

Whitepaper

Predictive analytics

Vincent Lukassen



Hoofdkantoor

Kruisboog 42
3905 TG Veenendaal
Tel. +31(0)318 - 55 20 20
Fax +31(0)318 - 55 23 55

Kenniscentrum

De Smalle Zijde 39
3903 LM Veenendaal
Tel. +31(0)318 - 50 11 19
Fax +31(0)318 - 51 83 59

info.nl@infosupport.com
www.infosupport.com
K.v.K. 3013 5370
BTW NL8062.30.277.B01

IBAN NL92 RABO 0305 9528 89
BIC RABONL2U
IBAN NL74 INGB 0004 7385 93
BIC INGBNL2A

Whitepaper

Predictive analytics

Titel	Whitepaper
Project/Onderwerp	Predictive analytics
Versie	1.1
Status	Definitief
Datum	19-jan-2015
Bestand	Whitepaper Predictive Analytics
Bedrijf	Info Support B.V.
Meer informatie	Voor vragen of meer informatie over deze whitepaper kunt u contact opnemen met Info Support door te bellen naar +31 (0) 318 55 20 20 en te vragen naar Sales Support & Marketing (Nederland) of te bellen naar +32 (0) 15 28 63 70 (België). U kunt ook een e-mail sturen naar sales@infosupport.com .

© Info Support B.V., Veenendaal 2015

Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze ook, zonder voorafgaande toestemming van **Info Support B.V.**

No part of this publication may be reproduced in any form by print, photo print, microfilm or any other means without written permission by **Info Support B.V.**

Prijsopgaven en leveringen geschieden volgens de Algemene Voorwaarden van **Info Support B.V.** gedeponeerd bij de K.v.K. te Utrecht onder nr. 30135370. Een exemplaar zenden wij u op uw verzoek per omgaande kosteloos toe.

Inhoudsopgave

1. Inleiding	3
1.1 Hoe word je president met predictive analytics?	4
2. Predictive analytics binnen BI	5
3. Het succes van predictive analytics	7
4. Aan de slag!	8
4.1 De lifecycle van predictive analytics	9
5. Conclusie	10
6. Over Info Support	11

1. Inleiding

Het is een vorm van BI die al een tijdje meegaat, maar steeds vaker op lijkt te duiken: predictive analytics. Het voornaamste verschil met de meeste andere business intelligence-modellen zit al in de naam besloten - *predictive*, oftewel voorspellend. Dat maakt het een van de meest praktische toepassingen van BI, maar ook een ingewikkelde.

Er zijn veel verschillende definities van predictive analytics in omloop, maar die van Eric Siegel vat het misschien wel het beste samen:

'Predictive analytics is een technologie die het toekomstige gedrag van mensen kan voorspellen op basis van ervaringen en data uit het verleden, met als doel om betere beslissingen te kunnen nemen.'

Het gaat dus om het voorspellen wat mensen gaan doen: welke consumenten besluiten om een bepaald product aan te schaffen of waarom willen ze een bepaald programma zien op tv? Maar ook (en deze vraag wordt met name beantwoord in de creditcard-industrie en door de politie): op welk moment gaan mensen frauderen of stelen?

Predictive analytics probeert dat allemaal te voorspellen op basis van het gedrag uit het verleden. Het klinkt misschien nogal futuristisch - wie de film *Minority Report* heeft gezien, zal daar aan denken. Maar het gebeurt al, zelfs al jaren.

Politiecorpsen in Amerika gebruiken predictive analytics om te voorspellen waar misdaden zullen gaan plaatsvinden. Hiervoor gebruiken ze allerlei sensoren, zoals camera's op straat, maar ook aangiften die zijn gedaan van misdaden en observaties van mensen waarvan crimineel gedrag wordt vermoed. Al die informatie wordt gecombineerd en in meldkamers wordt op stadsplattegronden real time geprojecteerd waar zich risicogebieden bevinden. De politie probeert dus te voorspellen waar misdaad gaat plaatsvinden en kan meer politie worden ingezet in wijken waar er sprake is van een verhoogd risico. IBM claimt zelfs dat met behulp van predictive analytics-oplossingen het misdaadcijfer in de stad Memphis met 30 procent is gedaald.

Predictive analytics biedt ook commerciële mogelijkheden. Zo ontdekte reisorganisatie Orbitz dat Apple-gebruikers vaak meer te besteden hebben dan gebruikers van andere systemen. Die informatie werd gebruikt door aan Apple-gebruikers duurdere hotelkamers aan te bieden; niet door ze twee keer zoveel te laten betalen voor dezelfde kamer, maar door andere, luxere kamers aan te bieden.

De nieuwste generatie televisies van Samsung beschikt over een systeem waarmee kan worden voorspeld wat mensen willen zien op tv. Ze krijgen een kijkadvies, gebaseerd op hun eerdere kijkgedrag. Netflix doet dat overigens ook.

Deze whitepaper gaat dieper in op het fenomeen predictive analytics. Welke rol speelt het binnen BI, en hoe pas je het toe?

1.1 Hoe word je president met predictive analytics?

Barack Obama heeft predictive analytics veelvuldig ingezet in zijn verkiezingscampagnes, zowel in 2008 en 2012. Op basis van bepaalde profielen wist het campagne team precies hoe groot de kans zou zijn dat een Amerikaan zou gaan stemmen voor de democraten of de republikeinen. Er wordt zelfs gezegd dat men van elk individu in Amerika wist wat deze zou gaan stemmen.

De methode: eerst werd in kaart gebracht wie lid was van een politieke partij, en dus zeker zou stemmen op de democraten of de republikeinen. Die informatie werd aangevuld met telefonisch onderzoek, waarin kiezers werd gevraagd waarop ze hadden gestemd. Met deze data in handen kon met behulp van predictive analytics worden bepaald welke eigenschappen (potentiële) stemmers op Obama hadden, en waarop het campagne team dus zijn pijlen moest richten.

In 2012 heeft het campagne team het nog fijnmaziger aangepakt: er werd ook gekeken naar de kans dat een kiezer kon worden beïnvloed. Men wist immers precies wie er in 2008 op Obama gestemd had. Daar schuilt natuurlijk ook een risico in.

Vergelijk het met een abonnement op een telecom-provider: op het moment dat je een berichtje krijgt dat je abonnement afloopt, ben je er dus ook van bewust dat je weer kunt kiezen (en dus ook kunt overlopen naar de concurrent).

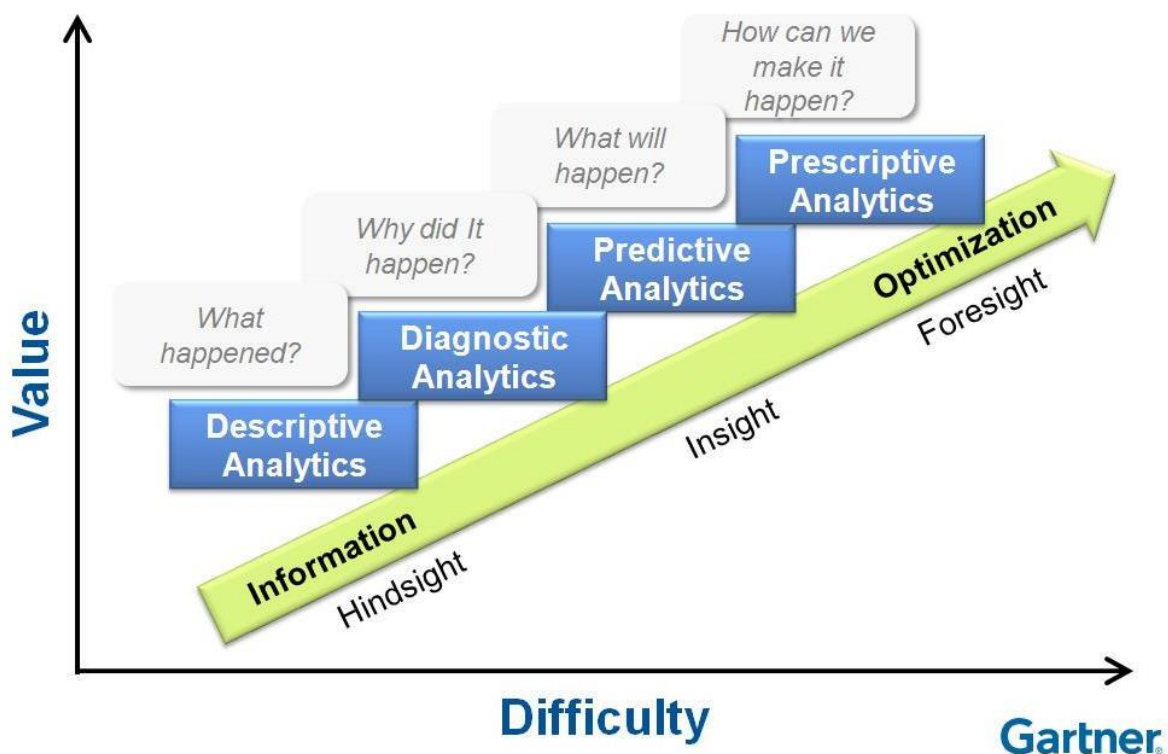
Daarom is er een kwadrant opgesteld, met twee kenmerken: stemde de kiezer al eerder op Obama, en is het de verwachting dat hij of zij dat weer gaat doen? Als het antwoord op beide vragen 'ja' was, dan werd deze kiezer niet benaderd. Alleen de kiezers van wie werd verwacht dat ze zouden kunnen worden overtuigd, werden proactief benaderd door het campagne team. Daarmee werden dus loyale kiezers uitgesloten, net als de kiezers die toch nooit op Obama zouden stemmen. Wat overbleef, waren de zwevende kiezers die nog beïnvloed zouden kunnen worden.

Op basis van analyse zijn kiezers heel gericht benaderd, onder meer met televisiereclames, die werden afgestemd op de profielen van de beïnvloedbare zwevende kiezers.

Zoals inmiddels bekend hadden de tegenstanders van Obama (Dick Cheney en Mitt Romney) het nakijken. Er werd gezegd dat hun campagne teams altijd een paar stappen achterliepen op de gerichte campagne van Obama.

2. Predictive analytics binnen BI

Om te kunnen begrijpen hoe predictive analytics zich verhoudt tot andere vormen van BI, nemen we het model van Gartner. Dit model onderscheidt vier soorten analytics: descriptive, diagnostic, predictive en prescriptive analytics.



- **Descriptive analytics** is de meest voorkomende vorm van BI, die eigenlijk altijd terugkijkt naar het verleden. Wat is er gebeurd, wie heeft wat gedaan, door wie en waar? De methode die wordt gebruikt is het verzamelen van veel data in een datawarehouse, waarna de gegevens worden opgeschoond, getransformeerd en in rapporten gevat om terug te kijken naar de afgelopen periode. Niet voor niets wordt dit ook wel eens 'rearview mirror BI' genoemd: het geeft een antwoord op de vraag 'wat is er gebeurd'.
- **Diagnostic analytics** kijkt ook naar het verleden, maar er wordt ook geanalyseerd *waarom* bepaalde dingen gebeurd zijn. Met behulp van social media-achtige toepassingen worden anderen betrokken in de vraag wat er in het verleden gebeurd is en waarom. De centrale vraag die hier wordt gesteld is: 'waarom gebeurde het?'.
- Natuurlijk is het ook belangrijk om te leren van het verleden, maar BI wordt pas echt nuttig als we er in de toekomst ook iets mee kunnen. Het volgende stapje is **predictive analytics**: dus voorspellingen kunnen doen op basis van data uit het verleden. Hierbij gaat het om de vraag 'wat gaat er gebeuren?'.

- Als je de toekomst kunt voorspellen, is **prescriptive analytics** de ultieme volgende stap: welke toekomstscenario's zijn er en hoe kunnen we die beïnvloeden? 'Hoe kunnen we het laten gebeuren?' of 'Wat kunnen we doen op het moment dat het gebeurt?' zijn de vragen die hier centraal staan. Deze techniek wordt onder meer gebruikt door creditcardmaatschappijen, om te analyseren of transacties echt zijn of dat er sprake is van fraude. Bedrijven willen fraude direct kunnen detecteren en er iets aan doen.

De BI-technieken waarmee je de toekomst kunt voorspellen bieden voor bedrijven natuurlijk meer waarde dan alleen maar het aantonen wat er gebeurd is. Dat maakt het tegelijkertijd ook ingewikkelder: wat er is gebeurd weet je zeker en staat vast, wat ons in de toekomst staat te wachten niet en kunnen we beïnvloeden.

3. Het succes van predictive analytics

Zoals gezegd wordt predictive analytics al jaren toegepast. Volgens Gartner staat predictive analytics dan ook aan de rechterkant van de hype-cycle, wat betekent dat het mainstream is geworden en dat we er allemaal al mee bezig zouden moeten zijn. Waarom hoor je er dan de laatste tijd ineens zo veel over?

Ten eerste gaan organisaties steeds meer het nut inzien van predictive analytics. Voorheen waren het met name creditcardmaatschappijen, marketingbureaus en verzekeringsmaatschappijen die inschattingen wilden maken van respectievelijk fraude, consumentengedrag of premies. Banken konden door middel van predictive analytics inschatten of klanten hun leningen zouden afbetalen.

Tegenwoordig hebben meer organisaties de mogelijkheid om deze vorm van BI in te zetten, omdat er betere en goed betaalbare software en hardware beschikbaar is om complexe analyses te maken.

Daarnaast zien steeds meer bedrijven de mogelijkheden van Big Data in. Er zijn ontelbaar veel openbare databronnen waar we informatie vandaan kunnen halen en vervolgens kunnen we die combineren met eigen data. Je hebt echter pas echt iets aan Big Data als je ze ook kunt analyseren - alleen gegevens opslaan is niet genoeg.

Succesfactoren

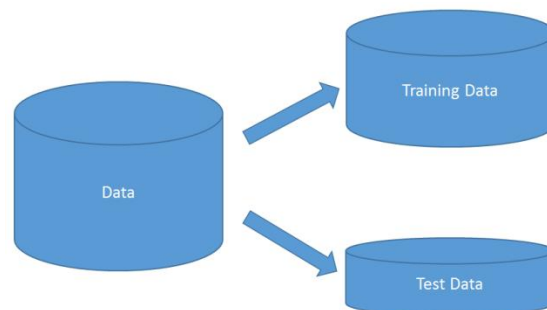
Wat zijn nu de voorwaarden voor het succesvol inzetten van predictive analytics? Er zijn vijf factoren te onderscheiden die de slagingskans beïnvloeden:

1. Het succes van predictive analytics valt of staat - zoals zo vaak in de IT - met het stellen van **heldere doelstellingen**. Wat willen we onderzoeken? Wat is het doel (kostenreductie, winstmaximalisatie of klanttevredenheid)? Hoe wil je daar komen? Stel, een telecombedrijf wil via predictive analytics de omzet vergroten, dan zou je als doelstelling kunnen formuleren dat je de kans wilt verkleinen dat klanten hun abonnement opzeggen.
2. Net zo belangrijk is het om vooraf te bepalen wat de organisatie uiteindelijk met de uitkomsten gaat **doen** en of zo'n vervolgactie ook **haalbaar** is.
3. Een volgende stap is het **vinden van de data**, die de basis vormen van de analyse. De juiste **kennis en tooling** is vervolgens nodig om de analyse te kunnen uitvoeren. Deze kennis heeft niet alleen betrekking op de data-analyse, maar zeker ook op de processen van het bedrijf.
4. De ervaring leert dat het zinvol is om **klein te beginnen** met het toepassen van predictive analytics. Een project dat eenvoudig uit te voeren is zorgt voor een lagere drempel om ook echt met de techniek aan de slag te gaan. Hoe meer beren er mogelijk op de weg zijn, des te groter de kans is dat die weg ophoudt voordat het doel is bereikt.
5. Tijdens het proces, tenslotte, is het met name belangrijk om goed te blijven **testen en monitoren** of de analyse nog steeds werkt. Iets dat vandaag geldt, hoeft morgen niet meer te gelden.

4. Aan de slag!

De belangrijkste basis om met predictive analytics aan de slag te gaan, is het **creëren van een productiedataset**. Van daaruit kunnen we een model bouwen om voorspellingen te kunnen doen.

Hiervoor moet de dataset worden **gesplitst** in trainingdata en testdata. Met de eerste set maken we het model, met de tweede set kunnen we het bewijzen. De trainingdata worden ingevoerd in een analysemodel, dat door middel van algoritmes verbanden probeert te vinden tussen data - ook wel *machine learning* genoemd. Met de testdata kan worden onderzocht of de voorspellingen ook uitkomen - en of het model dus werkt.



Het is belangrijk om binnen de analyse een splitsing te maken tussen bepaalde inputvariabelen en uitkomsten. Zo kun je bijvoorbeeld berekenen welke combinatie van gegevens in het verleden hebben geleid tot de aanschaf van een product.

Om oorzaak en verband écht goed van elkaar te kunnen scheiden, is het van groot belang om een zogeheten 'marker' te zetten in de dataset - dat is een datum uit het verleden. Alles voor die datum geldt als oorzaak, alles na die datum is een uitkomst.

In de dataset proberen we dus **verbanden** te ontdekken. Dat is nog best ingewikkeld, omdat een correlatie tussen twee variabelen nog niet betekent dat ze elkaar ook hebben beïnvloed. Om maar een simpel voorbeeld te noemen: het is heel goed denkbaar dat er een sterke correlatie is tussen het aantal ijsjes dat op een dag in Nederland wordt gekocht en het aantal openbare dronkenschappen. Maar dat wil niet zeggen dat mensen dronken worden van ijs of dat dronken mensen vaker ijs eten. Aan beide factoren ligt wellicht een andere oorzaak ten grondslag: mooi weer. Als de zon schijnt en de temperatuur hoog is, zullen mensen eerder ijs kopen en een terrasje opzoeken.

Het sleutelwoord hierbij is dus causaliteit. Als factor A en B met elkaar correleren, dan kan het zijn dat A beïnvloed wordt door B, maar het kan ook precies andersom zijn. Bovendien kunnen beide factoren, onafhankelijk van elkaar, worden beïnvloed door factor C.

Machine learning helpt bij het aantonen van causaliteit. Bijvoorbeeld door gebruik te maken van een zogeheten 'decision tree', waarin de oorzaken en gevolgen van verschillende factoren op een rij worden gezet.

Naast het decision tree model zijn er nog andere soorten analyses, waaronder basket analyses, waarin je een verband kunt aanbrengen tussen de aanschaf van verschillende producten: consumenten die A kochten, kopen ook vaak B. Dat kan bijvoorbeeld goed van pas komen bij de indeling van een winkel. Ook webshops maken hier vaak gebruik van: 'consumenten die dit boek aanschafte, kochten ook vaak dit boek'.

4.1 De lifecycle van predictive analytics

Hoe ziet de lifecycle van een predictive analytics-project eruit? Forrester identificeert de volgende fases:

1. Understand data: begrijp welke data je hebt en hoe je deze kunt analyseren
2. Prepare data: creëer een model op basis van deze data waarop je je predictions kunt gaan baseren
Predictive model: op basis van de analyse genereer je een of meerdere modellen
3. Evaluate: een model moet altijd worden geëvalueerd, eerst op basis van de testdata, vervolgens op basis van andere data. Je kiest vervolgens het model met de beste voorspellende waarde.
4. Deploy: vervolgens kun je het model in productie gaan nemen.
5. Monitor: je blijft een model continu monitoren, en daar aanpassen waar nodig. Wat vandaag de waarheid is, kan morgen weer veranderd zijn. Zo probeer je je model continu beter te maken.



Source: Forrester Research, Inc.

5. Conclusie

Je kunt met predictive analytics dus heel veel te weten komen over je doelgroep, en daar je acties zorgvuldig op afstemmen. Maar het gaat uiteindelijk natuurlijk niet om de analyse, maar om wat je ermee doet. Predictive analytics is slechts een rekenmodel. Bovendien moet je er altijd rekening mee houden dat mensen (en dus hun gedrag) kunnen veranderen - dus het model van vandaag hoeft morgen niet per se meer te gelden.

Het is daarom belangrijk om het rekenmodel voortdurend te blijven bijstellen, en om niet blind te varen op de uitkomsten van een analytisch model.

6. Over Info Support

Info Support is opgericht in 1986 en is met ruim 350 medewerkers in Nederland een vooraanstaand IT-dienstverlener op het gebied van IT-consultancy, software -ontwikkeling, opleidingen en beheer. Info Support is niet beursgenoteerd en financiert de verdere ontwikkeling van de organisatie op basis van een beheerste groei uit eigen middelen.

Onze drive achter de oplossingen die wij realiseren voor onze klanten is er sterk op gericht bedrijfsprocessen sneller en beter te maken. Info Support ontwikkelt en beheert solide en innovatieve softwareoplossingen die organisaties ondersteunen bij het realiseren van hun doelstellingen.



De kernwaarden Soliditeit, Integriteit, Vakmanschap en Passie typeren onze werkwijze, waarin we sociaal en solide management belangrijker vinden dan omzetmaximalisatie. Ons hoogste doel is dat we met opdrachtgevers en medewerkers willen bouwen aan langetermijnrelaties. Daarbij houden we ons aan gemaakte afspraken. Dit maken we in de praktijk waar, getuige de jarenlange relaties die we met onze klanten hebben. Info Support mag zich al 16 jaar op rij TOP-IT-werkgever van het jaar noemen.

Zie voor meer informatie www.infosupport.com.