



ID2214 Programming for Data Science 7.5 credits

Programmering för data science

This is a translation of the Swedish, legally binding, course syllabus.

Establishment

On 04/21/2020, the Head of the EECS School has decided to establish this official course syllabus to apply from autumn semester 2020, registration number J-2020-0592.

Grading scale

A, B, C, D, E, FX, F

Education cycle

Second cycle

Main field of study

Computer Science and Engineering

Specific prerequisites

Completed course in programming equivalent to ID1018/DD1310/DD1311/DD1312/DD1314/DD1315/DD1316/DD1318/DD1331/DD1337/DD100N.

Active participation in a course offering where the final examination is not yet reported in LADOK is considered equivalent to completion of the course. This applies only to students who are first-time registered for the prerequisite course offering or have both that and the applied-for course offering in their individual study plan.

Language of instruction

The language of instruction is specified in the course offering information in the course catalogue.

Intended learning outcomes

Having passed the course, the student should be able to

- account for and discuss the application of i) technologies to convert data to an appropriate format for data analysis ii) algorithms to analyse data through supervised and unsupervised machine learning as well as iii) technologies and performance metrics for evaluation of data analysis results
- implement and apply i) technologies to convert data to an appropriate format for data analysis ii) algorithms for supervised and unsupervised machine learning as well as iii) technologies and performance metrics for evaluation of data analysis results.

Course contents

Syntax and semantics for programming languages that are particularly suited for data science, e.g., Python.

Routines to import, combine, convert and make selection of data.

Algorithms for handling of missing values, discretisation and dimensionality reduction.

Algorithms for supervised machine learning, e.g., naïve Bayes, decision trees, random forests.

Algorithms for unsupervised machine learning, e.g., k-means clustering.

Libraries for data analysis.

Evaluation methods and performance metrics.

Visualisation and analysis of results of data analysis.

Examination

- INL1 - Assignment, 4.5 credits, grading scale: A, B, C, D, E, FX, F
- TEN1 - Examination, 3.0 credits, grading scale: A, B, C, D, E, FX, F

Based on recommendation from KTH's coordinator for disabilities, the examiner will decide how to adapt an examination for students with documented disability.

The examiner may apply another examination format when re-examining individual students.

If the course is discontinued, students may request to be examined during the following two academic years.

Written examination. Written assignments.

Ethical approach

- All members of a group are responsible for the group's work.
- In any assessment, every student shall honestly disclose any help received and sources used.
- In an oral assessment, every student shall be able to present and answer questions about the entire assignment and solution.