Emerging Technologies

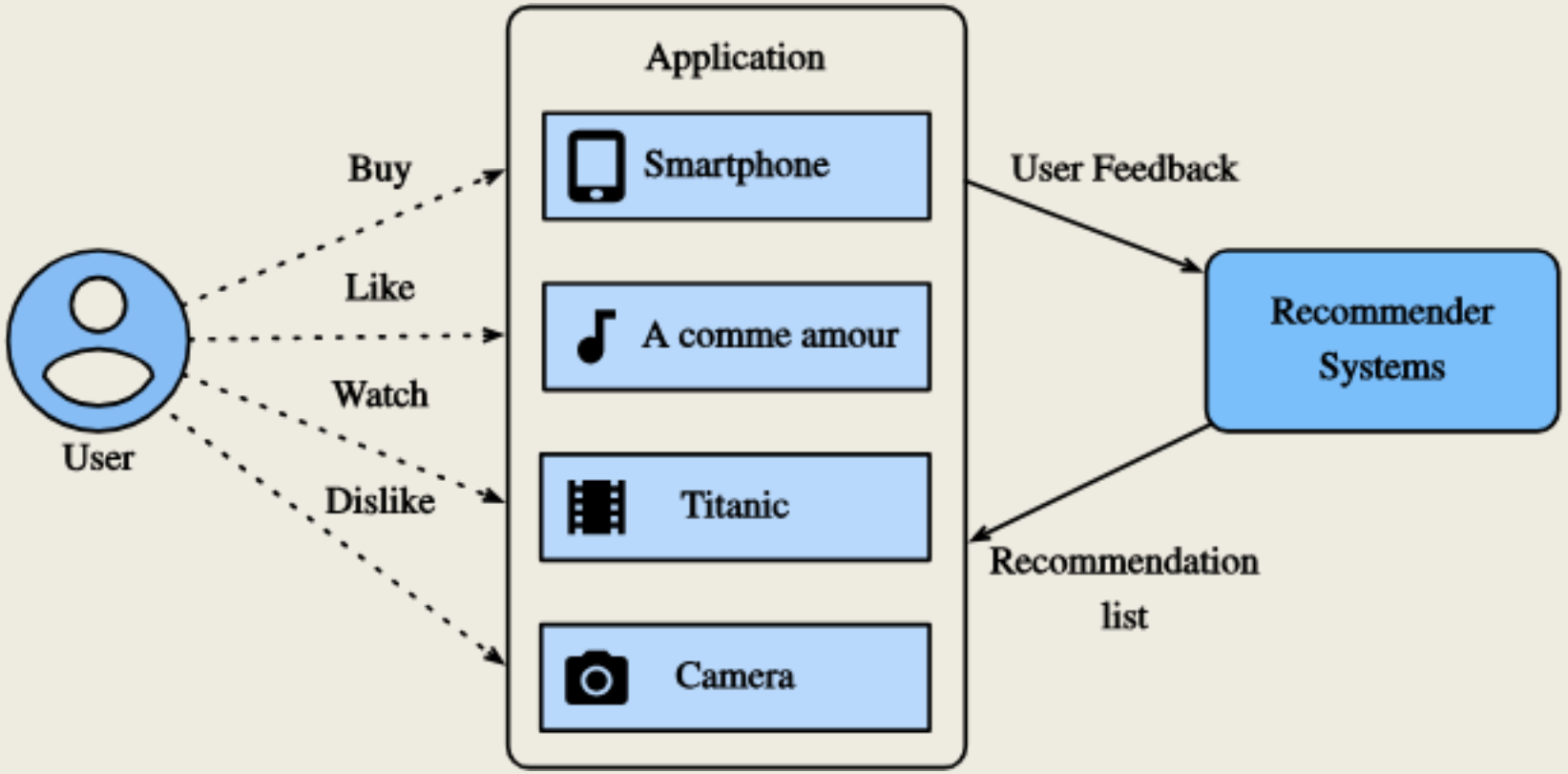
X.02. Recommendation Systems

LECTURER: NGUYỄN NGỌC TÚ

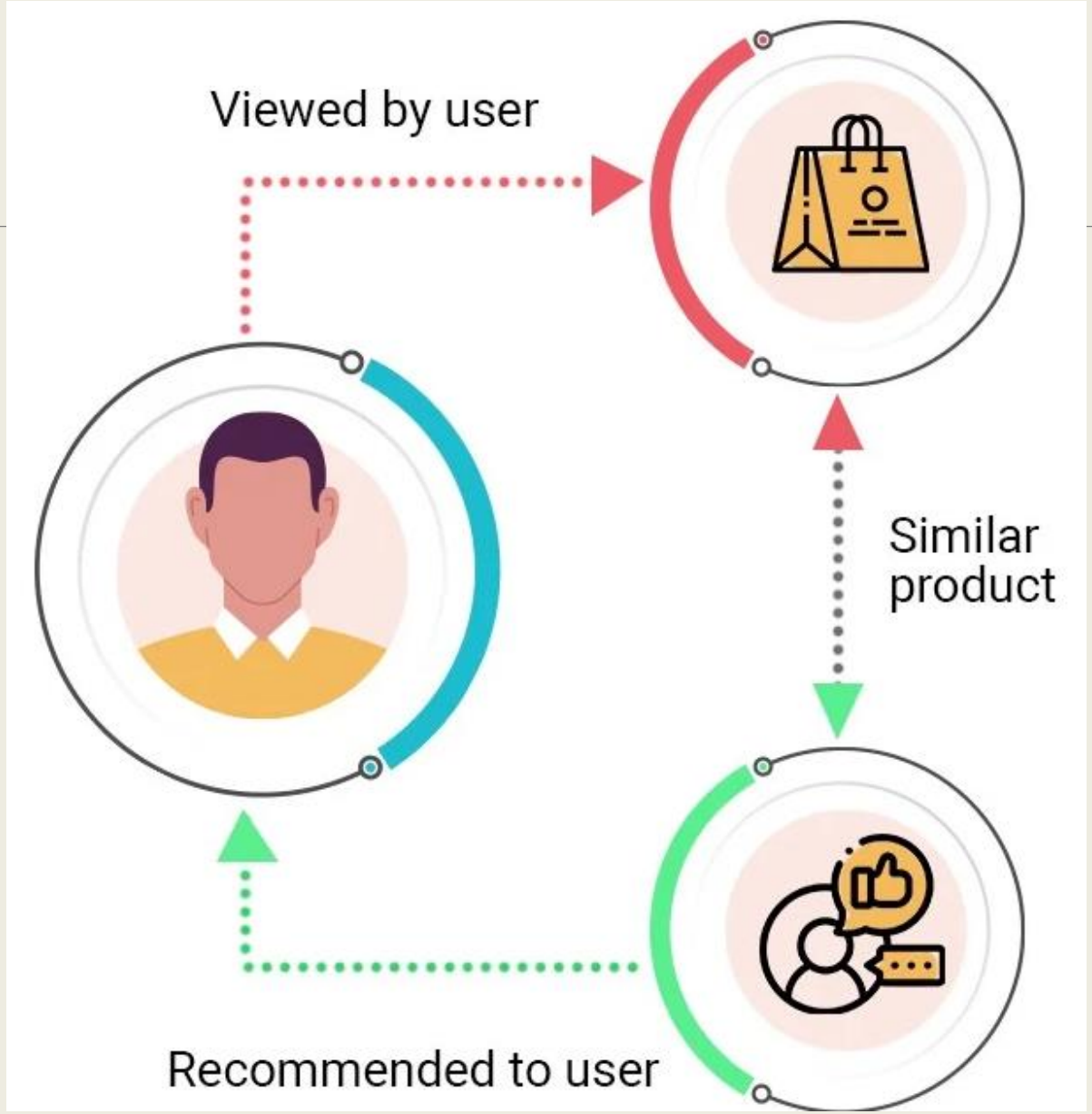
FACEGROUP: ATD.EMERGINGTECHNOLOGIES

RS – Recommendation System

Recommender Systems



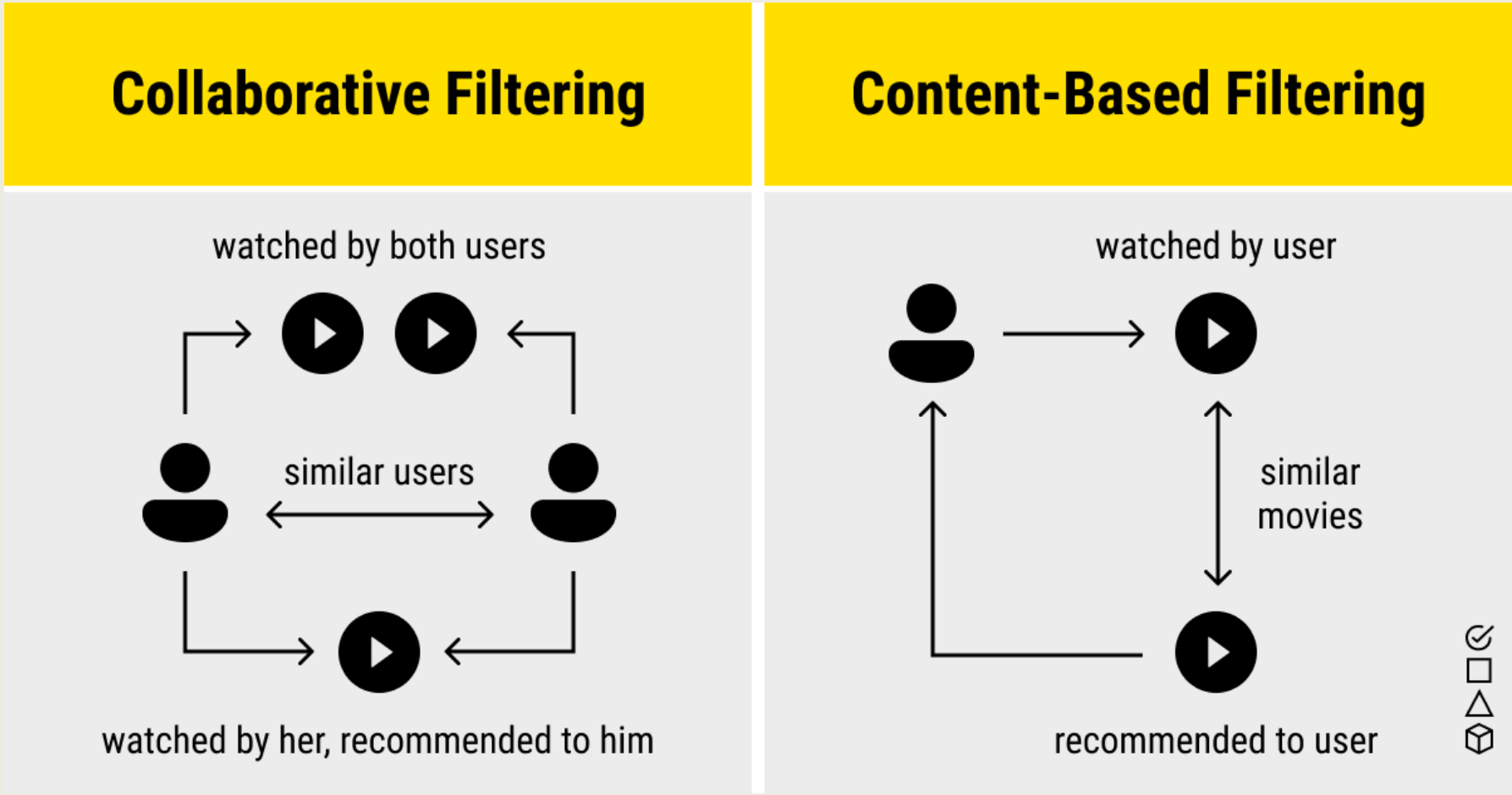
RS – Flow



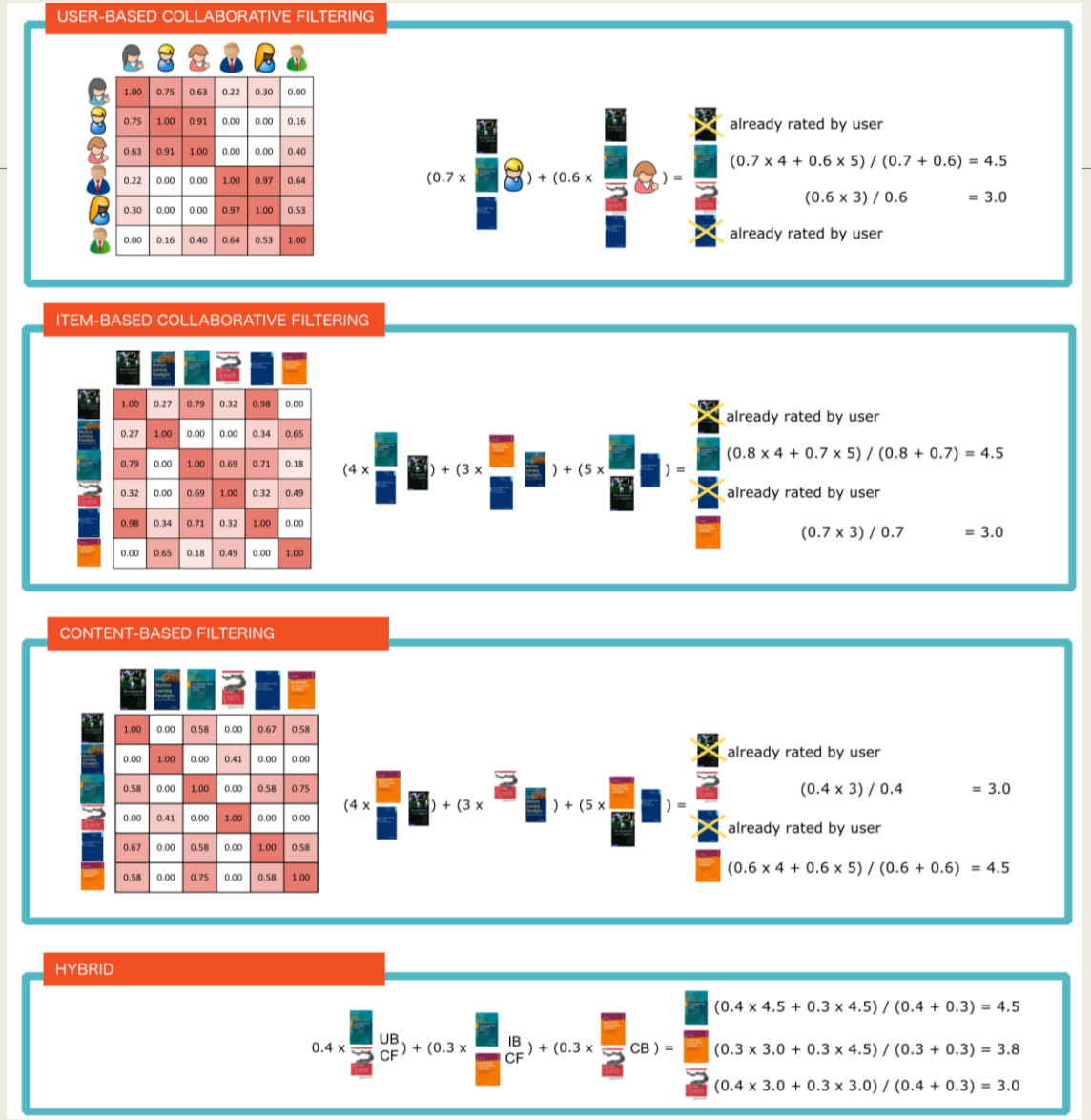
RS – Flow

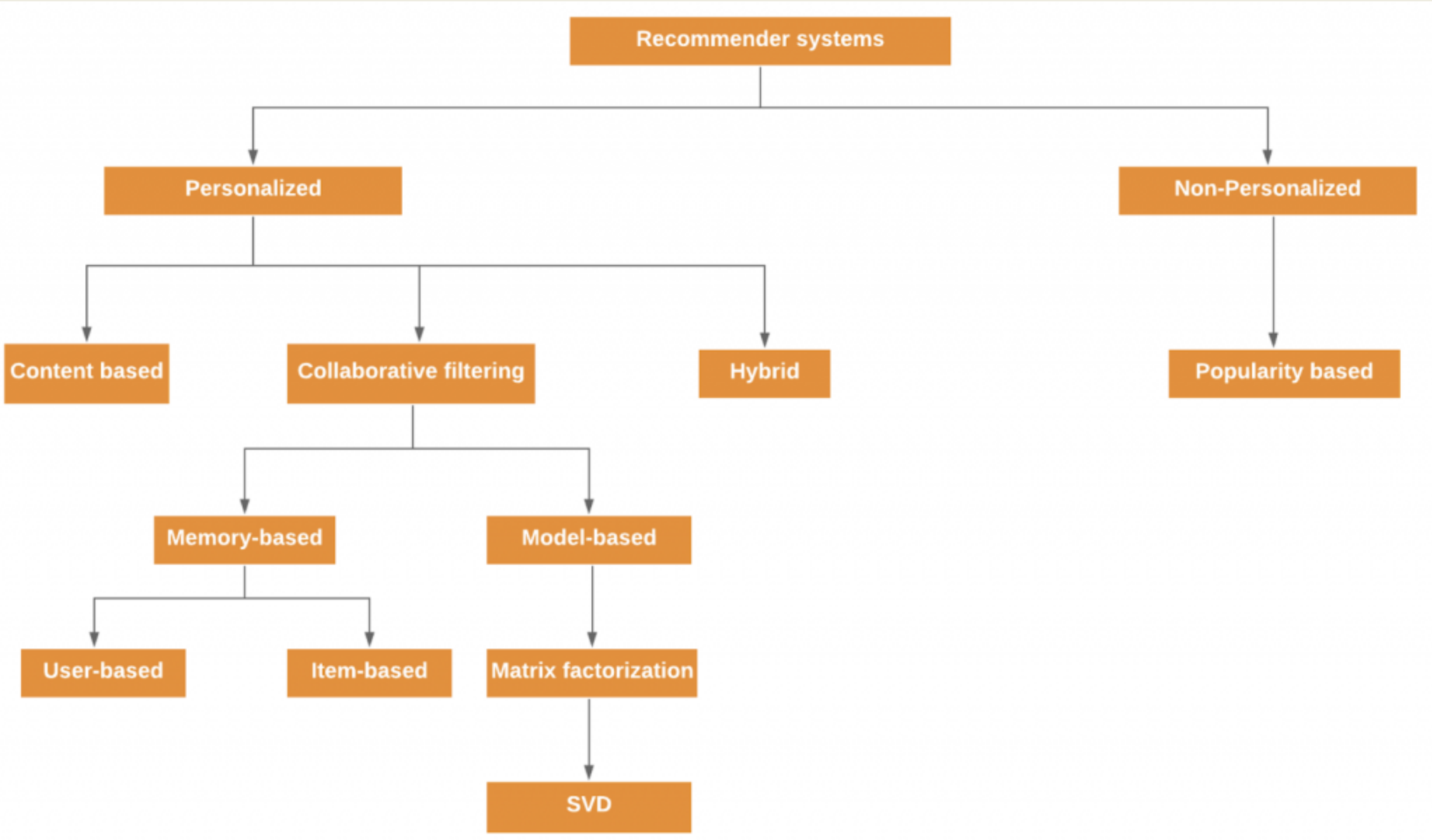
Flow for recommendation system using web scrapping

Types

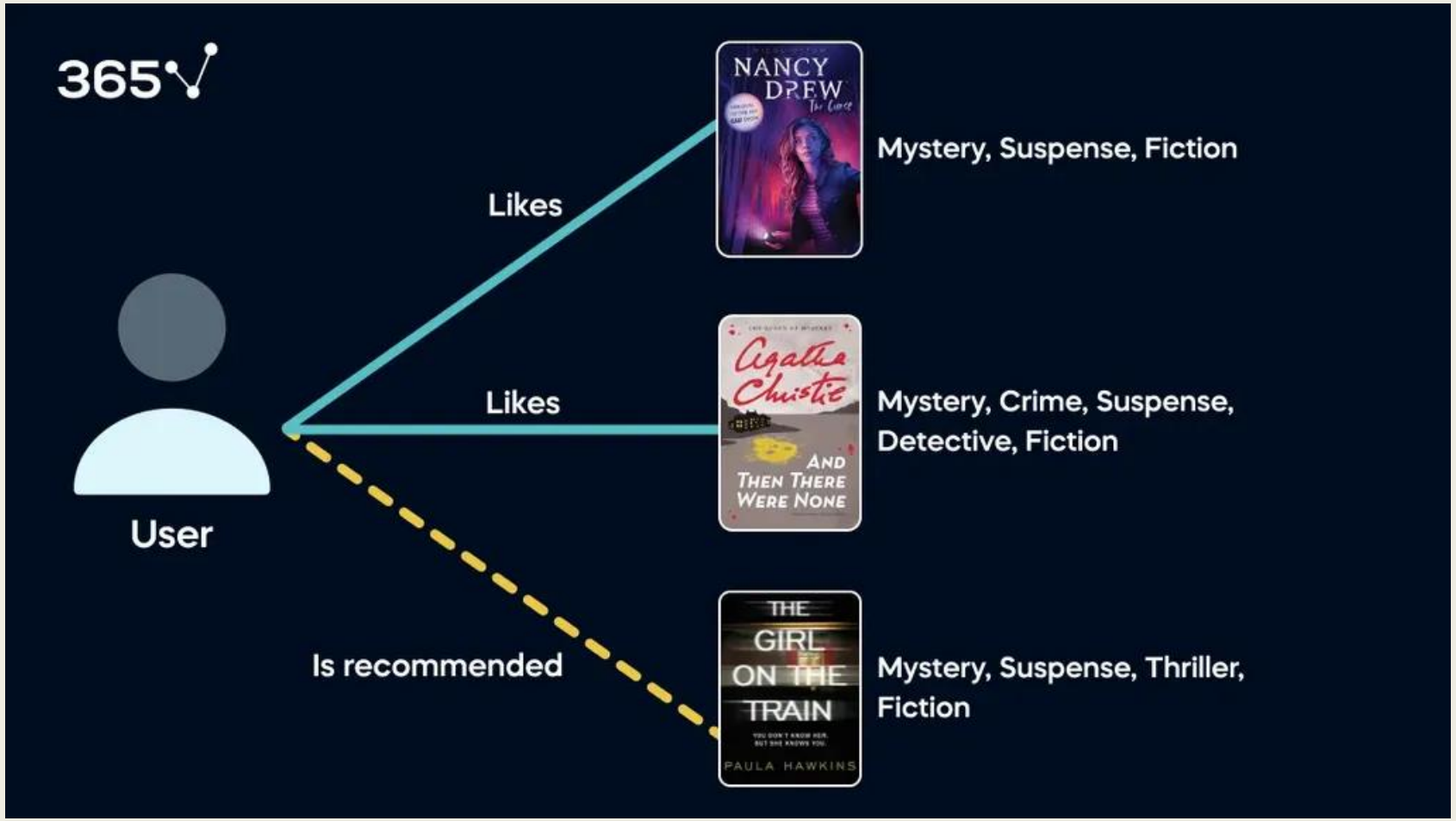

Collaborative Filtering
Content-Based Filtering
watched by both users
similar users
watched by her, recommended to him
watched by user
similar
movies
recommended to user

Algorithms

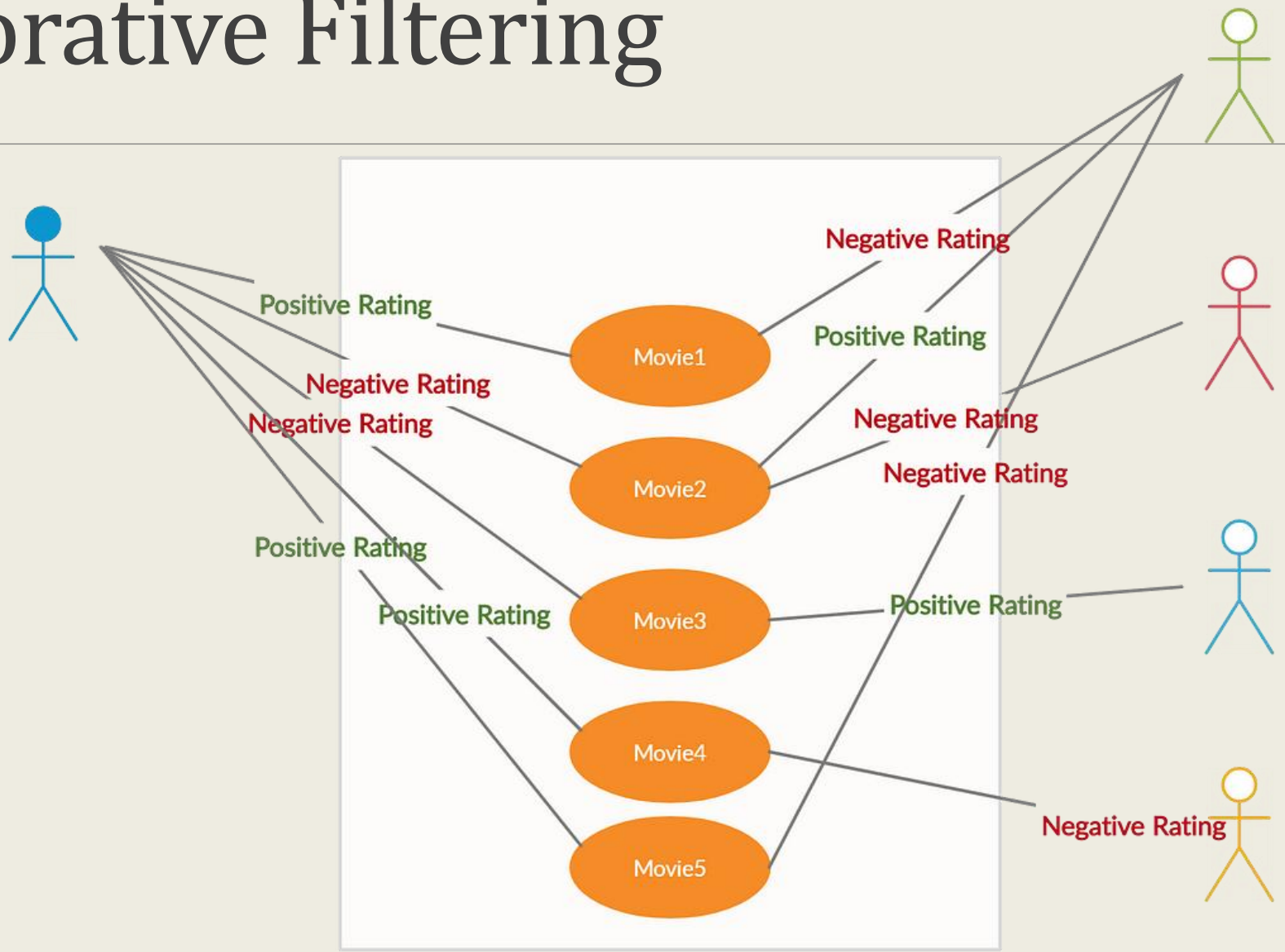

USER-BASED COLLABORATIVE FILTERING
1.00 0.75 0.63 0.22 0.30 0.00
0.75 1.00 0.91 0.00 0.00 0.16
0.63 0.91 1.00 0.00 0.00 0.40
0.22 0.00 0.00 1.00 0.97 0.64
0.30 0.00 0.00 0.97 1.00 0.53
0.00 0.16 0.40 0.64 0.53 1.00
(0.7 x) + (0.6 x) =
already rated by user
(0.7 x 4 + 0.6 x 5) / (0.7 + 0.6) = 4.5
(0.6 x 3) / 0.6 = 3.0
already rated by user
ITEM-BASED COLLABORATIVE FILTERING
1.00 0.27 0.79 0.32 0.98 0.00
0.27 1.00 0.00 0.00 0.34 0.65
0.79 0.00 1.00 0.69 0.71 0.18
0.32 0.00 0.69 1.00 0.32 0.49
0.98 0.34 0.71 0.32 1.00 0.00
0.00 0.65 0.18 0.49 0.00 1.00
(4 x) + (3 x) + (5 x) =
already rated by user
(0.8 x 4 + 0.7 x 5) / (0.8 + 0.7) = 4.5
already rated by user
(0.7 x 3) / 0.7 = 3.0
CONTENT-BASED FILTERING
1.00 0.00 0.58 0.00 0.67 0.58
0.00 1.00 0.00 0.41 0.00 0.00
0.58 0.00 1.00 0.00 0.58 0.75
0.00 0.41 0.00 1.00 0.00 0.00
0.67 0.00 0.58 0.00 1.00 0.58
0.58 0.00 0.75 0.00 0.58 1.00
(4 x) + (3 x) + (5 x) =
already rated by user
(0.4 x 3) / 0.4 = 3.0
already rated by user
(0.6 x 4 + 0.6 x 5) / (0.6 + 0.6) = 4.5
HYBRID
0.4 x UB CF) + (0.3 x IB CF) + (0.3 x CB) =
(0.4 x 4.5 + 0.3 x 4.5) / (0.4 + 0.3) = 4.5
(0.3 x 3.0 + 0.3 x 4.5) / (0.3 + 0.3) = 3.8
(0.4 x 3.0 + 0.3 x 3.0) / (0.4 + 0.3) = 3.0


Recommender systems
Personalized
Non-Personalized
Content based
Collaborative filtering
Hybrid
Popularity based
Memory-based
Model-based
User-based
Item-based
Matrix factorization
SVD

Content-Based Filtering



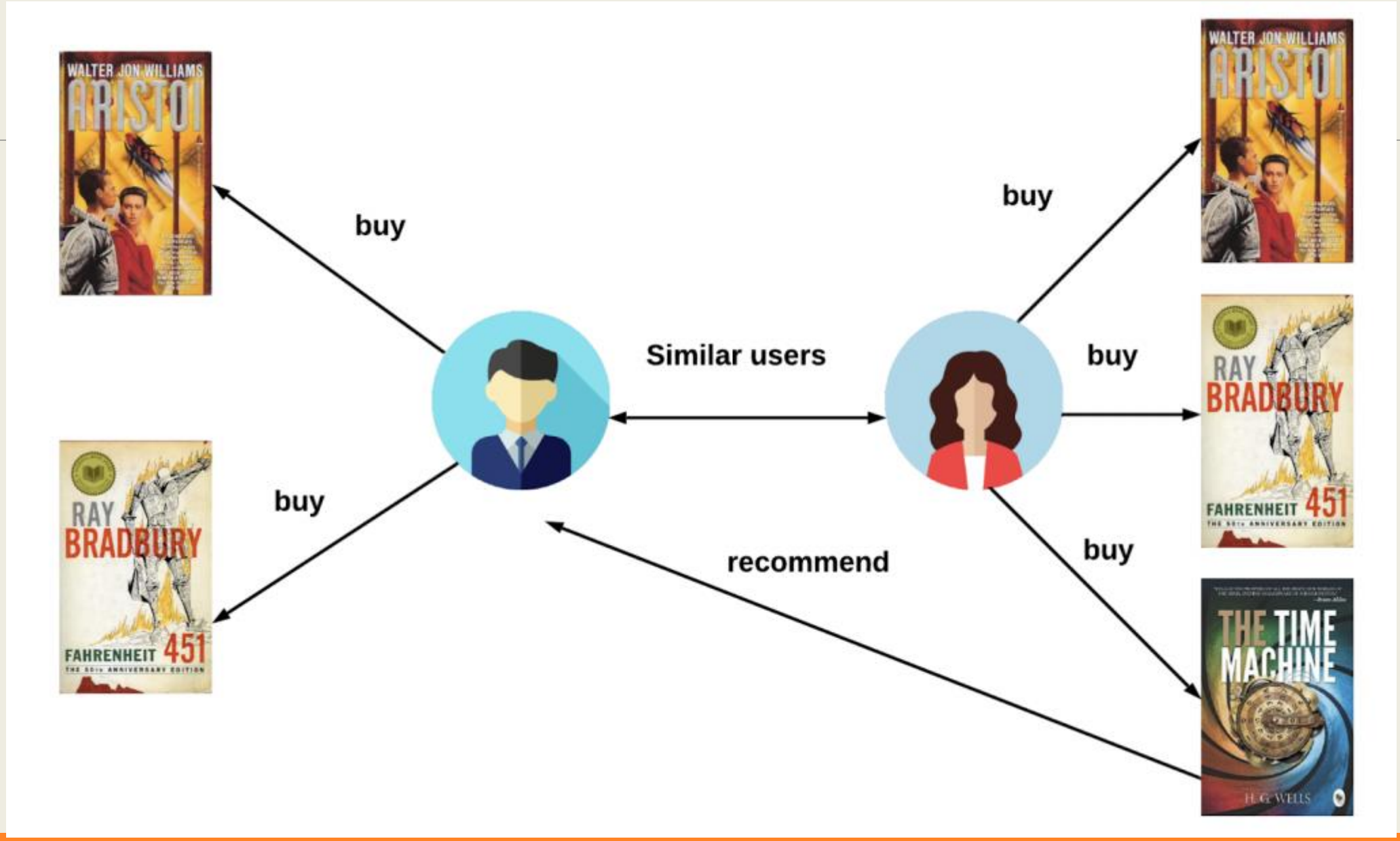
Collaborative Filtering



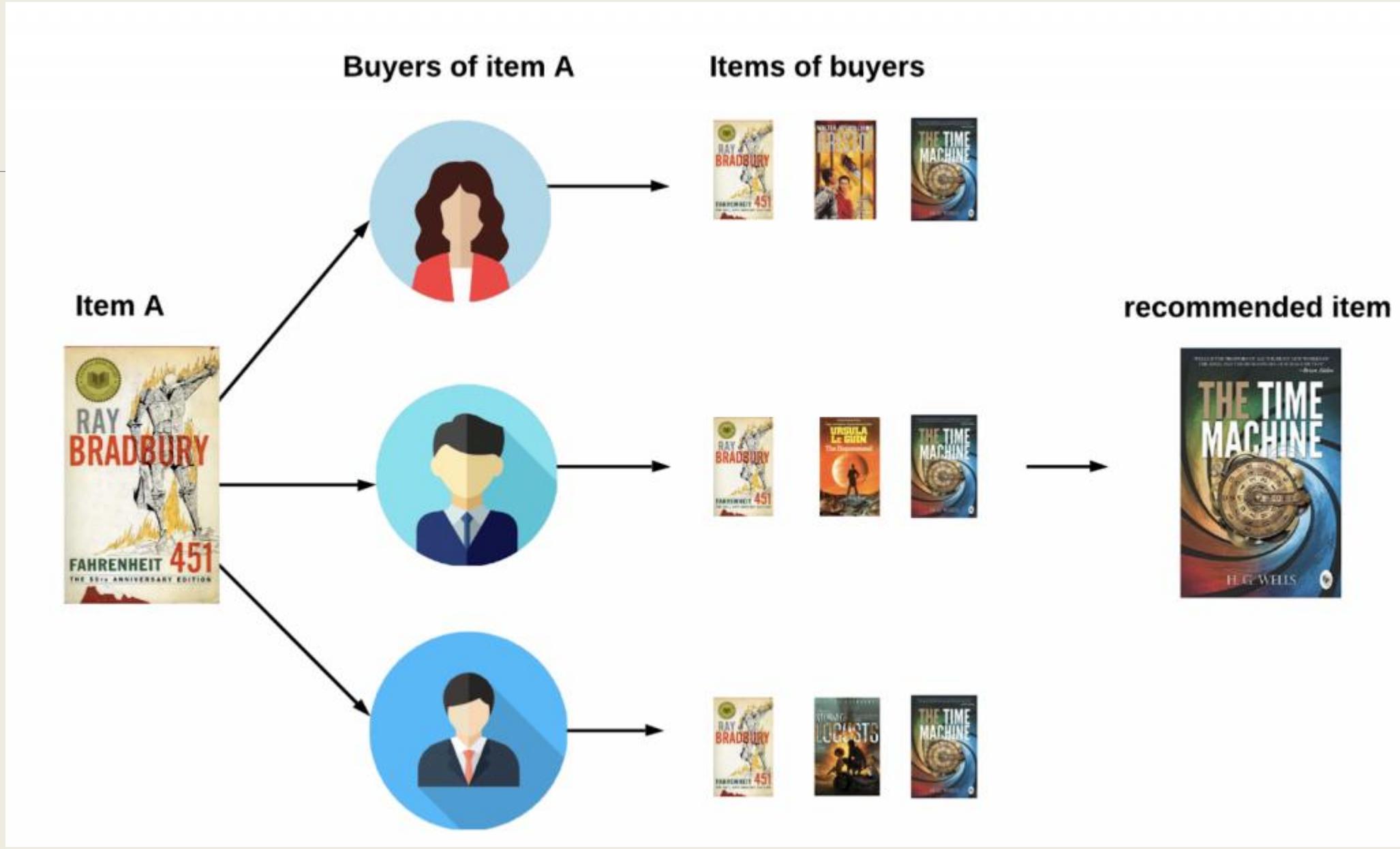
Collaborative Filtering – Teks

- Non-probabilistic algorithm
  - User-based nearest neighbor.
  - Item-based nearest neighbor.
  - Reducing dimensionality.
- Probabilistic algorithm
  - Bayesian-network model.
  - EM algorithm.
- …

Ex.


WALTER JON WILLIAMS
ARISTOI
buy
RAY
BRADBURY
FAHRENHEIT 451
buy
Similar users
buy
WALTER JON WILLIAMS
ARISTOI
buy
RAY
BRADBURY
FAHRENHEIT 451
buy
recommend
THE TIME MACHINE
H.G. WELLS

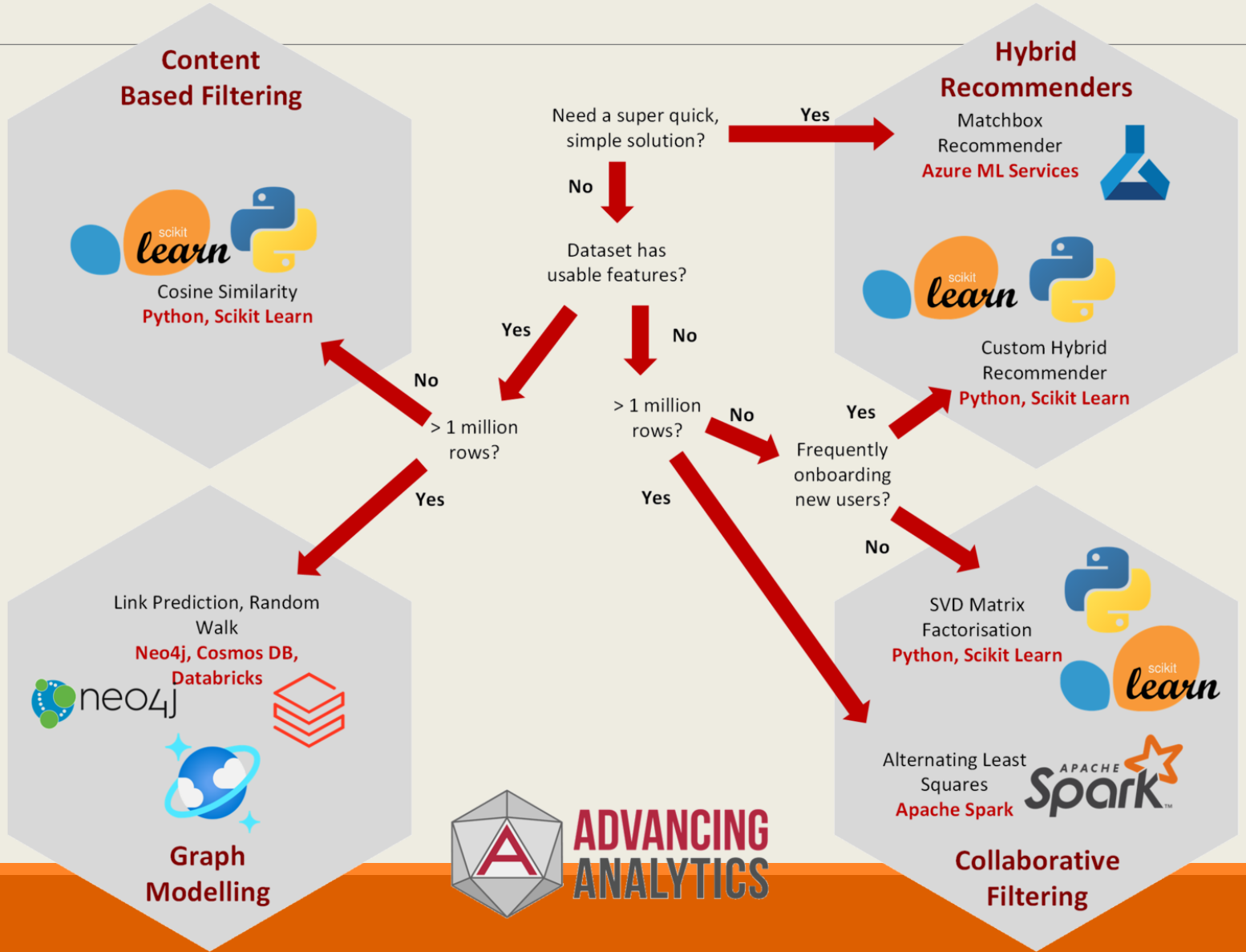
Ex.


Buyers of item A
Items of buyers
Item A
recommended item
RAY BRADBURY
FAHRENHEIT 451
THE TIME MACHINE
H.G. WELLS

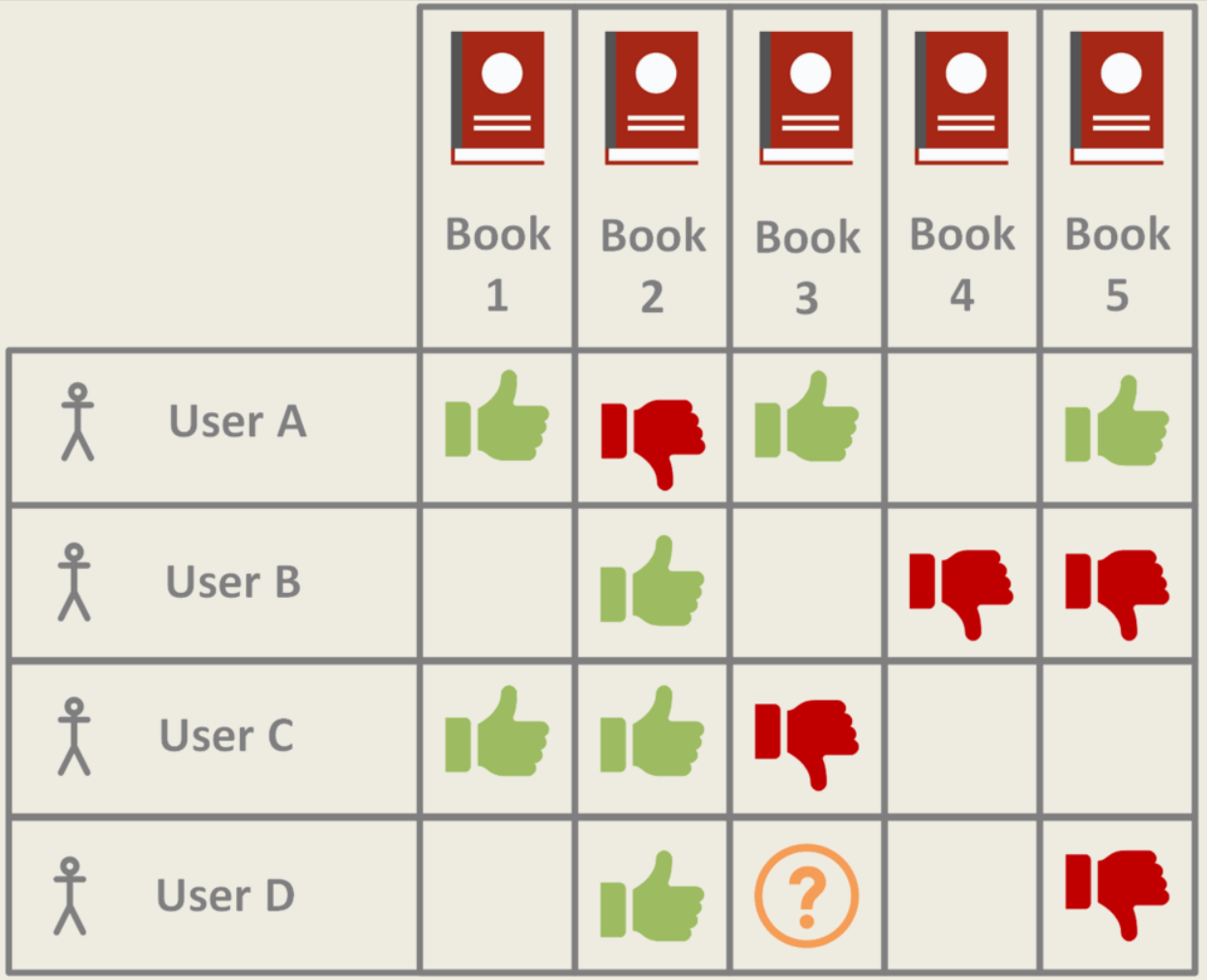
Ex. Benefits

Youtube	Amazon	Viblo
- **300 hours** of video are uploaded to YouTube every **minute**! - Almost **5 billion videos** are watched on Youtube every **single day.** - The total number of people who use YouTube **1,300,000,000 users**	- Sells more than **119,928,851 products** - In the U.S. alone, Amazon has over **150 million** monthly unique **visitors.**	- Over **26,000 verified** users - Over **21,000 published** posts - **1.5M Pageviews** per month

Good Practices



Data – Matrix



Data – Feature Vector

	A	B	C	D	E	F	item's feature vectors
Mưa nửa đêm	5	5	0	0	1	?	$\mathbf{x}_1 = [0.99, 0.02]$
Cỏ úa	5	?	?	0	?	?	$\mathbf{x}_2 = [0.91, 0.11]$
Vùng lá me bay	?	4	1	?	?	1	$\mathbf{x}_3 = [0.95, 0.05]$
Con cò bé bé	1	1	4	4	4	?	$\mathbf{x}_4 = [0.01, 0.99]$
Em yêu trường em	1	0	5	?	?	?	$\mathbf{x}_5 = [0.03, 0.98]$
User's models	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	$\leftarrow$ **need to optimize**

Exercise

- https://www.geeksforgeeks.org/recommendation-system-in-python/

- Data
  - userId movieId rating
  - movieId title genres