# Machine Learning Lab Course

## Organizational Meeting

Summer Term 2021

# Team

- Prof. Dr. Stephan Günnemann
- Aleksei Kuvshinov, Anna Kopetzki and Johannes Klicpera

This is a practical course (Praktikum) for **Master** students!

*Name of module: Large-Scale Machine Learning (IN2106, IN4192)*

Website: www.daml.in.tum.de/ml-lab

# Why attend our ML lab course?

1. Opportunity to **implement and apply** state-of-the-art ML algorithms.
2. Gain **hands-on experience** working on **real-world data**, solving **real-world tasks** by working on projects offered by our **industry partners** as well as academic projects.
   – successful projects might even qualify for a subsequent master thesis
3. Work on **large-scale problems** with the support of our **GPU computing resources.**

# Requirements

- Requirements for the lab course
  - **strong programming skills** (Python, deep learning frameworks Pytorch/Tensorflow, etc.)
  - strong knowledge in data mining/machine learning
  - you should have passed relevant courses (the more, the better)
    - Machine Learning for Graphs and Sequential Data, Machine Learning … (see the application form)
  - self-motivation

- Additional selection criteria
  - other **relevant** experience (projects in companies, experience as a HiWi)
    - you can send an overview of your experience to us (see end of slides)

# Organization

- Groups of 3 students
- This term we offer 3-4 different project
  - (new) for one topic 2 competing teams will be working on it
  - all groups can collaborate
- Students get access to our GPU compute servers
  - each of the servers has:
    - 4x NVIDIA GPU with 11GB RAM
    - 10-core CPU
    - 256 GB RAM
  - scale up your models and data!

# Organization

- Weekly meetings (around 2 hours)
  - **online**
  - groups present every week
  - each group should briefly report their progress and next steps
- Regular documentation of your work on wiki
- Code on git

# Projects – industrial and academic

## SIEMENS

**Few-Shot Similarity Detection in Tabular Data**

State-of-the-art approaches detect similarity through a comparison of vector representations of data points. Being able efficiently detect similarities in the data can save costs and optimize processes.

## TUM-DAML

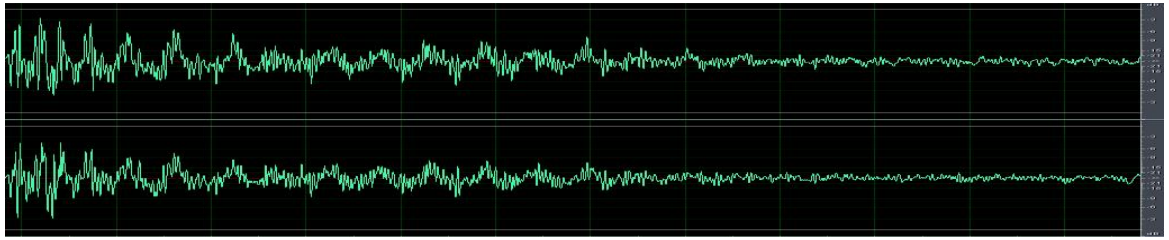
**Sample-efficient Graph Neural Networks (x2)**

- Unsolved problem: Accurate prediction with 100 samples
- Compete on molecular data
- Explore data selection, representation, models, layers, aggregation, training, …

## BMW

**Classification of Sound Signals**

Real-world setting: use mechanical noise to find out which part of the car is broken.



## CELONIS

**Motif Detection in User Trace Data**

The goal is to automatically find frequent sequential patterns, which are by definition a good candidate for automation, as a user is doing the same steps over and over again. As the data, capturing the clicks taken by the user, is never exactly the same sequence, the solution has to incorporate some edit distance to group those patterns.

**Registration via the matching system!**

Module name: Large-Scale Machine Learning (IN2106, IN4192)

**+ Fill out the application form (link is on the webpage)!**

www.daml.in.tum.de/ml-lab

- provide us with a list of your experiences with ML (courses, projects, etc.)
- please send us a short overview (bullet list, not a complete CV)

# On Friday, 12.02! Presentations with current lab participants.

bbb.rbg.tum.de/mar-g79-teu

- Denoising & Robust Classification of Sound Signals
- Transfer Learning: Day-Night Domain Shifts in Autonomous Driving
- Heat Demand Forecast for District Heating Systems
- Background-Masked Generative Models
- Generating Images with Normalizing Flows

## Questions in chat and live after each presentation!