

## Datascience

Portefolio

**Auteur:**

A.A Spaninks - 18169368

**Opleiding:**

Elektrotechniek, industriële automatisering

**Aanmaak datum:**

18 december 2019, Tilburg

**Instituut:**

De Haagse Hogeschool  
Academie voor toegepaste wetenschap  
Johanna westerdijkplein 75  
2521 EN, Den Haag

**Docent:**

Jeroen Vurens

**Opdrachtgever:**

Arie Taal

## Inhoud

1	Introductie .....	3
2	Datacamp.....	4
3	Domain knowledge.....	5
4	Data preprocessing .....	6
5	Communicaties .....	9
6	Contributies .....	10
7	Leerdoelen .....	12
8	Evaluatie project.....	13

## 1 Introductie

Dit is het verslag van de student Aurin Spaninks. In dit verslag wordt door de student geschreven over zijn prestaties tijdens de minor. De onderdelen die in het verslag voorkomen hebben betrekking tot het project en opdrachten die tijdens de minor zijn uitgevoerd. De onderdelen die de student heeft uitgevoerd en benoemd worden zijn: Domain knowledge, Data preprocessing en communicatie. Aan het einde van het verslag reflecteert de student op zijn prestaties en het geleverde werk. De onderdelen waar de student over gaat reflecteren zijn: de bijdragen aan het project, de gestelde leerdoelen en evaluatie van het project. De STARR-methode is gebruikt voor het reflecteren.

## 2 Datacamp

Voor de minor moeten de studenten de assignments maken op datacamp. In figuur 4 zijn de resultaten te zien van de student.

**Figuur 4**

My Assignments					
ASSIGNMENTS					
NAME	ASSIGNED AT	DUE BY	TYPE		
Intermediate Python	Sep 3, 2019	Sep 16, 2019, 09:00 CEST	Complete Course	Completed	>
Customizing plots	Sep 3, 2019	Sep 16, 2019, 09:00 CEST	Complete Chapter	Completed	>
Introduction and flat files	Sep 3, 2019	Sep 16, 2019, 09:00 CEST	Complete Chapter	Completed	>
Writing your own functions	Sep 3, 2019	Sep 23, 2019, 09:00 CEST	Complete Chapter	Completed	>
pandas Foundations	Sep 3, 2019	Sep 23, 2019, 09:00 CEST	Complete Course	Completed	>
Python Data Science Toolbox (Part 2)	Sep 3, 2019	Sep 30, 2019, 09:00 CEST	Complete Course	Completed	>
Plotting 2D arrays	Sep 3, 2019	Sep 30, 2019, 09:00 CEST	Complete Chapter	Completed	>
Statistical plots with Seaborn	Sep 3, 2019	Sep 30, 2019, 09:00 CEST	Complete Chapter	Completed	>
Cleaning Data in Python	Sep 3, 2019	Oct 7, 2019, 09:00 CEST	Complete Course	Completed	>
Statistical Thinking in Python (Part 1)	Sep 3, 2019	Oct 14, 2019, 09:00 CEST	Complete Course	Completed	>
Introduction to Python	Sep 3, 2019	Sep 9, 2019, 09:00 CEST	Complete Course	Completed	>

### 3 Domain knowledge

De energie transitie een belangrijk onderwerp om de klimaatveranderingen tegen te gaan. Energie transitie is dat gebruikelijk energiebronnen worden veranderd groene alternatieven zoals bijvoorbeeld: windturbines, zonnepanelen en geothermie. Dit betekent dat windturbines en zonnepanelen een belangrijke rol zullen spelen om elektrische energie op te wekken. Geothermie wordt gebruikt voor het verwarmen van de gebouwen doormiddel van aardwarmte. Geothermie wordt niet gebruikt voor het project maar hoort wel bij de energy transitie vanwege dat energy transitie niet alleen over het opwekken van elektrische energie gaat.

De student heeft meerdere bronnen gebruikt om meer inzicht te brengen in het onderwerp. Als eerste moet er een model gekozen om te gebruiken voor het zoeken naar de optimale configuratie. Een voorstel was om een genetisch algoritme te gebruiken. De persoon die het voorstelde heeft veel ervaring met dit soort problemen maar kon niet het bewijs leveren om de keuze te veroorloven. Als gevolg heeft de student bronnen gezocht om het argument te bekrachtigen. De student vond twee papers waarbij genetisch algoritmes worden gebruikt om de optimale configuratie te vinden van een Hybride zon/wind energie centrale. De papers zijn wel uitgebreider ze kijken naar meer uitgangswaarden en kijken ook naar de levensduur en onderhoud van componenten. Hieronder staat de referenties van [paper 1](#) en [paper 2](#).

- Hongxing Yanga,\*, Wei Zhoua, Lin Lua, Zhaohong Fangb, (19 September 2007), *Optimal sizing method for stand-alone hybrid solar–wind system with LPSP technology by using genetic algorithm*
- DhakerAbbesa, AndréMartineza, GérardChampenois (2 June 2013) *Lifecycle cost, embodied energy and loss of power supply probability for the optimal design of hybrid power systems*

De student heeft een [document](#) gemaakt om meer inzicht te geven over de werking van de zonnepanelen en windturbines. Dit document is gemaakt aan het begin van de minor om meer inzicht te creëren voor alle project deelnemers. De informatie die in het document staan zijn op de volgende bronnen gebaseerd:

- Vijayamohanan Pillai N, *Loss of Load Probability of a Power System*, Available at: <https://www.omicsonline.org/open-access/loss-of-load-probability-of-a-power-system-2090-4541-5-149.php?aid=35881> (Accessed: 13-01-2020).
- Nederlands Elektrotechnische Comité (Oktober 2015) *Nen 1010 - elektrische installaties voor laagspanning*

## 4 Data preprocessing

Het project heeft geen beschikking over een dataset om te gebruiken voor analyses. Er wordt namelijk een simulatie gebruikt voor het berekenen van energie. De student heeft de data gebruikt van de trainingen en simulatie.

De student heeft een dataset gemaakt die te zien is in figuur 2. De dataset bestaat uit configuraties die het resultaat aantonen van trainingen. Er zijn twee configuraties die de zonnepanelen op nul graden zet met een hogere prijs alle andere configuraties. Daarnaast is er een configuratie met nul windmolens en heeft ook een hogere prijs. De training is hoogstwaarschijnlijk in een lokaal minimum belandt voor deze specifieke trainingen.

De configuraties kunnen worden beschouwd als outliers als men de dataset niet wil gebruiken voor analyses om het algoritme te verbeteren. Is dit wel het geval dan kan het model verbeterd worden door meer data te generen met verschillende mutatiefactoren. Als er minder van deze outliers voor komen bij een ander mutatiefactor dan is het model verbeterd.

**Figuur 2**

	kosten	area zonnepanelen	Angle	org	kWh accu's	# windmolens	Cable area
1	481467430	817458	25	0	654695	20	49978
2	481530536	810540	26	0	648139	20	49866
3	481465812	817836	25	0	654512	20	49982
4	492024700	971776	0	0	607787	20	49999
5	481279408	812235	28	5	656706	20	49958
6	481574632	807266	27	0	659804	20	49828
7	481615111	804825	27	0	661065	20	49740
8	481295622	811137	26	9	657268	20	49567
9	492024657	971801	0	0	607775	20	49999
10	481566342	807899	27	0	659492	20	49786
11	1125517947	1158439	27	0	2263356	0	49954
12	481492769	812934	26	0	656907	20	49951
13	481557127	808333	27	0	659254	20	49863

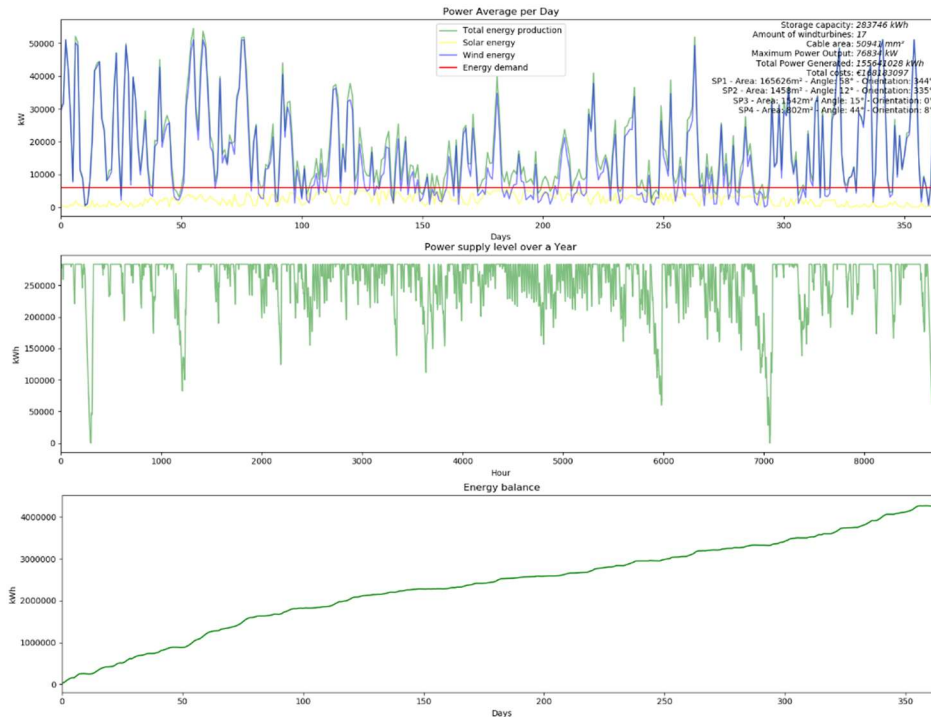
De student heeft ook gekeken hoe de parameters van de trainingen invloed hebben op de resultaten van de trainingen. Bijvoorbeeld in figuur 3 is een tabel met verschillende configuraties. In de blauwe kolom staan de verschillende accu prijzen en in de gele de grote van de accu. De accu prijs heeft invloed op het aantal accu's maar ook hoeveel zonnepanelen er worden neergezet als het aantal windturbines niet mogen veranderen. Bij de laatste configuratie valt het op dat de zonnepanelen op nul graden worden gezet. Dit is doordat zonnepanelen op nul graden meer energie opwekken aan de begin en einde van de dag. Het blijkt dus dat het efficiënter is om dan meer zonnepanelen neer te zetten op een inefficiënte hoek om zo meer energie op te wekken over de volledige dag.

**Figuur 3**

Wind Turbine			Solar panels								Accu	
			SP1			SP2						
Type	Aantal	Hoogte (m)	Terreingest.	Rendement	Area(m²)	Angle(degree)	Orientation(deg)	Area(m²)	Angle(degr	Orientation	€/kWh	kWh
4	7	135	0,12	15%	212492	47	0	0	0	0	0.2	4852090
4	7	135	0,12	15%	220619	8	0	0	0	0	5	5506982
4	7	135	0,12	15%	330611	63	0	0	0	0	25	2380840
4	7	135	0,12	15%	380875	55	0	0	0	0	50	2198149
4	7	135	0,12	15%	529695	53	0	0	0	0	100	1661501
4	7	135	0,12	15%	818851	38	0	0	0	0	200	1275773
4	7	135	0,12	15%	1640560	37	0	20785	0	0	400	738326
4	7	135	0,12	15%	1826655	27	0	0	0	0	800	684804
4	7	135	0,12	15%	2745704	18	0	7805	0	0	1600	549701
4	7	135	0,12	15%	10000000	0	0	8044321	0	0	100000	103982

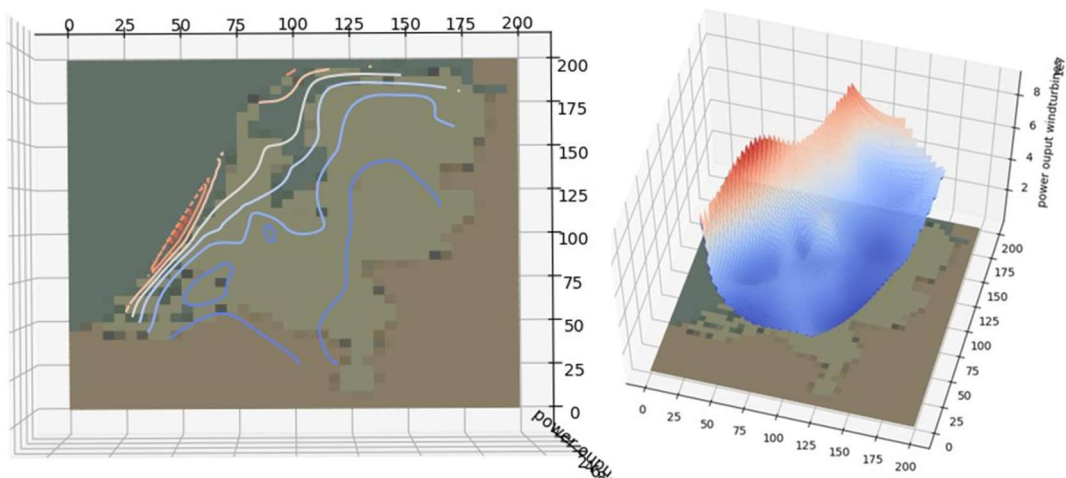
De student heeft ook data gevisualiseerd van de configuraties. Zo is in figuur 4 een voorbeeld een configuratie te zien. In figuur 4 is te zien de configuraties parameters, de energieproductie, de energiebalans en het batterij gebruik over een jaar.

**Figuur 4**

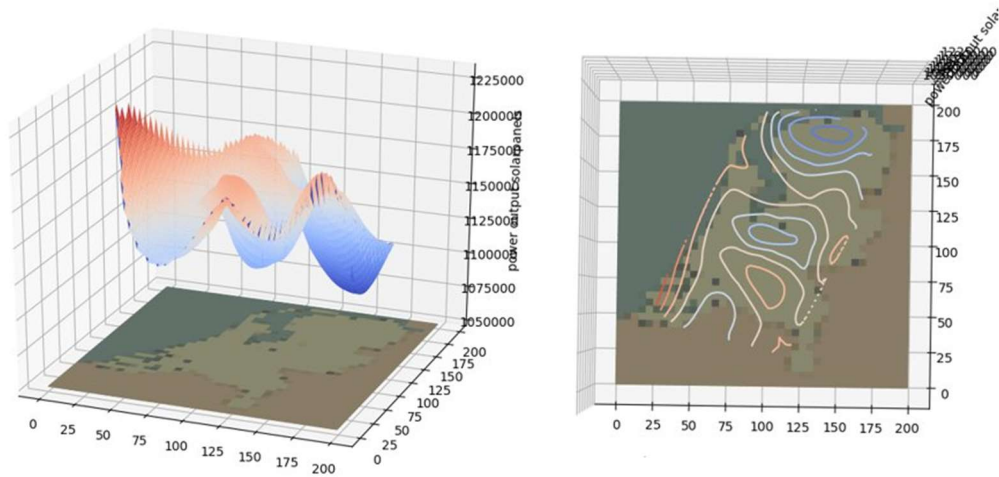


Daarnaast heeft de student ook een dataset gecreëerd met simulatie van een specifieke configuratie. Zo heeft de student met een projectgenoot de data gebruikt van alle weerstations die verspreid staan door Nederland om de energieproductie te visualiseren van zon en wind voor iedere locatie. In figuur 5 is de verspreiding windenergie te vinden en figuur 6 is verspreiding van zonne-energie te vinden.

**Figuur 5**



**Figuur 6**



De student heeft de code aangepast van een projectdeelnemer om de zonne-energie data te generen. Daarnaast heeft de student geholpen bij het schrijven van code voor de visualisaties en de interpolatie functies met een projectdeelnemer (figuur 5 en figuur 6). De student heeft ook de code gemaakt voor de visualisatie van een configuratie (figuur 4). De code is te vinden in de volgende links

- [Code 1](#)
- [Code 2](#)



## 5 Communicaties

The student heeft meerdere keren een presentatie voorbereid. Daarnaast heeft de student ook een presentatie gegeven aan de opdrachtgever: Arie Taal. In de volgende links staan een paar voorbeelden van een presentatie die voorbereid is of gemaakt door de student.

- [Presentatie 1](#)
- [Presentatie 2](#)

De student heeft specifieke onderdelen geschreven van de paper. De onderdelen zijn: "Windturbines" en "Cable area calculatie". De student heeft ook kleine aanpassingen of toevoegingen gemaakt aan de "introduction" en "Conclusion". De projectgroep heeft gezamenlijk de paper gecontroleerd en aangepast. De paper is te vinden in de volgende link.

## 6 Contributies

De projectgroep bestaat uit vijf projectleden die allemaal een andere achtergrond hebben. Dit is gunstig voor de projectgroep om dat er beschikking is over verschillende expertises. De projectgroep heeft namelijk beschikking over: één mechatronica engineer, één software engineer, twee netwerk en system engineers en één elektro engineer. Dit is een ideale groep om de optimale configuratie van een optimale configuratie van een hybride zon/wind energiecentrale te vinden.

Mijn taak binnen het project was om informatieve en betrouwbare bronnen te vinden om als groep de juiste beslissingen te maken in het project, vragen te beantwoorden omtrent elektrotechniek en het verduidelijken van het probleem met energietransitie. Over het algemeen het literatuuronderzoek uitvoeren en het vergelijken bestuderen van verschillende configuraties van de simulatie die beïnvloed worden door parameters van het genetische algoritme. Er zijn ook nog wat kleinere activiteiten geweest zoals het schrijven van een paper en het geven van korte presentaties aan de opdrachtgever.

Een van de projectleden had al veel ervaring met dit soort probleemstellingen en had al snel de oplossing gevonden om de optimale configuratie te vinden namelijk het gebruik van genetische algoritmes. Er was op dat moment nog geen bewijs dat dit de juiste methode was om de optimale configuratie te vinden. Mijn eerste actie was het zoeken naar literatuurbronnen die hetzelfde probleem wil oplossen voor het vinden van de optimale configuratie.

Eenmaal dat het genetische algoritme was gemaakt was mijn volgende actie om te kijken hoe de parameters van het genetische algoritme de optimale configuratie beïnvloeden. Hiervoor heb ik meerdere malen het algoritme uitgevoerd met verschillende parameters en de configuratie gesimuleerd om de configuratie te kunnen visualiseren.

Als laatste actie heb ik met een van mijn projectleden gekeken wat de energieopbrengst is van een voor geconfigureerde configuratie in verschillende locaties van Nederland. Dit is gevisualiseerd in een 3D plot om meer inzicht te krijgen in het verloop waar meer/minder energie wordt opgewekt.

Het resultaat is dat er papers zijn gevonden waarbij genetische algoritmes worden gebruikt om het de optimale configuratie te vinden dit is dan ook bewijs dat het mogelijk is. Echter is beschreven dat de meeste papers een andere soort genetisch algoritme om meerdere waardes van een configuratie zoals kosten, milieu en levensduur. Ons algoritme heeft dit niet nodig vanwege dat we alleen maar de kosten van belang zijn. Door meerdere malen het algoritme uit te voeren op verschillende parameters is er een spreadsheet gemaakt met alle verschillende optimale configuraties. Daarbij is er een visualisatie gemaakt die de energieopbrengst van zon/wind in een grafiek zet met daarbij de parameters van de configuratie.

Naar mijn mening heb ik het goed gedaan tijdens het project. De keuze om bewijs te leveren dat het gebruik van een genetische algoritme is goed. Hierdoor kunnen we veroorloven waarom we de keuze voor een genetisch algoritme een goede keuze was. De spreadsheet

van alle soorten verschillende configuraties had nog een keer moeten worden uitgevoerd na dat het genetisch algoritme was geoptimaliseerd. Als dit wel was gedaan had er meer inzicht kunnen gemaakt worden over de invloed van de parameters. Het beheren van mijn code had beter gekund. Ik had namelijk veel kleine scripts om data te verwerken of te visualiseren. Deze scripts raakte vaak verloren vanwege niet goed gebruik van de commit's waardoor ik mijn eigen scripts verwijderde. Voor de volgende keer moet ik mijn scripts apart opslaan in een map of meegeven aan de project folder op GitHub.

## 7 Leerdoelen

Tijdens het volgen van mijn opleiding is er een project voor “Predictive maintenance” uitgevoerd door mij. Dit project had beschikking over drie verschillende databases namelijk: een historian met sensor data, een database voor orders met machine statussen en de database voor onderhoudsbonnen. Het doel van het project was om onderhoud te kunnen voorspellen aan de hand van de data in de drie verschillende databases.

De uitvoering van het project was geen succes vanwege dat de projectleden geen ervaring met data science hebben. Het project had een hoge faalkans betreft de te kort komen van kennis, ijverigheid van de klant en geschikte tools voor de data-analyses. Vanwege dat het project met onvoldoende resultaat is afgesloten heb ik besloten om een data science minor te gaan volgen. Voordat de minor starten heb ik verschillende leerdoelen opgesteld betreft het maken en gebruiken van algoritmes, data verkenning en ook het gebruik van neural netwerken.

De acties die ik heb verricht waren bijvoorbeeld: onderzoeken verrichten op het gebied van elektriciteit voor domein kennis, Welke methodes gebruikelijk zijn voor de probleemstelling van het project, scripts schrijven in python, lessen volgen die te maken hebben met data science en algemene kennis over algoritmes en neurale netwerken.

Als resultaat heb ik geleerd dat je bij data science het heel belangrijk is om sceptisch te zijn met resultaten of aannames en als het bewijs wordt geleverd overtuigd worden van de feiten. Bijvoorbeeld in het project is er één project deelnemer die veel ervaring heeft met data science. De project deelnemer gaf al snel aan om genetische algoritmes gebruiken voor dit probleem. Dit is een aanname vanwege dat er geen bewijs was geleverd hiervoor. Ik ben mijzelf sceptisch gaan opstellen en heb een poging gedaan om bewijs te leveren voor deze aanname. Er papers gevonden met hetzelfde probleemstelling en wordt er gebruik gemaakt van genetische algoritmes. Met het bewijs gevonden Nu zijn de bewijs stukken gevonden om mij te overtuigen dat het gebruik van genetische algoritmes juist is.

Een ander voorbeeld is toen we de resultaten hadden gekregen van de beste configuratie voor een energiecentrale configuratie. De grote vraag die gesteld werd was: “Is dit wel de juiste configuratie?”. Hieruit leiden meerdere vragen zoals: “Waarom verschillen de hoek van de zonnepanelen veel en de oriëntaties bijna niets?”. “Waarom meer windmolens dan zonnepanelen?”. Door meerder configuraties te laten trainen met het genetische algoritme en extra data uit de simulatie te generen voor analyses is er meer inzicht gekomen over wat de juiste configuratie is. Hierbij hebben we ons zelf nogmaals sceptisch opgesteld om zeker te zijn dat de configuraties die we kregen uit de trainingen juist waren.

Het leren voor het gebruiken en programmeren van neurale netwerken is niet aanbod gekomen. Er was geen ruimte in het project om neurale netwerken toe te passen. Dit is immers niet erg sinds dat er een ander project is dat “Solar Spots” wil voorspellen die op de zon te zien zijn. Hiermee is er een mogelijkheid om een poging te nemen voor het gebruik van neurale netwerken.

## 8 Evaluatie project

De energie transitie is een belangrijk onderwerp om de klimaatveranderingen tegen te gaan. Het bedrijf Danone wil een grote stap nemen in een duurzamere wereld door het elektrische energie verbruik van een fabriek volledig voorzien van duurzame energie. Hiervoor wil het bedrijf een energiecentrale van windturbines en PV-panelen neerzetten om de stap naar een duurzamere wereld te realiseren. Het bedrijf wil weten wat de ideale configuratie van PV-panelen, windturbines en accu's voor de energiecentrale. Een project is opgezet om de optimale configuratie van een hybride zonne/wind energiecentrale te vinden.

De taak van de projectgroep is om de optimale configuratie te vinden. De projectgroep moet een onderzoek uitvoeren over het domein van de opdracht en de gebruikelijke methodes. Het resultaat dat de projectgroep moet leveren is een configuratie van zonnepanelen, windturbines en accu's. De configuratie bestaat uit vier zonnepanelen velden met een hoek en oriëntatie richting de zon in graden. Ook heeft een configuratie een aantal windturbines en accu's. De projectgroep moet wekelijks rapporteren van hun prestaties en als slot een paper schrijven over de resultaten.

De projectgroep heeft als eerste een plan van aanpak gemaakt om de probleemstelling, onderzoeksvragen en planning vast te leggen. Vervolgens werd er al snel een beslissing gemaakt om een genetisch algoritme te gebruiken om het probleem op te lossen. De beslissing was bepaald op de expertise van de een project deelnemer. Dit veroorlooft de keuze van een genetisch algoritme niet en is er een literatuuronderzoek opgestart om dit kunnen bevestigen. De projectgroep heeft op basis het literatuuronderzoek een applicatie gemaakt die de optimale configuratie kan vinden. De projectgroep heeft ook activiteiten gedaan buiten het project om. Bijvoorbeeld de intreedende gestaan om het project te vertegenwoordigen en een meegedaan aan een Techathon om in een dag een plan op te stellen om een wijk te verduurzamen.

Als resultaat heeft de projectgroep met het literatuuronderzoek meer inzicht gekregen over elektriciteit en zijn er papers gevonden met dezelfde probleemstelling. De papers omschrijven dat er een genetisch algoritme is gebruikt bij het vinden van de optimale configuratie. De projectgroep heeft met deze informatie een genetisch algoritme gemaakt die de configuratie gaat vinden. Als extra heeft de projectgroep weersdata gebruikt van meerdere weerstations om op verschillende plekken de optimale configuratie te vinden. Hierbij hebben ze gevonden dat de locatie van zeeland de beste locatie is om een duurzame energiecentrale te zetten.

De projectgroep heeft naar mijn mening de goede stappen en keuzes genomen om uit eindelijk op het juiste antwoordt te komen. De volgorde van het voorstel om de een genetisch algoritme te gebruiken is niet de correcte manier. Dit is echter wel goed opgelost doordat er een literatuuronderzoek is uitgevoerd om te onderzoeken wat gebruikelijk is. Een verbeterpunt van het project kan zijn het echt valideren van het model. De projectgroep had niet de kennis en vaardigheden om het resultaat van het genetisch algoritme te bevestigen. Dit komt hoogstwaarschijnlijk door het gebrek aan uitzoek werk om het te valideren. Echter als we het met een ander model moeten valideren had dit veel tijd in beslag genomen dat

misschien geen resultaat had opgeleverd. Tot slot is het project op een goed is afgerond met een mooi resultaat.