BOOK RECOMMENDATION

NEW WAY TO FIND LIFE

INTRODUCTION

 In this project, we developed a Book Recommendation System that suggests books based on the content of their descriptions. Rather than using user ratings or behavior, the system analyzes and compares book summaries to recommend similar titles.

PROCESS

Data Collection & Cleaning
- Dataset included: Titles, Authors, Descriptions
- Removed null entries and duplicates

🧼 NLP Preprocessing
- Lowercasing, punctuation & stopword removal
- Tokenization and lemmatization
- Vectorized using TF-IDF

📐 Similarity Computation
- Calculated Cosine Similarity between book vectors
- Returned top N similar books for any selected input

💻 Output
- Python-based terminal interface for testing
- Takes input book title and prints top recommendations

NLP TECHNIQUES

TF-IDF Vectorization

Used TfidfVectorizer to transform cleaned text into numerical features. TF-IDF gives weight to important words while reducing the impact of common words.

Cosine Similarity

Computes the similarity between TF-IDF vectors of books. Helps find books with similar themes or genres.

Sentiment ANALYSIS

Used TextBlob to assign a sentiment label to each book description. Polarity-based classification: Positive, Negative, or Neutral.

Named Entity Recognition (NER)

Used spaCy to extract named entities like names, locations, and dates from descriptions. Useful to identify key themes or figures in books. make it in one slide for ppt

PRODUCT #1

1. Purely Content-Based without Ratings
   - Unlike typical systems that rely on user feedback or ratings, this model recommends books solely based on content — making it ideal for new or unrated books.
2. TF-IDF on Book Descriptions
   - Utilizes TF-IDF vectorization to capture the essence of each book's description, allowing the system to understand content similarity at a meaningful level.
3. Cosine Similarity for Contextual Matching
   - Rather than matching keywords, it mathematically computes how similar books are based on description vectors, ensuring more accurate and relevant suggestions.
4. Scalable and Explainable
   - Easily scales to thousands of books, and each recommendation can be traced back to shared content themes, offering transparency over black-box algorithms.
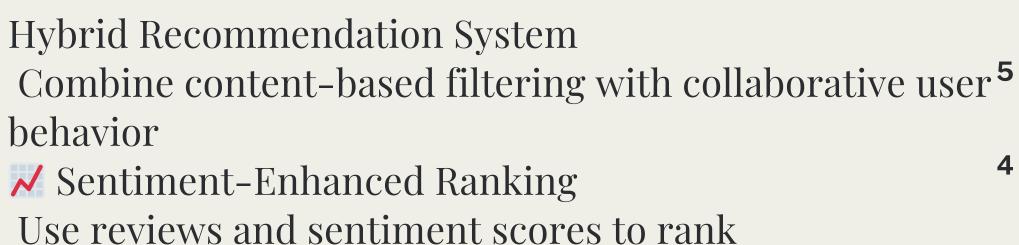5. Minimal Dependency on External Data
   - The model doesn't require user behavior, reviews, or external metadata — just the description of the book is enough to provide meaningful results.

CHALLENGES & SOLUTIONS

⚠ Inconsistent Descriptions

↠ Applied strict preprocessing to normalize varied writing styles

⚠ Empty or Duplicate Entries

↠ Cleaned dataset and dropped null values

⚠ Cold Start (No Ratings)

↠ Solved using content-based filtering (description only)

⚠ Too Generic Recommendations

↠ Enhanced model using fine-tuned TF-IDF parameters (e.g., max_df, min_df)

FUTURE SCOPE

Hybrid Recommendation System
 Combine content-based filtering with collaborative user behavior
📈 Sentiment–Enhanced Ranking
 Use reviews and sentiment scores to rank recommendations more accurately
🌐 Web App Deployment
 Use Streamlit or Flask to build an interactive user interface
📚 Genre & Author Filtering
 Allow users to filter recommendations by genre, length, or writing style
📱 Mobile Application Integration
Deploy as a mobile app for quick book discovery

Thank you!

SOUMYA KOTHARI

E23BCAU0041