COURSE SYLLABUS
Data Science Programming, 7.5 credits

Data Science Programming, 7,5 högskolepoäng

Course Code:	TDPS22	**Education Cycle:**	Second-cycle level
Confirmed by:	Dean Mar 1, 2021	**Disciplinary domain:**	Technology
Revised by:	Director of Education Oct 28, 2021	**Subject group:**	DT1
Valid From:	Jan 1, 2022	**Specialised in:**	A1F
Version:	2	**Main field of study:**	Computer Science

Intended Learning Outcomes (ILO)

After a successful course, the student shall

Knowledge and understanding
- display knowledge of notebook environments for writing, testing and debugging code
- demonstrate comprehension of data management and analysis in Python and R
- demonstrate comprehension of statistics, machine learning and model evaluation in Python and R
- show familiarity with the range of data science programming environments

Skills and abilities
- demonstrate the ability of writing well-structured data science programs in Python and R
- demonstrate skills of producing high quality data visualizations in Python and R

Judgement and approach
- demonstrate the ability to have good judgement with respect to selecting and evaluating programming constructs for solving data science problems

Contents

The course is a basic course in data science programming with Python and R. The course covers basic language features and concepts, including core libraries for data science programming, such as data management and augmentation, data analysis and visualization, machine learning and model evaluation, alternating theory with practice. After completing the course, the student shall have acquired a basic, but broad, knowledge in the field of data science programming. Specifically, the student should understand and be able to apply all theoretical concepts covered.

The course includes the following elements:
- Syntax and Semantics: basic language features for the programming languages Python and R
- Data Management: importing, exporting, transforming, representing and manipulating data
- Data Augmentation: missing value imputation, discretisation and dimensionality reduction
- Data Analysis and Visualization: libraries for statistical data analysis and visualization

- Machine Learning: libraries for supervised and unsupervised machine learning
- Evaluation and Performance: evaluation metrics and significance testing
- Overview of other (besides Python and R) Data Science Programming environments

Type of instruction

The teaching in the course consists of lectures, quizzes, assignments, workshops and tutoring.

The teaching is conducted in English.

Prerequisites

Passed courses at least 90 credits within the major subject Computer Engineering, Electrical Engineering (with relevant courses in Computer Engineering), or equivalent, or passed courses at least 150 credits from the programme Computer Science and Engineering, and completed courses Data Science, 7,5 credits and Machine Learning, 7,5 credits, or equivalent. Proof of English proficiency is required.

Examination and grades

The course is graded 5,4,3 or Fail.

Registration of examination:

Name of the Test	Value	Grading
Examination[I]	4.5 credits	5/4/3/U
Assignments	3 credits	U/G

[I] Determines the final grade of the course, which is issued only when all course units have been passed.

Course literature

The literature list for the course will be provided 8 weeks before the course starts.

Title: Learning Python, 5th ed, 2013.
Authors: Lutz, M.
Publisher: O'Reilly Media.
ISBN: 978-1449355739

Title: Python Data Science Handbook: Essential Tools for Working with Data, 1st ed, 2016.
Authors: VanderPlas, J.
Publisher: O'Reilly Media.
ISBN: 978-1491912058

Title: Hand-On Machine Learning with Scikit-Learn, Keras and TensorFlow: Concepts, Tools and Techniques to Build Intelligent Systems, 2nd ed, 2019.
Authors: Géron, A.
Publisher: O'Reilly Media.
ISBN: 978-1492032649

Title: Hands-On Programming with R: Write Your Own Functions and Simulations, 1st ed, 2014.
Authors: Grolemund, G.
Publisher: O'Reilly Media.
ISBN: 978-1449359010

Title: R for Data Science: Import, Tidy, Transform, Visualize, and Model Data, 1st ed, 2017.
Authors: Wickham, H. and Grolemund, G.
Publisher: O'Reilly Media
ISBN: 978-1491910399

Title: Machine Learning with R, the Tidyverse, and Mlr, 1st ed, 2020.
Authors: Rhys, H.I.
Publisher: Manning Publications.
ISBN: 978-1617296574