

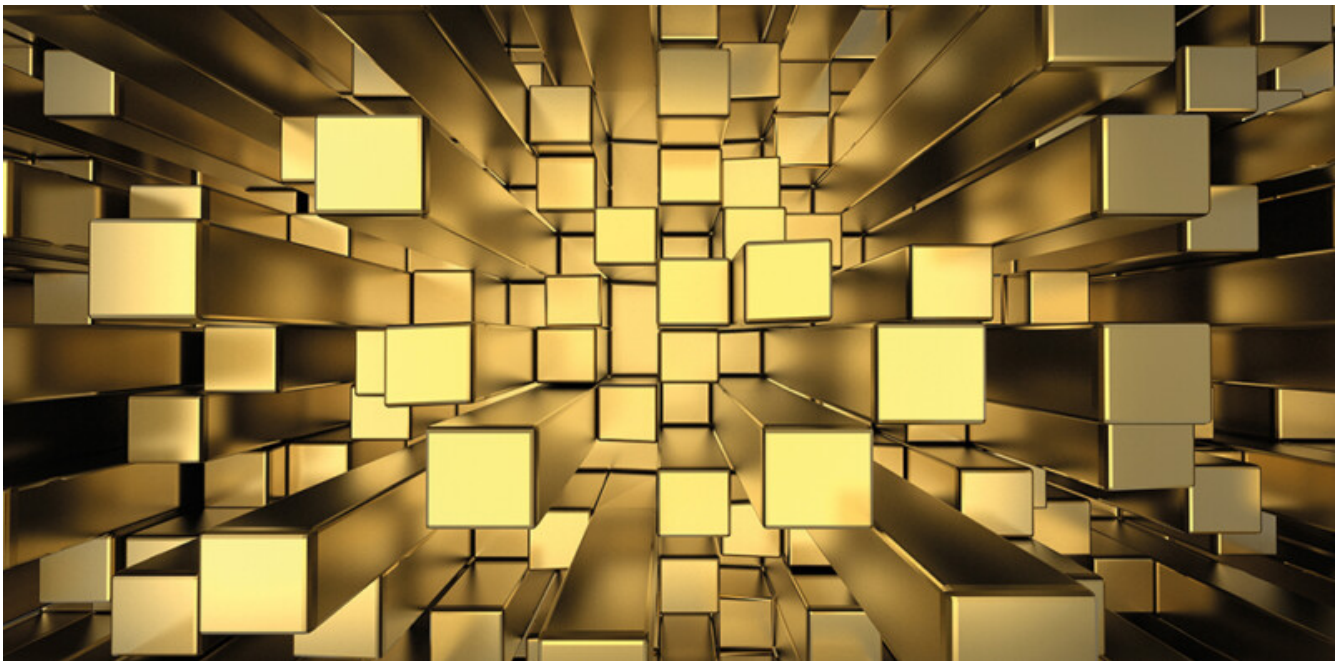
ARTIKELN | 1 APRIL 2019

DATAGEDREVEN WAARDECREATIE

Auteur: Hans van der Zwan

Beeld: Adobe Stock - Markus Winkler

Leestijd: 7 min



Hoe gaat waardecreatie op basis van data in zijn werk? Wat is belangrijk voor het welslagen van dataprojecten? Over het effectief mijnen van het nieuwe goud.

Op school leerden wij over de industriële revolutie en de grote verandering die deze voor iedereen teweegbracht. In de jaren tachtig van de vorige eeuw begon men te spreken over de digitale revolutie en over de overgang naar een informatiemaatschappij. Inmiddels beginnen wij te beseffen dat de digitale revolutie meer impact heeft op zowel organisaties, individuen als de maatschappij dan de industriële revolutie heeft gehad.

De mogelijkheden van het verzamelen en analyseren van data hebben impact op alle terreinen van ons leven, van hoe wij met elkaar communiceren tot hoe wij onze auto (niet) besturen, van hoe wij leren tot hoe politieke menings- en besluitvorming tot stand komt, van onze vakantieplanning tot de businessplanning van een onderneming. Wat dat laatste betreft wordt bedrijven voorgelaten werk te maken van het creëren van waarde uit data, want data is het nieuwe goud.

Niet alles is goud wat er blinkt

De big-data-revolutie verandert onze hele samenleving. Een transitie waar wij middenin zitten. Naast grote voordelen

AUDIT

MAGAZINE

zijn er schaduwzijden. Een gevolg van de big-data-revolutie is bijvoorbeeld het vormen van groepen van gelijkgestemden op social media. Daar waar de bedrijven achter deze media om het hardst roepen dat zij mensen met elkaar verbinden, ontstaat een segregatie in de maatschappij in min of meer homogene groepen. Groepen die zich vormen in netwerken op social media. Het risico van deze nieuwe groepsvorming is dat je als groepslid voortdurend gesterkt wordt in je mening en andere geluiden niet of nauwelijks meer hoort. Zo ontstaat er eerder polarisatie dan toenadering tussen deze groepen en de mensen die geen lid zijn van de groep.



Voor individuen betekent de digitale revolutie naast grote voordelen ook het inleveren van veel, heel veel privacy. Voor het schrijven van dit artikel heb ik mijn archief met persoonlijke data bij Google opgevraagd. Dit leverde een gezippt bestand van ruim 7 gigabyte op. Inmiddels zal ook ergens opgeslagen zijn dat dit bestand op 21 november 2018 is opgevraagd. Dit betreft dan een bedrijf waarvan bekend is dat persoonlijke data opgeslagen worden.

Veel minder bekend zijn de databedrijven die handelen in persoonlijke gegevens.¹ Interessant in dit verband is dat de grondlegger van Alibaba, het Chinese bedrijf waar je zo ongeveer alles kunt bestellen, in een interview in het *Financieel Dagblad* (27 oktober 2016) zei dat Alibaba geen retailer maar een databedrijf is. Data vormt het hart van de business en dat voor een bedrijf in een land waar GDPR en AVG niet het hart van de privacywetgeving zijn.

Data: het nieuwe goud

Ook bedrijven die zich niet als databedrijf afficheren, creëren waarde uit verzamelde data. Zo kan door analyse van historische klantgegevens marketingbudget gericht worden ingezet. Grootwinkel- en telecombedrijven zijn voorbeelden van bedrijven die hier hun voordeel mee doen. Bedrijven als Facebook, Google en Amazon weten uiteraard al langer hoe data om te zetten in goud. Niet voor niets zijn Apple, Alphabet (het moederbedrijf van Google), Microsoft, Amazon en Facebook tegenwoordig de bedrijven met de hoogste beurswaarde.²

Voor veel andere organisaties is het echter nog een vraag of, en zo ja hoe, het goud gemijnd kan worden uit de hoeveelheden data die in datawarehouses opgeslagen liggen. Om de slag niet te missen, worden datascientisten aangetrokken om de waardevolle informatie die in de data verborgen liggen naar boven te halen en de voordelen van de digitale revolutie te plukken. Dat dergelijke projecten niet altijd succesvol zijn, weten wij bijvoorbeeld van het project van de speciaal hiervoor in het leven geroepen afdeling 'de Broedkamer' bij de Belastingdienst. Dit project moest mede naar aanleiding van de uitzending van het onderzoeksjournalistieke tv-programma *Zembla* van 1 februari 2017 worden

stopgezet. Voor waardecreatie uit data is meer nodig dan een team dataspecialisten.

Medewerkers in data-analysetechnieken scholen helpt om draagvlak op de werkvloer te creëren voor een datagedreven organisatie

Van data naar goud: teamwork

Waardecreatie uit data is teamwork. Datascientisten zijn als de besten in staat data om te zetten in informatie, verbanden in data te ontdekken en te beschrijven met wiskundige modellen. Zij zijn niet de eerste aangewezenen om te bepalen wat de toegevoegde waarde is voor de organisatie, daarvoor is kennis van het domein waar de organisatie opereert onontbeerlijk. Een organisatie (meer) datagestuurd maken, dat wil zeggen data gebruiken om de organisatiedoelstellingen te halen, vereist samenwerking tussen datascientisten en de mensen die de businesskant van de organisatie kennen. Het is niet voor niets dat de laatste tijd, als het gaat om de aanpak van een project van waardecreatie uit data, veel aandacht wordt besteed aan het samenstellen van een gebalanceerd projectteam.³

Dat belang kan niet onderschat worden. Het is niet anders dan het altijd geweest is in de statistiek. De statistiek levert technieken om verbanden tussen variabelen meetbaar te maken, maar de statistiek doet geen uitspraak over causaliteit. Uitspraken daarover moeten uit andere disciplines komen, zoals de economie, de psychologie of de meteorologie. Met andere woorden, het is niet aan de statistiek om de causaliteit en relevantie van statistische verbanden te beoordelen. Dat vereist domeinkennis.

Elkaar moeilijk verstaan

Dataprojecten tot een succes maken vereist dus teamwork. Eén probleem dat herhaaldelijk gesignaleerd wordt, is dat datascientisten en de domeinkenners elkaars taal moeilijk verstaan. Een projectmanager is de eerste aangewezenen om de communicatie tussen beide groepen tot stand te brengen. Daarvoor is kennis van datascience, naast kennis van het domein van de organisatie, onontbeerlijk. Dat hoeft niet op het niveau van de datascientist te zijn, maar moet wel voldoende zijn om een oordeel te kunnen hebben over de inbreng van de datascientist. Zoals een accountant geen actuariael rekenaar hoeft te zijn om een oordeel te kunnen geven over een pensioenvoorziening op een balans, maar wel de concepten achter een berekening van zo'n voorziening moet begrijpen.

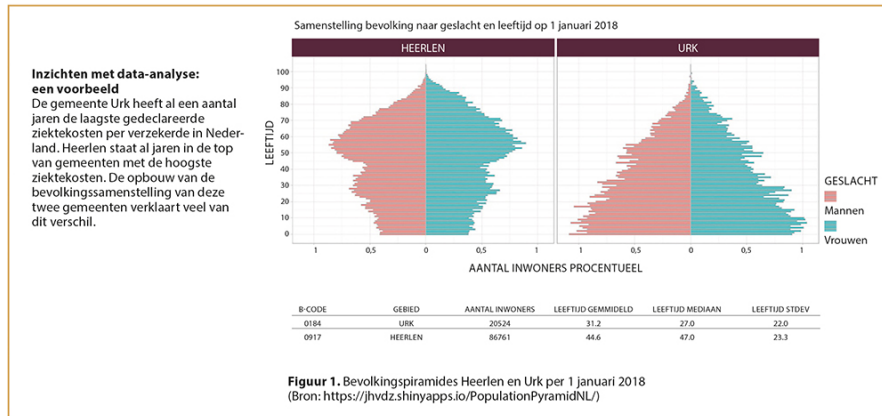
Steeds vaker wordt aanbevolen, medewerkers van de eigen organisatie (bij) te scholen in hedendaagse data-analysetechnieken. Juist daar waar organisaties nog zoekend zijn naar de antwoorden op de uitdagingen van het (big-)datatijdperk, kunnen relatief eenvoudige analysetechnieken tot verrassende inzichten leiden. Medewerkers daarin scholen en leren dergelijke inzichten uit de data te verkrijgen en presenteren, helpt om draagvlak op de werkvloer te creëren in een volgende stap naar een meer datagedreven organisatie.

Inzichten met data analyse: een voorbeeld

De gemeente Urk heeft al een aantal jaren de laagste gedeclareerde ziektekosten per verzekerde in Nederland. Heerlen staat al jaren in de top van gemeenten met de hoogste ziektekosten. De opbouw van de bevolkingssamenstelling van deze twee gemeenten verklaart veel van dit verschil (zie figuur 1).

AUDIT

MAGAZINE



Figuur 1. Bevolkingspiramides Heerlen en Urk per 1 januari 2018. (Bron: <https://jhvdz.shinyapps.io/PopulationPyramidNL/>)

Data-analyse, waar te beginnen

Door het vele hedendaagse onderzoek naar data-analysetechnieken neemt het arsenaal aan deze technieken enorm toe. Voor data-analisten/scientisten is een leven-lang-leren inmiddels voorwaarde om hun vak te kunnen blijven uitoefenen. Het is voor medewerkers aan een dataproject vanuit de organisatie veelal ondoenlijk deze ontwikkelingen bij te houden. Dat is ook niet nodig, basiskennis van statistische analysetechnieken en van veelgebruikte modellen uit het vakgebied machine learning (zie hierna), gevoegd bij domeinkennis, maken het mogelijk een waardevolle bijdrage te leveren. Een aantal onderwerpen waarvan (basis)kennis onontbeerlijk is:

Data-analyse (1): data verzamelen

Anders dan bij traditioneel onderzoek, waar de onderzoeksvraag bepalend is voor de te verzamelen data, beginnen projecten die moeten leiden tot een meer datagedreven organisatie met de vraag: welke waarde ligt er in de reeds aanwezige data besloten? Veelal betekent dit een keuze waar te beginnen. Dit vereist overleg tussen de datascientisten die weten wat analytisch mogelijk is en de domeindeskundigen die kunnen inschatten waar voor de organisatie de meeste winst te behalen is.

Data-analyse (2): data schonen

Data-analyse begint met data schonen. Het is altijd weer verbazingwekkend hoe data vervuild kunnen zijn. Hier kan een veelheid van oorzaken aan ten grondslag liggen, zoals invoerfouten, verandering van definities in de loop van de tijd, het ontbreken van waarden door een effect in meetapparatuur of het samenvoegen van data uit verschillende bronnen. Het is van groot belang de integriteit van de data waarmee verder gewerkt wordt vast te stellen. Het aloude adagium 'garbage in, garbage out' is ook in dit geval van toepassing.

Data-analyse (3): data wrangling

Een volgende stap is het transformeren van de data naar een formaat dat geschikt is voor data-analyse, bijvoorbeeld een datamatrix. Er dient zorg voor gedragen te worden dat numerieke en alfanumerieke variabelen als zodanig herkend worden door de gebruikte software. Berucht zijn de datumvariabelen, omdat er internationaal er geen consensus is over de notatie van een datum (in data-analyse wordt dit opgelost door datumvariabelen om te zetten naar het jjjjmmdd formaat).

Een data wrangler, een van de nieuwe beroepen die is ontstaan in het datatijdperk, heeft kennis van een diversiteit aan dataformaten en weet hoe deze naar gewenst formaat over te zetten.

AUDIT

MAGAZINE



Data-analyse (4): data-analyse en machine-learning

In de voorgaande fasen is basiskennis statistiek nodig, in de data-analysefase is meer nodig dan dat. Data-analyse begint veelal met visualisaties, het maken van tientallen grafieken om inzicht te krijgen in de data. Dit is een van die eenvoudige middelen die organisaties op het spoor kunnen zetten waar winst te behalen is.

Machine learning is de discipline die zich bezighoudt met het zoeken naar en beschrijven van structuren en verbanden in de data. Dit vakgebied op het snijvlak van wiskunde, statistiek en informatietechnologie staat momenteel centraal in de wereld van data-analyse.

Gezocht wordt naar modellen die de structuren in de data zo goed mogelijk beschrijven, zoals regressiemodellen en beslisbomen. Dergelijke modellen zijn niet nieuw, nieuw zijn wel de aanpassingen die erop gemaakt worden, bijvoorbeeld door het toevoegen van parameters indien gebruikte variabelen onderling sterk samenhangen. Nieuw is ook het door de inzet van software eindeloos variëren met parameters van de modellen. Veel onderzoek richt zich momenteel op het ontwerpen van nieuwe modellen en het toetsen van de bruikbaarheid. Opvallend is dat dit grotendeels empirisch onderzoek is dat in veel gevallen leidt tot goede voorspelmodellen, zonder dat duidelijk is waarom het model zo goed werkt. Dit speelt in het bijzonder bij het gebruik van zogenaamde neurale netwerken en 'deep learning'. Regelmatig valt dan de term 'black boxes'.

AUDIT

MAGAZINE



Zonder een expert te zijn op het gebied van machine learning kan met basiskennis van dit vakgebied gebruikgemaakt worden van de resultaten van diverse technieken. Net zomin als begrip van computersystemen nodig is om een computer te kunnen gebruiken, is kennis van algoritmen nodig om de analysetechnieken te kunnen gebruiken.

Data-analyse (5): presenteren en rapporteren

Het kunnen presenteren en rapporteren van de resultaten van analyses is van cruciaal belang om tot een datagedreven organisatie te komen. Noem het: het verhaal kunnen vertellen. En dat verhaal moet zowel verteld kunnen worden aan de kritische onderzoeker/wetenschapper als aan de mensen op de werkvloer die geen kennis hebben van data-analyse. Het eerste kan aan de datascientist in het team worden overgelaten, het tweede ligt meer op het pad van de data-analisten die zelf van de werkvloer komen. Zoals bij elke verandering is weerstand een natuurlijke reactie. Goed en begrijpelijk gepresenteerde resultaten van uitgevoerd data-onderzoek met aandacht voor de waarde die het heeft voor de organisatie, is de meest voor de hand liggende route om weerstand te overwinnen en dataprojecten tot een succes te maken.

Noten

1. Zembla besteedde er op 2 december 2015 aandacht aan in de aflevering 'Data: het nieuwe goud'.
2. <https://www.forbes.com/global2000/list> (geraadpleegd 21 november 2018).
3. <https://insidebigdata.com/2018/02/16/7-key-members-every-big-data-team> (geraadpleegd 21 november 2018).

Over

Hans van der Zwan werkt als docent data-analyse en machine learning en als programmamanager bij de Academie voor Masters & Professional Courses van De Haagse Hogeschool. Hij is verantwoordelijk voor de post-bacheloropleiding Big Data Analyst en MBA Big Data Analytics.

Tags: data analytics

AUDIT

MAGAZINE

Bron url: <https://auditmagazine.nl/artikelen/datagedreven-waardecreatie/>