Discover Track

The purpose of this track is to understand the scope and possible impact of data science. We discuss different views on data science, and on where and when data science can offer added value (and when not). Also, students learn to execute a data-science project by using the CRISP-DM process, and they get a hands-on, five days’ primer in machine learning.

Introduction (2 days)
- Creating value from data: towards a data-driven organization
- Entrepreneurship and intrapreneurship
- Data science in practice
- Executing a data-science project following the CRISP-DM process model: 1 day hands-on hackathon!

Executing a machine-learning project (5 days)
Participants are guided through a realistic data-science project where machine learning is applied to tackle a prediction problem. This case covers the phases of Business Understanding, Data Understanding, Data Preparation, Modeling and Evaluation of the CRISP-DM model. In this module, participants learn the basic techniques in machine learning:
- Practical machine learning in Python.
- Linear classifiers, nearest-neighbor classifiers, and decision tree classifiers.
- Random forests, support-vector machines, unsupervised learning.
- Cross-validation techniques: K-Folds, Leave One Out.
- Classification Metrics: Accuracy, Confusion Matrix, Precision, Recall.

The goal is that students can code an end-to-end machine learning pipeline that trains several classifiers and chooses the classifiers with the highest classification performance.