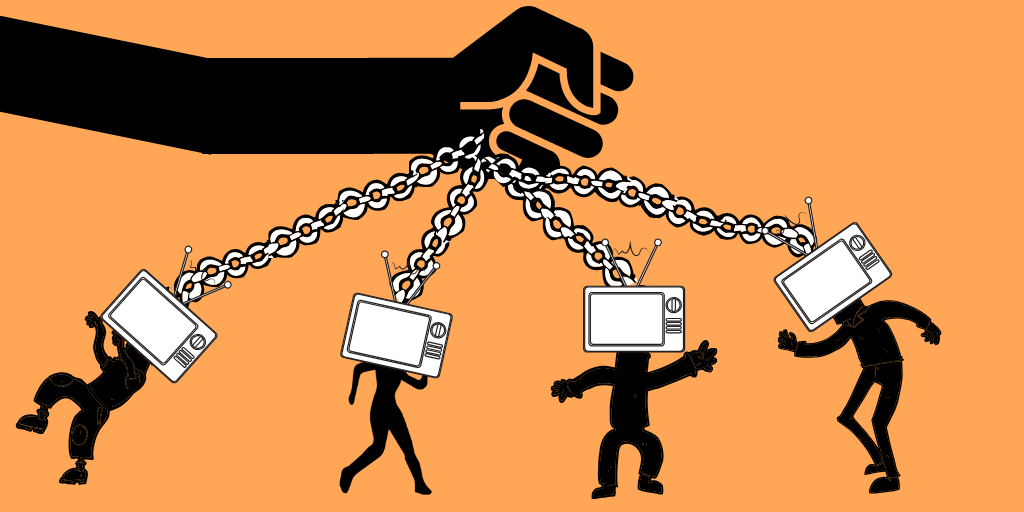
**Opgavens indhold:**

(mere eller mindre kronologisk)

* Forside
* Indholdsfortegnelse

* [] Indledning
* [] Problemformulering
* [] Den faglige gennemgang:
  + Redegørelse
  + Analyse
* [] Konklusion og diskussion

* [] Kilde og litteraturliste
* [] Bilag



# 

# Data misbruges til at profilere

Indledning til emne

Hvilke problemer er der?

**Redegørende**

Forklaring om data

* Hvad er data?
* Hvad er GDPR

Machine Learning

* Discrete
  + Supervised & unsupervised / Classification & Clustering

Hvordan gjorde Cambridge Analytica? (ikke detaljeret)

**Analyserende**

Hvorfor er --- etisk ukorrekt?

Er nogen love blevet brudt i processen?

Analyser, hvordan data faktisk blev brugt (brug evt.. billede)

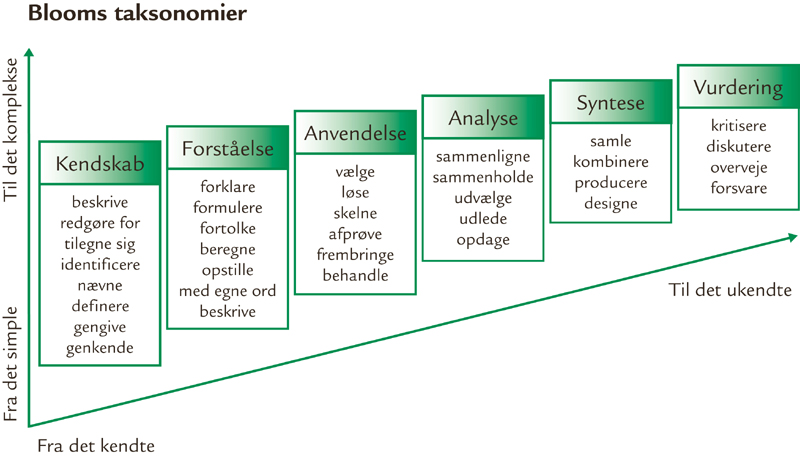
Snyd i demokrati (propaganda)

Hvorfor er det vigtigt at beskytte sin data?

Menneskelig rettighed

3 typer data

Løser GDPR problemet?



**Vurderende**

Hvordan er data-friheden blevet bedre?

Kan der ses nogen forskel?

Hvordan bliver valg-kampe anderledes?

Har GDPR gjort nok? Skulle man have gjort noget andet?

### Problemformulering

De digitale fodspor, som vi efterlader, når vi ‘liker’ og kommenterer posts på sociale medier, er i virkeligheden vigtig data, som bruges til at lave psykometriske profileringer af os. Denne form for data-analystik bruges ofte til at lave harmløse reklamer, som appelerer til sine målgrupper. Desværre har der i forbindelse med Trump-valget været stort fokus på misbrug af denne data. Når man ved nok om en person, kan man udnytte individuelles svagheder, og manipulere folk til at gøre nogle bestemte handlinger, til andres fordel. Denne slags data-misbrug kan ses i form af propaganda i medierne. Vdr. skandalen med Cambridge Analytica i 2016 så man også, at reklamer direkte var rettede mod letpåvirkede og mentalt svage mennesker (*persuadables*).

Hvordan kan det lykkes, at man kan lære nok om en person, til at man kan manipulere dem til at stemme anderledes i valg?

(Eventuelle arbejdsspørgsmål)

Hvilke metoder bruger data-analytikerer til at klassificere og profilere internetbrugere?

*Hvorfor kalder Christopher Wylie CA’s datamisbrug for psykologisk krigsførsel?*

Hvordan kan/bliver samme teknologi udnyttet andre steder i verden?

### Opgavens Pentagon

**• Hvad spørger du om? (Hvad er din problemformulering)?**

Jeg spørger ind til, hvordan man med en teknologi (Machine Learning) kan profilere folk, og hvordan dette kan bruges til at skabe propaganda.

*(se arbejdsspørgsmål)*

**• Hvorfor spørger du? (Hvad er de faglige problemstillinger? Hvad er interessant? Hvad undrer du dig over?)**

Data virker ligegyldig for mange mennesker, da det ikke direkte påvirker os. Dog er det interessant at se, hvorfor data er vigtigt at passe på, så vi ikke bliver udnyttede.

**• Hvad spørger du til? (Hvilket materiale skal du arbejde med? Hvad skal du analysere? Hvilken empiri vil du benytte)**

Cambridge Analytica har haft en lang historie med politiske kampagner i andre lande end blot USA og England. Samme strategier med brug af klassificering er også blevet brugt andre steder i verden. Jeg vil altså kigge på nogle er deres strategier

**• Hvad spørger du med? (Hvilke metoder og teori vil du anvende til din analyse?)**

Jeg har fundet flere dataset fra både personlighedstest og politiske præferencer. Ved at bruge klassificeringsalgoritmer (samt clustering), vil jeg prøve at vise, hvordan der er mønstre (klasser) mellem henholdsvis personligheder, og politiske præferencer afhængig af likede posts m.m.

**• Hvordan vil du gå frem?(Hvordan vil du strukturere og disponere din opgave?)**

I opgaven vil jeg komme ind på data. Dette inkluderer

1. Opsamling af data, og brug af andre menneskers data (inkl. brug af machine learning til at profilere internetbrugere)

Bagefter vil jeg komme ind på:

2. Vigtigheden af data, da vi nu ved, hvor stor en trussel data-misbrug kan være

For at forstå, hvor kraftfuld denne teknologi er, vil jeg kigge på:

3. Hvordan man bruger klassificering ved brug af big data, til at opnå gode ting i verden.

## K-means clustering

1. Input:

K = antal tyngdepunkter (‘*centroids’*)

Set punkter, x1 … xn

2) Placer tyngdepunkter tilfældigt på koordinatsættet

3) For hvert datapunkt, xi :

* Find nærmelse tyngdepunkt (euklidisk afstand)
* Sæt punktet xi til en gruppe (*‘cluster*’), j

4) For hvert ‘cluster’, j = 1 … K :

* Sæt et nyt tyngdepunkt, som svarer til gennemsnittet af alle datapunkternes positioner, xi, i nuværende cluster, j

5) Gentag fra *step 3* indtil ingen punkter ændrer sig

# Kilder

### Datasæt

Iris

<https://archive.ics.uci.edu/ml/datasets/Iris>

### Kode-snippet

Clustering i python

<https://datatofish.com/k-means-clustering-python/>

Iris datasæt i python

<https://github.com/vprusso/youtube_tutorials/blob/master/machine_learning/iris_classification/part_2.py>

### Billeder

### 

Figure 1 - https://www.netflix.com/watch/80117542?trackId=13752289&tctx=0%2C0%2C73724f3654e669fba64e68d5a801c05dfca6528c%3Ae24daacb7e5351f9501d2036e8d3c8c691178ca3%2C%2C

Figure 2 - https://www.youtube.com/watch?v=\_aWzGGNrcic

Figure 3 - https://www.youtube.com/watch?v=9991JlKnFmk

### Baggrundsviden

K-means clustering algoritme & forståelse af machine-learning algoritmer

<https://www.youtube.com/watch?v=9991JlKnFmk>

<https://www.youtube.com/watch?v=_aWzGGNrcic>

Facebook data til at forudse politisk interesse

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184562>

Cambridge Analytica og Facebook drama

<https://medium.com/@privacyint/what-zuckerberg-forgot-to-mention-profiling-7b7c596b9823>

<https://www.dr.dk/nyheder/viden/teknologi/overblik-saadan-fik-og-brugte-cambridge-analytica-50-millioner-profiler>

<https://www.netflix.com/watch/80117542?trackId=13752289&tctx=0%2C0%2C73724f3654e669fba64e68d5a801c05dfca6528c%3Ae24daacb7e5351f9501d2036e8d3c8c691178ca3%2C%2C>

<https://www.theguardian.com/commentisfree/2018/mar/23/plenty-more-like-cambridge-analytica-data-facebook>

<https://www.refinery29.com/en-us/2019/07/238738/what-happened-after-netflix-documentary-the-great-hack>

<https://en.wikipedia.org/wiki/Cambridge_Analytica#Elections>

Social profilering

<https://en.wikipedia.org/wiki/Social_profiling>

<https://www.researchgate.net/publication/293089145_Towards_semantic_dataset_profiling>

GDPR

<https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en>

<https://gdpr-info.eu/>

Data:

<https://blog.minitab.com/blog/understanding-statistics/understanding-qualitative-quantitative-attribute-discrete-and-continuous-data-types>

Menneskerettigheder

<https://amnesty.dk/om-amnesty/fns-verdenserklaering-om-menneskerettigheder>

<https://www.ohchr.org/EN/UDHR/Documents/UDHR_Translations/eng.pdf>