

NEDERLAND INVESTEERT TE WEINIG IN BASIS TECHNOLOGIEËN

OP ZOEK NAAR
EEN FUNDAMENT
VOOR ONZE
DIGITALE
SAMENLEVING

DE BODEMS VAN ONZE DATASCIENCE-PIZZA'S ZIJN VRIJWEL UITSLUITEND IN HANDEN VAN AMERIKAANSE BEDRIJVEN. ZIJ HEBBEN IN HOGE MATE DE CONTROLE OVER DE ELEKTRONISCHE PLATFORMEN DIE DE BASIS VORMEN VOOR ONZE ECONOMISCHE BEDRIJVIGHEID EN SOCIALE PROCESSEN, ZEGT WIL VAN DER AALST. HET BLIJFT MERKWAARDIG DAT IT GEEN TOPSECTOR IS.

door Wil van der Aalst beeld Shutterstock

“Unieke kans:
investeren in groene datascience.”

DANKZIJ DATASCIENCE (IN DE VOLKSMOND VAAK BIG DATA) ZIJN ER IN DE AFGELOPEN JAREN VELE NIEUWE DIENSTEN ONTSTAAN. Of het nu gaat om het boeken van een reis, het opsporen van criminelen, het behandelen van kanker of het zelfstandig laten rijden van auto's: data spelen een cruciale rol in onze samenleving. Ondanks dit belang investeren we als Europa, en dus ook Nederland, onvoldoende in de basistechnologieën die dit allemaal mogelijk maken.

Het maken van een datagedreven applicatie lijkt op het maken van een pizza: het gaat niet alleen om de toppings, ook de bodem is belangrijk. In dit geval bestaat de bodem onder meer uit databases, machine-learning/data-miningtechnieken en gedistribueerde infrastructures. We gaan er te gemakkelijk vanuit dat we een bodem van elders, vaak Amerika, kunnen gebruiken. Op deze manier creëren we een 'bodemloze' digitale samenleving waar we afhankelijk zijn van anderen. Tegelijkertijd zijn er geweldige kansen, onder meer op het gebied van 'groene datascience'. We kunnen bijvoorbeeld technieken ontwikkelen die slim gebruik maken van data zonder de gebruikelijke vormen van vervuiling, zoals het verlies van privacy en oneerlijke besluitvorming. Door te investeren in groene datascience kunnen we twee vliegen in een klap slaan.

Eenzijds creëert het een kans om zelf weer datasciencebodems te ontwikkelen. Anderzijds is het verantwoord gebruik van data cruciaal voor onze samenleving.

berekeningen te distribueren over grote hoeveelheden goedkope computers. Deze voorbeelden laten zien dat het niet alleen gaat om marketing en een goed

In de top 1000 meest geciteerde onderzoekers op IT-gebied staan dertien Nederlanders

AFHANKELIJK

Wie heeft de controle over onze data, software en infrastructures? Per internet-minuut worden er meer dan 150 miljoen e-mails gestuurd, 4 miljoen YouTube-video's bekeken en 3,5 miljoen zoekopdrachten via Google verstuurd. Deze hoeveelheden blijven groeien en zijn bijna niet te bevatten. Bijna al dit verkeer wordt gegenereerd door software ontwikkeld in Amerika (zie kader). Spotify is, als Zweeds bedrijf, een uitzondering op de regel. Minder zichtbaar zijn de onderliggende technieken die deze bedrijven groot hebben gemaakt. Alphabet zou niet bestaan zonder het PageRank-algoritme, bedacht door de oprichters van Google en vernoemd naar één van hen (Larry Page). Dankzij PageRank was het mogelijk zoekresultaten effectief en efficiënt te prioriteren. MapReduce was ook essentieel voor de groei van Google. Dit raamwerk stelde Google in staat

economisch klimaat. De technologische bodem is vaak essentieel voor het succes van dit soort IT-bedrijven. De bodems van onze datascience-pizza's zijn echter vrijwel uitsluitend in handen van Amerikaanse bedrijven. Ze hebben in hoge mate de controle over de elektronische platformen die de basis vormen voor onze economische bedrijvigheid en sociale processen (zie kader).

In China is men zich bewust van dit gevaar. Daarom probeert men Chinese alternatieven te creëren. Voorbeelden zijn Alibaba en Tencent. Alibaba is een tegenhanger van Amazon en heeft in China een marktaandeel van bijna 80 procent van de e-commercemarkt. Tencent is de Chinese tegenhanger van Facebook met producten als WeChat (sociaal netwerk) en QQ (messaging). Het sociale netwerk Hyves was een opmerkelijke poging om vanuit Nederland een sociaal platform aan te bieden. Hyves en andere soortgelijke

Europese platformen spelen echter geen rol van betekenis meer. Tegelijkertijd is er een stroom van nieuwe Amerikaanse platformen zoals bijvoorbeeld Uber (personenvervoer) en Airbnb (overnachtingen) die een toenemende invloed hebben op onze maatschappij.

WETENSCHAP

Het ontstaan van veel snelgroeiende databedrijven is direct te koppelen aan wetenschappelijk onderzoek. Larry Page en Sergey Brin, de oprichters van Google, ontwikkelden bijvoorbeeld hun eerste ideeën tijdens nooit afgemaakte promotietrajecten aan Stanford. Op datasciencegebied zien we dan ook een dominantie van de Amerikaanse universiteiten. De top 10 van meest geciteerde onderzoekers op IT-gebied

zijn allen verbonden aan Amerikaanse universiteiten (zie www.guide2research.com). In de top 250 staat slechts één Nederlander (ondergetekende, op plaats 13), dus slechts 0.4 procent. In de top 1000 staan dertien Nederlanders (1.3 procent). Dit illustreert dat het niveau in Nederland goed is, maar dat er onvoldoende ruimte is voor echt toponderzoek op het gebied van datascience. De financiering van IT-onderzoek in zijn algemeenheid blijft ver achter bij andere landen (zie kader). Het blijft merkwaardig dat IT geen topsector is en altijd in een ondergeschikte rol wordt geplaatst.

Financiering IT-onderzoek blijft ver achter bij andere landen

Voor het maken van een goede bodem is de keuze van toppings vaak irrelevant. Dit wil niet zeggen dat de toepassingen (dus de toppings) niet belangrijk zijn. Nieuwe toepassingen leveren nieuwe wetenschappelijke uitdagingen op, maar de fundamentele van datascience zijn vaak niet specifiek voor een toepassing. We ontwikkelen geen databasesystemen voor agrifood, spreadsheets voor water en deep-learningtechnieken voor de creatieve industrie. Door de ontwikkeling van de fundamentele systematisch te verspreiden over vele toepassingen kunnen we nooit een leidende positie innemen. Ondanks het toegenomen belang van Big Data en datascience telt Nederland bijvoorbeeld minder databasehoogleraren dan 25 jaar geleden. Ook op het gebied van onderwijs heeft de BV Nederland een probleem. Studenten met interesse op het gebied van datascience halen hun kennis vooral uit zogenaamde Massive Open Online Courses (MOOCs), ontwikkeld door toponderzoekers van prestigieuze Amerikaanse universiteiten. Op zich is dit geen probleem. Het is goed dat

PLATFORMEN IN AMERIKAANSE HANDEN

In 2001 waren General Electric (406 miljard dollar), Microsoft (365 miljard dollar), Exxon (272 miljard), Citi (261 miljard) en Walmart (260 miljard) volgens visualcapitalist.com de vijf meest waardevolle bedrijven. Vijftien jaar later waren dat Apple (582 miljard), Alphabet (556 miljard), Microsoft (452 miljard), Amazon (364 miljard) en Facebook (359 miljard). In mei 2017 was Apple het eerste bedrijf met een marktkapitalisatie van meer dan 800 miljard dollar. De vijf leidende bedrijven zijn Amerikaans, sterk data-georiënteerd en eigenaar van een of meer platformen. Apple heeft iOS en de App store met meer dan 2.2 miljoen apps. Alphabet, het moederbedrijf van Google, heeft de controle over platformen zoals YouTube en Android. De Google Play store heeft meer dan 2.8 miljoen apps. Microsoft heeft al jarenlang de controle over het operating system (Windows) en de basis-applicaties (Office) voor personal computers. Amazon heeft weer een heel ander platform in handen: het verkoopkanaal voor elk denkbaar product. Facebook biedt weer een sociaal platform, ook van een ongelofelijke omvang. De partijen die de controle hebben over deze platformen, zijn erg machtig omdat ze bepalen welke organisaties wel en welke geen toegang hebben. Dit heeft grote gevolgen voor onze digitale samenleving. De algoritmen, interfaces en ethische keuzes van een handjevol Amerikaanse bedrijven hebben een grote invloed op onze publieke belangen, zowel op sociaal als economisch gebied.

“Neerwaartse spiraal moet doorbroken worden.”

ONDERZOEKS-FINANCIERING

De budgetten van NWO zijn beperkt en er is vrijwel geen financiering beschikbaar voor de fundamentele van onze digitale samenleving. Er zijn weliswaar initiatieven als Commit2Data en op Europees niveau de Big Data Value Association (BDVA), maar deze zijn meer gericht op de toppings van de datascience-pizza. Commit2Data is bijvoorbeeld een meerjarig nationaal onderzoeks- en innovatieprogramma op basis van publiek-private samenwerking, gericht op de negen topsectoren (Agri & Food, Chemie, Creatieve Industrie, Energie, High Tech Systemen & Materialen, Life Sciences & Health, Logistiek, Tuinbouw & Uitgangsmaterialen en Water).

studenten actief op zoek gaan naar kennis. De Nederlandse universiteiten hebben echter moeite tegenwicht te bieden. Op de lange termijn kan het niet zo zijn dat de ontwikkeling van onderwijsmateriaal uitbesteed is aan Amerikaanse universiteiten. Waarom zouden studenten dan nog hier studeren? Ook hier speelt weer het probleem van de eerder genoemde ‘datascience-pizza’. Beleidsmakers denken dat we de bodem gewoon kunnen importeren en de competitie niet aan moeten gaan. Deze neerwaartse spiraal moet doorbroken worden door meer aandacht aan datascience-onderwijs te besteden. De nieuwe Bachelor- en Master-opleidingen onder de vlag van Jheronimus Academy of Data Science (JADS) zijn hier een mooi voorbeeld van.

Ook op onderwijsgebied heeft Nederland een probleem

GROENE DATASCIENCE

Op dit moment is er een unieke kans om de basis te versterken door te investeren in ‘groene datascience’, waarbij data op een positieve manier gebruikt kunnen worden zonder ‘vervuiling’, zoals oneerlijke en niet transparante beslissingen, misleidende of zelfs

foutieve antwoorden en schendingen van privacy en vertrouwelijkheid. Hoe maken we datascience-toepassingen eerlijk, accuraat, betrouwbaar en transparant? Responsible Data Science (RSD) is een breed samengesteld consortium van meer dan 25 onderzoeksgroepen die deze fundamentele vragen willen adresseren (www.responsible-datascience.org). Denk aan nieuwe algoritmen die automatisch beslissingen nemen, die voldoen aan een vooraf ingesteld eerlijkhedscriterium, of aan analysetechnieken die accurate antwoorden geven zonder de onderliggende data te bewaren. Dit vereist nieuwe pizzabodems. Deze kunnen in de toekomst een belangrijk concurrentievoordeel opleveren. Op dit moment is er al veel wantrouwen als het gaat om Big Data en datascience. De negatieve aspecten komen vaak in het nieuws zonder dat mensen beseffen dat veel dingen goedkoper, sneller en beter lopen dankzij het slim gebruik van data. We lopen het gevaar in een ‘datascience-winter’ terecht te komen waar mensen mordicus tegen het gebruik van data zijn. De oplossing ligt niet alleen in meer en betere wetgeving. We hebben positieve technologische oplossingen nodig om data in waarde om te zetten zonder de gebruikelijke vervuiling. Op deze wijze kunnen we onze digitale samenleving voorzien van onze eigen ‘groene’ bodem. 🌱

REACTIES EN BIJDAGEN

Voor reacties en nieuwe bijdragen van IT-experts: Henk Ester 020-2356415 h.ester@agconnect.nl



AUTEUR

WIL VAN DER AALST (www.vdaalst.com) is universiteitshoogleraar aan de Technische Universiteit Eindhoven. Daarbij heeft hij deeltijdaanstellingen bij Queensland University of Technology in Brisbane, Australië en Fondazione Bruno Kessler in Trento, Italië. Hij is ook wetenschappelijk directeur van het Data Science Center Eindhoven en betrokken bij de Jheronimus Academy of Data Science. Recent is hem de prestigieuze Duitse Alexander von Humboldt-wetenschapsbeurs toegekend, de meest waardevolle Duitse wetenschapsprijs. Hij ontvangt vijf miljoen euro om onderzoek te doen aan de Rheinisch-Westfälische Technische Hochschule (RWTH) in Aken.