DOOR WOUTER KARMAN
EN RIAN DE JONGE

AI: weet (je) wat je doet!?

De democratisering van Artificial Intelligence (AI) is volgens Gartner één van de toptrends van de komende jaren[1]. Dat klinkt uitermate positief in onze westerse oren: wat kan er tegen zijn als AI voor iedereen beschikbaar komt? Toch is niet vanzelfsprekend dat democratisering tot een succes leidt. Denk aan de problemen die kunnen ontstaan in 'de democratie' zelf, bijvoorbeeld door de impact van slecht uitgevoerde referenda. Dus, om deze 'democratisering van AI' tot een goed eind te brengen, hebben we een goede aanpak nodig. Kernpunt daarin is 'weten wat je doet', en dat is niet vanzelfsprekend[2].

In de recente geschiedenis vinden wij een vergelijkbare situatie: technieken voor regressieanalyse en hypothesetesten zijn door pakketten als SPSS ter beschikking gekomen van een brede groep wetenschappers en studenten. Dit heeft een schaduwzijde laten zien: hoe vaak komt het niet voor, dat conclusies zogezegd worden 'bewezen' op basis van onderzoek waarin hypotheses worden getoetst met statistische toetsen, waarbij niet wordt voldaan aan de onderliggende vereisten om deze toets te mogen gebruiken? In feite kan dan geen conclusie getrokken worden, laat staan dat sprake is van een bewijs.

Het belangrijkste vraagstuk van democratisering is kort gezegd: welke mogelijkheden en verantwoordelijkheden kun je op welke manier leggen bij welke groepen mensen en met welke waarborgen en ondersteuning. Hetzelfde geldt voor democratisering van AI. In dit artikel focussen we ons op het deel van AI dat zich richt op data science (bijvoorbeeld machine-learningalgoritmen). We bespreken vijf aspecten die van belang zijn voor actuarissen die een rol (willen) spelen in deze democratisering.

Dr. W. Karman is Advanced Analytics Consultant bij PwC.

Drs. R. de Jonge AAG is Advanced Analytics Consultant bij PwC.

1. HET TEAM

Bovengenoemd SPSS-voorbeeld laat zien wat gebeurt bij een mismatch van de benodigde kennis, ervaring en vaardigheden van de modelontwikkelaar en datgene wat nodig is om de gebruikte technieken beheerst uit te kunnen voeren.

Bij AI spelen dezelfde risico's zelfs in grotere mate, omdat de technieken van zichzelf nog minder transparant en beheersbaar zijn dan de klassieke statistische methoden. Zorg daarom dat je team bestaat uit een diversiteit aan professionals die daadwerkelijk voldoende begrip hebben 1) van de werkelijkheid die je wilt modelleren, 2) van de data die je gebruikt om die werkelijkheid te representeren, 3) van de methodieken en bijbehorende vereisten en aannames, 4) van de mogelijkheden die er zijn om de uitkomsten op kwaliteit en bruikbaarheid in de praktijk te toetsen en 5) van de technologie om de analyses mogelijk te maken. Samen kunnen data scientists, actuarissen, econometristen, domeinexperts en computer scientists overzicht houden over het complexe samenspel tussen data, methodiek, uitkomsten en de werkelijkheid. Spelenderwijs leren omgaan met methodieken is natuurlijk toe te juichen, maar voorkom dat een 'speelgoedmodel' gebruikt wordt voor 'het echte werk'. Wees ook eerlijk naar jezelf om te bepalen of je de juiste kennis en ervaring hebt om de methodieken toe te kunnen passen op de data en om conclusies te trekken. Anders moet je wellicht een andere rol kiezen in het team wanneer het gaat om de ontwikkeling van modellen die concrete impact hebben in de praktijk op échte beslissingen over échte klanten. Naast diversiteit is ook kwaliteit en beschikbaarheid van de teamleden een bepalende factor: het vasthouden en kunnen aantrekken van nieuwe talenten is een zorg voor veel bestuurders[3].

2. DE DATA

Meestal wordt de nadruk gelegd op datakwaliteit: voldoende data van goede kwaliteit verkrijgen is lastig en het in kaart brengen van de kwaliteit ook. Een ander aspect van data is net zo van belang en misschien nog lastiger te beheersen: welke (ongewenste) patronen en vooroordelen zitten in de data opgesloten? Ook daarvoor is voldoende begrip van de gebruikte data essentieel. Wanneer bijvoorbeeld een regressiemethodiek wordt toegepast, worden de onderkende patronen direct zichtbaar gemaakt in het resulterende model. Dan ben je in staat om erop in te grijpen omdat je ongewenste afhankelijkheden van bijvoorbeeld ras en geslacht relatief eenvoudig kunt corrigeren in het model dat je toepast.

Zeker bij meer complexe machine-learningalgoritmes krijg je niet direct een 'formule' die inzicht geeft in de patronen die zijn herkend en worden toegepast. Ook zijn de soorten verbanden die in modellen worden meegenomen complexer van aard, daardoor minder voor de hand liggend en dus slechter vooraf te vermoeden. Dit betekent dat je niet direct op basis van de uitkomsten van het model kunt zien of het model patronen (bijvoorbeeld vooroordelen) meeneemt die ongewenst zijn voor toekomstige besluitvorming. Over dit soort risico's is het nodige geschreven, denk bijvoorbeeld aan het boek *Weapons of Math Destruction*[4].

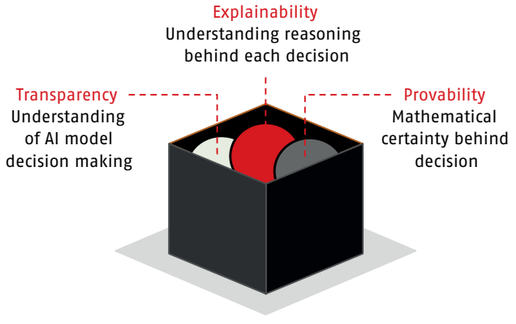
Hoewel wij de problemen volgend uit modellen die ongewenste vooroordelen weerspiegelen onderkennen, vinden wij niet dat dat mag leiden tot een demonisering van de algoritmen. Het is immers niet de schuld van het algoritme dat patronen in de data worden doorvertaald, maar de schuld van de mens zelf: het probleem ontstaat vanuit het (ongewenste) gedrag van de mensen die de data hebben gegenereerd en wordt in die gevallen door ontwikkelaars én gebruikers onvoldoende onderkend en gerepareerd. We moeten de resultaten kritisch tegemoet treden[5] en hebben voldoende expertise en inzicht nodig om dat te kunnen doen. Dit brengt ons terug bij het eerste aspect zoals hierboven benoemd: zorg voor voldoende expertise van zowel de algoritmen, als ook van de werkelijkheid en combineer die om het benodigde overzicht te kunnen behouden. De vraag is of je als actuaris beide rollen kunt vervullen, zoals we wel vaak proberen.

3. DE UITLEGBAARHEID AAN GEBRUIKERS

Transparantie en uitlegbaarheid zijn niet alleen nodig in modelontwikkeling, maar ook essentieel om enerzijds de kwaliteit van het model aan anderen te bewijzen en anderzijds vast te kunnen blijven stellen dat het model de benodigde kwaliteit behoudt tijdens gebruik. Het overgrote deel van bestuurders vindt uitlegbaarheid een voorwaarde om de adviezen van een model te vertrouwen[3]. Dit geldt des te meer als het model zelf beslissingen doorvoert.

Om toepassing van modellen in de dagelijkse praktijk mogelijk te maken, is vertrouwen van de gebruikers nodig. Hoeveel vertrouwen en wat daarvoor nodig is hangt af van het soort gebruik en van de gebruiker. Hoe dan ook zal je in enige mate moeten kunnen laten zien hoe de algoritmes werken, wat deze doen en aantonen dat je controle hebt over de werking en uitkomsten (zie figuur 1). Dit zal je zowel initieel bij inproductiename als ook continu tijdens de toepassing moeten kunnen en daarnaast zowel intern richting management, als ook extern richting toezichthouders, andere stakeholders en de samenleving. Zeker als je besluitvorming over (potentiële) klanten automatiseert in een model ben je verplicht aan die klant uit te kunnen leggen welke data op welke manier wordt afgewogen tot een oordeel.

Figuur 1:



Op verschillende manieren kun je anderen inzicht bieden in de werking van een algoritme. Meest voor de hand liggend is je te beperken tot data science technieken die goed interpreteerbaar zijn, zoals logistische regressie. Nadeel hiervan is dat dit complexere en minder intuïtieve technieken uitsluit, waardoor je mogelijk minder goede resultaten krijgt. Naast de gebruikelijke modelvalidatiemethodieken (gevoeligheidsanalyses, backtests, visualisaties, etc.) kun je onderbouwing zoeken via methodieken zoals LIME en SHAP[6]. Daarmee bepaal je op basis van simulatietechnieken welke parameterwaarden en andere factoren de meeste invloed hebben op uitkomsten en beslissingen. Naast al deze technieken is gebruik van visualisaties van groot belang om zelf na te gaan wat de data bevat en wat de uitkomsten representeren en ook om de complexiteit van de modellen te vertalen naar inzichten die begrijpelijk zijn voor domeinexperts en gebruikers. Dit geeft inzicht of er onverwachte patronen in het model zitten en of nader onderzoek daarnaar nodig is. Als je kunt uitleggen waarom het model bepaalde beslissingen maakt, is dat een waarborg dat dat ongewenste effecten vermeden worden en versterkt dat het vertrouwen in de toepassing van het model.

4. DE TOEPASSINGEN

Over dit aspect kunnen we kort zijn, maar dat maakt het niet minder belangrijk. Te vaak zien wij dat er op allerlei plekken in organisaties allerlei modellen worden toegepast, zonder dat er centraal overzicht wordt gehouden over welke modellen en algoritmen waar worden toegepast en op basis van welke data dat gebeurt. Gebrek aan overzicht en coördinatie kan leiden tot inefficiënties in gebruik van de beschikbare resources, maar ook tot inconsistenties tussen de modellen en aannames en daardoor tot suboptimale of verkeerde keuzes en slechtere uitlegbaarheid.

5. DE TECHNOLOGIE

Een draaischijftelefoon gebruik je niet om e-mail mee te versturen. Op dezelfde manier hangen toepassing, model en technologie van AI met elkaar samen: oude technologieën zijn in veel gevallen niet geschikt voor de ondersteuning van nieuwe algoritmen. Niet alleen de benodigde hoeveelheid rekenkracht is een factor, maar ook het verschil tussen piekbelasting en idle time. Ook wanneer in realtime resultaten nodig zijn, zoals bij pricing of bij fraudedetectie, is technologie nodig die voldoende schaalbaar is en is een implementatie nodig die volledig geautomatiseerd kan worden. Steeds vaker worden processen 'event driven' gemaakt, waarbij minder ruimte is voor handmatige berekeningsstappen. Het team heeft dus ook voldoende kennis nodig over bijvoorbeeld de dataopslag en het analyseplatform: er is geen data science zonder computer science.

HOE VOORWAARTS?

Als we al deze aspecten overzien, is het de vraag hoe ver we de democratisering van machine-learningtechnieken moeten laten gaan. Als zelfs iemand met een kwantitatieve achtergrond, zoals een actuaris, niet zomaar solo dit soort technieken kan toepassen, hoe kunnen we dan verwachten dat een gebruiker zonder die achtergrond dit wel doet zonder allerlei ongewenste en onbeheersbare risico's in onze organisaties te introduceren?

Tot slot, als actuaris zul je in aanraking komen met elk van deze vijf aspecten gezien de introductie van AI-technieken in onze sector. Waar de ene actuaris zich prettiger voelt bij de rol als gebruiker of sectorspecialist, heeft een ander voorkeur voor de meer technische kant. In al die rollen kan een actuaris floreren, zolang je zelf weet wat je kunt en je ervoor zorgt dat de rest van je team en organisatie jouw kwaliteiten aanvult: weet wat je doet en maak het inzichtelijk voor de anderen. ■

1 – https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018/

2 – zie ook https://hbr.org/2018/01/is-murder-by-machine-learning-the-new-death-by-powerpoint

3 – PwC CEO Survey 2018, https://www.pwc.nl/nl/actueel-en-publicaties/themas/economie /ceo-survey-2019.html

4 – https://weaponsofmathdestructionbook.com/

5 – zie ook het boek Hello World: How Algorithms Will Define Our Future and Why We Should Learn to Live with It http://www.hannahfry.co.uk/

6 – LIME: Local Interpretable Model-agnostic Explanations en SHAP: SHapley Additive exPlanations, zie bijvoorbeeld ook https://www.oreilly.com/ideas/testing-machine-learning-interpretability-techniques