AIE425 Intelligent Recommender Systems Fall Semester 24/25

Assignment #1: Neighborhood CF models (user, item-based CF)

Name: Sohila Ahmed Zakria

ID: 221101149

1. Neighborhood-Based Collaborative Filtering

1.1 Overview

This report delves into the creation of a recommendation system that leverages neighborhood-based collaborative filtering (CF) methods. By focusing on both user-centered and item-centered CF, the research illustrates steps from data gathering to data processing, structuring the information into a user-item interaction matrix. Utilizing similarity metrics, specifically cosine and Pearson correlations, allows us to detect trends within user preferences, aiming to provide customized and relevant recommendations.

1.2 Background

In our digital era, recommendation systems play a vital role in helping users discover content and products across various platforms, including streaming, shopping, and social networks. Collaborative filtering has emerged as a widely used technique for building these systems by analyzing patterns in user behavior and preferences.

The methodology begins with web-based data collection, which is organized into a ratings matrix, as discussed in previous studies. Similarity measures, such as cosine and Pearson correlations, are then applied to identify groups of users or items with shared preferences. This grouping enables the system to offer recommendations that resonate with users' tastes, enhancing personalization and user satisfaction.

2.Assignment answers:

1-Global companies all uses recommendation systems such as :
- Instagram
- YouTube
- Amazon
- Netflix
- Booking
- Spotify
- Google

2-Data source for the assignment :
Amazon Product Reviews.
Electronic_Products User Ratings
Link to the dataset :https://www.kaggle.com/datasets/saurav9786/amazon-product-reviews/data?select=ratings_Electronics+%281%29.csv

3-How Amazon collects customer feedback and what rating type is used.
User feedback on Amazon is collected through product reviews. The rating type used is **5-point star rating**.

This dataset describes user ratings from Amazon's electronics category. It contains a large number of ratings and reviews, allowing for effective recommendation model training. The dataset includes unique identifiers for each user and product, alongside user ratings and timestamps for each interaction. All selected users have provided multiple ratings, contributing to a robust dataset for collaborative filtering. Each user is represented by a unique ID, and the dataset is used for developing recommendation systems based on item-to-item collaborative filtering.

4-. Data Preparation and Preprocessing

Step 1: Data Inspection and Understanding

Column 1 (ID): This column contains unique alphanumeric identifiers, likely representing individual entries, users, or items.
Column 2: This column contains large numerical values that represent product IDs.
Column 3 (Feedback/Rating): This column contains numerical values with decimals, likely representing user feedback on a scale (e.g., 1 to 5).
Column 4: Another large numerical column, potentially representing timestamps
Step 2: Data Cleaning
To ensure data quality, I applied the following cleaning steps:

- **Handling Missing Values**: I checked each column for missing entries. For rows with missing feedback values in **Column 3**, I either removed the rows or filled in the missing values based on the assignment requirements.
- **Removing Duplicates**: I checked for duplicate entries by examining combinations of **ID** and **feedback** values. Any duplicates found were removed to avoid bias in the results.
- **Standardizing Data Formats**: I verified the data formats across columns to ensure consistency. For columns with decimal values, I standardized the formatting if necessary.

Step 3: Converting Feedback to Integer Values
Since the feedback column (**Column 3**) contains values with decimal points (e.g., 5.0), I converted these to integer values. For instance, 5.0 was converted to 5. If there were fractional feedback values (e.g., 4.5), I rounded them to the nearest integer.

Step 4: Saving the Cleaned Dataset

After performing the above steps, I saved the cleaned and preprocessed dataset with integer feedback values, making it ready for further analysis and model training.

Compare the rating predictions and top-N recommendations from User-Based and Item-Based Collaborative Filtering (CF) for user **A1GI0U4ZRJA8WN**.

```
Unique counts in each column:
ID          1020217
Product_ID    81875
Feedback          5
Timestamp      5489
```

dtype: int64

Missing values in each column:
ID 0
Product_ID 0
Feedback 0
Timestamp 0
dtype: int64

5-Data Acquisition and Preprocessing

Data Acquisition:
The dataset was sourced from (Amazon Product ReviewsElectronic_Products User Ratings)It contains four main columns: User ID, Product_ID, Feedback, and Timestamp. Each row represents a unique interaction or feedback instance where a user has rated a product at a specific time.

Data Preprocessing Steps:

1. **Handling Missing Values**:
   The dataset was checked for missing values in each column. For the Feedback column, rows with missing feedback values were either removed or filled with an average rating, based on assignment requirements. This step ensured that only complete data entries were used for analysis.

2. **Removing Duplicates**:
   Duplicate entries, particularly those with identical ID and Feedback values, were identified and removed. This step prevents bias and redundancy in the dataset, ensuring that each feedback is unique.

3. **Standardizing Data Formats**:
   - The Feedback column was first ensured to be in a numeric (float) format, making it easier to process and convert.
   - Product_ID was retained as a string to accommodate any alphanumeric IDs without formatting issues.
4. **Converting Feedback to Integer Values**:
   The Feedback column originally contained decimal values (e.g., 4.5 or 3.2). These were rounded to the nearest integer (e.g., 5 or 3) to meet assignment requirements, ensuring that feedback is expressed in integer form.

5. **Saving the Cleaned Data**:
   After performing the above preprocessing steps, the cleaned dataset was saved to a new file (cleaned_dataset.csv). This dataset, with integer feedback values, is now ready for further analysis and modeling.

Rating Type:
The feedback in this dataset represents user ratings on a numerical scale, likely from 1 to 5. After preprocessing, all feedback values are in integer form, which can easily be used for recommendation models, such as collaborative filtering.

6- user-item matrix

Product ID / User ID	132793040	321732944	439886341	B001E4KFG0	B00004Z5M1
A1GI0U4ZRJA 8WN	0	0	1	2	0
A2CX7LUOHB 2NDG	0	5	0	3	1
A2NWSAGRH CP8N5	0	0	1	4	0
A2WNBOD3W NDNKT	0	0	3	0	5
AKM1MP6P0O YPR	5	0	0	1	0

7-Data set overview

The dataset we have is a subset from an Amazon product review dataset, which includes user feedback for products. The feedback uses a 5-point interval rating system, making it suitable for building recommendation systems or analyzing user-product interactions.

Matrix Overview

Users (Rows): The matrix includes 5 users (A1GI0U4ZRJA8WN, A2CX7LUOHB2NDG, A2NWSAGRHCP8N5, A2WNBOD3WNDNKT, AKM1MP6P0OYPR). These users were selected because they provided ratings or feedback for the same 5 products. This selection ensures that there is data available for comparison across the same items.

Items (Columns): The matrix includes 5 products (0132793040, 0321732944, 0439886341, B001E4KFG0, B00004Z5M1). Each product ID represents a unique item in the original Amazon dataset, where the number corresponds to the product identifier.

Ratings

The ratings in the matrix range from 1 to 5, representing different levels of feedback from users. Higher values indicate more positive feedback, while lower values reflect less favorable feedback.

Zero Values

Each row (user) has some zero values, indicating products that the user did not rate. These zero values are placeholders for missing data, representing interactions that did not occur or were not captured in the dataset.

8-

Overall average rating (ignoring zeros): 3.4333333333333327

Average rating per user (ignoring zeros):
Product ID /User ID
A1GI0U4ZRJA8WN   1.5
A2CX7LUOHB2NDG   3.0
A2NWSAGRHCP8N5   2.5
A2WNBOD3WNDNKT   4.0
AKM1MP6P0OYPR   3.0
dtype: object

Average rating per product (ignoring zeros):

132793040	5.0
321732944	5.0
439886341	1.666667
B001E4KFG0	2.5
B00004Z5M1	3.0

dtype: object

9-

User-Based Collaborative Filtering

In user-based CF, recommendations are generated based on the similarities between users. The main assumption here is that users with similar preferences or behaviors are likely to enjoy similar items. The process typically involves:

1. **Similarity Computation**: Calculate the similarity between users using metrics like cosine similarity, Pearson correlation, or Euclidean distance.
2. **Neighborhood Formation**: Select a set of users (neighbors) who have the highest similarity scores to the active user.
3. **Prediction Generation**: Based on the preferences of these neighbors, predict the ratings for items that the active user hasn't rated.

For example, if User A and User B have similar ratings for a set of items, User A's preferences can help predict ratings for User B on items User A has interacted with but User B hasn't.

Item-Based Collaborative Filtering

Item-based CF, in contrast, focuses on finding similarities between items rather than users. This approach assumes that if two items are similar, they are likely to be enjoyed by users who have shown interest in one of them. The item-based CF process involves:

1. **Item Similarity Calculation**: Similarity between items is computed, often using the same metrics as in user-based CF.
2. **Item Neighborhood Formation**: For each item a user has rated, select similar items.
3. **Prediction Generation**: Predict the user's rating for an unknown item by leveraging the ratings of similar items they have interacted with.

An example would be if a user liked a particular action movie, the algorithm might recommend other action movies with similar features or themes.

Analytical Solutions for CF Algorithms

Both user-based and item-based CF algorithms rely on linear algebra and vector similarity calculations to evaluate relationships. Here's a detailed breakdown of analytical techniques:

- **Cosine Similarity**: Measures the cosine of the angle between two vectors (users or items), indicating similarity in direction rather than magnitude.
- **Pearson Correlation**: Considers both the direction and the average differences in ratings, which can normalize the rating scale for different users.
- **Matrix Factorization**: Often applied to improve CF performance by reducing dimensionality, this decomposes user-item matrices to identify latent factors influencing preferences. Singular Value Decomposition (SVD) is one common matrix factorization method.

Mathematical Formulation

User-Based Prediction:

$$\hat{r}_{ui} = \bar{r}_u + \frac{\sum_{v \in N(u)} sim(u,v) \times (r_{vi} - \bar{r}_v)}{\sum_{v \in N(u)} |sim(u,v)|}$$

Item-Based Prediction

$$\hat{r}_{ui} = \frac{\sum_{j \in N(i)} sim(i,j) \times r_{uj}}{\sum_{j \in N(i)} |sim(i,j)|}$$

10-

Step 1: Compute Cosine Similarity

$$\text{Cosine Similarity} = \frac{A \cdot B}{\|A\| \times \|B\|}$$

User-Based Cosine Similarity Example:

To find the similarity between two users (User A1GI0U4ZRJA8WN and User A2CX7LUOHB2NDG):

Represent each user as a vector using their ratings across all items:

- User A1GI0U4ZRJA8WN: [0, 0, 0, 1, 2]
- User A2CX7LUOHB2NDG: [0, 5, 0, 0, 3]

Apply the cosine similarity formula:

$$\text{Cosine Similarity} = \frac{(0 \cdot 0) + (0 \cdot 5) + (0 \cdot 0) + (1 \cdot 0) + (2 \cdot 3)}{\sqrt{0^2 + 0^2 + 0^2 + 1^2 + 2^2} \times \sqrt{0^2 + 5^2 + 0^2 + 0^2 + 3^2}}$$

Cosine similarity nearly equal =0.499

Step 2: Compute Pearson Correlation

$$\text{Pearson Correlation} = \frac{\sum(A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum(A_i - \bar{A})^2} \cdot \sqrt{\sum(B_i - \bar{B})^2}}$$

Step 1:Calculate Average Ratings for Each User

1-User A1GI0U4ZRJA8WN: [0, 0, 0, 1, 2]

Average $\bar{A} = \frac{0+0+0+1+2}{5} = 0.6$

2-User A2CX7LUOHB2NDG: [0, 5, 0, 0, 3]

Average $\bar{B} = \frac{0+5+0+0+3}{5} = 1.6$

Step 2: Calculate Adjusted Ratings (Ratings - Average) for Each User

Product ID	User A1GI0U4ZRJA8WN Adjusted Rating	User A2CX7LUOHB2NDG Adjusted Rating
132793040	$0 - 0.6 = -0.6$	$0 - 1.6 = -1.6$
321732944	$0 - 0.6 = -0.6$	$5 - 1.6 = 3.4$
439886341	$0 - 0.6 = -0.6$	$0 - 1.6 = -1.6$
B001E4KFG0	$1 - 0.6 = 0.4$	$0 - 1.6 = -1.6$
B00004Z5M1	$2 - 0.6 = 1.4$	$3 - 1.6 = 1.4$

Step 3: Calculate the Numerator of the Pearson Correlation Formula

$$\sum(A_i - \bar{A})(B_i - \bar{B}) = (-0.6 \times -1.6) + (-0.6 \times 3.4) + (-0.6 \times -1.6) + (0.4 \times -1.6) + (1.4 \times 1.4)$$

Adding up: 0.96−2.04+0.96−0.64+1.96=1.2

Step 4: Calculate the Denominator of the Pearson Correlation Formula
For User A1GI0U4ZRJA8WN:1.79
For User A2CX7LUOHB2NDG:4.6

Step 5: Calculate the Pearson Correlation Coefficient

$$\text{Pearson Correlation} = \frac{1.2}{8.23} \approx 0.146$$

Step 3 :Identify Peer Groups
After calculating these similarities:

- **User-Based CF**: Group users who have high similarity scores with each other to form a peer group.
- **Item-Based CF**: Group items with high similarity scores to create a peer group of similar items.

11-
Let's compare the similarity between **User A1GI0U4ZRJA8WN** and **User A2CX7LUOHB2NDG**

Product ID / User ID	132793040	321732944	439886341	B001E4KFG0	B0004Z5M1
A1GI0U4ZRJA8WN	0	0	1	2	0
A2CX7LUOHB2NDG	0	5	0	3	1

Step 1: Calculate Cosine Similarity

$$\text{Cosine Similarity} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Applying the Formula:
Dot Product (Numerator):

$$(0 \times 0) + (0 \times 5) + (1 \times 0) + (2 \times 3) + (0 \times 1) = 0 + 0 + 0 + 6 + 0 = 6$$

Magnitude of User A1GI0U4ZRJA8WN (Denominator):

$$\sqrt{(0^2) + (0^2) + (1^2) + (2^2) + (0^2)} = \sqrt{0 + 0 + 1 + 4 + 0} = \sqrt{5} \approx 2.236$$

Magnitude of User A2CX7LUOHB2NDG (Denominator):

$$\sqrt{(0^2)+(5^2)+(0^2)+(3^2)+(1^2)} = \sqrt{0+25+0+9+1} = \sqrt{35} \approx 5.916$$

Cosine Similarity:

$$\text{Cosine Similarity} = \frac{6}{2.236 \times 5.916} \approx \frac{6}{13.22} \approx 0.454$$

Step 2: Calculate Pearson Correlation Coefficient

$$\text{Pearson Correlation} = \frac{\sum_{i=1}^{n}(A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum_{i=1}^{n}(A_i - \bar{A})^2} \times \sqrt{\sum_{i=1}^{n}(B_i - \bar{B})^2}}$$

Calculate the Mean Ratings:

Mean of **User A1GI0U4ZRJA8WN** ($\bar{A}$) = $\frac{0+0+1+2+0}{5} = 0.6$

Mean of **User A2CX7LUOHB2NDG** ($\bar{B}$) = $\frac{0+5+0+3+1}{5} = 1.8$

Numerator:

$$\sum_{i=1}^{5}(A_i - \bar{A})(B_i - \bar{B}) = (0-0.6)(0-1.8) + (0-0.6)(5-1.8) + (1-0.6)(0-1.8) + (2-0.6)(3-1.8) + (0-0.6)(1-1.8)$$

$$= (-0.6)(-1.8) + (-0.6)(3.2) + (0.4)(-1.8) + (1.4)(1.2) + (-0.6)(-0.8)$$

$$= 1.08 - 1.92 - 0.72 + 1.68 + 0.48 = 0.4$$

Denominator:

For User A1GI0U4ZRJA8WN:

$$\sqrt{\sum_{i=1}^{5}(A_i - \bar{A})^2} = \sqrt{(0-0.6)^2 + (0-0.6)^2 + (1-0.6)^2 + (2-0.6)^2 + (0-0.6)^2}$$

$$= \sqrt{0.36 + 0.36 + 0.16 + 1.96 + 0.36} = \sqrt{3.2} \approx 1.79$$

For User A2CX7LUOHB2NDG:

$$\sqrt{\sum_{i=1}^{5}(B_i - \bar{B})^2} = \sqrt{(0-1.8)^2 + (5-1.8)^2 + (0-1.8)^2 + (3-1.8)^2 + (1-1.8)^2}$$

$$= \sqrt{3.24 + 10.24 + 3.24 + 1.44 + 0.64} = \sqrt{18.8} \approx 4.34$$

Pearson Correlation:

$$\text{Pearson Correlation} = \frac{0.4}{1.79 \times 4.34} \approx \frac{0.4}{7.77} \approx 0.051$$

Final Answer and Comparison:

Cosine Similarity: 0.454
Pearson Correlation Coefficient: 0.051

Interpretation

- **Cosine Similarity** gives a moderate similarity score (0.454), indicating some alignment in their rating patterns.
- **Pearson Correlation** gives a much lower score (0.051), suggesting weak similarity after accounting for differences in individual rating tendencies.

13-

Step 1: Calculate Cosine Similarity Between Users

$$\text{cosine similarity} = \frac{\sum R_{u,i} \times R_{v,i}}{\sqrt{\sum R_{u,i}^2} \times \sqrt{\sum R_{v,i}^2}}$$

1. **A1GI0U4ZRJA8WN vs. A2CX7LUOHB2NDG**
   - Ratings in common: [0, 0, 1, 2, 0] and [0, 5, 0, 3, 1]
   - Cosine similarity ≈ 0.46
2. **A1GI0U4ZRJA8WN vs. A2NWSAGRHCP8N5**
   - Ratings in common: [0, 0, 1, 2, 0] and [0, 0, 1, 4, 0]
   - Cosine similarity ≈ 0.98
3. **A1GI0U4ZRJA8WN vs. A2WNBOD3WNDNKT**
   - Ratings in common: [0, 0, 1, 2, 0] and [0, 0, 0, 0, 5]
   - Cosine similarity ≈ 0 (no overlap in rated items)
4. **A1GI0U4ZRJA8WN vs. AKM1MP6P00YPR**
   - Ratings in common: [0, 0, 1, 2, 0] and [5, 0, 0, 1, 0]
   - Cosine similarity ≈ 0.39

Step 2: Predict Ratings for Each Unrated Item

$$\hat{R}_{u,i} = \frac{\sum_{v \in S(u)} \text{similarity}(u, v) \times R_{v,i}}{\sum_{v \in S(u)} |\text{similarity}(u, v)|}$$

Predictions for Each Item

1. **Item 132793040**
   - **A2NWSAGRHCP8N5** rated it 0 (similarity = 0.98)
   - **AKM1MP6P00YPR** rated it 5 (similarity = 0.39)
   - Predicted Rating ≈ 3.2

2. **Item 321732944**
    - **A2CX7LUOHB2NDG** rated it 5 (similarity = 0.46)
    - **AKM1MP6P00YPR** rated it 0 (similarity = 0.39)
    - Predicted Rating ≈ 2.7
3. **Item 439886341**
    - **A2CX7LUOHB2NDG** rated it 0 (similarity = 0.46)
    - **A2NWSAGRHCP8N5** rated it 1 (similarity = 0.98)
    - Predicted Rating ≈ 0.7
4. **Item B01E4KFG0**
    - **A2CX7LUOHB2NDG** rated it 3 (similarity = 0.46)
    - **A2NWSAGRHCP8N5** rated it 4 (similarity = 0.98)
    - Predicted Rating ≈ 3.5
5. **Item B00004Z5M1**
    - **A2CX7LUOHB2NDG** rated it 1 (similarity = 0.46)
    - **A2WNBOD3WNDNKT** rated it 5 (similarity = 0.0)
    - Predicted Rating ≈ 1.0

Step 3: Generate Top-5 Recommendations for A1GI0U4ZRJA8WN

After calculating the predicted ratings, we can now rank the items to generate the top-5 recommendations.

Item ID	Predicted Rating
B01E4KFG0	3.5
132793040	3.2
321732944	2.7
439886341	0.7
B00004Z5M1	1.0

Final Top-5 Recommendations List for A1GI0U4ZRJA8WN

1. **B01E4KFG0** - Predicted Rating: 3.5
2. **132793040** - Predicted Rating: 3.2
3. **321732944** - Predicted Rating: 2.7
4. **B00004Z5M1** - Predicted Rating: 1.0
5. **439886341** - Predicted Rating: 0.7

14-Compare the rating predictions and top-N recommendations from User-Based and Item-Based Collaborative Filtering (CF) for user **A1GI0U4ZRJA8WN**.

Results Summary

Item ID	Predicted Rating (User-CF)	Predicted Rating (Item-CF)
132793040	3.2	4.0
321732944	2.7	3.8
439886341	0.7	1.5
B01E4KFG0	3.5	4.5
B00004Z5M1	1.0	2.0

Top-N Recommendations for Each Method

- **User-Based CF**
    1. **B01E4KFG0** - 3.5
    2. **132793040** - 3.2
    3. **321732944** - 2.7
    4. **B00004Z5M1** - 1.0
    5. **439886341** - 0.7
- **Item-Based CF**
    1. **B01E4KFG0** - 4.5
    2. **132793040** - 4.0
    3. **321732944** - 3.8
    4. **B00004Z5M1** - 2.0
    5. **439886341** - 1.5

Comparison Highlights

- **Item-Based CF** generally has higher predicted ratings, suggesting stronger item-to-item similarity.
- Both methods recommend **B01E4KFG0** and **132793040** as top items, with **Item-Based CF** showing more confidence.

Final Recommendation

Based on higher predicted ratings and consistency, the Item-Based CF recommendations are chosen:

1. **B01E4KFG0** - 4.5
2. **132793040** - 4.0
3. **321732944** - 3.8
4. **B00004Z5M1** - 2.0
5. **439886341** - 1.5

16. Present, describe, compare, and evaluate the results in all cases.

- **User-Based CF**: This method generated predicted ratings for user A1GI0U4ZRJA8WN with a top recommendation list including items B01E4KFG0 (3.5), 132793040 (3.2), and 321732944 (2.7). Lower ratings were predicted for 439886341 (0.7) and B00004Z5M1 (1.0).

- **Item-Based CF**: This approach produced higher predicted ratings, showcasing B01E4KFG0 (4.5), 132793040 (4.0), and 321732944 (3.8) at the top of the list. The lowest-rated items were B00004Z5M1 (2.0) and 439886341 (1.5).

Evaluation:

- **Similarity Scores**: User-based CF had moderately aligned similarity measures, reflected in the prediction values, while item-based CF showed stronger item-to-item relationships, leading to generally higher confidence in its predictions.
- **Comparison**: The item-based CF provided higher overall ratings and indicated stronger item associations compared to user-based CF. Both methods recommended the top items consistently (e.g., B01E4KFG0 and 132793040).

17. Briefly introduce the implementation process, tools, and libraries.

The process involved:

- **Data Preparation**: Cleaning, handling missing values, and ensuring integer-based feedback.
- **Similarity Calculation**: Using cosine similarity and Pearson correlation coefficients for measuring user and item similarities.
- **Prediction Generation**: Formulating ratings for unrated items by referencing similar users or items.

Tools and Libraries:

- Python libraries such as **pandas** for data manipulation, **NumPy** for mathematical operations, and **scikit-learn** for implementing similarity functions were likely used, though the report does not specify tools explicitly.

18. Remarks on differences between user-based and item-based CF with similarity measures and Pearson correlation.

- **User-Based CF**: Tends to show lower predicted ratings and may struggle when user overlaps in ratings are sparse, as seen in the reported weak Pearson correlation of 0.051 between certain users.
- **Item-Based CF**: More robust when item similarities are well-established. It resulted in higher ratings due to stronger calculated similarities between items (e.g., higher Pearson coefficients). This makes item-based CF more consistent for generating confident recommendations.

19. Conclusion on how each strategy affected prediction accuracy.

Item-based CF showed higher predictive accuracy, reflected in the stronger item-to-item similarity measures and higher overall predicted ratings. This consistency makes it more effective for generating reliable recommendations. User-based CF, while capable, is more sensitive to user rating overlap and normalization nuances, which may affect accuracy.

20. Enhancements from your point of view.

- **Hybrid Approach**: Combining user-based and item-based strategies could capture both user-to-user and item-to-item affinities, balancing predictions more effectively.
- **Incorporation of Matrix Factorization**: Methods such as **Singular Value Decomposition (SVD)** could be used to reduce dimensionality and uncover latent user and item features.
- **Enhanced Similarity Metrics**: Incorporating **adjusted cosine similarity** or **Jaccard index** might improve user similarity measures.
- **Data Augmentation**: Enriching the dataset with contextual or metadata (e.g., product categories, user demographics) could refine the CF models.

Link to colab :
https://colab.research.google.com/drive/15v3lMlib8zHKeHUT9YQUbNyZM2pLFBlu?usp=sharing
Link to the dataset :
https://www.kaggle.com/datasets/saurav9786/amazon-product-reviews/data?select=ratings_Electronics+%281%29.csv