DA ASSIGNMENT-3

TEAM ID: 18

TARUNKUMAR HOTHI-2023201042

ISWAR MAHAPATRO-2023201047

Association Rule Mining

Introduction

Association Rule Mining helps in identifying hidden relationships in datasets. It is popularly used for market basket analysis to discover product groupings bought together.

For this assignment, we developed a movie recommendation system using the FP-Growth algorithm, a method effective for mining frequent patterns.

Why FP-Growth?

- **Handling Large Datasets Efficiently**: FP-Growth processes large datasets faster than the Apriori algorithm as it avoids generating numerous candidate sets.
- **Compact Structure**: It uses an FP-Tree, minimizing database scans and efficiently compressing data.

- **No Candidate Set Generation**: FP-Growth directly generates frequent patterns without explicit candidate sets.
- **Scalability**: It scales well with increased data size, making it ideal for real-world applications like movie recommendations.

Building the Recommendation System

1. Data Preparation:

- We loaded and merged movie and rating datasets.
- Applied a threshold filter to remove low-rated movies (rating > 2) and users with fewer than 10 ratings.
- We split the data into training and testing sets by sampling a fraction of each user's
  data, and then converted the data into transactions.

2. FP-Tree Construction:

- Built an FP-Tree based on the frequency of movies watched together.
- Used a header table to track movies meeting the minimum support threshold.

3. Mining Frequent Patterns:

- Frequent itemsets were mined from the FP-Tree using conditional pattern bases.
- We recursively mined conditional FP-Trees to find all frequent itemsets.

4. Generating Association Rules:

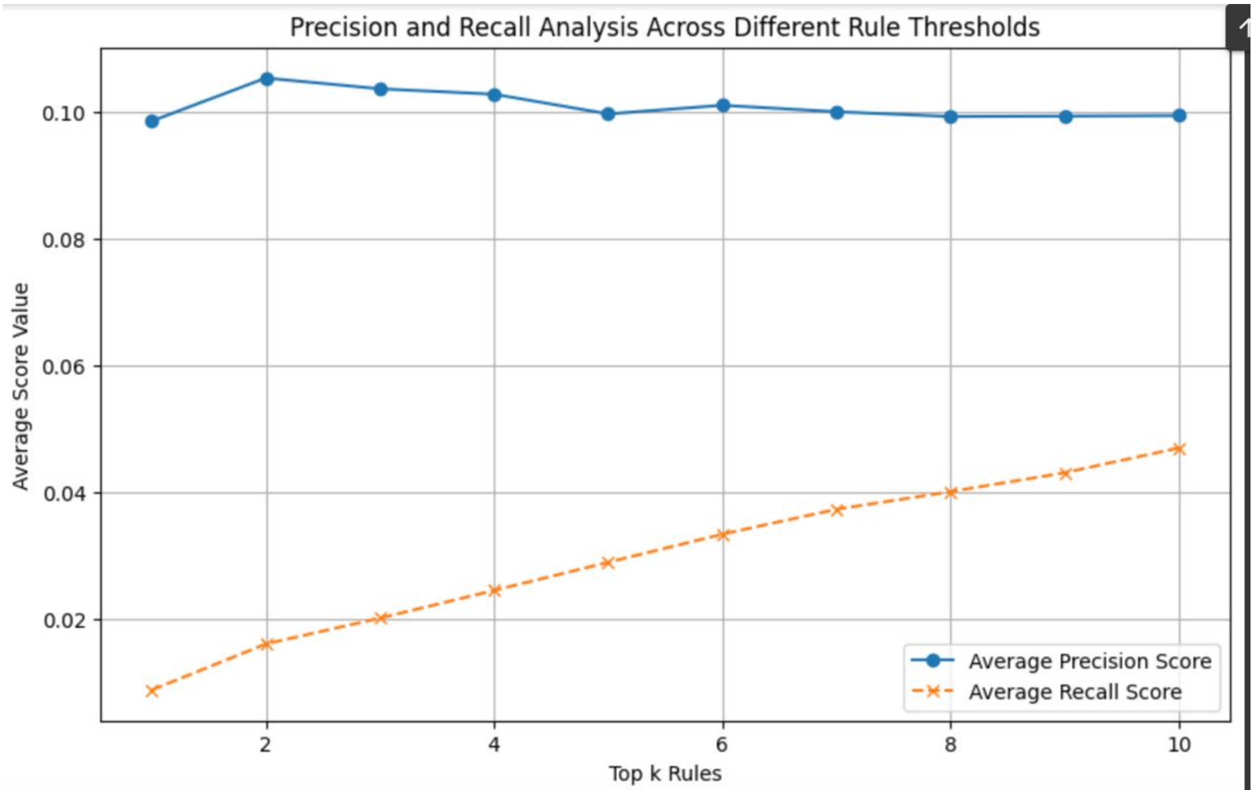
- Rules were generated from frequent itemsets with support and confidence metrics.
- Rules meeting a minimum confidence threshold were kept.

Evaluating the Recommendation System

- **Precision and Recall**:
    - We evaluated the system by calculating average precision and recall at different
      thresholds.
    - Precision remains stable at around 0.10 across various rule thresholds.
    - Recall improves as the number of rules increases, expanding the algorithm’s range of recommendations.

- **User-Specific Analysis**:
    - We analyzed the recommendation performance for specific users and we chose specific users to evaluate,and plotted for each user precision and recall.
    - We evaluated individual users' results. For instance, User 1 showed strong precision for top recommendations, while other users had varied results based on their history.

Results



- **Precision and Recall Analysis**:

  We plotted the average precision and recall by varying the number of rules from 1 to 10. Precision remains relatively constant at approximately 0.10 across different rule thresholds (k). This indicates that the FP-Growth algorithm consistently provides highly relevant and accurate recommendations, even as the number of rules increases.

- **Recall Performance**:

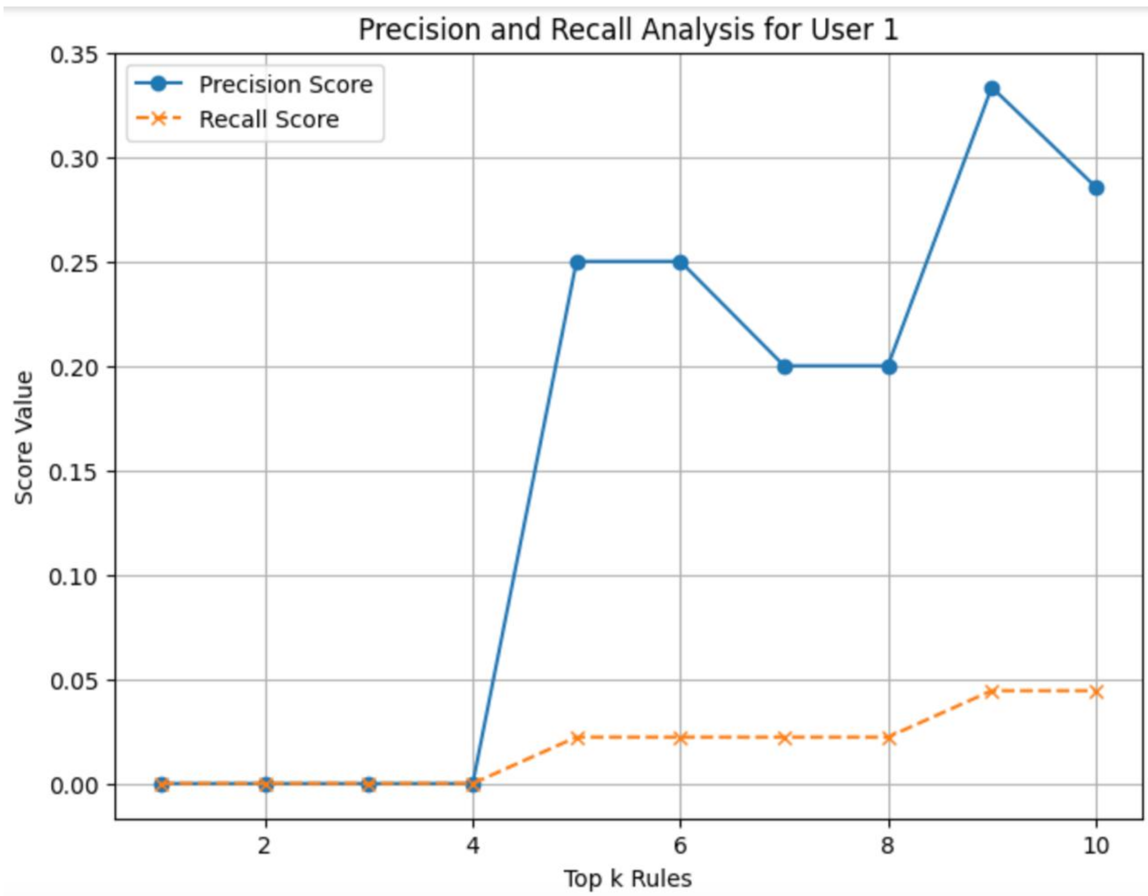
  As the number of top rules (k) grows, the system captures a broader range of relevant items, leading to improved recall. This highlights the algorithm's ability to expand its recommendation coverage over time without compromising accuracy.

- **Model Balance**:

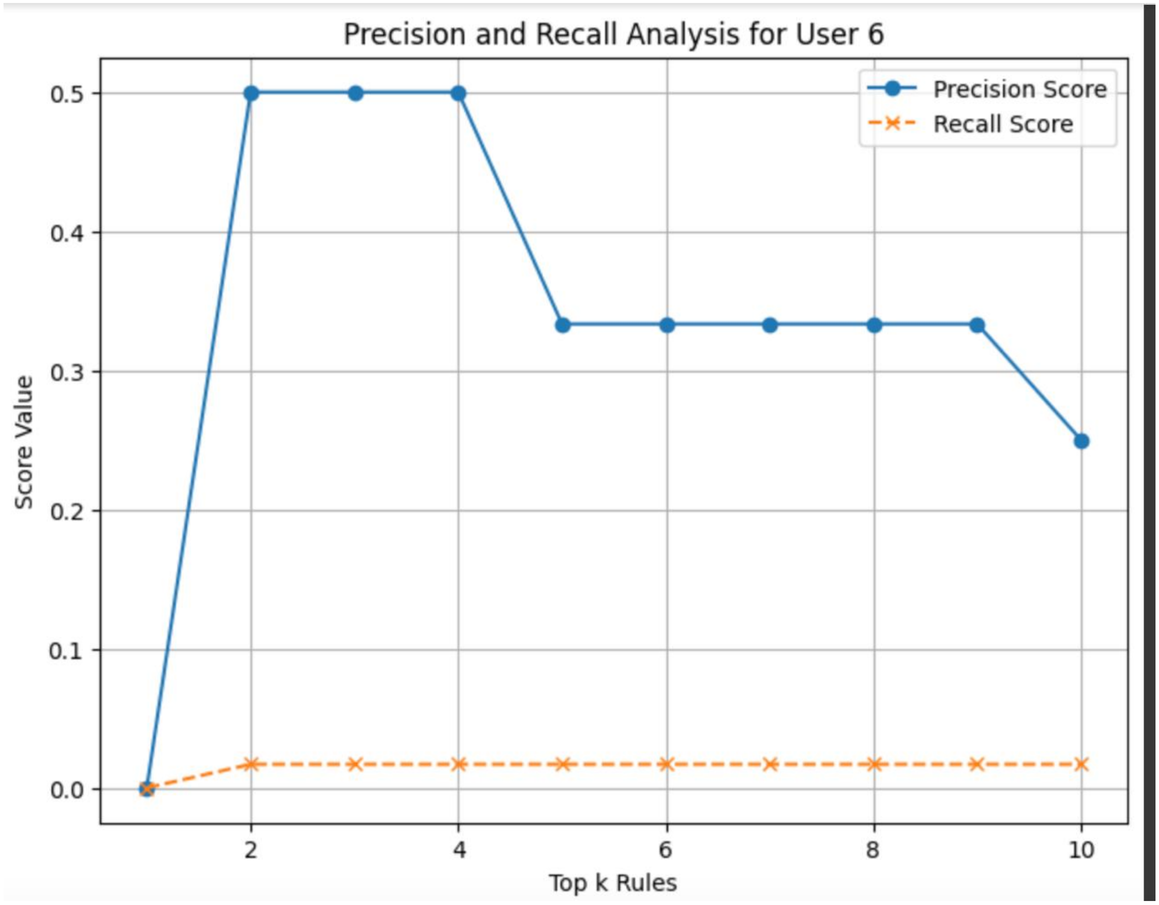
  The model strikes a strong balance between precision and recall, offering both accurate and comprehensive recommendations, which enhances the user experience.

User-Specific Performance:

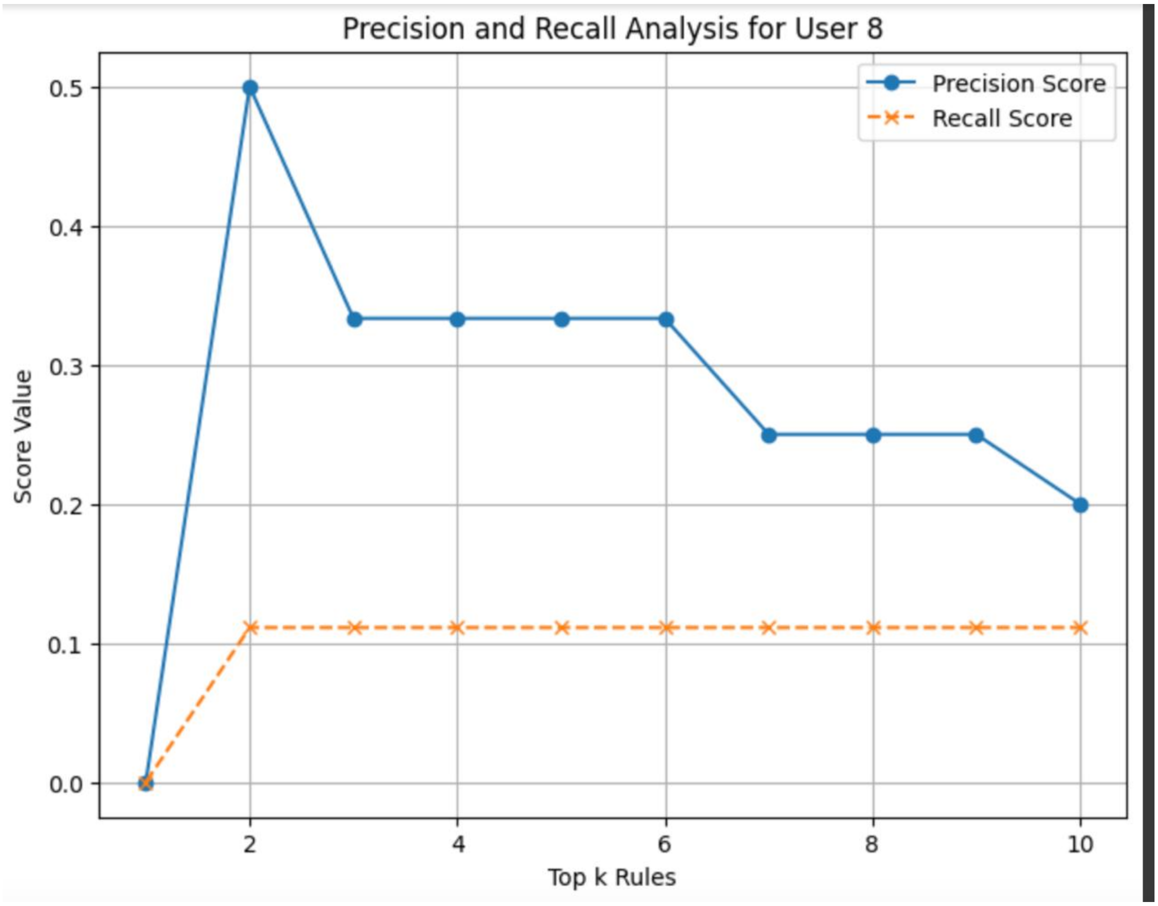
- **User 1**: The recommendation system delivers highly relevant suggestions, particularly at certain rule thresholds, reflecting the system's strong precision. However, recall could be further improved to capture a wider range of interests.



- **User 6**: For this user, the system provides precise recommendations at the top rule thresholds (k=2 to k=4), demonstrating the algorithm's effectiveness in delivering targeted suggestions early on.



- **User 8**: The system performs well in offering highly precise recommendations at the top rule threshold (k=2). Despite the early success, there is room for improving recall to capture a wider array of recommendations.



Conclusion

By utilizing the FP-Growth algorithm, we were able to efficiently process large datasets and generate meaningful recommendations. The algorithm excels in finding frequent patterns without requiring extensive computational resources.

We prepared the data, constructed the FP-Tree, mined frequent itemsets, and generated association rules. The system was thoroughly evaluated, confirming its ability to offer personalized movie recommendations based on users' viewing histories.