The effect of Tilburg gold on your health

Misja Mikkers & Gertjan Verhoeven

Introduction

Causality

DAGS

Machine learning and causality

Conclusion

Blog and code

The effect of Tilburg gold on your health

Pacmed meetup: Combining causal inference and machine learning in practice

Misja Mikkers & Gertjan Verhoeven

Dutch Healthcare Authority (NZa) & Tilburg University

December 11, 2019

TILBURG
UNIVERSITY

Understanding
Society

Nederlandse
Zorgautoriteit

Introduction

The effect of
Tilburg gold on
your health

**Misja Mikkers &
Gertjan
Verhoeven**

Example of a dataset dataset

The effect of
Tilburg gold on
your health

Misja Mikkers &
Gertjan
Verhoeven

Introduction

Causality

DAGS

Machine learning
and causality

Conclusion

Blog and code

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

TILBURG
UNIVERSITY

Understanding
Society

Nederlandse
Zorgautoriteit

Prediction versus understanding

The effect of
Tilburg gold on
your health

**Misja Mikkers &
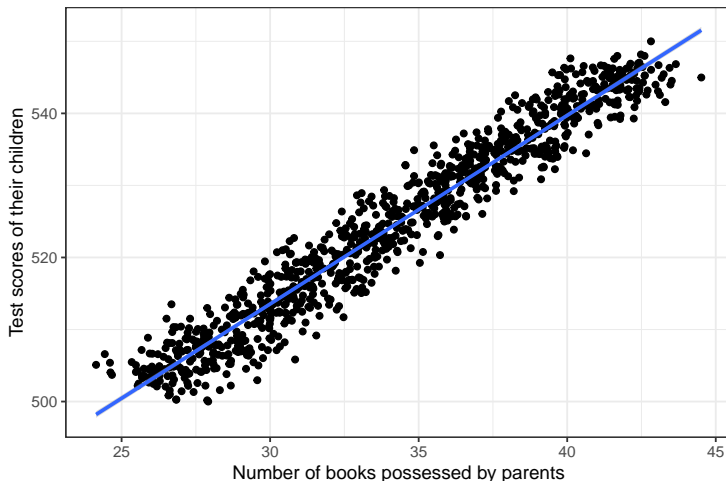Gertjan
Verhoeven**

Introduction

Causality

DAGS

Machine learning
and causality

Conclusion

Blog and code

Voorspellen versus begrijpen

The effect of
Tilburg gold on
your health

**Misja Mikkers &
Gertjan
Verhoeven**

TILBURG
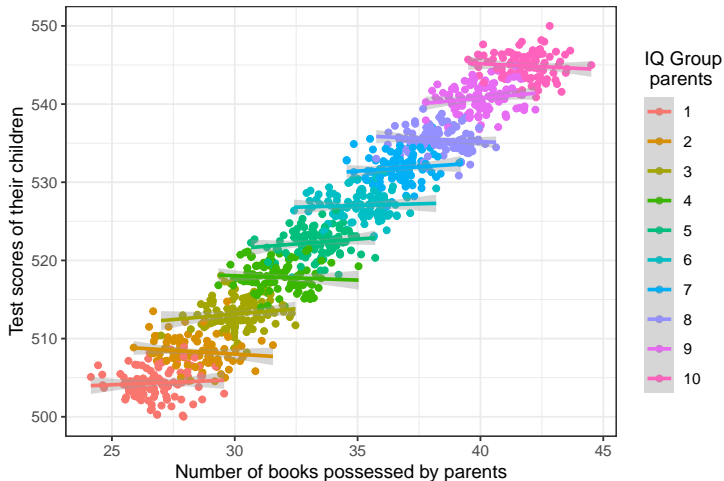UNIVERSITY

Understanding
Society

Nederlandse
Zorgautoriteit

DAG building blocks

The effect of
Tilburg gold on
your health

Misja Mikkers &
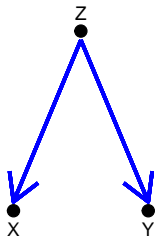Gertjan
Verhoeven

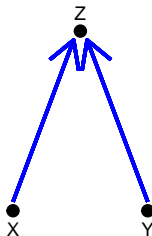Introduction

Causality

DAGS

Machine learning
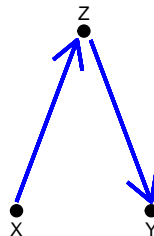and causality

Conclusion

Blog and code

Confounder

Collider

Mediator

Confounder

The effect of
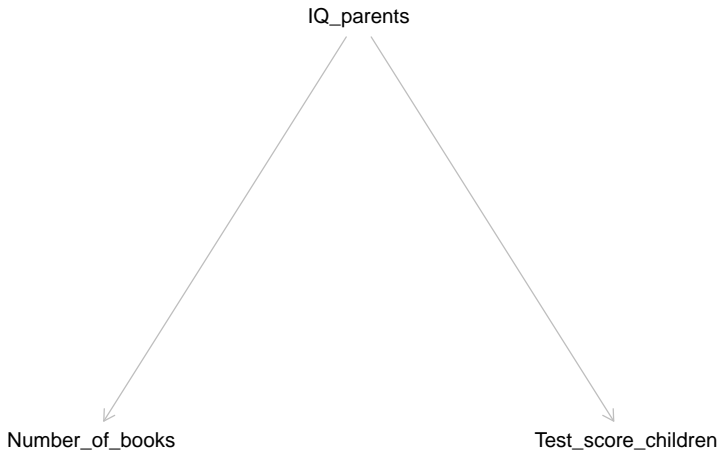Tilburg gold on
your health

Misja Mikkers &
Gertjan
Verhoeven

IQ_parents

Number_of_books   Test_score_children

Collider

The effect of
Tilburg gold on
your health

Misja Mikkers &
Gertjan
Verhoeven

Introduction

Causality

DAGS

Machine learning
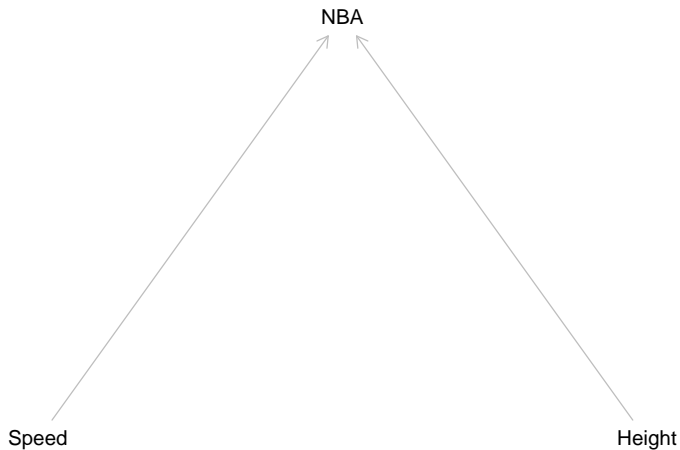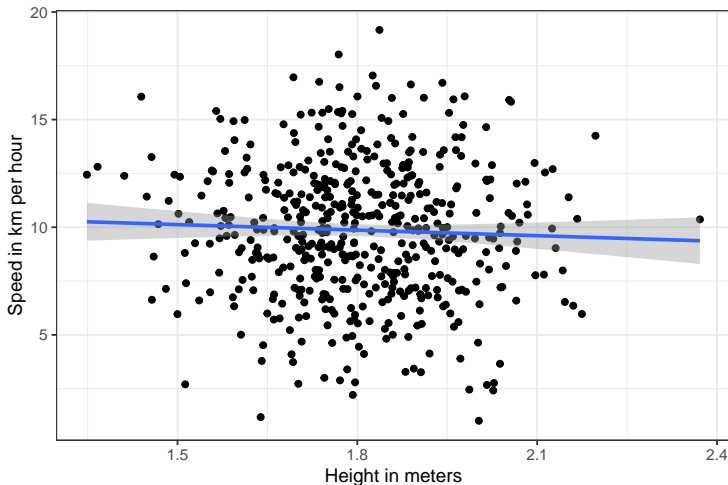and causality

Conclusion

Blog and code

NBA

Speed   Height

Does height causes speed?

The effect of
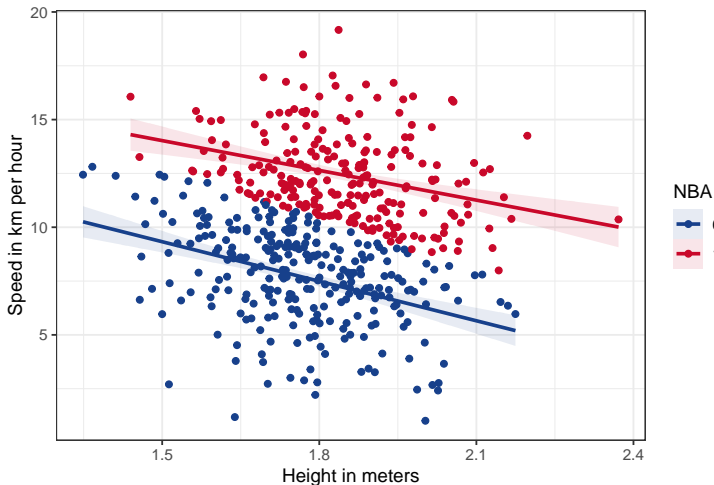Tilburg gold on
your health

Misja Mikkers &
Gertjan
Verhoeven

TILBURG
UNIVERSITY

Understanding
Society

Nederlandse
Zorgautoriteit

Does height causes speed?

The effect of
Tilburg gold on
your health

**Misja Mikkers &
Gertjan
Verhoeven**

TILBURG
UNIVERSITY

Understanding
Society

Nederlandse
Zorgautoriteit

Mediator

The effect of
Tilburg gold on
your health

**Misja Mikkers &
Gertjan
Verhoeven**

Budgets and points

The effect of
Tilburg gold on
your health

Misja Mikkers &
Gertjan
Verhoeven

Introduction

Causality

DAGS

Machine learning
and causality

Conclusion

Blog and code

TILBURG
UNIVERSITY

Understanding
Society

Budgets and Points

The effect of
Tilburg gold on
your health

Misja Mikkers &
Gertjan
Verhoeven

Introduction

Causality

DAGS

Machine learning
and causality

Conclusion

Blog and code

TILBURG
UNIVERSITY

Understanding
Society

Nederlandse
Zorgautoriteit

Example of a more complex DAG

The effect of
Tilburg gold on
your health

Misja Mikkers &
Gertjan
Verhoeven

Introduction

Causality

DAGS

Machine learning
and causality

Conclusion

Blog and code

Machine learning and causality

The effect of
Tilburg gold on
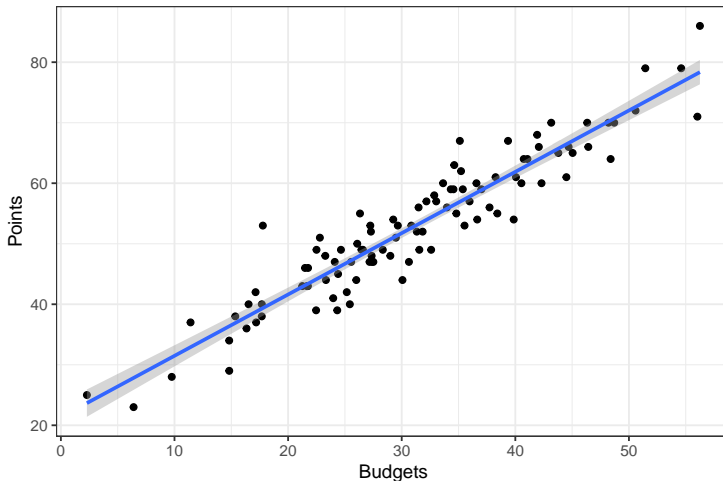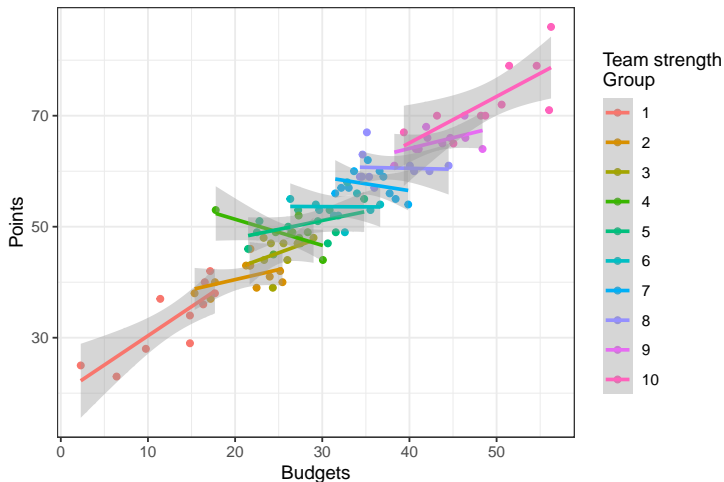your health

Misja Mikkers &
Gertjan
Verhoeven

Approach

The effect of
Tilburg gold on
your health

Misja Mikkers &
Gertjan
Verhoeven

Introduction

Causality

DAGS

Machine learning
and causality
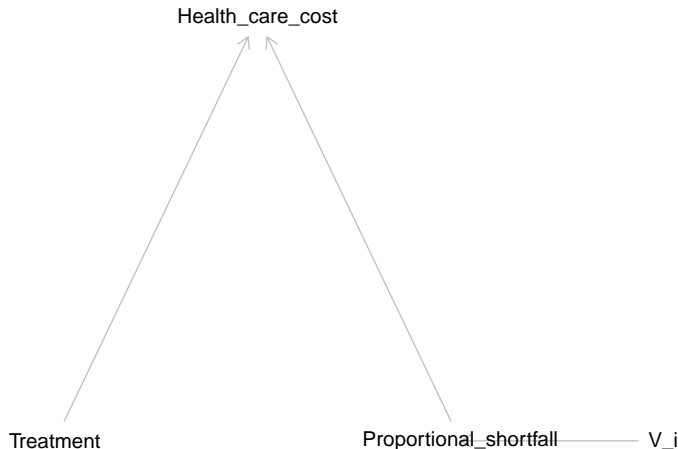
Conclusion

Blog and code

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")

Summary statistics

The effect of
Tilburg gold on
your health

Misja Mikkers &
Gertjan
Verhoeven

Introduction

Causality

DAGS

Machine learning
and causality

Conclusion

Blog and code

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

TILBURG
UNIVERSITY

Understanding
Society

Nederlandse
Zorgautoriteit

Predictions Random Forest

The effect of
Tilburg gold on
your health

Misja Mikkers &
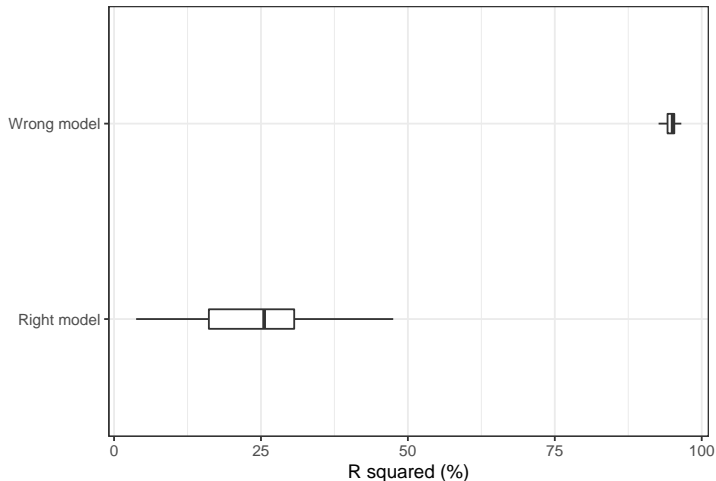Gertjan
Verhoeven

Introduction

Causality

DAGS

Machine learning
and causality

Conclusion

Blog and code

Average treatment effect

The effect of
Tilburg gold on
your health

**Misja Mikkers &
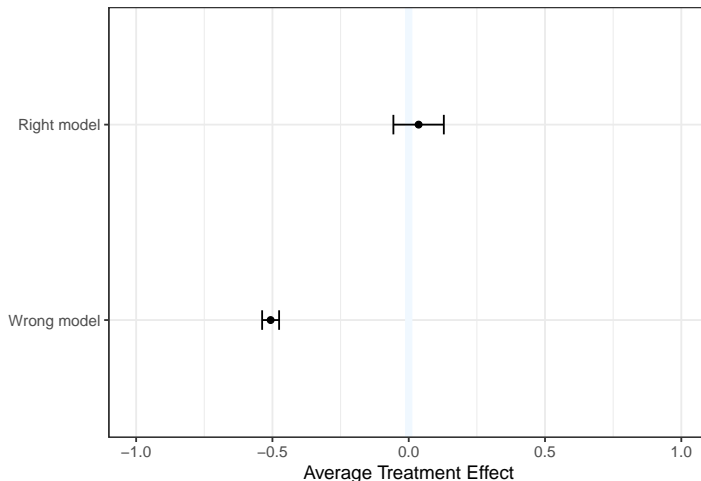Gertjan
Verhoeven**

Introduction

Causality

DAGS

**Machine learning
and causality**

Conclusion

Blog and code

Conclusion

The effect of
Tilburg gold on
your health

Misja Mikkers &
Gertjan
Verhoeven

- ▶ 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

TILBURG
UNIVERSITY

Understanding
Society

Nederlandse
Zorgautoriteit

Blog and code (in Dutch)

The effect of
Tilburg gold on
your health

Misja Mikkers &
Gertjan
Verhoeven

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

https://github.com/misjamikkers/Meetup_Informatieberaad

Nederlandse
Zorgautoriteit

TILBURG
UNIVERSITY

Understanding
Society