



Waardevolle modeldata

Voor mij is de actuaris het type collega dat geen genoeg neemt met onverklaarde afwijkingen, hoe klein ook. Elk resultaat wordt vanuit meerdere perspectieven geobserveerd en iedere waarde moet aantoonbaar in de brondata terug te vinden zijn. Als data engineer ben ik verantwoordelijk voor de aanlevering van data, zodat actuarissen daar hun analyses en wiskundige modellen op kunnen baseren. Hoewel hun scherpe oog angst kan inboezemen bij de partij die de data op een presenteerblaadje moet opdienen, is de fascinatie voor data ook onze gemene deler. De vraag is dan, wat is de exacte rol van business intelligence bij het creëren van waardevolle modeldata?

De bron van alle data is in de meeste gevallen een myriade aan applicaties, verspreid over de organisatie. De bronapplicaties produceren *data*, ook wel: 'een verzameling van ruwe feiten' (Introhive, 2024). Gedurende businessprocessen die plaatsvinden in deze applicaties, wordt data verwijderd, aangemaakt of gewijzigd. Voor ieder gegeven is slechts één applicatie leidend, ook al komt het gegeven in meerdere applicaties voor. Zo kan in meerdere applicaties vastliggen wat de echtelijke staat van een deelnemer is, maar moet er één applicatie worden aangewezen die als leidend geldt. De betekenis en waarde van een gegeven zijn dus niet dubbelzinnig. Het is belangrijk om dit te onderkennen, omdat er dan zekerheid is over de herkomst en juistheid van een data-element.

De output van actuariële modellen daarentegen bevat *informatie*, waarbij 'data is verwerkt, georganiseerd en in een betekenisvolle context is geplaatst' (Introhive, 2024). Dit kan verduidelijkt worden met behulp van een voorbeeld. Zo is de betekenisvolle context van de beoordeling 'negen' voor een examen belangrijk: in Nederland zou men zeer tevreden zijn met dit resultaat, maar op een Franse universiteit betekent het een onvoldoende.

Het nodigt uit om inzichten uit de filosofie te betrekken op het verband tussen (bron)data en informatie. David Hume (geboren in Edinburgh, 1711) was een belangrijk filosoof uit de tijd van de Verlichting en hoorde bij de stroming van het empirisme: kennis kan alleen ontstaan door ervaring (Hume, 1739). Hij werd onder andere bekend door zijn *is-ought* stelling. Volgens deze stelling kunnen normatieve conclusies nooit uit alleen feitelijkheden volgen. Een voorbeeld hiervan is de conclusie dat er in een bepaald jaar zo'n lage dekkingsgraad geldt dat er een aanvullende premie geheven mag worden. Deze conclusie kan niet alleen worden getrokken uit het feit 'de dekkingsgraad is 98

procent'. Hierbij is nog een normatieve premisse nodig: namelijk 'een dekkingsgraad van minder dan 100 procent is onvoldoende; het is nodig om aanvullende premie te heffen'.

Als we dit inzicht toepassen op data en informatie zal informatie niet uit louter feiten kunnen ontstaan, maar is er altijd een waardeoordeel bij betrokken. Het is interessant om na te gaan of dit ook het geval is bij de inputdata van actuariële modellen: hebben we het daar daadwerkelijk over een verzameling van ruwe feiten, of is er al sprake van een normatieve premisse?

Daarom vervult de afdeling business intelligence een essentiële rol tussen bronapplicaties enerzijds en actuariële modellen anderzijds. Data uit de verschillende bronapplicaties wordt door business intelligence verzameld en gecombineerd tot één datamodel en waar nodig aangevuld met statische data of data van buiten de organisatie. Het is het datamodel dat business value heeft en gebruikt kan worden door bijvoorbeeld actuaariaat.

In het datamodel wordt de betekenis van de feiten en de relaties tussen de feiten vastgelegd. Zo kan een persoon worden gedefinieerd aan de hand van een persoonsnummer. Een persoon heeft een relatie tot een arbeidsverhouding waarbij het persoonsnummer gerelateerd kan worden aan de attributen die een arbeidsverhouding uniek maken. Het is van belang dat in de breedte van de hele organisatie voor ieder inzicht dezelfde definitie van het feit en de relatie wordt gebruikt. Hieraan gelieerd is de term *single version of the truth* (King, 2003): dit principe zorgt ervoor dat analyses op basis van dezelfde data altijd tot consistente inzichten leiden. In de datamodellen van business intelligence wordt data geordend volgens de geldende business logica. Er vindt tevens filtering plaats op de gegevens die ertoe doen voor de organisatie: alleen gegevens die voor enig doel interessant zijn, worden bewaard. Bij een pensioenuitvoeringsorganisatie (PUO) zullen alleen de gegevens over een persoon worden vastgehouden die betrekking hebben op diens pensioenkeuzes, arbeidsverhouding, et cetera. Een modellenbureau zal deze interesse niet delen met de PUO en zal eerder attributen opnemen als lengte, of haar- en oogkleur. Het datamodel past dus bij de interesse van de organisatie en ontleent daaraan zijn waarde en business value.

De ordening en filtering worden in het datamodel expliciet gemaakt ten opzichte van hoe het in de bronapplicaties staat. Er is geen sprake meer van losse tabellen, maar van nauw verweven entiteiten met onderlinge relaties. Dit is een logische tussenstap gezien de enorme hoeveelheden data die in businessprocessen worden voortgebracht; er is behoefte aan vaststelling en samenhang.

De *is-ought* stelling lijkt daarmee ook van toepassing op de data input voor de actuariële modellen. De informatie die door actuariële modellen wordt voortgebracht, is niet alleen afgeleid van ruwe feiten, maar ook van normatieve premissen, die opgeslagen zijn in het datamodel van de (business intelligence) organisatie. Zo blijkt dat business intelligence niet alleen data ordent, maar ook betekenis toevoegt aan de modeldata van de actuaris. ■

Referenties

- Hume, D. (1739). *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects*. Oxford University Press (2nd ed. revised by P. H. Nidditch, 1975). <https://www.gutenberg.org/ebooks/4705>
- Introhive. (2024, July 1). *The Difference Between Data, Information & Intelligence*. <https://www.introhive.com/blog/difference-between-data-information-intelligence/>
- King, J. (2003, December 22). *Business Intelligence: One Version of the Truth*. *Computerworld*. <https://www.computerworld.com/article/1323267/business-intelligence-one-version-of-the-truth.html>

K. Vermeulen MSc is data engineer bij PGGM.

