

Scriptie



Afstudeerproject Software Engineering Hogeschool Windesheim

*HBO-ICT Voltijd
Software Engineering
Bachelor of Science (BSc)*

Opdracht: Smart Indoor Climate System (SICS)

Organisatie: Ultraware Consultancy and Development B.V.

Praktijkplaats: Assen

Opdrachtgever: Serge de Mul

Bedrijfsmentor: Martijn Gaasbeek

Schoolbegeleider: Wim Schepers

Versie: 1.2

Datum: 18 mei 2020

Student: Matthieu van Bekkum

Studentnummer: 1103729

Toestemming voor publicatie is afgegeven door Ultraware op 11 juni 2020.

VERSIEBEHEER

Versie	Datum	Beschrijving
0.1	27 januari 2020	Aanleiding & opzet document.
0.2	5 februari 2020	Opstellen onderzoeksvragen.
0.3	6 maart 2020	Aanpassen onderzoeksvragen aan huidige situatie.
0.4	10 maart 2020	Herformulering hoofd- en deelvragen met feedback Martijn.
0.5	24 maart 2020	Uitwerking eerste 2 deelvragen.
0.6	8 april 2020	Uitwerking deelvraag 3. Specificatie onderzoeksmethoden.
0.7	9 april 2020	Uitwerking lineair & niet-lineair probleem.
0.8	14 april 2020	Theoretisch onderzoek & intuïtie niet-lineaire classificatiealgoritmen.
0.9	22 april 2020	Algoritmeselectie baseren op factoren. Beoordeling toevoegen met beoordelingsschaal. Aanpassen deelconclusie 3.4.
1.0	24 april 2020	Verwerken feedback Wim Schepers.
1.1	6 mei 2020	Schrijven eindconclusie, discussie, aanbevelingen en samenvatting.
1.2	18 mei 2020	Toevoeging wat betreft de rol van het verminderen van het energieverbruik in het systeem.

Tabel 1. Versiebeheer.

VOORWOORD

Deze scriptie is opgesteld in het kader van de afstudeeropdracht genaamd Smart Indoor Climate System, een opdracht waar aan het gelijknamige systeem is gebouwd. In deze scriptie worden de mogelijkheden van het toepassen van kunstmatige intelligentie voor het voorverwarmen van de SIC-ruimte van Ultraware met de warmtepomp onderzocht. De SIC-ruimte is een ruimte in het bedrijfspand waar met sensoren wordt geëxperimenteerd. Dit onderzoek richt zich op het gebruiken van de data van die sensoren met de toepassing van kunstmatige intelligentie.

Het onderzoek is uitgevoerd door Matthieu van Bakkum voor de bacheloropleiding Software Engineering aan de Hogeschool Windesheim.

SAMENVATTING

Deze scriptie is opgezet als ondersteuning voor de ontwikkeling van het Smart Indoor Climate Project. Het doel is om kunstmatige intelligentie in te zetten bij het voorverwarmen van een bedrijfspand en het energieverbruik te minimaliseren met behoud van comfort. Maar hoe kan dit worden gerealiseerd? Deze scriptie toont het onderzoek wat uitgevoerd is om te bepalen welke vorm van kunstmatige intelligentie geschikt is, welke data er beschikbaar en relevant is, en hoe kunstmatige intelligentie uiteindelijk kan bijdragen aan deze bedrijfsdoelstelling.

De belangrijkste conclusie van deze scriptie is dat de algoritmen K-Nearest Neighbors en Naive Bayes het meest geschikt en toepasbaar zijn voor het toepassen van kunstmatige intelligentie bij het voorverwarmen van het bedrijfspand en het behalen van de bedrijfsdoelstelling.

Deze algoritmen kunnen door middel van Scikit-learn worden geïmplementeerd en het systeem en kunnen zo met data van eerder gedane metingen beslissingen nemen over de modus van de warmtepomp voor een nieuwe dag.

Ook het vergelijken van de resultaten van beide algoritmen en het bepalen van de meest energiezuinige optie hieruit blijkt uit het onderzoek, de zogenaamde combi-functionaliiteit.

Voor een preciezere uitgebreidere inhoudelijke samenvatting van de scriptie wordt verwezen naar de eindconclusie. Hier wordt men in een vogelvlucht meegenomen door het gehele onderzoek en worden de onderzoeksresultaten getoond.

INHOUDSOPGAVE

Versiebeheer	2
Voorwoord.....	3
Samenvatting.....	4
1. Figuren- en tabellenlijst.....	7
1.1. Figuren.....	7
1.2. Tabellen	7
2. Begrippen & afkortingen	8
3. Inleiding	9
4. Probleemstelling.....	10
5. Afbakening.....	11
5.1. Hoofdvraag	11
5.2. Deelvragen.....	11
6. Literatuurverkenning.....	13
6.1. Bachelorscriptie van medewerker Ultraware.....	13
6.2. Machine Learning A-Z™: Hands-On Python & R In Data Science van Udemy	13
7. Onderzoeksmethoden	14
7.1. Literatuuronderzoek.....	14
7.2. Exploratief onderzoek	14
8. Onderzoeksresultaten	15
8.1. Deelvraag 1: Hoe kan de warmtepomp worden aangestuurd?	15
8.1.1. Standen warmtepomp.....	15
8.1.2. Modussen warmtepomp	15
8.1.3. Verbruik standen warmtepomp	16
8.1.4. Verbruik modussen warmtepomp.....	16
8.1.5. Deelconclusie deelvraag 1	16
8.2. Deelvraag 2: Wat is de voor kunstmatige intelligentie relevante data en is deze beschikbaar?	17
8.2.1. Beschikbare data	17
8.2.2. Relevante data.....	17
8.2.3. Missende data	18
8.2.4. Omzetten beschikbare data naar relevante data.....	18
8.2.5. Deelconclusie deelvraag 2	18
8.3. Deelvraag 3: Welke vorm van kunstmatige intelligentie kan het beste worden toegepast?	19
8.3.1. Deelvraag 3.1: Welke vormen zijn er?.....	19
8.3.2. Deelvraag 3.2: Wat is de relatie tussen de attributen in de dataset?.....	20
8.3.3. Deelvraag 3.3: Welke soorten algoritmen zijn er en welke zijn relevant?	23

8.3.4.	Deelvraag 3.4: Welk(e) algoritme(n) is/zijn het best toepasbaar?	26
8.4.	Deelvraag 4: Hoe kan met deze kunstmatige intelligentie het energieverbruik worden verminderd?	34
8.4.1.	Impliciete vermindering energieverbruik via modus	34
8.4.2.	Combi-functionaliteit.....	34
8.4.3.	Deelconclusie deelvraag 4	34
8.5.	Deelvraag 5: Hoe kan kunstmatige intelligentie geïmplementeerd worden in het systeem?	35
8.5.1.	Deelvraag 5.1: Hoe kunnen de toepasbare algoritmen worden geïmplementeerd?	35
8.5.2.	Deelvraag 5.2: Hoe kunnen de resultaten van deze algoritmen worden geverifieerd?	36
9.	Eindconclusie	37
10.	Discussie	38
10.1.	Bepalen beoordelingscriteria algoritmeselectie.....	38
10.2.	Uitsluiten omgevingsfactoren	38
11.	Aanbevelingen	39
11.1.	Meenemen omgevingsfactoren	39
11.2.	Valideren resultaten algoritmen d.m.v. koppeling warmtepomp.....	39
11.3.	Uitbreiden keuze algoritmen.....	39
12.	Literatuurlijst	40

1. FIGUREN- EN TABELLENLIJST

1.1.FIGUREN

Figuur 2. Verschil tussen kunstmatige intelligentie, machine learning en deep learning (Touger, 2018).....	19
Figuur 3. Verschil tussen supervised en unsupervised learning (Seif, 2018).	23
Figuur 4. Voorbeeld logaritmisch verband temperatuur met delta tijd bij een constante stand van de warmtepomp.	25
Figuur 5. Visualisatie voorbeeld K-Nearest Neighbors algoritme (Eremenko & de Ponteves, 2020).	27
Figuur 6. Visualisatie voorbeeld kernel SVM-algoritme (Eremenko & de Ponteves, 2020).	28
Figuur 7. Visualisatie voorbeeld Naive Bayes algoritme (Eremenko & de Ponteves, 2020).	29
Figuur 8. Visualisatie voorbeeld beslissingsboom van het decision tree classificatiealgoritme (Eremenko & de Ponteves, 2020).	30
Figuur 9. Visualisatie voorbeeld verdeling data door decision tree classificatiealgoritme (Eremenko & de Ponteves, 2020).	30
Figuur 10. Visualisatie voorbeeld random forest classification (Abilash, 2018).	32

1.2.TABELLEN

Tabel 1. Versiebeheer.....	2
Tabel 2. Begrippen & afkortingen.	8
Tabel 3. Standen warmtepomp (Ultraware, 2019).	15
Tabel 4. Modussen van de warmtepomp voor het voorverwarmen tussen 6 en 9 (TextMechanic, 2018)....	15
Tabel 5. Verbruik per stand warmtepomp (Gaasbeek, 2020).	16
Tabel 6. Verbruik per modus warmtepomp.	16
Tabel 7. Beknopt overzicht beschikbare data (Tamminga, 2020).	17
Tabel 8. Voorbeeld overzicht relevante data.	17
Tabel 9. Voorbeeld overzicht relevante data.	20
Tabel 10. X- en Y-waarden voor machine learning model.	22
Tabel 11. Beoordelingsschaal per factor algoritme.....	26
Tabel 12. Beoordelingsschaal algoritme.....	26
Tabel 13. Beoordeling K-Nearest Neighbor.....	27
Tabel 14. Beoordeling Kernel Support Vector Machine (Kernel SVM).....	28
Tabel 15. Beoordeling Naive Bayes.	29
Tabel 16. Beoordeling Decision Tree Classification / beslissingsboomalgoritme.	31
Tabel 17. Beoordeling Random Forest Classification.	32
Tabel 18. Ranglijst toepasbare algoritmen op basis van theoretisch onderzoek.....	33

2. BEGRIPPEN & AFKORTINGEN

Begrip/Afkorting	Uitleg/betekenis
SIC	<i>Smart Indoor Climate</i> (Het slimme binnenklimaat voor bedrijfspanden wat beoogt wordt).
SICP	<i>Smart Indoor Climate Project</i> (het grotere samenwerkingsproject tussen Ultraware, de Hanzehogeschool Groningen en Sinuss).
SICS	<i>Smart Indoor Climate System</i> (dit afstudeerproject).
Basissysteem	Stuurt het binnenklimaat aan op alléén de temperatuur.
Uitbreidingen (op het basissysteem)	Stuurt het binnenklimaat naast temperatuur ook aan op een of meerdere van de volgende elementen: luchtvochtigheid, ventilatie, CO ₂ , schadelijke stoffen, etc.
SIC-kamer / SIC-ruimte	Experimenteerruimte bij Ultraware waar sensoren van Sinuss hangen. Hier worden metingen uitgevoerd die in de database worden opgeslagen.
MoSCoW	Staat voor must have, should have, could have en won't have. Geeft de mate van prioritering aan voor een bepaalde functionaliteit van de business.
HBO-i	De HBO-i stichting is het samenwerkingsverband van ICT-opleidingen in het hoger beroepsonderwijs in Nederland. Met als doel het ICT-onderwijs in Nederland te verbeteren. De 5 beroepscompetenties die ik gedurende de afstudeerstage ga aantonen komen van het HBO-i.
IDE	Staat voor Integrated development environment. Dit is het platform waarin de code wordt ontwikkeld.
PoC	<i>Proof of concept</i> . Dit is een simulatie van de werking van het systeem en wordt gebruikt om aan de business te laten zien dat de doelen behaald zijn. Het wordt nog niet ingezet voor de klant.
MVP	<i>Minimum viable product</i> . Dit is een product wat al daadwerkelijk kan worden ingezet voor de klant.
Pair programming	Met twee personen tegelijk aan dezelfde code werken.
Python	Programmeertaal waar het systeem in ontwikkeld zal worden.
Golang	Programmeertaal van Google waar veel mee gewerkt wordt bij Ultraware.
Artificiële intelligentie (AI) / kunstmatige intelligentie (KI)	Artificiële intelligentie (AI) of kunstmatige intelligentie (KI) is de wetenschap die zich bezighoudt met het creëren van een artefact dat een vorm van intelligentie vertoont (Kunstmatige intelligentie, 2020).
Machine learning (ML)	Automatisch leren of machinaal leren is een breed onderzoeksveld binnen artificiële intelligentie, dat zich bezighoudt met de ontwikkeling van algoritmes en technieken waarmee computers kunnen leren (Hulsen, s.d.).

Tabel 2. Begrippen & afkortingen.

3. INLEIDING

Bij Ultraware is men betrokken bij het Smart Indoor Climate Project. Dit project is opgezet om het binnenklimaat van bedrijfspanden te optimaliseren en te zorgen voor behoud van comfort en gezondheid. Het project is een samenwerking tussen Ultraware, de Hanzehogeschool Groningen en Sinuss. Deze bedrijven werken nauw samen voor dit project. Het project wordt uitgevoerd binnen ID3AS en gesubsidieerd door Interreg EDR. Tijdens mijn stage zal ik vooral te maken krijgen met Ultraware. Dit is het bedrijf waar ik formeel stageloopt en waar ook de meeste activiteiten rondom het project plaatsvinden. Voor het project zal ik bezig gaan met het optimaliseren van de warmtepomp. In het specifiek gaat het dan om het voorverwarmen van het bedrijfspand met deze warmtepomp. Op dit moment slaat deze aan met behulp van een timer, maar dit kan slimmer worden gemaakt. Ultraware wil graag dat er kunstmatige intelligentie/machine learning gebruikt zal worden om met data die door sensoren wordt verzameld het bedrijfspand op een slimme manier voor te verwarmen. Naast de warmtepomp kan er ook worden gedacht aan andere factoren die het binnenklimaat beïnvloeden, zoals de CO₂ concentratie en de luchtvochtigheid. Deze factoren vallen echter buiten de scope van deze opdracht maar kunnen in een latere fase of doorontwikkeling van dit project een rol gaan spelen. Hier hoeft slechts in beperkte mate rekening mee te worden gehouden. Het systeem is wel zo ontwikkeld dat het gemakkelijk uitbreidbaar en onderhoudbaar is.

In dit document worden de mogelijke toepassingen van kunstmatige intelligentie voor het voorverwarmen van het bedrijfspand onderzocht. Stapsgewijs wordt er toegewerkt naar een resultaat. Hoe werkt het voorverwarmen? En welke vormen van kunstmatige intelligentie zijn er? Uit dit onderzoek zal blijken welke methode het beste kan worden geïmplementeerd voor het behalen van het gewenste doel.

4. PROBLEEMSTELLING

Ultraware werkt samen met Sinuss en de Hanzehogeschool Groningen aan het Smart Indoor Climate Project (SICP). Het doel van dit project is het verbeteren van het binnenklimaat binnen bedrijfspanden. Dit afstudeerproject focust op één aspect hiervan: de warmtepomp. In de huidige situatie wordt de warmtepomp aangestuurd met een timer die aanslaat op een bepaalde tijd. Hierna gaat de warmtepomp aan in stand-by modus. Als er iemand de SIC-kamer binnenkomt gaat de warmtepomp in gewone modus. Dit zorgt wel voor een juist niveau van comfort en gezondheid alleen levert niet het meest zuinige verbruik op. De stand-by modus zal namelijk vaak aanstaan op momenten dat dit niet nodig is. De timer kan immers aanslaan op momenten dat niemand aanwezig is in de SIC-ruimte en hierdoor zal er wel energie worden verbruikt.

Het energieverbruik kan dus worden verminderd waarbij het niveau van comfort en gezondheid gelijk kunnen worden gehouden. Dit onderzoek richt zich op het vinden van een oplossing die antwoord geeft op deze probleemstelling.

5. AFBAKENING

Voor de afbakening van dit onderzoek is een hoofdvraag met deelvragen geformuleerd. Het onderzoek focust zich in de basis op het beantwoorden van deze vragen. De hoofdvraag en deelvragen vormen de rode draad in deze scriptie.

5.1. HOOFDVRAAG

Ultraware suggereert de toepassing van AI en/of machine learning op de beschikbare data van de metingen van de sensoren die in de SIC-ruimte aanwezig zijn. Echter blijft dit een suggestie en uit het onderzoek zal nader blijken of AI/machine learning daadwerkelijk de gewenste oplossing bieden voor de business. De hoofdvraag van deze scriptie richt zich hierop. Hetgeen waar Ultraware kennis van wil nemen kan worden samengevat in de volgende hoofdvraag.

Hoe kan kunstmatige intelligentie bijdragen aan het verminderen van het energieverbruik van de warmtepomp met behoud van comfort bij het voorverwarmen van het pand?

5.2. DEELVRAGEN

De hoofdvraag is breed geformuleerd en het antwoord op de hoofdvraag is ook direct de oplossing voor de business voor de probleemstelling die voor dit afstudeerproject is geformuleerd. In deze paragraaf wordt de hoofdvraag opgesplitst in behapbare deelvragen. Deze deelvragen zullen ieder een bepaalde bijdrage hebben bij het beantwoorden van de hoofdvraag. Iedere deelvraag is een stap naar de uiteindelijke oplossing. De deelvragen zijn hieronder opgesomd en er volgt ook een korte beschrijving per deelvraag. De tweede deelvraag heeft ook nog een aantal subdeelvragen.

Hoe kan de warmtepomp worden aangestuurd?

Om te bepalen hoe kunstmatige intelligentie kan bijdragen dient helder te zijn hoe de warmtepomp wordt aangestuurd. Werkt dit met een percentage of wordt er een temperatuur ingesteld? Deze vragen worden beantwoord in deze deelvraag.

Wat is de voor kunstmatige intelligentie relevante data en is deze beschikbaar?

Welke data is er nodig om te onderzoeken hoe kunstmatige intelligentie kan bijdragen aan het verminderen van het energieverbruik van de warmtepomp? Welke data is er allemaal beschikbaar? En welke data is hiervan relevant voor het aansturen van de warmtepomp? Dit zal worden onderzocht in de tweede deelvraag.

Welke vorm van kunstmatige intelligentie kan het beste worden toegepast?

Kunstmatige intelligentie is een erg breed begrip. Welke vormen zijn er van kunstmatige intelligentie en welke kan er worden gebruikt met de beschikbare dataset? Welke algoritmen kunnen relevante resultaten geven? Kortom, hoe kan kunstmatige intelligentie een rol spelen? Er zijn een aantal subdeelvragen opgesteld om deze deelvraag nog beter te kunnen beantwoorden.

- *Welke vormen zijn er?*

Welke vormen van kunstmatige intelligentie zijn er allemaal en welke vorm is het meest relevant in deze casus?

- *Wat is de relatie tussen de attributen in de dataset?*

In de vorige deelvraag is al bekeken welke data beschikbaar is. Nu gaat worden onderzocht hoe deze input kan geven voor een algoritme.

- *Welke soorten algoritmen zijn er en welke zijn relevant?*

Er zijn veel algoritmen maar deze zijn onderverdeeld in vele categorieën. Welke type algoritme is relevant in de casus?

- Aan de hand van welke criteria kan dit worden bepaald?
- Is supervised of niet supervised wenselijk?
- Is Real value based of categorical wenselijk?

- *Welk(e) algoritme(n) is/zijn het best toepasbaar?*

Nu duidelijk is welke categorie/type algoritme het meest relevant is kan er worden gekeken naar het best toepasbare algoritme binnen deze categorie.

Hoe kan met deze kunstmatige intelligentie het energieverbruik worden verminderd?

De toepassing van de kunstmatige intelligentie blijkt uit de vorige deelvraag, maar hoe kan deze implementatie bijdragen aan het verminderen van het energieverbruik van de warmtepomp?

Hoe kan kunstmatige intelligentie geïmplementeerd worden in het systeem?

Nu duidelijk is hoe kunstmatige intelligentie kan worden toegepast kan het worden geïmplementeerd. In deze deelvraag zal duidelijk worden hoe het eerder gedane onderzoek bijdraagt aan de implementatie van kunstmatige intelligentie in het eindproduct.

- *Hoe kunnen de toepasbare algoritmen worden geïmplementeerd?*

Hoe kunnen de algoritmen uit de vorige deelvraag gebruikt worden in het systeem en hoe zit deze implementatie er concreet uit?

- *Hoe kunnen de resultaten van deze algoritmen worden geverifieerd?*

Hoe kan worden aangetoond dat de algoritmen die geïmplementeerd zijn ook daadwerkelijk betrouwbare resultaten opleveren?

6. LITERATUURVERKENNING

Alvorens het onderzoek wordt uitgevoerd is er eerst een literatuurverkenning uitgevoerd waarbij is onderzocht wat er al eventueel bekend is over mogelijke toepassingen en wat er al eventueel gedaan is binnen de business case. Dit hoofdstuk gaat in op deze geraadpleegde onderzoeken.

6.1. BACHELORSRIPTIE VAN MEDEWERKER ULTRAWARE

Een anonieme¹ medewerker bij Ultraware vorig jaar zijn afstudeerproject uitgevoerd. Hiervoor heeft hij binnen het SICP een app ontwikkeld waarbij er feedback kan worden gegeven op de temperatuur in een bepaalde ruimte. Dit onderzoek heeft een aantal raakvlakken met mijn onderzoek. Zo heeft hij ook gekeken naar toepassingen van machine learning voor deze app. Ook heeft het onderzoek te maken met de temperatuur. De bachelorscriptie is daarom een goede start voor dit afstudeerproject.

6.2. MACHINE LEARNING A-Z™: HANDS-ON PYTHON & R IN DATA SCIENCE VAN UDEMY

De uitgebreide cursus Machine Learning A-Z™: Hands-On Python & R In Data Science is gebruikt in het onderzoek om basiskennis te verkrijgen van kunstmatige intelligentie en van machine learning. De cursus bestaat uit video's met uitleg omtrent kunstmatige intelligentie en machine learning en biedt ook programmatuur met voorbeelden van implementaties van algoritmen. Ook zijn er verwijzingen naar overige (wetenschappelijke) literatuur. Het is daarom ook het startpunt wat betreft het opdoen van kennis omtrent kunstmatige intelligentie en machine learning. Het is dus als exploratie gebruikt maar kan ook voor heel specifieke input dienen in het onderzoek. Zo zijn de populairste algoritmen voor machine learning opgenomen in deze cursus en beschreven. Daarom vormt het de ideale basisbron. Indien er specifieke elementen van de cursus zijn opgenomen in het onderzoek zal dit ook worden geciteerd.

¹Er is afgesproken met Ultraware om de medewerker te anonimiseren voor de publiceerbare scriptie.

7. ONDERZOEKSMETHODEN

Dit hoofdstuk beschrijft de gebruikte onderzoeksmethoden en daarbij waar en hoe deze zijn toegepast in de scriptie. Ook wordt er benoemd in welke (delen van de) deelvragen de onderzoeksmethoden gebruikt zijn.

7.1. LITERATUURONDERZOEK

Het literatuuronderzoek vormt een belangrijke bron voor deze scriptie. Voornamelijk alle informatie omtrent kunstmatige intelligentie en machine learning is geraadpleegd uit literatuur. Deze literatuur is voornamelijk op het internet te vinden maar er kunnen eventueel ook boeken zijn geraadpleegd. Indien een bron is gebruikt is deze geciteerd achter de gebruikte bron. Verder is de bron natuurlijk terug te vinden in de literatuurlijst aan het einde van dit document.

Literatuuronderzoek is de belangrijkste methode van onderzoek in deelvraag 4: “Welke vorm van kunstmatige intelligentie kan het beste worden toegepast?”. en bij de subdeelvraag van de een-na-laatste deelvraag “Welk(e) algoritme(n) is het best toepasbaar?”. Ook daar wordt bij een enkel onderdeel echter exploratief onderzoek gedaan. Hierover meer onder de paragraaf over exploratief onderzoek.

7.2. EXPLORATIEF ONDERZOEK

Het exploratieve onderzoek in deze scriptie baseert zich voornamelijk op het intuïtief benaderen van te onderzoeken aspecten. Vaak is dit in overleg met de business. Onder het exploratief onderzoek vallen voornamelijk het onderzoeken van het aansturen van de warmtepomp en het bepalen van de relevante data. Deze zaken zijn ook zeer specifiek voor deze business case. De resultaten van het exploratief onderzoek zijn vastgesteld door middel van het bevragen van medewerkers van het bedrijf en door het onderzoeken van de huidige opstelling van de SIC-ruimte, alsmede het analyseren van de beschikbare data in combinatie met een toelichting van een medewerker van het bedrijf. Vaak is er bij het exploratief onderzoek sprake van informatie die binnen het bedrijf bekend is maar wat boven water moet komen om het Smart Indoor Climate System goed te kunnen ontwerpen. Ook kan het exploratief onderzoek ontwerpkeuzes bevatten van de huidige systemen in het bedrijf die van invloed zijn op het te ontwikkelen systeem en daarmee een bijdrage leveren aan de kwaliteit van de oplossing.

Voor zover het mogelijk is wordt bij het exploratieve onderzoek uitgelegd wat de al bekende ontwerpkeuzes zijn waarvan het systeem afhankelijk gaat zijn. De naam van de medewerker wordt dan ook geciteerd om de bron van de informatie toe te lichten. Deze informatie kan echter zowel mondeling als schriftelijk zijn verkregen.

Exploratief onderzoek speelt een grote rol bij deelvraag 1: “Hoe kan de warmtepomp worden aangestuurd?” en deelvraag 2: “Wat is de voor kunstmatige intelligentie relevante data en is deze beschikbaar?”. Ook de subdeelvraag 4.2 “Wat is de relatie tussen de attributen in de dataset?” is op exploratieve wijze onderzocht.

8. ONDERZOEKSRÉSULTATEN

De onderzoeksvragen zoals geformuleerd bij de afbakening worden in dit hoofdstuk behandeld en beantwoord. Gezamenlijk leveren de deelvragen een bijdrage aan het beantwoorden van de hoofdvraag.

8.1. DEELVRAAG 1: HOE KAN DE WARMTEPOMP WORDEN AANGESTUURD?

Om te weten te komen hoe het aansturen van de warmtepomp energiezuiniger kan worden met behoud van comfort dienen we te begrijpen hoe het aansturen bij het voorverwarmen van het pand in zijn werk gaat. De warmtepomp heeft modussen en standen. Hierover wordt meer duidelijk in de komende paragrafen. Medewerkers van Ultraware (waaronder Martijn Gaasbeek) hebben informatie verstrekt over de werking van de warmtepomp van de SIC-ruimte.

8.1.1. Standen warmtepomp

De warmtepomp kent 4 standen die geconfigureerd kunnen worden. Deze standen zijn eenvoudig. De warmtepomp kan uitstaan of op de lage, gemiddelde en hoge stand. Bij iedere stand wordt de klep van de warmtepomp een bepaald percentage opengezet. Samengevat zijn hieronder de standen in een tabel te zien.

Stand	Stand klep
Uit	0%
Laag	10%
Midden	25%
Hoog	100%

Tabel 3. Standen warmtepomp (Ultraware, 2019).

8.1.2. Modussen warmtepomp

Het voorverwarmen van de SIC-ruimte gebeurt dagelijks tussen 6 uur en 9 uur 's ochtends. Dit kan per bedrijf verschillen maar in dit onderzoek gaan we van deze situatie uit. In deze tijdsspan wordt er door de warmtepomp verwarmt. De 4 standen van de warmtepomp zijn in de vorige paragraaf besproken. Hoe hoger de stand van de warmtepomp, hoe hoger het verbruik en hoe sneller de SIC-kamer opwarmt. De stand kan per halfuur worden gekozen. Indien we het aantal standen onderbrengen in variabele n en het aantal keuzemogelijkheden in variabele r . Kan met de formule $i=n^r$ het aantal modussen worden bepaald. De berekening wordt dan $4^6 = 4096$. Dit resulteert dus in 4096 beschikbare modussen. In onderstaande tabel worden een aantal van deze modussen getoond om een indruk te geven van de werking.

Nr. modus	6.00 uur	6.30 uur	7.00 uur	7.30 uur	8.00 uur	8.30 uur
1	Uit	Uit	Uit	Uit	Uit	Uit
2	Uit	Uit	Uit	Uit	Uit	Laag
3	Uit	Uit	Uit	Uit	Uit	Midden
4	Uit	Uit	Uit	Uit	Uit	Hoog
5	Uit	Uit	Uit	Uit	Laag	Uit
6	Uit	Uit	Uit	Uit	Laag	Laag
7	Uit	Uit	Uit	Uit	Laag	Midden
8	Uit	Uit	Uit	Uit	Laag	Hoog
9	Uit	Uit	Uit	Uit	Midden	Uit
...						
4094	Hoog	Hoog	Hoog	Hoog	Hoog	Laag
4095	Hoog	Hoog	Hoog	Hoog	Hoog	Midden
4096	Hoog	Hoog	Hoog	Hoog	Hoog	Hoog

Tabel 4. Modussen van de warmtepomp voor het voorverwarmen tussen 6 en 9 (TextMechanic, 2018).

8.1.3. Verbruik standen warmtepomp

Iedere stand van de warmtepomp heeft ook zijn eigen verbruik. In onderstaande tabel kan hier een voorbeeld van worden gevonden van de situatie in de SIC-kamer bij Ultraware (let op: deze data kan anders zijn als er sprake is van een andere warmtepomp en/of situatie. Het verbruik van de warmtepomp is daarom instelbaar in het systeem). Het verbruik zoals vermeld in onderstaande tabel is aangedragen door medewerker Martijn Gaasbeek en is representatief voor de warmtepomp in de SIC-kamer bij Ultraware.

Stand	Verbruik
Uit	0 kWh
Laag	10 kWh
Midden	20 kWh
Hoog	40 kWh

Tabel 5. Verbruik per stand warmtepomp (Gaasbeek, 2020).

8.1.4. Verbruik modussen warmtepomp

Bovenstaande verbruiksgegevens resulteren in een patroon van het verbruik over een bepaalde tijd. Dit wordt het verbruik van de modus genoemd. Dit kan worden uitgerekend. Dit gebeurt door voor ieder half uur het verbruik van een stand van de warmtepomp op te tellen. Als bijvoorbeeld het eerste halfuur de stand van de warmtepomp op hoog staat, zal er voor die periode $40 \times 0.5 = 20 \text{ kWh}$ bij het totaal op worden geteld. In een half uur wordt de helft van het aantal kWh verbruikt, vandaar dat dit maal 0.5 gedaan wordt. Over de hele voorverwarmingsperiode van 3 uur kan zo het totaalverbruik voor een bepaalde modus worden bepaald. Wanneer er dus de juiste modus wordt gekozen kan er op de meest effectieve manier worden voorverwarmd en daarmee kan het energieverbruik ook beperkt. De data van tabel 6 kan op deze wijze worden aangevuld zelfs als er geen informatie over beschikbaar is uit de API's.

Nr. modus	Verbruik
1	0 kWh
2	5 kWh
3	10 kWh
...	
4096	120 kWh

Tabel 6. Verbruik per modus warmtepomp.

8.1.5. Deelconclusie deelvraag 1

Het belangrijkste is het verschil tussen de stand en de modus. De stand van de warmtepomp bepaalt of de warmtepomp uit/laag/midden/hoog staat. De modus van de warmtepomp is het patroon van de stand over een tijdsverloop. Dit tijdsverloop is de tijd die gereserveerd is om het pand of de SIC-ruimte op te warmen. Voor de SIC-ruimte is het tijdsverloop van 6 uur tot 9 uur 's ochtends. Verder is het verbruik van belang. In de tabellen wordt duidelijk hoe het verbruik voor een bepaalde stand leidt tot een bepaald verbruik voor een modus. De modus en het tijdsverloop bepalen de aansturing van de warmtepomp.

8.2. DEELVRAAG 2: WAT IS DE VOOR KUNSTMATIGE INTELLIGENTIE RELEVANTE DATA EN IS DEZE BESCHIKBAAR?

Via de beschikbare API's kan veel data worden opgehaald. Er is veel data beschikbaar. Maar welke data is ook daadwerkelijk relevant en kan worden gebruikt om de doelstellingen te behalen? Bij deze deelvraag worden deze zaken onderzocht.

8.2.1. Beschikbare data

Via de API's is veel data beschikbaar. Deze worden beschikbaar gesteld via de sensoren. Welke metingen worden verzameld door Sinuss en zijn beschikbaar via de API's? Er volgt een overzicht van de beschikbare data. Deze is samengesteld door de documentatie van de API's te raadplegen (Tamminga, 2020). Belangrijk is te vermelden dat continue aan de API's wordt gewerkt en hierdoor kan de beschikbare data in de toekomst uitgebreider zijn dan wat er in dit onderzoek is vermeld.

API	Data
Motion	- Bevat data over de bewegingen in de ruimte.
Setpoints	- Bevat data over het licht in de ruimte. - Bevat data over de stroomvoorziening. - Bevat data over aan/uit van de warmtepomp. - Bevat data over de stand van de warmtepomp. - Bevat data over aan/uit van de ventilatie.
Temperature	- Bevat data over de binnentemperatuur.
Weather.com (extern)	- Bevat data over de buitentemperatuur.

Tabel 7. Beknopt overzicht beschikbare data (Tamminga, 2020).

Veel data is dus beschikbaar maar lang niet alle data is relevant. Bepaalde data kan wellicht relevant zijn maar zorgt voor een onnodige complexiteit van het systeem. We gaan kijken welke data nodig is om een efficiënt systeem te kunnen creëren die kan voldoen aan de bedrijfsdoelstelling.

8.2.2. Relevante data

Om te weten welke data nodig is om te bepalen hoe het pand het beste kan worden opgewarmd moeten we weten welke zaken intuïtief gezien invloed hebben op het verwarmen van een ruimte. Wat we willen weten is hoe de verwarming moet worden aangestuurd. Dit is dus de stand en modus van de warmtepomp. Hierover meer in de vorige deelvraag. Om te bepalen wat de beste modus is dient de effectiviteit van iedere modus te worden geanalyseerd. Daarom wordt een overzicht gemaakt per datum van de metingen en hoe de warmtepomp zich daarbij gedraagt. Deze data kan uit de beschikbare data worden gedestilleerd. Het is de relevante data. De volgende tabel laat een voorbeeld van zo'n overzicht zien.

Datum	Begin temp.	Eind temp.	Δ tijd	Buitemp. begin	Modus	Stroomverbruik
21-03-2020	15,0°C	19,5°C	3 uur	10°C	112210	203 kWh
22-03-2020	14,6°C	19,5°C	3 uur	9°C	233212	4090 kWh
23-03-2020	15,2°C	19,5°C	3 uur	12°C	111313	304 kWh
24-03-2020	15,9°C	19,5°C	3 uur	14°C	021200	60 kWh

Tabel 8. Voorbeeld overzicht relevante data.

8.2.3. Missende data

Uit de tabel met relevante data wordt duidelijk dat niet alle data die hierin staat ook beschikbaar is. Het stroomverbruik bijvoorbeeld, deze kan niet worden opgehaald door middel van de API's. Het zal daarom noodzakelijk zijn om deze uit te rekenen aan de hand van de uitgevoerde modus voor die datum. Het uitrekenen is al uitgelegd in de vorige deelvraag. De Δ tijd kan ook niet worden opgehaald maar wordt vooraf ingesteld. Voor het te bouwen systeem kan hier een configuratievariabele voor worden gekozen. De eindtemperatuur kan worden gemeten maar ook deze is vooraf geconfigureerd en er is afgesproken dat het om 9 uur 's ochtends 19,5 graden moet zijn. Dan is het pand of de ruimte succesvol voorverwarmd.

8.2.4. Omzetten beschikbare data naar relevante data

Uit het voorbeeld wordt duidelijk dat de relevante data die wel beschikbaar is vanuit de API's eerst gedestilleerd moet worden uit de beschikbare data. Dit is een heel proces. De metingen per API verschillen namelijk onderling. Bepaalde metingen worden ieder half uur gedaan en andere metingen iedere 10 seconden. Ook is het format niet gelijk en zijn er veel verschillen tussen deze data. De beschikbare data dient dus eerst te worden omgezet alvorens het kan worden gebruikt. Dit is een omvangrijke operatie. Er wordt in dit onderzoek niet verder op ingegaan hoe dit precies in zijn werk gaat. Hiervoor is in het systeem een omzet- en filterfunctionaliteit ontwikkeld. Voor meer informatie hierover wordt verwezen naar het technisch ontwerp van de applicatie.

8.2.5. Deelconclusie deelvraag 2

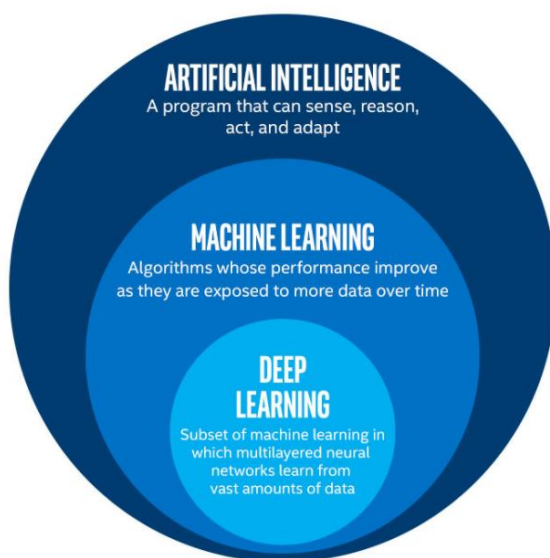
Er is veel data beschikbaar hoewel dit ook nog kan veranderen. Uit de beschikbare data is de relevante data gefilterd. Deze is te zien in tabel 8. Het stroomverbruik van de warmtepomp is een missend stuk in de data en deze wordt uitgerekend. Om de beschikbare data om te zetten naar relevante data is een groot proces nodig. Voor de werking hiervan kan het technisch ontwerp worden geraadpleegd.

8.3. DEELVRAAG 3: WELKE VORM VAN KUNSTMATIGE INTELLIGENTIE KAN HET BESTE WORDEN TOEGEPAST?

Kunstmatische intelligentie is een erg breed begrip. Het is daarom van belang welke vorm hiervan wordt gebruikt om de doelstellingen te behalen. Niet alle vormen van kunstmatische intelligentie dragen bij aan het behalen van deze doelstelling. Deze deelvraag is onderverdeeld in meerdere subdeelvragen. Deze beantwoorden allemaal een stukje van deze deelvraag. Na het beantwoorden van deze subdeelvragen zal een antwoord worden gegeven op de deelvraag.

8.3.1. Deelvraag 3.1: Welke vormen zijn er?

In het algemeen kan kunstmatische intelligentie worden onderverdeeld in drie lagen waarbij er een steeds specifiekere vorm van intelligentie wordt gebruikt. De drie lagen zijn kunstmatische intelligentie, machine learning en deep learning. In de volgende figuur wordt duidelijk hoe deze zich precies tot elkaar verhouden en wat het onderscheid tussen de lagen is.



Figuur 1. Verschil tussen kunstmatische intelligentie, machine learning en deep learning (Touger, 2018).

Kunstmatische intelligentie (artificiële intelligentie) is het algemene begrip. Dit gaat eigenlijk over alle algoritmen. Alle programmatuur die een vorm van intelligentie vertoont valt binnen deze categorie. Als onderdeel van kunstmatische intelligentie bestaat er machine learning. Machine learning focust zich op het leren d.m.v. data. Het werken met neurale netwerken wordt dan weer deep learning genoemd. Dit is weer een subonderdeel van machine learning.

Uit deelvraag 1 blijkt dat we te maken hebben met data uit vele metingen. Deze data wordt gebruikt om het energieverbruik te minimaliseren met behoud van comfort. Er wordt dus geleerd met de data om dit te bereiken. In het afstudeerproject zal er dus sprake zijn van machine learning. Een gewoon algoritme kan de data niet gebruiken om zichzelf te verbeteren over tijd dus simpele kunstmatische intelligentie is niet toereikend. Deep learning kan ook niet worden toegepast omdat er geen sprake is van neurale netwerken in de data. Het creëren van neurale netwerken zou een doel opzich kunnen zijn maar dit gaat voorbij de bedrijfsdoelstelling. Omdat er simpelweg geleerd wordt met de data gaat er dus met machine learning gewerkt worden om de doelstelling te bereiken. Hierbij wordt de data uit de eerste deelvraag gebruikt.

8.3.1.1. Deelconclusie deelvraag 3.1

Kunstmatische intelligentie heeft een tak machine learning. De machine learning heeft weer een tak deep learning. Voor deze business case zal machine learning worden toegepast omdat er met data van eerdere metingen wordt geleerd hoe de warmtepomp in de toekomst het beste kan worden aangestuurd.

8.3.2. Deelvraag 3.2: Wat is de relatie tussen de attributen in de dataset?

Om te bepalen welke relaties tussen de attributen in de dataset zijn kijken we eerst naar de relevante data. Het voorbeeld van de relevante data uit de tweede deelvraag gaan we nu analyseren om intuïtief een idee en beeld te vormen van de relaties die er zijn in de dataset. Op deze manier kunnen deze relaties in een later stadium worden gebruikt om de parameters van het algoritme goed in te kunnen richten.

Datum	Begintemp.	Eindtemp.	Δ tijd	Buitemtemp. begin	Modus	Stroomverbruik
21-03-2020	15,0°C	19,5°C	3 uur	10°C	112210	203 kWh
22-03-2020	14,6°C	19,5°C	3 uur	9°C	233212	4090 kWh
23-03-2020	15,2°C	19,5°C	3 uur	12°C	111313	304 kWh
24-03-2020	15,9°C	19,5°C	3 uur	14°C	021200	60 kWh

Tabel 9. Voorbeeld overzicht relevante data.

Er vallen een aantal relaties op in de relevante data. Allereerst de relatie van het verschil tussen de begin- en eindtemperatuur en de modus. De delta tijd speelt hier ook een rol in. In de volgende paragrafen worden de relaties nauwkeuriger bekeken.

8.3.2.1. Relatie delta temperatuur, delta tijd en modus

Het verschil tussen de begintemperatuur en de eindtemperatuur noemen we de delta temperatuur (Δ temperatuur). Deze wordt beïnvloed door twee factoren, de modus van de warmtepomp en de delta tijd. Allereerst bespreken we de modus van de warmtepomp. De delta temperatuur wordt voor een groot deel beïnvloed door de modus van de warmtepomp. Hoe harder de warmtepomp staat in die periode, hoe groter het verschil tussen de begintemperatuur en de eindtemperatuur. De delta temperatuur zal daardoor dus hoger uitvallen. Een tweede factor die daarbij een rol speelt is de delta tijd (Δ tijd). Indien deze groter is zal de warmtepomp langer aan kunnen staan waardoor delta temperatuur (Δ temperatuur) ook groter zal zijn. In de situatie van het Smart Indoor Climate System zal de delta tijd (Δ tijd) vrijwel altijd 3 uur zijn en kan daarom als factor worden genegeerd omdat het hierdoor een constante in de formule wordt. Echter dient het wel te worden benoemd aangezien dit zeker invloed kan hebben, zeker als de situatie wijzigt of er bij een ander bedrijf met een andere implementatie van het systeem bijvoorbeeld sprake zal zijn van een andere delta tijd (Δ tijd). Delta tijd (Δ tijd) is daarom wel instelbaar in het systeem zodat de toepassing ook in andere bedrijfssituaties gebruikt kan worden. Voor de testdata en het onderzoek blijven we echter uitgaan van Δ tijd = 3. Andere situaties vallen namelijk buiten de scope van het afstudeerproject.

8.3.2.2. Relatie delta temperatuur en begintemperatuur

Een tweede relatie die geconstateerd kan worden is de relatie tussen de delta temperatuur (Δ temperatuur) en de begintemperatuur. Aangenomen kan worden dat zodra de begintemperatuur hoger is de delta temperatuur lager uit zal vallen. Hoe zit dit precies? De meeste warmtepompen warmen op met warme lucht. De temperatuur van die verwarmende lucht is wellicht aanpasbaar maar hier zit een bepaald plafond aan. Dit zorgt ervoor dat indien de starttemperatuur laag is de warme lucht van de warmtepomp een grote invloed heeft op de temperatuur in die ruimte. Indien de starttemperatuur al relatief hoog is zal de warme lucht van de warmtepomp een relatief kleine invloed hebben op de temperatuur in de ruimte. Het resultaat is dat de delta temperatuur (Δ temperatuur) lager uit zal vallen.

8.3.2.3. Relatie modus warmtepomp en energieverbruik

De relatie tussen de modus van de warmtepomp en het energieverbruik is redelijk voor de hand liggend. Let bij de volgende uitleg goed op het verschil in definitie tussen de stand en modus van de warmtepomp zoals dit in de eerste deelvraag is behandeld.

Zoals al naar voren kwam in de eerste deelvraag heeft iedere stand van de warmtepomp een bepaald verbruik in kilowattuur. Daarmee kan ook het verbruik van de modus van de warmtepomp worden uitgerekend. Wanneer de gemiddelde stand van de warmtepomp over het tijdsverloop hoger is zal ook het energieverbruik hoger zijn. Dit is de situatie voor de modus van de warmtepomp. Simpel gezegd is het verbruik hoger als de warmtepomp harder heeft aangestaan. Het verband (de relatie) is dus lineair. Wanneer de modus hoger is (en de warmtepomp dus gedurende een langere tijd op een hogere stand staat) zal ook het energieverbruik hoger zijn.

8.3.2.4. Relatie delta temperatuur en omgevingsfactoren

Deze relatie is de lastigste. Dat er sprake is van een relatie van de delta temperatuur (Δ temperatuur) met meer factoren is duidelijk. De enige omgevingsfactor die gemakkelijk te vergelijken is, is de (absolute) buitentemperatuur. Deze is dan ook in de relevante data opgenomen, ook al wordt deze factor in het huidige systeem waarschijnlijk buiten beschouwing gelaten. Er zijn echter meer factoren die van invloed zijn maar die niet in de relevante data staan. In het volgende overzicht zijn deze factoren opgesomd:

- Geleiding warmte structuur gebouw
- Luchtvochtigheid
- Zuurstof en CO₂ gehalte
- Windkracht
- Windrichting

Deze factoren worden niet meegenomen maar hebben wel een invloed op de relatie van de delta temperatuur (Δ temperatuur). Doordat het algoritme echter gaat leren met de bekende data van een bepaald gebouw zullen deze factoren hier onzichtbaar in worden meegenomen. Het leren gebeurt dan specifiek voor dat gebouw waarbij deze overige factoren in de berekening zijn meegenomen.

Als voorbeeld kan er gekeken worden naar één van bovenstaande factoren, bijvoorbeeld de windkracht. Als deze erg hoog is zou er bij het gebouw sprake kunnen zijn van extra ventilatie door het gebouw of van snellere warmtegeleiding door de muren van het gebouw. Deze twee zaken kunnen de binnentemperatuur een klein beetje laten zakken of het opwarmen/voorverwarmen afremmen. Uit de bekende data uit het verleden zou het algoritme zo'n situatie kunnen herkennen en hierop kunnen inspelen waardoor de streef temperatuur alsnog bereikt wordt. Het kan zijn dat deze omgevingsfactoren echter dusdanig onzichtbaar zijn in de data uit het verleden dat dit een negatieve invloed kan hebben op de accuraatheid van het systeem.

Er kan worden besloten om de omgevingsfactoren wél in het systeem mee te nemen. Dit zal de accuraatheid van het systeem vergroten. Echter vergroot dit ook de complexiteit van het systeem (bij iedere extra factor die meegenomen wordt). De installatie van het systeem bij een ander bedrijf met nieuwe omstandigheden zou hierdoor ook meer tijd kosten. Deze omgevingsfactoren dienen dan namelijk van tevoren al worden ingesteld in het systeem. In het ontwerp van het huidige systeem is ervoor gekozen om te focussen op deze flexibiliteit en niet op de accuraatheid. Dit blijft echter een afweging ook in de toekomst, mocht het systeem later worden doorontwikkeld. Nieuwe factoren kunnen altijd in het systeem worden opgenomen.

8.3.2.5. *Relatie delta temperatuur en datum*

Tot slot is er nog een relatie tussen de datum en de delta temperatuur (Δ temperatuur). In de winter is er bijvoorbeeld sprake van een andere delta temperatuur (Δ temperatuur) dan in de zomer. De relatie is seizoensgebonden. Ook kan het zijn dat het voorverwarmen überhaupt minder nodig is in de zomer dan in de winter. Om de juiste keuze te maken is het dus van belang om de metingen die het meest recent zijn het zwaarst mee te laten wegen in het machine learning model. Gezien het feit dat de delta temperatuur in veel eerdergenoemde relaties voorkomt zal de datum hier redelijkerwijs ook invloed op uitoefenen. Daarom zal de datum worden meegenomen in het model.

8.3.2.6. *X- en Y-waarden machine learning model*

Nu bekend is welke relaties aanwezig zijn in de relevante data kan worden bepaald hoe een machine learning model zou kunnen worden opgebouwd. De X-waarden zijn hierbij de input en de Y-waarden de output van de data (Brownlee, How Machine Learning Algorithms Work (they learn a mapping of input to output), 2016). Voor de hand liggend is de modus van de warmtepomp de Y-waarde. De X-waarden zijn hierbij de begintemperatuur, de eindtemperatuur, delta tijd (Δ tijd) en de buitentemperatuur aan het begin. Waarden die niet worden opgenomen in het model zijn de dag van de week en het energieverbruik. De dag van de week (woensdag of vrijdag) heeft geen invloed op de berekening, maar de datum wel. Het energieverbruik kan later apart worden uitgerekend met een bepaalde Y-waarde en wordt daarom ook niet in het model opgenomen.

Machine learning model	
X-waarden	Y-waarden
- Datum - Begintemperatuur - Eindtemperatuur - Delta tijd (Δ tijd) - Buitentemperatuur begin	- Modus warmtepomp

Tabel 10. X- en Y-waarden voor machine learning model.

In bovenstaande tabel zijn de gekozen X- en Y-waarden voor het machine learning model wat de basis vormt voor het algoritme samengevat.

8.3.2.7. *Deelconclusie deelvraag 3.2*

Er zijn meerdere relaties te leggen in de relevante data. Uit de belangrijkste relatie zijn de X- en Y-waarden bepaald. Dit zijn de x-waarden: de datum, de begintemperatuur, de eindtemperatuur, delta tijd (Δ tijd) en de buitentemperatuur aan het begin. De Y-waarde is de modus van de warmtepomp.

8.3.3. Deelvraag 3.3: Welke soorten algoritmen zijn er en welke zijn relevant?

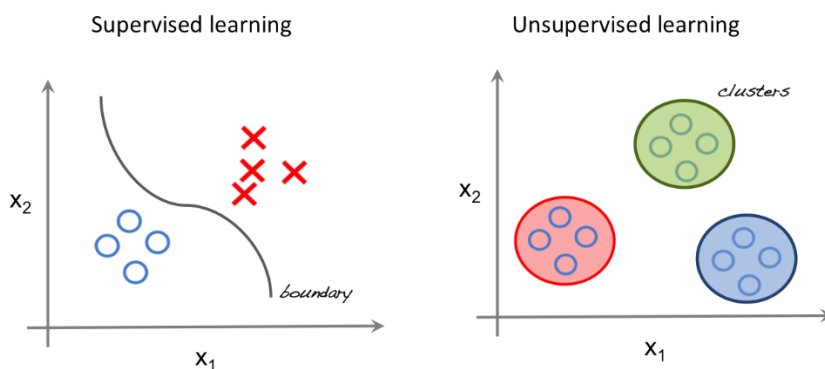
In deelvraag 3.1 bleek al dat er voor het business probleem voornamelijk gebruik zal worden gemaakt van machine learning. Maar welke soorten machine learning algoritmen zijn er? Hier wordt in deze deelvraag een antwoord op gegeven. Eerst volgt er een overzicht met de categorieën en soorten algoritmen waarna er gekeken wordt naar de juiste categorie en het beste soort algoritme voor het bestaande probleem.

8.3.3.1. Categorieën en soorten machine learning algoritmen

Bestaande algoritmen zijn onder te verdelen in verschillende categorieën en soorten. Door hier eerst naar te kijken kan er stap voor stap een geschikt algoritme worden uitgezocht. Allereerst wordt er gekeken naar supervised, unsupervised en semi-supervised learning. Dit zijn 3 categorieën waar algoritmen in kunnen worden verdeeld op basis van de manier waarop ze leren met de data. Daarna kijken we naar de 3 soorten algoritmen regressie, classificatie en clustering. De manier van voorspellen/toepassen van een nieuw datapunt verschilt bij deze soorten. Ten slotte kijken we naar lineaire en niet-lineaire algoritmen. Voor alle categorieën en soorten wordt bepaald welke er voor deze business case het meest geschikt is. Deze volgorde en manier van selecteren komen uit de modelselectie (Eremenko & de Ponteves, 2020).

8.3.3.2. Supervised, unsupervised en semi-supervised learning

Binnen de machine learning bestaan drie verschillende categorieën algoritmen. Allereerst is er supervised learning. Van supervised learning is sprake indien er bepaalde invoervariabelen (x) zijn en een uitvoervariabele (Y) is en er een algoritme gebruikt wordt om de toewijzingsfunctie van de invoer naar de uitvoer te leren (Brownlee, Supervised and Unsupervised Machine Learning Algorithms, 2016). Met andere woorden, er is al voorkennis beschikbaar van wat de outputwaarden voor onze data zou moeten zijn. Daarom is het doel van supervised learning om het algoritme te laten leren wat de gewenste output is, gegeven de voorkennis van gegevens en de bekende output. Bij unsupervised learning is geen sprake van een uitvoervariabele maar alléén van een invoervariabele. Als laatste is er ook nog zoiets als semi-supervised learning. Hier wordt in dit onderzoek verder geen aandacht aan besteed omdat dit niet vaak wordt gebruikt. Van semi-supervised learning is sprake als er relatief veel invoervariabelen (x) zijn ten opzichte van een beperkt aantal uitvoervariabelen (Y). Dit is dus een soort combinatie tussen supervised en unsupervised learning.



Figuur 2. Verschil tussen supervised en unsupervised learning (Seif, 2018).

Bovenstaand plaatje illustreert nogmaals vanuit een andere insteek het verschil tussen supervised en unsupervised learning. Bij supervised learning wordt er een grens bepaald tussen (delen van) de data en bij unsupervised learning is er sprake van het indelen in clusters. In het linkerplaatje wordt er een lijn getekend tussen datapunten waardoor er circels en kruisjes worden onderscheiden in de datapunten. In het rechterplaatje kan door middel van clustering onderscheid worden gemaakt tussen de groene, rode en blauwe datapunten.

8.3.3.3. Welke categorie algoritme is toepasbaar voor het bepalen van de modus van de warmtepomp?

Uit de uitleg van het supervised en unsupervised learning en de beschrijven van de bestaande relaties tussen de attributen van de relevante data kan worden geconcludeerd dat er voor deze situatie supervised learning nodig is. Er is namelijk al data beschikbaar en deze zal worden gebruikt om het algoritme steeds beter te trainen. Daarbij zal het algoritme kijken naar de al bekende input en output om zo op basis van nieuwe input een suggestie te kunnen doen van de output. Door het patroon in de eerdere data te herkennen kan worden bepaald wat voor de nieuwe situatie een gunstige output is.

8.3.3.4. Regressie, classificatie en clustering

Behalve de categorieën algoritmen zijn er ook soorten algoritmen die te classificeren zijn onder drie vormen. Dit zijn regressie, classificatie en clustering. Welk soort algoritme zal worden gekozen is afhankelijk van het soort probleem. Daarom zullen de drie klassen machine learning algoritmen hieronder verder worden uitgewerkt.

8.3.3.5. Regressie

Regressie: het voorspelt continu gewaardeerde output. De regressieanalyse is het statistische model dat wordt gebruikt om de numerieke gegevens te voorspellen in plaats van labels. Het kan ook de distributietrends identificeren op basis van de beschikbare gegevens of historische gegevens. Het voorspellen van iemands inkomen vanaf zijn leeftijd, onderwijs is een voorbeeld van een regressietaak (Kashid, 2018). Dit statistische model lijkt dus op het doortrekken van een lijn van een grafiek. Met regressie kan op een bepaald moment een waarde worden uitgerekend.

8.3.3.6. Classificatie

Classificatie: het voorspelt een discreet aantal waarden. Bij de classificatie worden de gegevens volgens verschillende parameters onder verschillende labels ingedeeld en vervolgens worden de labels voor de gegevens voorspeld. E-mails classificeren als spam of geen spam is een voorbeeld van een classificatieprobleem (Kashid, 2018).

8.3.3.7. Clustering

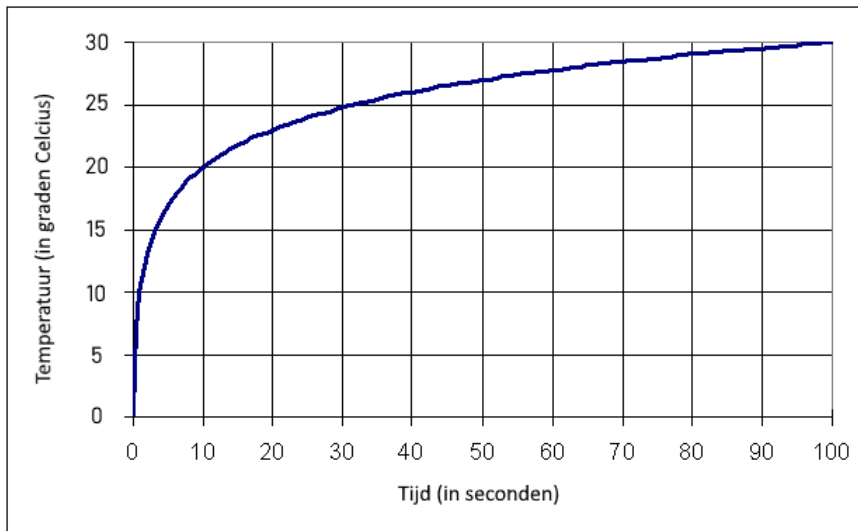
Clustering: Clustering is de taak om de dataset in groepen te verdelen, clusters genaamd. Het doel is om de gegevens zo op te splitsen dat punten binnen één cluster erg op elkaar lijken en punten in verschillende clusters verschillend zijn. Het bepaalt de groepering tussen niet-gelabelde gegevens (Kashid, 2018).

8.3.3.8. Welk soort probleem is het bepalen van de modus van de warmtepomp?

Nu helder is welke soorten algoritmen er zijn kan er gesteld worden dat voor het bepalen van de stand van de warmtepomp een classificatiealgoritme nodig is. De modus van de warmtepomp kan gezien worden als een klasse. De modussen die zijn gebruikt in de afgelopen tijd kunnen uit de data worden gedestilleerd en kunnen ook als klassen worden gezien. Door de modussen als klassen te behandelen kan een classificatiealgoritme worden geïmplementeerd. Op deze manier kan uiteindelijk worden bepaald binnen welke klasse de modus van de warmtepomp vandaag zou moeten vallen gelet op de eerdere gemaakte keuzes.

8.3.3.9. Lineair of niet-lineair probleem

Of er sprake is van een lineair of van een niet-lineair probleem hangt af van de x- en de Y-waarden, oftewel de te gebruiken input en de te berekenen output. De x- en Y-waarden zijn al eerder besproken in paragraaf 8.2.3.5. Er kan hieruit worden geconcludeerd dat de relatie tussen de x- en Y-waarden niet-lineair is. Wanneer het erg koud is in de ruimte en er wordt voorverwarmd zorgt dit voor een veel grotere stijging in temperatuur dan wanneer het al warm is in de ruimte. Als het heel warm is in de ruimte zou er mogelijkwerwijs niet eens een stijging in temperatuur worden waargenomen. Als het ware vlakt de lijn steeds meer af wanneer de streeftemperatuur bereikt wordt. Er is dus sprake van een logaritmisch verband die afhankelijk is van de starttemperatuur. Het logaritmische verband zorgt namelijk voor deze afvlakking (Murray, 2019). Doordat er een logaritmisch verband speelt in de waarden kan er geen sprake zijn van een lineair probleem (Eremenko & de Ponteves, 2020). Er is daarom sprake van een niet-lineair probleem.



Figuur 3. Voorbeeld logaritmisch verband temperatuur met delta tijd bij een constante stand van de warmtepomp.

Bovenstaand figuur laat zien hoe het voorverwarmen in het bedrijfspannd mogelijkwerwijs kan verlopen. Zodra de temperatuur hoger wordt zal er met dezelfde stand/modus van de warmtepomp een minder grote stijging worden bereikt naarmate men verder in de tijd komt. Dit komt door het logaritmisch verband van de temperatuur met delta tijd bij een constante stand van de warmtepomp. Bovenstaand figuur is slechts ter illustratie en geeft niet de werkelijke stijging weer, alleen een schets van de interactie.

8.3.3.10. Deelconclusie deelvraag 3.3

Voor de business case zal er een classificatiealgoritme kunnen worden gebruikt met een niet-lineair model aan de basis. Deze algoritmen zullen naar verwachting de meest waardevolle en bruikbare resultaten opleveren.

8.3.4. Deelvraag 3.4: Welk(e) algoritme(n) is/zijn het best toepasbaar?

Nu bekend is dat er een niet-lineair classificatie algoritme zal worden gebruikt, zullen alle mogelijke niet-lineaire classificatie algoritmen aan bod komen en zal worden onderzocht of deze een rol kunnen spelen voor de oplossing van het business probleem. Dit gebeurt op basis van theoretisch onderzoek. Het praktisch (experimenteel) onderzoek zal plaatsvinden in de volgende deelvraag.

8.3.4.1. Beoordeling algoritmen

In de paragraaf komen de algoritmen aan bod die aan de eerdergenoemde criteria voldoen. Ze zijn niet-lineair en behoren tot de classificatiealgoritmen. De algoritmen die aan bod komen worden beoordeeld op een aantal factoren. Aan de hand van de beoordeling wordt vervolgens bepaald welk algoritme het meest geschikt is voor de toepassing binnen de oplossing voor de business. De beoordeling wordt gebaseerd op de volgende factoren.

Intuïtieve toepasbaarheid

Kan het algoritme goed of niet goed worden toegepast in de bedrijfscasus? Dit is uiteraard een belangrijk eerste aspect voor het bepalen van de geschiktheid van een algoritme.

Complexiteit

De complexiteit van het algoritme speelt een rol. Wanneer het algoritme erg complex is kan het wellicht een minder goede keuze zijn. Dit beïnvloedt uiteraard ook de performance.

Performance

Of het algoritme heel snel of heel langzaam is speelt een rol bij de geschiktheid van het algoritme. Dit wordt dan ook meegenomen in de beoordeling.

Accuraatheid

In hoeverre zijn de resultaten in het algoritme accuraat? Of zijn de resultaten erg willekeurig? Ook hier wordt het algoritme op beoordeeld.

Inzichtelijkheid

Is het duidelijk wat het algoritme precies doet? Is het makkelijk uit te leggen hoe de resultaten zijn verkregen of is dit een grote black-box? De mate van inzichtelijkheid speelt mee bij de keuze voor een algoritme.

8.3.4.2. Beoordelingsschaal

Bovenstaande factoren worden op een schaal van 1 tot 5 beoordeeld. Hoe hoger de beoordeling, hoe beter het algoritme scoort op die factor. Er wordt ook een totaalscore bepaald die de uiteindelijke geschiktheid van het algoritme zal uitwijzen. Deze totaalscore wordt bepaald door de scores van de individuele factoren bij elkaar op te tellen. De intuïtieve toepasbaarheid telt hierbij dubbel. De reden hiervoor is dat deze zwaar meetelt. Als het algoritme slecht toepasbaar is zal deze niet gauw geschikt zijn. De totaalscore kan dus lopen van 0 tot 30.

Beoordelingsmodel per factor algoritme

0 - 5

Tabel 11. Beoordelingsschaal per factor algoritme.

Beoordelingsmodel algoritme

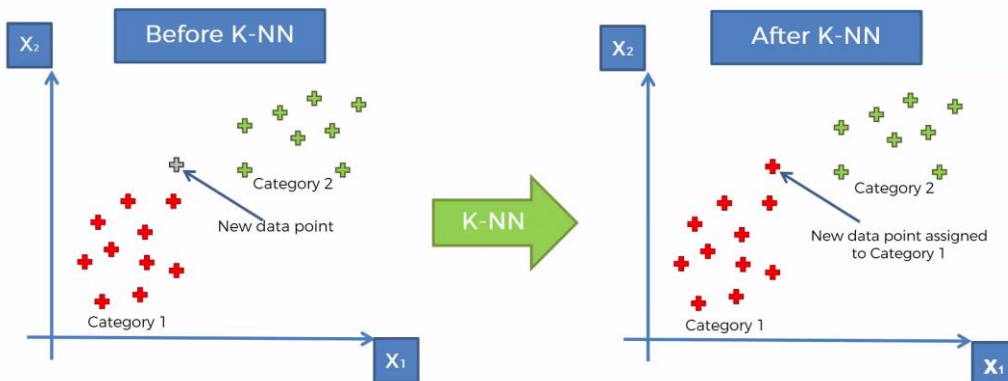
0 - 30

Tabel 12. Beoordelingsschaal algoritme.

8.3.4.3. K-Nearest Neighbors

Werking algoritme

Met K-Nearest Neighbors wordt van een nieuw datapunt onderzocht bij welke groep deze het dichtstbij zit. Aan de hand van eerder bekende data en de bijbehorende groep of categorie kan het algoritme beslissen hoe de nieuwe data wordt ingedeeld. In onderstaand figuur wordt dit geïllustreerd.



Figuur 4. Visualisatie voorbeeld K-Nearest Neighbors algoritme (Eremenko & de Ponteves, 2020).

Uit het figuur blijkt dat er een nieuw datapunt bij de al bestaande data wordt gevoegd. Deze is echter nog niet gecategoriseerd. Het algoritme bepaalt vervolgens aan de hand van de positie van het datapunt bij welke groep dit datapunt waarschijnlijk hoort. Dit gebeurt door te kijken naar de dichtstbijzijnde andere punten, de burens (neighbors). Er is ook een variabele k instelbaar. De grenzen waarmee de burens in groepen worden verdeeld worden bepaald door een variabele k . Een hoge waarde van k zorgt voor een minder willekeurige plaatsing van het nieuwe datapunt, maar zorgt ook voor minder duidelijke grenzen voor de burens (Everitt, Landau, Leese, & Stahl, 2011). Wanneer k gelijk is aan 1 wordt dit het Nearest Neighbors algoritme genoemd (zonder k).

Intuïtieve toepassing

Indien we de modus van de warmtepomp zien als een groep of categorie die door variabele k wordt bepaald, kan voor een nieuw datapunt met het K-Nearest Neighbors algoritme worden bepaald in welke categorie deze het beste zou kunnen vallen. Hierbij wordt er gekeken naar de afstand van het nieuwe datapunt ten opzichte van de groepen die eromheen staan. Er zou geëxperimenteerd kunnen worden met verschillende waarden voor k . De categorie bepaalt een bepaalde modus voor de warmtepomp waardoor er met het algoritme dus een modus wordt gekozen op basis van de locatie op de assen van de eerdere datapunten.

Beoordeling

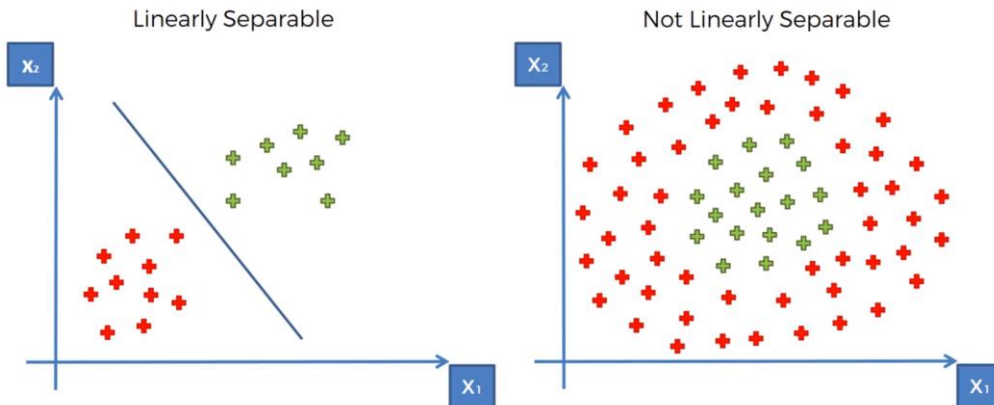
Factor	Score	Toelichting
Intuïtieve toepasbaarheid	5/5	Het algoritme is intuïtief toepasbaar en waarschijnlijk dus ook een goede optie voor de oplossing van de business.
Complexiteit	4/5	Weinig complexiteit, variabele k maakt het een 4. Voor Nearest Neighbors zou dit een 5 zijn.
Performance	5/5	Algoritme werkt snel en geeft snel resultaten terug.
Accuraatheid	5/5	Accuraatheid is instelbaar met variabele k .
Inzichtelijkheid	4/5	Berekening is inzichtelijk met duidelijke formule.
Totaal	28/30	Hoge score. Het algoritme is in grote mate geschikt.

Tabel 13. Beoordeling K-Nearest Neighbor.

8.3.4.4. Kernel Support Vector Machine (Kernel SVM)

Werking algoritme

Met het SVM-algoritme wordt er een lijn getrokken tussen bestaande data waarna op basis van die gezette lijn nieuwe datapunten kunnen worden ingedeeld. SVM is echter een lineair model. Er bestaat echter ook een niet-lineaire versie van. Dit wordt kernel SVM genoemd. Het kernel SVM-algoritme is in staat om ook andere vormen te tekenen in een bepaalde grafiek om zo de data te categoriseren. Op basis van die categorisatie kunnen nieuwe datapunten in de toekomst vervolgens worden ingedeeld.



Figuur 5. Visualisatie voorbeeld kernel SVM-algoritme (Eremenko & de Ponteves, 2020).

In bovenstaand figuur wordt het verschil tussen het SVM-algoritme en het kernel SVM-algoritme gevisualiseerd. Links is de data door middel van een lijn te categoriseren, rechts niet. Echter zou hier wel bijvoorbeeld een cirkel getekend kunnen worden om de data alsnog te kunnen categoriseren.

Intuïtieve toepassing

Aan de hand van het stellen van grenzen (dus het tekenen van figuren) in de bestaande data kan wellicht worden bepaald welke categorie er bij een bepaalde modus van de warmtepomp hoort. Wel is het zo dat er extreem veel modussen voor de warmtepomp zijn zoals al bleek uit de resultaten van de eerste deelvraag. Er zijn er 4096. Om nu zo'n groot aantal lijnen of vormen te maken is wellicht mogelijk maar het zou voor verwarring kunnen zorgen als bijvoorbeeld vormen door elkaar heen worden geplaatst. Daarom is het op het eerste gezicht onpraktisch gezien mogelijke overlapping bij de grote hoeveelheid en naar verwachting wordt de inzichtelijkheid zo ook minder transparant.

Beoordeling

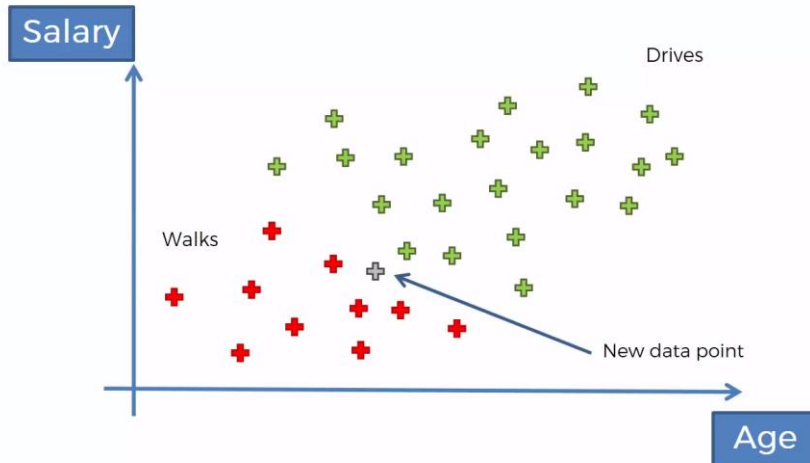
Factor	Score	Toelichting
Intuïtieve toepasbaarheid	2/5	Er wordt verwacht dat de grote hoeveelheid vormen
Complexiteit	3/5	Het tekenen van lijnen of figuren is geen complexe taak. Wel neemt de complexiteit toe wanneer de hoeveelheid lijnen of figuren toeneemt.
Performance	3/5	Goed maar wordt minder bij extreem veel lijnen of figuren.
Accuraatheid	3/5	Afhankelijk van vormen van de lijnen of figuren.
Inzichtelijkheid	3/5	De lijnen maken keuzes inzichtelijk, hoeveelheid maakt dit weer lastiger.
Totaal	16/30	Gemiddelde score. Vanwege de intuïtieve toepasbaarheid, de complexiteit en de inzichtelijkheid valt de score lager uit.

Tabel 14. Beoordeling Kernel Support Vector Machine (Kernel SVM).

8.3.4.5. Naïve Bayes

Werking algoritme

Het Naïve Bayes algoritme werkt met kansberekening. Hierbij wordt gebruik gemaakt van het Bayes-theoreem. Voor een bepaald nieuw datapunt wordt berekend wat de kans is dat deze bij de ene groep hoort of bij de andere groep. Hierbij wordt er gekeken naar de x en y positie van het nieuwe datapunt.



Figuur 6. Visualisatie voorbeeld Naïve Bayes algoritme (Eremenko & de Ponteves, 2020).

Intuïtieve toepassing

Het voordeel van het algoritme is dat er niet direct tot een bepaalde categorie of groep wordt ingedeeld, maar dat wordt teruggegeven wat de kans is dat het punt tot een bepaalde groep behoort. In het geval van het Smart Indoor Climate System betekent dit dat de kans kan worden bepaald in hoeverre een nieuw datapunt kan worden ingedeeld bij een bepaalde modus van de warmtepomp. Het voordeel hiervan is dat vergeleken kan worden welke modus van de warmtepomp het beste kan worden toegepast, mogelijk in combinatie met het energieverbruik.

Beoordeling

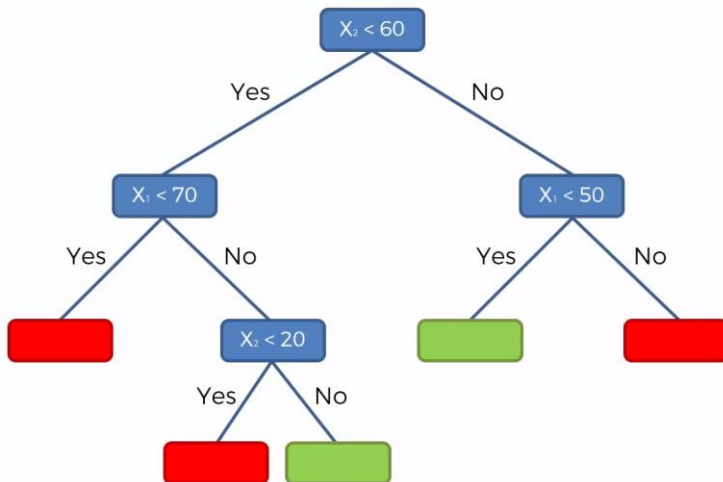
Factor	Score	Toelichting
Intuïtieve toepasbaarheid	4/5	Teruggeven van kans geeft mogelijkheden voor meenemen andere parameters.
Complexiteit	3/5	De formule die het algoritme gebruikt is redelijk complex. Bij grootschalige toepassing neemt de complexiteit toe.
Performance	4/5	Performance kan afhangen van extra parameters.
Accuraatheid	5/5	De kans wordt zeer accuraat teruggegeven en kan zo goed worden vergeleken.
Inzichtelijkheid	5/5	Door het teruggeven van kans is heel helder hoe de keuzes gemaakt worden. Er kan zo makkelijk met alternatieven worden vergeleken.
Totaal	25/30	Algoritme scoort hoog. Enige wat opvalt is de toenemende complexiteit bij grootschalige toepassing. Dit komt door de formule die gebruikt wordt.

Tabel 15. Beoordeling Naïve Bayes.

8.3.4.6. Decision Tree Classification / beslissingsboomalgoritme

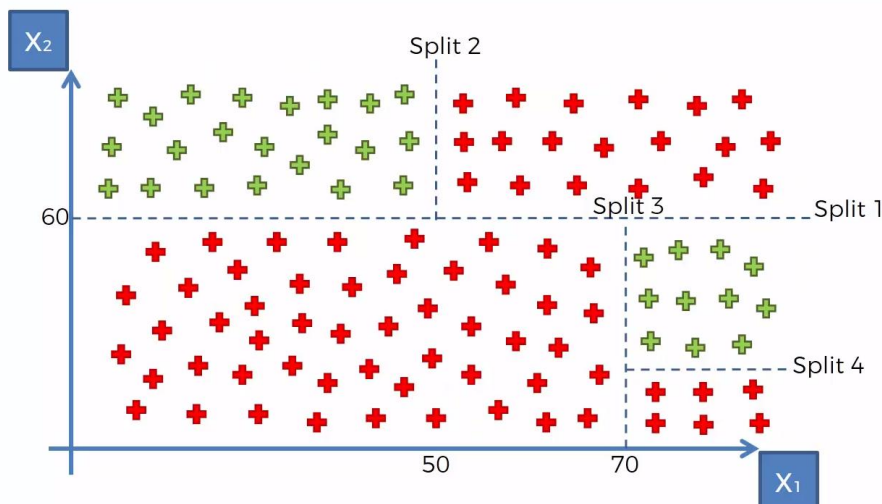
Werking algoritme

Het decision tree classification (beslissingsboomclassificatie) algoritme deelt de bestaande data in door lijnen te trekken in het bestaande dataveld. Dit gebeurt door de data aan een bepaalde randvoorwaarde te onderwerpen. Is $X_2 < 60$ bijvoorbeeld. De al bekende datapunten zijn tot bepaalde groepen gerekend waardoor kan worden uitgerekend wat de grenzen van deze groepen zijn. Het dataveld kan met het bepalen van de lijnen zo in vakken worden verdeeld waar de groepen zich in bevinden. Indien een nieuw datapunt wordt toegevoegd aan de dataverzameling is het op deze wijze gemakkelijk om deze toe te voegen aan een bestaande groep op basis van het vak waarbinnen dit nieuwe datapunt valt.



Figuur 7. Visualisatie voorbeeld beslissingsboom van het decision tree classificatiealgoritme (Eremenko & de Ponteves, 2020).

Bovenstaand figuur laat zien hoe de data wordt onderworpen aan bepaalde voorwaarden. Afhankelijk van het wel of niet voldoen aan de voorwaarde wordt de subset toegedeeld aan de groene groep of de rode groep. Omdat de beslissing de vorm heeft van een boom wordt dit de decision tree (beslissingsboom) genoemd.



Figuur 8. Visualisatie voorbeeld verdeling data door decision tree classificatiealgoritme (Eremenko & de Ponteves, 2020).

Als ondersteuning op de eerdergenoemde beslissingsboom wordt een visualisatie van de beslissingsstappen getoond. Iedere beslissing zorgt voor een splitsing in de data. Hierdoor kan de data uiteindelijk worden geclassificeerd.

Intuïtieve toepassing

Omdat de data in rechte vakken wordt opgedeeld zou het kunnen zijn dat hierdoor de toepassing van dit algoritme voor het Smart Indoor Climate System beperkt is. Echter zou hiervoor eerst moeten worden onderzocht hoe de datapunten van de al bestaande data precies samen clusteren, maar dit zal waarschijnlijk niet gemakkelijk op te delen te zijn in vakken. Dit komt doordat een punt met een lage binnentemperatuur en een hoge buitentemperatuur tot dezelfde modus kan worden gerekend als een punt met een hogere binnentemperatuur en een wat lagere buitentemperatuur. Het feit dat er rechte lijnen getrokken worden is iets wat daarom waarschijnlijk niet goed werkt voor deze specifieke situatie. Dit zou op een flexibelere manier moeten kunnen dan alleen vierkante vakken. Echter zijn er ook voordelen. Zo kunnen er duidelijke grenswaarden worden ingesteld. Dit zou met de temperatuur een goede mogelijkheid kunnen zijn. En beslissingsboom zorgt voor een heldere verdeling van de vakken die uit deze keuzes kunnen ontstaan.

Beoordeling

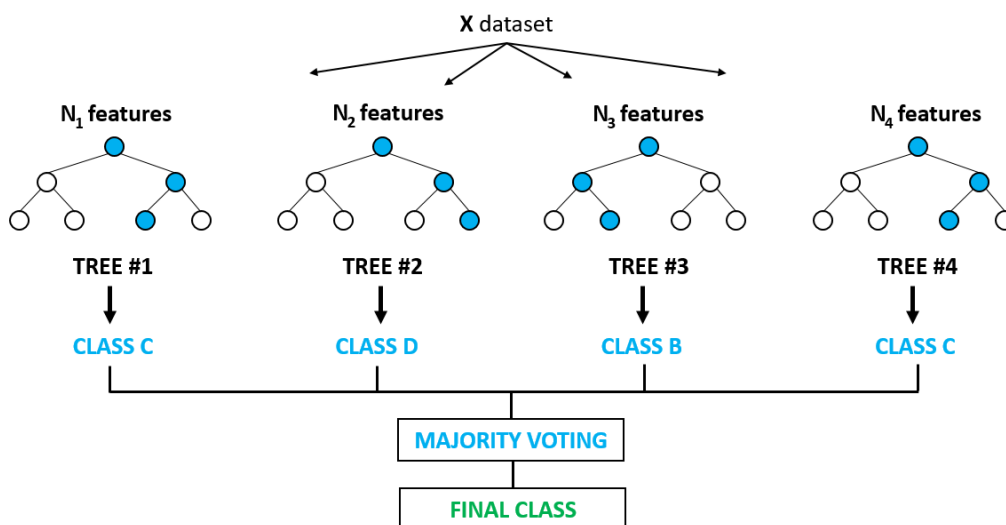
Factor	Score	Toelichting
<i>Intuïtieve toepasbaarheid</i>	3/5	Er zijn voor- en tegens. Rechte vakken zijn een nadeel maar de beslissingsboom maakt het algoritme juist weer aantrekkelijk.
<i>Complexiteit</i>	4/5	De beslissingsboom maakt de keuzes van het algoritme deelbaar in kleine stappen. Complexiteit kan echter toenemen bij een grote boom.
<i>Performance</i>	5/5	Aflopen van voorwaarden zorgt voor gemakkelijke stappen en dus een hoge performance.
<i>Accuraatheid</i>	2/5	Omdat de data alleen in rechthoekige blokken kan worden opgedeeld is het de vraag of dit accurate resultaten oplevert voor deze business case.
<i>Inzichtelijkheid</i>	5/5	De beslissingsboom zorgt voor een helder inzicht in de beslissingen die het algoritme maakt.
<u>Totaal</u>	22/30	Overwegend goede score waarbij voornamelijk de rechte vakken een negatieve invloed hebben. De performance en inzichtelijkheid zorgen weer voor een hogere score.

Tabel 16. Beoordeling Decision Tree Classification / beslissingsboomalgoritme.

8.3.4.7. Random Forest Classification

Werking algoritme

Met random forest classificatie wordt steeds een nieuwe beslissingsboom gemaakt van een willekeurig subset van de data. De manier waarop een boom zich gedraagt is hetzelfde als beschreven in het beslissingsboomalgoritme van de vorige deelparagraaf. Het verschil zit echter in het veelvoud aan bomen. Een enkele beslissingsboom geeft wellicht geen accurate informatie, maar de kracht van dit algoritme zit in het veelvoud waarmee de bomen van willekeurige subsets worden gemaakt. Door de gemiddelde resultaten van alle bomen te gebruiken in de berekening van een nieuw datapunt kan deze zeer nauwkeurig worden ingedeeld in een bepaalde categorie (Ho, 2016). Het is dus een complex algoritme wat de kracht van willekeurigheid en gemiddelden gebruikt om tot relevante resultaten te komen.



Figuur 9. Visualisatie voorbeeld random forest classification (Abilash, 2018).

Intuïtieve toepassing

De intuïtieve toepassing van dit algoritme is lastig. Dit heeft te maken met de ongekennde complexiteit van het algoritme. Door het doorlopen van meerdere bomen kan bij een groot aantal bomen en beslissingen de complexiteit exponentieel toenemen. In deze situatie zal experimenteel onderzoek nodig zijn om daadwerkelijk een goede inschatting te kunnen maken van de toepasbaarheid van dit algoritme. De hypothese is echter dat deze resultaten een stuk accurater zullen zijn dan bij het beslissingsboomalgoritme omdat de resultaten zijn gebaseerd op meerdere bomen dan slechts één boom.

Beoordeling

Factor	Score	Toelichting
Intuïtieve toepasbaarheid	4/5	Naar verwachting beter dan beslissingsboomalgoritme.
Complexiteit	2/5	Door de grote hoeveelheid bomen ontstaat al gauw complexiteit.
Performance	2/5	Veel bomen zorgen voor veel beslissingen wat de performance niet ten goede komt.
Accuraatheid	5/5	Accurater resultaat verwacht dan bij beslissingsboomalgoritme.
Inzichtelijkheid	2/5	De vele bomen maken de situatie minder inzichtelijk.
<u>Totaal</u>	19/30	Hoewel de resultaten waarschijnlijk wenselijker zijn is het algoritme complexer en minder inzichtelijk.

Tabel 17. Beoordeling Random Forest Classification.

8.3.4.8. Deelconclusie deelvraag 3.4

Om te bepalen welke algoritme het best toepasbaar is gebruiken we het beoordelingsmodel zoals dat is gebruikt voor ieder individueel algoritme. De algoritmen met een hogere score blijken uit dit onderzoek beter toepasbaar te zijn. Algoritmen met een lagere score zijn minder goed toepasbaar. Er is echter geen strakke scheiding tussen toepasbare en niet-toepasbare algoritmen. Het onderzoek adviseert op basis van de beoordeling het meest geschikte en toepasbare algoritme. Daaropvolgende algoritmen zijn echter ook goed toepasbaar. Er volgt een ranglijst met de algoritmen op basis van de bepaalde beoordelingen.

Nr.	Totaalbeoordeling	Algoritme	Toelichting
1	28/30	K-Nearest Neighbors	Het KNN-algoritme komt als beste uit de bus vanwege een goede score op complexiteit en inzichtelijkheid.
2	25/30	Naive Bayes	Groot voordeel van Naive Bayes is de kansberekening en vergelijkingsmogelijkheid.
3	22/30	Beslissingsboomalgoritme	Scoort hoog op inzichtelijkheid maar vraagtekens bij rechthoekige vakken.
4	19/30	Random Forest Classification	Hoewel deze waarschijnlijk betere resultaten oplevert dan het beslissingsboomalgoritme, is de complexiteit veel groter. Dit gaat ook ten koste van de inzichtelijkheid.
5	16/30	Kernel SVM	Scoort het laagst op alle onderdelen. Flexibiliteit van vormen is positief maar er zijn zorgen over overlapping door gigantische hoeveelheden modussen.

Tabel 18. Ranglijst toepasbare algoritmen op basis van theoretisch onderzoek.

Bovenstaande ranglijst geeft een samenvatting van het gedane onderzoek. Op basis van de factoren intuïtieve toepasbaarheid, complexiteit, performance, accuraatheid en inzichtelijkheid zijn de algoritmen beoordeeld. Om preciezer na te gaan hoe een score is opgebouwd wordt er verwezen naar de deelparagraaf voor het te raadplegen algoritme.

8.4. DEELVRAAG 4: HOE KAN MET DEZE KUNSTMATIGE INTELLIGENTIE HET ENERGIEVERBRUIK WORDEN VERMINDERD?

Uit de vorige deelvraag blijkt welke algoritmen er geschikt zijn voor het gebruik in het Smart Indoor Climate System. De algoritmen kunnen gemakkelijk worden ingezet om op basis van eerdere patronen een goede keuze te kunnen maken voor de modus van de warmtepomp van vandaag. Maar hoe speelt dit een rol in het verminderen van het energieverbruik?

8.4.1. Impliciete vermindering energieverbruik via modus

Impliciet wordt er al rekening gehouden met het energieverbruik. Dit omdat de modus een attribuut is in de trainingsdata. De modus houdt direct verband met het energieverbruik, zoals gezien kan worden in de eerste deelvraag waarbij de stand, modus en verbruik van de warmtepomp wordt uitgelegd.

Maar de vraag blijft, is deze impliciete benadering voldoende voor het actief verminderen van het verbruik van de warmtepomp? Dit is iets wat uit praktijkonderzoek zal moeten blijken. Hierbij gaat het valideren van de resultaten van de algoritmen d.m.v. van het koppelen van de warmtepomp helpen (zie de aanbevelingen voor meer informatie hierover).

Omdat het praktijkonderzoek nu niet mogelijk is (vanwege de coronacrisis en het niet in gebruik zijn van de warmtepomp) en hierdoor niet direct bewijs naar voren komt dat de implementatie zoals deze uit het onderzoek naar voren komt ook daadwerkelijk actief het energieverbruik van de warmtepomp vermindert, wordt het verstandig geacht om nog een actief element wat dit wél bewerkstelligt te implementeren.

8.4.2. Combi-functionaliteit

Hiervoor is de combi-functionaliteit bedacht. De combi-functionaliteit kan meerdere algoritmen in het systeem uitvoeren en de resultaten vergelijken op basis van het verbruik. Vervolgens kan uit deze mogelijk wisselende resultaten de meest energiezuinige optie worden uitgekozen.

Als er bijvoorbeeld door K-Nearest Neighbors als modus 001111 als resultaat wordt uitgebracht en door Naive Bayes 002233, dan kan door middel van de combi-functionaliteit de beste optie hiervan worden uitgekozen op basis van het laagste energieverbruik.

8.4.3. Deelconclusie deelvraag 4

Doordat de modus van de warmtepomp al een attribuut is in de trainingsdata en deze direct samenhangt met het verbruik van de warmtepomp speelt dit indirect een (passieve) rol bij het verminderen van het energieverbruik van de warmtepomp.

Omdat dit momenteel (nog) niet kan worden bewezen d.m.v. validatie door koppeling aan de warmtepomp, is een (actief) element aan het systeem toegevoegd.

Dit actieve element die aantoonbaar een lager energieverbruik oplevert is de combi-functionaliteit. Hiermee worden meerdere algoritmen uitgevoerd waarna hiervan vervolgens weer de meest energiezuinige optie wordt uitgekozen.

8.5. DEELVRAAG 5: HOE KAN KUNSTMATIGE INTELLIGENTIE GEÏMPLEMENTEERD WORDEN IN HET SYSTEEM?

Nu bepaald is welke algoritmen het meest toepasbaar en geschikt is voor het gebruik in het Smart Indoor Climate System kan dit in het systeem worden geïmplementeerd. Op deze wijze zal het systeem worden voorzien van het kunstmatige intelligentie element. De manier waarop dit gebeurd is door eerst de data op de juiste manier voor te bereiden voor het uitvoeren van het algoritme.

8.5.1. Deelvraag 5.1: Hoe kunnen de toepasbare algoritmen worden geïmplementeerd?

Voor de implementatie van de algoritmen is allereerst gekeken naar 2 Python modules voor het gebruik van machine learning. Scikit-learn en SciPy. Scikit-learn wordt meer gebruikt voor algemene toepassingen waarbij SciPy vaak wordt gebruikt voor wetenschappelijke toepassingen (scikit-learn vs SciPy: What are the differences?, s.d.).

8.5.1.1. Deelconclusie deelvraag 5.1

De implementatie voor het systeem neigt al gauw richting Scikit-learn. De grootste reden hiervoor is dat er bij Ultraware al expertise is op het gebied van Scikit-learn. Hierdoor is het gemakkelijker door te ontwikkelen wanneer het systeem in de toekomst wordt overgedragen aan andere medewerkers van Ultraware. Een tweede reden is dat Scikit-learn ook wordt gebruikt in de Udemy course (Eremenko & de Ponteves, 2020). Als laatste is de toepassing binnen het systeem praktisch en niet wetenschappelijk van aard. Er is daarom besloten om het algoritme in het systeem te implementeren met Scikit-Learn.

Voor de implementatie is de instructie vanuit de Udemy course gevolgd waarbij gebruik wordt gemaakt van Scikit-Learn (Eremenko & de Ponteves, 2020). Hierbij wordt de data allereerst op de juiste manier gepreprocessed waarna het algoritme ermee wordt getraind. Hierna wordt het algoritme uitgevoerd met de temperatuurdata van vandaag. Er wordt gebruik gemaakt van de X en y-waarden zoals deze beschreven zijn in paragraaf 8.3.2.5. Voor de exacte (technische kant van de) implementatie wordt verwezen naar het technisch ontwerp.

Twee algoritmen uit de eerdere deelconclusie worden geïmplementeerd in het systeem. Dit zijn K-Nearest Neighbors en Naive Bayes. Deze scoren namelijk het hoogst op de beoordeling. Ook is de combi-functionaliteit geïmplementeerd in het systeem. De afweging om twee algoritmen te implementeren en niet meer is vanwege de nog restende beschikbare tijd. Het implementeren van slechts één algoritme is ook minder wenselijk omdat resultaten dan niet onderling kunnen worden vergeleken. De combi-functionaliteit zou dan niet mogelijk zijn.

8.5.2. Deelvraag 5.2: Hoe kunnen de resultaten van deze algoritmen worden geverifieerd?

Om te controleren of de resultaten (de output) van de algoritmen betrouwbaar en correct zijn dienen deze te worden geverifieerd. Hiervoor zijn enkele opties overwogen. Allereerst het testen van verschillende scenario's. Hierbij wordt er gemanipuleerde data in het systeem ingeladen en een voorspelling gedaan van de output van het algoritme. Indien de hypothese van de output overeenkomt met de werkelijkheid kan worden geconcludeerd dat het algoritme naar verwachting functioneert. De werking en implementatie is dan geverifieerd. Een andere mogelijkheid is om de het systeem te implementeren en te koppelen aan de warmtepomp en over een bepaalde periode (bijv. over een maand) te kijken of het energieverbruik aantoonbaar lager uitvalt dan zonder de implementatie van het systeem.

8.5.2.1. *Deelconclusie deelvraag 5.2*

Voor beide aanpakken is wat te zeggen. Vanuit een wetenschappelijke benadering zou het uitvoeren van beide verificatiemogelijkheden het beste de betrouwbaarheid van de output aan kunnen tonen. Besloten is echter om de betrouwbaarheid van de resultaten van deze algoritmen alléén aan te tonen met de testscenario's. Dit is echter niet een geheel vrijwillige keuze. Gezien de omstandigheden omtrent het coronavirus is het momenteel namelijk erg ingewikkeld om de warmtepomp daadwerkelijk aan te sluiten. Het bedrijfspand wordt bijna niet gebruikt en dit zou ook onnodig verbruik opleveren.

Voor de daadwerkelijke scenariotesten en resultaten wordt verwezen naar het testplan en testrapport. Deze documenten zijn opgesteld om de resultaten te verifiëren en de betrouwbaarheid aan te tonen.

9. EINDCONCLUSIE

De eindconclusie wordt vormgegeven door stapsgewijs een beeld te geven bij het uitgevoerde onderzoek en een samenvatting te geven van de verschillende deelconclusies. Het doel van het onderzoek wordt neergezet en aan het einde van deze eindconclusie wordt uitgelegd hoe het doel van dit onderzoek is behaald.

Het doel van dit onderzoek is om te bepalen hoe een bedrijfspand door middel van kunstmatige intelligentie het beste kan worden voorverwarmd zodat het energieverbruik kan worden verminderd en het comfort kan worden behouden.

De warmtepomp heeft een bepaalde stand. Over een bepaald tijdsverloop wordt dit een modus genoemd. Uit de data van de API's kan worden geconcludeerd welke modus van de warmtepomp is gebruikt. Ook is data zoals de binnen- en buitentemperatuur beschikbaar en deze kan worden gebruikt om de warmtepomp slim aan te sturen.

Voor het gebruik van kunstmatige intelligentie in de toepassing van de aansturing van de warmtepomp met de oude data van binnen- en buitentemperatuur en bijbehorende modus is er sprake van de vorm machine learning. Er wordt immers geleerd met de oude data om een nieuwe beslissing te kunnen maken.

Vervolgens is de relatie in de relevante data (uit de API's) onderzocht. Hieruit blijkt dat de begintemperatuur, de eindtemperatuur, delta tijd (d tijd) en de buitentemperatuur begin invloed hebben op de modus van de warmtepomp. Deze relatie kan worden gebruikt om van dit verband te leren en hierin nieuwe situaties keuzes op te kunnen baseren.

Om deze keuzes te kunnen maken zal er een machine learning algoritme worden gebruikt. Om een toepasbaar en geschikt machine learning algoritme uit te kunnen kiezen wordt er eerst gekeken naar categorieën algoritmen. Hierna pas naar individuele algoritmen. Uit dit deelonderzoek blijkt dat het algoritme die de oplossing voor de business biedt een niet-lineair classificatiealgoritme is.

Hierna zijn alle niet-lineaire classificatiealgoritmen onder de loep genomen. Er zijn beoordelingscriteria opgesteld om al deze algoritmen te kunnen vergelijken. Uit het deelonderzoek blijkt dat K-Nearest Neighbors en Naive Bayes het beste scoren op deze beoordeling.

Met deze algoritmen treedt er impliciet een vermindering op van het energieverbruik, maar dit is zonder praktijktests lastig aan te tonen. Daarom is er een combi-functionaliteit ontwikkeld om op een actieve en expliciete manier het energieverbruik te verminderen.

Zowel K-Nearest Neighbors als Naive Bayes zijn geïmplementeerd in het systeem (SICS) met gebruik van de Scikit-learn module in Python. Ook is er een combi-functionaliteit geïmplementeerd om te zorgen voor expliciete aantoonbare vermindering van het energieverbruik. Dit wordt op basis van het onderzoek gezien als de meest geschikte manier om kunstmatige intelligentie in te zetten voor het voorverwarmen van een bedrijfspand en zo het energieverbruik te verminderen met behoud van comfort.

10. DISCUSSIE

De validiteit van dit onderzoek is zo veel mogelijk gewaarborgd. Er zijn echter een aantal punten in dit onderzoek waarbij er keuzes moesten worden gemaakt invloed zouden kunnen hebben op de validiteit. Deze punten worden in dit hoofdstuk besproken.

10.1. BEPALEN BEOORDELINGSCRITEIA ALGORITMESELECTIE

De beoordelingscriteria zijn op basis van intuïtie bepaald en zijn in overleg met Ultraware vastgesteld. Dit betekent echter niet dat deze beoordelingscriteria niet op een andere wijze hadden kunnen worden vastgesteld. Er hadden op basis van andere prioriteiten of een andere gebruikscasus ook andere beoordelingscriteria opgesteld kunnen worden. Mogelijkerwijs zou dit in het meest extreme geval tot een andere eindconclusie kunnen lijden. Echter is het selecteren van de beoordelingscriteria zorgvuldig uitgevoerd en hierdoor kan er voor deze specifieke toepassing worden bevestigd dat de algoritmen uit de eindconclusie wel degelijk geschikt zijn en een oplossing bieden voor de business.

10.2. UITSLUITEN OMGEVINGSFACTOREN

Dit is voornamelijk een keuze geweest en een geluid vanuit de business. De belangrijkste reden hiervoor is natuurlijk de schaarste aan tijd, maar ook de prioritering bij Ultraware. Het is daarom logisch dat dit punt vanuit een wetenschappelijk oogpunt ter discussie wordt gesteld. De omgevingsfactoren hadden meegenomen kunnen worden in het systeem. Dit had het systeem echter nog veel complexer gemaakt en was niet realistisch voor deze afstudeerperiode. Dat neemt niet weg dat dit meegenomen wordt in de aanbevelingen.

11. AANBEVELINGEN

Dit hoofdstuk bevat een aantal aanbevelingen die op basis van het uitgevoerde onderzoek worden gedaan aan het management van Ultraware voor de verdere doorontwikkeling van het systeem of bij het verwerken van dit onderzoek. De genoemde aanbevelingen zijn slechts op basis van het onderzoek in deze scriptie. Voor een volledig beeld van aanbevelingen en adviezen wordt verwezen naar het adviesrapport.

11.1. MEENEMEN OMGEVINGSFACTOREN

Om tot nog accuratere resultaten te komen in een toekomstige versie van het Smart Indoor Climate System, wordt aanbevolen toch de omgevingsfactoren mee te nemen in de berekening van de juiste modus van de warmtepomp. Hiervoor zou een heroverweging van de relevante data nodig zijn maar daar kan dit onderzoek wel een belangrijke basis voor bieden. De verwachting is dat dezelfde algoritmen hier ook het beste voor toepasbaar zijn maar toch wordt geadviseerd om ook hier opnieuw naar te kijken. Het zou afhankelijk kunnen zijn van nieuwe omgevingsfactoren welke algoritmen er het beste uit de bus komt. Overigens kan ook hier het onderzoek met de huidige onderzoeksresultaten waardevolle input geven als opstart naar deze uitbreiding.

11.2. VALIDEREN RESULTATEN ALGORITMEN D.M.V. KOPPELING WARMTEPOMP

Als de coronacrisis voorbij is en het gebruik van het systeem weer realistisch in de praktijk kan worden getest, wordt er aanbevolen om ook daadwerkelijk praktijktests te gaan doen. Door het systeem te koppelen aan de warmtepomp en over een periode het energieverbruik te vergelijken met een periode waarbij het systeem nog niet gekoppeld was (bijv. februari 2021 t.o.v. februari 2020) kan de werking van het systeem nog steviger worden bewezen en een overstap naar dit systeem worden gemotiveerd. Daarbij wordt ook aanbevolen om de resultaten te rapporteren, bijvoorbeeld door het uitbreiden van het testplan en testrapport.

11.3. UITBREIDEN KEUZE ALGORITMEN

Voor het systeem zou een uitbreiding van de keuzen voor de algoritmen een goede toevoeging zijn. Het zou namelijk kunnen dat in een bepaalde praktijksituatie een ander algoritme weer betere resultaten oplevert, iets wat alleen uit praktijktests zou kunnen blijken maar dit is wel aannemelijk. Bovendien kan de combi-functionaliteit dan een keuze maken uit nog meer resultaten wat mogelijk een nog zuiniger systeem oplevert.

12. LITERATUURLIJST

- Abilash, R. (2018, juli 31). *APPLYING RANDOM FOREST (CLASSIFICATION) — MACHINE LEARNING ALGORITHM FROM SCRATCH WITH REAL DATASETS*. Opgehaald van Medium.com: <https://medium.com/@ar.ingenious/applying-random-forest-classification-machine-learning-algorithm-from-scratch-with-real-24ff198a1c57>
- Brownlee, J. (2016, maart 11). *How Machine Learning Algorithms Work (they learn a mapping of input to output)*. Opgehaald van Machinelearningmastery.com: <https://machinelearningmastery.com/how-machine-learning-algorithms-work/>
- Brownlee, J. (2016, maart 16). *Supervised and Unsupervised Machine Learning Algorithms*. Opgehaald van Machinelearningmastery.com: <https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/>
- Duursma, J. (2020). *Wat is een neurale netwerk?* Opgehaald van Studio Overmorgen: <https://www.jarnoduursma.nl/is-neuraal-netwerk/>
- Eremenko, K., & de Ponteves, H. (2020, april 2). *Machine Learning A-Z™: Hands-On Python & R In Data Science*. Opgehaald van Udemy.com: <https://www.udemy.com/course/machinelearning/>
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Miscellaneous Clustering Methods*. Chichester: John Wiley & Sons, Ltd.
- Ho, T. K. (2016, april 17). *Random Decision Forests*. Opgehaald van <https://web.archive.org/web/20160417030218/http://ect.bell-labs.com/who/tkh/publications/papers/odt.pdf>
- Hulsen, P. (sd). *Wat is (het verschil tussen) Artificial Intelligence, Machine Learning en Deep Learning*. Opgehaald van cqm.nl: <https://cqm.nl/nl/nieuws/wat-is-het-verschil-tussen-artificial-intelligence-machine-learning-en-deep-learning>
- Kashid, V. (2018, maart 8). *What is the difference between regression, classification and clustering in machine learning?* Opgehaald van Quora.com: <https://www.quora.com/What-is-the-difference-between-regression-classification-and-clustering-in-machine-learning>
- Kunstmatige intelligentie*. (2020, januari 2). Opgehaald van Wikipedia: https://nl.wikipedia.org/wiki/Kunstmatige_intelligentie
- Murray, F. J. (2019, juni 14). *Logarithm, mathematics*. Opgehaald van Encyclopædia Britannica: <https://www.britannica.com/science/logarithm>
- scikit-learn vs SciPy: What are the differences?* (sd). Opgehaald van Stackshare.io: <https://stackshare.io/stackups/scikit-learn-vs-scipy>
- Scribbr. (sd). *Overzicht van onderzoeksmethoden en dataverzamelingmethoden*. Opgehaald van Scribbr.nl: <https://www.scribbr.nl/category/onderzoeksmethoden/>
- Seif, G. (2018, januari 21). *Deep Learning for Image Recognition: why it's challenging, where we've been, and what's next*. Opgehaald van Towardsdatascience.com: <https://towardsdatascience.com/deep-learning-for-image-classification-why-its-challenging-where-we-ve-been-and-what-s-next-93b56948fcef>
- Tamminga, E. (2020, Maart 10). *Controller front-end API*. Opgehaald van Sinuss Git: <https://git.sinuss.nl/SIC/controller/-/blob/master/proxy/swagger.yaml>

TextMechanic. (2018). *Combination Generator*. Opgehaald van TextMechanic.com:

<https://textmechanic.com/text-tools/combination-permutation-tools/combination-generator/>

Themische geleidbaarheid. (2019, november 19). Opgehaald van Wikipedia:

https://nl.wikipedia.org/wiki/Thermische_geleidbaarheid

Touger, E. (2018, augustus 3). *What's the Difference Between Artificial Intelligence (AI), Machine Learning, and Deep Learning?* Opgehaald van Prowess: <https://www.prowesscorp.com/whats-the-difference-between-artificial-intelligence-ai-machine-learning-and-deep-learning/>

Ultraware. (2019). *Machine learning voor een comfortabel en energiezuinig binnenklimaat*. Assen.

University of Minnesota Morris. (sd). *Classification & Regression Trees*. Opgehaald van University of Minnesota Morris: <http://mnstats.morris.umn.edu/multivariatestatistics/cart.html>