

Automatische bijsturing en spreiding van het elektriciteitsverbruik van ge- zinnen met behulp van artificiële intel- ligentie (AI).

Jan Tubeckx.

Scriptie voorgedragen tot het bekomen van de graad van
Professionele bachelor in de toegepaste informatica

Promotor: Dhr. F. Labijn
Co-promotor: Dhr. T. Tomme
Academiejaar: 2023–2024
Eerste examenperiode

Departement IT en Digitale Innovatie .



Woord vooraf

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Samenvatting

Door de omschakeling naar hernieuwbare energie wordt volop ingezet op elektrificatie. Dit vraagt enerzijds een modernisering van het distributienetwerk voor elektriciteit, maar ook de inzet van slimme technologieën om dit distributienetwerk te ontlasten. Een voorbeeld hiervan is de digitale elektriciteitsmeter die met behulp van allerhande apps 'slim' kan gemaakt worden en gezinnen zo meer inzicht kan geven in hun elektriciteitsverbruik zodat ze dit beter kunnen spreiden.

Dit onderzoek zal trachten te achterhalen hoe een app die een digitale elektriciteitsmeter aanstuurt kan geoptimaliseerd worden met behulp van artificiële intelligentie (AI). De meeste van de apps die momenteel bestaan geven immers enkel op een passieve manier weer wat het huidige verbruik en de huidige productie is. Er wordt van de verbruiker verwacht dat hij of zij op basis van deze inzichten de nodige acties onderneemt om het elektriciteitsverbruik bij te sturen. De Vlaamse Regulator van de Elektriciteits- en Gasmarkt (VREG) (2021) heeft vastgesteld dat het slim maken van digitale elektriciteitsmeters voorlopig niet tot de gewenste gedragsverandering en spreiding van het elektriciteitsverbruik leidt.

In het eerste deel van het onderzoek zal worden nagegaan welke apps er reeds ontwikkeld werden en wat hun tekortkomingen zijn. Het verdere onderzoek zal vervolgens een app gaan ontwikkelen die met behulp van artificiële intelligentie en een weer API het elektriciteitsverbruik automatisch en dus actief zal bijsturen en spreiden. Meer concreet zal AI worden gebruikt om voorspellingen te doen op basis van historische zonnestralingsgegevens. Dit resultaat zal daarna verder verfijnd worden met weersvoorspellingen. De app zal op basis van de gecombineerde voorspellingen elektrische toestellen automatisch gaan in- of uitschakelen. Met deze bachelorproef zal in de eerste plaats geprobeerd worden om twee slimme stekkers aan te sturen. Deze stekkers kunnen via wifi en dus een app bediend worden. Op die manier kunnen sanitaire toestellen die zelf geen wifi-verbinding kunnen maken, toch 'slim' gemaakt worden. De gebruiker van de app zal wel steeds de mogelijkheid geboden worden om manueel tussen te komen.

Het verwachte resultaat is dat het elektriciteitsverbruik op die manier wel meer gestuurd en gespreid zal worden. Dit zal niet alleen tot een ontlasting van het distributienetwerk voor elektriciteit leiden, maar zal ook de energiekosten voor de gezinnen drukken.

Inhoudsopgave

Lijst van figuren	vi
1 Inleiding	1
1.1 Probleemstelling	1
1.2 Onderzoeks vraag	4
1.3 Onderzoeksdoelstelling	4
1.4 Opzet van deze bachelorproef	5
2 Stand van zaken	6
2.1 Overzicht functionaliteiten bestaande apps	6
2.2 Voorspellingen maken met AI	7
2.2.1 Machine learning	7
2.2.2 Supervised machine learning: regressie modellen	9
2.2.3 Machine learning modellen evalueren	14
2.3 Elektriciteitsproductie voorspellen	15
2.3.1 Zonnestraling	16
2.3.2 Meteorologische data	16
3 Methodologie	18
4 Proof of Concept	19
4.1 Ontwikkeling van de mobiele app	19
4.1.1 Digitale elektriciteitsmeter uitlezen	19
4.1.2 Omvormer zonnepanelen uitlezen	20
4.1.3 Stroomproductie van zonnepanelen voorspellen met XGBoost .	21
4.1.4 Weergave uitgelezen data en voorspelling met een iOS app .	26
4.1.5 Aansturing slimme stekkers op basis van voorspelde stroomproductie	26
5 Conclusie	28
A Onderzoeksvoorstel	29
Bibliografie	30

Lijst van figuren

1.1	Verwachte evolutie van het aantal elektrische personenvoertuigen in Vlaanderen (Verdoodt, 2022)	2
1.2	Verwachte evolutie van het aandeel warmtepomp en -net bij residentiële verwarming in Vlaanderen (Verdoodt, 2022)	2
1.3	Verwachte evolutie van zonnepanelen in Vlaanderen (Verdoodt, 2022)	3
1.4	Verwachte evolutie van de elektrificatie van zwaar vachtvervoer in Vlaanderen (Verdoodt, 2022)	3
2.1	Grafische voorstelling beslissingsboom.	9
2.2	Grafische voorstelling van het Random Forest algoritme.	10
2.3	Grafische voorstelling Support Vector Machine.	11
2.4	Grafische voorstelling van het LSTM algoritme.	13
2.5	Grafische voorstelling van de structuur van het XGBoost algoritme.	14
4.1	Testopstelling: Raspberry Pi 5 via een RS422 serieel naar USB kabel aangesloten op een digitale elektriciteitsmeter.	20
4.2	Omvormer van de zonnepanelen met een wifi-stick.	21
4.3	Script voor het opvragen van de zonnestralingsdata.	23
4.4	Script voor het opvragen van de weerdata.	24
4.5	Heatmap van correlaties tussen de verschillende kenmerken van de dataset.	25
4.6	Screenshot van de Python code voor het aanmaken van features en lags.	26
4.7	Screenshot dagweergave.	27

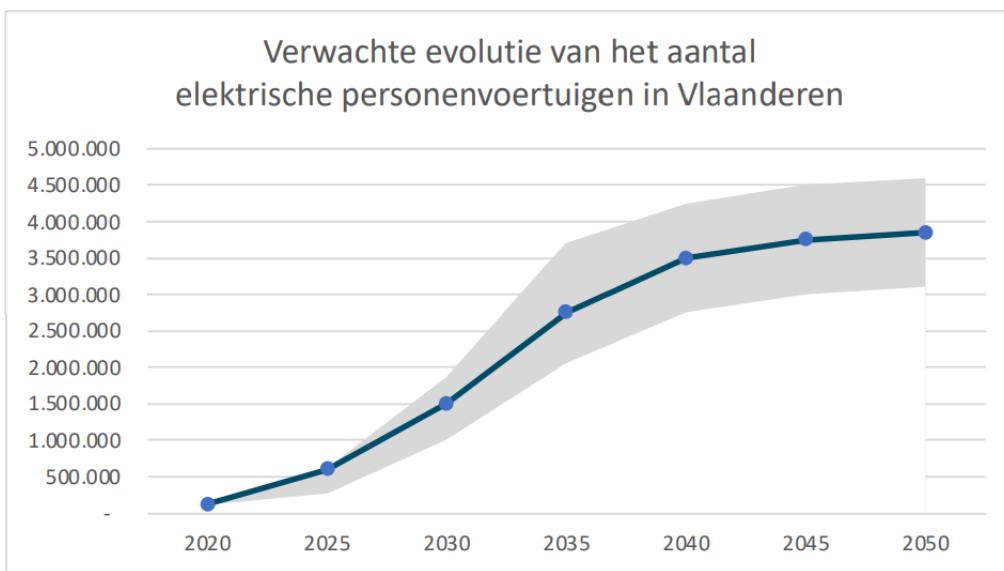
1

Inleiding

1.1. Probleemstelling

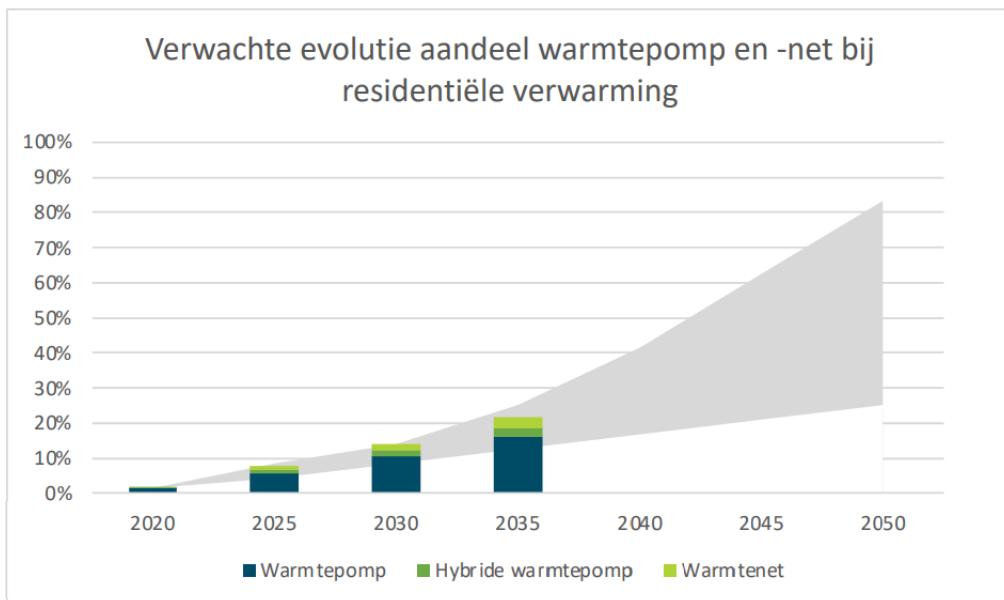
Anderhalf miljoen elektrische wagens, massaal veel warmtepompen en zonnepanelen gecombineerd met een toenemende elektrificatie van de industrie en het vrachtvervoer tegen 2030. Dat is de verwachting van de Vlaamse distributienetbeheerder Fluvius (Verdoodt, 2022). De energietransitie die vanuit Europa, België en Vlaanderen wordt aangestuurd, vormt een enorme uitdaging voor de distributienetten voor elektriciteit in Vlaanderen.

Om al te hoge investeringen in de distributienetten te vermijden, wordt door Fluvius ingezet op maatregelen die de belasting van het elektriciteitsnet kunnen verminderen en spreiden. Een van die maatregelen is de invoering en verplichting van de digitale elektriciteitsmeter. Tegen juli 2029 moeten alle Belgische huishoudens een digitale meter hebben. Deze digitale meters kunnen met behulp van apps 'slimmer' gemaakt worden. Zo kunnen gezinnen hun elektriciteitsverbruik in detail opvolgen en bijsturen wat hen meteen ook een besparing zal opleveren. Deze apps vereisen echter een actieve opvolging van de gebruiker. Die moet zelf toestellen in- of uitschakelen om het elektriciteitsverbruik bij te sturen. Uit onderzoek is echter gebleken dat deze actieve bijsturing na verloop van tijd vermindert (Wemyss e.a., 2019). Een enquête van de Vlaamse Regulator van de Elektriciteits- en Gasmarkt Vlaamse Regulator van de Elektriciteits- en Gasmarkt (VREG) (2021) waarin 1.000 Vlaamse gezinnen en 1.500 bedrijven bevraagd werden over hun ervaring en gedrag op de energemarkt toont zelfs aan dat van de 60% van de gezinnen met een digitale meter, slechts 8% hun energieverbruik effectief bijstuurt. Door de invoering van het capaciteitstarief, een andere maatregel waarmee de VREG het elektriciteitsnet hoopt te onlasten, zal de bijsturing en vooral spreiding van het elektriciteitsverbruik evenwel nog belangrijker worden voor gezinnen. Sinds



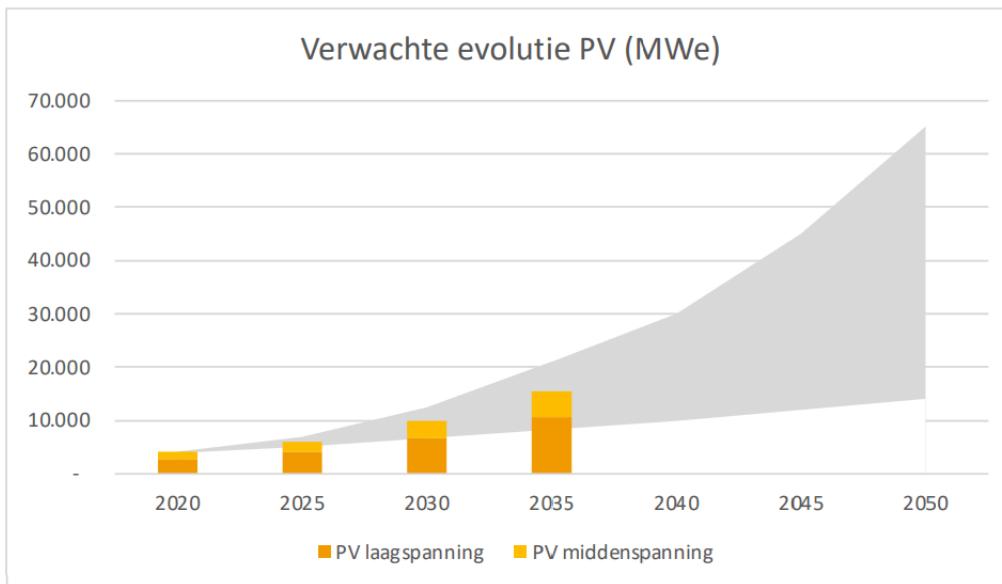
Figuur (1.1)

Verwachte evolutie van het aantal elektrische personenvoertuigen in Vlaanderen (Verdoodt, 2022).

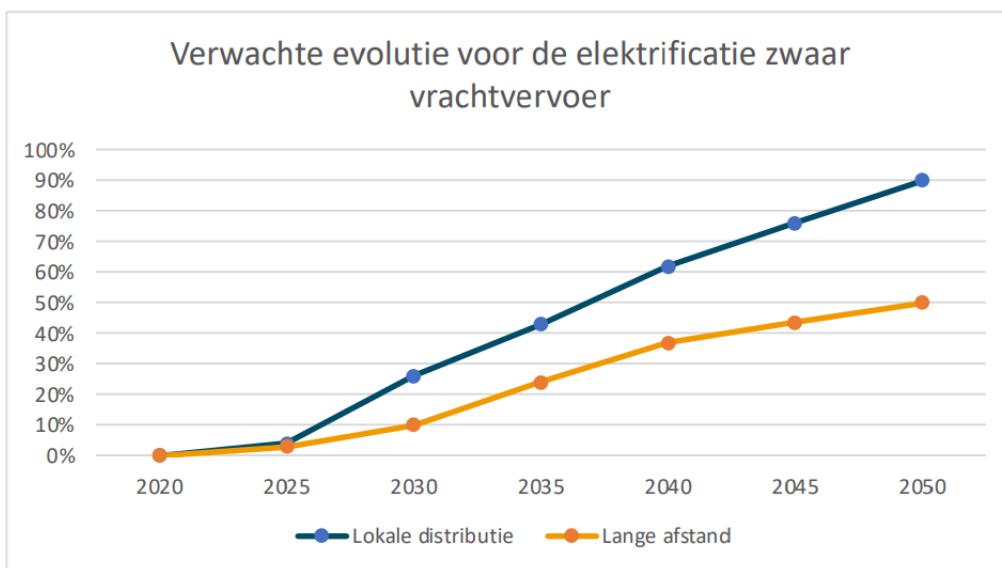


Figuur (1.2)

Verwachte evolutie van het aandeel warmtepomp en -net bij residentiële verwarming in Vlaanderen (Verdoodt, 2022).

**Figuur (1.3)**

Verwachte evolutie van zonnepanelen in Vlaanderen (Verdoodt, 2022).

**Figuur (1.4)**

Verwachte evolutie van de elektrificatie van zwaar vrachtvervoer in Vlaanderen (Verdoodt, 2022).

1 januari 2023 wordt nameijk een deel van de nettarieven die een gezin via haar elektriciteitsfactuur betaalt ook berekend op het maximale elektriciteitsverbruik. Er wordt dus voortaan ook gekeken naar de maximale capaciteit die de distributienetbeheerders ter beschikking moeten stellen. Bedoeling van het capaciteitstarief is dat gezinnen hun stroomverbruik beter gaan spreiden. Als iedereen op hetzelfde moment veel stroom verbruikt, kan het net overbelast raken. Het gevolg is dat netbedrijven dan extra moeten investeren in zwaardere elektriciteitskabels om die hogere verbruikspieken op te vangen (Selleslagh, 2022).

Daarom wordt meer en meer de vraag gesteld hoe nieuwe technologieën en technieken een oplossing kunnen bieden om gezinnen te helpen bij het spreiden van hun elektriciteitsverbruik. Kunnen apps die het elektriciteitsverbruik monitoren slimmer gemaakt worden zodat ze geen tussenkomst van een gebruiker meer behoeven? Kan bijvoorbeeld artificiële intelligentie (AI) daarbij een oplossing bieden? Om dit na te gaan, zal tijdens dit onderzoek als proof of concept een app ontwikkeld worden die het elektriciteitsverbruik van gezinnen automatisch zal sturen met behulp van artificiële intelligentie.

1.2. Onderzoeksvoraag

Kan het elektriciteitsverbruik van gezinnen automatisch bijgestuurd en gespreid worden met behulp van artificiële intelligentie om zo de elektriciteitskost te drukken?

Kan een app die de stroomproductie van zonnepanelen voorspelt door toepassing van machine learning (ML) ervoor zorgen dat het elektriciteitsverbruik beter wordt afgestemd op de zelf geproduceerde stroom?

1.3. Onderzoeksdoelstelling

Om na te gaan of het elektriciteitsverbruik van een gezin automatisch kan gestuurd en gespreid worden, zal een app ontwikkeld worden die in de eerste plaats het elektriciteitsverbruik en de zelf geproduceerde stroom van zonnepanelen inzichtelijk maakt. Om het elektriciteitsverbruik en de geproduceerde stroom zo goed mogelijk op elkaar af te stemmen zal met behulp van machine learning de stroomproductie van de volgende dag voorspeld worden. Hiervoor zal gebruik gemaakt worden van historische zonnestralingsgegevens verkregen via de CAMS Radiation Service (CRS) van de Copernicus Atmosphere Monitoring Service (CAMS). Deze voorspelling zal verder verfijnd worden met de weersvoorspelling van de Open-Meteo weer API. De app zal vervolgens op basis van de gemaakte voorspellingen twee slimme stekkers gaan aansturen, zodat het elektriciteitsverbruik zo nauwkeurig mogelijk samenvalt met de voorspelde elektriciteitsproductie.

1.4. Opzet van deze bachelorproef

De rest van deze bachelorproef is als volgt opgebouwd:

In Hoofdstuk 2 wordt vooreerst een overzicht gegeven van de apps die er momenteel bestaan om een digitale elektriciteitsmeter slimmer te maken. Er zal kort worden ingegaan op de functionaliteiten die ze bieden om het elektriciteitsverbruik van gezinnen bij te sturen en wat hun tekortkomingen zijn. Vervolgens zal een vergelijking gemaakt worden van de verschillende machine learning modellen die kunnen gebruikt worden om voorspellingen te maken.

In Hoofdstuk 3 wordt de ontwikkeling van een app als proof of concept besproken. Er zal worden toegelicht hoe de stroomproductie van zonnepanelen kan voorspeld worden en hoe deze voorspellingen kunnen gebruikt worden om slimmer stekkers en toestellen aan te sturen. Met de ontwikkeling van deze app zal de onderzoeks-vraag proberen beantwoord worden.

In Hoofdstuk ??, tenslotte, wordt de conclusie gegeven en een antwoord geformuleerd op de onderzoeks vragen. Daarbij wordt ook een aanzet gegeven voor toekomstig onderzoek binnen dit domein.

2

Stand van zaken

Met een digitale elektriciteitsmeter kunnen gezinnen hun elektriciteitsverbruik makkelijk opvolgen. Dat kan in de eerste plaats gratis via het online energieportaal [Mijn Fluvius](#). Daarnaast bestaan er ook heel wat gratis of betalende online apps die met een digitale meter kunnen verbonden worden om elektriciteitsverbruik op te volgen en eventueel bij te sturen. Zo'n slimme toepassingen heten in het jargon 'HEMS' (Home Energy Management System) of 'CEMS' (Customer Energy Management System). De aansluiting van deze apps gebeurt via de gebruikerspoorten (P1 en S1) van de digitale elektriciteitsmeter. Beide gebruikerspoorten zijn complementair en geschikt voor verschillende toepassingen. De P1-poort stuurt de elektriciteitsdata per seconde uit. Via de 'snelle' S1-poort worden ruwe data aan een zeer hoge frequentie ter beschikking gesteld aan een app of slimme thermostaat. Dit laat gedetailleerde verbruikersfeedback en sturing toe. Recente digitale elektriciteitsmeters hebben evenwel geen S1-poort meer. De verbinding tussen het meettoestel en een app gebeurt in de meeste gevallen via een wifiverbinding of 4G. Een overzicht van deze toepassingen vindt men op de website [www.maakjemeterslim.be](#).

2.1. Overzicht functionaliteiten bestaande apps

De bestaande apps bieden allemaal de mogelijkheid om elektriciteitsverbruik in real time op te volgen. Zo kan de gebruiker per dag, per week of per maand na gaan hoeveel elektriciteit er verbruikt en/of geproduceerd werd. Het meten van de energiekosten per huishoudtoestel is ook meestal standaard voorzien, maar geldt even vaak als een optie. De meeste apps sporen ook sluipterverbruik op en bieden gepersonaliseerde tips voor energiebesparing. Sommige apps bieden tenslotte ook de mogelijkheid om het gemeten elektriciteitsverbruik naast weerdata te leggen en op die manier na te gaan hoeveel stroom men verbruikt bij bepaalde weersomstandigheden (Deman, 2021). Toen ik vorig jaar mijn onderzoeksvoorstel voor deze

bachelorproef formuleerde en uitschreef voor het vak 'Research methods' was er nog geen enkele app die de inzichten in het elektriciteitsverbruik gebruikte om automatisch apparaten aan te sturen. Het was toen aan de gebruiker zelf om op basis van de gegevens die de app hem toonde actie te ondernemen en toestellen te gaan in- of uitschakelen. Zoals reeds eerder werd vermeld, leidde dit niet altijd tot de gewenste gedragsverandering (Wemyss e.a., 2019), (Mack e.a., 2019) en (Vlaamse Regulator van de Elektriciteits- en Gasmarkt (VREG), 2021). Ondertussen zijn een aantal apps daarom uitgebreid met de optie om (huishoud)apparaten automatisch in te schakelen via slimme stekkers. Zes apps voorzien momenteel in de optie om toestellen in te schakelen volgens elektriciteitstarieven of het overschat aan zelf geproduceerde stroom van zonnepanelen, daarbij telkens rekening houdend met het piekverbruik zodat een hoger capaciteitstarief vermeden wordt. Dit gebeurt met bijgeleverde slimme stekkers die via wifi kunnen aangestuurd worden. De automatische aansturing gebeurt echter steeds op het moment zelf, wanneer de app merkt dat er een overschat aan elektriciteitsproductie is of de elektriciteitsprijzen laag zijn. Tot op heden is er evenwel nog geen app die gebruik maakt van artificiële intelligentie om de zelf geproduceerde zonne-energie te voorspellen. Dit laat de gebruiker immers toe om de vaatwasser of wasmachine op voorhand in te laden, zodat de automatische aansturing ervan de volgende dag ook effectief kan gebeuren.

2.2. Voorspellingen maken met AI

Artificiële intelligentie (AI) verwijst naar de mogelijkheid van applicaties of machines om menselijke vaardigheden te tonen, zoals redeneren, leren en plannen. De term AI wordt vaak gebruikt om andere gerelateerde technieken aan te duiden, zoals machine learning (ML) en deep learning. ML en deeplearning kunnen inderdaad onder de verzamelnaam AI vallen, maar omgekeerd is dit niet altijd het geval. De verschillende technieken van AI hebben immers elk hun eigen doel en opzet. Zo is ML bijvoorbeeld gericht op de studie en de ontwikkeling van algoritmen die kunnen leren of hun prestaties kunnen verbeteren op basis van de data waarmee ze gevoed worden. Machine learning is daarom zeer geschikt om voorspellingen te maken op basis van data uit het verleden.

2.2.1. Machine learning

Machine learning kan opnieuw onderverdeeld worden in drie verschillende categorieën, namelijk supervised learning, unsupervised learning en reinforcement learning.

Supervised learning

Bij supervised machine learning wordt een algoritme door de mens aangeleerd welke conclusies het moet trekken. Er wordt daarbij gewerkt met gelabelde data, wat betekent dat voor elke invoer de gewenste uitvoer bekend is. Het doel van het

algoritme is om een model te bouwen dat de relatie tussen invoer en uitvoer begrijpt en deze kan toepassen op nieuwe, ongeziene gegevens. Indien een waarde of getal voorspeld wordt, spreekt men van een regressie. Heeft de voorspelling betrekking op een groep of categorie, dan is er sprake van classificatie (Brownlee, 2023).

Een voorbeeld van supervised learning is het voorspellen van huizenprijzen. Een model wordt daarbij getraind op een dataset van huizen waarvan de verkoopprijzen bekend zijn, samen met relevante kenmerken zoals locatie, grootte, aantal slaapkamers, enz. Het model leert vervolgens de relatie tussen deze kenmerken en de verkoopprijs van het huis. Als het model dan een nieuw huis met bijhorende kenmerken herkent, zal het de verkoopprijs ervan kunnen voorspellen.

Unsupervised learning

Unsupervised machine learning verloopt op een meer zelfstandige manier. Hierbij leert een algoritme om complexe processen en patronen te identificeren zonder de begeleiding van een mens. Bij unsupervised machine learning vindt training plaats op basis van data die geen labels of een specifieke, vooraf gedefinieerde output hebben (Brownlee, 2023).

Het misschien wel meest gekende voorbeeld van unsupervised learning zijn de aanbevelingen die streamingsplatformen zoals Netflix of Spotify aan hun kijkers aanbieden. Algoritmes analyseren het kijk- en luistergedrag van de gebruikers om daarin patronen en voorkeuren te ontdekken. Vervolgens kunnen dan gepersonaliseerde aanbevelingen gedaan worden.

Reinforcement learning

Bij reinforcement leert een algoritme of computersysteem hoe het moet handelen in een omgeving om een bepaald doel te bereiken. Dit gebeurt door het uitvoeren van acties en het ontvangen van feedback in de vorm van beloningen of straffen. Wat reinforcement learning specifiek maakt is dat het leerproces vergelijkbaar is met hoe een mens of dier zou leren door trial-and-error en beloningen (Efimov, 2024).

Reinforcement learning wordt gebruikt in robotica, bijvoorbeeld om robots autonom te laten navigeren in een omgeving. Om een robot obstakels te leren vermijden wordt gewerkt met beloningen en straffen. Voor elke correcte beweging richting het doel zal de robot een beloning ontvangen. Omgekeerd zal de robot bestraft worden voor elke incorrecte beweging die gemaakt wordt.

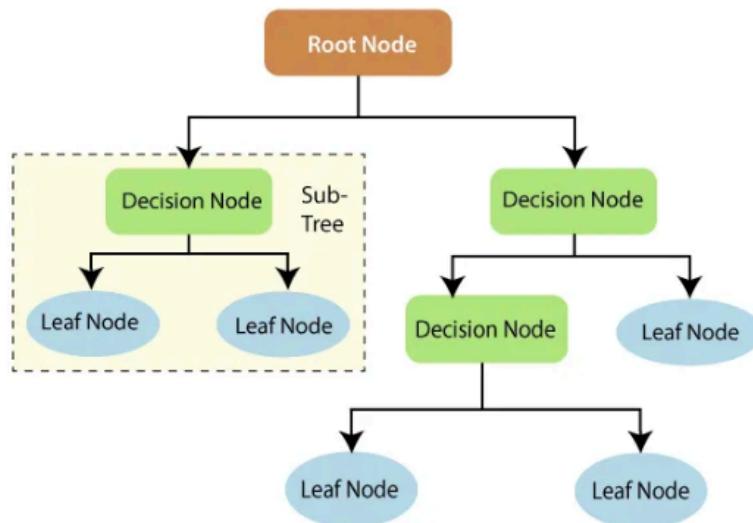
De applicatie die als proof of concept ontwikkeld wordt voor deze bachelorproef, zal de toekomstige stroomproductie van zonnepanelen gaan voorspellen. Hiervoor

zullen historische zonnestralingsgegevens gebruikt worden. Er zal dus gebruik gemaakt worden van de supervised learning methode en meer specifiek zal het regressiemodel toegepast worden. De waarde die moet voorspeld worden is de hoeveelheid zonnestraling uitgedrukt in Watt/m².

2.2.2. Supervised machine learning: regressie modellen

Decision Tree Regression (DTR)

Het Decision Tree model is een regressiealgoritme dat een dataset systematisch opdeelt in steeds kleinere homogene subgroepen op basis van de kenmerken van de gebruikte dataset. Het ontwikkelt daarbij een beslissingsboom, waarbij de interne knooppunten de kenmerken van de dataset vertegenwoordigen, de takken de beslissingsregels en elke bladknoop het resultaat. In de beslissingsboom zijn er twee knooppunten, namelijk de beslissingsknooppunten en de bladknooppunten. Beslissingsknooppunten worden gebruikt om een beslissing te nemen en hebben meerdere takken, terwijl Bladknooppunten de uitkomst van die beslissingen zijn en geen verdere takken bevatten. De beslissingen of de test worden uitgevoerd op basis van de kenmerken van de gegeven dataset. Het doel is om eenvoudige beslissingsregels te leren op basis van die specifieke kenmerken van de dataset om zo een model te ontwikkelen dat een bepaalde waarde voorspelt (Balakumar e.a., 2023).



Figuur (2.1)

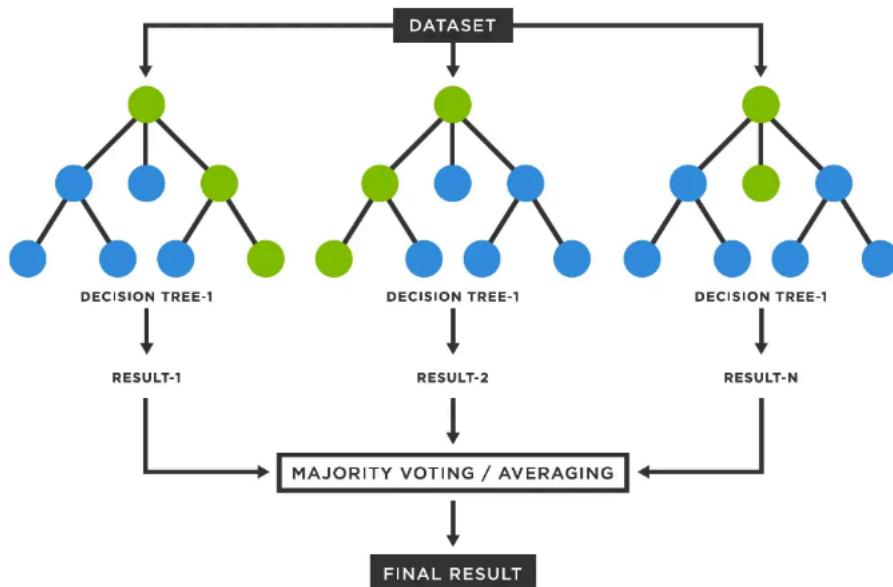
Grafische voorstelling beslissingsboom.

Decision trees zijn makkelijk te interpreteren omdat de geleerde beslissingsregels makkelijk te begrijpen en visualiseren zijn. Bovendien kan het DTR-algoritme niet-lineaire en dus meer complexe relaties tussen de inputvariabelen en de voorspelde

waarde ontdekken, wat het algoritme zeer bruikbaar maakt voor datasets die complexe patronen bevatten (Viswa, 2023).

Random Forest Regression (RFR)

Het Random Forest regressie model bestaat uit een verzameling van zogenaamde decision trees of beslissingsbomen. Een beslissingsboom is op zijn beurt een schema dat is opgebouwd uit een opeenvolging van binaire beslissingsregels. Het is een grafische voorstelling van een probleemstelling waarin verschillende mogelijke alternatieven met gebeurtenissen worden weergegeven en uitgewerkt. Het RFR-model maakt meerdere beslissingsbomen aan op basis van willekeurig gekozen subsets van de gebruikte dataset. Het model voegt vervolgens de uitkomsten van al deze beslissingsbomen samen om een algemene voorspelling te doen voor onbekende datapunten. Daarbij wordt telkens het gemiddelde of gewogen gemiddelde van elke beslissingsboom genomen. Op deze manier kan het grotere datasets verwerken en complexere verbanden vastleggen dan individuele beslissingsbomen. Het geheel van de voorspellingen van de beslissingsbomen zorgt voor een grotere accuraatheid dan de voorspelling van één enkele beslissingsboom. Over het algemeen kan gesteld worden dat hoe meer beslissingsbomen er in het RFR-model zitten, des te robuster het model zal zijn (Balakumar e.a., 2023).



Figuur (2.2)

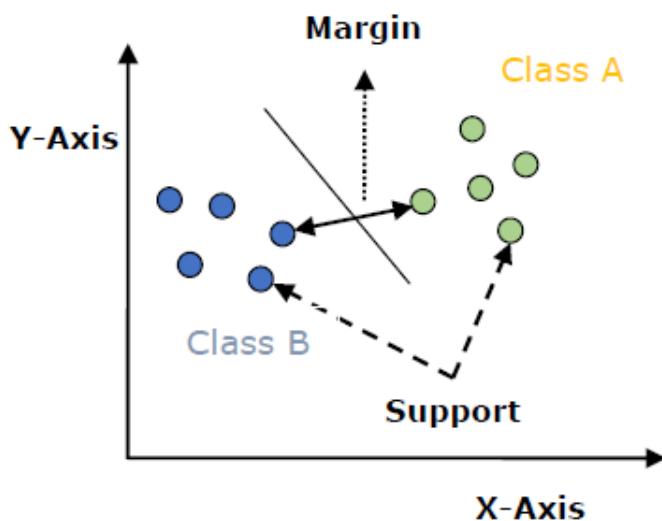
Grafische voorstelling van het Random Forest algoritme.

Het RFR-model wordt vaak gebruikt om continue waarden te voorspellen, zoals aandelenkoersen, tijdreeksen of verkoopprijzen. Het is minder gevoelig voor 'overfitting' (dit doet zich voor wanneer een ML-model de trainingsdata te goed leert, waardoor het nieuwe onbekende data slecht voorspelt) dan andere regressieme-

thoden omdat het meerdere willekeurige bomen onafhankelijk van elkaar opbouwt en een gemiddelde van de individuele voorspellingen neemt. Het is een goede keuze wanneer data gebruikt wordt met veel verschillende kenmerken of inputvariabelen (Sahai, 2023).

Support Vector Machine (SVM)

Het Support Vector Machine algoritme probeert data zo optimaal mogelijk in twee groepen te verdelen door het vinden van het beste hyperplane. Deze hyperplane is de meest optimale scheidingslijn tussen de twee groepen van data. Het algoritme maakt daarbij gebruik van support vectors. Dit zijn de datapunten die zich het dichtst bij de optimale scheidingslijn bevinden. Om de meest optimale scheidingslijn te achterhalen, berekent het algoritme de maximale marge of afstand tussen de datapunten van de twee groepen (Tziolis e.a., 2024). Omdat het algoritme zeer effectief is in het oplossen van complexere, niet-lineaire problemen, is het een model dat vaak gebruikt wordt om voorspellingen te maken op het gebied van hernieuwbare energie (Ahmad e.a., 2018).



Figuur (2.3)

Grafische voorstelling Support Vector Machine.

Het SVM-model wordt typisch gebruikt voor kleinere datasets omdat de trainigstijd van dit algoritme zeer lang kan zijn. Het wordt ook vooral toegepast op 'proper' datasets, dit zijn datasets die weinig afwijkingen of fouten bevatten.

Autoregressive Integrated Moving Average (ARIMA)

Het Autoregressive Integrated Moving Average (ARIMA) model is een populair voorspellingsmodel voor tijdreeksen. Het wordt veel gebruikt op verschillende gebieden om toekomstige waarden te analyseren en te voorspellen op basis van waar-

nemingen uit het verleden in een tijdreeksdataset.

Het ARIMA-model combineert drie belangrijke componenten: Autoregressie (AR), differentiëren (I) en voortschrijdend gemiddelde (MA).

AR - Autoregressief (p): Dit verwijst naar de afhankelijkheid van de huidige waarde in een tijdreeks van eerdere waarden. De volgorde van de AR-component, aangeduid met p, vertegenwoordigt het aantal vertraagde waarnemingen die in het model worden gebruikt.

I - Integratievevolgorde (d): Dit is het aantal verschillen dat we beschouwen tussen de huidige waarde en de waarde in het verleden om de tijdreeks stationair te maken. Stationair betekent dat de statistische eigenschappen van de reeks, zoals gemiddelde en variantie, constant blijven in de tijd.

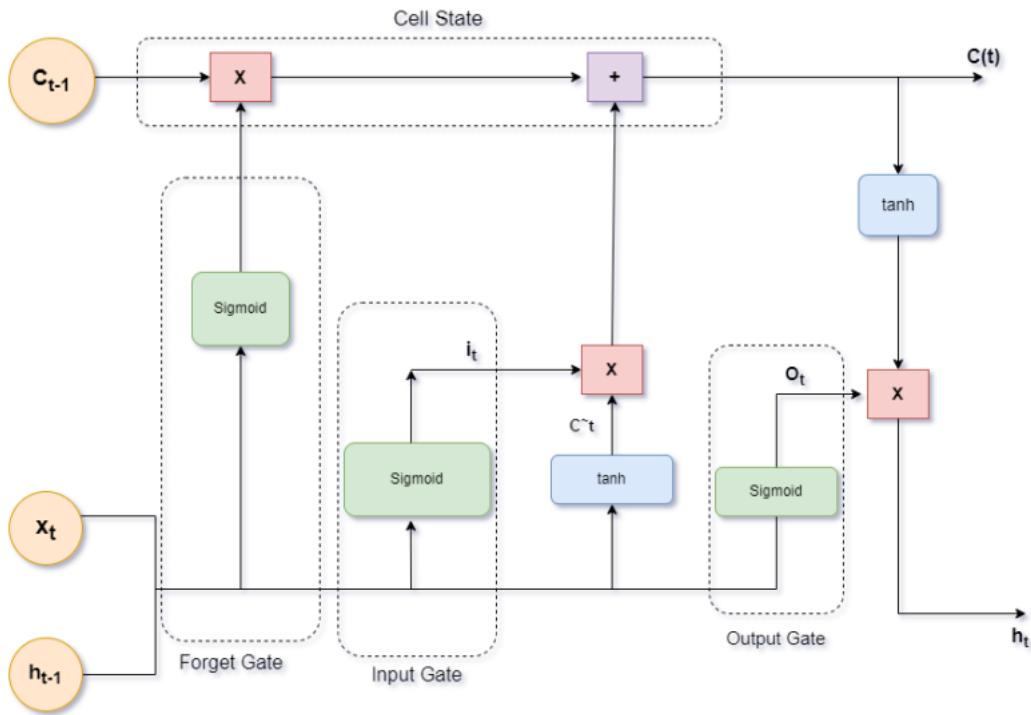
MA - Bewegend gemiddelde (q): Het vertegenwoordigt de afhankelijkheid tussen de huidige waarde en de restfouten van voorspellingen uit het verleden. Het berekent het gewogen gemiddelde van de vorige voorspellingsfouten. De volgorde van de MA-component, aangeduid met q, geeft het aantal vertraagde voorspellingsfouten weer die in het model worden gebruikt.

Long Short-Term Memory (LSTM)

Het Long Short-Term Memory (LSTM) is een type Recurrent Neural Network (RNN) dat wordt gebruikt om temporele afhankelijkheden in gegevens te modelleren. Concreet kan het zich gebeurtenissen uit het verleden of afhankelijkheden uit eerder waargenomen periodes herinneren.

Een LSTM is een type netwerk dat speciaal is ontworpen om een computer te helpen informatie langer te onthouden dan gewoonlijk mogelijk is met een terugkerend neuraal netwerk. Het lange-kortetermijngeheugenmodel werkt met behulp van 'cellen' die informatie gedurende langere tijd opslaan; wanneer een nieuwe invoer wordt gedetecteerd, worden er nieuwe cellen gemaakt die aan de bestaande cellen worden gekoppeld. Daarnaast kunnen de nieuwe cellen een deel van de informatie die in de bestaande cellen is opgeslagen 'vergeten', waardoor het netwerk een deel van wat het heeft geleerd kan 'vergeten'.

De belangrijkste componenten van de LSTM zijn de geheugencellen, vergeetpoorten en invoerpoorten. Geheugencellen zijn verantwoordelijk voor het langdurig opslaan van informatie, en de vergeet- en invoerpoorten beslissen welke informatie wel of niet in de cellen moet worden opgeslagen. Deze componenten zorgen ervoor dat LSTM-netwerken het verleden kunnen 'herinneren' en tegelijkertijd rekening kunnen houden met nieuwe input.

**Figuur (2.4)**

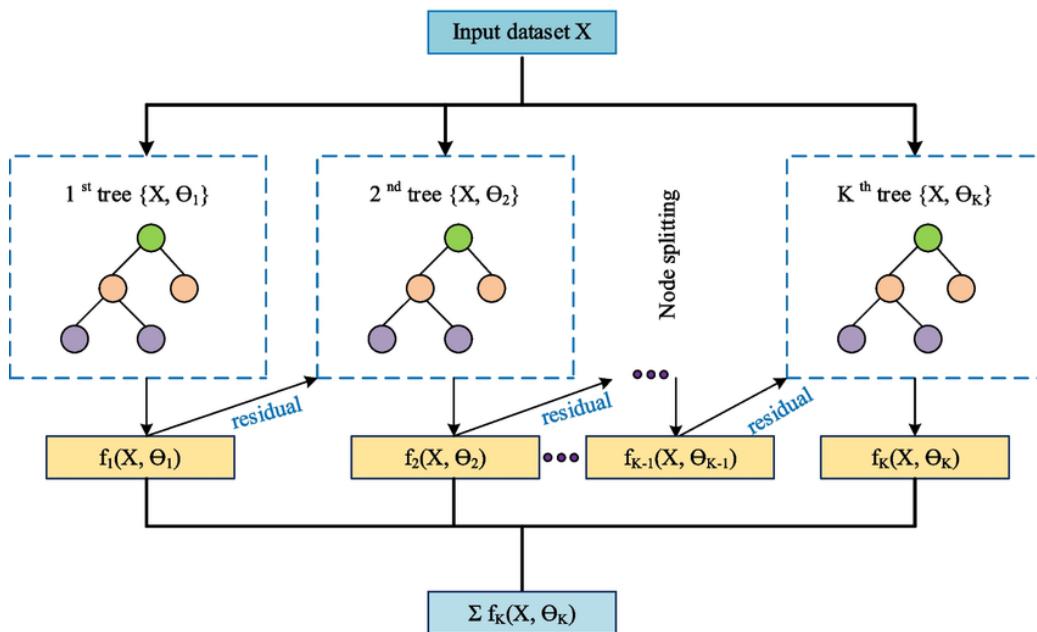
Grafische voorstelling van het LSTM algoritme.

LSTM's zijn zeer effectief vanwege hun vermogen om informatie gedurende langere tijd op te slaan en nieuwe input efficiënter te verwerken dan traditionele terugkerende neurale netwerken.

Extreme gradient Boosting (XGBoost)

Het XGBoost algoritme is een ensemble-leermethode die gradiëntversterking en beslissingsbomen combineert. Het concept van gradiëntversterking verwijst naar het opeenvolgend toevoegen van beslissingsbomen aan het model, waarbij elke volgende boom de fouten corrigeert die door de vorige zijn gemaakt. Dankzij dit iteratieve proces kan het XGBoost zijn voorspellingen voortdurend verbeteren en een hoge nauwkeurigheid bereiken. De werking van beslissingsbomen is hierboven reeds toegelicht.

Het XGBoost-model is specifiek ontworpen om de eigen prestaties te optimaliseren en grootschalige datasets efficiënt te verwerken. Door iteratief zwakke voor-

**Figuur (2.5)**

Grafische voorstelling van de structuur van het XGBoost algoritme.

spellingen toe te voegen, kan het XGBoost algoritme snelle en accurate voorspellingen maken. Het grote voordeel van XGBoost is dat het ingebouwde mechanismen heeft om ontbrekende waarden in de dataset te verwerken. Het kan tijdens het trainingsproces automatisch leren hoe er het beste met ontbrekende waarden kan worden omgegaan. Vermits de meeste datasets uit de echte wereld gegevens ontbreken, is het XGBoost algoritme dus zeer geschikt om voorspellingen te maken. Nog een voordeel van het XGBoost-model is dat het makkelijk schaalbaar is, omdat het gebruik maakt van parallelle verwerkingstechnieken waarbij de werklast over meerdere cores of machines verdeeld wordt. Dit maakt schnellere trainings- en voorspellingstijden mogelijk. Deze schaalbaarheid maakt XGBoost geschikt voor toepassingen die realtime voorspellingen vereisen.

2.2.3. Machine learning modellen evalueren

Er bestaan verschillende methodes waarmee de kwaliteit en nauwkeurigheid van een machine learning model kan gemeten worden. De meest gebruikte zijn:

Mean absolute error (MAE)

De Mean Absolute Error (MAE) of gemiddelde absolute fout is het gemiddelde van alle absolute voorspellingsfouten waarbij de voorspellingsfout het verschil is tussen de werkelijke en de voorspelde waarde. Het wordt berekend door de absolute verschillen tussen voorspelde waarden en werkelijke waarden bij elkaar op te tellen en te delen door het aantal waarnemingen. Door de absolute waarde van voorspellingsfouten te gebruiken wordt voorkomen dat positieve en negatieve fouten elkaar opheffen. Hoe kleiner de waarde van de MAE, des te beter de voorspellingen.

van een model overeenstemmen met de werkelijke waarden.

Mean Squared Error (MSE)

Mean Squared Error (MSE) of gemiddelde kwadratische fout is de gemiddelde kwadratische afwijking tussen de voorspelde waarden van een model en de werkelijke waarden. Een lagere MSE geeft aan dat de voorspellingen van een model dichter bij de werkelijkheid liggen en het model dus nauwkeuriger is.

Root mean squared error (RMSE)

Root Mean Squared Error (RMSE) of vierkantswortel van de gemiddelde kwadratische fout is de gemiddelde omvang van de positieve of negatieve verschillen tussen de waarden die door een model voorspeld zijn en de geobserveerde (feitelijke) waarden. Het kan beschouwd worden als de standaardafwijking van de voorspellingsfouten.

RMSE biedt één enkele waarde die de prestaties van een model vertegenwoordigt, waardoor een eenvoudige vergelijking tussen verschillende modellen of voorspellingstechnieken mogelijk is. Lagere RMSE-waarden duiden op betere modelprestaties. Een lagere RMSE geeft aan dat de voorspellingen van het model dichter bij de werkelijke waarden liggen, wat een hogere nauwkeurigheid suggereert.

R-squared (R^2)

De R-squared (R^2) of determinatiecoëfficiënt meet in hoeverre een model in staat is een bepaalde uitkomst te voorspellen. Het geeft aan hoe sterk het verband is tussen de werkelijke waarden en de voorspelde waarden van een model. R^2 is dan ook nauw verwant aan de correlatiecoëfficiënt. Het kwadraat van de correlatiecoëfficiënt tussen de onafhankelijke en afhankelijke variabelen is gelijk aan de R-kwadraatwaarde. Dit impliceert dat een hogere correlatie overeenkomt met een hogere R-kwadraatwaarde.

De laagst mogelijke waarde van R^2 is 0 en de hoogst mogelijke waarde is 1. Hoe beter een model is in het maken van voorspellingen, hoe dichter de determinatiecoëfficiënt R^2 bij het getal 1 zal liggen.

2.3. Elektriciteitsproductie voorspellen

Doordat het belang van hernieuwbare energie de laatste jaren enorm is toegenomen, is er al heel wat onderzoek verricht naar het voorspellen van de toekomstige stroomproductie van zonnepanelen met behulp van machine learning. In de meeste van deze onderzoeken wordt deze voorspelling gebaseerd op de voorspelling van de hoeveelheid zonnestraling. Een andere optie bestaat erin om de historische stroomproductie van de zonnepanelen te gaan analyseren en daaruit

een voorspelling te gaan maken (Wang e.a., 2022). Deze historische stroomproductie data is evenwel niet altijd voldoende vorhanden, zodat het soms niet mogelijk is om op basis daarvan voorspellingen te gaan doen.

2.3.1. Zonnestraling

Vermits zonnepanelen zonne-energie omzetten in elektriciteit, is het bijna een logische keuze om de toekomstige stroomproductie van zonnepanelen te gaan voorspellen op basis van de voorspelling van de hoeveelheid zonnestraling (Ledmaoui e.a., 2023). Meestal wordt daarbij de inkomende of globale horizontale instraling (GHI, Global Horizontal Irradiance) voorspeld. GHI is de totale hoeveelheid zonnestraling die het aardoppervlak bereikt en wordt uitgedrukt in W/m². De GHI bestaat uit directe normale instraling (DNI, Direct Normal Irradiance) en de diffuse horizontale instraling (DHI, Diffuse Horizontal Irradiance). DNI verwijst naar de grootste directe (90 graden) neerwaartse zonnestraling voor een bepaalde plaats. Deze directe normale instraling wordt gebruikt om de diffuse horizontale instraling te berekenen. De DHI is de zonnestraling die niet rechstreeks van de zon komt, maar verspreid is door de wolken en deeltjes in de atmosfeer. Deze zonnestraling komt in gelijke mate uit alle richtingen (Sehrawat e.a., 2023).

Historische data met betrekking tot zonnestraling kan vrij verkregen worden via de CAMS Radiation Service (CRS) van de Copernicus Atmosphere Monitoring Service ([CAMS](#)). Dit maakt deel uit van het European Earth observation programme Copernicus (EEC) dat informatiediensten biedt op basis van aardobservatiegegevens van satellieten en in-situ gegevens (niet uit de ruimte). Enorme hoeveelheden wereldwijde gegevens van satellieten en van meetsystemen op de grond, in de lucht en op zee worden gebruikt om informatie te verstrekken aan dienstverleners, overhedsinstanties, andere internationale organisaties en burgers. De aangeboden informatiediensten zijn gratis en open toegankelijk voor de gebruikers ervan. De geografische dekking van de gegevens is het gebied dat door de Meteosat (een serie geosynchrone weersatellieten van de Europese ruimtevaartorganisaties ESA en EUMETSAT) kan worden waargenomen. Dit gebied omvat Europa, Afrika en het Midden-Oosten.

2.3.2. Meteorologische data

De hoeveelheid stroom die zonnepanelen produceren is niet enkel afhankelijk van de hoeveelheid zonnestraling. Ook andere meteorologische gegevens spelen hierbij een rol. Zo zijn de temperatuur, de relatieve luchtvuchtigheid en de bewolkingsgraad factoren die het meeste invloed hebben op de stroomproductie van zonnepanelen (Sehrawat e.a., 2023).

Er zijn verschillende kanalen waarlangs historische weerdata kan verkregen wor-

den. Zo zijn er datasets van meteorologische gegevens die door het Koninklijk Meteorologisch Instituut (KMI) beschikbaar gesteld worden. Het aantal datasets is echter zeer beperkt en moeten handmatig gedownload worden. Ook via het CAMS kan historische weerdata opgevraagd worden, maar ook deze data is beperkt aangezien de meest recente weerdata van het jaar 2016 is.

Er bestaan tal van open source en dus gratis weer API's die makkelijk geïntegreerd kunnen worden met een app. Deze API's bieden voorspellingen aan van verschillend meteorologische gegevens, maar leveren vaak beperkte historische weerdata aan. Voor de historische weerdata werd [Meteostat](#) gekozen. Deze open-source API biedt een Python library waarmee met slechts een enkele HTTP-request historische weerdata voor een bepaalde locatie kan worden opgevraagd. Meteostat verzamelt historische weer- en klimaatdata van weerstations en verschillende nationale meteorologische instituten van over de hele wereld. De historische weerdata zal voor dezelfde periode als de zonnestralingsdata opgevraagd worden en op die manier de zonnestralingsdata verrijken, zodat de voorspelling van de toekomstige zonnestaling en stroomproductie van de zonnepanelen accurater wordt.

De voorspelling van de toekomstige elektriciteitsproductie van de zonnepanelen zal tenslotte nog proberen verbeterd worden door de voorspelde hoeveelheid zonnestraling te combineren met de weersvoorspellingen voor periode die voorspeld wordt. Omdat Meteostat enkel historische weerdata aanbiedt, werd voor de weersvoorspellingen een andere open-source weer-API gebruikt. [Open-Meteo API](#) werkt ook samen met verschillende nationale meteorologische diensten en biedt betrouwbare weersvoorspellingen aan tot 16 dagen in de toekomst. De aangeboden API-request kan makkelijk aangepast worden, zodat enkel de benodigde weerdata kan worden opgevraagd, wat een gunstig effect heeft op de performantie bij het oproepen van de data.

3

Methodologie

Supervised machine learning en meer specifiek regressie. (meest gebruikte machine learning techniek)

Multiple linear regression? Model dat gebruikt wordt wanneer meer dan 2 variabelen: historische productie + weersvoorspelling weer API

To do:

- Selecteer de beste ML-technieken om voorspellingen te maken
- Bepaal de accuraatheid van het regressiemodel (zie colab HoGent + rapport PXL)
- Kies model met beste score

Eerst zal de verzamelde data moeten gecleaned, getransformeerd en geanalyseerd worden.

Ensemble regressors en meer bepaald gradient boosting lijken het beste model te zijn want ANMAE (Absolute Normalized Mean Absolute Error) en PRMSE (??) laagste (Khasawneh e.a., [2024](#)) en (Tercha e.a., [2024](#)).

Korte termijn voorspelling elektriciteitsverbruik: ook daar eXtreme Gradient Boosting het beste ML algoritme (Irankhah e.a., [2024](#)).

Vaatwas en wasmachine verbruiken gemiddeld 1000 W per uur dus vanaf dan mogen deze toestellen ingeschakeld worden. Toestellen ingeven in app en daarachter zit bepaald wattage en vanf dat wattage zal het ingeschakeld worden.

4

Proof of Concept

4.1. Ontwikkeling van de mobiele app

Om de onderzoeksvraag van deze bachelorproef te kunnen beantwoorden werden een Python web server en een iOS app ontwikkeld als proof of concept. Deze app maakt in de eerste plaats het elektriciteitsverbruik en de stroomproductie van de zonnepanelen inzichtelijk. Daarnaast zal de app via de toepassing van machine learning de elektriciteitsproductie van de volgende dag voorspellen. Deze voor-spelling zal dan door de app gebruikt worden om de inschakeling van twee slimme stekkers te programmeren indien er voldoende stroomproductie voorspeld wordt.

4.1.1. Digitale elektriciteitsmeter uitlezen

Een mobiele app die het elektriciteitsverbruik wil bijsturen en beter spreiden zal de gebruiker ervan allereerst inzicht moeten geven in zijn of haar verbruik. Om de verbruiksgegevens te ontsluiten moet de digitale elektriciteitsmeter uitgelezen worden via de P1-poort. Hiervoor zijn verschillende mogelijkheden, van een dongle tot kleine toestellen die via een kabel met de P1-poort kunnen verbonden worden. Voor dit onderzoek is er gekozen voor een Raspberry Pi 5 minicomputer om de digitale meter uit te lezen. Deze singleboardcomputer die gebaseerd is op een ARM processor en de Linuxdistributie Ubuntu als besturingssysteem heeft, biedt meer mogelijkheden dan enkel het uitlezen van data. Omdat het verbruik ervan zeer laag is, kan dit toestel bovendien continu ingeschakeld blijven. Daarnaast heeft de Raspberry Pi wifi-ondersteuning zodat deze kan verbonden worden met het wifi-netwerk van de woning die als testomgeving gebruikt wordt. Dit maakt het mogelijk om een webserver op te zetten en meteen ook de omvormer van de zonnepanelen van de woning via het wifi-netwerk uit te lezen met de Raspberry Pi (zie volgende sectie). In de testwoning bevindt de digitale elektriciteitsmeter zich namelijk in de kelder, terwijl de omvormer van de zonnepanelen zich op zolder be-

vindt. Een situatie die wel vaker voorkomt en op deze manier kan worden opgelost.



Figuur (4.1)

Testopstelling: Raspberry Pi 5 via een RS422 serieel naar USB kabel aangesloten op een digitale elektriciteitsmeter.

Opslag uitgelezen elektriciteitsdata

Nadat de Raspberry Pi via de juiste kabel op de P1-poort werd aangesloten, kon de data van de digitale elektriciteitsmeter via een Python script worden uitgelezen. Hiervoor werd een bestaand Python script uitgebreid (Depuydt, 2021). Het bestaande script voorzag enkel in het uitprinten van de uitgelezen data in de console. Om de uitgelezen elektriciteitsdata later via een app te kunnen weergeven moest deze data worden weggeschreven naar een databank. Omdat de data via de P1-poort per seconde uitgelezen wordt, werd geopteerd voor InfluxDB. Deze NoSQL-databank is speciaal ontwikkeld voor time-series en is binnen bepaalde (ruime) grenzen gratis te gebruiken (Balis e.a., 2017) en (Struckov e.a., 2019). Het Python script werd tenslotte opgezet als een achtergrond service via de systeem manager voor Linux 'systemctl/systemd', zodat het voortdurend blijft uitgevoerd worden.

4.1.2. Omvormer zonnepanelen uitlezen

Zoals reeds vermeld bevindt de omvormer van de zonnepanelen zich op de zolderverdieping, terwijl de Raspberry Pi zich in de kelder bij de digitale elektriciteitsme-

ter bevindt. De enige mogelijkheid om een draadloze verbinding te maken tussen beide toestellen is via het lokale wifi-netwerk. De omvormer in casu heeft echter geen ingebouwde wifi-ondersteuning en diende uitgebreid te worden met een aparte wifi-stick. Via deze wifi-stick kon dan vervolgens de data van de omvormer uitgelezen worden. Dit gebeurt opnieuw via een Python script waarin een request wordt gestuurd naar de server van de producent waar de gegevens van de omvormer worden bijgehouden. Omdat de data van de digitale elektriciteitsmeter per seconde wordt uitgelezen, wordt ook de data van de omvormer per seconde opgevraagd. De ontvangen gegevens zijn de temperatuur, de huidige geproduceerde stroom en de totale dagproductie van de zonnepanelen.

**Figuur (4.2)**

Omvormer van de zonnepanelen met een wifi-stick.

Ook de data van de omvormer worden via het Python script weggeschreven naar een aparte InfluxDB databank en opgezet als een achtergrond service via de systeem manager voor Linux 'systemctl/systemd'. Zo blijft ook dit script automatisch uitgevoerd worden.

4.1.3. Stroomproductie van zonnepanelen voorspellen met XGBoost

Het oorspronkelijke idee was om de stroomproductie van de zonnepanelen te gaan voorspellen op basis van de historische data van die stroomproductie. Voor dit onderzoek was er echter geen of slechts beperkte historische data vorhanden, omdat de metingen van de omvormer van de zonnepanelen pas gestart zijn bij het begin van het onderzoek. Na uitgebreid literatuuronderzoek werd beslist om de

stroomproductie van de zonnepanelen te gaan voorspellen op basis van de hoeveelheid zonnestraling (Sehrawat e.a., 2023), (Ledmaoui e.a., 2023), (Wang e.a., 2022) en (Sansine e.a., 2022). Via de CAMS Radiation Service (CRS) van de Copernicus Atmosphere Monitoring Service ([CAMS](#)) kan voldoende betrouwbare historische data met betrekking tot zonnestraling verkregen worden.

Historische zonnestralingsdata verzamelen

De zonnestralingsdata die via de CAMS Radiation Service (CRS) kan bekomen worden omvat tijdreeksen van globale, directe en diffuse insstraling voor een tijdsperiode van 1 februari 2004 tot en met 2 dagen geleden. De granulariteit van de data varieert van 1 maand tot 1 minuut en kan via een API-request opgevraagd worden.

Er bestaat een open-source Python library 'pvlib' (Jensen e.a., 2023) die ontwikkeld werd om de opbrengst van zonnepanelen te simuleren. Deze library bevat ook tools om zonnestralingsdata op te vragen, waaronder de data van de CAMS Radiation Service (CRS). Omdat de reeds ontwikkelde scripts voor het uitlezen van de digitale elektriciteitsmeter en de omvormer van de zonnepanelen in Python geschreven zijn, worden ook de andere scripts in Python geschreven en kan de pvlib-library dus gebruikt worden. Om de zonnestralingsgegevens op te vragen moeten een aantal parameters worden meegegeven, waaronder de locatie (lengte- en breedtegraad), de start- en einddatum en de tijdsgranulariteit. Omdat van de gebruiker van de app niet kan verwacht worden dat hij of zij de geografische coördinaten van zijn of haar woning kent, wordt de Python library Geopy gebruikt om het ingevoerde adres van een gebruiker om te zetten in de correcte lengte- en breedtegraad coördinaten en deze vervolgens mee te geven in de API-call naar de CRS.

Voor dit onderzoek wordt de zonnestralingsdata voor een periode van iets meer dan 5 jaar opgevraagd, van 1 januari 2019 tot op heden om precies te zijn. Dat heden is evenwel de datum van 2 dagen eerder, aangezien het 2 dagen duurt vooraleer de CRS de waargenomen zonnestralingsdata verwerkt en opgeslagen heeft. Dit geeft voldoende data om via een machine learning algoritme de toekomstige hoeveelheid zonnestraling te gaan voorspellen. Voor de granulariteit werd geopteerd voor metingen om de 15 minuten. Zo zijn de metingen van de bekomen dataset voldoende gedetailleerd om accurate voorspellingen te kunnen maken, zonder dat de performantie in het gedrang komt. Bij metingen van 1 minuut blijkt de dataset te omvangrijk om er op een efficiënte en dus snelle manier berekeningen op uit te voeren.

```
11 # Get solar irradiance data for specified address
12 # User input
13 user_address = 'Toekomststraat 67, 9040 Sint-Amandsberg'
14
15 # Determine coordinates of entered address
16 loc = Nominatim(user_agent="Geopy Library")
17 getLoc = loc.geocode(user_address, timeout=None)
18 latitude = getLoc.latitude
19 longitude = getLoc.longitude
20
21 # past_date = (datetime.now()-timedelta(365*5)).date()
22 past_date = '2019-1-1'
23 current_date = datetime.now().date()
24
25 # Get solar irradiance data from Copernicus Atmosphere Monitoring Service (CAMS)
26 data = pvlib.iotools.get_cams(latitude,
27                                longitude,
28                                start=past_date,
29                                end=current_date,
30                                email='jan.tubeeckx@hotmail.com',
31                                identifier='cams_radiation',
32                                time_step='15min',
33                                time_ref='UT',
34                                verbose=False,
35                                integrated=False,
36                                label=None,
37                                map_variables=True,
38                                server='api.soda-solaridata.com',
39                                timeout=30)
```

Figuur (4.3)

Script voor het opvragen van de zonnestralingsdata.

Weerdata

Niet alleen de hoeveelheid zonnestraling is bepalend voor de stroomproductie van zonnepanelen, ook de weersomstandigheden oefenen een invloed uit. Vooral de omgevingstemperatuur, de relatieve luchtvochtigheid en de bewolkingsgraad zijn factoren die de stroomproductie van zonnepanelen kunnen vergroten of verkleinen (Sehrawat e.a., 2023).

Om de voorspelling van de toekomstige stroomproductie van de zonnepanelen accurater te maken, zal de zonnestralingsdata in de eerste plaats gecombineerd worden met historische weerdata voor dezelfde periode, van 1 januari 2019 tot op heden. Ook hiervoor werd gezocht naar een open-source API, waarmee de data via een Python script kan worden opgevraagd. Finaal is voor Meteostat gekozen. Deze API biedt een Python library waarmee met slechts een enkele HTTP-request historische weerdata voor een bepaalde locatie kan worden opgevraagd. Meteostat verzamelt historische weer- en klimaatdata van weerstations en verschillende nationale meteorologische instituten van over de hele wereld. Omdat ook voor deze API-call de

lengte- en breedtegraad van de gevraagde locatie moet worden ingegeven, wordt opnieuw gebruik gemaakt van de Python library Geopy waarmee een adres kan worden omgezet in de correcte lengte- en breedgraad coördinaten.

```

4   import openmeteo_requests
5   from datetime import datetime
6   from retry_requests import retry
7   from geopy.geocoders import Nominatim
8   from meteostat import Point, Hourly
9
10  # User input
11  user_address = "Toekomststraat 67, 9040 Sint-Amandsberg"
12  gradient = 45
13  orientation = -135
14
15  # Determine coordinates of entered address
16  loc = Nominatim(user_agent="Geopy Library")
17  getLoc = loc.geocode(user_address, timeout=None)
18  latitude = getLoc.latitude
19  longitude = getLoc.longitude
20
21  # Set time period
22  current_date = datetime.now().date()
23  year = current_date.year
24  month = current_date.month
25  day = current_date.day
26  start = datetime(2019, 1, 1)
27  end = datetime(year, month, day)
28
29  location = Point(latitude, longitude)
30
31  # Get daily data for 2018
32  data = Hourly(location, start, end)
33  weather_data = data.fetch()
34  weather_data.drop(columns=['snow', 'wdir', 'wspd', 'wpgr', 'tsun', 'coco'], inplace=True)
35  weather_data.rename(columns={'temp':'temperatuur', 'dwpt':'dauwpunt', 'rhum':'luchtvochtigheid',
36  | | | | | 'prcp':'neerslag', 'pres':'luchtdruk'}, inplace=True)
37  weather_data.index = pd.to_datetime(weather_data.index)
```

Figuur (4.4)

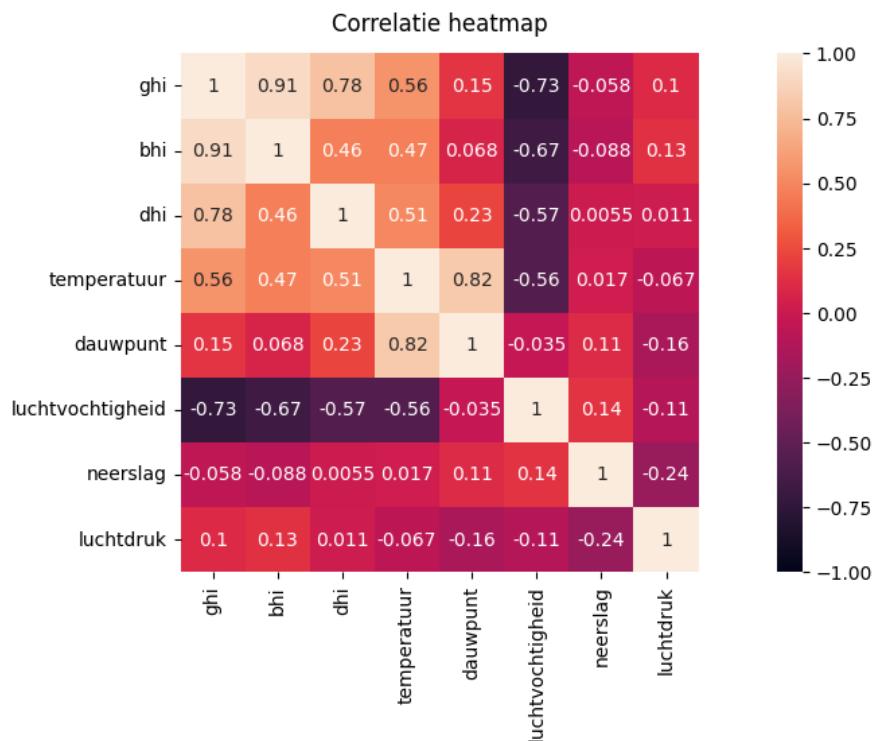
Script voor het opvragen van de weerdata.

Om de voorspelling van de stroomproductie van de zonnepanelen nog beter te maken, zal de bekomen voorspelling gecombineerd worden met de weersvoorspellingen voor de voorspelde periode. Aangezien Meteostat enkel historische data aanbiedt, moest een andere open-source weer-API gevonden worden. [Open-Meteo API](#) werkt ook samen met verschillende nationale meteorologische diensten waardoor het betrouwbare weersvoorspellingen aanbiedt. De aangeboden API-request is makkelijk in te stellen, zodat enkel de benodigde weerdata kan worden opgevraagd. Zo kan de data snel opgevraagd worden. Voor dit onderzoek worden enkel de voorspellingen van de omgevingstemperatuur, de relatieve luchtvochtigheid en de bewolkingsgraad opgehaald.

Voorspelling met XGBoost

Zeker heatmap met correlaties invoegen om zo de belangrijkste kenmerken, dit zijn de kenmerken met de grootste negatieve of positieve correlatie.

Om de voorspellingen zo accuraat mogelijk te maken, zal gebruik gemaakt worden van het Extreme Gradient Boosting (XGBOOST) algoritme (Ledmaoui e.a., 2023), (Wang e.a., 2022) en (Barrera-Animas e.a., 2022). De historische verbruiksdata, de productiedata van de omvormer van de zonnepanelen en de weerdata zullen als input gebruikt worden voor dit algoritme. De data van de digitale elektriciteitsmeter en de omvormer van de zonnepanelen wordt continue uitgelezen met een python script op een Raspberry Pi computer die via het wifi-netwerk verbonden is. Deze data wordt opgeslagen in een NoSQL-databank die speciaal voor time-series ontwikkeld is, nl. InfluxDB (Balis e.a., 2017) en (Struckov e.a., 2019). De app zal tenslotte ontwikkeld worden voor iOS en gebouwd worden met Xcode, SwiftUI en UIKit (**Allardice**) en (**Firtman2022**).



Figuur (4.5)

Heatmap van correlaties tussen de verschillende kenmerken van de dataset.

```

47  ## Apply feature engineering
48  # Create time series features based on time series index
49  def create_features(df):
50      df = df.copy()
51      df['hour'] = df.index.hour
52      df['dayofweek'] = df.index.dayofweek
53      df['quarter'] = df.index.quarter
54      df['month'] = df.index.month
55      df['year'] = df.index.year
56      df['dayofyear'] = df.index.dayofyear
57      df['dayofmonth'] = df.index.day
58      return df
59
60 historical_data = create_features(historical_data)
61
62 # Create lag features
63 def add_lag_features(df):
64     target_map = df[solar_irradiation].to_dict()
65     df['lag_1'] = (df.index - pd.Timedelta('72 hours')).map(target_map)
66     # df['lag_2'] = (df.index - pd.Timedelta('48 hours')).map(target_map)
67     # df['lag_3'] = (df.index - pd.Timedelta('7 days')).map(target_map)
68     # df['lag_4'] = (df.index - pd.Timedelta('14 days')).map(target_map)
69     return df
70
71 historical_data = add_lag_features(historical_data)

```

Figuur (4.6)

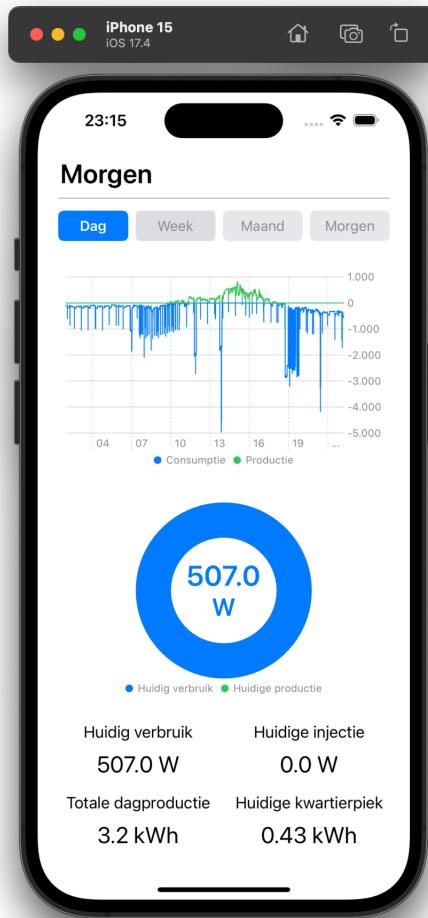
Screenshot van de Python code voor het aanmaken van features en lags.

4.1.4. Weergave uitgelezen data en voorspelling met een iOS app

4.1.5. Aansturing slimme stekkers op basis van voorspelde stroomproductie

De eerste slimme toestellen die via de app zullen worden beheerd zijn zonnepanelen en een warmtepomp (**Uytterhoeven2019**). Door de afschaffing van de virtueel terugdraaiende teller voor eigenaars van zonnepanelen, waarbij de teller van de elektriciteitsmeter terugdraait wanneer meer elektriciteit wordt opgewekt dan verbruikt, kan het verlies van dit voordeel opgevangen door het verbruik van de warmtepomp te laten samenvallen met de productiemomenten van de zonnepanelen (Selleslagh, 2021). Wanneer echter uit de weersverwachtingen blijkt dat er een aanzienlijke elektriciteitsproductie zal zijn, zal de app mede op basis van historiek van het elektriciteitsverbruik automatisch ook andere toestellen, zoals de vaatwas of wasmachine gaan inschakelen.

Om de sanitaire toestellen te kunnen gaan inschakelen, zal in de eerste plaats gekeken worden of deze toestellen van zichzelf reeds slim zijn. Meer concreet zal geverifieerd worden of ze over een Soft Real Time Operating System beschikken (Soft RTOS) en via het wifinetwerk kunnen communiceren met een app. Voor de sanitaire toestellen die niet slim zijn, zal gebruik gemaakt worden van slimme stopcontacten. Daarbij wordt er tussen het klassieke stopcontact en de stekker van het

**Figuur (4.7)**

Screenshot dagweergave.

toestel een apparaat geplaatst, waardoor een toestel op een eenvoudige manier slim kan gemaakt worden (**Jong2020**).

5

Conclusie

A

Onderzoeksvoorstel

Het onderwerp van deze bachelorproef is gebaseerd op een onderzoeksvoorstel dat vooraf werd beoordeeld door de promotor. Dat voorstel is opgenomen in deze bijlage.

Bibliografie

- Ahmad, M. W., Reynolds, J., & Rezgui, Y. (2018). Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. *Journal of Cleaner Production*, 203, 810–821. <https://doi.org/https://doi.org/10.1016/j.jclepro.2018.08.207>
- Balakumar, P., Vinopraba, T., & Chandrasekaran, K. (2023). Machine learning based demand response scheme for IoT enabled PVintegrated smart building. *Sustainable Cities and Society*, 89. <https://doi.org/https://doi.org/10.1016/j.scs.2022.104260>
- Balis, B., Bubak, M., Harezlak, D., Piotr, N., Pawlik, M., & Wilk, B. (2017). Towards an operational database for real-time environmental monitoring and early warning systems. *Procedia Computer Science*, 108, 2250–2259. <https://doi.org/https://doi.org/10.1016/j.procs.2017.05.193>
- Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinoshio, T. D., Delgado, J. M. D., & Adewale Akanbi, L. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Machine Learning with Applications*, 7(100204). <https://doi.org/https://doi.org/10.1016/j.mlwa.2021.100204>
- Brownlee, J. (2023, oktober 23). *Supervised and Unsupervised Machine Learning Algorithms*. Machine Learning Mastery. Verkregen april 30, 2024, van <https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/>
- Deman, K. (2021, augustus 4). Hoe maak je een digitale meter slim? Mijnenergie. Verkregen mei 15, 2023, van <https://www.mijnenergie.be/blog/hoe-maak-je-een-digitale-meter-slim/#:~:text=De%20website%20www.maakjemeterslim.be,real%20time%20op%20te%20volgen.>
- Depuydt, J. (2021, februari 4). Data lezen van de P1 poort op de Belgische digitale elektriciteitsmeter. Verkregen mei 21, 2023, van <https://jensd.be/1205/linux/data-lezen-van-de-belgische-digitale-meter-met-de-p1-poort>
- Efimov, V. (2024, april 9). *Reinforcement Learning, Part 1: Introduction and Main Concepts: Making the first step into the world of reinforcement learning*- Vyacheslav EfimovTowards Data Science. Towards Data Science. Verkregen april 30, 2024, van <https://towardsdatascience.com/reinforcement-learning-introduction-and-main-concepts-48ea997c850c>
- Irankhah, A., Yaghmaee, M. H., & Ershadi-Nasab, S. (2024). Optimized short-term load forecasting in residential buildingsbased on deep learning methods for

- different time horizons. *Journal of Building Engineering*. <https://doi.org/https://doi.org/10.1016/j.jobe.2024.108505>
- Jensen, A. R., Anderson, K. S., Holmgren, W. F., Mikofski, M. A., Hansen, C. W., Boeman, L. J., & Roel, L. (2023). pvlib iotools—Open-source Python functions for seamless access to solar irradiance data. *Solar Energy*, 266(112092). <https://doi.org/https://doi.org/10.1016/j.solener.2023.112092>
- Khasawneh, H. J., Ghazal, Z. A., Al-Khatib, W. M., Al-Hadi, A. M., & Arabiyat, Z. M. (2024). Creating optimized machine learning pipelines for PV power generation forecasting using the grid search and tree-based pipeline optimization tool. *Cogent Engineering*, 11(1). <https://doi.org/https://doi.org/10.1080/23311916.2024.2323818>
- Ledmaoui, Y., El Maghraoui, A., El Aroussi, M., Saadane, R., Chebak, A., & Chehri, A. (2023). Forecasting solar energy production: A comparative study of machine learning algorithms. *Energy Reports*, 10, 1004–1012. <https://doi.org/https://doi.org/10.1016/j.egyr.2023.07.042>
- Mack, B., Tampe-Mai, K., Kouros, J., Roth, F., Taube, O., & Diesch, E. (2019). Bridging the electricity saving intention-behavior gap: A German field experiment with a smart meter website. *Energy Research & Social Science*, 53, 34–46. <https://doi.org/https://doi.org/10.1016/j.erss.2019.01.024>
- Sahai, N. (2023, september 6). *Random Forest Regression – How it Helps in Predictive Analytics?* AnalyticsLabs. Verkregen april 30, 2024, van <https://www.analytixlabs.co.in/blog/random-forest-regression/>
- Sansine, V., Ortega, P., Hissel, D., & Hopuare, M. (2022). Solar Irradiance Probabilistic Forecasting Using Machine Learning, Metaheuristic Models and Numerical Weather Predictions. *Sustainability*, 14(22). <https://doi.org/https://doi.org/10.3390/su142215260>
- Sehrawat, N., Vashisht, S., & Singh, A. (2023). Solar irradiance forecasting models using machine learning techniques and digital twin: A case study with comparison. *International Journal of Intelligent Networks*, 4, 90–102. <https://doi.org/https://doi.org/10.1016/j.ijin.2023.04.001>
- Selleslagh, D. (2021, april 13). *Zo haalt u toch voordeel uit combinatie van warmtepomp en zonnepanelen.* Netto, De Tijd. Verkregen mei 25, 2023, van <https://www.tijd.be/netto/analyse/vastgoed/zo-haalt-u-toch-voordeel-uit-combinatie-van-warmtepomp-en-zonnepanelen/10297339.html>
- Selleslagh, D. (2022, december 15). *Wat betekent het nieuwe piektarief voor uw stroomfactuur?* Netto, De Tijd. Verkregen mei 15, 2023, van <https://www.tijd.be/netto/dossiers/uw-geld-in-2023/wat-betekent-het-nieuwe-piektarief-voor-uw-stroomfactuur/10434556.html>
- Struckov, A., Yufa, S., Visheratin, A. A., & Nasonov, D. (2019). Evaluation of modern tools and techniques for storing time-series data. *Procedia Computer Science*, 156. <https://doi.org/https://doi.org/10.1016/j.procs.2019.08.125>

- Tercha, W., Tadjer, S. A., Chekired, F., & Canale, L. (2024). Machine Learning-Based Forecasting of Temperature and Solar Irradiance for Photovoltaic Systems. *Energies*, 17. <https://doi.org/https://doi.org/10.3390/en17051124>
- Tziolis, G., Lopez-Lorente, J., Baka, M.-I., Koumis, A., Livera, A., Theocharides, S., Makrides, G., & Georghiou, G. E. (2024). Direct short-term net load forecasting in renewable integrated microgrids using machine learning: A comparative assessment. *Sustainable Energy, Grids and Networks*, 37(101256). <https://doi.org/https://doi.org/10.1016/j.segan.2023.101256>
- Verdoodt, B. (2022, oktober). *Investeringsplan Fluvius 2023-2032* (tech. rap.). Fluvius. <https://partner.fluvius.be/sites/fluvius/files/2022-06/investeringsplan-2023-2032.pdf>
- Viswa. (2023, juli 31). *Unveiling Decision Tree Regression: Exploring its Principles, Implementation*. Medium. Verkregen april 30, 2024, van <https://medium.com/@vk.viswa/unveiling - decision - tree - regression - exploring - its - principles - implementation-beb882d756c6>
- Vlaamse Regulator van de Elektriciteits- en Gasmarkt (VREG), ((2021). *Marktmonitor 2021* (tech. rap.). VREG. <https://www.vreg.be/sites/default/files/document/rapp-2021-18.pdf>
- Wang, X., Sun, Y., Luo, D., & Peng, J. (2022). Comparative study of machine learning approaches for predicting short-term photovoltaic power output based on weather type classification. *Energy*, 240(122733). <https://doi.org/https://doi.org/10.1016/j.energy.2021.122733>
- Wemyss, D., Cellina, F., Lobsiger-Kägi, E., de Luca, V., & Castri, R. (2019). Does it last? Long-term impacts of an app-based behavior change intervention on household electricity savings in Switzerland. *Energy Research & Social Science*, 47, 16–27. <https://doi.org/https://doi.org/10.1016/j.erss.2018.08.018>