**DIT852  Introduction to Data Science, 7.5 credits**

*Introduktion till Data Science, 7,5 högskolepoäng*  
*Second Cycle*

---

**Confirmation**

This course syllabus was confirmed by Department of Computer Science and Engineering on 2018-02-02 and was last revised on 2019-12-02 to be valid from 2020-08-31, autumn semester of 2020.

*Field of education:* Science 100%  
*Department:* Department of Computer Science and Engineering

**Position in the educational system**

The course is a compulsory course within the Applied Data Science Master's Programme.

The course can be part of the following programme: 1) Applied Data Science Master's Programme (N2ADS)

- **Main field of studies**  
  Data Science

- **Specialization**  
  A1N, Second cycle, has only first-cycle course/s as entry requirements

**Entry requirements**

To be eligible to the course, the student should have a Bachelor's degree. Specifically, at least 7,5 credits of successfully completed courses in programming are required.

Applicants must prove knowledge of English: English 6/English B or the equivalent level of an internationally recognized test, for example TOEFL, IELTS.

**Learning outcomes**

On successful completion of the course the student will be able to:
Knowledge and understanding
- describe fundamental types of analysis problems arising in data science
- give examples of data science applications from different contexts
- describe basic mathematical concepts in data science and understand how they can be realized in data science applications

Competence and skills
- use appropriate libraries and programming techniques to implement basic transformations, visualizations and analyses of example data
- apply simple machine learning methods implemented in a standard library
- identify appropriate types of analysis problems for concrete applications

Judgement and approach
- reflect on inherent limitations of results of data science methods
- critically analyze data science applications with respect to ethics and privacy

Course content
The course gives an introduction to applied data science using case studies from different application domains and a primer in using Python for data science.

In the course, the following topics will be covered:
- mathematical and statistical concepts useful in data science, such as basic set theory, logic, functions, and probability theory,
- case studies of data science applications displaying a range of application areas and fundamental types of analysis problems,
- a brief introduction to working in a Unix/Linux environment and programming to implement basic transformations, visualizations and analyses,
- use of a machine learning library to analyse data,
- demonstration of inherent limitations of computational analyses with examples,
- implications on privacy and ethical considerations.

Sub-courses
1. Written examination (*Skriftlig tentamen*), 4 credits
   Grading scale: Pass (G) and Fail (U)

2. Assignments (*Inlärningsuppgifter*), 3.5 credits
   Grading scale: Pass (G) and Fail (U)
Form of teaching
Lectures, computer sessions, and exercise sessions.

Language of instruction: English

Assessment
The course is examined by an individual written hall examination, as well as mandatory written assignments submitted as written reports, some of which will be carried out individually and others in groups of normally 2-4 students.

There will be non-obligatory individual assignments which grant bonus points for the written examination. These bonus points are valid for the whole academic year.

If a student, who has failed the same examined component twice, wishes to change examiner before the next examination, a written application shall be sent to the department responsible for the course and shall be granted unless there are special reasons to the contrary (Chapter 6, Section 22 of Higher Education Ordinance).

In cases where a course has been discontinued or has undergone major changes, the student shall normally be guaranteed at least three examination occasions (including the ordinary examination) during a period of at least one year from the last time the course was given.

Grades
The grading scale comprises: Pass (G) and Fail (U).
A Pass grade (G) for the entire course requires at least a Pass grade for all sub-courses.

Course evaluation
The course is evaluated through meeting after the course between teachers and student representatives. Further, an anonymous questionnaire is used to ensure written information. The outcome of the evaluations serves to improve the course by indicating which parts could be added, improved, changed or removed.

Additional information
The programming language Python will be used in this course and in some other courses in the Applied Data Science Master's programme for examples and assignments.

Course literature to be announced the latest 8 weeks prior to the start of the course.

The course replaces the course DIT851, 7.5 credits. The course cannot be included in a degree which contains DIT851. Neither can the course be included in a degree which is based on another degree in which the course DIT851 is included.

This course cannot be included in a degree which contains DIT405 Introduction to data
science and AI.