

NIET ALLE DATAWETENSCHAPPERS HOEVEN ECONOMETRISTEN TE ZIJN

Data scientists vinden is geen sinecure. Academische onderzoekers uit positieve wetenschappelijke disciplines nemen steeds vaker plaats in data science-teams en worden meer en meer betrokken bij toonaangevende big data-projecten.

✍ DRIES VAN NIEUWENHUYSE (ONDERZOEKER HOGENT) KENNETH VAN DEN BERGH (BICC MANAGER CREDENDO)



LEES DIT ARTIKEL INDIEN U WILT WETEN:

- Welke **types data scientists** er zoal kunnen zijn
- Waarom **kennis van de business van cruciaal belang** blijft
- Hoe u een **data scientist** kunt aanwerven

SCHAARSTE

In een tijd waarin de vraag naar data-wetenschappers stijgt, deels gestuwd door de groeiende groep analytische serviceproviders en andere bedrijven die geld willen uitgeven aan analytische inzichten, houdt het aanbod geen gelijke tred. Traditionele data-wetenschappers, met diploma's en ervaring in hogere wiskunde, informatica en economische disciplines, blijven schaars. Bedrijven zoeken dus elders en onderzoekers van de positieve wetenschappen kunnen hier uitermate goed van pas komen.

In de meeste wetenschappelijke disciplines is statistische analyse van grote gegevenssets gebruikelijk. Biologen die central *place foraging* bestuderen gaan bijvoorbeeld na in welke mate vogels verder vliegen van hun nest om grotere prooi te vangen en minder ver voor kleinere. Zo optimaliseren ze hun energievoorziening. Dergelijke geografische modellen, op basis van grote sets van waargenomen gegevens, zijn perfect bruikbaar voor de optimale inplanting van *retail outlets* in een gegeven stad. Het is dus geen wonder dat heel wat gerenommeerde gegevenswetenschappers

ook actief zijn in management consulting en een wetenschappelijke carrière combineren met een zakelijke.

VOOR ELK WAT WILS...

Natuurkundigen zijn misschien goed in het beantwoorden van vragen over procesoptimalisatie omdat ze veelal omgaan met gigantische hoeveelheden procesgegevens, gedragsecologen zijn misschien beter in het omgaan met multivariate klantgegevens, sociologen zijn misschien beter in het omgaan met niet-parametrische statistieken zoals gewone tellingen en scores op enquêtes, geografen zullen logischerwijs veel beter zijn in ruimtelijke analyse ...

Afhankelijk van de bedrijfsvraag zijn verschillende statistische technieken belangrijk en worden ze meestal beheerst door verschillende wetenschappers. Van daar dat niet alle (data) wetenschappers econometristen hoeven te zijn om waar-de toe te kunnen voegen.

IT'S ALL ABOUT DATA

Verschiedende wetenschappelijke disciplines hebben nogal te maken met specifieke data. 's Werelds grootste deeltjesversneller van het CERN in Zwitserland is een goed voorbeeld van waarom natuurkundigen goede gegevenswetenschappers kunnen zijn. De deeltjesversneller genereert gegevens met een snelheid van 1 MB per botsing, en dergelijke gebeurtenissen gebeuren met een snelheid van ongeveer 600 miljoen per seconde. Natuurkundigen schrijven algoritmen om de gegevens in realtime

te doorzoeken om alleen potentieel interessante gegevens te verzamelen en op te slaan. Het is niet moeilijk om te zien hoe dergelijke ervaring zich kan vertalen naar commerciële big data-projecten. Parametrische statistische methodes leunen het nauwst aan bij dergelijke procesdata.

Sociologen verzamelen gedragsgegevens die meestal in een discreet aantal mogelijkheden resulteren. Een zeven-delige Lickert-schaal is daar een mooi voorbeeld van. Dit zijn bijgevolg geen continue data hetgeen de onderzoekers tot het gebruik van andere methoden dwingt. *Purchase propensity* of de kans dat iemand een bepaald item wenst aan te schaffen, gaat specifieke methodes aanwenden die nauwer verwant zijn aan pure kansberekeningen. Logistische regressiemethoden zijn hier dan weer populair.

Biologische wetenschappers zijn op hun beurt uitermate vertrouwd met data sets met bijzonder veel variabelen zodat datareductie hen helemaal niet vreemd is. KPI-reductie door middel van factor analyse is hier een mooi voorbeeld van en levert de nodige en voldoende sturingsinstrumenten op.

ALLEN DAARHEEN?

Ondanks hun ervaring met cijfers betekent dit niet dat natuurkundigen, bio-wetenschappers, sociologen of andere positieve wetenschappers automatisch goede kandidaten zijn voor *data science*-jobs. Werken voor een onderneming biedt veel uitdagingen die verschillen



van academisch onderzoek, zelfs als de metier van gegevensanalyse vergelijkbaar is. Als je veel mensen binnenhaalt die geen idee hebben van het bedrijf, wordt het moeilijk om mensen naar praktische oplossingen te begeleiden.

DATA SCIENCE TEAM-SPELERS GEZOCHT

Bij het zoeken naar nieuwe medewerkers proberen we dus best de persoonlijkheden en vaardigheden van sollicitanten te beoordelen als onderdeel van een team, evenals hun gegevensanalysevaardigheden. Idealiter wordt de datawetenschapper als onderdeel van zijn sollicitatie gevraagd om een bedrijfsvraagstuk op te lossen met de meest geschikte techniek. Vervolgens wordt hij gevraagd om de uitkomst hiervan voor te stellen aan een of meerdere belanghebbende partijen of

stakeholders. In de bedrijfswereld is de presentatie immers minstens even belangrijk als de analyse. En hoewel sommige academici uitblinken in dit soort uitdagingen, doen anderen dat niet. Ook is het ervaringsniveau van de meeste academische onderzoekers, meer dan nodig is voor de meest complexe bedrijfsvraagstukken. We geloven dan ook dat bedrijven niet alleen op zoek moeten naar *data science*-ervaring maar naar de juiste mate van nieuwsgierigheid en onderzoeksdrang. Veel bedrijven kunnen rondkomen met het zelf uitbouwen van een team van bedrijfsanalisten, datavisualiseringsspecialisten, softwareontwikkelaars en systeemarchitecten om het werk van *data scientists* te doen. Door gebruik te maken van de veelheid van betalende en gratis opleidingen in specifieke domeinen kunnen werknemers

aangemoedigd worden om de achtergebleven vaardigheden bij te spijkeren.

TO PHD OR NOT TO PHD?

Je hebt niet per se een doctoraat en vijf tot tien jaar ervaring nodig om effectief te zijn als gegevenswetenschapper. Data science is niet bepaald rocket science maar eerder een metier. Dus in plaats van opzettelijk te zoeken naar academische onderzoekers om vacatures op het gebied van gegevenswetenschap te vervullen, kunnen bedrijven en organisaties die gericht zijn op het genereren van inkomsten door middel van inzichten, hun netwerken beter benutten en openstaan voor mogelijke kandidaten met een scala aan ervaringen. Een nauwe samenwerking tussen academici en bedrijven laat een gedegen kruisbestuiving toe en komt in toenemende mate voor. Wanneer starten wij ermee? ◀