



DATA & AI

CHURN PREDICTION HOS STARK

Ved Principal AI Specialist - Erik David Johnson

ERIK DAVID JOHNSON

- *Principal AI Specialist i Delegate*
- Humanistisk baggrund inden for sprog, kognitiv semantik og analytisk sprogfilosofi.
- Skiftet fokus mod IT med en kandidatgrad i *Software Engineering* fra ITU med specialisering i AI-Programming.
- Aktiv med artikler og foredrag om forretningsapplikationer af AI og forskning i sprogteknologi og stærk AI.
- Udpeget af regeringen til flere udvalg og kommissioner, inden for AI, Dataetik og disruption/innovation.

Se mere på www.erk.dk / [LinkedIn](#)



AI-TILTAG SOM DET OFTE SES I DAG

Fokus på det eksisterende data begrænser ambitioner og muligheder for AI-drevet innovation

Den strategi som jeg ser applikeret i markedet i dag ser ofte ud som nedenstående, hvor fokus og udgangspunkt er den tilgængelige data, som man så arbejder med i forskellige værktøjer i et forsøg på at finde nogle gode muligheder, som kunne være interessante at arbejde videre med. Fokus er her : *Hvad kan vi med vores data?*



Indsamling
af data



Eksperimenter
med data



Udvælg en
use case eller 2

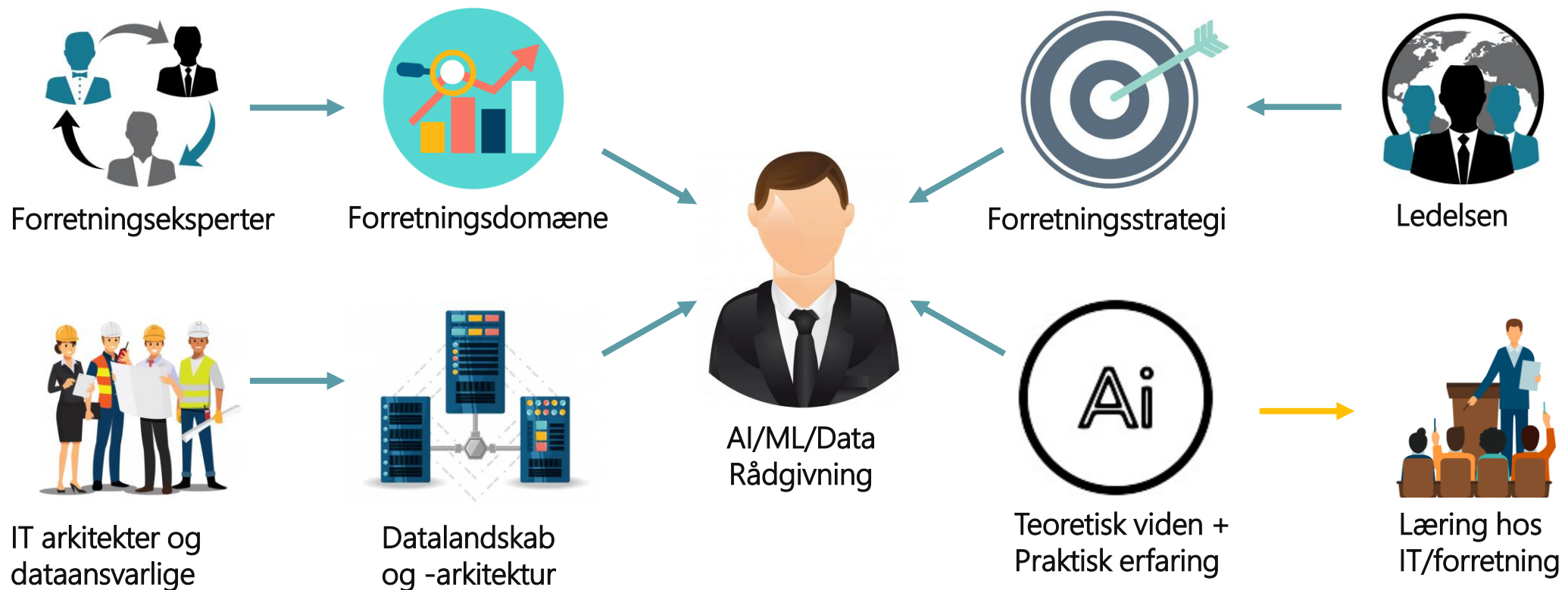


Byg og
afprøv PoC

AI EXPLORATION

Forretningsforståelse og dyb indsigt i AI leder til identifikation af forretningsmål

Jeg anbefaler en mere strategisk tilgang, hvor man ud fra en forståelse af AI, data, forretningsdomænet og jeres forretningsstrategi, starter med at finde frem til de rigtige forretningsmål. Fokus er her: *Hvad kan vi med vores forretning?*

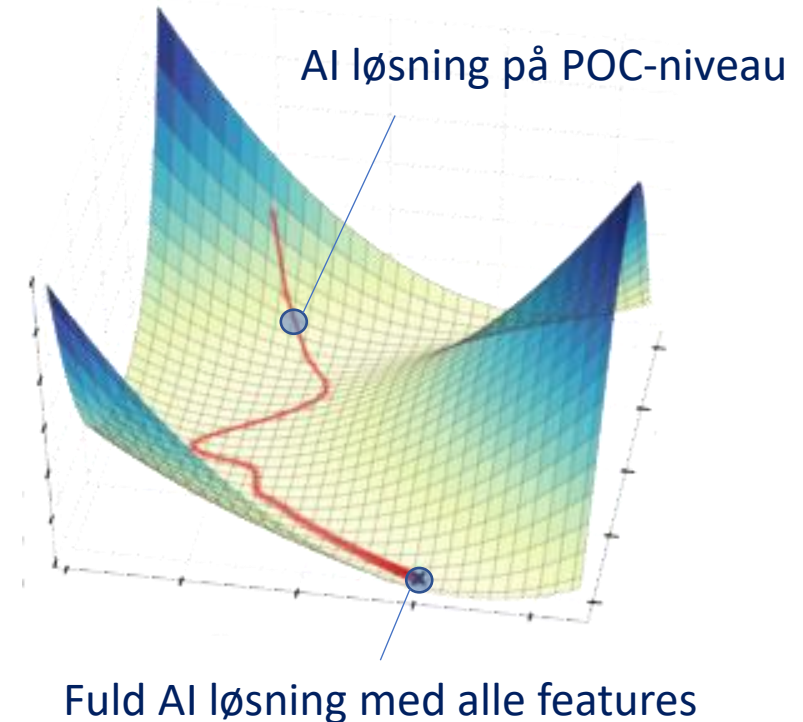


AI OG DIT (TRÆNING) DATA POTENTIALE

Når det giver mening at bruge AI, kan potentialet ikke evalueres i et DWH vha. statistik. De underliggende mønstre vil være for komplekse (ikke-lineære).

For at evaluere AI-potentialet må man bygge AI!

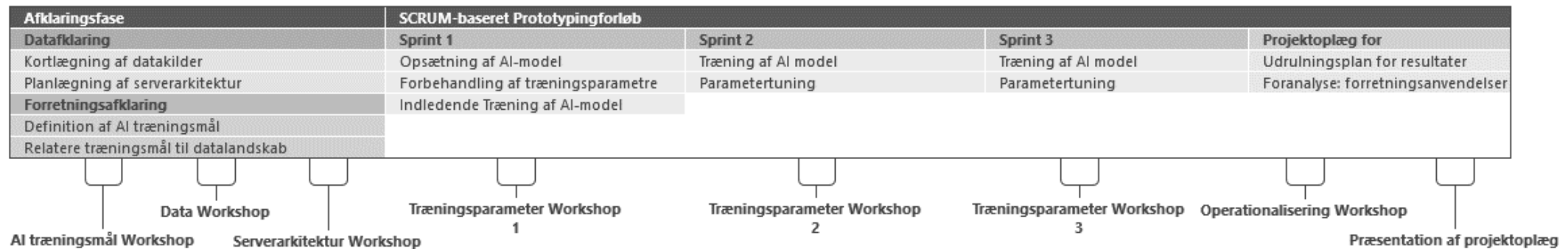
Den problemstilling løses ved at lave høj-risiko, høj-potentiale, kort **POC-forløb** inden for udvalgte forretningsmål.



ET AGILT POC PROJEKTFORMAT

Et typisk projekt på at udføre en AI POC vil være kort og agilt eksekveret

Omfanget af en POC afhænger af kompleksiteten af opgaven og det søgte forretningsmål, samt behovet for at arbejde med data inden det kan passes ind i en AI-model – også kaldet *Feature Engineering*.

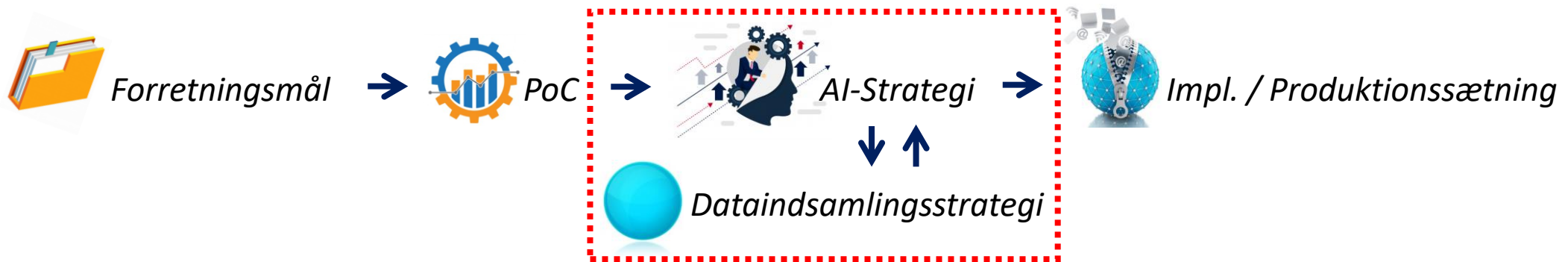


Undervejs vil det være nødvendigt at inddrage jeres medarbejdere i et begrænset omfang for bedre at forstå de forskellige arbejdsgange, og i et lidt større omfang at sparre med jeres dataansvarlige ressourcer.

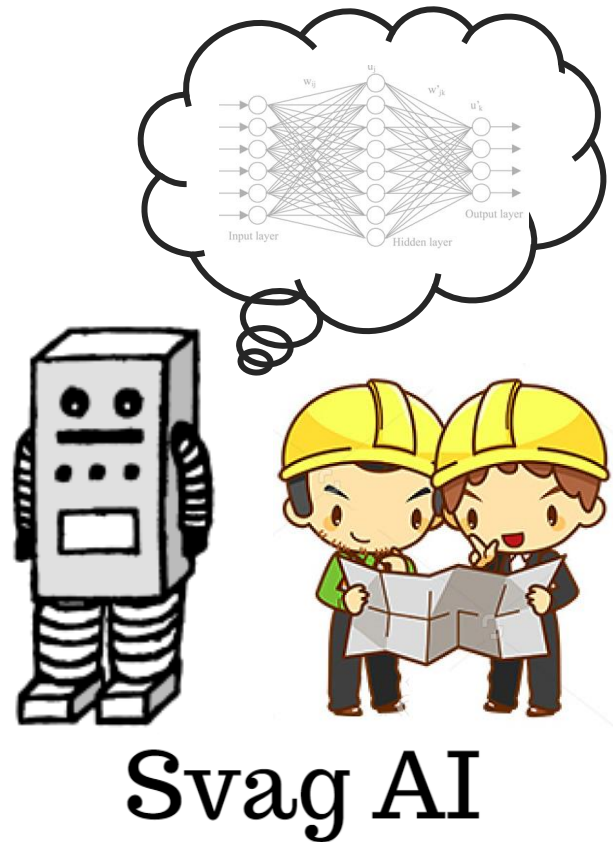
DATAINDSAMLINGSSTRATEGI – EN AI OPGAVE!

Eksekvering af dataindsamlingsstrategi er en opgave for AI-eksperter med domæneviden

En dataindsamlingsstrategi vil tage udgangspunkt i din forretning og nuværende data, og kan involvere alt fra IoT til Cloud arkitektur til Forretningsanalyse – alt med sigtekornet indstillet på det etablerede forretningsmål som AI skal hjælpe dig med at opnå.

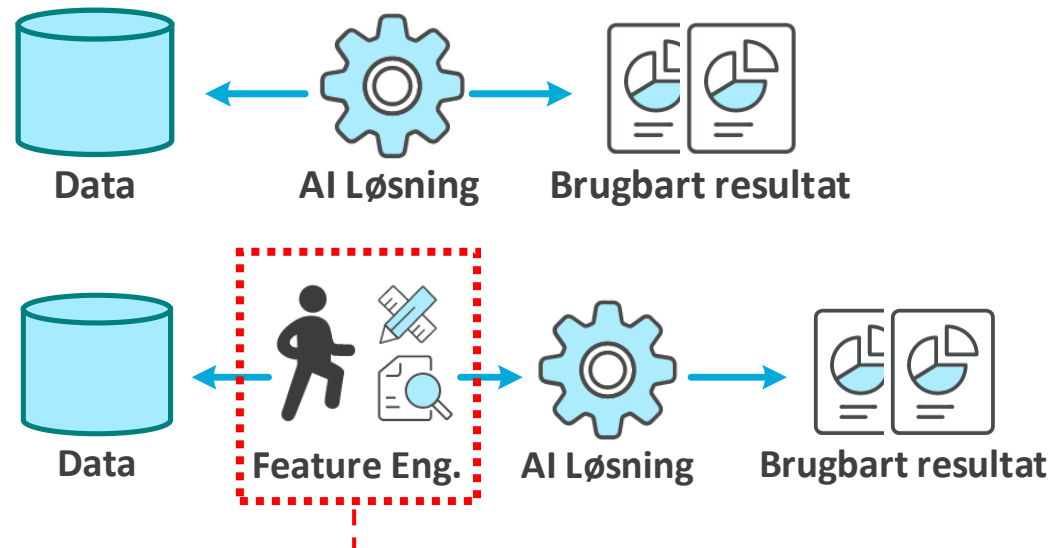


(KUNSTIG INTELLIGENS (AI) ER IKKE "PLUG'N'PLAY")



Kunstig Intelligens er et område betændt med 'hype', men den største hype er centreret omkring løfter fra konkurrerende firmaer med en AI-dimension, om at man kan sætte en AI direkte til data og få noget fornuftigt ud.

Dette er ikke tilfældet. De succesfulde, prisvindende, AI-drevne løsninger vi har lavet hos *rigtige danske virksomheder* har altid været ekstremt afhængige af kompetent 'feature-engineering' – noget som alle AI-eksperter kan genkende.



80-90% af indsatsen i et AI-projekt!



VI HAR ÅBEN
Man fredag
06.00 - 17.00
Lørdag
08.00 - 14.00
Søndag
10.00 - 14.00

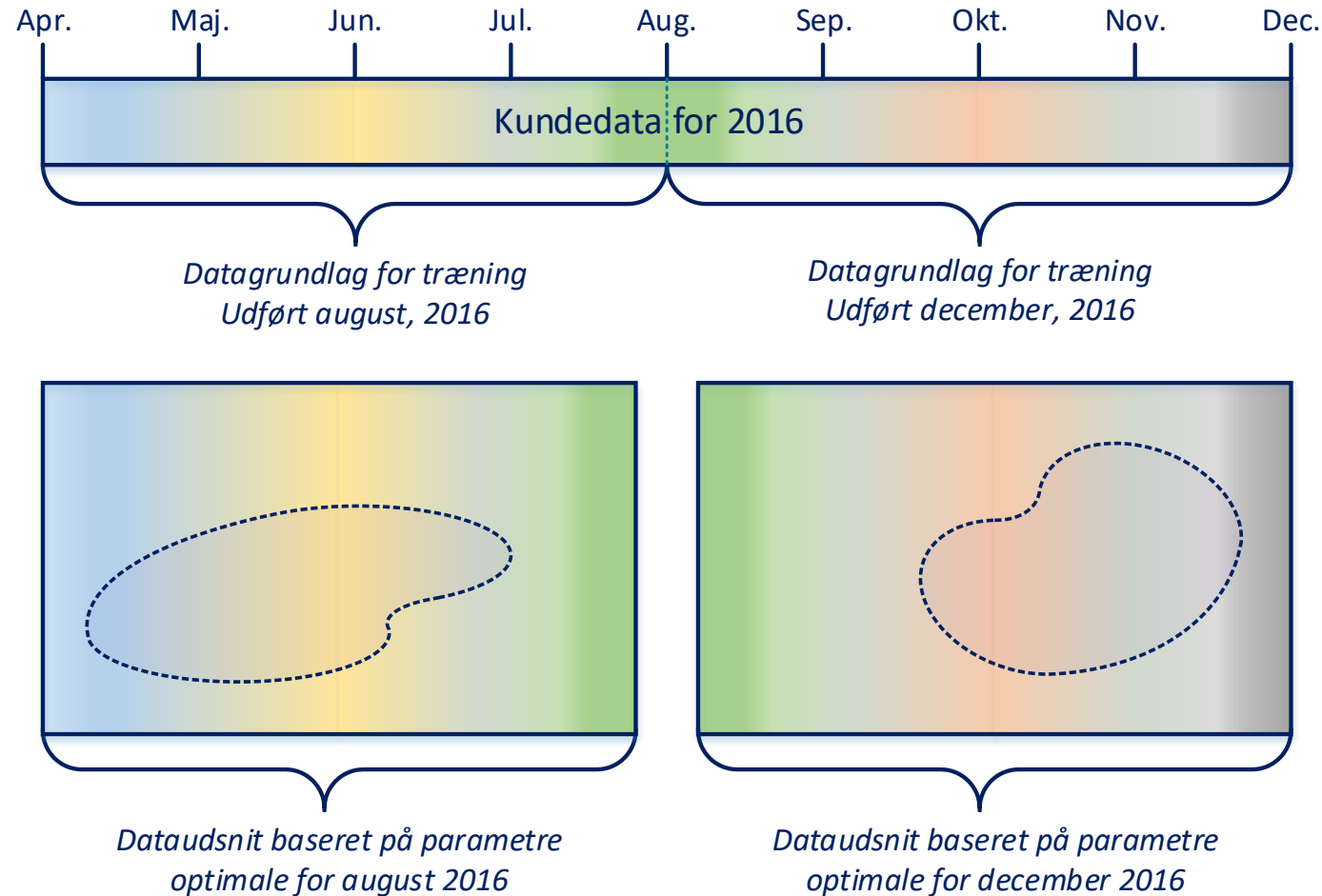
STARK

Churn Prediction



CASE - STARK

Selvlærende AI akklimerer sig et skiftende datalandskab

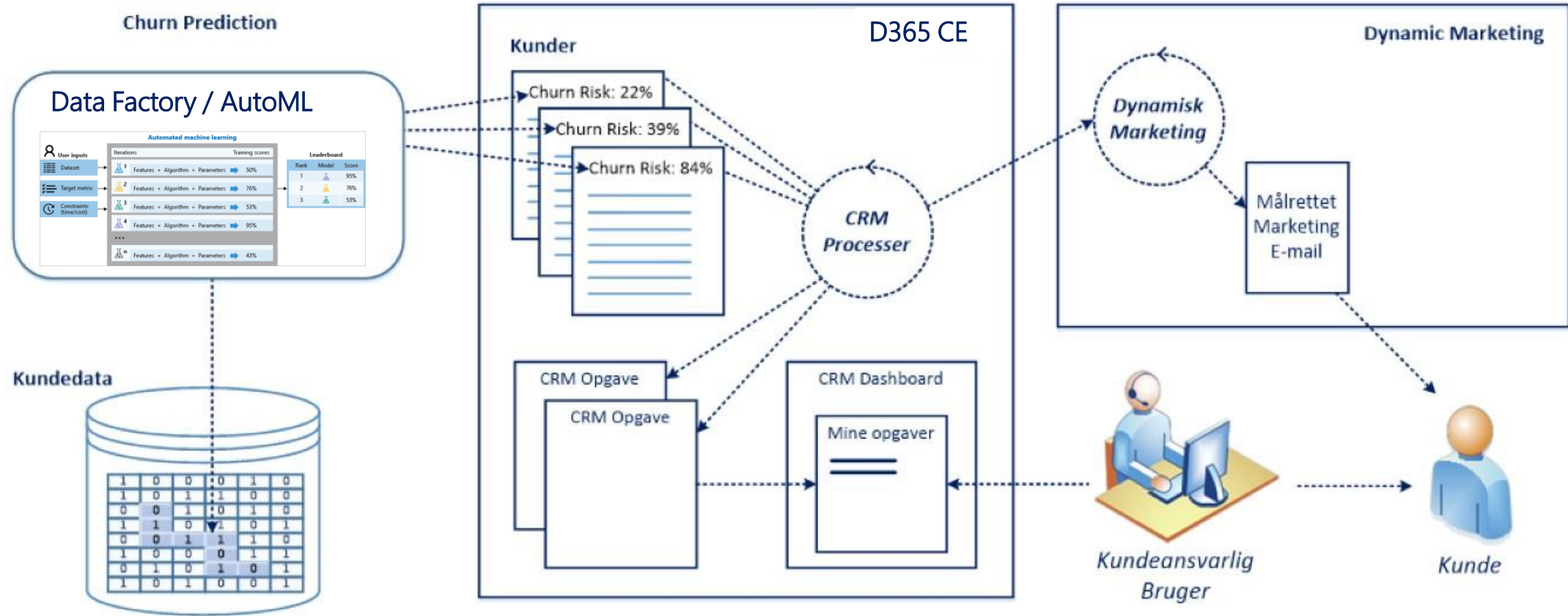


Træningen af den bagvedliggende kunstige intelligens udføres regelmæssigt. Derfor tilpasses modellen helt automatisk det aktuelle datalandskab.

Øverst: Et teoretisk eksempel på træning hvor man ser 4 måneder tilbage i tiden. Som regel vil den kunstige intelligens se mellem 6 og 24 måneder tilbage når den leder efter churn-mønstre i kundedata.

Nederst: En udvidelse til denne Churn Prediction kunne være Dynamisk Parameterudvælgelse gennem evolutionær programmering

Operationalisering i D365 CE & Marketing Automation

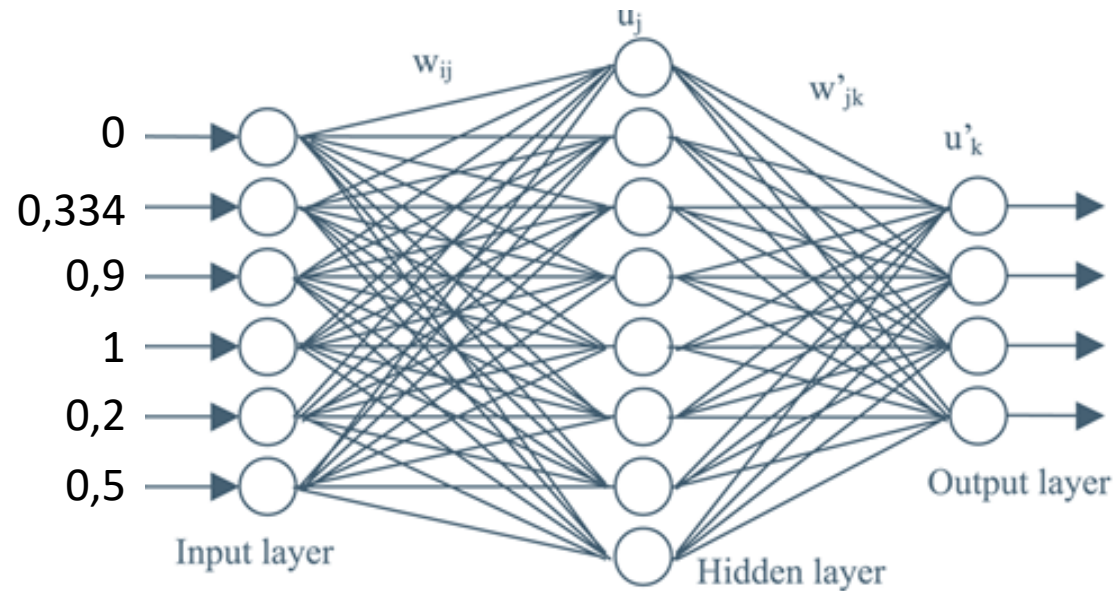


FEATURE ENGINEERING ER NØGLEN TIL SUCCES MED AI

Et eksempel fra AO Johansen Casen på Churn Prediction

Udfordring

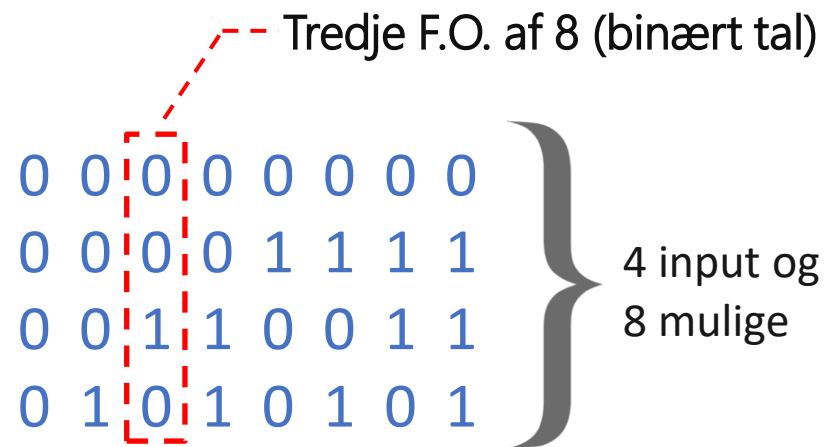
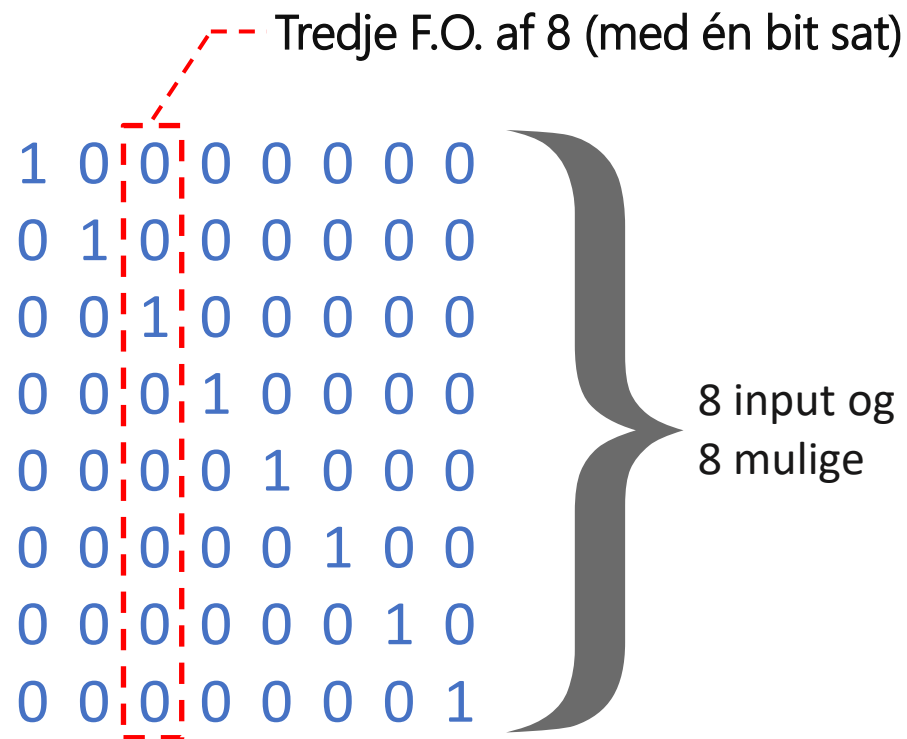
Der er 37 forskellige forretningsområder som skal ind i en AI-model som kun kan forstå til mellem 0 og 1 – i.e. et *kunstigt neuralt netværk*.



FEATURE ENGINEERING

Den akademiske tilgang

Den akademiske tilgang forlader sig på modellen, og repræsenterer de 8 forretningsområder (F.O.) som *binære* repræsentationer:

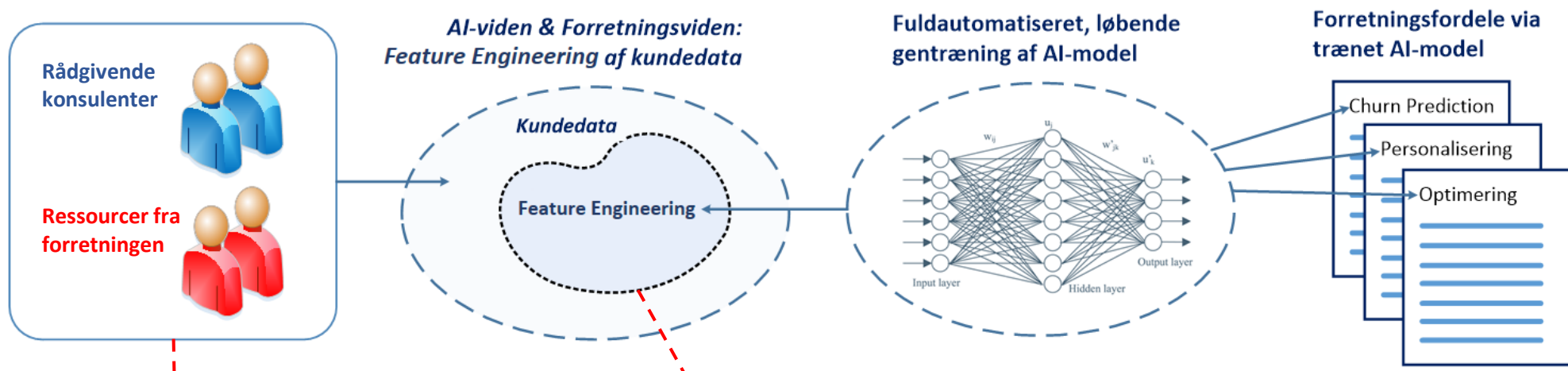


Fælles for dem er at de tager alle 8 forretningsområder ind i modellen

FEATURE ENGINEERING

Den praktiske løsning

Den praktiske tilgang til Feature Engineering sikrer at domæneeksperternes viden tages ind i den måde data repræsenteres på. De fremtidige, erfarne brugere er som regel også domæneeksperter!



Domæneeksperterne interviewes om deres viden og intuitioner, og man når frem til at de 8 forretningsområder har 3 segmenter der relaterer sig ensartet til Churn

$\left. \begin{matrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{matrix} \right\}$ Kun 2 input og 3 mulige

Dette udgør en kraftig og nødvendig reduktion af problemdomænet!

FEATURE ENGINEERING

Teorien fortæller os at vi skal omdanne data til noget en AI model kan arbejde med – for eks. at omdanne data til tal mellem 0 og 1, som et kunstigt neuralt netværk kan tage som input.

Praktisk erfaring peger på at dette ikke er nok, og at Feature Engineering skal inkludere domæne-eksperternes viden for at der kan laves en stærk, anvendelig løsning uden for laboratoriet.



*Du skal ikke træne en AI model til
at lære noget du allerede ved!*

*Den skal kun trænes til at fortælle
dig det du ikke ved.*

STARK FEATURE ENGINEERING

Feature engineering

- Feature typer:
 - Statisk – Kunde region
 - Historisk værdi - Betalingsbetingelser, kundestatus
 - Dage siden – Sidste tilbud, kreditnota
 - Aggregateret over interval - Har handlet online sidste år 360 dage, antal aktiviteter på kunde i CRM
 - Aggregateret over multiple intervaller – omsætning sidste fire 90 dages intervaller

DATAETIK

Etik er ofte forbrænder for lovgivningen, og Dataetik i en dansk kontekst kan potentielt blive en konkurrenceparameter. Dette er noget regeringen i høj grad stiler efter. **Bør du også det?**

SELVBESTEMMELSE

Mennesket skal bevare mest mulig kontrol over egne data.

LIGHED OG RETFÆRDIGHED

Teknologien må ikke diskriminere.

DIVERSITET

Flest mulige faggrupper med forskellig køn, alder, etnicitet osv. skal involveres i udviklingen af teknologiske løsninger.

ANSVARLIGHED

Alle led skal være ansvarlige for konsekvenserne af deres teknologiske løsninger.

VÆRDIGHED

Menneskets iboende værdighed skal vægtes over profit.

„Ethical behavior is about doing the right thing even when no one is watching – even when doing the wrong thing is legal.“

— SACHIN LODHA, TATA CONSULTANCY SERVICES, INDIEN

DATEETIK OG AI

- + Dataetisk Mærkeordning
- + Dataetisk Råd



Medlemskab

Medlemskab



Stakeholder

Leverandør



DATAETIK OG AI I PRAKSIS

Der er mange komplekse aspekter af dataetik og kunstig intelligens, men fra et *praktisk* perspektiv kan de fordeles på 3 hovedområder:



Gennemsigtighed



Bias



Personhenførbare data

Hvert af disse områder præsenterer en særlig etisk udfordring når det kommer til udviklingen af AI- og datadrevne løsninger.

Anbefaling nr. 3

**Dynamisk
værktøjskasse
til dataetik**



PERSONRELATERBAR DATA

Id	Duration(hrs)	# Packets	#NetFlows	Size	Bot	#Bots
1	6.15	71,971,482	2,824,637	52GB	Neris	1
2	4.21	71,851,300	1,808,123	60GB	Neris	1
3	66.85	167,730,395	4,710,639	121GB	Rbot	1
4	4.21	62,089,135	1,121,077	53GB	Rbot	1
5	11.63	4,481,167	129,833	37.6GB	Virus	1
6	2.18	38,764,357	558,920	30GB	Menti	1
7	0.38	7,467,139	114,078	5.8GB	Sogou	1
8	19.5	155,207,799	2,954,231	123GB	Murlo	1
9	5.18	115,415,321	2,753,885	94GB	Neris	10
10	4.75	90,389,782	1,309,792	73GB	Rbot	10
11	0.26	6,337,202	107,252	5.2GB	Rbot	3
12	1.21	13,212,268	325,472	8.3GB	NSISay	3
13	16.36	50,888,256	1,925,150	34GB	Virus	1

OPEN DATA DK



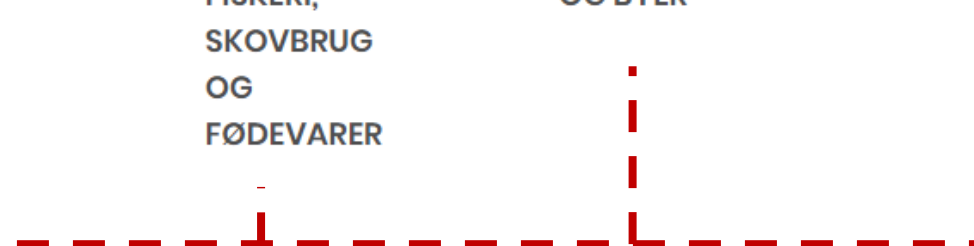
LANDBRUG,
FISKERI,
SKOVBRUG
OG
FØDEVARER



REGIONER
OG BYER

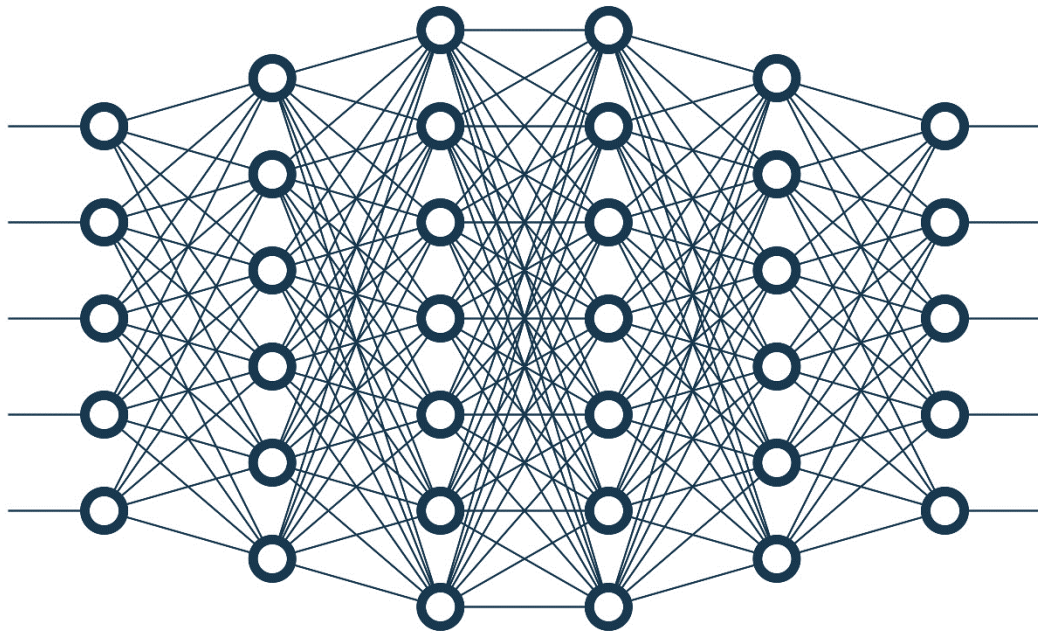


MILJØ



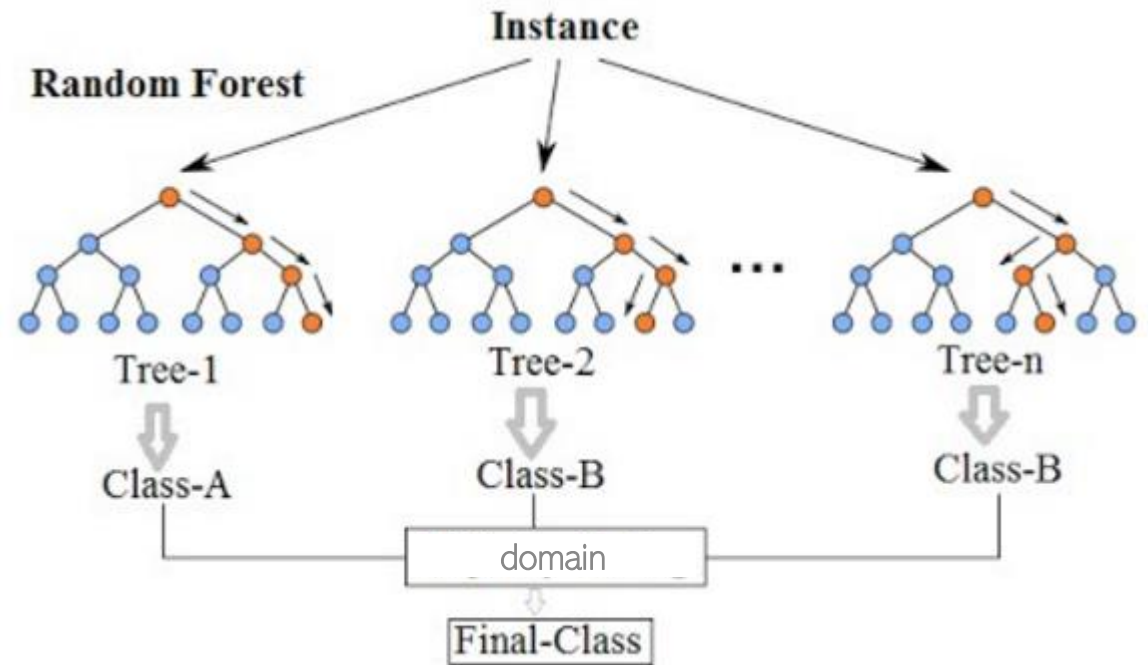
GENNEMSIGTIGHED

Deep Learning



-Hvilke features er vigtigst?

Random Forest



BIAS

Kort sagt er alle data og data-drevne løsninger noget der udspringer af menneskelig adfærd, og vores adfærd er styret af vores **foreløbige viden og antagelser om verden**. Vores antagelser er altså til stede i alle dele af en AI- og datadrevet løsning, med undtagelse af selve træningen af AI modellen, som finder mønstre i data.

BIAS



En løsning er at de folk der arbejder med Data og Datadrevne løsninger såsom AI, er i stand til at identificere de problemstillinger som kendetegner bias for data og løsninger inden for det givne domæne, og at **Feature Engineering sikrer at der ikke introduceres eller videreføres Bias**.