

The Problem we want to Solve ...

Netflix wants its users to enjoy new movies/series ...

but the TV screen can only display a small number of tiles ...

How can Netflix ensure that users enjoy the movie they watch?

If you leave it for the users to pick ...

- either they will have to scroll and search a lot (poor experience 🙄)
- or they might quickly choose a bad movie (poor experience 🙄)



Optimize User Experience by Predicting Movies

Netflix wants to optimize user experience by

- **predicting movies users will like** and
- **recommending** them to users.

People don't know what they want
until you show it to them.

-Steve Jobs

Companies that Care about this Problem

Airbnb

Instagram

Others:

•

Amazon

Netflix

Ebay

Spotify

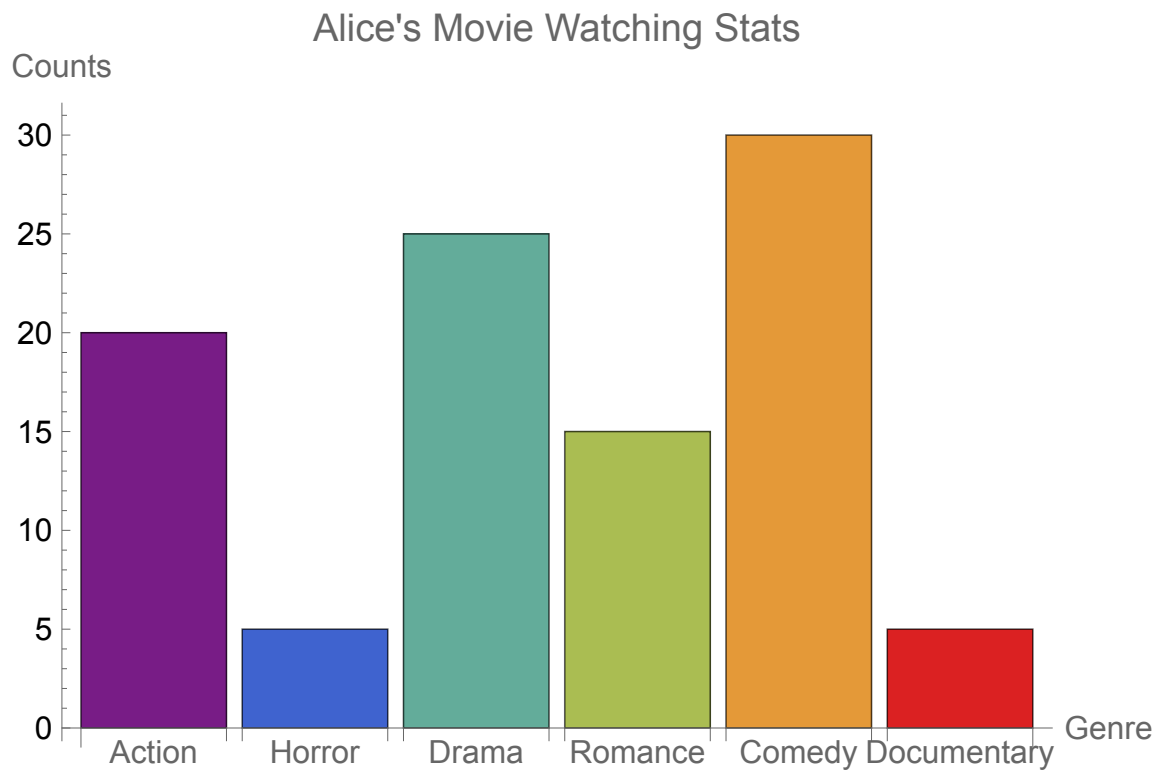
Facebook

 TikTok

Google



Example: What movie should Netflix recommend for Alice?



Ideas?

-

Some Hurdles in Designing Recommendation Engines

When trying to recommend movies:

- Say Alice watched **W**, **X**, **Y** and Bob watched **X**, **Y**, **Z**
- 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?

Some Hurdles in Designing Recommendation Engines

When trying to recommend movies:

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

Some Hurdles in Designing Recommendation Engines

When trying to recommend movies:

- 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 Foundation: Vector Spaces

n -dimensional Euclidean space,
where every element is represented by
a list of n real numbers

*Check Wolfram Notebook

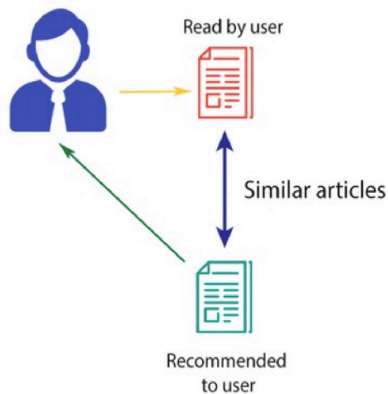
Quick Foundation: Vector Spaces

Check Wolfram Notebook

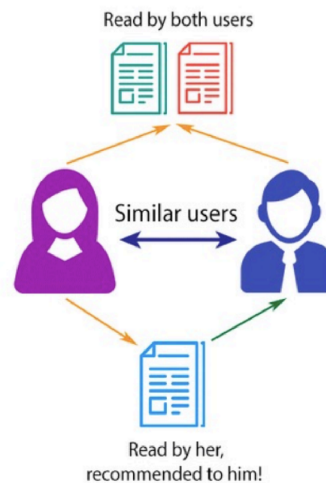
Three Main Types of Recommendation Engine Techniques

- Content based filtering
- Collaborative filtering
- Hybrid techniques

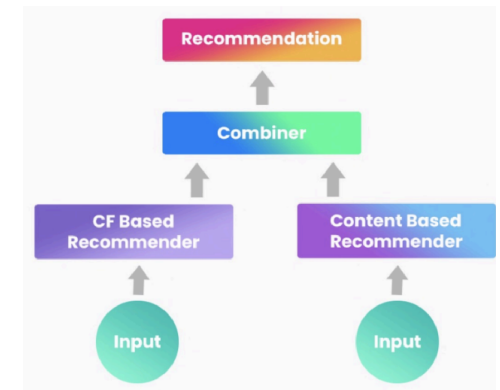
CONTENT-BASED FILTERING



COLLABORATIVE FILTERING



Hybrid
HYBRID



Content Based Filtering

Movies in Feature Space:

- Convert all movies into a point in a "feature space"
- Mark Alice's already-watched movies in that same "feature space"
- Find movies in the "neighborhood" of Alice's already-watched movies.

Collaborative Filtering

People in Feature Space

- Convert each person into a point in a "**feature space**"
- Find other people in the "**neighborhood**" of Alice
- Recommend movies they have watched

People in terms of Eigen Users

- Design M representative users—called **Eigen Users**
- Express any new user as **weighted combination** of eigen users
- Derive the recommendation from these weights

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!!

Social Implications (Privacy, Bias, Fairness ...)

- Companies can use data for shortlisting candidates for jobs ...
- Suppose an intelligent algorithm
 - uses data from the past candidates
 - who were or were not recruited
 - creates the Eigen Users from this data
 - selects candidates based on match with Eigen users
- What could be the problem with this approach? What kind of biases do you think may appear ...
 - When data is used to create the "representative" samples... the “Eigen Users”?
 - Are there other biases or issues with fairness?

Terminology You Should Know from These Slides

- N-dimensional vector space
- 1-dimensional data
- 2-dimensional data
- 3-dimensional data
- N-dimensional data
- Feature space
- Content based filtering
- Collaborative filtering
- Hybrid techniques
- Eigen users

Concepts You Should Know from These Slides

- problem solved by recommendation engines
- representing an entity in vector space
- representing and visualizing data samples in 1, 2, 3 ...
n-dimensional feature space
- nearness or distance of one sample from another
- nearness of content in content based filtering
- nearness of user in collaborative filtering
- imaginary “eigen users” as representative samples of
the user base