Machine Learning and Social Impact Workshop – Overview

The SIMLab will be hosting a workshop for UVU students in March/April 2022. It is divided into three sessions, to be held on March 25, April 1st, and April 8. Each session starts at 3 pm and is 90 minutes long.

The workshop introduces students to the fields of Machine Learning and Social Impact Measurement. The goal is to equip them with a working knowledge of concepts and tools that can be used to assess the social effects of programs carried out by governments, businesses, NGOs and other institutions.

The workshop will focus on quantitative methods, but it is not designed to delve into technical details. All the data handling and estimations will be done in R, a programming language and statistical package in widespread use in the academic and business communities.

At the end of the workshop students will be able to:

- Understand and employ basic R code.
- Inspect data utilizing visualization techniques.
- Apply select machine learning methods to data sets to make predictions.
- Compare different methods in terms of the quality of their predictions.
- Understand the causal inference problem and how it affects social impact measurement.
- Carry out matching procedures to reduce the effect of confounders.

- Estimate social impact (treatment effects) and interpret the results.
- Understand and interpret social impact reports published by businesses, governments and non-profits.
- Help design studies to assess social impact.

These are valuable and marketable skills that students can showcase when they participate in the job market. In addition to non-profits interested in evaluating the social impact of their activities and programs, they are in high demand by businesses eager to show corporate social responsibility. Managers are increasingly looking for ways to measure the social impact of their businesses, and investors make decisions taking that impact into account.

Session 1 (March 25):

Presentation (60 minutes):

- Introduction
- Data visualization
- Summary measures

Lab (30 minutes):

- Introduction to R
- R code for:
    - Importing data
    - Creating plots

Session 2 (April 1st):

Presentation (60 minutes):

- Machine Learning methods
    - Linear Regression
    - Classification trees

    - Validation
- Definition and examples of social impact
- Causal relationships
    - Randomized trials
    - Observational studies

Lab (30 minutes):

- R code for:
    - Linear regression
    - Classification trees
    - Validation

Session 3 (April 8):

Presentation (60 minutes):

- Preliminary analysis of treatment effects
- Matching procedure
- How confounders blur the line
- Estimation of treatment effects using matched data set

Lab (30 minutes):

- R code for
    - Boxplot analysis
    - Matching
    - Estimating treatment effects