

VOORWOORD

Elke dag worden er meer data aangemaakt dan in de hele menselijke geschiedenis tot aan het eind van de vorige eeuw. Meer dan 90 procent van alle data is in de laatste twee jaar gecreëerd. In 2020 wordt voor ieder mens op onze planeet elke seconde 1,7 megabyte aan data gecreëerd en opgeslagen. Welkom in de wereld van big data!

Waarom slaan we al die data op en wat doen we ermee? Data wordt de derde grote productiefactor genoemd naast arbeid en kapitaal. Big data zijn te beschouwen als een goudmijn, waaruit je door middel van data-analyse waardevolle inzichten haalt. Die inzichten kunnen leiden tot betere beslissingen voor je bedrijf, voor de maatschappij en voor je persoonlijke leven. Daarnaast spelen big data een grote rol bij de ontwikkeling van machine learning, de op dit moment meest succesvolle tak van kunstmatige intelligentie. En machine learning is op zijn beurt weer onmisbaar bij de analyse van big data. Samen maken data-analyse en machine

learning deel uit van data science, een multidisciplinair vakgebied dat diverse methoden, processen, algoritmes en systemen gebruikt om kennis en inzichten te halen uit data.

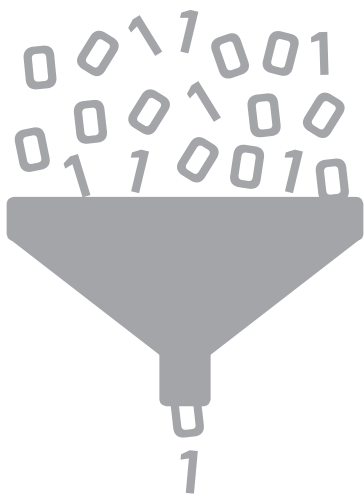
Dit is niet het eerste boek met de term 'big data' in de titel. Het is wel het eerste boek over big data dat je in een mum van tijd laat kennismaken met dit enorm interessante gebied en doet dat op een zeer toegankelijke manier. Als je nog weinig of niets weet van begrippen als algoritmes, big data, data science en machine learning, dan ben je hier aan het goede adres. Alles wat je nodig hebt om meer te weten zijn een flinke dosis nieuwsgierigheid en een open mind. Alle relevante vaktermen worden helder en duidelijk uit de doeken gedaan. Via tal van voorbeelden en anekdotes laat ik je de kracht zien van het gebruik van big data. Maar ook hoe je er zelf mee aan de slag kunt gaan: in je bedrijf en zelfs in je persoonlijke leven. Want vergis je niet, big data zijn niet alleen interessant voor de grote jongens zoals Google, Facebook, Amazon en Microsoft, die eindeloze budgetten hebben om alle kennis en inzicht te halen uit data. Het is ook zinvol voor de leverworstfabrikant, de 'bakker om de hoek' of de plaatselijke voet-

balvereniging. Het gebruik van data – big én small – zou eigenlijk net zo vanzelfsprekend moeten zijn als het bijhouden van de financiële administratie. Het beslissen en handelen op basis van data is veel effectiever dan het beslissen en handelen op basis van buikgevoelens, emotie, vooroordelen, veronderstellingen en routine ('Zo doen we het altijd, waarom moet het anders?'). Als je dit boek gelezen hebt dan weet je niet alleen waarom dat zo is, maar kun je bovendien zelf met data aan de slag in jouw organisatie en kun je het gebruik van data stapsgewijs professionaliseren.

De kracht van de 60-minutenserie is dat je in zeer korte tijd een goed overzicht krijgt over een belangrijke technologie die voor jou relevant kan zijn. Dat geldt ook voor *Meer weten met big data in 60 minuten*. De keerzijde van die medaille is dat je als schrijver slechts een tipje van de sluier kunt oplichten en feitelijk nergens diep op kunt ingaan, zeker bij dit omvangrijke vakgebied. Als het goed is krijg je tijdens het lezen trek in meer. Daarom heb ik in samenspraak met de uitgever een unieke constructie opgezet die je dat kosteloos aanbiedt. Gekoppeld aan dit boek wordt een aantal beknopte

e-books ontwikkeld waarin een bepaald aspect verder uitgediept wordt. Voorbeelden van die e-book zijn *Neurale netwerken doorgronden in 15 minuten* of *Starten met Power BI in 30 minuten*. Ze hebben dezelfde vormgeving en opmaak als dit boek, maar zijn alleen digitaal verkrijgbaar. Je kunt ze gratis downloaden via mijn website (dedata scientist.nl/ebooksbig-data) en via de uitgeverij (haystack.nl/ebooks-bigdata). Zo kan dit boek het begin zijn van een ontdekkingsstocht om nog meer te weten te komen over en met big data.

Dik Bijl
dedata scientist.nl



1

INLEIDING

Dit boek neemt je mee op een ontdekkingsreis naar de kracht van (big) data en laat je zien waarom data de derde grote productiefactor genoemd worden, naast kapitaal en arbeid. Ik begin met een aantal voorbeelden en anekdotes van hoe het gebruik van data heeft geleid tot nieuwe, niet voor de hand liggende kennis en inzichten. Ook toon ik hoe het gebruik van data en machine learning heeft geleid tot allerlei toepassingen, soms spectaculair en soms heel alledaags. Hoofdstuk 2 gaat van de voorbeelden over naar het grote plaatje: hoe big data de grondstof vormen voor data-analyses en voor machine learning en daarmee de digitale transformatie van onze maatschappij een extra impuls geven. Hoofdstuk 3 introduceert alle belangrijke begrippen rondom data-analyse, terwijl hoofdstuk 4 helder en inzichte-

lijk beschrijft wat machine learning is en hoe belangrijk dat is voor data-analyse. Daarna wordt het tijd te laten zien hoe je zelf aan de slag kunt gaan met (big) data. Eerst simpel en pragmatisch (hoofdstuk 5). Dan komt de volgende stap, en lees je hoe je het gebruik van data in je organisatie verder kunt professionaliseren (hoofdstuk 6). We beginnen dus eerst met een aantal voorbeelden die de kracht van big data aantonen.

Het eerste voorbeeld is echt een big data-klasieker en heeft in 2003 onbedoeld het wereldnieuws gehaald. Target is een grote Amerikaanse retailer in een markt waar Walmart de absolute leider is. Een boze man kwam een winkel van Target in Minneapolis binnen en eiste de manager onmiddellijk te spreken. De vader had een advertentieblaadje in zijn hand. Toen de manager verscheen, zei de vader kwaad: 'Kijk eens wat mijn tienerdochter van jullie toegestuurd kreeg: een blaadje met aanbiedingen voor babyspullen. Willen jullie soms dat zij zwanger wordt?' De verbouwereerde manager wist van niets, verontschuldigde zich en zei dat hij zou uitzoeken wat er mis was gegaan. Na een week belde hij de vader op om zich opnieuw te

verontschuldigen. Maar het was de vader zelf die zich bij hem verontschuldigde. Hij zei: 'Ik heb het er met mijn dochter over gehad en het blijkt dat zij inderdaad zwanger is.' Wat was er gebeurd?

Het is een bekend gegeven dat klanten redelijk trouw zijn en niet snel van hun voorkeur-retailer veranderen. Dat is een probleem in de moderne, hypercompetitieve markt. Op basis van een wetenschappelijk artikel uit de jaren tachtig wist het management van Target dat klanten eigenlijk alleen op belangrijke momenten in hun leven van leverancier wisselen. Een van die levensveranderende momenten is als iemand zwanger wordt. De data-analisten van Target kregen van het management de volgende vraag: 'Kunnen jullie achterhalen op basis van de verkooptransacties die we hebben of iemand zwanger is?' Target heeft net als andere retailers enorm grote databases met alle verkooptransacties in hun winkels. Als een klant opeens regelmatig babyspullen zoals luiers en babyshampoo begint aan te schaffen, dan kun je concluderen dat er een baby is geboren. Maar verandert het koopgedrag van diezelfde klant al eerder, al tijdens de eerste maanden van de zwangerschap?

Dat blijkt inderdaad het geval. Het data-analyseteam kwam met een tiental producten die klanten al relatief vroeg in hun zwangerschap bleken aan te schaffen, waaronder een bepaalde soort handcrème en wat eerste babyspulletjes. Met dat nieuwe en niet voor de hand liggende inzicht ging het management van Target aan de slag. Er kwam een campagne waarbij klanten met het koopprofiel van iemand die zwanger is, werden benaderd met speciale aanbiedingen en kortingen op babyspullen.

De campagne was er een van een zeer succesvolle reeks waarbij gewerkt werd met inzichten verkregen uit eigen databases. In de jaren na 2002 groeide Target met zo'n 20 procent per jaar. Het onbedoelde wereldnieuws van de boze vader toont overigens aan dat er ook risico's kunnen zitten aan 'data driven' handelen.

Een volgend voorbeeld stamt uit ongeveer dezelfde tijd. In het zuiden van de Verenigde Staten zijn er in een bepaald jaargetijde regelmatig orkanen. De orkanen, hun pad en hun kracht kunnen redelijk goed voorspeld worden (dankzij big data). Mensen in een orkaangebied weten enkele dagen van tevoren wanneer de orkaan komt. Ze bereiden zich daarop voor

door producten als flessenwater en zaklampen in te slaan. Winkelketens anticiperen daar weer op door die producten goed op voorraad te hebben. Walmart kwam door data-analyse tot een opzienbarende ontdekking. Het bleek dat klanten voorafgaand aan een orkaan onverwacht grote hoeveelheden Strawberry pop-tarts insloegen. Dat zijn in de VS zeer populaire, langdurig houdbare zoetige taartjes die je even opwarmt in de oven voordat je ze opeet. Voorafgaand aan een orkaan werden die achttien keer meer verkocht dan normaal. Op grond van dit door data-analyse gevonden inzicht, sloeg Walmart extra veel Strawberry pop-tarts in toen zich weer een orkaan aandeed. De omzet schoot omhoog.

Het volgende voorbeeld is een stuk dicht bij huis. Het gaat over de politie en betreft een toepassing binnen human resource management¹. De politie beschikt over een aantal helikopters en heeft daar uiteraard piloten voor nodig. Het bleek dat veel piloten al na een paar jaar de politie verlieten om een baan te accepteren in

1 Bron: 'AIHR Introduction Webinar HR Analytics' van AIHR Academy op YouTube

Ons leven wordt steeds meer bepaald door data. Je kunt er ook als organisatie je voordeel mee doen, maar dan moet je al die beschikbare data wel omzetten in waardevolle kennis. In dit boek lees je hoe je dit kunt doen.

Dik Bijl schrijft boeken en adviseert organisaties over belangrijke technologische ontwikkelingen voor mens en maatschappij. Hij is gespecialiseerd in big data en data science.

Weinig tijd, maar veel ambities? Informeer jezelf snel en grondig met de boeken in de serie *Digitale trends en tools in 60 minuten*.

