



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Bachelorarbeit

Elham Esmat

Konzeption und Implementierung eines Lernpro-
gramms für Big Data

Elham Esmat

Konzeption und Implementierung eines Lernprogramms für
Big Data

Bachelorarbeit eingereicht im Rahmen der Bachelorprüfung

im Studiengang Bachelor of Science Wirtschaftsinformatik
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Prof. Dr.-Ing. Olaf Zukunft
Zweitgutachter: Prof. Dr. Michael Köhler-Bußmeier

Abgegeben am 13.08.2018

Elham Esmat

Thema der Arbeit

Konzeption und Implementierung eines Lernprogramms für Big Data

Stichworte

Big Data, Lernprogramm, Mediendidaktik, E-Learning, Datenschutz, Android-App

Kurzzusammenfassung

Diese Bachelorarbeit befasst sich mit dem softwaretechnischen Entwurf und der Realisierung einer mobilen Android-Applikation, die als Lernprogramm für deutschsprachige Personen gedacht ist, welche sich kaum oder gar nicht mit Big Data und dem deutschen Datenschutzrecht auskennen. Dabei beschäftigt sich diese Arbeit mit der didaktischen Frage, wie Lerninhalte unter Verwendung von Medien gestaltet und präsentiert werden können. Mithilfe von durchgeführten Tests sollen sowohl die Gebrauchstauglichkeit der App als auch die Lernfortschritte bewertet werden.

Elham Esmat

Title of the paper

Design and implementation of an educational software for Big Data

Keywords

Big Data, educational software, media didactics, e-learning, data protection, Android app

Abstract

This bachelor thesis deals with the software design and development of an educational mobile Android application for a German-speaking audience that has minor or lacking knowledge about Big Data and the German data protection law. This thesis also addresses the question of how learning content can be shaped and presented to the students. Moreover, the usability of the app and the learning success is evaluated by means of conducted tests.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Inhaltlicher Aufbau dieser Arbeit	1
2	Einführung in „Big Data“	2
2.1	Begriffsklärung	2
2.2	Abgrenzung zu anderen Begriffen	5
2.3	Anwendung und Hürden von Big Data	7
2.3.1	Big Data-Lebenszyklus	7
2.3.2	Big Data-Nutzung in Unternehmen und Verwaltungen	8
2.3.3	Herausforderungen für Organisationen	10
2.4	Chancen und Risiken für Anwender und Betroffene	12
2.4.1	Das Potential von Big Data-Analysen	13
2.4.2	Maßgebliche Erfolgsfaktoren von Big Data-Anwendungen	15
2.4.3	Abschließende Beurteilung	17
2.5	Rechtliche Aspekte von Big Data	17
2.5.1	Datenschutz und Datensicherheit	17
2.5.2	Big Data im Kontext des Datenschutzrechts	18
3	Mediendidaktische Grundlagen	21
3.1	Lernen mit digitalen Medien	21
3.1.1	Fundamentale Lerntheorien	21
3.1.2	Medien, E-Learning & Lernprogramme	23
3.2	Lehren mittels digitaler Medien	25
3.2.1	Bewertung des Einsatzes digitaler Medien	25
3.2.2	Aufgabenbereich und Methoden der Mediendidaktik	26
4	Anforderungsanalyse	30

4.1	Bewertung und Auswahl der Werkzeuge	30
4.2	Zielgruppe.....	31
4.3	Marktanalyse.....	32
4.4	Anforderungen	33
4.4.1	Lerninhalte & Lernergebnisse.....	33
4.4.2	Nichtfunktionale & didaktische Anforderungen	35
5	Konzeption.....	38
5.1	Beschreibung der fachlichen Komponenten.....	38
5.2	Fachliche Architektur	39
5.3	Technische Konzeption	40
6	Realisierung	44
7	Evaluation	46
7.1	Beschreibung der Testteilnehmer	46
7.2	Usability-Test.....	46
7.2.1	Entwurf und Ablauf	46
7.2.2	Beobachtung und Feedback	47
7.3	Test des Lernerfolgs	47
7.3.1	Entwurf und Durchführung.....	47
7.3.2	Auswertung	48
7.4	Fazit	50
8	Ausblick	52
	Glossar	53
	Literaturverzeichnis.....	54
	Tabellenverzeichnis.....	57
	Abbildungsverzeichnis.....	58

1 Einleitung

1.1 Motivation

Big Data ist ein vielschichtiges Phänomen und spielt nicht nur in der Welt der Informatiker, Unternehmen und Geheimdienste eine wichtige Rolle. Auch stellt es die gesetzgebende und rechtsprechende Gewalt hinsichtlich datenschutzrechtlicher Aspekte vor eine neue Herausforderung. Insbesondere die jüngsten Geschehnisse rund um Abhörskandale und Datenklau in sozialen Netzwerken zeigen den Einfluss, den moderne Datenanalysetechnologien auf die Gesellschaft ausüben können.

Obwohl jeder von den Auswirkungen von Big Data betroffen ist, kennen laut einer Studie von (TNS-Infratest, 2016) nur 25% der befragten deutschen Staatsbürger den Begriff „Big Data“, und lediglich 9% können diesen beschreiben. Einerseits liegt dies daran, dass die Massenmedien Big Data oft als Schlagwort für Ereignisse im Zusammenhang mit digitalen Veränderungsprozessen verwenden und dieser Bezeichnung dadurch mehrere unterschiedliche Bedeutungen zugewiesen werden. Andererseits ist Big Data selbst in der Fachwelt nicht einheitlich definiert, so dass es Außenstehenden nicht leicht gemacht wird, sich darüber unmissverständlich zu informieren.

Es ist deshalb ein Bildungsangebot erforderlich, um diesen Menschen Big Data mitsamt seinen Facetten einfach und schnell erklären zu können. Im Zeitalter des Internets und der Smartphones können derartige Lernangebote als mobile Applikationen angeboten werden, so dass man ohne größeren Aufwand von überall und jederzeit darauf zugreifen kann.

Diese Bachelorarbeit hat das Ziel, genau solch eine Anwendung zu entwerfen und zu implementieren. Dabei sollen mediendidaktische Erkenntnisse dabei helfen, überwiegend technische und rechtliche Lerninhalte zielgruppengerecht als App zu gestalten.

1.2 Inhaltlicher Aufbau dieser Arbeit

Diese Bachelorarbeit ist in vier Blöcke aufgeteilt. Der erste Teil konzentriert sich einerseits auf die Begriffsdefinition und -abgrenzung, die Herausforderungen bei der Umsetzung sowie das aussichtsreiche und gefährliche Potential von Big Data. Andererseits wird zusätzlich der rechtliche und regulative Rahmen von Datenanalysen und seine Auswirkungen auf Big Data betrachtet. Im zweiten Teil wird untersucht, wie der Mensch (mit Medien) lernt und inwiefern didaktische Konzepte für das mediale Lernen entwickelt werden können. Der anschließende dritte Teil umfasst softwaretechnische Schritte zur Spezifikation, zum Entwurf und zur Implementierung eines Big Data-Lernprogramms. Schließlich wird diese Softwarelösung im letzten Teil hinsichtlich software-ergonomischer sowie didaktischer Merkmale bewertet und es werden Erweiterungsmöglichkeiten präsentiert.

2 Einführung in „Big Data“

2.1 Begriffsklärung

Die Bezeichnung „Big Data“ besitzt derzeit noch keine allgemeine Definition. Stattdessen gibt es in der Fachliteratur zahlreiche Beschreibungen davon, was Big Data alles umfasst (King, 2014 S. 34 ff.). Selbst unter ranghohen Wirtschaftsvertretern scheint es hiervon kein einheitliches Verständnis zu geben. Eine Online-Umfrage (SAP, 2012) von 154 Vorstandsmitgliedern verschiedener internationaler Unternehmen hat ergeben, dass die knappe Mehrheit der Befragten Big Data bloß als eine deutliche Erhöhung des Datenvolumens betrachtet, wohingegen sich fast ein Viertel auf das Modell der 3 Vs bezogen hat.

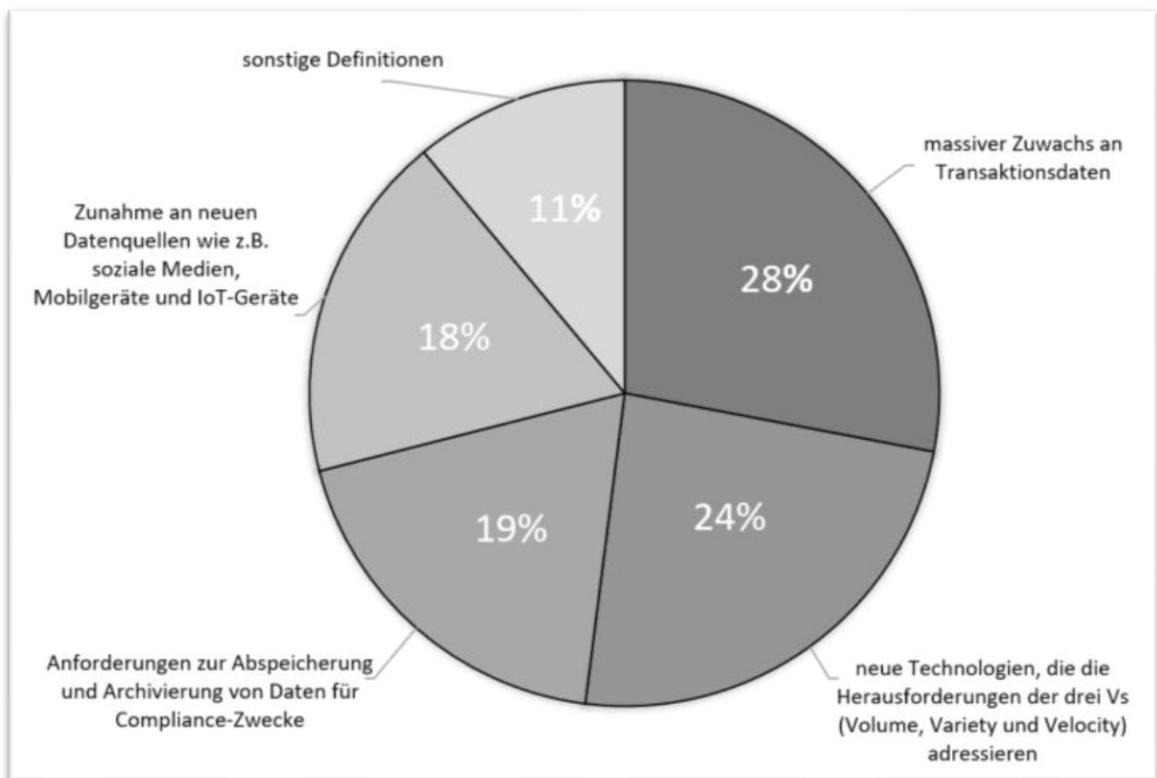


Abbildung 1: Big Data-Definitionen nach einer Umfrage (SAP, 2012) im April 2012 (eigene Darstellung)

Eine derartige Charakterisierung von Big Data ist die am häufigsten anzutreffende Definition, welche erstmals von Laney (Laney, 2001) erwähnt wird und sich auf das damalige Datenmanagement bezieht. Diese Eigenschaften, die sich gegenseitig beeinflussen können und daher nicht unabhängig voneinander sind, werden heutzutage wiederverwendet, um Big Data folgendermaßen zu beschreiben:

Volume

- beschreibt das Ausmaß an Daten und deren Zunahme
- das IT-gestützte Erfassen und Verarbeiten großer Datenmengen wird schon seit Jahrzehnten praktiziert, jedoch ist es heutzutage möglich, fast alles zu messen und zu protokollieren, was zu einer Anhäufung riesiger Datenmengen führt (jenseits des Terabyte-Bereichs)
- nach den Prognosen von Gantz und Reinsel wächst die Gesamtheit der Daten, die jemals erzeugt, gespeichert und kopiert worden ist, exponentiell (von ca. 30 Zetta-byte (ZB) im Jahr 2018 über knapp 50 ZB zwei Jahre später bis hin zu 163 ZB in 2025) (Gantz, et al., 2017 S. 7)
- es gibt keinen Schwellwert, der festlegt, wann Datenmengen als „Big Data“ klassifiziert werden, da Speicherkapazitäten in Zukunft zunehmen werden, so dass Volumen, die heutzutage als „riesig“ empfunden zukünftig als „gewöhnlich“ betrachtet werden (Gandomi, et al., 2015 S. 138-139)
- Außerdem ist die Größe der Datenmenge relativ zu betrachten und abhängig vom Anwendungsfall. Beispielsweise gibt es Anwendungen, die keine enormen Datenvolumen verarbeiten, stattdessen begrenzte Mengen an Daten geschickt miteinander kombinieren und trotzdem als Big Data-Anwendungen klassifiziert werden (van der Sloot, et al., 2016 S. 30).

Velocity

- repräsentiert die Geschwindigkeit der Zunahme, Veränderung und Verarbeitung der Datenmengen
- es wird alle zwei bis drei Jahre zu einer Verdopplung des Datenvolumens kommen (Gantz, et al., 2017 S. 7)
- hinsichtlich begrenzter Speichermöglichkeiten spielt es deswegen eine Rolle, zu klären, ob und welche Daten persistiert werden sollen und ob sie transient sind und schon bei der Erfassung (in „Echtzeit“) im Hauptspeicher analysiert werden können (Hurwitz, et al., 2013 S. 208-209)
- darüber hinaus verstehen Bachmann, Kemper und Gerzer unter Velocity „die Geschwindigkeit, mit der sich Daten und Beziehungen zwischen Daten sowie deren Bedeutung ändern“ (Bachmann, et al., 2014 S. 24-26)

Variety

- steht für die Heterogenität der Daten bezüglich der Quellen, Strukturen und Inhalte
- dies können Kunden- und Transaktionsdaten von Konsumenten, GPS-Bewegungsdaten, von Social Media-Anwendern selber überzeugte Nutzer-, Verhaltens- und Kommunikationsdaten, Sensordaten von IoT-Geräten, Metadaten, usw. sein, um einige Beispiele zu nennen

- solche Daten, die sowohl von Menschen als auch von Maschinen generiert werden, können in strukturierter (tabellarische Daten z.B. in relationalen Datenbanken), semistrukturierter (z.B. XML-Dateien) und in einer nicht vorher festgelegten Form (sogenannte „unstrukturierte“ Daten) auftreten
- die ersten beiden genannten Formen machen nur einen kleinen Anteil am gesamten Datenvolumen aus, während der Großteil sich aus unstrukturierten Daten (wie z.B. Bildern, Audio- und Videodateien) zusammensetzt (King, 2014 S. 37) & (Fasel, et al., 2016 S. 87).
- für die wirksame Verwertung dieser Arten von Daten stoßen relationale Datenbankmanagementsysteme und deren Verarbeitungstechnologien an ihre Grenzen

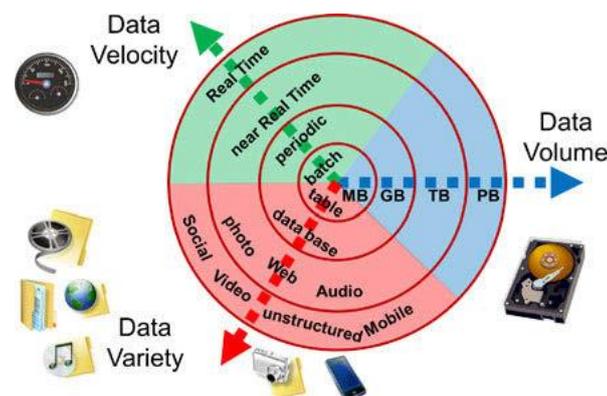


Abbildung II: graphische Darstellung des V-Modells aus (Klein, et al., 2013)

Zusätzlich werden manchmal folgende Eigenschaften zur Erweiterung des Modells herangezogen (Gandomi, et al., 2015 S. 138-139):

Value

- in der ursprünglichen Form (d.h. nach der Erfassung) haben Daten im Verhältnis zu ihrem Volumen einen relativ geringen Wert
- erst durch die Analyse dieser Datenmengen kann ein ökonomischer und gesellschaftlicher Mehrwert sowohl für Unternehmen und Verwaltungen als auch für Kunden und Bürger erzielt werden (van der Sloot, et al., 2016 S. 29-30)

Veracity

- beschäftigt sich mit der Frage, inwieweit die erfassten Daten verifiziert und validiert sind, um Analysen zu ermöglichen, deren Ergebnisse nachvollziehbar sind und nicht zu operationalen und strategischen Fehlentscheidungen führen (Hurwitz, et al., 2013 S. 207-208)
- die Datenanalysen sind daher mit einer Unzuverlässigkeit und Unvollständigkeit verbunden

- nach Angabe von (IBM) betragen die Kosten schlechter Datenqualität allein für die US-Wirtschaft 3,1 Billionen US-Dollar pro Jahr
- um eine hohe Datenqualität zu gewährleisten, muss daher eine Qualitätskontrolle der Daten erfolgen (Fasel, et al., 2016 S. 6) & (van der Sloot, et al., 2016 S. 19)

Letztendlich beschreibt das oben beschriebene Modell nicht nur die Merkmale von Big Data, sondern gibt auch die Anforderungen an Big Data-Anwendungen vor. Andererseits vernachlässigt diese überwiegend technische Definition die Sichtweise von Anwendern (z.B. Domänenspezialisten), welche deswegen mit den drei Fs wie folgt charakterisiert wird (Freytag, 2014 S. 100):

- Fast: Die Spezifikation von Aktivitäten (z.B. einer Analyse) soll zügig stattfinden. Außerdem soll das Ergebnis einer Aktivität so schnell wie möglich von der Ausführungsumgebung generiert werden können.
- Flexible: Bereits spezifizierte Aktivitäten sollen erweiterbar und an neuen Anforderungen anpassbar sein. Organisationen müssen in der Lage sein, auf die hohen Änderungsraten ihrer Geschäftsfelder (bspw. stetig ändernde Kundenanforderungen) schnellstens reagieren zu können.
- Focused: Die maßgeblichen Daten und Verarbeitungsschritte, die erforderlich sind, um das vordefinierte Ziel zu erreichen, sollen mit einem möglichst geringen Aufwand identifiziert und genutzt werden.

Alles in allem gibt es je nach Betrachtungsweise eine unterschiedliche Auffassung darüber, was „Big Data“ beschreibt. Oftmals wird diese Bezeichnung aus technischer Sicht definiert, was naheliegend ist, aber die rechtliche, gesellschaftliche und anwendungsbezogene Perspektive außer Acht lässt.

2.2 Abgrenzung zu anderen Begriffen

Neben der Big Data-Entwicklung gibt es weitere Trends wie Cloud-Computing und Internet of Things (IoT), die einen starken Bezug dazu haben und als Katalysatoren davon wirken. *Abbildung III* stellt den Zusammenhang dieser Begriffe vereinfacht dar.

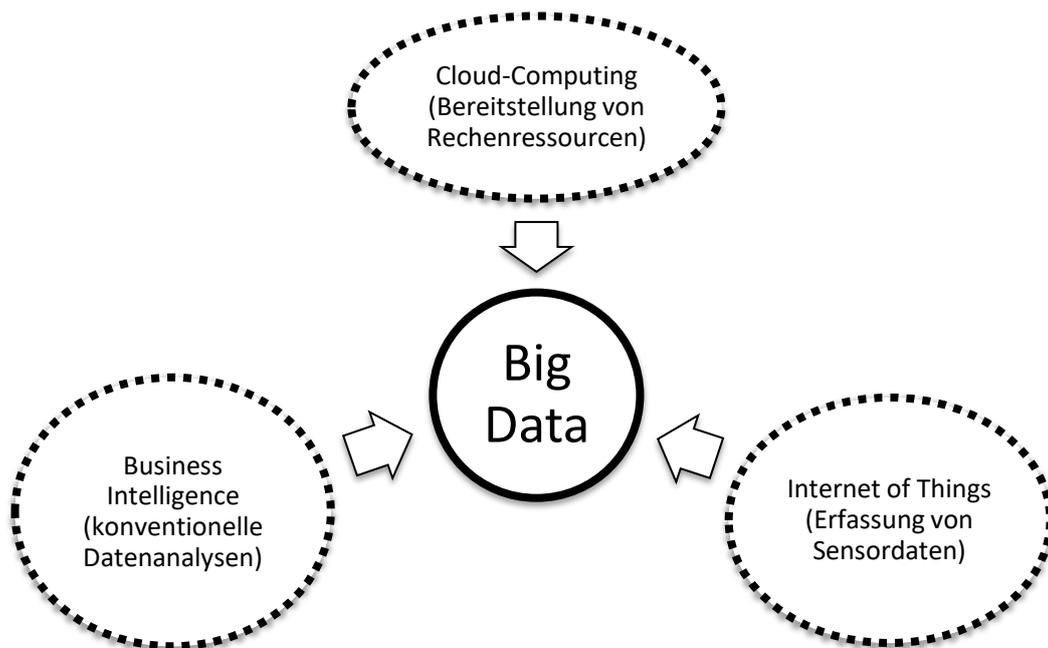


Abbildung III: Vereinfachter Zusammenhang zwischen Big Data, Cloud-Computing und Internet of Things (eigene Darstellung)

Cloud-Computing ermöglicht den auf Nachfrage verfügbaren Netzwerkzugang auf gemeinsam genutzte, konfigurierbare und skalierbare Rechenressourcen (wie z.B. Server, Speicherplatz, Anwendungen, usw.), so dass diese Anwendern als Dienstleistung („... as a Service“) zur Verfügung stehen. Dies kann auf Infrastruktur-, Plattform- und Softwareebene geschehen (NIST, 2011). Infolgedessen muss die Speicherung und Verarbeitung von Daten nicht mehr durch lokale IT-Infrastrukturen durchgeführt und kann stattdessen ausgelagert und über das Internet beansprucht werden. Im Gegensatz zu Big Data zielt Cloud-Computing auf die Umgestaltung von IT-Architekturen ab, wohingegen Big Data die Transformation von Entscheidungsfindungsverfahren anstrebt (van der Sloot, et al., 2016 S. 30). Durch Cloud-Computing und sukzessiv billiger werdender Hardware (Bureau of Labor Statistics, 2015) können Organisationen ihre IT-Systeme nicht nur *vertikal*, sondern auch *horizontal* skalieren.

Nach Sander Klous beschreibt das IoT-Paradigma die massive Zunahme von Sensoren, die in allen Arten von Produkten, Geräten und Maschinen eingebettet und über ein Netzwerk verbunden sind (van der Sloot, et al., 2016 S. 30). Dadurch können jederzeit automatisch Daten über sämtliche Dinge erhoben werden (z.B. der Energieverbrauch in Gebäuden), welche als Grundlage für die Kommunikation und Kooperation zwischen vernetzten IoT-Geräten dienen. Die Anzahl solcher Geräte wird sich voraussichtlich in den nächsten zehn Jahren vervielfachen, so dass IoT zunehmend zum exponentiellen Zuwachs an Daten beitragen wird (Gantz, et al., 2017 S. 11).

Außerdem ist es für ein besseres Verständnis von Big Data hilfreich, eine Abgrenzung zu Business Intelligence (BI) vorzunehmen. Lanquillon und Mallow beschreiben BI als einen Begriff, der die „Konzepte, Prozesse und Technologien zur systematischen Sammlung, Vereinheitlichung, Speicherung, Auswertung und Darstellung von Daten“ umfasst (Dorschel, 2015 S. 256). Der Schwerpunkt von klassischen BI-Verfahren ist die Beantwortung vorher definierter Fragen, indem (strukturierte) historische Daten untersucht werden. Mithilfe von Big Data soll BI dahingehend verändert werden, dass auch enorme und vielfältige (insbesondere unstrukturierte) Datenmengen nicht nur vergangenheitsbezogen, sondern auch gegenwarts- und zukunftsorientiert analysiert werden können. Daher können Big Data-Analysen als Erweiterung von BI betrachtet werden (Dorschel, 2015 S. 257 und S. 307-308). Dementsprechend kann man vier generelle Verarbeitungsarten unterscheiden (Freytag, 2014 S. 99):

1. **Track and Evaluate:** Zustände und Abläufe werden nachverfolgt, erfasst und ausgewertet, um Korrekturmaßnahmen einzuleiten oder Informationen zu übermitteln. Dies kann sowohl mithilfe historischer Daten als auch in Echtzeit durchgeführt werden (Beispiel: Auswertung der Positions- und Temperaturdaten von Wetterballons). Das Reporting („Was ist passiert?“) und Monitoring („Was passiert im Moment?“) können hier eingeordnet werden.
2. **Search and Identify (Information Retrieval):** Unter präziser oder ungenauer Angabe von Attributen werden aus einer Sammlung von Objekten die passenden gesucht, herausgefiltert und selektiert (Beispiel: eine Suchmaschine im Internet).
3. **Analyze:** Durch die systematische Untersuchung von Datenbeständen sollen neue Erkenntnisse abgeleitet werden (z.B. mittels Data Mining-Verfahren wie *Assoziationsanalysen* und *Clustering*), welche u.a. kausale Erforschungen („Warum ist das passiert / passiert das?“) unterstützen.
4. **Predict & Project:** auf der Grundlage der erzielten Ergebnisse der oben genannten Verfahren werden Vorhersagen (prädiktive Maßnahmen) und Vorausplanungen (präskriptive Handlungen) durchgeführt. Ein Beispiel hierfür sind Klimavorhersagen.

Die oben aufgeführten Verarbeitungsarten stellen lediglich eine Möglichkeit dar, um Datenanalysen zu kategorisieren. Alternativ kann man Analysevarianten (z.B. *Predictive Analytics*) bezüglich des Zeitbezugs, der Umsetzungskomplexität, dem potentiellen Wert und dem adressierten Zweck unterscheiden. In (Dorschel, 2015 S. 56 ff.) wird auf diese Unterscheidungsmöglichkeiten näher eingegangen.

2.3 Anwendung und Hürden von Big Data

2.3.1 Big Data-Lebenszyklus

Big Data umfasst mehr als nur den analytischen Aspekt. Dies geht aus dem in *Abbildung IV* aufgeführten Modell hervor, welches den Lebenszyklus von Big Data-Anwendungen darstellt. Dieser Zyklus bezieht sich auf die Merkmale von Big Data, die im V-Modell beschrieben sind und verknüpft diese miteinander.

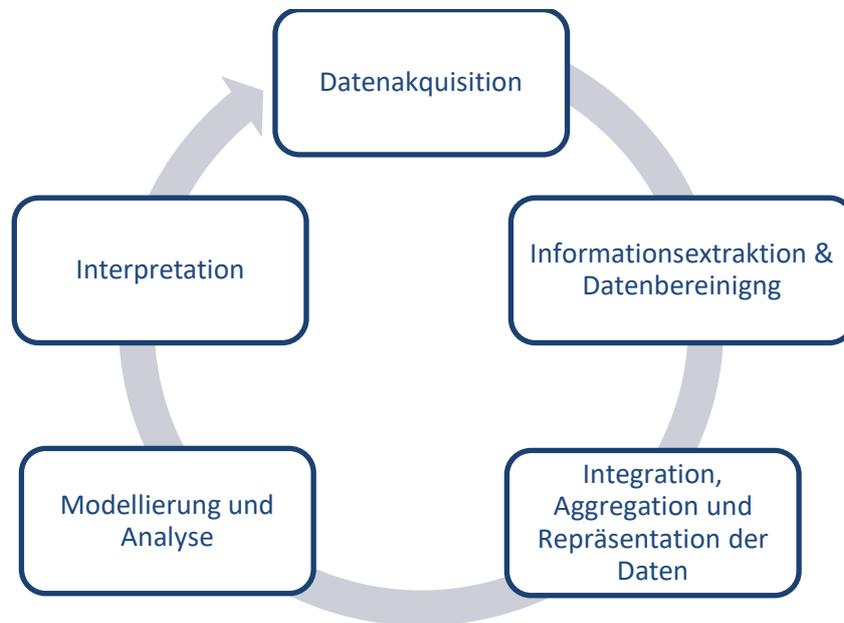


Abbildung IV: Eigene Darstellung des Big Data-Lebenszyklus-Modells aus (Jagadish, et al., 2014 S. 88-90)

Am Anfang werden Daten aus unterschiedlichen Quellen beschafft, unwichtige und fehlerhafte Daten aussortiert bzw. korrigiert, diese vorverarbeiteten Daten zusammengeführt und in einer für Analysezwecke optimierten Form aufbereitet. Die Ergebnisse dieser Analysen werden verifiziert, validiert und fließen schließlich in die Entscheidungsfindung ein. Je nach Anwendung (z.B. Echtzeitanalysen) kann diese Abfolge von Schritten ein kontinuierlicher Prozess sein. Eine umfassendere Erläuterung der einzelnen Schritte des Lebenszyklus ist in (Jagadish, et al., 2014 S. 88-90) zu finden.

2.3.2 Big Data-Nutzung in Unternehmen und Verwaltungen

Zwei im Auftrag von der Wirtschaftsprüfungsgesellschaft KPMG durchgeführten Umfragen deutscher Unternehmen und Verwaltungen haben ergeben, dass mehr als drei Viertel der befragten Unternehmen und Verwaltungen Big Data-Anwendungen bereits einsetzt, zukünftig einsetzen möchte oder den zukünftigen Einsatz davon zumindest diskutiert (vgl. *Abbildung V*). Anders formuliert heißt dies aber auch, dass ca. zwei Drittel der Unternehmen und Verwaltungen (noch) keine Big Data-Strategie umsetzt (KPMG (a), 2016 S. 36-37) & (KPMG (b), 2016 S. 26-27). Im Hinblick darauf lassen sich zweierlei Beobachtungen für die

Privatwirtschaft festhalten (KPMG (b), 2016 S. 38-39): Je größer ein Unternehmen ist, desto höher ist die Wahrscheinlichkeit, dass eine solche Strategie bereits erarbeitet wurde. Außerdem ist der Anteil erarbeiteter Strategien von der Branche abhängig (z.B. 56% im Mediensektor und 34% in der Automobilindustrie).

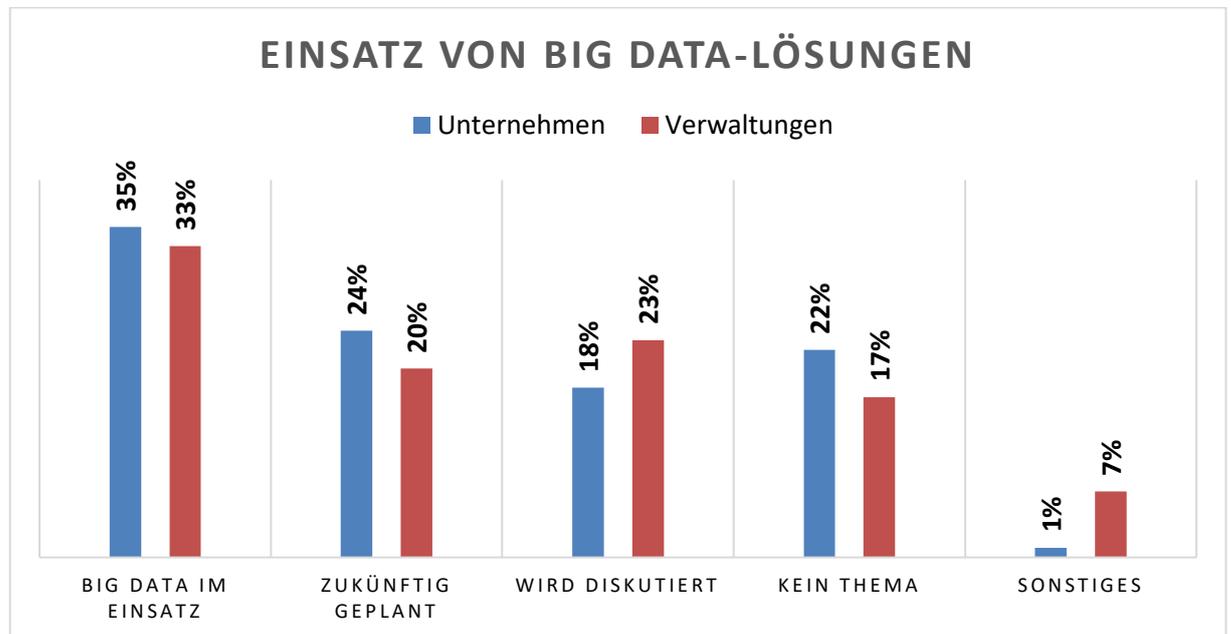


Abbildung V: Umfrageergebnisse deutscher Unternehmen und Verwaltungen, inwieweit diese Big Data einsetzen (Stand: 2016) (KPMG (a), 2016 S. 36-37) & (KPMG (b), 2016 S. 26-27)

Generell steht die große Mehrheit der Unternehmen dem Thema Big Data aufgeschlossen gegenüber, wohingegen nur weniger als ein Zehntel eine kritische bzw. ablehnende Haltung hat (vgl. *Abbildung VI*). Im öffentlichen Sektor ist der Anteil der kritischen Stimmen geringer und knapp die Hälfte der Verwaltungen ist an Big Data interessiert.

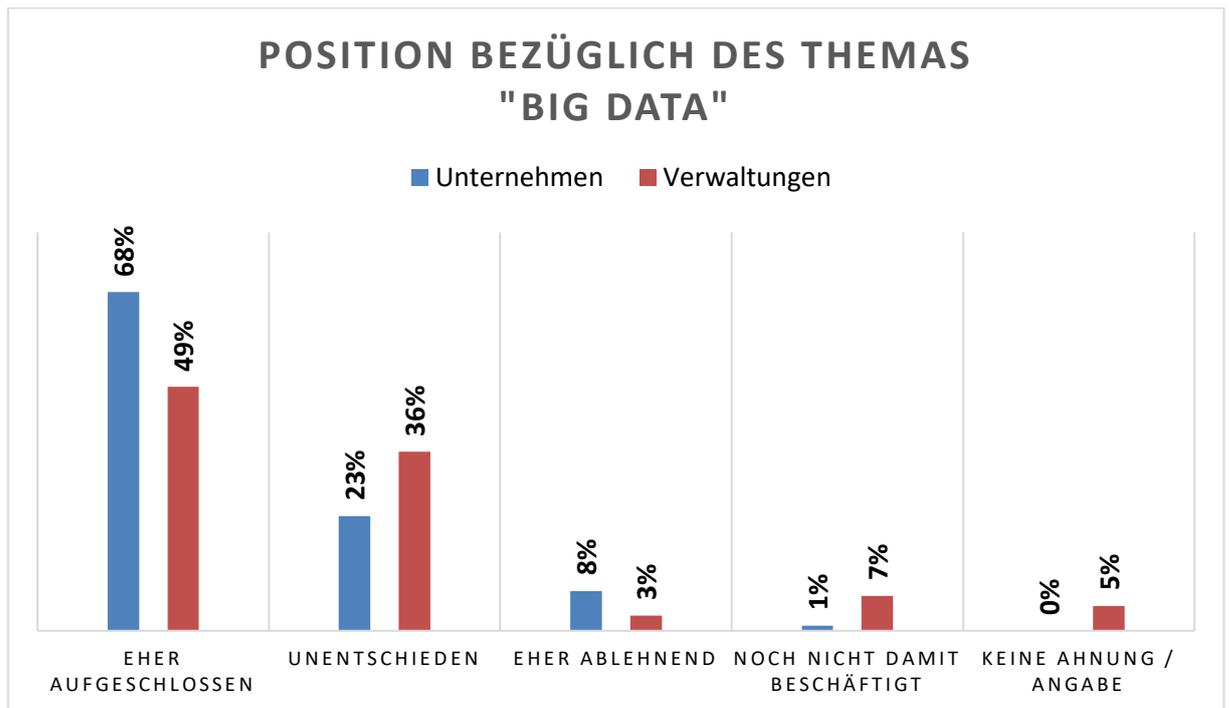


Abbildung VI: Umfrageergebnisse deutscher Unternehmen und Verwaltungen bezüglich ihrer Meinung zu Big Data (Stand: 2016) (KPMG (a), 2016 S. 25-26) & (KPMG (b), 2016 S. 35-36)

2.3.3 Herausforderungen für Organisationen

Trotz der Tendenz zur Verwendung von Big Data, gibt es Vorbehalte gegenüber fortgeschrittenen Datenanalysen. Ungefähr ein Drittel bis die Hälfte aller Unternehmen haben einen Mangel der Rechtsgrundlage, des Budgets, an Fachwissen und -kräften sowie Bedenken hinsichtlich Datenschutz und -sicherheit geäußert (KPMG (b), 2016 S. 26). Bei Verwaltungen ist dies ähnlich, gleichwohl diesbezüglich bis zu zwei Drittel betroffen sind. Aus Datenschutzgründen und der Befürchtung des Kontrollverlusts sind Verwaltungen überwiegend nicht zur Kooperation mit externen Dienstleistern bereit (KPMG (a), 2016 S. 21-23). Stefanie King verallgemeinert derartige Hürden von Big Data und unterscheidet diesbezüglich sechs sich überschneidende Kategorien: Daten, Ethik, Gesellschaft/Kultur, Organisation, Rechtslage und Technologie. In jeder Kategorie werden sowohl die Herausforderungen für Unternehmen und Verwaltungen als auch deren Lösungsansätze von King sehr detailliert beschrieben (King, 2014 S. 82 ff.). *Tabelle 1* soll deswegen einen vereinfachten Überblick ermöglichen. Manche Aspekte, die bereits genannt worden sind bzw. in späteren Abschnitten erläutert werden, werden der Vollständigkeit halber hier noch einmal aufgegriffen. Die hier erwähnten Lösungsansätze sind Beispiele, die sich ggf. auf mehrere Hürden beziehen und daher nicht 1:1 gegenübergestellt werden sollen.

Kategorie	
Hürden	Lösungsansätze
Daten	
<ul style="list-style-type: none"> • fehlende Verfügbarkeit und ein schlechter Zugang zu (aktuellen) Daten erschweren den Einstieg in Big Data • enorme Datenmengen müssen so gefiltert werden, dass Wesentliches von Unwesentlichem getrennt werden kann • Prüfung der Repräsentativität und Zweckhaftigkeit der Ergebnisse von Datenanalysen • Hohe Qualität von Daten, der gelieferten Analyseergebnisse und Möglichkeit der Verarbeitung von Daten • Datensicherheit muss gewährleistet werden, um Daten vor unerlaubtem Zugriff zu schützen 	<ul style="list-style-type: none"> • konkrete Projektziele definieren, um bereits bei der Datenerfassung nur die relevanten Daten zu sammeln, die zur Erreichung der Ziele dienen • um Ergebnisse zu validieren, sollen verschiedene Datenquellen herangezogen, miteinander kombiniert und Analysen wiederholt werden • Förderung von <i>Open Data</i>-Initiativen, die den Zugang zu Daten erleichtern
Ethik	
<ul style="list-style-type: none"> • Interessenskonflikte zwischen Organisationen („Datensammlern“) und „Datenlieferanten“ hinsichtlich der Datenanonymisierung, Privatsphäre und dem Datenmissbrauch • internationale Organisationen werden mit unterschiedlichen Wertevorstellungen konfrontiert 	<ul style="list-style-type: none"> • Vorteile von Datenanalysen gegen den Schutz von Individuen abwägen und auf dieser Basis Richtlinien etablieren • gemeinsame Wertevorstellungen ausarbeiten
Gesellschaft/Kultur	
<ul style="list-style-type: none"> • zum Schutz der Privatsphäre geben Individuen ihre Daten nicht oder nur spärlich preis, was die Arbeit der „Datensammler“ erschwert • dadurch können Dritte (z.B. B2B-Unternehmen) nicht auf diese Daten zugreifen, da sie diese Daten von den „Sammlern“ bekommen und keinen direkten Kontakt zu den Konsumenten haben 	<ul style="list-style-type: none"> • den Konsumenten wird der beiderseitige Nutzen von Datenanalysen vermittelt • der Konsument bekommt eine Gegenleistung für die Bereitstellung seiner Daten • in Absprache mit den Konsumenten stellen vertrauensvolle Vermittler (z.B. Verwaltungen) Daten an Dritte bereit
Organisation	
<ul style="list-style-type: none"> • Big Data-Projekte scheitern am Mangel an kompetenten Fachkräften, begrenzten Budgets und fehlendem Know-how im 	<ul style="list-style-type: none"> • Weiterbildung von Führungskräften und Angestellten, um Fachwissen zu vermitteln und den Stellenwert von Big

<p>Führungspersonal</p> <ul style="list-style-type: none"> • fehlende Bereitschaft / Möglichkeit der verschiedenen Arbeitsbereiche der Organisation, Daten miteinander auszutauschen 	<p>Data besser zu bestimmen</p> <ul style="list-style-type: none"> • Anpassung der Organisationsstruktur, so dass Datensilos beseitigt und zentrale Datenbestände etabliert werden • Outsourcing von Tätigkeiten
Rechtslage	
<ul style="list-style-type: none"> • internationale Rechtslage ist nicht einheitlich • die Eigentumsansprüche an Daten und das Recht an Analyseergebnissen sind (noch) nicht eindeutig geklärt • umfassende Kontrolle, ob und inwieweit Organisationen vorgegebene Richtlinien bezüglich der Privatsphäre, den Dateneigentumsrechten und dem Datenschutz einhalten 	<ul style="list-style-type: none"> • frühzeitige Anpassung von Organisationen auf aktuelle und zukünftige Vorgaben (z.B. durch die Umsetzung von <i>Data Governance</i> und <i>Compliance</i>) • Implementierung organisationsübergreifender Richtlinien
Technologie	
<ul style="list-style-type: none"> • eine effizientere Speicherung und (Echtzeit-) Verarbeitung von überwiegend semi- und unstrukturierten Daten ist erforderlich • Hard- und Software, welche die Anforderungen an Big Data-Systeme nicht ausreichend erfüllen • Systeme in den verschiedenen Arbeitsbereichen der Organisation, die miteinander inkompatibel sind • Aggregation und Visualisierung von Daten 	<ul style="list-style-type: none"> • Akquisition von Fachleuten, die sich mit der Auswahl von und dem Umgang mit modernen Datenverwaltungssystemen und Verarbeitungstechnologien (z.B. <i>NoSQL</i>-Datenbanksysteme) auskennen → derartige Lösungen sollen die konventionellen Systeme nicht ersetzen, sondern erweitern • Outsourcing der IT-Infrastruktur (z.B. durch die Nutzung von Cloud-Computing)

Tabelle 1: Hürden und Lösungsansätze von Big Data in Anlehnung an (King, 2014 S. 148 ff.)

Obwohl diese Herausforderungen den Fortschritt von Datenanalysen erschweren, wächst das Interesse und die Nachfrage nach Big Data-Lösungen. So soll der weltweite Umsatz bis 2020 auf 203 Milliarden US-Dollar steigen, mit einer erwarteten jährlichen Wachstumsrate von 11,7 % (IDC, 2016). Dies bestätigt den Eindruck, dass Big Data sukzessiv an Bedeutung gewinnt.

2.4 Chancen und Risiken für Anwender und Betroffene

Im Folgenden werden sollen sowohl die Vor- und Nachteile als auch Möglichkeiten und Gefahren der Big Data-Nutzung erläutert. Da Big Data-Systeme heutzutage in den verschiedensten Anwendungsbereichen vorkommen, werden die genannten Argumente anhand von aktuellen und historischen Beispielen näher beschrieben. Die hier verwendeten Beispiele sind größtenteils in (O'Neil, 2017) aufgeführt.

Dieser Abschnitt soll nicht nur die Aktualität von Big Data verdeutlichen, sondern auch die Möglichkeiten und den Einfluss moderner Datenanalysen hervorheben. Des Weiteren wird der Fokus auf die Betroffenen von Big Data-Anwendungen gelegt. Auf die rechtlichen Aspekte von Big Data wird im nächsten Abschnitt eingegangen, weswegen diese hier nur am Rande erwähnt werden.

2.4.1 Das Potential von Big Data-Analysen

Mithilfe von Big Data ist es möglich (Auswahl-)Prozesse automatisieren zu können, so dass nicht nur Zeit, sondern auch Geld gespart wird. Big Data-Systeme dienen v.a. zur Entscheidungsunterstützung. Zum Beispiel können Justizbehörden in den USA mithilfe von Modellen das Rückfallrisiko von verurteilten Personen besser einschätzen. Dadurch kann jeder Straftäter individuell und sachlicher beurteilt werden (O'Neil, 2017 S. 38).

Obwohl man sich durch diese Automatisierung mehr Objektivität, eine wissenschaftliche Fundierung und Fehlerminimierung erhofft, werden die entsprechenden Algorithmen letztendlich von Menschen programmiert, so dass voreingenommene Ansichten, Wertvorstellungen und fehlerhafte Annahmen nach wie vor einen Einfluss haben können. In dem oben genannten Beispiel fließen Informationen über das Umfeld des Täters (u.a. die Vorstrafen von Verwandten) zur Bestimmung des Rückfallrisikos mit ein. Dadurch wird dieser nicht nur wegen seiner Taten, sondern auch wegen seiner Herkunft beurteilt. Diese Bewertung, die auf Daten basiert, welche in der US-Justiz nicht als Beweismittel herangezogen werden dürfen, können die Strafzumessung beeinflussen (O'Neil, 2017 S. 41).

Hinsichtlich der Neutralität spielt es deswegen eine wichtige Rolle, welche Kriterien bzw. Eingangsdaten die Analyseverfahren beachten sollen.

Durch die Verwendung von Big Data können Organisationen ihren Umsatz und Konkurrenzfähigkeit erhöhen sowie ihre Kosteneffizienz und Produktivität verbessern. Dies wird durch verschiedene Studien belegt (King, 2014 S. 60 ff.). Zum Beispiel verwerten Speditionen Kamera-, GPS- und Fahrtenschreiber-Daten, um Unfälle zu verhindern, Fahrtwege zu optimieren, den Treibstoffverbrauch zu verbessern und ihre Fahrer zu beurteilen. Des Weiteren können Dienstleistungen und Produkte besser auf Individuen abgestimmt werden. In der Autoversicherungsbranche werden, ausgehend vom gemessenen Fahrverhalten, vorbildliche Fahrer mit Rabatten belohnt (O'Neil, 2017 S. 228-229). Gleichwohl diese Nutzung von Big Data Organisationen eine bessere Überprüfung und Steuerung ihrer Beschäftigten und Kunden ermöglicht, besteht die Gefahr, dass die Privatsphäre der betroffenen Personen verletzt wird und vorgegebene Verhaltensweisen aufgezwungen werden.

Ein weiteres Problem moderner Datenanalysen ist, dass der Fokus bei Personenuntersuchungen nicht auf eine individuelle Bewertung gelegt wird. Um Zeit und Geld zu sparen, wird stattdessen versucht, diese Menschen auf der Basis von *Korrelationen* in vorurteilsgeprägte Kategorien einzuordnen. Generell können erfasste Personendaten jedoch veraltet sein oder zu anderen gleichnamigen Personen gehören (O'Neil, 2017 S. 201-204 und 237-239).

Diese herangezogenen Korrelationen zur Kategorisierung werden üblicherweise von Maschinen entdeckt. Das hat einerseits den Vorteil, dass Muster erkannt werden, die einem Menschen vermutlich nicht auffallen würden, jedoch sind diese gefundenen Zusammenhänge nicht immer für die Anwender nachvollziehbar und überprüfbar, so dass unbewusst Entscheidungen getroffen werden können, die regelwidrig bzw. diskriminierend sein können.

Chance	Risiko
<ul style="list-style-type: none"> • Automatisierung von Geschäftsprozessen • Bessere Ausrichtung von Dienstleistungen und Produkten auf individuelle Bedürfnisse • Optimierung des Ressourcenmanagements • Erhöhung der Reaktionsfähigkeit und der Handlungsschnelligkeit von Organisationen 	<ul style="list-style-type: none"> • Verletzung der Privatsphäre • Benachteiligung bestimmter Personengruppen • Verwechslung von Korrelation und <i>Kausalität</i> • Einschränkung der persönlichen Entfaltung

Tabelle 2: potentielle Vor- und Nachteile von Big Data-Analysen zusammengefasst

Davon abgesehen kann Big Data die Ressourcenplanung optimieren: Beispielsweise gibt es Verfahren, die der Polizei in den USA dabei helfen vorherzusagen, an welchem Ort am wahrscheinlichsten eine Straftat passieren wird. Begrenzt verfügbares Personal kann so effektiver eingesetzt werden, so dass als Folge die Polizeiarbeit verbessert wird (O'Neil, 2017 S. 117-118 und 123). Wenn solche Prognoseprogramme jedoch verwendet werden, um einzelne potenzielle Straftäter zu identifizieren und abzuschrecken, können unschuldige Menschen, die in einer Gegend mit hoher Kriminalität wohnen, fälschlicherweise in den Fokus der Polizei geraten, so dass die freie Entfaltung ihrer Persönlichkeit beeinträchtigt wird (O'Neil, 2017 S. 141-142) & (van der Sloot, et al., 2016 S. 275-276).

Selbst das zukünftige Verhalten von Individuen, die keine persönlichen Informationen über sich preisgeben, kann von ähnlichen Personen abgeleitet werden, die wiederum ihre Informationen offengelegt haben (van der Sloot, et al., 2016 S. 267).

Big Data kann auch dafür eingesetzt werden, um politischen Einfluss auszuüben. Dies wird besonders am aktuellen Beispiel des sogenannten „Microtargeting“ deutlich. Damit ist ein datengestütztes Instrument für Wahlkampf- und Marketingzwecke gemeint, das eine per-

sonalisierte zielgerichtete Kontaktaufnahme mit Individuen und Gruppen ermöglicht (Kolany-Raiser, et al., 2018 S. 2). Auf der einen Seite können dadurch Kosten gespart werden, indem nur diejenigen Personen gezielt angesprochen werden, die als potentielle Käufer bzw. Wähler in Frage kommen, um letztendlich dadurch die Wahl- bzw. Kaufbereitschaft zu erhöhen (Kolany-Raiser, et al., 2018 S. 3-4) & (O'Neil, 2017 S. 259-261). Auf der anderen Seite besteht besonders im Wahlkampf die Gefahr, dass eine kritische Auseinandersetzung und Bewertung der Aussagen von Politikern nicht möglich ist, da jede Zielgruppe eine auf sie abgestimmte Botschaft erhält, welche von der außenstehenden Öffentlichkeit nicht empfangen werden kann. Diese Botschaften an die verschiedenen Zielgruppen können daher nur schwer miteinander verglichen werden, um Widersprüche bzw. Fehler aufzudecken. Schlimmstenfalls werden Zielgruppen manipuliert, um Ängste zu schüren oder den sogenannten *Bestätigungsfehler (confirmation bias)* auszunutzen (Kolany-Raiser, et al., 2018 S. 3-4) & (O'Neil, 2017 S. 254-255).

2.4.2 Maßgebliche Erfolgsfaktoren von Big Data-Anwendungen

Gerade im Baseball hat die Auswertung historischer Daten Spielern und Trainern dabei geholfen, eine geeignete Taktik für jede Spielsituation zu wählen und den Sport weiterzuentwickeln. Derartige statistische Systeme basieren auf relevanten (hinsichtlich der Leistungen von Spielern) und aktuellen Daten und nicht auf Näherungswerten. Diese Daten sind zentral einsehbar und dadurch ist Transparenz gewährleistet. Der Vergleich von Prognosen mit tatsächlichen Spielergebnissen ermöglicht Rückmeldungen an die Modelle, um Optimierungen vorzunehmen und diese weiterzuentwickeln.

Sollten diese Feedbacks ausbleiben oder nur unzureichend erfolgen, können falsche Annahmen nicht entdeckt und notwendige Anpassungen nicht durchgeführt werden, so dass diese Systeme selber bestimmen, was „gut“ oder „schlecht“ ist, ohne jemals zu erfahren, ob die gelieferten Ergebnisse tatsächlich mit der Realität übereinstimmen und bestätigt werden. Diese Systeme basieren jedoch auf Modellen, welche vereinfachte Abbilder des Anwendungsbereichs sind. Deswegen können sie lediglich herausfinden, wer oder was „gut“ oder „schlecht“ sein könnte.

Um die Analyseergebnisse validieren zu können, muss geprüft werden, welche Zuordnungen richtig oder falsch sind. Als Qualitätsmaße dienen hier *Precision* und *Recall*. Mithilfe einer Validierung kann u.a. festgestellt werden, ob Kausalität mit Korrelation verwechselt worden ist.

Zudem können Verfahren, die nur mit Näherungswerten als Indikatoren arbeiten, überlistet werden. Angenommen, ein Personaler möchte jemanden im Bereich Social Media-Marketing einstellen. Um nicht alle Bewerber auf ihre Tauglichkeit prüfen zu müssen, was aufwändig ist, wird eine Vorauswahl auf der Basis der *Twitter-Follower* getroffen. Die Annahme ist, dass Personen mit einer hohen Anzahl an Abonnenten erfolgreich in den sozialen Medien und deswegen geeigneter für die Stelle sind. Wenn jedoch diese Art der Vorauswahl manchen Bewerben bekannt ist, könnten sie ihre Followerzahl durch bezahlte Bots

stark verbessern, so dass letztendlich der beste Betrüger profitiert und nicht der geeignete Kandidat (O'Neil, 2017 S. 78-79).

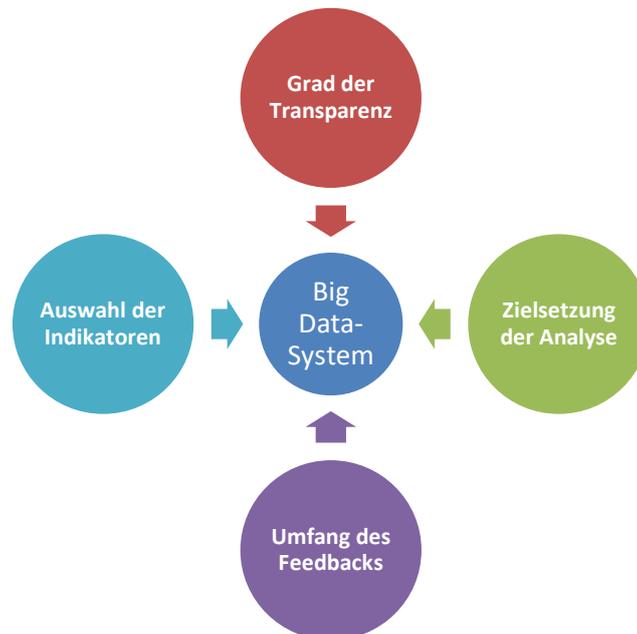


Abbildung VII: prägende Aspekte für den erfolgreichen Einsatz von Big Data-Systemen (eigene Darstellung)

Big Data-Analysen, die nicht transparent sind, führen dazu, dass betroffene Personen das Zustandekommen von (fehlerhaften) Ergebnissen nicht nachvollziehen und Analyseergebnisse ggf. nicht revidieren können, besonders wenn diese auf der Basis von korrekten Daten erfolgen (van der Sloot, et al., 2016 S. 267). Und im Falle einer Anfechtung liegt die Beweislast bei den Betroffenen, obwohl das eigentlich andersherum sein müsste (O'Neil, 2017 S. 21).

Auch die Zielsetzung entscheidet über den erfolgreichen und ethisch vertretbaren Einsatz von Big Data (vgl. *Abbildung VII*). Anhand der folgenden Beispiele ist dies gut erkennbar: Big Data kann dafür genutzt werden, Lieferketten von Unternehmen zu untersuchen, um mögliche Zwangsarbeit ausfindig machen zu können. Ein derartiges System kann von Hilfsorganisationen dafür eingesetzt werden, um betroffene Unternehmen darauf hinzuweisen, so dass diese nähere Untersuchungen und ggf. Anpassungen durchführen können (O'Neil, 2017 S. 293-294).

Anders betrachtet können Informationen über notleidende Menschen, z.B. diejenigen, die nach Informationen über Essensmarken im Internet suchen, dafür genutzt werden, bei diesen Personen Anzeigen von privaten Vermittlerdiensten zu schalten, welche nach der Vermittlung die erfassten Kontaktdaten von ihnen sammeln und für Werbezwecke an Dritte verkaufen. Bedürftigen werden dann Dienstleistungen aufgedrängt, die sie sich nicht leisten können (O'Neil, 2017 S. 279-280). Dadurch profitieren Suchmaschinen und Unternehmen

auf Kosten von Notleidenden. Sobald also Big Data-Anwendungen den alleinigen Fokus auf Effizienz und Profite legen, passiert dies oft auf Kosten der Gerechtigkeit.

2.4.3 Abschließende Beurteilung

Zusammenfassend kann man sagen, dass sich die Nutzung von Big Data in sämtlichen Bereichen der Wertschöpfungskette wiederfinden lässt und der Wirtschaft sowie dem öffentlichen Sektor ermöglicht, die Produktivität zu steigern, Kosten einzusparen und Geschäftsprozesse effizienter zu gestalten. Dennoch birgt die Datenverarbeitung im Zusammenhang mit Big Data Risiken, die die Gleichbehandlung und Privatsphäre von Menschen gefährden. Deswegen muss der potentielle Schaden, der durch Big Data-Systeme entstehen kann, insoweit abgewogen werden, dass die möglichen Kosten und Nutzen für Anwender und Betroffene gegenübergestellt werden und auf Grundlage dieser Betrachtung über den Umfang des Einsatzes derartiger Systeme entschieden wird.

2.5 Rechtliche Aspekte von Big Data

Am 25. Mai 2018 werden sowohl die EU-Datenschutz-Grundverordnung (DS-GVO) als auch die Neufassung des Datenschutzgesetzes der Bundesrepublik Deutschland (BDSG-neu) in Kraft treten und das aktuelle deutsche Datenschutzgesetz (BDSG) ersetzen. In den nachfolgenden Ausführungen wird erläutert, inwiefern diese Richtlinien die Nutzung von Big Data reguliert, welche Rechte Individuen besitzen, inwieweit Unternehmen und Verwaltungen davon betroffen sind und welche Fragestellungen noch offenbleiben.

2.5.1 Datenschutz und Datensicherheit

Nach Auffassung des Bundesverfassungsgerichts (BVerfG, 1983) ist der Datenschutz eng mit dem Prinzip der informationellen Selbstbestimmung verknüpft. Dieses stützt sich auf das Grundrecht der freien Entfaltung der Persönlichkeit und der Unantastbarkeit der Menschenwürde.¹ Demzufolge sollen Individuen selber entscheiden, ob bzw. inwieweit sie persönliche Informationen preisgeben und zur Verarbeitung zulassen möchten. Daher bezeichnet der Datenschutz den Schutz personenbezogener Daten von natürlichen Personen. Unter informationeller Selbstbestimmung ist aber kein „Eigentumsrecht“ der Individuen an ihren Daten gemeint: Beispielsweise „gehören“ resultierende Daten einer Fahrzeuginstandhaltung nicht dem Fahrzeughalter, sondern auch der Kfz-Werkstatt (van der Sloot, et al., 2016 S. 263).

Nach Art. 4 Nr. 1 DS-GVO sind „personenbezogene Daten“ im rechtlichen Sinne diejenigen Informationen, mit denen man *natürliche Personen* direkt oder indirekt eindeutig identifizieren kann (z.B. Kontaktdaten, Gesundheitsdaten, Kontonummern, usw.). Demgegenüber

¹ Art. 2 Abs. 1 Grundgesetz (GG) und Art. 1 Abs. 1 GG

sind „anonyme Daten“ Informationen, welche entweder keinen Bezug auf natürliche Personen haben oder personenbezogen sind, aber aufgrund ihrer Anonymisierung eine eindeutige Identifikation jener Personen nicht ermöglichen (König, et al., 2017 S. 163). Bei der Verarbeitung derartiger Daten greift der Datenschutz nicht.

Im Vergleich zum Datenschutz konzentriert sich die Datensicherheit darauf, den unberechtigten Datenzugriff durch Dritte, die Manipulation und den unwiederbringlichen Verlust von Daten, die nicht unbedingt einen Personenbezug haben müssen, zu verhindern. Beide Bezeichnungen haben jedoch einen Schnittbereich. So beinhaltet die Datensicherheit die technische und organisatorische Umsetzung des Datenschutzes. Gemäß der Anlage zu § 9 BDSG umfassen diese Maßnahmen u.a. eine Kontrolle des Zutritts zu Datenverarbeitungsanlagen, der Weitergabe und der Verfügbarkeit von personenbezogenen Daten.

2.5.2 Big Data im Kontext des Datenschutzrechts

Gemäß Art. 2 Nr. 1 DS-GVO und § 1 Abs. 1 BDSG-neu gilt das Datenschutzrecht sowohl für öffentliche Stellen als auch für private Organisationen, die personenbezogene Daten automatisiert verwalten oder verarbeiten. Für spezifische Bereiche gibt es spezielle Richtlinien wie z.B. die sogenannte „ePrivacy-Verordnung“, das Telemediengesetz und das Telekommunikationsgesetz, welche als Ergänzung dienen (König, et al., 2017 S. 155). Das Fundament davon bildet die Datenschutz-Grundverordnung, welche u.a. Grundsätze zur Verarbeitung von personenbezogenen Daten, Rechte der Betroffenen, Ausnahmeregelungen und (verschärfte) Sanktionen bei Rechtsverletzungen² festlegt. *Tabelle 3* liefert eine einfache (d.h. ohne Berücksichtigung von Details und Ausnahmen) Übersicht der Betroffenenrechte und Verpflichtungen für Verantwortliche.

Rechte von Betroffenen	Verpflichtungen der Verantwortlichen
<ul style="list-style-type: none"> • die Einwilligung zur Datenverarbeitung kann jederzeit widerrufen werden • gegen die Verarbeitung der Daten kann ein Widerspruch eingelegt werden • Auskunftsrecht über die Verwendungszwecke, Herkunft, Datenempfänger und Kategorien ihrer verarbeiteten Personendaten • Recht auf Berichtigung unrichtiger Angaben sowie Löschung und Verarbeitungseinschränkung ihrer personenbezogenen Daten 	<ul style="list-style-type: none"> • für die Datenverarbeitung wird eine Einwilligung der Betroffenen benötigt • personenbezogene Daten müssen rechtmäßig, nur für einen vorher festgelegten Zweck³ und für die Betroffenen transparent verarbeitet werden • die Daten müssen aktuell sowie richtig sein und auf eine sichere Weise (Datensicherheit) verarbeitet werden • die Menge an verarbeiteten Daten muss dem Zweck angemessen sein und darf nur bis zur Zweckerfüllung verwaltet

² nach Art. 83 DS-GVO können diese Geldbußen in Höhe von bis zu 20 Mio. € bzw. 4 % des weltweiten Jahresumsatzes sein

³ Explorative Big Data-Analysen sind deshalb nicht zulässig. vgl. (König, et al., 2017 S. 167)

<ul style="list-style-type: none"> • Recht darauf, die bereitgestellten personenbezogenen Daten in einem maschinenlesbaren Format zu erhalten und diese Dritten zur Verfügung zu stellen • Recht darauf, sich keinen Entscheidungen zu fügen, die auf einer reinen automatischen Verarbeitung beruhen und zu einer Beeinträchtigung führen 	<p>werden (danach müssen sie gelöscht werden)</p> <ul style="list-style-type: none"> • die Einhaltung der oben genannten Pflichten muss nachgewiesen werden • Information der Betroffenen über die Erhebung und Verarbeitung ihrer Daten • alle Datenempfänger müssen über die Löschung, Berichtigung und Verarbeitungseinschränkung informiert werden
--	---

Tabelle 3: Rechte von Betroffenen und Pflichten von Verantwortlichen nach Art. 5-23 DS-GVO (eigene Darstellung)

Wie Hornung und Herfurth bereits festgestellt haben, gibt es in der DS-GVO keinen direkten Bezug zu Big Data und dazugehörige Anwendungen (König, et al., 2017 S. 157). Im Zusammenhang mit dem Datenschutzrecht ist ein generelles Problem, dass Unternehmen zur Einwilligung der Datenverarbeitung *allgemeine Geschäftsbedingungen (AGB)* verwenden, welche jedoch von den wenigsten Betroffenen tatsächlich gelesen oder verstanden, aber dennoch akzeptiert werden (Vodafone, 2016 S. 8-9) & (Fasel, et al., 2016 S. 89). Dadurch wird es den Unternehmen leicht gemacht, sich die notwendige Zustimmung zu beschaffen.

Weitere Probleme, die insbesondere bei der Big Data-Nutzung entstehen, werden im Folgenden geschildert: Wie im letzten Abschnitt bereits erwähnt, greift der Datenschutz nur dann, wenn personenbezogene Daten verwendet werden. Durch eine Anonymisierung der Daten können sich Verantwortliche von den datenschutzrechtlichen Verpflichtungen befreien. Aufgrund der Datenvielfalt und der zur Verfügung stehenden Big Data-Technologien können anonymisierte Daten jedoch mit vielen weiteren Daten derart verknüpft werden, um eine Identifikation von Personen zu ermöglichen. Nach der DS-GVO wird das Interesse und v.a. der Aufwand der Verantwortlichen betrachtet, um Personen direkt oder indirekt zu identifizieren. Wann dieser „Grad der Deanonymisierung“ erreicht wird, ist den Verantwortlichen nicht immer klar, so dass dies zu einer gewissen Unklarheit führt (König, et al., 2017 S. 163-166).

Außerdem können Daten mithilfe von Big Data-Verfahren für mehrere Zwecke verwendet werden. Hinsichtlich des Grundsatzes zur Zweckbindung kann es für die Verantwortlichen anspruchsvoll sein, im Vorhinein all diese Zwecke anzugeben, da einerseits aktuelle und (noch nicht bekannte) zukünftige Verwendungszwecke mitberücksichtigt werden müssen und andererseits Verwendungszwecke sich zwischenzeitlich ändern können. Jedoch ist unter bestimmten Bedingungen eine Zweckänderung im Projektverlauf möglich (König, et al., 2017 S. 168-169). Beispielsweise ist die Zweckbindung bei Analysen zu wissenschaftlichen Forschungszwecken aufgelockert. Wie die Bezeichnung „wissenschaftliche Forschungszwecke“ ausgelegt ist, geht aus der DS-GVO nicht hervor, so dass hierin eine Möglichkeit besteht, personenbezogene Daten mithilfe von Big Data in der privaten Forschung umfassender zu verarbeiten (König, et al., 2017 S. 161-162).

Darüber hinaus stellt Big Data zwei Herausforderungen bezüglich der Gewährleistung von

Transparenz dar. Zum einen bestehen Interessenskonflikte zwischen Betroffenen, welche ihr Auskunftsrecht in Anspruch nehmen möchten, um sich über Big Data-Verfahren zu informieren, und Unternehmen, die ihre Betriebs- und Geschäftsgeheimnisse wahren möchten und deswegen auf die entsprechenden rechtlichen Beschlüsse verweisen. Laut Hornung und Herfurth können hier Geheimhaltungsvereinbarungen und unabhängige Vermittler Abhilfe schaffen (König, et al., 2017 S. 170-171). Zum anderen ist es schwierig, den Betroffenen die Datenverarbeitung nachvollziehbar zu erklären, wenn diese nicht das nötige Fachwissen besitzen, um die dahinterliegenden Algorithmen zu verstehen. Selbst wenn man sachverständige Dritte hinzuzieht, die als Berater dienen, muss geklärt werden, wer diese bezahlt (König, et al., 2017 S. 171-173).

Letztendlich stoßen die Ansätze des (neuen) Datenschutzrechts bei weiterführenden Fragestellungen insbesondere zum Thema „Big Data“ an ihre Grenzen. Dennoch ermöglicht die DS-GVO betroffenen Personen einen höheren Schutz und Verantwortlichen die Schaffung von mehr rechtlicher Klarheit.

3 Mediendidaktische Grundlagen

3.1 Lernen mit digitalen Medien

Dieser Abschnitt erläutert die grundlegenden Prinzipien des Lernens, definiert die Begriffe *Lernen*, *Medien* und *E-Learning* und bringt diese in Zusammenhang miteinander.

3.1.1 Fundamentale Lerntheorien

Das Lernen bezeichnet Verhaltensänderungen, die sowohl auf Umwelterfahrungen als auch auf den Erwerb und die Veränderung von Sach- und Handlungswissen (deklaratives und prozedurales Wissen) beruhen (Nieding, et al., 2015 S. 37-38 & 45).¹ Im Folgenden wird ausschließlich das Lernen von Lebewesen (insbesondere am Beispiel des Menschen), aber nicht *maschinelles Lernen* oder ähnliche Themen betrachtet.

Um die Prozesse und Art und Weise des Lernens (aus psychologischer Sicht) beschreiben und erklären zu können, sind Lerntheorien entwickelt worden. Im Zusammenhang mit medialem Lernen werden in der entsprechenden Literatur die klassischen Ansätze Behaviorismus, Kognitivismus und Konstruktivismus untersucht, welche sich nicht gegenseitig ausschließen. *Abbildung VIII* liefert eine kompakte Übersicht dieser Theorien:

In behavioristischen Theorien wird das Lernen als ein Prozess betrachtet, in dem Änderungen des beobachtbaren Verhaltens nicht durch innere Vorgänge, sondern durch Konsequenzen gesteuert werden, die wiederum durch Reaktionen auf vorausgegangene Umweltreize erfolgt sind (Kerres, 2012 S. 112-113), (Nieding, et al., 2015 S. 39) & (Minass, 2002 S. 14-15). Das Verhalten des (passiv agierenden) Lernenden kann daher belohnt werden, um dieses zu fördern und eine Assoziation zwischen dem Reiz und der Belohnung herzustellen.² Umgekehrt kann das Verhalten unter Verwendung von Bestrafungen auch abgewöhnt werden.

Der Kognitivismus geht davon aus, dass das Lernen nicht ausschließlich eine Verkettung von Reizen und Reaktionen, sondern ein Prozess ist, bei dem der Lernende Informationen verarbeitet und Vorgänge bezüglich der Wahrnehmung, des Denkens und des Gedächtnisses eine entscheidende Rolle spielen. Dementsprechend hängt der Lernerfolg von den bereits angeeigneten Erfahrungen und Kenntnissen des Lernenden ab. Die Lernenden speichern abstraktes Wissen ab, welches dann bspw. zur Problemlösung abgerufen werden kann (Kerres, 2012 S. 119 & 124) & (Nieding, et al., 2015 S. 42-43).

¹ (Stangl, 2018) liefert eine umfassende Übersicht über weitere Definitionen. Um ein einfaches und einheitliches Verständnis zu schaffen, beziehen sich die nachfolgenden Ausführungen lediglich auf die oben genannte Begriffsbeschreibung.

² Ein bekanntes Beispiel eines behavioristischen Ansatzes sind die Experimente des Psychologen Iwan Pawlow (Wikipedia, 2017)

Im Konstruktivismus wird beim Lernen das Wissen für jede Situation neu konstruiert, indem eine subjektive und individuelle Wahrnehmung und Interpretation eingehender Informationen auf der Grundlage des Vorwissens des Lernenden stattfindet. Demnach ist der Lernprozess in einen (soziokulturellen) Kontext eingebettet und ein Vorgang, um mit Situationen lösungsorientiert umgehen zu können (Meir, 2006 S. 14), (Kerres, 2012 S. 89-90 & 124-128) & (Nieding, et al., 2015 S. 48).

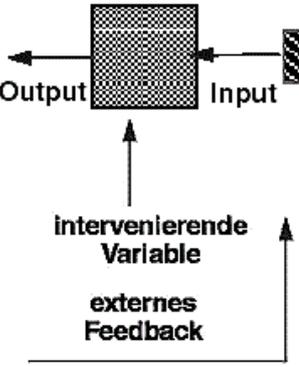
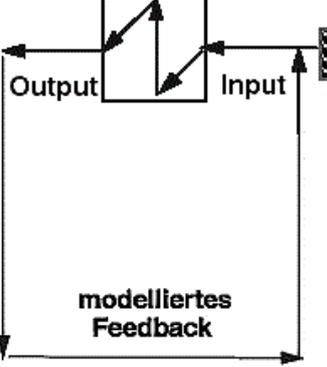
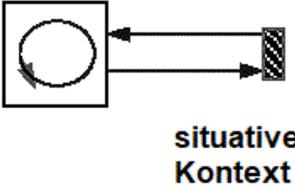
Behaviorismus	Kognitivismus	Konstruktivismus
<p data-bbox="411 577 528 645">Hirn ist „black box“</p>  <p data-bbox="304 1061 584 1173">Pro: es erfolgt stets eine Rückmeldung an den Lernenden</p> <p data-bbox="304 1305 584 1541">Contra: schwer übertragbar auf Situationen, in denen komplexere Lerninhalte angeeignet werden</p>	<p data-bbox="632 566 927 645">interne Verarbeitungsprozesse interessieren</p>  <p data-bbox="616 1061 951 1256">Pro: Fokus wird auf die Denk- und Verarbeitungsvorgänge gelegt, die maßgeblich für das Lernen sind</p> <p data-bbox="616 1305 951 1541">Contra: fehlende Berücksichtigung von sozialen Zusammenhängen, Motivation und Emotionen</p>	<p data-bbox="1015 566 1270 645">Hirn ist selbstreferentielles, zirkuläres System</p>  <p data-bbox="991 913 1326 947">selbstgesteuertes Lernen</p> <p data-bbox="983 1061 1342 1227">Pro: Beachtung der sozialen Umwelt sowie motivationaler und emotionaler Aspekte</p> <p data-bbox="983 1305 1302 1473">Contra: exploratives Lernen kann den Lernenden überfordern</p>

Abbildung VIII: Darstellung und Beurteilung von Lerntheorien, Grafik übernommen aus (Stangl, 2018), angepasst und ergänzt mit Informationen aus (Kerres, 2012 S. 112 ff.), (Meir, 2006 S. 11-16) & (Nieding, et al., 2015 S. 37-49)

Auf der Grundlage dieser Lerntheorien können Strategien zur Wissensvermittlung in Bezug auf multimedialem Lernen erarbeitet werden. Darüber hinaus kann das Lernen nach den Merkmalen „Zielorientierung“ und „Strukturiertheit“ unterteilt werden (vgl. *Tabelle 4*).

		Strukturiertheit	
		Nein	Ja
Zielorientierung	Nein	zufälliges Lernen	Rituale
	Ja	exploratives Lernen	unterweisendes Lernen

Tabelle 4: Unterscheidung der Lernarten, übernommen aus (Minass, 2002 S. 20)

Zufälliges Lernen geschieht ohne die Absicht zu lernen (implizit). Üblicherweise werden die meisten Fertigkeiten (z.B. das Erlernen der Muttersprache als Kleinkind) implizit angeeignet (Nieding, et al., 2015 S. 45). Im Vergleich dazu geben ritualisierte Handlungen (z.B. Tischgebete) ein geordnetes Sozialverhalten vor und dienen nicht zum Wissenserwerb, sondern zur Stärkung des Zusammengehörigkeitsgefühls. Durch exploratives Lernen wird zwar eine Lernabsicht verfolgt, jedoch findet der Wissenserwerb eher durch experimentelle Handlungen statt. Dahingegen ist der herkömmliche Unterricht und die autodidaktische Wissensa-neignung ein unterweisendes Lernen, da Lernziele formuliert werden und nach einem gewissen Plan gelernt wird (Minass, 2002 S. 20-21).

3.1.2 Medien, E-Learning & Lernprogramme

Wie die folgende Übersicht (*Abbildung IX*) zeigt, kann die Gesamtheit des Lernens weiter unterteilt werden. Einer dieser Teilbereiche ist das sog. *technologiebasierte Lernen*. Dieses beschreibt jedwedes Lernen, welches von Technologien in Form von analogen und digitalen Medien unterstützt wird (Minass, 2002 S. 31). Laut Meyer sind Medien „[...] Mittel zur Weitergabe und Verbreitung von Informationen durch Sprache, Gestik, Mimik, Schrift und Bild [...]“ (Bibliographisches Institut Mannheim, 1975). Dazu zählen bspw. nicht nur Printmedien, das Internet und Telefon sowie Computer, sondern auch der Mensch selber (z.B. durch Aufführung eines Theaterstücks).

Innerhalb des technologieunterstützten Lernens wird das E-Learning eingeordnet. Das E-Learning steht stellvertretend für alle Maßnahmen, die zur Vermittlung von Lerninhalten an Einzelpersonen bzw. Personengruppen unter Verwendung von digitalen Medien dienen (Kerres, 2012 S. 18) & (Minass, 2002 S. 23-27).

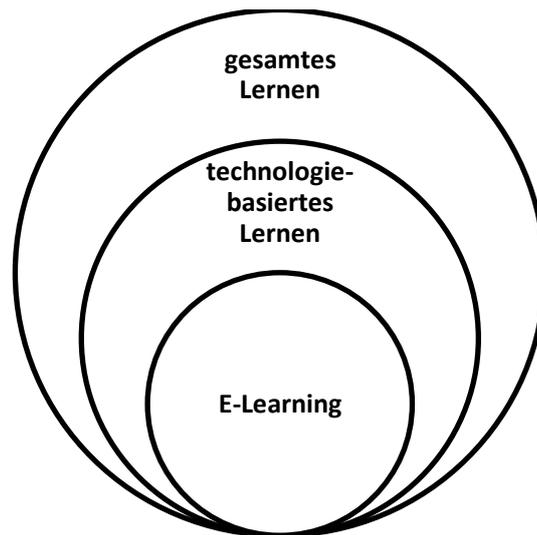


Abbildung IX: E-Learning im bildhaften Verhältnis zum gesamten Lernen, übernommen aus (Minass, 2002 S. 32)

Hierbei spielt der Umfang der Selbststeuerung des Lernenden eine tragende Rolle. Eine Person kann eher autodidaktisch, in einer sozialen Gruppe (z.B. über ein Internetforum) oder mit Unterstützung von Betreuern (z.B. mittels Onlinekursen) lernen (Kerres, 2012 S. 20-23). Lernprogramme sind eine spezielle Ausprägungsform des selbstgesteuerten Lernens mit digitalen Medien. Die nachfolgende Tabelle (Tabelle 5) zeigt, inwiefern bestimmte Aspekte der Selbststeuerung bei Lernprogrammen zum Ausdruck kommen.

	Lernprogramme
Inwieweit legt der Lernende die Lernziele und -inhalte selber fest?	Zwar sind die Ziele bereits im Voraus vorgegeben, jedoch können die Inhalte frei ausgewählt werden.
Bestimmt dieser selber die Lernmethoden?	Die Inhalte sind schon bereits in einer bestimmten Art und Weise aufbereitet.
Wählt dieser die Medien zum Lernen aus?	Indirekt, das Gerät, auf dem das Programm läuft, kann selber bestimmt werden (z.B. Tablet oder Smartphone).
Inwieweit legt der Lernende das Lerntempo und die Aufgaben zur Bearbeitung fest?	Das Tempo wird selber bestimmt und Übungsaufgaben werden zur Verfügung gestellt.
Werden der Ort und die Zeit des Lernens selber bestimmt?	Der Lernende kann diesbezüglich frei entscheiden.
Entscheidet der Lernende über die Sozialform (Gruppen-, Partner- und Einzelarbeit) des Lernens?	Das Programm richtet sich an Einzelpersonen, kann aber auch in der Gruppe benutzt werden.
Kontrolliert dieser seinen Lernfortschritt selber?	Testaufgaben helfen bei der Fortschrittsüberprüfung.

Tabelle 5: Grad der Selbststeuerung bei Lernprogrammen, basierend auf Informationen aus (Kerres, 2012 S. 6-7 & 23-24)

Lernprogramme können bei fehlender unterstützender Begleitung des Lernenden von Nachteil sein, wenn dieser Schwierigkeiten bekommt oder wenig motiviert bzw. diszipliniert ist. Darüber hinaus verhindert der Fokus auf das autodidaktische Lernen den unmittelbaren Austausch in einer sozialen Gruppe.

Trotzdem sticht diese Form des Lernens dadurch hervor, dass die Programme jederzeit und einfach zugreifbar sind. Außerdem benötigen derartige Systeme keine Verbindung zu anderen Geräten. Des Weiteren fördert die nicht erforderliche menschliche Betreuung die Selbstständigkeit von Personen, welche im Zuge des lebenslangen Lernens erforderlich ist (Kerres, 2012 S. 6-7 & 29-30).

3.2 Lehren mittels digitaler Medien

3.2.1 Bewertung des Einsatzes digitaler Medien

In Anknüpfung an die im vorherigen Abschnitt genannten Vor- und Nachteile von Lernprogrammen werden diese im Folgenden für das gesamte E-Learning verallgemeinert und die Gründe für das digitalmediengestützte Lernen beleuchtet. Die folgenden Ausführungen sind eine Zusammenfassung des vierten Kapitels aus (Kerres, 2012 S. 69-107) und mittels *Tabelle 6* dargestellt.

Pro	Contra
Die Einführung neuartiger digitaler Medien kann die Lernmotivation erhöhen.	Dieser Effekt ist jedoch kurzweilig. Besonders simple und unterhaltende E-Learning-Angebote können den Lernerfolg sogar schmälern.
Der Medieneinsatz <u>kann</u> im Vergleich zu anderen Lernangeboten das Verhältnis (Effizienz) zwischen monetären und zeitlichen Aufwendungen und dem monetären, quantitativen und qualitativen Nutzen der Lehre verbessern.	Die Beschaffung und der Einsatz dieser Medien bringt üblicherweise hohe Kosten mit sich.
Aufgrund der individuellen Anpassung an die Bedürfnisse der Lernenden ist die Aneignung von Lerninhalten durchschnittlich schneller. Studien zufolge kann die Lerndauer im Vergleich zum herkömmlichen Unterricht um bis zu 30 % verkürzt werden, was wiederum den Lernaufwand senkt.	Im Durchschnitt gibt es jedoch eine höhere Abbruchquote bei Lernenden (speziell bei denjenigen, die autodidaktisch lernen, aber kaum lerninteressiert und wenig selbstständig sind).
Die Auswertung mehrerer Einzelstudien belegt, dass	

<p>klassischer Unterricht in Kombination mit E-Learning, verglichen mit konventionellen Lernverfahren, zu einer leichten Steigerung des Lernerfolgs führt (insbesondere bei Kindern). Hierbei ist zu beachten, dass das bloße Einführen von digitalen Medien noch nicht viel bewirkt und der Lernerfolg überwiegend unabhängig von der Wahl des Mediums ist.</p>	
<p>E-Learning ermöglicht ein zeit- und standortunabhängiges Lernen außerhalb von Bildungseinrichtungen. Dies führt zu einer flexibleren Lernorganisation. Damit werden auch Personengruppen berücksichtigt, die konventionelle Lernangebote nicht wahrnehmen können (z.B. aufgrund fehlender Zeit, schwerer Erreichbarkeit der Einrichtungen, usw.).</p>	
<p>Mithilfe von E-Learning-Angeboten können Lerninhalte mithilfe von Audio- und Video-Elementen leicht veranschaulicht werden. Außerdem kann damit selbstgesteuertes bzw. kooperatives Lernen unterstützt werden.</p>	

Tabelle 6: Vor- und Nachteile des Einsatzes von E-Learning-Angeboten nach (Kerres, 2012 S. 69-107)

Letztendlich entscheidet die Lernbereitschaft des Lernenden und seine Akzeptanz des Mediums sowie die didaktische Aufbereitung der Lerninhalte durch Lehrende darüber, inwieweit digitale Medien dauerhaft zum Lernerfolg beitragen können (Kerres, 2012 S. 72 & 86).

3.2.2 Aufgabenbereich und Methoden der Mediendidaktik

Aus den oben genannten Ausführungen geht hervor, dass die Potentiale des E-Learning erst durch eine qualitativ hohe didaktische Konzeption erreicht werden. Die (gestaltungsorientierte) Mediendidaktik beschäftigt sich mit der Frage, wie mediengestütztes Lernen geplant und umgesetzt werden soll, um diese Vorteile wirksam ausnutzen zu können. Ziel der Mediendidaktik ist es nicht, Medienkompetenz aufzubauen, sondern diese vorauszusetzen, um Fachkompetenz mithilfe von Medien vermitteln zu können. Wie diese Vermittlung effizient gestaltet werden kann, ist Gegenstandsbereich der Mediendidaktik (Kerres, 2012 S. 36, 49 & 65-66).

Zum Beispiel muss man sich bei der Entwicklung eines Lernprogramms für Big Data damit auseinandersetzen, wie Big Data-Themen vermittelt werden können, welche Lernziele erreicht werden sollen und wie die Zielgruppe mitsamt Merkmalen aufgebaut ist. Hinzu kommen die üblichen Projektmanagement-Aspekte wie die Abwägung von Kosten und Nutzen, eine Machbarkeitsstudie und der Zweck des Medieneinsatzes. Des Weiteren spielen

die soziale, räumliche und zeitliche Lernorganisation sowie die Auswahl der didaktischen Methode eine zentrale Rolle (Kerres, 2012 S. 42 & 67-68).

Nach (Kerres, 2012 S. 47 & 297) sind didaktische Methoden Beschreibungen, „wie aus Lerninhalten Lernangebote werden“ und „wie Lernangebote aufbereitet werden können, um bestimmte Lernziele zu erreichen.“ *Tabelle 7* gibt einen schnellen Überblick über die hier behandelten Methoden.

Expositorische Methoden	einfache Lernziele (Sachwissen)
Prinzip: sequenzielle Präsentation im Vordergrund	
Chance: systematische Vermittlung von Fachwissen	
Herausforderung: Aktivierung von Lernprozessen durch Beispiele und Übungen	
Exploratives Lernen	
Prinzip: Selbststeuerung im Vordergrund	
Chance: Aktivierung von Interesse und des <i>Neugiermotivs</i>	komplexe Lernziele (Handlungswissen)
Herausforderung: Aufbereitung des Lernangebotes und Kontrolle des Lernfortschritts	
Problemorientierte Methoden	
Prinzip: Lernen mit Fällen, Projekten, Simulationen und Spielwelten	
Chance: Entwicklung von Kompetenzen im Umgang mit komplexen Problemen	
Herausforderung: Verzahnung mit abstrakter Wissensbasis	
Kooperative Methoden	
Prinzip: Lernen in Interaktion mit anderen	
Chance: Entwicklung von sozialen Kompetenzen	
Herausforderung: Sicherung des Lernziels für alle Beteiligten	

Tabelle 7: Übersicht der didaktischen Methoden und ihrer Eignung zur Vermittlung bestimmten Wissens, angepasst und übernommen aus (Kerres, 2012 S. 301-302)

Die meisten Lernangebote setzen expositorische Methoden um. Diese zeichnen sich dadurch aus, dass der Fokus auf die strukturierte und lineare Präsentation von abstrakten und konkreten Lerninhalten mittels Medien (z.B. Videos) gerichtet wird. Im Anschluss an diese Präsentation sollen Lernende das vermittelte Wissen anwenden, so dass diese selber aktiv werden. Hier sollte auf ein Gleichgewicht zwischen Präsentations- und Übungsphasen geachtet werden. Schließlich wird das Wissen mithilfe von Tests überprüft, um den Lernfortschritt zu überblicken und eine Rückmeldung zu erhalten (Kerres, 2012 S. 305 & 317-318).

Im Gegensatz zum fremdgesteuerten schrittweisen Lernweg bei expositorischen Methoden sind beim explorativen Lernen die Lernenden selbstgesteuerter und freier bezüglich der zeitlichen Abarbeitung von Lernaktivitäten. Dadurch soll die Lernmotivation erhöht und Lerninhalte sollen weniger voneinander abhängig werden. Hierbei ist zu beachten, dass der Lehrstoff in entsprechende Einheiten aufzubereiten ist und der Lernfortschritt schwieriger erfasst und überwacht werden kann (Kerres, 2012 S. 318-319 & 323).

Je nachdem wie der Lehrstoff gegliedert ist, wie die Zielgruppe aufgebaut, welches Vorwissen bereits vorliegt usw., muss entschieden werden, welche der beiden genannten didaktischen Methoden besser geeignet ist (Kerres, 2012 S. 335-336). Dies gilt auch für die nachfolgend vorgestellten Methoden (Kerres, 2012 S. 382-384).

Über expositorischen Methoden und dem explorativen Lernen kann insbesondere Sachwissen (z.B. Fakten) gut gelernt werden. Zur Vermittlung von Handlungswissen, sind problemorientierte und kooperative Methoden jedoch besser geeignet (Kerres, 2012 S. 339).

Mithilfe von problemorientierten Methoden sollen Lernende Wissen anhand von komplexen Anwendungssituationen (z.B. Projekte oder Fälle), bestehend aus einem Auftrag mit Meilensteinen, erlernen und auf andere Kontexte leichter übertragen können. Dadurch ist es nicht nur möglich fachliche Kompetenzen zu vermitteln, sondern auch die Selbst- und Zeitorganisation von Lernenden zu fördern. Problematisch hierbei könnte sein, dass die Komplexität der Aufgabe und das Hineinversetzen in diese Anwendungssituationen wegen mangelnder Erfahrung den Lernenden zur frühzeitigen Aufgabe zwingen kann. Lerninhalte müssen demnach so aufbereitet werden, dass sie an das Vorwissen der Lernenden anknüpfen und gelernte Erkenntnisse sich tatsächlich auf andere Situationen übertragen lassen (Kerres, 2012 S. 341-346, 355 & 358-359).

Zwei besondere didaktische Methoden im Zusammenhang mit Problemorientierung sind Computersimulationen und -spiele. Simulationen modellieren Anwendungssituationen und schaffen eine (visualisierte) immersive Lernumgebung (z.B. ein Flugsimulator), die dem Lernenden ermöglichen soll, u.a. zu experimentieren und die Auswirkungen bestimmter Aktionen zu erleben. Anhand dieser Simulationen können Lernende kontrolliert mit bestimmten Situationen und Problemen konfrontiert werden und dadurch erste praxistaugliche Erfahrungen sammeln. Man muss dennoch darauf achten, dass die Entwicklung einer Simulationsumgebung aufwändig ist, Lernende ohne unterstützende Anleitung schnell überfordert werden und beim Lernen falsche Schlussfolgerungen gezogen werden können (Nieding, et al., 2015 S. 163-168), (Minass, 2002 S. 84-86) & (Kerres, 2012 S. 359-366).

In Computerspielen kommt es zur immersiven Interaktion zwischen Lernenden (die Spieler) und der Spielumgebung. Die Spieler erlernen Sachwissen (die Spielregeln und -steuerung) und transferieren dieses auf Handlungswissen (zielgerichtete Handlungsweisen, um im Spiel voran zu kommen) und erhalten schließlich Rückmeldungen vom Spiel. Da Spiele von den Spielern nicht als Lernsituationen wahrgenommen werden, ist der Spaß und die Motivation des Spielens hoch. Beim Spielen hat das Lernen eine untergeordnete Rolle. Die Spieler lernen implizit und deswegen muss sichergestellt werden, dass festgelegte Lernziele tatsächlich erreicht werden. Außerdem muss darauf geachtet werden, dass die Umsetzung von explizitem Lernen in Computerspielen den Spielspaß beeinträchtigen kann, so dass Spieler dazu neigen, diese Lernsituationen zu umgehen, um weiterspielen zu können. Dadurch wird das explizite Lernen verhindert (Nieding, et al., 2015 S. 176-185), (Minass, 2002 S. 84-86) & (Kerres, 2012 S. 366-372).

Im Rahmen von spielbasierten Ansätzen fallen auch die Begriffe „Serious Games“ oder „Gamification“. Unter einem Serious Game versteht man Software, die nicht nur zur Unterhaltungszwecken, sondern auch zur Vermittlung von Lerninhalten dient, indem sie durchgespielt wird. Dahingegen wird Gamification als der Einsatz von Computerspielen-Bestandteilen (z.B. eine Fortschrittsanzeige in Form von Leveln) im spieluntypischen Kontext bezeichnet (Deterding, et al., 2011 S. 10). Ob eine Anwendung ein Spiel ist oder lediglich spieltypische Elemente aufweist, ist vorrangig von den Absichten der Entwickler und den Erfahrungen der Benutzer abhängig (Deterding, et al., 2011 S. 14).

Des Weiteren können problemorientierte Methoden um eine soziale Komponente erweitert werden, so dass Probleme kooperativ gelöst werden können. Dadurch können Kompetenzen wie Teamfähigkeit, Arbeitsteilung und Hilfs- sowie Kompromissbereitschaft begünstigt werden. Es lernt nicht jeder für sich selbst, sondern durch den Austausch in der Gruppe setzen sich Lernende mit den Ansichten anderer auseinander und profitieren durch diesen Wissensaustausch. Hier ist die Herausforderung für Lehrende sicherzustellen, dass alle Teilnehmer einer Gruppe zusammenarbeiten, gleich viel zur Problemlösung beitragen und die gleichen Lernziele erreicht haben (Kerres, 2012 S. 372-373 & 375-377).

4 Anforderungsanalyse

Dieses Kapitel dient vor allem dazu die Anforderungen systematisch zu erarbeiten und darzustellen. Dabei orientieren sich die nachfolgenden Abschnitte an dem Leitfaden für die mediendidaktische Entwicklung von Lernangeboten (Kerres, 2012 S. 493-499) und teilweise am Vorgehensmodell aus (Baumgartner, 1997). Es werden Bezüge zu den vorherigen Grundlagenkapiteln hergestellt und getroffene Entscheidungen auf dieser Grundlage begründet.

4.1 Bewertung und Auswahl der Werkzeuge

Um sämtliche und insbesondere spezifische Anforderungen zu ermitteln, wird von vornherein festgelegt, mit welchen Werkzeugen das Programm realisiert werden soll. Für die Umsetzung eines Big Data-Lernprogramms stehen zahlreiche Möglichkeiten zur Verfügung. Generell sollte darauf geachtet werden, dass das Programm auch auf leistungsschwachen und unterschiedlichen Systemen funktioniert, um möglichst viele potentielle Anwender aus der Zielgruppe zu erreichen (Löhr, 2012 S. 26). Außerdem ist zu beachten, dass die auszuwählenden digitalen Medien autodidaktisches Lernen und die anzuwendenden didaktischen Methoden unterstützen müssen.

Lernprogramme können beispielsweise als Anwendungen direkt auf dem Computer oder über einen Internetbrowser aufgerufen werden. Entwickler können die Lernsoftware mithilfe von sogenannten „Autorensystemen“ implementieren, von denen es unzählige gibt. Autorensysteme sind Software-Werkzeuge, um E-Learning-Angebote nach dem Baukastenprinzip zu entwickeln. Sie sind für Personen geeignet, die digitale Lernangebote realisieren möchten, aber über keine ausreichenden Programmierkenntnisse verfügen. Solche Werkzeuge gibt es für generelle und spezielle Anwendungszwecke und bilden eine Abstraktionsschicht zur herkömmlichen Programmierung. Nicht alle Autorensysteme sind kostenlos, für einige sind Lizenzen erforderlich, die über 1000 US-Dollar kosten können. Des Weiteren ist die Einarbeitung in Autorensystemen nicht immer trivial und erfordert einen nennenswerten Aufwand (Löhr, 2012 S. 26-27). Schließlich ist es nicht ganz einfach aus der Vielzahl der Angebote die entsprechenden Autorensysteme zu finden, die sämtliche Anforderungen an das Lernprogramm erfüllen können.

Alternativ können digitale Lernangebote eigenständig mithilfe einer Programmiersprache realisiert werden. Eine geeignete Programmiersprache ist „Java“, welche *objektorientiert* ist und sowohl für die Entwicklung von Desktopanwendungen als auch von mobilen Applikationen verwendet werden kann. Darüber hinaus gibt es für Java zahlreiche Programmbibliotheken und eine umfangreiche Internet-Gemeinschaft, die das Programmieren erleichtern. Bei dieser Art der Umsetzung hat der Entwickler durch die Einhaltung von Entwurfsprinzipien Einfluss darauf, dass die Software wiederverwendbar und erweiterbar ist.

Dennoch kann man nicht pauschal behaupten, dass man die eine Umsetzungsmöglichkeit immer der anderen vorziehen muss. Dies hängt maßgeblich von der Komplexität der späteren Lernsoftware, der Softwarekenntnisse der Entwickler, dem geplanten Aufwand und dem zur Verfügung stehenden Budget ab.

Dieses Lernprogramm soll als mobile Applikation für Android-Endgeräte entwickelt werden. „Android“ ist ein Betriebssystem von Google, welches speziell für mobile Endgeräte mit Touchscreen entwickelt worden ist. Es hat einen Marktanteil von ca. 86 % und ist damit das populärste Betriebssystem seiner Art (Gartner, 2018). Deswegen ist eine Android-App gut dafür geeignet, möglichst viele potentielle Benutzer zu erreichen. Derartige Apps werden u.a. nativ in Java programmiert. Für die App-Entwickler wird die kostenlose und mächtige Entwicklungsumgebung „Android Studio“ mitsamt Emulator für Endgeräte zur Verfügung gestellt, die das Programmieren erleichtert.

Im Vergleich dazu kann eine betriebssystemunabhängige Internetseite entwickelt werden, die als mobile Web-App ohne Installation auf beliebigen Endgeräten ausgeführt werden kann. Da ich jedoch im Gegensatz zur Android-Programmierung nicht ausreichend Erfahrung mit *HTML5* und *JavaScript* besitze, würde sich der dadurch entstehende Mehraufwand zur Berücksichtigung von lediglich 14 % der Anwender nicht lohnen. Zwar besteht die Möglichkeit sowohl native und als auch Web-Apps mithilfe von sogenannten App-Baukästen (ähnlich wie Autorensysteme) zu realisieren, aber diese sind üblicherweise kostenpflichtig und schränken die Freiheit des Entwicklers durch die Vorgabe von *Templates* ein.

4.2 Zielgruppe

Da im Rahmen dieser Arbeit bis auf den Entwickler keine weiteren *Stakeholder* aktiv sind, konzentriert sich dieser Abschnitt auf die Zusammensetzung der Zielgruppe des Lernprogramms.

Merkmal der Lernenden	Beschreibung
Anzahl der Teilnehmer	unbegrenzt
Geographische Verteilung	deutschsprachiger Raum
Alter	alle Personen ab 13 Jahren (die Applikation richtet sich v.a. an Heranwachsende und Erwachsene mittleren Alters)
Geschlecht	alle
Vorwissen	niedriges bzw. kein Vorwissen über Big Data (deshalb ist die App nicht für Informatiker oder informatik-affine Personen ausgerichtet)
Lernmotivation	sowohl ex- als auch intrinsisch
Lerndauer	maximal 20 Minuten pro thematische Einheit
Medienkompetenz	Teilnehmer sollen mit mobilen Endgeräten umgehen können
Lernorte	Zuhause, in der Schule, unterwegs, am Arbeitsplatz, usw.

Technische Ausstattung	Ein mobiles Android-Endgerät (ab Version „Jelly Bean“) ist erforderlich.
------------------------	--

Tabelle 8: Merkmale der Zielgruppe in Anlehnung an (Kerres, 2012 S. 258-264)

Wie *Tabelle 8* zeigt, kann die Zielgruppe aufgrund ihrer Größe und Vielseitigkeit nur grob charakterisiert werden, was eine genauere Segmentierung nicht ermöglicht. Deswegen sollen die obigen Ausführungen sowohl als Analyse der Zielgruppe verstanden werden als auch als Beschreibung davon, für wen das Lernprogramm geeignet ist. Letztendlich richtet es sich an alle Betroffenen und potentiellen Anwender von Big Data, die sich gar nicht oder nur kaum mit diesem Thema auskennen, sich aber dafür interessieren. Dennoch soll diese Zielgruppe dahingehend eingeschränkt werden, dass der Fokus auf junge Menschen und Personen mittleren Alters gerichtet wird, die ein aktuelles Android-Gerät besitzen. Des Weiteren wird auf die Barrierefreiheit der App (z.B. Eignung der App für Menschen mit Rot-Grün-Sehschwäche) der Einfachheit halber keine besondere Rücksicht genommen. Aufgrund der heterogenen Zielgruppe müssen die Inhalte des Lernprogramms möglichst unabhängig von der Altersgruppe, dem Geschlechts, dem Bildungsgrad und der Kultur gestaltet sein.

4.3 Marktanalyse

Bevor es zur Medienkonzeption kommt, sollte zunächst geprüft werden, ob bereits ähnliche Lernangebote existieren. Für die Schlagwörter „Big Data“ und „E-Learning“ liefern die gängigen Internet-Suchmaschinen überwiegend Ergebnisse darüber, welchen Einfluss Big Data auf die E-Learning-Branche haben könnte.

Dennoch gibt es Internetseiten, die Fernunterricht in Form von Onlinekursen (z.B. „edx.org“ oder „coursera.org“) anbieten. Darunter auch zahlreiche Angebote im Bereich Big Data und *Data Science*. Diese Kurse sind kostenlos (ausgestellte Zertifikate sind gebührenpflichtig) und spannen sich über einen Zeitraum von mehreren Wochen. Der angegebene Aufwand für die Lernenden beträgt einige Stunden pro Woche. Im Gegensatz zum Vorhaben dieser Bachelorarbeit sind die Kurse mehrheitlich technisch orientiert, lerninhaltlich tiefgehend und streben eine praxisbezogene Ausbildung von Lernenden an (z.B. durch die Verwendung bestimmter Software). Diese Produkte stehen deshalb nicht in Konkurrenz zu diesem Vorhaben und werden daher im Folgenden nicht weiter berücksichtigt.

Gerda Sieben berichtet in (Gapski, 2015 S. 131-139) über eine Arbeitshilfe für die Jugendarbeit namens „Big Data“.¹ Diese soll Medienpädagogen dabei unterstützen, Jugendliche darüber zu informieren, was Big Data eigentlich ist, welche Auswirkungen Big Data hat, wie man auf die Herausforderungen von Big Data reagieren kann, welche Spiele, Filme und andere Methoden es zu diesem Thema gibt und wo weitere Informationen zur Verfügung gestellt werden. Sie stellt demnach kein Konkurrenzprodukt dar, sondern aggregiert für

¹ diese Arbeitshilfe ist als Broschüre verfügbar: (jfc Medienzentrum e.V, 2015)

Jugendliche aufbereitetes Wissen, Ratschläge für Pädagogen und Lernende sowie Angebote zur Wissensvermittlung. Die Inhalte dieser Arbeitshilfe können daher für die spätere Konzeption und Entwicklung eines Big Data-Lernprogramms genutzt werden. Des Weiteren wird darin auf Spiele hingewiesen, die sich mit dem Thema „Big Data“ befassen. Eines davon ist ein Planspiel² und das andere ein Internetbrowserspiel, welches nur als Testversion zur Verfügung steht und scheinbar nicht mehr weiterentwickelt wird.³

Trotz alledem wird auf den gängigen Plattformen für mobile Applikationen nach derzeitigem Stand keine E-Learning-Anwendung angeboten, welche grundlegendes Sachwissen bezüglich Big Data und dem Datenschutz nicht nur an Jugendliche, sondern auch an weitere Bevölkerungsgruppen vermittelt (letzte Recherche: April 2018).

4.4 Anforderungen

4.4.1 Lerninhalte & Lernergebnisse

Auf der Basis der im vorherigen Abschnitt beschriebenen Merkmale lassen sich die Lerninhalte ausgestalten. Diese kann man in folgende Themenblöcke aufteilen:

- (1) Big Data - Definition und Abgrenzung zu Cloud-Computing, IoT und BI
- (2) Chancen und Risiken von Big Data-Analysen am Beispiel von Anwendungsszenarien
- (3) Herausforderungen bezüglich der Anwendung von Big Data
- (4) Datenschutz in Deutschland (Abgrenzung zur Datensicherheit, Rechte und Pflichten von Betroffenen und Verantwortlichen und Zusammenhang zu Big Data)

Diese Inhalte gilt es so zu vermitteln, dass die Lernenden die in *Tabelle 9* aufgeführten Lernergebnisse größtenteils erfolgreich erzielen. Für die didaktische Aufbereitung dieser Inhalte ist es erforderlich, anzugeben, wie sich das Lernen auf die Teilnehmer auswirken soll, d.h. welche Resultate erzielt werden sollen. Diese Lernergebnisse können anhand von Lehrzielen (einer Beschreibung von Zielen aus der Perspektive von Lehrenden) spezifiziert werden (Kerres, 2012 S. 278-279 & 294):

	Lernergebnis	funktionale Anforderung
	Das Lernprogramm wird von den Lernenden akzeptiert, genutzt und bewertet.	
Block (1)	Die Teilnehmer (Tn) sind in der Lage, die gängigste Definition von Big Data zu nennen und zu bewerten.	Dem Anwender wird die bekannteste Big Data-Definition (das V-Modell) angezeigt und erläutert.
	Tn können in eigenen Worten Big Data von Cloud-	Dem Anwender werden die

² http://www.jfc.info/data/Big_Data_Planspiel_V3.pdf (letzter Zugriff: 26.04.2018)

³ <http://datadealer.com/de> (letzter Zugriff: 26.04.2018)

	Computing, IoT und Business Intelligence abgrenzen.	Unterschiede zwischen Big Data und Cloud-Computing, IoT und BI gezeigt.
Block (2)	Tn sind vertraut mit einigen Vor- und Nachteilen der Big Data-Nutzung und können diese wiedergeben.	Dem Anwender werden Pro und Contra von Big Data anhand von Beispielen gezeigt.
Block (4)	Tn können die Unterschiede und Gemeinsamkeiten zwischen Datenschutz und Datensicherheit aufzeigen.	Der Anwender bekommt die Bedeutung der Begriffe „Datenschutz“ und „Datensicherheit“ erklärt.
	Tn sollen verstehen, wann der Datenschutz greift (personenbezogene Daten).	Der Anwender kann mithilfe von Beispielübungen sein Wissen über den Datenschutz anwenden.
	Tn können einige die in der DS-GVO genannten Rechte und Pflichten von Betroffenen und Verantwortlichen wiedergeben.	Der Anwender hat Zugriff auf die Betroffenenrechte und Verpflichtungen von Verantwortlichen gemäß der DS-GVO.
	Tn erkennen die Probleme des Datenschutzrechts bezüglich Big Data.	Der Anwender bekommt einen Überblick über die rechtliche Einordnung von Big Data in Deutschland.
Block (3)	Tn wissen grob, welche Prozessschritte in Big Data-Anwendungen durchlaufen werden.	Der Anwender kann auf eine Erläuterung des Big Data-Lebenszyklus mitsamt Herausforderungen zugreifen.
	Tn kennen einige Herausforderungen für die Prozessschritte des Lebenszyklus-Modells.	
	Tn sollen gelerntes Wissen unter Durchführung eines Tests wiedergeben können.	siehe <i>Tabelle 11</i>

Tabelle 9: Spezifikation der Lernergebnisse, ggf. Kategorisierung in Themenblöcke und daraus abgeleitete Anforderungen gemäß (Kerres, 2012 S. 280-285)

Daraus kann man die funktionalen Anforderungen an das Lernprogramm ableiten. Diese sind in *Tabelle 9* aufgelistet. Wie man sieht, sind sie fast bedeutungsgleich wie das entsprechende Lernergebnis, jedoch so umformuliert, dass sie beschreiben, was das Lernprogramm leisten soll. Diese Auflistung ist keinesfalls erschöpfend, da Anforderungen sich im zeitlichen Verlauf ändern oder neu hinzukommen können.

Schließlich muss geklärt werden, was in diesem Kontext unter „Lernerfolg“ zu verstehen ist. Nach Kerres ist der Lernerfolg nicht ausschließlich von der Erreichung der Lernergebnisse, sondern auch von anderen Faktoren abhängig. Dazu zählt das Kosten-Nutzen-Verhältnis,

die Nachfrage nach dem Lernangebot, die Akzeptanz und der Nutzungsumfang des Lernangebots sowie die Abbruchquote der Lernenden (Kerres, 2012 S. 89-90 & 281-282).

4.4.2 Nichtfunktionale & didaktische Anforderungen

Die in den obigen Ausführungen begonnene Anforderungsanalyse soll im Folgenden um technische, ergonomische und didaktische Aspekte des Lernprogramms erweitert werden. Die beiden nachfolgenden Tabellen fassen diese Anforderungen zusammen und richten diese bezüglich der Umsetzung einer Android-Applikation folgendermaßen aus:

Kategorie	Anforderung
Performanz	schnelle Ladezeiten des Hauptmenüs und anderer <i>Activities</i> bzw. <i>Fragmente</i> (unter zwei Sekunden)
	Programmbibliotheken zur Anzeige von hochauflösenden Bildern wie z.B. (Glide, 2018) verwenden
grafische Benutzerschnittstelle	pro Lerneinheit einheitliche Gestaltung des Layouts zur schnelleren Eingewöhnung beim Anwender
	einheitliche Abstände zwischen Bildschirm-elementen
	fachlich ähnliche Bildschirmelemente räumlich und zeitlich zusammenhängend gruppieren und präsentieren, um die Orientierung zu verbessern
	Bildschirmelemente entlang horizontaler und vertikaler Achsen ausrichten, um eine übersichtliche Struktur zu erzeugen
Typographie	Vermeidung von kursiven und komplett großgeschriebenen Texten sowie einheitliche Verwendung einer <i>serifenlosen</i> Schrift mit großem Schriftgrad (größer gleich 9 Punkt) zur besseren Lesbarkeit
	fett formatierter Text und Unterstreichungen sind zur Text hervorhebung geeignet
	linksbündigen Flattersatz oder, wenn umsetzbar, Blocksatz verwenden
Farbauswahl	einheitliche Farbverwendung
	auf stark <i>gesättigte Farben</i> verzichten und Vorder- und Hintergrund stark kontrastie-

	ren, um die Lesbarkeit zu erhöhen
	helle Farben zur Hervorhebung benutzen
Gestaltung von Texten, Grafiken, Audios und Videos	Bilder (bezüglich der Lesereihenfolge des Anwenders) vor dazugehörigen Texten platzieren
	zur besseren Übersicht Stichpunkte statt ganzer Sätze verwenden
	Bilder statt komplexe Animationen und Videos benutzen
Gebrauchstauglichkeit	Unterstützung von horizontaler (Landscape) und vertikaler (Portrait) Ausrichtung des Geräts
	die Bedeutung von Bildelementen soll mittels Icons und Kurzbeschreibungen verständlicher und intuitiver sein
Navigation	einheitliche Navigationshilfen im Randbereich des Bildschirms stets verfügbar machen
	aktuelle Position / aktuellen Stand im Lernprogramm anzeigen
Sonstiges	es kann eine Verbindung zum Internet hergestellt werden (z.B. um Updates einfacher durchführen zu können)
	das Lernprogramm soll auf allen Android-Geräten (ab Version „Jelly Bean“) funktionieren
	die App sollte nicht größer als 25 MB sein, damit man sie „schnell“ herunterladen und installieren kann

Tabelle 10: Nichtfunktionale Anforderungen, in Anlehnung an (Löhr, 2012 S. 19-21) und mit Informationen aus (Kerres, 2012 S. 158-160) und der Gestaltungsrichtlinien aus der Android-Dokumentation (Android, 2018) ergänzt

Wie im Abschnitt 3.2 *Lehren mittels digitaler Medien* beschrieben, ist die didaktische Aufbereitung der Lerninhalte entscheidend für den Lernerfolg. Deswegen sollte die Auswahl der Medien und anderer Werkzeuge der didaktischen Konzeption entsprechen.

Für ein Lernprogramm, das fundamentales Sachwissen an Personen mit wenig Vorwissen vermitteln und abfragen soll, eignet es sich ein expositorischer Ansatz. Eine ausschließliche Verwendung expositorischer Methoden birgt jedoch die Gefahr, den Lernenden nicht genügend Freiheiten und kaum einen Mehrwert gegenüber herkömmlichen Lernangeboten (z.B. Internetrecherchen) zu ermöglichen. Um das Abbruchrisiko zu senken, sollten Maßnahmen

getroffen werden, die zur Erhöhung der Lernmotivation und der Unterhaltung der Lernenden beitragen. Hierfür eignen sich explorative oder spielbasierte Ansätze.

Auf der Grundlage dieser Erkenntnisse können Anforderungen für die didaktische Konzeption festgehalten werden.

didaktische Anforderungen
Lernziele formulieren (vgl. <i>Tabelle 9</i>) und dem Anwender diese (schriftlich) mitteilen
komplexe Inhalte in Teile zerlegen, um den Anwender nicht zu stark zu belasten, da das Arbeitsgedächtnis eine begrenzte Kapazität hat
Informationen zu einem Thema sowohl in Bild- als auch in Textform (redundanzfrei) präsentieren (sog. „doppelte Kodierung“), um das Lernen zu unterstützen, indem es zu einer besseren Auslastung des Arbeitsgedächtnisses kommt
Aufmerksamkeit des Anwenders erregen und seine Lernmotivation erhöhen, indem z.B. ein konkreter Bezug der Lerninhalte auf die Alltagswelt des Anwenders hergestellt wird ⁴
Integration von Unterhaltungselementen (z.B. Gamification) zur Senkung des Abbruchrisikos
der Anwender kann das gelernte Wissen abfragen lassen und anwenden (z.B. mithilfe von Beispielen oder Frage-Antwort-Aufgaben) und erhält stets Rückmeldungen

Tabelle 11: Didaktische Anforderungen in Anlehnung an (Löhr, 2012 S. 25), angepasst und mit Informationen aus (Kerres, 2012 S. 158-160, 260-263 & 365-366) ergänzt

Die Ausführungen dieses Abschnitts schließen die Anforderungsanalyse ab und bilden die Vorbereitung für den fachlichen und technischen Entwurf des Lernprogramms.

⁴ Eine ausführliche Beschreibung von Gestaltungsmöglichkeiten bezüglich der Erhöhung der Lernmotivation ist in (Kerres, 2012 S. 262-263) nachlesbar.

5 Konzeption

5.1 Beschreibung der fachlichen Komponenten

Für den Entwurf des Lernprogramms ist es hilfreich, das System in (fachliche) Komponenten (auch Module genannt) aufzuteilen, um die zu implementierende Software wartbarer sowie erweiterbarer zu machen und Entwurfsprinzipien wie z.B. die Trennung von Zuständigkeiten umzusetzen. Im Folgenden wird der Aufbau und die Funktion dieser Komponenten im Einzelnen beschrieben.

Auf der Grundlage der in Abschnitt 4.4.1 vorgestellten Themenblöcke werden vier Hauptkomponenten abgeleitet, aus denen das System besteht:

Big Data: Begriffsklärung

Diese Komponente repräsentiert den Themenblock **(1)** und setzt sich aus drei Modulen zusammen. Das Modul „V-Modell“ soll in sequenzieller Art und Weise den Begriff Big Data erläutern. Hierfür ist geplant, die Inhalte in doppelter Kodierung (siehe *Tabelle 11*) zu präsentieren. Der gleiche expositorische Ansatz soll auch im zweiten Modul „Begriffsabgrenzung“ umgesetzt werden. Diese Subkomponente behandelt die Zusammenhänge und Unterschiede zwischen Big Data und Cloud-Computing, IoT sowie BI.

Schließlich sollen im dritten Modul die Inhalte der beiden anderen Module abgefragt werden. Dafür sind Multiple-Choice-Tests gut geeignet (Kerres, 2012 S. 425-426).

Datenschutzquiz

Diese Komponente bezieht sich auf den Themenblock **(4)** und dient zur Vermittlung von Gesetzesartikeln des Datenschutzrechts. Hierfür soll ein spielebasierter Ansatz in Form eines Quiz mit Jokern und Belohnungen umgesetzt werden. Die Inhalte werden nicht direkt vermittelt, sondern sollen durch die Durchführung des Quiz angeeignet werden. Dadurch soll der Anwender seine Vermutungen über den Datenschutz mit dem tatsächlich geltendem Recht abgleichen, um mögliche Missverständnisse auszuräumen. Außerdem sollen die eintönigen Gesetzesinhalte durch Gamification insoweit aufbereitet werden, um die Lernmotivation zu steigern.

Chancen und Risiken

Dieses Modul ist aus dem Themenblock **(2)** abgeleitet und besteht aus zwei Unterkomponenten. Die erste heißt „Anwendungsszenarien“ und soll mithilfe von Beispielszenarien die Möglichkeiten und Gefahren von Big Data-Anwendungen mit ein wenig Humor vermitteln. Im Sinne eines „Wahr/Falsch-Tests“ kann man den Anwender raten lassen, ob das Beispielszenario korrekt oder fiktiv ist. Damit könnte man diesen stärker miteinbeziehen und seine Aufmerksamkeit erregen (siehe *Tabelle 11*).

Ein zweites Modul „Persönlichkeitstest“ dient dazu, explorativ eine Anwendung zu benutzen, welche auf der Grundlage von digitalen Daten (z.B. aus sozialen Medien) einer Person ein psychologisches und demographisches Profil erstellt. Hiermit soll es für die Anwender möglich sein, Big Data-Analysen unmittelbar zu erfahren und ggf. mit bereits gemachten Erkenntnissen aus dem Themenblock **(4)** verknüpft.

Herausforderungen für Anwender

Der Themenblock **(3)** wird durch dieses Modul repräsentiert. Die Subkomponente „Lebenszyklus-Modell“ soll die Schwierigkeiten bei der Umsetzung eines Big Data-Projekts für jeden einzelnen Prozessschritt im Big Data-Lebenszyklus darstellen und ist besonders für anwendungs- bzw. technisch orientierte Lernende gedacht. Um den Lernerfolg sicherzustellen, soll dem Anwender in dieser Komponente zunächst empfohlen werden, die Lerneinheiten aus „Big Data: Begriffsklärung“ abzuschließen, bevor mit der Bearbeitung dieses Moduls begonnen wird. Die Ausrichtung des Moduls an ein bestimmtes Niveau an Vorwissen geht aus kognitiven Lerntheorien hervor (vgl. Abschnitt 3.1.1).

Schließlich soll auch hier das angeeignete Wissen mithilfe eines Multiple-Choice-Tests überprüft werden.

Alles in allem können die Hauptkomponenten des Systems als eigenständige Lerneinheiten betrachtet werden. Bis auf die bereits geschilderte Ausnahme spielt es keine Rolle, in welcher Reihenfolge die Module vom Anwender abgearbeitet werden. Außerdem muss nicht das gesamte Lernprogramm durchlaufen werden. Der Anwender kann daher diesbezüglich frei entscheiden, welche Module zum Lernen herangezogen werden.

Des Weiteren soll der Anwender durch die Beantwortung von Testfragen seinen eigenen Wissensstand beurteilen. Da die Tests hierbei eine sofortige Rückmeldung geben, wird ein behavioristischer Lernansatz umgesetzt, der diese Selbsteinschätzung unterstützen soll.

Darüber hinaus ist das System so aufgebaut, dass sowohl hierarchisch-sequentielle Lernpfade (z.B. im Modul „Big Data: Begriffsklärung“) als auch Möglichkeiten offenen Lernens (bspw. im Modul „Persönlichkeitstest“) angeboten werden.

5.2 Fachliche Architektur

Dieser Abschnitt befasst sich mit der Integration der oben beschriebenen Komponenten in die Systemarchitektur.

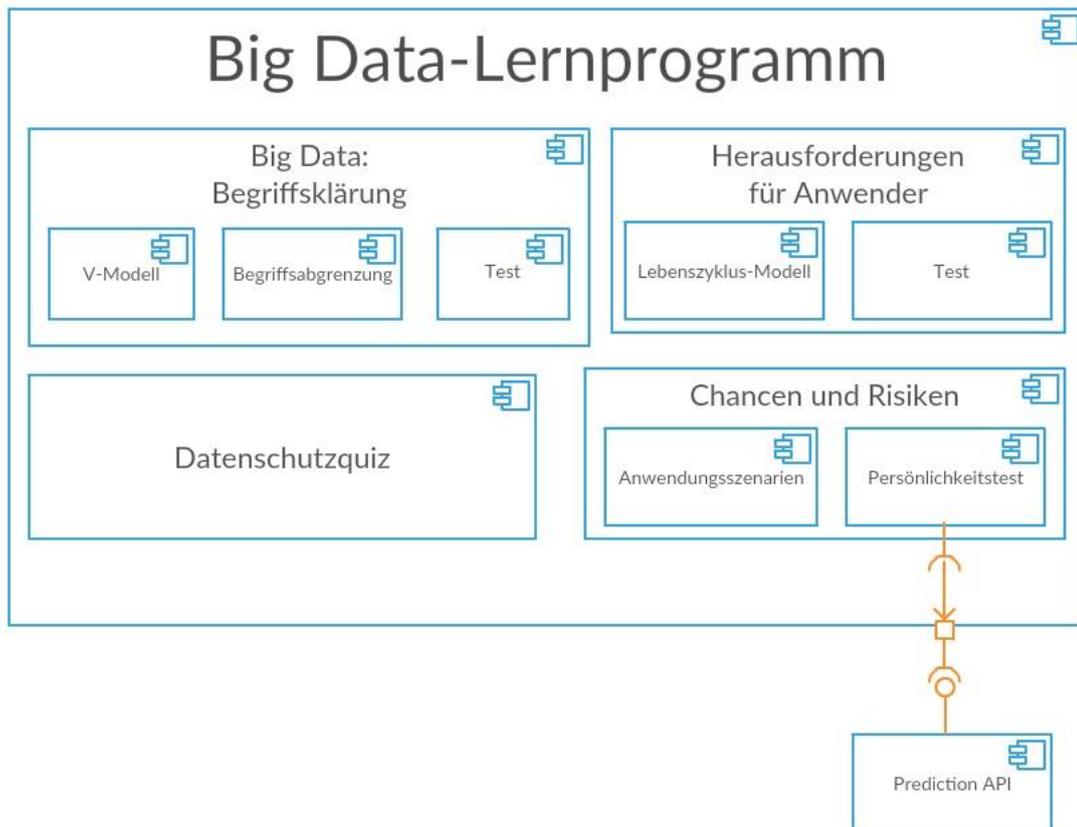


Abbildung X: fachliche Softwarearchitektur dargestellt als UML-Komponentendiagramm

Wie man in *Abbildung X* erkennen kann, gibt es zwischen den internen Komponenten keine Beziehungen bzw. Abhängigkeiten zueinander. Daher hat das System eine sehr geringe *Kopplung*, was das isolierte Testen und die Anpassbarkeit (z.B. die Erweiterung des Systems um weitere Module bzw. die Veränderung eines Moduls) erleichtert. Zwar sind die Module konzeptionell unabhängig voneinander, jedoch setzt das Modul „Herausforderungen für Anwender“ aus didaktischer Sicht die Inhalte von Themenblock **(1)** voraus (siehe Abschnitt 5.1), um erfolgreich bearbeitet zu werden. Diese lerninhaltlichen Verknüpfungen sind zu beachten, auch wenn sie in der Architektur nicht explizit dargestellt sind.

Darüber hinaus benötigt die Subkomponente „Persönlichkeitstest“ eine Netzwerkverbindung, um die externe Programmierschnittstelle „Prediction API“ (ApplyMagicSauce, 2018) aufrufen zu können.

5.3 Technische Konzeption

Um die Funktionsweise und Beschaffenheit des Lernprogramms besser zu veranschaulichen, sollen die folgenden Abbildungen diese in einer abstrahierten Form darstellen.

Einstiegspunkt des Lernprogramms ist die *Activity* „MainActivity“, welche während der Programmlaufzeit stets aktiv ist. Sie ermöglicht die jederzeitige Navigation durch das gesamte System und interagiert mit dem Anwender (z.B., wenn dieser von einer Lerneinheit zur nächsten wechselt). Die grafische Benutzerschnittstelle dieser Klasse ist so gestaltet, dass beim Programmstart zunächst nur die Navigationsleiste zu sehen ist. Der Hintergrund bleibt zunächst frei (siehe *Abbildung XI*).

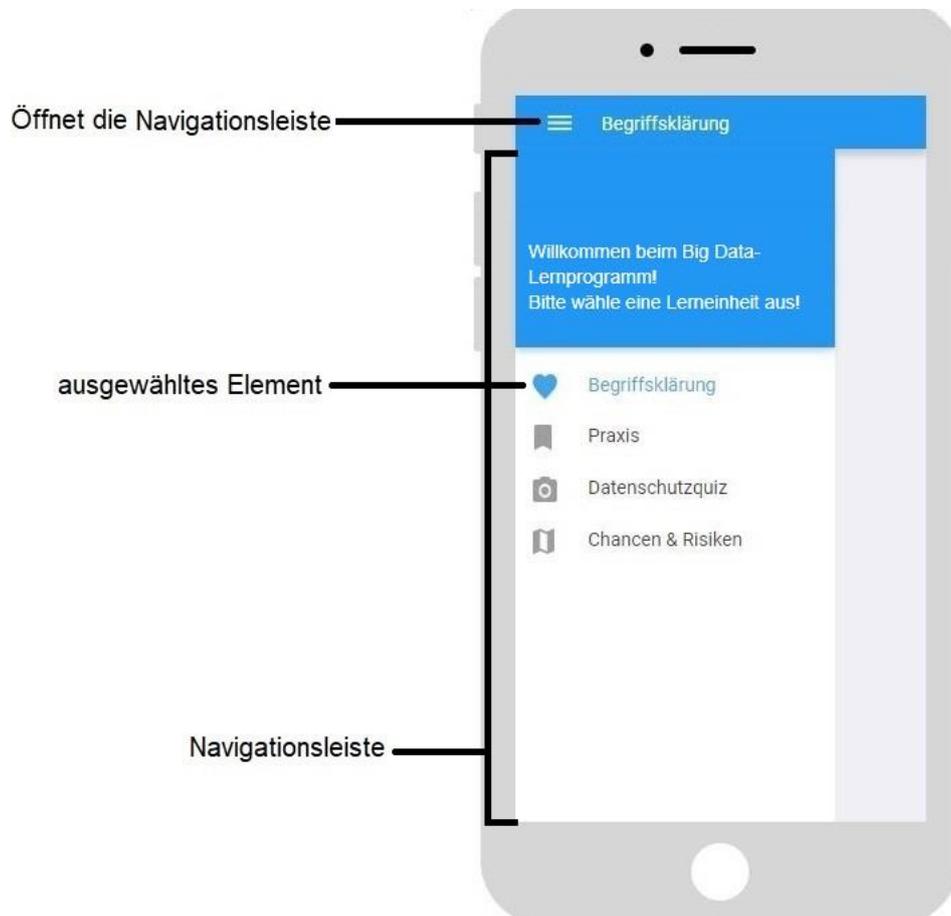


Abbildung XI: Mockup (Vorführmodell) des Startbildschirms der App. Die Symbole sind lediglich Platzhalter.

Sobald der Anwender eine Lerneinheit auswählt, werden die passenden Inhalte eines Moduls als Hintergrund hinzugefügt. Dies soll mittels sogenannter *Fragmente* (hier: Subklassen der Android-Klassen „ListFragment“ und „Fragment“) geschehen. Wenn der Anwender ein Navigationselement selektiert, wird das entsprechende Fragment aktiv. Jede Klasse „Fragment“ enthält die Anwendungslogik und UI eines Untermoduls. Außerdem haben die meisten Hauptmodule Menüs zur Navigation innerhalb der jeweiligen Lerneinheit (umgesetzt als „ListFragment“). Wenn innerhalb einer solchen „Menu-Klasse“ ein Menüelement ausgewählt wird, wird über die MainActivity das entsprechende Fragment dargestellt. Die oben

beschriebenen Ausführungen werden in *Abbildung XII* grafisch als UML-Diagramm dargestellt.

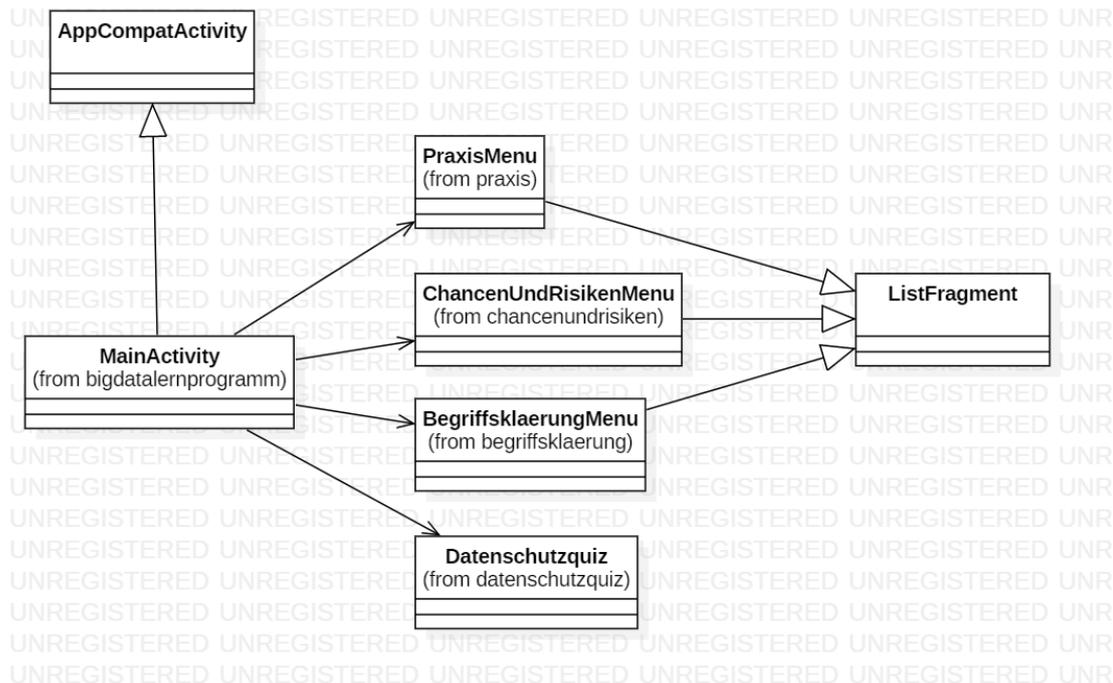


Abbildung XII: Ein UML-Klassendiagramm, welches die (Vererbungs-) Beziehungen zeigt. In Klammern steht die Paketzugehörigkeit (Teil 1 von 2). Anmerkung: der Bildhintergrund hat keine Bedeutung

Aus *Abbildung XIII* geht hervor, dass über die „Menu-Klasse“ eines Moduls sich die Unterkomponenten aufrufen lassen. Fast alle Fragmente lassen sich aufteilen in Klassen, die Lerninhalte vermitteln (hier „Info-Fragmente“ genannt) und welche, die Lerninhalte abfragen (hier „Quiz-Fragmente“ genannt). Info-Fragmente sind von der Gestaltung her gleich und lediglich dafür zuständig, verschiedene Texte und Bilder anzuzeigen.

Die Quiz-Fragmente haben ein Layout mit Auswahlbuttons und Textanzeigen. Außerdem besitzen sie jeweils eine einfache Java-Klasse, welche die Fragen, Antworten und weitere Fachlogik enthalten. „TextPrediction“ ist die einzige Fragment-Klasse, die über das Internet mit einer *API* kommuniziert und deswegen die Klasse „RequestQueue“ der *HTTP*-Programm-bibliothek „Google Volley“ (Volley, 2018) verwendet.

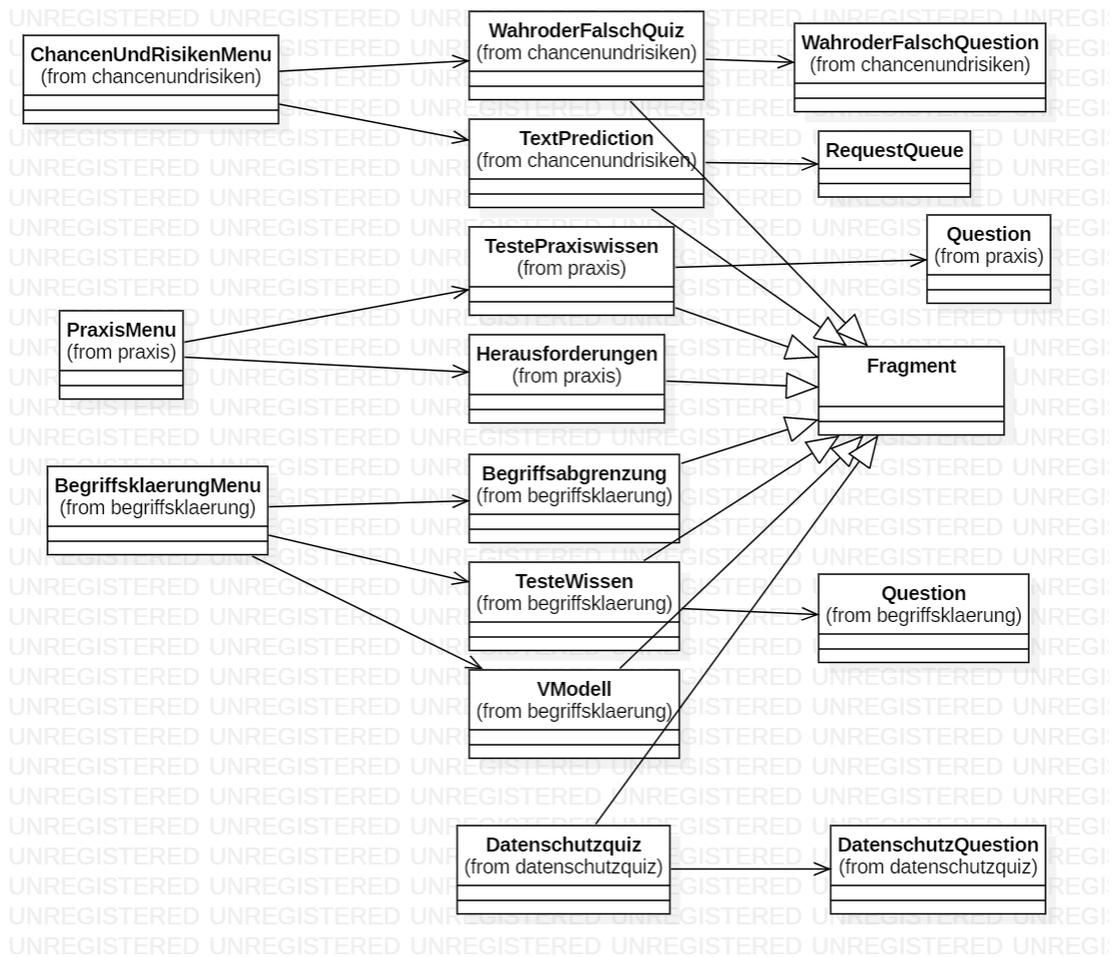


Abbildung XIII: UML-Klassendiagramm des Lernprogramms, welches die Klassen der Subkomponenten zeigt (Teil 2 von 2)

Damit ist der Entwurf abgeschlossen und dieses konzeptionelle Grundgerüst dient als Ausgangspunkt für die Implementierung des Lernprogramms.

6 Realisierung

Dieses Kapitel befasst sich damit, inwieweit die Anforderungen (siehe Kapitel 4.4) erfüllt und der Entwurf (siehe Kapitel 5) umgesetzt worden sind. Hierbei wird der Fokus besonders auf die Dinge, die gut bzw. weniger gut geklappt haben.

Besonders positiv ist die gut strukturierte und verständliche Dokumentation für Android-Entwickler aufgefallen. Diese hat sehr dabei geholfen, den Lebenszyklus von Activities und Fragmenten nachzuvollziehen. Außerdem wird viel Beispielcode bereitgestellt, so dass die Navigationsleiste schnell und simpel umgesetzt werden konnte. *Abbildung XIV* zeigt diese während der Programmaufzeit. Das Konzept aus *Abbildung XI* konnte so wie geplant umgesetzt werden.

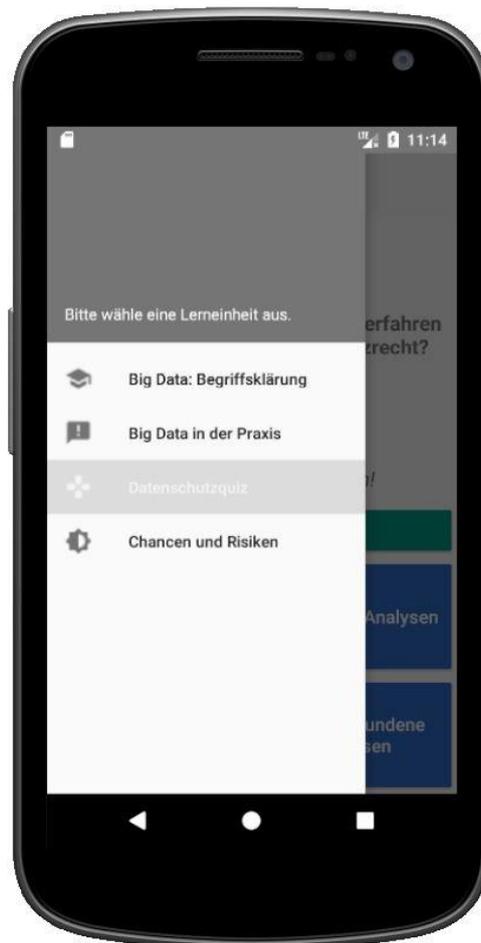


Abbildung XIV: Das Big Data-Lernprogramm ausgeführt im Emulator. Zu sehen ist die Navigationsleiste und das Datenschutzquiz im Hintergrund.

Des Weiteren konnten die Komponenten aus dem Entwurf 1:1 als Pakete mitsamt Klassen abgebildet werden. Durch die modulare Struktur sind auch umfangreichere nachträgliche Quellcodeveränderungen problemlos möglich gewesen, ohne die anderen Komponenten anpassen zu müssen.

Des Weiteren bietet Android viele Möglichkeiten hinsichtlich des *GUI*-Designs. In Kombination mit der Entwicklungsumgebung „Android Studio“ ist es möglich Gestaltungselemente wie in einem Baukasten per Mausklick zusammenzusetzen und mittels eines bereitgestellten Emulators von mobilen Endgeräten auszuprobieren. Außerdem hat Android Studio bei der Einhaltung der in (Android, 2018) beschriebenen Gestaltungsrichtlinien (Styleguides) unterstützt. Aufgrund des bereitgestellten Emulators konnte die Korrektheit und die Robustheit der App ausgetestet werden, was durch gewöhnliche automatisierte Tests (z.B. mithilfe von *JUnit*) nicht ausreichend möglich gewesen wäre.¹⁰ Hierbei ist anzumerken, dass diese Tests nicht erschöpfend sind. Sie sind so strukturiert, dass lediglich stichprobenartig verschiedene ausgedachte Anwendungsszenarien durchgespielt worden sind (z.B. die mehrfache Bearbeitung des Datenschutzquiz und bei jedem Durchlauf sind stets unterschiedliche Antworten ausgewählt worden). Diese Art des Testens hat dabei geholfen, einige Fehler aufzudecken.¹¹ Um jedoch umfassender und automatisiert testen zu können, empfiehlt es sich, zusätzlich dazu GUI-Testwerkzeuge wie z.B. (Espresso, 2018) oder (Selendroid, 2018) zu verwenden.

Nichtsdestoweniger ist Folgendes negativ anzumerken: Bei der Implementierung des Datenschutzquiz sind Anzeigefehler aufgefallen, nachdem bestimmte Buttons gedrückt wurden. Da zur Laufzeit jedoch keine *Exception* geworfen wurde und wegen der komplexen Struktur von Android-Projekten bestehend aus Java-Klassen und darin referenzierten *XML*-Dateien, ist es trotz *Debuggings* nicht möglich gewesen, den Fehler zu lösen bzw. überhaupt konkret zu lokalisieren. Weil es sich jedoch nicht um einen schwerwiegenden Fehler gehandelt hat, konnte dieses Problem durch eine alternative Implementierung umgangen werden.

Zum jetzigen Stand ist die Klasse „TextPrediction“ (siehe *Abbildung XII*) bzw. das Submodul „Persönlichkeitstest“ (siehe *Abbildung X*) noch nicht fertiggestellt, da momentan der zur Verfügung gestellte API-Key noch nicht freigeschaltet wurde.

Alles in allem sind jedoch zum derzeitigen Stand sämtliche (bisher erfassten) Anforderungen umgesetzt worden und bis auf die eben beschriebene Ausnahme konnten alle konzeptionellen Komponenten realisiert werden.

¹⁰ Das Programm besitzt keine automatisierten Unit-Tests, weil fast alle Klassen eine Benutzeroberfläche haben. Um diese zu testen wurden stattdessen manuelle GUI-Tests durchgeführt.

¹¹ Das hat (vorerst) scheinbar ausgereicht, da bei der späteren Evaluation der App durch echte Anwender kein Programmfehler (Bug) aufgetreten ist.

7 Evaluation

In diesem Kapitel wird beschrieben, wie die App sowohl in Bezug auf *Usability*-Aspekte als auch hinsichtlich des Erreichens des Lernerfolgs bewertet worden ist. Hierbei ist für jeden Gesichtspunkt ein eigener Test entwickelt und mit einigen Personen aus der Zielgruppe (siehe Abschnitt 4.2) durchgeführt worden.

Zunächst werden die Testteilnehmer kurz charakterisiert und anschließend für jeden Test jeweils der Entwurf, die Durchführung und Auswertung erläutert. Am Ende wird dann ein Fazit gezogen.

7.1 Beschreibung der Testteilnehmer

Es sind sieben freiwillige Teilnehmer herangezogen worden, die allesamt nach eigener Angabe wenig bzw. kein Wissen über Big Data gehabt haben. Das Alter der Teilnehmer erstreckt sich von 16 bis 58 Jahren. Außerdem sind beide Geschlechter vertreten gewesen (drei weibliche und vier männliche Testteilnehmer). Darüber hinaus haben alle Personen bestätigt, in der Lage zu sein, ein Android-Smartphone bedienen zu können.

Da nach (Kerres, 2012) die Motivation bzw. das Interesse eine maßgebliche Rolle beim Lernen spielen, sind nur diejenigen Personen berücksichtigt worden, die die Bereitschaft signalisiert haben, das Lernprogramm umfassend (d.h. alle Lerneinheiten) auszuprobieren.

7.2 Usability-Test

7.2.1 Entwurf und Ablauf

Mithilfe dieses Tests sollte geprüft werden, ob bzw. in welchen Bereichen der App die Anwender Bedienungsschwierigkeiten gehabt haben. Dabei lehnt sich dieser an entsprechende Tests an, die in (Knabbe, 2011) beschrieben sind. Analog dazu sind die folgenden Usability-Aspekte aus (Nielsen, 1993) berücksichtigt worden:

- **Learnability:** Wie schnell kann der Anwender selbstständig den Umgang mit dem System erlernen?
- **Errors:** Wie oft macht der Anwender Fehler¹² bei der Verwendung des Systems und wie schwerwiegend sind diese?
- **Satisfaction:** Wie zufrieden sind die Anwender während der Systemnutzung mit der gesamten Software bzw. mit einzelnen Komponenten?

¹² gemeint sind Fehler bei der Bedienung des Systems und nicht falsche Antworten bei den Übungsmodulen

Die Testteilnehmer haben zunächst das Lernprogramm komplett abgearbeitet. Während dieser Zeit sind diese von mir dahingehend beobachtet worden, in welchem Ausmaß die oben genannten Eigenschaften in Erscheinung getreten sind. Nach Abschluss der App haben die Teilnehmer jeweils ein eigenes Feedback hinsichtlich dieser Merkmale gegeben. Der zeitliche Ablauf des Tests wird in *Abbildung XV* zum besseren Verständnis visualisiert.

7.2.2 Beobachtung und Feedback

In Bezug auf „Learnability“ ist mir aufgefallen, dass die Testteilnehmer im Wesentlichen mit der App-Navigation zurechtkommen und die Funktionsweise von Icons, Buttons und anderen Gestaltungselementen umgehend begriffen haben. Dieser Eindruck ist durch das spätere Feedback bestätigt worden.

Während der Prüfung der „Error-Eigenschaft“ musste ich bei zwei Teilnehmern eingreifen, um ihnen kurz weiterzuhelfen. Diese wollten nämlich innerhalb der Lerneinheit „Big Data: Begriffsklärung“ zuerst den Wissenstest durchführen, ohne vorher die erforderlichen Submodule zur Aneignung der Lerninhalte abgearbeitet zu haben. Ansonsten ist aufgefallen, dass lediglich drei Personen gemerkt haben, dass die angezeigten Bilder eine Zoom-Funktion besitzen. Abhilfe können hier weitere Dialogfenster bei den betroffenen Modulen schaffen, die den Ablauf und die Funktionsweise erläutern.

Hinsichtlich der „Satisfaction-Eigenschaft“ haben alle Teilnehmer die spielerische Umsetzung des Datenschutzquiz und des Submoduls „Anwendungsszenarien“ (vgl. Abschnitt 5.1) gelobt und als „unterhaltsam“, „kreativ“ und „lustig“ empfunden. Nach Aussage der Teilnehmer werden die Informationen der beiden anderen Komponenten zwar übersichtlich und gut präsentiert, jedoch bieten sie keine Unterhaltung an. Außerdem haben drei Personen gemeint, dass die Komponente „Herausforderungen für Anwender“ „zu technisch“ und „uninteressant“ sei. Das liegt vermutlich daran, dass kein Testteilnehmer ein potentieller Anwender von Big Data ist, sondern eher zu den Betroffenen zählt. Das weitere Vorgehen mit dieser Komponente wird im nächsten Kapitel *Ausblick* erläutert.

7.3 Test des Lernerfolgs

7.3.1 Entwurf und Durchführung

Dieser Test hat überprüft, ob die meisten in Kapitel 4.4.1 aufgeführten Lernergebnisse erreicht wurden und das angeeignete Wissen Wochen später abgerufen werden konnte. Dabei sollte nicht nur die Lerneffektivität (d.h. das bloße Erreichen des Lernziels), sondern auch die Lerneffizienz (d.h. die Erreichung der Lernziele unter Berücksichtigung des zeitli-

chen Lernaufwands) betrachtet werden. Diesbezüglich ist ein Fragebogen¹³ mit einer Bearbeitungsdauer von maximal 30 Minuten und mit Aufgaben zum Testen der Lernergebnisse aus *Tabelle 9* erstellt worden. Einige Lernergebnisse bezüglich Themenblock **(3)** und **(4)** sind im Test nicht berücksichtigt worden (siehe *Tabelle 12*), da sie sich auf fortgeschrittenere Themen beziehen, aber der Test darauf abzielt, nur grundlegende Inhalte des Lernprogramms abzufragen, um die Teilnehmer nicht zu überfordern.

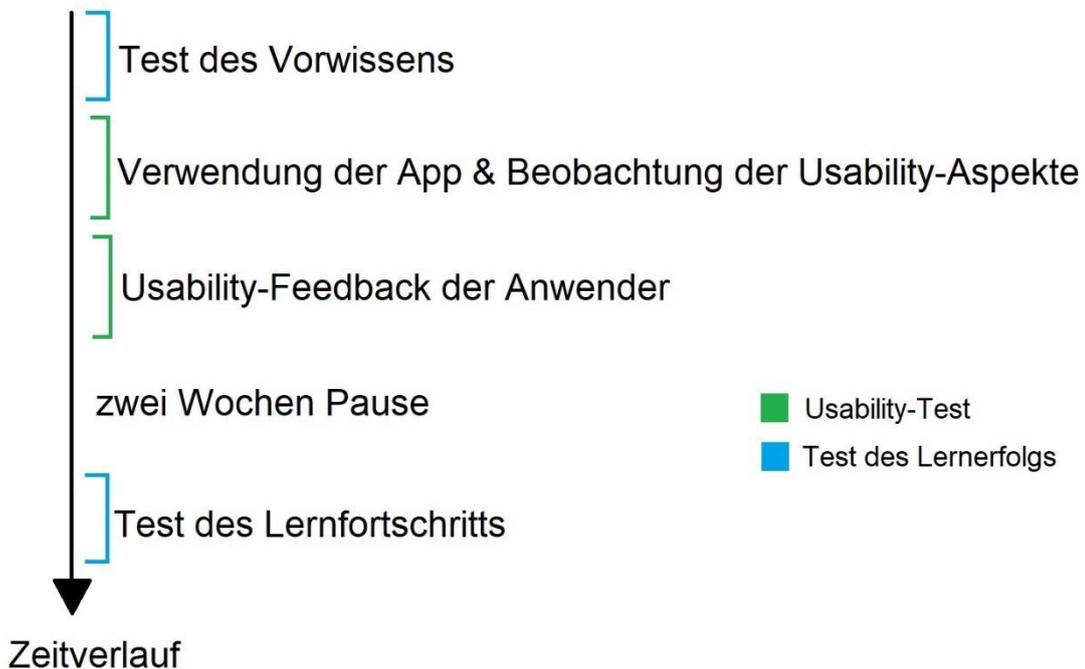


Abbildung XV: Der zeitliche Ablauf der Tests (nicht maßstabsgetreu).

Wie aus *Abbildung XV* zu entnehmen ist, ist dieser Test (im Folgenden „Test 1“ genannt) zunächst vor Bearbeitung des Lernprogramms durchgeführt worden, um das Vorwissen zu überprüfen. Zwei Wochen nach der Verwendung der App ist der gleiche Test noch einmal ausgehändigt worden (im Folgenden „Test 2“ genannt). Alle Testteilnehmer sind darüber nicht in Kenntnis gesetzt worden, um eine explizite Vorbereitung zu verhindern. Beide Testergebnisse haben anschließend als Grundlage zur Messung der Lernfortschritte gedient.

7.3.2 Auswertung

Vergleicht man die erzielten Ergebnisse von Test 1 und 2, so stellt man fest, dass bei Test 1 im Durchschnitt weniger Punkte pro Aufgabe erzielt worden sind als bei Test 2 (siehe *Abbildung XVI*). Erwartungsgemäß hat das kaum vorhandene Vorwissen der Teilnehmer zu ei-

¹³ zu finden ist dieser als Inhalt der beiliegenden CD-ROM

nem schwachen Abschneiden bei Test 1 geführt (insbesondere Aufgabe 5 ¹⁴ ist von vielen Teilnehmern nicht bearbeitet worden).

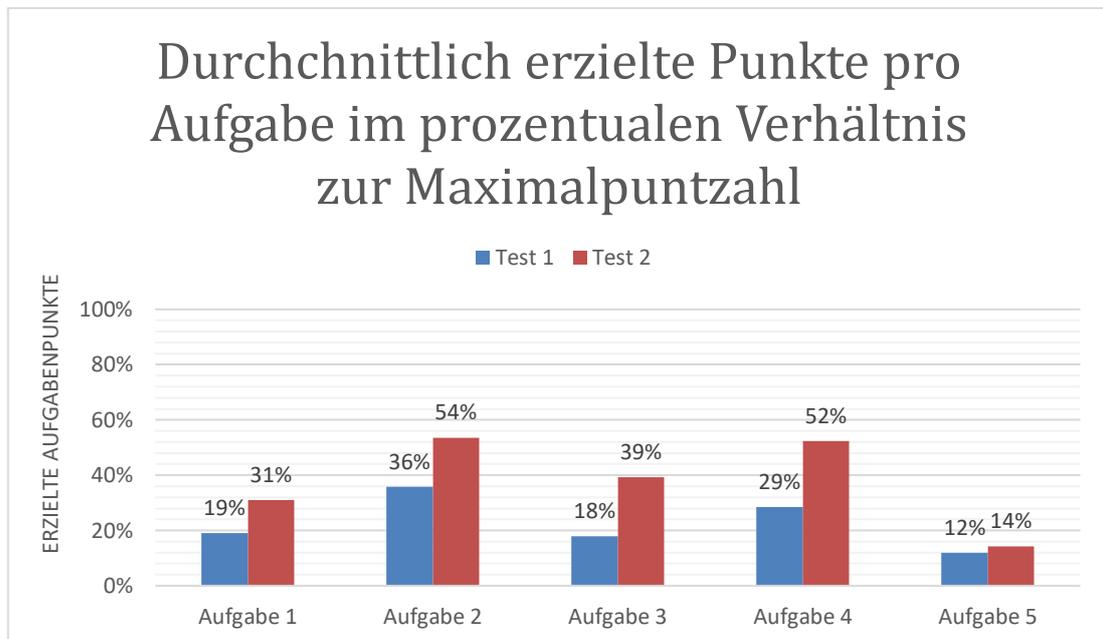


Abbildung XVI: Vergleich beider Testergebnisse. Z.B. haben die Teilnehmer des Tests 1 bei Aufgabe 1 im Durchschnitt 19% der erreichbaren Punkte erzielt.

Wie die *Abbildung XVI* zeigt, ist es bei jeder Aufgabe zu einer erkennbaren Verbesserung der Punktzahl gekommen. Besonders die Aufgaben über das Datenschutzrecht (Aufgabe 3 und 4) verzeichnen deutliche Steigerungen. Außerdem zeigen die Ergebnisse von Test 2, welche Aufgaben (und damit auch die entsprechenden Themenbereiche) relativ gut bzw. eher schlecht beherrscht wurden. Beispielsweise ist Aufgabe 2 ¹⁵, die einzige Übung gewesen, die von mehr als einem Teilnehmer komplett richtig gelöst worden ist. Dahingegen ist bei Aufgabe 5 zwar ein leicht verbessertes, jedoch ein weiterhin schwaches Ergebnis aufgefallen.

Daraus kann man schließen, dass das angeeignete Wissen nicht nur während der zweiwöchigen Pause teilweise beibehalten, sondern auch gezielt angewandt werden konnte. Es ist deshalb ein erster Lernerfolg erkennbar.

Lernergebnis	Art des Tests	Erfolg
Das Lernprogramm wird von den Lernenden akzeptiert, genutzt und bewertet.	Bestandteil des Feedbacks beim Usability-Test	Ja

¹⁴ eine Textaufgabe mit dem Fokus auf den rechtlichen Verpflichtungen von Big Data-Anwendern

¹⁵ eine Aufgabe über die Vor- und Nachteile von Big Data

Tn sind in der Lage, die gängigste Definition von Big Data zu nennen und zu bewerten.	Aufgabe 1 in Test 1 & 2	Ja, aber nicht alle Aspekte wurden benannt
Tn können in eigenen Worten Big Data von Cloud-Computing, IoT und Business Intelligence abgrenzen.		nur Cloud-Computing konnte korrekt definiert werden
Tn sind vertraut mit einigen Vor- und Nachteilen der Big Data-Nutzung und können diese wiedergeben.	Aufgabe 2 in Test 1 & 2	Ja
Tn können die Unterschiede und Gemeinsamkeiten zwischen Datenschutz und Datensicherheit aufzeigen.	Aufgabe 3 in Test 1 & 2	Ja, bei einigen Tn
Tn sollen verstehen, wann der Datenschutz greift.	Aufgabe 3 in Test 1 & 2	Ja, bei einigen Tn
Tn können einige die in der DS-GVO genannten Rechte und Pflichten von Betroffenen und Verantwortlichen wiedergeben.	Aufgabe 4 & 5 in Test 1 & 2	Betroffenenrechte konnten wiedergegeben werden, aber nicht die Pflichten für Verantwortliche
Tn erkennen die Probleme des Datenschutzrechts bezüglich Big Data.	Datenschutzquiz des Lernprogramms	die Übungsergebnisse werden nicht persistiert, deshalb ist keine Angabe darüber möglich
Tn wissen grob, welche Prozessschritte in Big Data-Anwendungen durchlaufen werden.	nicht abgefragt	keine Angabe darüber möglich
Tn kennen einige Herausforderungen für die Prozessschritte des Lebenszyklus-Modells.		
Tn sollen gelerntes Wissen unter Durchführung eines Tests wiedergeben können.	Test 1 & 2 sowie die Übungsmodule des Lernprogramms	Ja

Tabelle 12: Eine Übersicht darüber, welche Lernergebnisse auf welcher Weise und mit welchem Erfolg überprüft worden sind.

Die oben aufgeführte *Tabelle 12* fasst die Testergebnisse übersichtlich zusammen und zeigt ebenfalls auf, in welchen Bereichen noch Verbesserungsbedarf besteht.

7.4 Fazit

Der Usability-Test hat vor allem gezeigt, dass die Implementierung von spielerischen und humorvollen Elementen gut bei den Anwendern angekommen ist. Des Weiteren hat der andere Test erkennbare Lernfortschritte in den meisten Themenbereichen aufgezeigt. Dennoch ist dies unter Vorbehalt zu genießen, da erstens eine Stichprobe von sieben Teilnehmern nicht repräsentativ ist, um generelle Aussagen, über die Wirksamkeit des Lernprogramms zu treffen. Zweitens hat kein Teilnehmer ein sehr gutes Testergebnis¹⁶ erzielt, so dass der Lernerfolg noch ausbaufähig ist. Darüber hinaus sind der Einfachheit halber nicht alle Lerninhalte abgefragt worden (siehe *Tabelle 12*). Außerdem kann es sein, dass noch nicht alle Schwachstellen (aus Usability-Perspektive) identifiziert wurden. Da alle Teilnehmer alle Tests unter meiner Aufsicht durchgeführt haben, kann keine Aussage über die Abbruchquote beim selbstgesteuerten Lernen mit dieser App gemacht werden. Nichtsdestoweniger bin ich damit zufrieden, dass die App den Teilnehmern gefallen hat und damit Lernfortschritte erzielt worden sind.

¹⁶ d.h. größer gleich 90% der Maximalpunktzahl

8 Ausblick

Für den langfristigen Erfolg des Lernprogramms reicht die alleinige Akzeptanz durch die Anwender nicht aus. Entscheidend ist auch, wie gut dieses zugänglich ist und beworben wird, um ein möglichst großes Publikum zu erreichen (vgl. Abschnitt 4.4.1). Für die oben beschriebenen Tests sind die Teilnehmer persönlich angesprochen und assistiert worden und haben lediglich mit einem Smartphone-Emulator interagiert. Als Nächstes könnte man die App auf der Plattform „Google Play Store“ (GooglePlayStore, 2018) zum selbstgesteuerten Lernen zur Verfügung stellen. Dann wäre es möglich, die Anzahl der Installationen sowie die Sternebewertung zu betrachten, um die Popularität der App bewerten zu können. Diesbezüglich muss der Lernerfolg weiterhin messbar sein, weswegen handschriftliche und mündliche Tests ungeeignet sind. Stattdessen kann man den Test als Bestandteil der App realisieren und die Ergebnisse könnte man über das Internet an eine cloudbasierte Datenbank schicken. Dafür wäre z.B. „Firebase Realtime Database“ (Firebase, 2018) geeignet. Damit wäre die skalierbare Überprüfung des Lernerfolgs möglich. Des Weiteren könnte der Anwender dadurch eine Rückmeldung hinsichtlich seines Abschneidens im Vergleich zu den anderen bekommen.

Eine weitere Erweiterungsidee ist es, die Lerninhalte der App derart aufzuteilen, dass es einen eigenen Bereich für Big Data-Anwender und einen für Big Data-Betroffene gibt. Dementsprechend könnte man weitere technisch-orientierte Lerneinheiten für potentielle Big Data-Anwender einführen (z.B. über Konzepte von *NoSQL*-Systemen). Die Testergebnisse und das Feedback aus Kapitel 7 deuten darauf hin, dass dies eine sinnvolle Idee ist, da die gefragten Testteilnehmer Probleme mit diesen fortgeschrittenen und praxisbezogenen Inhalten gehabt haben. Des Weiteren könnte man für jeden Bereich einen eigenen Test zur Erfassung der Lernfortschritte entwickeln.

Schließlich kann man das Programm um automatisierte GUI-Tests erweitern (siehe Ausführungen in Kapitel 6), um die Korrektheit und Robustheit der Software noch besser überprüfen zu können.

Glossar

Vertikale Skalierung: hardwaretechnische Aufrüstung von Rechnern

Horizontale Skalierung: Erweiterung eines verteilten Systems mittels zusätzlicher Rechnerknoten

Assoziationsanalyse: eine regelbasierte Methode, um Zusammenhänge zwischen Variablen in einem Datenbestand zu entdecken

Clustering: eine Methode, um neue Klassen zu finden und Datensätze diesen zuzuordnen („Häufchenbildung“)

Predictive Analytics: Verwendung statistischer Methoden, um aktuelle und historische Daten zu analysieren und Vorhersagen zu zukünftigen Ereignissen zu treffen

Open Data: Daten, die der Öffentlichkeit frei zur Verfügung gestellt werden und ohne Einschränkung verbreitet und wiederverwendet werden können

B2B (business-to-business): geschäftliche Beziehungen zwischen Unternehmen

Data Governance: Maßnahmen, mit dem Ziel eine hohe Datenqualität während des kompletten Lebenszyklus der Daten zu gewährleisten

Compliance: Prüfung der Einhaltung von Vorschriften und Richtlinien

NoSQL: Sammelbegriff für Open Source-Datenbankmanagementsysteme, die moderne Anforderungen erfüllen sollen und schemafrei, horizontal skalierbar und nicht-relational sind

Korrelation: statistischer Zusammenhang zweier Variablen

Precision: Ein Klassifikationsmaß, das den Anteil der relevanten Instanzen unter den gefundenen Instanzen angibt

Recall: Ein Klassifikationsmaß, das den Anteil von gefundenen relevanten Instanzen in der Gesamtmenge aller relevanten Instanzen angibt

Kausalität: Ursache-Wirkung-Beziehungen zwischen Ereignissen

Bestätigungsfehler: Tendenz, diejenigen Informationen selektiv auszuwählen, welche die vorbestehenden Ansichten und Hypothesen bestätigen

Twitter: ein soziales Netzwerk zum Bloggen und Verbreiten von Nachrichten

Natürliche Person: rechtswissenschaftliche Bezeichnung für den Menschen als Träger von Pflichten und Rechten

Allgemeine Geschäftsbedingungen: einseitig vorformulierte Vertragsklauseln, die zum standardisierten und massenhaften Abschluss von Verträgen dienen

Maschinelles Lernen: ein Teilbereich der künstlichen Intelligenz, bei dem statistische Verfahren zur Erstellung von Vorhersagen und Entscheidungen über unbekannte Daten auf der Basis von vorliegenden Daten verwendet werden

Neugiermotiv: intrinsische Motivation, welche dazu veranlasst, sich mit neuen und unbekanntem Reizen zu beschäftigen

Objektorientierte Programmierung: ein Programmierparadigma, welches auf dem Konzept von „Objekten“ (d.h. eine Datenstruktur, welche aus Attributen und Methoden bestehen kann) beruht

Stakeholder: alle Personen bzw. Gruppen, die sich für den Verlauf und Ausgang des Projektes interessieren

Data Science: eine wissenschaftliche Disziplin, welche sich mit der Extraktion von Wissen und Information aus Daten beschäftigt

Activity: ein zentrales Konzept der Android-Programmierung, welches als Einstiegspunkt für die Interaktion zwischen App und Benutzer dient

Fragment: ähnlich wie eine Activity, aber derartig eingeschränkt, dass dessen Lebenszyklus an Activities gekoppelt ist, ein Fragment kann daher als eine „Unter-Activity“ betrachtet werden

Serife: kleine Querstriche an den Ecken und Enden von Buchstaben

Farbsättigung: beschreibt die Intensität und Reinheit eines Farbtons

Kopplung: ein Maß, welches die Stärke des Zusammenhangs zwischen mehreren Entwurfskomponenten angibt

API: englische Abkürzung für „Programmierschnittstelle“

HTTP: Das Hypertext Transfer Protocol ist ein Internetprotokoll zur Datenübertragung zwischen Anwendungen.

GUI: englische Abkürzung für „grafische Benutzerschnittstelle“

XML: Abkürzung für „Extensible Markup Language“, eine Auszeichnungssprache zur Gliederung und Formatierung der Inhalte von Textdateien in einer maschinenlesbaren hierarchischen Struktur.

JUnit: ein Programmiergerüst für Java zum automatisierten Testen von funktionalen Programmbestandteilen

Exception: englisch für „Ausnahme“, wird in Java im Fehlerfall erzeugt und dient zur Fehlerbehandlung

Debugging: ein Vorgang zur Fehleranalyse, bei dem man zur Laufzeit das Programm schrittweise durchlaufen sowie anhalten und man die Zuweisung von Variablenwerten nachverfolgen kann

Usability: englisch für „Gebrauchstauglichkeit“, beschreibt das Ausmaß, in welchem eine Software durch den Anwender verwendet werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen

HTML5: die fünfte Version einer Auszeichnungssprache zur Strukturierung und Repräsentation von Webinhalten

JavaScript: eine Programmiersprache, die dynamisches HTML auf Internetseiten ermöglicht

Templates: Vorlagen für Designs, Formatierungen usw.

Literaturverzeichnis

Android. 2018. Design for Android. [Online] 25. April 2018. [Zitat vom: 12. Juli 2018.] <https://developer.android.com/design/>.

- ApplyMagicSauce. 2018.** [Online] 2018. [Zitat vom: 27. Juli 2018.] <https://applymagicsauce.com/>.
- Bachmann, Ronald, Kemper, Guido und Gerzer, Thomas. 2014.** *Big Data - Fluch oder Segen? - Unternehmen im Spiegel gesellschaftlichen Wandels.* s.l.: mitp, 2014. 978-3826696909.
- Baumgartner, Peter. 1997.** Didaktische Anforderungen an (multimediale) Lernsoftware. [Buchverf.] Ludwig J. Issing, Klimsa und Paul. *Information und Lernen mit Multimedia.* s.l.: BeltzPVU, 1997.
- Bibliographisches Institut Mannheim. 1975.** *Meyers Enzyklopädisches Lexikon.* s.l.: Bibliographisches Institut, 1975. Bd. 15. 9783411012503.
- Bureau of Labor Statistics. 2015.** Long-term price trends for computers, TVs, and related items. [Online] 13. Oktober 2015. [Zitat vom: 27. April 2018.] <https://www.bls.gov/opub/ted/2015/long-term-price-trends-for-computers-tvs-and-related-items.htm>.
- BVerfG. 1983.** Volkszählungsurteil. [Online] 15. Dezember 1983. [Zitat vom: 27. April 2018.] <http://www.servat.unibe.ch/dfr/bv065001.html>.
- Deterding, Sebastian, et al. 2011.** From game design elements to gamefulness: defining "gamification" : MindTrek, 2011. Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. S. 9-15. 978-1-4503-0816-8.
- Dorschel, Joachim. 2015.** *Praxishandbuch Big Data - Wirtschaft – Recht – Technik.* s.l.: Gabler Verlag, 2015. 978-3-658-07288-9.
- Espresso. 2018.** [Online] 2018. [Zitat vom: 01. August 2018.] <https://developer.android.com/training/testing/espresso/>.
- Fasel, Daniel und Meier, Andreas. 2016.** *Big Data – Grundlagen, Systeme und Nutzungspotenziale.* s.l.: Springer Vieweg, 2016. 978-3-658-11588-3.
- Firestore. 2018.** [Online] 2018. [Zitat vom: 30. Juli 2018.] <https://firebase.google.com/docs/database/>.
- Freytag, Johann-Christoph. 2014.** Grundlagen und Visionen großer Forschungsfragen im Bereich Big Data. *Informatik-Spektrum.* 2014, 2. S. 97-104.
- Gandomi, Amir und Haider, Murtaza. 2015.** Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management.* 2015, Bd. 35, 2, S. 137-144.
- Gantz, John, Reinsel, David und Rydning, John. 2017.** [Online] 2017. [Zitat vom: 29. März 2018.] <https://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>.
- Gapski, Harald. 2015.** *Big Data und Medienbildung: Zwischen Kontrollverlust, Selbstverteidigung und Souveränität in der digitalen Welt.* s.l.: kopaed, 2015. 978-3867364034.
- Gartner. 2018.** Gartner Says Worldwide Sales of Smartphones Returned to Growth in First Quarter of 2018. [Online] 29. Mai 2018. [Zitat vom: 16. Juli 2018.] <https://www.gartner.com/newsroom/id/3876865>.

- Glide. 2018.** An image loading and caching library for Android focused on smooth scrolling. [Online] 2018. [Zitat vom: 09. Juli 2018.] <https://github.com/bumptech/glide>.
- GooglePlayStore. 2018.** [Online] 2018. [Zitat vom: 30. Juli 2018.] <https://play.google.com/store>.
- Hurwitz, Judith, et al. 2013.** *Big Data For Dummies*. s.l. : John Wiley & Sons, Inc., 2013. 978-1-118-50422-2.
- IBM.** The four Vs of Big Data. [Online] [Zitat vom: 27. April 2018.] http://www.ibmbigdatahub.com/sites/default/files/infographic_file/4-Vs-of-big-data.jpg.
- IDC. 2016.** Double-Digit Growth Forecast for the Worldwide Big Data and Business Analytics Market. [Online] 03. Oktober 2016. [Zitat vom: 27. April 2018.] <https://www.idc.com/getdoc.jsp?containerId=prUS41826116>.
- Jagadish, H. V., et al. 2014.** Big data and its technical challenges. *Communications of the ACM*. 2014, Bd. 57, 7, S. 86-94.
- jfc Medienzentrum e.V. 2015.** Big Data - Eine Arbeitshilfe für die Jugendarbeit. [Online] 2015. [Zitat vom: 27. April 2018.] http://www.jfc.info/data/Big-Data_Broschu__re_WEB_V9.pdf.
- Kerres, Michael. 2012.** *Mediendidaktik : Konzeption und Entwicklung mediengestützter Lernangebote*. s.l. : Oldenbourg, 2012. 978-3-486-27207-9.
- King, Stefanie. 2014.** *Big Data - Potential und Barrieren der Nutzung im Unternehmenskontext*. s.l. : VS Verlag für Sozialwissenschaften, 2014. 978-3-658-06585-0.
- Klein, Dominik, Tran-Gia, Phuoc und Hartmann, Matthias. 2013.** Big Data. *Informatik-Spektrum*. 2013, Bd. 35, 3.
- Knabbe, Martin. 2011.** Evaluierung von Grundsätzen für das Lernen in Videospielen. [Online] 26. April 2011. [Zitat vom: 12. Juli 2018.] <http://edoc.sub.uni-hamburg.de/haw/volltexte/2011/1196/pdf/Bachelorarbeit.pdf>.
- Kolany-Raiser, Barbara und Radtke, Tristan. 2018.** [Online] 2018. [Zitat vom: 27. März 2018.] http://www.abida.de/sites/default/files/16_Microtargeting.pdf.
- König, Christian, Schröder, Jette und Wiegand, Erich. 2017.** *Big Data - Chance, Risiken, Entwicklungstendenzen*. s.l. : Springer VS, 2017. 978-3-658-20082-4.
- KPMG (a). 2016.** [Online] 2016. [Zitat vom: 09. März 2018.] <https://assets.kpmg.com/content/dam/kpmg/pdf/2016/06/kpmg-mdws-2016-%C3%B6ffentlichersektor-sec.pdf>.
- KPMG (b). 2016.** [Online] 2016. [Zitat vom: 09. März 2018.] <https://cdn2.hubspot.net/hubfs/571339/LandingPages-PDF/kpmg-mdws-201-sec.pdf>.
- Laney, Doug. 2001.** [Online] 2001. [Zitat vom: 27. März 2018.] <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.
- Löhr, Bea Alice. 2012.** *Erstellung eines multimedialen Lernprogramms zur Auskultation von Hund und Katze*. s.l. : LMU München: Tierärztliche Fakultät, 2012.
- Meir, Susanne. 2006.** [Online] 2006. [Zitat vom: 19. April 2018.] https://lehrerfortbildung-bw.de/st_digital/elearning/moodle/praxis/einfuehrung/material/2_meir_9-19.pdf.
- Minass, Erik. 2002.** *Dimensionen des E-Learning*. s.l. : Smart Books, 2002. 978-3908492382.

- Nieding, Gerhild, Ohler, Peter und Rey, Günther Daniel. 2015.** *Lernen mit Medien.* s.l. : Schöningh, 2015. 9783825240011.
- Nielsen, Jakob. 1993.** *Usability Engineering.* San Francisco : Morgan Kaufmann, 1993. 978-0-12-518406-9.
- NIST. 2011.** The NIST Definition of Cloud Computing. [Online] 2011. [Zitat vom: 27. April 2018.] <https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf>.
- O'Neil, Cathy. 2017.** *Angriff der Algorithmen: wie sie Wahlen manipulieren, Berufschancen zerstören und unsere Gesundheit gefährden.* s.l. : Carl Hanser Verlag, 2017. 978-3446256682.
- SAP. 2012.** Small and midsize companies look to make big gains with big data. [Online] 26. Juni 2012. [Zitat vom: 27. April 2018.] <http://global.sap.com/news-reader/index.epx?PressID=19188>.
- Selendroid. 2018.** Selenium for Android. [Online] 2018. [Zitat vom: 01. August 2018.] <http://selendroid.io/>.
- Stangl, Werner. 2018.** [werner stangl]s arbeitsblätter. [Online] 2018. [Zitat vom: 27. April 2018.] <http://arbeitsblaetter.stangl-taller.at/LERNEN/Lerndefinitionen.shtml>.
- TNS-Infratest. 2016.** Kennen Sie den Begriff "Big Data"? [Online] Januar 2016. [Zitat vom: 30. Juli 2018.] <https://de.statista.com/statistik/daten/studie/25798>.
- van der Sloot, Bart, Broeders, Dennis und Schrijvers, Erik. 2016.** *Exploring the Boundaries of Big Data.* s.l. : WRR Verkenningen, 2016. 978 94 6298 358 8.
- Vodafone. 2016.** [Online] 2016. [Zitat vom: 03. April 2018.] <http://www.vodafone-institut.de/wp-content/uploads/2016/01/VodafoneInstitut-Survey-BigData-Highlights-de.pdf>.
- Volley. 2018.** An HTTP library that makes networking for Android apps easier and most importantly, faster. [Online] 2018. [Zitat vom: 14. Juli 2018.] <https://github.com/google/volley>.
- Wikipedia. 2017.** Pawlowscher Hund. [Online] 7. Dezember 2017. [Zitat vom: 27. April 2018.] https://de.wikipedia.org/wiki/Pawlowscher_Hund.

Tabellenverzeichnis

Tabelle 1: Hürden und Lösungsansätze von Big Data	12
Tabelle 2: potentielle Vor- und Nachteile von Big Data-Analysen	14
Tabelle 3: Rechte von Betroffenen und Pflichten von Verantwortlichen	19
Tabelle 4: Unterscheidung der Lernarten	23
Tabelle 5: Grad der Selbststeuerung bei Lernprogrammen	25
Tabelle 6: Vor- und Nachteile des Einsatzes von E-Learning-Angeboten	26
Tabelle 7: Übersicht der didaktischen Methoden	27
Tabelle 8: Merkmale der Zielgruppe	32

Tabelle 9: Spezifikation der Lernergebnisse	34
Tabelle 10: Nichtfunktionale Anforderungen	36
Tabelle 11: Didaktische Anforderungen	37
Tabelle 12: Übersicht der erreichten Lernergebnisse	50

Abbildungsverzeichnis

Abbildung I: Big Data-Definitionen	2
Abbildung II: graphische Darstellung des V-Modells	4
Abbildung III: Vereinfachter Zusammenhang zwischen den Fachbegriffen	6
Abbildung IV: Darstellung des Big Data-Lebenszyklus-Modells	8
Abbildung V: Umfrageergebnisse bezüglich des Big Data-Einsatzes	9
Abbildung VI: Umfrageergebnisse bezüglich Meinungen zu Big Data	10
Abbildung VII: Aspekte für den erfolgreichen Einsatz von Big Data-Systemen	16
Abbildung VIII: Darstellung und Beurteilung von Lerntheorien	22
Abbildung IX: E-Learning im bildhaften Verhältnis zum gesamten Lernen	24
Abbildung X: fachliche Softwarearchitektur als UML-Komponentendiagramm	40
Abbildung XI: Mockup des Startbildschirms der App	41
Abbildung XII: UML-Klassendiagramm Teil 1 von 2	42
Abbildung XIII: UML-Klassendiagramm Teil 2 von 2	43
Abbildung XIV: Big Data-Lernprogramm ausgeführt im Emulator	44
Abbildung XV: zeitlicher Ablauf der Tests	48
Abbildung XVI: Vergleich beider Testergebnisse	49

Versicherung über Selbstständigkeit

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne fremde Hilfe selbstständig verfasst und nur die angegebenen Hilfsmittel benutzt habe.

Hamburg, den _____

