Predictive maintenance voor actieve koolfiltervaten in biogasopwerkingssystemen

Afstudeeropdracht Lars van den Broek bij Bright

Concept afstudeerverslag
05-04-2023

L.A. van den Broek
428097
HBO-ICT Business
Saxion Hogescholen
Begeleider: Martin Wesselink

Voorwoord

Voor u ligt het afstudeerverslag “Predictive maintenance voor actieve koolfiltervaten in biogasopwerkingssystemen”. Dit afstudeerverslag is geschreven voor het afstuderen van de opleiding HBO-ICT Business aan Saxion University of Applied Science in Enschede. Dit verslag is het resultaat van het onderzoek dat ik sinds mei 2022 heb mogen doen bij Bright Renewables.

Deze afgelopen maanden zijn voor mij enorm leerzaam geweest. Ik ben veel te weten gekomen over de wereld van bio-energie, met in het bijzonder de opwerking van biogas tot biomethaan. Kennis die niks met mijn opleiding te maken heeft, maar die ik bijster interessant vind. Daarnaast heb ik ook veel geleerd over mijzelf en mijn manier van werken (en hoe dat te verbeteren). Het is ook een zware tijd geweest waarin ik een aantal weken door een coronabesmetting weinig tot niks heb kunnen doen, wat er ook toe heeft geleid dat ik dit verslag later kan presenteren dan dat ik had gewild. Al met al heeft deze tijd veel ups en downs gehad, maar ik kan met trots terugkijken op de werkzaamheden die ik heb verricht.

Natuurlijk heb ik dit niet alleen gedaan, en daarom zou ik graag van dit voorwoord gebruik willen maken om een aantal mensen te bedanken. Als eerste mijn begeleider vanuit Saxion, Martin Wesselink. Heel erg bedankt voor alle goede adviezen en begeleiding, vooral in de tijd dat ik ziek was. Wanneer ik in de stress zat, wist je me gerust te stellen en voor mij uit te zoeken wat mijn mogelijkheden waren voor uitstel vanwege mijn ziekte. Dat heeft vele zorgen weggenomen zodat ik mij kon focussen op beter worden.

Vervolgens wil ik ook graag mijn begeleider vanuit Bright bedanken, Vincent Kroeze. Vincent, je was een geweldige begeleider. Je enthousiasme en leergierigheid waren aanstekelijk, en ik heb sinds de middelbare school van niemand zoveel scheikundige lessen gehad. Je wist me altijd te motiveren om alles net wat dieper uit te zoeken, maar je wist me ook bij de les te houden wanneer ik te geïnteresseerd was in bijzaken.

Ook wil ik mijn andere collega’s binnen HoSt en Bright bedanken, met name de mensen met wie ik de afgelopen maanden een kantoor heb gedeeld. Bart, Bram, Cas, Erik, Niels, Remco, Simon, Sjoerd, Swen, Vivek, ik heb genoten van onze serieuze en minder serieuze momenten en ik keek elke dag uit naar onze wandeling over de campus.

Verder zijn er nog vrienden en familie die mij hebben ondersteund, van het nakijken van mijn verslagen op spelling en inhoudelijke input, tot het puur ondersteunen wanneer ik dat nodig had, opnieuw in het bijzonder tijdens de periode waarin ik ziek was. Annemieke, Juliane, Lotte, Petra, Reinier, Rhi, Theo, Wouter, heel erg bedankt voor al jullie hulp en steun.

Ik wens jullie veel leesplezier toe.

Lars van den Broek

Enschede, 26 maart 2023

Managementsamenvatting

Bright Renewables is een bedrijf dat actief is in de bio-energie-industrie. Een van de producten die het verkoopt zijn biogasopwerkingsinstallaties. Deze installaties hebben veel onderhoud nodig, waaronder het vervangen van actieve koolfiltervaten, die gebruikt worden om waterstofsulfide uit de biogas te filteren. Bright heeft de ambitie om het onderhoud via een predictive maintenance strategie te laten verlopen. Dit onderzoek heeft gekeken naar de mogelijkheid om door middel van data-analyses een predictive maintenance strategie op te stellen voor het onderhoud van de actieve koolfiltervaten in deze installaties.

Dit onderzoek heeft gebruik gemaakt van de CRISP-DM methode, met als doel om met behulp van de Orange software te voorspellen wanneer onderhoud nodig is. Hiervoor werd gebruik gemaakt van de machine learning modellen die beschikbaar waren in Orange.

Uit dit onderzoek is gebleken dat de voorspellingen, die gedaan worden door Orange, niet extreem nauwkeurig zijn, maar met behulp van menselijke interpretatie is het wel mogelijk om redelijke schattingen te maken, waardoor onderhoud ongeveer een week eerder kan worden ingepland dan dat nu het geval is. Er zitten wel significante foutmarges in dit model, waardoor het altijd nodig is om de voorspellingen kritisch te beoordelen.

De conclusie is daarom dat er voorzichtige voorspellingen kunnen worden gemaakt met behulp van een voorspellingsmodel binnen Orange, maar dat het van belang is om meer onderzoek te doen om te zorgen dat deze voorspellingen nauwkeuriger worden.

Inhoud

Figuren en tabellen lijst

Figuren

Tabellen

Begrippen, definities en afkortingen

Actief Koolfiltervat – Een vat gevuld met geactiveerd kool, dat gebruikt wordt om waterstofsulfide (H_2S) uit het biogas te filteren. De waterstofsulfide bindt aan het kool. Na verloop van tijd zal het kool verzadigd zijn, en zal het vat moeten worden vervangen.

API – Application Programming Interface – Een geformaliseerde manier om gegevens uit te wisselen tussen apps. In dit onderzoek heeft Ixon een programma met een API. Die API kan worden gebruikt om gegevens op te vragen.

ADLS Gen2 – Azure Data Lake Storage Gen2 – De tweede generatie van een Data Lake gemaakt door Microsoft Azure.

Biogasflow – De hoeveelheid biogas die op dat moment door de installatie stroomt, gemeten in Normaal kuub per seconde (Nm^3/s). Een Normaal kuub is de hoeveelheid kubieke meter die het gas had gehad bij een temperatuur van 273 K en een luchtdruk van 101,3 kPa; dit omdat de werkelijke dichtheid van gas sterk varieert afhankelijk van de temperatuur en de druk.

Biogasopwerking – Het purificeren het biogas. Met andere woorden, de gassen die niet methaan (CH_4) zijn uit het biogas filteren. Het doel is om het gas hiermee naar 99,5% methaan krijgen, waardoor het biomethaan mag worden genoemd en mag worden geïnjecteerd in het nationale gasnetwerk.

Classificatievragen – Vragen in een decision tree die bepalen welke tak er moet worden gevolgd voor de classificatie, bijvoorbeeld “Heeft attribuut A een waarde van 5 of hoger”. Na een flink aantal van dit soort vragen kom je uit op een classificatie.

Component-based – Programmeerbare software die niet met programmeertaal werkt, maar met voorgeprogrammeerde bouwstenen (components) die visueel aan elkaar kunnen worden gekoppeld.

CSV-bestand – Comma Seperated Value, een dataformat waarbij data van elkaar gescheiden wordt met een specifiek teken, vaak een komma, maar in Nederlandse datasets ook vaak een puntkomma, omdat onze decimalen met komma’s worden geschreven. CSV-bestanden zijn vaak eenvoudig in te lezen en nemen weinig ruimte in beslag.

Datawarehouses – Een vorm van data-opslag waarbij er zeer specifieke regels gelden om te zorgen dat data snel inzichtelijk is zonder dat het te veel opslag inneemt. Het nadeel hiervan is wel dat deze vorm van opslag minder flexibel is.

Doorslag – Wanneer een actief koolfiltervat vol begint te raken zal niet alle H_2S worden opgenomen. Daardoor zal een deel van de H_2S in het gas blijven. Dit wordt gemeten door de H_2S meter na het eerste filtervat. Wanneer deze meting meer dan 5ppm is (daaronder zou het een meetfout kunnen zijn) wordt dat een doorslag genoemd. Wanneer deze meting meer dan 100ppm is wordt dat een kritieke doorslag genoemd.

ERP-systeem – Enterprise resource planning systeem is software die wordt gebruikt binnen organisaties om alle processen binnen het bedrijf te ondersteunen. Het is een geavanceerd administratie systeem.

IoT – Internet of Things, apparaten die direct vanaf het internet kunnen worden aangesproken. IoT is vooral bekend van de smart lampen, maar in dit onderzoek gaat het om de installaties van Bright en de sensoren. Deze zijn ook direct van het internet aan te roepen.

Ixon – Bedrijf dat IoT oplossingen biedt. Ixon is de leverancier van Bright voor de sensoren in de installaties.

JSON-format – Java Script Object Notation, een dataformat waarbij data die met elkaar te maken heeft wordt geclusterd met elkaar (een zogenoemde Object).

MAE – Mean Absolute Error – De gemiddelde absolute afwijking. Een manier om te meten hoe betrouwbaar de voorspellingen zijn. Des te lager de MAE, des te beter. Een perfecte voorspelling zou een MAE van 0 hebben.

Membranen – Dunne vlakke structuur die twee ruimte van elkaar scheidt. Door Bright worden membranen gebruikt als filter.

MSE – Mean Square Error – Het gemiddelde van de tweede macht van de afwijking. MSE is een manier om te meten hoe betrouwbaar de voorspellingen zijn. Ook hier is lager beter, en een perfecte voorspelling zal een MSE van 0 hebben.

PLC – Programmable logic controller, een elektronisch apparaat met een microprocessor. In feite een kleine computer.

Powershell – Een oplossing van Microsoft die gebruikt wordt voor taakautomatisering, bestaande uit een scripttaal, een opdrachtregelshell en een framework voor configuratiebeheer dat werkt op Windows, Linux en macOS.

R2 – De determinatiecoëfficiënt. Dit duidt aan in welke mate een statistisch model een bepaalde uitkomst kan voorspellen. Hoe hoger de R2 des te beter het model de uitkomst voorspelt. Een R2 van 1 geeft een perfecte voorspelling weer.

RMSE – Rooted Mean Squared Error – De wortel van het gemiddelde van de tweede macht van de afwijking. De RMSE is een manier om te meten hoe betrouwbaar de voorspellingen zijn. Ook hier is lager beter, en een perfecte voorspelling zal een RMSE van 0 hebben.

SCADA-systeem – Supervisory Control And Data Acquisition – Een SCADA systeem bestaat uit een computer met daarop software voor het verzamelen, doorsturen, verwerken en visualiseren van meet- en regelsignalen.

VBA – Visual Basic for Applications – Een programmeertaal gebaseerd op Visual Basic, ontworpen om te gebruiken in de Office-applicaties van Microsoft.

XLSM-bestand – Een Excelbestand dat macro's geschreven in VBA bevat.

XLSX-bestand – Een Excelbestand zonder macro's.

1. Inleiding

1.1. Achtergrond

1.1.1. HoSt en Bright

De opdrachtgever is Bright Services, onderdeel van HoSt Holding. HoSt is een van de grootste spelers in de wereld op het gebied van bio-energietechnologieën. Bright Services is verantwoordelijk voor de service en het onderhoud van de biogasinstallaties van HoSt. Binnen Bright en HoSt zijn er een aantal medewerkers die specifiek belang hebben bij de resultaten van deze stage. Dit zijn Raymond Enkt, COO van HoSt Holding, Martijn Hiddink, Teamleader Mechanical design and ICT, en Vincent Kroeze, Proces Engineer. Raymond Enkt is verantwoordelijk voor alle bedrijfsprocessen binnen HoSt Holding. Hij is degene die deze opdracht heeft opgezet en hij is de product owner van dit project. Martijn Hiddink zal als Teamleader ICT verantwoordelijk zijn voor de opgeleverde producten na afloop van dit project. Het is van belang dat hij alle programma's begrijpt en daar eventuele latere aanpassingen kan doorvoeren. Verder is hij ook de persoon die kan assisteren met het integreren van de producten in de huidige ICT-omgeving van HoSt. Vincent Kroeze is proces engineer en is degene die momenteel de Proactive Maintenance Strategy in gang zet. Hij werkt onder andere met de IXON cloud en de data om pro actief apparatuur aan te merken voor onderhoud. Vincent heeft veel kennis over de context achter de data en kan daar dan ook in begeleiden. Vincent zal ook degene zijn wiens werkproces het meest zal veranderen door de implementatie van de op te leveren producten.

1.1.1.1 Organisatieomschrijving

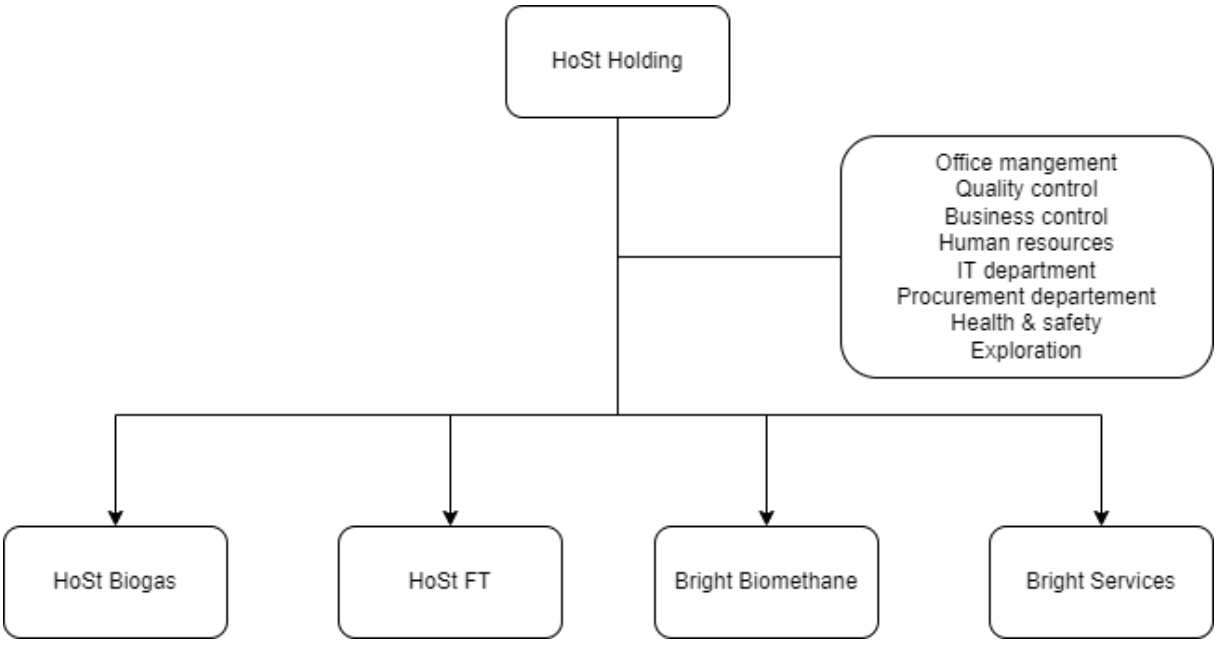
HoSt (Holec Stork) Holding is opgericht in 1991 en biedt totaaloplossingen voor bio-energiesystemen. Sinds 1999 is HoSt een zelfstandig bedrijf. HoSt draagt bij aan het succes van een circulaire economie door het produceren van duurzame energie, het oplossen van het afvalprobleem en het creëren van waardevolle eindproducten uit organisch afval. Dit doen ze door middel van hun biomassa-installaties. (Twilhaar, 2021) (HoSt Holding, 2022)

HoSt is de afgelopen jaren enorm gegroeid door zijn succes. In 2010 waren er totaal 20 mensen werkzaam binnen HoSt, maar ten tijden van het schrijven zijn er meer dan 240 medewerkers actief binnen HoSt in zowel binnen- als buitenland. De omzet in 2020 was circa 70 miljoen euro (Twilhaar, 2021).

HoSt bestaat uit een viertal bedrijven, te weten:

- HoSt Biogas — Ontwikkelt en bouwt biogasinstallaties
- HoSt FT — Ontwikkelt en bouwt biomassaketels
- Bright Biomethane — Ontwikkelt en bouwt gasopwerkingen
- Bright Services — Onderhoudt bovengenoemde installaties

In figuur 1 wordt een versimpeld organogram weergegeven waarin de organisatiestructuur van HoSt Holding en haar onderliggende B.V.'s wordt verduidelijkt.



Figuur 1 Versimpeld organogram HoSt

Hierin is te zien dat, hoewel alle B.V.'s apart opereren, vrijwel alle mogelijke staffuncties centraal worden geregeld. Hieronder valt ook de IT-afdeling, wat betekent dat alle oplossingen uiteindelijk grotendeels door HoSt Holding zullen worden onderhouden, ondanks dat de opdracht specifiek voor Bright Services wordt uitgevoerd.

1.1.1.2. Kernactiviteiten

Bright Services heeft als kernactiviteit het leveren van service en onderhoud aan de biogasinstallaties van HoSt Biogas en Bright Biomethane. Deze installaties wekken biogas op uit afvalstromen van andere bedrijven, zoals mest van boerderijvee. Om deze installaties werkend te houden, is er regelmatig onderhoud nodig. Zo moeten bepaalde onderdelen, zoals actieve koolfiltervaten, met enige regelmaat worden vervangen. Ook kan het voorkomen dat er acute problemen ontstaan die zo snel mogelijk moeten worden opgelost. Om dit onderhoud te ondersteunen zijn er werkvoorbereiders en proces engineers die de monteurs aansturen.

1.1.1.3. Situatie voorafgaand aan het onderzoek

Op dit moment is HoSt bezig met een transitie in de manier waarop interne data-opslag. Momenteel wordt alles nog met behulp van Exceldocumenten en VBA-scripts bijgehouden. Dit was een prima oplossing toen HoSt nog een klein bedrijf was, maar in de afgelopen jaren is HoSt zodanig gegroeid dat er nu zo'n tachtig programma's (voornamelijk Excel) worden gebruikt om het bedrijf te laten functioneren (Twilhaar, 2021).

Deze programma's zijn veelal gemaakt wanneer ze nodig waren en aangepast wanneer dat nodig was, vaak door verschillende mensen. Hierdoor is het overzicht vaak kwijt en zijn de programma's ook zeer foutgevoelig. Momenteel wordt het onderhoud van deze programma's uitgevoerd door een extern bedrijf: Fute.

Vanwege deze problemen, veroorzaakt door de enorme groei, is HoSt momenteel bezig met het implementeren van een ERP-systeem (Microsoft Dynamics 365). Voor deze implementatie hebben ze de externe partij 4PS ingeschakeld om te assisteren.

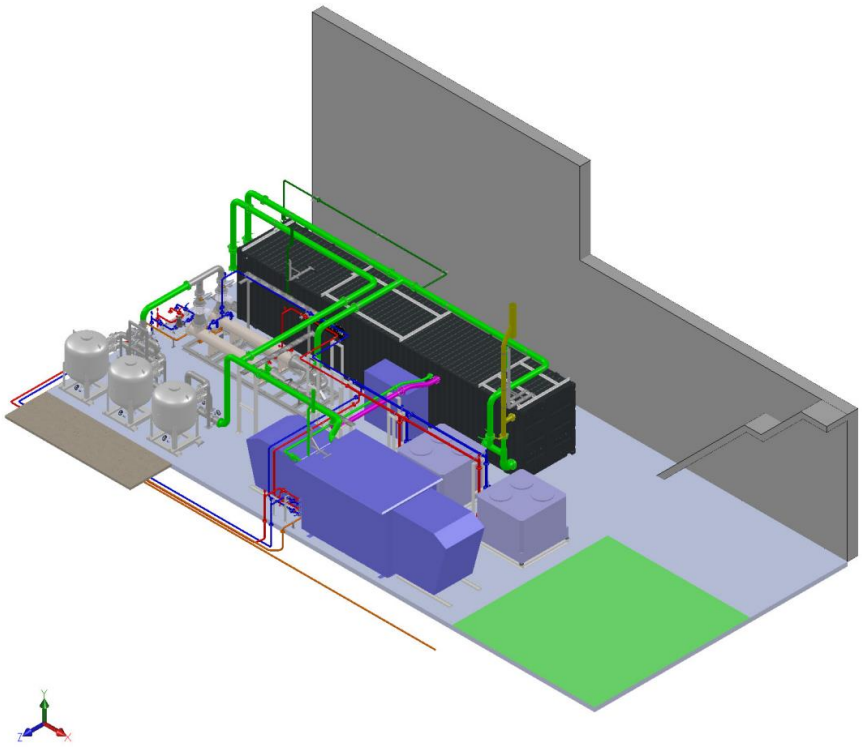
1.1.2. Biogasopwerkingsinstallatie

Een biogasopwerkingsinstallatie is een belangrijke component van een biogasketen. Het is een systeem dat biogas, dat wordt geproduceerd uit organisch materiaal, opwerkt tot een hogere kwaliteit, zodat het geschikt is voor gebruik als brandstof. Dit opgewerkte biogas kan worden gebruikt voor verschillende doeleinden, zoals verwarming, elektriciteitsopwekking of transport. Een biogasketen bestaat uit verschillende stappen, waaronder de aanvoer van organisch materiaal, de anaerobe vergisting, en de opwerking van het geproduceerde biogas tot biomethaan.

De rol van een biogasopwerkingsinstallatie in een biogasketen is om het geproduceerde biogas op te werken en te ontdoen van koolstofdioxide (CO_2) en andere verontreinigingen, zoals zwavelzuur (H_2S). Dit verhoogt het gehalte aan methaan (CH_4) tot ongeveer 89%, waardoor het biogas geschikt wordt voor gebruik als hernieuwbare energiebron en in het Nederlandse gasnet mag worden geïnjecteerd. De eisen van het percentage methaan kan verschillen per land.

Een van de belangrijkste onderdelen van een biogasopwerkingsinstallatie zijn de actieve koolfiltervaten. Deze filtervaten spelen een belangrijke rol bij het zuiveren van het biogas door het te ontdoen van verontreinigingen en geuren. Actieve kool is een poreus materiaal dat grote oppervlakken biedt voor adsorptie van verontreinigende stoffen.

Het onderhoud van een biogasopwerkingsinstallatie is van groot belang om een efficiënte werking van het systeem te garanderen. Het meest voorkomende onderhoud is het vervangen van de actieve koolfiltervaten, wat periodiek moet gebeuren. Dit is nodig omdat de actieve kool na verloop van tijd verzadigd raakt en niet meer in staat is om verontreinigingen uit het biogas te adsorberen. Het vervangen van de actieve kool is essentieel om de efficiëntie van het zuiveringsproces te waarborgen en om de levensduur van de biogasopwerkingsinstallatie te verlengen. Daarnaast is het belangrijk om regelmatig de filters te inspecteren op eventuele beschadigingen en om de drukverschillen over de filters te meten. Een te hoog drukverschil kan namelijk duiden op een verstopping van de filters, wat kan leiden tot een vermindering van de efficiëntie van het zuiveringsproces. (Bright Renewables, 2023) (Kroeze, Werking Biogasopwerkingsinstallatie, 2022)



Figuur 2 3D-tekening van een biogasopwerkingsinstallatie

In figuur 2 is een 3D-tekening van een biogasopwerkingsinstallatie te zien. Voor dit onderzoek zijn de koolfiltervaten (de drie grijze silo's links op de afbeelding) het belangrijkste onderdeel.

1.2. Onderzoek

1.2.1. Probleemstelling

De biogasinstallaties van Bright zijn complexe apparaten die veel onderhoud vereisen tijdens hun levensduur. Dit onderhoud wordt op dit moment voornamelijk reactive (wanneer de klant belt) en preventive gepland, bijvoorbeeld de compressor die, volgens voorschriften van de producent van dit onderdeel, elke 2000 uren moet worden onderhouden. Daarnaast wordt een aantal Key Performance Indicators (KPI's) nauwlettend in de gaten gehouden, zodat, wanneer die een grenswaarde bereiken, extra onderhoud kan worden ingepland. In het geval van actieve koolfiltervaten wordt er een KPI in de gaten gehouden die aangeeft hoeveel H_2S wordt gemeten na het eerste actieve koolfiltervat. Dit zou onder normale omstandigheden 0ppm moeten zijn, maar wanneer het eerste actieve koolfiltervat vol begint te raken zal deze meting boven de 5ppm uitkomen (tussen 0 en 5 kan er sprake zijn van een meetfout). Dit wordt 'doorslag' genoemd. Hoewel dit op het moment wel werkt, is dit geen efficiënte manier van handelen. Het geplande onderhoud vindt soms te vroeg plaats, waardoor niet het maximale uit alle apparatuur wordt gehaald, of het gebeurt te laat, waardoor de installatie niet altijd op volle kracht kan draaien. Het monitoren en interpreteren van de KPI's is ook tijdsintensief, en wanneer er grenswaardes worden bereikt, kan het voorkomen dat er geen onderhoudsmonteur op tijd beschikbaar is.

1.2.2. Doelstelling

Bright is op zoek naar een manier om het onderhoud van de actieve koolfiltervaten op een voorspellende manier plaats te laten vinden (predictive maintenance). Hierbij kan op basis van de data vooraf, en geautomatiseerd worden voorspeld wanneer er onderhoud nodig is. Deze voorspelling moet zodanig nauwkeurig zijn ,dat de melding niet te vroeg of te laat wordt gedaan, om zo het uiterste

uit de apparatuur te halen. Hiermee hoopt Bright de beschikbaarheid van de installaties te verhogen, de kosten van het onderhoud te verlagen en de werkdruk van de monteurs te verlagen.

1.2.3. Hoofdvraag

De hoofdvraag van dit project luidt als volgt:

“Hoe kan Bright op basis van data-analyses een predictive maintenance model opzetten voor de actieve koolfiltervaten?”

1.2.4. Deelvragen

Uit bovenstaande hoofdvraag volgen meerdere deelvragen:

1. “Hoe ziet het huidige werkproces van het onderhouden van de actieve koolfiltervaten eruit?”
2. “Welke data die in de installatie bij klanten wordt gemeten, is relevant voor het maken van voorspellingen voor het onderhoud aan de actieve koolfiltervaten?”
3. “Hoe kan de data die in de installaties bij klanten wordt gemeten, worden geëxtraheerd naar een data-opslag van Bright?”
4. “Hoe kan er een voorspellingsmodel worden gemaakt dat in staat is om het onderhoud van de actieve koolfiltervaten nauwkeuriger te voorspellen dan met de huidige methode?”
    a. “Welke variabelen kunnen het beste worden gebruikt voor dit model?”
    b. “Welke algoritmes zijn hiervoor het meest geschikt?”
5. “Hoe kunnen medewerkers op de voor hen ideale manier op de hoogte worden gesteld van deze voorspellingen?”
6. “Op welke manier moet het bedrijfsproces voor het plegen van onderhoud opnieuw worden ingericht om het in lijn te brengen met de strategieveranderingen van Bright?”
7. “Hoe kan ervoor worden gezorgd dat de medewerkers, die betrokken zijn bij dit bedrijfsproces, het nieuwe bedrijfsproces gaan volgen?”

1.2.5. Onderzoekopzet

In dit onderzoek zal gebruik worden gemaakt van het DOT-Framework (ICT research methods, 2022). Daarnaast zal voor de data-analyse gebruik worden gemaakt van de CRISP-DM methode. Voor meer uitleg over deze onderzoeksmethodes zie hoofdstuk 2.2.1. en 2.2.2..

2. Theoretisch kader

In dit hoofdstuk worden de kernbegrippen, theorieën en modellen besproken die relevant zijn voor het onderzoek.

2.1. Kernbegrippen

2.1.1. Predictive maintenance

Predictive maintenance is een vorm van onderhoud waarbij data-analyse en machine learning algoritmes worden gebruikt om te voorspellen wanneer er onderhoud nodig is. Deze voorspellingen worden gemaakt op basis van de gegevens die continu worden verzameld van de machine, zoals gassamenstelling, temperatuurmetingen en andere sensorgegevens. Door deze gegevens te analyseren kan het onderhoudsteam problemen detecteren voordat ze zich voordoen, en reparaties uitvoeren voordat de machine uitvalt.

Andere soorten onderhoud zijn reactive maintenance, planned maintenance en proactive maintenance. Reactive maintenance is het repareren van een machine nadat deze is uitgevallen, waardoor er vaak hoge kosten zijn en productieverlies optreedt. Planned maintenance is het uitvoeren van gepland onderhoud op vaste tijdsintervallen, ongeacht de staat van de machine. Proactive maintenance, ook wel bekend als online monitoring, is het monitoren van de prestaties van de machine om de gezondheid ervan te beoordelen en eventuele problemen te identificeren (Carvalho, et al., 2019).

2.1.2. Datamining

Datamining is een subdomein van kunstmatige intelligentie en kan worden gedefinieerd als een proces met als doel kennis te genereren uit data en deze op een begrijpelijke manier presenteren aan de gebruiker. Het genereren van kennis in de context van datamining kan worden vertaald als het ontdekken van nieuwe niet-triviale patronen, relaties en trends in de data, die nuttig zijn voor de gebruiker. Onder het dataminingproces vallen het verzamelen en selecteren van data, het voorbewerken van de data, de data-analyse zelf - inclusief de visualisatie van de resultaten, het interpreteren van de resultaten - en de toepassing van de nieuwe kennis. Voor de voorbewerking en de analyse van de data worden machine learning en statistische methodes gebruikt. De bevindingen uit datamining vallen onder twee categorieën: beschrijvend, waarbij patronen en relaties in data worden beschreven, en voorspellend, waarbij voorspellingen worden gemaakt op basis de relaties en trends van de data (Schuh, et al., 2019).

2.1.3. Data lake

Een data lake is een schaalbare opslaglocatie die een grote hoeveelheid ruwe data in hun originele format bewaart voor later gebruik. Daarnaast heeft een data lake verwerkingssystemen die gegevens kunnen verwerken zonder de gegevensstructuur aan te tasten. Data lakes worden doorgaans gebouwd om grote hoeveelheden ongestructureerde data te verwerken, waaruit verdere inzichten worden afgeleid. De data lakes gebruiken dus dynamische analytische toepassingen. De gegevens in de data lake worden direct toegankelijk zodra ze zijn gemaakt, in tegenstelling tot data warehouses waar de data zeer gestructureerd is, met statische analytische toepassingen en met minder snelle toegang tot de gegevens. Data lakes zijn vooral nuttig voor snel veranderende data (Miloslavskaya & Tolstoy, 2016).

2.1.4. ETL vs. ELT

ETL staat voor Extract, Transform, Load. Dit is de standaard manier van het inladen van data, waarbij de data eerst wordt geëxtraheerd van de bron en vervolgens wordt omgevormd zodat hij past in het

dataopslagmodel van de database. Vervolgens wordt de data in de database geladen, zodat hij gebruikt kan worden voor data-analyse.

ELT staat voor Extract, Load, Transform. Dit is een nieuwe manier van het inladen van data die vaker wordt gebruikt voor big data in data lakes. Hierbij wordt de data eerst geëxtraheerd van de bron, maar wordt vervolgens direct opgeslagen in de database als ruwe data. Hiervoor moet de database geschikt zijn, wat er op neerkomt dat hij een data lake moet zijn. Vervolgens wordt de data omgevormd zodat hij past binnen het data-analysemodel.

De ETL-methode is voornamelijk goed voor efficiënte data-opslag, omdat het omzetten van data wordt gedaan op basis van het dataopslagmodel, dat gemaakt wordt met normalisatie en opslagcapaciteit als belangrijkste factoren. De ELT methode is voornamelijk goed voor efficiënte data-analyse, omdat het omzetten van data gebeurt op basis van het analysemodel, dat gemaakt wordt met de analysetool als belangrijkste factor.

2.2. Theorieën en modellen

2.2.1. CRISP-DM

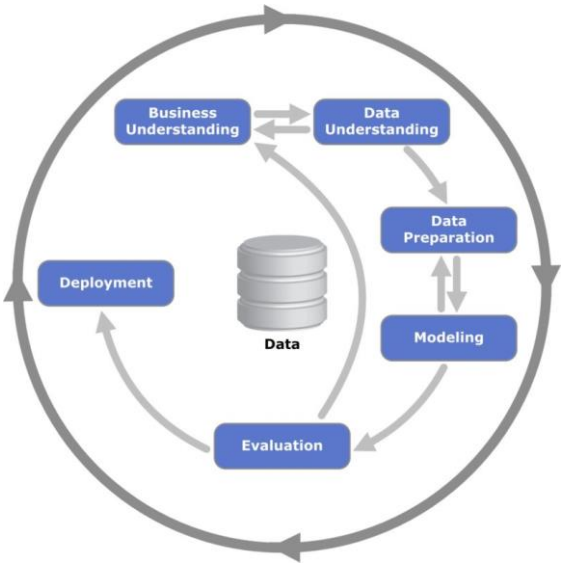
CRISP-DM is een industrieonafhankelijk procesmodel voor datamining. Het staat voor 'CRoss Industry Standard Process for Data Mining". Het bestaat uit zes iteratieve fasen, vanaf 'Business Understanding' tot 'Deployment' (zie Tabel 1). Tabel 1 geeft een korte beschrijving van het voornaamste idee, de taken en het resultaat (Schröer, Kruse, & Gómez, 2021).

Tabel 1 CRISP-DM procesmodelbeschrijvingen (Schröer, Kruse, & Gómez, 2021)

Fase	Korte beschrijving
Business Understanding	De bedrijfssituatie moet worden geëvalueerd om een overzicht te krijgen van de beschikbare en benodigde middelen. Het doel van de datamining bepalen is een van de belangrijkste aspecten in deze fase. Eerst moet het type datamining worden bepaald (bijvoorbeeld classificatie) en de dataminingsuccescriteria (zoals precisie). Daarnaast wordt een plan van aanpak gemaakt.
Data Understanding	De essentiële taken in deze fase zijn data verzamelen uit gegevensbronnen, deze data verkennen en beschrijven, en de kwaliteit ervan controleren. Dit gebeurt onder andere door de correlaties tussen de verschillende attributen te bepalen met behulp van statistische analyses.
Data Preparation	Gegevensselectie moet worden uitgevoerd door in- en uitsluitingscriteria te definiëren. Slechte gegevenskwaliteit kan worden opgelost door de gegevens op te schonen. Afhankelijk van het gebruikte model (gedefinieerd in de eerste fase) moeten afgeleide attributen worden geconstrueerd. Voor al deze stappen zijn verschillende methoden mogelijk die afhankelijk zijn van het model.
Modelling	De gegevensmodelleringsfase bestaat uit het selecteren van de modelleringstechniek en het bouwen van een testcase en het model. Alle dataminingtechnieken kunnen worden gebruikt. In het algemeen is de keuze afhankelijk van het bedrijfsprobleem en de gegevens. Belangrijker is om de keuze te verklaren. Bij het bouwen van het model is het gepast om het model te evalueren aan de hand van evaluatiecriteria en het beste te selecteren, in dit geval welk model het beste de onderhoudsbehoeftes kan voorspellen (predictive maintenance).
Evaluation	In de evaluatiefase worden de resultaten gecontroleerd op de gedefinieerde bedrijfsdoelen. De resultaten moeten worden geïnterpreteerd en er moeten verdere acties worden gedefinieerd. Daarnaast moet het proces in het algemeen worden beoordeeld.

Deployment	De implementatiefase kan verschillende uitwerkingen hebben. Het kan een eindrapport of een softwarecomponent zijn. De fase bestaat in brede zin uit het plannen van de implementatie en het monitoren en onderhouden ervan.

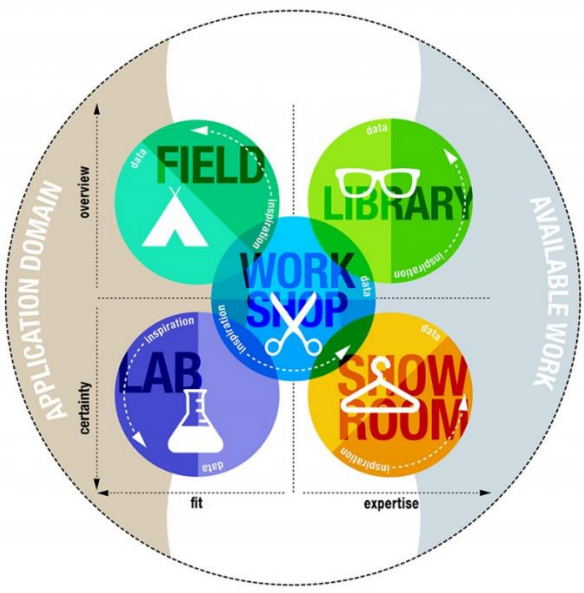
De stappen van CRISP-DM worden in de volgende volgorde gehanteerd (zie figuur 3).



Figuur 3 CRISP-DM Methode (cc-by-sa Kenneth Jensen)

2.2.2. DOT-Framework

De opzet van dit onderzoek wordt voornamelijk gebaseerd op het DOT-Framework. Het DOT-Framework bestaat uit vijf onderzoeksstrategieën. Deze strategieën zijn Library (Bieb), Field (Veld), Workshop (Werkplaats), Lab en Showroom (zie figuur 4). (Wikiwijs, 2022). Hiervoor is gekozen, omdat dit een zeer flexibel raamwerk is dat gemaakt is voor ICT-onderzoek. Doordat het verschillende methodes biedt binnen elke strategie, is het zeer eenvoudig om de juiste methodes te kiezen die passen bij het onderzoek.



Figuur 4 DOT-Framework (cc-by-sa HAN)

2.2.3. Orange

Orange is een open source machine learning en dataminingsoftware geschreven in Python. Het heeft een visueel programmeerbare front-end, bedoeld voor exploratieve data-analyses en -visualisatie en kan ook worden gebruikt als Python library. Het programma wordt onderhouden en ontwikkeld door de Bioinformatics Laboratory van de faculteit Computer and Information Science van de universiteit van Ljubljana. Orange is component-based. Deze componenten worden widgets genoemd, en worden gebruikt voor verschillende taken zoals simpele datavisualisatie en subsetselectie, tot empirische evaluaties van lerende algoritmes en voorspellende modellering (Naik & Samant, 2016).

2.2.4. Unsupervised learning

Unsupervised learning is een machinelearningmethodologie waarbij het algoritme zonder tussenkomst van mensen leert. Om te leren maakt hij gebruik van ongelabelde data die meestal ongestructureerd en onbewerkt is. Hierdoor heeft deze methodologie ook veel beperkingen. Hij heeft geen startpunt voor de training, waardoor hij goed is in slechts een aantal soorten taken. De meest gebruikte taken zijn clusteren en dimensiereductie (Sydorenko, 2021).

2.2.5. Supervised learning

In tegenstelling tot unsupervised learning, maakt supervised learning wel gebruik van gelabelde data. Dit betekent dat het model getraind wordt op basis van data die gestructureerd, opgeschoond en geannoteerd (aangegeven wat voor soort data het is) is. Dit annoteren, of labelen, is arbeidsintensief, maar daardoor is het wel veel makkelijker voor een algoritme om getraind te worden, waardoor er meer interessante bevindingen kunnen komen uit deze algoritmes (Sydorenko, 2021).

2.2.5.1. Classification

Classification is een subgroep van supervised learning modellen waarbij het doel is om objecten te verdelen in een beperkt aantal van tevoren bekende categorieën. Hiervoor moet het algoritme wel getraind zijn op genoeg objecten van elke categorie. Dit soort algoritmes worden bijvoorbeeld gebruikt voor het leren herkennen van objecten op foto's. Dit is hoe je telefoon weet dat je een foto maakt van een kat en niet, bijvoorbeeld, een hond. (Sydorenko, 2021)

2.2.5.2. Regression

Regression wordt gebruikt wanneer er sprake is van een doelwaarde die continu is, in plaats van categoriaal. Het gaat hier meestal om een numerieke waarde. Hierdoor zijn er vaak oneindig veel mogelijke antwoorden, en is een voorspeld antwoord vrijwel nooit precies goed. Het gaat erom om in de buurt van het juiste antwoord te komen. Hiervoor maakt het algoritme gebruik van correlatie tussen verschillende variabelen. Regression zoekt deze correlatie en op basis daarvan maakt hij voorspellingen voor de doelwaarde.

2.2.6. ARIMA

ARIMA is een model dat gebruikt kan worden voor het maken van voorspellingen van toekomstige waardes, gebaseerd op waardes van het verleden en het heden van dezelfde variabele. Om dit te kunnen bereiken is het nodig om een tijdsserie te hebben. Dit betekent dat het bekend moet zijn op welke tijd welke waarde is gemeten. Daarnaast is het ook van belang dat de waardes van de doelvariabele van het verleden en het heden bekend zijn in constante en reguliere intervallen (Master's in Data Science, 2023).

3. Methodologie

In dit hoofdstuk wordt kort de methodologie van de data-acquisitie en de data-analyse besproken en daarnaast wordt de methodologie van procesanalyse omschreven. In deze methodologie wordt gebruik gemaakt van het DOT-Framework (zie hoofdstuk 1.2.4).

3.1. Huidig werkproces

Deelvraag 1 gaat over het in kaart brengen van de huidige bedrijfsprocessen. Voor deze deelvraag wordt er gebruik gemaakt van de Library methode Expert Interview en de Field methode Interview.

3.1.1. Huidige werkwijze en achtergrond

De persoon binnen Bright die op het moment verantwoordelijk is voor het bedrijfsproces van onderhoud is de proces engineer Vincent Kroeze. Kroeze is bereikbaar voor zowel formele interviews als voor het beantwoorden van korte vragen. Omdat het hier om de specifieke processen van Bright gaat, is de proces engineer van Bright de meest geschikte persoon om dit interview mee te voeren (Expert Interview).

3.1.2. Stakeholders

Uit het Expert Interview zal een aantal medewerkers naar voren komen die stakeholders zijn in het huidige en toekomstige proces. Deze stakeholders zullen ook belangrijke informatie kunnen verstrekken. Deze stakeholders zullen worden geïnterviewd over hun rol binnen deze werkprocessen. Ook worden hun meningen en ideeën over het huidige werkproces meegenomen (Interview).

3.2. Relevante data

Deelvraag 2 gaat over de relevante data die er momenteel beschikbaar is. Voor deze deelvraag wordt er gebruik gemaakt van de Library methode Expert Interview en de Lab methode Data analytics.

3.2.1. Beschikbare data

Ook voor deze deelvraag is de proces engineer van Bright de meest geschikte expert, zowel vanwege de specifieke kennis waar hij over beschikt, als vanwege zijn algemene beschikbaarheid. Niet alleen weet hij welke data er wordt opgeslagen, ook weet hij hoe deze data wordt opgeslagen en wat deze data precies betekent. Verder kan hij ook de achtergrond en de scheikundige implicaties van de data verder uitleggen (Expert Interview). Daarnaast zal voor de beschikbare data ook worden gekeken naar de API om de antwoorden van de proces engineer te valideren (Data analytics).

3.3. Data-acquisitie en -opslag

Deelvraag 3 gaat over de data-acquisitie en -opslag. Voor deze deelvraag wordt er gebruik gemaakt van de Library methode Available product analysis en Community research en de Workshop methode Prototyping.

3.3.1. Beschikbare methodes

De data die wordt gemeten in de machines maakt gebruik van een Ixon cloud-oplossing. Ixon biedt zelf manieren aan om deze data uit de cloud te halen, bijvoorbeeld door middel van hun API (Ixon, 2022). Deze API is een voorbeeld van een available product. Ook de vervolgstappen, zoals het opslaan en transformeren van de data, zijn eerder uitgevoerd door andere ontwikkelaars en deze resources zijn veelal vrij toegankelijk (Available product analysis).

3.3.2. Prototype bouwen

Deze deelvraag kan het beste worden beantwoord door het daadwerkelijk uitvoeren van de data-acquisitie en -opslag. De opslag is op het moment qua prototyping out of scope, omdat de overgang naar de nieuwe database nog niet voltooid is. Het is daardoor niet mogelijk om op het moment een

werkend prototype te bouwen. Er zal wel een advies worden gegeven over de data-opslag. Het is wel essentieel om de acquisitie uit te voeren, ook voor het beantwoorden van de overige deelvragen. Hiervoor wordt gebruik gemaakt van de API van Ixon, in combinatie met PowerShell (Prototyping).

3.3.3. Bestaande oplossingen

Tijdens het maken van de prototypen is veel gebruik gemaakt van online resources zoals Stackoverflow en Reddit. Hier waren vaak oplossingen te vinden voor problemen die door andere gebruikers al zijn opgelost (Community research).

3.4. Data-analyse

Deelvraag 4 gaat over de data-analyse. Voor deze deelvraag wordt er gebruik gemaakt van de Library methode Literature research, de Workshop methode Prototyping en de Lab methode Data analytics. Verder wordt de CRISP-DM methode gebruikt.

3.4.1. Geschikte algoritmes

Er bestaan diverse algoritmes voor data-analyses. De algoritmes die worden getest zullen niet ontworpen worden door de onderzoekers; in plaats daarvan zal de literatuur erop worden nageslagen om te kijken welke mogelijke algoritmes zijn uitgevonden om deze analyses mee uit te kunnen voeren. Hierbij zal worden gekeken welke algoritmes geschikt zijn voor de specifieke data en analysedoelen van dit onderzoek (Literature research).

3.4.2. Prototype bouwen

Dit onderzoek focust zich op de vraag of het mogelijk is om met behulp van een data-analyse te voorspellen wanneer onderhoud zal moeten worden gepleegd (predictive maintenance), en een prototype van deze analyse zal daarmee de beste manier zijn om te bewijzen dat dat inderdaad kan. Mocht het prototype er niet in slagen de doorslag enigszins te voorspellen, dan zal dit niet betekenen dat het onmogelijk is, maar kan de conclusie wel worden getrokken dat het niet eenvoudig is en dat er meer onderzoek nodig is.

Deze data-analyse zal worden uitgevoerd met behulp van Orange. Daarmee kunnen een flink aantal data-analyse modellen worden getest om te kijken of die geschikt zijn voor het analyseren en voorspellen van de toekomstige data van Bright. Op basis hiervan kan een model worden gebouwd dat het doorslagmoment van de koolfilters met enige nauwkeurigheid kan voorspellen (Prototyping, Data Analytics).

3.5. Datavisualisatie

Deelvraag 5 gaat over de visualisatie van de data. Voor deze deelvraag wordt er gebruik gemaakt van de Field methode Interview en van de Workshop methode Prototyping.

3.5.1. Wensen van stakeholders

De datavisualisatie heeft als belangrijkste doel dat er voor de stakeholders van het nieuwe werkproces duidelijk wordt gemaakt wat de conclusies zijn van de data-analyse, en dan specifiek wanneer de voorspelde doorslag zal plaatsvinden. Hoe deze data inzichtelijk moet worden gemaakt heeft dus ook te maken met de voorkeuren en wensen van de stakeholders. Hierdoor is het belangrijk om met hen in gesprek te gaan (Interview).

3.5.2. Prototype bouwen

Net als met de data-acquisitie en de data-analyse zal ook hier een prototype voor worden gemaakt. Dit zal ook binnen Orange worden gemaakt, hoewel de uiteindelijke oplossing gezocht moet worden in een dashboardachtige omgeving, bij voorkeur binnen het ERP-systeem (Prototyping).

3.6. Implementatie

Deelvraag 6 gaat over de implementatie van het nieuwe werkproces en het opstellen van een implementatieadvies. Voor deze deelvraag wordt er gebruik gemaakt van de Field methode Interview en Task analysis en de Showroom methode product review.

3.6.1. Prototypen beoordelen

Wanneer de prototypen van 3.3.2, 3.4.3 en 3.5.2 klaar zijn, zullen deze worden getoond aan de proces engineer. Hij zal zijn kennis en kunde gebruiken om te controleren of de voorspellingen enigszins betrouwbaar, en dus geschikt voor implementatie zijn. Natuurlijk kunnen ze ook worden getest met bekende data, maar het is de proces engineer die het beste kan beoordelen of ze betrouwbaar zijn voor de toekomstige datapunten. Daarnaast zal hij ook kunnen beoordelen of ze goed genoeg zijn qua gebruikersgemak (Product Review).

3.6.2. Opinie stakeholders

Naast de proces engineer zullen ook de overige stakeholders gevraagd worden naar hun mening over de prototypen. Verder zal er ook worden gevraagd hoe zij verwachten dat de nieuwe methodes gebruikt kunnen worden in een nieuw werkproces (Interview).

3.6.3 Implementatie advies

Na afloop van dit onderzoek zullen er nog veel stappen doorlopen moeten worden voordat de resultaten van dit onderzoek kunnen worden geïmplementeerd in de dagelijkse werkzaamheden van Bright. Om deze implementatie in goede banen te leiden, wordt er een implementatieadvies opgesteld. Voor het opstellen van dit advies moeten die stappen in kaart worden gebracht (Task analysis).

3.7. Verandermanagement

De behoefte aan verandermanagement is in dit onderzoek beperkt, omdat er maar een klein aantal medewerkers is dat geraakt wordt door de veranderingen. Er wordt daarom gebruik gemaakt van een vereenvoudigde versie van het ADKAR-model (Mulder & Janse, 2022). Hierin staan de volgende stappen centraal.

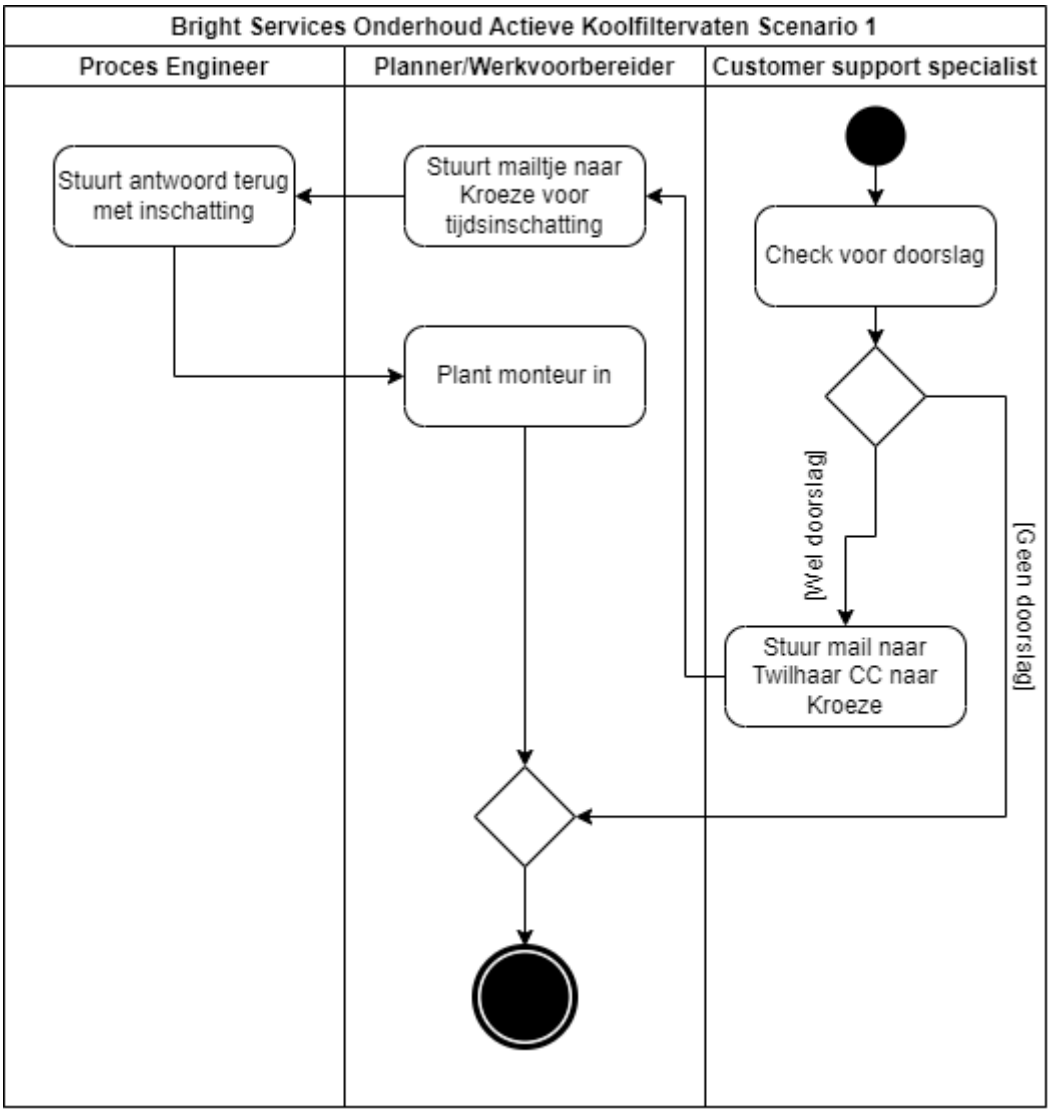
- Awareness: medewerkers moeten bewust worden van de noodzaak van de verandering;
- Desire: medewerkers moeten het verlangen krijgen om deel te nemen aan de verandering en ze volledig ondersteunen;
- Knowledge: door kennis van het veranderingsproces wordt het voor medewerkers duidelijk wat het doel is van de verandering;
- Ability: door het vermogen om nieuwe vaardigheden te leren en aan te sturen op ander gedrag, wordt de verandering beter geaccepteerd;
- Reinforcement: door de verandering te bekrachtigen, is het voor alle medewerkers duidelijk dat er geen weg meer terug is.

4. Resultaten

In dit hoofdstuk worden de resultaten van het onderzoek uiteengezet. Deze resultaten worden per deelvraag besproken.

4.1. Huidig werkproces

Het huidige werkproces voor het inplannen van de onderhoudsbeurten van de actieve koolfiltervaten heeft drie belangrijke actoren, oftewel medewerkers. Dit zijn de eerder genoemde planner/werkvoorbereider Ben Twilhaar, de proces engineer Vincent Kroeze en daarnaast ook de customer support specialist Ewout Goorhuis. De customer support specialist controleert elke dag op afstand de statussen van de installaties en is tevens bereikbaar voor de klanten. Mocht de customer support specialist een doorslag waarnemen (H_2S meting van meer dan 5ppm tussen de koolfilters), dan zal hij een mail sturen naar de planner en de proces engineer om hen hiervan op de hoogte te stellen. Vervolgens zal de planner met de proces engineer overleggen over een tijdsinschatting tot een kritieke doorslag (H_2S meting van meer dan 100ppm tussen de koolfilters). Met deze informatie plant de planner een monteur in (Kroeze, Huidige werkproces Bright Services onderhoud actieve koolfiltervaten, 2022).



Figuur 5 Stroomdiagram van het huidige werkproces van het inplannen van onderhoud op de actieve koolfiltervaten

4.2. Relevante data

Een uitgebreide beschrijving van de CRISP-DM methode die is gebruikt om deze data te achterhalen is te vinden in bijlage 1. Hier wordt slechts kort ingegaan op de verschillende stappen.

De eerste stap is Business Understanding. De gewenste output voor Bright is een proof of concept van een predictive maintenance model. De wens is om dit te doen voor de actieve koolfiltervaten op basis van een van de installaties. Op basis hiervan is het de bedoeling om te bepalen of predictive maintenance een realistische stap is om te nemen in het doel om de bedrijfsprocessen van Bright efficiënter te maken. Het dataminingdoel van dit onderzoek is om te kijken of de hoeveelheid dagen tot de doorslag plaats gaat vinden van tevoren kan worden voorspeld.

De tweede stap is Data Understanding. Er zijn totaal 103 variabelen die kunnen worden opgehaald uit de Ixon API. In bijlage 1 is een beschrijving van al die variabelen te vinden. In dit document wordt alleen ingegaan op de variabelen die relevant zijn voor dit onderzoek. In tabel 2 zijn de 11 variabelen met daarin hun code, beschrijving en type variabele en, wanneer het numerieke variabelen zijn, de classificatie continu of discreet. Daarna zal er voor de numerieke datapunten een lineaire regressie plaatsvinden om te bepalen in welke mate deze variabelen zijn gecorreleerd aan de dagen tot doorslag (zie tabel 3 en 4).

De relevantie van de variabelen is gebaseerd op de locatie van de sensoren in de keten, namelijk de sensoren die in de buurt zitten van de koolfiltervaten, en de sensoren die belangrijk zijn voor het maken van de afgeleide datapunten (Kroeze, Relevante data uit de Ixon API, 2022). Deze datapunten zijn:

Attribuutcode	**Beschrijving**	**Type**	**Continu/Discreet**
Tijd	Een timestamp van de meting	Interval	Continu
CF001	De biogasflow richting de koolfiltervaten (in Nm^3)	Ratio	Continu
CH4_BF	De CH_4 compositie in het ruwe biogas (in %)	Ratio	Continu
H2S_BF	De H_2S compositie in het ruwe biogas (in ppm)	Ratio	Continu
O2_BF	De O_2 compositie in het ruwe biogas (in %)	Ratio	Continu
CO2_BF	De CO_2 compositie in het ruwe biogas (in %)	Ratio	Continu
CH4_BT	De CH_4 compositie in enkel gefilterd biogas (in %)	Ratio	Continu
H2S_BT	De H_2S compositie in enkel gefilterd biogas (in %)	Ratio	Continu
O2_BT	De O_2 compositie in enkel gefilterd biogas (in %)	Ratio	Continu
CO2_BT	De CO_2 compositie in enkel gefilterd biogas (in %)	Ratio	Continu
SEQ.STATENR	Het statusnummer van de installatie (62 is bijvoorbeeld ‘aan’)	Nominal	-

Tabel 2 Beschrijving en classificatie van de relevante attributen uit Ixon

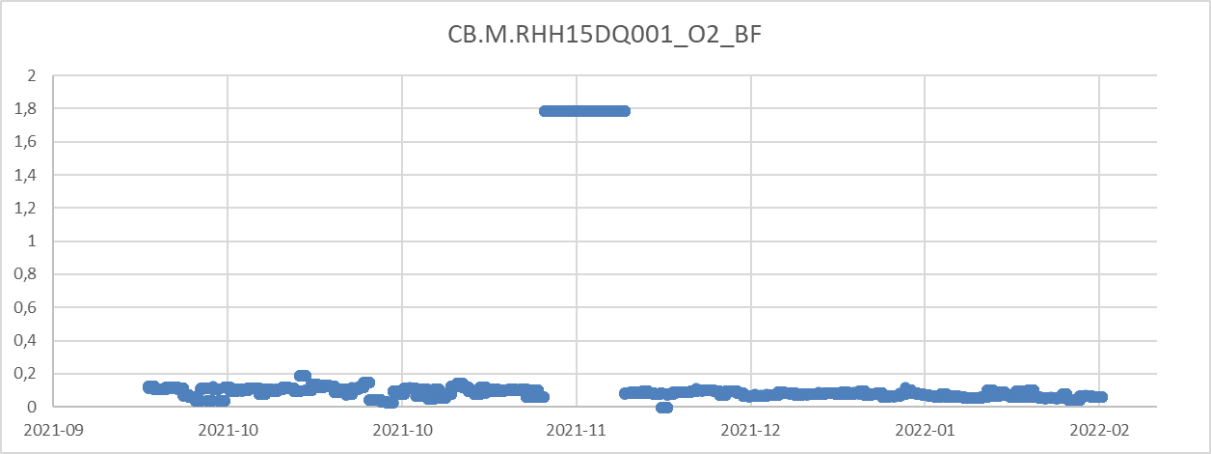
Bij de directe data die uit de Ixon API kan worden opgehaald zitten er geen datapunten die significant gecorreleerd zijn aan de dagen tot doorslag. Dit is berekend met de Pearson r correlatiefunctie, die beschikbaar is in Excel. De exacte functie is $Correl(X,Y) = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$, (Microsoft, 2023), waarbij x de te testen gemeten sensor waarde is en y de dagen tot doorslag. Hierbij is een score (correlatiecoëfficiënt) hoger dan 0,6 of lager dan -0,6 een indicatie van significant gecorreleerde waardes. Gezien het SEQ.STATENR geen numerieke waarde is, kan hier geen Pearson r correlatie op worden uitgevoerd, maar deze waarde is alsnog belangrijk, omdat die aangeeft wanneer de installatie aan staat, wat belangrijk is voor de afgeleide waardes. De correlatiecoëfficiënt van deze metingen ten opzichte van de dagen tot doorslag zijn:

Attribuutcode	Correlatiecoëfficiënt
A10CF001	-0,119516613
CH4_BF	-0,017195244
H2S_BF	-0,160101798
O2_BF	0,247093743
CO2_BF	-0,262650926
CH4_BT	0,195410121
H2S_BT	-0,427213973
O2_BT	0,083594382
CO2_BT	-0,203266955

Tabel 3 Correlatiecoëfficiënt van de relevante attributen uit Ixon

De derde stap van CRISP-DM is de Data Preparation. Hier wordt de selectie gemaakt, de data wordt opgeschoond en wordt er data geconstrueerd.

Bij de opschoning worden eventuele meetfouten verwijderd al dan niet vervangen door een schatting. De meetfouten zijn te herkennen aan hun afwijkende waardes. In figuur 6 is een voorbeeld van een meetfout te vinden. Hier is lijn bij 1,8% rond november 2021 een meetfout. Uit de overige tabellen is af te leiden dat er meerdere meetfouten zijn waargenomen rond november 2021 (zie hoofdstuk 3.2 van bijlage 1).



Figuur 6 Voorbeeld van een meetfout, concentratie O2 in ruw biogas (in %) gefilterd op status 62

Het construeren van data houdt in dat niet-beschikbare data wordt afgeleid van de beschikbare data. Omdat de doelwaarde (dagen tot doorslag) niet beschikbaar is in de originele dataset, moet die worden toegevoegd. Intuïtief is het logisch dat de hoeveelheid H_2S die in de filter is blijven hangen, en de tijd dat de installatie aan heeft gestaan beide relevant kunnen zijn voor het voorspellen van het moment waarop de filters moeten worden vervangen. Dit komt omdat de filters een bepaalde capaciteit en potentieel een bepaalde levensduur hebben. Vandaar dat ook actieve minuten sinds doorslag, en de totale hoeveelheid H_2S die door het eerste koolfiltervat is gestroomd, zullen worden geconstrueerd:

- Dagen tot doorslag (hierna DTD)
  De DTD kan alleen achteraf worden toegevoegd. Dit kan worden gebruikt voor het trainen van het datavoorspellingsmodel. Dit wordt berekend door te kijken naar hoe lang het duurt voor de volgende doorslag (H2S_BT met een waarde van boven de 5) plaatsvindt in de dataset.
- Actieve minuten sinds doorslag (hierna AMSD)

Dit wordt berekend door voor elke meting met een SEQ.STATENR van 62, 5 minuten toe te voegen aan de AMSD van de vorige meting. De AMSD reset naar 0 wanneer H2S_BT een waarde heeft van boven de 5 (doorslag).

- Totale hoeveelheid H_2S die door het eerste koolfiltervat is gestroomd sinds de laatste onderhoudsbeurt (hierna Totaal Nm^3 H_2S).
  Dit wordt berekend door de biogasflow (A10CF001) maal de H_2S-compositie in het ruwe biogas (H2S_BF delen door 1 miljoen, omdat dit gegeven wordt in ppm) op te tellen bij de waarde van Totaal Nm^3 H_2S van de vorige meting. De Totaal Nm^3 reset naar 0 wanneer er voor het eerst geen doorslag meer is.

Ook hiervan zal met behulp van de Pearson r correlatiefunctie in Excel een correlatiecoëfficiënt worden berekend:

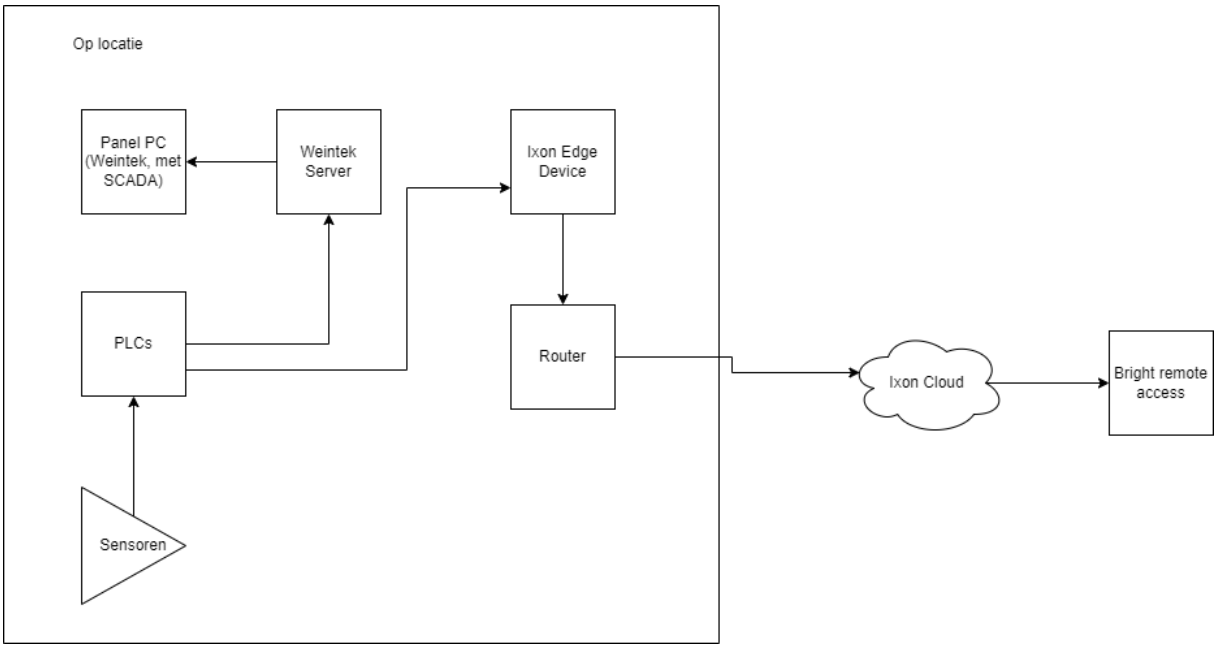
Attribuut	**Correlatiecoëfficiënt**
DTD	1
AMSD	0,191860259
Totaal Nm^3 H_2S	-0,827452537

Tabel 4 Correlatiecoëfficiënt van de afgeleide attributen

Uit deze correlatiecoëfficiënten blijkt dat de Totaal Nm^3 H_2S een sterke correlatie heeft met de DTD. De DTD heeft natuurlijk een correlatie van 1, omdat dat de doelwaarde is. De AMSD heeft, enigszins onverwacht, geen sterke correlatie.

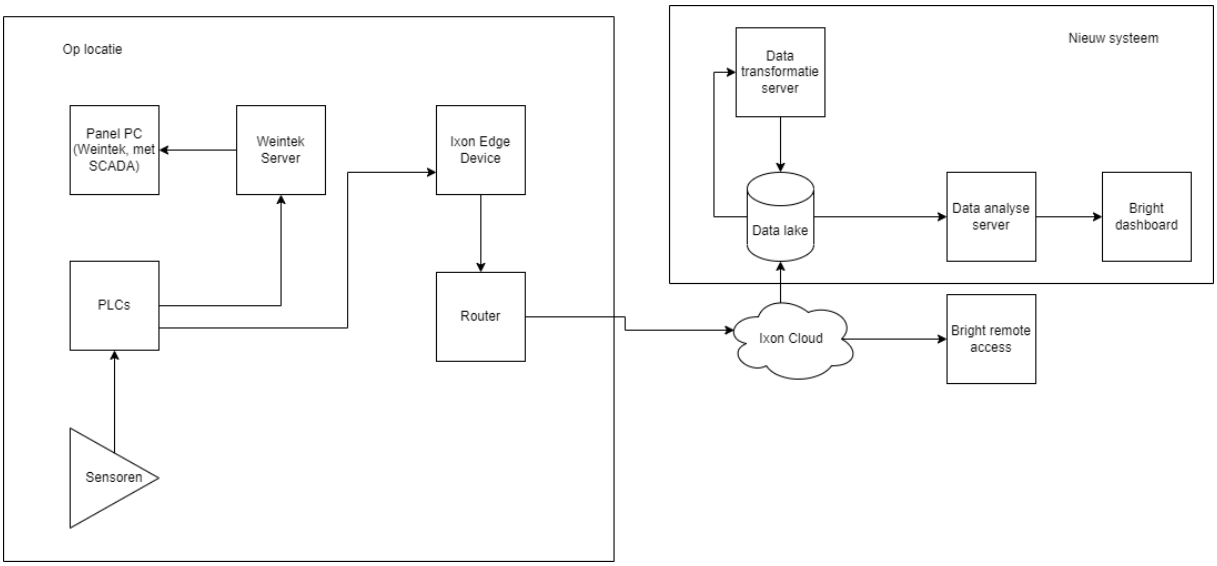
4.3. Data-acquisitie

De data die Bright verzamelt vanuit hun installaties met behulp van sensoren wordt door middel van een Ixon Cloud IoT-oplossing opgeslagen. Deze data is hieruit te halen met behulp van de API van Ixon (Ixon, 2022). Deze API is aangeroepen met PowerShell (zie bijlage 3 voor de gemaakte code). De opgevraagde data wordt geleverd in een JSON-format. Om data op te halen zijn wel de TagID's (attribuut binnen Ixon) nodig van de benodigde data. Deze TagID's zijn helaas niet eenvoudig in te zien, en niet elke sensor lijkt een TagID te hebben. Deze methode is op het moment dus niet geschikt om alle data op te halen. Er is ook een andere, handmatige manier om de data op te halen. Dit gebeurt door via Ixon een VPN-verbinding te maken met de installatie. Elke installatie heeft on-site een Ixon-router. Daar zit ook een Weintek IoT-installatie op, waar een lokaal SCADA-systeem op draait. Op dit SCADA-systeem wordt, mits ingesteld, een deel van de data uit de PLC opgeslagen (zie figuur 7 voor een schematische tekening van de huidige architectuur). Deze database kan als .db bestand worden gedownload. Op dit moment wordt de data door de proces engineer via een Pythonscript omgezet naar Panda Dataframe voor eigen gebruik, of naar XLSX of CSV wanneer de data door collega's wordt gebruikt (Kroeze, Handmatige data acquisitie, 2022). Dit is een omslachtige methode, maar deze methode heeft wel beschikking over alle benodigde data. Daarom is deze methode ook gebruikt voor dit onderzoek.



Figuur 7 Schematische weergave architectuur voor de huidige datavoorziening

In figuur 7 is de huidige schematische weergave van de architectuur voor de datavoorziening te zien. Om deze analyses te kunnen verwerken zal hier een extra datastroom aan moeten worden toegevoegd die de analyseserver voedt met data. Zie figuur 8 voor de voorgestelde nieuwe architectuur.

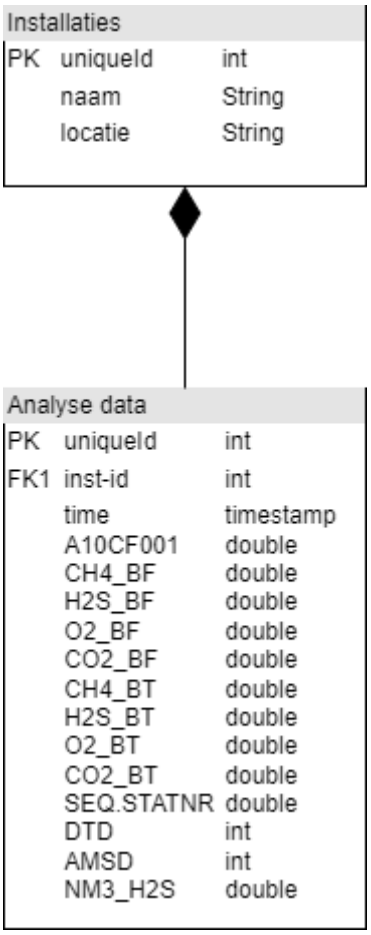


Figuur 8 Schematische weergave architectuur voor de toekomstige datavoorziening

Momenteel is Bright, samen met moederbedrijf HoSt, bezig om een nieuwe data-opslag te realiseren. Er zullen twee verschillende opslagsystemen worden gebruikt. Een van deze systemen zal worden gebruikt voor cruciale data, waardoor er gekozen is voor een veilig en betrouwbaar opslagsysteem. Daarnaast is er gekozen voor een apart systeem voor de big data die gebruikt wordt voor onder andere predictive maintenance. Omdat deze data niet cruciaal is voor de dagelijkse werkzaamheden, is prioriteit gegeven aan kostenbesparing. Daarom is er gekozen voor het minder betrouwbare, maar kostenefficiënte ADLS gen2 (Kroeze, Toekomstige ICT architectuur van HoSt en haar dochterondernemingen, 2022). Omdat een data lake data opslaat in zijn originele format, en ook werkt

met het JSON-format dat de API aanlevert, is het niet nodig om deze data te transformeren voordat hij wordt weggeschreven in de data lake (Miloslavskaya & Tolstoy, 2016).

De data zal worden opgeslagen in een datalake. Dit is een ongestructureerde vorm van opslag waarin de data in hun originele vorm via een ELT methode wordt opgeslagen. De data die vanuit de Ixon wordt opgehaald kan hier dus als de geëxporteerde JSON file ingeladen worden. In een datalake is het ook mogelijk om gedeeltelijke gestructureerde data op te slaan. Hierbij is het mogelijk om dit te doen voor elke usecase, gezien de opslag in een datalake lage kosten hebben. Dit betekent dat het normaliseren van data niet nodig is in een datalake. Om de data-analyse op een zo snel mogelijk te maken is het verstandig om de data, naast de JSON files, ook op te slaan op een meer gestructureerde manier. Ook is het mogelijk om afgeleide data apart op te slaan, om te voorkomen dat de berekeningen telkens opnieuw moeten plaatsvinden. De data voor de analyse zal op de volgende manier worden opgeslagen (zie figuur 9). Hierin worden er twee tabellen opgeslagen, de eerste met de informatie over de installaties, en de tweede met alle data die nodig is voor de analyse zoals bepaald in hoofdstuk 4.2. De analyse data wordt opgeslagen per tijdsinterval, dus voor elke data instantie in het 'installaties' tabel, zijn er meerdere data instanties in het 'analyse data' model.



Figuur 9 Gestructureerde data-opslag van de analyse data binnen de ongestructureerde datalake

Om het onderzoek te kunnen vervolgen zonder de permanente oplossingen die nog gaan komen, is de data via de handmatige methode, zoals hierboven beschreven, gebruikt om de data op te halen, en is de afgeleide data vervolgens met behulp van Excel berekend, zoals beschreven in 4.2. Dit is opgeslagen als CSV-bestand. Deze CSV-bestand kan vervolgens worden ingelezen door Orange (Orange, 2022).

4.4. Datavoorspellingsmodel

Een uitgebreide beschrijving van de CRISP-DM methode die is gebruikt om dit model te realiseren is te vinden in bijlage 1. Hier wordt slechts ingegaan op de belangrijkste bevindingen.

Stap 1 tot en met 3 van CRISP-DM zijn uitgelegd in hoofdstuk 4.2. Stap 4 van CRISP-DM betreft Modeling. Hierbij is het de bedoeling om een modeltechniek te kiezen, een test te ontwerpen, een model te bouwen en dit te beoordelen. Als eerste wordt er dus een modeltechniek gekozen.

De data van deze analyse is in de vorm van een tijdserie. Elke 5 minuten wordt er een sample gemaakt van de waardes van alle sensoren, en de afgeleide waardes worden berekend met zowel de waardes van de nieuwe sample, als met de waardes van de voorgaande samples.

Het doel van dit datavoorspellingsmodel is het voorspellen van de resterende tijd tot de doorslag plaats zal vinden, gebaseerd op de historische data: dit is dus een tijdserie. Verder zijn er gelabelde voorbeelden beschikbaar, waardoor supervised learning mogelijk is. Daarnaast is de doelwaarde een numerieke score, en geen categorie, dus vandaar dat regressieanalyses het meest geschikt zijn (Li, 2020) (Zie hoofdstuk 2.2.3 t/m 2.2.5). Bij veel tijdseries is ARIMA ook een geschikt model. In dit geval is ARIMA echter geen geschikt model, omdat ARIMA voornamelijk gebruik maakt van de patronen in het verloop van de doelwaarde (zie hoofdstuk 2.2.6). Aangezien de werkelijke doelwaarde in dit onderzoek (dagen tot doorslag) pas bekend is nadat de doorslag heeft plaatsgevonden, is er geen betrouwbare data beschikbaar voor het ARIMA-model om zijn voorspellingen op te baseren.

Binnen Orange zijn er zes regressie analysemodellen beschikbaar, namelijk Tree, Random Forest, Gradient Boosting, Linear Regression, AdaBoost en Neural Network. Er is besloten om al deze zes modellen te testen met de data uit hoofdstuk 4.2, waarbij de doelwaarde de DTD is, en alle overige variabelen worden gebruikt voor het voorspellingsmodel.

De volgende stap is het ontwerpen van de test. In totaal zijn er vijftien doorslagcycli geweest, waarvan de DTD dus bekend is. Omdat het voor het testen nodig is dat de DTD bekend is, is er besloten om de eerste vier te gebruiken als trainingsdata en de laatste te gebruiken als testdata. Dit wordt gedaan met behulp van Orange. Het gaat hier bijna 137.000 datarijen, met ieder 15 datapunten. Dat betekent dat elke doorslagcyclus gemiddeld ongeveer 137.000 datapunten bevat. Hierdoor zou de betrouwbaarheid, ondanks dat er slechts vijftien doorslagen hebben plaatsgevonden, redelijk hoog moeten zijn. Natuurlijk zal de betrouwbaarheid naar verloop van tijd hoger worden door de toevoeging van meer data.

Vervolgens moet het model, of in dit geval, moeten de zes modellen worden ontworpen. Dit gebeurt ook in Orange. Eerst wordt de doelwaarde gedefinieerd en vervolgens worden de trainingsdata en testdata opgesplitst. De trainingsdata wordt doorgestuurd naar de verschillende datavoorspellingsmodellen. De getrainde modellen worden vervolgens samen met de testdata beoordeeld door de 'Predictions'-module.

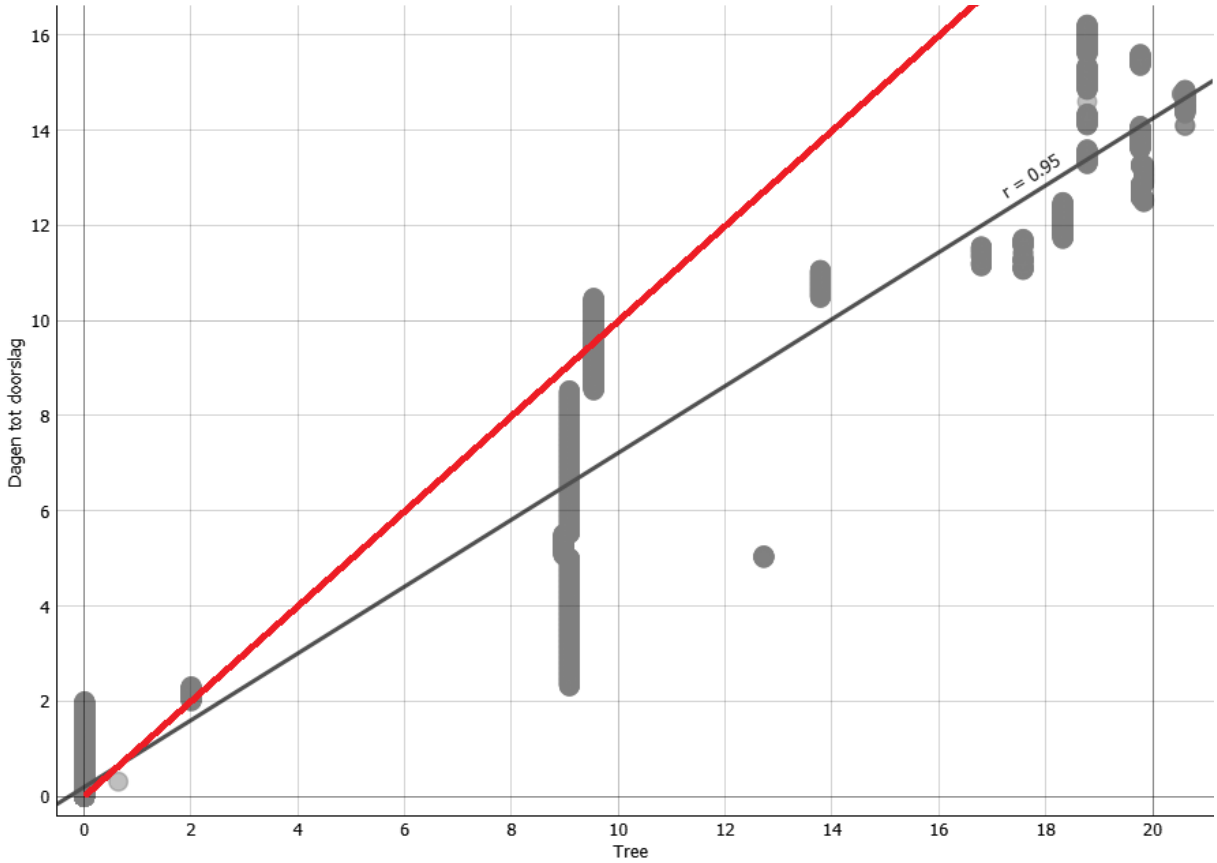
Hierna moeten de modellen worden geëvalueerd. Hieruit zijn de volgende scores gekomen.

Model	MSE	RMSE	MAE	R2	Opmerkingen
Tree	16.779	4.096	3.076	0.396	Individuele voorspellingen zijn niet zeer accuraat, maar de trendlijn wel
Random Forest	10.423	3.206	2.589	0.630	Voorspelling consistent te hoog, maar trendlijn behoorlijk accuraat
Gradient Boosting	14.789	3.846	3.185	0.468	Voorspelling consistent te hoog, maar trendlijn behoorlijk accuraat

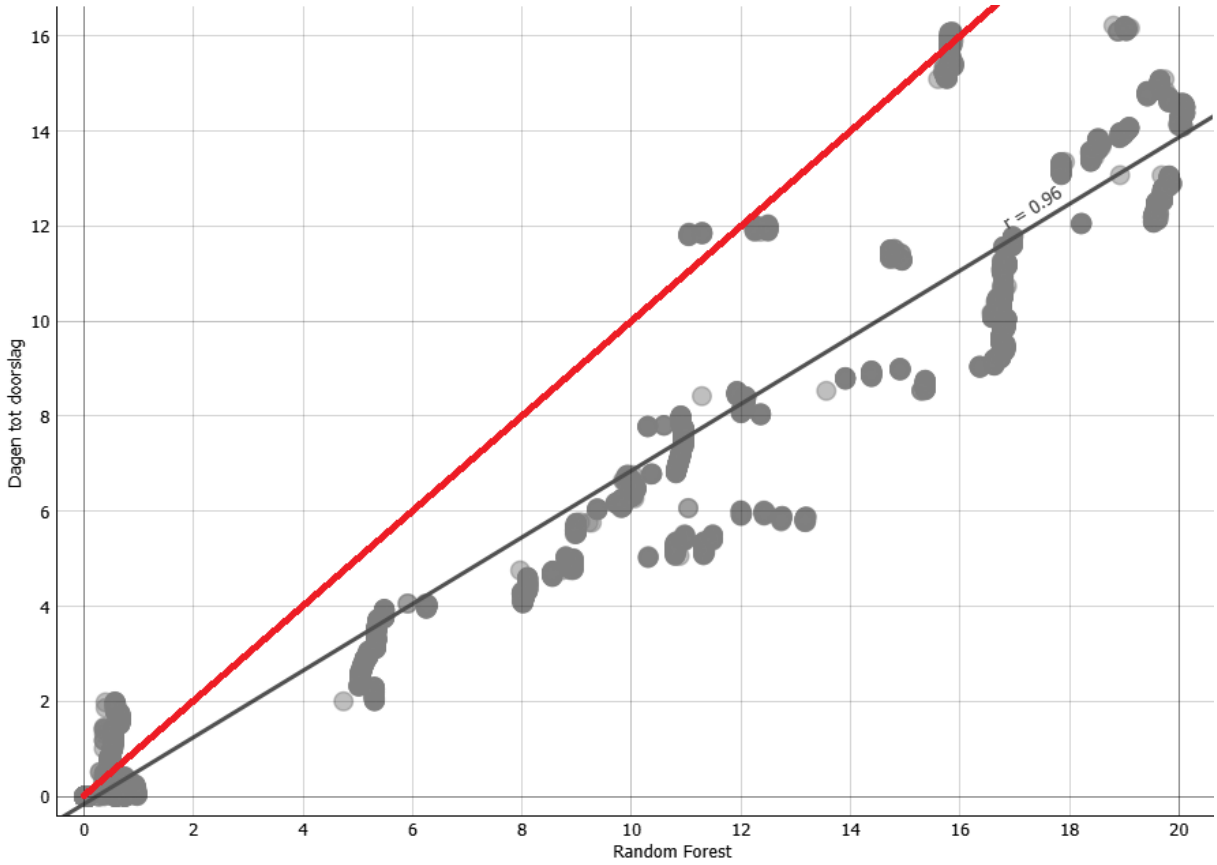
Linear Regression	35.728	5.977	5.253	-0.286	Zowel individuele voorspellingen als trendlijn consistent te hoog
AdaBoost	19.481	4.414	3.588	0.299	Voorspelling consistent te hoog, maar trendlijn behoorlijk accuraat
Neural Network	275101	524	524	-9900	Zeer inaccurate voorspellingen

Tabel 5 Test scores Orange modellen

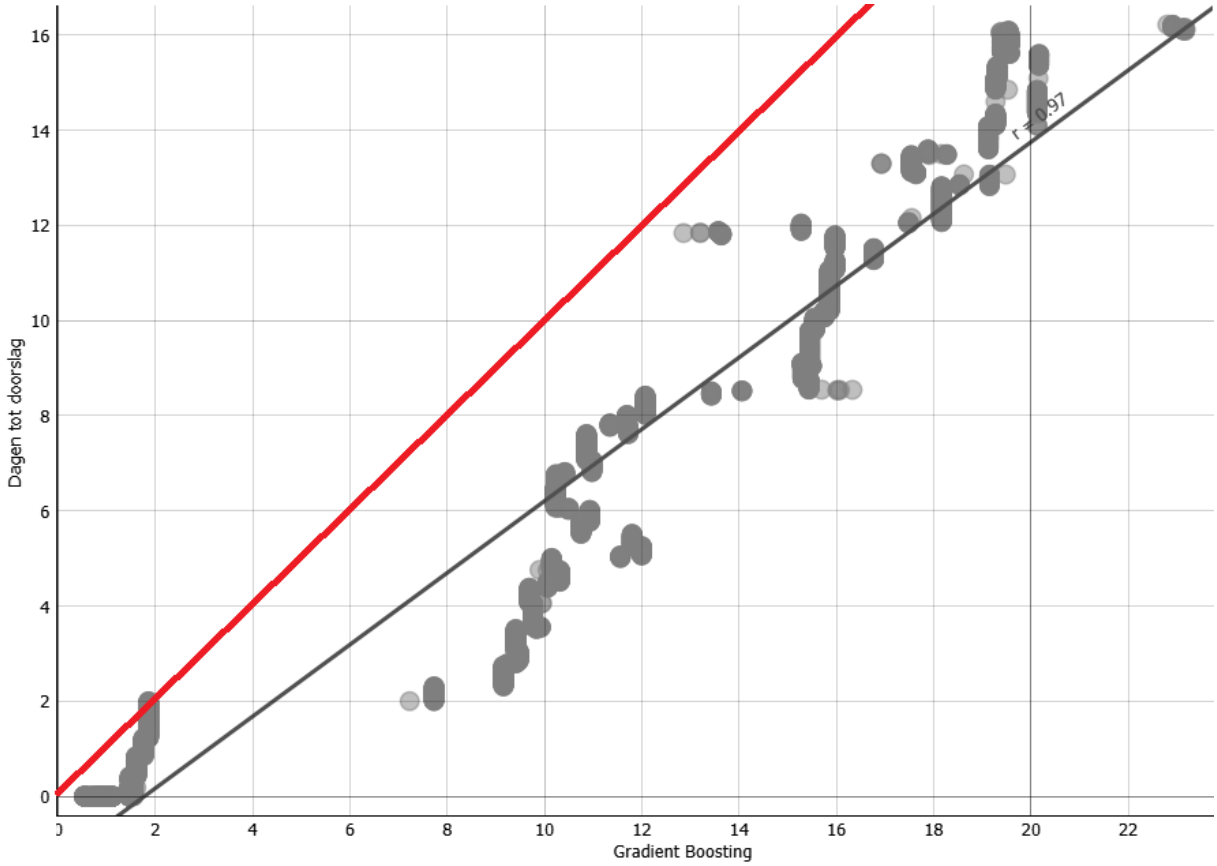
In bovenstaande tabel is een lage MSE, RMSE en MAE positief (zie Begrippen, definities en afkortingen). Hoe lager deze waardes zijn, Hoe accurater de voorspelling is. De MAE is de gemiddelde fout in de voorspelling. Met andere worden, een MAE van 3.076 betekent dat de voorspelling van de doorslag er gemiddeld iets meer dan 3 dagen naast zit. De R2 geeft aan in hoeverre de voorspellingen dezelfde lijn lijken te volgen als de werkelijke antwoorden. Een R2 van 1 betekent dat de voorspelling perfect is. In de opmerkingen wordt gesproken van een trendlijn die accuraat lijkt te zijn voor veel van deze modellen. In de volgende figuren worden van de verschillende relevante modellen (Tree (figuur 10), Random Forest (figuur 11), Gradient Boosting (Figuur 12) en AdaBoost (Figuur 13)) de trendlijn getoond. Hierbij worden er steeds twee lijnen getoond. De grijze lijn is de trendlijn die door de modellen wordt gegenereerd. In de volgende grafieken staan op de x-as het aantal dagen tot doorslag die voorspeld zijn met het model van die grafiek, en op de y-as het werkelijk aantal dagen tot doorslag. Elke punt in de scatterplot is een datapunt uit de testdata. De rode lijn is de lijn die perfect voorspellingen zou volgen (een DTD van x, en een voorspelling DTD van x).



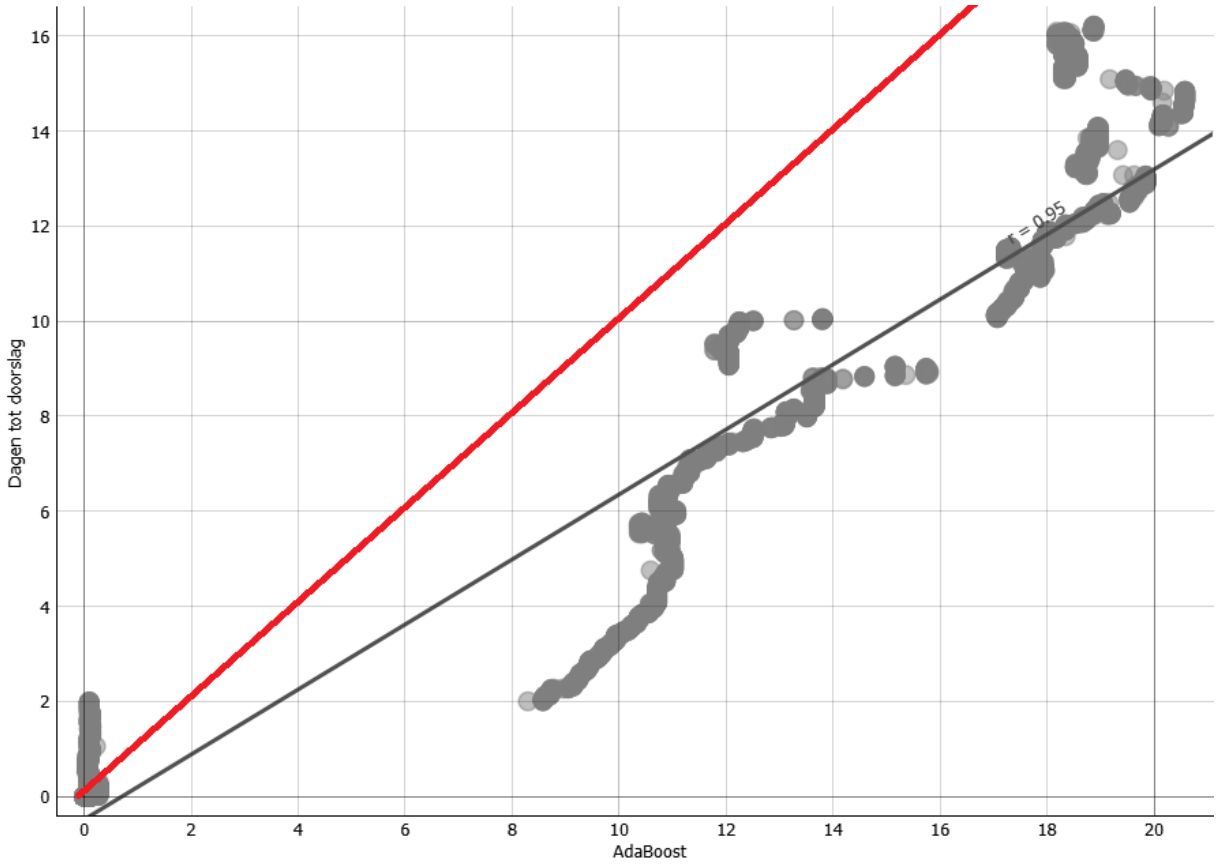
Figuur 10 Scatterplot voorspellingen met het Tree model



Figuur 11 Scatterplot voorspelling met het Random Forest model



Figuur 12 Scatterplot voorspellingen met het Gradient Boosting model



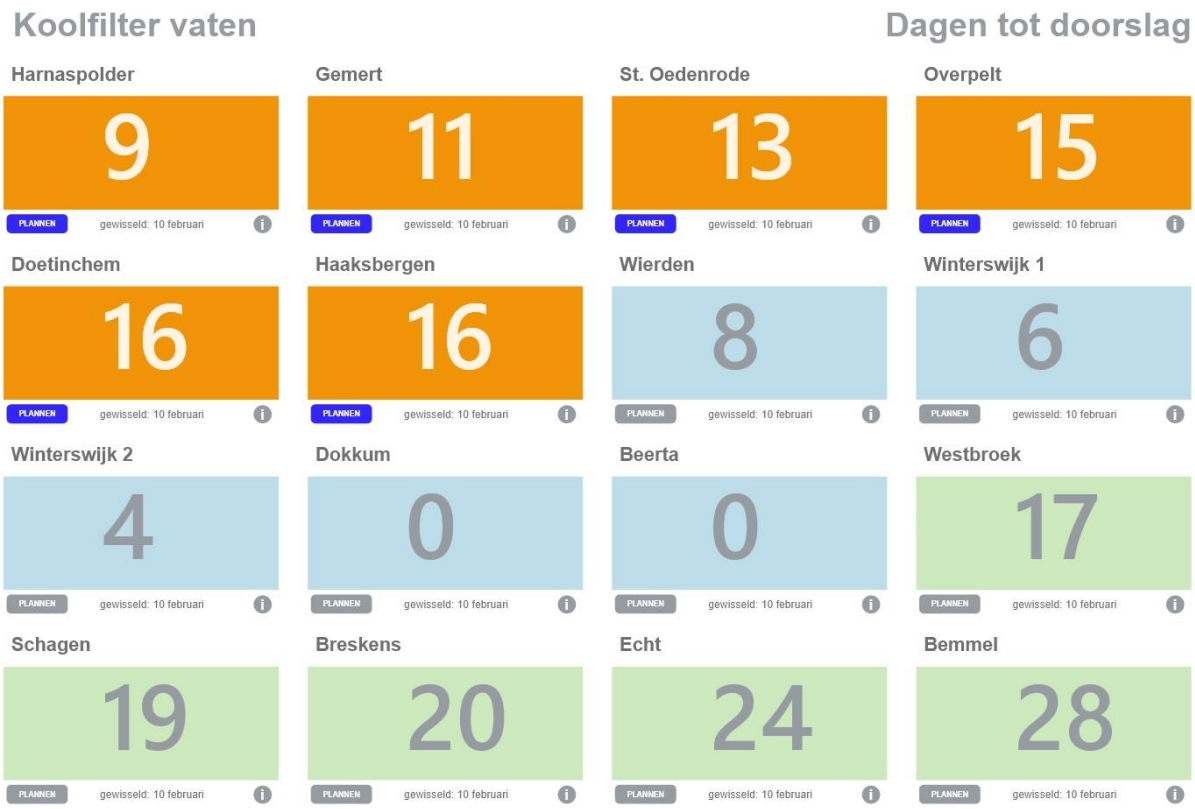
Figuur 13 Scatterplot voorspellingen met het AdaBoost model

Het is duidelijk te zien aan de verschillende modellen dat de voorspellingen twee dagen voor de doorslag zeer accuraat zijn. Dit komt doordat twee dagen voor de doorslag sporadisch een meting H_2S komt in de H2S_BT sensor. Dit is logisch, omdat die meting gezien wordt als doorslag als de sensor een waarde van boven de 5ppm meet. Het is duidelijk dat de modellen deze correlatie ook hebben gevonden. Het is echter niet op tijd om slechts twee dagen van tevoren een betrouwbare voorspelling krijgen, om een significante impact te hebben op het werkproces, omdat hiervoor minimaal een week nodig is (Kroeze, Twilhaar, & Goorhuis, Interview over mogelijke nieuwe werkwijzes, 2022). Wanneer deze data niet meegenomen wordt, zijn de trendlijnen een stuk minder accuraat. Alleen Random Forest geeft zonder deze laatste datapunten een accurate trendlijn. Dit komt overeen met de testscores uit tabel 5. Vandaar dat er is gekozen om Random Forest te gebruiken als voorspellingsmodel voor dit onderzoek.

4.5. Datavisualisatie

Uit de interviews met de stakeholders van dit proces (Kroeze, Twilhaar, & Goorhuis, Interview over mogelijke nieuwe werkwijzes, 2022) is gebleken dat de visualisatie van de data voor hen niet erg van belang is. In de kern was de wens om een datum, of een aantal mogelijke data waarin de doorslag zou kunnen plaatsvinden, te hebben. Hierbij willen ze zelf de inschatting maken welk risico ze nemen. Wel zouden ze graag de context achter die datum willen hebben.

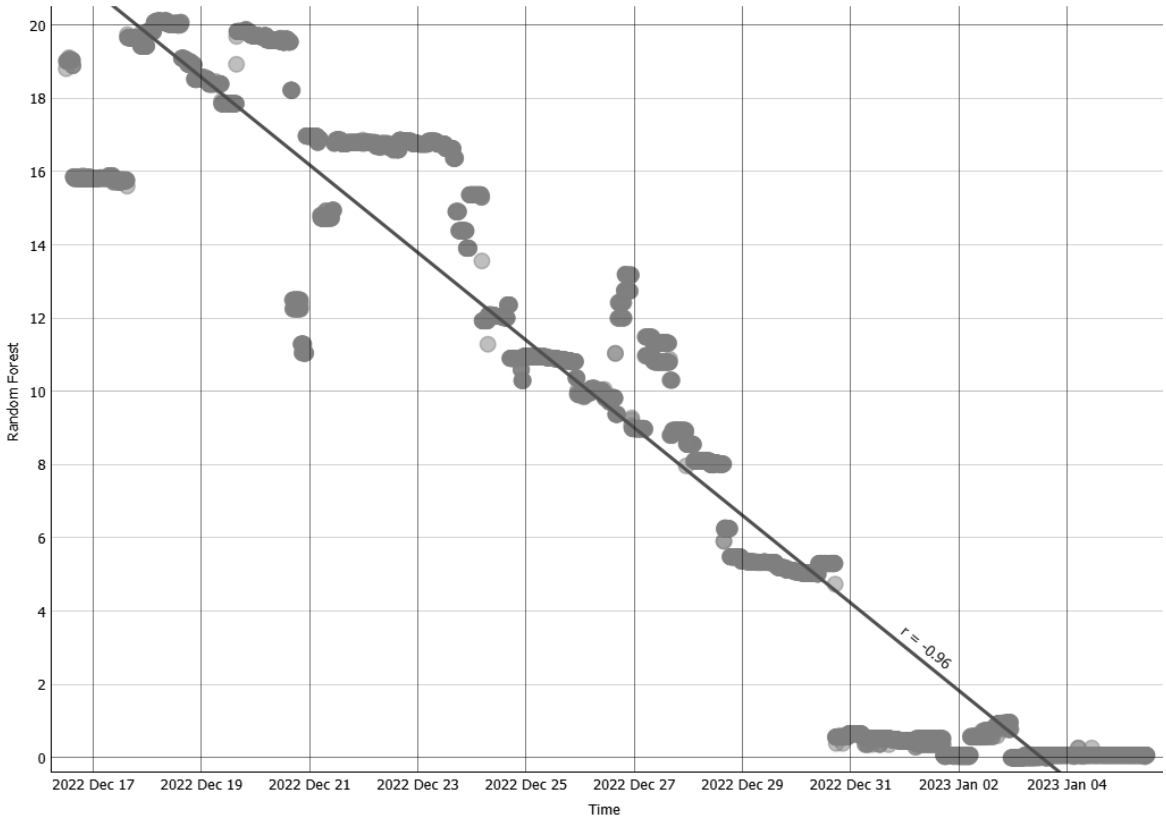
Vandaar dat een goede visualisatie om die datum en context duidelijk te krijgen wel degelijk van waarde is. Het geschatte aantal dagen tot de doorslag is het belangrijkste datapunt, dus datapunt krijgt zijn eigen visualisatie. Deze visualisatie zal samen met dezelfde visualisaties van alle andere installaties worden weergegeven in het hoofddashboard (zie figuur 14). De context achter deze graphic zal worden weergegeven in een detailsdashboard.



Figuur 14 Mockup hoofddashboard

Dit dashboard bevat 3 kleuren. Een feloranje kleur is voor installaties waar onderhoud voor moet worden gepland. Een licht blauwe is voor installaties waar onderhoud voor is gepland. Een licht groene kleur is voor installaties waar nog geen onderhoud voor hoeft worden gepland. Gezien er geen actie nodig is voor de blauwe en groene installaties zijn die kleuren subtieler, waardoor ze minder opvallen. De oranje installaties vereisen wel actie, vandaar dat er een felle kleur is gebruikt om de aandacht te trekken (Few, 2006) (Nussbaumer Knaflic, 2015). Mensen lezen (in ieder geval in het Nederlands en Engels) van links naar recht en van boven naar beneden, dus dat is ook de volgorde van de belangrijkste data punten. De installatie met de minste dagen tot doorslag staat links bovenaan. Verder zijn de knoppen voor het plannen grijs wanneer dit niet nodig is, en felblauw wanneer dit wel nodig is. Blauw contrasteert met oranje. Hoewel Nussbaumer Knaflic en Few aanraden om geen scroll te hebben, is het met de hoeveelheid installaties van Bright niet te doen om dat allemaal op één scherm te krijgen. Om te voorkomen dat dit een groot probleem is worden de installaties dus gesorteerd op de voorspelde hoeveelheid dagen tot doorslag (met uitzondering van de reeds ingeplande installaties), met andere woorden, van belangrijk naar onbelangrijk.

Dit detailsdashboard zal bestaan uit een viertal elementen. Het aantal dagen tot de doorslag zal natuurlijk onderdeel zijn van dit detailsdashboard, nu samen met de datum waarop deze doorslag zal plaatsvinden, volgens de schattingen. Een tweede onderdeel zijn de historische schattingen van de dagen tot doorslag in een scatterplot, met daarin een trendlijn getrokken die de x-as zal doorsnijden op de datum waarop de doorslag volgens de schattingen zal plaatsvinden (zie figuur 15). Hierbij zijn dus niet alleen alle inschattingen te zien, maar door de trendlijn ook een soort gemiddelde en een voorspelling richting de toekomst. Door te lezen waard de trendlijn de x-as doorsnijdt kan een datum van doorslag worden ingeschat.



Figuur 15 Datavisualisatie onderdeel scatterplot Random Forest

Voor de context achter deze data is ook de Nm^3 H_2S belangrijk, omdat dat het meest gecorreleerde datapunt is (zie tabel 2 en 3). Hiervan zal de huidige hoeveelheid worden getoond, alsook de geschatte capaciteit van de filter, gebaseerd op de historische data. Daarnaast zal het verloop van dit getal in een lijngrafiek worden weergegeven, met een staafdiagram voor de individuele dagen. Deze vier elementen samen vormen het detaildashboard (zie figuur 16).

Figuur 16 Mockup detaildashboard

Ook hier is er gekozen om de belangrijkste gegevens linksboven in de hoek te zetten, met weer een opvallende kleur oranje. Dit is om ervoor te zorgen dat de aandacht als eerste daarheen wordt getrokken (Few, 2006) (Nussbaumer Knaflic, 2015). De rest van de data is voornamelijk daar om context te bieden voor de data. Rechtsonder is er veel data te zien, vandaar dat de belangrijkste lijn daar ook een opvallende kleur krijgt.

4.6. Implementatie

4.6.1 Nieuw werkproces

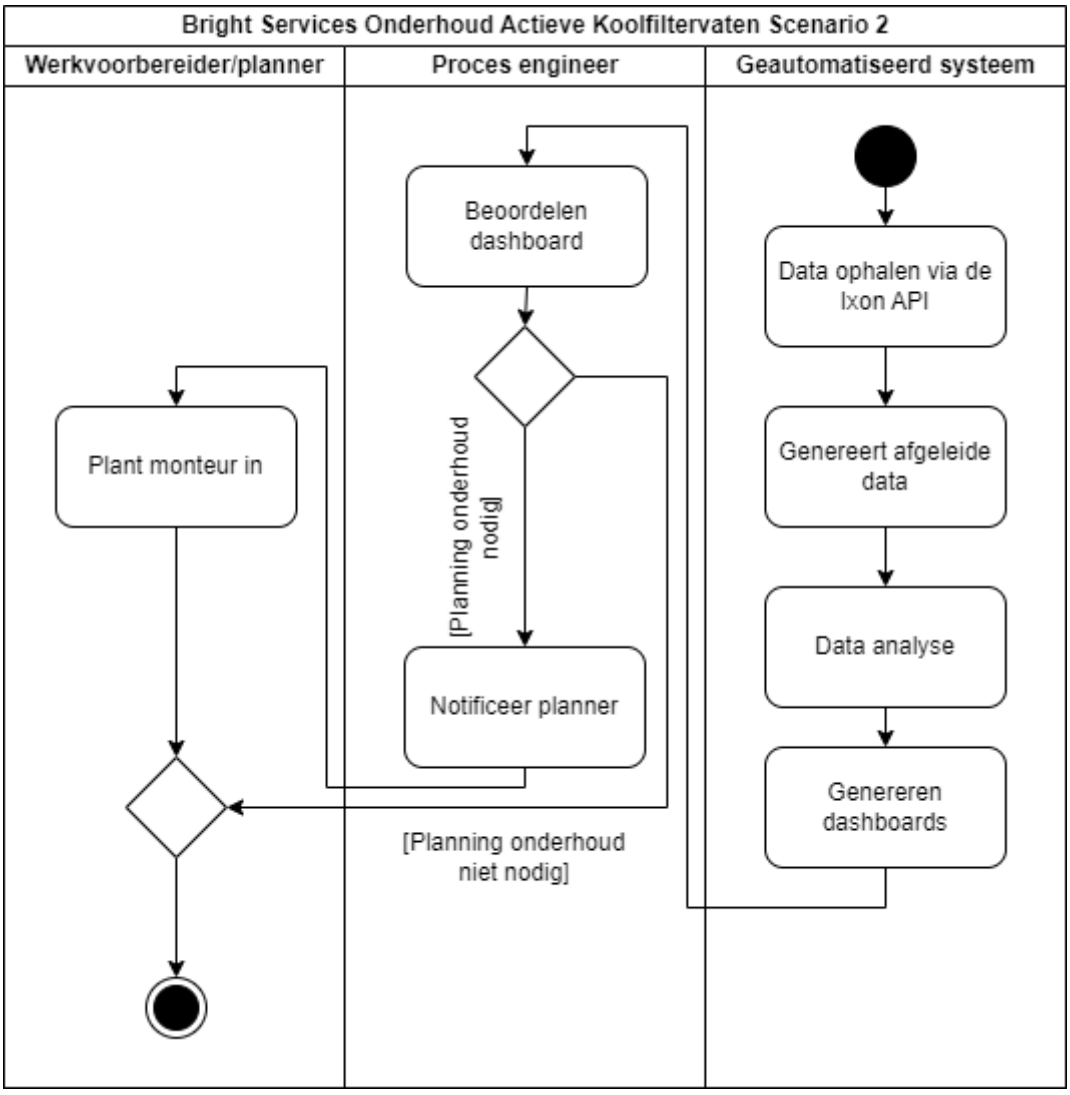
Uit de interviews met de stakeholders van dit proces (Kroeze, Twilhaar, & Goorhuis, Interview over mogelijke nieuwe werkwijzes, 2022) zijn de volgende punten naar voren gekomen.

Als eerste is de algemene mening dat de Customer Support Specialist niet nauw betrokken meer hoeft te zijn bij dit proces, als de data-analyse goed genoeg is om nauwkeurige voorspellingen te doen. Hoewel de Customer Support Specialist wel betrokken zou kunnen zijn bij het beoordelen van de data-analyses vanuit het dashboard is er door Bright aangegeven dat die verantwoordelijkheid bij voorkeur bij de proces engineer komt te liggen. Hij zal nog steeds de bevestiging van de daadwerkelijke doorslag kunnen doorgeven, maar deze bevestiging zal uiteindelijk geen invloed hebben op de planning, tenzij deze doorslag zeer veel afwijkt van de voorspelling (meer dan twee weken). Dit heeft geen significante invloed op de werkzaamheden van de Customer Support Specialist, omdat dit slechts een klein onderdeel was van zijn totale takenpakket.

Daarnaast gaf de planner aan dat hoe eerder de voorspellingen kunnen worden gedaan, hoe beter dit is voor zijn werkzaamheden. Wanneer de voorspellingen namelijk eerder bekend zijn bij de planner, wordt het makkelijker om het onderhoud in te plannen. Het streven van de planner is om twee weken

voor de doorslag met een nauwkeurigheid van vier dagen de doorslag te weten, omdat dat hem in staat stelt om één maand van tevoren een planning te maken voor de onderhoudsmonteurs. Dit omdat er, na de doorslag, nog twee weken zijn voordat de situatie kritiek wordt.

De proces engineer hoopt dat dit nieuwe proces minder werk voor hem is, omdat het huidige werkproces hem veel tijd kost en niet bij zijn expertisegebied past. Hij geeft aan dat hij hoopt dat het nieuwe proces betekent dat hij meer tijd overhoudt voor zijn andere werkzaamheden. Dit heeft geleid tot het volgende nieuwe werkproces (zie figuur 17). Op het moment dat er voldoende vertrouwen is in het systeem, zou er een geautomatiseerde notificatie vanuit het systeem rechtstreeks naar de planner kunnen, waardoor de proces engineer geen operationele betrokkenheid meer heeft.



Figuur 17 Stroomdiagram van het voorgestelde nieuwe werkproces van het inplannen van onderhoud op de actieve koolfiltervaten

4.6.2. Implementatieadvies

Voor een uitgebreide uitleg over het implementatieadvies, zie bijlage 1, hoofdstuk 6.

De implementatie kan het beste opgedeeld worden in vier fases. Deze fases zijn:

- Analysefase
- Definitiefase
- Bouwfase

- Opleveringsfase

In de analysefase worden de resultaten van dit onderzoek verder onderzocht, om te bevestigen dat dit ook voor alle andere installaties een realistisch beeld schetst. In deze fase wordt er ook een prototype gebouwd en wordt er een concept functioneel ontwerp gebouwd. Daarnaast wordt er een initiële kosten- batenanalyse gemaakt, waarna er een Go/No-Go moment wordt bereikt, waar besloten wordt of dit een kosteneffectief plan is.

In de definitiefase worden de wensen en eisen van de eindproducten opgesteld, alsook een inventarisatie gemaakt van de benodigde mensen en middelen. Hierbij wordt er niet alleen gekeken naar de interne beschikbaarheid, maar wordt er ook meegenomen welke uitbestedingsopties er zijn. Daarnaast wordt in deze fase een planning gemaakt. Ook wordt er hier bepaald welke producten er worden gebruikt om aan deze wensen en eisen te voldoen. Zo wordt er gekeken of er met Open Source software kan worden gewerkt, of dat er gebruik wordt gemaakt van een commercieel pakket zoals Azure Machine Learning.

In de bouwfase worden de eindproducten ontwikkeld. Gezien de huidige IT kennis binnen HoSt en Bright is het zeer waarschijnlijk dat hier een externe partij voor zal worden ingeschakeld. Er zal wel intern een productowner nodig zijn en een interne of onafhankelijke externe projectmanager. Daarnaast zal er in de bouwfase ook worden getest. Dit zal voornamelijk door de externe partij gebeuren, maar het is aan te raden om met enige regelmaat ook intern te testen, zodat er eventueel kan worden bijgestuurd.

In de opleveringsfase wordt de formele goedkeuring van de acceptatieomgeving gegeven door de verantwoordelijke stakeholders, waarna de omgeving wordt gemigreerd naar een productieomgeving. Tijdens deze fase zullen er ook instructies en trainingen komen voor de eindgebruikers. Er kunnen hierna nog steeds verdere ontwikkelingen worden gedaan.

4.7. Verandermanagement

Verandermanagement is gedaan met behulp van het ADKAR-model.

4.7.1. Bewustzijn (Awareness)

De medewerkers lijken zich al direct bewust te zijn van de noodzaak van de verandering. De proces engineer ergert zich enorm aan de hoeveelheid werk die hij nu extra moet doen om de schattingen te maken voor de planner, terwijl de planner zelf geïrriteerd is over de korte tijd die hij heeft om de planningen te maken. De customer support engineer heeft weinig last van de huidige situatie, maar merkt de frustratie bij zijn collega's.

4.7.2. Verlangen (Desire)

De proces engineer is degene die dit onderzoek heeft aangevraagd. Hij is bewust van de mogelijkheden van predictive maintenance en is daar ook zeer enthousiast over. HoSt is zelf een zeer innovatief bedrijf, waardoor dit soort innovaties ook erg goed passen in de cultuur van HoSt. Het verlangen bij de planner en de customer support engineer is minder groot dan bij de proces engineer, maar zij zullen uiteindelijk alleen verbeteringen ervaren, waar de proces engineer een significante verandering in zijn takenpakket krijgt.

4.7.3. Besef (Knowledge)

Alle medewerkers die betrokken zijn bij dit proces lijken zich zeer bewust van het einddoel. Zij worden hier ook enthousiast van. Ze zullen wel moeten leren omgaan met de nieuwe methodieken. De planner en de customer support engineer zullen niet veel extra kennis nodig hebben om hun werkzaamheden

uit te blijven voeren, maar zullen wel meegenomen worden in het proces en de uitleg. Zij zullen ook betrokken worden bij de presentatie van dit onderzoek.

De proces engineer zal moeten leren te vertrouwen op de data die gepresenteerd wordt door het systeem, en zal moeten leren om de data juist te interpreteren. Hiervoor zal hij extra kennis op het gebied van data-analyse moeten opdoen. Hij geeft aan bereid te zijn deze kennis tot zich te nemen, in de vorm van een cursus.

4.7.4. Vermogen (Ability)

De proces engineer zal verantwoordelijk worden voor het interpreteren van de data-analyse. In het begin zal deze interpretatie gecontroleerd moeten worden, doordat de customer support engineer het oude systeem als back-up in de gaten blijft houden, iets wat sowieso al gebeurt voor zijn andere werkzaamheden. Hierdoor is het mogelijk voor de proces engineer om zijn nieuwe vergaarde kennis te toetsen in de praktijk, zonder dat eventuele fouten direct ernstige gevolgen hebben. Verder is het verstandig dat dit nieuwe systeem in het begin op een beperkt aantal installaties wordt gebruikt, totdat de data-analyse en de interpretatie van de proces engineer erg betrouwbaar zijn.

4.7.5. Bekrachtiging (Reinforcement)

Op het moment dat de data-analyse zeer nauwkeurig is, kan het zijn dat de proces engineer niet meer nodig is voor het interpreteren van de data. Wanneer dit het geval is, kan het proces volledig worden geautomatiseerd (in dit geval hoeft alleen de planner nog actie te ondernemen, zodra het systeem automatisch een melding stuurt). Dit zal dan de nieuwe norm worden, waarbij de proces engineer alleen nodig is voor nieuwe installaties om het proces naar volledige automatisering te begeleiden.

5. Conclusie

In dit onderzoek is gezocht naar het antwoord op de vraag: “Hoe kan Bright op basis van data-analyses een predictive maintenance model opzetten voor de actieve koolfiltervaten?”. Hiervoor zijn er verschillende data-analyses getest om te kijken of deze geschikt zijn voor predictive maintenance.

Uit de resultaten blijkt dat het lastig is om voorspellingen te maken die ruim van tevoren zeer nauwkeurig zijn. Er is wel één data-analysemodel te vinden dat enigszins in de buurt komt van accurate voorspellingen, namelijk Random Forest. Dit model is alleen consistent te laat met zijn voorspelling (het voorspelt dat de doorslag later komt dan in werkelijkheid). De voorspellingen zijn twee dagen voor de doorslag van de H_2S-meter na het eerste actieve koolfiltervat zeer accuraat, maar dit is voor de behoeften van Bright te laat.

Wanneer de trendlijn van deze voorspellingen echter wordt doorgetrokken richting de 0 dagen tot de doorslag, lijkt deze voorspelling behoorlijk accuraat en bruikbaar. Hier is wel veel data voor nodig, wat weer leidt tot het punt dat de voorspelling aan de late komt.

Dit betekent niet dat er geen nuttige conclusies uit de data-analyse kunnen worden getrokken. Omdat de voorspelling de doorslag steeds later schat dan hij in werkelijkheid is, kan de proces engineer de voorspellingen op zodanige manier interpreteren dat er een periode kan worden gegeven waarin het waarschijnlijk is dat de doorslag plaatsvindt. Dit kan in het dashboard systeem meegenomen worden als een correctiefactor. Dit, gecombineerd met het feit de dag waarop het onderhoud moet worden gepleegd niet extreem nauwkeurig hoeft te zijn - een marge van vijf dagen is prima - betekent dat, met behulp van de data-analyse, het onderhoud ongeveer een week eerder kan worden ingepland dan zonder de data-analyse.

6. Discussie

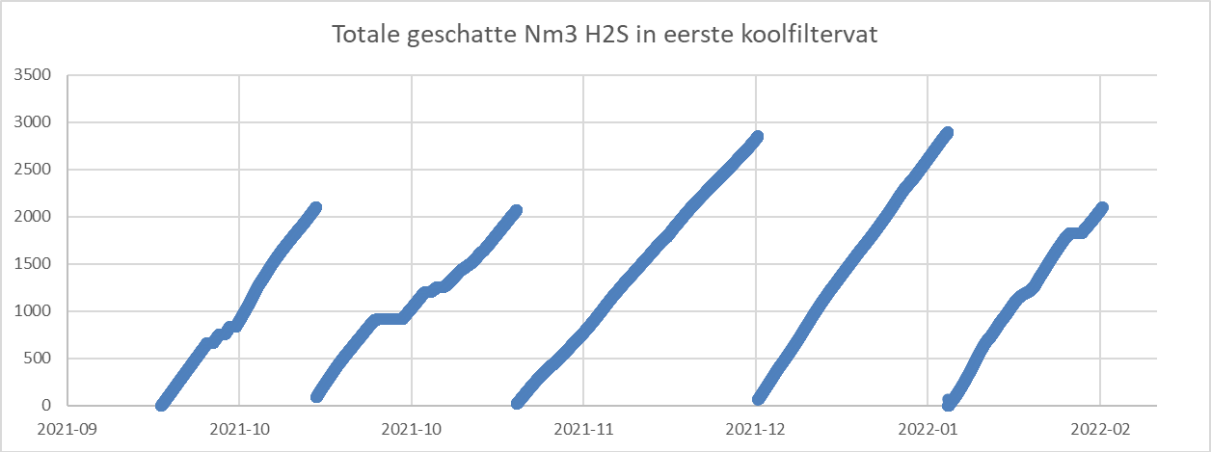
Dit onderzoek heeft een poging gedaan om de vraag “Hoe kan Bright op basis van data-analyses een predictive maintenance model opzetten voor actieve koolfiltervaten?” te beantwoorden. De conclusie was dat, hoewel de voorspellingen niet zo accuraat zijn als van tevoren gehoopt werd, er wel degelijk een voorspelling kon worden gemaakt voor de actieve koolfiltervaten van een van de installaties van Bright, die nuttig is voor het inplannen van het onderhoud.

Hier moeten wel een aantal kanttekeningen bij worden geplaatst. Als eerste heeft dit onderzoek plaatsgevonden bij één installatie, en één type onderhoud (namelijk, het vervangen van de actieve koolfiltervaten). Dat de conclusie voor deze installatie en dit onderhoud voorzichtig positief is, betekent niet dat deze modellen te gebruiken zijn voor alle installaties en alle soorten onderhoud die gepleegd moeten worden door Bright. Hiervoor zal een breder onderzoek nodig zijn bij meerdere installaties.

Daarnaast is de geautomatiseerde data-acquisitie die op het moment kan plaatsvinden niet voldoende om de benodigde data op te halen uit de systemen. Dit betekent dat er moet worden gewerkt met de handmatige methode, die veel meer tijd kost, om deze data-analysedata analyses dagelijks uit te kunnen voeren. Het is dus niet zeker of predictive maintenance op dit moment op een kosteneffectieve manier kan worden uitgevoerd. In overleg met Ixon kan dit wellicht worden opgelost.

Verder was de opgehaalde data niet altijd even nauwkeurig. Er was vaak sprake van meetfouten. Voor veel van deze datapunten zijn uiteindelijk schattingen gemaakt om de data-analyse alsnog te kunnen uitvoeren, maar deze schattingen zijn niet nauwkeurig. Hierdoor zal ook de data-analyse niet nauwkeurig zijn. Mocht er veel meer data beschikbaar zijn, dan zullen de meetfouten en de schattingen minder invloed hebben op het totale model en zullen de voorspellingen nauwkeuriger worden.

Er zijn ook een aantal andere inzichten naar voren gekomen tijdens dit onderzoek die niet binnen de scope van dit specifieke onderzoek vielen, maar wel interessant zijn om verder te onderzoeken. Zo was het bekend dat er factoren zijn die van invloed zijn op de totale capaciteit van de koolfiltervaten, maar wat opvallend was, was dat de capaciteit van de verschillende vaten niet evenredig was verdeeld. In plaats daarvan bleken ze allemaal rond een bepaald aantal vaste punten te liggen (zie figuur 18). Dit betekent dat onderzoek naar de factoren die deze capaciteit beïnvloeden zou kunnen helpen om de voorspellingen nauwkeuriger te maken. Daarnaast zou het kunnen dat de factoren te beïnvloeden zijn, waardoor de capaciteit van de koolfiltervaten kan worden verhoogd.



Figuur 18 Totale geschatte hoeveelheid (in Nm^3) H_2S in het eerste koolfiltervat

7. Aanbevelingen

In dit hoofdstuk zullen de aanbevelingen naar aanleiding van dit onderzoek worden uiteengezet. Als eerste zal er worden beschreven welke bedrijfsmatige veranderingen voordelig kunnen zijn voor Bright en daarna zal worden ingegaan op mogelijke vervolgonderzoeken die Bright kan uitvoeren.

7.1. Bedrijfsmatige veranderingen

Naar aanleiding van dit onderzoek zijn er een aantal bedrijfsmatige veranderingen die Bright kan doorvoeren om de bevindingen van dit onderzoek in de praktijk toe te kunnen passen.

Als eerste is het van belang dat het bedrijfsproces, zoals beschreven in paragraaf 4.6, wordt doorgevoerd. Hiervoor is het wel nodig om een data-analist beschikbaar te hebben en een dashboard te ontwikkelen. Momenteel lijkt de proces engineer de meest geschikte kandidaat om dit proces te begeleiden, maar omdat dit een significante hoeveelheid extra werk zal opleveren voor hem, zou dat betekenen dat hij een deel van zijn huidige takenpakket zou moeten overdragen aan een collega. In elk geval zal de proces engineer betrokken blijven bij het huidige proces, en is het voor hem nuttig om meer kennis over data-analyses op te doen. Hiermee kan hij de kwaliteit van de oplossing beter beoordelen.

Het is daarnaast nodig om de huidige functionaliteiten van Ixon te verbeteren, zodat de belangrijke data wel via de API van Ixon is op te halen. Momenteel is de IT-kennis binnen Bright en HoSt niet afdoende, waardoor het maar de vraag is of dit op het moment mogelijk is. Het is wellicht mogelijk dit uit te besteden aan Ixon, maar het blijft van belang dat Bright en HoSt zo snel mogelijk de IT-kennis op een hoger niveau brengen, bijvoorbeeld met een applicatiebeheerder. De applicatiebeheerder zou niet alleen kunnen helpen met Ixon, maar zal ook nuttig kunnen zijn in het ondersteunen van de IT-innovaties die Bright en HoSt momenteel aan het doorvoeren zijn, zoals de data lake, en het nieuwe ERP-systeem.

7.2. Vervolgonderzoek

Naar aanleiding van dit onderzoek zijn er een aantal vervolgonderzoeken die plaats kunnen vinden die voor Bright nuttig kunnen zijn.

In de discussie is een van de mogelijke vervolgonderzoeken al naar voren gekomen, namelijk een onderzoek naar de factoren die van invloed zijn op de capaciteit van de actieve koolfiltervaten. Hoewel dit onderzoek gebruik kan maken van data-analyses, zal de uitvoering beter gedaan kunnen worden door een chemisch technoloog.

Het is daarnaast nuttig om te onderzoeken in hoeverre de resultaten van dit onderzoek te reproduceren zijn voor de overige installaties van Bright. Hierbij kan niet alleen worden gekeken of er voor elke installatie een model te maken is, maar ook kan er worden gekeken of er een algemeen model kan worden gemaakt dat werkt voor elke installatie. Dat zou namelijk betekenen dat er ook predictive maintenance kan worden gepleegd op nieuwe installaties waar nog geen, of zeer weinig data voor beschikbaar is.

7.3. Prioriteiten

Van deze aanbevelingen is het aan te raden om als eerste het onderzoek naar de reproduceerbaarheid van dit onderzoek uit te voeren. Zolang het niet duidelijk is of dit onderzoek daadwerkelijk te reproduceren is voor alle installaties, is het niet nuttig om het bedrijfsproces in paragraaf 4.6 in te voeren.

Het op peil brengen van de IT-kennis binnen HoSt is belangrijk. Zonder deze kennis is de afhankelijkheid van derden zeer groot.

Bibliografie

Amala, G. (2019). Orange Tool Approach For Comparative Analysis of Supervised Learning Algorithm in Classifaction Mining. *Journal of Analysis and Computation*, 1-10.

Bright Renewables. (2023, 3 23). *Onze biogas naar groen gas membraantechnologie*. Retrieved from Bright Renewables: https://www.bright-renewables.com/nl/biogasopwerkingstechnologie/

Carvalho, T. P., Soares, F. A., Vita, R., Francisco, R. d., Basto, J. P., & Alcalá, S. G. (2019, November). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering, 137*, 3.

Few, S. (2006). *Information Dashboard Design; The Effective Visual Communication of Data.* Sebastopol, CA: O'Reilly Media, Inc.

HoSt Holding. (2022, mei 25). *Over HoSt | #1 in bioafval naar bio-energie*. Retrieved from HoSt: https://www.host.nl/nl/over-host-nr-1-in-afval-naar-bio-energie/

ICT research methods. (2022, juni 1). *Methods*. Retrieved from ICT Research Methods: https://ictresearchmethods.nl/Methods

Ixon. (2022, Oktober 21). *Endpoint List: DataList*. Retrieved from Ixon API Developer: https://developer.ixon.cloud/reference/post_data

Ixon. (2022, 10 1). *How to use teh APIv2?* Retrieved from Ixon Developer: https://developer.ixon.cloud/docs/how-to-use-the-apiv2#api-request-architecture

Kroeze, V. (2022, Augustus 12). Handmatige data acquisitie. (L. van den Broek, Interviewer)

Kroeze, V. (2022, Juni 6). Huidige werkproces Bright Services onderhoud actieve koolfiltervaten. (L. A. van den Broek, Interviewer)

Kroeze, V. (2022, Juni 22). Relevante data uit de Ixon API. (L. A. van den Broek, Interviewer)

Kroeze, V. (2022, September 12). Toekomstige ICT architectuur van HoSt en haar dochterondernemingen. (L. A. van den Broek, Interviewer)

Kroeze, V. (2022, 5 25). Werking Biogasopwerkingsinstallatie. (L. van den Broek, Interviewer)

Kroeze, V., Twilhaar, B., & Goorhuis, E. (2022, December 9). Interview over mogelijke nieuwe werkwijzes. (L. van den Broek, Interviewer)

Li, H. (2020, 12 9). *Which machine learning algorithm should I use?* Retrieved from SAS Blogs: https://blogs.sas.com/content/subconsciousmusings/2020/12/09/machine-learning-algorithm-use/

Master's in Data Science. (2023, 3 23). *What is ARIMA Modeling?* Retrieved from Master's in Data Science: https://www.mastersindatascience.org/learning/statistics-data-science/what-is-arima-modeling/

Microsoft. (2023, 3 20). *CORREL function*. Retrieved from Microsoft Support: https://support.microsoft.com/en-us/office/correl-function-995dcef7-0c0a-4bed-a3fb-239d7b68ca92

Miloslavskaya, N., & Tolstoy, A. (2016). Big Data, Fast Data and Data Lake Concepts. *Procedia Computer Science, 88*, 300-305.

Mulder, P., & Janse, B. (2022, 2 25). *ADKAR model*. Retrieved from Toolshero: https://www.toolshero.nl/verandermanagement/adkar-model/

Naik, A., & Samant, L. (2016). Correlation review of classification algorithm using data mining tool: WEKA, Rapidminer, Tanagra, Orange and Knime. *Proocedia Computer Science*(85), 662-668.

Nussbaumer Knaflic, C. (2015). *Storytelling with Data; a Data Visualization Guide for Business Professionals.* Hoboken: Wiley.

Orange. (2022, 12 21). *AdaBoost*. Retrieved from Orange Data Mining: https://orangedatamining.com/widget-catalog/model/adaboost/

Orange. (2022, November 11). *CSV File Import*. Retrieved from Orange datamining: https://orangedatamining.com/widget-catalog/data/csvfileimport/

Orange. (2022, 12 21). *Gradient Boosting*. Retrieved from https://orangedatamining.com/widget-catalog/model/gradientboosting/: https://orangedatamining.com/widget-catalog/model/gradientboosting/

Orange. (2022, 12 21). *Random Forest*. Retrieved from Orange Data Mining: https://orangedatamining.com/widget-catalog/model/randomforest/

Orange. (2022, 12 21). *Tree*. Retrieved from Orange Data Mining: https://orangedatamining.com/widget-catalog/model/tree/

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science, 181*, 526-534.

Schuh, G., Reinhart, G., Prote, J.-P., Sauermann, F., Horsthofer, J., Oppolzer, F., & Knoll, D. (2019). Data Mining Definitions and Applications for the Management of Production Complexity. *Procedia CIRP, 81*, 874-879.

Sydorenko, I. (2021, 5 3). *How to Choose the Right Machine Learning Algorithm: A Pragmatic Approach*. Retrieved from Label Your Data: https://labelyourdata.com/articles/how-to-choose-a-machine-learning-algorithm#unsupervised_ml_algorithms

Twilhaar, B. (2021). *Onderhoud onderhoud.* Enschede: Saxion.

Wikiwijs. (2022, juni 1). *DOT-framework Vijf onderzoeksstrategieen (Hoe)*. Retrieved from Wikiwijs: https://maken.wikiwijs.nl/127721/DOT_framework#!page-4502907