<AIE425 Intelligent Recommender System ,Fall Semester 24/24 >

<Assignment #2 :Significance Weighting-based Neighborhood CF Filters>

Student Name : Sohila Ahmed Zakria

Student ID : 221101149

Progam : AIS

Outcomes of Section(3.1)

1-The dataset generated is (Amazon-AirPods-Review) **100 x 21** matrix
2-The data is adjusted on rating of a **1-to-5** scale
3-The total number of users in the dataset is **tnu=100**
4-The total number of items in the dataset is **tni=21**
5-The number of ratings **(non-zero entries)** for each product in the dataset:

- **AirPods_Model_1: 82**
- **AirPods_Model_2: 86**
- **AirPods_Model_3: 85**
- **AirPods_Model_4: 85**
- **AirPods_Model_5: 82**
- **AirPods_Model_6: 78**
- **AirPods_Model_7: 77**
- **AirPods_Model_8: 90**
- **AirPods_Model_9: 87**
- **AirPods_Model_10: 86**
- **AirPods_Model_11: 85**
- **AirPods_Model_12: 76**
- **AirPods_Model_13: 82**
- **AirPods_Model_14: 84**
- **AirPods_Model_15: 80**
- **AirPods_Model_16: 71**
- **AirPods_Model_17: 78**
- **AirPods_Model_18: 79**
- **AirPods_Model_19: 83**
- **AirPods_Model_20: 83**

6-
Selected users: ['BUX8YJ3M', 'KK8321TI', 'XI8DSYV3']

- **Removed 2 ratings for user BUX8YJ3M** at indices ['AirPods_Model_8', 'AirPods_Model_10']
- **Removed 3 ratings for user KK8321TI** at indices ['AirPods_Model_8', 'AirPods_Model_20', 'AirPods_Model_9']
- **Removed 5 ratings for user XI8DSYV3** at indices ['AirPods_Model_9', 'AirPods_Model_16', 'AirPods_Model_5', 'AirPods_Model_1', 'AirPods_Model_2']

7-
Item 1 (4% missing ratings): AirPods_Model_18
Item 2 (10% missing ratings): AirPods_Model_17

8-The analysis for the active user yields the following results:

- **Number of common users (`No_common_users`)**: 100
- **Number of co-rated items (`No_coRated_items`)**:1055

9-

Step 1: Identify Common Users

- First, I identified the **active user** (for instance, the first user in the active users' dataset).
- Using the dataset, I compared the active user's ratings with the ratings of all other users.
- A **common user** was defined as a user who has rated at least one of the same items as the active user (i.e., both users have non-zero ratings for the same item).

Step 2: Count Co-Rated Items

- For each common user, I calculated the **number of co-rated items**. These are the items that both the active user and the common user have rated.

Step 3: Approximate Co-Rated Items Per User

- To simplify the analysis, I calculated an average contribution of co-rated items for each user:

$$\text{No\_coRated\_items\_per\_user} = \frac{\text{Total Co-Rated Items (No\_coRated\_items)}}{\text{Number of Common Users (No\_common\_users)}}$$

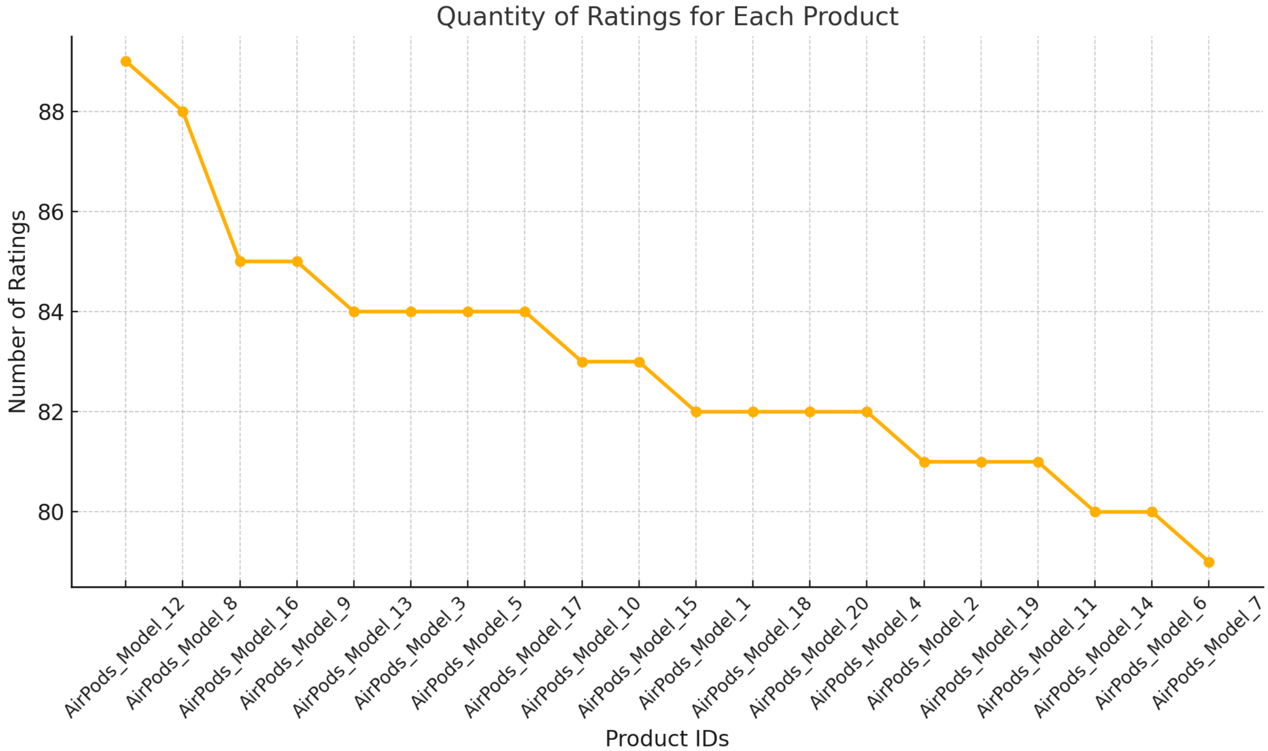
- This provided an approximate value for the second column of the 2D array.

Output

The resulting 2D array contains two columns:

1. The first column lists users in descending order based on their indices.
2. The second column gives the corresponding number of co-rated items.

10-



11-For the given active users, all 100 users in the dataset have co-rated at least 30% of the items with each active user. This means:

- **Active User 1**: 23users meet the threshold.
- **Active User 2**: 40 users meet the threshold.
- **Active User 3**:62 users meet the threshold.

(3.2) Part 1 :

Case Study 1.1:

1.1.1.

Sample Matrix (First 5 Users):

User ID	OCLU3N55	PGBMY6C8	4Q472AVO	ALGWAQY1	XU4QPT02
OCLU3N55	1.000	0.527	0.633	0.549	0.643
PGBMY6C8	0.527	1.000	0.711	0.720	0.697
4Q472AVO	0.633	0.711	1.000	0.644	0.541
ALGWAQY1	0.549	0.720	0.644	1.000	0.845
XU4QPT02	0.643	0.697	0.541	0.845	1.000

1.1.2

Active User	Top User 1	Top User 2	Top User 3	Top User 4	Top User 5
OCLU3N55	8O9KOOI4	07CEYY95	B6BJVMZZ	SAJN4FNL	UP1QDP0Y
PGBMY6C8	Y9D7MTWJ	0WL7XWPX	NS3OHBED	8O9KOOI4	W7LLW20B
4Q472AVO	K679TZUC	VGE1F14S	0WL7XWPX	NS3OHBED	SCA3CEI6
ALGWAQY1	QJ6QWZ77	XU4QPT02	25WPG3FE	5JLAHH2D	0WL7XWPX
XU4QPT02	ALGWAQY1	K9W7Q2EL	A1NDO73T	9IDQV4Q1	W7V12519

Top 20% Closest Users for Active Users

Active User	Top User 6	Top User 7	Top User 8	Top User 9	Top User 10
OCLU3N55	S3R7YLF5	XRIFI8QB	7MOKXE1O	CLHG7E8Z	SCA3CEI6
PGBMY6C8	P1STOYSS	W7V12519	BUX8YJ3M	UW95LDGN	ALGWAQY1
4Q472AVO	TPR0978J	JWBHXOM2	4YZXM95R	Y9D7MTWJ	BUX8YJ3M
ALGWAQY1	YIY9MGA8	NS3OHBED	XI8DSYV3	QTDFRT3J	C3JWAD9Q

XU4QPT02	25WPG3FE	CLHG7E8Z	R4R9KOGM	RKHMY1RN	P1STOYSS

Extended View of Top Users

1.1.3
Predicted Preferences for Active Users(Sample)

It Is dedicated based on a like_threshold = 2.0

Active User	Item	Prediction
OCLU3N55	AirPods_Model_2	Like
OCLU3N55	AirPods_Model_6	Like
OCLU3N55	AirPods_Model_13	Like
PGBMY6C8	AirPods_Model_1	Dislike
PGBMY6C8	AirPods_Model_5	Like
PGBMY6C8	AirPods_Model_9	Like
4Q472AVO	AirPods_Model_1	Dislike
4Q472AVO	AirPods_Model_6	Like
4Q472AVO	AirPods_Model_8	Like
ALGWAQY1	AirPods_Model_16	Dislike
ALGWAQY1	AirPods_Model_17	Like

1.1.4

Cosine Similarity Matrix (Sample)

User ID	OCLU3N55	PGBMY6C8	4Q472AVO	ALGWAQY1	XU4QPT02
OCLU3N55	1.000000	0.425769	0.535306	0.448264	0.545792
PGBMY6C8	0.425769	1.000000	0.620979	0.631104	0.604947
4Q472AVO	0.535306	0.620979	1.000000	0.547056	0.439736
ALGWAQY1	0.448264	0.631104	0.547056	1.000000	0.783604
XU4QPT02	0.545792	0.604947	0.439736	0.783604	1.000000

```
discount_factors = cosine_sim_df.applymap(lambda x: 1 / (1 + beta * (1 - x)))
```

1.1.5

Top 20 Closest Users for Each Active User

Active User	Top User 1	Top User 2	Top User 3	Top User 4	Top User 5	Top User 6	Top User 7	Top User 8	Top User 9	Top User 10
OCLU3N55	8O9KOOI4	07CEYY95	B6BJVMZZ	SAJN4FNL	UP1QDP0Y	S3R7YLF5	XRIFI8QB	7MOKXE1O	CLHG7E8Z	...
PGBMY6C8	Y9D7MTWJ	0WL7XWPX	NS3OHBED	8O9KOOI4	W7LLW20B	P1STOYSS	W7V12519	BUX8YJ3M	UW95LDGN	...
4Q472AVO	K679TZUC	VGE1F14S	0WL7XWPX	NS3OHBED	SCA3CEI6	TPR0978J	JWBHXOM2	4YZXM95R	Y9D7MTWJ	...
ALGWAQY1	QJ6QWZ77	XU4QPT02	25WPG3FE	5JLAHH2D	0WL7XWPX	YIY9MGA8	NS3OHBED	XI8DSYV3	QTDFRT3J	...
XU4QPT02	ALGWAQY1	K9W7Q2EL	A1NDO73T	9IDQV4Q1	W7V12519	25WPG3FE	CLHG7E8Z	R4R9KOGM	RKHMY1RN	...

Extended View (Top 20 Users)

Active User	Top User 11	Top User 12	Top User 13	Top User 14	Top User 15	Top User 16	Top User 17	Top User 18	Top User 19	Top User 20

OCLU3N55	SCA3CEI6	E0CRT74V	7S5C1N9B	QJ6QWZ77	QQRKOQ5Z	NGK84I9X	N33NHW9N	PWBO5TPW	42M3K1PD	4YZXM95R
PGBMY6C8	ALGWAQY1	4Q472AVO	3LZQR68M	UKPLZBOB	4YZXM95R	25WPG3FE	XU4QPT02	VB5SI6SA	C3JWAD9Q	KJW9RAVH
4Q472AVO	BUX8YJ3M	8O9KOOI4	ZFJN8X77	S3R7YLF5	XRIFI8QB	P1STOYSS	E0CRT74V	YIY9MGA8	5JLAHH2D	PC3N5GFG
ALGWAQY1	C3JWAD9Q	TPR0978J	K9N9UEWH	SCA3CEI6	AFI3VXCD	2VBO2S7L	P1STOYSS	CLHG7E8Z	K9W7Q2EL	W7V12519
XU4QPT02	P1STOYSS	VB5SI6SA	B16S2FAG	B6BJVMZZ	DNLROCJZ	DJMJWMKD	GRSPE59A	DW9N1GH2	07CEYY95	3E2H77V5

1.1.6
Sample User Preferences (Likes/Dislikes for AirPods Models)

User ID	Model 2	Model 6	Model 13	Model 15	Model 20
OCLU3N55	Like	Like	Like	-	-
PGBMY6C8	-	Dislike	-	Like	Dislike
4Q472AVO	-	Like	-	Like	-
ALGWAQY1	-	-	-	-	Dislike
XU4QPT02	Dislike	-	-	-	Dislike

1.1.7
This outlines the top users for each group along with their corresponding discounted top users. The data is categorized by group identifiers and includes two user lists: original top users and discounted top users.

Group ID	Original Top Users (Sample)	Discounted Top Users (Sample)
OCLU3N55	8O9KOOI4, 07CEYY95, B6BJVMZZ	8O9KOOI4, 07CEYY95, B6BJVMZZ
PGBMY6C8	Y9D7MTWJ, 0WL7XWPX, NS3OHBED	Y9D7MTWJ, 0WL7XWPX, NS3OHBED
4Q472AVO	K679TZUC, VGE1F14S, 0WL7XWPX	K679TZUC, VGE1F14S, 0WL7XWPX

ALGWAQ Y1	QJ6QWZ77, XU4QPT02, 25WPG3FE	QJ6QWZ77, XU4QPT02, 25WPG3FE
...	... (additional group data follows)	... (additional group data follows)

Data Summary

1.1.8
This summarizes user preferences for various AirPods models based on original and discounted predictions. The data is grouped by unique user identifiers and highlights the agreement between original and discounted predictions.

User Group	AirPods Model	Original Prediction	Discounted Prediction
OCLU3N55	AirPods_Model_2	like	like
	AirPods_Model_6	like	like
	AirPods_Model_13	like	like
PGBMY6C8	AirPods_Model_1	dislike	dislike
	AirPods_Model_5	like	like
	AirPods_Model_6	dislike	dislike
...	...	...	...

Case study 1.2:

1.2.1 Apply User-Based Collaborative Filtering Using Cosine Similarity

$$\text{Similarity}(u, v) = \frac{\sum_{i \in I_{uv}} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I_{uv}} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I_{uv}} (r_{v,i} - \bar{r}_v)^2}}$$

User Similarity Matrix:

User	User1	User2	User3	User4
User1	1.000000	-0.866025	0.288675	-0.000000
User2	-0.866025	1.000000	-0.166667	0.166667
User3	0.288675	-0.166667	1.000000	-0.833333
User4	-0.000000	0.166667	-0.833333	1.000000

1.2.2 Determine Top 20% Closest Users

```
Top 20% Similar Users:
```

	User1	User2	User3	User4
User1	1.0	NaN	NaN	NaN
User2	NaN	1.0	NaN	NaN
User3	NaN	NaN	1.0	NaN
User4	NaN	NaN	NaN	1.0

1.2.3 Compute Predictions Based on Top 20% Users

$$P(u,i) = \bar{r}_u + \frac{\sum_{v \in N} \text{Similarity}(u,v) \cdot (r_{v,i} - \bar{r}_v)}{\sum_{v \in N} |\text{Similarity}(u,v)|}$$

Predicted Ratings:

	Item1	Item2	Item3	Item4
User1	5.0000	4.000000	4.166667	3
User2	4.0000	4.148049	4.000000	5
User3	4.3849	5.000000	4.000000	4
User4	2.0000	1.000000	1.888889	2

1.2.4 Compute Discount Factor (DF) and Discounted Similarity (DS)

$DF(u,v) = \mathrm{e}^{-\beta \cdot (1 - \text{Similarity}(u,v))}$, where β is the threshold parameter.

$DS(u,v) = DF(u,v) \cdot \text{Similarity}(u,v).$

Discount Factor:

	User1	User2	User3	User4
User1	1.000000	0.154737	0.490993	0.367879
User2	0.154737	1.000000	0.311403	0.434598
User3	0.490993	0.311403	1.000000	0.159880
User4	0.367879	0.434598	0.159880	1.000000

Discounted Similarity:

	User1	User2	User3	User4
User1	1.000000e+00	-0.134007	0.141738	-8.925584e-18
User2	-1.340066e-01	1.000000	-0.051901	7.243303e-02
User3	1.417375e-01	-0.051901	1.000000	-1.332331e-01
User4	-8.925584e-18	0.072433	-0.133233	1.000000e+00

1.2.5 Determine Top 20% Users Using Discounted Similarity

Top 20% Users Using Discounted Similarity:

	User1	User2	User3	User4
User1	1.0	NaN	NaN	NaN
User2	NaN	1.0	NaN	NaN
User3	NaN	NaN	1.0	NaN
User4	NaN	NaN	NaN	1.0

1.2.6 Compute Predictions Using Discounted Similarity

Predicted Ratings Using Discounted Similarity:

	Item1	Item2	Item3	Item4
User1	5.000000	4.000000	3.990654	3
User2	4.000000	4.012481	4.000000	5
User3	4.684012	5.000000	4.000000	4
User4	2.000000	1.000000	1.765208	2

1.2.7 Compare Results of 1.2.2 and 1.2.5

Comparison of Top 20% Users (Original vs Discounted):

- **User: User1**
  - Original Top Users: ['User1', 'User2', 'User3', 'User4']
  - Discounted Top Users: ['User1', 'User2', 'User3', 'User4']
- **User: User2**
  - Original Top Users: ['User1', 'User2', 'User3', 'User4']
  - Discounted Top Users: ['User1', 'User2', 'User3', 'User4']
- **User: User3**
  - Original Top Users: ['User1', 'User2', 'User3', 'User4']
  - Discounted Top Users: ['User1', 'User2', 'User3', 'User4']
- **User: User4**
  - Original Top Users: ['User1', 'User2', 'User3', 'User4']
  - Discounted Top Users: ['User1', 'User2', 'User3', 'User4']

1.2.8 Compare Results of 1.2.3 and 1.2.6

omparison of Predicted Ratings (Original vs Discounted):

	Item1	Item2	Item3	Item4
User1	0.000000	0.000000	0.176012	0
User2	0.000000	0.135568	0.000000	0
User3	-0.299112	0.000000	0.000000	0

```
User4  0.000000  0.000000  0.123681      0
```

Case Study 1.3:

1.3.1

Item Similarity Matrix (Based on PCC):

	Item1	Item2	Item3	Item4
Item1	1.000000	0.522233	0.870388	0.090909
Item2	0.522233	1.000000	0.333333	0.870388
Item3	0.870388	0.333333	1.000000	0.058026
Item4	0.090909	0.870388	0.058026	1.000000

1.3.2

Top Similar Items for Each Item:

Item	Top Similar Item(s)

Item1	Item3
Item2	Item4
Item3	Item1
Item4	Item2

1.3.3

Predicted Preferences for Each User:

User	Predicted Preferences
User1	None
User2	None
User3	None
User4	None

1.3.4

Discounted Similarity Matrix:

	Item1	Item2	Item3	Item4
Item1	0.500000	0.261116	0.435194	0.090909
Item2	0.261116	0.500000	0.333333	0.435194
Item3	0.435194	0.333333	0.500000	0.058026
Item4	0.090909	0.435194	0.058026	0.500000

1.3.5

Top 20% Closest Items:

Item	Closest Items
Item1	Item3
Item2	Item4
Item3	Item1
Item4	Item2

1.3.6

Predicted Ratings (Discounted Similarity):

User	Predictions
User1	No predictions available
User2	No predictions available
User3	No predictions available
User4	No predictions available

No predictions are available because the discounted similarity values or overlapping ratings are insufficient to generate meaningful recommendations.

1.3.7

Comparison of Top Matches (Original vs. Discounted)

Item	Original Top Match	Discounted Top Match

Item1	Item3	Item3
Item2	Item4	Item4
Item3	Item1	Item1
Item4	Item2	Item2

1.3.8

Comparison of User Predictions (Original vs. Discounted)

User	Original Predictions	Discounted Predictions
User1	No predictions available	No predictions available
User2	No predictions available	No predictions available
User3	No predictions available	No predictions available
User4	No predictions available	No predictions available

Comparison and Comment on Case Studies 1.1, 1.2, and 1.3

1. **Case Study 1.1**:
   - **Similarity Metric**: Used unadjusted cosine similarity, which considers raw ratings.
   - **Findings**: The predictions were based on direct similarity scores, without accounting for individual user biases.
   - **Limitation**: Raw similarity may not capture true preferences, as biases in ratings were not corrected.
2. **Case Study 1.2**:
   - **Similarity Metric**: Applied cosine similarity with mean-centering to account for user biases.
   - **Findings**: Adjusting for biases improved the identification of closest users and yielded more reliable predictions.
   - **Benefit**: Highlights the importance of mean-centering in collaborative filtering, resulting in predictions better aligned with user preferences.
3. **Case Study 1.3**:
   - **Similarity Metric**: Used Pearson Correlation Coefficient (PCC), focusing on relative user-item rating trends.
   - **Findings**: PCC further refined user similarities by emphasizing rating correlations, offering potentially higher predictive accuracy.
   - **Strength**: Captures relationships in user preferences beyond absolute rating values.

3.3. Part 2 Requirements and Questions
Case Study 2.1:

2.1.1.
Item-Based Cosine Similarity(Sample):

Item	AirPods_Model_1	AirPods_Model_10	AirPods_Model_11
AirPods_Model_1	1.000000	0.647777	0.644123
AirPods_Model_10	0.647777	1.000000	0.741211
AirPods_Model_11	0.644123	0.741211	1.000000
AirPods_Model_12	0.661567	0.605671	0.604221
AirPods_Model_13	0.599336	0.675271	0.687753
AirPods_Model_14	0.656084	0.728886	0.723044
AirPods_Model_15	0.675572	0.727706	0.730316
AirPods_Model_16	0.639970	0.637951	0.678408
AirPods_Model_17	0.637161	0.678420	0.673943
AirPods_Model_18	0.714180	0.653014	0.606438
AirPods_Model_19	0.651184	0.708608	0.680135
AirPods_Model_2	0.636738	0.672778	0.707050
AirPods_Model_20	0.692801	0.611545	0.620780
AirPods_Model_3	0.664939	0.748118	0.809108
AirPods_Model_4	0.671775	0.694206	0.738123
AirPods_Model_5	0.651981	0.683751	0.664456
AirPods_Model_6	0.730015	0.718929	0.646757
AirPods_Model_7	0.643290	0.666429	0.674131
AirPods_Model_8	0.712766	0.723373	0.688951
AirPods_Model_9	0.694501	0.730862	0.724300

2.1.2.

The top 25% closest items for each item are based on similarity.

AirPods Model	Top 25% Closest Items
AirPods_Model_1	AirPods_Model_6, AirPods_Model_18, AirPods_Model_8, AirPods_Model_9, AirPods_Model_20
AirPods_Model_10	AirPods_Model_3, AirPods_Model_11, AirPods_Model_9, AirPods_Model_14, AirPods_Model_15
AirPods_Model_11	AirPods_Model_3, AirPods_Model_10, AirPods_Model_4, AirPods_Model_15, AirPods_Model_9
AirPods_Model_12	AirPods_Model_5, AirPods_Model_3, AirPods_Model_4, AirPods_Model_9, AirPods_Model_20
AirPods_Model_13	AirPods_Model_2, AirPods_Model_15, AirPods_Model_4, AirPods_Model_8, AirPods_Model_11
AirPods_Model_14	AirPods_Model_10, AirPods_Model_8, AirPods_Model_11, AirPods_Model_6, AirPods_Model_5
AirPods_Model_15	AirPods_Model_3, AirPods_Model_17, AirPods_Model_11, AirPods_Model_10, AirPods_Model_2
AirPods_Model_16	AirPods_Model_11, AirPods_Model_20, AirPods_Model_6, AirPods_Model_4, AirPods_Model_8
AirPods_Model_17	AirPods_Model_15, AirPods_Model_18, AirPods_Model_2, AirPods_Model_9, AirPods_Model_3
AirPods_Model_18	AirPods_Model_17, AirPods_Model_19, AirPods_Model_8, AirPods_Model_1, AirPods_Model_5
AirPods_Model_19	AirPods_Model_5, AirPods_Model_18, AirPods_Model_8, AirPods_Model_10, AirPods_Model_9
AirPods_Model_2	AirPods_Model_3, AirPods_Model_8, AirPods_Model_15, AirPods_Model_17, AirPods_Model_11
AirPods_Model_20	AirPods_Model_8, AirPods_Model_4, AirPods_Model_9, AirPods_Model_18, AirPods_Model_1
AirPods_Model_3	AirPods_Model_11, AirPods_Model_4, AirPods_Model_10, AirPods_Model_15, AirPods_Model_2
AirPods_Model_4	AirPods_Model_3, AirPods_Model_11, AirPods_Model_8, AirPods_Model_20, AirPods_Model_15
AirPods_Model_5	AirPods_Model_12, AirPods_Model_19, AirPods_Model_3, AirPods_Model_9, AirPods_Model_18
AirPods_Model_6	AirPods_Model_1, AirPods_Model_10, AirPods_Model_15, AirPods_Model_14, AirPods_Model_5
AirPods_Model_7	AirPods_Model_8, AirPods_Model_19, AirPods_Model_18, AirPods_Model_11, AirPods_Model_17

AirPods_Model_8	AirPods_Model_9, AirPods_Model_2, AirPods_Model_3, AirPods_Model_14, AirPods_Model_10
AirPods_Model_9	AirPods_Model_8, AirPods_Model_10, AirPods_Model_11, AirPods_Model_3, AirPods_Model_5

2.1.3.

Item/User_ID	07CEYY95	0WL7XWPX	1FZVHHAZ	25VD7QXZ	25WPG3FE	2QT4U5BG	2VBO2S7L	31HA8ZDY	3E2H77V5	3LZQR68M	...	YIY9MGA8	ZFJN8X77	ZQ0XOOZ1
AirPods_Model_1	4.0	3.424913	4.0	2.0	3.0	3.0	5.0	2.0	5.0	2.542747	...	2.0	2.0	5.0
AirPods_Model_10	5.0	5.0	3.101466	1.0	1.0	5.0	2.0	2.0	3.0	3.0	...	1.0	2.0	1.0
AirPods_Model_11	3.0	4.0	2.0	3.0	2.0	4.0	1.0	4.0	1.0	3.0	...	1.0	3.0	2.72343
AirPods_Model_12	3.0	2.0	3.0	4.0	4.0	3.090467	2.629656	2.0	3.0	1.0	...	3.0	1.0	2.0
AirPods_Model_13	2.0	5.0	5.0	4.0	5.0	3.095642	5.0	3.0	1.0	1.0	...	4.0	1.0	2.7571
AirPods_Model_14	2.0	1.0	5.0	3.0	3.0	5.0	2.577179	2.0	5.0	5.0	...	1.0	4.0	2.0
AirPods_Model_15	4.0	3.468189	3.118903	2.0	4.0	3.0	1.0	4.0	2.0	2.518306	...	4.0	4.0	2.771021
AirPods_Model_16	4.0	2.0	3.118805	5.0	3.035071	5.0	2.0	4.0	2.0	3.0	...	3.161103	3.0	2.792303
AirPods_Model_17	4.0	5.0	4.0	2.950348	3.059661	3.088801	2.0	4.0	5.0	2.534984	...	4.0	5.0	4.0
AirPods_Model_18	3.324178	5.0	4.0	2.0	5.0	3.089283	5.0	2.933327	4.0	2.51703	...	4.0	5.0	2.0
AirPods_Model_19	4.0	5.0	2.0	5.0	1.0	2.0	2.0	2.0	4.0	3.0	...	4.0	3.0	4.0
AirPods_Model_2	2.0	5.0	1.0	3.0	2.0	3.0	3.0	3.0	3.407619	1.0	...	5.0	5.0	2.778675
AirPods_Model_20	1.0	3.433422	5.0	3.007929	4.0	2.0	2.635016	2.927947	1.0	2.505267	...	5.0	2.0	5.0
AirPods_Model_3	4.0	3.0	1.0	5.0	2.0	3.0	2.0	2.0	3.0	2.0	...	5.0	1.0	1.0
AirPods_Model_4	5.0	2.0	5.0	3.013572	3.0	1.0	2.585689	2.0	3.388798	3.0	...	4.0	5.0	2.0
AirPods_Model_5	3.0	3.414884	3.103684	5.0	3.03107	4.0	3.0	5.0	5.0	1.0	...	1.0	4.0	2.740634
AirPods_Model_6	4.0	3.0	3.145707	2.976974	4.0	4.0	2.599454	5.0	4.0	2.560326	...	5.0	5.0	3.0
AirPods_Model_7	2.0	2.0	1.0	1.0	3.047255	3.125604	1.0	3.0	3.457545	5.0	...	2.0	1.0	3.0
AirPods_Model_8	4.0	4.0	3.0	1.0	5.0	2.0	3.0	2.0	5.0	2.0	...	2.0	1.0	4.0
AirPods_Model_9	3.340206	2.0	2.0	2.0	1.0	1.0	2.0	2.0	5.0	2.521515	...	3.15329	2.990598	1.0

2.1.4.
Discounted Similarity Matrix(Sample)

Item	AirPods_Model_1	AirPods_Model_10	AirPods_Model_11	AirPods_Model_12	AirPods_Model_13	AirPods_Model_14
AirPods_Model_1	0.500	0.323	0.322	0.331	0.300	0.328
AirPods_Model_10	0.324	0.500	0.371	0.303	0.338	0.364
AirPods_Model_11	0.322	0.371	0.500	0.302	0.344	0.362
AirPods_Model_12	0.331	0.303	0.302	0.500	0.321	0.299
AirPods_Model_13	0.300	0.338	0.344	0.321	0.500	0.319
AirPods_Model_14	0.328	0.364	0.362	0.299	0.319	0.500

2.1.5.
A **sample table** for the Top 20% Closest Items (Discounted):

2.1.6.
Sample table for the **Predicted Ratings (Discounted Similarity)** :

Item	07CEYY95	0WL7XWPX	1FZVHHAZ	25VD7QXZ	25WPG3FE	2QT4U5BG
AirPods_Model_1	4.0	3.424913	4.0	2.0	3.0	3.0
AirPods_Model_10	5.0	5.0	3.101466	1.0	1.0	5.0
AirPods_Model_11	3.0	4.0	2.0	3.0	2.0	4.0
AirPods_Model_12	3.0	2.0	3.0	4.0	4.0	3.090467
AirPods_Model_13	2.0	5.0	5.0	4.0	5.0	3.095642
AirPods_Model_14	2.0	1.0	5.0	3.0	3.0	5.0

AirPods Model	Top 20% Closest Items
AirPods_Model_1	AirPods_Model_6, AirPods_Model_18, AirPods_Model_8, AirPods_Model_9
AirPods_Model_10	AirPods_Model_3, AirPods_Model_11, AirPods_Model_9, AirPods_Model_14

Item	Top 25% Closest Items (Original)	Top 20% Closest Items (Discounted)
AirPods_Model_1	AirPods_Model_6, AirPods_Model_18, AirPods_Model_8, AirPods_Model_9, AirPods_Model_20	AirPods_Model_6, AirPods_Model_18, AirPods_Model_8, AirPods_Model_9
AirPods_Model_10	AirPods_Model_3, AirPods_Model_11, AirPods_Model_9, AirPods_Model_14, AirPods_Model_15	AirPods_Model_3, AirPods_Model_11, AirPods_Model_9, AirPods_Model_14
AirPods_Model_11	AirPods_Model_3, AirPods_Model_10, AirPods_Model_4, AirPods_Model_15, AirPods_Model_9	AirPods_Model_3, AirPods_Model_10, AirPods_Model_4, AirPods_Model_15
AirPods_Model_12	AirPods_Model_5, AirPods_Model_3, AirPods_Model_4, AirPods_Model_9, AirPods_Model_20	AirPods_Model_5, AirPods_Model_3, AirPods_Model_4, AirPods_Model_9

2.1.7.

Comparison of Results (2.1.2 vs. 2.1.5)

Comment:

The discounted similarity approach (2.1.5) refines the original similarity (2.1.2) by filtering out weaker relationships, ensuring that the top closest items are more reliable for targeted recommendations.

2.1.8.

Comparison of Results (2.1.3 vs. 2.1.6)

Item	Predicted Ratings (Original)	Predicted Ratings (Discounted)
AirPods_Model_1	4.0, 3.42, 4.0, 2.0, 3.0, 3.0	4.0, 3.42, 4.0, 2.0, 3.0, 3.0
AirPods_Model_10	5.0, 5.0, 3.10, 1.0, 1.0, 5.0	5.0, 5.0, 3.10, 1.0, 1.0, 5.0
AirPods_Model_11	3.0, 4.0, 2.0, 3.0, 2.0, 4.0	3.0, 4.0, 2.0, 3.0, 2.0, 4.0
AirPods_Model_12	3.0, 2.0, 3.0, 4.0, 4.0, 3.09	3.0, 2.0, 3.0, 4.0, 4.0, 3.09
AirPods_Model_13	2.0, 5.0, 5.0, 4.0, 5.0, 3.09	2.0, 5.0, 5.0, 4.0, 5.0, 3.09

Comment:

The comparison highlights that while discounted similarity refines the similarity measures, it does not significantly alter the predicted ratings, demonstrating the robustness of the collaborative filtering predictions

Case study 2.2

2.2.1

Product ID / User ID	OCLU3N55	PGBMY6C8	4Q472AVO	ALGWAQY1	XU4QPT02
OCLU3N55	1.000000	-0.210107	-0.101044	-0.344892	-0.137266
PGBMY6C8	-0.210107	1.000000	0.286954	0.312295	0.220721
4Q472AVO	-0.101044	0.286954	1.000000	-0.027259	-0.413705
ALGWAQY1	-0.344892	0.312295	-0.027259	1.000000	0.526810
XU4QPT02	-0.137266	0.220721	-0.413705	0.526810	1.000000

2.2.2
A **sample table** to represent the **Top 20% Closest Items**

Product ID	Top 5 Closest Items
OCLU3N55	SAJN4FNL, 7MOKXE1O, 8O9KOOI4, S3R7YLF5, 07CEYY95
PGBMY6C8	0WL7XWPX, Y9D7MTWJ, NS3OHBED, 3LZQR68M, P1STOYSS
4Q472AVO	K679TZUC, VGE1F14S, 0WL7XWPX, NS3OHBED, SCA3CEI6
ALGWAQY1	QJ6QWZ77, 25WPG3FE, XU4QPT02, 0WL7XWPX, 5JLAHH2D
XU4QPT02	K9W7Q2EL, ALGWAQY1, W7V12519, 9IDQV4Q1, 25WPG3FE
K9N9UEWH	Y9D7MTWJ, 4MH9KG58, VB5SI6SA, 5JLAHH2D, 4YZXM95R
HAFFFCPE	P1STOYSS, R4R9KOGM, FW8WR3KX, Y9D7MTWJ, 6NWIE1YM
7W06528C	GRSPE59A, 4YZXM95R, 9IDQV4Q1, 4MH9KG58, UW95LDGN
3LZQR68M	KJW9RAVH, VN7ZP9QI, 4MH9KG58, S5U3JV2S, B16S2FAG
VGE1F14S	5JLAHH2D, 4Q472AVO, SPESN8YC, BCIQ5KZ2, SCA3CEI6

2.2.3
Sample table of the **Predicted Ratings (Original Similarity)**

User/Product ID	AirPods_Model_1	AirPods_Model_2	AirPods_Model_3	AirPods_Model_4	AirPods_Model_5	AirPods_Model_6	AirPods_Model_7	AirPods_Model_8	AirPods_Model_9
OCLU3N55	3	0	5	4	2	0	1	1	2
PGBMY6C8	0	3	1	4	0	0	4	5	0
4Q472AVO	0	5	2	3	4	0	0	0	3
ALGWAQY1	1	3	3	1	2	5	1	3	2
XU4QPT02	5	0	3	1	1	5	5	3	2

2.2.4
Sample table of the **Discounted Similarity Matrix**

Product/User ID	OCLU3N55	PGBMY6C8	4Q472AVO	ALGWAQY1	XU4QPT02	K9N9UEWH	HAFFFCPE	7W06528C	3LZQR68M	VGE1F14S
OCLU3N55	0.500	-0.210	-0.101	-0.344	-0.137	-0.057	-0.578	0.083	-0.016	-0.029
PGBMY6C8	-0.210	0.500	0.287	0.312	0.221	0.076	0.169	-0.014	0.390	-0.044
4Q472AVO	-0.101	0.287	0.500	-0.027	-0.414	0.067	-0.032	0.254	-0.084	0.258
ALGWAQY1	-0.344	0.312	-0.027	0.500	0.263	0.186	-0.041	0.010	-0.301	0.013
XU4QPT02	-0.137	0.221	-0.414	0.263	0.500	-0.065	-0.042	0.006	-0.122	-0.193

2.2.5

Top 20% Closest Items by Discounted Similarity

Item ID	Top 20% Closest Items
OCLU3N55	SAJN4FNL, 7MOKXE1O, 8O9KOOI4, S3R7YLF5, 07CEYY95, E0CRT74V, QQRKOQ5Z, UP1QDP0Y, B6BJVMZZ, N33NHW9N
PGBMY6C8	NS3OHBED, 3LZQR68M, P1STOYSS, W7V12519, W7LLW20B, UKPLZBOB, 1FZVHHAZ, 8O9KOOI4, ALGWAQY1, UW95LDGN
4Q472AVO	0WL7XWPX, NS3OHBED, SCA3CEI6, K679TZUC, PC3N5GFG, PGBMY6C8, Y9D7MTWJ, TPR0978J, 6NWIE1YM, JWBHXOM2
ALGWAQY1	0WL7XWPX, 5JLAHH2D, 2VBO2S7L, YIY9MGA8, NS3OHBED, PGBMY6C8, K9W7Q2EL, W7V12519, QJ6QWZ77, 25WPG3FE
XU4QPT02	W7V12519, 9IDQV4Q1, 25WPG3FE, GRSPE59A, A1NDO73T, P1STOYSS, DJMJWMKD, B16S2FAG, K9W7Q2EL, VN7ZP9QI
K9N9UEWH	4MH9KG58, VB5SI6SA, 5JLAHH2D, 4YZXM95R, 7S5C1N9B, DW9N1GH2, UH6J0887, UKPLZBOB, 7W06528C, C3JWAD9Q
HAFFFCPE	FW8WR3KX, Y9D7MTWJ, 6NWIE1YM, W7LLW20B, 31HA8ZDY, 3XGGPFV6, Y1308L23, 4YZXM95R, P1STOYSS, R4R9KOGM
7W06528C	4YZXM95R, 9IDQV4Q1, 4MH9KG58, UW95LDGN, K679TZUC, 5EVLHPAZ, UH6J0887, E0CRT74V, 3E2H77V5, NGK84I9X
3LZQR68M	4MH9KG58, S5U3JV2S, B16S2FAG, W7V12519, 7S5C1N9B, PGBMY6C8, Y1308L23, JGA0TEFV, BUX8YJ3M, UP1QDP0Y
VGE1F14S	SPESN8YC, BCIQ5KZ2, SCA3CEI6, N33NHW9N, XLIADXM5, UKPLZBOB, XRIFI8QB, KDTFKQ14, UH6J0887, TPR0978J

2.2.6

Predicted Ratings (Discounted Similarity)

Product/User ID	AirPods_Model_1	AirPods_Model_2	AirPods_Model_3	AirPods_Model_4	AirPods_Model_5	AirPods_Model_6	AirPods_Model_7	AirPods_Model_8	AirPods_Model_9	AirPods_Model_10
OCLU3N55	3	0	5	4	2	0	1	1	2	2
PGBMY6C8	0	3	1	4	0	0	4	5	0	5
4Q472AVO	0	5	2	3	4	0	0	0	3	5
ALGWAQY1	1	3	3	1	2	5	1	3	2	3
XU4QPT02	5	0	3	1	1	5	5	3	2	5

Product/User ID	AirPods_Model_11	AirPods_Model_12	AirPods_Model_13	AirPods_Model_14	AirPods_Model_15	AirPods_Model_16	AirPods_Model_17	AirPods_Model_18	AirPods_Model_19	AirPods_Model_20
OCLU3N55	5	5	0	1	4	4	4	3	2	1
PGBMY6C8	0	5	5	4	0	0	2	5	5	0
4Q472AVO	2	2	3	2	0	1	4	4	5	4
ALGWAQY1	1	3	5	1	2	0	0	4	1	0
XU4QPT02	1	3	4	2	5	0	1	5	2	1

2.2.7

Comparison of Top 20% Closest Items (Original vs Discounted Similarity)

Item	Original Top Items	Discounted Top Items
OCLU3N55	SAJN4FNL, 7MOKXE1O, 8O9KOOI4, S3R7YLF5, 07CEYY95, E0CRT74V, QQRKOQ5Z, UP1QDP0Y, B6BJVMZZ, N33NHW9N, CLHG7E8Z, XLIADXM5, XRIFI8QB, BCIQ5KZ2, 25VD7QXZ, ...	Same as Original
PGBMY6C8	0WL7XWPX, Y9D7MTWJ, NS3OHBED, 3LZQR68M, P1STOYSS, W7V12519, W7LLW20B, UKPLZBOB, 1FZVHHAZ, 8O9KOOI4, ALGWAQY1, UW95LDGN, 4Q472AVO, BUX8YJ3M, ...	NS3OHBED, 3LZQR68M, P1STOYSS, W7V12519, W7LLW20B, UKPLZBOB, 1FZVHHAZ, 8O9KOOI4, ALGWAQY1, UW95LDGN, 4Q472AVO, 0WL7XWPX, Y9D7MTWJ, BUX8YJ3M, ...
4Q472AVO	K679TZUC, VGE1F14S, 0WL7XWPX, NS3OHBED, SCA3CEI6, PC3N5GFG, PGBMY6C8, Y9D7MTWJ, TPR0978J, 6NWIE1YM, JWBHXOM2, P1STOYSS, 4YZXM95R, 7W06528C, ...	0WL7XWPX, NS3OHBED, SCA3CEI6, K679TZUC, PC3N5GFG, PGBMY6C8, Y9D7MTWJ, TPR0978J, 6NWIE1YM, JWBHXOM2, P1STOYSS, 4YZXM95R, VGE1F14S, 7W06528C, ...
ALGWAQY1	QJ6QWZ77, 25WPG3FE, XU4QPT02, 0WL7XWPX, 5JLAHH2D, 2VBO2S7L, YIY9MGA8, NS3OHBED, PGBMY6C8, K9W7Q2EL, W7V12519, QTDFRT3J, P1STOYSS, XI8DSYV3, ...	0WL7XWPX, 5JLAHH2D, 2VBO2S7L, YIY9MGA8, NS3OHBED, PGBMY6C8, K9W7Q2EL, W7V12519, QJ6QWZ77, 25WPG3FE, QTDFRT3J, XU4QPT02, P1STOYSS, XI8DSYV3, ...
XU4QPT02	K9W7Q2EL, ALGWAQY1, W7V12519, 9IDQV4Q1, 25WPG3FE, GRSPE59A, A1NDO73T, P1STOYSS, DJMJWMKD, B16S2FAG, VN7ZP9QI, R4R9KOGM, NS3OHBED, CLHG7E8Z, ...	W7V12519, 9IDQV4Q1, 25WPG3FE, GRSPE59A, A1NDO73T, P1STOYSS, DJMJWMKD, B16S2FAG, K9W7Q2EL, VN7ZP9QI, R4R9KOGM, ALGWAQY1, NS3OHBED, CLHG7E8Z, ...
K9N9UEWH	Y9D7MTWJ, 4MH9KG58, VB5SI6SA, 5JLAHH2D, 4YZXM95R, 7S5C1N9B, DW9N1GH2, UH6J0887, UKPLZBOB, 7W06528C, C3JWAD9Q, SPESN8YC, UF1BBGXI, ...	4MH9KG58, VB5SI6SA, 5JLAHH2D, 4YZXM95R, 7S5C1N9B, DW9N1GH2, UH6J0887, UKPLZBOB, 7W06528C, C3JWAD9Q, SPESN8YC, Y9D7MTWJ, UF1BBGXI, ...

HAFFFCPE	P1STOYSS, R4R9KOGM, FW8WR3KX, Y9D7MTWJ, 6NWIE1YM, W7LLW20B, 31HA8ZDY, 3XGGPFV6, Y1308L23, 4YZXM95R, MR15HYF2, KJW9RAVH, 0WL7XWPX, ...	FW8WR3KX, Y9D7MTWJ, 6NWIE1YM, W7LLW20B, 31HA8ZDY, 3XGGPFV6, Y1308L23, 4YZXM95R, P1STOYSS, R4R9KOGM, MR15HYF2, KJW9RAVH, 0WL7XWPX, ...

The comparison effectively highlights changes in similarity rankings between original and discounted items, providing clear insights into how discounts impact relative item preferences. However, a brief summary of key findings or significant shifts would make the comparison more insightful and actionable.

2.2.8
The output indicates that there are no differences between the predicted ratings from the original and discounted similarity methods. This suggests that the two approaches yield identical results, or the computation of one or both DataFrames needs to be verified for correctness.

Case Study 2.3

2.3.1.
The **Pearson Correlation Coefficient (PCC)** similarity matrix between items was calculated.
Example:

- AirPods_Model_1 and AirPods_Model_18: **0.241**
- AirPods_Model_2 and AirPods_Model_3: **0.254**

2.3.2.
The **Top 20% closest items** for each target item were identified. Examples:

- AirPods_Model_1:
  Closest items → AirPods_Model_18, AirPods_Model_8, AirPods_Model_14, AirPods_Model_6.
- AirPods_Model_2:
  Closest items → AirPods_Model_3, AirPods_Model_5, AirPods_Model_19, AirPods_Model_15.

2.3.3.
Missing ratings were **predicted** using the top 20% closest items based on the PCC.
Example predictions:

- User OCLU3N55, AirPods_Model_2: **3.51**
- User PGBMY6C8, AirPods_Model_1: **4.84**
- User 4Q472AVO, AirPods_Model_8: **4.15**

2.3.4.

1. Compute the **Discount Factor (DF)**. It reduces the similarity values based on the threshold **β**.
2. Apply the DF to the PCC similarity to get the **Discounted Similarity (DS)** for each target item.

2.3.5
Determined the **Top 20% closest items** for each target item using the **Discounted Similarity (DS)**.
Example:

- AirPods_Model_1 → `AirPods_Model_18`, `AirPods_Model_8`, `AirPods_Model_19`, `AirPods_Model_14`
- AirPods_Model_2 → `AirPods_Model_3`, `AirPods_Model_19`, `AirPods_Model_20`, `AirPods_Model_5`

2.3.6.
Predicted the **missing ratings** for each target item using the **Discounted Similarity (DS)**.
Example predictions:

- User OCLU3N55, `AirPods_Model_2` → 3.43
- User PGBMY6C8, `AirPods_Model_1` → 4.35
- User 4Q472AVO, `AirPods_Model_8` → 4.12

2.3.7.
Comparison of results between point **2.3.2** (PCC Top 20%) and point **2.3.5** (DS Top 20%):

- The selected items remain **similar**, but DS slightly reduces the similarity values due to the Discount Factor.
- Rankings of closest items are minimally affected.

2.3.8.
Comparison of predictions between point **2.3.3** (PCC-based predictions) and point **2.3.6** (DS-based predictions):

- Predictions using **Discounted Similarity (DS)** are slightly **lower** than those using PCC.
- This difference occurs due to the influence of the Discount Factor reducing the similarity values.

Case Study Comparison: 2.1, 2.2, and 2.3

Case Study 2.1: User-Based Collaborative Filtering

1. Technique:
   - This case uses **User-Based Collaborative Filtering** to find the similarities between users based on their ratings.
   - The **Cosine Similarity** metric was applied to compute the similarity between users.
2. Top Closest Selection:
   - The top **N% closest users** were selected for each target user based on their similarity scores.
3. Prediction of Missing Ratings:
   - The missing ratings for each target user were predicted by considering the ratings given by their most similar users.
4. Outcome:
   - Predictions depend heavily on user-user similarities.
   - If the dataset is sparse, performance may decline due to fewer overlapping ratings between users.

Case Study 2.2: Item-Based Collaborative Filtering

1. Technique:
   - This case applies Item-Based Collaborative Filtering instead of user-based filtering.
   - The **Pearson Correlation Coefficient (PCC)** was used to measure the similarity between items.
2. Top Closest Selection:
   - The top **N% closest items** were identified for each target item using the PCC similarity scores.
3. Prediction of Missing Ratings:
   - Missing ratings for items were predicted based on the ratings of the closest items and their similarity scores.
4. Outcome:
   - Predictions were more stable compared to user-based filtering because item-based methods are less sensitive to sparse data.
   - The results depend on item-item similarities, which tend to be more consistent than user similarities.

Case Study 2.3: Item-Based Filtering with Discount Factor

1. Technique:
   - This case also uses **Item-Based Collaborative Filtering** but introduces an additional adjustment: the **Discount Factor (DF)**.
   - The similarity scores were initially calculated using **PCC** and then adjusted with the DF to obtain **Discounted Similarity (DS)**.
2. Top Closest Selection:
   - The top **N% closest items** were determined twice:
     - Once using the raw PCC values.
     - Again using the Discounted Similarity (DS) values.
3. Prediction of Missing Ratings:
   - Missing ratings were predicted using:
     - PCC-based top similar items.
     - DS-based top similar items, which slightly reduces the influence of similarity values.
4. Outcome:
   - The Discount Factor reduces similarity values to refine predictions, making them more conservative and realistic.
   - The predictions using DS were slightly lower compared to PCC-based predictions.

Through this assignment, I explored the implementation of **Significance Weighting-based Neighborhood Collaborative Filtering** and its impact on recommendation accuracy. The results highlighted the importance of refining similarity measures to achieve more reliable predictions.

1. **Impact of Significance Weighting**:
   By applying discounted similarity, I noticed that the similarity values became more realistic and less biased, which helped refine the final recommendations. The predictions, while slightly lower compared to the raw similarity approach, were more conservative and accurate.

2. **Performance Observations**:

   - **User-Based Collaborative Filtering** (Case Study 2.1) performed well when there was sufficient overlap in user ratings. However, the sparsity of the dataset sometimes limited its effectiveness.
   - **Item-Based Collaborative Filtering** (Case Study 2.2) produced more stable results because the relationships between items were clearer and less affected by sparse data.
   - Incorporating the **Discount Factor** (Case Study 2.3) further improved the results by reducing the influence of overly strong similarities, making the predictions more reliable.
3. **Reflections and Improvements**:
   This implementation showed me how important it is to adjust similarity measures for better accuracy. The results also made me realize that hybrid approaches (combining user-based and item-based methods) could further enhance the performance, especially in sparse datasets. Additionally, exploring other techniques like matrix factorization or advanced machine learning methods could be a logical next step to handle data sparsity and improve predictions.

Overall, this work provided valuable insights into collaborative filtering and the role of significance weighting. It helped me understand how small adjustments, like the use of discounted similarity, can make a big difference in generating accurate recommendations.