

Causaliteit en Machine learning

Presentatie voor het Zorginstituut

Misja Mikkers & Gertjan Verhoeven

Nederlandse Zorgautoriteit (NZa) & Tilburg University

9 januari 2020

Inleiding

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie

Blog en code



Voorbeeld van een dataset

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie

Blog en code

X_1	X_2	X_3	X_i	I	Y0	Y1
Man	9	14	1	0	67	NA
Vrouw	60	36	0	1	NA	113
Vrouw	7	2	1	1	NA	54
...

Voorspellen versus begrijpen

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

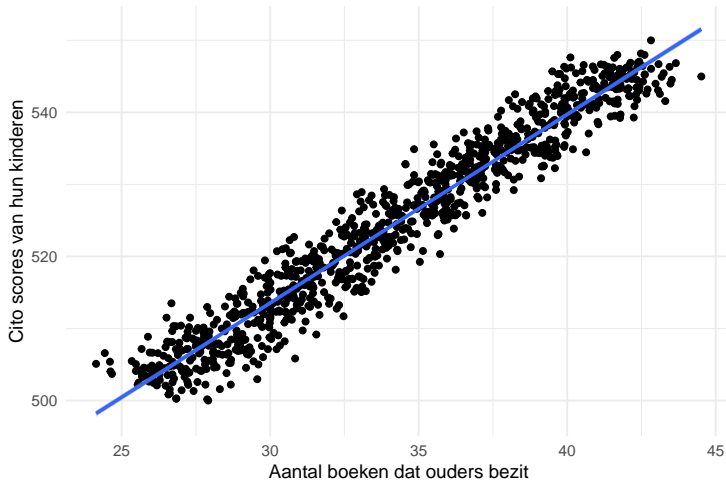
Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie

Blog en code



Voorspellen versus begrijpen

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

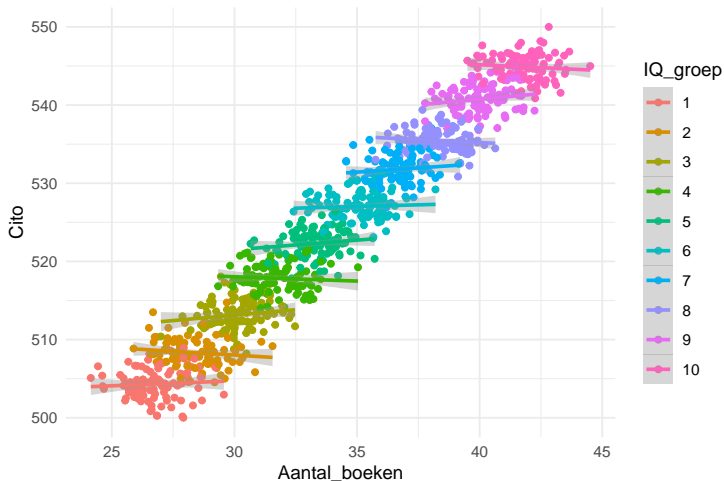
Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie

Blog en code



Basisvormen DAGs

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

Causaliteit

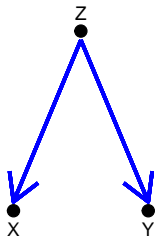
DAGs

Machine learning
en causaliteit

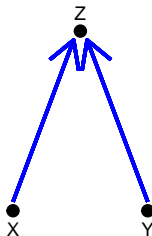
Conclusie

Blog en code

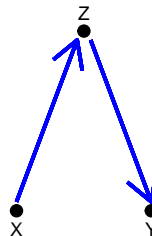
Confounder



Collider



Mediator



Confounder

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

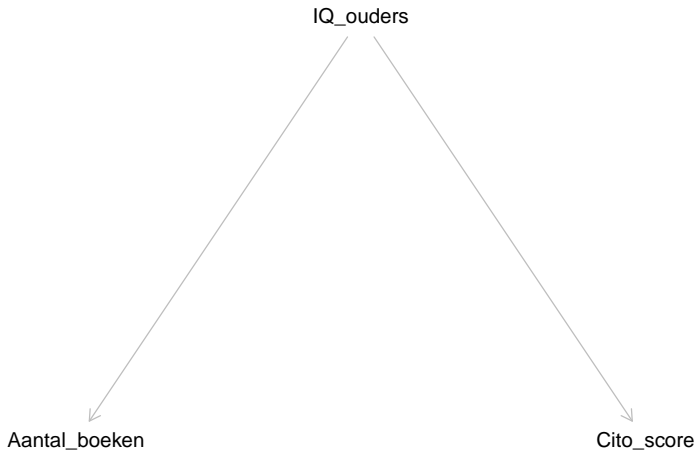
Causaliteit

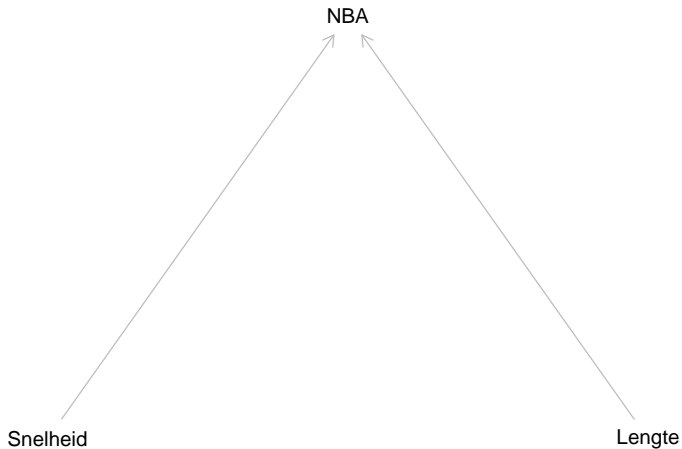
DAGS

Machine learning
en causaliteit

Conclusie

Blog en code

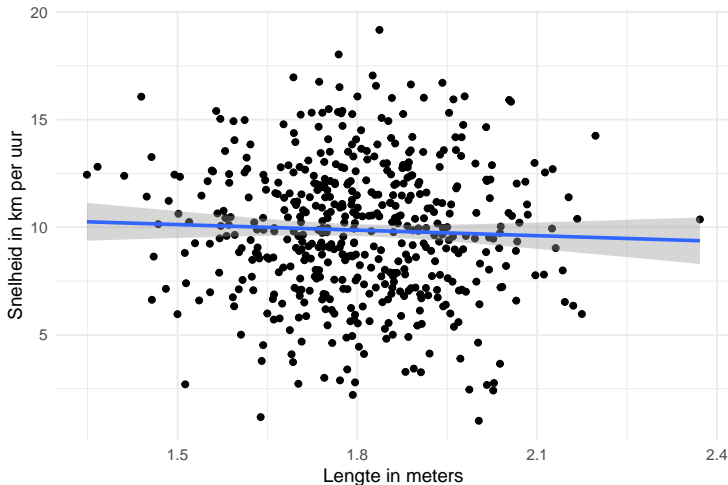




Veroorzaakt Lengte Snelheid?

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven



Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

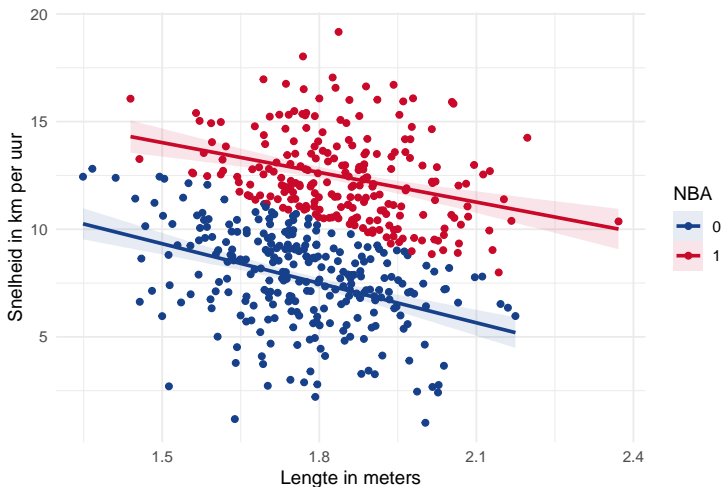
Conclusie

Blog en code

Veroorzaakt Lengte Snelheid?

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven



Inleiding

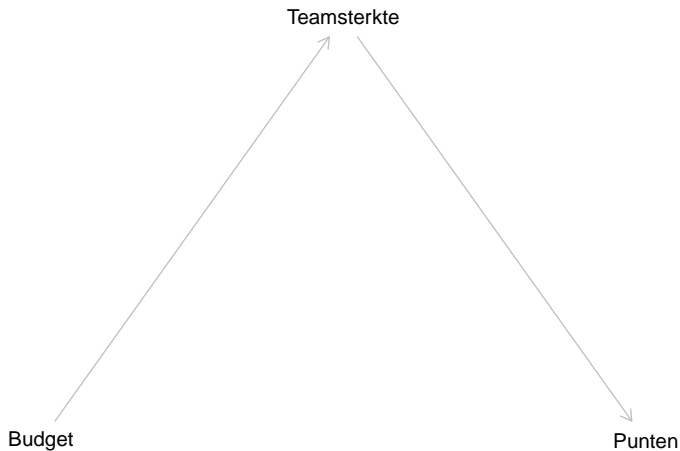
Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie

Blog en code



Budgetten en punten

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

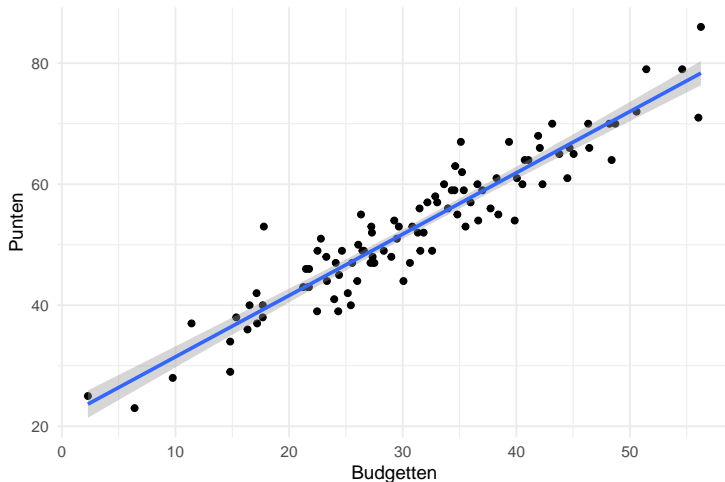
Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie

Blog en code



Budgetten en punten

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

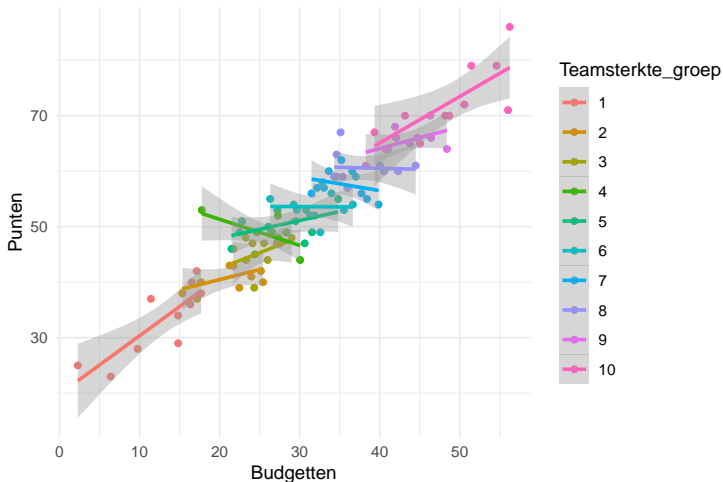
Causaliteit

DAGS

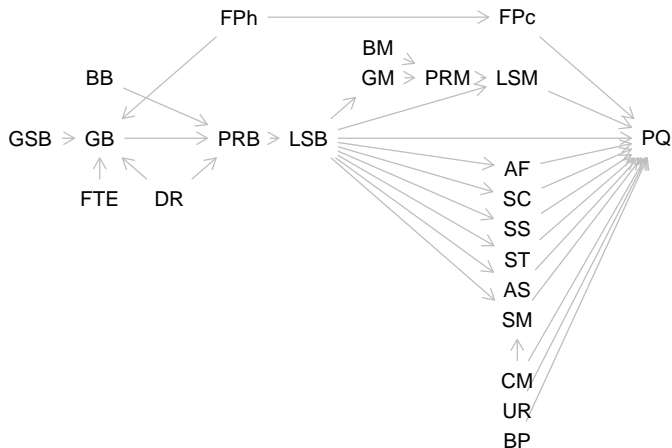
Machine learning
en causaliteit

Conclusie

Blog en code



Voorbeeld van een meer complexe DAG



Machine learning en causaliteit

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

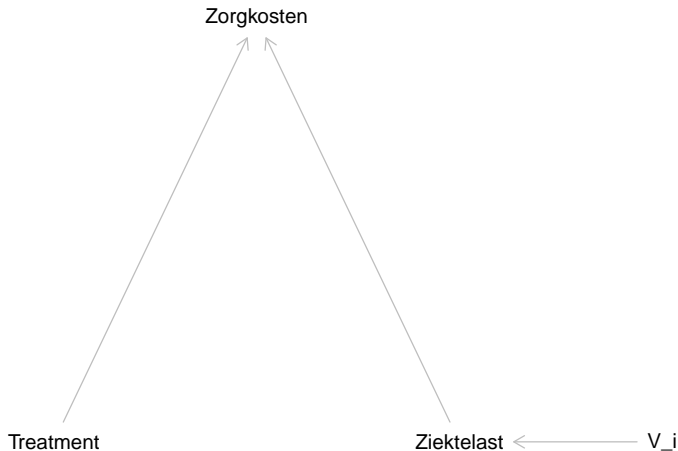
Causaliteit

DAGS

**Machine learning
en causaliteit**

Conclusie

Blog en code



De variabele *Ziektelast* is bepaald op basis van de volgende formule:

$$\text{Ziektelast} = \text{abs}(\text{scale}(V1^3 + 2 * V2 + 3 * V3^2 + 4 * V4 + 5 * V5 + 6 * V6 * V7)) + \epsilon$$

1. Fit een Random Forest model op de data 2. Bepaal het gemiddelde treatment effect met behulp van generalized random forests (grf)

We doorlopen deze procedure 2 keer:

- a. Een analyse waarbij we de ziektelast voorspellen met alle variabelen ("het verkeerde model")
- b. Een analyse waarbij we de ziektelast voorspellen met alle variabelen minus de collider *Zorgkosten* ("het goede model")

[Inleiding](#)

[Causaliteit](#)

[DAGS](#)

[Machine learning
en causaliteit](#)

[Conclusie](#)

[Blog en code](#)

Summary statistics

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie

Blog en code

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Ziektelast	1,000	0.834	0.562	0.003	0.494	0.945	2.737
Treatment	1,000	0.474	0.500	0	0	1	1
Zorgkosten	1,000	5.589	3.320	-0.069	3.460	7.192	16.431
V1	1,000	498.746	288.860	2	243	747.8	1,000
V2	1,000	492.975	293.667	1	236	741	1,000
V3	1,000	512.349	293.801	1	251.8	774	1,000
V4	1,000	489.041	288.225	1	234	729.2	1,000
V5	1,000	516.896	292.376	1	257	766	1,000
V6	1,000	497.232	284.475	1	265.8	736	1,000
V7	1,000	505.715	289.385	1	261.5	753.2	1,000
V8	1,000	501.330	276.913	3	267.8	744.2	998

Voorspellingen Random Forest

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

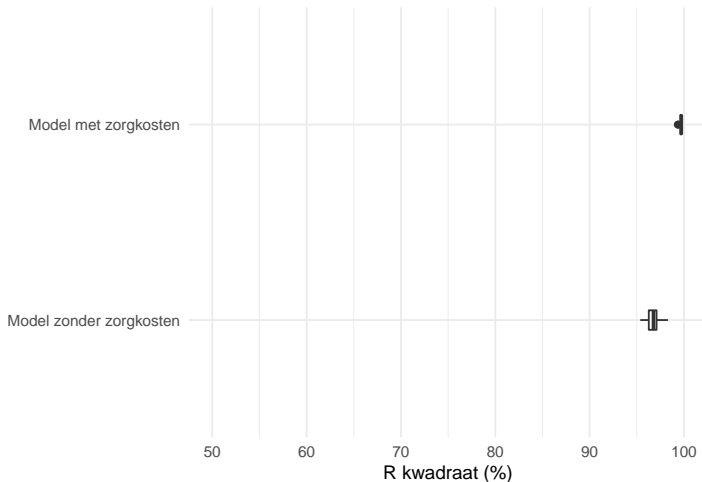
Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie

Blog en code



Schatting van het treatment effect

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

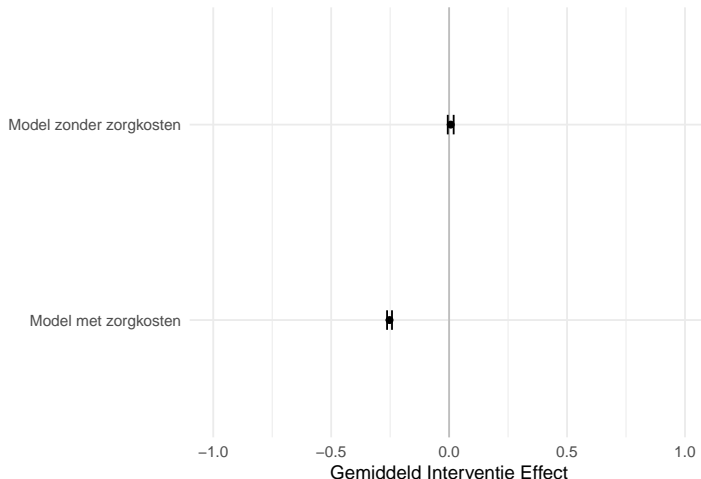
Causaliteit

DAGS

**Machine learning
en causaliteit**

Conclusie

Blog en code



- ▶ Het is essentieel om een causaal model te hebben
- ▶ Wanneer machine learning wordt gebruikt is de verleiding groot om alle variabelen in het model te gebruiken
- ▶ We hebben laten zien dat dit kan leiden tot misleidende conclusies

[https://misjamikkers.github.io/post/
causaliteit-en-machine-learning/](https://misjamikkers.github.io/post/causaliteit-en-machine-learning/)

https://github.com/misjamikkers/Meetup_Informatieberaad