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

**‘’Assignment #2: Significance Weighting-based Neighbourhood CF Filters.’’**

**‘’ Name: Mohamed Mosad Fawzy /ID: 221100611 program:AIS‘’**

**Table of content:**

**Introduction …………………………………………………………………….2**

**Explain the Dataset ……………………………..……….…….….…………..2**

**Outcomes of section 3.1 ……………………………………………….…….2**

**Part 1and2 requirements and questions outputs .……………………….4**

**Summary of the Comparison of part 1 and 2……………………………..15**

**General Comments and Summary……………………..………………...…16**

**Enhancement ……………………………………………………………….….17**

**Equation used in codes …………………………….………………….…….18**

**Conclusion…………………………………………….………………….……..18**

**Reference …………………………………………….………………………….19**

**Introduction:-**

The present study investigates advanced techniques for collaborative filtering to improve the fairness and reliability of recommender systems. Some subset of the dataset from Assignment 1 has been devised such that a square user-item matrix of sizes 100 matches with the given conditions. There were, therefore, three active users with varied counts of missing ratings and two target items of varying density of missing data. This arrangement allows an in-depth analysis of similarity measures and methods of prediction based on a few controlled parameters.

The analysis aims to assess the impact of different approaches, including original similarity measures, bias corrections, and significance weighting. These techniques were implemented on various case studies to evaluate their effect on prediction accuracy, the quality of top-N recommendations, and system robustness. These methods aim to investigate the best trade-off between prediction accuracy, computational efficiency, and practical applicability in recommender systems.

**Explain the Dataset:-**

I use the data from assignment 1 and manipulate it until I regenerate a user\_item matrix 100\*10 by this subset I achieve the conditions of Picking three users (U1, U2, U3) and consider them active users. One user with 2 missing ratings, another user with 3 missing ratings, and the other user with 5 missing ratings and Picking two items (l1 and I2) and consider them target items. One item had 4% missing ratings, and the other item had 10% missing ratings.

Until I fulfill the conditions.

**Outcomes of section 3.1 :-**

My dataset is already scaled (1 to 5) rating and I count the Total number of unique users (tnu): 100, Total number of unique items (tni): 10   
Ratings per product:

|  |
| --- |
| B07TDSJZMR 6 |
| B08637FWWF 9 |
| B07KJVGNN5 96 |
| B007HY7GC2 4 |
| B08KYJLF5T 11 |
| B09GBMG83Z 90 |
| B09FKT5PQ9 3 |
| B08THJD1MH 8 |
| B08FCQML37 3 |
| B08GYM3HVP 2 |

User with 2 missing ratings: ['AETE7Y3DZT6BLMWA6U27ADJDZ4LA']

User with 3 missing ratings: ['AEMJ2EG5ODOCYUTI54NBXZHDJGSQ']

User with 5 missing ratings: ['AFSKPY37N3C43SOI5IEXEK5JSIYA']

Item with 4% missing ratings: Index(['B07KJVGNN5']

Item with 10% missing ratings: Index(['B09GBMG83Z']

Sorted Co-Rating Statistics:

No\_common\_users No\_coRated\_items

0 100 8

1 100 7

2 91 5

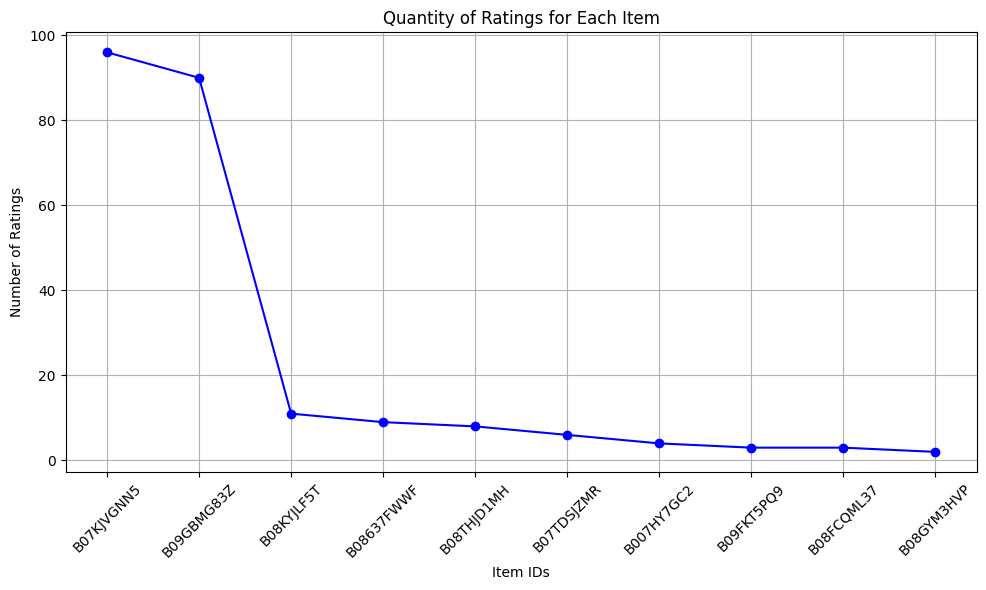
Sorted 2-D Array (No\_common\_users in descending order, No\_coRated\_items in corresponding order):

[[100 8]

[100 7]

[ 91 5]]

The curve that illustrates the quantity of ratings for every item in your dataset.

****

Thresholds (ß) for each active user:

User AETE7Y3DZT6BLMWA6U27ADJDZ4LA: ß = 25

User AEMJ2EG5ODOCYUTI54NBXZHDJGSQ: ß = 23

User AFSKPY37N3C43SOI5IEXEK5JSIYA: ß = 10

I saved this requirement in a text file called results.

**Part 1and2 requirements and questions outputs :-**

**Case 1.1**Cosine Similarity Matrix:

[[1. 0.80161343 0.57232262 ... 0.60955692 0.61300882 0.60955692]

[0.80161343 1. 0.48145555 ... 0.4330127 0.46948553 0.4330127 ]

[0.57232262 0.48145555 1. ... 0.37062466 0.41931393 0.37062466]

...

[0.60955692 0.4330127 0.37062466 ... 1. 0.98994949 1. ]

[0.61300882 0.46948553 0.41931393 ... 0.98994949 1. 0.98994949]

[0.60955692 0.4330127 0.37062466 ... 1. 0.98994949 1. ]]

Similarities of **AETE7Y3DZT6BLMWA6U27ADJDZ4LA** with all other users:

user\_id Similarity

|  |
| --- |
| AETE7Y3DZT6BLMWA6U27ADJDZ4LA 1.000000 |
| AEMJ2EG5ODOCYUTI54NBXZHDJGSQ 0.801613 |
| AE3TASYGLHHRHUJUDFTKFDMWFIYA 0.778142 |
| AFKZENTNBQ7A7V7UXW5JJI6UGRYQ 0.777854 |
| AG3ZLSFL6WEHCXA2SETWSPPDGTVQ 0.766613 |
| ... ... |
| AF7J5A2ME55LSCRVQWXB3BKE7CHQ 0.383131 |
| AGQSSZPF5DTU56OIYEZVVKTMKJIQ 0.383131 |
| AFAIJYOUO3NAWLBDIKTQSC3DASWA 0.383131 |
| AEVWAM3YWN5URJVJIZZ6XPD2MKIA 0.332261 |
| AFWVN52MRBWOTIK7UGXBWGOY4HBA 0.200083 |

[100 rows x 1 columns]

Similarities of **AEMJ2EG5ODOCYUTI54NBXZHDJGSQ** with all other users:

user\_id Similarity

|  |
| --- |
| AEMJ2EG5ODOCYUTI54NBXZHDJGSQ 1.000000 |
| AETE7Y3DZT6BLMWA6U27ADJDZ4LA 0.801613 |
| AFXDYTFXJLX4YG3UD7W23W7Z5Y4A 0.751041 |
| AG4COLOHCLHWUUYWCZ77TPJLE3EQ 0.720082 |
| AFCYUFW3NQ37UQXYVWL3LN4LAKLQ 0.715082 |
| ... ... |
| AGV4WWE4CB7LRPPV6PG7RCEFBT3A 0.102062 |
| AF7J5A2ME55LSCRVQWXB3BKE7CHQ 0.102062 |
| AGQSSZPF5DTU56OIYEZVVKTMKJIQ 0.102062 |
| AFAIJYOUO3NAWLBDIKTQSC3DASWA 0.102062 |
| AF4NIQPIZQ3S3G6KEVQW33DNHZSQ 0.102062 |

[100 rows x 1 columns]

Similarities of **AFSKPY37N3C43SOI5IEXEK5JSIYA** with all other users:

user\_id similarity

|  |
| --- |
| AFSKPY37N3C43SOI5IEXEK5JSIYA 1.000000 |
| AG6SNHIVEFMJIXCLZWKTGEOV5ZBQ 0.638690 |
| AELHB5QYXVSXZM263JIARJBWPOSA 0.620174 |
| AETE7Y3DZT6BLMWA6U27ADJDZ4LA 0.572323 |
| AFXF3EGQTQDXMRLDWFU7UBFQZB7Q 0.561179 |
| ... ... |
| AGC3Q6IXOVLTTDMS4Q55FPYUF6FQ 0.000000 |
| AHIL63TZRABKLCVM22CGMMKZDRXQ 0.000000 |
| AF7J5A2ME55LSCRVQWXB3BKE7CHQ 0.000000 |
| AGWGMDSRIXUSGG3AVYX65RVPWCLQ 0.000000 |
| AGQSSZPF5DTU56OIYEZVVKTMKJIQ 0.000000 |

[100 rows x 1 columns]

Top 20% closest users to **AETE7Y3DZT6BLMWA6U27ADJDZ4LA**:

user\_id Similarity

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | AEMJ2EG5ODOCYUTI54NBXZHDJGSQ 0.801613 | | AE3TASYGLHHRHUJUDFTKFDMWFIYA 0.778142 | | AFKZENTNBQ7A7V7UXW5JJI6UGRYQ 0.777854 | | AG3ZLSFL6WEHCXA2SETWSPPDGTVQ 0.766613 | | AGBOGNMAG3UCA6B4Z6VT62DVNOIA 0.753056 | | AFXF3EGQTQDXMRLDWFU7UBFQZB7Q 0.730286 | | AHVWKBFOVXQV6653O4XGCKOB6YVA 0.679861 | | AFD6OGB6AY4YPKN62LCTXGYR7KJA 0.666945 | | AEKGBJHGLLCKCRL3KMFZ4JUKVNSQ 0.664523 | | AG2UEDPK43QC5AU6HZTCBANPW2FQ 0.635350 | | AEQAYV7RXZEBXMQIQPL6KCT2CFWQ 0.625650 | | AEZGPLOYTSAPR3DHZKKXEFPAXUAA 0.621649 | | AEHWUHNTB5FX32HJ7UBOZ2WWUX3Q 0.613308 | | AHGAOIZVODNHYMNCBV4DECZH42UQ 0.613185 | | AFWTME2ROGUQI5J5FB3DWCLKZNBA 0.613009 |   AFQ7WYW4KSH4VI5OVXCP2GV6PBRA 0.613009 |
| AHY7ZJB523OPTIKXRI63PS2V6FSQ 0.611002 |
| AEYGPUCRKH7G4VM22FM3VAKSQ23Q 0.609557 |
| AF2T7NQ5S7SVEBHC7VGY23KZGQDA 0.609557 |

Top 20% closest users to **AEMJ2EG5ODOCYUTI54NBXZHDJGSQ:**

user\_id Similarity

|  |
| --- |
| AETE7Y3DZT6BLMWA6U27ADJDZ4LA 0.801613 |
| AFXDYTFXJLX4YG3UD7W23W7Z5Y4A 0.751041 |
| AG4COLOHCLHWUUYWCZ77TPJLE3EQ 0.720082 |
| AFCYUFW3NQ37UQXYVWL3LN4LAKLQ 0.715082 |
| AG3ZLSFL6WEHCXA2SETWSPPDGTVQ 0.668350 |
| AFXF3EGQTQDXMRLDWFU7UBFQZB7Q 0.629153 |
| AEQAYV7RXZEBXMQIQPL6KCT2CFWQ 0.625000 |
| AGBOGNMAG3UCA6B4Z6VT62DVNOIA 0.620543 |
| AGYPMYXECLNQ2I5WWFQ52COBAFHA 0.612596 |
| AHHFW36BP4VMQWC6V2NTKIXFAA2A 0.612596 |
| AHPFHP43AXWRYZZ4HPNCW7I7J3ZQ 0.591444 |
| AHMXJDRWRZ2OGHXU4SQDFS6KIBZA 0.589256 |
| AEVWAM3YWN5URJVJIZZ6XPD2MKIA 0.577651 |
| AG3VXRJ5OUQDF3UAEOEIIZ6Z5Z3A 0.577651 |
| AEZGPLOYTSAPR3DHZKKXEFPAXUAA 0.554257 |
| AHVWKBFOVXQV6653O4XGCKOB6YVA 0.535450 |
| AFPBVAEV3FCQOHKJJGAKQLZLTCZQ 0.520416 |
| AEWFEWBJVI2YN7WAWVXSOYT5MANA 0.520416 |
| AHB5CGLYN3Y6NIPHNQLYFJT2W2PQ 0.520416 |

Top 20% closest users to **AFSKPY37N3C43SOI5IEXEK5JSIYA:**

user\_id Similarity

|  |
| --- |
| AG6SNHIVEFMJIXCLZWKTGEOV5ZBQ 0.638690 |
| AELHB5QYXVSXZM263JIARJBWPOSA 0.620174 |
| AETE7Y3DZT6BLMWA6U27ADJDZ4LA 0.572323 |
| AFXF3EGQTQDXMRLDWFU7UBFQZB7Q 0.561179 |
| AG3ZLSFL6WEHCXA2SETWSPPDGTVQ 0.559342 |
| AFZK2BA7HVTZTYYVBQ6YYL5XPWLA 0.524142 |
| AHOT2ODB3FZ72IU5KMHHFPB6SNNA 0.524142 |
| AFPBVAEV3FCQOHKJJGAKQLZLTCZQ 0.513964 |
| AHB5CGLYN3Y6NIPHNQLYFJT2W2PQ 0.513964 |
| AEWFEWBJVI2YN7WAWVXSOYT5MANA 0.513964 |
| AFQQQ5LGNSQUEBGDCYBAZZE5T3DA 0.513964 |
| AEIPJBAN7A55Q5DFFPZSR2UV3OKA 0.508493 |
| AG5RLYHH277YN5CG5UIMLHMG4XWQ 0.497245 |
| AFUB7CHTXRPD447QVQCHBZVN2IPQ 0.497245 |
| AFA2U2ZWVRYX3ACXTUL7DU3ZUGLQ 0.497245 |
| AHV6QCNBJNSGLATP56JAWJ3C4G2A 0.486654 |
| AHEJ5LC7BSEADCIZQQQPZPVWOLCA 0.486654 |
| AEMJ2EG5ODOCYUTI54NBXZHDJGSQ 0.481456 |
| AEHOUKDUSHPOEDHAISNAVI7ANZHA 0.468807 |

--- Predictions using Original Similarity ---

Predictions for **AETE7Y3DZT6BLMWA6U27ADJDZ4LA**:

Item: B08FCQML37, Predicted Rating: 5.00, Classification: like

Item: B08GYM3HVP, Predicted Rating: 3.00, Classification: like

Predictions for **AEMJ2EG5ODOCYUTI54NBXZHDJGSQ**:

Item: B09FKT5PQ9, Predicted Rating: 3.38, Classification: like

Item: B08FCQML37, Predicted Rating: 2.40, Classification: dislike

Item: B08GYM3HVP, Predicted Rating: 3.00, Classification: like

Predictions for **AFSKPY37N3C43SOI5IEXEK5JSIYA**:

Item: B07TDSJZMR, Predicted Rating: 3.96, Classification: like

Item: B08637FWWF, Predicted Rating: 3.00, Classification: dislike

Item: B07KJVGNN5, Predicted Rating: 2.76, Classification: dislike

Item: B007HY7GC2, Predicted Rating: 2.84, Classification: dislike

Item: B08KYJLF5T, Predicted Rating: 2.79, Classification: dislike

--- Predictions using Discounted Similarity ---

Predictions for **AETE7Y3DZT6BLMWA6U27ADJDZ4LA:**

Item: B08FCQML37, Predicted Rating: 2.00, Classification: dislike

Item: B08GYM3HVP, Predicted Rating: 4.00, Classification: like

Predictions for **AEMJ2EG5ODOCYUTI54NBXZHDJGSQ:**

Item: B09FKT5PQ9, Predicted Rating: 5.00, Classification: like

Item: B08FCQML37, Predicted Rating: 2.00, Classification: dislike

Item: B08GYM3HVP, Predicted Rating: 4.00, Classification: like

Predictions for **AFSKPY37N3C43SOI5IEXEK5JSIYA:**

Item: B07TDSJZMR, Predicted Rating: 3.54, Classification: like

Item: B08637FWWF, Predicted Rating: 4.15, Classification: like

Item: B07KJVGNN5, Predicted Rating: 2.78, Classification: dislike

Item: B007HY7GC2, Predicted Rating: 3.31, Classification: like

Item: B08KYJLF5T, Predicted Rating: 4.18, Classification: like

**Case 1.2**

Similarities of **AETE7Y3DZT6BLMWA6U27ADJDZ4LA** with all other users:

user\_id

|  |
| --- |
| AEMJ2EG5ODOCYUTI54NBXZHDJGSQ -0.018248 |
| AFSKPY37N3C43SOI5IEXEK5JSIYA 0.148418 |
| AFKZENTNBQ7A7V7UXW5JJI6UGRYQ 0.193122 |
| AEVWAM3YWN5URJVJIZZ6XPD2MKIA -0.615967 |
| AHSPLDNW5OOUK2PLH7GXLACFBZNQ -0.167248 |
| ... |
| AEKLWIFD6GVHWJKIC4QDCDYRNYKQ -0.167248 |
| AG3ZLSFL6WEHCXA2SETWSPPDGTVQ 0.096561 |
| AFLX66DKF6R3H6OEOC3TIVAYXZIQ 0.000000 |
| AFQ7WYW4KSH4VI5OVXCP2GV6PBRA 0.167248 |
| AETGCWXC47MSMK6B2TLZ44KCFJZQ 0.000000 |

Top 20% closest users to **AETE7Y3DZT6BLMWA6U27ADJDZ4LA**:

user\_id

|  |
| --- |
| AFCYUFW3NQ37UQXYVWL3LN4LAKLQ 0.427890 |
| AEZGPLOYTSAPR3DHZKKXEFPAXUAA 0.386244 |
| AEQAYV7RXZEBXMQIQPL6KCT2CFWQ 0.386244 |
| AG6SNHIVEFMJIXCLZWKTGEOV5ZBQ 0.253472 |
| AFKZENTNBQ7A7V7UXW5JJI6UGRYQ 0.193122 |
| AFZUK3MTBIBEDQOPAK3OATUOUKLA 0.193122 |
| AE3TASYGLHHRHUJUDFTKFDMWFIYA 0.193122 |
| AHEJ5LC7BSEADCIZQQQPZPVWOLCA 0.167248 |
| AEIPJBAN7A55Q5DFFPZSR2UV3OKA 0.167248 |
| AHV6QCNBJNSGLATP56JAWJ3C4G2A 0.167248 |
| AHZOJQDTMJ6WIHY2RV7PB6K5TG7Q 0.167248 |
| AFPBVAEV3FCQOHKJJGAKQLZLTCZQ 0.167248 |
| AEHOUKDUSHPOEDHAISNAVI7ANZHA 0.167248 |
| AEHWUHNTB5FX32HJ7UBOZ2WWUX3Q 0.167248 |
| AHB5CGLYN3Y6NIPHNQLYFJT2W2PQ 0.167248 |
| AFUZHS2CHLPSFORO3LDM5VMVK3ZA 0.167248 |
| AFA2U2ZWVRYX3ACXTUL7DU3ZUGLQ 0.167248 |
| AF34HQQ3RZOFQDNN6TBKW523Z33A 0.167248 |
| AEWFEWBJVI2YN7WAWVXSOYT5MANA 0.167248 |
| AFUB7CHTXRPD447QVQCHBZVN2IPQ 0.167248 |

Similarities of **AEMJ2EG5ODOCYUTI54NBXZHDJGSQ** with all other users:

user\_id

|  |
| --- |
| AETE7Y3DZT6BLMWA6U27ADJDZ4LA -0.018248 |
| AFSKPY37N3C43SOI5IEXEK5JSIYA 0.152153 |
| AFKZENTNBQ7A7V7UXW5JJI6UGRYQ 0.551198 |
| AEVWAM3YWN5URJVJIZZ6XPD2MKIA 0.397475 |
| AHSPLDNW5OOUK2PLH7GXLACFBZNQ -0.763763 |
| ... |
| AEKLWIFD6GVHWJKIC4QDCDYRNYKQ -0.763763 |
| AG3ZLSFL6WEHCXA2SETWSPPDGTVQ 0.440959 |
| AFLX66DKF6R3H6OEOC3TIVAYXZIQ 0.000000 |
| AFQ7WYW4KSH4VI5OVXCP2GV6PBRA 0.763763 |
| AETGCWXC47MSMK6B2TLZ44KCFJZQ 0.000000 |

Top 20% closest users to **AEMJ2EG5ODOCYUTI54NBXZHDJGSQ**:

user\_id

|  |
| --- |
| AHEJ5LC7BSEADCIZQQQPZPVWOLCA 0.763763 |
| AEIPJBAN7A55Q5DFFPZSR2UV3OKA 0.763763 |
| AHV6QCNBJNSGLATP56JAWJ3C4G2A 0.763763 |
| AEHOUKDUSHPOEDHAISNAVI7ANZHA 0.763763 |
| AFUZHS2CHLPSFORO3LDM5VMVK3ZA 0.763763 |
| AHBI5SLZDP3Q3LZPETJLCHQFGLUA 0.763763 |
| AFUB7CHTXRPD447QVQCHBZVN2IPQ 0.763763 |
| AHB5CGLYN3Y6NIPHNQLYFJT2W2PQ 0.763763 |
| AHU2Y2ZFQKI3V3ARFDKZA6ER4NUQ 0.763763 |
| AFA2U2ZWVRYX3ACXTUL7DU3ZUGLQ 0.763763 |
| AEHWUHNTB5FX32HJ7UBOZ2WWUX3Q 0.763763 |
| AHZOJQDTMJ6WIHY2RV7PB6K5TG7Q 0.763763 |
| AEWFEWBJVI2YN7WAWVXSOYT5MANA 0.763763 |
| AFVTVAR5V6XKWMVHARHZXRCOSHIQ 0.763763 |
| AG5RLYHH277YN5CG5UIMLHMG4XWQ 0.763763 |
| AFPBVAEV3FCQOHKJJGAKQLZLTCZQ 0.763763 |
| AF34HQQ3RZOFQDNN6TBKW523Z33A 0.763763 |
| AHY7ZJB523OPTIKXRI63PS2V6FSQ 0.763763 |
| AFWTME2ROGUQI5J5FB3DWCLKZNBA 0.763763 |
| AFQ7WYW4KSH4VI5OVXCP2GV6PBRA 0.763763 |

Similarities of **AFSKPY37N3C43SOI5IEXEK5JSIYA** with all other users:

user\_id

|  |
| --- |
| AETE7Y3DZT6BLMWA6U27ADJDZ4LA 0.148418 |
| AEMJ2EG5ODOCYUTI54NBXZHDJGSQ 0.152153 |
| AFKZENTNBQ7A7V7UXW5JJI6UGRYQ 0.195180 |
| AEVWAM3YWN5URJVJIZZ6XPD2MKIA -0.108266 |
| AHSPLDNW5OOUK2PLH7GXLACFBZNQ -0.338062 |
| ... |
| AEKLWIFD6GVHWJKIC4QDCDYRNYKQ -0.338062 |
| AG3ZLSFL6WEHCXA2SETWSPPDGTVQ -0.292770 |
| AFLX66DKF6R3H6OEOC3TIVAYXZIQ 0.000000 |
| AFQ7WYW4KSH4VI5OVXCP2GV6PBRA 0.338062 |
| AETGCWXC47MSMK6B2TLZ44KCFJZQ 0.000000 |

Top 20% closest users to **AFSKPY37N3C43SOI5IEXEK5JSIYA**:

user\_id

|  |
| --- |
| AFXF3EGQTQDXMRLDWFU7UBFQZB7Q 0.442627 |
| AEZGPLOYTSAPR3DHZKKXEFPAXUAA 0.390360 |
| AEQAYV7RXZEBXMQIQPL6KCT2CFWQ 0.390360 |
| AHEJ5LC7BSEADCIZQQQPZPVWOLCA 0.338062 |
| AHV6QCNBJNSGLATP56JAWJ3C4G2A 0.338062 |
| AEIPJBAN7A55Q5DFFPZSR2UV3OKA 0.338062 |
| AFPBVAEV3FCQOHKJJGAKQLZLTCZQ 0.338062 |
| AFA2U2ZWVRYX3ACXTUL7DU3ZUGLQ 0.338062 |
| AHU2Y2ZFQKI3V3ARFDKZA6ER4NUQ 0.338062 |
| AEWFEWBJVI2YN7WAWVXSOYT5MANA 0.338062 |
| AEHWUHNTB5FX32HJ7UBOZ2WWUX3Q 0.338062 |
| AEHOUKDUSHPOEDHAISNAVI7ANZHA 0.338062 |
| AHB5CGLYN3Y6NIPHNQLYFJT2W2PQ 0.338062 |
| AFUB7CHTXRPD447QVQCHBZVN2IPQ 0.338062 |
| AFVTVAR5V6XKWMVHARHZXRCOSHIQ 0.338062 |
| AHBI5SLZDP3Q3LZPETJLCHQFGLUA 0.338062 |
| AHZOJQDTMJ6WIHY2RV7PB6K5TG7Q 0.338062 |
| AG5RLYHH277YN5CG5UIMLHMG4XWQ 0.338062 |
| AFQQQ5LGNSQUEBGDCYBAZZE5T3DA 0.338062 |
| AFWTME2ROGUQI5J5FB3DWCLKZNBA 0.338062 |

Predictions for **AETE7Y3DZT6BLMWA6U27ADJDZ4LA**:

Item: B08FCQML37, Predicted Rating: 2.38, Classification: dislike

Item: B08GYM3HVP, Predicted Rating: 3.77, Classification: like

Predictions for **AEMJ2EG5ODOCYUTI54NBXZHDJGSQ:**

Item: B09FKT5PQ9, Predicted Rating: 5.00, Classification: like

Item: B08FCQML37, Predicted Rating: 2.00, Classification: dislike

Item: B08GYM3HVP, Predicted Rating: 4.00, Classification: like

Predictions for **AFSKPY37N3C43SOI5IEXEK5JSIYA**:

Item: B07TDSJZMR, Predicted Rating: 5.00, Classification: like

Item: B08637FWWF, Predicted Rating: 2.00, Classification: dislike

Item: B07KJVGNN5, Predicted Rating: 4.09, Classification: like

Item: B08KYJLF5T, Predicted Rating: 5.00, Classification: like

Predictions for **AETE7Y3DZT6BLMWA6U27ADJDZ4LA** (**Discounted Similarity**):

Item: B08FCQML37, Predicted Rating: 1.78, Classification: dislike

Item: B08GYM3HVP, Predicted Rating: 4.00, Classification: like

Predictions for **AEMJ2EG5ODOCYUTI54NBXZHDJGSQ** (**Discounted** **Similarity**):

Item: B09FKT5PQ9, Predicted Rating: 5.00, Classification: like

Item: B08FCQML37, Predicted Rating: 2.00, Classification: dislike

Item: B08GYM3HVP, Predicted Rating: 4.00, Classification: like

Predictions for **AFSKPY37N3C43SOI5IEXEK5JSIYA** (**Discounted Similarity**):

Item: B07TDSJZMR, Predicted Rating: 5.00, Classification: like

Item: B08637FWWF, Predicted Rating: 2.00, Classification: dislike

Item: B07KJVGNN5, Predicted Rating: 4.09, Classification: like

Item: B08KYJLF5T, Predicted Rating: 5.00, Classification: like

**Case 1.3**

Top 20% Closest Users Based on PCC:

User: **AETE7Y3DZT6BLMWA6U27ADJDZ4LA**

user\_id

|  |
| --- |
| AHU2Y2ZFQKI3V3ARFDKZA6ER4NUQ 1.0 |
| AFPBVAEV3FCQOHKJJGAKQLZLTCZQ 1.0 |
| AEHOUKDUSHPOEDHAISNAVI7ANZHA 1.0 |
| AEIPJBAN7A55Q5DFFPZSR2UV3OKA 1.0 |
| AHEJ5LC7BSEADCIZQQQPZPVWOLCA 1.0 |
| AFA2U2ZWVRYX3ACXTUL7DU3ZUGLQ 1.0 |
| AHV6QCNBJNSGLATP56JAWJ3C4G2A 1.0 |
| AEHWUHNTB5FX32HJ7UBOZ2WWUX3Q 1.0 |
| AFUZHS2CHLPSFORO3LDM5VMVK3ZA 1.0 |
| AHB5CGLYN3Y6NIPHNQLYFJT2W2PQ 1.0 |
| AFUB7CHTXRPD447QVQCHBZVN2IPQ 1.0 |
| AEWFEWBJVI2YN7WAWVXSOYT5MANA 1.0 |
| AHBI5SLZDP3Q3LZPETJLCHQFGLUA 1.0 |
| AFVTVAR5V6XKWMVHARHZXRCOSHIQ 1.0 |
| AG5RLYHH277YN5CG5UIMLHMG4XWQ 1.0 |
| AF34HQQ3RZOFQDNN6TBKW523Z33A 1.0 |
| AETE7Y3DZT6BLMWA6U27ADJDZ4LA 1.0 |
| AHY7ZJB523OPTIKXRI63PS2V6FSQ 1.0 |
| AFQQQ5LGNSQUEBGDCYBAZZE5T3DA 1.0 |

User: **AEMJ2EG5ODOCYUTI54NBXZHDJGSQ**

user\_id

|  |
| --- |
| AG6SNHIVEFMJIXCLZWKTGEOV5ZBQ 1.0 |
| AEIPJBAN7A55Q5DFFPZSR2UV3OKA 1.0 |
| AHZOJQDTMJ6WIHY2RV7PB6K5TG7Q 1.0 |
| AHEJ5LC7BSEADCIZQQQPZPVWOLCA 1.0 |
| AFA2U2ZWVRYX3ACXTUL7DU3ZUGLQ 1.0 |
| AEHOUKDUSHPOEDHAISNAVI7ANZHA 1.0 |
| AEHWUHNTB5FX32HJ7UBOZ2WWUX3Q 1.0 |
| AELHB5QYXVSXZM263JIARJBWPOSA 1.0 |
| AFPBVAEV3FCQOHKJJGAKQLZLTCZQ 1.0 |
| AF34HQQ3RZOFQDNN6TBKW523Z33A 1.0 |
| AFUZHS2CHLPSFORO3LDM5VMVK3ZA 1.0 |
| AHU2Y2ZFQKI3V3ARFDKZA6ER4NUQ 1.0 |
| AHB5CGLYN3Y6NIPHNQLYFJT2W2PQ 1.0 |
| AFWTME2ROGUQI5J5FB3DWCLKZNBA 1.0 |
| AEMJ2EG5ODOCYUTI54NBXZHDJGSQ 1.0 |
| AFUB7CHTXRPD447QVQCHBZVN2IPQ 1.0 |
| AFWVN52MRBWOTIK7UGXBWGOY4HBA 1.0 |
| AHBI5SLZDP3Q3LZPETJLCHQFGLUA 1.0 |
| AFQQQ5LGNSQUEBGDCYBAZZE5T3DA 1.0 |

**User: AFSKPY37N3C43SOI5IEXEK5JSIYA**

**user\_id**

|  |
| --- |
| AFXF3EGQTQDXMRLDWFU7UBFQZB7Q 1.0 |
| AETE7Y3DZT6BLMWA6U27ADJDZ4LA 0.5 |
| AGUWL2R2JFLC3K65HLD6AHJV3KBA 0.0 |
| AEIPJBAN7A55Q5DFFPZSR2UV3OKA 0.0 |
| AHEJ5LC7BSEADCIZQQQPZPVWOLCA 0.0 |
| AFA2U2ZWVRYX3ACXTUL7DU3ZUGLQ 0.0 |
| AHHFW36BP4VMQWC6V2NTKIXFAA2A 0.0 |
| AHBZRDFYB2FWUAO63DCSF2VSTJ2Q 0.0 |
| AHU2Y2ZFQKI3V3ARFDKZA6ER4NUQ 0.0 |
| AF4NIQPIZQ3S3G6KEVQW33DNHZSQ 0.0 |
| AFCYUFW3NQ37UQXYVWL3LN4LAKLQ 0.0 |
| AEYGPUCRKH7G4VM22FM3VAKSQ23Q 0.0 |
| AHOEABHRAFWXIT4JZ5MKJ3FMASGA 0.0 |
| AG3VXRJ5OUQDF3UAEOEIIZ6Z5Z3A 0.0 |
| AHMXJDRWRZ2OGHXU4SQDFS6KIBZA 0.0 |
| AG535JFIAGQV4CY7TREDVYPDNGKA 0.0 |
| AHJQPUQLSQZE6LMIUMY7WNRXCQQQ 0.0 |
| AHRQPBQJJ2PJET5WBJIKNLAFHHSA 0.0 |
| AEWFEWBJVI2YN7WAWVXSOYT5MANA 0.0 |

Predictions for User: **AETE7Y3DZT6BLMWA6U27ADJDZ4LA**

Item: B08FCQML37, Predicted Rating: 2.37, Prediction: Dislike

Item: B08GYM3HVP, Predicted Rating: 1.00, Prediction: Dislike

Predictions for User: **AEMJ2EG5ODOCYUTI54NBXZHDJGSQ**

Item: B09FKT5PQ9, Predicted Rating: 4.86, Prediction: Like

Item: B08FCQML37, Predicted Rating: 1.67, Prediction: Dislike

Item: B08GYM3HVP, Predicted Rating: 4.00, Prediction: Like

Predictions for User: **AFSKPY37N3C43SOI5IEXEK5JSIYA**

Item: B07TDSJZMR, Predicted Rating: 5.00, Prediction: Like

Item: B08637FWWF, Predicted Rating: 1.00, Prediction: Dislike

Item: B07KJVGNN5, Predicted Rating: 3.33, Prediction: Like

Item: B007HY7GC2, Predicted Rating: 2.00, Prediction: Dislike

Item: B08KYJLF5T, Predicted Rating: 2.00, Prediction: Dislike

**Part 2 requirements and questions:-  
Case 2.1:**

Description:

•Loading the dataset, identifying items with missing values, and creating a cosine similarity matrix were involved.

•Predicted ratings were made for B07KJVGNN5, other target items.

Key Points:

•Missing Data Analysis: All other items had above 90 percent of missing values; B07KJVGNN5 had near to 4 percent of missing values.

•Similarity Matrix: A cosine similarity matrix identified the 25 most similar items.

•Predicted Ratings: Ratings were predicted for the item B07KJVGNN5 across 100 users; for instance, user ratings for B07KJVGNN5 ranged from 1.0 to 5.0.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Case 2.2:**

Description:

•This comparison used cosine similarity and analyzed the predicted ratings for the 20% closest items for the users.

Key Points:

•Item Similarity Comparison: Two lists of similar items were built by utilizing self (cosine similarity with bias adjustment) and other (alternative similarity measure).

For example, B07TDSJZMR was similar to [B07KJVGNN5, B08FCQML37] via "self," while it mapped to [B08637FWWF, B09FKT5PQ9] under the "other" method.

•Predicted Ratings: The matrix displays ratings that were predicted for different models for each user-item pair.

For instance, AEVWAM3YWN5URJVJIZZ6XPD2MKIA was shown to have a predicted rating of 1.0 for B07TDSJZMR and 2.932 for B08637FWWF using the "other" item.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Case 2.3:**

Description

•This case delves into the comparative analysis of similarity and rating predictions based on PCC and Discounted PCC.

Key Points:

•Similarity Comparison: The similarities were compared for the top 20% closest items under PCC and Discounted PCC.

Example: B07TDSJZMR was closest to [B07KJVGNN5, B09GBMG83Z] under PCC but [B007HY7GC2, B07TDSJZMR] under Discounted PCC.

•Predicted Ratings: Then, a comparison was undertaken to contrast the predicted ratings for various items using PCC and Discounted PCC.

Example: The predicted ratings for B07TDSJZMR and B08637FWWF were compared for each user by using both measures.

**Summary of the Comparison of part 1 and 2 :-**

Case Study 1.1 Observations

Original Similarity: The predictions are typically skewed right; most items therefore get classified as "like." This indicates a bias in similarity-based ratings without any further constraints.

Discounted Similarity: Ratings drop significantly for some users and items. In fact, this falloff shows the effect of applying a threshold to curb the overestimation of similarity. This sets up a semblance of balance, hurtling separation between "like" and "dislike."

Case Study 1.2 Observations

Mean-centering adjustments and bias correction create more stable and realistic predictions

Original Similarity: The ratings tend to be high, but with finer classifications than those in Case 1.1. For some users, there is still some preference bias, but this is doing less injury.

Discounted Similarity: Ratings and classifications are further adjusted, ensuring significant weighting plays a clearer role. Disliked items are recognized with greater confidence, though liked items remain rated high.

Case Study 1.3 Observations

The most restrictive yet precise approach

This results in lower predictions across the board than both cases 1.1 and 1.2. The early takeaway is that significance weighting more strongly reduces the negative impact of weak correlations.

Some borderline items classified as "like" in earlier cases are now relegated to "dislike" by certain users like AETE7Y3DZT6BLMWA6U27ADJDZ4LA and AFSKPY37N3C43SOI5IEXEK5JSIYA.

Impact of Applying Significance Weighting

Refinement of the Top-N List:

In Case 1.1, the top-N list consists of items with inflated similarities, therefore yielding relatively inaccurate recommendations.

In Case 1.2, mean-centering shifts correction for bias create a more balanced top-N list.

In Case 1.3, only strong correlations are permitted; thus, the top-N list is further refined, and weak candidates are excluded.

Rating Predictions:

The ratings are too optimistic in Case 1.1, making the predictions not reliable when it comes to practical applications.

In Case 1.2, the bias is tempered, promoting more accurate and varied ratings.

In Case 1.3, the way of significance weighting becomes stricter; this improves reliability at the cost of a lower average predicted rating, rendering them not as reliable.

**General Comments and Summary**

The implications of significance weighting are crucial to enhancing the quality of the top-N list, as well as the accuracy of predictions.

Case 1.1 is where this becomes our baseline, but without any constraints, it will lead to overestimations that will somewhat lower the efficiencies of recommendations.

Case 1.2 addresses the bias, bringing about an average between accuracy and predictability.

Case 1.3 with a more rigid disqualification of weak correlations has ultimately translated for the best accuracy along with a more reliable but conservative prediction.

Recommendation: A balance between bias correction and significance weighting (Case 1.2) may be generally ideal for most use cases, but perhaps it might be better to restrict to the strider course emphasized by (Case 1.3) for applications that desire maximum reliability.

**----------------------------------------------------------------------------------------------------------------**

Case 2.1. Baseline approach using cosine similarity. Items with high missing value percentages were identified, and B07KJVGNN5 (with fewer missing values) was chosen as the target item. An attempt at prediction based on the 25 most similar items failed to adjust the bias on the predictions or weighted these most similar items appropriately; as such, the predicted values were not precise.

Case 2.2. Improved approach with cosine similarity and bias adjustment. Predictions were compared using "self" (bias-adjusted Cosine similarity) against "other" (alternative similarity measures). Predictions proved more accurate, with similarities adjusted.

Case 2.3. Advanced approach using the Pearson Correlation Coefficient (PCC) and Discounted PCC. The Discounted PCC alleviated noise problems by strengthening the correlations and thus improving the similarity rankings. This led to the most accurate and consistent predictions.

Impact Analysis

Top-N list: Case 2.3 produced the most precise Top-N lists, armed with significance weighting, while Case 2.1 was unreliable.

Predicted Ratings: Case 2.3 produced the most accurate predictions. Case 2.2 was biased down, which led to more variance compared to Case 2.1.

Recommendation

Case 2.3 is preferable for accuracy and reliability.

Case 2.2 is a fine alternative for implementations with less complexity and somewhat less accuracy.

**Enhancement:-**

The gradual incorporation of bias correction and stricter significance weighting produces more precise and reliable predictions, with the ability to cater to varying usage requirements (e.g., balancing predictability vs. reliability).

When the baseline approaches to bias-corrected and significance-weighted methods are followed, then in Case 2, prediction accuracy is gradually increased while recommendation reliability also steadily improves. Only a minor degenerated discount similarity correlation (discounted PCC) (Case 2.3) is better since it is thus more precise and stable, useful for accurate top-N recommendations and predictions.

Summary of Enhancements;

1. Accuracy: By bias adjustment, filtering out aberrant correlations, and applying discounted significance weights.
2. Reliability: By subjecting these correlations to a stricter filtering process and noise reduction techniques.
3. Flexibility: By fairly balancing between prediction accuracy and computational complexity such that any application needs are catered with ease.

**Equation used in codes:-**

**Conclusion :-**

This analysis showed the importance of bias adjustment and significance weighting to improve recommendation prediction accuracy. The significance weighting reduces the influence of weak correlations, thereby better ranking the similar ratings and increasing the reliability of the predictions. Bias adjustments would modulate overly optimistic predictions and neutralize very high and low ratings for liked and disliked items.

Key Findings:

The influence of significance weighting: Cases 1.3 and 2.3 give rise to most accurate predictions and Top-N lists excluding weak correlations, albeit with lower average ratings.

Cases 1.2 and 2.2 presented a more balanced approach between accuracy and computational efficiency.

Recommendations: For many applications, bias adjustment and moderate significance weighting would be the ideal compromise, with more stringent methods for high-accuracy needs.

Fundamental improvements would be dynamic weighting, hybrid models, and real-life validation for improvement in performance and user satisfaction. Therefore, importance-weighted biasline adjustments are critical to ensure robust and reliable recommendation systems.

**Reference :-**

**[1] R. J. Moen, "Average cosine similarity of user pairs as a function of the number of edits from time of," *ResearchGate*. Available:** [**https://www.researchgate.net/figure/Average-cosine-similarity-of-user-pairs-as-a-function-of-the-number-of-edits-from-time-of\_fig1\_220271994**](https://www.researchgate.net/figure/Average-cosine-similarity-of-user-pairs-as-a-function-of-the-number-of-edits-from-time-of_fig1_220271994)**. [Accessed: 16-Dec-2024].**

**[2] D. K. L. S. "2D Array in Java – Two Dimensional and Nested Arrays," *freeCodeCamp*. Available:** [**https://www.freecodecamp.org/news/2d-array-in-java-two-dimensional-and-nested-arrays/**](https://www.freecodecamp.org/news/2d-array-in-java-two-dimensional-and-nested-arrays/)**. [Accessed: 16-Dec-2024].**

**[3] "Prediction equation," *ScienceDirect*. Available:** [**https://www.sciencedirect.com/topics/computer-science/prediction-equation**](https://www.sciencedirect.com/topics/computer-science/prediction-equation)**. [Accessed: 16-Dec-2024].**

**[4] "Pearson’s correlation coefficient," *Britannica*. Available:** [**https://www.britannica.com/topic/Pearsons-correlation-coefficient**](https://www.britannica.com/topic/Pearsons-correlation-coefficient)**. [Accessed: 16-Dec-2024].**

**[5] "Recommender System Series Part 2: Neighborhood-Based Collaborative Filtering," *Encora Insights*. Available:** [**https://insights.encora.com/insights/recommender-system-series-part-2-neighborhood-based-collaborative-filtering**](https://insights.encora.com/insights/recommender-system-series-part-2-neighborhood-based-collaborative-filtering)**. [Accessed: 16-Dec-2024].**

**[6] D. R. G. M. de M. and R. B. G. S. S. J. P. G. P. H. D. S. J., "A survey on recommender systems based on collaborative filtering techniques," *Journal of Big Data*, vol. 8, no. 1, pp. 1-24, 2021. Available:** [**https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00538-3**](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00538-3)**. [Accessed: 16-Dec-2024].**

**[7] "Item-to-Item Based Collaborative Filtering," *GeeksforGeeks*, [Online]. Available:** [**https://www.geeksforgeeks.org/item-to-item-based-collaborative-filtering/**](https://www.geeksforgeeks.org/item-to-item-based-collaborative-filtering/)**. [Accessed: Nov. 5, 2024].**

**[8] "Pandas Documentation," [Online]. Available:** [**https://pandas.pydata.org/docs/**](https://pandas.pydata.org/docs/)**. [Accessed: Nov. 5, 2024]**