Project – Book Recommendation System
Team Member Details

1) Veerraju Chowdary Rimmalapudi (700731964)

Motivation:

In the past shopping used to happen in physical stores. Users must select the products which are available in the store. Nowadays shopping is no longer a mundane task with the advent of internet shopping including personal preferences that can be personal brands or choices people are preferring more personalized products to be reviewed. This case is applicable to food, beauty and many more. Recent applications of recommendation systems help industries to send the recommendations to customers with the history of the customer behavior.

Significance:

The number of avid readers across the world is increasing. The main reason for this increase is the evolution of different reading modes like e-books. To reach out to more readers publishers are looking for the best platforms which have a strong book recommender system in built.

For the industries mainly for e-commerce vendors huge revenue is generated using recommendation systems. Top industries are investing in building robust recommender systems for the users.

Objectives:

Book recommendation is a challenging task since the reader picks up the book because of various reasons that could be recommended by a friend, he/she read the same author in the past. Existing methods of content based, and collaborative filtering have their own limitations. Recommending similar items continuously makes the customer choices limited to certain preferences.

The objective of the project is to:

- Building a hybrid recommender system that combines the results of the existing methods.

Features:

Main features of the project are:

- Building a hybrid recommender system to meet the reader criteria.

- Using mixed hybrid recommendation technique. Main advantage of using mixed type of hybrid recommendation system is it generates the candidates simultaneously and trains to the best model for better performance.

Dataset:

Book recommendation system uses 3 files: books.csv, users.csv and ratings.csv files.

Books files contain ISBN of each book, title of the book, book author, year of publication, publisher of the book and image URLs. Since ISBN is unique column to identify the book invalid ISBNs are removed from the data. The original source of the data is collected from AWS. In the User file demographic information of the user is mentioned age and location along with the user id Which is anonymized as integers. The ratings file contains information about the book. These ratings are. reviewed on the scale of 1-10. Higher ratings are considered as higher appreciation which is known as explicit ratings whereas "0" ratings are implicit ratings.

Dataset link

Detail design of Features:
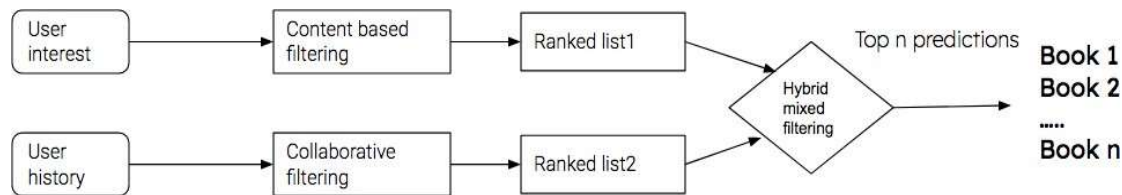
fig.1.Block diagram of hybrid mixed filtering

The basic approaches of recommender systems have pros and cons. Content, and collaborative filtering methods are not considered as strong when we consider the recommendations individually. So here we are proposing a mixed hybrid recommendation system which gives the results based on the mixed method which is robust compared to the individual methods.

Analysis:

Nowadays recommender systems are everywhere it online grocery shopping, apparel and many more. In the past we have implemented content based filtering and collaborative filtering to avoid the shortcomings of individual methods. We are deploying hybrid filter methods [1] [2]. There are multiple types of hybrid recommendation systems namely: switching, weighted, and mixed etc., Mixed approach selects the best model and fits the training set to that model. This process is automatic which is an advantage of mixed hybrid method [3].

```
Books by same Author:

Harry Potter and the Goblet of Fire (Book 4)
Harry Potter and the Order of the Phoenix (Book 5)
Harry Potter and the Sorcerer's Stone (Book 1)


Books by same Publisher:

The Seeing Stone
The Story of the Seagull and the Cat Who Taught Her To Fly
The Slightly True Story of Cedar B. Hartley: Who Planned to Live an Unusual Life
```

Fig.2. Results of content-based filtering.

```
Input Book:
Harry Potter and the Sorcerer's Stone (Harry Potter (Paperback))

RECOMMENDATIONS:

Harry Potter and the Prisoner of Azkaban (Book 3)
Harry Potter and the Goblet of Fire (Book 4)
Harry Potter and the Order of the Phoenix (Book 5)
```

Fig.3. results of collaborative filtering.

```
Input Book:

Harry Potter and the Sorcerer's Stone (Harry Potter (Paperback))

Recommended Books:

Harry Potter and the Goblet of Fire (Book 4)
Harry Potter and the Prisoner of Azkaban (Book 3)
Harry Potter and the Sorcerer's Stone (Book 1)
```
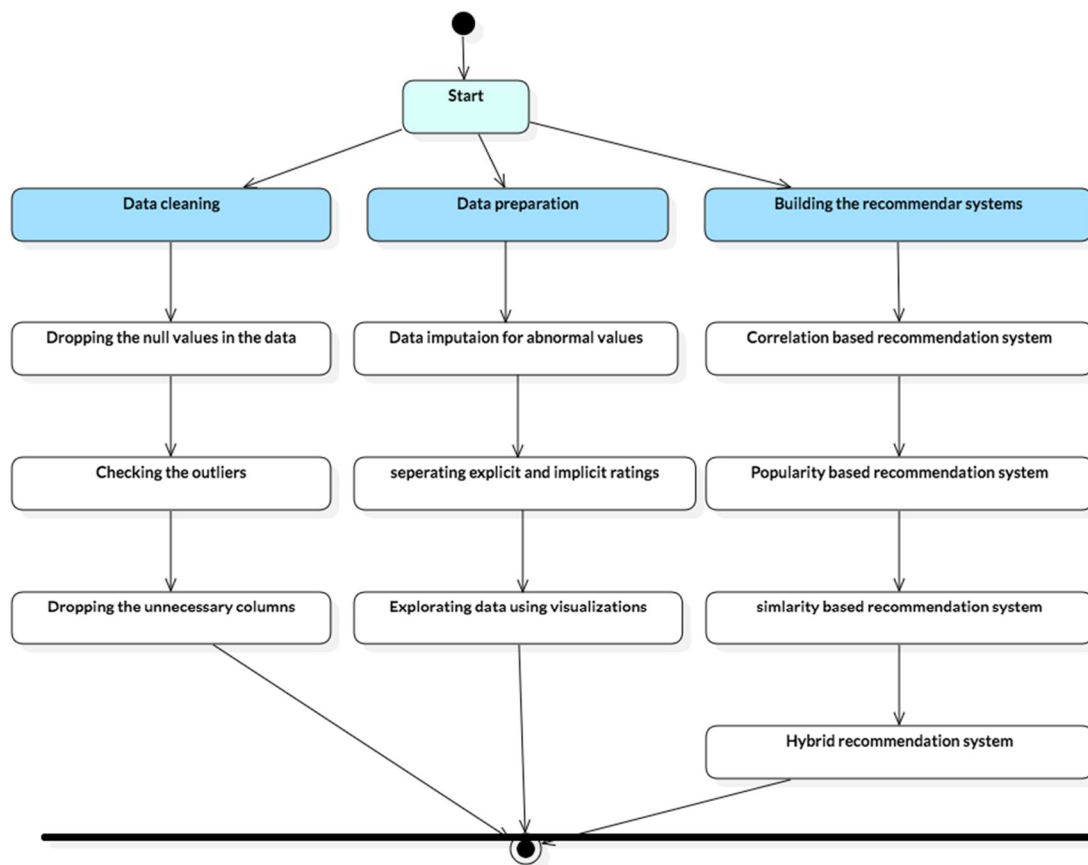
Fig.4. Results of hybrid filtering.

Implementation:

Implementation of book recommendation systems is divided into stages: data cleaning, data preparation and building recommendation systems.

Since the dataset consists of 3 files of information each file needs to be cleaned separately. **Data cleaning** includes removing the null values, examining the distribution numerical columns to check any abnormalities in the data. In this stage we retain the necessary features and remove the unnecessary features. We have dropped the image URL columns.

recommendation system which gives the results based on the mixed method which is robust compared to the individual methods.

At the stage of data cleaning, we have identified some abnormal values in age and publishing year columns. The age column contains a considerable amount of 0 values in the data, so we have ignored the rows with age range 0-80 as normal range and ignored the other values in the dataset. Another column with abnormal values is the publishing year which is greater than the current year and the range is between 0-2050 which is abnormal. We have considered an imputation method to fill the abnormal values that are mean. The ratings column also divided into implicit and explicit we have built the recommendation system on explicit ratings and ignored the implicit ratings.

After the data preparation various recommender systems are built on the final data. The first method is popularity-based recommendation system:

- All the ratings are summed up and based on the user preference top n number of books are displayed the popular books to read.
- Similar author recommendation system displays the books of same authors from the popular books.
- Similar publisher recommendation system displays the books of same publisher from the popular books.
- User-item collaborative filtering system recommends the books based on the cosine similarity.
- Correlation based recommendation system recommends books based on Pearson's correlation.
- Content based recommendation system recommends the list of books which are higher than the set threshold of popular ratings.
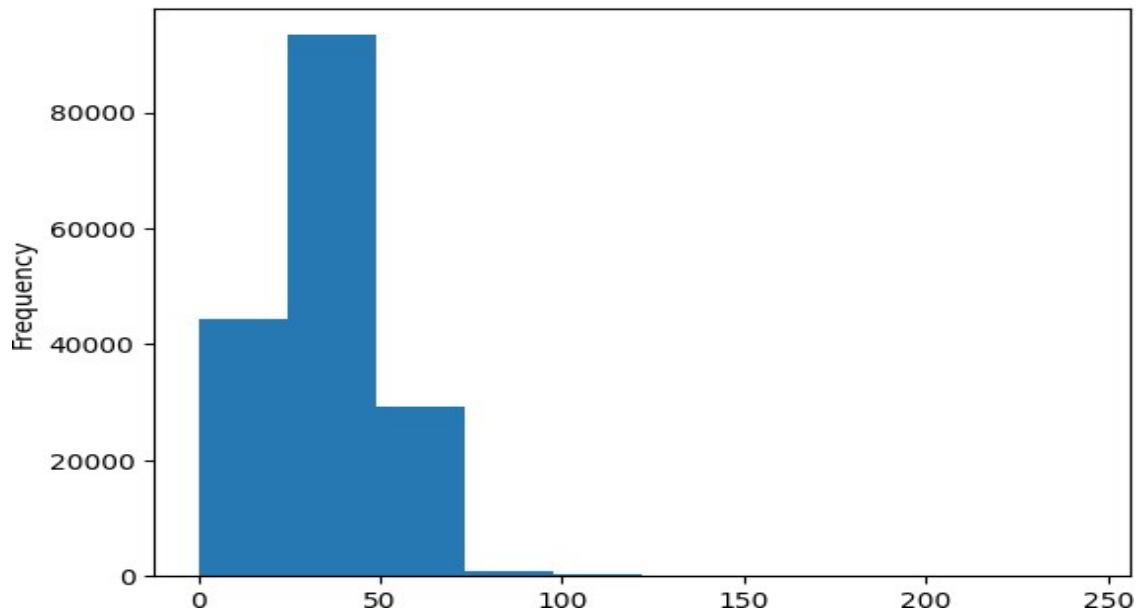
Hybrid recommender system recommends the list of books by combining content based filtering and collaborative filtering suggestions.
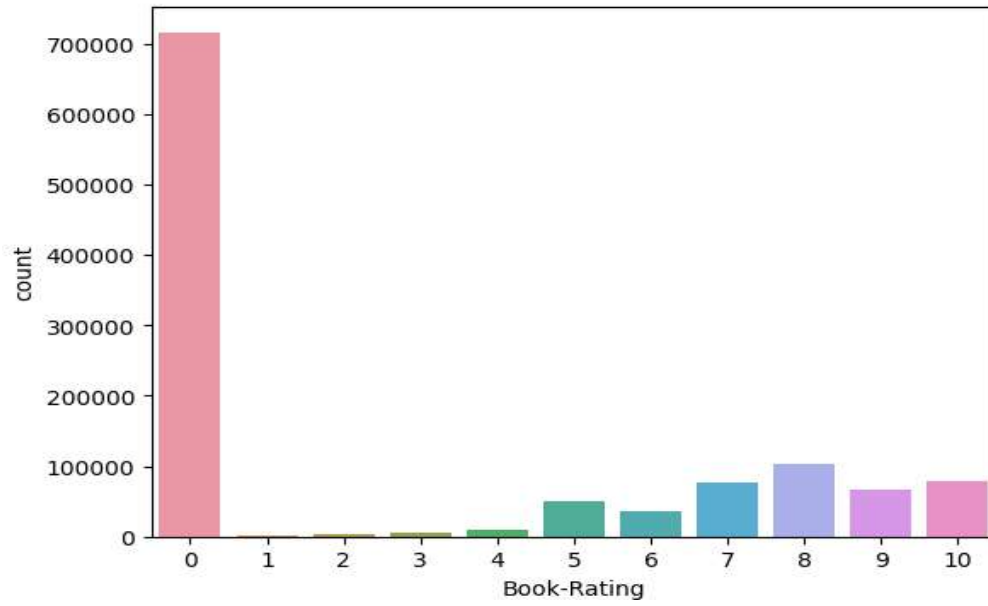
Preliminary Results:

As a preliminary analysis we have used exploratory data analysis to understand the structure of the data.
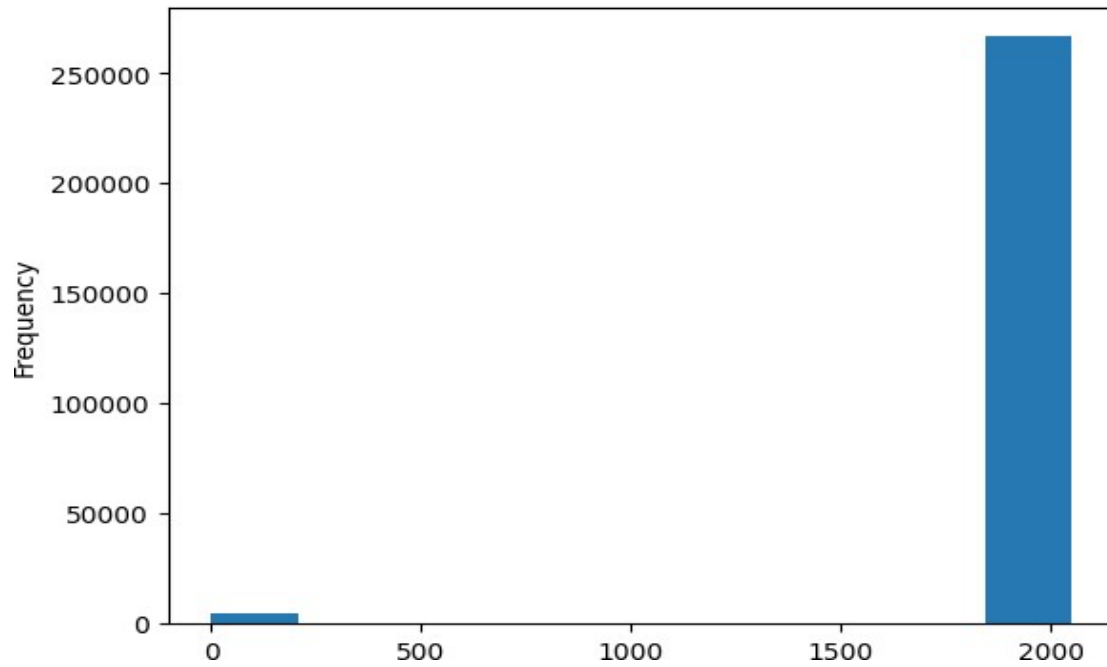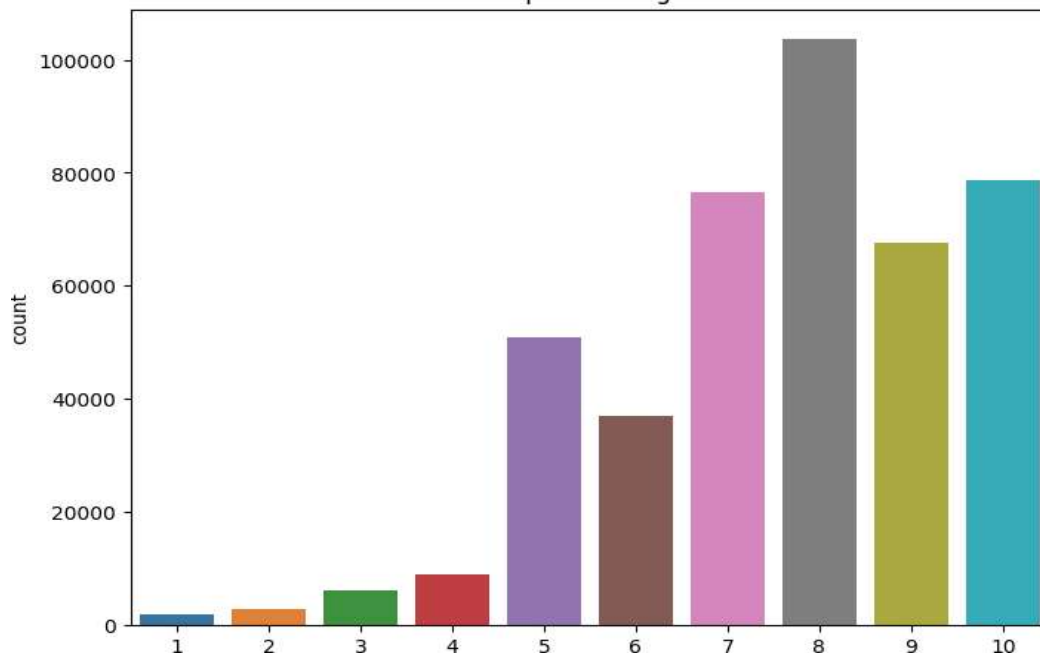

Fig.1.Histogram of year of publication

Fig.2.distribution of age of the users

Fig.3.countplot for book-ratings (explicit and implicit)

Fig.4.countplot of explicit book ratings

Project Management:

https://docs.google.com/spreadsheets/d/1F0gLEcUkL7izPY345es0xiVTJa-at-LrZty7oPO3ZPQ/edit#gid=0

REFERENCES:

[1].R. Ramesh and S. Vijayalakshmi, "Improvement to Recommendation system using Hybrid techniques," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2022, pp. 778-782, doi: 10.1109/ICACITE53722.2022.9823879.

[2].R. Turnip, D. Nurjanah and D. S. Kusumo, "Hybrid recommender system for learning material using content-based filtering and collaborative filtering with good learners' rating," 2017 IEEE Conference on e-Learning, e-Management and e-Services (IC3e), Miri, Malaysia, 2017, pp. 61-66, doi: 10.1109/IC3e.2017.8409239.

[3].P. Arunruviwat and V. Muangsin, "A Hybrid Book Recommendation System for University Library," 2022 26th International Computer Science and Engineering Conference (ICSEC), Sakon Nakhon, Thailand, 2022, pp. 291-295, doi: 10.1109/ICSEC56337.2022.10049318.