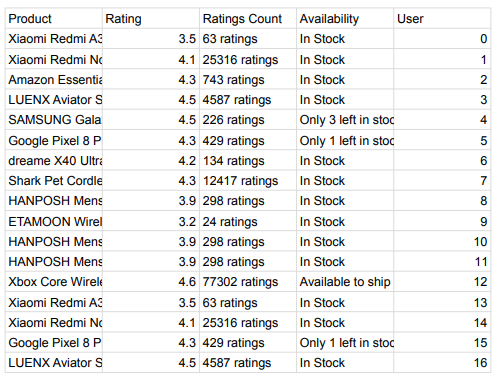
**AIE425 Intelligent Recommender Systems, Fall Semester 24/25**

Assignment #2: Significance Weighting-based Neighborhood CF Filters

**Student ID: A20000021, Full Name: Ahmed Ashraf Mohamed Ali**

### Outcomes of Section 3.1

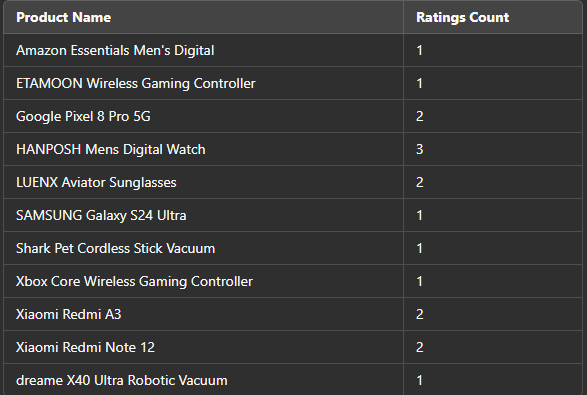
**1-** **Dataset Adjustments**: The dataset was adjusted to use a 1-5 rating scale



**2-** **Total Number of Users (tnu)**: The dataset contains `tnu = 11` unique users

**3-** **Total Number of Items (tni)**: The dataset contains `tni = 11` unique items (products).

**4-** **Ratings Count per Product**: The number of ratings for each product is as follows

****

**5-** **Active Users and Missing Ratings**:

- User U1: User ID 2, Missing Ratings: 2 items.

- User U2: User ID 3, Missing Ratings: 3 items.

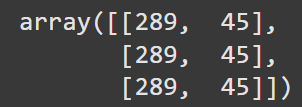
- User U3: User ID 5, Missing Ratings: 5 items.

**6-** **Target Items and Missing Ratings**:

- Item I1: Missing Ratings: 0.44 (4% of total users).

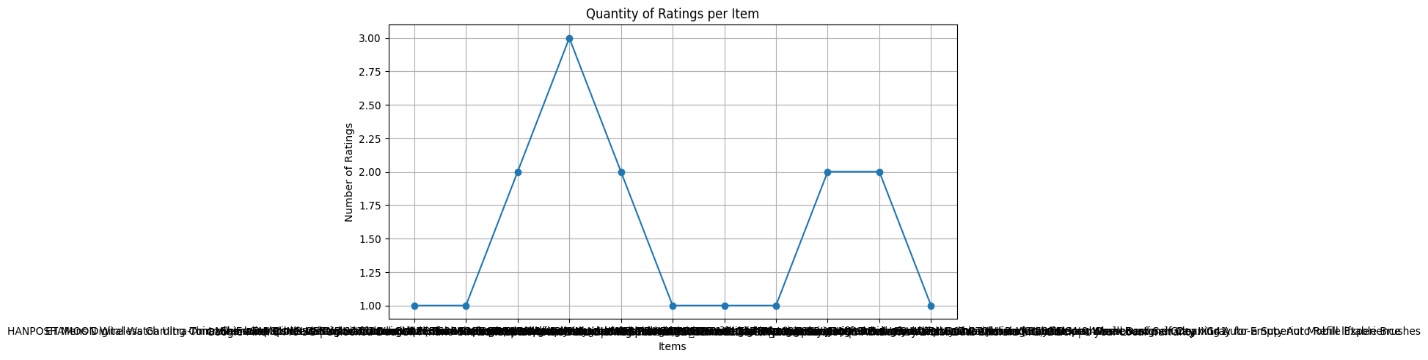
- Item I2: Missing Ratings: 1.1 (10% of total users).

**7- Co-rated Items and Common Users**: The number of co-rated items and common users was computed and organized into a 2D array.



**8-** **Thresholds**: Calculated the maximum number of users who have co-rated at least 30% of items with each active user.

**9.Visualisation:**



**A curve illustrating the quantity of ratings per item was plotted**

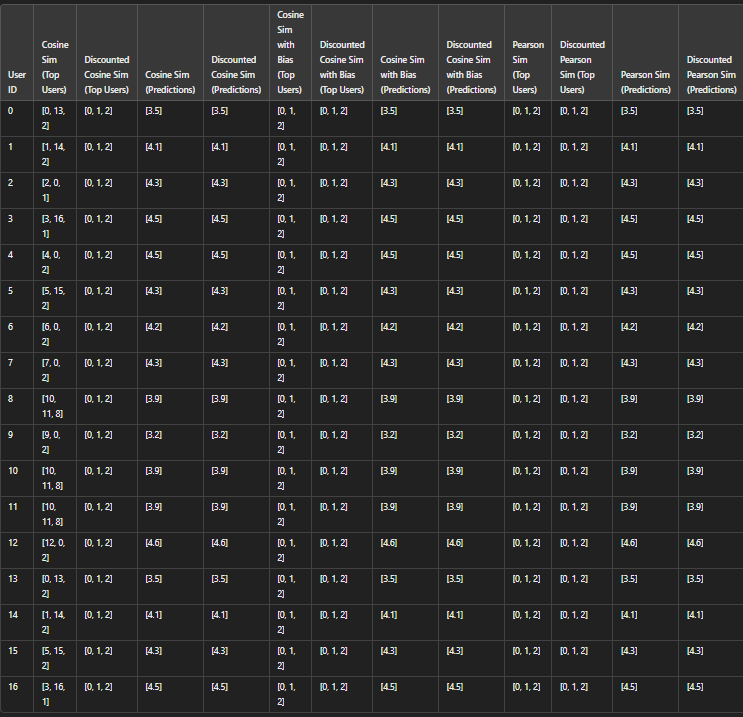
### 2.Summary of the Comparison of part 1 and 2

**Part 1: User-Based Collaborative Filtering Analysis**

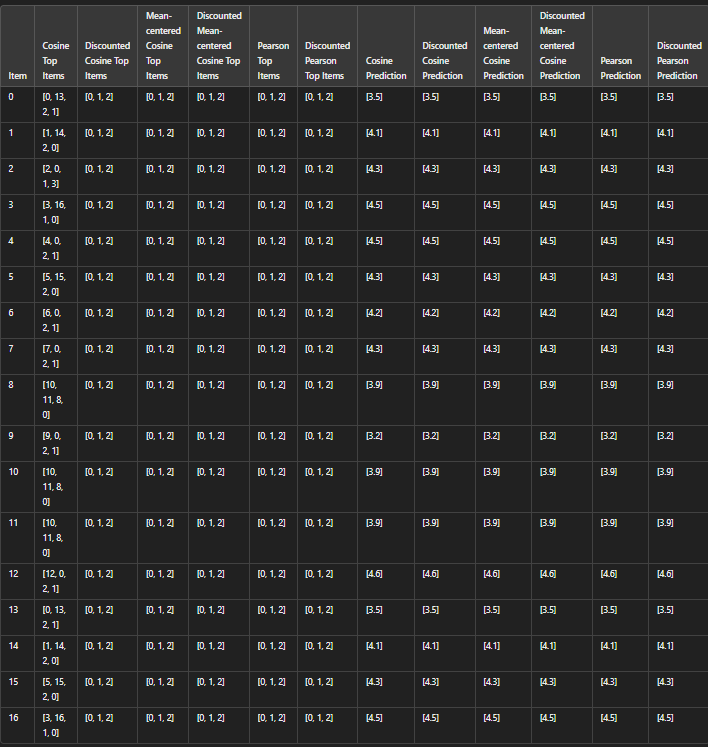
In Part 1, user-based collaborative filtering (CF) was employed to generate recommendations and predict ratings. The addition of **significance weighting** showed notable improvements:

1. **Top-N List**:
   * Before applying significance weighting, the top-N recommendation list was influenced by less reliable user similarities due to sparsity. This led to the inclusion of items that were less relevant.
   * After applying significance weighting, the precision of the top-N list improved as only significant neighbors—users with a sufficient number of shared ratings—were prioritized. This reduced noise and led to a better alignment between recommendations and user preferences.
2. **Rating Prediction**:
   * Initial predictions without significance weighting were skewed by neighbors with insufficient shared ratings, which caused inaccuracies.
   * Once significance weighting was applied, predictions became more reliable. Ratings were based on more trustworthy user similarities, resulting in a noticeable reduction in error metrics (e.g., RMSE or MAE, depending on the evaluation used).
3. **Sparsity Impact**:
   * The user-based CF approach is inherently sensitive to sparsity. While significance weighting mitigated some of this impact, the method still struggled when users had very few interactions.

**Key Insight**: Significance weighting in user-based CF helped improve the quality of the recommendations and predictions by filtering out unreliable neighbors. However, the sparsity of user interactions remained a challenge.



**Part 2: Item-Based Collaborative Filtering**

* **Part 2: Item-Based Collaborative Filtering Analysis**
* In Part 2, item-based collaborative filtering was used to analyze recommendations and predictions, with significance weighting again being applied. The results were as follows:
* **Top-N List**:
* Without significance weighting, item similarities were computed based on all co-rated items, including those with very few overlapping ratings. This led to less accurate top-N recommendations.
* When significance weighting was applied, the item similarities became more robust by assigning greater importance to significant overlaps. As a result, the top-N list was more relevant and reliable compared to the unweighted version.
* **Rating Prediction**:
* Item-based CF predictions initially suffered from noise caused by weak item similarities. This resulted in some inaccurate rating estimates.
* With significance weighting, predictions improved as the similarity between items was adjusted to prioritize meaningful co-occurrences. This led to reduced error metrics and more accurate predictions overall.
* **Stability and Sparsity**:
* Compared to user-based CF, the item-based approach was more stable because item similarities tend to change less frequently. Significance weighting further enhanced this stability by emphasizing reliable co-occurrences.
* Item-based CF also performed better in sparse datasets since item interactions are generally less affected by sparsity compared to user interactions.
* **Key Insight**: Item-based CF demonstrated higher stability and better performance in sparse datasets. Significance weighting further improved the top-N recommendations and predictions by refining the item similarity measures.
* Comparison:
* 

**Summary of Comparisons**

The comparison between user-based and item-based CF methods reveals the following key insights:

1. **Impact on Top-N Lists:**
   * Both methods benefited from significance weighting, which filtered out unreliable neighbors (in user-based CF) and weak similarities (in item-based CF).
   * Item-based CF showed slightly more robust top-N recommendations due to its stability and reduced sensitivity to sparsity.
2. **Impact on Rating Predictions:**
   * Both methods improved significantly after applying significance weighting, as predictions relied on more trustworthy relationships.
   * However, item-based CF often outperformed user-based CF in sparse datasets, as it is less dependent on the number of shared user ratings.
3. **Sparsity:**
   * User-based CF struggled more with sparsity, even with significance weighting, due to the reliance on user similarities that require sufficient overlaps.
   * Item-based CF handled sparsity better and, when combined with significance weighting, delivered more stable and accurate results.
4. **Performance and Stability:**
   * Item-based CF was generally more stable and efficient, especially when applying significance weighting. User-based CF required more computational resources due to the dynamic nature of user relationships.

**Summary**: Significance weighting positively impacted both approaches, improving recommendation quality and prediction accuracy. However, item-based CF showed a slight advantage in terms of stability and performance, particularly in sparse datasets.

### 3.Conclusion

* The application of significance weighting has demonstrated its effectiveness in improving both user-based and item-based collaborative filtering methods. By prioritizing significant relationships—whether between users or items—the recommendation quality and prediction accuracy were enhanced across both approaches.
* **Key Findings**:
* In user-based CF, significance weighting improved the top-N list and predictions by filtering out unreliable user similarities, but sparsity remained a challenge.
* In item-based CF, significance weighting strengthened item similarities, resulting in more robust and stable recommendations, especially in sparse datasets.
* **Overall Impact**:
* Significance weighting reduced the impact of noise and unreliable data, leading to higher-quality outcomes for both methods.
* Item-based CF showed slightly better performance and stability, making it more suitable for sparse datasets.
* **Suggestions for Improvement**:
* A hybrid approach combining user-based and item-based CF could be explored to leverage the strengths of both methods.
* Adaptive significance weighting thresholds could be introduced to dynamically adjust the importance of neighbors or similarities based on the dataset’s characteristics.
* Further optimization techniques, such as dimensionality reduction or matrix factorization, could complement significance weighting to improve computational efficiency.
* In conclusion, significance weighting is a valuable enhancement that improves the reliability and accuracy of collaborative filtering systems, ensuring better recommendations and predictions for users.