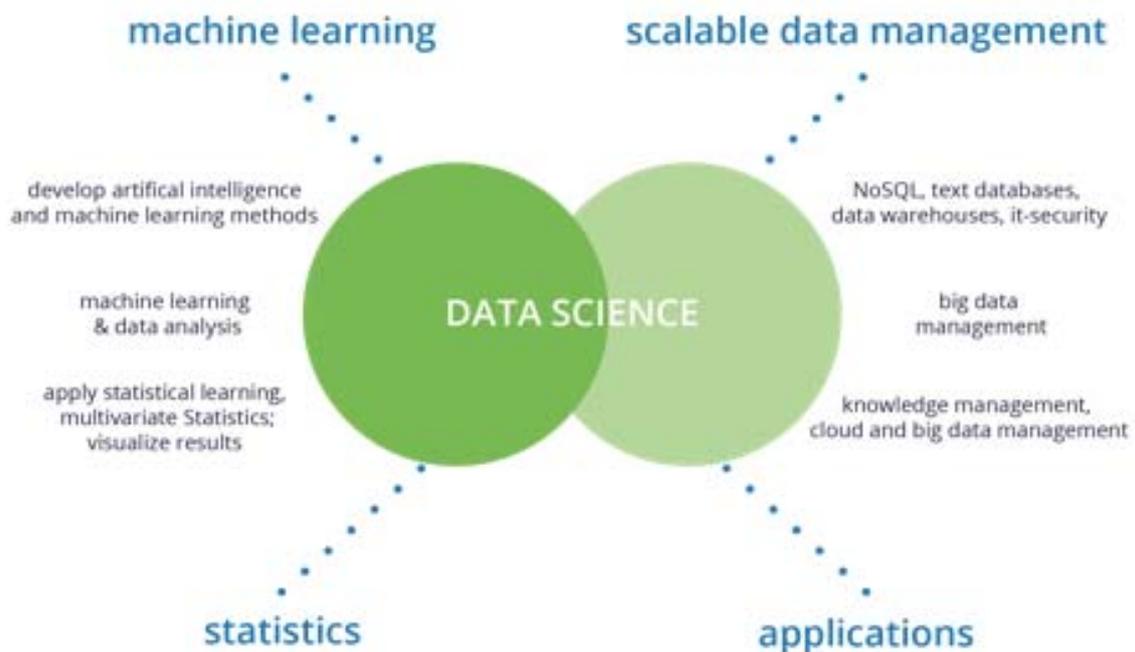


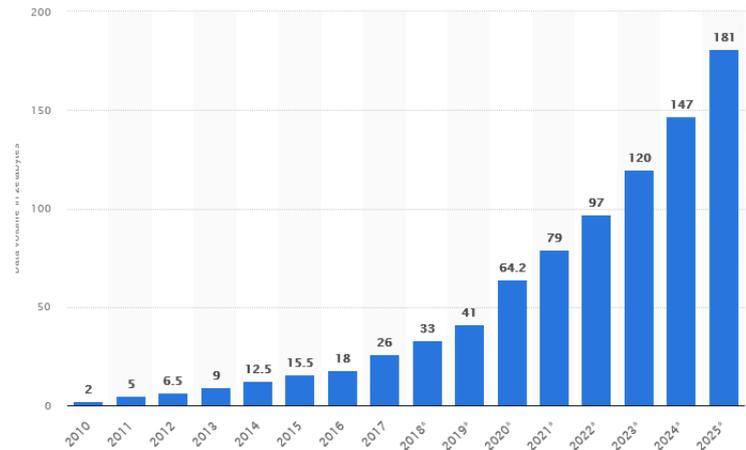
Master DATA SCIENCE

Prof. Dr. Erhard Rahm
Mai 2023

DATA SCIENCE



- massives Wachstum an Daten ("Big Data")
- zunehmende Digitalisierung in Wirtschaft und allen Lebensbereichen
- Erzeugung von Massendaten durch Smartphones (Fotos, Videos, Bewegungsdaten), Nutzung sozialer Netze und Internet, Sensoren bei der Produktion ...



Quelle: Statista

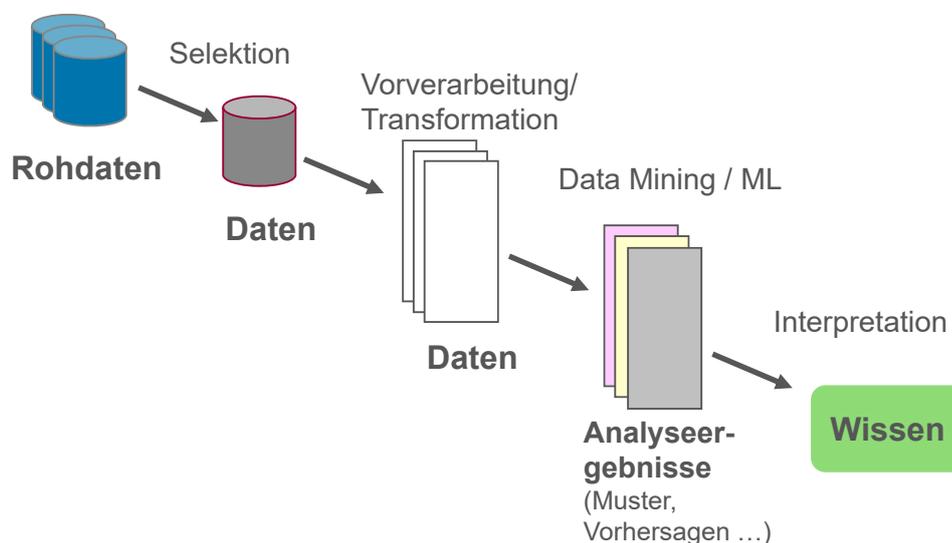
wachsende Datenmengen (in Zettabyte)

1 Zettabyte sind 1 Milliarde Terabyte (10^{21} B)



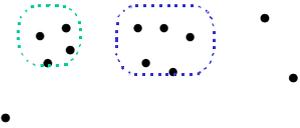
DATENANALYSE / KNOWLEDGE DISCOVERY

- (semi-)automatische Extraktion von Wissen aus Daten
- Kombination von Verfahren zu Datenbanken, Statistik (Data Mining) und KI (maschinelles Lernen)



Clusteranalyse

- Objekte (Kunden, Produkte, ...) werden aufgrund von Ähnlichkeiten in Klassen eingeteilt (Segmentierung)



Assoziationsregeln

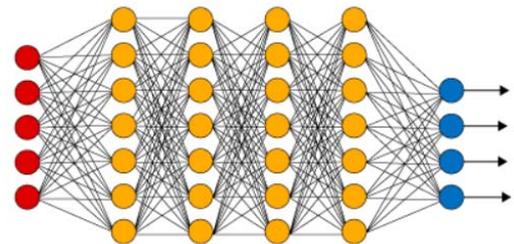
- Warenkorbanalyse (z.B. Kunde kauft A und B => Kunde kauft C)
- Nutzung für Kaufvorhersagen / Recommendations, Produkt-Bundling, ...

Klassifikation

- Zuordnung von Objekten zu Gruppen/Klassen mit gemeinsamen Eigenschaften bzw. Vorhersage von Attributwerten
- Verwendung von Stichproben (Trainingsdaten)
- Ansätze: Entscheidungsbaum-Verfahren, **neuronale Netze**, statistische Auswertungen

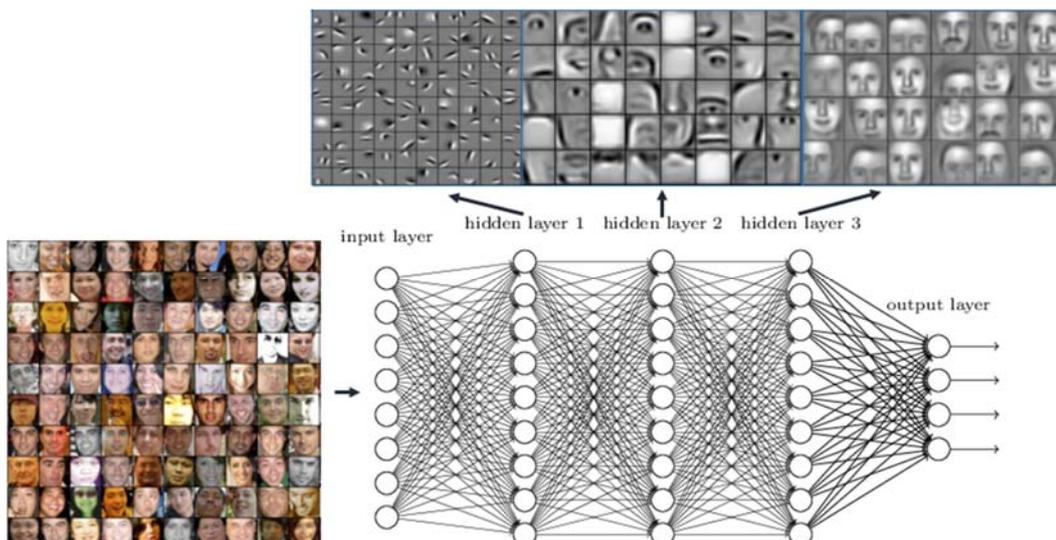
weitere Ansätze:

- genetische Algorithmen
- Regressionsanalyse zur Vorhersage numerischer Attribute ...



GESICHTSERKENNUNG

- neuronale Netze lernen Merkmale der Eingabedaten in Form von aufeinander aufbauenden Konzepten.
- hierarchische Repräsentation der Daten (Farbwerte der Pixel):
Kanten -> Teile des Gesichts -> gesamtes Gesicht



- KI erzeugt Dokumente, Programme, Fotos etc., in hoher Qualität
- ChatGPT und Varianten machen Nutzung sehr einfach
 - Nutzung sehr großer „Sprachmodelle“
 - Reinforcement Learning ermöglicht kontinuierliche Verbesserung durch Nutzer-Feedback



Ein Data Science Studium sollte eine breite Palette von Fähigkeiten und Kenntnissen abdecken, um einen erfolgreichen Karriereweg in der Branche zu ermöglichen. Hier sind einige wichtige Themen, die in einem Data Science Studium abgedeckt werden sollten:

1. Programmierung: Eine solide Kenntnis von Programmiersprachen wie Python und R ist unerlässlich, um Daten zu sammeln, zu verarbeiten und zu analysieren.
2. Mathematik und Statistik: Eine solide Kenntnis von Mathematik und Statistik ist für die Arbeit in der Datenanalyse unerlässlich. Themen wie lineare Algebra, Multivariable Analysis, Statistiken und Wahrscheinlichkeit sind wichtig.
3. Machine Learning: Das Verständnis von Machine Learning Algorithmen und Techniken, wie neuronale Netze, Entscheidungsbäume und Clustering ist ein wichtiger Bestandteil eines Data Science Studiums.
4. Datenbanken und Datenmanagement: Die Verwaltung von großen Datenmengen und die Arbeit mit Datenbanken sind wichtige Fähigkeiten für Data Scientists.
5. Data Visualization: Die Fähigkeit, Daten auf visuelle und ansprechende Weise darzustellen und zu kommunizieren, ist für eine effektive Datenanalyse unerlässlich.
6. Geschäftswissen: Data Scientists sollten auch in der Lage sein, Daten in den Kontext des Geschäfts zu stellen und die Auswirkungen auf das Unternehmen zu verstehen.
7. Big Data Technologien: Das Verständnis von Big Data Technologien wie Hadoop, Spark und NoSQL-Datenbanken ist für die Arbeit mit großen Datenmengen unerlässlich.
8. Ethik und Datenschutz: Data Scientists müssen auch über die ethischen und Datenschutzfragen im Zusammenhang mit der Arbeit mit Daten Bescheid wissen.

Diese Liste ist nicht abschließend, aber sie gibt einen guten Überblick darüber, was in einem Data Science Studium gelernt werden sollte.

- hoher und wachsender Bedarf an Master-Absolvent(inn)en mit vertieften Data-Science-Kenntnissen
- zukunftsorientierte Vertiefung aufbauend auf Bachelorabschluss
- Studienangebot mit vielen Berufsperspektiven
 - Forschung / Promotion (z.B. in Leipzig)
 - IT-Firmen, Startups
 - Beratung
 - alle Unternehmen/Organisationen mit Daten / IT-Abteilung ...
- Uni Leipzig bietet einzigartiges Angebot in Kombination mit KI-Zentrum ScaDS.AI

ZUGANGSVORAUSSETZUNGEN

- Bachelor Informatik, Wirtschaftsinformatik oder Data Science oder
- sonstiger erster berufsqualifizierender Hochschulabschluss sowie Kenntnisse in
 - Mathematik / theoretische Informatik im Umfang von 20 LP, darunter Wahrscheinlichkeitstheorie
 - Grundlagen Informatik/Programmierung im Umfang von mind. 40 LP, darunter MUP1, MUP2, ADS1, ADS2, DBS1, objektor. Programmierung
- englische Sprachkenntnisse
- Deadlines: 15.3. für SS, 15.9. für WS

- neuer Studiengang seit SS20
 - Beginn im WS oder SS
- Umfang: 4 Semester, 120 LP
- Hauptinhalte
 - Skalierbares Datenmanagement („Big Data“), mind. 20 LP
 - Datenanalyse / Machine Learning, mind. 20 LP
 - mind. 10 LP Statistik-bezogene Module
 - Ergänzungs-/Anwendungsmodule
 - zusätzliche Data-Science-Module
 - Module aus anderen Masterstudiengängen mit Data-Science-Bezug, auch Medizininformatik, Bioinformatik, Digital Humanities und Erdwissenschaften
 - viele Praktika möglich
- Schwerpunktsetzung/Vertiefung möglich in
 - Skalierbares Datenmanagement
 - Datenanalyse
 - mehr als minimaler Umfang von je 20 LP

GROBABLAUF

1. Semester		2. Semester		3. Semester	4. Semester
Skalierbare Datenbanktechnologien 1		Skalierbares Datenmanagement	Skalierbares Datenmanagement	Vertiefung	Mastersem. Data Science (5 LP)
Datenanalyse		Datenanalyse		Vertiefung	Masterarbeit (25 LP)
Ergänzung	Ergänzung	Ergänzung	Anwendung/ Ergänzung	Anwendung/ Ergänzung	

Schwerpunkt Big Data (Skalierbares Datenmgmt)

1. Semester		2. Semester		3. Semester		4. Semester	
Skalierbare Datenbanktechnologien 1		SDBT2	Big Data Praktikum	Prakt. Data Wareh./ Mining	Medical Data Science	Masterseminar (5 LP)	Masterarbeit (25 LP)
Künstl. neuronale Netze u. maschin. Lernen		Multivariate Statistik und Data Mining		Mathem. Foundations of DS	Advanced NLP		
IT-Sicherheit	Prakt. IT-Sicherheit	Data Prep. & Cleaning	Aktuelle Trends in DS	Verfahren und Anwendungen in den Digital Humanities			

Schwerpunkt Datenanalyse

Skalierbare Datenbanktechnologien 1		SDBT2	Big Data Praktikum	Mathem. Foundations of DS	Advanced NLP	Masterseminar (5 LP)	Masterarbeit (25 LP)
Künstl. neuronale Netze u. maschin. Lernen		Multivariate Statistik und Data Mining		Mediz. Bildverarbeitung u. Bildaufnahme			
Sequenzanalyse und Genomik		Grundl. komplexer Systeme	Aktuelle Trends in DS	Wiss. Visualisierung			

SKALIERBARES DATENMANAGEMENT

- **10-INF-DS-01 Skalierbare Datenbanktechnologien 1 (10 LP), Pflicht, WS**
- 10-INF-DS-101 Skalierbare Datenbanktechnologien 2 (5 LP), SS
- 10-INF-DS-102 Big Data Praktikum (5 LP), SS
- 10-INF-DS-103 Praktikum Data Warehouse & Data Mining (5 LP), WS
- 10-INF-DS104 Data Preparation & Cleaning (5 LP), SS
- 10-202-2503 Grundlagen der IT-Sicherheit (5 LP), WS
- 10-202-2502 Praktikum der IT-Sicherheit (5 LP), WS
- **10-INF-DS105 Verschlüsseltes Datenmanagement auf privaten Daten (5 LP), SS**
- 10-xxx-xxxx Datenschutz von Anonymisierung bis Zweckbindung (10 LP), SS
- 09-MIN-WA004 Medical Data Science (5 LP), WS

- 10-202-2128 Künstliche neuronale Netze, Deep Learning, maschinelles Lernen u. Signalverarbeitung (10 LP), WS
- 10-202-2104 Neuromorphe Informationsverarbeitung (10 LP), SS
- 10-202-2201 Wissenschaftliche Visualisierung (10 LP), WS, mit Praktikum
- 10-202-2204 Medizinische Bildverarbeitung und Bildaufnahme (10 LP), SS
- 10-202-2351 Advanced natural language processing (5 LP), WS
- 10-202-2340 Applied language technologies (5 LP), SS

- 10-INF-DS201 Mathematical foundations of data analysis (5 LP), WS
- 10-INF-DSxx Introductory Computing for Data Science (5 LP), SS
- 10-INF-DSxx Optimization for Machine Learning (5 LP), SS

- 10-INF-DS202 Probabilistic machine learning (5 LP), SS
- 09-INF-BI01 Statistisches Lernen (10 LP), WS, mit Praktikum
- 07-202-1103a Advanced Statistics (5 LP), WS
- 07-202-2302 Multivariate Statistik und Data Mining (10 LP), SS

- 10-INF-DS-301 Aktuelle Trends in Data Science (5 LP), SS

Data-Science-relevante Angebote aus Master Informatik, z.B.

- 10-202-2137 KI und Ethik (5 LP), WS
- 10-202-2131 Soziale Netzwerke (5 LP), WS
- 10-202-2346 Automated Software Engineering (10 LP), SS
- 10-202-2345 Software Engineering für KI-Systeme (10 LP), SS

Medizin/Bioinformatik

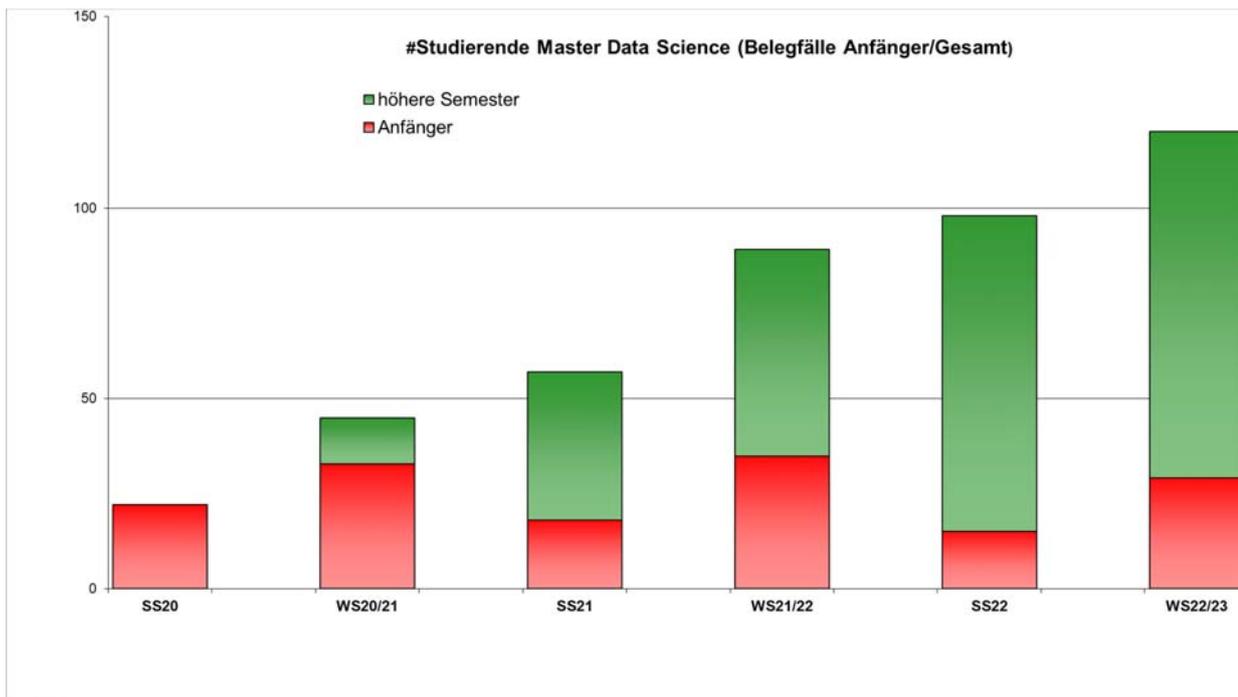
- 10-xxx-xxxx Künstliche Intelligenz im Gesundheitswesen (5 LP), SS
- 10-202-2207 Sequenzanalyse und Genomik (10 LP), WS
- 10-202-2205 Graphen und biologische Netze (10 LP), WS

Digital Humanities

- 10-DIH-0001 Verfahren und Anwendungen in den Digital Humanities (10 LP), WS

Earth Sciences

- 12-111-1034 Data Assimilation (5 LP), SS
- 12-GEO-M-DS02 Spatio-temporal data (5 LP), SS
- 12-GEO-M-SK01 Research Data Management and Social Responsibility (5 LP), WS



KI-ZENTREN

KI-Strategie des Bundes:
Einrichtung von 5 Zentren
für Künstliche Intelligenz
(neben DFKI) :

- Berlin (BIFOLD)
- Dortmund / Bonn (ML2R)
- Dresden / Leipzig (ScaDS.AI)
- München (MCML)
- Tübingen (tuebingen.ai)



SCADS.AI IN LEIPZIG

- Standort: Löhrs Carré (Humboldtstr. 25)
- Flächen für
 - derzeit 110 Arbeitsplätze
 - Graduiertenschule
 - Living Lab
 - Transferzentrum
 - KI-Professuren



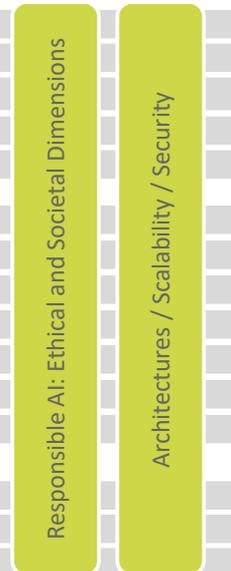
Research Areas



Topic Areas

- Life Science & Medicine
- Environment & Earth Sciences
- Software Engineering
- Physics / Chemistry
- Engineering / Business
- Understanding Language
- Methods and Hardware for Neuro-Inspired Computing
- Graph-based Artificial Intelligence
- Knowledge Representation & Engineering
- Scalable Visual Computing
- Federated, Efficient Learning
- Math Foundations & Statistical Learning
- Big Data Analytics
- Open Data & Open Models
- Data Quality & Data Integration

Crosscutting Topics



KI-PROFESSUREN

UNIVERSITÄT
LEIPZIG

5 neue KI-Professuren in Leipzig

- 1. Professur seit April 2022 besetzt
- KI-Humboldt-Professur seit Mai 2022
Sayan Mukherjee (von Duke Univ.)



2022

AI for Data Privacy and Security

Humboldt professorship

2023

Multi-modal Machine Learning

2024
/25

Computer Vision

Advanced Machine Learning



WEITERE INFOS / KONTAKT

UNIVERSITÄT
LEIPZIG

<https://www.uni-leipzig.de/studium/vor-dem-studium/studienangebot/studiengang/course/show/data-science-m-sc>

<https://dbs.uni-leipzig.de/de/study/mds>

Studienbüro: Augustusplatz 10, 04109 Leipzig, Raum A510
studienbuero@mathematik.uni-leipzig.de



Fragen ?

