



## RECOMMENDATION ENGINES: THE PROBLEM WE WANT TO SOLVE

- Netflix wants its users to enjoy movies ... but the TV screen can only display a small number of movies ...
- How can Netflix ensure that users enjoy the movie they watch
- If you leave it for the users to pick ...  
either he will have to scroll and search a lot (poor experience)  
or she might quickly choose a bad movie (poor experience).
- Netflix wants to optimize user experience by predicting movies users will like ... and recommending them to users.



## WHAT COMPANIES CARE ABOUT THE PROBLEM ?

- Netflix (movies)
- Amazon (shopping)
- Search engines (ranking news items)
- Spotify (recommending music)
- Google news (customizing news recommendations)
- Yelp (recommending restaurants and services)
- Goodreads (book recommendations)
- ... many many more

Airbnb

Instagram

Amazon

Netflix

Ebay

Spotify

Facebook

Twitter

Google

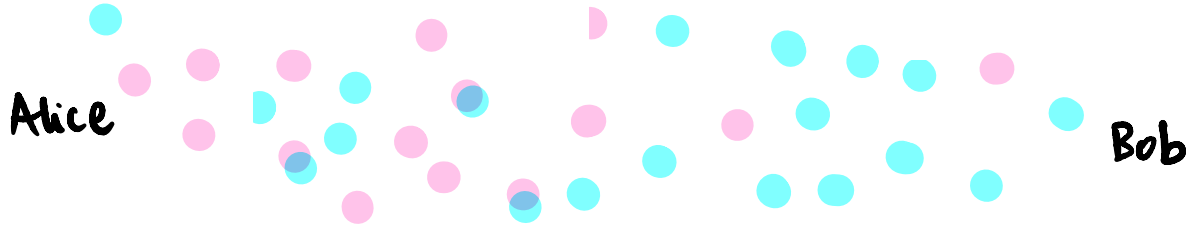
Uber

ISN'T THIS PROBLEM EASY TO SOLVE? HOW WOULD YOU DO IT ?

Propose an idea

## SOME HURDLES IN DESIGNING RECOMMENDATION ENGINES

- Say Alice watched W, X, Y ... Bob watched X, Y, Z ... and now Steve is a new user who has watched X and Y ...
- What would you recommend to Steve?
- Would you take some average of W and Z? What does that mean?



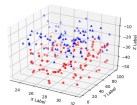
- If Steve watched Terminator, Matrix, and Borne Identity ... are you only going to recommend action movies?
- Are you sure Steve may not like comedy? Or Sci-Fi?

- When you are starting out as a company, you don't have much user data ... what do you do?

- How do you know your recommendation worked well or not?

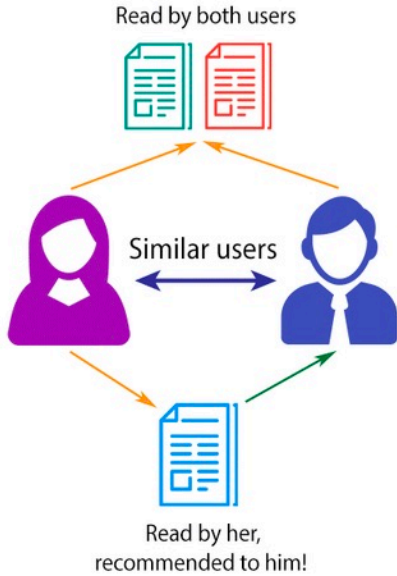
## Quick Foundations : Vectors, Vector spaces

↳ We want to represent movies as vectors or data points → "data science"

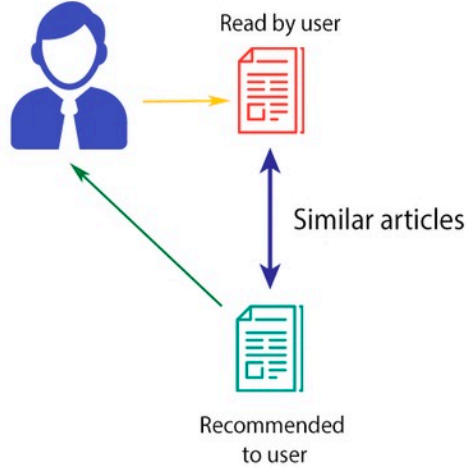


# 3 TYPES OF RECOMMENDATION ENGINE TECHNIQUES

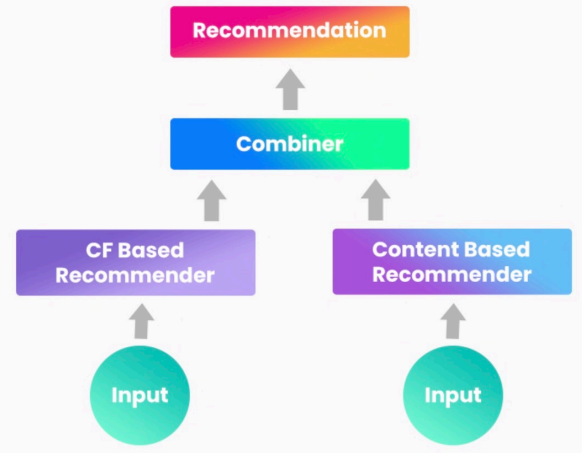
## COLLABORATIVE FILTERING



## CONTENT-BASED FILTERING



## HYBRID







## (2) COLLABORATIVE FILTERING

- Design  $M$  representative users — called EIGENUSERS
- Express any new user as weighted combination of eigenusers.
- Derive the recommendation from these weights.

movie

Terminator  
Matrix  
Titanic  
Conjuring

user

	1	2	3	...	$n$	
Alice	1	4/5	5/5	3/5	...	1/5
Bob	2	2/5	4/5	1/5		5/5
John	3					
⋮						
⋮				...		
⋮						
Susan	$m$					

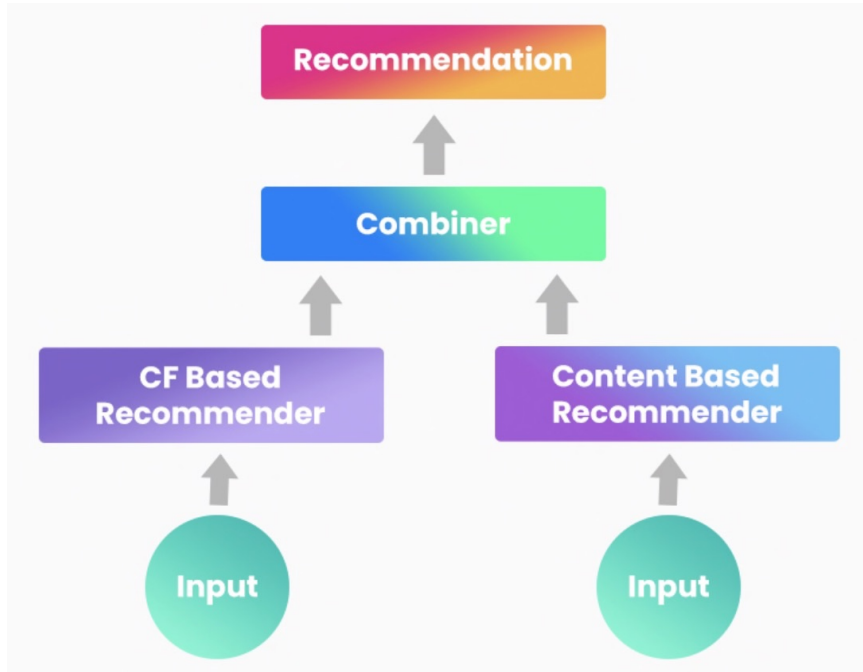
rating of movie  $j$  by user  $i$

$$m > n$$

Steve

4/5	5/5	3/5	...	X
-----	-----	-----	-----	---

### (3) HYBRID TECHNIQUES



## SOCIAL IMPLICATIONS (PRIVACY, BIAS, FAIRNESS ...)

Companies need data for content-based or collaborative filtering. Where are they getting the data?

- Cookies in your browser
- Your visited websites
- Your shopping patterns
- Your search queries in the Internet

This data is feeding recommendation engines ... but also leaking a lot of information about you to the Internet.

What if tomorrow, a Government says ... you have been eating junk food, so we are revoking your medical insurance

Companies using data for shortlisting candidates for a job ...

- Suppose the intelligent algorithm uses data from the past candidates who were, or were not, recruited.
- Trains the eigenusers from this data

What's the problem?

What kind of other biases can you think of ... when data is used to create the "representative" samples ... the EIGENITEMS ? Are there other biases or fairness