

Why am I seeing this ad?

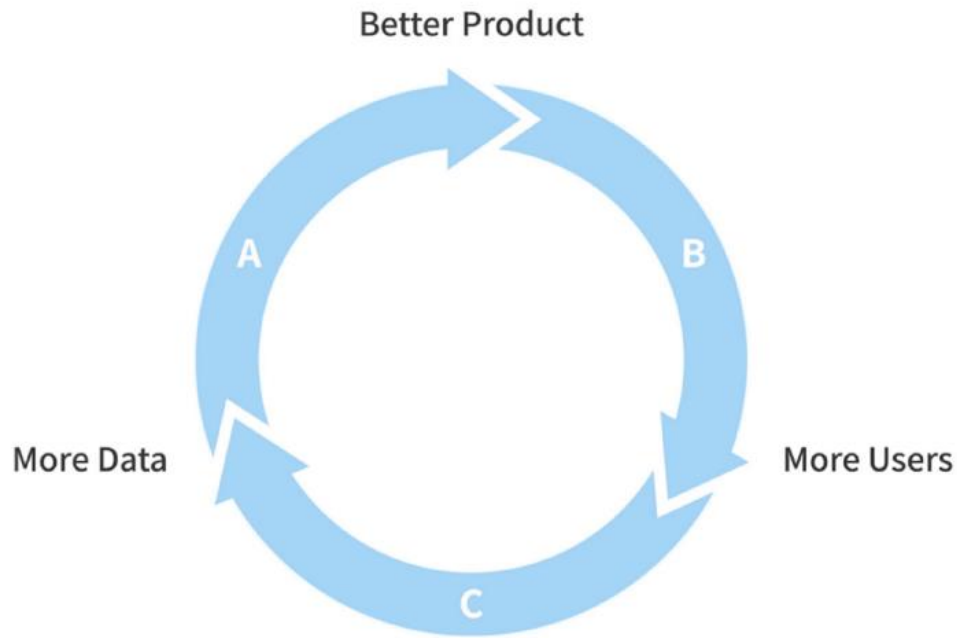
An overview of the most valuable AI Algorithm

Agenda

- Business Value
- Problem Definition
- Data
- Algorithms
- Demo

Business Value

- Recommenders are the single most important algorithm.
- The more they are used, the more value they generate.
- Recommenders are a renewable data resource that provides deep customer insights.



What problem does this solve?

- Songs on Spotify
- Movies on Netflix
- Videos on YouTube
- Related Posts on Twitter/Instagram/LI
- Similar dishes on Ubereats
- And yes, Ads on Facebook...



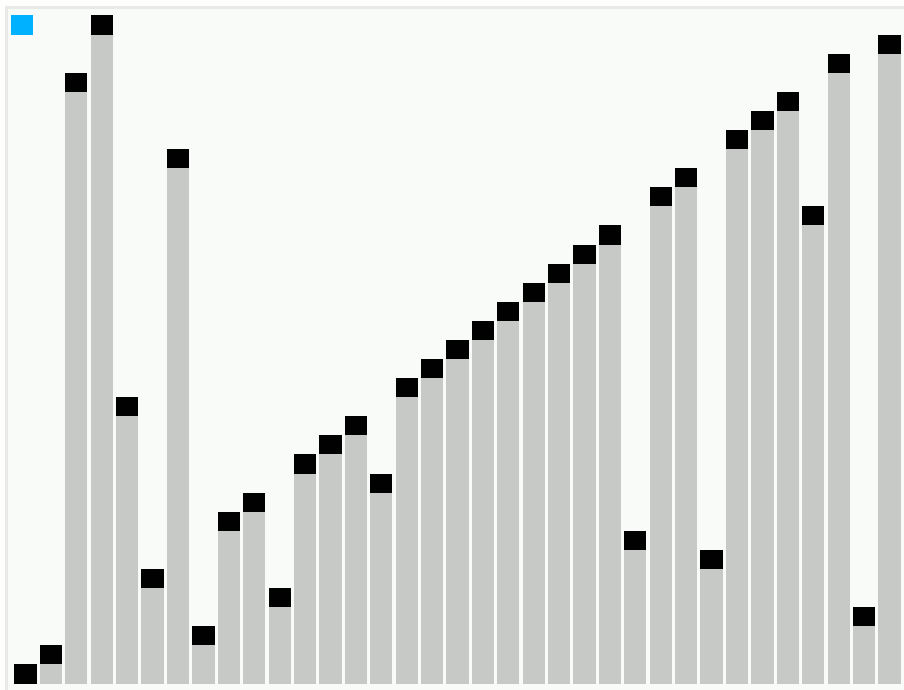
How do you get the data?



Explicit data: How you intentionally interacted with the site. (what you liked, purchased, etc.)

Implicit data: Data from the items and how you implicitly interacted with the site (what you clicked on, search logs, etc.)

Algorithms



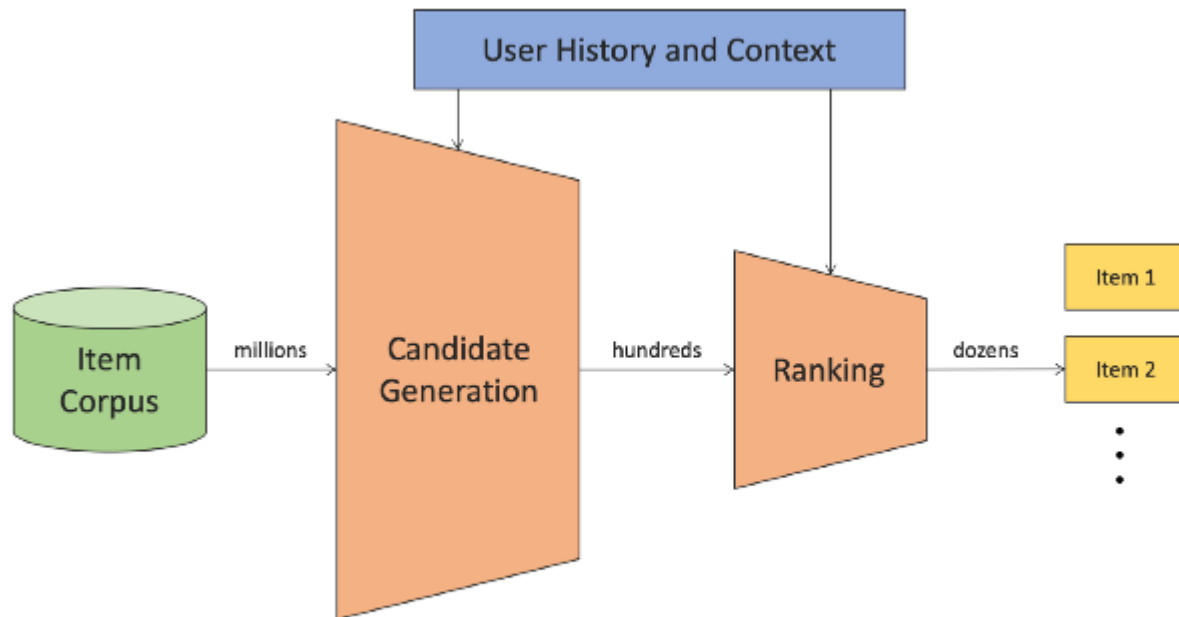
<https://www.cs.umd.edu/~samir/498/Amazon-Recommendations.pdf>

- Collaborative filtering
- Content-based filtering
- Social and demographic recommenders
- Contextual recommendation

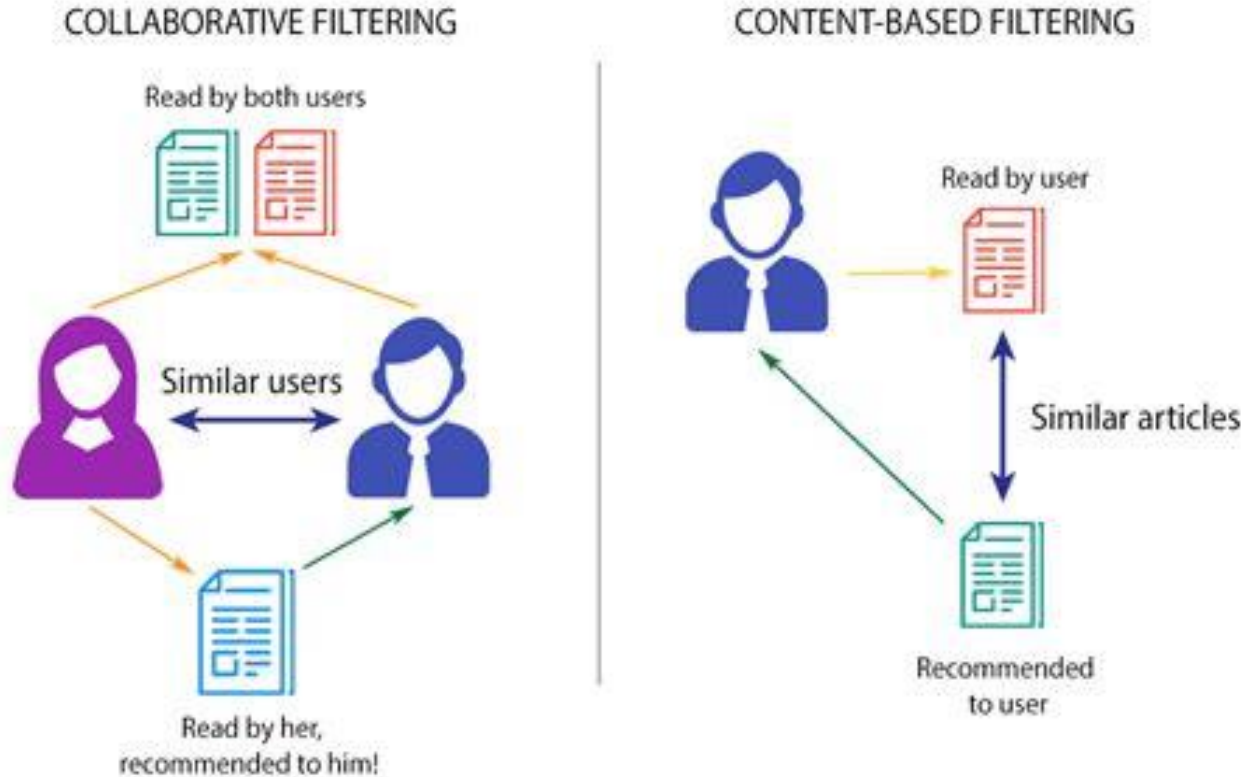
The Essence of Recommenders

All Recommenders have two things in common:

- They generate candidates
- They rank the candidates



Demo of a Content Based Recommender



The Data

<https://bit.ly/ai4nh>

or

<https://repl.it/@ImmanuelKant/ContentRecommenderExample>

TF-IDF Vectorizer

TF (term frequency) of a word is the number of times it appears in a document. When you know it, you're able to see if you're using a term too often or too infrequently.

IDF (inverse document frequency) of a word is the measure of how significant that term is in the whole corpus.


$$w_{x,y} = tf_{x,y} \times \log \left(\frac{N}{df_x} \right)$$

TF-IDF

Term x within document y

$tf_{x,y}$ = frequency of x in y

df_x = number of documents containing x

N = total number of documents

TF-IDF calculation

How to make a vector from sentences

Julie loves John more than Linda loves John

Jane loves John more than Julie loves John

John 2 2

Jane 0 1

Julie 1 1

Linda 1 0

likes 0 1

loves 2 1

more 1 1

than 1 1

the two vectors are,

Item 1: [2, 0, 1, 1, 0, 2, 1, 1]

Item 2: [2, 1, 1, 0, 1, 1, 1, 1]

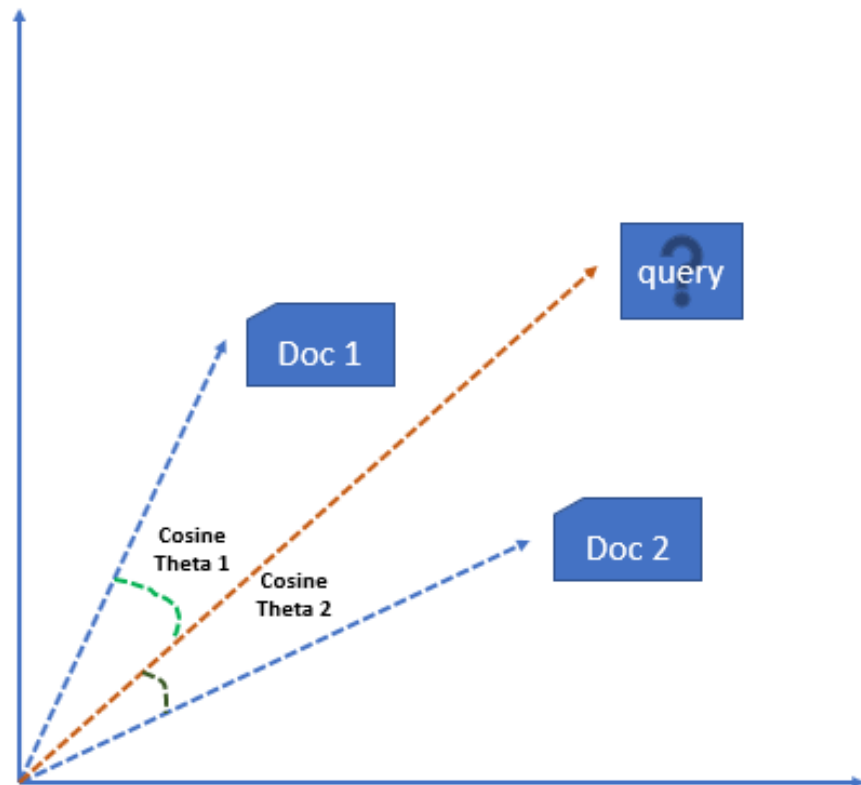
The cosine angle (the smaller the angle) between the two vectors' value is 0.822 which is nearest to 1.

(i.e.: the sentences are similar)

The big idea...

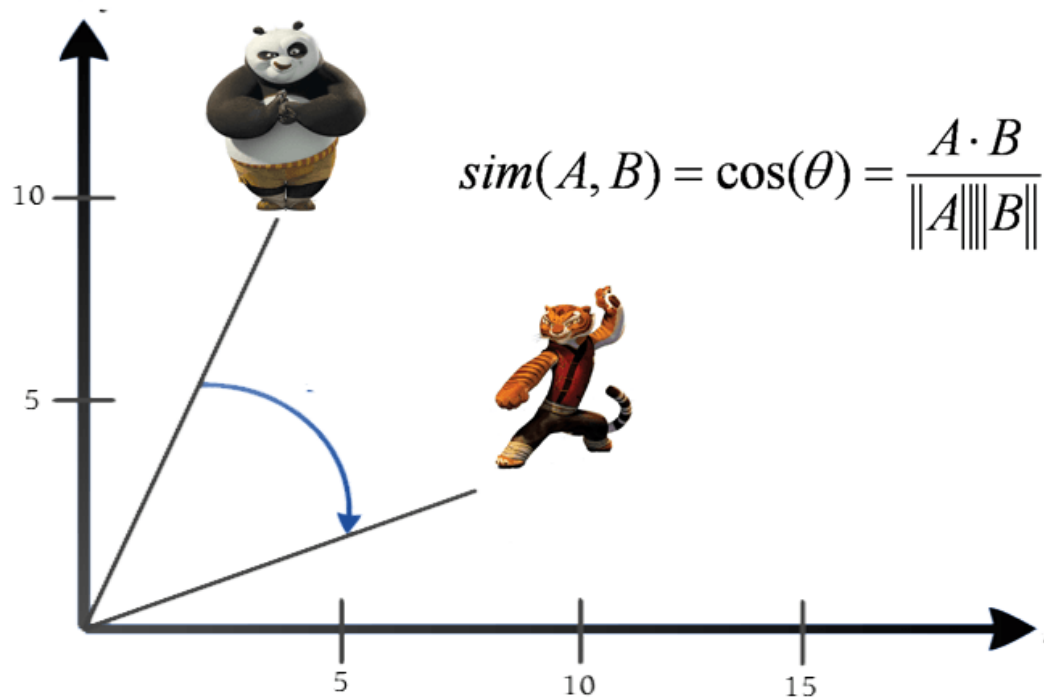
Every item in your set is a unique vector.

By calculating the cosine between the vectors, you can create a score that determines how “similar” the items are.



Cosine Similarity

Cosine Similarity



Recommendation



Summary and Question and Answer

