CP421 Project Proposal

Hybrid Personalized Book Recommendations Project

Group Members: Jose Manuel Ruiz , Usama Mohiuddin, Nathan Boshart, Filip Kossowski, Kevin Lee, Mohamad el mansour

Due Date: 2023-11-13

Objective:
Our goal is to create an application that takes in the user preferences and geographical location such as taking in the users age, geographical location, to recommend books. With these inputs, this application will output a list of recommended books and it will also output an accuracy so that the user can see how accurate this recommended book was compared to other peoples ratings. This rating will help users gauge how well these recommendations match general preferences, enhancing user trust in our system.

Methodology:
An integration between Collaborative filtering and content-based filtering that brings a Hybrid Approach for our recommendation technique. Collaborative filtering will be used to identify books preferred by similar users. To implement collaborative filtering, we will analyze user patterns, identifying books liked by similar users, and to do this we either apply Pearson correlation or cosine similarity to measure the similarity. Meanwhile Content-Based filtering will used to analyze books such as the author, title and year of publication to find books with similar characteristics. We will start of content based filtering by using the cosine similarity to determine the similarity between the user preferences and check the similarity to the ratings being pulled from the dataset. To summarize collaborative filtering is based on user behaviour and and finding similarities between the users. While content based filtering relies on the characteristics of the items themselves.

Selecting Features:
The provided dataset provides us with several attribute options for distinguishing users preferences. We are provided with the age and location of the user, these attributes will help us create a feature to distinguish users based on commonalities. These user attributes will be combined to create a feature used in our Collaborative filtering methods. User attributes like age and location are helpful to find similarities between users.

The dataset also includes ratings provided by users that will be used to create features for recognizing user preferences. These features will be used in our Collaborative filtering methods since they require us to analyze user similarity to provide accurate book recommendations. The ratings feature is also helpful because it is a quantitative value, a 1 to 5 scale, and will not require much preprocessing to extract.

Similarly, we will use attributes provided for each book to build features for our Content-Based filtering methods. Author, title and year of publication data will be used to build features to distinguish books based on content.

Data Handling:
In the Data Handling phase, we will use Pandas for loading, exploring, and preprocessing our dataset. Key steps include cleaning the data to remove duplicates and handle missing values, transforming data types for efficiency, and feature engineering to extract relevant information. We will also normalize ratings to mitigate scale biases and potentially reduce the dataset size by filtering out less relevant entries. Finally, the dataset will be split into training and testing sets to enable effective model training and validation.

Conclusion:
Our project aspires to bridge the gap between users and their next favorite read. By combining innovative data mining techniques and user-centric design, we aim to create a book recommender system that is not only efficient but also a delight to use.