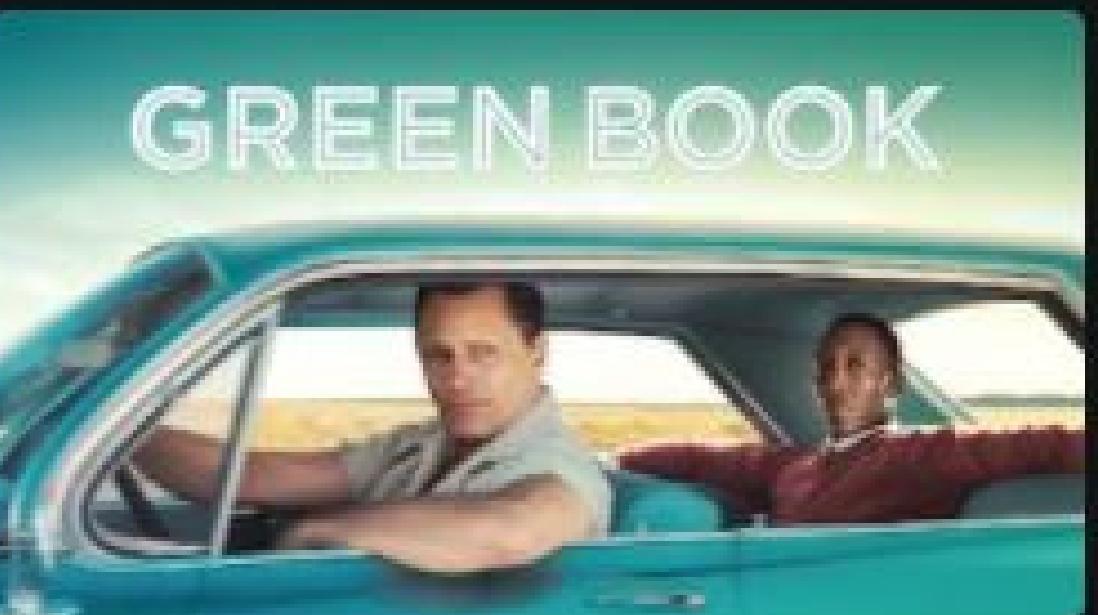


POL SERRAMALERA GUERIN

DECEMBER, 2020

MOVIE RECOMMENDER SYSTEM

Recomendaciones para Pol



Las 10 más populares en España hoy



THE PURPOSE OF A RECOMMENDATION
SYSTEM IS TO SEARCH FOR CONTENT THAT
WOULD BE INTERESTING TO AN INDIVIDUAL



KEY APPLICATIONS



Movie
platforms



E-commerce



Social media
users



Music
platforms



Courses and
articles

WORKFLOW

DATASET
PREPROCESSING

EVALUATING
MODELS

BUILDING MODELS

FUTURE
IMPROVEMENTS



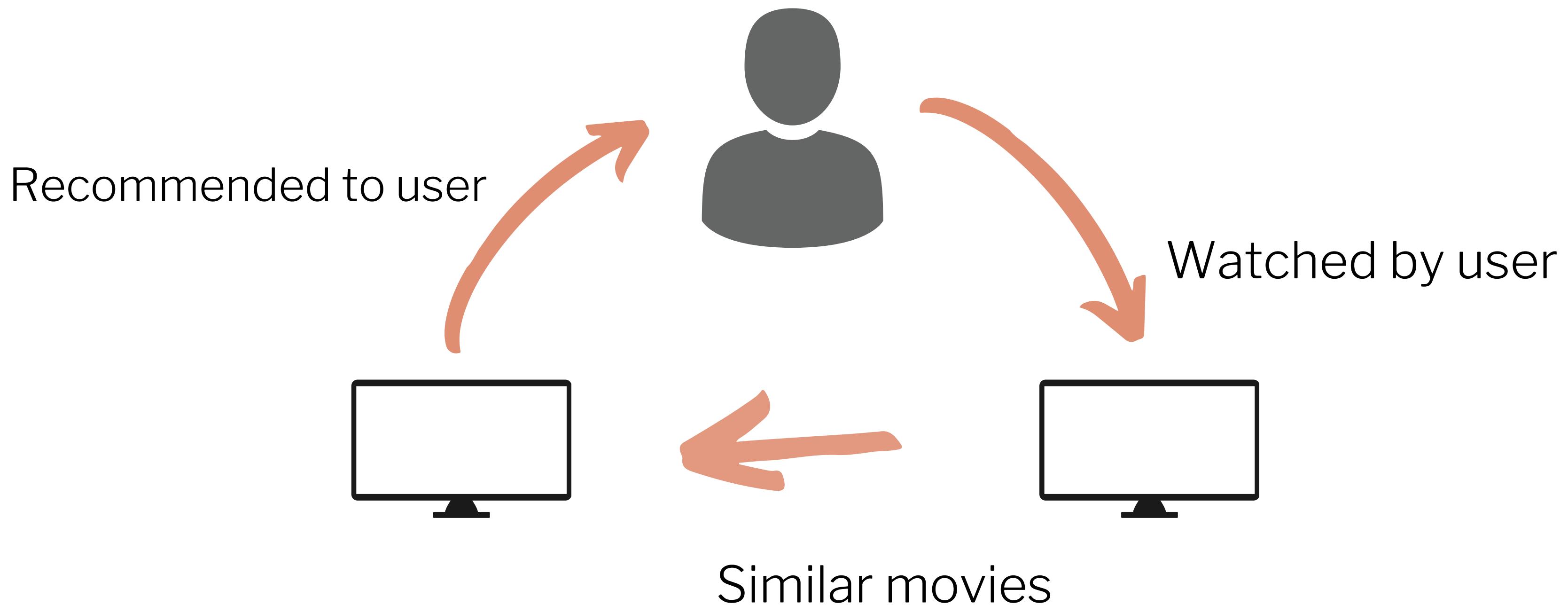
+100k movies
and thousands
of users

IMDB movie database
with thousands of
movies and several
features such as
genre, actors,
description, etc

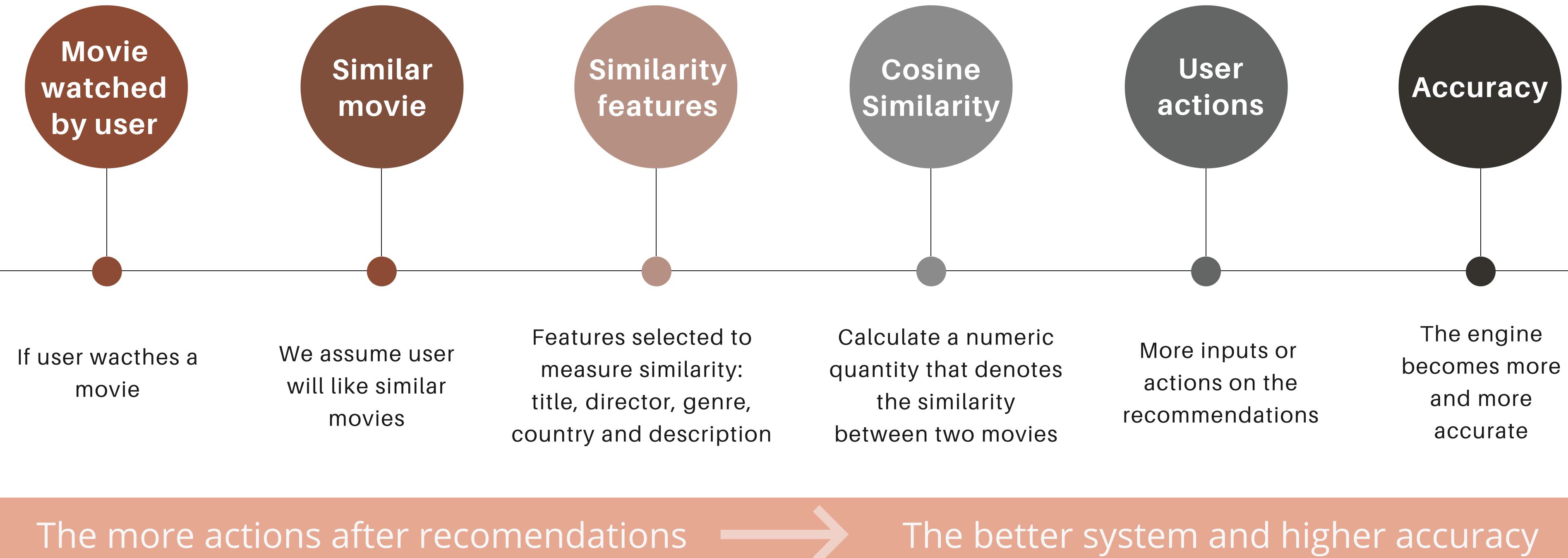
Movielens dataset with
+1000 users, movies
and ratings

BUILDING THE MODELS

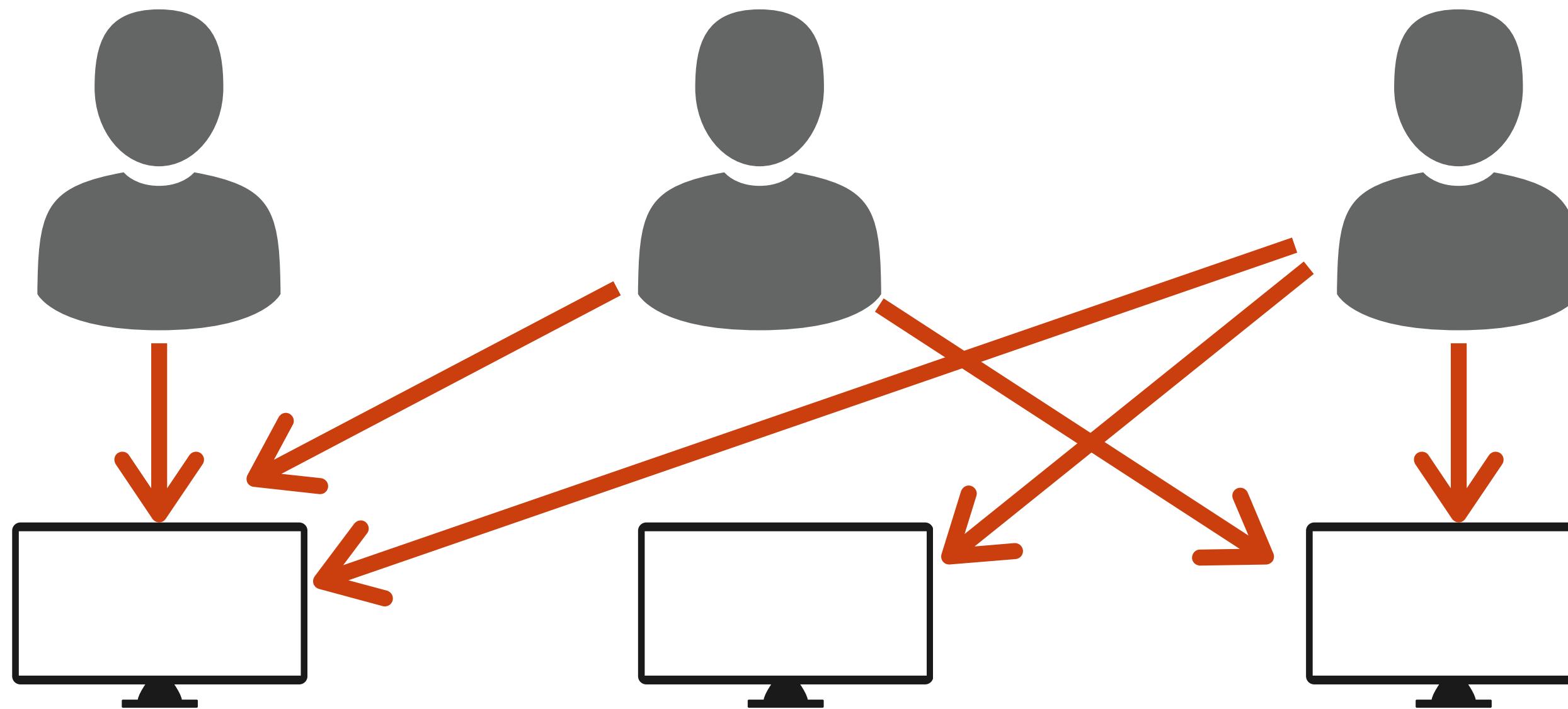
1. CONTENT-BASED FILTERING



CONTENT-BASED FILTERING

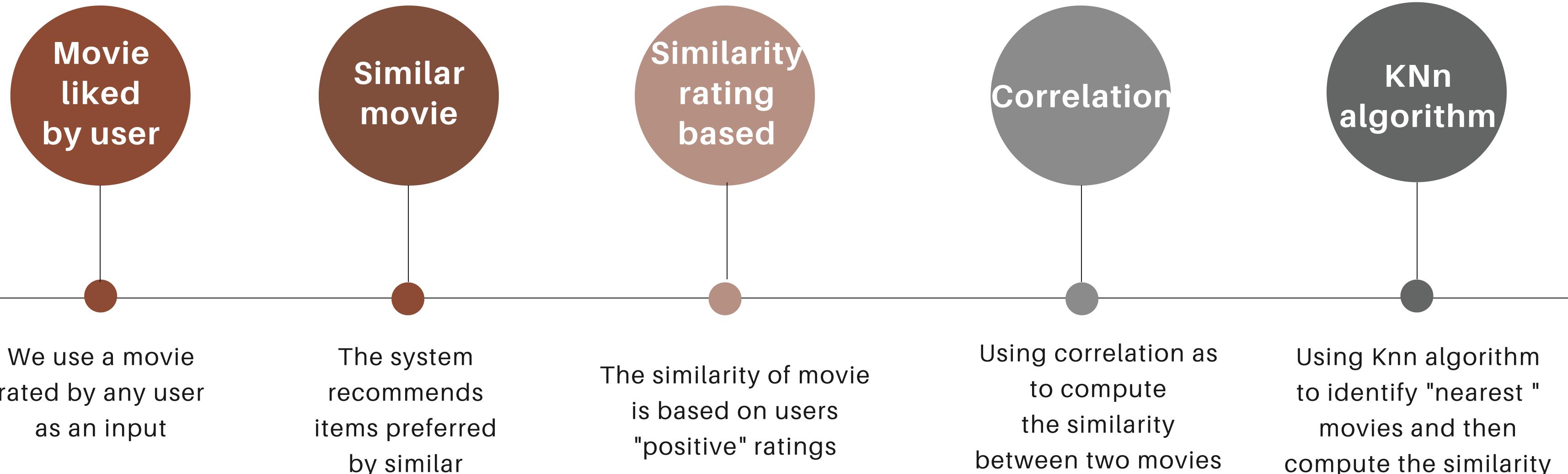


2. COLLABORATIVE FILTERING



Similar movies based on users ratings

COLLABORATIVE FILTERING



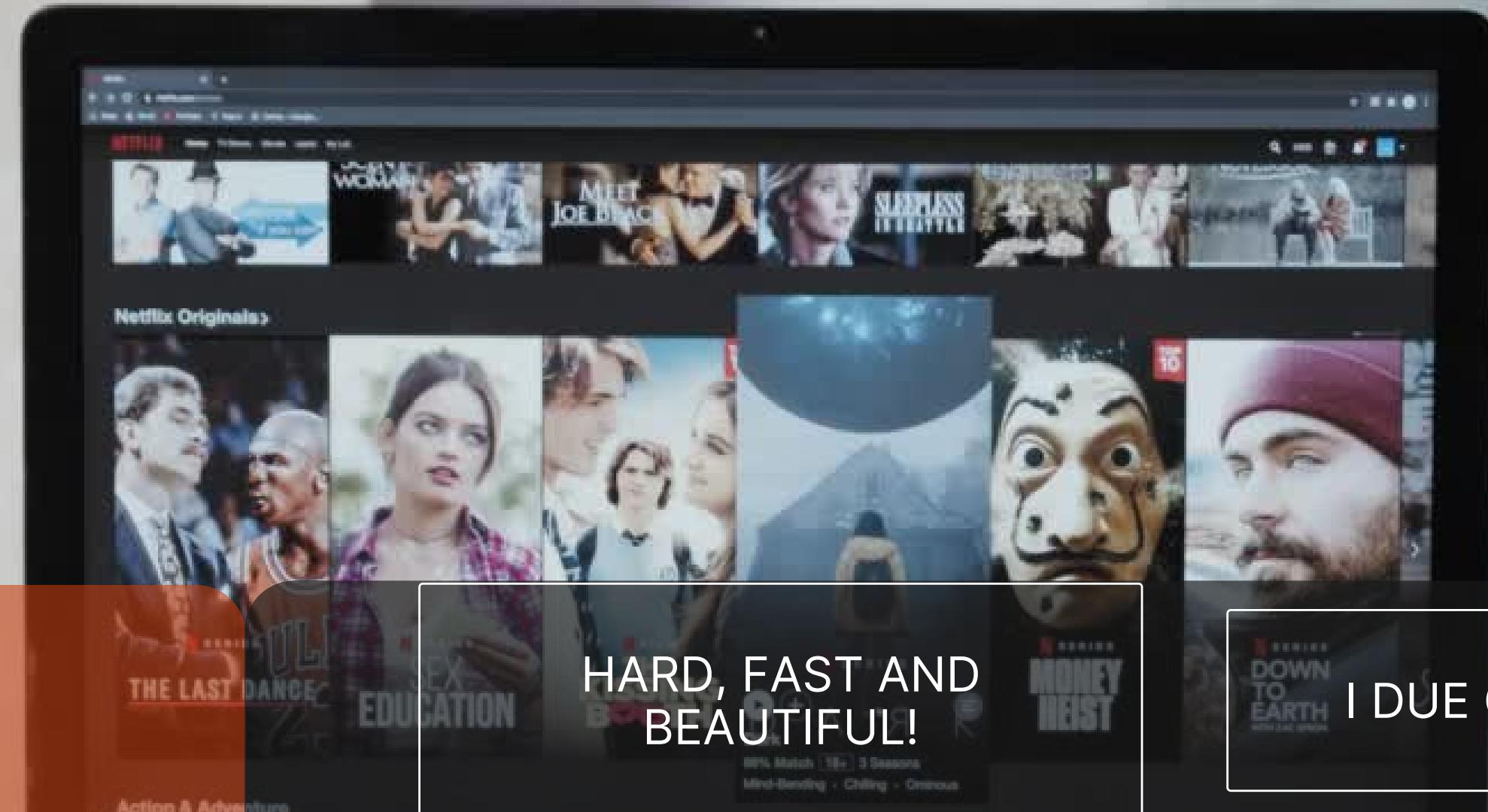
The more user interact with movies



The better system and higher accuracy

Content-based filtering

Movie inputed: *cinderella*



LA DEA INGINOCCHIATA

YOUNG MAN OF
MANHATTAN

I DUE COMBATTENTI

RAFFLES

SUCCESS AT ANY PRICE

Collaborative filtering (Corr)

Movie input: *Cinderella*



SNOW WHITE AND THE
SEVEN DWARFS

DUMBO

PINOCCHIO

BEAUTY AND THE BEAST

LION KING

FANTASIA

Collaborative filtering (KNn)

Movie input: *Cinderella*



SNOW WHITE AND THE
SEVEN DWARFS

PINOCCHIO

DUMBO

BEAUTY AND THE BEAST

FANTASIA

E.T. THE EXTRA-
TERRESTRIAL

MODEL EVALUATION

Select the first rated movie of each user from the test set and pass it to the model as an input



Check if users have watched them in the future



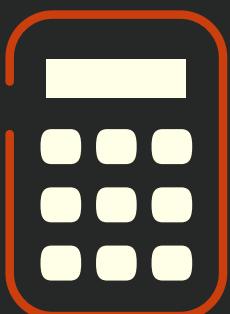
Split the dataset used in training and test set (80-20)



Get TOP 10 recommendations back



Calculate the total number of users that watched at least one of the movies recommended and the total movies recommended and watched afterwards



CF USING CORRELATION

43

users with at least 1 movie match

243

total matched movies

5%

of tested users have a match

CF USING NEAREST NEIGHBOURS

59

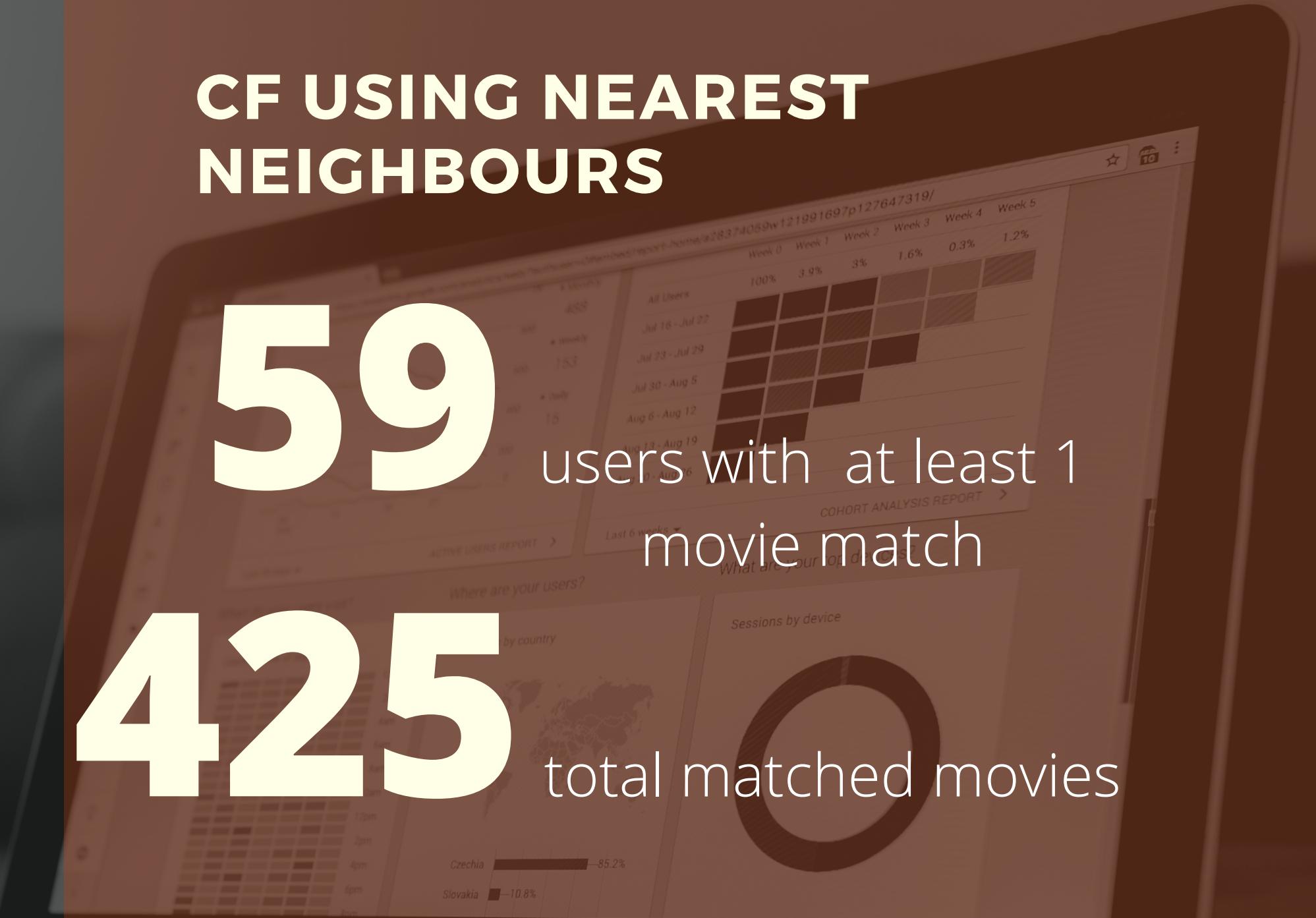
users with at least 1 movie match

425

total matched movies

7%

of tested users have a match



Conclusions and improvements

ACCURACY IMPROVES ONCE THE SYSTEM IS RUNNING



The more interactions the better system accuracy as the systems "memorizes" more data



RATINGS SCALABILITY FOR BETTER ACCURACY

Normalizing the ratings of each user



COLD START PROBLEM

The introduction of new users or new movies there means insufficient data on these new entries for the CF to work accurately



CLOUD COMPUTING FOR SYSTEM IMPROVEMENT

This recommender systems take a lot of computational power



Thank you