



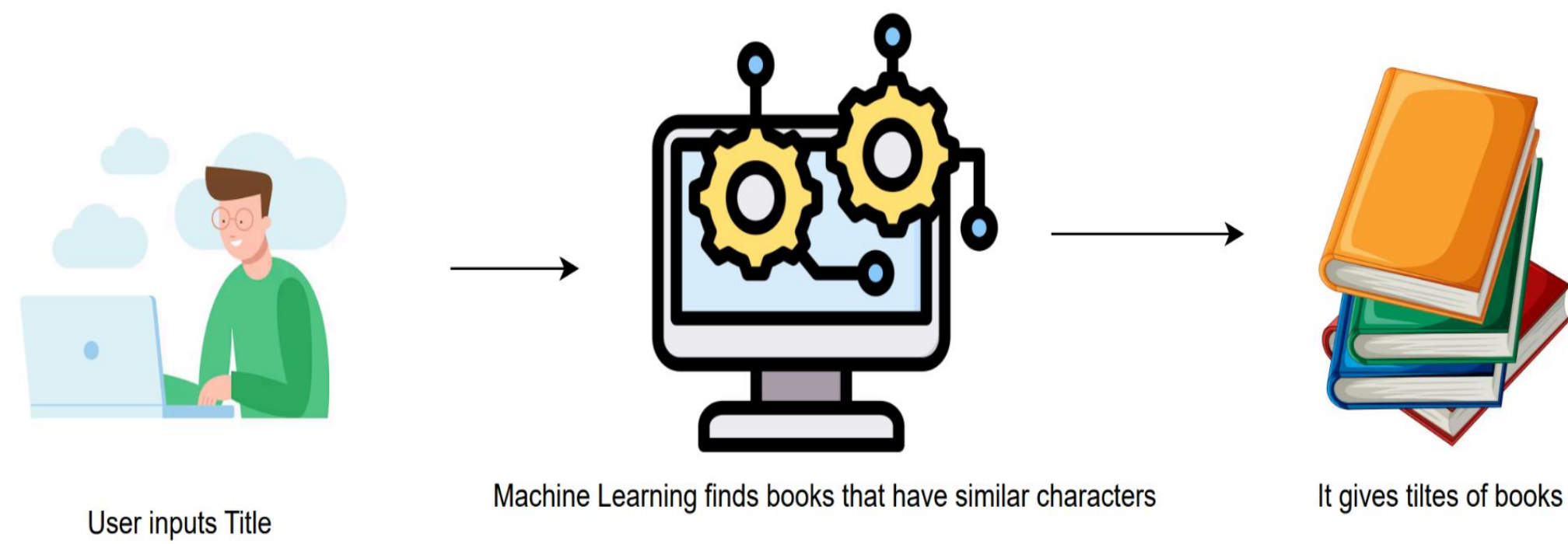
Hybrid Book Recommendation Based on Character

Musa Tora



Graphical Abstract

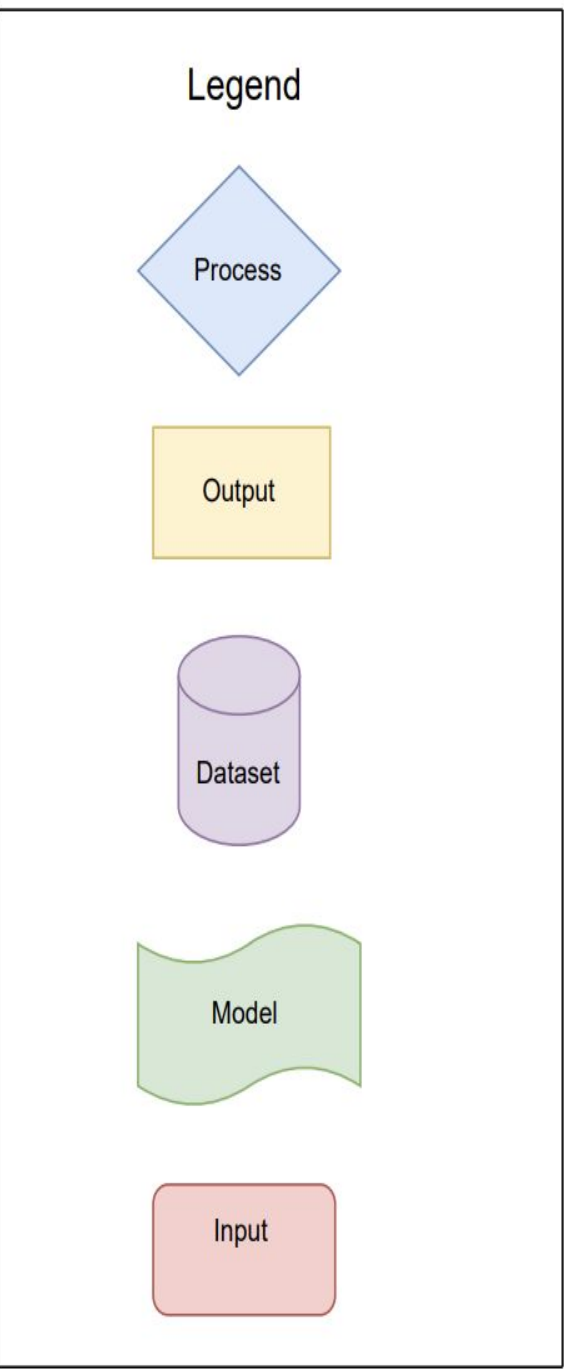
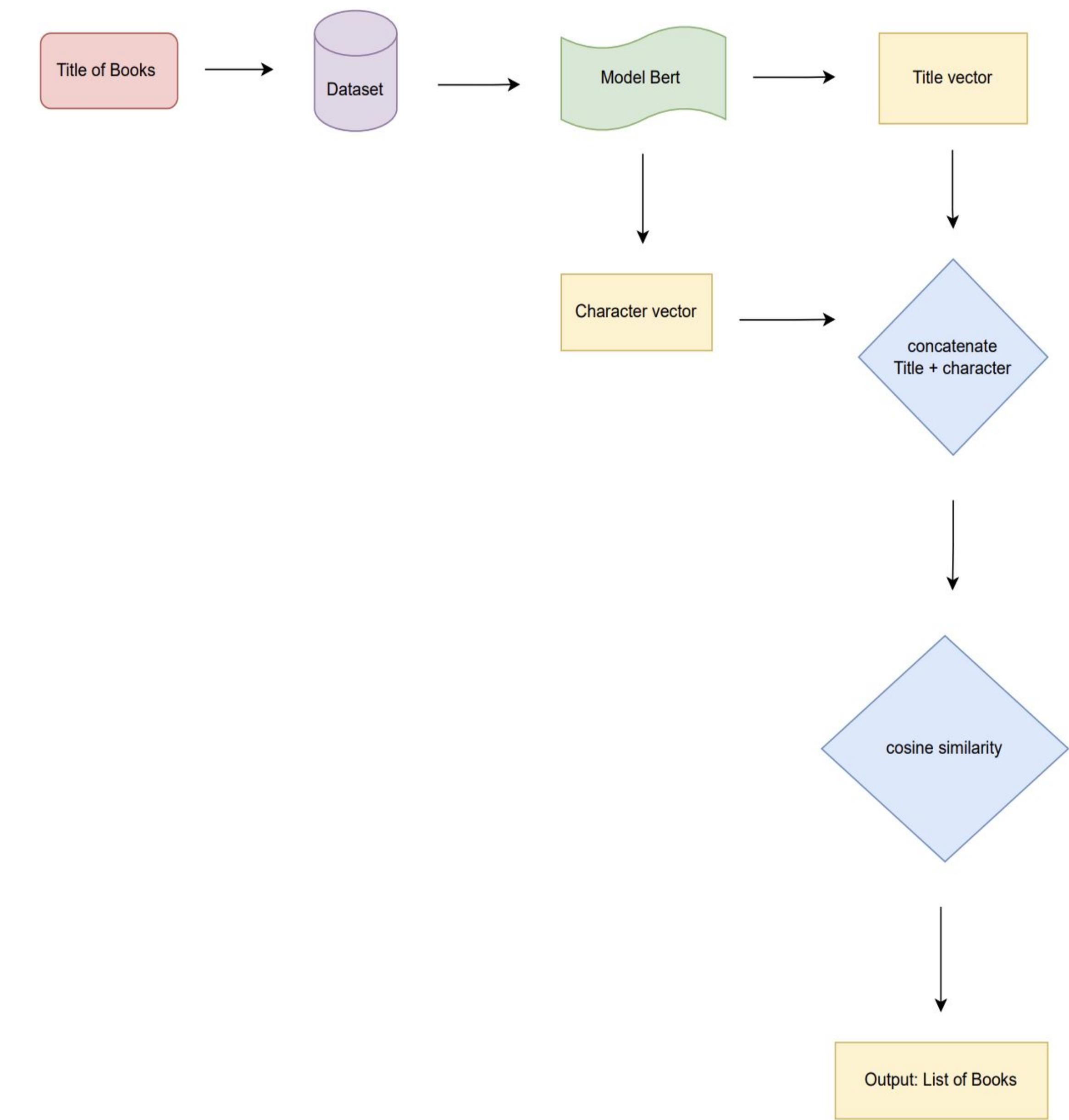
We recommend books by matching similar character profiles across different works. Each book is represented as a single vector that fuses two signals: BERT-based title embeddings and character embeddings from Project Gutenberg, concatenated into a unified semantic representation. We then compute cosine similarity to rank candidates by proximity in this vector space, highlighting books whose character profiles align with a reader's history. This fused, character-aware approach yields more personalized and narratively relevant recommendations than title-only methods.



Introduction

Most recommenders lean on genre or collaborative filtering, which can create echo chambers. This project instead centers on characters their personalities, values, and arcs on surface books that feel emotionally similar even across genres. Each book is represented by a hybrid vector that concatenates BERT title embeddings with precomputed character embeddings, and cosine similarity ranks the closest matches. The result is more personalized, diverse recommendations that preserve emotional continuity while encouraging discovery. If time permits, we'll fine-tune on character descriptions and add theme clustering or a Streamlit UI.

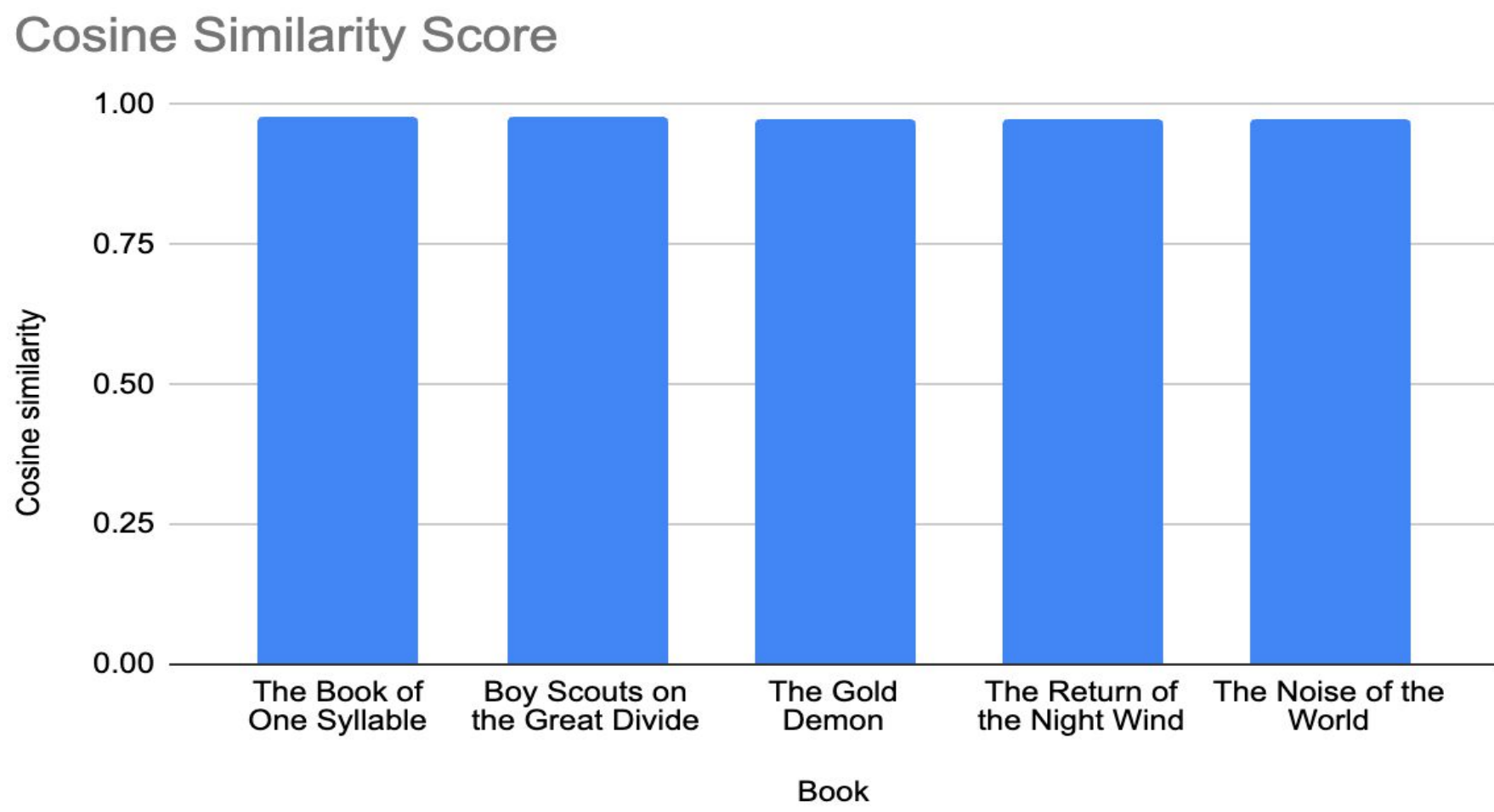
Data Architecture



Methods

This project uses Project Gutenberg (plus Charembench) for book metadata and character-level information (descriptions, relationships). Book titles are encoded with Sentence-BERT to produce contextual embeddings, and character descriptions produce complementary character embeddings capturing narrative traits. The system is built in Python 3 within Jupyter Notebook, using sentence-transformers for SBERT, NumPy/pandas for data handling, and scikit-learn to compute cosine similarity for ranking. For a user-facing experience, the model is deployed as a Streamlit web app where users enter a title and receive dynamic recommendations.

Results



Discussion & Future Work

Most current systems rely on metadata or descriptions and overlook how characters shape reader preferences. Combining character and title embeddings offers a more human-centered way to recommend books. Future work includes fine-tuning BERT on literature-specific data and working on including books after 1960s in the dataset