

Het draait om de robuustheid van het systeem

Martin Heijnsbroek
zaterdag 22 oktober 2016

Ik hoop dat het algoritme het gedrag van onze klanten goed zal voorspellen'. Dit klinkt als een logische verwachting voor een manager die net geïnvesteerd heeft in het aannemen van een geweldige data scientist. Die manager wil hier op zijn minst een nauwkeurige voorspelling aan overhouden, en geeft dus het nieuwe talent een stevige prikkel om de accuraatheid van zijn algoritme aan te tonen.

Ons talent kan op zijn beurt de verleiding bijna niet weerstaan: zal hij alle alfa's in de directie even laten zien hoe fantastisch zijn model past op de data uit het verleden? Gelukkig houdt hij zich in, en begaat hij deze statistische doodzonde niet. Want hij weet dat een perfecte 'fit' op het verleden geen enkele garantie biedt voor de toekomst. Ook gaat hij niet rondstroom met indrukwekkende p-waarden, terwijl hij weet dat het opentrekken van zijn statistisch arsenaal met 100% zekerheid alle monden in de directie zal snoeren.

De data scientist houdt zich dus in, en ziet dat zijn argumenten over de betrouwbaarheid van het algoritme de verkeerde discussie gaat opleveren. Ten eerste zit een algoritme er altijd wel ergens naast. En ook al klopt deze vandaag aardig, dan nog kan dat binnen een dag veranderen. Als je het inzicht in het gedrag van klanten namelijk vertaalt naar nieuwe campagnes die zijn gericht op die klanten, dan kan je daarmee ook het gedrag van diezelfde klanten veranderen, en moet je het model weer aanpassen. Dat gebeurt ook als er iets verandert in de markt, bijvoorbeeld door een nieuwe concurrent: de voorspelkracht van je algoritme kan afnemen.

Daarom kun je je als manager beter afvragen hoe robuust het *systeem* is waarin het algoritme opereert dan hoe sterk het model op een bepaald moment is. Bijvoorbeeld door uit te vinden hoe goed het model zich automatisch aanpast op nieuwe data, of hoe de voorspelfouten worden gerapporteerd en besproken. Of hoe je ervoor kunt zorgen dat er geen 'garbage-in-probleem' ontstaat, door handmatige feeds van excelbladen die zeer foutgevoelig zijn.

Ik zou me ook afvragen hoe je de kennis over de ontwikkeling van het algoritme gaat documenteren, zodat je niet afhankelijk bent van één briljant individu. En, misschien wel het meest belangrijk voor het behouden van je baan: ervoor zorgen dat de data en modellering in lijn zijn met de data-wetgeving.

Het wordt dus hoog tijd dat managers begrijpen waar het echt om draait als we met betrouwbare modellen willen werken: namelijk dat het relatief eenvoudig is om eenmalig te scoren met een algoritme, in vergelijking met het bouwen van een robuust systeem waarmee je de concurrentie wil verslaan.

'Ik voorspel dat managers over tien jaar een stuk wijzer praten over algoritmen.'

'Ik voorspel dat de manager dan al is vervangen door een algoritme.'



*Martin
Heijnsbroek*

*Jan Fred
van Wijnen,
chef Morgen*