



# Datacoalitie Datagedreven onderzoek

*Studenten nóg beter begeleiden  
met een datagedreven aanpak!*

16 mbo-instellingen onderzoeken samen met OCW, DUO en MBO Raad hoe data kan worden ingezet om de studieloopbaan van studenten te verbeteren en wat de toegevoegde waarde is om dit met een datagedreven aanpak te doen.

16 mbo-instellingen:

Firda

roc van  
twente



curio

Noorderpoort

deltion  
college

ROC NIJMEGEN

Graafschap  
College

Aventus<sup>A</sup>  
De school waar jij 't maakt

ONDERWIJSGROEP TILBURG

KONING  
WILLEM I  
COLLEGE

albeda

roc  
MONDRIAAN

Alfa - college

ZADKINE

gildeopleidingen

MBO Raad:



DUO:



OCW:



Projectorganisatie

hutspot



# Agenda

- 1 **Introductie**
- 2 **Werkwijze datacoalitie**
- 3 **De ontwikkelde methodiek: de uitnodigingsregel**
- 4 **Kunnen we dit repliceren bij andere instellingen?**
- 5 **Wat zijn de uitkomsten / geleerde lessen?**
- 6 **Hoe gaan we als datacoalitie verder?**

# Werkwijze van de datacoalitie

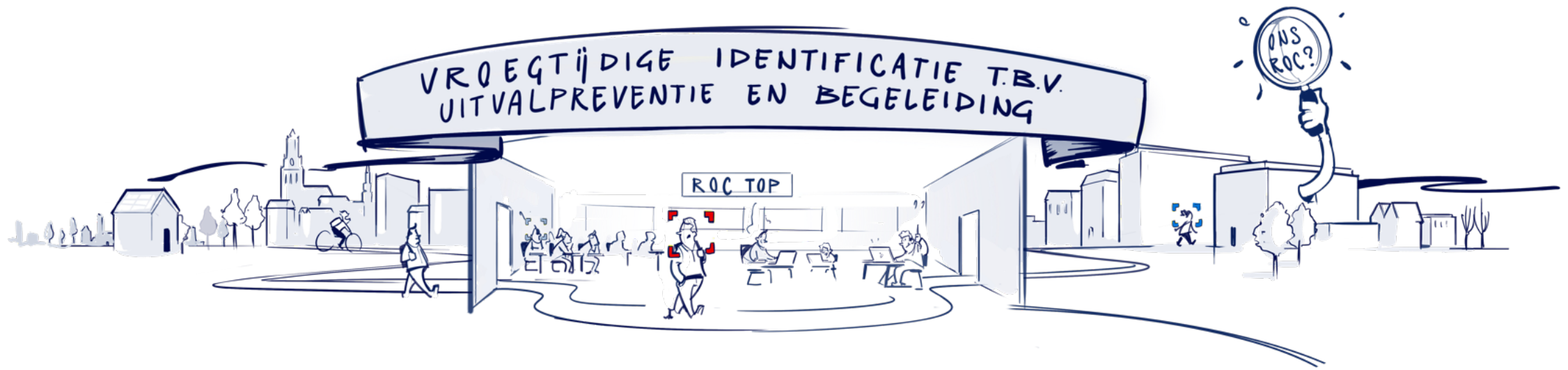


## Ontstaan van de coalitie

## Werkwijze

- Praktische onderzoeksvragen leiden tot werkgroepen
- MBO instellingen kiezen hoe en in welke mate ze actief participeren
- Praktische resultaten & lessons learned
- Opbrengsten gedeeld met de sector





## Onderzoeksgroep 3

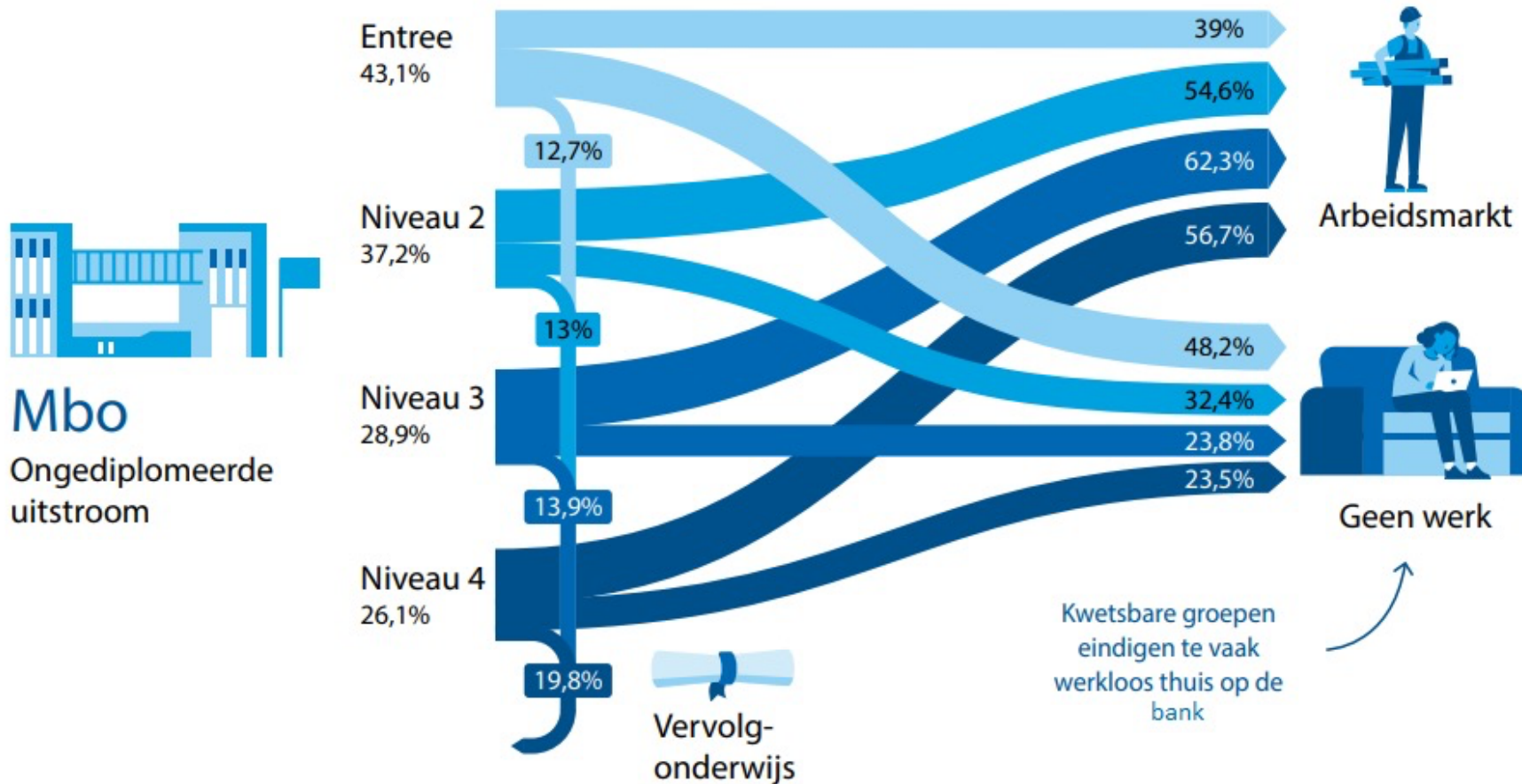
Vroegtijdige identificatie  
t.b.v. uitval preventie

# Aanleiding onderzoek Irene Eegdeman

*“Wij hebben zoveel uitval, wil jij even uitzoeken  
wat we daaraan kunnen doen?”*



# Studentuitval in beeld

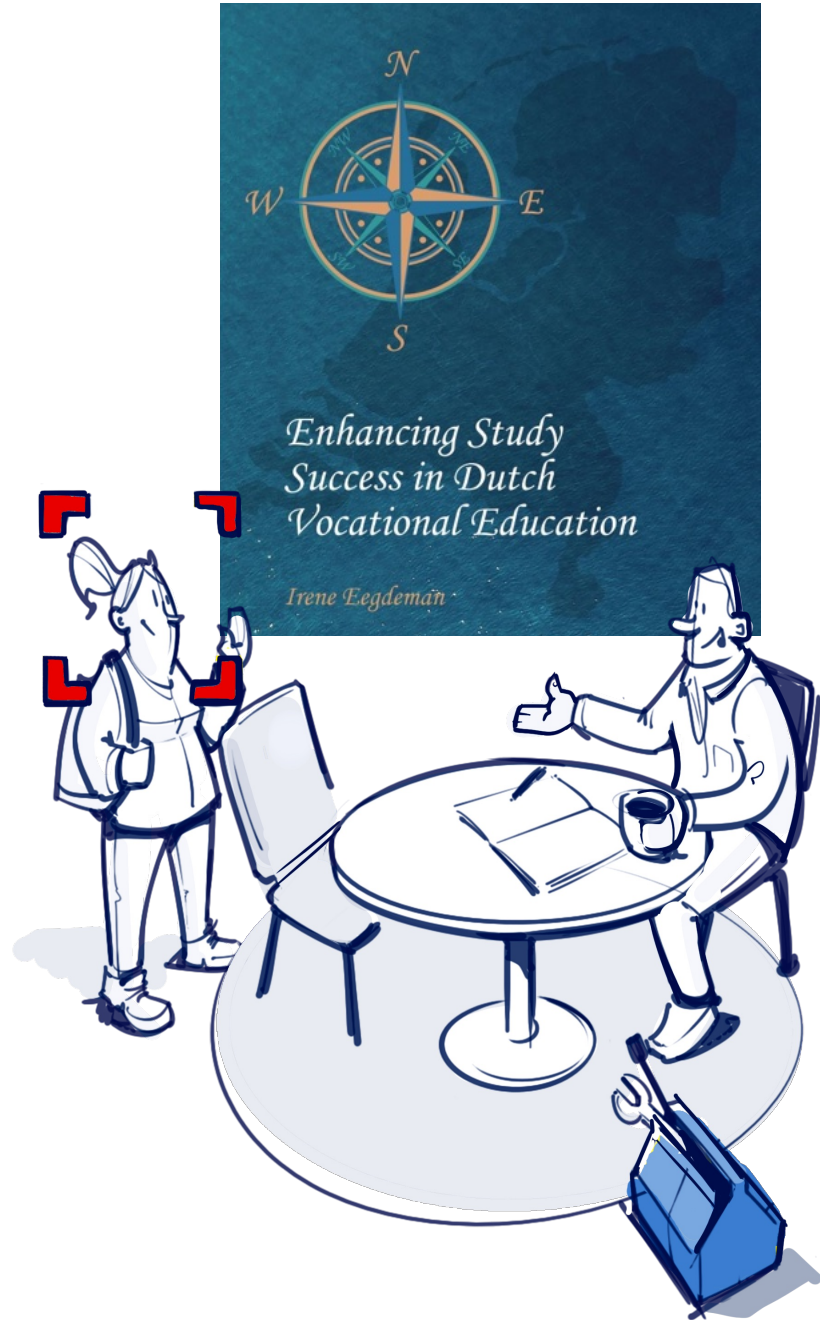


Bron: CBS (2022a, 2022b)

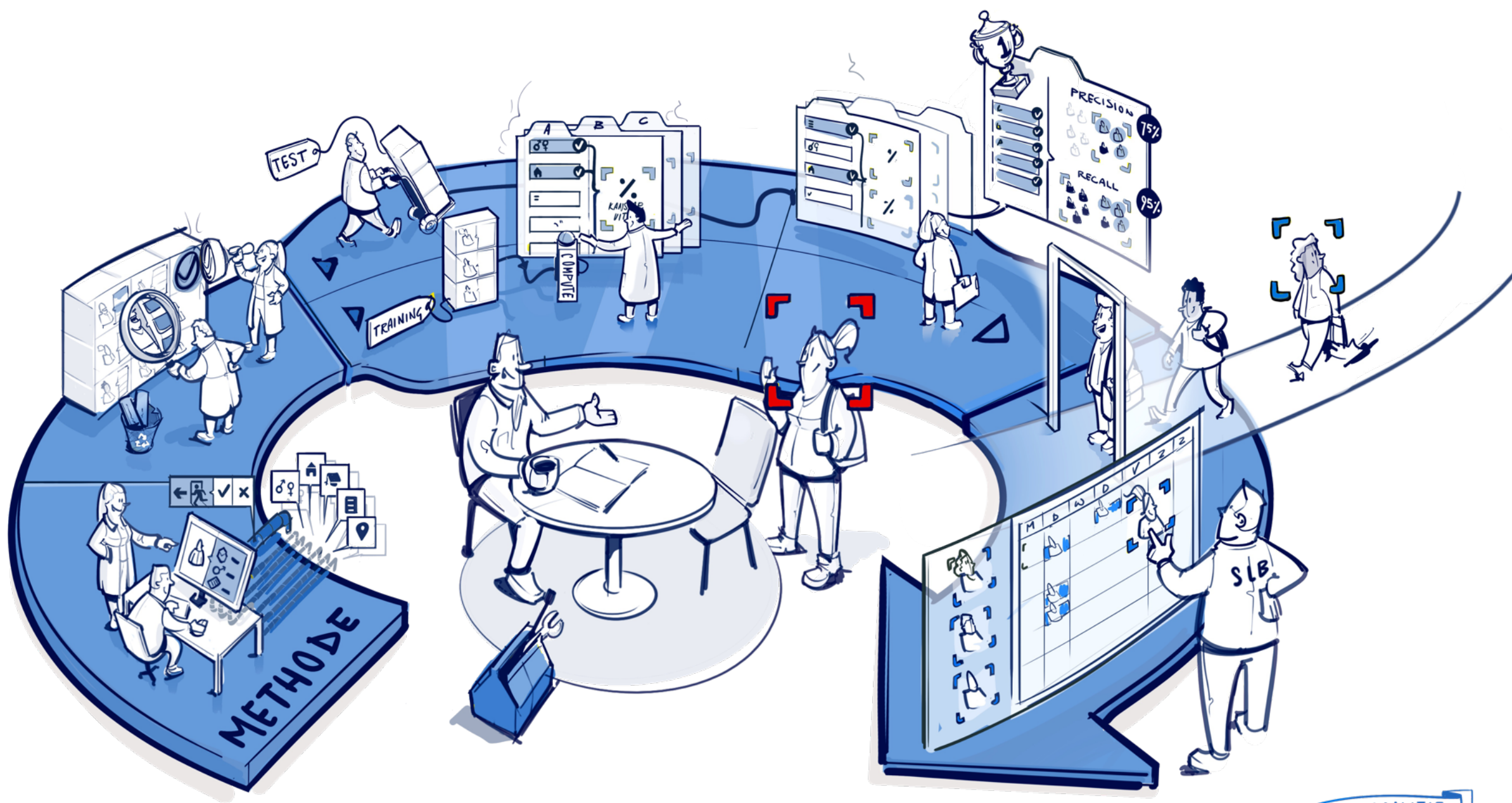
# Uitnodigingsregel

*We zijn dus op zoek naar een manier om studentuitval vroeg in beeld te hebben...*

*Kunnen we bepalen wie we voor bijvoorbeeld een interventie zouden moeten uitnodigen?*







# Data verzamelen

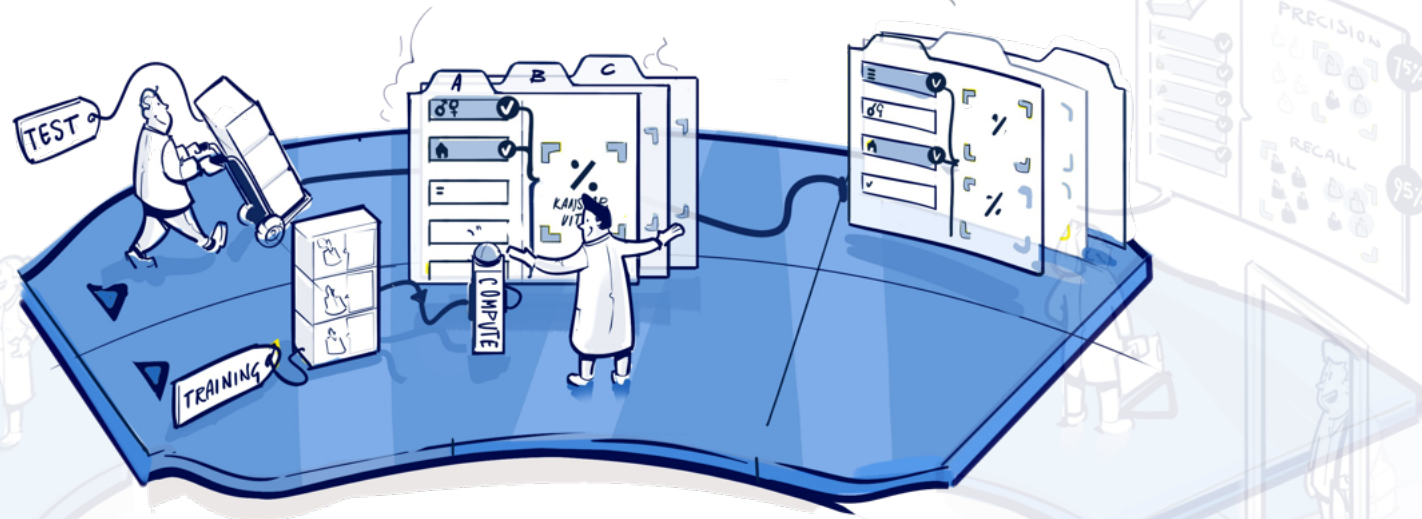
**Vroeger:** Op basis van theorie een keuze maken voor variabelen in een voorspelmodel.

**Nu:** Opkomst van Machine Learning in het onderwijs, gebruik alle data en laat de computer beslissen wat de beste variabelen zijn.

## Welke data?

- Vooropleidingsdata inclusief diploma & resultaten
- Opleidingsdata
  - Formatieve en summatieve resultaten
  - Niveau en variant
  - Opleidingsnaam
- Presentie / verzuim





# Uitkomst waarden

Vooral geïnteresseerd naar goede voorspelling van **uitval** dus dan kijken naar **sensitiviteit** (identificatie van de uitvallers) en **precisie** (correct voorspelde uitvallers)

**Geen arbitraire grens van 0.5 meer om 'de uitvaller' te bepalen:**

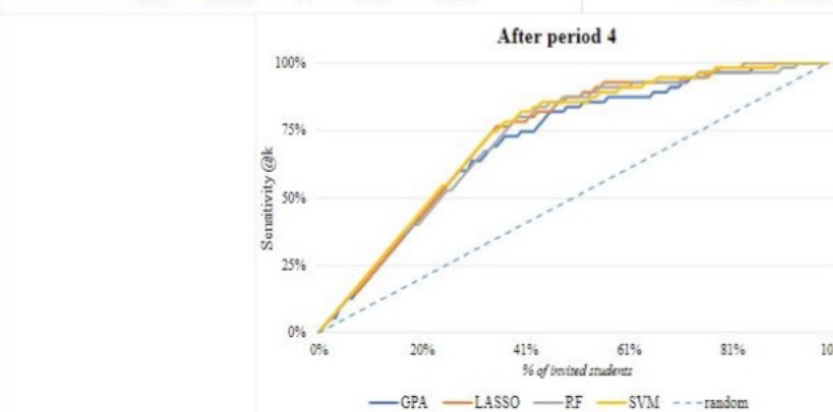
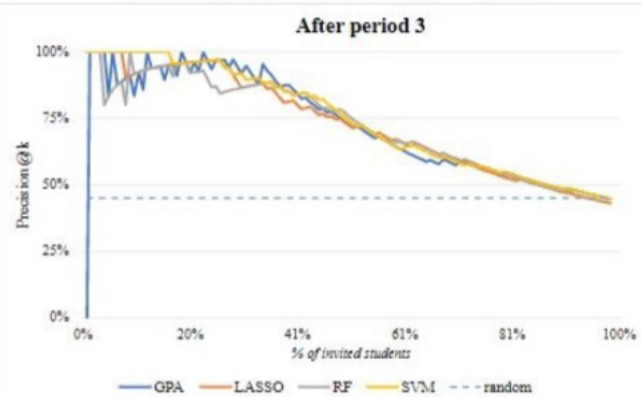
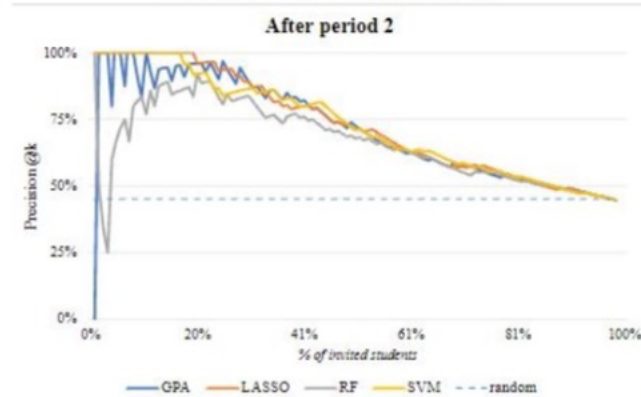
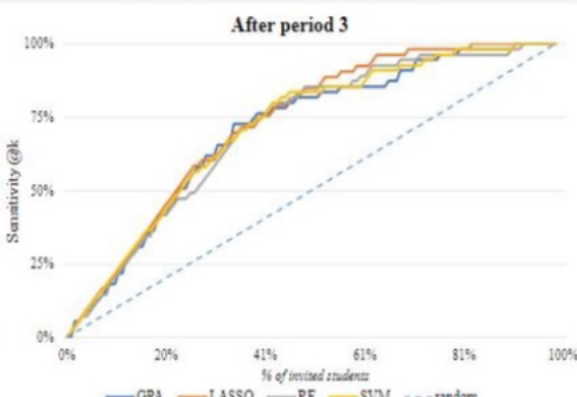
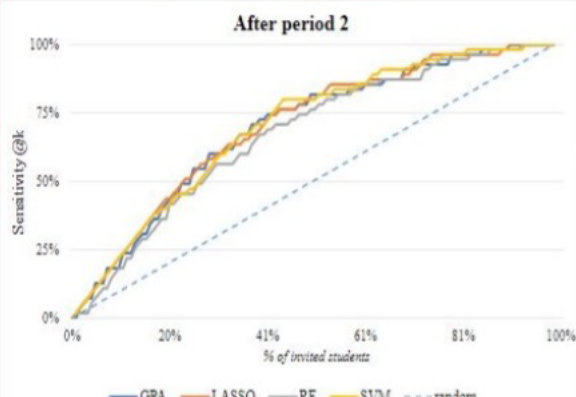
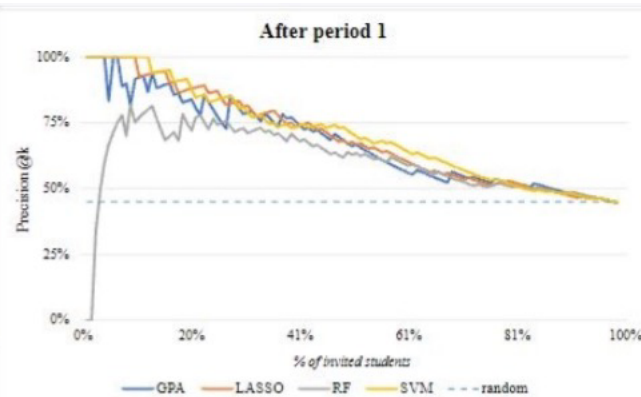
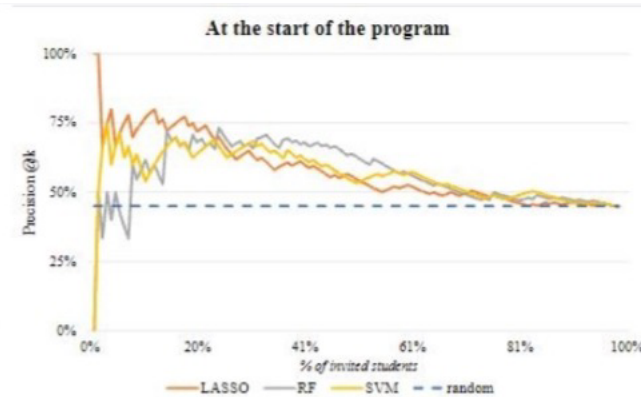
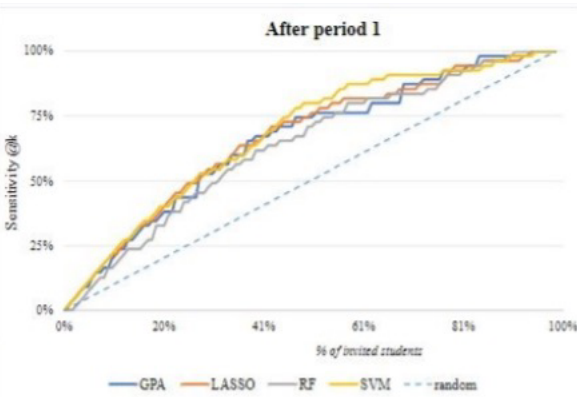
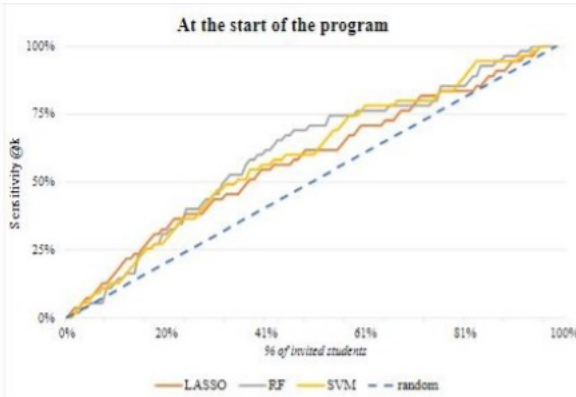
Ik wil jullie laten zien dat je zelf kan bepalen hoeveel % van de uitvallers je zou willen bereiken en hoeveel studenten je daarvoor moet uitnodigen



# Uitnodigingsregel

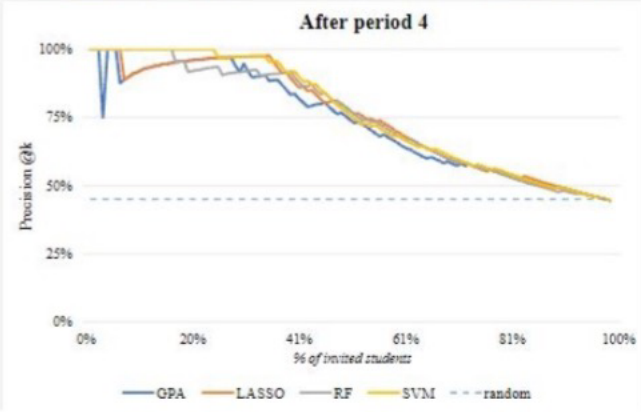
- Studenten sorteren op uitvalskans (hoog naar laag)
- Periodiek data toevoegen
- Dynamisch de prestatie van het model meten
- Meerdere modellen geschat (LASSO, Random Forest, SVM)
- Kijken bij % uitgenodigden, niet bij een kans van 0.5
- Kijken op 5 momenten in het jaar





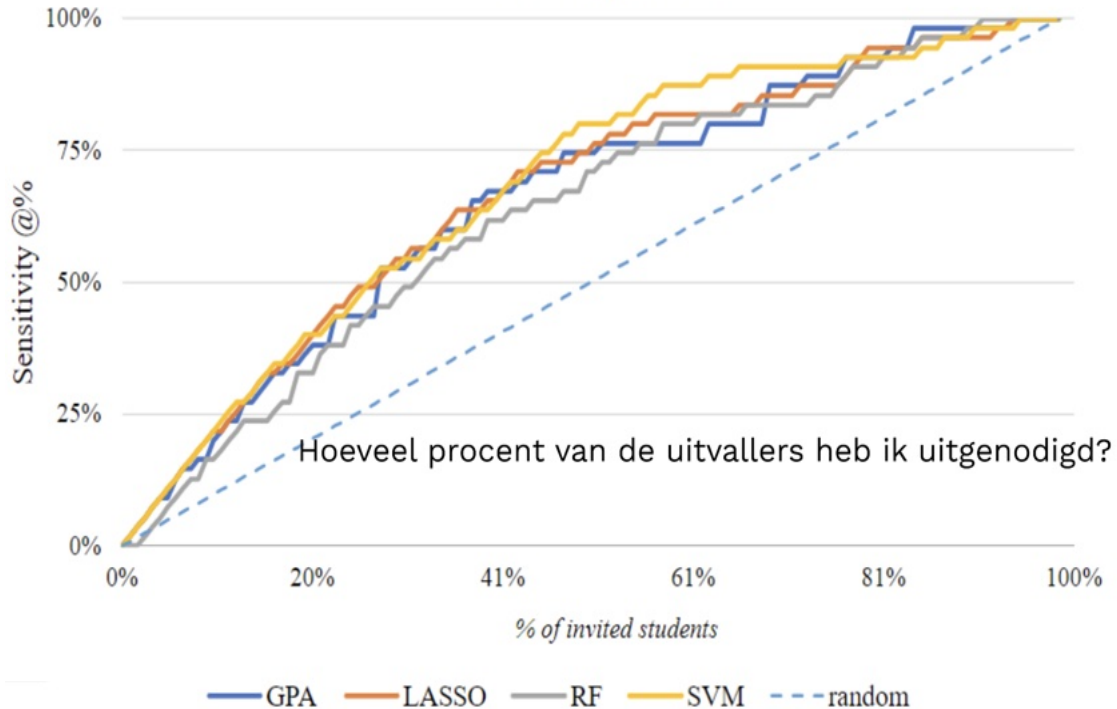
Sensitiviy: hoeveel % van de uitvallers heb je te pakken?

Precision: hoeveel % van de uitgenodigde studenten is uitvaller?

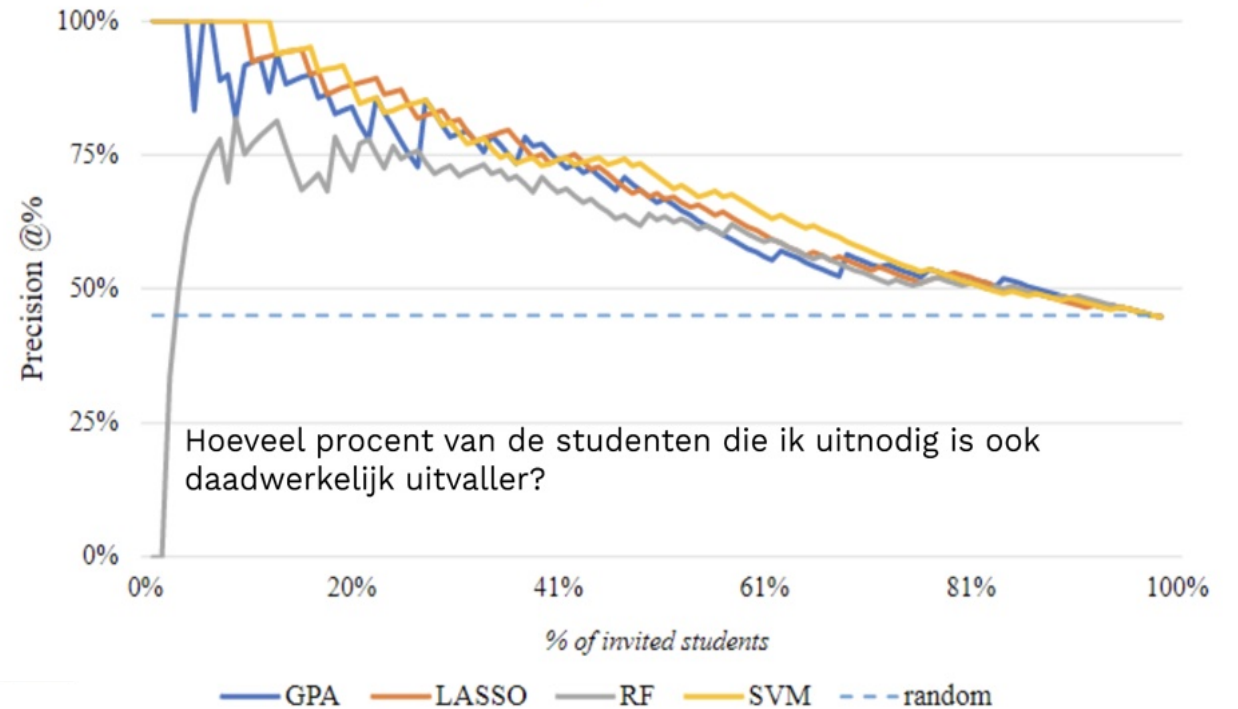


# Uitnodigingsregel

After period 1



After period 1



# Toepassen

- Bijvoorbeeld volgorde van SLB gesprek bepalen
- Waar moeten we ons op richten in de eerste 10 weken?
- Geïnformeerd keuzes maken (eerste 10%, of juist tweede 10%?)

→ *Welke gevolgen heeft handelen op basis van modellen?* ←





# vroegtijdige identificatie t.b.v. uitvalpreventie en begeleiding



De data wordt verdeeld in een trainingsdataset en een testdataset. De trainingdataset wordt gebruikt om verschillende algoritmes te trainen en vergelijken.

3.

Met behulp van de testdataset wordt gecontroleerd of het algoritme geen verkeerde verbanden heeft gelegd.

4.

5.

Om te zien of het algoritme goed genoeg werkt wordt er een nieuwe voorspelling gemaakt op de studenten uit de testdataset. Zo wordt er getest of de voorspellingen van het algoritme kloppen.

2.

De data wordt gecontroleerd op kwaliteit: is het beschikbaar, is de kwaliteit op orde, is de data eenduidig vastgelegd?

1.

Een onderzoeker verzamelt samen met een medewerker historische data over een groep studenten. Bijvoorbeeld het type vooropleiding, eindcijfers VO, geslacht, woonplaats, formatieve resultaten en verzuimgegevens. Hoe meer data, hoe beter! Ook moet natuurlijk bekend zijn of er sprake was van uitval of niet.



6.

Het beste algoritme kan nu gebruikt worden. De SLB'er kent zijn studenten. Het algoritme helpt de SLB'er om vroegtijdig te bepalen welke student(en), op basis van een verwachte uitvalkans, extra aandacht of hulp nodig hebben.

## Hoe werkt de methode Eegdemans?

**Met behulp van data brengen we studenten in het vizier die een verhoogde kans hebben op uitval. Als we weten welke studenten de meeste kans hebben om uit te vallen uit de opleiding kunnen mentoren en SLB'ers gericht vroegtijdig studenten uitnodigen.**



# Replicatieonderzoek

# Replicatieonderzoek

*Kunnen we het onderzoek op een andere instelling herhalen, voor dezelfde én andere opleidingen?*

*En hoe generaliseerbaar zijn de resultaten?*



# Aanpak: deelnemende instellingen

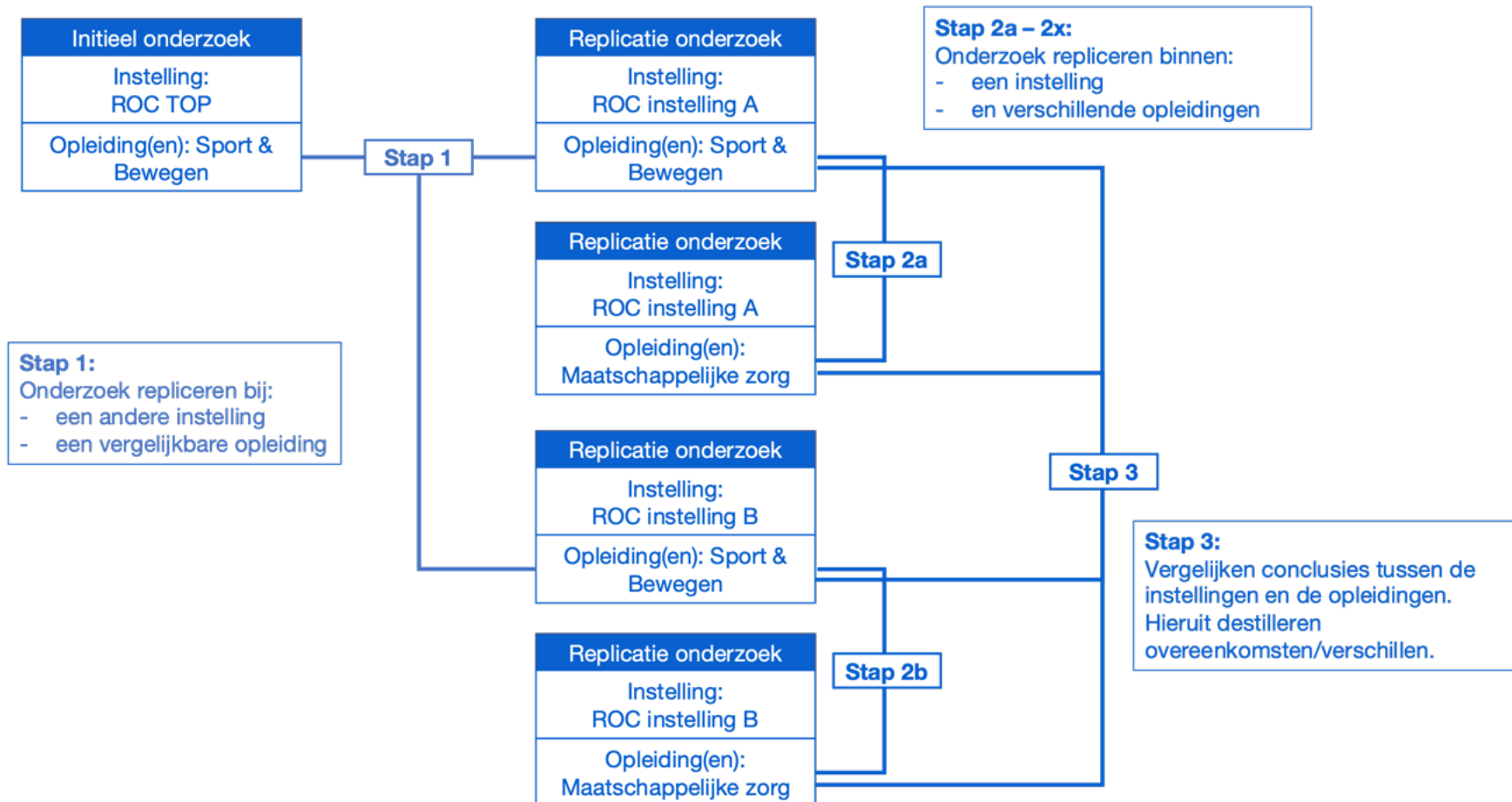
## Opleidingen:

- Sport en bewegen
- Maatschappelijke zorg
- MBO-verpleegkundige

## Instellingen

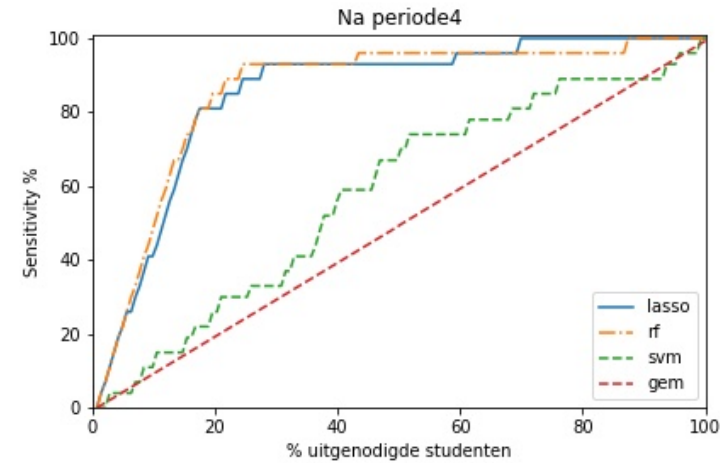
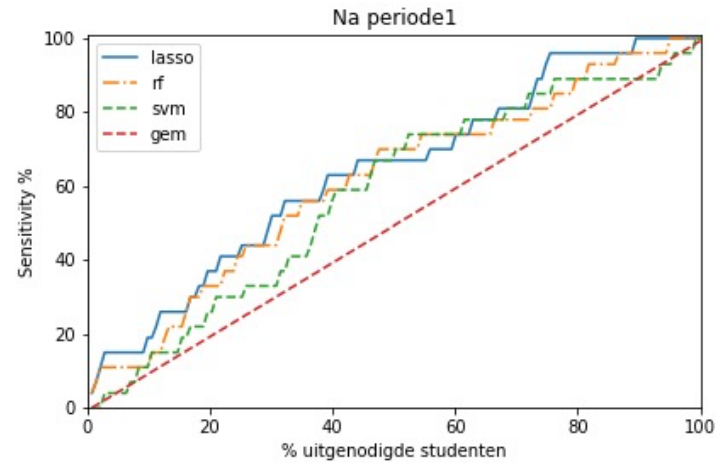
1. ROC van Amsterdam-Flevoland
2. Albeda
3. CURIO
4. Onderwijsgroep Tilburg
5. Noorderpoort
6. Gilde Opleidingen
7. Alfa College
8. ROC Mondriaan
9. Graafschap college

# Aanpak: onderzoeksopzet

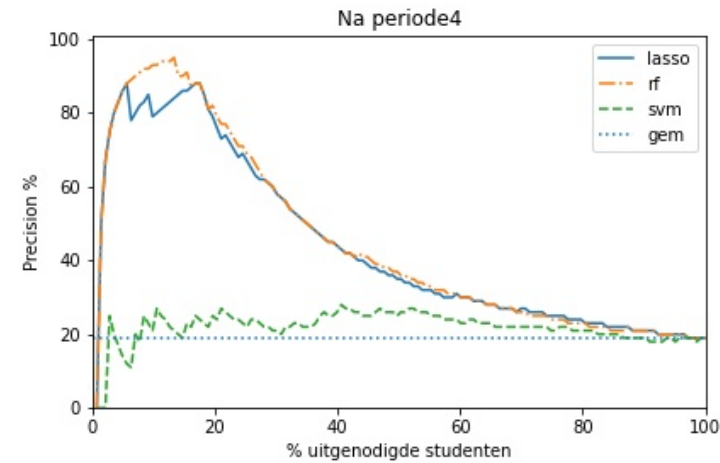
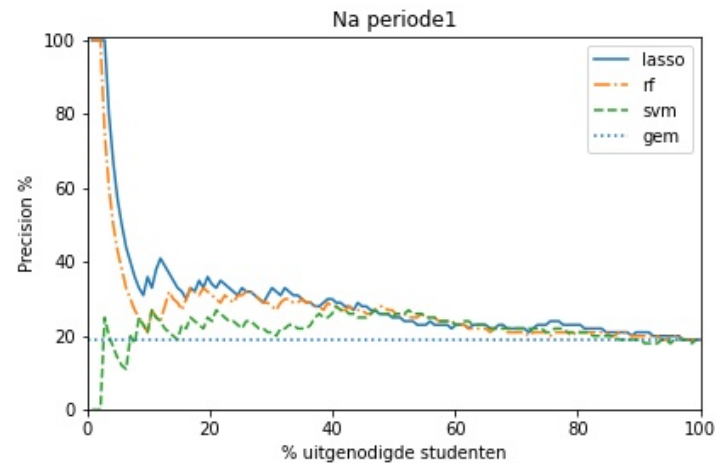


# Resultaten

## Sensitivity

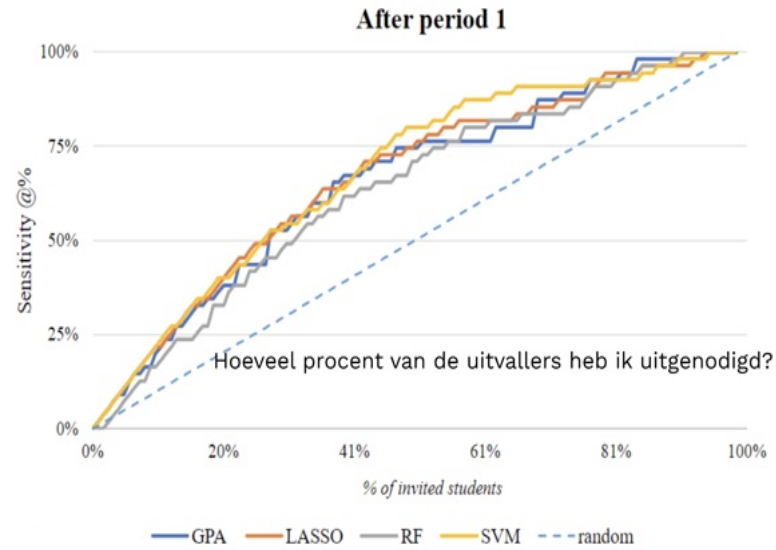


## Precision



De resultaten verbeteren naarmate je verder in het jaar bent

# Resultaten- Sensitivity

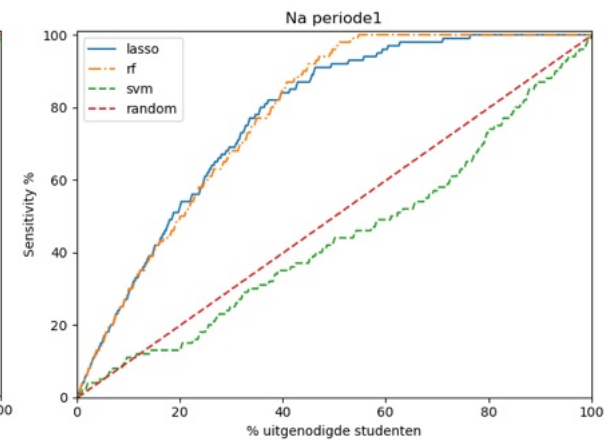
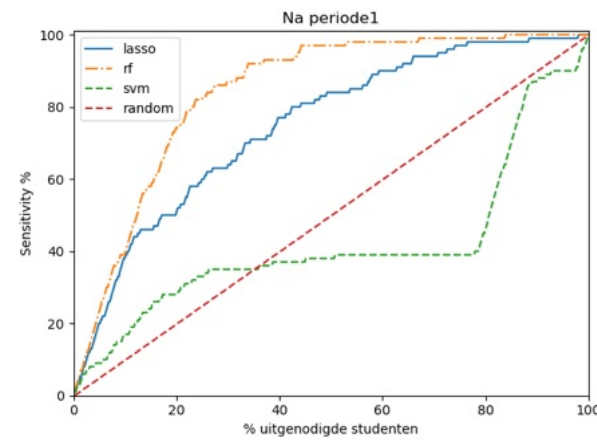
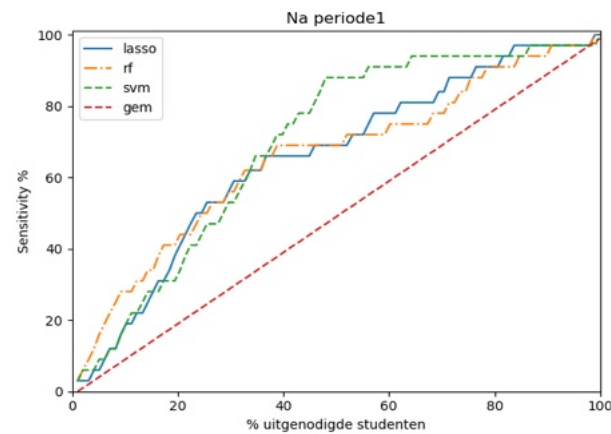
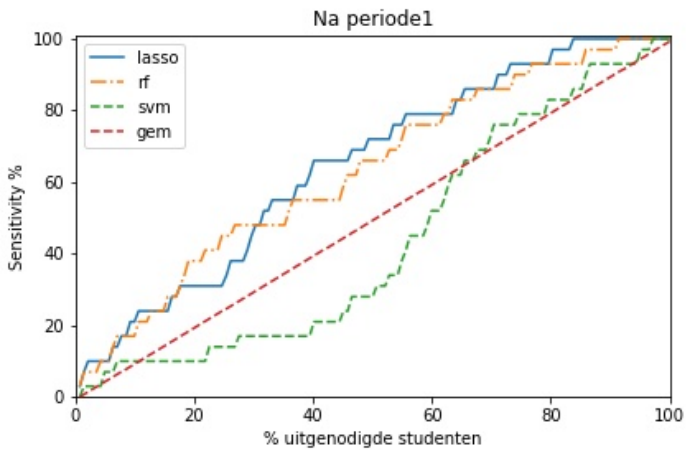


ROCvA/F

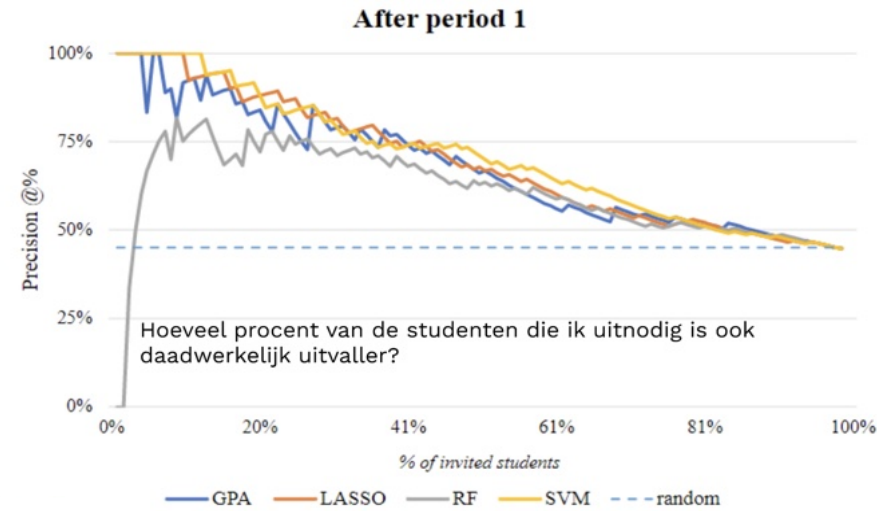
ROC Mondriaan

Onderwijsgroep Tilburg

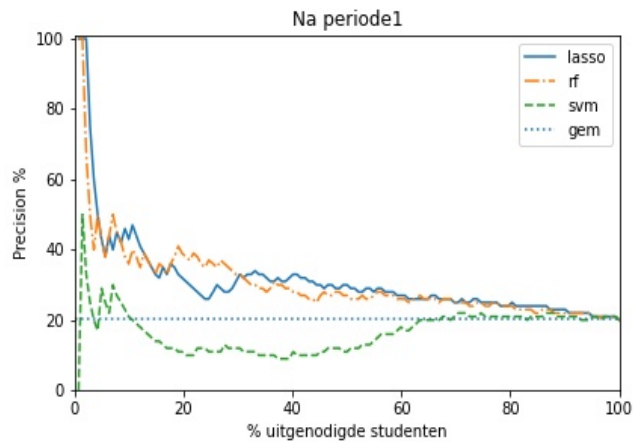
Gilde opleidingen



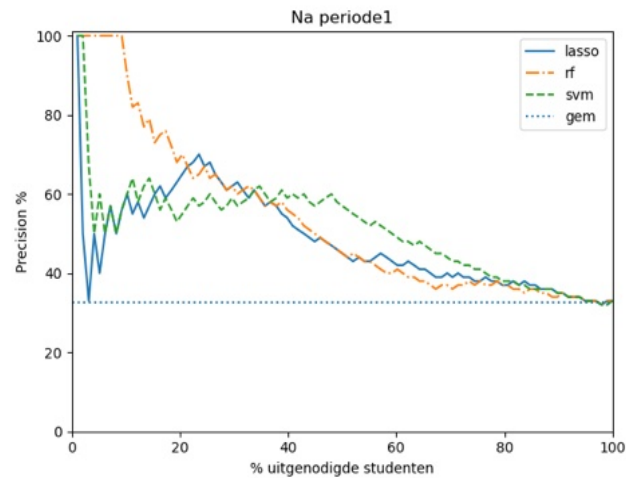
# Resultaten- Precision



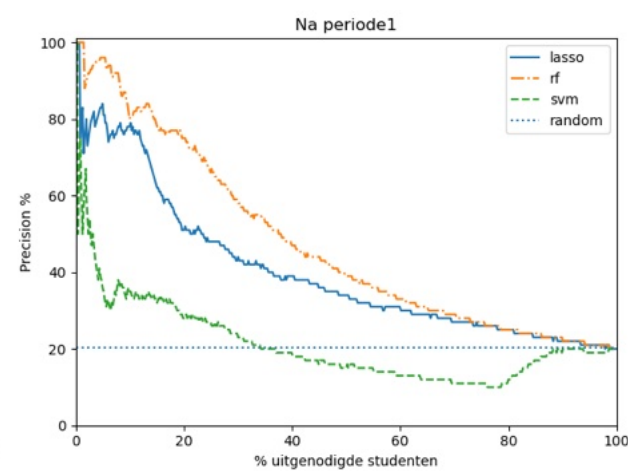
ROCvA/F



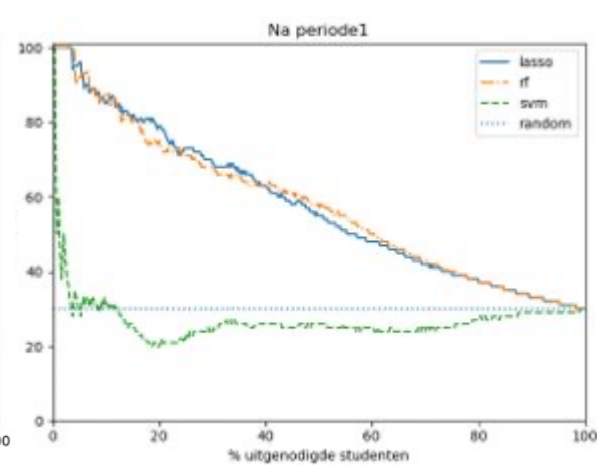
ROC Mondriaan



Onderwijsgroep Tilburg



Gilde opleidingen



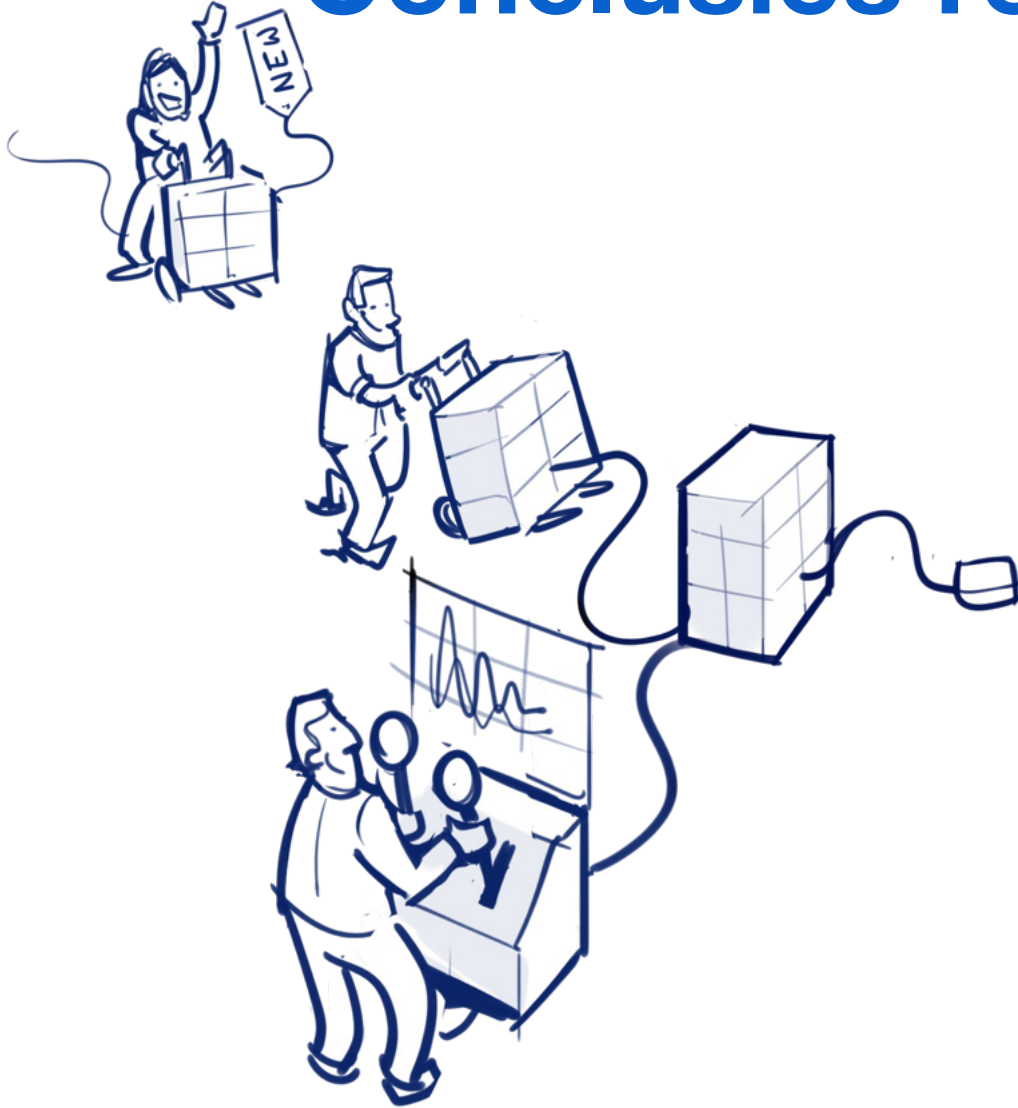
# Conclusies replicatie onderzoek

## Conclusie

- Resultaten worden beter naar mate het jaar vordert
- Resultaten verschillen tussen opleidingen en tussen instellingen

Het is mogelijk om een robuust model te ontwikkelen dat in een vroeg stadium potentiële uitvallers kan identificeren. Maar er is ruimte voor verbetering....

# Conclusies replicatie onderzoek

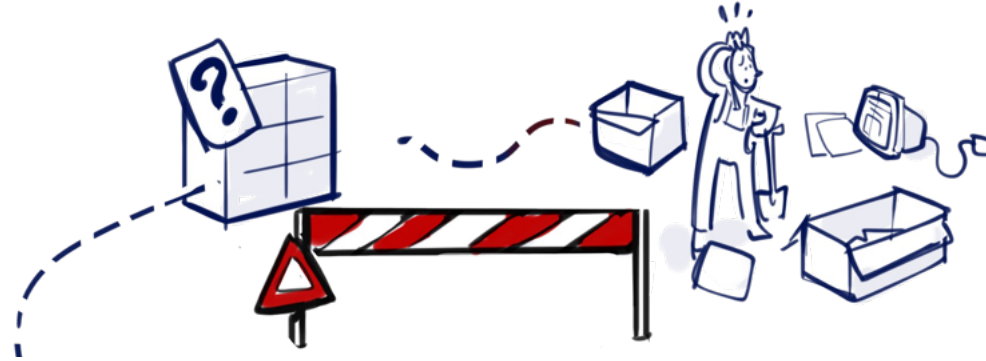


## Wat kan er nog beter?

- Datakwaliteit is niet overal hetzelfde... Sommige opleidingen zijn heel compleet, andere minder.
- Meer data en met name meer detaildata toevoegen (o.a. presentie)
- Recentere jaren (grotere groep studenten om op te trainen)
- Hyper-parameters en instellingen van de modellen verfijnen



# Wat was de grootste uitdaging?



**Binnen niet alle instellingen is een datascientist (skills) voorhanden**

- Er zijn nog geen gebaande paden naar de data die er in de organisatie is

***Databeschikbaarheid & kwaliteit is het grootste aandachtspunt***

- Vooropleidingscijfers, motivatie, en intaketoetsen waren niet (overal) beschikbaar.
- Andere opleidingen zijn mogelijk relevanter (bv. meer dropout) of effectiever (bv. meer data beschikbaar)

# Hoe gaan we als datacoalitie verder?



## 1 Model verder verbeteren:

Wat is het effect van meer data toevoegen? Met name detailinformatie over resultaten en Aan- en Afwezigheid.

Welke verbetering kunnen we behalen met het finetunen van de parameters? Bijvoorbeeld de diepte van de beslisbomen?



## 2 Model naar de praktijk brengen? Hoe kijken onderwijsteams naar dergelijke modellen en wat hebben zij nodig om dit naar de onderwijspraktijk te brengen?



## 3 Hoe kunnen we verder aan de slag met de kwaliteit van de data? Hoe breng je dat effectief in kaart om verbeterkansen te identificeren?

# Zelf aan de slag?

**Oproep aan de sector:** werk samen, datagedreven, aan onderzoek naar het terugbrengen van verzuim.

## Wat moet je daarvoor als school doen?

1. **Breng data voor de toekomst op orde**
  - Organiseer een structurele vastlegging van studieresultaten (formatief en summatief)
  - Zorg voor behoud van data voor onderzoek
  - Organiseer toegang tot de data. Vastleggen is belangrijk maar toegang is essentieel
2. **Investeer in datascience kennis en kunde**
  - Bouw kennis en kunde op
3. **Breng 'datagedreven werken' naar de teams**
4. **Voer het ethische gesprek als organisatie en met je student**

