CAS Applied Data Science Study Plan

UNIVERSITÄT BERN

02.02.2021

The Applied Data Science Course is a university study program leading to a "Certificate of Advanced Studies in Applied Data Science" awarded by the University of Bern as laid out in the regulations of the Faculty of Science of the University of Bern for the certificate course Applied Data Science of 15. April 2021.

1. Course objectives

About

In most disciplines, the amount of available data is growing rapidly. Handling and processing the data remains a challenge and the demand for data science expertise is increasing dramatically across all domains. This CAS offers internal and external students, researchers, and employees the opportunity to complement their domain-specific core competences with a formal deepening and broadening of their applied data science knowledge and skills. The format is designed to align with the participants' main study and professional activities. The teaching and learning approaches are team and discussion oriented and designed to develop practical competency.

Competence objectives

The course competences are developed in six modules. At the end graduates will be:

1. familiar with different data sources and data types, and able to develop data management plans;
2. able to describe, extract and present scientific knowledge from data using statistical methods;
3. able to process data with machine learning tools and methods;
4. familiar with best practices for data management, analytics, and science;
5. able to analyse and communicate data science challenges and use a wide range of data science tools and methods;
6. able to train and apply deep neural networks for a range of tasks;
7. able to perform data science projects with machine learning.

2. Duration, outcomes and objectives of the modules

Duration and scope

The CAS consists of 18 days of courses (126 hours attendance) and 16 ECTS credit points (total effort about 480 hours) whereof 4 ECTS result from the CAS project work.
Attendance at one introductory *About the CAS Applied Data Science* meeting before course admission is mandatory.

Module 1

Data acquisition and management

In which participants learn to understand data sources and types, how to design data flows and models, and manage and visualise data algorithmically

ECTS	2 ECTS credit points (incl. self-study and report)	**Duration**	3 days = 21 hrs. attendance
Assessment	Conceptual Design Report	**Req. attendance**	80%
Learning outcomes	Graduates will • know data sources, types, sizes, storage systems, common infrastructures and acquisition techniques; • be able to analyse data flows; • be able to develop conceptual, logical and physical data models; • know how to use databases and MySQL; • be able to design data management plans; • will be able to collect, manage and visualise data.		
Learning objectives	• Definitions for data and science • Types of sensors and data • Quality of data • Infrastructures for data • Data flows and models • Algorithmic data management, filtering, cleaning and visualization • MySQL		
Learning and teaching methods	• Online platform • Lectures • Hands-on tutorials • Discussions • Report writing		
Prerequisites	None		
Teaching language	English		

Module 2

Statistical Inference for Data Science

In which we recall or become familiar with typical statistical concepts for describing and analysing data.

ECTS	2 ECTS credit points (incl. self-study and project)	**Duration**	3 days = 21 hours attendance
Assessment	Poster presentation	**Req. attendance**	80%
Learning outcomes	Graduates will/can • know the importance of statistical inference for data science and where to apply it; • understand the basic theoretical concepts;		

	• apply the theoretical concepts with software on data; • draw scientific conclusions from statistical analysis results.
Learning objectives	• Probabilities and distributions • Significance and p-values • Estimation paradigms (least squares, maximum likelihood) • Uncertainties (standard errors, confidence intervals) • Basic regression • Hypothesis testing
Learning and teaching methods	• Online platform • Lectures • Inverted classroom with hands-on tutorials • In-person classes for discussions, feedback and deepening of knowledge • Presentation of results in poster session
Prerequisites	• Basic experience with a programming language providing statistical libraries (Python, R, …) • Knowledge in mathematics and programming at the level of a university introduction lecture
Teaching language	English

Module 3

Data Analysis with Machine Learning

In which participants learn about standard machine learning techniques and how to apply them.

ECTS	2 ECTS credit points (incl. self-studies and project)	**Duration**	3 days = 21 hrs attendance
Assessment	Presentation	**Req. attendance**	80%
Learning outcomes	Graduates will • know the basic concepts of machine learning: training, testing, over/ and underfitting, performance measures; • can apply software to perform linear regression, decision trees and random forest, neural networks (ML software is accessed via Python.		
Learning objectives	• Machine learning overview, classification and regression • Linear model and logistic regression • Trees and forests • PCA and embeddings • Neural networks		
Learning and teaching methods	• Online platform • Lectures • Hands-on tutorials • Project work • Presentations		
Prerequisites	• Python programming experience		

	• Knowledge in mathematics and programming at the level of an introductory lecture
Teaching language	English

Module 4

AI philosophy, ethics and peer exchange

In which participants reflect upon philosophical and ethical aspects of Artificial Intelligence (AI) and share and discuss data science experiences

ECTS	2 ECTS credit points (incl. self-studies	**Duration**	3 days = 21 hours attendance
Assessment	presentation	**Req. attendance**	80%
Learning outcomes	Graduates will/can • know the philosophical conceptions of AI; • know main moral challenges related to AI and discuss solutions from the perspective of ethics; • use distributed VCS software and platforms; • present and discuss data science topics.		
Learning objectives	• History of AI and philosophical conceptions of AI • Ethical challenges due to AI in the perspective of ethical theories, machine ethics		
Learning and teaching methods	• Online platform • Seminar discussions • Oral presentations		
Prerequisites	• Data analysis experience		
Teaching language	English		

Module 5

Best Practices for Data Science

In which best practices and tools for data and code management, resource usage and collaboration are presented, discussed and applied for the CAS projects

ECTS	2 ECTS credit points (incl. self-study + module work)	**Duration**	3 days = 21 hours attendance
Assessment	Project presentation	**Req. attendance**	80%
Learning outcomes	Graduates will • know best practices for scientific computing; • be able to use distributed Version Control Software (VCS); • know basic cyber security challenges; • know legal aspects regarding intellectual property and licensing;		

	• be able to document and publish software projects with VCS and websites.
Learning objectives	• Best practices for scientific computing • Collaborative distributed version control, code review (course tool is Git and related platforms) • Basic cyber security • Intellectual property and licenses Documentation
Learning and Teaching methods	• Online platform • Lectures • Hands-on tutorials • Inverted classroom • Project work • Presentations
Prerequi-sites	• Python programming experience • Knowledge in mathematics and programming at the level of an introductory lecture
Teaching language	English

Module 6

Deep Learning

In which participants learn about deep learning techniques and how to apply them.

ECTS	2 ECTS credit points (incl. self-studies and project)	**Duration**	3 days = 21 hours attendance
Assessment	Project presentation	**Req. attendance**	80%
Learning outcomes	Graduates will • understand different models for supervised, unsupervised and reinforced learning; • be able to apply software for deep learning (course software is TensorFlow); • be able to train, tune and assess deep networks.		
Learning objectives	• TensorFlow • Training of deep models • Tuning of deep models • Model performance assessment		
Teaching methods	• Online platform with course material • Theoretical and applied lectures • Hands-on tutorials • Project work with machine learning software • Project report • Oral presentations		
Prerequi-sites	• Python programming and data handling • Basic linear algebra, analysis, statistics and machine learning		
Teaching language	English		

Project Work

CAS Project Work

In which participant teams perform and present a comprehensive data science project based on all CAS modules.

ECTS	4 ECTS credit points	**Duration**	120 hours
Assessment	Written report	**Req. attendance**	No attendance
Learning outcomes	Graduates will • be able to perform and present a data science project with machine learning.		
Learning objectives	• Selection of objectives from all modules		
Teaching methods	• Project work with remote supervision • Written report		
Prerequisites	• Module 1 - 6		
Teaching language	English		

3. Performance assessment

Assessment

The performance assessment includes 80% attendance and passed performance assessments from all modules and the CAS project report.

a) For Module 1 a written data science concept design report is assessed.

b) For Module 2 a poster presentation is assessed.

c) For Module 3, 4, 5 and 6 oral presentations are assessed.

For the CAS project a written report is assessed.

4. Final regulations

Entry into force

The present plan shall enter into force on 1. May 2021.

02.02.2021

Released by the program management:

Prof. Dr. Thomas Wihler (MAI)

15.04.2021

Adopted by the Faculty of Science, University of Bern

The Dean

Prof. Dr. Zoltan Balogh