

# 1. Introduction

Practical Data Science (PDS)



Lecture: **Prof. Dr. Gunther Gust**



- Inhaber des Lehrstuhls für Wirtschaftsinformatik und Künstliche Intelligenz im Unternehmen (seit 2022)
- Forschung und Lehre über die Anwendung von KI in Unternehmen und Gesellschaft
  - Smart Grids
  - Smart Mobility
  - Smart Cities & Urban Analytics
- Ausbildung: KIT, ITESM Monterrey, Uni Freiburg, Lawrence Berkeley National Laboratory
- Stipendiat Stusti & Heinrich Böll Stiftung

Exercise: **Viet Nguyen**



- Computer Science and Engineering (Bachelor) at the Frankfurt University of Applied Sciences and the Vietnamese-German University
- Data Engineering and Analytics (Master) at the Technical University of Munich (graduated with distinction)
- Data Science Intern: Knorex, Ho Chi Minh, Vietnam
- Software Engineer: mesoneer, Ho Chi Minh, Vietnam



**This course introduces data science tools, as well as advanced methodologies in an applied manner**

SPOTLIGHT ON BIG DATA

Spotlight

ARTWORK Tamar Cohen, Andrew J Buboltz  
2011, silk screen on a page from a high school  
yearbook, 8.5" x 12"

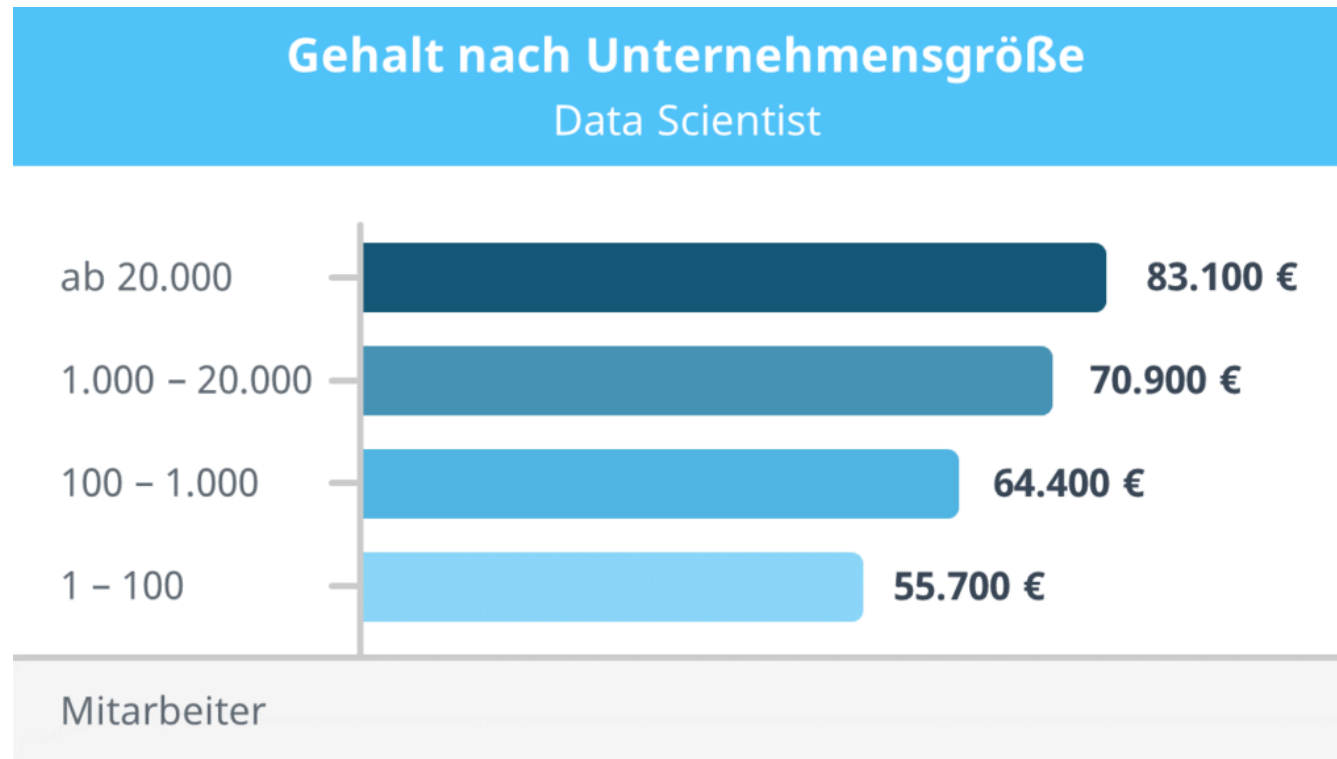
# Data Scientist:

## *The Sexiest Job of the 21st Century*

**Meet the people who  
can coax treasure out of  
messy, unstructured data.**  
by Thomas H. Davenport  
and D.J. Patil



# It prepares you for promising positions in practice



## Learning objectives for this course

- The foundations, **frameworks and applications** of the emerging field of data science
- Design, implement, and evaluate the core algorithms underlying an **end-to-end data science workflow**, including data import, analysis, and presentation of information
- Leverage the **Python** application programming interface (API) ecosystem and data infrastructure that supports data acquisition, storage, retrieval and analysis
- The application of a data-based **analytical approach** to identify and solve problems
- Implementation and execution skills for **data-driven business analytics**

# Course Schedule

Oct – Dec











1. Introduction, Jupyter Notebooks, Version Control (**today**)
2. Descriptive Analytics
3. Machine Learning Introduction
4. Feature Engineering
5. Deep Learning on Tabular Data
6. Deep Learning for Computer Vision
7. Deep Learning for Image Segmentation, Geocoding & Capstone Project
8. Data to Production
9. Deep Learning for Natural Language Processing
10. NLP with Huggingface

---

Jan & Feb

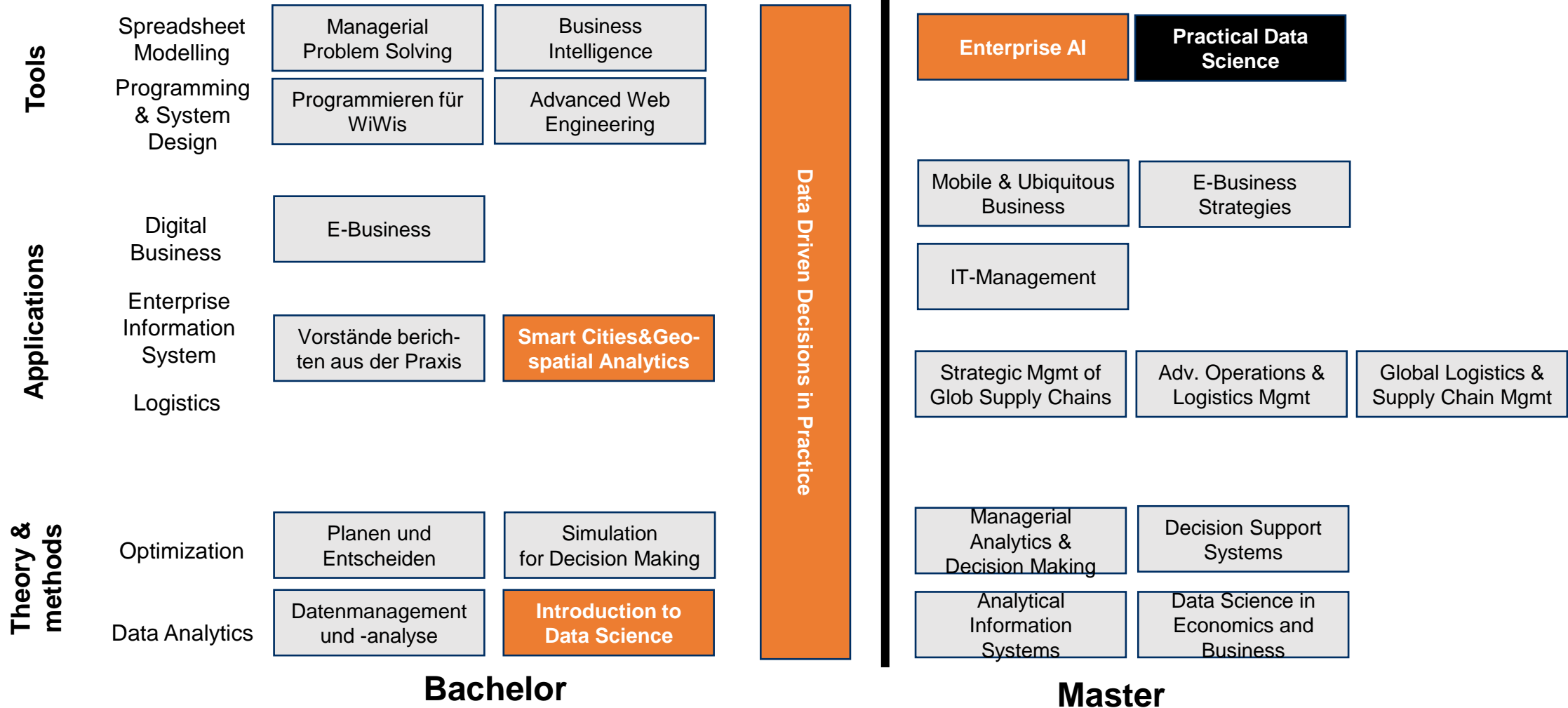
11. Guest lecture
12. Group Work on Capstone Project
13. Final presentations (End of semester)

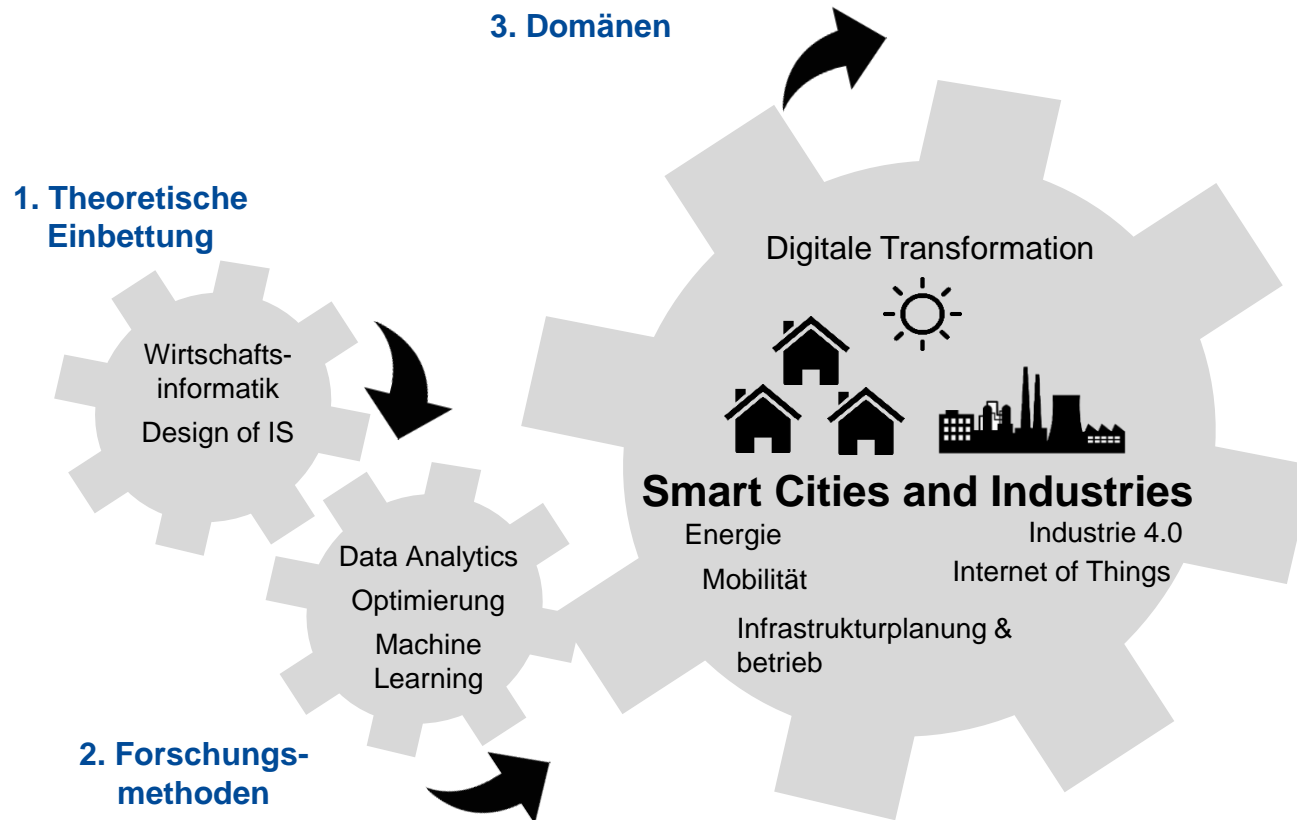
# Why Python?

Oct 2021	Oct 2020	Change	Programming Language		Ratings	Change
1	3	↑		Python	11.27%	-0.00%
2	1	↓		C	11.16%	-5.79%
3	2	↓		Java	10.46%	-2.11%
4	4			C++	7.50%	+0.57%
5	5			C#	5.26%	+1.10%
6	6			Visual Basic	5.24%	+1.27%
7	7			JavaScript	2.19%	+0.05%
8	10	↑		SQL	2.17%	+0.61%
9	8	↓		PHP	2.10%	+0.01%
10	17	↑↑		Assembly language	2.06%	+0.99%



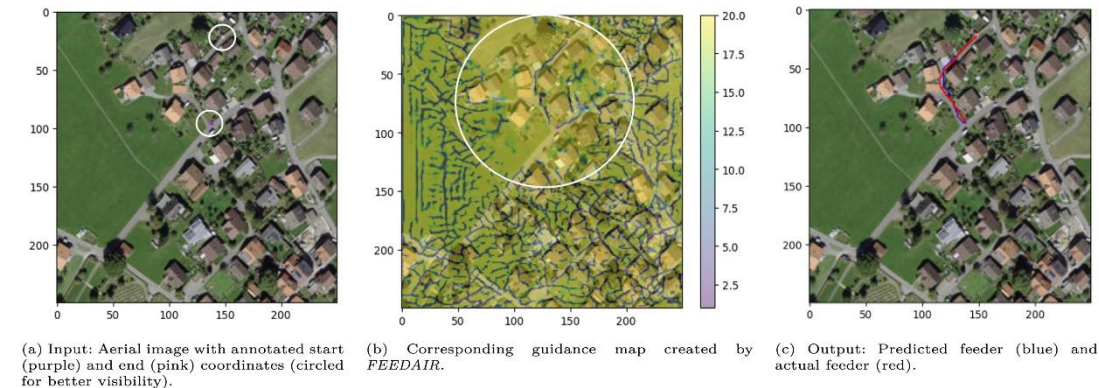
# The course relates to several other courses of the D3 teaching program





## Aktuelle Projekte und Themen (Auswahl)

- Computer vision-based planning of electricity networks
- Dynamic electricity pricing & load flexibility
- Earthquake monitoring based on spatial sensor data and graph-convolutional neural networks
- Mixed autonomous urban mobility systems
- Multi-modal urban analytics



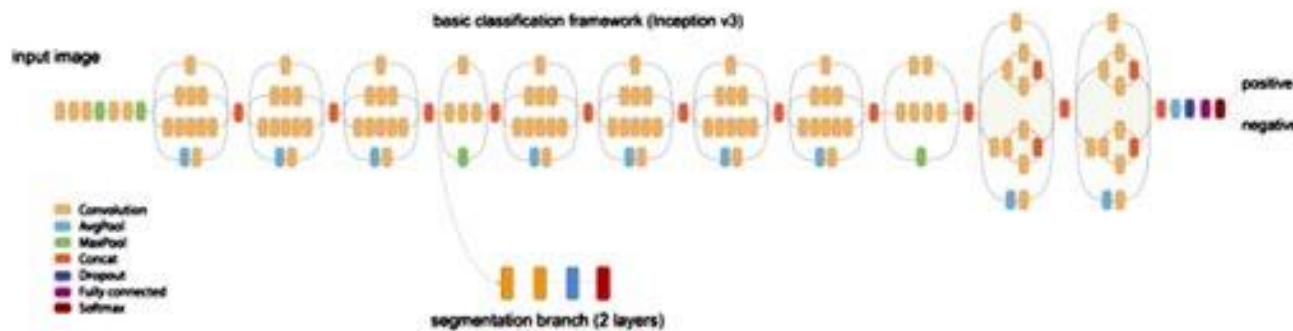
## Deep Solar

### Objectives

- Detect solar PV systems based on areal images
- Geocode the PV systems and estimate their size

### Results

- First exhaustive data set containing the geolocation of PV systems for large regions of Germany
- The data set serves as the basis for further analyses about the **adoption of PV systems**
- In a follow-up project we plan to combine the PV data set with 3D city models



# What can we learn from aerial images about residential electricity consumption?

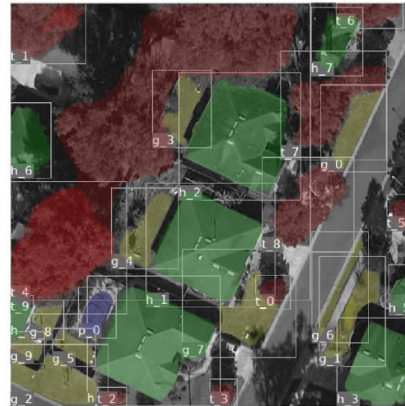
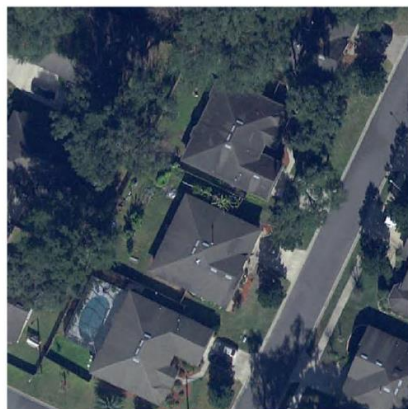
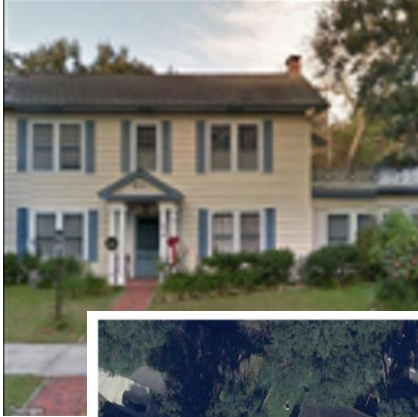
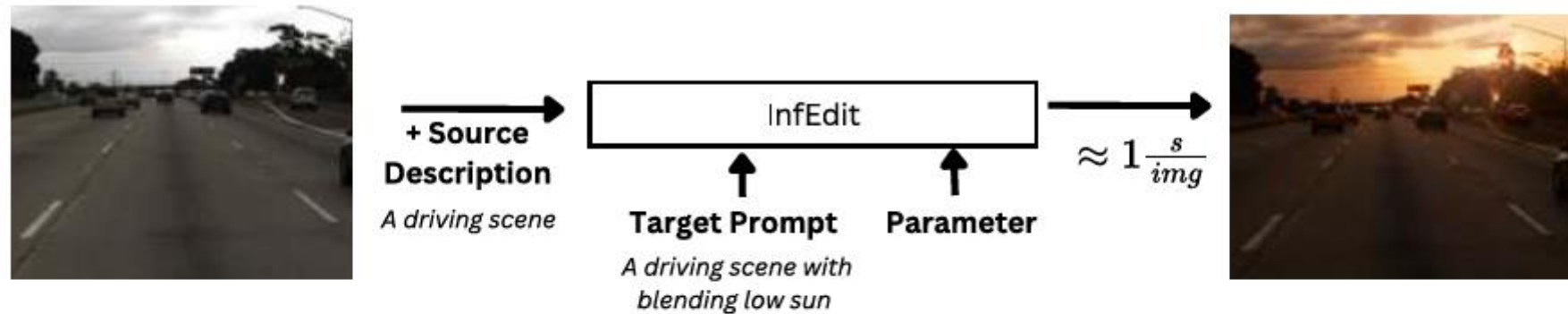


Image-based model	
Bathrooms	x
Bedrooms	x
Building compactness	2.0%
Building footprint	14.3%
Building perimeter	0.0%
Building size	12.1%
Building stories	3.9%
Construction year (actual)	0.0%
Construction year (effective)	2.3%
Exterior wall material	x
Floor material	x
Heating system	x
Interior wall material	x
Land cover (grass)	0.0%
Land cover (tree)	27.8%
NDVI	x
Parcel size	x
Pool size	37.6%
Zip Code	x

ROSENFELDER M, WUSSOW M, GUST G, CREMADES R, NEUMANN D (2021): Predicting Residential Electricity Consumption Using Aerial and Street View Images. Applied Energy.



# Using generative AI to create artificial training images for autonomous driving



**Figure 7:** Generative Image Augmentations with InfEdit



## Ways to work with us

- **Hiwi jobs** (Bachelor's and Master's level)
  - Women are particularly encouraged to apply (see also Women@WIWI program)
- **Seminar thesis**
  - Bachelor seminar
  - Master seminar
  - Project seminar
- **Bachelor's thesis**
- **Master's thesis**
  - Scholarships available
- **Doctorate**
  - Applications are accepted on an on-going basis

- **Lecture**

- Mondays 14:15 – 15:45 (approx.)
- Theoretical foundations
- Implementation examples using Python

- **Exercise**

- Directly after lecture
- Weekly assignments to be solved on your own
  - Published after lecture
  - Due 1.5 weeks later
  - Complete 4 out of 5 assignments to be eligible for final project
  - Meaningful participation in the WueCampus forum (or Discord Server) **before Nov 30** counts as 1 assignment

- **Bonus** (0,3 bonus for the exam)

- Complete 5 out of 5 assignments

- **Final project**

- Work on a real-world problem set
- Groups of 3 students