

Beter ingericht analyseproces levert hoger informatieniveau

# Wegwijzers naar Intelligence

Peter van Til en Jeroen Vuurens

**Business Intelligence belooft ons informatie die leidt tot 'betere' business decisions en competitive advantage. Maar kan die belofte worden waargemaakt? Het grotere potentieel van een duurdere BI-oplossing ten opzichte van rapportage-tools wordt zelden benut.**

BI-oplossingen worden vaak ingezet voor de gebruikelijke management-rapportage. Het probleem ligt bij de uitvoering van het BI-proces. Het is moeilijk intelligence uit een datawarehouse te halen. Een gestructureerd proces met technieken om *information overload* terug te dringen, helpt daarbij.

## De meerwaarde van een hoger intelligence-niveau is dat je de juiste vragen kan stellen

"Mensen zijn beperkt in hun vermogen om absolute beslissingen te nemen. Dit komt doordat mensen over het algemeen niet meer dan 5 tot 9 'chunks of information' tegelijk kunnen onthouden." (George A. Miller, *The magical number 7, plus or minus two: Some Limits on Our Capacity for Processing Information*, *The Psychological Review*, 1956, vol. 63 pp. 81-97.) Volgens het onderzoek van Miller kunnen mensen cognitief grote hoeveelheden getallen niet aan. Dat uit zich als mensen proberen rationele beslissingen te nemen over veel data. Onderzoek wijst uit dat we bij teveel informatie meer succesvol zijn met beslissen op gevoel dan met rationeel denken. In afbeelding 1 staat een klein voorbeeld waarin het al niet meevalt om snel te zien in hoeverre resultaten opmerkelijk te noemen zijn. Dat probleem wordt groter naarmate er meer getallen staan.

Gebrekkig gebruik van instrumenten levert al snel teveel cijfers op en zorgt daarmee voor *information overload*. Dat maakt de productie van *actionable intelligence* voor de hogere managementlagen complex en tijdrovend. Goed gebruik van instrumenten kan echter helpen om grote hoeveelheden data verwerkbaar te maken.

## Actionable Business Intelligence

Echte intelligence staat zelden op scherm of papier. Het is de kennis en ervaring van de manager die hem in staat stelt de informatie te interpreteren en daaruit conclusies te trekken. Hoe lager het niveau van de management-informatie, hoe meer de intelligence gecreëerd wordt in het hoofd van de manager. De afhankelijkheid van de organisatie van de manager wordt groter bij een lager informatieniveau. Dit remt bovendien de ontdekking van nieuwe intelligence af. Hoe hoger het intelligence-niveau dat BI kan leveren, hoe kleiner die nadelen worden.

De laatste jaren neemt de vraag naar een hoger informatieniveau toe. Men wil actionable BI: intelligence waar je wat mee kunt. Neem geen genoegen met: "In 2005 hebben we een recordomzet geboekt", maar verlang antwoord op de vraag achter deze vraag: "Waardoor wordt dat veroorzaakt?" Door het zoeken naar de oorzaak komen we intelligence op het spoor, gebeurtenissen die hebben bijgedragen aan het succes. Dat is het soort inzicht waarmee we betere beslissingen kunnen nemen in de toekomst. Een beter ingericht analyseproces kan een hoger intelligence-niveau leveren en tegelijk efficiënter zijn.

## Information overload terugbrengen

Kritisch bekeken zeggen losse meetwaarden weinig, zoals "We hebben in 2004 voor product 302 een omzet gehaald van 4,536 Mln". Een meetwaarde krijgt pas betekenis als we het in een context beschouwen. In een complexe context zoals afbeelding 1,

	Product 302	Product 600	Product 601	Product 700
	T-Shirt	Sweatshirt	Sweatshirt	Shorts
2004	4536	5472	5184	6660
2005	8400	30816	30240	17280
2006	16968	37152	30528	44100

Afbeelding 1: Information Overload.

	Product 302		Product 600		Product 601		Product 700	
	T-shirt		Sweatshirt		Sweatshirt		Shorts	
	echt	verwacht	echt	verwacht	echt	verwacht	echt	verwacht
2004	4536	2753	5472	6762	5184	6072	6660	6265
2005	8400	10929	30816	26839	30240	24103	17280	24866
2006	16968	16222	37152	39839	30528	35777	44100	36910

**Afbeelding 2:** Sheet met verwachte waarden.

	Product 302		Product 600		Product 601		Product 700	
	T-shirt		Sweatshirt		Sweatshirt		Shorts	
	echt	verwacht	echt	verwacht	echt	verwacht	echt	verwacht
2004	4536	2753	5472	6762	5184	6072	6660	6265
2005	8400	10929	30816	26839	30240	24103	17280	24866
2006	16968	16222	37152	39839	30528	35777	44100	36910

**Afbeelding 3:** Kruistabel met wegwijzers.

gaan mensen automatisch schatten omdat de context teveel getallen bevat. We kunnen de context vereenvoudigen door elke meetwaarde tegen één referentiegetal af te zetten. In afbeelding 2 is dat gedaan door de verwachte omzet ernaast te zetten. De verwachte omzet per product kan statistisch worden berekend aan de hand van het percentage ten opzichte van de totale omzet van dat product en het percentage ten opzichte van de totale omzet in die periode. Op basis van die twee kansen kan de verwachte waarde worden uitgerekend:

Verwacht(Product 302, 2004) =

$$P(\text{Product 302}) * P(2004) * \text{totale omzet.}$$

We kunnen de information overload verder terugbrengen door visualisatie. In een analyse-tool kunnen de feiten die sterk afwijken van de verwachte waarde, worden gekleurd zoals in afbeelding 3. Met een calculatie en een opmaakinstelling hebben we als het ware *wegwijzers* gemaakt. De analist ziet in één oogopslag waar de betere en slechte resultaten zitten.

Er zijn verschillende formules mogelijk om wegwijzers te maken. Beter dan procentuele of absolute afwijking van de verwachte waarde, is het gebruik van een formule om de *associatie* te berekenen zoals in afbeelding 4. Hoe verder de associatie van de 0 verwijderd is, hoe significanter de meetwaarde afwijkt van de verwachte waarde.

In de statistiek spreekt men van associatie als er een samenhang blijkt te zijn tussen variabelen op nominale schaal. De associatie kan in één cijfer worden uitgedrukt door een berekening uit te voeren over de resultaten over die twee variabelen. De uitkomst is een associatiecoëfficiënt die kan variëren tussen 0 (onafhankelijk) en +1 of -1 (volledige positieve/negatieve samenhang). Hoe verder

de associatie verwijderd is van 0, hoe kleiner de kans op toeval en hoe waarschijnlijker dat er een reden voor de afwijking is.

Wat is het rendement van uw Business Intelligence-proces? Levert het Actionable BI op? Is het efficiënt? We gaan structuur brengen in het BI-proces, zodat de efficiency en effectiviteit worden verhoogd. Dat doen we door het analyseproces te beschouwen in drie delen:

- Signaleren, het opmerkelijkste resultaat vinden (bij een bepaalde invalshoek);
- Verklaren, verbanden opzoeken die het signaal (mede) kunnen veroorzaken;
- Concluderen, de verklaringen omzetten naar conclusies.

## Signaleren

Het startpunt is een kruistabel zoals in afbeelding 4 over twee dimensies van een bepaalde *business value*. De business value kan bijvoorbeeld worden gekozen aan de hand van een doel, succesfactor of KPI. We gebruiken associaties als wegwijzers om de meest significante afwijking naar boven of beneden te markeren.

In afbeelding 4 is te zien dat producten 600 en 601 beide relatief hoge associaties laten zien in 2005. Beide producten zijn 'Sweatshirts'. Kunnen we nu het beste deze twee producten samen nemen (en toewerken naar een conclusie voor Sweatshirts) of de twee producten apart onderzoeken? Door de twee producten samen te nemen tot het producttype sweatshirts krijgen we een associatie van 0.18 (zie afbeelding 5). Hoe hoger de associatie, hoe hoger de afwijking van de verwachte waarde en hoe hoger

	Product 302		Product 600		Product 601		Product 700	
	T-shirt		Sweatshirt		Sweatshirt		Shorts	
	omzet	associatie	omzet	associatie	omzet	associatie	omzet	associatie
2004	4536	0.08	5472	-0.04	5184	-0.03	6660	0.01
2005	8400	-0.07	30816	0.08	30240	0.12	17280	-0.15
2006	16968	0.02	37152	-0.05	30528	-0.10	44100	0.13

**Afbeelding 4:** Associaties.

	Sweatshirt		Rest	
	associatie	omzet	associatie	omzet
2005	0.18	61056	-0.18	76944
Rest	-0.18	78336	0.18	168406

**Afbeelding 5:** Clusteren.

dus het potentieel. Dus samennemen en zoeken naar verklaringen voor Sweatshirts biedt betere kansen.

Over het algemeen zoeken we bij het signaleren naar de hoogste associatie. Voor het vinden van de hoogste associatie kan men drillen, clusteren en splitsen. Is de associatie op een ander niveau of bij een categorie-indeling hoger, dan werken we daar mee verder. Associaties wijzen de gebruiker in de juiste richting. Door te clusteren kan het overzicht na elke stap worden vergroot (het gevonden signaal wordt afgezet tegen de rest). Een eerlijke vergelijking kan alleen worden gemaakt als de overzichten van alle stappen dezelfde data omvatten. Clusters worden dus steeds afgezet tegen de rest om geen data uit te filteren.

## Verklaren

Bij het verklaren proberen we verbanden op te sporen met het gevonden resultaat. Waardoor wordt de hogere omzet van Sweatshirts in 2005 van afbeelding 5 veroorzaakt? Hoe weet je dat je de beste verklaringen hebt gevonden? We zetten het gevonden signaal (hogere omzet in 2005 van Sweatshirt) op één as. Vervolgens toetsen we andere dimensies door die op de andere as te zetten. We gebruiken weer associaties om mogelijke oorzaken/verbanden snel op te sporen.

Indien tussen hoge associaties een verband zichtbaar lijkt, kan dat worden getoetst door die samen te nemen (clusteren).

In afbeelding 6 vinden we de positieve associaties in naastgelegen maanden. Door die samen te nemen in afbeelding 7 zien we de associatie stijgen. Een gegronde conclusie waarom er juist in de periode Aug-Dec meer is verkocht levert meer potentiële business value dan elke maand apart. Op een soortgelijke manier kan men proberen meer medeorzaken te traceren.

## Concluderen

De analyse houdt echter niet op na het 'ophoesten' van getallen. De behoefte aan informatie op een hoger intelligence-niveau is een behoefte om beslissingen te kunnen nemen. De zoektocht is pas klaar als we een conclusie, een aannemelijke oorzaak hebben gevonden. Hopelijk hebben we tijdens het verklaren verschillende samenhangen gevonden die we in hypothesen kunnen bundelen en vervolgens onderzoeken. In deze case bleek het te gaan om een uitzonderlijk zacht najaar en winter in de Verenigde Staten, waardoor meer mensen in een sweatshirt over straat liepen zonder

	Aug-Dec		Rest	
	associatie	omzet	associatie	omzet
2005 Sweatshirt	0.24	47520	-0.24	13536
Rest	-0.24	145408	0.24	178278

**Afbeelding 7:** Verklaringen clusteren.

winterjas. Er werd ook een hoge associatie met twee uitgevoerde kleuren gevonden. Vermoedelijk is de modegevoeligheid hoger als men geen jas over de kleding draagt.

Zoals in dit voorbeeld te zien is wordt het trekken van bruikbare conclusies gevoed door de gevonden mogelijke verklaringen. Indien er ook feiten over het weer en de mode in het datawarehouse staan, dan zouden die samenhangen waarschijnlijk ook als mogelijke verklaringen uit de analyse komen. Indien onze analysemogelijkheden echter zo beperkt zijn dat we dat niet in één sheet op bovenstaande manier kunnen onderzoeken, dan kunnen we de kleuren en het tijdstip in elk geval apart onderzoeken.

Antwoord op de vraag achter de vraag staat echter lang niet altijd in het datawarehouse. Misschien bevat ons datawarehouse geen data over weer en mode, of werd de hogere verkoop in werkelijkheid vooral veroorzaakt door het faillissement van een concurrent of een succesvolle reclamecampagne? Een associatie hoeft niet altijd een oorzakelijk verband te zijn. Men doet er goed aan de bevindingen voor te leggen aan deskundigen, die de verklaring kunnen verifiëren en in een context plaatsen. De meerwaarde van een hoger intelligence-niveau is dat je de juiste vragen kan stellen. En je hebt je opengesteld voor het vinden van onverwachte verbanden die kunnen leiden tot nieuwe inzichten en betere beslissingen.

## Conclusies

Het is zeker mogelijk om met de huidige BI-tools tot een hoger intelligence-niveau te komen. Een goede aanpak is echter noodzakelijk voor een kwalitatief goed en efficiënt analyseproces. Een combinatie van tools kan daarbij handig zijn, zoals het signaleren met een OLAP-tool en het verklaren met een datamining tool. Het doorontwikkelen van software biedt hopelijk in de toekomst betere ondersteuning. Ooit zal de manager zelf de knoppen bedienen en zelf actionable intelligence kunnen opsporen. Tot die tijd is het een klus voor een BI-analist die met de juiste technieken betere resultaten kan bereiken.

### Peter van Til en Jeroen Vuurens

Peter van Til ([peter.van.til@sogeti.nl](mailto:peter.van.til@sogeti.nl)) is BI-consultant bij Sogeti Nederland B.V. Jeroen Vuurens ([j.b.p.vuurens@hhs.nl](mailto:j.b.p.vuurens@hhs.nl)) is docent Bedrijfskundige Informatica aan De Haagse hogeschool.

	December		November		Oktober		September	
	associatie	omzet	associatie	omzet	associatie	omzet	associatie	omzet
2005 Sweatshirt	0.19	9504	0.10	10656	0.01	7776	0.06	9216
Rest	-0.19	11760	-0.10	30162	-0.01	37824	-0.06	33228

**Afbeelding 6:** Verklaring zoeken.