Lab session 1

Machine Learning for Behavioral Data (CS-421)

February 23, 2022

Today

- Welcome!
- Quiz and survey
- Setting up your environment
- Project presentations

Team

Instructor

Tanja Käser
tanja.kaeser@epfl.ch

Teaching Assistants

Vinitra Swamy, Paola Mejia
vinitra.swamy@epfl.ch, paola.mejia@epfl.ch

Student Assistant

Francesco Salvi
francesco.salvi@epfl.ch

Lab Sessions

- Wednesday, 8:15-10:00
- Content:
  - Introduction
  - Project office hours

Week	Lecture	Project Hours	Milestones
1	Introduction	Detailed project presentation	-
2	Data Exploration	Introduction to tasks for M2	*M1: Preferences on team members and data sets (Feb 28, 23:59)*
3	Regression	Office hours	
4	Classification	Office hours	
5	Model Evaluation	Introductions to tasks for M4	*M2: **Individual** exploration of selected data set (March 21, 23:59)*
6	Time Series Prediction	Individual discussion with teams	*M3: selection of research question and approach (March 28, 23:59)*
7	Time Series Prediction	Office hours	
8	Time Series Prediction	Office hours	
9	Spring Break	Spring Break	
10	Guest Lecture: Neuroscience	Office hours	
11	Unsupervised Learning	Individual discussion with teams	*M4: submission of results for first research question (May 2, 23:59)* *M5: ideas for extension and approach (May 2, 23:59)*
12	Unsupervised Learning	Office hours	
13	Ethical Machine Learning	Office hours	
14	Ethical Machine Learning	Office hours	
15	Project Presentations	Project Presentations	*M6: Final Presentation*
16	Semester End	Semester End	*M7: Hand in report and code base (June 12, 23:59)*

Project

- Teams of 3 people
- We will provide data sets
- We will provide example research questions
- You will suggest an additional analysis/extension to the selected research question
- We will give feedback during the semester (see milestones)
- We will do project office hours (during lab sessions)
- You will do a presentation in the last week of the semester
- Final project (Code + Report)

Quiz

Quiz

https://www.python.org/

https://github.com/

https://www.anaconda.com/products/individual

https://jupyter.org/

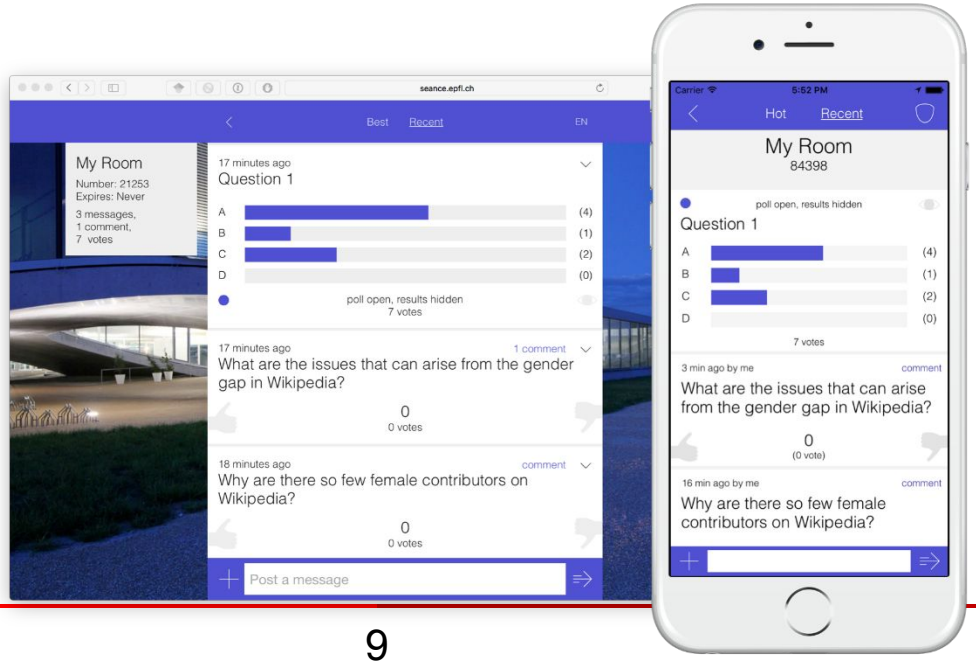
https://noto.epfl.ch/

https://colab.research.google.com/

http://speakup.info/

SpeakUp

https://go.epfl.ch/mlbd-lecture



Python

SpeakUp: How much do you know about Python?

A: It’s a family of nonvenomous snakes with 10 genera and 42 species.

B: I have heard about the programming language Python.

C: I have used Python a few times (e.g. for courses).

D: I use Python on a regular basis.

Jupyter

SpeakUp: How much do you know about Jupyter?

A: It’s the largest planet of our solar system.

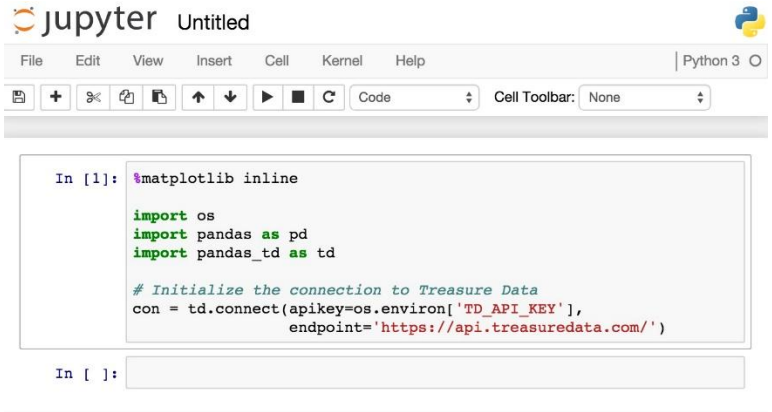
B: I have heard about Jupyter notebooks.

C: I have used Jupyter notebooks a few times (e.g. for courses).

D: I use Jupyter notebooks on a regular basis.

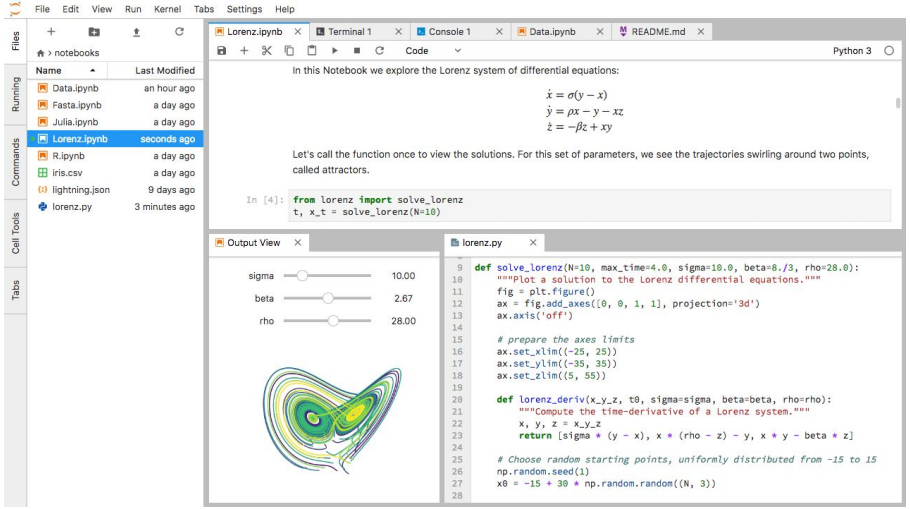
Jupyter

Jupyter notebook



Tutorial: https://www.dataquest.io/blog/jupyter-notebook-tutorial/

JupyterLab



Why JupyterLab: https://towardsdatascience.com/jupyterlab-a-next-gen-python-data-science-ide-562d216b023d

Anaconda (local env)

SpeakUp: How much do you know about Anaconda?

A: It’s the heaviest and one of the longest known snake species.

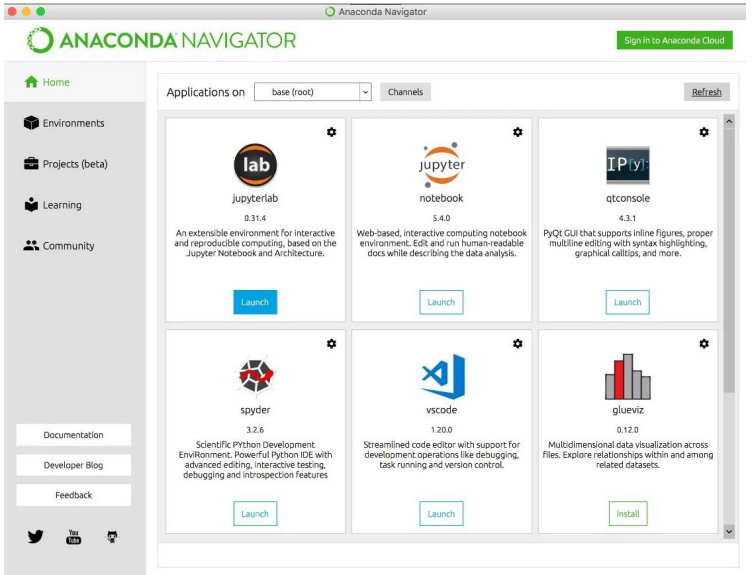
B: I have heard about Anaconda.

C: I have used Anaconda a few times.

D: I use Anaconda on a regular basis.

Anaconda (local env)

- You have the full control
- Works offline
- https://www.anaconda.com/products/individual



- **Tutorial**: https://www.edureka.co/blog/python-anaconda-tutorial/

Google Colab (online env)

SpeakUp: How much do you know about Colab?

A: It’s an abbreviation for an artist group from New York.

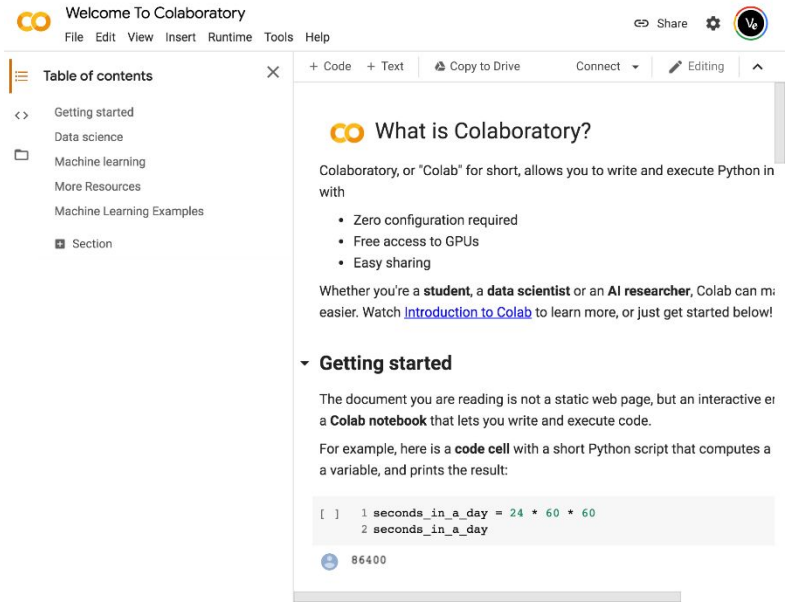
B: I have heard about Colab.

C: I have used Colab a few times.

D: I use Colab on a regular basis.

Google Colab (online env)

- Ready environment
- Uses Google's infrastructure
- Collaborative functionality
- Requires Google account
- https://colab.research.google.com/



- **Video**: https://www.youtube.com/watch?v=inN8seMm7UI

EPFL Noto (online env)

SpeakUp: How much do you know about Noto?

A: It’s a city in Sicily declared a UNESCO world heritage in 2002.

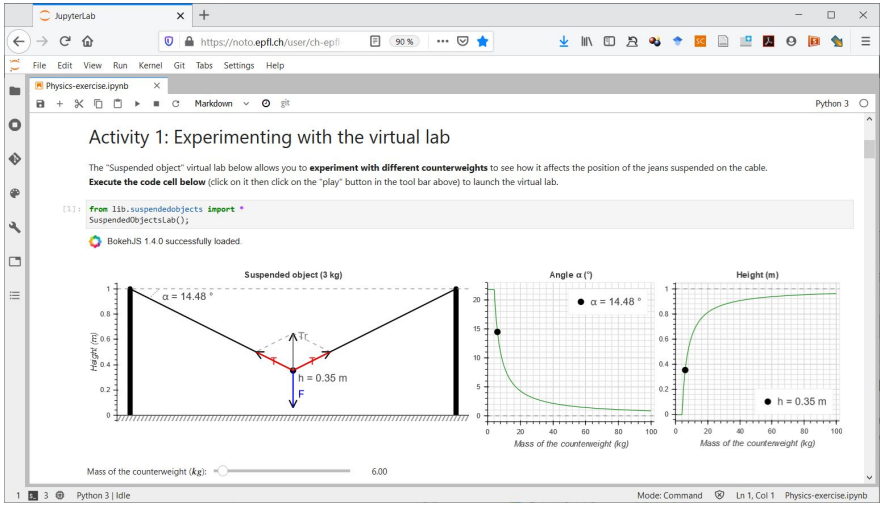
B: I have heard about Noto.

C: I have used Noto a few times.

D: I use Noto on a regular basis.

EPFL Noto (online env)

- Ready environment
- Login with your Gaspar
- https://noto.epfl.ch/



GitHub

SpeakUp: How much do you know about GitHub?

A: Git…..what?

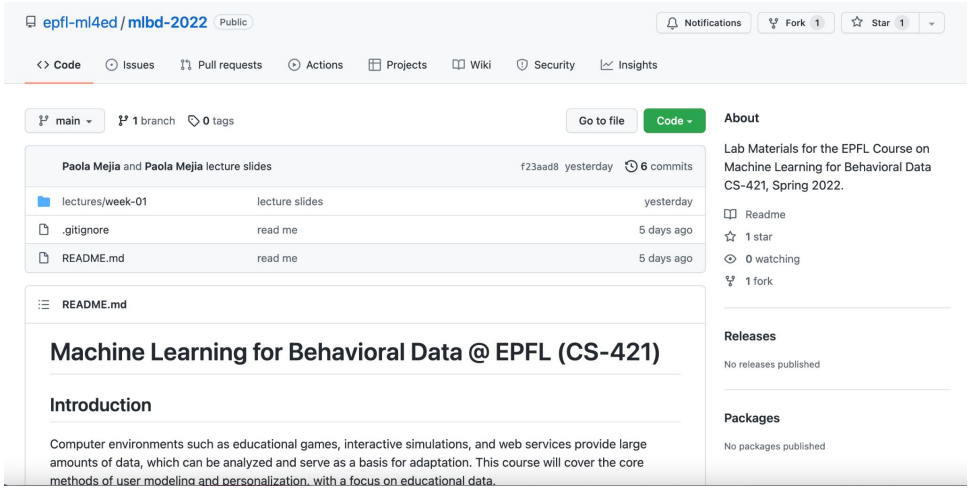
B: I have heard about GitHub.

C: I have used GitHub a few times.

D: I use GitHub on a regular basis.

GitHub

- Share files and code
- Version control
- The course repository: https://github.com/epfl-ml4ed/mlbd-2022



- **Tutorial**: https://www.edureka.co/blog/how-to-use-github/

(Demo)

Setting up the environment

- Set up an environment on which you can
  - Run Jupyter notebooks in Python
  - Connect to course repository: https://github.com/epfl-ml4ed/mlbd-2022

- We will use https://noto.epfl.ch/
  - But you are free to use whatever you want (e.g. Anaconda, Colab etc.)
  - It's your responsibility to have a working environment
- **Task**: Pull course's GitHub repository

Noto

- Using Noto:
  - Go to https://noto.epfl.ch/
  - Login with your GASPAR
  - Go to Git → Clone
  - Clone the course repository: https://github.com/epfl-ml4ed/mlbd-2022
  - Read sample tutorial (lectures/01-tutorial.ipynb)

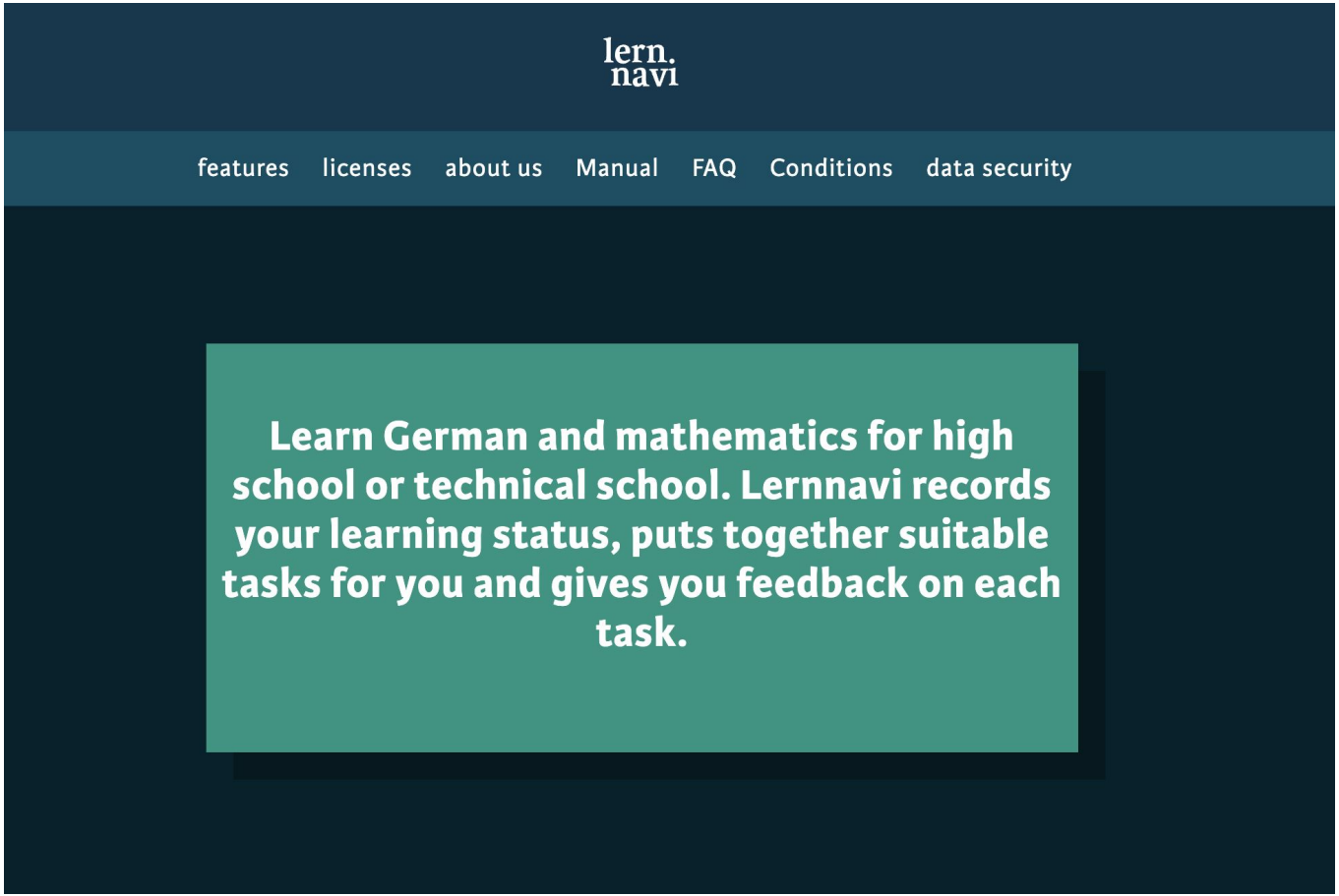
Anaconda

- Virtual environment:
  - https://janakiev.com/blog/jupyter-virtual-envs/
  - Create virtual environment: `python -m venv myenv`
  - Activate virtual environment: `source myenv/bin/activate`
  - add to Jupyter: `python -m ipykernel install --user --name=myenv`

Start-up Presentations

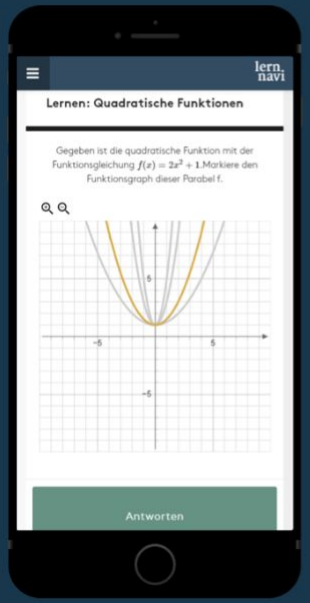
Two EdTechCollider start-ups have provided us with real-world data:

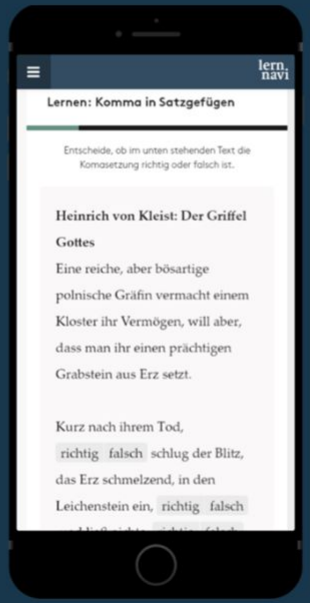
- Lernnavi
- Classtime



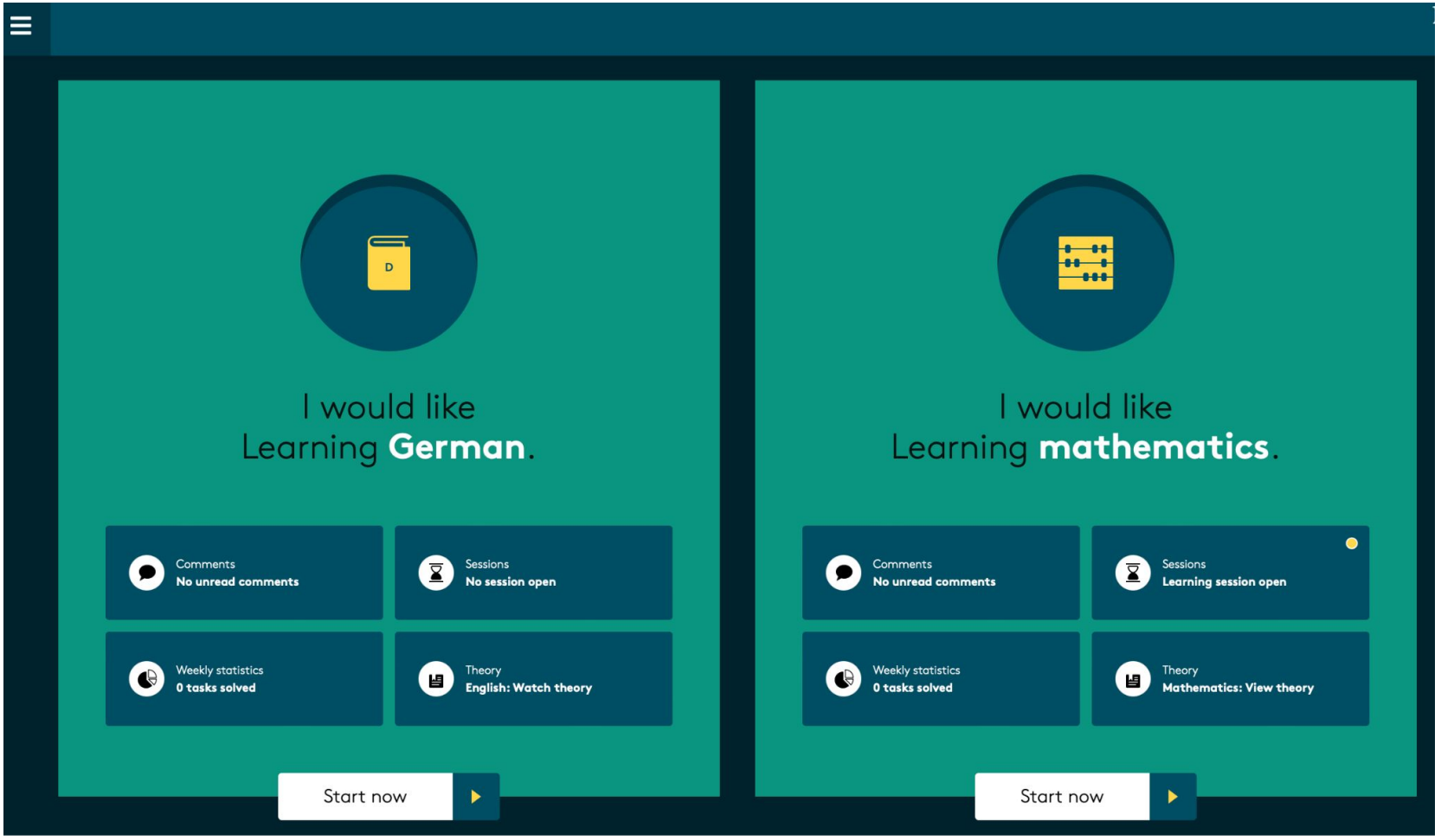
https://www.lernnavi.ch/

German and mathematics





Lernnavi is an instrument for promoting part of the basic subject-related study skills in German and mathematics. The development teams created the reference framework for German and mathematics based on the appendix to the framework curriculum. The cantonal student councils and the VSDL (Association of Swiss German teachers) or the DMK (German-Swiss Mathematics Commission) and the core group for canon mathematics were able to comment on these in the consultation process. Thanks to this approach and the broad support, a good fit with the framework curriculum is ensured.


I would like
Learning German.
Comments
No unread comments
Sessions
No session open
Weekly statistics
0 tasks solved
Theory
English: Watch theory
Start now
I would like
Learning mathematics.
Comments
No unread comments
Sessions
Learning session open
Weekly statistics
0 tasks solved
Theory
Mathematics: View theory
Start now

Demo

- Demo accounts (send email or raise hand)

Lernnavi | Overview

- ~13,000 students
- ~3,000,000 events
- Detailed table description

userId	timestamp	category	action	eventType	transactionToken	sessionId
387699	1621541425644	TASK	VIEW_QUESTION	VIEW	a2434f9f-e25f-4247-b462-d33015c851c3	34333.0
387699	1621541428812	TASK	SUBMIT_ANSWER	CLICK	a2434f9f-e25f-4247-b462-d33015c851c3	34333.0
387699	1621541430337	TASK	NEXT	CLICK	a2434f9f-e25f-4247-b462-d33015c851c3	34333.0
387699	1621541430564	TASK	VIEW_QUESTION	VIEW	5822efa9-e1e5-4c5d-8808-b0807d9578ab	34333.0
387699	1621541431405	SESSION	CLOSE	CLICK	5822efa9-e1e5-4c5d-8808-b0807d9578ab	34333.0

Lernnavi | Project Ideas

- **Time series analysis** of students.
- Prediction of **students' knowledge gain**
- Does **gender / geographic region** have considerable differences in student performance?
- Prediction of students' engagement.
- What are the traits of the most successful students?

Classtime

PRICING USE CASES RESOURCES English Teacher Login

The clear path to student success.

Classtime is a solution for classrooms that complements in-class teaching with immediate feedback on students' level of understanding.

Teachers

Start now

Students

Enter your code here Join

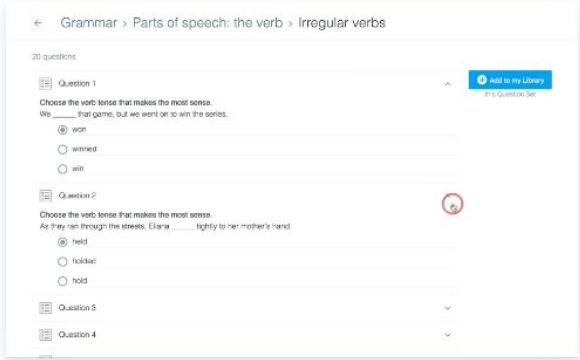
or log in with your student account

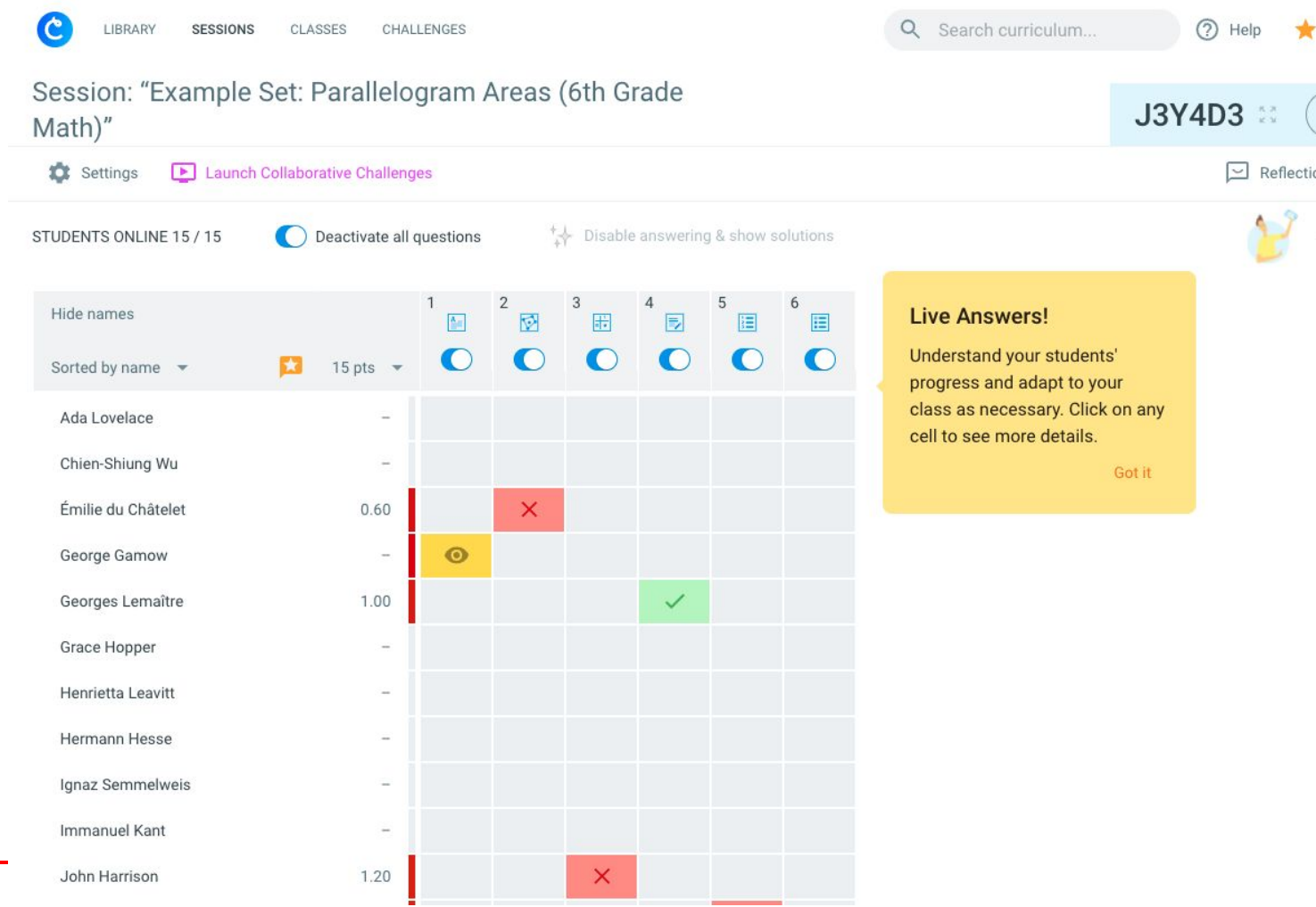
https://www.classtime.com

Question libraries

Take advantage of our free libraries with more than 50,000 standards-aligned questions, handcrafted for your students' success by education experts.

> Khan Academy
> Public Library




LIBRARY
SESSIONS
CLASSES
CHALLENGES
Search curriculum...
Help
Session: “Example Set: Parallelogram Areas (6th Grade Math)”
J3Y4D3
Settings
Launch Collaborative Challenges
Reflectio
STUDENTS ONLINE 15 / 15
Deactivate all questions
Disable answering & show solutions
Hide names
Sorted by name
15 pts
1
2
3
4
5
6
Ada Lovelace –
Chien-Shiung Wu –
Émilie du Châtelet 0.60
George Gamow –
Georges Lemaître 1.00
Grace Hopper –
Henrietta Leavitt –
Hermann Hesse –
Ignaz Semmelweis –
Immanuel Kant –
John Harrison 1.20
Live Answers!
Understand your students' progress and adapt to your class as necessary. Click on any cell to see more details.
Got it

Demo

- Students can create free accounts to simulate:
  - the teacher experience
  - the student experience

http://classtime.com → Teacher Login

sessions (meta-information about session format, teacher)
questions (all questions asked across sessions)
answers (each row represents a student answer to a question)
reflection_questions (4 reflection questions available to ask students)
reflection_answers (each row represents a student answer to a reflection questions)

Classtime | Overview

- ~6,000,000 students
- ~7,000,000 questions
- 390,000 Classtime sessions
- 4 reflection questions

Detailed table description

(second part)

Classtime | Project Ideas

- **Time series analysis** of students in sessions
- Prediction of **reflection response**
- Does **language / geographic region** have considerable differences in student performance?
- Do **specific teachers** have high student engagement? What are the traits of the most successful classrooms?

Milestone M1

https://go.epfl.ch/mlbd-m1

Fill out with team and start-up preference

Deadline: Monday, Feb 28th, 23:59

Feedback

We are actively looking for feedback to improve

https://go.epfl.ch/mlbd-feedback

Questions?