Question A: Recommender Systems

1. Different Types of Recommender Systems

- **Content-Based Filtering**: Content-based filtering focuses on recommending items like those the user has already shown interest in. This method uses item features, such as book genres, authors, keywords, or other characteristics, to match items with users' past interactions (GeeksforGeeks, 2024). For instance, if a user has shown interest in historical fiction, the recommender system would suggest similar books in the same genre.
- **Collaborative Filtering**: This type of system relies on the collective preferences and behaviors of users. It makes recommendations based on the premise that users with similar preferences will enjoy similar items (GeeksforGeeks, 2024). For example, if several users with similar reading habits have all highly rated a particular book, that book might be recommended to others with similar preferences.
- **Recommendation for Barnes & Noble**: For a company like Barnes & Noble with a diverse customer base and a vast catalog, a **hybrid approach** is ideal. A hybrid system can utilize both content-based and collaborative filtering methods, improving accuracy by considering both item characteristics and user preferences (Michael, 2023). This approach would allow Barnes & Noble to recommend books that align with individual users past interactions while also factoring in trends among similar type of users.

2. Data Mining Algorithms for Recommender Systems

- **K-Nearest Neighbors (K-NN)**: This algorithm identifies and recommends items preferred by users who are "nearest neighbours" to the target user in terms of taste or browsing history (GeeksforGeeks, 2021). It is commonly used in collaborative filtering and could identify customers with similar reading habits, characteristics helping to tailor personalized book recommendations.
- **Matrix Factorization Techniques (e.g., Singular Value Decomposition - SVD)**: SVD and other matrix factorization methods break down large datasets to identify patterns, even in sparse data (when users have rated only a few items). SVD helps in predicting users'

preferences by learning underlying patterns in their interactions with items, making it a powerful tool for creating recommendations (GeeksforGeeks, 2021).

- **Decision Trees and Random Forests**: Decision trees can be used to analyze customer characteristics and behaviors, while random forests (ensemble methods) offer robust predictions based on these patterns (GeeksforGeeks, 2021). For Barnes & Noble, these models could help predict specific book preferences based on individual customer data.

3. Required Data for the Recommender System

- **User Data**: Information about users, such as demographic details, reading history, previous purchases, and any ratings they have given to books. This data helps understand each customer's preferences (Rina Caballar, 2024).
- **Item Data**: Details about each book, including title, author, genre, keywords, and user ratings. This metadata is essential for content-based filtering and provides context to match books with user preferences (Rina Caballar, 2024).
- **User-Item Interaction Data**: Data on specific interactions between users and items, including clicks, views, ratings, and purchases. This information helps collaborative filtering algorithms understand user preferences based on past behavior (Rina Caballar, 2024).
- **Data Volume**: An effective recommender system generally requires a large dataset. Ideally, data from hundreds of thousands or even millions of users would be available, with each user having multiple interactions recorded. This allows the system to identify patterns and make precise recommendations (Rina Caballar, 2024).

4. Measuring Success of the Recommender System

- **Click-Through Rate (CTR)**: CTR indicates how often users click on recommended items. A higher CTR suggests that the recommendations are relevant and engaging (Rina Caballar, 2024).
- **Conversion Rate**: Measures the percentage of recommendations that result in a purchase, providing insight into the system's effectiveness in driving sales (Rina Caballar, 2024).

- **User Engagement Metrics**: These metrics track user interactions with recommended items, including the amount of time spent browsing suggested books. High engagement signals that the recommendations are keeping users interested (Rina Caballar, 2024).
- **User Satisfaction Surveys**: Direct feedback from users provides a qualitative measure of the recommender system's success, showing if users find the recommendations useful (Rina Caballar, 2024).
- **Diversity and Novelty**: These metrics assess the range of items recommended and the system's ability to introduce new, relevant items that users might not otherwise have found (Rina Caballar, 2024).

Question B: Chestnut Ridge Laptop – Determining the Ideal Product

1. Baseline Laptop Configuration

The baseline laptop was identified by filtering the dataset to find the configuration where all dummy-coded attributes are set to zero. The baseline configuration is as follows:

- **RAM**: 4GB
- **Screen Size**: 13 inches
- **Battery Life**: 6 hours
- **Price**: $800
- **Hard Drive**: SATA

This serves as a reference for comparison with other configurations.

2. Data Cleaning Process

The data was checked for missing values and correct data types for each column. There were no missing values, and all columns were in the expected data types (integer type except for the dtype column, which is of object type), indicating the dataset was ready for analysis.

3. Average Partworths for Each Attribute Level

Average partworths were calculated across all 20 participants. Here are the results:

- **8GB RAM**: Increased ratings on average compared to the baseline 4GB.
- **16GB RAM**: Highest average rating among RAM configurations.
- **Screen Sizes (14-inch and 15-inch)**: The 14-inch screen size has a higher average rating than the 15-inch, suggesting it may be more preferred.
- **Battery Life (8-hour and 10-hour)**: The 10-hour battery life received a higher average rating.
- **Prices ($850 and $900)**: Lower prices tend to increase ratings, with $800 as the baseline.
- **Hard Drive (SSD)**: The SSD configuration showed a higher rating than SATA.

These partworths represent the perceived utility of each feature level, guiding which configurations are more appealing.

4. Most Important Attribute

Price was identified as the most influential attribute, with the highest range in partworths. This was decided by calculating the range of average ratings across levels for each attribute, and price showed the largest range, indicating its significant impact on participant preferences.

5. Optimal Laptop Specification Based on Average Partworths

To maximize consumer rating, the ideal laptop configuration, based on the highest average partworths, includes:

- **RAM**: 8GB
- **Screen Size**: 15 inches
- **Battery Life**: 10 hours
- **Price**: $850
- **Hard Drive**: HDD

This configuration balances consumer preference in key areas, bearing the highest expected ratings.

6. Dollar Value of Specific Attribute Levels

The dollar value for each specified attribute level was derived by comparing average ratings and translating differences into dollar amounts. Here are the results:

- **16GB RAM**: $20.67
- **15-inch Screen Size**: -$18.91
- **SSD Hard Drive**: $11.78
- **10-hour Battery Life**: -$11.14

Negative values for the 15-inch screen and 10-hour battery life indicate these attributes were less preferred, relative to their alternatives, indicating consumers may not be willing to pay extra for them.

7. Interpretation of Negative Dollar Values

The negative dollar values for the 15-inch screen and 10-hour battery life were further investigated:

- **15-inch Screen Size**: The 15-inch screen received an average rating lower than the 14-inch, leading to a negative valuation.
- **10-hour Battery Life**: Although the 10-hour battery life received a slightly higher rating than the 8-hour option, the difference was minimal, resulting in a small dollar value.

These findings suggest that, while longer battery life and larger screens might seem beneficial, they may not hold as much value to consumers in this context.

References

GeeksforGeeks. (2021, July 26). Retrieved from https://www.geeksforgeeks.org/data-mining-and-recommender-systems/

GeeksforGeeks. (2024, May 20). Retrieved from https://www.geeksforgeeks.org/what-are-recommender-systems/

Michael, M. (2023, February 27). *aporia*. Retrieved from What are recommender systems? Use cases, types, and techniques: https://www.aporia.com/learn/recommender-systems/what-are-recommender-systems-use-cases-types-and-techniques/

Rina Caballar, C. S. (2024, June 19). *IBM*. Retrieved from What is a recommendation engine? : https://www.ibm.com/think/topics/recommendation-engine