* Updated table.

	U1	U2	U3	U4	U5
TDN	0	0	-1	0	1
TOB	1	0	0	0	-1
TLOTR	1.33	0	-0.67	-0.67	0
TGF	1.75	-1.25	-0.25	-0.25	0
TGTB&TU	0	0	0	0	0

1. $\cos(TOB, TDN) = \frac{(1)(-1)}{\sqrt{2} \times \sqrt{2}} = -0.5$

2. $\cos(TOB, TLOTR) = \frac{(1)(1.33)}{\sqrt{2} \times \sqrt{1.33^2 + (0.67)^2 \times 2}} = 0.57$

3. $\cos(TOB, TGF) = \frac{(1)(1.75)}{\sqrt{2} \times \sqrt{1.75^2 + 1.25^2 + (0.25)^2 \times 2}} = 0.55$

4. $\cos(TOB, TGTB\&TU) = 0$

∴ Similarity of TOB with other movies is
$[-0.5, \boxed{0.57}, \boxed{0.55}, 0]$

... 2 movies with highest similarity = TLOTR, TGF

∴ To predict the rating, we take weighted avg.

$\mu_{TOB, 3} = \frac{0.57 \times 2 + 0.55 \times 3}{0.57 + 0.55} = 2.49 \approx 2.5$

∴ Rating of movie of TOB by User 3 is 2.5

PAGE No.
DATE

Q2) b) TGF movie – user U5
using Item – Item collaborative filtering.

	U1	U2	U3	U4	U5
TDN	4	0	3	0	5
TOB	4	3	0	0	2
TLOTR	4	0	2	2	0
TGF	5	2	3	3	2
TGTB&TU	4	4	0	0	0

* finding centered cosine similarity for TOB with all movies.

(A) Mean of Row.

Row 1 = 4 + 3 + 5 / 3 = 4
Row 2 = 4 + 3 + 2 / 3 = 3
Row 3 = 4 + 2 + 2 / 3 = 2.67
Row 4 = 5 + 2 + 3 + 3 / 4 = 3.25
Row 5 = 4 + 4 / 2 = 4

(B) Substract the mean with respective row values.

Updated table.	U1	U2	U3	U4	U5
TDN	0	0	-1	0	1
TOB	1	0	0	0	-1
TLOTR	1.33	~~-1.25~~ 0	-0.67	-0.67	0
TGF	1.75	-1.25	-0.25	-0.25	0
TGTB&TU	0	0	0	0	0

Q2) Collaborative Filtering to calculate ratings.

a. TOB movie — user U3

Using Item - Item collaborative filtering.

	U_1	U_2	U_3	U_4	U_5
TDN	4	0	3	0	5
TOB	4	3	[?]	0	2
TLOTR	4	0	2	2	0
TGF	5	2	3	3	0
TGTB&TU	4	4	0	0	0

* Finding centered cosine similarity for TOB with all movies.

(A) Mean of Row.

Row 1 $= \frac{4+3+5}{3} = 4$

Row 2 $= \frac{4+3+2}{3} = 3$

Row 3 $= \frac{4+2+2}{3} = 2.67$

Row 4 $= \frac{5+2+3+3}{4} = 3.25$

Row 5 $= \frac{4+4}{2} = 4$

(B) Subtract the mean with respective row values.

∴ Cosine Similarity of user U_1 & U_3 is the highest = 0.791.
Thus, they are the most similar.

3. Centered Cosine Similarity:
Calculate mean of each row & substract mean of that row from all the readings of that row.

Mean of
Row 1 $= \frac{4+4+4+5+4}{5} = 4.2$

Row 2 $= \frac{3+2+4}{3} = 3$

Row 3 $= \frac{3+2+3}{3} = 2.67$

Row 4 $= \frac{2+3}{2} = 2.5$

Row 5 $= \frac{5+2}{2} = 3.5$

* Revised ratings of all users will be
..(subtract mean from respective row values)

u_1 [-0.2, -0.2, -0.2, 0.8, -0.2]

u_2 [0, 0, 0, -1, 1]

u_3 [0.33, 0, -0.67, 0.33, 0]

u_4 [0, 0, -0.5, 0.5, 0]

u_5 [1.5, -1.5, 0, 0, 0]

2. Cosine Similarity Method:

Formula: $\frac{\sum A \cdot B}{\sqrt{\sum A^2} \times \sqrt{\sum B^2}}$

$U_1 = [4, 4, 4, 5, 4]$
$U_2 = [0, 3, 0, 2, 4]$
$U_3 = [3, 0, 2, 3, 0]$
$U_4 = [0, 0, 2, 3, 0]$
$U_5 = [5, 2, 0, 0, 0]$

a) $c(U_1, U_2) = \frac{U_1 \cdot U_2}{\|U_1\| \|U_2\|} = \frac{4\times0 + 4\times3 + 4\times0 + 5\times2 + 4\times4}{\sqrt{4^2+4^2+4^2+5^2+4^2} \times \sqrt{3^2+2^2+4^2}}$

$= \frac{38}{\sqrt{89} \times \sqrt{29}} = 0.748$

b) $c(U_1, U_3) = \frac{4\times3 + 4\times0 + 4\times2 + 5\times3 + 4\times0}{\sqrt{4^2+4^2+4^2+5^2+4^2} \times \sqrt{3^2+2^2+3^2}} = 0.791$

c) $c(U_1, U_4) = \frac{4\times0 + 4\times0 + 4\times2 + 5\times3 + 4\times0}{\sqrt{4^2+4^2+4^2+5^2+4^2} \times \sqrt{2^2+3^2}} = 0.6762$

d) $c(U_1, U_5) = \frac{4\times5 + 4\times2 + 4\times0 + 5\times0 + 4\times0}{\sqrt{4^2+4^2+4^2+5^2+4^2} \times \sqrt{5^2+2^2}} = 0.551$

e) $c(U_2, U_3) = \frac{0\times3 + 3\times0 + 0\times2 + 2\times3 + 4\times0}{\sqrt{3^2+2^2+4^2} \times \sqrt{3^2+2^2+3^2}} = 0.238$

f) $c(U_2, U_4) = \frac{0\times0 + 3\times0 + 0\times2 + 2\times3 + 4\times0}{\sqrt{3^2+2^2+4^2} \times \sqrt{2^2+3^2}} = 0.31$

g) $c(U_2, U_5) = \frac{0\times5 + 3\times2 + 0\times0 + 2\times0 + 4\times0}{\sqrt{3^2+2^2+4^2} \times \sqrt{5^2+2^2}} = 0.21$

h) $c(U_3, U_4) = \frac{3\times0 + 0\times0 + 2\times2 + 3\times3 + 0\times0}{\sqrt{3^2+2^2+3^2} \times \sqrt{2^2+3^2}} = 0.769$

i) $c(U_3, U_5) = \frac{3\times5}{\sqrt{3^2+2^2+3^2} \times \sqrt{5^2+2^2}} = 0.594$

j) $c(U_4, U_5) = \frac{0}{\sqrt{2^2+3^2} + \sqrt{5^2+2^2}} = 0$

Q2) Collaborative filtering to calculate rating.

Q. TOB movie — user U3

Using Item - Item collaborative filtering.

	U_1	U_2	U_3	U_4	U_5
TDN	4	0	3	0	5
TOB	4	3	[2]	0	2
TLOTR	4	0	2	2	0
TGF	5	2	3	3	0
TGTBRTU	4	4	0	0	0

* Finding centered cosine similarity for TOB with all movies.

(A) Mean of Row.

Row 1 = $\frac{4+3+5}{3} = 4$

Row 2 = $\frac{4+3+2}{3} = 3$

Row 3 = $\frac{4+2+2}{3} = 2.67$

Row 4 = $\frac{5+2+3+3}{4} = 3.25$

Row 5 = $\frac{4+4}{2} = 4$

(B) Subtract the mean with respective row values.

SMA Assignment 3

Atharva Pawar (9427)
Comps - A

PAGE No.
DATE

Q1. Dataset : Reviews for the movies.

Q Find Similar users using follow methods.

1. Jaccard Similarity
2. Cosine similarity
3. Centered cosine Similarity

* Table :

	TDN	TOB	TLOTR	TGF	TGTB&TU
U1	4	4	4	5	4
U2		3		2	4
U3	3		2	3	
U4			2	3	
U5	5	2			

1. Jaccard Similarity Method : $Jacc = \frac{|A \cap B|}{|A \cup B|}$

$J(U_1, U_2) = \frac{3}{5} = 0.6$

$J(U_1, U_3) = \frac{3}{5} = 0.6$

$J(U_1, U_4) = \frac{2}{5} = 0.4$

$J(U_1, U_5) = \frac{2}{5} = 0.4$

$J(U_2, U_3) = \frac{1}{5} = 0.2$

$J(U_2, U_4) = \frac{1}{4} = 0.25$

$J(U_2, U_5) = \frac{1}{4} = 0.25$

$J(U_3, U_4) = \frac{2}{3} = 0.66$

$J(U_3, U_5) = \frac{1}{4} = 0.25$

$J(U_4, U_5) = \frac{0}{4} = 0$

∴ Jaccard Similarity of users U3 & U4 is the highest, thus users U3 & U4 are the most similar.

(Cosine Tech)

1. $c(u_1, u_2) = \frac{(0.8)(-1) + (0.2)(1)}{\sqrt{(-0.2)^2 \times 4 + (0.8)^2} \times \sqrt{(-1)^2 + 1^2}} = -0.791$

2. $c(u_1, u_3) = \frac{(-0.2)(0.33) + (-0.2)(-0.67) + (0.8)(0.33)}{\sqrt{(0.2)^2 \times 4 + (0.8)^2} \times \sqrt{(0.33)^2 \times 2 + (0.67)^2}} = 0.455$

3. $c(u_1, u_4) = \frac{(-0.2)(-0.5) + (0.8)(0.5)}{\sqrt{(0.2)^2 \times 4 + (0.8)^2} \times \sqrt{(0.5)^2 \times 2}} = 0.791$

4. $c(u_1, u_5) = \frac{(-0.2)(1.5) + (-0.2)(-1.5)}{\sqrt{(0.2)^2 \times 4 + (0.8)^2} \times \sqrt{(1.5)^2 \times 2}} = 0$

5. $c(u_2, u_3) = \frac{(-1)(0.33)}{\sqrt{2} \times \sqrt{(0.33)^2 \times 2 + (0.67)^2}} = -0.286$

6. $c(u_2, u_4) = \frac{(-1)(0.5)}{\sqrt{2} \times \sqrt{(0.5)^2 \times 2}} = -0.5$

7. $c(u_2, u_5) = 0$

8. $c(u_3, u_4) = \frac{(-0.67)(-0.5) + (0.33)(0.5)}{\sqrt{(0.33)^2 \times 2 + (0.67)^2} \times \sqrt{(0.5)^2 \times 2}} = 0.866$

9. $c(u_3, u_5) = \frac{(0.33)(1.5)}{\sqrt{(0.33)^2 \times 2 + (0.67)^2} \times \sqrt{(1.5)^2 \times 2}} = 0.286$

10. $c(u_4, u_5) = 0$

Users 3 & user 4 have highest centered cosine similarity, thus they are most similar.

Q2) Collaborative filtering to calculate rating.

Q. TOB movie — user U3

Using Item - Item collaborative filtering.

	U_1	U_2	U_3	U_4	U_5
TDN	4	0	3	0	5
TOB	4	3	[2]	0	2
TLOTR	4	0	2	2	0
TGF	5	2	3	3	0
TGTBRTU	4	4	0	0	0

* Finding centered cosine similarity for TOB with all movies.

(A) Mean of Row.

Row 1 = $\frac{4+3+5}{3} = 4$

Row 2 = $\frac{4+3+2}{3} = 3$

Row 3 = $\frac{4+2+2}{3} = 2.67$

Row 4 = $\frac{5+2+3+3}{4} = 3.25$

Row 5 = $\frac{4+4}{2} = 4$

(B) Subtract the mean with respective row values.

Q3) dB : Song Vector Matrix & Preference of Alice. Content based Filtering to Rec song to Alice. Cosine similarity.

dB	Song	Tempo	Duration	Loudness	Danceability	Energy
(SA)	Song A	0	1	0	0.6	0.8
(SB)	Song B	0	0	1	0.5	0.7
(SC)	Song C	0	0	1	0.5	0.7
(UP)	Alice Pref	0	0	0	0.8	0.7

⇒ Formula : $\frac{P_1 \cdot P_2}{\|P_1\| \cdot \|P_2\|}$

$$\cos(UP, SA) = \frac{0.6 \times 0.8 + 0.8 \times 0.7}{\sqrt{1^2 + 0.6^2 + 0.8^2} \times \sqrt{0.8^2 + 0.7^2}} = 0.42$$

$$\cos(UP, SB) = \frac{0.5 \times 0.8 + 0.7 \times 0.7}{\sqrt{1^2 + 0.5^2 + 0.7^2} \times \sqrt{0.8^2 + 0.7^2}} = 0.373$$

$$\cos(UP, SC) = \frac{0.5 \times 0.8 + 0.7 \times 0.7}{\sqrt{1^2 + 0.5^2 + 0.7^2} \times \sqrt{0.8^2 + 0.7^2}} = 0.373$$

∵ the similarity of Alice perference & Song A is more (0.42),
∴ It is more likely to be recommended to Alice.