

## Datamining legt ook online de ziel van de klant bloot

# Schatgraven op internet

In het online kanaal vindt al jarenlang enorm veel interactie met de klant plaats. Toch worden de gegevens die online het gedrag van klanten beschrijven, nog steeds nauwelijks gebruikt ten behoeve van analyses en geïntegreerde marketing. De traditionele statistische technieken om klantgedrag en klantbehoeften te voorspellen zijn echter ook geschikt om toe te passen op online klantgegevens.

Het gedrag van klanten wordt offline al jaren met grote nauwkeurigheid voorspeld. Vaak wordt per klant in realtime de beste aanbieding doorgerekend. De voorspellingen zijn gebaseerd op grote hoeveelheden gedragsgegevens, contactgegevens, productgegevens en persoonsgegevens van klanten. Deze schijnbaar onoverzichtelijke bron wordt behapbaar gemaakt door toepassing van dataminingtechnieken die patronen en profielen kunnen opsporen in de gegevens.

### Verborgen

Datamining is het toepassen van statistische algoritmen om (verborgen) structuren en patronen in grote databestanden te vinden. Voorbeelden in de marketing zijn het vinden van klanten die met een grote mate van waarschijnlijkheid zullen reageren op een direct-mailactie, of het segmenteren van een klantenbestand aan de hand van de waarschijnlijkheid dat klanten zullen opzeggen, om zo input te krijgen voor behoudsactie.

Het dataminingproces is grofweg in drie fases in te delen: het definiëren van een dataminingvraag, het kiezen van de juiste dataminingtechniek, en het samenstellen van een dataminingtabel.

Het kiezen van een geschikte dataminingvraag is een belangrijke stap, omdat dit de methode van analyse en de richting van de uitkomsten van de analyse bepaalt. Een miningvraag kan zijn: welk gedrag vertonen bezoekers op de website vlak voordat ze een abonnement opzeggen (zodat je potentiële opzeggers op tijd kunt benaderen), of welke producten kan ik het beste aanbieden aan klanten die recent een doorlopende reisverzekering hebben afgesloten?

De tweede stap in het dataminingproces betreft de keuze van de juiste dataminingtechniek. Iedere dataminingtechniek heeft specifieke voor- en nadelen. Zo zal de ene techniek wellicht tot betere voorspellingen leiden, terwijl een andere techniek juist beter interpreteerbaar en uitlegbaar is.

De derde stap, het samenstellen van een dataminingtabel, is een zeer belangrijke en zal verreweg de meeste tijd in beslag nemen (tot wel tachtig procent). Allereerst dient de dataminingtabel op het juiste niveau te worden samengesteld. Meestal zul je klantgedrag willen voorspellen

## “Datamining kan ook succesvol worden toegepast op webdata”

en zul je dus je miningtabel op klantniveau moeten samenstellen. Vervolgens moet de miningtabel worden aangevuld met variabelen die mogelijk een verband (correlatie) met de dataminingvraag kunnen hebben. Dit kunnen transactiegegevens, productbezitgegevens en klantcontactgegevens zijn, maar ook samengetrokken variabelen zoals klantwaarde, kanaalpreferentie per klant, kosten per klant, klachten per klant en koophistorie per klant.

Het samenstellen van een geschikte miningtabel (met internetgegevens) is tot op heden vaak een probleem geweest. De meeste webanalytics-pakketten leggen gegevens namelijk op sessieniveau vast en rapporteren ook op sessieniveau. Omdat we met datamining meestal uitspraken doen over behoefte en gedrag van klanten, dient de miningtabel meestal ook op klantniveau te worden opgesteld. Daarom zal er in het onderliggende datamodel van de webanalytics-pakketten nog een aggregatie van sessieniveau naar klantniveau moeten plaatsvinden.

### Online analyses

In het grote scala aan statistische technieken en algoritmes onderscheiden wij de drie groepen dataminingtechnieken die zich het best lenen voor online analyses: associatieve technieken, technieken waarbij de te



## Case 1: websitebezoek segmenteren

Een grote aanbieder van vakantie-reizen wilde achterhalen wat bezoekers op de website doen. Om hier inzicht in te krijgen is een clusteranalyse uitgevoerd, waarbij alle sessies over de laatste maand zijn meegenomen. Uit de clusteranalyse blijkt dat de volgende groepen sessies bestaan: korte sessies waarbij slechts op de aanbiedingen werd gefocussed, de zogenaamde *pricehunters*; extreem korte sessies, waarin slechts één of twee pagina's zijn bekeken, zogenaamd 'verkeerd verbonden'; lange sessies die zich verspreiden over de gehele website, de zogenaamde oriëntatiesessie; middellange sessie waarbij

de manier van boeken wordt onderzocht door de bezoekers, de zogenaamde 'boekingsvoorbereiding-sessie'; korte sessie waarin de aanbiedingen worden bekeken en de boekingsmodule wordt doorgelopen, meestal eindigend in een boeking, de zogenaamde 'boekers'.

Een nieuwe clusteranalyse op klantniveau liet zien dat deze verschillende sessies door alle klanten kunnen worden toegepast, en afhankelijk zijn van de fase van het beslisproces waarin de klant zich bevindt. Dit inzicht werd gebruikt om een nieuwe indeling van de website vorm te geven waarbij elk van dit type sessies optimaal werd ondersteund. Met andere woorden: de bezoeker werd gefaciliteerd in het behalen van zijn – of haar doelen op de website.

voorspellen variabele een binaire variabele is, en technieken waarbij de te voorspellen variabele een continue variabele is.

Associatieve technieken zijn de meest bekende vorm van datamining. Dit zijn technieken waarmee overeenkomsten binnen groepen en juist verschillen tussen verschillende groepen kunnen worden gebaseerd op geselecteerde variabelen. Bij associatieve technieken wordt datamining gebruikt om associaties en verbanden te leggen zonder dat er een doelvariabele is. Voorbeelden van dergelijke technieken zijn de clusteranalyse en de apriori-analyse.

Clusteranalyse is vooral bekend uit marktonderzoek en wordt veel ingezet voor klantsegmentatieonderzoeken. Het doel van de analyse is niet het voorspellen van klantgedrag, maar het zoeken naar een beperkt aantal homogene groepen die onderling sterk verschillen. Online kan deze techniek heel goed worden ingezet om websitebezoek te segmenteren. Case 1 laat dat zien.

## Apriori-analyse

De aprioritechniek is uitermate geschikt om er transactiedata mee te analyseren. De basket-analyse is hiervan het bekendste voorbeeld. Binnen (offline) retail is deze techniek veel ingezet om te bepalen welke artikelen veel samen worden gekocht, om zo de winkel optimaal in te richten (zodat bijvoorbeeld bier en zoutjes in hetzelfde schap staan). In online retail zijn de mogelijkheden van deze techniek nog vele malen groter, omdat je online geen last hebt van fysieke beperkingen.

Een bekend en succesvol online voorbeeld van toepassing van apriori-datamining is Amazon. Door uitgebreide analyse van transactiedata

## Case 2: Opzeggingen voorspellen

Een uitgever van een dagblad heeft de laatste jaren steeds meer moeite om abonnees te behouden. Om de uitstroom te verlagen wil de uitgever graag van tevoren weten welke abonnees een verhoogde kans hebben om de krant op te zeggen. Hiervoor wordt een churnmodel ontwikkeld. Veel van de klanten maken gebruik van de website van het dagblad om op de hoogte te blijven van het laatste nieuws. Gemiddeld wordt de website ongeveer drie keer per dag bekeken. Op internet zijn voor abonnees extra diensten gratis beschikbaar gesteld. Hiervoor moet wel op de site worden ingelogd, waardoor bekend is wie, wanneer, hoe vaak en wat op de site bezoekt. Deze gegevens kunnen worden gebruikt om opzeggers te voorspellen.

Er is een miningtabel opgesteld met alle klanten die in de afgelopen drie maanden hebben opgezegd. Vervolgens wordt een model gemaakt om te voorspellen bij welke klanten een grote kans bestaat dat zij hun abonnement opzeggen. Hieruit blijkt dat weglopers in de maand voordat zij opzeggen, meerdere keren de contactpagina met de algemene voorwaarden bekijken.

Klanten die hetzelfde gedrag gaan vertonen als de opzeggers, en dus een risicogroep zijn, komen nu in een aparte retentiegroep terecht. Daar worden extra marketingmiddelen ingezet om de klant alsnog te bewegen het abonnement te verlengen.

## De praktijk leert dat organisaties vaak nog niet klaar zijn voor structurele datamining van internetgegevens

wordt bij ieder product een extra suggestie gegeven. ('Mensen die dit boek kochten, hebben ook dit boek aangeschaft.')

## Doelvariabelen

Een voorbeeld van een miningvraag waarbij de doelvariabele een binaire variabele is, betreft het voorspellen van de kans op weglopen. De doelvariabele is in dit geval wegliep (ja/nee). Hierbij zijn technieken als logistische regressie, kansbomen en neurale netwerken geschikte algoritmen.

Is de doelvariabele een continue variabele is, dan is een voorbeeld van een miningvraag het voorspellen van de hoogte van de verwachte omzet per klant voor specifieke online marketingcampagnes. In dit geval is de doelvariabele de omzet per klant, deze kan een onbeperkt aantal waarden aannemen.

## Klein beginnen

Zoals we hebben laten zien in de voorbeelden kan datamining ook succesvol worden toegepast op webdata en zijn de inzichten die je ermee kunt verwerven van grote waarde. Investeer echter niet meteen veel geld in systemen en tools, maar begin klein.

Webanalytics is voor veel organisaties nog relatief nieuw. Richt je eerst op de (quick) wins die met webanalytics te bereiken zijn. Als dat proces onder de knie is, kun je als bedrijf eens naar datamining gaan kijken. Neem eens een steekproef uit je webdata en probeer daar een miningtabel van te maken. Levert dit nieuwe inzichten op, dan kun je hiermee verder gaan.

De praktijk leert dat organisaties vaak nog niet klaar zijn voor structurele datamining van internetgegevens, zoals dit voor offline direct marketing al wel is ingebakken in organisaties. Het advies is dan ook om bij deze afdelingen aansluiting te zoeken en langzaam het webkanaal te ontsluiten en onderdeel te maken van het offline datawarehouse. Online marketing hoeft in dit geval niet opnieuw het wiel uit te vinden, maar kan snel de inhaalslag maken en profiteren van het voorbereidende werk van databasemarketing in de afgelopen twintig jaar.

Bart Clement ([bart.clement@dikw.nl](mailto:bart.clement@dikw.nl)) is consultant advanced intelligence bij DIKW Consulting.

Bram van Essen ([bram.vanessen@vemic.nl](mailto:bram.vanessen@vemic.nl)) is zelfstandig consultant marketing intelligence.