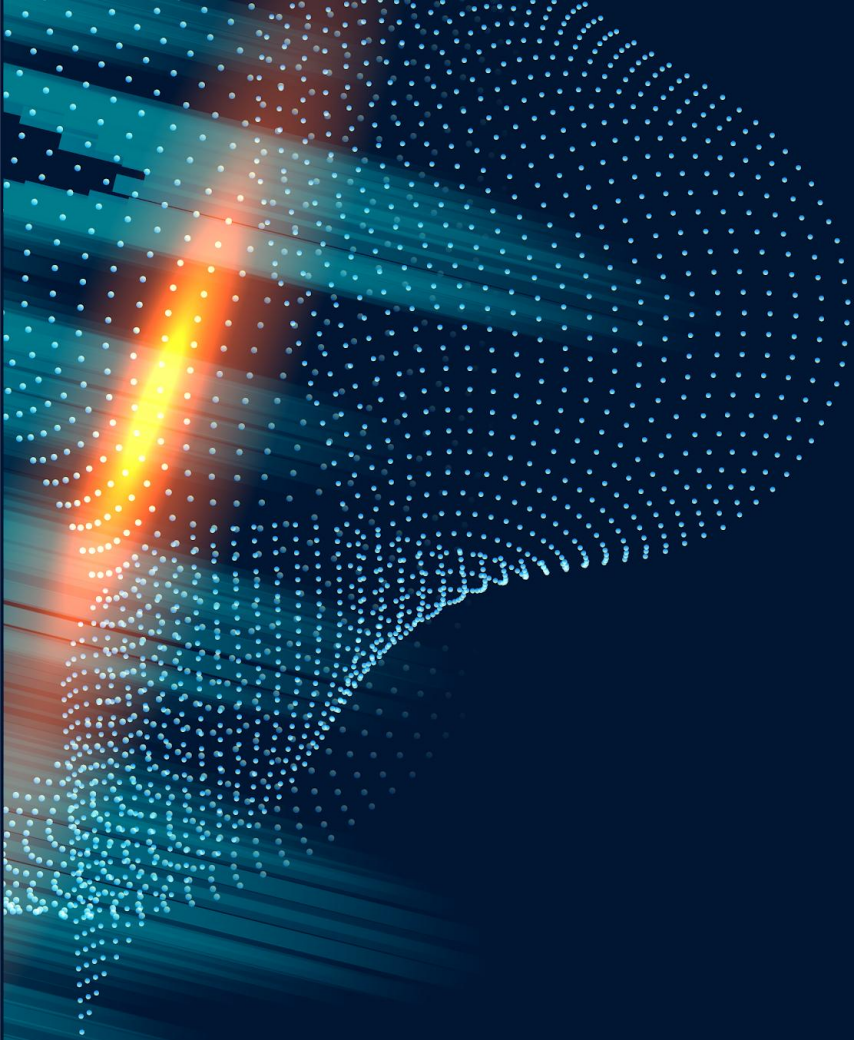
SCOPE!

Novelty of the project opted

SCOPE

A book recommendation system is a type of recommendation system where we have to recommend similar books to the reader based on his interest. The books recommendation system is used by online websites which provide ebooks like google play books, open library, good Read's, etc.

DATASET!

Novelty of the project opted

DATASET

For this Project, we used Book-Crossing(BX) Dataset. This Dataset contains 278,858 users(anonimyzed but with demographic information) providing 1,149,780 ratings(explicit and implicit) about 271,379 books.

Dataset Description:

i. BX-Books.csv has 8 columns which include information about books.We can uniquely identify each book with the help of ISBN. The tile, author and publisher and year of publication of each book has been listed. The last three columns include URLs for different sizes of images.

ii. BX-Users.csv: lists the age and location of a user along with their User-ID.

iii.BX-Book-Ratings.csv: contains book ratings given by users. The User-ID and ISBN have also been provided.

IMPLEMENTATION DETAILS!

Novelty of the project opted

IMPLEMENTATION DETAILS!

1. Modules used:
   pycountry_convert, wordcloud, surprise, numpy , pandas, gradio,
   matplotlib,missingno,seaborn.

2.Dataset Collection:
   BX_book.csv

3.Model:
   1.Collabrative filtering
   2.Content based
   3.Item Based
   4.Machine Learning based

4. UI:
   using Gradio.

PROJECT
DEMONSTRATION

PROJECT DEMONSTRATION

For this Project, we used Book-Crossing(BX) Dataset. This Dataset contains 278,858 users(anonimyzed but with demographic information) providing 1,149,780 ratings(explicit and implicit) about 271,379 books.

Dataset Description:

.BX-Books.csv has 8 columns which include

information about books.We can uniquely identify each book with the help of ISBN. The tile, author and publisher and year of publication of each book has been listed. The last three columns include URLs for different sizes of images.
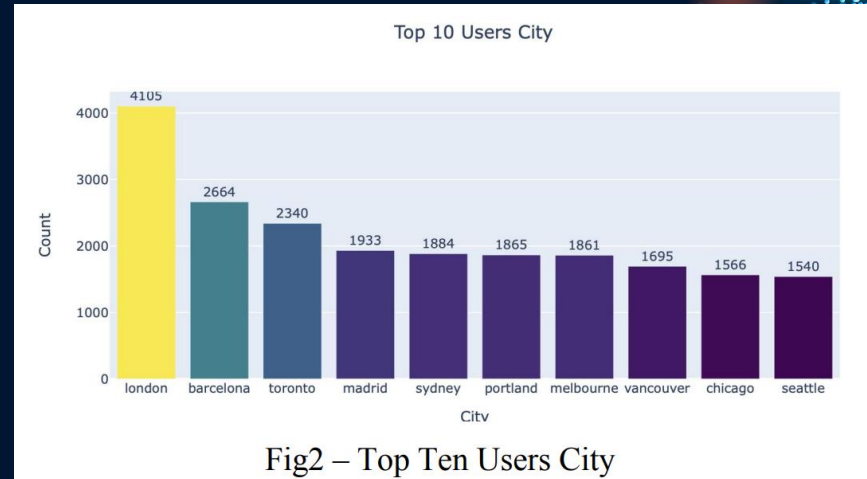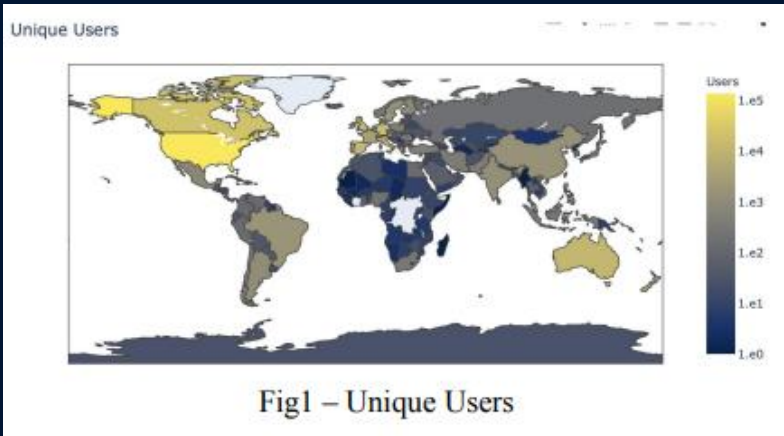
ii.BX-Users.csv: lists the age and location of a user along

with their User-ID.

iii.BX-Book-Ratings.csv: contains book ratings given by

users. The User-ID and ISBN have also been provided.

Exploratory Data Analysis

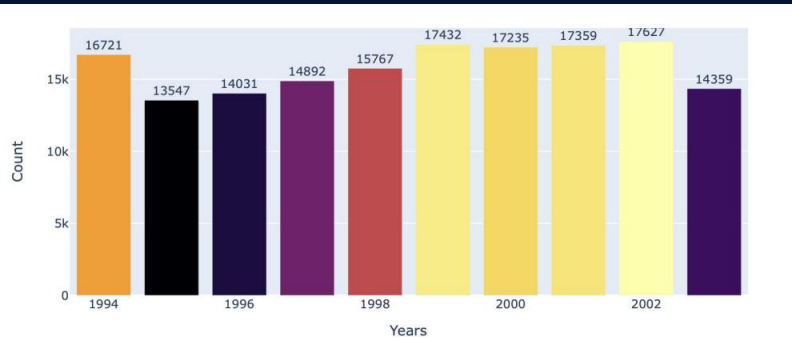Fig1 – Unique Users

Fig2 – Top Ten Users City

Exploratory Data Analysis

Fig3 – Top Ten Number Of Books Published By years
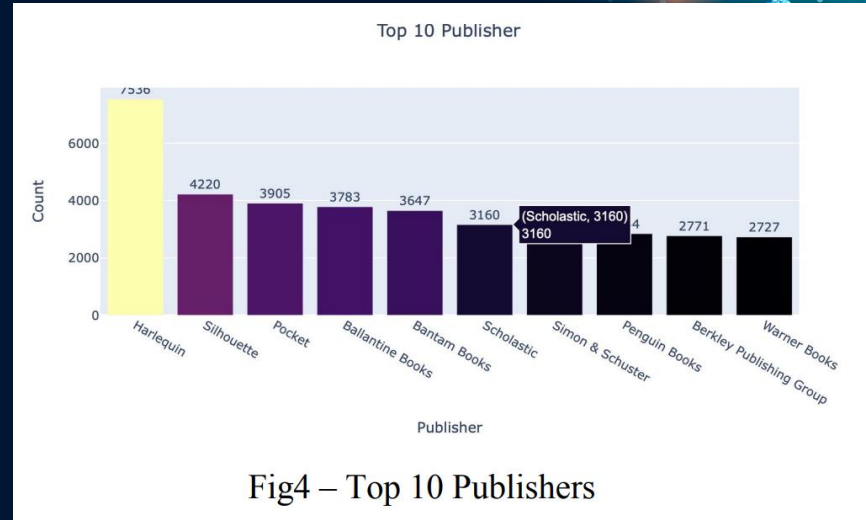
Fig4 – Top 10 Publishers
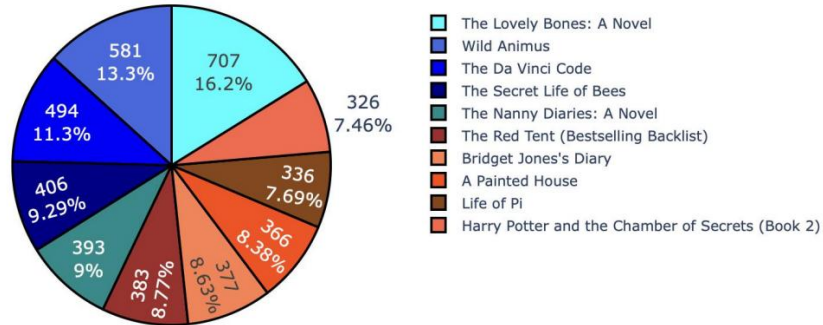
Exploratory Data Analysis

Fig5 – Most Reviewed Books
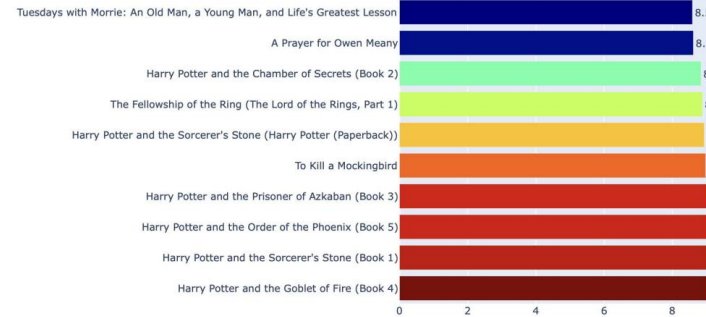

Fig6 – Top Ten Best Rated Books

PROJECT DEMONSTRATION

Modules used: pycountry_convert, wordcloud, surprise, numpy , pandas, gradio, matplotlib,missingno,seaborn.

We used:

1. Collaborative filtering and in that user- user based and item - item based filtering.

    1. KNN Algorithm for Collaborative-Based Filtering. Collaborative Filtering Using k-Nearest Neighbors (kNN) kNN is a machine learning algorithm to find clusters of similar users based on common book ratings, and make predictions using the average rating of top-k nearest neighbors.

    2. By iterating the columns, we are subtracting the mean values. compute_similarity() is calculating the cosine similarity of two users. create_similarity_matrix() is creating the cosine similarity matrix between all the users. get_neighbors() is calculating the neighbours on the basis of matrix created. recommend() is recommending the top users that a user is matched with.

    3. DRAWBACKS: Scalability and Scarcity

PROJECT DEMONSTRATION

2. CONTENT-BASED FILTERING :

      1. Content-based filtering, also referred to as cognitive filtering, recommends items based on a comparison between the content of the items and a user   profile.

      2. The content of each item is represented as a set of descriptors or terms, typically the words that occur in a document

      3.After cleaning and preparing the data, we can use TF- IDF representation to make our recommender system.

PROJECT DEMONSTRATION

3.Model Based :
        1. We will limit the books in this whose ratings are given atleast 5 times. We are normalizing the ratings on the basis of gaussian normalization.

        2. Baselines can be estimated in two different ways:
                1. Stochastic Gradient Descent (SGD)
                2. Alternating Least Squares (ALS)

* With the implementation of these models we can print the required output in our UI

* For UI we have used Gradio.

PROJECT DEMONSTRATION

3.Model Based :
        1. We will limit the books in this whose ratings are given atleast 5 times. We are normalizing the ratings on the basis of gaussian normalization.

        2. Baselines can be estimated in two different ways:
                1. Stochastic Gradient Descent (SGD)
                2. Alternating Least Squares (ALS)

* With the implementation of these models we can print the required output in our UI
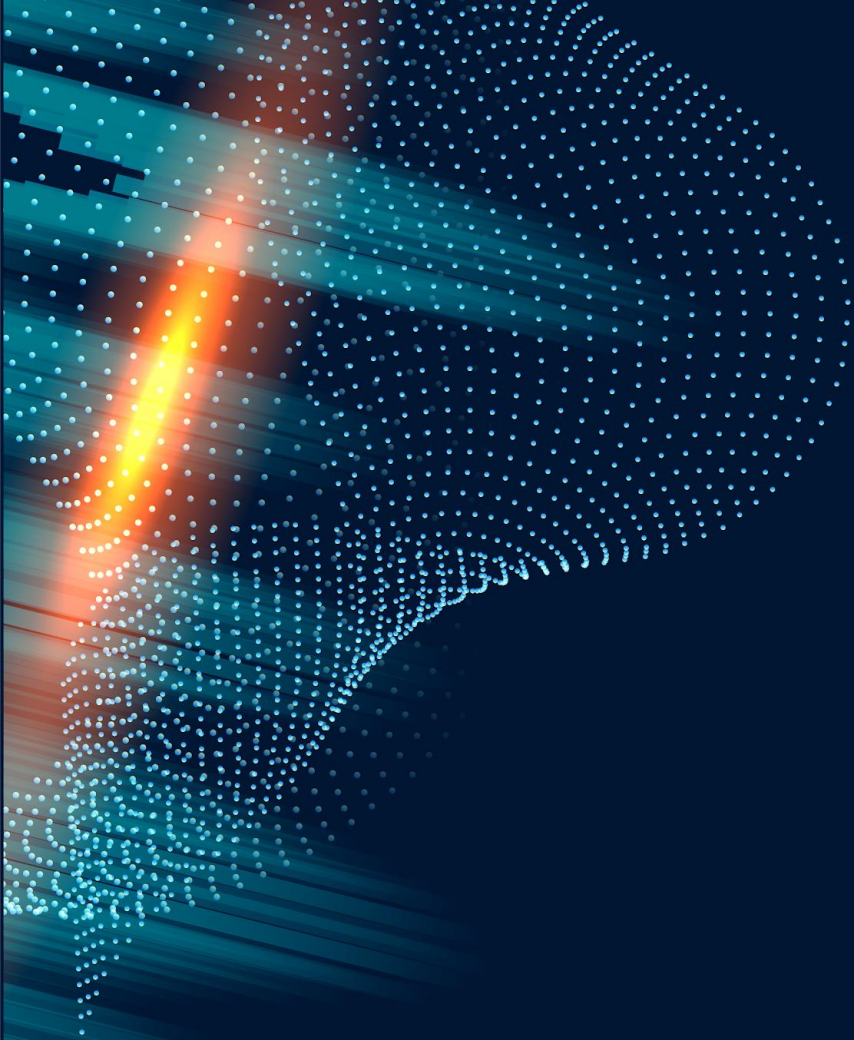
* For UI we have used Gradio.

RESULT AND DISCUSSION

Result and Discussion

- Using Collabarative filtering

User - based

	ISBN	bookTitle	bookAuthor	yearOfPublication	publisher
52	1558744150	Chicken Soup for the Woman's Soul (Chicken Sou...	Jack Canfield	1996	Health Communications
97	0842329129	Left Behind: A Novel of the Earth's Last Days ...	Tim Lahaye	1996	Tyndale House Publishers
134	0061009059	One for the Money (Stephanie Plum Novels (Pape...	Janet Evanovich	1995	HarperTorch
219	0345370775	Jurassic Park	Michael Crichton	1999	Ballantine Books
267	080410753X	The Kitchen God's Wife	Amy Tan	1992	Ivy Books
385	059035342X	Harry Potter and the Sorcerer's Stone (Harry P...	J. K. Rowling	1999	Arthur A. Levine Books
489	0399146431	The Bonesetter's Daughter	Amy Tan	2001	Putnam Publishing Group
786	043935806X	Harry Potter and the Order of the Phoenix (Boo...	J. K. Rowling	2003	Scholastic
1726	0811801802	Sabine's Notebook: In Which the Extraordinary ...	Nick Bantock	1992	Chronicle Books
1779	0877017883	Griffin & Sabine: An Extraordinary Corresp...	Nick Bantock	1991	Chronicle Books

Result and Discussion

- Collabarative filtering

Item-based

	ISBN	bookTitle	bookAuthor	yearOfPublication	publishe
2949	0385419287	The 776 Stupidest Things Ever Said	Ross Petras	1993	Main Street Book
1217	0140183515	Just So Stories (Penguin Twentieth-Century Cla...	Rudyard Kipling	1990	Penguin Book
1757	0394747232	Maus a Survivors Tale: My Father Bleeds History	Art Spiegelman	1986	Pantheon Book
3296	0373078110	Hidden Star (The Stars Of Mithra) (Harlequin ...	Nora Roberts	1997	Silhouett
2864	0373078358	Secret Star (The Stars Of Mithra) (Harlequin ...	Nora Roberts	1998	Silhouett
1183	0373218419	Blithe Images	Nora Roberts	2003	Silhouett
2791	0373218486	Going Home: Unfinished Business/ Island of Flo...	Nora Roberts	2002	Silhouett
2477	0373218540	Dangerous	Nora Roberts	2002	Silhouett
383	0373242328	The Perfect Neighbor (The Macgregors) (The Mac...	Nora Roberts	1999	Silhouett
1553	0373243286	Irish Rebel (Special Edition, 1328)	Nora Roberts	2000	Silhouett

Result and Discussion

- Content - based filtering

	ISBN	bookTitle	bookAuthor	yearOfPublication	publisher
277	0385486804	Into the Wild	Jon Krakauer	1997	Anchor
428	1559029838	Call of the Wild	Jack London	1994	Selldowns/no More Orders
2616	0515132292	Wild	Lori Foster	2002	Jove Books
2060	1551668777	So Wild A Heart	Candace Camp	2002	Mira
3247	0843953004	Wild Desire	Phoebe Conn	2003	Leisure Books
3390	0399149279	Wild Pitch	Mike Lupica	2002	Putnam Publishing Group
2612	0743437128	Wild Orchids : A Novel	Jude Deveraux	2003	Atria Books
907	0380812037	On a Wild Night (Cynster Novels)	Stephanie Laurens	2002	Avon
1858	0373240880	Waiting For Nick (Those Wild Ukrainians) (Sil...	Nora Roberts	1997	Silhouette
957	0671769316	NEW ROADSIDE AMERICA : THE MODERN TRAVELER'S G...	Doug Kirby	1992	Fireside

RESULT AND DISCUSSION

- Using machine learning based models

	ISBN	Rating	bookTitle	bookAuthor	yearOfPublication	publisher
3367	0345342968	3.035087	Fahrenheit 451	RAY BRADBURY	1987	Del Rey
675	0446310786	3.051375	To Kill a Mockingbird	Harper Lee	1988	Little Brown &; Company
2739	0345361792	3.032526	A Prayer for Owen Meany	John Irving	1990	Ballantine Books
6117	059035342X	3.043820	Harry Potter and the Sorcerer's Stone (Harry P...	J. K. Rowling	1999	Arthur A. Levine Books
6730	0671027360	3.032648	Angels & Demons	Dan Brown	2001	Pocket Star
3227	0316666343	3.047491	The Lovely Bones: A Novel	Alice Sebold	2002	Little, Brown
1892	0385504209	3.061188	The Da Vinci Code	Dan Brown	2003	Doubleday
5715	0142001740	3.057554	The Secret Life of Bees	Sue Monk Kidd	2003	Penguin Books
8264	0156027321	3.034544	Life of Pi	Yann Martel	2003	Harvest Books
4925	043935806X	3.031039	Harry Potter and the Order of the Phoenix (Boo...	J. K. Rowling	2003	Scholastic

RESULT AND DISCUSSION

EXPECTED OUTPUT IN THE UI

FUTURE WORK

FUTURE WORK

We can create the hybrid model i.e. it is special type of recommendation system which can be considered as the combination of the content and collaborative filtering method. Combining collaborative and content-based filtering together may help in overcoming the shortcoming we are facing at using them separately and also can be more effective in some cases.

THANK YOU