

Whitepaper

CRISP-DM

Sinds 2014 houdt KIM Plus Delta (KPD) zich bezig met data analyse en voorspellend onderhoud (Predictive Maintenance) voor de industrie en vastgoed.

Hiervoor hebben we een strategie ontwikkeld op basis van het CRISP DM model. Het data-analyseteam van KPD vertelt je graag over hun aanpak en ervaringen bij elke stap. Dit past perfect binnen de filosofie van ons bedrijf waarin: kennisdelen centraal staat.

In deze paper wordt een korte beschrijving gegeven van de mogelijkheden van de data mining methode CRISP-DM. Wij leggen je uit waarom wij binnen KIM Plus Delta de CRISP-DM methode toepassen om snel tot concrete resultaten te komen in voorspellend onderhoud.

Veel leesplezier,

Marco Wessels en Dirk Rösler
Partners KIM Plus Delta BV

KIM Plus Delta

Predictive Maintenance en de oneindige mogelijkheden van de datarevolutie

De boost in de beschikbaarheid van schaalbare computing heeft de afgelopen decennia geleid tot een gelijk opgaande groei van het gebruik van kunstmatige intelligentie (AI) en Machine Learning (ML). Een van de toepassingen waarbij ML-tools ontegenzeggelijk een grote belofte zijn, is Predictive Maintenance (PdM). Door PdM toe te passen, kan worden voorspeld wanneer een machine onderhoud nodig heeft. Hierdoor kan onvoorziene productiestilstand worden voorkomen, downtime van apparatuur worden verminderd en kunnen onderhoudsprogramma's worden geoptimaliseerd.

Predictive Maintenance: onontbeerlijk maar onbekend

Hoewel de potentiële voordelen van PdM alom worden erkend, hebben veel onderhoudsorganisaties problemen met het opzetten van een succesvol PdM-programma. Eén van de belangrijkste redenen hiervoor is dat de klassieke stakeholders in de industriële onderhoudssector niet voldoende bekend zijn met de nieuwe beschikbare technologieën op het gebied van data science. Onderhoudsmanagers weten vaak niet waar te beginnen met het opzetten van een PdM-programma. Er is doorgaans veel engineeringkennis ten aanzien van de werking van machines en hun faalmechanismen, maar het hele proces van data-acquisitie tot faalpatroonherkenning en -voorspelling is een geheel andere discipline.

Het is in de huidige tijd, waarin de druk op de marges steeds hoger wordt en efficiencyverbeteringen noodzakelijk zijn, echter onontbeerlijk om concrete stappen te maken en te investeren in PdM.

Ook voor machinebouwers (OEMs) wordt het steeds belangrijker om te kunnen voorspellen wanneer onderhoud aan hun installaties nodig is. De tijd dat een OEM zijn businessmodel kon invullen met overhauls en verkoop van spare parts, is voorbij. Afnemers vragen steeds vaker een garantie voor een hoge productiebetroouwbaarheid. En in deze sterk veranderende markt bewegen veel OEMs zich richting 'productie als een service', waarbij de beschikbaarheid van de machine centraal staat. Beter kunnen voorspellen wanneer een onderdeel kapot gaat, zorgt voor een economisch geoptimaliseerde service.

KIM Plus Delta

Een planmatige aanpak is noodzakelijk om het uiteindelijke doel te realiseren: meer inzicht in de staat van je machinepark, een efficiënter onderhoudsplan en betere afstemming tussen onderhoud en productie.

Data mining maakt goud van procesinformatie

Hoe je het ook bekijkt, PdM is inherent een zogenaamde Data Mining activiteit. Dat betekent dat het belangrijkste doel is om *data* om te zetten in nauwkeurige *informatie* over de gezondheidsstatus van assets. Hoewel het modieuze idee tegenwoordig bestaat dat data goud is, is dit even onzinnig als stellen dat aarde goud is. Alleen de juiste 'grondstof' zal, mits goed behandeld en bewerkt, tot iets van echte waarde leiden.

Machines geven potentieel een schat aan informatie. Met behulp van allerlei sensoren kunnen veel verschillende datapunten uit apparatuur worden verkregen. Temperatuur, voltages, druk of trillingen kunnen relatief gemakkelijk worden gemeten. Maar het hebben van data alleen is niet genoeg om een voorspellend onderhoudsprogramma op te baseren. Sterker nog, juist in deze tijd van data (of option) overload is het ontzettend belangrijk om heel gericht en gestructureerd te werk te gaan.

Veel vragen, even zoveel mogelijkheden

In een machinepark kunnen erg veel verschillende typen storingen plaatsvinden. Deze kunnen vaak op evenveel manieren worden gedetecteerd. Om een beeld te krijgen van wat de oorzaak van een storing is, zullen machines of het productieproces onder een vergrootglas moeten worden gehouden. Er zijn veel vragen die zich meteen al opdringen. Welke faalvormen kunnen worden voorspeld en welke wil men weten? Wat moet er gemeten worden? Hoeveel data is er nodig en met welke frequentie moet er gemeten worden? Welke software is hiervoor nodig? Hoe kunnen deze metingen worden gecombineerd of verwerkt? En tot slot: welk model wordt er toegepast voor de voorspelling? Het spectrum aan potentiële mogelijkheden is vaak overweldigend. Om het overzicht niet te verliezen en goede analyse van de onderhoudssituatie te maken is een goed gedefinieerd, stapsgewijs proces noodzakelijk. Gelukkig zijn er beproefde methodes beschikbaar die hierin als duidelijke routekaart gebruikt kunnen worden.

KIM Plus Delta

Stapsgewijs naar een antwoord

Het volgen van een stapsgewijze processtroom helpt bij het opzetten van een efficiënte transformatie van data naar informatie. Hierdoor wordt de focus op het einddoel niet uit het oog verloren. Het risico van data mining is vaak dat heel interessante wiskundige problemen worden opgelost, maar nutteloze modellen voor de startvraag worden gecreëerd.

De vraag is natuurlijk uit welke stappen zo'n processtroom bestaat en hoe die wordt opgezet. Gelukkig is het niet nodig om vanaf nul te beginnen. Er is een hele data-industrie die voorop loopt in de toepassing van data mining en de toepassing van Machine Learning. Voor de onderhoudsindustrie is data mining relatief nieuw, maar er kan worden gebouwd op beproefde methodes die in andere sectoren (informatica, bancaire sector, automotive, etc.) al breed worden toegepast.

CRISP-DM: Data mining met focus op business case

Een van de meest succesvolle data mining methodes is de *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Deze methode is in de jaren negentig ontwikkeld door een consortium bestaande uit onder andere IBM, SPSS, DaimlerChrysler en NCR. CRISP-DM geeft een duidelijk en uitgewerkt plan om een goede focus te krijgen op data mining in een business context. Eén van de grote, en in onze optiek doorslaggevende, voordelen van het CRISP-DM model ten opzichte van andere data mining methodes is de sterke focus op de beoogde toegevoegde waarde en business case. Met andere woorden: deze data mining methode begint bij het definiëren van een business case en helpt stapsgewijs om inzicht te krijgen in de betekenis van proces- en machinedata voor deze specifieke business case. Zo helpt CRISP-DM bij het nemen van business beslissingen. In tegenstelling tot andere methodes wordt eerst de doelstelling bepaald – wat moet worden bereikt? Een hogere productie? Minder ongeplande downtime? En waarom precies? Daarna wordt pas nagedacht over welke data hiervoor nodig is en hoe die gebruikt kan worden om de informatie te verkrijgen die het fundament vormt voor de business case. In Predictive Maintenance vertaalt dit zich naar de volgende concrete vragen:

- Welke storingen leveren de meeste problemen op en zorgen voor het meeste productieverlies?

KIM Plus Delta

- Hoe komen die storingen tot stand? Wat is het achterliggende faalmechanisme?
- Hoe kunnen deze faalmechanismen gedetecteerd worden?

De CRISP-DM methode volgt in principe het bekende en beproefde V-model uit de systems engineering discipline. Dit betekent dat het vanuit een globaal businessperspectief steeds dieper inzoomt op de echte technische ontwikkeling om vervolgens weer terug te gaan naar een evaluatie en implementatie op businessniveau.

Zes fases met concrete resultaten

Het CRISP-DM-model omvat een aantal duidelijk gedefinieerde stappen die helpen bij het verkrijgen van inzicht in het proces of machinepark. De methode bestaat uit zes opeenvolgende fases. Iedere fase is opgedeeld in een aantal taken die specifieke concrete resultaten opleveren. Door de continue referentie aan de business case is de kwaliteitsborging van het hele proces dus gebaseerd op de uiteindelijke toegevoegde waarde.

1. Begrijp de business

De eerste stap van dit data mining model is het bepalen van de business case. Wat wil men met deze exercitie bereiken? De doelen moeten heel concreet en kwantitatief worden gesteld. Als men bijvoorbeeld de stilstand van een machine wil verminderen, moet er eerst bepaald worden wat dat zou opleveren aan winst. De Business Case wordt opgebouwd door, bijvoorbeeld via FMECA (Failure Mode Effect & Criticality Analysis) of RCA (Root Cause Analysis), vanuit duidelijke KPI's in te zoomen op het daadwerkelijke probleem. Welke machine geeft de grootste problemen? Welke faalvorm is hierin dominant? Welk fysisch gedrag en onderdelen zijn hierbij betrokken? Een dergelijke analyse, die wordt opgenomen in het projectplan, zal leiden tot een duidelijke motivatie en eerste invulling van het gehele data mining proces.

2. Begrijp de data

Een van de resultaten uit stap 1 is het besluit welke gegevens uit het proces of de machine moeten worden verzameld. In de "Data Understanding" stap wordt een initiële dataset geformuleerd en bestudeerd op zowel zijn karakteristieken als inhoud. Bij karakteristieken kunnen we denken aan zaken als format, volume en

KIM Plus Delta

frequentie. Dit soort eigenschappen hebben verregaande invloed op het hele analysevraagstuk. Het inhoudelijke gedeelte in deze stap bestaat uit een eerste grafische en statistische analyse van de data. Deze stap is essentieel voor het bepalen van de richting van het uiteindelijke modelleren.

3. Data voorbereiden

De Data Preparation fase vormt het praktische hart van data mining. Het gaat er in deze fase om dat de data die uit het proces of de apparatuur wordt verkregen bruikbaar wordt gemaakt voor toepassing in de data mining modellen. In deze stap wordt beslist welk deel van de beschikbare data gebruikt wordt. Ook moet een consistente dataset worden gecreëerd en moeten uitschieters uit de dataset worden gefilterd. Soms kan het nodig zijn om verschillende datasets samen te voegen of het format van de data aan te passen. Zogenaamde data transformation technieken kunnen hier van onschatbare waarde zijn en het uiteindelijke model sterk versimpelen of inzichtelijker maken. De “Data Preparation” stap is in de praktijk één van de meest bewerkelijke activiteiten binnen de CRISP-DM methode.

4. Modelleren

In deze fase kan eindelijk begonnen worden met het modelleren van de data. Het data science gedeelte, dat over het algemeen het meest tot de verbeelding spreekt, komt dus redelijk laat in het hele proces. Op basis van de twee voorgaande fases is waarschijnlijk al gekozen welk model (of modellen) in aanmerking komt en nu moeten daarin verdere keuzes gemaakt worden. Het is belangrijk om de aannames, waarop deze modelkeuze gebaseerd is, goed in kaart te brengen. Uiteraard dienen alle gebruikelijke technieken te worden toegepast met trainingsdataset en testdatasets. De evaluatie die in deze fase gedaan moet worden, is vooral technisch van aard en kan leiden tot bijvoorbeeld hyper parameter tuning. Het model moet hier expliciet vanuit de data science beoordeeld en gevalideerd worden.

5. Evaluatie

Tijdens de evaluatiefase worden de verkregen modellen vanuit een businessperspectief onder de loep genomen. De eerdere stappen dienen nu geëvalueerd te worden met betrekking tot de initiële vraagstelling. Krijgt men

KIM Plus Delta

met dit model ook antwoord op de hoofdvraag? Leveren de modellen praktische resultaten op voor de vooraf gestelde business case? Zijn deze resultaten robuust? Het is het beste om de modellen ook in de praktijksituatie te testen. In deze fase worden de modellen, die voldoen aan alle vooraf gestelde criteria, geselecteerd. Ook worden de acties voor de volgende fase bepaald. Mocht de uitkomst van deze fase negatief zijn, dan moet men terug naar de 'tekentafel'. De eerdere stappen moeten dan worden doorgelicht om te achterhalen waar het proces is gaan afwijken van de oorspronkelijke doelstelling.

6. Gebruik

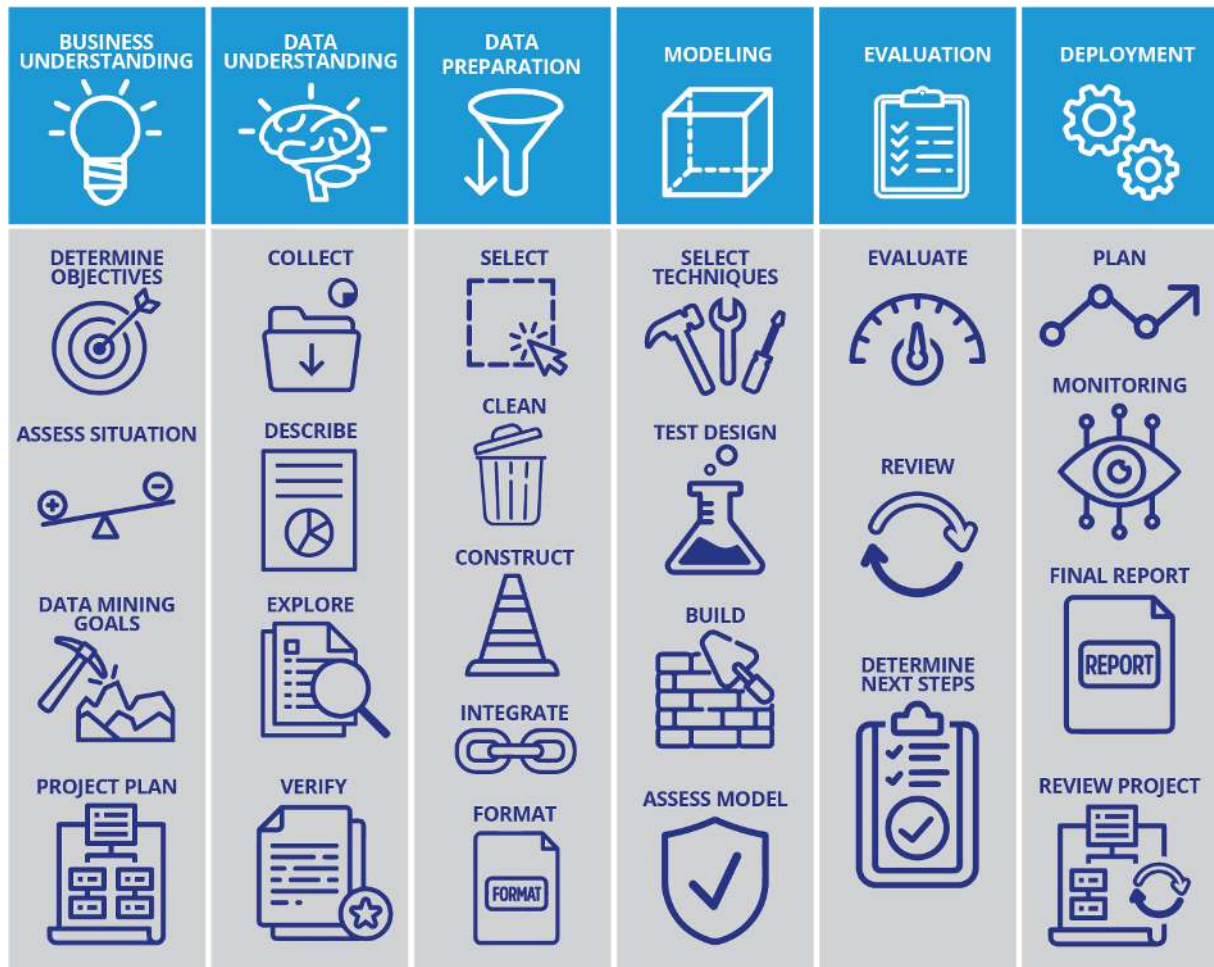
Als in de evaluatiefase de modellen goedgekeurd zijn kan de daadwerkelijke implementatie plaatsvinden. Het is belangrijk dat de resultaten ook daadwerkelijk gebruikt worden, anders is alle moeite voor niks geweest. Een model maken om in een la te leggen, daar heeft niemand iets aan.

In het voorbeeld van het reduceren van ongeplande downtime werd bijvoorbeeld gevonden dat een bepaalde aandrijving de grootste problemen oplevert. Het bleek verder dat bepaalde frequentiekenmerken in de stroom de kop op steken ruim voordat hij daadwerkelijk functioneel faalt. De stroomdata van deze aandrijving kan gemakkelijk worden geregistreerd en een model voor het spectrale karakter kan ook opgezet worden. Als dit model goed gebruikt wordt gaan alarmbellen rinkelen als de aandrijving in zijn faalregime komt. De aandrijving kan zo tijdig worden vervangen en de productie downtime worden verminderd.

Continu schaven

Data mining is een proces van continu schaven en bijstellen. Er moet worden geschakeld tussen de verschillende fases om resultaten te staven en eventuele fouten uit het proces te halen. Het is ook belangrijk om te onthouden dat data mining een cyclische activiteit is. Het is verstandig om de verschillende stappen periodiek te herzien. Als er aanpassingen in een machine worden gedaan of een productielijn wordt aangepast, werkt het model misschien niet meer optimaal.

KIM Plus Delta



Op deze manier leidt CRISP-DM de gebruiker vanaf de initiële business case door het gehele data mining proces heen. Het is taak om consequent de verschillende stappen nauwkeurig uit te voeren. Als data inzichtelijk en behapbaar worden, is voor vrijwel iedere PdM kwestie een oplossing te vinden

Multidisciplinair team is een must

In het licht van alle CRISP-DM stappen en activiteiten willen we de lezer wijzen op een vaak onderbelicht punt van PdM: het brede spectrum van een reële implementatie. Predictive Maintenance raakt aan vele disciplines. De oorsprong van onderhoud ligt natuurlijk bij maintenance engineers en managers (zogenaamde domeinkennis experts) maar het is duidelijk dat er nieuwe rollen en functies nodig zijn om PdM te gaan toepassen. De meest voor de hand liggende rol is natuurlijk die van data scientist. Een andere cruciale rol is die van data engineer, i.e., de verantwoordelijke voor alle datalogistiek binnen de organisatie. Deze nieuwe werknemers moeten plaatsnemen in multidisciplinaire teams met de meer traditionele engineers en gezamenlijk zullen ze

KIM Plus Delta

verantwoordelijk zijn voor het succes van PdM. In een ideale situatie worden experts uit verschillende disciplines betrokken bij het maken van de vertaling van data naar 'actionable information'. Operators en monteurs hebben technische kennis van de apparatuur en zien op welke punten deze vaak faalt. Ook hebben ze inzicht in de specifieke faalmechanismen. Een elektrotechnisch engineer kan voorstellen doen om sensoren te plaatsen of data te ontsluiten uit PLC systemen. Een data engineer zal de data uit de sensoren vervolgens in een database of data lake opslaan en uiteindelijk kan een data scientist zich dan buigen over de data-specifieke analyse en het echte modelleren. Pas na dit gehele traject kan de onderhoudsmanager tezamen met dit team beslissen over de te nemen acties.

We kunnen dit vertalen naar het V-model van CRISP-DM: De Business Case en Data Understanding fase zullen beide veel input vergen van domeinkennis stakeholders. De technische deep dive van Data Transformation en Modelling zal meer een pure data science aangelegenheid zijn waar de data scientist wat geïsoleerder aan werkt. In de Evaluation en Deployment fase komen alle disciplines weer samen om alles binnen de business doelstelling te beoordelen en uiteindelijk te implementeren.

De realiteit is dat de hierboven beschreven multidisciplinariteit lang niet altijd aanwezig is en het zal duidelijk zijn dat er een grote commitment nodig is vanuit hoger management om deze organisatie op te zetten en alle partijen klaar te stomen voor deze nieuwe onderhoudsmethode.