Machine Learning for Behavioral Data (CS-421)

Introduction

February 18, 2025



Today

- What is ML for Behavioral Data?
- Course Logistics
- Active Learning
- Project
 - Example Projects
 - EdTech StartUp(s)

Today

- What is ML for Behavioral Data?
- Course Logistics
- Active Learning
- Project
 - Example Projects
 - EdTech StartUp(s)

This will be an interactive course...

More on this later

For now: take your phone (or laptop) and join us on

SpeakUp

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

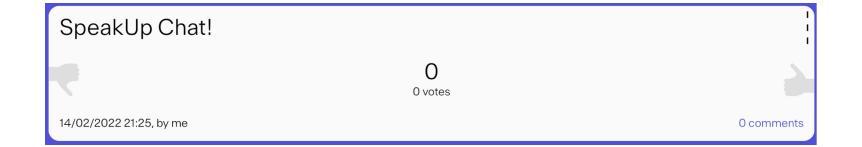


About Me

- Assistant professor at EPFL since May, 2020
- Head of the ML4ED lab
- In the past, I was a
 - senior data scientist at the SDSC
 - postdoc at Stanford University
 - postdoc at ETH Zurich/consultant for Disney research
 Zurich
 - PhD student at ETH Zurich

Students – Shake Hands





Human

Interaction with Environment



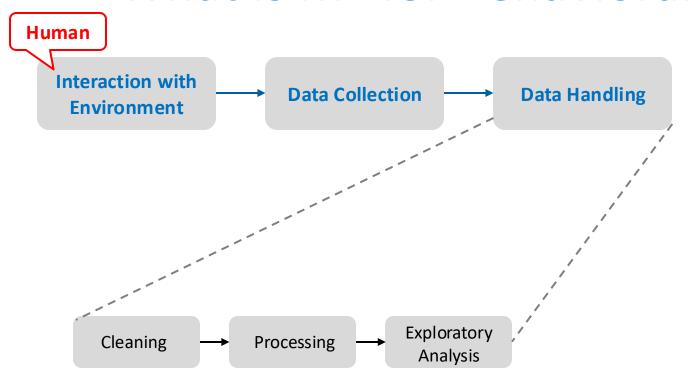


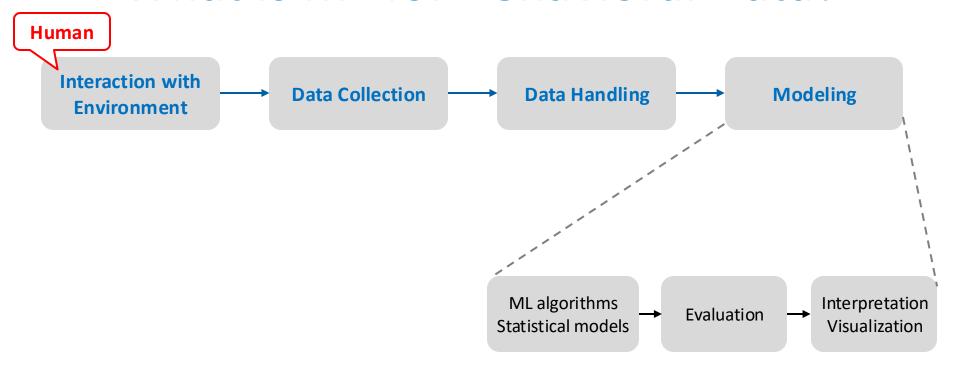


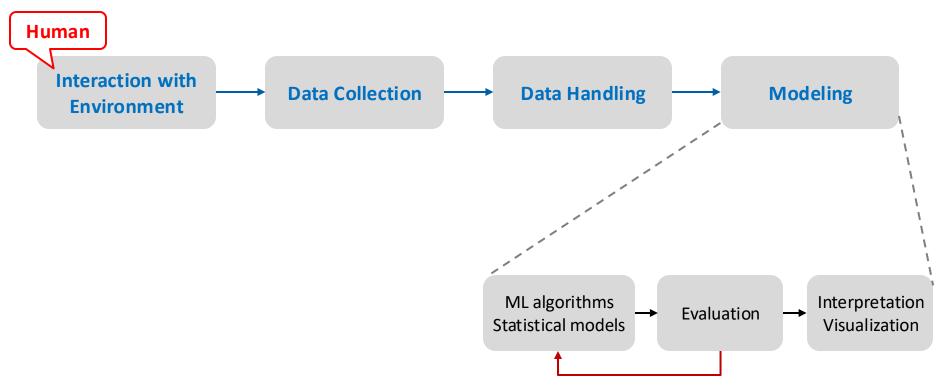


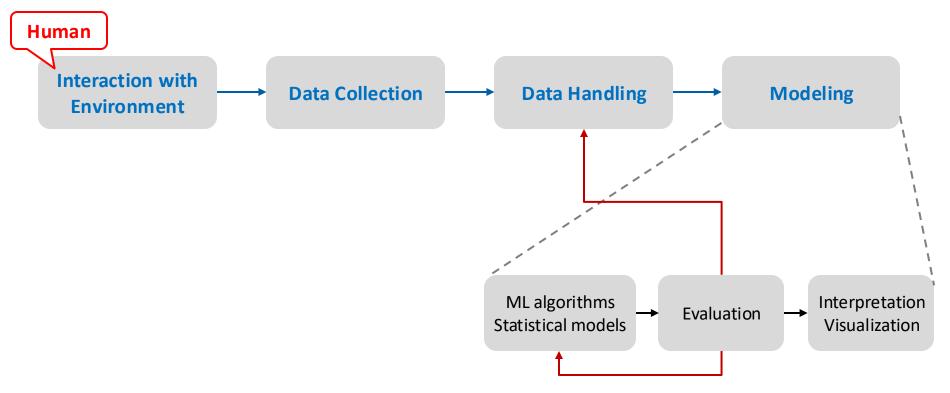
Interaction with Environment Data Collection

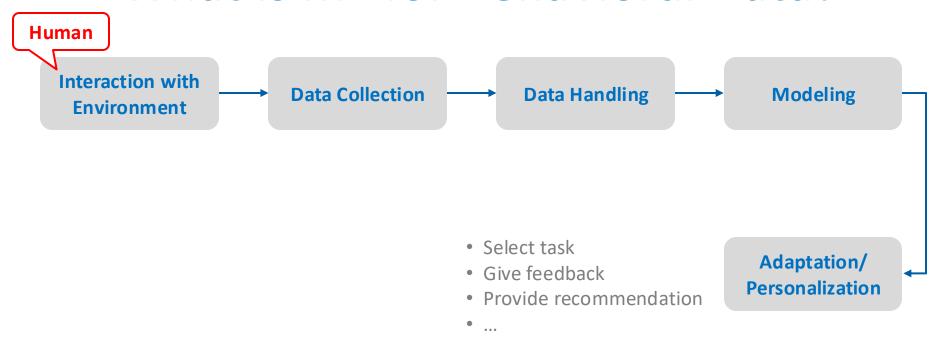
- Clickstream
- Text
- Categorical Data
- Images
- Video
- Sensor Data
- •

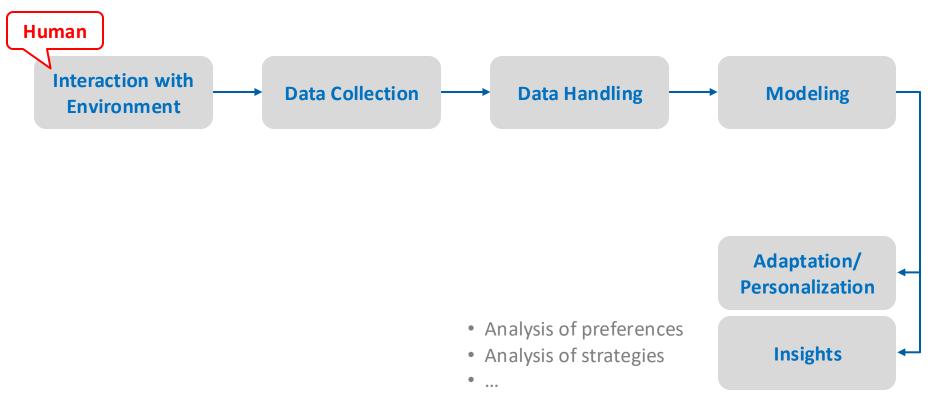


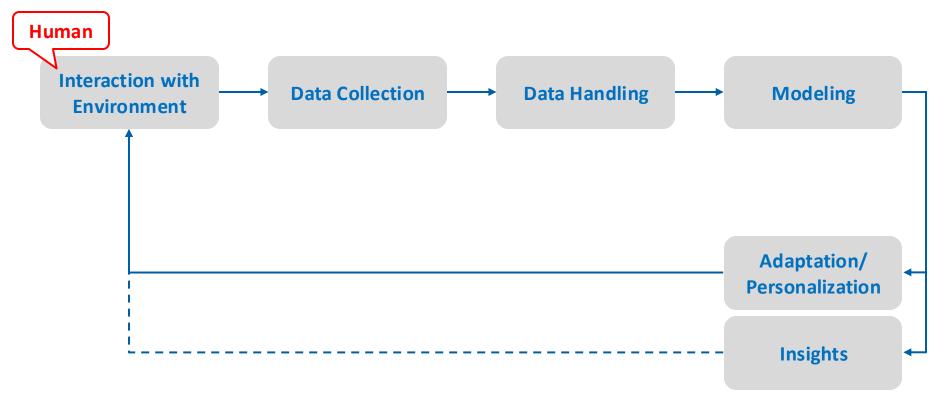












Week	Lecture/Lab
1	Introduction
2	Data Exploration
3	Regression
4	Classification
5	Model Evaluation
6	Time Series Prediction
7	Time Series Prediction

Week	Lecture/Lab
1	Introduction
2	Data Exploration —
3	Regression
4	Classification
5	Model Evaluation
6	Time Series Prediction
7	Time Series Prediction

- Exploring & visualizing data
- Time Series Exploration

Week	Lecture/Lab
1	Introduction
2	Data Exploration
3	Regression —
4	Classification
5	Model Evaluation
6	Time Series Prediction
7	Time Series Prediction

- Generalized Linear Models
- Mixture Models
- Regression for time series

Week	Lecture/Lab
1	Introduction
2	Data Exploration
3	Regression
4	Classification
5	Model Evaluation
6	Time Series Prediction
7	Time Series Prediction

- Random Forest, nearest neighbors, etc.
- Classifying time series data

Week	Lecture/Lab
1	Introduction
2	Data Exploration
3	Regression
4	Classification
5	Model Evaluation —
6	Time Series Prediction
7	Time Series Prediction

- Cross validation, bootstrap, information scores
- Error metrics & visualization

Week	Lecture/Lab
1	Introduction
2	Data Exploration
3	Regression
4	Classification
5	Model Evaluation
6	Time Series Prediction
7	Time Series Prediction

Complete pipeline for one use case:

- Data exploration
- Prediction
- Model evaluation

Week	Lecture/Lab
1	Introduction
2	Data Exploration
3	Regression
4	Classification
5	Model Evaluation
6	Time Series Prediction
7	Time Series Prediction

Supervised learning on time series:

- Probabilistic graphical models
- Neural networks: LSTM, GRU, etc.

Week	Lecture/Lab
8	Time Series Prediction
9	Unsupervised Learning
10	Spring Break
11	Unsupervised Learning
12	Ethical Machine Learning
13	Ethical Machine Learning
14	Reserve
15	Poster Presentations

Week	Lecture/Lab
8	Time Series Prediction
9	Unsupervised Learning
10	Spring Break
11	Unsupervised Learning
12	Ethical Machine Learning
13	Ethical Machine Learning
14	Reserve
15	Poster Presentations

- K-Means, Spectral Clustering
- Choosing the optimal K*
 Clustering time-series data

Week	Lecture/Lab
8	Time Series Prediction
9	Unsupervised Learning
10	Spring Break
11	Unsupervised Learning
12	Ethical Machine Learning
13	Ethical Machine Learning
14	Reserve
15	Poster Presentations

- Fairness
- Explainability

Lecture/Lab

- Monday, 13:15 16:00
- INJ 218
- Lecture + practice session
- Slides will be uploaded to our GitHub
- Jupyter Notebooks will be uploaded to our GitHub
- Recording: we will make the recordings from the past year available

Project

- Teams of 3 people
- We will provide the data sets (from EdTech Start-Ups)
- 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)
- You will do a poster presentation in the last week of the semester
- Final project (Code + Report) delivered by June 6, 2025 23:59 CET

Project (Office) Hours

- Tuesday, 16:15-17:00
- INJ 218
- Content:
 - Introduction to project tasks
 - Individual feedback meetings with teams
 - Drop-in office hours for questions regarding the lecture or project

Project Schedule

Week	Project Hours	Milestones
1	Environment setup	
2	Introduction to tasks for M2	M1: preferences on team members and start-up
3	Office hours – Individual Drop-In	
4	Introduction to tasks for M4	M2: individual exploration of selected data set
5	Office hours – Individual Drop-In	M3: selection of research question and approach
6	Individual discussion with teams	
7	Office hours – Individual Drop-In	
8	Office hours – Individual Drop-In	

Project Schedule

Week	Project Hours	Milestones
9	Team Coaching	
10	Spring Break	
11	Office hours – Individual Drop-In	M4: submission of results for first research question M5: ideas for extension (+ approach)
12	Individual Discussions with teams	
13	Office hours – Individual Drop-In	
14	Office hours – Individual Drop-In	
15	Poster Presentations	M6: poster session (in person, on campus)
16		M7: Hand in report and code base

Grading

• 50% Project

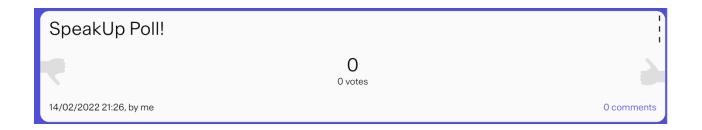
- Teams of 3 people
- All milestones are mandatory
- Attendance of all individual feedback meetings and team coaching meetings is mandatory
- 15% individual exploration (M2), 25% intermediate results (M4),
 20% poster presentation (M6), 40% final results (M7)
- 50% Final Exam (exam session)
 - Individually
 - Mix of conceptual and coding questions

Course Goals

- Explain the main machine learning approaches to personalization, describe their advantages and disadvantages and explain the differences between them
- Implement algorithms for these machine learning models
- Apply them to real-world data
- Assess / evaluate their performance

Which ML courses have you taken?

- A. Introduction to Machine Learning
- B. Machine Learning
- C. Applied Data Analysis
- D. Other



Course Prerequisites

- Probabilities and statistics
- Programming:
 - Project: Python
 - Exam: Python



Foundations of machine learning

Important Websites

- Moodle: https://moodle.epfl.ch/course/view.php?id=16434
 - Contains all important information
 - Use forum for questions
 - For more personal questions contact teaching assistants
- Project:
 - GitHub: https://github.com/epfl-ml4ed/mlbd-2025
 - EPFL Noto: https://noto.epfl.ch/

Team

Instructor



Tanja Käser tanja.kaeser@epfl.ch

Teaching Assistants







Bahar Radmehr, Seyed Parsa Neshaei, Paola Mejia bahar.radmehr@epfl.ch, seyed.neshaei@epfl.ch, paola.mejia@epfl.ch

Feedback

- We are committed to providing the best possible version of the course
- If you want to give us feedback, there will be a link on Moodle:

Feedback

We are fully committed to providing the best possible version of the course and we appreciate all constructive feedback. We are looking forward to reading your comments and improving based on them.

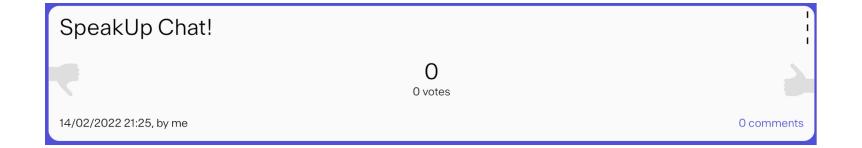
Feedback link (anonymous)

Questions?

Today

- What is ML for Behavioral Data?
- Course Logistics
- Active Learning
- Projects: EdTech StartUp(s)

Active learning – what is it?



Active learning – what is it?

- Activities that students do to construct knowledge and understanding
 - Read
 - Write
 - Explore
 - Discuss
 - **–** ...

Active learning in this course

SpeakUp

Collecting Ideas

Polls

Think — Pair - Share

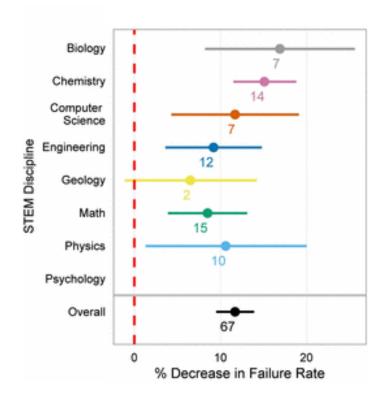
Jupyter Notebook

Demonstration

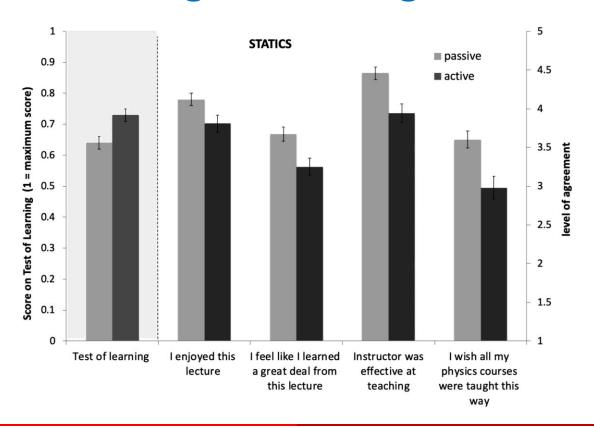
Worked examples

Small coding tasks

Active learning increases performance



Watch out: Feeling-of-Learning can deceive you!



The lecture will be interactive, thus

- we expect you to attend the lecture
- we expect you to participate in all the activities

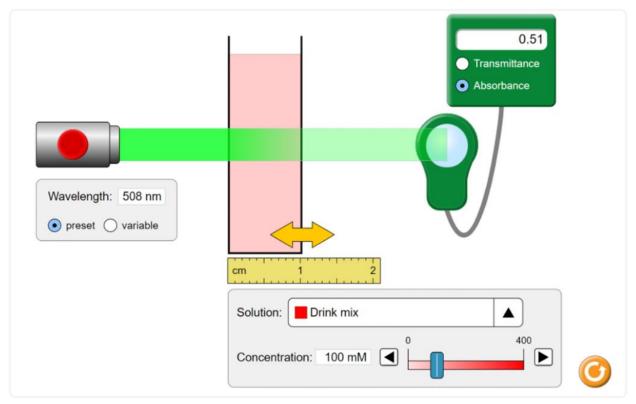
Important: bring your laptop!

Questions?

Today

- What is ML for Behavioral Data?
- Course Logistics
- Active Learning
- Project
 - Example Projects
 - EdTech StartUp(s)

Inquiry learning in simulations



Inquiry Learning – the data

- 505 apprentices
- 11 vocational schools from Switzerland



Inquiry learning – put in practice



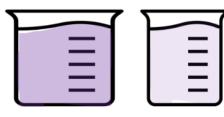
Recently, Migros has realized that some producer add artificial substances to juices to make them look more colorful. Yet, these substances can be harmful for consumers' health!

Migros has sent us samples to have us check which juices contained added artificial substances.

We can perform this check by comparing the samples to references (natural juices).



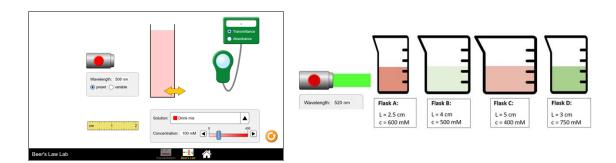




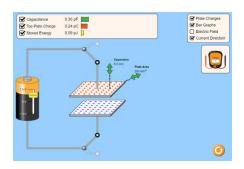
Lemon Juice

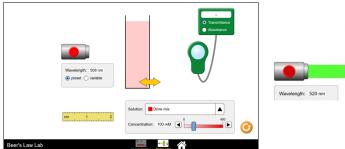
Grape Juice

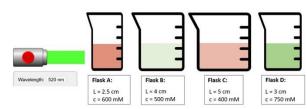
Identifying successful inquiry



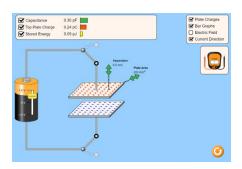
Identifying successful inquiry

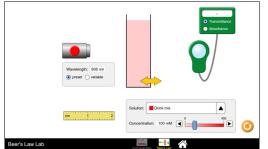


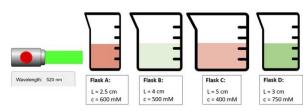


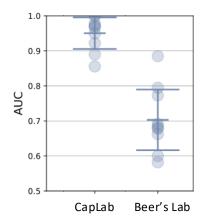


Identifying successful inquiry





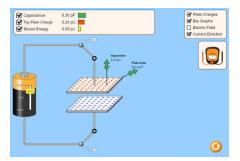


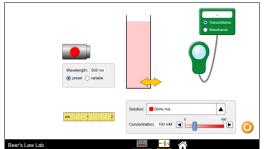


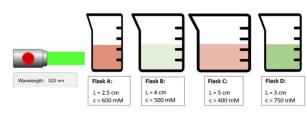
Cock, J., Marras, M., Giang, C., Käser, T. (2021). *Early Prediction of Conceptual Understanding in Interactive Simulations*, EDM, https://educationaldatamining.org/EDM2021/virtual/static/pdf/EDM21_paper_203.pdf

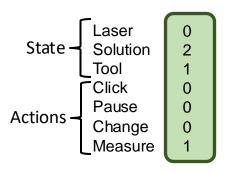


Generalizing features across domains

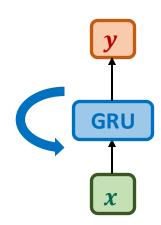






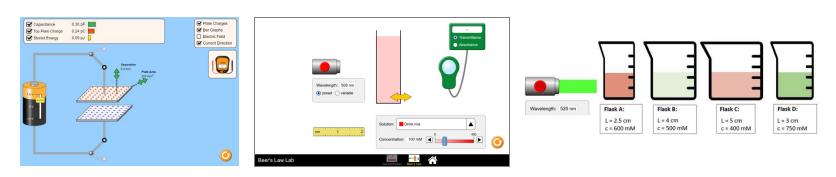


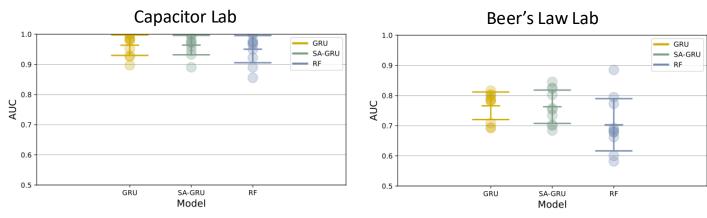
Input: State-Action combination at time step *t*





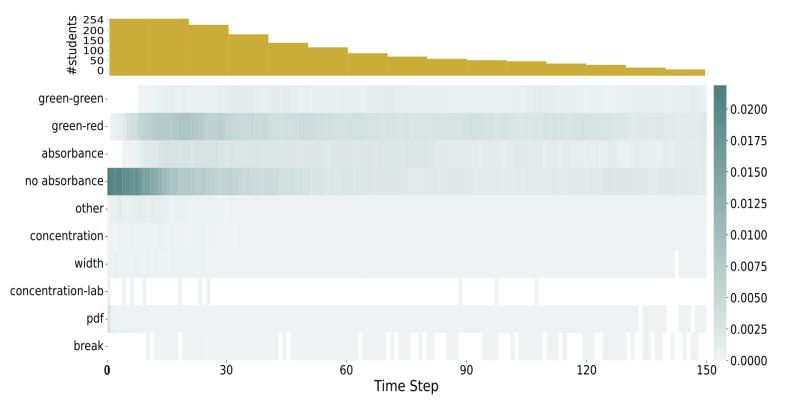
Generalizing features across domains





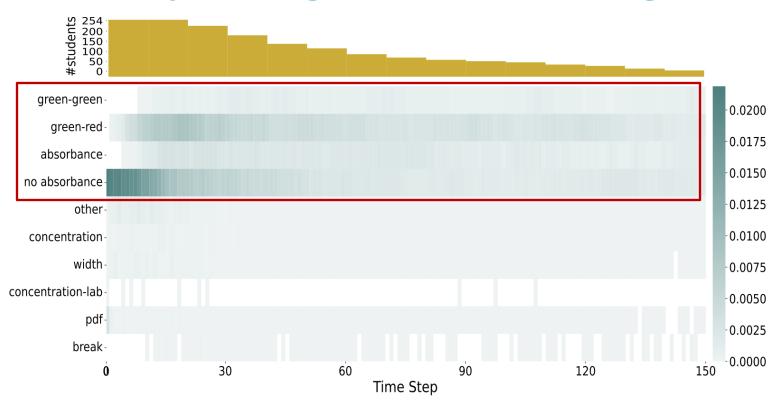
Cock, J., Marras, M., Giang, C., Käser, T. (2022). *Generalisable methods for early prediction in interactive simulations for education*, EDM, https://doi.org/10.48550/arXiv.2207.01457

Interpreting attention weights



Cock, J., Marras, M., Giang, C., Käser, T. (2022). *Generalisable methods for early prediction in interactive simulations for education*, EDM, https://doi.org/10.48550/arXiv.2207.01457

Interpreting attention weights



Cock, J., Marras, M., Giang, C., Käser, T. (2022). *Generalisable methods for early prediction in interactive simulations for education*, EDM, https://doi.org/10.48550/arXiv.2207.01457

Identifying struggling students in MOOCs



Digital Signal Processing



Villes Africaines



Venture

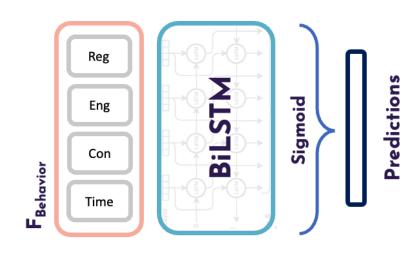
Identifying struggling students



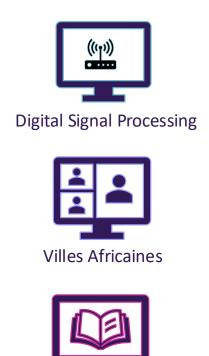
Digital Signal Processing



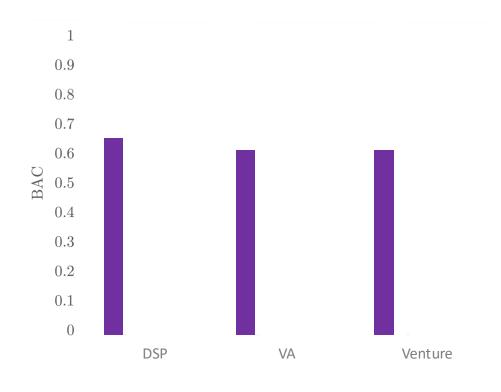




Transferring models between courses



Venture



Transfer between iterations









Transferring from a corpus









Including meta-information into training



Digital Signal Processing



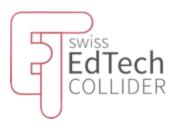
Venture



Today

- What is ML for Behavioral Data?
- Course Logistics
- Active Learning
- Project
 - Example Projects
 - EdTech StartUp(s)

Swiss "EdTech" Hub

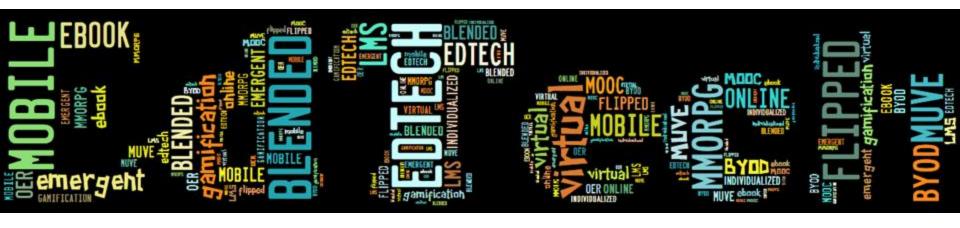


Why a Swiss "EdTech" Hub?

To support the digital transformation in **ED**ucation with **TECH**nological solutions



EdTech Market – Highly Fragmented





Large diversity in the use of technology-enhanced solutions



Early Childhood Education



Compulsory Education



Upper Secondary / Higher Education



University/VET



Continuous Training & Education



Corporate Training & Learning



Mission and Vision

- Bring players in EdTech together in one place in order to create a market place around Education and EdTech
- Focus on future learning solutions / future skills
- Long-term partnerships (not a short-term incubator)
- Help to accelerate growth
- Sustainability / long term positive impact on society



93 StartUps (status: 01.2022)





The two participating StartUps

- GoGymi (Jan Bühlmann)
- Taskbase (Léonore Guillain)

Up next...

- Detailed information regarding the project: milestones, guidelines, grading, data sets, etc. [lab session today]
- Setting up GitHub and Jupyter notebook for the lecture and project [project hours today]

Remember

- Register for the course on IS Academia
- Bring your laptop!
- You find everything on...

Moodle:

https://moodle.epfl.ch/course/view.php?id=16434