Smart Reads: *A Hybrid System for Personalized Book Recommendations*

Project Synopsis

Name	Shreeja S Maiya
USN	231VMTR00316
Elective	Artificial Intelligence
Date of Submission	24/02/2025

Executive Summary

The growing availability of books across various platforms has made it increasingly difficult for readers to discover titles that match their preferences. With millions of options, finding the right book becomes a daunting task. The goal of this project is to develop a Hybrid Book Recommendation System that combines both **Collaborative Filtering** (CF) and **Content-Based Filtering** (CBF) to provide personalized book suggestions tailored to individual preferences.

The hybrid system will leverage Collaborative Filtering to analyse user behaviour and Content-Based Filtering to evaluate book metadata, such as title, author, and genre. By merging these two methods, the system will offer more accurate, diverse, and relevant recommendations, addressing the cold-start problem and offering suggestions even for new users with limited ratings.

The system will be implemented using Python libraries such as Pandas, Scikit-learn, and Surprise, with TF-IDF for content analysis. The expected outcome is to improve the book discovery process, offering readers recommendations that are personalized, novel, and diverse.

Objectives and Scope

- Develop a Hybrid Book Recommendation System that integrates Collaborative Filtering and Content-Based Filtering to provide personalized book recommendations.
- Address the **cold-start problem** by combining content-based features (title, author, genre) with user ratings.
- **Enhance accuracy and diversity** in recommendations by comparing the hybrid model with traditional filtering techniques.

Methodology

1. **Data Collection and Preprocessing:**
   - Datasets Used: Books.csv, Users.csv, and Ratings.csv.
   - Data Cleaning: Handling missing values, merging datasets based on ISBN and User ID. Feature Extraction: TF-IDF will be used to convert book titles, authors, and genres into numerical vectors.
2. **Collaborative Filtering (CF):**
   - User-Based CF: Recommends books based on similar users' preferences.
   - Item-Based CF: Suggests books based on item similarity. Cosine Similarity: Measures relationships between users or books.
3. **Content-Based Filtering (CBF):** Uses TF-IDF Vectorization to analyse book descriptions, titles, and genres. *Cosine Similarity* is used to find similar books based on metadata.
4. **Hybrid Model:** Merges CF and CBF predictions using a weighted average method. Balances both user behaviour and book content for precise recommendations.
5. **Recommendation Generation:** Final book recommendations are generated based on the combined results of CF and CBF.

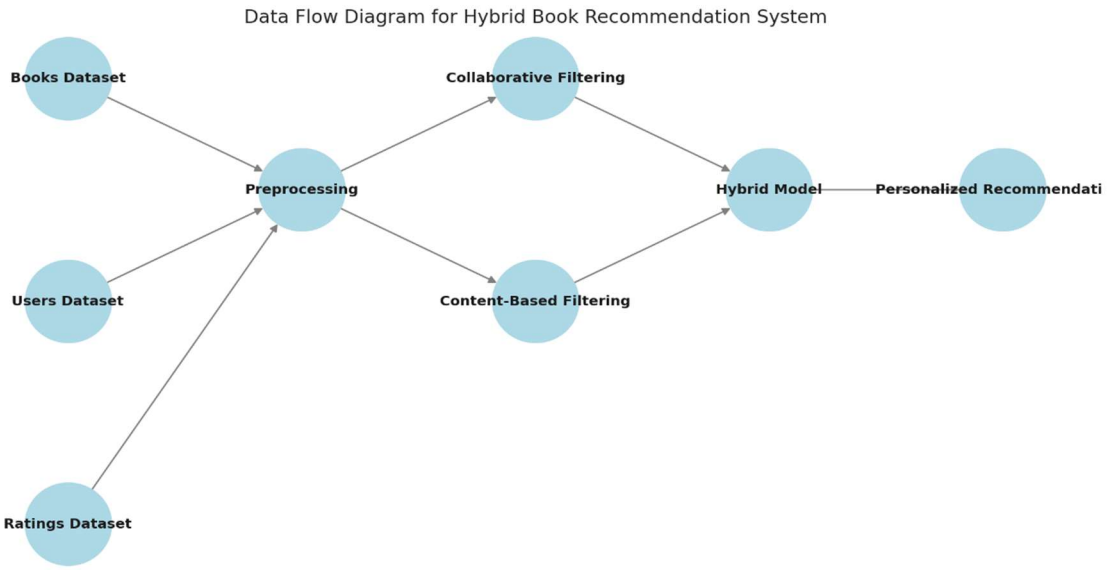

Figure 1 illustrates the data flow process from collection to recommendation generation.

Preliminary Findings & Expected Results

Current recommendation systems often rely solely on Collaborative Filtering or Content-Based Filtering, each with limitations:

- Collaborative Filtering: Faces cold-start issues for new users/books.
- Content-Based Filtering: Provides narrow recommendations, limiting diversity.

Expected Outcomes:

- Higher accuracy in recommendations by combining CF and CBF.
- Overcomes the cold-start problem by integrating book metadata.
- Enhances recommendation diversity, helping users discover new books.

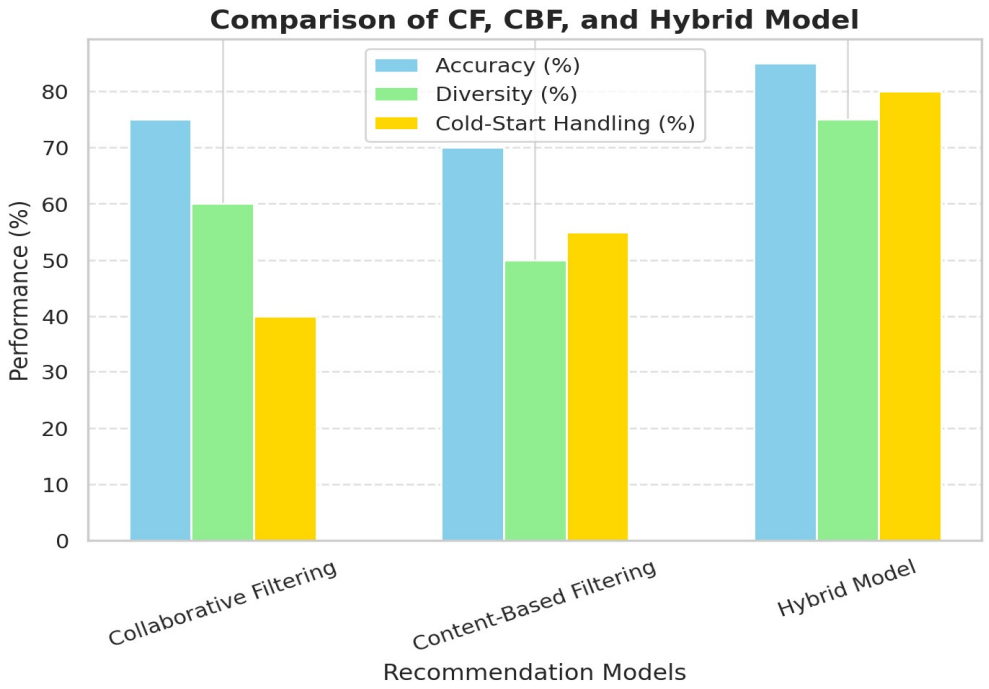

Figure 2 presents a comparison of the accuracy and diversity of CF, CBF, and Hybrid models, demonstrating the improved performance of the Hybrid approach.

Conclusion

The proposed Hybrid Book Recommendation System leverages both Collaborative Filtering and Content-Based Filtering to provide users with accurate, diverse, and personalized book recommendations. By addressing the cold-start problem, the model ensures effective recommendations even for new users or books. The implementation of this system will enhance the book discovery experience, helping users find books tailored to their preferences efficiently.