

Actuariële analyse in het AI-tijdperk: ChatGPT als innovatieve partner

Ruim een jaar geleden werd ChatGPT online gezet door OpenAI. Sindsdien heeft vrijwel iedereen van deze innovatie gehoord en velen maken er ook gebruik van. ChatGPT is een krachtig AI-model dat in staat is om natuurlijke taal te begrijpen en te produceren op een mensachtige manier.

Ook voor actuarissen geldt dat integratie van geavanceerde technologieën, zoals ChatGPT, de efficiëntie en effectiviteit ten goede kan komen. Dit artikel bespreekt twee mogelijke toepassingen: databewerking en code-optimalisatie. Om te onderzoeken hoe ChatGPT actuarissen kan ondersteunen in hun dagelijkse werkzaamheden, hebben wij een aantal testen gedaan om te beoordelen wat de kwaliteit en de praktische toepasbaarheid is van de antwoorden die ChatGPT teruggeeft. Hierbij vergelijken we versie 3.5 met 4 van ChatGPT. Versie 3.5 is vrij toegankelijk voor iedereen, terwijl ChatGPT-4 alleen beschikbaar is via een betaald abonnement. De antwoorden van GPT-4 zijn over het algemeen nauwkeuriger en relevanter, dankzij een beter contextueel begrip. Bovendien biedt deze versie meer functionaliteiten, wat hieronder verder toegelicht zal worden.

DATABEWERKING

Voor actuarissen is data-analyse een cruciaal onderdeel van hun werk. ChatGPT kan hen helpen bij het herkennen van patronen, het opruimen van gegevens en het doen van suggesties voor modellen. In ChatGPT is het mogelijk om zelf datasets in te voeren. Bij GPT-3.5 is dit bewerkelijker en beperkt tot bestanden met een kleine omvang. Je moet het bestand opslaan als een CSV-bestand, openen met Wordpad of Notepad, alles selecteren en vervolgens in het Chat-invoerscherm van ChatGPT plakken. GPT-4 biedt daarentegen de mogelijkheid om bestanden in te lezen.

We hebben tests uitgevoerd om te beoordelen hoe ChatGPT deze data kan interpreteren en analyseren. Allereerst creëren we een gegevensreeks op basis van een CBS-dataset met het aantal overledenen (gesplitst naar man/vrouw en totaal) op weekbasis in de jaren 2018 t/m 2022. De namen van de variabelen staan in de eerste rij van het bestand, de eerste kolom geeft de betreffende week aan. Hier zie je een deel van de set:

Perioden	Aantal Overleden totaal	Aantal overleden man	Aantal overleden vrouw
2019 week 52	3028	1471	1557
2019 week 53 (2 dagen)	858	412	446
2020 week 1 (5 dagen)	2245	1092	1153
2020 week 2	3364	1677	1687
2020 week 3	3157	1530	1627
2020 week 4	3046	1470	1576
2020 week 5	3164	1578	1586

Op de vraag hoe de toename van het aantal doden in de weken 13 t/m 15 van 2020 verklaard zou kunnen worden, geeft ChatGPT aan dat dit samenhangt met de eerste periode van de COVID-19-pandemie. Het hoge aantal COVID-19-besmettingen resulteerde in een toename van het aantal sterfgevallen.

In de dataset zijn sommige weken incompleet, zoals week 53. Om te testen of ChatGPT dit herkent en kan interpreteren, stellen we de volgende vraag: 'Kun je een nieuwe dataset maken die alleen de complete weken bevat?' ChatGPT genereert vervolgens een dataset die vrij is van onvolledige weken. Dit toont aan dat ChatGPT in staat is om deze situatie te herkennen en correct te interpreteren. Hierbij zijn de antwoorden van GPT-3.5 en GPT-4 van vergelijkbare kwaliteit.

Een andere dataset bevat fictieve polis- en ziektegegevens van verzekerden. Hierbij wordt elke rij geassocieerd met één maand exposure van een verzekerde. Bij het inlezen interpreteert ChatGPT dit als volgt:

'Het lijkt erop dat de dataset bestaat uit informatie over individuen verspreid over verschillende jaren en maanden. De kolommen bevatten het jaar van de polis, de maand van de polis, de leeftijd in maanden en een label 'ziek' dat aangeeft of de persoon ziek is (1) of niet (0).'

Vervolgens vragen we aan GPT-3.5 hoe de invalideringskans kan worden berekend. ChatGPT legt uit dat dit kan met de formule 'aantal zieken / totaal aantal individuen'. We vragen om deze kans daadwerkelijk te berekenen op basis van de dataset. ChatGPT antwoordt dat hiervoor het aantal waarnemingen nodig is. Blijkbaar is het nog niet duidelijk dat iedere rij in deze set 1 waarneming vertegenwoordigt. We vermelden dit daarom nogmaals expliciet aan ChatGPT. Volgens ChatGPT is de kans dan gelijk aan het aantal ziektegevallen gedeeld door het totaal aantal waarnemingen, namelijk 53/881. Hoewel de dataset inderdaad 881 waarnemingen bevat, is het onduidelijk waar het aantal 53 vandaan komt (aangezien de dataset 3 rijen heeft waarbij de kolom 'ziek' gelijk is aan 1). Zelfs als we expliciet aangeven dat het aantal zieken gelijk is aan het aantal waarnemingen waarbij 'ziek' gelijk is aan 1, blijft ChatGPT rekenen met 53 zieken. We hebben deze test opnieuw uitgevoerd met GPT-4, en daar begrijpt ChatGPT aanvankelijk ook niet hoe het aantal zieken kan worden bepaald. Als er echter expliciet wordt aangegeven hoe het aantal zieken uit de dataset kan worden afgeleid, geeft ChatGPT wel het juiste antwoord terug.

ChatGPT lijkt vaak goed in staat te zijn om de inhoud van datasets te herkennen maar kan deze niet altijd goed interpreteren. Vooral bij berekeningen zijn de antwoorden niet altijd betrouwbaar. GPT-4 presteert hierbij beter dan GPT-3.5.

OPTIMALISEREN CODERING

In de dagelijkse praktijk van een actuaaris vormt programmeren een substantieel onderdeel. ChatGPT is een nuttige tool voor het opzetten van scripts, efficiënter maken van code, en debuggen. Waar je rekening mee moet houden is dat de informatie die ChatGPT heeft niet altijd up-to-date is. De laatste update van GPT-3.5 was bijvoorbeeld in januari 2022. GPT-3.5 kan daarom geen recente packages in programmeertalen herkennen. Hoewel het optimaliseren van codes zeker mogelijk is met GPT-3.5, zijn de resultaten van de betaalde versie GPT-4 nauwkeuriger. Met GPT-4 heb je bovendien toegang tot de 'GPT store', waar je uit talrijke voorgeprogrammeerde GPT's kunt kiezen afhankelijk van de gewenste toepassing. Zo is bijvoorbeeld de Grimoire GPT gespecialiseerd in het optimaliseren van code.

We vragen Grimoire om te helpen bij het schatten van een model op basis van een arbeidsongeschiktheidsverzekeringsdataset. We geven een beschrijving van de dataset (inclusief de variabelen) en stellen voor om een logistiek regressiemodel te gebruiken. Grimoire reageert met het volgende plan van aanpak:

1 Data laden en voorbereiden:

- Laden van de dataset in een DataFrame
- Omzetten van datatypes indien nodig (bijvoorbeeld leeftijd als integer, verzekerd bedrag als float).

2 Data verkennen:

- Analyseren van de verdeling van de klassen (0 voor niet ziek, 1 voor ziek).
- Statistische samenvatting en visualisatie van de features (leeftijd, geslacht, beroeps categorie, wachtdagen, verzekerd bedrag).

3 Data schoonmaken en voorbereiden:

- Behandelen van ontbrekende of onvolledige gegevens.
- Normaliseren of standaardiseren van numerieke variabelen indien nodig.

4 Feature engineering:

- Creëren van extra features of aanpassen van bestaande indien van toepassing (bijvoorbeeld leeftijdscategorieën).

5 Logistische regressiemodel:

- Opsplitsen van de dataset in trainings- en testsets.
- Implementeren van een logistische regressiemodel met behulp van een geschikte bibliotheek (zoals scikit-learn in Python).
- Trainen van het model op de trainingsset.

6 Model evaluatie:

- Testen van het model op de testset.
- Berekenen van evaluatiestatistieken (zoals nauwkeurigheid, precisie, recall, F1-score, en ROC-curve).

7 Model optimalisatie:

- Eventueel uitvoeren van hyperparameter tuning om modelprestaties te verbeteren.

We vragen Grimoire om R-code te genereren voor de verschillende stappen. Grimoire levert per stap een code met een uitgebreide beschrijving, en geeft ook aan welke delen van de code eventueel zelf nog aangepast moeten worden.

Hoewel GPT-3.5 ook in staat is om dit te doen, zijn de antwoorden minder gedetailleerd. De antwoorden zijn soms ook minder passend (je moet vaker doorvragen voordat je een geschikt antwoord krijgt).

AANDACHTSPUNTEN BIJ HET GEBRUIK

Uiteraard is het essentieel om antwoorden kritisch te evalueren. Hoewel AI-modellen zoals ChatGPT context steeds beter begrijpen, hebben ze nog beperkingen in het herkennen van complexe nuances, wat kan leiden tot verkeerde interpretaties. ChatGPT is getraind op grote datasets, maar als die onjuistheden of vooroordelen bevatten, kan dit de output beïnvloeden. Bovendien opereren veel AI-modellen waaronder ChatGPT als black boxes, waardoor resultaten moeilijk te herleiden en valideren zijn.

Verder zal kunstmatige intelligentie binnenkort gereguleerd worden door de komst van de AI-act, de Europese Wet op Kunstmatige Intelligentie. Het doel van de AI-act is om misbruik van persoonsgegevens – als gevolg van inzet van AI – te voorkomen. High-risk AI-toepassingen, zoals systemen voor acceptatie en premiestelling bij levens- en zorgverzekeraars, worden aan strenge eisen onderworpen. Hoewel dit waarschijnlijk geen directe impact heeft op ChatGPT zelf, moeten actuarissen die AI in hun modellen gebruiken hier rekening mee houden.

CONCLUSIE

ChatGPT kan actuarissen op verschillende manieren ondersteunen in hun werk. Het is zeker nuttig om zelf te experimenteren hoe deze tool je kan helpen om het werk efficiënter in te richten. Hierbij is het cruciaal om de resultaten altijd zorgvuldig te evalueren. ChatGPT is een hulpmiddel, de expertise van actuarissen blijft onverminderd van belang! ■

Drs. I van Oostrom AAG is als actuaaris werkzaam op de afdeling Product en Pricing Inkomens bij a.s.r.

Dit artikel is geschreven uit naam van de AG-Commissie Verzekeringen.

