

deVerdieping

DATA-ANALYSE Om **afwijkend gedrag** bij burgers te herkennen, koppelt de overheid allerlei gegevens uit haar databases en laat er analyses op los. Wat betekent dat voor de relatie tussen burger en die overheid?

tekst **Kristel van Teeffelen** illustratie **Johan Kleinjan**

Op zoek naar de risicoburger

Een hypothetisch voorbeeld: de 18-jarige Milan woont in de Apeldoornse wijk Osseveld. Hij is gestopt met zijn opleiding. Zonder diploma is het moeilijk een baan vinden. Veel heeft Milan niet te doen, hij hangt met zijn vrienden rond in de buurt, een Vindex-wijk. Klachten van buurtbewoners over geluidshinder, rommel en vernielingen stromen binnen bij gemeente en politie. Die besluit extra rondjes te rijden in de wijk.

Zo is dat altijd gegaan: er is overlast, waarna er wordt ingegrepen. Maar, zo dachten ze in Apeldoorn: wat als je al vóórdat er problemen zijn kunt optreden? De gemeente begint daarom een pilot waarbij ze gegevens analyseert over bijvoorbeeld schooluitval, leefsituatie in een buurt, eerdere overlast en faciliteiten in de wijk. Daar moeten dan buurten in beeld komen met relatief meer kans op problemen. Buurtwerkers kunnen vervolgens jongeren als Milan begeleiden nog voordat zij het verkeerde pad kiezen.

De gedachte achter de pilot in Apeldoorn is dat risico's zoveel mogelijk uitgesloten moeten worden. Voorkomen is beter dan genezen, dat idee is erg populair bij de overheid, zegt Marc Schuilenburg, filosoof en criminoloog aan de Vrije Universiteit in Amsterdam. "Er heerst een bijna kinderlijk enthousiasme om daarvoor data-analyses in te zetten", zegt hij, wijzend naar de uitdijende lijst projecten.

Uitkeringsfraude

De pilot in Apeldoorn staat niet op zichzelf. Capelle aan den IJssel koppelde als één van de eerste gemeenten databestanden over onder meer belastingen aan die van het UWV, om uitkeringsfraude te ontdekken. De 'digitale stofkam' verving de 'ouderwetse' controle op het aantal tandenborstels.

De gemeente Alphen aan den Rijn zet data-analyse in om zogenoemde spookadressen op te sporen. Ze legt gegevens van de energiemaatschappij naast het aantal bewoners dat op een adres staat ingeschreven: kijkt het

stroomgebruik af van wat je logischerwijs verwacht, dan komt dat op de computer van ambtenaren naar boven.

Ook de politie heeft allerlei projecten lopen. Zo worden algoritmes in verschillende regio's ingezet om te bepalen waar meer gesurveilleerd wordt, omdat de kans op woninginbraken in een bepaalde buurt groter is volgens berekeningen van de computer. Ook loopt al een aantal jaar het project ProKid: een methode om vroegtijdig kinderen tot twaalf jaar op te sporen die het risico lopen een criminele carrière te ontwikkelen. Zodat jeugdzorg kan ingrijpen.

Naast het grote enthousiasme waarmee de overheid data koppelt – gedreven door het verlangen om geld te besparen en criminaliteit en fraude terug te dringen – klinken er ook zorgen. Volgens de Wetenschappelijke Raad voor het Regeringsbeleid (WRR) kan het ertoe leiden dat mensen het gevoel krijgen dat de overheid ze in de gaten houdt. En, zo schreef de WRR eind vorige maand in het rapport 'Big data in een vrije en veilige samenleving', het kan zijn dat burgers daardoor hun gedrag aanpassen.

Jaap-Henk Hoepman, privacyonderzoeker aan de Nijmeegse Radboud Universiteit, deelt die zorg: "Mensen kunnen zich bekeken voelen, en niet alleen degenen die grensoverschrijdend gedrag vertonen. Wat als je gewoon veel doucht, omdat je een huidziekte hebt, en in de systemen naar boven komt als iemand die voor een eenpersoonshuishouden wel heel veel water verbruikt? Krijg je dan het gevoel je gedrag aan te moeten passen om bezoeken van ambtenaren te voorkomen? We noemen dat een *chilling effect*."

Is dat de opsporing van relatief kleine vergrijpen waard, vraagt Hoepman zich af. Niet dat hij uitkeringsfraude of woninginbraken wil bagatelliseren, maar de maatschappelijke kosten die daar tegenover staan, moeten we niet onderschatten, stelt hij. Hij vergelijkt het met de controle op te hard rijden: technisch is het mogelijk om in alle auto's zouden een kastje te plaatsen dat automatisch een boete geeft als je te hard rijdt. "Daar hoor je niemand over", zegt Hoepman. "We voelen meteen dat het te ver gaat. Toch is het te vergelijken

Data-analyse wordt al gebruikt om probleemwijken in beeld te krijgen, en om toekomstige criminelen te herkennen

Wat als je gewoon veel doucht vanwege een huidziekte, maar in de systemen naar bovenkomt als iemand die mogelijk illegalen huisvest?

ken met de inzet van data-analyse om fraude te voorkomen. Al die brokjes data samen zorgen ervoor dat je in de gaten wordt gehouden."

En dan maakt het algoritme ook wel eens fouten. Hoewel de kracht van analyse van grote hoeveelheden data er vooral in zit dat tot dan toe onbekende patronen aan de oppervlakte komen. Om een voorbeeld te noemen: in zwarte BMW's die op bepaalde tijden op bepaalde plekken rijden, is bovengemiddeld veel kans op de aanwezigheid van drugs. Vertaald naar het individu zegt dat niet altijd wat: niet elke bestuurder van een zwarte BMW heeft kwaad in de zin.

"Uit de eerste evaluatie van het ProKid-project bleek dat in een derde van de gevallen onterecht een kind was aangemerkt als risicogeval", zegt Marc Schuilenburg. "Ik vind dat veel."

Voor burgers kan de impact door de verkeerde analyse groot zijn: jeugdzorg komt over de vloer, of je bent altijd de gene die aan de kant wordt gezet bij een verkeerscontrole. En dat terwijl je niets verkeerd hebt gedaan, je bent alleen volgens het algoritme een risicogeval. "Het idee dat je onschuldig bent tot het tegendeel is bewezen, komt daardoor op de helling", zegt Schuilenburg. "Dat is een grondbeginsel van ons rechtssysteem."

Precies dat is het bezwaar van de D66-fractie in de Apeldoornse gemeenteraad tegen de inzet van big data-analyse om overlast van jongeren tegen te gaan. Jongeren als Milan worden op basis van stigma's en beslissingen van computers in hokjes geplaatst, vreest de partij. Terwijl ze misschien nooit iets verkeerd hebben gedaan, of zullen doen.

Er bestaat bovendien een risico dat het altijd dezelfde personen, groepen of wijken zijn die een rode vlag achter hun naam krijgen, waarschuwt Frederik Zuiderveen Borgesius, die aan de Universiteit van Amsterdam onderzoek doet naar privacy en profilering. "Hoewel een algoritme objectief lijkt, is dat het zeker niet. Zo hangt veel af van de data die je erin stopt. Surveilleer je vaker in bepaalde wijken, dan worden daar meer gegevens over incidenten verzameld, waardoor die buurt waar-

schijnlijk weer uit de algoritmes rolt als risicogebied."

Een algoritme kan zo een zelfversterkend effect hebben. "Dat leidt ertoe dat bepaalde mensen of wijken steeds vaker de klos zijn. En laten we ervanuit gaan dat dat niet de witte mannen op de Amsterdamse Zuidas zijn", aldus Schuilenburg.

Discriminerende factor

Die discriminerende factor is het gevolg van de werkwijze bij data-analyses, zegt Arre Zuurmond. Hij is de gemeentelijk ombudsman van Amsterdam en was daarvoor bijzonder hoogleraar ICT en de toekomst van het openbaar bestuur in Leiden. Te vaak gooit een gemeente nu gewoon wat data bij elkaar, die worden gezeefd en dan is het afwachten wat er blijft liggen, zegt hij. "In plaats daarvan zouden gemeenten moeten denken in handelingspatronen, waarbij je uitgaat van concreet gedrag van een individu en het er niet toe doet of iemand jong, oud, man of vrouw, blank of zwart is. Als iemand voor de derde keer in vijf jaar een nieuw paspoort komt aanvragen, dan kan dat aanleiding zijn voor verder onderzoek. Zo weten we dat bolletjesslikkers stempels in hun paspoort voor de uitkeringsinstantie willen verbergen. Je kunt als gemeente bij de zoveelste aanvraag kijken of iemand een uitkering krijgt en zo nodig navraag doen bij de Marechaussee. Dat is een andere aanpak dan dat je iemand extra controleert omdat hij nou eenmaal in een bepaald profiel past."

Ondanks de bezwaren, wijst alles erop dat het koppelen en analyseren van databestanden om de risicoburger te ontdekken bij de overheid alleen maar zal toenemen (zie kader). Wat betekent dat? Minder fraude, minder overlast en daarmee een prettigere samenleving, zal de een zeggen. Maar ook een verschuiving van het machts-evenwicht tussen overheid en burger, zo waarschuwt de WRR. Overheden kijken steeds dieper in het leven van burgers, terwijl ze zelf juist minder transparant worden. Zo houden ze details over data-analyses vaak geheim om burgers met slechte bedoelingen niet in de kaart te spelen. Die disbalans gaat wringen, verwacht de WRR.