

Eindrapportage

CASUS ZONNECEL-ANALYSE II

Finn Alberts, Laurent Dassen, Dajmen Graus en Bram
Verheijen
ZUYD HOGESCHOOL | HBO ICT

**ZU
YD**

Inhoudsopgave

1 Aanleiding	3
2 Doelstelling	4
3 Requirements.....	5
4 Data	6
5 Aanpak	7
6 Resultaten	8
6.1 Beoordeling en cleaning van de dataset.....	8
6.1.1 Data van de temperatuursensoren.....	8
6.1.2 Data van het weerstation.....	8
6.1.3 Data van de zonnepanelen	9
6.1.4 Algemene stappen voor datapreparatie.....	10
6.1.5 Beoordeling dataset.....	10
6.2 Data-analyses	11
6.2.1 Verband sensortemperatuur – spanning.....	11
6.2.2 Verband sensortemperatuur – efficiëntie	12
6.2.3 Verband lichtinstraling – vermogen.....	13
6.2.4 Verband lichtinstraling – stroomsterkte	14
6.2.5 Verband sensortemperatuur – lichtinstraling – vermogen	15
6.3 Datavisualisatietools	18
6.3.1 Grafana.....	18
6.3.2 Streamlit.....	18
6.3.3 Tableau.....	19
6.3.4 Kibana.....	19
6.3.5 Power BI	20
6.3.6 Dash	21
6.3.7 Vergelijking en gekozen tool.....	22
6.4 Gerealiseerde dashboard.....	23
7 Discussie.....	25
8 Conclusie	26
9 Verwijzingen.....	27

1 Aanleiding

Het bijzonder lectoraat duurzame energie van Zuyd Hogeschool doet op de Chemelot Campus in Sittard onderzoek naar de efficiëntie van zonnecellen, die daarnaast nog een andere functie hebben. Op dit moment wordt onderzoek gedaan naar het verwerken van zonnecellen in geluiddempende wanden. Hierbij worden veel metingen gedaan bij de zonnepanelen. Momenteel worden deze metingen nog niet uitgevoerd voor deze geluiddempende wanden. Er is wel data beschikbaar van het Rolling Solar project, een project waarbij zonnepanelen in wegdek werden verwerkt. Deze data is (qua structuur) vergelijkbaar met de data die straks zal worden gemeten bij de zonnepanelen in de geluidwanden.

Omdat het opwekkend vermogen van de zonnepanelen mede afhankelijk is van het weer, worden ook weersmetingen gedaan. Het lectoraat wil graag het rendement van de zonnepanelen verhogen en voorspellingen doen voor het rendement van nieuwe panelen.

Oprachtgevers in deze zijn Fallon Colberts en Chris Mass-Protzen. Daarnaast is Sowande Boksteen vanuit het lectoraat Data Intelligence betrokken.

2 Doelstelling

Het doel van dit project is het vinden van verbanden in de data en het visualiseren daarvan. Daarnaast is het doel om aan de hand van deze informatie voorspellingen te kunnen doen voor nieuwe metingen. Deze nieuwe metingen kunnen ook voor nieuwe panelen (op nieuwe locaties) zijn of voor panelen met andere functies (dan wegdek of geluidswand).

3 Requirements

Om de doelstelling te concretiseren, worden in samenspraak met de opdrachtgever eisen opgesteld. Middels wekelijks overleg worden deze opgehaald en gevalideerd. Vervolgens worden deze geprioriteerd middels de MoSCoW-methode (Wikipedia-community, 2021). De gespecificeerde eisen zijn te vinden in Tabel 1. Een definitie van begrippen is te vinden in Bijlage A – Woordenlijst requirements.

Tabel 1 Eisen

Nummer	Eis	Prioriteit
F1	De data van de zonnecellen en weerstations wordt gevisualiseerd in een dashboard.	Should
F2	De gebruiker kan variabelen ingeven voor invoer en voor uitvoer, op basis waarvan naar correlaties wordt gezocht.	Should
F3	De correlaties binnen deze data van de zonnecellen worden gevisualiseerd.	Must
F4	De correlatie tussen sensortemperatuur en spanning wordt gevisualiseerd.	Must
F5	De correlatie tussen sensortemperatuur en efficiëntie wordt gevisualiseerd.	Must
F6	De correlatie tussen lichtinstraling en vermogen wordt gevisualiseerd.	Must
F7	De correlatie tussen lichtinstraling en stroomsterkte wordt gevisualiseerd.	Must
F8	Voor de verschillende correlaties worden modellen getraind.	Must
F11	De gebruiker kan op basis van een getraind model een voorspelling opvragen aan de hand van nieuwe invoerwaarden via een user interface.	Should
F9	De ruwe data wordt gevisualiseerd in het dashboard.	Should
N1	Er wordt rekening gehouden met missende data.	Must
F10	De data van verschillende metingen kan gelijktijdig worden gevisualiseerd in één grafiek, mits de assen hetzelfde zijn.	Should
N2	De data van de zonnecellen en weerstations wordt realtime opgehaald.	Could
N3	Het dashboard is onafhankelijk van het door de gebruiker gebruikte platform te gebruiken.	Could
N4	Er wordt enkel gebruik gemaakt van open source protocollen en software.	Could
N5	De ruwe data moet niet uit het dashboard geëxtraheerd kunnen worden, met uitzondering van afbeeldingen.	Should
N6	Alleen geautoriseerde gebruikers hebben toegang tot het dashboard.	Won't
F11	Het maximale toelaatbare verschil in temperatuur van de sensoren op de zonnepanelen, op basis waarvan data wordt gecleaned, is verstelbaar.	Should
N7	Het dashboard is webbased.	Should
F12	De door het model benaderde formule voor de verschillende correlaties is te zien in het dashboard.	Must
F13	Het gemiddelde weerbericht over de geselecteerde periode wordt weergegeven in het dashboard.	Could

4 Data

De data die tijdens dit project zal worden gebruikt, is de data van het Rolling Solar project. De opbouw van deze datasets is vergelijkbaar met de opbouw van de datasets van het project met de geluiddempende wanden.

Er zijn drie datasets:

- Een dataset met data van de zonnepanelen. Deze bevat de spanning en stroomsterkte van de zonnepanelen.
- Een dataset met data van de temperatuursensoren. Deze bevat de temperatuur van de zonnepanelen.
- Een dataset met weersdata. Deze bevat de Global Horizontal Irradiance (GHI, ofwel lichtinstraling), omgevingstemperatuur, luchtvochtigheid, temperatuur waarbij dauw ontstaat, luchtdruk, windrichting en windsnelheid. Het meetstation voor deze weerdata staat op de Chemelot Campus en dus in de buurt van de zonnepanelen.

Iedere dataset meet iedere minuut, 24 uur per dag, 7 dagen in de week. De datasets zijn beschikbaar per dag en per week. Voor dit project wordt gewerkt met datasets per week, omdat hierbij ook de fase¹ vermeldt staat, welke noodzakelijk is voor de analyses. Bij de datasets per dag is deze fase nog niet toegevoegd.

De eerste metingen zijn van week 1 van 2022. Er komt nog steeds wekelijks nieuwe data beschikbaar.

Een gedetailleerde specificatie van de structuur van de datasets is te vinden in hoofdstuk 6.1 Beoordeling en cleaning van de dataset.

¹ Tijdens het RollingSolar project werden verschillende zonnepanelen getest. In fase A werden silicium zonnepanelen gebruikt. In fase B waren dit dunne film zonnepanelen.

5 Aanpak

Het project zal in drie stappen worden doorlopen: data cleaning, data-analyse en datavisualisatie. Omdat de data die wordt gebruikt de data van het Rolling Solar project is en niet de data van het project met de geluiddempende wanden, is het belangrijk om hierbij op te merken dat de resultaten van de analyse niet centraal staan, maar in plaats daarvan de manier waarop tot deze resultaten wordt gekomen. Op deze manier kunnen dezelfde analyses in de toekomst worden uitgevoerd voor de data van de geluidswanden.

Tijdens de data cleaning zal de dataset worden geanalyseerd en de structuur onderzocht. Hierbij wordt tevens de volledigheid en kwaliteit van de data beoordeeld. Hiervoor wordt gekeken naar onder andere missende waardes en (duidelijke) foutieve metingen. Tijdens de data cleaning wordt tevens gekeken hoe met deze foutieve/missende data moet worden omgegaan. Dit levert naast een opgeschoonde dataset ook een beoordeling van de dataset op en een methode om toekomstige datasets van deze projecten te cleanen. Deze beoordeling zal de doorslag geven of de data-analyses wel of niet (goed) kunnen worden uitgevoerd op basis van de data.

Bij de data-analyse worden de daadwerkelijke analyses gedaan. Hiermee wordt onderzocht op welke manier de verschillende features verbonden zijn met elkaar. Net als bij de data cleaning wordt hierbij gedocumenteerd welke analyses zijn uitgevoerd en hoe deze voor toekomstige datasets kunnen worden uitgevoerd. Hierbij wordt minimaal het verband sensortemperatuur-spanning, sensortemperatuur-efficiëntie, lichtinstraling-vermogen en lichtinstraling-stroomsterkte geanalyseerd.

Voor de datavisualisatie zal een dashboard worden gerealiseerd waarin de resultaten van de data-analyse kunnen worden weergegeven. Hierbij wordt het dashboard gesorteerd van algemeen naar steeds specifieker. Daar de resultaten van de analyse eenzelfde vorm zullen hebben voor zowel het Rolling Solar project als het project met de geluiddempende wand, zal dit dashboard ook geschikt zijn voor dit nieuwe project. Om een goede keuze te kunnen maken voor de te gebruiken tool voor het realiseren van het dashboard, zullen verschillende dashboardtools met elkaar worden vergeleken. Hierbij worden de tools op verschillende criteria beoordeeld.

6 Resultaten

6.1 Beoordeling en cleaning van de dataset

Om de dataset als 'bruikbaar' te classificeren moet de dataset gecontroleerd en beoordeeld worden. De dataset bestaat uit een drietal sets: de temperatuursensoren, het weerstation en de metingen van de zonnepanelen. Deze drie sets hebben een aantal kolommen waarin waardes staan vermeld. De datasets meten alle drie iedere minuut. Hieronder zullen de verschillende datasets worden toegelicht, met daarbij de manier waarop deze is opgeschoond (gecleaned).

6.1.1 Data van de temperatuursensoren

De dataset van de temperatuursensoren bevat de metingen van de temperatuursensoren op de zonnepanelen. Dit zijn dus de temperaturen van de zonnepanelen zelf en niet van de omgevingstemperatuur.

Tabel 2 Temperatuurset beoordeling

Kolomnaam	Beoordeling
Sensortype	Sensortype geeft aan of de temperatuursensor tot fase A of tot fase B behoort. Deze kolom bevat geen lege waardes. Er is geen koppeling mogelijk naar de losse zonnepanelen in de "Power"-dataset.
HumanDate	De datum en tijd (tot op seconde nauwkeurig) van de meting. Deze kolom bevat geen lege waardes.
ROMID	Het ROMID is een uniek ID per sensor. Deze kolom bevat geen lege waardes.
Temperaturer	De gemeten temperatuur. Hier zijn veel NaN-waardes aanwezig, met name bij fase A.

De temperatuur die door de sensoren wordt gemeten bevat uitschieters. Om deze uitschieters eruit te filteren, wordt een maximaal verschil in temperatuur binnen één minuut gedefinieerd. Is een temperatuurverschil groter, dan wordt deze waarde omgezet naar NaN. Het maximaal toegestane temperatuurverschil staat ingesteld op 2 graden, maar in de toekomst moet deze waarde kunnen worden aangepast binnen het dashboard. Zie eis F11

Per fase zijn meerdere temperatuursensoren aanwezig. Er wordt daarom per fase een gemiddelde genomen per minuut. Dat wil zeggen dat een dataframe wordt aangemaakt, met daarin per minuut twee gemiddeldes: één voor fase A en één voor fase B.

6.1.2 Data van het weerstation

De dataset van het weerstation bevat verschillende weervariabelen. Dit weerstation staat op de Chemelot Campus en daarmee dus in de buurt van de zonnepanelen.

Tabel 3 Weerstationset beoordeling

Kolomnaam	Beoordeling
Sensortype	Altijd "Testmeteo".
HumanDate	De datum en tijd (tot op seconde nauwkeurig) van de meting. Deze kolom bevat geen lege waardes.
GHI	Global Horizontal Irradiance, ookwel lichtinstraling. Van weken 18 t/m 26 in 2022 zijn de waardes -100. Dit betekent dat het weerstation geen GHI-waardes heeft gemeten in die weken. Vanaf week 27 en verder staat vaak het getal 65 535 en 65 534 als waarde voor de GHI. Deze waardes

	worden geplaatst in de nacht wanneer er geen sprake is van zonnestraling, ofwel GHI.
Tamb	Omgevingstemperatuur. Deze kolom bevat geen lege waardes.
RH	Luchtvochtigheid. Deze kolom bevat geen lege waardes.
dew	De temperatuur waarbij dauw ontstaat. Deze kolom bevat geen lege waardes.
Pamb	De luchtdruk. Deze kolom bevat geen lege waardes.
Wdir	De windrichting in graden. Deze kolom bevat weinig foute waardes. Op een paar plekken -100. Dit moet worden gezien als NaN.
Wspeed	De windsnelheid. Deze kolom bevat weinig foute waardes. Op een paar plekken -100. Dit moet worden gezien als NaN.

Omdat bij enkele kolommen te zien is dat -100 wordt gebruikt als aanduiding voor NaN, wordt voor alle kolommen -100 omgezet naar NaN, ook in de kolommen waar nu nog geen waardes van -100 te zien zijn.

Daarnaast worden waardes van boven de 2000 voor de GHI omgezet naar NaN. De waarde van 2000 is gekozen, omdat de maximumwaarde op aarde rond de 1360 W/m² ligt (Newport, sd).

6.1.3 Data van de zonnepanelen

De data van de zonnepanelen bevat verschillende variabelen, over de opgewekte energie door de zonnepanelen.

Tabel 4 Zonnepanelenset beoordeling

Kolomnaam	Beoordeling
Sensortype	Altijd "Power".
HumanDate	De datum en tijd (tot op seconde nauwkeurig) van de meting. Deze kolom bevat geen lege waardes.
ID	Het ID van het zonnepaneel. Dit zijn waardes van 101 t/m 130. Zonnepanelen met IDs van 101 t/m 118 behoren tot fase A. Zonnepanelen met ID 119 en hoger behoren tot fase B. Deze kolom bevat geen lege waardes.
Voltage	Deze kolom bevat de gemeten spanning in volt. Omdat er sprake is van een wisselspanning bevat deze kolom zowel positieve en negatieve waardes. Deze kolom bevat geen lege waardes. Wel bevat deze kolom soms (in de nacht) hele kleine waardes, welke kunnen worden gezien als ruis.
Current	Deze kolom bevat de stroomsterkte in milliampère. Omdat er sprake is van een wisselspanning bevat deze kolom zowel positieve en negatieve waardes. Deze kolom bevat geen lege waardes.
Power	Deze kolom is altijd 0.
Energy	Deze kolom is altijd 0.

Omdat er sprake is van een wisselspanning, waardoor negatieve waardes in de data voorkomen, wordt van zowel voltage als current de absolute waarde genomen. De voltages kunnen af en toe metingen geven van hele kleine getallen (kleiner dan 1). Dit gebeurt ook regelmatig tijdens de nacht, wanneer het voltage zeker 0 moet zijn. Daarom worden deze waardes gezien als ruis en worden voltages kleiner dan 0 omgezet naar NaN. Daarnaast wordt power berekent middels $P =$

$U \cdot I$, oftewel $Power = Voltage \cdot \frac{Current}{1000}$ (current wordt gedeeld door 1000 om van milliampère naar ampère te gaan).

Net als bij de metingen in de temperatuursensor, bevat power regelmatig uitschieters. Echter, deze uitschieters schieten altijd naar 0. Aangezien power wel daadwerkelijk 0 kan zijn, wordt een waarde van 0 uitsluitend omgezet in NaN, wanneer de waarde ervoor groter als 20 watt was. Dit betekent dus dat uitschieters naar 0, waarbij de uitschieter groter is dan 20 watt, naar NaN worden omgezet.

De energy kolom kan worden gebruikt om het aantal w/m^2 op te slaan. Echter wordt deze alleen gebruikt om vervolgens de efficiëntie (rendement) te berekenen. Er wordt daarom direct een nieuwe kolom efficiency gecreëerd en de energy kolom wordt niet gebruikt (en blijft dus op 0). Voor het berekenen van de efficiency wordt gebruik gemaakt van de volgende formule $Efficiency = \frac{Power}{GHI \cdot Oppervlakte\ paneel}$. De oppervlakte van een paneel van fase A is hierin $1,33\ m^2$ en fase B is hierin $1,88\ m^2$. Indien GHI 0 is, wordt de efficiency op NaN gezet.

6.1.4 Algemene stappen voor datapreparatie

Er zijn ook een aantal stappen die voor alle datasets worden uitgevoerd. De eerste stap hierin is het verwijderen van de secondes bij alle HumanDate kolommen. Alle datasets bevatten metingen per minuut. Deze metingen zijn echter niet altijd in dezelfde seconde uitgevoerd. Door de secondes te verwijderen, kunnen de verschillende datasets met elkaar worden gelinkt per minuut.

Daarnaast bevat de dataset ook enige ruis. Deze ruis ontstaat deels door natuurlijke omstandigheden en deels door onnauwkeurigheden van de sensoren. Om deze ruis te verminderen, wordt de data gegroepeerd in groepen van 15 minuten. Voor iedere kolom wordt het gemiddelde genomen over deze 15 minuten.

6.1.5 Beoordeling dataset

De dataset biedt voldoende mogelijkheden voor de uit te voeren analyses. Hierbij zijn wel een aantal belangrijke data cleaning stappen nodig, welke hierboven beschreven zijn.

Daarnaast moet er rekening worden gehouden met defecte sensoren. Zo zijn er in de datasets weken, waarin de sensor voor lichtinstraling defect was. Dit maakt de analyses hier onmogelijk. Het zorgen voor kwalitatief goede, werkende, sensoren is daarom van groot belang voor de analyses. Mochten de sensoren beschadigd raken, is het aan te raden deze zo snel mogelijk te repareren of te vervangen. Als dit niet mogelijk is, is het belangrijk om dit defect te documenteren, zodat vreemde patronen in de data daarmee wel verklaard kunnen worden.

6.2 Data-analyses

Na het uitvoeren van de data cleaning, kunnen de data-analyses worden uitgevoerd. Hiervoor worden modellen getraind. Deze modellen kunnen, nadat ze getraind zijn, op basis van nieuwe data voorspellingen doen. Dit is hetzelfde als het vinden van verbanden/correlaties. De data op basis waarvan de voorspelling wordt gedaan worden de features genoemd. De voorspelde/berekende waarde wordt de afhankelijke variabele genoemd. (Shmueli, Bruce, Gedeck, & Patel, 2020)

Voor alle modellen wordt gebruik gemaakt van multiple linear regression. Multiple linear regression neemt meerdere (onafhankelijke) variabelen en bepaald daarmee een afhankelijke variabele (Shmueli, Bruce, Gedeck, & Patel, 2020). Als voorbeeld kan worden genomen dat het opgewekte vermogen door een zonnepaneel afhankelijk is van (onder andere) sensortemperatuur en lichtinstraling. De algemene formule voor multiple linear regression is te zien in Vergelijking 1.

Vergelijking 1 Algemene formule voor multiple linear regression

$$\hat{y} = b_0 + \sum b_i x_i$$

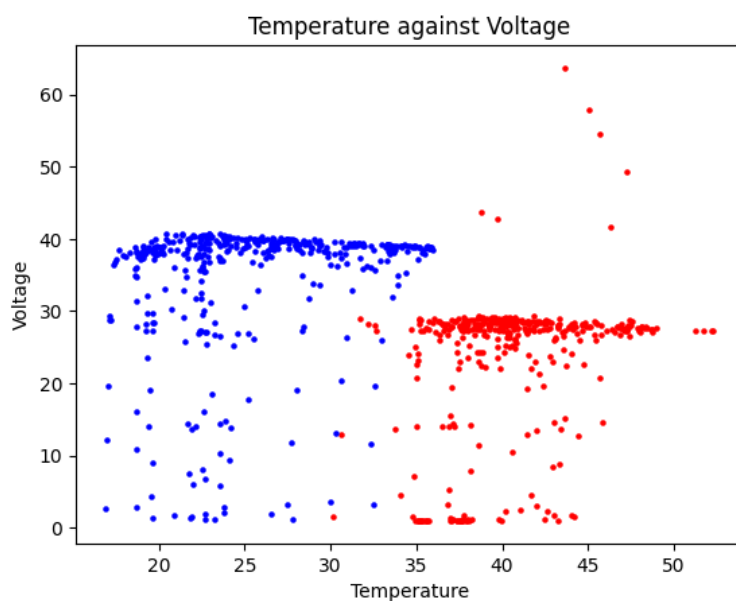
Hierin is \hat{y} de voorspelde waarde, b_0 de bias (waarde als alle features 0 zijn), b_i de coëfficiënt voor een feature en x_i de waarde voor een feature. Tijdens het trainen van het model worden de waarden voor b zo goed mogelijk door het model bepaald.

Er is gekozen voor multiple linear regression, omdat de te onderzoeken verbanden lineair lijken te zijn, wanneer de variabelen tegen elkaar worden uitgezet. Daarnaast geeft de literatuur aan voor de verbanden lichtinstraling-vermogen (Lopes, et al., 2018) en lichtinstraling-stroomsterkte (Richardson, et al., 2019) dat deze lineair zijn.

In onderstaande hoofdstukken zal gebruik worden gemaakt van week 36 van 2022 voor visualiseren en evalueren van de analyses. De resultaten van deze week zijn vergelijkbaar met andere weken.

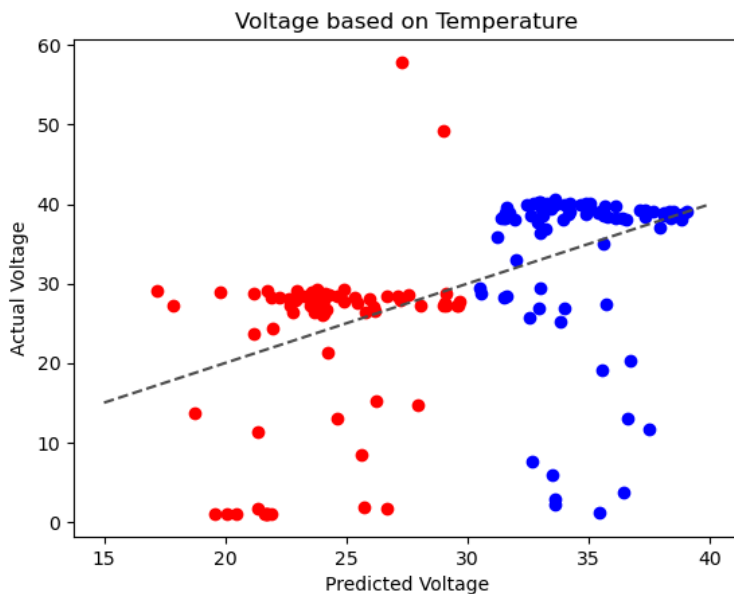
6.2.1 Verband sensortemperatuur – spanning

Het eerste onderzochte verband is het verband tussen sensortemperatuur (in graden Celsius) en spanning (in volt). In Figuur 1 zijn deze variabele uitgezet tegen elkaar te zien.



Figuur 1 Spanning uitgezet tegen sensortemperatuur fase A (rood) en fase B (blauw) in week 36 van 2022

Op basis hiervan is een model getraind, waarbij sensortemperatuur fungeert als feature en spanning als afhankelijke variabele. Het resultaat hiervan is te zien in Figuur 2. Daar de voorspellingen bij een perfect model een-op-een overeenkomen met de werkelijkheid, is in deze figuur ook de lijn $y = x$ getekend.

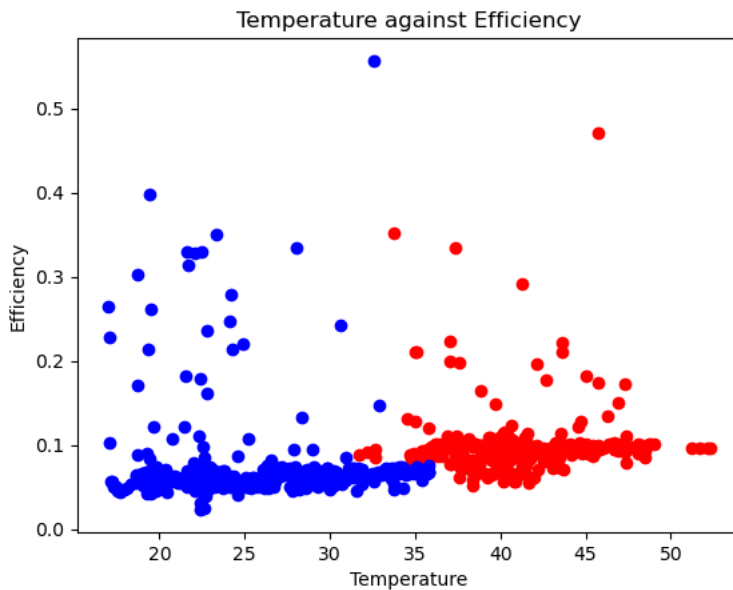


Figuur 2 Voorspellingen voor voltage op basis van sensortemperatuur fase A (rood) en fase B (blauw) in week 36 van 2022

Zoals in Figuur 2 te zien is, is het voor een multiple linear regression model lastig om het voltage te voorspellen, enkel op basis van sensortemperatuur. Het model geeft dan ook een R^2 -score van 11,0% voor fase A en -3,3% voor fase B. De R^2 -score geeft voor een model aan hoeveel procent van de ruis door het model kan worden verklaard. De root mean squared error (RMSE) ligt voor beide rond de 10,4 volt. Dat wil zeggen dat de testdata gemiddeld 10,4 volt van het daadwerkelijke voltage af lag. Een mogelijke verklaring voor deze lage correlatie is dat het voltage erg consistent is. Dat wil zeggen: deze verandert nauwelijks.

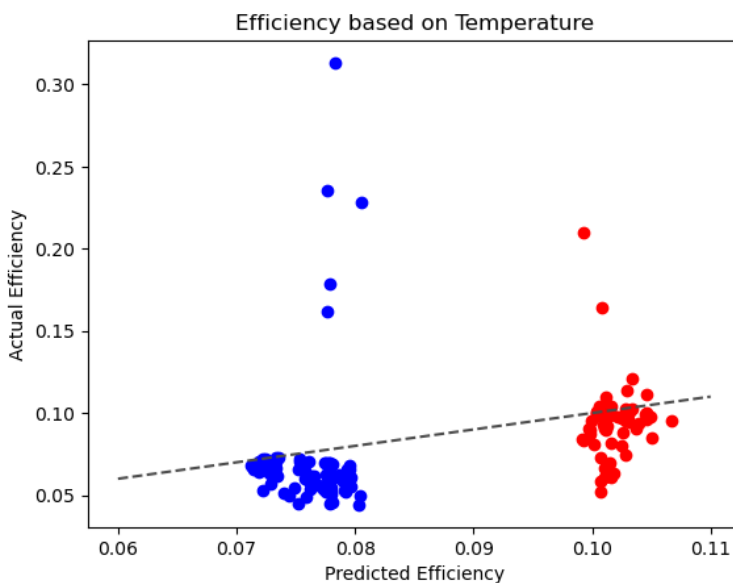
6.2.2 Verband sensortemperatuur – efficiëntie

Het tweede onderzochte verband, is het lineaire verband tussen de sensortemperatuur (graden Celsius) en de efficiëntie (procent). In Figuur 3 zijn deze twee variabelen tegen elkaar uitgezet te zien.



Figuur 3 Efficiëntie uitgezet tegen sensortemperatuur fase A (rood) en fase B (blauw) in week 36 van 2022

Het multiple linear regression model neemt hier de sensortemperatuur als feature en de efficiëntie als afhankelijke variabele. Dit geeft resultaten zoals te zien in Figuur 4. Hierin is wederom de lijn $y = x$ weergegeven, waar alle punten in een perfect model op moeten liggen.

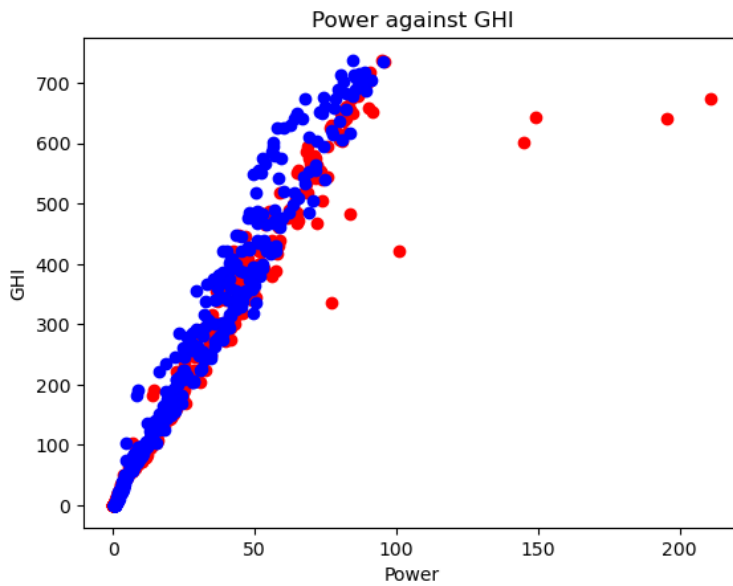


Figuur 4 Voorspellingen voor efficiëntie op basis van sensortemperatuur fase A (rood) en fase B (blauw) in week 36 van 2022

Zoals te zien in Figuur 4 is het model slecht in staat op de efficiëntie goed te voorspellen. De R^2 -score is slechts -12,8% voor fase A en 0,4% voor fase B. De RMSE ligt voor beide rond de 3,3%. Net als bij het voltage is de efficiëntie ook zeer consistent en verandert dus nauwelijks. Dat is een mogelijke verklaring waarom het model minder goed presteert.

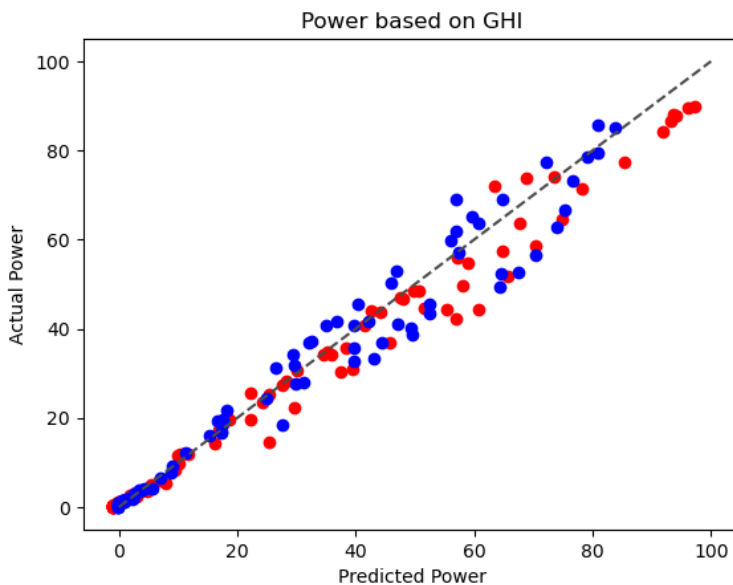
6.2.3 Verband lichtinstraling – vermogen

In Figuur 5 is het opgewekte vermogen (in watt) uitgezet tegen de gemeten lichtinstraling (in W/m^2) uitgezet.



Figuur 5 Lichtinstraling uitgezet tegen opgewekt vermogen fase A (rood) en fase B (blauw) in week 36 van 2022

In het multiple linear regression model wat hiervoor is getraind, is de lichtinstraling gebruikt als feature en het opgewekte vermogen als afhankelijke variabele. In Figuur 6 zijn de voorspelling van het model weergegeven, samen met de lijn $y = x$.



Figuur 6 Voorspellingen voor opgewekt vermogen op basis van lichtinstraling fase A (rood) en fase B (blauw) in week 36 van 2022

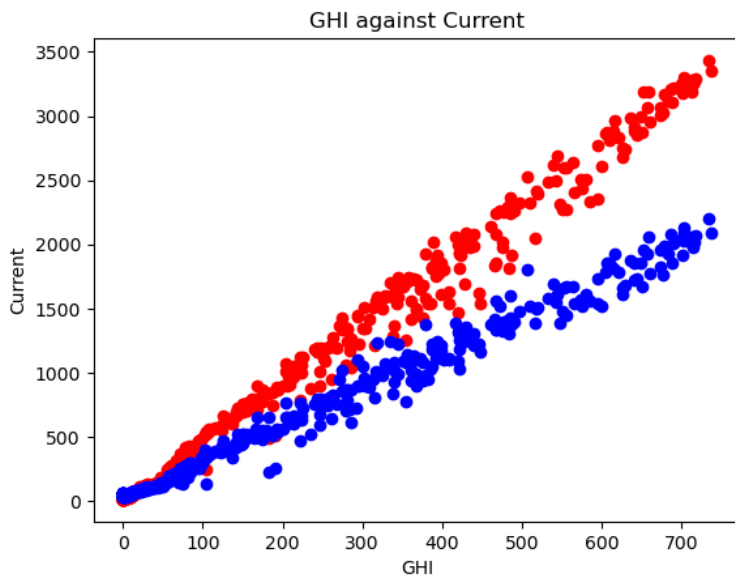
Dit model geeft behoorlijke goede voorspellingen, wat laat zien dat er een sterk verband is tussen lichtinstraling en opgewekt vermogen. De R^2 -score is 96,9% voor fase A en 95,9% voor fase B, wat dit zeer sterke verband verder aantoont. De RMSE ligt op 4,9 watt voor fase A en 5,3 watt voor fase B.

6.2.4 Verband lichtinstraling – stroomsterkte

Figuur 7 laat de lichtinstraling (in W/m^2) zien, uitgezet tegen stroomsterkte (in milliampère).

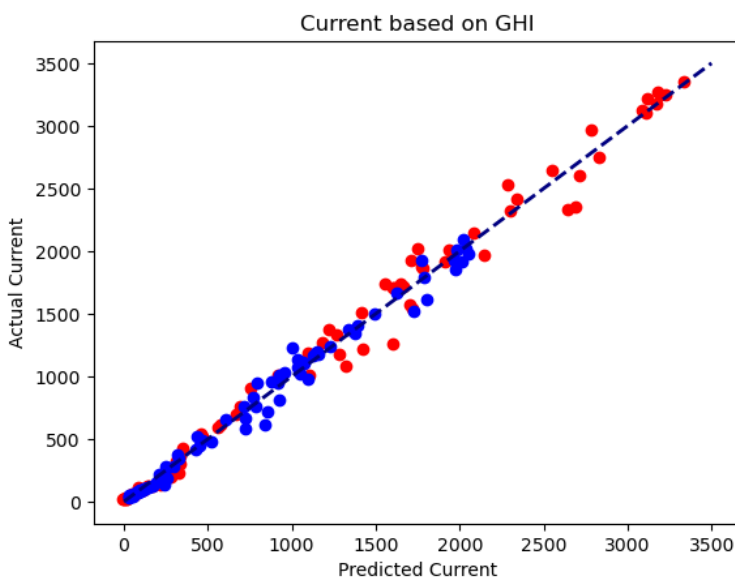
Opvallend is dat fase A sneller stijgt dan fase B, wat een indicatie kan zijn dat fase A beter presteert

dan fase B. Echter kan deze conclusie nog niet worden getrokken, omdat voltage ook een rol speelt bij het uiteindelijke opgewekte vermogen (en daarmee de efficiëntie).



Figuur 7 Stroomsterkte uitgezet tegen lichtinstraling fase A (rood) en fase B (blauw) in week 36 van 2022

Het getrainde multiple linear regression model neemt de lichtinstraling als feature en bepaalt hiermee de afhankelijke variabele stroomsterkte. In Figuur 8 zijn de voorspellingen van dit model te zien. De lijn $y = x$ is hierin wederom ook te zien.



Figuur 8 Voorspellingen voor stroomsterkte op basis van lichtinstraling fase A (rood) en fase B (blauw) in week 36 van 2022

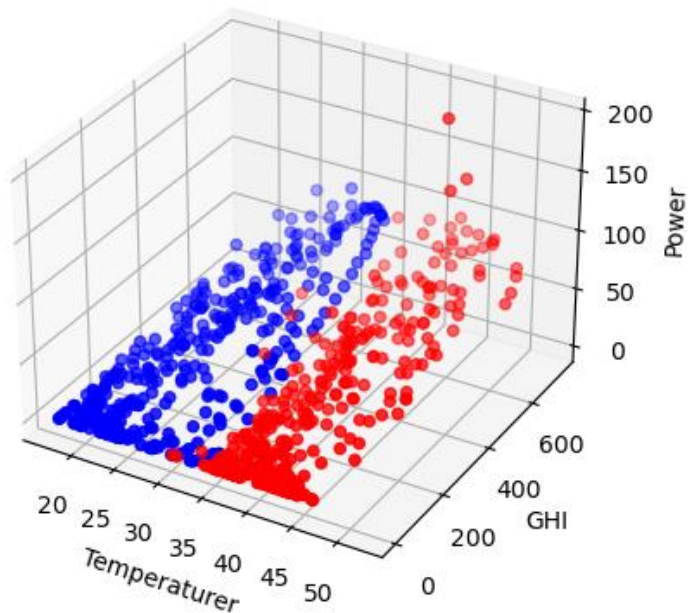
Er is een zeer sterk verband te zien. De R^2 -scores voor dit model zijn 98,8% voor fase A en 98,6% voor fase B, wat bijzonder hoge correlaties aantoont. De RMSE ligt op 116,5 mA voor fase A en 74,6 mA voor fase B.

6.2.5 Verband sensortemperatuur – lichtinstraling – vermogen

Daar sensortemperatuur de spanning en efficiëntie beïnvloed en lichtinstraling de stroomsterkte (en daarmee ook het opgewekte vermogen), is ook een model getraind waarin alle drie deze variabelen

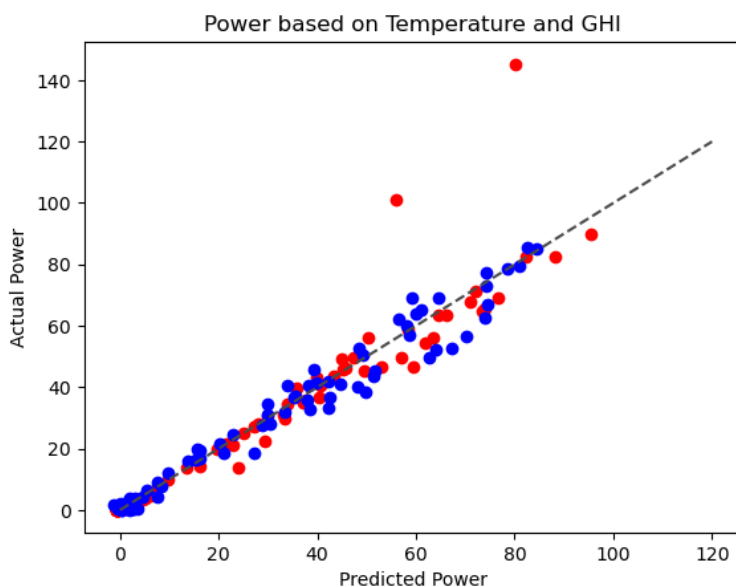
tegen elkaar zijn uitgezet en waarin het verband is onderzocht. In Figuur 9 is een 3D-grafiek te zien van deze variabelen (sensortemperatuur in graden Celsius, lichtinstraling in W/m² en opgewekt vermogen in watt). Hierin is duidelijk een 'schuinliggend vlak' te zien.

Power against Temperature against GHI in 3D



Figuur 9 Opgewekt vermogen uitgezet tegen sensortemperatuur en lichtinstraling fase A (rood) en fase B (blauw) in week 36 van 2022

Het multiple linear regression model wat hiermee is getraind, neemt de lichtinstraling en de sensortemperatuur en voorspelt op basis daarvan het opgewekte vermogen. Dit geeft een voorspellingen zoals te zien in Figuur 10.



Figuur 10 Voorspellingen voor opgewekt vermogen op basis van sensortemperatuur en lichtinstraling fase A (rood) en fase B (blauw) in week 36 van 2022

Er is een sterke correlatie te zien, waarbij de R^2 -score voor fase A op 89,4% ligt en voor fase B op 96,5%. De RMSE ligt op 9,8 watt voor fase A en 4,9 watt voor fase B. De resultaten zijn hiermee vergelijkbaar met de resultaten van de correlatie tussen enkel lichtinstraling en opgewekt vermogen. Dit toont aan dat de sensortemperatuur slechts een zeer beperkte invloed heeft op het opgewekt vermogen. Ook als gekeken wordt naar de formule die het model teruggeeft is dit terug te zien. De formule is te zien in Vergelijking 2. De coëfficiënt voor de temperatuur is slechts 0,0156 voor fase A en 0,3796 voor fase B. Aangezien de temperaturen van de zonnepanelen ongeveer tussen de 20 °C en de 50 °C zitten, is de impact van de temperatuur slechts ongeveer $50 \cdot 0,3796 = 18,98$ watt in het slechtste geval.

Vergelijking 2 Formule voor vermogen op basis van sensortemperatuur en lichtinstraling voor week 36 van 2022

$$\text{Fase A: } P = 0,0156T + 0,1341GHI - 1,1793$$

$$\text{Fase B: } P = 0,3796T + 0,1129GHI - 8,6232$$

6.3 Datavisualisatietools

Het visualiseren van de data wordt gedaan middels een dashboard. Voor het realiseren van zo'n dashboard zijn verschillende tools beschikbaar. Om een goede keuze te kunnen maken voor de te gebruiken tool, worden meerdere tools bekeken en met elkaar vergeleken.

6.3.1 Grafana

Grafana is een tool voor multi-platform open source analytics en interactieve visualisatie webapplicaties. Het biedt grafieken en waarschuwingen voor het web wanneer verbonden met ondersteunde gegevensbronnen. (Grafana, sd)

Grafanalib

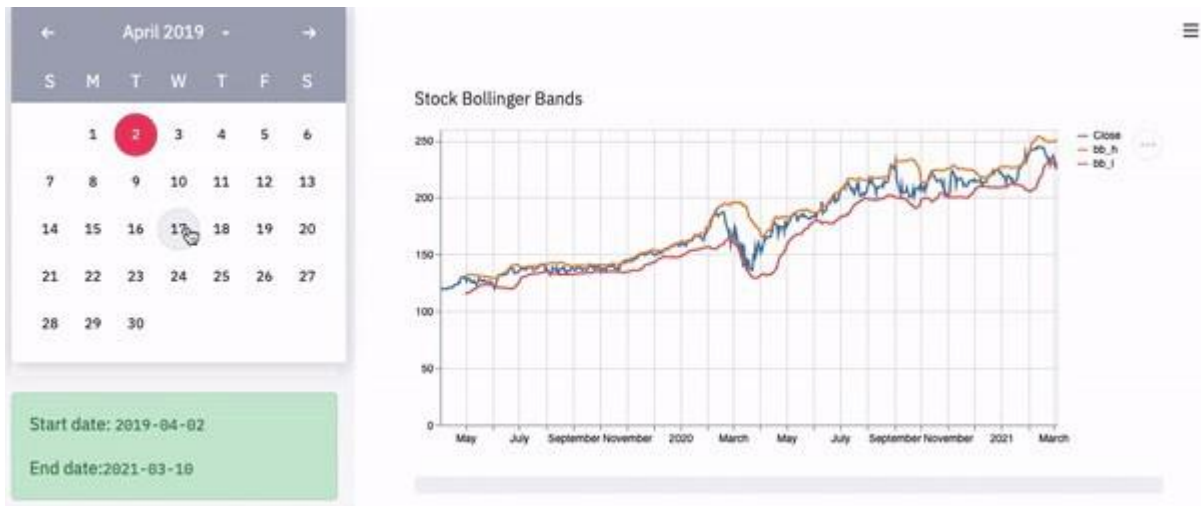
Met Grafanalib is het mogelijk om Grafana-dashboards te genereren op basis van eenvoudige Python-scripts. Grafana migreert dashboards naar de nieuwste versie van het Grafana-schema bij het importeren, wat betekent dat dashboards die met Grafanalib zijn gemaakt, worden ondersteund door alle versies van Grafana. Echter is de documentatie hiervan zeer gelimiteerd. (Grafanalib contributors, 2023)



Figuur 11 Grafana Dashboard (Grafana, sd)

6.3.2 Streamlit

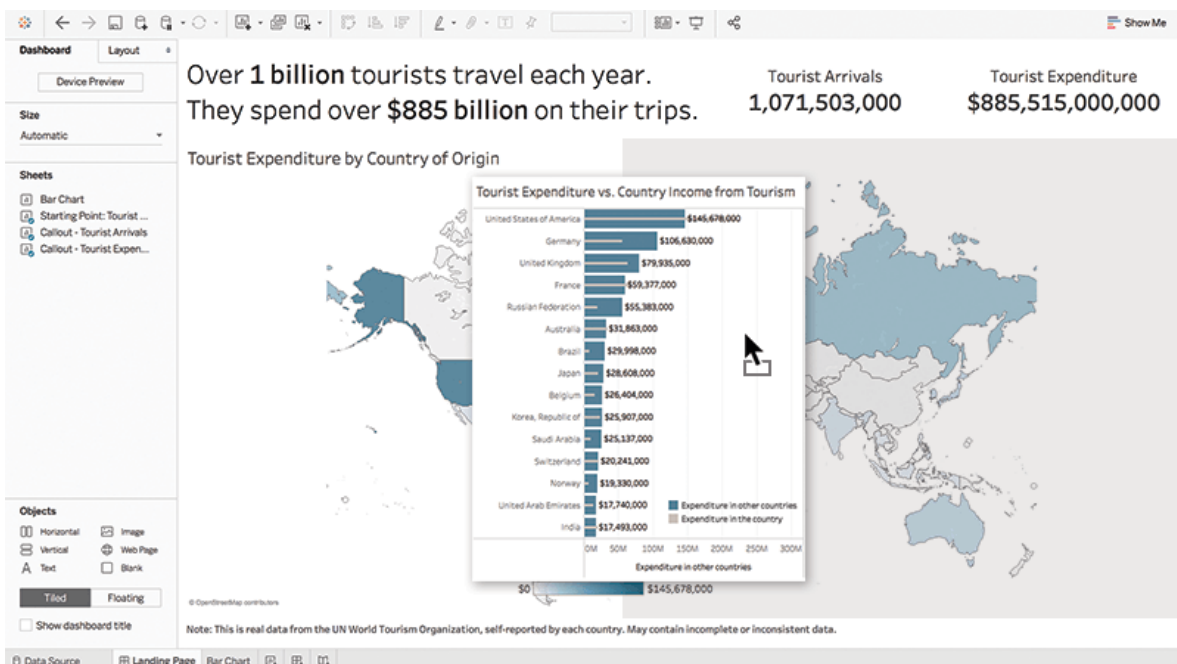
Streamlit is een opensource dashboard tool die het mogelijk maakt om snel en gemakkelijk interactieve dashboards te bouwen. Het maakt gebruik van Python, waardoor het gemakkelijk is om data science toe te passen. Streamlit biedt een aantal handige functies, zoals het makkelijk toevoegen van visuele elementen, zoals grafieken, kaarten en het invoervelden waar gebruikers gegevens kunnen invoeren om de resultaten te personaliseren. (Streamlit, sd)



Figuur 12 Streamlit Dashboard (Kuo, 2021)

6.3.3 Tableau

Tableau is een dashboard tool voor interactieve dashboards voor data-analyse en -visualisatie. Het maakt gebruik van een drag-and-drop interface waarmee gebruikers visuele elementen, zoals grafieken en kaarten, kunnen toevoegen aan hun dashboard zonder dat ze code hoeven te schrijven. Tableau biedt ook uitgebreide opties voor het filteren, groeperen en op andere manieren manipuleren van gegevens, zodat gebruikers diepgaande inzichten kunnen krijgen in hun data. (Tableu, sd)

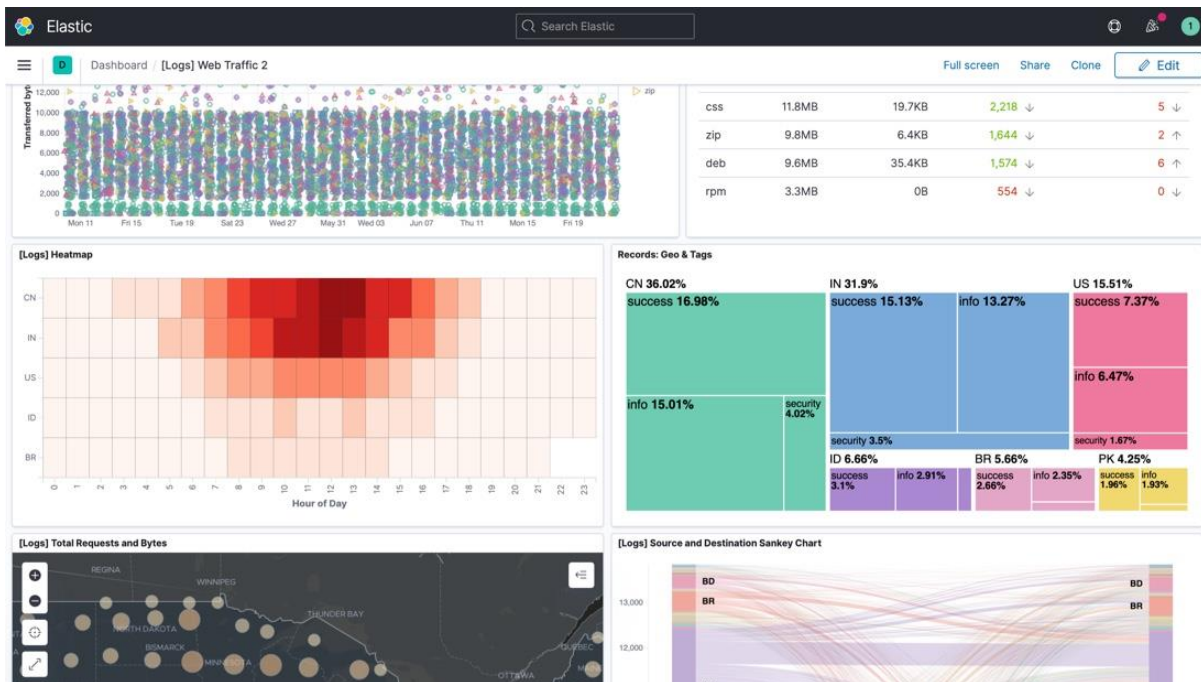


Figuur 13 Tableau Dashboard (Tableu, sd)

6.3.4 Kibana

Kibana is een open source dashboard tool die wordt gebruikt voor het analyseren en visualiseren van gegevens in Elasticsearch. Elasticsearch is een open-source zoektechnologie die gebruikmaakt van Lucene, een indexerings- en zoekmachinebibliotheek, om gegevens te indexeren en te zoeken. Dit maakt het mogelijk om snel grote hoeveelheden gegevens te doorzoeken en inzichten te verkrijgen. Kibana biedt gebruikers de mogelijkheid om real-time interactieve dashboards te maken

en te delen met anderen, zodat zij gegevens kunnen onderzoeken en inzichten kunnen verkrijgen. Het biedt ook verschillende functies voor het zoeken, filteren en samenwerken aan gegevens. (Kibana, sd)



Figuur 14 Kibana Dashboard (Kibana, sd)

6.3.5 Power BI

Power BI is een oplossing voor bedrijfsanalyse die is ontwikkeld door Microsoft. Het richt zich op het bieden van interactieve visualisatie en business intelligence-mogelijkheden aan eindgebruikers. Power BI biedt ook verschillende functies voor het verbinden en samenwerken aan gegevens, zodat gebruikers gegevens kunnen combineren en analyseren vanuit verschillende bronnen. (Microsoft Power BI, sd)

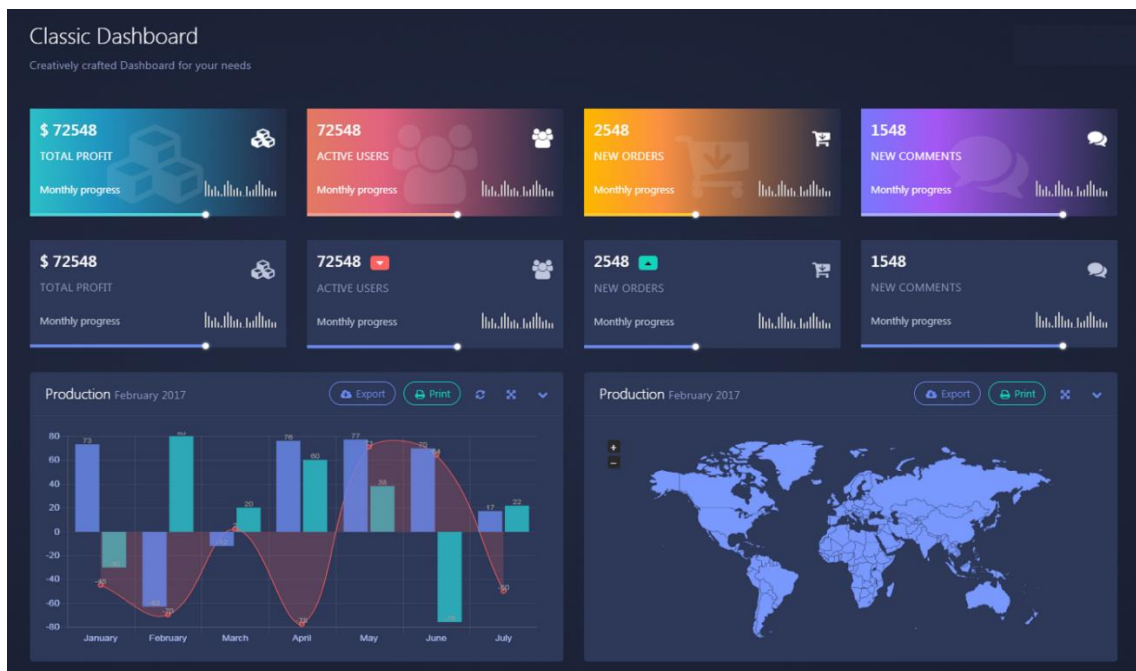


Figuur 15 Power BI Dashboard (maggiesMSFT, et al., 2023)

6.3.6 Dash

Dash is een open-source dashboard tool voor Python die wordt gebruikt voor het bouwen van interactieve web-based dashboards. Het maakt gebruik van de populaire web-ontwikkeling framework Flask en de datavisualisatie library Plotly om interactieve dashboards te bouwen. Dit maakt het eenvoudig om web-applicaties te bouwen met behulp van Python. Plotly biedt verschillende soorten grafieken en visualisaties, waaronder lijngrafieken, staafgrafieken en kaarten. Dit maakt het mogelijk om gegevens op een visueel aantrekkelijke en interactieve manier te presenteren.

Dash biedt gebruikers de mogelijkheid om real-time dashboards te maken en te delen met anderen, zodat zij gegevens kunnen onderzoeken en inzichten kunnen verkrijgen. Het biedt ook verschillende functies voor het samenwerken aan gegevens en het bouwen van aangepaste visualisaties. (Plotly, sd)



Figuur 16 Dash Dashboard (Timotic, 2018)

6.3.7 Vergelijking en gekozen tool

Voor het kiezen voor de optimale tool voor het project, worden de tools met elkaar vergeleken op de volgende aspecten:

- **Open source:** Hierbij wordt onderzocht of de tool open source is. Het gebruik van open source tools is een van de eisen vanuit de requirements (eis N4).
- **Documentatie:** Hierbij wordt onderzocht tot hoeverre mate er documentatie over de tool te vinden is. Onder andere wordt gekeken naar websites en video's die uitleg geven over het gebruik van de tool. Goede en complete documentatie is belangrijk om de tool goed te kunnen gebruiken.
- **Webapplicatie:** Hierbij wordt onderzocht of de tool kan worden gebruikt als een webapplicatie en gemakkelijk het zou zijn om de tool in de vorm van een website te implementeren. Een webapplicatie is een van de eisen vanuit de requirements (eis N7).
- **Python:** Hierbij wordt onderzocht tot in hoeverre mate de programmeertaal Python compatibel is met de tool. Python is de standaard in de industrie voor data science applicaties (Institute of data, 2020).
- **Data visualisatie:** Hierbij wordt onderzocht tot hoeverre mate de tool geschikt is voor het visualiseren van data.
- **Data science:** Hierbij wordt onderzocht in hoeverre mate de tool geschikt is voor het implementeren van de data science algoritmen welke nodig zijn om tot de visualisaties te komen.
- **Authenticatie en autorisatie:** Hierbij wordt onderzocht of de tool de mogelijkheid aanbiedt voor het integreren van authenticatie en autorisatie. Authenticatie en autorisatie is een van de eisen vanuit de requirements (eis N6).

Afhankelijk van de mate waarin de tool aan een onderdeel voldoet, wordt de tool in de categorie een kleur toegekend. Voldoet de tool helemaal aan een aspect, krijgt het de kleur groen. In het geval dat de tool niet voldoet aan een bepaald aspect, krijgt het de kleur rood. In de gevallen waarbij de tool niet direct aan het aspect voldoet of als de documentatie hierover niet duidelijk genoeg is, krijgt

het de gele kleur. Afhankelijk van deze vergelijking, kan een conclusie worden getrokken van welke tool het meest compatibel met deze opdracht.

Tabel 5 Beoordeling dashboardtools op criteria

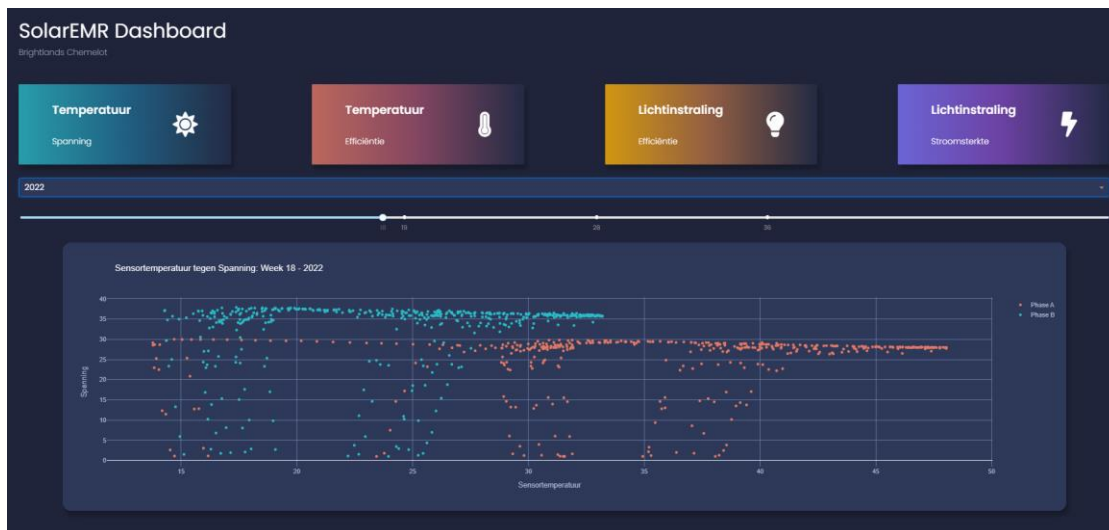
	Grafana ²	Streamlit ³	Tableau ⁴	Kibana ⁵	Power BI ⁶	Dash ⁷
Open source	Green	Green	Green	Green	Red	Green
Documentatie	Green	Green	Green	Green	Green	Green
Webapplicatie	Red	Green	Yellow	Red	Red	Green
Python	Yellow	Green	Yellow	Yellow	Green	Green
Data visualisatie	Green	Yellow	Green	Green	Green	Green
Data science	Red	Green	Yellow	Red	Yellow	Green
Authenticatie	Green	Green	Yellow	Green	Green	Green

De uiteindelijke dashboard tool die blijkt het meest geschikt te zijn voor het project, is Dash. Deze tool voldoet aan alle categorieën in vergelijking met de andere tools. In alle aspect scoort Dash groen, waaruit blijkt dat deze tool meer dan geschikt is voor het creëren van een dashboard voor dit project.

6.4 Gerealiseerde dashboard

Het in Dash gerealiseerde dashboard laat de vier verschillende verbanden zien. Bovenin het scherm kan worden gekozen voor één van de vier verbanden, welke vervolgens wordt weergegeven. Middels een dropdown-menu en slider kan voor een jaar en week worden gekozen.

De eerste grafiek op het scherm laat de twee variabelen uitgezet tegen elkaar zien. Voor fase A en fase B worden verschillende kleuren gebruikt. Een schermafbeelding is te zien in Figuur 17.



Figuur 17 Schermafbeelding van dashboard met selectieknoppen

² (Grafana, sd)

³ (Streamlit, sd) en (Streamlit, sd)

⁴ (Tableau, sd) en (Tableau, sd)

⁵ (Kibana, sd)

⁶ (Microsoft Power BI, sd)

⁷ (Plotly, sd) en (Plotly, sd)

Onder deze grafiek wordt een tweede grafiek weergegeven, waarin de voorspelling van het model te zien zijn. In deze grafiek wordt een voorspelde waarde uitgezet tegen de daadwerkelijke waarde. Omdat in een perfect model de voorspelde waarde precies overeenkomt met de daadwerkelijke waarde, zouden voor sterke verbanden alle punten op een rechte lijn $y = x$ moeten liggen. Ook hierin worden fase A en fase B in verschillende kleuren weergegeven.

Rechts van de tweede grafiek wordt de formule weergegeven die het model heeft benaderd voor de verschillende fases. Ook wordt hier de R^2 -score en de RMSE voor de twee fases weergegeven. Dit is te zien in Figuur 18.



Figuur 18 Schermafbeelding van het dashboard met de verschillende grafieken

7 Discussie

Een discussiepunt voor binnen dit project is dat uitsluitend onderzoek is gedaan naar lineaire verbanden. Zoals beschreven in hoofdstuk 6.2 Data-analyses is voor multiple linear regression gekozen, vanwege het theoretische verband wat te vinden is in de literatuur. Ook het patroon wat te zien is in een plot van de data lijkt voor alle vier lineair te zijn. Echter, het onderzoeken van andere soorten verbanden, zoals bijvoorbeeld exponentiële, zou mogelijk een toegevoegde waarde kunnen hebben gehad voor dit project.

8 Conclusie

De datasets van het RollingSolar project vereisen een aantal belangrijke data cleaning stappen, om tot een bruikbare dataset te komen. Aan de hand van deze nieuwe datasets kunnen de verschillende verbanden worden onderzocht.

Voor het vinden van de verbanden wordt gebruik gemaakt van multiple linear regression, omdat bij het plotten van de data er een lineair verband lijkt te zijn. Daarnaast geeft de literatuur voor twee van de vier verbanden ook aan dat dit verband lineair moet zijn. Bij de verbanden lichtinstraling-vermogen en lichtinstraling-stroomsterkte zijn zeer sterke verbanden te zien (R^2 -score groter dan 95%). Voor de verbanden sensortemperatuur-spanning en sensortemperatuur-efficiëntie kunnen geen lineaire verbanden worden gevonden.

De vier (al dan niet aanwezige) verbanden worden gevisualiseerd in een dashboard. Hiermee is de doelstelling, zoals gedefinieerd in hoofdstuk 2 Doelstelling, bereikt. Kijken naar de eisen zoals gesteld in hoofdstuk 3 Requirements, wordt aan alle must eisen voldaan. Ook aan een aantal should and could eisen wordt voldaan, maar er zijn ook nog een aantal should, could en won't eisen die nog niet zijn gerealiseerd.

Aanbevolen wordt om te starten met de uitrol van het dashboard naar het web. Hiermee wordt het dashboard beschikbaar om te gebruiken. Gelijktijdig wordt aanbevolen om de ontwikkeling van het dashboard voort te zetten en de nog niet gerealiseerde eisen te realiseren. Hierbij wordt aangeraden om eis N2, het realtime ophalen van de data, als eerste te realiseren. Daarmee wordt veel handwerk (het handmatig uploaden van datasets van nieuwe weken) weggenomen.

9 Verwijzingen

- Grafana. (sd). *Grafana*. Opgehaald van Grafana Labs: <https://grafana.com/>
- Grafanalib contributors. (2023, januari 9). *weaveworks/grafanalib*. Opgehaald van Github: <https://github.com/weaveworks/grafanalib>
- Institute of data. (2020, februari 10). *Why is the Data Science Industry demanding Python?* Opgehaald van Institute of data: <https://www.institutedata.com/blog/why-is-the-data-science-industry-demanding-python/>
- Kibana. (sd). *Your window into the Elastic Stack*. Opgehaald van Kibana: <https://www.elastic.co/kibana/>
- Kuo, C. (2021, maart 6). *Building a Stock Market App with Python Streamlit in 20 Minutes*. Opgehaald van Medium: <https://python.plainenglish.io/building-a-stock-market-app-with-python-streamlit-in-20-minutes-2765467870ee>
- Lopes, F. M., Silva, H. G., Salgado, R., Cavaco, A., Canhoto, P., & Collares-Pereira, M. (2018). Short-term forecasts of GHI and DNI for solar energy systems operation: assessment of the ECMWF integrated forecasting system in southern Portugal. *Solar Energy Volume 170*, pp. 14-30.
- maggiesMSFT, JustPies, PRMerger14, TimShererWithAquent, davidiseminger, v-hearya, & v-kents. (2023, januari 6). *Inleiding tot dashboards voor Power BI-ontwerpers*. Opgehaald van Microsoft Learn: <https://learn.microsoft.com/nl-nl/power-bi/create-reports/service-dashboards>
- Microsoft Power BI. (sd). *Maak direct impact met je data*. Opgehaald van Microsoft Power BI: <https://powerbi.microsoft.com/nl-nl/>
- Newport. (sd). *Introduction to Solar Radiation*. Opgehaald van Newport: <https://www.newport.com/t/introduction-to-solar-radiation>
- Plotly. (sd). *Authentication*. Opgehaald van Plotly Documentation: <https://dash.plotly.com/authentication>
- Plotly. (sd). *Low-Code Data Apps*. Opgehaald van Plotly: <https://plotly.com/>
- Richardson, W. J., Cañadillas, D., Moncada, A., Guerrero-Lemus, R., Shephard, L., Vega-Avila, R., & Krishnaswami, H. (2019, februari 17). Validation of All-Sky Imager Technology and Solar. *Applied Sciences*, p. 684.
- Shmueli, G., Bruce, P. C., Gedeck, P., & Patel, N. R. (2020). *Data mining for business analytics*. Hoboken: Wiley.
- Streamlit. (sd). *A faster way to build and share data apps*. Opgehaald van Streamlit: <https://streamlit.io/>
- Streamlit. (sd). *Authentication without SSO*. Opgehaald van Streamlit documentation: <https://docs.streamlit.io/knowledge-base/deploy/authentication-without-sso>
- Tableau. (sd). *Authentication*. Opgehaald van Tableau documentation: https://help.tableau.com/current/server/en-us/security_auth.htm

Tableau. (sd). *Dashboards done right* . Opgehaald van Tableau: <https://www.tableau.com/learn/get-started/dashboards>

Tableau. (sd). *Het meest toonaangevende analyseplatform ter wereld*. Opgehaald van Tableau: <https://www.tableau.com/nl-nl>

Timotic, M. (2018, juni 6). *Phases of an app development lifecycle*. Opgehaald van TMS: <https://tms-outrsource.com/blog/posts/app-development-lifecycle/>

Wikipedia-community. (2021, april 30). *MoSCoW-methode*. Opgehaald van Wikipedia: <https://nl.wikipedia.org/wiki/MoSCoW-methode>

Bijlage A – Woordenlijst requirements

Term	Beschrijving
Cleanen	Het opschonen van de data om deze klaar te maken voor analyses.
Correlatie	Een samenhangend verband.
Dashboard	Een interactieve pagina voor het weergeven van datavisualisaties.
Efficiëntie	De verhouding tussen de energie van de zonnestraling en de energie opgewekt door het zonnepaneel.
Energie	Aantal watt per vierkante meter.
Gebruiker	De persoon die interacteert met het dashboard.
Lichtinstraling	De hoeveelheid zonnestraling die de zonnepanelen treft uitgedrukt in W/m ² .
Missende data	Ontbrekende metingen in de dataset.
Model	Een model dat gebruikmaakt van gegevens om voorspellingen of inzichten te genereren.
Open source	Software waarvan de source code beschikbaar is en welke vrij te gebruiken is, zonder aanvullende restricties.
Platform	Het platform vanaf waar de applicatie wordt benaderd, zoals bijvoorbeeld Windows, MacOS of Linux.
Ruwe data	De data die door de meetapparatuur wordt verzameld.
Spanning	De spanning in volt.
Stroomsterkte	De stroomsterkte in ampère.
Temperatuur	De temperatuur in graden Celsius.
Webbased	Via een webpagina.
Weerstation	Meetstation voor het meten van diverse weersvariabelen.
Zonnecel	Een paneel wat elektrische energie opwekt aan de hand van zonnestraling.