

Contact Information

Prof. Dr. Tillmann Schwörer

Professor of Applied Data Science Head of Study Program Data Science

Fachhochschule Kiel

University of Applied Sciences

Study Program Data Science

Grenzstraße 3

D-24149 Kiel

Phone: +49 431 210 - 2541

Email: tillmann.schwoerer@fh-kiel.de

Office: C12 - 2.16

How to best reach me?

You have questions, problems, issues that you would like to discuss with me? Related to this lecture, or related to this study program?

- Just ask me during our regular lecture times
- ▶ I prefer if we discuss it face to face, e.g. in the break or right after the lecture
- ▶ If it deserves more time and thought, we can still schedule a separate (Zoom) meeting or correspond by email

Course website

Course website "Tools and Programming Languages for Data Science" on Moodle Lehre: https://learn.fh-kiel.de/course/view.php?id=14335

- Contents will be uploaded during the semester
- **Course materials**: slides, data, scripts, notebooks, etc.
- Examination-related aspects
 - Information about examination tasks: dates, data, task sheets
 - Submission of your solutions

Announcements

- e.g. changes regarding lecture date or location
- e.g. relevant information / updates related to exam questions
- e.g. important information on short notice
- You will receive these announcements also to your student email address

Lecture Date and Location

- Time and place
 - Monday, 12:45 4:00 PM in O15-3.47
 - Wednesday, 12:45 4:00 PM in O15-3.47
- Lecture Period
 - 12 dates in the period 15.09. 27.10.2024
- ▶ The lectures of this course will terminate after the first half of the semester
- ▶ There are examination tasks both during and after the end of the lecture period
- ► The course will be intense and may require significant time investments
 - regular lecture times
 - work on problemsets and projects
 - self-study time

Workload

Number of SWS 6 4 SWS

Credits 5,0 Credits

Contact hours 6 48 Hours

Self study **1**02 Hours

Basic Python programming skills are recommended. Students with little or no Python programming skills are strongly advised to participate at the Pre-Course Programming, taking place in the week prior to the start of the regular programming course.

See Modul Description

Prerequisites of this course

- ➤ A proper installation of Python and a suitable IDE and basic python knowledge are an important prerequisite for this course
- We covered and practiced these fundamental aspects in the Precourse Programming last week
- Previous knowledge of specific Python data science packages is not needed

Precourse Programming

If you have missed the Precourse Programming, make sure that you are familiar with its contents:

- Installation of Python and Visual Studio Code
- Package installation: pip, virtual environments, requirements files
- Primitive types
- Collections: list, tuple, dict, set
- Functions
- Control flow: for, while, if-else statement
- List and dictionary comprehensions
- Important Iterators: range, zip, enumerate, ...
- Some Python coding conventions: indentation, naming, docstrings, type hints, ...

You find a link to the Moodle page of the Precourse Programming on our regular "Tools and Programming for Data Science" Moodle page.

Zoom

- We use Zoom meetings ocassionally in the lectures to easily share the screen and discuss code examples and solutions from students
- ► In the first 2 weeks of the semester, lectures will be streamed via Zoom to smoothen the study start
 - for those students who cannot yet be in Kiel due to Visa issues or similar reasons
- Zoom Meeting Link: see course Moodle page

Zoom rules

Lecture streaming is a special offer that you should not take for granted. Please respect the following zoom rules:

- Join the lecture on time
- Turn on your video for a more engaging environment, if possible
- Do not record or share the lectures
- Mute yourself when your not contributing
- Inform me (ideally by voice) if I forget to share screen
- If you must leave the lecture early, inform me beforehand

Literature

▶ The main materials are the ones uploaded to the course Moodle page

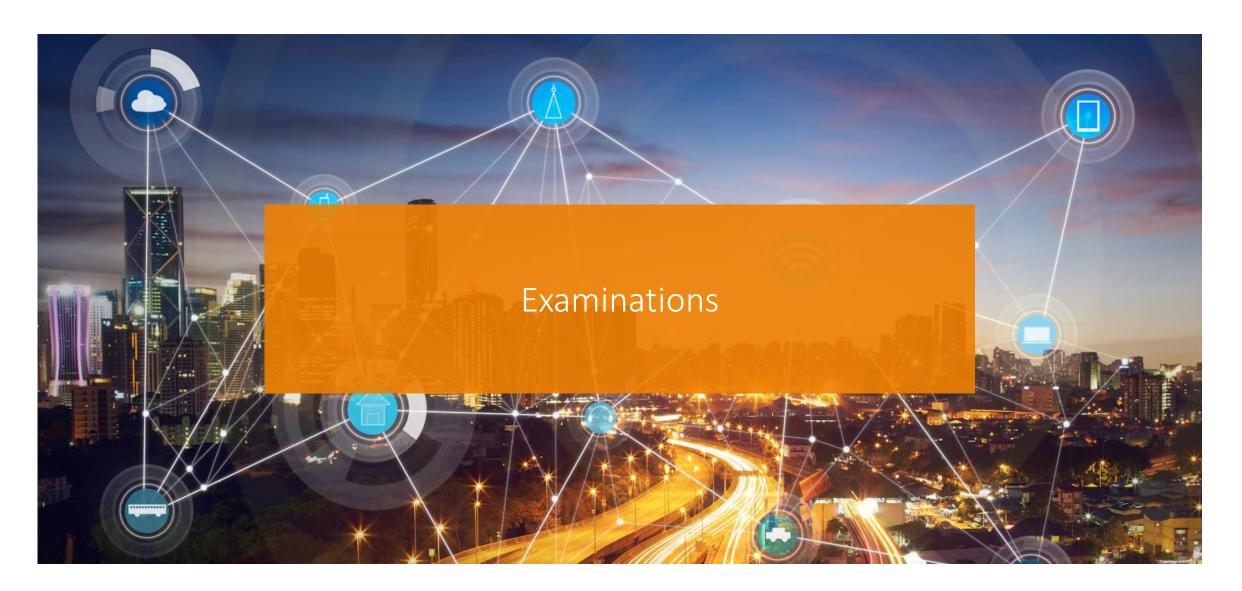
Additional literature

VanderPlas: Python Data Science Handbook. O'Reilly, first edition.

For Python foundations see:

- VanderPlas: A Whirlwind Tour of Python. O'Reilly, first edition.
- Precourse Programming





Examination registration

- ▶ Registration period for examinations at the Faculty of Media: 01.11. 15.11.2024
 - After this period you won't be able to to register for (or unregister from)
 examinations
 - Examination website: <u>qis.fh-kiel.de</u>
 - ✓ Registration
 - ✓ Grades
- ► For formal exam-related questions consult:
 - <u>Examination office of Faculty of Media</u> (Sandra Kalwis, or Antonia Weckwerth)
 - Prüfungsordnung Data Science (so far, German only)
 - Prüfungsverfahrensordnung Kiel UAS (so far, German only)

Portfolio examination: task overview

▶ The examination consists of a Portfolio of multiple tasks

Task	Dates	Weight
Project 1	Submission date: 13.10.2024	30 %
Project 2	Submission date: 27.10.2024	30 %
Programming Test	Date: 18.11.2024 12:45 PM	40 %

Note:

- I highly value active participation and contributions to the lecture
- Outstanding participation during the semester will be honored with bonus points

Portfolio examination: Projects

The 2 Projects test your ability to combine your knowledge from different topics of the course to solve typical Data Science programing problems in a close to real-life scenario

- You have sufficient time to solve it
- You don't need to know or memorize everything → you may look up information and investigate
- ▶ The task will be more comprehensive and difficult compared to regular problemsets and the programming test. The approach won't be straightforward \rightarrow you may need to decide between different alternatives and justify your decisions
- ► For a good grade I expect high quality solutions, a clean and well structured submission, pythonic code, well formulated answers

Portfolio examination: Projects

Grading

- Completeness, correctness and quality of solutions
- Clean and well formatted Jupyter Notebooks
- Pythonic code (following Python conventions and best practices)
- Timely submissions

Other aspects

- ◆ The projects are NO team exercises → solutions must be your own ones
- When the project tasks are published, this will be announced in class and on Moodle
- Part of the project work can be done during the lecture times (I will notify you when this is the case). The other parts should be done at home

Portfolio examination: Programming Test

The Programming Test tests your ability to **solve typical programming tasks** of a Data Scientist **on your own** and in a **limited amount of time**

- Questions will require in particular your expertise in data reading, data cleaning, data analysis using the Pandas library and basic data visualization
- ► The full scope of topics covered in the Programming Test will be communicated beforehand
- The questions will be easier and straightforward to solve compared to the Projects
- ▶ The main focus is on the correctness/suitability of your Python code and explanations
- In general, I do not judge the efficiency and conciseness of your code. However, I will deduct points if your code is chaotic and non-Pythonic

Portfolio examination: Programming Test

- Duration: 1 hour
- You are allowed
 - to use lecture materials, books, cheatsheets, function documentation, etc. either stored on your laptop or written on paper
- You are NOT allowed
 - to use Al-assistance: Chat GPT, Github Copilot, or similar
 - to use the internet
 - ✓ you will be asked to turn off WiFi before the beginning of the test

Any attempt to cheat will result in a score of 0 points and a report to the examination office

Grading scheme

Erreichte	Numerische	Deutsche	Englische	Definition
Leistung in	Bewertung	Bezeichnung	Bezeichnung	
Prozent				
(NMP) *)				
≥ 95	1,0	Sehr gut	very good	Eine hervorragende
				Leistung
≥ 90	1,3	"	п	п
≥ 85	1,7	Gut	good	Eine Leistung, die
				erheblich über den
				durchschnittlichen
				Anforderungen liegt
≥ 80	2,0	"	п	п
≥ 75	2,3	11	п	п
≥ 70	2,7	Befriedigend	satisfactory	Eine Leistung, die
				durchschnittlichen
				Anforderungen genügt
≥ 65	3,0	"	п	п
≥ 60	3,3	11	"	п
≥ 55	3,7	Ausreichend	pass	Eine Leistung, die trotz
				ihrer Mängel noch den
				Anforderungen genügt
≥ 50	4,0	"	11	п

Asking questions

- Sometimes tasks are not entirely clear and can be interpreted in different ways. If this is the case, please ask so that I can clarify. In general, if a task can be interpreted in different ways, than different solutions will be counted as correct.
- If it is relevant to all students, I will inform all students
- You may also ask me during the classes, because I am here to help. However, very likely you will only receive hints.

ChatGPT and other Al assistance

- ► Al-tools such as ChatGPT and Github Copilot ...
 - write code
 - explain code
 - find errors
 - ... in surprising quality and speed
- ▶ Am I allowed to use AI-tools?
 - For the programming test: No
 - Otherwise: Yes, the use of AI assistance is allowed

But use it wisely !!!

ChatGPT and other Al assistance

- ▶ Use AI tools to enhance your own understanding and become a better programmer
 - Verify that you understand topics on your own
 - Challenge yourself to solve programming problems on your own
 - Compare your solutions with AI-based suggestions and learn from the differences
 - Use AI tools as sparring partners to get explanations and fill knowledge gaps
 - Don't take correctness of AI suggestions for granted
 - Consult complementary help from real persons (students, professors, stackoverflow, etc.) and documentation
- ▶ It is your programming skills not your prompting skills that will have the largest influence on your course grade.

Chat GPT Bullshit Bingo

Best of Chat GPT Bullshit Bingo Winter Semester 2023/24:

"In this project, we embarked on a journey to dissect ..."

"Employing advanced text mining techniques I embark on topic modeling endeavors..."

"This study embarked on an exploratory journey into the world of ..."

"Moreover, exploration of the unique words and patterns embedded in the language of hate speech will be embarked on to better understand this..."

"Let me take you on an exciting journey I've embarked on ... So, I rolled up my sleeves and plunged into this vast ocean of ... I'm using some cool data science tricks to make sense of it all."



Data Scientist embarking on exciting endeavors







Brainstorming "Data Science"

How would you define data science?

Brainstorming "Data Science"

Data:

- Describe data sets that you have worked on yourself before?
- What are typical data sets and data sources in your domain?
- What are characteristics of these data?

Data Science Goals/Problems/Products

- Describe a data science problem that you are familar with or have addressed on your own
- What are other goals or problems that a data scientist addresses?
- How does the output or product a a data scientist look like?

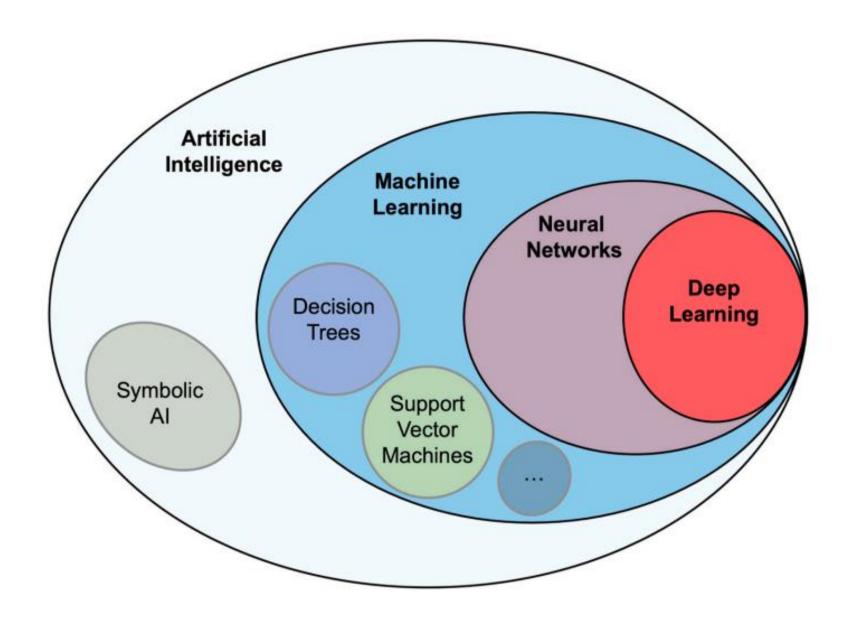
Technologies/Processes/Algorithms

- How do data scientists reach their goals or solve their problems?
- Which are the steps or processes involved in a data scientist work?
- Which are relevant technologies or algorithms?

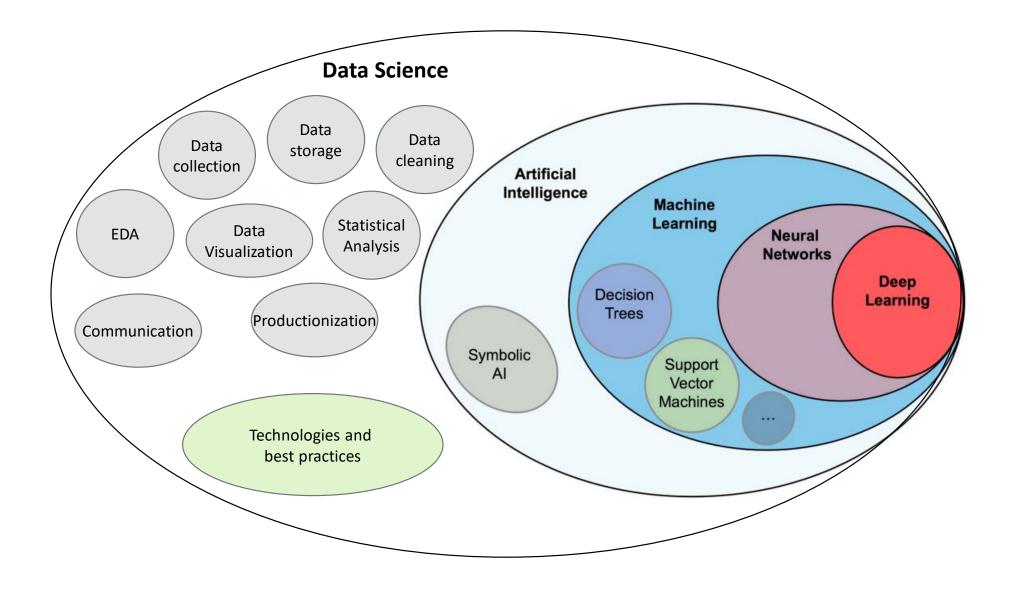
A data science definition

Data Science combines methods, processes and technologies from the field of mathematics, statistics and computer science with domain knowledge to extract insights from data with the aim to solve real-world problems and inform decision-making.

How is data science related to artificial intelligence?

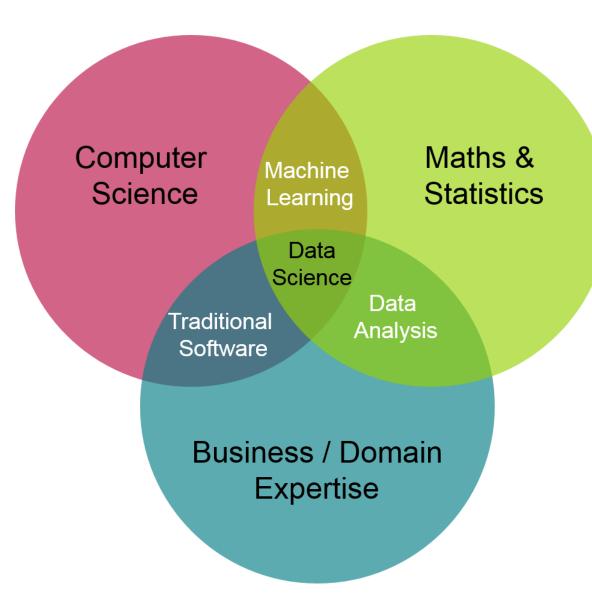


How is data science related to artificial intelligence?



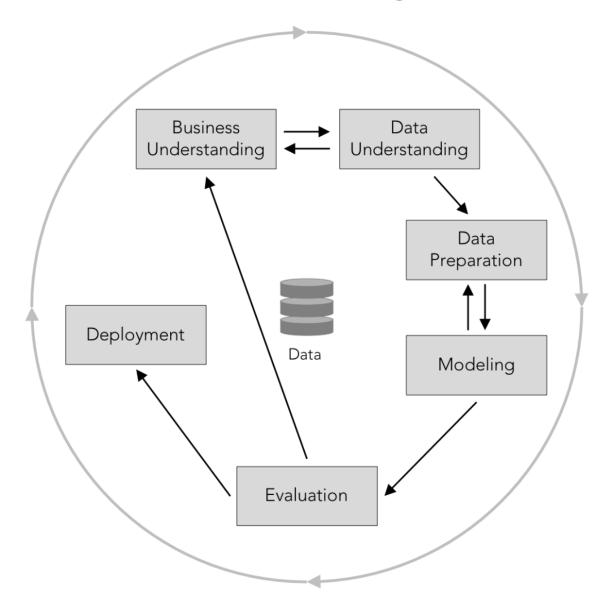
Data science skill set

- You are coming from different backgrounds and thus have different sets of previous expertise
- Recommendations:
 - Identify where your knowledge gaps and invest time to fill these gaps
 - Apply data science in the domain where you are an expert
- ► There are different roles with different focus in the job market: data analyst, data engineer, data scientist, machine learning engineer, data science architect, ...

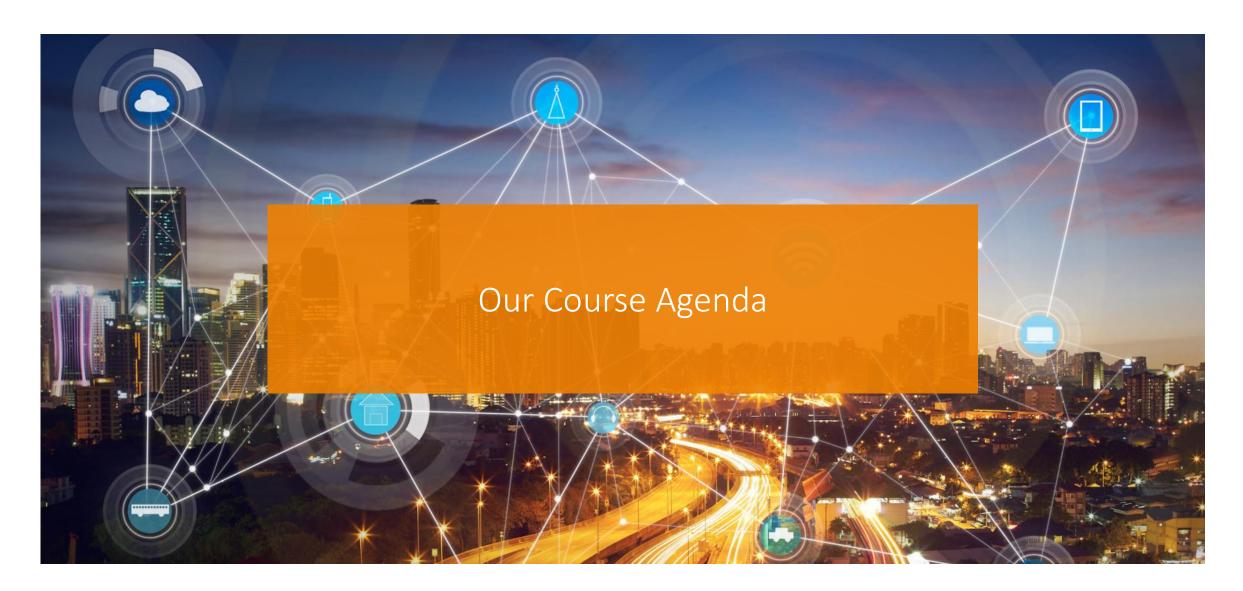


CRISP-DM

Cross-Industry Standard Process for Data Mining







Our course agenda

- Introduction and overview
- NumPy: Basic data handling with Numpy arrays
- Pandas
 - Exploratory data analysis
 - Data consolidation
 - Data cleaning
- Data visualization using Matplotlib and Seaborn
- Interacting with APIs
- Interacting with SQL databases
- Version Control with Git and GitHub
- Advanced Python

Python foundations



Operators

Functions

Control flow and iterators

Programming concepts & paradigms

See also Precourse Programming

Tooling

Installation

Visual Studio Code

Jupyter Notebooks

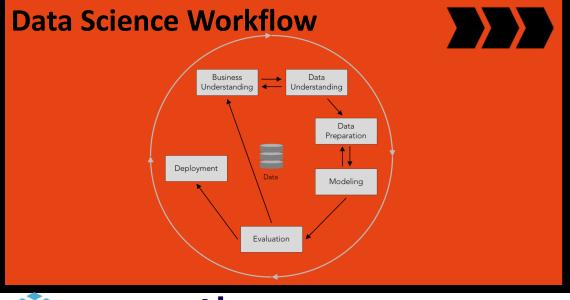
Packages

Virtual Environments

Git and Github



Python









Why Programming?

What are the benefits of being able to program?

- ▶ Better understand algorithms and scientific methods
- Have full control over each step
- Have access to latest scientific methods
- Build entire data products
- Make yourself independent of proprietary tools and services
- Make less errors
- Make projects reproducible and portable

Why Python?

What are the benefits of learning Python in the context of data science?

- Standard language in Data Science
- Large number of specialised data science libraries
- Access to state-of-the-art algorithms
- Open source and free of charge
- Large community
- Integration with databases
- Focus on readibility
- Easy to learn
- Job opportunities