

Book Recommendation System Using Big Data

Date: 06/02/2025

Authors:

Liav Cohen 209454693

Noam Tsfati 314952912

Alon Gottlib 208290684

Problem Description

Developing a book recommendation system using collaborative and content-based filtering. Utilizing Big Data technologies to provide personalized recommendations in near real-time.

Project and Big Data Relevance

- The project leverages Big Data techniques to efficiently process and analyze large-scale user interactions and book metadata.
- Technologies like Apache Kafka, PySpark, and PostgreSQL enable scalable recommendation models.

Dataset:
Goodbooks-
10k

- 10,000 books and 6 million ratings
- Metadata: genres, authors, titles
- Stored in PostgreSQL for structured querying

Methodology: Alternating Least Squares (ALS)

- Matrix factorization for collaborative filtering
- Used by Netflix and Amazon
- Identifies latent factors in user-book interactions
- Optimized for large datasets

Content-Based Filtering

- Uses metadata (genres, authors) for recommendations
- TF-IDF and cosine similarity to find similar books
- Hybrid approach:
Combines with ALS

Evaluation Metrics

- Precision: Measures relevance of recommendations
- Recall: Measures coverage of relevant recommendations
- RMSE: Measures prediction accuracy

Explainability: Why Was This Recommendation Made?

- Recommendations explained by analyzing similar books
- Uses cosine similarity of genre vectors
- Displays top 3 similar books with similarity scores
- Helps users understand and trust recommendations

Academic References & Insights

- Big Data Classification Challenges (Suthaharan) - Discusses scalability and data handling challenges
- Predicting Amazon Ratings (Woo & Mishra) - Machine learning models for rating prediction and recommendations
- 3. Amazon Recommendation System - Collaborative filtering and NLP for product recommendations

Results and Conclusion

- Expected improvements in recommendation accuracy
- Evaluated against traditional static models
- Future Work: Enhancing explainability and fairness