

item-based collaborative filtering

Data Matrix:

User	Book X	Book Y	Book Z
U1	4	5	?
U2	5	3	4
U3	2	4	3

Predict how U1 would rate Book Z.

Book X and Z:

- Ratings: $X = [5, 2]$, $Z = [4, 3]$ (U2 and U3)
- Numerator = $(5 \times 4) + (2 \times 3) = 20 + 6 = 26$
- Denominator = $\sqrt{(25 + 4)} \times \sqrt{(16 + 9)} = \sqrt{29} \times \sqrt{25} \approx 5.385 \times 5 = 26.93$
- Cosine similarity $\approx 26 / 26.93 \approx \mathbf{0.965}$

Book Y and Z:

- Ratings: $Y = [3, 4]$, $Z = [4, 3]$ (U2 and U3)
- Numerator = $(3 \times 4) + (4 \times 3) = 12 + 12 = 24$
- Denominator = $\sqrt{(9 + 16)} \times \sqrt{(16 + 9)} = \sqrt{25} \times \sqrt{25} = 5 \times 5 = 25$
- Cosine similarity = $24 / 25 = \mathbf{0.96}$

U1's known ratings:

- Book X: 4
- Book Y: 5

Prediction:

$$\frac{(4 \times 0.965) + (5 \times 0.96)}{0.965 + 0.96} = \frac{3.86 + 4.8}{1.925} \approx \frac{8.66}{1.925} \approx \boxed{4.5}$$

User	Show 1	Show 2	Show 3
U1	2	?	4
U2	3	4	5
U3	1	3	2

Predict how **U1** would rate **Show 2**.

Show 2 vs Show 1:

- Ratings: Show 1 = [3, 1], Show 2 = [4, 3] (U2, U3)
- Numerator = $(3 \times 4) + (1 \times 3) = 12 + 3 = 15$
- Denominator = $\sqrt{(9 + 1)} \times \sqrt{(16 + 9)} = \sqrt{10} \times \sqrt{25} \approx 3.16 \times 5 = 15.8$
- Cosine similarity $\approx 15 / 15.8 \approx \mathbf{0.949}$

Show 2 vs Show 3:

- Ratings: Show 3 = [5, 2], Show 2 = [4, 3] (U2, U3)
- Numerator = $(5 \times 4) + (2 \times 3) = 20 + 6 = 26$
- Denominator = $\sqrt{(25 + 4)} \times \sqrt{(16 + 9)} = \sqrt{29} \times \sqrt{25} \approx 5.385 \times 5 = 26.93$
- Cosine similarity $\approx 26 / 26.93 \approx \mathbf{0.965}$

Known ratings:

- Show 1: 2
- Show 3: 4

$$\frac{(2 \times 0.949) + (4 \times 0.965)}{0.949 + 0.965} = \frac{1.898 + 3.86}{1.914} \approx \frac{5.758}{1.914} \approx \boxed{3.0}$$

User	Apples	Bananas	Cherries
U1	5	2	?
U2	3	1	4
U3	4	2	5

Predict U1's rating for **Cherries**

Similarity (Cherries, Apples):

Apples: [3, 4], Cherries: [4, 5] (U2, U3)

$$\text{Cosine Numerator} = 3 \times 4 + 4 \times 5 = 12 + 20 = 32$$

$$\text{Denominator} = \sqrt{(9+16)} \times \sqrt{(16+25)} = \sqrt{25} \times \sqrt{41} \approx 5 \times 6.4 \approx 32$$

$$\text{Cosine similarity} \approx 1.00$$

Similarity (Cherries, Bananas):

Bananas: [1, 2], Cherries: [4, 5]

$$\text{Cosine Numerator} = 1 \times 4 + 2 \times 5 = 4 + 10 = 14$$

$$\text{Denominator} = \sqrt{(1+4)} \times \sqrt{(16+25)} = \sqrt{5} \times \sqrt{41} \approx 2.24 \times 6.4 \approx 14.3$$

$$\text{Cosine similarity} \approx 0.98$$

U1 rated Apples = 5, Bananas = 2

$$\text{Prediction} = \frac{(5 \times 1.00) + (2 \times 0.98)}{1.00 + 0.98} = \frac{5 + 1.96}{1.98} = \frac{6.96}{1.98} \approx \boxed{3.52}$$

User	FIFA	Minecraft	Zelda
U1	?	3	4
U2	5	2	3
U3	4	1	2

Similarity (FIFA, Minecraft):

FIFA = [5, 4], Minecraft = [2, 1]

Cosine Numerator = $5 \times 2 + 4 \times 1 = 10 + 4 = 14$

Denominator = $\sqrt{(25+16)} \times \sqrt{(4+1)} = \sqrt{41} \times \sqrt{5} \approx 6.4 \times 2.24 \approx 14.3$

Cosine similarity ≈ 0.98

Similarity (FIFA, Zelda):

FIFA = [5, 4], Zelda = [3, 2]

Cosine Numerator = $5 \times 3 + 4 \times 2 = 15 + 8 = 23$

Denominator = $\sqrt{41} \times \sqrt{(9+4)} = \sqrt{41} \times \sqrt{13} \approx 6.4 \times 3.6 \approx 23$

Cosine similarity ≈ 1.00

U1 ratings: Minecraft = 3, Zelda = 4

$$\text{Prediction} = \frac{(3 \times 0.98) + (4 \times 1.00)}{0.98 + 1.00} = \frac{2.94 + 4.00}{1.98} = \frac{6.94}{1.98} \approx \boxed{3.51}$$

User	Pizza	Pasta	Sushi	Burger
U1	5	4	?	3
U2	3	2	4	4
U3	4	3	5	2
U4	2	2	3	1

Predict **U1's rating for Sushi**

Sushi & Pizza:

Sushi = [4, 5, 3], Pizza = [3, 4, 2] (U2, U3, U4)

$$\text{Numerator} = 3 \times 4 + 4 \times 5 + 2 \times 3 = 12 + 20 + 6 = 38$$

$$\text{Denominator} = \sqrt{(9+16+4)} \times \sqrt{(16+25+9)} = \sqrt{29} \times \sqrt{50} \approx 5.38 \times 7.07 \approx 38.03$$

$$\text{Cosine} \approx 38 / 38.03 \approx \mathbf{1.00}$$

Sushi & Pasta:

Pasta = [2, 3, 2], Sushi = [4, 5, 3] (U2, U3, U4)

$$\text{Numerator} = 2 \times 4 + 3 \times 5 + 2 \times 3 = 8 + 15 + 6 = 29$$

$$\text{Denominator} = \sqrt{(4+9+4)} \times \sqrt{(16+25+9)} = \sqrt{17} \times \sqrt{50} \approx 4.12 \times 7.07 \approx 29.15$$

$$\text{Cosine} \approx 29 / 29.15 \approx \mathbf{0.995}$$

Sushi & Burger:

Burger = [4, 2, 1], Sushi = [4, 5, 3]

$$\text{Numerator} = 4 \times 4 + 2 \times 5 + 1 \times 3 = 16 + 10 + 3 = 29$$

$$\text{Denominator} = \sqrt{(16+4+1)} \times \sqrt{(16+25+9)} = \sqrt{21} \times \sqrt{50} \approx 4.58 \times 7.07 \approx 32.38$$

$$\text{Cosine} \approx 29 / 32.38 \approx \mathbf{0.896}$$

Known U1 ratings:

- Pizza = 5
- Pasta = 4
- Burger = 3

$$\text{Prediction} = \frac{(5 \times 1.00) + (4 \times 0.995) + (3 \times 0.896)}{1.00 + 0.995 + 0.896} = \frac{5 + 3.98 + 2.69}{2.891} = \frac{11.67}{2.891} \approx \boxed{4.04}$$

User	Rock	Jazz	EDM	Classical
U1	4	3	5	?
U2	5	2	4	4
U3	2	4	1	5
U4	3	3	2	3

Predict **U1's** rating for **Classical**

Classical & Rock:

$$\text{Rock} = [5, 2, 3], \text{Classical} = [4, 5, 3]$$

$$\text{Numerator} = 5 \times 4 + 2 \times 5 + 3 \times 3 = 20 + 10 + 9 = 39$$

$$\text{Denominator} = \sqrt{(25+4+9)} \times \sqrt{(16+25+9)} = \sqrt{38} \times \sqrt{50} \approx 6.16 \times 7.07 \approx 43.56$$

$$\text{Cosine} \approx 39 / 43.56 \approx \mathbf{0.895}$$

Classical & Jazz:

$$\text{Jazz} = [2, 4, 3], \text{Classical} = [4, 5, 3]$$

$$\text{Numerator} = 2 \times 4 + 4 \times 5 + 3 \times 3 = 8 + 20 + 9 = 37$$

$$\text{Denominator} = \sqrt{(4+16+9)} \times \sqrt{(16+25+9)} = \sqrt{29} \times \sqrt{50} \approx 5.38 \times 7.07 \approx 38.02$$

$$\text{Cosine} \approx 37 / 38.02 \approx \mathbf{0.973}$$

Classical & EDM:

$$\text{EDM} = [4, 1, 2], \text{Classical} = [4, 5, 3]$$

$$\text{Numerator} = 4 \times 4 + 1 \times 5 + 2 \times 3 = 16 + 5 + 6 = 27$$

$$\text{Denominator} = \sqrt{(16+1+4)} \times \sqrt{(16+25+9)} = \sqrt{21} \times \sqrt{50} \approx 4.58 \times 7.07 \approx 32.38$$

$$\text{Cosine} \approx 27 / 32.38 \approx \mathbf{0.834}$$

- Rock = 4, Jazz = 3, EDM = 5

$$\text{Prediction} = \frac{(4 \times 0.895) + (3 \times 0.973) + (5 \times 0.834)}{0.895 + 0.973 + 0.834} = \frac{3.58 + 2.92 + 4.17}{2.702} = \frac{10.67}{2.702} \approx \boxed{3.95}$$

Users like or dislike articles on a news site (1 = liked, 0 = not interacted):

User	Article A	Article B	Article C	Article D
U1	1	1	0	?
U2	1	0	1	1
U3	1	1	1	0

Task:

Predict whether U1 would like **Article D** using **Jaccard similarity** between articles.

Jaccard similarity between two binary vectors (items) is:

$$\text{Jaccard}(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Where:

- $|A \cap B|$ = number of users who liked both articles
- $|A \cup B|$ = number of users who liked either article

Get U1's liked items

U1 liked:

- **Article A**
- **Article B**

We will compare **Article D** with these two.

Create Binary Vectors for Articles

- Article A = [1, 1] (liked by U2 and U3)
- Article B = [0, 1]
- Article D = [1, 0]

Compute Jaccard Similarities

Jaccard(D, A):

$$D = [1, 0], A = [1, 1]$$

$$\cap = [1 \ \& \ 1, 0 \ \& \ 1] = [1, 0] \rightarrow \text{Sum} = 1$$

$$\cup = [1 \ | \ 1, 0 \ | \ 1] = [1, 1] \rightarrow \text{Sum} = 2$$

$$J(D, A) = \frac{1}{2} = 0.5$$

Jaccard(D, B):

$$D = [1, 0], B = [0, 1]$$

$$\cap = [0, 0] = 0$$

$$\cup = [1, 1] = 2$$

$$J(D, B) = \frac{0}{2} = 0.0$$

U1 liked A and B, so we average the similarities of D with A and B:

$$\text{Avg Similarity} = \frac{J(D, A) + J(D, B)}{2} = \frac{0.5 + 0.0}{2} = 0.25$$

We define a **threshold**, e.g., **0.3**, to recommend or not:

- Since $0.25 < 0.3 \rightarrow$ **Do not recommend** Article D to U1

U1 is unlikely to like Article D, based on low average similarity with their liked articles (A and B).