ASSIGNMENT # 03: BEE-14

CS-471 Machine Learning

Submission Deadline: 19th Dec 2025

Assignment Title: Book Recommendation System (Content-Based + Collaborative Filtering)

Objective:

You will build a **simple recommender system** by implementing:

1. Content-Based Filtering (CBF)
2. Collaborative Filtering (CF)

and compare their recommendations.

Part1: Dataset Selection

You will use the **Book-Crossing Dataset**, which contains:

- 278,000 users
- 270,000 books
- 1.1 million ratings
- Metadata fields:
    - Book-Title, Book-Author, Year-Of-Publication, Publisher, Book-Description, ISBN, User-ID, Book-Rating

You can download it by searching: **"Book-Crossing Dataset Kaggle"**.

Part 2: Data Preprocessing

- Load the Book-Crossing dataset (ratings + books).
- Clean the dataset: Remove missing values, Remove books with no ratings, Convert ratings to numeric, Choose a subset for faster processing (optional):

Example: books with at least 50 ratings
users with at least 30 ratings

(Document all preprocessing decisions in your notebook.)

Part 3: Content Based Filtering

Objective:

Recommend books similar to a given book using metadata.

Steps:

1. Use Book Title + Author + Publisher
   (optional: Description if available)
2. Create a combined text feature:
3. Compute TF-IDF vectors.
4. Compute cosine similarity between books.
5. For a chosen book: Find top-10 most similar books.

Deliverables:

- Explain how you created text features
- Show TF-IDF matrix shape

- Show top-10 similar books for any selected book

Part 4: Collaborative Filtering
Objective:
Recommend books based on user–item ratings.
Method: Item-Item Collaborative Filtering

- Compute book-book similarity based on user ratings
- Recommend items similar to the books the user rated highly

Steps:

- Create pivot table (users × books)
- Use cosine similarity
- For a selected user:
  Recommend top-5 books not rated by the user

Deliverables:

- Similarity matrix snippet
- Top-5 recommended books for one user

Part 6: Submission Requirements
Submit:

- Jupyter Notebook (.ipynb)
- Short report (2 pages max)
- Screenshots of results inside notebook

Note: Your submitted code should be neat and clean with proper comments added.
