

SMITZH

VORTECH

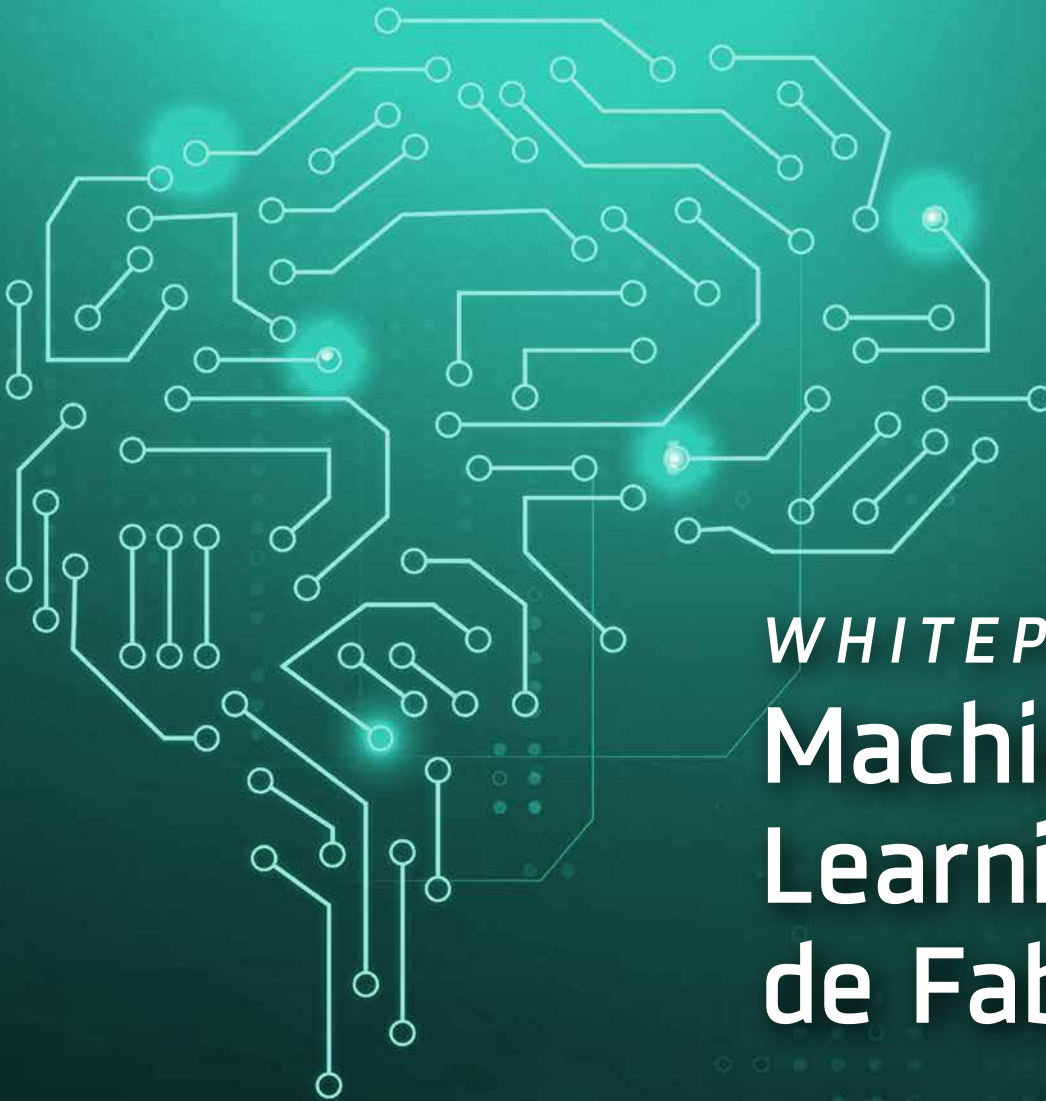


UReason
INTELLIGENCE TO ACT



Innovation
Quarter

Airborne



WHITEPAPER
**Machine
Learning in
de Fabriek**

Hoe kun je data uit machines en machine learning gebruiken om processen te optimaliseren? En hoe pak je zo iets aan? In dit whitepaper leest u hoe dat in z'n werk gaat aan de hand van een voorbeeld. Dat voorbeeld is ontleend aan een project dat is uitgevoerd door een samenwerking van bedrijven in het kader van Smart Manufacturing Industriële Toepassing in Zuid-Holland (SMITZH).

Inleiding

SMITZH is er “voor alle ondernemers die sneller, goedkoper, duurzamer of hoogwaardiger willen produceren door middel van digitalisering”. Veel ondernemers zullen inderdaad sneller, goedkoper etc. willen produceren en willen best met digitalisering aan de slag. Maar dan komt direct de vraag op: hoe dan?

Dit whitepaper probeert een antwoord op die vraag te geven aan de hand van een voorbeeld: het monitoren van kritische parameters in de productie van composieten om zo afkeuren van producten te voorkomen.

We nemen u mee langs de vijf stappen van onze aanpak:

- Stap 0: verzamel de juiste expertise
- Stap 1: wat wil je bereiken?
- Stap 2: de data
- Stap 3: het algoritme
- Stap 4: deployment
- Stap 5: gebruik

Maar voordat we de aanpak gaan uitwerken beschrijven we eerst het vraagstuk dat we als voorbeeld gebruiken.

De voorbeeldcase: Productie van composieten

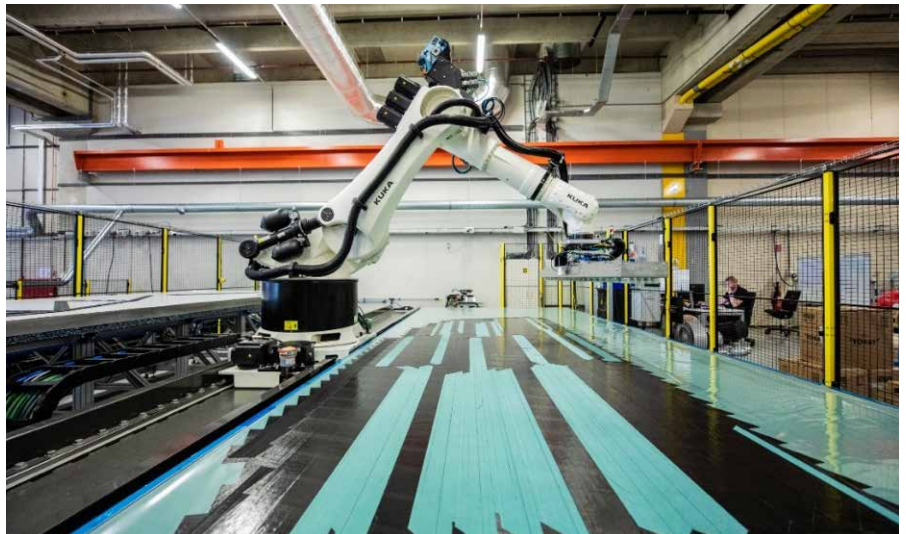
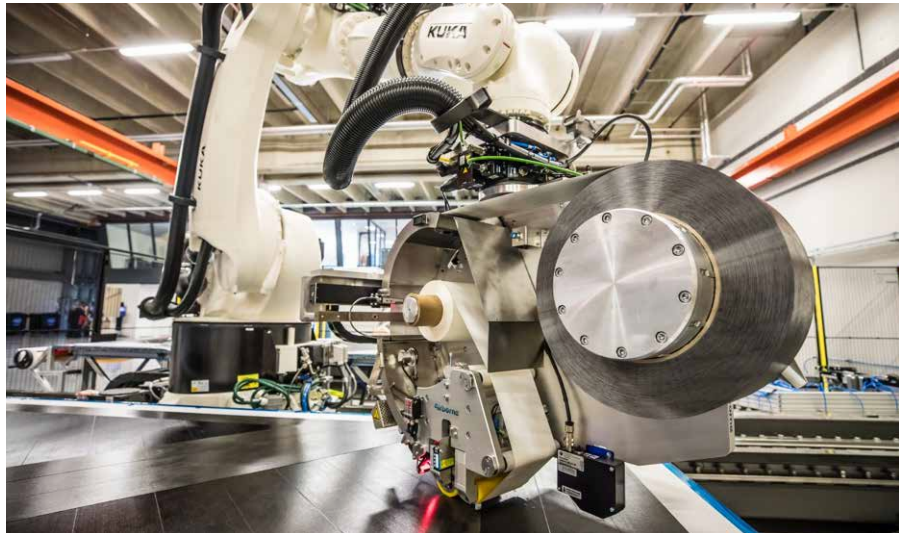
Het voorbeeld dat in dit whitepaper centraal staat is de Automated Laminating cel van Airborne Composites Automation. Daarmee worden halffabricaten voor hoogwaardige composiet onderdelen gemaakt. Hoogwaardige composieten worden vanwege hun eigenschappen veel in de luchtvaart en high-end automotive gebruikt. Het composiet in deze toepassingen bestaat uit een kunststof hars die versterkt wordt met koolstofvezels en wordt aangeleverd als prepreg tape op een rol, waarbij de lange vezels van tevoren geïmpregneerd zijn met de hars.

Het proces in de laminating cel bestaat uit verschillende stappen, eerst wordt een basis laminaat opgebouwd en daarna worden de juiste vormen hieruit gesneden en op een aparte tafel op elkaar gelegd om het product te maken. In deze case is dieper ingegaan op het maken van het basislaminaat. Om dit te maken worden lange stroken tape strak naast elkaar gelegd om zo een laag te vormen. Het laminaat bestaat uit verschillende lagen waarbij de tape telkens in een andere richting wordt gelegd.

Bij dit proces is de afstand tussen 2 tapes in een laag een maatstaf voor de kwaliteit van het eindproduct. Dit gat moet binnen de toleranties vallen van het specifieke product en tegelijkertijd is een overlap van 2 tapes niet toegestaan omdat dit lokaal een verdikking in het product veroorzaakt.

Het doel van het project dat in dit whitepaper centraal staat is om het proces zodanig te gaan beheersen dat deze afstand tussen 2 tapes zo vaak mogelijk aan de eisen voldoet. Daarom wordt gekeken of er een machine learning model gemaakt kan worden dat die afstand kan voorspellen op basis van de instellingen en sensorwaarden van de laminating cell. Als dat lukt, dan heeft Airborne een tool in handen om de instellingen van de machine zodanig te kiezen dat het leggen van de stroken zo goed mogelijk gebeurt.

De automated laminating cel van Airbone Composites Automation.



Stap 0

Verzamel de juiste mensen

Om een project zoals dit tot een goed einde te brengen is veel expertise nodig. Die is vaak bij verschillende mensen te vinden. De volgende rollen zijn in elk geval van belang:

a. De gebruiker

Voorop in het hele proces staat de gebruiker. Dat is meestal degene die verantwoordelijk is voor het proces waar het om gaat. Maar degene die dagelijks met het proces bezig is, de operator, is net zo belangrijk. De kennis over het proces zit vaak bij beiden. Waar de verantwoordelijke meestal het beste weet wat voor de business de belangrijke aspecten zijn, weet de man bij de machine veel meer over de praktische aspecten van het proces. Naast deze twee mensen is ook de IT-verantwoordelijke essentieel: uiteindelijk gaat de machine learning applicatie onderdeel uitmaken van de IT-systemen rondom het proces.

Deze mensen zijn overigens niet alleen essentieel voor het ontwikkelen van de machine learning toepassing. Ze zijn minstens zo belangrijk voor de acceptatie ervan. Degene die ermee moet werken zal de voordelen moeten zien en zich er comfortabel bij moeten voelen. Degene die verantwoordelijk is voor het proces zal de nieuwe toepassing moeten 'verkopen', zowel naar het hogere management als naar de werkvloer. En de IT-verantwoordelijke zal ervoor moeten zorgen dat de mensen van zijn afdeling de nieuwe toepassing kunnen beheren. In de voorbeeld-case waren diverse mensen van Airborne betrokken in de genoemde rollen.

b. De automatiseerder

De meetgegevens moeten uit de machine gehaald worden, opgeslagen en bewerkt worden en leiden tot nieuwe stuursignalen in het proces. Dit is het werk van de automatiseerder die alles weet van de software om data te lezen uit de machine, de workflow tools kent om de data te bewerken en aan te bieden aan een algoritme en de resultaten te visualiseren, en tenslotte weet hoe de resultaten omgezet kunnen worden tot acties in het proces. In het project dat hier als voorbeeld wordt gebruikt was UReason de partij die deze hele infrastructuur ingericht heeft.

*Voor een project
als dit is allerlei
expertise nodig.*

c. De datascientist

Voor het ontwikkelen van een geschikt algoritme is een datascientist nodig met kennis van data science en machine learning. Die datascientist moet ook goed in staat zijn om het proces te begrijpen waar het om gaat. Dat vergt ook gevoel voor techniek. Zeker als het gaat om het begrijpen van de data is het heel belangrijk dat de datascientist snapt wat de data voorstelt. Maar ook in het bouwen van het algoritme is begrip van het proces essentieel. In het voorbeeldproject was VORtech de partij die de data science en de techniek bij elkaar bracht.

Naast de genoemde partijen deden in dit project ook nog TNO, Produmize (leverancier van smart industry oplossingen voor het MKB) en Focus-ON (autonome control-oplossingen voor industrie 4.0) mee. De laatstgenoemden hebben kennis en ervaring gedeeld met de bovengenoemde partijen, wat bij dit soort nieuwe technologie enorm behulpzaam is.

Stap 1

Wat wil je bereiken?

*Een vraagstuk
moet heel
specifiek gemaakt
worden.*

De eerste en misschien wel belangrijkste stap is om vast te stellen wat het vraagstuk is. Dat klinkt eenvoudig, maar in het voorbeeld dat we in deze whitepaper behandelen is daar toch de nodige discussie over geweest. De mensen van Airborne hadden een aantal ideeën ter verbetering van de Automated Laminating Cell en het gebruik ervan. Een van de ideeën was om te voorspellen wanneer er onderhoud nodig is. Dan zou je op tijd onderdelen kunnen vervangen of schoonmaken voordat het apparaat storing geeft. Een ander idee, daaraan gerelateerd, was of je zou kunnen nagaan of er patronen te vinden zijn in ongeplande stilstand. Tenslotte was er nog een aantal ideeën rondom de kwaliteitsborging. Het materiaal dat verwerkt wordt is duur dus het levert veel voordeel op als je kunt voorkomen dat het proces fout gaat.

De keuze voor het uiteindelijke vraagstuk kwam tot stand in een gesprek tussen alle betrokken partijen. In die discussie werd enerzijds gekeken naar de business waarde (wat levert het op) maar ook naar de data (hebben we de data om dit vraagstuk aan te pakken). Om dat laatste te kunnen beoordelen gaf Airborne ook een introductie over de Automated Laminating Cell zodat alle betrokkenen goed inzicht hadden in de werking ervan.

Op grond van de inzichten van de betrokken partijen maakte Airborne uiteindelijk de keuze om te onderzoeken of fouten bij het proces voorspeld zouden kunnen worden. Zo'n vraagstuk moet dan heel specifiek gemaakt worden zodat de datascientist gericht aan het werk kan. In dit geval leidde dat tot de vraag: wat is de voorspelde afmeting van de 'gap' (de afstand tussen 2 tapes) op een willekeurig tijdstip in de komende minuut, gerekend vanaf het moment waarop de voorspelling gemaakt wordt?

Als dat zou lukken dan zou dat op twee manieren voordeel opleveren. Ten eerste zou het inzicht geven in de factoren die de fouten veroorzaken. Daarmee kunnen er in het proces of in de machine misschien aanpassingen gedaan worden. Ten tweede zou er een applicatie ontwikkeld kunnen worden die real-time voorspelt dat de gap waarde buiten de toegestane limiet zal komen. Hiermee zouden de operators dan tijdig kunnen ingrijpen en het proces kunnen stoppen. Op termijn zou het misschien mogelijk zijn om het voorspelalgoritme direct in de besturing op te nemen zodat het robot pad automatisch gecorrigeerd wordt.

Stap 2

De data

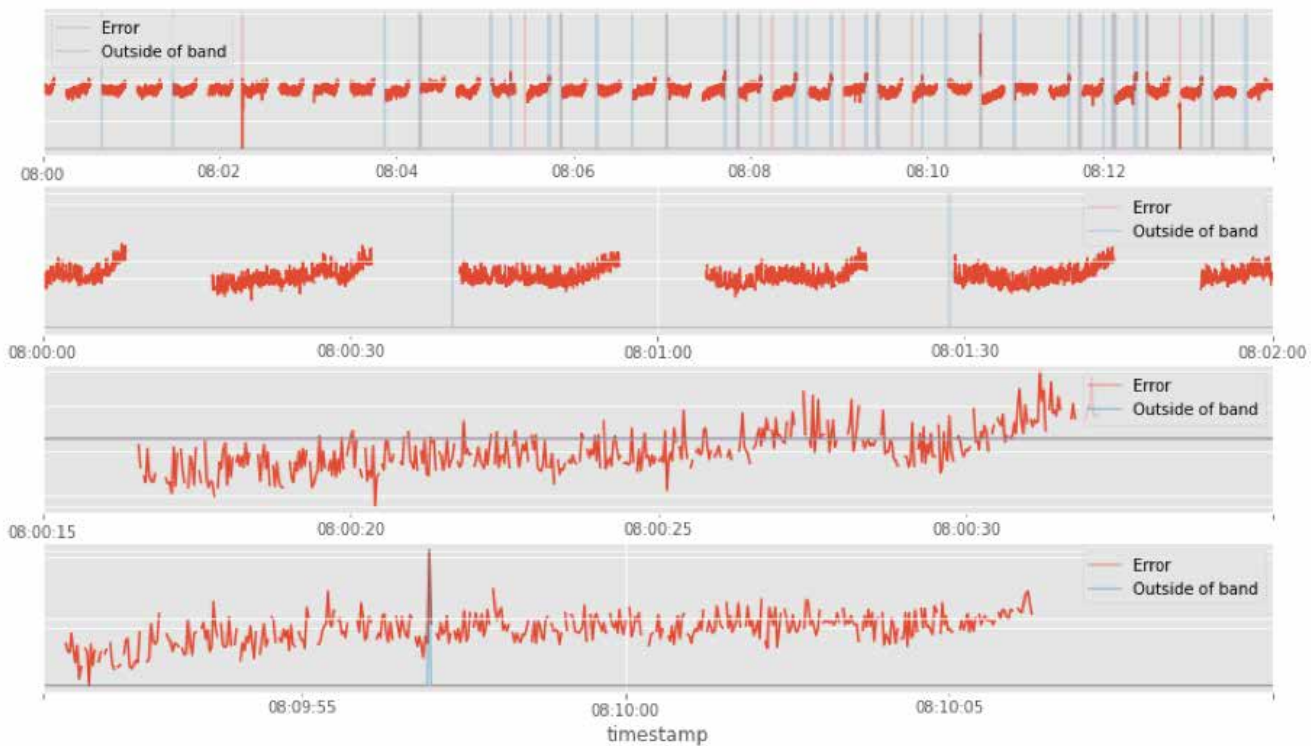
Nadat in stap 1 al is vastgesteld dat de benodigde gegevens inderdaad voorhanden zijn, wordt er in deze tweede stap gekeken wat de kwaliteit van de gegevens is. Dit is ook weer een gezamenlijke exercitie, nu van de datascientist en de gebruiker. De datascientist verkent de gegevens: ze maakt grafieken van de belangrijkste tijdseries en gaat na of er bijzonderheden zijn. De gebruiker is nodig om toe te lichten wat er allemaal in de data te zien is en wat de gegevens precies betekenen.

De gebruiker is nodig om toe te lichten wat de data betekent.

In het voorbeeld dat we hier bespreken (en eigenlijk in elk project) waren er zeer zeker bijzonderheden. Een belangrijk artefact was bijvoorbeeld dat de verschillende sensoren een andere klok hanteerden. De tijdstippen van de metingen moesten dus achteraf op elkaar aangesloten worden. Dat lukte redelijk goed omdat het proces bestaat uit processtappen waarvan het begin en einde in de data te herkennen zijn. Daardoor is goed te bepalen welke verschuiving de verschillende klokken hebben.

Een tweede issue qua tijdstippen was dat niet alle gegevens op vaste tijdstippen gemeten werden. Sommige gegevens werden alleen bij events vastgelegd. Toch wil de datascientist bij voorkeur alle waarden op vaste, equidistante tijdstippen hebben. Dat betekent dat de data geresampled moet worden. Ook moest er iets gebeuren met tijdstippen waarop geen waarde gemeten was en waarbij de juiste invulling niet voor de hand lag. Dit soort aanpassingen moet zorgvuldig gebeuren om geen artefacten te introduceren die de verdere analyse kunnen verstoren. Tenslotte moest ook de data van het begin en einde van een processtap weggegooid worden omdat die de analyse verstoorden. Ook daarbij is zorgvuldigheid een vereiste: er moet vastgesteld worden dat de weggegooid informatie niet nodig is voor de verdere analyse.

Al deze bewerkingen zijn vastgelegd in Python-scripts zodat deze herhaald kunnen worden voor nieuwe data. Een bijkomend voordeel van het vastleggen is dat gemakkelijk een review gedaan kan worden door een collega om te zien of er geen denkfouten zijn gemaakt.



Een kleine uitsnede van de gegevens op verschillende detailsniveaus. De onderste twee afbeeldingen geven verschillende tijdstippen weer.

Stap 3

Het algoritme

Nu het vraagstuk in stap 1 heel specifiek gedefinieerd is en de gegevens in stap 2 beschikbaar zijn gemaakt in de juiste vorm en met de juiste kwaliteit, kan het algoritme ontwikkeld worden. Dit is een taak voor de datascientist.

*Het algoritme
moet het beter
doen dan
de baseline.*

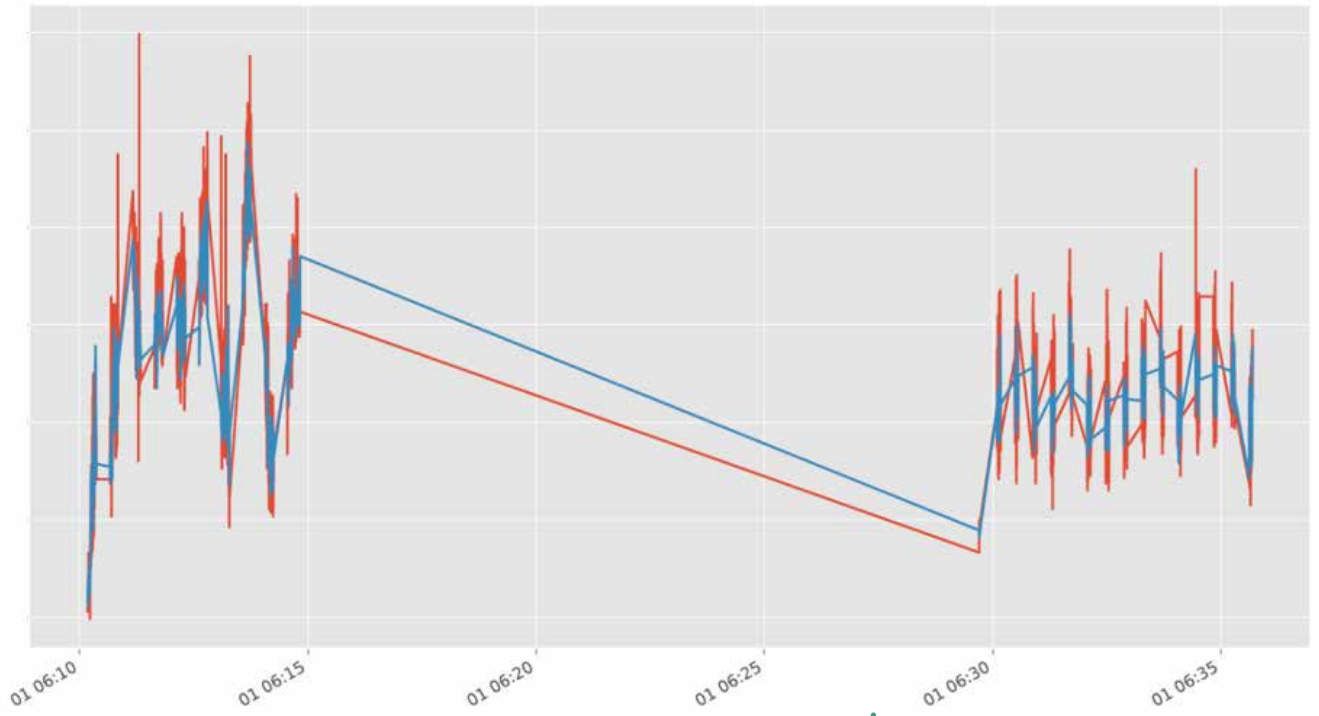
De eerste stap daarin is het vaststellen van een baseline. In dit geval: hoe goed kun je de 'gap' nu al voorspellen zonder geavanceerde algoritmen. Als een complexer algoritme het niet beter doet dan zo'n simpele benadering dan heeft het geen zin om dat complexere algoritme te gebruiken. Een bekend voorbeeld onder datascientists is bijvoorbeeld dat de weersvoorspelling voor morgen vaak al heel goed is als je eenvoudigweg voorspelt dat het hetzelfde zal zijn als vandaag. Zeker in landen met een heel constant klimaat zal een complex algoritme best moeite hebben om het beter te doen.

Ook in dit geval is als eerste zo'n baseline vastgesteld: de 'gap' zal net zo groot blijven als hij was op het moment dat de voorspelling gemaakt wordt. Het zal duidelijk zijn dat het niet uitdagend zou moeten zijn om iets beters te maken. Maar helaas komen we nog steeds situaties tegen waarbij een datasciencespecialist met heel veel geavanceerde algoritmes een voorspelling doet die niet beter is dan wat men met het 'boerenverstand' kan doen. Dus het vaststellen van de baseline is toch belangrijk.

Een datascientist kan uit heel veel benaderingen kiezen om een algoritme te ontwikkelen. Afhankelijk van het vraagstuk ligt een aantal methoden voor de hand. Meestal blijven er dan nog genoeg keuzemogelijkheden over. Op grond van zijn ervaring zal de datascientist daar één of twee benaderingen uit kiezen om uit te proberen. In dit geval is gekozen voor een random forest. Dat is een relatief simpele en robuuste methode die vaak als eerste geprobeerd wordt.

Een mooi aspect van de random forest is dat het als bijproduct ook iets zegt over het relatieve belang van de verschillende gegevens voor de uitkomst (de feature importance). In dit geval is dat een groot voordeel, want het geeft aan wat de belangrijkste factoren voor productiefouten zijn. Dat is dan weer van belang voor Airborne omdat ze daarmee suggesties krijgen over wat er aan het proces en de machine verbeterd zou kunnen worden. Daarnaast geeft de feature importance ook suggesties voor de datascientist om het model verder te verbeteren.

De keuze voor een random forest bleek meteen al afdoende: de 'gap' wordt goed genoeg voorspeld voor de beoogde toepassing en inderdaad veel beter dan met het baseline model.



De gemeten (rood) en voorspelde (blauw) gap voor een tweetal cycli.

Stap 4 Deployment

In Stap 2 is programmatuur gemaakt om de data te verwerken en in stap 3 is programmatuur gemaakt voor het voorspellingsalgoritme. Daarna komt het punt waarop dit alles in de fabriek geïntroduceerd wordt. Hier is de automatisering in de lead, samen met de IT-verantwoordelijke bij de gebruiker.

*Met PMML kan
een algoritme
eenvoudig
geïntegreerd
worden.*

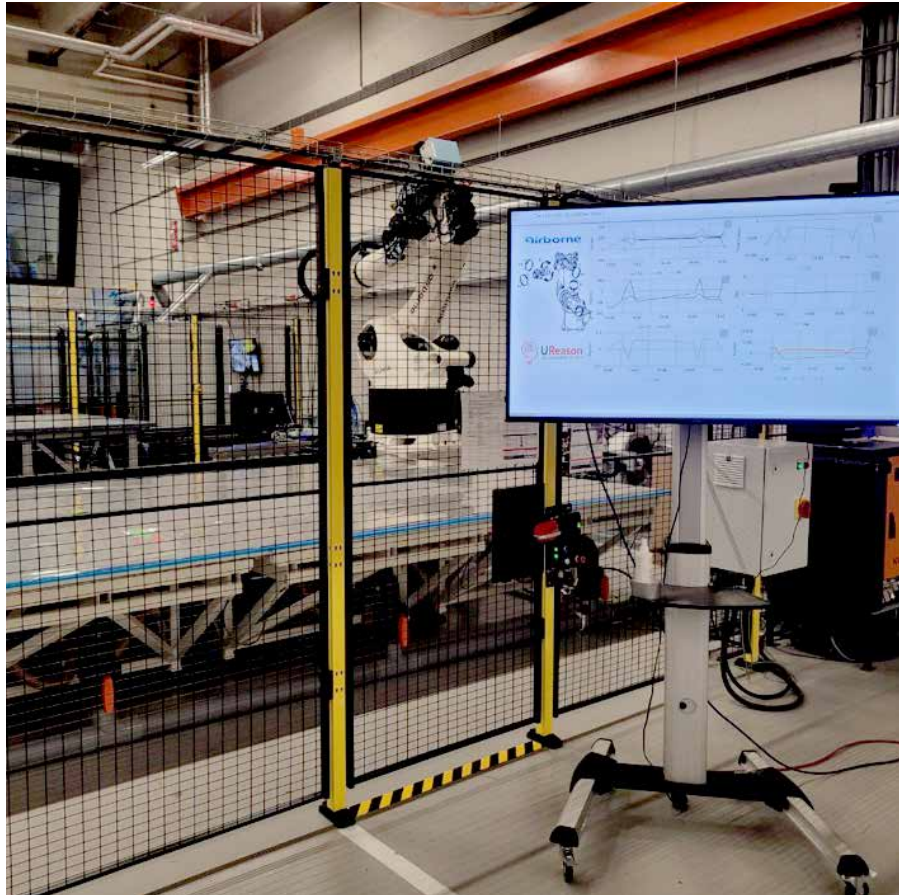
De software voor het verzamelen en bewerken van de gegevens, voor het draaien van het algoritme en voor het presenteren van de uitkomsten moet ingebracht worden in de IT-systemen bij de gebruiker.

In dit geval is ervoor gekozen om de toepassing te installeren op een aparte laptop die gekoppeld is aan de Automated Laminating Cell. Daarmee dringt de nieuwe applicatie niet direct door in de IT-systemen van het bedrijf als geheel en kan het beheer heel lokaal gehouden worden. Als de toepassing later opgeschaald wordt kan alsnog besloten worden om de applicatie op een hoger niveau te integreren.

Als softwareplatform is gekozen voor APM Studio van UReason. Dit is een krachtig softwareplatform waarmee alle benodigde functionaliteit snel kan worden gerealiseerd. Het is eenvoudig om in een grafische interface een workflow te definiëren waarin alle stappen van de slimme applicatie geconfigureerd worden. Dus zowel het ophalen van de gegevens, het bewerken ervan en het draaien van het algoritme, als ook het presenteren en distribueren van de uitkomsten.

Een sterk aspect is verder dat APM Studio PMML ondersteunt. Dat is een gestandaardiseerde manier om een machine learning model met een eenvoudige, geautomatiseerde stap te integreren in de operationele omgeving onafhankelijk van de analyse-omgeving waarin het model getraind en gevalideerd is. Er hoefde dus niets gewijzigd te worden aan de programmatuur.

*De uiteindelijke APM-studio
opstelling bij Airborne Composites
Automation.*



Stap 5 Gebruik

Nadat de operationele omgeving ingericht is en in gebruik is gesteld, kan de slimme applicatie daadwerkelijk gebruikt gaan worden. Hierbij is het met name van belang dat de operators die bij de machine staan ermee leren werken. Maar minstens zo belangrijk is om ook hun feedback te verzamelen. Werkt het voor hen? Wanneer wel en wanneer niet? Wat zou het gebruik voor hen makkelijker maken? Het gebeurt zelden dat een toepassing meteen de eerst keer raak is. En het gebeurt maar al te vaak dat de operators de nieuwe applicatie links laten liggen omdat ze er eerder last van hebben dan voordeel.

Bij de ingebruikname is het ook van belang om vast te stellen of het voor de business de verwachte waarde opbrengt. Soms is dat niet zo en valt het resultaat tegen. Dan is het van belang om te evalueren en te leren: wat hadden we kunnen doen om eerder te zien dat het niet zou werken? Maar ook als het wel succesvol is, moet daarbij stilgestaan worden. Want een succesvolle introductie van een slimme applicatie motiveert enorm voor het zetten van verdere stappen.

Op dit moment draait de voorbeeldapplicatie inderdaad bij Airborne.

Geïnteresseerden kunnen de applicatie op verzoek gaan bekijken.

Meer weten?

Dit whitepaper heeft in grote lijnen geschetst hoe de ontwikkeling van een slimme applicatie in z'n werk gaat. De stappen zijn in veel van dit soort trajecten min of meer hetzelfde al zal de invulling ervan de ene keer anders zijn dan de andere keer.

Wilt u meer weten over het ontwikkelen van een slimme applicatie voor uw onderneming, neem dan contact op met:

Mark Roest, VORtech

mark.roest@vortech.nl

06-4478 4413

Jules Oudmans, UReason

joudmans@ureason.com

06 185 565 50

Dankwoord

Het voorbeeldproject dat in dit whitepaper beschreven is, werd financieel ondersteund door het programma SMITZH-3, een samenwerkingsverband van TNO en Innovation Quarter, dat op haar beurt mede gefinancierd wordt door de provincie Zuid-Holland.