



Teknologiforståelse *gennem* matematik
Hvordan ved din streamingtjeneste, hvad du vil se?

Udgangspunkt i forskning

LEMMA 8. *There are g urns, and exactly $d > 0$ of them are not empty. Furthermore, assume one can check if a specific urn is empty in constant time. Let $\gamma \in (0, 1)$ be a parameter. Then one can output a number $Z \geq 0$, such that $Z \in [0, 1]$, and $\mathbb{E}[Z] \in I = [\frac{1}{d\Delta} - \gamma, \frac{1}{d\Delta}]$, where $\Delta = \lceil \ln \gamma^{-1} \rceil + 4 = \Theta(\log \gamma^{-1})$. The expected running time of the algorithm is $O(g/d)$.*

Alternatively, the algorithm can output a bit X , such that $\mathbb{P}[X = 1] \in I$.

PROOF. We modify the algorithm of Lemma 7, so that it outputs $i/(g\Delta)$ instead of i/g . If the algorithm does not stop in the first $g\Delta + 1$ iterations, then the algorithm stops and outputs 0. Observe that the probability that the algorithm fails to stop in the first $g\Delta$ iterations, for $p = d/g$, is $(1 - p)^{g\Delta} \leq \exp(-\frac{d}{g}g\Delta) \leq \exp(-d\Delta) \leq \exp(-\Delta) \ll \gamma$.

Let Z be the random variable that is the number output by the algorithm. Arguing as in Lemma 7, we have that $\mathbb{E}[Z] \leq 1/(d\Delta)$. More precisely, we have $\mathbb{E}[Z] = \frac{1}{d\Delta} - \sum_{i=g\Delta+1}^{\infty} \frac{i}{g\Delta} (1-p)^{i-1} p$. Let

$$\begin{aligned} \sum_{i=gj+1}^{g(j+1)} \frac{i}{g} (1-p)^{i-1} p &\leq (j+1) \sum_{i=gj+1}^{g(j+1)} (1-p)^{i-1} p = (j+1)(1-p)^{gj} \sum_{i=0}^{g-1} (1-p)^i p \\ &\leq (j+1)(1-p)^{gj} \leq (j+1) \left(1 - \frac{d}{g}\right)^{gj} \leq (j+1) \exp(-dj). \end{aligned}$$

Let $g(j) = \frac{j+1}{\Delta} \exp(-dj)$. We have that $\mathbb{E}[Z] \geq \frac{1}{d\Delta} - \beta$, where $\beta = \sum_{j=\Delta}^{\infty} g(j)$. Furthermore, for $j \geq \Delta$, we have

$$\frac{g(j+1)}{g(j)} = \frac{(j+2) \exp(-d(j+1))}{(j+1) \exp(-dj)} \leq \left(1 + \frac{1}{\Delta}\right) e^{-d} \leq \frac{5}{4} e^{-d} \leq \frac{1}{2}.$$

As such, we have that

$$\beta = \sum_{j=\Delta}^{\infty} g(j) \leq 2g(\Delta) \leq 2 \frac{\Delta+1}{\Delta} \exp(-d\Delta) \leq 4 \exp(-\Delta) \leq \gamma,$$

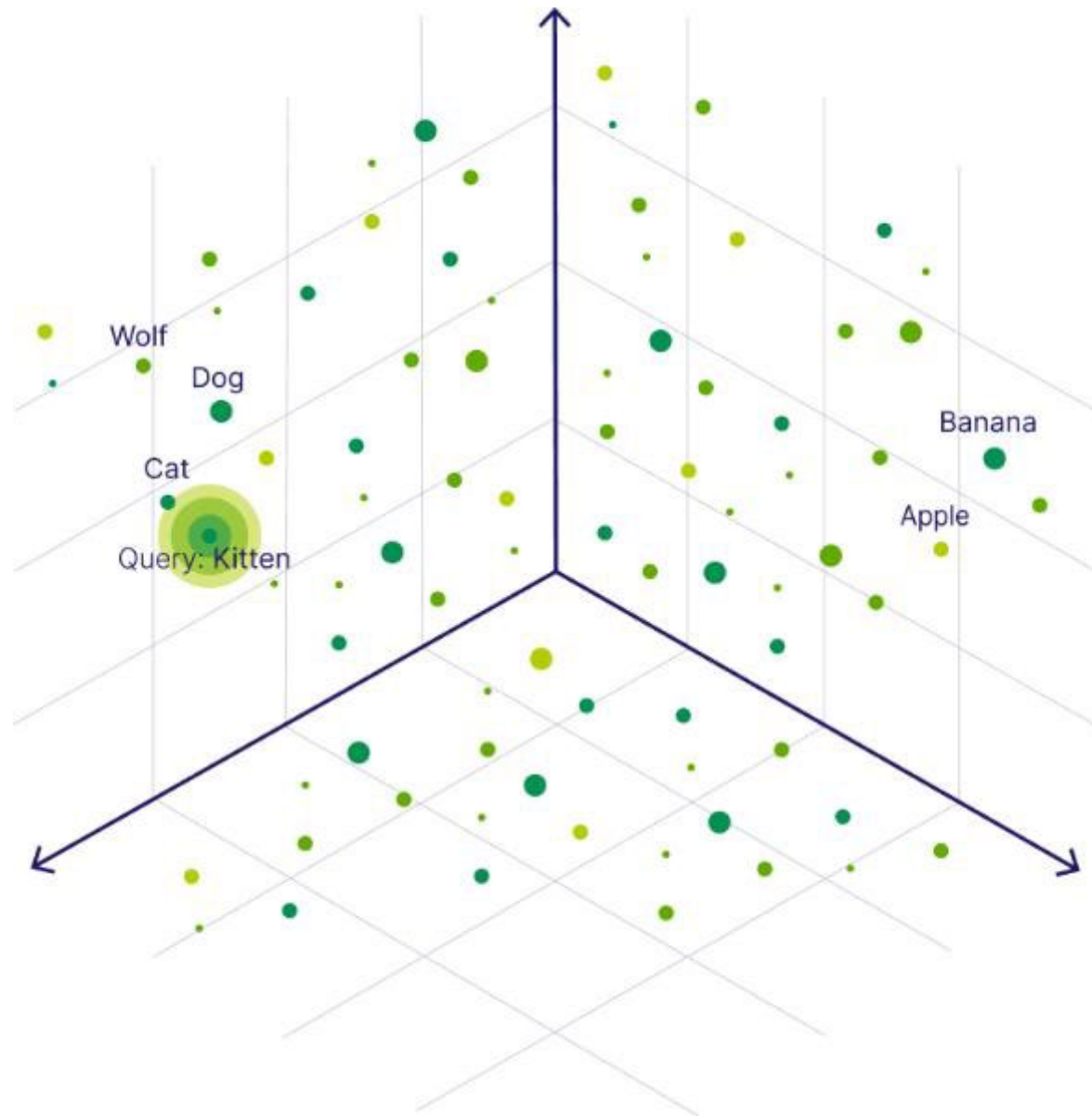
by the choice of value for Δ . This implies that $\mathbb{E}[Z] \geq 1/(d\Delta) - \beta \geq 1/(d\Delta) - \gamma$, as desired.

The alternative algorithm takes the output Z , and returns 1 with probability Z , and zero otherwise. $\square$

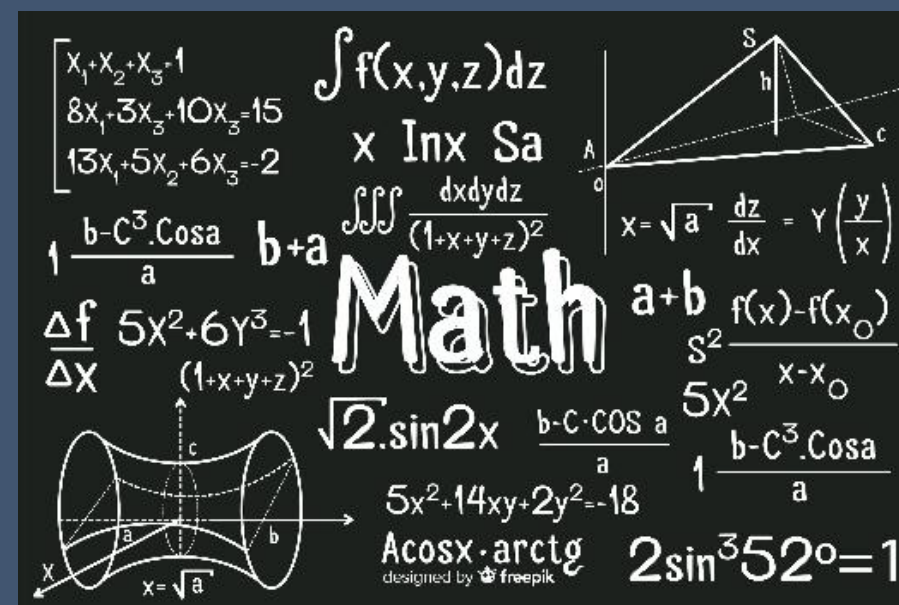
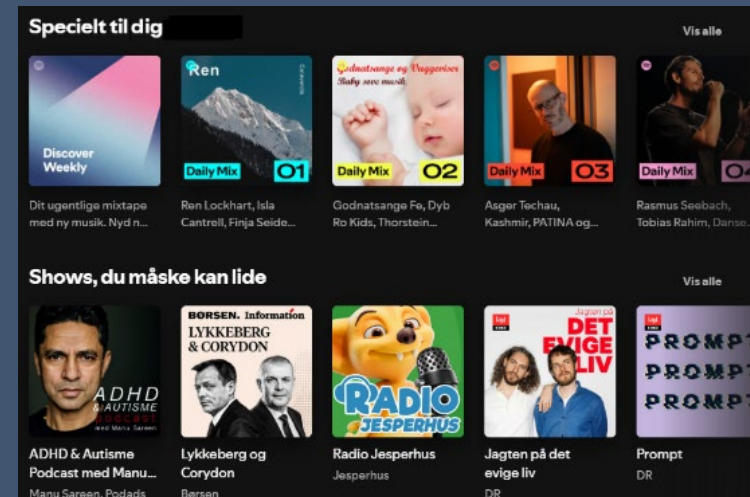
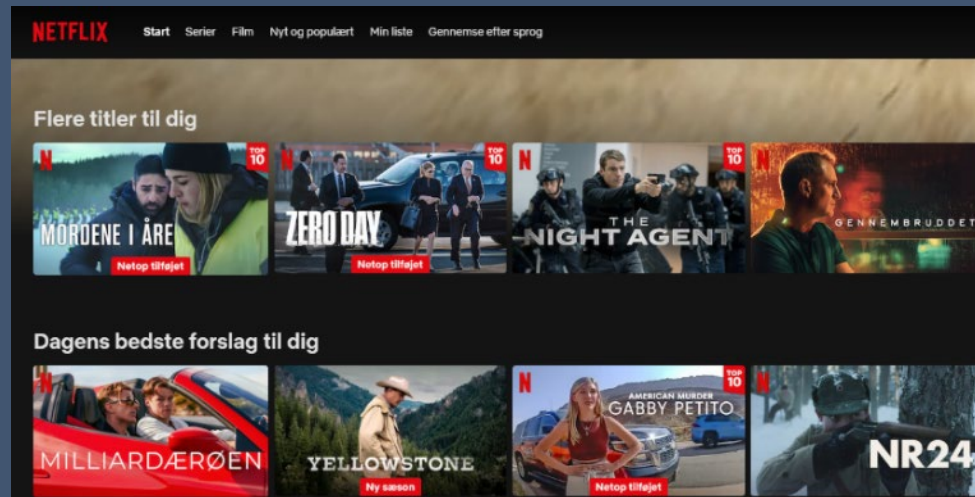
- De-kontekstualiseret
- De-personaliseret (Brousseau, 2003)

Aumüller et al. (2022)

- Similarity search i mange dimensioner
- Nærmeste nabo algoritmer bliver langsomme
- Der findes hurtigere metoder, men er de fair?



Streamingtjenesternes anbefalinger



Overblik

- ▶ Anbefal fire film til Lucca og snak om anbefalingsalgoritmer
- ▶ Video: Hvad betyder K-NN?
- ▶ Øvelse med K-NN og Disney+
- ▶ Video: Hvad betyder A-NN?
- ▶ Øvelse med A-NN og Spotify
- ▶ Opsamling og dilemmaer

Læringsmål

- I kan udregne afstanden mellem to punkter i et koordinatsystem
- I kan opsamle og sortere data i regneark
- I kender til nogle af de algoritmer, som bruges til anbefalinger; K-NN og A-NN
- I forstår, hvordan sange, film mv. kan modelleres i et koordinatsystem vha. deres egenskaber



Rasmus Pagh

- ▶ Professor i datalogi på Københavns Universitet
- ▶ Forsker i privacy (beskyttelse af personlige data) og maskinlæring. Bl.a. hvordan algoritmer og datastrukturer fungerer.

—

Datastrukturer er den måde, data bliver gemt på. Lidt som en hylde med bøger: Du kan sætte en ny bog op, fjerne en bog, søge efter en bestemt bog og se, om hylden er tom.

—

Algoritmer beskriver præcis, hvad computeren skal gøre. Ligesom en opskrift når du laver mad.



Jeres anbefaling

- ▶ Udgangspunkt i Frost
- ▶ Hvilke film anbefalede I?
- ▶ Hvordan kom I frem til jeres anbefaling?



Pinocchio
(1940)



Løvernes Konge
(1994)



Pocahontas
(1995)



Toy Story
(1995)



Herkules
(1997)



Mulan
(1998)



Tarzan
(1999)



Toy Story 2
(1999)



Fantasia
(2000)



Monsters, Inc.
(2001)



Find Nemo
(2003)



De utrolige
(2004)



Biler
(2006)



Ratatouille
(2007)



Wall-E
(2008)



Op
(2009)



To på flugt
(2010)



Modig
(2012)



Frost
(2013)



Maleficent
(2014)



Big Hero 6
(2014)



Eventyret om
Askepot (2015)



Inderst inde
(2015)



Zootopia
(2016)



Find Dory
(2016)



Vaiana
(2016)



Skønheden og
udyret (2017)



Frost 2
(2019)



Encanto
(2021)



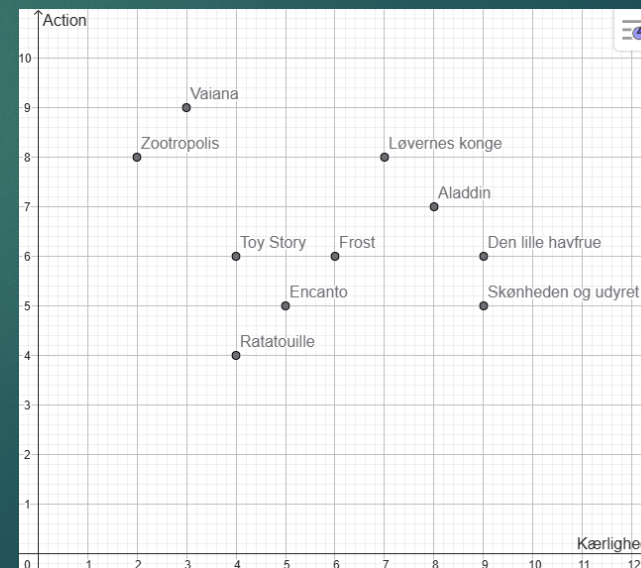
Snehvide
(2025)

Video om K-NN algoritmen

▶ <https://youtu.be/SnxK9yjPhGg>

Arbejd med K-NN

- ✓ Vælg 10 film og bedøm kærlighed og action fra 0-10.
- ✓ Indsæt dem som punkter i koordinatsystemet.
- ▶ Vælg hvilken film, du så sidst, og marker den tydeligt.
- ▶ Tal i gruppen/klassen om, hvordan I kan beregne afstanden.
- ▶ Beregn afstanden fra jeres valgte punkt til hvert af de andre punkter.
- ▶ Find ved hjælp af jeres udregninger de tre nærmeste punkter. Synes I, det nærmeste punkt det bedste forslag?
- ▶ EKSTRAOPGAVE: Hvad er den største afstand, der kan være mellem to film?



Video om A-NN algoritmen

▶ <https://youtu.be/3hdoVk2Wqz0>



Happiness

- ▶ Giver sangen dig et smil på læben?
- ▶ Bliver du glad af at høre sangen?
- ▶ Sange med lav happiness føles mere triste.
- ▶ Sang med høj happiness:
<https://youtu.be/TUVcZfQe-Kw?list=RDTUVcZfQe-Kw&t=131>



Danceability

- ▶ Tempo i sangen
- ▶ Hvor stabil rytmen i sangen er
- ▶ Styrken på beatet
- ▶ Sang med høj danceability:
https://youtu.be/WWZ_nM4IWd4?list=RDWWZ_nM4IWd4&t=101

Arbejd med A-NN

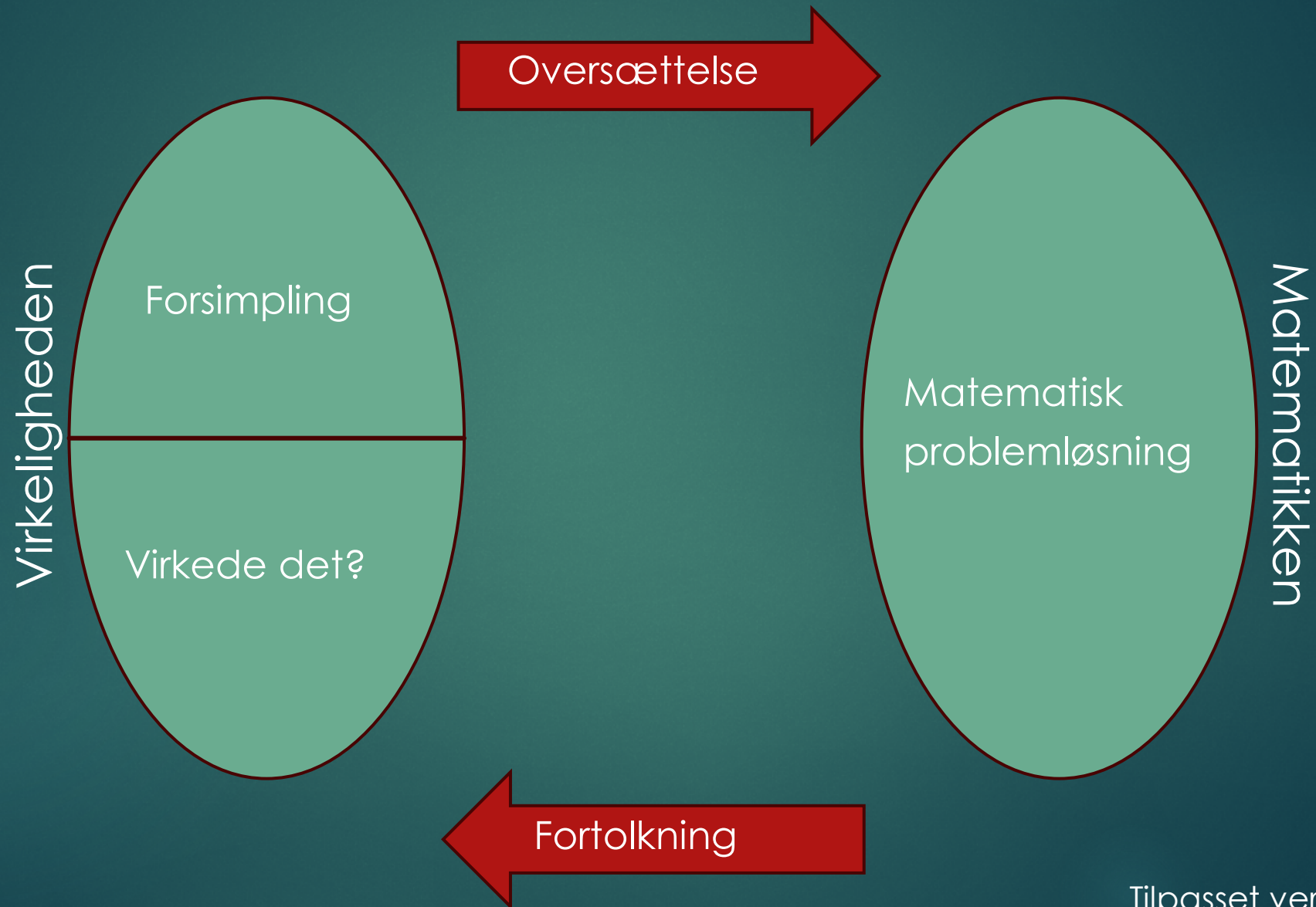
- ▶ Musikken er lavet om til punkter.
- ▶ Koordinatsystemet deles ind i felter.
- ▶ Vælg det musiknummer, du hørte sidst, og skift punktets farve til rød.
- ▶ Hvis der er punkter i samme boks, som det valgte nummer, så sæt et kryds i regnearket.
- ▶ Opdater 10 gange og sæt krydser i regnearket hver gang.
- ▶ Sorter sangene i regnearket, så dem, der har været i samme felt som sangen flest gange, ligger øverst.
- ▶ Anbefalingen er det punkt, som ligger i samme felt flest gange.

[geogebra.org/classic/ttbrxy6j](https://www.geogebra.org/classic/ttbrxy6j)

Sheets: kortlink.dk/2sh4u

Excel: kortlink.dk/2sh4p

Matematisk modellering



Opsamling



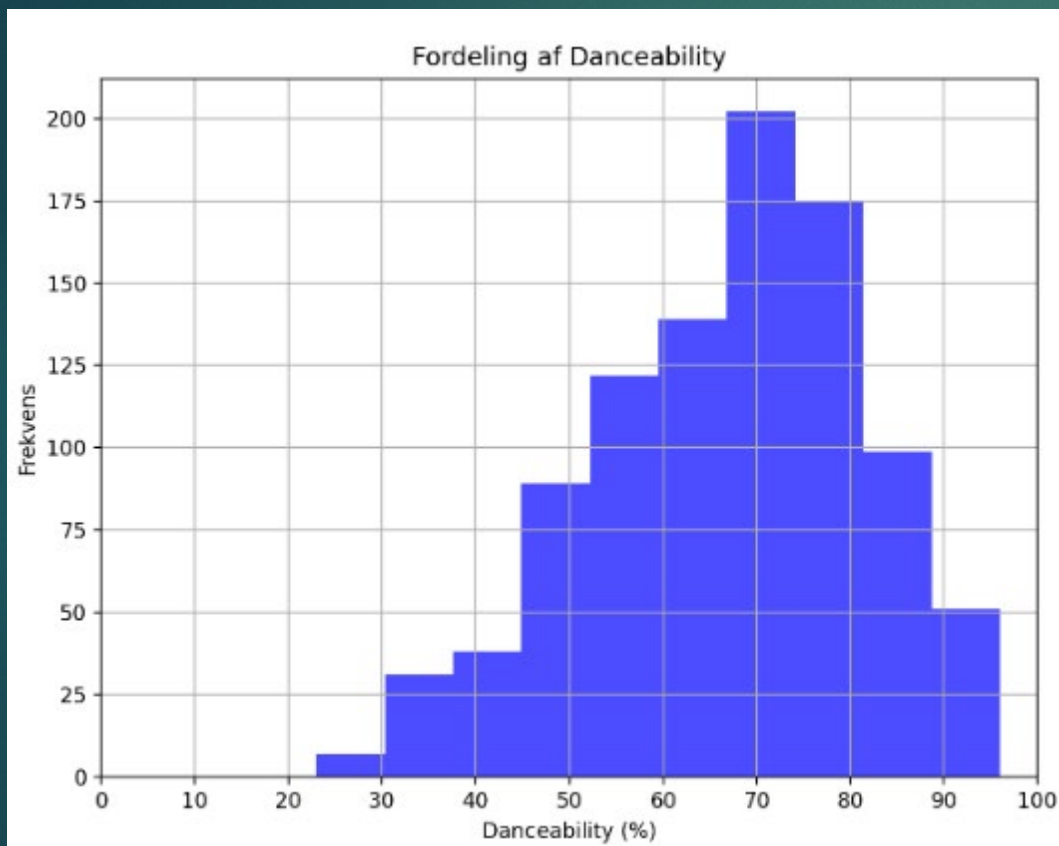
Rasmus Pagh siger: *“Det er ønskeligt at anbefale brugeren noget, der minder om emnet, de kigger på, men også noget, der kan give dem et andet perspektiv.”*

► Hvordan kan det opnås?

Spotify anbefaler musik, som brugere med samme smag som dig godt kan lide.

► Hvad kan det betyde for nye ukendte kunstnere?

Algoritmernes magt



Du er ny kunstner og udgiver et nummer med danceability på 10.

- ▶ Hvor mange andre numre ligner din musik?
- ▶ Hvor ofte vil du blive foreslået?
 - ▶ Brug det, I har lært om K-NN og A-NN, til at begrunde jeres svar

Undersøg forløbene

- De to betaudgaver ligger på
 - www.edtalk.dk/albatros-1
 - www.edtalk.dk/albatros-2
- Kig dem igennem
- Forbered din undervisning med dem
- Stil spørgsmål og få gode råd

Mere fra ALBATROS

- Hør vores podcasts om teknologiforståelse i matematik
- Afprøv betaudgaver af vores forløb
- Mød os på BigBang i Odense 18.-19. marts 2026
- Kontakt: pws@ind.ku.dk

