Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
1. Puroshotam Kumar Singh: purshotamsingh61@gmail.com • Exploratory data analysis – Users, Books and Ratings data analysis. • Data Wrangling – checking missing values, outliers, features modification. • Fitting Models – Collaborative Filtering(NMF, SVD and kNN algorithms). • Presentation, Technical documentation. 2. Vikram Pandey: pandey.vicky@yahoo.com • Exploratory data analysis – Users, Books and Ratings data analysis. • Data Wrangling – checking missing values, outliers, features modification. • Fitting Models – Collaborative Filtering(NMF, SVD and kNN algorithms). • Presentation, Technical documentation.
Please paste the GitHub Repo link.
Puroshotam's Github Link:- https://github.com/PuroshotamSingh/Book-Recommendation-System Vikram's Github Link:- https://github.com/vickypandey07/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. 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. 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 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. Ratings 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, 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 Popularity Based technique to tackle the Cold-Start problem. In this approach, most popular books were recommended to every new user who just signed in.

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

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

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