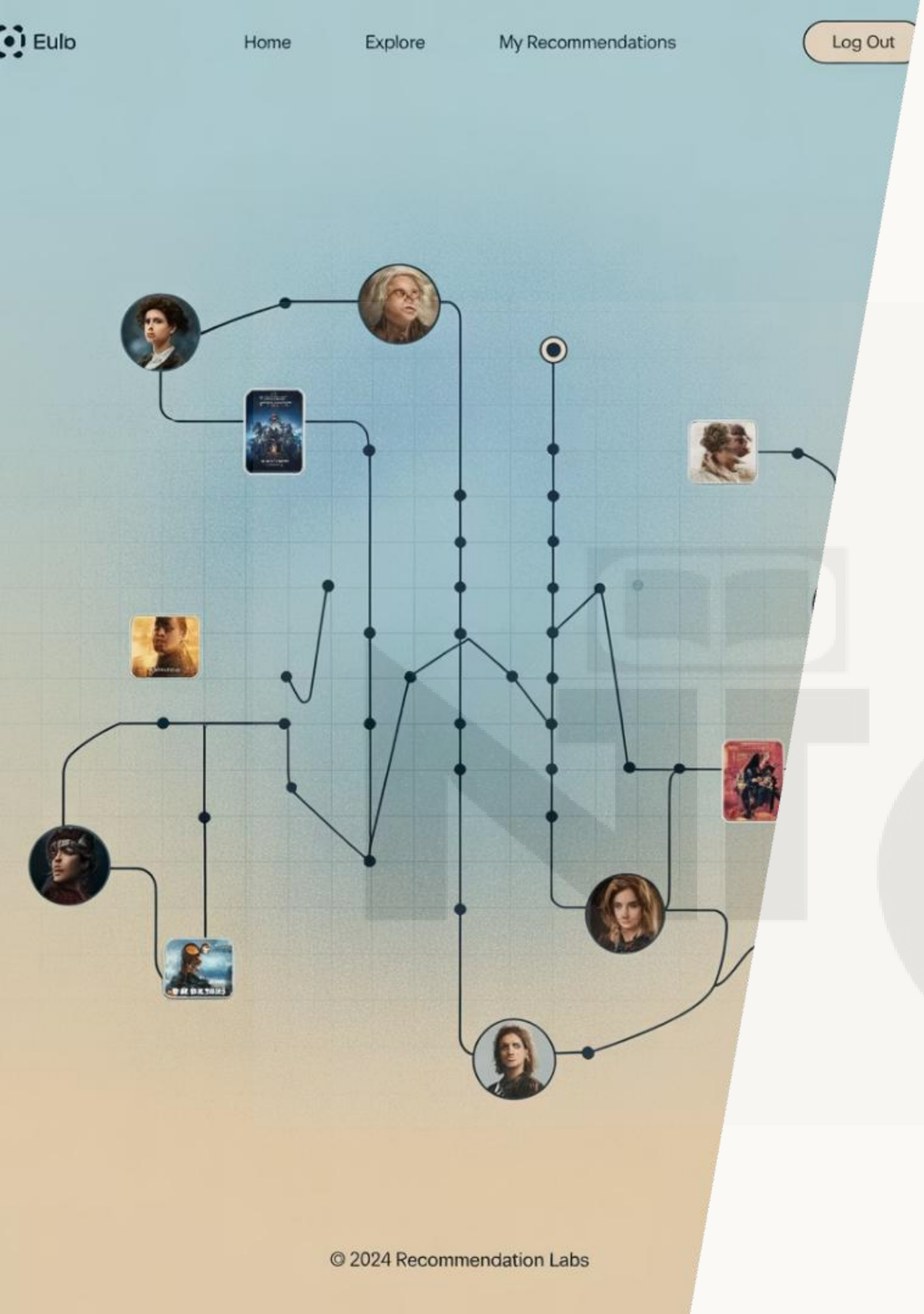


User-Based Collaborative Filtering

Step-by-step approach to predict user preferences based on similar users' ratings

MUKESH KUMAR

What is User-Based Collaborative Filtering?

Definition

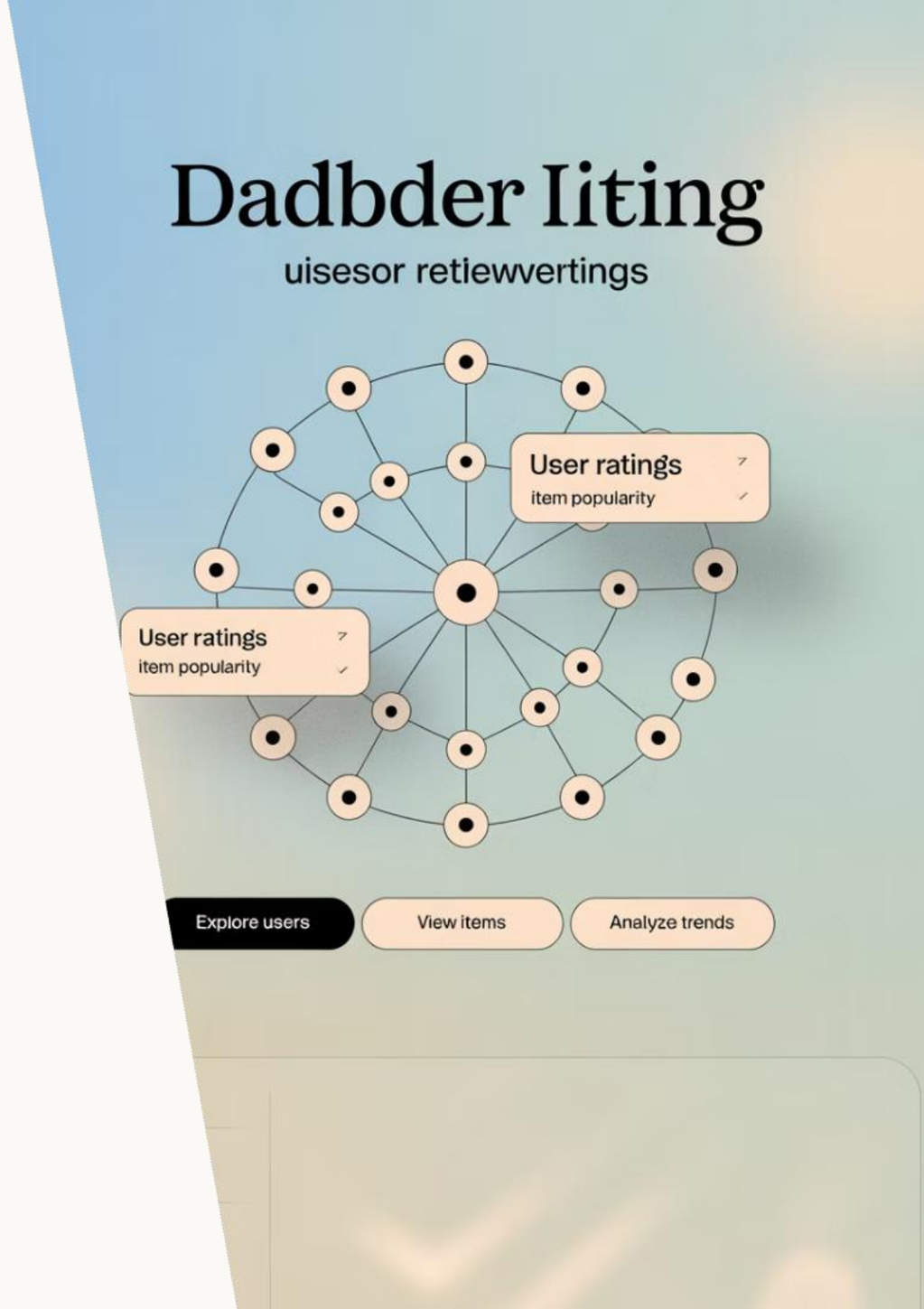
Recommends items based on similar users' preferences

Benefits

Discovers unexpected items users might enjoy

Applications

Movies, products, music, and content recommendations



Our Example Setup

4 Movies

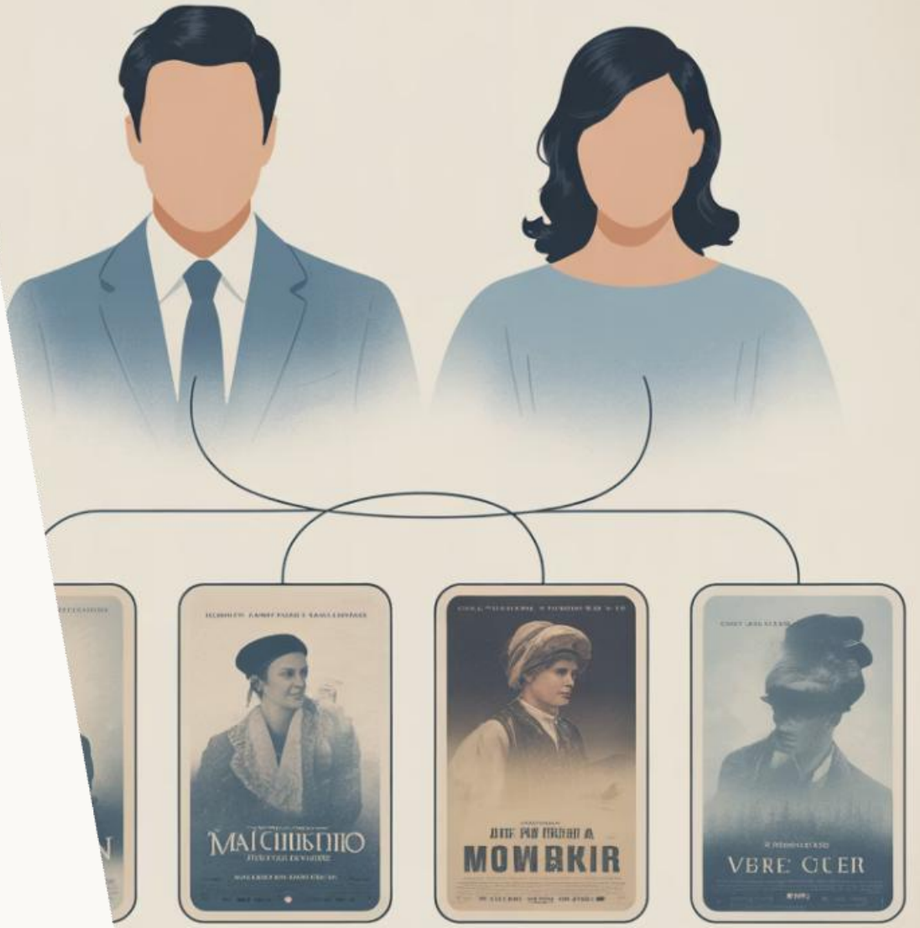
Movies A, B, C, and D

2 Users

User 1 and User 2 with ratings

Goal

Predict next movie for User 1



User-Item Rating Matrix

User / Movie	Movie A	Movie B	Movie C	Movie D
User 1	5	3	0	0
User 2	4	0	4	2

Step 1: Compute User Vectors Vectors

Identify vectors

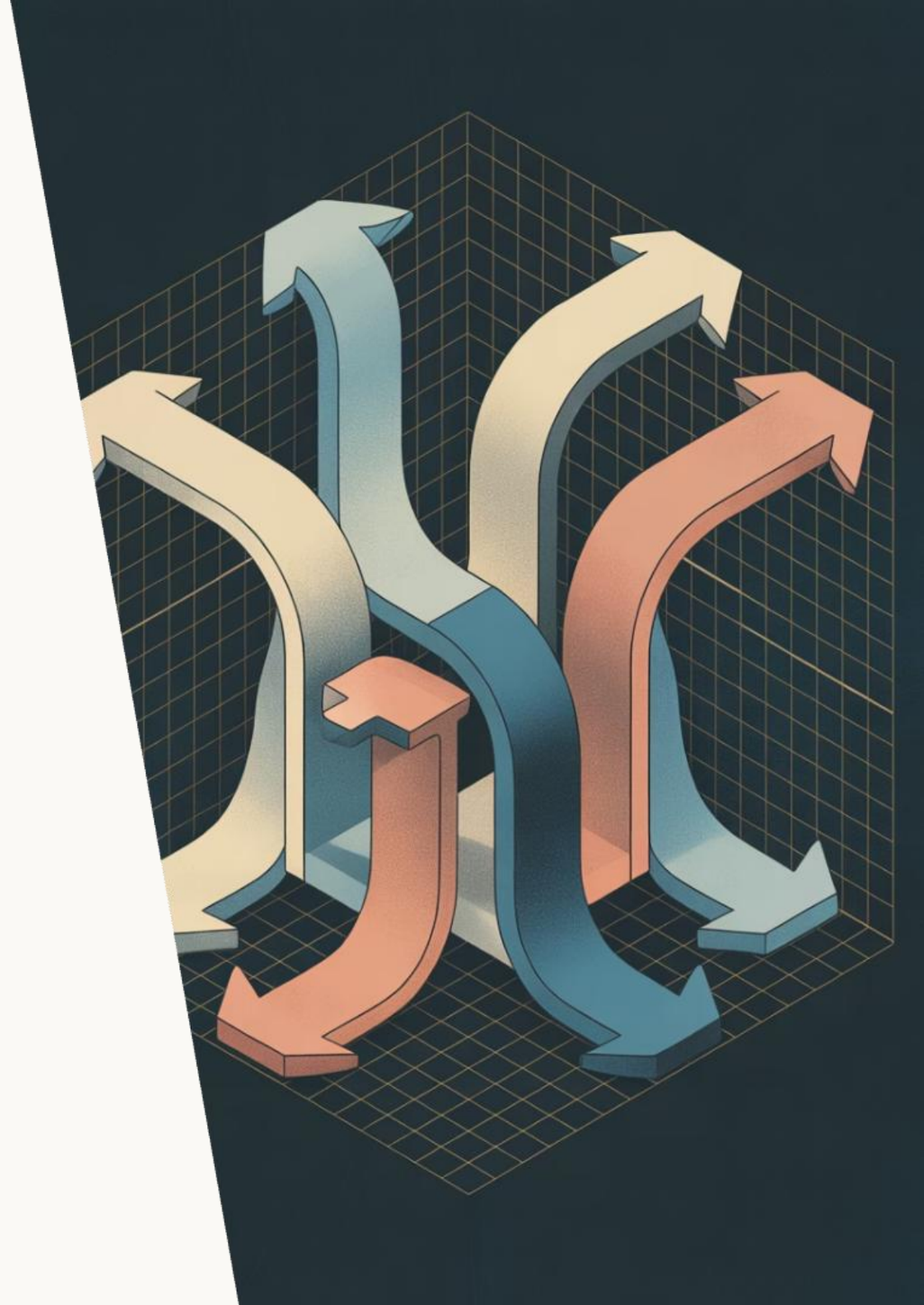
Extract ratings as numerical arrays

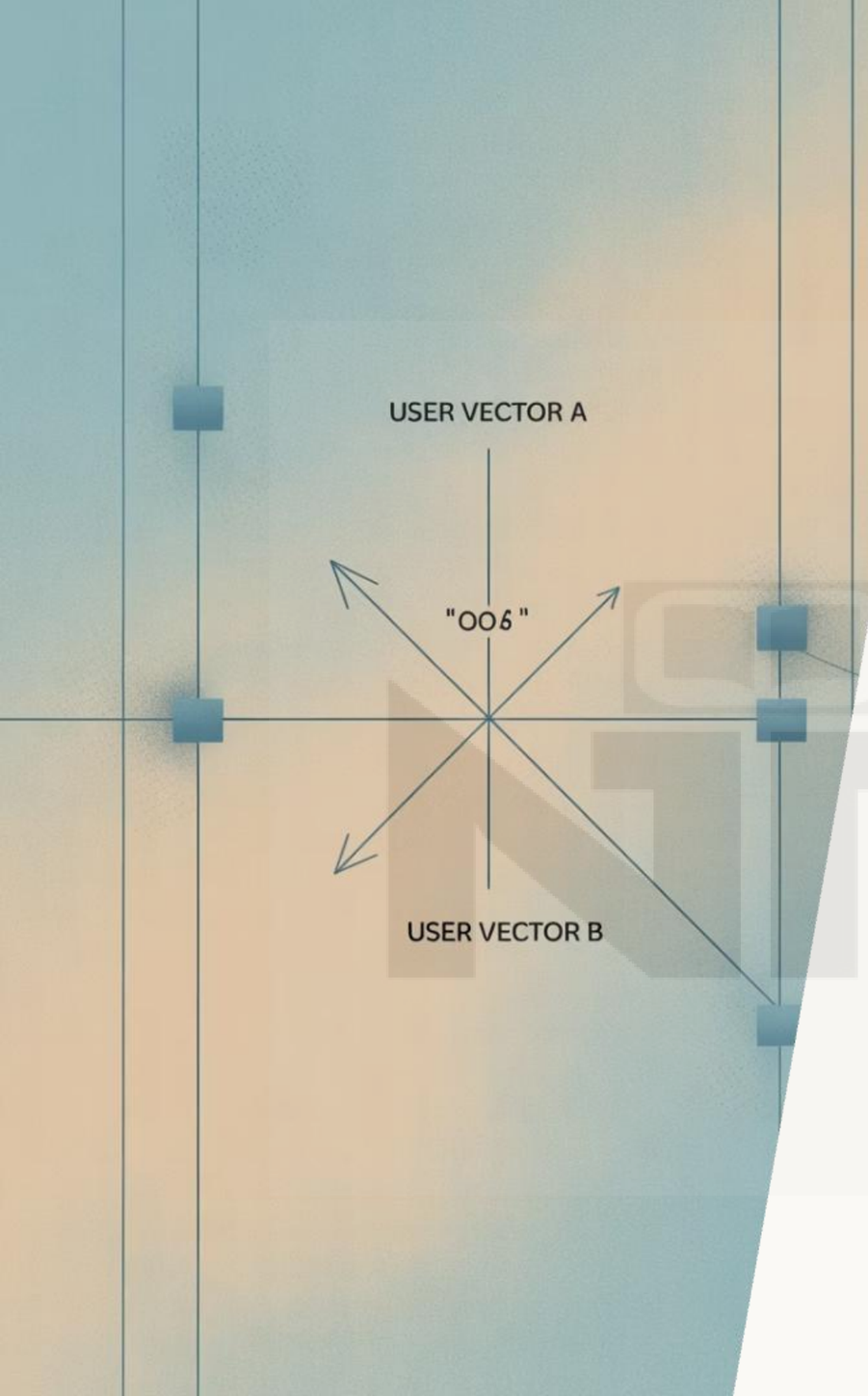
User 1 vector

[5, 3, 0, 0]

User 2 vector

[4, 0, 4, 2]





Step 2: Calculate Similarity

Formula:

$$\text{cosine}(U_1, U_2) = \frac{U_1 \cdot U_2}{||U_1|| \times ||U_2||}$$

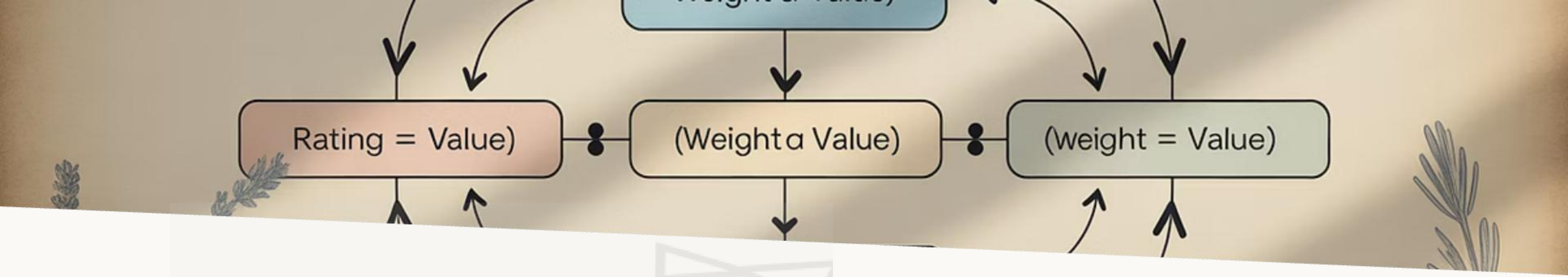
Calculation:

$$U_1 \cdot U_2 = (5 \times 4) + (3 \times 0) + (0 \times 4) + (0 \times 2) = 20$$

$$||U_1|| = \sqrt{5^2 + 3^2 + 0 + 0} = \sqrt{34} \approx 5.83$$

$$||U_2|| = \sqrt{4^2 + 0 + 4^2 + 2^2} = \sqrt{16 + 0 + 16 + 4} = \sqrt{36} = 6$$

$$\text{cosine}(U_1, U_2) = \frac{20}{5.83 \times 6} \approx \frac{20}{34.98} \approx 0.572$$



Step 3: Predict Ratings

Use similarity-weighted sum of ratings from similar users (only User 2 here):

$$\text{PredictedRating}_{U1,MovieX} = \frac{\sum_{\text{neighbors}} \text{Similarity}(U1, U_i) \times \text{Rating}(U_i, MovieX)}{\sum_{\text{neighbors}} \text{Similarity}(U1, U_i)}$$

- For Movie C:

$$\frac{0.572 \times 4}{0.572} = 4$$

- For Movie D:

$$\frac{0.572 \times 2}{0.572} = 2$$

Final Recommendation

4.0

Movie C Rating

Highest predicted score

2.0

Movie D Rating

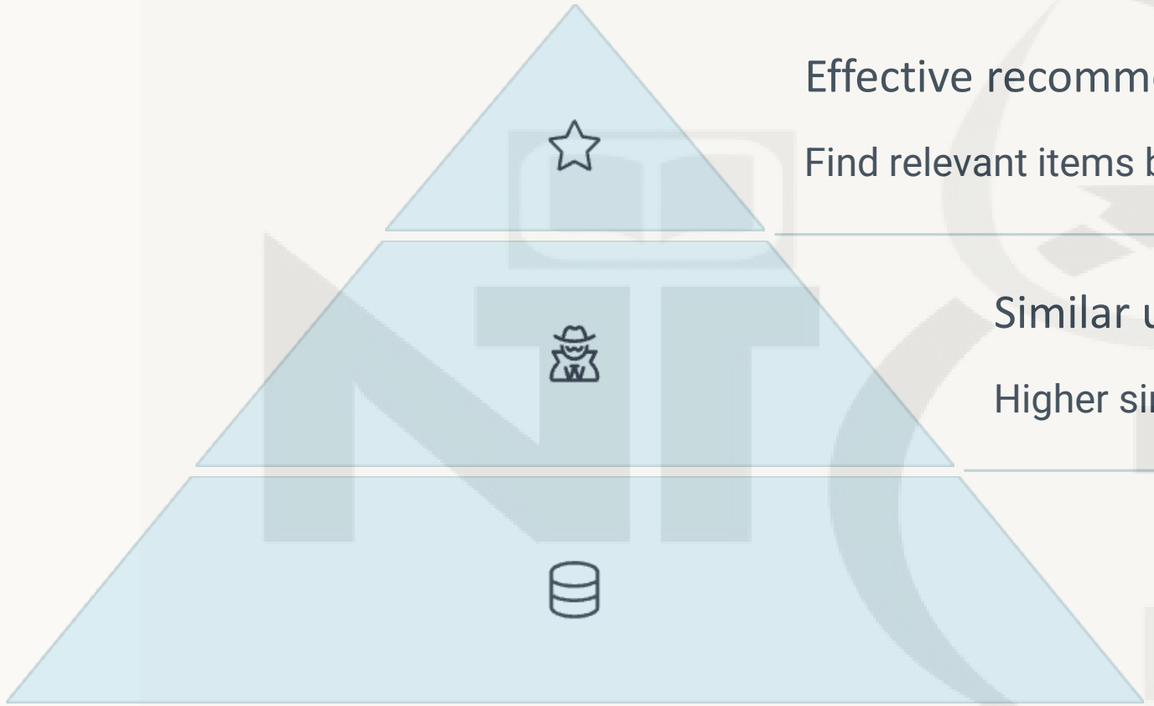
Lower predicted interest

1

Final Pick

Recommend Movie C to User 1

Key Takeaways



Effective recommendations

Find relevant items based on similar users

Similar users matter

Higher similarity = stronger influence

Sufficient data needed

Requires overlapping preferences to work well