Bachelor-Studiengang Data Science

Prof. Dr. Reimar Hofmann Studiendekan

Agenda

- Vorstellung
- Was ist Data Science?
- Themen und Themenwünsche
- Organisatorisches und Studium allgemein
- Einstieg in die beschreibende Statistik

Mein Hintergrund

 Studium: Informatik mit Nebenfach Mathematik an der TU München und University of Edinburgh



Siemens Forschung (Maschinelles Lernen)



 Gründung des StartUp Panoratio (In-Memory Datenanalyse)



 Seit 2009: Hochschule Karlsruhe: Mathe 1, Statistik, Data Science, Entrepreneurship



WS 2013: Google/Zürich (In-Memory Datenanalyse)



• Seit 2020: Studiendekan im Bachelor-Studiengang Data Science



- Warum studieren Sie Data Science?
- Ihre Erwartungen an dieses Studium?
- Was möchten Sie lernen?

• ...

Bitte an der Umfrage teilnehmen:



https://shorturl.at/cdnUZ
oder im ILIAS-Ordner dieses Kurses



Data Science

Mögliche Definition*:

Data Science beschäftigt sich mit generalisierbaren Systematiken, mit denen aus Daten Wissen gewonnen werden kann.

Insbesondere:

- Erkenntnisgewinn
- Vorhersagen
- Entscheidungsoptimierung

Der Begriff kann sowohl manuelle als auch maschinelle Verfahren umfassen.

* frei nach Dhar, Vasant (2013), Communications of the ACM, Vol. 56 No. 12, Pages 64-73

Die Treiber

- Digitalisierung
- Neue Verfahren und Technologien
- Big Data
- Maschinelles Lernen
- Deep Learning
- •



Data Analytics

Data Analytics beschäftigt sich mit Verfahren und Vorgehensweisen, wie aus Daten Erkenntnisse gewonnen werden können. Der Fokus liegt dabei auf *manuellen Schritten*.

Mögliche Ziele:

- Erkenntnisgewinn (als Selbstzweck)
- Grundlage für menschliche Entscheidungen/Optimierungen.
- Kombination von datenbasiertem mit menschlichem Wissen
- Oft Teilschritt beim maschinellen Lernen
 (z.B. feature Engineering, Interpretation und Bewertung von Ergebnissen)

Maschinelles Lernen (früher: Data Mining)

Maschinelles Lernen beschäftigt sich mit Verfahren und Vorgehensweisen, wie aus Daten Erkenntnisse gewonnen werden können. Der Fokus liegt dabei auf *manuellen Schritten*.

Mögliche Ziele:

- Erkenntnisgewinn (als Selbstzweck)
- Grundlage für menschliche Entscheidungen/Optimierungen.
- Kombination von datenbasiertem mit menschlichem Wissen
- Oft Teilschritt beim maschinellen Lernen (z.B. feature Engineering, Interpretation und Bewertung von Ergebnissen)

Künstliche Intelligenz

Definition früher (1990er Jahre):

Maschine, die den Turing-Test besteht, d.h. die ein Mensch im Chat nicht von einem Menschen unterscheiden kann.

Definition heute:

Wird heute für alles benutzt, was irgendwie mit Daten zu tun hat, z.B.

- Empfehlungssysteme auf Netflix,
- Vorhersage, welche Kunden Ihr Abo kündigen werden ...



Buzzword Bingo

Die Modebegriffe im IT-Umfeld ändern sich alle 5-10 Jahre. Die zugrundeliegenden Konzepte bleiben aber oft 20-30 Jahre gültig.

Artificial Intelligence

MIS (Management Information System)

EIS (Executive Information System)

DSS (Decision Support System) **Da**

Data Mining

KDD (Knowledge Discovery in Databases)

Data Warehouse

Business Intelligence

Predictive Analytics

Ad-Hoc Reporting

Real-Time DWH Self-Service BI

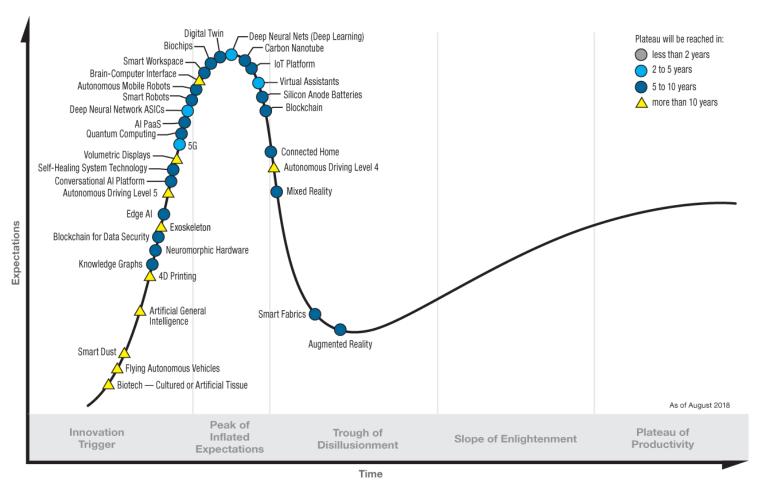
Deep Learning

Data Science Big Data



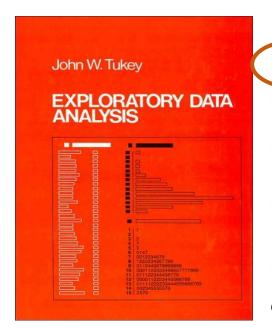


Gartner Hype Cycle for Emerging Technologies, 2018





Die Themen sind gar nicht so neu ...



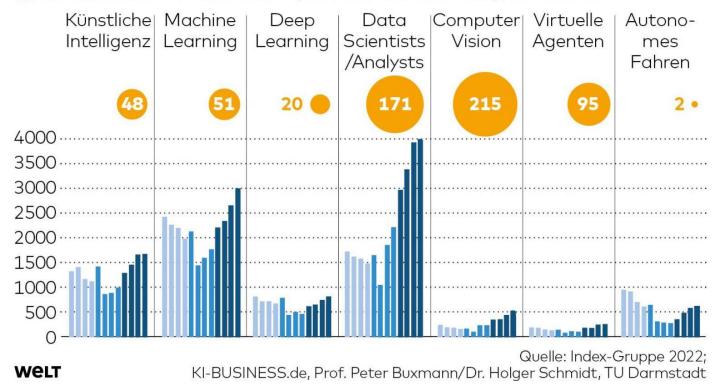
tatistician, interested in inferences from the particular to the general. But as I have watched mathematical statistics evolve, I have had cause to wonder and doubt... I have come to feel that my central interest is in *data analysis*... Data analysis, and the parts of statistics which adhere to it, must...take on the characteristics of science rather than those of mathematics... data analysis is intrinsically an empirical science... How vital and how important... is the rise of the stored-program electronic computer? In many instances the answer may surprise many by being 'important but not vital,' although in others there is no doubt but what the computer has been 'vital.'" In 1947, Tukey coined the term "bit" which Claude Shannon used in his 1948 paper "A Mathematical Theory of Communications." In 1977, Tukey published *Exploratory Data Analysis*, arguing that more emphasis needed to be placed on using data to suggest hypotheses to test and that Exploratory Data Analysis and Confirmatory Data Analysis "can—and should—proceed side by side."

Quelle: Forbes (http://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/#1db981b69fd2)

Es geht heute immer noch darum, die Ziele umzusetzen, an denen man seit 60 Jahren arbeitet. Die **Methoden entwickeln sich** *langsam und stetig* **weiter**, die Tools und vor allem die Rechenleistung wechseln etwas schneller. Die **Buzzwords wechseln viel schneller und radikaler**, da die Hersteller und Berater ständig Overselling betreiben, und deshalb die Begriffe bald mit enttäuschten Erwartungen assoziiert werden und damit verbrannt sind – und das, obwohl in dem Gebiet gute Fortschritte gemacht werden.

KI-Experten sind stark nachgefragt

Ausgeschriebene Positionen zwischen Q1 2019 und Q4 2021 in Deutschland Wachstum über zwei Jahre (Q4 2019 bis Q4 2021), in Prozent



Quelle: Bis zu 200.000 Euro im Jahr – so kommen Sie an die lukrativen KI-Jobs., Zeitung die WELT vom 11.03.2022, https://www.welt.de/wirtschaft/karriere/plus237471693/KI-Gehaelter-Bis-zu-200-000-Euro-so-kommen-Sie-an-die-lukrativen-Jobs.html



Was macht ein Data Scientist?

Alter: 35

Beruf: Data Scientist

Sparte:



Berufliche Ziele

- "etwas Sinnvolles tun"
- Durch Data Science Behandlung und Lebensqualität der Patienten verbessern

Arbeitsaufgaben:

- Data Scientist in einem Klinikverbund
- Zuständig für Gestaltung datengetriebener Ansätze zur Verbesserung der Abläufe und/oder der Behandlungsqualität
- Ergänzt die klassische IT-Abteilung
- Schnittstellenfunktion zu vielen Bereichen

- Datenbestände aus unterschiedlichen Quellen zusammenschließen interdisziplinäres Umfeld, Lösungsgestaltung (führt den Design-Prozess)
- Zusammenarbeit mit Fachexperten, Pflegeexperten, Medizinern und dem Controlling/der medizinischen Kodierung
- Projektorientiert



Alter: 30

Beruf: Data Scientist Sparte: Industrial IoT



Berufliche Ziele

- Reelle Datenprobleme in all ihrer theoretischen und praktischen Komplexität lösen
- Die Digitalisierung mitgestalten

Arbeitsaufgaben:

- Data Scientist bei einem Anbieter für Industrial IoT Lösungen
- Erfasst und modelliert kundenspezifische Probleme; entwickelt Lösungen zur Datenerfassung, -integration und –interpretation. Entwickelt, darauf aufbauend, Data Science und ML Software und Algorithmen. Integriert diese in IoT Lösungen im Produktionsumfeld

- Interdizsiplinäre, technische Arbeit im Kontext Machine Learning, Informatik, IoT
- Arbeitet vorrangig mit Kollegen aus dem Data Science Team sowie Backend-Entwicklern zusammen.
- Zusammenarbeit mit dem technischen Vertrieb und gelegentlich den Kunden zur vertieften technischen Umsetzung der kundenspezifischen Lösungen



Alter: 40

Beruf: Lead Data Scientist

Sparte: Elektronik



Berufliche Ziele

- Mit Daten neue Lösungen finden: Risikoanalyse, Forecasting
- *Industrie*probleme lösen
- Dieselben Ansätze auf ganz unterschiedliche Bereiche anwenden

Arbeitsaufgaben:

- Lead Data Scientist bei einem Hersteller für halbleiterbasierte Elektronik-Produkte
- Verantwortet datengestützte Innovationen der Produkte, von Fertigungsoptimierung bis hin zu innovativen Anwendungen wie Connected Agriculture
- Leitet ein Team von Data Scientists
- Hält Vorlesungen und betreut Studenten im Bereich Data Science

- Interagiert mit unterschiedlichen Abteilungen um "Daten-Probleme" aufzugreifen und deren Lösungen zu konzeptionieren und umzusetzen.
- Hauptaufgabe: Theoretische Modellierung der praktischen Anforderungen, in der Regel mit Hilfe mathematischer und Machine-Learning Konzepte.
- Umsetzung erfolgt durch das Data Science Team, mit Hilfe der IT.



Alter: 35

Beruf: Data Scientist Sparte: E-Commerce



Berufliche Ziele

- Aus Daten unternehmerische Erkenntnisse und Handlungsempfehlungen ableiten
- In einem interdisziplinären Umfeld tätig sein

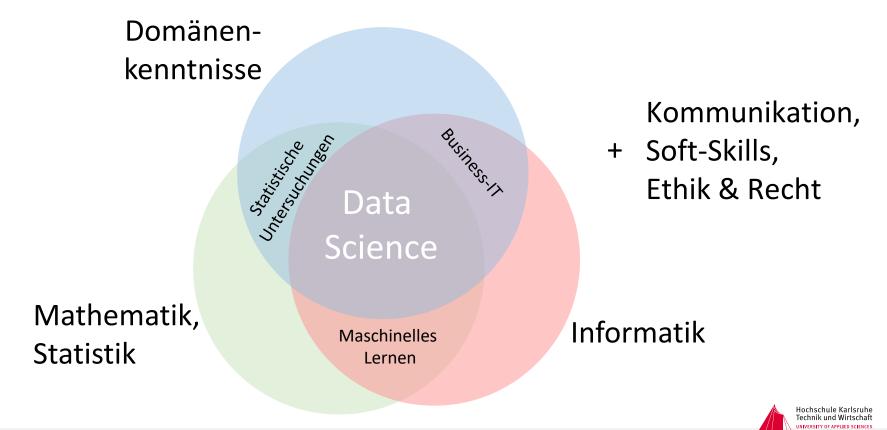
Arbeitsaufgaben:

- Einziger Data Scientist bei der E-Commerce-Abteilung eines grossen Online-Bekleidungsshop
- Web-Analysen zur Feststellung der Performance von Vermarktung und E-Commerce Produkten.
- Marketing-Kampagnen, Analysen, Executive Performance Reports, ...
- Schnittstellenfunktion zwischen Produkt, Analysetechnik und Geschäftsführung

- Arbeitet produktseitig mit Werbeexperten, Produkt- und Vermarktungsspezialisten, von denen die meisten kaum technische oder Data Science Kenntnisse haben
- Erstellt Handlungsempfehlungen und Analysen für die Geschäftsführung
- Arbeitet technikseitig mit wenigen internen Datenbankspezialisten und vielen externen Anbietern von BI- und Analyse Tools

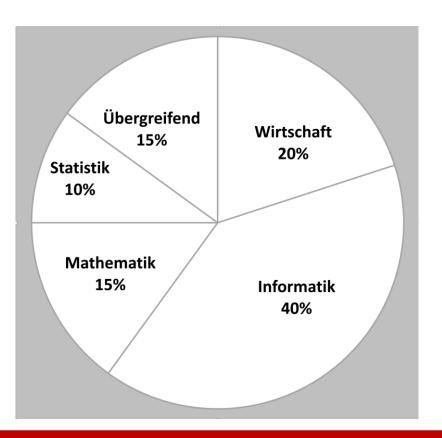


Was muss ein Data Scientist können?





Aufbau des Studiums



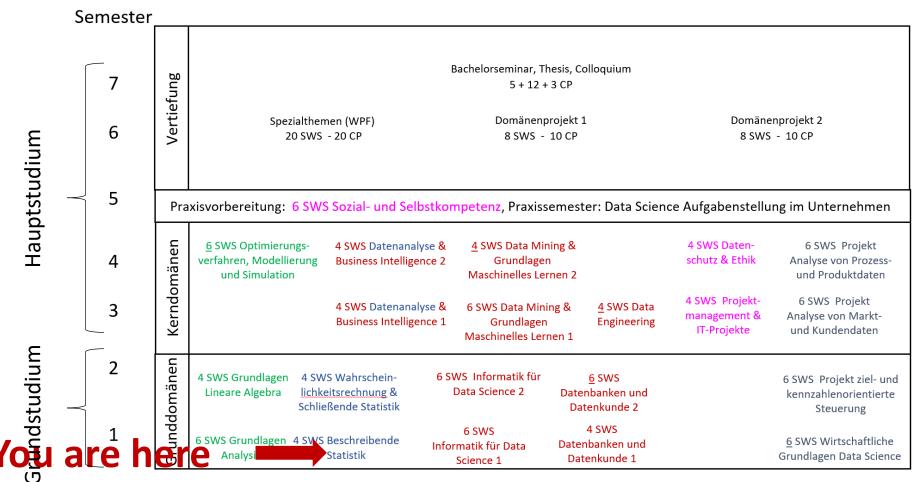
Ab dem 2. Semester jedes Semester ein Data Science Projekt, z.B.

- Ziel- und kennzahlenorientierte Steuerung in Unternehmen
 (2. Sem)
- Analyse von Markt- und Kundendaten (3. Sem)
- Analyse von Prozess- und Produktdaten (4. Sem)

Anwendungsorientiertes Lernen:

Alle Themen (auch Mathematik, Statistik, Informatik) werden möglichst an Hand solcher Projekte vermittelt

Studiengang Data Science



Mathe

Statistik

Informatik

Übergreifend

Wirtschaft

Was Sie gelernt haben sollten

- Abgrenzung zwischen den Begriffen "Data Science", "Data Analytics" und "Maschinelles Lernen"
- Sich von Modebegriffen nicht beeindrucken zu lassen