

Introduction to Data Science

BDS

2024-2025

Ernst C. Wit

Schedule

Monday 9:00 – 10:45, room C1.03, East Campus

Instructor

Lecturer: Ernst C. Wit, wite@usi.ch, D5.08

Assistant: Lovnesh Bhardwaj, lovnesh.bhardwaj@usi.ch

Office hour: Monday 11:00 – 12:00 (outside C1.03, right after class)

Course description

What is data science? This class will give you a first introduction to this interdisciplinary field. From an epistemological perspective, the interplay of reality, data, models and knowledge matured over the period from the 17th until the 20th century. While this was going on, practical demands from taxation, criminality and hygiene pushed the envelope of how noisy empirical evidence translates into firm knowledge. In the 21st century the scale in which certain types of data have grown meant that modelling paradigms shifted from explanatory to purely predictive. In this course, we answer 4 fundamental questions: (i) what is data, (ii) why do we need data, (iii) how do we get data and (iv) how do we use data.

Evaluation

The final grade will be a weighted average of the midterm exam and the final exam:

$$\text{Final grade} = 0.3 \times \text{midterm} + 0.7 \times \text{final exam}$$

Programming

The course will use the R computational environment, besides Python.

Background reading

- Kelleher, John D., and Brendan Tierney. Data.science. MIT press, 2018.
- Stigler, Stephen M. Statistics.on.the.table;The.history.of.statistical.concepts.and.methods. Harvard University Press, 2002.

Plagiarism

Principle: A student should never take someone else's material and present it as his or her own. Doing so means committing plagiarism.

The term "material" here refers to ideas, words, code, or any other piece of intellectual work, including suggestions and corrections regarding the student's own work. Whenever external material is used verbatim or with modifications, the student must clearly identify the external material and acknowledge its source. Failing to do so means committing plagiarism. In any case in which external material is used by a student in homeworks and/or exams, the work will be evaluated based on the added value produced by that student.

Committing plagiarism on an assignment or an exam will result in failing that assignment or that exam, and the Dean's office will be notified. Penalties may be escalated in accordance with the regulations of the Faculty of Informatics.

Outline of Course

- Week 1: What is data?
- Week 2: Why using data?
- Week 3-6: How to get data?
 - Population and samples
 - Random sampling and probability
 - Biases and experimental design
- **Week 7: Midterm exam**
- Week 8-13: What to do with data
 - Exploration
 - Prediction
 - Explanation and Causality
 - Pitfalls
- **Exam period: Final exam**