BOOK RECOMMENDATION SYSTEM

Submitted for  
  
Natural Language Processing – CBCA275 / CBSC360

Submitted by:

(Roll Number) [Your Name]

(Roll Number) [Teammate Name]

Submitted to:  
Dr. Shakshi Sharma

Jan–May 2025

School of Computer Science and Engineering

Bennett University

# INDEX

|  |  |  |
| --- | --- | --- |
| Sr. No. | Content | Page No. |
| 1 | Abstract | 1 |
| 2 | Introduction | 2 |
| 3 | Related Work | 3 |
| 4 | Methodology | 4 |
| 5 | Hardware/Software Required | 5 |
| 6 | Experimental Results | 6 |
| 7 | Conclusions | 7 |
| 8 | Future Scope | 8 |
| 9 | GitHub Link of the Complete Project | 9 |

# Abstract

Recommendation systems are essential in helping users navigate vast content libraries. This project presents a Content-Based Book Recommendation System using Natural Language Processing (NLP) techniques. We employ a combination of TF-IDF vectorization, Sentiment Analysis, and Named Entity Recognition (NER) to deeply understand the textual content of book descriptions. By calculating Cosine Similarity between TF-IDF vectors, we generate recommendations based on content relevance. Sentiment scores help gauge the emotional tone of descriptions, while NER identifies key people, places, and organizations mentioned in the text. The system runs through a Streamlit interface, enabling user-friendly interaction and live book suggestions.

# Introduction

Book recommendation systems help readers discover new titles based on interests, genres, and themes. Unlike traditional collaborative filtering, this project takes a content-based approach, focusing on what a book is about, rather than how others rated it. By analyzing the description of the books, we determine their textual similarity. We further enhance recommendation quality by integrating Sentiment Analysis to detect tone and Named Entity Recognition to extract meaningful entities. This multi-technique approach ensures deeper text understanding and smarter recommendations.

# Related Work

Earlier systems relied heavily on user data such as ratings or purchase history. Examples include:  
- Amazon & Goodreads: Use hybrid systems combining collaborative and content filtering.  
- Google Books: Focuses more on metadata and keyword search.  
- Research Papers: Some use advanced models like BERT for semantic similarity, or VADER for sentiment scoring.  
  
Our approach builds upon these ideas by using TF-IDF, Sentiment Analysis, NER, and Cosine Similarity, all lightweight and explainable techniques suitable for scalable deployment.

# Methodology

Dataset:  
The Goodreads dataset is used, which includes:  
- Title, Author  
- Description/Summary  
- Genres, Ratings  
  
Steps:  
1. Data Preprocessing:  
 - Lowercasing, punctuation removal  
 - Stopword removal using NLTK  
 - Handling null values and encoding corrections  
  
2. TF-IDF Vectorization:  
 - Used to transform book descriptions into weighted numerical vectors.  
  
3. Cosine Similarity:  
 - Measures similarity between TF-IDF vectors.  
  
4. Sentiment Analysis:  
 - Applied using TextBlob to determine polarity.  
  
5. Named Entity Recognition (NER):  
 - Using spaCy to extract entities for context.  
  
6. Recommendation Logic:  
 - Fetch top N similar books based on similarity scores.  
  
7. Streamlit Interface:  
 - UI where users enter a book title and get recommendations.

# Hardware/Software Required

Hardware:  
- Minimum 4GB RAM (8GB recommended)  
- Internet access  
  
Software:  
- Python 3.x  
- Libraries: pandas, numpy, sklearn, nltk, textblob, spacy, streamlit  
- Platform: Google Colab / Jupyter / VS Code  
- Hosting: Ngrok

# Experimental Results

Sample Test 1  
Input Book: To Kill a Mockingbird  
- Sentiment: Positive  
- Entities: “Maycomb”, “Atticus Finch”, “Tom Robinson”  
- Recommended Books:  
 - The Catcher in the Rye  
 - Of Mice and Men  
 - Go Set a Watchman  
  
Sample Test 2  
Input Book: Pride and Prejudice  
- Sentiment: Neutral to Positive  
- Entities: “Elizabeth Bennet”, “Mr. Darcy”, “Netherfield”  
- Recommended Books:  
 - Emma  
 - Sense and Sensibility  
 - Northanger Abbey

# Conclusions

This project successfully demonstrates how combining multiple NLP techniques can enhance recommendation quality. By using TF-IDF, Cosine Similarity, Sentiment Analysis, and Named Entity Recognition, the system understands both the tone and context of books. This leads to smarter and more relevant recommendations for users.

# Future Scope

- Implement genre-based filtering  
- Use transformer models like BERT  
- Add user login for personalization  
- Integrate live APIs from Goodreads or Google Books  
- Build a full-fledged app for public use

# GitHub Link of the Complete Project

[Paste your GitHub repo link here – make sure it includes: code, dataset, README, and your final presentation]