



## Master thesis (30 ECTS)

### Working Title

Approaches to synthetic data generation based on real geotechnical datasets

### Project objectives

In machine learning, the development of reliable methods capable of yielding accurate predictions and classifications is highly dependent on the quality of the data. One of the main difficulties in developing accurate machine learning-based models in geoen지니어ing is the extreme sparsity and imbalance in data. However, different procedures of synthetic data generation (e.g. Oversampling, GANs) can help overcome the limitations arising from training machine learning models using an imbalanced dataset.

This master thesis starts with a literature research on generative adversarial networks (GANs) with a focus on conditional GANs (CGANs). You will be provided with the code of an existing GAN for the generation of synthetic TBM data. Your task will be to modify it into a conditional GAN, capable of generating synthetic TBM data based on conditions (i.e., rock mass classification). With this CGAN it should then be possible to generate synthetic data for underrepresented classes (i.e., data from fault zones) to overcome the imbalance problem.

### Student has enthusiasm for

machine learning applications for geotechnical purposes, geotechnical data, programming

### Start

August 2022

### Project term (min. / max.)

6 / 9 months

### Coop. with external institution

no

### Possibility of remuneration

no

### Contact person(s)

Paul Unterlaß – unterlass@tugraz.at