BookWise Recommender: Personalized Book Recommendations

A Machine Learning Project

Anoush A Shirazi

August 2025

Project Overview

- **Objective**: Build a book recommendation system using collaborative filtering, content-based filtering, and popularity-based methods.
- **Dataset**: Books, users, and ratings datasets.
- **Technologies**: Python, Pandas, Scikit-learn, SciPy, Streamlit, Matplotlib, Seaborn.
- **Deliverables**: Modular code, Jupyter notebooks, Streamlit web app, visualizations, and demo media.

Methodology

1. **Data Preprocessing**: Cleaned datasets by removing duplicates, handling missing values, and filtering ratings.
2. **Models**:
   - Collaborative Filtering (SVD): Latent factor modeling.
   - Content-Based: TF-IDF and cosine similarity on book titles.
   - Popularity-Based: Top-rated books by count.
   - Hybrid: Combines all three for robust recommendations.
3. **Evaluation**: RMSE on a sample of ratings.
4. **Visualization**: Heatmaps, PCA plots, rating distributions.
5. **Deployment**: Streamlit app for interactive recommendations.

Results

- **Personalized Recommendations**: Accurate suggestions using hybrid model.
- **Model Performance**: Competitive RMSE (exact value depends on dataset size).
- **User Insights**: Visualized rating patterns and user demographics.
- **Interactive App**: Streamlit interface for real-time recommendations.
- **Scalability**: Modular code adaptable to other domains (e.g., movies, e-commerce).

Visualizations

- **Rating Distribution**: Shows user rating patterns.
- **Heatmap**: User-book interaction matrix for top users and books.
- **PCA Plot**: User clustering in SVD-reduced space.

See demo/rating_distribution_plot.png for an example visualization.

Streamlit App

- Interactive web interface for recommendations.
- Features:
  - Select User ID and Book Title.
  - Display recommendations from all models.
  - Evaluate model performance (RMSE).
- Screenshot available in `demo/streamlit_screenshot.png`.
- Demo video in `demo/Streamlit_Demo.mp4`.

Conclusion

- **Achievements**: End-to-end ML pipeline with robust models and interactive deployment.
- **Skills Demonstrated**: Machine learning, NLP, data visualization, web development.
- **Business Impact**: Enhanced user experience, scalable framework, actionable insights.
- **GitHub Repository**: https://github.com/anoushirazi/ml-book-recommender