



GESELLSCHAFT  
FÜR INFORMATIK

PERSONNEL  
ANALYSIS  
DATA  
SEARCHING  
VERIFICATION  
CODING  
TRAINING

## EMPFEHLUNGEN

# MASTERSTUDIENGÄNGE „DATA SCIENCE“ –

AUF BASIS EINES BACHELORS IN  
(WIRTSCHAFTS-) INFORMATIK ODER MATHEMATIK

JULI 2021

[WWW.GI.DE](http://WWW.GI.DE)

# EMPFEHLUNGEN FÜR MASTERSTUDIENGÄNGE „DATA SCIENCE“ – AUF BASIS EINES BACHELORS IN (WIRTSCHAFTS-) INFORMATIK ODER MATHEMATIK

---

Empfehlungen der Gesellschaft für Informatik e.V.  
erarbeitet von der Task Force „Data Science“

Die Empfehlungen wurden am 26. Juni 2021 vom Präsidium  
der GI verabschiedet

An diesen Empfehlungen haben mitgewirkt:

Ziawasch Abedjan (Hannover), Thomas Bendig (Berlin),  
Ulf Brefeld (Lüneburg), Joachim Bürkle (Frankfurt), Jörg Desel (Hagen),  
Stefan Edlich (Berlin), Thomas Eppler (Albstadt-Sigmaringen),  
Michael Goedicke (Duisburg-Essen), Nils Hachmeister (Bielefeld),  
Jens Heidrich (Kaiserslautern), Stephan Höppner (Berlin),  
Stefan M. Kast (Dortmund), Daniel Krupka (Berlin), Klaus Lang (Bingen),  
Peter Liggesmeyer (Kaiserslautern), Julia Meisner (Berlin),  
Ingo Scholtes (Wuppertal), Marina Tropmann-Frick (Hamburg)

Die Empfehlungen wurden von Jörg Desel koordiniert.

# INHALT

	Vorwort . . . . .	4
<b>1</b>	Vorbemerkung . . . . .	5
<b>2</b>	Anforderungen aus der Berufspraxis . . . . .	6
<b>3</b>	Begriffliche Einordnung. . . . .	7
<b>4</b>	Data-Science-Kompetenzen . . . . .	8
<b>5</b>	Lehrinhalte aus Teildisziplinen. . . . .	9
	5.1 Mathematik und Statistik. . . . .	9
	5.2 Informatik und Programmierung. . . . .	10
	5.3 Data-Science-Methoden . . . . .	11
	5.4 Interdisziplinäre Aspekte . . . . .	12
	5.5 Anwendung von Data Science in der Domäne . . . . .	13
<b>6</b>	Curriculumsempfehlungen für Masterstudiengänge „Data Science“ . . . . .	14
	6.1 Mathematik und Statistik. . . . .	14
	6.2 Informatik und Programmierung. . . . .	15
	6.3 Data-Science-Methoden . . . . .	15
	6.4 Interdisziplinäre Aspekte . . . . .	15
	6.5 Anwendung von Data Science in der Domäne . . . . .	15
	6.6 Tabellarische Darstellung des Curriculums. . . . .	16
	6.7 Abschließende Empfehlungen . . . . .	16
	Literatur . . . . .	17
	Autorenschaft . . . . .	19

## VORWORT

Daten bilden die Grundlage einer zunehmend digital vernetzten Gesellschaft. In der Wirtschaft ist Data Science der Schlüssel der digitalen Transformation. Daten innovativ, verantwortungsvoll und gemeinwohlorientiert zu nutzen, kann das Zusammenleben in Deutschland, in Europa und in der Welt bedeutsam verbessern und natürliche Ressourcen schützen [Bundesregierung2021]. Der Forschung an und mit Daten kommt dabei eine zentrale Bedeutung zu – seien es Satellitenbilder als Datenquellen für Navigationssysteme oder für die Klimaforschung, Studiendaten zur Unterstützung von Mediziner\*innen bei der Diagnose oder Daten über Konsument\*innenverhalten zur Berechnung der individuell passendsten Angebote. Die Menge der täglich neu generierten Daten ist enorm, genau wie der potenzielle Nutzen für Wissenschaft, Forschung, Unternehmen und Verwaltung.

Die Fähigkeit, diese Daten zu analysieren und nutzbar zu machen, ist eine wichtige Voraussetzung, um wertvolle Erkenntnisse in der Wissenschaft gewinnen und nutzbringende Anwendungen für Gesellschaft, Staat und Wirtschaft generieren zu können. Der interdisziplinäre Forschungszweig Data Science, also das Management und die Analyse von Daten, gilt daher schon heute als eine der wichtigsten Schlüsseldisziplinen für Wissenschaft und Wirtschaft. Auch für die weitere Anwendung von Künstlicher Intelligenz (KI) und Lernenden Systemen stellt die Verfügbarkeit von Daten und die Datenverwaltung eine zentrale Voraussetzung dar.

Die Bundesregierung hat deshalb die Nationale Forschungsdateninfrastruktur (NFDI) auf den Weg gebracht, die die dezentral, projektförmig und/oder temporär



PETER LIGGESMEYER, MICHAEL GOEDICKE, JÖRG DESEL

gelagerten Datenbestände von Wissenschaft und Forschung in Deutschland erschließen soll. Als übergreifendes Forschungsdatenmanagement soll die NFDI Standards im Datenmanagement setzen und Forschungsdaten nachhaltig nutz- und weiterverwendbar machen.

Mit diesen curricularen Empfehlungen für Data-Science-Masterstudiengänge aufbauend auf einem Bachelor in der (Wirtschafts-) Informatik und der Mathematik sollen Mindeststandards definiert werden und Empfehlungen für die Akkreditierung solcher Studiengänge vorgelegt werden.

### PETER LIGGESMEYER

Past-Präsident der Gesellschaft für Informatik, Sprecher GI-Task Force „Data Science“

### MICHAEL GOEDICKE

Vize-Präsident der Gesellschaft für Informatik, Sprecher GI-Task Force „Data Science“

### JÖRG DESEL

Sprecher GI-Fachbereich

„Informatik und Ausbildung / Didaktik der Informatik“

## 1 VORBEMERKUNG

Aufgrund der wachsenden Bedeutung von Daten, dem wissenschaftlich fundierten Umgang mit diesen (Data Science) und der großen Relevanz für die Informatik hat der Vorstand der Gesellschaft für Informatik (GI) 2018 die Task Force „Data Science/Data Literacy“ ins Leben gerufen, die sich gemeinsam mit Partnern anderer Wissenschaftsdisziplinen aus den Natur- und Lebenswissenschaften das Ziel gesetzt hat, Empfehlungen für die Entwicklung von Studien- und Weiterbildungsangeboten auszusprechen.

Unter Beteiligung der Plattform Lernende Systeme wurde das Arbeitspapier „Data Science: Lern- und Ausbildungsinhalte“ (Dezember 2019) erstellt, welches ein Set an Lern- und Ausbildungsinhalten definiert und dieses in 14 Kompetenzfelder strukturiert [GI2019]. Darüber hinaus wurden drei idealtypische Personas mit verschiedenen Bildungs- bzw. Ausbildungszielen definiert: (A) Master in Data Science, um später als Data Scientist in der Industrie oder Forschung tätig zu sein, (B) Master in einer Wissenschaft aus einer Anwendungsdomäne (z. B. Naturwissenschaften), um Data-Science-Kompetenzen für die Domäne zu erwerben, und (C) Berufstätige mit informatischen und mathematischen Kenntnissen, um Data-Science-Kompetenzen für die praktische Anwendung im Job zu erwerben.

Das Ziel dieser Publikation liegt darin, die Empfehlungen zu Lern- und Ausbildungsinhalten für Persona A (Master in Data Science mit einem Bachelor in Informatik, Mathematik oder verwandten Studiengängen) weiter zu konkretisieren und Empfehlungen für ein Curriculum für Data-Science-Studiengänge aus Sicht der Informatik vorzulegen. Es soll als Diskussionsgrundlage dienen, um mit anderen beteiligten Fachdisziplinen abgestimmte ASIIN-Empfehlungen für die Akkreditierung solcher Studiengänge zu erarbeiten.

Für angenommene Vorqualifikationen aus Bachelorstudiengängen in Informatik und Wirtschaftsinformatik werden die „Empfehlungen für Bachelor- und Masterprogramme im Studienfach Informatik an Hochschulen“ [GI2016] sowie die „Rahmenempfehlung für die Ausbildung in Wirtschaftsinformatik an Hochschulen“ [GI2017] herangezogen. Die aktuellen fachspezifisch Ergänzenden Hinweise (FEH) der ASIIN beziehen sich für Informatik und Wirtschaftsinformatik auf dieselben Inhalte. Für angenommene Vorqualifikationen aus Bachelorstudiengängen der Mathematik verwenden wir die ASIIN-FEH „Zur Akkreditierung von Bachelor- und Masterstudiengängen der Mathematik“ vom 09.12.2016 [ASIIN2016].

Wir betrachten Data-Science-Masterstudiengänge (Persona A aus dem Arbeitspapier zu Lern- und Ausbildungsinhalten), die entweder auf einem Bachelorabschluss Mathematik, Informatik, Wirtschaftsinformatik oder einem vergleichbaren Bachelor mit Kompetenzen in Mathematik, Statistik und Informatik aufbauen (wie z. B. die Bachelorstudiengänge der Statistik sowie der Bio- oder Geoinformatik). Streng konsekutive Data-Science-Studiengänge (Bachelor und Master) werden hier nicht adressiert. Je nach vorliegender Vorqualifikation kann auf entsprechenden Kompetenzen aufgebaut werden. Sollte ein Masterstudiengang „Data Science“ alternative Zugangsmöglichkeiten haben, muss gewährleistet werden, dass die jeweils fehlenden Kompetenzen nachgeholt werden.

Wir betrachten nicht explizit Bachelorstudiengänge „Data Science“, auch wenn diese partiell die hier vorgestellten Inhalte eines Masterstudiums „Data Science“ behandeln können. Da diese nicht auf den oben genannten Grundlagen aufbauen und zusätzliche Bachelor-typische Module zu überfachlichen Kompetenzen beinhalten, können grundsätzlich in den etwas längeren Bachelorstudiengängen (sechs oder sieben Semester) dieselben Inhaltsbereiche behandelt werden wie in den Masterstudiengängen – allerdings in einer niedrigeren Durchdringung. Bei Masterstudiengängen sollte die Mehrzahl der Kompetenzen mit den Lernzielen „Anwenden“ und „Analysieren“ (nach der Anderson-Krathwohl-Taxonomie) vermittelt werden. Bei Bachelorstudiengängen reichen die Lernziele „Verstehen“ und „Anwenden“ bei der Mehrzahl der Kompetenzen aus. Spätestens bei Seminaren oder der Abschlussarbeit unterscheiden sich Bachelor- und Masterstudiengänge grundsätzlich.

Auch wenn wir keine konkreten Empfehlungen für die Aufteilung von Lerninhalten zwischen Bachelor und Master für konsekutive Studiengänge geben, können Lerninhalte analog zur Betrachtung von Bachelorstudiengängen in Data Science behandelt werden. Ein\*e wie oben beschriebene\*r Bachelorabsolvent\*in bringt – auf einer niedrigeren Ebene der kognitiven Durchdringung – bereits die wesentlichen Kompetenzen mit, die im Masterstudium entsprechend vertieft werden können. Wir empfehlen, entsprechende Vertiefungsmöglichkeiten in allen im Arbeitspapier genannten Kompetenzbereichen vorzusehen.



## 2 ANFORDERUNGEN AUS DER BERUFSPRAXIS

Aufgrund veränderter Anforderungen in der Arbeitswelt verlangen auch immer mehr Unternehmen, Verwaltungen und andere Organisationen Data-Science-Kompetenzen von ihren Beschäftigten. Es gibt nicht nur eine große Nachfrage, sondern auch einen spürbaren Mangel an qualifizierten Data Scientists, die von vielen Unternehmen und Organisationen händeringend gesucht werden. In der jüngeren Vergangenheit ist hier ein großer Kampf um die besten Talente entbrannt.<sup>1</sup>

Laut einer IDC-Studie ist der prozentuale Anteil der Data Scientists an der gesamten Beschäftigtenzahl zwischen 2013 und 2019 lediglich von 3,0 % auf 3,7 % gestiegen.<sup>2</sup> Der Bedarf an Fachkräften im Bereich Data Science wuchs in diesem Zeitraum dagegen enorm. Der IDC-Studie zufolge konnten im Jahr 2019 in Deutschland über 100.000 Data-Science-Stellen nicht besetzt werden. Für das Jahr 2025 werden 126.000 unbesetzte Arbeitsplätze vorausgesagt.

In den USA sieht die Situation sehr ähnlich aus: Seit 2017 hat das US-Amerikanische Jobportal *Glassdoor* Data Scientist als gefragtestes Berufsbild in den USA identifiziert. Darüber hinaus hat das *U.S. Bureau of Labor Statistics* festgestellt, dass die Nachfrage nach Data-Science-Fähigkeiten bis 2026 zu einem Anstieg der Beschäftigung in diesem Bereich um 27,9 % führen wird. In den USA erwartet man folglich zwei bis vier Millionen neue Stellenausschreibungen im Bereich Data Science.

Die deutsche Universitäts- und Hochschullandschaft reagiert auf diese Entwicklung. Über den *Hochschulkompass*, ein Informationsportal der Hochschulrektorenkonferenz, lassen sich 2021 über 200 Studiengänge ermitteln, die Data Science im Titel enthalten.<sup>3</sup> 2018 lag die Anzahl an Bachelor- und Masterstudiengängen sowie verschiedener Kontaktstudienangebote bzw. Zertifikatskurse für Data Science noch bei rund 30, wie eine Bestandsaufnahme des Studienangebots für Data Science an den deutschen Hochschulen durch das Institut für Hochschulentwicklung ergab [*Lübcke2018*].

Die Gründe des Fachkräftemangels, englisch auch *„Data Science Gap“*<sup>4</sup> genannt, sind jedoch vielfältig, und dementsprechend gestaltet sich seine Überwindung herausforderungsreich:

An erster Stelle zu nennen sind die hohen Anforderungen an das Kompetenzprofil von Data Scientists (vgl. Kapitel 4): Neben fundierten mathematischen und technischen Kompetenzen fordern Unternehmen und Organisationen zumeist domänenspezifische Expertise sowie eine gute Kenntnis wirtschaftlicher Zusammenhänge, um aus den Ergebnissen die richtigen Schlüsse zu ziehen und sie gewinnbringend zu kommunizieren [*Lübcke2018*]. Auf mathematisch-technischer Ebene erstreckt sich die Liste wichtiger Kompetenzfelder vom Zugriff, der Transformation und Aufbereitung von Daten über den Umgang mit großen Datenmengen und komplexen Datenstrukturen, die Identifikation, Selektion und Gewichtung von Features im Bereich Data Engineering bis hin zur Auswahl, Anpassung und Integration von Tools und von Data-Science-Funktionen in das Software- bzw. Systems Engineering. Neben den für diese Fertigkeiten notwendigen Abschlüssen in Informatik oder Mathematik sollen Data Scientists auch über Soft Skills wie Kommunikationsstärke und Kreativität verfügen, wie eine Analyse des Jobportals *joblift* ermittelte.<sup>5</sup>

Zwar weisen einige größere Unternehmen und Organisationen umfangreiche Data Labs auf, die differenzierte Kompetenzprofile erlauben; zumeist wird vom Berufsbild Data Scientist jedoch erwartet, von einem vertiefenden Verständnis von Datenbanken bis hin zur fachspezifischen Datenvermittlung ein breites Kompetenzprofil vorhanden ist.<sup>6</sup> Zum einen liegt dies an der noch relativen Neuheit der Profession selbst, wodurch sich Unternehmen und Organisationen in vielen Fällen erst Klarheit über ihre tatsächlichen Bedarfe schaffen müssen; es ist zu erwarten, dass sich die berufspraktischen Anforderungen an Data Scientists zukünftig verfeinern [*Lübcke2018*]. Zum anderen ist die Ausbildung eines breiten Kompetenzprofils aber insofern grundsätzlich sinnvoll, als das Berufsbild Data Scientist aufgrund einer rasanten technologischen Entwicklungsgeschwindigkeit einer hohen Änderungsdynamik unterliegt und sich die Schwerpunkte in den Aufgaben und Fertigkeiten laufend verschieben (vgl. Kapitel 3).

1 <https://quanthub.com/data-scientist-shortage-2020/> (27. Januar 2021).

2 <https://datalandscape.eu/study-reports/final-study-report-european-data-market-monitoring-tool-key-facts-figures-first-policy> (27. Januar 2021).

3 Suchergebnisse für „Data Science“ über: <https://www.hochschulkompass.de/home.html> (10. Juni 2021).

4 <https://towardsdatascience.com/the-data-science-gap-5cc4e0d19ee3> (12. April 2021).

5 <https://joblift.de/Presse/data-scientist> (04. Februar 2021).

6 <https://www.cio.de/a/was-ein-data-scientist-wirklich-koennen-muss,3577657> (04. Februar 2021).

### 3 BEGRIFFLICHE EINORDNUNG

Der Begriff ‚Data Science‘ wird in der Praxis häufig recht schwammig verwendet und adressiert unterschiedliche Kompetenzfelder in verschiedener Tiefe. Weitestgehende Einigkeit herrscht allerdings darüber, dass es bei Data Science in der Hochschullehre im Wesentlichen um die Vermittlung von Fachwissen im Hinblick auf den Umgang mit Daten geht – insbesondere mit Big Data. Eine von vielen Praktiker\*innen genutzte Definition stammt von Drew Conway.<sup>7</sup> Diese stellt Data Science als eine domänenübergreifende Disziplin dar, die in der Schnittmenge grundlegende informatische und mathematische Kenntnisse erfordert sowie Sachkenntnis in Bezug auf die Anwendungsdomäne.

Neben Studiengängen mit dem Namen Data Science kamen in der letzten Zeit verwandte Studiengänge auf den Markt, die gewisse Anteile von Data Science besonders vertiefen oder die Perspektive aus einer Disziplin betonen und deshalb andere Namen tragen wie zum Beispiel Data Engineering, Data Analytics oder Machine Learning. Dabei wird Data Science manchmal als übergeordneter und manchmal als untergeordneter Begriff verwendet.

Eine einheitliche Abgrenzung verschiedener Studienfelder im Bereich Data Science kann aber gegenwärtig schon deshalb keinen Bestand haben, weil fortwährende technologische und wissenschaftliche Entwicklungen zu einer ständigen Verschiebung von Schwerpunkten und Aufgabenfeldern führen. Dies lässt sich beispielhaft an dem mit Data Science verknüpften Berufsbild illustrieren.

So hat sich in den letzten Jahren das Verständnis für die Bedeutung der Datenqualität, der Vermeidung systematischer Verzerrungen (Bias), der Daten-Kuration und der Anpassung von Modellen an sich verändernde Daten vertieft. Entsprechend wurde die Notwendigkeit erkannt, diesen Themen mehr Zeit und Arbeit zu widmen [Sambasivan2021]. Gleichzeitig ist das Verständnis für die Besonderheiten verschiedener Modelle gewachsen und parallel nimmt die (Weiter-)Entwicklung von einschlägigen ML-Toolboxen Data Scientists zunehmend Komplexität und Arbeit ab, sodass in die eigentliche Modellbildung heute deutlich weniger Zeit investiert werden muss als noch vor einigen Jahren.

Diese und andere parallele Entwicklungen ziehen wiederum Veränderungen von Organisationsstrukturen und Berufsprofilen nach sich und können durch zukünftige wissenschaftliche und technische Entwicklungen auch noch gänzlich andere Verläufe nehmen.

Die Begrifflichkeiten und Herausforderungen rund um das Thema Data Science sind daher – getrieben von Technik und Wissenschaft – im Fluss und werden es auf absehbare Zeit bleiben [Hachmeister2021]. Eine zu enge begriffliche Fassung bei der Konzeption von Studiengängen läuft daher immer Gefahr von dieser Dynamik überholt zu werden.

Aus diesen Gründen raten wir zur Verwendung eines breiten Oberbegriffs als Titel entsprechender Studiengänge. Etwaige besondere Schwerpunkte in Richtung einzelner Kompetenzfelder – auf dem Arbeitsmarkt werden die oben genannten und weitere Spezialisierungen ja durchaus nachgefragt – sollten über entsprechende Vertiefungen in den Studiengängen kenntlich gemacht werden. Weiterhin empfehlen wir den Titel „Data Science“ für Studiengänge, sofern dies inhaltlich weitgehend gerechtfertigt ist, da dieser a) im allgemeinen Sprachgebrauch ohnehin eine breite Verwendung findet und b) als Oberbegriff eine große Strahlkraft bei Arbeitgeber\*innen und Studieninteressierten hat.<sup>8</sup>

<sup>7</sup> <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram> (15. März 2021).

<sup>8</sup> Dessen ungeachtet beziehen sich die vorliegenden Empfehlungen auch auf verwandte Studiengänge mit abweichenden Namen bzw. können für diese angepasst werden.

## 4 DATA-SCIENCE-KOMPETENZEN

Absolvent\*innen eines Masterstudiums „Data Science“ müssen den *Data Science Lifecycle* (z. B. nach CRISP-DM - *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment*<sup>9</sup> - oder nach der Data-Science-Prozesskette der Plattform *Lernende Systeme [LernendeSysteme2020]*) eines Data-Science-Projekts vollständig und selbständig bewältigen. Das beinhaltet, dass sie die für die einzelnen Schritte notwendigen Werkzeuge und Methoden hinsichtlich ihrer Eignung bewerten, aussuchen und einsetzen sowie auf eine übergreifende Werkzeugunterstützung zurückgreifen können.

Hierfür ist eine Vielzahl unterschiedlicher Kompetenzen erforderlich:

Auf mathematisch-technischer Ebene müssen die mathematischen und algorithmischen Grundlagen aller Schritte des *Data Science Lifecycles* nachvollzogen werden können, was die Beherrschung von mindestens einer für Data Science relevanten Programmiersprache (z. B. Python, R, Julia, Matlab, Octave) und Grundkenntnis in der Softwareentwicklung und dem Datenmanagement voraussetzt.

Da Data Science ein anwendungsbezogenes Fach ist, welches das Ziel hat, handlungsfähige Ergebnisse zu Fragestellungen einer bestimmten Branche zu generieren, müssen Absolvent\*innen darüber hinaus wenigstens eine

Anwendungsdomäne für Data Science und die in dieser Domäne relevanten Fragestellungen sowie wichtige Akteur\*innen kennen und in der Lage sein, die spezifischen Charakteristika der Daten in der entsprechenden Domäne zu beurteilen. Die Ebene des Domänenwissens beinhaltet auch, die Rahmenbedingungen der eigenen Domäne von jeweils anderen unterscheiden zu können und fähig zu sein, sich in andere Domänen einzuarbeiten. Data Scientists müssen auch imstande sein, das für verschiedene Anwendungsdomänen relevante Verständnis der Daten zu entwickeln und die Qualität sowie die Ergebnisse von Auswertungen der Daten (statistisch) interpretieren zu können. Nur so können Data Scientists den letzten Schritt des *Data Science Lifecycles* beherrschen: die disziplinübergreifende und zielgruppengerechte Kommunikation der generierten Ergebnisse.

Im Rahmen der Arbeit innerhalb von Unternehmen und Organisationen müssen Absolvent\*innen schließlich zum einen die Rolle von Data-Science-Anwendungen in diesen Unternehmen und Organisationen beurteilen können, zum anderen die ethischen und juristischen Aspekte von Data-Science-Projekten kennen und ihr eigenes Handeln danach ausrichten. Und natürlich sollte ein Data Scientist in allen genannten Punkten stets interessiert daran sein, Entwicklungen zu erkennen, zu verstehen und das eigene Kompetenzportfolio entsprechend zu erweitern.

<sup>9</sup> <https://www.ibm.com/docs/de/spss-modeler/SaaS?topic=dm-crisp-help-overview> (21. Juni 2021).



## 5 LEHRINHALTE AUS TEILDISZIPLINEN

Es geht hier nicht darum, was in einem Masterprogramm Data Science gelehrt werden soll, sondern um die Inhalte, die zusammen mit einem zugrundeliegenden Bachelorprogramm gelehrt werden sollen. In Kapitel 6 schlagen wir vor, wie ein Masterprogramm strukturiert sein kann, das auf unterschiedlichen konkreten Bachelorprogrammen aufbaut. Der Bezug zu anderen Studiengängen und die Angaben von Leistungspunkten (LP) sollen lediglich ein Indikator für den Grad der Vertiefung darstellen. Es ist in verschiedenen Kombinationen möglich, Kompetenzen gleichzeitig in mehreren Bereichen zu erwerben (zum Beispiel bei einer Kombination der Lehrinhalte ‚Algorithmen‘ und ‚Data-Science-Methoden‘), sodass die Summe der LP auf diese Weise reduziert werden kann.

### 5.1 MATHEMATIK UND STATISTIK<sup>10</sup>

Die Bewältigung der Komplexität von Data Science-Aufgaben erfordert Kenntnisse aus mehreren Gebieten der Mathematik und Informatik. Es ist von enormer Bedeutung, neben anderen Disziplinen generelle mathematische Konzepte zu beherrschen, die für viele Methoden aus dem Data-Science-Bereich eine Grundlage bilden. Des Weiteren kommt es auf die analytische Denkweise und das Abstraktionsvermögen an – Eigenschaften, die mit dem Studium mathematischer Konzepte vermittelt werden.

#### **Lineare Algebra**

Lineare Algebra gehört zur unverzichtbaren Basis der mathematischen Grundausbildung. Angefangen mit Lösungen linearer Gleichungssysteme bis zu den linearen Abbildungen ist dieser Zweig der Mathematik eine Grundvoraussetzung für das Verständnis und die Arbeit mit vielen Machine-Learning-Algorithmen. Themen, die durch lineare Algebra abgedeckt werden, umfassen u. a. Vektoren, Matrizen und Tensoren, lineare Abhängigkeit und Determinanten, Eigenwerte und Singulärwertzerlegung.  
*Empfohlener Umfang: 5 LP*

#### **Optimierung und Simulation**

Inhalte zur Linearen Optimierung werden oft mit Linearer Algebra kombiniert, zur Nichtlinearen Optimierung oft mit Diskreter Mathematik oder mit Numerik.  
*Empfohlener Umfang: 2,5 LP*

#### **Analysis**

Aus der Data-Science-Perspektive ist die Analysis ein weiterer unverzichtbarer Zweig der Mathematik. Zentrale Fragen der Analysis drehen sich um Funktionen: Wie kann das Änderungsverhalten einer Funktion erfasst, verstanden, beschrieben und vorhergesagt werden. Benötigte Kenntnisse umfassen u. a. Mengen und Zahlen, reellwertige Funktionen, Grenz- und Extremwerte und Reihen sowie Stetigkeit und Differenzierbarkeit.

*Empfohlener Umfang: 2,5 LP*

#### **Stochastik**

Die Stochastik bildet einen mathematischen Rahmen für den Umgang mit Unsicherheiten und unsicheren Aussagen. Sie vereint die Gebiete der Wahrscheinlichkeitsrechnung, der Kombinatorik und der mathematischen Statistik. Statistik und Wahrscheinlichkeitstheorie sind grundlegende Komponenten der Datenanalyse und bilden einen wesentlichen Teil der Kompetenzen und Kenntnisse in der Data Science. Dieser Bereich soll den Studierenden einen Einblick in die wichtigsten Statistik- und Data-Analytics-Paradigmen und -Denkschulen bieten. Sie können separat oder als Teil anderer Data-Analytics-bezogener Module oder Kurse unterrichtet werden. Zentrale Konzepte sind hier bspw. nichtparametrische, multivariate, hochdimensionale und bayesianische Verfahren, High-Dimensional Probability, Applied Probability, Zufallsmatrix-Theorie sowie Markow-Prozesse.

*Empfohlener Umfang: 5 LP*

#### **Angewandte Statistik**

Eine wesentliche Aufgabe der Statistik ist es, Daten zu beschreiben und Erkenntnisse für die Entscheidungsfindung in geeigneter Weise aufzuarbeiten und darzustellen. Zu den für Data Science relevanten Inhalten gehören u. a. die beschreibende/deskriptive Statistik, die Konfidenz- und Faktorenanalyse sowie die Anwendung mathematischer Methoden der schließenden/induktiven Statistik, die es erlauben Hypothesen zu testen und Parameter zu schätzen.

*Empfohlener Umfang: 5 LP*

<sup>10</sup> Dieses Kapitel adressiert die Kompetenzfelder 1 und 2 aus [GI2019].

## 5.2 INFORMATIK UND PROGRAMMIERUNG<sup>11</sup>

Um Data-Science-Methoden anwenden und dazugehörige Systeme entwickeln und erstellen zu können, sind neben mathematischen Kompetenzen auch Informatik- und Programmierkenntnisse wichtig. Informatik- und Programmiermethoden sind im *Data Science Lifecycle* auf jeder Stufe in unterschiedlicher Granularität von Bedeutung. Programmier- und Datenbankkenntnisse als auch Software-Engineering-Methoden sind in der *Design&Build*-Phase notwendig, während IT-Infrastrukturen vor allem in der *Release*- und der *Integrate*-Phase gefragt sind.

### Grundlagen Programmierung

Data Scientists müssen die grundlegenden Programmiermethoden und -verfahren kennen und souverän anwenden können. Dazu gehören Konzepte imperativer Programmiersprachen, deren Anwendung zur Programmerstellung sowie Qualitätssicherungsverfahren zur Sicherstellung einer hohen Programmgüte.

*Empfohlener Umfang: 5 LP*

### Fortgeschrittene Programmierung

Während Programmiergrundlagen unabhängig von speziellen Data-Science-Programmiersprachen und -bibliotheken die Basis für die Programmentwicklung legen, benötigen Data Scientists weitergehende Fähigkeiten, um Data-Science-Methoden anwenden und Programme für die Datenanalyse entwickeln zu können. Dazu zählen Programmiersprachen, z. B. Python, R oder Julia, aber auch spezielle Bibliotheken für konventionelle Programmiersprachen. Die Anwendung von Data-Science-Bibliotheken erfordert zudem weitergehende Programmierkonzepte, z.B. aus objektorientierter oder funktionaler Programmierung sowie Parallelisierung.

*Empfohlener Umfang: 2,5 LP*

### IT-Infrastrukturen

Für die Operationalisierung des Data-Science-Lebenszyklus sind Fähigkeiten im Umgang mit IT-Infrastrukturen erforderlich. Dazu gehören das Container-Management und Cloud Computing genauso wie Virtualisierung und Streaming.

*Empfohlener Umfang: 2,5 LP*

### Software Engineering

Die Entwicklung und Erstellung von Data-Science-Lösungen erfordert Konzepte für die ingenieurmäßige Entwicklung softwarebasierter Systeme wie Anforderungsma-

agement, Architektur und Entwurf, Qualitätssicherung (Verifikation und Validation), Software-Projektmanagement und Software-Entwicklungsprozesse.

*Empfohlener Umfang: 5 LP*

### Datenbanken und Big Data

Datenintensive Anwendungen aus dem Data-Science-Bereich bauen auf Datenverarbeitungsketten (Data Pipelines) auf. Data Scientists benötigen Verständnis und Anwendungswissen der Funktionsweise relevanter Algorithmen und Datenstrukturen zur Aufbereitung, Speicherung und Verarbeitung großer, heterogener und dynamischer Datenbestände, u. a. verteilt und in Echtzeit. Außerdem wird ein sicherer Umgang mit unterschiedlichen Datenquellen vorausgesetzt (z. B. Datenbanken, *Repositories*, *Data Warehousing*, NoSQL-Komponenten).

*Empfohlener Umfang: 5 LP*

### Data Governance und IT-Sicherheit

Um die Qualität, den Schutz und die Sicherheit der Daten gewährleisten zu können und für die Einhaltung rechtlicher Vorgaben zu sorgen, werden Data Governance- und Data Policy-Verfahren verwendet. Wichtige Aspekte sind Fragen von Sicherheit, Schutz, Zuverlässigkeit, Integrität, Vertraulichkeit und Verfügbarkeit der genutzten IT-Systeme und der Daten. Dies umfasst die Spezifikation der Anwendungsanforderungen, den Entwurf und die Analyse von Verfahren zur Lösung der gestellten Aufgaben, die Entwicklung von Datenstrukturen und Algorithmen, deren Implementierung in Software und Hardware und den Nachweis dafür, dass das so konstruierte System die gestellten Anforderungen erfüllt.

*Empfohlener Umfang: 5 LP*

### Algorithmen und Komplexität

Die Komplexität eines Algorithmus beschreibt das Laufzeitverhalten und den Speicherbedarf bei Eingaben wachsender Größe. Die Komplexitätsanalyse und der Entwurf möglichst effizienter Algorithmen zusammen mit geeigneten Datenstrukturen gehören zu den Grundbausteinen jedes Informatikstudiums. Die Verarbeitung sehr großer Datenmengen in Data-Science-Projekten ist nur mit effizienten Algorithmen möglich. Data Scientists müssen daher die Komplexität der von ihnen verwendeten Algorithmen einschätzen können sowie Konzepte für den Entwurf effizienter Algorithmen beherrschen, besonders im Bereich Machine Learning.

*Empfohlener Umfang: 5 LP*

<sup>11</sup> Dieses Kapitel adressiert die Kompetenzfelder 3, 4, 5 und 7 aus [GI2019].

### 5.3 DATA-SCIENCE-METHODEN<sup>12</sup>

Viele im Bereich Data Science verwendete Methoden sind nicht spezifisch für diesen Anwendungsbereich entstanden und können daher auch in anderen Bereichen (oft Statistik) vorkommen. Data Scientists müssen aber ein tiefes Verständnis und Anwendungskompetenzen der Methoden haben, wie sie im sog. *Data Science Lifecycle* vorkommen.

Deshalb werden diese Methoden hier als Kerngebiet des Data-Science-Studiums aufgeführt. Der Data Science Lifecycle wird in unterschiedlichen Formen veröffentlicht, seine Schritte haben immer wieder unterschiedliche Namen und beruhen auf verschiedenen Methoden, die wir im Folgenden teilweise nebeneinander aufführen. Neben den einzelnen Kompetenzbereichen ist die Verwendung der Methoden im Data Science Lifecycle ein wesentlicher Lehrinhalt, der aber nicht gesondert aufgeführt wird. Zu allen Methoden sind neben dem Verständnis der Grundlagen die Kenntnis von und der Umgang mit entsprechenden Werkzeugen – auch im Zusammenspiel der Methoden – notwendige Kompetenzen.

#### **Datenintegration und Datenqualität**

Für die Datenintegration werden Daten aus verschiedenen Quellen gesammelt und in unterschiedlichen Formaten miteinander verbunden, aufbereitet und integriert. Zur Anwendung kommen Methoden und Werkzeuge wie Data Pipelines, ETL-Werkzeuge (Extrahieren/Transformieren/Laden), Data-Cleaning-Algorithmen, Matching-Regeln und skalierbare Data-Preparation- sowie Feature-Engineering-Verfahren. Weiterhin ist es erforderlich, dass für die weiteren Schritte des Data Science Lifecycles die Qualität der Daten validiert wird. Dies setzt den Umgang mit fehlenden, fehlerhaften und verzerrten Daten voraus.  
*Empfohlener Umfang: 5 LP*

#### **Datenvisualisierung und Explorative Datenanalyse**

Sowohl die Anwendung von Visualisierungswerkzeugen als auch die theoretischen Grundlagen wie *Perception Theory* und *Editorial Thinking* werden hier vermittelt. Weiterhin werden auch Bezüge zu den Bereichen der Computergrafik hinsichtlich *Overlays*, Interaktivität und Annotationen hergestellt. Explorative Analysemethoden und Methoden der multivariaten Datenvisualisierung gehören ebenfalls zum Modulinhalt.  
*Empfohlener Umfang: 2,5 LP*

#### **Data Mining**

Dieser Bereich soll die Studierenden mit den wichtigsten Data-Mining-Algorithmen und verwandten Methoden der Wissensextraktion, Wissensrepräsentation und Argumentation vertraut machen. Es werden mathematische Modelle und Algorithmen behandelt, die auf große Datenmengen angewendet werden können. Zu den Methoden gehören Prozesse für Häufigkeits- und Assoziationsanalysen (*Rule Mining*), *Knowledge Discovery in Databases* (KDD), *Information Retrieval*, Text-, Web- und Process-Mining, Zeitreihenanalyse und das entsprechende Reporting.

*Empfohlener Umfang: 5 LP*

#### **Maschinelles Lernen**

Den Data Scientists steht eine große Auswahl an fertigen Bibliotheken für Maschinelles Lernen zur Verfügung. Dennoch muss das Studium über die einfache Anwendung von Algorithmen hinausgehen. Neue Probleme, mit denen Studierende konfrontiert sind, erfordern oftmals ein tiefes Verständnis der theoretischen Grundlagen sowohl einfacher als auch fortgeschrittener Algorithmen. Diese Kompetenzen decken die Verwendung, die Analyse, den Entwurf und die Anpassung von Algorithmen für Maschinelles Lernen ab. Es werden wichtige Fragestellungen zur Behandlung und Modellierung von Daten und zur Definition von Zielfunktionen sowie Optimierungsstrategien behandelt. Aspekte, die hier eine Rolle spielen, sind Inferenzmethoden, probabilistische und statistische Modelle/Konzepte, Regressionsmodelle und Machine-Learning-Anwendungen (wie *Natural Language Processing*) sowie einschlägige Programmiersprachen (z. B. Python, R, Julia) und Bibliotheken (z. B. SciKitLearn, Dask).

*Empfohlener Umfang: 10 LP*

#### **Deep Learning und weitere KI-Methoden**

Deep Learning setzt Neuronale Netze mit mehreren Schichten (*Deep Neural Networks*) ein, z. B. zur Bilderkennung oder zum *Natural Language Processing*. Ausgebildete Data Scientists kennen sich mit verschiedenen Deep-Learning-Netzen und -Architekturen aus. Gefordert werden neben der Beherrschung von Deep Learning auch Kenntnisse weiterer Methoden der Künstlichen Intelligenz wie dem *Reinforcement Learning* und generativer Modelle.

*Empfohlener Umfang: 5 LP*

<sup>12</sup> Dieses Kapitel adressiert die Kompetenzfelder 8, 9, 10, 11 und 12 aus [GI2019].

#### 5.4 INTERDISZIPLINÄRE ASPEKTE<sup>13</sup>

Zentrales Wesensmerkmal der Data Science ist ihre Interdisziplinarität. Durch die zunehmende Digitalisierung und Datafizierung [Filipovic2015] aller Wirtschafts- und Lebensbereiche und der damit einhergehenden wissenschaftlichen Auseinandersetzung mit diesen Entwicklungen und Phänomenen werden große Datenmengen erfasst, ausgewertet und für spezifische Zwecke nutzbar gemacht. Der wissenschaftliche und ökonomische Umgang mit Daten erfolgt also nicht im „luftleeren“ Raum, sondern immer in bestimmten Anwendungskontexten. Welchen Wert Daten und welche sozialen Konsequenzen die Ergebnisse von Data-Science-Projekten entwickeln, wird maßgeblich durch den jeweiligen Kontext ihrer Nutzung bestimmt. Dadurch können Daten, die für einen bestimmten Zweck erhoben und verarbeitet wurden, in einem anderen Anwendungsfall eine ganz andere, teils missbräuchliche und schädliche Wirkung entfalten [BMI2019].

Das Ziel einer verantwortungsvollen Datennutzung [Bundesregierung2021] macht vor diesem Hintergrund nicht nur die Aneignung domänenspezifischen Wissens (vgl. Kapitel 5.5) unabdingbar, sondern erfordert auch die fundierte Kenntnis interdisziplinärer Aspekte wie Datenethik oder Data Privacy. So können Data Scientists einerseits dem Anspruch gerecht werden, die Ergebnisse sowie den Nutzen ihrer Arbeit effektiv innerhalb eines ökonomischen oder wissenschaftlichen Umfeldes zu kommunizieren und sich andererseits ihrer gesellschaftlichen Verantwortung in der Verarbeitung sensibler Daten und in der Bereitstellung von Ergebnissen mit teils beträchtlichen sozialen Folgen bewusst zu sein und entsprechend zu handeln.

##### **Datenethik**

Data Scientists müssen von der Datenherkunft bis hin zur Verwertung der Ergebnisse in der Lage sein, den gesamten Lebenszyklus aus ethischer Perspektive zu hinterfragen. Untersucht werden sollen u. a., ob in den ausgewählten Datensätzen systematische Verzerrungen (Bias) vorliegen, beispielsweise bestimmte Personengruppen unter- oder überrepräsentiert sind oder unzulässig diskriminiert werden [BMI2019]. Zudem werden die Rolle der Algorithmen bei der Verschärfung der vorliegenden Verzerrungen und vorgeschlagene Lösungsansätze

diskutiert. Aspekte, die im Rahmen eines Studiums vermittelt werden sollen, sind ein entsprechendes Problembewusstsein, die Bedeutung der Erklärbarkeit von Modellen, die Auswirkung von Verzerrungen in Trainingsdaten, eine Technikfolgenabschätzung oder *Algorithmic-Accountability*-Kriterien wie *Fairness*, *Trustworthiness* etc. [AIHLEG2019].

*Empfohlener Umfang: 5 LP*

##### **Data Privacy und weitere juristische Aspekte**

Um Datenmissbrauch zu verhindern, müssen Data Scientists mit den wichtigsten Datenrechten und -pflichten vertraut sein [BMI2019]. Dafür muss mindestens eine vertiefende Auseinandersetzung mit dem nationalen und internationalen Datenschutzrecht erfolgen, die Spezifika personenbezogener bzw. nichtpersonenbezogener Daten, unterschiedliche Datenlizenzen bzw. Nutzungsbestimmungen sowie Aspekte der *Data Compliance* bekannt sein und entsprechende Handlungskompetenzen erworben werden.

*Empfohlener Umfang: 2,5 LP*

##### **Projektmanagement, Soft Skills, interdisziplinäre Kommunikation**

Als Teil von Organisationen müssen Data Scientists fähig sein, Projekte planen, durchführen und evaluieren zu können. Der Nutzen von Data Science sowie die Vorteile und Risiken der generierten Ergebnisse müssen für den jeweiligen Anwendungskontext erkannt und interdisziplinär verständlich kommuniziert werden. Sinnvoll dafür ist die Vermittlung von Kenntnissen aus den Bereichen Projektmanagement, Marketing und Kommunikation.

*Empfohlener Umfang: 2,5 LP*

##### **Datenökonomie**

Data Scientists müssen die ökonomischen Mechanismen und Zielsetzungen kennen, die die Sammlung und Analyse von Daten antreiben. Dabei geht es nicht nur darum, neue digitale Wertschöpfungsmodelle und Wertschöpfungspotenziale von Daten erkennen und ausschöpfen zu können. Es sollte auch ein kritisches Bewusstsein für die Marktakteur\*innen und Machtverhältnisse in der Datenökonomie und deren gesellschaftlichen Einfluss gefördert werden.

*Empfohlener Umfang: 2,5 LP*

<sup>13</sup> Dieses Kapitel adressiert das Kompetenzfeld 6 aus [GI2019].

## 5.5 ANWENDUNG VON DATA SCIENCE IN DER DOMÄNE<sup>14</sup>

Data Science ist ein anwendungsbezogenes Fach. Der Anwendungskontext von Data Science differiert in den verschiedenen Anwendungsdomänen erheblich. Data Scientists müssen in der Lage sein, selbstständig ein Verständnis für die relevanten Fragestellungen innerhalb einer Anwendungsdomäne zu gewinnen, die Möglichkeiten der Gewinnung von Daten und die Bewertung ihrer Qualität zu eruieren, ein mögliches Vorgehen juristisch, ethisch und ökonomisch zu beurteilen, ein Data-Science-basiertes Projekt selbstständig und unter Verwendung der geeigneten Methoden durchzuführen und die Ergebnisse adressatengerecht zu präsentieren. Dazu ist die selbstständige praktische Durchführung eines Projekts dieser Art im Studium in wenigstens einer Fachdomäne im Team mit Domänenexpert\*innen erforderlich. In einer wissenschaftlichen Domäne spielt das jeweilige Forschungsdatenmanagement eine wichtige Rolle.

Data Scientists sind oft keine Expert\*innen in einer Domäne, sondern müssen die Vielfältigkeit der Fragestellungen und der Rahmenbedingungen in unterschiedlichen Domänen kennen und in der Lage sein, sich selbstständig in eine neue Domäne einzuarbeiten. Dazu muss ein Domänenprojekt innerhalb des Data-Science-Studiums durchgeführt werden. Das Projekt kann auch als Vorstufe für die Masterarbeit verwendet werden. Im Projekt werden Fragestellungen aus der Domäne gemeinsam mit einschlägigen Expert\*innen unter Verwendung von Data-Science-Werkzeugen bearbeitet.

Interdisziplinäre Data Labs, in denen unter Anleitung und selbstständig an konkreten Beispielprojekten gearbeitet wird, eignen sich zur agilen Adaption an methodische Entwicklungen, die teilweise schneller verlaufen, als universitäre Curricula geändert werden können.

*Empfohlener Umfang: min. 15 LP*

---

<sup>14</sup> Dieses Kapitel adressiert die Kompetenzfelder 13 und 14 aus [GI2019].

## 6 CURRICULUMSEMPFEHLUNGEN FÜR MASTERSTUDIENGÄNGE „DATA SCIENCE“

In den hier betrachteten Empfehlungen für Masterstudiengänge „Data Science“ wird ein Teil der erforderlichen Kompetenzen in einem Bachelorstudiengang erworben, dessen erfolgreicher Abschluss Zugangsvoraussetzung für das Masterprogramm ist. Je nach gewählten Zugangsvoraussetzungen gestalten sich daher die Curricula für Masterstudiengänge „Data Science“ unterschiedlich. Abhängig davon, welcher Bachelorstudiengang absolviert wurde, müssen einzelne Studierende unterschiedliche Kompetenzen zu Beginn des Masterstudiengangs erwerben.

Im Folgenden betrachten wir den (häufigen) Fall, dass der Abschluss eines Bachelorstudiengangs in Informatik, Mathematik oder verwandten Fächern eine Zugangsvoraussetzung darstellt. Die dort bereits erlernten Inhalte variieren je nach Hochschule, können aber aufgrund der einschlägigen Empfehlungen der Gesellschaft für Informatik bzw. anderer Fachgesellschaften oder Akkreditierungsagenturen eingeschätzt werden. Wir gehen zudem davon aus, dass ein\*e Bewerber\*in für ein Masterstudium „Data Science“ Wahlmöglichkeiten zur Abdeckung der notwendigen Kompetenzen weitgehend ausgenutzt hat.

Sollte ein Masterstudiengang mehrere Zugangsmöglichkeiten besitzen, so bringen Studierende mit einschlägigem Vorstudium bereits relevante Vorkenntnisse mit, während Studierende aus anderen Fächern diese Inhalte vor oder während des Masterstudiums nachholen müssen. Bei der Curriculumsentwicklung eines solchen Masterstudiengangs können nicht alle eventuell aufzuholenden Inhalte in regulären Modulen vermittelt werden, wenn der Durchschnitt der Vorkenntnisse der Studierenden zu gering ist – je nach Vorstudium würden Studierende zu viel wiederholen müssen. Umgekehrt kann aber auch nicht die Vereinigung der Inhalte aller zugelassenen Bachelorprogramme vorausgesetzt werden, denn dann müsste jede\*r Studierende trotz erfüllter Zugangsvoraussetzungen die Inhalte der jeweils anderen Bachelorstudiengänge nachholen.

Wir empfehlen, für diese Konstellationen passende Angebote zu machen. Sollten die noch fehlenden Inhalte im Rahmen des Masterstudiums nachgeholt werden, so kann dies durch sogenannte Synchronisationsmodule geschehen, also alternative Module, die je nach Bachelorabschluss komplementär zu den bereits erworbenen Kompetenzen sind. Eine andere Möglichkeit sind Module, die inhaltlich Bachelormodule umfassen, aber über

diese hinausgehen, sodass der individuelle Aufwand zur Bearbeitung der Module vom Vorwissen abhängt. Auf den durchschnittlichen individuellen Workload hat dies nur geringe Auswirkungen. Je nach Vorstudium stehen Modulen mit höherem Bearbeitungsaufwand andere Module gegenüber, deren Bearbeitungsaufwand aufgrund der Vorkenntnisse reduziert ist.

Im Folgenden wird für die Bachelorstudiengänge Mathematik und Informatik dargestellt, welche der im vorhergehenden Abschnitt beschriebenen Kompetenzen dem Bachelor- und welche dem Masterstudiengang zugeordnet werden können und wie groß jeweils der verbleibende Umfang für Mastermodule ist.<sup>15</sup> Exemplarisch betrachten wir zusätzlich Bachelorstudiengänge der Wirtschaftsinformatik. Dies lässt sich entsprechend auf andere einschlägige Studiengänge mit Mathematik-, Statistik- und/oder Informatikbezug anwenden. Wir verwenden die folgenden Abkürzungen:

- BM** – Masterstudium Data Science nach Bachelorstudium Mathematik
- BI** – Masterstudium Data Science nach Bachelorstudium Informatik
- BW** – Masterstudium Data Science nach Bachelorstudium Wirtschaftsinformatik

### 6.1 MATHEMATIK UND STATISTIK

#### Lineare Algebra

Kann für **BM**, **BI** und **BW** als bekannt vorausgesetzt werden.

#### Optimierung und Simulation

Kann für **BM** und **BW** als bekannt vorausgesetzt werden, für **BI** nur vereinzelt.

*Empfohlener Umfang: 2,5 LP für BI.*

#### Analysis

Kann für **BM** und **BI** als bekannt vorausgesetzt werden, für **BW** nur vereinzelt.

*Empfohlener Umfang: 2,5 LP für BW.*

#### Stochastik

Kann für **BM** als bekannt vorausgesetzt werden, die Kompetenzen aus **BI** und **BW** reichen im Allgemeinen nicht aus.

*Empfohlener Umfang: 5 LP für BI und BW.*

<sup>15</sup> Auf eine wiederholte ausführliche Beschreibung der einzelnen Studieninhalte wird verzichtet, sie sind in Kapitel 5 zu finden.



### Angewandte Statistik

Kann für **BW** vorausgesetzt werden. Grundlagen können für **BM** und **BI** vorausgesetzt werden, wobei viele Informatikstudiengänge den entsprechenden Empfehlungen nicht folgen (vor einigen Jahren gehörte Statistik noch nicht in den üblichen Fächerkanon der Informatik). Explizite Statistikveranstaltungen kommen im Mathematikstudium meist nicht vor, Grundlagen der Statistik werden dort aber als Teil der Stochastik mitbehandelt. Darüber hinaus werden auch fortgeschrittene Inhalte der Statistik benötigt, die Absolvent\*innen eines Bachelorstudiums Mathematik sich selbst erarbeiten können.

*Empfohlener Umfang: 2,5 LP für **BI**.*

## 6.2 INFORMATIK UND PROGRAMMIERUNG

### Grundlagen Programmierung

Konzepte imperativer Programmierung können für **BM, BI** und **BW** als bekannt vorausgesetzt werden.

### Fortgeschrittene Programmierung

Im Bereich Data Science werden spezielle Programmiersprachen wie Python oder R verwendet. Für **BI** und **BW** kann vorausgesetzt werden, dass Absolvent\*innen sich diese Sprachen ohne Aufwand selbst aneignen können. Zudem sind Fertigkeiten im Umgang mit Programmierungsumgebungen notwendig.

*Empfohlener Umfang: 2,5 LP für **BM** (bei praktischen Übungen mehr).*

### IT-Infrastrukturen

Kann nur für **BI** vorausgesetzt werden.

*Empfohlener Umfang: 2,5 LP für **BM** und **BW**.*

### Software Engineering

Kann für **BI** und **BW** als größtenteils bekannt vorausgesetzt werden, auch wenn nicht alle für Software Engineering relevanten Konzepte (z. B. Anforderungserhebung oder Entwicklungsprozesse) überall in gleicher Tiefe gelehrt werden.

*Empfohlener Umfang: 5 LP für **BM**.*

### Datenbanken und Big Data

Kann für **BI** und **BW** als bekannt vorausgesetzt werden, auch wenn nicht alle für Data Science relevanten Konzepte (NoSQL, Cloud, Big Data) überall gelehrt werden.

*Empfohlener Umfang: 5 LP für **BM**.*

### Data Governance und IT-Sicherheit

Kann für **BI** und **BW** als bekannt vorausgesetzt werden, auch wenn nicht alle für Data Science relevanten Konzepte überall gelehrt werden.

*Empfohlener Umfang: 5 LP für **BM**.*

### Algorithmen und Komplexität

Kann für **BI** als bekannt vorausgesetzt werden, unter Berücksichtigung entsprechender individueller Ausrichtung im Mathematikstudium auch für **BM**.

*Empfohlener Umfang: 5 LP für **BW***

## 6.3 DATA-SCIENCE-METHODEN

Data-Science-Methoden kommen in den hier betrachteten Bachelorstudiengängen in nicht nennenswertem Umfang vor.

### Datenintegration

*Empfohlener Umfang: 5 LP*

### Datenvisualisierung

*Empfohlener Umfang: 2,5 LP*

### Data Mining

*Empfohlener Umfang: 5 LP*

### Maschinelles Lernen

*Empfohlener Umfang: 10 LP*

### Deep Learning und weitere KI-Methoden

*Empfohlener Umfang: 5 LP*

## 6.4 INTERDISZIPLINÄRE ASPEKTE

Interdisziplinäre Aspekte kommen in den hier betrachteten Bachelorstudiengängen in nicht nennenswertem Umfang vor.

### Datenethik

*Empfohlener Umfang: 5 LP*

### Data Privacy und weitere juristische Aspekte

*Empfohlener Umfang: 2,5 LP*

### Projektmanagement, Soft Skills, interdisziplinäre Kommunikation

*Empfohlener Umfang: 2,5 LP*

### Datenökonomie

*Empfohlener Umfang: 2,5 LP*

## 6.5 ANWENDUNG VON DATA SCIENCE IN DER DOMÄNE

*Empfohlener Umfang: 15 LP*

## 6.6 TABELLARISCHE DARSTELLUNG DES CURRICULUMS

Für die drei betrachteten Möglichkeiten eines Masterstudiums „Data Science“ nach Bachelor Mathematik (BM), Informatik (BI) und Wirtschaftsinformatik (BW) werden im Folgenden die Summen der erforderlichen Leistungspunkte berechnet.

KOMPETENZEN	LP	BM	BI	BW
<b>Fortgeschrittene Mathematik und Statistik</b>	<b>20</b>	<b>0</b>	<b>10</b>	<b>7,5</b>
Lineare Algebra	5			
Optimierung und Simulation	2,5		2,5	
Analysis	2,5			2,5
Stochastik	5		5	5
Angewandte Statistik	5		2,5	
<b>Informatik und Programmierung</b>	<b>30</b>	<b>20</b>	<b>0</b>	<b>7,5</b>
Grundlagen Programmierung	5			
Fortgeschrittene Programmierung	2,5	2,5		
IT-Infrastrukturen	2,5	2,5		2,5
Software Engineering	5	5		
Datenbanken und Big Data	5	5		
Data Governance und IT-Sicherheit	5	5		
Algorithmen und Komplexität	5			5
<b>Data-Science-Methoden</b>	<b>27,5</b>	<b>27,5</b>	<b>27,5</b>	<b>27,5</b>
Datenintegration	5	5	5	5
Datenvisualisierung	2,5	2,5	2,5	2,5
Data Mining	5	5	5	5
Maschinelles Lernen	10	10	10	10
Deep Learning und weitere KI-Methoden	5	5	5	5
<b>Interdisziplinäre Aspekte</b>	<b>12,5</b>	<b>12,5</b>	<b>12,5</b>	<b>12,5</b>
Datenethik	5	5	5	5
Data Privacy und weitere juristische Aspekte	2,5	2,5	2,5	2,5
Projektmanagement	2,5	2,5	2,5	2,5
Datenökonomie	2,5	2,5	2,5	2,5
Domänenprojekt	15	15	15	15
<b>Summe</b>	<b>105</b>	<b>75</b>	<b>65</b>	<b>70</b>

### 6.7 ABSCHLIESSENDE EMPFEHLUNGEN

Die Kompetenzen eines Masterstudiengangs „Data Science“ erfordern – zusammen mit einem vorausgehenden Bachelorstudiengang – für die zwingend notwendigen Bestandteile 105 LP, davon fallen 30-40 LP in die Bachelorstudiengänge Mathematik, Informatik oder Wirtschaftsinformatik. Geht man von einer zusätzlichen Vertiefung oder einem Nebenfach im Bereich der Informatik in einem Mathematikstudium (und entsprechenden Vertiefungen in einem Bachelorstudium der Wirtschaftsinformatik) aus, so verbleiben 65 LP für verpflichtende Module im Masterstudium „Data Science“. Umfasst die Zugangsvoraussetzung weitere, verwandte Bachelorstudiengänge, dürfte diese Zahl eher höher ausfallen (bei einem Bachelorstudiengang „Data Science“ oder Statistik eher geringer). In diesen Fällen sollte den Studierenden die Möglichkeit angeboten werden, fehlende Kompetenzen durch zusätzliche Brückenkurse zu erwerben.

Bei einem viersemestrigen Studium stehen 120 LP zur Verfügung, bei einem dreisemestrigen Studium 90 LP. Es bleibt daher nur wenig Raum (25-55 LP) für vertiefende Module und ein Abschlussmodul, was bei einem nicht streng konsekutiven Masterstudiengang allerdings nicht überrascht. Das Abschlussmodul kann entweder eine Abschlussarbeit im Umfang von 30 LP sein, oder aber eine Kombination aus Abschlussarbeit mit 15 LP und einem Data Science Projekt in der Domäne. Unter den weiteren Modulen sollte verpflichtend ein Seminar sein. Die Beschäftigung mit dem aktuellen Forschungsstand bereitet auf die Abschlussarbeit vor. Die Präsentation und ggfs. Ausarbeitung stellen zusätzliche relevante Prüfungsformen dar. Ergänzende Module können zur Profilbildung eines Masterstudiengangs „Data Science“ dienen. Dabei sollen den Studierenden in angemessener Weise Wahlmöglichkeiten angeboten werden.

## LITERATUR

[AIHLEG2019] High-Level Expert Group on Artificial Intelligence (2019): Ethics Guidelines for Trustworthy AI, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>. (Abgerufen am 21.06.2021).

[ASIIN2016] ASIIN e.V. (2016): Fachspezifisch Ergänzende Hinweise des Fachausschusses 12 – Mathematik – zur Akkreditierung von Bachelor- und Masterstudiengängen Mathematik, [https://www.asiin.de/files/content/kriterien/ASIIN\\_FEH\\_12\\_Mathematik\\_2016-12-09.pdf](https://www.asiin.de/files/content/kriterien/ASIIN_FEH_12_Mathematik_2016-12-09.pdf). (Abgerufen am: 10.06.2021).

[BMI2019] Bundesministerium des Innern, für Bau und Heimat (2019): Gutachten der Datenethikkommission, <https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.html>. (Abgerufen am: 13.04.2021).

[Bundesregierung2021] Die Bundesregierung (2021): Datenstrategie der Bundesregierung, <https://www.bundesregierung.de/breg-de/suche/datenstrategie-der-bundesregierung-1845632>. (Abgerufen am: 21.06.2021).

[Filipovic2015] Filipovic, A. (2015): Die Datafizierung der Welt. Eine ethische Vermessung des digitalen Wandels, *Communicatio Socialis*, 48, 1, 6–15. DOI: 10.5771/0010-3497-2015-1-6.

[GI2016] Gesellschaft für Informatik e.V. (2016): Empfehlungen für Bachelor- und Masterprogramme im Studienfach Informatik an Hochschulen, [https://dl.gi.de/bitstream/handle/20.500.12116/2351/58-GI-Empfehlungen\\_Bachelor-Master-Informatik2016.pdf?sequence=1&isAllowed=y](https://dl.gi.de/bitstream/handle/20.500.12116/2351/58-GI-Empfehlungen_Bachelor-Master-Informatik2016.pdf?sequence=1&isAllowed=y). (Abgerufen am: 10.06.2021).

[GI2017] Gesellschaft für Informatik e.V. (2017): Rahmenempfehlungen für die Ausbildung in Wirtschaftsinformatik an Hochschulen, <https://dl.gi.de/handle/20.500.12116/2352>. (Abgerufen am: 10.06.2021).

[GI2019] Gesellschaft für Informatik e.V. (2019): Arbeitspapier Data Science: Lern- und Ausbildungsinhalte, [https://gi.de/fileadmin/GI/Allgemein/PDF/GI\\_Arbeitspapier\\_Data-Science\\_2019-12\\_01.pdf](https://gi.de/fileadmin/GI/Allgemein/PDF/GI_Arbeitspapier_Data-Science_2019-12_01.pdf). (Abgerufen am: 10.06.2021).

[Hachmeister2021] Hachmeister, N., Theiß, J., Weiß, K., & Decker, R. (2021): Balancing Plurality and Educational Essence: Higher Education Between Data-Competent Professionals and Data Self-Empowered Citizens, *Data*, 6, 2, 1-15. DOI: <https://doi.org/10.3390/data6020010>.

[LernendeSysteme2020] Lernende Systeme – Die Plattform für Künstliche Intelligenz (2020): Von Daten zu KI. Intelligentes Datenmanagement als Basis für Data Science und den Einsatz Lernender Systeme, [https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG1\\_Whitepaper\\_Von\\_Daten\\_zu\\_KI.pdf](https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG1_Whitepaper_Von_Daten_zu_KI.pdf). (Abgerufen am 21.06.2021).

[Lübcke2018] Lübcke, M.; Wannemacher, K.: Vermittlung von Datenkompetenzen an den Hochschulen: Studienangebote im Bereich Data Science, *Forum Hochschulentwicklung* 01/2018.

[Sambasivan2021] Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P. K., & Aroyo, L. M. (2021): “Everyone wants to do the model work, not the data work”: Data Cascades in High-Stakes AI, *CHI Conference on Human Factors in Computing Systems*, CHI '21.

## AUTORENSCHAFT

Besonderer Dank gilt folgenden Autorinnen und Autoren (in alphabetischer Reihenfolge), die diese Empfehlungen entwickelt haben:

Abedjan, Ziawasch  
(Leibniz Universität Hannover / GI-FG Datenbanksysteme)

Bendig, Thomas  
(adesso SE)

Brefeld, Ulf  
(Leuphana Universität Lüneburg /  
Plattform Lernende Systeme)

Bürkle, Joachim  
(DB Systel / Deutsche Bahn)

Desel, Jörg  
(FernUniversität Hagen / Sprecher GI-Fachbereich  
Informatik und Ausbildung – Didaktik der Informatik &  
GI-FG Informatik in Studiengängen an Hochschulen /  
Studienkommission Fakultätentag Informatik)

Edlich, Stefan  
(Beuth Hochschule, Berlin)

Eppler, Thomas  
(Hochschule Albstadt-Sigmaringen)

Goedicke, Michael  
(Universität Duisburg-Essen / GI-Vizepräsident &  
Sprecher GI-Task Force „Data Science“)

Hachmeister, Nils  
(Universität Bielefeld / Bielefeld Center for Data Science)

Heidrich, Jens  
(Fraunhofer-Institut für Experimentelles Software  
Engineering IESE / GI-FG Measurement & Data Science)

Höppner, Stephan  
(Atos Information Technology GmbH /  
GI-FG Vorgehensmodelle)

Kast, Stefan M.  
(TU Dortmund / Gesellschaft Deutscher Chemiker)

Krupka, Daniel  
(Gesellschaft für Informatik)

Lang, Klaus  
(TH Bingen / Vorsitzender Fachbereichstag Informatik)

Liggesmeyer, Peter  
(Fraunhofer-Institut für Experimentelles Software  
Engineering IESE / GI-Past-President & Sprecher  
GI-Task Force „Data Science“)

Meisner, Julia  
(Gesellschaft für Informatik)

Scholtes, Ingo  
(Universität Wuppertal / GI-Präsidium)

Tropmann-Frick, Marina  
(HAW Hamburg / GI-FG Data Science for Data Engineering)



GESELLSCHAFT  
FÜR INFORMATIK

---

## GESELLSCHAFT FÜR INFORMATIK E. V. (GI)




### **Geschäftsstelle Bonn**

Wissenschaftszentrum  
Ahrstr. 45  
53175 Bonn  
Tel.: +49 228 302-145  
Fax: +49 228 302-167  
E-Mail: [bonn@gi.de](mailto:bonn@gi.de)

### **Geschäftsstelle Berlin**

Spreepalais am Dom  
Anna-Louisa-Karsch-Str. 2  
10178 Berlin  
Tel.: +49 30 7261 566-15  
Fax: +49 30 7261 566-19  
E-Mail: [berlin@gi.de](mailto:berlin@gi.de)

[gs@gi.de](mailto:gs@gi.de)  
[www.gi.de](http://www.gi.de)

 [/informatikradar](https://twitter.com/informatikradar)  
 [/company/gesellschaft-fuer-informatik](https://www.linkedin.com/company/gesellschaft-fuer-informatik)  
 [/WirSindInformatik](https://www.youtube.com/WirSindInformatik)