

Voorspellend onderhoud: Van theorie naar praktijk

Slim gebruik van data uit de productieprocessen gecombineerd met machine learning kunnen leiden tot predictive maintenance. Jules Oudmans van softwarebedrijf UReason legt uit in welke situaties het gebruik van data zinvol is, hoe je aan de slag kunt met datastromen en wat het kan opleveren. Van theorie naar praktijk.

Tekst **Evi Husson**

► “Minder storingen, geen ongeplande downtime, betere planning van werkzaamheden en onderhoud. ‘Slimme’ of ‘intelligente’ machines kunnen bijdragen aan het bereiken van deze doelen. Het gaat dan niet louter om het verzamelen van data die machines en processen genereren. ‘Intelligente systemen’ ontstaan op het moment dat een machine of een (computer)proces een afweging kan maken over wat de beste volgende actie is. Die progressie gaat verder op het moment dat er voldoende beschikbare data is om machine learning algoritmes te trainen om iets te kunnen voorspellen. Denk bijvoorbeeld aan kennis over hoe de kwaliteit van olie in een bepaalde motor of transformator zal evolueren op basis van historische data zodat tijdig het juiste onderhoud kan worden uitgevoerd.” Aan het woord is Jules Oudmans van UReason, een softwarebedrijf dat productiebedrijven helpt met het optimaliseren van het asset performance management systeem op basis van machine learning en kennismodellen.

Machine learning

“Machine learning kan helpen om te voorspellen wanneer onderhoud nodig is.” Even terug naar de basis. Machine learning, een vorm van artificiële intelligentie (AI) is een techniek die op basis van data algoritmen opbouwt die een nuttige uitspraak kunnen doen over die data. Het kan bijvoorbeeld bepaalde afwijkende waarden uit de data halen of aangeven dat een event staat te gebeuren.

Beter inplannen onderhoud

Oudmans: “In de industrie wordt supervised learning (zie kader, Subcategorieën van machine learning red.) regelmatig toegepast

met het doel om vroegtijdig fouten te detecteren om onverwachte storingen te voorkomen. Een voorbeeld. Wanneer een pomp faalt kan de productie stilvallen. Dit kan worden voorkomen door op basis van data uit het verleden modellen - zoals neurale netwerken - te trainen zodat al vroegtijdig afwijkingen kunnen worden gedetecteerd die zullen leiden tot falen. Denk aan de vorming van schalie op de pomp. Het bedrijf kan met deze informatie twee zaken doen: het kan onderhoud vervroegd inplannen om uitval te voorkomen of het kan een aantal parameters in het productieproces aanpassen zodat het moment van de verwachte uitval vooruit kan worden geschoven.”

Bepaal je doel en breng het gedetailleerd in kaart

Alles start dus met data. “Dat ligt iets genuanceerder. Je moet over voldoende betrouwbare historische data beschikken maar vooral ook een doel hebben wat je met die data wilt doen. Daar wringt in de praktijk regelmatig de schoen. Bedrijven verzamelen data, maar *what's next?*”

Stel dat je ongeplande uitval van bedrijfskritische machines wilt aanpakken. Belangrijk is dan om eerst een selectie aan faalmodi in kaart te brengen. De volgende stap is nagaan wat in het verleden de oorzaken zijn geweest van deze faalmodi en welke effecten ze hebben teweeggebracht. Daarna moet worden gekeken in welke data deze afwijkingen zijn te constateren. Op basis van de data en de functionele degradaties en falen kunnen modellen worden opgesteld die worden getraind om falen - nog voor de storing optreedt - te herkennen.”



Foto: credit

Van theorie naar praktijk

Oudmans legt de theorie opnieuw uit aan de hand van het voorbeeld van de pomp. “Een bedrijf wil voorkomen dat een bepaalde pomp van een bedrijfskritische machine stilvalt. De pomp kan meerdere faalmodi hebben, denk aan een slechte uitlijning van de as, een te lage druk of te weinig aanzuigkracht. Oorzaken die

“Je moet over voldoende betrouwbare historische data beschikken maar vooral ook een doel hebben wat je met die data wilt doen.”

in het verleden hieraan ten grondslag lagen zijn cavitatie, waaierslijtage, een verstopte aanzuigleiding, ventielen die niet volledig openen, enzovoort. Meetbare effecten, afhankelijk van de oorzaak zijn geluidstoename, meer trillingen, een lekkende afdichting, lagerslijtage, afname van de efficiëntie van de pomp of erosievorming op de waaier. De faalmodi, oorzaken en effecten zijn vaak af te leiden uit verschillende databronnen zoals de OEE database (Overall Equipment Effectiveness), de data die trillingsensoren genereren, het CMMS systeem (Computerized Maintenance Management Systeem) en andere historische procesdata.

Door (statistische) verbanden tussen verschillende gegevensverzamelingen te leggen, wordt duidelijk hoe data kan leiden tot nuttige informatie. Is dit helder, dan kan een model met algoritmes worden opgesteld dat het systeem aanleert om afwijkingen in een vroeg stadium te herkennen.”

Subcategorieën van machine learning

Er zijn verschillende soorten machine learning:

- 1) Supervised learning: Hierbij wordt aan een algoritme de verwachte uitkomst gegeven vanuit historische data. Het doel is om een model te creëren waarmee je op basis van historische data de gewenste output met goede precisie kunt classificeren.
- 2) Unsupervised learning: Hierbij worden geen resultaten aan het algoritme gegeven maar is het doel om een structuur te vinden in de input. Het systeem zoekt structuren in data.
- 3) Hybride of zelflerende vorm: Dit is semi-supervised learning, waarbij modellen lerend worden gemaakt. Door voldoende herhaling zal het systeem geleidelijk aan zelf patronen leren herkennen
- 4) Reinforcement learning: Hierbij wordt een beloning gegeven als het algoritme de juiste uitkomst heeft.



Maar hoe begin je?

Verbanden leggen tussen verschillende gegevensverzamelingen is niet eenvoudig. Daarom is in de loop der jaren een aantal standaard methodes opgesteld die daarbij kunnen helpen. Een ervan is het CRISP DM-model, wat staat voor Cross-Industry Standard Process for Data Mining. Deze methode wordt vaak gebruikt om heel gericht te zoeken naar (statistische) verbanden tussen verschillende gegevensverzamelingen. In de basis zijn daarbij zes stappen belangrijk: *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation* en *deployment*. (zie kader: In zes stappen van ruwe data naar inzichten)

Oudmans: "In de praktijk komt deze werkwijze erop neer dat je op basis van de data en kennis die vaak in het bedrijf beschikbaar is, kenbaar maakt wat het probleem is dat je wilt oplossen. Na selectie en voorbereiding van data wordt een model gevormd en getraind. Dit trainen gebeurt met data uit het verleden. Een bedrijf heeft in het verleden historische faaldata van een pomp verzameld. Uit de data zijn de functionele degradaties af te leiden. Het model dat is opgesteld wordt vervolgens gevoed met deze bekende dataset en wordt getraind om de verschillende faalmodi te herkennen. Vervolgens geven we aan hetzelfde algoritme een andere historische dataset die hij nog niet eerder heeft gezien zodat we kunnen nagaan of het model met een goede nauwkeurigheid een voorspelling kan doen van de faalmodi waarin we zijn geïnteresseerd."

Een positief resultaat is hier erg belangrijk. "Het model zou minimaal een nauwkeurigheid van tachtig procent moeten hebben. Is het percentage tachtig procent of lager, dan betekent dit dat

het model twee op tien keer (of vaker) een fout denkt te detecteren maar ernaast zit. In twee op tien gevallen is er dus sprake van een *false positive*. Is dit het geval, dan is aan te raden om vooralsnog de stap naar predictief onderhoud niet te maken, tot er voldoende nauwkeurigheid is."

Zijn de resultaten wel positief, dan kan worden gestart met de uitrol en worden verbindingen gemaakt met de realtime databronnen. Hierna volgt de evaluatie en finetuning.

"Is in twee op tien gevallen er sprake van een false positive, dan is het aan te raden om vooralsnog de stap naar predictief onderhoud niet te maken."

Kennis

Het bepalen van een concreet doel en betrouwbare data zijn belangrijk. Kennis is een derde aspect dat belangrijk is om succes te boeken. "Bedrijven die veronderstellen dat ze zelf geen kennis nodig hebben omdat machine learning-modellen



‘zelflerende’ systemen zijn, komen van een koude kermis thuis. Kennis is wel degelijk erg belangrijk. Heb je een hoge kennis en veel data van hoge kwaliteit, dan kun je goede beslissingen nemen op basis van de resultaten die de getrainde weergeven. Is je kennis en expertise minder goed, maar heb je wel veel kwalitatieve data, dan kun je nog steeds goede beslissingen nemen, maar de kwaliteit van de beslissingen is minder goed. Hetzelfde geldt voor veel kennis en een lage kwaliteit aan data.” Weinig kennis en weinig betrouwbare data staat gelijk aan een traject dat gedoemd is te mislukken. “In de praktijk zien we veel bedrijven die of te weinig data hebben of te weinig betrouwbare data. Dit komt wellicht omdat er op het gebied van faaldata of functionele degradatie nog geen strakke discipline is om dit op te slaan terwijl dit erg waardevol kan zijn. Daarom is het aan te bevelen om data rond degradaties nauwkeurig te verzamelen en documenteren. Op die manier zal de stap van preventief naar voorspellend onderhoud sneller kunnen worden gemaakt. ”



In 6 stappen van ruwe data naar inzichten

Het CRoss-Industry Standard Process for Data Mining-model wordt vaak gebruikt om inzichten te krijgen uit data. Ruwweg bestaat dit model uit zes stappen.

1) Begrijp je business

De eerste fase richt zich op de precieze beschrijving van het probleem of de projectdoelstellingen vanuit zakelijk en economisch perspectief. Er moet een projectplan voor het project worden ontwikkeld op basis van de concrete zakelijke en analytische doelstellingen. In het projectplan zijn onder meer de vereiste tijd, personeel en materiële middelen vermeld.

2) Begrijp je data

Na het formuleren van de analytische doelstellingen is de volgende stap het maken van een selectie van de relevante datasets. Deze fase draagt bij aan het identificeren van bestaande correlaties uit de gegevens. Ook wordt duidelijk of er mogelijke kwaliteitsgebreken in de gegevens zijn. De fase bestaat uit de volgende vier stappen: het verzamelen (1), beschrijven (2), analyseren (3) en evalueren (4) van de data. Er moet eerst worden gekeken of de data betrouwbaar kunnen worden verzameld. Welke bronnen zijn er beschikbaar? In welke mate dragen de data van deze bronnen bij aan het behalen van de vooropgestelde doelen? Is de data bruikbaar of moet deze op een andere manier worden gegenereerd? Problemen die zich voordoen tijdens het verzamelen, moeten worden gedocumenteerd of opgelost om mogelijke afwijkingen in de toekomst te voorkomen. De verzamelde data per databron moet vervolgens goed worden gedocumenteerd: data, tijd, formaat, hoeveelheid,... Tot slot moet worden bepaald of de datasets voldoende en bruikbaar zijn voor de analyse.

3) Voorbereiding data

Deze fase omvat alle activiteiten om de uiteindelijke dataset of datasetselectie te maken die in de modelleringssoftware wordt geladen voor analyse. De datakwaliteit en de technische voorwaarden spelen hierbij een grote rol. Aan het einde van dit proces moet duidelijk zijn welke gegevensreeksen in de analyse worden opgenomen of uitgesloten. Be-

langrijk daarbij is dat er een schone dataset wordt geselecteerd waarbij per datakolom dezelfde meeteenheid en datatypes voorkomen zodat alles klopt.

4) Modellering

Dit is vaak een erg uitdagende stap. Er moet een modelleertechniek worden geselecteerd waarmee het model kan worden gemaakt. Vaak wil men een bepaalde trend of bepaalde afwijkende eigenschappen van normaal gedrag kunnen waarnemen binnen bepaalde datasets die duiden op een specifieke faalmodus.

Om te controleren of het model geschikt is, zal aan de hand van een testmodel – waarmee ook wordt getraind – de kwaliteit en nauwkeurigheid worden nagegaan. Het is belangrijk dat foutpercentages laag zijn. Na het trainen en controleren van het testmodel volgt een modelvalidatie en een evaluatie.

5) Evaluatie

De resultaten die voortvloeien uit het testmodel worden vergeleken met de vooropgestelde doelstellingen. Bedrijven zullen in deze fase moeten beslissen of het model voldoende geschikt is om indien nodig verder te verfijnen en vervolgens te implementeren in de bedrijfsprocessen.

6) Uitrol

De uitrolfase is de fase die niet langer offline, maar online/in real time plaatsvindt. De opgedane kennis en geconstrueerde modellen zullen in de praktijk worden gebracht. Het opstellen van een implementatiestrategie, het monitoren van de resultaten van de modellen, het updaten en finetunen horen bij de uitrol.

Uit de praktijk blijkt dat veel tijd – vaak meer dan de helft van de tijd – nodig is voor de voorbereiding van de gegevens. Slechts 20 tot 30 procent van de geschatte tijd wordt besteed aan het bepalen van de relevante datasets. De rest van de tijd gaat op aan het bepalen van het bedrijfsvraagstuk, het meten van de prestaties en de implementatie van de gecreëerde modellen.