



Why Am I Seeing This Ad?

*Exploring the Secrets
Behind AI & Algorithms*

PRESENTER: Dr. ERNESTO LEE | Chief Data Scientist, Florida Center for Advanced Analytics and Data Science

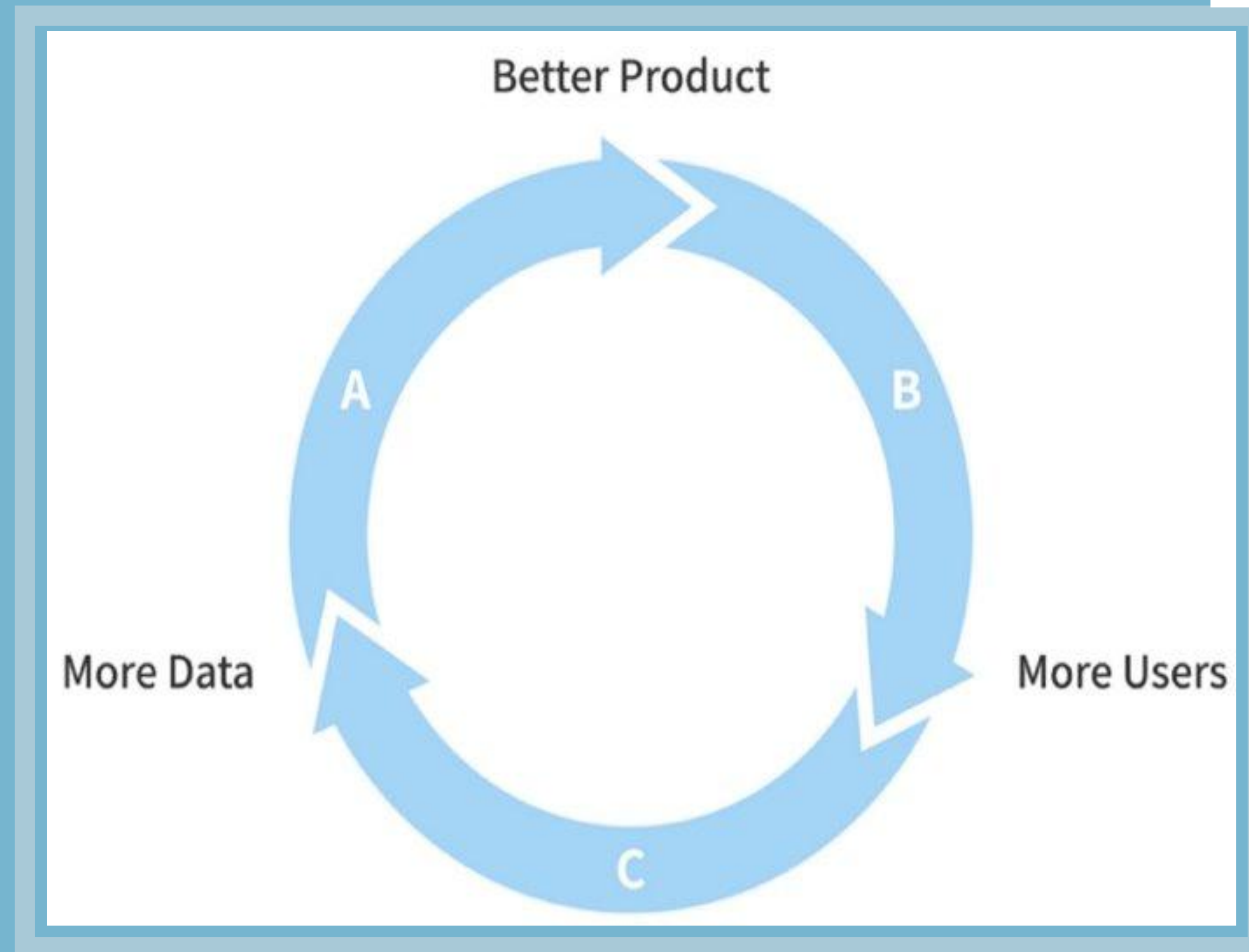
The Business Value

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 PROBLEMS DOES THIS SOLVE?

SONGS ON SPOTIFY

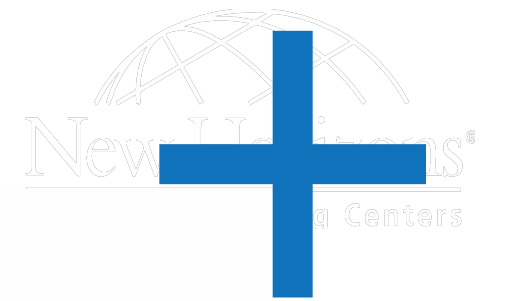
MOVIES ON NETFLIX

VIDEOS ON YOUTUBE

RELATED POSTS ON
TWITTER/INSTAGRAM/LI

SIMILAR DISHES ON UBER EATS

AND YES, ADS ON FACEBOOK...



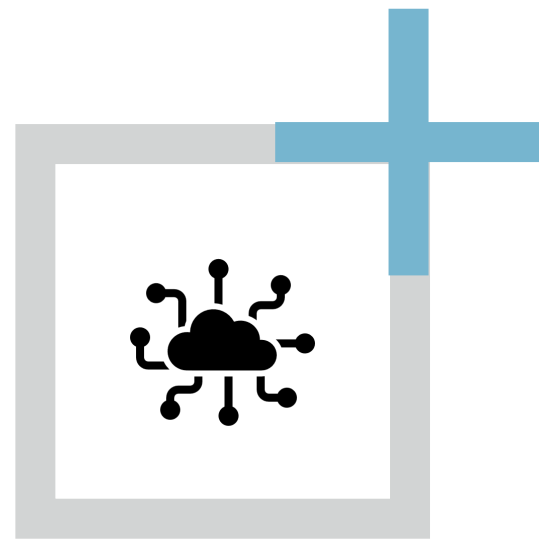
MakeAGIF.com



HOW DO YOU GET THE DATA?

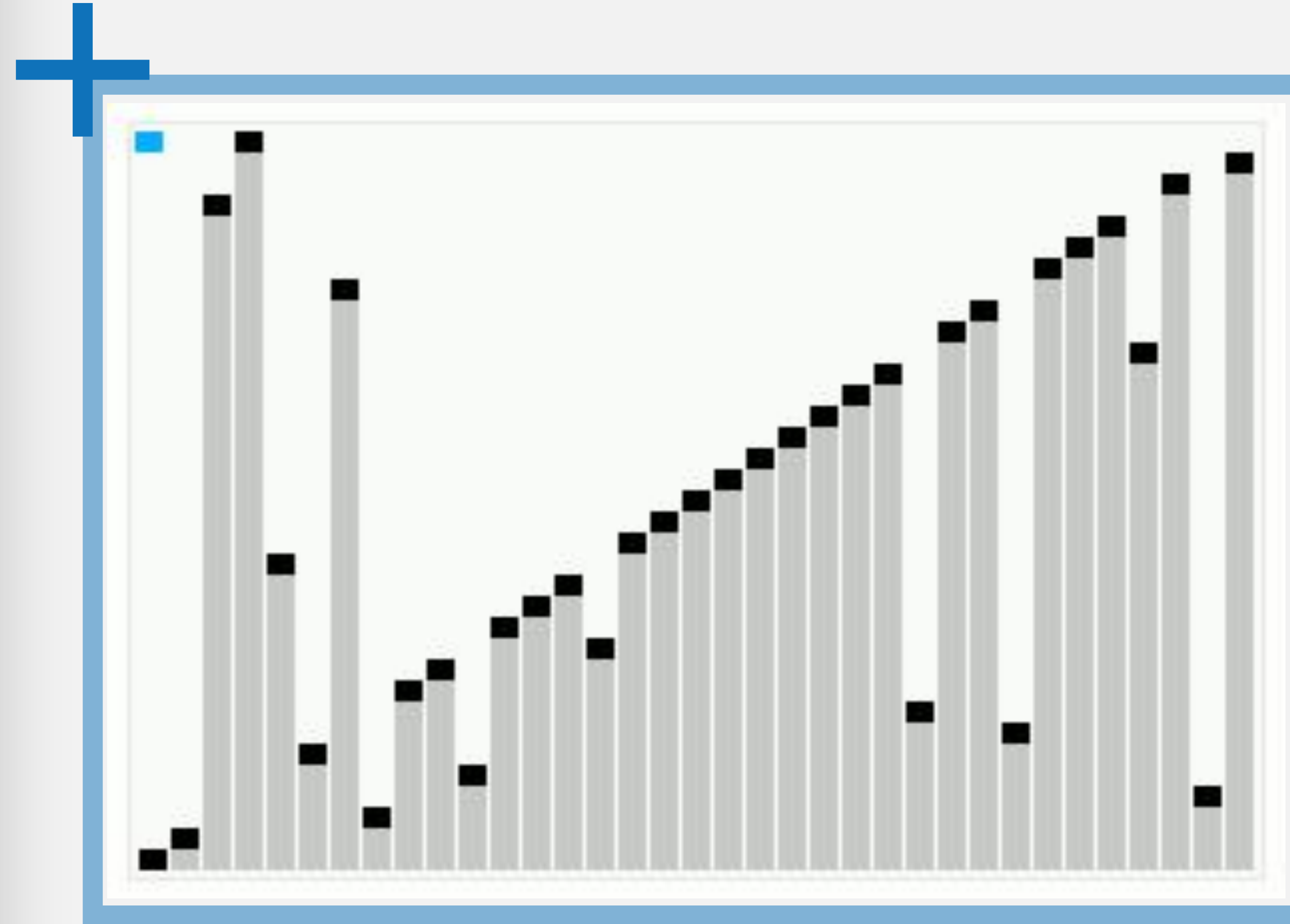
EXPLICIT DATA: HOW YOU INTENTIONALLY INTERACTED WITH THE SITE. (WHAT YOU LIKED, PURCHASED, ETC.)

IMPLICIT DATA: DATA FROM THE ITEMS AND HOW YOU IMPLICITLY INTERACTED WITH THE SITE (WHAT YOU CLICKED ON, SEARCH LOGS, ETC.)



ALGORITHMS

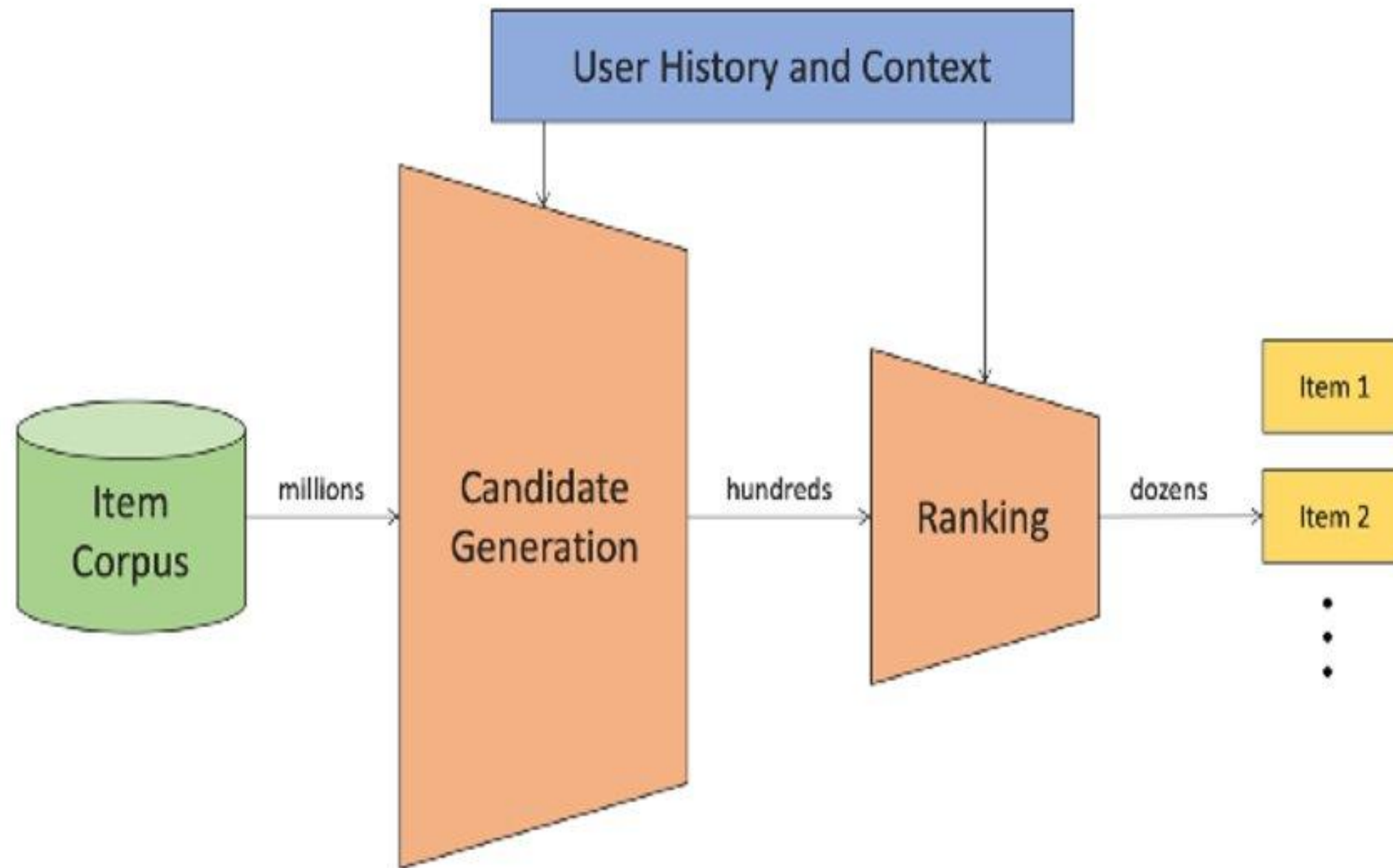
- 1 Collaborative Filtering
- 2 Content-based Filtering
- 3 Social and Demographic recommenders
- 4 Contextual recommendation



THE ESSENCE OF RECOMMENDERS

ALL RECOMMENDERS HAVE TWO THINGS IN COMMON:

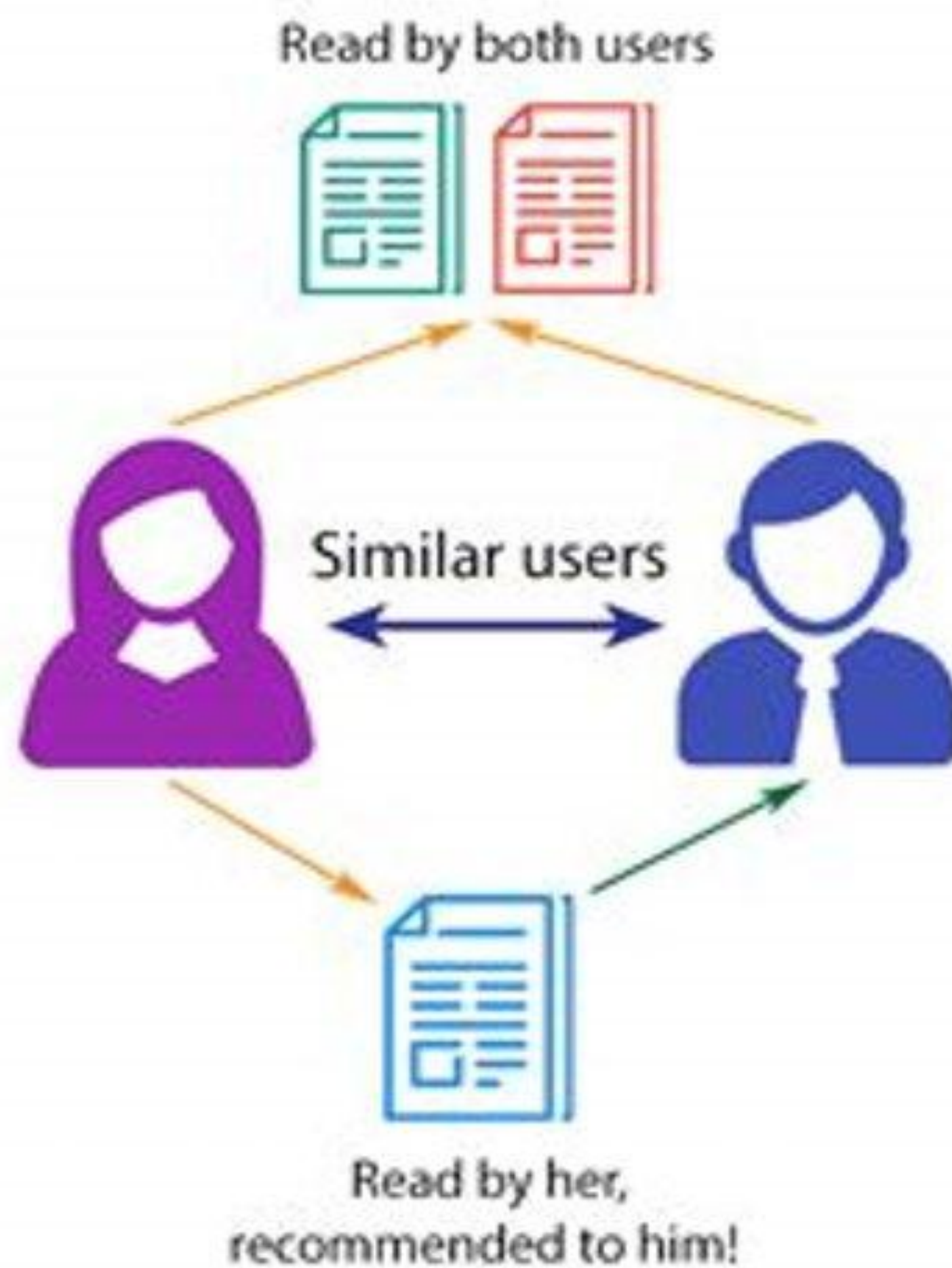
- THEY GENERATE CANDIDATES
- THEY RANK THE CANDIDATES



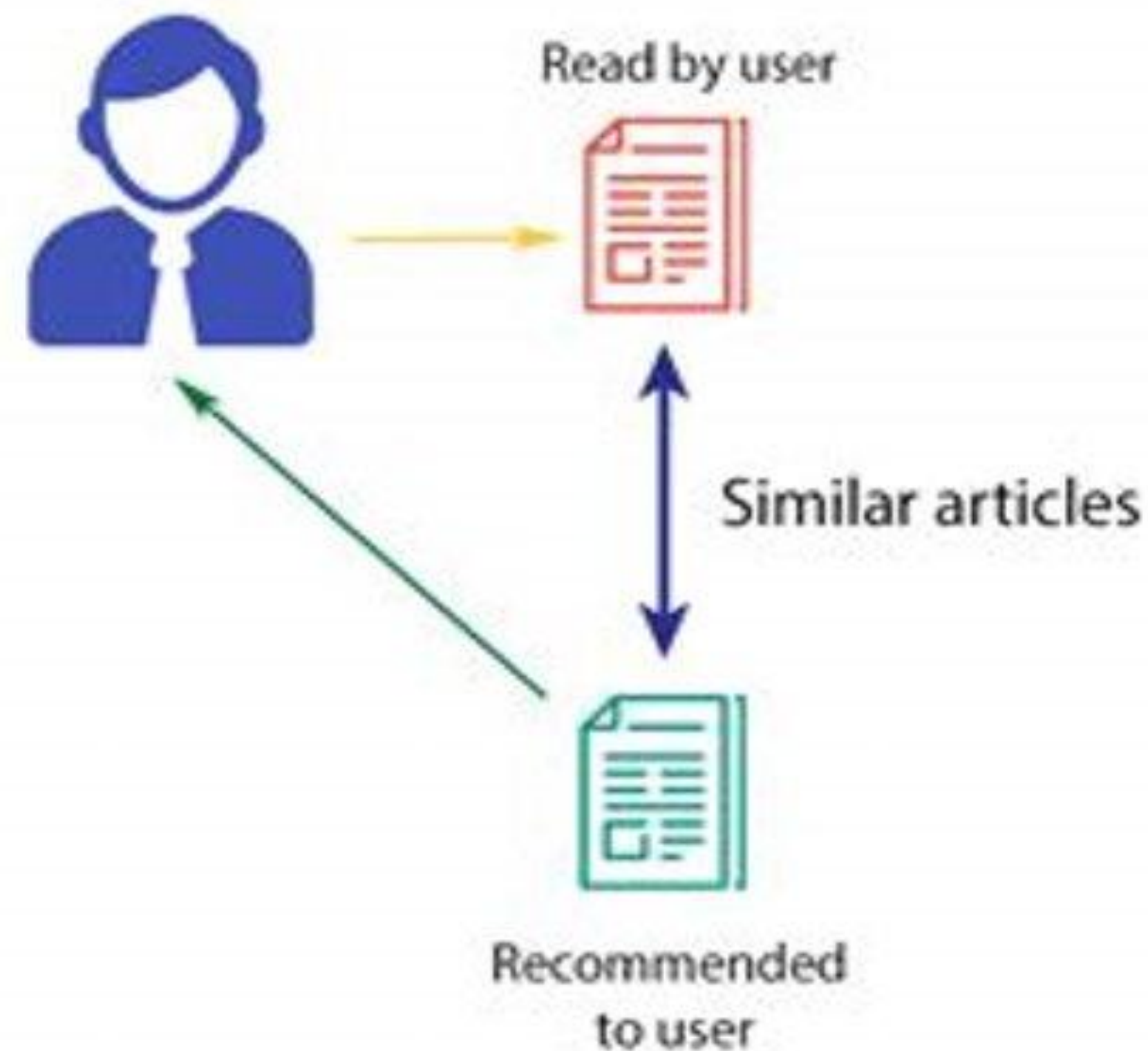
DEMO | Content Based Recommender

[HTTPS://REPL.IT/@IMMANUELKANT/CONTENTRECOMMENDEREXAMPLE](https://repl.it/@ImmanuelKant/contentrecommenderexample)

COLLABORATIVE FILTERING



CONTENT-BASED FILTERING



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

Julie loves John more than
Linda loves John

Jane loves John more than
Julie loves John

John 2 2

Jane 0 1

Julia 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)

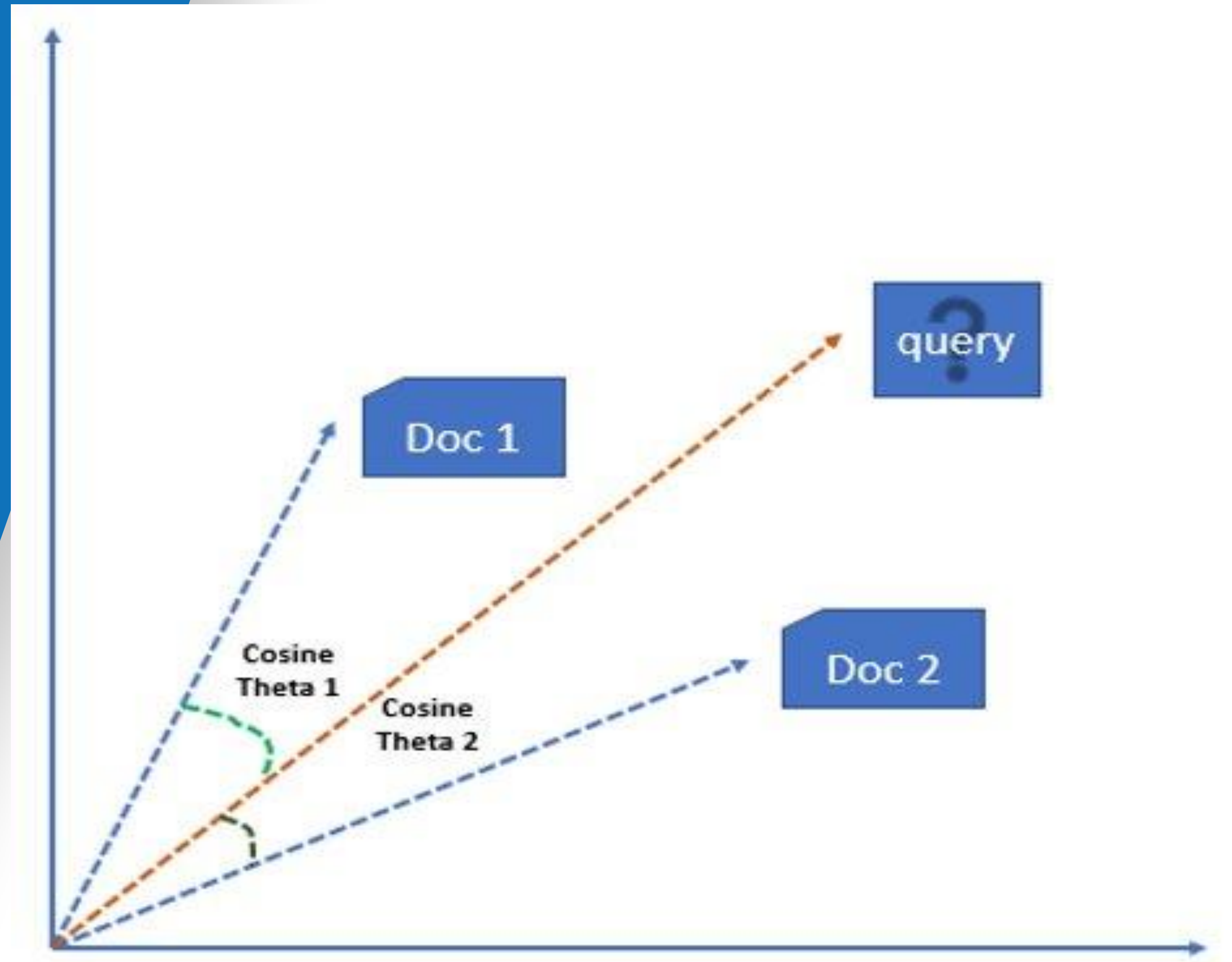
HOW TO MAKE A VECTOR FROM SENTENCES

THE BIG IDEA

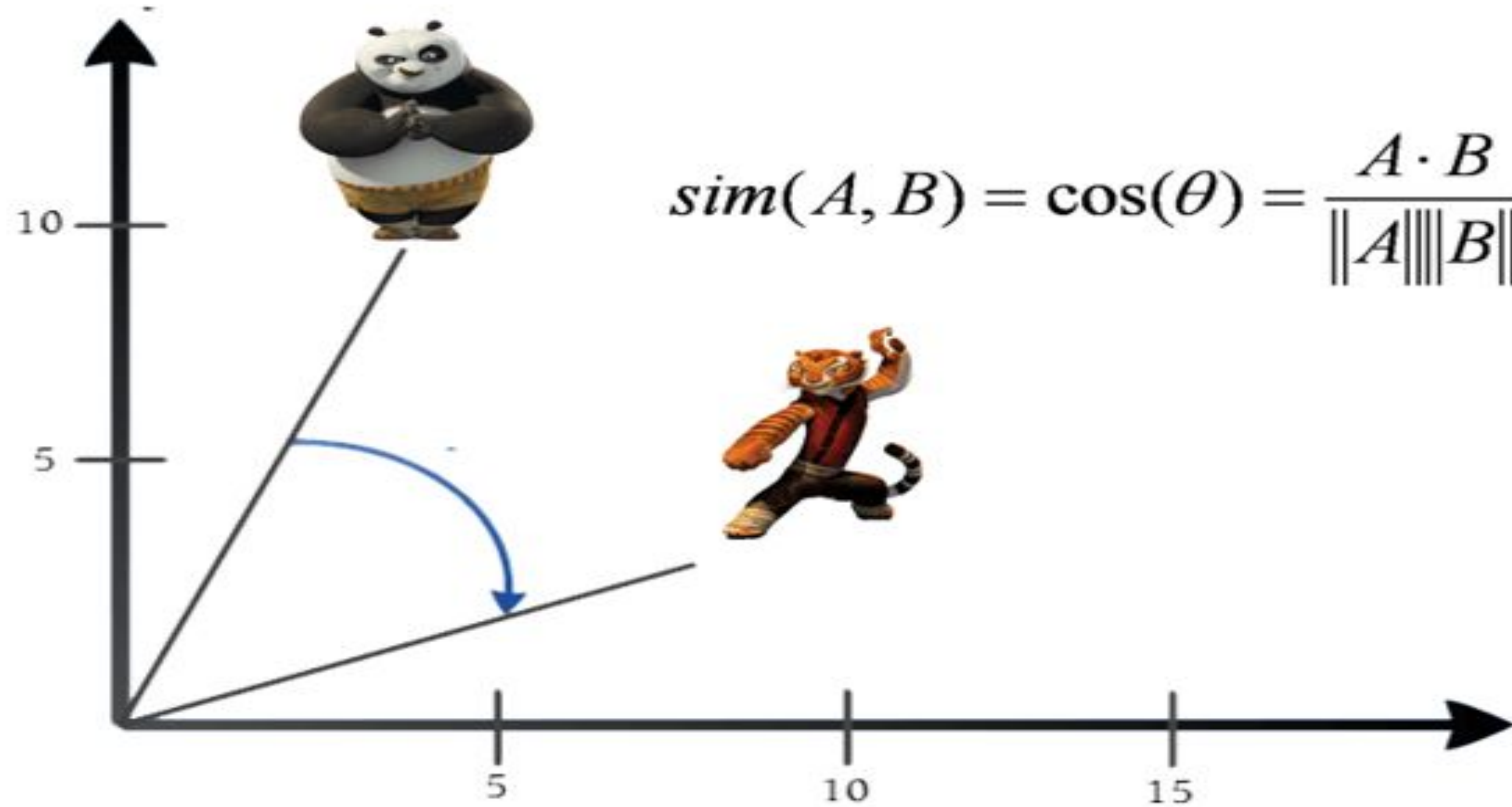


Every item in your set is a unique vector.

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



Cosine Similarity



+ COSINE SIMILARITY



RECOMMENDATION



THANK YOU



