Competence Document



University of Applied Sciences

Maurits Krijnen

Version 1.0

Date: 20/06/22

Table of Contents

1.	Persona	4
	1.1 Introduction	4
	About me	4
	From web design to application development	4
	Fontys	4
	1.2 Skills and Challenges	5
	Substantive knowledge	5
	Professional skills	5
2.	Context	6
	Learning Goals for this semester	6
	In which context would you like to work?	6
	Conclusion	6
3.	Project / Challenge description	7
	Project Eye disease detection using ML image recognition:	7
	Machine Learning with Self reinforcement learning	7
4.	Competence profile	8
	Current profile	8
	Intended development	9
	Intended development	9
	Intended development Description:	9
	Final development	9
5.	KPI-table with proof	.10
6.	Sprint retrospectives	.11
	Datasets	.11
	Design Documentation	.11
	Finishing Documentation	.11
	Machine Learning start	12
	Stakeholder meetings	.12
	Freaky Friday	.12
	Model issues	.12
	Freaky Friday	13

Switching Models	13
More Delays	13
8. Feedpulse	16

1. Persona

1.1 Introduction

About me

My name is Maurits Krijnen, 23 years old currently living in Deurne. I usually spend my free time playing video games with friends, playing guitar, or coding a personal project. I have a background of VMBO-T following 2 mbo-4 courses. The first was a financial course that I didn't really enjoy but did finish, after which I decided to switch to media/application development where I discovered my enjoyment of programming.

From web design to application development

After finishing my MBO internship at a 5-man website building company, my preference switched towards wanting to work at more 'technical' IT projects. I felt that the realisation challenges in web development were not fulfilling or stimulate me creatively.

I do still practice with web in the form of MVC languages such as Ruby on Rails(RoR), but this is usually as a frontend for projects such as my Travelling Salesman Web application.

Fontys

Semester 1 & 2

During semester 1 I used my web knowledge to help the new students during the proftaak. Making and hosting a website with Laravel. In semester 2 I made a mobile app with react native, which I had not attempted before. For my solo projects I created a Tetris game in C# for semester 1. And in semester 2 I made created a web app using the public API from League of Legends.

I debated between choosing media or software but chose software in the end.

This decision was made based on that I get the more motivation and creativity when trying to solve complex problems related to data, compared to trying to design a website/game.

Semester 3

For semester 3 I went with an Open Learning semester. I did a group project for canvas which aimed at creating a new (possibly automated) feedback system. In this project I created documentation and added new C# functions to an existing project.

For my personal (freaky Friday) project I worked on researching Machine Learning(ML) and created a Travelling Salesman application in ruby. I had some success with this project but had too little time to get it finished as I was only working 1 day a week on this.

Semester 4

For semester 4 I continued my research into ML by following an Artificial Intelligence specialisation semester. Here I learned a lot about how to use data and a methodology used for AI projects. While this was a very informative semester the technical realisation requirements of the semester were very low, and due to this I didn't learn what I originally signed up for.

Semester 5

I applied for an internship at ASML to help create a data visualisation dashboard for the CS productivity team, important goals were performance and its accessibility. I learned a lot of technical and communicative skills during my time there. I was also successful in applying the knowledge I gathered in semester 3 & 4 to create an efficient dashboard.

This semester

My goals for this semester are to continue my learning into Machine Learning or creating other data related IT solutions.

1.2 Skills and Challenges

Substantive knowledge

I am confident in my ability to use a wide variety of coding/programming languages. Having experience in C#, ASP.net, JavaScript, html/CSS, ruby, ruby on rails, php, Laravel, and recently python(and jupyter notebook/pandas). Having worked a lot in web development I also have experience with MySQL, SQLite3, and apache servers. I also worked with an Ubuntu OS for my programming which made me pickup on some Linux knowledge as well. Think of basic navigation, nano, git, SQL from command line etc.

Professional skills

From internships I have some experience in communicating with customers and shareholders and converting their project goals into easy to understand requirements. I have also become a lot more comfortable doing presentations over the years of going to different MBO schools

I have experience with UML documentation as I had to work with it in my MBO course and previous semesters. Examples of what I've worked with are Flowcharts/activity diagrams, use cases, class diagrams. Also included was having classes on doing Database normalization.

Organizing work/environment/team

Before my time at Fontys I had spent close to no time working in groups. But after the group projects and my internship at ASML I have more experience in organizing agile/scrum environments and executing group work activities.

2. Context

Learning Goals for this semester

I have been interested in creating algorithms driven by complex data and more specifically machine learning for a while. I have completed some projects related to this topic in semester 3 and 4 but these were mostly proof of concepts with no real application. For this semester I am looking towards creating an application with some real world application.

Topics I am interested in

IT solutions for handling data, algorithms, machine learning, OOP.

In which context would you like to work?

Health

I very much like the idea of using Machine Learning to create apps that improve the health/medical sector for both patients and personnel. I feel like a lot of advancements could be made in this area, a simple google search reveals a few machine learning projects that are already in use(https://builtin.com/artificial-intelligence/machine-learning-healthcare).

For me personally, the most interesting ones are those that try to diagnose or predict the development of a patient. Some examples are: kensci.com, pathai.com and Project Inner eye. These projects show that there is a growing market for these IT solutions, especially with world wide shortages of specialized medical personnel.

If an AI could take over the job of accurately discerning what is wrong with a patient, there would be fewer human errors thus increasing the quality of healthcare. Furthermore, lowering the work load of medical personnel will allow doctors to focus their attention on other matters, saving costs for the hospital in the process.

Data

What I learned about myself after my ASML internship is that I don't just find motivation in creating ML solutions, but in IT solutions handling data in general. I find Financial data or consumer data much less interesting then technical applications however. And find that this creativity is important for my personal motivation towards completing a project.

Conclusion

In my intake letter in semester 3 I was convinced that including the context in this semester was too far fetched. And that I would be focusing on a general proof of concept of machine learning.

For this semester I am confident in my ability to complete a ML solution with a real world health context. I already have a project in mind that I wish to discuss as well.

3. Project / Challenge description

Project Eye disease detection using ML image recognition:

Using Machine Learning to compare eye images and try to categorize if there are diseases present. Different data sets that can be used are found below.

Most of these use Fundus photography that are taken with a machine that is often available in both optician shops and hospitals.

https://en.wikipedia.org/wiki/Fundus photography

https://www.kaggle.com/donkeys/retinopathy-train-2015?select=trainLabels.csv

Machine Learning with Self reinforcement learning

Self-play is where the algorithm learns by playing against itself without requiring any direct supervision, has become the new weapon in modern Reinforcement Learning (RL) for achieving superhuman performance in practice.

My goal is to investigate the working of this concept and create a simple game with an ML AI that has learned from self play.

4. Competence profile

Current profile

	Managing	Analysing	Advising	Designing	Realising	Communi- cation	Learning	Judge- ment
User interaction								
Business processes								
Software								
Hardware								
Infrastructure								

Description

UI: Having worked directly with customers and web development, I can say with confidence that I can give advise and design User interaction at a level that is ready for the workplace. I have a fair amount of knowledge on how to manage and analyse what a user would want in such an app as well. As stated in my Persona I have a wide area of expertise making me capable of realising most requests myself.

Software: I can perform all software tasks on a basic level. Some examples are (Simple games using c# and ruby, data gathering applications using API's and showing results on a website(Ruby on Rails/MVC app) and creating a neural network for the traveling salesman problem without using ML libraries.

I can make apps in C, C#, Python, ruby, and JS and have worked on researching and creating Machine Learning solutions in my AI specialisation semester.

Business: Some knowledge on how to interact with customers and document what they want in some contexts.

Hardware: Close to no experience or knowledge on this subject.

Infra: Limited experience or knowledge on this subject. Some knowledge on setting up VM's/webservers and interacting with them. I also have some general knowledge of Linux by working on an ubuntu OS.

Intended development

Intended development

	Managing	Analysing	Advising	Designing	Realising	Communi- cation	Learning	Judge- ment
User interaction								
Business processes								
Software								
Hardware								
Infrastructure								

Intended development Description:

My intention is still to focus on developing myself in the designing and realising of software problems. For this reason, I want to work on Machine Learning since It is one of the more advanced software problems, that is also very relevant in IT today.

I also want to improve my User interaction designs as they are important for showing the ML results.

Final development

	Managing	Analysing	Advising	Designing	Realising	Communi- cation	Learning	Judge- ment
User interaction								
Business processes								
Software								
Hardware								
Infrastructure								

5. KPI-table with proof

Rating system: O = Outstanding G = Good S= Satisfactory U= Unsatisfactory

КРІ	Proof	Rating
TI-2.1	https://fhict.instructure.com/courses/10683/assignments/201492 https://fhict.instructure.com/courses/10683/assignments/201354	G G
Analysis-U2.3	https://fhict.instructure.com/courses/10683/assignments/201492	S
Advise-U2.2	https://fhict.instructure.com/courses/10683/assignments/201492	S
Advise-S3.1	https://fhict.instructure.com/courses/10683/assignments/201493	S
Analysis-S3.1	https://fhict.instructure.com/courses/10683/assignments/201630	G
FOO-2.1	https://fhict.instructure.com/courses/10683/assignments/201354	G

6. Sprint retrospectives

Sprint 1

Datasets

At the start of the semester I had already performed some research for datasets that could possibly be used for the Eye trainer project.

To summarise, I had gathered 4 datasets with different kinds of images and/or amounts of images. I had a short meeting with the stakeholder were I showed the datasets and asked for his opinion on how useful a diagnose of each would be. Afterwards I wrote a short document on this interview and prepared the information I had gathered for the sprint 1 delivery.

More on the datasets here <u>Canvas assignment Debriefing Dataset interview</u>

Design Documentation

After some initial 1 on 1 meetings with established I started with the design process. Here the goal was to create documentation that could form the basis for interaction between the developer and the stakeholders going forward. To achieve this I created a requirements document, a prototype/wireframe and then a class diagram based on the previous 2 documents. These documents can be found on my personal course.

As the documentation is not fully completed as of the sprint 1 delivery, we decided to finish the documents in week 1 of sprint 2 and confirm the quality with a stakeholder interview.

Sprint 2

Finishing Documentation

For week 1 I finished the documentation and held a short interview with the stakeholder as was planned. I implemented the feedback that was given shortly afterwards. Most of the feedback was related to unnecessary functionality that would not add anything to a normal user.

Also the project will go by the name of "DRP Detector" from now on, as the previous title was misleading according to the stakeholder.

More on the interview here <u>Canvas assignment Debriefing Dataset interview</u>

Machine Learning start

After the model was done I got started with a TensorFlow model tutorial. At the start of sprint 3 I will be attempting to convert the code into accepting the DRP dataset.

Stakeholder meetings

After some delays we setup a meeting between me, the supervising teacher and the stakeholder where we discussed the goals of the project. Here we discussed what functionality would benefit the Eye health care the most and how to achieve this. Some of the main goals we discussed were saving costs, faster processing of patients and less workload for specialised personnel.

The way of achieving this was to try and automate the diagnosing of Diabetic retinopathy and automatically redirecting the patient to the correct Health department. Usually these images need to be evaluated by medical eye professionals. But the creation process of the images can be done by regular personnel. This creates an opportunity to automate the first process and save time and costs.

Freaky Friday

On week 1 of sprint 2 I was able to find a side project I was interested in, creating a simple game that 2 ML models could train themselves on using a neural network. For this sprint I was only able create a basic version of the game itself but hope to get started with the model soon.

Sprint 3

Model issues

As the realisation process has fully started I've been focused on creating the ML model. However it's unfortunately been going slower than predicted. And with the initial delays on the technical documentation and Stakeholder meetings in sprint 1/2 the model is behind on schedule.

The main issue is that the TensorFlow only accepts 2D NumPy arrays of the same size. And the only way to import all images into the same list seems to be by using the Sckit-image collection library, as it converts the images into 2D(sometimes 3D) NumPy arrays. Despite this the list type is a child class of the NumPy array, making the TensorFlow model reject it with an error.

Later this sprint I realised this can be fixed by converting them into NumPy arrays, but this takes a long time to do even for low amounts of images. Despite the efficiency being low, this did allow me to generate the first successful model just before we had some days off.

The results of the model were concerning however, as only category 0 had any results. As it turned out, the dataset was "Biased". After some research it turns out this is a common problem in Machine Learning where 1 category is many times more present then others. I did some testing at the end of the sprint and discovered this was heavily present in our dataset. For sprint 4 the goal.

Freaky Friday

As the progress on the DRP detector model has been going slow I have spent little time developing the Self training project as my focus has shifted for this sprint. The goal is to resume the FF project after the model issues have been resolved.

Sprint 4

Switching Models

After talking about the model issues with some friends I suddenly came to a possible solution for the efficiency problem the model was having. While the **scikit**-image collection library had very limited functionality, the same group is also known for the **scikit**-learn library. I started to wonder if their ML models had any compatibility for the image collection library, and as it turns out, they do. Using a simple SVM model and changing the code a bit I got the model to accept the image collection without converting any data to NumPy arrays. Which decreased the model training time significantly, as the NumPy conversion was the part that took the longest

I had considered other libraries/models at the start, but never reconsidered them after the I got the tutorial TensorFlow model to work. And looking back on it I should've considered using scikit-learn models from the start, but at least its working now.

More Delays

While switching models was the right call to improve efficiency, this didn't improve the models performance. It took close to a week to implement the new model and on top of this I also had to miss out on meetings due to being sick. Once the may holiday came around I spent most of it with a fever and the week after didn't go much better.

During the last week of Sprint 4 discussed my progress with the supervising teacher and had a discussion on where to focus development during the last sprint of the semester.

Freaky Friday

Just like in sprint 3 I worked very little on my Friday project. Truthfully I had forgotten about it for quite a bit while I was trying to figure out the main project. I'll be attempting to get a simple prototype ready for sprint 5.

Sprint 5

Fixing the dataset bias and model Analysis

I managed to fix the data bias problem by sorting the images by categories into different folders, and then load equal amounts of all of them. This led to a fully working model that was able to predict all categories slightly better then a person guessing. While this is not near the level we want it to be its a significant improvement from before.

In order to better focus on where the model's problems lie, I wrote a short analysis paper on the model performance and dataset issues. It can be read here.

Freaky Friday and COVID

After I fixed the dataset bias I had planned to spent extra time creating the self learning machine learning model. Unfortunately shortly after I got started with the model I caught a bad case of COVID. While not bad enough to be hospitalized, this made it so that I was unable to work on the project for an entire week worth of time, which is unfortunate as I had planned for the final week(s) to be spent fixing documents and creating a simple dashboard for the DRP detector project.

After I recovered from covid I felt very lost with what I had made so far and decided to quickly create a different game to create the algorithm for. Initially I thought any game would do (ex I chose tic tac toe) but that wasn't the case. For example I learned that in order to get the algorithm to learn it needs a performance metric to know how well its performing. Tic tac toe only has 1 metric and that is win or lose. After realising this I tried some alternatives including a chess library that allowed me to quickly setup all te requirements. However this also was a problem as chess is too complicated for me to set a performance metric. Win or lose is once again 1 metric but board positions could also serve as a metric. However I'm not an avid chess player so I am unable to judge the performance of certain board positions.

In the end I wasn't able to complete the entirety of my freaky Friday project. However as I learned a lot from the research I did on how to solve various problems in Machine Learning I still see it as a successful project.

Repairing the documents

As the final week of the project had arrived and I was still down with COVID, there were some choices that had to be made. I contacted my teacher for a meeting as I had recovered enough to be able speak and work on the project.

We decided priority should be on improving the products that I had received feedback on. For example I had received feedback on the analysis document I wrote in week 1/2 of sprint 5 that I hadn't had the time to work on because of COVID.

This also meant that there would be no desktop application for the DRP detector project.

7. Evaluation and Reflection

Intro

When the semester started I had a very clear idea of my learning goals for the semester. When I saw there were no group projects available that served my learning goals I actively pursued creating my own group project. While this was all going well, I underestimated how much work it would take to create the final product in 1 semester while working alone.

I was asked to find other group members to join me for the project but I didn't really know anyone that was interested. And at the same time I was interested in working on all parts of the project myself. So I decided with the project forming deadline coming up fast to work on the project alone. I was still concerned with creating a Freaky Friday project but thought this could come later as I was very excited to get to working on the DRP detector.

Reflecting back

Looking back on the semester the project itself would likely have had a better final product if I had more group members or adjusted the scope of the project to working on it alone. And while I could not have foreseen the delays with falling ill twice this semester it does reveal the downsides of the tight schedule I was holding myself to. If the scope was adjusted from the start or if I had a group of students working on the project the impact of the delays would've been reduced.

Delays and illness

While my learning goals were ambitious and I was motivated to work on the project because of them there were multiple times that delays disrupted the workflow I was in. The first was when the model was having issues performing in sprint 3. I was focusing on researching the issues online since they seemed to have a cause that would be common for more AI projects. However as started getting ill around the spring break I lost track of my progress.

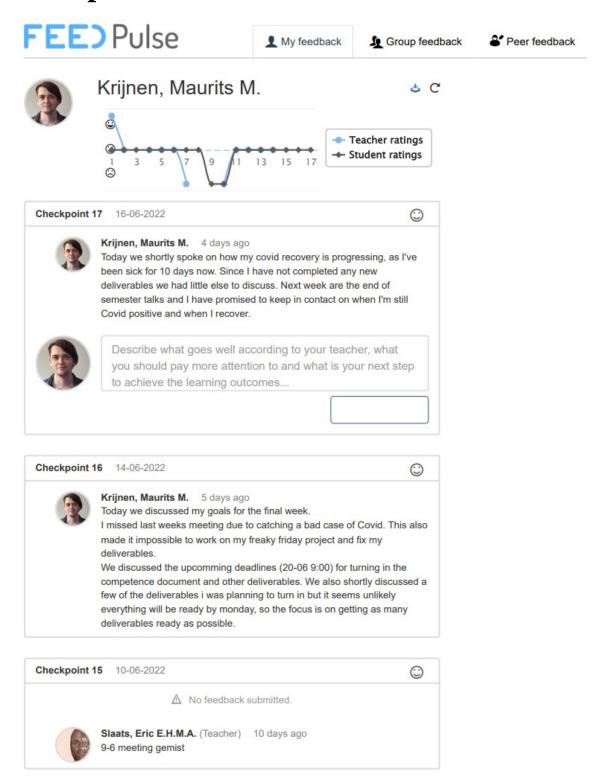
When I recovered from being sick I was lost on both the working of my code and the cause/solution to the problem I was trying to solve at the time. I also was behind on filling in feed-pulses and updating my personal course. While updating these areas the time since I had written the code increased and I started to remember less of what I was working on before.

A very similar situation happened when I fell ill with COVID in sprint 5 as it had been a while since I had worked on my Freaky Friday project. Leading to extended delays here as well.

Conclusion

In general I feel that while the delays could not have been prevented, the severity of the delays could've somewhat been reduced if I was more disciplined in keeping notes. I personally feel that I have learned and grown a lot as a machine learning student from all the research I've done during this semester. Even if the final products might not reflect this directly.

8. Feedpulse









Krijnen, Maurits M. 17 days ago

Had a discussion today where we spoke about my recent deliverables and did some reflecting on the semester. Since we're approaching the last few weeks of the semester I asked some questions about the final deadlines and turn in, apparently there is still an email coming that will depict the exact final day. It was recommended to me to have everything ready by Thursday however.

As I turned in my DRP dataset analysis document last week I received feedback on the document and we went over the various changes that should be made. We also discussed the progress of my freaky Friday project. I have promised to turn it in on Monday so we can make some iterative changes during the week.

I also updated my competence document and we spoke a bit about what I should add in the reflection section.



Slaats, Eric E.H.M.A. (Teacher) 6 days ago Ondanks veel uitval door ziekte toch booruitgang

Checkpoint 13 24-05-2022







Krijnen, Maurits M. a month ago

Today we discussed my progress on finishing the model and my deliverables in the short term.

To start we spoke about the evaluation document ive been working on. In this document I do analysis on the poor performance of the model and give some advice on how to proceed going forward. This document is planned to be turned in today.

After this I went over the conversation I had with the stakeholder last Friday. I discussed the models performance and the binary solution we discussed last week. The stakeholder mentions that a binary system does not give any value to the diagnosing of DRP and as such has been dismissed.

I also discussed the large amount of faulty images and asked if this was common in the field. Their response was that these faulty images are usually retaken and not submitted. This and other minor reasons have made me believe that the dataset currently in use is of lower quality then expected. If a better dataset was used the models performance would increase by quite a bit.

We also shortly discussed the added value of creating some research questions to aid me in improving the model. And the possibility of creating a research document on the findings.







Krijnen, Maurits M. a month ago

Today we discussed my plan on how to finish up the project in the coming weeks. There are about 4 weeks remaining and while the model is now working as intended, its performance isn't very good. We talked about various reasons that could be the cause for this such as the differences between category 1-2-3 being relatively small, leading to the idea of converting the labels to a binary system. We decided it would be for the best if I wrote an analysis/advice document on the topic.

Since the project deadline is approaching, I discussed the topic of whether to start focusing on the desktop application or continue improving the ML model. Eventually we decided that it should still be possible to do both with priority on improving the model since quality has priority over quantity in this case.

Checkpoint 11 12-05-2022







Krijnen, Maurits M. a month ago

Since the last meeting I've worked on various parts of the model to diagnose the problems it has and try to improve its results. We've discussed the progress and the current problems I'm working on specifically the recent one on the category 0 bias.

Some examples of progress that was discussed:

I created a new model with the sckit-learn library that work a lot more efficiently on how much time and RAM that is used while training the model.

Created an evaluation table using sckit-learn that shows a lot more information on the models performance. This confirmed my suspicion that large amount of category 0 present in the dataset causes the model to fail.

After the evaluation tool I started to do some research on how others fixed this problem (will create a assignment on this soon) and came to various websites (https://machinelearningmastery.com/what-is-imbalanced-classification/ &

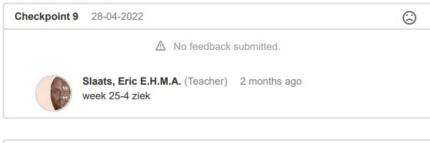
https://www.tensorflow.org/tutorials/structured_data/imbalanced_data). I've been trying to implement some of the suggestions but with no succes yet due to the examples being too different from my project.

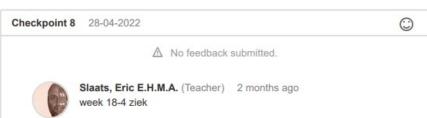
I said that my next goal was to try to create a dataset of even partitions per category. Since this would reduce the dataset to about 3500 images from 40.000, a suggestion was made to edit the images in different rotations to increase the data available.

Finally we spoke about the assignments in my personal course. I've not updated these in a while but promised to update them this week so they can be graded.

Since not a lot of time is left the goal is to finish with the model soon and get to working on the full application.









I was able to get the model to work with a test data set of 100 images and discussed some of the issues I was having. The main problem we discussed was the size of the numpy array that I wanted to load into the model was to large for my RAM. I was considering writing it to disk but a different solution is available at Fontys through the number cruncher at the delta location this would be a lot faster then doing it myself. I received the email address of Jelle Maas who I can contact for permission/reserving the machine.

We also discussed some other issues with the model. For example the accuracy of the initial test was 78%, but from my analysis this mostly due to the large amount of category 0 (no dr) in the data set. Most of the not 0 images were wrong, leading to a very low recall performance where those that are sick would be diagnosed wrong.

This is likely a common problem in these kinds of ML algorithms so I will do some research on solutions.

Finally we shortly spoke on my canvas course, since I've been working on the ML model in the notebook I haven't been as active on my course and will deliver my current progress instead of waiting till its done.



Checkpoint 6 04-04-2022







Krijnen, Maurits M. 2 months ago

Today we talked about the progress of creating a Proof of Concept for the ML model. I am working with a tensorflow tutorial to get a POC ready to convert into a model that can take in the fundus images.

We discussed some possible ways to improve the model with data cleaning and found an algorithm that can detect the location of veins in the

I also scheduled another meeting with the stakeholder to discuss these ideas and if he himself has any knowledge that might be helpful.

Checkpoint 5 28-03-2022







Krijnen, Maurits M. 3 months ago

Today we reviewed the progress of sprint 1 with the stakeholder. I held a short presentation on the sprint goals, what the progress for sprint 1 is and what goals I'm aiming for in sprint 2.

We also discussed the topic of holding user tests at the stakeholders location(they could do it themselves). And the possibility of the stakeholder retrieving more images for testing.

Checkpoint 4 24-03-2022







Krijnen, Maurits M. 3 months ago

We discussed the KPI's I've added and where to make some changes. We also spoke a bit on the general progress of the documents and training architecture of the AI, and what expectations there are for the next sprint demo. The stakeholder meeting planned for 18-03-2022 was moved to 28-03-2022.

Checkpoint 3 17-03-2022







A No feedback submitted.



Slaats, Eric E.H.M.A. (Teacher) a month ago no Entry this checkpoint

Checkpoint 2 10-03-2022







Krijnen, Maurits M. 3 months ago

Today we spoke about my progress on the technical documentation since the last meeting. We discussed the importance of creating Rubrics early on so that I can better define the rating of an assignment. This also helps with adding outcomes which I struggle with.

We also decided that setting sprint goals will be important for this project since I am working as the only developer. During our meetings we will review if I am staying focused on the sprint goals or if there is some adjustments to be made.

There is a stakeholder meeting planned for 11:30 18-03-2022. I have a short meeting planned with the stakeholder this evening, where I will gather some feedback on the project requirements and functionalities planned.

Checkpoint 1 24-02-2022







Krijnen, Maurits M. 4 months ago

We discussed the interview I had with the stakeholder on the datasets I found (see assignment debriefing interview). After this we decided on a direction for the project while brainstorming a bit for possibilities of increasing the scope of the project by adding more data or creating multiple networks. We planned a meeting with the stakeholder for the Monday after sprint 1 but this hasn't been confirmed yet.