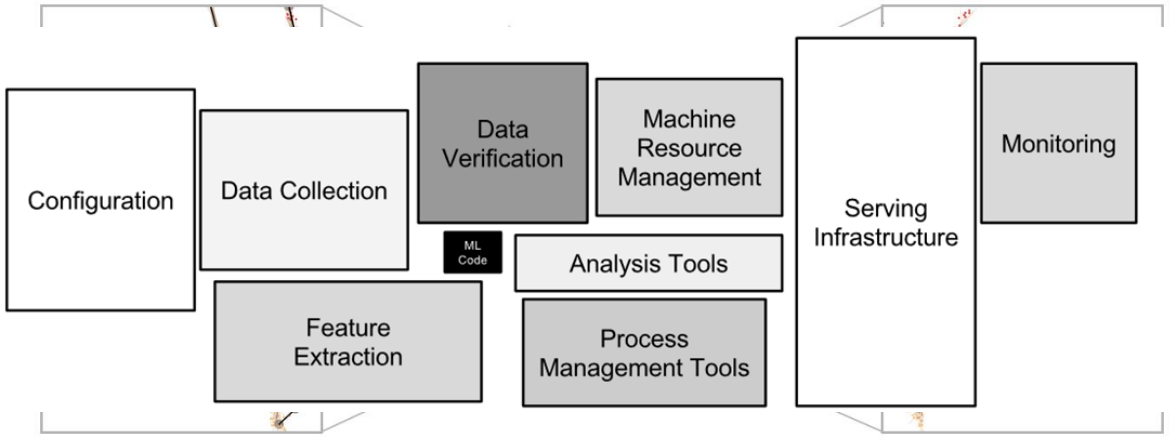
02476 Machine Learning Operations

Nicki Skafte Detlefsen

Intro to the course

Who am I

- Bachelor, master, PhD from DTU
- Currently: Postdoc
- Old focus:
  - Inductive biases in deep learning
  - Probabilistic generative models
  - Manifold learning
- New focus:
  - MLOps



PIONEER CENTRE FOR ARTIFICIAL INTELLIGENCE

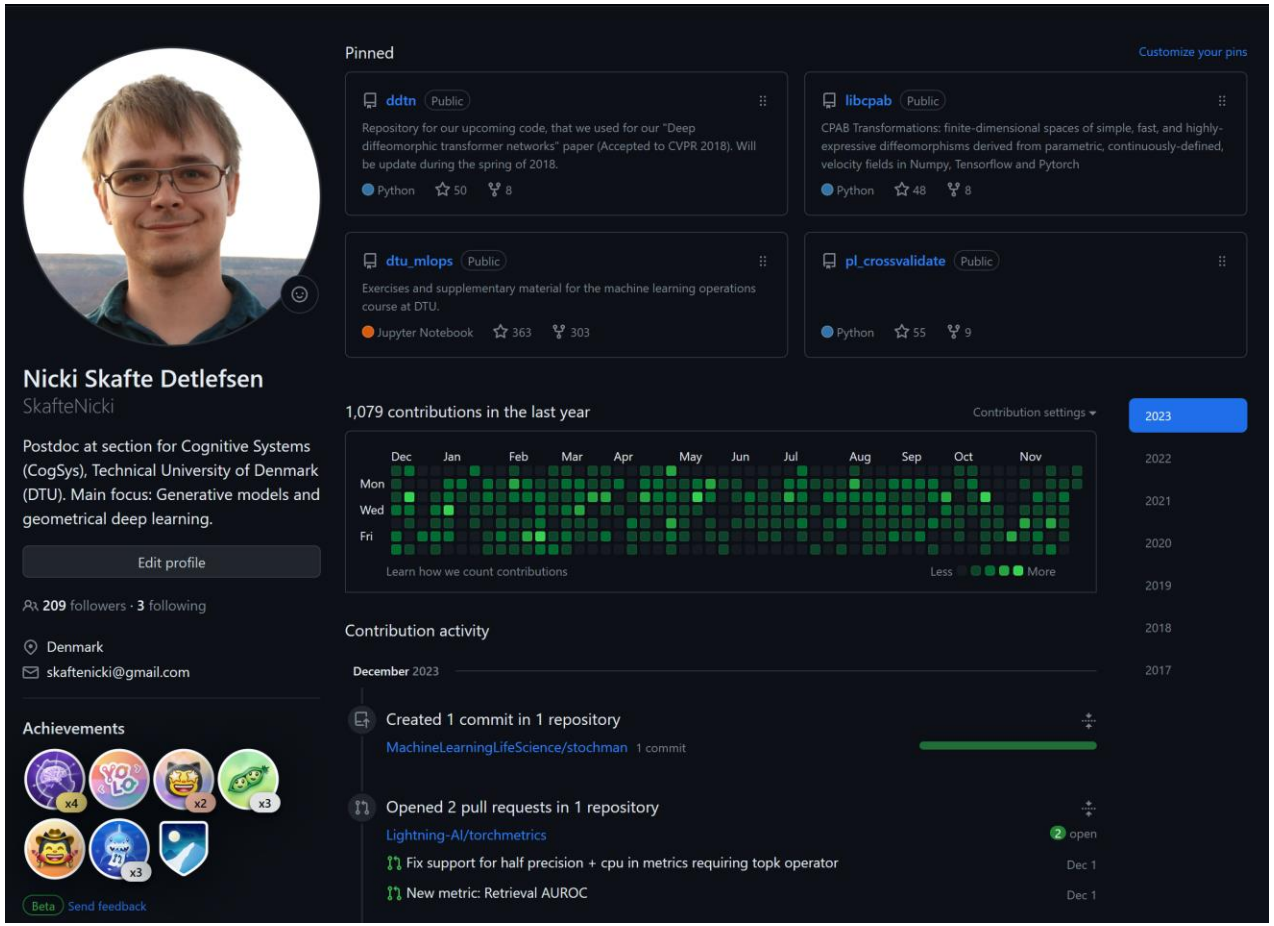
My secret identity

💻 Eager open-source contributor

- Numpy
- Scikit-learn
- Pytorch

⚡ ML Engineer at https://lightning.ai/

- Pytorch-lightning
- Torchmetrics



Who else to know about

Søren Hauberg
Co-responsible

- August Leander Høeg
- Diogo Oliveira Marques Adegas
- Anders Gjølbye Madsen
- Amin Hasanpour
- Laurits Fredsgaard Larsen
- Shah Bekhsh
- Lina Skerath
- Fabian Scott
- Junaid Ahmed Qazi
- Ana Marija Pavicic

TAs are available from 10:00-16:30 every day

In general there is 1-4 on duty, either online or physical on campus

Course setting

- 5 ECTS
- 3 weeks period
- Level: Master
- Grade: Pass/not passed
- Type of assessment:
  – Project report

Recommended prerequisite

- General understanding of machine learning (datasets, probability, classifiers, overfitting etc.)
- Basic knowledge about deep learning (backpropagation, convolutional neural network, auto-encoders etc.)
- Coding in Pytorch

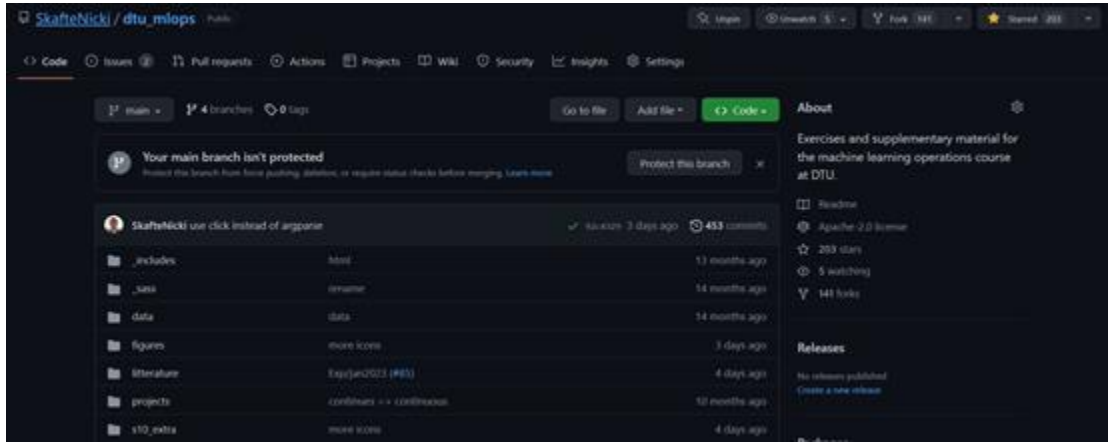
Course webpage

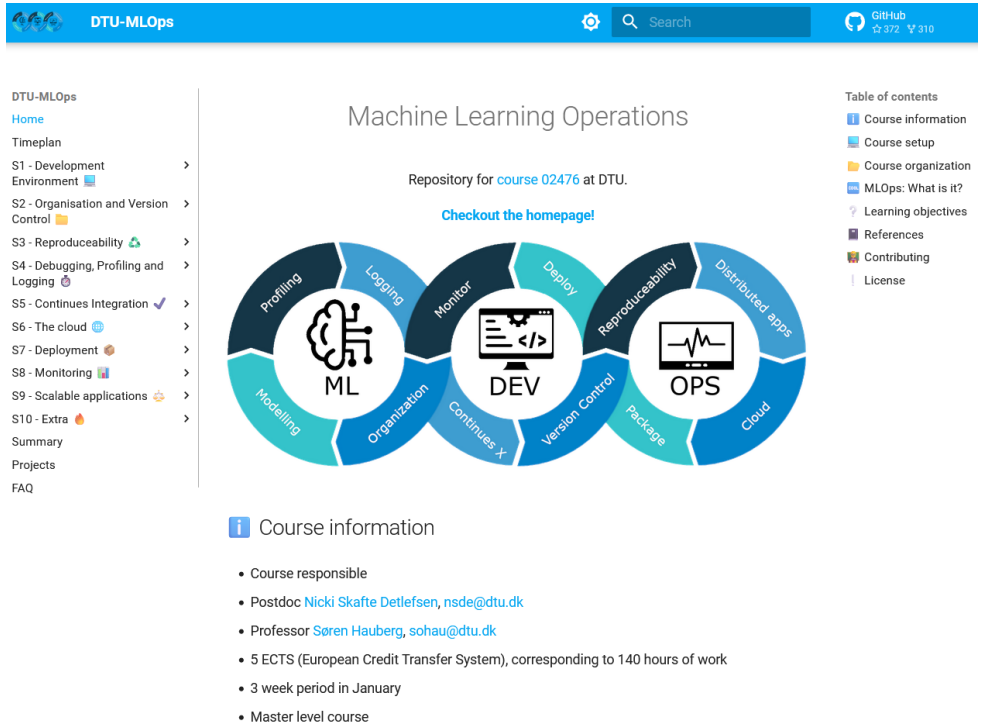
Github:

https://github.com/SkafteNicki/dtu_mlops

Rendered page:

https://skaftenicki.github.io/dtu_mlops/





Communication

Join the slack channel

https://join.slack.com/t/dtumlops/shared_invite/zt-2utq0bupc-5gTBkFPjaTLkQVBc2C4Qqg

General announcements

- Asking questions
- Communication with team members

For non-public info we use DTU learn

https://learn.inside.dtu.dk

What is this course about?

What is this course:

> *Introduce the student to several tools and software development practices that will help them organize, scale, deploy and monitor machine learning models either in a research or production setting. To provide hands-on experience with a number of frameworks, both local and in the cloud, for working with large scale machine learning pipelines.*

Keywords

- ⭐ Organization
- ⭐ Scalability
- ⭐ Reproducibility
- ⭐ Hands-on experience

What this course is not

⚠️ How different machine learning models works

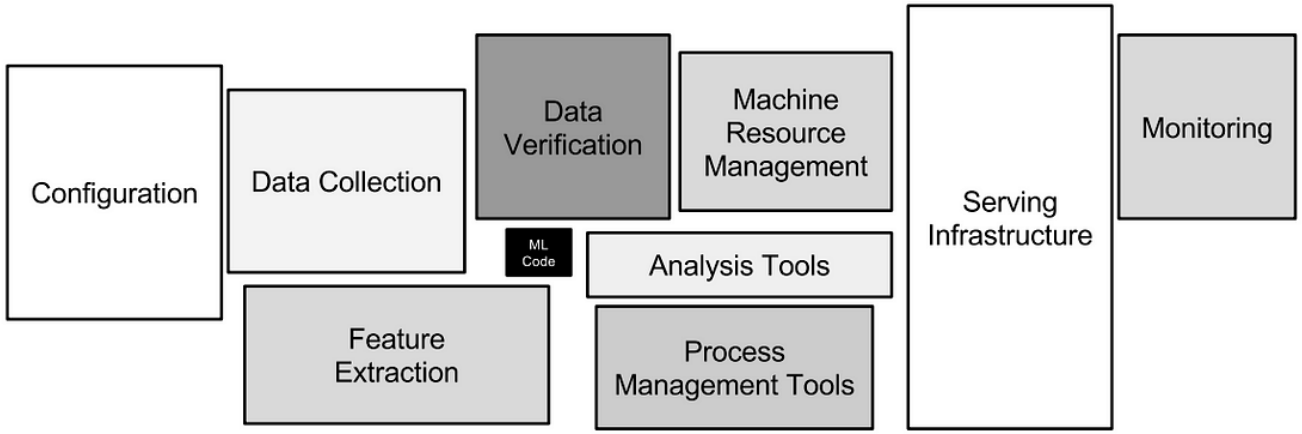


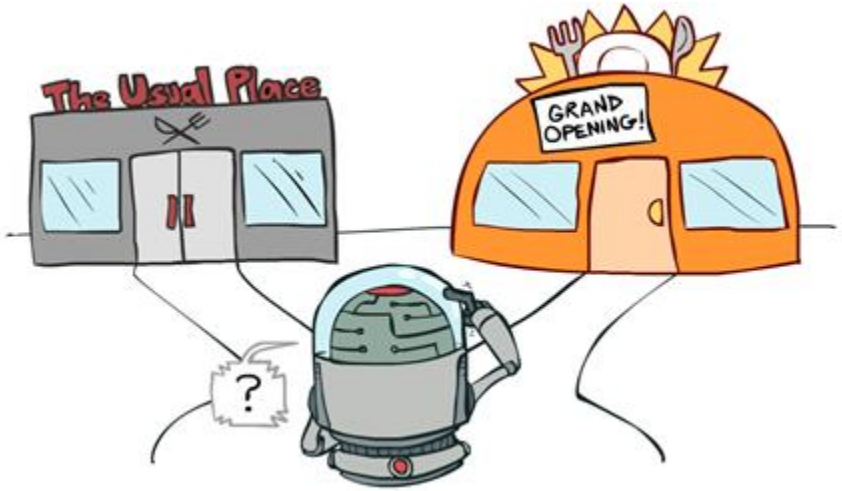
Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

The teaching method of this course

The course is centered around two principals:

- *Learning by doing*
- *Hybrid learning*

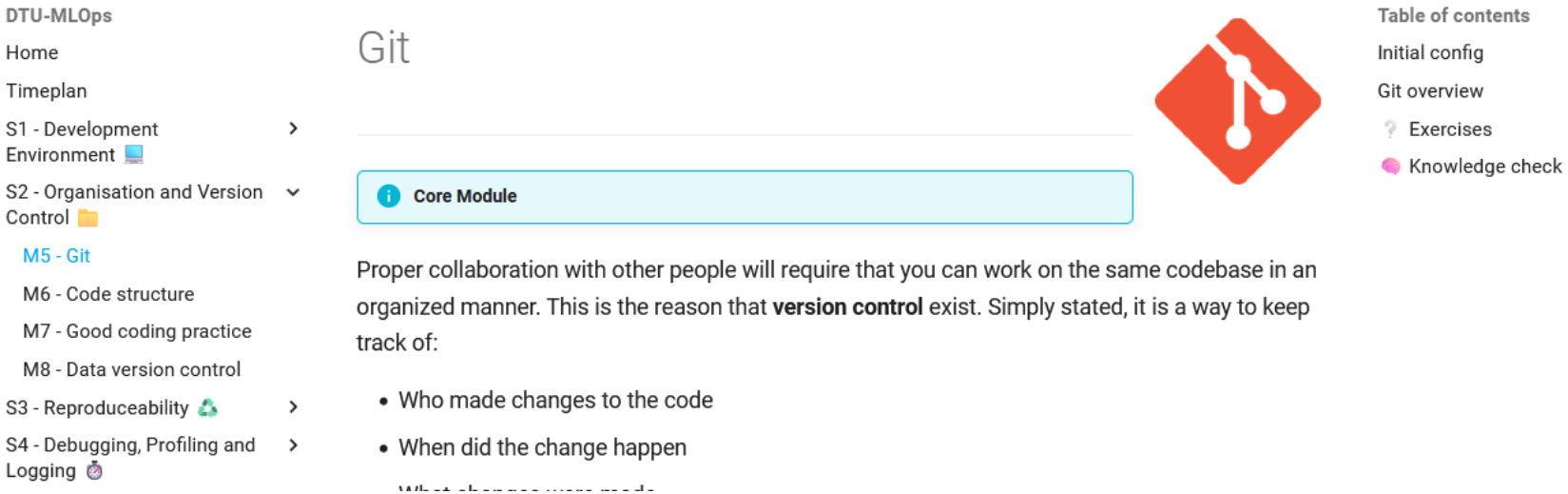
We provide lectures, exercises and guidance but encourage self study.



Exploitation vs Exploration

Organization of material

- 1 day = 1 session (S)
- 1 session = multiple modules (M)
- Core modules:
  - Essential in some way
- All other modules are highly recommended
- S10 contains additional modules



What I hope from this course

- Have fun!
- That you get to fill your toolbox with useful frameworks
- (Maybe) Learn something along the way

A typical day in this course

⚡ Exercise days:

- Meet in at 9:00
- Lecture for 30-45 mins
  - I am still learning how to do lectures
  - Lectures are not meant to give teach you anything, but provide some context to the topic of the day
- Exercises until 14:00-17:00
  - Remember to take a lunch break
  - Workload will depend on you

🔥 Project days

- Sometimes a small lecture or company presentation
- Rest of the day you work on projects
- Office hour (may be virtual)

Projects 🔥

Approximately 1/3 of the course time is spend on project work

More info here: https://skaftenicki.github.io/dtu_mlops/pages/projects

Already now you are recommended to think about forming groups

- 4 people (3 and 5 is also acceptable)
- Thursday we will do some speed dating to form groups for people not already having one.
- Also feel free to write in the #find-a-group slack channel.

Prompt:

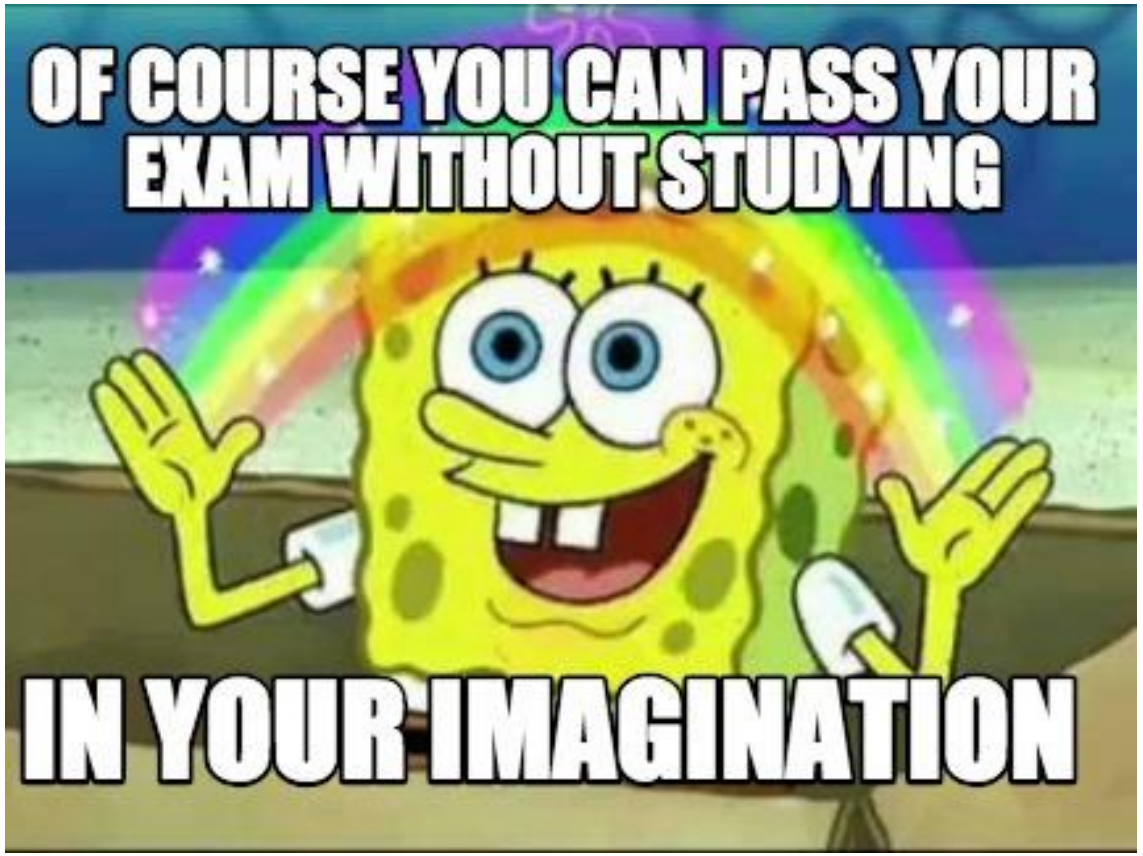
Group of students working hard on a project

How to pass

- Do the exercises
- In the final project:

  Show that you can use the tools you learn about throughout the course

We still have a 100% pass rate after approximately ~350 students.



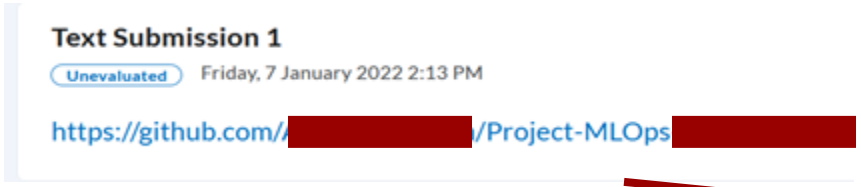
Exam

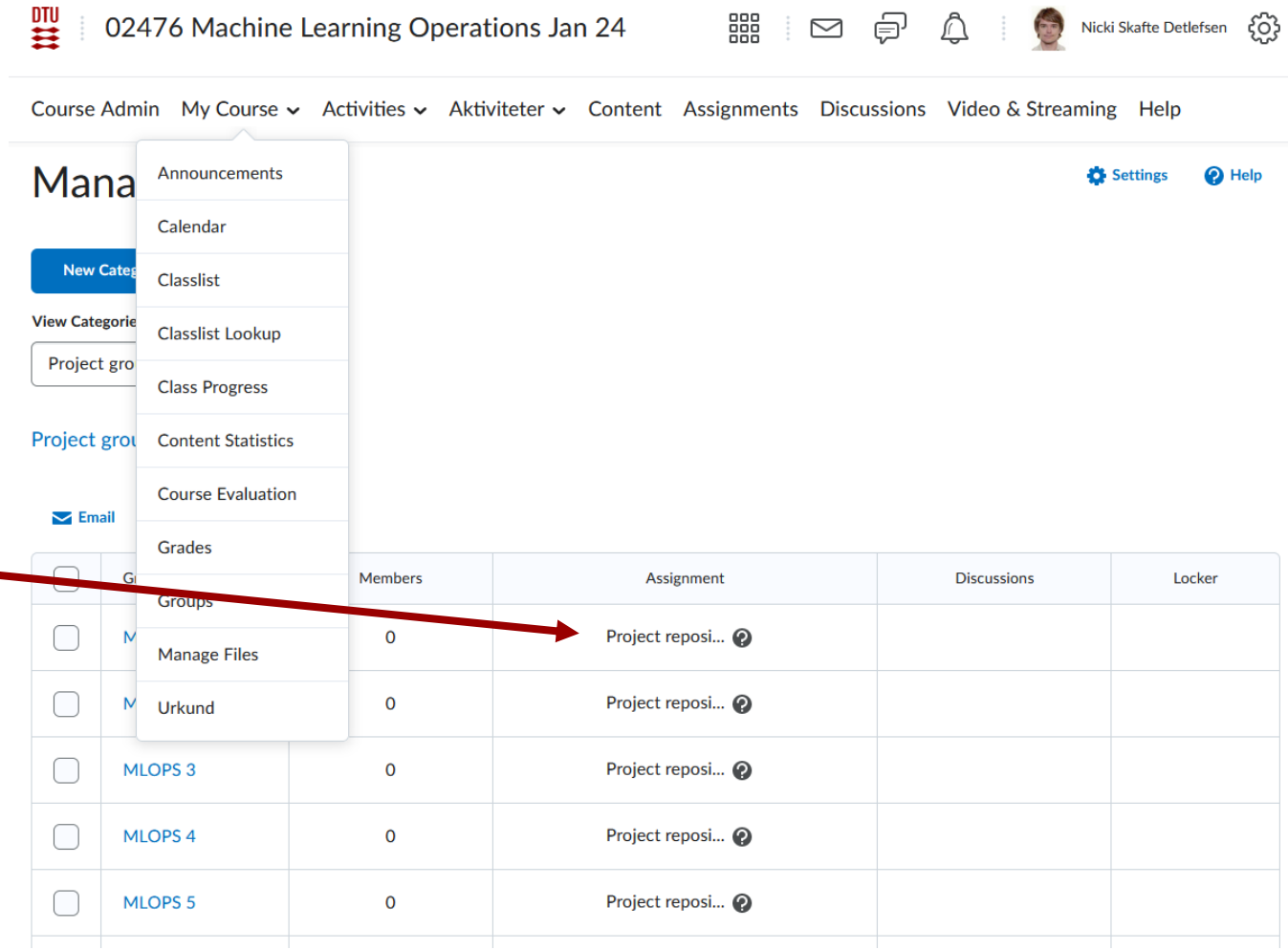
The exam only consist of a written part: A template with ~30 questions that you can fill out as you work on your projects. It will be part of your project Github repository.

More on this on Friday ⚠️.

One hand-in during the course

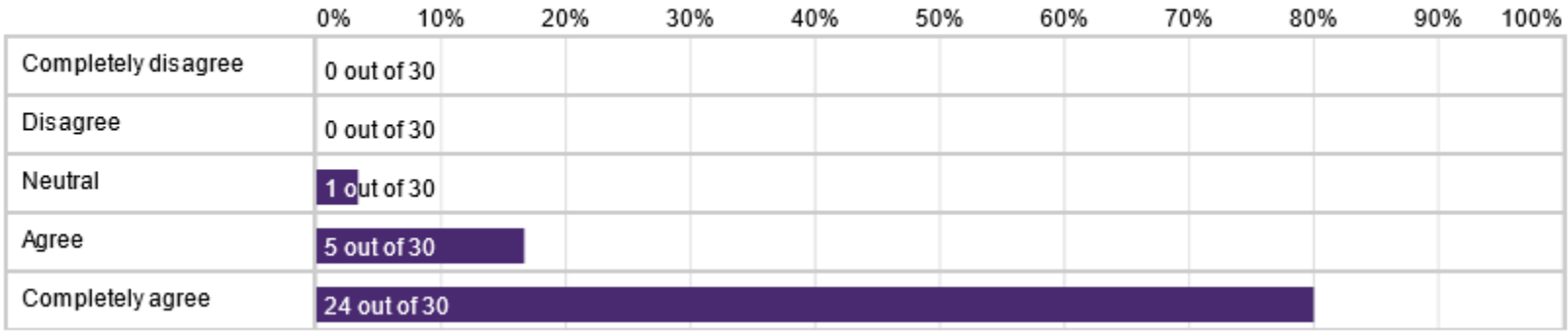
- Signup as a group
- Hand-in the link to your Github project repository



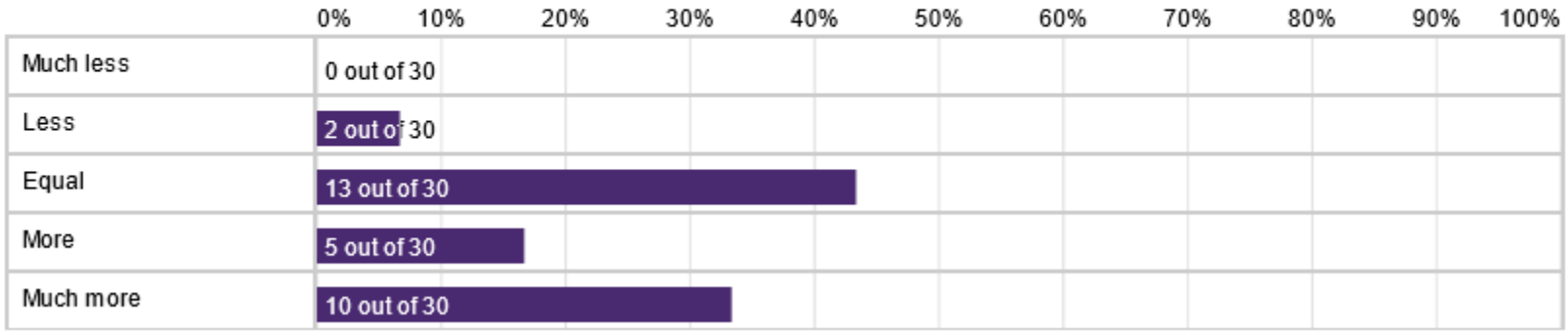


The course in 2 statistics

1.1 I have learned a lot from this course.

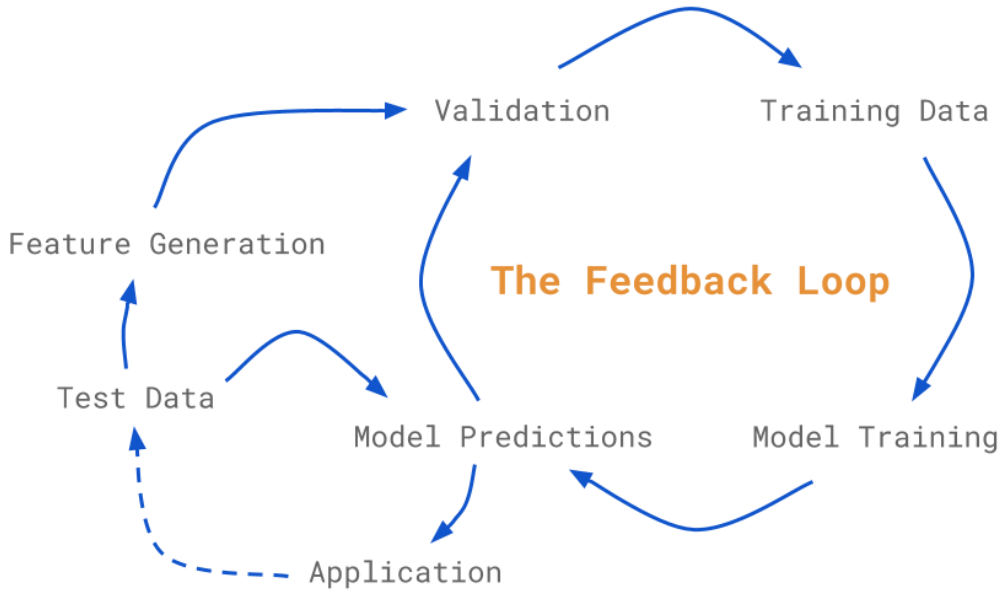


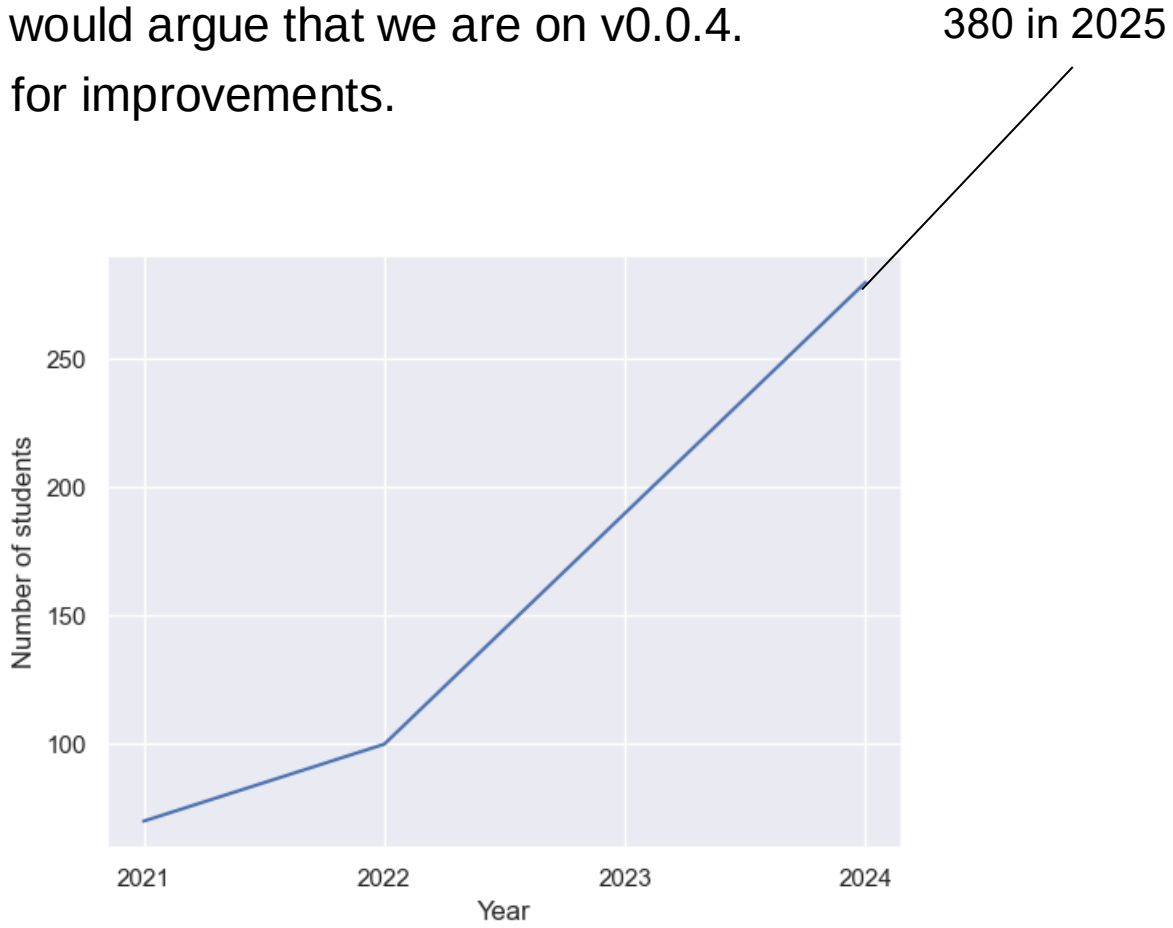
2.1 5 ECTS credits correspond to nine working hours per week for the 13-week period (45 working hours per week for the three-week period).
I think the time I have spent on this course is



It is not a perfect course

Some would say we are on v4.0 of the course; I would argue that we are on v0.0.4.
Please come talk to me if you have suggestions for improvements.





FAQ

❓ Can I work from home

💡 Yes, but note that

* for the project days you need to agree with your group on this
* we have limited TA resources and will priorities students on campus
* the oral exam is takes place physically

❓ Can I use ChatGPT or similar

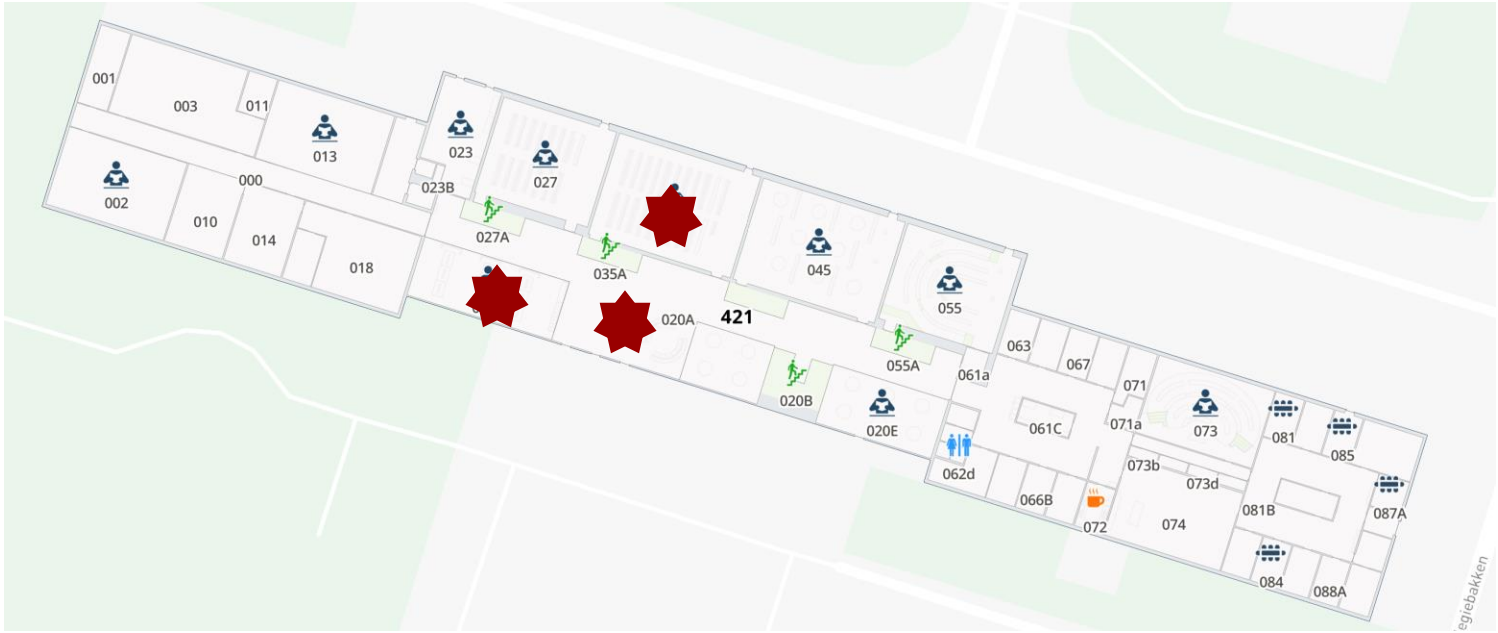
💡 Yes all you want, but make sure you still learn something

❓ What if I become sick during the course

💡 If you can work from home, then that is the best option. Second best option, is to make sure you still contribute to the final project but skip doing some of the exercises

How to get help ?

- We have auditorium 72, group area mid, group area west, but use whatever space you can find in the building
- Nicki will be in the auditorium from 8-14
- TAs will be around from 11-16:30 in auditorium + group areas



Memes

Let’s try to have some fun while learning