Book Recommendation System

From
Aashutosh Kumar Sah

To
Mindrisers Institute of Technology

Project Overview

The Book Recommendation System suggests books based on collaborative filtering using cosine similarity. The model recommends similar books based on user ratings and behavior. The complete data pipeline was built from scratch and deployed as an interactive web app using Streamlit on localhost.

Objectives

To build a predictive model and deploy it as an interactive web application using Streamlit, following the complete data science life cycle.

Dataset Description

The dataset is taken from Kaggle - Book Recommendation Dataset and contains the following CSV files:

File	Description
Books.csv	Contains book metadata (title, author, etc.)
Users.csv	User demographic info
Ratings.csv	Book ratings by users

Data Collection and Exploration

- Loaded all 3 datasets using Pandas.
- Inspected shape, null values, and structure.
- Printed basic statistics for each dataset.

Key Observations:

- Books and ratings datasets had nulls in some important fields like ISBN or Book-Rating.
- Ratings.csv had over a million entries.

Data Cleaning and Transformation

Steps Taken:

- Dropped null values from important fields.
- Removed duplicates.
- Merged Books.csv and Ratings.csv using ISBN.
- Filtered users with more than 200 ratings (active users).
- Filtered books with at least 50 ratings (popular books).
- Created a pivot table for books vs. users.

Exploratory Data Analysis (EDA)

Visual Insights:

- Top 10 Most Rated Books visualized using a bar chart.

- Book Ratings Distribution shows most ratings were 0 or in low range.
- User Age Distribution shows peak reader ages between 20–40.

Tools used: matplotlib, seaborn

Feature Selection

- Selected active users and popular books to create a more relevant and denser pivot matrix.
- Created a pivot table with Book-Title as rows and User-ID as columns.
- Used this matrix as input for similarity calculation.

Model Development

- Used **Cosine Similarity** from sklearn.metrics.pairwise to find similar books based on user ratings.
- Stored the similarity matrix and implemented a recommend() function.

Model Characteristics:

- Collaborative filtering (user-based and item-based behavior)
- No content-based filtering (e.g., genres, summaries)

Model Evaluation

Since this is a recommendation system, we evaluated by manually checking:

- **Recommendation quality** (using recommend("1984"), etc.)
- **Popularity-based recommendation** using books with high ratings and counts.

Model Deployment – Streamlit Web Application

Features Implemented:

- Select a book to get 5 similar recommendations.
- Shows image, title, author (with by), year, and publisher.
- Book search with filtering.
- Vertical layout with responsive cards.

- Footer: "Made with ❤️ by Aashutosh Kumar Sah".

To Run the App:

- Place Books.csv, Ratings.csv in a data/ folder.
- Run this in terminal:

Step 1:

```
pip install streamlit
pip install scikit-learn
```

Step 2:

```
streamlit run app.py
```

Conclusion

This project successfully implemented:

- End-to-end data science lifecycle.
- Collaborative filtering model for recommendations.
- Beautiful, user-friendly app using Streamlit.
- Practical use of cosine similarity and EDA for feature understanding.