



DATA SCIENCE VOORKOMT VERDRINKEN IN DATA



Dhr. Arjan de Heer, Programmamanager
DataLab bij JIVC

JIVC DATA LAB



DATA SCIENCE FOR OUR DEFENCE

De minister en de CDS waren tijdens de presentatie van de DefensieVisie 2035 meer dan duidelijk: Defensie moet een hoogtechnologische organisatie worden die informatiegestuurd opereert, zelfstandig en schaalbaar kan optreden en flexibel kan reageren op wisselende omstandigheden, zowel op nationaal als internationaal niveau. Deze toekomstvisie en dit nieuwe 'profiel' vragen niet om data science, maar éisen het. Het is dan ook niet voor niets dat aan het DataLab JIVC, hét IT-bedrijf van Defensie onlangs structureel financiële middelen zijn toegekend waarmee het kennis kan opbouwen, borgen en beschikbaar stellen aan alle onderdelen van de krijgsmacht. →



Data science is een term die door sommige mensen nog met de nodige scepsis wordt bekeken. Een buzzwoord? Waait wel weer over? Een hype? Zeker niet.

Data science is letterlijk een voorwaarde voor informatiegestuurd optreden (IGO) zoals beschreven staat in de Defensie-Visie 2035. Onze CDS gaf het tijdens de persconferentie ook treffend aan: “Waar we vroeger te weinig informatie hadden, hebben we nu feitelijk overschotten en moeten we oppassen hierin niet te verdrinken.” Hiermee raakt hij de kern van het probleem dat wij in het DataLab aanpakken.

Van data naar informatie

Om het geschetste probleem in zijn volle omvang te begrijpen eerst een korte terugblik in de tijd. Al langer waren technici en wetenschappers zich bewust van de macht van data. Hoe meer data – losse getallen – hoe meer informatie. Om tot het laatste te komen is het echter nodig om de data samen te voegen en te interpreteren. In sommige gevallen is dit eenvoudig en leert een reeks geluids- of trillingsopnames dat een vijandelijk voertuig dichterbij komt. Met een beetje extra moeite is ook op te maken of hij recht op je afkomt of verderop voorbijrijdt. Of je ziet vanuit twee getallenreeksen hoe een oplopend toerental van een motor gepaard gaat met een steeds grotere warmteontwikkeling.

Sneller en goedkoper

Nog interessanter wordt het wanneer veel meer data afkomstig van verschillende bronnen kunnen worden gecombineerd en gebruikt om informatie te maken. Hier schiet het menselijk brein al snel tekort. Trillingen die iedere 10 microseconden worden gemeten, de informatie van vijftig drones die gelijktijdig binnenkomt, de realtime locatie, beschikbaarheid en conditie van tienduizenden laarzen, helmen en trackers, het combineren van locatiedata, weersomstandigheden, radarbeelden en het oorlogsrecht...

Mensen kunnen dat met hun hersenen al snel niet meer behappen. Niet alleen de grote hoeveelheid data (big data) maar zeker niet de snelheid waarmee de data beschikbaar komt. Maar dat willen we wel: uit big data precies de informatie halen die relevant is, mogelijke verbanden herkennen en deze gebruiken om toekomstige situaties te voorspellen. Hiervoor is reken capaciteit nodig die deze big data kan herkennen, waarden toe kennen, categoriseren, combineren, uitsluiten, analyseren en nog veel meer.

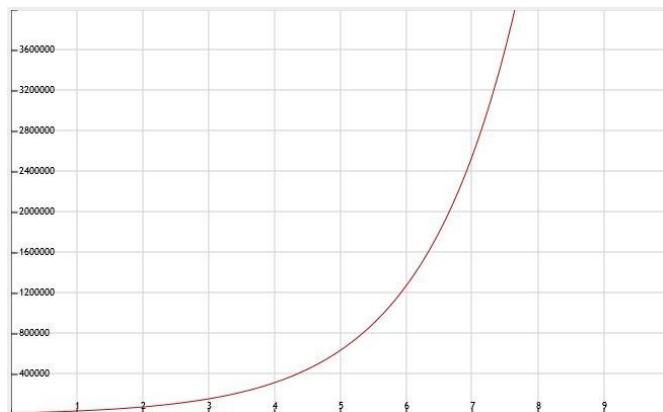
En precies dát is wat er de afgelopen jaren – conform de wet van Moore – beschikbaar is gekomen: Een dermate grote reken capaciteit dat computers in staat zijn om véél data (1), gelijktijdig (2) te kunnen verwerken tot informatie. Dit is ook terug te zien in consumentenproducten zoals smartphones en tablets die steeds kleiner en dunner worden terwijl de functionaliteit alleen maar toeneemt. Of aan een flight simulator die in de nieuwste versie bijna niet meer van echt is te onderscheiden door de grote reken capaciteit in combinatie

met de huidige generatie grafische kaarten.

Parallel aan de ontwikkeling van reken capaciteit, zijn de ontwikkelingen op het vlak van sensoren en datacommunicatie minimaal zo snel gegaan. Dit betekent dat de data inmiddels veel goedkoper zijn te genereren dan vroeger (door goedkopere sensoren) en bovendien snel en betrouwbaar in grote hoeveelheden (bandbreedte) zijn te versturen, het liefst draadloos. Hetzij beveiligd, hetzij via standaard open communicatieprotocollen.

Exponentiële groei

Het is interessant om te vermelden dat deze ontwikkelingen onder de zogenaamde ‘exponentiële groei’ vallen waarvan hieronder een bijbehorende kromme te zien is. Kenmerkend voor exponentiële groei is de bijzonder geringe impact aan het begin, daarna een omslagpunt waarbij veranderingen wel degelijk merkbaar zijn en vervolgens het traject waarbij de grafiek steeds steiler gaat lopen. Dit is het deel van de kromme waarin ongeloof optreedt bij mensen: zoveel, zo lang, zo vaak? Dit is op dit moment ook het geval bij de hoeveelheid data die we op dit moment gelijktijdig kunnen genereren, versturen en verwerken. Maar dan? De belangrijkste beperking om deze



data voor nuttige doeleinden te gebruiken, ligt op dit moment in – wederom – de beperkingen van de mens. De creatieve geest van de mens die kan bepalen waar de behoeften liggen en waar de successen zijn te boeken door data te verzamelen en te verwerken tot informatie; de techniek in hardware is in principe volledig beschikbaar. De software (rekenalgoritmes) is per toepassing te schrijven.

Toepassingen Defensie

Tot zover de geschiedenis en de theorie. Maar waar begint nu precies de data science en wat gaat Defensie daarmee doen? Om met de laatste vraag te beginnen: De belangrijkste toepassingen voor Defensie liggen bij het informatiegestuurd optreden (IGO), op het logistieke vlak en in het onderhoud. Eigenlijk kennen we allemaal de kleinere voorbeelden wel. De marechaussee die software inzet om via gezichtsherkenning mogelijke criminelen in grote mensenmassa's te her-



Met behulp van algoritmes en tekstherkenning ondersteunt het DataLab de KMar in de inzet van algoritmes om automatisch kruisverbanden in proces-verbalen te leggen.

kennen. Planningssoftware waarmee voertuigen binnen Defensie op het juiste tijdstip op de juiste locatie beschikbaar zijn. Sensoren die in hydraulische systemen van marineschepen de conditie van de olie monitoren waardoor in een vroeg stadium aankomende problemen zijn te tackelen vóórdat zij tot storingen of uitval leiden.

Voor al deze toepassingen is het noodzakelijk dat een specifiek algoritme de data – feitelijk de losse getallen – omzet naar relevante informatie voor een specifieke persoon onder specifieke omstandigheden. Wanneer het gaat om het combineren van data uit verschillende bronnen die middels algoritmes worden geanalyseerd, geïnterpreteerd en omgezet naar informatie, spreken we van data science. Vrij vertaald datawetenschap en per definitie een interdisciplinair onderzoeksveld waarin wetenschappelijke methoden worden gebruikt om kennis en informatie uit data te onttrekken.

Voor bepaalde applicaties bestaan de algoritmes al. Zeker in het kader van logistiek en onderhoud zijn er binnen de industrie al ontelbaar veel voorbeelden waarbij data wordt gebruikt om

onderhoud uiteindelijk zo betrouwbaar mogelijk te voorspellen. Maar ook met betrekking tot complexere zaken zijn er in de civiele wereld al mooie oplossingen ontwikkeld. Zo zijn er in studentenstad Eindhoven proeven gedaan met een systeem dat is geïnstalleerd in een uitgaansgelegenheid en op basis van datum (tentamens geweest of vakantie), weersomstandigheden, tijdstip van de dag, lichtniveau en geluid voorspellingen doet over de kans dat er op straat 'relletjes' ontstaan.

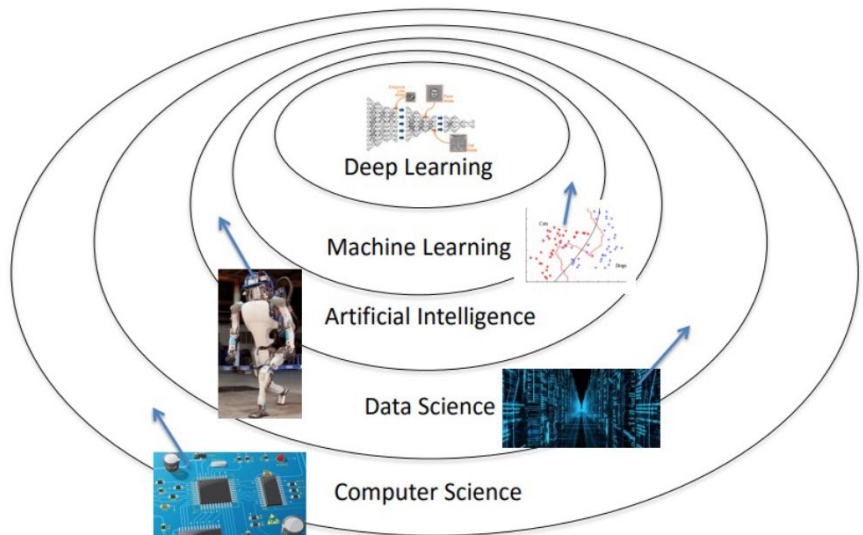
DataLab

Peter Güldenpfennig, Jr Innovatie-manager: “Bij Defensie willen we echter meer, en zeker in het kader van informatiegestuurd optreden waar we gegevens willen verzamelen die door een mens niet middels onze zintuigen zijn te verzamelen en zeker niet in korte tijd te verwerken. Een domein dat binnen de industrie niet bekend is en waar we dus onze eigen oplossingen moeten ontwikkelen en testen. Speciaal hiervoor is een aantal jaar geleden het DataLab opgezet. DataLab valt binnen The Next Step onder het thema ‘Informatiegestuurd optreden’ en is onderdeel van JIVC. Hier doen we onder andere onderzoek naar Data Science en Artificial Intelligence. Zoals in onderstaande figuur te zien is dat eigenlijk nog het voorstadium van wat op dit moment het hoogst haalbare is: Deep learning. Wanneer je dáár bent aangeland zit je op een niveau waar technische systemen – zoals robots en autonome systemen – (bijna) in staat zijn om op het cognitieve niveau van mensen te functioneren. Zo ver zijn we dus nog (lang) niet, maar door ons nu te focussen op data science, zijn we die weg wel ingeslagen.”

Logische start

Het DataLab komt overigens niet uit de lucht vallen. Drie jaar geleden hebben

Van CS naar DL





Datascience in RAS project - op weg naar IGO

RAS staat voor 'Robots and Autonome Systems' en is een project waarin wordt onderzocht in hoeverre dit type systemen een bijdrage kan leveren aan de taken van de krijgsmacht. In een eerste fase is onderzoek gedaan naar de bruikbaarheid van bestaande oplossingen, eventueel met kleine aanpassingen. Deze worden vervolgens onder praktijkomstandigheden getest in hoofdzakelijk Oirschot.

Informatie vergaren

In een eerste stap richting serieuze data science worden de autonome systemen ingezet om data te verzamelen en door te geven aan de commandant. Hiermee wordt zijn letterlijke en figuurlijke blikveld breder omdat autonome systemen op locaties kunnen komen die voor mensen mogelijk te gevaarlijk zijn of zeer moeilijk bereikbaar of omdat de systemen apparatuur mee kunnen nemen die voor mensen te zwaar is.

Projectleider Ikol Martijn Hädicke: "Het ontsluiten van databronnen met behulp van robots is nog relatief eenvoudig en levert vooral een bijdrage in de hoeveelheid informatie die per tijdseenheid beschikbaar komt. Daarbij zitten we nu nog in het stadium dat één mens één autonoom systeem of robot bedient en vervolgens de opgehaalde data bekijkt, interpreteert en gebruikt. Met behulp van data science is het de bedoeling om uiteindelijk met één man bijvoorbeeld vijftig drones aan te sturen. Dit heeft pas zin wanneer de grote hoeveelheid informatie die hiermee in één keer beschikbaar komt, snel en betrouwbaar kan worden beoordeeld door specifieke software. Afkomstig van één bron spreken we eerst nog van data-analyse. Het combineren van data afkomstig van meer bronnen data science. Dit laatste is noodzakelijk om de commandant de informatie te geven die hij nodig heeft."

Adaptatie

In hetzelfde project zijn onlangs verrassende resultaten geboekt ten aanzien van de adaptatie van het werken met deze nieuwe systemen. In hoeverre is een mens



immers bereid om te 'luisteren' naar wat een algoritme hem aangeeft of adviseert?

Hädicke: "Om nieuwe technieken te laten werken, is het in alle gevallen belangrijk om de eindgebruiker in het ontwikkelproces mee te nemen. In een test hebben we een aantal drones beschikbaar gesteld aan jonge pelotonscommandanten met een specifieke opdracht en vervolgens geobserveerd hoe zij omgaan met deze nieuwe mogelijkheden."

In eerste instantie laten ze mogelijkheden vrijwel onbenut door gebrek aan ervaring, kennis en tijd. Zij doen eerst op de traditionele manier wat hen gevraagd is. In dit geval het verzamelen van informatie voor hun commandant. Na een nacht – niet slapen – worden de jonge militairen geconfronteerd met het feit dat zij voor 8.00 's morgens zelf niet op pad mogen. Op dat moment wordt de meerwaarde van de drones ontdekt – zij mogen namelijk wél op pad – en wordt een creatieve ader aangeboord. Daarbij ontdekken ze de mogelijkheden van de drones en het feit dat deze controleerbaar en voorspelbaar zijn. Hädicke: "In feite doen ze hetzelfde – hun opdracht uitvoeren – maar nu met andere middelen waarmee ze flexibeler zijn, een groter effect kunnen creëren en bovendien hun mensen veilig kunnen houden. Deze relatief snelle adaptatie is wat dat betreft een verrassend positieve uitkomst en biedt perspectief voor de mensen die oplossingen ontwikkelen gebaseerd op data science."



De veelvoud aan systemen binnen Defensie leveren data op. Hieruit informatie destilleren is een van de doelen van de verschillende data-initiatieven binnen Defensie.

verschillende 'high potentials' al gewerkt aan oplossingen voor enkele tientallen use cases waarbij gebruik is gemaakt van data-analyse en data science. Op basis van de goede resultaten meende het JIVC dat het tijd werd om het onderwerp data science gestructureerd aan te pakken; de start van het DataLab. Inmiddels werken we hier met zo'n tien mensen die onderzoek doen naar de techniek en hiermee kennis opbouwen. Daarnaast ligt een belangrijke taak in het borgen van deze kennis en om onderzoek te doen naar aanverwante thema's zoals ethiek, cyber security, privacy, veiligheid...

Je werkt hier immers wel met data science binnen de militaire context waarbij je áltijd te maken hebt met de veiligheid van je eigen mensen én die van burgers. Het wordt niet voor niets gedefinieerd als een interdisciplinair onderzoeksveld.

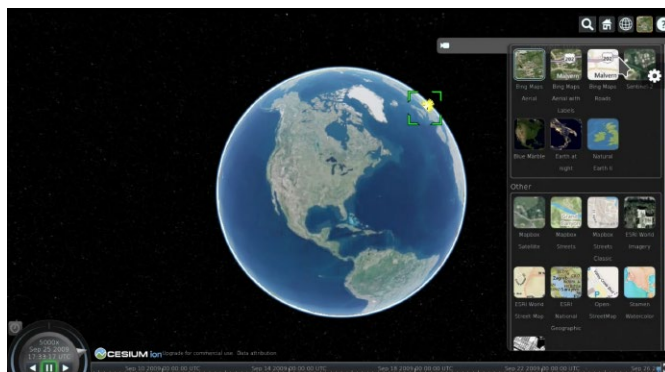
Last but not least zijn we defensiebreed dé vraagbaak voor alle vraagstukken die mogelijk met data science zijn aan te pakken. Of het nu gaat om personeel, materieel of videoanalyse. Hierop gaan we overigens niet zitten wachten. Joerie Troost, adviseur en data scientist binnen Defensie vult aan: "We hebben de afgelopen tijd veel bezoeken aan uiteenlopende locaties gebracht om te praten over de mogelijkheden voor een specifieke eenheid. We willen daarbij het liefst dat mensen zélf met vraagstukken komen zodat er ook draagvlak is om er – in overleg met opco's – aan te werken. Daarbij gaan we in eerste instantie uit van data die al beschikbaar zijn; hiermee is op dit moment al grote stappen te zetten. Belangrijk is dan wel dat eenheden zich bewust zijn van de spreekwoordelijke stip op de horizon: wáár wil men naartoe, welke rol kan data science hierin betekenen en hoe kan het DataLab daarin een aandeel leveren."

En nu dóórpakken

In de afgelopen drie jaar heeft het DataLab niet alleen al goed werk geleverd, maar is ook het besef gegroeid dat IGO de toekomst is. In dat kader meldde de CDS in de persconferentie waarin de DefensieVisie werd gepresenteerd: "We zullen in

staat moeten zijn om sneller gedifferentieerde informatie naar het krijgsmachtonderdeel te sturen die het nodig heeft. Dit om ervoor te zorgen dat de commandant ter plekke voor de komende dertig minuten over de inlichtingen beschikt die hij of zij nodig heeft. Enerzijds om sneller beslissingen te kunnen nemen, anderzijds om eerder de juiste beslissingen te nemen. Om dit proces in te richten gaat het over IT, opslaan, zoeken en technologie."

Peter Güldenpfennig: "Ik denk dat hiermee de relevantie van onderzoek naar en de toepassing van data science binnen Defensie is aangetoond en dat het terecht is dat er nu structureel geld beschikbaar komt om het DataLab op te schalen van nu tien man naar mogelijk tot vijftig medewerkers. Hiermee kunnen we aanzienlijk meer onderzoek doen, onze taken uitbreiden met de beveiliging van data en meer projecten opstarten om ideeën uit te werken en te valideren. Mits we uiteraard kunnen beschikken over de beste mensen. We werken daarbij overigens ook samen binnen de gouden driehoek met onder meer TNO en het bedrijfsleven.



Een initiatief binnen het DataLab is The GeoStack Project, een open source-project waarbij je geodata op kaarten kunt plotten.

Een focus ligt daarbij op de eerder gedefinieerde aandachtsgebieden voor Defensie: IGO, onderhoud en logistiek. Vooral IGO is complex omdat je hier met veel databronnen te maken hebt. Data afkomstig van bijvoorbeeld je eigen materieel (schip, vliegtuig, auto), van autonome systemen en robots die gegevens verzamelen maar ook van GIS, weersvoorspellingen, medische data uit de kleding van onze militairen, onderschepte berichten van de tegenstander...

Om deze data gelijktijdig om te zetten naar relevante informatie voor de commandant ter plaatse vergt nog wel wat maar biedt ook ongekende mogelijkheden. Kortom: een vakgebied dat nog in de kinderschoenen staat en waarbinnen we grote ontwikkelingen kunnen verwachten.

Om die reden is het belangrijk om de opgebouwde kennis van het DataLab te kunnen borgen en hiermee een solide basis op te bouwen voor toekomstig gebruik." 