

Machine learning in de Zorg

“Kunstmatige intelligentie is de nieuwe elektriciteit. Net zoals elektriciteit vrijwel elke industrie in de vorige eeuw heeft getransformeerd, zal kunstmatige intelligentie vrijwel elke industrie in de huidige eeuw transformeren” (Andrew Ng, Stanford AI professor)

Judith Houtepen, MBA, AAG
Sjors Altemühl, MSc, AAG

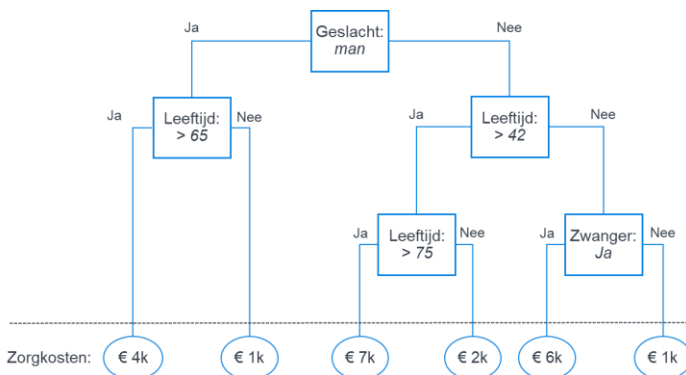


Kunstmatige intelligentie is een breed vakgebied. Het deelgebied van **machine learning** ligt echter dicht aan tegen technieken waar actuarissen al decennialang gebruik van maken. Actuarissen hebben dan ook een uitstekende basis om machine learning toe te voegen aan hun bestaande skill set.

Net als traditionele statistische modellen doen machine learning algoritmes een voorspelling over de waarde van een onzekere variabele (bv. zorgkosten) op basis van een reeks kenmerken. In formulevorm: $Y = f(X_i)$.

De meeste traditionele technieken gaan uit van vaststaande **lineaire relaties** waarvoor coëfficiënten worden geschat. Ook bij het risicovereveningsmodel geldt dat de compensatie die voor een verzekerde vanuit de verevening ontvangen wordt de som is van de normbedragen van verschillende aan gezondheid gerelateerde vereveningskenmerken. In formulevorm: $Y = \sum \beta_i X_i$.

Door de rigide structuur focussen traditionele technieken zich op ‘gemiddelde’ verbanden. Subtielere verbanden of kruiseffecten (bv. afwijkingen in kosten tussen kinderen en volwassenen voor specifieke aandoeningen) worden veelal genegeerd. ‘Tree-based’ machine learning algoritmes zijn voor actuarissen goed toegankelijk. Deze algoritmes schatten **beslisbomen** om tot een voorspelling te komen. Beslisbomen zijn niet-lineair in hun opzet en hierdoor goed in staat om kruiseffecten op te pakken. En in tegenstelling tot traditionele technieken bepalen deze algoritmes bovendien zelf welke kruiseffecten relevant zijn om op te nemen.



Voorbeeld van een simpele beslisboom

Om ook subtiele verbanden op te vangen wordt bij machine learning de schattingsprocedure een groot aantal keer herhaald (typisch 100 tot 1000 keer). Bij **random forest** gebeurt dit door met behulp van bootstrapping nieuwe versies van de

oorspronkelijke dataset te creëren. Elke schattingsprocedure kan leiden tot een andere beslisboom met andere voorspellingen. De uiteindelijke voorspellingswaarde is het gemiddelde van alle voorspellingen van de individuele beslisbomen.

Met machine learning wordt de rol van de **actuaris** cruciaal dan ooit. Machine learning is namelijk geen vervanging van traditionele statistische of actuariële technieken. Gaat het om pure voorspelkracht op individueel niveau dan is machine learning vaak superieur. Maar bij machine learning is het niet altijd eenvoudig om te herleiden hoe tot een bepaalde voorspelling wordt gekomen. Wanneer uitlegbaarheid belangrijk is zijn traditionele technieken dan ook vaak geschikter. Ook als het vooral gaat om voorspelkracht op geaggregeerd niveau dan is de toegevoegde waarde van machine learning soms beperkt. Het is de taak van de actuaire om hier keuzes in te maken.

Een plek binnen het bestaande zorglandschap waar machine learning waarde kan toevoegen zijn analyses naar **subpopulaties** (bv. regio's of aandoeningen). Het risicovereveningsmodel schat per vereveningskenmerk een normbedrag, maar deze normbedragen representeren enkel gemiddelde verbanden die gelden op populatieniveau. Wanneer deze normbedragen gebruikt worden voor het voorspellen van zorgkosten van hele specifieke subpopulaties kunnen er aanzienlijke verschillen ontstaan tussen de voorspelde en daadwerkelijke zorgkosten. De risicoverevening is ook niet voor dit doeleinde opgezet, maar is erop gericht om verzekeraars op totaalniveau te compenseren voor verwachte verschillen in zorgkosten als gevolg van de samenstelling van hun verzekerdenpopulatie. Ook laat de risicoverevening sommige kenmerken bewust buiten beschouwing.

Machine learning kan de betrouwbaarheid van analyses naar subpopulaties vergroten. Machine learning modellen hebben een flexibele structuur en komen tot voorspellingen op basis van microverbanden. Hierdoor is er minder risico dat zorgkostenvoorspellingen voor een specifieke subgroep worden beïnvloed door andere delen van het verzekerdenbestand, en stijgt de betrouwbaarheid van analyses naar subpopulaties. Machine learning is dan ook een nuttige tool om beter inzicht te krijgen in de zorgdynamiek van een populatie.



Volg ons op LinkedIn:

<https://www.linkedin.com/company/milliman-benelux/>