

The good, the bad and the data

Martin Heijnsbroek

7 oktober 2017

Om beurten schrijven vier experts uit de datawereld over een brandende kwestie. Deze week: Martin Heijnsbroek over de redenen waarom wij leken wat meer op data scientists zouden moeten letten.

Hoe verzin je het: fouten maken in zo iets simpels als het controleren van de boekhouding? En toch weten die aardige accountants het te presteren om elke dag weer in opspraak te komen, binnen en buiten Nederland. Het is natuurlijk triest en niet goed te praten, maar gelukkig betalen de grotere kantoren in Nederland zo nu en dan ook een echt fikse boete. Dus duiken de grote accountancykantoren al jaren massaal de consultancy en de wereld van data science in.

Je kunt je echter wel afvragen of we hier blij mee moeten zijn, want waar zit volgens jou het echte risico voor de maatschappij? Bij die accountant, of bij de data scientist, die over het paard getilde figuur met complexe modellen en kunstmatige intelligentie die ons het beloofde digitale land gaat binnenrijden?

Het antwoord lijkt me eenvoudig: het risico zit bij de data scientist. In tegenstelling tot wat mensen soms geloven, is data science - of het nu gaat over statistiek, machine learning of kunstmatige intelligentie - geen exacte wetenschap. Dat betekent dat twee data scientists die werken aan eenzelfde probleem vaak tot totaal verschillende output kunnen komen. Er bestaat eenvoudig niet zo iets als de beste voorspelling van het weer of het kredietrisico: het hangt er gewoonweg van af wat er in de toekomst gaat gebeuren.

Daarnaast zijn sommige technieken in mijn geliefde discipline relatief nieuw, en sterk groeiend in veel functies en sectoren - en daarmee zijn ze dus nog niet uitgekristalliseerd. We hebben vandaag de dag discussies over welke techniek goed is in welke situatie, hoe je in de praktijk om moet gaan met het ontwikkelen van je algoritme, het vastleggen van je code, en welke test wel of niet relevant is. En daarmee heeft de data scientist heel veel ruimte om zijn persoonlijke oordeel mee te laten wegen in het ontwikkelen en parametriseren van de algoritmes. Helaas krijg je daarbij ook zijn agenda, politieke voorkeur of persoonlijke bias mee cadeau.

Niet alleen is de discipline van data science minder uitgekristalliseerd dan accountancy, ik denk ook dat het risico van een op hol geslagen data scientist ook groter is. We ontwikkelen nu chatbots die straks met ons, of onze kinderen op Twitter en Facebook in gesprek gaan treden, en die we niet meer kunnen onderscheiden van een 'echte' mens. Tsja, die platform-economie - het klinkt leuk maar het geeft uiteindelijk enorm veel macht aan de partij die de meeste data en daardoor ook beste algoritmes heeft op haar platform. Wat zal het resultaat zijn van die winner-takes-all-dynamiek in data science?

Los van de spannende digitale ontwikkelingen, en die wat onervaren data scientist, ligt overigens het grootste risico bij jezelf, als FD-lezer. Door de complexiteit van de algoritmes en de onderliggende data, en vaak een totaal onbegrip van data science, vindt je het natuurlijk lastig om door te vragen bij de data scientist. Die overigens ook nog eens in staat is om zijn werk nog complexer te presenteren dan dat eigenlijk al is.

Dus stel je die data scientist niet de vragen die juist nodig zijn om de risico's in te dammen. Bijvoorbeeld: 'Waarom gebruiken we deze black-box machine learning-

techniek?' Of: 'Zijn er ook technieken die transparanter zijn?' En: 'Wat zijn de meest extreme uitkomsten van deze optimalisatie voor onze klanten?' Of nog simpeler: 'Hoe gaan we de nauwkeurigheid van dit model monitoren?'

Een beetje meer kennis en een beetje minder naïviteit bij de FD-lezer en vele managers is in mijn ogen het belangrijkste medicijn op dit moment. Net als bij die bankbestuurder in 2008, die met iets meer kennis van zaken had kunnen weten dat al die hypotheekportefeuilles niet zo gezond waren. Of, zoals Socrates al zei: het kwaad komt door onwetendheid, het goede door meer kennis.

Martin Heijnsbroek is managing partner bij MIcompany.