This is going to be a list of all the work items I have completed in dev Ops in the past Semester

**Learning on ML:**

1. Datacamp course: Introduction to python

- Screenshot of completed courses

In the first weeks of our project, we planned one day a week for the Datacamp courses. Since I already had a small coding project in Python, which was self-thought, the year before, I already knew the basic syntax of the language. Though, having a structured course layed out, which shows one the proper techniques, which are used in the Machinelearning field.

From week to week I felt more secure in the language, and I was able to make use of the API’s that got introduced to us in the course. Especially working in Data Science shows the beauty of an object orientated coding language like python. With the right API’s like Numpy and SKLearn it’s possible to attack unimaginable big problems with just a few lines of code.

70 Writing a script to get basic information about the data depending on exercises

TODO:

-Edit printouts, make clear what’s happening

-In Local directory and Getting to know

After I got to know our data through the experiments in blender, where at this point I was only able to load in one file at a time, I decided to read some basic information from the data. Instead of only using one file at a time, I wanted to calculate the mean for one axes, the X-axes of the right thorax, for all of the files contained in one given folder. Eventhough the mean of a dataframe can easily be calculated with the df.mean() function, I found this exercise really helpful, because it helped me understand the datastructure and its dimensions. Also it got me used to working with classes and functions without getting to complicated for the start.

122 Creating dataset that can be used for a ML Model

TODO

-what sprint?

At this point of time, I already worked with our data quite a bit, I plotted some easy graphs, I created my own anymations in Blender, and I could load in different files at once.

Now I wanted to load in the data to feed it to any ML model, but not by storing all the data in the memory, but by saving only the paths to the data in memory, to safe a lot of computing power. The whole groups agreed on this standard, so it was everybodys responsibility to get used to coding like this.

In main\_raphy.py I was able to make use of the already written controller scripts, that were implemented in the main branch of our GitHub repository.

The moment I got this script working, our main structure in the masterbranch changed again, so I continued working with the code that was already supplied by my colleagues. In src2/main. A np\_combination\_train and np\_combination\_test gets created. Those are X and y for the most of the machinelearning models. This is the first time, I used the datastructure that was planned for the model from the beginning, even from the last group.

Since I’ve been working with the 65 column data structure, I can say with certainty, that I understood the structure and know how it gets created.

Also to mention is that I prepared my data on my own, not for the logistic regression but for the CNN we’ve been training.

**125**. creating dataframe of 650 columns for the ml model

Explain this datastructure here:

Right here Im not explaining the full structure of the data, since this already has been done in our researchpaper

**123**. Splitting the data in a test and training set

The train and test set basically was split manually at the beginning. I picked out random patients out of different patientgroups, to be excluded of the training set, to be put into the test set. This is done to have a proper validation of the model. If the same data is used for training and validation, one cannot say with certainty, if the model is good, since the model already knows the test data from the training.

154 Prepare Data for RNN’S (CNN)

927ce917

-Picture of datastructure RGB

At first it was thought, that we’re going to use an RNN as an unsupervised training method, but we decided to use a CNN instead. That is because of the Datastructure we’re working with. Since there is, for every bone in every patient, an X Y and Z Euler-Angle, we can put these three values in one RGB pixel. Eventhough it isn’t necessary for a Nural Network to have an Input like this, we thought it would be a nice way of preprocessing the data, without loosing any of the information. In order to achieve something like this, the rotation first need to be normalized and the mapped on a scale from 0 to 1

As it is already described in the picture above, the “width” of the inputdata are the eight bones of the upper body, for 5 different exercises performed by one patient. Therefore one picture is one iteration of all the exercises one Patient has done( AB1, AF1,RF1,AH1…) the arrays shape is (40,100,(3)) Where 40 is 8 bones x 5 exercises and the 100 is the resampled framescount(length of exercise) the third(3) dimension is represented in the picture by the colour at the given pixel. These tensors then get stored in one bigger Tensor, which then gets fed to the CNN.

111. Finish ML Jupyter Notebooks

Done till 3.3 auto MPG regression

What kind of proof?

Is it necessary?

In preparation for the test, I started working through all the Machinelearing Notebooks provided on the datascience server. Therefore, I simultaneously had the lecture and the notebook open, this way I was able to understand and see the techniques used in the lectures by myself on my own computer. The given problems were built up gradually and of course followed the weekly lectures in the same paste.

Since our Dataset was quite different than for example the MNIST or other examples used throughout the lectures, it was quite hard to make the transition from big and structured datasets to ours, which is really small and only contains features that are from the same type. Nevertheless it was a good practice and it helped me understand the bigger picture of Data Science.

Excel sheet understanding last groups work

-described by task 122 partly

-Excel sheet with progress

In order to get everybody on the same page with their machine learning skills by week 10, we created a task, as a researcher I want to understand the steps the last group took. Therefore we created a Excel sheet to keep track on progress.

This task contains multiple subtasks, that have either been completed just to complete the subtask or in order to accomplish something else in the project. Here I will give links to the sections in my portfolio where I completed those tasks.

**Visualisation:**

20. Making a wireframe model based on data

-Master branch contains all blend files

Before any sort of a model could be created, it was upon me, to figure out what the data actually means. With some information from the LUMC and a paper describing the WU standard for Euler Angles in bone structures, we came up with a factsheet to describe which columns is responsible for which bone and in this bone which axes.

Description of the factsheet:

Afterwards, the struggle began to find the best “resting position” for the armature(the skeleton). This basically went down through trial and error. I first started with only one side(right side) and if I saw movement that made sense for my eyes I tried to construct the left arm based on the right one.

Master branch contains all blend files

35. Refine the 3D model in Blender

-Commit in task

-Add Blender Screenshots here

-Add

ON the second phase of the refinement, I added functions, to sort the different armatures in the scene, by either their exercise- or patient-group. I also automated the script to load in multiple files at the same time. Basically the script was created to function by the click of one button.

The user is able to choose which files he/she wants to read in, and by what the user wants to order the different “patients”. After all the animations have been created, one has the possibility to switch into the orthographic view, in order to compare the different “patients” without having any perspective issues.

This way of view the “patients” enables the possibility to see the exercises in 2D and 3D but also with and without perspective(with perspective helps to understand the movement, without makes it easy to compare the “patients to eachother”), all this can be done in realtime and full moveability of the camera in the same viewport.

Result:

Euler angles were in the way of getting a proper visualization. We also didn’t want to make more assumptions than we needed too. Maybe continued in chapter

64. Prepare Exercises for LUMC

-Scan in pages of LUMC

In order to validate the 3D visualization made in Blender, we either needed to check back with our client at the LUMC and get a verification that the animations are correct. The other possibility was, to create out own exercises, which one of the group members would perform while hooked up to the same Flock of Birds system, that has been use to collect the patientdata we received, at the LUMC.

To make sure, that the exercises will be performed exactly how I envisioned them beforehand, I created a protocol. This protocol contains hand drawn descriptions of the exercise and also step by step instructions for the “patient”. We didn’t want to rely on filenames and the protocol, we also filmed all the execises that had been performed by our group at the LUMC.

**71** cutting recorded videos of Lennart and Brice

-Files locally

Once we visited the LUMC and recorded our own movement data, It was about time to validate the script I created earlier. The first step was, to cut the recorded Video files to the right length and name them with the corresponding filenames of the .CSV – files.

Even though we had a protocol, to keep track in which order the tasks have been performed, it still was wearisome work, because this wasn’t the time for mistakes, since the recording should proof the reliability of the script that was planned to be used to label our dataset. To Accomplish this task I used Adobe Premiere.

After hours of trial and error, I came to the conclusion, that I won’t be able to get a proper representation of the data, using Blender. That’s because of the Euler angles that are used to describe the patients movement. In the Euler Angles nature lays, that the three angels can only describe ONE rotation in 3d space. This only applies, if the three axes are staying in the same order. This being said: X = 34°, Y = 45°, Z = 180° is different to : Z = 180°, X = 34°, Y = 45°

127 Visualize the dataset for the ML model in scatterplot

-python\src2\fd\visualize.py

-PG1-3 vs 4

Commit 42948dc5

After I made my first contact with the data in Blender and some minor experiments with Machinelearning, I started creating some scatterplots, to further understand the data.

In src2\main.py a visualization gets shown, after the training and test data gets created. This visualization shows the data exactly how it gets handed to a Machinelearnign model.

This graph was more a proof of concept, than an actual tool for visualization. This came, when I integrated the Visualization class into the main branch. At this point of the project we realized that patientgroup 4 we received from the LUMC wasn’t what we expected it to be. The values in pg4 were completely different, then the ones for pg 1-3.

-master2.0

-screenshot of configurations for the Visualize class

In order to contribute to the project, I created a class Visualize in the master branch, that anybody could make use of, just by turning it on in the config and by changing a few parameters in the definition of the class.

The user can choose which Petientgroups, out of these groups which patients, what exercises and what bones should be displayed in the graphs. If none of the parameters gets set, the script will grab all the information out of the config. Also the length of the exercise doesn’t matter, since the script gets all it’s information out of the config.

I created a visualize\_exercise function to compare different exercises between Patientgroups and patients and bones.

152 Create visualization to compare RAW and CSV files

Branch Raw visualization

All the efforts to get a real 3D visualization in Blender didn’t bring any good results and just let us continue guessing. Since Assumptions aren’t any good in machine learing, we needed a visualization that we could trust. Therefore eddy wrote a script, based on the last groups work, to visualize the RAW files we received from the LUMC. I used this as a base for my script.

98 Check for flipped sensors

-Answer in issue

166 Create poster for presentation

-Poster in cloud

**Paper**:

26. Read Paper

66. Find research paper that help answer a sub-question

-excel sheet with subquestions

-answer one subquestion in detail

**174** Summarize answer of sub-question

-Answer somewhat in issue

173 Start with the paper

189 Visualize: Explain the graphs we’re using

181 Data: Occupied space

180: Data: Frame gEn

I cannot make it any further, I feel sooooo distracted by the work Im doing. Maybe writing 2000 words in two days wasn’t the right thing to do at this point of time. Im still hesitating if I want to prepare myself for tomorrow or if I just let happen what is about to happen. Maybe its my purpose to go out of my way in order to help the project to go along.

The big question is, what’s going to happen if I don’t do anything against it?

The sense is lost,