

Wat datamining retailers oplevert

28-04-2016 11:15

Door Richard van Hoorn

Freelance marketing intelligence analist

Big data, iedereen heeft het erover en ook retailers denken na over toepassingsmogelijkheden van datagedreven marketing. Maar wat levert de inzet van datamining in retail op? Kunnen retailers door data te gebruiken de jaarlijkse 1,5 miljard aan reclamebestedingen slimmer inzetten?

Laten we uit de vele toepassingsmogelijkheden van retail datamining een klassieker voorzien van een eenvoudig rekenvoorbeeld. Het voorspellen van respons om de conversie van een direct mail campagne te verhogen. In dit voorbeeld stuurt een retailer een direct mail naar twintigduizend klanten, waarvan klantgegevens bekend zijn. Klanten die responderen hebben een netto klantwaarde van vijftig euro, dit is opbrengst per kopende klant. De direct mail kost precies 1 euro aan druk- en portokosten, per verzonden exemplaar. Bij een respons van drie procent zijn de opbrengsten $3\% \text{ van } 20.000 \times 50 = 30.000$. Daar gaan de kosten van de DM vanaf. Dit is $20.000 \times 1 = 20.000$. Er is dus tienduizend euro netto aan de testcampagne verdiend. Als er nu nog tweehonderdduizend vergelijkbare klantadressen beschikbaar zijn, moet je de actie dan grootschalig herhalen? Meteen doen, zal menig retailer zeggen, want je verdient door je aanbod naar tien keer zoveel adressen te sturen honderdduizend euro. Eerst eens even in de data graven, zou de data-analist zeggen. En hij heeft gelijk, het nettoresultaat van de actie kan oplopen naar 130 duizend euro.

Een kijkje onder de motorkap

Hoe gaat de analist te werk? De twintigduizend klanten van de testcampagne worden gesplitst in twee groepen. De eerste groep wordt gebruikt om het verband te vinden tussen de klanten die reageren op de directe mail en hun kenmerken, zoals eerdere aankopen en persoonsgegevens. Wat zijn gemiddelde eigenschappen van die kopers, in vergelijking met de niet-kopers? Hieruit volgt een voorspellend model, in datamining ook wel algoritme genoemd. Deze formule wordt op de tweede groep, waarvan de respons bekend is, losgelaten om te controleren of het model werkt. Na een aantal testen en verbeteringen kan de analist per klant voorspellen hoe groot de kans is dat iedere individuele klant op de actie reageert. Voor deze voorspelling is geen glazen bol nodig, gewoon gezond verstand en een computer. Het is toegepaste statistiek, geen hocus pocus.

Kleinere actie, hoger resultaat

De voorspelling kan nu veilig worden losgelaten op het bestand met tweehonderdduizend vergelijkbare adressen. Per adres is bekend hoe groot de kans op respons is. Van 1 tot 99 procent kans op respons worden de klanten op een rijtje gezet. Vanzelfsprekend worden vervolgens de klanten met een lage kans op respons niet gemaïld, want dat kost 1 euro per klant en levert naar verwachting niets op. De klanten met een hoge kans op respons ontvangen een aanbieding in de brievenbus. Deze voorspellende analyses zijn betrouwbaar en worden niet alleen gebruikt om respons te voorspellen, maar ook om te voorspellen welke klanten dreigen weg te lopen. Ook voorspellen wat en wanneer er mogelijk gekocht wordt is mogelijk. Meer verkopen, klanten winnen en behouden, dat is het doel van retail datamining.

De uitkomst

Stel dat de analist adviseert om 65 procent van de tweehonderdduizend te mailen. De omvang van de mailing wordt dan 130 duizend. Dit levert een besparing van $70.000 \times 1 = 70.000$ op. Bij een volledig betrouwbare voorspelling zou de respons dan stijgen van drie procent naar ruim 4,6 procent. Maar het netto aantal kopers blijft natuurlijk niet gelijk. Van de zeventigduizend die niet gemaïld worden zijn er

waarschijnlijk toch een paar verloren kopers. Een voorspelling is nooit volledig betrouwbaar.

Als we uitgaan van een respons van 4% over de 130.00 geadresseerden dan is de uitkomst:

Retail campagne			
Netto klantwaarde	€	50,00	
Prijs per DM	€	1,00	
		Vorige campagne	Nieuwe campagne
		Zonder analyse	Zonder analyse Na analyse
Aantal adressen		20000	200000 130000
Respons		3%	3% 4%
Opbrengsten	€	30.000	€ 300.000 € 260.000
Kosten	€	20.000	€ 200.000 € 130.000
Resultaat	€	10.000	€ 100.000 € 130.000

De opbrengsten van de datagedreven campagne bij een respons van vier procent zijn 260 duizend euro in plaats van driehonderdduizend. Omdat de kosten sterker zijn gedaald, tweehonderdduizend naar 130 duizend, is de winst gestegen van honderdduizend naar 130 duizend. Een gebruikelijk resultaat, dat nog hoger wordt door de mail te personaliseren, zoals beschreven in mijn vorige blog, de slimme actiefolder.

Geen software investeringen nodig

Moet je nu, enthousiast geworden door het lonkende datagoud, meteen als bezetene gaan investeren in computers en software? Nee, niet nodig. Het goede nieuws is dat je voor datagedreven marketing in retail nauwelijks software-investeringen nodig hebt. Data-analyses kunnen prima uitgevoerd worden met opensource oplossingen. Iedereen met een laptop en toegang tot de kassadata kan de bovenstaande analyse uitvoeren. Voor veel retailers is het wel van belang om de klanten te kunnen identificeren en de data voldoende verrijkt te hebben. Ook om analyses automatisch online uit te voeren geen hoge investeringen nodig. Wel moet je voldoen aan alle privacyregels. Nu al mag je niet zomaar alle persoonsgegevens vastleggen en gebruiken. Nieuwe Europese regels geven aan dat straks voor het profileren ook toestemming gegeven moet worden. Dit alles hoeft niemand te weerhouden om vanaf vandaag de reclamekosten naar beneden te brengen door persoonlijk en relevant te communiceren.

Richard van Hoorn gaat in een reeks blogs in op de toepassingen van datamining binnen retail. [Kijk hier voor meer informatie.](#)