CRITERIA TO BE GRADED (HAS TO BE DONE)

|  |
| --- |
| 1.     The contents of your personal portfolio reflect your contribution to the project, your abilities and what you have learned (y/n). |
| 2.     Your portfolio consists of materials that you either realized individually, or in case of a group effort, a clear statement of what your contribution is in this group effort (y/n). |
| 3.     The (digital) portfolio is written in a very easily accessible way (y/n). |
| 4.     The main document is a reader's guide (index) that shortly introduces your contributions and links to pages where the contributions are described in detail (y/n). |
| 5.     Every contribution should be accessible from the reader's guide in a single click (y/n). |
| 6.     The portfolio consists of links to the Python Notebooks or other evidence material about your contribution on the project that you have finished yourself. |

Datacamp

Notebooks

* Transfering data to csv (first approaches)
* K-nearest neighbors->KNearstNeighboors
* Helped detecting sprints->PredictSprintsTrial(found detect peaks function rest of the code was almost already done the for was by jake and I added regression to it)
* Detect rotations->DetectRotations/DetectRotationsV2
* SGD->LinearModelsStory
* 2d cnn->GPU all links
* The sprints cript (I gave some ideas major part of the coding was martijn)

Jupyter Notebooks

Filter them

Presentations

Internal-> 2 one alone another with jake

External->Last one with martijn

Research Paper-> Summary took part in summary, introduction, problem description and research questions

REFLECTION AND EVALUATION

-Reflection on own contribution to the project(needs improvement)

*Situation*

In this minor the project I was assigned to was project Wheels, the task assigned to the group was to improve the outcomes of collecting data of wheelchair basketball players from the Dutch national team through sensors in the wheelchair.

*Task*

This was my first time working in a task with such level of importance and with a group of more than 3 people. Knowing this I couldn’t take a main role as I knew almost anything about what to do or how to do it so I had to listen many times as at first, I couldn’t add much, though throughout the project this was improved. Even though my situation, I expected myself to be helpful and at least have some general knowledge on all the aspects in which the project is involved.

*Action*

To make my role in the project more relevant I tried to work on many different aspects of it such as model preparation, research in models and data preprocessing.

*Reflection*

I think that my contribution in some parts due to lack of knowledge and not a complete domain of the language could have been better, but I think I tried my hardest for getting results good enough and tried to make something relevant for the project. I think in the future I should try to get more involved in the decision making and understand better the line in which the project is going, this means to know hat has to be done and why, this could help me to take a more important role in future projects.

STARR, reflected contribution and 400+ words(263)

-Reflection on own learning objectives

*Situation*

For the first time thanks to this minor, I got hands into python, also was my first contact with data science and my first project with this level of seriousness, so the situation was really challenging from the first moment

*Task*

From the first moment my objectives were to get some basic knowledge of python and learn everything I could of machine learning methods, and I wanted to understand how a real project works too, as they all are really valuable for my future.

*Action*

In order to achieve my goals, I tried from the first moment to take notes of everything that was new for me, which made me have a slower pace compared to others but would help me to understand everything better. Also, I tried to assist to all the lectures and workshops possible as even if they were optional, I knew they would be helpful too

*Reflection*

In my opinion my first approach to all these new things was good, but I think I went too slow on the python learning, which delayed me in other tasks. So definitely the approach I wanted to take was that, but I should have distinguished better in what parts I should be more conscientious in order to optimize time and efforts

STARR, reflected learning obectives and 400+ words(213)

-Evaluation on group project as a whole

*Situation*

For this project our group had 2 electrical engineers, 2 software engineers, 1 informatic engineer and Daan(ask him what were his studies about. This was a really good starting point as we had great variety of fields of expertise which could provide us with different points of view on the problems that could show up.

*Task*

As good it was to have a great variety of points of view for a same point, because that would help us have more than one opinion on one problem, our main deal would be also to reach a point of agreement between all of us, because as we all come from different fields it would be very possible that our opinions on some actions would be opposed to each other

*Action*

To avoid any kind of opinion difference, we just wanted to know each other’s opinions before making any decision, so we could make sure that everybody could understand why we are doing something

*Reflection*

In my opinion the group was good, everybody tried to do its hardest and to communicate with each other. Of course, there were points of non-agreement, but I think that they were solved easily, just by exposing what each other means with their idea and then see how that would affect the course of the project.

STARR, reflected group project and 400+ words(213)

**Predictive Analytics**->Knn, SGD(linear models), ´

*KNN*

As a first approach of the project, we decided to split the group into machine learning and data team. The machine learning team was in charge of making the first approaches to any usable method we could use. So, we decided to try make work a decision tree, knn and gaussian naïve bayes. My task was to make a working model of K-nearest Neighbors. This selection was made since they were the first models we were told about, so it would be a very naïve approach to the project in order to get used to the project, python and how machine learning works. Although some research was conducted before to see some examples of the model, also the datacamp resources were helpful(Supervised Learning with scikit-learn)

<https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>

<https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn>

<https://paperswithcode.com/method/k-nn>

<https://www.kdnuggets.com/2020/04/introduction-k-nearest-neighbour-algorithm-using-examples.html>

Filter links

As this was a first approach to machine learning the model was at its simplest so no tunning was done, just tried to work over to small sets of neighbors and see which one provided us with better accuracy. Then the accuracy of both train and test sets was plotted to see the difference between newighbours and apporpiately choose the optimum one

*SGD*

At this moment we wanted to try new models and we decided we wanted to see how linear models adapted to our dataset. At first, we researched on basic models(explain what models), also there was a team who made learning lab presentation based on svm which is a really well known linear model. Out of all the methods svm was really suitable as it’s a good method for two-class problems and also can perform well in time series dataframe, but for us there was only one problem which was the data frame’s size as it isn’t a good method for big amounts of data. Then we looked for similar options and we found out that the sgd was the one that could fit better out of the linear models we knew. After conducting some research and getting a working code we decided to tune the model through gridsearch providing a wide range of possible parameters and obtaining the best ones out of it, I also balanced the data(tag balanced=TRUE) which provided with not really good results so we decided to drop the model as the accuracy was not what we expected(have to explain why it didn’t work in the end)the data of the dataframe was not absolute??

Show image of confusion matrix/precision or whatever I have

*2d CNN*

After having our neural network and being in the middle of the 1d CNN development we decided to start working on the 2d CNN. For this task lot of time was required as I didn’t find that much information about how neural networks worked and how to make our data work into a 2d CNN. The main problem with this method was how to adapt our data to a 2d cnn, in the internet there aren’t that many resources about time series data being used for a 2d cnn so at first one option considered was converting the data into a 3d matrix, this method wouldn’t provide with any success, so we decided to make the graphs into images and then save them all, as we would have images and its information ready; the problem comes when there would be lots of problems when labeling the information, as we would need to make sure that only one plot occurred per image and also if a way of making the machine learning code know wether it comes from a previous sprint or its has just started ( Another idea we had was taking the plots into images which was achieved, we have the data almost ready as we only needed one of the most important parts the labels, in this case wether something was a sprint or not, this was impossible as we would need to make sure that there was not more than one sprint per image but also the sprints length don’t have same length so some would take more than 10 secs and others just 3 secs) the code was taken from some webs and videos of youtube(look for them) and also some of the previous code made for the 1d cnn was used, so the model would have been functional if it wasn’t for our dataset problems .so it was almost impossible for us and also we were told in our internal presentation to drop it.

(explain better)

**Domain** **knowledge**

When assigned to the task of making a functional 2d CNN to detect sprints with our data, first there was some research needed as I had no previous knowledge about image recognition. 2d CNN consists of ->explain whats it

Explain steps of 2d cnn based on documentation

Explain what are tensors, pooling, layers etc

**Data preprocessing**->Detect rotations

When choosing tasks for next sprints I wanted to change from machine learning tasks to data tasks, so I got into the rotation detection. In this task we already had divided the data into chunk of n, also we were using the max values of the chunks among those values because the chunks wore amongst one second so we chose the best value for it(explain better whats the best value and why) I was required to look at different of our features as wheelRotSpeed, … Also I needed to establish parameters for when the rotation started, mainly the criteria was the frame rotationspeed going above a certain value, but also I tried to work on different calculations that could helpto get more accurate on the detection of rotations.

All the measing values had been replaced by NaN at first instance so first I detected Nan from the datframe they were replaced by 0. Explain possibility of rotations not ending in the same plot

Once I had the conditions for a rotations, I plotted them i order o see if they were actually rotations, after trying out many combinations I found out the one providing better results, which aren’t perfect as we are working with max values we can’t find exactly all the rotations and sometimes finds rotations which aren’t