Please upload your first draft of your final report here. Unless your client specifies otherwise, write in English and use the following format.

1. Cover page and table of contents
2. High-level introduction to the problem and product vision (500–1500 words)
3. Mid-to-high-level description of proposed solution and developed product (1000–2000 words)
4. Low-level documentation of implementation (no word limit; may also be an appendix or a link to an online repository)
5. High-level conclusion that summarises the product capabilities and includes ideas for future improvement (500–1500 words)

If your client *does* specify otherwise, please use the format your client requests. Your report should still be 2000–5000 words, however, excluding the product documentation.

Let op: er staat dan wel leuk tussen de 2000-5000 woorden maar voor de volle punten moeten we er minstens 3500 hebben.

Table of Contents

Iedereen leest de tekst door **2**
1 | Introduction (500-1500)(569) **3**
2 | Solution (1000-2000) **5**
R 5
Recommendation for the measurements 5
Proposed Solution 7
3 | Implementation (Mees) **7**
4 | Conclusion (500-1500) (Jannes) **8**
5 | References **10**

Iedereen leest de tekst door

1 | Introduction (500-1500)(569)

One of the goals of horticultural production in greenhouses is to increase the sustainable income of the grower. The investment costs for conventional plant production as well as labour and energy costs are much lower comparing to the investment costs for greenhouses. This can only be balanced out with a better utilization of the yielding potential of plants, higher labour productivity and higher energy efficiency (*Tantau, H. J. (1991)*). Another goal of horticultural production in greenhouses has to do with the global population rapidly increasing, together with the demand for healthy fresh food. The horticulture industry can play an important role in providing food, but encounters difficulties finding skilled staff to manage crop production (*Hemming et al. (2019*)).

In the competitive horticulture industry, small improvements can make or break the competitive advantage of the yield. In order to stay ‘ahead of the curve’, calculated decisions need to be made about the greenhouse climates (*Sigrow, n.d.*).

Sigrow is a company in the Netherlands which provides state of the art sensor technology tailored to a greenhouse. These sensors make it possible for Sigrow’s clients to stay ahead of the curve and make those calculated decisions. Since the horticulture industry is very competitive, little is known about the possibilities to use artificial intelligence to optimize plant growth in greenhouses. Sigrow has little knowledge about artificial intelligence as well and that is why a collaboration of the two fields was needed.

Over the years, Sigrow has gathered information about the environment in the greenhouses, such as the temperature, air humidity, light intensity and amount of carbon dioxide (CO2) in the air. These four features are the most important factors for plants to grow, apart from water and nutrition (*Hemming et al. (2019*)). The task Sigrow had for us was to optimize the plant growth by optimizing the greenhouse climate using Artificial Intelligence (AI). This task was dubbed the ‘Sigrow Challenge’.

When Sigrow provided the data there were two significant problems:

1. The data lacked information about the plants.
2. The amount of data was scarce.

Sigrow has only collected data about the environment of the plants, but not the plants themselves. The data did not contain any information about which kind of plant it was, nor the height, stem diameter or weight. In addition to this, the amount of data was so scarce that it was impossible to split the dataset into a training and test set.

Because of these two problems, a new challenge was set up. Instead of optimizing the plant growth by optimizing the greenhouse climate, Sigrow wanted a recommendation. This recommendation had to meet a number of requirements. First, it had to contain a comprehensive analysis about the measurements Sigrow is currently taking. This includes advice about which measurements need to be done and why. Second, it had to contain a simple overview of possible feature optimization techniques which Sigrow could use, imagining the two problems were not there. At last, the recommendation had to contain an understandable explanation about several machine learning techniques.

Besides the recommendation, Sigrow also asked for a implementation design for these feature optimization techniques. Since the data is not usable, the results would be unreliable, but an implementation, imagining the data is complete, would be.

By providing Sigrow a recommendation about how to optimize the plant growth using machine learning and implementing these techniques for them, we hope Sigrow can optimize the plant growth in the future and stay ahead of the curve.

2 | Solution (1000-2000)

- intro
- Goede dataset fixen & preprocessen (The measurements)(Otto)
- Proposed solution (in cooperation with CEO): (Riemer)
    - 1. Recommendation
    - 2. Implementation of several proposed techniques
- outro

In the introduction, two main problems were explained. Since the first problem was insurmountable to solve in four weeks time, it was decided to search for an external dataset. This dataset had to meet two requirements. The first requirement was that some of the features in the external dataset had to meet some of the features of Sigrow's dataset. This way it was possible to come up with reliable solutions to optimize the plant growth and give Sigrow the right recommendations. The second requirement for the external dataset was that it had to contain enough information about the plant growth and the greenhouse climate that it can be splitted into a training and test set. In the end, the external dataset of the autonomous greenhouse challenge was used.

This dataset contained information about cucumbers and contained features such as temperature, air humidity, light intensity and the amount of CO_2 in the air, all similar to Sigrow's features. In addition to Sigrow's dataset, information about the weight of the cucumbers and the number of cucumbers was available in this external dataset.

This section is divided into three parts. First the recommendation about the measurements will be discussed. Second the recommendation about the feature optimization techniques and last some artificial intelligence methods for Sigrow to look into.

R

Recommendation for the measurements

There are multiple features which could be measured to get more data regarding the plants. These features consist for example of the height, stem diameter, number of leaves or weight. The more features there are regarding the plants, the better the results.

To optimize the plant growth, output features are essential. If Sigrow can measure the aforementioned features, it is possible to use artificial intelligence. Artificial intelligence consists mostly of using data to make accurate predictions. This means that not only the output features are important, but the amount of data is significant as well. The more data there is, the more accurate artificial intelligence can make predictions.

If Sigrow wants to optimize the plant growth in order to optimize production, it is also necessary to measure the production. Sigrow can use a similar approach as in the autonomous greenhouse challenge, in which the cucumber production is measured by dividing the cucumbers in three classes. The first class is A, which contains the cucumbers with a weight above 375 grams and has no defects. Class B contains the cucumbers with a weight between 300-374 grams or has some defects. At last class C contains all the cucumbers with a weight below 300 grams. Ofcourse, in order to successfully classify the plants, each kind of plant has its own threshold values. Below is an overview of the thresholds for the cucumber distribution.

Cucumber distribution among three different classes, like with the greenhouse challenge:

Class	Weight	Defects
A	> 375 grams	None
B	300-374 grams	At most a few
C	< 300 grams	No limit

These are all the features Sigrow uses. This is not part of the recommendation but it is handy for analyzing.

Number	Feature name	Explanation	Measuring unit
1	humid	Air humidity	%
2	par	Par (light)	µmol/m2/s
3	temp	Air temperature	°C
4	soil_temp	Soil temperature	°C
5	soil_hum	Soil humidity	%
6	soil_ec	Soil bulk	mS/m
7	rssi	Radio strength	-
8	batt	Battery	mV
9	co2	CO2	ppm
10	par_top	PAR top	µmol/m2/s
11	par_bottom	PAR bottom	µmol/m2/s
12	rad_top	Radiation top	W/m2
13	rad_bottom	Radiation bottom	W/m2

14	dli	Day Light Integral	mol/m2
15	pore_ec	Soil Pore Ec	mS/cm
16	soil_bec_cm	S Bulk EC	mS/cm
17	soil_temp_f	Soil temperature F	°F
18	tmp_f	Air temperature F	°F
19	g_water	Absolute humidity	g/m3
20	t_leaf	Leaf temperature	°C
21	vpd	Vapour Pressure Deficit	KPa
22	weight_t	Weight total	g
23	weight_1	Weight 1	g
24	weight_2	Weight 2	g
25	weight_3	Weight 3	g
26	weight_4	Weight 4	g
27	wli	Week Light Integral	mol/m2
28	vbat	Battery Percentage	%
29	flow_in	Flow in	liters
30	flow_out	Flow out	liters

Proposed Solution

3 | Implementation (Mees)

4 | Conclusion (500-1500) (Jannes)

The final product that was delivered was two fold. The first part consisted of researching and explaining Artificial Intelligence techniques that might prove useful to Sigrow in the future. The techniques that were discussed in this report were polynomial fitting, support vector regression, time series forecasting and deep deterministic policy gradient. These techniques themselves can be divided into two groups. The first group being the techniques that can be executed after a growth cycle of the crops. They can be used to determine in hindsight what were the best values for some selected features like humidity, light and CO2 levels. These techniques are support vector regression and polynomial fitting. The second group of techniques include time series forecasting and deep deterministic policy gradient. These are techniques that can be used while the crops are growing to determine the best possible action one can take based on the current environment of the crops. The second part of the final product was determining the optimal values of the selected features. The results of this implementation can be seen previously. Below the individual conclusions for the examined algorithms can be seen.

The first technique that was discussed was support vector regression.

Like Polynomial Regression, this technique is relatively easy to implement. It does not require a lot of data and/or runtime to acquire results. However, the results can be unstable or less informative. A regression approach to the Sigrow Challenge is a trade-off of difficulty and quality. SVR methods can perform better with a lot of data.

The second technique that was discussed was polynomial regression.

Polynomial regression is a rather easy way of modeling data. Not much computing power, advanced know-how and time is needed to gather results from this implementation. For Sigrow this could be a reasonable option. The acquired datasets that we have now do not contain much information about the plants. It is difficult to implement certain techniques that require vast amounts of such data. The regression technique can acquire results with limited data. These results can be unstable, but they can be of use nevertheless. Therefore polynomial regression is a solid pick for working with such limited datasets.

Another technique that was discussed was time series forecasting.
CONCLUSIE TIMESERIES

The last technique that was discussed was deep deterministic policy gradient. Reinforcement learning is a very computing expensive technique since significant computing power is needed. However, the main problem for Sigrow is the amount of training data needed. Deep reinforcement learning would be an excellent technique for learning the optimal environment for plant growth, assuming the changes reported about the measurements are made. For the Sigrow challenge the agent would need to find the correct action, setting the humidity at the correct level for example, in the current state, the observed environment.

The conclusion derived from our implementation is as follows:
CONCLUSION PROGRAMMEER WERK

When evaluating everything together the most important conclusion is that data is one of the most important factors in machine learning and artificial intelligence by extension. The main hurdle that was encountered was the lack of usable data. Future research should focus on obtaining reliable data. One way this was solved in a different situation was using generative adversarial networks [2] to create and discriminate fake data. This might be of use as well in the Sigrow challenge.

5 | References

[1] Alhnaity, B., Pearson, S., Leontidis, G., & Kollias, S. (2019). Using Deep Learning to Predict Plant Growth and Yield in Greenhouse Environments. *arXiv preprint arXiv:1907.00624*.

[2] Hemming, S., de Zwart, F., Elings, A., Righini, I., & Petropoulou, A. (2019). Remote control of greenhouse vegetable production with artificial intelligence—greenhouse climate, irrigation, and crop production. *Sensors*, *19*(8), 1807.

[3] Tantau, H. J. (1991). Optimal control for plant production in greenhouses. *IFAC Proceedings Volumes*, *24*(11), 1-6.