TIJD VOOR IMPLEMENTATIE VAN VERANTWOORDE DATADIENSTEN

DE TRIAS ANALYTICA VOOR RESPONSIBLE DATA SCIENCE

TNO innovation for life

WHITEPAPER
September 2018

Thymen Wabeke
Victor Klos
Tjerk Timan

MANAGEMENTSAMENVATTING

Organisaties ontwikkelen in toenemende mate datadiensten die gebruik maken van zelflerende algoritmes. Deze algoritmes zijn veelbelovend omdat ze automatisch regels en patronen extraheren uit data waarmee beslissingen of voorspellingen gemaakt worden. Het gebruik van zelflerende algoritmes brengt tevens nieuwe uitdagingen met zich mee. Dit whitepaper introduceert deze uitdagingen en beschrijft hoe organisaties ze kunnen adresseren. We richten ons daarbij op drie invalshoeken, namelijk regulering omtrent data, randvoorwaarden voor het toepassen van zelflerende algoritmes en technieken om gevoelige data te analyseren. Het adresseren van deze uitdagingen draagt bij aan een verantwoorde ontwikkeling en inzet van datadiensten. Data en datadiensten gaan hand in hand. Het verkrijgen en verwerken van gegevens is aan regels gebonden, met name waar het persoonsgegevens betreft. Dit whitepaper benoemt de belangrijkste begrippen uit de Algemene Verordening Gegevensbescherming (AVG) die sinds 25 mei 2018 van kracht is. Transparantie van het analyseproces en de uitlegbaarheid van geautomatiseerde beslissingen zijn met name relevant voor zelflerende algoritmes.

Vervolgens richten we ons op het vakgebied machine learning. Dit vakgebied houdt zich bezig met de ontwikkeling en toepassing van zelflerende algoritmes. De regels en patronen die deze algoritmes extraheren zijn nooit perfect. En er zijn uitdagingen die organisaties moeten adresseren wanneer ze zelflerende algoritmes verantwoord willen toepassen (zoals sampling bias en overfitting). Deze uitdagingen – en eventuele consequenties – zijn niet alleen relevant voor de betreffende data scientists, maar voor de hele organisatie.

De gegevens die datadiensten verwerken zijn vaak (privacy)gevoelig. Het derde deel van dit whitepaper richt zich op het analyseren van gevoelige data. Er bestaan technieken waarmee organisaties hun databronnen kunnen anonimiseren. Daarnaast is het soms mogelijk om een datadienst anders te ontwerpen zodat meerdere organisaties gezamenlijk een analyse kunnen uitvoeren zonder gevoelige gegevens te hoeven delen.

Wat ons betreft is de tijd rijp om te innoveren met de technieken die in het whitepaper aan bod komen door deze toe te passen. TNO wil graag de volgende stap zetten en bijdragen aan de ontwikkeling en implementatie van verantwoorde datadiensten. Daarvoor gaan we graag de samenwerking met u aan.

INHOUDSOPGAVE

INTRODUCTIE

Organisaties ontwikkelen in toenemende mate datadiensten die gebruik maken van zelflerende algoritmes. Deze algoritmes zijn veelbelovend omdat ze automatisch regels en patronen extraheren uit data waarmee beslissingen of voorspellingen gemaakt worden. Het gebruik van zelflerende algoritmes brengt tevens nieuwe uitdagingen met zich mee. In dit whitepaper introduceren we deze uitdagingen en beschrijven we hoe organisaties ze kunnen adresseren.

Hoe kunnen organisaties op een verantwoorde manier datadiensten ontwikkelen en inzetten?

Voorafgaand aan het schrijven van dit document heeft een aantal interviews plaats gevonden met stakeholders van het Nationaal Cyber Security Centrum (NCSC)[1]. Het doel van deze interviews was om inzicht te krijgen in de uitdagingen die organisaties tegenkomen tijdens big data projecten. Naar aanleiding van deze interviews is besloten om bovenstaande hoofdvraag te benaderen vanuit drie invalshoeken. Ten eerste, wat is de regulering omtrent de verwerking van persoonsgegevens. Ten tweede, hoe werken de methodes achter zelflerende algoritmes en wat zijn randvoorwaarden voor een verantwoorde toepassing. Ten derde, welke technieken helpen bij het anonimiseren van persoonsgegevens en hoe kunnen partijen gezamenlijk datadiensten ontwikkelen zonder gevoelige informatie te hoeven delen.

AFBAKENING

Het whitepaper richt zich op drie invalshoeken:
(1) regulering omtrent data,
(2) randvoorwaarden voor het toepassen van zelflerende algoritmes en
(3) technieken om gevoelige data te analyseren.

Deze invalshoeken zijn gekozen naar aanleiding van de organisaties die we hebben gesproken. Uiteraard zijn er ook andere onderwerpen te bedenken die relevant kunnen zijn voor de ontwikkeling van een verantwoorde datadienst, bijvoorbeeld *interoperabiliteit* en *gebruikersvriendelijkheid*. Deze onderwerpen vallen echter buiten de scope van dit whitepaper.

Het whitepaper beschrijft de relevante kernbegrippen uit de *Algemene Verordening Gegevensbescherming* (AVG), maar is geen juridisch document of AVG-handleiding. Geïnteresseerde lezers verwijzen we graag naar de handleidingen van de Autoriteit Persoonsgegevens en de Norwegian Data Protection Authority.[2]

1 Het NCSC is een informatieknooppunt en expertisecentrum voor cybersecurity. De primaire doelgroep van het NCSC is de rijksoverheid en de vitale infrastructuur.
2 Zie de documenten https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/avg_in_een_ notendop.pdf en https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf.

LEESWIJZER

Het whitepaper adresseert de uitdagingen die zich vaak voordoen bij de ontwikkeling van datadiensten vanuit drie invalshoeken. Iedere invalshoek heeft zijn eigen hoofdstuk. De hoofdstukken vormen één geheel, maar kunnen ook los van elkaar gelezen worden.

Hoofdstuk ‘Regulering in het data-landschap’ richt zich op de regulering omtrent datadiensten. We benoemen de belangrijkste begrippen uit de AVG. Daarnaast besteden we aandacht aan juridische aspecten die van belang zijn voor zelflerende algoritmes.

Data-analyses worden relevant zodra data voor dit doel verwerkt mag worden. Hoofdstuk ‘Leren van data’ begint met een introductie van het vakgebied *machine learning*. Dit vakgebied houdt zich bezig met de ontwikkeling en toepassing van zelflerende algoritmes om inzichten te verkrijgen uit data. Vervolgens richten we ons op uitdagingen die organisaties moeten adresseren wanneer ze zelflerende algoritmes verantwoord willen toepassen.

In veel gevallen zijn de gegevens die een datadienst analyseert (privacy)gevoelig. Hoofdstuk ‘Gevoelige gegevens analyseren’ richt zich op dit onderwerp. Het eerste deel van dit hoofdstuk beschrijft technieken om een databron te ontdoen van gevoelige gegevens. Daarna introduceren we technieken waarmee meerdere organisaties gezamenlijk een datadienst kunnen ontwikkelen zonder gevoelige gegevens te hoeven uitwisselen. Ten slotte beschrijven we twee raamwerken om de gevoeligheid van een databron te evalueren.

⟩ REGULERING IN HET DATA-LANDSCHAP

Data en datadiensten gaan hand in hand. Het verkrijgen en verwerken van data is aan regels gebonden, met name waar het persoonsgegevens betreft. Een van de uitgangspunten van de Europese dataregulering is dat data makkelijk kan worden uitgewisseld of verhandeld binnen Europa, met als doel om de data-economie te bevorderen en administratieve processen te vergemakkelijken (European Commission 2018). Tegelijkertijd zet Europa stevig in op het beschermen van persoonsgegevens. Dit blijkt ook uit de *Algemene Verordening Gegevensbescherming* (AVG) die onlangs van kracht is geworden (European Commission 2016).

VERWERKEN VAN PERSOONSGEGEVENS

De AVG heeft alleen betrekking op de verwerking van persoonsgegevens. Het is dus allereerst relevant om te weten wat de wetgever verstaat onder de begrippen *verwerking* en *persoonsgegevens* (uit Artikel 4 van de AVG):

1 ‘persoonsgegevens’: alle informatie over een geïdentificeerde of identificeerbare natuurlijke persoon (‘de betrokkene’); als identificeerbaar wordt beschouwd een natuurlijke persoon die direct of indirect kan worden geïdentificeerd, met name aan de hand van een identificator zoals een naam, een identificatienummer, locatiegegevens, een online identificator of van een of meer elementen die kenmerkend zijn voor de fysieke, fysiologische, genetische, psychische, economische, culturele of sociale identiteit van die natuurlijke persoon.

2 ‘verwerking’: een bewerking of een geheel van bewerkingen met betrekking tot persoonsgegevens of een geheel van persoonsgegevens, al dan niet uitgevoerd via geautomatiseerde procedés, zoals het verzamelen, vastleggen, ordenen, structureren, opslaan, bijwerken of wijzigen, opvragen, raadplegen, gebruiken, verstrekken door middel van doorzending, verspreiden of op andere wijze ter beschikking stellen, aligneren of combineren, afschermen, wissen of vernietigen van gegevens.

We kunnen hieruit opmaken dat de definities van persoonsgegevens en verwerking erg breed zijn. Datadiensten hebben dus al snel te maken met de verwerking van persoonsgegevens. In sommige gevallen is het mogelijk om data te *anonimiseren*. Dit betekent dat individuen niet meer geïdentificeerd kunnen worden, de data niet langer persoonsgegevens zijn en de AVG dus niet meer van toepassing is.
In hoofdstuk ‘Gevoelige gegevens analyseren’ gaan we verder in op anonimiseringstechnieken.

Een database met daarin persoonsgegevens hoeft geen probleem te zijn. Er zijn duidelijke voorwaarden om persoonsgegevens te mogen verwerken. Het begrip *doelbinding* is hierbij met name van belang. Doelbinding betreft het principe dat dataverzameling plaatsvindt met een specifiek, expliciet en legaal doel en dat data niet verder gebruikt mag worden voor andere doeleinden. Aan deze doelbinding hangt een aantal randvoorwaarden, zoals *noodzakelijkheid* en *proportionaliteit*. Noodzakelijkheid richt zich op de vraag of de dataverwerking noodzakelijk is om het gestelde doel te bereiken. Proportionaliteit gaat over de vraag of de data die verwerkt wordt echt allemaal nodig is om het doel te bereiken.

In veel gevallen – vooral in de publieke sector – vormt het verwerken van persoonsgegevens geen probleem, omdat het doel legitiem is en vaak ook proportioneel. De verwerking valt dan al binnen de doelbinding van een organisatie. Wanneer dit niet het geval is kunnen organisaties een individu om een geïnformeerde toestemming vragen. *Geïnformeerde toestemming* betekent het op de hoogte brengen en het geven van een keuze aan gegevenssubjecten om hun persoonsgegevens voor een bepaald doel te gebruiken. Het is hierbij overigens van belang dat de aangeboden datadienst gelijk moet blijven na het al dan niet verlenen van toestemming.

PRIVACY BY DESIGN AND DEFAULT

De AVG bepaalt niet alleen óf data verwerkt mag worden maar stelt ook voorwaarden aan hóe data verwerkt moet worden. Zo gaat de wetgever uit van het principe *'privacy by design and default'*. Dit principe stelt ten eerste dat waar mogelijk privacy-verhogende maatregelen gebruikt moeten worden. Daarnaast verlangt de wetgever dat datadiensten niet meer persoonsgegevens verwerken dan noodzakelijk is. Dit laatste begrip wordt ook wel *dataminimalisatie* genoemd. In 'Het Blauwe Boekje' wordt het principe van *privacy by design and default* vertaald naar acht ontwerpstrategieën die ontwerpers en bouwers van datadiensten kunnen hanteren (Hoepman 2018).

Met betrekking tot de manier waarop zelflerende algoritmes gegevens verwerken zijn twee andere onderwerpen ook belangrijk, namelijk *transparantie* en de *uitlegbaarheid* van algoritmische besluitvorming.

AVG bepaalt niet alleen óf data verwerkt mag worden maar stelt ook voorwaarden aan hóe data verwerkt moet worden

TRANSPARANTIE EN UITLEGBAARHEID

De AVG hamert erg op transparantie. Hiermee bedoelt de wetgever dat een organisatie moet kunnen uitleggen welke persoonsgegevens ze verwerkt en met welk doel. Daarnaast bevat de AVG een artikel over *algoritmische besluitvorming* (artikel 22). Dit artikel richt zich specifiek op zelflerende algoritmes die automatisch beslissingen nemen. Bijvoorbeeld de beslissing om een klant een bepaalde hypotheekrente aan te bieden.

De wetgever redeneert vanuit het principe dat een individu recht heeft op een logische uitleg van iedere algoritmische beslissing. Er is nog veel onduidelijkheid over hoe dit principe zal worden gehandhaafd, aangezien er in de dagelijkse praktijk erg veel automatische beslissingen plaatsvinden zonder dat we ons afvragen hoe dit precies werkt.[3] Welke beslissingen moeten organisaties kunnen uitleggen? En hoe gedetailleerd moet een uitleg zijn?

3 De blogpost "Is there a 'right to explanation' for machine learning in the GDPR?" gaat dieper in op de onduidelijkheid rondom de uitlegbaarheid van algoritmische besluitvorming. https://iapp.org/news/a/ is- there- a- right- to- explanation- for- machine- learning- in- the- gdpr/

Er zijn verschillende niveaus waarop algoritmische beslissingen uitgelegd kunnen worden, bijvoorbeeld op organisatorisch of beleidsmatig niveau. Hierbij gaat het bijvoorbeeld om een uitleg over de manier waarop een organisatie de uitkomsten van zelflerende algoritmes gebruikt en wie eindverantwoordelijk is. Het is ook mogelijk om uitleg te geven over de (technische) opzet van een datadienst. Hierbij gaat het om het beschrijven van de stappen en algoritmes die gebruikt worden om tot een beslissing te komen. Ten slotte is het mogelijk om de totstandkoming van individuele beslissingen te verklaren. Hierbij wordt bijvoorbeeld uitgelegd hoe de combinatie van bepaalde eigenschappen van een individu en de regels van een algoritme resulteren in een bepaalde beslissing. Het uitleggen van uitkomsten op dit niveau is een grote onderzoeksuitdaging en wordt ook wel *explainable artificial intelligence* (XAI) genoemd.

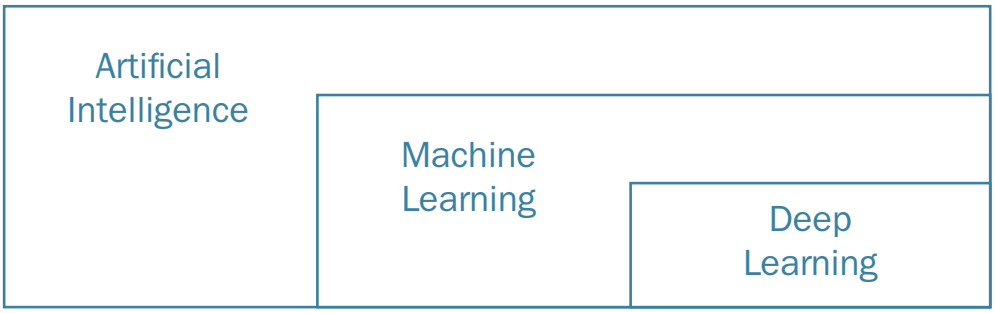
Zoals gezegd is de AVG niet heel duidelijk in wat wordt verlangd op het gebied van uitlegbaarheid. De geest van de wet is echter wel helder: het gaat erom dat organisaties ethisch handelen. Zaken als discriminatie en ongewenste profiling dienen bijvoorbeeld voorkomen te worden. In de toekomst zullen audits van zelflerende algoritmes daarom steeds belangrijker worden (Casey, Farhangi, and Vogl 2018). Organisaties kunnen zich hierop voorbereiden door grip te krijgen op hun analyseprocessen en te begrijpen hoe algoritmische beslissingen tot stand komen.

Het in kaart brengen van alle data- en analyseprocessen binnen een organisatie is een goede eerste stap. Een *Data Protection Impact Assessment* (DPIA) kan hierbij helpen, omdat dit instrument inzicht geeft in de risico's die de verwerking van persoonsgegevens opleveren en welke maatregelen moeten worden genomen om deze risico's af te dekken. Daarnaast is het belangrijk dat verschillende mensen in de organisaties begrijpen hoe zelflerende algoritmes werken en wat de randvoorwaarden zijn. Op deze manier kunnen organisaties gezamenlijk en weloverwogen bepalen hoe deze algoritmes worden ingezet. In het volgende hoofdstuk nemen we zelf het voortouw door ons te richten op de kracht en uitdagingen van de methodes achter zelflerende algoritmes.

Organisaties kunnen zich voorbereiden door grip te krijgen op hun analyseprocessen en te begrijpen hoe algoritmische beslissingen tot stand komen

LEREN VAN DATA

Datadiensten maken in toenemende mate gebruik van zelflerende algoritmes om inzichten te verkrijgen uit data. Deze algoritmes zijn veelbelovend omdat ze geautomatiseerd regels en patronen extraheren uit data. Om te kunnen redeneren over de opbrengsten en uitdagingen van zelflerende algoritmes is het handig om te begrijpen hoe de onderliggende methodes ongeveer werken. Omdat er veel termen geïntroduceerd worden beginnen we met het helder uitleggen van onze scope (zie ook figuur 1).



FIGUUR 1: ARTIFICIAL INTELLIGENCE OMVAT DE VAKGEBIEDEN MACHINE LEARNING EN DEEP LEARNING.

Het veld van *artificial intelligence* omvat de zoektocht naar intelligente en bewuste computers. Daar zijn we nog ver van verwijderd, en volgens velen zal generieke artificial intelligence er nooit komen. Wel worden er goede praktische resultaten behaald met *machine learning*. Dit is 'computers laten leren zonder expliciet programmeren', vrij naar Arthur Samuel (Samuel 1959). Een specifieke vorm daarvan is *deep learning*, dat de laatste jaren sterk in opkomst is. De intuïtie achter deep learning is geïnspireerd op de werking van de hersenen.

Dit hoofdstuk begint met een laagdrempelige introductie van het veld *machine learning*, informeel en zonder wiskunde. Voor de geïnteresseerde lezer is er een lijst opgenomen met referenties naar aanvullend materiaal.[4] Vervolgens beschrijven we een aantal uitdagingen die organisaties moeten adresseren wanneer ze machine learning verantwoord willen toepassen. Deze uitdagingen – en eventuele consequenties – zijn niet alleen relevant voor de betreffende data scientists, maar voor de hele organisatie.

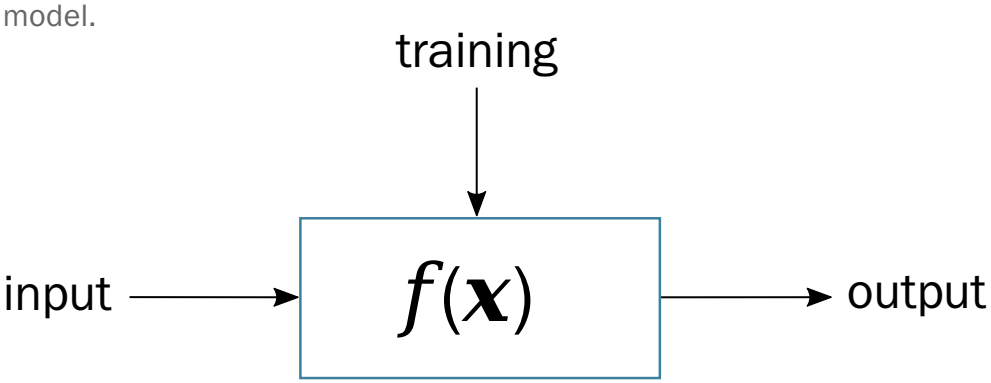
MACHINE LEARNING

Computerprogramma's worden regelmatig gebruikt om beslissingen of voorspellingen te maken. Om dit te kunnen doen bevatten deze programma's functies die een invoer vertaalt naar een uitvoer. Een voorbeeld maakt veel duidelijk. Stel een bank automatiseert het doen van hypotheekaanbiedingen aan klanten. De invoer van het programma bestaat uit gegevens over een huis (zoals vloeroppervlakte, aantal kamers enz.), de taxatiewaarde, inkomensgegevens van de klant en de huidige rentestand. De te betalen rente is een mogelijke uitvoer.

4 Wie meer wil weten na het lezen van dit hoofdstuk verwijzen wij graag naar één van de volgende de boeken: 'Artificial Intelligence: a Modern Approach' (Russell and Norvig 2013), 'Learning from Data' (Abu- Mostafa 2012) en 'Deep Learning: a Practitioner's Approach' (Patterson 2017).

Nu is dit een redelijk voor de hand liggend voorbeeld: veel is bekend en voorspelbaar in de wereld van de financiële dienstverlening. Soms is het niet zo eenvoudig om een dergelijke functie te programmeren. Bijvoorbeeld omdat het niet precies duidelijk is hoe het een met het ander samenhangt. Hoe beslis je – op basis van regels – bijvoorbeeld of een bepaald weefsel geïnfecteerd is, of wat de kwaliteit van een bepaalde partij appels is?

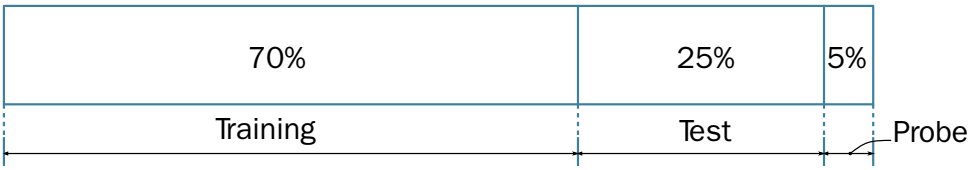
In plaats van kennisgebaseerd op zoek te gaan naar regels van de functie die invoer vertaalt naar uitvoer, hanteert machine learning een ander paradigma. De computer zoekt zelf op basis van voorbeelden naar de optimale parameters van de functie (zie figuur 2). In de terminologie van machine learning wordt deze functie aangeduid als model.



FIGUUR 2: MACHINE LEARNING IS DATA-GEDREVEN ZOEKEN NAAR DE BESTE FUNCTIE DIE INPUTS OMREKENT NAAR OUTPUTS.

Het ‘leren’ van het model gebeurt in de trainingsfase. Tijdens deze fase worden er voorbeelden aangeboden aan het model. Die voorbeelden zijn steeds in de vorm ‘bij deze *input* verwacht ik deze *output*’. Het model zoekt de parameters die de beste afstemming opleveren tussen input en output, en kan na de training zelfstandig nieuwe inputs vertalen in outputs.

Er moet aan een aantal basisvoorwaarden voldaan worden om tot een betrouwbaar model te komen. Ten eerste, moet de data die gebruikt wordt representatief zijn voor het onderliggende fenomeen. In het voorbeeld van hypotheekaanbiedingen betekent dit dat alle relevante gegevens over huizen en klanten bekend moeten zijn en dat deze gegevens een goede afspiegeling van de werkelijkheid moeten zijn. De verdeling tussen mannen en vrouwen in de data moet bijvoorbeeld representatief zijn voor de werkelijke populatie. Ten tweede, is het belangrijk dat het model correct wordt geëvalueerd. Een voorwaarde is dat de voorbeelden die gebruikt worden tijdens de trainingsfase niet gebruikt mogen worden om een model te testen. Daarom wordt een dataset vaak verdeeld in een training- en testset (zie figuur 3). Soms wordt er ook een aparte probeset achtergehouden als extra controle, bijvoorbeeld tijdens oplevering. De ontwikkelaars van het model hebben daar dan geen toegang toe.



FIGUUR 3: OPDELEN VAN EEN DATASET IN DRIE DELEN: TRAITRAININGS-NG-, TEST- EN PROBESET.

Een goed model wordt verkregen als:
1 het algoritme past bij de data *(oftewel 'kies het juiste gereedschap')*;
2 de data juist is geselecteerd *(een goede representatie van de werkelijkheid)*;
3 de trainingsfase de juiste duur heeft
(te kort betekent dat niet alle aanwezige informatie wordt meegenomen, te lang zorgt ervoor dat het model de aanwezige ruis leert).

Er worden drie soorten machine learning onderscheiden. Als eerste is er *supervised learning*, waarbij tijdens de training bekend is welke output wordt verwacht bij welke input. De verwachte output wordt ook wel waarheid of label genoemd. Hier tegenover staat staat *unsupervised learning*, waarbij er helemaal geen sprake is van een waarheid maar waar wel interesse is om de interne structuren van de data te doorgronden. Clustering is een voorbeeld van zo'n analyse. Ertussenin zit *reinforcement learning*, waarbij er alleen een uitkomst op termijn bekend is, zoals bij spellen of beurskoersen. In 'Veel gebruikte machine learning methodes' bespreken we enkele veelgebruikte machine learning methodes in meer detail.

Machine learning brengt uitdagingen met zich mee, wat betekent dit voor organisaties?

Het toepassen van machine learning brengt uitdagingen met zich mee. In het vervolg van dit hoofdstuk richten we ons op deze uitdagingen en wat dat betekent voor organisaties.

VAKMANSCHAP VEREIST

Het zou mooi zijn als er één perfect model bestond dat gebruikt kon worden voor allerlei toepassingen. Helaas is dit om meerdere reden onmogelijk. Ten eerste, laat de *No Free Lunch Theorem* zien dat vooraf niet bepaald kan worden welke machine learning methode het meest geschikt is voor een bepaald probleem (Wolpert and Macready 1997). Dit betekent dat men altijd meerdere methodes moet uitproberen en vergelijken. Ten tweede, is aangetoond dat een model nooit perfect zal zijn. De *Bayes Error Rate* stelt namelijk dat een model altijd een zeker aantal fouten zal maken − behalve in triviale gevallen (Fukunaga 1990). Ten derde, is het onmogelijk om een model te trainen op toekomstige data. Doordat de wereld verandert is het waarschijnlijk dat de oorspronkelijke data − en daarmee het getrainde model − na verloop van tijd niet meer representatief zal zijn. Dit heeft tot gevolg dat een model na verloop van tijd meer fouten zal maken en aanpassingen noodzakelijk zijn.

Bovenstaande redenen laten zien dat het toepassen van machine learning altijd aandacht en vakmanschap vereist. Op het moment van implementatie, maar ook in de toekomst wanneer een model wordt gebruikt in een productieomgeving. In de volgende paragrafen besteden we aandacht aan vier uitdagingen die hierbij vaak terugkomen, namelijk *sampling bias*, *overfitting*, *fairness*, en *foutmarges*.[5] De eerste twee uitdagingen kunnen data scientists grotendeels zelf adresseren terwijl de inzet van verschillende personen in een organisatie vereist is voor de laatste twee.

SAMPLING BIAS

Tijdens de Amerikaanse presidentsverkiezingen in 1948 kopte de ‘Chicago Daily Tribune’ dat de uitslag bekend was. De krant had onderzoek gedaan door telefonische enquêtes en wist met grote zekerheid dat Dewey zou winnen. De uiteindelijke uitslag was heel anders: niet Dewey maar Truman had de verkiezingen gewonnen.

PRESIDENT TRUMAN LACHT TRIOMFANTELIJK NA DE STATISTISCHE BLUNDER VAN DE ‘CHICAGO DAILY TRIBUNE’ DIE VEROORZAAKT WERD DOOR SAMPLING BIAS.

5 Naast de voorbeelden in dit white paper is deze poster over data fallacies ook zeer de moeite waard. https://data-literacy.geckoboard.com/assets/pdf/data-fallacies-to-avoid.pdf

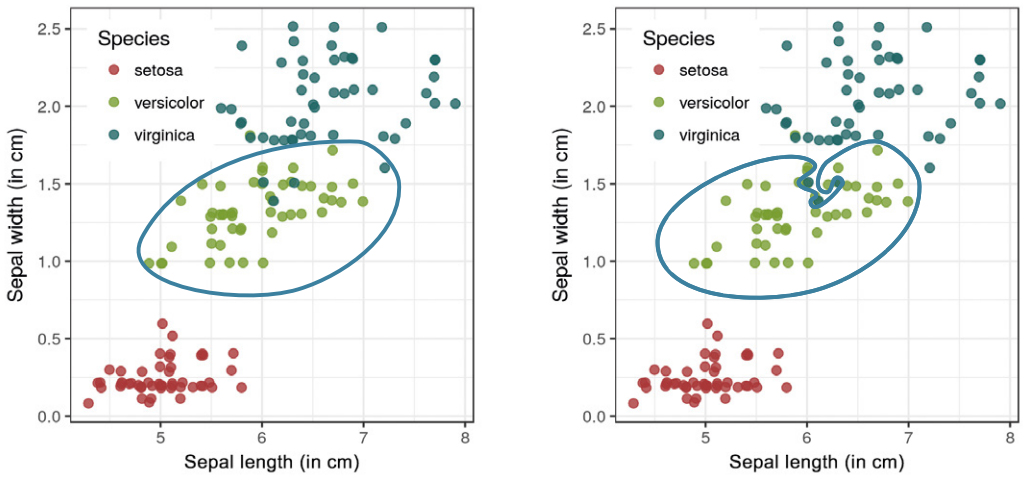
De gemaakte fout staat bekend als *sampling bias*. Door mensen telefonisch te enquêteren werd de gestelde vraag beantwoord door mensen die significant rijker zijn dan gemiddeld, want telefoonaansluitingen werden pas later gemeengoed.

Bij het verzamelen van data moet daarom geprobeerd worden om sampling bias te vermijden als dat mogelijk is. Wanneer dat niet mogelijk is dient er rekening te worden gehouden met het feit dat er bias in de dataset zit, bijvoorbeeld door er voor te compenseren.

OVERFITTING

Tijdens de trainingsfase stemt een model zich af op de voorbeelden die worden aangeboden. Teveel afstemming is ongewenst, omdat het model dan slecht zal presteren wanneer het in een aanraking komt met nieuwe, ongeziene voorbeelden. Dit probleem wordt ook wel *overfitting* genoemd.

We kunnen dit probleem ook illustreren middels een voorbeeld. In figuur 5 staan twee classificatiemodellen van de bekende Iris-dataset met daarin de hoogte en breedte van 150 bloembladen (Ronald Fisher, 1936). In de figuur *(links)* een voorbeeld van een goed model: het model is afgestemd op de data en negeert de details die er niet toe doen; in dit geval de drie blauwe punten die qua eigenschappen erg lijken op de groene bloemen. Daarnaast een model dat lijdt aan overfitting: door teveel rekening te houden met de details in de trainingset wordt het model te nauwkeurig. Tenminste: nauwkeurig op de trainingset, want bij toepassing in de praktijk zal het model tegenvallen.



FIGUUR 5: (LINKS) DE IRIS-DATASET MET EEN NATUURLIJKE CLASSIFICATIE. DE ORANJE LIJN REPRESENTEERT DE GRENZEN VAN EEN KLASSE (IN DIT VOORBEELD DE BLOEMSOORT) DAT DOOR HET MODEL IS GEVONDEN. DE KLEUR VAN DE PUNTEN STAAT VOOR DE DAADWERKELIJKE KLASSE. (RECHTS) DEZELFDE DATASET MET EEN CLASSIFICATIE DIE HET GEVOLG IS VAN OVERFITTING.

FAIRNESS IS EEN AFWEGING

Wie op internet zoekt naar plaatjes van een ‘CEO’ vind bijna uitsluitend blanke mannen die ouder zijn dan 55 jaar. Dit is omdat in werkelijkheid het merendeel van de *Chief Executive Officers* precies dat is: blanke mannen van boven de 55 jaar. Toch kan deze uitkomst beschouwd worden als ongewenst.

In tegenstelling tot bij het voorbeeld over de verkiezingen van president Truman is er hier geen sprake van *selection bias*; de trainingsset in zijn geheel is biased en de resulterende datadienst is intrinsiek oneerlijk, ofwel *unfair*.

Resteert de vraag wat hieraan te doen is. Welk deel van de trainingsdata dient dan buiten beschouwing te worden gelaten? Op basis van welke vragen – en antwoorden – kan hierin een heldere keuze worden gemaakt, die niet alleen technisch uitvoerbaar is maar ook resulteert in een eerlijke datadienst? Data scientists kunnen hierover meedenken, maar uiteindelijk is fairness een afweging voor de organisatie als geheel.

INTERPRETATIE VAN FOUTMARGE

Data scientists zoeken naar een model dat een goede beschrijving oplevert van de werkelijkheid. Echter, een model is nooit 100% nauwkeurig. Stel dat een data scientist een model oplevert met de claim ‘mijn model is 96% nauwkeurig’. Wat betekent dit dan? Een verstandige manier om hiernaar te kijken is ‘mijn model zit er in minstens 4% van de gevallen naast, en straks in de praktijk waarschijnlijk meer’. Aan anderen binnen het bedrijf de opdracht om te onderzoeken hoe dit uitpakt in die gevallen en voor welke klanten dat relevant is.

Positieve of negatieve effecten van een datadienst zijn ook relevant voor andere stakeholders

MACHINE LEARNING IN ORGANISATIES

Zoals we hierboven hebben gezien, is het niet eenvoudig om een goed functionerende datadienst te ontwikkelen. Bovendien is een goed functionerende datadienst nooit perfect: modellen zijn vereenvoudigingen, modellen die nu representatief zijn, zijn dat in de toekomst minder. Data scientists kunnen een deel van deze uitdagingen zelf adresseren. Het machine learning paradigma en de randvoorwaarden die aan het begin van dit hoofdstuk aan bod kwamen zal iedereen die zelf modellen ontwikkeld bekend voorkomen. Toch is de uiteindelijk geleverde datadienst niet alleen een zaak voor de ontwikkelaars. De positieve of negatieve effecten van een datadienst zijn namelijk ook relevant voor andere stakeholders waaronder product managers, marketeers, leidinggevenden en klanten. Het is daarom belangrijk dat verschillende medewerkers de randvoorwaarden en uitdagingen van machine learning begrijpen. Op deze manier kan de organisatie als geheel weloverwegen keuzes maken over de inzet van machine learning.

GEVOELIGE GEGEVENS ANALYSEREN

In veel gevallen zijn de gegevens die datadiensten verwerken (privacy)gevoelig. In dit hoofdstuk introduceren we technieken om een databron te ontdoen van gevoelige gegevens. Vervolgens kijken we naar methodes waarmee meerdere organisaties gezamenlijk een datadienst kunnen ontwikkelen zonder dat ze gevoelige gegevens hoeven te delen. Tot slot tonen we twee raamwerken om te evalueren in hoeverre een geanonimiseerde databron nog gevoelige gegevens bevat.

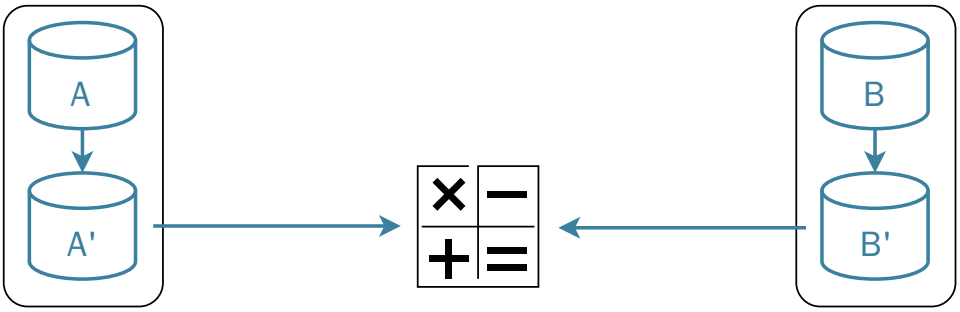
DATA ONTDOEN VAN GEVOELIGE INFORMATIE

Anonimiseren en pseudonimiseren zijn methodes om data te ontdoen van identificeerbare gegevens. Het onderscheid tussen de twee methodes is belangrijk wanneer sprake is van een databron met persoonsgegevens. Technieken die anonimiseren zorgen er voor dat het onmogelijk wordt om gegevens aan een individu te koppelen. Dit betekent dat data niet langer persoonsgegevens zijn en de AVG dus niet meer van toepassing is. Een technologie die pseudonimiseert maakt het veel lastiger om individuen te identificeren, maar dit is (in theorie) niet onmogelijk. Dit betekent dat de AVG nog steeds van toepassing is en pseudonimiseren wordt gezien als beveiligings-maatregel om ongewenste herleiding tegen te gaan. Het is nog niet altijd duidelijk of technologieën die we hier beschrijven worden beschouwd als anonimiseren of pseudonimiseren.

Stel dat informatie wordt verzameld over het aantal personen dat last heeft van een seksueel overdraagbare aandoening (SOA). Individuen willen niet delen of ze een SOA hebben omdat dit een gevoelig gegeven is. Daarom verbergt ieder individu zijn antwoord door twee munten te werpen. De eerste worp bepaalt of hij eerlijk zal antwoorden. Bij kop antwoordt het individu oprecht of hij een SOA heeft, maar bij munt bepaalt de tweede worp het antwoord. Een individu antwoordt namelijk ‘ja’ als de tweede worp kop is en ‘nee’ bij munt. Door het antwoord op deze manier te bewerken kan ieder individu veilig de gevoelige informatie delen. De ontvanger weet immers niet of een antwoord oprecht is of bepaald door toeval. Tegelijkertijd is de kansverdeling van eerlijke antwoorden wel bekend; de kans op kop en munt is beide 50%. De partij die data ontvangt kan de bewerkte antwoorden en de kansverdeling met elkaar te combineren waardoor wel degelijk inzichten kunnen worden verkregen over het aantal SOAs in de hele populatie.

Randomized response, een voorbeeld van een anonimiseringstechniek die gevoelige informatie ontdoet van de verwijzing naar een individu

Bovenstaande paragraaf beschrijft een *randomized response*. Dit is een voorbeeld van een anonimiseringstechniek die gebruikt kan worden om gevoelige informatie van een individu te verbergen . Hierdoor ontstaat een afgeleide, ongevoelige databron die veilig gedeeld kan worden met andere partijen. De ontvangende partij heeft geen toegang meer tot de gevoelige informatie terwijl hij wel in staat is om de gewenste data-analyse uit te voeren (zie tevens figuur 6).



FIGUUR 6: PARTIJEN BEWERKEN HUN DATABRONNEN EN ANALYSEREN MET ONGEVOELIGE DATA.

In sommige gevallen is het niet gewenst om databronnen te ontdoen van gevoelige informatie, bijvoorbeeld omdat het dataverlies dat hierdoor ontstaat onacceptabel is. Als dit het geval is kunnen partijen er voor kiezen te analyseren zonder gevoelige data te delen.

ANALYSEREN ZONDER GEVOELIGE DATA TE DELEN

Stel dat een groep mensen hun totale gewicht wil berekenen, zonder hun gewicht met elkaar of een centrale partij te hoeven delen. De deelnemers splitsen daarom hun gewicht in twee willekeurige getallen. De enige voorwaarde is dat de twee getallen bij elkaar opgeteld het werkelijke gewicht vormen. Iemand die 80 kilogram weegt kan bijvoorbeeld zijn gewicht splitsen in de getallen 30 en 50 of -542 en 622. Deze twee getallen zijn apart van elkaar betekenisloos en kunnen dus veilig gedeeld worden. Tegelijkertijd zijn het nog steeds getallen waarmee gerekend kan worden. Deelnemers kunnen hun betekenisloze getallen uitwisselen en vervolgens ieder apart de som van de betekenisloze getallen die ze ontvangen hebben uitrekenen. Door deze tussen-uitkomsten weer bij elkaar op te tellen wordt het werkelijke antwoord verkregen, namelijk het totale gewicht van de groep.

Secret sharing: deelnemers splitsen gevoelige gegevens in betekenisloze getallen die veilig gedeeld kunnen worden

Bovenstaande paragraaf beschrijft *secret sharing* . Dit is een voorbeeld van een techniek waarmee organisaties enkel (tussen)uitkomsten met elkaar hoeven te delen. De oorspronkelijke, gevoelige data die nodig is voor een analyse wordt niet of alleen op een onleesbare manier gedeeld (zie figuur 7).

FIGUUR 7: PARTIJEN VOEREN IEDER EEN DEEL VAN DE ANALYSE UIT EN DELEN ENKEL ONGEVOELIGE (TUSSEN) UITKOMSTEN MET ELKAAR.

Er zijn grofweg twee strategieën om gezamenlijk analyses uit te voeren zonder de onderliggende databronnen te hoeven delen. In sommige gevallen kunnen analyses in verschillende onderdelen gesplitst worden zodat organisaties enkel ongevoelige tussenuitkomsten met elkaar delen. Daarnaast is het mogelijk om de gevoelige databron op een onleesbare manier te delen, bijvoorbeeld door deze te vercijferen. Hieronder geven we een voorbeeld van beide strategieën.

ONGEVOELIGE TUSSENUITKOMSTEN DELEN

Concepten als *distrubuted learning* en *federated learning* gaan er vanuit dat partijen los van elkaar machine learning modellen trainen die later gecombineerd worden. Het uitgangspunt is dat de losse modellen inaccuraat zijn maar het gecombineerde model wel goed presteert. De uitdaging is om modellen op de juiste manier te combineren. Men kan er bijvoorbeeld voor kiezen om modelparameters samen te voegen of om de uitkomsten van twee modellen te middelen.

Een bekende toepassing van *federated learning* is Google's aanpak voor het voorspellen van toetsaanslagen (McMahan et al. 2016). De privacygevoelige toetsaanslagen van smartphonegebruikers werden in dit onderzoek niet naar een centrale server gestuurd. In plaats daarvan werd op iedere smartphone een apart model getraind. Vervolgens werden de minder gevoelige modelparameter op een veilige manier gedeeld en samengevoegd tot een universeel voorspellingsmodel.

Secure multi-party computation technieken zoals homomorfe vercijfering maken het mogelijk om onleesbare data te analyseren

ONLEESBARE DATA DELEN

Secure multi-party computation technieken zoals *homomorfe vercijfering* maken het mogelijk om onleesbare data te analyseren. Homomorfe vercijfering is bijvoorbeeld toegepast in het TNO-project PRANA-DATA. In dit project is een proof of concept ontwikkeld waarbij een arts die een baby behandelt inzichten krijgt over het verwachte groeipatroon en eventuele complicaties die kunnen optreden (bijvoorbeeld obesitas). Deze voorspellingen werden berekend door de medische records van vergelijkbare kinderen samen te voegen. De medische records zijn gevoelig en waren daarom vercijferd middels homomorfe vercijfering. Hierdoor was het mogelijk om de aggregatie te berekenen zonder toegang te hebben tot de gevoelige data.

Zowel de data die gebruikt wordt tijdens een analyse, als de berekeningen en de tussenuitkomsten zijn onleesbaar wanneer gebruik wordt gemaakt van technieken als homomorfe vercijfering. Men zou kunnen stellen dat de onleesbare gegevens geen persoonsgegevens meer zijn en daarom buiten de AVG vallen (Spindler and Schmechel 2016). De technieken geven echter geen garanties over de (on)gevoeligheid van een uitkomst nadat deze is ontcijferd. Het is dus mogelijk dat een uitkomst nog steeds herleidbaar is naar een individu. Om deze reden wordt secure multi-party computation soms gebruikt in combinatie met andere anonimiseringstechnieken.

RAAMWERKEN OM DE GEVOELIGHEID TE TOETSEN

De twee oplossingsrichtingen die tot nu toe zijn beschreven, verwijderen gevoelige gegevens en enkel (tussen)uitkomsten delen, trachten databronnen op zo'n manier te analyseren dat de gevoeligheid van data gerespecteerd wordt. Het is belangrijk om te evalueren in hoeverre de technieken hierin slagen. Met andere woorden, wanneer is een databron voldoende geanonimiseerd? In de rest van dit hoofdstuk beschrijven we twee formele raamwerken die gebruikt kunnen worden om deze vraag te beantwoorden: *k-anonymity* en *differential privacy*.

Wanneer is een databron voldoende geanonimiseerd om gedeeld te worden? Formele raamwerken kunnen helpen bij het beantwoorden van deze vraag

K-ANONYMITY: KOPPELING NAAR INDIVIDU VERDWIJNT IN DE GROEP

k-anonymity is een bekend raamwerk om te toetsen of gegevens herleidbaar zijn tot individuen. Dit raamwerk vereist dat de gegevens die gebruikt kunnen worden om een individu te identificeren nooit uniek mogen zijn (Sweeney 2002). Iedere combinatie van identificeerbare gegevens moet minimaal k keer voorkomen; vandaar de naam *k*-anonymity. Het effect hiervan is dat een individu verdwijnt in de groep en het lastiger wordt om te bepalen aan welk individu gevoelige informatie toebehoort.

k-anonymity heeft een paar belangrijke nadelen. Het raamwerk gaat er bijvoorbeeld vanuit dat de gegevens die gebruikt kunnen worden om iemand te identificeren vooraf bekend zijn. Voor bepaalde gegevens zoals postcodes is dit eenvoudig maar voor veel andere gegevens niet. Het is onderzoekers bijvoorbeeld gelukt om 95% van de personen in een databron met gegevens van telefoonmasten te identificeren (Montjoye et al. 2013). Hiervoor waren vier locaties per persoon al voldoende; een combinatie van gegevens die op het eerste gezicht misschien niet uniek lijkt.

Een ander nadeel van *k*-anonymity is dat het werkt volgens een *release and forget* aanpak. Een databron wordt eenmalig geanonimiseerd en vervolgens gedeeld met een of meerdere partijen. Andere partijen kunnen hier vervolgens alles mee doen. Een bekend voorbeeld van dit nadeel is de Netflix-prijs. Netflix publiceerde in 2006 een databron met daarin onder andere de tijdstippen waarop geanonimiseerde gebruikers films bekeken. Onderzoekers zochten vervolgens naar openbare recensies op filmwebsite IMDB. De identiteit van gebruikers in de Netflix-databron kon alsnog worden achterhaald door de tijdstippen waarop recensies waren gepubliceerd te vergelijken met tijdstippen waarop films werden bekeken (Narayanan and Shmatikov 2008).

DIFFERENTIAL PRIVACY: VERBERG INFORMATIE VAN EEN INDIVIDU

Differential privacy is een ander raamwerk om te toetsen in hoeverre informatie herleidbaar is tot een individu. Dit raamwerk stelt dat de informatie van een enkel individu de uitkomst van een analyse niet te veel mag beïnvloeden (Dwork and Roth 2013). Differential privacy wordt de laatste jaren steeds meer gebruikt, onder andere door Google en Apple om gebruikersstatistieken over hun software te ontdoen van verwijzingen naar individuele gebruikers. De toegevoegde waarde van het dit raamwerk kan het beste worden uitgelegd aan de hand van een voorbeeld.

Stel dat een grote groep collega's op 31 december hun gemiddelde inkomen berekent. Aangezien het een gemiddelde betreft blijven de individuele inkomens onbekend. Op 1 januari berekenen de collega's nogmaals het gemiddelde inkomen met als enige verschil dat er nu een nieuwe medewerker in dienst is. De collega's kunnen de gemiddelde inkomens van 31 december en 1 januari met elkaar vergelijken en op deze manier achterhalen wat het inkomen is van de nieuwe medewerker. Differential privacy beschermt individuele data zodat het inkomen van de nieuwe medewerker verborgen blijft, bijvoorbeeld door kleine hoeveelheden ruis aan de uitkomst toe te voegen.

De mate waarin individuele datapunten verborgen worden, heeft invloed op het spanningsveld tussen databescherming en de analysekwaliteit. Differential privacy maakt dit spanningsveld expliciet. Het is mogelijk om te kiezen voor meer bescherming, maar dit betekent tegelijkertijd dat de datakwaliteit afneemt. Het is dus belangrijk om per use case een goede afweging te maken tussen datakwaliteit en bescherming.

Het is belangrijk om per use case een goede afweging te maken tussen datakwaliteit en bescherming.

CONCLUSIE: ‘TIJD VOOR IMPLEMENTATIE VAN VERANTWOORDE DATADIENSTEN’

Hier zijn verschillende technische maatregelen beschreven die gebruikt kunnen worden om gevoelige data te analyseren of om de gevoeligheid van data te evalueren. Er wordt veel onderzoek gedaan naar deze technologieën, maar het aantal implementaties in operationele omgevingen blijft achter. Dat is jammer want de technieken kunnen een positieve bijdrage leveren aan de ontwikkeling van nieuwe, verantwoorde datadiensten.

Evalueer en monitor zowel het implementatieproces als de uiteindelijke oplossing met verschillende stakeholders

Wat ons betreft is de tijd rijp om te innoveren en de technieken toe te passen. Zoals beschreven, is het belangrijk om hierbij de hele organisatie te betrekken. Evalueer en monitor zowel het implementatieproces als de uiteindelijke oplossing met verschillende stakeholders. Op deze manier wordt helder in hoeverre gebruikers, klanten en juristen de technologie en de daaruit voorkomende datadienst percipiëren en vertrouwen. Dit helpt de eigen organisatie, maar kan ook bijdragen aan de ontwikkeling van standaardprocedures, handvatten en (open source) bouwblokken zodat toekomstige datadiensten sneller van de grond komen.

TNO wil graag de volgende stap zetten en technieken voor verantwoorde datadiensten verder ontwikkelen en implementeren. Daarvoor gaan we graag de samenwerking met u aan.

REFERENTIES

Abu-Mostafa, Yaser S. 2012. Learning from Data. AMLBook. https://www.xarg.org/ref/a/B0759M2D9H/.

Casey, Bryan, Ashkon Farhangi, and Roland Vogl. 2018. 'Rethinking Explainable Machines: The Gdpr's' Right to Explanation'Debate and the Rise of Algorithmic Audits in Enterprise.'

Dwork, Cynthia, and Aaron Roth. 2013. 'The Algorithmic Foundations of Differential Privacy.' Foundations and Trends in Theoretical Computer Science 9 (3-4). Now Publishers: 211–407. https://doi.org/10.1561/0400000042.

European Commission. 2016. 'Data Protection.' https://eur-lex.europa.eu/legal-content/NL/TXT/?uri=OJ:L:2016:119:TOC.

European Commission 2018. 'Mid-Term Review of the Digital Single Market (Dsm) – a Good Moment to Take Stock.' https://ec.europa.eu/digital-single-market/en/content/mid-term-review-digital-single-market-dsm-good-moment-take-stock.

Fukunaga, Keinosuke. 1990. 'Chapter 3 - Hypothesis Testing.' In Introduction to Statistical Pattern Recognition (Second Edition), edited by Keinosuke Fukunaga, Second Edition, 51–123. Boston: Academic Press. https://doi.org/10.1016/B978-0-08-047865-4.50009-0.

Hoepman, Jaap-Henk. 2018. 'Privacyontwerpstrategieën (Het Blauwe Boekje).' https://www.cs.ru.nl/~jhh/publications/pds-boekje.pdf.

McMahan, H. Brendan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. 2016. 'Communication-Efficient Learning of Deep Networks from Decentralized Data.' Proceedings of the 20 th International Conference on Artificial Intelligence and Statistics (AISTATS) 2017. JMLR: W&CP volume 54.

Montjoye, Yves-Alexandre de, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. 2013. 'Unique in the Crowd: The Privacy Bounds of Human Mobility.' Scientific Reports 3 (1). Springer Nature. https://doi.org/10.1038/srep01376.

Narayanan, Arvind, and Vitaly Shmatikov. 2008. 'Robust de-Anonymization of Large Sparse Datasets.' In Security and Privacy, 2008. SP 2008. IEEE Symposium on, 111–25. IEEE.

Patterson, Josh. 2017. Deep Learning: A Practitioner's Approach. O'Reilly Media. http://shop.oreilly.com/product/0636920035343.do.

Russell, Stuart, and Peter Norvig. 2013. Artificial Intelligence: A Modern Approach, 3/E. Pearson Education. https://www.amazon.com/dp/B00566HTI4.

Samuel, Arthur L. 1959. 'Some Studies in Machine Learning Using the Game of Checkers.' IBM Journal of Research and Development 3 (3). IBM: 210–29.

Spindler, Gerald, and Philipp Schmechel. 2016. 'Personal Data and Encryption in the European General Data Protection Regulation.' J. Intell. Prop. Info. Tech. & Elec. Com. L. 7. HeinOnline: 163.

Sweeney, Latanya. 2002. '*K*-Anonymity: A Model for Protecting Privacy.' International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10 (05). World Scientific: 557–70.

Wolpert, D.H., and W.G. Macready. 1997. 'No Free Lunch Theorems for Optimization.' IEEE Transactions on Evolutionary Computation 1 (1). Institute of Electrical; Electronics Engineers (IEEE): 67–82. https://doi.org/10.1109/4235.585893.

VEEL GEBRUIKTE MACHINE LEARNING METHODES

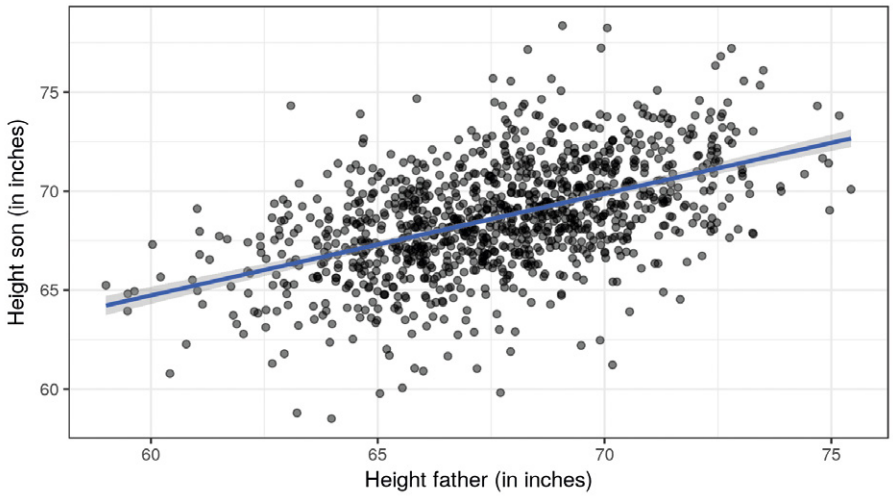
In deze appendix bespreken we enkele veelgebruikte machine learning methodes. Het betreft twee *supervised learning* methodes en een beroemde nieuwe techniek genaamd *deep learning*.

SUPERVISED LEARNING

Onder supervised learning vallen alle methodes om een model te vinden die de invoer het beste omzet naar de uitvoer en tevens uitgaan van een bekende waarheid. Die waarheid wordt meestal gegeven door een mens en dat is ook waar het woord ‘supervised’ op slaat. Het meest eenvoudige voorbeeld is de lineaire regressie, dus daar beginnen we mee. Vervolgens kijken we naar classificatie, het proces om invoer in te delen in een bepaalde categorie (klasse).

REGRESSIE ANALYSE

Bij een simpele lineaire regressie is het model waar we naar op zoek zijn een rechte lijn die de dataset beschrijft. Zo’n model kunnen we bijvoorbeeld trainen voor de bekende dataset van Pearson (rond 1899). Deze dataset bevat iets meer dan 1000 datapunten met daarin de lengte van vaders en hun zonen. Figuur 9 toont het regressiemodel dat is getraind op deze dataset. De lijn geeft een ruwe beschrijving van de dataset. Zo kan worden afgelezen dat als een vader 75 inch lang is (1m90), de verwachting is dat hij een zoon heeft van rond de 72,5 inch (1m84).



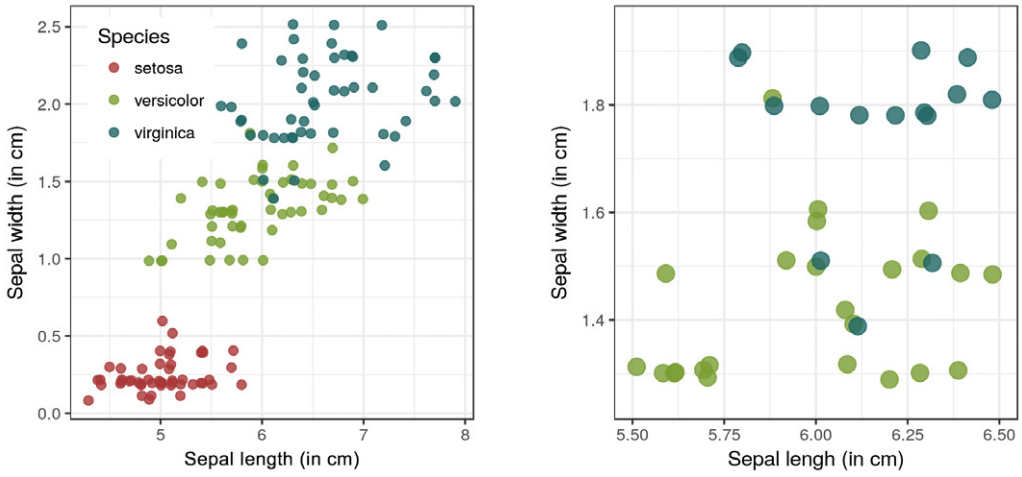
FIGUUR 9: EEN LINEAIR REGRESSIEMODEL DAT IS GETRAIND OP DE PEARSON DATASET.

In het voorbeeld hierboven is de regressie lineair met twee dimensies, dus een rechte lijn. In de praktijk worden vaak complexere regressies gedaan met meerdere dimensies. De grote toepasbaarheid en inzichtelijkheid maakt dat regressies populaire machine learning methodes zijn.

CLASSIFICATIE

Classificatie is het proces om van een bepaalde input te bepalen in welke vaste categorie die zit. Bijvoorbeeld van welke soort bloem een bloemblad afkomstig is, of van welke soort een appel is.

Om dit te illustreren gebruiken we de beroemde *Iris* dataset (Ronald Fisher, 1936). In onderstaand figuur *a* (links) staat van 150 bloembladen de hoogte uitgezet tegen de breedte:



De aangegeven kleuren geven de werkelijke soort aan. Als we inzoomen (figuur *b*, rechts) dan is duidelijk te zien dat de soorten niet heel eenvoudig te scheiden zijn op basis van deze meetgegevens. Dit is normaal in alle realistische datasets. Een poging om alsnog te proberen de klassen te scheiden leidt al snel tot *overfitting*.

DEEP LEARNING

Deep learning is ‘the new kid on the block’. Het fundament onder deep learning is al ouder; de eerste experimenten dateren al uit de zestiger jaren van de vorige eeuw. Wel is de toepassing een aantal jaar geleden pas echt goed mogelijk geworden door de komst van grootschalige en goedkope computerkracht in de vorm van videokaarten (meer specifiek: GPU’s). Sindsdien heeft ook de innovatie in dit veld een grote vlucht genomen.

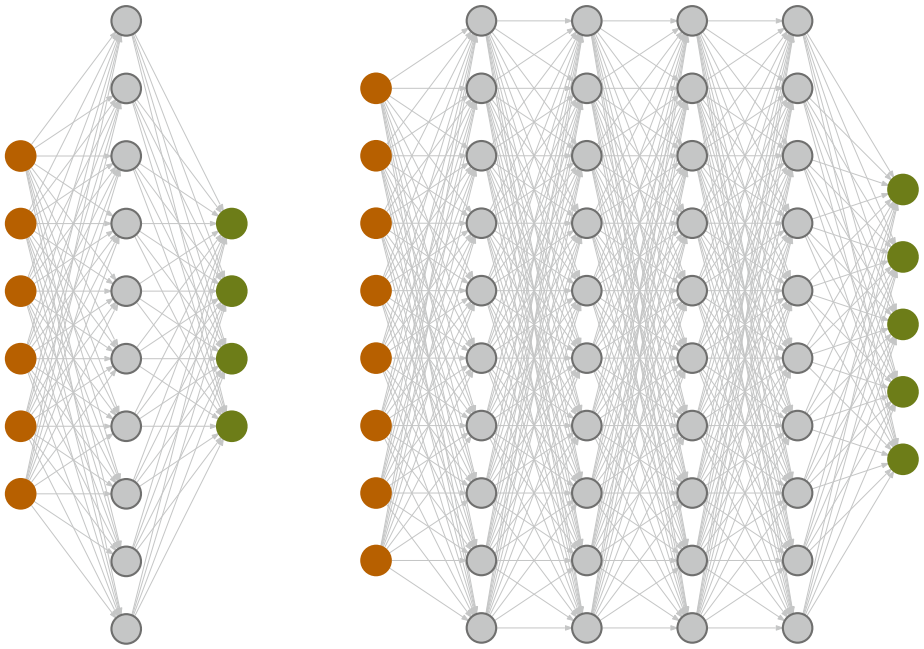
Ook voor deep learning geldt de hiervoor behandelde machine learning indeling.
In deze sectie kijken we daarom naar wat deep learning speciaal maakt. Om te beginnen een korte introductie in de neurale netwerken, dan volgen twee aansprekende voorbeelden (*convolutional neural networks en generative adverserial networks*).

NEURALE NETWERKEN

Neurale netwerken zijn opgebouwd uit *perceptrons*, of neuronen. Deze worden gestapeld tot netwerken met één 'verborgen laag', zie de figuur hieronder (a, links). Dit was tot begin jaren negentig ook de maximale grootte die getraind kon worden.

De *input layer* heeft één neuron voor ieder element van de invoer. Dit kunnen bijvoorbeeld de pixels uit een plaatje zijn of letters uit het alfabet, dat hangt helemaal af van de toepassing. De *output layer* heeft één neuron voor ieder mogelijk resultaat. Bijvoorbeeld een neuron voor kat en een neuron voor hond is voldoende om een kat-of-hond detector te maken.

Het netwerk wordt getraind door links een bepaald input aan te bieden en tegelijkertijd rechts de gewenste uitkomst. Het neurale netwerk verandert de eigenschappen (gewichten) van alle neuronen een klein beetje, en schuift zo per aangeboden voorbeeld iets op richting eindresultaat. Bij het uiteindelijke gebruik kun je dan links een onbekende input aanbieden en rechts uitlezen wat het netwerk daar van 'vindt'.



FIGUUR 10: (LINKS) KLEIN NEURAAL NET MET ENKELE VERBORGEN LAAG, B (RECHTS) GROTER NEURAAL NETWERK MET MEERDERE VERBORGEN LAGEN. BIJ BEIDE: BLAUWE NEURONEN ZIJN INPUTS, GROENE NEURONEN ZIJN OUTPUTS EN DE NEURONEN UIT DE HIDDEN LAYERS ZIJN LICHTGRIJS.

Nadat de belangstelling voor neurale netwerken een tijd weg was werd er begin deze eeuw opnieuw mee geëxperimenteerd. Dit komt doordat de hoeveelheid data en beschikbare rekenkracht sterk is toegenomen. Inmiddels zijn de netwerken een stuk groter geworden en bevatten ze vaak tientallen verborgen lagen van duizenden neuronen. Met deze netwerken kunnen indrukwekkende resultaten behaald worden. Zo kon door de inzet van neurale netwerken in een nieuw domein (spraakherkenning, beeldherkenning) de bestaande nauwkeurigheid zo maar met 15% worden overtroffen.

CONVOLUTIONAL NEURAL NETWORKS

Deze netwerken zijn vooral geschikt gebleken voor de herkenning van objecten in plaatjes, tekst en geluid. In het geval van plaatjes herkennen de eerste lagen in het netwerk patronen zoals arceringen, hoeken, randen en dergelijke, waarna de rest van het netwerk deze onderdelen koppelt om uiteindelijk hele objecten te classificeren. Op deze website staat een interactieve demo die zeer inzichtelijk is en die zeker de moeite van het bekijken waard is.

GENERATIVE ADVERSARIAL NETWORKS

Dit is de technologie achter de successen van de computer op het gebied van spellen. In 1997 werd Kasparov – de toenmalig wereldkampioen schaken – verslagen door Deep Blue. Deze software was nog gebaseerd op het afgaan van zoveel mogelijk potentiële stellingen om de beste zet te vinden. Afgelopen jaar versloeg Google's AlphaGo de Go wereldkampioen Ke Jie in een spel dat vele malen complexer is dan schaken.

De software van Google is gebaseerd op *generative adversarial network*s, oftewel GAN's. Het basisprincipe is bijzonder eenvoudig: neem twee neurale netwerken en stel ze beide zo in dat ze alleen legale zetten in overweging nemen . Train ze vervolgens op de uitkomsten van de spellen die ze tegen elkaar spelen. Iedere uitgevoerde zet heeft een groter of kleiner gevolg voor de uitslag. Een uitslag die in de toekomst ligt weliswaar, maar waarvan wel bepaald kan worden of die goed is (winst van de partij) of slecht (verlies van de partij). Zonder ooit een mens het spel Go te hebben zien spelen werd zo toch software gemaakt die de wereldkampioen versloeg.

Contact
Jean-Louis Roso
Sr. Business Development Manager
UNIT ICT
Den Haag, New Babylon
jean-louis.roso@tno.nl
+31 888 66 72 43

TNO innovation for life

TNO.NL

18-0032