Book Recommendation System

CAPSTONE -2 Project Proposal - by Debisree Ray

The Problem:

Online recommendation systems are the 'in' thing to do for many e-commerce websites. A recommendation system broadly recommends products to customers best suited to their tastes and traits. This project is focused around building various kinds of book recommendation engines; namely the Simple Generic Recommender, the Content-Based Filter and the User-Based Collaborative Filter. The performance of the systems will be evaluated in both a qualitative and quantitative manner.

The Client:

Any E-commerce business website or online book-selling portal is the potential customer.

The Data:

The data is from the Book Crossing dataset. This dataset has been compiled by Cai-Nicolas Ziegler in 2004, and it comprises of three tables for users, books, and ratings. All three (following) files are available in the CSV dump file (BX-CSV-Dump.zip).

- **BX-Book-Ratings.csv** (referred to as the rating file) has the following data fields:

User ID:	The ID of the reviewer
ISBN:	International Standard Book Number (Unique no. identifying the book)
Book Rating:	Numeric (1-10) showing the rating

- **BX-Users.csv** (referred to as the users' file) has the following data fields:

User ID:	The ID of the reviewer
Location:	City of the reviewer
Age:	Age of the reviewer

- **BX-Books.csv** (referred to the books file) has the following data fields:

ISBN:	International Standard Book Number (Unique no. identifying the book)
Book Title:	Title of the book
Book Author:	Author name
Year of Publication:	Year
Publisher:	Publisher name/company
Image-URL-S:	URL
Image-URL-M:	URL
Image-URL-L:	URL

The Business Problem and the Approach:

The goal of the project is to create a recommendation system of various kinds of book recommendation engines; namely the Simple Generic Recommender, the Content-Based Filter and the User-Based Collaborative Filter.

The first step is the data wrangling and the pre-processing. Next, is to apply the right ML model to build the system. The associated business problems which can be addressed are as follows:

1) Can the ratings depend on the age of the reviewers?

2) What are the unique features affecting the user rating?
3) Is there any relation between the cities and the rating?
4) Which location has the most and the least no. of the reviewers?
5) Do the year of publication and the publisher's house name has any effect on the rating?

Deliverable: The deliverables will be the code(s) on GitHub in the form of Jupyter Notebooks, and a slide deck. This will include a report and I intend to write a blog explaining the code and the results.