

# Forudsigelse af forurening på Regionsniveau

Ida Dyrhund Damgaard, Region Midtjylland

REGION  
SJÆLLAND

*- vi er til for dig*



Region Nordjylland



Region  
Hovedstaden



Region  
Syddanmark

# Hvor startede vi?



# Projekt organisering



Tvær regional projektgruppe:

- it kyndige
- dataansvarlige
- sagsbehandlere

ITK – tværregional finansiering  
GovTech Midtjylland

# Use cases

Hvordan kan data og machine learning skabe værdi for bedre beslutninger og prioritering af indsatser?



I samarbejde med



Fra data  
til bedre  
beslutninger



Store mængder  
af historiske data



Machine learning  
modeller



Indsigt og  
forudsigelser



Bedre prioritering  
og målrettede indsatser



# User cases

Hvordan kan data og machine learning skabe værdi for bedre beslutninger og prioritering af indsatser?



I samarbejde med



## Fra data til bedre beslutninger



Store mængder af historiske data



Machine learning modeller



Indsigt og forudsigelser

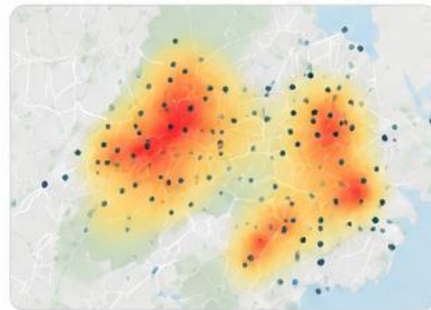


Bedre prioritering og målrettede indsatser



### 1. Forureningsforudsigelse

Kan vi forudsige, hvor der er forurenet og dermed prioritere lokaliteter bedre?

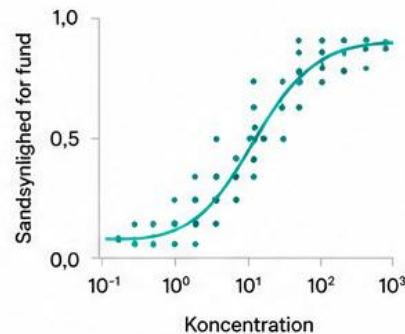


Sandsynlighed for forurening



### 2. Sandsynlighed for fund

Hvor sandsynligt er det et fund i en prøve af en given parameter gruppe?



Eksempel på output

72%

Sandsynlighed for fund



### 3. PFAS-risiko

Kan vi ved hjælp af gamle prøver sige om der vil være PFAS til stede på lokaliteten?



Historiske prøver



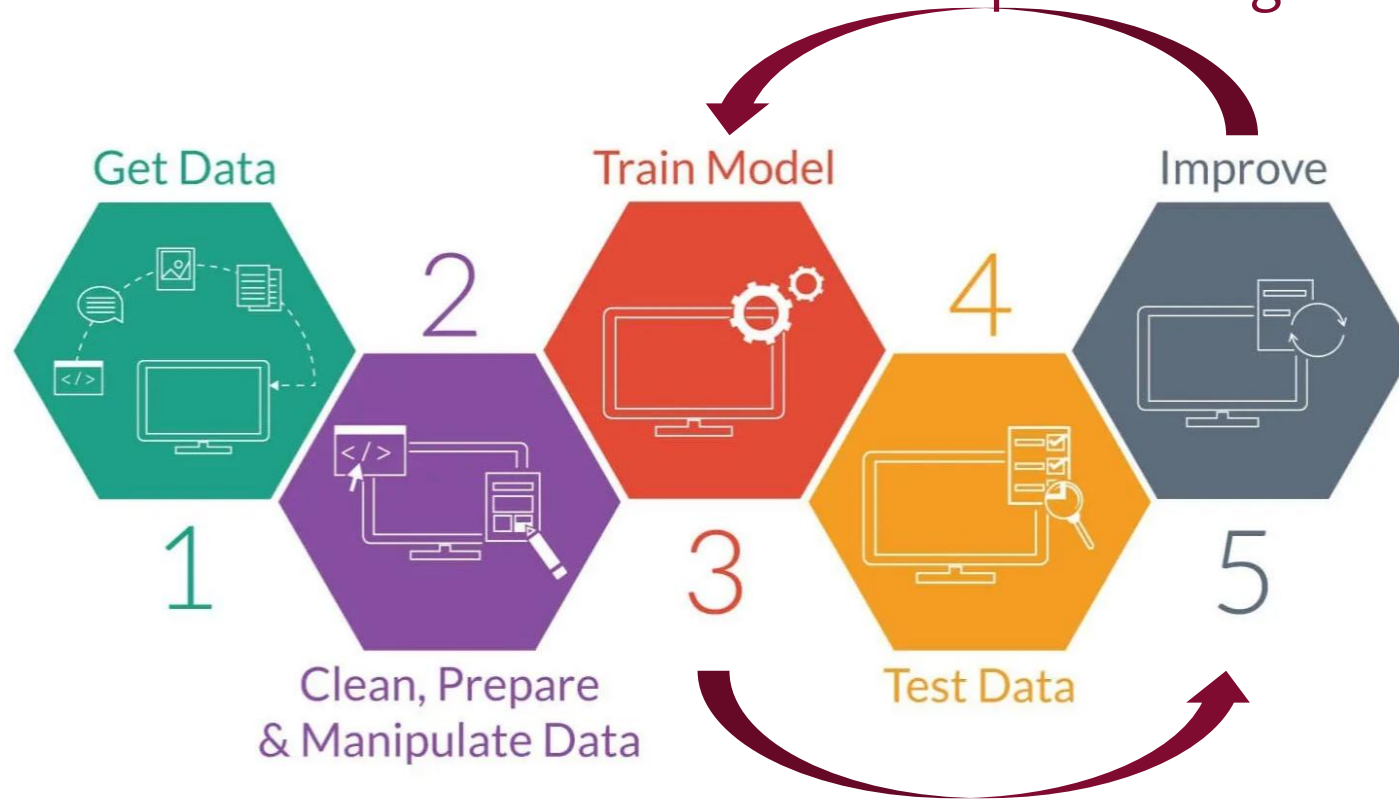
Sandsynlighed for PFAS

65%

Risiko for PFAS til stede

# Machine Learning

Find mønstre i store data sæt. Både antal datapunkter og observationer.



Både automatisk og manuel

Fra data  
til bedre  
beslutninger



Store mængder  
af historiske data



Machine learning  
modeller



Indsigt og  
forudsigelser



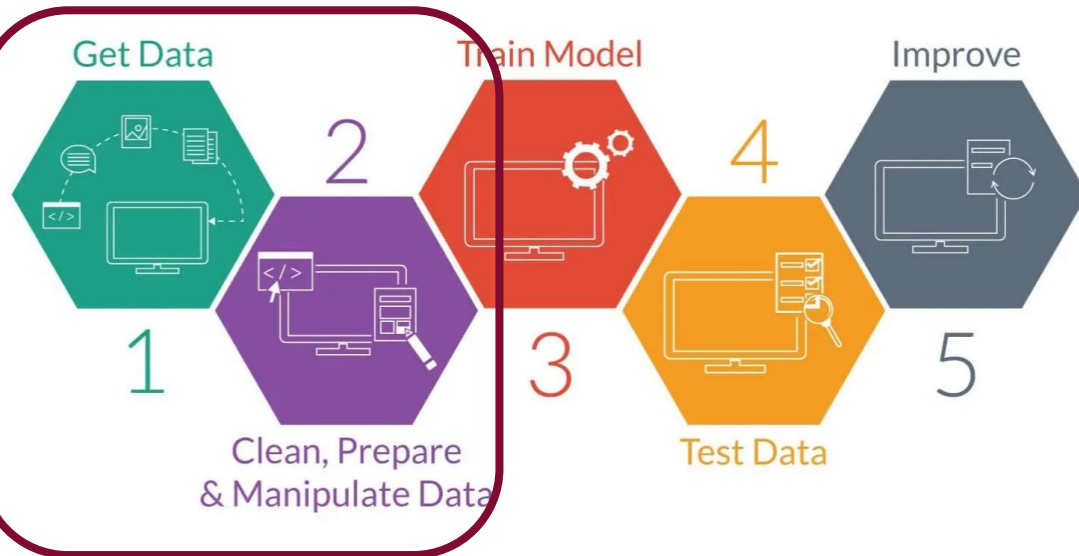
Bedre prioritering  
og målrettede indsatser



# Konceptuelt overblik over dataindsamling

Vi kunne også se på:

- Omsætning
- Ansatte
- Selskabstype



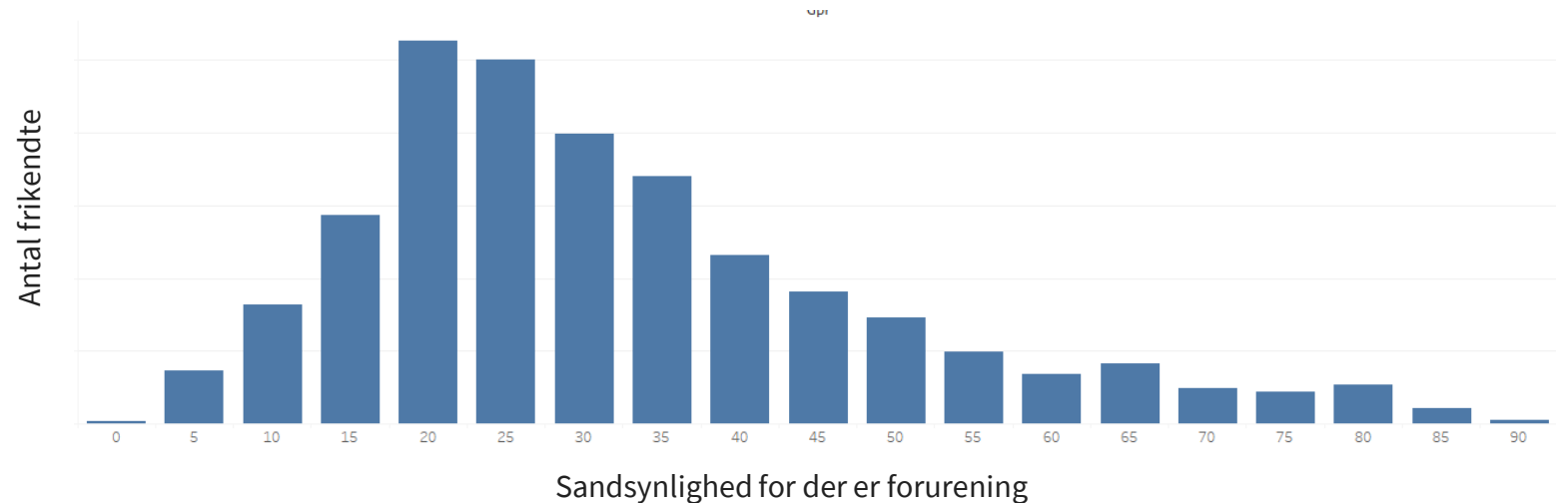
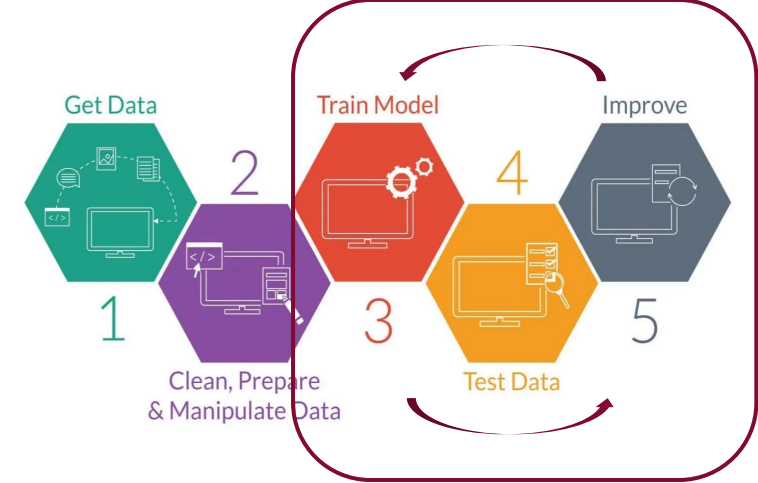
- Jupiter (analysedata)
  - DK Jord (aktiviteter, brancher, drift osv.)
  - Grundrisk (infiltrationsrater, jordbund)
- betydeligt hvordan data er registreret

# Validering af modellen mod data vi kender resultatet på

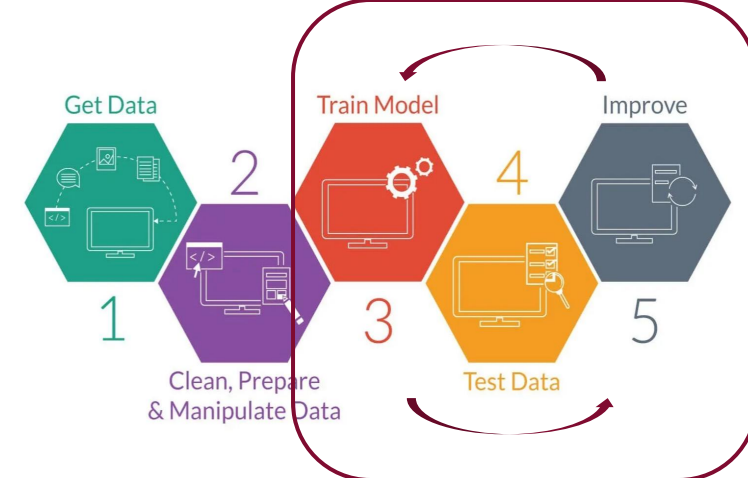
Hvor tit tager algoritmen fejl?

- dvs. siger en ren grund er forurenet

Så skal vi se på, hvor ofte modellen vil sige en "grund vi ved er frikendt" er forurenet (repræsentativ datamængde)



# Validering af modellen mod data vi kender resultatet på



Hvor tit tager algoritmen fejl?

- dvs. siger en ren grund er forurenset

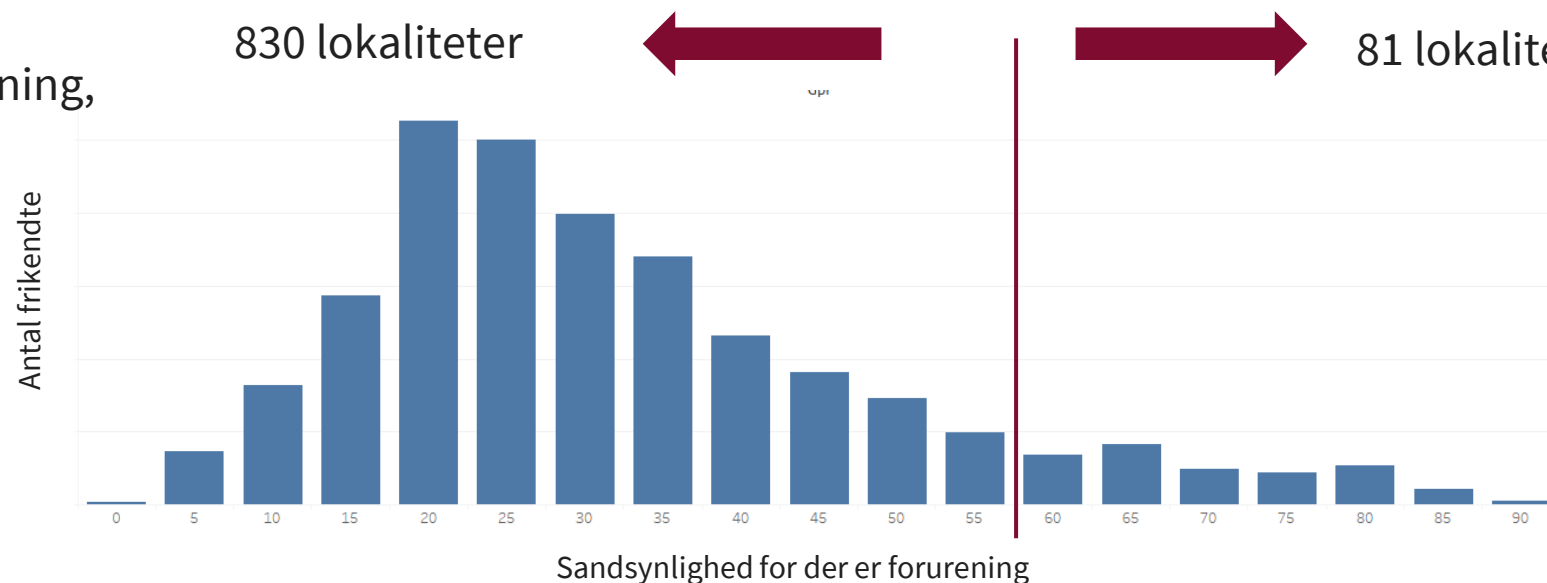
Så skal vi se på, hvor ofte modellen vil sige en "grund vi ved er frikendt" er forurenset (repræsentativ datamængde)

Omkring 60% sandsynlighed for forurening, er der ikke længere mange

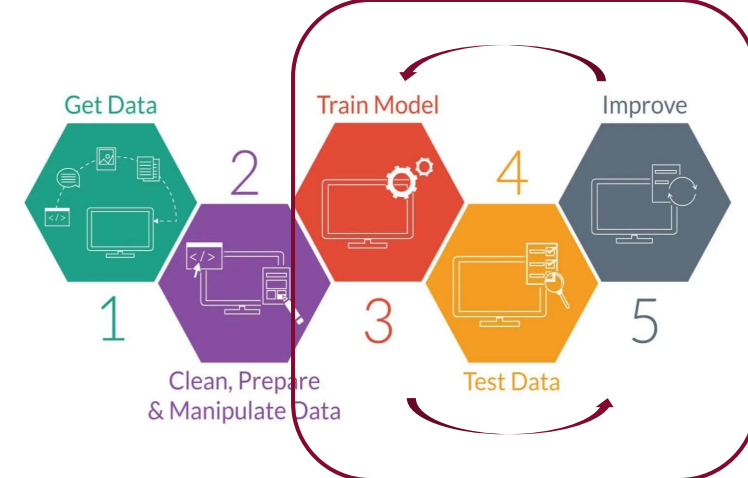
90% nøjagtighed

830 lokaliteter

81 lokaliteter



# Validering af model mod data vi kender resultatet på



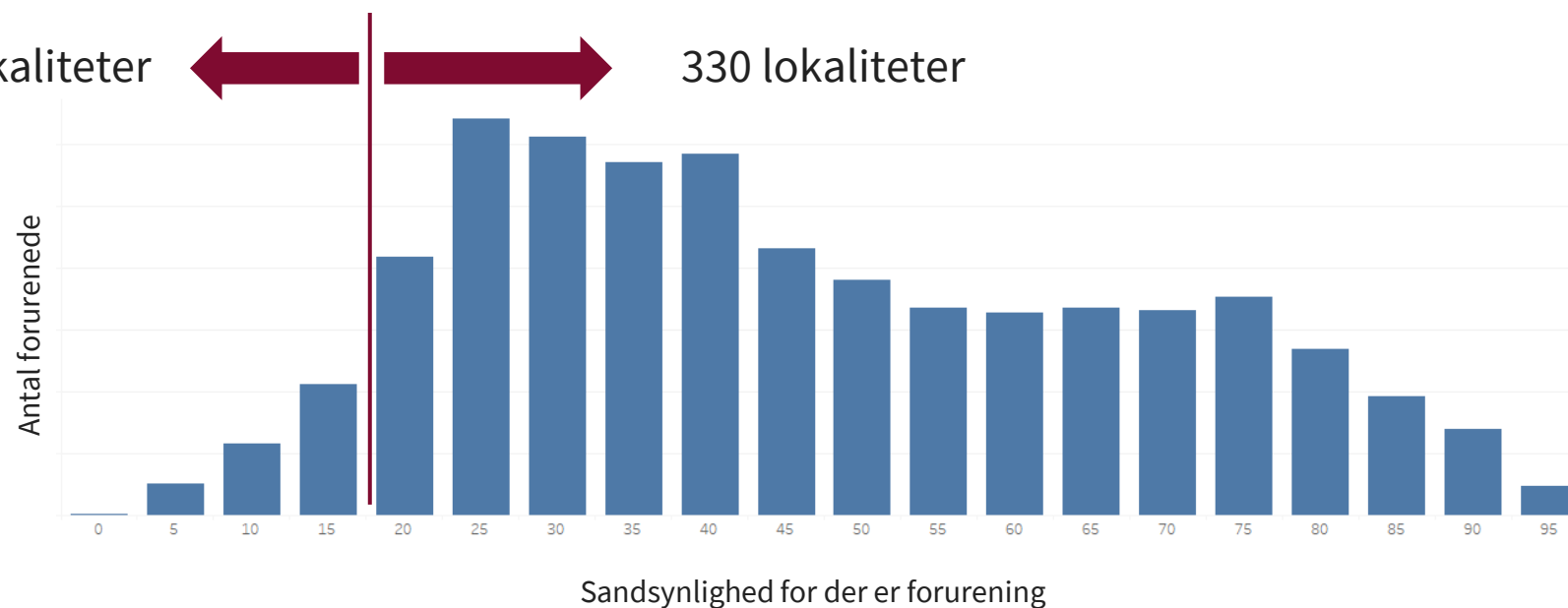
Hvor ofte ER der forurenset selv om modellen siger der IKKE er forurenset?

Vi kigger på grunde vi ved der er forurenede og kører modellen mod det

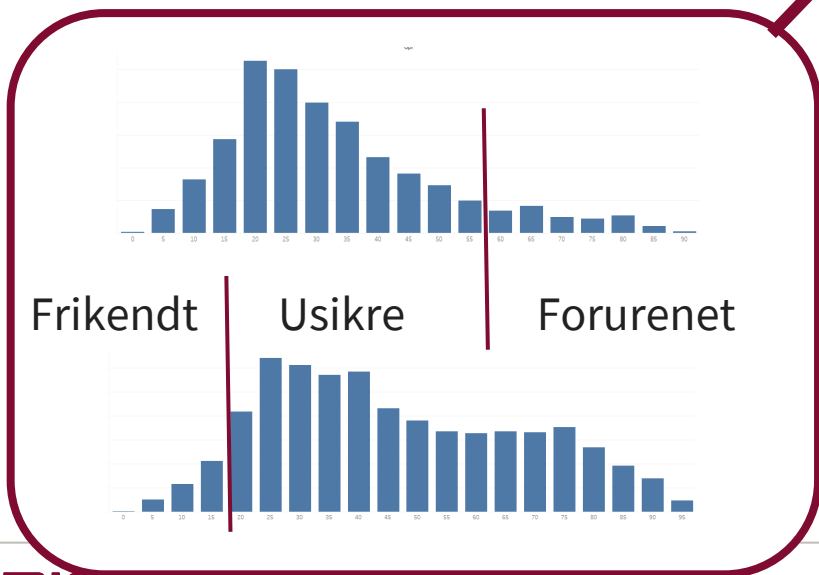
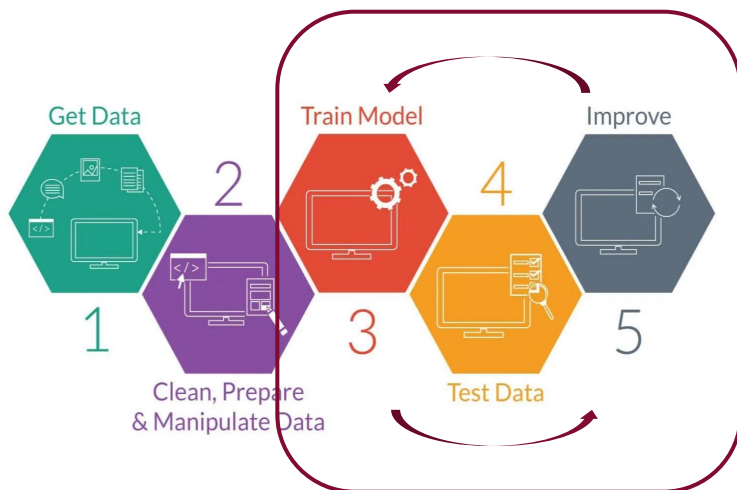
95% nøjagtighed

20 lokaliteter

330 lokaliteter



# Hvad kigger vi på?



## Overfladebehandling af metal

Risiko for forurening  
**45%**

Vi er usikre på om der findes noget på lokationen.

Vis udslagsgivende parametre

Vigtig info: Det er sammenhængen mellem de forskellige faktorer ved netop denne lokalitet, der definerer, hvor meget hver enkelt faktor betyder i negativ eller positiv retning. De enkelte faktorer vil have anden vægtning i andre sager selvom værdien kan være det samme. Eks. Branchekoden losseplads kan betyde lidt i en sag, mens det betyder meget i en anden sag.

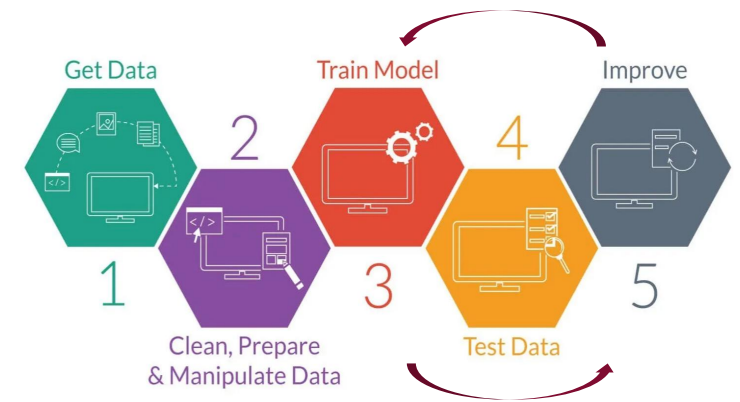
- Taler imod forurening.
- Taler for forurening.

	<b>Aktivitetstype</b> Metal, maling og lakering af
	<b>Branchekode</b> Overfladebehandling af metal
	<b>Driftsperiode</b> 41.03 år
	<b>Driftsalder</b> 29.7 år
	<b>Jordart</b> Ukendt
	<b>Infiltration</b> 72 mm
	<b>Grundvandsdybde</b> 0.65 meter

JAR prediktion per aktivitet

Vi kunne også se på:  
- Kommune nummer  
- Nedbør

# Use cases og modellens nøjagtighed



**1. Forureningsforudsigtelse**  
Kan vi forudsige, hvor der er forurenet og dermed prioritere lokaliteter bedre?

Sandsynlighed for forurening

Høj

Lav

90-95% præcision

**2. Sandsynlighed for fund**  
Hvor sandsynligt er det et fund i en prøve af en given parameter gruppe?

Eksempel på output

**72%**

Sandsynlighed for fund

Lav præcision, pga. for lidt data i Jupiter

**3. PFAS-risiko**  
Kan vi ved hjælp af gamle prøver sige om vi vil være PFAS til stede i vores vand?

Historiske prøver

Sandsynlighed for PFAS

**65%**

Risiko for PFAS til stede

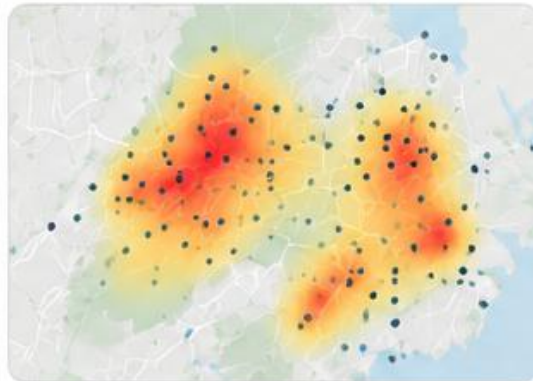
Lav præcision, pga. for lidt data

# Hvordan kan vi bruge algoritmen?



## 1. Forureningsforudsigelse

Kan vi forudsige, hvor der er forurenet og dermed prioritere lokaliteter bedre?



Sandsynlighed  
for forurening



Prioriterer årligt ca. :

- 150 indledende undersøgelser
- 20 videregående grundvandsundersøgelser

Fra data  
til bedre  
beslutninger



Store mængder  
af historiske data



Machine learning  
modeller



Indsigt og  
forudsigelser



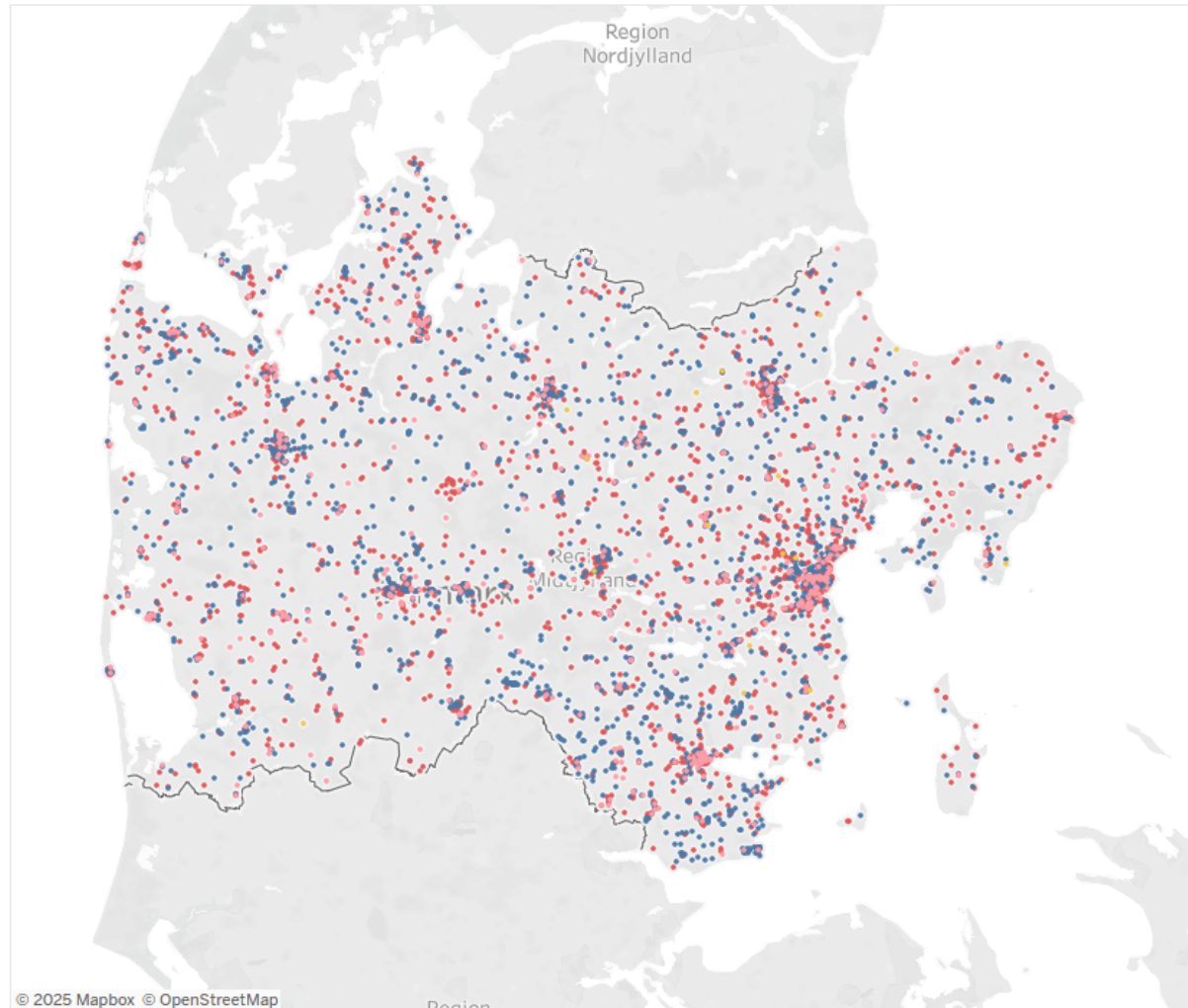
Bedre prioritering  
og målrettede indsatser



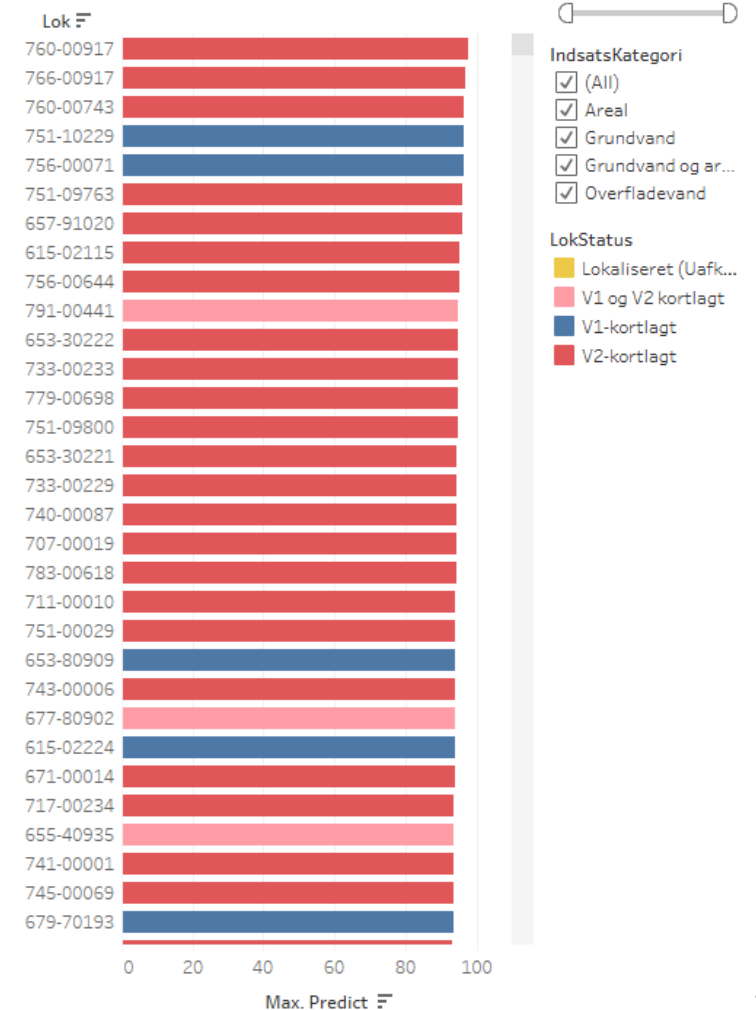
# Hvordan kan vi bruge modellen?



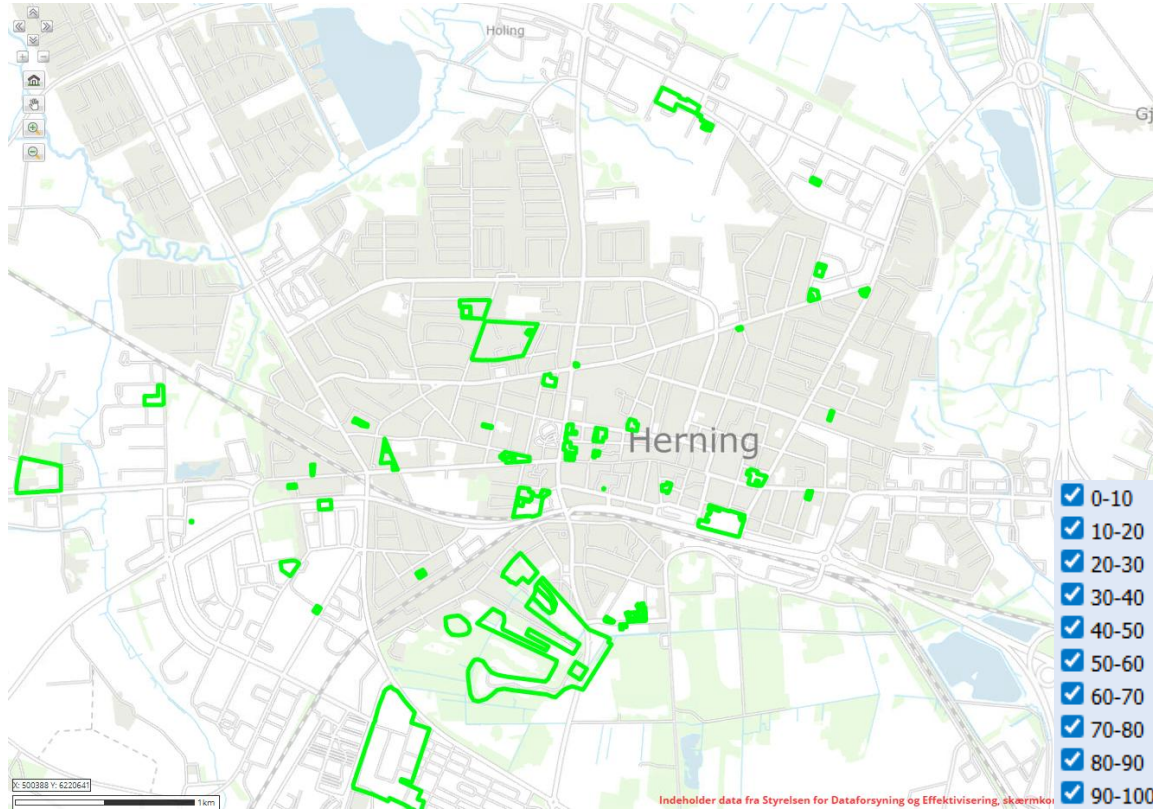
Kort



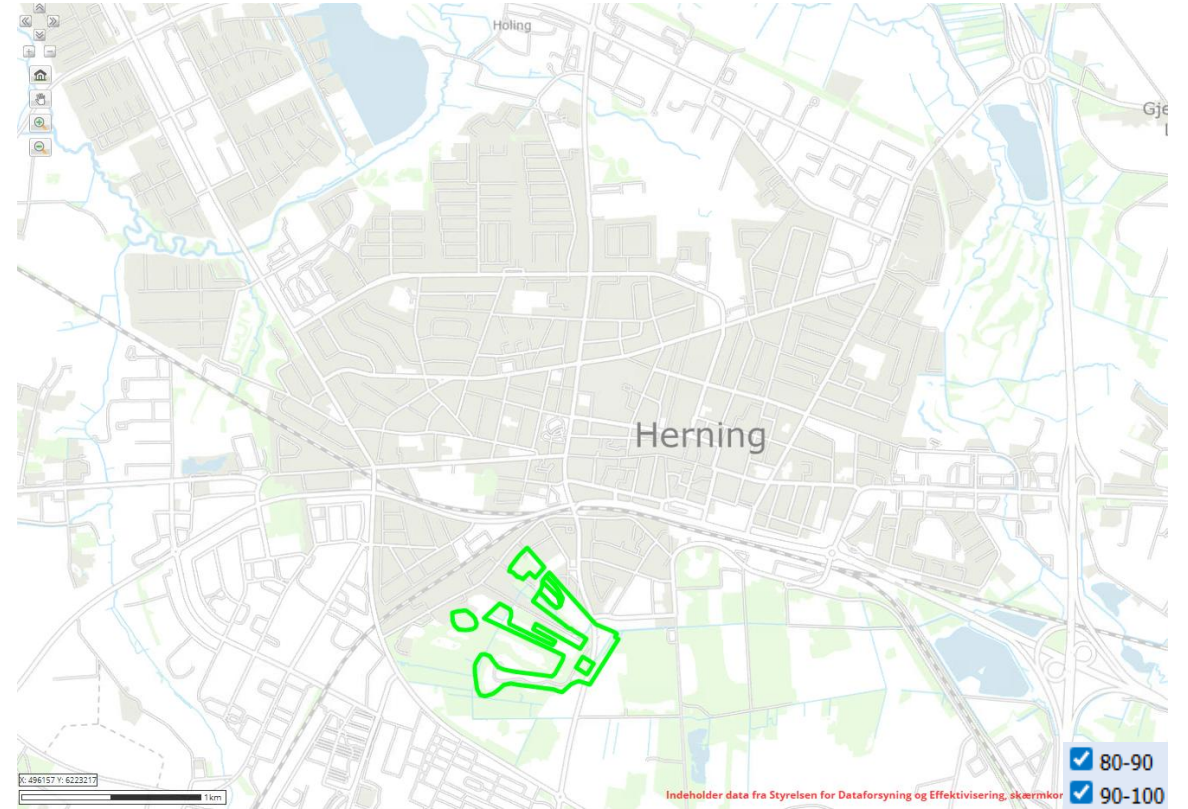
Prædiktion



# Eksempel - Herning



Alle lokaliteter i indsats, der forudsiges forurennet



Alle lokaliteter i indsats, der med 80-100% sandsynlighed vurderes forurennet

# Hvad har vi lært indtil nu?



## POTENTIALE

 **Algoritmen giver brugbare forudsigelser**

---

 Første år har vi brugt modellen

 "Hit raten" i forhold til forudsigelse ses senere



## BEGRÆNSNINGER

 **Få lokaliteter tilbage**  
I forhold til områdebaseret tilgang bliver forudsigelsen ikke så afgørende i alle tilfælde

---

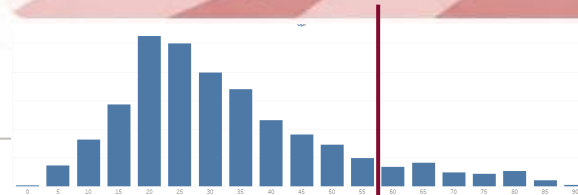
 **Algoritmen mangler faglig kontekst**

- Kender ikke forureningsstofferne fysisk-kemiske egenskaber
- Kan forudsige på ja / nej niveau
- Ikke på risikobillede



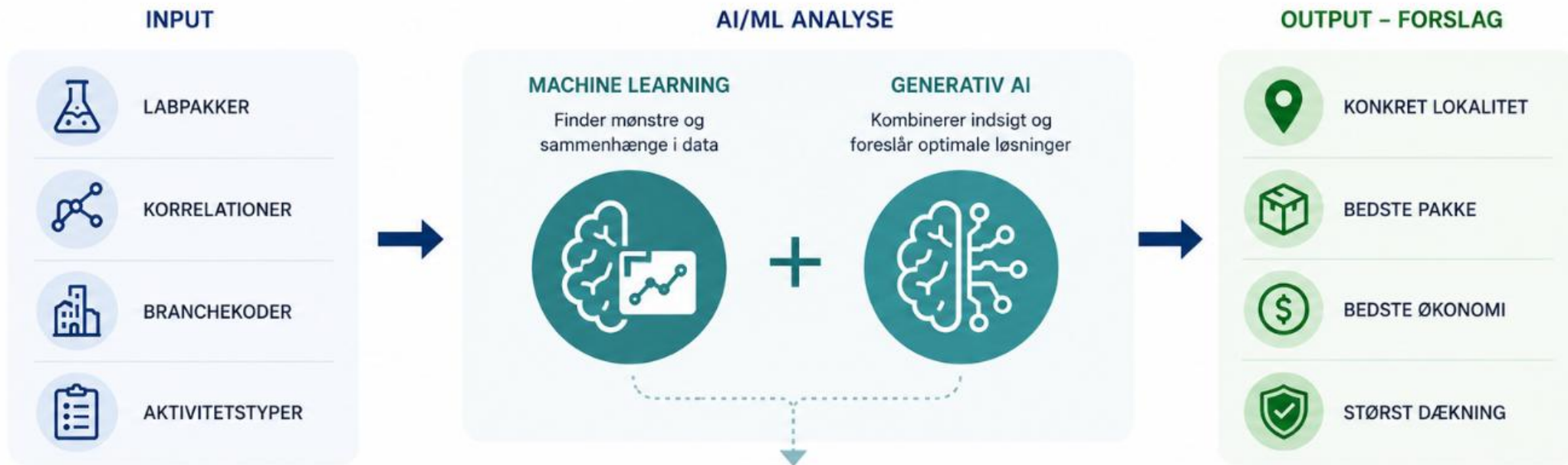
## DET STRATEGISKE SPØRGSMÅL

**Tør vi være mere dristige?**



## AI-optimering af laboratoriepakker

Machine learning og Generativ AI arbejder sammen om bedre prioritering



# Tak Spørgsmål?



Jeg har anvendt ChatGPT som hjælpemiddel i forbindelse med udarbejdelsen af disse slides. ChatGPT har på baggrund af mine beskrivelser udarbejdet flere figurer.