MAB MAANDBLAD VOOR ACCOUNTANCY EN BEDRIJFSECONOMIE

Maandblad voor Accountancy en Bedrijfseconomie 92(5/6) (2018): 143–146
DOI 10.5117/mab.92.26691

Column

De belofte van Big Data

Tom Groot

Received 15 May 2018 | **Accepted** 14 June 2018 | **Published** 23 July 2018

Bedrijven en organisaties in de overheids- en non-profitsector worden in toenemende mate geconfronteerd met nieuwe ontwikkelingen in de informatie- en communicatietechnologie. Het gaat van nieuwe IT-toepassingen in de uitvoering van productietaken en verbeterde interne communicatie binnen bedrijven tot de ontwikkeling van nieuwe producten en diensten op basis van real-time gebruikersinformatie en nieuwe vormen van klantcontact via sociale media. Bedrijven verwerken momenteel in twee jaar tijd meer gegevens dan in de voorgaande 2000 jaren en dit volume wordt elk jaar exponentieel groter (Kyed, Fillela and Venugopal 2013). Organisaties die gebruik maken van big data en deze analyseren ten behoeve van prestatieverbetering bereiken 5 tot 6 procent hogere productiviteit (Brynjolfsson, Hammerbacher and Stevens 2011). Een recente studie laat zien dat organisaties die big data en artificiële intelligentie op een doelgerichte manier gebruiken in hun ondernemingsstrategie gemiddeld 10% hogere operationele winstmarge laten zien (Bughin et al. 2017). De winstpotentie blijkt het grootst te zijn bij organisaties in de financiële sector, retail, onderwijs, professionele dienstverlening en gezondheidszorg: opvallend veel bedrijven buiten de traditionele maakindustrie!

Big data en data analytics

Big data is een containerbegrip en dat maakt het moeilijk om precies te omschrijven waar we het nu over hebben. Data is de grondstof voor informatie, big data zijn gegevens die aan vier kenmerken voldoen: een groot volume, hoge omloopsnelheid, grote variëteit en onzekere waarheidsgetrouwheid. In het Engels spreekt men ook wel van de vier V's: huge *volume*, high *velocity*, huge *variety*, and uncertain *veractity* (Vasarhelyi et al. 2015). Veel gegevens komen van transactieverwerkende systemen, van sensoren zoals camera's en microfoons, van het internet of things, van sociale media en van andere private of publieke databanken. Dit verklaart het grote volume, de grote variëteit, het hoge verversingstempo en de gebrekkige controle op databetrouwbaarheid. Internet heeft de mogelijkheid geboden om deze gegevensbronnen met elkaar te verbinden en op deze wijze big data-systemen te creëren. Aanvankelijk gebruikten organisaties het internet om afzetmogelijkheden te vergroten en efficiëntie te verhogen. We zouden deze fase Big Data 1.0 kunnen noemen. Vervolgens begonnen bedrijven gebruik te maken van de interactieve mogelijkheden van internet en van sociale media om doelgericht op gedragingen en opvattingen van individuele klanten te reageren. Een bekend voorbeeld is Amazon dat op basis van aankopen van klanten automatisch nieuwe suggesties genereert van soortgelijke producten of klanten die naast de huidige aankoop ook andere zaken hebben aangeschaft. Dit is Big Data 2.0. Momenteel is fase 3.0 van Big Data aangebroken: de fase van *data science.* In deze fase zoeken bedrijven naar data en analysemethoden die beslissingen ondersteunen en (geautomatiseerd) verbeteren (Provost and Fawcett 2013). De belangrijkste uitdaging van deze fase is een verdieping van de data-analyse, zodat de besluitvorming in organisaties meer op data wordt gebaseerd dan op subjectieve inschattingen en persoonlijke overtuigingen.

De McKinsey Advanced Analytics-groep heeft op basis van een onderzoek onder meer dan 100 data analytics-projecten in bedrijven een inschatting gegeven van de mate waarin zij de bedrijfsprestaties in een periode van drie jaar kunnen verbeteren (Hürtgen and Mohr 2018). Daarbij maakt zij een onderscheid tussen projecten die de opbrengsten verhogen (de zogenaamde "top line growth") en kosten verlagen (de "bottom line reduction"). Het daaruit volgend overzicht is indrukwekkend (zie tabel 1), omdat het laat zien dat er op tal van terreinen winst te boeken is. Tevens wordt duidelijk dat deze verbeteringen een grote invloed kunnen hebben op omzet, marge en kosten. Overigens is het goed te bedenken dat elke inschatting van resultaatverbetering op zichzelf staat, de effecten van de verschillende maatregelen kunnen dan ook niet bij elkaar worden opgeteld.

Tevens valt op dat de maatregelen betrekking hebben op verschillende bedrijfsfuncties. Zo zien we verbeteringen in productsamenstelling, marketing (bundeling van producten, prijsstelling en promotie), logistiek (voorraadbeheer en ruimtegebruik), procesmanagement (onder-

Tabel 1. Bewezen waardecreatie op basis van Data Analytics.

Opbrengstverhoging (top line growth)			Kostenverlaging (bottom line reduction)	
Maatregel	Geschat effect		Maatregel	Geschat effect kostenreductie
	op omzet	op marge		
Optimaliseren productenpakket	2,0 %	1,0 ppt *)	Preventief onderhoud	20–00% van onderhoudskosten
Cross- en Upselling van producten	2,0 %		Marketing uitgaven	5–50% van marketingkosten
Vasthouden bestaande klanten	1,5 %		Voorspelling afzet	20–00% opslag- kosten
Prijsstelling	2,0 %	1,0 ppt	Voorkomen van fraude	1–1% van verlies door fraude
Verbeteren voorraadbeheer en bestellingen	2,0 %	0,5 ppt	Beheer dubieuze debiteuren	10–00% van verliezen door oninbaarheid
Optimaliseren product-promotie	1,5 %	1,0 ppt	Planning inzet personeel	10–00% van personeelskosten
Optimaliseren gebruik ruimte en opslag	1,5 %		Optimaliseren van de supply chain	10–00% van logistieke kosten

*): ppt = procentpunt.
Bron: McKinsey 2018.

houd en optimaliseren supply chain), financiën (beheer dubieuze debiteuren en voorkómen van fraude) en personeelsmanagement. Hieruit wordt ook duidelijk hoezeer Big Data-analyse functie-overschrijdend kan werken.

Data analytics-processen en -technieken

Er is nog een andere verandering die Big Data teweeg brengt en dat is de wijze van probleemoplossing. In het algemeen zijn we getraind om problemen op een systematische wijze op te lossen, waarbij de te volgen stappen in een logische volgorde worden afgewerkt. Men begint met een probleemdefinitie, kiest een analysemethode en daarbij behorende data, en genereert een (zoveel mogelijk geoptimaliseerde) oplossing. Dit is een recht-doorzee, lineaire oplossingsstrategie. Bij Big Data werkt de aanpak veelal minder lineair en meer iteratief. Dit is goed zichtbaar gemaakt in het *Cross Industry Standard Process for Data Mining* (CRISP-DM) dat veelvuldig wordt gebruikt in data mining-projecten (Shearer 2000). CRISP is gebaseerd op vijf projectfasen: (1) probleemanalyse gebaseerd op inzicht in bedrijfsprocessen; (2) inzicht in de mogelijkheden en beperkingen van beschikbare data; (3) prepareren van data voor analyse (selectie en schoning van data); (4) modelbouw en -oplossing; (5) evaluatie van oplossingen uit de modelanalyses op basis van de doelstellingen uit de probleemanalyse; en (6) gebruik van de modeloplossingen in concrete managementbeslissingen. Door de complexiteit van bedrijfsproblemen en de grote variëteit van bedrijfsdata ligt het niet voor de hand dat deze cyclus lineair wordt doorlopen. Integendeel, CRISP beveelt juist aan om tussen elk tweetal terugkoppelingen uit te voeren, zodat probleemanalyses kunnen worden bijgesteld op basis van beperkingen van aanwezige data, bij het prepareren van data rekening wordt gehouden met de eisen van analysemodellen, en modellen kunnen worden aangepast als blijkt dat de aard van de analyses niet geheel voldoet aan de eisen die beslissers daaraan stellen.

De data mining-processen volgen een overeenkomstige structuur. Allereerst dienen data voor analyse *bereikbaar* te worden gemaakt. Dit kan op veel verschillende manieren, bijvoorbeeld door het automatiseren en robotiseren van (administratieve) processen. De verschillende databronnen moeten met elkaar in verbinding worden gebracht door het gebruik van data warehouses. Die bronnen kunnen kwantitatieve datasets zijn, maar ook tekst, beeld- en geluidopnames. Vervolgens kunnen verschillende technieken worden gebruikt om data te *verkennen,* zoals descriptieve statistische analyses (gemiddelde, modus, mediaan en spreiding), visuele dataverkenningstechnieken met een graphical user interface (GUI) zoals OLAP (on-line analytical processing) en queries die kenmerken genereren van deelpopulaties (bijvoorbeeld: wat zijn kenmerken van onze klanten die na één aankoop niet meer bij ons terugkeren?). Een stap verder gaan *segmentatie-, classificatie- en clustering*-technieken die pogen samenhangen in datasets zichtbaar te maken. Zo kan men proberen te achterhalen welke factoren meespelen in het succes van opvallend goed-scorende verkoopmedewerkers. Deze technieken worden vooral *inductief* en *verkennend* gebruikt: men probeert verbanden op te sporen die voorheen nog niet bekend waren. Ten slotte zijn er de *associatie*-technieken die beogen veronderstelde verbanden te vinden. Te denken valt aan correlatie en verschillende varianten van regressieanalyse. Vooral in deze laatste categorie vinden we de gebruikelijke statistische technieken die in de huidige bedrijfseconomische opleidingen worden onderwezen.

Big data-uitdagingen

Uit dit overzicht blijken big data niet alleen nieuwe mogelijkheden te openen, maar ze stellen ons ook voor nieuwe uitdagingen. Door de grote variëteit en veranderlijkheid van data is een extra aandachtspunt de grote variatie in datadefinities en databetrouwbaarheid. Dit is een onderwerp waarop accountants goede diensten kunnen bewijzen: ze zijn immers getraind in het beoordelen

van consistentie en betrouwbaarheid van gegevens. Een andere uitdaging is het goed kiezen van bedrijfsproblemen die door big data kunnen worden opgelost. Uit de opsomming van verbeteringsmogelijkheden van McKinsey wordt duidelijk dat big data voor elke bedrijfsfunctie relevant kan zijn. Management accountants zijn bij uitstek in staat om deze specifieke bedrijfsproblemen hun plaats te geven in de totale planning en control-cyclus van de gehele organisatie. Dit kan helpen bij een goede prioriteitsstelling: welke problemen zijn het meest belangrijk en hoeveel middelen kunnen worden ingezet voor de oplossing hiervan? Ten slotte dient voldoende kennis in huis te zijn om de geavanceerde analysetechnieken op een verantwoorde manier toe te passen. Zo is het van groot belang om analysemethoden te gebruiken die passen bij het bedrijfsprobleem dat moet worden opgelost. Grote datasets en geavanceerde data-analysemethoden stellen ons voor weer nieuwe problemen. Datasets met veel waarnemingen vergroten de kans op het vinden van "false positives": verbanden en verschillen die statistisch significant zijn, maar vooral berusten op toeval. Bouwers en gebruikers van modellen dienen een scherp oog te houden voor het onderscheid tussen nonsense-verbanden en betekenisvolle relaties. Een ander bekend probleem is het risico van "overfitting": er is momenteel geen beperking meer aan de complexiteit van big data-modellen. Hierdoor kunnen ze met veel verschillende omstandigheden rekening houden. Modellen worden op een specifieke dataset ontwikkeld (de zogenaamde "trainingdata") zodat ze alle kenmerken van die dataset goed weergeven. Echter, een model met een goede representativiteit van trainingdata hoeft nog niet in staat te zijn generaliseerbare uitspraken over nieuwe gevallen te doen (de "use data"). Misschien hebben nieuwe gevallen wel kenmerken die niet in de trainingdata voorkomen. Een complex model dat goed rekening houdt met de specifieke kenmerken van de trainingdata en niet goed de doelvariabele van nieuwe, nu nog onbekende gevallen voorspelt is "overfitted". Ontwikkelaars moeten dus een afweging maken tussen betrouwbaarheid van modellen in de testfase en generaliseerbaarheid voor nieuwe gevallen. Hiervoor is inhoudelijke kennis van het vakgebied nodig en technisch inzicht in de werking van data analytics-methoden. Er zijn momenteel weinig goed opgeleide bedrijfskundigen die aan deze eisen voldoen.

Big data en accounting

Een laatste uitdaging geldt de beroepsgroep van accountants. Zoals uit het voorgaande duidelijk wordt is juist voor management accountants, controllers en auditors een grote taak weggelegd. Zij zijn bij uitstek in staat om de integriteit van gegevens te beoordelen, relevante gegevens te selecteren voor de interne besluitvorming en beheersing, en de betrouwbaarheid en volledigheid van de verslaggeving te bewaken. Zij zouden in de ontwikkelingen van big data in bedrijven een richtinggevende rol kunnen spelen. Het kan best zijn dat accountants nu al in specifieke bedrijfssituaties een belangrijke bijdrage leveren. Er bestaat echter weinig inzicht in de rol die accountants in deze ontwikkeling spelen. In de wetenschappelijke literatuur zien we verontrustend weinig ontwikkelingen op dit gebied. Een notoire uitzondering is een speciale editie van Accounting Horizons (twee zeer lezenswaardige bijdragen zijn Vasarhelyi et al. 2015; Warren et al. 2015). Mijn oproep is dat accountants hun plaats in de ontwikkeling van big data-toepassingen moeten innemen. Dit zal wel betekenen dat ze hun kennis op het gebied van data-analysetechnieken sterk dienen te verbreden en verdiepen. Hierin ligt ook een mooie uitdaging voor de opleidingen van accountants en controllers.

■ **Prof. dr. T.L.C.M. Groot** is hoogleraar Management Accounting aan de Faculteit Economie en Bedrijfskunde van de Vrije Universiteit Amsterdam.

Literatuur

- Brynjolfsson E, Hammerbacher J, Stevens B (2011) Competing through data: Three experts offer their game plans. McKinsey Quarterly 2011 (4 October): 36–67. https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/competing-through-data-three-experts-offer-their-game-plans
- Bughin J, Hazan E, Ramaswamy S, Choi M, Allas T, Dahlström P, Henke N, Trench M (2017) Artificial Intelligence: The next digital frontier? McKinsey Global Institute. https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx
- Hürtgen H, Mohr N (2018) Achieving business impact with data. Düsseldorf: Digital/McKinsey. https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20analytics/our%20insights/achieving%20business%20impact%20with%20data/achieving-business-impact-with-data_final.ashx
- Kyed A, Fillela G, Venugopal C (2013) The future revolution on Big Data. International Journal of Advanced Research in Computer and Communication Engineering 2(6): 2446–6451. https://www.ijarcce.com/upload/2013/june/44-Abdul%20Raheem-The%20Future%20Revolution%20on%20Big%20Data.pdf
- Provost F, Fawcett T (2013) Data science for business: What you need to know about data mining and data-analytic thinking. Se-

bastopol, USA: O'Reilly Media. http://shop.oreilly.com/product/0636920028918.do

- Shearer C (2000) The CRISP-DM model: The new blueprint for data mining. Journal of Data Warehousing 5(4): 13–32. https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf
- Vasarhelyi MA, Kogan A, Tuttle BM (2015) Big data in Accounting: An overview. Accounting Horizons 29(2): 381–196. https://doi.org/10.2308/acch-51071
- Warren Jr JD, Moffitt KC, Byrnes P (2015) How big data will change accounting. Accounting Horizons, 29(2): 297–707. https://doi.org/10.2308/acch-51069