



INNOvember

powered by
Rijks Innovatie Community



Open the black box



Open the black box



Joke
De Boer

CJIB

Facilitator



Willy
Tadema

CJIB

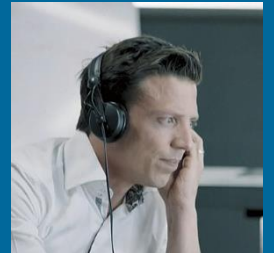
XAI
intro



Gerben
Kooistra

CJIB

XAI
verdieping



Selmar
Smit

TNO

Fairness



Huishoudelijke mededelingen



Vragen in de chat



Audio op “Mute”

- Wanneer je aan het woord bent, stel jezelf even voor (naam, organisatie en functie).
- Heb je bandbreedte problemen? Zet je video dan uit.
- Doet je geluid het niet? Probeer dan even opnieuw in te loggen.
- De presentaties worden nagestuurd.



Stelling: Een algoritme moet volledig transparant zijn

- Eens,
- Oneens,

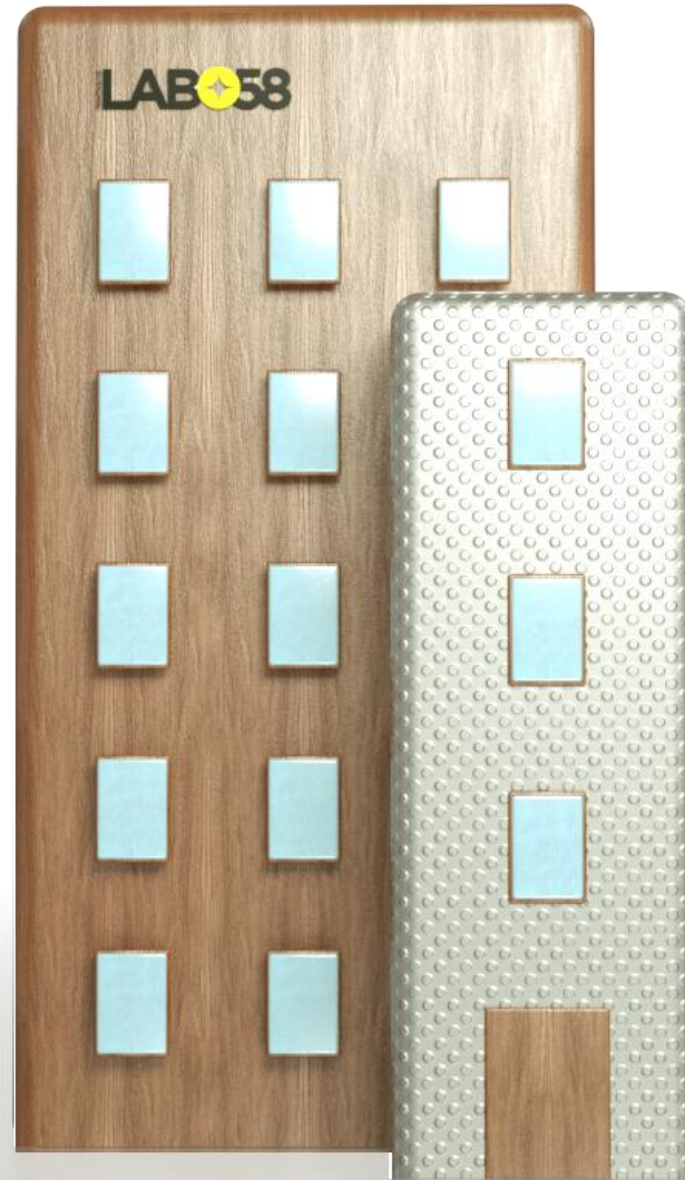


INNOvember

Spreker: Willy Tadema

powered by
Rijks Innovatie Community

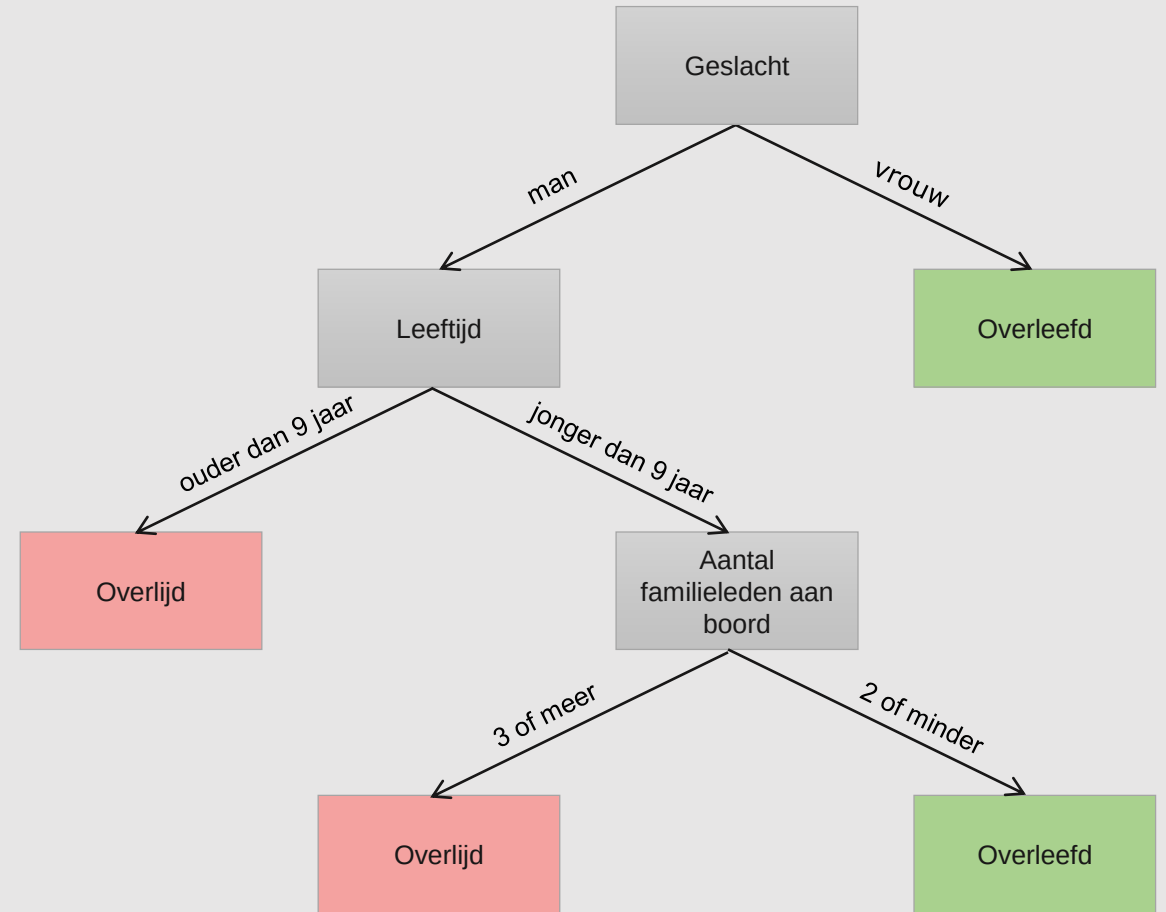
INNOVATIE **LAB58**





Eenvoudige ML algoritmen

- Regressie of decision trees
- Technisch transparant
- Goed uitlegbaar



Voorbeeld van een decision tree voor de kans op overleving aan boord van de Titanic

Transparantie & Uitlegbaarheid

Belangrijk voor:

- Detecteren van fouten
- Opsporen van ongewenste discriminatie
- Acceptatie van het model
- Informatie waarop burger kan acteren
- 'A right to explanation'

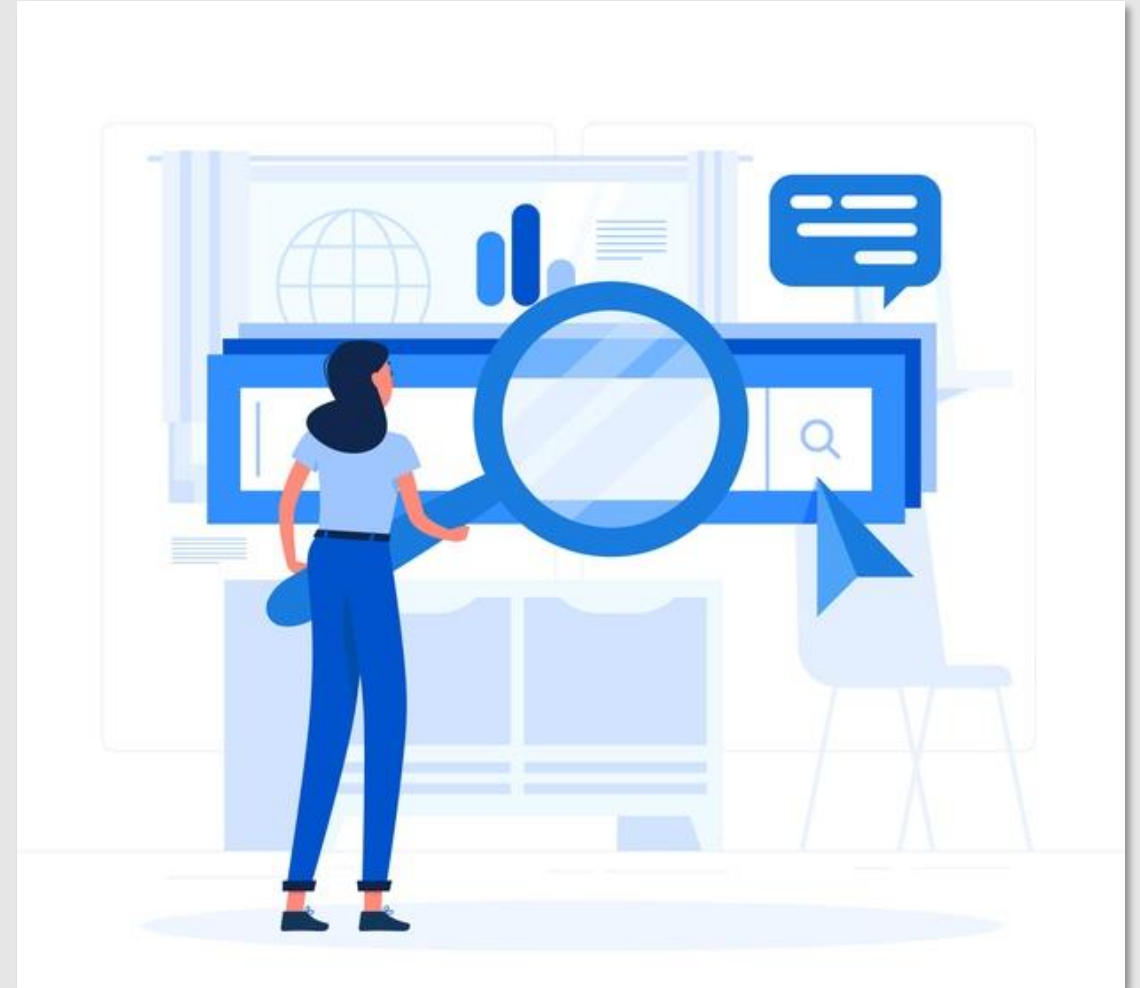
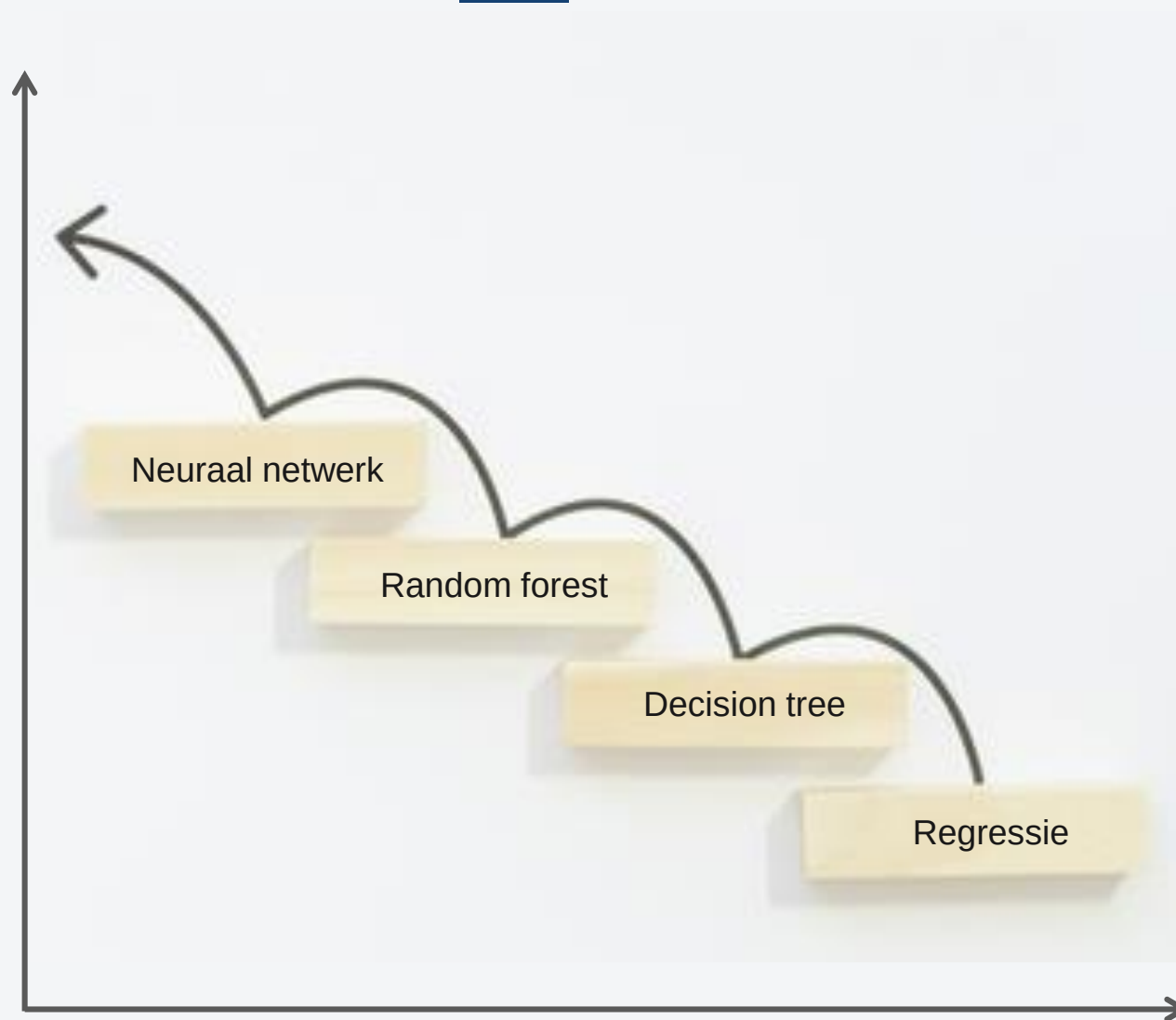


image: Freepik.com

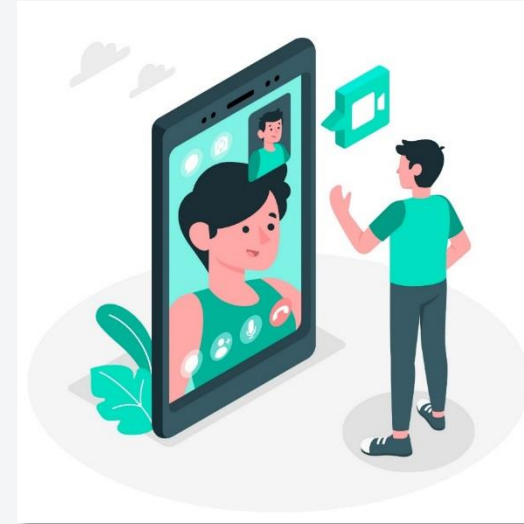


Meer complexiteit +
betere prestaties



Meer transparantie + betere uitlegbaarheid

Wat doe je bij de vraag voor één miljoen?



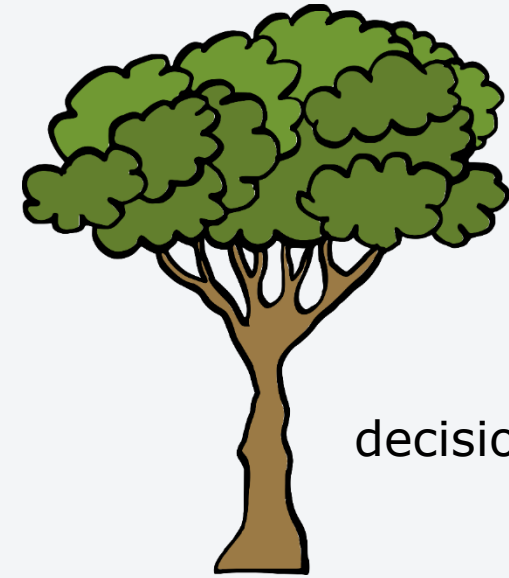
Bel je een vriend...



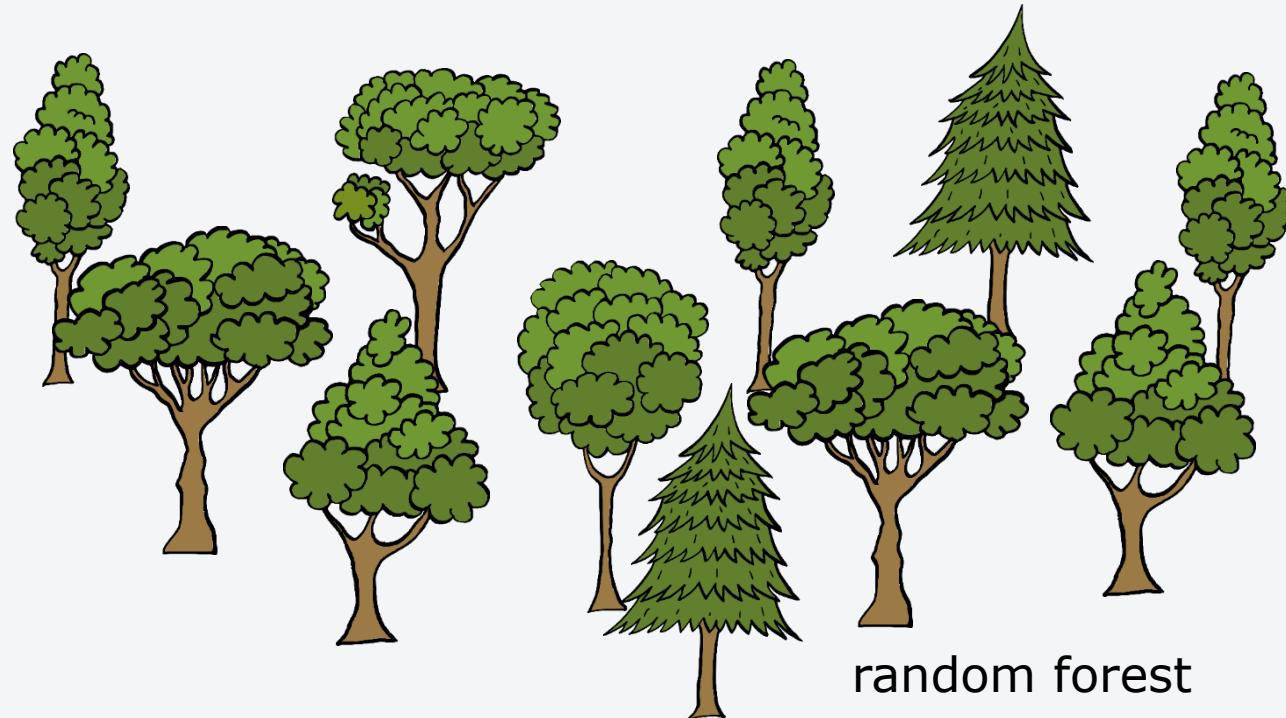
...of laat je het publiek stemmen?

Black box model: random forest

- Bel een vriend = decision tree
- Laat het publiek stemmen = random forest
- Random forest presteert beter, maar is niet transparant



decision tree



random forest



Open the black box

- Technische transparantie = interne werking
→ auditeerbaarheid
- Uitlegbaarheid = interpreteerbaarheid
→ methoden voor eXplainable AI (XAI)
- Afgestemd met juristen

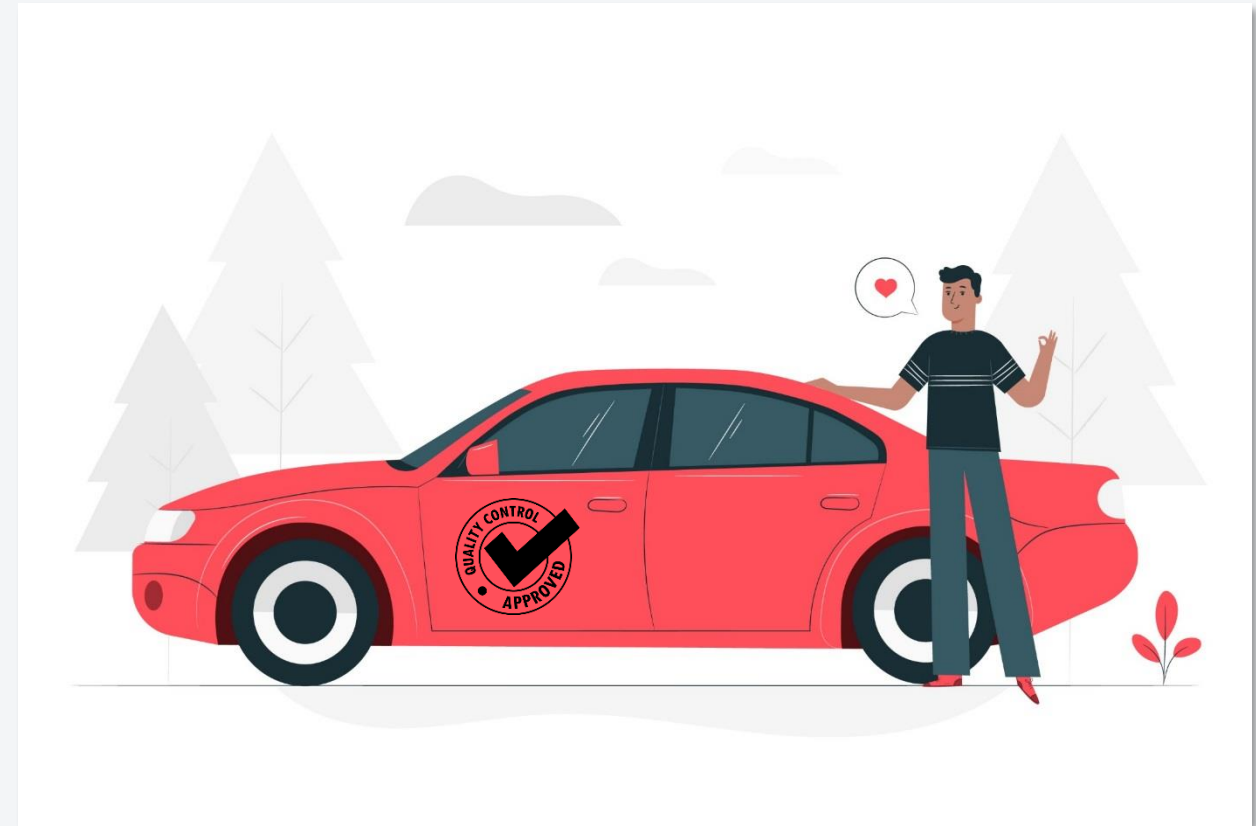


image: Freepik.com



INNOvember

powered by
Rijks Innovatie Community



Wie had de meeste kans om de ramp te overleven?



INNOvember

Spreker: Gerben Kooistra

powered by
Rijks Innovatie Community

Waarom eerst modellen zoals Random Forest?

- Meeste winst mee te behalen; vervolgstap naar Deep Learning.
- Verschillende onderzoeken wijzen uit dat soortgelijke modellen goed presteren voor voorspellen menselijk gedrag.
- Tot nu toe ook gebleken bij onderzoeken op eigen data.



De Titanic dataset

- Kunnen we voorspellen wie de scheepsramp gaat overleven?
- Kunnen we onze voorspelling uitleggen?
- Vertaling naar tegenwoordige situaties.



P. Id	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.25		S
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs)	female	38	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2.	7.925		S

Wat is voor ons “Uitlegbaarheid”?

- Uitlegbaarheid = Interpreteerbaarheid
- Interpreteerbaarheid op:
 - Overall niveau (Business) - Hoe gedraagt model zich?
 - Predictie niveau (Burger) - Hoe kom je bij de voorspelling?



Model Agnostische methodes

- Focus op 'Model Agnostische methodes'.
- Op elk model toepasbaar.
- Benader elk model als 'Black Box'; geen assumptie op interne structuur.
- Behulp van de 'Dalex package/library'. Beschikbaar in Python en R.



Hoe maak je een model voor de Burger inzichtelijk?

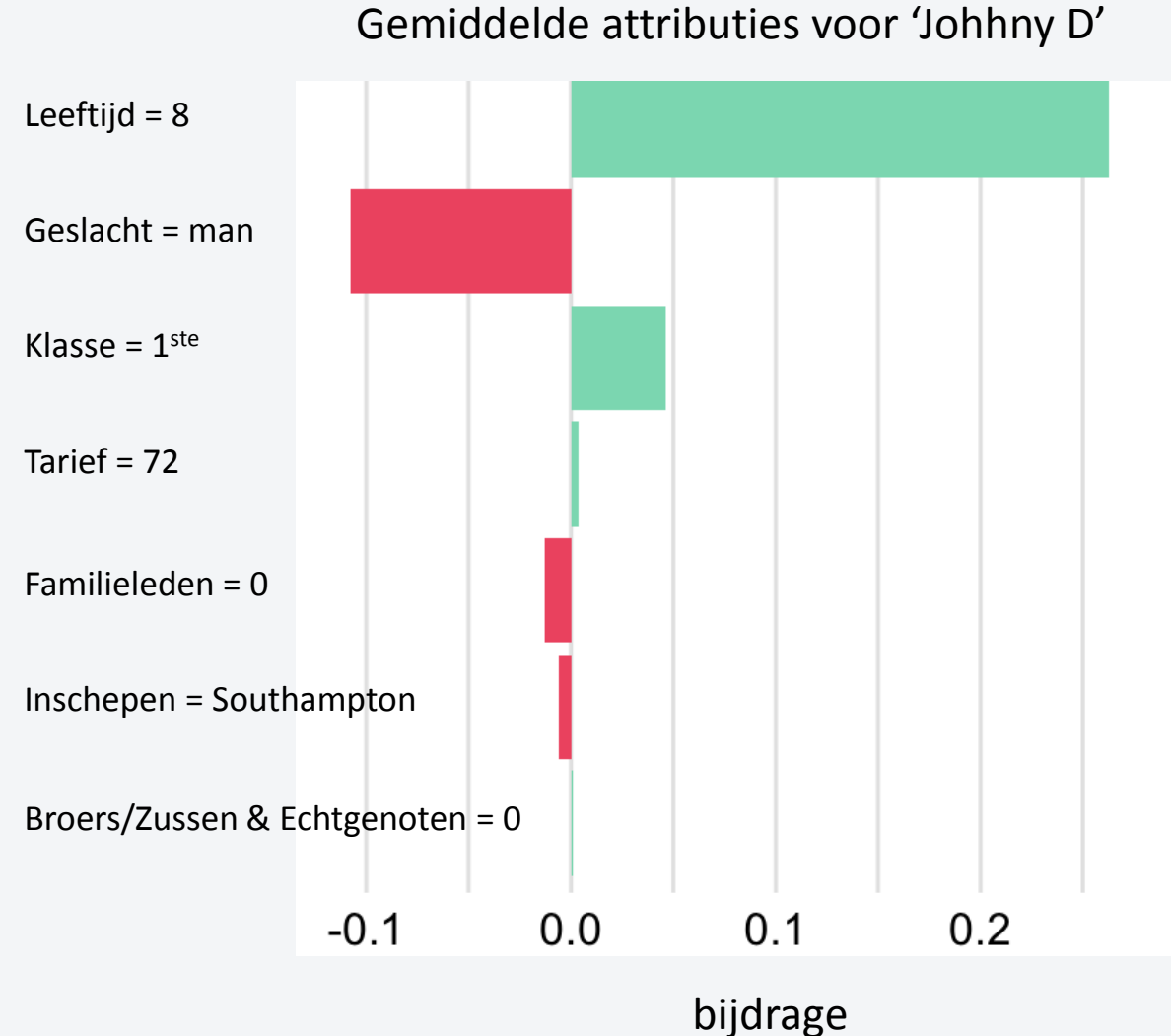
- Hoe (= op welke manier) voorspelt het model individueel?
- Belangrijk voor de burger
- Technieken waarmee dit bijvoorbeeld kan:
 - Shapley waarden
 - 'What-if' analyse





Inzichtelijk maken van de invloed van elke variabele

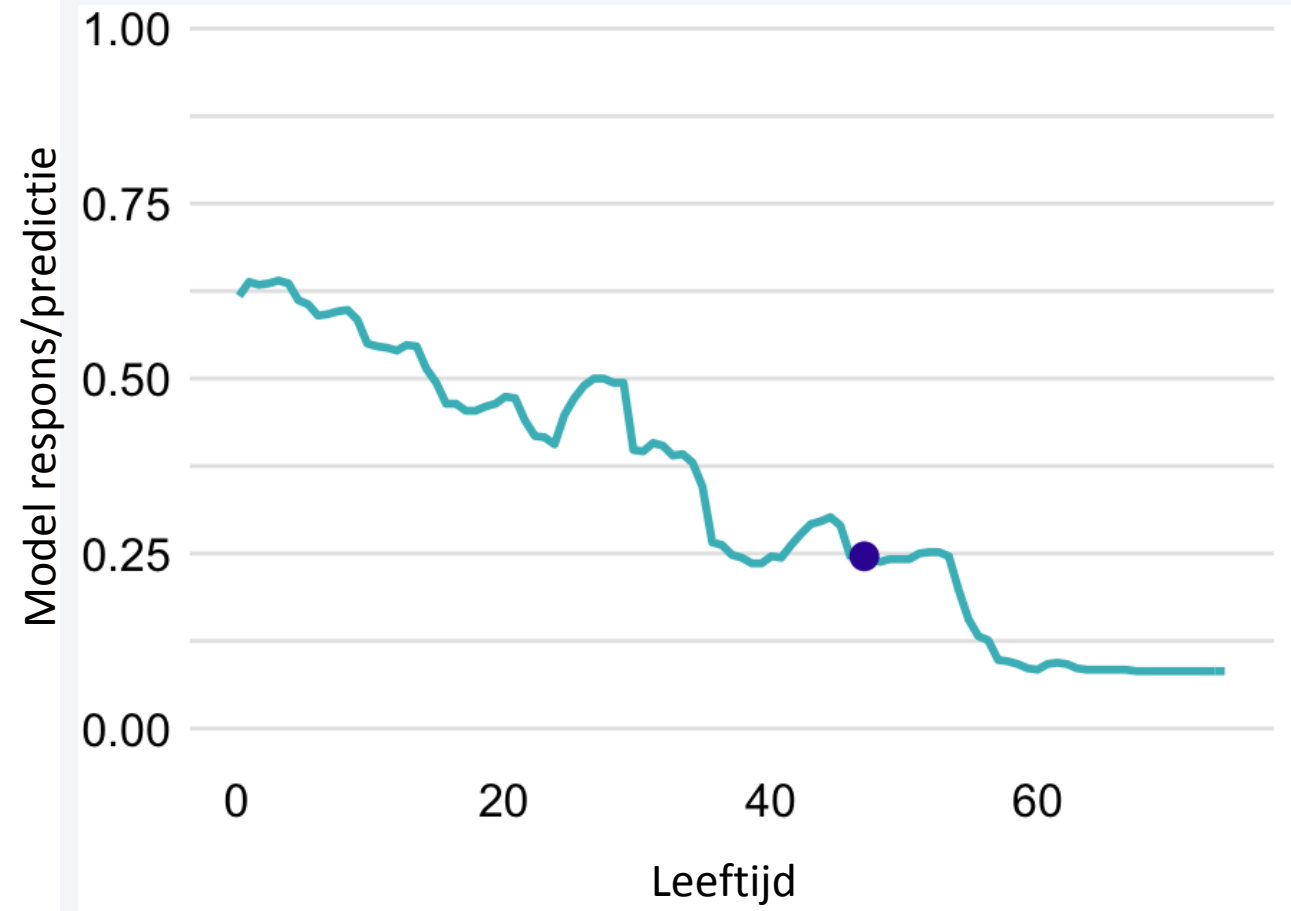
- Met een Shapley plot wordt weergegeven hoeveel invloed elke variabele heeft op de voorspelling.
- Burger krijgt inzicht in welke eigenschappen belangrijkst zijn bij voorspelling



Wat als de waarde van variabele verandert

- What-if analyse
- Een burger krijgt inzicht in hoe bijv. verandering in gedrag de voorspelling beïnvloed.

Ceteris Paribus voor 'Titanic Random Forest model'



Hoe maak je model voor de Business inzichtelijk?

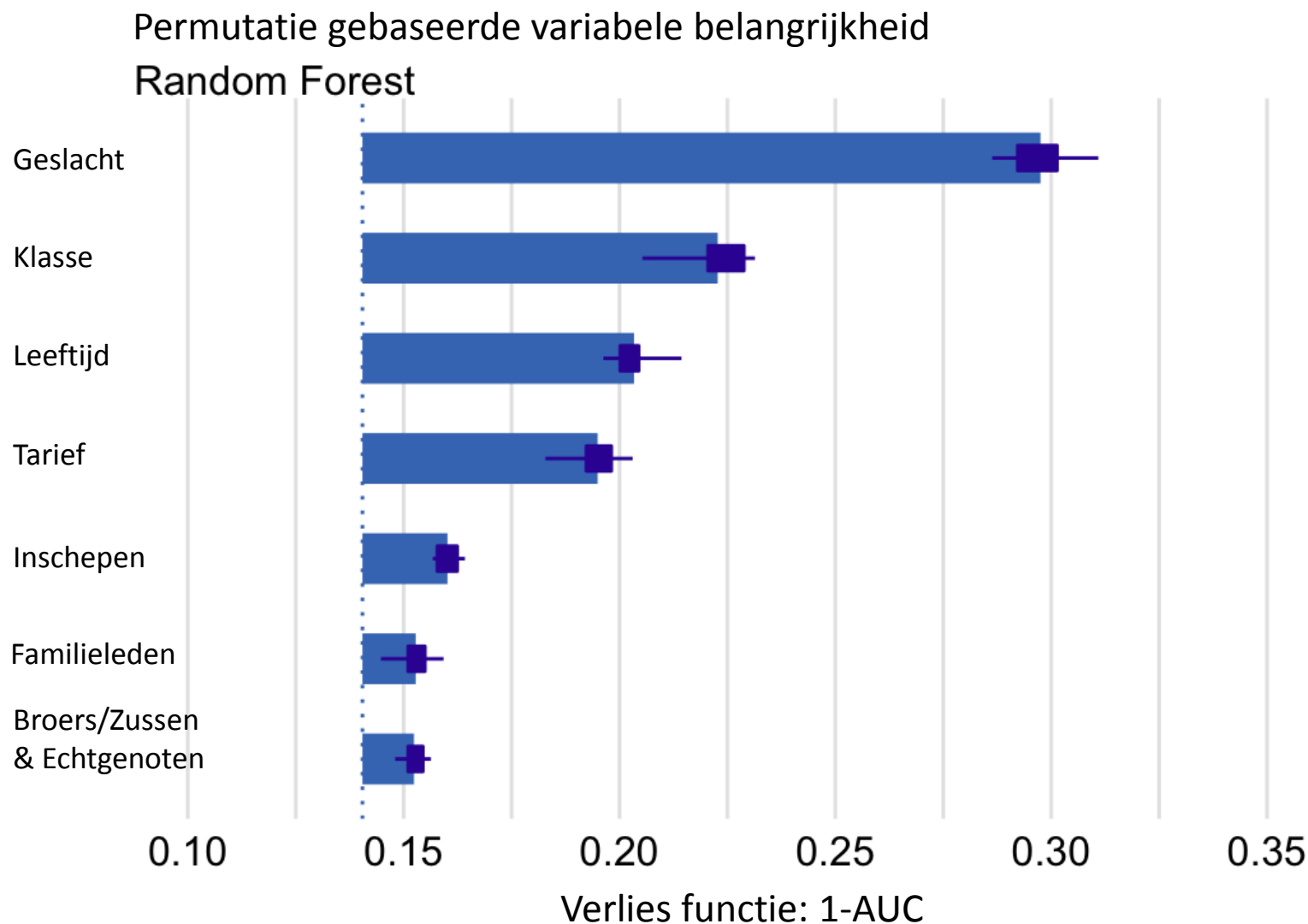
- Hoe (= op welke manier) werkt/voorspelt het model overall?
- Belangrijk voor de Business
 - Nieuwe inzichten verkrijgen.
 - Weten wat beperkingen van model zijn.
- Technieken waarmee dit bijvoorbeeld kan:
 - Belangrijkheid variabelen
 - Partial Dependence profielen





Hoe belangrijk is elke variabele?

- Plot laat zien hoe belangrijk een variabele is voor de correcte voorspelling van het model.
- Geeft business inzicht in belangrijkste voorspellers.

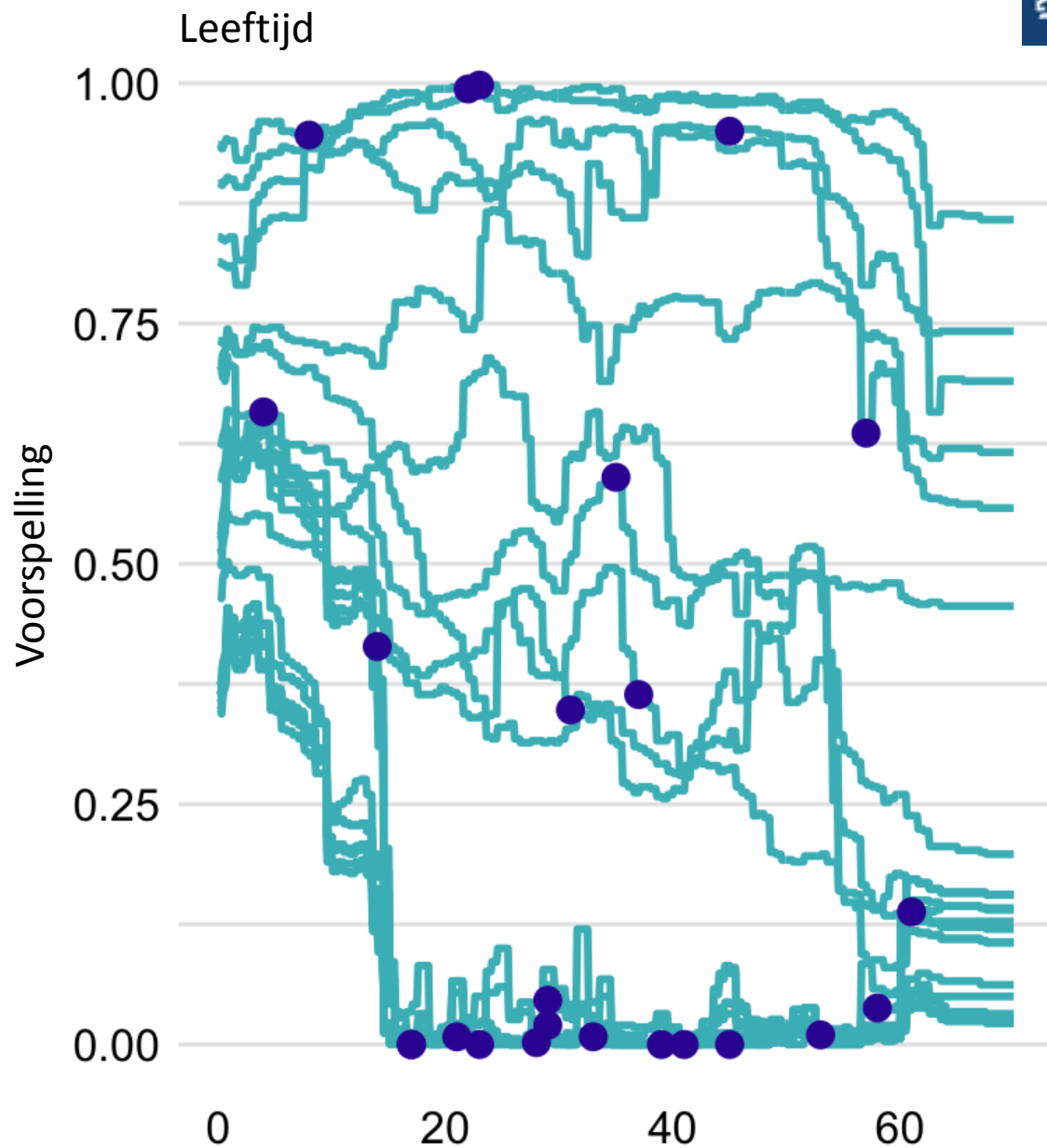


Partial Dependence profiel

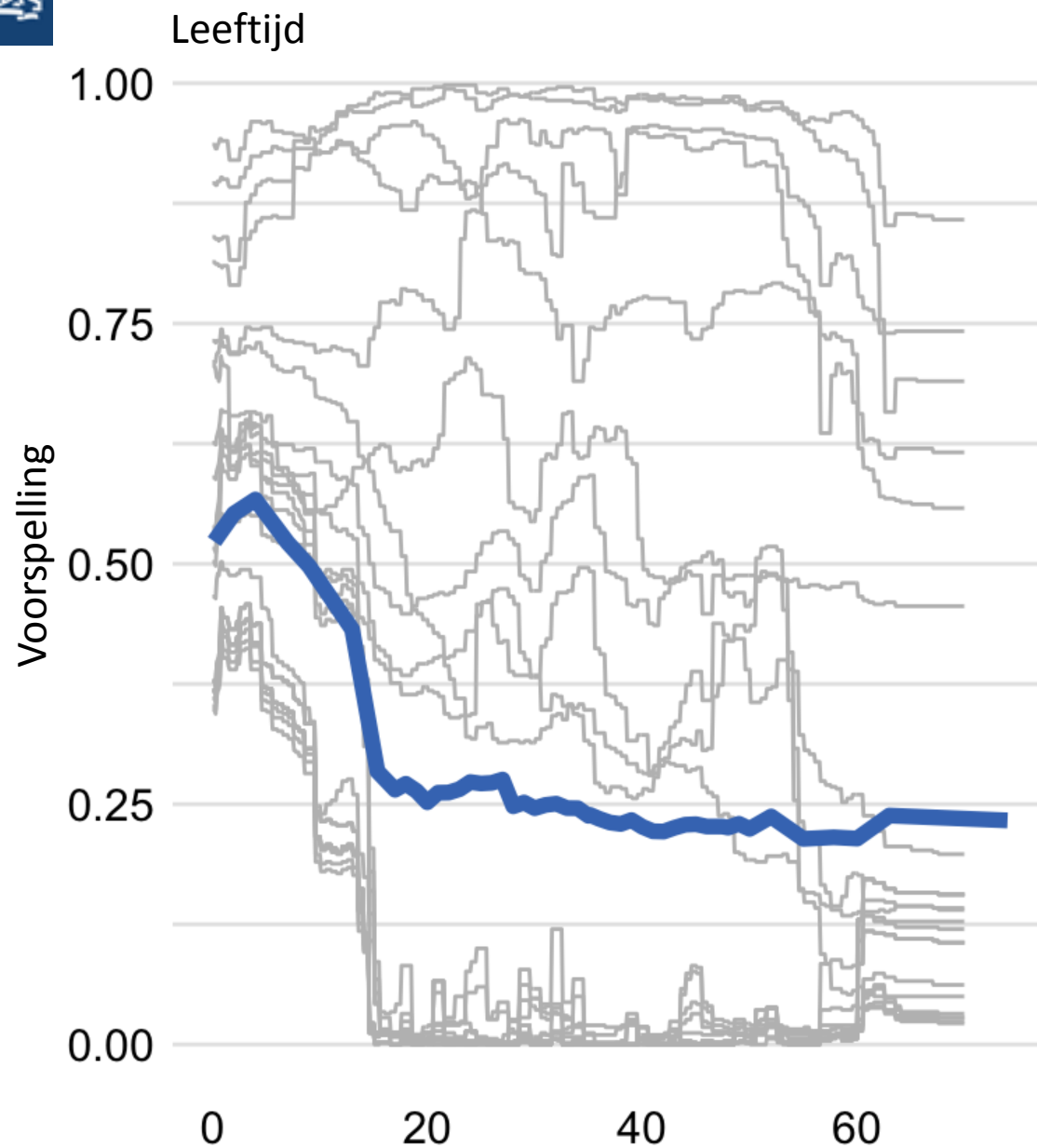
- Een Partial Dependence profiel is niks anders dan een gemiddelde Ceteris-paribus ('What-if') profiel van een variabele .
- Laat zien hoe de uitkomst van een model zich als functie gedraagt bij een specifieke variabele.
- Laat zien hoe stabiel/instabiel een model is.
- Partial Dependence profielen zijn zeer handig om modellen te vergelijken.



Ceteris-Paribus profiel ('What-if' analyse)



Partial-dependence profiel



Meer weten?

- Interessante boeken:
 - [Explanatory Model Analysis - Przemyslaw Biecek and Tomasz Burzykowski](#)
 - [Interpretable Machine Learning - Christoph Molnar](#)
- Wij geven workshops vanaf januari 2021 en helpen jou en je collega's (en je organisatie) graag verder op weg met XAI.
 - PostbusInnovatielab058@cjib.nl





INNOvember

powered by
Rijks Innovatie Community



Stelling: Algoritmes mogen niet discrimineren.

- Eens,
- Oneens,



INNOvember

Spreker: Selmar Smit

powered by
Rijks Innovatie Community

FAIRNESS

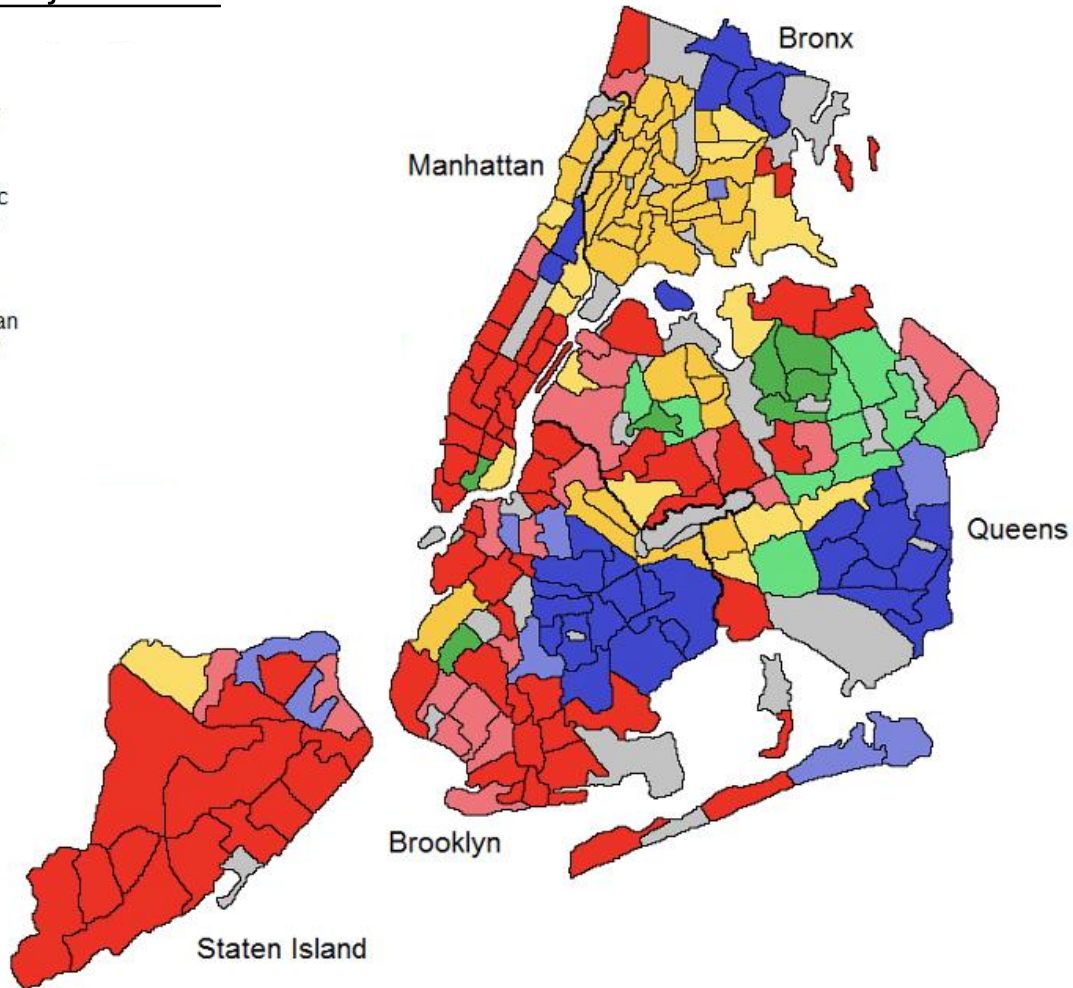
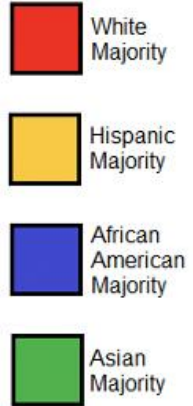
◆ INNOvember

TNO

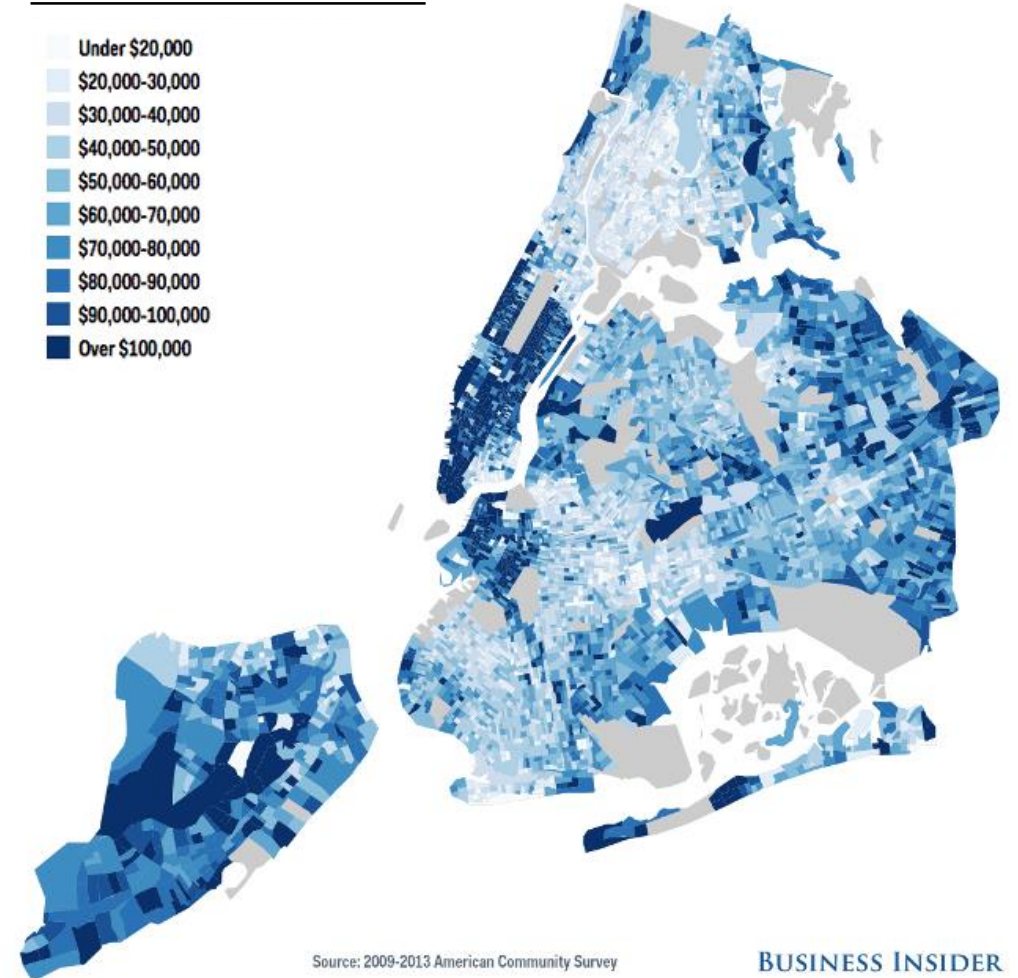
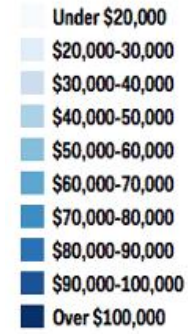
Selmar smit

selmar.smit@tno.nl

Race and Ethnicity



Median Household Income

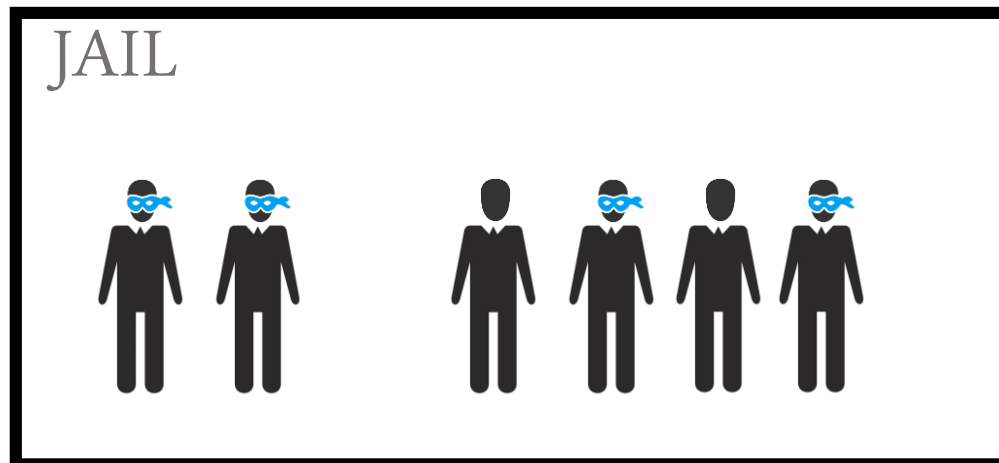


Source: 2009-2013 American Community Survey

BUSINESS INSIDER

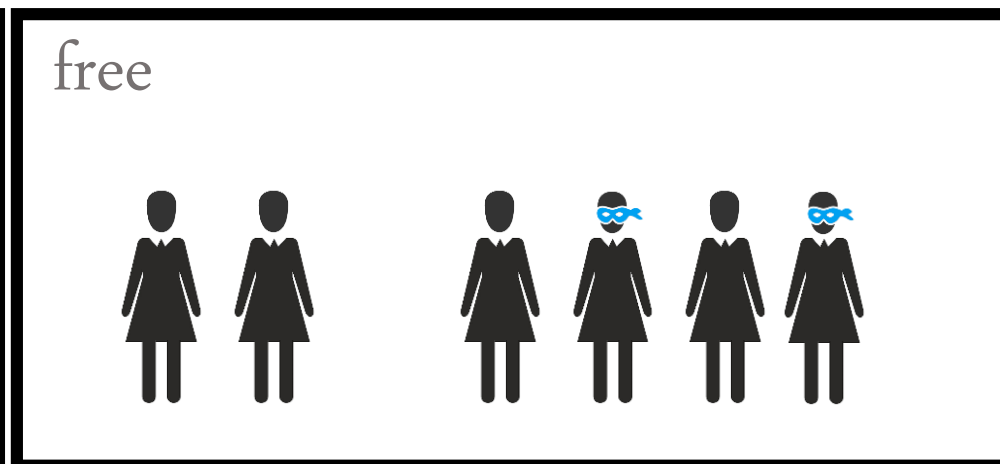
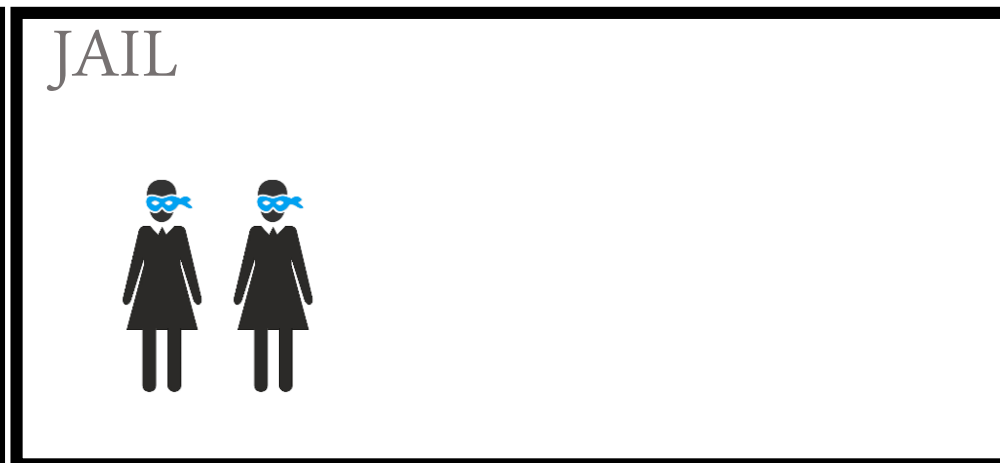
men

accuracy 75%



women

accuracy 75%



Compas data

predicted risk of recidivism versus 4 year numbers

Caucasian			Total
High Risk	81	195	276
Low Risk	1139	461	1600

African-American			Total
High Risk	284	741	1025
Low Risk	990	532	1522

Accuracy

predicted risk of recidivism versus 4 year numbers

Caucasian			Total
High Risk	81	195	276
Low Risk	1139	461	1600

71%

African-American			Total
High Risk	284	741	1025
Low Risk	990	532	1522

68%

High Risk Percentage

predicted risk of recidivism versus 4 year numbers

Caucasian			Total
High Risk	81	195	276
Low Risk	1139	461	1600

15%

African-American			Total
High Risk	284	741	1025
Low Risk	990	532	1522

40%

High Risk when Wrong

predicted risk of recidivism versus 4 year numbers

Caucasian			Total
High Risk	81	195	276
Low Risk	1139	461	1600

15%

African-American			Total
High Risk	284	741	1025
Low Risk	990	532	1522

35%

High Risk with no Reoffending

predicted risk of recidivism versus 4 year numbers

Caucasian			Total
High Risk	81	195	276
Low Risk	1139	461	1600

7%

African-American			Total
High Risk	284	741	1025
Low Risk	990	532	1522

22%

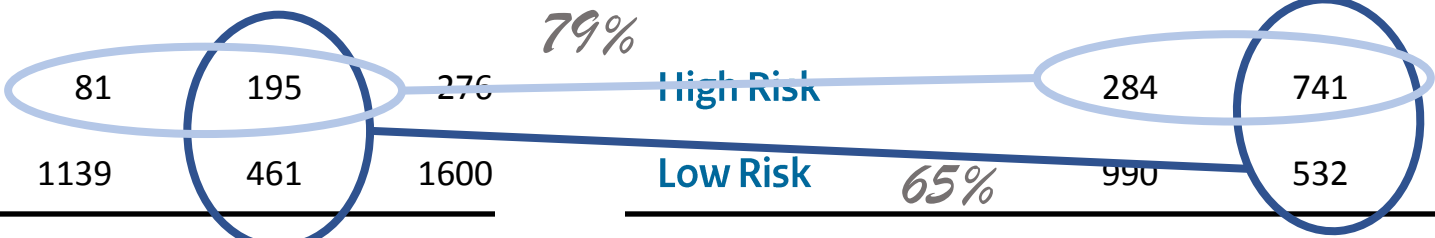
High Risk vs. Reoffending

predicted risk of recidivism versus 4 year numbers

Caucasian			African-American		
			Total		Total
High Risk	81	195	276	High Risk	1025
Low Risk	1139	461	1600	Low Risk	1522

79%

65%



Reoffending with Low Risk

predicted risk of recidivism versus 4 year numbers

Caucasian			Total
High Risk	81	195	276
Low Risk	1139	461	1600
			28%

African-American			Total
High Risk	284	741	1025
Low Risk	990	532	1522
			35%

Not Reoffending with High Risk

predicted risk of recidivism versus 4 year numbers

Caucasian			Total
			30%
High Risk	81	195	276
Low Risk	1139	461	1600

African-American			Total
			27%
High Risk	284	741	1025
Low Risk	990	532	1522

Not Reoffending with High Risk

predicted risk of recidivism versus 4 year numbers

Caucasian			Total
High Risk	81	195	276
Low Risk	1139	461	1600

30%

African-American			Total
High Risk	284	741	1025
Low Risk	990	532	1522

27%

predicted risk of recidivism versus 4 year numbers z

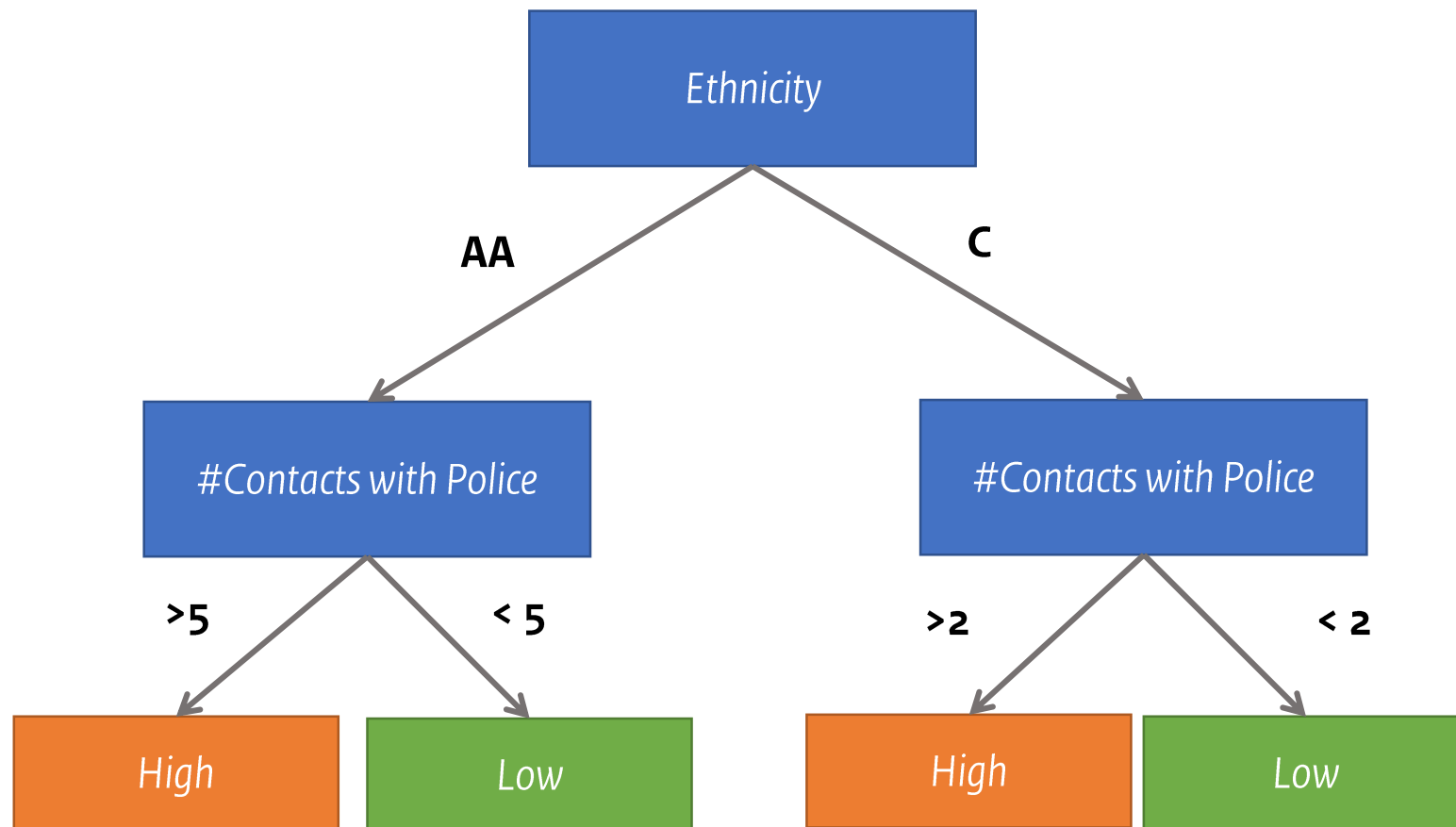
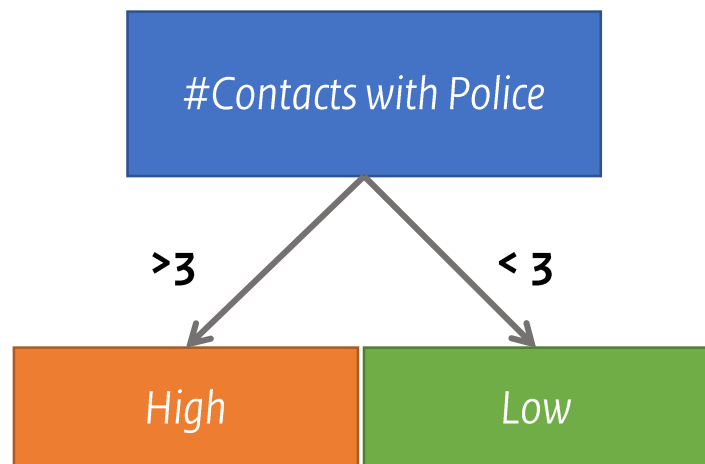
Caucasian			Total
High Risk	81	195	276
Low Risk	1139	461	1600

7%

African-American			Total
High Risk	284	741	1025
Low Risk	990	532	1522

22%

High Risk with no Reoffending



Not Reoffending with High Risk

predicted risk of recidivism versus 4 year numbers

Caucasian			Total
High Risk	148	231	361
Low Risk	1072	460	1532

39%

African-American			Total
High Risk	155	612	767
Low Risk	1119	532	1651

20%

predicted risk of recidivism versus 4 year numbers z

Caucasian			Total
High Risk	148	231	361
Low Risk	1072	460	1532

12%

African-American			Total
High Risk	155	612	767
Low Risk	1119	532	1651

12%

High Risk with no Reoffending

*there is no such thing as fairness
only acceptable* levels of unfairness*



INNOvember

powered by
Rijks Innovatie Community



Stelling: Een algoritme moet volledig transparant zijn

- Eens,
- Oneens,



Stelling: Algoritmes mogen niet discrimineren.

- Eens,
- Oneens,



www.INNOvember.nl

powered by
Rijks Innovatie Community