

Book Recommendation

Abstract

Using a dataset from Kragger, we will build a recommendation on books. The recommendation systems will be based on collaborative filtering at our first attempt, first with item-based then with user-based. Next, we will utilize a content-based approach using the books' titles and authors. Finally, we will combine these techniques to form a hybrid method.

Problem Statement and Business Case

Recommendation systems are one of the most integral parts of many online systems. A good recommendation system not only increases a company's short-term revenue, but also supports customer satisfaction, which is vital to the company's long-term benefit. In 2016, Netflix estimated that its recommendation is worth \$1 billion per year in revenue. Netflix even held a competition in designing recommendation systems with a grand prize of \$1 million. With people now spending more time online due to Covid, a sound recommendation system is even more essential to keep people engaged. In this project, we will build a recommendation system on books.

Data Science Workflow

- Collaborative Filtering: Item-base vs. User-base
- Content-based approach: Title of the book, Author
- Model-based approach: clustering, dimension reduction on the consumer-item matrix
- Help readers find good books

Data Collection

[Book-Crossing: User review ratings | Kaggle](#)

This data contains over 1 million customer ratings on 200,000 books from 200,000 customers. Each line in the data files corresponds to an individual rating for a book. Some helpful features are book_ratings, book_title, book_author, user_id, user_location, user_age.

Data Processing, Machine Learning

- Collaborative Filtering
- Content-based System
- Hybrid techniques
- Clustering, dimensionality reduction

Deliverables

- Jupyter notebook
- 10-12 pages report