Centraal Planbureau

Rechtvaardige algoritmes

Selectie-algoritmes kunnen worden bijgestuurd om representatief te selecteren met behoud van effectiviteit. Belangrijke voorwaarde is wel dat er van tevoren goed wordt nagedacht over de verhoudingen waarin verschillende groepen in de selectie moeten worden opgenomen.

Een economische kijk op selectie-algoritmes maakt de uitruil tussen effectiviteit en representativiteit zichtbaar. Dit helpt bij de afweging hoe algoritmes in te zetten.

CPB - juli 2023

Brinn Hekkelman, Mark Kattenberg, Bas Scheer

Inhoudsopgave

1 Over deze publicatie 4
2 Inleiding 5
3 Achtergrond 6
4 Selectie-algoritmes in economisch perspectief 8
4.1 De afruil tussen effectiviteit en representativiteit 8
4.2 Sturen op representativiteit in de selectie 9
4.2.1 Weglaten van gevoelige informatie uit de dataset 10
4.2.2 Eisen stellen aan de selectie op basis van objectieve statistieken 10
4.2.3 Bijsturen van de representativiteit in de selectie 11
4.3 De drie pijlers van elk datagedreven selectie-algoritme: uitkomstmaat, data en voorspelmethode 12
4.3.1 De uitkomstmaat 12
4.3.2 De data 13
4.3.3 De voorspelmethode 14
4.4 Samenvatting 14
5 Casus: de selectie van geneeskundestudenten 14
5.1 Resultaten 16
5.1.1 Selectie met gewogen loterij en ongewogen loterij 16
5.1.2 Datagedreven selectie 17
5.1.3 Resultaten voor andere uitkomstmaten 19
5.1.4 Resultaten voor andere representativiteitsmaten 21
5.2 Samenvatting 24
6 Discussie 24
Literatuur 27
7 Bijlagen 29
7.1 Opzet selectie-algoritme 29
7.2 Resultaten voor volledige dataset: 1987 - 2004 31

Samenvatting

In selectie-algoritmes voor werkprocessen en besluitvorming zitten steeds vaker voorspelmodellen. Denk daarbij bijvoorbeeld aan HR-afdelingen die algoritmes gebruiken om de meest geschikte sollicitanten te selecteren of aan financiële instelling en uitvoeringsorganisaties die datagedreven algoritmes gebruiken om fraude op te sporen. Deze voorspelmodellen zijn vaak effectief omdat ze toegespitst zijn op het selecteren van de kandidaten die het hoogst scoren op een bepaalde uitkomst.

De focus van voorspelmodellen op een specifieke uitkomst kan er echter ook toe leiden dat bepaalde groepen worden over- of ondergerepresenteerd in de selectie. De vrees bestaat dat selectie-algoritmes hierdoor bepaalde groepen benadelen, zoals minderheden, vrouwen of mensen met een laag inkomen. Dit kan bijvoorbeeld gebeuren als door een algoritme bestaande vooroordelen worden versterkt. Hoe dit risico te beheersen, is een belangrijk discussiepunt in het huidige maatschappelijk debat. In deze publicatie dragen we bij aan dit maatschappelijk debat door een economisch perspectief te bieden op het gebruik van selectie-algoritmes. Dit is een aanvulling op bestaande processen, zoals het Toetsingskader algoritmes en de IAMA.

We laten zien dat selectie-algoritmes kunnen worden gebruikt om representatief te selecteren met behoud van effectiviteit. Als voorbeeld illustreren we hoe je op representativiteit kan bijsturen in de selectie van geneeskundestudenten. In plaats van gevoelige informatie weg te laten, sturen we de uitkomsten van selectie-algoritmes bij op een gewenste mate van representativiteit. Wat de juiste mate van representativiteit is en op welke uitkomst gestuurd moet worden, zijn vragen die subjectief beantwoord moeten worden.

Gevoelige informatie weglaten uit voorspelmodellen heeft niet altijd het gewenste effect. Deze gegevens weglaten lijkt een voor de hand liggende en simpele oplossing, maar het biedt geen garantie voor succes. Onderzoek laat zien dat in sommige gevallen het onthouden van gevoelige informatie – zoals etniciteit – juist leidt tot meer uitsluiting van minderheden. Door gevoelige informatie niet mee te nemen kun je het zicht op benadeling van bijvoorbeeld geneeskundestudenten met een migratieachtergrond verliezen.

Een belangrijk inzicht uit economisch wetenschappelijke studies is dat rechtvaardigheidsoverwegingen, zoals kansengelijkheid of representativiteit, direct in de uitkomsten van een selectie-algoritme verwerkt kunnen worden. Als je weet in welke verhouding je kandidaten met een migratieachtergrond in de selectie van geneeskundestudenten wilt opnemen, dan kun je dat als voorwaarde opleggen aan een selectie-algoritme. Zo kan de voorspelkracht van een voorspelmodel worden benut zonder dat dit ten koste gaat van de representativiteit van studenten met een migratieachtergrond. Dit bijsturen van het selectie-algoritme vereist wel dat er expliciet een keuze wordt gemaakt over de gewenste representativiteit in de uiteindelijke selectie.

Wat de juiste representativiteit van de selectie is, staat niet vast. Als je bij migratieachtergrond voor proportionaliteit kiest, doe je dat dan op basis van de personen die zich aanmelden of op basis van de hele bevolking? En richt je alleen op migratieachtergrond of ook op bijvoorbeeld genderverhoudingen? Omdat er geen objectief juist antwoord op dit soort vragen bestaat, vereist de opzet van het algoritme een discussie.

We laten zien dat een selectie-algoritme voor geneeskundestudenten op basis van een bijgestuurd voorspelmodel betere uitkomsten geeft dan loting, voor elke mate van representativiteit. Bij de selectie van geneeskundestudenten werd in het verleden een gewogen loting toegepast. Ter vergelijking gebruiken we een voorspelmodel om te selecteren op verwachte uitkomsten en vervolgens sturen we in de selectie het aandeel personen met een migratieachtergrond bij. Door het voorspelmodel bij te sturen daalt het verwachte slagingspercentage – de effectiviteit – slechts beperkt, terwijl de representativiteit hetzelfde wordt als bij loting. Ten opzichte van de loting zorgt bijgestuurde selectie voor 300 extra afgestudeerde basisartsen per jaar.

1 Over deze publicatie

Algoritmes worden steeds vaker gebruikt, maar hun inzet is nooit neutraal of objectief. Dit komt bijvoorbeeld door de keuze voor welke doel zo'n algoritme ontwikkeld wordt. Of, bij een voorspellende selectie-algoritme, de keuze welke uitkomst en *input* data het voorspellende algoritme gebruikt. Deze data reflecteert gewenst én ongewenst gedrag van mensen, waardoor het risico bestaat dat deze algoritmes ongewenst gedrag versterken. De laatste tijd zijn er in Nederland helaas meerdere voorbeelden geweest waarin dit gebeurd lijkt te zijn.

In deze publicatie behandelen we het tegengaan van vooringenomenheid (*bias*) van algoritmes vanuit economisch perspectief. Het onderwerp sluit aan bij de wens van het CPB om maatschappelijk relevante publicaties te schrijven binnen het thema 'divergentie en ongelijkheid'.[1] Als economen hebben we de kennis om over dit onderwerp te schrijven, omdat we vaak modelleren. Hierbij komen we veel in aanraking met het opsporen en tegengaan van *bias*.[2]

Het economisch perspectief om algoritmes eerlijker te maken heeft toegevoegde waarde. Het bijsturen van de representativiteit van de selectie wordt niet besproken in de bestaande hulpmiddelen en handreikingen om te voorkomen dat voorspellende algoritmes discrimineren. Wij denken dat het opnemen van deze methode toegevoegde waarde heeft, omdat het een economische oplossing biedt om discriminatie door algoritmes te beperken. Hierin staat de afruil tussen effectiviteit en representativiteit centraal. Deze methode kan worden gebruikt in het onderdeel van de IAMA waarin besproken wordt of, en in hoeverre, de inbreuk van voorspellende algoritmes op mensenrechten opweegt tegen de effectiviteit van het algoritme.

Het door ons gehanteerde economisch perspectief verschilt fundamenteel van het juridisch perspectief. Het economisch perspectief gaat over de vraag 'hoe de effectiviteit van algoritmes te maximaliseren gegeven een gewenste mate van representativiteit'. Hieruit volgt dat omstreden variabelen, zoals tweede nationaliteit of leeftijd, gebruikt mogen worden als zij de voorspelkracht van het model verbeteren, mits je voldoet aan de eisen aan representativiteit. Dit verschilt sterk van het juridische perspectief dat voorschrijft aan welke wettelijke eisen een algoritme, en de door haar gebruikte data, moeten voldoen om ingezet te mogen worden. In deze publicatie laten we zien dat het tegengaan van vooringenomenheid van algoritmes negatieve gevolgen heeft voor hun verwachte effectiviteit.

We presenteren de 'bijgestuurde selectiemethode' om de economische uitruil inzichtelijk te maken. We maken de methode concreet met een rekenvoorbeeld waarin we deze methode gebruiken om studenten te selecteren voor de opleiding Geneeskunde. Hieruit blijkt dat het mogelijk is om een algoritme bij te sturen, zodat het evenveel studenten met een migratieachtergrond toelaat als bij loting. Tegelijkertijd zorgt het bijgestuurde algoritme echter voor meer afgestudeerde studenten.

[1] Zie het werkplan 2023 van het Centraal Planbureau.
[2] Ter illustratie: David Card, Joshua Angrist en Guido Imbens ontvingen in 2022 de Nobelprijs voor de economie voor hun methodologische bijdragen aan het analyseren van causale effecten. Hun werk richt zich op de vraag hoe causale effecten zonder *bias* gemeten kunnen worden als observationele data gebruikt worden.

2 Inleiding

In selectie-algoritmes voor werkprocessen en besluitvorming zitten steeds vaker voorspelmodellen. Denk daarbij bijvoorbeeld aan HR-afdelingen die algoritmes gebruiken om de meest geschikte sollicitanten te vinden of het gebruik van datagedreven algoritmes om fraude op te sporen door financiële instellingen en uitvoeringsorganisaties. Deze voorspelmodellen zijn vaak effectief omdat ze toegespitst zijn op het selecteren van de kandidaten die het hoogst scoren op een bepaalde uitkomst, waardoor de verwachte uitkomst zo hoog mogelijk is. De voorspelkwaliteit van voorspelmodellen is over de jaren steeds beter geworden dankzij de toegenomen hoeveelheid beschikbare data en rekenkracht. Daardoor worden ze voor steeds meer taken ingezet. In het kader 'wat is een selectie-algoritme?' gaan we in op het onderscheid tussen selectie-algoritmes en voorspelmodellen.

De focus van voorspelmodellen op een specifieke uitkomst kan er echter ook toe leiden dat bepaalde groepen worden over- of ondergerepresenteerd in de selectie. Als er in de data namelijk een correlatie tussen de uitkomst en een groep bestaat, zal het voorspelmodel dat gebruiken en personen uit die groep vaker of minder vaak selecteren. De vrees bestaat dat voorspellende selectie-algoritmes hierdoor bepaalde groepen, zoals minderheden, vrouwen of lage inkomens, benadelen. Dit kan bijvoorbeeld gebeuren als bestaande vooroordelen worden versterkt. Hoe dit risico te beheersen, is een belangrijk discussiepunt in het huidige maatschappelijk debat. In deze publicatie dragen we bij aan dit maatschappelijk debat door een economisch perspectief te bieden op het gebruik van selectie-algoritmes.

Een belangrijk inzicht uit economisch wetenschappelijke studies is dat rechtvaardigheidsoverwegingen, zoals kansengelijkheid of representativiteit, direct in de uitkomsten van een selectie-algoritme verwerkt kunnen worden. Een voorspelmodel zal een selectie maken waarbij de verwachte effectiviteit het hoogst is, ongeacht de representativiteit van die selectie, terwijl een selectie die alleen representatief is geen goede verwachte effectiviteit hoeft te hebben. Echter, als je weet in welke verhoudingen je verschillende groepen in de selectie wilt opnemen, dan kun je die voorwaarden opleggen aan de uitkomsten van het selectie-algoritme. Zo kan de voorspelkracht van gebruikte voorspelmodellen worden benut zonder dat dit ten koste gaat van de representativiteit van de geselecteerde groep. Deze aanpak vereist dat er keuzes worden gemaakt over de gewenste uitkomsten van een selectie-algoritme. Vragen die hierbij een rol kunnen spelen zijn: welke groepen neem je mee als je kijkt naar representativiteit? Richt je je op genderverhoudingen, op migratieachtergrond, of op allebei? En wat is het aandeel van elke groep dat je in de selectie wilt garanderen? Omdat er geen objectief juist antwoord op dit soort vragen bestaat, moet daarin subjectief keuzes gemaakt worden. Daarom vereist de opzet van het algoritme een discussie.

Door het selectie-algoritme uitgebreid te testen, kun je nagaan of er specifieke groepen worden over- of ondergeselecteerd. Ook kun je zo in kaart brengen hoe de gekozen uitkomstmaten veranderen als je de selectie van het algoritme aanpast. Bovendien kan aan de hand van deze informatie worden bepaald welke voorwaarden er moeten worden opgelegd aan het selectie-algoritme om ongewenste selectie te verminderen. Pas na deze checks kan het in de praktijk worden ingezet. Uiteraard moeten ook de uitkomsten uit de praktijk worden gemonitord, zodat het selectie-algoritme continu kan worden aangepast op basis van de meest recente uitkomsten.

Deze publicatie is een verrijking van de bestaande richtlijnen om de risico's van selectie-algoritmes te ondervangen. De publicaties over bestaande richtlijnen worden in hoofdstuk 3 besproken. Het economisch perspectief op algoritmes bespreken we in hoofdstuk 4. In hoofdstuk 5 volgt een casus waarin we de inzichten uit hoofdstuk 4 toepassen op de selectie van geneeskundestudenten. Hoofdstuk 6 sluit af met een discussie.

Wat is een selectie-algoritme?

Een algoritme is een vaste set stappen die doorlopen wordt om een uitkomst te bereiken. Een selectie-algoritme is dus een vaste set stappen om kandidaten te selecteren. Algoritmes worden vaak vergeleken met een recept voor een bepaald gerecht, waarbij de handelingen die in het recept beschreven staan de stappen van het algoritme zijn. Het gerecht is dan de uitkomst van het algoritme. Een selectie-algoritme kan heel simpel zijn, bijvoorbeeld "selecteer willekeurig honderd kandidaten" of "selecteer de honderd kandidaten met het hoogste eindexamencijfer".

Een selectie-algoritme kan gebruikmaken van een voorspelmodel. Een voorspelmodel is een (mogelijk complex) algoritme dat probeert om een bepaalde uitkomst zo goed mogelijk te voorspellen, gegeven een aantal inputvariabelen. Wanneer een voorspelmodel gebruikt wordt in een selectie-algoritme, selecteert dat selectie-algoritme doorgaans die kandidaten die volgens het voorspelmodel de beste verwachte uitkomst hebben.

Niet ieder selectie-algoritme gebruikt een voorspelmodel. Loting en gewogen loting zijn immers ook selectie-algoritmes. Zelfs een selectie-algoritme zonder voorspelmodel kan *bias* versterken. Zo worden mensen met een migratieachtergrond minder vaak geselecteerd voor de studie geneeskunde als de loting gewogen wordt naar eindcijfer.

3 Achtergrond

De Nederlandse Grondwet verbiedt discriminatie, al zijn er wettelijke uitzonderingen. Artikel 1 van de Nederlandse Grondwet stelt dat: '*Allen die zich in Nederland bevinden, worden in gelijke gevallen gelijk behandeld. Discriminatie wegens godsdienst, levensovertuiging, politieke gezindheid, ras, geslacht of op welke grond dan ook, is niet toegestaan*'. Hiermee is directe en indirecte discriminatie 'op welke grond dan ook' in principe verboden.[3] Tegelijkertijd zijn er wettelijke uitzonderingen – arbeid door kleine kinderen is verboden – en biedt het non-discriminatiebeginsel de mogelijkheid om onderscheid te maken op basis van persoonskenmerken, mits het onderscheid gemaakt wordt op 'relevante factoren' en het onderscheid 'objectief gerechtvaardigd' kan worden.[4] Dit houdt in dat het gebruik van selectie-algoritmes in principe verboden is, tenzij het gemaakte onderscheid objectief gerechtvaardigd is en de gebruikte kenmerken relevante factoren zijn. De afgelopen jaren zijn selectie-algoritmes veel in het nieuws geweest, omdat men vreest dat ze discrimineren. Bijvoorbeeld het anti-fraudesysteem SyRI[5], het gebruik van tweede nationaliteit door de Belastingdienst[6] of het risico dat algoritmes in het algemeen discrimineren[7].

[3] Men spreekt van directe discriminatie wanneer personen in vergelijkbare situaties ongelijk behandeld worden. Indirecte discriminatie betreft de situatie waarin 'een ogenschijnlijk neutrale bepaling, maatstaf of handelswijze een groep personen in vergelijking met andere groepen bijzonder treft' (Van der Sloot e.a., 2022).
[4] Het maken van onderscheid is 'objectief gerechtvaardigd' wanneer er een legitieme reden is om onderscheid te maken, wanneer het maken van onderscheid proportioneel is [gegeven het doel] en wanneer er geen minder ingrijpende middelen tet beschikking staan om het doel te bereiken (Van der Sloot e.a., 2022).
[5] "Anti-fraudesysteem SyRI moet van tafel, overheid maakt inbreuk op privéleven" (NOS, 2020)
[6] "Institutioneel racisme bij fraudedienst van de fiscus. dit was er aan de hand" (NU.nl, 2022)
[7] "Slecht nieuws: algoritmes discrimineren anno 2021 nog steeds" (Brandpunt, 2021)

De afgelopen jaren is er een sterke beleidsreactie geweest om de risico's van deze algoritmes in kaart te brengen en in te dammen. Zo is de Autoriteit Persoonsgegevens sinds kort ook toezichthouder op algoritmes. Daarnaast zijn er verschillende publicaties uitgekomen die moeten helpen om discriminatie door algoritmes tegen te gaan. Daarin wordt besproken hoe algoritmes zouden moeten worden ontworpen en gecontroleerd, zoals in de Handreiking non-discriminatie *by design* (Van der Sloot et al., 2021) of Aandacht voor Algoritmes (Algemene Rekenkamer, 2023). Deze publicaties zijn uitgebreid met procedurele waarborgen vooraf, zoals de Impact Assessment Mensenrechten en Algoritmes (IAMA, zie Gerards et al., 2021) en de Data Privacy Impact Assessment (DPIA, zie Autoriteit Persoonsgegevens, 2023). Deze publicaties en procesafspraken hebben tot doel om vooraf duidelijkheid te scheppen of de inzet van een algoritme gewenst en toegestaan is. Dit vereist vaak een afweging tussen de mate waarin een algoritme haar doel bereikt en de verwachte impact op mensenrechten. Hieronder lichten we dit toe aan de hand van de IAMA.

De IAMA is een vragenlijst over de doelen, de vormgeving en het gebruik van algoritmes en hoe al deze aspecten doorwerken op mensenrechten. De IAMA structureert de discussie hierover en helpt zo de besluitvorming over de inzet van algoritmes binnen (specifiek) overheidsorganen. Aan de hand van de IAMA wordt een discussie gevoerd over de verwachte effectiviteit en impact op mensenrechten. Dit gebeurt *voordat* het algoritme gebruikt wordt, volgens het adagium 'bezint eer ge begint'.

In de regel zal een algoritme dat mensen selecteert vaak bepaalde mensenrechten raken. Dit geldt bijvoorbeeld voor het recht op een gelijke behandeling of het recht op persoonsgegevens-bescherming. Dit gebeurt wanneer algoritmes profileren op basis van tweede nationaliteit[8] of als algoritmes getraind en ingezet worden op data met persoonskenmerken. De afweging tussen enerzijds de doelen van het gebruik van het algoritme en anderzijds het raken van mensenrechten is dan ook een belangrijk onderdeel van de IAMA. Om deze afweging goed te kunnen maken, is niet alleen nodig om te weten *of* mensenrechten met het algoritme geraakt worden, maar juist ook *in hoeverre* mensenrechten geraakt worden. Om de doelstelling van het algoritme goed af te kunnen wegen tegen de impact op de mensenrechten is het bijvoorbeeld van belang om te weten of het algoritme vijf of vijftig procentpunt meer mensen met een migratieachtergrond selecteert. Helaas is het moeilijk om vooraf te bepalen in hoeverre het gebruik van een algoritme bepaalde groepen benadeelt, wat de kracht van de IAMA beperkt.

De IAMA onderkent dat het moeilijk is om verschillende algoritmes te vergelijken als ze variëren in de mate waarin mensenrechten geraakt worden. Dit komt ook doordat algoritmes kunnen verschillen in de mate waarin doelen behaald worden. De IAMA benadrukt dat hierover een 'open discussie' gevoerd moet worden:

"*Er moet een afweging gemaakt worden tussen het relatieve gewicht van het grondrecht dat op het spel gaat en het relatieve gewicht van de maatschappelijke doelen die worden nagestreefd.*" (IAMA, p77)

"*De vraag kan rijzen wat de trade-off moet zijn als een alternatief iets minder effectief lijkt te zijn, maar wel het grondrecht minder lijkt aan te tasten. Hiervoor zijn er geen harde criteria beschikbaar – hierover zal vooral een open discussie moeten worden gevoerd.*" (IAMA, p76-77)

Het is moeilijk om vooraf een open discussie te voeren over botsende grondrechten, vooral wanneer niet bekend is in hoeverre deze grondrechten geraakt worden. Het is achteraf eenvoudiger om een inschatting te maken van de mate waarin een algoritme haar doelstelling behaalt en het effect dat ze heeft op

[8] Voorspelmodellen mogen in principe geen gebruik maken van persoonskenmerken als ze daardoor onderscheid maken op basis van geslacht, etniciteit, leeftijd, seksuele voorkeur of levensovertuiging, omdat dan sprake is van discriminatie. Dit mag echter wel indien het 'relevante factor' is voor het verkrijgen van goede voorspellingen. Zie Van der Sloot e.a., 2022.

mensenrechten. In deze publicatie beschrijven we hoe de inbreuk op mensenrechten via de 'bijgestuurde selectiemethode' beperkt kan worden. Daarnaast laten we zien hoe hiermee een bepaalde mate van gelijke behandeling gegarandeerd kan worden, ook als voorspellende algoritmes gebruikmaken van vervuilde data die dus ongewenste patronen bevatten. Ongewenste patronen ontstaan bijvoorbeeld als discriminatie in de samenleving terugkeert in de data. Bovendien laten we zien hoe eisen op het gebied van gelijke behandeling doorwerken op de voorspelde effectiviteit van het algoritme. Hiermee kwantificeert de bijgestuurde selectiemethode de uitruil tussen de voorspelde effectiviteit van algoritmes en haar doorwerking op de mensenrechten, specifiek het recht op gelijke behandeling. Dit begrenst niet alleen de impact op mensenrechten via het recht op een gelijke behandeling, maar het kan ook procedurele grondrechten en het recht op behoorlijk bestuur versterken.

De in deze publicatie beschreven 'bijgestuurde selectieprocedure' vult 'procedures vooraf' zoals de IAMA en DPIA daarmee aan. Ook vult ze bestaande 'achteraf-procedures' aan, zoals het Toetsingskader algoritmes (Algemene Rekenkamer, 2023). Dit toetsingskader stelt onder andere voor om data en voorspellingen van algoritmes te testen op discriminatie, maar gaat niet in op de uitruil tussen de voorspelde effectiviteit van het algoritme en de impact op mensenrechten. Daarnaast onderscheidt deze publicatie zich van eerder werk door te focussen op de inzichten vanuit een economische invalshoek, ondersteund met wetenschappelijke artikelen uit dit veld. We richten ons hierbij op selectie-algoritmes. Aan de hand van een casus, de selectie van geneeskundestudenten, laten we zien hoe deze inzichten in de praktijk kunnen worden ingezet.

4 Selectie-algoritmes in economisch perspectief

In dit hoofstuk benaderen we selectie-algoritmes vanuit economisch perspectief. Eerst gaan we nader in op de discussie over effectiviteit en representativiteit door te bekijken hoe die twee met elkaar interacteren. Vervolgens bespreken we verschillende methoden om selectie-algoritmes bij te sturen in zowel effectiviteit als representativiteit, waarbij ook de door ons gebruikte methode voor de casus in hoofdstuk 5 aan bod komt. Om het plaatje compleet te maken, lichten we de drie belangrijkste factoren van selectie-algoritmes toe.

4.1 De afruil tussen effectiviteit en representativiteit

De kwaliteit van een selectieprocedure hangt niet alleen af van de effectiviteit, maar ook van de representativiteit. Een selectieprocedure heeft als doel om een goede selectie te maken. Wat een goede selectie is kan op twee vlakken beoordeeld worden: hoe geschikt zijn de mensen in de selectie (de effectiviteit) en welke kenmerken hebben zij (de representativiteit)? Hoe geschikt iemand is wordt beoordeeld door middel van een bepaalde uitkomst. Omdat dit doorgaans pas in de toekomst duidelijk wordt, moet de selectieprocedure op basis van andere informatie een inschatting maken.

Om de effectiviteit te maximaliseren, kan een selectie-algritme gekozen worden dat een voorspelmodel gebruikt om de uitkomsten zo goed mogelijk te voorspellen. Een voorspelmodel gebruikt hiervoor 'voorspellers' of *input* data: kenmerken van de kandidaten die iets zeggen over wat de uitkomst van hun selectie waarschijnlijk zal zijn. Zo'n voorspelmodel is volledig gericht op het maximaliseren van de gekozen uitkomst.

Naast effectiviteit kan een selectieprocedure ook beoordeeld worden op representativiteit van de selectie. Wordt bijvoorbeeld een subgroep met een bepaald achtergrondkenmerk vaker of minder vaak dan gemiddeld geselecteerd? Het kan bijvoorbeeld onwenselijk zijn dat kandidaten uit een bepaalde regio ondergerepresenteerd worden in de selectie, of het kan bijvoorbeeld wenselijk zijn dat kandidaten met een beperking overgerepresenteerd worden in de selectie.

De selectie die een voorspelmodel maakt, houdt in principe geen rekening met de representativiteit van deze selectie. Wanneer wel rekening gehouden wordt met de representativiteit van de selectie, bijvoorbeeld door de representatie van bepaalde subgroepen te vergroten,verandert dit de gemaakte selectie. Door zo de focus (deels) te verleggen, zal de gewijzigde selectie minder hoog scoren op de uitkomstmaat. In sectie 4.2.3 bespreken we manieren om bij te sturen op representativiteit van de selectie bij het gebruik van voorspelmodellen. Omdat een voorspelmodel enkel de effectiviteit maximaliseert, zal het bijsturen op representativiteit de voorspelde effectiviteit verminderen. Dit neemt niet weg dat de veranderde selectie wenselijk kan zijn, wanneer de grotere representativiteit opweegt tegen het verlies aan effectiviteit.

Maatschappelijke voorkeuren bepalen hoe effectiviteit en representativiteit tegen elkaar afgewogen zouden moeten worden. Waar bijvoorbeeld bij preventief fouilleren de representativiteit een belangrijke rol speelt, is het bij sport juist de effectiviteit die een doorslaggevende rol speelt: de beste sporters worden geselecteerd voor de wedstrijden op de hoogste niveaus. Het beoordelen van de representativiteit én van de effectiviteit van een selectie vergt een subjectieve discussie. Het staat echter vast dat effectiviteit en representativiteit van een selectieprocedure niet los van elkaar staan. Er moet daarom gekozen worden wat de beste selectie is met inachtneming van de afruil tussen effectiviteit en representativiteit. In veel gevallen zal dit betekenen dat een balans tussen de twee gevonden moet worden. Het is niet nodig, en in veel gevallen ook niet wenselijk, om ofwel effectiviteit ofwel representativiteit te maximaliseren en de ander te negeren. Overigens staan de twee niet lijnrecht tegenover elkaar: vaak is het meer een nadruk die op het een of het ander ligt. Het kan zelfs zo zijn dat een selectie zowel qua effectiviteit als qua representativiteit de beste selectie is.

De afruil tussen effectiviteit en representativiteit kan er op termijn anders uitzien. Een voorspelmodel voorspelt uitkomsten in de toekomst op basis van data uit het verleden. Deze data kan bijvoorbeeld de informatie bevatten dat minderheden minder goede uitkomsten hebben, omdat ze door de bestaande meerderheid gemarginaliseerd worden. In dat geval zou het maken van een meer representatieve selectie met een groter aandeel minderheden ertoe kunnen leiden dat de uitkomsten voor de geselecteerde minderheden in de nieuwe situatie beter zijn. De voorspelde effectiviteit is dus altijd een verwachting op basis van data uit het verleden en biedt geen garantie voor toekomstige effectiviteit van een nieuwe selectiecompositie.

4.2 Sturen op representativiteit in de selectie

Er zijn verschillende manieren om representativiteit na te streven in de selectie. In deze publicatie behandelen we er drie:

1. Weglaten van gevoelige informatie uit de dataset
2. Eisen stellen aan de selectie op basis van objectieve statistieken
3. Bijsturen van de representativiteit in de selectie

Deze onderdelen keren niet allemaal terug in de hulpmiddelen om voorspellende algoritmes eerlijker te maken. Zo staan in de IAMA (Gerards et al., 2021) en Handreiking non-discriminatie *by design* (van der Sloot et al., 2021) geen richtlijnen om specifieke verklarende variabelen niet te gebruiken. In plaats daarvan stellen ze voor om te testen of de gebruikte data (voorspellers en uitkomsten) compleet en accuraat zijn en of hun

waarden beïnvloed worden door ongewenste patronen in de samenleving (*measurement bias*). Dit vereist veel kennis over de maatschappelijke processen waaronder de data tot stand komen. Het testen van de werking van algoritmes achteraf is een belangrijk onderdeel binnen de IAMA (Gerards et al., 2021), de Handreiking non-discriminatie *by design* (van der Sloot et al., 2021) en het Toetsingskader Algoritmes (Algemene Rekenkamer, 2023). Deze publicaties stellen voor om algoritmes (statistisch) te testen op hun correcte werking en *fairness bias* en de uitslag van deze tests mee te nemen in de vraag of het algoritme ingezet moet worden.

Het bijsturen van de representativiteit van de selectie wordt niet besproken in de eerdergenoemde hulpmiddelen bedoeld om te voorkomen dat voorspellende algoritmes discrimineren. We bespreken de methode hier wel, omdat het optimaliseren van de effectiviteit gegeven de gewenste mate van representativiteit een economische oplossing biedt om discriminatie door algoritmes te beperken. Bovendien kan deze methode gebruikt worden in het onderdeel van de IAMA waarin besproken wordt of, en in hoeverre, de inbreuk van voorspellende algoritmes op mensenrechten opweegt tegen de effectiviteit van het algoritme.

4.2.1 Weglaten van gevoelige informatie uit de dataset

Het weglaten van gevoelige informatie, zoals persoonskenmerken, is geen garantie dat een voorspelmodel niet discrimineert. Toch wordt dit vaak voorgesteld als oplossing om te voorkomen dat een voorspelmodel direct of indirect discrimineert.[9] Dit lijkt een voor de hand liggende en simpele oplossing, maar biedt geen garantie voor succes. Vaak is gevoelige informatie terug te vinden in andere data die dan als *proxy* van de gevoelige informatie fungeren. Dit leidt tot een zichzelf herhalend probleem waarbij steeds meer data weggelaten moeten worden of waarbij niet meer duidelijk is of de gevoelige informatie uit de data is verdwenen of niet (Kleinberg, Ludwig, Mullainathan, & Sunstein, 2018).

Daarnaast kan gevoelige informatie een voorspelmodel ook representatiever maker. Onderzoek heeft uitgewezen dat in sommige gevallen het meenemen van gevoelige informatie – zoals etniciteit – juist kan leiden tot minder benadeling van minderheden (Rambachan, Kleinberg, Ludwig, et al., 2020). In de literatuur wordt dit *bias reversal* genoemd (Rambachan & Roth, 2019). Hierbij worden bestaande vooroordelen (*bias)* ongedaan gemaakt, omdat bijvoorbeeld de resultaten uit de data voor een onterecht overgeselecteerde groep tegenvallen. Door gevoelige informatie helemaal niet mee te nemen, en er ook niet naar te kijken in de resulterende selectie, kun je het zicht op de mogelijke over- of onderselectie van bepaalde groepen verliezen. Dan ontstaat er juist het risico dat eventuele discriminatie niet wordt ontdekt, terwijl algoritmes juist nuttig kunnen zijn bij het detecteren daarvan (Kleinberg et al., 2020).

Wel kan toegang tot gevoelige informatie onwenselijk zijn vanwege operationele risico's. Bijvoorbeeld om te voorkomen dat persoonsgegevens via een datalek bij de uitvoerder van het algoritme op straat terechtkomen. De beheerder van de data moet zeker kunnen zijn dat de data veilig zijn bij de uitvoerder van het algoritme. Bij zeer gevoelige data kan het hierom een oplossing zijn om ieder gebruik ervan te verbieden. Operationele risico's maken geen deel uit van dit onderzoek, maar worden wel behandeld in andere publicaties over algoritmes, zoals het Toetsingskader algoritmes (Algemene Rekenkamer, 2023).

4.2.2 Eisen stellen aan de selectie op basis van objectieve statistieken

Er zijn veel statistische toetsen ontwikkeld om te beoordelen of de voorspellingen van een voorspelmodel *bias* bevatten.[10] Zulke toetsen vergelijken bijvoorbeeld of de uitkomst vaker optreedt bij

[9] Zo dringt Amnesty er op aan om mensenrechten te beschermen door een verbod op het gebruik van etniciteit en nationaliteit bij het maken van risicoprofielen voor wetshandhaving (Amnesty, 2023). Kamerleden van Denk en D66 hebben dit ook voorgesteld in een (verworpen) motie, zie Tweede Kamer (2021).
[10] Zie (Kleinberg et al., 2016).

mensen voor wie het risico hoog is en of deze relatie voor verschillende groepen even sterk is (*calibration test*). Een andere mogelijkheid is om te kijken of een algoritme vaker aangeeft dat een risico optreedt bij mensen voor wie het risico zich gemanifesteerd heeft. De mate waarin dit het geval is, mag niet verschillen voor verschillende groepen (*balance of the positive class*). Weer een andere toets gaat na of het tegenovergestelde het geval is (*balance of the negative class*). Technisch lijkt het daarmee mogelijk om te beoordelen of een algoritme discrimineert, maar helaas kunnen deze drie tests nooit gelijktijdig waar zijn, behalve in heel triviale gevallen.Daar komt bij dat groepsverschillen in voorspellingen het gevolg kunnen zijn van de gekozen uitkomst, de gekozen data (zowel de voorspellers als de patronen erin), de gekozen schattingsmethode en van structurele verschillen tussen groepen.[11] Als er structurele verschillen zijn tussen groepen is het daarom niet duidelijk of het verwerpen van een statistische toets het gevolg is van *bias* of van structurele verschillen.

4.2.3 Bijsturen van de representativiteit in de selectie

Door de representativiteit van de selectie expliciet bij te sturen, kan deze worden gegarandeerd. Als je weet welk aandeel van de selectie je een bepaalde groep wilt laten krijgen, kun je dat als voorwaarde opleggen aan het selectie-algoritme. Je gebruikt dan een voorspelmodel om een *ranking* van kandidaten naar de verwachte uitkomst te maken, waarna je een selectie maakt op basis van zowel de *ranking* als de voorwaarden op het aandeel. Zo kan de effectiviteit van een voorspelmodel worden benut zonder dat dit ten koste gaat van de representativiteit van de aangewezen groep. Het voordeel hiervan is dat we ons bij het voorspelmodel volledig kunnen richten op het maken van de meest effectieve selectie van de kandidaten. De representativiteit wordt immers gewaarborgd *nadat* het voorspelmodel zijn werk heeft gedaan.

Deze manier om representativiteit in de selectie te borgen, is optimaal vanuit welvaartseconomisch oogpunt wanneer men een voorkeur heeft voor effectiviteit én representativiteit. In het bijzonder leidt deze methode tot een grotere welvaart dan wanneer relevante variabelen weggelaten worden (Kleinberg, Ludwig, Mullainathan, & Rambachan, 2018; Rambachan, Kleinberg, Mullainathan, et al., 2020). Een belangrijke veronderstelling hierbij is dat het voorspelmodel de effectiviteit van een bepaalde groep structureel verkeerd inschat (wat zorgt voor de vooringenomenheid), maar dat de *ranking* van personen binnen de groep accuraat is.

Het is dus mogelijk om effectief te selecteren met voorspelmodellen zonder de representativiteit te schaden. Het moet dan wel vooraf duidelijk zijn in welke verhoudingen je verschillende groepen in de selectie wilt opnemen, waarbij meerdere indicatoren gebruikt kunnen worden om deze groepen te vormen (Bansak et al., 2018). Dit vereist consensus over welke samenstelling de selectie moet hebben. Vanwege de afruil tussen effectiviteit en representativiteit speelt ook de effectiviteit mogelijk een rol (Kleinberg, Ludwig, Mullainathan, & Sunstein, 2018). Wat zijn de gevolgen van de voorwaarden aan de selectie voor de effectiviteit? Dit kan soms pas achteraf goed bepaald worden, maar in het algemeen leiden voorwaarden die de selectie in grotere mate veranderen tot lagere effectiviteit.

Een aantal vragen moet beantwoord worden om de bijgestuurde selectiemethode te gebruiken. Welke groepen neem je mee als je kijkt naar representativiteit? Richt je je op genderverhoudingen, op migratieachtergrond, of op allebei? En wat is het aandeel van elke groep dat je in de selectie wilt garanderen? Streef je bijvoorbeeld proportionaliteit na of hanteer je een ondergrens voor het aandeel? En als je kiest voor proportionaliteit, doe je dat dan op basis van de hele bevolking of op basis van de personen die zich aanmelden?Voor dit soort vragen bestaat geen objectief juist antwoord. Dat er keuzes gemaakt moeten worden is duidelijk, maar wat de juiste antwoorden zijn, dat is subjectief. Zodra deze keuzes gemaakt zijn, kunnen we representativiteit direct in de selectie bijsturen.

[11] Dit wordt het meest toegankelijk uitgelegd in (Kleinberg, Ludwig, Mullainathan, & Sunstein, 2018)

4.3 De drie pijlers van elk datagedreven selectie-algoritme: uitkomstmaat, data en voorspelmethode

Er zijn drie belangrijke factoren die de effectiviteit van een voorspelmodel bepalen. Dit zijn de gekozen uitkomstmaat, de gebruikte *input* data en de gekozen voorspelmethode. In de komende paragraven leggen we uit waarom.

4.3.1 De uitkomstmaat

Het vinden van een geschikte uitkomstmaat is erg belangrijk voor de effectiviteit van de selectieprocedure. Een voorspelmodel focust immers op het maximaliseren van de gekozen uitkomstmaat. In gevallen waarin deze keuze niet goed wordt gemaakt kan dat slecht uitpakken. Hoe dit fout kan gaan, is het makkelijkst te illustreren met een voorbeeld: het gebruik van algoritmes in de gezondheidszorg in de VS (Obermeyer et al., 2019). In de VS wordt gebruik gemaakt van commerciële algoritmes om te bepalen wie aanspraak maakt op bepaalde gezondheidszorg. Obermeyer et al. onderzochten een veelgebruikt algoritme en stelden vast dat het algoritme het risiconiveau van zwarte patiënten lager inschatte dan voor witte patiënten, ook al waren de patiënten even ziek. Aangezien patiënten met een hoger risiconiveau meer zorg krijgen, werden zwarte patiënten hierdoor achtergesteld. Dit gebeurde zonder dat het algoritme wist of een patient wit of zwart is.

De achterstelling van zwarte patienten kwam niet voort uit het gebruik van een indicator voor het behoren tot een minderheidsgroep, maar kwam voort uit een verkeerd gekozen uitkomstmaatstaf. Het voorspelmodel van het algoritme had namelijk gezondheidskosten als uitkomstmaat gekregen. Dit veroorzaakte een *bias*: zwarte patiënten in de VS hebben minder toegang tot zorg en maken daardoor minder gebruik van gezondheidszorg, waardoor hun gezondheidskosten lager liggen bij gelijke gezondheid. Door de uitkomstmaat van het voorspelmodel te vervangen en de daadwerkelijke gezondheid van de patiënten als uitkomstmaat te nemen, viel deze *bias* weg.De juiste uitkomstmaat kiezen, kan dus veel problemen bij de toepassing van selectieprocedures voorkomen. Eventuele *bias* in de uitkomstmaat, zoals in de casus van Obermeyer et al. (2019), moeten namelijk achteraf gecorrigeerd worden door de uitkomstmaat te vervangen of door de representativiteit van de selectie bij te sturen (zoals behandeld in sectie 4.2.3).

Ten slotte is het bij de keuze voor de uitkomstmaat belangrijk om verschillende belanghebbenden van een selectieprocedure te raadplegen. Belanghebbenden kunnen immers verschillende perspectieven hebben op de geschiktheid van een uitkomstmaat. Neem bijvoorbeeld de selectie van geneeskundestudenten. Een opleidingscoördinator zal graag zien dat er veel studenten nominaal de studie afronden. Dus zal diegene het slagingspercentage bij nominale studieduur als geschikte uitkomstmaat zien. Terwijl voor een ziekenhuis vooral het arbeidsaanbod van afgestudeerden belangrijk is. Het bestuur van een ziekenhuis zou dus liever de kans dat iemand gaat werken als dokter als uitkomstmaat nemen.

Ook komt de door de gebruiker gekozen uitkomst niet altijd overeen met het publieke belang (Rambachan, Kleinberg, Mullainathan, et al., 2020). Dit geeft het belang aan van inspraak in de vormgeving van selectiemethoden, zoals ook wordt beschreven in de eerder besproken beleidspublicaties (Gerards et al., 2021; van der Sloot et al., 2021). Dat kan ook helpen om mogelijke *biases* in uitkomstmaten, zoals in het voorgaande voorbeeld, vroeg te herkennen.

4.3.2 De data

Voor een effectief voorspelmodel is het belangrijk dat er veel en goede data beschikbaar zijn, met informatie die helpt bij het voorspellen van de gekozen uitkomstmaat. Een goede voorspeller is zowel voorspellend als motiverend. Een voorbeeld hiervan is het eindexamencijfer. Het eindexamencijfer geeft een goede indicatie wie er de meeste kans maakt om een studie succesvol af te ronden. Bovendien, door bekend te maken dat het cijfer opgenomen wordt in een voorspelmodel geef je scholieren een extra prikkel om goed hun best te doen op school voor een hoger cijfer.

Wel geldt hierbij de beperking dat goede voorspellers niet perfect zijn. Het eindexamencijfer voorspelt de kans dat iemand zijn studie afrondt, omdat het informatief is over iemands intellectuele capaciteiten. Tegelijk kunnen scholieren met lagere intellectuele capaciteiten een hoger gemiddeld cijfer halen voor het eindexamen als ze examentraining krijgen of als iemand uit hun omgeving hen helpt bij het leren dan iemand met hogere intellectuele capaciteiten die deze ondersteuning niet krijgt. Gerelateerd aan dit punt is dat de geobserveerde waarden van de data het gevolg kunnen zijn van ongewenste processen in de maatschappij. Zo melden sommige mensen zich bijvoorbeeld vaker bij de huisarts dan anderen, waardoor medische aandoeningen vaker gevonden en geregistreerd worden. Geregistreerde medische aandoeningen meten dan iemands gezondheid met niet-willekeurige meetfouten. Dit zorgt voor selectieve labels, waardoor een voorspelmodel getraind op mensen die vaak naar de huisarts gaan niet representatief is voor mensen die dat niet doen. Het probleem van selectieve labels en *measurement error* komt vaak voor bij het gebruik van voorspelmodellen (Kleinberg, Lakkaraju, Leskovec, Ludwig, et al., 2018).

Een goede voorspeller moet daarnaast niet zomaar aan te passen zijn door de kandidaat om zo meer kans te maken op selectie. Als kandidaten dat doen, zal die voorspeller snel zijn voorspellende waarde verliezen. Net als bij uitkomstmaten kunnen slechte voorspellers ervoor zorgen dat er niet geselecteerd wordt op datgene wat eigenlijk gewenst is. Als kandidaten bijvoorbeeld een bepaalde certificering die als voorspeller gebruikt wordt illegaal kunnen kopen, verliest deze certificering de eigenlijke waarde als voorspeller van competentie. Het is daarom van belang om de kwaliteit van voorspellingen te blijven monitoren.

Zogeheten 'exogene voorspellers', voorspellers waarvan de waarde niet door de kandidaat te beïnvloeden zijn, kunnen niet door de kandidaat aangepast worden om meer kans te gaan maken in de selectie. Informatie uit registers van de overheid zijn hiervoor geschikt. De buurt waar iemand is geboren kan in dat opzicht een goede voorspeller zijn, omdat dit gegeven correleert met niet-waargenomen kenmerken van de kandidaat. Keuzes van de ouders buiten beschouwing gelaten, kan een kandidaat deze voorspeller nooit aanpassen. Dit voorkomt dat de voorspeller zijn voorspellende waarde verliest als gevolg van strategisch gedrag onder kandidaten. Maar aan de andere kant kan deze voorspeller deterministisch of fatalistisch beleefd worden. Zodra een kandidaat geboren wordt, staat de voorspeller al vast en daarmee ook een deel van de toekomstige selectieprocedure. Dit kan demotiverend zijn en als oneerlijk gezien worden, zelfs als de voorspeller gebruikt wordt om ongelijkheid te voorkomen.

Tot slot kan het voor de representativiteit belangrijk zijn dat juist gevoelige informatie die iets over bepaalde gevoelige groepen zegt beschikbaar is. Dit hebben we al benoemd in sectie 4.2.1. Zonder deze data kan het selectie-algoritme niet worden bijgestuurd op representativiteit en kan ook niet worden gecontroleerd of de uiteindelijke selectie representatief is.

4.3.3 De voorspelmethode

Ook de gekozen voorspelmethode is belangrijk voor de effectiviteit van de selectieprocedure. Dankzij de toename in de beschikbaarheid van data en rekenkracht wordt er steeds meer gebruikgemaakt van geavanceerde voorspelmodellen, zoals *machine learning*-modellen. De ontwikkeling van nieuwe methoden heeft geleid tot betere voorspelkwaliteit van voorspelmodellen. Deze geavanceerde methoden werken echter alleen als ze goed worden afgesteld (*getraind*) en als er veel data beschikbaar zijn. Ook is het bij de geavanceerde methoden vaak lastiger om erachter te komen waarom het voorspelmodel niet goed werkt, bijvoorbeeld als het voor sommige groepen mensen vreemde voorspellingen geeft. Het is daarom goed om verschillende methoden, ook simpele, te testen en de resultaten te vergelijken.

Geavanceerde voorspelmodellen moeten onderhouden worden. Als die over langere tijd worden ingezet moeten de data van de voorspelmodellen regelmatig worden geüpdatet om ervoor te zorgen dat voorspellingen realistisch blijven. De voorspellende waarde van de *input* kan immers over tijd veranderen. Dit kun je doen door een deel van de plaatsen willekeurig toe te wijzen, zodat je de uitkomsten van die groep kunt gebruiken voor het updaten van het voorspelmodel. Het is belangrijk om deze onderhoudskosten in het oog te houden bij het opzetten van een selectie-algoritme.

4.4 Samenvatting

Selectie-algoritmes kunnen beoordeeld worden op zowel effectiviteit (hoe geschikt is de selectie) als representativiteit (wat zijn de kenmerken van degenen in de selectie). Tussen effectiviteit en representativiteit kan een afruil bestaan. Wanneer – gegeven de gebruikte uitkomst en het gekozen algoritme – de *input*data aangepast worden in een poging om het algoritme eerlijker te maken, zal de voorspelde effectiviteit van het algoritme in de regel afnemen.

Bij deze afruil is het vaak niet nodig of wenselijk om ofwel alleen effectiviteit ofwel alleen representativiteit te maximeren. Mede daarom biedt het simpelweg weglaten van gevoelige informatie geen goede oplossing, dat neemt namelijk de mogelijkheid tot bijsturen weg. Een bewust bijgestuurde middenweg kan juist het beste zijn, omdat zo de effectiviteit gemaximaliseerd wordt gegeven de mate van representativiteit en de gebruikte uitkomstmaat, data en het algoritme. Vanuit welvaartseconomisch oogpunt is het beter om zoveel mogelijk informatie te gebruiken om de effectiviteit van selectie-algoritmes te maximaliseren en vervolgens te bepalen wat de representativiteit in de selectie gaat zijn. Die aanpak geeft veel mogelijkheid tot bijsturing en laat bovendien het gebruik van geavanceerde voorspelmodellen toe. De drie belangrijkste factoren die de effectiviteit van zulke voorspelmodellen bepalen zijn de uitkomstmaat, de data en de voorspelmethode. Vaak kiest de ontwikkelaar van het selectie-algoritme deze.

5 Casus: de selectie van geneeskundestudenten

In dit hoofdstuk passen we de inzichten uit het voorgaande hoofdstuk toe op een casus: de selectie van geneeskundestudenten. Hierbij vergelijken we de uitkomsten van de selectie op basis van een gewogen loterij, zoals de selectie in het verleden plaatsvond, met die van een datagedreven voorspelmodel. We beoordelen de selectiemethoden op twee verschillende uitkomstmaten: slagingspercentage en arbeidsdeelname in de medische sector. We laten daarbij ook zien of de verschillende methodes bepaalde

groepen over- of onderselecteren. We kijken daarbij naar: migratieachtergrond, gender en het inkomen van de ouders.

Hieronder beschrijven we de onderzoeksopzet. Voor onze casus hebben we data tot onze beschikking van de centrale gewogen loterij uitgevoerd door de Dienst Uitvoering Onderwijs (DUO) tussen 1987 en 2004, zoals ook gebruikt in het onderzoek van Ketel et al. (2016). We hebben de dataset van DUO verrijkt met administratieve data van het CBS. De dataset is uitgebreid met informatie over de migratieachtergrond, geslacht en het ouderinkomen van de aspirant studenten die meededen aan de loterij. Ook hebben we informatie toegevoegd over behaalde diploma's en arbeidsdeelname in de periode na de loterij. Bij de arbeidsdeelname nemen we zowel werknemers als zelfstandigen mee die actief zijn in de medische sector.

In het verleden werd een gewogen loterij gebruikt om de beperkte opleidingsplaatsen van de studie geneeskunde te verdelen. Omdat het aantal aanmeldingen voor geneeskunde structureel hoger lag (en nog steeds ligt) dan het aantal beschikbare plaatsen, moesten studenten geselecteerd worden.[12] Tussen 1987 en 2004 werd dit gedaan met een centrale gewogen loterij. Het idee van de gewogen loterij is dat iedere kandidaat kans maakt op toelating tot de studie, maar dat kandidaten met een hoger eindexamencijfer meer kans maken dan een kandidaat met een laag cijfer.

In onze casus vergelijken we vier selectie-algoritmes. Dit zijn: een loterij, een gewogen loterij, een simpel voorspelmodel en een uitgebreid voorspelmodel. In de loterij wordt toelating willekeurig toegewezen: elke kandidaat heeft evenveel kans om toegelaten te worden. Bij de gewogen loterij is toelating ook willekeurig, maar maken kandidaten die in een hogere cijfercategorie vallen meer kans op toelating. We hebben de gewogen loterij toegepast zoals die tussen 1987 en 2004 daadwerkelijk heeft plaatsgevonden.[13]

Bij de selectie met een voorspelmodel gebruiken we de kenmerken van de kandidaten om te bepalen wie er toegelaten wordt. Dit gebeurt in twee stappen: eerst geven we het model de kenmerken en de uitkomst van eerdere toegelaten personen. Hier bepaalt het voorspelmodel welke kenmerken relevant zijn en leiden tot een hogere kans op een goede uitkomst, zoals bijvoorbeeld afstuderen binnen de tijd die daarvoor staat. In de tweede stap passen we het model toe op kenmerken van nieuwe aanmelders. Hier selecteert het voorspelmodel nieuwe aanmelders die op basis van hun kenmerken de meeste kans hebben op een goede uitkomst. Per jaar zijn er ongeveer 2800 plekken beschikbaar voor de geneeskundestudie.

Het simpele voorspelmodel selecteert alleen op basis van cijfercategorie, het uitgebreide voorspelmodel gebruikt aanvullende kenmerken. Het simpele voorspelmodel gebruikt dus dezelfde informatie als de gewogen loterij. Het uitgebreide voorspelmodel maakt niet alleen gebruik van cijfercategorie, maar ook van aanvullende kenmerken: vooropleiding (bijv. vwo of propedeuse hbo), leeftijd bij aanmelding, geslacht, migratieachtergrond en ouderinkomen.

De lotingsdata zijn bij uitstek geschikt om de voorspelmodellen te schatten, omdat de selectie willekeurig tot stand is gekomen. Gegeven hun eindcijfer verschillen ingelote en uitgelote studenten niet van elkaar op het gebied van achtergrondkarakteristieken, zoals motivatie of aanleg. Dat betekent dat het voorspelmodel geschat op de uitkomsten van de ingelote studenten ook bruikbaar is voor studenten die uitgeloot zijn. We kunnen er dus van uitgaan dat als we ons voorspelmodel schatten alleen op ingelote

[12] In ons voorbeeld is sprake van een vast aantal beschikbare plaatsen (*numerus fixus*). In andere toepassingen kan dit minder vast liggen. In dat geval wordt er vaak een selectiecriterium gebruikt om mensen te selecteren: als je een bepaalde waarde van de uitkomstmaat overstijgt, dan wordt je geselecteerd. Bijvoorbeeld: als je gemiddelde eindexamencijfer hoger is dan een 8, dan wordt je toegelaten. Ook in dat soort gevallen kan de bijgestuurde selectiemethode worden toegepast, met als verschil dat de hoogte van het selectiecriterium moet worden gekozen in plaats van het aantal beschikbare plaatsen.

[13] Voor de technische details van het selectieproces in de casus, zie bijlage 1 (sectie 7.1).

studenten, waarvoor we de uitkomsten kennen, de resultaten hetzelfde zijn als wanneer we het voorspelmodel toepassen op studenten waarvoor we de uitkomsten niet kennen: nieuwe aanmeldingen. Om deze analyse correct uit te voeren, moeten we er dus wel rekening mee houden dat de loterij gewogen is, wat betekent dat de toelatingskans afhangt van het eindexamencijfer. Studenten met hogere cijfers hadden een grotere kans op toelating. Bij het schatten van het voorspelmodel kunnen we hiervoor corrigeren, aangezien we precies weten hoe de kansen verschillen tussen cijfercategorieën. Door studenten die minder kans maakten op toelating meer gewicht te geven, corrigeren we voor het feit dat er te weinig van deze studenten in de dataset met ingelote studenten zitten. Zo zorgen we ervoor dat de dataset waarop we het voorspelmodel trainen vergelijkbaar is met de data van alle aanmeldingen.

Ten slotte beperken we ons tot de periode 2000-2004 om er zeker van te zijn dat de studentenpopulatie in onze dataset over de jaren vergelijkbaar is. Als we de hele dataset zouden meenemen, lopen we het risico dat het voorspelmodel niet goed wordt geschat, bijvoorbeeld doordat de voorspellende waarde van de *inputs* van het voorspelmodel over tijd veranderen. Om dit soort tijdseffecten zo veel mogelijk uit te sluiten, laten we de resultaten zien op de meest recente vijf jaar van onze dataset. De hoofdanalyse op de hele periode is beschikbaar in de bijlage, sectie 7.2; de resultaten daarvan zijn vergelijkbaar met de analyse op de beperkte dataset.

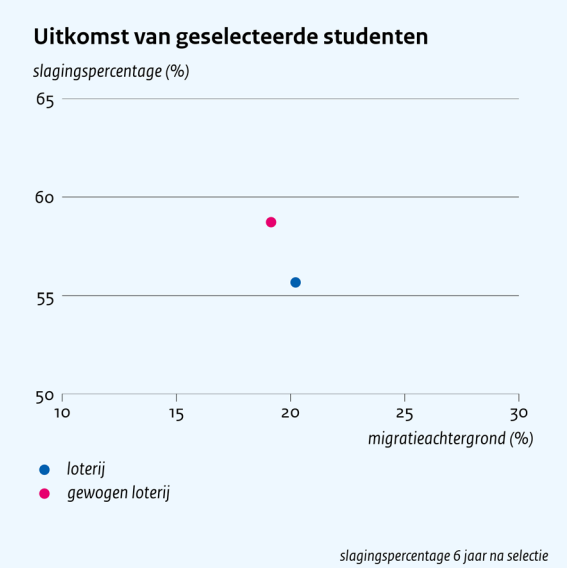
5.1 Resultaten

5.1.1 Selectie met gewogen loterij en ongewogen loterij

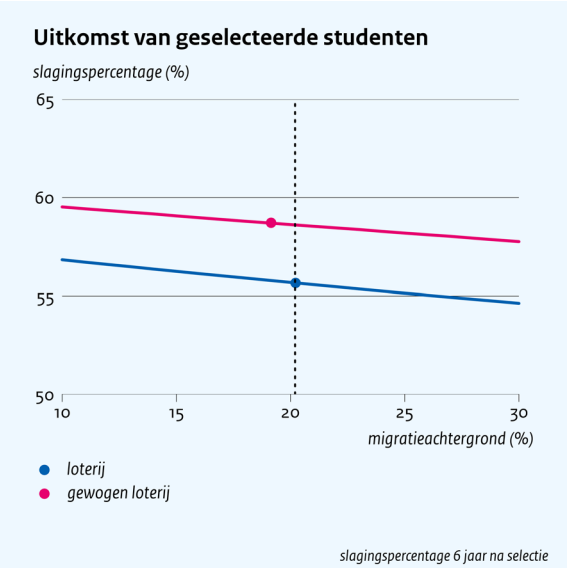
Zoals eerder gezegd werd voor de selectie van geneeskundestudenten gebruikgemaakt van een gewogen loterij om te bepalen wie er toelating kreeg tot de studie. Daarbij gold dat scholieren met een hoog eindexamencijfer meer kans maakten dan scholieren met een laag cijfer. Een indirect effect van deze opzet is dat personen met een migratieachtergrond minder kans maakten op toelating, omdat ze gemiddeld lagere cijfers haalden. Daardoor lag het aandeel personen met een migratieachtergrond lager in de selectie dan in de aanmeldingen.Dit zien we ook terug in onze resultaten in figuur 1: de selectie met een gewogen loterij heeft een kleiner aandeel studenten met een migratieachtergrond (19,1%) dan de selectie met een ongewogen loterij (20,2%). Daar staat tegenover dat het slagingspercentage hoger ligt bij de gewogen loterij (58,9%) dan bij de ongewogen loterij (55,8%). Dit komt doordat de gewogen loterij meer kandidaten met een hoog eindexamencijfer selecteert, en die hebben een hogere slagingskans. Bij een studentenpopulatie van 2800 per jaar betekent dat 87 extra afgestudeerde basisartsen.

Het is echter mogelijk om het aandeel studenten met een migratieachtergrond in de selectie bij te sturen, zonder dat dit veel afbreuk doet aan het slagingspercentage. Zo kan met elke selectiemethode elk aandeel studenten met een migratieachtergrond worden bereikt. Hoe dat werkt, staat beschreven in sectie 4.2.3. Deze resultaten zijn te zien in figuur 2: de lijn geeft het slagingspercentage voor elk aandeel studenten met migratieachtergrond in de selectie. Elk punt op de lijn kan worden gekozen door vast te leggen dat het bijbehorende aandeel studenten met migratieachtergrond moet worden geselecteerd. Bijvoorbeeld: als we bepalen dat het aandeel studenten met migratieachtergrond gelijk moet zijn aan 20,2% -het aandeel in de aanmeldingen – dan heeft de selectie van de gewogen loterij een slagingspercentage van 58,6% en die van de ongewogen loterij 55,7%. Het bijgestuurde gewogen lotingalgoritme heeft dus een hoger slagingspercentage dan het selectie-algoritme met ongewogen loting, terwijl het aandeel studenten met migratieachtergrond gelijk is. We zien zelfs dat de lijn voor de gewogen loterij voor elk aandeel met migratieachtergrond boven de lijn van de loterij ligt. Dit betekent dat de gewogen loterij voor elk gekozen aandeel met migratieachtergrond een hoger slagingspercentage geeft dan de ongewogen loterij. Dit komt doordat de gewogen loterij meer scholieren selecteert met een hoog eindexamencijfer en die scholieren hebben een hoger slagingspercentage.

Figuur 1 - Gewogen loterij geeft hoger slagingspercentage, maar minder studenten met migratieachtergrond



Figuur 2 – Door het aandeel studenten met een migratieachtergrond bij te sturen zijn meer uitkomsten mogelijk



5.1.2 Datagedreven selectie

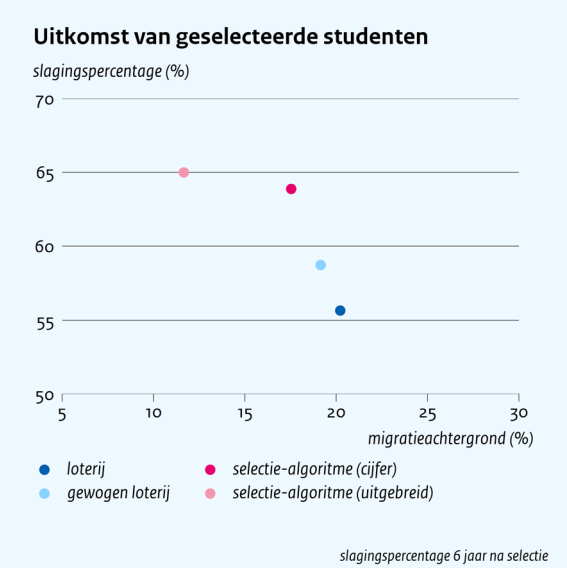
De resultaten tonen dat de datagedreven voorspelmodellen effectiever studenten selecteren dan de (gewogen) loterij. Figuur 3 laat de uitkomsten van de verschillende selectiemethoden zien, als we geen restricties opleggen op het aandeel studenten met een migratieachtergrond. De selectie van het simpele voorspelmodel op basis van eindexamencijfer heeft een hoger slagingspercentage dan de gewogen loterij (63,9% om 58,7%), maar selecteert minder studenten met migratieachtergrond (17,5% om 19,1%). Het simpele voorspelmodel behaalt dit hogere slagingspercentage door kandidaten met het hoogste cijfer voorrang te geven. Dit is anders dan de gewogen loterij, waarin kandidaten met een hoger cijfer slechts een grotere kans maken op toelating, maar daarvan niet zeker zijn. Aangezien studenten met een hoger eindexamencijfer vaker nominaal afstuderen, zorgt het selectie-algoritme voor een hoger slagingspercentage. Het uitgebreide voorspelmodel haalt een nog wat hoger slagingspercentage dan het simpele voorspelmodel (65,0% om 63,9

%), maar selecteert een stuk minder studenten met migratieachtergrond: slechts 11,7%. Het meenemen van meer kenmerken gaat dus ten koste van de representativiteit als het selectie-algoritme niet wordt bijgestuurd.

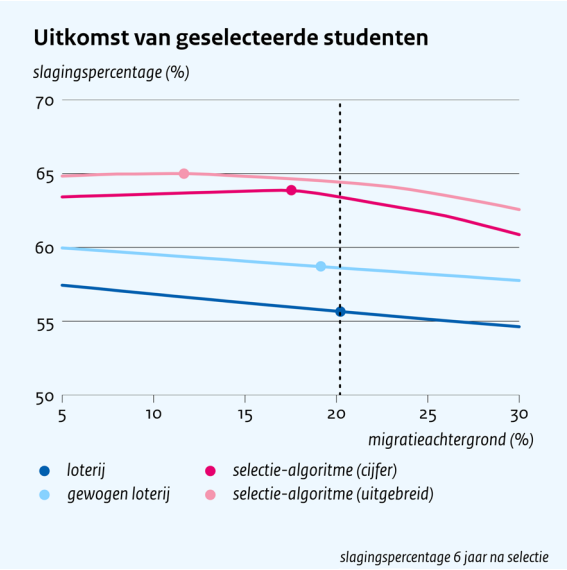
Voorspelmodellen kunnen representatiever gemaakt worden door ze bij te sturen. Net zoals we in de vorige sectie deden voor de gewogen en ongewogen loterij, kunnen we ook bij het simpele en uitgebreide voorspelmodel het aandeel studenten met migratieachtergrond in de selectie bijsturen. Zo kunnen we ervoor zorgen dat de selectie niet ten koste gaat van kansengelijkheid voor mensen met een migratieachtergrond. Deze resultaten zijn te zien in figuur 4: de lijn geeft het slagingspercentage gegeven het opgelegde aandeel studenten met een migratieachtergrond in de selectie. Elk punt op de lijn kan dus worden gekozen door vast te leggen dat het bijbehorende aandeel studenten met migratieachtergrond moet worden geselecteerd.

De door ons gevonden effecten van bijsturing zijn dan fors. Als we bijvoorbeeld bepalen dat het aandeel met migratieachtergrond gelijk moet zijn aan 20,2% – het aandeel in de aanmeldingen – dan geeft het simpele voorspelmodel op basis van het eindexamencijfer een slagingspercentage van 63,4%. Het slagingspercentage van de selectie van het uitgebreide voorspelmodel is nog wat hoger: 64,4%. Dat is aanzienlijk hoger dan de slagingspercentages van de selecties van de gewogen en ongewogen loterij, die 58,6% en 55,7% waren. Dit verschil is het gevolg van het gericht selecteren van kandidaten – apart onder kandidaten met een migratieachtergrond en onder kandidaten zonder – op basis van verwachte uitkomst, in plaats van willekeurig. Bij een studentenpopulatie van 2800 per jaar betekent dat in de selectie van het uitgebreide voorspelmodel ongeveer 162 meer basisartsen per jaar nominaal afstuderen dan in de selectie van de gewogen loterij. Dat is een stijging van bijna tien procent, wat een aanzienlijk effect heeft op het arbeidsaanbod van artsen, ook omdat dit effect ieder jaar optreedt. Het gebruik van een bijgestuurd selectie-algoritme zorgt naar verwachting dus voor meer nominaal afgestudeerde basisartsen dan een loting, terwijl het aandeel studenten met een migratieachtergrond gelijk is.

Figuur 3 – Datagedreven selectie resulteert in een hoger slagingspercentage dan (gewogen) loterij, maar minder studenten met migratieachtergrond



Figuur 4 – Als het aandeel met migratieachtergrond wordt bepaald, dan resulteert datagedreven selectie in een hoger slagingspercentage voor elk aandeel migratieachtergrond



5.1.3 Resultaten voor andere uitkomstmaten

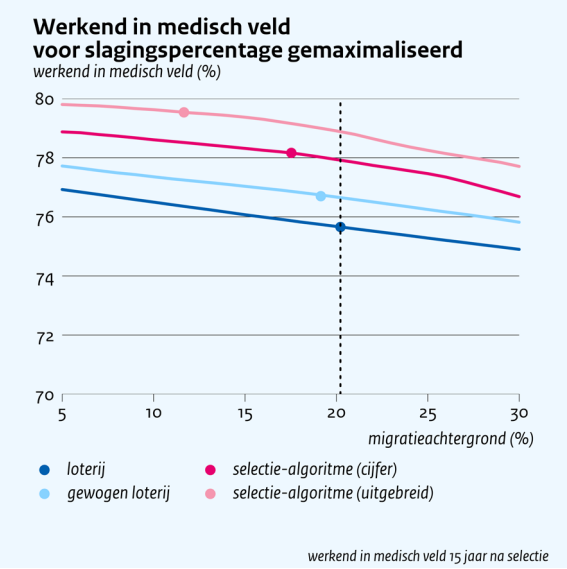
De gekozen uitkomstmaat kan doorwerken op de vooringenomenheid van een selectie-algoritme. Zoals genoemd in sectie 4.3.1 is de keuze van de uitkomstmaat één van de belangrijkste factoren voor een voorspelmodel. Een uitkomstmaat wordt doorgaans gekozen, omdat die een goed meetbare indicator voor een achterliggend doel is. Het voorspelmodel maximaliseert voor de uitkomstmaat, niet het achterliggende doel zelf, en dus moet goed gekeken worden of de keus van uitkomstmaat de juiste is om het doel te bereiken. De selectie die de ene uitkomst maximaliseert, hoeft niet de beste selectie te zijn voor een andere uitkomst. Daarnaast kunnen verschillende partijen verschillende achterliggende doelen hebben waarop andere uitkomstmaten beter op sturen. De door de gebruiker van het voorspelmodel gekozen uitkomstmaat hoeft dus niet het altijd meest maatschappelijk wenselijke doel te bereiken.

Voor dit onderzoek leggen we daarom twee uitkomstmaten naast elkaar. We vergelijken de effecten als 'slagen voor de studie zes jaar na selectie' gebruikt wordt als uitkomstmaat en als 'werken in het medisch veld vijftien jaar na selectie' gekozen wordt. Allebei deze uitkomstmaten zijn een indicator voor het achterliggende doel: kandidaten selecteren die goed als arts zorg kunnen bieden aan de samenleving. Wat we kunnen doen is een selectie maken die gemaximaliseerd wordt voor de ene uitkomst, en dan kijken hoe diezelfde selectie scoort op de andere uitkomst.

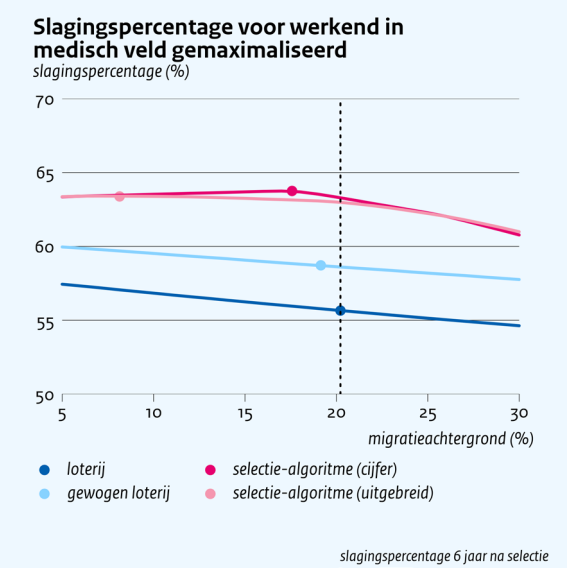
Selecteren voor de ene uitkomstmaat betekent dus niet dat er optimaal geselecteerd wordt voor een andere uitkomstmaat. Figuur 5 toont het aandeel kandidaten in de selectie dat na vijftien jaar werkzaam is in een medisch veld, waarbij de selectie gemaakt werd om het aandeel kandidaten dat na zes jaar geslaagd is voor de studie te maximaliseren. De punten geven per voorspelmodel aan wat de optimale representatie van kandidaten met een migratieachtergrond is voor een maximaal slaginspercentage. Zoals we zien zijn er voor de uitkomst 'werkend in medisch veld' echter hogere waarden haalbaar (in dit geval bij minder kandidaten met een migratieachtergrond). Selecteren voor de ene uitkomstmaat betekent dus niet dat er optimaal geselecteerd wordt voor een andere uitkomstmaat. We zien echter wel dat het merendeel van de geselecteerde studenten uiteindelijk werkzaam is in het medisch veld, ongeveer 78 à 80%. De studenten die zijn geselecteerd op basis van hun hoge verwachte slagingskans zijn meestal uiteindelijk ook werkzaam in het medisch veld, al had dat percentage nog hoger kunnen zijn geweest.

Dit resultaat volgt ook uit een andere analyse. Dit doen we door werken in een medisch veld vijftien jaar na selectie als uitkomstmaat nemen en vervolgens kijken hoe deze selectie het doet op slagingspercentage. Figuur 6 toont dit. Blijkbaar is het aantal kandidaten met een migratieachtergrond in de selectie voor wie de uitkomst 'werkend in medisch veld' maximaal is, ook min of meer het optimale aantal kandidaten voor wie het slagingspercentage optimaal is. Als we deze resultaten vergelijken met figuur 4, die het slagingspercentage weergeeft voor selecties waarbij dat ook de uitkomstmaat was, zien we wel dat het slagingspercentage in absolute zin iets lager is. De reden is opnieuw dat deze selectie niet gemaakt is met het slagingspercentage als expliciete uitkomstmaat, en dus niet het maximale slagingspercentage behaalt. Het verschil is echter klein, omdat voordat je werkzaam bent in de medische sector je doorgaans ook al een geneeskundediploma hebt behaald.

Figuur 5 – Een selectie om het slagingspercentage te maximaliseren hoeft een andere uitkomst, werken in een medisch veld, niet te maximaliseren.



Figuur 6 – Slagingspercentage voor selecties waarbij juist werken in een medisch veld vijftien jaar na selectie als uitkomstmaat genomen werd.



In ons geval leidt het gebruik van de twee uitkomstmaten tot selecties die op beide uitkomstmaten redelijk goed scoren, maar leidt het gebruik zonder bijsturing tot een andere mate van representativiteit. Geslaagd zijn voor de studie zes jaar na selectie en werken in een medisch veld vijftien jaar na selectie representeren immers hetzelfde achterliggende doel: kandidaten selecteren die goed als arts zorg kunnen bieden aan de samenleving. Toch kan de representativiteit verschillen als het voorspelmodel niet bijgestuurd wordt. Het uitgebreide voorspelmodel selecteert ongeveer 8 procent mensen met een migratieachtergrond als het moet voorspellen of iemand werkt in een medisch veld, terwijl dit percentage bijna 13 procent is als het afstuderen binnen zes jaar na selectie moet voorspellen. Dit is belangrijk, omdat partijen verschillende achterliggende doelen kunnen hebben die door de ene uitkomstmaat beter benadrukt worden dan door de andere. Een universiteit bijvoorbeeld vindt het misschien vooral belangrijk dat studenten snel (nominaal) afstuderen en zal dan de eerste uitkomstmaat prefereren. Maar misschien is het voor de samenleving wel belangrijker dat kandidaten daadwerkelijk in een medisch veld komen te werken en zou de tweede uitkomstmaat geprefereerd kunnen worden.

Daarbij komt nog dat de beperkingen van deze uitkomstmaten niet vergeten mogen worden. Het is bekend dat mensen met een migratieachtergrond geregeld met discriminatie op de arbeidsmarkt te maken krijgen (Bertrand & Mullainathan, 2004; Blommaert et al., 2014). Dat kan ervoor zorgen dat het goede kandidaten toch niet lukt om een baan als arts te bemachtigen. Gezien vanuit de uitkomstmaat 'werken in een medisch veld vijftien jaar na selectie' hebben die kandidaten dan een slechte uitkomst, terwijl dat niet betekent dat ze een slechte arts zouden zijn. Daarnaast is het niet zeker dat andere kwaliteiten die de samenleving graag bij een arts ziet, zoals vriendelijk met patiënten omgaan, goed worden weergegeven door een uitkomstmaat als deze.

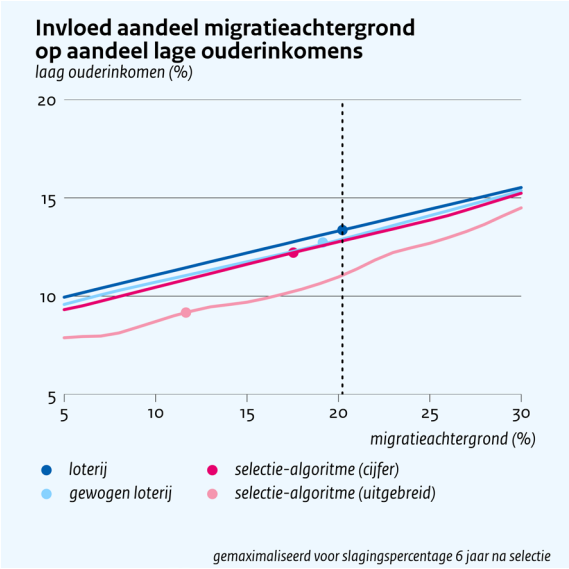
5.1.4 Resultaten voor andere representativiteitsmaten

Wanneer de representatie van de ene groep in een selectie wordt vastgelegd, kan dat invloed hebben op de representatie van andere groepen. In onze casus sturen we het voorspelmodel bij door op te leggen welk aandeel van de selectie een migratieachtegrond moet hebben, maar daarbij bestaat het risico dat dit ten koste gaat van andere groepen. Bijvoorbeeld dat hierdoor vooral meer buitenlandse studenten met Europese landen als herkomst worden toegelaten, ten koste van niet-Europese landen. Om hierop zicht te krijgen, kijken we naar het aandeel van bepaalde groepen in de selectie wanneer het aandeel van andere groep in de selectie

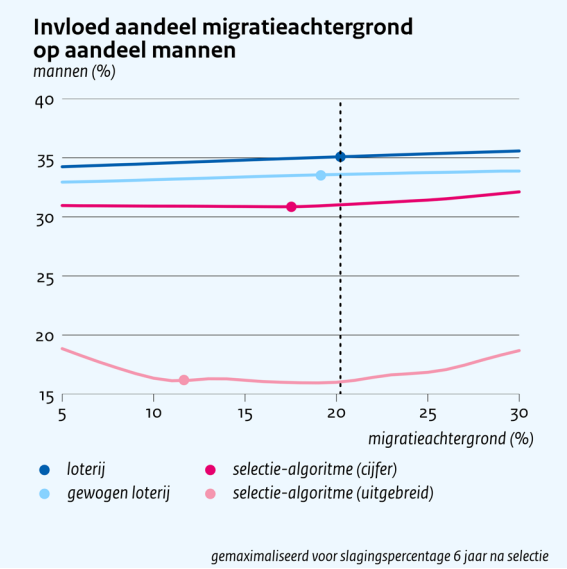
wordt vastgelegd. Figuur 7 toont het aandeel kandidaten in de selectie van wie de ouders een (relatief) laag inkomen hebben voor verschillende aandelen van kandidaten met een migratieachtergrond in de selectie. We zien dat als we het selectie-algoritme meer kandidaten met een migratieachtergrond laten selecteren, we ook meer kandidaten met (relatief) lage ouderinkomens selecteren. Dit komt doordat ouders van kandidaten met een migratieachtergrond vaker een laag inkomen hebben.

Wanneer meerdere kenmerken van de selectie belangrijk zijn, zullen deze allemaal meegegeven moeten worden. Figuur 8 laat zien dat het geselecteerde aandeel kandidaten met een migratieachtergrond weinig invloed heeft op het geselecteerde aandeel mannen. Wel zien we dat het uitgebreide voorspelmodel een stuk minder mannen selecteert dan het simpele. Een verklaring hiervoor is dat vrouwen vaker nominaal afstuderen dan mannen ook als hun eindcijfer iets lager is. Als we ook meer mannen zouden willen selecteren, zullen we dus expliciet het aandeel mannen in de selectie moeten bijsturen. Dit laat zien hoe belangrijk het is om naar subgroepen te kijken en daarop mogelijk bij te sturen. Wanneer je dit niet expliciet aan het selectie-algoritme meegeeft, zal het er geen rekening mee houden. Bansak et al., 2018 laat zien dat een forse stijging in verwachte effectiviteit ook mogelijk is als in meerdere dimensies eisen voor representativiteit opgelegd worden.

Figuur 7 – Een groter aandeel migratieachtergrond selecteren, leidt tot ook een groter aandeel lage ouderinkomens selecteren.

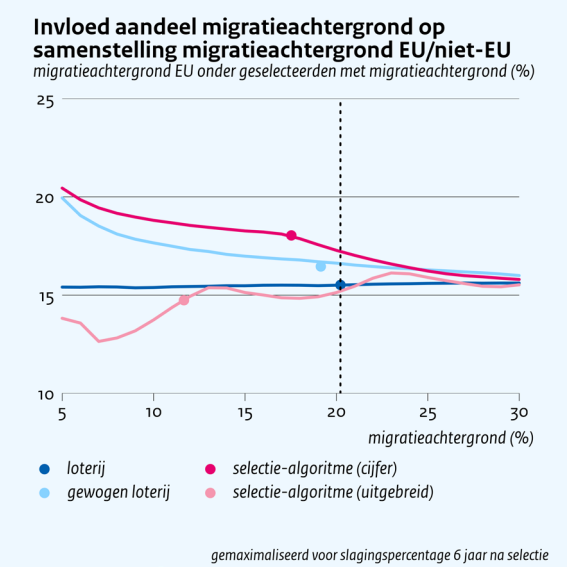


Figuur 8 – Het selecteren van een bepaald aandeel migratieachtergrond heeft weinig invloed op het aandeel geselecteerde mannen.



De selectie-algoritmes selecteren dus niet disproportioneel veel studenten met een Europese herkomst. Figuur 9 toont het percentage kandidaten met een EU-migratieachtergrond in het totaal aantal kandidaten met een migratieachtergrond (EU en niet-EU). Als we het selectie-algoritme bijsturen om meer kandidaten met een migratieachtergrond te selecteren, dan zien we dat het uitgebreide voorspelmodel niet voornamelijk kandidaten met een migratieachtergrond uit de EU selecteert. In tegendeel, de voorspelmodellen die puur op cijfer selecteren hebben juist een hoger percentage kandidaten met een migratieachtergrond buiten de EU in de selectie.

Figuur 9 – Het aandeel met herkomst EU in studenten met migratieachtergrond neemt niet toe als we het selectie-algoritme meer studenten met een migratieachtergrond laten selecteren



5.2 Samenvatting

De resultaten laten duidelijk zien dat het bijgestuurde selectie-algoritme dat gebruik maakt van een voorspelmodel de potentie heeft om effectiever te zijn dan een loterij, zonder dat dit ten koste gaat van kansengelijkheid. Soms leidt het zelfs tot betere representativiteit. Om dit toe te passen moeten er wel expliciete keuzes worden gemaakt over welke uitkomsten worden nagestreefd, en welke kenmerken van kandidaten aan een voorspelmodel gegeven worden. Ook moet vastgesteld worden voor welke groepen de representatie in de gaten gehouden moet worden. Dat is de belangrijkste uitdaging voor de toepassing van datagedreven selectie. Als over die keuzes geen overeenstemming kan worden bereikt, kan een ongewogen loterij worden gebruikt om kansengelijkheid te bereiken. Het nadeel daarvan is echter dat de effectiviteit fors lager uitvalt dan met de datagedreven selectie.

Bij datagedreven selectie moet het voorspelmodel onderhouden worden. De uitkomsten moeten gemonitord worden en de data geüpdatet om ervoor te zorgen dat voorspellingen realistisch blijven. De voorspellende waarde van de *input* kan immers over tijd veranderen. Dit kun je doen door een deel van de plaatsen - bijvoorbeeld 5% - via een loterij te laten gebeuren, zodat je de uitkomsten van die groep kunt gebruiken voor het updaten van het voorspelmodel. Voor de simpele datagedreven selectie op basis van alleen eindcijfer is dit minder belangrijk, omdat het aannemelijk is dat een hoger eindexamencijfer over het algemeen samenhangt met een hogere slagingskans.

De selectie-algoritmes die we in dit onderzoek hebben gebruikt, kunnen waarschijnlijk nog effectiever worden ingezet door onderwijsinstellingen zelf. Die hebben namelijk meer gedetailleerde informatie over kandidaten en die specifieke gegevens kunnen goede voorspellers zijn. Bijvoorbeeld het precieze eindexamencijfer – terwijl wij alleen een beperkt aantal cijfercategorien tot onze beschikking hebben – en uitslagen van toelatingstoetsen die ook als *input* voor de voorspelmodellen kunnen dienen. Ook data die de mate van motivatie van kandidaten weergeven, zouden een grote verbetering als gevolg kunnen hebben. Onze resultaten geven dus waarschijnlijk een ondergrens voor de potentie die datagedreven selectie-algoritmes hebben.

6 Discussie

Selectie-algoritmes worden steeds vaker gebruikt en daarmee groeit de vrees dat zij discrimineren en ongewenste patronen in de samenleving versterken. Er zijn daarom veel regelgevingen en processen ontworpen om discriminatie door algoritmes te beperken. Deze maatregelen bestrijken een breed palet en gaan zowel over het toezicht op het trainen, testen en gebruik van algoritmes in het Toezichtskader algoritmes (Algemene Rekenkamer, 2023) als over het in kaart brengen hoe het gebruik van algoritmes mensenrechten raken en in hoeverre dit zich verhoudt tot de doel van het algoritme via de IAMA (Gerards et al., 2021).

Het gebruik van selectie-algoritmes heeft voor- en nadelen. Een voordeel van het gebruik van selectie-algoritmes is dat ze 'gelijke gevallen gelijk behandelen'. Stel een instelling hanteert een algoritme om een risico-inschatting te maken. Personen met dezelfde set aan *input*-data zullen dan dezelfde risicovoorspelling krijgen. Hiermee sluiten ze aan bij artikel 1 van de Nederlandse grondwet, die het maken van onderscheid in

gelijke gevallen verbiedt.[14]Een nadeel van het gebruik van selectie-algoritmes is dat verschillende gevallen een verschillende behandeling kunnen krijgen. Dit kan ook gebeuren als mensen verschillen op kenmerken waarvan we vinden dat die niet relevant zouden moeten zijn. Op de arbeidsmarkt vindt discriminatie plaats tegen mensen met een migratieachtergrond (Bertrand & Mullainathan, 2004; Blommaert et al., 2014). Vertaald naar de casus uit hoofdstuk 5 kan dit ervoor zorgen dat afgestudeerde studenten met een migratieachtergrond minder vaak doorstromen tot arts of medisch specialist. Hierdoor kan een selectie-algoritme, dat studenten moet selecteren die arts worden, een lagere succeskans geven aan mensen met een migratieachtergrond dan aan mensen zonder migratieachtergrond, ook als ze in alle andere kenmerken identiek zijn (en even geschikt zijn voor de taak).

Daarnaast kunnen ongewenste patronen in de samenleving ervoor zorgen dat mensen als 'gelijke gevallen' gezien worden, terwijl ze eigenlijk verschillen. Scholieren die bijles gevolgd hebben, zullen gemiddeld hogere cijfers halen dan even intelligente scholieren die dat niet gedaan hebben. Hierdoor kan het zijn dat de capaciteiten van twee personen verschillen, terwijl het eindcijfer hetzelfde is. Dit probleem speelt niet alleen als selectie-algoritmes gebruikt worden, maar ook als mensen selecties maken op basis van kenmerken.

In deze publicatie behandelen we het tegengaan van discriminatie door algoritmes als een economisch probleem. We benadrukken dat voorspellende selectie-algoritmes enkel tot doel hebben om de effectiviteit van een bepaalde selectie te maximaliseren. Zij houden hierdoor geen rekening met de representatitiviteit van de selectie: de mate waarin de kenmerken van de geselecteerde groep overeenkomen met die van de niet-geselecteerde groep. Omdat selectie-algoritmes – gegeven de gekozen uitkomst en de *input*data – de effectiviteit maximaliseren, leidt een aanpassing van de selectie ten gunste van de representativiteit tot een daling in verwachte effectiviteit. Het tegengaan van discriminatie door algoritmes beïnvloedt dus zowel de representativiteit als de effectiviteit.

In deze publicatie presenteren we de methode van 'bijgestuurde selectie' als alternatief om voorspellende algoritmes rechtvaardiger te maken. Andere manieren om discriminatie tegen te gaan vergelijken vaak impliciet twee selectie-algoritmes op basis van representativiteit. Dit gebeurt bijvoorbeeld als getest wordt of het gebruik van persoonskenmerken leidt tot uitsluiting, zoals voorgesteld in de Handreiking non-discriminatie *by design*. De effectiviteit van beide selectie-algoritmes – met en zonder persoonskenmerken – wordt daarbij niet meegenomen in de afweging. Bij de bijgestuurde selectiemethode wordt de effectiviteit wel meegenomen in de vergelijking. In deze methode focust het voorspellend selectie-algoritme zich enkel op het maken van een zo goed mogelijke risico-inschatting. Er worden daarbij – vanuit het oogpunt van representativiteit – geen restricties opgelegd aan de *input*data of het te gebruiken algoritme. In plaats daarvan wordt, nadat een risicoinschatting is gemaakt, de representativiteit van de selectie gegarandeerd door op te leggen welke kenmerken de geselecteerde groep moet hebben. De methode van bijgestuurde selectie maximaliseert op deze wijze de effectiviteit gegeven een gewenste mate van representativiteit.

Het gebruik van de bijgestuurde selectiemethode maakt het daarmee mogelijk om de inbreuk van algoritmes op mensenrechten te kwantificeren én om deze te beperken of zelfs tegen te gaan. Wel vereist deze methode dat gebruikers een expliciete keuze maken over welke kenmerken de geselecteerde groep moet hebben. Dit kan door te kijken naar een of naar meerdere kenmerken. Het vastleggen van eisen op een bepaald kenmerk kan doorwerken op andere kenmerken van de selectie. Zo laat onze casus zien dat het omhoog bijsturen op het aantal studenten met een migratieachtergrond ook het aandeel studenten uit gezinnen met een laag inkomen vergroot. Als een volledig voorspelmodel gebruikt wordt, is het aandeel mannen in de selectie echter laag, en dit aandeel stijgt niet als bijgestuurd wordt op het aandeel studenten met een migratieachtergrond. Er moet dus goed nagedacht worden over op welke kenmerken de selectie

[14] Wanneer mensen een risicoinschatting maken bestaat het risico dat gelijke gevallen ongelijk behandelt worden (de Beer, 2021). Ook de mate waarin mensen in staat zijn om hun beslissingen te onderbouwen blijkt in de prakijk vaak tegen te vallen (Kleinberg, Ludwig, Mullainathan, & Sunstein, 2018).

bijgestuurd wordt. Dit is een normatieve keuze die binnen organisaties en in de maatschappij bepaald moet worden.

De bijgestuurde selectiemethode vult bestaande regelgeving zoals de IAMA en het Toetsingskader algoritmes aan, maar zij vervangt deze dus niet. De bijgestuurde selectiemethode maakt het mogelijk om de uitruil tussen effectiviteit en representativiteit te kwantificeren. Dit geeft een invulling aan het laatste onderdeel van de IAMA, waarin de (mogelijke) inbreuk van een algoritme op de mensenrechten afgewogen wordt op het doel van het algoritme. De bijgestuurde selectiemethode maakt het immers mogelijk om na te gaan welke kenmerken de selectie heeft, en hoe de effectiviteit van het algoritme verandert als de representativiteit aangepast wordt. Tegelijkertijd kan de bijgestuurde selectiemethode gebruikt worden om algoritmes te toetsen op een eventuele discriminerende werking, wat nauw aansluit bij het Toetsingskader algoritmes.

Literatuur

Algemene Rekenkamer. (2023). *Toetsingskader Algoritmes*. Algemene Rekenkamer. https://www.rekenkamer.nl/onderwerpen/algoritmes/algoritmes-toetsingskader

Autoriteit Persoonsgegevens. (2023). *Data protection impact assessment (DPIA)*. https://www.autoriteitpersoonsgegevens.nl/nl/zelf-doen/data-protection-impact-assessment-dpia

Bansak, K., Ferwerda, J., Hainmueller, J., Dillon, A., Hangartner, D., Lawrence, D., & Weinstein, J. (2018). Improving refugee integration through data-driven algorithmic assignment. *Science*, *359*(6373), 325–329.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, *94*(4), 991--1013.

Blommaert, L., Coenders, M., & Tubergen van, F. (2014). Discrimination of Arabic-named applicants in the Netherlands: An internet-based field experiment examining different phases in online recruitment procedures. *Social Forces*, *92*(3), 957--982.

Brandpunt. (2021). *Slecht nieuws: Algoritmes discrimineren anno 2021 nog steeds*. https://www.npo3.nl/brandpuntplus/algoritmes-discrimineren-2021

Gerards, J., Schaefer, M., Vankan, A., Muis, I., & others. (2021). *Impact Assessment Mensenrechten en Algoritmes*.

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.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *The Quarterly Journal of Economics*, *133*(1), 237–293.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Rambachan, A. (2018). Algorithmic fairness. *Aea Papers and Proceedings*, *108*, 22–27.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Sunstein, C. R. (2018). Discrimination in the Age of Algorithms. *Journal of Legal Analysis*, *10*, 113–174.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Sunstein, C. R. (2020). Algorithms as discrimination detectors. *Proceedings of the National Academy of Sciences*, *117*(48), 30096–30100.

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). Inherent trade-offs in the fair determination of risk scores. *ArXiv Preprint ArXiv:1609.05807*.

NOS. (2020). *Anti-fraudesysteem SyRI moet van tafel, overheid maakt inbreuk op privéleven*. https://nos.nl/artikel/2321704-anti-fraudesysteem-syri-moet-van-tafel-overheid-maakt-inbreuk-op-priveleven

NU.nl. (2022, May 30). *Institutioneel racisme bij fraudedienst van de fiscus. Dit was er aan de hand*. https://www.nu.nl/binnenland/6203770/institutioneel-racisme-bij-fraudedienst-van-de-fiscus-dit-was-er-aan-de-hand.html

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.

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

Rambachan, A., Kleinberg, J., Mullainathan, S., & Ludwig, J. (2020). *An economic approach to regulating algorithms*. National Bureau of Economic Research.

Rambachan, A., & Roth, J. (2019). Bias in, bias out? Evaluating the folk wisdom. *ArXiv Preprint ArXiv:1909.08518*.

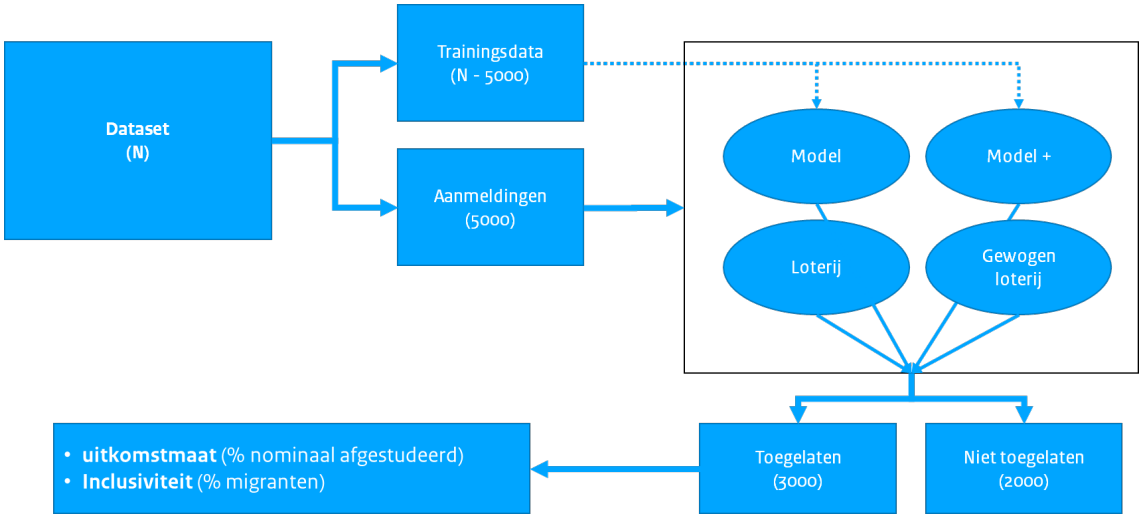
van der Sloot, B., Keymolen, E., Noorman, M., Weerts, H., Wagensveld, Y., & Visser, B. (2021). *Handreiking non-discriminatie by design*.

7 Bijlagen

7.1 Opzet selectie-algoritme

In het onderstaande schema staat de opzet van het selectie-algoritme waarop onze resultaten in hoofdstuk 5 gebaseerd zijn. De dataset bestaat uit iedereen die tussen 2000 en 2004 is toegelaten tot de opleiding geneeskunde via de gewogen loterij. Uit deze dataset trekken we 5000 personen die zich aanmelden voor de opleiding (aanmeldingen). De overige data gebruiken we voor training van de twee datagedreven voorspelmodellen: een voorspelmodel op basis van eindexamencijfer in categorieën (Model) en een uitgebreid voorspelmodel (Model +) met aanvullende kenmerken: vooropleiding (bijv. vwo of propedeuse hbo), leeftijd bij aanmelding, geslacht, migratieachtergrond en ouderinkomen. De loterijen hebben geen training nodig, omdat ze willekeurig aanmeldingen selecteren, waarbij de gewogen loterij gebruikmaakt van gewichten die zijn vastgesteld bij het opzetten van de selectieprocedure in 1975.

Schematische weergave van het selectie-algoritme



We specificeren de voorspelmodellen als volgt. De modellen zijn geschat met *Ordinary Least Squares* (OLS), waarbij de volgende specificaties zijn gebruikt:

$$Model: Y_i = \beta_0 + \sum_{g=B}^{F} \beta_g C_i + \varepsilon_i,$$

Y_i is hierbij de uitkomstmaat voor persoon i, bijvoorbeeld of er een diploma geneeskunde is behaald zes jaar na toelating. C_i is de cijfercategorie is waar persoon i in valt. Deze categorieën lopen van A (hoge cijfers) naar F (lage cijfers, zie tabel B1 hieronder. Categorie A is weggelaten uit de specificatie en is dus opgenomen in de constante β_0. Het uitgebreide voorspelmodel voegt extra kenmerken toe aan de specificatie:

$$Model+: Y_i = \beta_0 + \sum_{g=B}^{F} \beta_g C_i + \beta_{eu} M_{EU,i} + \beta_{ol} M_{OL,i} + \beta_{ink} Ink_i + \sum_{l=(0,19]}^{(23,\infty)} \beta_l L_i + \beta_{vwo} Opl_{vwo,i} + \beta_{wo} Opl_{wo,i} + \beta_{hbo} Opl_{hbo,i} + \beta_{hboprop} Opl_{hboprop,i} + \beta_{woprop} Opl_{woprop,i} + \beta_{int} Opl_{int,i} + \varepsilon_i,$$

waarbij $M_{EU,i}$ en $M_{OL,i}$ staan voor migratieachtergrond uit de EU en Overige Landen (OL). Ink_i is een indicator voor laag ouderinkomen, met peiljaar 2005. L_i is een indicator voor leeftijdscategorie, opgedeeld in vier categorieën: jonger dan 19, ouder dan 19 en jonger dan 21, ouder dan 21 en jonger dan 23, en ouder dan 23. De categorie ouder dan 19 en jonger dan 21 is uit de specificatie weggelaten en is opgenomen in de constante: β_0. De overige termen zijn indicatoren voor de vooropleiding (aangeduid met Opl) op het moment dat de kandidaten zich aanmelden. $Opl_{vwo,i}$ staat voor vooropleiding vwo, $Opl_{wo,i}$ voor wo, $Opl_{hbo,i}$ voor hbo en $Opl_{int,i}$ voor een internationaal diploma. De toevoeging $prop$ staat voor propedeuse: in dat geval heeft de aanmelder alleen de propedeuse van de opleiding gevolgd: $Opl_{hboprop,i}$ voor de hbo-propedeuse en $Opl_{woprop,i}$ voor de wo-propedeuse.

Tabel B1. Indeling cijfercategorieën

Cijfer-categorie	Gemiddeld eindexamencijfer
A	Cijfer ≥ 8.5
B	8.0 ≤ Cijfer < 8.5
C	7.5 ≤ Cijfer < 8.0
D	7.0 ≤ Cijfer < 7.5
E	6.5 ≤ Cijfer < 7.0
F	Cijfer < 6.5

We selecteren kandidaten volgens onderstaande procedure. Nadat de voorspelmodellen zijn getraind, krijgen ze de 5000 aanmelders – voor het eerst – te zien en worden deze gerangschikt op basis van de verwachte score op de uitkomstmaat ($\hat{Y}_i$). Vervolgens worden de beschikbare plaatsen vergeven aan de hoogst gerangschikte aanmelders. In onze casus gaan we ervan uit dat er 3000 plaatsen beschikbaar zijn, dus 60% van de aanmelders kan worden toegelaten. Dit is ongeveer gelijk aan het aandeel toelatingen zoals dat in de UWV-dataset voorkomt tussen 2000 en 2004. Ten slotte registeren we de uitkomst, bijvoorbeeld het slagingspercentage, en de kenmerken van de personen, zoals de migratieachtergrond, van de studenten die zijn geselecteerd voor toelating.

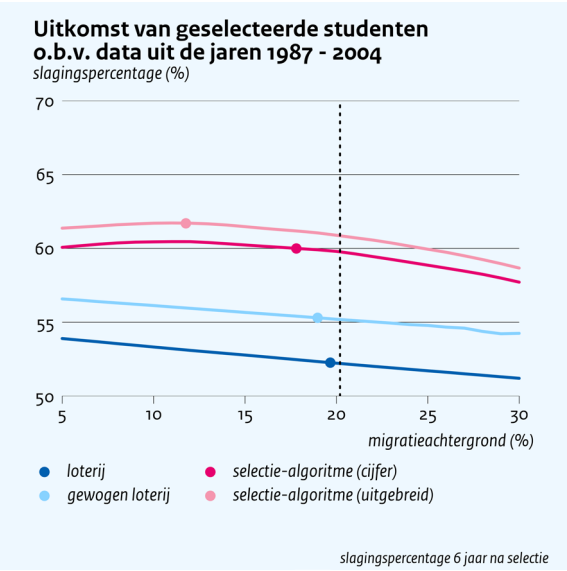
Weging verandert het selectieproces van de loterij. In de ongewogen loterij worden de plaatsen simpelweg willekeurig verdeeld: de 5000 aanmelders worden willekeurig gerangschikt en de 3000 bovenste worden toegelaten. In de gewogen loterij gaat dit anders, daar wordt eerst bepaald hoeveel plaatsen elke cijfercategorie beschikbaar heeft, conform de gewichten zoals die in de daadwerkelijk uitgevoerde gewogen loting werden gebruikt. Vervolgens worden de aanmelders binnen elke cijfercategorie willekeurig gerangschikt en worden de bovensten van elke cijfercategorie toegelaten, conform het aantal beschikbare plaatsen per cijfercategorie.

Het bijsturen van het aandeel van een bepaalde groep in de selectie, doen we door het selectieproces op te delen in twee groepen. Voor studenten met een migratieachtergrond doen we dat als volgt. We verdelen de

aanmeldingen in de groep met en zonder migratieachtergrond en rangschikken de aanmelders binnen deze groepen. Vervolgens worden de hoogst gerangschikte aanmelders geselecteerd op basis van het aantal plaatsen dat per groep beschikbaar is. Zetten we het aandeel studenten met een migratieachtergrond bijvoorbeeld op 20%, dan zijn er 600 plaatsen (0,2 * 3000) beschikbaar voor deze groep en worden de bovenste 600 toegelaten. De overige 2400 plaatsen worden toebedeeld aan de hoogst gerangschikte aanmelders zonder migratieachtergrond. Zo wordt gegarandeerd dat 20% van de plaatsen naar studenten met een migratieachtergrond gaat en dat alleen de hoogst gerangschikte personen worden toegelaten.

Om er zeker van te zijn dat onze resultaten niet gebaseerd zijn op een toevalstreffer voeren we deze exercitie meerdere keren uit. Daarbij trekken we telkens weer willekeurig 5000 aanmelders, trainen we de voorspelmodellen, maken we de rankschikkingen, bepalen we wie er worden toegelaten en noteren we de uitkomsten. Dit hebben we duizend keer uitgevoerd en de resultaten van hoofdstuk 5 tonen het gemiddelde van deze duizend simulaties.

7.2 Resultaten voor volledige dataset: 1987 - 2004

Resultaten op basis van de volledige dataset zijn vergelijkbaar met de resultaten die in de hoofdtekst gepresenteerd worden. In sectie 5.2 lieten we de resultaten zien van verschillende selectiemethoden met als uitkomst het behalen van een diploma zes jaar na selectie, uitgezet tegen het aantal geselecteerde kandidaten met een migratieachtergrond. Aan het eind van sectie 5.1 gaven we aan dat we ons beperken tot de laatste vijf jaar van de loting om zeker te weten dat de studentenpopulatie in de dataset vergelijkbaar is over de jaren. Dezelfde selectiemethoden kunnen uiteraard ook worden toegepast op de gehele dataset van 1987 tot en met 2004. Figuur 10 toont het equivalent van figuur 4 voor de gehele dataset van 1987 - 2004. Over de hele breedte ligt het slagingspercentage iets lager, maar de verhoudingen tussen de selectiemethoden zijn nagenoeg hetzelfde.

Figuur 10 – Slagingspercentage tegen migratieachtergrond voor verschillende selectiemethoden, op basis van data uit de hele lotingsperiode van 1987 - 2004