JACOBS UNIVERSITY

Study Program Handbook

Data Engineering

Master of Science

Subject-specific Examination Regulations for Data Engineering

The subject-specific examination regulations for Data Engineering 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 (Appendix 1).

Upon graduation students in this program will receive a Master of Science (MSc) degree with a scope of 120 ECTS credit points (for specifics see chapter 2 of this handbook).

Valid for all students starting their studies in Fall 2022

Study Program Chair: Prof. Dr. Stefan Kettemann

http://www.jacobs-university.de/data-engineering

dataengineering@jacobs-university.de

Version	**Valid as of**	**Decision**	**Details**
Fall 2022 – V1	Sep 01, 2022	May 22, 2019	V1 Originally approved by the Academic Senate

Contents

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.

The graduate program in Data Engineering 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. Thus, the program provides the essentials for paving the way to a successful career: computer skills and mathematical understanding paired with practical experience in selected application fields.

The program is embedded into the School of Computer Science & Engineering at Jacobs 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 & Engineering, it includes contributions from and supports applications in the two other research schools: The School of Science (bioactive substances), and the School of Business, Social & Decision Sciences (in modern societies).

Moreover, the Data Engineering program attracts students with diverse career goals, backgrounds, and prior work experience. Therefore, the program offers four focus tracks within which the students can choose to specialize further: Computer Science, Geo-Informatics, Bio-Informatics and Business & Supply Chain Engineering. These tracks are a preparation for the Advanced Projects within the Discovery Area and the Master Thesis.

In particular, one specialization track is Computer Science providing them the skills to go beyond a mere usage of existing toolboxes, and develop innovative data analysis techniques of their own design.

Another specialization track is Bioinformatics and the analysis of biomedical data. Integration and model-based interpretation of high-throughput data are severe bottlenecks in biomedical and pharmaceutical research. Data Engineering prepares students for the novel computational challenges in these fields.

A third specialization track is Geo-Informatics which provides an introduction to Geographic Information System techniques, principles of spatial analysis, and data mining with integration of remote sensing and GPS. It thereby provides an early exposure to earth science data and its handling.

Students can also choose the specialization track of Business & Supply Chain Engineering. A vast amount of data is collected as part of business processes in particular along supply chains. In this specialization track students will concentrate on the full data analysis cycle including pre-processing of data, data analysis and deployment of model results within the business process.

The graduate program in Data Engineering is tailored to a diverse student body (see also Section 1.3) with a wide variety of interests, academic backgrounds, and previous experiences. Small group sizes, a low student-to-teacher ratio, and personalized supervision/advising allow the program to cater 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

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;
- the legal foundations of Data Engineering;
- 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

Upon completion of this program, students will be able to

- critically assess and creatively apply technological possibilities and innovations driven by big data;
- use sensors and microcontrollers to collect data and to transmit them to databases on servers or the internet in general;
- set up and use databases to efficiently and securely manage and access large amounts of data;
- apply statistical concepts and use statistical models in the context of real-life data analytics;
- use, adapt and improve visualization techniques to support data-based decision-making;
- design, implement and exploit various representations of data for classification and regression including supervised machine learning methods and core ideas of deep learning;
- apply and critically assess data acquisition methods and analytical techniques in real life situations, organizations and industries;
- 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;
- professionally communicate their conclusions and recommendations, the underlying information and their reasons to both specialists and non-specialists, clearly and unambiguously on the basis of the state of research and application;

- assess and communicate social, scientific and ethical insights that also derive from the application of their knowledge and their decisions;
- engage ethically with the academic, professional and wider communities and actively contribute to a sustainable future;
- take responsibility for their own learning, personal development, and role in society, evaluating critical feedback and self-analysis;
- take on lead responsibility in a diverse team;
- adhere to and defend ethical, scientific and professional standards.

1.3 Target Audience

The Data Engineering 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 proceed to research-oriented work towards a master or ultimately a PhD degree. Typical examples are:

- a bachelor in computer science who wants to acquire skills in data analysis and micro/macroeconomics for a career in computational finances;
- 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.

In order to facilitate the integration of students with diverse backgrounds, we offer remedial courses in the first semester. Placement tests in the orientation week before the beginning of the first semester help students to identify contents that they need to refresh or remedy.

1.4 Career Options

The demand for Data Engineers is massive. Typical fields of work encompass the finance 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 are able to visualize the findings clearly.
- Graduates of the program work as data analysts, data managers, data architects, business consultants, software and web developers, or system administrators;
- an MSc degree in Data Engineering also allows students to move on to a PhD and a career in academia and research institutions.

The employability of Data Engineering graduates is promoted by organizing contacts with industry and research institutes throughout the curriculum. In the first semester, in the Current Topics in Data Engineering seminar, companies and research groups introduce their field of interest. The advanced projects, in the second and third semesters can be combined with internships in research institutes or companies. In the second and third semester, the participation in public big data challenges is organized as an integral part of the curriculum.

1.5 Admission Requirements

Applicants need to submit the following documents in order to be considered for admission:

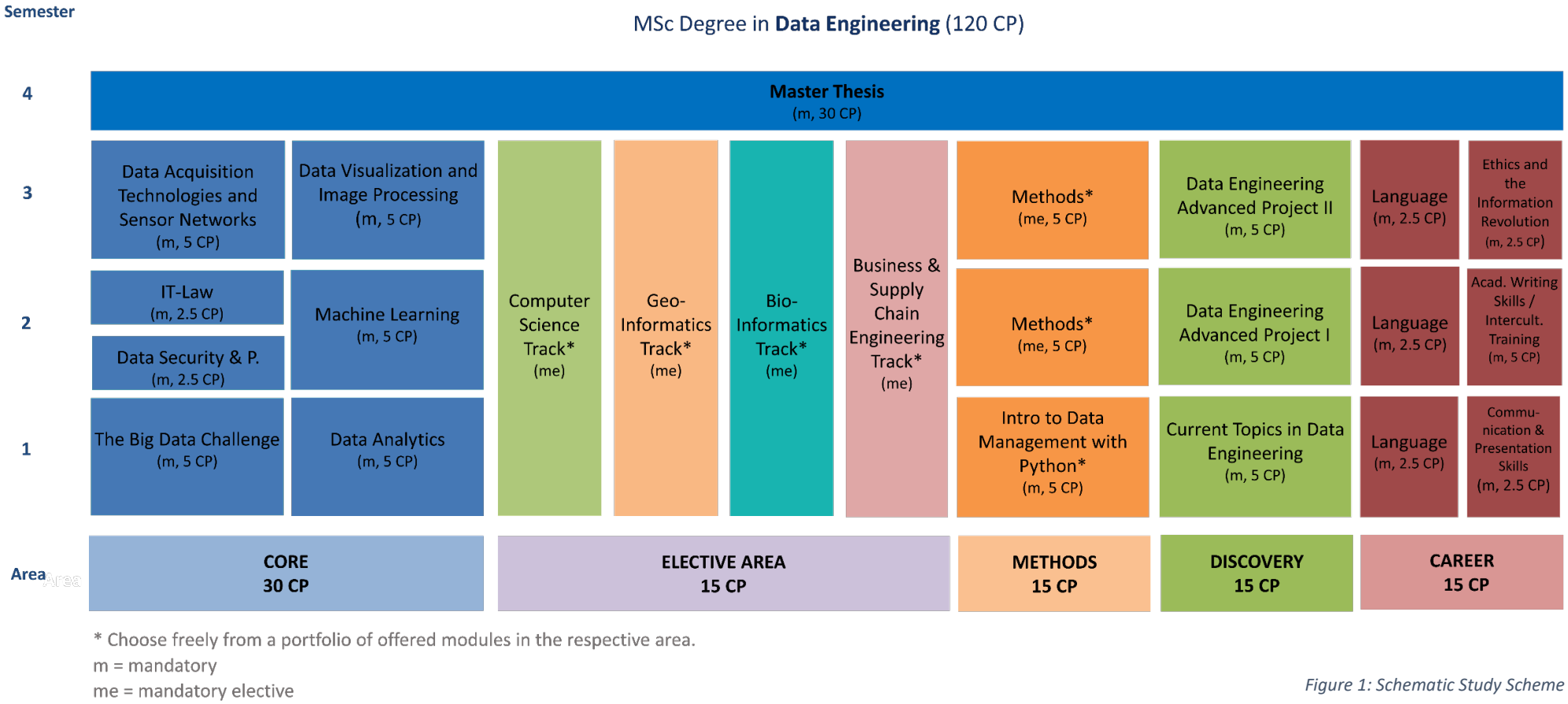
- Letter of motivation
- Curriculum vitae (CV)
- Certified university transcripts in English or German
- Bachelor's degree certificate or equivalent (may be handed in later)
- Two letters of recommendation
- Language proficiency test results (TOEFL, IELTS or equivalent) as outlined on the website.

Please visit http://jacobs-university.de/study/graduate/application-information for more details on the application process.

2 The Curriculum

2.1 The Curriculum at a Glance

The Data Engineering graduate program is composed of foundational lectures, specialized modules, industry seminars and applied project work, leading to a master thesis that can be conducted in research groups at Jacobs University, at external research institutes or in close collaboration with a company. The program takes four semesters (two years). The following table shows an overview of the modular structure of the program. The program is sectioned into five areas (Core, Elective, Methods, Discovery, and Career) and the Master Thesis. All credit points (CP) are ECTS (European Credit Transfer System) credit points. In order to graduate, students need to obtain 120 CP.



* Choose freely from a portfolio of offered modules in the respective area.
m = mandatory
me = mandatory elective

Figure 1: Schematic Study Scheme

See Chapter 3 “Modules” of this handbook for the detailed module descriptions or refer to CampusNet (https://campusnet.jacobs-university.de).

2.2 Study and Examination Plan

MSc Degree in Data Engineering Matriculation Fall 2022							
Module Code	**Program-Specific Modules**	**Type**	**Assessment**	**Period[1]**	**Status[2]**	**Semester**	**CP**
Semester 1							**30**
	CORE Area						**10**
MCO003-BigData	**Module: Big Data Challenge**				**m**	**1**	**5**
MCO003-051003	Big Data Challenge	Lecture	Term paper (Project report)	During semester			
MCO011-DataAnaDE	**Module: Data Analytics**				**m**	**1**	**5**
MCO011-340131	Data Analytics	Lecture	Written examination	Examination period			
	Elective Area				**me**		**5**
	- students choose one module from those listed below						
	Methods Area						**5**
MMM014-IntroDataMan	**Module: Introduction to Data Management with Python**				**m**	**1**	**5**
MMM014-350200	Introduction to Data Management with Python	Lecture/Tutorial	Written examination / Programming assignments	Examination period / During semester			
	Discovery Area						**5**
MRD004-CurTopDE	**Module: Current Topics in Data Engineering**				**m**	**1**	
	Current Topics in Data Engineering	Colloquium	Poster Presentation	During semester			
	Career Area						**5**
MCA006-Commun	**Module: Communication and Presentation Skills for Executives**				**m**	**1**	**2.5**
MCA006-051464	Communication and Presentation Skills for Executives	Seminar	Oral presentation	During semester			
JTLA-xxx	**Module: Language 1**				**m**	**1**	**2.5**
	German is the default language. Native German speakers take modules in another offered language.						
JTLA-xxx	Language 1	Seminar	Various	Various	me		
Semester 2							**27.5**
	CORE Area						**7.5**
MCO013-MachLearn	**Module: Machine Learning**				**m**	**2**	**5**
MCO013-320372	Machine Learning	Lecture	Written examination	Examination period			
MDSSB-LAW-01	**Module: IT Law**				**m**	**2**	**2.5**
MDSSB-LAW-01	IT Law	Lecture	Term paper	Examination period			
	Elective Area				**me**		**5**
	- Students choose a module from those listed below.						
	Methods Area				**me**		**5**
	- Students choose a module from those listed below.						
	Discovery Area						**5**
MRD005-AdvProj1	**Module: Data Engineering Advanced Project I**				**m**	**2**	**5**
MRD005-34001	Data Engineering Advanced Project I	Lecture & Seminar	Term paper (Project report)	flexible			
	Career Area						**5**
MCA008-AcaWri	**Module: Academic Writing Skills/Intercultural Training**				**m**	**2**	**2.5**
MCA008-340231	Academic Writing Skills/Intercultural Training	Lecture	Term Paper	During semester			
JTLA-xxx	**Module: Language 2**				**m**	**2**	**2.5**
JTLA-xxx	Language 2	Seminar	Various	Various	me		

Semester 3							**32.5**
	CORE Area						**12.5**
MCA005-DataSecurity	**Module: Data Security and Privacy**				**m**	**1 or 3**	**2.5**
MCA005-340251	Data Security and Privacy	Lecture	Written examination	Examination period			
MCO014-DataVisImage	**Module: Data Visualization and Image Processing**				**m**	**3**	**5**
MCO014-340231	Data Visualization and Image Processing	Lecture	Written examination	Examination period			
MCO015-DataAquiSens	**Module: Data Acquisition Technologies and Sensor Networks**				**m**	**3**	**5**
MCO015-340112	Data Acquisition Technologies and Sensor Networks	Lecture & Lab	Term paper (Project report)	During semester			
	Elective Area				**me**		**5**
	- Students choose a module from those listed below.						
	Methods Area				**me**		**5**
	- Students choose a module from those listed below.						
	Discovery Area						**5**
MRD006-AdvProj2	**Module: Data Engineering Acvanced Project II**				**m**	**3**	**5**
MRD006-340002	Data Engineering Acvanced Project II	Project work	Term paper (Project report)	flexible			
	Career Area						**5**
MDSSB-EIR-01	**Module: Ethics and the Information Revolution**				**m**	**3**	**2.5**
MDSSB-EIR-01	The Information Revolution	Seminar	Term paper (Project report)	During semester			2.5
JTLA-xxx	**Module: Language 3**				**m**	**3**	**2.5**
JTLA-xxx	Language 3	Seminar	Various	Various			
Semester 4							**30**
	Master Thesis						**30**
MMT003-MasterThesis	**Module: Master Thesis MSc DE**				**m**	**4**	**30**
MMT003-340003	Master Thesis						
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 http://www.jacobs-university.de/academic-calendar).

[2] m = mandatory, me = mandatory elective

Elective Area *Students choose 15 CP of manadatory electives*							
	Computer Science Track						**20**
MECS001-StatMod	**Module: Principles of Statistical Modeling**				**me**	**2**	**5**
MECS001-340101	Principles of Statistical Modeling	Lecture	Project Report	During semester			
MECS002-NetworkTheo	**Module: Network Theory**				**me**	**1 or 3**	**5**
MECS002-340212	Network Theory	Lecture	Written examination	Examination period			
MCO012-AdvDataBase	**Module: Advanced Databases**				**me**	**2**	**5**
MCO012-340152	Advanced Databases	Lecture	Written examination	Examination period			2.5
MCO012-340153	Advanced Databases Lab	Lab	Lab project	During semester			2.5
MECS004-ParDisCom	**Module: Parallel and Distributed Computing**				**me**	**3**	**5**
MECS004-30040	Parallel and Distributed Computing	Lecture	Written examination	Examination period			
	Geoinformatics Track						**10**
MEGI001-GeoInf	**Module: Geoinformatics**				**me**	**1 or 3**	**5**
MEGI001-210213	Geo-Information Systems	Lecture	Term paper	Examination period	m		2.5
MEGI001-210103	Introduction to Earth System Data	Lecture			m		2.5
MEGI002-GeoInfLab	**Module: Geoinformatics Lab**				**me**	**2**	**5**
MEGI002-210214	Geoinformatics Lab	Lecture	Term paper	Examination period			
	Bio-Informatics Track						**15**
MEBI001-IntroSysBio	**Module: Introduction to Systems Biology**				**me**	**2**	**5**
MEBI001-550432	Introduction to Systems Biology	Lecture	Written examination	Examination period			
MDE-BIO-03	**Management and Analysis of Biological and Medical Data**				**me**	**1 or 3**	**5**
MDE-BIO-03	Management and Analysis of Biological and Medical Data	Seminar	Oral Examination	Examination period			
	Business & Supply Chain Engineering Track						**10**
MESC001-DataMin	**Module: Data Mining**				**me**	**2**	**5**
MESC001-340122	Data Mining	Lecture	Term paper (Project report)	During semester			
MCO008-DataAnaSCM	**Module: Data Analytics in Supply Chain Management**				**me**	**1 or 3**	**5**
MCO008-051008	Data Analytics in Supply Chain Management	Lecture	Term paper (Project report)	During semester			
Total CP							**65**

Methods Area							
Students take "Introduction to Data Management with Python" in the first semester and choose 2 modules from the list below in semester 2 and 3.							
							20
MMM004-ModDynSys	**Module: Modeling and Control of Dynamical Systems**				**me**	**2**	**5**
MMM004-340103	Modeling and Control of Dynamical Systems	Seminar	Written examination	Examination period			
MMM005-ModSigProc	**Module: Modern Signal Processing**				**me**	**2**	**5**
MMM005-340153	Modern Signal Processing	Seminar	Oral presentation	During semester			
MMM007-NetBioMed	**Module: Network Approaches in Biology and Medicine**				**me**	**3**	**5**
MMM007-550443	Network Approaches in Biology and Medicine	Lecture	Oral presentation	During semester			
MMM008-ApplDynSys	**Module: Applied Dynamical Systems**				**me**	**2**	**5**
MMM008-110231	Applied Dynamical Systems	Lecture	Term paper (Project report)	During semester			
	Remedial Courses (Methods Area)						**10**
MMM009-CalLinAlg	**Module: Calculus and Linear Algebra for Graduate Students**				**me**	**1**	**5**
MMM009-340181	Calculus and Linear Algebra for Graduate Students	Lecture	Written examination	Examination period			
MMM011-ProbabGS	**Module: Probabilities for Graduate Students**				**me**	**1**	**5**
MMM011-340171	Probabilities for Graduate Students	Lecture	Written examination	Examination period			
Total CP							**30**

Figure 2: Study and Examination Plan

2.3 Core Area (30 CP)

This area is the centerpiece of the Data Engineering program. The six mandatory modules in the Core Area cover essential methods of data engineering. They provide the foundations for further, more advanced courses and applied projects by introducing the fundamental concepts, methods and technologies used in data engineering. The modules are intensive courses accompanied by hands-on tutorials and labs.

Core Modules					
Module Title	**Module No.**	**Semester**	**Mandatory**	**Instructor**	**CP**
The Big Data Challenge	MDE-CO-01	1	yes	Wilhelm	5
Data Analytics	MDE-CO-02	1	yes	Wilhelm	5
Machine Learning	MDE-CO-04	2	yes	Kettemann	5
Data Security and Privacy	MDE-CO-03	1,3	yes	Zaspel	2.5
IT Law	MDSSB-LAW-01	2	yes	Brockmann/Kettemann	2.5
Data Visualization and Image Processing	MDE-CO-05	3	yes	Kettemann	5
Data Acquisition Technologies and Sensor Networks	MDE-CO-06	3	yes	Hu	5

2.4 Elective Area (15 CP)

The Data Engineering program attracts students with diverse career goals, backgrounds, and prior work experience. Therefore, modules in this area can be chosen freely by students depending on their prior knowledge and interests. Students can choose to strengthen their knowledge by following one of four suggested focus tracks and electing the modules offered therein: Computer Science, Geo-Informatics, Bio-Informatics and Business & Supply Chain Engineering. These tracks are a preparation for the Advanced Projects within the Discovery Area and the Master Thesis.

Students may choose any combination of the modules listed below. Each track may be followed completely and/or complemented with other modules (as necessary in case of the tracks with 10 CP). In addition to the modules offered within these focus tracks, 3rd year modules from the undergraduate curriculum or other graduate programs at Jacobs University can be taken with the approval of the program coordinator. Please see CampusNet (https://campusnet.jacobs-university.de) for current offerings.

To enhance flexibility, students may transfer modules between the Elective and the Methods Areas (except for remedial modules) after consulting their academic advisor.

Elective Modules					
Computer Science Track					
Module Title	**Module No.**	**Semester**	**Mandatory**	**Instructor**	**CP**
Principles of Statistical Modeling	MDE-CS-03	2	no	Kettemann	5
Advanced Data Bases	MDE-CS-04	2	no	Baumann	5
Network Theory	MDE-CS-01	1,3	no	Kettemann	5
Parallel and Distributed Computing	MDE-CS-02	3	no	Zaspel	5
Geo-Informatics Track					
Module Title	**Module No.**	**Semester**	**Mandatory**	**Instructor**	**Credits**
Geo Informatics	MDE-GEO-01	1	no	Unnithan	5
Geo-Informatics Lab	MDE-GEO-02	2	no	Unnithan	5
Bio-Informatics Track					
Modeling and Analysis of Complex Systems	MDE-BIO-01	1,3	no	Merico	5
Introduction to Systems Biology	MDE-BIO-02	2	no	Hütt	5
Business & Supply Chain Engineering Track					
Data Mining	MDE-BSC-01	2	no	Wilhelm	5
Data Analytics in Supply Chain Management	MSCM-CO-07	3	no	Wicaksono	5

2.5 Methods Area (15 CP)

In the Methods Area advanced concepts, methods and technologies of data engineering are introduced with a view towards industrial applications. Students can choose freely from the modules in this area. To enhance flexibility, students may transfer modules between the Elective and the Methods Areas (except for remedial modules) after consulting their academic advisor.

Methods Modules					
Module Title	**Module No.**	**Semester**	**Mandatory**	**Instructor**	**CP**
Introduction to Data Management with Python	MDE-MET-03	1	yes	Brandt	5
Modeling and Control of Dynamical Systems	MDE-MET-04	2	no	Bode	5
Modern Signal Processing	MDE-MET-07	2 (biannually)	no	Abreu	5
Network Approaches in Biology and Medicine	MDE-MET-05	1,3	no	Hütt	5
Applied Dynamical Systems	MDE-MET-06	2 (biannually)	no	Oliver	5

Within the Methods Area Jacobs University offers special remedial modules, which are recommended to refresh knowledge or to fill knowledge gaps, preparing students to successfully take the Data Engineering Core Area modules. Based on a placement test in the orientation week, the academic advisor will propose which of the modules are useful depending on prior knowledge of the student.

Remedial Modules (Method Area)					
Module Title	**Module No.**	**Semester**	**Mandatory**	**Instructor**	**CP**
Calculus and Linear Algebra for Graduate Students	MDE-MET-01	1	no	Gorbovickis	5
Probabilities for Graduate Students	MDE-MET-02	1	no	Bode	5

2.6 Discovery Area (15 CP)

This area features in the first semester a Project Seminar introducing the students to Current Topics and Challenges in Data Engineering, which is followed by two advanced projects in Data Engineering in semesters 2 and 3, each of which is worth 5 CP. The projects can be done in the research groups at Jacobs University or during internships at companies. The projects are supervised by Jacobs University faculty.

Discovery Modules					
Module Title	**Module No.**	**Semester**	**Mandatory**	**Instructor**	**CP**
Current Topics in Data Engineering	MDE-DIS-01	1	yes	Kettemann, DE faculty	5
Data Engineering Advanced Project I	MDE-DIS-02	2	yes	entire faculty	5
Data Engineering Advanced Project II	MDE-DIS-03	3	yes	entire faculty	5

2.7 Career Area (15 CP)

In this area students acquire skills to prepare them for a career as data engineers in industry.

Career Modules					
Module	**Module No.**	**Semester**	**Mandatory**	**Instructor**	**CP**
Language Skills	MCA002	1, 2, 3	Yes	LCC	7.5
Communication & Presentation Skills for Executives	MDE-CAR-01	1	yes	Kettemann	2.5
Academic Writing Skills/Intercultural Training	MDE-CAR-02	2	yes	Kettemann	2.5
Ethics and the Information Revolution	MDSSB-EIR-01	3	yes	Brockmann	2.5

2.8 Master Thesis (30 CP)

In the fourth semester, students conduct research and write a master thesis guided and supported by their academic advisor.

Module Title	**Module No.**	**Semester**	**Mandatory**	**Instructor**	**Credits**
Master Thesis	MDE-THE-01	4	yes	entire faculty	30

3 Data Engineering Modules

3.1 Core Area (30 CP)

3.1.1 Big Data Challenge

<table>
<tr><td colspan="3">Module Name
Big Data Challenge</td><td>Module Code
MDE-CO-01</td><td>Level (type)
Year 1
(CORE)</td><td>CP
5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-CO-01 | Big Data Challenge | Lecture | 5</td></tr>
<tr><td>Module Coordinator

Prof. Dr. Adalbert F.X. Wilhelm</td><td colspan="3">Program Affiliation

▪ MSc Data Engineering</td><td colspan="2">Mandatory Status

Mandatory for DE and SCM</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites
☒ None

Co-requisites
☒ None

Knowledge, Abilities, or Skills
▪ Researching information, assessing sources and report writing</td><td>Frequency

Annually
(Fall)

Duration

1 semester</td><td colspan="2">Forms of Learning and Teaching

▪ Lecture (17.5 hours)
▪ Project work (90 hours)
▪ Private study (17.5 hours)

Workload

125 hours</td></tr>
<tr><td colspan="6">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.</td></tr>
<tr><td colspan="6">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. Technological research quite naturally 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.
In this module students receive an overview of big data by looking at it from various perspectives, primarily the business and societal points of view. The focus is not on the technical methods and skills, but on case studies that show big data and data engineering in a cross-section.</td></tr>
</table>

Intended Learning Outcomes

Upon completion of this module, students will be able to

- contribute knowledgeably to the current debate about big data, digitalization and industry 4.0;
- explain and discuss pros and cons of digitalization from a business perspective as well as a societal perspective;
- perform a SWOT analysis on current big data initiatives;
- evaluate technological possibilities and innovations driven by big data;
- assess the business opportunities of current big data developments.

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

- For DE: This module provides an overview on practical big data applications. The computational details will then be studied in MDE-CS-04.
- For SCM: Concepts are applied in MSCM-CO-03 Trends & Challenges in Supply Chain Management. Project management concepts taught in MSCM-CO-01 will be applied. Academic writing skills taught in MSCM-CAR-01 facilitate the completion of the tasks in this module.

Examination Type: Module Examination

Assessment Type: Project Report — Length: 2.500 words
Weight: 100%

Scope: All intended learning outcomes of this module.

3.1.2 IT Law

<table>
<tr><td colspan="3">Module Name
IT Law</td><td>Module Code
MDSSB-LAW</td><td>Level (type)
Year 1 (CORE)</td><td>CP
2.5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDSSB-LAW-01 | IT Law | Lecture | 2.5</td></tr>
<tr><td>Module Coordinator

Prof. Dr. Adalbert F.X. Wilhelm</td><td colspan="3">Program Affiliation

▪ MSc Data Science for Society and Business</td><td colspan="2">Mandatory Status

Mandatory for DE, mandatory elective for DSSB</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites: ☒ None
Co-requisites: ☒ None
Knowledge, Abilities, or Skills: ☒ None</td><td>Frequency

Annually (Spring)</td><td colspan="2">Forms of Learning and Teaching

▪ Lecture (17.5 hours)
▪ Private study (45 hours)</td></tr>
<tr><td colspan="3"></td><td>Duration

1 semester</td><td colspan="2">Workload

62.5 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation

Read the Syllabus.</td></tr>
<tr><td colspan="6">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 is in conflict 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 looks into 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).</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

By the end of this module, students should be able to
• identify legal questions and implications in relation to digital transformation technologies/IT law/ AI and algorithms
• understand fundamental national and international legal frameworks related to the use of data
• know the relevant IP rights regarding data and algorithms
• understand and critically assess legal regulations about data privacy and data protection
• recognize and explain the types of bias inherent in data processing
• explain the legal concerns related to data-based automatic decision making
• understand how to comply to the GDPR and assess its impact on individuals, firms, and organizations
• understand and critically evaluate the liabilities and available remedies with regard to data
• explain and develop potential future IT regulation mechanisms</td></tr>
</table>

Indicative Literature

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

Usability and Relationship to other Modules

- For DSSB students: It is one of the three Career modules (IT Law, Language III, and Ethics and the Information Revolution) that can be chosen for replacement by the internship. Students need to replace 10 CP for the internship.

Examination Type: Module Examination

Assessment Type: Term Paper

Length: 3.500 words
Weight: 100%

Scope: All intended learning outcomes of this module.

3.1.3 Data Security and Privacy

Module Name	*Module Code*	*Level (type)*	*CP*
Data Security and Privacy	MDE-CO-03	Year 1 or 2 (CORE)	2.5

Module Components

Number	*Name*	*Type*	*CP*
MDE-CO-03	Data Security and Privacy	Lecture	2.5

Module Coordinator	*Program Affiliation*	*Mandatory Status*
Prof. Dr. Peter Zaspel	▪ MSc Data Engineering	Mandatory for DE

Entry Requirements			*Frequency*	*Forms of Learning and Teaching*
Pre-requisites	*Co-requisites*	*Knowledge, Abilities, or Skills*	Annually (Fall)	▪ Seminar (17.5 hours) ▪ Private Study (45 hours)
☒ None	☒ None	☒ None	*Duration*	*Workload*
			1 semester	62.5 hours

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

After successful completion of this module, students will be able to:

- analyze and develop principles for public key encryption;
- assess and choose appropriate techniques for authentication;
- understand the design of internet standards;
- summarize and communicate the principles behind encryption using shared keys;
- critically assess and identify how security issues are solved and how this will impact the 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

N.A.

Examination Type: Module Examination

Assessment type: Written examination

Duration: 90 minutes
Weight: 100%

Scope: All intended learning outcomes of this module.

3.1.4 Data Analytics

<table>
<tr><td colspan="3">Module Name
Data Analytics</td><td>Module Code
MDE-CO-02</td><td>Level (type)
Year 1 (CORE)</td><td>CP
5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-CO-02 | Data Analytics | Lecture | 5</td></tr>
<tr><td>Module Coordinator

Prof. Dr. Adalbert F.X. Wilhelm</td><td colspan="3">Program Affiliation

▪ MSc Data Engineering</td><td colspan="2">Mandatory Status

Mandatory for DE
Mandatory elective for DSSB</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites
☒ None

Co-requisites
☒ None

Knowledge, Abilities, or Skills
☒ None</td><td>Frequency

Annually (Fall)

Duration

1 semester</td><td colspan="2">Forms of Learning and Teaching

▪ Lecture (17.5 hours)
▪ Tutorials (17.5 hours)
▪ Private study (90 hours)

Workload

125 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation
Read the Syllabus.
Take the free online course: Introduction to Data Science at https://cognitiveclass.ai/courses/data-science-101/</td></tr>
<tr><td colspan="6">Content and Educational Aims

This module introduces concepts and methods of data analytics. The objective of the module is to present methods for gaining insight 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 course takes an applied approach and combines the theoretical foundation of data analytics with a practical exposure to the data analysis process.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

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

▪ explain advanced data analytics techniques in theory and application;
▪ apply data analytics methods to real-life problems using appropriate tools;
▪ evaluate and compare different data analytics algorithms and approaches;
▪ apply statistical concepts to evaluate data analytics results.</td></tr>
<tr><td colspan="6">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)</td></tr>
<tr><td colspan="6">Usability and Relationship to other Modules</td></tr>
</table>

In this module students will learn concepts and various techniques for data analysis. They will be rigorously applied in MDE-CS-03 as well as in the applied projects MDE-DIS-02 and MDE-DIS-03, and typically also in the master thesis.
Examination Type: Module Examination Assessment Type: Project Report — Length: 20 pages Weight: 100% Scope: All intended learning outcomes of this module.

3.1.5 Machine Learning

<table>
<tr><td colspan="3">Module Name
Machine Learning</td><td>Module Code
MDE-CO-04</td><td>Level (type)
Year 1 (CORE)</td><td>CP
5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-CO-04 | Machine Learning | Lecture | 5</td></tr>
<tr><td>Module Coordinator

Prof. Dr. Stefan Kettemann</td><td colspan="3">Program Affiliation

▪ MSc Data Engineering</td><td colspan="2">Mandatory Status

Mandatory for DE

Mandatory Elective for CSSE and DSSB</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites: ☒ None
Co-requisites: ☒ None
Knowledge, Abilities, or Skills:
▪ Basic linear algebra, calculus and probability theory, as typically acquired in entry modules in BSc studies</td><td>Frequency

Annually (Spring)</td><td colspan="2">Forms of Learning and Teaching

▪ Lectures (35 hours)
▪ Private Study, incl. exercises and exam preparation (90 hours)</td></tr>
<tr><td colspan="3"></td><td>Duration

1 semester</td><td colspan="2">Workload

125 hours</td></tr>
<tr><td colspan="6">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.</td></tr>
<tr><td colspan="6">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.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

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

▪ design, implement and exploit elementary supervised ML methods for classification and regression with expert care given to dimension reduction preprocessing and regularization;
▪ understand and practically use PCA and linear regression;</td></tr>
</table>

- 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

This module is a natural companion to the "Principles of Statistical Modeling" (PSM) module MDE-CS-03. The ML module focuses on practical ML skills, whereas PSM module on rigorous mathematical formalism and analysis. For students not familiar with graph theory, it is recommended to take the first semester course MDE-CS-01 Network Theory, which introduces concepts used in this Machine Learning module.

Examination Type: Module Examination

Assessment Type: Written Exam

Duration: 120 minutes
Weight: 100%

Scope: All intended learning outcomes of this module.

3.1.6 Data Visualization and Image Processing

<table>
<tr><td colspan="3">Module Name
Data Visualization and Image Processing</td><td>Module Code
MDE-CO-05</td><td>Level (type)
Year 2 (CORE)</td><td>CP
5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-CO-05 | Data Visualization and Image Processing | Lecture | 5</td></tr>
<tr><td>Module Coordinator

Prof. Dr. Stefan Kettemann</td><td colspan="3">Program Affiliation

▪ MSc Data Engineering</td><td colspan="2">Mandatory Status

Mandatory for DE</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites
☒ None

Co-requisites
☒ None

Knowledge, Abilities, or Skills
▪ Basic linear algebra, calculus and programming skills</td><td>Frequency

Annually (Fall)

Duration

1 semester</td><td colspan="2">Forms of Learning and Teaching

▪ Lectures (35 hours)
▪ Private Study, incl. exercises and exam preparation (90 hours)

Workload

125 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation

Read the syllabus.</td></tr>
<tr><td colspan="6">Content and Educational Aims

This module introduces the basic concepts of (1) data visualization and (2) image processing.

(1) Computer-based visualization systems provide visual representations of datasets intended to help people carry out certain task more effectively. These datasets can come from very diverse sources, such as scientific experiments, simulations, medical scanners, commercial databases, financial trans-actions, health records, social networks and the like. In the This module deals with effective visual mappings as well as interaction principles for various data, to develop an understanding of the perceptual and cognitive aspects of visual representations. Students learn how to evaluate visualization systems.

(2) The second half of the module focuses on image processing and delves into questions of how we can digitally process image data. Topics include for instance sampling and quantization strategies, image segmentation, image transformations, noise reduction and feature extraction.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

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

▪ represent and interact with various data visually;
▪ evaluate visual depictions of data and find possible improved presentations;
▪ assist users in visual data analysis;
▪ understand transforms and being able to apply them to 2D images.</td></tr>
<tr><td colspan="6">Indicative Literature

M. O. Ward, G. Grinstein, D. Keim, Interactive Data Visualization: Foundations, Techniques, and Applications, Second Edition, Matthew O. Ward, Georges Grinstein, Daniel Keiml, 2015, ISBN, 9781482257373.

A. C. Telea, Data Visualization: Principles and Practice, Second Edition, A K Peters, 2014, ISBN, 9781466585263.</td></tr>
</table>

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 MDE-CO-02 and the Data Mining module MDE-BSC-01.

Examination Type: Module Examination

Assessment Type: Written Exam

Duration: 120 minutes
Weight: 100%

Scope: All intended learning outcomes of this module.

3.1.7 Data Acquisition Technologies and Sensor Networks

<table>
<tr><td colspan="3">Module Name
Data Acquisition Technologies and Sensor Networks</td><td>Module Code
MDE-CO-06</td><td>Level (type)
Year 2 (CORE)</td><td>CP
5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-CO-06 | Data Acquisition Technologies and Sensor Networks | Lecture and Lab | 5</td></tr>
<tr><td>Module Coordinator

Dr. Fangning Hu</td><td colspan="3">Program Affiliation

▪ MSc Data Engineering</td><td colspan="2">Mandatory Status

Mandatory for DE</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites
☒ None

Co-requisites
☒ None

Knowledge, Abilities, or Skills
▪ The students should be familiar with at least some of the following topics: basic electrical circuits, microcontrollers, HTML, PHP, SQL, C, and Python.</td><td>Frequency

Twice per year

Duration

1 semester</td><td colspan="2">Forms of Learning and Teaching

▪ Lectures and Lab (35 hours)
▪ Private Study (90 hours)

Workload

125 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation

Read the syllabus.
A lab manual will be provided, reading the lab manual before each lab session is recommended.</td></tr>
<tr><td colspan="6">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.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

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

▪ acquire data from different sensors and use a microcontroller to process them;
▪ transmit data from the microcontroller to a database on a server
▪ collect data from web browsers and transmit them to a database on a server
▪ visualize the data on computers or smart devices
▪ set up a wireless sensor network and communicate data among different components.</td></tr>
</table>

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 will later be processed and analyzed by techniques studied in the Data Analytics module MDE-CO-02, the Machine Learning module MDE-CO-04, and the Data Analytics in Supply Chain Management module MSCM-CO-07.

Examination Type: Module Examination

Assessment Type: Project Report

Length: 20 pages
Weight: 100%

Scope: All intended learning outcomes of this module.

3.2 Elective Area (15 CP)

3.2.1 Computer Science Modules

3.2.1.1 Principles of Statistical Modeling

<table>
<tr><td colspan="3">Module Name
Principles of Statistical Modeling</td><td>Module Code
MDE-CS-03</td><td>Level (type)
Year 1 (Elective)</td><td>CP
5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-CS-03 | Principles of Statistical Modeling | Lecture | 5</td></tr>
<tr><td>Module Coordinator

Prof. Dr. Stefan Kettemann</td><td colspan="3">Program Affiliation

▪ MSc Data Engineering</td><td colspan="2">Mandatory Status

Mandatory Elective for DE</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites
☒ None

Co-requisites
☒ None

Knowledge, Abilities, or Skills
▪ Basic linear algebra, calculus and probability theory, as typically acquired in entry modules in BSc studies</td><td>Frequency

Annually (Spring)

Duration

1 semester</td><td colspan="2">Forms of Learning and Teaching

▪ Lectures (35 hours)
▪ Private Study (90 hours)

Workload

125 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation

Read the syllabus.</td></tr>
<tr><td colspan="6">Content and Educational Aims

This module introduces the basic concepts of statistical modeling. The focus is on a thorough understanding of fundamental concepts: the nature of probability spaces and random variables; distributions and their representations; design and critical assessment of real-life samples; statistical hypothesis testing; statistical decision-making; strategies for estimator design. This module is distinguished from standard probability courses for non-mathematical audiences in that the mathematical model of „probability" is rigorously introduced, including sigma-fields.

The primary educational aim is to lift students to a level of mastery and understanding of the intricate formalism of probability and statistics that enables them to read the respective scientific literature and to adapt existing algorithms or even develop new algorithms. This module is thus targeted at students who want to go beyond a mere mechanical use of existing statistical toolboxes, and develop innovative data analysis techniques of their own design.

The secondary educational aims are to enable students to (i) understand the substantial differences between methodological approaches and fundamental mindsets in statistics vs. machine learning and (ii) understand the differences between and respective advantages and disadvantages of classical frequentist vs. Bayesian modeling methods.</td></tr>
</table>

Intended Learning Outcomes

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

- correctly and insightfully use the core formalism of probability theory;
- understand the (basic) formalism used in the scientific literature of machine learning and statistics;
- decide which type of approach is indicated to address a given modeling task (frequentist vs. Bayesian; black-box-modeling in machine learning spirit vs. statistical decision procedures; maximum-likelihood vs. Bayesian vs. unbiasedness criteria for procedure selection);
- appreciate the importance of being exact and circumspective in setting up statistical modeling procedures.

Indicative Literature

H. Jäger, Principles of Statistical Modeling, online tutorial http://minds.jacobs-university.de/teaching/courses/t2019psm/

V. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, 1995.

R. J. Hyndman, G. Athanasopoulos Forecasting, Principles and Practice, , online script, https://otexts.com/fpp2/.

Usability and Relationship to other Modules

The Machine Learning module MDE-CO-04 and the Data Analytics module MDE-CO-02 are complementary in that they introduce and focus on practical tools and techniques, whose theoretical foundations only can become fully clear in this module.

Examination Type: Module Examination

Assessment Type: Project Report

Length: 10 pages
Weight: 100%

Scope: All intended learning outcomes of this module.

3.2.1.2 Network Theory

<table>
<tr><td colspan="3">Module Name
Network Theory</td><td>Module Code
MDE-CS-01</td><td>Level (type)
Year 1/2 (Elective)</td><td>CP
5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-CS-01 | Network Theory | Lecture | 5</td></tr>
<tr><td>Module Coordinator

Prof. Dr. Stefan Kettemann</td><td colspan="3">Program Affiliation

▪ MSc Data Engineering</td><td colspan="2">Mandatory Status

Mandatory Elective for DE</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites
☒ None

Co-requisites
☒ None

Knowledge, Abilities, or Skills
▪ Basic linear algebra, calculus and probability theory, as typically acquired in entry modules in BSc studies</td><td>Frequency

Annually (Fall)

Duration

1 semester</td><td colspan="2">Forms of Learning and Teaching

▪ Lectures (35 hours)
▪ Private Study, incl. exercises and exam preparation (90 hours)

Workload

125 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation

Read the syllabus. Refresh your Linear Algebra. Read the first two chapters of the primary book Networks: An Introduction by M.E.J Newman, ISBN 9780199206650 (2010)</td></tr>
<tr><td colspan="6">Content and Educational Aims

The theory of networks - as diverse as power grids, computer networks, social networks, and biological networks - has emerged in recent years as a highly dynamic and rapidly developing discipline. The study of networks is broadly interdisciplinary and important developments have occurred in many fields, including mathematics, physics, computer and information sciences, biology, and the social sciences. This module introduces this field, starting with a review of the diverse realizations of networks. We then teach how to measure the structure of networks and introduce methods for analyzing network data.

We introduce graph theory, which forms the basis of network theory. Then, we review computer algorithms and spectral methods to analyze networks. We introduce various mathematical models of networks, including random graph models and generative models, and conclude with more recent theories that model the dynamical processes taking place on networks.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

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

▪ measure structure of networks;
▪ analyze network data;
▪ perform the modeling of dynamic processes on networks;</td></tr>
</table>

- communicate in scientific language using advanced field-specific technical terms.

Indicative Literature

M. Newman, Networks an Introduction, Oxford Univ. Press, 2010, ISBN: 9780199206650.

A.-L. Barabasi, Network Science, Cambridge University Press, Cambridge, 2016, ISBN-10: 1107076269.

Usability and Relationship to other Modules

This course prepares for the courses MDE-CO-04 Machine Learning and MDE-CS-03 Principles of Statistical Modeling.

Examination Type: Module Examination

Assessment Type: Written Exam

Duration: 120 minutes
Weight: 100%

Scope: All intended learning outcomes of this module.

3.2.1.3 Advanced Databases

<table>
<tr><td colspan="2">Module Name
Advanced Databases</td><td>Module Code
MDE-CS-04</td><td>Level (type)
Year 1 (Elective)</td><td>CP
5</td></tr>
<tr><td colspan="5">Module Components

Number | Name | Type | CP
MDE-CS-04-A | Advanced Databases | Lecture | 2.5
MDE-CS-04-B | Advanced Databases Lab | Lab | 2.5</td></tr>
<tr><td>Module Coordinator

Prof. Dr. Peter Baumann</td><td colspan="2">Program Affiliation

▪ MSc Data Engineering</td><td colspan="2">Mandatory Status

Mandatory Elective for DE and CSSE</td></tr>
<tr><td rowspan="2">Entry Requirements

Pre-requisites

Co-requisites

☒ None

Knowledge, Abilities, or Skills

• 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</td><td>Frequency

Annually (Spring)</td><td colspan="2">Forms of Learning and Teaching

▪ Lecture (40 hours)
▪ Lab (40 hours)
▪ Private study (45 hours)</td></tr>
<tr><td>Duration

1 semester</td><td colspan="2">Workload

125 hours</td></tr>
<tr><td colspan="5">Recommendations for Preparation

N.A.</td></tr>
<tr><td colspan="5">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 Core lecture Data Management 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.</td></tr>
<tr><td colspan="5">Intended Learning Outcomes

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

▪ Summarize the state of the art in data management for particularly large and complex data
▪ Establish criteria for selecting adequate scalable data management technology based on various criteria
▪ Establish a state of the art database schema for a given application scenario
▪ Tune a relational database for best performance on some given query workload</td></tr>
</table>

- Adequately consider security aspects in databases
- 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

Pre-requisite Introduction to Data Management with Python.

Examination Type: Module Component Examinations

Module Component 1: Lecture

Assessment Type: Written Exam — Duration: 120 min, Weight: 67%

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

Module Component 2: Lab

Assessment Type: Lab 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%.

3.2.1.4 Parallel and Distributed Computing

<table>
<tr><td colspan="3">Module Name
Parallel and Distributed Computing</td><td>Module Code
MDE-CS-02</td><td>Level (type)
Year 2
(Elective)</td><td>CP
5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-CS-02 | Parallel and Distributed Computing | Lecture | 5</td></tr>
<tr><td colspan="2">Module Coordinator

Prof. Dr. Peter Zaspel</td><td colspan="2">Program Affiliation

▪ MSc Data Engineering</td><td colspan="2">Mandatory Status

Mandatory elective for DE, CSSE, RIS (BSc) and CS (BSc)</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites: ☒ None
Co-requisites: ☒ None
Knowledge, Abilities, or Skills:
▪ Basic knowledge in C/C++
▪ Mandatory proficiency in Python</td><td>Frequency

Annually (Fall)

Duration

1 semester</td><td colspan="2">Forms of Learning and Teaching

▪ Lecture (35 hours)
▪ Private study (90 hours)

Workload

125 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation
If no knowledge in C/C++ is present, interested students are encouraged get a basic understanding of C/C++ (via online material) in order to better understand some of the discussed concepts.</td></tr>
<tr><td colspan="6">Content and Educational Aims

In the 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 dataanalysis (OpenMP / MPI) and aim at understanding performance and scalability in this field (weak vs. strong scaling, Amdahl's law).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.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

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

• understand theory and fundamentals of parallelization models (shared-/distributed memory, SIMD, SIMT)
• explain and apply parallel programming methodologies (OpenMP / MPI)
• describe and analyze performance and scalability (weak vs. strong scaling, ...)
• Understand basic principles of distributed and cloud computing
• use distributed processing frameworks (Spark / Hadoop MapReduce / Dask) for scalable distributed calculations
• develop scalable machine learning and data processing on Big Data</td></tr>
<tr><td colspan="6">Indicative Literature

Zaccone, Python Parallel Programming Cookbook, O'Reilly.

J.C. Daniel, Data Science with Python and Dask, Manning Publications.</td></tr>
</table>

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

Usability and Relationship to other Modules

N.A.

Examination Type: Module Examination

Assessment Type: Written Exam

Duration: 120 minutes
Weight: 100%

Scope: All intended learning outcomes of this module.

3.2.2 Geoinformatics Track

3.2.2.1 Geoinformatics

<table>
<tr><td colspan="3">Module Name
Geoinformatics</td><td>Module Code
MDE-GEO-01</td><td>Level (type)
Year 1
(Elective)</td><td>CP
5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-GEO-01-A | Geo-Information Systems | Lecture | 2.5
MDE-GEO-01-B | Introduction to Earth System Data | Lecture | 2.5</td></tr>
<tr><td>Module Coordinator

Prof. Dr. Vikram Unnithan</td><td colspan="3">Program Affiliation

▪ MSc Data Engineering
▪ BSc Earth & Environmental Science</td><td colspan="2">Mandatory Status

Mandatory elective for DE and DSSB</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites

☒ None

Co-requisites

☒ None

Knowledge, Abilities, or Skills

▪ Basic computer skills, basic working knowledge of Linux OS and Python</td><td>Frequency

Annually (Fall)

Duration
1 semester</td><td colspan="2">Forms of Learning and Teaching

▪ Lecture attendance (40 hours)
▪ Practical assignments (40 hours)
▪ Private study (45 hours)

Workload

125 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation

▪Read the Syllabus
▪Geographic Information Systems and Science, 2nd Edition (2005) Paul A. Longley, Michael F. Goodchild, David J. Maguire, David W. Rhind. Wiley, 560 p. ISBN 0470721448
▪Python Data Science Handbook, Jake VanderPlas, 2016 - https://jakevdp.github.io/PythonDataScienceHandbook/</td></tr>
<tr><td colspan="6">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.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes</td></tr>
</table>

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

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

Indicative Literature

The course is based on a self-contained, detailed set of online lecture notes.

Nevertheless, the following provides a good overview of the material covered:

P. A. Longley, M. F. Goodchild, D. J. Maguire, D. W. Rhind, Geographic Information Systems and Science, 2nd Edition, Wiley, 2005, 560 p. ISBN 0470721448.

Jake VanderPlas, Python Data Science Handbook, 2016, https://jakevdp.github.io/PythonDataScienceHandbook/.

Usability and Relationship to other Modules

- This module is a natural companion to the "Principles of Statistical Modeling" (PSM) module MDE-CS-03.
- The ML module focuses on practical ML skills, whereas PSM module on rigorous mathematical formalism and analysis.
- For students not familiar with graph theory, it is recommended to take the first semester course MDE-CS-01 Network Theory, which introduces concepts used in this Machine Learning module.

Examination Type: Module Examination

Assessment Type: Term Paper — Length: 20 pages, Weight: 100%

Scope: All intended learning outcomes of this module.

3.2.2.2 Geoinformatics Lab

<table>
<tr><td colspan="3">Module Name
Geoinformatics Lab</td><td>Module Code
MDE-GEO-02</td><td>Level (type)
Year 1 (Elective)</td><td>CP
5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-GEO-02 | Geoinformatics Lab | Lecture | 5</td></tr>
<tr><td>Module Coordinator

Prof. Dr. Vikram Unnithan</td><td colspan="3">Program Affiliation

▪ MSc Data Engineering
▪ BSc Earth & Environmental Science</td><td colspan="2">Mandatory Status

Mandatory elective for DE and DSSB</td></tr>
<tr><td colspan="3" rowspan="2">Entry Requirements

Pre-requisites
☒ None

Co-requisites
▪ Geoinformatics

Knowledge, Abilities, or Skills
▪ Basic computer skills</td><td>Frequency

Annually (Spring)</td><td colspan="2">Forms of Learning and Teaching

▪ Lecture attendance (40 hours)
▪ Practical assignments (40 hours)
▪ Private study (45 hours)</td></tr>
<tr><td>Duration

1 semester</td><td colspan="2">Workload

125 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation

▪Read the Syllabus.
▪Python Data Science Handbook, Jake VanderPlas, 2016 - https://jakevdp.github.io/PythonDataScienceHandbook/
▪Geospatial Data and Analysis, Bill Day, Jon Bruner, Aurelia Moser, 2017, O'Reilly Media, Inc. ISBN: 9781491984314</td></tr>
<tr><td colspan="6">Content and Educational Aims

This lab module provides the necessary hands-on skills and expertise needed to gather, analyse, and model geospatial and /or temporal data. Integration, analysis, management and visualization of large volumes of spatial data from multiple sources at a variety of scales form a part of the assignments and lab work. Students may also have to design, integrate and implement a variety of sensors to gather, process, visualize and analyze environmental, oceanographic or other geo data. Theoretical concepts are demonstrated, and practical training provided using state of-the-art software and hardware. Examples of applications to various fields such as geo-and bio-sciences, data management, habitat management, risk assessment and geo-marketing are discussed and the role of the internet in data mining and Web GIS illustrated.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

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

▪ design, implement and exploit elementary supervised ML methods for classification and regression with expert care given to dimension reduction preprocessing and regularization;
▪ understand and practically use PCA and linear regression;
▪ understand the core ideas behind feedforward neural networks and the backpropagation algorithm, as the basis for accessing "deep learning" methods.</td></tr>
<tr><td colspan="6">Indicative Literature

J. VanderPlas, Python Data Science Handbook, 2016, https://jakevdp.github.io/PythonDataScienceHandbook/

B. Day, J. Bruner, A. Moser, Geospatial Data and Analysis, O'Reilly Media, 2017, ISBN: 9781491984314</td></tr>
</table>

Usability and Relationship to other Modules

- MDE-GEO-01 ideally a pre-requisite but due to schedule constraints it is co-requisite
- Uses and builds on concepts from all CORE modules, in particular MDE-CO-01, MDE-CO-02, MDE-CO-05 and MDE-CO-06

Examination Type: Module Examination

Assessment Type: Term Paper

Duration: 20 pages
Weight: 100%

Scope: All intended learning outcomes of this module.

3.2.3 Bio-Informatics Track

3.2.3.1 Introduction to Systems Biology

<table>
<tr><td colspan="3">Module Name
Introduction to Systems Biology</td><td>Module Code
MDE-BIO-02</td><td>Level (type)
Year 1/2 (Elective)</td><td>CP
5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-BIO-02 | Introduction to Systems Biology | Lecture | 5</td></tr>
<tr><td colspan="2">Module Coordinator

Prof. Dr. Marc-Thorsten Hütt</td><td colspan="2">Program Affiliation

• MSc Data Engineering</td><td colspan="2">Mandatory Status

Mandatory Elective for DE</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites | Co-requisites | Knowledge, Abilities, or Skills
☒ None | ☒ None | ☒ None</td><td>Frequency

Annually (Spring)</td><td colspan="2">Forms of Learning and Teaching

• Lectures (35 hours)
• Private Study (80 hours)
• Exam and preparation (10 hours)</td></tr>
<tr><td colspan="3"></td><td>Duration

1 semester</td><td colspan="2">Workload

125 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation

Read the syllabus.</td></tr>
<tr><td colspan="6">Content and Educational Aims

Systems Biology aims to understand the functioning of a cell due to the concerted action of its constituents. At the same time, many spatial and temporal scales contribute to cellular organization, which render it a complex interplay of regulatory processes. It seems, therefore, futile to address this problem of system understanding without the appropriate toolbox. This module provides the mathematical and conceptual toolbox for “doing Systems Biology”.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

Upon completion of this module, students will be able to:
• describe the key goals and methods of Systems Biology;
• analyze metabolic fluxes;
• recognize and apply models of signal transduction pathways;
• analyze gene regulatory systems;
• analyze gene expression patterns.</td></tr>
<tr><td colspan="6">Indicative Literature

E. Klipp, R. Herwig, A. Kowald, C. Wierling and H. Lehrach Systems Biology in Practice: Concepts, Implementation and Application, Wiley-VCH, 2005.

U.Alon, An Introduction to Systems Biology: Design Principles of Biological Circuits. Chapman & Hall/CRC, 2006.

B. O. Palsson, Systems Biology – Properties of reconstructed networks, Cambridge University Press, 2006.</td></tr>
</table>

Usability and Relationship to other Modules
N.A.
Examination Type: Module Examination
Assessment Type: Written Exam Duration: 120 minutes Weight: 100% Scope: All intended learning outcomes of this module.

3.2.3.2 Modeling and Analysis of Complex Systems

<table>
<tr><td colspan="3">Module Name
Modeling and Analysis of Complex Systems</td><td>Module Code
MDE-BIO-01</td><td>Level (type)
Year 1/2 (Elective)</td><td>CP
5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-BIO-01 | Modeling and Analysis of Complex Systems | Lecture | 5</td></tr>
<tr><td>Module Coordinator

Prof. Dr. Agostino Merico</td><td colspan="3">Program Affiliation

▪ MSc Data Engineering
▪ BSc Earth & Environmental Science</td><td colspan="2">Mandatory Status

Mandatory Elective for DE and DSSB</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites
☒ None

Co-requisites
☒ None

Knowledge, Abilities, or Skills
▪ Analysis, Basic Calculus, and Linear Algebra</td><td>Frequency

Annually (Fall or Spring)

Duration

1 semester</td><td colspan="2">Forms of Learning and Teaching

▪ Lecture attendance (35 hours)
▪ Practical exercises, private study incl. exam preparation (90 hours)

Workload

125 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation

Read the Syllabus.</td></tr>
<tr><td colspan="6">Content and Educational Aims

This is a hands-on module on the mathematical and computational modeling of various complex systems, covering diverse fields of the natural and social sciences. The module starts with an introduction to mathematical modeling. The elements of a model are presented and the steps to follow when constructing a model are reviewed, from formulating the question, determining the basic constituents of a model, and qualitatively and quantitatively describing the relevant system to analyzing the equations with various checks and balances. An introduction are provided on Python, the programming language constituting the main computational tool adopted in the module. To put into practice the theory on the basics of modelling and Python programming, a number of classical models in ecology are reviewed, coded, and numerically analyzed. This will build up the skills for developing models that describe different complex systems and the associated processes. In particular, different ial equation models are developed. They describe:

(1) the dynamics of diseases such as HIV, (2) the microbial growth in batch and chemostat cultures, (3) the dynamics of plankton ecosystems in the oceanic mixed layer, and (4) examples of life acting as a regulating force at a planetary scale. In addition, the lecturer introduces Agent-Based Modelling techniques with applications to cultural segregation problems and spatially explicit predator-prey interactions.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

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

▪ independently design and develop models (from the basic conceptual aspects, to the mathematical equations and the numerical code) for tackling problems in the natural and social sciences
▪ undertake numerical equilibria and stability analysis, to evaluate model performance, and to identify uncertainties in model results.</td></tr>
<tr><td colspan="6">Indicative Literature

The course is based on a self-contained, detailed set of online lecture notes and practical exercises.</td></tr>
</table>

Usability and Relationship to other Modules
N.A.
Examination Type: Module Examination
Assessment Type: Written Exam Duration: 120 minutes Weight: 100% Scope: All intended learning outcomes of this module.

Module Name Management and Analysis of Biological and Medical Data			***Module Code*** MDE-BIO-03	***Level (type)*** Year 1/2 (Elective)	***CP*** 5
Module Components					
Number	*Name*			*Type*	*CP*
MDE-BIO-03	Management and Analysis of Biological and Medical Data			Seminar	5
Module Coordinator Prof. Dr. Marc-Thorsten Hütt	***Program Affiliation*** • MSc Data Engineering (DE)			***Mandatory Status*** Mandatory elective for DE	
Entry Requirements *Pre-requisites* ☒ none	*Co-requisites* ☒ none	*Knowledge, Abilities, or Skills*	***Frequency*** Annually (Spring/Fall)	***Forms of Learning and Teaching*** • Lectures and plenary discussions (20 hours) • Project work (105 hours)	
			Duration 1 semester	***Workload*** 125 hours	

Recommendations for Preparation

Content and Educational Aims

High-throughput technologies have turned biological and medical research into 'big data' endeavors. It is indispensable to be able to navigate the rich, intricate landscape of biological and medical databases and to contextualize and analyze information from diverse sources. Here we explore examples of recent databases in Biology and Medicine selected from the current database issue of the journal *Nucleic Acids Research*. Typical examples include databases for gene expression, metabolic systems, genome-wide association studies and epigenetic information.

In the first sessions of the course, we define small research projects based on the selected databases. In the rest of the course these research projects will be pursued in small groups and the results will be reported and discussed.

Intended Learning Outcomes

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

1. identify and process a variety of data formats and data standards in biology and medicine
2. access and use the main bioinformatics databases
3. download and analyze diverse biological and medical data
4. derive research questions from scientific publications
5. apply concepts from data science to biological and medical databases.

Indicative Literature

Usability and Relationship to other Modules

Examination Type: Module Examination

Assessment Type: Oral Exam — Duration/Length: — Weight: 100%

Scope: All intended learning outcomes of this module.

3.2.4 Business and Supply Chain Engineering Track

3.2.4.1 Data Mining

<table>
<tr><td colspan="3">Module Name
Data Mining</td><td>Module Code
MDE-BSC-01</td><td>Level (type)
Year 1 (CORE)</td><td>CP
5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-BSC-01 | Data Mining | Lecture | 5</td></tr>
<tr><td>Module Coordinator

Prof. Dr. Adalbert F.X. Wilhelm</td><td colspan="3">Program Affiliation

▪ MSc Data Engineering</td><td colspan="2">Mandatory Status

Mandatory for DE and DSSB</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites
• Data Analytics

Co-requisites
• Machine Learning

Knowledge, Abilities, or Skills
▪ Knowledge of Data Analytics software/ programming languages such as R or Python</td><td>Frequency

Annually (Spring)

Duration

1 semester</td><td colspan="2">Forms of Learning and Teaching

▪ Lecture (17.5 hours)
▪ Project work (90 hours)
▪ Private study (17.5 hours)

Workload

125 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation

Practice data analysis tasks. Read the Syllabus.</td></tr>
<tr><td colspan="6">Content and Educational Aims

The focus of this module is on practical applications of algorithms and computational paradigms that allow computer-based search and detection of data patterns and regularities. Students learn how to use such tools to perform predictions and make forecasts. Students will study data mining as the core component in the knowledge discovery in database process which deals with extracting useful information from raw data. This knowledge discovery process includes data selection, cleaning, coding, using different statistical and machine learning techniques, and visualization of data and generated patterns and structures. The module aims tot provide an overview of all these issues and illustrates the whole process by examples.
A major component of the module is group-based participation in a data analysis competition. This competition allows students to apply the concepts learned in class and to develop the computational skills to analyze data in a collaborative setting.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

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

▪ be able to implement and apply advanced data mining methods with appropriate tools
▪ be able to evaluate and compare the suitability, scalability and efficiency of different methods in practical settings
▪ have gained experience in performing a full cycle of data mining and data analysis
▪ have acquired practical skills to tackle data mining problems</td></tr>
<tr><td colspan="6">Indicative Literature

G. James, D. Witten, T. Hastie, R. Tibshiran, Introduction to Statistical Learning with R by Springer, 2013 (ISLR).
J. VanderPlas, Python Data Science Handbook, 2016 - https://jakevdp.github.io/PythonDataScienceHandbook/.</td></tr>
<tr><td colspan="6">Usability and Relationship to other Modules

This module builds on the core module data analytics MDE-CO-02 and prepares students for applied projects in data analysis as well as a master thesis in this field.</td></tr>
</table>

Examination Type: Module Examination

Assessment Type: Term Paper (Project Report)	Length: 20 pages Weight: 100%

Scope: All intended learning outcomes of the module.

3.2.4.2 Data Analytics in Supply Chain Management

<table>
<tr><td colspan="3">Module Name
Data Analytics in Supply Chain Management</td><td>Module Code
MSCM-CO-07</td><td>Level (type)
Year 2 (CORE)</td><td>CP
5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MSCM-CO-07 | Data Analytics in Supply Chain Management | Lecture | 5</td></tr>
<tr><td colspan="2">Module Coordinator

Prof. Dr.-Ing. Hendro Wicaksono</td><td colspan="2">Program Affiliation

• MSc Supply Chain Management</td><td colspan="2">Mandatory Status

Mandatory elective for SCM and DE</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites
• MSCM-MET-01 – Programming in Python

OR

• MDE-MET-03 Intro to Data Management with Python

Co-requisites
☒None

Knowledge, Abilities, or Skills
• Basics of statistical analytics and machine learning
• Basics of database and SQL
• Basics of programming skills, such as R, Python, and Java</td><td>Frequency

Annually (Fall)

Duration

1 semester</td><td colspan="2">Forms of Learning and Teaching

• Lecture and feedback sessions (35 hours)
• Group Work (45 hours)
• Private Study (45 hours)

Workload

125 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation

Sanders, N. Big data driven supply chain management: a framework for implementing analytics and turning information into intelligence, Pearson Education, 2014.</td></tr>
<tr><td colspan="6">Content and Educational Aims

In recent years, big data has become a significant topic in supply chain management, as the amount of data generated in supply chain management practices has grown exponentially. Data analytics are techniques that apply data mining, statistical analysis, predictive analytics, and machine learning to uncover hidden patterns, correlations, trends, and other business-valuable information and knowledge from data.

The module focuses on the supply chain management scenarios that generate and consume data intensively and require data analytics to improve the decision-making process through descriptive, predictive, and prescriptive analytics. These include:

• Descriptive statistics on and historical insight into companies' production, financial, operations, sales, customers, etc.
• Forecasting customer behavior, purchasing patterns, production performance, energy consumption, etc.
• Prescriptive analytics for assessing the offer that should be made to a certain customer, to decide on the shipment strategy for each location, to determine the most efficient material flow in a factory, etc.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

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

• identify scenarios in supply chain management and evaluate the opportunities and challenges of data analytics applications;
• determine the objective of data analytics in different scenarios and the data sources required to achieve that objective;</td></tr>
</table>

- apply methods and tools to collect and integrate data from different sources in the context of supply chain management;
- apply machine learning and statistical analytics methods and tools to uncover hidden patterns, correlations, trends, and knowledge that are useful for improving supply chain management processes;
- evaluate data analytics results in different scenarios and solve the problems that might occur throughout the entire data analytics process, from data collection to analysis;
- develop deployment architecture concepts by integrating existing tools/software;
- develop business model and ecosystem concepts.

Indicative Literature

N.A.

Usability and Relationship to other Modules

Programming methods, such as R and Python, taught in MSCM-MET-01 Programming in Python and MSCM-MET-03 Programming in R as well as project management concepts taught in MSCM-CO-01 will be applied. Academic writing skills taught in MSCM-CAR-01 facilitate the completion of tasks in this module.

Examination Type: Module Examination

Assessment Type: Project Report

Length: 2.500 words
Weight: 100%

Scope: All intended learning outcomes of the module.

3.4 Methods Area (15 CP)

3.4.1 Introduction to Data Management with Python

<table>
<tr><td colspan="2">Module Name

Introduction to Data Management with Python</td><td>Module Code

MDE-MET-03</td><td>Level (type)

Year 1 (Methods)</td><td>CP

5</td></tr>
<tr><td colspan="5">Module Components

Number | Name | Type | CP
MDE-MET-03-A | Introduction to Data Management with Python - Lecture | Lecture | 2.5
MDE-MET-03-B | Introduction to Data Management with Python -Tutorial | Tutorial | 2.5</td></tr>
<tr><td>Module Coordinator

Dr. Carlos Brandt</td><td colspan="2">Program Affiliation

▪ MSc Data Engineering</td><td colspan="2">Mandatory Status

Mandatory for DE</td></tr>
<tr><td colspan="2">Entry Requirements

Pre-requisites: ☒ None
Co-requisites: ☒ None
Knowledge, Abilities, or Skills: ☒ None</td><td>Frequency

Annually (Fall)</td><td colspan="2">Forms of Learning and Teaching

▪ Lecture attendance (17.5hours)
▪ Tutorial attendance (17.5 hours)
▪ Private Study (90 hours)</td></tr>
<tr><td colspan="2"></td><td>Duration

1 semester</td><td colspan="2">Workload

125 hours</td></tr>
<tr><td colspan="5">Recommendations for Preparation

None.</td></tr>
<tr><td colspan="5">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.</td></tr>
<tr><td colspan="5">Intended Learning Outcomes

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

• explain and apply fundamental concepts of imperative programming using Python
• understand and use basic data structures
• summarize and apply fundamental algorithms (e.g. sorting)
• execute basic data analysis tasks (average, min, max, ...)
• Understand and implement linear algebra operations using NumPy/SciPy
• explain fundamentals of relational databases
• describe and use SQL to create, modify and query data from relational databases</td></tr>
</table>

- understand and apply DataFrames and data analysis using Pandas
- 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 to courses like “Advanced Databases” or “Machine Learning”.

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: Tutorial

Assessment Type: Practical Assessment (Programming Assignments) Weight: 50%

Scope: All intended learning outcomes of this module.

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

3.4.2 Modeling and Control of Dynamical Systems

<table>
<tr><td colspan="3">Module Name
Modeling and Control of Dynamical Systems</td><td>Module Code
MDE-MET-04</td><td>Level (type)
Year 1/2 (Methods)</td><td>CP
5.0</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-MET-04 | Modeling and Control of Dynamical Systems | Seminar | 5.0</td></tr>
<tr><td>Module Coordinator

Dr. Mathias Bode</td><td colspan="3">Program Affiliation

• MSc Data Engineering</td><td colspan="2">Mandatory Status

Mandatory Elective for DE</td></tr>
<tr><td colspan="2">Entry Requirements

Pre-requisites

☒ None</td><td>Co-requisites

☒ None</td><td>Knowledge, Abilities, or Skills

• Basic linear algebra, calculus, probability concepts and programming skills as taught in introductory modules.</td><td>Frequency

Annually (Spring)

Duration

1 semester</td><td>Forms of Learning and Teaching

• Lectures (35 hours)
• Private Study (90 hours)

Workload

125 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation

Read the book: "Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering" by Steven H. Strogatz, in particular parts I+II.In order to prepare, please, read chapters 1,2+5.</td></tr>
<tr><td colspan="6">Content and Educational Aims

Predictions based on the past, with or without additional input information? This is the topic of our module on dynamical systems. In many cases these forecasts are (almost) exact; in others we can only get probabilistic information. Based on a large set of examples, the module is going to discuss these so-called deterministic and stochastic systems. Topics we cover include:

• Deterministic low-dimensional dynamical systems.
• Control of deterministic linear systems.
• Linear prediction of stochastic time series.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

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

• understand and apply fundamental concepts of deterministic and stochastic dynamical systems,
• solve linear ordinary differential equations with constant coefficients,
• understand and apply fundamental concepts from linear control theory
• understand and apply (conditional) means, variances, and covariances in order to predict the behavior of simple stochastic systems.</td></tr>
<tr><td colspan="6">Indicative Literature

S. H. Strogatz, Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering, 2nd edition, 2015.

S. Zak, Systems and Control, Oxford University Press, 2003.

H. Stark & J. Woods, Probability and Random Processes with Applications to Signal Processing, Westview Press, 2002.</td></tr>
<tr><td colspan="6">Usability and Relationship to other Modules

Complementary to the machine Learning module MDE-CO-04 this module focuses on a theory-based design of models. Such models, if available, are usually “smaller” and easier to parameterize.</td></tr>
</table>

Examination Type: Module Examination

Assessment Type: Written Exam

Duration: 120 minutes
Weight: 100%

Scope: All intended learning outcomes of this module.

3.4.3 Modern Signal Processing

Module Name	*Module Code*	*Level (type)*	*CP*
Modern Signal Processing	MDE-MET-07	Year 1/2 (Methods)	5.0

Module Components

Number	*Name*	*Type*	*CP*
MDE-MET-07	Modern Signal Processing	Seminar	5.0

Module Coordinator	*Program Affiliation*	*Mandatory Status*
Prof. Dr. Giuseppe Abreu	• MSc Data Engineering	Mandatory Elective for DE

Entry Requirements			*Frequency*	*Forms of Learning and Teaching*
Pre-requisites	*Co-requisites*	*Knowledge, Abilities, or Skills*	Bi-annually (Fall)	• Lectures (35 hours) • Private Study (90 hours)
☒ None	☒ None	☒ None		
			Duration	*Workload*
			1 semester	125 hours

Recommendations for Preparation

Read the Syllabus.

Content and Educational Aims

This module aims to introduce students to a modern perspective of signal processing, which in the recent years has seen significant changes due to the emergence of new mathematical and algorithmic tools. At the core of this new perspective is the departure from canonical compact orthonormal representations (of which Fourier analysis is the primary example) and minimalistic sampling (of which the Nyquist rate is the primary example) towards sparse, non-orthogonal signal representations, typically resulting from oversampling and the use of redundant bases. Another major aspect in which modern signal processing differs from its classical counterpart is the significantly larger role played by numerical methods. Indeed, traditional signal processing was developed during an era when computers were either non-existent or incipient, thus relying fundamentally on tools such as algebraic geometry and harmonic analysis, and consequently typically leading to techniques that yield exact results and even closed-form solutions under ideal conditions (e.g., in the absence of distortion), which are then brought to practical applications by means of suitable algorithmic approximations. By contrast, modern signal processing embraces numerical methods and 'algorithms' at its core, thus relying heavily on tools such as convex optimization, non-convex optimization (e.g. genetic algorithms) and machine learning which, albeit not necessarily empirical, take full advantage of the computational power of modern computers.

This module explores several of the aforementioned tools, discussing concrete examples such as isotropic embedding (which can be used for wireless localization), matrix completion (which can be used for data compression), and the design of tight frames (which can be used to increase robustness to distortion).

Intended Learning Outcomes

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

- understand the fundamental principles behind modern signal processing algorithms;
- gain a new perspective of signal processing problems through the prism of new algorithms in which signals are treated as data;
- practice on how to address both new and "old'" signal processing problems via the new tools of modern signal processing.

- further develop their Matlab programming skills (or an equivalent programming language with sufficient support of for mathematical libraries);
- gain a deeper and a modern understanding of crucial mathematical tools such as linear algebra (vectors and matrices) and functional analysis (Hilbert spaces, inner products, basic calculus), in the context of their application to data engineering.

Indicative Literature

P. Walk and P. Jung, Compressed Sensing: Applications to Communication and Digital Signal Processing, Springer, 2019.

S. Oh, Matrix Completion: Fundamental Limits and Efficient Algorithms, Stanford University, 2010.

J. Dattorro, Convex Optimization and Euclidean Distance Geometry, Meboo Publishing, 2008.

I. Rish, G. Grabarnik, Sparse Modeling: Theory, Algorithms, and Applications, CRC Press, 2014.

S. S. Foucart and H. Rauhut, A Mathematical Introduction to Compressive, Birkhäuser, 2013.

Usability and Relationship to other Modules

-

Examination Type: Module Examination

Assessment Type: Oral Presentation — Duration: 30 minutes, Weight: 100%

Scope: All intended learning outcomes of this module.

3.4.4 Network Approaches in Biology and Medicine

<table>
<tr><td colspan="2">Module Name

Network Approaches in Biology and Medicine</td><td>Module Code

MDE-MET-05</td><td>Level (type)

Year 1/2 (Elective)</td><td>CP

5</td></tr>
<tr><td colspan="5">Module Components

Number | Name | Type | CP
MDE-MET-05 | Network Approaches in Biology and Medicine | Lecture | 5</td></tr>
<tr><td>Module Coordinator

Prof. Dr. Marc-Thorsten Hütt</td><td colspan="2">Program Affiliation

▪ MSc Data Engineering</td><td colspan="2">Mandatory Status

Mandatory elective for DE and DSSB</td></tr>
<tr><td colspan="2">Entry Requirements

Pre-requisites
☒ None

Co-requisites
• Geoinformatics

Knowledge, Abilities, or Skills
▪ Analysis, Basic Calculus, and Linear Algebra</td><td>Frequency

Annually (Spring)

Duration

1 semester</td><td colspan="2">Forms of Learning and Teaching

▪ Lecture attendance (35 hours)
▪ Private study (90 hours)

Workload

125 hours</td></tr>
<tr><td colspan="5">Recommendations for Preparation

Read the Syllabus.</td></tr>
<tr><td colspan="5">Content and Educational Aims

'Network science' employs the formal view of graph theory to understand the design principles of complex systems. Abstracting cellular processes in from biology into networks can contribute to an understanding of how such cellular systems function. Over the last two decades, this approach has revolutionized the way we think about biological systems.

Here, the application of network analysis to biology and medicine are discussed. In this module standard networks considered in Systems Biology (gene regulatory networks, metabolic networks, signaling networks and protein-protein interaction networks), in which each link corresponds to a specific biological process are discussed. It is enhanced by the discussion of relational networks, which are capable of serving as very efficient sources of data integration and interpretation: the diseasome, a network where a disease is linked to a gene, in which there is data evidence relating the gene to the disease; and the drug-target network, where drugs and proteins linked by drug-target associations.
In addition to standard review articles and textbooks on Network Science, material from recent scientific literature is incorporated in the module.</td></tr>
<tr><td colspan="5">Intended Learning Outcomes

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

▪ understand the basic principles of network science applications to Biology and Medicine;
▪ use and access the main bioinformatics databases to obtain biological networks;
▪ analyze biological networks;
▪ combine multiple data analysis tools for a comprehensive analysis of molecular data;
▪ describe in some detail essential facts and theoretical concepts derived from recent scientific literature;
▪ identify open questions from the scientific literature and synthesize information from the literature into a scientific presentation.</td></tr>
<tr><td colspan="5">Indicative Literature

A.-L. Barabási, Network science. Cambridge University Press, 2016.</td></tr>
</table>

Alon, U. (2007). Network motifs: theory and experimental approaches. Nature Reviews Genetics, 8(6):450–461.
A.-L. Barabási (2012), The network takeover. Nature Physics, 8(1):14–16.

A.-L. Barabási, N. Gulbahce and Loscalzo (2011). Network medicine: a network-based approach to human disease. Nature reviews. Genetics, 12(1):56.

Barabasi, A.-L. and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. Nature reviews. Genetics, 5(2):101.

Radde, N. E. and Hütt, M.-T. (2016). The physics behind systems biology. EPJ Nonlinear Biomedical Physics, 4(1):7.
Strogatz, S. H. (2001). Exploring complex networks. Nature, 410(6825):268.

and recent scientific literature.

Usability and Relationship to other Modules

This module is recommended to be taken together with the elective modules in the Bio-Informatics track.

Examination Type: Module Examination

Assessment Type: Oral Presentation — Duration: 30 minutes
Weight: 100%

Scope: All intended learning outcomes of this module.

3.4.5 Applied Dynamical Systems

<table>
<tr><td colspan="3">Module Name
Applied Dynamical Systems</td><td>Module Code
MDE-MET-06</td><td>Level (type)
Year 1/2 (Methods)</td><td>CP
5.0</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-MET-06 | Applied Dynamical Systems | Lecture | 5.0</td></tr>
<tr><td colspan="2">Module Coordinator

Prof. Dr. Marcel Oliver</td><td colspan="2">Program Affiliation

• MSc Data Engineering
• BSc Mathematics</td><td colspan="2">Mandatory Status

Mandatory Elective for DE</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites
☒ None

Co-requisites
☒ None

Knowledge, Abilities, or Skills
• Analysis, basic Calculus and Linear Algebra</td><td>Frequency

Annually (Spring)

Duration

1 semester</td><td colspan="2">Forms of Learning and Teaching

• Lectures (35 hours)
• Private Study (90 hours)

Workload

125 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation

Read the Syllabus.</td></tr>
<tr><td colspan="6">Content and Educational Aims

This module is a first hands-on introduction to theory and applications of dynamical systems. A crucial component of this class is the use of computer experiments to foster intuitive understanding and develop students' skills in using the computer to bridge the gap between mathematical idea and concrete implementation and application.

Topics include nonlinear oscillators, coupled pendula, and pattern formation in chemical reactions. A main focus of the lab is the development of standard tools for the numerical solution of differential equations, the application of automated tools for bifurcation analysis, and continuation methods. Further topics include agent-based models and pseudo-spectral PDE solvers for reaction-diffusion equations.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

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

• apply fundamental concepts of deterministic and stochastic modeling;
• implement standard mathematical software;
• design, conduct, and interpret controlled in-silico scientific experiments;
• demonstrate the mastery of numerical methods to solve differential equations.</td></tr>
<tr><td colspan="6">Indicative Literature

J. Sethna, Statistical Mechanics: Entropy, Order Parameters, and Complexity, Oxford University Press, 2006.

Steven Strogatz, Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering, Westview Press, second edition, 2014.</td></tr>
<tr><td colspan="6">Usability and Relationship to other Modules

This module is complementary to the module MDE-MET-04 Modeling and Control of Dynamical Systems.</td></tr>
<tr><td colspan="6">Examination Type: Module Examination

Assessment Type: Term Paper (Project Portfolio)
Length: 20 pages
Weight: 100%

Scope: All intended learning outcomes of this module.</td></tr>
</table>

3.4.6 Remedial Modules

3.4.6.1 Calculus and Linear Algebra for Graduate Students

<table>
<tr><td colspan="3">Module Name
Calculus and Linear Algebra for Graduate Students</td><td>Module Code
MDE-MET-01</td><td>Level (type)
Year 1 (Methods)</td><td>CP
5.0</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-MET-01 | Calculus and Linear Algebra for Graduate Students | Lecture | 5.0</td></tr>
<tr><td colspan="2">Module Coordinator

Prof. Dr. Igors Gorbovickis</td><td colspan="2">Program Affiliation

• MSc Data Engineering</td><td colspan="2">Mandatory Status

Mandatory Elective for DE</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites

☒ None

Co-requisites

☒ None

Knowledge, Abilities, or Skills

• Mathematics at High School level</td><td>Frequency

Annually (Fall)

Duration

1 semester</td><td colspan="2">Forms of Learning and Teaching

• Lectures (35 hours)
• Private Study (90 hours)

Workload

125 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation

Read the Syllabus.</td></tr>
<tr><td colspan="6">Content and Educational Aims

This module offers a highly structured introduction to the fundamentals of two major pillars of mathematical modelling and analysis: Single and multivariable calculus on the one hand and linear algebra on the other.

It is a gateway for graduate students who have not been exposed to the topics so far, or who were exposed long ago and needs to be refreshed.

Topics include sequences, series, limits, derivatives, Taylor series, and integrals as well as vectors, matrices, determinants, eigenvalues, eigenvectors, scalar products, and norms. The module focuses on practical experience rather than on mathematical rigor.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

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

• apply the fundamental concepts of calculus and linear algebra in structured situations;
• understand and use vectors and matrices, calculate determinants, eigenvalues and eigenvectors in simple cases;
• calculate derivatives and simple integrals;
• explain the importance of the methods of calculus and linear algebra in problems arising from applications;
• understand the methods of calculus and linear algebra used in more advanced modules as well as in scientific literature.</td></tr>
<tr><td colspan="6">Indicative Literature

G. Strang, Introduction to Linear Algebra, 5th edtion, Wellesley-Cambridge Press, 2016, ISBN: 978-09802327-7-6.</td></tr>
<tr><td colspan="6">Usability and Relationship to other Modules</td></tr>
</table>

This module introduces and refreshes the essential Calculus and Linear Algebra required in most of the modules of the data engineering program. There is a placement test offered in the orientation week before the start of the first semester to help all students to find out if they need to take this remedial course.

Examination Type: Module Examination

Assessment Type: Written Exam

Duration: 120 minutes
Weight: 100%

Scope: All intended learning outcomes of this module.

3.4.6.2 Probabilities for Graduate Students

<table>
<tr><td colspan="3">Module Name
Probabilities for Graduate Students</td><td>Module Code
MDE-MET-02</td><td>Level (type)
Year 1 (Methods)</td><td>CP
5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-MET-02 | Probabilities for Graduate Students | Lecture | 5</td></tr>
<tr><td colspan="2">Module Coordinator

Dr. Mathias Bode</td><td colspan="2">Program Affiliation

• MSc Data Engineering</td><td colspan="2">Mandatory Status

Mandatory Elective for DE</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites

☒ None

Co-requisites

☒ None

Knowledge, Abilities, or Skills

☒ None</td><td>Frequency

Annually (Fall)

Duration

1 semester</td><td colspan="2">Forms of Learning and Teaching

• Lectures (35 hours)
• Private Study (90 hours)

Workload

125 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation

Read the Syllabus.</td></tr>
<tr><td colspan="6">Content and Educational Aims

This module offers a highly structured introduction to 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 long ago and needs to be refreshed. The module starts 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. Block three continues with the essential ideas of expected values, moments, and estimation.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

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

• understand the fundamental concepts of probabilities and combinatorics and to apply them in structured situations,
• apply important probability laws (Binomial, Poisson, Normal),
• understand and apply probability distributions and densities,
• understand and apply means, variances, and covariances – also in the context of simple estimation contexts.</td></tr>
<tr><td colspan="6">Indicative Literature

H. Stark, J. W. Woods, Probability and Random Processes with Applications to Signal Processing, Third Edition, 2002.</td></tr>
<tr><td colspan="6">Usability and Relationship to other Modules

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. There is a placement test offered in the orientation week before the start of the first semester to help all students to find out if they need to take this remedial course.</td></tr>
</table>

Examination Type: Module Examination

Assessment Type: Written Exam

Duration: 120 minutes
Weight: 100%

Scope: All intended learning outcomes of this module.

3.5 Discovery Area (15 CP)

3.5.1 Current Topics in Data Engineering

<table>
<tr><td colspan="3">Module Name
Current Topics in Data Engineering</td><td>Module Code
MDE-DIS-01</td><td>Level (type)
Year 1 (Discovery)</td><td>CP
5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-DIS-01 | Current Topics in Data Engineering | Colloquium | 5</td></tr>
<tr><td>Module Coordinator

Prof. Dr. Stefan Kettemann</td><td colspan="3">Program Affiliation

• MSc Data Engineering</td><td colspan="2">Mandatory Status

Mandatory for DE</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites | Co-requisites | Knowledge, Abilities, or Skills
☒ None | ☒ None | ☒ None</td><td>Frequency

Annually (Fall)

Duration

1 semester</td><td colspan="2">Forms of Learning and Teaching

• Colloquium (17.5 hours)
• Private Study (107.5 hours)

Workload

125 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation

Read the Syllabus.</td></tr>
<tr><td colspan="6">Content and Educational Aims

This module introduces current topics and challenges of data engineering. Lectures are taught by faculty members and invited experts from companies, presenting selected fields of their research activities and interest in data engineering. For each field an overview of the scientific background, the motivation and major challenges is provided together with a list of references. This is complemented by an in-depth discussion of the specific research topics. Each student will then select one field of the faculty presentations and will prepare a term paper in the form of a master thesis proposal, which will be presented as a poster at the end of the module. The module will additionally feature tutorials providing the students with scientific skills.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

Upon completion of this module, students will be able to:
• describe a current topic in Data Engineering;
• research and read scientific literature;
• communicate in scientific language using field specific-technical terms.</td></tr>
<tr><td colspan="6">Indicative Literature

The literature is provided by each instructor of the current topics lecture in the slides, which are provided immediately after each lecture to all students by pdf on a teamwork space created by the instructor of record Prof. Kettemann.</td></tr>
<tr><td colspan="6">Usability and Relationship to other Modules

This module particularly prepares for the Advanced Project modules MDE-DIS-02 and MDE-DIS-03 and also gives the students an orientation with respect to which methods are required to master current developments in data engineering.</td></tr>
</table>

Examination Type: Module Examination

Assessment Type: Poster Presentation

Duration: 120 minutes
Weight: 100%

Scope: All intended learning outcomes of this module.

3.5.2 Advanced Project 1

Module Name	*Module Code*	*Level (type)*	*CP*
Advanced Project 1	MDE-DIS-02	Year 1 (Discovery)	5

Module Components

Number	*Name*	*Type*	*CP*
MDE-DIS-02	Advanced Project 1	Lecture and Seminar	5

Module Coordinator	*Program Affiliation*	*Mandatory Status*
Prof. Dr. Stefan Kettemann	• MSc Data Engineering	Mandatory for DE

Entry Requirements			*Frequency*	*Forms of Learning and Teaching*
Pre-requisites	*Co-requisites*	*Knowledge, Abilities, or Skills*	Annually (Spring)	• Lectures (17.5 hours) • Seminar (35 hours) • Private Study (72.5 hours)
☒ None	☒ None	☒ None	*Duration*	*Workload*
			1 semester	125 hours

Recommendations for Preparation

Read the Syllabus.

Content and Educational Aims

This module aims to provide the student with an in-depth understanding and command of one of the data analytics or data management techniques that are represented by the research groups of the faculty of DE. The subdiscipline involved (e.g. database management, machine learning, statistical data analysis, information theory, data acquisition, or big data technologies) changes from year to year and from hosting group to hosting group. The detailed structure and schedule depend on the specific demands and options of the hosting group.

An Advanced Project module typically begins with an introduction to the concerned technology or method. This leads the student to a level of competence with which he/she can insightfully apply the respective methods to practical, real-life tasks. Typically, the second half of the module is devoted to individual (or if indicated, group) projects in which a nontrivial data management/ analysis task, given by the instructor, is completed. The project outcome is a technical report (target size: 20 pages) together as well as with a presentation to the Data Engineering program students and faculty.

Intended Learning Outcomes

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

- understand current technical/scientific literature, and distinguish good from second-rate publications
- write / configure computer programs / tools specifically for the subject area
- master relevant data pre/ postprocessing routines specifically for the subject area
- design and schedule a complex DE project, including escape options, keep milestones/timelines
- consistently apply scientific language to communicate in writing his/her understanding clearly and precisely to a non-expert audience.

Indicative Literature

The literature is provided individually to each student by each instructor of the respective advanced project.

Usability and Relationship to other Modules

The students can choose a project, ideally on a topic and with a supervisor they already encountered during the 1st semester module MDE-DIS-01 Current Topics in Data Engineering.

Examination Type: Module Examination

Assessment Type: Term Paper (Project Report)	Duration: 20 pages Weight: 100%

Scope: All intended learning outcomes of this module.

3.5.3 Advanced Project 2

<table>
<tr><td colspan="3">Module Name
Advanced Project 2</td><td>Module Code
MDE-DIS-03</td><td>Level (type)
Year 2 (Discovery)</td><td>CP
5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-DIS-03 | Advanced Project 2 | Project Work | 5</td></tr>
<tr><td colspan="2">Module Coordinator

Prof. Dr. Stefan Kettemann</td><td colspan="2">Program Affiliation

• MSc Data Engineering</td><td colspan="2">Mandatory Status

Mandatory for DE</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites

☒ None

Co-requisites

☒ None

Knowledge, Abilities, or Skills

☒ None</td><td>Frequency

Annually (Fall)

Duration

1 semester</td><td colspan="2">Forms of Learning and Teaching

• Supervised Study, Research and Project Work (125 hours)

Workload

125 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation

Read the Syllabus.</td></tr>
<tr><td colspan="6">Content and Educational Aims

This module aims to providing the student with an in-depth understanding and command of one of the data analytics or data management techniques that are represented by the research groups of the faculty of DE. The subdiscipline involved (e.g. database management, machine learning, statistical data analysis, information theory, data acquisition, or big data technologies) changes from year to year and from hosting group to hosting group. The detailed structure and schedule depend on the specific demands and options of the hosting group. An Advanced Project module typically begins with a taught introduction to the concerned technology or method. This will lead the student to a level of competence with which he/she can insightfully apply the respective methods to practical, real-life tasks. Typically, the second half of the module is devoted to individual (or if indicated, group) projects in which a nontrivial data management/analysis task, given by the instructor, is worked out. The project outcome is a technical report (target size: 20 pages) together with a presentation to the Data Engineering program students and faculty.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

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

• understand current technical/scientific literature, and distinguish good from second-rate publications;
• write / configure computer programs / tools specifically for the subject area;
• master relevant data pre/postprocessing routines specifically for the subject area;
• design and schedule a complex DE project, including escape options, keep milestones/timelines;
• hone technical writing skills;
• communicate technical results to a non-expert audience.</td></tr>
<tr><td colspan="6">Indicative Literature

The literature is provided individually to each student by each instructor for the respective advanced project.</td></tr>
<tr><td colspan="6">Usability and Relationship to other Modules

The students can build on the project they worked on in the module MDE-DIS-02 Advanced Project 1. However, they are also free to choose another project topic with a different supervisor.</td></tr>
</table>

Examination Type: Module Examination

Assessment Type: Term Paper (Project Report)	Length: 20 pages Weight: 100%

Scope: All intended learning outcomes of this module.

3.6 Career Area (15 CP)

3.6.1 Language Skills

The descriptions of the language modules are provided in a separate document, the "Language Module Handbook" that can be accessed from here: https://www.jacobs-university.de/study/learning-languages

3.6.2 Academic Writing Skills/Intercultural Training

<table>
<tr><td colspan="3">Module Name
Academic Writing Skills/Intercultural Training</td><td>Module Code
MDE-CAR-02</td><td>Level (type)
Year 1 (CAREER)</td><td>CP
2.5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-CAR-02 | Academic Writing Skills/Intercultural Training | Seminar | 2.5</td></tr>
<tr><td colspan="2">Module Coordinator

Prof. Dr. Stefan Kettemann</td><td colspan="2">Program Affiliation

• MSc Data Engineering</td><td colspan="2">Mandatory Status

Mandatory for DE</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites: ☒ None
Co-requisites: ☒ None
Knowledge, Abilities, or Skills: ☒ None</td><td>Frequency

Annually (Spring)

Duration

1 semester</td><td colspan="2">Forms of Learning and Teaching

• Lectures (17.5 hours)
• Private Study (45 hours)

Workload

62.5 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation

Read the Syllabus.

Fraedrich, J. & Ferrell, O.C. (2014): Business Ethics: Ethical Decision Making & Cases. Cengage Learning.</td></tr>
<tr><td colspan="6">Content and Educational Aims

The academically rigorous nature of graduate studies requires students to master academic writing skills and techniques. In this introductory course, students in DE master's program will learn the foundations of academic writing at a graduate level, with special focus on writing academic essays, identifying organizational patterns of academic texts, and formulating arguments to produce cohesive and coherent academic papers. Through the process of drafting, continuous feedback and editing, students will improve their writing skills. This course will also help students develop their research skills by highlighting techniques of finding and evaluating sources, and utilizing citation and referencing styles. As graduate students, adhering to The Code of Academic Integrity is a requirement. Hence, this course will incorporate a session on scholarly and intellectual standards set by Jacobs University. The second part of this course is a training seminar. It will give answers to frequently asked questions by students on the topics of working and living in Germany. Here the students will find information on employment and how to get access to the German labor market. The seminar also provides an overview of labor conditions in Germany, the multifaceted forms of employment, business cultures and useful tips and information for the job entry in a German company.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

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

• structure their ideas to write clear summaries, coherent paragraphs and cohesive literature reviews;
• write different segments of an academic paper employing writing styles that display advanced grammar and precise and concise language use;
• successfully find and evaluate sources for research;
• use citation and referencing styles applicable for their discipline;
• unintentional plagiarism and adhere to the code of academic integrity.</td></tr>
</table>

- understand labor conditions in Germany.
- understand the typical business cultures in German companies.

Indicative Literature

The literature is provided individually to each student by each instructor for the respective advanced project.

Usability and Relationship to other Modules

Advanced Project 1, Advanced Project 2, Master thesis

Examination Type: Module Examination

Assessment Type: Term Paper (Report)

Length: 10 pages
Weight: 100%

Scope: All intended learning outcomes of this module.

3.6.3 Communication & Presentation Skills for Executives

<table>
<tr><td colspan="3">Module Name

Communication & Presentation Skills for Executives</td><td>Module Code

MDE-CAR-01</td><td>Level (type)

Year 2 (CAREER)</td><td>CP

2.5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDE-CAR-01 | Communication & Presentation Skills for Executives | Seminar | 2.5</td></tr>
<tr><td>Module Coordinator

Prof. Dr. Stefan Kettemann,
Prof. Dr. Hilke Brockmann</td><td colspan="3">Program Affiliation

▪ MSc Data Engineering</td><td colspan="2">Mandatory Status

Mandatory elective for DE and DSSB</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites: ☒ None
Co-requisites: ☒ None
Knowledge, Abilities, or Skills:
▪ Analysis, Basic Calculus, and Linear Algebra</td><td>Frequency

Annually (Fall)

Duration

1 semester</td><td colspan="2">Forms of Learning and Teaching

▪ Seminar (17.5 hours)
▪ Private study (45 hours)

Workload

62.5 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation

Read the Syllabus</td></tr>
<tr><td colspan="6">Content and Educational Aims

An executive career in an international business environment requires excellent communication and presentation skills. Managers have to communicate effectively with a large variety of target audiences, often in different languages and with different cultural backgrounds. This is true for employees and/or direct reports, business partners as well as customers. The ability to present and communicate succinctly and confidently while being culturally aware and building rapport and trust with different audiences is crucial. In this interactive module, students are introduced to the basics of effective presentation and communication techniques. They learn how to present themselves, their business project, or academic work, with impact, tailoring both the content and their delivery style to different types of audiences.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

Upon completion of the module, students will be able to

▪ act as effective communicators – in both group and individual situations;
▪ understand interpersonal communication models and group dynamics in presentations;
▪ enjoy the process of presenting;
▪ understand the importance of building rapport and trust with audiences;
▪ use presentation software (PowerPoint, Prezi) confidently and in a visually pleasant way;
▪ learn how to structure presentations in a coherent manner and develop captivating narratives;
▪ work with different presentation formats (Ignite, Pecha Kucha, Pitching etc.);
▪ understand and apply the basics of logical reasoning in oratory (deductive/inductive);
▪ develop oratory and rhetorical skills drawing on Aristotle's teaching of logos, ethos and pathos;
▪ understand and apply the basics of interpersonal communication (Johari Window, 4-Ears model etc.);
▪ give and receive constructive feedback;
▪ present themselves in different business situations;</td></tr>
</table>

- collaborate effective in intercultural teams.

Indicative Literature

This course utilizes lecture formats, case studies and interactive presentations, discussions, role play and peer-to-peer coaching. The course will also use internet resources, videos, and home assignments to illustrate and practice specific communication aspects.

Usability and Relationship to other Modules

This module is recommended to be taken together with the elective modules in the Bio-Informatics track.

Examination Type: Module Examination

Assessment Type: Oral Presentation

Duration: 15 minutes
Weight: 100%

Scope: All intended learning outcomes of this module.

3.6.4 Ethics and the Information Revolution

<table>
<tr><td colspan="3">Module Name
Ethics and the Information Revolution</td><td>Module Code
MDSSB-EIR-01</td><td>Level (type)
Year 2 (Career)</td><td>CP
2.5</td></tr>
<tr><td colspan="6">Module Components

Number | Name | Type | CP
MDSSB-EIR-01 | Ethics and the Information Revolution | Seminar | 2.5</td></tr>
<tr><td>Module Coordinator

Prof. Dr. Hilke Brockmann</td><td colspan="3">Program Affiliation

• MSc Data Science for Society & Business</td><td colspan="2">Mandatory Status

Mandatory DE, mandatory elective for DSSB</td></tr>
<tr><td colspan="3">Entry Requirements

Pre-requisites
☒ None

Co-requisites
☒ None

Knowledge, Abilities, or Skills
☒ None</td><td>Frequency

Annual (Fall)

Duration

1 semester</td><td colspan="2">Forms of Learning and Teaching

• Seminar (17.5 hours)
• Private study (45 hours)

Workload

62.5 hours</td></tr>
<tr><td colspan="6">Recommendations for Preparation

Read the Syllabus.
Binns (2018) Fairness in Machine Learning: Lessons from Political Philosophy. Proceedings of Machine Learning Research 81:1-11.</td></tr>
<tr><td colspan="6">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, in-classroom discussions and interactions are indispensable for identifying possible dilemmas and conflict of interests and for balancing contradictions to derive practical solutions and policy advice.</td></tr>
<tr><td colspan="6">Intended Learning Outcomes

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

• report on major ethical theories relevant to digital technologies
• integrate different ethical standpoints and arguments to address concrete societal problems
• assess the societal and ethical implications of digitization
• deal with legal aspects of ethics by applying means to prevent and deal with violations of privacy and transparency
• apply actions to contribute to the transition to a more just and trustworthy digital transformation as a part of one’s job
• implement justice and social equality as dimensions of ethics and sustainability</td></tr>
<tr><td colspan="6">Indicative Literature

Binns (2018) Fairness in Machine Learning: Lessons from Political Philosophy. Proceedings of Machine Learning Research 81:1-11.</td></tr>
</table>

Usability and Relationship to other Modules It is one of the three Career modules (IT Law, Language III, and Ethics and the Information Revolution) that can be chosen for replacement by the internship. Students need to replace 10 CP for the internship.
Examination Type: Module Examination Assessment Type: Term Paper (report) Length: 20 pages Weight: 100% Scope: All intended learning outcomes of the module.

3.7 Master Thesis (30 CP)

Module Name	*Module Code*	*Level (type)*	*CP*
Master Thesis	MDE-THE-01	Year 2	30

Module Components

Number	*Name*	*Type*	*CP*
MDE-THE-01	Master Thesis	N.A.	30

Module Coordinator	*Program Affiliation*	*Mandatory Status*
Prof. Dr. Stefan Kettemann	• MSc Data Engineering	Mandatory for DE

Entry Requirements

Pre-requisites	*Co-requisites*	*Knowledge, Abilities, or Skills*
• MDE-DIS-02 Advanced Project I • MDE-DIS-03 Advanced Project II	☒ None	• Proficiency in the area of the chosen thesis topic.

Frequency	*Forms of Learning and Teaching*
Annually (Spring)	• Private Study (750 hours)
Duration	*Workload*
1 semester	750 hours

Recommendations for Preparation

Read the Syllabus.

Content and Educational Aims

The aim of this module is to train students to motivate, design, carry out and document a research project in one of the areas represented by the research groups of the faculty of DE. Some familiarity with the requisite data engineering techniques will typically have been acquired in one of the preceding Advanced Projects (MDE-DIS-02 or MDE-DIS-03). The thesis topic is determined in mutual agreement with the module instructor. They may arise from the ongoing research in the instructor's own research group, but it is also possible for a student to adopt a topic of his/her own choice provided the instructor agrees to supervise it. 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 likely give substantial guidance for (i) and (iii), whereas the other aspects will be addressed with larger degrees of self-guidance. A research proposal document summarizing (i) – (iv) is expected as an interim result and milestone (target size: 10 pages). In the first weeks of the course, an intense taught tutorial on scientific working and writing is held. 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.

Intended Learning Outcomes

Discipline-Specific Skills (subject area depending on research discipline of the hosting group):

- understanding, at a professional level, of a circumscribed segment of the hosting group's research area;
- ability to apply specific and selected DE techniques, as required for the project, at a professional level;
- general professional skills;
- designing and carrying out the full cycle of a scientific research project in a professional manner;
- formulating a research proposal such that that it could serve as a funding proposal;

- writing a research thesis such that it could be submitted to a scientific publication venue, or as a project report to a funding agency or industrial client;
- presentation of project results for specialists and non-specialists.

Indicative Literature

N.A.

Usability and Relationship to other Modules

The master thesis can build on the advanced project courses MDE-DIS-02 Advanced Project 1 and MDE-DIS-03 Advanced Project 2 but the students are free to choose a different topic and a different supervisor for the master thesis.

Examination Type: Module Examination

Assessment Component 1: Thesis	Length: 30 – 60 pages Weight: 75%

Scope: All intended learning outcomes of this module.

Assessment Component 2: Oral Examination (Defense)	Duration: 20 minutes Weight: 25%

Scope: Mainly presentation of project results but the presentation touches all intended learning outcomes

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

4 Data Engineering Graduate Program Regulations

4.1 Scope of These Regulations

The regulations in this handbook are valid for all students who entered the Data Engineering graduate program at Jacobs University in Fall 2021. In case of conflict between the regulations in this handbook and the general Policies for Master Studies, the latter apply (see http://www.jacobs-university.de/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, Jacobs 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.

4.2 Degree

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

4.3 Graduation Requirements

In order to graduate, students need to obtain 120 CP. In addition, the following graduation requirements apply:

- In each module, students need to obtain a minimum amount of CP as indicated in chapter 2 of this handbook.
- Students need to complete all mandatory components of the program as indicated in chapter 2 of this handbook.

5 Appendices

5.1 Intended Learning Outcomes Assessment-Matrix

Data Engineering (M.Sc.)					MCO003 – BigData Challenge	MCO011-340131 Data Analytics	Data Engineering in Society	CO22-320372 Machine Learning	MCO014 – DataVisImage	MCO015 – DataAcquiSens	MECS001-340101 Princ.Stat.Model.	MECS002 – NetworkTheo	MECS003 – DataComp	Advanced Databases	Parallel and Distributed Computing,	MEGI001 – Geoinf	MEGI002 – GeoinfLab	Introduction to Data Management with Python	MEBI001-550432 Introduction to Systems Biology	MDE-BIO-03	MDE-BIO-031Management and Analysis	MESC001-340122 Data Mining	MCO008-051008 Data Analytics in Supply Chain Management	MMM005-340153Modern Signal Processing	MMM007-550443Network Approaches in Biology and Medicine	CO18-110231 Applied Dyn. Systems	MMM009-340181Calculus and Linear Algebra for Graduate Students	MMM010 Computational Methods for Data Engineering	MMM011-340171 Probabilities for Graduate Students	MRD004 – CurTopDE	MRD005 – AdvProj1	MRD006 – AdvProj2	Master's Thesis	MCA002 – Language MA	MCA006- Communication & Presentation Skills f. Ex.	MDSSB-EIR-01Ethics and the Information Revolution	MCA008-Academic Writing Skills/Intercultural Training
Semester					1	1	2	2	3	1, 2 or 3	2	1 or 3	1 or 3	2	3	1 or 3	2	1	2	1 or 3	2	3	2	1or 3	2	1	1	1	2	3	4	1,2,3	1,3	1,3	2
Mandatory/ optional					M	M	M	M	M	M	O	O	0	O	O	O	O	M	O	O	O	O	O	O	O	O	O	M	M	M	M	M	M	M	M
Credits					5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	30	5	5	5	2.5
Competencies*																																					
Program Learning Outcomes	**A**	**E**	**P**	**S**																																	
ILO 1 critically assess and creatively apply technological possibilities and innovations driven by big data	x	x			x	x						x		x	x	x		x		x	x	x		x				x	x	x	x		x	x	
ILO 2 use sensors and microcontrollers to collect data and to transmit them to databases on servers or the internet in general	x	x								x							x																			
ILO 3 set up and use databases to efficiently and securely manage and access large amounts of data	x	x				x						x	x	x	x		x	x		x	x	x		x			x			x	x	x				
ILO 4 apply statistical concepts and use statistical models in the context of real-life data analytics	x	x				x					x	x	x			x				x	x	x		x	x		x	x		x	x	x				
ILO 5 use, adapt and improve visualization techniques to support data-based decision making	x	x							x	x		x				x	x	x												x	x	x				
ILO 6 design, implement and exploit various representations of data for classification and regression including supervised machine learning methods and	x	x						x			x	x	x	x		x						x	x	x						x	x	x				
ILO 7 apply and critically assess data acquisition methods and analytical techniques in real life situations,	x	x		x	x	x		x							x	x						x		x				x	x	x	x				
ILO 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	x	x		x				x				x	x			x			x			x								x	x	x				
ILO 9 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	x	x	x	x	x								x			x				x	x	x	x	x	x				x	x	x	x				x
ILO 10 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				x					x	x	x	x		x	x	x
ILO 11 engage ethically with academic, professional and wider communities and actively contribute to a sustainable future			x	x			x																						x	x	x	x			x	x
ILO 12 take responsibility for their own learning, personal development and role in society, evaluating critical feedback and self-analysis			x	x			x									x				x		x							x	x	x	x		x	x	x
ILO 13 take on lead responsibility in a diverse team		x	x				x										x												x	x	x	x				
ILO 14 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	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Assessment Type																																					
oral examination																				x			x	x									x	x	
final written exam						x		x	x			x	x						x							x	x	x							
project					x					x	x					x	x				x	x			x					x	x	x		x	x	
essay																																				x
lab report																																					
poster presentation																													x								
presentation																																					
various																															x						
*Competencies: A-scientific/academic proficiency; E-competence for qualified employment; P-development of personality; S-competence for engagement																																					