

ANCHORMEN



RAOUL GROULS

BASICS 2



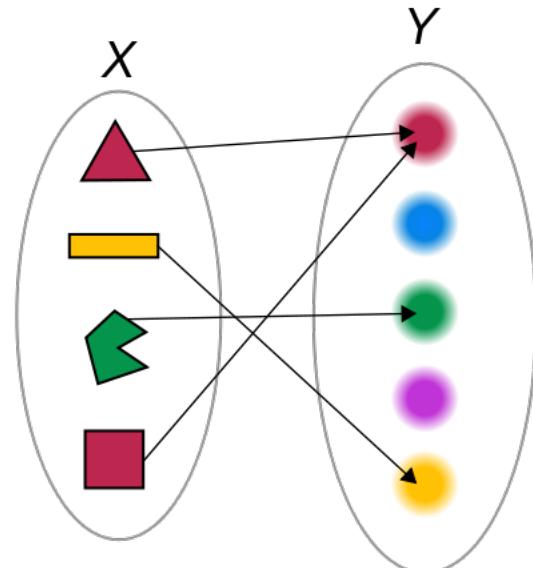
MACHINE LEARNING



Leer een functie f die input X afbeeldt op Y .

X zijn “features”, oftewel kenmerken

Y zijn uitkomsten (labels of getallen).



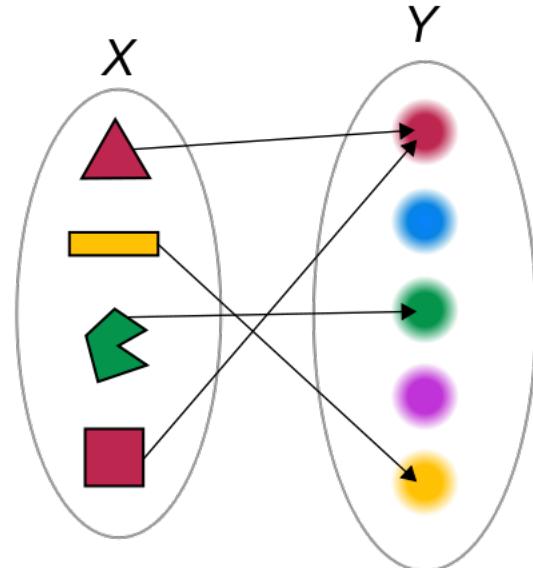
MACHINE LEARNING



In het voorspellen worden fouten gemaakt.

We onderscheiden:

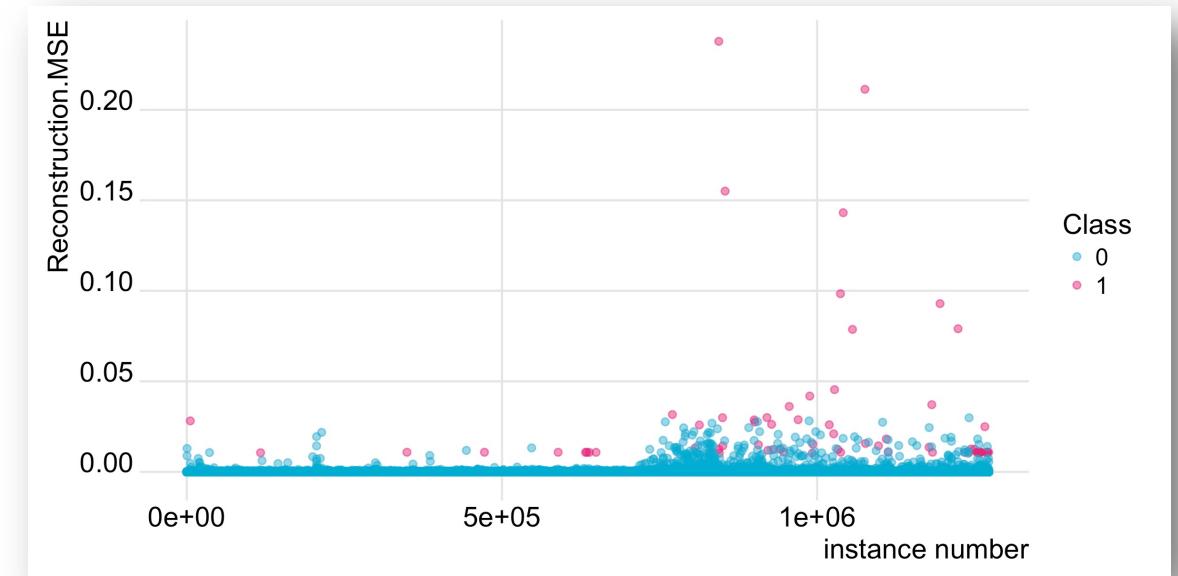
- Anomalies
 - Natuurlijke variantie
 - Noise
 - Bias
 - Variance



ANOMALIES

Noise is een fout in de labels (Y) of in de meting van de features (X).

Anomalies zijn een breder concept dat ook de natuurlijke variatie binnen een populatie omvat.



Fraude detectie

ANOMALIES

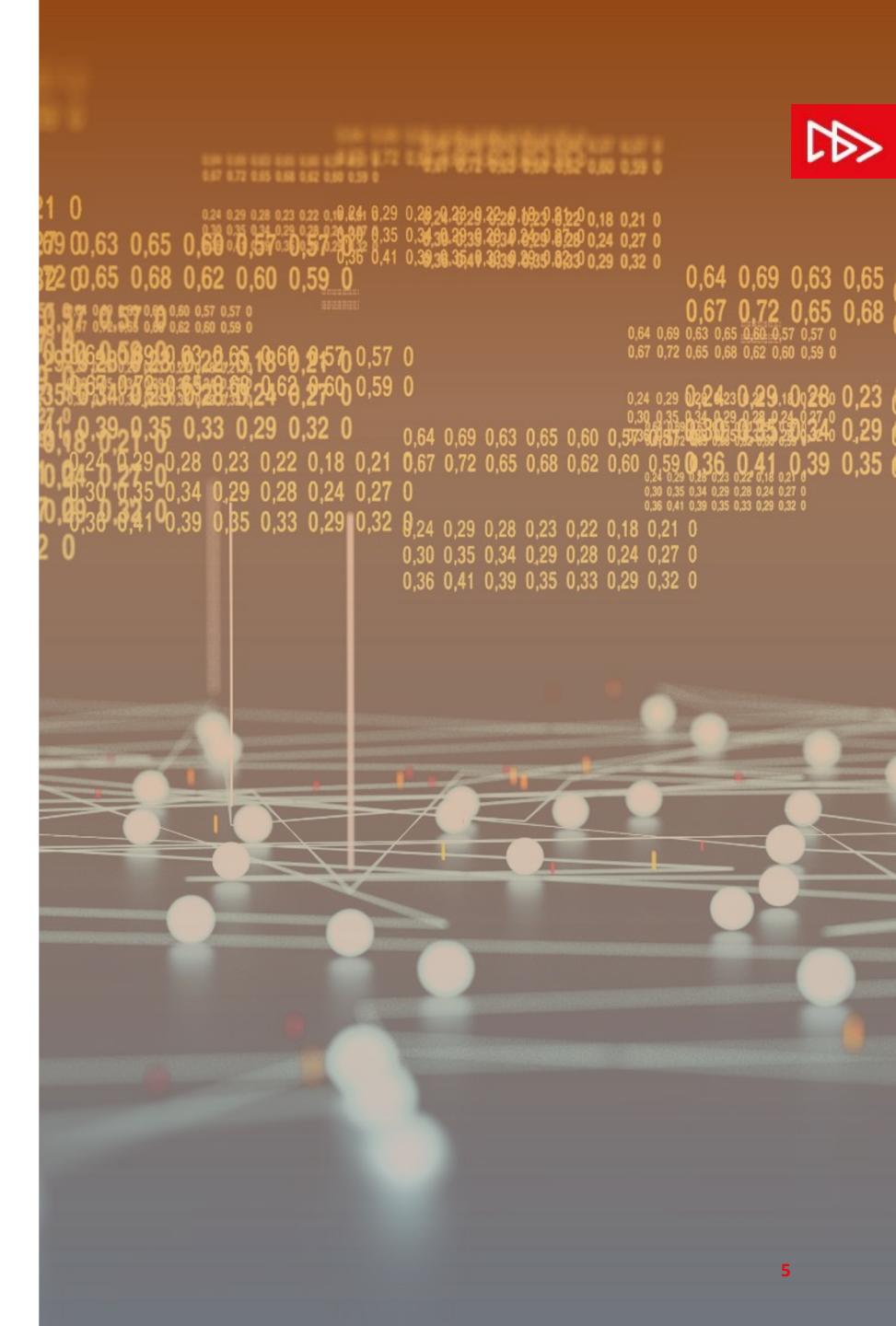


Als ik de lengte van een persoon opmeet, en als waarde 2.50m vind:

- Is dit altijd een anomalie? Waarom wel of niet?
- Is dit altijd noise? Wanneer wel of niet?

En als ik 1.75m vindt;

- Kan dit een anomalie zijn?
- En noise?



NOISE

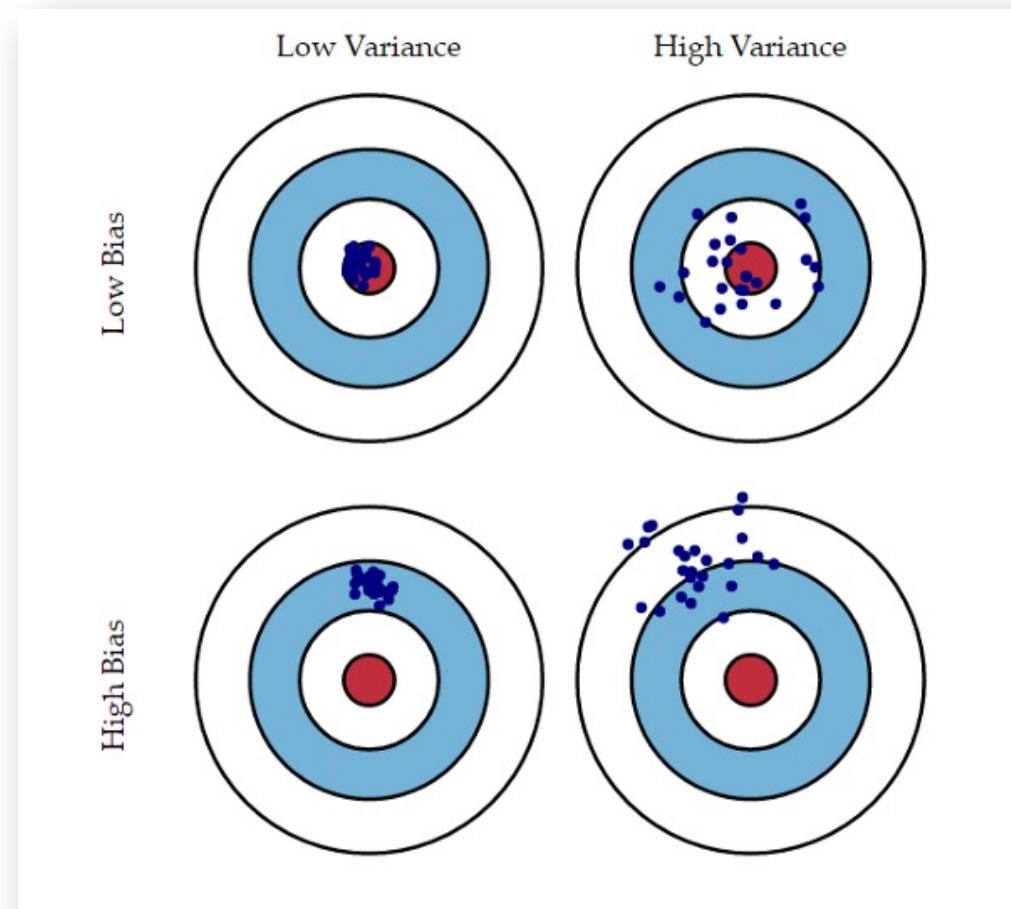


Data = echte signaal + noise

Noise is toegevoegde informatie zonder betekenis.

Twee soorten:

- Random error (variance)
- Systematic error (bias)

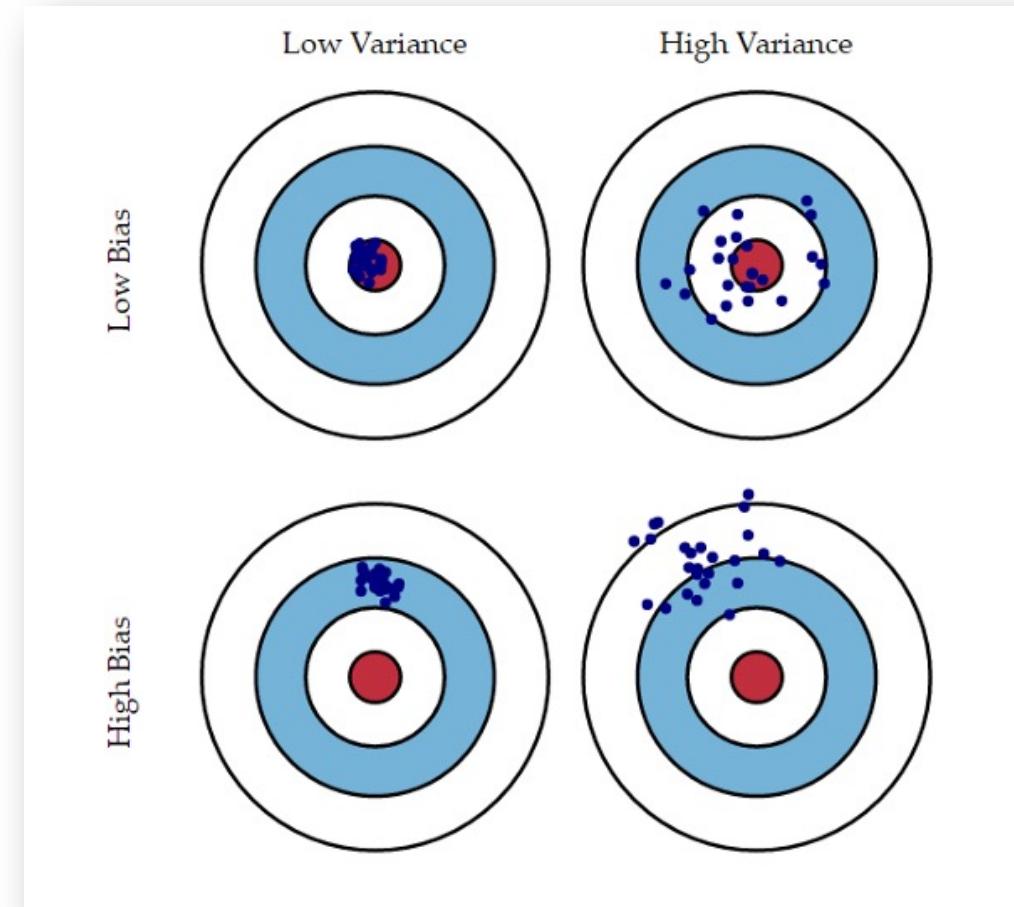


NOISE



Als we het over de error hebben:

- **Variance:** de metingen zijn niet precies
- **Bias:** de metingen hebben een structurele afwijking



NOISE

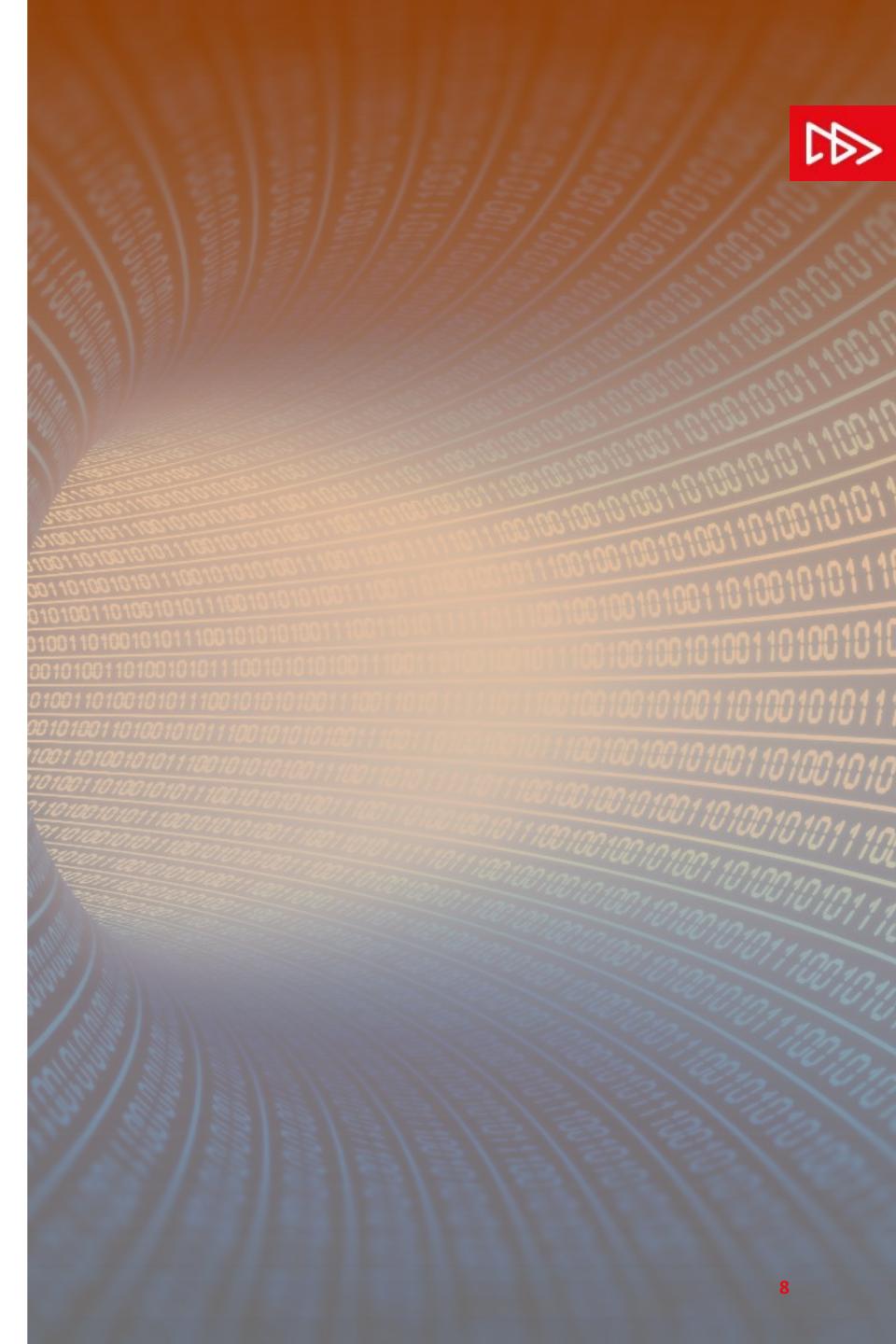


Bij het voorspellen van stadsdrukte:

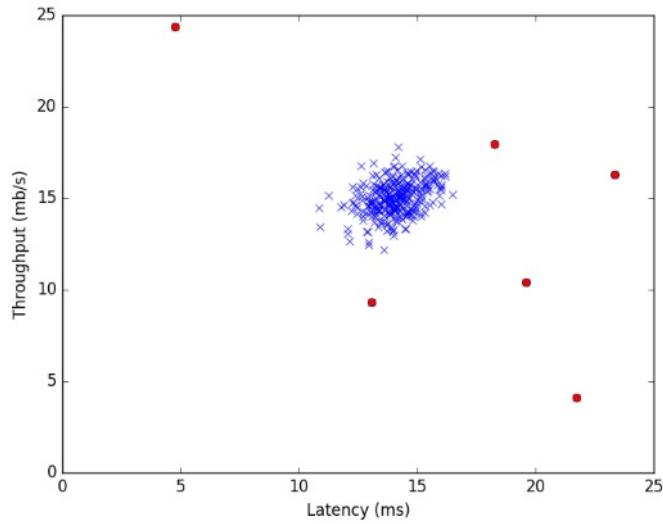
Wat is hier bias? En wat is variance?

En bij het voorspellen van fraudekans?

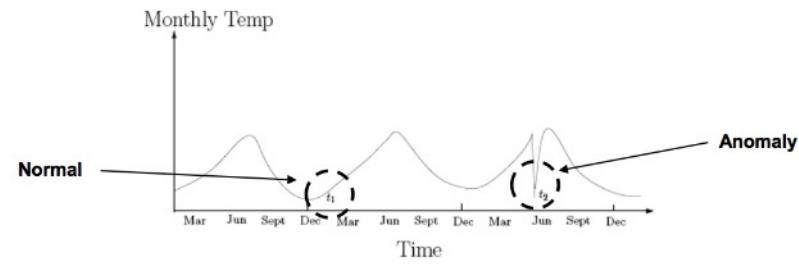
Kan bias zich ook “verstoppen” als variance?



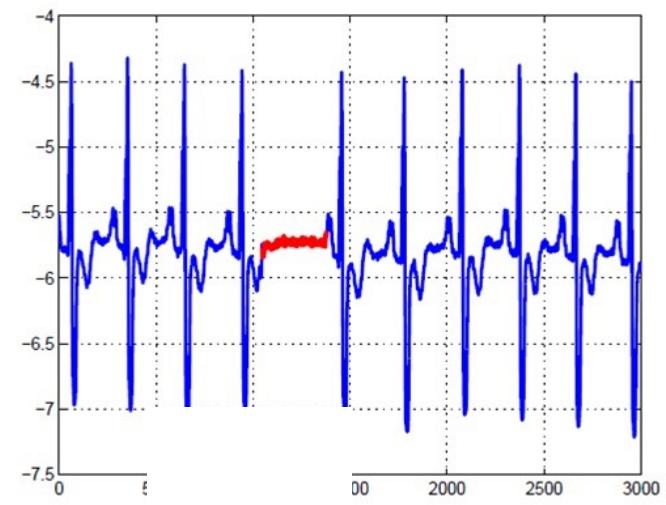
TYPES OF ANOMALIES



Point



Contextual



Collective

CONFUSION MATRIX



Type I Error



False positive

Type II Error



False negative

CONFUSION MATRIX



Een classificatiemodel kan op twee manieren een fout maken.

De “verwarring” die ontstaat bij het maken van verkeerde voorspellingen kan worden weergeven in een confusion matrix:

- FP is een type I error
- FN is een type II error

		Actual Value (as confirmed by experiment)	
		positives	negatives
Predicted Value (predicted by the test)	positives	TP True Positive	FP False Positive
	negatives	FN False Negative	TN True Negative

PRECISION VS RECALL



Recall wordt ook wel sensitivity genoemd

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

	Actual positive	Actual negative
Predicted positive	Precision	
Predicted negative		

	Actual positive	Actual negative
Predicted positive	recall	
Predicted negative		



PRECISION VS RECALL

Strategie: schieten met hagel

Je selecteert heel veel gevallen, bijvoorbeeld wel 90% (TP+FP), zodat je alle TP gevallen vindt maar ten koste van heel veel FP, en heel weinig FN.

	Actual positive	Actual negative
Predicted positive	Precision	
Predicted negative		

In welke gevallen zou je zo iets willen doen?

Heb je nu een hoge of lage precision?

En een hoge of lage recall?

	Actual positive	Actual negative
Predicted positive	recall	
Predicted negative		

PRECISION VS RECALL



Strategie: scherpschutter

Je selecteert maar 1% ipv 10%, maar als je iemand selecteert maak je eigenlijk nooit een fout.

In welke gevallen zou je zo iets willen doen?

Heb je nu een hoge of lage precision?

En een hoge of lage recall?

	Actual positive	Actual negative
Predicted positive	Precision	
Predicted negative		

	Actual positive	Actual negative
Predicted positive	recall	
Predicted negative		

PRECISION VS RECALL



Stel, je hebt een kinderfeestje waarbij je met een groep kinderen gaat zwemmen. Je moet nu 8 kinderen verzamelen.

1. Je laat per ongeluk één kind achter, dat wel mee had moeten komen.
2. Je neemt per ongeluk een kind mee, dat niet bij het feestje hoort.

Welke van de twee fouten is een FP? En welke een FN?

CONFUSION MATRIX

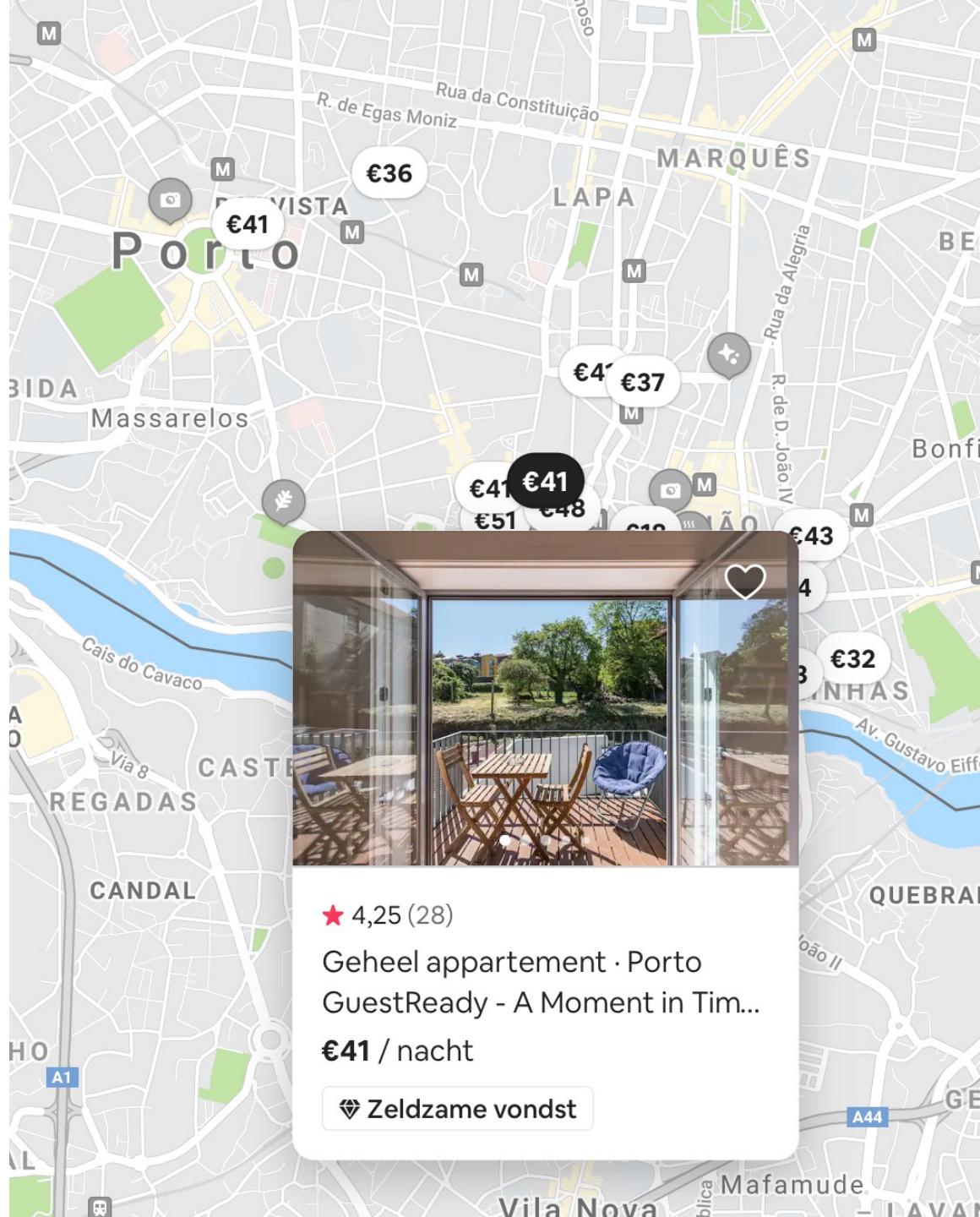
AirBnB heeft software die background checks doet via analyse van o.a. social media om in kaart te brengen hoe betrouwbaar je persoonlijkheid is (o.a. narcisme, psychopathie, etc).

Wat betekent hier een FP? En een FN?

Welke fout is voordeliger voor AirBnB?

En voor de verhuurder?

En voor de huurder?



PRECISION VS RECALL

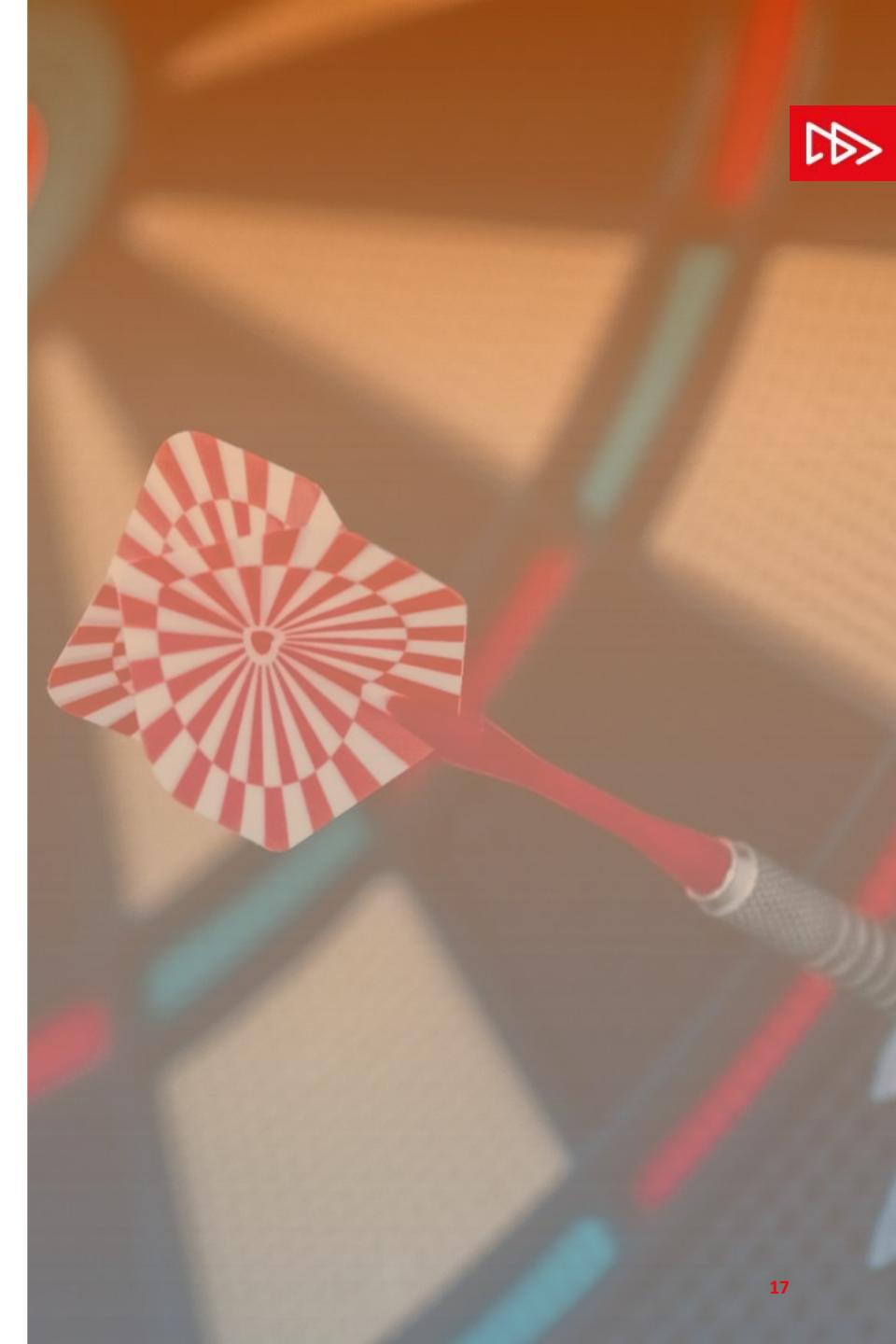


Het soort fout dat je maakt heeft gevolgen voor de performance. Elk type fout heeft een andere consequentie.

Welk type fout zwaarder weegt, hangt af van de context! Dit is vaak een ethische afweging die de opdrachtgever moet maken (en niet de programmeur).

Conclusie 1: je **wilt dus niet perse de hoogste performance**. Soms is de ene soort fout erger dan de andere soort fout!

Conclusie 2: de opdrachtgever moet richting geven aan de verhouding precision-recall.





ALGORITMES NEMEN MENSELIJKE BIAS OVER

Toepassing: COMPAS algoritme gebruikt in de US om de kans op recidive te voorspellen.

Bias: algoritme bleek racistisch. Zwarten werden 2x zo vaak fout gemisclassificeerd.

Probleem: de ontwerper van het algoritme weigert de broncode van zijn algoritme te laten zien.

Oorzaak: overgenomen bias. Er wordt een vragenlijst gebruikt waarbij in de antwoorden de bias van de samenleving doorklinkt. Bijvoorbeeld zwarten en blanken roken evenveel marihuana, maar zwarten worden er bijna 10x zo vaak om veroordeeld.

Bronnen:

<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>

<https://github.com/propublica/compas-analysis>





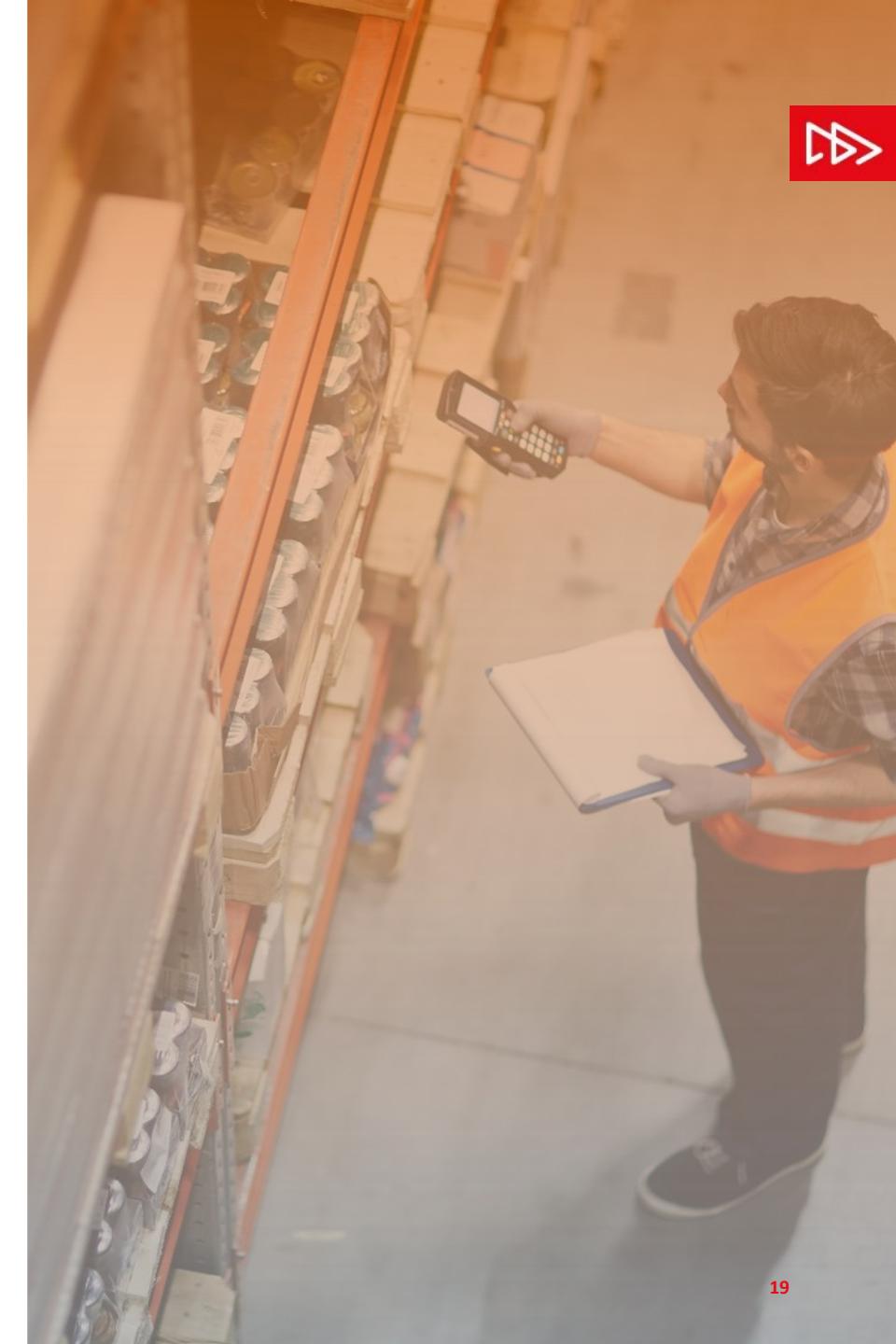
ALGORITMES NEMEN MENSELIJKE BIAS OVER

Toepassing: Amazon gebruikt algoritme voor HR om sollicitanten ratings te geven.

Bias: Algoritme had een voorkeur voor mannen.

Oorzaak: overgenomen bias. Het algoritme was getraind door de observatie van patronen in 10 jaar van sollicitaties bij Amazon.

Bron:<https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scaps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>



ALGORITMES NEMEN MENSELIJKE BIAS OVER

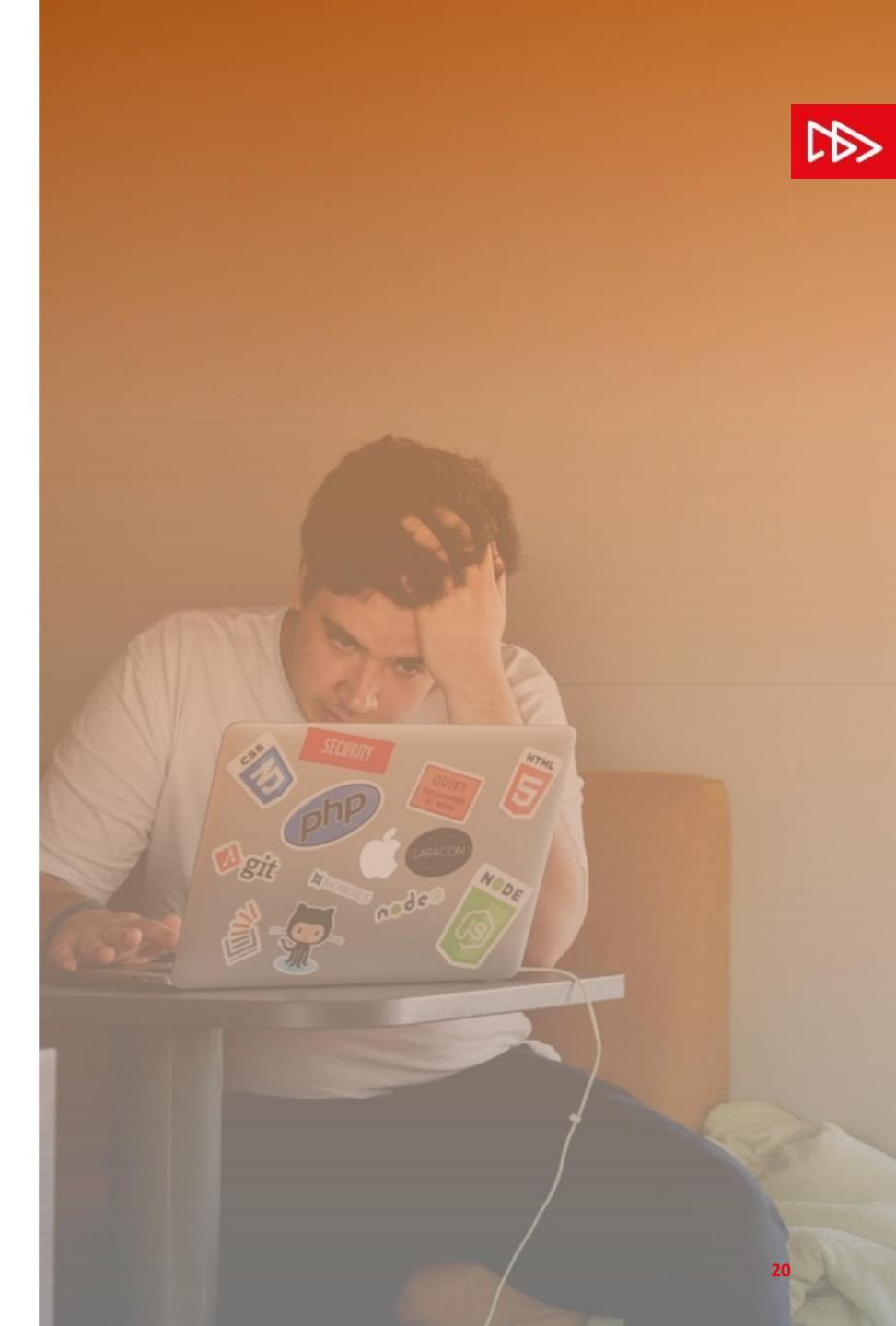


Maar lost het afschaffen van het algoritme het probleem op?

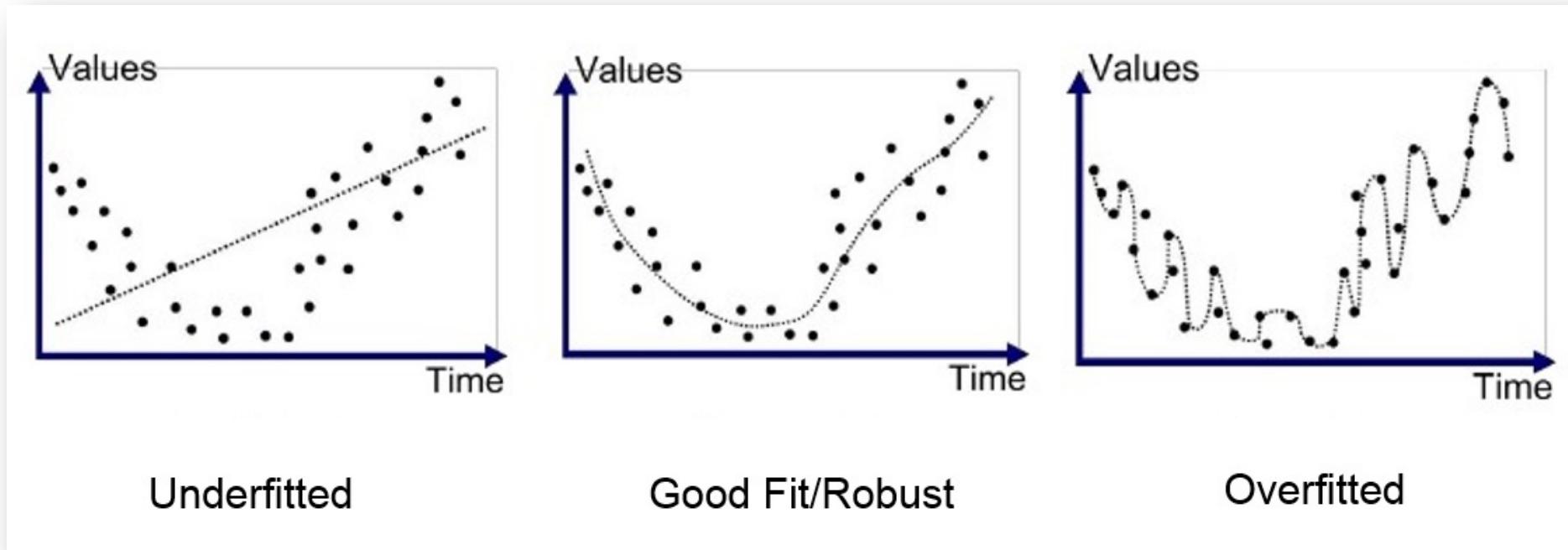
Wellicht accurater: het algoritme legt de menselijke bias bloot.

Hoe bevraag je een algoritme?

Suggestie: laat het algoritme vertellen: als deze feature anders was geweest, was je in een andere categorie terecht gekomen.



GENERALISATIE

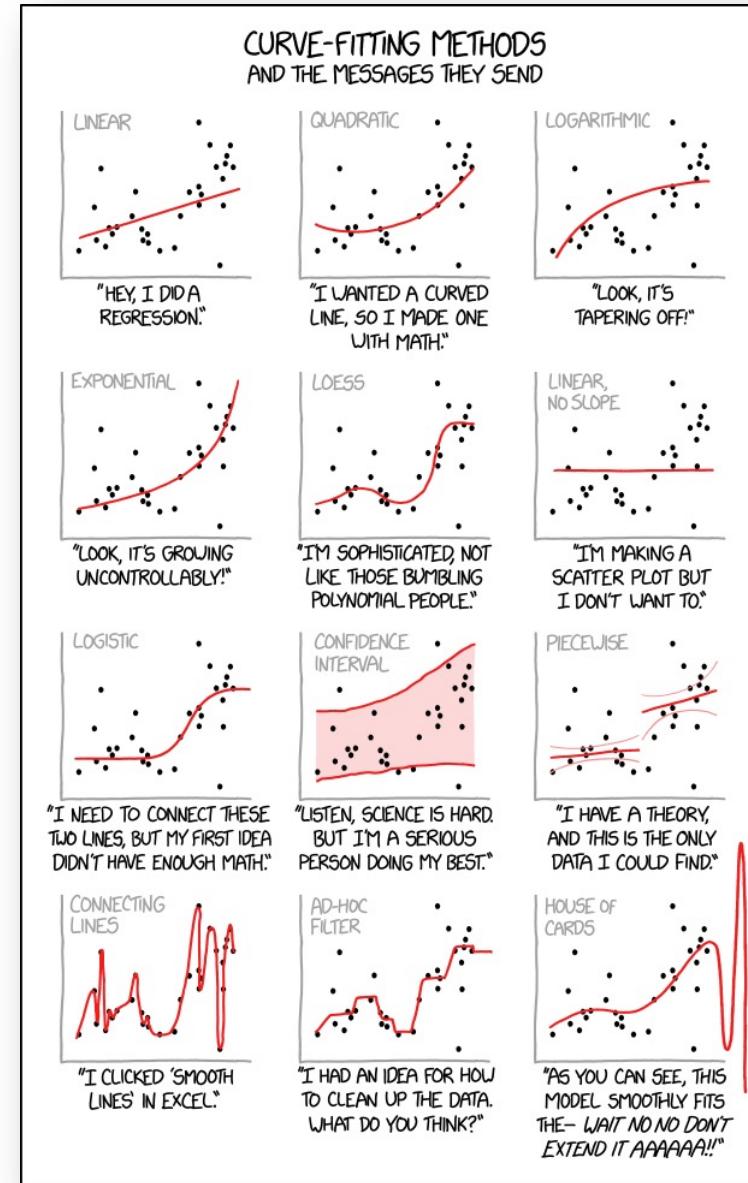


Underfitted

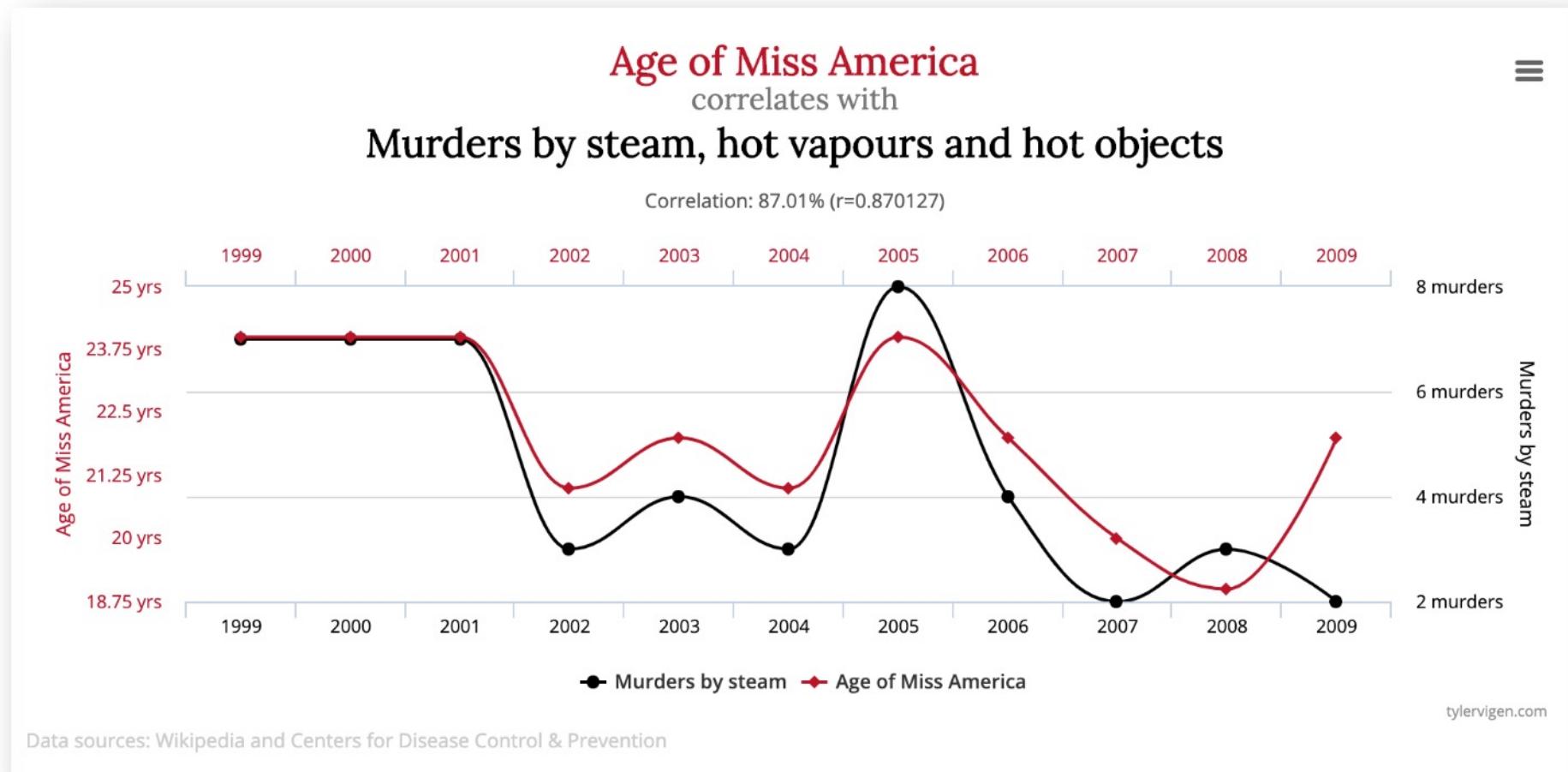
Good Fit/Robust

Overfitted

CURVE-FITTING METHODS



SPURIOUS CORRELATIONS



Bron: <http://tylervigen.com/spurious-correlations>

SPURIOUS CORRELATIONS



Dit komt vaker voor dan je zou denken!

Statstiek is een vakgebied dat zich slecht leent voor intuitief gebruik.

Bij hoeveel mensen is de kans dat twee willekeurig gekozen mensen dezelfde verjaardag hebben, groter dan 50%? En groter dan 99%?

A) 23

B) 57

C) 183

D) 365



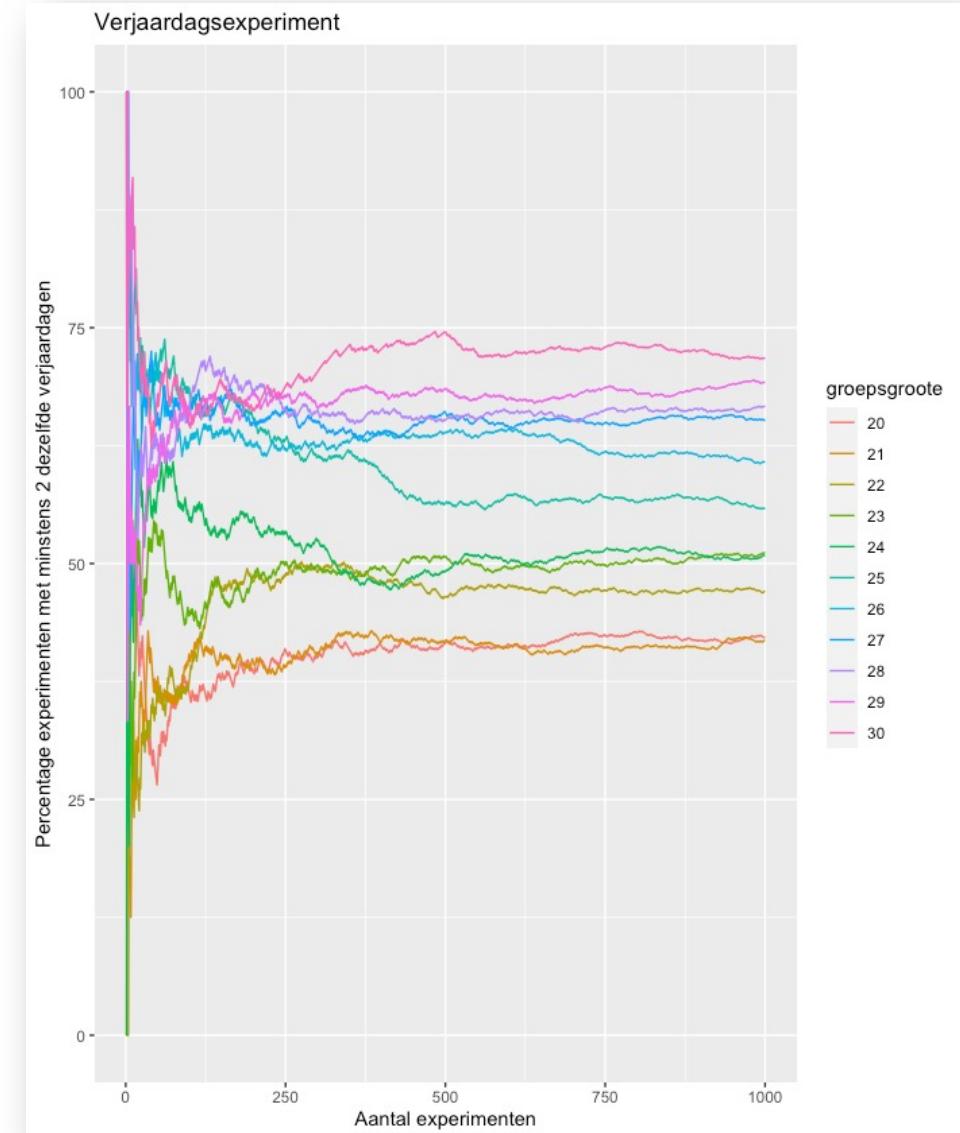
SPURIOUS CORRELATIONS



50% : 23 mensen

99% : 57 mensen

Toelichting: <https://nl.wikipedia.org/wiki/Verjaardagenparadox>



CORRELATIE IS GEEN CAUSATIE

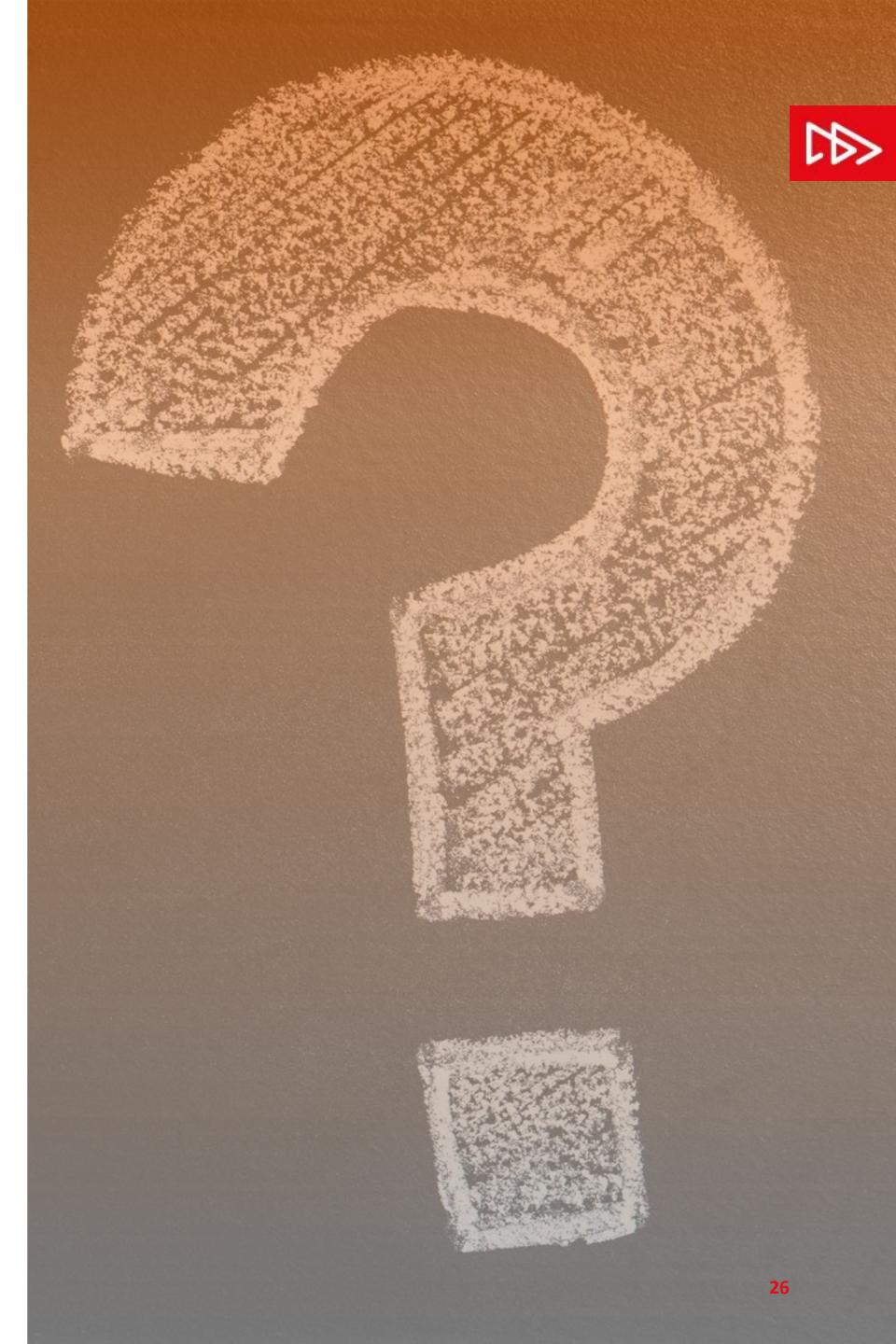


Er bestaat een correlatie tussen ijsconsumptie en verdrinkingsdood.

Dit wordt duidelijk veroorzaakt door een derde factor, namelijk: de temperatuur.

Kunnen we nu ijsconsumptie gebruiken om voorspellingen te doen over het aantal verdrinkingsdoden?

En zo ja, wat is een mogelijk probleem?



CORRELATIE IS GEEN CAUSATIE



Ja, dat kan heel goed!

De meeste voorspellende algoritmes zijn gebaseerd op correlaties ipv causaties.

Er ontstaat pas een probleem als je:

- De relatie vervolgens omdraait: B.v. Ik laat mijn kind geen ijs meer eten, zodat de kans dat hij verdrinkt minder groot is.
- Niet doorhebt dat de correlatie in feite een ongewenste bias weergeeft.





ANCHORMEN
data activators

THANK YOU FOR YOUR ATTENTION

-  Pedro de Medinaalaan 11,
1086 XK Amsterdam

 020 - 773 1972

 www.anchormen.nl