

## 6 Big data en gelijke behandeling

*Hans Lammerant, Peter Blok en Paul de Hert*

### 6.1 Inleiding

Big data kan vanuit het oogpunt van gelijke behandeling zowel voor- als nadelen hebben. Wanneer besluitvorming op meer data is gebaseerd en de analyse (deels) geautomatiseerd wordt, kan deze meer geobjectiverd worden en kan de rol van menselijke vooroordelen geëlimineerd of verminderd worden. Anderzijds, big data kan ook juist die vooroordelen opnemen in geautomatiseerde besluitvorming en zo aan het oog onttrekken. Eveneens maakt de grotere beschikbaarheid van data een meer gedifferentieerde behandeling van mensen mogelijk en de vraag is of de gronden op basis waarvan dat gebeurt altijd even wenselijk zijn. Hoe big data-praktijken juridisch moeten worden beoordeeld en gereguleerd vanuit het oogpunt van gelijke behandeling is daarom een zeer actueel probleem.

In dit hoofdstuk bekijken we allereerst de voornaamste basisbegrippen inzake gelijke behandeling en discriminatie (paragraaf 6.2). Vervolgens kijken we naar het Nederlandse en Europese juridisch kader inzake non-discriminatie (paragraaf 6.3-6.5) en hoe die regels moeten worden toegepast op big data-praktijken (paragraaf 6.6). Vervolgens gaan we in op de vraag hoe big data, en meer bepaald data mining, bedoeld of onbedoeld kan leiden tot discriminerende besluitvorming (paragraaf 6.7) en op de verhouding tussen het non-discriminatierecht en het gegevensbeschermingsrecht (paragraaf 6.8). Ten slotte denken we na over de vraag hoe discriminatie door big data-praktijken voorkomen kan worden (paragraaf 6.9).

### 6.2 Discriminatie

Om onze vraagstelling te begrijpen over het verband tussen big data en discriminerende praktijken, helpt het om een korte omweg te maken richting sociologie en rechtswetenschap en te kijken naar de respectievelijke concepten.

Het beginsel van gelijkheid is een erkend ethisch en juridisch beginsel. Zonder te stellen dat mensen van nature gelijk zijn, vergt het een gelijke behandeling van mensen als uitgangspunt. Innig daarmee verbonden is het verbod op discriminatie dat eraan in de weg staat dat individuen ongerechtvaardigd of nadelig behandeld worden op grond van hun lidmaatschap van een bepaalde groep of categorie. Centraal hierbij staat de term 'niet-gerechtvaardigd'. Elke besluitvorming over mensen brengt een verschil in behandeling teweeg en het is zaak om de niet-gerechtvaardigde van de gerechtvaardigde te onderscheiden. Bepaalde gronden voor besluitvorming worden als niet-gerechtvaardigd beschouwd: bepaalde fysieke of culturele eigenschappen bijvoorbeeld met

betrekking tot geslacht, etnische origine, religie of seksuele voorkeur. Het non-discriminatierecht heeft zich geleidelijk uitgebreid, zonder dat evenwel gesteld kan worden dat werkelijk alles dat als sociaal-discriminerend gezien kan worden, nu echt verboden is.

Discriminatie als sociaal fenomeen is complex en gehuld in vele verschijningen. In de sociologie worden drie discriminerende mechanismen onderscheiden: vooroordeel, rationele discriminatie en niet-intentionele of onbedoelde discriminatie. Vooroordelen zijn onfaire of onredelijke negatieve houdingen tegenover een groep die de basis kunnen vormen voor stereotypering op grond waarvan discriminerende behandelingen kunnen ontstaan. Rationele discriminatie vertrekt niet van dergelijke negatieve houdingen ten aanzien van groepen, maar vertrekt van een rationele redenering waarbij gebrek aan kennis over een individu wordt gecompenseerd met het gebruik van eerder opgedane kennis over de gemiddelde eigenschappen of resultaten in de groep waartoe het individu behoort. Opnieuw zien we hier gebruik van een stereotypering, maar ditmaal één die gerechtvaardigd kan worden op rationele, doch niet altijd op morele of juridische gronden. Te denken valt aan werkgevers die vrouwen achterstellen vanwege het 'risico' op zwangerschappen en verzekeraars die het geslacht in aanmerking nemen om risico's in te schatten.

Ten slotte is er niet-intentionele discriminatie, met name wanneer er geen intentie of wil is om te discrimineren maar waar, als gevolg van gebrek aan bewustzijn, bepaalde praktijken worden uitgeoefend met een discriminerend resultaat. Een voorbeeld hiervan zijn beoordelingstesten met impliciete culturele vooroordelen gebruikt in sollicitatie-procedures.

Zoals we in onze inleiding schreven, vormt de toepassing van big data-technieken vanuit het oogpunt van gelijke behandeling niet alleen een risico, maar ook een kans om discriminatie te verminderen en gelijkheid te bevorderen. Menselijke vooroordelen kunnen leiden tot discriminatie. Computersystemen waarmee big data wordt verwerkt hebben deze in beginsel niet, althans niet als die vooroordelen er niet in geprogrammeerd zijn. Toepassing van die systemen kan dus voorkomen dat beslissingen bewust of onbewust worden genomen op basis van vooroordelen. Daarnaast wezen we erop dat zogeheten rationele discriminatie optreedt als een gebrek aan kennis over een individu wordt aangevuld met kennis over een groep. Ook in dat opzicht kan de toepassing van big data discriminatie verminderen. Het verzamelen van big data kan er immers toe leiden dat er een veel completer beeld over het individu bestaat, waardoor beslissingen kunnen worden genomen op basis van gegevens die beter recht doen aan het individu. Zo kan een persoon die in het verleden geen lening kreeg omdat de kredietinstelling geen gegevens had waaruit bleek dat die persoon kredietwaardig was, die lening nu wel krijgen omdat meer bronnen worden meegenomen.<sup>1</sup> Ten slotte kunnen big data-toepassingen actief worden ingezet om materiële ongelijkheid te verminderen. *Georgia State University* heeft bijvoorbeeld het slagingspercentage van minder kansrijke groepen

1 Zie White House, Executive Office of the President, *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*, mei 2016, p.11-12, online beschikbaar via: [https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016\\_0504\\_data\\_discrimination.pdf](https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf).

aanzienlijk weten te verhogen door de vorderingen van studenten nauwgezet bij te houden en extra begeleiding te bieden waar dat nodig was.<sup>2</sup>

Tegelijkertijd vergroot de opkomst van big data het risico op ongelijke behandeling, simpelweg omdat die technieken meer differentiatie mogelijk maken. Zo was het voor een verzekeraar tot recent niet zinvol om onderscheid te maken in de hoogte van premies, omdat verzekeraars niet of nauwelijks inzicht hadden in de risico's op het niveau van individuen. De opkomst van big data verandert dat<sup>3</sup> en zal tot gevolg hebben dat in de toekomst vaker onderscheid zal worden gemaakt tussen verzekeringsnemers, bijvoorbeeld op grond van allerlei gegevens die ze zelf aanleveren via apps (over rijgedrag) en wearables. Die nieuwe vormen van onderscheid kunnen problematisch worden vanuit het oogpunt van gelijke behandeling als de gebruikte gegevens of de daarop gebaseerde besluitvormingsmodellen een bias bevatten. Dat brengt mee dat de regels voor gelijke behandeling de komende jaren in belang zullen toenemen. Een ander voorbeeld betreft de ontwikkelingen op het gebied van marketing. Waar voorheen het hele publiek vaak dezelfde reclameboodschap kreeg gepresenteerd, kan met behulp van big data een gedetailleerd onderscheid worden gemaakt tussen verschillen klanten en kan iedere klant een gepersonaliseerd aanbod worden voorgeschoteld. Die praktijk hoeft niet op gespannen voet te staan met de regels voor gelijke behandeling als de goederen of diensten wel voor iedereen beschikbaar blijven. Dat kan anders worden als het aanbod zodanig wordt gepersonaliseerd dat bepaalde voordelen feitelijk alleen voor bepaalde groepen beschikbaar zijn.<sup>4</sup>

### 6.3 Het non-discriminatierecht

In het non-discriminatierecht moet een onderscheid worden gemaakt tussen enerzijds het algemene gelijkheidsbeginsel en anderzijds de specifieke regels voor het discrimineren op specifieke gronden. Het algemene gelijkheidsbeginsel vereist dat vergelijkbare situaties niet verschillend en verschillende situaties niet gelijk worden behandeld. Dit algemene beginsel is uitgewerkt voor een aantal specifieke persoonskenmerken, zoals ras, geslacht en godsdienst.

Die specifieke kenmerken zijn vanuit het oogpunt van discriminatie 'verdacht', omdat ze in het algemeen irrelevant zijn of horen te zijn bij het maken van beslissingen. Daarom bestaan er specifieke, relatief strenge regels voor het maken van onderscheid op die specifieke gronden.

2 White House, Executive Office of the President, *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*, mei 2016, p.17, online beschikbaar via: [https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016\\_0504\\_data\\_discrimination.pdf](https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf).

3 Zie Rathenau Instituut, *Berekende risico's. Verzekeren in de datagedreven samenleving*, Den Haag 2015, online beschikbaar via: <https://www.rathenau.nl/nl/publicatie/berekende-risicos-verzekeren-de-datagedreven-samenleving>; Verbond van Verzekeraars, *Grip op data. Green paper Big Data*, april 2016, online beschikbaar via: <https://www.verzekeraars.nl/actueel/nieuwsberichten/Documents/2016/April/Grip%20op%20data%20-%20green%20paper%20Big%20Data.pdf>.

4 Zie over deze wijze van risicoselectie bij verzekeraars: Nederlandse ZorgAutoriteit, *Samenvattend rapport Rechtmatige uitvoering Zorgverzekeringswet 2013, Onderdelen acceptatieplicht, verbod op premiedifferentiatie en zorgplicht*, oktober 2014, online beschikbaar via: [https://www.nza.nl/1048076/1048181/Samenvattend\\_rapport\\_Rechtmatige\\_uitvoering\\_Zorgverzekeringswet\\_2013.pdf](https://www.nza.nl/1048076/1048181/Samenvattend_rapport_Rechtmatige_uitvoering_Zorgverzekeringswet_2013.pdf).

Het algemene gelijkheidsbeginsel staat onder meer in het eerste artikel van de Nederlandse Grondwet, artikel 26 van het Internationaal Verdrag voor de Burgerlijke en Politieke Rechten (IVBPR), het twaalfde protocol bij Europees Verdrag voor de Rechten van de Mens (EVRM) en de artikelen 20 en 21 van het Handvest van de Grondrechten van de Europese Unie en wordt door het Hof van Justitie bovendien gezien als algemeen beginsel van Unierecht.<sup>5</sup> Die grondrechten en beginselen schrijven voor dat allen gelijk zijn voor de wet en dat eenieder aanspraak heeft op gelijke behandeling. Ze verbieden niet alle vormen van onderscheid, maar alleen onderscheid dat niet is gebaseerd op “redelijke en objectieve gronden”.<sup>6</sup> Dat betekent dat alleen sprake is van discriminatie als het gemaakte onderscheid geen gerechtvaardigde doelstelling heeft of als er geen redelijke verhouding bestaat tussen de maatregel die het onderscheid maakt en het daarmee gerechtvaardigde doel.<sup>7</sup> Dit algemene gelijkheidsbeginsel heeft in de praktijk niet zo veel impact, onder meer omdat slechts marginaal getoetst wordt als niet de hierna te bespreken specifieke gronden van discriminatie aan de orde zijn.<sup>8</sup>

De meeste van de hiervoor genoemde artikelen uit de grondrechtenverdragen laten al zien dat de bescherming met name wordt geboden voor een aantal specifieke persoonskenmerken. Daarbij noemt niet elke rechtsbron uitdrukkelijk dezelfde specifieke kenmerken, maar bijvoorbeeld ras, geslacht en godsdienst komen voor in elk verdrag. Een van de meeste uitgebreide opsommingen staat in artikel 21 van het EU Handvest. Dat bevat een verbod op discriminatie, “met name op grond van geslacht, ras, kleur, etnische of sociale afkomst, genetische kenmerken, taal, godsdienst of overtuigingen, politieke of andere denkbeelden, het behoren tot een nationale minderheid, vermogen, geboorte, een handicap, leeftijd of seksuele geaardheid”.

Artikel 19(1) van het Verdrag betreffende de Werking van de Europese Unie geeft de Unie de bevoegdheid om passende maatregelen te nemen om discriminatie op grond van geslacht, ras of etnische afstamming, godsdienst of overtuiging, handicap, leeftijd of seksuele geaardheid te bestrijden. Specifieke Europese richtlijnen werken dit mandaat uit, doch met telkens, afhankelijk van de specifieke materie, grote of kleine verschillen.<sup>9</sup> Deze richtlijnen richten zich uitsluitend op gelijke behandeling op de arbeidsmarkt.

5 Zie over de bronnen van non-discriminatie recht in het kader van de EU: A. Ellis & P. Watson, *EU Anti-discrimination Law*, Oxford University Press, 2012, hoofdstuk 1 en 3.

6 Mensenrechtencomité van het IVBPR, 9 april 1987, *NJCM-Bulletin* 1987/5, p. 377-89; HR 7 mei 1993, ECLI:HR:1993:ZC0955, *AB* 1993/440, r.o. 3.2.

7 EHRM 29 april 2008, nr. 13378/05, *EHRC* 2008/80, r.o. 60; HR 22 november 2013, ECLI:HR:2013:1212.

8 R. Gellert et al., ‘A Comparative Analysis of Anti-Discrimination and Data Protection Legislations’, in: B. Custers, T. Calders, B. Schermer & T. Zarsky, *Discrimination and Privacy in the Information Society*, Springer, 2012.

9 Richtlijn 2000/43/EG van de Raad van 29 juni 2000 houdende toepassing van het beginsel van gelijke behandeling van personen ongeacht ras of etnische afstamming; Richtlijn 2000/78/EG van de Raad van 27 november 2000 tot instelling van een algemeen kader voor gelijke behandeling in arbeid en beroep; Richtlijn 2004/113/EG van de Raad van 13 december 2004 houdende toepassing van het beginsel van gelijke behandeling van mannen en vrouwen bij de toegang tot en het aanbod van goederen en diensten; Richtlijn 2006/54/EG van het Europees Parlement en de Raad van 5 juli 2006 betreffende de toepassing van het beginsel van gelijke kansen en gelijke behandeling van mannen en vrouwen in arbeid en beroep.

Alleen voor geslacht is er een richtlijn die zich ook uitstrekt tot het leveren van goederen en diensten.<sup>10</sup>

In Nederland is bescherming tegen discriminatie op specifieke gronden, mede ter implementatie van de bovengenoemde Europese richtlijnen, uitgewerkt in een reeks wetten. De belangrijkste wet in dit verband is de Algemene wet gelijke behandeling (AWGB). Die wet verbiedt discriminatie op grond van godsdienst, levensovertuiging, politieke gezindheid, ras, geslacht, nationaliteit, hetero- of homoseksuele gerichtheid of burgerlijke staat. De AWGB is van toepassing op het gebied van arbeid (bijvoorbeeld werving en selectie, bemiddeling, ontslag, arbeidsvoorwaarden en toegang tot vrije beroepen) en op sociaaleconomisch terrein (zoals sociale zekerheid en het aanbieden van goederen en diensten). Daarnaast bestaan de Wet gelijke behandeling op grond van leeftijd (WGBL), de Wet gelijke behandeling op grond van handicap of chronische ziekte (WGBH/CZ), de Wet gelijke behandeling van mannen en vrouwen (WGB), de Wet onderscheid arbeidsduur (WOA) en de Wet onderscheid bepaalde en onbepaalde tijd (WOBOT), die – zoals de namen van de wetten duidelijk maken – discriminatie verbieden op grond van respectievelijk leeftijd, handicap, geslacht, arbeidsduur en deeltijdarbeid.

Het toepassingsbereik van deze wetten hangt niet alleen af van de gronden waarop onderscheid wordt gemaakt, maar ook van het gebied waarop dat gebeurt. De grenzen van die gebieden zijn wat ruimer getrokken dan het bereik van de Europese richtlijnen. Hierna volgt voor het overzicht een schema.

Voor de volledigheid wordt erop gewezen dat bepaalde vormen van discriminatie ook strafbaar zijn gesteld. Op grond van artikel 429quater van het Wetboek van Strafrecht is discriminatie wegens ras, godsdienst, levensovertuiging of hetero- of homoseksuele gerichtheid in de uitoefening van een ambt, beroep of bedrijf strafbaar. Het begrip ‘discriminatie’ wordt in artikel 90quater van het Wetboek van Strafrecht gedefinieerd als elke vorm van onderscheid die ten doel of gevolg kan hebben dat de erkenning, het genot of de uitoefening van fundamentele rechten teniet wordt gedaan of aangetast. Aangenomen wordt dat alleen *ongerechtvaardigd* onderscheid onder dit begrip valt en dat voor de rechtvaardigingsgronden kan worden aangesloten bij de hiervoor besproken gelijkebehandelingswetgeving.<sup>11</sup>

10 Richtlijn 2004/113/EG van de Raad van 13 december 2004 houdende toepassing van het beginsel van gelijke behandeling van mannen en vrouwen bij de toegang tot en het aanbod van goederen en diensten.

11 Zie M.L.M. Hertogh & P.J.J. Zoontjes (red.), *Gelijke behandeling: principes en praktijken*, Wolf Legal Publishers, 2006, p. 29-45.

Wet	Grond	Gebied							
		Arbeid	Vrij beroep	Vakbond	Onderwijs	Wonen	Vervoer	Overig goed/dienst	Sociale voorzieningen
AWGB	Godsdienst	X	X	X	X	X	X	X	X
	Politieke gezindheid	X	X	X	X	X	X	X	X
	Ras	X	X	X	X	X	X	X	X
	Geslacht	X	X	X	X	X	X	X	X
	Zwangerschap	X	X	X	X	X	X	X	X
	Nationaliteit	X	X	X	X	X	X	X	X
	Homo/hetero	X	X	X	X	X	X	X	X
	Burgerlijke staat	X	X	X	X	X	X	X	X
WGBH/ CZ	Handicap	X	X	X	X	X	X		
WGBL	Leeftijd	X	X	X	X				
WGB	Geslacht	X							
WOA	Voltijd/deeltijd	X							
WOBOT	Vast/tijdelijk contract	X							

#### 6.4 Directe en indirecte discriminatie

Alle genoemde wetten en richtlijnen verbieden zowel directe als indirecte discriminatie en ook in het kader van de rechtspraak over fundamentele rechten komen beide vormen voor.<sup>12</sup> Het onderscheid tussen de twee vormen van discriminatie is van belang omdat de rechtvaardigingsgronden voor directe discriminatie veel beperkter zijn. Directe discriminatie mag nooit, tenzij een beroep kan worden gedaan op de in de wetgeving nauw omschreven uitzonderingen voor specifieke situaties, zoals positieve discriminatie van vrouwen (zie bijvoorbeeld artikel 2 lid 2 AWGB). Indirecte discriminatie is toegelaten in meer gevallen. Hier geldt de algemene rechtvaardigingsgrond dat het onderscheid is toegestaan als het objectief gerechtvaardigd wordt door een legitiem doel en de middelen voor het bereiken van dat doel passend en noodzakelijk zijn.

De begrippen directe en indirecte discriminatie zijn gedefinieerd in de wetten en richtlijnen. Van directe discriminatie is sprake wanneer iemand op grond van een van de specifieke kenmerken ongunstiger wordt behandeld dan een ander in een vergelijkbare situatie wordt, is of zou worden behandeld. Indirecte discriminatie doet zich voor wanneer een ogenschijnlijk neutrale bepaling, maatstaf of handelwijze personen met een van de specifieke kenmerken in vergelijking met andere personen bijzonder benadeelt.

<sup>12</sup> Zie over het onderscheid onder meer European Union Agency for Fundamental Rights: *Handbook on European non-discrimination law*, Imprimerie Centrale, 2010, p. 22-31; A. Ellis & P. Watson, *EU Anti-discrimination Law*, Oxford University Press 2012, hoofdstuk 4; voor de bescherming tegen indirect discriminatie onder het EVRM zie EHRM 13 november 2007, 57325/00, ECLI:CE:ECHR:2007:1113JUD005732500, *D.H. v. Tsjechië*.

Bij directe discriminatie gaat het er dus kort gezegd om dat een onderscheid wordt gemaakt op basis van een van de specifieke kenmerken, bijvoorbeeld als een vrouw voor een baan wordt geweigerd op basis van haar geslacht.<sup>13</sup> Daarbij moet niet alleen komen vast te staan dat een negatieve beslissing over een vrouw is genomen, maar ook dat die beslissing anders zou zijn uitgevallen voor een man die zich in een vergelijkbare situatie bevond.<sup>14</sup> Zo wees het Hof van Justitie een klacht over discriminatie op grond van geslacht af in een zaak waarin een vrouw minder betaald kreeg dan mannelijke collega's. De vrouw bevond zich naar het oordeel van het Hof niet in een vergelijkbare situatie als de mannen omdat alleen zij als uitzendkracht werkte en de mannen rechtstreeks in dienst waren.<sup>15</sup>

Uit rechtspraak van het Hof van Justitie volgt dat ook sprake is van directe discriminatie als onderscheid wordt gemaakt op basis van een kenmerk dat onlosmakelijk is verbonden met een van de verboden gronden, bijvoorbeeld als een vrouw wordt geweigerd niet zozeer omdat ze vrouw is, maar omdat een toekomstige zwangerschap de werkgever ziekingeld zou kosten.<sup>16</sup> Al voordat zwangerschap als afzonderlijke discriminatiegrond was opgenomen in een richtlijn, kwalificeerde het Hof dat onderscheid als directe discriminatie op grond van geslacht omdat het motief rechtstreeks verband hield met het geslacht van de sollicitant.<sup>17</sup>

In civiele procedures is het in beginsel aan het slachtoffer van discriminatie om te bewijzen dat er sprake is van een onderscheid op een verboden grond. De Hoge Raad heeft echter geoordeeld dat de rechter cijfermatige, statistische verschillen kan aanmerken als voldoende motivering voor de stelling dat er is gediscrimineerd. Ook kan hij aan die cijfers het vermoeden ontleenen dat er sprake is van directe discriminatie en op de wederpartij de last leggen door middel van specifiek op het slachtoffer gerichte argumenten aan te tonen dat de achterstelling berust op rechtens aanvaardbare gronden.<sup>18</sup> Die uit de rechtspraak van de Hoge Raad blijvende omkering van de bewijslast komt ook naar voren in de rechtspraak van het Europese Hof voor de Rechten van de Mens<sup>19</sup> en is inmiddels ook opgenomen in de Europese richtlijnen en de gelijkebehandelingswetten. Die schrijven voor dat de verweerder moet aantonen dat het beginsel van gelijke behandeling niet werd geschonden als de klager feiten kan aanvoeren die discriminatie doen vermoeden. Aldus kan een ondoorzichtig beleid ertoe leiden dat directe discriminatie wordt aangenomen.<sup>20</sup>

13 Zie de definitie in onder meer artikel 1 lid 1 sub b AWGB.

14 Zie European Union Agency for Fundamental Rights, *Handbook on European non-discrimination law*, Imprimerie Centrale 2010, paragraaf 2.2.2.

15 HvJ EU 13 januari 2004, C-256/01, ECLI:EU:C:2004:18, *Allonby*.

16 European Union Agency for Fundamental Rights, *Handbook on European non-discrimination law*, Imprimerie Centrale 2010, paragraaf 2.2.3.

17 HvJ EU 8 november 1990, C-177/88, ECLI:EU:C:1990:383, *Dekker*, 10.

18 HR 10 december 1982, NJ 1983/687; zie ook EHRM 13 november 2007, 57325/00, ECLI:CE:ECHR:2007:1113JUD005732500, *D.H. v. Tsjechië*.

19 EHRM 6 January 2005, 58641/00, *Hoogendijk v. the Netherlands*; European Union Agency for Fundamental Rights, *Handbook on European non-discrimination law*, Imprimerie Centrale 2010, paragraaf 5.3.

20 HvJ EU 10 oktober 1989, C-109/88, ECLI:EU:C:1989:383.

Bij indirecte discriminatie wordt geen onderscheid gemaakt op basis van een van de specifieke gronden, maar hierbij is het gevolg van het onderscheid wel dat een groep personen met een van de specifieke kenmerken wordt benadeeld. Aangenomen wordt dat van indirecte discriminatie sprake is als personen met een ‘verboden kenmerk’ in overwegende of disproportionele mate deel uitmaken van de groep van benadeelde personen. Daarbij wordt zowel gelet op de mate van oververtegenwoordiging van die groep als op de continuïteit van een verschil in behandeling.<sup>21</sup>

Eenieder die slachtoffer is van een discriminatoire praktijk kan een beroep doen op de genoemde wetten en richtlijnen, ook als die persoon zelf niet behoort tot de groep met een van de specifieke kenmerken. Dat blijkt uit het arrest *Chez* van het Hof van Justitie. Het *Chez*-arrest betrof het plaatsen van elektriciteitsmeters op grote hoogte in hoofdzakelijk door Roma bewoonde wijken met als reden fraude te voorkomen. Die praktijk werd gekwalificeerd als discriminatoir omdat die de Roma-bevolking bijzonder benadeelt en was daarmee onrechtmatig ten opzichte van iedere inwoner van de wijk, ook inwoners die geen Roma-achtergrond hadden.<sup>22</sup>

### 6.5 Rechtvaardiging

Zowel bij directe als bij indirecte discriminatie staan er rechtvaardigingsgronden ter beschikking en moet een proportionaliteitsafweging gemaakt worden. De belangrijkste consequentie van het verschil tussen directe en indirecte discriminatie is dat die beschikbare rechtvaardigingsgronden variëren. Bij indirecte discriminatie is er sprake van een algemene rechtvaardigingsgrond vervat in de definitie. Bij directe discriminatie gaat het om een beperkt aantal nauw omschreven rechtvaardigingsgronden, die bovendien specifiek zijn naargelang het beschermde kenmerk en voor een bepaalde context.

Als er sprake is van directe discriminatie is het onderscheid alleen toegestaan als het onderscheid valt onder een van de specifieke uitzonderingen die in de betreffende wetten staan opgesomd. Zo voorziet de AWGB, in overeenstemming met Richtlijn 2000/43, in een uitzondering op het verbod van ongelijke behandeling gebaseerd op een kenmerk in verband met ras of etnische afstamming wanneer dit berust op “wezenlijke en bepalende beroepsvereisten” (artikel 2 lid 4 sub b AWGB). Wat daaronder moet worden verstaan is limitatief uitgewerkt in artikel 2 van het Besluit gelijke behandeling:

- a. “de uitoefening van het beroep of de activiteit van acteur, danser of kunstenaar, voor zover het beroep of de activiteit betrekking heeft op het vertolken van een bepaalde rol;
- b. de uitoefening van het beroep of de activiteit van mannequin die bepaalde kledingstukken moet tonen door deze te dragen, van model voor beeldend kunstenaars, fotografen, cineasten, kappers, grimeurs en schoonheidsspecialisten, voor zover in redelijkheid eisen kunnen worden gesteld aan uiterlijke kenmerken.”

21 Zie A.W. Heringa, *Algemene Wet Gelijke Behandeling*, Deventer: Kluwer, 1994, p. 43-45; J.H. Gerards & A.W. Heringa, *Wetgeving gelijke behandeling*, Deventer: Kluwer, 2003, paragraaf 2.2.3.

22 HvJ EU 16 juli 2015, C-83/14, ECLI:EU:C:2015:480, *CHEZ Razpredelenie Bulgaria AD*, r.o. 60.



Niet voor alle specifieke gronden is de beoordeling even streng. Voor bijvoorbeeld leeftijdsdiscriminatie geldt een ruimere, open rechtvaardigingsgrond. Die luidt hetzelfde als de rechtvaardigingsgrond voor indirecte discriminatie, die hierna zal worden besproken: onderscheid op basis van leeftijd mag als het objectief gerechtvaardigd wordt door een legitiem doel en de middelen voor het bereiken van dat doel passend en noodzakelijk zijn (artikel 7 lid 1 sub b WGBL). Uit rechtspraak van het Hof van Justitie blijkt dat een leeftijdsvereiste bij fysiek zware beroepen als brandweerman, piloot of politiemann op die grond in principe gerechtvaardigd kan zijn.<sup>23</sup> De rechtvaardiging moet wel met objectieve gegevens worden verantwoord, net als het niet voorhanden zijn van minder ingrijpende maatregelen (bijvoorbeeld fysieke tests).

De definitie van indirecte discriminatie in de richtlijnen beschrijft dat een rechtvaardiging mogelijk is “door een legitiem doel en de middelen voor het bereiken van dat doel passend en noodzakelijk zijn”. Ten eerste moet er sprake zijn van een legitiem doel. Aangezien indirecte discriminatie een algemene rechtvaardigingsgrond bevat, zijn in dit geval zeer verschillende doelen mogelijk. Ten tweede moeten de gebruikte middelen passend en noodzakelijk zijn.<sup>24</sup> Passend wil zeggen dat de maatregel geschikt is om het nagestreefde legitieme doel te bereiken.<sup>25</sup> Geschikt betekent niet dat doelstellingen volledig bereikt moeten worden. Een merkbaar verschil, bijvoorbeeld reductie in fraude of verlaging van een bepaald risico, is voldoende om als geschikt gezien te worden.<sup>26</sup> Noodzakelijk impliceert dat de legitieme doelstellingen niet door “een minder ingrijpende, maar even geschikte maatregel” kunnen worden bereikt.<sup>27</sup>

Ten derde moet nagegaan worden of de door de gebruikte middelen veroorzaakte nadelen niet onevenredig afbreuk doen aan het gewaarborgde recht op gelijke behandeling en of die middelen geen excessieve inbreuk maken op de legitieme belangen van de betrokkenen.<sup>28</sup> Dit impliceert dat een evenwicht moet worden gevonden tussen de verschillende belangen en dat het nagestreefde legitieme doel zo veel mogelijk in overeenstemming moet worden gebracht met het principe van gelijke behandeling.<sup>29</sup>

## 6.6 Big data

Bij de toepassing van het non-discriminatierecht op big data-praktijken is het zinvol om een onderscheid te maken tussen twee fasen in een besluitvormingsproces waarin geautomatiseerde gegevensverwerkingen kunnen worden ingezet. Ten eerste kan de

23 HvJ EU 12 januari 2010, C-229/08, ECLI:EU:C:2010:3, *Wolf*; HvJ EU 13 september 2011, C-447/09, ECLI:EU:C:2011:573, *Prigge e.a.*; HvJ EU 13 november 2014, C-416/13, ECLI:EU:C:2014:2371, *Vital Pérez*.

24 HvJ EU 16 juli 2015, C-83/14, ECLI:EU:C:2015:480, *CHEZ Razpredelenie Bulgaria AD*, r.o. 118.

25 Conclusie A-G HvJ EU 12 maart 2015, C-83/14, ECLI:EU:C:2015:170, *CHEZ Razpredelenie Bulgaria AD*, 121.

26 Conclusie A-G HvJ EU 12 maart 2015, C-83/14, ECLI:EU:C:2015:170, *CHEZ Razpredelenie Bulgaria AD*, 123.

27 Conclusie A-G HvJ EU 12 maart 2015, C-83/14, ECLI:EU:C:2015:170, *CHEZ Razpredelenie Bulgaria AD*, 126.

28 HvJ EU 16 juli 2015, C-83/14, ECLI:EU:C:2015:480, *CHEZ Razpredelenie Bulgaria AD*, 123.

29 Conclusie A-G HvJ EU 12 maart 2015, C-83/14, ECLI:EU:C:2015:170, *CHEZ Razpredelenie Bulgaria AD*, 131.

besluitvorming zelf worden geautomatiseerd, bijvoorbeeld door een computer standaard een bepaalde beslissing (zoals wel/geen lening) te laten nemen als aan bepaalde gedefinieerde voorwaarden is voldaan, of door de computer op basis van die variabelen een bepaalde waarde (zoals een *credit score*) te laten berekenen die de basis vormt voor verdere besluitvorming. In zijn meest eenvoudige vorm (als P, dan Q) ziet zo'n besluitvormingsmodel of algoritme er bijvoorbeeld als volgt uit: als een persoon in een bepaalde wijk woont, krijgt die persoon geen lening. Big data-besluitvormingsprocessen werken op dezelfde manier, met dit verschil dat er met veel meer variabelen rekening wordt gehouden dan één kenmerk. Er wordt doorgaans gewerkt met profielen, die zijn opgebouwd uit een waaier van variabelen. Dankzij big data-toepassingen worden die profielen steeds uitgebreider en kunnen de consequenties die aan bepaalde (combinaties van) variabelen moeten worden verbonden steeds nauwkeuriger worden berekend.

Big data-systemen kunnen ook worden gebruikt in de fase voorafgaand aan de besluitvorming, te weten bij het ontwikkelen van de besluitvormingsmodellen of algoritmes. In deze fase worden de data-analysetechnieken gebruikt die Custers heeft beschreven in hoofdstuk 2.4.3 van dit boek. Het systeem gaat dan in één of meerdere datasets op zoek naar correlaties tussen bepaalde gegevens, bijvoorbeeld naar gegevens die correleren met het al dan niet terugbetalen van een lening. Als het systeem vaststelt dat er een verband bestaat tussen enerzijds de schermgrootte van een televisie die iemand koopt en anderzijds de tijd die het duurt dat een lening voor een televisie wordt terugbetaald, kan de variabele 'schermgrootte' worden toegevoegd aan het profiel dat een elektronicawinkel gebruikt om te bepalen onder welke voorwaarden iemand een televisie op krediet kan kopen.

Bij de beantwoording van de vraag of een bepaalde big data-praktijk in overeenstemming is met het non-discriminatie-recht moet ten eerste worden beoordeeld of de specifieke wetgeving op het gebied van gelijke behandeling, zoals de Algemene Wet Gelijke Behandeling, van toepassing is. Als dat zo is, moet de big data-toepassing immers worden getoetst aan het soms vrij restrictieve regime van die wetgeving. Zoals hiervoor is uiteengezet hangt het toepassingsbereik van de verschillende wetten af van zowel de grond (bijvoorbeeld ras of leeftijd) als het gebied (bijvoorbeeld arbeid of de levering van goederen/diensten). Zo zal big data veelal worden gebruikt in het kader van de levering van goederen en diensten, zoals verzekeringen en kredieten. Op dat terrein is de AWGB van toepassing, maar gelden de overige gelijkebehandelingswetten niet. Verder is het belangrijk te wijzen op het feit dat deze regels niet van toepassing zijn als de gegevensverwerking niet leidt tot een verschil in behandeling van personen. Een verzekeringsbedrijf kan bijvoorbeeld gegevens over geslacht verwerken in modellen om haar risicoblootstelling op bedrijfsniveau in te schatten of om gemiddelde verzekeringspremies te berekenen. De regels inzake gelijke behandeling worden pas van toepassing wanneer er sprake is van differentiatie op individueel niveau.<sup>30</sup>

Als een van deze specifieke wetten van toepassing is, moet ten tweede worden beoordeeld of er sprake is van directe discriminatie. Directe discriminatie zal zich kunnen voordoen

30 Zie bijvoorbeeld Europese Commissie, Richtsnoeren betreffende de toepassing van Richtlijn 2004/113/EG van de Raad op verzekeringen, in het licht van het arrest van het Hof van Justitie van de Europese Unie in zaak C-236/09 (*Test-Achats*), 13 januari 2012.

als een van de ‘verboden’ gronden is opgenomen in het model dat wordt gebruikt in de fase van besluitvorming, bijvoorbeeld als ras is opgenomen in het profiel op basis waarvan het beslismodel bepaalt of iemand in aanmerking komt voor een lening. Het bijzondere aan big data-toepassingen is dat het altijd gaat om een veelheid aan factoren. Een beslissing zal dus nooit worden gebaseerd enkel op de verboden grond, maar op basis van een veelheid van variabelen waaronder een verboden grond. Dat brengt mee dat het bij big data-toepassingen voor een slachtoffer van discriminatie bijzonder lastig zal zijn om – zonder kennisgeving van het model – directe discriminatie aan te tonen. Daarvoor moet immers niet alleen komen vast te staan dat een negatieve beslissing is genomen, maar ook dat die beslissing anders zou zijn uitgevallen voor iemand die zich, afgezien van de verboden grond, in eenzelfde situatie bevindt. Strikt genomen zou dat betekenen dat een voorbeeld moet worden gevonden van een persoon van wie alle in het profiel opgenomen kenmerken identiek zijn, met uitzondering van het kenmerk ras.

Daarbij moet wel worden bedacht dat van het bedrijf of de organisatie die het beslismodel toepast, kan worden verlangd te bewijzen dat de verboden grond *niet* in het beslismodel voorkomt zodra er een vermoeden van discriminatie is. Als het bedrijf of de organisatie geen inzage kan of wil geven in de werking van het model, bijvoorbeeld vanwege de concurrentiegevoeligheid daarvan, bestaat voor het bedrijf of de organisatie dus het risico dat directe discriminatie wordt aangenomen ook als er geen verboden gronden in het beslismodel staan.

Als er sprake is van directe discriminatie moet vervolgens worden beoordeeld of het onderscheid valt onder een van de wettelijke rechtvaardigingsgronden. Omdat die rechtvaardigingsgronden beperkt zijn, is directe discriminatie in het algemeen moeilijk te rechtvaardigen, zelfs als er een objectieve, door zorgvuldige big data-analyses ondersteunde basis bestaat voor het onderscheid. De wet staat bijvoorbeeld niet toe dat een autoverzekeraar bij het bepalen van de premie onderscheid maakt op basis van geslacht. Een door big data gevoede analyse die aantoont dat vrouwen echt minder vaak ongelukken veroorzaken, rechtvaardigt dat onderscheid niet.<sup>31</sup> Bij leeftijdsdiscriminatie ligt dat anders, omdat in dat kader een ruimere rechtvaardigingsgrond beschikbaar is. Big data-analyses die aantonen dat er daadwerkelijk een correlatie bestaat tussen de geschiktheid voor een beroep en leeftijd kunnen dan wel behulpzaam zijn, mits die analyses zelf geen bias bevatten (zie daarvoor de volgende paragraaf).

Als er geen directe discriminatie is, moet ten derde worden getoetst of er sprake is van indirecte discriminatie. Indirecte discriminatie zal zich voordoen als er geen verboden gronden in het beslismodel zijn opgenomen, maar er een correlatie bestaat tussen de wel in het model gebruikte variabelen en de verboden gronden. Zo kan een onderscheid op basis van iemands adres meebrengen dat bepaalde raciale groepen worden benadeeld omdat groepen met een bepaald ras oververtegenwoordigd zijn in bepaalde wijken. Het feit dat het bedrijf of de organisatie niet de intentie heeft om te discrimineren op grond van ras en zich daar ook niet van bewust is, maakt dat niet anders. Voor indirecte discriminatie volstaat dat het beslismodel personen met een van de specifieke kenmerken

31 Zie het reeds genoemde *Test Achats*-arrest van het HvJ EU.

in vergelijking met andere personen bijzonder benadeelt, tenzij die benadeling objectief gerechtvaardigd is.

Ook voor indirecte discriminatie geldt dat in beginsel het slachtoffer de discriminatie moet bewijzen, maar dat die bewijslast omdraait als het slachtoffer een vermoeden van discriminatie hard kan maken. Normaliter kan dat vermoeden worden gebaseerd op statistische gegevens die aantonen dat de neutrale grond waarop een bedrijf of organisatie een onderscheid maakt, resulteert in een oververtegenwoordiging van bepaalde groepen. Bij big data-toepassingen is er echter niet één grond bepalend voor het onderscheid, maar een veelheid aan factoren en is het model of algoritme waarin die zijn opgenomen niet bekend bij het slachtoffer. Dat maakt dat bij big data-toepassingen zelfs een vermoeden van indirecte discriminatie lastig te onderbouwen zal zijn. Dat kan ervoor pleiten bij dergelijke toepassingen eerder de bewijslast om te draaien, bijvoorbeeld door een vermoeden van discriminatie aan te nemen in de situatie dat het slachtoffer kan wijzen op enkele voorbeelden waarin personen met een beschermd kenmerk nadeliger werden behandeld dan anderen en het bedrijf geen inzicht geeft in zijn besluitvormingsmodel.

Als een beslismodel personen met een van de specifieke kenmerken bijzonder benadeelt, omdat de als negatief gewaardeerde variabelen vaker voorkomen bij die groep personen, zal het bedrijf dat het model toepast een objectieve rechtvaardiging moeten kunnen aanvoeren. In dat kader zal ten eerste duidelijk moeten worden gemaakt dat het onderscheid een legitiem doel dient, bijvoorbeeld het voorkomen dat leningen niet worden terugbetaald. Daarnaast zal moeten worden aangetoond dat het gebruikte beslismodel en de daarin opgenomen variabelen een geschikt middel zijn om dat doel te bereiken. De voor big data-toepassingen belangrijkste, maar ook lastigste vraag is onder welke omstandigheden het feit dat het beslismodel is gebaseerd op correlaties die big data-analyses hebben gevonden, kan worden aanvaard als objectieve rechtvaardiging. Het korte antwoord op die vraag luidt dat die analyses behulpzaam kunnen zijn, maar dat ze indirecte discriminatie niet altijd kunnen rechtvaardigen. Bij de opzet van de analyses worden namelijk keuzes gemaakt die het discriminatoire effect van het resulterende beslismodel kunnen beïnvloeden. Als gekozen had kunnen worden voor een minder discriminatoire, maar even effectieve methode, zal dat afbreuk doen aan de rechtvaardiging van het onderscheid. In de navolgende paragraaf zal dat concreter worden gemaakt door te laten zien hoe keuzes in de opzet van de analyse een discriminatoire effect kunnen hebben.

Als ten vierde een bepaalde praktijk buiten het toepassingsbereik van de specifieke gelijkebehandelingswetten valt, zal nog moeten worden getoetst aan het algemene gelijkheidsbeginsel. Ook in dat kader geldt dat het onderscheid objectief gerechtvaardigd moet zijn en dat de lat hoger ligt voor gronden die onlosmakelijk met een persoon verbonden zijn, zoals geslacht, ras en etnische afkomst. Of al dan niet statistische gegevens aanvaardbaar zijn of dat meer vaststaande gegevens nodig zijn, is onderdeel van de proportionaliteitsafweging. Hoe groter de impact, hoe meer zekerheid benodigd is alvorens een rechtvaardiging te aanvaarden.

## 6.7 Discriminatoire effecten van data mining

Keuzes bij het opstellen van een besluitvormingsmodel en de in dat verband gebruikte data-analyses kunnen op verschillende manieren leiden tot discriminatie. Meestal gaat het om vormen van niet-intentionele discriminatie, maar het technologisch ondersteund beslisproces kan ook gebruikt worden om een vorm van directe discriminatie te verhullen. We kunnen de sluiptwegen van mogelijke discriminatie in dit proces nagaan op drie niveaus:

1. *het model als representatie van de realiteit*: weerspiegelt het besluitvormingsmodel adequaat de onderliggende realiteit of zorgt het voor onbedoelde bias? Dit heeft betrekking op de variabelen die we in het resulterende model gebruiken en die worden aangereikt via onze probleemdefinitie en via de beschikbare data;
2. *de data als representatie van de realiteit*: weerspiegelt de training data waarmee het model wordt opgebouwd adequaat die onderliggende realiteit? Deze dataset kan op verschillende manieren onderhevig zijn aan een bias;
3. *de data als representatie van ongewenste informatie*: de dataset kan een weerspiegeling zijn van vooroordelen en historische discriminatie waarvan juist afstand genomen moet worden. Gevolg is dat modellen op basis van die dataset ook die bias zullen vertonen en, wanneer hiervoor niet gecorrigeerd wordt, kunnen leiden tot ongelijke behandeling.

In het volgende gaan we dieper in op deze sluiptwegen.

### 6.7.1 Sluipweg 1: het model als representatie van de realiteit

Weerspiegelt het besluitvormingsmodel adequaat de onderliggende realiteit? Het model moet een relatie leggen tussen een probleem- of vraagstelling, uitgedrukt in één of meerdere te verklaren doelvariabelen (betaalt iemand zijn lening terug?), en een reeks predictoren of verklarende variabelen (in welke wijk woont iemand?).

Een eerste mogelijke bron van discriminatie vormt de omschrijving van de problemen waarvoor een oplossing wordt gezocht. Aan de *data miner* of data-analist komt de moeilijke taak toe om dit probleem te vertalen in een meer formele vraagstelling waar computers mee aan de slag kunnen, onder meer door het bepalen van de te modelleren doelvariabelen. Voor zo'n te verklaren doelvariabele moet ook een serie mogelijke waarden of 'labels' die deze variabele kan hebben, worden vastgelegd. Die labels kunnen een aantal discrete waarden of klassen zijn (bijvoorbeeld kredietwaardig of niet) of een te schatten waarde (bijvoorbeeld een risico score). Deze vertaalslag van een probleem naar concrete te modelleren variabelen en de waarden die die variabelen kunnen krijgen kan niet-intentioneel uitdraaien op een beslissysteem dat negatief uitdraait voor bepaalde groepen. Definities van doelvariabelen – met andere woorden wat is een 'goed' resultaat of wat is een 'risico' – kunnen subjectief zijn of worden beïnvloed door andere factoren, zoals de noodzaak een probleem te vertalen in meetbare en in de database vindbare criteria. Een voorbeeld vormen diploma's of academische graden. Deze kunnen wijzen op bepaalde competenties, maar deze competenties kunnen ook op andere manieren worden verkregen, zoals ervaring of zelfstudie. Het meetbare criterium van een diploma

is bijgevolg niet zaligmakend en kan andere sociale factoren weerspiegelen, zoals het hebben van de economische middelen om te betalen voor een academische studie. Bovendien kan het hebben van een diploma variëren tussen leeftijdsgroepen of geografische herkomst, afhankelijk van hoe toegankelijk het formele academische of hoger onderwijs was of is. Het hebben van een diploma kan gebruikt worden om competenties te meten, maar onder omstandigheden kan er dus een bias optreden met discriminerende effecten als resultaat. Een ander voorbeeld is het gebruik van gegevens over de lengte van vorige aanstellingen als graadmeter voor het nemen van beslissingen over het aannemen van nieuw personeel. Wanneer de gegevens weerspiegelen dat vrouwen kortere aanstellingen hebben als gevolg van zwangerschappen of kinderopvang, kan dit criterium leiden tot een onbedoelde maar indirecte discriminatie van vrouwen. Het probleem kan ook opduiken zonder dat big data-technieken worden gehanteerd, maar er is een niet geringe kans dat bij het gebruik van technologisch ondersteunde beslissingsprocessen de discriminatie heviger en minder zichtbaar wordt.<sup>32</sup>

Een tweede onderdeel van het model, en dus ook een mogelijke bron van discriminatie, zijn de voorspellende variabelen of predictoren. Welke die variabelen kunnen zijn, wordt aangereikt via de beschikbare attributen in de dataset. In de klassieke statistische vraagstelling wordt op voorhand een model vastgelegd in een statistische hypothese, die vervolgens op de data getest wordt. Hier wordt dus een expliciete keuze gemaakt van relevante attributen door de data-analist. Bij data mining beschikt men over een grote reeks mogelijke variabelen aanwezig in de dataset. Welke variabelen relevant zijn en als predictor in het model worden opgenomen, wordt nu geselecteerd via het data mining-algoritme. Dit data mining-algoritme stelt dus tegelijk de hypothesen op en selecteert welke het meest geschikt zijn om als model te functioneren. Bij sommige algoritmes wordt het ondoorzichtig welke variabelen effectief gebruikt worden in een model of hoe het algoritme tot een bepaalde selectie van predictoren heeft besloten. Op het niveau van de beschrijving van de realiteit door het model stelt zich dezelfde vraag naar het adequaat zijn van de predictoren en of deze voor bepaalde groepen of categorieën tot onbedoelde discriminatie kunnen leiden. Aangezien deze predictoren via de data worden aangereikt zitten we hier al bij de volgende vraag.

### 6.7.2 *Sluipweg 2: de data als representatie van de realiteit*

Weerspiegelt de training data waarmee het model wordt opgebouwd adequaat de onderliggende realiteit? Het uitwerken van een model op basis van data steunt op twee vooronderstellingen met betrekking tot die data. Primo, dat de kenmerken van de populatie waarop het model gebouwd wordt, dezelfde blijven bij de populatie waarop het model in de toekomst zal worden toegepast. Secundo, dat de initiële gegevens een betrouwbare weergave van deze populatie geven. Wanneer deze vooronderstellingen niet blijken te kloppen, kan het model mislukken en kunnen onbedoelde discriminerende effecten optreden door een model gesteund op een specifieke realiteit op een populatie buiten deze specifieke context toe te passen. Bekijken we achtereenvolgens beide aannames.

32 S. Barocas & A.D. Selbst, 'Big Data's Disparate Impact', *California Law Review* 2016, p. 671-732.

De eerste vooronderstelling houdt geen stand wanneer relevante factoren veranderen ten opzichte van de training data. Denk bijvoorbeeld aan veranderende economische (bijvoorbeeld inkomensniveaus en werkgelegenheidsgraad in tijden van crisis versus betere economische tijden) of andere omstandigheden (bijvoorbeeld veranderingen in de beschikbaarheid van technologie, zoals smartphones). Deze veranderingen worden niet weerspiegeld in het model, waardoor het een meer van de realiteit afwijkende beschrijving vormt. Modellen zijn dus maar zolang houdbaar als de data representatief blijven en het gebruik van verouderde modellen kan leiden tot een bias ten aanzien van bepaalde groepen.

De tweede vooronderstelling kan op verschillende manieren geschonden zijn. Ten eerste kan de genomen steekproef een inaccurate weerspiegeling zijn van de onderliggende werkelijkheid. Ten tweede kunnen de attributen of mogelijke voorspellende variabelen in de dataset weinig geschikt zijn om de populatie of een deel ervan op een relevante wijze te beschrijven.<sup>33</sup>

Bij een inaccurate steekproef kunnen bepaalde delen van de gekozen populatie onder- of oververtegenwoordigd zijn in de gegevens, wat de resultaten van het beslismodel negatief kan beïnvloeden. Dit laatste kan het onbedoelde gevolg zijn van de gekozen wijze van verzamelen van gegevens.

---

Zo wordt bij bepaalde crisisinterventies, bijvoorbeeld bij natuurrampen, gebruikgemaakt van Twitter-data. Hierdoor kan een vertekend beeld ontstaan over een populatie omdat teveel wordt uitgegaan van mensen met internet of smartphones. De kansarme bevolking die niet over deze instrumenten beschikt, komt dan niet in beeld.

Een ander voorbeeld dat in de literatuur terugkomt is dat van *Street Bump* in Boston, een programma dat de wegkwaliteit meet via gegevens over plotselinge veranderingen gesignaleerd door middel van versnellingsmeters in smartphones. Dergelijke slimme initiatieven geven een onvoldoende beeld vanwege de ondervertegenwoordiging van arme buurten waar auto's en smartphones aanwezig zijn.

Soms is een bias het rechtstreekse gevolg van bestaande discriminerende praktijken. Bijvoorbeeld raciale profilering door de politie kan leiden tot een oververtegenwoordiging van bepaalde bevolkingsgroepen in strafzaken en veiligheidsstatistieken.<sup>34</sup>

---

Met het tweede probleem waren we bij de vorige vraag geëindigd. Een slordige selectie van attributen te gebruiken als input-variabelen, ook wel 'feature selectie', kan discriminerend werken voor een groep, met name als deze attributen of kenmerken deze groep niet goed karakteriseren en als factoren die verantwoordelijk zijn voor statistische variatie binnen deze groepen hetzij ontbreken of te grofmazig worden binnengebracht. In beginsel houdt elke representatie door een aantal attributen of kenmerken een reductie in van een veel complexere realiteit. Het is immers onmogelijk om met alle

---

33 T. Calders & I. Žliobaitė, 'Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures', in: B. Custers, T. Calders, B. Schermer & T. Zarsky, *Discrimination and Privacy in the Information Society*, Springer, 2012.

34 *Ibid.*

factoren rekening te houden. Het terugbrengen van een verschijnsel tot een reeks kenmerken, zelfs een heel uitgebreide, is dus altijd een vorm van stereotypering. Een voorstelling kan daarom falen in het onderscheidingsvermogen en niet voldoende in detail gaan bij het onderscheiden van uitkomsten met betrekking tot bepaalde groepen. Toegeschreven kenmerken zijn dan niet meer dan ‘*coarse proxies*’ of grove benaderingen van bepaalde fenomenen en kunnen discriminatie tevens in de hand werken.

---

De controversiële praktijk in de Verenigde Staten van *redlining* is gebaseerd op een dergelijke grove *proxy*. Financiële instellingen gebruiken postcodes of geografische criteria als een gemakkelijk beschikbare indicatie voor kredietwaardigheid en weigeren leningen of andere diensten aan mensen in bepaalde wijken. De postcode wordt dan een indicator over het inkomen of het vermogen om leningen terug te betalen. Het is een goedkope, maar zeer grove indicator, zeker in vergelijking met een individueel zorgvuldig onderzoek naar de kredietwaardigheid van een persoon.

---

In de literatuur worden nog meer voorbeelden gegeven van discriminerende praktijken gebaseerd op het gebruik van dergelijke *proxies*.<sup>35</sup> Dit probleem wordt versterkt door het feit dat attributen waarmee personen gerepresenteerd worden gecorreleerd kunnen zijn. Bij zulke gecorreleerde attributen is het veel moeilijker na te gaan welke bijdragen aan het resulterende model en op welke wijze. In deze situatie is het moeilijker om na te gaan of er sprake is van ontbrekende of onvolledige data, waardoor informatie over bepaalde groepen op een gebrekkige manier in de data aanwezig kan zijn.<sup>36</sup>

### 6.7.3 *Sluipweg 3: de data als representatie van ongewenste informatie*

Weerspiegelen de data informatie die niet als basis voor besluitvorming gebruikt mag worden en daarom niet in het model mag worden opgenomen? Vroegere discriminerende praktijken waarvan we nu afstand nemen of vooroordelen die aanwezig blijven in de dagelijkse praktijk kunnen weerspiegeld worden in de gebruikte data.

Hierboven werd al het probleem aangehaald van etnische profilering, die zorgt voor een over-representatie van bepaalde bevolkingsgroepen in data met betrekking tot criminaliteit. Maar subjectieve bias kan ook ingang vinden via de *labeling* van data (het benoemen van de uitkomsten). In voorkomend geval schuilt het probleem niet in de methode van dataverzameling, maar schuilt de discriminatie in de gebruikte labels (de uitkomsten voor de doelvariabele). Er is geen probleem wanneer de labels een weergave zijn van objectieve metingen en waarnemingen. Zelfs wanneer er een bias is in de verzameling van de gegevens, dan nog kunnen de labels correct zijn. Bijvoorbeeld de uitkomst van een alcoholtest, ook al werd die uitgevoerd op grond van etnische profilering bij het verzamelen van gegevens. Maar wanneer het plakken van labels op een uitkomst steunt op een subjectieve menselijke beoordeling, is het duidelijk dat vooroordelen

---

35 S. Barocas & A.D. Selbst, ‘Big Data’s Disparate Impact’, *California Law Review* 2016, p. 671-732.

36 T. Calders & I. Žliobaitė, ‘Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures’, in: B. Custers, T. Calders, B. Schermer & T. Zarsky, *Discrimination and Privacy in the Information Society*, Springer, 2012.



kunnen binnensluipen. Zo zullen gegevens over aanstellingen en toelatingen uit het verleden ook iets vertellen over de toentertijd gehanteerde vooroordelen van hen die betrokken waren bij deze procedures.

Het onderscheid tussen verschillende bronnen van partijdigheid of bias is belangrijk omdat elk een andere aanpak vereist om ze te corrigeren. Over- of ondervertegenwoordiging van bepaalde populaties kan worden gecorrigeerd door middel van de invoering van wegingsfactoren om de uitkomst te corrigeren, terwijl vooroordelen binnengeslopen in de labels zelf minder eenvoudig vast te stellen en te corrigeren zijn.<sup>37</sup>

## 6.8 Technische maatregelen en ‘non-discrimination by design’

Het juridisch kader legt de big data-gebruiker op om ervoor zorg te dragen dat zijn big data-toepassing niet leidt tot een discriminerende behandeling en, bij een vermoeden van discriminatie, om dit voldoende inzichtelijk en controleerbaar te maken. Het is dus aan te bevelen om hieraan reeds bij het ontwerp van zulk een systeem aandacht te besteden en de nodige maatregelen te treffen om eventuele discriminatie te kunnen opsporen en te kunnen voorkomen. Deze behoefte heeft de laatste jaren geleid tot vernieuwend onderzoek naar technische maatregelen daarvoor. Verscheidene statistische testmethodes zijn ontwikkeld om na te gaan of een discriminerende bias aanwezig is in oude datasets of op basis daarvan ontwikkelde modellen. Eveneens is een reeks methodes ontwikkeld om discriminatie te voorkomen of tegen te gaan bij data mining en de daarmee ontwikkelde modellen. Pre-processing-technieken verwijderen of compenseren de discriminatie in de training dataset, waarna die gebruikt kan worden met standaard data mining-technieken. In-processing technieken passen de data mining-algoritmes aan opdat ze de aanwezige bias compenseren. Post-processing technieken corrigeren de ontwikkelde modellen om de discriminerende bias te verwijderen.<sup>38</sup> Om dat te kunnen doen, zullen de technieken vaak gebruik moeten maken van gegevens over de ‘verboden’ gronden, zoals ras en godsdienst.<sup>39</sup> In deze fase van de data-analyse kan het gebruik van gegevens over die kenmerken dus juist bijdragen aan het voorkómen van discriminatie.

Sommige van deze technieken kunnen afgesteld worden, zodat ze niet alle verschillen verwijderen, wat neerkomt op positieve discriminatie, maar enkel de verschillen die niet verklaarbaar zijn op basis van bepaalde objectieve factoren. Bijvoorbeeld inkomensverschillen kunnen voortkomen uit discriminatie tussen bepaalde groepen, maar ook uit verschillen in scholingsgraad en diploma’s. Die verschillen in scholing kunnen ook een gevolg zijn van discriminatie in het onderwijssysteem, maar een werkgever die rekening houdt met diploma’s zal doorgaans geen ontoelaatbare ongelijke behandeling kunnen worden verweten, zelfs niet als dit tot een ondervertegenwoordiging van bepaalde groepen leidt. In dit geval is het verschil verantwoord door een legitiem doel,

<sup>37</sup> *Ibid.*

<sup>38</sup> Voor meer informatie over deze technieken, zie onder meer B. Custers, T. Calders, B. Schermer & T. Zarsky (eds.), *Discrimination and privacy in the information society: Data mining and profiling in large databases*, Springer Science & Business Media, 2012

<sup>39</sup> I. Zliobaité & B. Custers, ‘Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models’, *Artificial Intelligence Law* 2016/24, p. 183-201.

met name het werven van door scholing voldoende bekwaam personeel, en is er geen sprake van laakbare indirecte discriminatie.

Dit soort technische maatregelen is nog volop in ontwikkeling en staat, ook al zijn er al bruikbare methoden voorhanden, nog in de kinderschoenen. Dit valt te vergelijken met de situatie toen de eerste *privacy by design*-principes werden geformuleerd en enkel een reeks *privacy-enhancing techniques* (PET) beschikbaar waren. Ondertussen is *privacy by design* verder ontwikkeld tot uitgewerkte ontwerpmethodologieën met precieze technische doelstellingen, ontwerppatronen en -methodologieën, auditing- en assessment-procedures. In de komende jaren kan *non-discrimination by design* zich ook op deze manier ontwikkelen en gebruikmaken van de *lessons learned* met *privacy by design*. Anderzijds, een basis gereedschapskist aan methodes om discriminerende bias op te sporen en te voorkomen is voorhanden. De gebruiker en ontwerper van big data-toepassingen heeft, gezien zijn juridische verplichtingen, alle redenen om hiernaar te kijken.

### 6.9 Discriminatie en het gegevensbeschermingsrecht

Het voorgaande toont aan dat de regels op het gebied van gelijke behandeling een eigen rol hebben bij de regulering van big data-praktijken. Daarnaast kunnen de regels rond gelijke behandeling invulling geven aan de algemene principes van het gegevensbeschermingsrecht (zie over het gegevensbeschermingsrecht: hoofdstuk 3 van dit boek). De Algemene Verordening Gegevensbescherming bepaalt bijvoorbeeld dat persoonsgegevens enkel mogen worden verzameld voor “gerechtvaardigde doeleinden” en moeten worden verwerkt op een wijze die ten aanzien van de betrokkene “rechtmatig” en “behoorlijk” is (artikel 5 lid 1 sub a en b van Verordening 2016/679, hierna: AVG). Dit impliceert dat persoonsgegevens niet mogen worden verzameld of worden gebruikt om te discrimineren en dat de verwerking niet mag leiden tot een discriminerende behandeling. Een ander voorbeeld is dat de Verordening bepaalt dat er passende beveiligingsmaatregelen moeten worden genomen, rekening houdend met onder meer “risico’s voor de rechten en vrijheden van personen”. Overweging 75 maakt duidelijk dat daarbij met name moet worden gedacht aan de situatie dat de verwerking kan leiden tot discriminatie. Ook artikel 22 AVG verplicht de verantwoordelijke om bij geautomatiseerde besluitvorming met betrekking tot individuen, wanneer die toegelaten is, passende maatregelen ter bescherming van de rechten en vrijheden en gerechtvaardigde belangen van de betrokkene te nemen. In al deze gevallen kunnen de regels op het gebied van gelijke behandeling verduidelijken wanneer er sprake is van discriminatie en daarmee wanneer de abstracte principes van de Verordening zijn geschonden. Die regels maken bijvoorbeeld duidelijk dat het risico op discriminatie niet is beperkt tot de categorie gegevens die de AVG vanwege het potentieel discriminatoire karakter ervan als “bijzonder” aanmerkt. Zo vallen bepaalde gronden voor onderscheid die de wetgever in het kader van gelijke behandeling als ‘verdacht’ aanmerkt, zoals geslacht en burgerlijke staat, niet in de categorie bijzondere persoonsgegevens. Bovendien kan ook worden gediscrimineerd zonder bijzondere persoonsgegevens te verwerken. Van indirecte discriminatie is juist sprake als de beslissing om een persoon anders dan anderen te behandelen niet is gebaseerd op een bijzonder gegeven als ras of godsdienst, maar op

beleid dat tot gevolg heeft dat personen met een bepaald ras of een bepaalde godsdienst wel worden achtergesteld.

Omgekeerd kunnen de regels van het gegevensbeschermingsrecht een rol spelen bij de implementatie en handhaving van het non-discriminatie recht. Zoals opgemerkt in hoofdstuk 3 kan het gegevensbeschermingsrecht worden gezien als ‘voorportaal’ voor de bescherming van het non-discriminatie recht, omdat bij big data-toepassingen de verwerking van persoonsgegevens voorafgaat aan besluitvorming die kan leiden tot discriminatie. De middelen die kunnen worden ingezet ter handhaving van het gegevensbeschermingsrecht, zoals de bevoegdheden van de Autoriteit Persoonsgegevens, kunnen aldus ten goede komen aan de naleving van het non-discriminatie recht. Daarnaast voorziet het gegevensbeschermingsrecht in transparantie verplichtingen inzake de verwerking van persoonsgegevens die nuttig kunnen zijn bij het ontdekken van discriminerende big data-praktijken. Zo heeft bij profilering de betrokkene het recht op “nuttige informatie over de onderliggende logica” (artikel 13 lid 2 sub f, 14 lid 2 sub g en 15 lid 1 sub h AVG). Dat vereist op zijn minst een duidelijke uiteenzetting van de criteria waarop de beslissingen zijn gebaseerd.<sup>40</sup> Dit ondersteunt de conclusie dat een onderneming die big data gebruikt de werking en logica van een big data-systeem inzichtelijk moet kunnen maken en zal moeten kunnen aantonen dat dit niet leidt tot een discriminerende behandeling.

#### 6.10 **Besluit: big data vereist een zorgvuldige en transparante controle op onbedoeld insluipen van discriminatie**

Het gebruik van data mining en big data kan leiden tot meer geobjectiverde besluitvorming en behandeling van personen, maar kan ook op verschillende manieren leiden tot discriminatie, zelfs onbedoeld. In deze bijdrage bespreken we drie sluiptwegen waardoor bias kan binnendringen in besluitvorming op basis van data mining en aanleiding kan geven tot onbedoelde ongelijke behandeling. Ten eerste via het model als beschrijving van de realiteit, waarbij zowel de te verklaren variabelen als de predictoren kunnen falen als representatie van bepaalde groepen. Dit kan aanleiding geven tot bias ten aanzien van die groepen. Ten tweede via de training data die gebruikt worden om het model te bouwen, als potentieel gebrekkige weergave van de realiteit. Die data kunnen falen als weergave van de (relevante) realiteit voor bepaalde groepen. Externe veranderingen kunnen maken dat de dataset een verouderde weergave is. Een foutieve steekproef kan aanleiding geven tot over- of onderrepresentatie van bepaalde groepen of de gebruikte categorieën kunnen te grofmazig en weinig representatief zijn. Ten derde kan de dataset ook elementen weergeven die niet in het model opgenomen mogen worden. Sommige predictoren kunnen gecorreleerd zijn met beschermden gronden of de dataset kan een weerspiegeling zijn van vooroordelen en historische discriminatie waarvan juist afstand moet worden genomen. Het gebruik van big data brengt daarom een risico op onbedoelde discriminatie met zich mee.

<sup>40</sup> See Article 29 Data Protection Working Party, ‘Opinion 03/2013 on purpose limitation’, WP 203, 2 april 2013, p. 47.

Het gelijkheidsbeginsel vereist dat ongelijke behandeling objectief te verantwoorden is. Dit geldt *a fortiori* voor de beschermde gronden uit de discriminatiewetgeving. De big data-operator moet aantonen dat er sprake is van een legitiem doel en dat de middelen, het model met zijn variabelen, voor het bereiken van dat doel passend en noodzakelijk zijn. Rationele discriminatie, waarbij ongelijke behandeling verantwoord wordt op basis van andere doelstellingen, op basis van geslacht, ras en etnische afkomst is zeer moeilijk te verantwoorden. In het algemeen is het verstandig deze gronden te weren als expliciete criteria voor besluitvorming. Onduidelijkheid over hoe de big data-toepassing tot bepaalde beslissingen komt, neemt de verantwoordelijkheid daarvoor niet weg en kan leiden tot aansprakelijkheid voor discriminatie. Daarom is het aangegeven om reeds bij aanvang van het ontwerp aandacht te hebben voor mogelijke discriminatie en de nodige technische maatregelen te nemen om dit na te gaan en te voorkomen.