

Geo Data Science

Organizational Details

Prof. Dr. Martin Kada

Chair Methods of Geoinformation Science (GIS)
Institute of Geodesy and Geoinformation Science

Copyright Notice

The teaching materials for this course and all elements contained therein are protected by international copyright laws. They may only be used for study purposes for the corresponding course.

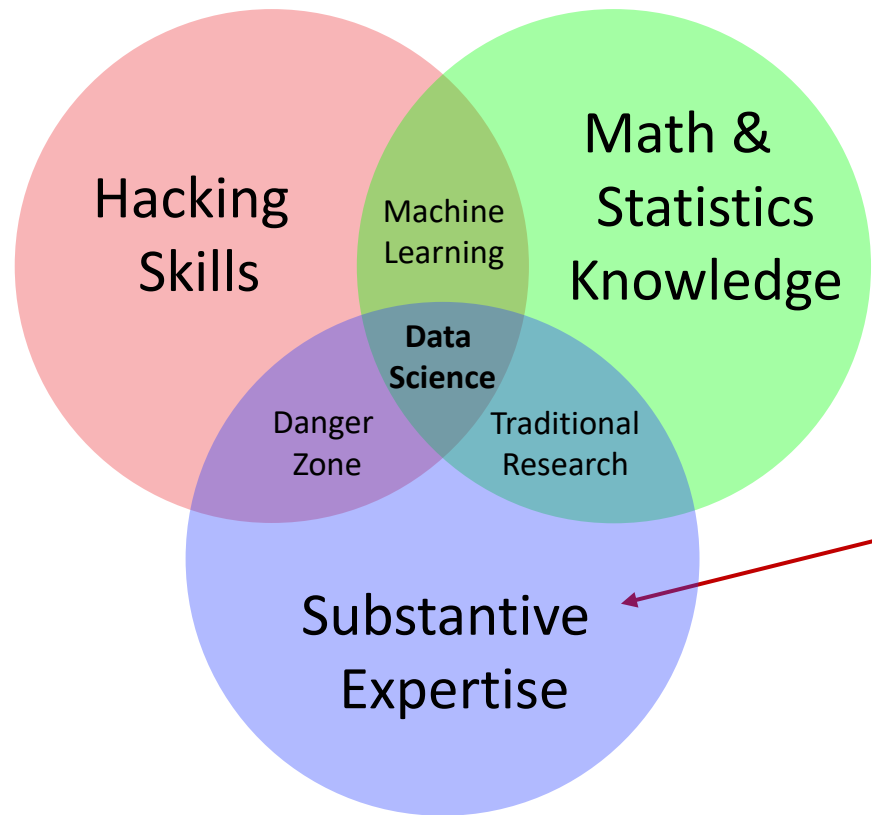
Any reproduction and redistribution of the course materials without written permission is prohibited, other than the following: You may print or download them for your own personal use while attending the course.

What is (Geo) Data Science?

“A data scientist is someone who knows more statistics than a computer scientist and more computer science than a statistician”

Geo Data + Data Science = Geo Data Science

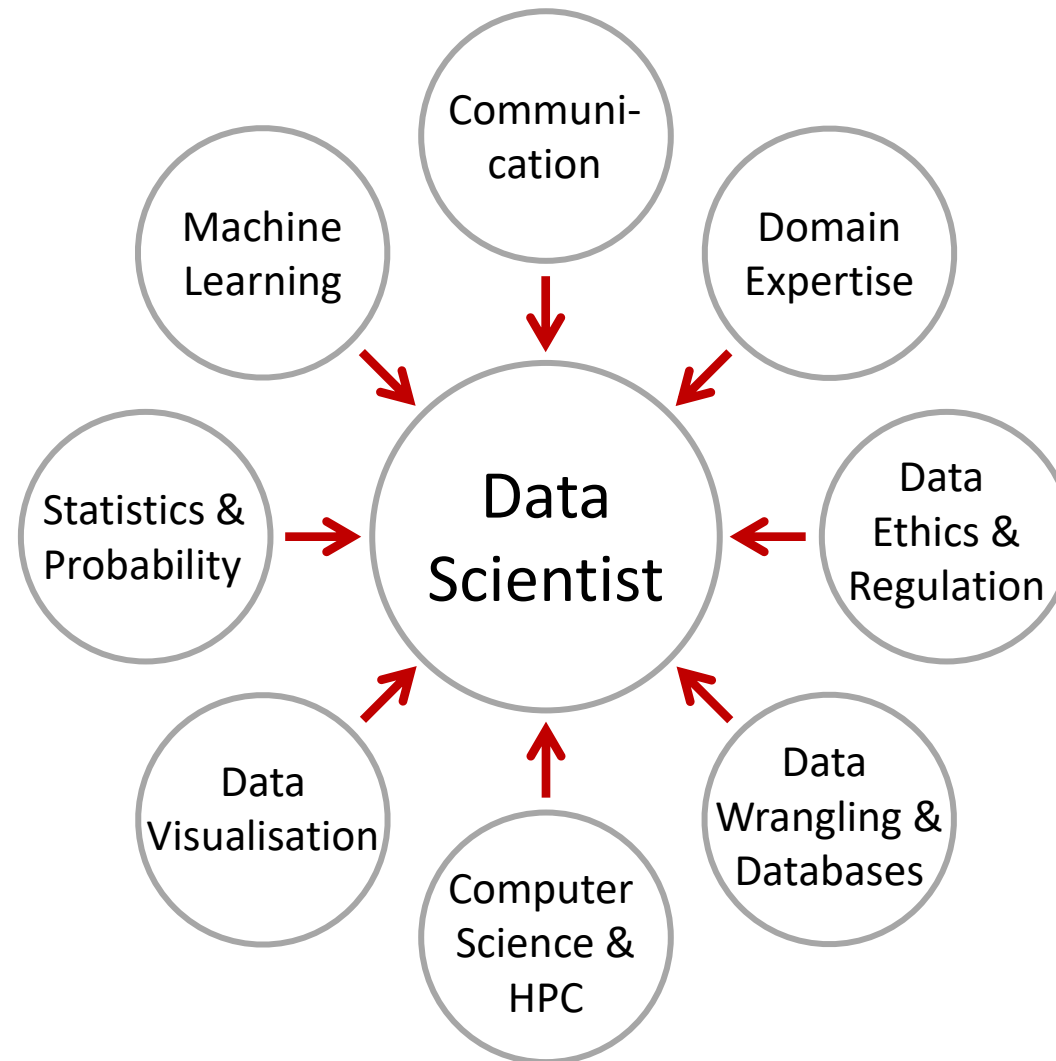
What is (Geo) Data Science?



Chair "Methods of Geoinformation Science"

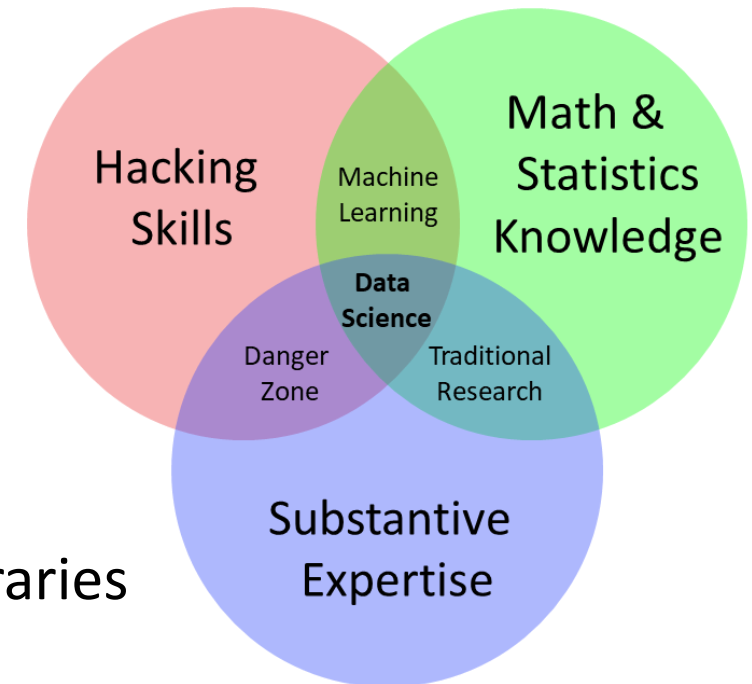
- 3D city models
- Aerial point cloud processing
- Geometric reconstruction of 3D building models
- Cartographic 3D generalization techniques

Skills of a Data Scientist



Course Content

- Data-oriented programming
- Exploratory thematic and spatial data analysis
- Data visualization and map generation
- **Supervised and unsupervised machine learning**
- Practical programming exercises in Python using numerical, scientific, and machine learning libraries (e.g. NumPy, SciPy, pandas, GeoPandas, scikit-learn)



Schedule & Team

- Monday 08h – 10h
 - Lilli Kaufhold
 - H6131 / H6134 (PC Pool)

Lilli Kaufhold
Exercises



- Monday 10h – 12h
 - Prof. Dr. Martin Kada
 - H6131

Prof. Dr.
Martin Kada
Lectures



- There may be deviations from the schedule,
but these will be posted in a timely manner on ISIS
 - Check regularly on ISIS and your TU Berlin emails for ISIS messages

Contact

- Contact Lilli Kaufhold for exercise related issues:
 - lilli.kaufhold@tu-berlin.de
- Contact Matthias Druve for technical issues with JupyterHub:
 - matthias.druve@tu-berlin.de
- Contact Almut Gothe for examination related issues
 - almut.gothe@tu-berlin.de
- In all other cases, contact Martin Kada
 - martin.kada@tu-berlin.de
- Please do not use ISIS messages to contact us, better write an email directly to us

- This course can be taken in the master's program Geodesy and Geoinformation Science in both major and minor subject area
 - 4 hours (total) of lectures & exercises per week
 - 6 credit points (ECTS)
 - 30-minute oral examination (no prerequisites)
- Other degree programs
 - Environmental Planning (M.Sc.)
 - Ecology and Environmental Planning (B.Sc. + M.Sc.)
 - Civil Systems Engineering (M.Sc.)
 - Check with your study program



Major & Minor Subject in GIT

Summer term

Major GIT (21CP):

Geographical
Information Systems
(3CP)

Summer & winter term

(Specialization) module
of your choice (6CP)

(Specialization) module
of your choice (6CP)

Winter term

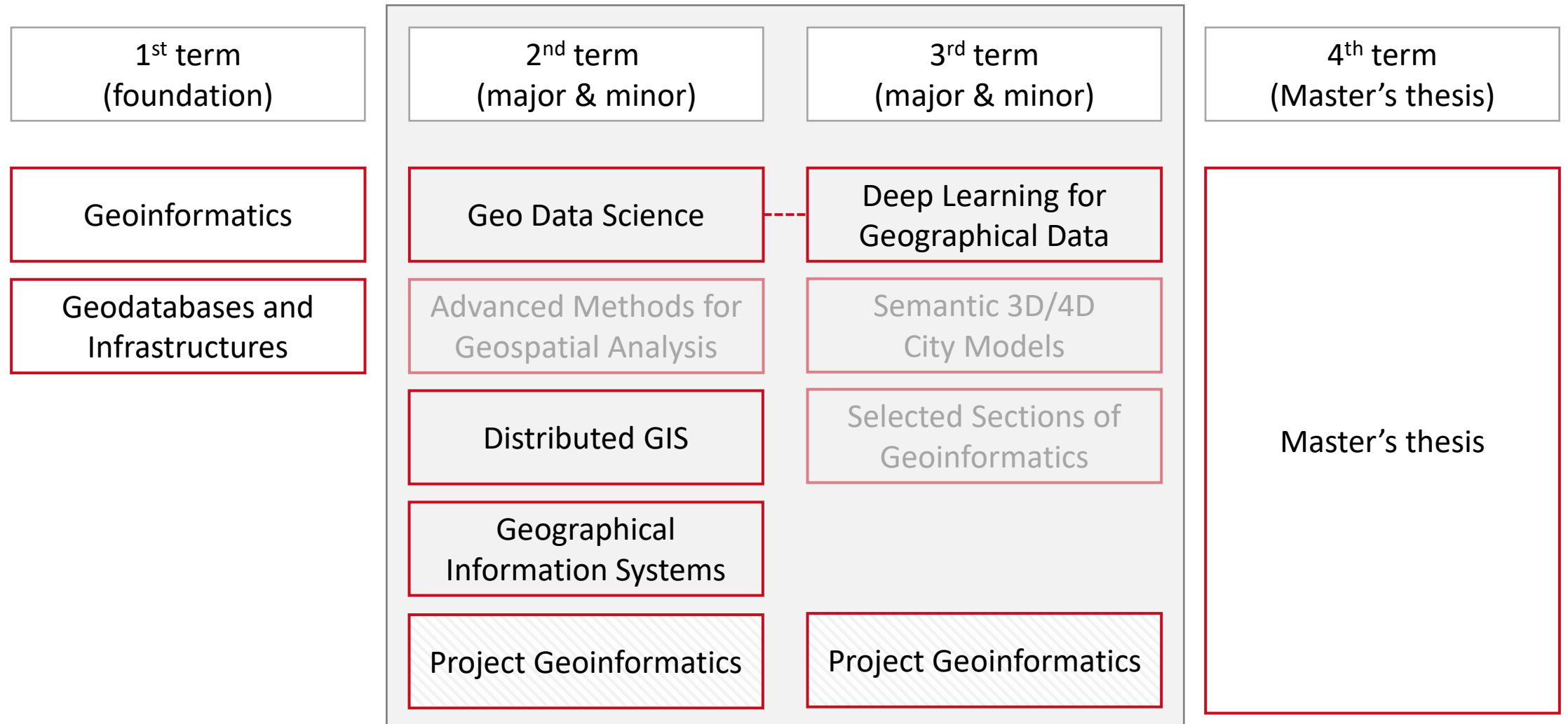
Project Geoinformatics
(6CP) (mandatory!!!)

Minor GIT (9CP):

Geographical
Information Systems
(3CP)

(Specialization) module
of your choice (6CP)

Modules offered by the Chair GIS



Learning Outcomes & Detailed Content

- For **learning outcomes** and more detailed **content**, please refer to the module description from the official module catalogue of the master's program "Geodesy and Geoinformation Science"
- The module description is also available on the ISIS page of the course

09.04.2021, 08:37:26 Uhr ModultBeschreibung #51433 / 1 Seite 1 von 2

Geo Data Science

Module title: Geo Data Science	Credits: 6	Responsible person: Kada, Martin
Website: No information	Office: H 12	Contact person: No information
	Display language: Englisch	E-mail address: martin.kada@tu-berlin.de

Learning Outcomes

Students have a profound understanding of the scientific fields of data science, big data and machine learning in general and applied to geographical data. They have acquired theoretical and practical knowledge in geo data management, manipulation, and visualization as well as familiarity with big data technologies. Students understand the mathematical background, the working principles, and applications of machine learning algorithms. They are able to transfer real-world problems from the geo-sciences into machine learning models, find and develop solution strategies, and implement them programmatically using Python in conjunction with the respective standard software libraries.

Content

Exploratory thematic and spatial data analysis (with Python), statistical analytics, correlation, data manipulation and cleaning, feature extraction from geographical data, supervised vs. unsupervised learning, linear and polynomial regression, regularized linear models, logistic and multinomial logistic regression, cost functions, model training and fine-tuning, gradient descent, learning curves, performance measures, support vector machines, decision trees, random forests, ensemble learning, dimensionality reduction, segmentation and clustering (k-means, hierarchical clustering, DBSCAN), privacy and ethics in data science, Python numerical, scientific, and machine learning libraries (e.g. NumPy, SciPy, pandas, GeoPandas, scikit-learn).

Module Components

Course Name	Type	Number	Cycle	SWS
Geo Data Science	UE		SS	2
Geo Data Science	VL		SS	2

Workload and Credit Points

Geo Data Science (Übung)	Multiplier	Hours	Total
Attendance	15.0	2.0h	30.0h
Pre/post processing	15.0	4.0h	60.0h
			90.0h

Geo Data Science (Vorlesung)	Multiplier	Hours	Total
Attendance	15.0	2.0h	30.0h
Pre/post processing	15.0	4.0h	60.0h
			90.0h

The Workload of the module sums up to 180.0 Hours. Therefore the module contains 6 Credits.

Description of Teaching and Learning Methods

Lectures (45%), exercises (45%), and independent reading (10%).

Requirements for participation and examination

Desirable prerequisites for participation in the courses:
Profound knowledge of geographical data representations and linear algebra, basic knowledge and experience in programming with Python.

Mandatory requirements for the module test application:
No information

Module completion

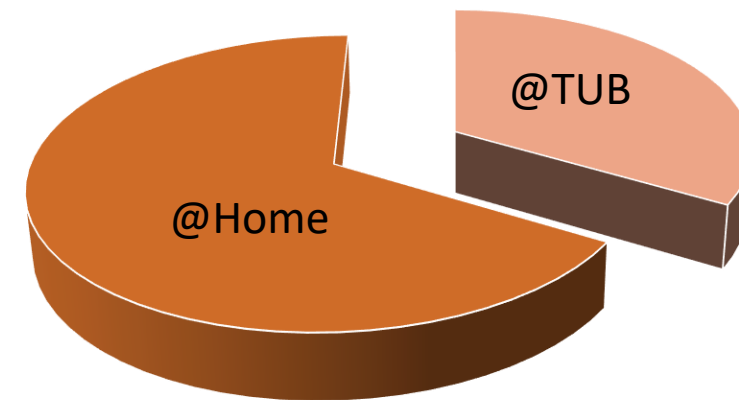
Grading: graded	Type of exam: Oral exam	Language: English	Duration/Extent: 30 Minutes
--------------------	----------------------------	----------------------	--------------------------------

- Lectures (45%)
 - Lectures in presence in H 6131
- Exercises (45%)
 - Exercises in presence in H 6131 (H 6134) with programming homework
- Independent reading (10%)
 - Book chapters, scientific papers, etc.

- Our PC pool is too small (19 workplaces) to effectively conduct hands-on exercises there (with the expected student group size)
- Format:
 - Presentation of the exercise tasks
 - Review of the necessary (Python) programming skills
 - Introduction to the used programming libraries
 - Overview and discussion of the possible solutions with the students
 - Exercise notebooks (in Jupyter format) for working at home
 - Discussions on the solutions in the following week

Workload

- 6 CP → 180 academic hours
- 4 SWS → 4 academic hours per week by attendance in class
→ 8 academic hours per week for homework and exam preparation
- 1 academic hour → 45 real minutes
180 academic hours → 135 real hours



Access to JupyterHub

- Send an email with the topic “exercise server” to the following email address asking for access to JupyterHub providing us your **family name, first name, matriculation number, TU Berlin email address, possibly TU Berlin user name, and the course(s) you want to use the server for**

andreas.fuls@tu-berlin.de

- **Please do not send us your password!!!**
- Access will be granted as soon as possible (within office hours), but it may from time to time also take a few days

Prerequisites

- No prerequisites w. r. t. previous courses and knowledge
- Helpful are
 - Basic knowledge of programming and Python
 - Some familiarity with geo data

- Oral examination:
 - 30 minutes @ Kaiserin-Augusta-Allee 104-106, TU Berlin
 - Covering the content of the lectures and the exercises
 - Please note that there is also a lot of course content given within the exercises
- Exam dates (subject to changes):
 - 22.07.2024 (early exam period for students who need to quickly finish the module)
 - 05.08.2023 – 09.08.2024 (first regular exam period)
 - 30.09.2023 – 02.10.2024 (second regular exam period)
 - Exam dates (the particular day within the exam period) and times are assigned per exam period in the order of registration, we cannot consider individual preferences
 - Best to keep your schedule flexible during exam week in case of last-minute changes

Registration & De-Registration

- Registration:
 1. MOSES (once the registration is activated)
 2. You get an email from almut.gothe@tu-berlin.de with the day and time of your exam (typically) within the next seven days
 3. **You need to confirm your examination time slot by replying to this email**
 - Please register at least 2 weeks prior to the exam period, as otherwise the time slots may be filled up (first-come-first-serve principle)
- De-Registration:
 - You can de-register from an exam without providing a reason up to three days before the oral examination

Withdrawal from Examination

- Withdrawal up to the actual day of the examination:
 - Due to illness or other (*relevant*) reason
 - Withdrawal must be submitted to the examiner and the examination office on the day of the examination before the beginning of the exam at the latest
 - See the TU Berlin website for further information:
<https://www.tu.berlin/en/pruefungen/mounts/examinations/module-examinations/>
 - The examiner reports "did not appear" to the grading system and the examination office clears the examination attempt upon receipt of the required documents that provide proof of the reason

Repetition of Exams

- In case you failed the exam, you can repeat it up to two times in the next announced exam period for oral examinations
- There is generally no second attempt to improve a grade

Tips for Your Learning Success

- If you do not understand something within the lectures or exercises or just need additional information:
 - Ask right away within the lecture or exercise!
 - Otherwise, please consider to do some research on the internet. There are fantastic online resources (blogs, tutorials, lectures, talks, code, etc.) on the subject of data science and machine learning
- Our philosophy is that after the foundation modules, it is reasonable to expect that students in a master's program should also be able to learn certain topics on their own, of course with some guidance

Tips for Your Learning Success

- Take the exercises seriously, they help in understanding the topics and for being well prepared for the oral examination as well as the master thesis
 - Some content is only taught within the exercises and is expected to be known during the examination

What comes next?

- In the study program Geodesy and Geoinformation Science:
 - Project Geoinformatics
 - Deep Learning for Geographical Data
 - Master's thesis in the subject area of Geoinformation Technology (GIT)
- Machine and Deep Learning topics are always welcome in these modules

Consultation Hours

- Consultation hours by appointment (via email)
 - Write an email asking for an appointment
 - We typically reply within 3 to 5 working days (but could also take longer)
 - Depending on our schedule, you will be given an appointment for the following week or the week after

- (Geo) data science is an exciting and up-and-coming area of study, for which you will find plenty of information to learn from that helps you to better understand the topics, and with a lot of further materials to study and try out
- Become active yourself and immerse yourself in the topic, you will not only gain a better understanding of geo data science, but also learn how to master new subject fields, which is an invaluable skill for your future professional life

Thank you for your attention!