



Rechtvaardige Selectie- Algoritmes

Brinn Hekkelman,
Bas Scheer,
Mark Kattenberg

b.hekkelman@cpb.nl
www.cpb.nl | @CPBnl

Insteek

- Algoritmes kunnen heel nuttig en goed ingezet worden
- Zorg wel dat je zelf in controle blijft!

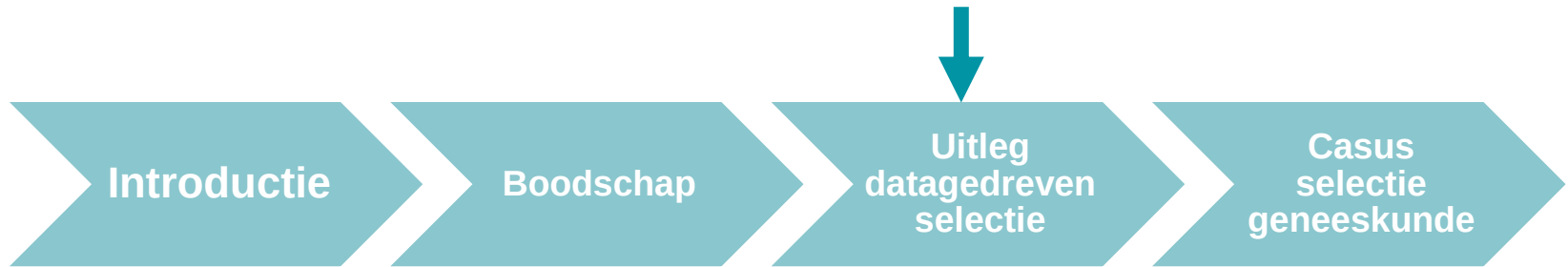


Selectieprocedures gebruiken steeds vaker datagedreven selectie

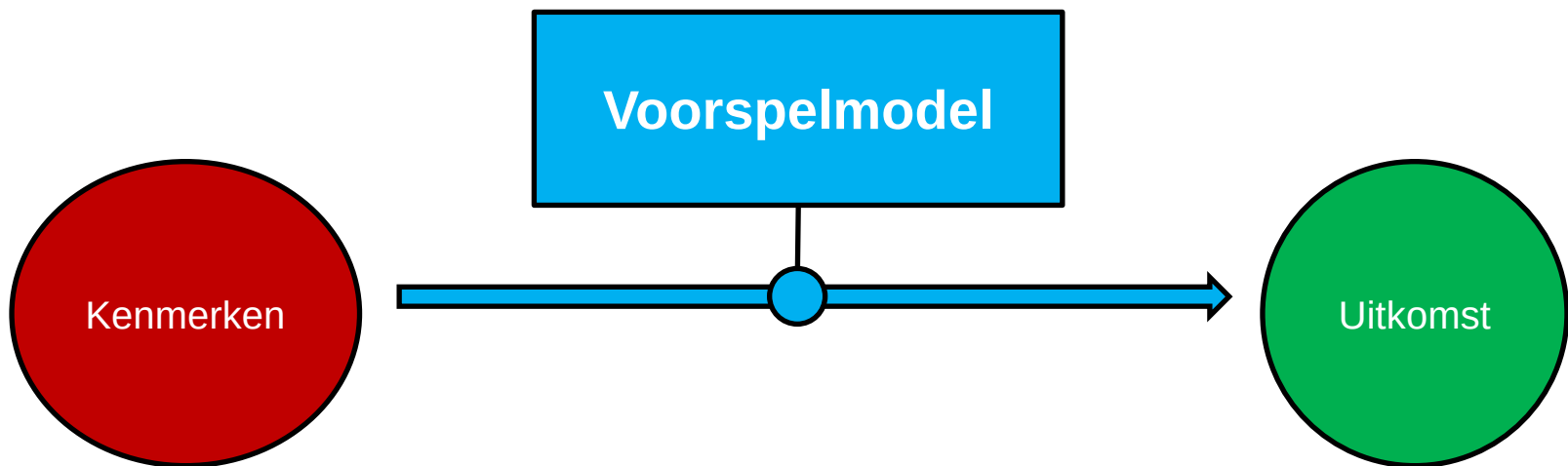
- Denk aan:
 - Datagedreven fraudedetectie door financiële instellingen en uitvoeringsorganisaties
 - AI-oplossingen voor het vinden van de meest geschikte sollicitant
- De angst bestaat dat deze systemen bestaande vooroordelen in stand houden
 - Of zelfs verergeren
 - En zo leiden tot discriminatie en uitsluiting
- Veel voorkomende oplossing: gevoelige informatie weglaten in selectieproces
 - Gedachte is: “als je iets niet meeneemt, dan kan je er niet op discrimineren”

Onze boodschap

- Het weglaten van gevoelige informatie is geen garantie voor succes
 - Zelfs al observeer je iets niet direct, via via kan het toch in de data sluipen
- Datagedreven selectie zonder uitsluiting van kwetsbare groepen is wel mogelijk
 - Maar dan moet je actief bijsturen in het selectie-algoritme voordat je het gaat inzetten
- Keuzes welke kwetsbare groep te beschermen zijn subjectief van aard
 - Dus: selectie-algoritmes zijn nooit helemaal objectief



Datagedreven selectie met een **voorspelmodel**: hoe werkt dat precies?



Zodat het **voorspelmodel** met **kenmerken** de **uitkomst** kan voorspellen



Dat voorspelmodel kun je inzetten in een selectie-algoritme:

1. Train **voorspelmodel** op **kenmerken** en **uitkomsten** (verzamelde data)
2. Maak **voorspellingen** met **voorspelmodel** en **kenmerken** (nieuwe data)
3. Selecteer de **100 hoogste voorspellingen**
(bij 100 beschikbare plaatsen)

Waarom zorgt dat voor een risico op over- of onderselectie?

- Voorspelmodellen vinden kandidaten die **hoogst scoren** op verwachte **uitkomst**
- Dankzij die gerichte opzet zijn voorspelmodellen **effectief**:
 - Selecteren van de groep kandidaten die ervoor zorgen dat de uitkomst zo hoog mogelijk is
- Maar: **groepen** kunnen worden **over- of ondergerepresenteerd** in de selectie
 - Een groep die minder hoog scoort op de uitkomst, wordt minder geselecteerd
 - Ook als de lagere score resultaat van discriminatie is

Hoe kunnen we dat dan proberen tegen te gaan?

- **Gevoelige informatie weglaten uit voorspelmodel** ✖
- **De verhoudingen in de selectie bijsturen** ✔

✗ Gevoelige informatie weglaten werkt niet per se

- Gevoelige informatie kan **indirect worden meegenomen** via andere informatie
 - Bijvoorbeeld: migratieachtergrond hangt samen met woonplaats
- Gevoelige informatie weglaten kan zelfs bestaande **bias verergeren**
 - Stel: een politieagent fouilleert zonder goede reden vaker migranten
 - Dan zal een voorspelmodel opmerken dat de *success rate* (daadwerkelijk iets vinden) laag ligt bij migranten
 - En dus aangeven minder vaak migranten te fouilleren
- “Bias reversal”
 - (Arnold et al., 2022; Rambachan et al., 2020; Rambachan and Roth, 2019; Kleinberg et al., 2018)

Recent voorbeeld: fraudecontrole DUO

- Fraude met uitwonende beurs:
“woonde student wel uit huis?”
- “De afgelopen tien jaar ondersteunden de advocaten in totaal **376** studenten die beschuldigd werden van frauderen met de uitwonende beurs. Bij **367** van hen [...] ging het om studenten met een **migratieachtergrond**.”
- Omdat DUO niet naar migratieachtergrond keek konden ze zich nu onmogelijk verdedigen



✓ De verhoudingen in de selectie bijsturen

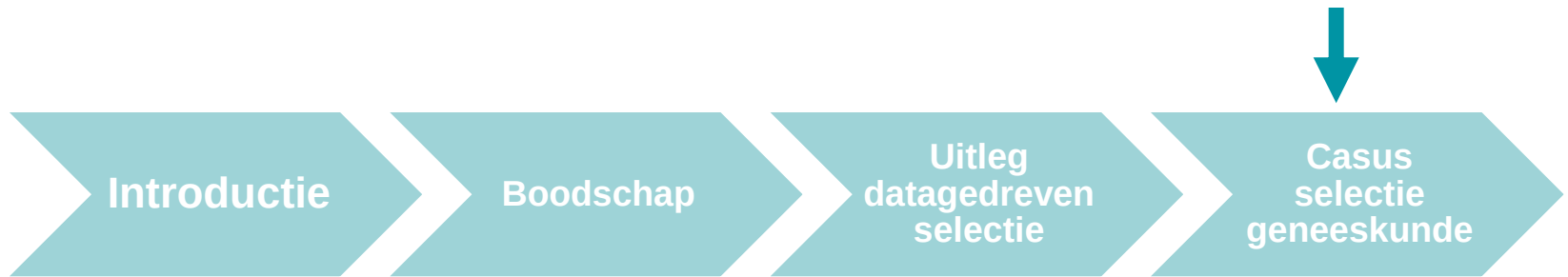
1. Train **voorspelmodel** op kenmerken en uitkomsten
(verzamelde data)
2. Maak **voorspellingen** met **voorspelmodel** en
kenmerken (nieuwe data)
3. Selecteer **beste 20 uit groep A**, en **beste 80 uit groep B**

(20 + 80 = 100 beschikbare plaatsen)

Aandeel groep A is
vastgezet op 20%

Door de selectie bij te sturen kan je representativiteit garanderen

- Maar: je moet het eens worden over **welk aandeel** en **welke groepen**
- Gegeven dat je daar uit komt, krijg je moois:
- **Effectiviteit** van voorspelmodel benut met waarborg van **representativiteit**



Casus: selectie van geneeskundestudenten

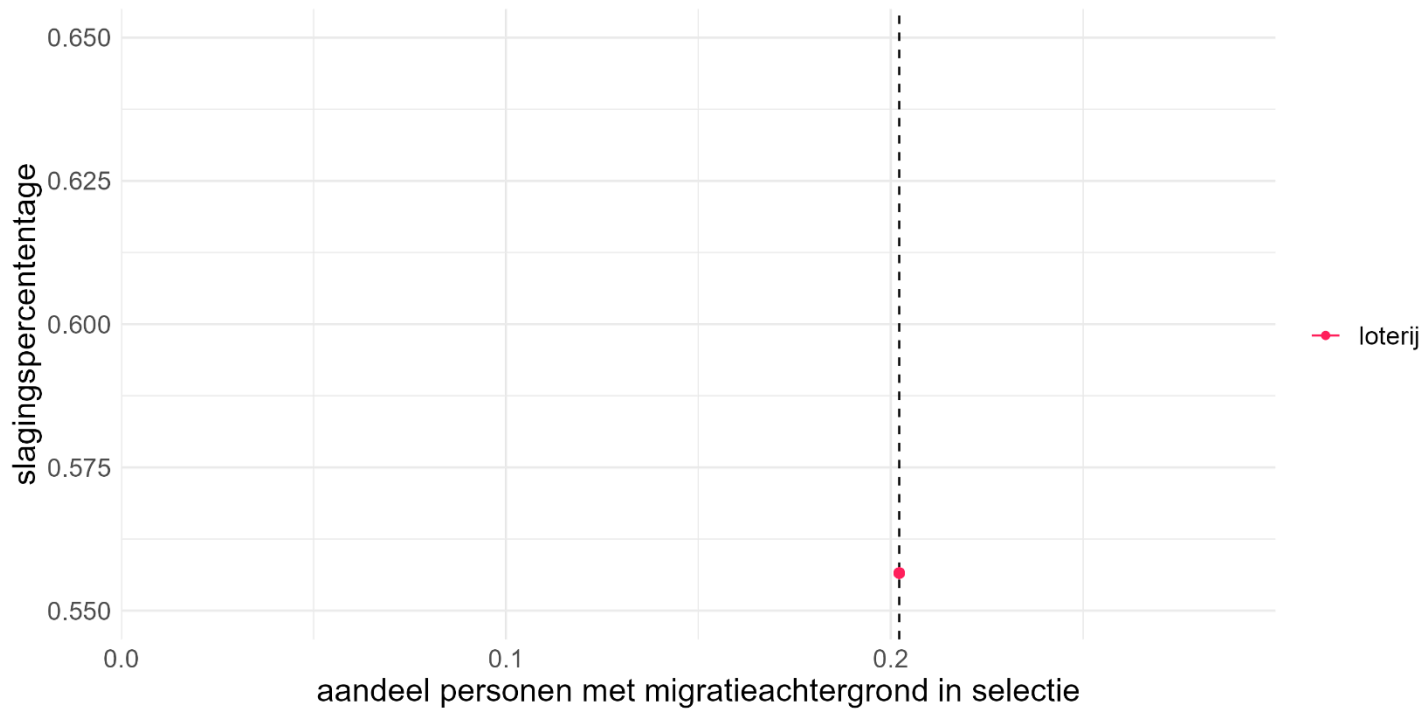


We vergelijken uitkomsten van verschillende selectiemethoden

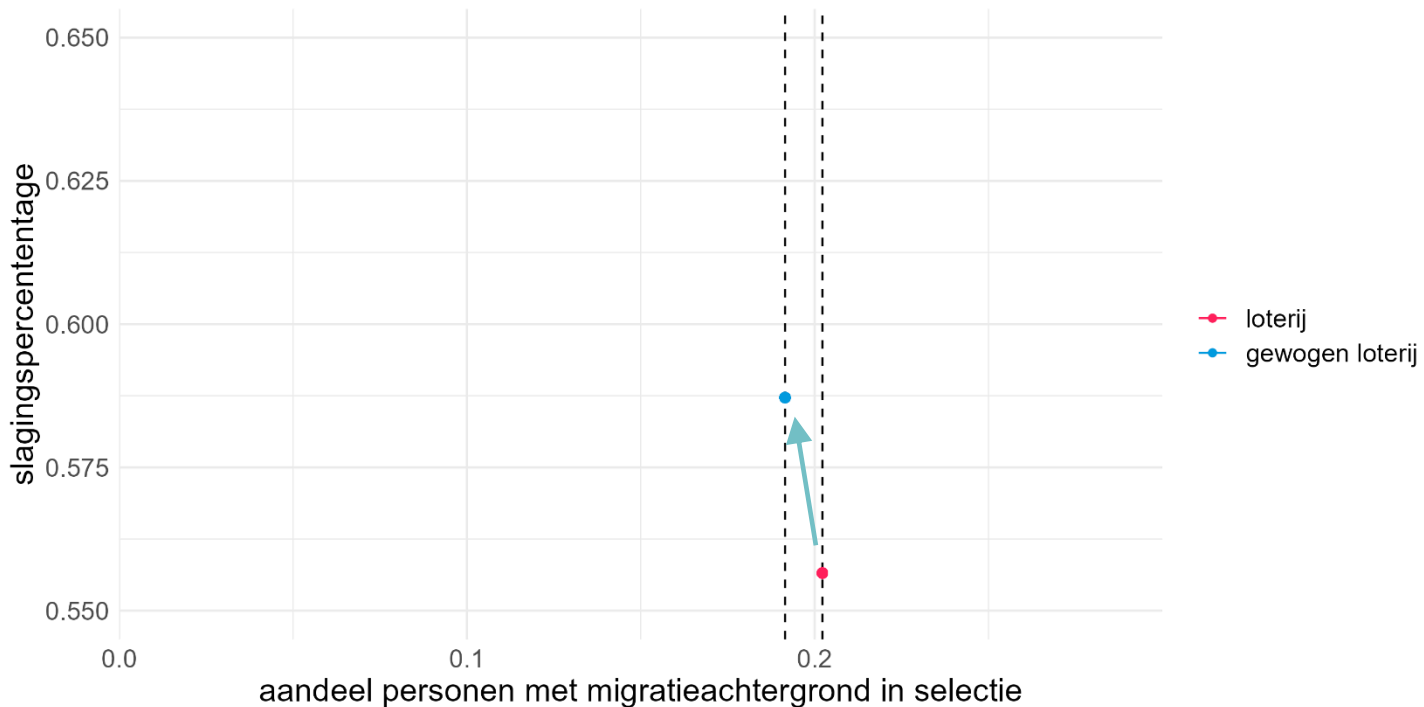
Selectiemethoden	Uitkomst	Beschermde groep
Loterij	Slagingspercentage (6 jaar na selectie)	Personen met migratieachtergrond
Gewogen Loterij	BIG-registratie	Personen met ouders met laag inkomen
Voorspelmodel (cijfer) (alleen cijfercategorie)	Werkzaam in zorgsector	Mannen
Voorspelmodel (uitgebreid) (cijfer, leeftijd, vooropleiding, geslacht, migratieachtergrond, ouderinkomen)	Aantal uren werkzaam in zorgsector	

Data van Ketel et al. (2016)

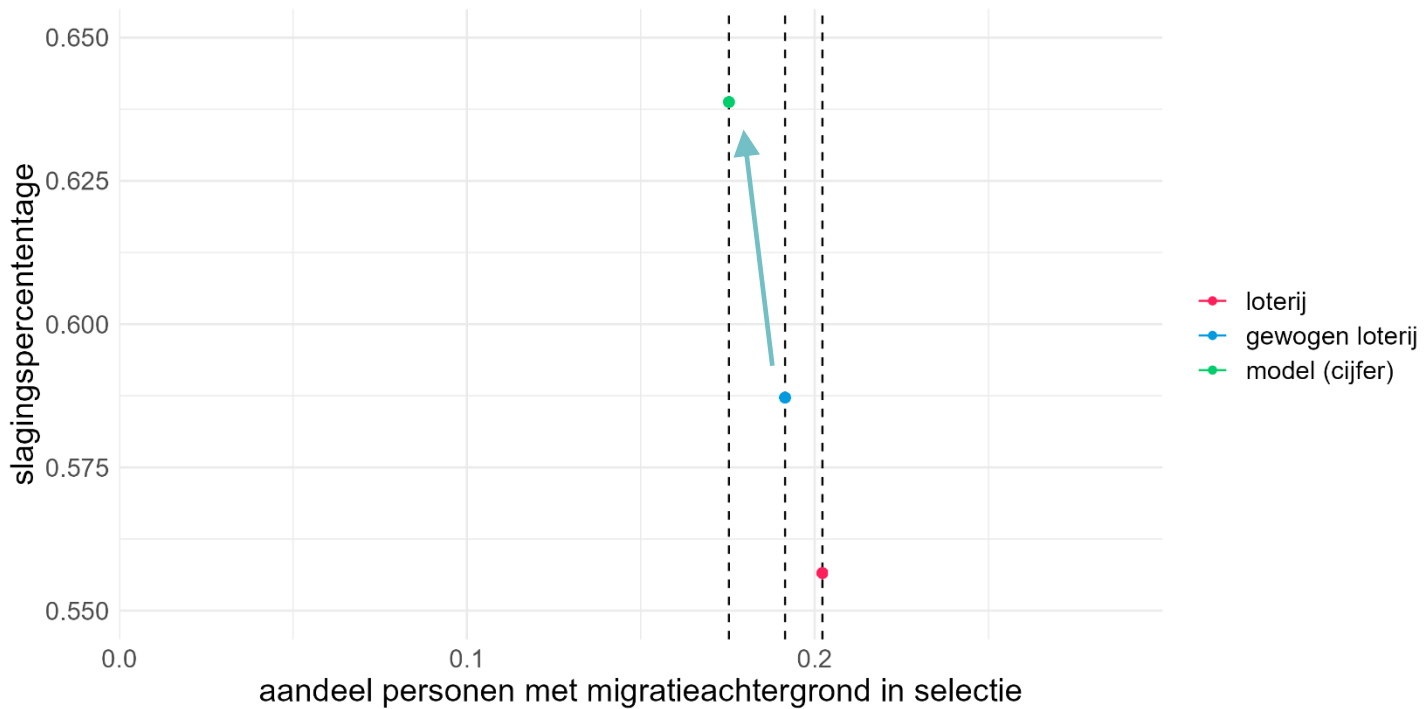
Loterij: slagingspercentage 56%, aandeel migranten 20%



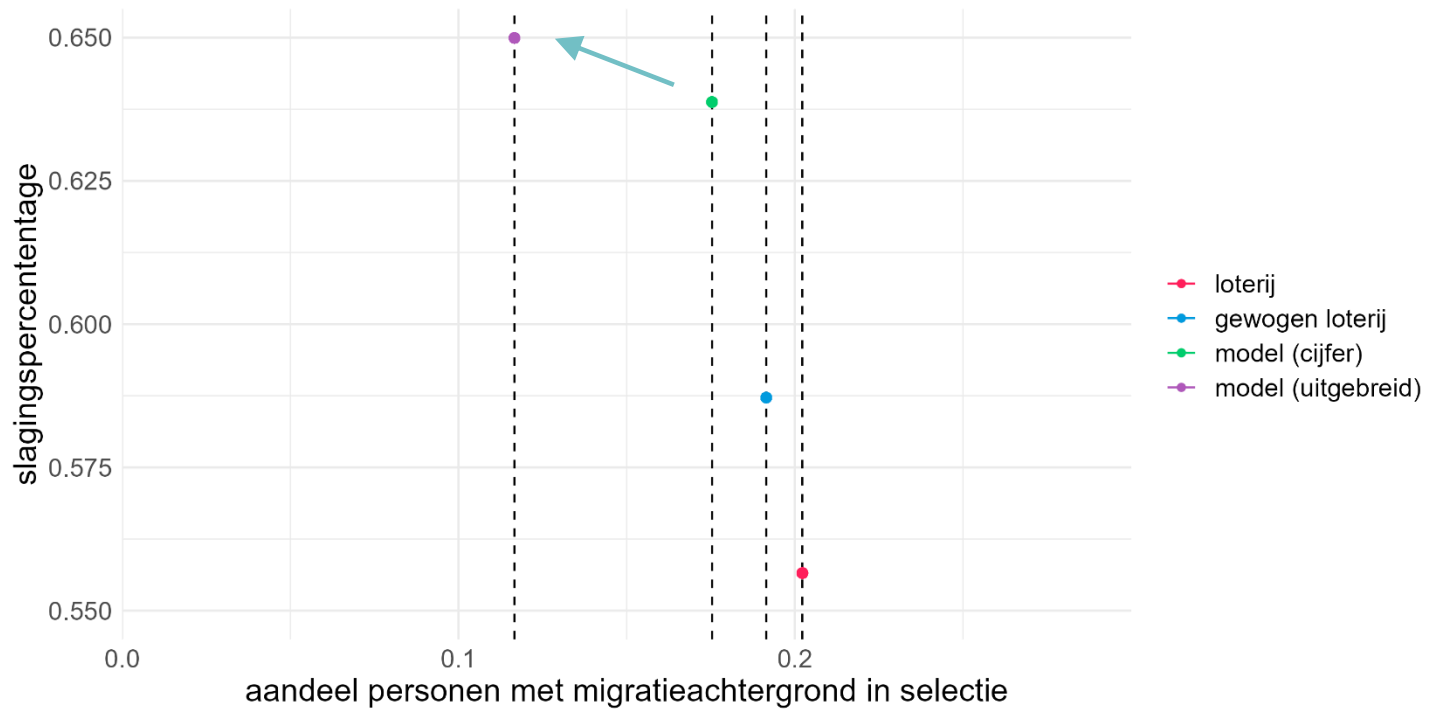
Gewogen loterij (cijfer): slagingspercentage 59%, aandeel migranten 19%



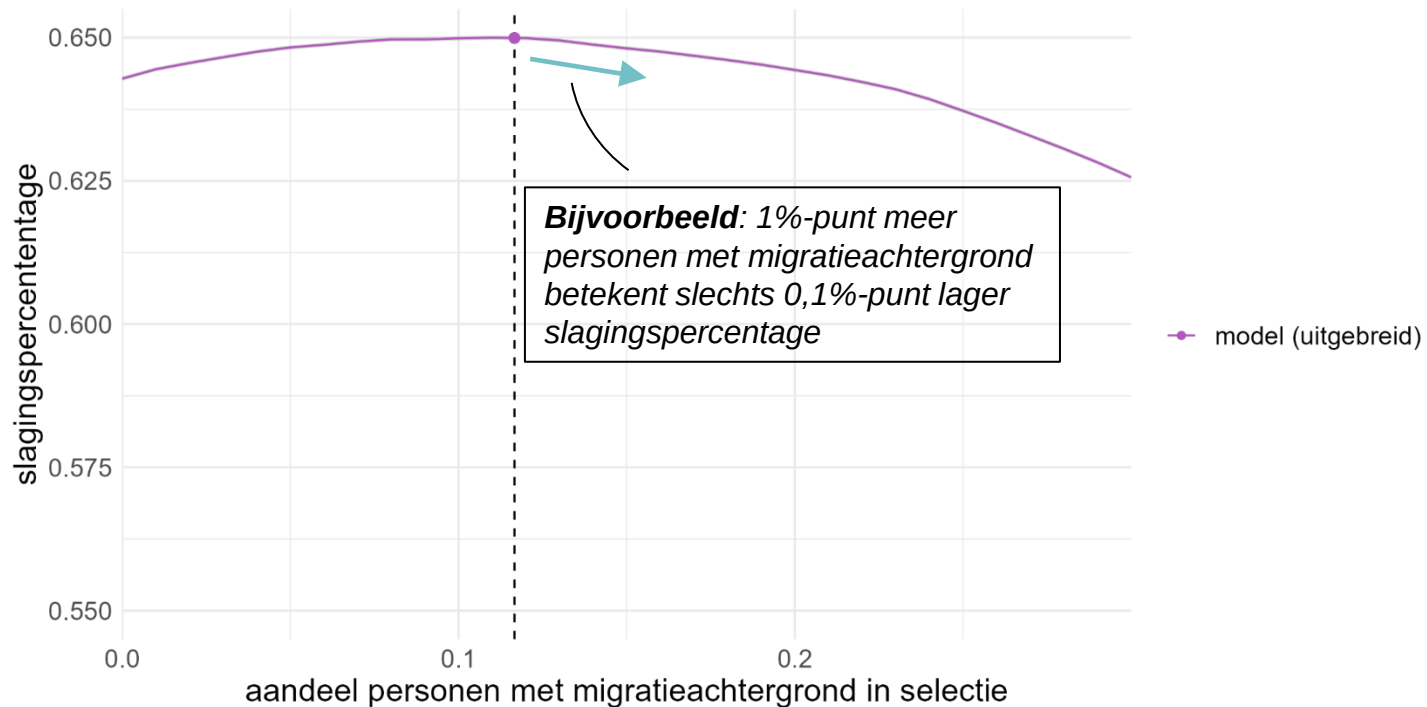
Voorspelmodel (cijfer): slagingspercentage 64%, aandeel migranten 17%



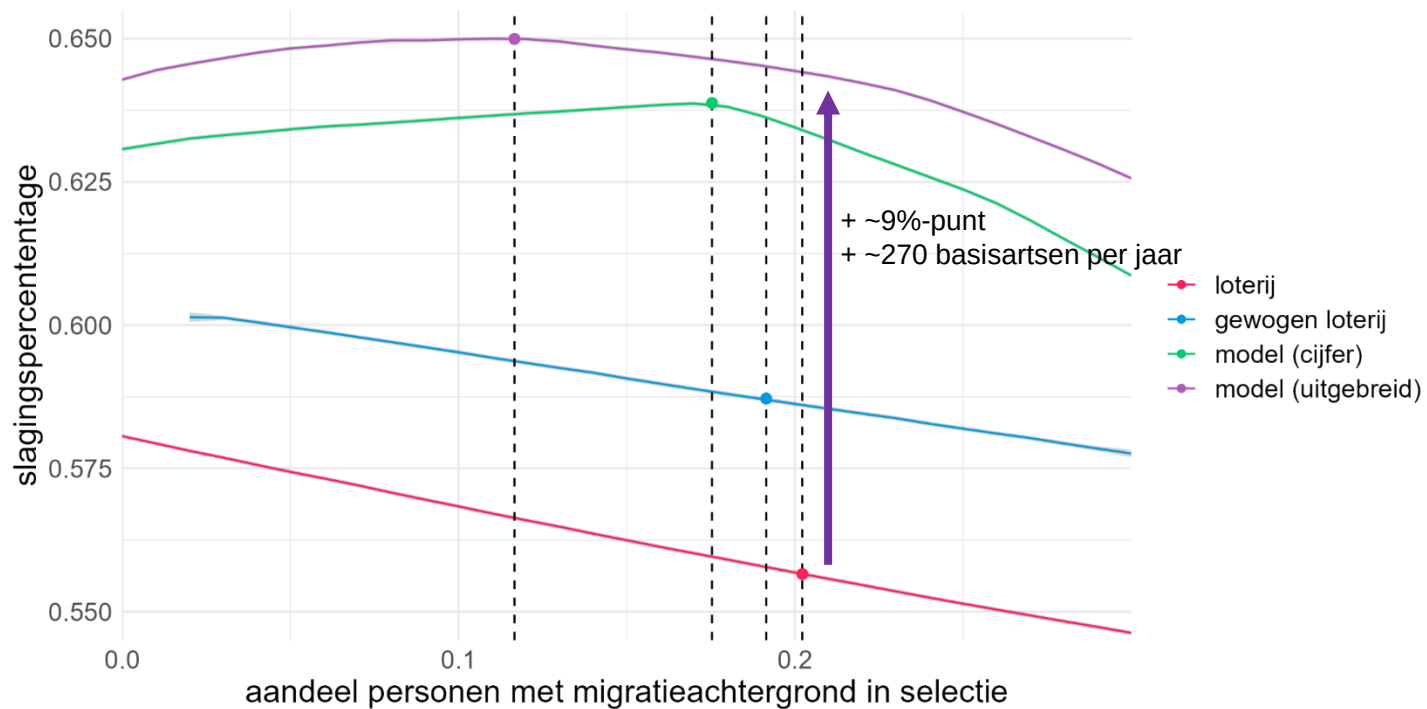
Voorspelmodel (uitgebreid): slagingspercentage 65%, aandeel migranten 12%



We kunnen het aandeel migranten in de selectie bijsturen



De selecties van de voorspelmodellen halen hoger slagingspercentage voor elk aandeel migranten



Onze boodschap

Mooi nieuws

Selectie door middel van algoritmes is mogelijk zonder uitsluiting van kwetsbare groepen

- Combinatie van voorspelkracht met representativiteit
- Door selectie-algoritmes te testen kun je in kaart brengen hoe de uitkomst en representativiteit samenhangen

Bijvoorbeeld: 1%-punt meer personen met migratieachtergrond betekent slechts 0,1%-punt lager slagingspercentage

Moeilijk nieuws

Om uitsluiting te voorkomen moet je:

- Weten wat je precies wilt nastreven qua representativiteit
- Gevoelige informatie meenemen om selectie bij te sturen
- Verantwoordelijkheid niet bij het algoritme leggen



Discussie



Tips / take-aways / wijsheden

- Doe altijd ook een aselechte steekproef om een unbiased overzicht te houden
- Ook als je gevoelige data niet in het algoritme kan of mag meenemen, probeer dan in ieder geval te kijken hoe de verdeling in je selectie eruit ziet
- Wees transparant en expliciet over je keuzes wanneer je algoritmes inzet
- De wereld is de wereld, een versimpeling via een model verandert dat niet



Referenties

Kleinberg, J., Ludwig, J., Mullainathan, S., & Rambachan, A. (2018, May). Algorithmic fairness. In *Aea papers and proceedings* (Vol. 108, pp. 22-27).

Rambachan, A., Kleinberg, J., Ludwig, J., & Mullainathan, S. (2020, May). An economic perspective on algorithmic fairness. In *AEA Papers and Proceedings* (Vol. 110, pp. 91-95).

Rambachan, A., Kleinberg, J., Mullainathan, S., & Ludwig, J. (2020). An economic approach to regulating algorithms (No. w27111). National Bureau of Economic Research.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453.

Ketel, N., Leuven, E., Oosterbeek, H., & van der Klaauw, B. (2016). The returns to medical school: Evidence from admission lotteries. *American Economic Journal: Applied Economics*, 8(2), 225-254.

Het selectie-algoritme met loting:

1. Orden **kandidaten willekeurig**
2. Selecteer de **X bovenste kandidaten**

(X = aantal beschikbare plaatsen)

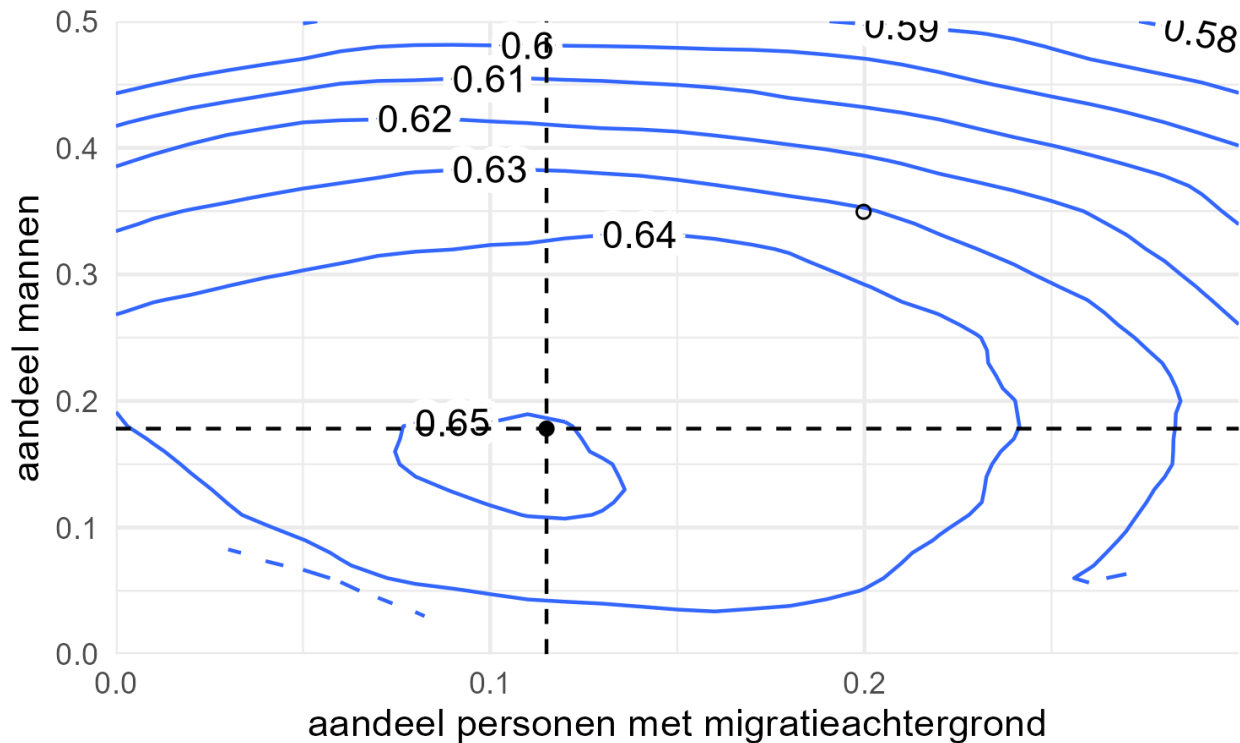
Het datagedreven selectie-algoritme:

1. Train **voorspelmodel** op **kenmerken** en **uitkomsten**
(van in het verleden geselecteerde geneeskundestudenten)
2. Maak **voorspellingen** met **voorspelmodel** en **kenmerken**
(van nieuw aangemelde kandidaten voor geneeskunde)
3. Selecteer **X** hoogste **voorspellingen**
(**X** kandidaten met beste verwachte uitkomst)

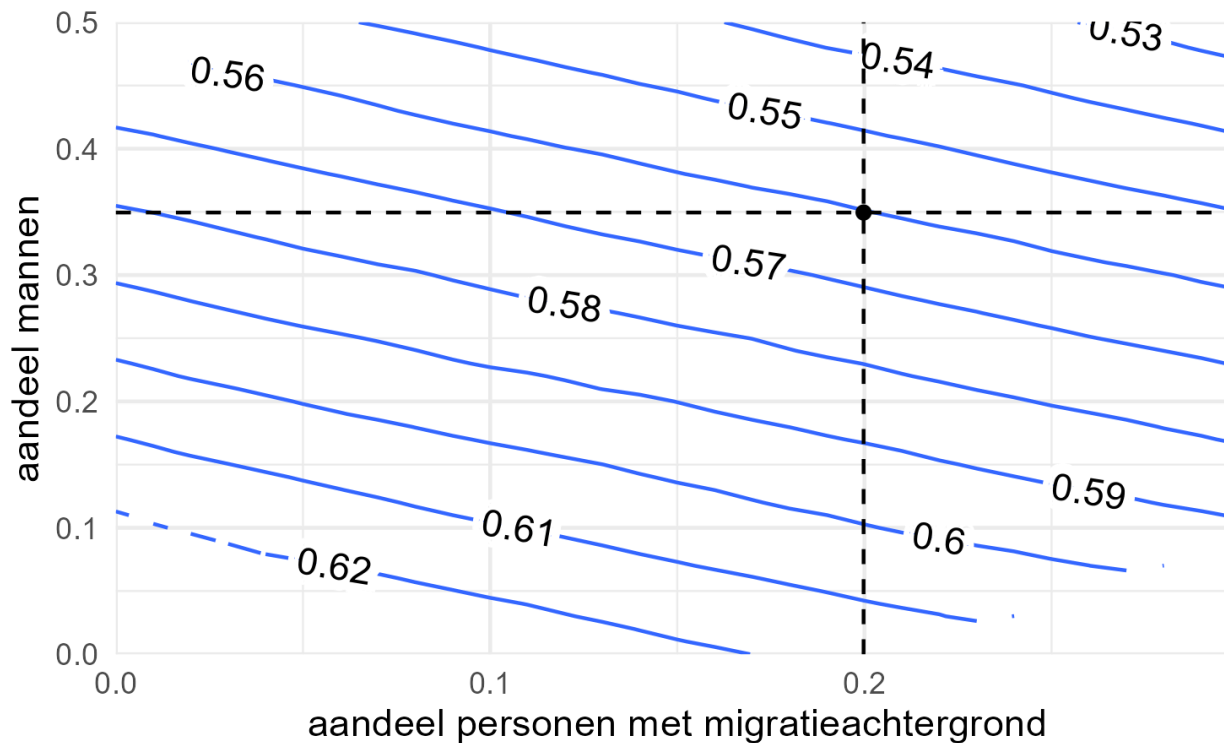
De keuze voor de juiste uitkomst blijft van belang

- Het voorspelmodel richt zich immers puur op die uitkomst
- Bijvoorbeeld:
 - Een model bepaalt het **risico op ziekte** door **zorguitgaven** te voorspellen
 - Mensen met hoog risico krijgen preventieve zorg toegewezen
 - **Maar!** Minderheden hebben vaak minder toegang tot zorg, dus minder **zorguitgaven**
 - Dus krijgen minderheden minder preventieve zorg
- Bias verdwijnt als je uitkomst vervangt met daadwerkelijke **gezondheidsuitkomsten**
- Dit is echt gebeurd in de V.S. (Obermeyer et al., 2019)

Het is mogelijk om op meerdere achtergrondkenmerken tegelijkertijd te selecteren – uitgebreid model



Het is mogelijk om op meerdere achtergrondkenmerken tegelijkertijd te selecteren – loterij





Centraal Planbureau

www.cpb.nl | [@CPBnl](https://twitter.com/CPBnl)

65358979323846264338327950288419716939937510582097494459230781640
520899862803482534211706798214808651328230664709384460955058223172
9408128481117450284102701938521105559644622948954930381964428810975
933446128475648233786783165271201909145648566923460348610454326648
936072602491412737245870066063155881748815209209628292540917153643
92590360011330530548820466521384146951941511633072703657595919530921
7381932611793105118548074462379962749567351885752724891227938183011
29833673362440656643086021394946395224737190702179860943702770539
217176293176752384674818467669405132000568127145263560827785771342757
896091736371787214684409012249534301465495853710507922796892589235420
1995611212902196086403441815981362977477130996051870721134999999837297
804995105973173281609631859502445945534690830264252230825334468503526
19311881710100031378387528865875332083814206171776691473035982534904287
55468731159562863882353787593751957781857780532171226806613001927876611
195909216420198938095257201065485863278865936153381827968230301952035
3018529689957736225994138912497217752834791315155748572424541506959508
2953311686172785588907509838175463746493931925506040092770167113900984
882401285836160356370766010471018194295559619894676783744944825537977
472684710404753464620804668425906949129331367702898915210475216205696
602405803815019351125338243003558764024749647326391419927260426992279
678235478163600934172164121992458631503028618297455570674983850549458
858692699569092721079750930295532116534498720275596023648066549911988
183479775356636980742654252786255181841757467289097777279380008164706
00161452491921732172147723501414419735685481613611573525521334757418494
684385233239073941433345477624168625189835694855620992192221842725502
54256887671790494601653466804956272327917860857843838279679766814541
0095388378636095068006422512517392984896084128488626945604241965
15047
478