

Disziplin-übergreifendes Seminar (03-M-GS-13)

Forschungsdatenmanagement und Data Science

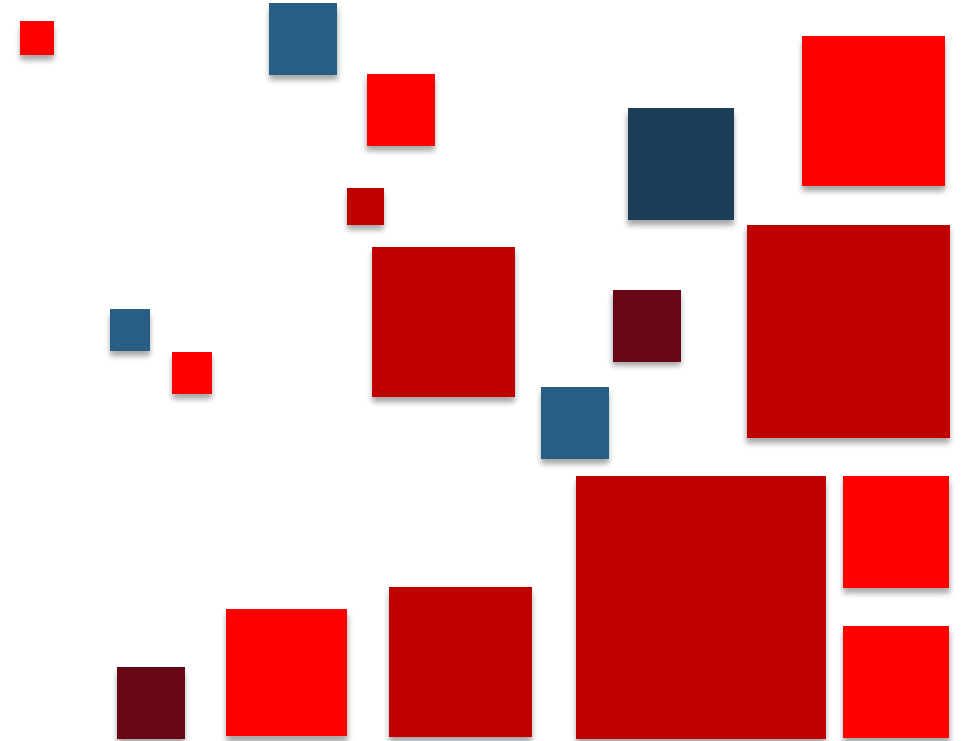
Prof. Dr. Iris Pigeot

Prof. Dr. Frank Oliver Glöckner

Prof. Dr. Rolf Drechsler

Dr. Tanja Hörner

et al.



**UBremen
Research
Alliance**

Data Train

Training in Research Data Management
and Data Science

Termin	Thema	Expert*in
04.11.2020	Einführungsvortrag	Prof. Dr. Iris Pigeot
11.11.2020	Anwendungsbeispiel aus der freien Wirtschaft	Sonja Sievi, Georg Anspach von Broecker
18.11.2020	Daten- und Informationsmanagement	Prof. Dr. Frank Oliver Glöckner, Dr. Ivaylo Kostadinov
25.11.2020	Datenrecht – Datenschutz und Lizenzen	Prof. Dr. Benedikt Buchner
03.12.2020	Data Engineering	Prof. Dr. Adalbert Wilhelm
09.12.2020	Datenbanken und Big Data Management	Prof. Dr. Sebastian Maneth
16.12.2020	Management von Software Entwicklungen	Prof. Dr. Rainer Koschke
06.01.2021	Erkenntnisse in der digitalen Welt	Prof. Dr. Dr. Norman Sieroka
13.01.2021	Grundlagen der Informatik für Data Science	Prof. Dr. Rolf Drechsler
20.01.2021	Data Science Grundlagen	Dr. Juliane Jarke, Dr. Hendrik Heuer
29.01.2021	Programmiersprachen	Prof. Dr. Christoph Lüth
03.02.2021	Künstliche Intelligenz	Prof. Dr. Michael Beetz
10.02.2021	Maschinelles Lernen	Dr. Marvin Wright
17.02.2021	Datenvisualisierung	Dr. Georgia Albuquerque

Veranstalter*innen

Prof Dr. Iris Pigeot



Professorin für Statistik an der Universität Bremen

Direktorin des Leibniz-Instituts für Präventionsforschung und Epidemiologie – BIPS

Stv. Vorsitzende des U Bremen Research Alliance e.V.

Prof Dr. Frank Oliver Glöckner



Professor für Erdsystem Datenwissenschaften an der Universität Bremen

Leiter PANGAEA (MARUM) und des Datenzentrums am Alfred-Wegener-Institut (AWI)

Außerord. Prof. für Bioinformatik an der Jacobs University

Prof Dr. Rolf Drechsler



Professor für Rechnerarchitektur an der Universität Bremen

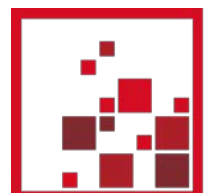
Direktor der Cyber-Physical Systems Arbeitsgruppe am Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI)

Sprecher des Data Science Centers der Universität Bremen (DSC@UB)

Dr. Tanja Hörner



Koordinatorin der disziplinübergreifenden Doktorand*innen Ausbildung „Data Train – Training in Research Data Management and Data Science“ der U Bremen Research Alliance



**UBremen
Research
Alliance**

Data Train

Training in Research Data Management
and Data Science

Veranstaltungskonzept

Expert*innen Team FDM



Softwaremanagement
Prof. Dr. Rainer Koschke
Universität Bremen, FB03



Erkenntnisse in der digitalen Welt
Prof. Dr. Norman Sieroka
Universität Bremen, FB09



Datenbanken und Big Data
Prof. Dr. Sebastian Maneth
Universität Bremen, FB03



Datenrecht/ Datenschutz/ Lizenzen
Prof. Dr. Benedikt Buchner
Universität Bremen, FB06



Data Engineering
Prof. Dr. Adalbert Wilhelm
Jacobs University



Daten- u. Informationsmanagement
Prof. Dr. Frank Oliver Glöckner, AWI/PANGAEA, UB
Dr. Ivaylo Kostadinov, GFBio e.V.



**UBremen
Research
Alliance**

Data Train
Training in Research Data Management
and Data Science

Veranstaltungskonzept



Visualisierung

Dr. Georgia Albuquerque
Deutsches Zentrum für Luft
und Raumfahrt



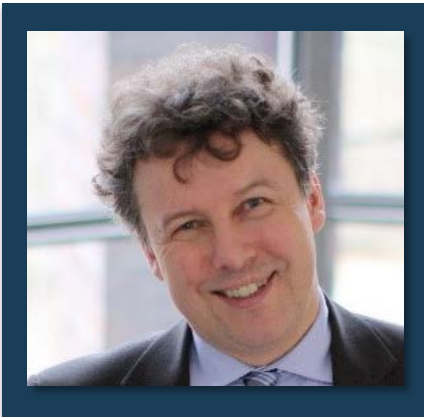
Maschinelles Lernen

Dr. Marvin Wright
Leibniz-Institut für Präventionsforschung und
Epidemiologie – BIPS



Data Science Grundlagen

Dr. Juliane Jarke, UB (FB03), ifib
Dr. Hendrik Heuer, UB (FB03), ifib



Grundlagen der Informatik

Prof. Dr. Rolf Drechsler
Universität Bremen, FB03
Data Science Center



Programmiersprachen

Prof. Dr. Christoph Lüth
Deutsches Zentrum für Künstliche Intelligenz,
Universität Bremen, FB03



Künstliche Intelligenz

Prof. Dr. Michael Beetz
Universität Bremen, FB03, IAI

Veranstaltungskonzept



**UBremen
Research
Alliance**

Data Train

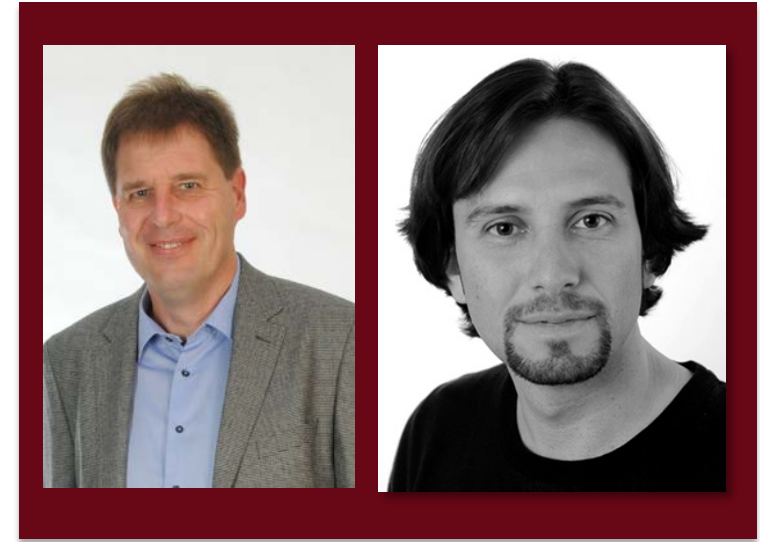
Training in Research Data Management
and Data Science

FDM1: Daten- und Informationsmanagement

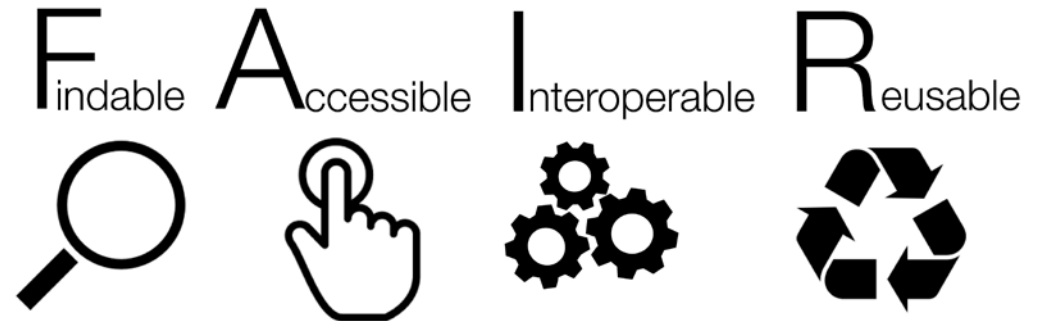
Dozenten:

Ivaylo Kostadinov (ikostadi@gfbio.org),

Frank Oliver Glöckner (fog@uni-bremen.de)



- Forschungsdaten: Lebenszyklus & Herausforderungen
- Datenmanagementpläne
- FAIR Data Prinzipien
- Metadaten: Signifikanz & Standardisierung
- Archivierung, Publikation, Zitation



**UBremen
Research
Alliance**

Data Train

Training in Research Data Management
and Data Science

Themenvorstellung

FDM 2: Datenrecht –Datenschutz und Lizenzen

Dozent:

Benedikt Buchner (bbuchner@uni-bremen.de)



- Rechtliche Rahmenbedingungen des FDM:
Wem „gehören“ Daten?
- Der Schutz personenbezogener Forschungsdaten
 - Datenschutz und Forschung
 - Einwilligung
- Der urheberrechtliche Schutz von Forschungsdaten
 - Urheberrecht und Datenbankschutz
 - Lizenzen



**UBremen
Research
Alliance**

Data Train

Training in Research Data Management
and Data Science

Themenvorstellung

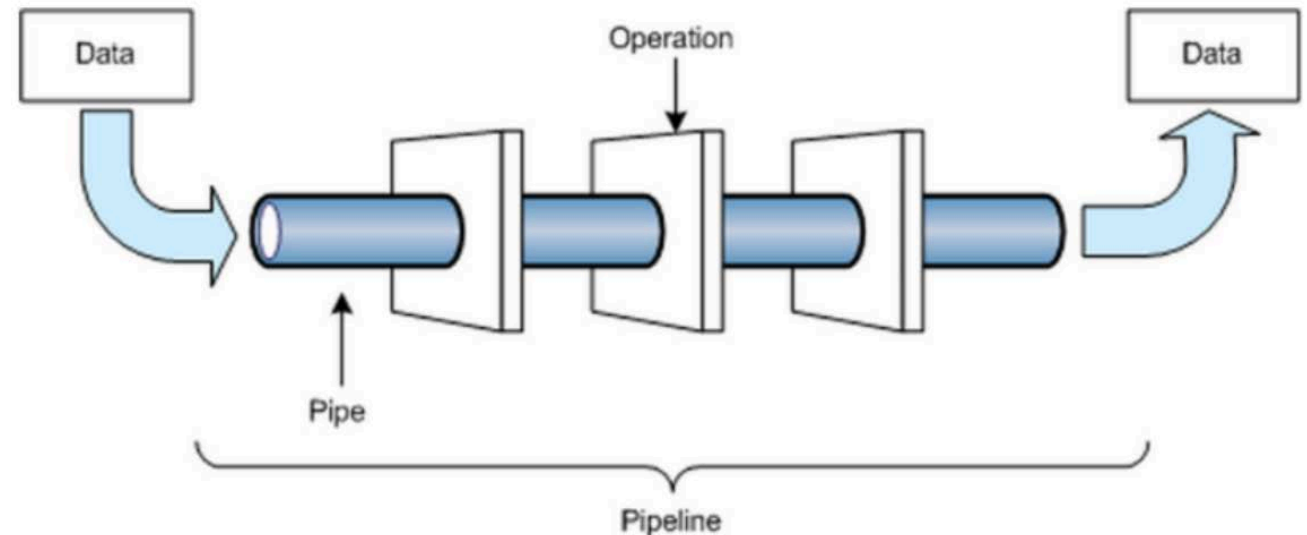
FDM 3: Data Engineering

Dozent:

Adalbert F.X. Wilhelm (a.wilhelm@jacobs-university.de)



- Data Engineering
 - Rolle und Aufgaben
- Data Pipelines
- Extract, Transform, Load (ETL) - Pipeline
- Daten- und Prozessarchitektur
- Datentransformation
 - Feature engineering
- Anwendungsbeispiele:
 - Human Activity Recognition
 - Biomedical Text Data



Quelle: <https://bit.ly/2HZp2Pn>

Themenvorstellung



U Bremen
Research
Alliance

Data Train

Training in Research Data Management
and Data Science

FDM 4: Datenbanken und Big Data Management

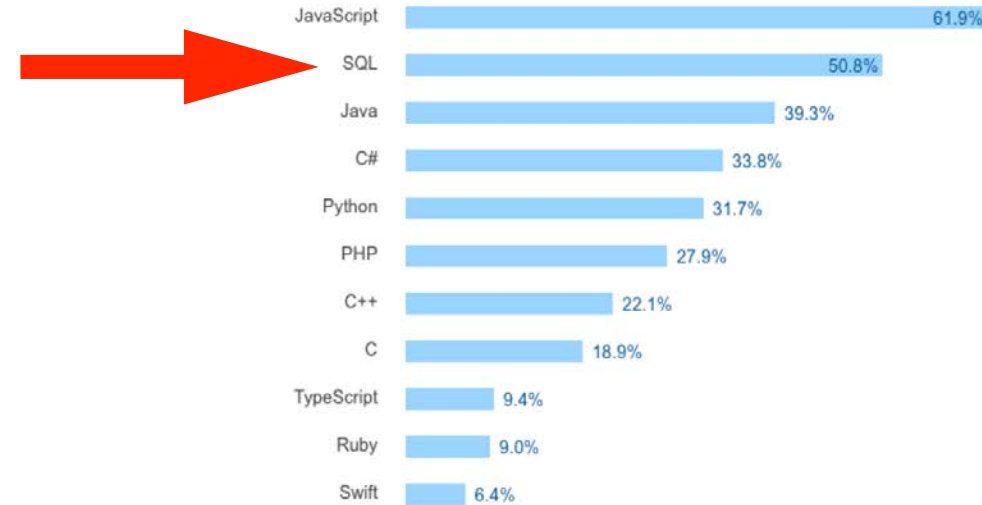
Dozent:

Sebastian Maneth (maneth@uni-bremen.de)



- Datenbankgrundlagen
 - Transaktionen, ACID, SQL
- Effiziente Auswertung via Indizes
 - B-trees
- Schnelle Volltextsuche
 - Suffix Trees & Arrays
 - Borrows Wheeler Transform

Stack Overflow Developer Survey (March 2017)



Most Popular Technologies — Programming Languages¹

¹ <https://stackoverflow.com/insights/survey/2017>



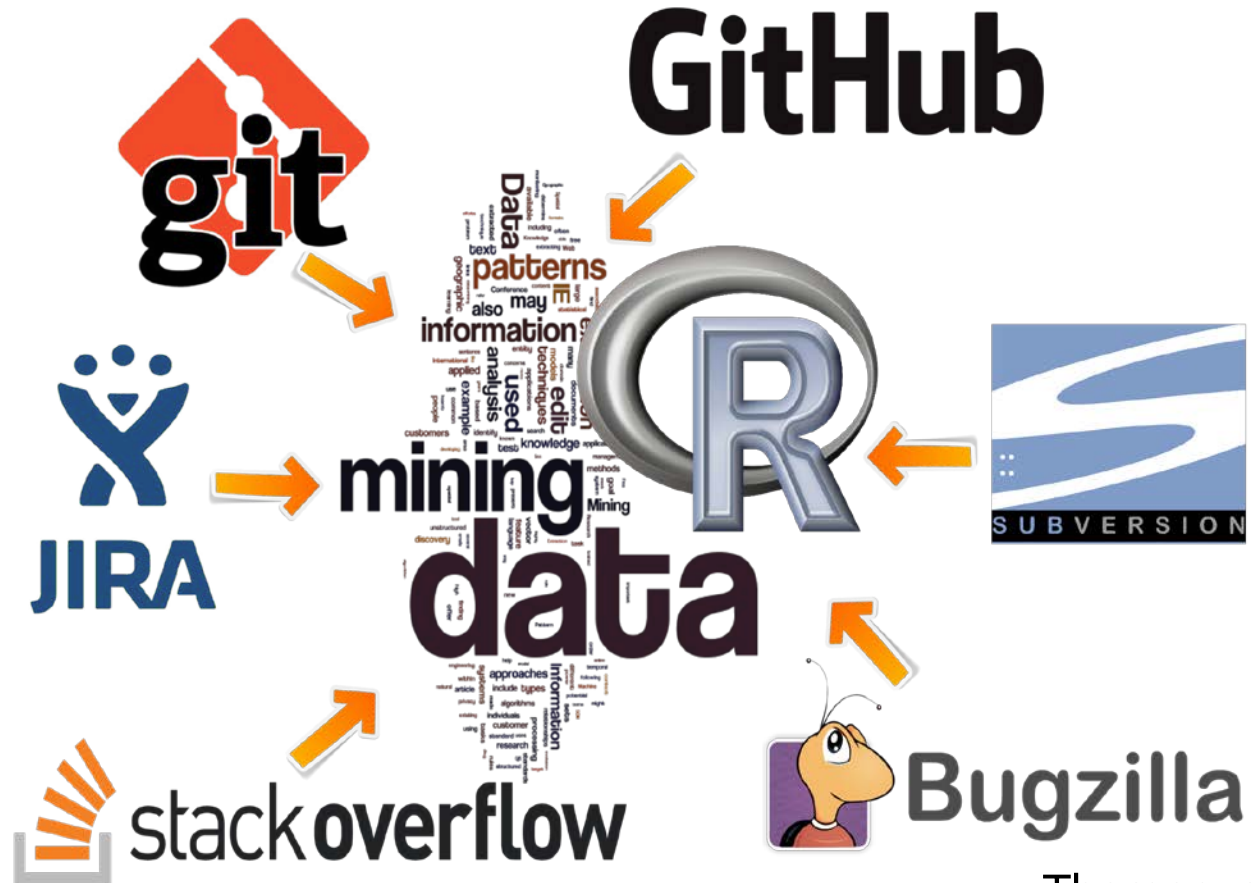
FDM 5 Data Science in der Softwaretechnik

Dozent:

Rainer Koschke (koschke@uni-bremen.de)



- Software-Analytics
- Software-Repository-Mining
- Software-Metriken
- Software-Visualisierung



**UBremen
Research
Alliance**

Data Train

Training in Research Data Management
and Data Science

Themenvorstellung

FDM: 6 Erkenntnisse in der digitalen Welt

Dozent:

Norman Sieroka (sieroka@uni-bremen.de)

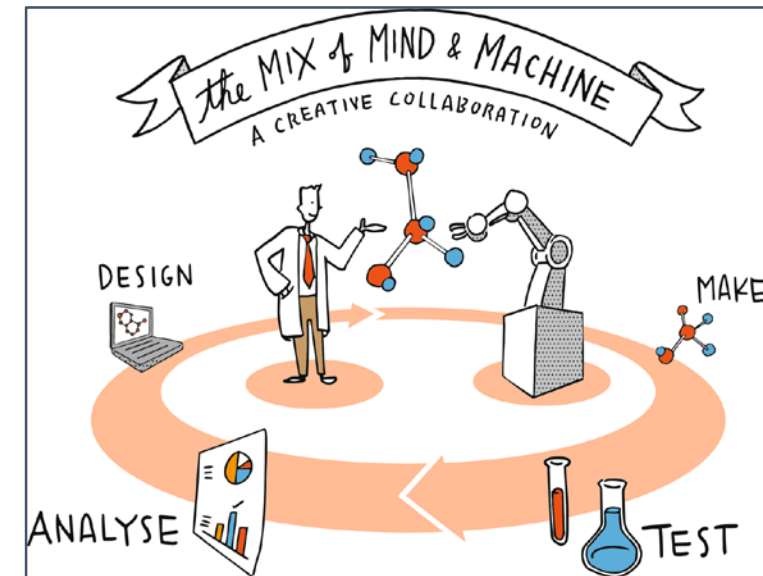


Philosophische Fragen und gesellschaftliche Herausforderungen von "DataScience,,

- Quantifizierung (in) der Wissenschaft – warum? was ist neu, was ist altbekannt?
- Standardfragen im Kontext künstlicher Intelligenz ... und was wir eigentlich fragen sollten

Es geht nicht darum, Ihnen eine fertige Theorie zu liefern,
... sondern darum, Sie zur kritischen Reflexion anzuregen!

Und wir werden alles "in unseren eigenen digitalen Welt"
(selbst entwickeltes Online-Tool) erarbeiten und dokumentieren!



**UBremen
Research
Alliance**

Data Train

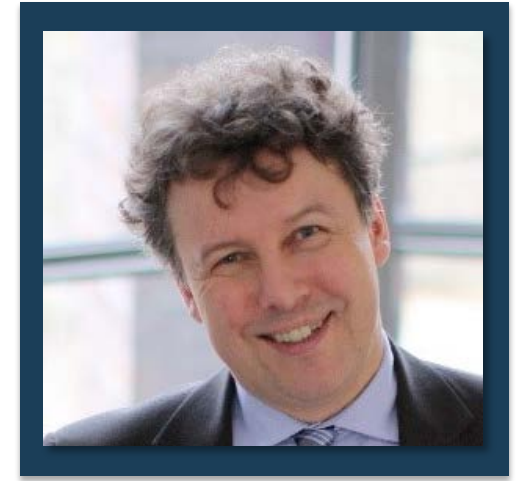
Training in Research Data Management
and Data Science

Themenvorstellung

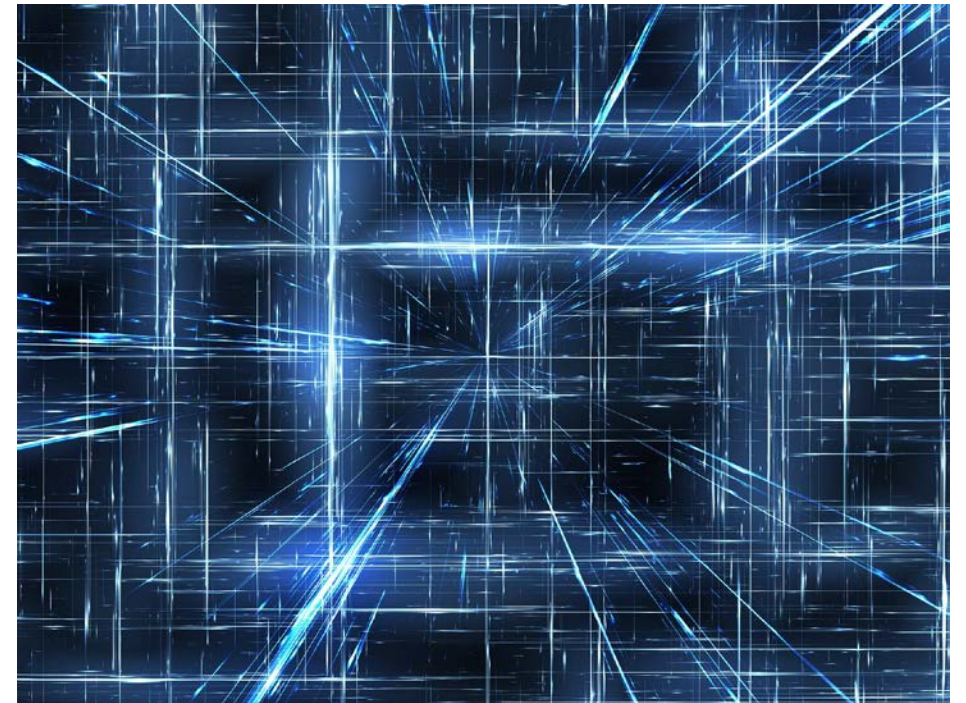
DS 1: Grundlagen der Informatik für Data Science

Dozent:

Rolf Drechsler (drechsler@uni-bremen.de)



- Informatik und ihre Teildisziplinen
 - Angewandt, technisch, praktisch, theoretisch
- Algorithmen
- Programmiersprachen
- Datenspeicherung und -verarbeitung
 - Datenstrukturen
- Anwendungsbeispiel: Sortieren
 - Bubble Sort, Merge Sort, Quicksort



**UBremen
Research
Alliance**

Data Train

Training in Research Data Management
and Data Science

Themenvorstellung

DS 2: Data Science Grundlagen

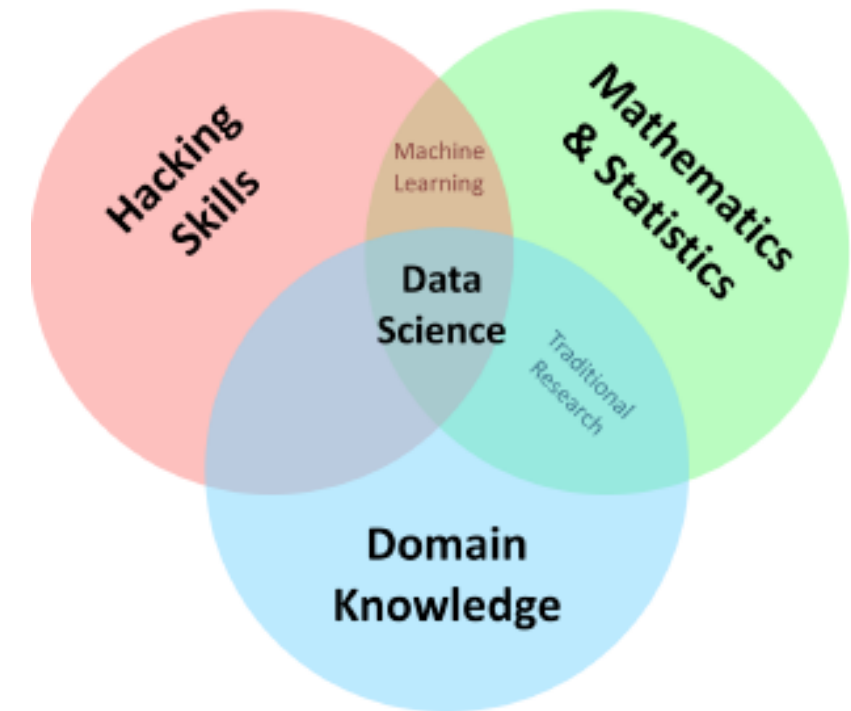
Dozent*innen:

Juliane Jarke (jjarke@ifib.de)

Hendrik Heuer (hheuer@ifib.de)



- Begriffsdefinition: Data Science als Anwendung des Maschinellen Lernens
- Unterscheidung zwischen Überwachtem Lernen (Klassifikation, Regression) und Unüberwachtem Lernen (Clustering, Dimensionsreduktion)
- Live-Coding eines Systems, das Kaufempfehlungen für Autos macht
- Informatik & Gesellschaftsperspektive mit besonderem Fokus auf Ethical, Legal, and Social Implications (ELSI) & Fairness, Accountability, & Transparency in ML (FATML)



Themenvorstellung

DS 3: Programmiersprachen

Dozent:

Christoph Lüth (clueth@uni-bremen.de)



- Was ist eigentlich eine Programmiersprache?
 - Was kennzeichnet eine Programmiersprache? Wozu dient sie, was muss sie alles können, und warum ist HTML keine? Und was hat Turing damit zu tun?
- Was gibt es für Programmiersprachen?
 - Es gibt ca. 700 Sprachen, wie behält man da den Überblick? Wir unterscheiden Sprachen nach ihrem Abstraktionsgrad, ihrem Programmierparadigma (imperativ, prozedural, objektorientiert, funktional, logisch...), oder ihrem Anwendungsgebiet.
- Welche Programmiersprachen sollte man kennen?
 - Wir stellen einige wichtige Sprachen vor: C, Java, Python kennen viele, Prolog, Haskell oder Scala vielleicht nicht.



**UBremen
Research
Alliance**

Data Train

Training in Research Data Management
and Data Science

Themenvorstellung

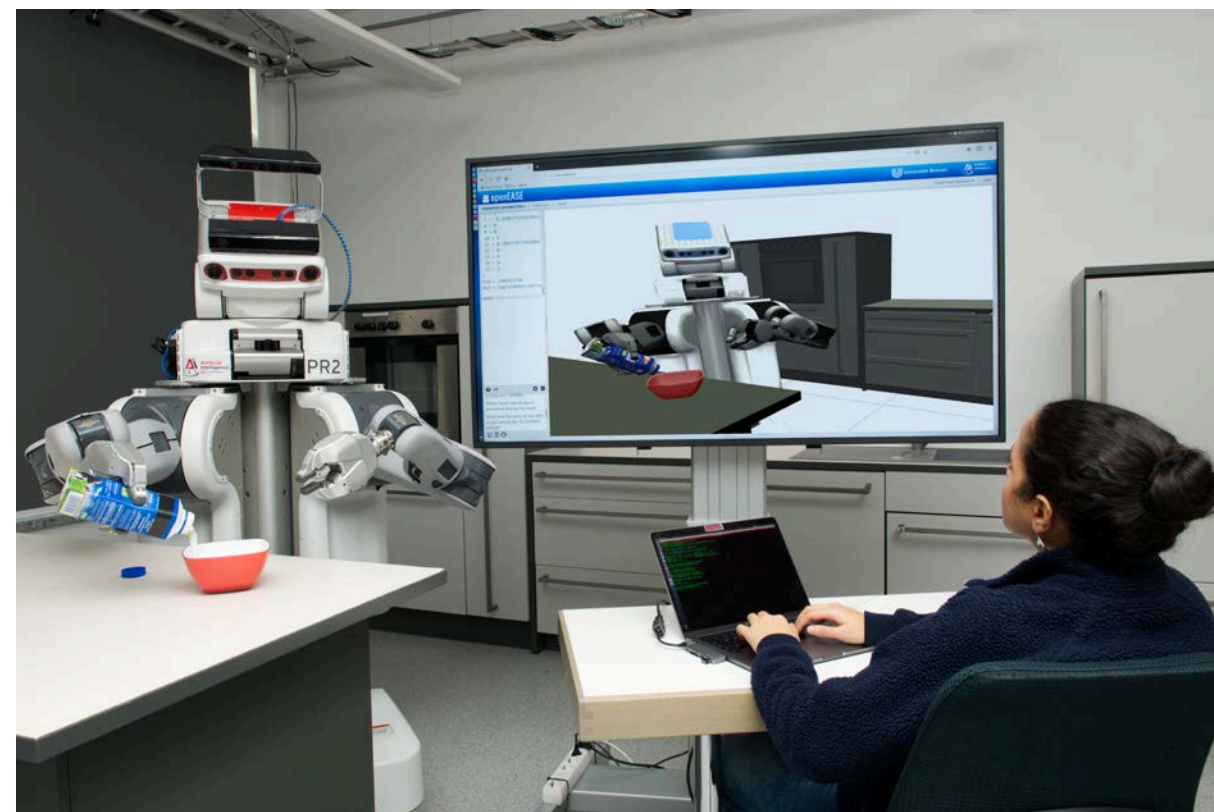
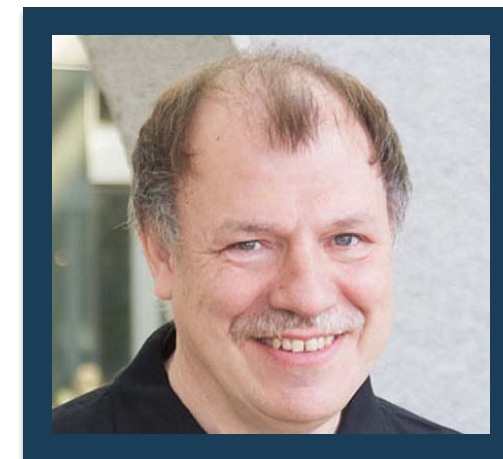
DS 4: Open KI-fähige Robotik Forschung

Dozent:

Michael Beetz (beetz@cs.uni-bremen.de)

In dem Seminar wird überblicksartig dargestellt, wie KI-fähige Roboter demokratisiert werden. Durch:

- Frei zugängliche Roboterdaten über eine öffentliche Cloud Infrastruktur
- Open-source Software
- „Laboratory without walls“
- Reproduzierbare Experimente
- Starkes Engagement frei zugänglicher Lehre und Ausbildung



**UBremen
Research
Alliance**

Data Train

Training in Research Data Management
and Data Science

Themenvorstellung

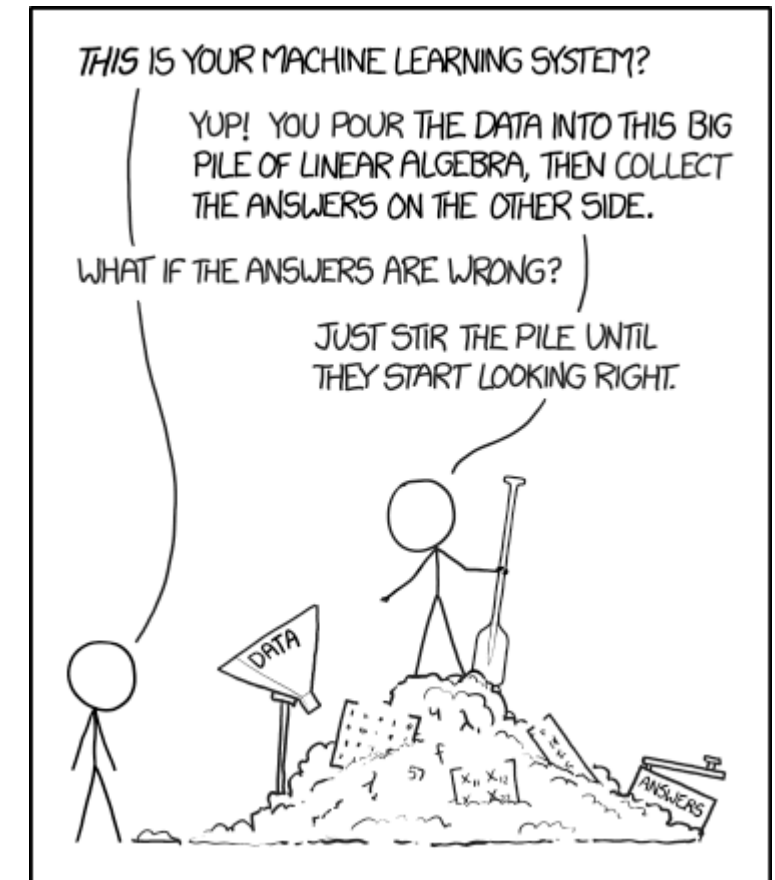
DS 5: Maschinelles Lernen

Dozent:

Marvin N. Wright (wright@leibniz-bips.de)



- Überblick maschinelle Lernverfahren
 - Nearest Neighbors, Gradient Boosting, Random Forests, Neural Networks, Support Vector Machines, u.v.m.
- Überwachtes und nicht überwachtes Lernen
- Klassifikation, Regression und Spezialfälle
- Parametertuning und Modellevaluation
- Praktisches Beispiel in R mit mlr3
- Ausblick: Interpretierbares maschinelles Lernen



<https://xkcd.com/1838/>

Themenvorstellung



**U Bremen
Research
Alliance**

Data Train

Training in Research Data Management
and Data Science

DS 6: Datenvisualisierung

Dozent:

Georgia Albuquerque (georgia.albuquerque@dlr.de)



- Visuelle Repräsentationen
 - “Information visualization”
 - “Scientific visualization”
- Multidimensionale Visualisierung
- Beispiele multidimensionaler Visualisierungen
 - “Scatterplot Matrices”
 - “Parallel Coordinates”
- “Interaction”

