National University of Sciences and Technology (NUST)
School of Electrical Engineering and Computer Science

Smart Book Recommendation System

Book recommendation system by using data big data mining techniques

Instructor: Dr. Muhammad Daud Abdullah Asif

Submitted By:

SHEEZA ASLAM

374382

1) Introduction

Project objective

The main objective of this project is to create a book recommendation system using data mining techniques. The system aims to recommend books to users based on their previous reviews, identifying patterns and relationships in the data to provide personalized recommendations.

Importance of book recommendations

Book recommendations are important in online and digital book retail libraries because they help users discover new books, improve user engagement and can lead to increased revenue or borrowing rate. Personalized recommendations improve user experience by recommending books that match the user's interests and reading habits.

2) Project Scope and Objectives

Project scope

This project focuses on creating a recommendation system using user review data. The scope includes data loading cleaning frequent pattern mining association rules and KNN to find similar books.

Objective

The objective of the project is

- To identify the set of items that are frequent in a user's book reviews.
- To create association rules that represent relationships between books.
- Use these rules to create recommended book recommendations personalization.
- Apply KNN to search for books with similar characteristics.

3) Implementation Details:

To implement personalized book recommendations the system uses the following steps:

Data Collection: Gather user ratings and other relevant data' such as book genres and publication years frequent pattern mining use FP growth to identify frequent itemsets representing sets of books that are commonly rated together.
Association Rules: Generate association rules from frequent itemsets to discover relationships between books.
Personalized Recommendations: Use the association rules to suggest books to users based on their previous ratings and the patterns identified in the data.

4) Load and Clean Data

Data source

Data for this project comes from Excel files, including book information, user reviews and users. Book data contains details such as ISBN and book title, while rating data includes user IDs and book ratings.

Data cleaning

This includes removing duplicates, filling in values missing and ensure consistency of data types. These steps are essential to ensure data is clean and ready for analysis.

Data reduction

To reduce memory usage and processing time, the project focuses on users with minimum transaction volume and limit analysis to the highest-rated books. This approach allows you to focus on active users and relevant books.

5) Frequent Pattern Mining and Association Rules

FP-Growth algorithm

FP-Growth algorithm is used to find frequent sets in data. Frequent item sets represent sets of books that are often ranked together by users. The “support” metric indicates the proportion of a user’s transactions that contain a particular group of items.

Association Rules

These rules are derived from frequent groups of items. They describe relationships between groups of items, indicating which books can be ranked together. The main metrics of an association rule include “confidence” (the probability of an outcome given the premises) and “gain” (the degree of association where a value greater than 1 indicates a relationship) powerful system.

Top Recommendations

By sorting Using link rules based on “lift”, you can make top recommendations, indicating which books closely related and therefore likely to be recommended.

6) Data visualization

Purpose of data visualization

Data visualization helps understand frequent sets of items and association rules by providing a visual representation of patterns and relationships system in data. This can make the results more intuitive and easier to interpret.

Visualization types

Bar charts: These are used to visualize support for frequent sets of items, shows which book series are most popular among users.
Scatter plot: It is used to visualize association rules plots confidence versus gain to show the robustness and reliability of the rules.
br>Interpret Visualizations: These visualizations help identify meaningful patterns and relationships in the data. For example, a bar chart can show which book series are frequently ranked together, while a scatter chart can show which association rules are strong and reliable.

7) Export results for Further Analysis

Export data: Export results to an Excel or CSV file allowing for further analysis or integration with other systems. This step is important for sharing results and continuing analysis in different environments.

8) KNN for Nearest Book Recommendations

Purpose of KNN: k-nearest neighbor (KNN) is used to find books with similar features. This approach can recommend similar books based on a specific feature vector such as rating or book genre.

KNN application: KNN is applied to find the closest books based on certain feature vectors. This involves adapting the KNN model to the book's characteristics and querying the nearest neighboring books.

Result: The KNN output shows the index of the closest books most suggest books that share similar characteristics and can be recommended together.

9)Making the report on book recommendations

In the end, we produced an excel report called book_recommendation.xlsx that included details on frequently occurring itemsets and the values that support them. There are two main columns in this report:

Support: Shows how frequently a certain itemset appears in the dataset. A higher support value indicates greater commonality of the itemset, indicating popular book pairings.

Itemsets: Books that commonly appear together are referred to as itemsets. Book recommendations can be made using the information in this column, which is used to identify common book categories. Python and the pandas library-more especially, the ExcelWriter function for exporting data to an Excel format-were used to construct the report to learn more about correlations and conduct additional analysis on the most common itemsets, utilize the book_recommendation.xlsx file.

9) Use Case

Personalized Book Recommendations

The use case revolves around creating a personalized book recommendation system. The primary objective is to provide users with book suggestions that align with their interests and preferences. thereby enhancing the user experience and driving engagement.

Description

Personalized book, recommendations leverage user data such as previous ratings, to identify patterns and relationships among books. By applying data mining techniques like frequent pattern mining and association rules, the system can suggest books that users are likely to enjoy based on their reading habits.

Business Goals

The business goals for this use case include:

Improving customer satisfaction: By offering personalized book recommendations, users are more likely to find books they enjoy, leading to greater satisfaction and positive user experiences.
Increasing User Engagement: Personalized recommendations encourage users to interact with the platform more frequently, exploring new books and genres.
Boosting Sales or Borrowing Rates: For online book retailers personalized recommendations can lead to increased sales. For digital libraries, they can result in higher borrowing rates.

10) Stakeholders:

Several stakeholders are involved in this use case, each playing a specific role:

Users: The End-users who receive book recommendations. Their feedback and engagement drive the success of the system.
Data Analysts: Responsible for analyzing user data and identifying patterns to improve recommendations.
System Administrators: Ensure the recommendation system runs smoothly and securely.
Marketing Teams: Use personalized recommendations to target users with promotional campaigns.

11) Benefits

Personalized book recommendations offer numerous benefits including:

Enhanced User Experience: By suggesting books tailored to users preferences, the system provides a more personalized and enjoyable experience.
Increased Customer Retention: When users receive recommendations that resonate with them they re more likely to return to the platform, leading to higher retention rates.
Boosted Sales or Borrowing: For online book retailers, personalized recommendations can drive sales by suggesting books that users are likely to purchase. In digital libraries, they can increase borrowing rates by encouraging users to explore new books.
Support for Marketing Campaigns: Personalized recommendations can be used by marketing teams to create targeted campaigns, promoting specific books or genres based on user preferences.

12) Challenges and Considerations

While personalized book recommendations offer many benefits, there are challenges to consider

Data Privacy: Handling user data requires careful attention to privacy and security.
Algorithm Complexity: Frequent pattern mining and association rules can be computationally intensive, requiring efficient algorithms and data reduction techniques.
Diverse User Preferences: Users have varying tastes and preferences, making it important to create a flexible recommendation system that adapts to different needs.

13) Conclusion

Personalized book recommendations are a valuable use case for online book retailers and digital libraries. By leveraging data mining techniques and association rules, the system can provide tailored suggestions that enhance the user experience, increase engagement, and drive sales or borrowing rates. The success of this use case relies on careful data analysis, a robust recommendation algorithm, and a user-centric approach to delivering personalized content

14) OUTPUT

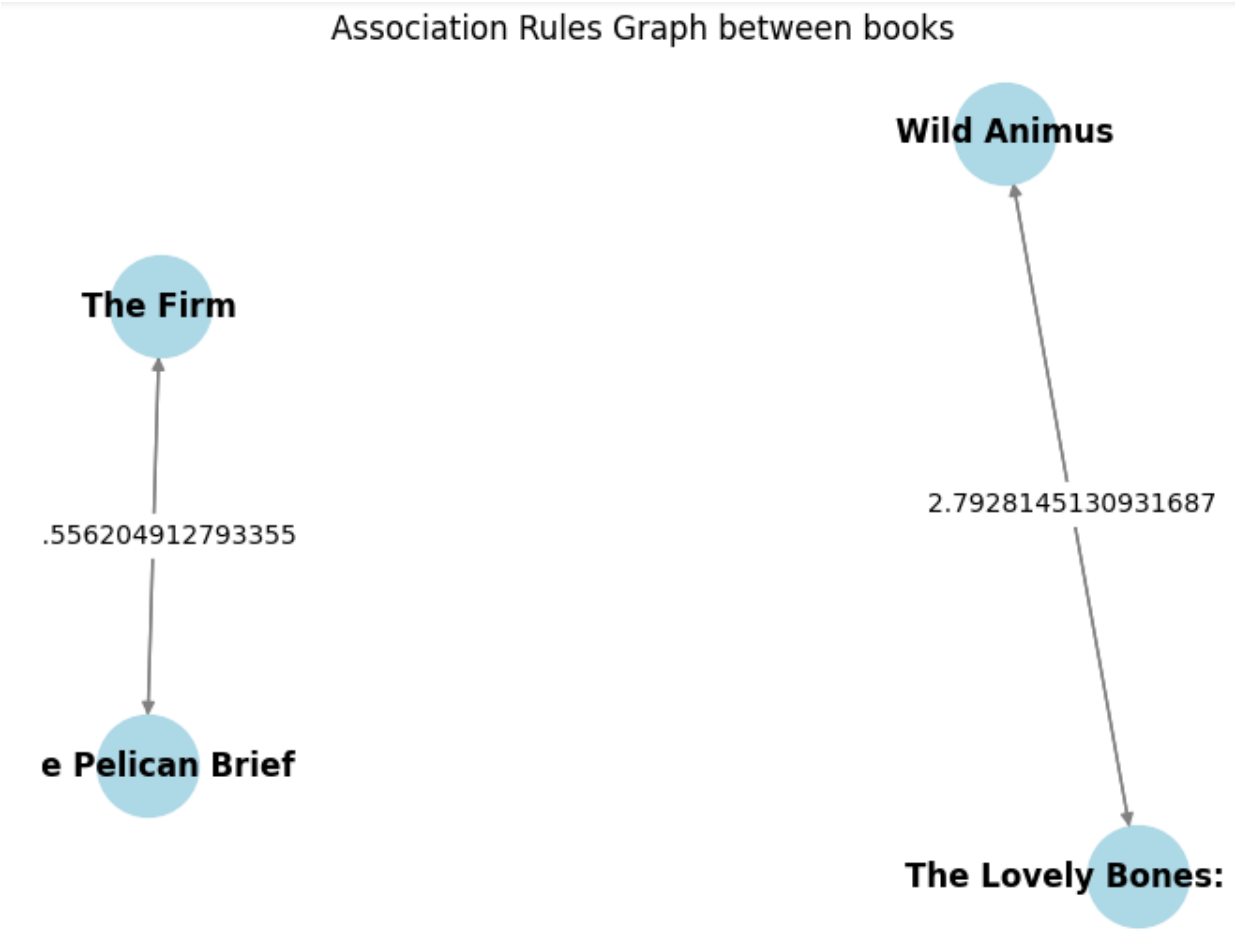
1. Top 50 Books by Average Rating

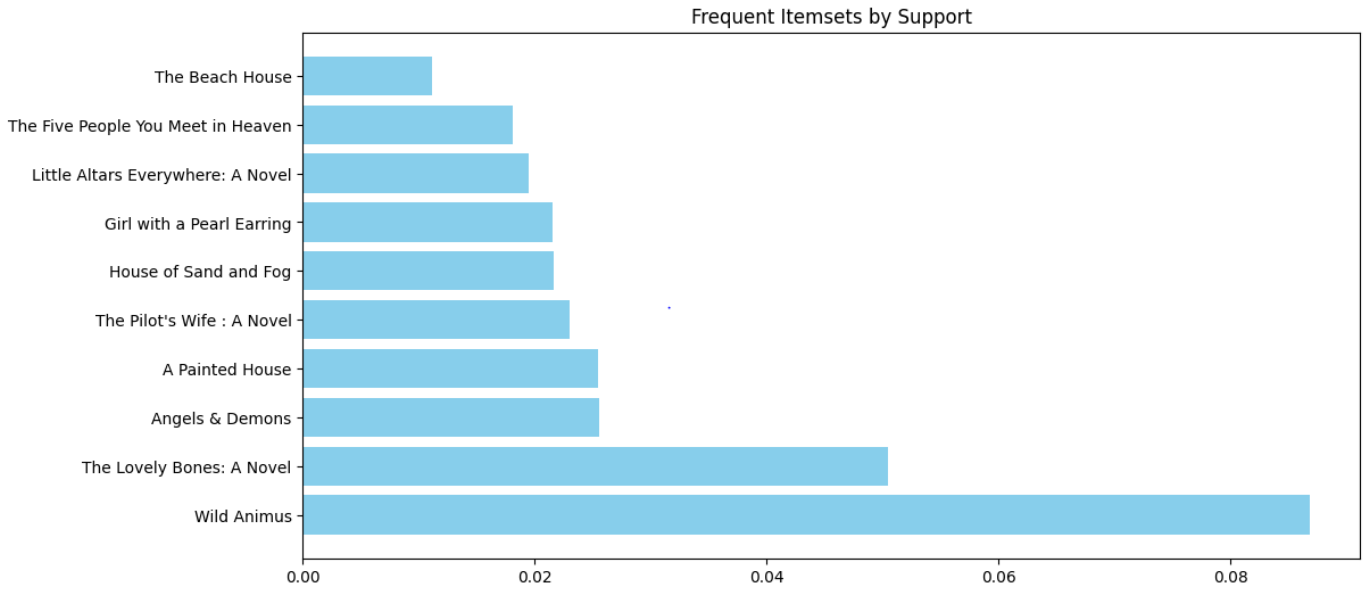
	ISBN	Book-Title	Book-Rating
4	000225056X	The International Garlic Cookbook	10.0
5	000255755X	Come Through Marjorie's Garden Gate: Spend a Y...	10.0
6	000599134X	Collins Bible Handbook	10.0
9	000639194X	Firewing	10.0
11	000648185X	Escape from the Big Muddy	10.0
12	000673765X	Apple Bough	10.0
13	000692347X	The Clue of the Broken Locket (The Nancy Drew ...	10.0
14	000710331X	Tarot	10.0
15	000710698X	Pagan Pathways, New Edition	10.0
16	000711365X	Central Fells: Lakeland Fellranger (Lakeland F...	10.0
18	000715111X	The Real Witches Book of Spells and Rituals (R...	10.0
20	002011690X	Complete How-To Book of Indiancraft: 68 Projec...	10.0
21	002028361X	Deathbird Stories (Collier nucleus fantasy &am...	10.0
22	002863196X	Complete Idiot's Guide to Getting Your Romance...	10.0
23	002934770X	Dearest Friend	10.0
24	003008539X	Three With a Bullet	10.0
25	006009298X	Temptation of a Proper Governess	10.0
26	006009608X	Princess in Waiting (The Princess Diaries, Vol...	10.0
27	006013738X	The survival of the wisest	10.0
28	006015425X	Illywhacker	10.0
30	006024416X	Betsy-Tacy and Tib	10.0
31	006026425X	Farmer Boy (Little House)	10.0
32	006039174X	Every Knee Shall Bow : The Truth & Tragedy...	10.0
34	006059330X	A Lady of His Own	10.0
35	006061272X	My angel's name is Fred: Tales of growing up C...	10.0
37	006091548X	Predictive astrology: Understanding transits a...	10.0
39	006093624X	The Adventures of Flash Jackson : A Novel	10.0
40	006095177X	The Holistic Pediatrician: A Parent's Comprehe...	10.0
41	006095616X	The Dance of Connection: How to Talk to Someon...	10.0
42	006098712X	Days of Our Lives: The Complete Family Album	10.0
44	006099701X	The Book of Laughter and Forgetting	10.0
45	006100328X	Doorways in the Sand	10.0
46	006105531X	A Diversity of Dragons (Pern)	10.0
47	006105965X	Hammer Down (Nascar, No. 5)	10.0
48	006250973X	Jesus: The Evidence	10.0
49	006270107X	The Dictionary of American Slang	10.0

2. Visualize Association Rules with NetworkX

```
Association Rules:
                                antecedents  ... zhangs_metric
0                               [Wild Animus]  ...      0.703008
1  [The Lovely Bones: A Novel]  ...      0.676091
2                      [The Pelican Brief]  ...      0.966429
3                               [The Firm]  ...      0.966226

[4 rows x 10 columns]
Top Recommendations:
                                antecedents  ... zhangs_metric
2                      [The Pelican Brief]  ...      0.966429
3                               [The Firm]  ...      0.966226
0                               [Wild Animus]  ...      0.703008
1  [The Lovely Bones: A Novel]  ...      0.676091
```




Frequent Itemsets by Support
The Beach House
The Five People You Meet in Heaven
Little Altars Everywhere: A Novel
Girl with a Pearl Earring
House of Sand and Fog
The Pilot's Wife : A Novel
A Painted House
Angels & Demons
The Lovely Bones: A Novel
Wild Animus
0.00
0.02
0.04
0.06
0.08

3. Apply KNN for Nearest Book Recommendations

```
book_features = np.array([
    [1, 0, 1],
    [0, 1, 1],
    [1, 1, 0]
])

knn = NearestNeighbors(n_neighbors=2)
knn.fit(book_features)

query = np.array([[1, 1, 0]])
distances, indices = knn.kneighbors(query)

print("Nearest books based on KNN:")
print(indices)
```

```
Nearest books based on KNN:
[[2 1]]
```

4. Export Results for Further Analysis

book_recommendations.xlsx

	A	B	C
1	**support**	**Itemsets**	
2	0.01807	['The Five People You Meet in Heaven']	
3	0.02554	['Angels & Demons']	
4	0.01112	['The Beach House']	
5	0.01952	['Little Altars Everywhere: A Novel']	
6	0.08687	['Wild Animus']	
7	0.05051	['The Lovely Bones: A Novel']	
8	0.02544	['A Painted House']	
9	0.02297	["The Pilot's Wife : A Novel"]	
10	0.02168	['House of Sand and Fog']	
11	0.02158	['Girl with a Pearl Earring']	
12	0.02024	['Summer Sisters']	
13	0.01854	['The Testament']	
14	0.01668	['The Street Lawyer']	
15	0.01586	['The Partner']	
16	0.01524	['Me Talk Pretty One Day']	
17	0.01395	['Lucky : A Memoir']	
18	0.01344	['A Map of the World']	
19	0.01313	['Congo']	
20	0.01133	['Deception Point']	
21	0.01092	['2nd Chance']	
22	0.01045	['Roses Are Red (Alex Cross Novels)']	
23	0.0104	['The Bean Trees']	
24	0.01432	['Wicked: The Life and Times of the Wicked Witch of the West']	