

AI-precisie met klassieke uitlegbaarheid

Explainable boosting machines (EBM)



9 september 2025
Triple A - Risk Finance

SPREKER



Robert Mooijer

2020 - heden
Senior consultant
Non-Life Insurance

Agenda

1

Introductie

2

Klassiek vs A.I.

3

Theorie: EBM en interacties

4

Interactie detectie met FAST-algoritme

5

Conclusie

6

Afsluiting

INTRODUCTIE

Voorspellende modellen

- Cruciaal in de verzekeringswereld.
- O.a. premiestelling, assumpties en fraudedetectie.

Opmars A.I. modellen.

Voordelen:

- Sterkere **voorspelkracht** t.o.v. klassieke modellen.
- Detectie van patronen die mensen niet kunnen zien.
- Simpele toepassing.

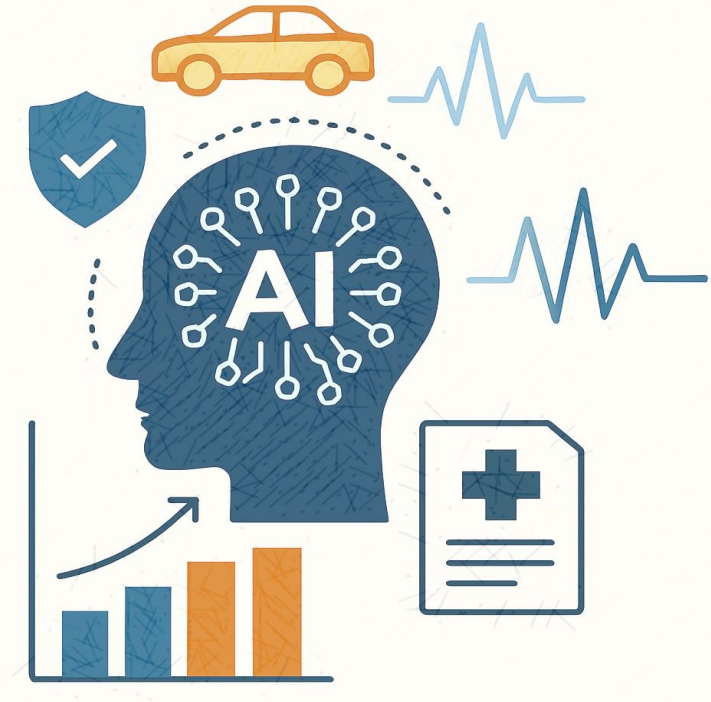
Nadelen:

- Beperkte **uitlegbaarheid**.
- Statistische onderbouwing ontbreekt.
- Expert judgement lastig.

Vandaag:

- Slimmere 'klassieke modellen' die de voorspelkracht van A.I. modellen benaderen, zonder de nadelen.

➤ Explainable boosting machines



MODELLEN

		Model	Uitlegbaarheid	Voorspelkracht	Overfitting	Non-Lineariteit	Onderhoud
Klassiek		OLS (Ordinary Least Squares)	✓✓✓	✗	✓✓✓	✗✗✗	✓✓✓
		GLM (Generalized Linear model)	✓✓✓	✓	✓✓	✗✗	✓✓
		GAM (Generalized Additive Model)	✓✓	✓✓	✗	✓✓	✓
A.I.		GBM (Gradient Boosting Machines)	✗	✓✓✓	✗✗	✓✓✓	✗✗
		Neural Networks	✗✗✗	✓✓✓	✗✗✗	✓✓✓	✗✗✗



THEORIE (EBM)

- Startpunt:
 - Bestaand model
 - GAM/GLM
 - Verklarende variabelen bekend
- Doel:
 - Hogere voorspelkracht
 - Klassieke context
- Hoe:
 - Toevoegen 2D interacties

Basis model:

$$g(E[y]) = \beta_0 + \sum f_i(x_i)$$



Intercept



Individuele
componenten
(1^e orde)

Toevoeging interactie effecten:

$$= \beta_0 + \sum f_i(x_i) + \sum f_{i,j}(x_i, x_j)$$



Interactie
componenten
(2^e orde)

THEORIE (INTERACTIES)



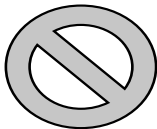
Realistischer modellering
Variabelen beïnvloeden elkaar



Verbeterde voorspelkracht
Betere fit en minder bias



Inzichten in risicoprofielen
Combinaties van kenmerken tonen verborgen risico's



Voorkomt onderschatting
Risico's bij specifieke klantgroepen

THEORIE (INTERACTIES)



Leeftijd bestuurder × Autotype

Jonge bestuurders in een sportauto hebben een **veel hoger risico** dan je zou verwachten op basis van leeftijd en autotype afzonderlijk.



Woningtype × Locatie

Appartementen in stedelijke gebieden kunnen **lager** brandrisico hebben door betere brandpreventie dan appartementen op het platteland.



Leeftijd polishouder × Gezinsgrootte

Een persoon met kinderen (beginnend bestuurders) heeft over het algemeen een verhoogd risico t.o.v. een persoon zonder.



INTERACTIE DETECTIE

$$g(E[y]) = \beta_0 + \sum f_i(x_i) + \boxed{\sum f_{i,j}(x_i, x_j)}$$

Opties:

- Fit voor elke mogelijke interactie
- FAST¹

INTERACTIE DETECTIE

Basismodel

Variabelen:

LeeftijdBestuurder
AutoType
GezinsGrootte

Verbetering?

Indien nieuwe model verbetering oplevert in RSS.



FAST algoritme

Rangschik op basis van hoogste voorspelkracht:

1. LeeftijdBestuurder x AutoType
2. LeeftijdBestuurder x GezinsGrootte
3. AutoType x GezinsGrootte

Modelleer 'beste' interactie in basismodel

LeeftijdBestuurder x AutoType

INTERACTIE DETECTIE

Basismodel

Variabelen:

LeeftijdBestuurder

AutoType

GezinsGrootte

LeeftijdBestuurder x AutoType

Verbetering?

Indien nieuwe model verbetering oplevert in RSS.



FAST algoritme

Rangschik op basis van hoogste voorspelkracht:

1. LeeftijdBestuurder x AutoType
2. LeeftijdBestuurder x GezinsGrootte
3. AutoType x GezinsGrootte

Modelleer 'beste' interactie in basismodel

LeeftijdBestuurder x AutoType

INTERACTIE DETECTIE

Basismodel

Variabelen:

LeeftijdBestuurder

AutoType

GezinsGrootte

LeeftijdBestuurder x AutoType

Verbetering?

Indien nieuwe model verbetering oplevert in RSS.



FAST algoritme

Rangschik op basis van hoogste voorspelkracht:

1. ~~LeeftijdBestuurder x AutoType~~
1. LeeftijdBestuurder x GezinsGrootte
2. AutoType x GezinsGrootte

Modelleer 'beste' interactie in basismodel

LeeftijdBestuurder x AutoType

INTERACTIE DETECTIE

Basismodel

Variabelen:

LeeftijdBestuurder

AutoType

GezinsGrootte

LeeftijdBestuurder x AutoType

Verbetering?

Indien nieuwe model verbetering oplevert in RSS.



FAST algoritme

Rangschik op basis van hoogste voorspelkracht:

1. ~~LeeftijdBestuurder x AutoType~~
1. LeeftijdBestuurder x GezinsGrootte
2. AutoType x GezinsGrootte

Modelleer 'beste' interactie in basismodel

LeeftijdBestuurder x AutoType
LeeftijdBestuurder x GezinsGrootte

INTERACTIE DETECTIE

Basismodel

Variabelen:

LeeftijdBestuurder

AutoType

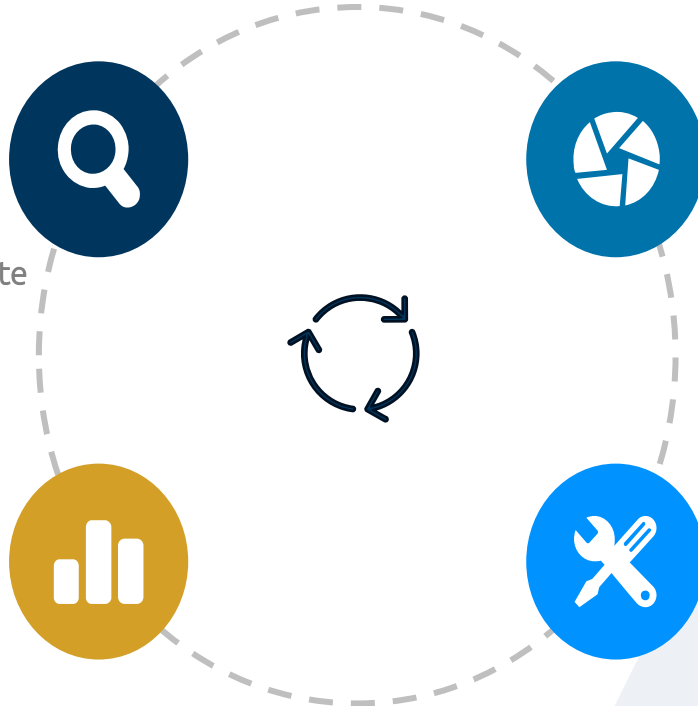
GezinsGrootte

LeeftijdBestuurder x AutoType

LeeftijdBestuurder x GezinsGrootte

Verbetering?

Indien nieuwe model verbetering oplevert in RSS.



FAST algoritme

Rangschik op basis van hoogste voorspelkracht:

1. ~~LeeftijdBestuurder x AutoType~~
1. LeeftijdBestuurder x GezinsGrootte
2. AutoType x GezinsGrootte

Modelleer 'beste' interactie in basismodel

LeeftijdBestuurder x AutoType
LeeftijdBestuurder x GezinsGrootte

FAST



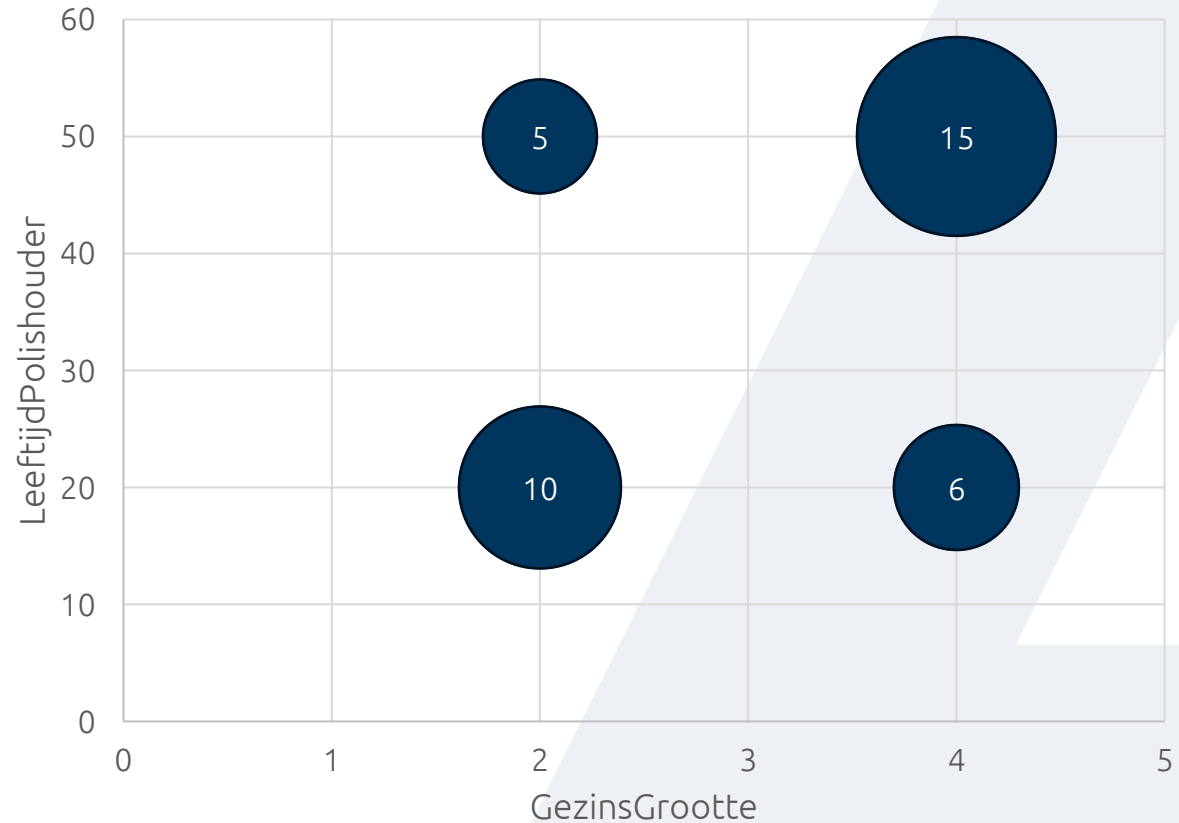
Features: 2



$N = 4$



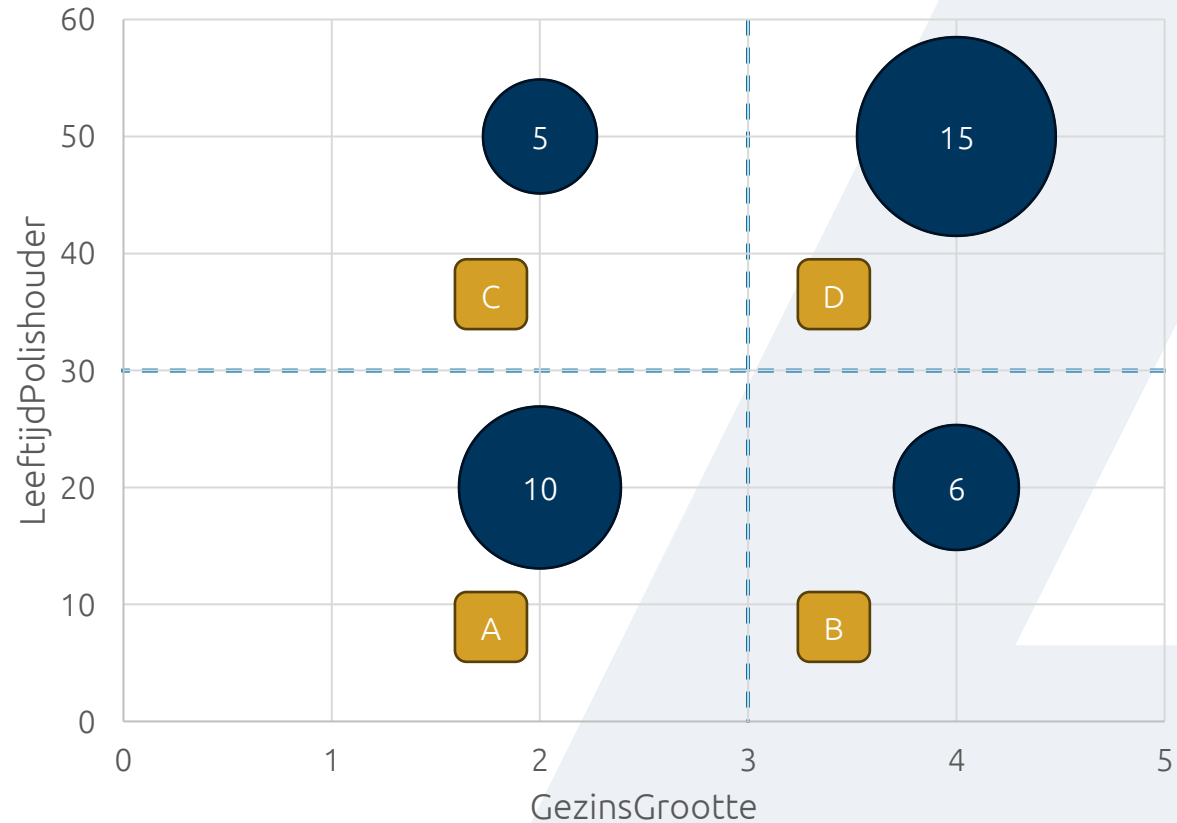
Doelvariabele: r



FAST

Categoriseer in groepen

- 2 per variabele



FAST

Categoriseer in groepen

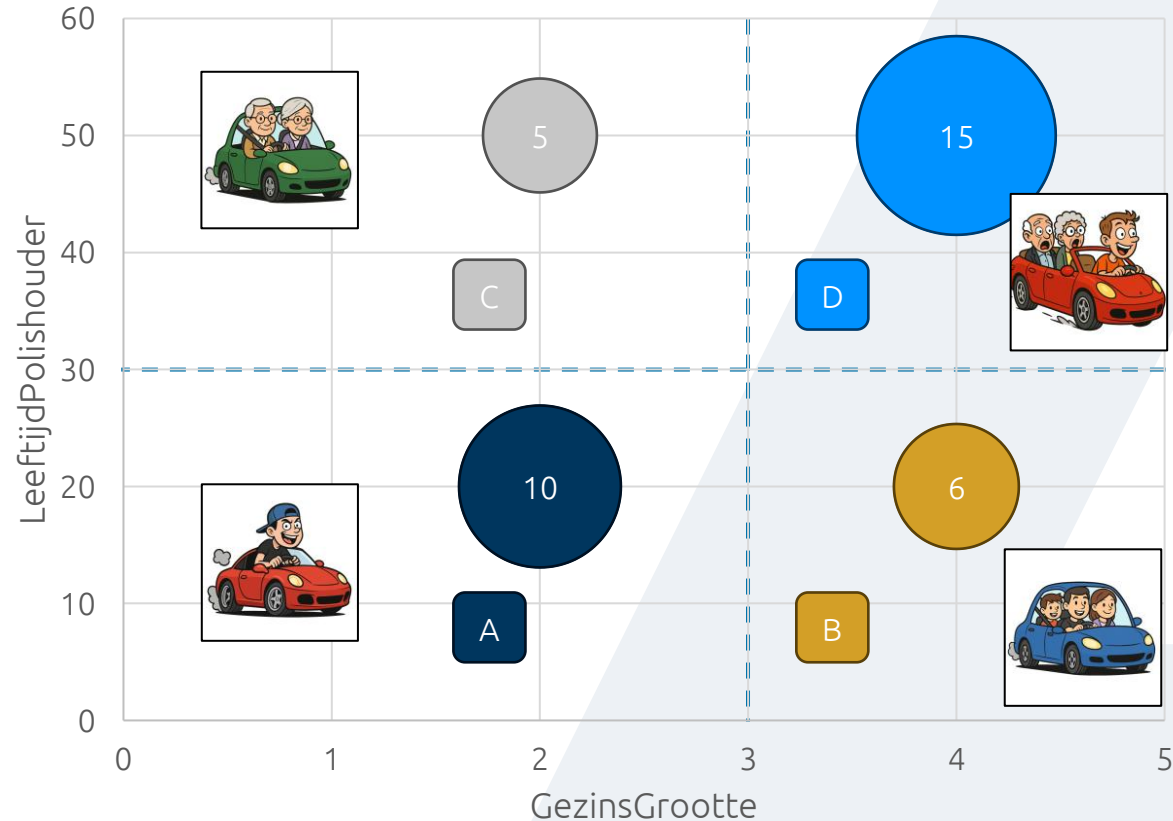
- 2 per variabele

Met kinderen:

- Leeftijd 50 > Leeftijd 20

Zonder kinderen:

- Leeftijd 20 > Leeftijd 50



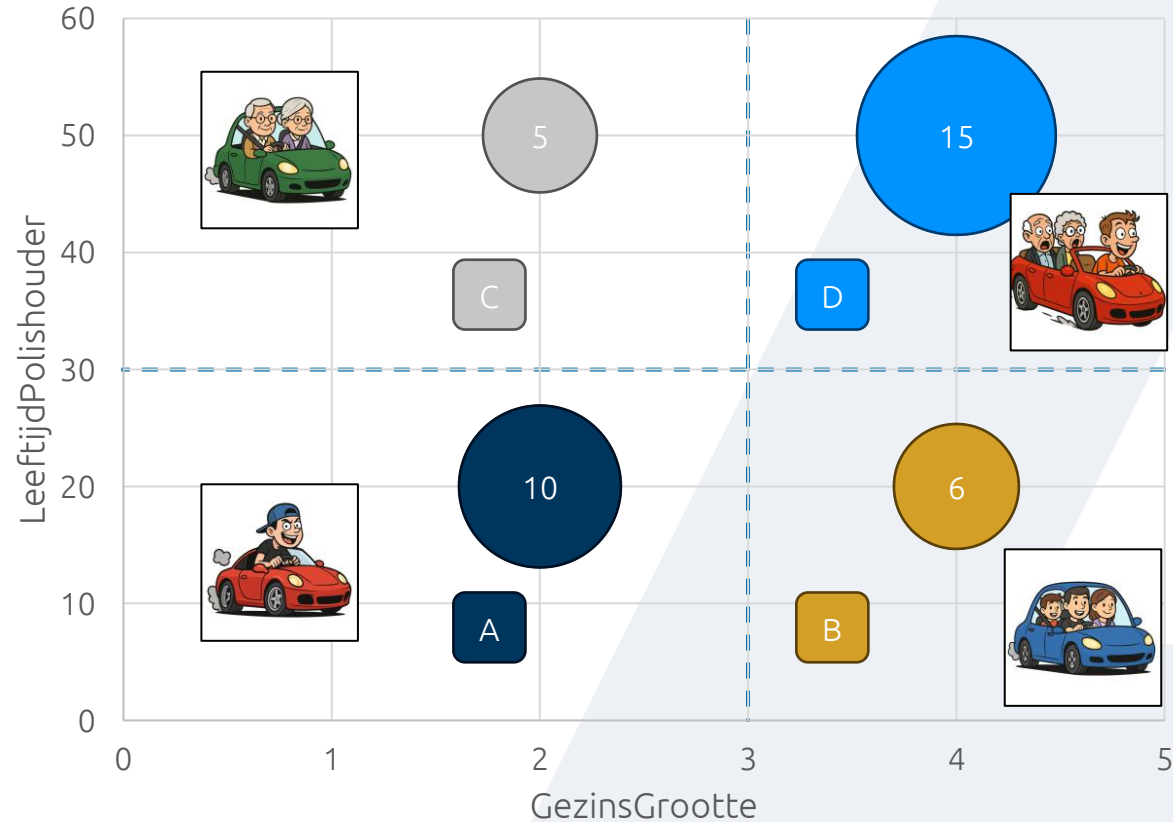
FAST

Loss functie (RSS):

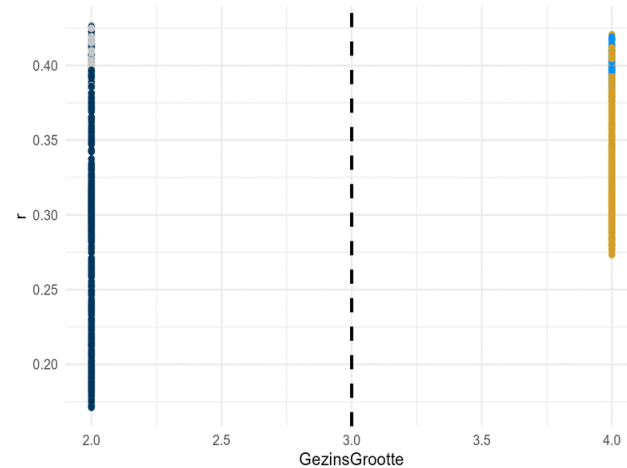
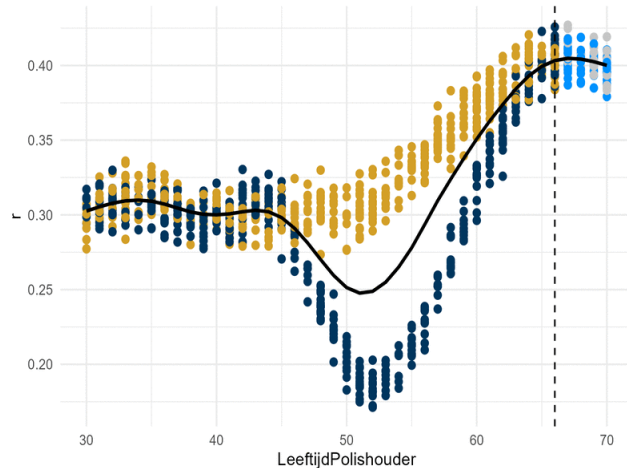
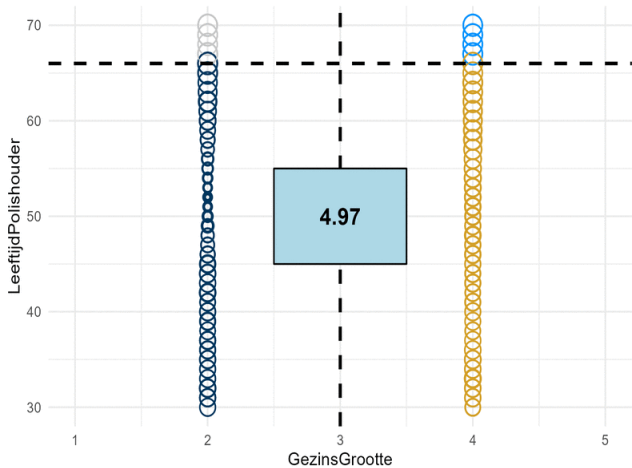
$$\sum (y - f(\text{Leeftijd}, \text{Gezinsgrootte}))^2$$

Voorspelling ($\hat{y}$):

$$f(\text{Leeftijd}, \text{Gezinsgrootte}) = \text{Gem}(\text{kwadrant})$$



FAST



Jong/geen kinderen

Ouder/geen kinderen

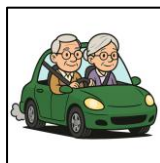
Jong/wel kinderen

Ouder/wel kinderen

A



C



B

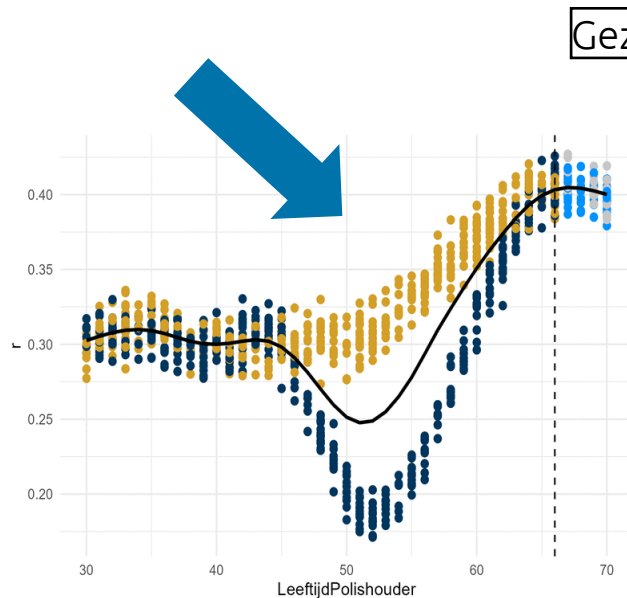
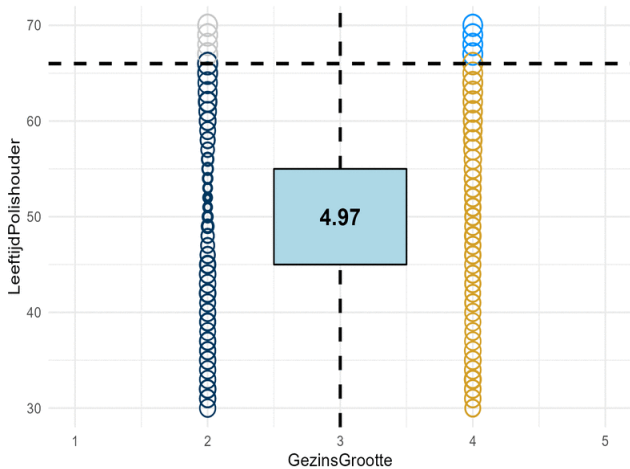


D

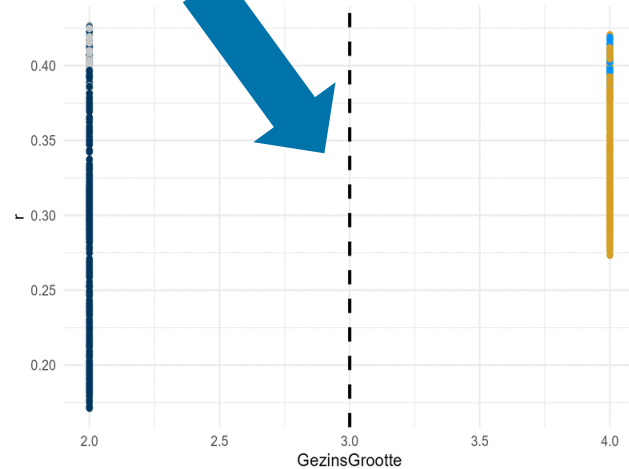


LeeftijdPolishouder = 52 (ouder)

FAST



GezinsGrootte = 3



Jong/geen kinderen

Ouder/geen kinderen

Jong/wel kinderen

Ouder/wel kinderen

A



C



B

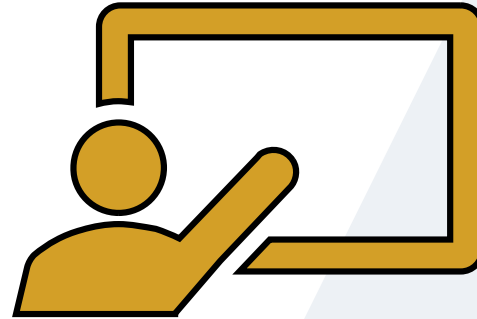


D



CONCLUSIE

- Waarom zou je dit willen gebruiken?
 - Hogere voorspelkracht
 - Inzicht in verborgen effecten
 - Mogelijke effecten:
 - Anti-selectie
 - Opzeggingen
 - Uitlegbaar
 - Keuzevrijheid



BEDANKT VOOR UW AANDACHT

Bezoekadres

Hoogoorddreef 54
1101 BE Amsterdam Zuidoost

www.aaa-riskfinance.nl
info@aaa-riskfinance.nl