



Profielen bijstandsinstream Eindhoven

Hanneke Posthumus

Sophie Doove

Lydia Geijtenbeek

Rosa Meijer

1. Managementsamenvatting

Eindhoven heeft via een *machine learning* algoritme inzichten vergaard in het sociaal domein. Dit is gebeurd in het Urban Data Center Eindhoven, een samenwerking tussen de gemeente Eindhoven en het CBS. Hiervoor is een onderzoeksbestand gebruikt dat het CBS in opdracht van VNG Realisatie heeft samengesteld en dat informatie bevat over *life events* en een groot aantal achtergrondkenmerken van personen die gebruik maken van voorzieningen in het sociaal domein¹. Dit onderzoek is een verkenning van de mogelijkheden van machine learning (meer specifiek: classificatiebomen) voor het voorspellen van instroom of uitstroom in het sociaal domein. Hiertoe heeft het CBS een classificatieboom algoritme gebruikt om instroom in de algemene bijstand te voorspellen, op een dataset die de variabelen uit het bestand van VNG Realisatie combineert met instroomgegevens over 2015.

De uitkomst van het algoritme bestaat uit een classificatieboom, waaruit vijf profielen zijn gedestilleerd. Binnen deze profielen is de kans op instroom in Eindhoven gemiddeld 47%, en tezamen beschrijven de profielen 23% van de Eindhovense instroom in 2015. Dit zijn de profielen:

1. Herinstromers die recent zonder inkomen of in de WW kwamen te zitten.
2. Herinstromers die langer geleden zonder inkomen of in de WW kwamen te zitten, en geen partner hebben met inkomen uit werk, pensioen of een arbeidsongeschiktheidsuitkering.
3. Mensen die zich afgelopen jaar inschreven in de BRP en in een institutioneel huishouden zaten. Dit zijn waarschijnlijk grotendeels vergunninghouders.
4. Personen die recent zonder inkomen of in de WW kwamen te zitten, niet in de bijstand zaten, zonder partner met arbeidsongeschiktheidsuitkering, met sociale huurwoning, en het voorgaande jaar een laag inkomen en weinig vermogen hadden.
5. Personen die langer geleden in de WW kwamen, onder de 52 zijn, en ingeschreven staan bij het CWI², en verder net als profiel 4: niet in de bijstand zaten, zonder partner met arbeidsongeschiktheidsuitkering, met sociale huurwoning, en het voorgaande jaar een laag inkomen en weinig vermogen hadden.

Aangezien bijna de helft van de personen in deze profielen binnen een jaar instroomt in de bijstand, zou je met deze profielen in de hand heel gericht personen willen benaderen die dreigen in te stromen en daarmee eventueel instroom voorkomen. Het gaat hierbij om flinke aantallen: in januari 2015 bevatten deze profielen ruim 600 Eindhovenaren waarvan er ruim 300 later dat jaar zouden instromen. Voor profielen 4 en 5 geldt echter dat deze grotendeels gebaseerd zijn op databronnen die niet bekend zijn bij de gemeente (zoals sociaal economische status of inkomen), terwijl het voor vergunninghouders in profiel 3 lastig is proactief beleid te voeren. Daarmee zijn voor de gemeente met name herinstroomprofielen 1 en 2 geschikt om proactief beleid op te voeren. Het gaat hierbij om grofweg 100 van de 1300 jaarlijkse instromers in de bijstand in Eindhoven.

¹ Een toelichting op dit onderzoek staat op de CBS website: <https://www.cbs.nl/nl-nl/maatwerk/2018/13/longitudinale-analyse-sociaal-domein-2015>

² Tegenwoordig UWV Werkbedrijf.

Inhoudsopgave

1.	Managementsamenvatting	2
2.	Inleiding	4
3.	Methode en data	5
3.1	Onderzoeksopzet	5
3.2	Methode: classificatieboom	5
3.3	Data	6
3.4	Variabelen	7
4.	Resultaten	10
4.1	Classificatieboom	10
4.2	Vijf profielen	12
4.3	Profiel 1	13
4.4	Profiel 2	14
4.5	Profiel 3	15
4.6	Profiel 4	16
4.7	Profiel 5	17
5.	Conclusie	19
6.	Reflectie	20
6.1	Technische keuzes en beperkingen	20
6.2	Aanbevelingen	21
7.	Bijlagen	22
7.1	Tabellen en grafieken	22
7.2	Achtergrond bij het classificatieboom algoritme	25
7.3	Achtergrond bij operationalisering van de variabelen	26

2. Inleiding

Sinds de decentralisaties van de Wmo2015, Participatiewet en Jeugdwet hebben gemeenten er belangrijke taken bijgekregen. Zo zijn gemeenten sindsdien verantwoordelijk voor de financiering en uitvoering van regelingen in het sociaal domein die voorheen onder de verantwoordelijkheid vielen van de Rijksoverheid. Sindsdien hebben gemeenten een enorme behoefte aan feitelijke informatie over het gebruik van verschillende voorzieningen in het sociaal domein. VNG Realisatie heeft het CBS daarom opdracht gegeven om speciaal voor gemeentes een onderzoeksbestand samen te stellen. Dit bestand bevat informatie over *life events* (ofwel variabelen die een gebeurtenis beschrijven) van alle Nederlanders en daarnaast een groot aantal achtergrondkenmerken van personen die gebruik maken van voorzieningen in het sociaal domein³.

De gemeente Eindhoven biedt ondersteuning aan personen die deelnemen aan regelingen binnen het sociaal domein, maar wil liever nog mensen helpen om te voorkomen dat het zo ver komt. Een datagerichte aanpak kan helpen om potentiële instromers vroegtijdig te signaleren. Je kan bijvoorbeeld een *machine learning*-algoritme (zoals een classificatieboom algoritme) gebruiken, om personen te verdelen in groepen met een grote of juist kleine instroomkans. Hieruit kan je vervolgens profielen maken met achtergrondkenmerken van personen die een grote kans hebben om in te stromen. Met deze instroomprofielen in de hand kan de gemeente Eindhoven vervolgens preventief beleid ontwikkelen voor groepen met de grootste instroomkans. Gezien deze kans is vanuit het Urban Data Center Eindhoven – waarin de gemeente en het CBS samenwerken aan de beantwoording van beleidsvragen – besloten een classificatieboom te maken. Om te beginnen is gekozen om profielen vast te stellen voor de instromers in de bijstand: Wat zijn hun profielen? En hoe vaak komen deze in Eindhoven voor? En hoe zijn deze verdeeld over de werkgebieden?

Dit onderzoek is een eerste verkenning van de mogelijkheden van een dergelijke datagerichte aanpak. Het onderzoeksbestand voor VNG Realisatie is hierbij het uitgangspunt, omdat dit gericht is op personen in het sociaal domein en de life events mogelijk samenhangen met instroom in het sociaal domein. Het voornaamste doel van dit onderzoek is om te bepalen in hoeverre een classificatieboom en de instroomprofielen die daaruit voortkomen, kunnen helpen bij beleid gericht op preventie van instroom in het sociaal domein. De uitkomsten van dit onderzoek laten zien wat er wel en niet mogelijk is met een dergelijke datagerichte aanpak en welke aanvullingen gewent zijn om deze optimaal te laten aansluiten op de alledaagse werkelijkheid van het beleid.

Dit rapport is als volgt opgebouwd. In hoofdstuk 3 beschrijven we de onderzoeksmethode en lichten we het gebruik van classificatiebomen toe. Ook beschrijven we de databronnen en variabelen die we hierbij gebruiken en deels zelf afgeleid hebben. In hoofdstuk 4 volgen de uitkomsten, waarbij we zowel de classificatieboom als de profielen laten zien, met daarbij aantallen voor Eindhoven en haar werkgebieden. We sluiten af met de conclusies en reflecties in hoofdstuk 5 en 6 en de bijlages in hoofdstuk 7.

³ Een toelichting op dit project staat op de CBS website: <https://www.cbs.nl/nl-nl/maatwerk/2018/13/longitudinale-analyse-sociaal-domein-2015>

3. Methode en data

3.1 Onderzoekopzet

Het doel van dit onderzoek is om te bepalen of classificatiebomen inzichten opleveren die bruikbaar zijn om de instroom in het sociaal domein te beperken. Dit is onderzocht met een testcase: Is het mogelijk om met classificatiebomen te identificeren welke personen in Eindhoven een grote kans hebben om in te stromen in de bijstand? Een belangrijke beperking is dat de aantallen instromers in de bijstand in Eindhoven te klein zijn om gedetailleerde profielen mee te kunnen bepalen die statistisch betrouwbaar zijn. Er is dan een groot risico op 'overfitting', ofwel dat je profielen vindt die toevallig voor 2015 relatief veel instromers bevatten, maar die in een toekomstige dataset niet algemeen geldig zullen blijken.

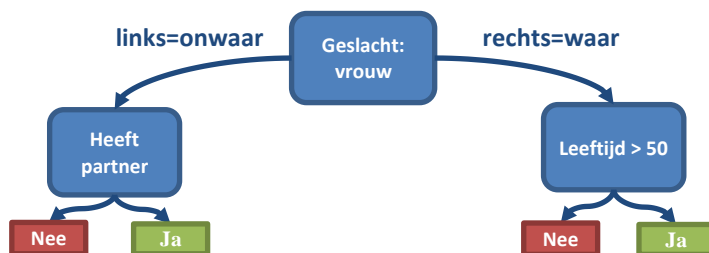
Daarom hebben we in dit onderzoek gekozen voor een tweetraps-aanpak:

1. In de eerste trap stellen we risicoprofielen op voor instroom in de bijstand in heel Nederland. Voor de bepaling van deze risicoprofielen gebruiken we een classificatieboom algoritme, dat aan de hand van persoons- en life eventvariabelen groepen probeert te maken met een hoog of juist laag risico op instroom (zie sectie 3.2). De resulterende classificatieboom gebruiken we om profielen te bepalen van personen met een groot instroomrisico.
2. In de tweede trap gebruiken we de profielen om in kaart te brengen hoeveel Eindhovenaren er in elk profiel zitten en wat voor Eindhoven de instroomkans is per profiel. Ook geven we de aantallen instromers per profiel voor heel Eindhoven en voor elk van haar werkgebieden.

3.2 Methode: classificatieboom

Een classificatieboom probeert met een algoritme een populatie in zo homogeen mogelijke groepen te verdelen aan de hand van een aantal verklarende variabelen. In dit geval betekent dit dat het algoritme groepen probeert te maken met meer of juist minder instromers dan gemiddeld in Nederland. Het gebruikte algoritme begint met de hele populatie en verdeelt deze vervolgens in steeds kleinere en homogener groepen aan de hand van de verklarende variabelen. In elke stap probeert het algoritme alle variabelen uit, meet voor elke variabele hoeveel homogener de groepen door splitsing daarop zouden worden en kiest de variabele uit die het hoogste scoort. Met deze variabele maakt de boom twee nieuwe groepen, die vervolgens opnieuw verbeterd worden in de volgende stap. Het resultaat van dit hele proces kan je uitbeelden als een diagram met een boomstructuur, vandaar de naam. Het classificatieboom algoritme kan op verschillende manieren worden ingesteld. De keuzes die wij hebben gemaakt, lichten we toe in Bijlage 5.2.

Figuur 1 geeft een voorbeeld van een heel kleine hypothetische classificatieboom. In deze boom is 'Geslacht' zichtbaar het kenmerk dat van alle variabelen dat de populatie het best kan splitsen in groepen met vooral "Ja" en "Nee". Vervolgens geldt voor vrouwen dat de variabele 'Leeftijd' het beste de groepen onderscheidt, en voor mannen is dat of ze een partner hebben ('Heeft partner'). Figuur 2 bevat een vergelijkbare figuur voor de uiteindelijke classificatieboom van dit onderzoek, en sectie 4.1 geeft hier een toelichting op.



Figuur 1 Voorbeeld van een hypothetische classificatieboom

3.3 Data

Populatie

De populatie voor dit onderzoek bestaat uit alle personen die per 1-1-2015 stonden ingeschreven in de Basisregistratie Personen (BRP), exclusief groepen die vanwege hun achtergrondkenmerken niet in aanmerking komen om in de bijstand in te stromen. Dit betekent concreet dat de volgende groepen zijn uitgesloten van de populatie:

- Personen die op 1-1-2015 jonger zijn dan 18 jaar, want onder de 18 heb je geen recht op bijstand.
- Personen die op 1-1-2015 ouder zijn dan 65 jaar en 3 maanden, want vanaf die leeftijd heb je recht op een AOW-uitkering.
- Personen die op 1-1-2015 reeds een bijstandsuitkering ontvangen.

De totale onderzoekspopulatie bestaat uit 10 079 000 personen⁴, waarvan er 87 000 tussen februari en december 2015 minimaal een keer instroomden in de bijstand.

Samenstelling bestand

Voor dit onderzoek hebben we twee bestanden gebruikt die in opdracht van VNG Realisatie zijn verzameld voor gebruik door gemeentes. Allereerst gaat het om het *peilmomentenbestand* met daarin voor elke regeling in het sociaal domein of een persoon daar gebruik van maakte op 1-1-2015. Daarnaast is het *life-eventbestand* gebruikt met daarin voor verschillende life events (zoals geboorte, verhuizing) de laatste gebeurtenis in 2014 en 2015.

Naast deze bestanden is voor dit onderzoek ook gebruik gemaakt van het Stelsel van Sociaal-Statistische Bestanden (SSB). Hieruit is iedereen geselecteerd die per 1-1-2015 stond ingeschreven in de BRP. Uit dit bestand hebben we variabelen gehaald van allerlei relevante achtergrondkenmerken, zoals geslacht of leeftijd, maar ook andere relevante variabelen zoals inkomen of woningeigenschappen.

Tot slot hebben we zelf nog nieuwe life-eventvariabelen bepaald, voor gebeurtenissen waarvan we in een vooronderzoek zagen dat ze vaak voorkomen vlak voordat iemand instroomt. Het gaat hierbij om uitstroom uit de bijstand, verandering van positie in het huishouden en verandering van sociaal economische klasse (SEC). De sociaal economische klasse geeft aan of iemand een voltijds opleiding volgt en wat zijn of haar voornaamste inkomensbron is.

De variabelen die in dit onderzoek zijn gebruikt worden hieronder uitgebreider toegelicht. Tabel 1 bevat een compleet overzicht van alle gebruikte variabelen.

⁴ Aantallen worden afgerond op 1000 tallen, tenzij anders aangegeven.

3.4 Variabelen

Doelvariabele: instroom

Het doel van dit onderzoek is om te bepalen of classificatiebomen inzicht kunnen geven in de profielen van instromers in het sociaal domein, waarbij de bijstand als testcase dient. De afhankelijke variabele is dan ook 'instroom', die aangeeft of iemand instroomt in de bijstand, waarbij we ons beperken tot bijstandsvormen die vallen onder de Participatiewet. Concreet kijken we naar instroom in de bijstand in de periode 1-2-2015 tot en met 31-12-2015. We hebben gekozen voor deze periode omdat per 1-1-2015 de Participatiewet is ingegaan, waardoor instroom in 2014 of eerder misschien niet vergelijkbaar is met huidige of toekomstige instroom. Daarnaast bevat het onderzoeksbestand dat is samengesteld voor VNG Realisatie waar we op voortbouwen, juist over deze periode informatie. Verder hebben we januari 2015 buiten beschouwing gelaten, omdat de invoering van de Participatiewet per 1-1-2015 mogelijk heeft geleid tot administratieve correcties in januari. We definiëren een persoon als 'instromer' wanneer hij of zij voorkomt in het bijstandsregister, terwijl dit de voorgaande kalendermaand niet het geval was. Naast de variabele 'instroom' is de variabele 'instroommaand' afgeleid. Deze gebruiken we om te bepalen hoe ver voor het moment van instroom een andere gebeurtenis (life event) plaatsvond. Ons onderzoeksbestand bevat 87 000 instromers.

Controlegroep

Aangezien we in dit onderzoek instromers willen vergelijken met niet-instromers hebben we ook een *controlegroep* nodig. Om te voorkomen dat het bestand te groot wordt, nemen we hiertoe een aselechte steekproef uit de onderzoekspopulatie van niet-instromers. De controlegroep is vijf keer zo groot als het aantal instromers, en bestaat uit 437 000 personen.

Voor elk van de niet-instromers in de controlegroep maken we bovendien een *fictieve instroommaand* aan, in exact dezelfde verhouding als de daadwerkelijke instroommaanden van instromers. De fictieve instroommaanden zijn nodig, omdat de meeste gebeurtenis-variabelen (life events) een relatieve tijdsaanduiding hebben ten opzichte van de instroommaand (i.e. het aantal maanden vóór instroom). Dankzij de fictieve instroommaanden kunnen we deze variabelen op dezelfde manier bepalen voor de controlegroep als voor de instromers. Voor meer toelichting, zie Bijlage 5.3.

Variabelen voor classificatie

In dit onderzoek waren er 33 variabelen waaruit het classificatieboom algoritme kon kiezen om instromers te onderscheiden van niet-instromers, zie Tabel 1 voor een overzicht hiervan. Voor het lezen van deze tabel is het handig om te beseffen dat deze variabelen zijn onder te verdelen in drie soorten:

1. Variabelen die de situatie aangeven op een bepaald *peilmoment*, zoals leeftijd, of woonplaats. Hier is het peilmoment steeds 1-1-2015.
2. Variabelen die betrekking hebben op een bepaalde *periode*, zoals loon of inkomen. Hier gaat het dan steeds om het voorgaande jaar, ofwel de periode 1-1-2014 tot en met 31-12-2014. Hieronder vallen ook variabelen die aangeven of iemand in een periode gedetineerd is geweest, zorg met verblijf heeft ontvangen, of voortijdig schoolverlater was⁵.
3. Variabelen die een *verandering* beschrijven, zoals een huwelijk of verlies van een baan. Analoog aan het bestand voor VNG Realisatie noemen we deze veranderingen *life events*.

⁵ Deze laatste drie zijn *geen* life events, omdat het niet gaat om het moment dat het gebeurt (bijvoorbeeld het moment dat iemand in de gevangenis terecht komt), maar of iemand in een bepaald jaar één of meer dagen in een bepaalde situatie heeft verkeerd (dus of iemand tijdens dit jaar gedetineerd is geweest).

Deze life events zijn altijd bepaald binnen een vooraf afgebakende periode. Meestal is dat '12 maanden voor instroom'. Binnen de life-eventvariabelen is onderscheid te maken tussen drie typen variabelen. Dit onderscheid wordt in Bijlage 5.3 toegelicht.

Tabel 1 Variabelen die zijn gebruikt in het classificatieboom algoritme

Nr	Naam	Soort variabele	Toelichting
1	Leeftijd*	moment 1-1-2015	
2	Geslacht	moment 1-1-2015	
3	Herkomst*	moment 1-1-2015	Nederland, Suriname, Turkije, Marokko, Ned. Antillen, overige westers, overig niet-westers, onbekend
4	Burgerlijke staat*	moment 1-1-2015	Gehuwd, partner, verweduwd, gescheiden, ongehuwd
5	Heeft een partner	moment 1-1-2015	Woont samen, al dan niet geregistreerd
6	Eigendom woning*	moment 1-1-2015	Eigen bezit, sociale huur over overige huur
7	Woningoppervlakte*	moment 1-1-2015	
8	Netto vermogen*	moment 1-1-2015	Vermogen min schuld
9	Overige schuld	moment 1-1-2015	Schuld van het huishouden, excl. Hypotheek
10	Ingeschreven CWI*	moment 1-1-2015	Ingeschreven bij CWI, tegenwoordig UWV Werkbedrijf
11	Indicatie WSW	moment 1-1-2015	Indicatie Wet Sociale Werkvoorziening
12	Inkomen*	periode 2014	Persoonlijk inkomen uit werk ⁶
13	Huishoudinkomen*	periode 2014	Totaal inkomen uit werk in huishouden ⁶
14	Soort baan*	periode januari 2015	DGA, stagiair, uitzend, oproep, WSW, anders
15	Soort contract	periode januari 2015	Vast of tijdelijk
16	Gewerkte dagen	periode januari 2015	
17	Zorg met verblijf	periode 2014	
18	Voort. schoolverlater	periode 2014	
19	Gedetineerd	periode 2010-2014	
20	Uitstroom uit bijstand*	life event II	Laatste uitstroom voor instroom
21	Sociaal economische status (SEC)*	life event III	Laatste verandering in SEC ⁷ , inclusief SEC voor en na verandering. De categorieën zijn: "werk of pensioen",

⁶ We hadden ook beschikking over andere inkomensvariabelen dan primair inkomen uit werk, zoals bruto inkomen, of besteedbaar inkomen. Omdat in deze laatste twee echter ook het bijstandsbedrag is verrekend, hebben we die buiten beschouwing gelaten.

⁷ Naast sociaal economische status (SEC) hadden we ook beschikking over variabelen over deelname op 1-1-2015 aan allerlei regelingen (zoals IOW, IVA, WGA, WAZ, WAO, WAJONG, IOAZ, IOAW, BBZ, SRG), maar het zou verwarrend kunnen zijn om life events en peilmomentvariabelen over vergelijkbare onderwerpen te hebben. Bovendien bleken deze variabelen niet relevant bij het onderscheiden van instromers en niet-instromers.

Nr	Naam	Soort variabele	Toelichting
			“WW”, “bijstand”, “uitkering ziekte/AO”, “overige sociale voorziening”, “studie”, “geen inkomen”.
22	Sociaal economische status (SEC) partner*	life event III	Idem, voor eventuele partner. Anders leeg.
23	Instroom in BRP*	life event II	Inschrijving bij Nederlandse gemeente, na immigratie of langdurig verblijf buitenland
24	Diploma ontvangen	life event I	
25	Instroom in BRP vanuit instelling*	life event II	Inschrijving bij Nederlandse gemeente na verblijf in institutioneel huishouden. Veelal asielzoekers.
26	Verhuisd	life event I	
27	Kind geboren	life event I	
28	Gehuwd	life event I	
29	Gescheiden	life event I	
30	Verweduwd	life event I	
31	Verlies familielid	life event I	Overlijden eerstegraads familielid
32	Start schuldsanering (WSNP)	life event I	
33	Plaats in huishouden ⁸	life event III	Kind, alleenstaande, partner, ouder met partner, alleenstaande ouder, institutioneel huishouden, overig/onbekend

* Variabele die door de boom is gekozen

⁸ We hadden ook informatie over veranderingen in het type huishouden, maar de plaats in het huishouden leek ons relevanter. Bovendien bleek type huishouden niet relevant voor het onderscheiden van instromers.

4. Resultaten

4.1 Classificatieboom

Toelichting bij de figuur

De classificatieboom die dit onderzoek oplevert, is te zien in Figuur 2⁹. De boom in deze figuur zit als volgt in elkaar. De boom staat in feite op zijn kop, en bestaat uit knopen (beslisregels), takken (pijlen), en bladeren (uitkomsten). De beslisregels zien er uit als blauwe (of lichtblauwe) afgeronde rechthoeken. De beslisregels bestaan steeds uit één criterium, dat 'waar' of 'onwaar' is. Aan een beslisregel-rechthoek zitten altijd twee pijlen naar beneden. Personen waarvoor het criterium 'waar' is, volgen de pijl naar rechtsonder. Personen waarvoor het criterium niet geldt, volgen de pijl naar linksonder. Dit gaat zo door totdat je een rode, oranje of groene rechthoek bereikt. Dit zijn de uitkomsten. Elke uitkomst-rechthoek staat voor een groep personen. Voor de groene "Ja"-uitkomsten geldt dat meer dan de helft van de personen die hier uitkomen de komende 12 maanden instroomt. Voor de rode en oranje "Nee"-uitkomsten geldt dat de meerderheid juist *niet* instroomt.

Met deze boom kun je dus testen of een persoon of groep het komende jaar waarschijnlijk wel of niet instroomt in de bijstand. Dit doe je door van boven naar onderen door de boom te 'lopen'. Je begint bij de bovenste knoop, ofwel de wortel. De wortel van de classificatieboom in dit onderzoek heeft als criterium "Laatste uitstroom \leq 12 maanden". Als het waar is volg je de pijl naar rechts, anders links. Zo ga je via de pijlen van criterium naar criterium totdat je bij een blad uitkomt. Is het blad groen? Dan is de kans groot (ofwel groter dan 50 procent) dat deze persoon het komende jaar instroomt. Is het blad rood of oranje? Dan is de kans groter dat deze persoon juist niet instroomt.

Eigenschappen van de boom

De boom is zo samengesteld dat hoe hoger een criterium in de boom staat, des te belangrijker bijbehorende variabele is om instromers van niet-instromers te onderscheiden. Het bovenste criterium gaat over hoe lang geleden iemand is uitgestroomd is. Duur na laatste uitstroom is dus zeer relevant voor instroom. De herkomst van een persoon komt alleen helemaal onderaan de boom voor in een criterium. Dit betekent dat herkomst nauwelijks relevant is voor het bepalen of iemand gaat instromen of niet. Van de 33 variabelen waar de boom uit kon kiezen, zijn er uiteindelijk 18 helemaal niet gebruikt. Deze variabelen, zoals geslacht of burgerlijke staat, blijken nog minder relevant voor uitstroom.

Verder zijn de criteria steeds zó geformuleerd, dat de kans op instroom groter is wanneer aan een criterium voldaan is. De personen die niet aan dat criterium voldoen hebben een kleinere instroomkans en volg je naar links. Degenen die juist wel aan dat criterium voldoen hebben een grotere instroomkans en volg je naar rechts. Zodoende zitten in de rechter zijtakken relatief de meeste instromers.

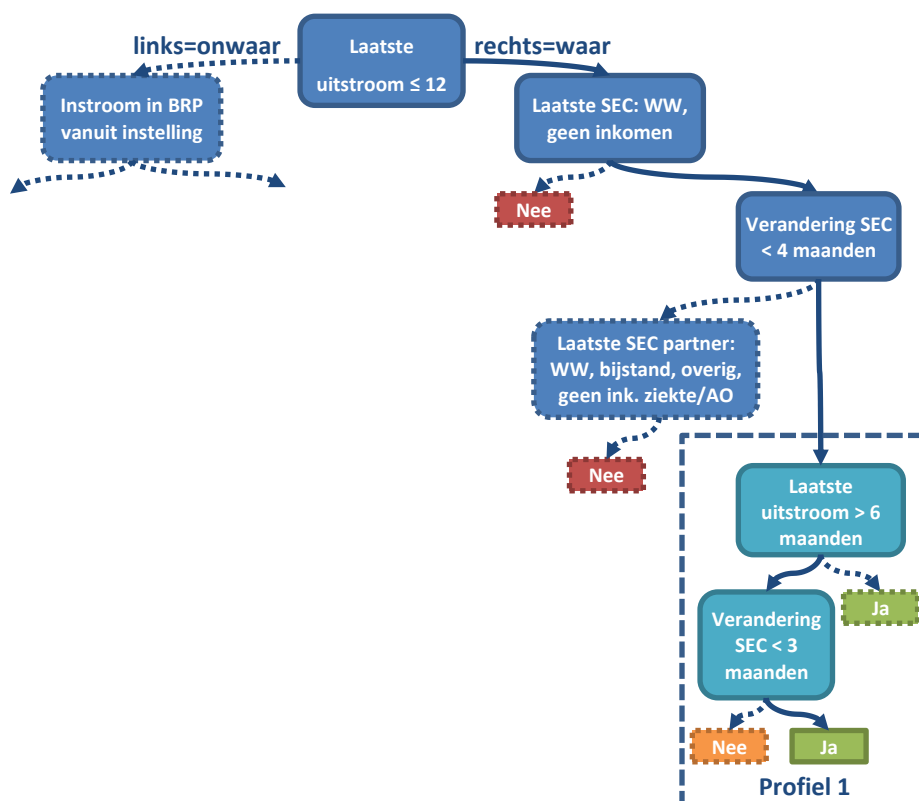
⁹ De validiteit van de boom is gecontroleerd en goed bevonden. Zie voor een toelichting Bijlage 5.2.



Figuur 2 Classificatieboom voor instroom in bijstand

Voorbeeld: uitstroomer met kort een baan

In Figuur 2 worden 5 profielen onderscheiden. Als voorbeeld toont Figuur 3 het pad van personen die horen bij profiel 1. Denk bijvoorbeeld aan iemand die in de bijstand zat en 3 maanden geleden een baan kreeg voor 1 maand, waarna hij of zij zonder inkomen kwam te zitten. Gaat deze persoon instromen? Om deze vraag te beantwoorden begin je bij de bovenste knoop. Deze heeft als criterium "laatste uitstroom \leq 12 maanden", wat staat voor "de laatste uitstroom uit de bijstand van deze persoon was 12 maanden of minder geleden". Aangezien de persoon 3 maanden geleden uitstroomde is dit waar. Je volgt dus de rechter pijl naar "Laatste SEC: WW, geen inkomen", wat betekent dat de laatste 'sociaal economische status' van deze persoon 'WW', of 'geen inkomen' was. De persoon had geen inkomen, dus ook dit is waar. Volg wederom de rechter pijl, naar "Verandering SEC < 4 maanden". Omdat de persoon 2 maanden geleden van baan naar geen inkomen was gegaan, is ook dit waar en volgen we wederom de meest rechterpijl, die in dit geval recht naar onderen gaat, naar "Laatste uitstroom > 6 maanden". Dit laatste criterium is lichtblauw in plaats van donkerblauw en bevindt zich binnen de stippellijnen van 'Profiel 1'. Deze persoon past dus binnen het eerste profiel. Je kan eventueel nog verder lopen. Aangezien de laatste uitstroom 3 maanden geleden was, is het criterium "Laatste uitstroom > 6 maanden" niet waar, en loop je naar links, naar "Verandering SEC < 3 maanden". De laatste verandering was maar 2 maanden geleden, dus dit is waar. De pijl naar rechts komt uit bij een "Ja" knoop. Deze persoon zit dus niet alleen in profiel 1, daarnaast heeft hij of zij ook een kans van meer dan 50 procent om daadwerkelijk in te stromen.



Figuur 3 Voorbeeld van een pad in de boom naar 'Profiel 1'

4.2 Vijf profielen

In principe zou je elk blad met "Ja", en dus vooral instromers, in de boom van Figuur 2 kunnen zien als één profiel. Vaak zitten er echter twee of meer bladeren vlak bij elkaar in de boom, waardoor het de vraag is of het zinvol is om deze als afzonderlijke profielen te zien. Bovendien bevatten sommige bladeren te weinig personen om ook voor Eindhoven tot zinvolle gegevens te komen. Om deze reden

hebben we steeds groepen van aangrenzende knopen en bladeren samengenomen tot één profiel. Uiteindelijk resulteerde deze aanpak in vijf verschillende profielen.

Deze profielen bevatten samen 18,5 procent van alle instromers. Blijkbaar zijn de andere instromers onvoldoende verschillend van de niet-instromers. Of preciezer geformuleerd: de overige 81,5 procent van de instromers verschilt op grond van de gebruikte variabelen onvoldoende van de niet-instromers in de populatie om te kunnen onderscheiden aan de hand van een profiel van voldoende omvang. Er zijn twee manieren om tot meer profielen en meer instromers te komen, maar die hebben allebei belangrijke nadelen:

- 1 Je kan kleine bladeren toestaan, van minder dan 300 personen. Je loopt dan echter wel een groot risico op 'overfitting', ofwel dat je een profiel maakt voor mensen die in 2015 toevallig instroomden, maar normaal gesproken niet vaker dan gemiddeld in de bijstand komen. Bovendien zijn de profielen dan te klein om voor Eindhoven te kunnen publiceren.
- 2 Je kan een profiel maken met relatief weinig instromers, door een deel van de boom te nemen met enkel "Nee"-bladeren. Nu zijn de profielen zo gekozen dat meer dan een derde van de personen die daarin zitten ook echt instromen, maar je zou de lat lager kunnen leggen. Profielen waarbij slechts een kleine minderheid instroomt de bijstand geven echter nauwelijks richting of focus voor beleid.

De uiteindelijke vijf profielen kan je als volgt kort samenvatten. Profielen 1 en 2 bevatten 'draaideur' bijstandontvangers, profiel 3 vooral vergunninghouders, en profielen 4 en 5 personen die instromen vanuit werk of WW. Hieronder volgt een nadere omschrijving van elk van de profielen. Daarbij geven we niet alleen aan wat de voorwaarden zijn om in een bepaald profiel terecht te komen, maar ook het aantal personen en het aandeel instromers per profiel in 2015 en de verdeling hiervan over de Eindhovense werkgebieden. De aantallen per profiel in Eindhoven en haar werkgebieden zijn overigens ook terug te vinden in Tabel 2 en

Tabel 3 van de bijlage. De achtergrondkenmerken per profiel zijn opgenomen in Figuur 11 tot en met 15 van de bijlage.

4.3 Profiel 1

Dit zijn personen die maximaal een jaar geleden zijn uitgestroomd uit de bijstand. Daarnaast hebben ze recentelijk te maken gehad met een overgang naar WW of een situatie waarin zij geen inkomen ontvingen. Je kan hierbij denken aan iemand die 5 maanden geleden uit de bijstand is gestroomd omdat ze een baan vond, maar die baan na 3 maanden weer verloor. Vanwege haar arbeidsverleden heeft ze geen recht op WW. Na twee maanden zonder inkomen stroomt geraakt ze weer in de bijstand.

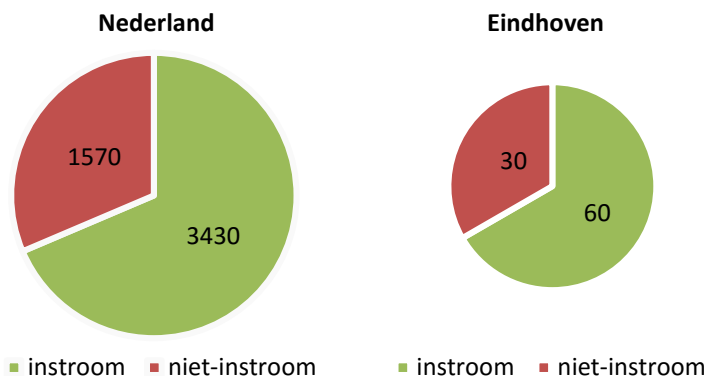
Wat hun achtergrondkenmerken betreft zijn personen in profiel 1 'doorsnee' instromers en lijken ze erg op het gemiddelde van de vijf profielen: iets meer dan de helft is man, iets minder dan de helft heeft een Nederlandse achtergrond, de leeftijdsgroep 28 tot 37 komt het vaakst voor en 40 procent woont alleen. Tot slot hoorde twee derde bij de armste 20 procent van de onderzoekspopulatie.

Profiel 1

- Uitstroom ≤ 12 maanden voor instroom
- Laatste SEC: WW, geen inkomen
- Verandering SEC < 4 maanden voor instroom

Figuur 4 Karakterisering van de personen in profiel 1

In heel Nederland vallen er in 2015 ongeveer 5000 personen in profiel 1, waarvan 3420 (ofwel ruim twee derde) instromers. Hiermee is profiel niet bijzonder groot, maar bevat het wel relatief veel instromers. In Eindhoven wonen ongeveer 90 personen met dit profiel.



Figuur 5 Aantal personen in profiel 1, naar instroom

De Eindhovenaren in profiel 1 zijn evenredig verdeeld over de werkgebieden, dus iets meer in grote werkgebieden Stratum en Woensel Zuidwest en iets minder in het kleinere Meerhoven of Woensel Zuidoost. Overigens heeft Meerhoven ook relatief weinig instromers, namelijk slecht 0,4 procent van de bevolking. Het instroompercentage is met 1,3 procent het hoogste in de werkgebieden Woensel Noord Ontginning, Woensel Noordoost en Woensel Zuidwest.

4.4 Profiel 2

Ook dit zijn personen die maximaal een jaar geleden zijn uitgestroomd uit de bijstand. Wel hebben zij al langer te maken met een situatie waarin zij WW ontvangen of geen inkomen hebben dan de personen in profiel 1. Daarnaast hebben zij, als zij al een partner hebben, partners met een zwakke economische positie: partners die geen werk hebben, geen pensioen ontvangen en geen uitkering ontvangen voor arbeidsongeschiktheid. Je kan hier denken aan iemand waarvan de partner een bijstandsuitkering ontvangt, maar zelf 11 maanden geleden een baan vond en daardoor uitstroomde uit de bijstand. Die baan is na een maand weer gestopt en sindsdien zit deze persoon alweer 10 maanden zonder inkomen. Terugkeer naar de bijstand volgt.

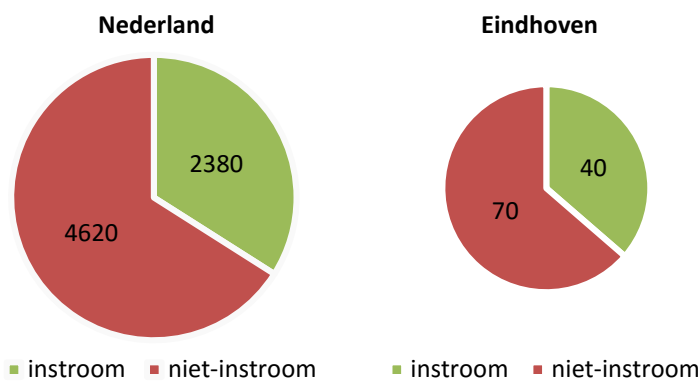
Ook voor personen in profiel 2 geldt dat de achtergrondkenmerken erg doorsnee zijn voor een instromer; deze zijn vrijwel hetzelfde als voor personen in profiel 1, behalve dat ze nog minder vaak WW ontvangen (15 procent) voordat ze instroomden.

Profiel 2

- Uitstroom \leq 12 maanden voor instroom
- Laatste SEC: WW, geen inkomen
- Verandering SEC $>$ 3 maanden voor instroom
- Laatste SEC partner: *geen* werk/pensioen en geen ziekte/AO

Figuur 6 Karakterisering van de personen in profiel 2

Ongeveer 7000 Nederlanders vallen in 2015 onder profiel 2, waarvan iets meer dan een derde instroomt. Door het relatief lage instroompercentage, is dit profiel wat aantallen instromers betreft het op één na kleinste. In Eindhoven gaat het om ongeveer 110 personen, die wederom evenredig verdeeld zijn over de werkgebieden.



Figuur 7 Aantal personen in profiel 2, naar instroom

4.5 Profiel 3

Het derde profiel bevat vooral statushouders zonder inkomen. Het gaat hier voornamelijk om personen die recentelijk in een institutioneel huishouden zaten en nieuw zijn ingeschreven in de BRP. Deze combinatie komt vooral voor bij statushouders. Personen in asielzoekerscentra worden namelijk beschouwd als personen die in een institutioneel huishouden zitten. Daarnaast worden personen die nieuw in Nederland komen wonen (zoals asielzoekers), nieuw in de BRP ingeschreven. Pas als asielzoekers een verblijfsvergunning krijgen, worden zij statushouder en kunnen zij een uitkering krijgen.

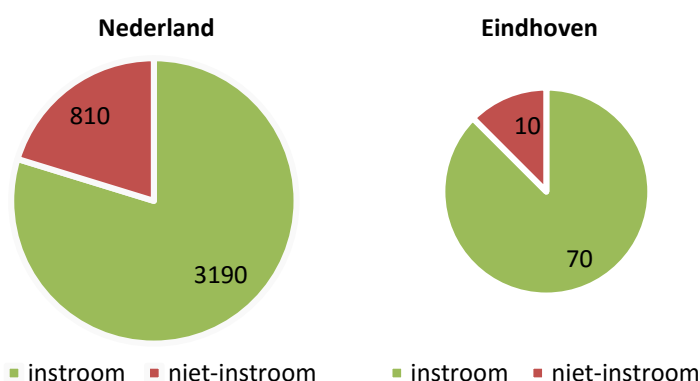
Ook aan de achtergrondkenmerken van personen in dit profiel is duidelijk dat het hier vooral om statushouders gaat. Zo zijn vijf van de zes personen in dit profiel afkomstig uit een niet-westers land, anders dan Marokko, Turkije, Suriname of de Antillen. Van de instromers is dit zelfs 95 procent. Verder is driekwart van de personen in dit profiel een man en is eveneens driekwart tussen de 18 en 37 jaar oud. Dit komt inderdaad overeen met statushouders, waarvan bekend is dat deze vaker man en jong zijn. Bovendien ontving niemand in dit profiel een inkomen voordat ze instroomden.

Profiel 3

- Uitstroom > 12 maanden voor instroom of nooit
- Instroom in BRP vanuit een instelling
- Laatste SEC: geen inkomen

Figuur 8 Karakterisering van de personen in profiel 3

Profiel 3 is met ongeveer 4000 personen in 2015 in Nederland en 80 in Eindhoven één van de twee kleinste profielen. Het is echter wel het profiel met het grootste aandeel instromers, namelijk drie kwart in Nederland en zelfs meer in Eindhoven. Van de mensen die voldoen aan dit profiel, is de kans het grootste dat deze ook inderdaad gaan instromen. Overigens woonde de overgrote meerderheid van de personen in dit profiel in 2015 in het werkgebied Woensel Acht & Aanschot, waar tot voor kort het AZC gevestigd was.



4.6 Profiel 4

Nu blijven de mensen over die niet recent zijn uitgestroomd uit de bijstand en ook niet zijn ingestroomd in de BRP vanuit een instelling. Deze vinden we in profielen 4 (Figuur 9) en 5 (Figuur 10). Voor beide profielen geldt dat personen geen inkomen hadden of WW of bijstand¹⁰ ontvingen, en hun partner (als ze die al hebben) geen uitkering ontvangt voor ziekte of arbeidsongeschiktheid. Bovendien hebben personen in beide profielen een laag vermogen en wonen ze in een sociale huurwoning.

Het verschil tussen profielen 4 en 5 zit hem er vooral in dat personen in profiel 4 recent veranderd zijn van sociaal economische status (SEC). Vóór deze verandering had meer dan de helft van de instromers in profiel 4 nog inkomen uit werk, ontving 10 procent een uitkering wegens arbeidsongeschiktheid of ziekte en volgde bijna 20 procent een opleiding. Na de verandering ontving iets minder dan de helft een WW-uitkering, terwijl een kleine meerderheid voordat ze de bijstand ingingen geen inkomen had. Bij deze laatste groep kan je denken aan werkloze ex-scholieren, personen zonder recht op WW zoals zelfstandigen of mensen die zelf hun baan hebben opgezegd. Een voorbeeld van iemand die binnen profiel 4 zou vallen is iemand die twee jaar gewerkt heeft en daarna 3 maanden WW ontving. Als de WW afloopt, heeft de persoon recht op bijstand

¹⁰ Wij onderzoeken of personen instromen in de algemene bijstand volgens de Participatiewet. Er zijn echter vormen van bijstand, zoals bijstand elders verzorgenden en bijstand voor adreslozen, die wel onder de SEC-categorie bijstand vallen, maar niet als instroom tellen. Als personen vanuit zo'n andere bijstandsvorm instromen in de algemene bijstand, beschouwen we ze als instromer. Zie ook de laatste paragraaf van bijlage 7.3.

omdat zijn of haar partner geen inkomen heeft, ze geen eigen woning hebben en samen nauwelijks spaargeld hebben.

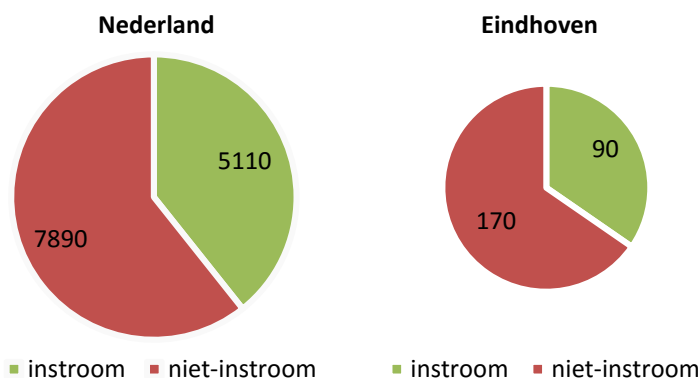
Profiel 4 bestaat voor 40 procent uit jongeren (18-27 jaar), waarmee dit een relatief jong profiel is. Verder wonen relatief veel personen (20 procent) in dit profiel bij een van hun ouders, al is nog steeds de meerderheid (40 procent) alleenstaand.

Profiel 4

- Uitstroom > 12 maanden voor instroom of nooit
- *Niet* instroom in BRP vanuit een instelling
- Laatste SEC: WW, bijstand¹⁰, geen inkomen
- Laatste SEC partner: *niet* ziekte/AO
- Verandering SEC ≤ 3 maanden voor instroom
- Vermogen < 6 400 euro
- Woningeigendom: sociale woning
- Inkomen huishouden < 29 700 euro

Figuur 9 Karakterisering van de personen in profiel 4

Met 13 000 personen in 2015 waarvan ruim 5 duizend instromers is profiel 4 verreweg het grootste profiel, zowel in personen als instromers. Dat geldt ook voor Eindhoven, waar maar liefst 280 personen binnen dit profiel vallen waaronder een derde instromers. Ook dit profiel is vrij evenredig verdeeld over de Eindhovense werkgebieden. Opvallend is echter wel dat het relatief kleine Woensel Noord Ontginning maar liefst 40 personen kent in dit profiel, bijna evenveel als in het veel grotere Tongelre Centrum en Woensel Zuidwest.



4.7 Profiel 5

Net als profiel 4, bevat profiel 5 personen die niet recent (of nooit) zijn uitgestroomd uit de bijstand of recent zijn ingestroomd in de GBA, maar wel een laag vermogen hebben, in een sociale huurwoning wonen en geen partner hebben met een uitkering voor ziekte of arbeidsongeschiktheid. Een belangrijk onderscheid is echter dat profiel 5 mensen beschrijft waarbij de laatste verandering van de SEC langer geleden was en dat ze sindsdien WW ontvingen. Bovendien zijn personen in profiel 5 jonger dan 52 jaar, staan ze ingeschreven bij het CWI (tegenwoordig het UWV werkbedrijf) en ontving hun huishouden in 2014 zeer weinig inkomen uit werk (minder dan 7 duizend euro), waarvan zij zelf minder dan 3 duizend euro. Denk hierbij aan iemand die een baan had en daarna voor langere tijd WW ontving. De persoon probeert vanuit de WW een baan te vinden en schrijft zich

ook in bij het CWI (tegenwoordig UWV werkbedrijf), maar helaas lukt dat niet. Als de WW afloopt, heeft de persoon recht op een uitkering omdat hij of zij alleen woont en nauwelijks vermogen bezit.

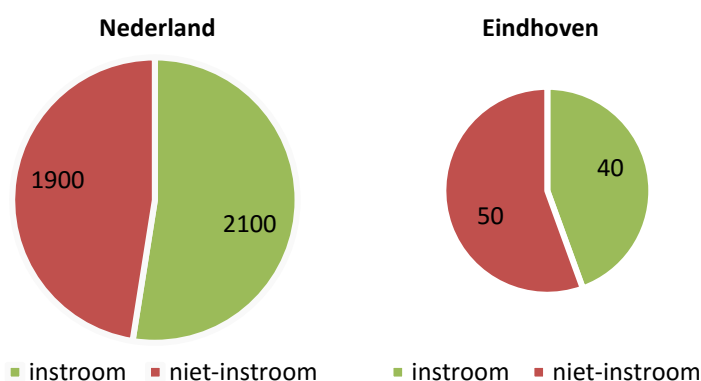
Profiel 5 is het enige instroomprofiel dat voor meer dan de helft (namelijk 60 procent) bestaat uit mensen met een Nederlandse achtergrond. Dat is echter nog steeds minder dan de personen die in geen enkel profiel vallen, want daarvan heeft bijna 80 procent een Nederlandse achtergrond. Verder valt op dat dit profiel vrijwel geen jongeren bevat, maar wel voor meer dan de helft bestaat uit personen tussen de 37 en 48 jaar. Ondanks dat personen in dit profiel WW ontvingen, behoorde meer dan 80 procent in dit profiel tot de armste huishoudens.

Profiel 5

- Uitstroom > 12 maanden voor instroom of nooit
- *Niet* instroom in BRP vanuit een instelling
- Laatste SEC: WW
- Laatste SEC partner: *niet* ziekte/AO
- Verandering SEC > 3 maanden voor instroom
- Woningeigendom: sociale woning, onbekend
- Ingeschreven bij CWI
- Leeftijd < 52
- Inkomen huishouden < 7 000 euro
- Vermogen < 4 700 euro
- Inkomen < 2 700 euro

Figuur 10 Karakterisering van de personen in profiel 5

Profiel 5 is met 4000 personen in 2015 samen met profiel 3 het kleinste. Ook het aantal personen dat instroomt, is met 2100 personen, het kleinst. Ook voor Eindhoven is dit één van de kleinere profielen. Zowel profiel 5 als profiel 2 hebben voor Eindhoven slechts rond de 40 instromers.



5. Conclusie

Om gemeenten te voorzien in hun behoefte aan informatie over het gebruik van verschillende voorzieningen in het sociaal domein, heeft het CBS in opdracht van VNG Realisatie een onderzoeksbestand samengesteld¹¹. Door dit bestand uit te breiden met personen die geen gebruik maken van regelingen in het sociaal domein, kan deze ook gebruikt worden om met een machine learning algoritme te voorspellen wat voor soort personen een grote kans hebben om in of uit te stromen in het sociaal domein. Dit biedt mogelijkheden voor data gestuurd preventief en activerend beleid. Deze kans is opgepakt in dit onderzoek van het Urban Data Center Eindhoven van het CBS.

In dit onderzoek is als testcase gekeken naar de instromers in de bijstand: Wat zijn hun profielen en hoe vaak komen die in Eindhoven voor? Aan de hand van een classificatieboom zijn vijf profielen onderscheiden waarmee je bijna 1 op de 5 personen die het komende jaar gaan instromen in de bijstand kan identificeren. In Eindhoven ligt dit percentage iets hoger. Profielen 1 en 2 bevatten 'draaideur' bijstandontvangers die kort geleden ook al in de bijstand zaten. Profiel 3 bevat vooral vergunninghouders en profielen 4 en 5 personen die instromen vanuit werk of WW. De profielen zijn uitgebreid toegelicht in hoofdstuk 3.

De profielen laten zien voor welke groepen proactief arbeidsmarktbeleid waardevol lijkt. Zo lijkt het zinvol om beleid te voeren op personen die recentelijk uit de bijstand zijn uitgestroomd (profielen 1 en 2) en vergunninghouders (profiel 3). Dit bevestigt echter wat de gemeente Eindhoven al wist. Dat geeft weliswaar vertrouwen in de methode, maar levert helaas geen nieuwe inzichten op. Voor profielen 4 en 5 geldt dat de gemeente hier lastig zelf direct beleid op kan voeren omdat deze niet weet wie er WW ontvangt.

Daarnaast demonstreert dit onderzoek de mogelijkheden en beperkingen van classificatiebomen, waarover meer in het hierop volgende reflectiehoofdstuk. Door de lessen uit dit onderzoek te gebruiken, kunnen gemeentes in de toekomst met classificatiebomen bruikbaarere profielen maken van in- en uitstromers uit andere regelingen en voorzieningen binnen het sociaal domein. Dit kan gemeentes helpen bij de verdere ontwikkeling en uitvoering van meer data gedreven beleid.

¹¹ Een toelichting op dit project staat op de CBS website: <https://www.cbs.nl/nl-nl/maatwerk/2018/13/longitudinale-analyse-sociaal-domein-2015>

6. Reflectie

Dit onderzoek heeft de inzichten in het beleidsveld verruimd. De uiteindelijke bruikbaarheid van de resultaten is echter nog wat beperkt. Zo valt enkel een minderheid van de instromers binnen een profiel, lijken de gevonden profielen achteraf gezien voor de hand liggend en beschikt de gemeente in veel gevallen niet over de informatie die nodig is om te bepalen tot welk profiel iemand hoort. In deze reflectie beschrijven we allereerst waarom het niet eenvoudig is om het onderzoeksresultaat te verbeteren. Daarna volgt een lijst van aanbevelingen, waarmee een toekomstig onderzoek bruikbaarere resultaten op kan leveren.

6.1 Technische keuzes en beperkingen

In dit onderzoek hebben we profielen gemaakt waarmee je 18,5 procent van de totale instroom (en 22,6 van de Eindhovense instroom) met redelijke zekerheid (bijna de helft stroomt in) kan voorspellen. Dat lijkt misschien weinig. Echter, onder de hele onderzoekspopulatie was de instroom in de bijstand in 2015 minder dan 0,9%, terwijl de instroomkans in de vijf profielen samen bijna de helft is. Dat is ruim vijftig keer zo veel en, vanuit statistisch oogpunt gezien, een hele prestatie. Ook is de uitkomst verklaarbaar doordat instroom in de bijstand niet volledig wordt bepaald door iemand zijn achtergrondkenmerken, maar bijvoorbeeld ook zijn karakter. Wellicht is het mogelijk om met andere keuzes een paar procentpunt meer instromers te kunnen 'vangen' in instroomprofielen, maar veel meer dan het huidige resultaat is op basis van administratieve data hoogst onwaarschijnlijk. Hieronder een aantal mogelijkheden die in toekomstig onderzoek kunnen worden overwogen:

- 1 Selectie van profielen. Je kan uit dezelfde boom 'grotere' profielen halen, zodat er meer instromers in een profiel vallen. Dat gaat echter ten koste van het percentage instromers binnen de profielen. Ofwel: meer kwantiteit gaat ten koste van de kwaliteit. Doordat het aantal personen dat in de classificatieboom eindigt bij een 'nee'-knoop veel groter is dan bij 'ja'-knopen, levert dit hoogstens een paar procentpunt meer instromers op, in ruil voor een aanzienlijk kleinere instroomkans.
- 2 Variabelen. Dit onderzoek maakte gebruik van de variabelen in het bestand voor VNG Realisatie, omdat deze deels toegespitst zijn op het sociaal domein. Dit bleken echter niet de beste variabelen om instroom te voorspellen. Tijdens het onderzoek bleek dat vooral eerdere uitstroom en verandering in sociaal economische status relevant waren, waardoor deze variabelen tijdens het onderzoek zijn toegevoegd. Ook in toekomstig onderzoek is het belangrijk om kritisch te zijn op de geschiktheid van variabelen.
- 3 Parameterinstellingen. Voor dit onderzoek hebben we een ondergrens genomen van 300 personen per knoop. Als je kleinere knopen toestaat, kan dit aanvullende kleine profielen opleveren, waardoor er iets meer instromers binnen een profiel terechtkomen. Echter voor Eindhoven zou zo'n klein profiel dan waarschijnlijk uit één of slechts enkele instromers bestaan. Dat zijn geen aantallen waar de gemeente beleid op kan voeren. Daarnaast geven kleinere knopen statistisch onbetrouwbare resultaten en is het risico bijvoorbeeld groot dat het instroompercentage in andere jaren veel lager is.¹² Tot slot is bij kleinere aantallen een regionale opsplitsing niet mogelijk, terwijl daar in dit onderzoek wel om gevraagd was door de gemeente.
- 4 Meer data. Door meerdere jaren aan instroom te combineren, wordt de groep instromers groter en ontstaan er mogelijk iets gedetailleerdere profielen. Er zitten echter ook nadelen

¹² Knopen van minder dan 300 personen kunnen namelijk minder dan 15 niet-instromers bevatten, omdat we bij de niet-instromers vanwege de grootte van het bestand een steekproef hebben genomen van ongeveer 1 op 20. Hierdoor zou het onderscheid tussen instroom en niet-instroom te veel van toeval afhankelijk zijn, met een groot risico op overfitting,

aan. Allereerst moet je je dan afvragen in hoeverre instroom in eerdere jaren (vóór de Participatiewet) relevant zijn voor de toekomst. Daarnaast moeten er verschillende aanvullende keuzes gemaakt worden, bijvoorbeeld of je instroom binnen één jaar of binnen vijf jaar wilt voorspellen, welke data je kiest als peilmoment voor welke instromer, wat te doen met herhaalde instroom en hoe de controlegroep samen te stellen.

6.2 Aanbevelingen

Dit onderzoek was een experiment en dus ook deels bedoeld om van te leren. Vandaar dat we naast de uitkomsten van het onderzoek ook een aantal aanbevelingen delen. Hopelijk zullen daardoor resultaten van toekomstige soortgelijke onderzoeken beter bruikbaar zijn in de praktijk:

- 1 Interactief proces. In dit onderzoek is een groot aantal keuzes gemaakt, over: populatie, definitie van instroom, gebruikte variabelen en parameters van het algoritme. Nu lag een deel van deze keuzes vooraf vast, bijvoorbeeld dat het VNG Realisatiebestand uitgangspunt was, dat het ging om uitstroom in 2015, of dat er regionaal uitgesplitst moet worden. Achteraf bleken sommige van deze keuzes niet helemaal aan te sluiten bij de wensen van gemeente Eindhoven. Dit valt te ondervangen door vooraf minder vast te leggen en de gemeente tijdens het proces steeds keuzes voor te leggen, met daarbij een onderbouwing van de voor- en nadelen. Dat is wel arbeidsintensiever en kost meer tijd, omdat er dan meer overleg nodig is, dat bovendien goed voorbereid moet worden.
- 2 Variabelen van de gemeente. Eén van de nadelen van de huidige profielen is dat een aantal sleutelvariabelen, zoals sociaal economische klasse, niet beschikbaar zijn bij gemeentes. De gemeente heeft daardoor weinig handvatten om met de profielen aan de gang te gaan. Als gemeentes zelf variabelen aandragen, bijvoorbeeld het gebruik van sociale regelingen waarover zij zelf informatie in huis hebben, dan kan ze ook makkelijker mensen benaderen aan de hand van profielen.
- 3 Inclusief proces. Bij de keuze van variabelen, definities en instellingen is het van meerwaarde ervaringsdeskundigen uit de praktijk te betrekken. Met name gemeenteambtenaren 'uit het veld' (consulenten, case managers, specialisten inkomen) of medewerkers van het UWV hebben waarschijnlijk kennis die hierbij nuttig is. Bovendien kunnen zij in een vroeg stadium terugkoppeling geven over de praktische toepassingen van het resultaat. Ook hierbij geldt overigens dat dit arbeidsintensief is en tijd kost.
- 4 Type regeling. De algemene bijstand is weliswaar de grootste regeling die onder de Participatiewet valt, maar de groep van instromers is ook relatief diffuus. Daardoor is het wellicht lastiger om instromers in profielen samen te vatten. Voor andere regelingen met een meer homogene groep van instromers is het wellicht eenvoudiger om via een algoritme voor een groter aandeel instromers profielen vast te stellen.
- 5 Verwachtingen. Stem aan het begin van het onderzoek goed de verwachtingen af. De onderzoeksresultaten lijken achteraf vaak erg voor de hand liggend, omdat je er makkelijk een verhaal bij kan bedenken. Dat wil echter niet zeggen dat je alle resultaten ook daadwerkelijk vooraf had kunnen voorspellen. Door de verwachtingen vooraf expliciet te maken (bijvoorbeeld: welke profielen verwacht je), wordt beter duidelijk welke uitkomsten wel of juist niet nieuwe informatie opleveren. Bovendien zijn expliciete verwachtingen een waardevol middel om de plausibiliteit van de resultaten te toetsen.

7. Bijlagen

7.1 Tabellen en grafieken

Tabellen met aantallen per profiel

De onderstaande tabellen geven de aantallen personen en instromers die horen bij elk van de vijf profielen die in sectie 4.2 zijn beschreven. Deze aantallen geven we voor zowel Eindhoven als heel Nederland weer. De aantallen niet-instromers voor heel Nederland zijn echter minder nauwkeurig, omdat deze op basis van een steekproef zijn berekend. De aantallen voor Nederland en Eindhoven staan in Tabel 2, terwijl Tabel 3 deze laatste onderverdeelt naar Eindhovense werkgebieden. Deze aantallen zijn gebruikt voor taartdiagrammen met aantallen per profiel, in alinea 4.2.

Tabel 2 Aantal personen en instromers per profiel, Nederland en Eindhoven, 2015 (februari-december)

	Nederland			Eindhoven			
	totaal personen ¹³	instroom	van alle instroom ¹⁴	totaal personen	instroom	van alle instroom ¹⁴	van profiel ¹⁴
	<i>aantal</i>	<i>aantal</i>	%	<i>aantal</i>	<i>aantal</i>	%	%
Profiel 1	5 000	3 430	3,9	90	60	4,5	67
Profiel 2	7 000	2 380	2,7	110	40	2,6	32
Profiel 3	4 000	3 190	3,6	80	70	5,4	86
Profiel 4	13 000	5 110	5,8	260	90	6,9	35
Profiel 5	4 000	2 100	2,4	90	40	3,3	47
Alle profielen	33 000	16 210	18,5	640	300	22,6	47
Totale populatie	10 079 000	87 480	100,0	139 980	1 340	100,0	1

Tabel 3 Aantal personen per profiel en instromers, per werkgebied in Eindhoven, 2015 (februari-december)

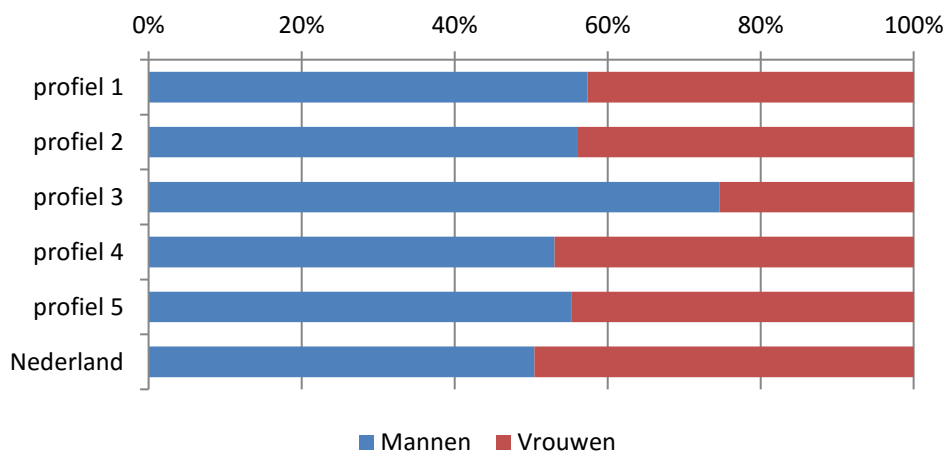
	personen					instroom		
	totaal	Profiel 1	Profiel 2	Profiel 3	Profiel 4	Profiel 5	totaal	% alle personen ¹⁴
	<i>aantal</i>						<i>aantal</i>	%
Eindhoven	139 980	90	110	80	260	90	1 340	1,0
Gestel	16 910	10	20	0	30	10	190	1,1
Meerhoven	6 400	0	0	0	10	0	20	0,4
Stratum	21 150	20	20	0	30	10	160	0,8
Striip oost	15 020	10	10	0	30	10	120	0,8
Tongelre Centrum	18 880	10	10	0	40	10	180	1,0
Woensel Acht & Aanscot	21 650	10	10	80	20	10	160	0,7
Woensel Noord								
Ontginning	8 620	10	10	0	40	10	110	1,3
Woensel Noordoost	7 320	10	10	0	20	10	100	1,3
Woensel Zuidoost	6 670	0	10	0	10	0	60	0,9
Woensel Zuidwest	17 360	20	20	0	50	10	240	1,4

¹³ Afgerond op duizendtallen, omdat het gaat om een schatting op basis van een steekproef. In sectie 7.2 wordt uitgelegd waarom er met een steekproef is gewerkt.

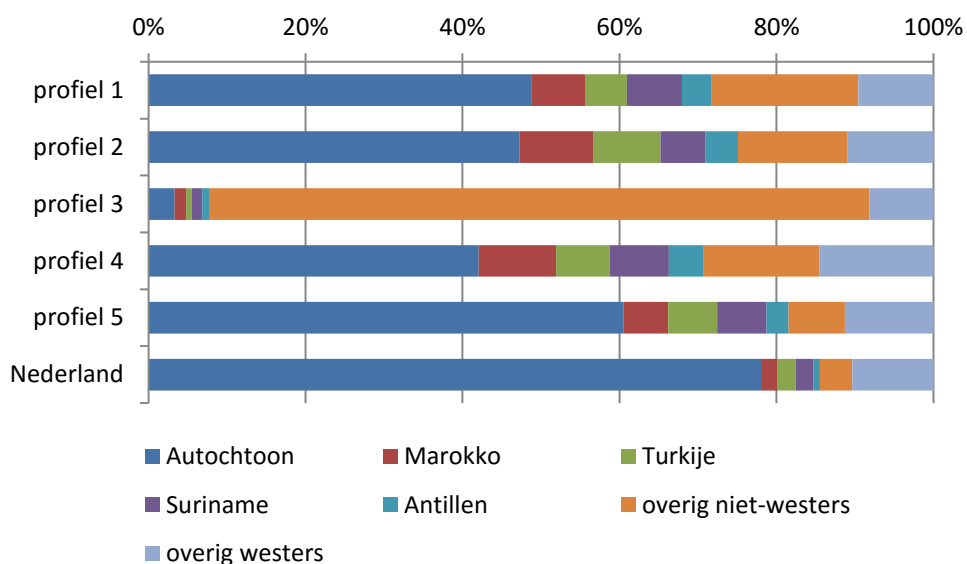
¹⁴ De percentages zijn berekend op basis van de niet-afgeronde cijfers.

Grafieken met achtergrondkenmerken per profiel

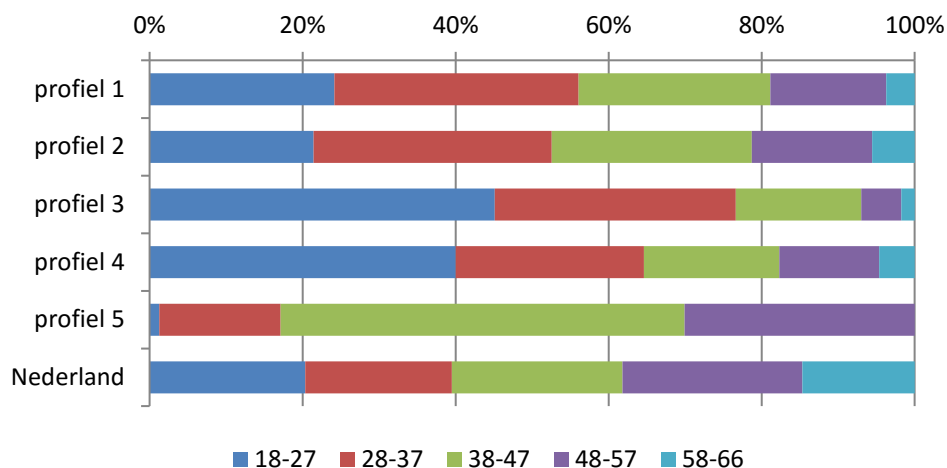
De onderstaande staafdiagrammen geven voor alle vijf de profielen weer hoe de samenstelling van de groep is wat betreft geslacht, herkomst, leeftijd, huishoudsituatie en besteedbaar huishoudinkomen. Deze zijn gebruikt voor de beschrijvingen de personen per profiel, in alinea 4.2.



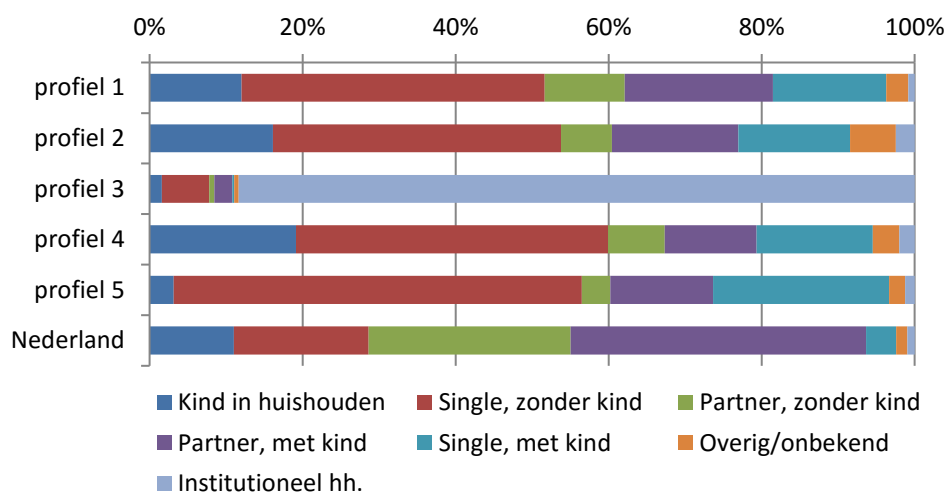
Figuur 11 Geslacht per profiel en voor de hele populatie, 2015



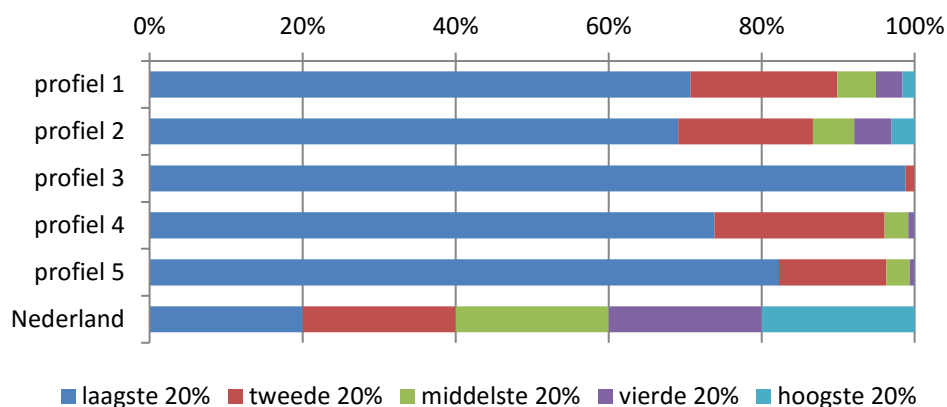
Figuur 12 Herkomst per profiel en voor de hele populatie, 2015



Figuur 13 Leeftijd per profiel en voor de hele populatie, 2015



Figuur 14 Laatste huishoudsituatie voor instroom, per profiel en voor de hele populatie, 2015



Figuur 15 Besteedbaar huishoudinkomen in 5 groepen, per profiel en voor de hele populatie, 2015

7.2 Achtergrond bij het classificatieboom algoritme

Keuzes en instellingen

Bij het maken van deze boom hebben we niet de hele populatie van Nederland gebruikt. Dit zou namelijk een bestand zijn van ruim 10 miljoen personen met elk 120 variabelen, wat te veel rekenkracht kost voor het algoritme dat een classificatieboom berekend. De gebruikte populatie voor de boom bevat alle instromers uit de onderzoekspopulatie¹⁵ en vijf keer zo veel niet-instromers. Bovendien zijn alle inwoners van Eindhoven opgenomen, ongeacht of ze instroomden. Aan het boomalgoritme geven we een extra variabele (de zogenaamde *prior*) mee die corrigeert voor het feit dat we met een steekproef werken en de boom optimaliseert naar homogeniteit in de echte aantallen en dus niet de aantallen in de steekproef. Hierdoor gaat het algoritme door met splitsen van een groep totdat niet enkel in de steekproef, maar ook in de populatie de meerderheid in een knoop uit instromers bestaat.

Daarnaast heeft het algoritme parameters waarmee je eisen kan stellen aan de structuur van de boom. We hebben daarbij aangegeven dat de boom maximaal 30 lagen diep mag zijn, en dat een groep altijd minimaal 300 personen moet bevatten. Bij deze grootte kan je namelijk redelijkerwijs aannemen dat het niet puur toeval is dat mensen met deze eigenschappen instromen en heb je bovendien ook voor de gemeente Eindhoven nog voldoende vulling. Verder hebben we aangegeven hoezeer de homogeniteit moet verbeteren om een splitsing ook daadwerkelijk uit te voeren via de *complexity parameter*. Hierbij hebben we aangegeven dat ook kleine verbeteringen nog meegenomen worden.

Validatie

We hebben de boom uitgebreid getest. Allereerst hebben we vastgesteld dat de boom onafhankelijk was van de gebruikte parameters. Vervolgens hebben we de boom onderworpen aan tienvoudige kruisvalidatie. Dit betekent dat we 10 keer een boom hebben gemaakt op basis van 90% van de dataset en deze 10 bomen hebben vergeleken met de uiteindelijke boom. Dit leverde in alle gevallen erg vergelijkbare bomen op. Bovendien bleek dat de takken die relatief het vaakst voorkwamen in de testbomen ook allemaal in de uiteindelijke boom zitten. Verder hebben we de verhoudingen van de profielen van Nederland en Eindhoven vergeleken en deze bleken zeer goed overeen te komen. Tot slot hebben we de uitkomsten tijdens een expertoverleg aan beleidsdeskundigen van de gemeente Eindhoven voorgelegd. Zij beoordeelden de resultaten als plausibel. Daarom concluderen we dat de boom in Figuur 2 een betrouwbare classificatie maakt van instromers in de bijstand.

¹⁵ Zie sectie 2.3 voor een omschrijving van de onderzoekspopulatie.

7.3 Achtergrond bij operationalisering van de variabelen

Operationalisering life events

Bij het maken van variabelen voor gebeurtenissen hebben we een aantal keuzes gemaakt, die relevant kunnen zijn voor de interpretatie hiervan, namelijk:

- Als een bepaalde soort verandering meerdere keren voorkomt, nemen we alleen de laatste mee in onze analyses. Als bijvoorbeeld iemand twee keer trouwt, dan telt enkel het laatste huwelijk als ‘huwelijk’ life event.
- Voor indicatoren met een life-eventvariabele hebben we de overeenkomstige peilmomentvariabelen buiten het bestand gelaten, om te voorkomen dat verschillende peilmomenten door elkaar heen lopen¹⁶.

Verschillende typen life events

We kijken naar verschillende soorten veranderingen, die op verschillende manieren zijn afgeleid. We onderscheiden drie verschillende typen life events:

- I. *Gebeurtenis met datum*. Bijvoorbeeld een huwelijk. Dit zijn de eenvoudigste life events. Deze variabelen geven het aantal maanden weer tussen instroom en de datum van de gebeurtenis. Als de gebeurtenis niet voorkwam in de 12 maanden voor instroom, dan heeft de life-eventvariabele de waarde ‘99’.
- II. *Entree of exit*. Bij deze life events wordt per maand gekeken of iemand minimaal één dag in een bepaald register zit. Als iemand een bepaalde kalendermaand geen enkele dag in het register zit, maar de maand er na wel, dan telt deze als instromer. Wie andersom in een kalendermaand een of meer dagen in een register zit, maar de maand er na geen enkele dag, telt als uitstromer. Concrete voorbeelden hiervan zijn instroom naar bijstand en uitstroom vanuit bijstand en instroom in de Basisregistratie Personen (BRP). Net als bij type I geven deze variabelen het aantal maanden tussen het life event en instroom weer, en ‘99’ als er geen entree of exit was in de 12 maanden voor instroom.
- III. *Verandering van status*. Bijvoorbeeld een verandering van de plaats in het huishouden, zoals bij een kind dat zelfstandig gaat wonen en daarmee van ‘inwonend kind’ verandert naar ‘alleenstaande’. Deze life events bestaan uit drie variabelen. Allereerst weer een indicator die het aantal maanden tot de laatste verandering weergeeft (met ‘99’ bij geen verandering). Daarnaast zijn er twee variabelen die informatie geven over de aard van de verandering: de eerste geeft aan wat de situatie vóór verandering was en de tweede geeft de situatie na verandering weer.

Fictieve instroommaand voor controlegroep

Om te zorgen dat de life-eventvariabelen ook voor andere analyses gebruikt kunnen worden, en omdat deze deels uit het bovengenoemde bestand voor VNG Realisatie komen, hebben we deze niet direct relatief ten opzichte van de instroommaand berekend. In plaats daarvan hebben we de laatste verandering in 2014 en in 2015 bepaald en aan de hand daarvan de laatste verandering in 12 maanden voor instroom benaderd. Dat werkt als volgt:

1. Als de laatste verandering in 2015 vóór de instroommaand was, dan nemen we deze. Dit is ook echt de laatste verandering voor instroom en geen benadering.
2. Als de laatste verandering in 2015 op of na de instroommaand was, dan nemen we de laatste verandering in 2014. In veel gevallen is dit ook daadwerkelijk de laatste verandering

¹⁶ Een voorbeeld: voor ‘plaats in huishouden’ hebben we variabelen die de plaats in het huishouden weergeven voor en na de laatste (eventuele) verandering van de plaats in het huishouden, maar geen variabele voor de plaats in het huishouden op 1-1-2015.

voor instroom. Alleen als er tussen januari 2015 en de instroommaand nog een andere verandering was, dan missen we deze. In deze gevallen is het life event een benadering.

3. Hierop zijn twee uitzonderingen. Bij de life events 'verandering van sociaal economische categorie' en 'uitstroom' hebben we voor instromers de laatste echte verandering voor instroom berekend, omdat dit de kernvariabelen in de analyse zijn waarvoor we geen benadering willen gebruiken.

Instroom vanuit sociaal economische categorie bijstand?

In de life-eventvariabelen over de sociaal economische categorie (SEC), kan het voorkomen dat 'laatste SEC voor instroom' in de categorie 'bijstand' valt. Er zijn twee redenen waarom dit tóch kan gebeuren:

1. De variabele 'instroom' gaat over personen die instromen in de algemene bijstand. Er zijn echter ook andere vormen van bijstand, zoals bijstand voor elders verzorgenden of bijstand voor adreslozen, die wel onder de SEC-categorie bijstand vallen, maar niet als instroom tellen. Het gaat om ongeveer 10 procent van alle bijstandsuitkeringen in 2015. Voor personen met deze uitkeringen is het dus mogelijk om in te stromen in de bijstand terwijl de laatste SEC voor instroom ook al 'bijstand' was.
2. Instroom in januari 2015 hebben we buiten beschouwing gelaten, zie de tekst over instroom op pagina 6, sectie 3.3. Als iemand in januari instroomde, dan telt deze niet als instromer maar behoort tot de controlegroep. Als een dergelijke januari-instromer bovendien later dat jaar geen andere verandering in SEC had, dan is de laatste SEC voor de fictieve instroommaand 'bijstand'.