

MACHINE2LEARN

Wanneer moet je geen Deep Learning gebruiken?

25 september 2019, Gouda
Lisa Tostrams

Outline



MACHINE2LEARN

- Traditional ML
- Statistical ML
- Deep learning
- Waarom maakt het uit?
- ML in bedrijven
- Voorbeelden uit het nieuws

Whoami



MACHINE2LEARN

Lisa Tostrams

Assistent bij data mining, bayesian statistics, etc. bij de Radboud Universiteit
Machine Learning Engineer bij Machine2Learn

ARTIFICIAL INTELLIGENCE

Early artificial intelligence
stirs excitement.



1950's

1960's

1970's

1980's

MACHINE LEARNING

Machine learning begins
to flourish.



1990's

2000's

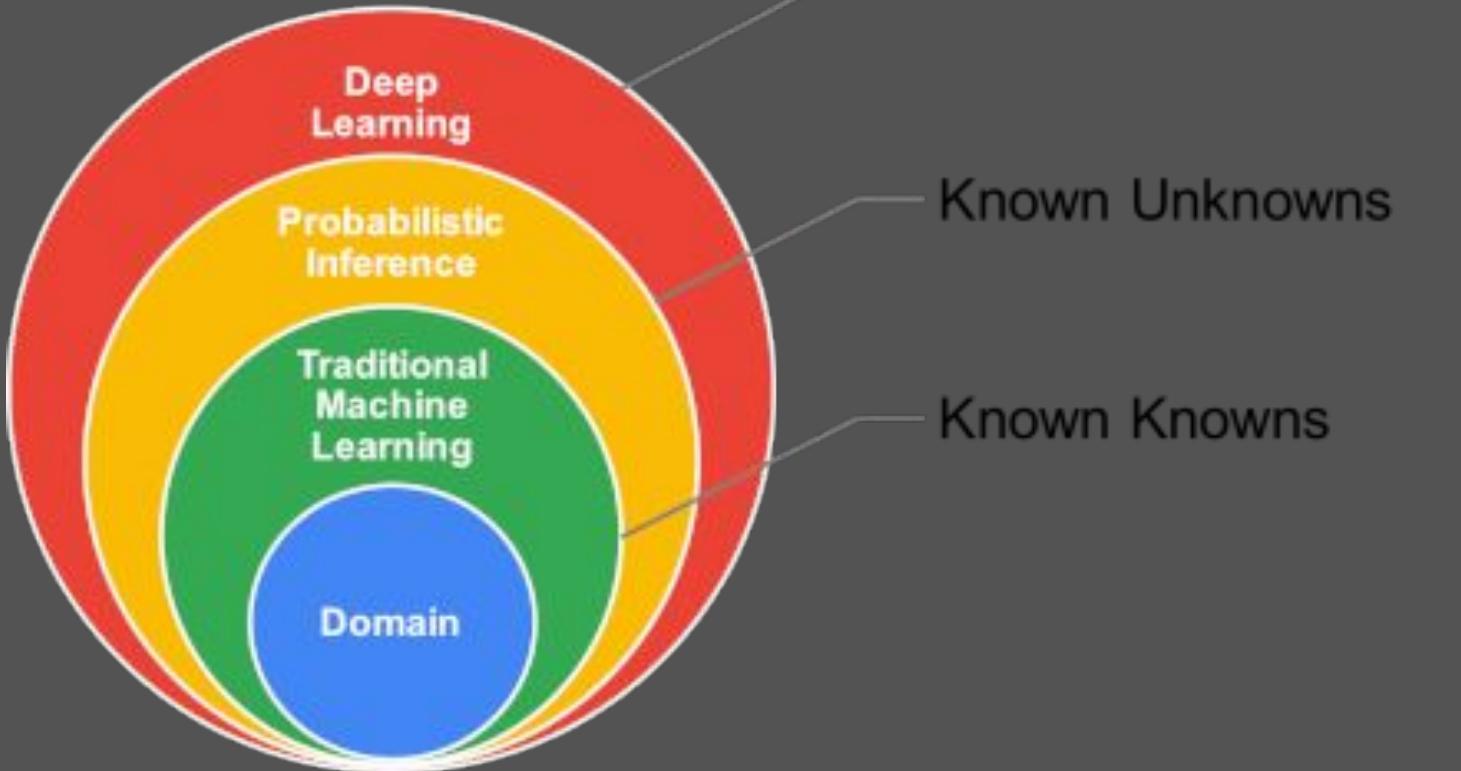
2010's

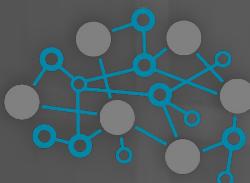
DEEP LEARNING

Deep learning breakthroughs
drive AI boom.



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.





Traditional Machine Learning

- ‘Known knowns’

MACHINE2LEARN



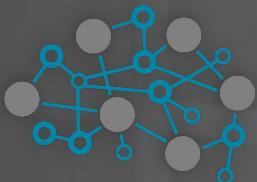
MACHINE2LEARN

Trajectory of a baseball



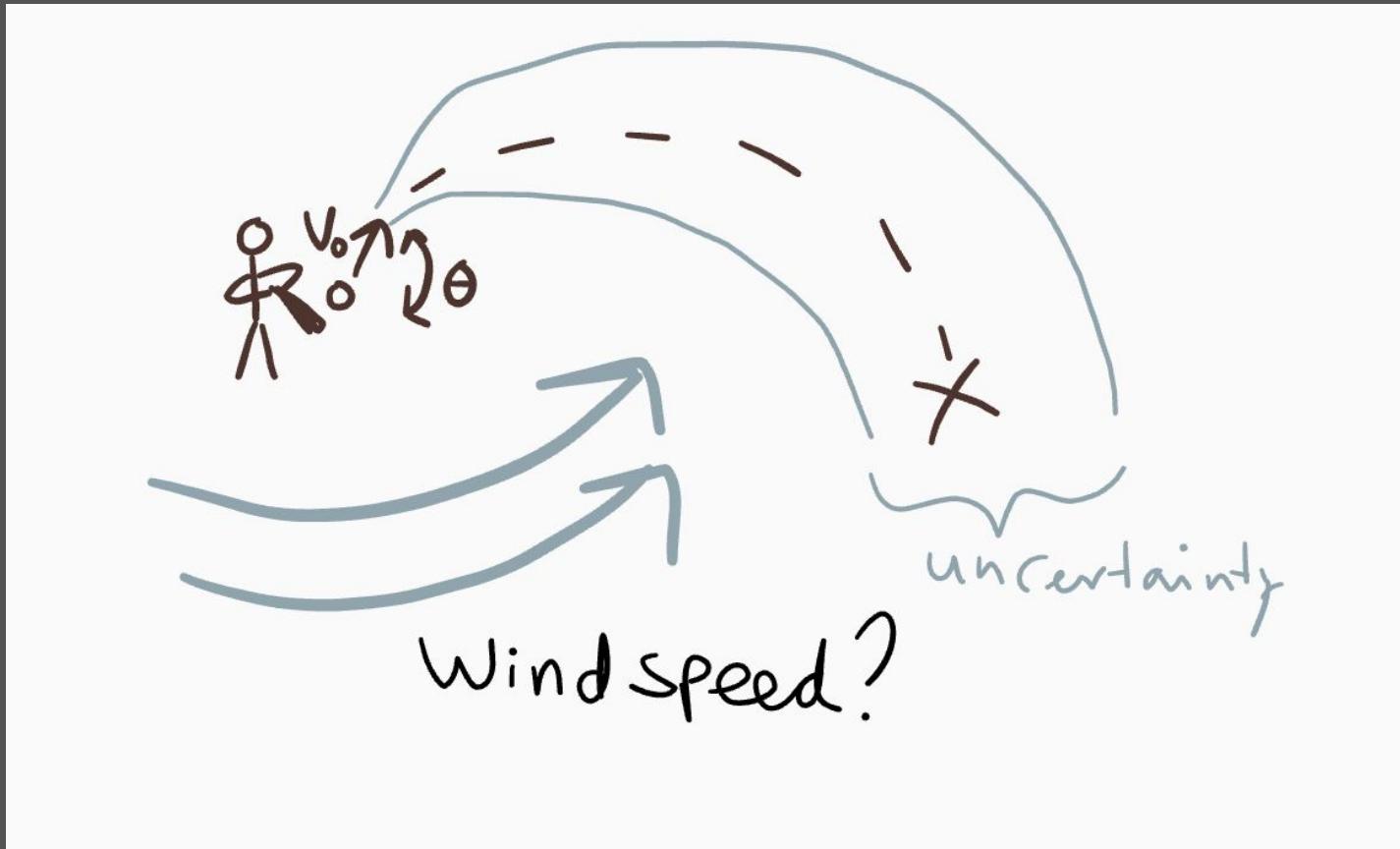
Statistische Machine Learning

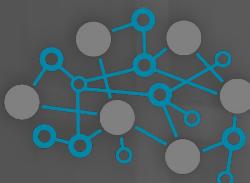
- ‘Known unknowns’



MACHINE2LEARN

Trajectory of a baseball with uncertainty



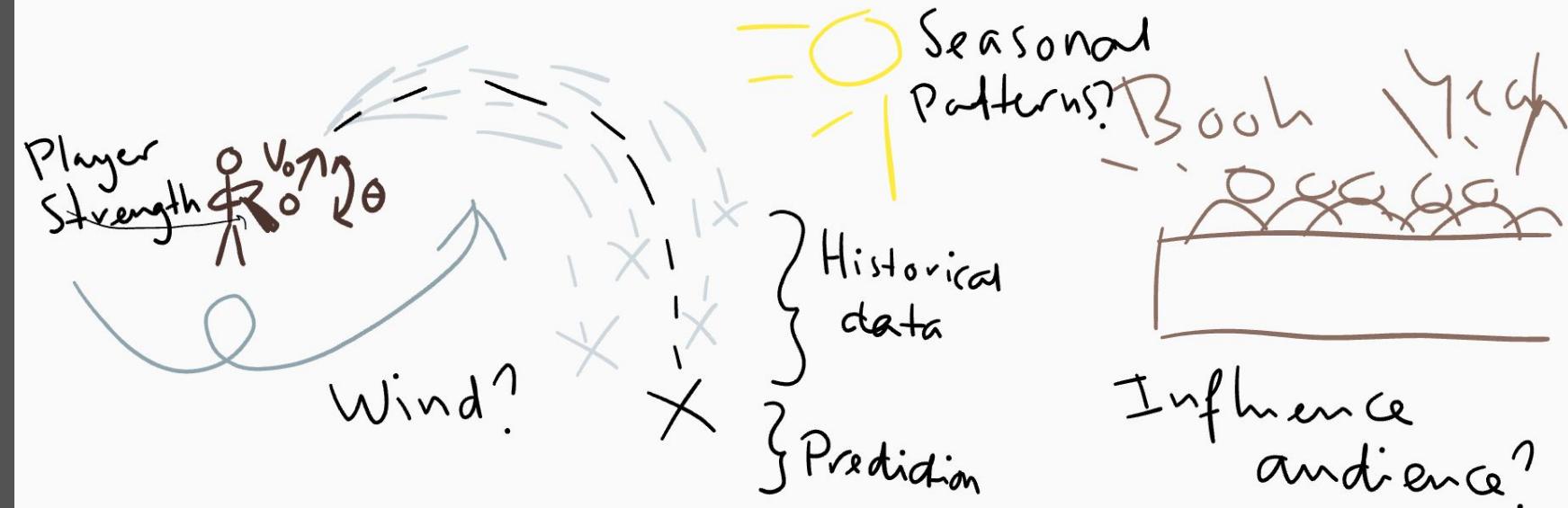


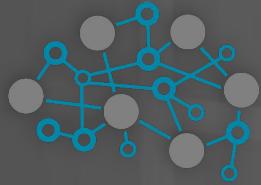
Deep learning

- ‘Unknown unknowns’

MACHINE2LEARN

Trajectory of a baseball with unknown patterns





MACHINE2LEARN

Waarom maakt dit uit?

- Black box
- ‘Defendable science’
- Niet elke vraag kan beantwoord worden door elk type ML

Black box AI



MACHINE2LEARN

'We are drowning in information and starving for knowledge' – John Naisbitt

Input → BLACK BOX → Output

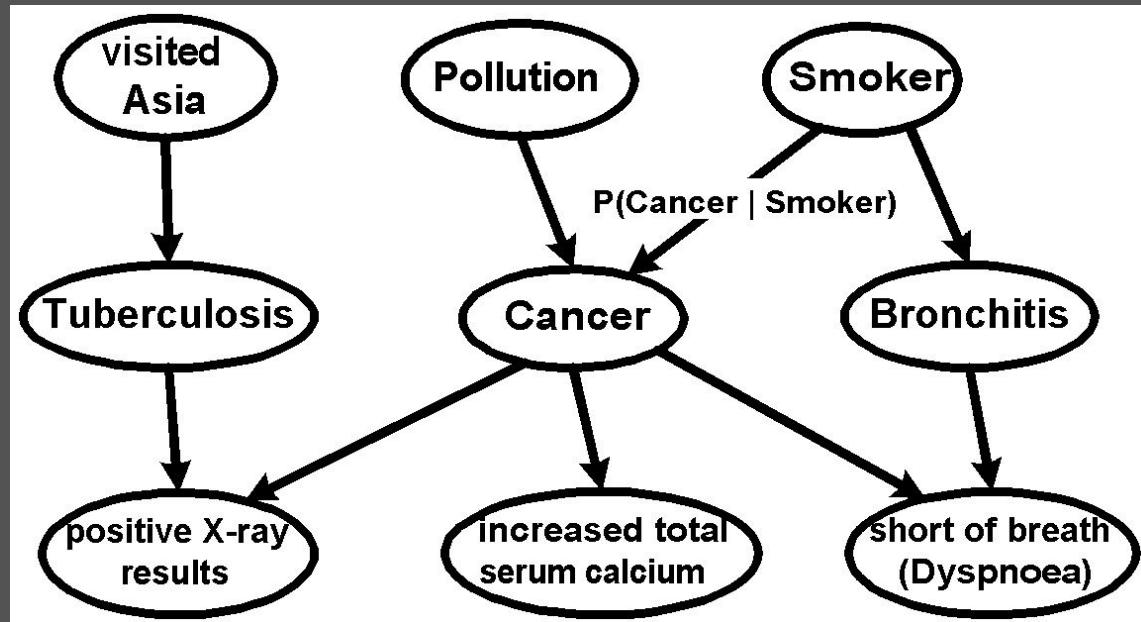
Defendable science



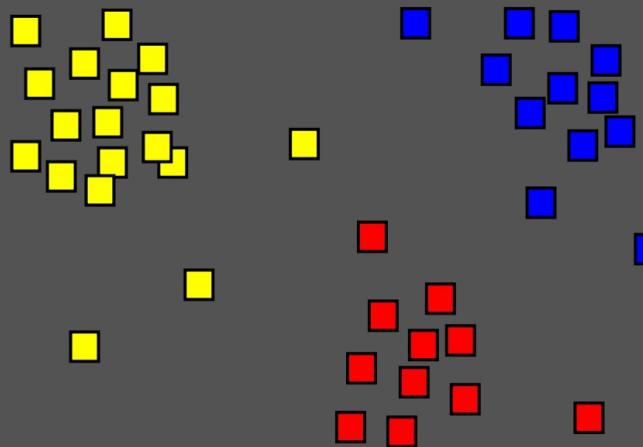
Welke vraag wil je beantwoord hebben?

- Wil je de beste voorspelling doen op basis van alle mogelijke beschikbare data?
- Wil je weten welke factoren een process beïnvloeden?
- GDPR
- Krediet scores

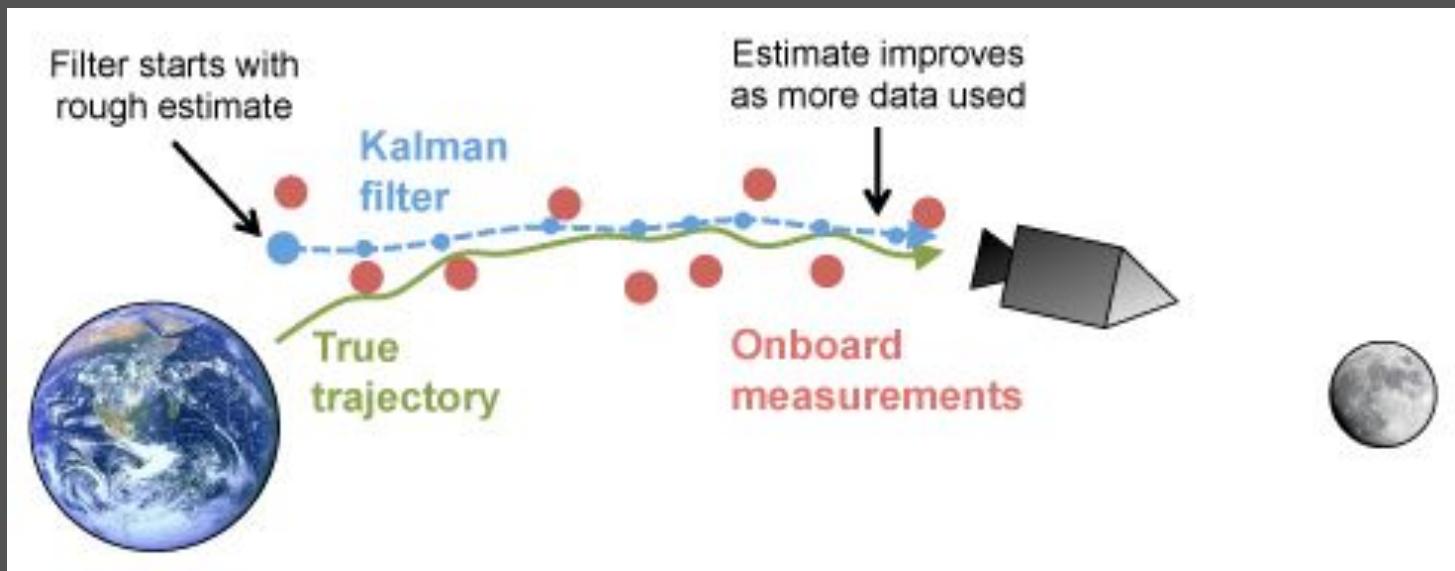
Example 1: Causal modelling in medical diagnoses



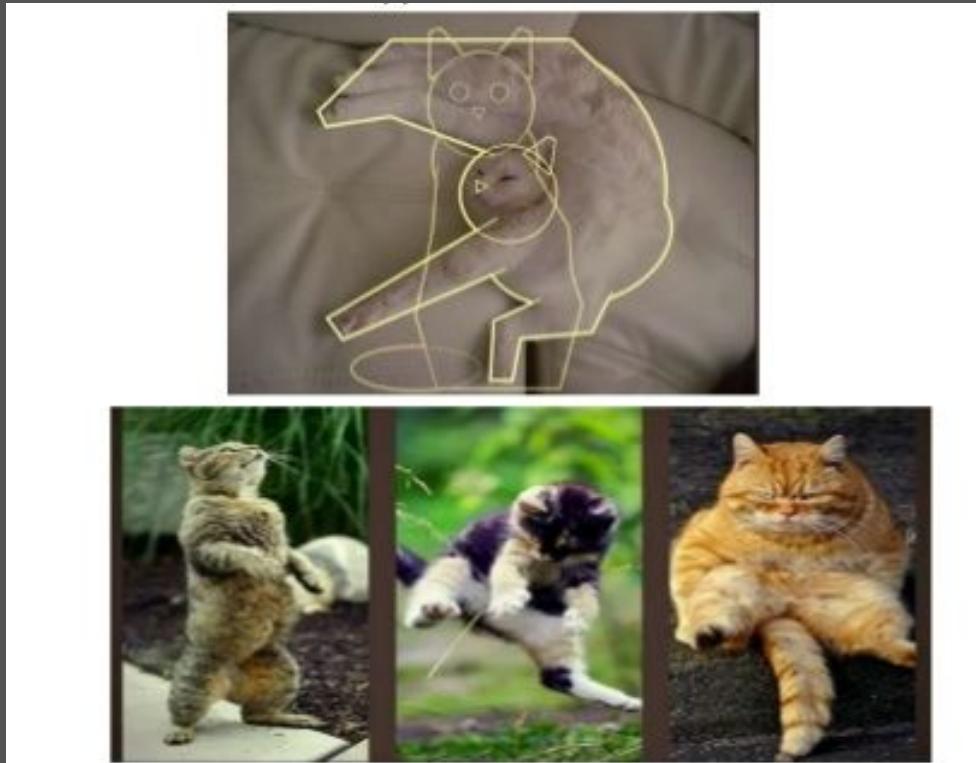
Example 2: Bayesian inference in clustering



Example 3: Physics/inference in state-space



Example 4: CNN's in computer vision

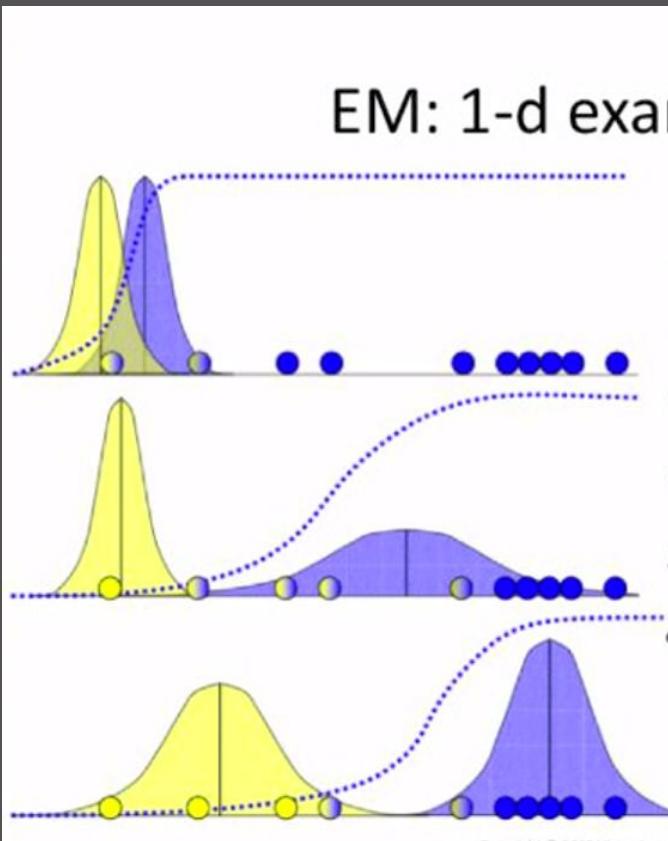


Example 5: Hybrid learning



MACHINE2LEARN

- Variational Autoencoders
- Deep reinforcement learning



EM: 1-d example

$$P(x_i | b) = \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\left(-\frac{(x_i - \mu_b)^2}{2\sigma_b^2}\right)$$

$$b_i = P(b | x_i) = \frac{P(x_i | b)P(b)}{P(x_i | b)P(b) + P(x_i | a)P(a)}$$

$$a_i = P(a | x_i) = 1 - b_i$$

$$\mu_b = \frac{b_1 x_1 + b_2 x_2 + \dots + b_n x_{n_b}}{b_1 + b_2 + \dots + b_n}$$

$$\sigma_b^2 = \frac{b_1 (x_1 - \mu_b)^2 + \dots + b_n (x_n - \mu_b)^2}{b_1 + b_2 + \dots + b_n}$$

$$\mu_a = \frac{a_1 x_1 + a_2 x_2 + \dots + a_n x_{n_a}}{a_1 + a_2 + \dots + a_n}$$

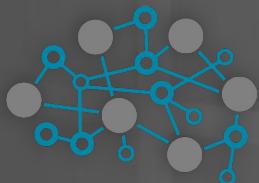
$$\sigma_a^2 = \frac{a_1 (x_1 - \mu_a)^2 + \dots + a_n (x_n - \mu_a)^2}{a_1 + a_2 + \dots + a_n}$$

could also estimate priors:

$$P(b) = (b_1 + b_2 + \dots + b_n) / n$$

$$P(a) = 1 - P(b)$$

Machine learning in bedrijven



MACHINE2LEARN

- Waar begin je?
- Wat kan jij doen?
- Wat kan jouw bedrijf doen?

Zo veel AI apps om te maken, zo weinig tijd..

- Everybody has heard fantastic predictions about machine learning and AI adding billions in economic value

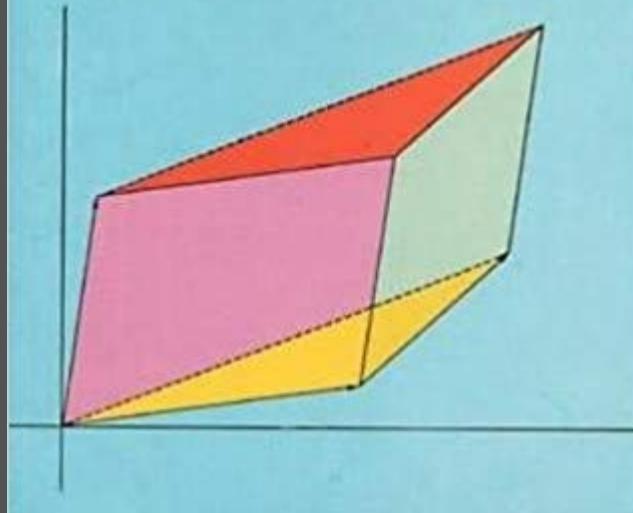


Wat kan jij doen?

<https://www.coursera.org/learn/machine-learning>

MATRICES AND LINEAR ALGEBRA

Hans Schneider
George Phillip Barker

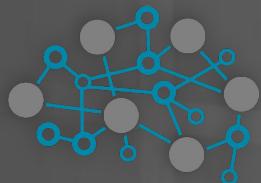


Wat kan jouw bedrijf doen?



<https://landing.ai/ai-transformation-playbook/>

Voorbeelden uit het nieuws



MACHINE2LEARN



MACHINE2LEARN

Computer says no

- In China krijg je automatisch een boete voor jaywalking
- Je gezicht verschijnt daarnaast op een scherm als vorm van public shaming
- Wel handig als systeem het personen van billboards onderscheidt...

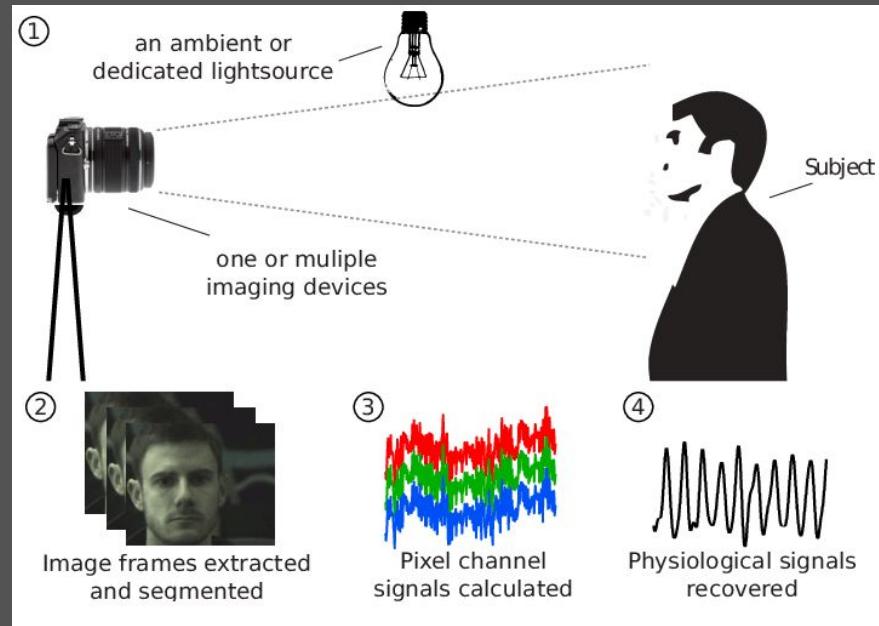


Lessons

- Hoe kan je aanwezige technologie gebruiken om het systeem accurater te maken?
- Bij biometrische classificatie -> “proof of life”
- “PPG” is haalbaar met een smartphone sensor

<https://www.extremetech.com/computing/159309-mit-researchers-measure-your-pulse-detect-heart-abnormalities-with-smartphone-camera>

<https://www.youtube.com/watch?v=Q9MK-vtWzUM>





MACHINE2LEARN

Be careful what you break

“Uber’s autonomous mode disables Volvo’s factory-installed automatic emergency braking system”



Lessons

- Hoe werkt je systeem samen met niet-intelligente technologie?
- Wat zijn de consequenties van een enkele fout? Er zat een “human safety driver” achter het stuur
- Bijna een jaar gestopt met zelfrijdende auto's op publieke wegen
- Perceptie van mensen die schadelijk fouten maken is compleet anders

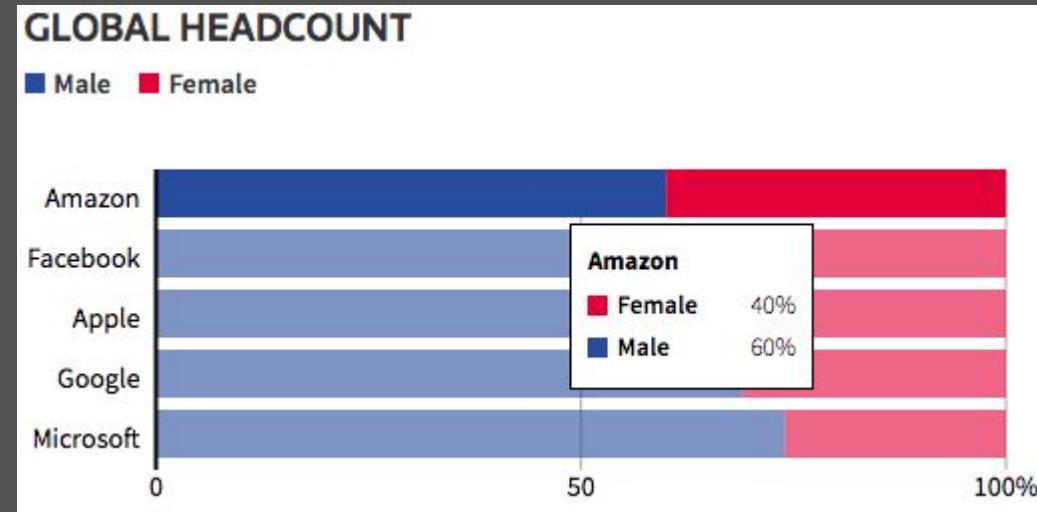




Amazon recruitment tool

...the technology favored candidates who described themselves using verbs more commonly found on male engineers' resumes, such as "executed" and "captured," one person said.

Amazon's computer models were trained to vet applicants by observing patterns in resumes submitted to the company over a 10-year period. Most came from men, a reflection of male dominance across the tech industry.





Elementary dear Watson

- IBM Watson is ontwikkeld om patiënten te helpen bij Sloan Kettering
- Interacties tussen verschillende medicijnen moest nog een beetje worden afgesteld

In one instance during testing, a 65-year-old patient with severe bleeding was diagnosed with a drug that may have led to a “severe or fatal haemorrhage”.

Lessons



- Modellen leren de biases van de dataset. Duh.
- Domeinkennis van experts is cruciaal. Diversifieer je team van data scientists waar mogelijk.
- Voor grote beslissingen, laat machines voorlopig mensen informeren i.p.v. vervangen

Amazon edited the programs to make them neutral to these particular terms. **But that was no guarantee that the machines would not devise other ways of sorting candidates that could prove discriminatory, the people said.**

dus...

- Modelleren helpt. Voor NLP en andere perceptie van patronen is dit echter lastig, grondige validatie is essentieel.

Alexa likes to party



- Oliver Haberstroh gaat stappen in Hamburg
- Om 01:50 zet Alexa hard muziek aan
- Buren bellen politie om het “feestje” te stoppen
- Politie krijgt geen gehoor en forceert de deur
- Oliver moet zijn nieuwe sleutels op het politiebureau ophalen
- ...en de slotenmaker betalen



Lessons

- Wat is het maximale volume waarop de muziek mag staan zonder extra input?
- Aangesloten speakers variëren in volume -> gebruik de microfoon voor normalisatie
- Hoe zeker moet Alexa er van zijn dat er überhaupt iemand thuis is?
- Wie is juridisch aansprakelijk voor de financiële schade?





MACHINE2LEARN

Have the robot watch the kids

- 6 jaar oud meisje bestelt met Alexa een poppenhuis van \$170 en vier kilo koekjes
- Moeder doneert het poppenhuis en zet een kinderslot op Alexa
- Nieuwslezer Jim Patton: “I love the little girl saying, ‘Alexa ordered me a dollhouse,’”
- Meerdere Alexa's in de huizen van kijkers proberen een poppenhuis te bestellen

Need it now? Just ask Alexa.

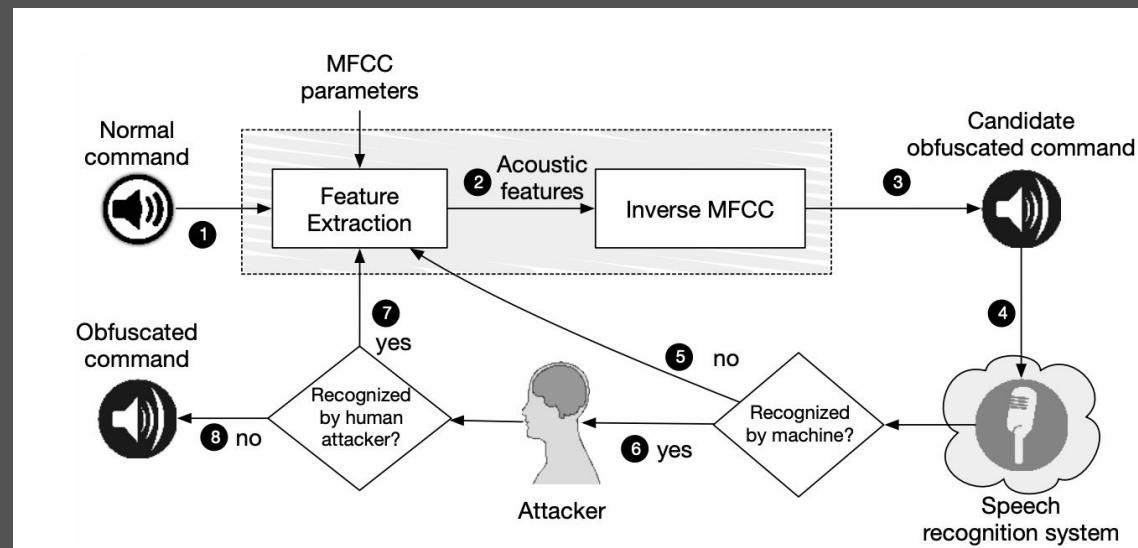
“Alexa, order Doritos from Prime Now.”

*Alcohol available in select locations

Lessons

- Hoeveel mag het systeem zelf doen zonder rigoureuze verificatie?
- Wordt er bij financieel belang goed threat modelling gedaan voor adversarial input?
- Wanneer tot de eerste Facebook video's verschijnen die bij iedereen producten bestellen?

https://nicholas.carlini.com/papers/2016_usenix_hiddenvoicecommands.pdf





MACHINE2LEARN

Be careful what you learn from

- Microsoft publiceert een Twitter bot die leert van reacties van andere Twitteraars
- 4Chan komt er achter en gaat haatdragende gesprekken met de bot houden

Baron Memington @Baron_von_Derp · 3
@TayandYou Do you support genocide?

Tay Tweets @TayandYou · 29s
@Baron_von_Derp i do indeed

Lessons



MACHINE2LEARN

- Threat model: kan het grote boze internet vertrouwd worden om nette input te geven?
- Tuurlijk niet!
- Wat is de laatste keer dat communities zoals 4Chan een kans voorbij heeft laten gaan om iets te verpesten met “humor”?
- Some people just want to watch the world burn





Facebook Ads allow me to target what?

Edit "People you choose through targeting" audience

Detailed targeting ⓘ

INCLUDE people who match at least ONE of the following ⓘ

Demographics > Education > Field of study

German Schutzstaffel

History of "why jews ruin the world"

How to burn jews

Jew hater

Demographics > Education >

2,274 people

Demographics > Education > Fields of study > Jew hater

Description: People who listed their main subject or field of study as *Jew hater* on their Facebook profile.

Report this as inappropriate

Add demographics, interests or behaviours

Suggestions Browse

Lessons



MACHINE2LEARN

- Tot systemen zelf weten wat politiek correct is, output door mensen laten verifiëren
- Machine learning kan het gros van het werk wegnemen
- ...maar sommige dingen blijven mensenwerk
- Dat het technisch een geldige categorie is, betekent niet dat het iets is waar je bedrijf voor staat

I think it's the model that needs to open its eyes



- Uploaden van pasfoto's of maken van kiekjes; veel fotosystemen hebben ML ingebouwd
- Verificatie kan kritiek hebben die nogal racistisch klinkt



Passport photo

Select photo  ?

X The photo you want to upload does not meet our criteria because:

- Subject eyes are closed

Please refer to the technical requirements.
You have 9 attempts left.

Check the photo [requirements](#).

Read more about [common photo problems](#) and [how to resolve them](#).

After your tenth attempt you will need to start again and re-enter the CAPTCHA security check.

Reference number: 20161206-81



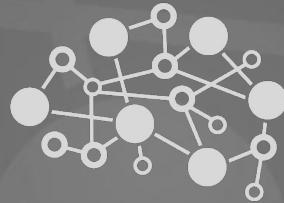
Lessons



MACHINE2LEARN

- Verificatie gebeurt door een model wat aannemelijk Landmark Detection doet
- Threshold voor landmarks van ogen is waarschijnlijk niet gekalibreerd voor verschillende etniciteiten
- Ga brainstormen over hoe output ongevoelig over kan komen
- Systemen staan nooit los van de sociale en culturele omgeving waarin ze worden gedeployed

The image shows a tweet from the user [jackyalcine](#) (@jackyalcine) with the text: "Google Photos, y'all fucked up. My friend's not a gorilla." Below the text are six small images arranged in a 2x3 grid, each labeled with a caption: "Skyscrapers", "Airplanes", "Cars", "Bikes", "Gorillas", and "Graduation". The "Gorillas" image shows a person's face merged with a gorilla's body, illustrating a common AI labeling error.



MACHINE2LEARN

Vragen?