



Inbound marketing, webanalyses en mestfraudeurs

De groeiende macht van Analytics

Hans Lamboo

Begin september vond in Kopenhagen het congres A2010 plaats, een congres over Analytics georganiseerd door SAS. Wat direct de conclusie kan zijn is dat de adoptie begint toe te nemen en het domein Analytics een flinke professionaliseringsslag doormaakt, zowel aan de gebruikers- als aan de softwarekant.

Het domein Analytics behoort tot de wereld van Business Intelligence. Op zich niet merkwaardig, omdat het gaat om ondersteuning van het besluitvormingsproces, globaal de functie van BI. Volgens de oude definitie van TDWI is *reporting* de laagste vorm van Analytics, daarboven bevinden zich werkgebieden zoals datamining, forecasting en predictive analysis die op dit moment meestal worden aangeduid als Advanced Analytics.

Er is een aantal grote verschillen tussen reporting en Advanced Analytics. Bij reporting is de inbreng van Analytics beperkt; het gaat meestal om het aanbieden van materiaal aan de eindgebruiker die zelf de analyse doet door te goochelen met meestal financiële gegevens – om inzicht te krijgen in ontwikkelingen in het verleden en op basis daarvan besluiten te nemen die de toekomstige ontwikkelingen positief moeten beïnvloeden. Bij Advanced Analytics ligt dat anders: het gaat om het kijken naar toekomstige ontwikkelingen en het vinden van patronen of samenhang, die bepaalde beslissingen *mogelijk* maken. De software analyseert, de gebruiker besluit. Het doel van de analyse, de hoeveelheden data, het aantal variabelen, en de complexe bewerkingen die de data moeten ondergaan aler een betrouwbare uitkomst is verkregen, vragen om volkomen andere competenties van de analisten dan bij reporting. Kennis van de data uiteraard, kennis van statistiek, van econometrie, marketing, producten en klanten, maar ook wiskundige kennis en een vleugje psychologie: het ontwikkelen van een betrouwbaar model is een taak waar niet iedereen zich zomaar aan kan wagen. In het geval van reporting is sprake van een grote groep gebruikers die van hoog tot laag verspreid is door de gehele onderneming – en tegenwoordig ook vaak daarbuiten – waarbij de gebruikers verschillende rollen kunnen spelen en elk voor zich hun conclusies uit de informatie zullen trekken; voor Advanced Analytics is meestal een speciale groep of afdeling ingericht, die zich dagelijks bezig houdt met

het ontdekken van patronen, het zoeken naar samenhang en het ontwikkelen van modellen, meestal op klantgegevens. En in tegenstelling tot reporting: op dit moment vrijwel altijd nog voor intern gebruik. Er zijn inmiddels websites die analytics gebruiken om de klant naar de juiste plaats te leiden, of ruwe data en een analytische functionaliteit beschikbaar stellen aan de bezoeker. Dat aantal is groeiende.

Aanjagers

Er zijn drie aanjagers voor Advanced Analytics. In de eerste plaats het beroemde boek 'Competing on Analytics' van Tom Davenport, waarin hij aantoont dat de enige informatie die een bedrijf exclusief bezit, de gegevens over de eigen klanten zijn; slimme, zorgvuldige analyse kan tot groot competitief voordeel leiden. De tweede driver is de toename van beschikbaarheid van data, onder meer door het internet; de derde is de noodzaak van het vinden van onzichtbare verbanden, zoals bij fraudedetectie.

Veel organisaties hebben moeten constateren dat de *Outbound* marketingcampagnes – waar buiten de muren van de onderneming gezocht wordt naar nieuwe klanten – steeds minder opleveren. Tom Davenport indachtig worden de pijlen meer en meer op de eigen klantenkring gericht, de *Inbound* campagnes. Een van de targets daarbij is het voorkomen van *churning*; het weglopen van de klant. Er zijn modellen ontwikkeld die vrij nauwkeurig kunnen bepalen of een klant het profiel van een *churner* heeft. In feite kan de gehele klantenkring worden opgedeeld in vier segmenten: de klanten die blijven, ongeacht of je actie onderneemt of niet en aan de andere kant de klanten die opzeggen – wat je ook doet. Daartussen zit de groep die een duwtje in de rug nodig heeft, en een groep klanten die opzeggen zodra je contact met ze zoekt, omdat ze daar niet van gediend zijn. Tricky dus. In de eerste en laatste groep hoeft je geen tijd en

energie te steken, je zoekt de groep die een incentive nodig heeft om klant te blijven, met uitsluiting van de groep die juist vertrekt bij contact. Bij zo'n model wil je de nettogroep dus scherp in beeld brengen. *Net lift*, heet dat in Analytics-termen. De volgende praktijkvoorbeelden van toepassing van Advanced Analytics geven een indruk van het belang, de rol die IT speelt en de professionaliteit.

ING Retail

Hans de Wit is Researchmanager bij ING Retail op de afdeling Customer Intelligence. Hij houdt zich vooral bezig met klantgedrag en innovatie. "Onze afdeling Customer Intelligence werkt met bijna 90 mensen binnen de afdeling Marketing. De meesten houden zich bezig met multi-campaign management, gebruikmakend van de analytics die de onderzoekers aanleveren." Voor zo'n churningmodel draaien hij en zijn mensen de hand niet om. "Maar het is een vak apart. Het gaat er vooral om: hoe en wanneer pas je ze toe." Bij ING Retail is men op dit moment meer gefocust op Next Best Action, waarbij afnemers van een bepaald product die hetzelfde profiel hebben als kopers van een ander product dat product op elk contactmoment aantrekkelijk krijgen aangeboden.

"Analytics spelen bij ING al een belangrijke rol en dat zal alleen maar toenemen," voorspelt De Wit. "ING heeft zo'n 8 miljoen klanten. Dat betekent dat we steeds meer analytische modellen zullen moeten ontwikkelen. We zetten nu de eerste stappen om dat geautomatiseerd te kunnen doen."

De analyses vinden plaats in een daartoe speciaal ingericht Oracle Datawarehouse, met daarin gegevens over alle klanten, hun producten, hun limieten, hun betalingsgedrag, de historie vanaf 2000. Sommige gegeven worden dagelijks, andere wekelijks of maandelijks bijgewerkt. Bij elkaar goed voor zo'n 2 Terabyte aan data.

"Mocht ik nieuwe data nodig hebben dan geef ik een briefing aan mijn collega's van Informatie Management, die in overleg met IT ervoor zorgen dat het datawarehouse aan onze wensen wordt aangepast." Daarbij wordt niet over een nacht ijs gegaan, want voor de geleverde IT-diensten krijgt De Wit gewoon een rekening. "We hebben de beschikking over een omgeving waar we met nieuwe variabelen kunnen experimenteren. Op basis daarvan besluiten we of we ons datawarehouse laten aanpassen. We hebben dan immers de business case." De samenwerking tussen de researchers, informatie managers en IT verloopt uitstekend. Momenteel wordt gewerkt aan de bouw van een nieuw datawarehouse voor ING Zakelijk, en het ligt in de bedoeling over enige tijd over te stappen op een grid. "Ik ben erg tevreden over de responstijden. IT houdt ook een oogje in het zeil waar het de netheid van onze query's betreft."

De Wit gebruikt zelden externe data, behalve bijvoorbeeld bij zakelijke kredietrisico's. "Daar hebben we een model voor ontwikkeld; je wilt een klant die er slecht voorstaat immers geen aantrekkelijk aanbod doen voor een grote lening." Bij sommige analyses



Dr. Natasha Zharinova, hoofd Business Analytics bij Marktplaats.nl:
"We werken in een snel voortrazende en veranderende omgeving".

worden ook gegevens uit Sociale Media ingezet. "Die verzamelen we niet zelf, dat doet een externe partij. We zijn wel bezig om Text Mining op poten te zetten, maar data die opgeslagen worden binnen de ING moeten aan allerlei eisen voldoen, waardoor dat erg duur is. De externe partij haalt de data dus van het web en slaat die op, zodat wij onze analyses erop kunnen loslaten."

Marktplaats.nl

"Analytics zijn van vitaal belang voor ons bedrijf en het zal de komende jaren alleen maar nog belangrijker worden," zegt Dr. Natasha Zharinova, hoofd Business Analytics bij Marktplaats.nl. Of beter gezegd: bij eBay's onderdeel Classifieds Group, waar naast Marktplaats.nl nog soortgelijke websites in 20 landen als Denemarken, Italië, Australië en Canada worden geëxploiteerd. Bij een internetbedrijf gaat alles snel, zowel ten goede als ten kwade. Monitoring was dan ook de eerste stap om te kunnen zien of alles gaat zoals verwacht. Vervolgens stelde men de vraag: wat wordt er eigenlijk precies verwacht? Wat mógen we verwachten? Dat is waar Analytics ten tonele verscheen.

"Onze informatie komt niet binnen als mooie rechte lijn, dat is logisch. We hebben te maken met seizoensinvloeden, de dag van de week, tijd van de dag. Het aantal geplaatste advertenties kan per dag 100 procent verschillen, het kan een regenachtige zondag zijn of een drukke dinsdagmiddag in de vakantie. Als het monitoringsysteem een alert gaf, wilden we graag zeker weten of het een positief of negatief alert was, of we blij konden zijn of

ons juist zorgen moesten gaan maken." Daarvoor moest eerst helder worden waar dat aan afgelezen kon worden.

Zharinova staat aan het hoofd van de speciale afdeling Business Analytics, van een groep van ongeveer tien mensen, inclusief de datawarehouse-engineers, die zich overigens in Shanghai bevinden. Het datawarehouse staat fysiek in de VS en is een *shared infrastructure* voor alle gebruikers binnen de eBay-organisatie. "Het mooie is dat het dezelfde tabellen zijn. Selecteer een site naar keuze en je kunt dezelfde vragen stellen – en je krijgt dus vergelijkbare antwoorden."

Analytics werkt nauw samen met de verschillende business units.

"Dat vind ik een belangrijke ontwikkeling," meent Zharinova.

"We werken in een snel voortrazende en veranderende omgeving. We moeten vaak plannen bijstellen. Aan de andere kant worden er voortdurend nieuwe business ideeën geïntroduceerd.

Het mes snijdt aan twee kanten: de business kan mij dingen vragen uit te zoeken, ik kan mijn eigen analyses voorleggen.

Ik maak een statistisch model met bepaalde variabelen dat een afwijking aantoont, de business zal vervolgens willen weten wat de afwijking heeft veroorzaakt."

Haar team meet alles wat er maar te meten valt op de sites: de activiteit van verkopers, van kopers, van bezoekers. Privacy staat daarbij hoog in het vaandel, eBay loopt al jaren voorop als het om privacy gaat. De user experience en het vertrouwen van de gebruikers in de websites mogen onder geen beding in gevaar komen.

Omdat de Classifieds Group onderdeel is van eBay, waartoe ook onder meer PayPal behoort, kan gebruik gemaakt worden van de volledige hardware en infrastructuur van de enterprise, die zorg draagt voor beheer, onderhoud en beschikbaarheid. Voor Analytics wordt gebruik gemaakt van een database in het Teradata Datawarehouse, dat geheel ontworpen is door het team van Zharinova. "Wij zijn geheel verantwoordelijk voor het datawarehouse en ons team doet alles van datamodellering tot ETL. De data worden dagelijks uit de productiesystemen gerepliceerd en in het datawarehouse geladen. Daarop draaien we onze analyses." Omdat de productiedatabases continu gerepliceerd worden is het mogelijk ook naar de live data te kijken, maar dat wordt zelden gedaan.

Bovenop het Teradata datawarehouse draaien MicroStrategy en SAS. MicroStrategy wordt bij eBay enterprisewide gebruikt voor rapportages en staat dus alle units ter beschikking. "SAS is een heel ander verhaal," vertelt Zharinova. "Dat kwam aan de orde toen we analyses gingen maken en ons op enig moment realiseerden dat een Excel-bestand van 90 MB niet de juiste oplossing was." Dat de keus op het SAS-platform viel, komt volgens haar door de functionaliteit voor het maken van *time series forecasting* en dat er geen limiet is aan de hoeveelheid opgeslagen data en dat je daar snelle analyses op kunt doen. Ook de integratie met MS Office speelde een rol: "In de wekelijkse management updates worden vooral PowerPoint slides gebruikt. Die hoeven eigenlijk alleen maar geactualiseerd te worden, iets wat met de add-in van SAS heel gemakkelijk is."

Het datawarehouse is groot. Heel groot. Alleen al het gedeelte

van Marktplaats.nl is al bijna 1 Terabyte, voor het gehele eBay concern bevat het datawarehouse minstens 1 Petabyte. Een recente hardware update zorgde voor een aanmerkelijk betere performance. "Mijn datawarehouse engineers in Shanghai maken zich vooral zorgen om CPU-gebruik en schrokken enorm toen ze sommige query's van onze collega's in Amsterdam zagen. Toen we ze dat uitgelegd hadden, hebben ze ons geholpen met de query efficiency die bijna 70 procent hoger kwam te liggen!" Voor Natasha Zharinova is het ontdekken van dingen die anders over het hoofd gezien zouden zijn, de reden om Analytics in te zetten. "Het bevestigen of juist ontkrachten van vermoedens is een belangrijke taak van Analytics. En soms ontdekken we merkwaardig gedrag waar we iets mee kunnen. Dat maakt het een heel interessant vak."

Ministerie van LNV

Bij de afdeling Analyse & Selectie van het Ministerie van Landbouw, Natuur en Voedselkwaliteit is fraudedetectie de belangrijkste drijfveer om Analytics te doen. Deze afdeling is onderdeel van de Dienst Regelingen, één van de uitvoeringsorganisaties van LNV en bevindt zich in Assen. Ooit werd de vestiging in Assen bestempeld tot algemeen Mestloket en sindsdien is het kantoor ook gegroeid tot het centrale dataverzamel-punt van de Dienst Regelingen.

Jantien Meijer is medewerker Analyse & Selectie binnen de Dienst Regelingen van het Ministerie. Analyse & Selectie werd opgericht bij de invoering van het nieuwe mestbeleid per 1 januari 2006 om de voorselectie te maken voor andere afdelingen, waaronder de Algemene Inspectie Dienst (AID), de inspecteurs die in het veld de fysieke controles doen.

"Analyse & Selectie vormt de brug tussen de administratieve kant en de AID aan de fysieke kant," zegt Meijer. "De data die wij hier centraal verzamelen komen uit heel verschillende bronnen. Allereerst van de boeren zelf die jaarlijks opgave moeten doen van land, opstallen, vee en personeel, en de claim die zij leggen op subsidies. Daarnaast heeft DR nog registraties van andere landbouwgegevens, zoals bijvoorbeeld runderen, geiten, schapen en mestvervoer en krijgen we data van verschillende andere organisaties, zoals bijvoorbeeld van die organisaties die zich bezighouden met de registratie van varkens en pluimvee." De data worden eerst opgeslagen in verschillende databases, om-dat vanuit de historie verschillende softwarepakketten worden gebruikt. De voor A&S meest relevante data, zo'n 150 Gigabyte, worden vervolgens in een speciaal datawarehouse geladen. Dat is ontwikkeld met Oracle Warehouse Builder. Daarnaast draait een apart datawarehouse met data omtrent dierziekten. Het is de bedoeling uiteindelijk te komen tot één datawarehouse binnen SAS. Als de analisten andere data in het datawarehouse zouden willen hebben, wordt daartoe een verzoek ingediend bij een andere afdeling van de Dienst Regelingen, die vervolgens samen met de Functioneel Applicatie Beheerder gaat kijken op welke wijze data kunnen worden verkregen. Meijer heeft zelf heel weinig met IT te maken.

"Het primaire doel van onze analyses is om overtreders te kunnen aanwijzen," legt ze uit. "Dat kan variëren van simpele overtreders, die hun formulier niet of te laat hebben ingediend, tot grootschalige fraudeurs. Daarnaast berekenen wij de normen uit de Meststoffenwet, waarvoor we onder andere de hoeveelheid dieren nodig hebben, de hoeveelheid grond, de hoeveelheden aangevoerde en afgevoerde mest. Bovendien zijn er allerlei uitzonderingsregels van toepassing."

Erg diep zit het Ministerie volgens haar nog niet in de Analytics, het werk is nog vrij basaal. "We krijgen af en toe wel vragen en daar hebben we onze statistische modellen voor paraat. Maar echte analyse zoals het integreren van vaste statistieken en uiteindelijk ook gaan voorspellen, forecasting, dat vind ik een uitdagende kwestie en ik hoop dat Analyse & Selectie dat in de toekomst gaat doen."

Steeds meer werk doet Analyse & Selectie voor het Gemeenschappelijk Landbouwbeleid, op grond waarvan subsidies worden verstrekt, wat voor deze afdeling in Assen een heel nieuw werkgebied is. Daarvoor moeten gewogen steekproeven worden genomen vanuit de verschillende datasets. Dat heeft niet direct met overtreders op dat moment te maken, maar er wordt wel gekeken naar mogelijke risico's, die vervolgens worden gemeld bij de AID. Bij de analyses ten behoeve van de Meststoffenwet worden direct (potentiële) overtreders gesignaleerd en voor verder onderzoek aangemeld.

"Datakwaliteit is echt een issue. Bij Analyse & Selectie zijn we ons daar al van het begin af van bewust," stelt Meijer. "Terwijl de mensen die de data inwinnen zullen zeggen dat het er voor 95 procent goed inzit en dat goed genoeg vinden, is helaas de kans groot dat nu net in die 5 procent de bedrijven zitten die als

overtreders aangewezen worden, terwijl dat misschien helemaal niet het geval is." Een bijkomend probleem is dat de dienst die de data inwint dat ook voor andere afdelingen doet; helaas worden binnen de verschillende regelingen vaak net iets andere definities gehanteerd. Even de datasets aanpassen is er dus niet bij. "We hameren al een tijd op datakwaliteit. Maar we zijn een kleine en relatief jonge afdeling, dus onze impact is niet groot. Inmiddels is de notie van het gebrek aan datakwaliteit in de hele organisatie aan het doordringen en zijn we een lang en vermoeiend traject gestart binnen de hele dataketen. Het duurde al een paar maanden voordat iedereen dezelfde definitie van datakwaliteit had geaccepteerd." Maar, zo weet Meijer, voor het uitvoeren van complexe Analytics als forecasting is de datakwaliteit van immens belang.

"We staan hier nog voor heel wat uitdagingen, maar Analyse & Selectie heeft beslist een gouden toekomst voor zich," meent Jantien Meijer.

Conclusie

Verskillende stadia van volwassenheid dus bij de toepassing van Analytics, maar een domein dat inmiddels de interesse van velen heeft gewekt. De ontwikkelaars van hardware en software zijn er klaar voor en bieden steeds geavanceerder modellen en tools. De komende jaren zal de belangstelling alleen maar groeien. En hoewel de rol van de IT-afdeling in het Analytics-domein bescheiden is, zullen de inrichting van speciale, soms zeer grote datawarehouses en de ontwikkelingen op het web IT'ers zeker voor nieuwe uitdagingen stellen.

Hans Lamboo is hoofdredacteur van Database Magazine.

Update

DB/M migreert online artikelarchief

Op de website van Database Magazine is de oude vertrouwde zoekfunctie weer beschikbaar, die u toegang geeft tot het grootste gedeelte van ons artikelarchief. Op onze website www.dbm.nl ziet u middenin de pagina het blok DB/M-artikelen, waar enkele van de meest recente te zien zijn. Klikte u vervolgens in de header op 'zoeken', dan komt u in een scherm waar u op auteur, trefwoord en editie kunt zoeken. Van artikelen die voor het jaar 2000 werden gepubliceerd is helaas niet altijd een PDF-bestand beschikbaar. Voor het gebruik van het online archief heeft u zich alleen maar te registreren – abonnees hebben altijd toegang en

bovendien tot alle artikelen. Gewerkt wordt nog aan de 'lees'-functie, die

het mogelijk maakt om een artikel in HTML-formaat op het scherm te krijgen.

The screenshot shows the Database Magazine website interface. At the top, there is a navigation bar with links for Home, Nieuwsbrief, Het Blad, Speciaal, Guides, Library, Vacatures, BI-platform, English, Contact, Mijn profiel, and Uitloggen. Below this, there are several sections: 'DBM blog - laatste entry' with a recent article from 30-09-2010; 'DBM activiteiten' listing events like Cloud Expo NL and BPMN 2.0; 'DBM nieuws | alle items' with a list of news items including a SAS award and Oracle BI updates. On the right side, there is a search box labeled 'DBM zoeken' and a sidebar with a 'DBM' logo and a 'DBM agenda' section. The main content area features a search bar and a list of articles under the heading 'Online Archief'.