

Adviesdocument

Risicoprofilering heronderzoek bijstandsuitkering

Hoofdpunten normatieve adviescommissie








- > **Algoritmische profilering is onder strikte voorwaarden mogelijk**
Mits zorgvuldig toegepast kan algoritmische profilering verantwoord worden ingezet voor heronderzoek naar de rechtmatigheid van bijstandsuitkeringen
- > **Profilering moet geen verdenking zijn**
Heronderzoek moet meer op dienstverlening en minder op wantrouwen zijn gebaseerd
- > **Diversiteit aan selectiemethoden**
Om tunnelvisie en feedbackloops te vermijden is het wenselijk om algoritmische profilering te combineren met expert-gedreven profilering en aselecte steekproeven
- > **Verantwoord gebruik van profileringscriteria**
Met oog op het voorkomen van (proxy)discriminatie en andere onwenselijke vormen van onderscheid maakt de commissie per variabele een afweging over de geschiktheid voor profileringsdoeleinden (zie [Infographic](#))
- > **Uitlegbaarheidsvereisten machine learning**
Het is noodzakelijk dat de selectie van inwoners door de gehele besluitvormingsketen uitlegbaar is. Complexe trainingsmethoden om profileringscriteria te selecteren, zoals het onderhavige xgboost-algoritme, worden te complex geacht om te voldoen aan uitlegbaarheidsvereisten

Samenvatting advies









De commissie oordeelt dat algoritmische risicoprofilering onder strikte voorwaarden kan worden ingezet voor het selecteren van burgers met een bijstandsuitkering voor heronderzoek. Het oogmerk van heronderzoek is een leidende factor bij de beoordeling van profileringscriteria. Indien het bijstandsheronderzoek minder op basis van verdenking en meer op basis van dienstverlening zou worden ingevuld, dan laat de commissie meer ruimte voor het gebruik van profileringsvariabelen om hulpbehoevende groepen gericht van dienst te kunnen zijn. Voor de verschillende variabelen die door de Gemeente Rotterdam zijn gebruikt in de periode 2017-2021 geeft de commissie een onderbouwd oordeel waarom deze variabelen wel of niet wenselijk zijn om als selectie criterium voor profilering te worden gebruikt (zie [Infographic](#)). Gecombineerde inzet van verschillende selectiemethoden (waaronder expert-gedreven profilering en aselecte steekproeven) wordt wenselijk geacht om tunnelvisie en verkeerde feedbackloops te doorbreken. De commissie adviseert om zwaardere eisen te stellen aan algoritmische variabelenselectie dan voor profileringscriteria geselecteerd door domeinexperts. De commissie stelt dat algoritmes die ter ondersteuning worden ingezet om burgers te selecteren voor heronderzoek uitlegbaar moeten zijn. Complexe trainingsmethoden, zoals het xgboost-algoritme gebruikt door de Gemeente Rotterdam, kunnen niet voldoen aan deze uitlegbaarheidsvereisten. Dit advies is gericht aan alle Nederlandse en Europese gemeenten die profileringsmethoden in de context van sociale voorzieningen overwegen of toepassen.

Infographic – Suggestie verantwoorde profileringscriteria

Per variabele adviseert de adviescommissie over de geschiktheid van profileringscriteria in de context van heronderzoek bij standsuitkeringen. Onderstaand advies is een suggestie. Organisaties moeten zelf per geval en per context blijven afwegen of onderscheid op basis van bepaalde criteria wenselijk is.

Wel te gebruiken variabelen		Niet te gebruiken variabelen	
Leeftijd		Postcode, wijk	
No show bij afspraak met gemeente	 	Geslacht, gender	
Reminders voor aanleveren van informatie	 	Reden voor afspraak met gemeente (jaarlijks gesprek, intake)	
Participatie in traject naar werk (training, werkplaats, maatschappelijke taak)	  	Type contact (mailen, bellen, appen, post)	
Type woonvorm (met huisgenoot, met partner)	 	Laaggeletterdheid	
Kostendelersnorm	 	ADHD	
		Bezoekt psycholoog	
		Aantal kinderen	
		Sector (werk)ervaring (horeca, bouw, logistiek)	
		Assertiviteit	 
		Professioneel voorkomen	

Legenda

	Wettelijk verboden		Veranderlijk
	Proxy-discriminatie		Subjectief
	Inhoudelijk verband met doel		Onduidelijke variabele
	Geen inhoudelijk verband met doel		Beheersbare risico's

Box 1

Algo prudentie: Case-based normatief advies voor ethische algoritmes

Algorithm Audit heeft geen mandaat om juridische bindende of officiële uitspraken te doen. In onze case studies geven wij niet-bindend ethisch advies. Ethisch advies gaat vaak verder dan dat wat nodig is om de wet na te leven. Toch dient bij gebrek aan wettelijke regelgeving of heldere normen van toezichthouders ons onafhankelijke ethische advies als een wegwijzer voor organisaties. Onze casuïstiek kan ook helpen bij het uitwerken van officiële standaarden of het ondersteunen van toekomstige besluiten van rechtsorganen. In dat opzicht is ons ethisch advies relevant binnen het juridische domein.

Inhoudsopgave

Hoofdpunten normatieve adviescommissie	2
Samenvatting advies	2
Infographic – Suggestie verantwoorde profileringscriteria	3
Voorwoord	5
Scope van advies	7
Algemene overwegingen	7
Transparantie en uitlegbaarheid	9
Expert-gedreven versus algoritmisch profileren	11
Variabelenselectie voor profilering	12
Samenstelling normatieve adviescommissie	16

Over Algorithm Audit

Algorithm Audit is een Europees kennisplatform voor AI bias testing en normatieve AI standaarden. De doelen van de stichting zijn driedelig:



Normatieve adviescommissies

Adviseren over ethische kwesties in concrete algoritmische toepassingen door het samenbrengen van deliberatieve, diverse adviescommissies, met *algoprudentie* als resultaat (zie Box 1)



Technische hulpmiddelen

Implementeren en testen van technische methoden voor bias-detectie en -mitigatie, zoals onze [bias scan tool](#)



Kennisplatform

Samenbrengen van kennis en experts om het collectieve leerproces over verantwoorde inzet van algoritmes aan te jagen, bijvoorbeeld ons [AI Policy Observatory](#) en [position papers](#)

Voorwoord

Dit advies is de uitkomst van een deliberatieve afweging van een onafhankelijke normatieve adviescommissie. Stichting Algorithm Audit heeft dit advies opgesteld op basis van gesprekken die zijn gevoerd tijdens een fysieke commissiebijeenkomst. Tijdens deze bijeenkomst zijn verschillende ethische kwesties besproken rondom de toepassing van algoritmische risicoprofilering voor heronderzoek bij bijstandsuitkeringen.

De specifieke casus die ten grondslag ligt aan dit advies is een algoritme dat de Gemeente Rotterdam in de periode 2017-2021 heeft ingezet voor dit doel. Dit algoritme is een aantal keer in opspraak geraakt, o.a. door een rapport van de Rotterdamse Rekenkamer¹, dat stelde dat het gebruik van de variabele 'laaggeletterdheid' een risico vormt voor proxydiscriminatie. Het journalistieke onderzoekscollectief VPRO Argos/Lighthouse Reports heeft het model onderzocht en kritiek geuit op o.a. het gebruik van subjectieve persoonskenmerken en de vooringenomenheid van de trainingsdata². In Nederland staat de controverse over gemeentelijke profileringsalgoritmen tegen de achtergrond van het kinderopvangtoeslagenschandaal, waar duizenden onschuldige burgers in grote problemen zijn geraakt door onterechte fraudeverdenkingen van de Belastingdienst. Omdat discriminerende profileringsalgoritmen hierbij een rol speelden, voelt men een grote urgentie om na te denken hoe (en of) dergelijke systemen verantwoord kunnen worden ingezet. In de praktijk is het lastig om onderzoek te verrichten naar profileringsmodellen bij de overheid, aangezien vaak onvoldoende zicht is op de precieze werking ervan. Door Wet openbaarheid van bestuur (Wob)- en Wet open overheid (Woo)-verzoeken van VPRO Argos/Lighthouse Reports zijn technische documentatie en evaluatiemetrieken over het Rotterdamse model openbaar gemaakt, waardoor juist relatief veel bekend is over dit algoritme³.

Vanwege zowel het controversiële karakter van het algoritme, het grote belang van vraagstukken rondom de inzet van risicoprofilering en (semi-)automatische besluitvorming door gemeenten, en de beschikbaarheid van informatie over dit algoritme, achtte Algorithm Audit dit een geschikte casus voor een uitgebreid onderzoek en een evaluatief oordeel door een normatieve adviescommissie. Alhoewel de casus-Rotterdam de basis vormt, zijn de vragen en afwegingen die in dit rapport centraal staan algemeen van toepassing. Het advies is daarom gericht aan alle Nederlandse en Europese gemeenten, die vroeg of laat zullen moeten nadenken over automatisering en risicoprofilering in de context van sociale voorzieningen.

Op basis van een uitgebreid onderzoek naar deze casus heeft Algorithm Audit ethische kwesties geïdentificeerd, die wij als meest urgent en belangwekkend hebben ingeschat. In dit onderzoek zijn naast domeinexperts ook burgers geconsulteerd op basis van hun ervaringen met bijstandsheronderzoek. De volledige uitwerking hier van is te vinden in een apart

¹ Gekleurde Technologie, Rotterdamse Rekenkamer 2021 <https://rekenkamer.rotterdam.nl/onderzoeken/algoritmes/>

² Inside the Suspicion Machine, Wired 2023 <https://www.wired.com/story/welfare-state-algorithms/>

³ Wob-verzoek VPRO Argos/Lighthouse Reports

<https://www.vpro.nl/dam/jcr:c87f2d6c-3f9c-4498-9a9c-f3bc5483a437/Downloads%20Model%20Rotterdam.zip>

document (Probleemstelling⁴). Dit document fungeerde als basis voor de bespreking en het oordeel van de adviescommissie. Als hoofdvraag geldt welke kenmerken ('variabelen') geschikt zijn om te dienen als criteria voor risicoprofielen, gericht op het selecteren van burgers met een bijstandsuitkering voor heronderzoek. In de Probleemstelling wordt dit besproken aan de hand van drie deelvragen:

1. Het probleem van variabelen die gelden als zogenaamde 'proxy' voor beschermde gronden, en welke daarvan moeten worden gemeden om proxy-discriminatie te mitigeren;
2. Het probleem van subjectieve en irrelevante variabelen of anderszins onwenselijke vormen van onderscheid;
3. De afweging tussen algoritmische profilering en expert-gedreven profilering.

Rondom deze kernvragen kwam de adviescommissie ook te spreken over andere aspecten die van belang werden geacht voor een verantwoord gebruik van risicoprofilering door gemeenten. Deze hebben o.a. betrekking op uitlegbaarheidsvereisten en de wijze waarop de gemeente de burger benadert.

Dit advies is een uitwerking van een deliberatieve afweging tussen experts en belanghebbenden die samen de adviescommissie vormen. De commissie is een diverse groep waarin de verschillende relevante expertises en belangengroepen vertegenwoordigd zijn. De exacte samenstelling is te vinden in de sectie [Samenstelling normatieve adviescommissie](#). Zowel de adviescommissie als Algorithm Audit zelf zijn volledig onafhankelijk. Het onderzoek en het advies zijn niet in opdracht van de Gemeente Rotterdam uitgevoerd. Het advies van de commissie is niet bindend, maar dient als een normatieve richtlijn voor alle partijen die worstelen met ethische dilemma's rondom algoritmische profilering in de context van sociale voorzieningen.

⁴ Probleemstelling Risicoprofilering voor heronderzoek bijstandsuitkering, AA:2023:02:P.

Betrokkenen bij totstandkoming advies

Voor dit onderzoek zijn verschillende belanghebbenden betrokken. De samenstelling van de normatieve adviescommissie wordt vermeld in de sectie Samenstelling normatieve adviescommissie.



Individen onderworpen aan algoritme



Onderzoeksjournalisten



Belangenvertegenwoordigers van benadeelden



Gemeentelijke instanties (Rotterdamse Rekenkamer en Ombudsman)



Gemeente Rotterdam



Juridische experts en wetenschappers

Scope van advies

Er spelen vele factoren een rol bij verantwoorde risicoprofilering voor bijstandsheronderzoek en het gebruik van algoritmes daarvoor. Dit advies beperkt zich tot een aantal belangrijke hete hangijzers waarvoor, naar het oordeel van Algorithm Audit, dringend normatieve en publiek toegankelijke besluitvorming nodig is. Dit is voornamelijk het geval bij die kwesties waar bestaande regelgeving, handreikingen en implementatiekaders geen pasklaar antwoord bieden. Met betrekking tot deze casus zijn dat de volgende aspecten:

- > doel van heronderzoek en de in dat kader wenselijke vormen van profilering;
- > transparantie- en uitlegbaarheidsvereisten voor (algoritmische) risicoprofilering;
- > bepalen welke variabelen wel en niet wenselijk zijn om te worden gebruikt voor risicoprofilering;
- > afweging tussen algoritmische en expert-gedreven profilering.

De focus van de adviescommissie op bovenstaande kwesties wil niet zeggen dat andere aspecten minder belangrijk zijn. Zoals elke organisatie die algoritmische systemen toepast moeten gemeenten zorg dragen voor goede governance met betrekking tot dataverwerking, besluitvormingsprocessen en het beleggen van rollen en verantwoordelijkheden. Ook de controle van de datakwaliteit is van groot belang. Voldoende aandacht voor deze onderdelen van verantwoord algoritmegebruik is verre van vanzelfsprekend. Zoals het onderzoek van VPRO Argos/Lighthouse Reports laat zien is in deze casus sprake van niet-representatieve trainingsdata. Deze casus laat ook zien hoe de governance en de verantwoording over algoritmes systematisch te kort schiet. Het is zorgwekkend dat algoritmische systemen enkele jaren als 'pilot' actief zijn en in de praktijk een belangrijke rol zijn gaan spelen in de interne werkprocessen van gemeenten. Intern toezicht, zoals een functionaris gegevensbescherming, zou bijzondere aandacht moeten schenken aan voorafgaande toetsing van pilots en bijbehorende gegevensverwerking. De reden dat het advies van de commissie niet ingaat op deze aspecten van governance en datakwaliteit is dan ook niet dat ze minder urgent zouden zijn, maar dat bestaande kaders hiervoor al voldoende concrete handvatten bieden. Dit in tegenstelling tot de normatieve kwesties die in dit advies worden behandeld, waar concreet toepasbare normen nog grotendeels ontbreken. Het ontwikkelen en formuleren van zulke publiek toegankelijke, normatieve standaarden is dan ook het doel van de normatieve adviescommissies van Algorithm Audit.

Algemene overwegingen

Bij het bepalen hoe profileringsalgoritmes verantwoord kunnen worden ingezet is het van belang om mee te wegen wat het doel van de profilering is. Oftewel, wat is de handeling die volgt op de profilering, en wat is daarvan de impact op de betrokken personen. Algoritmische systemen functioneren nooit op zichzelf, en moeten dus worden beoordeeld in de institutioneel-maatschappelijke context waarin ze zijn ingebed. In de context van profilering voor de selectie tot het bijstandsheronderzoek is het daarom belangrijk om na te denken over het precieze oogmerk van heronderzoek.

Heronderzoek kan worden benaderd als een strenge controle, waar de inwoner die wordt

geselecteerd voor heronderzoek al bij voorbaat wordt verdacht en gewantrouwd. Weliswaar is een bijstandsheronderzoek formeel gezien geen fraudeonderzoek, aangezien dit een aparte procedure is die pas op basis van concrete aanwijzing in een apart proces wordt gestart. Van fraudeverdenking is daarom op papier geen sprake bij een heronderzoek. Tegelijkertijd kan de procedure in de praktijk vanuit het perspectief van de burger wel zo uitwerken. Een uitnodiging tot heronderzoek kan overkomen als een verdenking en wordt door de burger (onder andere daarom) als stressvol ervaren. Een streng controlerende benadering van heronderzoek is gebaseerd op wantrouwen en heeft een sterk antagonistisch karakter: de gemeente staat tegenover de burger. Een andere benaderingswijze is gebaseerd op een coöperatieve houding: de gemeente werkt samen met de burger. Heronderzoek staat dan in het teken van dienstverlening. Het doel is om de inwoner te helpen met een correcte aanvraag en de administratieve afhandeling daarvan. Vetrekpunt is hier het inzicht dat aanvraagprocedures en officiële documenten ingewikkeld kunnen zijn voor inwoners en dat een fout snel wordt gemaakt. Het doel van heronderzoek is dan om deze fouten te herstellen en om samen tot een correcte bijstandsaanvraag te komen. Contact met inwoners tijdens een heronderzoek kan voor de gemeente waardevolle feedback opleveren over de informatievoorziening naar burgers toe, bijvoorbeeld als jargon of complexe webformulieren een obstakel vormen.

De commissie schat de ethische risico's van profilering lager in als het heronderzoek een dienstverlenend oogmerk heeft. In het tegenovergestelde geval, van een heronderzoek gebaseerd op verdenking, zijn de risico's groter en stelt de commissie hogere eisen. Bij een dienstverlenend karakter zijn bepaalde vormen van onderscheid legitiem, die bij een streng controlerende benadering juist ontoelaatbaar zijn⁵. Het identificeren van groepen die meer baat hebben bij hulp heeft namelijk een volstrekt andere impact op de verhouding tussen burger en gemeente dan een profilering gericht op strenge controle.

De verschillende benaderingswijzen kunnen niet slechts een andere 'framing' zijn van de praktijk van bijstandsheronderzoek. Weliswaar is de manier waarop de gemeente communiceert over heronderzoek een belangrijk onderdeel van dienstverlenend beleid. Voor aanvragers van bijstand is het van belang hoe de gemeente hen benadert: op basis van verdenking of op basis van samenwerking en vertrouwen. Het helpt als de gemeente in brieven en in persoonlijk contact benadrukt dat heronderzoek geen verdenking bij voorbaat betekent, maar een dienst om tot correcte bijstandsaanvragen te komen. Tegelijkertijd kan het niet bij communicatie blijven. De dienstverlenende benadering moet concreet worden geïmplementeerd. De commissie heeft verschillende suggesties:

- > **Coöperatieve houding:** Als een correcte bijstandsuitkering wordt gezien als een gedeelde verantwoordelijkheid van inwoner en gemeente, vraagt dit extra stappen van de gemeente. Van inwoners mag worden gevraagd dat zij correcte gegevens en documenten aanleveren als zij aanspraak willen maken op publieke middelen. Deze in-

⁵ Onder Uitvoeringswet AVG (UAVG) artikel 25 sub a zijn uitzondering mogelijk voor onderscheid op basis van beschermde persoonskenmerken in het geval van positieve discriminatie.

formatieplicht is niet alleen een wettelijke vereiste⁶, het is een redelijke voorwaarde die de burger begrijpt. Tegelijkertijd moet men oog hebben voor de ongelijke machts- en kennisverhouding tussen burger en gemeente. Burgers staan in hun eentje tegenover een complexe overheidsorganisatie waarvan ze de procedures vaak niet kennen. Een al te antagonistische verhouding tussen burger en gemeente is dan funest en voedt wederzijds wantrouwen. Het is zaak dat de gemeente zich minder tegenover en meer naast de burger opstelt, en altijd oog houdt op de proportionaliteit van de maatregelen en vereisten.

- > **Servicegericht, niet bestraffend:** De bijstandsprocedure moet volstrekt helder zijn, en er moet een laagdrempelige mogelijkheid bestaan om hulp te krijgen. Naast digitale formulieren moeten er altijd alternatieve mogelijkheden bestaan om aanvragen te doen of om met de gemeente in contact te komen, om te waarborgen dat (digitale) laaggeletterdheid geen achterstelling betekent. Aan de inwoner moet regelmatig terugkoppeling worden gevraagd wat er verbeterd kan worden. Fouten in aanvragen moeten niet direct leiden tot een bestraffing van de inwoner, maar moeten in de eerste plaats worden opgevat als een falen van het systeem, dat in samenwerking met de burger kan worden opgelost⁷. De commissie adviseert een coulante opstelling tegenover fouten en onvolledigheden. Als alternatief voor een bestraffende benadering is een positieve benadering mogelijk die de inwoner stimuleert om mee te helpen bij heronderzoek. Er valt te denken aan een kleine financiële vergoeding of een andere beloning wanneer de inwoner meewerkt bij het onderzoek⁸. De ambtenaren die zijn betrokken bij bijstandsheronderzoek moeten worden getraind in een servicegerichte benadering in plaats van een wantrouwende benadering.

Deze algemene overwegingen dienen ter aanbeveling om de procedure rondom heronderzoek te verbeteren, maar ze dienen ook als achtergrond voor de volgende aanbevelingen die gericht zijn op profileringsmethoden. De commissie gaat in haar oordeel uit van een situatie die meer wordt gekarakteriseerd door strenge controle dan door dienstverlening. In een dergelijke situatie is het ingrijpend, vervelend en stressvol om te worden geselecteerd voor heronderzoek. Bovendien is er hier sprake van een antagonistische verhouding tussen burger en gemeente, die snel kan leiden tot onderling wantrouwen. Het is dan zaak om strenge eisen te stellen aan risicoprofilering, om elke schijn van vooringenomenheid en willekeur te vermijden en om de transparantie naar de burger te vergroten.

Transparantie en uitlegbaarheid

Voor de inzet van profileringsmethoden is het noodzakelijk dat beslissingen uitlegbaar zijn. Nederlandse gemeenten zijn hiertoe verplicht volgens de Algemene wet bestuursrecht

⁶ Participatiewet art. 17 eerste lid, met inachtneming van ECLI:NL:CRVB:2022:1395.

⁷ In het Franse bestuursrecht biedt *droit à l'erreur* burgers het recht te goeder trouw een administratieve fout te maken.

⁸ Hierbij moet uiteraard worden gelet op regels in de bijstand inzake het ontvangen van vergoedingen en giften.

(Awb)⁹ en de Algemene verordening gegevensbescherming (AVG)¹⁰. Maar de uitlegbaarheid van een beslissing, zoals de uitnodiging voor heronderzoek, is ook een belangrijk principe om de legitimiteit van en het vertrouwen in de overheid te waarborgen. De inzet van (semi-)geautomatiseerde besluitvorming kan moeilijkheden opleveren voor betekenisvolle uitlegbaarheid. Daarom stelt de commissie in de context van algoritme-gestuurde selectie voor heronderzoek eisen aan de uitlegbaarheid van het algoritmische model.

Xgboost onvoldoende uitlegbaar

De commissie is van mening dat het xgboost-algoritme – de machine learning methode gebruikt door de gemeente Rotterdam in de periode 2017-2021 voor het samenstellen van risicoprofielen – niet betekenisvol uitlegbaar is aan burgers. Selectiecriteria en drempelwaarden worden door xgboost gekozen aan de hand van aggregatiestatistiek over honderden of duizenden verschillende beslisbomen¹¹. Een individuele selectiebeslissing op basis van een xgboost-risicoprofiel kan daarom door de gemeente alleen worden gemotiveerd in ingewikkelde statistische termen. De commissie vindt een dergelijke uitleg te complex¹².

De commissie adviseert gemeenten daarom, indien zij algoritmische profilering willen inzetten, om minder complexe (machine learning) algoritmes te gebruiken voor de selectie van variabelen en/of drempelwaarden voor profileringscriteria¹³. Een vereiste hiervoor is dat men kan vertrouwen op de kwaliteit van de historische data waarop het algoritme getraind wordt.

Informatievoorziening en bezwaar

Betekenisvolle uitlegbaarheid is belangrijk vanuit het oogpunt van bezwaar. Een geselecteerde inwoner kan bezwaar maken als men het oneens is met een toegekende categorie, en kan de legitimiteit van deze categorie bevragen. Voor profielen gegenereerd door xgboost is dit alleen mogelijk door de kansverdeling over alle betrokken beslisbomen te bevragen. Dit maakt het voor inwoners praktisch onmogelijk om in bezwaar te gaan. De commissie adviseert gemeenten selectiecriteria expliciet te vermelden wanneer zij naar burgers communiceren waarom zij zijn geselecteerd voor heronderzoek. Mits gepaard met mogelijkheden voor bezwaar en beroep en een gepaste vorm van communicatie kan dit

⁹ Verbod op vooringenomenheid Awb 2:4, zorgvuldigheidsbeginsel Awb 3:2-3:4, motiveringsbeginsel Awb 3:47. Zie ook: *In alle openheid: transparant algoritmegebruik door de overheid*, College van de Rechten van de Mens. (2023) <https://publicaties.mensenrechten.nl/publicatie/bf15558a-1b17-43d7-a60e-df9ff8847491>

¹⁰ Artikel 13.2.f, artikel 14.2.g en artikel 15.1.h eisen dat data subjects recht hebben op “betekenisvolle informatie over de toegepaste logica” met betrekking tot profilering.

¹¹ Zie [Figuur 3](#) in Probleemstelling AA:2023:02:P⁴.

¹² Een dergelijke uitleg is mogelijk ook in strijd met gegevensbeschermingsrechten (AVG) van burgers en motiveringsverplichtingen voor Nederlandse publieke organisaties (Awb). Concrete rechterlijke uitspraken hierover zijn nog niet beschikbaar.

¹³ Denk aan parametrische en non-parametrische statistische methoden die wel middels discrete categorisering uitlegbaar zijn, bijv. logistische regressie of L1-regularisatiemethoden.

bijdragen aan de behoorlijkheid van gemeentelijke besluiten en dus aan het vertrouwen van burgers in de overheid.

Expert-gedreven versus algoritmisch profileren

Bij de beoordeling van algoritmische profileringsmethoden is het van belang om mee te wegen wat mogelijke alternatieven zouden zijn. In de context van deze casus is profilering door domeinexperts het meest directe alternatief voor algoritmische profilering. Risicoprofielen die handmatig worden samengesteld door domeinexperts zijn betekenisvoller uitlegbaar dan profilering door het xgboost-algoritme (zie [Box 2](#)). Dit neemt niet weg dat men ook voor handmatig samengestelde profielen moet kunnen motiveren waarom bepaalde variabelen en drempelwaarden zijn gekozen als selectie criterium. Hiervoor grijpt men vaak terug op statistische onderbouwing, waardoor het verschil tussen handmatige en (uitlegbare) algoritmische profilering in de praktijk minder groot is dan het lijkt. Het is daarnaast belangrijk om te beseffen dat ook bij simpele en handmatig samengestelde profielen risico's bestaan op (proxy-)discriminatie en vooringenomenheid. Handmatige profilering vereist een soortgelijke afweging van gebruikte variabelen als voor algoritmische methoden, een afweging die bovendien inhoudelijk en waardegedreven moet worden onderbouwd. Het advies over verantwoorde variabelenselectie hieronder dient als een voorzet hoe een dergelijke afweging kan worden gemaakt.

De commissie acht het wenselijk om verschillende selectiemethoden naast elkaar te gebruiken. Het laten domineren van één soort profileringsmethode kan tunnelvisie en verkeerde feedbackloops in de hand werken. Dit betekent dat één bepaalde groep er overmatig uit wordt gepikt, waarbij andere groepen de dans ontspringen. Hieruit kan een

Box 2

Uitlegbaarheidsvereisten voor risicoprofilering

Selectie door een xgboost-profiel kan alleen worden uitgelegd met behulp van de kansverdeling over betrokken beslisbomen:

“U bent geselecteerd omdat X/N beslisbomen u op basis van uw leeftijd hebben geselecteerd en Y/N beslisbomen op basis van uw woonvorm hebben geselecteerd”.

In het algemeen dient profilering makkelijk uitlegbaar te zijn aan de hand van discrete categorisering, zoals bij:

“U bent geselecteerd omdat uw leeftijd tussen A en B is en u woonvorm C heeft”.

feedbackloop ontstaan: een oververtegenwoordiging van een bepaalde groep in de dataset leidt dan tot steeds hogere risicoscores en dus een vooringenomenheid met betrekking tot die groep. Het rouleren van risicoprofielen en het parallel gebruiken van meerdere selectiemethoden moet zulke patronen doorbreken en risico's mitigeren. Met name een doorlopend gebruik van aselechte steekproeven is hier van belang. De resultaten van aselechte steekproeven kunnen dienen om mogelijke vooringenomenheid van profileringsmethoden voortdurend te monitoren.

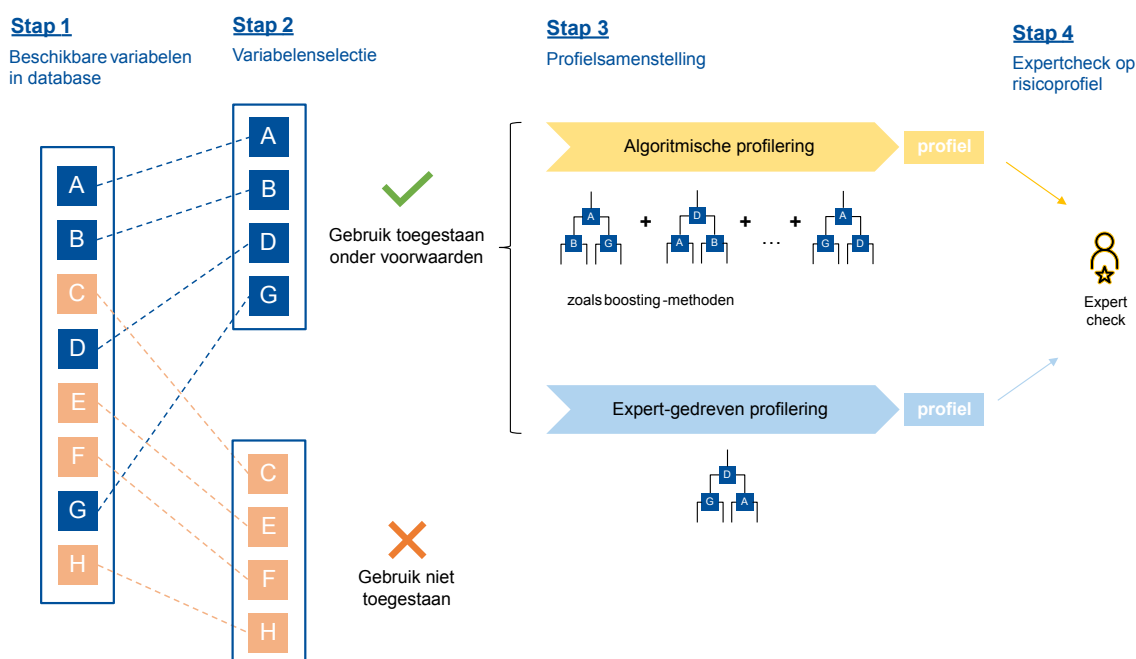
Variabelenselectie voor profilering

Voordat risicoprofielen worden samengesteld door machine learning algoritmen (of door domeinexperts) is het belangrijk om stil te staan bij de vraag welke variabelen in aanmerking komen voor gebruik in zulke risicoprofielen. Om discriminerende vooringenomenheid en andere vormen van ethisch onwenselijk onderscheid te vermijden moeten sommige variabelen op voorhand worden uitgesloten. Een schematisch overzicht van dit proces is weergegeven in [Figuur 1](#). De commissie geeft enkele algemene adviezen over de selectie van geschikte variabelen, om vervolgens een concreet oordeel te vellen over individuele variabelen die uit de casus-Rotterdam naar voren kwamen.

(Proxy-)discriminatie

Bij het overwegen van variabelen is het noodzakelijk om persoonskenmerken te vermijden die gelden als beschermde grond (zoals nationaliteit en religie) en die tot discriminatie kunnen leiden. Dit is in de meeste gevallen ook wettelijk verboden¹⁴. Ook het gebruik van kenmerken die geen klassieke beschermde grond vormen in non-discriminatierecht, zoals mentale gezondheid, kan wettelijk verboden zijn. Daarnaast kunnen ogenschijnlijk onschuldige variabelen als zogenaamde 'proxy' fungeren voor beschermde persoonskenmerken, omdat zij daarmee een sterke statistische correlatie vertonen. Dit betekent dat het gebruik van zulke proxy-variabelen alsnog kan leiden tot discriminatoire vooringenomenheid van het model. Dit geldt bijvoorbeeld voor de kenmerken laaggeletterdheid of postcode (die gelden als proxy's voor o.a. afkomst). In het algemeen is zulke proxy-discriminatie duidelijk onwenselijk. Aan de andere kant kunnen er gegronde redenen bestaan om een dergelijk onderscheid te maken, om doelgroepen gericht op te zoeken. In het geval van profilering voor heronderzoek is het duidelijk dat laaggeletterdheid een indicatie kan zijn voor het risico op fouten in een bijstandsaanvraag. Het meenemen van deze variabele in een risicoprofiel zou de gemeente in staat stellen om een hulpbehoevende doelgroep extra te helpen. Zoals echter hierboven is uitgelegd, gaat de commissie in haar oordeel niet uit van een dienstverlenende inzet van risicoprofilering, maar van een streng controlerende inzet. In dat geval is de afweging volstrekt anders, omdat discriminatoire vooringe-

¹⁴ Algemene wet gelijke behandeling (AWGB), Europese Verdrag van de Rechten van de Mens (EVRM), Algemene verordening gegevensbescherming (AVG). Bij toetsing aan artikel 14 EVRM en artikel 1 EVRM Protocol 12 bestaat enige ruimte om onderscheid op grond van afkomst te rechtvaardigen, zij het wel dat er dan sprake moet zijn van zeer zwaarwegende redenen. Zie ook Discriminatie door Risicoprofielen, College van de Rechten van de Mens (2021) <https://www.mensenrechten.nl/actueel/nieuws/2021/11/30/discriminatie-door-risicoprofielen-voorkom-etnisch-profileren>



Figuur 1 – Schematische weergave van variabelenselectie voor risicoprofilering

nomenheid vermeden moet worden als het gaat om de selectie van inwoners voor ingrijpend en stressvol heronderzoek. Voor interne diagnostiek, om te onderzoeken waarom bepaalde doelgroepen meer moeite hebben met bijstandsaanvragen, ziet de commissie meer ruimte om zulke kenmerken mee te nemen.

Onveranderlijke persoonskenmerken

Bij het selecteren van geschikte variabelen zou een eerste overweging kunnen zijn om onveranderlijke persoonskenmerken (zoals de meeste beschermde gronden) uit te sluiten en alleen keuzeafhankelijke kenmerken mee te nemen. De fundering hiervoor is dat personen in ieder geval nooit op basis van kenmerken mogen worden beoordeeld waar zij geen keus in hebben gehad. Het probleem is echter dat dit onderscheid in de praktijk lastig te maken valt. Leeftijd is bijvoorbeeld een onveranderlijk persoonskenmerk waar wel vaak onderscheid op wordt gemaakt – met reden. Het is staande praktijk dat personen boven een pensioengerechtigde leeftijd worden uitgezonderd van bijstandsheronderzoek. En veranderlijke kenmerken zijn in principe, maar vaak niet in de praktijk toe te rekenen aan iemands vrije keuze, zoals de plek waar iemand woont (dus o.a. postcode). Het voorbeeld van postcode laat ook zien dat een onderscheid tussen onveranderlijke en keuzeafhankelijke variabelen geen pasklaar antwoord biedt op het vraagstuk van proxy-discriminatie.

Inhoudelijke relevantie

Een ander mogelijk te generaliseren criterium voor variabelenselectie is het inhoudelijke verband met bijstandsonderzoek en risico's op onrechtmatige toekenning. Variabelen kunnen worden uitgesloten die inhoudelijk niet redelijkerwijs te relateren zijn aan dit doel. Maar ook dan moet de relevantie van factoren in algemene zin worden afgewogen tegenover het risico op proxy-discriminatie en andere vormen van onwenselijk onder-

scheid. Deze afweging kan alleen per individuele variabele worden gemaakt. Een voorbeeld van een indicatie die redelijkerwijs verband houdt met bijstandsonderzoek is een 'no show': het onaangekondigd niet komen opdagen bij een afspraak met de gemeente waaruit blijkt dat een inwoner onvoldoende meewerkt.

Subjectiviteit

Een volgende overweging is de objectiviteit en de onveranderlijkheid van data voor bepaalde kenmerken. Het is belangrijk om in het oog te houden dat verkregen data nooit één-op-één samenvalt met de realiteit, en dat dit voor sommige kenmerken erg problematisch kan zijn. Allereerst zijn sommige kenmerken lastig te kwantificeren of zijn ze subjectief. Dit geldt met name voor persoonlijke eigenschappen die niet objectief vast te stellen zijn, zoals de kenmerken 'assertiviteit' of 'verzorgdheid' die in de casus naar voren komen. Ook protocollen die de scoring van zulke eigenschappen moeten standaardiseren kunnen nooit de subjectiviteit volledig wegnemen. Een andere variabele die de commissie op voorhand uitsluit zijn open tekstvelden waar ambtenaren naar eigen inzicht overige notities kunnen vermelden¹⁵. In de vertaling naar kwantitatieve waarden worden zulke notities gerepresenteerd in binaire vorm (wel of niet aanwezige notitie) die de inhoudelijke betekenis niet weergeven.

Veranderlijkheid

Daarnaast is de verkregen data voor bepaalde kenmerken erg veranderlijk en niet representatief. Door de veranderlijkheid van de werkgelegenheid is data uit het ene jaar over het aandeel in bijstandsaanvragen van bepaalde beroepen of competenties niet representatief voor het volgende jaar. Ook voor personeigenschappen moet een sterk voorbehoud worden gemaakt: een enkele meting mag niet tot onbepaalde tijd worden opgeslagen en worden gebruikt als objectieve score van een bepaalde eigenschap. Een incidentele aanvaring met een ambtenaar mag geen eeuwig vonnis betekenen voor iemands risicoscore. Er dienen protocollen in werking te zijn die bepalen wanneer verkregen data moet worden gewist, en op welke manier een inwoner inzicht kan krijgen in deze gegevens om het aan te vechten. Aangezien dit onder governance valt gaat de commissie hier niet verder op in.

Feedback loops

Een laatste advies over variabelenselectie geldt de noodzaak om variabelen en risicoprofielen te rouleren. Het is onwenselijk als jaar in jaar uit dezelfde kenmerken bepalen wie wordt geselecteerd voor heronderzoek. Naast dat dit op zichzelf als onwenselijk en oneerlijk kan worden beschouwd, bestaat hier ook het gevaar van feedbackloops. Overmatige concentratie op bepaalde groepen en kenmerken kan leiden tot een zichzelf versterkend proces, waarbij gevonden onrechtmatigheden voor een overgerepresenteerde groep onrecht leiden tot een steeds grotere risicoscore voor diezelfde groep.

¹⁵ Zie Hof van Justitie van de Europese Unie (HvJ-EU), C 817/19, Ligue des droits humains ASBL v Conseil des ministres, 21 June 2022, ECLI:EU:C:2022:491, mn. 130-140.

(On)verantwoord gebruik van variabelen

Gegeven bovenstaande overwegingen acht de commissie sommige variabelen¹⁶ verantwoord en andere variabelen onverantwoord om te gebruiken voor profilering in de context van heronderzoek bij bijstandsuitkeringen. Een overzicht voor gebruik van 17 verschillende variabelen als selectie criterium kan worden gevonden in de [Infographic](#). Dit advies wordt gemotiveerd aan de hand van verschillende beweegredenen die hierboven zijn besproken. Een overzicht van beweegredenen is gegeven in de legenda. Beheersbare risico's refereren naar adequate algoritmebeheersmethoden op onder andere organisatorisch en technisch vlak¹⁷.

¹⁶ Gebaseerd op de lijst met variabelen vermeld in Appendix B – Dataverzameling uit de Probleemstelling AA:2023:02⁴

¹⁷ Zie bijvoorbeeld Onderzoekskader Algoritmes, Auditdienst Rijk (2023)

<https://www.rijksoverheid.nl/documenten/rapporten/2023/07/11/onderzoekskader-algoritmes-adr-2023>

Samenstelling normatieve adviescommissie

Dit advies is de uitkomst van een gezamenlijk deliberatief proces. Specifieke uitspraken in dit document corresponderen daarom niet noodzakelijkerwijs met het standpunt van individuele leden van de adviescommissie. Commissieleden kunnen niet individueel verantwoordelijk worden gehouden voor dit advies.

Datum

De normatieve adviescommissie is op 30 juni 2023 fysiek bijeengekomen in Rotterdam. Dit adviesdocument is op 30 augustus 2023 goedgekeurd door alle leden van de adviescommissie.

Samenstelling normatieve adviescommissie

De normatieve adviescommissie voor deze casus bestaat uit:

- Abderrahman El Aazani, onderzoeker bij Ombudsman Rotterdam-Rijnmond
- Francien Dechesne, Associate Professor Law and Digital Technologies, Universiteit Leiden
- Maarten van Asten, wethouder Financiën, Digitalisering, Sport en Evenementen Gemeente Tilburg
- Munish Ramlal, ombudsman Metropool Amsterdam
- Oskar Gstrein, Assistant Professor Governance and Innovation, Rijksuniversiteit Groningen.

Dankwoord

Naast vele anderen die we hebben gesproken of die kritisch hebben meegelezen, willen we de volgende personen en organisaties speciaal bedanken voor hun bijdrage aan het onderzoek:

- Hinda Haned, professor Responsible Data Science, Universiteit van Amsterdam
- Participatieraad Amsterdam
- Lighthouse Reports
- VPRO Argos.

Financiering van Algorithm Audit

Stichting Algorithm Audit wordt ondersteund door onafhankelijke publieke en filantropische subsidies. Subsidies worden ingezet voor gevraagd en ongevraagd onderzoek en het opstellen van probleemstellingen, waarin ethische kwesties die zich voordoen bij de inzet van algoritmes worden gespecificeerd. De experts en belanghebbenden die deelnemen in een normatieve adviescommissie wordt een vergoeding aangeboden. We delen de inzichten voortkomend uit ons werk met zowel de samenleving als de internationale AI auditing community om publieke kennis op te bouwen over de verantwoorde inzet van algoritmen. Werken zonder winstoogmerk dient de doelen van de stichting het beste.



Structurele partners van Algorithm Audit

SIDNfonds

SIDN Fonds

Het SIDN Fonds staat voor een sterk internet voor iedereen. Het Fonds investeert in projecten met lef en maatschappelijke meerwaarde, met als doel het borgen van publieke waarden online en in de digitale democratie.

European
Artificial Intelligence
& Society Fund

European AI&Society Fund

Het European AI&Society Fund ondersteunt organisaties uit heel Europa die AI beleid vormgeven waarin mens en maatschappij centraal staan. Het fonds is een samenwerkingsverband van 14 Europese en Amerikaanse filantropische organisaties.



Ministerie van Binnenlandse Zaken en
Koninkrijksrelaties

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties

Het ministerie van BZK maakt zich sterk voor een democratische rechtsstaat, met een slagvaardig bestuur. Ze borgt de kernwaarden van de democratie. BZK staat voor een goed en digitaalvaardig openbaar bestuur en een overheid waar burgers op kunnen vertrouwen.



 www.algorithmaudit.eu

 www.github.com/NGO-Algorithm-Audit

 info@algorithmaudit.eu

Stichting Algorithm Audit is geregistreerd bij de Kamer
van Koophandel onder nummer 83979212