

Capstone Project

Submission

Instructions:

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

Team Member's Name, Email and Contribution:

Rahul Shah: shahrahoool010@gmail.com

1. Exploratory Data Analysis
2. Data Wrangling – handling missing values, treating outliers, features modification.
3. Model Building, Model Evaluation
4. Presentation, Technical documentation.

Please paste the GitHub Repo link.

Github Link:- <https://github.com/rahoool010/Book-Recommendation-System>

Please write a short 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. Recommendation systems have been an important part of e-commerce on the web for the customer to suggest items what they would be interested.

In a very general way, 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 really 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 main objective is to create a book recommendation system for users.

Techniques Implemented:

- Popularity Based System
- Collaborative Filtering using kNN
- Collaborative Filtering using SVD

Throughout the study, we performed various steps to build a book recommender system. We started with data wrangling in which we tried to handle null values and performed feature modifications. Next, we did some exploratory data analysis and tried to draw observations from it. Finally, A book recommendation system was designed using different filtering techniques. After implementing Collaborative Filtering model (SVD matrix factorization), we were satisfied with the results. We observe that, we got Recall@10 (43%) and Recall@15 (52%), which is fair enough for such a large dataset.

Conclusions from analysis includes:

- “Selected Poems” were read more by the users.
- Majority of the users were from USA.
- “Harlequin” has published the most number of books.
- Among authors, “Agatha Christie” has written the most number of books followed by “William Shakespeare” and “Stephen King”.

In this study, we limit ourselves to collaborative filtering methods. However the information about the books can be easily obtained using the ISBN number of the book, to overcome the cold start problem when the new user becomes a member or when a new book is released. A hybrid of content based and collaborative filtering methods can be used to generate more relevant recommendations.