

Data Domino

Om beurten schrijven vier experts uit de datawereld over een brandende kwestie. Deze week: Max Welling over hoe beginners in big data hier functioneel mee kunnen omgaan.

Eerst je data inventariseren, daarna verder kijken

Veel beroeps-groepen worstelen met het thema big data en kunstmatige intelligentie. Dat merkte ik de laatste maanden op symposia van advocaten, rechters, onderwijzers en artsen. De grote vragen die daar aan bod komen zijn: 'Hoe blijf ik bij in een wereld die snel verandert?' En: 'Hoe profiteer ik hiervan op een ethisch verantwoorde manier?'

Ter verduidelijking: big data en kunstmatige intelligentie zijn twee kanten van dezelfde munt. Met de term big data bedoelen we vaak het verzamelen en opslaan van data, en het doen van relatief eenvoudige analyses, terwijl we bij de term kunstmatige intelligentie meer denken aan zeer geavanceerde algoritmen. Op het symposium Big Data: Ramp of Rijkdom? op 23 november bij de Raad

voor de Rechtspraak in Den Haag werd de noodzaak om vaart te maken wel gevoeld. De advocaten zijn ten slotte de rechters al een eind voor met het op IBM's supercomputer Watson gebaseerde expertsysteem 'Ross' dat in no time alle relevante processtukken opsnort. Nog even en de rechter wordt in de rechtszaal om de oren geslagen met zijn eigen inconsistenties. De rechterlijke macht

zit op een goudmijn aan data: vele jaren aan procesdossiers en vonnissen. Om de transparantie van de rechtspraak te vergroten vraagt men zich af of deze stukken niet openbaar gemaakt moeten worden. Maar kan dat? Immers, die data bevatten persoonlijke gegevens en anonimiseren is niet eenvoudig. En van wie zijn die data eigenlijk? Bovendien lopen rechters het risico te worden gepro-

fileerd. Bijvoorbeeld: rechter X geeft systematisch zwaardere straffen dan gemiddeld. Die spiegel kan confronterend zijn, maar bevordert wel de objectiviteit van de rechtspraak. Een dag later, bij het symposium Meetbaar Beter in Apeldoorn, werden de parallellen met de gezondheidszorg al snel duidelijk. Ook hier was een gevoel van urgentie te bespeuren. Eén arts vertelde dat het

ophangen van een A4'tje in de operatiekamer met een checklist van vier afvinkpunten de postoperatieve overlevingskansen omhoog deed schieten. Tel daarbij op het feit dat de helft van alle medische procedures niet wetenschappelijk is getoetst, en je begrijpt dat er hier nog wel wat te winnen valt met big data. Dat ook privacy een serieus issue is in de zorg, werd pijnlijk onderstreept door

Max Welling

Max Welling



'Data-analyse kan onthullen dat de ene rechter zwaarder straft dan de andere.'

'Interessante info als je een schadeclaim wilt incasseren.'



Jan Fred van Wijnen, chef Morgen

een artikel dat op die dag verscheen: 'Ziekenhuizen lekken dagelijks privacy-gevoelige informatie'. Na al dat kletsen op symposia moeten de handen uit de mouwen gestoken. Maar hoe doe je dat? Een big-data-avontuur begint altijd bij de bron. Inventariseer dus eerst wat je aan data in huis hebt. Bedenk dan welk probleem je wilt oplossen. Betrek externe expertise in dit proces: acade-

mische partners of start-ups, gespecialiseerd in 'data science', kunstmatige intelligentie en privacy. Formuleer met hen kleine behapbare projecten en ga experimenteren. Falen is niet erg, het doel is te leren. Zo wordt uw organisatie langzaam maar zeker 'data savvy'. Max Welling is hoogleraar machine learning (UvA) en medeoprichter van start-up Scyfer.