

Causaliteit en Machine learning

Misja Mikkers & Gertjan Verhoeven

Nederlandse Zorgautoriteit (NZa) & Tilburg University

23 september 2019

Inleiding

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie



Voorbeeld van een dataset

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie

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

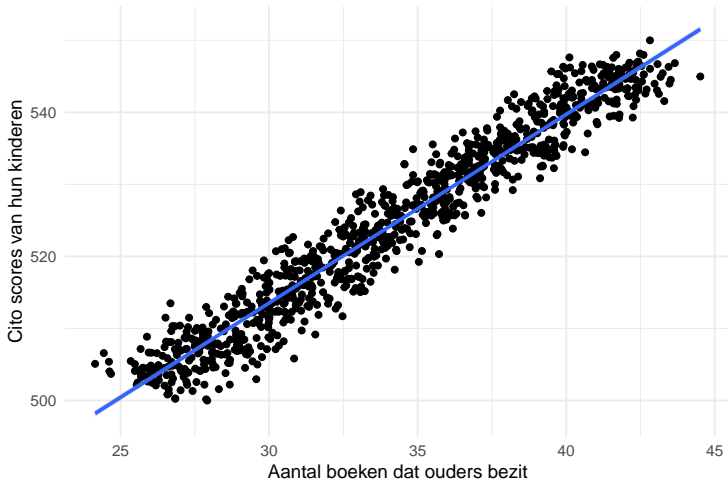
Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie



Voorspellen versus begrijpen

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

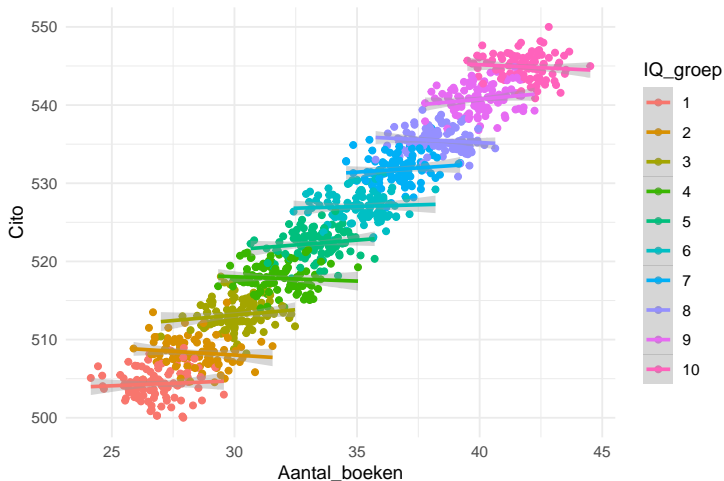
Inleiding

Causaliteit

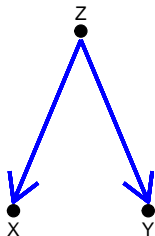
DAGS

Machine learning
en causaliteit

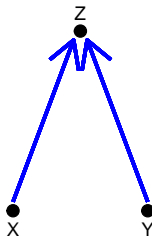
Conclusie



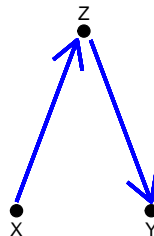
Confounder



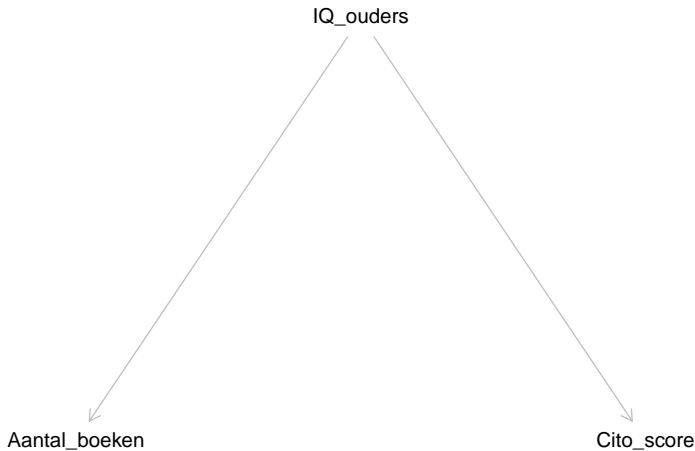
Collider

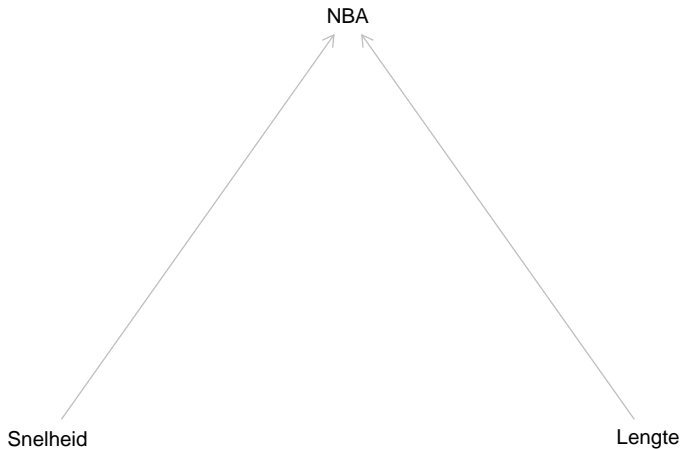


Mediator



Confounder

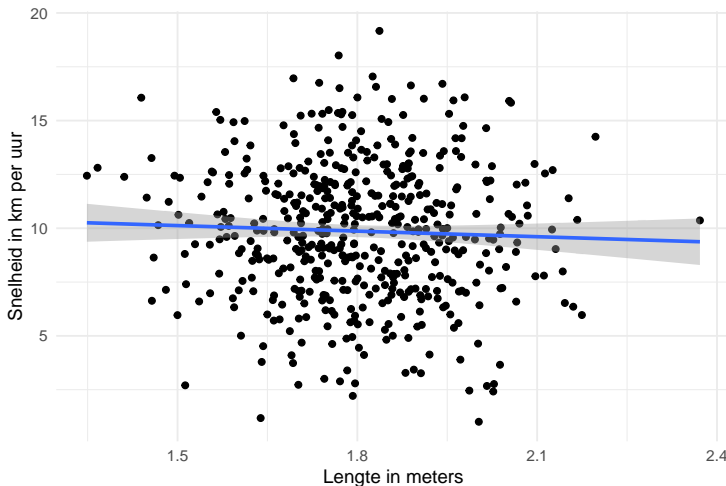




Veroorzaakt Lengte Snelheid?

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven



Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie

Veroorzaakt Lengte Snelheid?

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

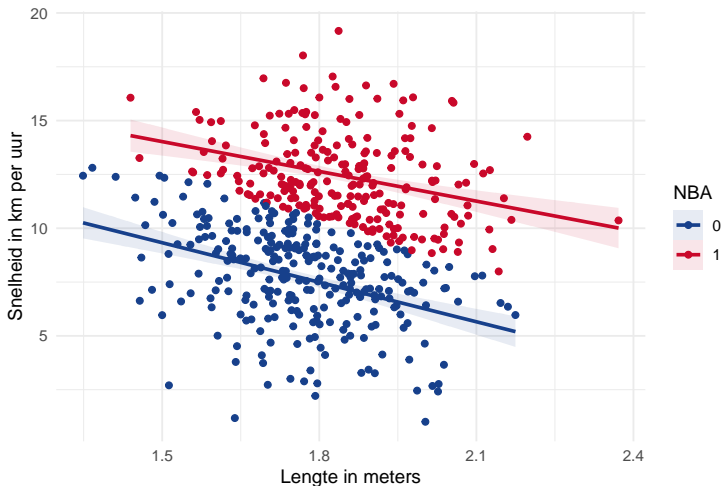
Inleiding

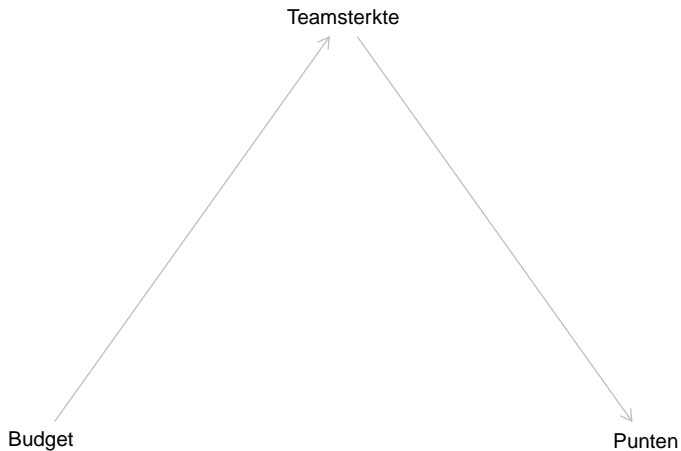
Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie





Budgetten en punten

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

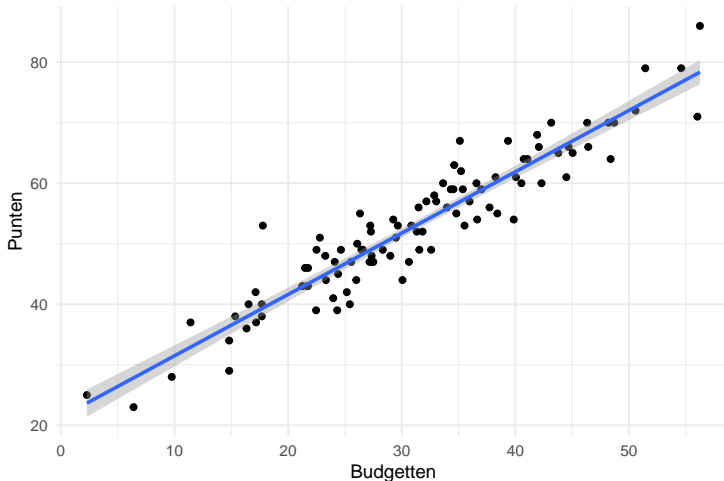
Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie



Budgetten en punten

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

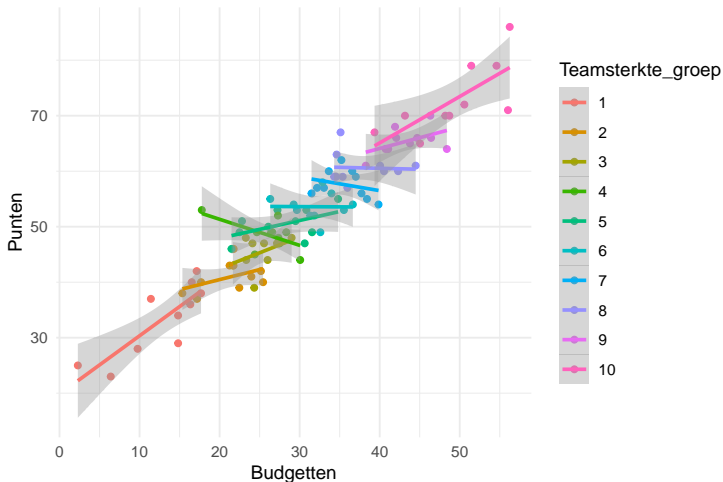
Inleiding

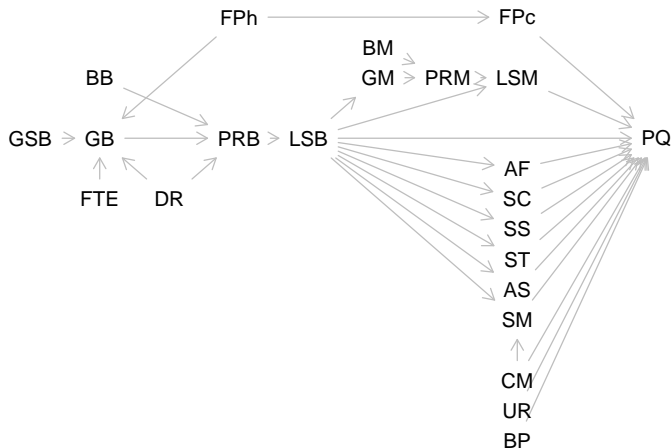
Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie





Machine learning en causaliteit

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

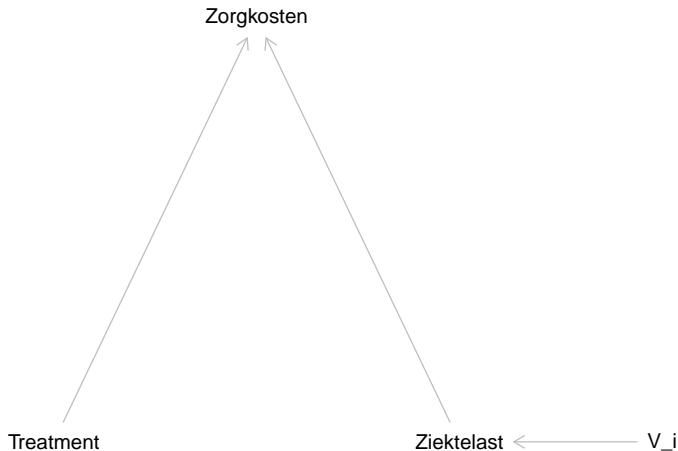
Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie



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:

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

Summary statistics

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Ziektelast	1,000	0.830	0.558	0.005	0.493	0.936	2.695
Treatment	1,000	0.493	0.500	0	0	1	1
Zorgkosten	1,000	5.629	3.197	-0.228	3.639	7.169	16.276
V1	1,000	497.768	287.495	1	254	747.2	1,000
V2	1,000	498.749	285.709	2	253.5	746.2	1,000
V3	1,000	499.119	292.200	4	233.8	746	1,000
V4	1,000	504.484	294.009	1	243.8	769	1,000
V5	1,000	489.628	282.146	1	255	734.2	1,000
V6	1,000	506.039	284.987	2	270.5	762.2	1,000
V7	1,000	496.600	283.415	1	265	735.2	998
V8	1,000	498.522	287.912	1	250.8	747.5	999

Voorspellingen Random Forest

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

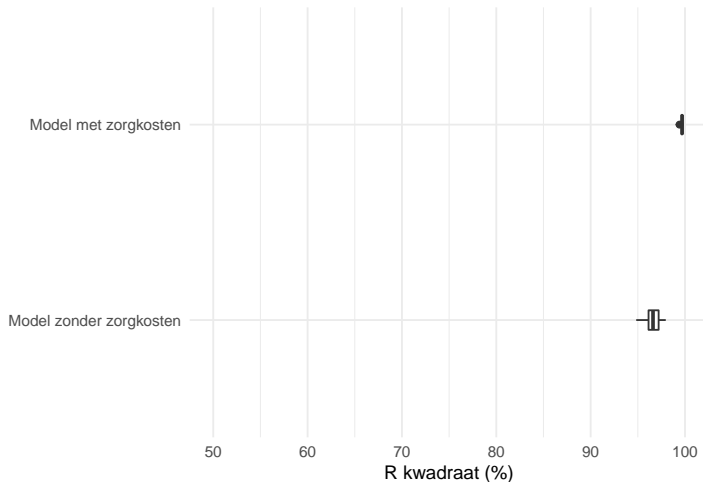
Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie



Schatting van het treatment effect

Causaliteit en
Machine learning

Misja Mikkers &
Gertjan
Verhoeven

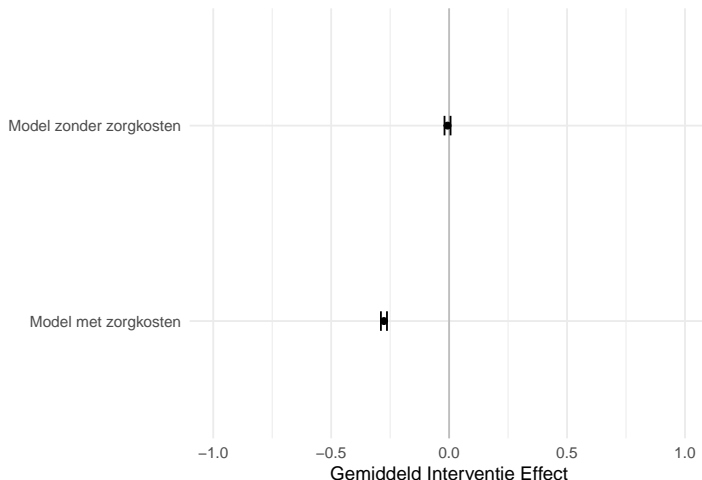
Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie



- ▶ 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