# The effect of Tilburg gold on your health

Presentation Zorginstituut: Combining causal inference and machine learning in practice

Misja Mikkers & Gertjan Verhoeven

Dutch Healthcare Authority (NZa) & Tilburg University

January 09, 2020

The effect of Tilburg gold on your health

Misja Mikkers & Gertjan Verhoeven

Introduction

DAGS

and causality

onclusio





#### Introduction

The effect of Tilburg gold on your health

#### Misja Mikkers & Gertjan Verhoeven

#### Introduction

Causali

#### AGS

Machine learning and causality

#### Conclusion







# Example of a dataset dataset

The effect of Tilburg gold on your health

Misja Mikkers & Gertjan Verhoeven

Introduction

Caasan

AGS

Machine learning and causality

onclusion

Blog and code

χ_1	X_2	X_3	^_I	ı	ΥU	ΥI
Male	9	14	1	0	67	NA
Female	60	36	0	1	NA	113
Female	7	2	1	1	NA	54

**\**/0

1/1





### Prediction versus understanding



Misja Mikkers & Gertjan Verhoeven

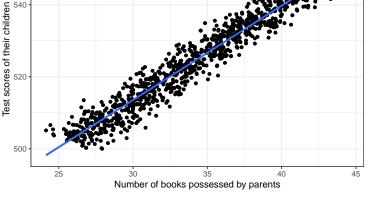
Introductio

Causality

AGS

Machine learning and causality

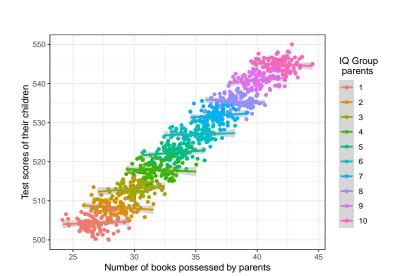
Conclusion







## Voorspellen versus begrijpen



The effect of Tilburg gold on your health

Misja Mikkers & Gertjan Verhoeven

Introduction

Causality

AGS

Machine learning and causality

Conclusion





# **DAG** building blocks

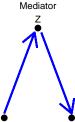
The effect of Tilburg gold on your health

Misja Mikkers & Gertjan Verhoeven

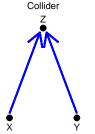
**DAGS** 

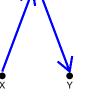
Machine learning and causality

Blog and code



Confounder









#### Confounder

The effect of Tilburg gold on your health

Misja Mikkers & Gertjan Verhoeven

Introduction

Causalit

DAGS

Machine learning and causality

Conclusion

Blog and code

TILBURG

IQ\_parents Number\_of\_books Test\_score\_children



#### **Collider**

The effect of Tilburg gold on your health

Misja Mikkers & Gertjan Verhoeven

ntroduction

Causalit

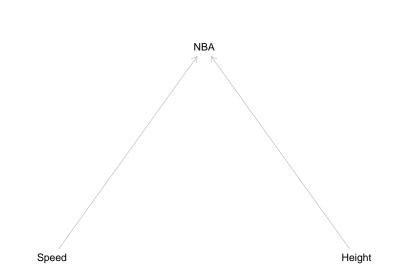
DAGS

Machine learning and causality

Conclusion

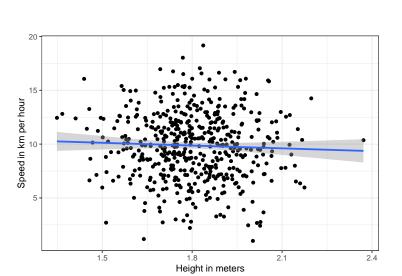
Blog and code

TILBURG UNIVERSITY





### Does height causes speed?



The effect of Tilburg gold on your health

Misja Mikkers & Gertjan Verhoeven

Introduction

Causali

DAGS

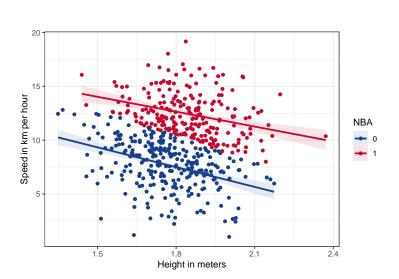
Machine learning and causality

Conclusion





### Does height causes speed?



The effect of Tilburg gold on your health

Misja Mikkers & Gertjan Verhoeven

Introduction

Causaii

DAGS

Machine learning and causality

Conclusion





#### Mediator

The effect of Tilburg gold on your health

Misja Mikkers & Gertjan Verhoeven

Introduction

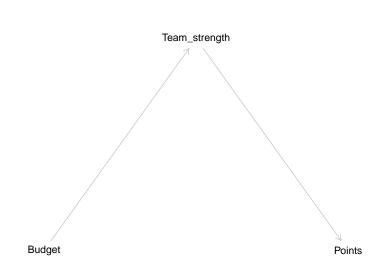
Causali

DAGS

Machine learning and causality

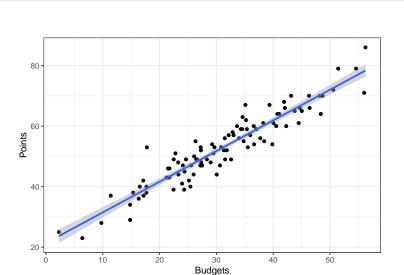
Conclusion







# **Budgets and points**



The effect of Tilburg gold on your health

Misja Mikkers & Gertjan Verhoeven

Introduction

Causali

DAGS

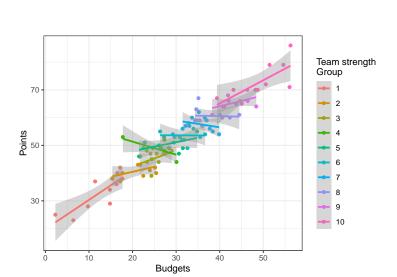
Machine learning and causality

Conclusion





### **Budgets and Points**



The effect of Tilburg gold on your health

Misja Mikkers & Gertjan Verhoeven

Introduction

ausanı

DAGS

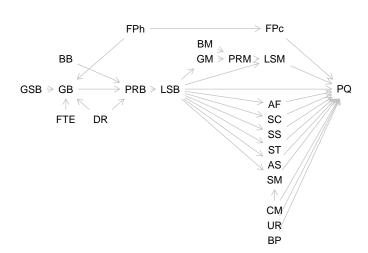
Machine learning and causality

Conclusion





### **Example of a more complex DAG**



The effect of Tilburg gold on your health

Misja Mikkers & Gertjan Verhoeven

Introduction

Caasan

DAGS

Machine learning and causality

Conclusion





# Machine learning and causality

The effect of Tilburg gold on your health

Misja Mikkers & Gertjan Verhoeven

Introduction

ausanty

AGS

Machine learning and causality

Conclusion

Blog and code

University

Health\_care\_cost Proportional\_shortfall Treatment



Blog and code

Misja Mikkers &

The variable *Proportional Shortfall* is based on:

Proportional Shortfall =  $abs(scale(V1^3 + 2 * V2 +$  $3 * V3^2 + 4 * V4 + 5 * V5$  $+6 * V6 * V7)) + \epsilon$ 

1. Fit Random Forest model on the data 2. Determine the average treatment effect with generalized random forests (grf)

We will fit 2 models for each of these steps:

a. An analysis with all variables ("the wrong model") b. An analysis with all variables, except the collider Health care cost ("the right model")





#### Misja Mikkers & Gertjan Verhoeven

Introduction

Lausality

**DAGS** 

Machine learning and causality

Conclus

Statistic	N	Mean	St. Dev.	Min	Max
Proportional Shortfall	1,000	1.114	0.859	0.003	4.778
Treatment	1,000	0.474	0.500	0	1
Health care cost	1,000	6.976	4.746	-1.578	27.557
V1	1,000	498.746	288.860	2	1,000
V2	1,000	492.975	293.667	1	1,000
V3	1,000	512.349	293.801	1	1,000
V4	1,000	489.041	288.225	1	1,000
V5	1,000	516.896	292.376	1	1,000
V6	1,000	497.232	284.475	1	1,000
V7	1,000	505.715	289.385	1	1,000
V8	1,000	501.330	276.913	3	998





#### **Predictions Random Forest**

The effect of Tilburg gold on your health

#### Misja Mikkers & Gertjan Verhoeven

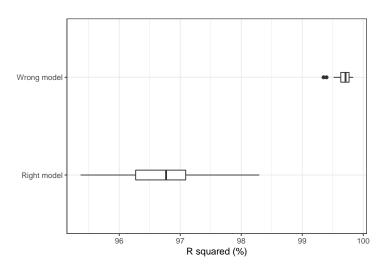
Introduction

ausanı

AGS

Machine learning and causality

Conclusion







## Average treatment effect





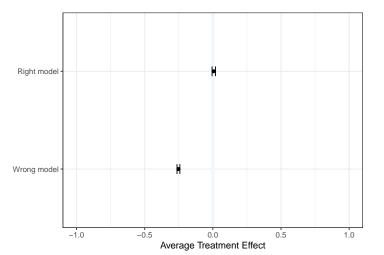
Introductio

Causani

DAGS

Machine learning and causality

Conclusion







#### Conclusion

The effect of Tilburg gold on your health

Misja Mikkers & Gertjan Verhoeven

and causality

Conclusion

- Causal models are necessary for inference
- It is tempting to use all variables in a machine learning model
- However, this could lead to misleading conclusions





# Blog and code (in Dutch)

The effect of Tilburg gold on your health

Misja Mikkers & Gertjan Verhoeven

and causality

Blog and code

https://misjamikkers.github.io/post/ causaliteit-en-machine-learning/

https://github.com/misjamikkers/Meetup\_Informatieberaad



