

Datamining vaak onbetrouwbaar

Gevonden verbanden uit onsamenhangende gegevens zijn vaak niet causaal van aard

**Lambert van der Bruggen,
John Poppelaars en
Harry Veenendaal**

Toverwoord van veel ondernemingen en overheidsorganisaties is datamining, een methode om uit een database met veel onsamenhangende gegevens, bepaalde verbanden te ontdekken. Maar bij verkeerde toepassing is het een onbetrouwbare methode en dat gebeurt in de praktijk vaak.

Financiële dienstverleners gebruiken datamining om bijvoorbeeld de kredietwaardigheid van cliënten te toetsen. Ziekenhuizen trachten vooraf een bepaalde zorgvraag vast te stellen en politiediensten zoeken combinaties van persoonskenmerken om verdachten op te sporen.

De meeste gevonden verbanden hebben geen enkele causale rela-

tie met elkaar, maar worden toch als vaststaand feit aangenomen met een voorspellende waarde. Dit komt door gebrek aan economische kennis ter zake.

Het is steeds gemakkelijker verbanden te ontdekken en zelf te interpreteren zonder dat statistische experts nodig zijn. De vele ICT-systemen van bedrijven en instanties nodigen daartoe uit, zij bevatten immense hoeveelheden gegevens die ooit al dan niet met een bepaald doel zijn vastgelegd. De vraag is of dergelijke resultaten van datamining meer zeggen over de kwaliteit van de verzamelde data en de gebruiker dan over de mogelijk verborgen verbanden.

Tussen ontdekken en interpreteren zit een wereld van verschil. Een voorbeeld: neem data rond de ontwikkeling van vogels over een bepaalde periode. Voeg gegevens toe over de kindergeboorte over diezelfde periode. Ongetwijfeld

kan een verband worden geconstateerd tussen toename van het aantal ooievaars en stijging van het aantal baby's. Van een causaal verband is geen sprake.

Een ander voorbeeld: Euronext maakt gebruik van datamining voor het opsporen van verdachte transacties bij beursgenoteerde ondernemingen. Via geavanceerde patroonherkenning worden koers- en omzetafwijkingen geconstateerd die mogelijk duiden op handel met voorkennis. Daarop wordt de Autoriteit Financiële Markten ingeschakeld die al dan niet een onderzoek instelt. De

ICT-systemen met heel veel gegevens maken het steeds gemakkelijker allerlei verbanden te leggen

AFM kan op korte termijn de handel in een aandeel stilleggen, zoals recent enkele malen gebeurde. Een verband tussen die afwijkingen en handel met voorkennis is in eerste instantie net zo onzeker als dat tussen ooievaar en baby.

Een eindgebruiker die geen diepgaande kennis heeft van econometrie is nauwelijks in staat juiste conclusies te trekken. Datamining is in zekere zin goochelen met cijfers, met de wens als vader van de gedachte. Ondernemingen en overheidsinstanties moeten daarom waken voor een tunnelvisie, verbanden niet klakkeloos aannemen en de relatieve waarde van datamining onderkennen.

Ir. L.J.J. van der Bruggen is vicepresident Business Consulting, drs. John Poppelaars is Business Development Manager en mr. drs. H.F. Veenendaal is Legal Counsel bij Ortec.

