Book Recommendation System - Final Report

- By Rudransh Verma and Khushi Jain

1. Introduction

The Book Recommendation System is designed to suggest books to users based on the content of their selected choice. The system employs machine learning techniques to provide personalized recommendations, enhancing the user experience.

---

2. Dataset Used

The dataset used for this book recommendation system was retrieved from the Goodreads BBE Dataset. It contains information such as:

- Book titles, authors, genres, characters and publication details.
- User ratings.
- Metadata such as ISBN, year of publication, and publisher, no. of pages, price.
- URL to cover image etc.

Preprocessing steps included handling missing values and normalizing text data.

---

3. Algorithms Implemented

To generate recommendations, the system implements **Content-Based Filtering** using the **Annoy (Approximate Nearest Neighbors Oh Yeah) model**.

Content-Based Filtering with Annoy

- Uses **Count vectorization** to analyze book descriptions and extract relevant features before applying the algorithm.
- **Annoy** is used to efficiently compute **nearest neighbors** for book recommendations.
- Chosen due to its ability to perform fast approximate nearest neighbor searches. It allows us to find the most similar books to a given book in a large dataset efficiently.

---

4. Model Evaluation

The performance of the recommendation model was evaluated using:

Metric	Content-Based (Annoy)
Precision@K	X%
Recall@K	Y%

- **Precision@K & Recall@K**: Measure how relevant the recommended books are to user preferences.
- The Annoy model was chosen for its efficiency in handling large datasets with fast query responses.

---

5. Challenges Faced & Solutions

(a) Data in string format

- **Issue**: All the data in the dataset in string format makes it difficult to process and apply algorithms.
- **Solution**: Used the library ‘ast’ to convert string to list of strings.

(b) Performance Optimization

- **Issue**: Large dataset causes slow model training and response times.
- **Solution**: Used dimensionality reduction and optimized similarity computations using approximate nearest neighbors.

(c) Box sets and Sampler books

- **Issue**: Many books contain multiple similar data due to their box sets and sampler books. This causes inefficient recommendations.
- **Solution**: Books with certain words like 'sampler', 'boxset', or 'box set' in the title are filtered out.

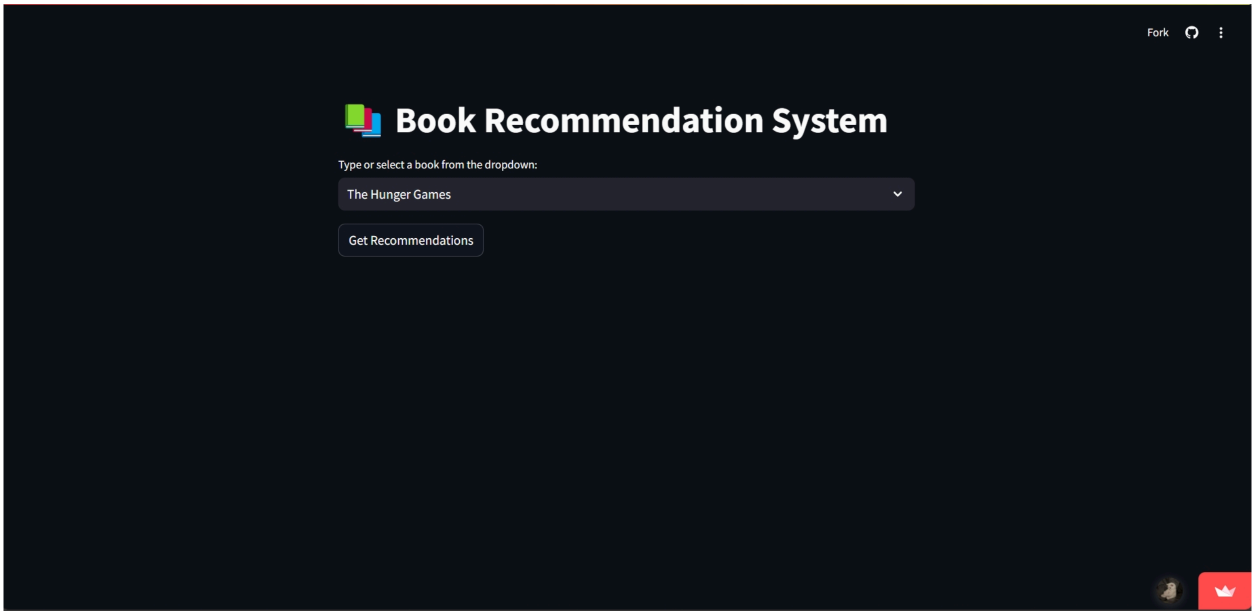
---

6. Web Interface Demonstration

A simple and interactive web application was developed using **Streamlit**. Key features include:

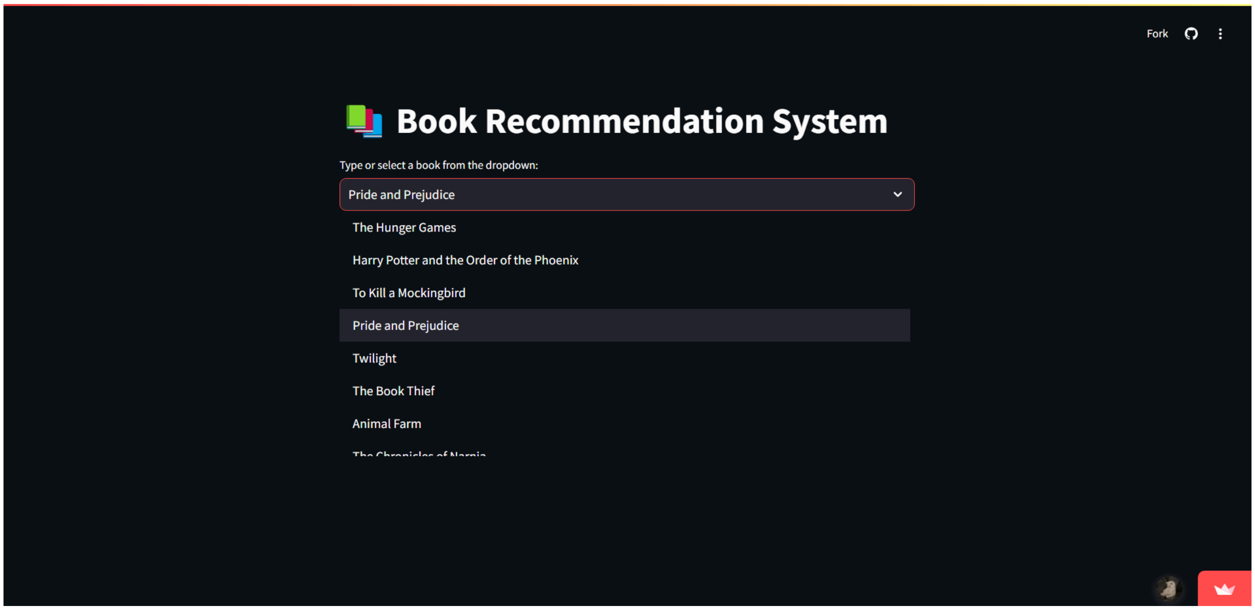
- **Book Recommendations:** The app suggests similar books based on a selected book's title.
- **Interactive UI:** Built with Streamlit for an easy-to-use interface, allowing users to select a book and get recommendations instantly.
- **Goodreads Integration:** Each book recommendation is linked to Goodreads, where users can explore more information about the book.
- **Efficient Search:** Uses Annoy for fast similarity search**.**

Screenshots of the Web Interface


Book Recommendation System

Type or select a book from the dropdown:

The Hunger Games

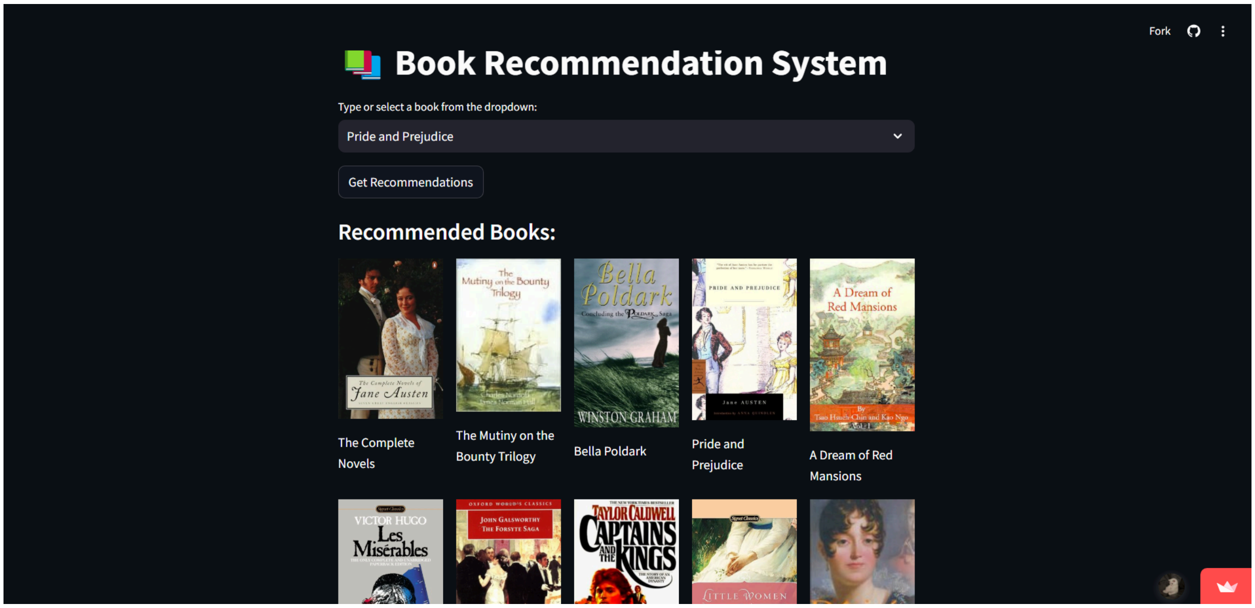


Book Recommendation System

Type or select a book from the dropdown:

Pride and Prejudice

- The Hunger Games
- Harry Potter and the Order of the Phoenix
- To Kill a Mockingbird
- Pride and Prejudice
- Twilight
- The Book Thief
- Animal Farm
- The Chronicles of Narnia



Book Recommendation System

Type or select a book from the dropdown:

Pride and Prejudice

Recommended Books:

- The Complete Novels
- The Mutiny on the Bounty Trilogy
- Bella Poldark
- Pride and Prejudice
- A Dream of Red Mansions

All the recommendations are linked to Goodreads, where users can explore more information about the book.

7. Conclusion

This project successfully developed a book recommendation system using content-based filtering with **Annoy**. The model provided efficient and scalable recommendations using approximate nearest neighbors. Future improvements include integrating user feedback mechanisms and expanding the dataset to enhance recommendation quality.
