

PARELTJES VINDEN IN DE DATABERG



De databerg wordt ongeveer iedere 14 maanden twee keer zo hoog. Databestanden met aankoopgegevens, bezochte locaties, klantgegevens en surfgedrag herbergen een schat aan informatie over trends. Door grote databestanden met elkaar te koppelen, kunnen bedrijfsprocessen worden verbeterd en nieuwe diensten ontwikkeld. Big data wordt big business. Dr. Sandjai Bhulai (Vrije Universiteit) belicht praktijkvoorbeelden.

Wat doet u als de pinautomaat bij uw bank niet werkt? Belt u het servicenummer op uw bankpas? Waarschijnlijk niet. Teveel gedoe. De kans is groot dat u geïrriteerd doorloopt naar de volgende automaat. En het zou zomaar kunnen dat u er en passant een tweet uitgooit om uw irritaties kracht bij te zetten. Welnu, dat laatste is voer voor sentimentanalyses en ... betere dienstverlening. Met een beetje geluk en technisch inzicht is de locatie van de pinautomaat veel sneller gevonden dan op de gebruikelijke manier. Want hoeveel tijd was er overheen gegaan voordat uw klacht uiteindelijk bij de dienstdoende servicemonteur van de bank terecht zou zijn gekomen?

Big data, big business

De koppeling van grote databestanden gaat leiden tot ontwikkeling van allerlei nieuwe producten en diensten. Big data wordt big business. Iedereen roept het. Maar hun beste praktijkvoorbeelden geven organisaties niet graag prijs. Daarom gaan we te rade bij een 'onafhankelijke' deskundige. Dr. Sandjai Bhulai promoveerde in 1994 op 'hoogdimensionale' (zeg maar: complexe) computersystemen. Daarna maakte hij wiskundige modellen voor telecommunicatiebedrijf Lucent. Sinds 2003 is Bhulai universitair hoofddocent aan de VU, met een focus op contactcenters en

telecombedrijven. Verder is hij verbonden aan het CWI (Centrum Wiskunde & Informatica) en heeft hij zijn eigen bedrijf, Prompt Business Analytics, dat software levert aan de zorgsector. Volgens Bhulai moeten organisaties die mee willen in de snelle ontwikkelingen beter gebruikmaken van de nieuwe kansen door koppeling van databestanden.

Voorspellend algoritme

Zomaar wat willekeurige cijfers. Volgens Kwik-Fit Autoservice (182 vestigingen) rijden 440.000 Nederlandse automobilisten in de zomer door op winterbanden. De ANWB heeft bijna 4 miljoen leden, waarvan er jaarlijks 1 miljoen telefonisch contact opnemen. ING heeft 2.400 geldautomaten (waarvan 700 bij Albert Heijn). In Nederland hebben ongeveer 2,5 miljoen mensen een contract voor groene stroom. Het HagaZiekenhuis in Den Haag telde in 2011 ruim 34.000 dagbehandelingen. Aan al deze namen en cijfers hangen kengetallen en databestanden. Hoewel de cijfers op zichzelf niets met elkaar te maken hebben, zou een beetje wiskundige er met een voorspellend algoritme zomaar nieuwe dienstverlening of efficiëntere organisatiestructuren uit kunnen halen. En dat is in deze tijd van bezuinigingen waar alle bedrijven en organisaties naar op zoek zijn. Het moet efficiënter, slimmer en klantgerichter in 2013 en het nieuwe groeipotentieel is niet ver weg. Met andere woorden: het ligt voor het grijpen op uw eigen harde schijf. Onderzoek toont echter aan dat we veel minder datagedreven zijn dan onze directeuren denken. Softwareleverancier QlikTech vroeg 1.000 kantoormedewerkers in Groot Brittannië naar hun datagebruik en constateerde dat 90 procent van hen meer data verwerkt dan een jaar geleden maar dat slechts een kwart ook echt nieuwe dingen heeft ontdekt over de bedrijfsvoering. Bhulai: "De valkuil is dat organisaties gegevens gaan verzamelen en dan niet weten wat ze er mee aan moeten. Je moet heel goed voor ogen hebben wat je van je klanten wilt weten."

E-thermostaat

Veel gegevens zijn openbaar en makkelijk toegankelijk. Zo kunnen de gegevens van het KNMI door iedereen geraadpleegd worden. Dat deed Bhulai voor energiebedrijf Essent, dat eind 2011 de E-thermostaat introduceerde. Met de E-thermostaat kunnen klanten van Essent op afstand met een smartphone of computer de temperatuur in hun huis regelen. Bhulai kreeg anonieme klantgegevens van Essent over gemeten binnentemperaturen en de instelling van de thermostaten op gegeven tijdstippen. Deze gegevens koppelde hij aan de 23 bekende variabelen van het KNMI. Bhulai: "De vraag was wat de klant zou gaan instellen en of we trends konden detecteren in diens gedrag. Daaruit kwam naar voren dat de windsnelheid veel effect kan hebben op wat de klant instelt. En daaruit konden we afleiden dat 80 procent van de huizen slecht geïsoleerd is. Daar zou Essent een nieuwe dienst omheen kunnen bouwen." Zoals Essent u in het najaar een energiescan op uw tochtige woning kan aanbieden, kan Albert Heijn met behulp van de klantenkaart uw ideale looproute in de winkel

vaststellen. En Google weet waarschijnlijk meer van uw depressieve klachten dan uw eigen levenspartner. Hij of zij krijgt te horen dat u 's ochtends geen zin hebt om op te staan. Terwijl u in Google al lang hebt gezocht op specifieke zoektermen als 'manisch depressief, 'angstremmer' of 'alternatieve geneeswijze'.

Nieuwsdetector Twitter

Slim gebruikmaken van big data werd enkele jaren geleden nog business intelligence genoemd. Maar business intelligence ging over het verleden en was daarom niet slim genoeg. Inmiddels zijn we een stap verder. Het nieuwe vakgebied heet business analytics en voorspelt de toekomst. Die is strategisch vele malen belangrijker. Een knap staaltje toekomst voorspellen is de manier waarop de VU gebruikmaakt van Twitter. Een wiskundige als Bhulai ziet Twitter als een enorme interface, een speeltuin vol semantiek, nullen en enen, waaruit bruikbare informatie te halen valt. Bhulai: "In Nederland worden vijf miljoen tweets per dag verstuurd. Daarmee zijn we internationaal een van de koplopers. Hoe kunnen we in al die tweets nieuws en trends ontdekken? Dat leek ons voor de VU een interessant vraagstuk. Het kan gaan om een kapotte pinautomaat, de vermissing van personen, een brand of een aardbeving. We willen er bovenop zitten als de eerste berichten zich aandienen. We detecteren trends in wording."

Dashboard

Bhulai en zijn team ontwikkelden een dashboard. Daarvoor moesten wel wat hobbels genomen worden. Zo wilde Twitter slechts een klein deel van de data geven. Daarom schakelde Bhulai een team van semantici in, waarmee in totaal 300 typisch Nederlandse woorden werden geselecteerd zoals 'want', 'ook' en 'omdat'. Daarmee wordt meer dan 90 procent van alle tweets binnengehaald. Op basis van clustering van woorden die met elkaar te maken hebben kan Bhulai de versnelling meten van berichtgeving over een bepaald onderwerp. Spam wordt eruit gefilterd en het echte nieuws blijft over. NU.nl werkte mee aan een test en haalde er gemiddeld twee nieuwsitems per dag uit. Achter zijn beeldscherm toont Bhulai de tien trending topics op 20 maart 2012. Op één staat #PrayForMexico op de dag van de zware aardbeving in de buurt van de stad Acapulco. Een grafiek toont de ontwikkeling van de berichtenstroom. Bhulai: "Kijk, 's ochtends om twee minuten over half acht kwamen de eerste tweets binnen. Op dat moment waren we er al bij want we willen inzicht in de versnelling. Zo ontdekken we nieuws voordat het nieuws is."

Met één muisklik

Ook contactcenters kunnen veel beter gebruikmaken van de nieuwe mogelijkheden door hun eigen bestanden te koppelen aan bijvoorbeeld social media. Bhulai: "Aan ons als wiskundigen de taak alle verschillende datastromen op één plek zien te krijgen in één datawarehouse. Zodra er gebeld wordt door een klant, moet de klantadviseur al

volledig geïnformeerd zijn. Hoe was het klantgedrag op internet? Welke concurrenten zijn op het web bezocht? Hoe was het surfgedrag op de eigen website? Is de klant al in de webshop geweest? Wanneer was de klant in de fysieke winkel? Welke transacties hebben daar plaatsgevonden? Welke gegevens staan er op de klantenkaart? Welk second best offer kan de klant worden aangeboden? Veel bedrijven zijn nog lang niet zover dat ze al die datastromen met één muisklik beschikbaar hebben.”

Telco en energie

Telecombedrijven en energiebedrijven zijn volgens Bhulai goed op weg om uit hun databergen nieuwe markt- en klantgegevens te halen. Sectoren die het minder goed doen zijn de kledingretail (“die kennen hun klanten niet”) en de zorgsector. De vraag is of ze zich kunnen veroorloven om deze nieuwe kansen onbenut te laten. De zorgsector staat voor de enorme uitdaging om met beperkte middelen de groeiende vraag naar zorg op te vangen. De data die zorginstellingen hebben, herbergt een schat aan informatie die gebruikt kan worden om de organisatie te verbeteren, efficiënter te werken, kosten te besparen en de klant beter te helpen.

Bhulai: “We weten bijvoorbeeld dat ziekenhuizen die patiënten opnemen vaak geen idee hebben wanneer ze ze weer ontslaan. Toch valt dat met big data heel goed te voorspellen. Op die voorspellingen kunnen ze hun capaciteit instellen. Met big data krijgen zorginstellingen prestaties inzichtelijk op het gebied van capaciteitsmanagement, patiëntenzorg, wachttijden en behandelduur. Door toepassing van wiskundige beslismodellen kunnen ze analyseren en berekenen welke scenario’s gevolgd moeten worden om tot de gewenste wachttijd en capaciteit te komen. Beleidskeuzes worden straks meer gemaakt op basis van feiten en minder met de onderbuik.”

Valkuil

De databerg wordt steeds groter. Het datavolume verdubbelt iedere 14 maanden. De grote kunst is te zorgen dat relevante data in iedere laag van de organisatie terechtkomt. Van de directeur tot de klantadviseur, allemaal halen ze hun eigen gegevens uit een en dezelfde informatiebron. Het contactcenter krijgt belangrijke klantinformatie, de afdeling Projecten krijgt informatie om urenrekeningen te kunnen maken, de afdeling Service krijgt informatie over het soort onderhoud en het beschikbare personeel en de afdeling Finance krijgt gegevens om liquiditeitsprognoses te kunnen maken. Bhulai: “Elke afdeling heeft zo haar eigen stuurinformatie. De valkuil is dat alle afdelingen hun eigen toko runnen, en alle leden van de directie dat ook doen. De CIO, CFO, CEO, CCO en CTO moeten hetzelfde belang dienen en elkaar helpen om toegevoegde waarde.

Tekst: Ton Verheijen