Study Program Handbook

Minor in Data Science

Subject-specific Examination Regulations for a Minor in Data Science (Fachspezifische Prüfungsordnung)

The subject-specific examination regulations for a Minor in Data Science are defined by this program handbook and are valid only in combination with the General Examination Regulations for Undergraduate degree programs (General Examination Regulations = Rahmenprüfungsordnung).

Version	Valid as of	Decision	Details
Fall 2022 – V1	Sep 01, 2022	May 25, 2022	V1 Approved by the Academic Senate
		Jun 09, 2022	V1.1 Editorial changes (Insertion of Module numbers; error correction in Modules "Applied Machine Learning" and "Introduction to Data Science")

CONTENTS

1 Program Overview

1.1 Concept

1.1.1 The Jacobs University Educational Concept

Jacobs University aims to educate students for both an academic and a professional career by emphasizing four core objectives: academic quality, self-development/personal growth, internationality and the ability to succeed in the working world (employability). Hence, study programs at Jacobs University offer a comprehensive, structured approach to prepare students for graduate education as well as career success by combining disciplinary depth and interdisciplinary breadth with supplemental skills education and extra-curricular elements.

In this context, it is Jacobs University's aim to educate talented young people from all over the world, regardless of nationality, religion, and material circumstances, to become citizens of the world who are able to take responsible roles for the democratic, peaceful, and sustainable development of the societies in which they live. This is achieved through a high-quality teaching as well as manageable study loads and supportive study conditions. Study programs and related study abroad programs convey academic knowledge as well as the ability to interact positively with other individuals and groups in culturally diverse environments. The ability to succeed in the working world is a core objective for all study programs at Jacobs University, both in terms of actual disciplinary subject matter and also to the social skills and intercultural competence. Study-program-specific modules and additional specializations provide the necessary depth, interdisciplinary offerings and the minor option provide breadth while the university-wide general foundation and methods modules, mandatory German language requirements, and an extended internship period strengthen the employability of students. The concept of living and learning together on an international campus with many cultural and social activities supplements students' education. In addition, Jacobs University offers professional advising and counseling.

Jacobs University's educational concept is highly regarded both nationally and internationally. While the university has consistently achieved top marks over the last decade in Germany's most comprehensive and detailed university ranking by the Center for Higher Education (CHE), it has also been listed by the renowned Times Higher Education (THE) magazine as one of the top 300 universities worldwide (ranking group 251-300) in 2019, 2020 and 2021. The THE ranking is considered as one of the most widely observed university rankings. It is based on five major indicators: research, teaching, research impact, international orientation, and the volume of research income from industry.

1.1.2 Program Concept

Digitally transforming and data-driven societies reveal novel opportunities and challenges that are only partially acknowledged in disciplinary study programs. The Minor in Data Science at Jacobs University puts inter-disciplinarity, diversity, and an internationally shared expert knowledge and understanding into action. Our students obtain a solid foundation in this emerging field, covering both technical and non-technical aspects. With an emphasis on online and asynchronous learning, the Minor in Data Science is open to students from all Jacobs University Bachelor programs allowing for a minor in their curricular structure.

1.2 Specific Advantages of Data Science at Jacobs University

The small campus of Jacobs University with its broad spectrum of disciplines including Social Sciences, Natural Sciences, Engineering and Computer Science provides ample opportunities for collaborative and interdisciplinary data-driven projects. The Minor in Data Science offers students a natural route to the in-house graduate programs Data Engineering (DE) and Data Science for Society and Business (DSSB). By advancing the cooperation of Jacobs University with partners from academia, industry and business, Data Science opens new opportunities for our students.

Asynchronous learning is promoted through online and hybrid course formats in order to manage individual preferences and heterogeneous expertise. Theory and applications are integrated in courses combining lectures with practical tutorials and research projects in order to deepen an understanding through immersion. An innovative teaching approach switches flexibly between these formats and tailor's education towards a personalized experience.

1.3 Program-Specific Educational Aims

1.3.1 Qualification Aims

The Minor in Data Science offers additional interdisciplinary qualification and training in universally applied digital skills for undergraduate students from all majors. Graduates are well prepared to either enter a thriving job market or to successfully pursue their academic studies in many dynamically growing research fields. Particular attention is paid to the following components.

- Societal and legal implications of data science and digitization, business and career opportunities.
- Problem-oriented learning through hands-on computational exercises with Python and other up-to-date software.
- Theoretical background: data operations and analysis pipeline, computer science basics, statistics, and maths.
- Data science applications and case studies from social sciences, natural sciences, and engineering.
- Preparation for life-long digital learning through online/asynchronous course formats.

General qualification follows the Digital Education Plan of the European Union (EU), accelerating the transformation of an open and inclusive higher education in the EU but also beyond https://education.ec.europa.eu/document/digital-education-action-plan.

1.3.2 Intended Learning Outcomes

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

1. explain societal implications of the digital transformation,
2. understand the legal data protection framework,
3. know the data analysis pipeline and basics of database management,
4. identify important problem types and solution approaches in data analytics,
5. enumerate and describe main Machine Learning (ML) tasks and applications,
6. process ordered data sets such as time series and images,

7. prepare unstructured data sets for processing and analysis,
8. explain matrix decompositions and their usage in data science,
9. apply important ML techniques to classification and regression problems,
10. compose Python scripts and functions to solve data science tasks,
11. apply functions from the libraries NumPy, SciPy, Pandas to data science problems,
12. address Machine Learning tasks by means of the Python library SciKit-learn (sklearn).

1.4 Target Group

The Minor in Data Science is targeted towards students of all areas interested in quantitative scientific analysis, interpretation, and decision-making. There are no explicit prerequisites but a sound foundation in Mathematics is helpful. Full synergy within the minor option for all Bachelor students in all programs, except for those programs excluding a minor track in their curricular structure, or supporting only specific minor subjects (presently IEM, ECE, and RIS).

1.5 More Information and Contact

For more information, please contact the study program chair:

TBD

https://www.jacobs-university.de/study/undergraduate/programs/major-minor-combination

2 The Curricular Structure

2.1 General

The curricular structure provides multiple elements for enhancing employability, interdisciplinarity, and internationality. The unique Jacobs Track, offered across all undergraduate study programs, provides comprehensive tailor-made modules designed to achieve and foster career competency. Additionally, a mandatory internship of at least two months after the second year of study and the possibility to study abroad for one semester give students the opportunity to gain insight into the professional world, apply their intercultural competences and reflect on their roles and ambitions for employment and in a globalized society.

All undergraduate programs at Jacobs University are based on a coherently modularized structure, which provides students with an extensive and flexible choice of study plans to meet the educational aims of their major as well as minor study interests and complete their studies within the regular period.

The framework policies and procedures regulating undergraduate study programs at Jacobs University can be found on the website (https://www.jacobs-university.de/academic-policies).

2.2 The Jacobs University 3C Model

Jacobs University offers study programs that comply with the regulations of the European Higher Education Area. All study programs are structured according to the European Credit Transfer System (ECTS), which facilitates credit transfer between academic institutions. The three-year under-graduate program involves six semesters of study with a total of 180 ECTS credit points (CP). The undergraduate curricular structure follows an innovative and student-centered modularization scheme - the 3C-Model - that groups the disciplinary content of the three study years according to overarching themes:

Year I:
CHOICE
Students have the CHOICE to decide on their major after the first year of study.

Year II:
CORE
Students study the CORE elements of their major and may choose a minor.

Year III:
CAREER
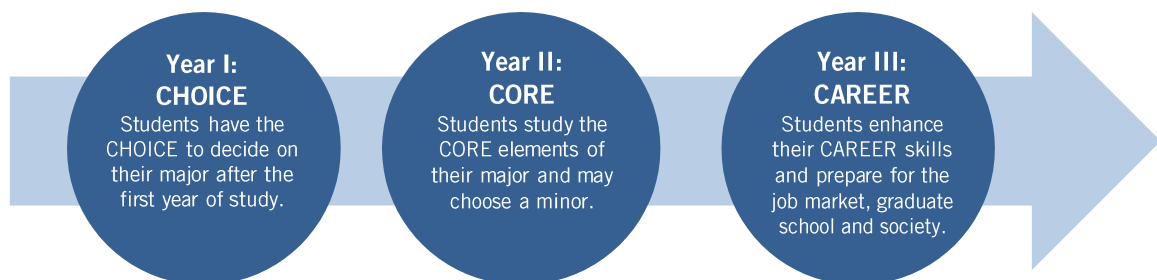Students enhance their CAREER skills and prepare for the job market, graduate school and society.

Figure 1: The Jacobs University 3C-Model

2.3 Module Requirements for a Minor in Data Science

A minor in Data Science requires 30 CP. The default option to obtain a minor in Data Science is pictured in the schematic study plan in chapter 4. It includes the following CHOICE and CORE modules:

- CHOICE Module: Introduction to Data Science (7.5 CP)
- CHOICE Module: Data Structures and Processing (7.5 CP)
- CORE Module: Data Analytics and Modeling (7.5 CP)
- CORE Module: Applied Machine Learning (7.5 CP)

3   Minor in Data Science Undergraduate Program Regulations

3.1 Scope of these Regulations

The regulations in this handbook are valid for all students who entered their undergraduate studies at Jacobs University in Fall 2022. In case of a conflict between the regulations in this handbook and the general Policies for Bachelor Studies, the latter applies (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.

3.2 Degree

After successful completion, the minor in Data Science will be listed on the final transcript under PROGRAM OF STUDY and BA/BSc – [name of the major] as "(Minor: Data Science)."

4   Schematic Study Plan for a Minor in Data Science

Figure 2 shows schematically the sequence and types of modules required for the study program.

Minor in Data Science at Jacobs University (30 CP)

Bachelor Thesis / Seminar (m, 15 CP)			**Big Questions** (me, 5 CP)	**Big Questions** (me, 2.5 CP)
Study Abroad Option (22.5 CP) **Specialization** (me, 3 x 5 CP)			**Community Impact Project** (m, 5 CP)	**Big Questions** (me, 2.5 CP)
Internship/Startup (Summer) (15 CP)				
CORE (program specific, m, 7.5 CP)	**CORE** (program specific, m, 7.5 CP)	**CORE*** Applied Machine Learning (me, 7.5 CP)	**Methods/Skills** (shared with other programs, 5.0 CP)	**Language** (me, 2.5 CP)
CORE (program specific, m, 7.5 CP)	**CORE** (program specific, m, 7.5 CP)	**CORE*** Data Analytics and Modeling (me, 7.5 CP)	**Methods/Skills** (shared with other programs, 5.0 CP)	**Language** (me, 2.5 CP)
CHOICE (program specific, m, 7.5 CP)	**CHOICE** (program specific / individual selection, m / me, 7.5 CP)	**CHOICE*** Data Structures and Processing (me, 7.5 CP)	**Methods/Skills** (Shared with other programs, 5.0 CP)	**Language** (me, 2.5 CP)
CHOICE (program specific, m, 7.5 CP)	**CHOICE** (program specific / individual selection, m /me, 7.5 CP)	**CHOICE*** Introduction to Data Science (me, 7.5 CP)	**Methods/Skills** (Shared with other programs, 5.0 CP)	**Language** (me, 2.5 CP)
CHOICE / CORE 90 CP			**JACOBS TRACK 45 CP**	

Year 3, **Year 2**, **Year 1**, **Area**

* mandatory for Data Science minor students
m = mandatory
me = mandatory elective

Figure 2: Schematic Study Plan for a Minor in Data Science

5   Module Descriptions

5.1. Introduction to Data Science

Module Name		Module Code	Level (type)	ECTS
		CH-700	Year 1	7.5
Introduction to Data Science				

Module Components				
Number	Name		Type	ECTS
CH-700-A	Introduction to Data Science		Lecture	7.5

Module Coordinator	Program Affiliation	Mandatory Status
	• Minor in Data Science	Mandatory for CSSE and Minor in Data Science

Entry Requirements			Frequency	Forms of Learning and Teaching
Pre-requisites	Co-requisites	Knowledge, Abilities, or Skills	Annually (Fall)	• Lectures (hybrid / online) (52.5 hours) • Private Study (135 hours)
☐ None	☐ None		**Duration** 1 semester	**Workload** 187.5 hours

Recommendations for Preparation
None.

Content and Educational Aims
The module introduces data science with an integrated presentation of three essential components, namely, (1) societal/legal implications and business opportunities, (2) technical/theoretical background and case studies, (3) an introduction to the Python coding environment. The first component entails a conceptual introduction to the opportunities and the challenges of a digitally transformed and data-driven society, presentations on industry standards and legal frameworks, and discussions of critical issues such as cybersecurity and surveillance. The second component includes topics such as data science terminology, digital data and their representations, and introductions to exploratory data analysis and prominent supervised and unsupervised learning tasks. The third component offers an introduction to the Python ecosystem of data representation, processing, analysis, and visualization, starting with Jupyter notebooks, installing suitable environments, and introductions to data science related packages such as NumPy, SciPy, Matplotlib, Seaborn, and Pandas. Fundamental data science concepts are summarized and illustrated using real-world data from various disciplines. Flexible educational formats (mostly online and hybrid) allow for asynchronous learning. Lectures are combined with an exposure to Python programming and data processing and visualization environments, including hands-on practicals, examples, and exercises.

Intended Learning Outcomes

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

- explain societal implications of the digital transformation,
- understand the legal data protection framework,
- carry out basic data processing and visualization tasks,
- apply fundamental data science methods to structured data,
- understand the logic of Python scripts and functions,
- compose Python code using templates

Indicative Literature

Ani Adhikari, John DeNero, David Wagner. Computational and Inferential Thinking: The Foundations of Data Science. Originally developed for the UC Berkeley course Data 8: Foundations of Data Science. An online version of the textbook is available at https://inferentialthinking.com/.

The Alan Turing Institute, Data Science for the Social Good.

Philip D . Brooker. Programing with Python for Social Scientists. Sage 2020.

Shin Takahasi, Iroha Inoue. The Manga Guide to Linear Algebra. Trend-Pro 2012.

Steven S. Skiena. The Data Science Design Manual. Springer 2017.

Jake Vanderplas. Python Data Science Handbook. O'Reilly 2016. An online version is available at https://jakevdp.github.io/PythonDataScienceHandbook/.

Shoshana Zuboff. The Age of Surveillance Capitalism. London: Profile 2019.

Usability and Relationship to other Modules

-

Examination Type: Module Examination

Type: Written Examination   Duration/Length: 180 min
Scope: All intended learning outcomes of the module.   Weight: 100 %

Module achievement: 50% of the assignments need to be correctly solved.

5.2 Data Structures and Processing

Module Name		Module Code	Level (type)	CP
Data Structures and Processing		CH-701	Year 1	7.5

Module Components					
Number	Name			Type	CP
CH-701-A	Data Structures and Processing			Lecture	7.5

Module Coordinator	Program Affiliation	Mandatory Status
	• Minor in Data Science	Mandatory for Minor in Data Science

Entry Requirements			Frequency	Forms of Learning and Teaching
Pre-requisites	Co-requisites	Knowledge, Abilities, or Skills	Annually (Spring)	• Lectures (hybrid / online) (52.5 hours)
☒ none	☒ none	•		• Private Study (135 hours)
			Duration	**Workload**
			1 semester	187.5 hours

Recommendations for Preparation
Required for solving the coding assignments are Python skills at the level achieved after successful completion of the module Introduction to Data Science.

Content and Educational Aims
In this module, data structures and the data analysis pipeline are introduced in three parts. The first part gives an overview of the data analysis pipeline from capturing and processing to storing and analyzing data. Database concepts and management as well as the basic distinction between structured and unstructured data are reviewed, including an introduction to the relational data model, supplemented by examples of how specific disciplinary databases are handled. The second part is concerned with different types of structured data, starting with time series and images as examples of ordered data vectors and data matrices, respectively, and addressing both numeric and text data. Particular emphasis will be on tables and their higher-dimensional extensions, allowing for multivariate correlation and regression studies. The third part deals with unstructured data as obtained from web scraping and text mining. Unstructured data need to be prepared for subsequent analyses and use through operations such as merging, ordering, transforming, and resampling. Flexible educational formats (mostly online and hybrid) allow for asynchronous learning. Lectures are combined with Python exercises with particular emphasis on the Pandas package. Disciplinary applications and case studies are immersed as bridging elements.

Intended Learning Outcomes
Upon completion of this module, students will be able to: 1. enumerate and explain key operations along the data analysis pipeline,   2. understand the basics of database management and important data models,   3. process ordered data sets such as time series and images,   4. prepare unstructured data sets for processing and analysis,   5. apply the Pandas package to process and display time series, images, and tables,   6. use Python tools to prepare and process unstructured data.

Indicative Literature

- Ani Adhikari, John DeNero, David Wagner. Computational and Inferential Thinking: The Foundations of Data Science. Originally developed for the UC Berkeley course Data 8: Foundations of Data Science. An online version of the textbook is available at https://inferentialthinking.com/.
- Steven S. Skiena. The Data Science Design Manual. Springer 2017.
- Jake Vanderplas. Python Data Science Handbook. O'Reilly 2016. An online version is available at https://jakevdp.github.io/PythonDataScienceHandbook/.

Usability and Relationship to other Modules

-

Examination Type: Module Examination

Type: Written Examination	Duration/Length: 180 min
Scope: All intended learning outcomes of the module.	Weight: 100 %

Module achievement: 50% of the assignments need to be correctly solved.

5.3 Data Analytics and Modeling

Module Name		Module Code	Level (type)	CP
Data Analytics and Modeling		CO-710	Year 2	7.5

Module Components				
Number	Name		Type	CP
CO-710-A	Data Analytics and Modeling		Lecture	7.5

Module Coordinator	Program Affiliation	Mandatory Status
	• Minor in Data Science	Mandatory for Minor in Data Science

Entry Requirements			Frequency	Forms of Learning and Teaching
Pre-requisites	Co-requisites	Knowledge, Abilities, or Skills	Annually (Fall)	• Lectures (hybrid / online) (52.5 hours)
☒ none	☒ none	•		• Private Study (135 hours)
			Duration	**Workload**
			1 semester	187.5 hours

Recommendations for Preparation

Required for solving the coding assignments are Python skills at the level achieved after successful completion of the module Introduction to Data Science. Furthermore, students are encouraged to review first-year level statistics and linear algebra.

Content and Educational Aims

The module offers an introduction to the principles of data analytics and predictive data modeling and is structured into four parts. First, essential concepts from statistics are reviewed in the data modeling context, illustrating key ideas including randomness, distributions, and confidence regions. Examples and case studies are discussed to distinguish between proper and improper uses of statistics. Basic linear algebra is reviewed in the second part of the module, emphasizing vectors, distances, linear equations, matrices, and inversion. Key ideas such as the least squares approach are motivated with geometrical principles. The third part of the module is concerned with matrix decompositions such as the Singular Value Decomposition (SVD) and its close relatives Principal Component Analysis (PCA) and Empirical Orthogonal Function (EOF) analysis. The fourth part clarifies the distinction between linear and nonlinear modeling, and introduces key nonlinear techniques. Flexible educational formats (mostly online and hybrid) allow for asynchronous learning. Lectures are combined with Python exercises. Disciplinary applications and case studies are immersed as bridging elements.

Intended Learning Outcomes

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

1. identify important problem types and solution approaches in data analytics,
2. understand how key concepts from statistics and linear algebra enter data science,
3. explain matrix decompositions and their usage in data science,
4. discuss regularization concepts and optimality criteria in data analytics,
5. know the basics of nonlinear modeling and related computational approaches,

6. convert data structures to Python/NumPy arrays for usage in data modeling,
7. apply Python statistics and linear algebra tools in data analytics and modeling.

Indicative Literature

- Ani Adhikari, John DeNero, David Wagner. Computational and Inferential Thinking: The Foundations of Data Science. Originally developed for the UC Berkeley course Data 8: Foundations of Data Science. An online version of the textbook is available at https://inferentialthinking.com/.
- Steven S. Skiena. The Data Science Design Manual. Springer 2017.
- Gilbert Strang: Linear Algebra and Learning from Data. Wellesley-Cambridge 2019. See https://math.mit.edu/~gs/learningfromdata/.
- Joe Suzuki: Statistical Learning with Math and Python. Springer 2021.
- Jake Vanderplas. Python Data Science Handbook. O'Reilly 2016. An online version is available at https://jakevdp.github.io/PythonDataScienceHandbook/.

Usability and Relationship to other Modules

-

Examination Type: Module Examination

Type: Written Examination   Duration/Length: 180 min
Scope: All intended learning outcomes of the module.   Weight: 100 %

Module achievement: 50% of the assignments need to be correctly solved.

5.4 Applied Machine Learning

Module Name	Module Code	Level (type)	CP
Applied Machine Learning	CO-711	Year 2	7.5

Module Components			
Number	Name	Type	CP
CO-711-A	Applied Machine Learning	Lecture	7.5

Module Coordinator	Program Affiliation	Mandatory Status
	• Minor in Data Science	Mandatory for Minor in Data Science

Entry Requirements			Frequency	Forms of Learning and Teaching
Pre-requisites	*Co-requisites*	*Knowledge, Abilities, or Skills*	Annually (Fall)	• Lectures (hybrid / online) (52.5 hours)
☒ none	☒ none	•		• Private Study (135 hours)
			Duration	**Workload**
			1 semester	187.5 hours

Recommendations for Preparation
Required for solving the coding assignments are Python skills at the level achieved after successful completion of the module Introduction to Data Science. Furthermore, students are encouraged to review first-year level statistics and linear algebra.

Content and Educational Aims
The module provides a hands-on introduction to Machine Learning (ML), emphasizing practical aspects of workflows and applications. Topics include k-Means clustering, Nearest Neighbor (NN) and Naive Bayes techniques, Decision Trees, Support Vector Machines (SVMs). Particular emphasis is on Neural Networks and Deep Learning. Theoretical concepts such as distance metrics, graphs, and networks are reviewed. Flexible educational formats (mostly online and hybrid) allow for asynchronous learning. Lectures are combined with Python exercises with particular emphasis on the SciKit-learn package. Disciplinary applications and case studies are immersed as bridging elements.

Intended Learning Outcomes
Upon completion of this module, students will be able to: 1. enumerate and describe main Machine Learning (ML) tasks and applications,   2. discern and explain important ML approaches to classification and regression,   3. apply Nearest Neighbor and Naive Bayes techniques to classification problems,   4. employ Decision Trees and Support Vector Machines to solve data science problems,   5. understand and use Neural Network and Deep Learning techniques,   6. address Machine Learning tasks by means of the Python library SciKit-learn (sklearn).

Indicative Literature
• Steven S. Skiena. The Data Science Design Manual. Springer 2017.   • Joe Suzuki: Statistical Learning with Math and Python. Springer 2021.   • Jake Vanderplas. Python Data Science Handbook. O'Reilly 2016. An online version is available at https://jakevdp.github.io/PythonDataScienceHandbook/.

Usability and Relationship to other Modules
•

Examination Type: Module Examination

Type: Written Examination   Duration/Length: 180 min
Scope: All intended learning outcomes of the module.   Weight: 100 %

Module achievement: 50% of the assignments need to be correctly solved.

6   Appendix

6.1   Intended Learning Outcomes Assessment-Matrix

Minor in Data Science					Introduction to Data Science	Data Structures and Processing	Data Analytics and Modeling	Applied Machine Learning
Semester					1	2	3	4
Mandatory (m) for a Minor in Data Science					m	m	m	m
Credits					7.5	7.5	7.5	7.5
	Competencies*							
Program Learning Outcomes	A	E	P	S				
Explain societal implications of the digital transformation	x	x	x		x			
Understand the legal data protection framework	x	x			x			
Know the data analysis pipeline and basics of database management	x	x	x	x	x	x	x	x
Identify important problem types and solution approaches in data analytics	x	x			x		x	x
enumerate and describe main Machine Learning (ML) tasks and applications	x	x			x		x	x
Process ordered data sets such as time series and images	x	x				x		
Prepare unstructured data sets for processing and analysis	x					x		
Explain matrix decompositions and their usage in data science	x						x	
Apply important ML techniques to classification and regression problems	x	x	x	x	x		x	x
Compose Python scripts and functions to solve data science tasks	x	x	x		x	x	x	x
Apply functions from the libraries NumPy, SciPy, Pandas to data science problems	x	x	x				x	
Address Machine Learning tasks by means of the Python library SciKit-learn (sklearn)	x	x	x	x				x
Assessment Type								
Oral examination								
Written examination					x	x	x	x
Project								
Term papaer, essay								
(Lab) Report								
Poster presentation								
Presentation								
Thesis								
Various								
Module achievements/bonus achievements					x	x	x	x

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