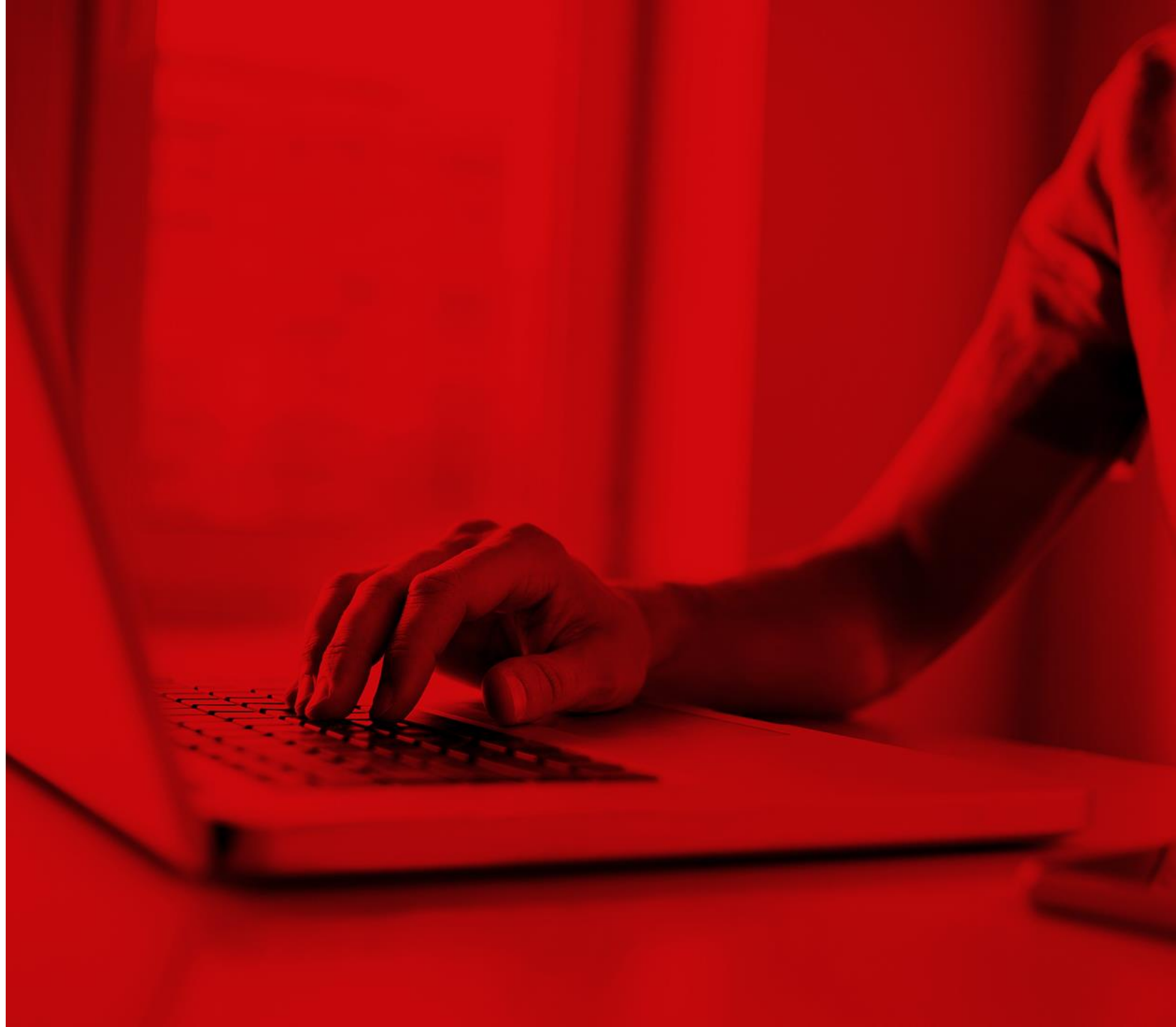


**EN
INTRODUKSJON
TIL
ANBEFALINGER**





Håkon Dissen

Konsulent hos Itera

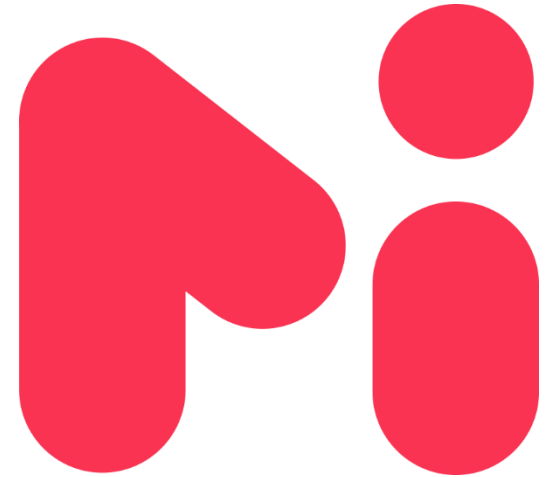
Master Datateknikk NTNU
Intelligente Systemer

hakon.dissen@itera.no

DEN NYE MÅTEN Å SE TV

HVA ER DET?

- En intern start-up
- Jobber etter lean startup prinsipper
 - Build: Bygge MVP (eksperiment)
 - Measure: Teste og validere hypoteser
 - Learn: Pivotere
- Skal utvikle et nytt produkt for å dekke markedetsbehov i TV-bransjen
 - Samle innhold og tjenester på ett sted
 - Umiddelbar avspilling av relevant, personlig tilpasset innhold
 - Tilby innhold til Cord Cutters



PRODUKTET, FOR EN STUND SIDEN





AGENDA

Hva er en anbefaler?

1. Hvorfor
2. Hvordan?

Content based filtering

1. Vektorprodukt
2. Fordeler og ulemper

Collaborative filtering

1. I sin simpleste form
2. Nabolagsbasert
3. Modellbasert
4. Fordeler og ulemper

Oppsummering(?)

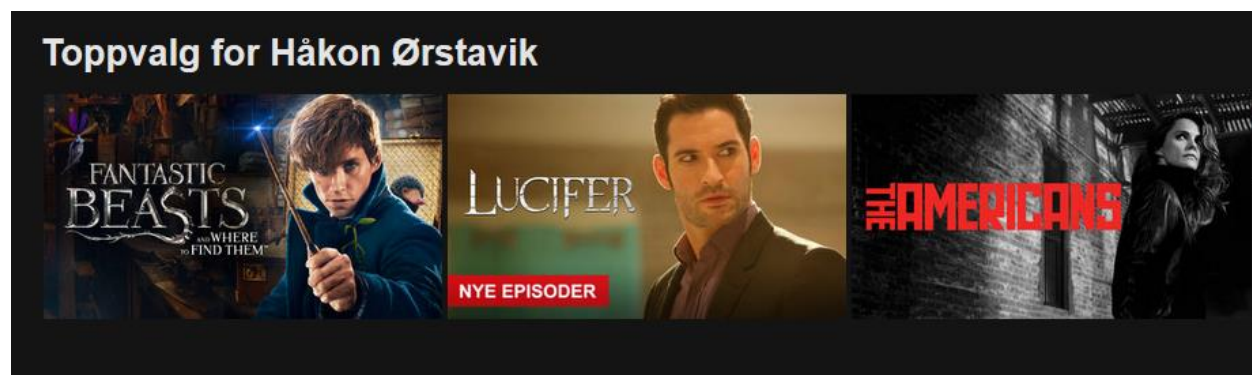
HVA GJØR EN ANBEFALINGSMOTOR

- Finn meg noe som ligner på dette.

Customers who bought this item also bought

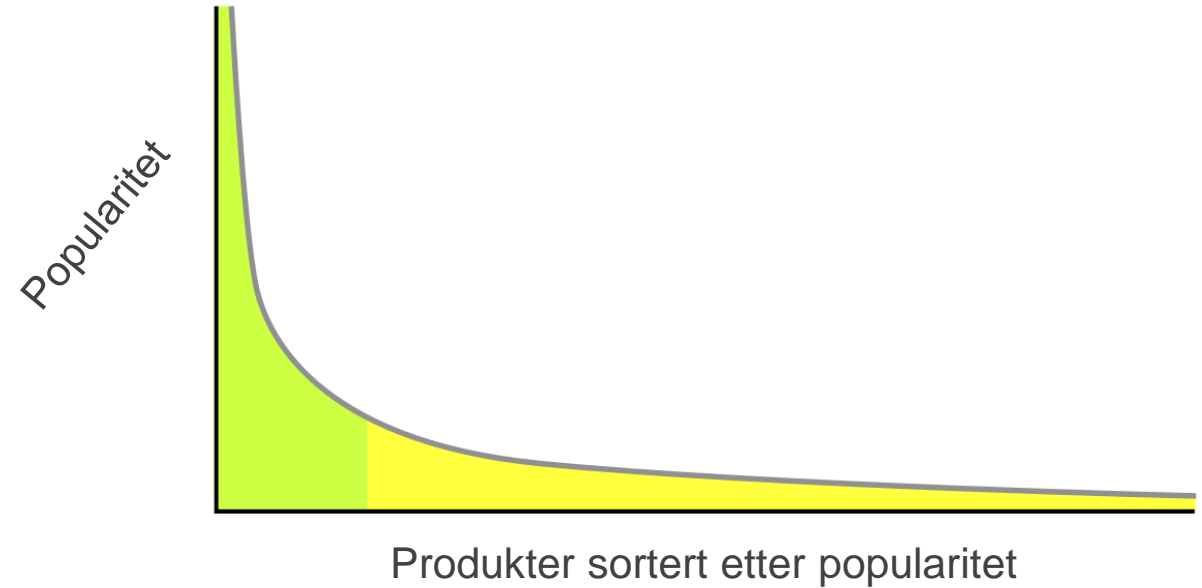


- Lag en liste som passer for meg.



MOTIVASJON FOR ANBEFALINGER

- Treffe brukere med relevant innhold
- Øke salg
- Treffe flere brukere
- Utnytte større del av sortimentet

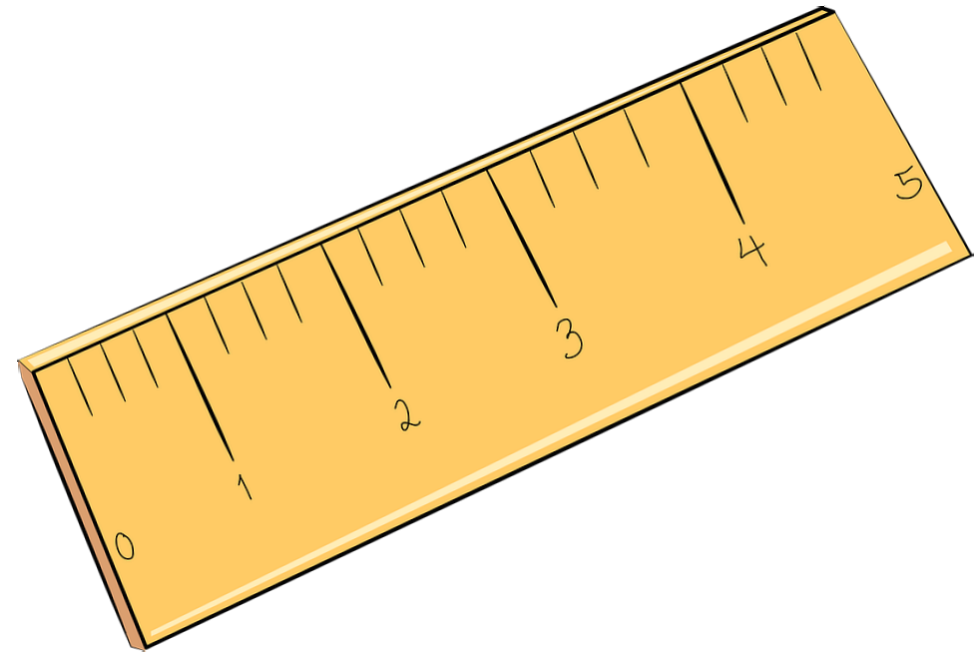


TRE DATATYPER

- Brukerdata
 - Implisitt: Interaksjoner mellom brukere og produkter/systemet. (vær kreativ)
 - Eksplisitt: Eksplisitte tilbakemeldinger fra brukere.
- Metadata
 - Data som beskriver hver ting. (Spilletid, sideantall, word bag osv)
 - Gjøres ofte om til features.
- Kontekst
 - Hvem er brukeren?
 - Tid og sted

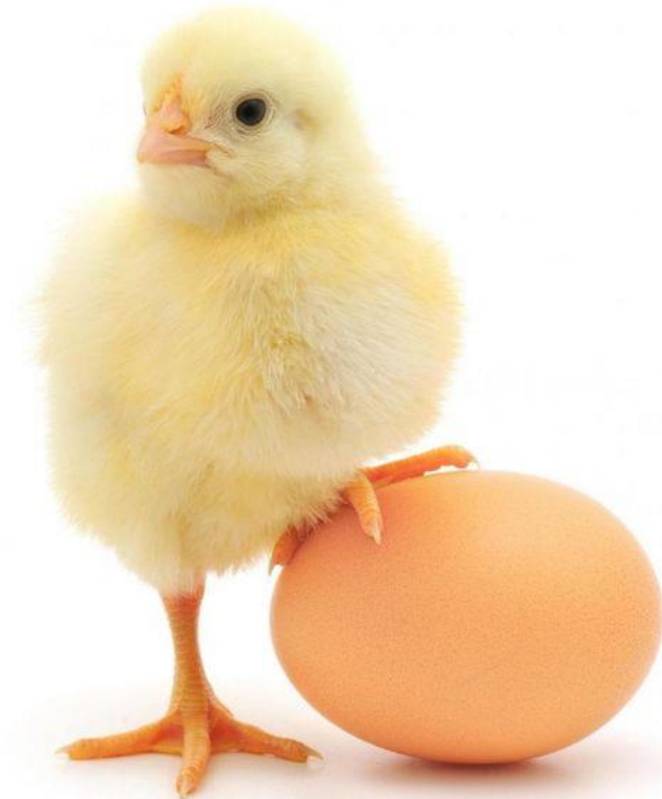
HVORDAN MÅLER VI ANBEFALEREN?

- Precision
- Coverage
- Metodespesifikke målinger
 - RMSE og venner
- Brukertestester



COLD START

- Nye brukere
 - Hva liker de?
 - Intervjurunde
- Nytt innhold
 - Hvem liker denne filmen?



EKSTERNT INNHOLD

- Content is king
- Må lage anbefalinger mellom flere domener





AGENDA

Hva er en anbefaler?

1. Hvorfor
2. Hvordan?

Content based filtering

1. Vektorprodukt
2. Fordeler og ulemper

Collaborative filtering

1. I sin simpleste form
2. Nabolagsbasert
3. Modellbasert
4. Fordeler og ulemper

Oppsummering(?)

CONTENT BASED FILTERING

- Alt innhold har attributer (*features*)
 - Krever *feature engineering*
- Lager en modell av brukeren basert på adferd
- Baserer seg på at du liker ting som har tilsvarende *features*
- Har mange likheter med søk. Du søker med en artikkel eller en brukerprofil.

```
{
  "budget": 28000000,
  "genres": [
    "Action",
    "Thriller"
  ],
  "homepage": "",
  "id": 562,
  "imdb_id": "tt0095016",
  "original_language": "en",
  "original_title": "Die Hard",
  "overview": "NYPD cop, John McClane's plan to reconcile with his estranged wife is ... bad guys down.",
  "popularity": 4.115053,
  "release_date": "1988-07-15",
  "revenue": 140767956,
  "runtime": 131,
  "tagline": "40 Stories. Twelve Terrorists. One Cop.",
  "title": "Die Hard",
  "video": false,
  "vote_average": 7.4,
  "vote_count": 3524
}
```

CONTENT BASED FILTERING

Features: (old, violent, colourful, animated)

Metropolis : (1.0, 0.5, -1.0, 0.0)

Inception : (0, 0.5, 0.5, 0.0)

Askepott : (0.2, 0.5, 1.0, 1.0)

Min profil : (0.5, 1.0, -1.0, -1.0)

Metropolis : 2

Inception : 0

Askepott : -1.4



AGENDA

Hva er en anbefaler?

1. Hvorfor
2. Hvordan?

Content based filtering

1. Vektorprodukt
2. Fordeler og ulemper

Collaborative filtering

1. I sin simpleste form
2. Nabolagsbasert
3. Modellbasert
4. Fordeler og ulemper

Oppsummering(?)

COLLABORATIVE FILTERING

- Lager anbefalinger kun basert på brukerdata.
 - Eksplisitte tilbakemeldinger
 - Implisitte tilbakemeldinger
- Baserer seg på at brukere som har noen felles interesser har flere felles interesser.

People who liked this also liked... [Learn more](#)



Breaking Bad (2008–2013)
Certificate: 15 Crime | Drama | Thriller
★★★★★ 9,5/10

A high school chemistry teacher diagnosed with inoperable lung cancer turns to manufacturing and selling methamphetamine in order to secure his family's future.

[Add to Watchlist](#)
[Next »](#)

Stars: Bryan Cranston, Aaron Paul, ...

MADE FOR HÅKON

Discover Weekly

Your weekly mixtape of fresh music. Enjoy new discoveries and deep cuts chosen just for you. Updated every Monday, so save your favourites!

Made for Håkon Dissen by Spotify • 30 songs, 2 hr 18 min

CO-OCCURENCE

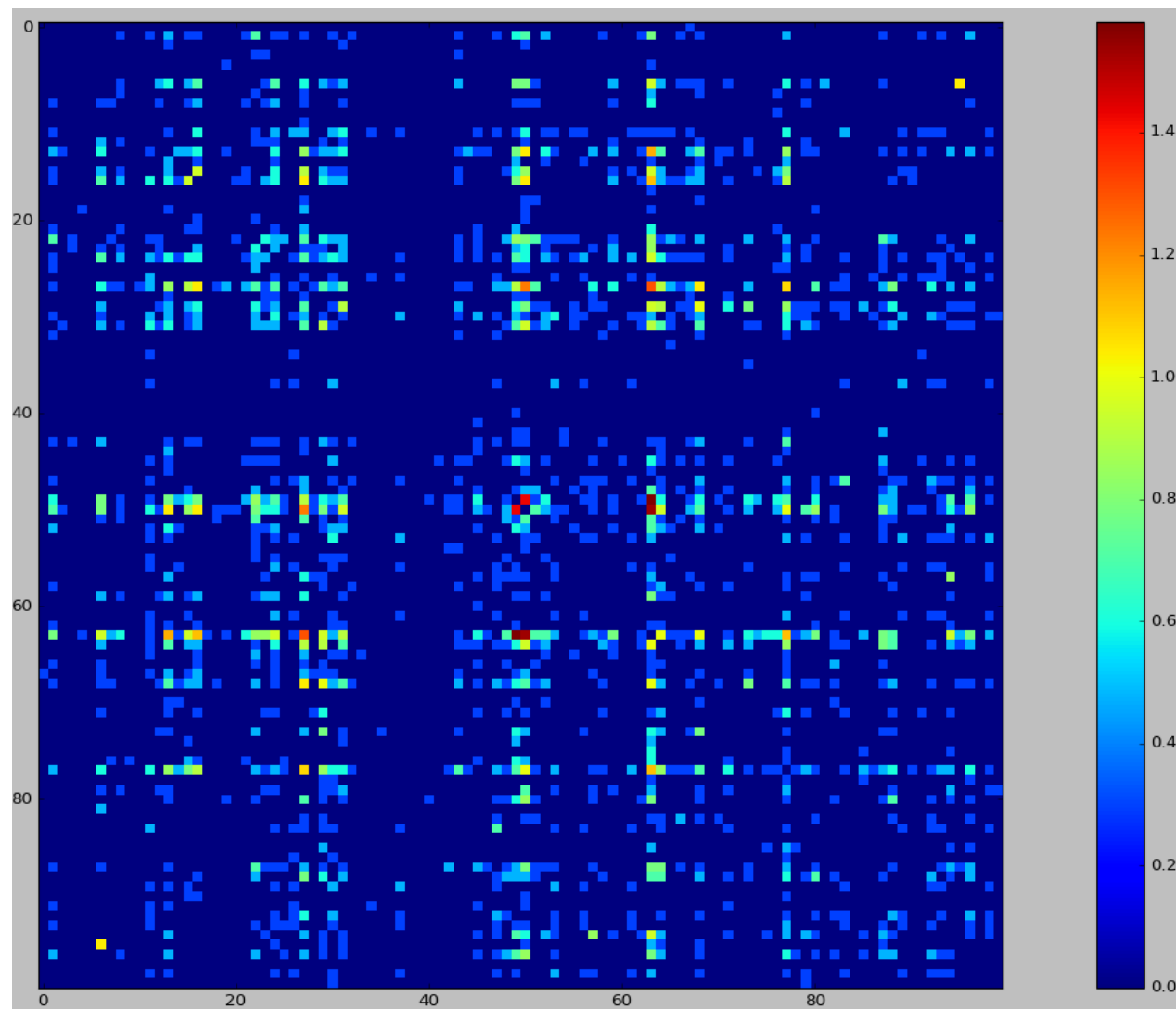
u : bruker

i : ting (f.eks film)

$$\mathbf{H} = \begin{matrix} & i_1 & i_2 & i_3 & i_4 & i_5 \\ \begin{matrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{matrix} & \begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

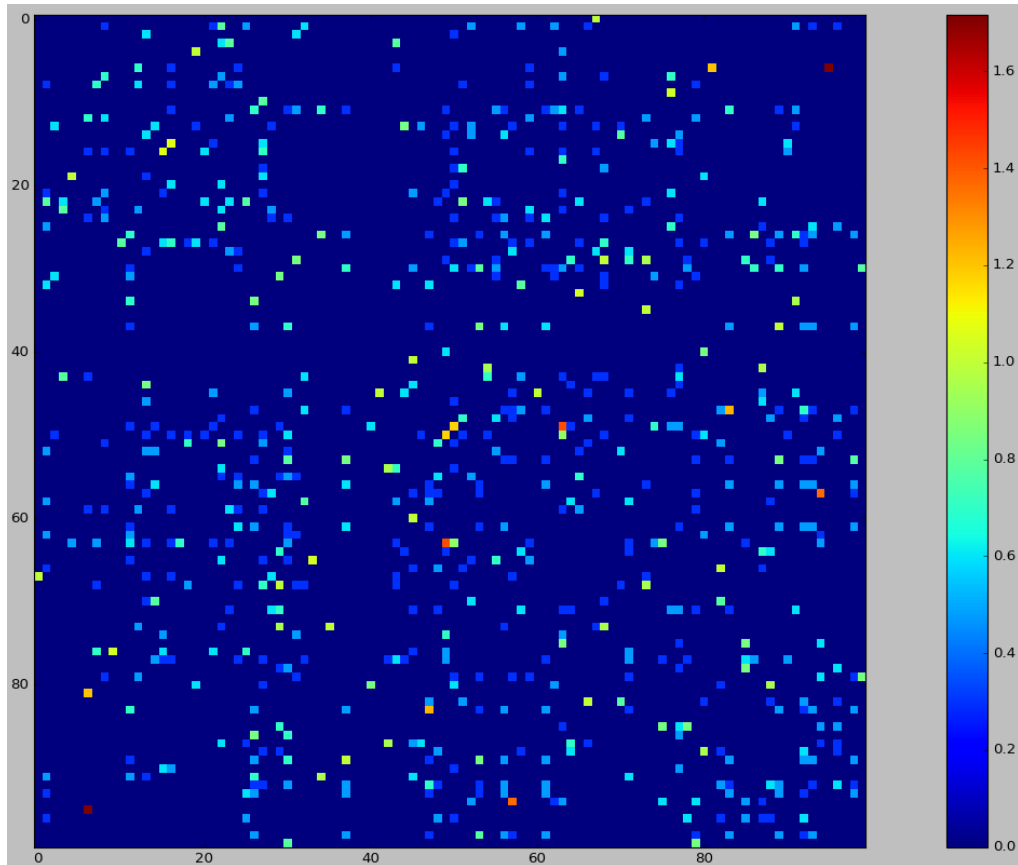
$$\mathbf{C} = \mathbf{H}^T \mathbf{H} = \begin{matrix} & i_1 & i_2 & i_3 & i_4 & i_5 \\ \begin{matrix} i_1 \\ i_2 \\ i_3 \\ i_4 \\ i_5 \end{matrix} & \begin{pmatrix} 3 & 1 & 1 & 1 & 2 \\ . & 2 & 1 & 2 & 1 \\ . & . & 1 & 1 & 0 \\ . & . & . & 2 & 1 \\ . & . & . & . & 4 \end{pmatrix} \end{matrix}$$

CO-OCCURRENCE

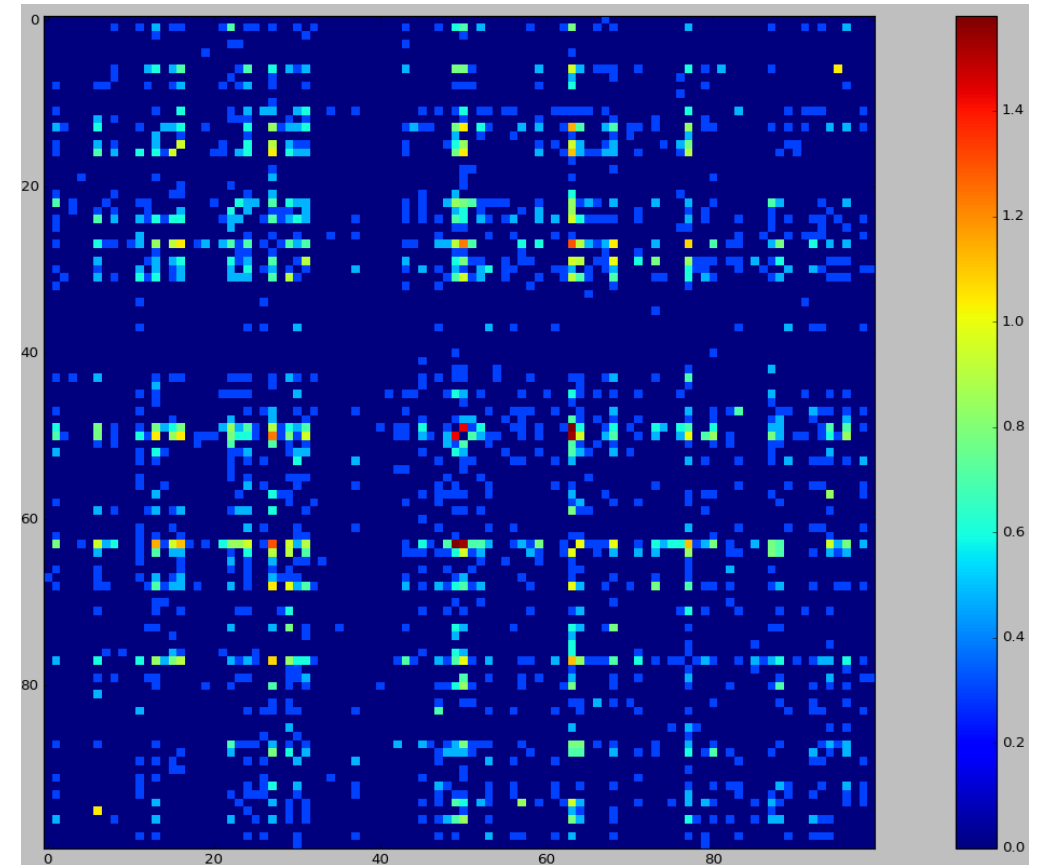


LOG LIKELIHOOD RATIO

LLR



CO-OC



OTTOBOT MUVP



Håkon Dissen 12:27 PM ☆

[@ottobot](#) american dad



ottobot APP 12:27 PM

Dude, you should totally watch:
Futurama,
South Park,
Simpsons,
Family Guy,
The Big Bang Theory



Håkon Dissen 12:27 PM ☆

[@ottobot](#) håndball-em: frankrike-norge



ottobot APP 12:27 PM

Dude, you should totally watch:
Håndball-EM: Norge-Russland,
Håndball-EM: Tsjekkia - Norge,
Håndball-EM: Kroatia-Norge,
Håndball-EM: Norge-Romania,
Håndball-EM: Ungarn - Norge



Håkon Dissen 12:37 PM ☆

[@ottobot](#) the kardashians



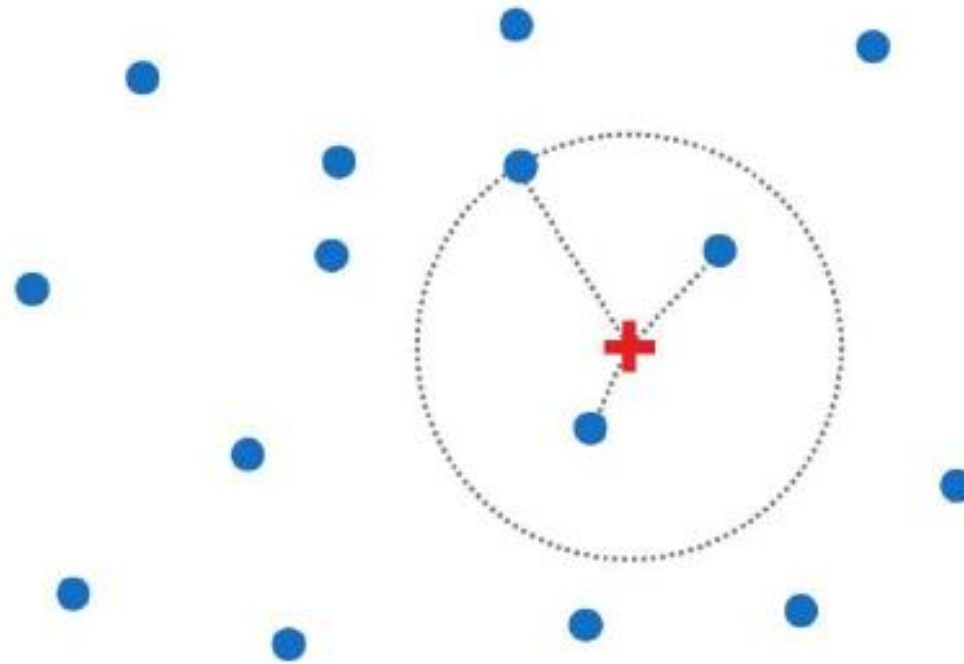
ottobot APP 12:37 PM

Dude, you should totally watch:
Kjolen som sier ja,
Say Yes to the Dress,
The Real Housewives of Orange County,
Svenske Hollywoodfruer,
Millionær søker kjæreste

CF: NABOLAG

User id	Terminator	Gladiator	Alien
1	4	1	4
2	5	2	5
3	1	4	1
4	5	?	?

CF: NABOLAG



Illustrasjon fra: https://www.researchgate.net/figure/280620610_fig1_Figure-2-Illustration-of-the-kNN-search-strategy-from-45-The-red-cross-indicates

CF: NABOLAG

User id	Terminator	Gladiator	Alien
1	4	1	4
2	5	2	5
3	1	4	1
4	5	$= (1 + 2) / 2 = 1.5$	$= (4 + 5) / 2 = 4.5$

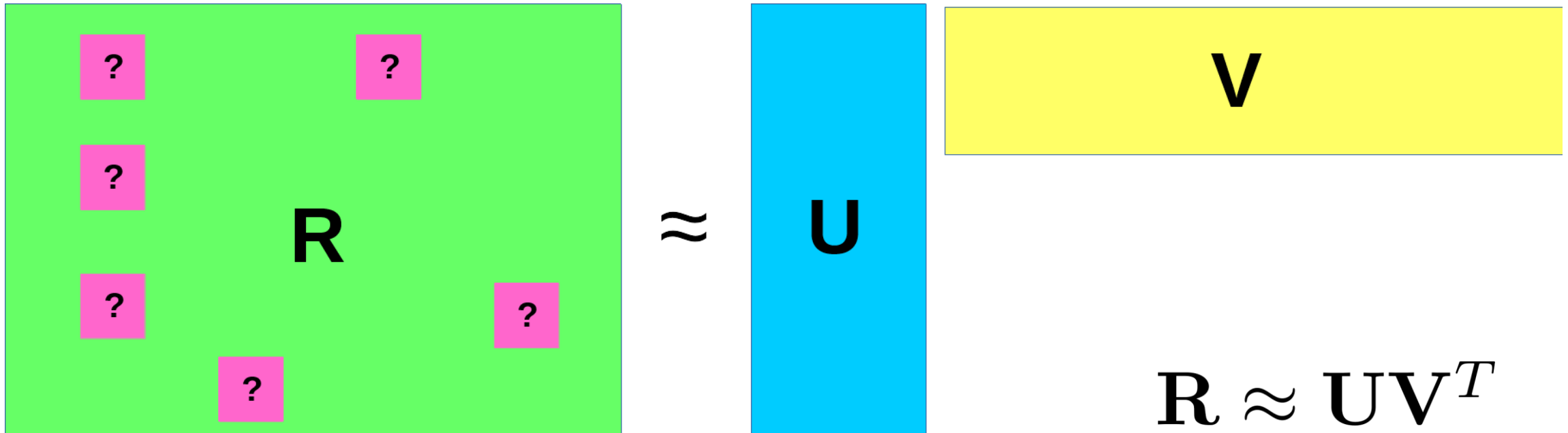
CF: MODELLBASERT

R =

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1	0	3	0	3	0
User 2	4	0	0	2	0
User 3	0	0	3	0	0
User 4	3	0	4	0	3
User 5	4	3	0	4	0

$0 = ukjent$

CF: MODELLBASERT



$$R \approx UV^T$$

$$r_{ij} \approx u_i v_j$$

NERO	JULIUS CAESAR	CLEOPATRA	THE NOTEBOOK	PRETTY WOMAN	CASABLANCA
1	1	1	0	0	0
1	1	1	0	0	0
1	1	1	0	0	0
1	1	1	1	1	1
-1	-1	-1	1	1	1
-1	-1	1	1	1	1
-1	-1	-1	1	1	1

\approx

HISTORY	ROMANCE
1	0
1	0
1	0
1	1
-1	1
-1	1
-1	1

\times

NERO	JULIUS CAESAR	CLEOPATRA	THE NOTEBOOK	PRETTY WOMAN	CASABLANCA	
1	1	1	0	0	0	HISTORY
0	0	1	1	1	1	ROMANCE

$$\mathbf{R} \approx \mathbf{U}\mathbf{V}^T$$

$$\mathbf{r}_{ij} \approx \mathbf{u}_i \mathbf{v}_j$$



AGENDA

Hva er en anbefaler?

1. Hvorfor
2. Hvordan?

Content based filtering

1. Vektorprodukt
2. Fordeler og ulemper

Collaborative filtering

1. I sin simpleste form
2. Nabolagsbasert
3. Modellbasert
4. Fordeler og ulemper

Oppsummering(?)

SAMMENLIKNING

COLLABORATIVE FILTERING

- Benytter seg av brukerdata
- Sliter med cold-start
 - Både for artikler og brukere
- Bruker lang tid på å komme i gang
- Krever ikke feature engineering

CONTENT BASED FILTERING

- Benytter seg av metadata
- Sliter også med cold-start, men bare for brukere
- Rask på å komme i gang
- Krever feature engineering / feature extraction



itera

MAKE A DIFFERENCE