Data Engineering Technologies (online)

Master of Science

Subject-specific Examination Regulations for Data Engineering Technologies (DET)

The subject-specific examination regulations for Data Engineering Technologies (DET) are defined by this program handbook and are valid only in combination with the General Examination Regulations for Master degree programs ("General Master Policies").

This handbook also contains the program-specific Study and Examination Plan (in chapter 2.2).

Upon graduation students in this program will receive a Master of Science (MSc) degree with a scope of 120 ECTS credit points (CP) (for specifics see chapter 3 of this handbook).

Valid for all students starting their studies in Fall 2024

Version	**Valid as of**	**Decision**	Details
Fall 2024 – V1	Sept 01, 2024	Jan 31, 2024	Originally approved by the Academic Senate

Contents

1 Program Overview 1

1.1 Concept 1

1.2 Qualification Aims 1

1.2.1 Educational Aims 2

1.2.2 Intended Learning Outcomes 2

1.3 Online teaching and learning 3

1.3.1 General Framework 3

1.3.2 Student Workload 3

1.3.3 Academic Tutors 4

1.3.4 Assessment and Grading 4

1.3.5 Learning Management Software 4

1.4 Target Audience 4

1.5 Career Options 5

1.6 Admission Requirements 5

1.7 More information and contacts 6

2 The Curriculum 7

2.1The Curriculum at a Glance 7

2.2. Study and Examination Plan 8

2.3. Core Area (45 CP) 10

2.4.Methods Area (22.5 CP) 10

2.5. Foundation Area (22.5 CP) 10

2.6. Master Thesis (30 CP) 11

3 Data Engineering Technologies Graduate Program Regulations 12

3.1. Scope of these Regulations 12

3.2. Degree 12

3.3. Graduation Requirements 12

3.4 Other Program-specific Policies & Practices 12

4 Modules Description 13

4.1 Core Area (45 CP) 13

4.1.1 Big Data Challenge for DET 13

4.1.2 Data Analytics 16

4.1.3 IT Law 18

4.1.4 Ethics and the Information Revolution 20

4.1.5 Data Security and Privacy 22

4.1.6 Statistical and Machine Learning 24

4.1.7 Data Acquisition Technologies 26
4.1.8 Image Processing 28
4.2 Methods Area (22.5 CP) 30
4.2.1 Data Base Management Tools in Python 30
4.2.2 Advanced Data Bases 32
4.2.3 Parallel and Distributed Computing 34
4.3 Foundation Area (22.5 CP) 36
4.3.1 Mathematics for Graduate Students 36
4.3.2 Text Analysis and Natural Language Processing 38
4.3.3 Visual Communication and Data Storytelling 40
4.5 Master Thesis (30 CP) 42
5 Appendix 44
5.1 Intended Learning Outcomes Assessment Matrix 44

1 Program Overview

1.1 Concept

Today we are “drowning in data and starving for information”, while acknowledging that “data is the new gold”. However, deriving value from all the data now available requires a transformation in data analysis, in how we see, maintain, share and understand data. Data Engineering is an emerging profession concerned with the task of acquiring large collections of data and extracting insights from them. It is driving the next generation of technological innovation and scientific discovery, which is expected to be strongly data-driven.

This consecutive, application-oriented online graduate program in Data Engineering Technologies offers a fascinating and profound insight into the methods and technologies of this rapidly growing area. The program combines the big data aspects of “Data Analytics” as well as of “Data Science” with the technological challenges of data acquisition, curation, and management via databases and warehouses, big data pipelines and cloud computing. Thus, the program provides the essentials for paving the way to a successful career in data engineering: computer skills and mathematical understanding paired with acquiring experience in selected application fields.

The program is embedded into the School of Computer Science and Engineering at Constructor University. This school investigates the mobility of people, goods, and information. Even though the Data Engineering program is centered in the School of Computer Science and Engineering, it includes contributions from and supports applications in the two other research schools: The School of Science, and the School of Business, Social & Decision Sciences.

Moreover, the Data Engineering Technologies program attracts students with diverse career goals, backgrounds, and prior work experience. The graduate program in Data Engineering Technologies is thus tailored to a diverse student body (see also Section 1.3). It is open to the 21-year-old student who has just graduated with a Bachelor degree, as well as a person who already has been employed in a data-intensive company and who wants to keep up with current data engineering practices.

1.2 Qualification Aims

The program is an online program with optional blended elements, e.g., in summer. Lectures incorporate asynchronous material and primarily follow a flipped classroom model, i.e., including application components in the spirit of problem-based- as well as project-based-learning. Practical components, particularly labs, projects, and thesis are based on remote access and distributed development. Tutoring includes virtual study groups, peer evaluation and mentoring by faculty. Performance evaluations are conducted as online e-exams.

The remote work aspects include collaborative software development and remote access to physical devices for, e.g., control, monitoring and maintenance. Due to the aspects of independent, self-governed knowledge acquisition, the students are prepared for life-long learning, where additional knowledge and skills need to be acquired or updated in a regular fashion, especially in Data Engineering Technologies.

1.2.1 Educational Aims

The program aims to provide an in-depth understanding of the essential aspects of data-based decision-making and the skills required to apply and implement these powerful methods in a successful and responsible manner. Apart from the necessary programming skills, this comprises:

- methods of data acquisition both from the internet and from sensors;
- methods to efficiently store and access data in large and distributed data bases;
- statistical model building including a wide range of data mining methods, signal processing, and machine learning techniques;
- visualization of relevant information;
- construction and use of confidence intervals, hypothesis testing, and sensitivity analyses;
- know the legal foundations of data engineering;
- methods to ensure data security and privacy;
- awareness of the societal and ethical implications of digitization;
- scientific qualification;
- competence to take up qualified employment in Data Engineering;
- competence for responsible involvement in society;
- personal growth.

1.2.2 Intended Learning Outcomes

By the end of the online program students will be able to:

1. critically assess and creatively apply technological possibilities and innovations driven by big data;
2. acquire data from sensors and use microcontrollers to process data and to transmit them to databases on servers or the internet in general;
3. set up and use databases to efficiently and securely manage and access large amounts of data;
4. apply statistical concepts and use statistical models in the context of data analytics;
5. use, adapt and improve visualization techniques to support data-based decision-making;
6. design, implement and exploit various representations of data for classification and regression including supervised machine learning methods and core ideas of deep learning;
7. apply and critically assess data acquisition methods and analytical techniques in practical scenarios in organizations and companies;
8. independently investigate complex problems and undertake scientific or applied research into a specialist area utilizing appropriate methods, also taking methods and insights of other disciplines into account;
9. professionally communicate their conclusions and recommendations, the underlying information and their reasons to both specialists and non-specialists, both clearly and unambiguously based on the state of research and application;
10. assess and communicate social, scientific and ethical insights that also derive from the application of their knowledge and their decisions;
11. engage ethically with the academic, professional and wider communities and actively contribute to a sustainable future;

12. take responsibility for their own learning, personal development, and role in society, evaluating critical feedback and self-analysis;
13. adhere to and defend ethical, scientific and professional standards.

1.3 Online teaching and learning

1.3.1 General Framework

Constructor University online study programs focus on the holistic learning success of students and offer a variety of synchronous and asynchronous formats that align with problem- and project-based learning.
The online master program in DET applies proven and effective teaching and learning modalities that engage distance learners and support a vibrant learning community. This means that students participate in online courses with predominantly asynchronous lectures and learning activities that are complemented by synchronous tutorials and hands-on sessions.
Students are guided and supported by faculty as well as experienced tutors and lecturers to transfer acquired knowledge into practice. The hands-on elements include dedicated collaboration with other students using tools and concepts that enable distributed work from different places and different time-zones, including remote access to physical devices and set-ups.
Students enrolled in online study programs will find their course materials such as videos, case studies, scholarly articles, websites, podcasts, online games etc. on a Learning Management Software (LMS) platform provided by Constructor University.

1.3.2 Student Workload

Module sizes range from 2.5 to 7.5 CP. Studying in an online program at Constructor University involves students actively participating in reading, preparing assignments, meeting with peers on task/group projects, synchronous tutor sessions, and watching the required videos.

The terms used in the module data sheets that refer to student workload are defined as follows:

- Asynchronous Self-study = time that that student uses in predefined study contents on digital platforms. Main goal is to acquire content and methods.

- Interactive Learning = time that students spend in a synchronous manner with tutors and in study groups and working on group projects.

- Independent Study = time that students use with recommended further study content and first application of acquired knowledge.

- Assessment preparation = Application of acquired knowledge to specific problems that serve as examples of typical exam questions or writing term papers, designing presentations etc.

1.3.3 Academic Tutors

Academic tutors specifically support the instructor of records and students within the graduate program in their asynchronous teaching and learning. They hold tutorial sessions for online students (individually or in groups) and serve as a first point of contact for student concerns and questions regarding asynchronous learning material and their learning process. In this way, we guarantee that all students, regardless of the global time zone in which they live, can be fully supported by Constructor University.

1.3.4 Assessment and Grading

In Constructor University's online study programs, we particularly emphasize formative forms of assessment. Formative assessment is used to monitor and evaluate how students are learning as they work through a module or study program. It is designed to help students learn more effectively by giving them feedback on their performance and on how it can be improved and/or maintained. It may be marked pass-fail, complete-incomplete, or other rating scale as part of the requirement to qualify for or participate in the final assessment. There are also similar assessment formats, so-called summative assessment with a final grade at the end of the course as in the on-campus teaching, e.g. written exams, presentations, and lab reports.
Any type of assessment may be conducted electronically or complemented by electronic and online assessment and submission elements. This includes computerized testing in a test center, video interviews, online/electronic submission and other formats which use electronic systems and/or devices. For computerized assessments, students will be offered an introduction to the system used to familiarize themselves with it.

1.3.5 Learning Management Software

Constructor University's online classes are supported by technology that includes a learning management system (LMS) and additional education technology tools that may be integrated into the LMS or offered as an alternative environment for students to engage in or to apply their knowledge and skills and to participate in simulations. The LMS includes discussion forums, assignments and quizzes, a gradebook, calendars, instructor and student dashboards. Additional tools offered may include video or document annotations, virtual labs for a variety of technical skills, gamified experiences, and more. The LMS and some associated tools enable timely communication to the students that can support time management and motivation to engage in their course work. The students will have access to applications that enable group work and peer-to-peer communication.

1.4 Target Audience

The Data Engineering Technologies graduate program is targeted towards students who have completed their BSc in areas such as computer science, physics, applied mathematics, statistics, electrical engineering, communications engineering or related disciplines, and who want to deepen their knowledge and data engineering skills. Typical examples are:

- a bachelor in business with a solid statistics and analysis foundation and programming experience;
- a bachelor in geology who wants to become a data scientist and needs to deepen his/her mathematical and statistical skills;

- a student with a bachelor or master degree in one of the natural sciences who wishes to boost his/her career in empirical research or industrial research and development, where professional handling of very large-scale data collections has become a prime bottleneck for success;
- a bachelor in mathematics or physics who wants to capitalize on his/her theoretical knowledge of modeling methods by learning about the hands-on side of data analysis, interesting fields for applications, and options for employment;
- a student with an undergraduate degree in the life sciences wishing to expand their skill sets towards computational methods and to specialize in bioinformatics and the analysis of biomedical data.

To facilitate the integration of students with diverse backgrounds, introductory courses to data base management with Python and mathematics for graduate students in the first semester ensure that all students have or acquire the minimum required skills in programming and mathematics, which are the prerequisites to succeed in this graduate program.

1.5 Career Options

The demand for Data Engineers is massive. Typical fields of work encompass the finance sector, the energy sector, the automotive and health industry as well as retail and telecommunications. Companies and institutions in almost every domain need:

- experts for data acquisition who find out how to collect the data needed;
- experts for data management who know how to store, enhance, protect and process large amounts of data efficiently;
- experts for data analysis who evaluate and interpret the collected data correctly and can visualize the findings clearly.
- Graduates of the program work as data engineers, data analysts, data managers, data architects, business consultants, software and web developers, or system administrators;
- an MSc degree in Data Engineering Technologies also allows students to move on to a PhD and a career in academia and research institutions.

The Career Service Center (CSC) helps students in their career development. It provides students with high-quality training and coaching in CV creation, cover letter formulation, interview preparation, effective presenting, business etiquette, and employer research as well as in many other aspects, thus helping students identify and follow up on rewarding careers after graduating from Constructor University. Furthermore, the Alumni Office helps students establish a long-lasting and global network which is useful when exploring job options in academia, industry, and elsewhere.

1.6 Admission Requirements

The Data Engineering Technologies graduate program requires students to hold a bachelor's degree. Applicants need to prove a strong interest in the contents of the study program in a motivation letter. The general "Admission and Enrollment Policies" of Constructor University apply (see https://constructor.university/student-life/student-services/university-policies/academic-policies).

Additionally, participants should possess elevated analytical, problem solving and verbal communication skills which must be substantiated in recommendation letters.

Admission to Constructor University is selective and based on a candidate's university achievements, recommendations and self-presentation. Students admitted to Constructor University demonstrate exceptional academic achievements, intellectual creativity, and the desire and motivation to make a difference in the world.

The following documents need to be submitted with the application:

- Letter of motivation
- Curriculum vitae (CV)
- Official or certified copies of university transcripts
- Bachelor's degree certificate or equivalent
- Language proficiency test results (minimum score of 90 (TOEFL), 6.5 (IELTS) or 110 (Duolingo)).
- Copy of Passport
- Letter of recommendation (optional)

Formal admission requirements are subject to higher education law and are outlined in the Admission and Enrollment Policy of Constructor University.

For more detailed information about the admission visit:
https://constructor.university/admission-aid/application-information-graduate

1.7 More information and contacts

For more information on the study program please contact the Study Program Chair:

Prof. Dr. Markus Wenzel

Professor of Computer Science

Email: mwenzel@constructor.university

or visit our program website:
Data Engineering Technologies | Constructor University

For more information on Student Services please visit:

https://constructor.university/student-life/student-services

2 The Curriculum

2.1The Curriculum at a Glance

The Data Engineering Technologies curriculum is divided into four semesters and takes two years to complete. Each semester is composed of core, methods, and foundation modules, leading to a master thesis that may be conducted in collaboration with an industry partner. The modules are grouped into four areas, including Core (45 CP), Methods (22.5CP), Foundations (22.5 CP) and Master Thesis (30 CP), as outlined in the Schematic Study Plan (see figure 1). Detailed module descriptions in their latest version are available in the catalogue on CampusNet. The default module size is 7.5 CP, with smaller 2.5 CP modules being possible as justified exceptions, e.g., if the learning goals are more suitable for 2.5 CP and the overall student workload is balanced.

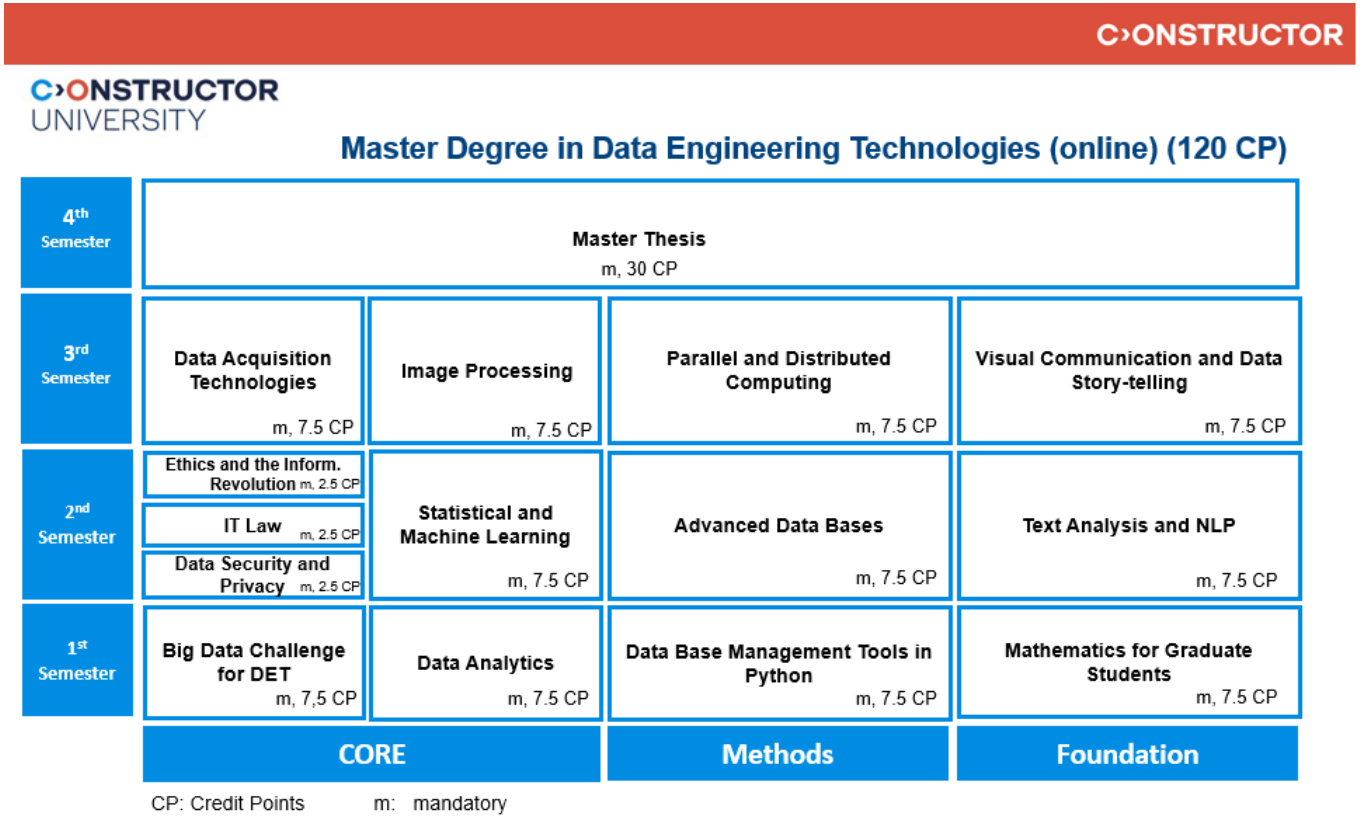

Figure 1: Schematic Study Scheme

2.2. Study and Examination Plan

MSc Degree in Data Engineering Technologies Matriculation Fall 2024							
Module Code	**Program-Specific Modules**	**Type**	**Assessment**	**Period**[1]	**Status**[2]	**Semester**	**CP**
Semester 1							**30**
	CORE Area						**15**
	Module: Big Data Challenge for DET				**m**	**1**	**7.5**
	Big Data Challenge	Lecture (online)	Project report	During Semester			2.5
	Big Data Challenge Tutorial	Tutorial (online)					2.5
	Big Data Challenge Seminar	Seminar (online)	Program code				2.5
	Module: Data Analytics				**m**	**1 or 2**	**7.5**
	Data Analytics	Lecture (online)	Project report	During Semester			5
	Data Analytics Tutorial	Tutorial (online)					2.5
	Methods Area						7.5
	Module: Data Base Management Tools in Python				**m**	**1**	**7.5**
	Data Base Management Tools in Python	Lecture (online)	Written examination	Examination period, during Semester			2.5
	Data Base Management Tools in Python - Programming Tutorial	Tutorial (online)	Program code				2.5
	Data Base Management Tools in Python Tutorial	Tutorial (online)					2.5
	Foundation Area						**7.5**
	Module: Mathematics for Graduate Students				**m**	**1**	**7.5**
	Mathematics for Graduate Students	Lecture (online)	Written examination	Examination period			
Semester 2							**30**
	CORE Area						**15**
	Module: Statistical and Machine Learning				**m**	**2**	**7.5**
	Statistical and Machine Learning	Lecture (online)	Written examination	Examination period			5
	Statistical and Machine Learning Tutorial	Tutorial (online)					2.5
	Module: IT Law				**m**	**2**	**2.5**
	IT Law	Lecture (online)	Term paper	During Semester			
	Module: Data Security and Privacy				**m**	**2**	**2.5**
	Data Security and Privacy	Lecture (online)	Written examination	Examination period			
	Module: Ethics and the Information Revolution				**m**	**2**	**2.5**
	Ethics and the Information Revolution	Seminar (online)	Term paper	During Semester			2.5
	Methods Area						**7.5**
	Module: Advanced Data Bases				**m**	**2**	**7.5**
	Advanced Data Bases	Lecture (online)	Written examination	Examination period			2.5
	Advanced Data Bases Lab	Lab (online)	Lab report	During Semester			2.5
	Advanced Data Bases Tutorial	Tutorial (online)					2.5
	Foundation Area						**7.5**
	Module: Text Analysis and Natural Language Processing				**m**	**2**	**7.5**
	Text Analysis and Natural Language Processing	Seminar (online)	Project report	During Semester			5
	Text Analysis and Natural Language Processing Tutoiral	Tutorial (online)					2.5

Semester 3							**30**
	CORE Area						**15**
	Module: Image Processing				**m**	**3**	**7.5**
	Image Processing	Lecture (online)	Written examination	Examination period			5
	Image Processing Tutorial						2.5
	Module: Data Acquisition Technologies				**m**	**3**	**7.5**
	Data Acquisition Technologies	Lecture & Lab (online)	Project report	During Semester			5
	Data Acquisition Technologies Tutorial	Tutorial (online)					2.5
	Methods Area						**7.5**
	Module: Parallel and Distributed Computing				**m**	**3**	**7.5**
	Parallel and Distributed Computing	Lecture (online)	Written examination	Examination period			5
	Parallel and Distributed Computing Tutoiral	Tutorial (online)					2.5
	Foundation Area						**7.5**
	Module: Visual Communication and Data Storytelling				**m**	**3**	**7.5**
	Visual Communication and Data Storytelling	Lecture (online)	Project report	During Semester			5
	Visual Communication and Data Storytelling Tutorial	Tutorial (online)					2.5
Semester 4							**30**
	Module Master Thesis DET (online)				**m**	**4**	**30**
	Master Thesis MSc DET (online)	Thesis	Thesis				
	Master Thesis Defense		Oral examination				
Total CP							**120**

[1] Each lecture period lasts 14 semester weeks and is followed by reading and examination days. Written examinations are centrally scheduled during weeks 15 and 16. For all other assessment types, the timeframes indicated in the above table stipulate the period during which module work has to be handed in or presented. Specific information on dates of topic announcement as well as submission deadlines is communicated in the syllabus which is made available to the students at the beginning of each semester. Academic dates are published in the university-wide Academic Calendar (see https://constructor.university/student-life/student-services/university-policies/academic-policies).

[2] m = mandatory

Figure 1: Study and Examination Plan

2.3. Core Area (45 CP)

This area is the centerpiece of the Data Engineering Technologies (online) program. The eight mandatory modules in the Core Area provide the foundations for further, more advanced courses and applied projects by introducing the fundamental concepts, methods and technologies used in data science and engineering. The modules are intensive courses accompanied by hands-on tutorials.

To pursue a DET (online) master, the following Core modules (45 CP) need to be taken as mandatory modules (m):

- CORE Module: Big Data Challenge for DET (m, 7.5 CP)
- CORE Module: Data Analytics (m, 7.5 CP)
- CORE Module: Statistical and Machine Learning (m, 7.5 CP)
- CORE Module: Data Security and Privacy (m, 2.5 CP)
- CORE Module: IT Law (m, 2.5 CP)
- CORE Module: Ethics and the Information Revolution (m, 2.5 CP)
- CORE Module: Image Processing (m, 7.5 CP)
- CORE Module: Data Acquisition Technologies (m, 7.5 CP)

2.4. Methods Area (22.5 CP)

In the Methods Area advanced concepts, methods and technologies of data engineering are introduced with a view towards industrial applications.

To pursue a DET (online) master, the following Methods modules (22.5 CP) need to be taken as mandatory modules (m):

Methods Module: Data Base Management Tools in Python (m, 7.5 CP)
Methods Module: Advanced Data Bases (m, 7.5 CP)
Methods Module: Parallel and Distributed Computing (m, 7.5 CP)

2.5. Foundation Area (22.5 CP)

The foundation modules provide a highly structured introduction to the fundamentals of mathematical modelling and analysis, text mining and natural language processing, and data visualization and data storytelling.

To pursue a DET master (online), the following Foundation modules (22.5 CP) need to be taken as mandatory modules (m):

- Foundation Module: Mathematics for Graduate Students (m, 7.5 CP)
- Foundation Module: Text Analysis and Natural Language Processing (m, 7.5 CP)
- Foundation Module: Visual Communication and Data Storytelling (m, 7.5 CP)

2.6. Master Thesis (30 CP)

In the fourth semester, students conduct research and write a mandatory master thesis guided and supported by their thesis supervisors, worth 30 credit points.

- Thesis Module: Master Thesis (m, 30 CP)

The Master thesis provides an opportunity for students to develop their interests in a specific subject area or specialization, and to demonstrate their ability to undertake independent research. Before being eligible to submit the final thesis, students must present and submit a research proposal in advance.

3 Data Engineering Technologies Graduate Program Regulations

3.1. Scope of these Regulations

The regulations in this handbook are valid for all students who entered the DET graduate online program at Constructor University in Fall 2024. In case of conflict between the regulations in this handbook and the general policies for master online studies, the latter shall apply (seehttps://constructor.university/student-life/student-services/university-policies/academic-policies).

In exceptional cases, certain necessary deviations from the regulations of this study handbook might occur during the course of study (e.g., change of the semester sequence, assessment type, or the teaching mode of courses).

In general, Constructor University Bremen reserves therefore the right to change or modify the regulations of the program handbook also after its publication at any time and in its sole discretion.

3.2. Degree

Upon successful completion of the study program, students are awarded a Master of Science (MSc) degree in Data Engineering Technologies.

3.3. Graduation Requirements

In order to graduate, students need to obtain 120 credit points by completing all mandatory components of the program as indicated in chapter 2 of this handbook.

3.4 Other Program-specific Policies & Practices

Close contact and cooperation between program representatives and students are crucial. Therefore, regular meetings are held to continuously evaluate the program, its modules and workshops, supervision, and opportunities. In doing so, the study program chair and involved faculty gain important insights into students' experiences, demands, and overall impressions of the program. On the module component level, students are asked to perform module component evaluations to ensure that the modules are high-quality and that lecturers can make any necessary changes.

The study program chair intensively uses this feedback as well as feedback from industry partners to improve the learning environment, the program's offering, and its progress. The current program was shaped through input from previous experiences and discussions with several stakeholders, including students and industry practitioners.

In exceptional cases, certain necessary deviations from the regulations of this study handbook might occur during the course of study (e.g., change of the semester sequence, assessment type, or the teaching mode of courses). Constructor University Bremen reserves therefore the right to modify the regulations of the program handbook.

4 Modules Description

4.1 Core Area (45 CP)

4.1.1 Big Data Challenge for DET

Module Name Big Data Challenge for DET		**Module Code** XX	**Level (type)** Year 1 (Core)	**CP** 7.5
Module Components				
Number	Name		Type	CP
XX-A	Big Data Challenge		Lecture (online)	2.5
XX-B	Big Data Challenge Tutorial		Tutorial (online)	2.5
XX-C	Big Data Challenge		Seminar (online)	2.5
Module Coordinator Prof. Dr. Adalbert F.X. Wilhelm	**Program Affiliation** • MSc Data Engineering Technologies online (DET) (online)		**Mandatory Status** Mandatory for DET (online)	
Entry Requirements Pre-requisites ☒ None	Co-requisites ☒ None	Knowledge, Abilities, or Skills • Researching information, assessing sources and scientific report writing	**Frequency** Annually (Fall)	**Duration** 1 semester

Student Workload				
Asynchronous Self Study	Interactive Learning	Exam Preparation	Independent Study	Hours Total
35 h	35 h	90h	27.5h	187.5 h

Recommendations for Preparation

- Read the Syllabus.
- Read Susan Ettlinger (2015). What Do we do with all this Big Data? Altimeter. https://www.prophet.com/2015/01/new-research-what-do-we-do-with-all-this-big-data/
- Watch corresponding TEDTalk.

Content and Educational Aims

Big data is one of the buzz words of the current decade and refers to the collection and exploration of complex data sets. This complexity of big data is typically described by the four V's: Volume, Velocity, Variety, and Veracity. From a business perspective, big data is often portrayed as a sea of big opportunities. The public debate is torn between the two poles portrayed by the writers George Orwell and Aldous Huxley: complete surveillance resulting in oppression on the one end, and irrelevance and narcissism on the other. Quite naturally, technological research is mostly concerned with the

technical feasibility of different approaches, the continuously increasing challenges with respect to the four V's, and the creative solutions needed to tackle them.
This module will equip students with the fundamental knowledge needed to harness the power of Big Data by providing an overview on key concepts of Big Data Analytics, including data collection, storage, processing, and governance.
In the lecture component we will explore the foundations, methodologies, and techniques used to extract valuable insights from massive datasets. The course begins with an overview of Big Data and its implications in various industries. Students will gain a solid understanding of the challenges posed by Big Data, such as data quality, scalability, and privacy, and explore solutions to overcome these obstacles. Next, we will delve into the fundamentals of data processing and storage technologies, including Hadoop, and Spark. Students will learn how to design distributed systems that can handle vast amounts of data efficiently. Data governance is a crucial aspect of any Big Data course. It plays a vital role in ensuring the quality, integrity, and security of data throughout its lifecycle. In this course, students will gain a comprehensive understanding of data governance principles and best practices.
The tutorial offers students the possibility to foster their knowledge by asking questions, discussing specific issues, and by collaborating and debating with their peers. Particularly, the tutorial will guide and support the students through their project work.
The seminar component offers the students the chance to dig deeper into the software and programming aspect of big data. Based on the Hadoop ecosystem, students will familiarize themselves with different software components of the big data analysis pipeline such as Airflow, Kubernetes and further popular DevOps. During the Seminar, students will implement data access, data preprocessing, and data analysis along given practical examples.

Intended Learning Outcomes

Upon completion of this module, students will be able to:

1. contribute knowledgeably to the current debate about big data, digitalization and industry 4.0;
2. explain and discuss pros and cons of digitalization from a business perspective as well as a societal perspective;
3. understand the fundamental concepts and challenges of Big Data Analytics;
4. evaluate technological possibilities and innovations driven by big data;
5. assess the business opportunities of current big data developments;
6. design and implement scalable data processing and storage solutions;
7. demonstrate a deep understanding of the software and programming aspect of big data analysis using the Hadoop ecosystem;
8. identify and explain the purpose and functionality of key software components such as Airflow, Kubernetes, and popular DevOps tools used in the big data analysis pipeline.

Indicative Literature

McLellan (2013): Big Data: An Overview
https://www.zdnet.com/article/big-data-an-overview/

V. Mayer-Schönberger & K.Cukier: "Big Data: A Revolution That Will Transform How We Live, Work, and Think", 2013.

S. Akter & S. Fosso Wamba, Big data analytics in e-commerce: A systematic review and agenda for future research, 2016. Electronic Markets, 26 173-194.

Z. Lv, H. Song, P. Basanta-Val, A. Steed and M. Jo. "Next-Generation Big Data Analytics: State of the Art, Challenges, and Future Research Topics," in IEEE Transactions on Industrial Informatics, vol. 13, no. 4, pp. 1891-1899, Aug. 2017.

B.Balamurugan, A. R. Nandharini: "Big Data: Concepts, Technology, and Architecture

Usability and Relationship to other Modules

This module provides an overview on practical big data applications and tools.

Examination Type: Module Component Examinations

Module Component 1: Lecture & Tutorial

Assessment Type: Project Report Length: 4000 words
Weight: 67 %

Scope: Intended learning outcomes 1-5 of the module

Module Component 2: Seminar

Assessment Type: Program Code Weight: 33%

Scope: Intended learning outcomes 6-8 of the module

Completion: To pass this module, the examination of each module component has to be passed with at least 45%.

4.1.2 Data Analytics

Module Name	**Module Code**	**Level (type)**	**CP**
Data Analytics	XXX	Year 1 (CORE)	7.5

Module Components

Number	Name	Type	CP
XXXX	Data Analytics	Lecture (online)	5
XXXX	Data Analytics Tutorial	Tutorial (online)	2.5

Module Coordinator	**Program Affiliation**	**Mandatory Status**
Prof. Dr. Adalbert F.X. Wilhelm	MSc Data Engineering Technologies online (DET) (online)	Mandatory for DET (online), and for DSSB (online)

Entry Requirements			**Frequency**	**Duration**
Pre-requisites	Co-requisites	Knowledge, Abilities, or Skills	Biannually (Fall and Spring)	1 semester
☒ None	☒ None	☒ None		

Student Workload

Asynchronous Self Study	Interactive Learning	Assessment Preparation	Independent Study	Hours Total
35 h	35h	60 h	57.5 h	187.5 h

Recommendations for Preparation

Read the Syllabus.
Take the free online course “Introduction to Data Science” at https://cognitiveclass.ai/courses/data-science-101/

Content and Educational Aims

This module introduces the concepts and methods of data analytics. The objective of the module is to present methods for gaining insights from data and drawing conclusions for analytical reasoning and decision making. The module comprises a broad spectrum of methods for modelling and understanding complex datasets. Comprising both descriptive and predictive analytics, the standard portfolio of supervised and unsupervised learning techniques is introduced. Automatic analysis components, such as data transformation, aggregation, classification, clustering, and outlier detection, will be treated as an integral part of the analytics process.
As a central part of this module, students are introduced to the major concepts of statistical learning, such as cross-validation, feature selection, and model evaluation. The module combines and applies the theoretical foundation of data analytics with a practical exposure to the data analysis process.

Intended Learning Outcomes

By the end of this module, students will be able to

1. explain advanced data analytics techniques in theory and application
2. apply data analytics methods to real-life problems using appropriate tools
3. evaluate and compare different data analytics algorithms and approaches
4. apply statistical concepts to evaluate data analytics results

Indicative Literature

G. James, D. Witten, T. Hastie, Rob Tibshirani: Introduction to Statistical Learning with R by Springer, 2013. (ISLR)
A. Telea, Data Visualization: Principles and Practice, Wellesley, Mass.: AK Peters, 1st edition, 2008. (DV)
M. Ward, G. Grinstein, D. Keim, Interactive Data Visualization: Foundations, Techniques, and Applications. AK Peters, 1st edition, 2010. (IDV)

Usability and Relationship to other Modules

- In this module, students will learn the concepts and various techniques of data analysis. They will be applied in other Data Engineering Technologies modules, and typically also in the master thesis.

Examination Type: Module Examination

Assessment Type: Project Report Length: 4000 words

Weight: 100%

Scope: All intended learning outcomes of the module.

Completion: To pass this module, the examination has to be passed with at least 45%.

4.1.3 IT Law

Module Name			**Module Code**	**Level (type)**	**CP**
IT Law				Year 1 (Core)	2.5
Module Components					
Number	Name			Type	CP
	IT Law			Lecture (online)	2.5
Module Coordinator Prof. Dr Hilke Brockmann/ Prof. Dr. Stefan Kettemann	**Program Affiliation** • MSc Data Science for Society and Business online (DSSB) (online)			**Mandatory Status** Mandatory for DET (online) and DSSB (online)	
Entry Requirements Pre-requisites ☒ None	Co-requisites ☒ None	Knowledge, Abilities, or Skills ☒ None	**Frequency** Annually (Spring)	**Duration** 1 semester	

Student Workload

Asynchronous Self Study	Interactive Learning	Assessment Preparation	Independent Study	Hours Total
17.5 h	10 h	10 h	25 h	62.5 h

Recommendations for Preparation

None.

Content and Educational Aims

Digital information, the Internet, and applications like YouTube or social networking tools like Instagram, Facebook, or Twitter have disrupted legal systems (Murray 2016). IT law is not limited to one legal area but encompasses civil, public, and criminal laws. It spans from human rights law to intellectual property law, contract and consumer protection law, privacy law, data protection law, and other legal domains. Moreover, the global exchange of data conflicts with the territorial principle of jurisdiction. In addition, IT regulations are in a constant flux to keep up with the accelerated pace of technological progress. This module investigates the most important areas of IT law. It provides the participants with a sound understanding of legal principles and regulations, and sheds light on international as well as European ICT policies and governance. A special focus will be given to the European General Data Protection Regulation (GDPR).

Intended Learning Outcomes

By the end of this module, students should be able to

1. identify legal questions and implications in relation to digital transformation technologies/IT law/ AI and algorithms
2. understand fundamental national and international legal frameworks related to the use of data
3. know the relevant IP rights regarding data and algorithms
4. understand and critically assess legal regulations about data privacy and data protection
5. recognize and explain the types of bias inherent in data processing
6. explain the legal concerns related to data-based automatic decision making
7. understand how to comply to the GDPR and assess its impact on individuals, firms, and organizations
8. understand and critically evaluate the liabilities and available remedies with regard to data
9. explain and develop potential future IT regulation mechanisms

Indicative literature

Lloyd (2020). Information Technology Law. Oxford: Oxford University Press (9th ed).

Usability and Relationship to other modules

IT Law provides the fundamental knowledge on the legal framework when dealing with data. This knowledge will be expanded and applied in courses such as "Ethics and the Information revolution" and needs to be taken into account in the master thesis module.

Examination Type: Module Examination

Assessment Type: Term Paper Length: 3500 words
Weight: 100%

Scope: All intended learning outcomes of the module.

Completion: To pass this module, the examination has to be passed with at least 45%.

4.1.4 Ethics and the Information Revolution

Module Name	**Module Code**	**Level (type)**	**CP**
Ethics and the Information Revolution		Year 1 (Core)	2.5

Module Components

Number	Name	Type	CP
	Ethics and the Information Revolution	Seminar (online)	2.5

Module Coordinator	**Program Affiliation**	**Mandatory Status**
Prof. Dr. Hilke Brockmann	• MSc Data Science for Society & Business online (DSSB) (online)	Mandatory for DET (online) and DSSB (online)

Entry Requirements			**Frequency**	**Duration**
Pre-requisites	Co-requisites	Knowledge, Abilities, or Skills	Annually (Spring)	1 semester
☒ None	☒ None	☒ None		

Student Workload

Asynchronous Self Study	Interactive Learning	Assessment Preparation	Independent Study	Hours Total
17.5 h	10 h	10 h	25 h	62.5 h

Recommendations for Preparation

Read the Syllabus.
Binns (2018) Fairness in Machine Learning: Lessons from Political Philosophy. Proceedings of Machine Learning Research 81:1-11.

Content and Educational Aims

Many data specialists claim that we are at the cusp of an information revolution. Based on inventions dating back to WWII, IT innovations have re-organized our society around one "big metadata computer" that is permanently computing data and associating metadata about everything we do. Digital technologies also have the potential to disrupt the ethical standards and rules of our society. In this module, we discuss whether we have to forfeit privacy in times of big data, if machines compromise our identity, and if shared data enables institutions to abuse their power and undermine the civil society?

The module pursues three goals. 1. Participants will immerse themselves and learn about core ethical theories. 2. They will integrate this theoretical knowledge and develop a "Big Data Ethics," which they 3. will put into practice. For the second and third purposes, discussions and interactions are indispensable for identifying possible dilemmas and conflict of interests and for balancing contradictions to derive practical solutions and policy advice.

Intended Learning Outcomes

By the end of the module, students will be able to

1. report on major ethical theories relevant to digital technologies
2. integrate different ethical standpoints and arguments to address concrete societal problems
3. assess the societal and ethical implications of digitization
4. deal with legal aspects of ethics by applying means to prevent and deal with violations of privacy and transparency
5. apply actions to contribute to the transition to a more just and trustworthy digital transformation as a part of one's job
6. implement justice and social equality as dimensions of ethics and sustainability

Indicative Literature

Binns (2018) Fairness in Machine Learning: Lessons from Political Philosophy. Proceedings of Machine Learning Research 81:1-11.

Examination Type: Module Examination

Assessment Type: Term Paper Length: 3000 words
Weight: 100%

Scope: All intended learning outcomes of the module.

Completion: To pass this module, the examination has to be passed with at least 45%.

4.1.5 Data Security and Privacy

Module Name	**Module Code**	**Level (type)**	**CP**
Data Security and Privacy	XX	Year 1 (Core)	2.5

Module Components

Number	Name	Type	CP
XX	Data Security and Privacy	Lecture (online)	2.5

Module Coordinator	**Program Affiliation**	**Mandatory Status**
NN	• MSc Data Engineering Technologies online (DET) (online)	Mandatory for DET (online) and DSSB (online)

Entry Requirements			**Frequency**	**Duration**
Pre-requisites	Co-requisites	Knowledge, Abilities, or Skills	Annually (Spring)	1 semester
☒ None	☒ None	☒ None		

Student Workload

Asynchronous Self Study	Interactive Learning	Assessment Preparation	Independent Study	Hours Total
17.5 h	10 h	10 h	25 h	62.5 h

Recommendations for Preparation

Read the syllabus.

Content and Educational Aims

Data Security and Privacy introduces concepts of data security. Basic cryptographic mechanisms are introduced, and it will be explained how these mechanisms can be used to protect data during transmission over the Internet or while data is stored on computing systems. The module component will also introduce the technical aspects of data privacy and concepts such as anonymity, linkability, observability and pseudonymity.

Intended Learning Outcomes

Upon completion of this module, students will be able to:

1. analyze and develop principles of symmetric and asymmetric encryption
2. understand the use of cryptographic hash functions to ensure data integrity
3. summarize and communicate the principles of key management approaches
4. assess and choose appropriate techniques for authentication
5. critically assess how data security and privacy issues are solved and how this will impact the data security of applications

Indicative Literature

D. R. Stinson, Cryptography: Theory and Practice, ISBN, 1-58488-206-9, Chapman & Hall. 4th edition, 2018.

https://ebookcentral.proquest.com/lib/jacob/detail.action?docID=5493336

Usability and Relationship to other Modules

Examination Type: Module Examination

Assessment Type: Written examination Duration: 90 minutes
Weight: 100%

Scope: All intended learning outcomes of the module

Completion: To pass this module, the examination has to be passed with at least 45%.

4.1.6 Statistical and Machine Learning

Module Name	**Module Code**	**Level (type)**	**CP**
Statistical and Machine Learning	XX	Year 1 (Core)	7.5

Module Components

Number	Name	Type	CP
XX-A	Statistical and Machine Learning	Lecture (online)	5
XX-B	Statistical and Machine Learning Tutorial	Tutorial (online)	2.5

Module Coordinator	**Program Affiliation**	**Mandatory Status**
NN	• MSc Data Engineering Technologies online (DET) (online)	Mandatory for DET (online)

Entry Requirements			**Frequency**	**Duration**
Pre-requisites	Co-requisites	Knowledge, Abilities, or Skills	Annually (Spring)	1 semester
☒ None	☒ None	☒ Basic linear algebra, calculus and probability theory, as typically acquired in entry modules in BSc studies		

Student Workload

Asynchronous Self Study	Interactive Learning	Exam Preparation	Independent Study	Hours Total
35h	35 h	30 h	87.5 h	187.5 h

Recommendations for Preparation

Read the Syllabus.
Highly recommended: Mitchell, Tom M.: Machine Learning (McGraw-Hill, 1997) IRC: Q325.5.M58 1997. This standard, classical textbook gives a very accessible overview of ML.

Content and Educational Aims

Machine learning (ML) is a module that concerns algorithms that are fed with (large quantities of) real-world data, and which return a compressed "model" of the data. An example is the "world model" of a robot: the input data are sensor data streams, from which the robot learns a model of its environment. Another example is a spoken language model: the input data are speech recordings, from which ML methods build a model of spoken English -- useful, for instance, in automated speech recognition systems. There are many formalisms in which such models can be cast, and an equally large diversity of learning algorithms. At the same time, there is a relatively small number of fundamental challenges that are common to all of these formalisms and algorithms.

The module introduces such fundamental concepts and illustrates them with a choice of elementary model formalisms (linear classifiers and regressors, radial basis function networks, clustering, neural networks). Furthermore, the module also (re)introduces required mathematical material from probability theory and linear algebra. The main educational aims are twofold: to make students fully aware of the two main hurdles for obtaining good models from data: (i) the "curse of

dimensionality" and (ii) the bias-variance dilemma and to provide standard tools to cope with these difficulties, namely (i') dimension reduction by feature extraction, for example via PCA or clustering, and (ii') cross-validation and regularization.

Intended Learning Outcomes

Upon completion of this module, students will be able to:

1. design, implement and exploit elementary supervised ML methods for classification and regression with expert care given to dimension reduction preprocessing and regularization;
2. understand and practically use PCA and linear regression;
3. understand the core ideas behind feedforward neural networks and the backpropagation algorithm, as the basis for accessing "deep learning" methods.

Indicative Literature

T. M. Mitchel, Machine Learning, McGraw-Hill, 1997, IRC: Q325.5.M58.

Usability and Relationship to other Modules

In this module students will learn concepts and various techniques of Machine Learning. These skills will be expanded on in subsequent core and method courses and typically in the master thesis.

Examination Type: Module Examination

Assessment Type: Written Examination Duration: 120 minutes
Weight: 100%

Scope: All intended learning outcomes of the module

Completion: To pass this module, the examination has to be passed with at least 45%.

4.1.7 Data Acquisition Technologies

Module Name	**Module Code**	**Level (type)**	**CP**
Data Acquisition Technologies	XX	Year 2 (Core)	7.5

Module Components

Number	Name	Type	CP
XX-A	Data Acquisition Technologies	Lecture and Lab (online)	5
XX-B	Data Acquisition Technologies Tutorial	Tutorial (online)	2.5

Module Coordinator	**Program Affiliation**	**Mandatory Status**
NN	• MSc Data Engineering Technologies online (DET) (online)	Mandatory for DET (online)

Entry Requirements			**Frequency**	**Duration**
Pre-requisites	Co-requisites	Knowledge, Abilities, or Skills	Annually (Fall)	1 semester
☒ None	☒ None	The students should be familiar with at least some of the following topics: basic electrical circuits, microcontrollers, HTML, PHP, SQL, C, and Python.		

Student Workload

Asynchronous Self Study	Interactive Learning	Exam Preparation	Independent Study	Hours Total
35 h	35 h	60 h	57.5h	187.5 h

Recommendations for Preparation

Read the Syllabus.
A lab manual will be provided, reading the lab manual before each lab session is recommended.

Content and Educational Aims

- Medical monitoring, smart cars, smart grids, smart homes, and ubiquitous connections to the internet everywhere: There will be an ocean of data not only entered by humans but also automatically pouring in from billions of sensors deployed in a plethora of devices. How are such data collected, and how can they be made available to you, to your doctor, or to other users? These are only some of the questions to be addressed. This module offers a hands-on introduction to the technology behind the scenes. Topics include microcontrollers; how to program them; the way they interact with sensors and actuators; and the wireless techniques they use to communicate with each other, with other computers, and with the internet.
- As the module covers a wide range of platforms, it also utilizes aspects from a variety of different languages and devices. To be successful, it helps to be familiar with basic electrical circuits, microcontrollers, HTML, PHP, SQL, C, and Python. Although there will be a lot of support, it is recommended to be familiar with at least a few of these aspects.

Intended Learning Outcomes

Upon completion of this module, students will be able to:

1. acquire data from different sensors and use a microcontroller to process them;
2. transmit data from the microcontroller to a database on a server
3. collect data from web browsers and transmit them to a database on a server
4. visualize the data on computers or smart devices
5. set up a wireless sensor network and communicate data among different components.

Indicative Literature

M. Kooijman, Building wireless sensor networks using Arduino: leverage the powerful Arduino and XBee platforms to monitor and control your surroundings, Packt Publishing, 2015 ISBN:9781784397159 1784397156.

H. E Williams, D. Lane, Web database applications with PHP and MySQL, O'Reilly Media, 2004, ISBN: 0596005431 9780596005436.

Usability and Relationship to other Modules

This module offers the techniques of wireless acquisition of the data that can be processed and analyzed by techniques learnt in other DET modules and applied in the master thesis.

Examination Type: Module Examination

Assessment Type: Project Report Length: 20 pages
Weight: 100%

Scope: All intended learning outcomes of the module

Completion: To pass this module, the examination has to be passed with at least 45%.

4.1.8 Image Processing

Module Name Image Processing			**Module Code** XX	**Level (type)** Year 2 (Core)	**CP** 7.5
Module Components					
Number	Name			Type	CP
XX-A	Image Processing			Lecture (online)	5
XX-B	Image Processing Tutorial			Tutorial (online)	2.5
Module Coordinator Prof.Dr. Markus Wenzel	**Program Affiliation** • MSc Data Engineering Technologies online (DET) (online)			**Mandatory Status** Mandatory for DET (online)	
Entry Requirements			**Frequency** Annually Fall	**Duration** 1 semester	
Pre-requisites	Co-requisites	Knowledge, Abilities, or Skills			
☒ None	☒ None	• Basic linear algebra, calculus and programming skills			

Student Workload				
Asynchronous Self Study	Interactive Learning	Exam Preparation	Independent Study	Hours Total
35 h	35 h	30 h	87.5 h	187.5 h

Recommendations for Preparation
Read the Syllabus.

Content and Educational Aims

This module introduces the basic concepts of image processing.

This course introduces the basic concepts and applications of image processing for data engineers. This module's aim is to enable students:

- to transform images using different methods,
- to work with digital images conceived of as two- or higher-dimensional signals,
- and to perform operations on them to extract useful information.

This course introduces sampling and quantization strategies, elementary image transformations, and higher-level information extraction techniques like noise reduction, image segmentation, and feature extraction leading up to image analysis automation using state-of-the-art methods from machine learning and deep learning research.

This module consists of quizzes following each lecture. Each quiz will consist of a set of questions, some in open form, some in multiple-choice form, and some in gap text form, depending on the content of the lecture. Other formats might substitute those formats dynamically. Quizzes are conducted online.

Intended Learning Outcomes

Upon completion of this module, students will be able to:

1 Understand and describe types of imaging physics, implemented in common sensor technologies used to acquire images (e.g. radiation, emission, and transmission imaging, and examples of each)
2 explain basic concepts of image processing like pixel-wise transforms, neighborhood transforms, and how elementary transforms can be composed into image analysis, on the example of morphological image processing;
3 apply sampling and quantization strategies,
4 Understand and compare common methods for image segmentation, like region growing, connected components, or the watershed transform,
5 Know algorithms for edge, corner, and interest point detection like Canny edges, Harris corners, others
6 compare noise reduction methods by their principles, and select appropriate methods based on a given use case, like Gaussian blurring, bilateral filters, anisotropic diffusion filters
7 perform feature extraction from images using methods from categories like shape features, intensitiy features, texture features, interest point descriptors;
8 Apply machine learning to features derived from images,
9 Know common techniques for deep-learning-based image processing, like CNNs for classification and segmentation, GANs for image generation, and Transformers for multi-purpose image processing;
10 design and implement their own image processing algorithms in Python and apply these to real world examples.

Indicative Literature

A. Bovik (ed.): The Essential Guide to Image Processing, 2nd editiion. Academic Press, 2009

M. Sonka, V. Hlavac, R. Boyle: Image Processing, Analysis, & Machine Vision, 4th edition. Cengage Learning, 2014

Usability and Relationship to other Modules

As this module introduces visualization techniques for data sets, it builds on courses introducing data systems, particularly the Data Analytics module.

Examination Type: Module Examination

Assessment Type: Written Examination Duration: 120 minutes
Weight: 100%

Scope: All intended learning outcomes of the module

Completion: To pass this module, the examination has to be passed with at least 45%.

4.2 Methods Area (22.5 CP)

4.2.1 Data Base Management Tools in Python

Module Name Data Base Management Tools in Python			**Module Code** XX	**Level (type)** Year 1 (METHODS)	**CP** 7.5
Module Components					
Number	Name			Type	CP
XX-A	Data Base Management Tools in Python			Lecture (online)	2.5
XX-B	Data Base Management Tools in Python – Programming Tutorial			Tutorial (online)	2.5
XX-C	Data Base Management Tools in Python Tutorial			Tutorial (online)	2.5
Module Coordinator NN	**Program Affiliation** • MSc Data Engineering Technologies (DET) (online)			**Mandatory Status** Mandatory for DET (online) and for DSSB (online)	
Entry Requirements Pre-requisites ☒ None	Co-requisites ☒ None	Knowledge, Abilities, or Skills ☒ None	**Frequency** Annually (Fall)	**Duration** 1 semester	

Student Workload				
Asynchronous Self Study	Interactive Learning	Exam Preparation	Independent Study	Hours Total
35 h	35 h	30 h	87.5 h	187.5 h

Recommendations for Preparation
Read the Syllabus.

Content and Educational Aims

- This module introduces data engineering students to the field of data management with Python. Data management describes the vast field of methodologies to collect, store, process and provision data. The aim of this module is to focus on a very applied view of these tasks.
- Since Python has become the de-facto standard in the field, the initial part of the module is concerned with a basic introduction into core concepts of imperative programming in Python.
- Data structures and fundamental algorithms are discovered in a hands-on fashion. These will also include basic numerical and data analysis tasks based on NumPy/SciPy. One source from which we can collect and in which we can store data are relational databases.
- The course introduces the Structured Query Language (SQL) to get access to this data source. More recently, data is frequently stored in Data Frames, a data structure provided by Pandas, a Python library. Pandas also provides functionality to carry out data analysis tasks. Provisioning of data analysis outputs will be done by basic2D visualization techniques.

Intended Learning Outcomes

Upon completion of this module, students will be able to:

1. explain and apply fundamental concepts of imperative programming using Python
2. understand and use basic data structures
3. summarize and apply fundamental algorithms (e.g. sorting)
4. execute basic data analysis tasks (average, min, max, ...)
5. Understand and implement linear algebra operations using NumPy/SciPy
6. explain fundamentals of relational databases
   describe and use SQL to create, modify and query data from relational databases
7. understand and apply DataFrames and data analysis using Pandas
8. visualize simple data by different types of 2D plots using Matplotlib

Indicative Literature

Jake VanderPlas, Python Data Science Handbook, O'Reilly.

Cay S. Horstmann, Rance D. Necaise, Python for Everyone, 3rd Edition, Wiley.

Usability and Relationship to other Modules

The course provides the necessary background knowledge for

- DET: all subsequent Data Engineering Technologies courses, in particular to “Advanced Databases” and “Statistical and Machine Learning”
- DSSB (online): efficient data handling and processing typically required in the master thesis module.

Examination Type: Module Component Examinations

Module Component 1: Lecture

Assessment Type: Written Examination Duration: 120 minutes
Weight: 50%

Scope: All intended learning outcomes of this module excluding practical aspects.

Module Component 2: Programming Tutorial

Assessment Type: Program Code Weight: 50%

Scope: All intended learning outcomes of the module

Completion: To pass this module, the examination of each module component has to be passed with at least 45%.

4.2.2 Advanced Data Bases

Module Name Advanced Data Bases			**Module Code** XX	**Level (type)** Year 1 (METHODS)	**CP** 7.5
Module Components					
Number	Name			Type	CP
XX-A	Advanced Data Bases			Lecture (online)	2.5
XX-B	Advanced Data Bases Lab			Lab (online)	2.5
XX-C	Advanced Data Bases Tutorial			Tutorial (online)	2.5
Module Coordinator NN	**Program Affiliation** • MSc Data Engineering Technologies online (DET) (online)			**Mandatory Status** Mandatory for DET (online)	
Entry Requirements			**Frequency**	**Duration**	
Pre-requisites	Co-requisites	Knowledge, Abilities, or Skills	Annually (Spring)	1 semester	
☒ Data Base Management Tools in Python	☒ None	• mandatory knowledge of SQL • working knowledge about fundamental data structures, such as trees • working knowledge of computer architectures • good command of at least one programming language, as several languages will be used in the lab			

Student Workload				
Asynchronous Self Study	Interactive Learning	Exam Preparation	Independent Study	Hours Total
35 h	35 h	60 h	57.5 h	187.5 h

Recommendations for Preparation
Read the Syllabus.

Content and Educational Aims

This course deepens knowledge and skills in managing and serving Big Data with emphasis on flexibility and scalability. As a result of this course, students will know the state of the art in data management for particularly large and complex data, including in cloud-based data setups

Based on the Data Engineering Technologies lecture on Data Base Management Tools in Python the course starts with a reinspection of classical SQL, preparing an overview of SQL query processing. Based on this understanding opportunities of optimization and parallelization are discussed. Subsequently, novel developments in Big Data services are discussed. NoSQL approaches with their new data models are inspected, such as documents, graphs and arrays. This is contrasted with NewSQL and their novel techniques for competitive performance. Dedicated architectures are discussed, such as

MapReduce. This leads to general scalability considerations, with an emphasis on large-scale parallel and distributed processing. Throughout the course practical considerations play an important role, including practitioner hints on database modeling, tuning, and security. Practical guided hands-on exercises complement this.

Intended Learning Outcomes

Upon completion of this module, students will be able to:

1. Summarize the state of the art in data management for particularly large and complex data
2. Establish criteria for selecting adequate scalable data management technology based on various criteria
3. Establish a state-of-the-art database schema for a given application scenario
4. Tune a relational database for best performance on some given query workload
5. Adequately consider security aspects in databases
6. Develop applications using Web and database technology

Indicative Literature

McLellan (2013): Big Data: An Overview, https://www.zdnet.com/article/big-data-an-overview/

S. Akter & S. Fosso Wamba, Big data analytics in e-commerce: A systematic review and agenda for future research, 2016. Electronic Markets, 26 173-194.

Z. Lv, H. Song, P. Basanta-Val, A. Steed and M. Jo. "Next-Generation Big Data Analytics: State of the Art, Challenges, and Future Research Topics," in IEEE Transactions on Industrial Informatics, vol. 13, no. 4, pp. 1891-1899, Aug. 2017.

Usability and Relationship to other Modules

Examination Type: Module Examination

Module Component 1: Lecture

Assessment Type: Written Examination Duration: 120 min
Weight: 67%

Scope: Intended learning outcomes (1,2,3,4,5).

Module Component 2: Lab

Assessment Type: Laboratory Report
Weight: 33%

Scope: Intended learning outcomes (3,4,5,6).

Completion: To pass this module, the examination of each module component has to be passed with at least 45%.

4.2.3 Parallel and Distributed Computing

Module Name Parallel and Distributed Computing		**Module Code** XX	**Level (type)** Year 2 (METHODS)	**CP** 7.5
Module Components				
Number	Name		Type	CP
XX-A	Parallel and Distributed Computing		Lecture (online)	5
XX-B	Parallel and Distributed Computing Tutorial		Tutorial (online)	2.5
Module Coordinator NN	**Program Affiliation** • MSc Data Engineering Technologies online (DET) (online)		**Mandatory Status** Mandatory for DET (online)	
Entry Requirements Pre-requisites: ☒ None Co-requisites: ☒ None Knowledge, Abilities, or Skills: ▪ Mandatory proficiency in Python		**Frequency** Annually (Fall)	**Duration** 1 semester	

Student Workload				
Asynchronous Self Study	Interactive Learning	Exam Preparation	Independent Study	Hours Total
35 h	35 h	30 h	87.5 h	187.5 h

Recommendations for Preparation

Read the Syllabus.

Content and Educational Aims

In recent years, the development of parallel and cloud computing has opened the door for Big Data analysis and processing. This module aims at providing an overview and introduction to the vast field of parallel and cloud computing.

- In traditional parallel computing, we aim to develop notions for different parallelization models (shared-memory, distributed- memory, SIMD, SIMT), get to know appropriate programming methodologies for high performance data analysis (OpenMP / MPI) and aim at understanding performance and scalability in this field (weak vs. strong scaling, Amdahl's law). In the tutorial a basic introduction in C/C++ is provided.
- This fundamental knowledge will then be carried over to recent developments in cloud computing, where distributed processing frameworks (Spark / Hadoop MapReduce / Dask), based on appropriated deployment infrastructures, are in the process to become De Facto standards for Big Data processing and analysis. We will approach these technologies from a practical point of view and aim at developing the necessary knowledge to carry out scalable machine learning and data processing on Big Data.

Intended Learning Outcomes

Upon completion of this module, students will be able to:

1. understand theory and fundamentals of parallelization models (shared-/distributed memory, SIMD, SIMT)
2. explain and apply parallel programming methodologies (OpenMP / MPI)
3. describe and analyze performance and scalability (weak vs. strong scaling, ...)
4. Use GPU programming
5. Understand basic principles of distributed and cloud computing
6. use distributed processing frameworks (Spark / Hadoop MapReduce / Dask) for scalable distributed calculations
7. develop scalable machine learning and data processing on Big Data

Indicative Literature

Zaccone, Python Parallel Programming Cookbook, O'Reilly.

J.C. Daniel, Data Science with Python and Dask, Manning Publications.

Z. Radtka, D. Miner, Hadoop with Python. Hadoop with Python, O'Reilly.

Usability and Relationship to other Modules

Examination Type: Module Examination

Assessment Type: Written Examination Duration: 120 minutes
Weight: 100 %

Scope: All intended learning outcomes of the module

Completion: To pass this module, the examination has to be passed with at least 45%.

4.3 Foundation Area (22.5 CP)

4.3.1 Mathematics for Graduate Students

Module Name	**Module Code**	**Level (type)**	**CP**
Mathematics for Graduate Students	XX-01	Year 1 (Foundation)	7.5

Module Components

Number	Name	Type	CP
XX-01	Mathematics for Graduate Students	Lecture (online)	7.5

Module Coordinator	**Program Affiliation**	**Mandatory Status**
NN	• MSc Data Engineering Technologies online (DET) (online)	Mandatory for DET (online) and DSSB (online)

Entry Requirements			**Frequency**	**Duration**
Pre-requisites	Co-requisites	Knowledge, Abilities, or Skills	Annually (Fall)	1 semester
☒ None	☒ None	☒ Mathematics at High School level		

Student Workload

Asynchronous Self Study	Interactive Learning	Exam Preparation	Independent Study	Hours Total
35 h	35 h	30 h	87.5 h	187.5 h

Recommendations for Preparation
Read the Syllabus.

Content and Educational Aims

- This module offers a highly structured introduction to the fundamentals of three major pillars of mathematical modelling and analysis: Single and multivariable calculus, linear algebra, as well as the fundamentals of combinatorics and probabilities as they are used for statistical modeling and estimation.
- It is a gateway for graduate students who have not been exposed to the topics so far, or who were exposed to them long ago and need a refresher.
- The module starts with an introduction to linear algebra, including matrices, determinants, eigenvalues, eigenvectors, scalar products, and norms. It continues with single and multivariable calculus, including sequences, series, limits, derivatives, Taylor series, and integrals. The module focuses on practical experience rather than on mathematical rigor.
- The module continues then with the concept of probabilities, including joint, conditional and total probabilities with a focus on independence, which leads us to a discussion of Bayes's theorem. We shall then proceed to factorials, and binomial coefficients, with many applications to be followed by the binomial law, and its Poisson and Normal approximations. A second block covers random variables with their distributions and density functions. Here we are going to discuss continuous random variables in detail. It concludes with the essential ideas of expected values, moments, and estimation.

Intended Learning Outcomes

Upon completion of this module, students will be able to:

1. calculate derivatives and simple integrals;
2. apply the fundamental concepts of calculus and linear algebra in structured situations;
3. understand and use vectors and matrices, calculate determinants, eigenvalues and eigenvectors in simple cases;
4. explain the importance of the methods of calculus and linear algebra in problems arising from applications;
5. understand the methods of calculus and linear algebra used in more advanced modules as well as in scientific literature.
6. understand the fundamental concepts of probabilities and combinatorics and to apply them in structured situations,
7. apply important probability laws (Binomial, Poisson, Normal),
8. understand and apply probability distributions and densities,
9. understand and apply means, variances, and covariances – also in the context of simple estimation contexts.

Indicative Literature

G. Strang, Introduction to Linear Algebra, 5th edtion, Wellesley-Cambridge Press, 2016, ISBN: 978-09802327-7-6.

H. Stark, J. W. Woods, Probability and Random Processes with Applications to Signal Processing, Third Edition, 2002.

Usability and Relationship to other Modules

This module introduces and refreshes the essential Calculus and Linear Algebra required in most of the modules of the data engineering program. Familiarity with probability-related concepts is the basis to understand the foundations of stochastic modelling and the data analytics and machine learning techniques which form a central part of data engineering.

Examination Type: Module Examination

Assessment Type: Written Examination Duration: 120 min

Weight: 100%

Scope: All intended learning outcomes of the module

Completion: To pass this module, the examination has to be passed with at least 45%.

4.3.2 Text Analysis and Natural Language Processing

Module Name	**Module Code**	**Level (type)**	**CP**
Text Analysis and Natural Language Processing		Year 1 (Foundation)	7.5

Module Components

Number	Name	Type	CP
	Text Analysis and Natural Language Processing	Seminar (online)	5
XXXXX	Text Analysis and Natural Language Processing Tutorial	Tutorial (online)	2.5

Module Coordinator	**Program Affiliation**	**Mandatory Status**
NN	• MSc Data Science for Society and Business (online) (DSSB)	Mandatory for DET (online) and DSSB (online)

Entry Requirements			**Frequency**	**Duration**
Pre-requisites	Co-requisites	Knowledge, Abilities, or Skills	Annually (Spring)	1 semester
☒ None	☒ None	☒ Programming skills in R or Python at an intermediate level		

Student Workload

Asynchronous Self Study	Interactive Learning	Exam Preparation	Independent Study	Hours Total
35 h	67.5 h	0 h	85 h	187.5 h

Recommendations for Preparation
None.

Content and Educational Aims

This module will teach the fundamentals of text mining, natural language processing, and automated content analysis using Python. Students will learn the entire text analysis pipeline, from basic web scraping techniques for collecting text data from social media, over text representations and ontologies, to text mining algorithms and efficient representation of analysis results. Students will be exposed to theoretical and methodological foundations of text mining, such as word frequencies, ontologies, bag-of-word, as well as the application of machine learning algorithms for text and sentiment analysis. The module will introduce exemplary studies on text and sentiment analysis and provide an opportunity for hands-on programming to realize different analyses. The module covers a spectrum of text mining methods, from basic lexicographic measures to more complex statistical learning algorithms such as sentiment analysis and topic modeling.

Intended Learning Outcomes

By the end of this module, students should be able to

1. explain the concept of “text as data”
2. use basic methods for information extraction and text data retrieval
3. process and prepare text data for statistical modeling and automated content analysis
4. perform different text analyses using text mining packages in R
5. interpret diverse text analytical measures
6. undertake a knowledgeable automated content analysis with text data

Indicative Literature

Lane, Howard, Hapke (2019) Natural Language Processing in Action. Shelter Ilant: NYy

Usability and Relationship to other Modules

This module translates the insights from Big Data Challenge DET into text analysis.

Examination Type: Module Examination

Assessment Type: Project Report Length: 4000 words
Weight: 100%

Scope: All intended learning outcomes of the module.

Completion: To pass this module, the examination has to be passed with at least 45%.

4.3.3 Visual Communication and Data Storytelling

Module Name	**Module Code**	**Level (type)**	**CP**
Visual Communication and Data Storytelling		Year 2 (Foundation)	7.5

Module Components

Number	Name	Type	CP
	Visual Communication and Data Storytelling	Lecture (online)	5
	Visual Communication and Data Storytelling Tutorial	Tutorial (online)	2.5

Module Coordinator	**Program Affiliation**	**Mandatory Status**
Prof. Dr. Jan Lorenz	• MSc Data Science for Society and Business (DSSB) (online)	Mandatory for DET (online) and DSSB (online)

Entry Requirements			**Frequency**	**Duration**
Pre-requisites	Co-requisites	Knowledge, Abilities, or Skills	Annually (Fall)	1 semester
☒ None	☒ None	☒ None		

Asynchronous Self Study	Interactive Learning	Assessment Preparation	Independent Study	Hours Total
35 h	62.5 h	60 h	30 h	187.5 h

Recommendations for Preparation
Read the syllabus and search for appropriate online example cases.

Content and Educational Aims

Data is often intuitively communicated using statistical graphs and visualization dashboards. Effective communication using visuals and dashboards has become a key qualification for modern business intelligence professionals. This module introduces the basic ideas and concepts of data visualization and data storytelling. Computer-based visualization systems provide visual representations of datasets to process data more effectively. These datasets may come from different sources, such as scientific experiments, simulations, medical scans, commercial databases, financial transactions, health records, and social networks. They also cater to different audiences. Students will learn about the theory of graphical design and the science of visual perception to make compelling visual representations with static and interactive maps for a scientific and non-scientific audience. Students learn to design elegant data visualizations that support the exchange of information and corroborate the data findings. Students also learn to evaluate visualization systems from both the designer's and audience's perspective. Visualization skills are further elaborated with the support of selected online programming snippets.

Topics:
- Theory of graphical design
- Grammar of graphics
- Science of visual perception
- Exploratory data analysis and static graphics in R
- Scientific storytelling for various formats and audiences
- Visualization programming

Intended Learning Outcomes

By the end of this module, students should be able to

1. visually represent various data sources
2. choose suitable visual representations for different data sets
3. evaluate visual depictions of data
4. assist users in visual data analysis
5. target visual representations to different audiences

Indicative Literature

Dykes (2019) Effective Data Storytelling: How to Drive Change with Data, Narrative, and Visuals. Hoboken, NJ: Wiley.

Nussbaumer, Knaflic (2015) Storytelling with Data: A Data Visualization Guide for Business Professionals. Hoboken, NJ: Wiley.

Usability and Relationship to other Modules

Can be used in all modules, particularly in the thesis modules.

Examination Type: Module Examination

Assessment Type: Project Report Length: 5000 words
Weight: 100%

Scope: All intended learning outcomes of the module.

Completion: To pass this module, the examination has to be passed with at least 45%.

4.5 Master Thesis (30 CP)

Module Name	**Module Code**	**Level (type)**	**CP**
Master Thesis DET		Year 2	30

Module Components

Number	Name	Type	CP
	Thesis	Thesis	30

Module Coordinator	**Program Affiliation**	**Mandatory Status**
NN	• MSc Data Engineering Technologies online (DET) (online)	Mandatory for DET (online) students

Entry Requirements			**Frequency**	**Duration**
Pre-requisites	Co-requisites	Knowledge, Abilities, or Skills	Annually (Spring)	1 semester
☒ Successful completion of at least 75 CP	☒ None	• Proficiency in the area of the chosen thesis topic		

Student Workload

Asynchronous Self Study	Interactive Learning	Assessment Preparation	Independent Study	Hours Total
10h	15 h	25 h	700 h	750 h

Recommendations for Preparation

- Identify an area or a topic of interest and contact potential supervisors
- Create a research proposal including a research plan to ensure timely submission.
- Ensure you possess all required technical research skills or are able to acquire them on time.
- Review again the University's Code of Academic Integrity and Guidelines to Ensure Good Academic Practice

Content and Educational Aims

The Master thesis allows students to develop their interests in a specific subject area or specialization, and demonstrate their ability to undertake independent research.

The selected topic of the thesis, as well as the approach must be related to a data engineering problem of current interest. The Study Program Chair has to approve the topic to ensure it is embedded in the program's overall topic, its aims and goals.

The thesis work comprises the full cycle of a scientific research endeavor: (i) identifying a relevant open research question, (ii) carrying out a literature survey to put the planned work in its context and relate it to the state of the art (SoA), (iii) formulate a concrete research objective, (iv) design a research plan including a statement of criteria to evaluate the success of the project, (v) carry out the plan (with the possibility to change the original plan when motivated), (vi) document the results, (vii) analyze the results with respect to the SoA, the original objective, and the success criteria, and (viii) document all of this in a thesis report. All of this work should be done with as much self-guidance as can be reasonably expected. The instructor will give substantial guidance for (i) and (iii), whereas the other aspects will be addressed with larger degrees of self-guidance.

In the first week of the course, an intense taught tutorial on scientific working, writing and presentation is given. The subsequent weeks follow a seminar style where students present and discuss literature as well as their own results to date. The project consists of the proposal, a thesis report (target size: 30–60 pages, and an oral presentation at the end of the course.
After 4 weeks the student has to submit a project proposal document motivating the master thesis project and defining milestones, with target length of 10 pages, including references. That proposal has to be defended by the student in an oral presentation in the same week.

Before submitting the final version of the master thesis student is asked to given an oral defense of the thesis to both thesis supervisors.

Intended Learning Outcomes

Upon completion of this module, students will be able to

1. independently develop research questions guided by gaps in existing knowledge and determine appropriate research strategies and plans;
2. independently choose and justify appropriate research methods to new unsolved problems or issues;
3. critically assess scientific results and literature;
4. summarize the current state of knowledge in their chosen specialization area;
5. independently apply appropriate knowledge, methods and competencies acquired during their studies;
6. develop conclusions based on their own analysis;
7. use individual feedback to develop and mature within the field of their specialization;
8. effectively communicate and discuss their research results to various audiences;
9. take into consideration social and ethical consequences of their activities.
10. Formulate a research project proposal.
11. presentation of project results for specialists and non-specialists.

Indicative Literature

N. A.

Usability and Relationship to other Modules

Examination Type: Module Examination

Assessment Component 1: Thesis

Length: 30-50 pages
Weight: (75%)

Assessment Component 2: Oral Examination (Master Thesis Defense)

Duration: 20 minutes
Weight: (25%)

Module Achievement: Completion of proposal and proposal presentation are pre-requisites prior to the submission of the thesis.

Scope: All intended learning outcomes of the module

Completion: This module is passed with an assessment-component weighted average grade of 45% or higher.

5 Appendix

5.1 Intended Learning Outcomes Assessment Matrix

Data Engineering Technologies (M.Sc.)					Big Data Challenge for DET	Data Analytics	IT Law	Data Security and Privacy	Statistical and Machine Learning	Ethics and the Information Revolution	Data Acquisition Technologies	Image Processing	Data Base Management Tools in Python	Advanced Data Bases	Parallel and Distributed Computing	Mathematics for Graduate Students	Text Analysis and NLP	Visual Communication and Data Storytelling	Master Thesis
Semester					1	1,2	2	2	2	**2**	3	3	1	2	3	1	2	3	4
Mandatory/ Mandtory elective					m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
Credits					7.5	7.5	2.5	2.5	7.5	2.5	7.5	7,5	7,5	7,5	7,5	7.5	7.5	7.5	30
Competencies*																			
Program Learning Outcomes	**A**	**E**	**P**	**S**															
Critically assess and creatively apply technological possibilities and innovations driven by big data	x	x			x	x							x	x	x		x		x
Use sensors and microcontrollers to collect data and to transmit them to databases on servers or the internet in general	x	x									x								x
Set up and use databases to efficiently and securely manage and access large amounts of data	x	x				x							x	x	x		x		x
Apply statistical concepts and use statistical models in the context of real-life data analytics	x	x				x										x	x		x
Use, adapt and improve visualization techniques to support data-based decision making	x	x									x	x	x				x	x	x
Design, implement and exploit various representations of data for classification and regression including supervised machine learning methods and core ideas of deep learning	x	x							x					x			x		x
Apply and critically assess data acquisition methods and analytical techniques in real life situations, organizations and industries	x	x		x	x	x			x						x		x		x
Independently investigate complex problems and undertake scientific or applied research into a specialist area utilizing appropriate methods, also taking methods and insights of other disciplines into account	x	x		x					x										x
Professionally communicate their conclusions and recommendations, the underlying information and their reasons to specialists and non-specialists, both clearly and unambiguously on the basis of the state of research and application	x	x	x	x	x												x	x	x
Assess and communicate social, scientific and ethical insights that also derive from the application of their knowledge and their decisions		x	x	x	x												x	x	x
Engage ethically with academic, professional and wider communities and actively contribute to a sustainable future			x	x														x	x
Take responsibility for their own learning, personal development and role in society, evaluating critical feedback and self-analysis			x	x												x			x
Adhere to and defend ethical, scientific and professional standards	x	x	x	x	x	x			x		x	x	x	x	x	x	x	x	x
Assessment Type																			
Oral examination																			x
Written examination								x	x			x	x	x	x	x			
Project assessment																			
Project report					x	x					x						x	x	
Practical assessment																			
Program Code					x								x						
Term paper							x			x									
Laboratory report														x					
Poster presentation																			
Presentation																			
Thesis																			x
Module Achievements																			x

*Competencies: A-scientific/academic proficiency; E-competence for qualified employment; P-development of personality; S-competence for engagement in society

Figure 3: ILO Assessment Matrix