

Effectief inzetten van resultaten van datamining is de grootste horde

Broodnodige intelligentie voor CRM

Peter van der Putten

Het succes van een onderneming wordt bepaald door de kwaliteit van de beslissingen die het neemt bij elk klantcontact. Binnen klantrelatiemanagement (CRM) wordt datamining een steeds belangrijker instrument om input te leveren voor deze beslissingen, in de vorm van voorspellingen van klantgedrag. Het daadwerkelijk op grote schaal operationeel inzetten, volgen en onderhouden van deze modellen is hierbij de grootste uitdaging - een probleem dat pas schoorvoetend erkend en geadresseerd wordt door de belangrijkste spelers in de datamining- en CRM-markt.

Operationele CRM-pakketten zoals Siebel, Clarify, Chordiant en Peoplesoft stroomlijnen de klantcontact-processen. Maar optimaliseren deze systemen ook echt de klantrelatie, zowel voor de klant als voor het bedrijf? Geeft deze software werkelijk inzicht in de potentiële waarde van klanten, een indicatie van mogelijke product-interesse in plaats van productbezit of schattingen van risico's als slecht betalingsgedrag? Het moge duidelijk zijn dat alle insiders, de innovatieve CRM-gebruikers en de CRM-leveranciers inbegrepen, ervan overtuigd zijn dat analytische CRM en voorspellende datamining in het bijzonder een noodzakelijke aanvulling zijn om de stap te maken van efficiënte klant-processen naar een optimale klantrelatie (zie ook kader: hoe werkt een voorspellingsmodel?). Maar zoveel overeenstemming als er bestaat over de noodzaak,

zoveel wederzijds onbegrip is er over hoe dit het meest doelmatig in praktijk te brengen valt.

KOUDWATERVREES BIJ KLANTEN

Reporting, warehousing, OLAP en datamining zijn allemaal business intelligence technologieën om gegevens voor strategische doeleinden te analyseren en exploiteren. Het is dan ook begrijpelijk dat bij veel bedrijven het misverstand ontstaat dat er eerst een volledig datawarehouse moet zijn voordat een bedrijf kan starten met 'geavanceerde analyses' als datamining. Toch is dit geen juiste opvatting. Het verdient de aanbeveling om voorafgaand en tijdens de ontwikkeling van een warehouse enkele ad hoc datamining projecten uit te voeren.

Een voorbeeld uit de praktijk is een telecombedrijf dat een datawarehouse wil opzetten om klantbehoud te verbeteren. In eerste instantie wordt er een ontwerp

Ik moet de eerste business-analist die vloeiend SQL spreekt nog tegenkomen

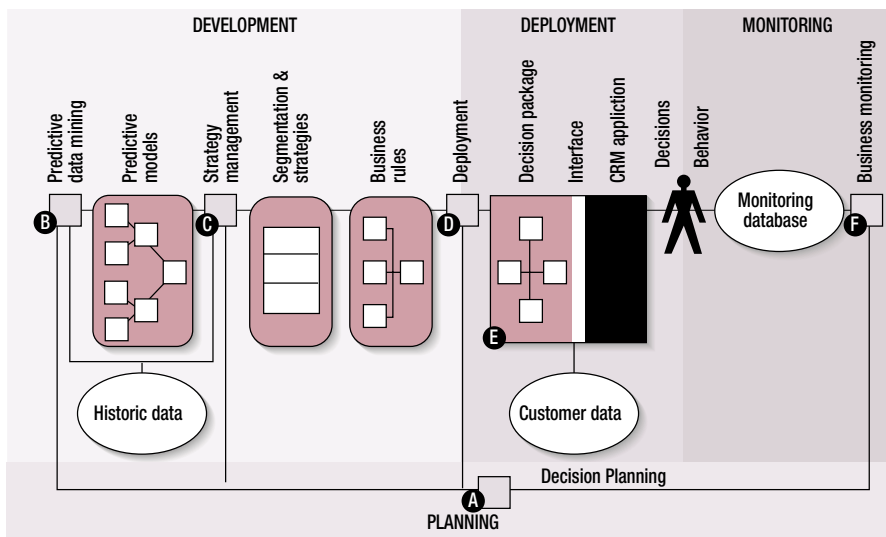
gemaakt waarbij enkele tientallen tot honderden klantvariabelen worden onttrokken uit operationele systemen. Na een datamining-analyse van een paar dagen op een eenmalig aangemaakte extractie blijkt dat de modellen die het meest accuraat weglooptgedrag van klanten voorspellen,

gebruik maken van minder dan vijftien variabelen. In de eerste fase hoeft er dus maar een zeer beperkte datamart ingericht te worden, tegen een fractie van de kosten van het complete ontwerp en tegen lagere risico's op verstoring van bronsystemen en operationele datastores.

Een vergelijkbare redenering gaat tot op zekere hoogte op voor operationele CRM systemen. Het is zeer wel mogelijk geld te verdienen met analytical CRM en datamining voordat het operationele CRM-systeem volledig uitgerold is, bijvoorbeeld door zich eerst te richten op 'ouderwetse' low-tech kanalen als direct mailings, ATM's, legacy branchsystemen enzovoort.

LOSSEN LEVERANCIERS DE JUISTE PROBLEMEN OP?

Een logische wedervraag is natuurlijk of de leveranciers zich wel bezighouden met het oplossen van de problemen die werkelijk belangrijk zijn voor de klant. Ook voor een zeer ervaren dataminer vormen de statistisch georiënteerde tool-sets als SPSS Clementine, SAS Enterprise Miner, IBM Intelligent Miner en het open source pakket WEKA op het eerste gezicht het Land van de Onbegrensde Mogelijkheden. Tientallen datamining-algoritmen, zo mogelijk nog meer geavanceerde statistische instellingen, rapporten en knoppen waar aan gedraaid kan worden. En er komen bij elke release meer en meer *low-level features* bij. Maar de vraag is natuurlijk of de CRM- of marketing-analist hier

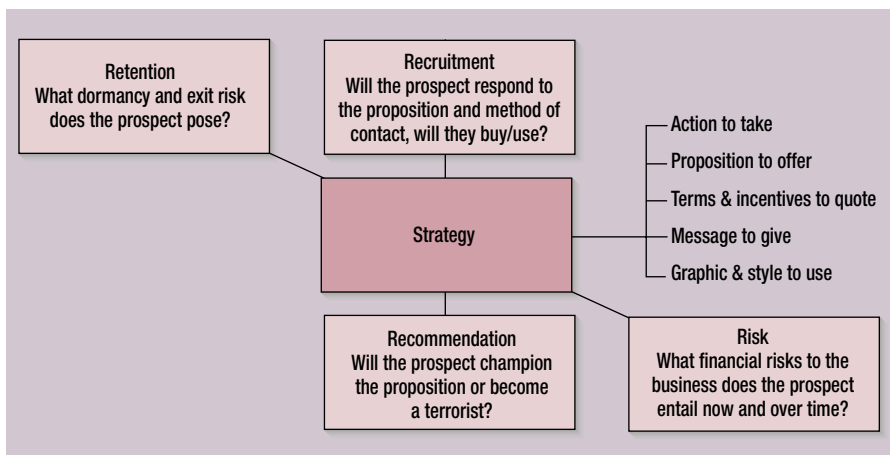


FIGUUR 1: VISIE OP HET INZETTEN VAN DATAMINING TER OPTIMALISATIE VAN CRM.

blij van wordt, de man of vrouw die maar beperkte tijd heeft om in zijn eentje voorspellingsmodellen voor respons, interesse, risico en klantbehoud up to date te houden, en dat voor misschien wel tien of meer producten.

Microsoft, Oracle en, veelzeggend genoeg, IBM zoeken het daarom in de richting van het onder de motorkap van de database onderbrengen van rudimentaire datamining-functionaliteit zoals voorspellen en segmenteren, en het beschikbaar stellen van deze functionaliteit door middel van eigen gebrouwen uitbreidingen op SQL of kleine applets. Voorbeelden hier van zijn Microsoft SQL Server 2000 Data Mining, Oracle 9i Data Mining, en de datamining-extenders die eind dit jaar uitkomen voor IBM's DB2. Het automatiseren, onder de motorkap brengen en tegelijkertijd ontsluiten van

datamining-resultaten is natuurlijk een lovenswaardig doel, maar het is de vraag of met het onderbrengen van mining-features in de database alle problemen de wereld uit zijn. Is het wel terecht en eerlijk om de DBA of database-developer verantwoordelijk te houden voor het ontwikkelen en onderhouden van veilige, betrouwbare datamining-modellen? Zal een database-omgeving, hoe uitgebreid ook, daar ooit voldoende ondersteuning voor kunnen bieden? Datamining is geen risicovrij proces, noch een magisch gereedschap dat simpelweg data verwerkt en goudklompjes oplevert. En de hamvraag: is het niet logisch dat, als datamining zoals vaak wordt gezegd inderdaad een business-activiteit is, dit daarom ook door business-analisten uitgevoerd dient te worden in plaats van de IT hier mee op te zadelen? Ik moet



FIGUUR 2: DIMENSIES WAAROP KLANTGEDRAG VOORSPELD KAN WORDEN.

de eerste business-analist die vloeiend SQL spreekt nog tegenkomen.

GEDEELDE VISIE OP DE INZET VAN DATAMINING

Kortom, zowel klanten als leveranciers richten zich teveel op het oplossen van de verkeerde problemen. Wat er met name aan schort is een door beide partijen gedeelde visie op de inzet van datamining voor CRM. Een visie met een sterke business-focus, die de nadruk legt op de lifecycles nadat een voorspellingsmodel is ontwikkeld (zie figuur 1). In de rest van dit artikel zal een mogelijke invulling hiervan besproken worden. Dit wordt geïllustreerd met een aantal voorbeelden, maar het is van belang te onderkennen dat dit een productonafhankelijke procesbeschrijving is. Databases en aan databases gerelateerde technologie spelen in dit proces een belangrijke rol. De kern van de visie is dat de business-gebruiker zelf strategieën voor het optimaliseren van klantcontact definieert en implementeert tot op elke klant 'touchpoint' en de prestaties van de strategieën volgt.

DIMENSIES VAN KLANTGEDRAG

Idealiter worden klanten beoordeeld op verschillende dimensies van klantgedrag: productinteresse en klantwaarde, trouw, risico's zoals slecht betaalgedrag en doorverwijswaarde (zie figuur 2). Per dimensie kan er weer verder ingezoomd worden. Als een klant dreigt op te stappen naar een concurrent bijvoorbeeld, kan ingeschat worden welke klantbehoud-aanpak de meeste kans van slagen zal hebben. Of per productfamilie kan bepaald worden welke producten het meest interessant zijn voor een klant. De gebruiker moet dus zeer snel betrouwbare en goed voorspellende modellen kunnen produceren. Er is dus behoefte aan een wizard die het gehele modelontwikkelproces ondersteunt en automatiseert, zonder verlies van controle door de gebruiker. Kortom, de dataminingtool is eerder een modellenfabriek dan een generieke gereedschapkast met allerlei toeters en bellen.

Ondanks dat databases op zichzelf tekort schieten voor het ontwikkelen van modellen door business-gebruikers, kunnen miningtools onderhuids gebruik maken van de datamining-functionaliteit die databases bieden. Als men als leverancier of klant niet met handen en voeten gebonden wilt zijn aan een database-leverancier dan zal men gebruik moeten maken van 'open' datamining-API's zoals XML for Analysis geïnitieerd door Microsoft, of de Java Data Mining (JDM) API als voorgesteld door IBM, Oracle en Sun. Helaas is dit nog toekomstmuziek.

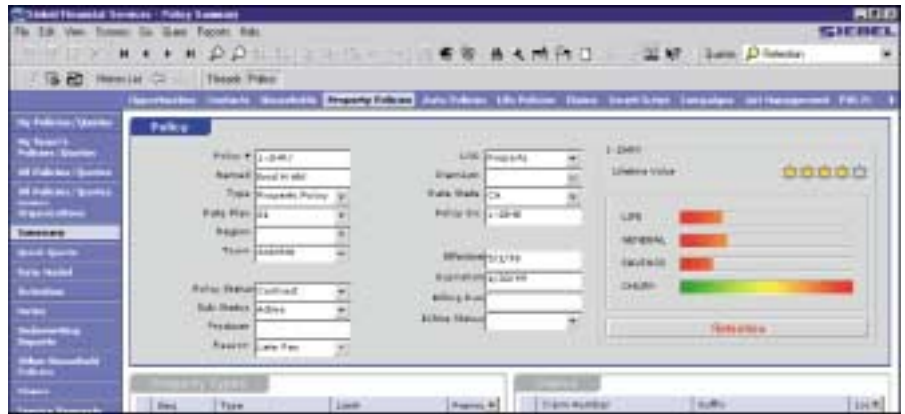
Maar de vraag is natuurlijk of de CRM- of marketing-analist hier blij van wordt

Deze API's zijn nog in een zeer pril stadium gemeten naar ondersteuning door leveranciers; de adoptie door klanten is nog vrijwel nul.

Zonder data valt er natuurlijk weinig te minen. Extraheer-, transformatie- en laad-gereedschappen (ETL) kunnen ingezet worden om snel datasets met historisch en te voorspellen klantgedrag aan te maken. Ideaal voor datamining is een enkele gedeneraliseerde tabel, met een record per klant dat alle klantkenmerken bevat. Dit wijkt dus sterk af van de sterschema's die voor OLAP doeleinden gebruikelijk zijn of klassieke relationele database schema's.

OOK REGELS ZIJN VAN BELANG

Om een optimale beslissing te kunnen maken moeten de verschillende klant-dimensies ten opzichte van elkaar afgewogen worden. Dit wordt bepaald door bedrijfsbeleid, strategie en regels. Bijvoorbeeld, tot op welke hoogte is het bedrijf bereid om risico te lopen door een lening aan te bieden aan een klant met een hoge voorspelde interesse in dat product? Zijn er bepaalde criteria om een lening af te mogen nemen, ongeacht voorspelde interesse? Als voor een klant evenveel interesse voorspeld wordt in product A en B, heeft het bedrijf dan een voorkeur



FIGUUR 3: INTELLIGENTE INTERACTIE DOOR HET AFWEGEN VAN VERSCHILLENDE VOORSPELLINGEN.

voor een van die twee producten, en welk besliscriterium wordt hierbij gehanteerd? Gaan we een klant die op het punt staat op te stappen hoe dan ook een product proberen te slijten, of is een andere actie geboden? Welke acties worden er gekoppeld aan de gevonden klantsegmenten?

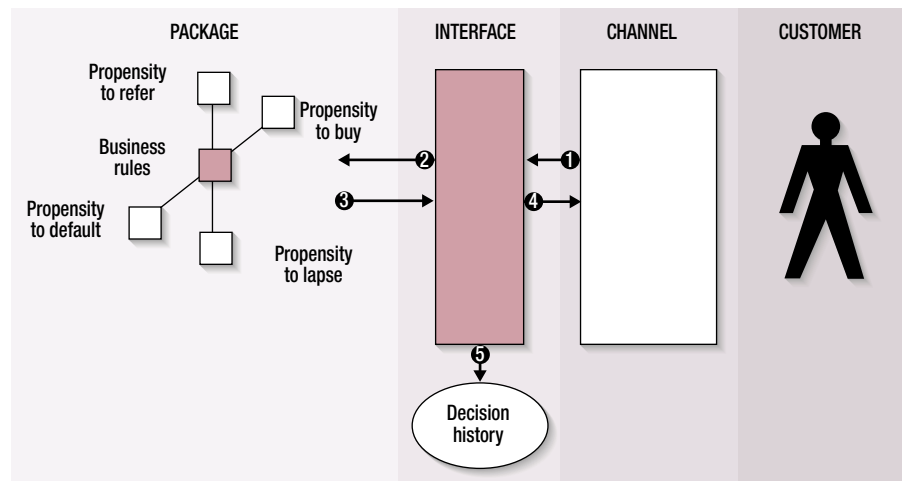
Een voorbeeld is te zien in figuur 3, waarbij een intelligente adviescomponent geïntegreerd is met een CRM-systeem (combinatie van Siebel Financial Services en OMEGA). Deze klant is geïnteresseerd in diverse productfamilies, maar heeft ook een hoog risico om de relatie op te zeggen. Omdat het om een hoogwaardige klant gaat wordt geadviseerd om te proberen deze klant te behouden - door op de 'Retention' knop te klikken wordt een klantbehoud-script geactiveerd.

De voorspellingsmodellen, regels en scripts om deze beslissing te kunnen maken moeten *soft coded* beschikbaar zijn, liefst in een enkel strategie-object en ze moeten te wijzigen zijn door gebruikers

met de juiste permissies. Met andere woorden, deze strategieën dienen opgeslagen te worden in een database, bij voorkeur in een open formaat zoals XML of een superset van PMML (predictive modeling markup language).

DOMME KANALEN

Een klant wil consistent benaderd worden, ongeacht het gebruikte kanaal. Bedrijven willen een strategie eenvoudig kunnen aanpassen en de wijzigingen direct terug zien over de kanalen heen. Niemand is gebaat bij een intelligente website in combinatie met een call center zonder enige vorm van personalisatie. Dit pleit ervoor geen intelligentie in de kanalen zelf te stoppen, maar de kanalen te laten interacteren met een centraal opgeslagen, herbruikbare klant-interactie-strategie. Dit wordt ook wel het 'thin channel' paradigma genoemd.



FIGUUR 4: IMPLEMENTATIE VAN KLANTINTERACTIE STRATEGIEËN.

Hoe werkt een voorspellingsmodel?

Voorspellingsmodellen kunnen worden gebruikt om op basis van klantkenmerken een voorspelling te maken van toekomstig klantgedrag. Deze modellen worden ontwikkeld op basis van historische gegevens. Hierbij geldt het adagium 'Garbage in, garbage out'. Des te beter en relevanter de data, des te beter de voorspellingen. Stel dat we interesse in wodka willen voorspellen. We beschikken over het volgende historische bestand:

Geslacht	Leeftijd	Grote_stad	Koopt_wodka
M	24	Ja	Ja
V	27	Ja	Ja
M	28	Nee	Nee
M	39	Ja	Nee
M	60	Nee	Ja
V	61	Nee	Nee

Speciale algoritmen kunnen nu gebruikt worden om het verband te ontdekken tussen de eerste drie velden en het Koopt_wodka veld. De formule die dit verband berekent is het voorspellingsmodel, en kan toegepast worden op klanten die nog geen wodka gekocht hebben (zie onder). Door klanten uit het historische bestand achter te houden kan de prestatie van het model getest worden op 'nieuwe klanten' die niet gebruikt zijn om het voorspellingsmodel te bouwen.

Geslacht	Leeftijd	Grote_stad	Kans_op_Koopt_wodka
M	25	Ja	0.7
V	35	Nee	0.1
M	55	Nee	0.6
			...

Een illustratie van dit concept is te zien in figuur 4. In een *inbound*, real-time scenario klopt een klant aan bij het bedrijf via een kanaal naar keuze (stap 1,2). De klant wordt beoordeeld op de verschillende klantdimensies en via business rules wordt de beste volgende actie gespecificeerd (stap 3) en doorgegeven aan het aanroepende systeem (stap 4). Vervolgens worden de kenmerken van de klant en de gegenereerde adviezen of beslissingen weggeschreven naar een monitoring datamart (stap 5). In een batch *outbound* scenario, bijvoorbeeld het selecteren van te mailen klanten uit een database geldt hetzelfde model, maar worden de klantcontacten geïnitieerd door het aanroepende systeem zelf.

Moderne EAI technologieën als SOAP, XML, J2EE of .Net applicatieservers en webservices spelen een belangrijke rol om goed te kunnen integreren met bestaande systemen en kanalen, zowel voor batch als voor realtime-toepassingen. Maar met

name de integratie met ouderwetse maar alomtegenwoordige kanalen, zoals legacy branchsystemen en het toevoegen van scores aan de centrale klantdatabase, verdient ook de aandacht. Contentmanagement-databases kunnen ingezet worden voor centrale opslag van presentatieonderdelen zoals scripts, afbeeldingen, schermonderdelen enzovoort.

MONITORING EN MODELBEHEER

Uiteindelijk is het van belang dat de performance van de CRM-activiteiten én van de voorspellingsmodellen gevolgd wordt. Ligt de respons en verkoop per kanaal nog op hetzelfde niveau, zijn de risico's binnen de klant- en productportfolio niet te veel toegenomen en voorspellen de modellen nog binnen betrouwbare marges? Indien de voorspellende kracht terugvalt kan bijvoorbeeld beoordeeld worden of de klan-

ten die zich aandienen substantieel andere eigenschappen of gedrag vertonen. Dit is dan een belangrijk signaal om over te gaan tot ontwikkeling van een nieuw model en klantstrategie. Het verdient wel de aanbeveling deze nieuwe strategie eerst een tijd parallel als 'uitdager' strategie te laten draaien, om te beoordelen of de resultaten in de praktijk werkelijk beter zijn. Deze methode staat bekend als 'champion-challenger' testen. ETL, OLAP en reporting spelen in de monitoring fase een belangrijke rol, om snel en eenvoudig de monitoring database te kunnen bevragen en eenvoudig de set aan rapporten uit te breiden.

SAMENVATTING

Over het belang van datamining zijn klanten en leveranciers het meer dan eens, maar de gezamenlijke visie over de invoering ervan kan nog verbeterd worden, en hopelijk draagt dit artikel hier aan bij. De doelgerichtheid van analytical CRM en datamining in het bijzonder bevordert dat datawarehouse- en operationele CRM-trajecten een sterke business-focus houden, met een incrementele scope en ROI gaandeweg de invoering. De invoering van datamining behelst meer dan het aanschaffen van een aantal miningmodulen binnen een relationele database. Het is in ieder geval aan te bevelen een degelijke procesarchitectuur op te zetten met veel nadruk op implementatie en exploitatie van modelresultaten en strategieën. Net als bij ieder ander databasetraject dient de verantwoordelijkheid voor de inhoud - de modellen en strategieën in dit geval - bij de business te liggen. Maar om het echt optimaliseren van de klantrelatie te laten slagen zal database technologie een doorslaggevende rol spelen. ●

Peter van der Putten (peter.van.der.putten@kiq.com) is principal consultant datamining & decisioning bij KIQ (<http://www.kiq.com>) en onderzoeker op dit gebied bij het Leiden Institute of Advanced Computer Science. KIQ is ontwikkelaar van de OMEGA Active Decision Management Suite en mede bedenker van de Fly By Wire visie op intelligente geoptimaliseerde CRM.