

DE OVERHEID IS WATCHING YOU

VDAB OMARMT DATAMINING



Steven Genbrugge

Adviseur Studiedienst
Vlaams ABVV

Niemand ontsnapt aan het **vangnet van Big Data**. Ons consumentengedrag wordt vastgelegd via onze e-commerce aankopen en digitale klantenkaarten, onze clicks en likes worden geregistreerd, en op basis van onze vriendenlijst, muziekvoorkeur en andere info bepaalt Facebook onze seksuele geaardheid. Op de werkvloer kan de nood aan onderhoud van de machine voorspeld worden, en worden onze prestaties en afwezigheden nauwgezet geregistreerd. We zijn ook vertrouwd met de Siri-stem die onze vragen met tal van persoonlijke suggesties beantwoordt en we schrikken er niet meer van als bol.com aangeeft dat een hippe zitbank ook wel interessant voor ons is net nadat we een forse grasmaaier kochten. Ons rechtvaardigheidsgevoel krijgt zelfs een opkikkertje als we lezen dat controlediensten Financiën data-analyse toepassen om fiscale fraude te detecteren. En in onze mailbox worden vervelende mails geweerd door hen het label spam te geven. Kortom we maken allemaal ongewild gebruik van technieken zoals datamining en dataprofiling.

Data verzamelen doen we natuurlijk al sinds mensenheugenis. Het analyseren van data is onderdeel van zowat elke wetenschap en bestaat er in om beschrijvend of diagnostisch om te gaan met bepaalde vraagstukken. Aan de hand van een salaris enquête zal men beschrijven hoeveel elke beroeps categorie verdient, en gaat men op zoek naar oorzaken en gevolgenfactoren die deze loonverschillen kunnen duiden, zoals een verhoogde vraag naar arbeidskrachten, gevaarlijk werk, hoge productiviteit.

Met datamining zoeken we naar kennis, verbanden en dus eigenlijk verborgen patronen in grote databanken.

Met **datamining** en **profiling** gaan we echter een stapje verder. Als je weet waarom je in het verleden iets gebeurd is, kan je ook een poging doen om de toekomst te voorspellen. En dat kan bestaan uit een heel eenvoudige koppeling van gegevens. Als we weten dat 40% van de arbeiders metaal een reis gepland hebben naar Frankrijk dan kunnen we redelijk eenvoudig een voorspelling doen over het reisgedrag van deze groep in het volgend jaar. En dan kunnen we de marketing en inkoop van reizen hierop afstemmen. Een bekend voorbeeld hiervan is Amazon dat de pakjes die nog niet besteld zijn toch al op basis van deze predictive analyse opstuurt naar zijn verdeelcentra. Het voorspellen van het gedrag van mensen en ze indelen in groepen wordt als profiling benoemd. Met datamining zoeken we naar kennis, verbanden en dus eigenlijk verborgen patronen in grote databanken.

Digitalisering is al lang geen vooruitgangsgedachte meer, digitalisering is een werkelijkheid.

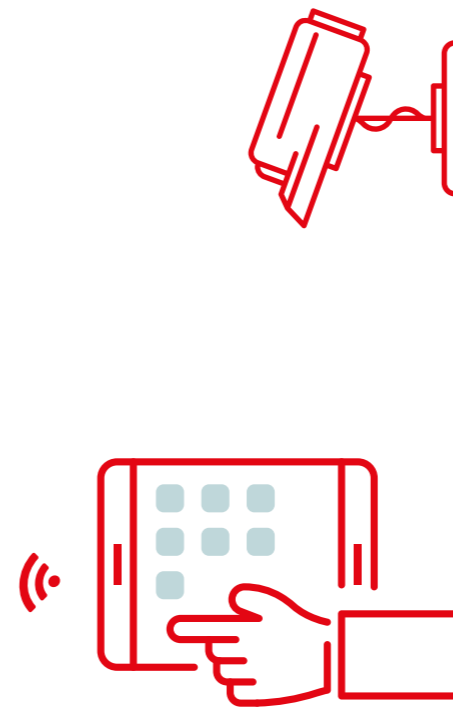
Ook de overheid ontdekte dat er met de gigantische stroom informatie waarover ze beschikken heel wat toepassingen binnen

het bereik liggen om hun dienstverlening of opdracht te kunnen verbeteren en dit vooral binnen een context van **besparingen**. De fiscus bepaalt op basis van datamining de risicoprofielen die vatbaar zijn voor controle. Een aanpak die als vorm van efficiëntie in de behandeling van de onderzoeken in elk geval ook aangeeft dat de overheid al lang niet meer in staat is om alle belastingplichtige personen en vennootschappen te controleren. En bij gebrek aan voldoende personeel dus risico's op ongelijke en oneerlijke benadering installeert. Om te vermijden dat (dezelfde groep) bedrijven en personen steeds opnieuw de dataminingsradar kunnen ontlopen moeten voldoende menselijke factoren worden ingebouwd. Voldoende inspecteurs die hun terreinkennis kunnen inzetten aanvullend op deze digitale selecties en hen de mogelijkheid blijven bieden om open, doelgerichte controles te laten uitvoeren zijn essentieel om deze gevaren te kunnen corrigeren.

Verbeteren van kennis, manipuleren van gedrag

Maar ook in andere bevoegdheidssterreinen graaft men naarstig naar en in de dataschaten. Een van de instellingen die fel inzet op digitale processen is de Vlaamse Dienst voor Arbeidsbemiddeling en Beroepsopleiding (VDAB), vooral door het ontwikkelen van nieuwe toepassingen die heel wat voorspellende informatie aanleveren. VDAB beschikt niet alleen over honderdduizenden digitale dossiers van werkzoekenden maar ze ontvangen jaarlijks ook tienduizenden nieuwe vacatures. VDAB zorgt voor een **automatisch vacatureaanbod** door in enkele milliseconden honderdduizenden werkzoekenden te matchen met de tienduizenden vacatures.

Daarnaast beheert de overheid in tal van andere levensdomeinen databanken die de burgers en diensten toelaten dezelfde informatie niet altijd opnieuw te moeten opgeven. Door het koppelen van deze databanken kan de VDAB ook andere parameters uit de loopbanen kruisen met hun bemiddelingsgegevens. En nog meer tot de verbeelding spreekt de registratie van het **clickgedrag** van diezelfde mensen wanneer ze gebruik maken van de tools die VDAB beschikbaar stelt. Bij het navigeren wordt het gedrag van de surfer nauwgezet bijgehouden: welke vacatures klikt men aan, hoeveel bekijkt men er op de website, hoeveel tijd besteed hij aan een kijkje nemen... Deze gigantische berg aan dossier- en arbeidsmarktinformatie vormen samen met de matchings-, navigatie- en historische gegevens, de Big Data van de VDAB. De VDAB heeft nu in samenwerking met de KU Leuven en de Vlerick Business School algoritmes ontwikkeld die een inzicht in het **zoekgedrag** van de werkzoekende genereren. Via een statistische benadering kunnen er aanbevelingen worden geformuleerd die de klant moet vooruithelpen in zijn zoektocht. Zo kunnen er vacatures worden aange-reikt die in het verleden een grote matchbaarheid hadden bij soortgelijke profielen of kunnen er suggesties op maat worden voorgesteld om andere acties te ondernemen in functie van het gekozen jobdoelwit. Hetzelfde systeem van data-analyse kan ook gebruikt worden om onpopulaire vacatures bij de juiste werkzoekenden aan te bevelen. Het opent ook de mogelijkheid om de huidige onevenwichten in vacaturebereik bij te stellen. Uit interne audits blijken namelijk heel wat vacatures momenteel quasi (+ 70%) niet bekeken, en slechts een beperkt deel van



“
De digitalisering is een realiteit en het graven en exploiteren van dit nieuwe goud zal nog exponentieel toenemen.



dit aanbod wordt veelvuldig aangeklikt. De data-analyse opent de deur om commerciële vacatures waar de werkgever expliciet een VDAB-tussenkomst vraagt naar werkzoekende te pushen. Al deze hulpmiddelen hebben één ding gemeen: ze voorspellen ons gedrag. Ze doen dit door data te verzamelen in **grote datasets**, deze te analyseren en vervolgens patronen en verbanden voor te stellen die een voorspellende uitspraak doen over het gedrag van de klant/gebruiker. Het opgraven van deze informatieschat heet **Datamining**, het proces om via deze algoritmes voorspellende modellen te creëren die door een voortdurende stroom aan nieuwe data gevoed en aangepast worden, omschrijft men als **Machine Learning**.

Een dam opwerpen tegen deze ontwikkelingen is niet mogelijk noch wenselijk. De vloedstroom aan data zijn belangrijke bouwstenen in de ontwikkeling van nieuwe technologieën. Maar innovaties worden geboren vanuit een rationaliteit. Het is aan de beleidsmakers om er ook een laagje moraliteit op aan te brengen. Anders kunnen het disruptieve mechanismen worden die de rechten en vrijheden van de burgers, werkzoekenden bruuskeren. Een kritische lezing is op zijn plaats om de onderhuidse gevaren van die wonderbaarlijke technologieën goed te kunnen inschatten.

Is de dataverstrekker op de hoogte?

De databergen zijn ondertussen zo immens dat het onmogelijk is deze te onderwerpen aan een menselijke controle. Het gevolg hiervan is dat we veroordeeld zijn om slaafs de correctheid van de algoritmes te volgen.

Volgens professor Max Welling, hoogleraar Machine Learning aan de Universiteit van Amsterdam is het naïef om naast de comfortoportuniteiten blind te blijven voor de gevaren van de dataficatie: privacy schendingen, misbruik van gegevens, het trekken van verkeerde conclusies, de ontmenselijking van de dienstverlening, de verdringing van arbeidsplaatsen door automatische systemen.

VDAB legt de werkzoekende conform de wet op de privacy wel voor dat de gegevens door hen worden verwerkt en dat de ze op elk moment kunnen worden bekeken en aangepast. Maar het geeft de werkzoekende geen inzicht in de wijze waarop VDAB de registraties uitvoert en op welke manier de verwerking van de gegevens verlopen. Nog minder zijn ze er zich van bewust dat het klikgedrag gecaptureerd wordt in databronnen.

Relativiteit van de data

Om een goede datamining te kunnen uitvoeren heb je een goede grondstof nodig, wat zich digitaal vertaalt in betrouwbare, correcte geordende en geregistreerde data. Nadat het basismateriaal uitgezuiverd is worden algoritmes op de databases afgevoerd om op zoek te gaan naar statistische verbanden. Het gevaar van een algoritme is dat het de indruk wekt dat het patroon een objectief gegeven is. Maar de voorspelling is in werkelijkheid afhankelijk van het invoeren van correcte data. En laat dit nou net mensenwerk zijn. Enerzijds kunnen menselijke inschattingen, misplaatste vooringenomenheid blindelings over het hoofd worden gezien wat dan uiteindelijk uitmondt in een

slecht instrument. Slecht maar net omwille van zijn immense dataomvang nog moeilijk controleerbaar en herstelbaar. Voldoende mensen en middelen moeten vrijgemaakt worden om een open controle en toepassing van democratische spelregels op deze processen te kunnen blijven uitvoeren.

De grootste struikelblok in deze analyse zit in de begripsverwarring tussen 'een correlatie' en 'causaal verband'. Je kan vaststellen dat een bepaald profiel meer kans op werk heeft als hij zou solliciteren naar een job in een andere sector omdat er een correlatie is vastgesteld tussen dat profiel en een bepaalde beroeps categorie. Maar een causaal verband kunnen we er nooit aan vastknopen. Dit zou namelijk betekenen dat je het switchen naar andere job doelwitten op basis van die analyse resulteert in het bemachtigen van die job. We hoeven geen wetenschappelijke expert te zijn om te beseffen dat toeval, omstandigheden, motivationele, inhoudelijke en ook wel subjectieve elementen de cruciale rol spelen in het aanwervingsproces. De data analyse schiet te kort in het overbruggen van die subjectieve elementen maar draagt wel het gevaar in zich ten minste de suggestie op te waken dat er een causaal verband is tussen zich heroriënteren naar andere jobs en het aangeworven worden.

Vertrouwen

Een belangrijke bedenking met een brede maatschappelijke dekking is het gevaar dat burgers, sociaal verzekeren en dus ook werkzoekenden zich afkeren van een overheid die zich steeds dieper ingraaft

in onze **persoonsgebonden data**. Zo dreigt het gevaar dat de overheidsdiensten het vertrouwen verliezen, noodzakelijk om een goede klantenbinding met het doelpubliek te kunnen realiseren. Zoals de VDAB die met zijn determinerende zoekpatronen, het gevoel stimuleert dat de overheid via uitdijende dossiervorming, clicks en andere data steeds alwetender wordt en steeds minutieuzer het zoekgedrag onder de loep kan houden. Maar anderzijds krijgt die werkzoekenden het alsmoer moeilijker om de structuur, doelstellingen en informatiestromen van de VDAB te kunnen begrijpen. Die mentale twist vormt zo een voedingsbodemp die het vertrouwen ondermijnt. Transparantie en duidelijke meerwaarden zullen in een communicatielijnt moeten geplaatst worden om het vertrouwen van de burgers te kunnen behouden. Zoniet riskeert men processen waarbij men doelbewust dataregistratie probeert te manipuleren om te ontsnappen aan deze autonomiebeperking.

Controle

VDAB geeft aan dat de persoonlijke data van die werkzoekende in de analyse enkel als statistisch materiaal worden aangewend. Bedrijven, werknemers en werkzoekenden krijgen wel gepersonaliseerde suggesties maar zonder de bescherming van de privacy te schenden.

Maar de VDAB wil een stapje verder gaan dan enkel suggesties te lanceren die het bemiddelen faciliteert. Het zal ook een tool

worden die de mate van beschikbaarheid in kaart kan brengen. En het zal een mogelijkheid bieden om moeilijke vacatures dwingend aan te bieden bij de werkzoekenden. Als de nieuwe technologie een draagvlak verdient, zal ze er moeten voor zorgen dat niet alleen de plichten maar ook de rechten worden gevrijwaard en dat de aangereikte tools op een evenwichtige en wederkerige wijze kunnen worden ingezet voor de noden van zowel de vraag- als aanbodzijde van de arbeidsmarkt. De verleiding is dan ook groot om dergelijke digitale systemen in te zetten **ter vervanging van een menselijke tussenkomst** en dit volop in te zetten om besparingsoperaties te organiseren. We verwachten van de overheid dat de werkzoekenden kwalitatief worden opgevolgd en dat personen die niet beschikbaar zijn geresponsabiliseerd worden. Maar niet door algoritmes, wel door **motiverende gespreksvoering en empathische ondersteuning en adequate begeleiding**.

En wiens keuzevrijheid staat hier ter discussie? **Datamining versterkt de macht van het getal**. Keuzes die niet kunnen buigen op de topnoteringen via de artificiële intelligentie

zullen richting exit worden geleid.

Het lijkt het ons ook redelijk dat vacaturematching niet alleen vanuit het werkzoekendenperspectief wordt bekeken maar dat ook de data-analyse losgelaten wordt op de impact van de werkgeverszijde. Zichtbare onevenwichten in het niet weerhouden van bepaalde groepen (leeftijd, geslacht, afkomst,...) verdienen evenzeer een sturend beleid met bijvoorbeeld quota.

Win-win voor iedereen

De digitalisering is een realiteit en het graven en exploiteren van dit nieuwe goud zal nog exponentieel toenemen. De overheid moet gebruik maken van de merites van deze technologieën om wins voor haar burgers te realiseren. Maar ze moet ook maatstaven ontwikkelen zodat deze platformen betrouwbaar zijn in analyse, rechtvaardig in hun output, evenredig in gebruik en transparant in hun opzet. Pas dan zal de nieuwe technologie bij de overheid een breed maatschappelijk draagvlak verdienen waar alle burgers beter van worden.

Steven Genbrugge
 Adviseur Studiedienst Vlaams ABVV