

# Datamining, de huidige stand van zaken

---

Gert Haanstra en Melbert van Emmerik, CustomerBase Jaarboek 2000, F&G Publishing, Amersfoort, Januari 2000, blz 64 t/m 75

*De mogelijkheden van datamining worden uitgebreider en interessanter. Enerzijds worden de tools beter en krijgen de mensen die ermee werken meer ervaring. Anderzijds wordt er meer data in databases opgeslagen door de enorme groei in de hedendaagse communicatie mogelijkheden, wat meer en meer informatie over klanten, klantcontacten en transacties genereert. Het gaat zelfs zover dat alle boodschappen in de supermarkt door middel van klantenkaarten geregistreerd kunnen worden. Hetzelfde geldt voor iedere klik die klanten op jouw internet site maken. De opslag van al deze informatie vormt niet meer het grootste probleem, De opslagsystemen worden steeds goedkoper en meer geavanceerd. Maar wat moet je met al die informatie? Door middel van datamining gaan steeds meer bedrijven zoeken naar mogelijkheden om de verkregen informatie te gebruiken in hun marketing beleid. Doel van dit stuk is het geven van een overzicht van de (on)mogelijkheden van datamining, afgezet tegen de ontwikkeling die datamining doormaakt.*

## Datamining in een breder verband

Datamining is de zoektocht naar belangrijke verbanden en patronen binnen een analysebestand. Het is een essentieel onderdeel van een veel groter proces: het Knowledge Discovery in Databases (KDD) proces. Dit laatste proces omvat het gehele proces van commerciële vraagstelling naar de uiteindelijke marketing activiteit die op specifieke groepen klanten wordt ondernomen.

Het KDD-proces begint met het selecteren van die gegevens uit de grote hoeveelheid data die relevant zijn voor het beantwoorden van de vraagstelling. Na de selectie van relevante gegevens vindt een data manipulatie slag plaats. Hierbij worden gegevens geschoond, consistent gemaakt en gecombineerd tot nieuwe informatie. Dit is met name in het beginstadium een bijzonder tijdrovende aangelegenheid. Welke gegevens zijn überhaupt binnen een organisatie aanwezig? Waar zijn alle gegevens opgeslagen? Vaak zijn er verschillende systemen of databases waardoor informatie op verschillende manieren is opgeslagen. Het koppelen van gegevens uit deze verschillende systemen kan dan problemen opleveren. Daarom is het streven dit manipulatieproces zo veel mogelijk te automatiseren. Waarbij veelal de gemanipuleerde data in datawarehouses wordt opgeslagen. Het automatisch opslaan van informatie in datawarehouses is dus bijzonder belangrijk. Daarmee krijgt de informatie over de tijd een consistent karakter (en wordt de kans op fouten kleiner) en is de informatie voor de dataminingfase direct toegankelijk. Vervolgens begint de zoektocht naar belangrijke verbanden en patronen binnen het analyse bestand. Deze dataminingfase resulteert uiteindelijk in de kennis omtrent de initiële vraagstelling.

De term datamining wordt vaak verward met KDD (Knowledge Discovery in Databases). Strikt genomen is het zoeken van relevante informatie uit databases specifiek gestuurd door de kennis en ervaring in het bedrijf geen onderdeel van datamining. Het creatieve proces van het selecteren en manipuleren van data om nieuwe afgeleide informatie te verkrijgen zou je databasemining kunnen noemen. Het resultaat is een analysebestand dat gebruikt wordt om verbanden en patronen binnen het analyse bestand te vinden. Het analyse pakket automatisch laten zoeken naar deze verbanden en patronen heet datamining. Zowel databasemining en datamining zijn onderdelen van het Knowledge Discovery in Databases proces. In dit artikel ligt de nadruk op de rol van datamining, hoewel het onoverkomelijk is om ook andere facetten van het KDD-proces de revue te laten passeren.

## Ontwikkelingen in datamining

### *Pioniersfase*

Datamining begint met het uitproberen van de mogelijkheden. Zo'n eerste stap heeft als functie om het bedrijf warm te maken voor datamining. Wat kan datamining voor het bedrijf betekenen? Levert het geld op? Zo ja, hoeveel? Welke investeringen gaan ermee gepaard? En op welke gebieden vereist het organisatorische aanpassingen? Bedrijven begrijpen dat klantinformatie geld oplevert, en proberen met behulp van datamining de benodigde informatie uit de databases te krijgen. Enerzijds om de effectiviteit van de marketing inspanningen te vergroten. Anderzijds om nieuwe verkoopmogelijkheden te ontdekken in het klantenbestand. Een eerste grote succes prikkelt het bedrijf om aandacht te geven aan de mogelijkheden.

Er wordt het een en ander uitprobeerde met de gegevens uit bestaande databases: een grote analyse, het bouwen van een scoremodel, het vinden van nieuwe klantsegmenten. Bij een gedegen teststrategie en voldoende aandacht voor de creatie van afgeleide voorspellende klantkenmerken kan het succes van datamining niet uitblijven. In dit stadium staan alle stadia in het KDD-proces centraal. In de ideale situatie werken verschillende disciplines samen om hun eerste resultaten te behalen. IT-specialisten om ruwe data uit de diverse transactie bestanden beschikbaar te stellen. Marketeers en/of domein deskundigen om input te geven aan de bruikbaarheid van de klantinformatie. En analisten voor het bewerken van de data en het ontwikkelen van een voorspellingsmodel. Duidelijk is dat datamining in deze eerste fase vanaf de zijlijn meedoet aan het verkoopproces. Bij aantoonbaar succes gaat de deur open voor een structurele rol in het marketing & verkoop proces. Vaak wordt datamining voor verschillende toepassingen in een soort proeftuin setting uitprobeerde om een breder draagvlak te creëren om de vaak hoge investeringskosten het hoofd te kunnen bieden. De eerste fasen van het KDD-proces worden vervolgens voor een groot deel door IT-specialisten geautomatiseerd door de bouw van datawarehouses. Het datawarehouse bevat alle relevante afgeleide data die gebruikt worden in datamining. Wanneer het datawarehouse volledig geïmplementeerd is zal de nadruk steeds meer op datamining komen te liggen.

### *Productiefase*

Komt er een vervolg na de 'pioniersfase', dan staat in de tweede fase de integratie van datamining in het marketing- & verkoopproces centraal. Enerzijds om datamining consequent te gebruiken en maximaal te profiteren van de opbrengsten. Anderzijds om ontwikkelde voorspellingsmodellen continu te evalueren en verder aan te scherpen. De integratie moet uiteindelijk leiden tot structurele verbeteringen in de actieresultaten.

Het streven is continuïteit. Continuïteit in het verbeteren van de kwaliteit van de modellen. Continuïteit in het creëren van meer modellen voor nieuwe doelgroepen en voor andere producten. En continuïteit in het binnenhalen van meer omzet. Ofwel, continuïteit in het optimaliseren van het commerciële proces.

In de praktijk betekent dit het opzetten van testcycli om voortdurend de juiste mix van klanten, producten, kanalen en timing te achterhalen. Verschillende acties worden specifiek opgezet om nieuwe modellen te bouwen. Daarnaast worden bestaande modellen aangescherpt door ze te gebruiken en tegen oude modellen te testen. Uiteindelijk leiden deze cycli tot een set kwalitatief goede voorspellingsmodellen voor de meeste 'standaard' acties.

Voorheen gaven analisten voornamelijk vrijblijvend advies aan marketeers over de te benaderen klantgroepen. Met een integratie van datamining, ontstaat een duidelijke verantwoordelijkheid in het continu vinden van nieuwe, beter scorende doelgroepen in de database. Een discipline die in eerste instantie ligt bij analisten, maar die een sterke samenwerking met data specialisten en marketeers vereist voor maximaal succes. Op basis van het profiel van gevonden klantgroepen wordt tezamen met de marketeers een geschikte campagne bedacht.

Het zoeken van een goede combinatie tussen enerzijds campagne management systemen en anderzijds geavanceerde datamining tools is derhalve een pre. Centraal hierin staat een datawarehouse waarbij alle relevante informatie over klanten is verzameld. Het datawarehouse biedt zowel de informatie die nodig is om datamining toe te passen als de informatie die nodig is om klanten voor verschillende campagnes te selecteren. Afhankelijk van kanaalvoorkeuren, recente communicatie uitingen, specifieke klantinformatie en de scores op verschillende scoremodellen worden klanten voor campagnes geselecteerd.

Vereiste is dat de informatie in het datawarehouse zeer up-to-date en betrouwbaar is. Bij voorkeur worden de scores voor de verschillende voorspellingsmodellen berekend op het moment van selectie. Dit houdt in dat het campagne management niet voorberekende scores voor verschillende campagnes gebruikt, maar dat de scores dynamisch beschikbaar zijn. Het campagne management systeem roept het model aan, berekend de scores en wijst vervolgens klanten toe aan verschillende campagnes.

Het uiteindelijk resultaat is goed lopend contact management proces. Een voortdurende bewaking van de kwaliteit van selecties. Een effectieve verdeling van klanten onder acties van marketeers. Kortom, een instrument om op efficiënte wijze bestaande klanten te bewerken.

In deze tweede fase zorgt het gebruik van datamining voor een sterke leercurve in de organisatie. Het meest duidelijk is de stijging in de effectiviteit van acties. Met als

voorwaarde: een procedurele aanpassing van het selectie/analyse proces, dat het gebruik van score modellen waarborgt. De invalshoek in deze fase is nog sterk gericht op outbound marketing activiteiten. Door de opkomst van meer interactieve media en ook de groeiende behoeften van klanten om zelf het initiatief te nemen om contact te zoeken met bedrijven via deze kanalen, moeten toepassingen van datamining verder uitgebreid worden om ook de meer interactieve contact momenten te benutten.

### *Interactiefase*

Waar in de vorige fase vooral de marketeers en analisten de touwtjes in handen hadden, neemt de klant steeds vaker het roer over. Met de komst van nieuwe en de verdere ontwikkeling van bestaande interactieve media, zoals call-centers, voice-response-units en internet, ligt het contact initiatief meer en meer bij de klant. De tijd om klanten goed voorbereid te benaderen met een specifiek aanbod is er opeens niet meer. De klant of een prospect neemt met jou contact op. Het ultieme moment om te marketen is daar. Wat moet je doen?

Of het nu om een call-center gaat of om internet, het concept blijft hetzelfde. De bestaande kennis over de betreffende klant moet onmiddellijk leiden tot een commerciële suggestie op maat. Op zich geen probleem. Een selectie van de beste benadering op basis van een batterij scores lijkt een oplossing. Maar wat als de klant zijn beklag komt doen of een product bestelt, waardoor de voorgestelde benadering zinloos blijkt? De klant neemt het initiatief tot het contact en zal daarom eerst iets van jou verlangen. Deze informatie is essentieel in het bepalen van een eventuele commerciële benadering richting die klant. Voorberekende scores op basis van datamining modellen zijn niet toereikend. Het online scoren van datamining modellen met zowel bestaande als zojuist verkregen klantinformatie is derhalve een must.

Datamining in het productieproces stelt hoge eisen aan de voorspellingsmodellen. Deels om rekening te houden met de dynamiek, waarbij laatste wijzigingen leiden tot nieuwe uitkomsten. En deels vanwege de directe feedback op de gekozen benadering. Het verschil tussen actie en het meten van de reactie reduceert zich van enkele weken, of zelfs maanden, tot een enkele seconden. Dit vereist een snelle doorberekening van modellen en een zeer valide en robuuste voorspelling. In de praktijk betekent dit dat, bijvoorbeeld, een agent aan het call-center het advies krijgt om de klant een aanbieding te doen. Het model geeft een kans van slagen van 80% aan. De informatie over de klant wijzigt zich tijdens het gesprek en de scoringskansen worden tijdens het gesprek binnen enkele seconden opnieuw vastgesteld. Als de scoringskans van het betreffende aanbod zakt naar bvb. 20%, zal een ander aanbod met een hogere kans van slagen op het scherm van de agent verschijnen. Uiteindelijk bepaalt de agent altijd zelf of de geboden commerciële suggestie daadwerkelijk uitvoerbaar is.

Belangrijk voordeel van het inzetten van modellen in het interactieve productieproces is dat de communicatie met de klant verschuift van monoloog naar dialoog. Beschikt men in het klantcontact over de juiste informatie dan ontstaat er een effectieve dialoog met de klant. Hierbij krijgt de klant inzicht in het bedrijfsproces, en merkt het dat het hier invloed op heeft. Directe gevolg is een vergroting van de klanttevredenheid, en daarmee samenhangend, een

duurzame relatie met de klant. Dus het inzetten van datamining in het productieproces stelt een bedrijf in staat nieuwe klanten te werven, langdurige, succesvolle relaties met klanten op te bouwen en deze te onderhouden.

## Een stap richting Customer Relationship Management

Duidelijk is dat de huidige datamining technologie klaar is om ingezet te worden in het versterken en het verdiepen van de individuele relatie met de klant. Ofwel, datamining kan worden gebruikt in het Customer Relationship Management (CRM) concept. Centraal in dit concept is de interactie met de klant tijdens een contactmoment. De klantwaarde kan worden verhoogd en de loyaliteit worden vergroot door in te gaan op de eigen wensen en behoeften van iedere klant. Juist in klantcontacten ondersteunt datamining het beslissingsproces door gebruik te maken van de kennis over de klant.

Op basis van specifieke klantkenmerken vindt een differentiatie van klanten plaats op basis van gedrag, kanaalvoorkeur en specifieke wensen en behoeften. Met name in dit traject speelt datamining een kritische rol. Op het moment van interactie speelt differentiatie een grote rol. Op welke wijze moeten bepaalde klanten benaderd worden, kan ik een aanbod doen en zo ja welk aanbod en wat zijn goede argumenten om bij die specifieke klant te gebruiken? Datamining kan worden ingezet om de effectiviteit van de differentiatie en interactie te evalueren om het vervolgens te optimaliseren. In dit hoofdstuk gaan we in op de eisen voor een succesvolle implementatie van datamining in CRM.

Voor de implementatie van datamining binnen customer relationship management is het belangrijk om een onderscheid te maken tussen inbound en outbound contacten. Bij outbound contacten (mail, leads, calls) vindt het toewijzen van klanten aan verschillende campagnes plaats in batch processen. Een campagne management systeem of een scheduler met verschillende query jobs heeft bijvoorbeeld de gehele nacht de tijd om klanten te selecteren op basis van specifieke klantkenmerken. Bij inbound contacts is die tijd er niet meer. Wil je klanten gedifferentieerd aanspreken, dan is het noodzakelijk om alle relevante klantinformatie up-to-moment paraat te hebben. Sterker nog, de nieuw informatie die gedurende de interactie tot stand komt, kan zeer essentieel zijn in het vervolg van de interactie. Door middel van online scoring van voorspellingsmodellen wordt nog tijdens het gesprek de vervolgstap bepaald. Dit betekent dat binnen enkele seconden data verwerkt moet zijn en dat scores op verschillende modellen doorgerekend moet worden.

De meeste bedrijven hebben operationele systemen en datawarehouses op gescheiden platformen geïnstalleerd. De reden hiervoor is dat een datawarehouse primair gebruikt wordt voor het klaarzetten van ruwe (geschoonde en consistente), afgeleide en gesommeerde data voor analyses en rapportages. Door deze informatie op een afzonderlijk platform beschikbaar te stellen, is er geen gevaar voor overbelasting van operationele systemen. Onder andere doordat up-dates van datawarehouses periek in batch processing liep. Tegenwoordig is er een grote behoefte aan up-to-moment klantinformatie. Door het meer interactieve karakter van online klantcontacten bij call-centers of op internet, moeten bedrijven zowel beschikken over up-to-date informatie als over zoveel mogelijk afgeleide klantinformatie als over de meest geschikte actie voor dat moment. Dit heeft belangrijke



consequenties voor het onderhoud van datawarehouses, aangezien datawarehouses eveneens up-to-date moeten zijn.

Welke eisen het gebruik van datamining in CRM omgeving stelt aan de infrastructuur, laten we zien in een voorbeeld. Een voorspeller in een specifiek model kan 'het aantal stortingen over de laatste maand' zijn. Dit kenmerk is verkregen door het aantal stortingen in het gegeven tijdinterval te tellen in de transactiesystemen. De storting moet eerst in het transactiesysteem worden verwerkt. Vervolgens moet het datawarehouse gemuteerd worden op basis van de nieuwe informatie. Daarna worden opnieuw de scoremodellen doorlopen, voordat we weten of de mutatie van invloed is op het aanbod dat we online willen gaan doen. Het zal duidelijk zijn dat dit geen haalbare kaart is tijdens hetzelfde telefoongesprek van een call center agent met een klant. Een mogelijkheid is om genoeg te nemen met informatie die een dag oud is. Wanneer een call-center agent een melding krijgt dat een bepaald product geschikt is voor die klant, dan zal die agent afhankelijk van het gesprek bepalen of de voorgestelde suggestie nog hout snijdt. Een andere optie is om voor interactieve toepassingen alleen gebruik te maken van scoremodellen die minder dynamische of kritische klantkenmerken bevatten. Een derde optie is om gegevens via triggers in de database af te tappen waardoor een snellere herberekening van scoremodellen mogelijk is.

## De noodzakelijke functionaliteiten

Om CRM volledig en goed te kunnen uitvoeren, zijn een aantal tools nodig. In een z.g. contact management systeem (CMS) is campagne management naadloos geïntegreerd met datamining en online scoring.

Ten behoeve van de ontwikkeling van score modellen kan een datamining tool niet meer bestaan zonder KDD functionaliteiten als het toegankelijk maken van databases en verregaande datamanipulatie mogelijkheden. Belangrijk zijn de volgende onderdelen:

- 'data access' voor het flexibel en dynamisch verkrijgen van gegevens uit het datawarehouse,
- 'data preparatie' voor het bewerken, aggregeren en mergen van gegevens uit verschillende tabellen,
- 'behavioural scoring' voor de ontwikkeling van score modellen,
- 'export' van model definities in het datawarehouse voor gebruik door het contact management systeem.

Vervolgens zijn de volgende campagne management functionaliteiten gewenst:

- 'event triggering' om belangrijke events te herkennen en daar met campagnes op in te springen,
- 'targeting systeem' om te beslissen welke campagnes het beste passen bij welke klanten,
- 'scheduler' om de selecties van verschillende campagnes in de tijd te automatiseren,
- 'channel management' voor distributie en coördinatie van verschillende , multi-channel campagnes,

- 'decision support' om prioriteiten te stellen aan het gebruik van score modellen en het gebruik van de verschillende distributie kanalen,
- 'reporting' om enerzijds vooraf de (financiële) consequenties van campagnes te bepalen en om anderzijds de campagnes te monitoren.

In een multi-channel situatie spelen vele aspecten een rol, waaronder:

- Directheid (b.v. mailings, huis-aan-huis inserts, coupons),
- Differentiatie (b.v. campagne A voor klant X en campagne B voor klant Z),
- Persoonlijk (b.v. franchise, tussenpersonen, call-centers, gepersonaliseerde internetpagina's),
- Massaal (b.v. radio/tv-advertenties die specifieke doelgroepen aanspreken, nieuwsbladen),
- Electronisch (b.v. telefoon, voice-response, sms, internet, email),
- Interactief (b.v. internet, call-center).

## Geschikte analyse technieken

De gebruikelijke analyse technieken worden bekeken aan de hand van de nieuwe eisen die vanuit CRM aan datamining technieken worden gesteld. Met datamining tools moet het o.a. mogelijk zijn om toekomstige trends en toekomstig klantgedrag te voorspellen. Waarbij het mogelijk moet zijn om pro-actieve en kennisgedreven beslissingen te ondersteunen. Een belangrijk en vaak veronachtzaamd criterium voor de keuze van de tools of technieken is de interpreteerbaarheid van de segmenten en het gemaakte model. Naast de kwaliteit van de voorspelling (b.v. het aantal keren dat de voorspelling ook daadwerkelijk uitkomt), moet duidelijk zijn wie voor welk aanbod geschikt is. En vooral de reden waarom iemand het aanbod krijgt. De meeste leveranciers van datamining tools geven aan de interpreteerbaarheid van hun oplossingen/technieken hoog in het vaandel te hebben.

De combinatie van verschillende meer traditionele statistische methoden en nieuw verkregen inzichten hebben tot een grote diversiteit aan datamining technieken en algoritmen geleid. Veelal staan datamining leveranciers in nauw contact met verschillende universitaire onderzoekscentra om snel nieuwe algoritmen in hun datamining oplossingen te integreren. In het verleden leidde dit op een gegeven moment tot steeds minder inzichtelijke methoden, omdat voortdurend op zoek ging naar de best voorspellende datamining technieken. Tegenwoordig is er het besef dat voor marketing toepassingen de interpreteerbaarheid van datamining uitkomsten zeer relevant is. Je wilt niet alleen weten welke campagnes het beste aansluiten bij welke klanten, maar je bent ook geïnteresseerd in de onderliggende redenen. Met name bij de meer marketing gerichte leveranciers is inmiddels de trend om een hoge prioriteit toe te kennen aan de interpreteerbaarheid bij een aanbieden van datamining toepassingen. Enkele technieken die door veel datamining tools ondersteund worden, zullen op basis van deze criteria worden beschreven.

### *Regressie analyse*

Meest traditioneel zijn de regressie technieken. Het meest inzichtelijke is de multi-pele regressie waarbij een lineaire functie een schatting geeft van hetgeen je wilt voorspellen,

bijvoorbeeld respons op een mailing. Daarnaast wordt ook wel logistische regressie gebruikt wanneer de te voorspellen variabele, zoals respons, slechts twee uitkomsten omvat (ja/nee). Logistische regressie is wat moeilijker interpreteerbaar, maar stelt minder hoge eisen aan de kwaliteit van de data.

### *Decision trees, decision lists of decision rules*

Waarschijnlijk de techniek met de meest inzichtelijke resultaten. Bij beslissingstechnieken wordt niet zozeer een voorspelling gedaan, maar wordt de dataset in segmenten opgesplitst. Deze segmenten worden expliciet beschreven aan de hand van de kenmerken van de klant(groep). Elk segment heeft een bepaalde gemiddelde score, bijvoorbeeld gemiddeld respons. De vergelijking van de score tussen de segmenten resulteert in uitspraken over de meest kansrijke segmenten voor een bepaalde actie. Bij decision trees wordt het gehele analysebestand in segmenten opgesplitst. Bij decision rules worden de beste segmenten uit de dataset geïdentificeerd. Tussen de verschillende segmenten kan overlap ontstaan. Overlap is uitgesloten bij decision lists aangezien hier stapsgewijs naar het meest kansrijke segment wordt gezocht. Eenmaal een segment gevonden dan worden de klanten in dat segment uitgesloten bij de zoektocht naar het volgende meest kansrijke segment. Beslissingstechnieken spreken marketeers bijzonder aan, aangezien het resultaat een serie klantprofielen is met een bijbehorende score. Bijvoorbeeld, mannen onder 25 jaar met een lease auto hebben een schaderisico van, laten we zeggen, 45% op jaarbasis.

### *K-nearest neighbour*

De K-nearest neighbour gaat uit van de gedachte dat klanten die op elkaar lijken hetzelfde gedrag vertonen. Wanneer je wilt voorspellen hoe iemand zich zal gaan gedragen, kijken we naar het K-aantal (b.v. 10) klanten dat het dichtst bij hem in de dataspace zitten, met andere woorden het meest dezelfde kenmerken bezitten. Naarmate meer van zijn burens bepaald gedrag vertonen, zal het waarschijnlijker zijn dat de klant in kwestie ook dat gedrag gaat vertonen.

### *Neurale netwerken*

De verschillende varianten van neurale netwerken zijn geïnspireerd op het menselijk brein dat een onnoemelijk hoeveelheid neuronen bevat die allemaal tot elkaar in verbinding staan via synapsen. Het basisprincipe komt erop neer dat een neurale netwerk eveneens neuronen (nodes) die alle mogelijke klantkenmerken weergeven (input nodes). Via een aantal tussenliggende lagen staan deze input nodes in verbinding met de output nodes (hetgeen je wilt voorspellen). De tussenliggende lagen bestaan uit een groot aantal lineaire functies waarbij input nodes via variërende gewichten de output nodes voorspellen. Gedurende het trainen worden de gewichten en de functies steeds aangepast en wordt het effect van het resultaat gemeten in termen van verschillen tussen voorspelde en werkelijk uitkomst. Het grote probleem bij neurale netwerken is dat het een ondoorzichtige black-box is, dat wel oplossingen verstrekt, maar niet de totstandkoming van de oplossing. Bovendien is het minder geschikt voor een dynamische omgeving omdat het netwerk hiervoor opnieuw getraind moet worden.



## *Genetische algoritmen*

Genetische algoritmen zijn geïnspireerd door de evolutie theorie, waarbij oplossingen in verschillende stadia geoptimaliseerd worden. In eerste instantie worden een aantal ruwe oplossingen geboden. Het algoritme gaat verder met de beste oplossingen en maakt vervolgens combinaties van de verschillende oplossingen. Delen van de ene oplossing worden gebruikt in andere oplossingen en andersom. Dit proces herhaalt zich waarbij verschillende generaties oplossingen ontstaan totdat er uiteindelijk een bevredigde oplossing uitrolt. Het proces van voortborduren op eerdere oplossingen maakt het herleiden van het eindresultaat naar de oorspronkelijke klantkenmerken moeilijk. Hetgeen de interpreteerbaarheid van de uiteindelijke oplossing naar klantkenmerken uiteindelijk niet ten goede komt.

## *Association rule*

Association rules zijn altijd gebaseerd op binaire kenmerken (ja/nee). Er wordt gezocht naar IF-THEN regels. Een uitkomst is bijvoorbeeld; als geslacht is 'man' en aankoop is 'luiers', dan 50% koopt eveneens bier. Zo kun je oneindig veel associaties maken, maar uiteraard ben je geïnteresseerd in de regels die een voldoende groot deel van het klantbestand identificeren en die een hoge uitkomst opleveren. Resultaten zijn makkelijk en eenduidig te interpreteren.

## **Conclusie**

Datamining maakt een snelle ontwikkeling door. Echter, slechts weinig bedrijven weten deze mogelijkheden volledig te benutten. Bij een aantal bedrijven zijn marketing datawarehouses inmiddels operationeel, al dan niet na een lange lijdensweg. Nu is het stadium bereikt waarin de aandacht uitgaat naar de inrichting van het campagne management proces inclusief het effectieve gebruik van scoremodellen. Ondertussen gaat de ontwikkeling van datamining steeds verder. Een volledige integratie in het customer relationship management concept is inmiddels mogelijk. Meer directe kanalen, meer interactiviteit en meer initiatief van de klant. Denken bedrijven dat ze er zijn met een goed lopend campagne management systeem en een volwaardig datawarehouse, dan komen ze bedrogen uit. Want hoe kun je interactieve contactmomenten commercieel uitbuiten wanneer klanten zelf contact zoeken? Wanneer klanten een call-center bellen, zijn zij bereid tot interactie. Wanneer klanten de internetsite bezoeken staan ze open voor informatie. Dit zijn bij uitstek de momenten om de relatie met de klant te verstevigen.

Veel bedrijven zijn nu in staat om gedifferentieerde outbound marketing activiteiten te ontplooiën. De toekomst ligt in de inbound contactmomenten. Inbound contactmomenten komen bij uitstek tot stand via call-centers, via email en via internet. Call-center agents moeten derhalve uitgerust worden met informatie om deze moment optimaal te benutten. De uitdaging hierbij ligt in de snelheid waarmee de informatie uit het gesprek wordt verwerkt in de score van de klant (en de aanbieding die deze klant krijgt). Bij internet ligt de uitdaging meer in de grote hoeveelheid gedragsinformatie afkomstig van het internet bezoek. Wat doen individuele klanten op de site en in hoeverre zijn wij in staat dat te beïnvloeden en commercieel uit te buiten?

De uitdaging voor de toekomst ligt initieel op een betere differentiatie van klanten op basis van de huidige kennis van klanten en op de ontwikkeling van meer gedifferentieerde klantconcepten. Daarnaast komt weer nieuwe informatie vrij over interactieve contactmomenten met klanten. Deze informatie overstijgt, wat volume betreft, waarschijnlijk zeer snel de huidige informatie over onze klanten. Het intelligent omgaan met deze nieuwe vloedgolf aan klantinformatie in het contact met klanten zal een hot issue worden. Welke informatie is relevant? Welke informatie is vervolgens af te leiden? En hoe kan die informatie worden gebruikt om de interactie met klanten te faciliteren om zodoende de customer relationship optimaal te managen.

