

# De Juridische Voorspelindustrie: onzinnige hype of nuttige ontwikkeling?

Floris Bex\* & Henry Prakken\*

*De laatste tijd wordt veelvuldig gediscussieerd over het gebruik van voorspelalgoritmes in het recht - er wordt zelfs beweerd dat de invoering van 'de robotrechter' een kwestie van tijd is. In dit artikel bespreken we of, en zo ja hoe, voorspelalgoritmes van nut kunnen zijn voor de rechtsgemeenschap, in het bijzonder voor de rechtspraak.*

## 1. Inleiding

De juridische 'voorspelindustrie' rukt op: onder de noemer 'Artificiële Intelligentie (AI) in de Rechtspraak'<sup>1</sup> wordt veelvuldig gediscussieerd over het gebruik van AI-algoritmes voor het voorspellen van uitspraken in rechtszaken.<sup>2</sup> Ondanks de ronkende teksten in de media over de mogelijkheden die deze algoritmen bieden met betrekking tot robotrechters<sup>3</sup> en 'rechtbankshoppen',<sup>4</sup> zijn er ook sceptische geluiden te horen vanuit de rechtswetenschap.<sup>5</sup> Een veelgenoemd bezwaar is dat het *voorspellen* van een beslissing in een rechtszaak met behulp van statistische correlaties niet hetzelfde is als het *nemen* van deze beslissing op basis van redengevende verbanden. Verder moeten de gerapporteerde resultaten van de algoritmen ook niet overschat worden. Als voorbeeld: ook zonder AI-algoritme kan zeer nauwkeurig voorspeld worden dat een willekeurige strafzaak tot een schuldigbevinding zal leiden, simpelweg omdat meer dan 90% van de strafzaken die voor de rechter komen tot schuldigbevinding leiden. Zijn voorspelalgoritmes een onzinnige en tijdelijke hype of kunnen zij wel degelijk van nut zijn voor de rechtsgemeenschap? In dit artikel trachten wij een antwoord op deze vraag te geven.

---

\* Prof.dr. F.J. Bex is bijzonder hoogleraar data science en de rechtspraak aan het Department of Law, Technology, Markets, and Society, Tilburg University, wetenschappelijk directeur van het Nationaal Politielab AI bij het Innovation Centre for AI (ICAI) en universitair docent AI bij het departement Informatica, Universiteit Utrecht. Prof.dr.mr. H. Prakken is hoogleraar rechtsinformatica en juridische argumentatie aan de Faculteit der Rechtsgeleerdheid, Rijksuniversiteit Groningen en universitair hoofddocent AI bij het departement Informatica, Universiteit Utrecht.

<sup>1</sup> C. Prins & J. van der Roest, 'AI en de rechtspraak', *NJB* 2018/206; H. Prakken, 'Komt de robotrechter er aan?', *NJB* 2018/207; S. Verberk, M. Noordergraaf & C.E. du Perron (red.) 'Algoritmes in de rechtspraak. Wat artificiële intelligentie kan betekenen voor de rechtspraak', *Rechtstreeks* 2019, afl. 2.

<sup>2</sup> Publicaties over algoritmen die gerechtelijke uitspraken voorspellen: D.M. Katz, M.J. Bommarito & J. Blackman, 'A general approach for predicting the behavior of the Supreme Court of the United States', *PLoS ONE* 2017, afl. 4; N. Aletras e.a., 'Predicting judicial decisions of the European Court of Human Rights', *PeerJ Computer Science* 2016-2; M. Medvedeva, M. Vols & M. Wieling, 'Using machine learning to predict decisions of the European Court of Human Rights', *Artificial Intelligence and Law* 2019, doi.org/10.1007/s10506-019-09255-y. Publicaties die deze voorspelalgoritmen (kritisch) bespreken: F. Pasquale & G. Cashwell, 'Prediction, persuasion, and the jurisprudence of behaviourism' *University of Toronto Law Journal* 2018-68.supplement 1, p. 63-81; K.D. Ashley, 'A Brief History of the Changing Roles of Case Prediction in AI and Law', *Law in Context* 2019, afl. 1, p. 93-112; D.L. Chen, 'Machine Learning and Rule of Law', in: M.A. Livermore & D.N. Rockmore (red.), *Law as Data*, Santa Fe: Santa Fe Institute Press 2018, p. 429-438.

<sup>3</sup> F. Jensma, 'Big data kunnen de rechter verdringen', *NRC* 28 oktober 2017; H.J. van den Herik, 'In 2030 zullen computers rechtspreken', *Mr. Online* 31 oktober 2016, www.mr-online.nl/in-2030-zullen-computers-rechtspreken/.

<sup>4</sup> C. Driessen, 'Wie personeel wil lozen na een ruzie moet bij rechtbank Den Haag zijn', *NRC* 11 september 2019; E. Kreulen, 'De rechtspraak is met deze nieuwe robot niet langer digibeeft', *Trouw* 26 augustus 2019.

<sup>5</sup> Prakken 2018, Pasquale & Cashwell 2018.

Eerst bespreken we verschillende typen juridische voorspelalgoritmes en het onderscheid tussen ‘algoritmische deskundigen’ en ‘algoritmische uitkomstvoorspellers’. Dan bespreken we enkele belangrijke aandachtspunten bij het bepalen van de kwaliteit van voorspelalgoritmes – bijvoorbeeld verschillende manieren om de kwaliteit te evalueren en voorwaarden waaraan de data moet voldoen om tot een geldige voorspelling te komen. Vervolgens komen we toe aan onze hoofdvraag, hoe de verschillende typen voorspelalgoritmes nuttig kunnen zijn voor de rechtswetenschap, rechtzoekenden en de rechtspraak. We zullen betogen dat er zeker sprake is van een hype rond voorspelalgoritmen, maar dat dergelijke AI-algoritmen niet als onzinnig afgedaan moeten worden, aangezien zij op verschillende manieren van nut kunnen zijn voor het recht. Zogenaamde *algoritmische uitkomstvoorspellers* kunnen helpen bij kwantitatieve analyses van het recht en de rechtspraak, en zogenaamde *algoritmische deskundigen* kunnen rechters ondersteunen bij het beantwoorden van bepaalde feitelijke vragen in een rechtsgeding.

## 2. Verschillende juridische voorspelalgoritmes

Voorspelalgoritmes zijn veelal typische voorbeelden van *supervised machine-learning*-algoritmes. Zo’n algoritme krijgt eerst een grote hoeveelheid historische gevallen - de trainingsdata - te zien, met de kenmerken en de uitkomst van die gevallen (bijvoorbeeld bij het voorspellen van recidive of de persoon recidive gepleegd heeft, of bij het voorspellen van rechterlijke uitspraken wat de uitspraak was). Uit deze trainingsdata kan het algoritme de mogelijk zeer complexe relaties tussen deze kenmerken en de mogelijke uitkomsten leren, om daarmee voor nog ongeziene gevallen de uitkomst te ‘voorspellen’.<sup>6</sup> Een eenmaal getraind algoritme wordt geëvalueerd aan de hand van testdata: historische gevallen waarvan alleen de kenmerken (en dus niet de uitkomst, welke wel bekend is) aan het algoritme getoond worden.

### 2.1 Algoritmische deskundigen: voorspellen van voor een beslissing relevante feiten

Sommige juridische voorspelalgoritmes maken inschattingen op punten die relevant zijn voor een rechterlijke beslissing, welke anders door de rechter zelf of door een menselijke deskundige moeten worden gemaakt. Bekende voorbeelden zijn algoritmes die de kans op recidive voorspellen,<sup>7</sup> en algoritmes die de verwachte milieu-impact schatten van activiteiten waarvoor een milieuvergunning wordt aangevraagd, zoals het AERIUS-systeem dat stikstofuitstoot schat.<sup>8</sup> We zullen dit soort systemen *algoritmische deskundigen* noemen.

### 2.2 Algoritmische uitkomstvoorspellers: voorspellen van rechterlijke uitspraken

Ook zijn er *algoritmische uitkomstvoorspellers*, algoritmes die uitkomsten van rechtszaken voorspellen. Er zijn grofweg drie typen: voorspellers op basis van niet-inhoudelijke kenmerken van de zaak, voorspellers op basis van de tekstuele beschrijving van de zaak en voorspellers op basis van juridisch relevante factoren.

*Voorspellen op basis van niet-inhoudelijke kenmerken van de zaak*

Sommige algoritmes doen hun voorspellingen op basis van niet-inhoudelijke kenmerken van de zaak. Een voorbeeld daarvan is het algoritme dat uitspraken van het Amerikaanse

---

<sup>6</sup> ‘Voorspellen’ staat hier tussen aanhalingstekens aangezien de zaken waarvoor het algoritme de uitspraak voorspelt ook zaken uit het verleden kunnen zijn waarvan we al weten wat de uitspraak was.

<sup>7</sup> Zie bijvoorbeeld R. Berk e.a., ‘Fairness in criminal justice risk assessments: the state of the art’, *Sociological Methods & Research* 2018, <https://doi.org/10.1177/0049124118782533>.

<sup>8</sup> Aeries, Rekeninstrument voor de leefomgeving, [www.aerius.nl/nl](http://www.aerius.nl/nl).

hoogrechtshof voorspelt<sup>9</sup> op basis van gegevens die beschikbaar zijn in een database<sup>10</sup> over het hof, zoals het soort zaak, de datum waarop de zaak diende en bij welke lagere rechtbank de oorspronkelijke zaak diende. Dit algoritme, dat 70% van de uitspraken correct voorspelde, kan de voorspelde uitkomsten niet op een juridisch zinvolle manier uitleggen, want de kenmerken op basis waarvan het de uitkomst voorspelt zijn niet gerelateerd aan de inhoudelijke merites van de zaak: een – sterk versimpeld – voorbeeld van een uitleg van het algoritme is ‘ik voorspel dat het hof in deze zaak het oordeel van de lagere rechtbank bevestigt, want dat doet het meestal in economische zaken waar dhr. Roberts de *chief justice* is’.

#### *Voorspellen op basis van de tekstuele beschrijving van de zaak*

Andere algoritmes voorspellen uitkomsten met behulp van een statistische analyse van de tekst van jurisprudentie, waarbij statistische verbanden gelegd worden tussen de frequentie van woordcombinaties en de uitkomst van een zaak. Een voorbeeld is het algoritme dat voorspelt of het Europese Hof voor de Rechten van de Mens (EHRM) voor een specifiek artikel uit het gelijknamige Verdrag zal beslissen of dat artikel geschonden is<sup>11</sup> op basis van beschrijvingen van het EHRM van de procedurele voorgeschiedenis en de feiten die aanleiding waren om de zaak bij het Hof aan te brengen.<sup>12</sup> Het algoritme voorspelde 75% van de beslissingen correct. Hoewel het lijkt alsof het algoritme inhoudelijk naar de zaak kijkt (voorgeschiedenis, feiten), zeggen de gevonden statistische verbanden niets over de juridisch relevante gronden voor de uitkomst van een zaak. Daarom kan ook dit algoritme de voorspelde uitkomsten niet op een juridisch zinvolle manier uitleggen. Zo waren de drie woordcombinaties met de hoogste voorspellende waarde voor ‘schending’ respectievelijk ‘district prosecution office’, ‘the district prosecutor’ en ‘the first applicant’. Dat is juridisch niet erg informatief.

#### *Voorspellen op basis van juridisch relevante factoren*

Een derde aanpak voorspelt uitkomsten op basis van de juridisch relevante factoren in een zaak. Bekend is het onderzoek van Ashley en collega’s naar jurisprudentie over misbruik van bedrijfsgeheimen in het Amerikaanse recht<sup>13</sup>. Zowel voor eerder besliste zaken als voor te voorspellen zaken worden van tevoren handmatig de juridisch relevante factoren aangegeven. Bijvoorbeeld of de gebruikte informatie concurrentievoordeel opleverde, of er een ‘non-disclosure agreement’ getekend was, of het product ‘reverse-engineerbaar’ was, d.w.z. of het nagemaakt kon worden op basis van openbare informatie, en of de informatie in onderhandelingen geopenbaard was. Vervolgens worden deze factoren gerelateerd aan uitspraken in de eerder besliste zaken. Het leggen van deze relaties kan handmatig gebeuren, maar ook met een *machine-learning* algoritme dat de (statistische) relaties tussen factoren en

---

<sup>9</sup> Katz e.a. 2017. Soortgelijke commerciële voorspelalgoritmes zijn in de Verenigde Staten ontwikkeld voor zaken bij lagere rechters op basis van data over bijvoorbeeld de rechters, de advocaten en de procespartijen, bijvoorbeeld Lex Machina (<https://lexmachina.com>), LexPredict ([www.lexpredict.com](http://www.lexpredict.com)) en Premonition (<https://premonition.ai/>). In Nederland worden ook dergelijke data verzameld om mogelijke voorspelalgoritmes op te trainen, zie artikelen in *NRC Handelsblad* en *Trouw*, noot 4.

<sup>10</sup> The Supreme Court Database, <http://scdb.wustl.edu/>.

<sup>11</sup> Medvedeva e.a. 2019.

<sup>12</sup> De teksten van de zaken zijn beschikbaar in de HUDOC-database, <https://hudoc.echr.coe.int/>.

<sup>13</sup> K.D. Ashley & S. Brüninghaus, ‘Automatically classifying case texts and predicting outcomes’, *Artificial Intelligence and Law* 2009, afl. 17, p. 125-165. Saillant detail is overigens dat wanneer de relaties tussen factoren en uitspraken handmatig door de mens werden ingevoerd de nauwkeurigheid van een voorspelling het hoogst was, namelijk 91%.

uitspraken automatisch leert. Een dergelijk algoritme kan een uitkomst dan inhoudelijk uitleggen op een manier die voor juristen vertrouwd is, omdat de voorspelling gedaan is op basis van juridisch relevante factoren. Bijvoorbeeld: ‘ik voorspel dat de eiser in deze zaak gelijk krijgt, aangezien in eerdere zaken waar de informatie concurrentievoordeel opleverde ook de eiser gelijk kreeg, hoewel het product in beide zaken reverse-engineerbaar was’. Ook kan als verdere uitleg naar soortgelijke precedentes met dezelfde uitkomst verwezen worden. Dergelijke algoritmes wisten tussen de 82% en 88% van de uitspraken in zaken over misbruik van bedrijfsgeheimen correct te voorspellen.

Een groot nadeel van deze benadering is dat het handmatig aangeven van de juridisch relevante factoren enorm arbeidsintensief is en bovendien een vorm van juridische interpretatie is. Er wordt gewerkt aan taalanalyse-algoritmes voor het automatisch herkennen van factoren in teksten, maar dit onderzoek staat nog in de kinderschoenen.<sup>14</sup>

Samengevat hebben algoritmische uitkomstvoorspellers nogal wat beperkingen als het gaat om toepassing in de rechtspraak. Of ze kunnen hun voorspellingen niet uitleggen, of ze vereisen een substantiële menselijke voorbewerking van de data waarmee ze werken. En in beide gevallen is hun voorspellende kracht vooralsnog bescheiden, zoals we hieronder nader uitleggen.

### 3. Algemene aandachtspunten bij voorspelalgoritmes

Hoe goed zijn juridische voorspelalgoritmes tegenwoordig in het voorspellen van rechtszaken? Een aantal bekende kwesties uit de statistiek en *data science* maakt een antwoord op deze vraag niet eenvoudig: wat zijn de criteria voor het meten van de kwaliteit van voorspellingen en hoe staat het met de kwaliteit en beschikbaarheid van de data waarmee juridische voorspelalgoritmes werken? Toch kan gezegd worden dat de resultaten vooralsnog niet spectaculair zijn.

#### *Hoe beoordelen we de kwaliteit van voorspellingen?*

Als eerste rijst de vraag welke beoordelingsmetriek we het beste kunnen gebruiken om de kwaliteit van voorspellingen te meten. Vaak wordt de *accuracy* (nauwkeurigheid) genoemd: hoeveel procent van de zaken uit de testset voorspelt het algoritme correct? Deze nauwkeurighedsmaat is betrekkelijk, aangezien in de realiteit de verdeling van de uitspraken vaak scheef is: het EHRM beslist in gemiddeld 76% van de zaken dat het artikel geschonden is. Als we dus altijd ‘schending’ gokken hebben we een nauwkeurigheid van 76%, hoger dan de 75% van het getrainde voorspelalgoritme.<sup>15</sup> Verder beantwoorden de hierboven besproken algoritmes een ja/nee-vraag, dus zelfs met het opgooien van een muntje scoort men al 50%. Tenslotte moeten we oppassen met de interpretatie van het begrip nauwkeurigheid. Een nauwkeurigheid van bijvoorbeeld 80% betekent niet dat de kans dat een willekeurige rechter in een individuele zaak dezelfde beslissing zou nemen 80% is.

Er zijn veel (statistische) metrieken voor algoritmen die wel met scheve verdelingen kunnen werken, of waarmee kansen voor een individueel geval bepaald kunnen worden. Het lastige is

---

<sup>14</sup> Zie hoofdstuk 10 in K.D. Ashley, *Artificial Intelligence and Legal Analytics. New Tools for Law Practice in the Digital Age*, Cambridge: Cambridge University Press 2017 en M. Schraagen e.a., ‘Argumentation-driven information extraction for online crime reports’, *CKIM 2018 International Workshop on Legal Data Analysis and Mining (LeDAM 2018)*, *CEUR Workshop Proceedings* 2019.

<sup>15</sup> Dit betekent overigens niet dat voorspelalgoritmes bij zulke scheve verdelingen zinloos zijn, of niets geleerd hebben – een ‘altijd schending’ gok zal immers nooit een ‘niet-schending’ zaak herkennen waar het algoritme dat wel doet.

dat hun toepassing en interpretatie niet triviaal is, zeker niet voor juristen. Vaak hangt de kwaliteit van een algoritme ook af van waar het algoritme voor gebruikt wordt en hoe erg een bepaalde fout is: het niet herkennen van een tumor is erger dan het missen van een spambericht.<sup>16</sup> Ook is relevant hoe goed mensen dezelfde taak uitvoeren. Een imperfect algoritme kan nog steeds nuttig zijn als het beter presteert dan mensen bij dezelfde taak.

*Welke data is er voorhanden om het algoritme te trainen?*

De kwaliteit van voorspellingen hangt onder meer af van de kwaliteit en de beschikbaarheid van de data. Bij het voorspellen van rechterlijke uitspraken is een probleem dat vaak maar een klein deel van de uitspraken publiek beschikbaar is.<sup>17</sup> Ook kan men zich afvragen of er wel echt van voorspellen sprake is als het algoritme op de tekst van een te voorspellen uitspraak toegepast wordt, aangezien een uitspraak ná de beslissing geschreven wordt en de rechter dus in de uitspraak deels ‘naar de beslissing toe schrijft’.<sup>18</sup> Beter zou zijn om te voorspellen op basis van het zaaksdossier dat de rechter onder ogen krijgt, alleen zijn die dossiers al helemaal niet in groten getale voorhanden.

Een ander bekend probleem uit de statistiek is dat van *overfitting* van een algoritme op de data, waar het algoritme te zeer toegespitst is op specifieke elementen uit de data. Als er bijvoorbeeld in de trainingset veel zaken zitten over een bepaald land uit een bepaalde periode, en dat land heeft tijdens die periode met een opstand of burgeroorlog te kampen gehad, dan kan een algoritme onterecht concluderen dat zaken tegen dat land een hoge kans op succes hebben.

Ook is een algoritme dat leert van zaken uit het verleden niet altijd goed generaliseerbaar naar de toekomst, omdat de typen zaken maar ook juridische, ethische en maatschappelijke opvattingen veranderen. Dit is goed te zien in de experimenten met het EHRM-algoritme:<sup>19</sup> als het algoritme alleen getraind werd op zaken die chronologisch gezien vóór de te voorspellen uitspraak vielen, daalde de nauwkeurigheid van het algoritme tussen de 58% en 68%, afhankelijk van hoe recent de trainingszaken waren.

#### **4 Mogelijk nut van uitkomstvoorspellingen**

We hebben gezien dat de huidige algoritmische uitkomstvoorspellers nog lang niet perfect voorspellen en dat hun kwaliteit niet eenvoudig te bepalen of te garanderen is. Bovendien hebben we gezien dat ze óf hun voorspellingen niet in juridisch zinvolle termen kunnen uitleggen óf dat ze een substantiële handmatige voorbewerking van de data vereisen. Kunnen deze algoritmes toch nuttig zijn? We bespreken dit voor drie doelgroepen, de rechtswetenschap, rechtzoekenden en de rechtspraak, waarbij onze voornaamste aandacht voor de rechtspraak zal zijn.

---

<sup>16</sup> Zie voor een dergelijke discussie over beoordelingsmetrieken binnen het juridische domein J. Bijlsma, F.J. Bex & G. Meynen, ‘Artificiële intelligentie en risicotaxatie: drie kernvragen voor strafrechtjuristen’, *NJB* 2019/2778.

<sup>17</sup> In 2018 verscheen slechts 4,6 procent van alle uitspraken op rechtspraak.nl, zie *NRC Handelsblad*, noot 4.

<sup>18</sup> Medvedeva e.a. 2019 gebruiken, in tegenstelling tot Aletras e.a. 2016, alleen informatie uit de te voorspellen uitspraak die in principe vóór de beslissing bekend was. Echter, de zaaksbeschrijvingen in de te voorspellen uitspraak zijn niet geheel gelijk aan de beschrijvingen die enkele jaren voor de uitspraak naar de procespartijen gestuurd zijn.

<sup>19</sup> Medvedeva e.a. 2019.

## 4.1 De rechtswetenschap

Algoritmes en statistieken kunnen rechtswetenschappers inzicht geven in wat rechterlijke uitspraken beïnvloedt, en hoe deze uitspraken veranderen over de jaren.<sup>20</sup> Zo kan het algoritme dat uitspraken van het Amerikaanse Hoogerechtshof voorspelt helpen om inzicht te krijgen in de invloed van politieke voorkeuren van rechters of de presidenten die hen benoemd hebben. Ook ongewenste trends of invloeden kunnen zo ontdekt worden, zoals dat bepaalde rechtbanken strenger oordelen in vergelijkbare zaken, of de invloed van het tijdstip van beraadslaging (vlak voor de lunch of niet) op hoe streng rechters oordelen.<sup>21</sup> De angst kan bestaan dat dit het vertrouwen in de rechtspraak kan ondermijnen, omdat het laat zien dat rechters ook onderhevig zijn aan typisch menselijke vooroordelen. Ons inziens kan dergelijk onderzoek de kwaliteit van de rechtspraak juist verbeteren. Wel is het dan belangrijk verder te kijken dan de soms tendentieuze berichtgeving in de media.<sup>22</sup>

## 4.2 Rechtzoekenden

Voor rechtzoekenden kan informatie over de kans op succes of de verwachte hoogte van een schadevergoeding nuttig zijn voor bijvoorbeeld de beslissing om te gaan procederen of om op een schikkingsvoorstel in te gaan. Voor beslissingen over procederen hoeft de voorspelling niet perfect te zijn; een bedrijf dat regelmatig procedeert kan al financieel voordeel hebben als het algoritme iets accurater is dan een menselijke medewerker die dezelfde inschatting maakt. Om dezelfde reden hoeft zo'n voorspelling niet op inhoudelijke gronden gebaseerd te zijn – een rechtzoekende zal echter baat hebben bij een algoritme dat een (juridische) uitleg geeft van de beslissing, aangezien ze die ook in de zaak kan aanvoeren.<sup>23</sup> Als nadeel van statistieken wordt wel het 'rechtbankshoppen' genoemd: de rechtzoekende maakt haar zaak aanhangig bij de rechtbank waar ze het meeste kans maakt. Echter, 'rechtbankshopping' is niet meer dan wat advocaten van oudsher al voor hun cliënten doen, namelijk inschatten waar deze de meeste kans maken. En in het kader van de rechtsgelijkheid is het ook zaak dat de rechtspraak, eventueel gebruikmakend van dezelfde statistieken, zorgt dat er zo weinig mogelijk verschil is tussen rechtbanken.

## 4.3 De rechtspraak

Kunnen algoritmische uitkomstvoorspellers in de toekomst algoritmische rechters worden? Sommigen denken van wel. Hierbij wordt wel eens verwezen naar het medische domein, waarin breed geaccepteerd is dat bijvoorbeeld een menselijke oncoloog een datagebaseerd

---

<sup>20</sup> Het (kwantitatief) bestuderen van de rechtspraak is niet nieuw; het vakgebied van de jurimetrie (recent herdoopt tot *empirical legal studies*) bestaat al tientallen jaren; zie bijv. J. Jacobs & M. Vols, 'Juristen als rekenmeesters: Over de kwantitatieve analyse van jurisprudentie', in: P.A.J. van den Berg & G. Molier (red.), *In dienst van het recht: Opstellen aangeboden aan prof. mr. J.G. Brouwer ter gelegenheid van zijn afscheid als hoogleraar Algemene Rechtswetenschap aan de Rijksuniversiteit Groningen (Brouwer bundel)*, Den Haag: Boom Juridisch 2017, p. 89-104. Wat nieuw is, is het gebruik van moderne methoden uit de *machine-learning* en taaltechnologie in aanvulling op meer conventionele statistiek.

<sup>21</sup> S. Danziger, J. Levav & L. Avnaim-Pesso, 'Extraneous factors in judicial decisions', *Proceedings of the National Academy of Sciences of the United States of America* 2011, afl. 17, p. 6889–6892.

<sup>22</sup> Zo is er bijvoorbeeld in de pers veelvuldig geschreven over zeer strenge 'hungry judges' (Danziger e.a. 2011), maar worden deze resultaten bestreden door andere wetenschappers (K. Weinshall-Margel & J. Shapard, 'Overlooked factors in the analysis of parole decisions', *Proceedings of the National Academy of Sciences of the United States of America* 2011, afl. 42; A. Glöckner, 'The irrational hungry judge effect revisited', *Judgment and Decision Making* 2016, afl. 6, p. 601-610).

<sup>23</sup> Vergelijk bijvoorbeeld het systeem van [www.magontslag.nl](http://www.magontslag.nl), waarin via expliciete juridische regels omtrent ontslag een mogelijke uitkomst in een ontslagzaak wordt gegeven, met de meer statistische voorspellingen van LexIQ (*NRC Handelsblad* en *Trouw* 2019, noot 4).

voorspelalgoritme voor het herkennen van huidkanker moet raadplegen als bewezen is dat dit algoritme beter presteert dan de mens.<sup>24</sup> Maar deze analogie klopt niet, want anders dan in het medische voorbeeld, doen een juridisch voorspelalgoritme en een rechter verschillende dingen.

In het medische voorbeeld verrichten mens en algoritme dezelfde taak, namelijk het herkennen van kanker op foto's van bijvoorbeeld moedervlekken. Ook worden de inschattingen van zowel de mens als het algoritme vergeleken met dezelfde (objectieve) waarheid: door de cellen onder een microscoop te bekijken kan met zekerheid bepaald worden of het om kanker gaat. Het gaat hier dus om het vergelijken van een menselijke en een algoritmische deskundige aan de hand van dezelfde standaard.<sup>25</sup> Dan is een vergelijking van hoe goed mens en algoritme het doen zinvol, en kan het algoritme het beter doen dan de arts, namelijk door kwaadaardige moedervlekken die de arts mist als zodanig te bestempelen. In het geval van een algoritmische uitkomstvoorspeller gebeurt er echter iets anders. Deze voorspelt namelijk welke diagnose de menselijke arts (zonder microscoop) zou stellen, en dan is het zinloos om te zeggen dat het algoritme het beter doet dan de desbetreffende arts. Sterker nog: ook een correcte voorspelling van een foute diagnose door de arts geldt voor zo'n algoritme als een succes. Om dezelfde reden is het zinloos om een juridische uitkomstvoorspeller te vergelijken met een de rechter, want ook hier geldt dat een correcte voorspelling van een foutieve juridische beslissing voor het voorspelalgoritme als een succes telt. Dus de accuratesse van een algoritmische uitkomstvoorspeller kan geen maatstaf zijn voor de juridische kwaliteit van de voorspelde beslissingen, want onder de correct voorspelde beslissingen kunnen best juridische incorrecte of discutabele beslissingen zijn.

Een fundamenteel probleem voor juridische uitkomstvoorspellers is dat rechters niet voorspellen aan de hand van statistische correlaties, maar beslissen op basis van redengevende verbanden. De artsen en voorspelalgoritmes in de bovenstaande voorbeelden zoeken naar *statistische* verbanden tussen de kenmerken en de uitkomsten. Rechters echter zoeken niet naar statistische maar naar *redengevende* verbanden. Een voorbeeld: stel dat een strafrechter het redengevend vindt of iemand door een onvoorwaardelijke straf zijn baan zou verliezen. Werkloosheid correleert statistisch met andere factoren, zoals woonplaats of opleidingsniveau, dus een datagestuurde voorspelalgoritme zal een statistisch verband vinden tussen iemands woonplaats en het al of niet krijgen van een onvoorwaardelijke straf. Maar voor de rechter is iemands woonplaats natuurlijk geen redengevend feit. Een motivering als 'u krijgt een onvoorwaardelijke straf, want u woont in de Schilderswijk, maar de medeverdachte krijgt een voorwaardelijke straf, want hij woont in Wassenaar' zal in het algemeen niet acceptabel bevonden worden.

Ook is er in het recht vaak geen duidelijke objectieve waarheid – juist hierom is het belangrijk dat rechters hun beslissingen uitvoerig motiveren, zodat de uitspraak *inhoudelijk* getoetst kan worden. Voorspelalgoritmes zouden hun voorspellingen dus ook inhoudelijk moeten kunnen motiveren. Maar eerder zagen we dat voorspelalgoritmes gebaseerd op *machine learning* vaak juist niet hun uitkomsten in juridisch zinvolle termen kunnen uitleggen of motiveren. Alleen algoritmes die hun voorspellingen baseren op juridische relevante factoren kunnen dit. Echter, deze algoritmes vereisen, zoals we in par. 2.2 hebben gezien, dat alle juridisch relevante factoren in het zaaksdossier handmatig worden aangegeven en gekwalificeerd. Daarvoor moet

---

<sup>24</sup> J. Susskind, *Future Politics: Living Together in a World Transformed by Tech*, Oxford: Oxford University Press 2018. Zie ook A. Karsemeijer, 'Zijn dit de langetermijneffecten van algoritmen?' *Rechtstreeks* 2019, afl. 2, p. 35-38.

<sup>25</sup> Zo kunnen we ook menselijke en algoritmische deskundigen vergelijken die recidive voorspellen; we weten immers van personen of ze wel of niet opnieuw gerecidiveerd hebben na hun vrijlating. Zie Bijlsma e.a. 2019.

een rechter inhoudelijk over de zaak nadenken, net zoals hij altijd al heeft gedaan. Dat beperkt het nut van dit soort algoritmes, hoewel het denkbaar is dat ze rechters toch van nuttige informatie kunnen voorzien, zeker als het in de toekomst mogelijk wordt om de relevante factoren uit de tekst van jurisprudentie te leren.

Ook is er in principe geen bezwaar tegen de inzet van algoritmische deskundigen in de rechtszaal. Omdat algoritmische deskundigen adviseren op gebieden waarop de rechter geen expert is, heeft het minder zin om een voorspelling inhoudelijk te controleren. Zinvoller is om inhoudelijke (technische) experts te laten vaststellen of het algoritme in het algemeen wel van voldoende kwaliteit is.

## 5 Conclusies en aanbevelingen

In dit artikel hebben we besproken wat voorspelalgoritmes kunnen betekenen in het recht, met name in de rechtspraak. Cruciaal bleek het verschil tussen algoritmische deskundigen en algoritmische uitkomstvoorspellers. Het opvatten van een voorspelling van *algoritmische uitkomstvoorspellers* – in het bijzonder datagestuurde uitkomstvoorspellers – als redengevend voor een rechterlijke uitspraak is principieel onjuist; het verwacht voorspellen met beslissen, in het bijzonder statistische met redengevende verbanden. Maar tegen het gebruik van *algoritmische deskundigen* in rechtsgedingen zijn geen principiële bezwaren. Hetzelfde geldt voor het toepassen van statistiek en algoritmische uitkomstvoorspellers door de rechtswetenschap en de rechtzoekende. Zo kunnen voorspelalgoritmen rechtswetenschappers inzicht geven in wat rechterlijke uitspraken beïnvloedt en hoe deze uitspraken veranderen over de jaren, en kunnen ze nuttig zijn voor burgers en de advocatuur om in te schatten wat de kansen in de rechtszaal zijn. Er is alleen het praktische probleem van het vaststellen van de kwaliteit van dergelijke algoritmes. Dat probleem is allesbehalve triviaal, gezien de vele valkuilen waarover de data science ons leert.

De enorme aandacht voor voorspelalgoritmen is zeker een hype, maar onzinnig willen wij hem niet noemen. In ons artikel hebben we verschillende manieren besproken waarop ze nuttig kunnen zijn voor de rechtsgemeenschap. Verder is het onderzoek naar zulke algoritmen, en in bredere zin AI-algoritmen voor het recht, hard nodig. De rechtsgemeenschap vaart wel bij gedegen kwantitatieve analyses met behulp van AI, en loopt hierin ver achter op bijvoorbeeld de medische wetenschap. De razendsnelle ontwikkelingen binnen de taaltechnologie maken het bovendien mogelijk de rechtspraak beter te digitaliseren, bijvoorbeeld door het zoeken naar of samenvatten van relevante documenten, en het vinden van soortgelijke zaken te ondersteunen. Belangrijk is dat de rechtsgemeenschap samen met AI-ontwikkelaars en -onderzoekers optrekt in het ontwikkelen, valideren en uitleggen van algoritmes.

Tot slot bevelen wij het volgende aan met betrekking tot het gebruik van voorspelalgoritmes in de rechtspraak.

Datagestuurde uitkomstvoorspellers hebben geen plaats in een rechtsgeding. ‘Inhoudelijke’ voorspellers hebben dat mogelijk wel, maar alleen als ze hun uitkomsten op juridisch relevante gronden kunnen motiveren en als hun juridische kwaliteit voldoende vaststaat. En de rechter mag zich alleen door deze inhoudelijke motiveringen laten leiden en niet door statistische metrieken zoals nauwkeurigheid. Niet een nauwkeurigheidspercentage of kansinschatting, maar alleen inhoudelijke argumentatie kan redengevend zijn voor een beslissing.



Voorspellingen van algoritmische deskundigen kunnen nuttig zijn voor rechters, maar ze mogen alleen gebruikt worden als de rechter voldoende zicht heeft op de kwaliteit van het algoritme. Het is daarom belangrijk dat er deugdelijke validatieprocedures komen en regels omtrent de toelaatbaarheid van algoritmische deskundigen in rechtsgedingen, net zoals er regels en procedures zijn omtrent menselijke deskundigen.