

Capstone Project-4

Book Recommendation System

Instructions:

- i) Please fill in all the required information.
- ii) Avoid grammatical errors.

Team Member's Name, Email, and Contribution:
<p>1. Prerit Tyagi</p> <ul style="list-style-type: none">• Exploratory data analysis – Users, Books, and Rating data analysis.• Data Wrangling – checking missing values, outliers, and feature modifications.• Fitting Models – Collaborative Filtering(NMF, SVD, and kNN algorithms).• Presentation video• Technical documentation.• Project summary
Please paste the links.
<p>GitHub Repo link:- https://github.com/Preritp2/Book_Recommendation_System_Capstone_Project_4</p> <p>Drive link:- https://drive.google.com/drive/folders/1eCER6Wp1PWIXuY1Mae41_bBnVidBOmj?usp=sharing</p>
Please write a summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)

During the last few decades, with the rise of YouTube, Amazon, Netflix, and many other such web services, recommender systems have taken more and more place in our lives. From e-commerce (suggest to buyers articles that could interest them) to online advertisement (suggest to users the right contents, matching their preferences), recommender systems are today unavoidable in our daily online journeys.

Generally, recommender systems are algorithms aimed at suggesting relevant items to users (items being movies to watch, text to read, products to buy, or anything else depending on industries).

Recommender systems are 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.

The Book-Crossing dataset comprises 3 files.

- **Users dataset contains User-ID, Location, and Age.**
- **Books dataset contains ISBN, Book-Title, Book-Author, Year-Of-Publication.**
- **Publisher. The rating dataset contains Book-Rating expressed on a scale from 1-10.**

The problem statement was to build a book recommendation system for users.

We started with data wrangling in which we tried to handle null values, and outliers and performed feature modifications. Next, we did some exploratory data analysis on all 3 datasets and tried to draw observations from the features we had in the datasets.

Next, we implemented the Popularity Based technique to tackle the Cold-Start problem. This approach recommended the most popular books to every new user who just signed in.

After that, we applied Collaborative Filtering (CF); a Model-based Memory based approach. In the model-based approach, we implemented Non-negative Matrix Factorization (NMF) and Singular Value Decomposition (SVD). And in Memory based approach we implemented User-Item and Item-Item algorithms to recommend books to similar users using cosine similarity and kNN algorithm.

It was observed that in the case of Model-based CF, the SVD technique worked way better than NMF with a lower Mean Absolute Error (MAE). And in the case of Memory based CF, item-item performed better than user-item because of lower computation requirements.

However, there can be more modifications to this analysis. One could implement Content-Filtering based recommendation system if more information regarding the book's dataset is given, namely features like Genre, Description, etc.