



Diese Arbeit wurde vorgelegt am
Lehrstuhl A für Mathematik

Sicherheit der Privatsphäre in sozialen Netzwerken - Wie Mathematik die Nutzer ausspioniert.

Ein Lehr-Lern-Modul im Rahmen eines mathematischen
Modellierungstages für Schülerinnen und Schüler der Sekundarstufe I

Privacy's security in sozial networks - How mathematics spy on the users.

A teaching and learning module in the context of a mathematical modeling
day for pupils at German secondary level I

Masterarbeit
Mathematik

Dezember 2018

Vorgelegt von
Presented by

Nils Steffen



Erstprüfer
First examiner

Prof. Dr. Sebastian Walcher
Lehrstuhl A für Mathematik
RWTH Aachen University

Zweitprüfer
Second examiner

PD. Dr. Olaf Wittich
Lehrstuhl A für Mathematik
RWTH Aachen University

Koreferent
Co-supervisor

Maike Sube
Lehrstuhl A für Mathematik
RWTH Aachen University

Inhaltsverzeichnis

Tabellenverzeichnis	V
Abbildungsverzeichnis	VI
1. Einleitung	1
2. Didaktischer Hintergrund	3
2.1. Das Schülerlabor CAMMP	3
2.1.1. Vorstellung des Schülerlabors	3
2.1.2. Ziele von CAMMP	4
2.2. Mathematische Modellierung	4
2.2.1. Das mathematische Modell	5
2.2.2. Modellierungskreisläufe	6
2.2.3. Der PPDAC Zyklus	10
3. Fachlicher Hintergrund	12
3.1. Problemstellung	12
3.2. Das soziale Netzwerk Friendster	13
3.2.1. Das Unternehmen Friendster Inc.	13
3.2.2. Das soziale Netzwerk als ungerichteter Graph	14
3.2.3. Informationen zum verwendeten Datensatz	17
3.2.4. Aufbereitung der Daten	18
3.2.5. Codierung der Informationen in MATLAB	21
3.3. Heuristiken definieren und anwenden	22
3.3.1. Vorgehen zur Untersuchung der Fragestellung	22
3.3.2. Zu untersuchende Heuristiken	25
3.3.3. Ergebnisse der verwendeten Heuristiken	27
3.4. Lagemaße	29
3.4.1. Das arithmetische Mittel	29
3.4.2. Der Median	30
3.4.3. Der Modus	31
3.5. Streuungsmaße	32
3.5.1. Die Spannweite	32
3.5.2. Der Quartilsabstand	33
3.5.3. Die Standardabweichung	34
3.6. Boxplot	34
3.7. Bedingte Wahrscheinlichkeit	36
4. Didaktisch-methodisches Konzept	39
4.1. Ziele und curriculare Einbindung des entwickelten Lernmoduls	39
4.2. Gemeinsamkeiten und Unterschiede zum bestehenden Lernmodul Sicherheit der Privatsphäre in sozialen Netzwerken für die Sek. II	41
4.3. Der Modellierungskreislauf als Struktur des entwickelten Lernmoduls	42

4.4.	Ablauf des CAMMP days	44
4.5.	Vorstellung der Materialien	47
4.5.1.	Modellierungsvortrag	48
4.5.2.	Einführungsvortrag	48
4.5.3.	Einführung in MATLAB	49
4.5.4.	Arbeitsblatt 1 Daten filtern	51
4.5.5.	Erste Sicherungsphase	52
4.5.6.	Zwischenvortrag	53
4.5.7.	Arbeitsblatt 2 Vorhersageregeln anwenden	54
4.5.8.	Zweite Sicherungsphase	57
4.5.9.	Arbeitsblatt 3 Werbung schalten	58
4.5.10.	Dritte Sicherungsphase	60
4.5.11.	Abschlusspräsentation	61
4.5.12.	Material für die Dozenten	62
4.6.	Die Vorzüge der Einbindung der Computersoftware MATLAB in das entwickelte Lernmodul	63
5.	Durchführung und Evaluation des entwickelten Lernmoduls	65
5.1.	Rahmenbedingungen	65
5.2.	Leitende Gesichtspunkte der Beobachtung und der Evaluation	65
5.3.	Vorstellung des Evaluationsbogens	66
5.4.	Vorstellung der gesammelten Beobachtungen und der Ergebnisse aus der Evaluation	67
5.5.	Aus der Evaluation und Beobachtung resultierende Verbesserungen zum entwickelten Lernmodul	72
6.	Ausblick	74
	Literatur	76
	Anhang	80
A.	Begriffe aus der Wahrscheinlichkeitsrechnung	80
B.	Methodisches Konzept	81
C.	Einführungsvortrag	83
C.1.	Folien zum Einführungsvortrag	83
C.2.	Notizen zum Einführungsvortrag	92
D.	Einführung in MATLAB	97
E.	Arbeitsblatt 1	98
E.1.	Zusatzmaterial von Arbeitsblatt 1	103
E.1.1.	Antwortblatt	103

E.1.2. Besprechungsfolie	105
E.2. Musterlösung für Arbeitsblatt 1	107
F. Arbeitsblatt 2	109
F.1. Zusatzmaterial von Arbeitsblatt 2	119
F.1.1. Übersicht über die Vorhersageregeln 1-6	119
F.1.2. Übersicht über die Vorhersageregeln 7-10	120
F.1.3. Antwortblatt	121
F.1.4. Hilfekarte 1 - Wiederholung	124
F.1.5. Hilfekarte 2 - Boxplot	126
F.1.6. Besprechungsfolie	130
F.2. Musterlösung für Arbeitsblatt 2	132
G. Arbeitsblatt 3 für Niveau 1	137
G.1. Zusatzmaterial von Arbeitsblatt 3 für Niveau 1	142
G.1.1. Antwortblatt für Niveau 1	142
G.1.2. Besprechungsfolie für Niveau 1	144
G.2. Musterlösung für Arbeitsblatt 3 Niveau 1	146
H. Arbeitsblatt 3 für Niveau 2	148
H.1. Zusatzmaterial von Arbeitsblatt 3 für Niveau 2	153
H.1.1. Antwortblatt für Niveau 3	153
H.1.2. Zusatzarbeitsblatt für Niveau 2	155
H.1.3. Antwortblatt zur Zusatzaufgabe	158
H.1.4. Hilfekarte 3 - Baumdiagramm	159
H.1.5. Besprechungsfolie für Niveau 2	162
H.2. Musterlösung für Arbeitsblatt 3 Niveau 2	164
I. Abschlussvortrag	167
I.1. Folien zum Abschlussvortrag	167
I.2. Notizen zum Abschlussvortrag	170
J. Evaluation	173
J.1. Evaluationsbogen	173
J.2. Ergebnisse der Evaluation	179

Tabellenverzeichnis

1.	Übersicht des gefilterten Datensatzes aus dem Lernmodul für die Sek. II (Sube, 2016)	18
2.	Übersicht des Datensatzes, nachdem die beiden Filterungen durchgeführt wurden.	20
3.	Ermittelte, gerundete Trefferquoten nach Anwendung der einzelnen Heuristiken auf den Datensatz mit 44747 Nutzern.	28
4.	Durchschnittliche Intervallbreiten, in denen das Alter der Nutzer liegt, welches mit den Heuristiken zur Spannweite, zum Quartilsabstand und zur Standardabweichung vorhergesagt wurde. Die Werte beziehen sich wieder auf den Datensatz bestehend aus 44747 Nutzern.	29

Abbildungsverzeichnis

1.	Sieben-schrittiger Modellierungskreislauf nach Blum und Leiss (vgl. Blum, 2006, S. 9)	6
2.	Eingesetzter Modellierungskreislauf im Schülerlabor CAMMP	8
3.	Verwendetes Modellierungsspirale im Schülerlabor CAMMP (vgl. Frank et al., 2018, S. 140)	9
4.	PPDAC-Zyklus nach Wild und Pfannkuch (vgl. Wild & Pfannkuch, 1999, S. 225)	10
5.	Beispielhafte Veranschaulichung eines sozialen Netzwerks als ungerichteter Graph. Die Nutzer stellen die Knoten dar und die Freundschaftsbeziehungen zwischen den einzelnen Nutzern stellen die Kanten des Graphen dar. (vgl. Sube, 2016, S. 18).	14
6.	Beispielhafte Veranschaulichung der Freundschaftsbeziehungen in einem sozialen Netzwerk. In orange sind die Freunde 1. Ordnung und in blau die Freunde 2. Ordnung von Peter dargestellt (vgl. Sube, 2016, S. 24).	16
7.	Die Altersverteilung im Datensatz aus dem Lernmodul für die Sekundarstufe II (Sube, 2016) zeigt die absoluten Häufigkeiten der Nutzer mit einem Alter zwischen 0 und 100 Jahren. Auffällig sind sowohl Altersangaben von 0 Jahren als auch Angaben um die 90 Jahre.	19
8.	<i>oben</i> : Absoluten Häufigkeiten der Nutzer mit einer gewissen Anzahl an Freunden nach der ersten Filterung; <i>unten</i> : Ausschnitt der absoluten Häufigkeiten der Nutzer mit einer gewissen Anzahl an Freundes-Freunden nach der ersten Filterung.	20
9.	Gespeicherte Informationen in der Computersoftware MATLAB am Beispiel des Nutzers mit der ID 10.	21
10.	Darstellung der Basisvariante eines Boxplots bestehend aus Minimum, Maximum, unteres sowie oberes Quartil und Median. Die Verteilung der Daten wurde dem zugehörigen Boxplot in einem Punktdiagramm gegenübergestellt.	35
11.	Zweistufiges Baumdiagramm zur Bestimmung der theoretischen Trefferwahrscheinlichkeit $P(R)$ mit Hilfe der Vorhersageregeln zum Zufall eine richtige Zuordnung in die beiden Altersintervalle zu tätigen. In grün sind die Wahrscheinlichkeiten der Pfade markiert, die zu einer richtigen Zuordnung bzw. Vorhersage führen.	37
12.	Darstellung der Gemeinsamkeiten und Unterschiede der Materialien für <i>Niveau 1</i> (für eine siebte Klasse) und <i>Niveau 2</i> (für eine achte oder höhere Klasse).	42
13.	Modellierungskreislauf für das entwickelte Lernmodul	43
14.	Beispielhafter Auszug aus dem MATLAB Live Skript, welches die SuS im entwickelten Lernmodul eine Einführung in den Umgang mit dem Programm MATLAB gibt.	47

15. Die drei zentralen, im MATLAB Live Skript verwendeten Icons. *Links*: Icon, welches den SuS einen Arbeitsauftrag signalisiert. *Mitte*: Icon, das vor dem grauen Kasten mit dem enthaltenen Code steht. *Rechts*: Icon, welches die Existenz einer Hilfekarte für die aktuelle Aufgabe aufzeigt. . 51
16. Ausgabe von MATLAB bei einer Aufgabe innerhalb des zweiten Arbeitsblatts zur Bestimmung des arithmetischen Mittels, Medians und Modus zu den Altersangaben der Freunde 1. Ordnung eines ausgewählten Nutzers im Datensatz. MATLAB zeigt an, dass die berechneten Ergebnisse zum arithmetischen Mittel und zum Median noch nicht korrekt sind. . 64

1. Einleitung

In der heutigen Welt ist die Mathematik in zahllosen Bereichen nicht mehr wegzudenken. Effiziente organisatorische Abläufe im Verkehr oder in der Kommunikationstechnik, die Navigation mit dem GPS-Gerät, die Wetterprognose für die nächste Woche oder die Suchanfrage mit Google sind nur ein paar von zahlreichen Beispielen, auf welche die Gesellschaft ohne die Mathematik verzichten müsste. Quasi jedes technische Gerät, das unser Leben erleichtert und für unseren Wohlstand steht, basiert auf Mathematik. In unserer Informations- und Wissensgesellschaft hat sich die mathematische Bildung schon lange zu einer Schlüsselqualifikation etabliert (vgl. Engel, 2018, S. 3). Doch obwohl Mathematik alltäglich ist und in fast allen Lebens- und Berufssparten genutzt wird, werden die Lehrkräfte in den Schule von Seiten der Schülerinnen und Schüler¹ immer wieder mit den Fragen konfrontiert: „Warum soll ich das lernen?“, „Wozu brauche ich das eigentlich später?“. Auch in der Öffentlichkeit ist dieses Bild der Mathematik wiederzufinden. Viele Menschen setzen sie nur mit der Manipulation von Zahlen gleich und sehen die enge Einbettung in die Gesellschaft nicht (vgl. Loos & Ziegler, 2015, S. 9). So hat die Mathematik schon seit einiger Zeit mit einem Relevanz-Paradoxon zu kämpfen (vgl. Engel, 2018, S. 3). Zwar bildet sie die Grundlage unserer technologischen Gesellschaft, jedoch bleiben den meisten Menschen in vielen Bereichen des alltäglichen Lebens die mathematischen Bezüge verborgen (vgl. Engel, 2018, S. 3). Für die Vermittlung eines angemessenen Bildes von Mathematik muss dieses Paradoxon wenigsten teilweise aufgedeckt werden.

Aus diesem Grund spielen schon seit vielen Jahren in den didaktischen Diskussionen die Einbindung von Realbezügen sowie Modellierung im Mathematikunterricht eine bedeutende Rolle. Die SuS sollen bereits in den Schulen erfahren, dass Mathematik Anwendung in ihrem Leben findet und weitaus mehr darstellt als ein Konstrukt aus Formeln und Symbolen. Der Einbezug von anwendungsorientierten Problemen soll den SuS verdeutlichen, welchen Einfluss die Mathematik auf unsere Welt nimmt. Darüber hinaus können nach Westermann durch Realbezüge im Mathematikunterricht ein tieferes Verständnis sowie ein nachhaltigeres Behalten mathematischer Begriffe geschaffen und die Motivation seitens der SuS erhöht werden (vgl. Westermann, 2011, S. 149). Das Schülerlabor CAMMP (Computational and Mathematical Modeling Program) der RWTH Aachen hat sich zur Aufgabe gesetzt, den SuS durch die Erstellung und Vermittlung authentischer Problemstellungen ein Gefühl für die gesellschaftliche Bedeutung der Mathematik zu vermitteln sowie beim Aufbau von Modellierungskompetenzen zu unterstützen. In unterschiedlichen Lernmodulen lösen die SuS mit Hilfe ihrer bereits erlernten mathematischen Kompetenzen und dem Einsatz der Computersoftware MATLAB realitätsnahe Fragestellungen. Durch die aktive Auseinandersetzung mit lebensnahen Problemstellungen lernen die SuS sowohl die Grundlagen der Modellierung als auch die Anwendung ihrer Fähigkeiten auf außermathematische Kontexte kennen. Diese Arbeit beschäftigt sich mit der Gestaltung eines neuen Lernmoduls für die Se-

¹Im weiteren Verlauf der Masterarbeit wird für eine einfachere Lesbarkeit die Abkürzung „SuS“ verwendet.

kundarstufe I zum Thema „Sicherheit der Privatsphäre in soziale Netzwerke - Was Facebook, Instagram und Co. über uns wissen“. Nach der aktuellen JIM-Studie (2017) sind für die Jugendlichen soziale Netzwerke, wie Facebook oder Instagram, ein selbstverständlicher Teil des Alltags (vgl. Feierabend et al., 2017, S. 3). Eine Kommunikation ohne Messenger wie WhatsApp ist heutzutage kaum mehr vorstellbar. Durch die Nutzung sozialer Netzwerke geben die Nutzer auch viele Daten über sich preis. Die Betreiber solcher Netzwerke sammeln die angegebenen Informationen und generieren sogar mit bestimmten Heuristiken Informationen, die ein Nutzer in seinem Profil nicht angegeben hat (vgl. Garcia et al., 2014, S. 2). In dem entwickelten Lernmodul sollen die SuS auf diese Möglichkeit aufmerksam gemacht werden und erfahren, wie mit Hilfe einfacher mathematischer Regeln Vorhersagen über das Alter einer Person getroffen werden können. Dabei arbeiten die SuS mit realen Daten des mittlerweile abgeschalteten Netzwerks Friendster. Zudem sollen die SuS im Laufe des Moduls kritisch über ihren eigenen Umgang mit sozialen Netzwerken reflektieren.

Das erste Kapitel dieser Arbeit stellt kurz das Schülerlabor CAMMP vor, in dessen Rahmen das Lernmodul durchgeführt wird, und konzentriert sich im Anschluss auf den didaktischen Hintergrund der mathematischen Modellierung, welcher eine zentrale Rolle im gesamten Lernmodul einnimmt. Im nächsten Kapitel werden die fachlichen Grundlagen des entwickelten Moduls dargestellt. Die Struktur und das didaktisch methodische Konzept des erarbeiteten Lernmoduls wird im dritten Kapitel vorgestellt. Im Anschluss erfolgt die Beschreibung der ersten Durchführung mit der dazugehörigen Evaluation sowie die daraus resultierenden Verbesserungen. Zum Schluss wird ein kurzer Ausblick über weitere Optimierungsmöglichkeiten gegeben.

2. Didaktischer Hintergrund

In diesem Kapitel wird zunächst kurz das Schülerlabor CAMMP vorgestellt, in dessen Rahmen die Erprobung und Durchführung des entwickelten Workshops stattfindet. Im Anschluss wird auf das didaktische Konzept des Modellierens eingegangen, welches ein zentrales Element des erstellten Lernmoduls sowie aller bestehenden CAMMP Projekte darstellt.

2.1. Das Schülerlabor CAMMP

Die nachfolgenden Informationen zu den beiden Kapitel zur Vorstellung des Schülerlabors und dessen Ziele sind der Internetseite vom Schülerlabor entnommen (vgl. CAMMP).

2.1.1. Vorstellung des Schülerlabors

CAMMP steht für „Computational and Mathematical Modeling Program“ und ist seit 2013 ein mathematisches Schülerlabor an der RWTH Aachen. Organisiert wird CAMMP vom Lehrstuhl A für Mathematik, vom Lehrstuhl für Mathematik CCES und der Graduiertenschule AICES. Das Schülerlabor hat sich zur Aufgabe gesetzt, SuS anhand von praktischen Beispielen die Grundlagen der mathematischen Modellierung näher zu bringen sowie ihnen die Bedeutsamkeit der Mathematik in unserer heutigen Welt aufzuzeigen. Sowohl SuS als auch Lehrkräfte haben hier die Möglichkeit mit realen, authentischen Problemen konfrontiert zu werden, diese mathematisch zu modellieren und mit Hilfe von digitalen Werkzeugen sowie mathematischen Methoden zu lösen. Dabei wird an den aktuellen Kernlehrplänen angeknüpft, aber auch neue, für das jeweilige Lernmodul notwendige Mathematik vermittelt.

Ein Angebot des Schülerlabors ist der CAMMP day, ein mathematischer Modellierungstag, in dessen Rahmen sich abhängig vom Lernmodul SuS aus der Mittel- oder Oberstufe einen Tag lang an der RWTH Aachen mit einer lebensnahen Fragestellung aus Alltag, Industrie oder Wissenschaft beschäftigen. So werden beispielsweise Workshops zu Themen wie GPS, Solarkraftwerk oder Shazam angeboten. In kleinen Gruppen von zwei bis drei Personen arbeiten die SuS möglichst selbstständig an der Lösung der Problemstellung und nutzen dazu größtenteils Inhalte aus dem Mathematikunterricht. Darüber hinaus nutzen die SuS zur Unterstützung bei der Lösung des mathematischen Problems die Software MATLAB. Während des gesamten Tages werden die SuS von wissenschaftlichen Mitarbeitern und Mitarbeiterinnen von CAMMP unterstützt. Ein solcher CAMMP day wurde im Rahmen der vorliegenden Masterarbeit für die Mittelstufe zum Thema Sicherheit der Privatsphäre in sozialen Netzwerken erstellt.

Neben dem CAMMP day bietet das Schülerlabor auch zwei mal im Jahr eine CAMMP week an, bei der SuS eine Woche lang ein Problem aus der Wissenschaft oder Wirtschaft erarbeiten. Die Fragestellungen werden dabei von Partnerfirmen und Instituten der Universität gestellt und sind Bestandteil aktueller Forschungsfragen. Eine Unter-

stützung bei der Erarbeitung der verschiedenen Probleme erhalten die SuS die gesamte Woche über von Lehrkräften, studentischen Hilfskräften, Doktoranden, Professoren sowie Firmenvertretern. Am Ende der Woche findet eine Abschlussveranstaltung statt, bei der den Problemstellern die Ergebnisse der Woche präsentiert werden.

2.1.2. Ziele von CAMMP

Das Schülerlabor CAMMP hat das Ziel, anhand von ausgewählten Beispielen den Umgang mit mathematischer Modellierung zu fördern sowie den SuS die gesellschaftliche Bedeutung von Modellierung und Mathematik aufzuzeigen. Auch sollen neben den inhaltlichen Kompetenzen die im Kernlehrplan verankerten, prozessbezogenen Kompetenzbereiche gefördert werden. Dazu gehören im besonderen Maße das Nutzen von Werkzeugen durch den Einsatz von digitaler Software wie z.B. MATLAB oder GeoGebra, Kommunizieren und Argumentieren sowie Problemlösen und vor allem Modellieren. Für den letzten Bereich lernen die SuS im Laufe ihres Workshops den Modellierungskreislauf kennen und lösen das gestellte Problem durch das zum Teil mehrmalige Durchlaufen dieses Kreislaufes. Durch das Arbeiten in Gruppen erweitern die SuS zusätzlich ihre Teamfähigkeit, indem sie zusammen an einer Fragestellung arbeiten und an auftretenden Problemen gemeinsam wachsen. Ein weiteres Ziel von CAMMP ist es den SuS eine Berufs- und Studienorientierung anzubieten und sie über ein Studium im Bereich CES (Computational Engineering Science), Mathematik oder Informatik zu informieren. So erhalten die SuS die Möglichkeit, im Rahmen des Schülerlabors einen Einblick in diese Bereiche zu bekommen und zu entscheiden, ob ein Studium in einem dieser Fächer von Interesse wäre.

Jedoch profitieren nicht nur SuS von dem Schülerlabor. Auch Lehrkräfte sollen durch die Teilnahme an den Angeboten von CAMMP fortgebildet und für ihren eigenen Unterricht inspiriert sowie motiviert werden. So erhalten die Lehrkräfte durch das Schülerlabor Anregungen, wie komplexe und authentische Problemstellungen in den Unterricht eingebunden werden können. Aber auch die Motivation und das Interesse an der Mathematik, welches die SuS durch das Arbeiten an lebensnahen Situationen bei CAMMP aufbauen, kann die Lehrperson durch das Aufgreifen der mathematischen Inhalte im eigenen Unterricht nutzen. Darüber hinaus arbeitet CAMMP momentan an der Entwicklung von Materialien, welche auch in der Schule eingesetzt werden können, um das Einbinden von authentischen Problemen und der Kompetenz des Modellierens in der Schule zu erleichtern. Auch in der Lehrerbildung ist CAMMP aktiv, indem es angehende Lehrkräfte ein Lehr-Lern-Labor anbietet. Hier wird Lehramtsstudenten die Möglichkeit geboten, in praxisorientierten Lehrveranstaltungen erste Unterrichtserfahrungen mit SuS zu sammeln und sich im Bereich der mathematischen Modellierung fortzubilden.

2.2. Mathematische Modellierung

Schon seit einigen Jahren stellen die Einbindung von Realitätsbezügen und Modellieren im Mathematikunterricht wichtige Punkte in der didaktischen Diskussion dar. Denn die

Mathematik spielt in der heutigen Zeit eine bedeutsame Rolle in vielen Wissenschaften und in der Wirtschaft. Doch um die Zusammenhänge zwischen der Welt und der Mathematik zu erfahren, müssen in den Schulen zunächst die nötigen Voraussetzungen geschaffen werden (vgl. Maaß, 2011, S. 7). So haben in den letzten Jahren Realitätsbezüge und die Förderung von mathematischen Modellierungskompetenzen verstärkt einen Einzug in deutsche Curricula und Bildungsstandards für den Mathematikunterricht aller Schulstufen gefunden (vgl. Borromeo Ferri et al., 2013, S. V). Seit 2003 hat die Kultusministerkonferenz (KMK) in den Bildungsstandards für die Sekundarstufe I die Kompetenz des Modellierens als eine von sechs allgemeinen Kompetenzen, welche die SuS im Laufe ihrer Schulzeit erwerben sollen, integriert (vgl. KMK Mathematik Sek. I, 2004, S. 8). In der aktuellsten Version der Bildungsstandards für die Sekundarstufe II wird dieser Kompetenzbereich ebenfalls aufgelistet (vgl. KMK Mathematik Sek. II, 2015, S. 15). Durch diese Beschlüsse ist auch in den aktuellen Kernlehrplänen für die Sekundarstufe I und II das Modellieren als eine prozessbezogene Kompetenz verankert.

Im Folgenden wird auf den Begriff des mathematischen Modells eingegangen sowie die didaktische Einbindung im Unterricht in Form von Modellierungskreisläufen diskutiert.

2.2.1. Das mathematische Modell

In der Regel sind realistische Problemstellungen zu komplex, um darauf direkt mathematische Methoden anwenden zu können. Es ist oft sinnvoller, Vereinfachungen zu treffen und ein passendes Modell zu entwickeln, mit dem mathematisch gearbeitet werden kann (vgl. Maaß, 2011, S. 13). Ein mathematisches Modell ist demnach eine vereinfachte Darstellung der Realität, welches nur gewisse, für die jeweilige Fragestellung relevante Teilaspekte berücksichtigt (vgl. Borromeo Ferri et al., 2013, S. 12). Bei der Modellbildung findet demnach immer ein Prozess der Reduktion von Komplexität statt, sodass ein Modell nie alle Merkmale des originalen Vorgangs abbildet. Ein gutes Modell sollte jedoch nur so einfach wie nötig sein und die Aspekte, die modelliert werden müssen, sollten bestmöglich der Realität entsprechen (vgl. Bruder et al., 2015, S. 32). Darüber hinaus sind Modelle nie eindeutig. Je nachdem welche Annahmen und Vereinfachungen getroffen werden, kann eine Situation durch verschiedene Modelle beschrieben werden. Jedes Modell sollte dabei widerspruchsfrei, in sich stimmig und zweckmäßig sein (vgl. Borromeo Ferri et al., 2013, S. 13). Das Ziel eines Modells ist es, zu einer überprüfbaren Aussage für die Realität zu kommen (vgl. Bruder et al., 2015, S. 34).

Modelle werden zu bestimmten Zwecken gebildet. Es wird zwischen den folgenden zwei Ausprägungen von Modellen unterschieden: Auf der einen Seite gibt es normative Modelle, welche vorschreiben, wie in einer bestimmten Situation in der Realität zu verfahren ist. Als Beispiel können hier die Spielregeln für ein Fußballturnier genannt werden. Auf der anderen Seite gibt es deskriptive Modelle, welche die Realität abzubilden versuchen. Hierunter fallen Modelle, die beschreiben und erklären. Unter günstigen Umständen können aus diesen Modellen auch solche entstehen, die Entwicklungen vorhersagen. Als Beispiel für solche Modelle können Klimamodelle genannt werden,

welche die aktuelle Situation darstellen und die zukünftige Entwicklung vorhersagen (vgl. Bruder et al., 2015, S. 34).

2.2.2. Modellierungskreisläufe

Modellieren ist ein komplexer Prozess, weshalb dieser häufig idealisiert als Kreislauf dargestellt wird. Einerseits kann durch diese Darstellung der Begriff des Modellierens an sich veranschaulicht werden. Andererseits stellt ein solcher Kreislauf den SuS eine Hilfe für die Bearbeitung von Modellierungsaufgaben dar (vgl. Borromeo Ferri et al., 2013, S. 14). Modellierungskreisläufe existieren in verschiedenen Versionen mit unterschiedlich vielen Schritten.

Eine ausführliche Version liefert der sieben-schrittige Modellierungskreislauf von Blum und Leiss (siehe Abbildung 1), welcher sich in der deutschsprachigen Mathematikdidaktik bewährt hat und hier kurz vorgestellt wird. Er beschreibt detailliert die einzelnen Schritte, die die Lernenden beim Lösen von Modellierungsaufgaben durchlaufen, und differenziert vor allem diejenigen Prozesse, welche für SuS potenzielle kognitive Hürden darstellen können (vgl. Hinrichs, 2008, S. 19). Dieser Kreislauf dient zudem als Grundlage für den vom Schülerlabor CAMMP verwendeten Modellierungskreislauf (siehe Abbildung 2).

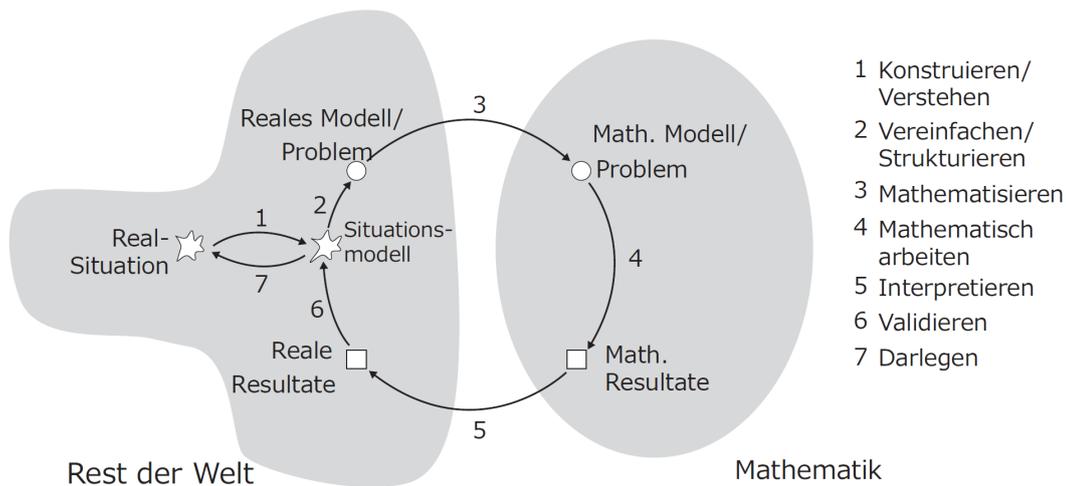


Abbildung 1: Sieben-schrittiger Modellierungskreislauf nach Blum und Leiss (vgl. Blum, 2006, S. 9)

Wie viele Modellierungskreisläufe hat auch dieser Kreislauf als Ausgangspunkt die reale Situation. Die SuS müssen zunächst einmal *verstehen*, was die Fragestellung des Problems aussagt. Anschließend müssen nötige Informationen bezüglich des Problems gesammelt werden, um sich eine Vorstellung von der Situation der Aufgabe zu verschaffen und ein Situationsmodell zu erhalten.

Ziel ist es nun, das Situationsmodell zu mathematisieren. Da das Modell in der Regel noch zu komplex ist, um es direkt mathematisch zu beschreiben, muss zunächst eine

Vereinfachung stattfinden, indem bestimmte Eigenschaften weggelassen werden (vgl. Barzel et al., 2011, S. 157). Anhand von konkreten Leitfragen für die Modellierung werden hier wichtige von unwichtigen Informationen getrennt. So muss z.B. die Frage geklärt werden, welche Mathematik in dem Zusammenhang bereits bekannt ist oder wie viel Zeit für die Lösung investiert werden kann (vgl. Hinrichs, 2008, S. 21). Zusätzlich werden in diesem Schritt auch idealisierte Annahmen getroffen, indem bestimmte Eigenschaften angenommen und hinzugefügt werden (vgl. Barzel et al., 2011, S. 157). Bei der Vereinfachung besteht jedoch immer die Gefahr einer zu starken Reduktion der Annahmen, sodass die reale Situation zu stark verfälscht abgebildet wird. Demnach ist die Konstruktion eines realen Modells immer mit einem Kompromiss zwischen der Einfachheit des Modells und der Adäquatheit für die Problemstellung verbunden (vgl. Hinrichs, 2008, S. 21).

Als nächster Schritt folgt die *Mathematisierung* der vereinfachten Situation, indem das Modell in ein mathematisches Problem überführt wird. Dabei findet eine Übersetzung aller Annahmen und Vereinfachungen, die im vorherigen Schritt getroffen wurden, von der Alltagssprache in die Sprache der Mathematik statt (vgl. Maaß, 2011, S. 14). Auf dieses mathematische Modell können die SuS ihre bekannten mathematischen Methoden anwenden und weiterentwickeln.

Der nächste Schritt behandelt das *mathematische Arbeiten*, in dem eine Lösung des Problems im mathematischen Modell gesucht wird. Durch die Anwendung heuristischer Strategien und mathematischer Algorithmen gelangen die SuS zu einem mathematischen Resultat (vgl. Maaß, 2011, S. 14). Da die Mathematik in der Regel zu komplex ist, um eine Lösung mit Stift und Papier zu erhalten, wird zum Lösen meistens auf den Computer als Hilfsmittel zurückgegriffen. So kommt auch bei CAMMP die Computersoftware MATLAB zum Einsatz.

Es folgt eine wichtige Phase, in der wieder der Transfer von der Mathematik auf die reale Situation stattfindet. Hier erfolgt im Hinblick auf die ursprüngliche Fragestellung die *Interpretation* der Ergebnisse, um ein reales Resultat zu erhalten (vgl. Maaß, 2011, S. 14).

Mit der Interpretation ist die Modellierung jedoch noch nicht abgeschlossen. Der gesamte Modellierungsprozess muss kritisch reflektiert und das Resultat auf unterschiedliche Weise *validiert* werden, um zu überprüfen, ob die Lösung alle vorhandenen Bedingungen erfüllt. Hier wird der Kreislaufcharakter des Modellierungsprozesses deutlich. Sind die Abweichungen im Rahmen der Validierung zu groß, müssen einzelne Schritte oder sogar der gesamte Prozess erneut durchgeführt werden. Beispielsweise müssen die getroffenen Annahmen verändert oder andere mathematische Strategien verwendet werden (vgl. Hinrichs, 2008, S. 27).

In der letzten Phase, der *Darlegung*, sollen die Ergebnisse nachvollziehbar dargestellt, die ursprüngliche Problemstellung beantwortet und das Vorgehen erläutert werden. So spricht dieser Schritt vornehmlich kommunikative und argumentative Kompetenzen an (vgl. Hinrichs, 2008, S. 28).

Der Modellierungskreislauf von Blum und Leiss ist unter kognitiven Gesichtspunkten erstellt worden und verfolgt das Ziel, bei den SuS ein Verständnis ihrer Modellierarbeit

auf der Metaebene zu schaffen (vgl. Frank et al., 2018, S. 140). Für die SuS ist dieser Kreislauf häufig schon zu komplex, denn sie haben Schwierigkeiten, die sieben Schritte des Modellierungskreislaufs zu unterscheiden (vgl. Hinrichs, 2008, S. 19). Aus diesem Grund verwendet das Schülerlabor CAMMP in Anlehnung an Blum und Leiss ein didaktisch reduziertes Schemata, welches den SuS eine einfachere Orientierungshilfe bei der Bearbeitung von Modellierungsaufgaben gibt.

In dem Modellierungskreislauf von CAMMP (siehe Abbildung 2) wird auch zwischen Realität und Mathematik unterschieden. Ausgangspunkt ist auch hier das **reale Problem**, welches die SuS zunächst, wie von Blum und Leiss beschrieben, verstehen müssen. Durch das Treffen von *Vereinfachungen* und *Annahmen* findet eine Überführung in ein **vereinfachtes Problem** statt. Der Schritt „vereinfachen/ strukturieren“ aus dem Modell von Blum und Leiss ist hier zu finden. Für dieses vereinfachte Problem wird durch *mathematisches Beschreiben* ein **mathematisches Modell** aufgestellt. Diesen Schritt nennen Blum und Leiss „mathematisieren“. Auf das mathematische Modell können mathematische Methoden angewendet werden, um eine **mathematische Lösung** zu *berechnen*. Bei Blum und Leiss ist dieser Schritt unter „mathematisch arbeiten“ zu finden. Wie auch beim Modellierungskreislauf von Blum und Leiss findet am Ende ein Rückschluss zu der realen Ausgangssituation statt, indem die Lösung bezogen auf das reale Problem *interpretiert* und *validiert* wird. Die Schritte „interpretieren“ und „validieren“ aus dem Modellierungskreislauf von Blum und Leiss wurden hier zusammengefasst. Der Schritt „darlegen“ wird im Modellierungskreislauf von CAMMP nicht explizit erwähnt. Diese Kompetenz wird jedoch im Rahmen CAMMP an vielen Stellen abgefragt.

Auch im entwickelten Lernmodul kommt dieser Kreislauf zum Einsatz. Einerseits durchlaufen die SuS während der Bearbeitung des Lernmoduls den gesamten Modellierungskreislauf selber und können dadurch die einzelnen Schritte eigenständig nachvollziehen. Andererseits wird er den SuS im Laufe des Tages mehrmals präsentiert, um ihnen ein Bewusstsein für ihre aktuelle Position im Modellierungskreislauf zu vermitteln.

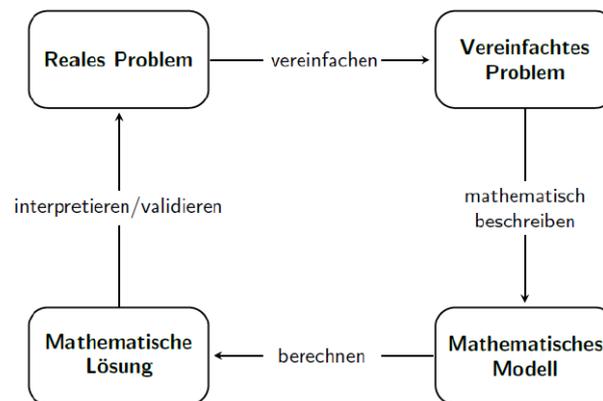


Abbildung 2: Eingesetzter Modellierungskreislauf im Schülerlabor CAMMP

Durch das mehrmalige Durchlaufen des Modellierungskreislaufes wird das entwickel-

te Modell mit jedem Durchgang erweitert. Mit jedem Durchlauf vergrößert sich das Verständnis über die zu untersuchende Fragestellung und das Wissen über das Problem wird vielschichtiger. Um diese wachsende Komponente zu verdeutlichen, stellen Büchter und Leuders die Modellierung in einer nach außen ausbreitenden Spirale dar (vgl. Büchter & Leuders, 2005, S. 76). Diese Idee griff auch das Schülerlabor CAMMP mit der von der Initiative Computer-Base Math eingeführten Solution Helix of Math auf (siehe Abbildung 3). Im Gegensatz zu Büchter und Leuders wird diese Spirale immer enger. Dies soll den Aspekt der Annäherung an eine optimale Lösung durch Wiederholung der Modellierungsschritte verdeutlichen (vgl. Frank et al., 2018, S. 139). Darüber hinaus beinhaltet die von CAMMP verwendete Spirale den Aspekt computergestütztes, mathematisches Arbeiten, da die SuS in den verschiedenen Modulen ein digitales Werkzeug, die Computersoftware MATLAB, als Unterstützung nutzen. Denn in der heutigen Zeit sind die meisten Anwendungen von Mathematik nicht mehr ohne moderne Computertechnologie möglich, sodass der Computereinsatz bei realitätsbezogenen Problemen eine sinnvolle Unterstützung bietet (vgl. Greefrath & Weitendorf, 2013, S. 181).

Durch eine Verminderung des Rechenaufwands kann der Fokus der SuS auf den Modellierungsprozess an sich sowie die problemorientierte Bearbeitung der Aufgabe gelenkt werden. Zudem wäre der Aufwand ohne Computereinsatz bei einem Großteil der Aufgaben, welche die SuS im Rahmen eines Lernmodul bearbeiten, zu zeitaufwändig, um in der vorgesehenen Zeit bewältigt zu werden. Insbesondere kann durch den Computereinsatz ein komplexes Problem mit Hilfe gängiger Schulmathematik bearbeitet werden, indem die mathematischen Inhalte, welche über den Schulstoff hinaus gehen, vom Computer übernommen werden. Digitale Werkzeuge können zudem die Aufgabe des Visualisierens übernehmen, indem beispielsweise gegebene Daten mit Hilfe einer Computeralgebraanwendung in ein Koordinatensystem dargestellt werden (vgl. Greefrath & Weitendorf, 2013, S. 182-183).

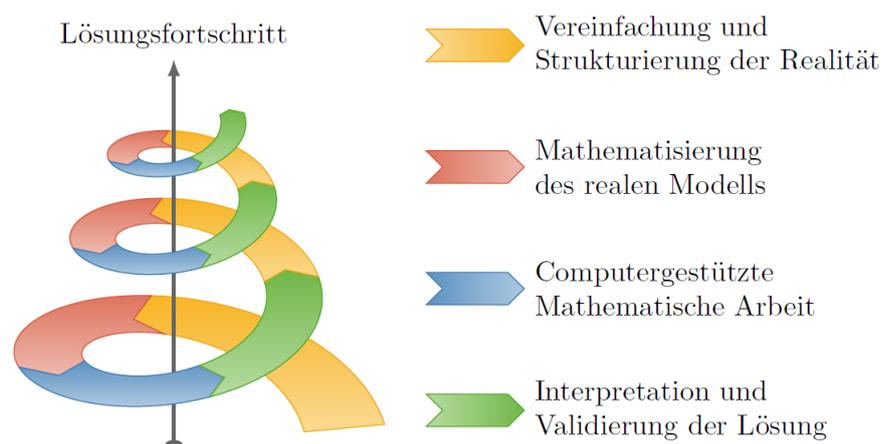


Abbildung 3: Verwendete Modellierungsspirale im Schülerlabor CAMMP (vgl. Frank et al., 2018, S. 140)

2.2.3. Der PPDAC Zyklus

In der heutigen Zeit basiert die Entscheidungsfindung in der Industrie, der Medizin, der Wirtschaft und in anderen Bereichen der Gesellschaft zunehmend auf dem Erheben sowie Analysieren und Interpretieren von Daten, weshalb Statistiken einen wesentlichen Bestandteil des heutigen Informationszeitalters darstellen (vgl. Frischemeier, 2017, S. 9). Aber auch durch die Medien findet eine tägliche Konfrontation mit Daten im Alltag statt, deren Interpretation das Leben maßgeblich beeinflussen können, wie beispielsweise eine politische Entscheidung, die auf Grund der vorliegenden Daten getroffen werden. Als ein aktuelles Beispiel sind hier die Klimamodelle zum Klimawandel zu erwähnen, auf deren Grundlage politische Entscheidungen und Ziele festgelegt werden. Die Orientierung in einer solchen Welt der Daten setzt eine gewisse Datenkompetenz der SuS voraus. (vgl. Frischemeier, 2017, S. 9). Für die Vermittlung einer solchen Kompetenz haben Wild und Pfannkuch den PPDAC-Zyklus (Problem - Plan - Daten - Analyse - (C)Konklusionen) entwickelt, welcher den SuS auf einer Metaebene Anhaltspunkte für das Durchlaufen einer Datenanalyse gibt. Auch bei der Konzeption des entwickelten Lernmoduls dieser Masterarbeit wurden einzelnen Schritte dieses Zyklus berücksichtigt.

Der PPDAC-Zyklus (siehe Abbildung 4) soll den SuS bei statistischen Untersuchungen als Werkzeug zum Lösen von realen Problemen dienen (vgl. Wild & Pfannkuch, 1999, S. 225). Dabei zeigt der Zyklus eine strukturierte Darstellung des Ablaufs statistischer Untersuchungen und gibt Lernenden eine Orientierung bei der Durchführung statistischer Projekte. Aber auch Lehrende kann er als Planungsinstrument solcher Projekte unterstützen.

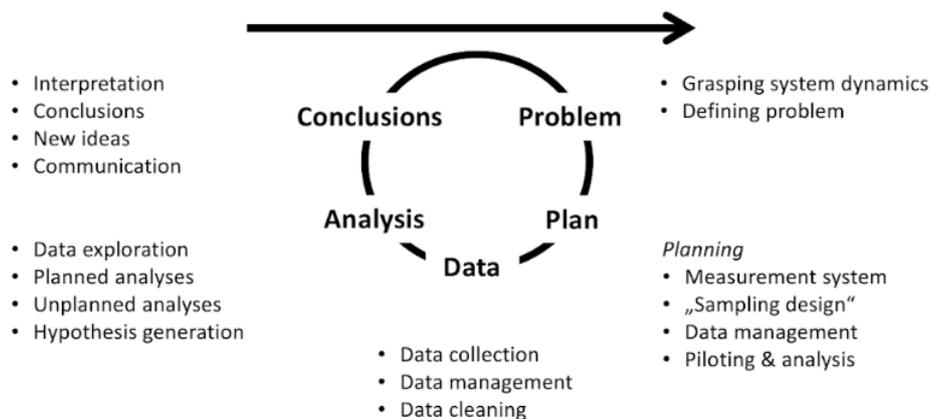


Abbildung 4: PPDAC-Zyklus nach Wild und Pfannkuch (vgl. Wild & Pfannkuch, 1999, S. 225)

Am Anfang des Zyklus steht die Phase des *Problems*, bei der die genaue Definition des Problems im Vordergrund steht. Es wird nach einer Fragestellung gesucht, die im Laufe der Untersuchung beantwortet werden soll. Hier sollen bereits verschiedene Einflussfaktoren bedacht und abgewogen werden. Eine mögliche Problem- bzw. Fragestellung

für eine statistische Erhebung könnte beispielsweise das Freizeitverhalten von Studierenden an einer Universität sein (vgl. Frischemeier, 2017, S. 211). Die nächste Phase besteht in der *Planung* der Datenerhebung, bei der das Konstruieren der Messinstrumente sowie das Aufstellen eines Stichprobenplans im Mittelpunkt steht. Mit Bezug zum genannten Beispiel, die Freizeitaktivität der Studierenden zu ermitteln, kann hier für die Befragung der Studierenden ein Fragebogen erstellt werden. Unter anderem muss entschieden werden, ob der Fragetyp offen oder geschlossen sein soll und mögliche Fehlerquellen bei der Itemkonstruktion diskutiert werden (vgl. Frischemeier, 2017, S. 211). Ebenso muss der zeitliche Aspekt berücksichtigt werden und sich gegeben falls nur auf eine konkrete Zielgruppe unter den Studierenden geeinigt werden. Auch kann hier schon eine erste Pilotierung erfolgen, indem beispielsweise der entwickelte Fragebogen an einzelne Studierende getestet wird. Nach der Planung erfolgt die *Erhebung der Daten*, in der die Daten gesammelt, verwaltet und bereinigt werden. In der darauffolgenden *Analyse* werden Hypothesen generiert und sowohl geplante als auch ungeplante Analysen durchgeführt. Eine mögliche Hypothese, welche im Rahmen des aufgestellten Beispiels nachgegangen werden kann, ist beispielsweise die Fragestellung, ob Studierende, die viel Fernsehen schauen, weniger lesen (vgl. Frischemeier, 2017, S. 224). Der Abschluss des Zyklus bildet das *Fazit*. Hier werden die Ergebnisse interpretiert sowie kommuniziert, Schlussfolgerungen gezogen und über mögliche neue Ideen nachgedacht.

3. Fachlicher Hintergrund

In diesem Kapitel wird die Problemstellung für das entwickelte Lernmodul dargestellt und die mathematischen Grundlagen, welche für den CAMMP day benötigt werden, vorgestellt und erläutert. Dabei fokussiert sich die Problemstellung darauf, Informationen von Nutzern eines sozialen Netzwerks vorherzusagen, welche diese nicht in ihren Profilen angegeben haben.

3.1. Problemstellung

Soziale Netzwerke wie Twitter, Facebook oder WhatsApp sind aus unserem heutigen Leben nicht mehr wegzudenken. Laut der aktuellen JIM-Studie (2017) nutzen 89 % der 12- bis 19-Jährigen den Kommunikationsdienst WhatsApp täglich. Instagram kommt auf 44 % und Snapchat auf 39 %. Auch ist zu verzeichnen, dass der prozentuale Anteil an Nutzern bei den drei genannten Beispielen bis zum siebzehnten Lebensjahr jährlich ansteigt (vgl. Feierabend et al., 2017, S. 35-36).

Dies zeigt eindrucksvoll, dass soziale Netzwerke in der heutigen Zeit einen selbstverständlichen Teil des Alltags der Jugendlichen darstellen. Jedoch darf in einer Gesellschaft, welche zunehmend auf Informations- und Kommunikationstechnologien basiert, der Schutz der Privatsphäre nicht vergessen werden. Bereits Garcia, Sergiol und Schweitzer machten 2014 in ihrem Artikel „Online Privacy as a Colective Phenomenon“ auf die bereits bestehende Möglichkeit aufmerksam, Informationen von Nutzern zu sammeln, welche diese nicht mit der Öffentlichkeit teilen wollen (vgl. Garcia et al., 2014, S. 2). Dabei wird in dem Artikel zwischen der Erstellung eines *Partial Shadow Profiles* und eines *Full Shadow Profiles* unterschieden. Garcia, Sargiol und Schweitzer definieren *Partial Shadow Profiles* wie folgt (vgl. Garcia et al., 2014, S. 4):

Definition *Partial Shadow Profiles*:

Für jeden Nutzer eines sozialen Netzwerks wird ein Profil angelegt, welches alle Informationen beinhaltet, die freiwillig angegeben wurden. Werden über einen Nutzer zusätzliche, persönliche Informationen erzeugt und gespeichert, welche er nicht selber angeben hat und teilen wollte, so wird von einem Partial Shadow Profile gesprochen.

Informationen, welche von Nutzern sozialer Netzwerke nicht preisgegeben wurden, können bereits ohne technischen Aufwand über die soziale Online-Interaktion oder durch Informationsangaben von Freunden, erstellt werden. Bereits im Rahmen der Masterarbeit „Wie sicher ist meine Privatsphäre in Online Netzwerken? ...und was hat das mit Mathe zu tun“ wurde gezeigt, dass mit Hilfe einfacher Heuristiken und den Informationen der Freunde, die nicht angegebene Sexualität des Nutzers mit einer Genauigkeit von etwa 90 % bestimmt werden kann (siehe Sube, 2016).

Während sich *Partial Shadow Profiles* auf Nutzer beziehen, welche bereits Mitglied eines sozialen Netzwerks sind, liegt der Fokus bei *Full Shadow Profiles* auf Personen, die noch mit keinem sozialen Netzwerk in Berührung gekommen sind (vgl. Garcia et al., 2014, S. 6):

Definition *Full Shadow Profiles*:

Full Shadow Profiles sind erstellte Profile von Personen, die kein Mitglied eines sozialen Netzwerks sind.

Solche *Full Shadow Profiles* können zum Beispiel dann erstellt werden, wenn ein Benutzer eines Netzwerkes seine Kontaktliste mit einem sozialen Netzwerk teilt (vgl. Garcia et al., 2014, S. 6). Anhand der Telefonnummern und ggf. Email-Adressen können die Betreiber Personen herausfinden, denen kein Konto zugeordnet werden kann. Wenn diese „Nicht-Nutzer“ in vielen Kontaktlisten von sozialen Netzwerk-Benutzern erscheinen, können aus deren Angaben ebenfalls persönliche Informationen ermittelt werden (vgl. Garcia et al., 2014, S. 6). So steht beispielsweise in der WhatsApp Datenschutzrichtlinie geschrieben: „Im Einklang mit geltenden Gesetzen stellst du uns regelmäßig die Telefonnummern in deinem Mobiltelefon-Adressbuch zur Verfügung, darunter sowohl die Nummern von Nutzern unserer Dienste als auch die von deinen sonstigen Kontakten. Möglicherweise stellst du uns auch eine E-Mail-Adresse zur Verfügung.“ (WhatsApp Inc., 2018).

In beiden Fällen geht es also um die Vorhersage von Informationen, welche Personen nicht angegeben haben. Darauf aufbauend verfolgt der entwickelte CAMMP day anhand realistischer Nutzer-Daten des sozialen Netzwerks Friendster folgende zentrale Fragestellung:

Wie gut kann man das Alter eines Nutzers in einem sozialen Netzwerk vorhersagen, obwohl dieses aus den Angaben in seinem Profil nicht ablesbar ist?

3.2. Das soziale Netzwerk Friendster

Im CAMMP day arbeiten die SuS mit realen Nutzer-Profilen aus dem sozialen Netzwerk Friendster. Zunächst werden ein paar Informationen über dieses Netzwerk vorgestellt. Anschließend erfolgt eine Darstellung der Daten, mit denen die SuS arbeiten.

3.2.1. Das Unternehmen Friendster Inc.

Das Unternehmen Friendster Inc. gründet 2002 das vor allem im englischsprachigen und asiatischen Raum beliebte soziale Netzwerk Friendster. Es besaß im Jahr 2004 knapp 350 Millionen Nutzer und galt damit als größtes soziales Netzwerk im Internet (vgl. Klanert, 2014). Doch durch die aufkommenden Konkurrenten MySpace und vor allem Facebook stoppte das Wachstum von Friendster, sodass das soziale Netzwerk bereits ab dem Jahr 2006 eine immer geringere Beliebtheit erfuhr (vgl. McMillan, 2013). Aktuell ist die Funktionalität von Friendster eingestellt. Auf ihrer offiziellen Internetseite ist von einer Pause bis zum Jahr 2015 die Rede (vgl. Friendster Inc, o.J.), welche anscheinend bis heute anhält. Jedoch wurden alle öffentlich zugänglichen

Informationen gespeichert. Diese Informationen können frei verfügbar in einem Internetarchiv² heruntergeladen werden. Sie zeigen eine Momentaufnahme des damaligen sozialen Netzwerks.

3.2.2. Das soziale Netzwerk als ungerichteter Graph

In dem entwickelten Lernmodul wird das soziale Netzwerk, wie in Abbildung 5 beispielhaft angedeutet, als ein ungerichteter Graph dargestellt. Dazu werden zunächst kurz die Grundbegriffe der Graphentheorie vorgestellt:

Definition Graph, Knoten, Kante (vgl. Hartmann, 2015, S. 264):
 Ein (ungerichteter) **Graph** G besteht aus einer Menge $V = V(G)$, den **Knoten**, und aus einer Menge $E = E(G)$ von ungeordneten Paaren $k = \{v_i, v_j\}$ mit $v_i, v_j \in V$, den **Kanten** von G . Man schreibt $G = (V, E)$.

Dabei werden $v_i \in V$ und $v_j \in V$ einer Kante $k = \{v_i, v_j\}$ als *Endknoten* bezeichnet, für die $v_i \neq v_j$ gilt (vgl. Hartmann, 2015, S. 264). Für den Fall $v_i = v_j$ wird von einer *Schlinge* gesprochen (vgl. Hartmann, 2015, S. 264). Nebenbei ist von einem *gerichteten* Graphen die Rede, wenn die Kanten einen Richtungssinn aufweisen und der Graph als Kanten geordnete Paare $k = [v_i, v_j]$ besitzt (vgl. Hartmann, 2015, S. 287). In diesem Fall werden die Kanten als Pfeile dargestellt und v_i als *Anfangsknoten* und v_j als *Endknoten* bezeichnet (vgl. Tittmann, 2014, S. 133).

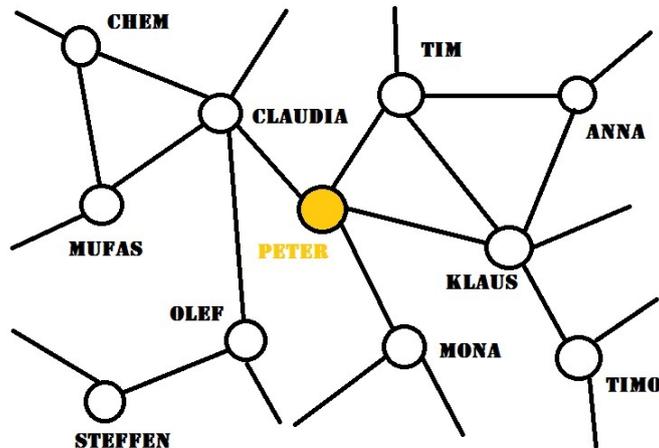


Abbildung 5: Beispielhafte Veranschaulichung eines sozialen Netzwerks als ungerichteter Graph. Die Nutzer stellen die Knoten dar und die Freundschaftsbeziehungen zwischen den einzelnen Nutzern stellen die Kanten des Graphen dar. (vgl. Sube, 2016, S. 18).

Damit stellen in dem ungerichteten Graphen des sozialen Netzwerks in Abbildung 5

²Das Archive kann unter der Internetadresse <https://archive.org/details/archive-team-friendster> aufgerufen werden.

die Nutzer die Knoten dar und die Freundschaftsbeziehungen zwischen den Nutzern werden durch die Kanten symbolisiert. Freundschaften mit sich selbst ergeben hier keinen Sinn, weshalb in dem Graph keine Schlingen auftreten. Darüber hinaus ist eine Freundschaft kommutativ. Das heißt, wenn Nutzer A mit Nutzer B befreundet ist, dann ist auch Nutzer B mit Nutzer A befreundet. Die Kanten bestehen demnach, wie in der obigen Definition erwähnt, aus ungeordneten Paaren und weisen keine Richtung auf. Es liegt also ein ungerichteter Graph vor. Auch sind im Fall eines sozialen Netzwerks die Mengen V und E endlich, da die Anzahl der Nutzer und der Freundschaftsbeziehungen endlich sind.

Es folgt eine weitere Definition, um die Freundschaft zwischen zwei Nutzern mit einem Fachbegriff ausdrücken zu können:

Definition adjacente Knoten (vgl. Tittmann, 2014, S. 133):
Zwei Knoten $x, y \in V$, die durch eine Kante $k = \{x, y\}$ verbunden sind, heißen adjazent (benachbart) in G .

Befreundete Nutzer können demnach als adjazente Nutzer bezeichnet werden. Im Folgenden wird auf den Abstand zweier Knoten zueinander eingegangen, um eine Definition für den Grad der Freundschaft aufzustellen:

Definition Kantenfolge (vgl. Tittmann, 2014, S. 133):
Eine Kantenfolge in einem Graphen $G = (V, E)$ ist eine alternierende Folge von Knoten v_i und Kanten k_i

$$u = v_0, k_1, v_1, k_2, v_2, k_3, \dots, k_{k-1}, v_{k-1}, k_k, v_k = v$$

für die stets $k_i = \{v_{i-1}, v_i\}$, $1 \leq i \leq k$ gilt.

In Abbildung 5 sieht eine beispielhafte Kantenfolge zwischen Peter und Olaf wie folgt aus: *Peter, Freundschaftsbeziehung zwischen Peter und Claudia, Claudia, Freundschaftsbeziehung zwischen Claudia und Olaf, Olaf*.

Aus der Definition wird ersichtlich, dass innerhalb einer Kantenfolge Knoten mehrmals auftreten können. Für die Beschreibung des sozialen Netzwerks als ungerichteter Graph wird aus diesem Grund der *Weg*, als ein Spezialfall der Kantenfolge, betrachtet:

Definition Weg (vgl. Hartmann, 2015, S. 270):
Sei $G = (V, E)$ ein Graph. Ein Weg von u nach v mit $u, v \in V$ ist eine Kantenfolge von u nach v , in der alle Knoten verschieden sind.

Die oben beschriebene Kantenfolge zwischen Peter und Olaf stellt demnach auch einen Weg zwischen diesen beiden Nutzern dar. In der folgenden Definition wird die Länge einer Kantenfolge und damit auch eines Weges festgehalten:

Definition Länge einer Kantenfolge (vgl. Hartmann, 2015, S. 271):
 Die Anzahl der Kanten einer Kantenfolge oder eines Weges heißt Länge der Kantenfolge beziehungsweise des Weges.

Die Länge des oben beschriebenen Weges zwischen Peter und Olaf beträgt demnach zwei. Mit diesem Wissen kann nun abschließend der Abstand zwischen zwei Knoten definiert werden:

Definition Abstand zweier Knoten (vgl. Tittmann, 2014, S. 133):
 Der Abstand $d(u, v)$ zweier Knoten $u \in V$ und $v \in V$ eines Graphen $G = (V, E)$ ist die Länge eines kürzesten, u und v verbindenden Weges. Existiert kein Weg von u nach v , so wird $d(u, v) = \infty$ gesetzt.

Der kürzeste Weg muss dabei nicht eindeutig bestimmt sein. Abbildung 6 zeigt beispielhaft, dass der Abstand zwischen Peter und Anna zwei ist und auf zwei verschiedene Wege erreicht werden kann. Zum einen gibt es den Weg *Peter, Freundschaftsbeziehung zwischen Peter und Tim, Tim, Freundschaftsbeziehung zwischen Tim und Anna, Anna*, und zum anderen gibt es den Weg *Peter, Freundschaftsbeziehung zwischen Peter und Klaus, Klaus, Freundschaftsbeziehung zwischen Klaus und Anna, Anna*. Mit diesem Hintergrund können die Begriffe *Freunde 1. Ordnung* und *Freunde 2. Ordnung* definiert werden:

Definition Freunde 1. oder 2. Ordnung (vgl. Sube, 2016, S. 24):
 Ausgehend von einem Nutzer werden alle Personen als **Freunde 1. Ordnung** bezeichnet, deren Knoten den Abstand 1 zum Knoten des betrachteten Nutzers aufweisen. Dementsprechend werden alle Personen als **Freunde 2. Ordnung** des Nutzers bezeichnet, deren Knoten den Abstand 2 zum Knoten des betrachteten Nutzers haben.

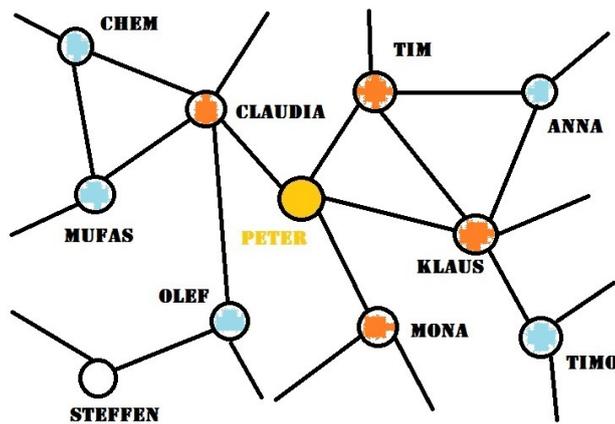


Abbildung 6: Beispielhafte Veranschaulichung der Freundschaftsbeziehungen in einem sozialen Netzwerk. In orange sind die Freunde 1. Ordnung und in blau die Freunde 2. Ordnung von Peter dargestellt (vgl. Sube, 2016, S. 24).

Personen, die in direkter Verbindung zu einem Nutzer stehen, werden also als Freunde 1. Ordnung bezeichnet, da sie einen Abstand der Ordnung 1 vom Nutzer aufweisen. Die Freunde 1. Ordnung von Peter sind in Abbildung 6 in orange eingezeichnet. In blau sind die Freunde 2. Ordnung von Peter dargestellt. Freunde 2. Ordnung sind Personen, die in direkter Beziehung zu den Freunden 1. Ordnung eines Nutzers stehen und vom Nutzer den Abstand der Ordnung 2 haben.

In dem entwickelten CAMMP day für die Mittelstufe kommen die SuS mit der hier beschriebenen Graphentheorie nicht in Berührung. Sie werden im Lernmodul die anschaulichen Begriffe *Freunde* und *Freundes-Freunde* für die hier definierte Nomenklatur Freunde 1. Ordnung und Freunde 2. Ordnung benutzen. Im weiteren Verlauf der Masterarbeit werden diese Begriffe für die Beschreibung des entwickelten Materials synonym verwendet.

3.2.3. Informationen zum verwendeten Datensatz

Da das entwickelte Lernmodul auf dem CAMMP day „Wie sicher ist meine Privatsphäre in Online Netzwerken? ...und was hat das mit Mathe zu tun“ für die Sekundarstufe II (Sube, 2016) aufbaut und eine Weiterentwicklung für die Sekundarstufe I darstellt, wird der dort eingesetzte Datensatz als Ausgangsbasis verwendet. Für diesen Datensatz wurden die ersten 1 Millionen Profile analysiert, unter denen sich vor allem Nutzer aus den USA befinden (vgl. Garcia et al., 2014, S. 3). Der Informationsgehalt der einzelnen Profile ist unterschiedlich und richtet sich nach den öffentlich gemachten Angaben der Nutzer. Folgende Informationen können in einem Profil enthalten sein (Sube, 2016, S. 20):

- Nutzer ID (gibt die Mitgliedsnummer an, Beitrittsreihenfolge ersichtlich)
- Name als Wort
- Geburtsdatum
- Geschlecht (männlich, weiblich oder unspezifisch)
- Interessen (Freunde, Fans, Partner für Aktivitäten, Nur herumschauen, Frauen treffen, Männer treffen, Beziehung mit Frauen, Beziehung mit Männern, Beziehung mit Frauen und Männern, Frauen und Männer treffen)³
- Beziehungsstatus (allein stehend, verheiratet, in einer Beziehung, in Gemeinschaft lebend, es ist kompliziert)

Da im Rahmen des Lernmoduls für die Sekundarstufe II Vorhersagen über die sexuelle Orientierung getätigt werden, sind nur die Profile interessant, welche diese Information öffentlich angegeben haben. Aus diesem Grund wurden in einer ersten Filterung alle Nutzer herausgenommen, bei denen die sexuelle Orientierung nicht bestimmt werden

³Hier können mehrere Aspekte angegeben werden.

konnte. Darüber hinaus werden im Lernmodul für die Sekundarstufe II die Vorhersagen auf Grundlage der Angaben der Freunde eines Nutzers getroffen. Im Rahmen einer zweiten Filterung wurden deshalb nur die Nutzer beibehalten, welche nach der ersten Filterung mindestens eine Freundschaft zu einem Nutzer aufweisen konnten (vgl. Sube, 2016, S. 20).

Konkret blieb nach den Filterungen ein Datensatz bestehend aus 56284 Nutzern übrig. Dieser Datensatz bildet die Grundlage für das entwickelte Lernmodul für die Sekundarstufe I, muss jedoch für die aktuelle Fragestellung, das Alter eines Nutzers vorherzusagen, weiter aufbereitet werden.

3.2.4. Aufbereitung der Daten

Tabelle 1 gibt eine Übersicht über den zu Grunde liegenden, gefilterten Datensatzes aus 56284 aus dem bestehenden Lernmodul für die Sekundarstufe II. Anzumerken ist hier bereits die geringe durchschnittliche Anzahl der Freunde 1. Ordnung von 3.26, welche die reale Situation in heutigen sozialen Netzwerken nicht widerspiegelt. Eine Tatsache, auf die im Ausblick, am Ende dieser Arbeit nochmal genauer eingegangen wird (siehe Kapitel 6).

Tabelle 1: Übersicht des gefilterten Datensatzes aus dem Lernmodul für die Sek. II (Sube, 2016)

Gesamtanzahl der Nutzer:	56284
Durchschnittliche Anzahl Freunde 1. Ordnung:	3.26
Durchschnittliche Anzahl Freunde 2. Ordnung:	19.16
Minimales Alter im Datensatz:	0
Maximales Alter im Datensatz:	91

Sowohl Tabelle 1 als auch eine Betrachtung der Altersverteilung in Abbildung 7 lassen die Vermutung aufkommen, dass die Angaben der Nutzer im sozialen Netzwerk Friendster nicht alle wahrheitsgetreu getätigt wurden. Aus diesem Grund werden die knapp 1000 Nutzer mit einem Alter von null Jahren aus dem Datensatz herausgenommen. Ebenfalls sind Altersangaben um 90 Jahre anzuzweifeln. Natürlich ist es prinzipiell möglich, dass Personen in diesem Altersbereich Teil eines sozialen Netzwerks sind. Jedoch ist zu bedenken, dass Friendster als eines der ersten sozialen Netzwerke online war und gerade die älteren Personen an der damals gerade aufkommenden Digitalisierung größtenteils nicht teilgenommen haben. So ist es zu dieser Zeit unüblich Personen in diesem Alter in einem sozialen Netzwerk anzutreffen. Besonders auffällig ist zudem der starke Anstieg der Nutzer von 89 auf 90 sowie 91 Jahre. So gibt es fast sechs Mal so viele Nutzer mit einem Alter von 90 Jahre als 89-Jährige. Auch die Tatsache, dass die Anzahl der Personen im vorliegenden Datensatz ab 28 Jahre mit zunehmenden Alter

zurückgeht und plötzlich bei 90 Jahren wieder stark ansteigt, wirkt seltsam. Genau diese Thematik diskutieren die SuS auch innerhalb des Lernmoduls. Dort dürfen sie selber entscheiden, ob sie Nutzer ab einer gewissen oberen Altersgrenze aussortieren. Für den weiteren Verlauf der Arbeit wird ein Datensatz betrachtet, der Nutzer mit einem Alter von 90 Jahren und älter nicht berücksichtigt.

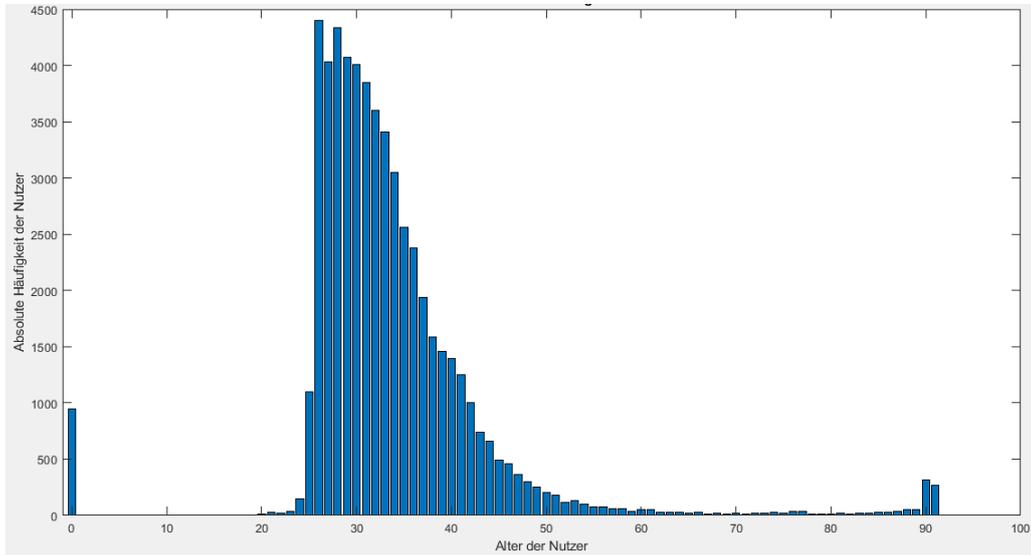


Abbildung 7: Die Altersverteilung im Datensatz aus dem Lernmodul für die Sekundarstufe II (Sube, 2016) zeigt die absoluten Häufigkeiten der Nutzer mit einem Alter zwischen 0 und 100 Jahren. Auffällig sind sowohl Altersangaben von 0 Jahren als auch Angaben um die 90 Jahre.

Nach der ersten Filterung ist erneut eine Betrachtung der Anzahl der Freunde 1. Ordnung nötig. Abbildung 8 (oben) zeigt, dass durch das Aussortieren der Nutzer nun wieder Profile mit null Freunden 1. Ordnung zu finden sind. Für das entwickelte Lernmodul werden ebenfalls die Altersangaben der Freunde 2. Ordnung benötigt. Nutzer mit null Freunden 2. Ordnung haben damit ebenfalls keinen Informationsgehalt und werden, wie die Nutzer mit null Freunden 1. Ordnung, aus dem Datensatz herausgenommen.

Insgesamt bleiben nach den beiden Filterungen 44747 der anfänglichen 56284 Nutzer übrig. Nach der Filterung wird die vereinfachte Annahme getroffen, dass die Altersangaben der verbleibenden Nutzer korrekt sind. Tabelle 2 gibt eine Übersicht über den Datensatz, der im weiteren Verlauf der Arbeit betrachtet wird und auf dem auch die erstellten Musterlösungen basieren. Wie bereits erwähnt, können die später erzeugten Datensätze der SuS im CAMMP day von diesem abweichen, da sie im Rahmen der ersten Filterung selbstständig die Altersgrenze festlegen.

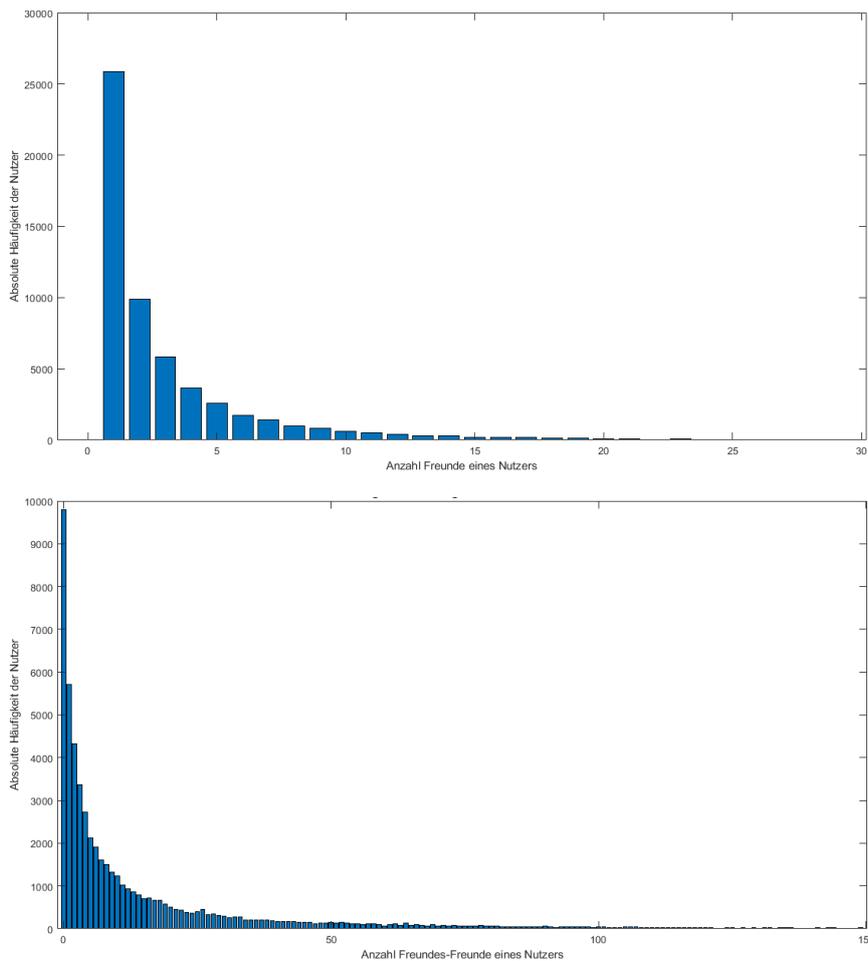


Abbildung 8: *oben*: Absoluten Häufigkeiten der Nutzer mit einer gewissen Anzahl an Freunden nach der ersten Filterung; *unten*: Ausschnitt der absoluten Häufigkeiten der Nutzer mit einer gewissen Anzahl an Freundes-Freunden nach der ersten Filterung.

Tabelle 2: Übersicht des Datensatzes, nachdem die beiden Filterungen durchgeführt wurden.

Gesamtanzahl der Nutzer:	44747
Durchschnittliche Anzahl Freunde 1. Ordnung:	3.68
Durchschnittliche Anzahl Freunde 2. Ordnung:	23.06
Minimales Alter im Datensatz:	18
Maximales Alter im Datensatz:	89

3.2.5. Codierung der Informationen in MATLAB

```
Valid: 1
ID: 10
Age: 42
Gender: 2
RelationshipStatus: 2
Interests: [4 6 2 3]
NumberFriends: 4
NumberFriendsSecondOrder: 9
Friends: [4×1 double]
FriendsSecondOrder: [9×1 double]
Hetero: 1
OriginalID: 220
NumberGayFriends: 1
NumberHeteroFriends: 3
NumberGayFriendsSecondOrder: 2
NumberHeteroFriendsSecondOrder: 9
AgeFriends: [40 38 40 34]
AgeFriendsSecondOrder: [41 37 35 39 32 38 41 27 37]
```

Abbildung 9: Gespeicherte Informationen in der Computersoftware MATLAB am Beispiel des Nutzers mit der ID 10.

Zu jedem Nutzer im Datensatz werden die Eigenschaften, wie in Abbildung 9 beispielhaft für den Nutzer mit der ID 10 dargestellt, in englischer Sprache mit Hilfe der Computersoftware MATLAB gespeichert. Die Codierung der Eigenschaften erfolgt nach dem gleichen Schema wie in dem bereits von Sube entwickelten Lernmodul für die Sekundarstufe II. Zahlen zu Eigenschaften, die kein Nutzer besitzt, werden hier ebenfalls nicht angegeben (Sube, 2016, S. 28):

- Gender: 1 - Male; 2 - Female
- Relationship Status: 0 - Unknown; 1 - It's complicated; 2 - Single; 4 - In Relationship; 5 - Married
- Interest: 2 - Friends; 3 - Activity Partners; 4 - Dating Men; 5 - Dating Women; 6 - Relationship with Men; 7 - Relationship with Women
- Hetero: 0 - gay; 1 - hetero

Es ist noch zu erwähnen, dass die gespeicherten Nutzer IDs nicht mehr den originalen IDs aus der Friendster Datenbank entsprechen. Nach jeder Filterung werden die Nutzer IDs neu vergeben, sodass diese keine Lücken aufweisen. Für den betrachteten Datensatz aus 44747 Nutzern ist die Nutzer ID i demnach immer ein Zahl zwischen 1 und 44747. Die gespeicherten Eigenschaften eines Nutzers können mit Hilfe der ID in MATLAB durch den Befehl $Users(i)$ mit $i \in [1, 44747]$ abgerufen werden, wenn der Datensatz bereits eingelesen wurde.

3.3. Heuristiken definieren und anwenden

Ziel des entwickelten Lernmoduls ist es, mit Hilfe einfacher, mathematischer Methoden Regeln aufzustellen, welche das Alter einer Person gut vorhersagen. Im Folgenden wird das Vorgehen zur Untersuchung der Fragestellung kurz dargestellt und die einzelnen Vorhersageregeln, welche im Laufe des CAMMP days zum Einsatz kommen, vorgestellt.

3.3.1. Vorgehen zur Untersuchung der Fragestellung

Die verwendeten Regeln, welche das Alter einer Person vorhersagen sollen, werden in Heuristiken vorgegeben, welche für das entwickelte Lernmodul bereits als Funktionen in MATLAB programmiert sind. Das Wort Heuristik stammt von dem griechischen Wort „heuriskein“ ab und bedeutet „finden“ oder „entdecken“. Unter Heuristiken werden Verfahren verstanden, die zum Lösen eines Problems genutzt werden können und neue Erkenntnisse gewinnen. Dazu werden unter Anderem Vermutungen, Analogien, Generalisierungen oder Betrachtungen von Zusammenhängen genutzt (vgl. *Meyers Enzyklopädisches Lexikon (Band 11)*, 1974, S. 819).

Im entwickelten Lernmodul sollen die SuS die vorgegebenen Heuristiken unter methodischer Anleitung auf ihre Effektivität überprüfen, indem sie die einzelnen Regeln auf ihren Datensatz anwenden. Dazu gehen die SuS wie folgt vor:

1. Es wird eine Heuristik zur Vorhersage des Alters ausgewählt.
2. Zunächst wird aus dem vorliegenden Datensatz die Stichprobengröße N festgelegt, auf welche die Heuristik angewendet werden soll. Die Auswahl der Personen erfolgt dabei zufällig nach dem Prinzip „Ziehen ohne Zurücklegen“.
3. Die Heuristik wird auf die Stichprobe angewendet und damit das Alter der Nutzer in dieser Stichprobe vorhergesagt.
4. Für jeden Nutzer wird die getroffene Vorhersage durch einen Vergleich mit der angegebenen Altersangabe überprüft.

Im letzten Schritt können genau zwei Fälle eintreten:

- Die getroffene Vorhersage ist richtig oder
- die getroffene Vorhersage ist falsch.

Das beschriebene Vorgehen kann in Anlehnung an (Cramer & Kamps, 2017, S. 201-202) mathematisch durch ein stochastisches Zufallsexperiment modelliert werden:

Modell eines N-fach wiederholten, stochastisch unabhängigen Zufallsexperiments:

Für jeden der N Nutzer wird unabhängig der anderen Nutzer mit einer ausgewählten Heuristik das Alter vorhergesagt. Es können dabei die beiden Ereignisse „die Vorhersage ist richtig“ und „die Vorhersage ist falsch“ eintreten. Weiter seien $X_1, \dots, X_N \sim \text{bin}(1, p)$ identisch verteilte, stochastisch unabhängige, binomialverteilte Zufallsvariablen. Dabei gibt X_i , $i \in \{1, \dots, N\}$ das Ergebnis der i-ten Einschätzung eines Nutzers an und kann die Werte

$$X_i = \begin{cases} 0, & \text{falls die Vorhersage falsch ist} \\ 1, & \text{falls die Vorhersage richtig ist} \end{cases}, \quad i \in \{1, \dots, N\}$$

annehmen. Die Wahrscheinlichkeit $P(X_i = 1) = p$, $p \in [0, 1]$, eine richtige Vorhersage zu treffen, ist dabei für jeden Nutzer gleich und nicht bekannt. Diese Wahrscheinlichkeit p wird als *Trefferwahrscheinlichkeit* bzw. *Trefferquote* bezeichnet.

Die unbekannte Trefferwahrscheinlichkeit kann für eine große Stichprobengröße mit Hilfe der relativen Häufigkeit näherungsweise bestimmt werden, wie die schwache Version vom Gesetz der großen Zahlen später zeigen wird. Die relative Häufigkeit wird dabei durch Zählen der Anzahl an richtigen Vorhersagen bestimmt. Zunächst werden die Begriffe absolute und relative Häufigkeit, bezogen auf die getroffene Vorhersage, definiert.

Absolute Häufigkeit (vgl. Cramer & Kamps, 2017, S. 16):

Für das Merkmal „getroffene Vorhersage eines Nutzers“ mit den möglichen Merkmalsausprägungen $u_1 = 0$, für eine falsche Vorhersage, und $u_2 = 1$, für eine richtige Vorhersage, liegt die Urliste x_1, \dots, x_N vor.

Die Zahl n_j , $j \in \{1, 2\}$ gibt die Anzahl der Merkmalsausprägung u_j in der Urliste an und heißt absolute Häufigkeit der Beobachtung u_j . Es gilt:

$$n_j = |\{i \in \{1, \dots, N\} | x_i = u_j\}|, \quad j \in \{1, 2\}$$

Relative Häufigkeit (vgl. Cramer & Kamps, 2017, S. 17):

Die absolute Häufigkeit der Merkmalsausprägung u_j in der Urliste sei durch n_j gegeben, $j \in \{1, 2\}$. Der Quotient

$$f_j = \frac{n_j}{N}$$

mit der gewählten Stichprobengröße N , heißt relative Häufigkeit der Merkmalsausprägung u_j .

Als nächstes wird gezeigt, dass die relative Häufigkeit f_2 für die Merkmalsausprägung u_2 , eine richtige Vorhersage zu treffen, für eine große Stichprobe N eine gute Näherung für die unbekannte Trefferwahrscheinlichkeit darstellt. Denn mit Hilfe des schwachen Gesetzes der großen Zahlen ist es möglich, einen Zusammenhang zwischen

den relativen Häufigkeiten und den entsprechenden Wahrscheinlichkeiten im zugehörigen stochastischen Modell herzustellen. Das schwache Gesetz der großen Zahlen lautet:

Definition Schwaches Gesetz der großen Zahlen (vgl. Cramer & Kamps, 2017, S. 246):

Gegeben seien stochastisch unabhängige und identisch verteilte Zufallsvariablen X_1, \dots, X_n mit Erwartungswert $E(X_i) = \mu$ und Varianz $V(X_i) \leq M < \infty \forall i \in \{1, \dots, n\}$ für ein konstantes $M > 0$. Dann gilt:

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| \geq \varepsilon\right) \leq \frac{M}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0 \quad \forall \varepsilon > 0$$

Das schwache Gesetz der großen Zahlen besagt also, dass mit wachsender Stichprobe die Wahrscheinlichkeit, dass das arithmetische Mittel der Zufallsvariablen vom Erwartungswert der Verteilung um mindestens ε abweicht, gegen Null geht.

In dem oben aufgestellten Modell eines N -fach wiederholten, stochastisch unabhängigen Zufallsexperiments werden N unabhängige Versuchswiederholungen durchgeführt und die Zufallsvariablen X_i , $1 \leq i \leq N$ sind stochastisch unabhängig sowie binomialverteilt. Das heißt, es gilt (vgl. Cramer & Kamps, 2017, S. 247):

$$X_i \sim \text{bin}(1, p), \quad E(X_i) = P(A_i) = p, \quad \text{Var}(X_i) = p(1 - p), \quad \forall i \in \{1, \dots, N\}.$$

A_i beschreibt dabei das Ereignis, für den i -ten Nutzer eine richtige Vorhersage zu treffen. Weiter tragen in dem aufgestellten Modell nur die Zufallsvariablen mit $X_i = 1$ für $i \in \{1, \dots, N\}$ zum arithmetischen Mittel bei, womit das arithmetische Mittel mit der relativen Häufigkeit f_2 , eine richtige Vorhersage zu treffen, gleichzusetzen ist. Nach dem schwachen Gesetz der großen Zahlen konvergiert dann die relative Häufigkeit $f_2 = \frac{1}{N} \sum_{i=1}^N X_i$ stochastisch gegen die unbekannte Trefferwahrscheinlichkeit p . Das bedeutet, dass für eine relativ große Stichprobe N die relative Häufigkeit f_2 für die Trefferwahrscheinlichkeit p angenähert werden kann.

Der Vollständigkeit halber wird noch kurz auf den Begriff der Binomialverteilung eingegangen:

Definition Binomialverteilung (vgl. Cramer & Kamps, 2017, S. 171):

Es liegt eine Reihe von $n \in \mathbb{N}$ gleichartigen und unabhängigen Zufallsexperiment vor, die jeweils nur zwei mögliche Ergebnisse haben (Erfolg bzw. Treffer und Niete bzw. Nichttreffer). Diese Art von Zufallsexperimenten werden Bernoulli-Experimente genannt. Dann gibt die Binomialverteilung

$$P(\text{„genau } k \text{ Treffer“}) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$$

die Wahrscheinlichkeit $p \in [0, 1]$ an, mit der sich insgesamt genau $k \in \{1, \dots, n\}$ Treffer ergeben.

3.3.2. Zu untersuchende Heuristiken

Im Laufe des entwickelten Lernmoduls kommen unterschiedliche Heuristiken in Form von Vorhersageregeln zum Einsatz, die im Folgenden kurz vorgestellt werden. Es gibt einerseits Heuristiken, welche mit Hilfe von Lagemaßen das exakte Alter einer Person vorhersagen und andererseits Heuristiken, die unter der Verwendung von Streuungsmaßen das Alter eines Nutzers innerhalb eines gewissen Intervalls vorhersagen. Außerdem werden Heuristiken verwendet, die einen Nutzer innerhalb oder außerhalb eines vorgegebenen Altersintervall einordnen.

Heuristiken, die das exakte Alter eines Nutzers vorhersagen:

Als Lagemaß zur Vorhersage des exakten Alters werden das arithmetische Mittel, der Median und der Modus verwendet. Auf die einzelnen Lagemaße wird in Abschnitt 3.4 genauer eingegangen. Für jedes Lagemaß gibt es eine Heuristik, welche das Alter der Freunde 1. Ordnung eines Nutzers betrachtet, und eine, die das Alter der Freunde 2. Ordnung berücksichtigt. Die einzelnen Regeln können wie folgt definiert werden, wobei die Regel für die Freunde 1. und 2. Ordnung jeweils zusammengefasst sind:

Vorhersageregeln zum arithmetischen Mittel:

Das Alter des Nutzers entspricht dem auf eine ganze Zahl gerundeten arithmetischen Mittel des angegebenen Alters der Freunde (bzw. der Freundes-Freunde).

Vorhersageregeln zum Median:

Das Alter des Nutzers entspricht dem auf eine ganze Zahl gerundeten Median des angegebenen Alters der Freunde (bzw. der Freundes-Freunde).

Vorhersageregeln zum Modus:

Das Alter des Nutzers entspricht dem Modus des angegebenen Alters der Freunde (bzw. der Freundes-Freunde). Bei mehreren Modi ist das Alter des Nutzers das auf eine ganze Zahl gerundete arithmetische Mittel dieser Werte.

Heuristiken, die das Alter eines Nutzers innerhalb eines Intervalls vorhersagen:

Als Streuungsmaß zur Vorhersage des Alters eines Nutzers innerhalb eines Intervalls werden die Spannweite, der Quartilsabstand und die Standardabweichung betrachtet. Auch hier gibt es zu jedem genannten Streuungsmaß jeweils eine Heuristik für die Freunde 1. und 2. Ordnung. Die Größe des vorhergesagten Intervalls, in dem das Alter der Nutzer liegen soll, hängt dabei von den Altersangaben der Freunde 1. oder 2. Ordnung der jeweiligen Person ab. Eine detaillierte Darstellung der drei Streuungsmaße wird im Abschnitt 3.5 vorgenommen. Die einzelnen Regeln werden wie folgt definiert, wobei auch hier wieder die Regeln für die Freunde 1. und 2. Ordnung zusammengefasst

sind:

Vorhersageregeln zur Spannweite:

Das Alter des Nutzers liegt innerhalb der Spannweite des angegebenen Alters der Freunde (bzw. der Freundes-Freunde).

Vorhersageregeln zum Quartilsabstand:

Das Alter des Nutzers liegt innerhalb des Quartilsabstands des angegebenen Alters der Freunde (bzw. der Freundes-Freunde).

Vorhersageregeln zur Standardabweichung:

Es wird für den Nutzer das arithmetische Mittel \bar{x} und die Standardabweichung s des angegebenen Alters der Freunde (bzw. der Freundes-Freunde) bestimmt. Das Alter des Nutzers liegt dann innerhalb der Werte $\bar{x} - s$ und $\bar{x} + s$.

Heuristiken, die das Alter eines Nutzers innerhalb eines vorgegebenen Intervalls vorhersagen:

In dem entwickelten Lernmodul sollen Nutzer mit einem Alter innerhalb des vorgegebenen Intervalls $[25, 30]$ vorhergesagt werden. Dazu werden Heuristiken

- zum Zufall,
- zum Anteil der Freunde 1. Ordnung eines Nutzers mit einem Alter innerhalb des gesuchten Intervalls $[25, 30]$ und
- zum arithmetischen Mittel

aufgestellt. Die Heuristik zum Zufall lautet:

Vorhersageregeln zum Zufall:

Mit der Wahrscheinlichkeit q wird dem Nutzer ein Alter zwischen 25 und 30 Jahre zugeordnet. Mit der Wahrscheinlichkeit $1 - q$ ist der Nutzer nicht zwischen 25 und 30 Jahre alt.

In den verwendeten Datensätzen aus dem entwickelten Lernmodul sind die Altersangaben jedes Nutzers bekannt und werden als richtig angenommen. Die SuS wählen dort q datenbasiert als relative Häufigkeit in ihrem Datensatz Nutzer mit einem Alter innerhalb des Intervalls $[25, 30]$ anzutreffen. In dem Datensatz bestehend aus 44747 Nutzern besitzen gerundet 42 % der Nutzer ein Alter zwischen 25 und 30 Jahre. Aus diesem Grund wird dort $q = 0.42$ gewählt und somit einem zufälligen Nutzer mit einer Wahrscheinlichkeit von 42 % ein Alter innerhalb des Intervalls $[25, 30]$ zugeordnet. Als nächstes wird die Heuristik zum Anteil der Freunde 1. Ordnung vorgestellt. Die Idee dahinter ist, dass Nutzer einen höheren Anteil an Freunden ihres eigenen Alters besitzen. Die hier angegebenen Anteile können und sollen von den SuS im CAMMP

day variiert werden. Durch Anwendung der Heuristik sollen die SuS die Anteile herausfinden, welche zur besten Trefferquote für eine richtige Vorhersage führen.

Vorhersageregeln zum Anteil der Freunde:

Wenn mehr als 40% der Freunde eines Nutzers ein Alter zwischen 25 und 30 Jahren besitzen, dann befindet sich das Alter dieses Nutzers ebenfalls innerhalb dieser Altersspanne. Wenn weniger als 20% der Freunde ein Alter zwischen 25 und 30 Jahren besitzen, dann befindet sich das Alter dieses Nutzers nicht in dieser Altersspanne. Sonst lasse den Zufall entscheiden.

Die zufällige Entscheidung orientiert sich dabei an der eben aufgestellten Vorhersageregeln zum Zufall. Das heißt, liegt in dem verwendeten Datensatz aus 44747 Personen für einen Nutzer der Anteil der Freunde mit einem Alter innerhalb des Intervalls [25, 30] zwischen 20 und 40 Prozent, so wird diesem Nutzer mit der Wahrscheinlichkeit $q = 0.42$ ein Alter zwischen 25 und 30 Jahre zugeordnet.

Zum Schluss wird nochmal die bereits erwähnte Heuristik zum arithmetischen Mittel aufgegriffen. Jedoch soll hier keine Bestimmung des exakten Alters stattfinden, sondern überprüft werden, ob das arithmetische Mittel innerhalb des Intervalls [25, 30] liegt.

Vorhersageregeln zum arithmetischen Mittel im vorgegebenen Intervall:

Es wird für den Nutzer das arithmetische Mittel des angegebenen Alters der Freunde (bzw. der Freundes-Freunde) bestimmt. Liegt dieser Wert zwischen 25 und 30 Jahren, dann befindet sich das Alter des Nutzers ebenfalls innerhalb dieser Altersspanne.

3.3.3. Ergebnisse der verwendeten Heuristiken

Die vorgestellten Heuristiken bzw. Vorhersageregeln wurden auf den Datensatz mit 44747 Nutzern angewendet. Tabelle 3 zeigt die resultierenden Trefferquoten der jeweiligen Heuristiken, die sich nach der Anwendung ergeben. Es ist zu erkennen, dass die Heuristiken, welche auf Lagemaßen beruhen und das exakte Alter vorhersagen, nur geringe Trefferquoten erzielen. Mit einer Trefferquote von 19 % ist die Regel, welche den Median der Altersangaben der Freunde 1. Ordnung eines Nutzers bestimmt, unter den Lagemaßen die beste Heuristik. Generell ergeben sich bei den Lagemaßen für die Untersuchung der Freunde bessere Ergebnisse als bei den Heuristiken, welche die Freunde 2. Ordnung betrachten. Hier kann die Vermutung aufgestellt werden, dass die Freunde 1. Ordnung genauere Informationen über den Nutzer preisgeben als die Freunde 2. Ordnung.

Bei den Heuristiken zu den Streuungsmaßen ergibt sich genau das umgekehrte Bild. Die Vorhersageregeln zu den Freundes-Freunden erzielen deutlich bessere Trefferquoten als zu den Freunden 1. Ordnung. Der Grund dafür liegt an den größeren Intervallbreiten, in denen das vorhergesagte Alter der Nutzer liegen soll, die sich bei der Betrachtung der Freunde 2. Ordnung ergeben. Tabelle 4 zeigt für die Heuristiken zu den Streuungsmaßen die durchschnittlichen Intervallbreiten, sowohl bei Betrachtung

der Freunde, als auch bei Betrachtung der Freundes-Freunde. Ein Vergleich der beiden Tabellen zeigt, dass je größer die Intervallbreite ist, desto höher ist die Trefferquote für eine richtige Vorhersage des Alters. Die höchste Trefferwahrscheinlichkeit ergibt mit 76 % die Vorhersage zur Spannweite der Freunde 2. Ordnung. Jedoch liegt hier auch die durchschnittliche Intervallbreite, in der das vorhergesagte Alter des Nutzers liegt, bei 20 Jahre. Eine Halbierung der Intervallbreite auf 10 Jahre bei der Regel zur Standardabweichung der Freundes-Freunde wirkt sich nur minimal auf die Trefferquote aus. Diese liegt bei der Standardabweichung der Freunde 2. Ordnung immer noch bei 72 %. Mit einer durchschnittlichen Intervallbreite von 5 Jahren erzielt die Heuristik zum Quartilsabstand der Freunde eine Trefferquote von 56 %. In dem entwickelten Lernmodul sollen die SuS die Trefferquoten mit den durchschnittlichen Intervallbreiten in Beziehung setzen und ausgehend aus ihren herausgefundenen Ergebnissen eine begründete Einschätzung für die ihrer Meinung nach beste Vorhersageregel abgeben. Für die Vorhersage, ob sich das Alter eines Nutzers inner- oder außerhalb des Intervalls [25, 30] befindet, liefern die beiden letzten Heuristiken in Tabelle 3 zu den Freunden 1. Ordnung mit 80 % eine gleich gute Trefferquote. Sie ist knapp 30 % besser als eine zufällige Zuordnung der Nutzer in eine der beiden Altersgruppen. Insgesamt stellt sie die beste, erzielte Trefferquote dar, welche mit den aufgestellten Heuristiken für diesen Datensatz erzielt werden konnte.

Tabelle 3: Ermittelte, gerundete Trefferquoten nach Anwendung der einzelnen Heuristiken auf den Datensatz mit 44747 Nutzern.

Untersucht: Heuristik:	Freunde 1. Ordnung	Freunde 2. Ordnung
Arithmetisches Mittel	17 %	11 %
Median:	19 %	14 %
Modus:	18 %	14 %
Spannweite:	53 %	76 %
Quartilsabstand:	47 %	56 %
Standardabweichung:	53 %	72 %
Zufällige Zuordnung:	51 %	nicht untersucht
Anteil der Freunde:	80 %	nicht untersucht
Arithmetisches Mittel im vorgegebenen Intervall:	80 %	75 %

Tabelle 4: Durchschnittliche Intervallbreiten, in denen das Alter der Nutzer liegt, welches mit den Heuristiken zur Spannweite, zum Quartilsabstand und zur Standardabweichung vorhergesagt wurde. Die Werte beziehen sich wieder auf den Datensatz bestehend aus 44747 Nutzern.

Untersucht:	Freunde 1. Ordnung	Freunde 2. Ordnung
Heuristik:		
Spannweite:	6 Jahre	20 Jahre
Quartilsabstand:	3 Jahre	5 Jahre
Standardabweichung:	5 Jahre	10 Jahre

3.4. Lagemaße

Lagemaße sind in der Statistik gängige Hilfsmittel, um eine Reduktion der in den Daten enthaltenen Informationen vorzunehmen und einen Datensatz mit nur wenigen Kenngrößen zu beschreiben. Sie dienen dabei der Beschreibung eines Zentrums der vorhandenen Daten in einem Datensatz (vgl. Cramer & Kamps, 2017, S. 23).

Wie bereits erwähnt werden in dem entwickelten Lernmodul zur Vorhersage des exakten Alters Heuristiken zum arithmetischen Mittel, Modus und Median aufgestellt. Diese Lagemaße werden im Folgenden vorgestellt.

3.4.1. Das arithmetische Mittel

Das arithmetische Mittel \bar{x} , auch bekannt unter den Begriffen Mittelwert oder Durchschnitt, ist ein Lagemaß zur Beschreibung metrischer Daten. Als metrische Daten werden Merkmalsausprägungen bezeichnet, die sich durch Zahlen erfassen lassen und bei denen Abstände zwischen diesen Zahlen sinnvoll interpretiert werden können (vgl. Cramer & Kamps, 2017, S. 8). Eine formale Definition für das arithmetische Mittel lautet nach Cramer und Kamps:

Definition arithmetisches Mittel (vgl. Cramer & Kamps, 2017, S. 28):
 Sei x_1, \dots, x_n ein Datensatz aus n Werten eines metrischen Merkmals. Das arithmetische Mittel \bar{x} ist definiert durch

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i.$$

In Worten lässt sich das arithmetische Mittel darstellen als

$$\bar{x} = \frac{\text{Summe aller Werte}}{\text{Anzahl aller Werte}}.$$

Für ein besseres Verständnis der Lage- und Streuungsmaße wird der metrische Beispieldatensatz

$$8, 10, 13, 13, 10, 9, 11, 15, 10 \quad (1)$$

betrachtet, anhand dessen die vorgestellten Lage- und Streuungsmaße beispielhaft bestimmt werden. Für diesen Datensatz ergibt sich für das arithmetische Mittel der Wert

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{8 + 9 + 11 + 3 \cdot 10 + 2 \cdot 13 + 15}{9} = 11.$$

Beim arithmetischen Mittel handelt es sich um einen repräsentativen Wert, der den gesamten Datensatz berücksichtigt und in einer Kennzahl zusammenfasst (vgl. Krüger et al., 2015, S. 57). Da für die Berechnung alle Werte des Datensatzes berücksichtigt werden, haben Ausreißer einen großen Einfluss auf das arithmetische Mittel. Als Ausreißer werden Werte bezeichnet, die in Relation zu der Mehrzahl der Daten im Datensatz verhältnismäßig groß oder klein sind (vgl. Cramer & Kamps, 2017, S. 33).

3.4.2. Der Median

Der Median \tilde{x} wird auch als Zentralwert bezeichnet und ist nur für ordinalskalierte und metrische Merkmale anwendbar (vgl. Mittag, 2017, S. 62). Von ordinalen Merkmalen ist die Rede, wenn die Werte eines Datensatzes geordnet werden können und hinsichtlich ihrer Größe vergleichbar sind. Das heißt, es kann unterschieden werden, ob eine Ausprägung im Vergleich zu einer anderen Ausprägung kleiner bzw. schlechte, gleich oder größer bzw. besser ist (vgl. Cramer & Kamps, 2017, S. 8). Für die Bestimmung des Medians wird zunächst die Rangwertreihe eines Datensatzes definiert:

Definition Rangwertreihe (vgl. Cramer & Kamps, 2017, S. 18):

Seien x_1, \dots, x_n Beobachtungswerte eines ordinalskalierten oder metrischskalierten Merkmals. Dann heißt die aufsteigende Auflistung der Beobachtungswerte

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

Rangwertreihe.

Der Beispieldatensatz (1) besitzt die Rangwertreihe

$$x_{(1)} = 8 \leq 9 \leq 10 \leq 10 \leq 10 \leq 11 \leq 13 \leq 13 \leq 15 = x_{(9)}. \quad (2)$$

Der Median wird auf der Basis der Rangwertreihe eines Datensatzes definiert. Er liegt immer „in der Mitte“ der Daten, in dem Sinne, dass 50 % aller Daten größer oder gleich und 50 % aller Daten kleiner oder gleich dem Median sind (vgl. Cramer & Kamps, 2017, S. 25). Da das Alter, welches im entwickelten Lernmodul betrachtet wird, ein metrisches Merkmal darstellt, wird hier nur auf die Definition des Medians für diesen

Merkmalstyp eingegangen:

Definition Median (vgl. Cramer & Kamps, 2017, S. 27):

Sei $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ die Rangwertreihe eines metrisch skalierten Datensatzes x_1, x_2, \dots, x_n . Dann ist der Median \tilde{x} definiert durch

$$\tilde{x} = \begin{cases} x_{(\frac{n+1}{2})}, & \text{falls } n \text{ ungerade,} \\ \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}), & \text{falls } n \text{ gerade.} \end{cases}$$

In dem Beispieldatensatz (1) ist $n = 9$ ungerade. Mit Blick auf die Rangwertreihe 2 ergibt sich damit für diesen Datensatz der Median

$$\tilde{x} = x_{(\frac{9+1}{2})} = x_{(5)} = 10.$$

Ist die Stichprobengröße N ungerade, so ist der Median immer eindeutig bestimmt und es kommt nur ein Wert in Frage. Bei einer geraden Stichprobengröße wird für metrische Merkmale der Median aus dem arithmetischen Mittel der beiden mittleren Werte bestimmt, um auch in diesem Fall einen eindeutigen Wert zu erhalten (vgl. Mittag, 2017, S. 62-63).

Der Median verhält sich im Gegensatz zum arithmetischen Mittel robust gegenüber Ausreißern, also gegenüber besonders großen oder kleinen Werten im Datensatz (vgl. Cramer & Kamps, 2017, S. 33).

3.4.3. Der Modus

Der Modus x_{mod} oder auch Modalwert ist ein Lagemaß, welches zur Beschreibung aller Datensätze verwendet werden kann und beschreibt den Wert, der am häufigsten im Datensatz vorkommt (vgl. Mittag, 2017, S. 62). Dieser Wert weist demnach die größte absolute bzw. relative Häufigkeit auf. Formal kann der Modus wie folgt definiert werden:

Definition Modus (vgl. Cramer & Kamps, 2017, S. 23-24):

In einem Datensatz haben die verschiedenen Merkmalsausprägungen u_1, \dots, u_m die absoluten Häufigkeiten n_1, \dots, n_m und relativen Häufigkeiten f_1, \dots, f_m . Jede Merkmalsausprägung u_j , $j \in \{1, \dots, m\}$ mit der Eigenschaft

$$n_j = \max\{n_1, \dots, n_m\} \text{ bzw. } f_j = \max\{f_1, \dots, f_m\}$$

wird als Modus x_{mod} bezeichnet.

Der Modus im dargestellten Beispieldatensatz 1 beträgt 10, da dieser Wert dort am häufigsten auftritt.

3.5. Streuungsmaße

Die alleinige Angabe eines Lagemaßes zur Beschreibung der Verteilung eines Datensatzes ist in der Statistik unzureichend. Denn zwei Datensätze können in einem Lagemaß, wie beispielsweise dem arithmetische Mittel \bar{x} , übereinstimmen, jedoch lässt diese Kenntnis offen, wie stark die einzelnen Werte in den Datensätzen von diesem Mittelwert abweichen bzw. streuen. Diese Streuung innerhalb eines Datensatzes wird durch Streuungsmaße quantifiziert (vgl. Mittag, 2017, S. 23-24). Unter der Berücksichtigung der auftretenden Größenordnungen im Datensatz gilt, dass je größer der Wert eines Streuungsmaßes ist, desto mehr streuen die Daten im Datensatz. Zudem setzen Streuungsmaße einen Abstandsbegriff voraus, wodurch diese nur bei metrisch skalierten Merkmalen bestimmt werden können (vgl. Cramer & Kamps, 2017, S. 34). Um ein Intervall vorherzusagen, in dem das Alter eines Nutzers liegt, werden im Lernmodul die Streuungsmaße Spannweite, Quartilsabstand und Standardabweichung verwendet und nun kurz vorgestellt.

3.5.1. Die Spannweite

Die Spannweite R ergibt sich aus der Differenz aus dem größten Wert (Maximum) und dem kleinsten Wert (Minimum) innerhalb eines Datensatzes:

Definition Spannweite (vgl. Mittag, 2017, S. 68):

Für einen metrischen Datensatz x_1, \dots, x_n mit Rangwertreihe $x_{(1)} \leq \dots \leq x_{(n)}$ ist die Spannweite R definiert als

$$R = x_{(n)} - x_{(1)}.$$

Dabei wird $x_{(n)}$ als *Maximum* und $x_{(1)}$ als *Minimum* bezeichnet.

Für den Beispieldatensatz (1) ergibt sich mit der Betrachtung seiner Rangwertreihe (2) eine Spannweite von

$$R = x_{(n)} - x_{(1)} = 15 - 8 = 7.$$

Aus der Definition wird bereits das Problem ersichtlich, dass sich die Spannweite bei der Berechnung nur auf die äußersten Werte eines Datensatzes stützt und damit höchst anfällig gegenüber Ausreißern ist. Liegen beispielsweise alle Werte eng beieinander und nur ein Wert liegt als Ausreißer weit von diesen entfernt, würde die Interpretation der berechneten Spannweite auf eine hohe Streuung schließen lassen. Dass die meisten Werte jedoch nicht stark streuen, würde nicht berücksichtigt werden. Das heißt, die Spannweite kann schnell einen verfälschten Eindruck vom Ausmaß der Streuung der Daten innerhalb eines Datensatzes hinterlassen (vgl. Cleff, 2008, S. 68). Beispielsweise würde die Hinzunahme des Werts 25 zum Datensatz (1) zu einer Spannweite von $R = 25 - 8 = 17$ führen, sodass durch einen einzelnen Wert eine deutlich größere Spannweite bewirkt wird.

3.5.2. Der Quartilsabstand

Da sich der Quartilsabstand Q aus dem unteren und oberen Quartil zusammensetzt, wird zunächst das p -Quartil eingeführt. Ein p -Quartil \tilde{x}_p hat die Eigenschaft, dass mindestens $p \cdot 100\%$ der Werte der Rangwertreihe $x_{(1)} \leq \dots \leq x_{(n)}$ eines Datensatzes kleiner oder gleich und mindestens $(1 - p) \cdot 100\%$ der Werte größer oder gleich \tilde{x}_p sind (vgl. Mittag, 2017, S. 73). Das p -Quartil kann folgendermaßen definiert werden:

Definition p -Quartil (vgl. Cramer & Kamps, 2017, S. 28):

Für einen metrischen Datensatz x_1, \dots, x_n sei die Rangwertreihe $x_{(1)} \leq \dots \leq x_{(n)}$ gegeben. Das p -Quartil \tilde{x}_p für $p \in (0, 1)$ ist gegeben durch

$$\tilde{x}_p = \begin{cases} x_{(k)}, & \text{falls } np < k < np + 1, \quad np \in \mathbb{N} \\ \frac{1}{2}(x_{(k)} + x_{(k+1)}), & \text{falls } k = np, \quad np \in \mathbb{N}. \end{cases}$$

Aus der Definition wird ersichtlich, dass der Median einen Spezialfall des p -Quartils darstellt und als 0.5-Quartil angesehen werden kann. Auch beim p -Quartil für metrische Merkmale wird im zweiten Fall die Eindeutigkeit durch die Berechnung des arithmetischen Mittels der beiden Werte, zwischen denen das p -Quartil liegt, erreicht. Das 0.25-Quartil wird als *unteres Quartil* und das 0.75-Quartil als *oberes Quartil* bezeichnet. Der Quartilsabstand wird als Differenz aus oberen und unterem Quartil definiert:

Definition Quartilsabstand (vgl. Cramer & Kamps, 2017, S. 35):

Der Quartilsabstand Q eines metrischen Datensatzes x_1, \dots, x_n ist definiert als

$$Q = \tilde{x}_{0.75} - \tilde{x}_{0.25}.$$

Dabei bezeichnet $\tilde{x}_{0.75}$ das obere und $\tilde{x}_{0.25}$ das untere Quartil der Daten.

Es wird erneut der Datensatz (1) mit seiner zugehörigen Rangwertreihe (2) betrachtet und der Quartilsabstand bestimmt:

$$Q = \tilde{x}_{0.75} - \tilde{x}_{0.25} = x_{(7)} - x_{(3)} = 13 - 10 = 3.$$

Aus der Definition des Quartilsabstands ist zu erkennen, dass die Daten nur in Form der beiden Quartile eingehen und die Ausreißer eines Datensatzes in der Regel nicht in die Berechnung mit einbezogen werden. Aus diesem Grund ist der Quartilsabstand gegenüber Ausreißer nicht besonders anfällig. Wird beispielsweise zum Datensatz (1) die beiden Ausreißer 1 und 25 hinzugefügt ergibt sich mit $x_{(1)} = 1 \leq 8 \leq 9 \leq 10 \leq 10 \leq 10 \leq 11 \leq 13 \leq 13 \leq 15 \leq 25 = x_{(11)}$ eine neue Rangwertreihe. Der Quartilsabstand dieses Datensatzes liegt bei $Q = \tilde{x}_{0.75} - \tilde{x}_{0.25} = x_{(11)} - x_{(3)} = 13 - 9 = 4$, sodass die beiden hinzugenommenen Ausreißer den Quartilsabstand nicht stark beeinflussen.

3.5.3. Die Standardabweichung

Die Standardabweichung s misst, wie stark die einzelnen Daten im Datensatz vom Mittelwert \bar{x} abweichen und ist wie folgt definiert:

Definition Standardabweichung (vgl. Mittag, 2017, S. 69):

Die Standardabweichung s eines metrischen Datensatzes x_1, \dots, x_n mit zugehörigem arithmetischem Mittel \bar{x} ist definiert durch

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Für den Beispieldatensatz (1) ergibt sich mit dem bereits berechneten arithmetischem Mittel $\bar{x} = 11$ die Standardabweichung

$$s = \sqrt{\frac{1}{9} \sum_{i=1}^9 (x_i - 11)^2} \approx 2.11$$

Der Ausdruck $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ unterhalb der Wurzel wird als *Varianz* bezeichnet und gibt den mittleren quadratischen Abstand der einzelnen Daten vom arithmetischem Mittel an. Den normalen Abstand zum arithmetischem Mittel zu wählen ist nicht sinnvoll, da aus der Definition des Mittelwerts klar wird, dass $\sum_{i=1}^n (x_i - \bar{x}) = 0$ ergibt. Das arithmetische Mittel wird auf Grund dieser Eigenschaft auch als Schwerpunkt des Datensatzes interpretiert (vgl. Cleff, 2008, S. 59). Durch das Ziehen der Wurzel wird die Standardabweichung in der Einheit angegeben, in der auch die Ausgangsdaten gemessen werden. Damit stellt die Standardabweichung im Vergleich zur Varianz ein wesentlich anschaulicheres Streuungsmaß dar (vgl. Mittag, 2017, S. 69). Da die Abweichung der Daten vom Mittelwert quadriert werden, fallen Ausreißer besonders stark ins Gewicht und beeinflussen die Standardabweichung (vgl. Cleff, 2008, S. 66). So führt beispielsweise die Hinzunahme des Werts 25 zum Beispieldatensatz (1) zu einer Standardabweichung von $s \approx 5.12$. Der Ausreißer hat also zu einer mehr als doppelt so großen Standardabweichung geführt.

3.6. Boxplot

Der Boxplot stellt eine graphische Methode zur Darstellung unterschiedlicher Lage- und Streuungsmaße eines Datensatzes dar. Dabei können die Lage- und Streuungsmaße, die im Boxplot zum Einsatz kommen, verschieden gewählt werden. Die in Abbildung 10 gezeigte Version stellt nur eine mögliche Variante dar. Sie besteht aus den beiden Extremwerten $x_{min} = x_{(1)}$ und $x_{max} = x_{(n)}$, dem unteren Quartil $\tilde{x}_{0.25}$, dem oberen Quartil $\tilde{x}_{0.75}$ und dem Median \tilde{x} . In der Literatur wird diese Variante des Boxplots als Basisvariante vorgestellt (vgl. Mittag, 2017, S. 75). Bei diesem Boxplot wird der Kasten bzw. die Box vorgegeben durch das untere Quartil $\tilde{x}_{0.25}$ als linker Rand und

das obere Quartil $\tilde{x}_{0.75}$ als rechter Rand. Damit stellt die Länge des Kastens den Quartilsabstand dar. Der Median befindet sich innerhalb der Box und wird in Form eines Striches eingezeichnet. Von dem Kasten führen jeweils links und rechts zwei Linien weg, die als Whisker bezeichnet werden. Die linke Linie endet beim Minimum $x_{(1)}$ des Datensatzes und die rechte beim Maximum $x_{(n)}$. Der Abstand zwischen diesen beiden Extremwerten stellt die Spannweite dar (vgl. Cramer & Kamps, 2017, S. 41).

Das unteren Quartils $\tilde{x}_{0.25}$ ist so definiert, dass höchstens 25 % aller Werte kleiner als $\tilde{x}_{0.25}$ sind. Ebenso gilt für das obere Quartil $\tilde{x}_{0.75}$, dass höchstens 25 % aller Werte größer als $\tilde{x}_{0.75}$ sind (vgl. Krüger et al., 2015, S. 126). Demnach liegen unterhalb sowie oberhalb der Box jeweils höchstens 25 % und innerhalb der Box mindestens 50 % der Daten.

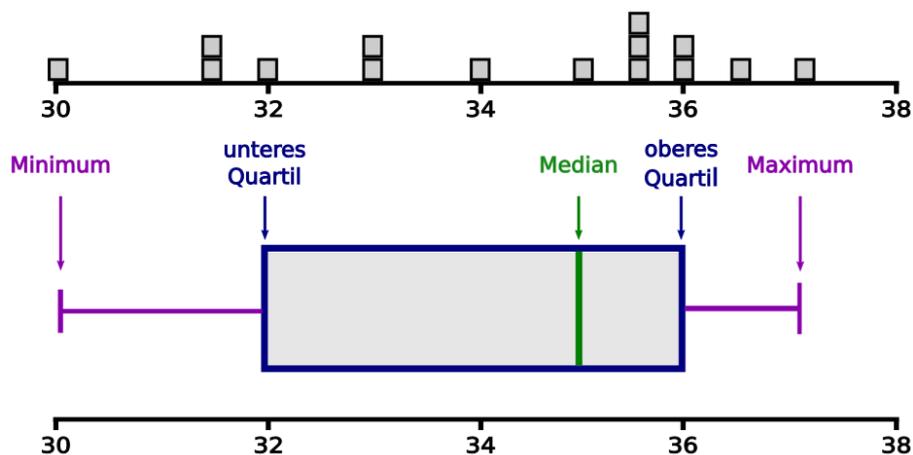


Abbildung 10: Darstellung der Basisvariante eines Boxplots bestehend aus Minimum, Maximum, unteres sowie oberes Quartil und Median. Die Verteilung der Daten wurde dem zugehörigen Boxplot in einem Punktdiagramm gegenübergestellt.

Ein Boxplot gibt einerseits einen Überblick über die Verteilung der Daten und eignet sich andererseits als Vergleich zweier oder mehrerer Datensätze. Liegt beispielsweise ein symmetrischer Boxplot vor, das heißt, der Median liegt in der Mitte der Box und die beiden Whiskers sind gleich lang, kann auf eine symmetrische Verteilung der Daten im Datensatz geschlossen werden. Streuen die Werte kaum, ist dies an den kurzen Whiskers sowie an einer kurzen Box zu erkennen. Auch eine fehlende Symmetrie einer Verteilung lässt sich an Hand eines Boxplots erkennen. Ist beispielsweise der Median innerhalb der Box stark nach rechts verschoben und ggf. auch der rechte Whisker im Vergleich zum linken kurz, wird von einer linksschiefen Verteilung der Daten im Datensatz gesprochen. Das bedeutet, die 50 % der Daten oberhalb des Medians liegen innerhalb eines kleinen Intervalls, während sich die anderen 50 % der Daten auf ein großes Intervall verteilen. Der umgekehrte Fall wird als rechtsschiefe Verteilung bezeichnet (vgl. Cleff, 2008, S. 57). Abbildung 10 zeigt den Boxplot einer linksschiefen Verteilung. Außerdem wurde in dieser Abbildung, wie von Krüger, Sill und Sikora emp-

fohlen, die Verteilung der Daten dem zugehörigen Boxplot in einem Punktdiagramm gegenübergestellt (vgl. Krüger et al., 2015, S. 123).

3.7. Bedingte Wahrscheinlichkeit

Mit der bedingten Wahrscheinlichkeit wird der Einfluss von Vorinformationen in ein stochastisches Modell beschrieben (vgl. Cramer & Kamps, 2017, S. 189). Mit ihr kann die Wahrscheinlichkeit für ein Ereignis A unter der Voraussetzung angegeben werden, dass ein Ereignis B bereits eingetreten ist. Die bedingte Wahrscheinlichkeit kann wie folgt definiert werden:

Definition Bedingte Wahrscheinlichkeit (vgl. Cramer & Kamps, 2017, S. 189): Sei (Ω, Σ, P) ein Wahrscheinlichkeitsraum mit Ereignismenge Ω , σ -Algebra Σ und Wahrscheinlichkeitsmaß P . Für zwei Ereignisse $A, B \in \Sigma$ mit $P(B) > 0$ ist die bedingte Wahrscheinlichkeit von A unter B definiert durch

$$P_B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Dabei bezeichnet $P(A \cap B)$ die Wahrscheinlichkeit, dass A und B gemeinsam auftreten.

Die in der Definition erwähnten Begriffe Ereignismenge, σ -Algebra und Wahrscheinlichkeitsmaß können bei Bedarf im Anhang (siehe Abschnitt A) nachgeschlagen werden. Mit der Definition der bedingten Wahrscheinlichkeit wird klar, dass

$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

ist. Diese Erkenntnis basiert auf dem *Satz von Bayes*. Mit ihm lässt sich die Wahrscheinlichkeit von einem Ereignis A unter der Bedingung, dass Ereignis B eingetreten ist, durch die Wahrscheinlichkeit von B unter der Bedingung, dass A eingetreten ist, berechnen:

Definition Satz von Bayes (vgl. Krüger et al., 2015, S. 149): Sei (Ω, Σ, P) ein Wahrscheinlichkeitsraum. Für zwei Ereignisse $A, B \in \Sigma$ mit $P(B) > 0$ gilt

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}.$$

In Abschnitt 3.3.2 wurde bereits die Zufallsregel vorgestellt, die einem Nutzer im Datensatz mit der Wahrscheinlichkeit q ein Alter innerhalb des Intervalls $[25, 30]$ zuordnet. Dabei wurde die Wahrscheinlichkeit q als relative Häufigkeit interpretiert, mit der im betrachteten Datensatz Nutzer zwischen 25 und 30 Jahren anzutreffen sind. In dem betrachteten Datensatz bestehend aus 44747 Nutzern liegt die relative Häufigkeit bei etwa $q = 0.42$. Im Rahmen des entwickelten CAMMP days sollen die SuS für die Zufallsregel die theoretische Trefferquote bzw. die theoretische Wahrscheinlichkeit bestimmen,

mit der die Zufallsregel einem Nutzer das richtige Altersintervall vorhersagt. Dazu betrachten sie das in Abbildung 11 dargestellte zweistufige Baumdiagramm. In der ersten Stufe stehen die beiden Ereignisse I = „Das zugeordnete Alter liegt innerhalb des Intervalls $[25, 30]$ “ und \bar{I} = „Das zugeordnete Alter liegt nicht innerhalb des Intervalls $[25, 30]$ “. Die zweite Stufe des Baumdiagramms enthält die beiden Ereignisse R = „Die getroffene Zuordnung ist richtig“ und \bar{R} = „Die getroffene Zuordnung ist falsch“. Wie oben bereits erwähnt ist die Wahrscheinlichkeit für das Ereignis I durch die relative Häufigkeit q gegeben, also ist $P(I) = q = 0.42$. Demnach ist $P(\bar{I}) = 1 - q = 0.58$. Bei den Wahrscheinlichkeiten in der zweiten Stufe des Baumdiagramms handelt es sich um bedingte Wahrscheinlichkeiten, da sie von dem Eintreffen der Ereignisse I und \bar{I} beeinflusst werden. Es ist $P_I(R) = q = 0.42$. Der Grund dafür ist, dass im Datensatz aus 44747 Nutzern etwa 42 % der Personen ein Alter zwischen 25 und 30 Jahren aufweisen. Da das Ereignis I eingetreten ist, ist die getroffene Zuordnung zu 42 % korrekt. Mit einer analogen Begründung ist $P_{\bar{I}}(R) = 1 - q = 0.58$. Mit Hilfe der Definition der bedingten Wahrscheinlichkeit ergibt sich dann für die Wahrscheinlichkeit $P(R)$, mittels der Zufallsregel eine richtige Zuordnung in eine der beiden Altersgruppen zu tätigen,

$$P(R) = P(I \cap R) + P(\bar{I} \cap R) = P(I) \cdot P_I(R) + P(\bar{I}) \cdot P_{\bar{I}}(R) \approx 0.51.$$

Damit liegt die theoretische Trefferquote, mit der Zufallsregel eine richtige Vorhersage zu treffen, bei etwa 51 %.

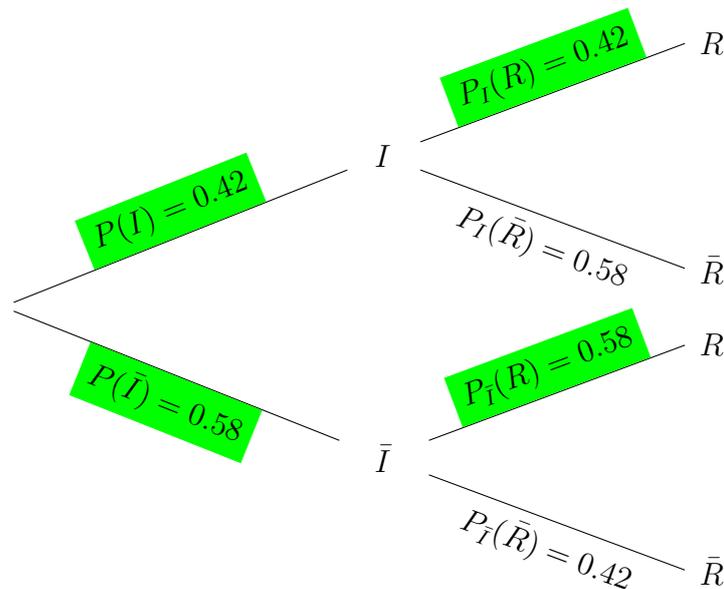


Abbildung 11: Zweistufiges Baumdiagramm zur Bestimmung der theoretischen Trefferwahrscheinlichkeit $P(R)$ mit Hilfe der Vorhersageregeln zum Zufall eine richtige Zuordnung in die beiden Altersintervalle zu tätigen. In grün sind die Wahrscheinlichkeiten der Pfade markiert, die zu einer richtigen Zuordnung bzw. Vorhersage führen.

Da die SuS in der Mittelstufe noch nicht mit bedingten Wahrscheinlichkeiten in Berührung gekommen sind, berechnen sie im entwickelten Lernmodul die theoretische Trefferwahrscheinlichkeit $P(R)$ mit dem zweistufigen Baumdiagramm und den Pfadregeln. Mit Hilfe der Pfadregeln können Wahrscheinlichkeiten in einem mehrstufigen Zufallsexperiment, welches durch ein Baumdiagramm dargestellt ist, berechnet werden. Insgesamt gibt es zwei Pfadregeln, die Produktregel zur Berechnung zusammengesetzter Ereignisse eines mehrstufigen Vorgangs und die Summenregel zur Berechnung eines Ereignisses, welches sich aus den Ereignissen mehrerer Pfade zusammensetzt (vgl. Krüger et al., 2015, S. 135, 138):

1. Pfadregel (Produktregel):

Die Wahrscheinlichkeit eines zusammengesetzten Ereignisses in einem mehrstufigen Zufallsexperiment ist gleich dem Produkt der Wahrscheinlichkeiten längs des Pfades, der zu diesem Ereignis führt.

2. Pfadregel (Summenregel):

Die Wahrscheinlichkeit eines Ereignisses in einem mehrstufigen Zufallsexperiment ist gleich der Summe der Wahrscheinlichkeiten der für dieses Ereignis günstigen Pfade.

Die günstigen Pfade zur Berechnung der theoretischen Trefferquote $P(R)$ sind in Abbildung 11 in grün dargestellt. Mit der Produkt- und Summenregel ergibt sich für die theoretische Trefferwahrscheinlichkeit $P(R)$ dann ebenfalls der Wert

$$P(R) = 0.42^2 + 0.58^2 = 0.1764 + 0.3363 = 0.5128 \approx 0.51.$$

4. Didaktisch-methodisches Konzept

Dieses Kapitel dient der Vorstellung des didaktisch-methodischen Konzepts des Lernmoduls. Es werden ein Überblick über die Ziele und die Struktur des CAMMP days gegeben sowie der Aufbau und die didaktischen Vorüberlegungen zum entwickelten Material erläutert.

4.1. Ziele und curriculare Einbindung des entwickelten Lernmoduls

Die Ziele des entwickelten CAMMP days stützen sich größtenteils auf die im Kernlehrplan Mathematik der Sekundarstufe I für das Land Nordrhein-Westfalen verankerten Kompetenzen und Ziele sowie auf Winters (1995) angestrebten Grunderfahrungen für den Mathematikunterricht. Zunächst werden die Ziele des Lernmoduls stichpunktartig aufgelistet, bevor die Parallelen zu Winters (1995) Grunderfahrungen und zum Kernlehrplan aufgezeigt und noch weitere, zu vermittelnde Kompetenzen genannt werden:

- Die SuS sollen anhand des im Lernmodul durchgenommenen Beispiels die Relevanz der Mathematik und des Modellierens in der heutigen Welt erkennen.
- Die SuS sollen über ihren Umgang mit sozialen Netzwerken reflektieren.
- Die SuS sollen eine Realsituation in ein mathematisches Modell übersetzen können.
- Die SuS sollen erklären können, wie mit einfachen mathematischen Methoden nicht angegebene Informationen eines Nutzers in einem sozialen Netzwerk vorhergesagt werden können.
- Die SuS sollen Lagemaße und Streuungsmaße zur Vorhersage des Alters einer Person nutzen können.

Ein wichtiges Ziel des Lernmoduls ist es, den SuS die Relevanz der Mathematik in der heutigen Welt und vor allem in ihrem alltäglichen Leben zu zeigen. Auch Winter sieht die Fähigkeit, die Rolle zu erkennen, welche die Mathematik in der Welt spielt, als einen wichtigen Aspekt der mathematischen Grundausbildung (vgl. Winter, 1995, S. 37). Die SuS bearbeiten dazu ein reales Problem, welches sich an ihrer Lebenswelt orientiert. Denn nach der JIM-Studie (2017) sind in der heutigen Zeit soziale Netzwerke längst ein fester Bestandteil im Leben der Jugendlichen (vgl. Feierabend et al., 2017, S. 35-36). Das erarbeitete Modul soll den SuS einen Überblick geben, wie mit Hilfe einfacher, mathematischer Regeln Vorhersagen über persönliche Informationen eines Nutzers erstellt werden können, die dieser nicht preisgegeben hat. Damit unterstützt der CAMMP day auch eine der drei genannten Grunderfahrungen von Winter: Mit Hilfe der Mathematik sollen die SuS Erscheinungen aus Natur, Gesellschaft und Kultur wahrnehmen und verstehen (vgl. Winter, 1995, S. 37). Am Ende des Modellierungstages sollen die SuS die Fragestellung „Wie gut kann man das Alter eines Nutzers in einem sozialen Netzwerk vorhersagen, obwohl dieses aus den Angaben in seinem Profil

nicht ablesbar ist?“ beantworten und das Vorgehen zur Beantwortung der Frage erklären können. Auch soll den SuS klar werden, dass ihre Schulmathematik bereits für eine grobe Antwort dieser komplexen Fragestellung ausreicht und eine Verwendung im Alltag findet.

Der mathematische Hintergrund richtet sich an SuS ab der siebten Klasse. Das entwickelte Lernmodul bietet den SuS sowohl die Möglichkeit, bereits gelernte Inhalte zu wiederholen und zu vertiefen, als auch noch unbekannte Inhalte zu lernen. Die Mathematik des Moduls greift auf die Bestimmung von relativen und absoluten Häufigkeiten und auf das Interpretieren von statistischen Darstellungen wie beispielsweise Säulendiagrammen zurück, was laut Kernlehrplan in der sechsten Klasse durchgenommen wird. Auch die Berechnung von Median und arithmetischem Mittel sollte bereits in der sechsten Klasse behandelt worden sein (vgl. KLP Mathematik Sek. I, 2007, S. 23). Der Modus wird im Kernlehrplan nicht erwähnt. Diesen sollen die SuS am Ende des CAMMP days als ein weiteres Lagemaß anwenden können. Auch greift das entwickelte Lernmodul die Kompetenzen im Bereich Stochastik auf, die nach dem Kernlehrplan bis zum Ende der achten Klasse erworben sein sollten. So sollen die SuS den Median, die Spannweite und das untere sowie obere Quartil zur Darstellung von Häufigkeitsverteilungen als Boxplot nutzen können und die Spannweite sowie den Quartilsabstand in statistischen Darstellungen interpretieren. Zudem sollen die SuS aus einer achten oder höheren Klasse Pfadregeln zur Bestimmung von Wahrscheinlichkeiten bei zweistufigen Zufallsexperimenten nutzen (vgl. KLP Mathematik Sek. I, 2007, S. 28).

Durch das Durchlaufen des Modellierungskreislaufs sollen die SuS das Konzept der mathematischen Modellierung kennenlernen und ihre Modellierungskompetenz weiter aufbauen. Hier soll ein weiterer Beitrag zur mathematischen Grundausbildung geleistet werden. Winter versteht darunter unter anderem die Fähigkeit Modelle zu bilden und zu nutzen, um reale, komplexe Probleme mathematisch zu beschreiben (vgl. Winter, 1995, S. 38). Auch der Kernlehrplan zählt das Modellieren als eine der vier prozessbezogenen Kompetenzen auf (vgl. KLP Mathematik Sek. I, 2007, S. 11). Im Laufe des Modellierungstages lernen die SuS eine Realsituation in ein mathematisches Modell zu übersetzen und zu mathematisieren. Sie wenden ihr mathematisches Wissen auf dieses Modell an und setzen die aus dem mathematischen Modell gewonnenen Lösungen am Ende in Bezug zur realen Situation. Diese Kompetenzen sollen die SuS auch laut Kernlehrplan bis zum Abschluss der achten Klasse erreicht haben (vgl. KLP Mathematik Sek. I, 2007, S. 26).

Das erarbeitete Modul soll die SuS ebenfalls für das Thema Sicherheit der Privatsphäre in sozialen Netzwerken sensibilisieren und die folgenden an Grimm, Keber und Zöllner angelegten Kompetenzen zum Thema Privatsphäre weitergeben (vgl. Grimm et al., 2015, S.11-12):

- Ethische Kompetenz:
Die SuS sollen über die Wichtigkeit zum Schutz der Privatsphäre reflektieren können.
- Strukturelle Kompetenz:
Die SuS können erklären, wer private Daten zu welchem Zweck erhebt und ggf.

auch weiter gibt. Ebenso werden sie dafür sensibilisiert, dass in der heutigen Zeit auch private Informationen, welche sie nicht veröffentlicht haben, herausgefunden werden können.

- Risikokompetenz:
Die SuS können die Folgen, welche die Preisgabe von privaten Informationen nach sich ziehen, abschätzen.
- Handlungskompetenz
Die SuS können Möglichkeiten zum Schutz ihrer Privatsphäre ergreifen.

4.2. Gemeinsamkeiten und Unterschiede zum bestehenden Lernmodul Sicherheit der Privatsphäre in sozialen Netzwerken für die Sek. II

Das entwickelte Lernmodul baut auf dem bereits bestehenden CAMMP day von Sube (Sube, 2016) auf und befasst sich ebenfalls mit der Fragestellung, wie mit Hilfe einfacher, mathematischer Methoden Vorhersagen über Informationen von Nutzern sozialer Netzwerke getroffen werden können, welche diese in ihren Profilen nicht angegeben haben. Während sich Sube auf die Vorhersage der sexuellen Orientierung fokussiert, wird in dem hier dargestellten Lernmodul eine Vorhersage über das Alter der Nutzer getroffen. Wie auch bei Sube werden den SuS dazu verschiedene Vorhersageregeln vorgestellt, welche sie durch die Anwendung auf ihren Datensatz auf ihre Effektivität überprüfen sollen. Genau wie bei Sube greift auch dieses Lernmodul auf reale Nutzerprofile aus dem sozialen Netzwerk Friendster zurück. Als Güte für die Vorhersageregeln benutzt Sube das Cohen's Kappa, welches einen Bezug zur relativen Häufigkeit der richtigen Kategorisierungen und der Häufigkeit an Übereinstimmungen durch eine zufällig getroffene Kategorisierung angibt. Das Ziel besteht in der Überprüfung, ob die aufgestellten Vorhersageregeln besser sind als der Zufall, also einer zufälligen Zuordnung der Nutzer in eine sexuelle Orientierung (vgl. Sube, 2016, S. 40). In dem entwickelten Lernmodul für die Mittelstufe wird dagegen die Trefferquote, also der Anteil an richtigen Vorhersagen, als Gütekriterium verwendet.

Anders als bei Subes CAMMP day sollen sich die SuS in diesem Lernmodul zunächst selber mit den Daten aus Friendster auseinandersetzen und bezogen auf die zu untersuchende Fragestellung aufbereiten. Das heißt, die SuS sollen sich im Rahmen des ersten Arbeitsblatts eine Übersicht über den Datensatz verschaffen und beispielsweise Nutzer mit vermeintlich falschen Angaben aus diesem entfernen. Damit wurde die dritte Phase des vorgestellten PPDAC Zyklus (Abbildung 4) in das Lernmodul integriert, in der sich die SuS intensiv mit den Daten auseinandersetzen sollen.

Das übergeordnete Thema Sicherheit der Privatsphäre in sozialen Netzwerken wird in diesem Lernmodul genau wie bei Sube mit den SuS diskutiert. Auch in diesem Lernmodul sollen die SuS reflektieren, was für sie Sicherheit in sozialen Netzwerken bedeutet, welche Vor- und Nachteile sie in dem Sammeln von Nutzerdaten sehen und welche Daten sie über sich selber preisgeben. Ebenfalls sollen die SuS am Ende des CAMMP

days über ihren Umgang mit sozialen Netzwerken reflektieren.

Wie im bereits bestehenden CAMMP day für die Sekundarstufe II wurden auch für dieses Lernmodul zwei verschiedene Versionen entwickelt. So gibt es das *Niveau 1* für SuS aus einer siebten Klasse und das *Niveau 2* für SuS aus einer achten oder höheren Klasse. Die Materialien unterscheiden sich, wie in Abbildung 12 dargestellt, nur im dritten Aufgabenblatt. Die ersten beiden Arbeitsblätter sind für beide Niveaus identisch. Im *Niveau 2* untersuchen die SuS die im Abschnitt 3.3.2 aufgeführte Heuristik zum Zufall. In diesem Zusammenhang sollen sie auch Wahrscheinlichkeiten mit Hilfe eines zweistufigen Baumdiagramms bestimmen. Da die SuS in der siebten Klasse noch nicht mit mehrstufigen Zufallsexperimenten in Berührung gekommen sind, wurde im *Niveau 1* diese Aufgabe ersetzt. Im *Niveau 1* testen die SuS statt der Heuristik zum Zufall eine Vorhersageregeln, welche untersucht, ob das arithmetische Mittel der Freunde 1. Ordnung in einem vorgegebenen Intervall liegt. Zudem wurde im *Niveau 2* eine zweite Zusatzaufgabe, die sich ebenfalls mit dem zweistufigen Baumdiagramm befasst, hinzugefügt. Für eine bessere Übersicht wurden beide Zusatzaufgaben auf ein extra Zusatzaufgabenblatt gesetzt. Der genaue Aufbau vom dritten Aufgabenblatt ist für das erste und zweite Niveau in Abschnitt 4.5.9 dargestellt.

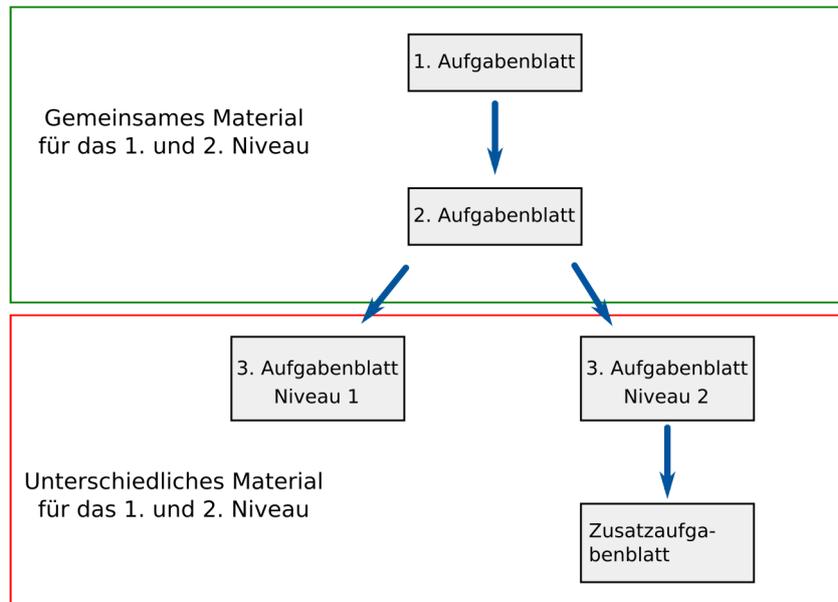


Abbildung 12: Darstellung der Gemeinsamkeiten und Unterschiede der Materialien für *Niveau 1* (für eine siebte Klasse) und *Niveau 2* (für eine achte oder höhere Klasse).

4.3. Der Modellierungskreislauf als Struktur des entwickelten Lernmoduls

Im entwickelten Lernmodul wird die zentrale Frage- bzw. Problemstellung

„Wie gut kann man das Alter eines Nutzers in einem sozialen Netzwerk vorhersagen, obwohl dieses aus den Angaben in seinem Profil nicht ablesbar ist?“

mit Hilfe des Modellierungskreislaufes gelöst:

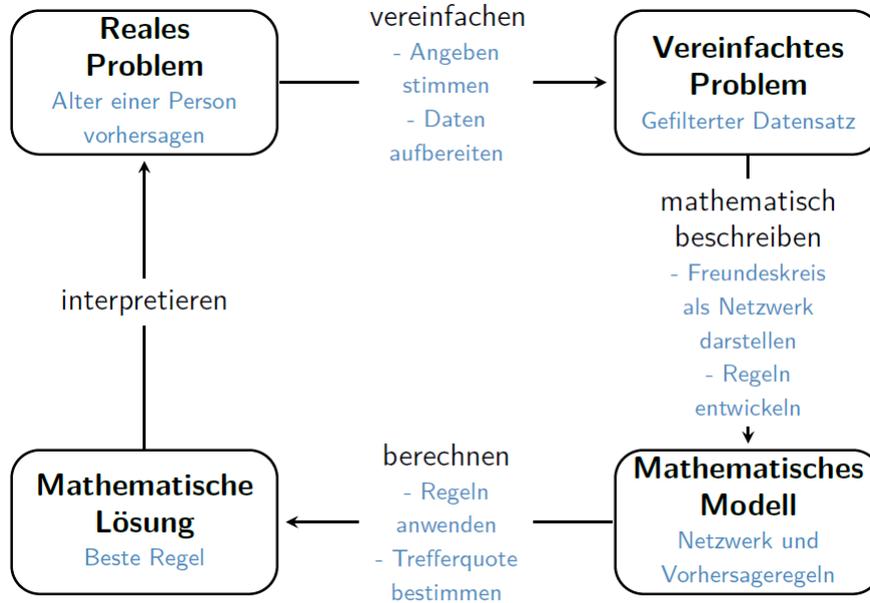


Abbildung 13: Modellierungskreislauf für das entwickelte Lernmodul

Das *reale Problem* im Lernmodul besteht darin, dass unbekanntes Alter eines Nutzers vorherzusagen. Für diese Vorhersage werden zunächst Vereinfachungen getroffen, indem ein Datensatz aus dem sozialen Netzwerk Friendster betrachtet wird und die SuS aus diesem Datensatz Nutzer mit vermeintlich falschen Angaben herausfiltern. Auch wird die *Vereinfachung* getroffen, dass die Angaben der verbleibenden Nutzer nach der Filterung korrekt sind. So erhalten die SuS einen gefilterten Datensatz, der das *vereinfachte Problem* darstellt. Dieses Problem wird anschließend durch die Darstellung des sozialen Netzwerks als Graph und das Aufstellen von Vorhersageregeln *mathematisch beschrieben*. Das *mathematische Modell* besteht dann aus dem Graph und den einzelnen Regeln. Im Schritt *berechnen* werden die Vorhersageregeln auf den gefilterten Datensatz angewendet und die Güte mit Hilfe der Trefferquote bestimmt. Das Ergebnis der Anwendung stellt die *mathematische Lösung* dar, welche die SuS in Bezug zum *realen Problem interpretieren*. Im Laufe des Lernmoduls werden die SuS durch die Betrachtung neuer Vorhersageregeln ihr *mathematisches Modell* immer wieder anpassen, um das aufgestellte Modell zu verbessern und eine für die Fragestellung noch bessere Lösung zu erhalten.

4.4. Ablauf des CAMMP days

Der CAMMP day nimmt insgesamt sechs Zeitstunden in Anspruch. In dieser Zeit sind verschiedene Vorträge der Dozenten, gemeinsame Besprechungen der Lösungen mit den SuS und Pausen enthalten. Insgesamt haben die SuS etwa vier Stunden Zeit, die konzipierten Materialien des Lernmoduls zu bearbeiten. Dabei arbeiten sie in Zweier- bis Dreier-Gruppen.

Im Folgenden wird der genaue Ablauf für den entwickelten CAMMP day vorgestellt. Eine stichpunktartige Zusammenfassung mit einem groben zeitlichen Ablauf der einzelnen Phasen für die Dozenten, welche das Lernmodul durchführen, ist als methodisches Konzept im Anhang (siehe B) zu finden.

1. Begrüßung:

Der CAMMP day startet mit einer kurzen *Begrüßung* in Form einer Präsentation, in der sich die Dozenten kurz vorstellen und den SuS das Konzept des Schülerlabors CAMMP näher bringen. Auch wird den SuS der Ablauf des Tages transparent gemacht.

2. Modellierungsvortrag:

Im Anschluss an die Begrüßung erfolgt der etwa 15 minütige *Modellierungsvortrag* eines Doktoranden, welcher den SuS die Idee und das Konzept hinter der mathematischen Modellierung erklärt. Mit Hilfe eines aktuellen Forschungsgebiets, wie beispielsweise der Strahlentherapie, präsentiert der Doktorand den SuS die einzelnen Schritte des vorgestellten CAMMP Modellierungskreislaufes aus Abschnitt 2.2.2.

3. Einführungsvortrag:

Im *Einführungsvortrag* (siehe Abschnitt 4.5.2) wird das Thema „Sicherheit der Privatsphäre in sozialen Netzwerken - Was Facebook, Instagram und Co. über uns wissen“ des CAMMP days vorgestellt. Mit den SuS wird über die Thematik Sicherheit und über die Möglichkeit persönliche Informationen von Nutzern zu generieren, die sie nicht mit der Öffentlichkeit teilen wollten, diskutiert. Auch werden den SuS Gründe aufgezeigt, warum Betreiber von sozialen Netzwerken überhaupt an persönlichen Informationen der Nutzer interessiert sind. Im Anschluss wird den SuS die Fragestellung „Wie gut kann man das Alter eines Nutzers in einem sozialen Netzwerk vorhersagen, obwohl dieses aus den Angaben in seinem Profil nicht ablesbar ist?“ vorgestellt, welche sie im Laufe des Tages beantworten sollen. Der Einführungsvortrag wird etwa nach der Hälfte der Folien unterbrochen und führt in die *erste Arbeitsphase*. Bevor die SuS jedoch mit der ersten Arbeitsphase starten, erhalten sie eine kurze *Einführung in das Programm MATLAB* (siehe Abschnitt 4.5.3), mit dem sie während des Tages arbeiten.

4. Erste Arbeitsphase:

In der *ersten Arbeitsphase* (siehe Abschnitt 4.5.4) beschäftigen sich die SuS mit der Aufbereitung des Datensatzes bestehend aus den 56284 Nutzern. Sie verschaffen sich einen Überblick über den Datensatz und nehmen zwei Filterungen vor.

Zum einen filtern sie selbstständig Nutzer mit einem vermeintlich falschen Alter und zum anderen nehmen sie Personen aus dem Datensatz heraus, welche nach der ersten Filterung keine Freunde 1. oder 2. Ordnung mehr besitzen. Auf Grund der ersten Filterung, in der die SuS die Altersgrenzen selber festlegen können, erhalten die einzelnen Gruppen am Ende der Arbeitsphase ihren persönlichen Datensatz, mit dem sie im weiteren Verlauf des CAMMP days arbeiten werden. Schnelle SuS können sich bereits Gedanken über mögliche Vorhersageregeln zur Bestimmung des Alters einer Person machen.

5. Erste Sicherungsphase:

Es folgt die *erste Sicherung* (siehe Abschnitt 4.5.5), in der mit den SuS über den Ausgangsdatensatz sowie über die beiden vorgenommenen Filterungen diskutiert wird. Vor allem bei der ersten Filterung sollen die Gruppen ihre gewählten Altersgrenzen begründet erörtern. Aber auch das Prinzip hinter der zweiten Filterung wird mit den SuS besprochen. Direkt im Anschluss fährt der Dozent mit dem zweiten Teil des *Einführungsvortrags* fort. Hier werden den SuS nun die ersten Vorhersageregeln zur Bestimmung des exakten Alters vorgestellt, die sie in der *zweiten Arbeitsphase* anwenden werden. Ausgehend von der Tatsache, dass die einzelnen Gruppen während der ersten Arbeitsphase unterschiedlich große Datensätze erzeugt haben, wird gemeinsam mit den SuS die Trefferquote, als ein geeignetes Gütekriterium zum Vergleich der Ergebnisse untereinander eingeführt. Danach beginnt für die SuS die *zweite Arbeitsphase*.

6. Zweite Arbeitsphase:

Nach einer kurzen *Pause* wenden die SuS in der *zweiten Arbeitsphase* (siehe Abschnitt 4.5.7) die Vorhersageregeln zu den drei Lagemaße arithmetisches Mittel, Median und Modus auf ihren Datensatz an. Dabei dürfen sie in ihrer Gruppe entscheiden, ob sie die Regeln für die Freunde 1. oder 2. Ordnung betrachten möchten. Sie bestimmen die Trefferquoten der drei Vorhersageregeln und interpretieren die Ergebnisse. Im nächsten Schritt lernen die SuS mit Hilfe eines Boxplots die Spannweite und den Quartilsabstand als zwei Streuungsmaße kennen. Zu diesen beiden Streuungsmaßen werden ebenfalls Vorhersageregeln, welche das Alter eines Nutzer innerhalb eines Intervalls vorhersagen sollen, aufgestellt und auf den Datensatz angewendet. Schnelle SuS lernen mit der Standardabweichung ein drittes Streuungsmaß kennen, zu dem ebenfalls eine Vorhersageregeln überprüft werden soll.

7. Zweite Sicherungsphase:

Im Rahmen der *zweiten Sicherungsphase* (siehe Abschnitt 4.5.8) werden die Trefferquoten der einzelnen Vorhersageregeln auf einer vorbereiteten Folie gesammelt und mit den SuS interpretiert. Das Erstellen eines Boxplots wird mit den SuS besprochen und auf der Folie eingezeichnet. Auch wird mit den SuS über das Thema Robustheit gegenüber Ausreißer von arithmetischen Mittel, Median, Spannweite und Quartilsabstand diskutiert.

Die Besprechung der *zweiten Arbeitsphase* soll noch vor der einstündigen *Mittagspause* erfolgen. Sie kann jedoch auch nach der Besprechung der Vorhersageregeln zu den drei Lagemaßen unterbrochen und nach der *Mittagspause* fortgeführt werden.

8. **Dritte Arbeitsphase:**

Die *dritte Arbeitsphase* unterscheidet sich wie bereits angesprochen je nachdem ob SuS aus einer siebten Klasse oder SuS aus einer achten oder höheren Klasse den entwickelten CAMMP day besuchen. Aus diesem Grund wurden zwei verschiedene Arbeitsblätter entworfen. In beiden Varianten versetzen sich die SuS in die Rolle eines Betreibers eines sozialen Netzwerks, der personalisierte Werbung an die Personengruppe mit einem Alter zwischen 25 und 30 Jahren verschicken möchte.

Auf dem Arbeitsblatt für die siebte Klasse (siehe Abschnitt 4.5.9) lernen die SuS zunächst eine Vorhersageregeln kennen, welche das arithmetische Mittel der Freunde 1. Ordnung bestimmt und dem Nutzer ein Alter zwischen 25 und 30 Jahren zuordnet, wenn das arithmetische Mittel innerhalb des Intervalls $[25, 30]$ liegt. Als eine weitere Regel zur Vorhersage der richtigen Altersgruppe betrachten die SuS den Anteil der Freunde 1. Ordnung eines Nutzers, deren Alter innerhalb des Intervalls $[25, 30]$ liegt. Schnelle SuS können im Rahmen einer Zusatzaufgabe mit den bereits kennengelernten Vorhersageregeln das tatsächliche Alter der aussortierten Nutzer mit einem angegebenen Alter von null Jahren bestimmen. Das Arbeitsblatt für die achte oder höhere Klasse (siehe Abschnitt 4.5.9) unterscheidet sich nur in der ersten Vorhersageregeln und durch eine weitere Zusatzaufgabe vom Arbeitsblatt für die siebte Klasse. So beschäftigt sich die erste Vorhersageregeln mit der im Abschnitt 3.3.2 vorgestellten Zufallsregeln, bei der mit der Wahrscheinlichkeit q einem Nutzer ein Alter zwischen 25 und 30 Jahre zugeordnet wird. Hier sollen die SuS die Trefferwahrscheinlichkeit für eine richtige Zuordnung in die beiden Altersgruppen sowohl experimentell als auch theoretisch bestimmen. In der zweiten Zusatzaufgabe sollen die SuS überlegen, mit welcher Wahl von q sie die beste theoretische Trefferquote erhalten würden.

9. **Dritte Sicherungsphase:**

In der *dritten Sicherungsphase* (siehe Kapitel 4.5.10) werden wieder die Ergebnisse zu den beiden Vorhersageregeln gesammelt und interpretiert. Außerdem wird für das zweite Niveau detailliert auf die Erstellung eines zweistufigen Baumdiagramms eingegangen und die anschließende Berechnung der Wahrscheinlichkeiten mit den SuS Schritt für Schritt besprochen.

10. **Tagesabschluss:**

Der *Tagesabschluss* beginnt mit der *Abschlusspräsentation* (siehe Abschnitt 4.5.11), in der die SuS kurz die Ergebnisse des Tages zusammenfassen. Auch wird hier das Thema Sicherheit der Privatsphäre in sozialen Netzwerken nochmal aufgegriffen, über das die SuS erneut diskutieren. Im Anschluss findet die *Verabschiedung* der

SuS statt, bei der sie auf die weiteren Angebote von CAMMP hingewiesen werden. Zum Schluss erfolgt die *Evaluation*, um einerseits den SuS die Möglichkeit zum Feedback zu geben und andererseits dem CAMMP Team mögliche Verbesserungen im durchgeführten Lernmodul aufzuzeigen.

4.5. Vorstellung der Materialien

Hauptbestandteil der im Rahmen des Lernmoduls entwickelten Materialien stellen zwei Präsentationen und drei Arbeitsblätter mit zugehörigen Antwortblättern dar. Die Einführungspräsentation dient dabei zur Motivation der SuS und zur Einführung in die Problemstellung. Am Ende des CAMMP days regt die Abschlusspräsentation die SuS dazu an, über die im Laufe des Tages gewonnenen Erkenntnisse zu diskutieren. Eine Besonderheit dieses entworfenen Lernmodul stellt die Realisierung der Arbeitsblätter im MATLAB Live Skript dar.

Das MATLAB Live Skript (siehe Abbildung 14) ist ein Arbeitsblatt, welches die Arbeitsaufträge in Form von Text, Bildern oder Gleichungen mit dem MATLAB Code kombiniert. Solche Arbeitsblätter sind für die SuS interaktiv. Zum einen müssen sie bei der Bearbeitung der Aufgaben nicht ständig zwischen Aufgabenblatt und Computer wechseln und zum anderen erhalten sie die Ausgabe der Computersoftware direkt unterhalb der bearbeiteten Aufgabe (siehe Abbildung 14). Bei klassischen MATLAB-Dateien erscheint die Rückmeldung dagegen nicht aufgabenbezogen im Command Window oder bei Graphiken jeweils in einem neuen Fenster. Dadurch fällt es den SuS oft schwer, die Übersicht zu behalten und den einzelnen Ausgaben der Computersoftware die zugehörigen Aufgaben zuzuordnen. Ein weiterer Vorteil des Live Skripts liegt in der Einsparung von Papier, welches ökonomischer ist und dabei hilft, die Ordnung am Arbeitsplatz der SuS zu bewahren.

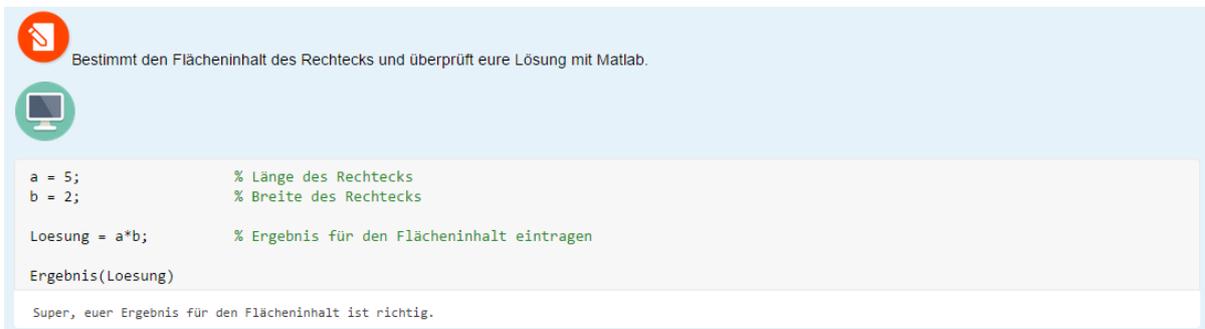


Abbildung 14: Beispielhafter Auszug aus dem MATLAB Live Skript, welches die SuS im entwickelten Lernmodul eine Einführung in den Umgang mit dem Programm MATLAB gibt.

Zur Sicherung der zentralen Ergebnisse wurde für jedes Arbeitsblatt ein Antwortblatt erstellt, auf dem die SuS ihre Antworten eintragen. Einerseits können die SuS so während den Sicherungsphasen den Laptop schließen, um ihre Aufmerksamkeit auf die

Besprechung zu lenken. Andererseits helfen sie die SuS dabei ihre Ergebnisse zu strukturieren. Zudem können sie auch die wichtigsten Resultate aus dem CAMMP day mit nach Hause nehmen. Außerdem wurden für ausgewählte Aufgaben Hilfekarten als MATLAB Live Skript erstellt, um das unterschiedliche Leistungsniveau der SuS zu berücksichtigen sowie das eigenständige Lernen zu fördern (vgl. Meyer, 2004, S. 104). Für die Dozenten, welche das entwickelte Lernmodul durchführen, wurde zu jedem Arbeitsblatt eine Musterlösung und eine Folie zur Besprechung der Ergebnisse während den Sicherungsphasen erstellt. Die Folien orientieren sich dabei an den Antwortblättern der SuS.

Im Anhang sind alle erstellten Materialien zu finden. Im Folgenden werden die entwickelten Materialien und die dazugehörigen didaktischen Überlegungen, in der Reihenfolge, wie sie auch im CAMMP day zum Einsatz kommen, einzeln vorgestellt. Ebenso wird darauf eingegangen, auf welchen Schritt im Modellierungskreislauf sich einzelne Aufgaben beziehen.

4.5.1. Modellierungsvortrag

Der Modellierungsvortrag wurde nicht im Rahmen dieser Arbeit entworfen, stellt aber einen wichtigen Bestandteil für den CAMMP day dar und wird deshalb hier kurz vorgestellt. Der Vortrag wird von einem Doktoranden der RWTH Aachen, welcher im CAMMP Projekt involviert ist, gehalten und dient der Vorstellung des mathematischen Modellierungskreislaufes (siehe Abbildung 2). Mit Hilfe des Kreislaufs erläutert der Doktorand bzw. die Doktorandin die einzelnen Schritte des Modellierungsprozesses und kombiniert diese Schritte mit einem interessanten Thema aus der eigenen Forschung. So sehen die SuS gleich, dass Modellieren eine zentrale Rolle in der heutigen Zeit spielt. Auch stellt der Doktorand bzw. die Doktorandin zu Beginn des Vortrags kurz seinen bzw. ihren Werdegang vor, um den SuS einen möglichen Weg nach ihrer Schullaufbahn aufzuzeigen.

4.5.2. Einführungsvortrag

Duit und Wodzinski ordnen dem Einstieg in einer Lernsequenz die Funktionen zu, die SuS neugierig zu machen sowie das Interesse am Thema zu wecken, an Vorwissen, Alltagserfahrungen sowie Schülervorstellungen anzuknüpfen und über das Kommende zu informieren (vgl. Duit & Wodzinski, 2006, S. 9). Ausgehend von diesen Punkten wird den SuS im Einführungsvortrag (siehe Anhang C.1) die Problemstellung vorgestellt, mit der sie sich während des CAMMP days auseinandersetzen. Um das Interesse der SuS zu wecken und ihnen die Relevanz des Themas Sicherheit der Privatsphäre in sozialen Netzwerken aufzuzeigen, wird sich zunächst ein Überblick darüber verschafft, wer von den anwesenden SuS ein soziales Netzwerk nutzt. Durch diesen Einstieg sollen bereits zu Beginn möglichst alle SuS aktiviert werden. Anschließend werden die aktuellen Nutzerzahlen bekannter sozialer Netzwerke vorgestellt. Die Alltagserfahrungen und Schülervorstellungen werden im Anschluss in einer Diskussion mit eingebunden. Die SuS sollen darüber diskutieren, welche Informationen sie in einem sozialen Netz-

werk preisgeben und wie sie die Sicherheit ihrer Privatsphäre in solchen Netzwerken einschätzen. Sie werden darauf aufmerksam gemacht, dass ihre Daten von den Betreibern dieser Netzwerke analysiert werden und sollen gemeinsam im Plenum überlegen, welche Informationen von besonderem Interesse sind. Nach der Diskussion wird den SuS der Begriff des Schattenprofils vorgestellt. Um einen Alltagsbezug herzustellen, wird den SuS anhand von drei Beispielen erläutert, wie einfach Informationen, die Nutzer nicht freiwillig angegeben haben, erzeugt werden können. Auch wird den SuS beispielhaft der finanzielle Wert ihrer Informationen genannt. So sollen diese und auch die kommenden Informationen eine emotionale Bedeutung erhalten und nicht gleich wieder aus dem Arbeitsspeicher verschwinden, wodurch sich die Wahrscheinlichkeit für eine Überführung in das Langzeitgedächtnis erhöht (vgl. Heymann, 1998, S. 7). Es folgt nun die Vorstellung der zentrale Fragestellung „Wie gut kann man das Alter eines Nutzers in einem sozialen Netzwerk vorhersagen, obwohl dieses aus den Angaben in seinem Profil nicht ablesbar ist?“, welche die SuS im Laufe des Tages beantworten sollen. Diese Fragestellung stellt im vorgestellten Modellierungskreis von Abbildung 2 das reale Problem dar. Im Bezug zum PPDAC Zyklus (siehe Abbildung 4) befinden sich die SuS in der Phase des Problems, in der die Problemstellung genau definiert wird. Die SuS werden hier wieder mit in die Präsentation eingebunden, indem sie über die Relevanz, das Alter einer Person zu kennen, diskutieren. Im Anschluss wird die Problemstellung des Lernmoduls anhand des Modellierungskreislaufs durchgegangen. Das reale Problem, das Alter einer Person in einem sozialen Netzwerk vorherzusagen, wird durch die Betrachtung eines kleinen Datensatzes bestehend aus realen Nutzerprofilen des sozialen Netzwerks Friendster vereinfacht. Das heißt, die Betrachtung eines ausgewählten Datensatzes stellt im Modellierungskreislauf das vereinfachte Modell dar. Das Arbeiten mit realen Nutzerprofilen wirkt für die SuS einerseits authentisch und zeigt ihnen andererseits, wie einfach es ist, an öffentliche Informationen von Personen zu gelangen.

Den SuS wird nun mitgeteilt, dass die Daten zunächst bezogen auf die Fragestellung gefiltert werden müssen. An dieser Stelle wird der Einführungsvortrag unterbrochen und die erste Arbeitsphase gestartet. Dort bearbeiten die SuS das erste Arbeitsblatt mit dem Titel „Daten filtern“, welches im Abschnitt 4.5.4 vorgestellt wird. Die Unterbrechung des Vortrags soll den positiven Nebeneffekt bewirken, dass die Konzentration und Aufnahmefähigkeit der SuS nicht nachlässt. Denn dies könnte dazu führen, dass sie die präsentierten Informationen nicht mehr verarbeiten können. Ein Wechsel der Lernform in eine aktive Phase kann damit einer Ermüdung der Konzentrationsfähigkeit vorbeugen, sodass die SuS im zweiten Teil der Präsentation wieder ihre volle Aufmerksamkeit auf die dort vermittelten Informationen lenken können (vgl. Gudjons, 2005, S. 14).

4.5.3. Einführung in MATLAB

Die SuS kommen in den meisten Fällen während eines CAMMP days erstmalig mit der Computersoftware MATLAB in Berührung. Um den SuS den Einstieg in die Arbeit mit diesem Programm zu erleichtern, wird es vor der ersten Benutzung, also vor

der Bearbeitung des ersten Arbeitsblatts, kurz vorgestellt. Durch diese Einführung soll sichergestellt werden, dass die SuS die nötigen Voraussetzungen zur Bearbeitung der kommenden Arbeitsaufträge besitzen und durch die Verwendung von MATLAB nicht überfordert sind. Denn eine Überforderung führt zu einer Hemmung der Motivation und wirkt sich negativ auf den Lernprozess aus (vgl. Deci & Ryan, 1993, S. 231).

Für die Einführung in den Umgang mit MATLAB öffnet der Dozent das Live Skript *Einfuehrung.mlx* (siehe Anhang D). Zunächst wird den SuS erklärt, dass sie sich MATLAB wie einen „großen Taschenrechner“ vorstellen können. Anschließend erfolgt eine Vorstellung der Oberfläche des Programms. Für die SuS sind dabei zwei Bereiche interessant. Zum einen der *Current Folder-Bereich* auf der linken Seite. Hier werden die für den CAMMP day benötigten Dateien und Ordner angezeigt. Dort können die SuS die benötigten Arbeitsblätter und auch Hilfekarten auswählen. Zum anderen das *Live Skript* bzw. der *Live Editor* in der Mitte, welcher auch den größten Platz in MATLAB einnimmt. In diesem Bereich können die SuS die Arbeitsblätter lesen sowie interaktiv bearbeiten. Der *Workspace-Bereich* und das *Command Window* werden für das Lernmodul nicht benötigt und minimiert.

Nachdem die Oberfläche von MATLAB vorgestellt wurde, wird mit den SuS eine beispielhafte Aufgabe durchgegangen, damit ihnen der spätere Aufbau der Arbeitsblätter klar wird. Zunächst erfolgt in Form eines Textes und eventuell eines Bildes die Aufgabenstellung. In dem Fall der Einführung handelt die Aufgabe über die Bestimmung des Flächeninhaltes eines Rechtecks mit gegebenen Seitenlängen a und b . Unterhalb der Aufgabenstellung befindet sich, durch einen gräulichen Kasten hervorgehoben, der Bereich, in dem der Code steht. Zur besseren Abgrenzung gibt es ein Icon in Form eines Notizblocks, welches die SuS auf einen Arbeitsauftrag aufmerksam macht, und ein Icon in Form eines Computerbildschirms, das den Code-Bereich anzeigt. Im Laufe des CAMMP days finden die SuS auch ein Icon, welches eine Person mit einem Fragezeichen zeigt und angibt, dass zu dieser Aufgabe eine Hilfekarte bereitsteht. Die drei Icons sind in Abbildung 15 dargestellt. Im Kasten, in dem der Code steht, finden die SuS an bestimmten Stellen den Eintrag NaN, die Abkürzung für „Not a Number“. Für den Eintrag NaN tragen die SuS ihre Lösung ein. Zum besseren Verständnis wird in grün die Bedeutung zu jeder für die SuS relevanten Zeile im Code beschrieben. Die Realisierung des Codes für die Einführung in MATLAB ist in der bereits vorgestellten Abbildung 14 dargestellt.

Zum Überprüfen der Eingabe klicken die SuS anschließend auf das Symbol *Run Section*, welches sich an der oberen Leiste von MATLAB befindet. Ist das Ergebnis des Flächeninhalts korrekt, erhalten die SuS auch diese Rückmeldung von MATLAB genau unterhalb des grauen Kastens mit dem Code. Ebenso verhält es sich bei einer falschen Eingabe.



Abbildung 15: Die drei zentralen, im MATLAB Live Skript verwendeten Icons. *Links:* Icon, welches den SuS einen Arbeitsauftrag signalisiert. *Mitte:* Icon, das vor dem grauen Kasten mit dem enthaltenen Code steht. *Rechts:* Icon, welches die Existenz einer Hilfekarte für die aktuelle Aufgabe aufzeigt.

4.5.4. Arbeitsblatt 1 | Daten filtern

Nach der Einführung in die Computersoftware MATLAB steigen die SuS in das erste Arbeitsblatt ein. Wie von Wild und Pfannkuch vorgeschlagen sollen sich die SuS zunächst im ersten Arbeitsblatt (siehe Anhang E) einen Überblick über den Datensatz aus 56284 Nutzerprofilen verschaffen und diesen bezogen auf die Fragestellung bereinigen, bevor Analysen durchgeführt werden (vgl. Wild & Pfannkuch, 1999, S. 225). Nachdem die SuS den Datensatz in das Programm MATLAB geladen haben, lassen sie sich zunächst die Altersverteilung in Form eines Säulendiagramms in dem Programm anzeigen. Dieses Diagramm zeigt für jedes Alter zwischen 0 und 100 Jahren die absoluten Häufigkeiten der Nutzer im Datensatz, wie es bereits in Abbildung 7 dargestellt wurde. Die SuS sollen das Säulendiagramm ausgehend von Leitfragen beschreiben und Nutzer mit einem auffälligen Alter identifizieren. Anschließend können sie begründet eine untere und obere Altersgrenze festlegen, um entsprechende Nutzer unter- bzw. oberhalb dieser Grenze aus dem Datensatz herauszunehmen. Personen, die ein Alter von 0 Jahren angegeben haben, müssen von allen SuS aus dem Datensatz gefiltert werden. Ob die SuS jedoch Altersangaben um die 80 oder 90 Jahre für realistisch halten, dürfen sie selber entscheiden. Die Begründung ihrer Altersgrenzen halten die SuS auf dem zugehörigen, ersten Antwortblatt (siehe Anhang E.1.1) fest. Durch diese freie Wahl bekommen die SuS bei der Erstellung des Datensatzes die Möglichkeit zur Mitbestimmung und erhalten so ihren persönlichen Datensatz, mit dem sie im weiteren Verlauf des CAMMP days arbeiten werden. Diese Mitbestimmung kann nach Deci und Ryan die Motivation der SuS steigern (vgl. Deci & Ryan, 1993, S. 233).

Unabhängig von den getroffenen Altersgrenzen müssen die SuS im Rahmen einer zweiten Filterung Nutzer mit einer Anzahl von null Freunden 1. Ordnung und null Freunden 2. Ordnung aus dem Datensatz nehmen. Dazu betrachten die SuS zwei weitere Säulendiagramme. Das erste Säulendiagramm zeigt ihnen die absolute Häufigkeit der Nutzer in ihrem Datensatz mit einer gewissen Anzahl an Freunden 1. Ordnung und das zweite Diagramm berücksichtigt die Freunde 2. Ordnung. Mit Hilfe dieser Säulendiagramme sollen die SuS erkennen, dass es Nutzer gibt, die null Freunde 1. oder 2. Ordnung besitzen und somit das Vorhersagen von Informationen erschweren. Diese Nutzer werden im Rahmen einer zweiten Filterung aus dem Datensatz herausgenommen. Dazu

müssen die SuS den Code, der hinter der zweiten Filterung steckt, mit Worten beschreiben. Um sie zu unterstützen, sind die einzelnen Schritte des Codes bereits als Sätze im Live Skript vorgegeben. Es gibt insgesamt acht Schritte, wie beispielsweise „Den aktuellen Datensatz laden.“ oder „Die Anzahl N der Nutzer bestimmen.“, welche die SuS in die richtige Reihenfolge bringen müssen. Mit dieser Aufgabe sollen die SuS den Ansatz hinter der Filterung verstehen und diese nicht nur stupide vom Programm MATLAB durchführen lassen. In der nächsten Aufgabe folgt eine Anwendung des Gelernten. Anhand eines ausgedachten kleinen Datensatzes aus sechs Nutzern sollen die SuS das kennengelernte Prinzip der Filterung anwenden. Durch die Anwendung erhalten die SuS ebenfalls einen neuen Blick auf ihr Gelerntes, wodurch sie ein tieferes Verständnis erlangen sollen. Auch bekommen sie eine Rückmeldung über ihren Lernerfolg, indem sie erkennen, ob sie das Gelernte bereits verstanden haben (vgl. Heymann, 2012, S. 9). Zum Schluss lassen sich die SuS die Informationen ihres gefilterten Datensatzes anzeigen und notieren Stichprobengröße, maximales sowie minimales Alter und die durchschnittliche Anzahl der Freunde 1. sowie 2. Ordnung auf ihrem Antwortblatt. Damit haben die SuS den dritten Schritt des PPDAC Zyklus selbstständig durchlaufen und erhalten mit ihrem individuellen bereinigten Datensatz ihr vereinfachtes Problem im Modellierungskreislauf.

Jeder Schüler und jede Schülerin bringt seine eigenen Lernvoraussetzungen mit und arbeitet in seinem persönlichen Lerntempo (vgl. Bruder & Reibold, 2010, S. 6). Dies muss bei der Gestaltung des Lernmoduls berücksichtigt werden, um die SuS individuell nach ihren Voraussetzungen zu fördern. Aus diesem Grund besitzt dieses Arbeitsblatt, genau wie die beiden Folgenden, eine Zusatzaufgabe. Auf der einen Seite haben schnelle SuS so die Möglichkeit, sich noch intensiver mit der Thematik zu beschäftigen, und auf der anderen Seite entstehen keine Lernpausen oder Langeweile, wenn bereits fertige SuS auf andere Gruppen warten müssen. In der Zusatzaufgabe des ersten Aufgabenblatts können sich SuS bereits über mögliche Vorhersageregeln zur Bestimmung des Alters eines Nutzers Gedanken machen und diese als „Wenn,...,dann...“-Aussagen auf ihrem Antwortblatt notieren. Die Zusatzaufgabe greift einerseits die vierte Phase des PPDAC Zyklus auf, die Analyse. So können die SuS beispielsweise bereits Hypothesen aufstellen, ob es effektiver ist, das exakte Alter eines Nutzers vorherzusagen oder ein Intervall zu bestimmen, in dem das Alter des Nutzers liegen soll. Andererseits beginnen die SuS mit dieser Aufgabe durch das Formulieren von möglichen Vorhersageregeln ihr vereinfachtes Modell mathematisch zu beschreiben, um, wie im Modellierungskreislauf (Abbildung 2) vorgesehen, zu einem mathematischen Modell zu gelangen.

4.5.5. Erste Sicherungsphase

Nach der Bearbeitung eines jeden Arbeitsblatts erfolgt eine Sicherungsphase. Dadurch werden die zentralen Ergebnisse regelmäßig wiederholt und zusammengefasst. Außerdem soll so sichergestellt werden, dass sich die SuS danach wieder alle auf dem gleichen Wissensstand befinden (vgl. Meyer, 2004, S. 65).

Das erste Arbeitsblatt wird mit Hilfe einer vorgefertigten Overheadfolie (siehe Anhang E.1.2) besprochen. Anhand der Graphik zur Altersverteilung (siehe Abbildung

7) diskutieren die SuS, welche Nutzer im Datensatz eine falsche Altersangabe getätigt haben könnten. Die Gründe der SuS werden auf der Folie notiert und beispielhaft drei Altersschranken zur Filterung der Nutzer mit einem Alter unterhalb der Schranke a und oberhalb der Schranke b notiert. Anschließend erfolgt eine Diskussion über die Säulendiagramme zu den Anzahlen der Freunde 1. und 2. Ordnung der Nutzer. Auch hier werden die Gründe, welche Nutzer sich eher weniger für eine Vorhersage einer Information wie des Alters eignen, notiert. Auch wird mit den SuS das Vorgehen der zweiten Filterung anhand der sechs fiktiven Nutzer besprochen. Zunächst wird eine Liste erstellt, die für jeden Nutzer eine Eins als Eintrag enthält. In einer zweiten Liste bekommen diejenigen Nutzer eine Null als Eintrag, welche null Freunde 1. oder 2. Ordnung besitzen. In dem konstruierten Beispiel sind das der zweite und der fünfte Nutzer. Die letzte Liste besteht dann nur noch aus den Nutzer IDs derjenigen Personen, die in der zweiten Liste keine Null als Eintrag besitzen. Den SuS kann in diesem Kontext erklärt werden, dass es in der Informatik und auch Mathematik eine gängige Methode ist, wahren Aussagen, wie das Vorhandensein mindestens eines Freundes oder Freundes-Freunde, mit einer Eins zu codieren. Falsche Aussagen werden dagegen mit einer Null codiert. Um den SuS zu zeigen, dass die Gruppen durch die individuelle Wahl der Altersgrenze unterschiedliche Datensätze erzeugt haben, wird für drei Gruppen eine Übersicht ihrer gefilterten Daten auf der Folie notiert.

Die Ideen zu möglichen Vorhersageregeln, welche sich schnelle SuS bereits in der Zusatzaufgabe überlegt haben, können sie im nachfolgenden Zwischenvortrag einbringen. Dort sollen die SuS in einer kurzen Diskussion zunächst eigene Vorschläge tätigen, wie das Alter einer Person vorhergesagt werden kann. Auch wird, wie im vierten Schritt des PPDAC Zyklus vorgesehen, mit den SuS Hypothesen aufgestellt, welche Regeln zu guten Vorhersagen führen. Diese Hypothesen gehen die SuS in der darauf folgenden Arbeitsphase nach.

4.5.6. Zwischenvortrag

Nach der Besprechung des ersten Arbeitsblatts wird der Vortrag aus der Einführung fortgeführt. Die SuS werden nun, wie von Duit und Wodzinski empfohlen, über das kommende Vorgehen informiert (vgl. Duit & Wodzinski, 2006, S. 9). Nachdem die SuS ihre eigenen Ideen zum Vorhersagen des Alters vorgetragen haben, werden die ersten Vorhersageregeln, welche sich auf die Lagemaße arithmetisches Mittel, Median und Modus stützen, vorgestellt und das Vorgehen zum Anwenden bzw. Überprüfen der Regeln erläutert. Anhand einer Regel wird dieses Vorgehen für einen fiktiven Nutzer mit den SuS besprochen. Dadurch soll sichergestellt werden, dass sie das Vorgehen für das kommende zweite Arbeitsblatt verstanden haben und ihnen bewusst ist, was von ihnen verlangt wird. Denn ohne ein Verständnis für das Vorgehen ist eine erfolgreiche Arbeitsphase nicht möglich (vgl. Serve, 2011, S. 235). Es erfolgt erneut ein Bezug zum Modellierungskreislauf. Die Vorhersageregeln und die Darstellung des sozialen Netzwerks als Graph, in dem die Freunde 1. und 2. Ordnung eines Nutzers dargestellt sind, stellen in diesem Kreislauf das mathematische Modell dar. Um eine mathematische Lösung zu erhalten, müssen diese Regeln auf die Nutzer im Datensatz angewendet

werden. Zum Schluss erfolgt, bezogen auf das reale Ausgangsproblem, das Alter eines Nutzers in einem sozialen Netzwerk vorherzusagen, eine Interpretation der Ergebnisse. Bevor die SuS mit dem zweiten Arbeitsblatt starten, wird ihnen in Erinnerung gerufen, dass die einzelnen Gruppen auf Grund der Filterung in der ersten Arbeitsphase einen unterschiedlichen Datensatz besitzen und ein Kriterium eingeführt werden muss, die Ergebnisse der Gruppen zu den Vorhersageregeln untereinander zu vergleichen. Nach einer kurzen Bedenkzeit wird gemeinsam mit den SuS der Begriff der Trefferquote in der Gestalt

$$\text{Trefferquote} = \frac{\text{Anzahl der Nutzer, die richtig eingeschätzt wurden}}{\text{Gesamtanzahl der Nutzer, die überprüft wurden}}$$

erarbeitet, welche sie in der zweiten Arbeitsphase für die einzelnen Vorhersageregeln bestimmen sollen. Damit endet der Vortrag und die SuS bearbeiten das zweite Arbeitsblatt.

4.5.7. Arbeitsblatt 2 | Vorhersageregeln anwenden

Auf dem zweiten Arbeitsblatt (siehe Anhang F) sollen die SuS zunächst die im Einführungsvortrag kennengelernten Vorhersageregeln zum arithmetischen Mittel, Median und Modus auf die Nutzer in ihrem Datensatz anwenden und die Trefferquoten bestimmen. Hier durchlaufen die SuS also den im Modellierungskreislauf (Abbildung 2) vorkommenden Schritt „berechnen“, um zu einer mathematischen Lösung zu gelangen. Dem vorangestellt sollen die SuS als erstes die drei Lagemaße wiederholen bzw. gegebenenfalls den Modus als ein neues Lagemaß kennenlernen. Dazu berechnen sie die Lagemaße für die Altersangaben der Freunde 1. Ordnung eines selbst ausgewählten Nutzers in ihrem Datensatz. Ob ihre Ergebnisse richtig sind, können sie gleich mit MATLAB überprüfen. Mit dieser ersten Aufgabe wird zum einen den SuS die Möglichkeit eines aktiven Erinnerns gegeben, indem sie bereits erworbenes Wissen aus dem Unterricht in einem neuen Kontext anwenden. Zum anderen wird gleich zu Beginn sichergestellt, dass die SuS die Berechnung der drei Lagemaße und damit die Anwendung der Vorhersageregeln verstanden haben. Nach Heymann sind dies zwei Aspekte, die bei der Gestaltung einer effektiven Lernsequenz berücksichtigt werden sollten (vgl. Heymann, 1998, S. 8). Zu dieser Aufgabe existiert eine Hilfekarte (siehe Anhang F.1.4), welche die SuS jeweils mit einem Beispiel bei der Berechnung der drei Lagemaße anleitet. Durch die Hilfekarte wird die Heterogenität der SuS berücksichtigt und das eigenständige Lernen gefördert. Nach dieser Wiederholung können die SuS entscheiden, ob sie die Vorhersageregeln zu den Freunden 1. oder 2. Ordnung untersuchen möchten. Eine Betrachtung aller sechs Regeln nimmt zu viel Zeit in Anspruch und birgt die Gefahr, dass die SuS das Anwenden der Regeln als monoton empfinden und sie dadurch an Motivation verlieren (vgl. Deci & Ryan, 1993, S. 226). Darüber hinaus können die SuS Hypothesen aufstellen, ob eine Betrachtung der Freunde 1. Ordnung oder der Freunde 2. Ordnung zu genaueren Ergebnissen führt, was den Gedanken der vierten Phase im PPDAC Zyklus (Abbildung 4) aufgreift. Haben sich die SuS entschieden, wenden sie die drei Regeln zunächst auf einzelne Nutzer an und notieren auf

dem zweiten Antwortblatt (siehe Anhang F.1.3), ob die Regeln für den ausgewählten Nutzer zu einer richtigen oder falschen Vorhersage führen. Anschließend wenden sie die Regeln auf alle Nutzer in ihrem Datensatz an und bestimmen die Trefferquote, eine richtige Vorhersage zu erzielen. Für eine selbstgewählte Regel sollen die SuS zudem die Stichprobengröße N mehrmals variieren und die Trefferquoten für die gewählte Stichprobe auf ihrem Antwortblatt notieren. Darauf aufbauend sollen die SuS begründen, welche Stichprobengröße eine aussagekräftige Trefferquote repräsentiert.

Anschließend wird die Robustheit gegenüber Ausreißern vom Median und vom arithmetischen Mittel thematisiert, um zu begründen, warum die Vorhersageregeln zum Median eine bessere Trefferquote erzielt als die Regeln zum arithmetischen Mittel. Zunächst berechnen die SuS die beiden Lagemaße für die Altersangaben der Freunde vom ausgedachten Nutzer Kim im Jahr 2003. Die Altersangaben von Kims Freunden befinden sich alle in der Nähe von Kims eigenem Alter. Beide Lagemaße kommen auf den Wert 22, der auch dem wahren Alter von Kim entspricht. Es wird nun das konstruierte Szenario betrachtet, dass Kim im Jahr 2004 seine beiden Eltern als Freunde hinzugewonnen hat. Die Ausreißer in den Daten werden also durch die Altersangaben von Kims Eltern repräsentiert. Eine erneute Bestimmung des Medians führt zum gleichen Wert wie aus dem Jahr 2003. Das arithmetische Mittel liefert nun aber den Wert 30 und weicht damit um 8 Jahre deutlich vom wahren Alter ab. So sehen die SuS anschaulich, welchen Effekt Ausreißer auf die beiden Lagemaße haben. Diese Erkenntnis halten sie auch auf ihrem Antwortblatt fest.

Die SuS haben durch die herausgefundenen Trefferquoten eine erste mathematische Lösung erhalten und befinden sich nun im letzten Schritt des Modellierungskreislaufes, der Interpretation der Ergebnisse bezogen auf das reale Problem. Dabei werden sie feststellen, dass die Trefferquoten zu den Lagemaßen unter 20 % liegen und somit für eine Vorhersage des exakten Alters unzureichend sind. Hier lernen die SuS die bereits im Modellierungsvortrag erwähnte Notwendigkeit kennen, den Modellierungskreislauf mehrfach durchlaufen zu müssen, um zu einer zufriedenstellenden Lösung zu gelangen. Dazu wird für das weitere Vorgehen das mathematische Modell durch die Betrachtung neuer Vorhersageregeln angepasst. Aus diesem Grund lernen die SuS im Verlauf des Arbeitsblattes zwei weitere Regeln kennen, die nicht das exakte Alter eines Nutzers bestimmen, sondern ein Intervall vorhersagen, in dem das Alter der Person liegt. Diese Regeln basieren auf die Spannweite und auf dem Quartilsabstand. Um den SuS ein Gefühl für diese beiden Streuungsmaße zu vermitteln, zeichnen sie zunächst zu den Altersangaben der Freunde 1. Ordnung der fiktiven Nutzerin Julia einen Boxplot. Mit einem Boxplot können die SuS den Median und die beiden Streuungsmaße auf einen Blick erfassen. Zur Orientierung und als Erinnerung sehen die SuS auf ihrem Arbeitsblatt die Graphik 10 aus dem Abschnitt 3.6, die den Aufbau eines Boxplots mit den dazugehörigen Kenngrößen für die Erstellung zeigt. Wie in didaktischen Büchern der Stochastik empfohlen, wird dem zugehörigen Boxplot in dieser Graphik auch die Verteilung der Daten in einem Punktdiagramm gegenübergestellt (vgl. Krüger et al., 2015, S. 123). Für SuS, die sich bei der Zeichnung eines Boxplots unsicher sind, existiert eine Hilfekarte (siehe Anhang F.1.5). In dieser Hilfekarte werden die SuS bei der Erstellung des Boxplots Schritt für Schritt interaktiv angeleitet. Die Hilfekarte ist in die fünf

Schritte

1. Die Daten sortieren
2. Maximum und Minimum bestimmen
3. Den Median bestimmen
4. Das untere Quartil und das obere Quartil bestimmen
5. Den Boxplot zeichnen

unterteilt. Das heißt, je nachdem bei welchem Schritt die SuS Schwierigkeiten haben, können sie den entsprechenden Abschnitt in der Hilfekarte auswählen, um eine Unterstützung zu erhalten. Jeder Abschnitt beginnt zunächst mit einem kurzen Informationstext, der das Vorgehen zum jeweiligen Schritt. Anschließend sollen die SuS das erhaltene Wissen auf den Datensatz mit den Altersangaben der Freunde von Julia anwenden. Die Computersoftware MATLAB gibt ihnen dann gleich eine Rückmeldung, ob ihre Lösung richtig ist. Mit der Hilfekarte passt sich die Aufgabe, einen Boxplot zu zeichnen, bis zu einem gewissen Grad den unterschiedlichen Lernvoraussetzungen der SuS an. Auch können sie den Schwierigkeitsgrad der Aufgabe in gewissen Grenzen an ihren individuellen Lerntypen anpassen, da sie selbst entscheiden, wann und in welchem Umfang sie die Hilfen nutzen. Diese Entscheidung beruht auf den kognitiven und motivationalen Voraussetzungen der SuS. Somit bietet die Hilfekarte eine Möglichkeit zur Differenzierung (vgl. Wodzinski, 2013, S. 46).

Nach der Erstellung des Boxplots überlegen die SuS, wie viel Prozent der Werte mindestens innerhalb des gezeichneten Kastens des Boxplots liegen. Auch sollen sie die Spannweite und den Quartilsabstand zu den betrachteten Daten von Julias Freunden bestimmen. Eine Definition zur Berechnung der beiden Werte finden sie auf dem Arbeitsblatt. Wie schon beim arithmetischen Mittel und Median diskutieren die SuS auch zu den beiden Streuungsmaßen das Thema Robustheit gegenüber Ausreißern. Nachdem sich die SuS intensiv mit den beiden Streuungsmaßen beschäftigt haben und sichergestellt ist, dass sie mit diesen Begriffen umgehen können, erhalten sie eine Übersicht über die neuen Vorhersageregeln (siehe Anhang F.1.2). Denn diese neuen Regeln beziehen sich genau auf die beiden vorgestellten Streuungsmaße. Wieder können die einzelnen Gruppen entscheiden, ob sie die Freunde 1. oder 2. Ordnung betrachten wollen. Sie wenden ihre ausgewählten Vorhersageregeln an, bestimmen die Trefferquoten und lassen sich von MATLAB in einem weiteren Schritt die durchschnittlichen Intervallbreiten, in denen das vorhergesagte Alter mit den neuen Regeln liegt, anzeigen. Aus diesen Informationen sollen die SuS interpretieren, welche Regel sich am Besten für eine Vorhersage des Alters eignet. Ihre Ergebnisse halten die SuS auf ihrem Antwortblatt fest. Dabei sind die SuS erneut an dem Punkt im Modellierungskreislauf angekommen, an dem sie ihre mathematischen Lösungen bezogen auf die Ausgangsfragestellung interpretieren müssen.

Die Zusatzaufgabe in diesem Arbeitsblatt befasst sich mit der Standardabweichung als

drittes Streuungsmaß. Mit Hilfe der vorgegebenen Formel sollen die SuS zunächst die Standardabweichung eines vorgegebenen Datensatzes bestimmen. Anschließend wenden sie die Vorhersageregeln zur Standardabweichung auf ihren Datensatz an und bestimmen, wie bei den beiden vorherigen Streuungsmaßen, die Trefferquote sowie die durchschnittliche Intervallbreite.

4.5.8. Zweite Sicherungsphase

Auch die Sicherung des zweiten Arbeitsblatts erfolgt mit einer vorgefertigten Overheadfolie (siehe Anhang F.1.6). Zunächst wird mit den SuS diskutiert, welche Regel sie zur Bestimmung des exakten Alters am Effizientesten vermuten. Die Begründungen werden auf der Folie gesammelt. Im nächsten Schritt werden die Trefferquoten zu jeder Regel gesammelt und mit den SuS besprochen, welchen Einfluss die Wahl der Stichprobengröße auf die Trefferquote hat. Die Gründe, eine möglichst große Stichprobengröße zu wählen, werden ebenfalls auf der Folie notiert. Anschließend wird das Thema Robustheit gegenüber Ausreißern bezüglich des arithmetischen Mittels und des Medians angesprochen. Dazu werden zunächst das arithmetische Mittel und der Median der Altersangaben von Kims Freunden im Jahr 2003 und im Jahr 2004 aufgeschrieben. Danach gehen die SuS in der Besprechung auf den Einfluss der Ausreißer, die im Jahr 2004 dazugekommen sind, ein. Dieser Einfluss wird auch auf der Folie festgehalten. Bevor mit den SuS auf die neuen Regeln eingegangen wird, sollen sie kurz zusammenfassen, zu welchen Ergebnissen die Vorhersageregeln zur Bestimmung des exakten Alters der Nutzer kommen und ob die herausgefundenen Trefferquoten für eine richtige Vorhersage zufriedenstellend sind. Die zweite Sicherungsphase kann aus zeitlichen Gründen optional an dieser Stelle unterbrochen und nach der Mittagspause fortgeführt werden. Denn eine Teilsicherung der ersten Vorhersageregeln, welche im Rahmen des Zwischenvortrags vorgestellt wurden, ist nun erfolgt.

Auf Grund der niedrigen Trefferquoten für die Vorhersage des exakten Alters erkennen die SuS die Notwendigkeit neue Regeln aufzustellen. Die Besprechung dieser neuen Regeln startet mit dem Boxplot zu den Altersangaben von Julias Freunden. Dieser Boxplot wird von einem Schüler bzw. einer Schülerin auf der Overheadfolie aufgezeichnet und das Vorgehen zur Erstellung eines Boxplots erklärt. Mit Hilfe des Boxplots werden die Begriffe Spannweite und Quartilsabstand aufgegriffen und im Sachzusammenhang interpretiert. Auch erfolgt hier eine Diskussion über die Robustheit gegenüber Ausreißern bezüglich der beiden Streuungsmaße. Die Interpretation und Diskussion wird stichpunktartig auf der Folie festgehalten. Anschließend werden die neuen Vorhersageregeln zur Spannweite und zum Quartilsabstand thematisiert. Die SuS nennen zunächst wieder ihre begründeten Vermutungen für die beste Regel. Danach werden die herausgefundenen Trefferquoten zu den einzelnen Regeln auf der Folie notiert. Ebenfalls werden die durchschnittlichen Intervallbreiten, in denen das Alter der Nutzer vermutet wird, für jede Regel festgehalten. Ausgehend von den Trefferquoten und den Intervallbreiten diskutieren die SuS abschließend, welche der untersuchten Vorhersageregeln sie einem Betreiber eines sozialen Netzwerkes für den realen Einsatz empfehlen würden. Wie in der letzten Phase des PPDAC Zyklus, dem Fazit, vorgesehen, interpretieren

und kommunizieren die SuS hier ihre herausgefundenen Lösungen. Darüber hinaus werden bereits über neue Ideen nachgedacht, um eine noch genauere Vorhersage über das Alter eines Nutzers zu treffen.

Zu der Zusatzaufgabe, welche die Standardabweichung als eine weitere Vorhersageregeln betrachtet, erfolgt keine Sicherung im Plenum, da diese Aufgabe nur von einem kleinen Teil der SuS bearbeitet wird. Die SuS können jedoch bei Fragen die Aufgaben mit den Dozenten besprechen. Eine Überprüfung der Aufgaben erfolgt zudem mit MATLAB.

4.5.9. Arbeitsblatt 3 | Werbung schalten

Im dritten Aufgabenblatt begeben sich die SuS in die Rolle eines Betreibers eines sozialen Netzwerks, der den Auftrag erhalten hat, Werbung an Personen zwischen 25 und 30 Jahren zu verschicken. Um diesen Auftrag zu erfüllen, überprüfen die SuS zwei Ideen ihrer Mitarbeiter. Je nach Niveau unterscheidet sich die Idee des ersten Mitarbeiters:

Drittes Aufgabenblatt für Niveau 1:

Im ersten Niveau (siehe Anhang G) befasst sich die Idee des ersten Mitarbeiters mit einer Vorhersageregeln, die untersucht, ob das arithmetische Mittel der Freunde 1. Ordnung im vorgegebenen Altersintervall $[25, 30]$ liegt. Wenn dies der Fall ist, wird dem Nutzer ebenfalls ein Alter innerhalb dieses Intervalls zugeordnet. Die SuS wenden diese Vorhersageregeln auf ihren Datensatz an und tragen ihre herausgefundenen Trefferquoten auf das zugehörige Antwortblatt (siehe Anhang G.1.1) ein. Um einer aufkommenden Langeweile bzw. Monotonie bei den SuS vorzubeugen, werden im dritten Arbeitsblatt die Trefferquoten gleich von MATLAB angegeben und müssen nicht mehr einzeln berechnet werden. Außerdem formulieren die SuS eine Vorhersageregeln für die Freunde 2. Ordnung und wenden diese Regeln ebenfalls auf ihren Datensatz an.

Die anschließende Vorhersageregeln eines zweiten Mitarbeiters ist für beide Niveaus identisch. Hier werden die Anteile der Freunde 1. Ordnung eines Nutzers betrachtet, die ein Alter zwischen 25 und 30 Jahren aufweisen. Zunächst wenden die SuS diese Regeln mit den Parametern $[0.4, 0.2]$ an. Das bedeutet, wenn mehr als 40 % der Freunde ein Alter innerhalb des Intervalls $[25, 30]$ besitzen, wird dem Nutzer ebenfalls dieses Alter zugeordnet. Haben weniger als 20 % der Freunde ein Alter innerhalb dieses Intervalls, wird dem Nutzer ein Alter außerhalb des Intervalls vorhergesagt. Liegt der Anteil zwischen 20 % und 40 % wird eine zufällige Zuordnung, im Sinne der im Abschnitt 3.3.2 vorgestellten Vorhersageregeln zum Zufall, vorgenommen. Um sicherzustellen, dass die SuS das Vorgehen zu dieser Vorhersageregeln verstanden haben, wenden sie die Regeln zunächst auf den fiktiven Nutzer Peter an. Sie bestimmen den Anteil der Freunde 1. Ordnung von Peter, die ein Alter zwischen 25 und 30 Jahre haben und treffen basierend auf diesem Anteil mit der Regeln des zweiten Mitarbeiters eine Vorhersage für das Alter von Peter. Meyer spricht in diesem Zusammenhang von einer klaren Strukturierung der Lernsequenz (vgl. Meyer, 2004, S. 30). Den SuS soll die Aufgabenstellung klar sein und sie sollen zu jeder Zeit wissen, was von ihnen verlangt wird. Anschließend wenden sie die Vorhersageregeln auf ihren Datensatz an. Die Trefferquote, welche MATLAB nach

der Anwendung ausgibt, übertragen die SuS auf ihr Antwortblatt. Im weiteren Verlauf dürfen sie eigene Parameter wählen und überprüfen, ob sie mit ihrer Wahl eine noch bessere Trefferquote erzielen können. Dieser kleine Wettkampf unter den Gruppen, die beste Trefferquote herauszufinden, soll die Motivation der SuS erhöhen, was zur einer intensiveren Bearbeitungen der Aufgabe führen soll (vgl Aebli, 1933, S. 343). Für die gewählten Parameter halten die SuS auf ihrem Antwortblatt außerdem fest, ob eher Nutzer mit einem Alter innerhalb des Intervalls $[25, 30]$ oder außerhalb des Intervalls besser vorhergesagt werden.

Schnelle SuS haben im Rahmen einer Zusatzaufgabe die Möglichkeit mit ihren gelernen Vorhersageregeln das vermeintlich richtige Alter der aussortierten Nutzer mit einer Altersangabe von null Jahre vorherzusagen. Sie können auf diese Nutzer ihre Regeln zu den Lagemaßen und den Streuungsmaßen anwenden und die jeweiligen Ergebnisse vergleichen. Mit dieser Aufgabe sollen die SuS erkennen, dass sie mit ihren herausgefundenen Trefferquoten nun eine Aussage über das unbekannte Alter eines Nutzers treffen können bzw. überprüfen können, ob das angegeben Alter eines Nutzer korrekt ist.

Drittes Aufgabenblatt für Niveau 2:

Im zweiten Niveau (siehe Anhang H) befasst sich die Idee des ersten Mitarbeiter mit der im Abschnitt 3.3.2 vorgestellten Vorhersageregeln zum Zufall. Das heißt, mit einer Wahrscheinlichkeit q wird dem Nutzer ein Alter innerhalb des Intervalls $[25, 30]$ zugeordnet und mit einer Wahrscheinlichkeit $1 - q$ ein Alter außerhalb des Intervalls. Als Wahrscheinlichkeit q wählen die SuS die relative Häufigkeit, einen Nutzer mit einem Alter zwischen 25 und 30 Jahren in ihrem Datensatz zu finden. Anschließend wenden sie mit diesem q die Vorhersageregeln zum Zufall auf die Nutzer in ihrem Datensatz an. Diese Anwendung führen die SuS mehrmals mit jeweils unterschiedlichen Stichprobengrößen durch. Die daraus resultierenden Trefferquoten notieren sie auf dem zugehörigen dritten Antwortblatt (siehe Anhang H.1.1). Anschließend sollen die SuS mit Hilfe eines zweistufigen Baumdiagramms die theoretische Trefferwahrscheinlichkeit bestimmen, mit der die Vorhersageregeln zum Zufall eine richtige Zuordnung in die beiden Altersgruppen vornimmt. Diesen Wert vergleichen die SuS dann mit den bereits experimentell herausgefundenen Trefferquoten. Sie sollen erkennen, dass je größer sie die Stichprobe wählen, sich die experimentelle Trefferquote immer mehr der theoretischen Trefferquote annähert. Für SuS, die Schwierigkeiten bei der Erstellung des Baumdiagramms haben, existiert eine Hilfekarte (siehe Anhang H.1.4). Genau wie die Hilfekarte zum Boxplot leitet diese Hilfekarte die SuS Schritt für Schritt an. Folgende fünf Schritte können dazu separat aufgerufen werden:

1. Die zwei Stufen im Baumdiagramm darstellen
2. Die Wahrscheinlichkeit an den Ästen eintragen
3. Die richtigen Pfade aussuchen

4. Die Wahrscheinlichkeit der richtigen Pfade bestimmen
5. Die gesuchte Wahrscheinlichkeit berechnen

Wieder können die SuS nach ihrem persönlichen Leistungsniveau entscheiden, bei welchen Schritten sie eine Unterstützung benötigen. Vor allem soll aber durch diese Hilfekarte ein Ausstieg der SuS auf Grund der erhöhten Schwierigkeit vermieden werden. Wenn SuS der Aufgabe nicht mehr folgen können, brechen sie den Lernprozess ab und steigen gedanklich aus. Einerseits bleibt so Lernzeit ungenutzt und andererseits wird dieser Ausstieg oft mit einem Misserfolg verbunden (vgl. Wodzinski, 2013, S. 47). Mit dieser Hilfekarte soll genau diesen Aspekten entgegengewirkt werden, indem auch leistungsschwächere SuS Teilleistungen erbringen können und dadurch ein Kompetenzerlebnis erfahren.

Wie bereits erwähnt, ist die Idee des zweiten Mitarbeiters auf diesem Arbeitsblatt identisch zu der aus dem ersten Niveau. Zusätzlich zur Zusatzaufgabe aus dem ersten Niveau existiert im zweiten Niveau eine weitere Aufgabe. Beide Aufgaben stehen auf einem extra Aufgabenblatt (siehe Anhang H.1.2) zu dem ein eigenes Antwortblatt (siehe Anhang H.1.3) existiert. In der zweiten Zusatzaufgabe sollen die SuS selber Wahrscheinlichkeiten x wählen, mit denen einem Nutzer ein Alter zwischen 25 und 30 Jahren zugeordnet wird, und die zugehörige theoretische Trefferwahrscheinlichkeit berechnen. Zudem sollen sie herausfinden, bei welcher Wahl von x sie die höchste theoretische Trefferquote erhalten.

4.5.10. Dritte Sicherungsphase

Für die Sicherung der Ergebnisse aus dem dritten Arbeitsblatt existiert ebenfalls eine bereits vorbereitete Overheadfolie für Niveau 1 (siehe Anhang G.1.2) und für Niveau 2 (siehe Anhang H.1.5).

Für das erste Niveau werden als erstes die herausgefundenen Trefferquoten der SuS auf der Folie festgehalten und die Vorhersageregeln für die Freunde 2. Ordnung aufgeschrieben. Für die Besprechung der Vorhersageregeln des zweiten Mitarbeiters, welche im Niveau 2 identisch verläuft, wird zunächst auf das ausgedachte Beispiel des Nutzers Peter eingegangen. Anhand dieses Beispiels sollen die SuS die Regel des Mitarbeiters nochmal in eigenen Worten erklären und eine Vorhersage für Peters Alter treffen. Danach werden die Trefferquoten zu den gewählten Parametern der SuS auf der Folie gesammelt. Dabei gehen die SuS jeweils darauf ein, welche der beiden Altersgruppen besser vorhergesagt wird. Abschließend sollen die SuS die herausgefundenen Ergebnisse bezogen auf die reale Ausgangsfragestellung, das Alter eines Nutzers vorherzusagen, interpretieren.

Im zweiten Niveau unterscheidet sich die Besprechung am Anfang, da die SuS hier die Vorhersageregeln zum Zufall betrachtet haben. Da die einzelnen Gruppen unterschiedliche Datensätze besitzen, werden zunächst für drei Gruppen die herausgefundenen relativen Häufigkeiten, einen Nutzer mit einem Alter zwischen 25 und 30 Jahren im Datensatz vorzufinden, auf der Folie notiert. Anschließend werden für vier verschiedenen Stichprobengrößen die Trefferquoten, eine richtige Vorhersage mit einer zufälligen

Zuordnung zu landen, gesammelt. Zur besseren Übersicht schreibt der Dozent die Stichprobengrößen aufsteigend auf. Die Sicherung des zweistufigen Baumdiagramms erfolgt durch einen Schüler bzw. eine Schülerin, der bzw. die das Baumdiagramm auf der Folie aufzeichnet und dabei das Vorgehen erklärt. Mit den SuS wird die Berechnung der theoretischen Trefferwahrscheinlichkeit, für einen zufällig ausgewählten Nutzer eine richtige Vorhersage zu treffen, ausführlich anhand des gezeichneten Baumdiagramms besprochen und auf der Folie notiert. Im Anschluss erfolgt der Rückschluss auf die herausgefundenen Trefferquoten zu den unterschiedlich gewählten Stichprobengrößen. Mit den SuS wird besprochen, dass sich der Wert der experimentellen Trefferquote bei einer wachsenden Stichprobengröße dem Wert der theoretischen Trefferwahrscheinlichkeit annähert. Die Besprechung der Vorhersageregeln des zweiten Mitarbeiters erfolgt dann analog zu der aus dem ersten Niveau.

Wie in der zweiten Sicherungsphase werden auch hier die Zusatzaufgaben nicht mit allen SuS besprochen, da diese nicht von jeder Gruppe bearbeitet werden. Auch hier können die SuS ihre Lösungen jederzeit schon während der dritten Arbeitsphase mit den Dozenten besprechen.

4.5.11. Abschlusspräsentation

In der Abschlusspräsentation (siehe Anhang I.1) sollen nochmal die Ergebnisse des Tages zusammengefasst und eine Diskussion zum Thema Sicherheit der Privatsphäre in sozialen Netzwerken geführt werden. Zunächst sollen die SuS auf die Ausgangsfragestellung „Wie gut kann man das Alter eines Nutzers in einem sozialen Netzwerk vorhersagen, obwohl dieses aus den Angaben in seinem Profil nicht ablesbar ist?“ eingehen und eine Antwort formulieren. Hier sollen die SuS auch nochmal kurz auf die Ergebnisse der einzelnen überprüften Vorhersageregeln eingehen. Anschließend wird den SuS der Modellierungskreislauf gezeigt. Sie sollen hier zunächst selber ihr Vorgehen mit Hilfe des Kreislaufes beschreiben. und erkennen, dass sie im ersten Arbeitsblatt durch die Filterung der Nutzer mit einer vermeintlich falschen Altersangabe und den daraus resultierenden gefilterten Datensatz ein vereinfachtes Modell aufgestellt haben. Durch das Aufstellen von Vorhersageregeln haben sie ihr vereinfachtes Modell mathematisch beschrieben und sind durch die Anwendung der Regeln auf die Nutzer in ihrem Datensatz zu mathematischen Lösungen gelangt. Die SuS sollen vor allem erkennen, dass sie während des CAMMP days den Modellierungskreislauf mehrfach durchlaufen haben, indem das mathematische Modell immer wieder angepasst wurde. Zum Beispiel wurde erkannt, dass die Trefferquoten zu den Vorhersageregeln, die das exakte Alter vorhersagen sollen, für die reale Situation zu gering waren. Demnach wurde das mathematische Modell durch neue Regeln angepasst. Die Trefferquoten dieser neuen Regeln wurden ebenfalls bestimmt und die mathematischen Lösungen erneut auf die reale Situation bezogen. In diesem Zusammenhang sollen die SuS außerdem überlegen, wie das aufgestellte Modell, eine Vorhersage zum Alter einer Person zu treffen, noch weiter verbessert werden kann. Beispielsweise könnten die SuS auf die Ideen kommen, die Freunde 3. Ordnung oder einen größeren Datensatz zu betrachten.

Der zweite Teil der Präsentation stellt eine Diskussion dar, um gemeinsam mit den

SuS nochmal zum Thema Sicherheit der Privatsphäre in sozialen Netzwerken ins Gespräch zu kommen. Als erstes wird der Frage nachgegangen, ob es Personen gibt, die vom Sammeln der Daten ausgeschlossen sind. Hier soll die Möglichkeit aufgezeigt werden, Informationen über Personen herauszubekommen, die kein Mitglied eines sozialen Netzwerks sind. Als Beispiel wird zum einen die Weitergabe von Kontaktbuchinformationen durch Nutzer, die Mitglieder in sozialen Netzwerken sind, genannt. Über diese Angaben können auch auf private Informationen von Personen geschlossen werden, die kein Teil eines sozialen Netzwerkes sind. Zum anderen sammeln und analysieren nicht nur soziale Netzwerke Informationen über Personen. Auch Online-Giganten wie Google oder Amazon gehen dieser Tätigkeit nach. Mit dieser ersten Frage wird den SuS aufgezeigt, dass eigentlich alle Personen vom Datensammeln und der Vorhersage persönlicher Informationen betroffen sind. Als nächstes werden die SuS mit der provokativen Aussage „Wenn du nicht für einen Dienst bezahlst, dann bist du nicht der Kunde, sondern das Produkt.“ konfrontiert. Denn die Dienste von Facebook, Twitter oder auch Google sind kostenlos, obwohl diese Unternehmen unter anderem Kosten für Mitarbeiter und für die Betreuung der Server aufbringen müssen. Die SuS sollen überlegen, ob sie einverstanden sind, mit ihren Daten für diese Dienste zu bezahlen. Anschließend diskutieren die SuS über mögliche Vor- und Nachteile, die das Sammeln von Daten mit sich bringt. Hier sollen sowohl Vorteile, wie das Verhindern von möglichen Terroranschlägen oder die Unterstützung der Polizeiarbeit bei der Fahndung eines Verdächtigen, als auch Nachteile, wie beispielsweise die Gefahr der Manipulation durch persönlich angepasste Werbung und Informationen, thematisiert werden. Auch sollen die SuS über Handlungsmöglichkeiten zur digitalen Selbstverteidigung reflektieren. Dazu diskutieren sie über Möglichkeiten, die Daten und damit die Privatsphäre zu schützen. Hier können beispielsweise das Nutzen mehrere Suchmaschinen über Google hinaus, die Verwendung mehrerer Email-Adressen oder das Deaktivieren von Zugriffsberechtigungen installierter Apps auf dem Smartphone genannt werden. Abschließend werden die SuS nach den Konsequenzen gefragt, die sie aus dem CAMMP day mitnehmen. Hier sollen die SuS nochmal für sich über den zukünftigen Umgang mit sozialen Netzwerken reflektieren.

4.5.12. Material für die Dozenten

Zur Unterstützung der Dozenten bei der Durchführung des entwickelten Lernmoduls, wurden ebenfalls Materialien entworfen. Dazu gehört zunächst das methodische Konzept (siehe Anhang B). Es besteht aus einer Auflistung aller Materialien, die während des CAMMP days zum Einsatz kommen, und gibt einen groben tabellarischen Ablauf des Lernmoduls vor. Die Dozenten können in Form einer Checkliste abhaken, ob sie alle Materialien für den CAMMP day vorliegen haben und den Ablauf der einzelnen Phasen während der Durchführung nachverfolgen.

Auch existieren für die entworfenen Präsentationen Notizen zu den jeweiligen Vorträgen (siehe Anhänge C.2 und I.2). Sie geben den Dozenten eine Übersicht, was zu den einzelnen Folien der Vorträge erzählt werden soll, und vermitteln ihnen dadurch eine gewisse Sicherheit beim Präsentieren.

Darüber hinaus existieren zu allen Arbeitsblättern Musterlösungen (siehe Anhang E.2, F.2, G.2 und H.2). Sie enthalten sowohl Lösungswege als auch Begründungen zu den einzelnen Aufgaben. Ebenso gibt es einen Mustercode, welcher die Musterlösung des Codes darstellt. Diesen können die Dozenten bei auftretenden Fehlermeldungen mit dem Code der SuS abgleichen, um eine schnelle Lösung des Problems zu finden.

4.6. Die Vorzüge der Einbindung der Computersoftware MATLAB in das entwickelte Lernmodul

Besonders im Umgang mit realen Problemstellungen kann die Einbindung bzw. der Einsatz digitaler Werkzeuge wie MATLAB eine sinnvolle Unterstützung für die Einführung komplexerer Anwendungen und Modellierungen darstellen. Digitale Werkzeuge wie MATLAB verringern den Rechenaufwand und bieten die Möglichkeit, in einer angemessenen Zeit ein realistisches Datenvolumen zu bearbeiten (vgl. Greefrath & Weitendorf, 2013, S. 182) So ist es den SuS möglich, die aufgestellten Vorhersageregeln zur Bestimmung des Alters auf reale Nutzerprofile des sozialen Netzwerks Friendster anzuwenden. Auch die Stichprobengröße kann schnell und unkompliziert variiert werden. Durch den Einsatz von MATLAB werden die SuS also auch entlastet, wodurch der Fokus auf den Aufbau der Modellierungskompetenz sowie auf das problemorientierte Bearbeiten der Aufgaben gelegt werden kann (vgl. Kaiser et al., 2015, S. 371).

Ein weiterer Vorzug der Einbindung digitaler Werkzeuge im Kontext von Modellierungsproblemen stellt die Möglichkeit zum Recherchieren von Informationen im Zusammenhang mit dem Anwendungskontext dar (vgl. Greefrath & Weitendorf, 2013, S. 183). Die SuS können jeder Zeit selbstständig im Internet nach benötigten Informationen, wie beispielsweise einer Formel zur Berechnung einer Aufgabe, suchen.

Digitale Werkzeuge können zudem die Aufgabe der Visualisierung übernehmen (vgl. Greefrath & Weitendorf, 2013, S. 182). So können sich die SuS beispielsweise im Rahmen des ersten Arbeitsblatts mit Hilfe von MATLAB die Altersverteilung in ihrem Datensatz als Säulendiagramm anzeigen lassen und dies als Ausgangspunkt für die Entwicklung eines mathematischen Modells benutzen.

Ein weiterer Vorteil der Einbindung digitaler Medien liegt in der Unterstützung von Kontrollprozessen. Mit Hilfe von Überprüfungsfunktionen können bearbeitete Aufgaben kontrolliert bzw. überprüft werden (vgl. Greefrath & Weitendorf, 2013, S. 183). Diese Kontrollinstanz übernimmt auch MATLAB im Rahmen des entwickelten Lernmodus. Die Rückmeldung der SuS erfolgt dabei immer direkt unterhalb der bearbeiteten Aufgabe. Müssen die SuS mehrere Lösungen auf einmal überprüfen, gibt MATLAB eine differenzierte Ausgabe zu der Richtigkeit der Lösungen an. Abbildung 16 zeigt beispielhaft eine solche differenzierte Rückmeldung bei einer noch nicht korrekten Eingabe der SuS.

```
Nutzer          = 8;           % Nutzer, dessen Freunde angezeigt werden sollen
arithmetisches_Mittel = 38;      % Euer berechnetes arith. Mittel für den Nutzer
Median          = 33;          % Euer berechneter Median für den Nutzer
Modus           = 33.5;        % Euer berechneter Modus für den Nutzer
```

```
Daten
Check_Loesung1(Users, Nutzer, arithmetisches_Mittel, Median, Modus)
```

```
Euer Ergebnis zum arithmetischen Mittel ist noch nicht korrekt.
Euer Ergebnis zum Median ist noch nicht korrekt
```

```
-----
Die Altersangaben der Freunde von Nutzer 8 lauten:
33 43 36 41 31 31 38 37 36
-----
```

Abbildung 16: Ausgabe von MATLAB bei einer Aufgabe innerhalb des zweiten Arbeitsblatts zur Bestimmung des arithmetischen Mittels, Medians und Modus zu den Altersangaben der Freunde 1. Ordnung eines ausgewählten Nutzers im Datensatz. MATLAB zeigt an, dass die berechneten Ergebnisse zum arithmetischen Mittel und zum Median noch nicht korrekt sind.

5. Durchführung und Evaluation des entwickelten Lernmoduls

Das entwickelte Lernmodul wurde im Rahmen der Masterarbeit mit einer Schülergruppe durchgeführt. In diesem Kapitel werden die Beobachtungen und Reaktionen der ersten Durchführung beschrieben. Wie für die CAMMP days üblich, wurde auch zu diesem Lernmodul ein Evaluationsbogen an die teilnehmenden SuS ausgeteilt. Anhand ausgewählter Schwerpunkte werden die Ergebnisse dieser Evaluation vorgestellt und die sowohl aus der Evaluation als auch aus der Beobachtung resultierenden Verbesserungen für das entworfene Lernmodul aufgezeigt.

5.1. Rahmenbedingungen

Der entwickelte CAMMP day wurde erstmalig am 20.11.2018 mit einer achten Klasse des Kaiser-Karls-Gymnasiums aus Aachen durchgeführt. Es nahmen insgesamt 22 SuS teil, wobei der männliche und weibliche Anteil an Teilnehmern ausgeglichen war. Nach Aussagen der Lehrkraft handelt es sich bei der achten Klasse um einen eher schwächeren Kurs, bei der die Heterogenität der SuS deutlich zu spüren ist. So gibt es in der Klasse sowohl SuS, welche die achte Klasse bereits wiederholen und deren Leistung wieder im defizitären Bereich liegt, als auch SuS, deren Leistung sich im Einser-Bereich einordnet. Der CAMMP day begann um 8:15 Uhr und endete gegen 13:30 Uhr. Auf Grund der verkürzten Zeit ist die große Mittagspause von einer Stunde entfallen. Stattdessen wurden zwei kleinere Pausen à 15 Minuten eingelegt.

5.2. Leitende Gesichtspunkte der Beobachtung und der Evaluation

Sowohl die Fragen des Evaluationsbogens als auch die während der Durchführung gesammelten Beobachtungen orientieren sich an den folgenden Schwerpunkten:

- **Interesse:**
Hat das Thema „Sicherheit der Privatsphäre in sozialen Netzwerken“ das Interesse der SuS geweckt, sie für den CAMMP day begeistert und sie zum Nachdenken angeregt?
- **Verständnis:**
Haben die SuS verstanden, wie mit einfachen mathematischen Regeln Vorhersagen über Informationen in sozialen Netzwerken getroffen werden können, welche die Nutzer in ihren Profilen nicht angegeben haben?
- **Schwierigkeit der Aufgaben:**
Ist der Schwierigkeitsgrad unter Berücksichtigung der Heterogenität der Gruppe angemessen? Waren die im Lernmodul vorkommenden Lage- und Streuungsmaße den SuS bereits vor dem CAMMP day bekannt? Hatten sie Schwierigkeiten mit diesen Lage- und Streuungsmaße umzugehen?

- **Modellieren und Mathematik:**

Wird den SuS durch die gewählte Thematik im Lernmodul die Rolle bzw. die Bedeutung des Modellierens und der Mathematik in der Gesellschaft bewusst und leistet das Lernmodul einen Beitrag zum Aufbau von Modellierungskompetenzen?

- **Umgang mit MATLAB:**

Sind bei den SuS beim Umgang mit dem Programm MATLAB Probleme oder Schwierigkeiten aufgetreten?

- **Leerlauf:**

Hatten die SuS den Eindruck, dass sie zeitweise keinen Arbeitsauftrag hatten, und ist dadurch Langeweile aufgetreten?

5.3. Vorstellung des Evaluationsbogens

Der Evaluationsbogen, den die SuS ausgefüllt haben, bestand sowohl aus geschlossenen als auch aus offenen Fragen. Zu dem bereits bestehenden Fragebogen, der nach jedem durchgeführten CAMMP day ausgeteilt wird, wurden folgende, speziell auf das entwickelte Lernmodul angepasste Fragen in den Fragebogen eingearbeitet:

- Der CAMMP day hat mir Spaß gemacht.
- Das Thema „Sicherheit der Privatsphäre in sozialen Netzwerken“ hat mich interessiert.
- Ich habe verstanden, wie mit einfachen, mathematischen Regeln Vorhersagen über das Alter getroffen werden können.
- Ich habe auf Blatt 1 verstanden, warum die Daten gefiltert werden müssen.
- Ich wusste bereits, dass Informationen vorhergesagt werden können.
- Ich werde in Zukunft mehr über das Thema „Sicherheit in sozialen Netzwerken“ nachdenken.
- Ich wusste bereits vor dem CAMMP day, dass Mathematik in meinem Alltag eine wichtige Rolle spielt.
- Durch den CAMMP day habe ich erkannt, dass Mathematik eine wichtige Rolle in der Gesellschaft hat.

Der gesamte Fragebogen ist im Anhang J.1 zu finden.

Der Evaluationsbogen besteht aus mehreren Teilen. Zu Beginn werden die SuS nach dem Geschlecht gefragt. Anschließend wird die Bewertung des entwickelten CAMMP days vorgenommen. Hier müssen die SuS zu verschiedenen Aussagen, wie beispielsweise „Das Thema ‚Sicherheit der Privatsphäre in sozialen Netzwerken‘ hat mich interessiert.“

oder „Die Hilfekarten waren hilfreich.“ Stellung nehmen. Dabei können sie zwischen vier Antwortmöglichkeiten unterscheiden: „Trifft gar nicht zu (- -)“, „Trifft eher nicht zu (-)“, „Trifft zum Teil zu (+)“ und „Trifft voll zu (++)“. Auch können die SuS hier im Rahmen zweier geschlossener Fragen, bei denen eine Mehrfachnennung möglich ist, angeben, ob sie bereits vor der Durchführung des Lernmoduls mit den vorgekommenen Lage- und Streuungsmaßen vertraut waren und ob sie nach dem CAMMP day ein besseres Verständnis dafür besitzen. Im nächsten Teil des Evaluationsbogens können die SuS in offenen Fragen angeben, ob ihnen im Lernmodul etwas besonders schwer oder leicht gefallen ist, an welchen Stellen sie sich mehr Unterstützung gewünscht hätten und ob ihnen etwas besonders gut oder schlecht gefallen hat. Auch werden sie nach möglichen Verbesserungsvorschlägen und nach ihrem persönlichen Lernzuwachs gefragt. Zum Schluss werden die SuS gebeten eine abschließende Schulnote für den CAMMP day und für die Dozenten zu geben.

5.4. Vorstellung der gesammelten Beobachtungen und der Ergebnisse aus der Evaluation

In diesem Abschnitt werden die Beobachtungen der ersten Durchführung sowie die Ergebnisse der Evaluation anhand der leitenden Gesichtspunkten vorgestellt. Auf Grund der kleinen Stichprobe von 22 SuS sei anzumerken, dass die hier aufgeführten Ergebnisse in keinsten Weise empirisch erforscht sind. Außerdem handelt es sich bei den getätigten Angaben in der Evaluation um Selbsteinschätzungen der SuS.

Interesse:

Bei der Durchführung des Lernmoduls war ein großes Interesse sowie eine hohe Motivation zu beobachten. Gleich zu Beginn der Einführungspräsentation haben die SuS interessiert zugehört. Auch haben sich viele SuS an den Diskussionen in dieser Präsentation beteiligt und ihre Gedanken sowie Erfahrungen zum Thema „Sicherheit der Privatsphäre in sozialen Netzwerken“ eingebracht. Auch in der Abschlusspräsentation am Ende des Lernmoduls war eine hohe Beteiligung zu verzeichnen. Die Äußerungen der SuS und die hohe Beteiligung lässt darauf schließen, dass sich viele SuS bereits schon mit der Thematik „Sicherheit der Privatsphäre in sozialen Netzwerken“ auseinandergesetzt haben und sich für diese Thematik interessieren. Auch die Lehrkraft teilte mit, dass sie es erstaunlich fand, wie viel die SuS bereits zu diesem Thema wussten. Während den Arbeitsphasen haben die SuS die Aufgaben größtenteils mit großem Eifer bearbeitet. Vier SuS mussten ab und zu motiviert werden an den Aufgaben weiter zu arbeiten.

Das hohe Interesse an der Thematik spiegelt auch die Evaluation wieder. Hier stimmten 76 % der SuS der Aussage voll zu, dass sie das Thema „Sicherheit der Privatsphäre in sozialen Netzwerken“ interessiert hat und etwa 20 % stimmten ihr zum Teil zu. Nur ein Teilnehmer bzw. eine Teilnehmerin gab an, dass sie dieser Aussage eher nicht zustimmen würde. Der Aussage „Der CAMMP day hat mir Spaß gemacht“ stimmten über 60 % voll und knapp 40 % zum Teil zu. Ebenfalls gaben 76 % der Personen an,

dass ihnen die Veranstaltung gut gefallen hat und knapp 15 % hat die Veranstaltung teilweise gut gefallen. Nur ein Schüler bzw. eine Schülerin gab an, dass ihm bzw. ihr die Veranstaltung nicht gut gefallen hat. Auch Aussagen zu der offenen Frage „Hat dir etwas an diesem Kurs besonders gut gefallen?“ wie „Das Thema“, „Das Thema war sehr aktuell und betrifft alle.“ oder die abschließenden Kommentare der SuS wie „Der Tag hat mit Spaß gemacht“ zeigen, dass ein hohes Interesse an der Thematik des Lernmoduls besteht. Die Lehrkraft gab ebenfalls an, dass die Aktualität und die Bedeutsamkeit der Thematik Stärken des CAMMP days sind.

Verständnis:

Ein Ziel des Lernmoduls ist es, den SuS ein Verständnis dafür zu vermitteln, wie mit Hilfe einfacher, mathematischer Regeln Vorhersagen über Informationen von Nutzern sozialer Netzwerke getätigt werden können, welche diese in den Profilen nicht angegeben haben. In der Evaluation stimmten etwa 62 % voll, 24 % zum Teil, 10 % eher nicht und 5 % gar nicht zu, am Ende des CAMMP days dieses Verständnis erlangt zu haben. Außerdem gaben etwa 57 % der SuS an, auf dem ersten Arbeitsblatt verstanden zu haben, warum der bereitgestellte Datensatz zunächst gefiltert werden muss. Etwa 38 % stimmten dieser Aussage zum Teil zu. Nur ein Schüler bzw. eine Schülerin war der Meinung, dass er bzw. sie den Grund der Filterung auf dem ersten Arbeitsblatt eher nicht verstanden hat. Diese Ergebnisse aus der Evaluation spiegeln auch die Beobachtungen während des CAMMP days wieder. In der Besprechung zum ersten Arbeitsblatt haben sich viele SuS beteiligt und sinnvolle Gründe für die Notwendigkeit einer Bereinigung des Datensatzes genannt. Auch konnten die SuS am Ende des CAMMP days noch weitere, sinnvolle Regeln aufstellen, mit denen das unbekanntes Alter eines Nutzers in einem sozialen Netzwerk vorhergesagt werden kann.

Das entwickelte Lernmodul soll ebenfalls bei den SuS den Umgang mit den vorkommenden Lage- und Streuungsmaße verbessern. Beim arithmetischen Mittel und der Spannweite gaben knapp 80 %, beim Quartilsabstand etwa 90 % und beim Modus knapp 75 % der SuS an, dass sich durch den CAMMP day der Umgang mit diesen mathematischen Begriffen verbessert hat. Laut der Evaluation kannten vor dem CAMMP day das arithmetische Mittel etwa 74 %, den Median 90 %, den Modus 42 %, den Quartilsabstand 47 % und die Spannweite 68 % der teilgenommenen SuS. Vor allem beim Modus und beim Quartilsabstand konnte durch das entwickelte Lernmodul ein besseres Verständnis bei den SuS hergestellt werden.

Zusammenfassend kann anhand der Evaluation festgehalten werden, dass die Ziele, den SuS die zentrale Idee hinter der Vorhersage von Informationen in sozialen Netzwerken zu vermitteln, ein besseres Verständnis für die in den Vorhersageregeln vorkommenden mathematischen Grundlagen herzustellen und den SuS die Idee der Aufbereitung eines Datensatzes zu veranschaulichen, größtenteils als erfüllt angesehen werden können.

Schwierigkeit der Aufgaben:

Auf Grund der hohen Heterogenität in der Klasse haben die SuS unterschiedliche Fortschritte im Arbeitsmaterial erzielt. Alle SuS haben das zweite Arbeitsblatt an dem Tag geschafft. 12 SuS fingen mit der Bearbeitung des dritten Arbeitsblatts an, wovon

4 SuS dieses Blatt beendet und mit der ersten Aufgabe auf dem Zusatzaufgabenblatt angefangen haben. Dass die SuS nicht alle das dritte Arbeitsblatt geschafft haben, ist jedoch keinesfalls schlimm, da sie bereits am Ende des zweiten Aufgabenblatts durch die Anwendung der Regeln zu den Lage- und Struungsmaße ein Verständnis für das Vorhersagen des Alters haben und den Modellierungskreislauf, wie im didaktisch-methodischen Konzept erläutert (siehe Abschnitt 4.3), zwei Mal durchliefen.

Im ersten Aufgabenblatt kamen alle SuS bis zur fünften Aufgabe, der zweiten Filterung, etwa gleich schnell voran. Die fünfte Aufgabe, bei der die SuS den Code für die zweite Filterung in die richtige Reihenfolge bringen sollen, hat etwa mehr als der Hälfte der SuS Schwierigkeiten bereitet. Hier mussten die betreuenden Dozenten bei vielen SuS unterstützend eingreifen. Der Hauptgrund für die aufgetretenen Schwierigkeiten bestand darin, dass die SuS vom Programm MATLAB keine genau Rückmeldung für die acht einzeln zu sortierenden Schritte erhalten haben. Es wurde ihnen lediglich angezeigt, wie viele der acht Schritte bereits insgesamt an der richtigen Stelle stehen. So konnten die SuS nicht genau feststellen, wo genau ihr Fehler lag.

Im zweiten Arbeitsblatt hat sich dann die Heterogenität der Klasse besonders gezeigt. Etwa ein Viertel der Klasse wusste nicht, wie das arithmetische Mittel zu berechnen ist. Beim Median war dies bei etwa der Hälfte der Klasse der Fall, obwohl in der Evaluation 90 % der SuS angegeben dieses Lagemaß bereits vor dem CAMMP day gekannt zu haben. Diese SuS mussten zunächst die bereitstehende Hilfekarte durchlesen. Beim Zeichnen des Boxplots auf dem zweiten Arbeitsblatt musste etwa die Hälfte der Klasse auf die dafür entwickelte Hilfekarte zurückgreifen. Ein Viertel der SuS hat dabei den vollen Umfang der Hilfekarte genutzt und jeden der dort aufgeführten Schritte zum Zeichnen eines Boxplots durchgearbeitet. Die Zusatzaufgabe auf dem zweiten Aufgabenblatt haben 6 SuS bearbeitet.

Auf dem dritten Aufgabenblatt hatten alle SuS, welche dieses bearbeitet haben, Schwierigkeiten beim Erstellen des zweistufigen Baumdiagramms. Dies lag vermutlich daran, dass diese Thematik noch nicht in der Schule besprochen wurde. Auch mit der dafür vorgesehenen Hilfekarte brauchten die SuS noch Unterstützung von den Dozenten.

Insgesamt sind die SuS jedoch gut mit dem entwickelten Material zurecht gekommen. Nur sehr selten haben SuS Verständnisfragen zu den Aufgaben auf den Arbeitsblättern gestellt. Auch in der Evaluation ist der freie Kommentar „Die Aufgabenstellung war sehr gut und gut zu verstehen“ zu finden. Ebenfalls gaben etwa 67 % der SuS in der Evaluation an, dass die Inhalte im Lernmodul klar vermittelt wurden. Knapp 20 % stimmten dieser Aussage zum Teil, 10 % eher nicht und eine Person gar nicht zu. Nach Angaben der Evaluation scheint auch der Schwierigkeitsgrad der Aufgaben angemessen. Nur vier SuS empfanden die Aufgaben größtenteils zu einfach und sieben Teilnehmer bzw. Teilnehmerinnen waren der Meinung, dass die Aufgaben größtenteils zu schwer waren.

Die Heterogenität der Klasse wurde durch den Einsatz der Hilfekarten gut aufgefangen. Mit Hilfe dieser Karten konnten die SuS selbstständig die Aufgaben lösen, bei denen sie Schwierigkeiten hatten. Ebenfalls konnten bei den SuS aufgekommene Fragen oft schon mit den Hilfekarten gelöst werden. Auch in den Kommentaren der Evaluation wurde diese Unterstützung positiv erwähnt. So schreiben die SuS beispielsweise: „Die

Hilfekarten haben einem sehr geholfen die Aufgaben zu lösen.“ oder „Es war immer Hilfe da, wenn man sie brauchte“. Der Aussage „Die Hilfekarten waren hilfreich“ stimmen etwa 57 % voll, 33 % zum Teil, 5 % eher nicht und 5 % gar nicht zu.

Modellieren und Mathematik:

Ein Ziel des entwickelten Lernmoduls wie auch vom Schülerlabor CAMMP ist es, einen Beitrag zum Aufbau der Modellierungskompetenz zu leisten und den SuS ein Verständnis für die Bedeutung von Modellieren sowie der Mathematik in der Gesellschaft zu vermitteln. In der Evaluation stimmten knapp 43 % der SuS zu, durch den CAMMP day das mathematische Modellieren besser verstanden zu haben. Etwa 38 % stimmten dieser Aussage zum Teil zu. Nur ein Schüler bzw. eine Schülerin gab an, gar kein besseres Verständnis für das Modellieren erhalten zu haben, und etwa 14 % sind der Meinung, nur eine geringe Verbesserung ihrer Modellierungskompetenz erfahren zu haben. Dies deckt sich auch mit den Beobachtungen. Am Ende des CAMMP day zeigte sich, dass ein Großteil der SuS die einzelnen Schritte des Modellierungskreislaufs verstanden hatten. Denn im Rahmen der Abschlusspräsentation sollten die SuS nochmal abschließend ihr Vorgehen anhand des Modellierungskreislaufs erklären. Dies hat sehr gut funktioniert. Die SuS konnten die einzelnen Aufgaben des Lernmoduls den jeweiligen Schritten im Kreislauf zuordnen.

Der Aussage „Durch den CAMMP day habe ich erkannt, dass Mathematik eine wichtige Rolle in der Gesellschaft hat.“ stimmten etwa 62 % voll, 33 % zum Teil, und 5 % eher nicht zu. Im Vergleich zu der Aussage „Ich wusste bereits vor dem CAMMP day, dass Mathematik in meinem Alltag eine wichtige Rolle spielt.“ (33 % trifft voll zu, 38 % trifft zum Teil zu, 14 % trifft eher nicht zu, und 14 % trifft gar nicht zu) ist bei den SuS durch das entwickelte Lernmodul ein stärkeres Bewusstsein für die Bedeutsamkeit der Mathematik in der Gesellschaft entstanden. Insgesamt kann für die evaluierte Schülergruppe festgehalten werden, dass beim Großteil der SuS die Modellierungskompetenz verbessert und ein Verständnis für die Bedeutung des mathematischen Modellierens sowie generell für die Mathematik hergestellt wurde.

Umgang mit MATLAB:

Für die SuS ist es natürlich zunächst schwierig, ein Programm zu bedienen, mit dem sie zuvor noch nicht in Berührung gekommen sind. Aus diesem Grund ist es wichtig, den Umgang mit MATLAB so einfach wie möglich zu gestalten. Daher bekommen die SuS vor dem ersten Umgang mit MATLAB eine kurze Einführung in das Programm, bei der ihnen die wichtigsten Funktionen für den Tag gezeigt werden. Diese Einführung wurde von den SuS in der Evaluation positiv angemerkt. Über 80 % gaben an, dass sie die Einführung in Matlab als hilfreich empfanden, und knapp 20 % stimmten dieser Aussage zum Teil zu. Während der Durchführung des CAMMP days traten nur sehr selten Verständnisfragen zu MATLAB oder zur Eingabe von Lösungen in das Programm auf. Diese Beobachtung zeigt sich ebenfalls in der Evaluation. Dort gaben etwa 33 % der SuS an, beim entwickelten Lernmodul keine Probleme gehabt zu haben, die Aufgaben mit MATLAB umzusetzen. Etwa 38 % stimmten dieser Aussage zum Teil zu. Nur 2 SuS hatten laut Evaluation ganzheitlich und 4 SuS teilweise Schwierigkeiten

beim Umgang mit MATLAB.

Sowohl die Lehrkraft als auch die SuS lobten in Form eines mündlichen Feedbacks die Rückmeldung des Programms MATLAB über die Korrektheit ihrer Lösung. Es wäre jeder Zeit klar, welche Aufgaben bereits korrekt sind oder an welchen Stellen ein Fehler aufgetreten ist. Auch die Icons im MATLAB Live Skript wurden positiv hervorgehoben. Sie hätten eine klare Abgrenzung von Aufgabenstellung und Eingabebereich der Lösung im Code geschaffen.

Leerlauf:

Auch bei einer heterogenen Schülergruppe soll verhindert werden, dass bei den SuS während der Bearbeitung des Lernmoduls Leerläufe vorkommen. Das heißt, sie sollen nicht auf SuS, die noch nicht mit der Bearbeitung der Aufgaben fertig sind, warten müssen. Waren SuS mit den Aufgaben auf einem Arbeitsblatt fertig, konnten sie die dazugehörige Zusatzaufgabe bearbeiten oder bereits mit dem nächsten Aufgabenblatt anfangen. Phasen, bei denen SuS keinen Arbeitsauftrag und damit einen Leerlauf hatten, wurden nicht beobachtet. Auch in der Evaluation gaben etwa 48 % der SuS an, während des CAMMP days keinen Leerlauf gehabt zu haben. Etwa 38 % stimmten dieser Aussage zum Teil und fast 10 % eher weniger zu. Nur ein Schüler bzw. eine Schülerin gab an, während des CAMMP days Leerlauf gehabt zu haben. Hier deckt sich die Evaluation nicht ganz mit den Beobachtungen. Vermutlich haben ein paar SuS auch Phasen in denen sie nicht weiter kamen und auf die Hilfe eines Dozenten warteten oder selbst eingelegte Pausen als Leerlauf interpretiert und in der Evaluation angegeben.

Fazit:

Insgesamt kann die erste Durchführung des entwickelten Lernmoduls als gelungen angesehen werden. Der programmierte MATLAB-Code lief bis auf einen kleinen Fehler reibungslos durch, die SuS haben die Aufgabenstellungen auf den Arbeitsblättern verstanden und hatten nur selten Schwierigkeiten die Aufgaben mit MATLAB zu bearbeiten. Durch die Rückmeldung von MATLAB, an welchen Stellen die SuS einen Fehler begangen haben, und den entwickelten Hilfekarten konnten die SuS selbstständig an den Aufgaben arbeiten und mussten nur selten einen Betreuer um Hilfe bitten. Besonders positiv ist die Relevanz und das Interesse der SuS an das Thema „Sicherheit der Privatsphäre in sozialen Netzwerken“ hervorzuheben. Dies spiegelt sich zum einen in der Evaluation wieder, zeigte sich andererseits aber auch an den angeregten Diskussionen zu dieser Thematik während der Einführungs- und Abschlusspräsentation. Am Ende der Evaluation konnten die SuS dem entwickelten Lernmodul eine Schulnote geben. 21 von den 22 Teilnehmern und Teilnehmerinnen sind dieser Aufforderung nachgekommen. Die Note *sehr gut* wurde 12 Mal, die Note *gut* 6 Mal, die Note *befriedigend* 2 Mal und die Note *ausreichend* 1 Mal vergeben. Insgesamt ergibt sich damit ein Notendurchschnitt von 1.6, was als ein sehr gutes Resultat für die erste Durchführung angesehen werden kann.

5.5. Aus der Evaluation und Beobachtung resultierende Verbesserungen zum entwickelten Lernmodul

Ausgehend von den Ergebnissen der Beobachtungen und der Evaluation der ersten Durchführung wurden Verbesserungen im entwickelten Lernmodul vorgenommen. Auf dem ersten Arbeitsblatt wurde zunächst der einzige, aufgetretene Fehler im programmierten MATLAB Code behoben. Dort konnten die SuS im Rahmen der dritten Aufgabe bei der ersten Filterung, in der Nutzer mit einem vermeintlich falschen Alter aus dem Datensatz aussortiert werden, die obere Altersgrenze nicht über 90 Jahre wählen. Haben die SuS ein Alter über 90 Jahre gewählt, wurde ihnen eine Fehlermeldung angezeigt und die Filterung nicht durchgeführt. Dieser Fehler im Code ist nun behoben. Ebenfalls wurde die zweite Filterung auf dem ersten Arbeitsblatt überarbeitet. Dort wurden im ersten Teil der fünften Aufgabe die einzelnen Schritte, die den Code hinter der zweiten Filterung mit Worten beschreiben und welche die SuS in die richtige Reihenfolge bringen sollen, ausführlicher bzw. klarer formuliert. Auch findet dort nun eine differenziertere Rückmeldung über MATLAB statt. Vorher wurde den SuS nur angezeigt, wie viele Schritte sich insgesamt in der richtigen Reihenfolge befinden. Da die Aufgabe viel Zeit in Anspruch nahm und einige SuS Schwierigkeiten bei der Sortierung hatten, zeigt MATLAB den SuS nun für jeden einzelnen Schritt an, ob dieser korrekt ist.

Für das zweite Arbeitsblatt wurde auf dem zugehörigen Antwortblatt die Formel für die Trefferquote notiert. Zwar wird die Formel im Rahmen des Zwischenvortrags mit den SuS hergeleitet und besprochen, jedoch ist während der ersten Durchführung des CAMMP days aufgefallen, dass die Formel für die Trefferquote bei ein paar SuS bei der Bearbeitung des zweiten Arbeitsblatts nicht mehr präsent war. Diese ist nun gleich oben auf dem Antwortblatt zu finden. Außerdem wird sie den SuS nun auch nochmal im Live Skript, wenn sie die Formel zum ersten mal benötigen, angezeigt. So soll sichergestellt sein, dass alle SuS die benötigten Voraussetzungen zur Bearbeitung der Aufgaben besitzen.

In der ersten Aufgabe des zweiten Arbeitsblatts, bei der die SuS das arithmetische Mittel, den Median und den Modus zu den Altersangaben der Freunde 1. Ordnung eines Nutzers bestimmen sollen, haben viele SuS, wie in den bereits vorgestellten Vorhersageregeln, in MATLAB die auf eine ganze Zahl gerundeten Werte eingetragen. Da das Programm jedoch den exakten Wert überprüft, wurde den SuS dafür ein Fehler rückgemeldet. Dies führte bei den SuS zur Verwirrung, da sie im vorherigen Vortrag mitgeteilt bekommen haben, nur die gerundeten Werte zu betrachten. Der Code für diese Aufgabe wurde überarbeitet. Nun werden auch gerundete Werte akzeptiert und von MATLAB als korrekt rückgemeldet.

Weiter wurde im zweiten Arbeitsblatt bei der dritten Aufgabe die Anzahl zu überprüfender Nutzer reduziert. Um mit den Vorhersageregeln vertraut zu werden, sollen die SuS in dieser Aufgabe die Regeln zunächst auf einzelne Nutzer anwenden und überprüfen, ob für die ausgewählten Nutzer das Alter mit den gewählten Regeln richtig vorhergesagt wird. Da die SuS in der ersten Durchführung mit den Vorhersageregeln zu den Lagemaßen sehr gut zurecht kamen, sollen sie nun nur noch zwei statt vier

Nutzer betrachten. So soll außerdem Zeit gespart und aufkommende Langeweile vorgebeugt werden.

Eine weitere Änderung wurde in Aufgabe 6b) auf dem zweiten Arbeitsblatt vorgenommen. Hier sollen die SuS angeben, wie viel Prozent der Daten bei einem Boxplot mindestens innerhalb der Box liegen. Etwas mehr als die Hälfte der SuS hatten bei dieser Aufgabe Schwierigkeiten. Zur Unterstützung wird den SuS nun zunächst eine Definition des unteren und oberen Quartils gegeben. Aus diesen Definitionen können die SuS ableiten, wie viel Prozent der Werte unterhalb sowie oberhalb des Kastens im Boxplot liegen.

Die letzte Änderung besteht in der Auswahl der beiden entwickelten Niveaus. Angedacht war, Niveau 1 für eine siebte Klasse und Niveau 2 für eine achte oder höhere Klasse zu wählen. Nach der ersten Durchführung hat sich jedoch gezeigt, dass die SuS der achten Klasse Schwierigkeiten bei der Erstellung eines zweistufigen Baumdiagramms hatten, da diese Thematik noch nicht im Unterricht behandelt wurde. Niveau 2 sollten demnach nur Klassen bearbeiten, welche bereits zweistufige Zufallsexperimente und Baumdiagramme im Unterricht besprochen haben. Wenn Lehrkräfte das entwickelte Lernmodul mit ihrer Klasse durchführen wollen, müssen sie eine Auskunft darüber geben, ob diese Themen bereits behandelt wurden. Das Niveau wird dann ausgehend von der Angabe der Lehrkraft ausgewählt. Die neue Aufteilung der Niveaus wurde in das methodische Konzept eingearbeitet.

6. Ausblick

Im letzten Abschnitt dieser Masterarbeit sollen noch Möglichkeiten zur Weiterentwicklung und Optimierung des Lernmoduls aufgezeigt werden.

Zunächst sollte sich darum bemüht werden, den SuS einen größeren Datensatz zur Verfügung zu stellen. Im Abschnitt 3.2.4 wurde bereits gezeigt, dass die durchschnittliche Anzahl an Freunden 1. Ordnung im momentanen Datensatz bei 3.68 liegt. Diese Zahl spiegelt nicht ansatzweise die reale Situation in aktuellen sozialen Netzwerken wieder. Zudem würde eine größere Anzahl an Freunden 1. Ordnung eventuell zu genaueren Trefferquoten bei den Vorhersageregeln zu den Lage- und Struungsmaßen führen. Leider gab es Schwierigkeiten, eine größere Datenmenge an Nutzerprofilen des sozialen Netzwerks Friendster herunterzuladen. An dieser Problematik wird aber momentan gearbeitet. Das Lernmodul ist so konzipiert, dass der verwendete Datensatz jeder Zeit ohne Probleme ausgetauscht werden kann. Sobald ein größerer Datensatz vorliegt, sollte dieser in das Lernmodul eingebunden und die Trefferquoten der aufgestellten Vorhersageregeln erneut überprüft werden.

Eine weitere Möglichkeit zur Weiterentwicklung des Lernmoduls besteht darin, noch neue Vorhersageregeln aufzustellen und auf ihre Effektivität zu überprüfen. Beispielsweise sind in den heruntergeladenen Nutzerprofilen auch die Interessen der Nutzer gespeichert. Hier könnte untersucht werden, ob gewisse Interessen mit bestimmten Altersgruppen korrelieren und daraus eventuell neue Vorhersageregeln abgeleitet werden können. Auch könnten die bereits bestehenden Vorhersageregeln miteinander kombiniert werden, um eventuell noch bessere Trefferquoten zu erzielen.

Wie in der Evaluation bereits angesprochen, hatten die SuS Schwierigkeiten bei der Erstellung des zweistufigen Baumdiagramms auf dem dritten Arbeitsblatt, da diese Thematik noch nicht im Unterricht behandelt wurde. Bei einer erneuten Durchführung des entwickelten Lernmoduls mit einer Klasse, die bereits Baumdiagramme im Unterricht thematisiert hat, sollte beobachtet werden, wie diese SuS mit der Aufgabe zurecht kommen. Fällt vielen SuS die Bearbeitung der Aufgabe trotzdem noch schwer, könnte diese aus dem regulären Aufgabenpool herausgenommen und als Zusatzaufgabe eingesetzt werden. An diese Aufgabe könnten sich dann die leistungsstärkeren SuS wagen.

Während des CAMMP days gab es eine kleine Schülergruppe von vier SuS, welche nur sehr schwer für die Mathematik hinter dem Lernmodul zu begeistern war. Diese vier SuS kamen nur langsam mit den Aufgaben voran und mussten öfters zum Weiterarbeiten motiviert werden. Ein kurzes Gespräch mit den SuS am Ende des CAMMP days hat ergeben, dass sie zwar an der Thematik des Lernmoduls interessiert waren, sich jedoch lieber weniger stark mathematisch mit dem Thema auseinandergesetzt hätten. Hier wäre zu überlegen, den SuS ab einem gewissen Punkt (beispielsweise nach der Anwendung der Vorhersageregeln zum exakten Alter auf dem zweiten Arbeitsblatt) die Wahl zu geben, ob sie weitere Vorhersageregeln mathematisch erarbeiten oder lieber Informationen zum Thema „Sicherheit der Privatsphäre in sozialen Netzwerken“ recherchieren möchten. Dazu müsste ein Arbeitsblatt erstellt werden, welches die SuS bei der Recherche anleitet, sodass ein zielloses Suchen im Internet vermieden wird. Diese

Informationen könnten die SuS dann in einer kurzen Präsentation vorbereiten und den anderen Teilnehmern bzw. Teilnehmerinnen am Ende des CAMMP days vorstellen. Durch diese Wahl würde auch eine Differenzierung nach dem Interesse stattfinden und den SuS eine Möglichkeit zur Mitbestimmung gegeben.

Ein aktuelles Ziel vom Schülerlabor CAMMP ist es, authentische Modellierungsaufgaben in die Schulen zu bringen, sodass SuS und Lehrkräfte auch in den Schulen an realen Fragestellungen arbeiten können. Sowohl die Thematik als auch die im entwickelten Lernmodul vorkommenden mathematischen Kompetenzen, welche auch im Kernlehrplan für die Sekundarstufe I zu finden sind, eignen sich für den Schuleinsatz. Denkbar wäre es im Rahmen einer weiteren Abschlussarbeit aufbauend auf dieser Masterarbeit Material, für eine Unterrichtreihe zu den Themen wie absolute und relative Häufigkeit, arithmetisches Mittel, Median, Quartilsabstand, Spannweite und der Erstellung eines Boxplots zu entwickeln.

Abschließend kann festgehalten werden, dass das Ziel der Arbeit, ein Lernmodul zum Thema „Sicherheit der Privatsphäre in sozialen Netzwerken“ zu entwickeln, als erreicht angesehen werden kann. Die Materialien haben sich im Rahmen der ersten Durchführung bewährt und können nach den bereits umgesetzten Verbesserungen weiteren Schülergruppen angeboten werden.

Literatur

- Aebli, H. (1933). *Zwölf Grundformen des Lehrens. Eine Allgemeine Didaktik auf psychologischer Grundlage*. Stuttgart: Klett-Cotta.
- Barzel, B., Elschenbroich, H. J., Hefendehl-Hebeker, L., Heintz, G., Heske, H., Hußmann, S., Lambert, A., Leuder, T., Ludwig, M., von Saint-George, G. & Westermann, B. (2011). *Mathematik Didaktik - Praxishandbuch für die Sekundarstufe I und II* (Cornelsen, Hrsg.).
- Büchter, A. & Leuders, T. (2005). *Mathematikaufgaben selbst entwickeln, Lernen fördern - Leistung überprüfen*. Berlin: Cornelsen.
- Blum, W. (2006). Modellierungsaufgaben im Mathematikunterricht - Herausforderungen für Schüler und Lehrer. In A. Büchter, H. Humenberger, S. Hussmann & S. Prediger (Hrsg.), *Realitätsnaher Mathematikunterricht - vom Fach aus und für die Praxis. Festband für Hans-Wolfgang Henn zum 60. Geburtstag*. Hildesheim: Franzbecker.
- Borromeo Ferri, R., Greefrath, G. & Kaiser, G. (2013). *Mathematisches Modellieren für Schulen und Hochschule. Theoretische und didaktische Hintergründe*. Wiesbaden: Springer Verlag.
- Bruder, R., Hefendehl-Hebeker, L., Schmidt-Thieme, B. & Weigand, H.-G. (2015). *Handbuch der Mathematikdidaktik*. Heidelberg: Springer Spektrum.
- Bruder, R. & Reibold, J. (2010). Weil jeder anders lernt. Ein alltagstaugliches Konzept zur Binnendifferenzierung. In *Mathematik lehren, Heft 162*. (S. 2-9). Seelze: Friedrich.
- CaMMP Team. (o.J.). *CaMMP Webseite*. Zugriff auf <https://blog.rwth-aachen.de/campp/> (letzter Aufruf am 18.09.2018)
- Cleff, T. (2008). *Deskriptive Statistik und moderne Datenanalyse. Eine computergestützte Einführung mit Excel, SPSS und STATA*. Wiesbaden: Gabler.
- Cramer, E. & Kamps, U. (2017). *Grundlagen der Wahrscheinlichkeitsrechnung und Statistik. Eine Einführung für Studierende der Informatik, der Ingenieur- und Wirtschaftswissenschaften* (4. Aufl.). Berlin, Heidelberg: Springer Spektrum.
- Deci, E. L. & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. In *Zeitschrift für Pädagogik, Heft 39*. (S. 223-238). Weinheim: Beltz.
- Duit, R. & Wodzinski, C. T. (2006). Guten Unterricht planen – Kategorien fachdidaktischen Denkens bei der Planung von Unterricht. In *Naturwissenschaften im Unterricht Physik, Heft 92*. (S. 9-11). Seelze: Friedrich.

- Engel, J. (2018). *Anwendungsorientierte Mathematik: Von Daten zur Funktion*. (2. Aufl.). Berlin: Springer Spektrum.
- Feierabend, S., Plankenhorn, T. & Rathgeb, T. (2017). *JIM-Studie 2017. Jugend, Information, (Multi-) Media*. Stuttgart: Medienpädagogischer Forschungsverband Südwest.
- Frank, M., Richter, P., Roeckerath, C. & Schönbrodt, S. (2018). Wie funktioniert eigentlich GPS? - ein computergestützter Modellierungsworkshop. In G. Greefrath & H.-S. Siller (Hrsg.), *Digitale Werkzeuge, Simulationen und mathematisches Modellieren. Didaktische Hintergründe und Erfahrungen aus der Praxis*. (S. 137-164). Wiesbaden: Springer Spektrum.
- Friendster Inc. (o.J.). *Webseite*. Zugriff auf <http://www.friendster.com/> (letzter Aufruf am 17.08.2018)
- Frischemeier, D. (2017). *Statistisch denken und forschen lernen mit der Software TinkerPlots*. Wiesbaden: Springer Spektrum.
- Garcia, D., Sargiol, E. & Schweitzer, F. (2014). *Online Privacy as a Collective Phenomenon*. Zugriff auf <https://arxiv.org/pdf/1409.6197v1.pdf> (6.6.2016) (letzter Aufruf am 17.05.2018)
- Greefrath, G. & Weitendorf, J. (2013). Modellieren mit digitalen Werkzeugen. In Borromeo Ferri, R., Greefrath, G. & Kaiser, G. (Hrsg.), *Mathematisches Modellieren für Schulen und Hochschule. Theoretische und didaktische Hintergründe* (S. 181-201). Wiesbaden: Springer Verlag.
- Grimm, P., Keber, T. O. & Zöllner, O. (2015). Anonymität und Transparenz in der digitalen Gesellschaft. Einleitende Bemerkungen zur Digitalen Ethik . In P. Grimm, T. O. Keber & O. Zöllner (Hrsg.), *Anonymität und Transparenz in der digitalen Gesellschaft*. (S. 7-20). Stuttgart: Franz Steiner.
- Gudjons, H. (2005). Methoden und Strategien intelligenten Übens. In *Pädagogik, Heft 11 / 2005*. (S. 12-15). Weinheim: Beltz.
- Hartmann, P. (2015). *Mathematik für Informatiker* (6. Aufl.). Wiesbaden: Springer Fachmedien.
- Heymann, H. W. (1998). Üben und Wiederholen – neu betrachtet. In *Pädagogik, Heft 10 / 1998*. (S. 5-11). Weinheim: Beltz.
- Heymann, H. W. (2012). Schüler beim Aufbau von Kompetenzen unterstützen. Üben, Anwenden, Vertiefen – Gelingensbedingungen für nachhaltiges Üben. In *Pädagogik, Heft 12 / 2012*. (S. 5-11). Weinheim: Beltz.
- Hinrichs, G. (2008). *Modellierung im Mathematikunterricht*. Heidelberg: Spektrum Akademischer Verlag.

- Kaiser, G., Blum, W., Ferri, R. B. & Greefrathi, G. (2015). Anwendung und Modellieren. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme & H.-G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik*. (S. 357-384). Berlin, Heidelberg: Springer Spektrum.
- Klanert, J. (2014). *Wandel durch das Internet: Das Online-Verhalten der User wird immer komplexer*. Zugriff auf https://www.focus.de/digital/experten/klanert/wandel-durch-das-internet-das-online-verhalten-der-user-wird-immer-komplexer_id_4078023.html (letzter Aufruf am 14.08.2018)
- Krüger, K., Sill, H.-D. & Sikora, C. (2015). *Didaktik der Stochastik in der Sekundarstufe I*. Berlin and Heidelberg: Springer Spektrum.
- Kultusministerkonferenz der Länder in der Bundesrepublik Deutschland. (2004). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss*. München: Luchterhand.
- Kultusministerkonferenz der Länder in der Bundesrepublik Deutschland. (2015). *Bildungsstandards im Fach Mathematik für die Allgemeine Hochschulreife*. München: Carl Link.
- Loos, A. & Ziegler, G. M. (2015). Gesellschaftliche Bedeutung der Mathematik. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme & H.-G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik*. (S. 3-18). Berlin and Heidelberg: Springer Spektrum.
- Maaß, K. (2011). *Mathematisches Modellieren. Aufgaben für die Sekundarstufe I*. (5. Aufl.). Berlin: Cornelsen Scriptor.
- McMillan, R. (2013). *The Friendster Autopsy: How a Social Network Dies*. Zugriff auf <https://www.wired.com/2013/02/friendster-autopsy/> (letzter Aufruf am 14.08.2018)
- Meyer, H. (2004). *Was ist guter Unterricht?* (12. Aufl.). Berlin: Cornelsen Scriptor.
- Meyers Enzyklopädisches Lexikon (Band 11)* (9. Aufl.). (1974). Mannheim and Wien and Zürich: Bibliographisches Institut AG.
- Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen. (2007). *Kernlehrplan für das Gymnasium - Sekundarstufe I (G8) in Nordrhein-Westfalen - Mathematik*. Frechen: Ritterbach.
- Mittag, H.-J. (2017). *Statistik. Eine Einführung mit interaktiven Elementen*. (5. Aufl.). Berlin and Heidelberg: Springer Spektrum.
- Serve, H. J. (2011). Das Unterrichtsprinzip der Übung. . In N. Seibert & H. J. Serve (Hrsg.), *Prinzipien guten Unterrichts. Kriterien einer zeitgemäßen Unterrichtsgestaltung*. (S. 221-243). München: PrimS.

- Sube, M. (2016). *Wie sicher ist meine Privatsphäre in Online Netzwerken? ...und was hat das mit Mathe zu tun?* Zugriff auf <https://blog.rwth-aachen.de/campp/files/2016/10/thesis-soziale-netzwerke.pdf> (letzter Aufruf am 20.08.2018)
- Tittmann, P. (2014). *Einführung in die Kombinatorik* (2. Aufl.). Berlin, Heidelberg: Springer Spektrum.
- Westermann, B. (2011). Anwendungen und Modellbildung. In T. Leuders (Hrsg.), *Mathematikdidaktik. Praxishandbuch für die Sekundarstufe I und II*. Berlin: Cornelsen.
- WhatsApp Inc. (2018). *Aktualisierte WhatsApp Nutzungsbedingungen und Datenschutzrichtlinie*. Zugriff auf <https://www.whatsapp.com/legal/?eea=1#privacy-policy-information-we-collect> (letzter Aufruf am 17.07.2018)
- Wild, C. J. & Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. In *International Statistical Review / Revue Internationale de Statistique* (S. 223-248). Auckland: International Statistical Institute.
- Winter, H. (1995). Acht Thesen zum allgemeinbildenden Mathematikunterricht. Eine komprimierte Zusammenfassung der Habilitationsschrift Allgemeinbildung und Mathematikunterricht. In *Mitteilung der Gesellschaft für Didaktik der Mathematik, Nr. 61* (S. 37-46).
- Wodzinski, R. (2013). Lernen mit gestuften Hilfen. In *Physik Journal 03 / 2013*. (S. 45-49). Weinheim: Wiley-VCH.

Anhang

A. Begriffe aus der Wahrscheinlichkeitsrechnung

Folgende Definitionen sind von (Cramer & Kamps, 2017) entnommen:

Definition Ereignismenge (vgl. Cramer & Kamps, 2017, S.160):

Die Menge aller möglichen Ereignisse eines Zufallsvorgangs (Zufallsexperiments) wird Ereignismenge (oder Grundraum) genannt und mit dem Buchstaben Ω bezeichnet.

Definition σ -Algebra (Cramer & Kamps, 2017, S.174):

Seien $\Omega \neq \emptyset$ und $\Sigma \subseteq \text{Pot}(\Omega)$ ein System von Teilmengen von Ω . Σ heißt σ -Algebra von Ereignissen über Ω , falls gilt:

- (i) $\Omega \in \Sigma$
- (ii) $A \in \Sigma \Rightarrow A^c \in \Sigma$ für jedes $A \in \Sigma$
- (iii) für jede Folge A_1, A_2, \dots von Mengen aus Σ gilt: $\bigcup_{n=1}^{\infty} A_n \in \Sigma$.

Definition Wahrscheinlichkeitsmaß (vgl. Cramer & Kamps, 2017, S.174):

Sei Σ eine σ -Algebra über $\Omega \neq \emptyset$. Eine Abbildung $P : \Sigma \rightarrow [0, 1]$ mit

- (i) $P(A) \geq 0 \forall A \in \Sigma$
- (ii) $P(\Sigma) = 1$ und
- (iii) $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$ für jede Wahl paarweise disjunkter Mengen aus Σ

heißt Wahrscheinlichkeitsmaß auf Ω .

B. Methodisches Konzept

Methodisches Konzept für die Dozenten

Wichtig: Es wurden zwei verschiedene Niveaus für diesen CAMMP day entwickelt. *Niveau 1* richtet sich an SuS ab einer siebten Klasse und *Niveau 2* an SuS, die bereits mehrstufige Zufallsexperimente und Baumdiagramme im Unterricht behandelt haben. Das Material für die beiden Niveaus unterscheidet sich jedoch nur im dritten Aufgabenblatt. Für Niveau 2 wurde für dieses Aufgabenblatt zudem noch ein extra Zusatzaufgabenblatt erstellt. Der Ablauf des CAMMP days ist für beide Niveaus gleich. Die Unterschiede im Material sind für Niveau 2 in rot gekennzeichnet.

Für den CAMMP day werden folgende Materialien benötigt:

Arbeitsblätter:

- Regeln 1-6.pdf (sind einlamiert)
- Regeln 7-10.pdf (sind einlamiert)
- AB 1_Daten filtern.pdf
- AB 2_Vorhersageregeln anwenden.pdf
- AB 3_Werbung schalten.pdf (für Niveau 1 und Niveau 2 unterschiedlich)
- AB 3_Zusatzaufgaben.pdf (nur für Niveau 2)

Folien für die Besprechungen:

- AB 1_Daten filtern_Besprechung.pdf
- AB 2_Vorhersageregeln anwenden_Besprechung.pdf
- AB 3_Werbung schalten_Besprechung.pdf (für Niveau 1 und Niveau 2 unterschiedlich)

Präsentationen:

- opening presentation.pdf
- closing presentation.pdf
- Berufs-Studieninformation presentation.pdf
- Modellierungsvortrag.pdf (wird vom Doktoranden organisiert)
- Einfuehrungsvortrag_Soziale Netzwerke.pdf
- Abschlussvortrag_Soziale Netzwerke.pdf

Hinweise zu den Präsentationen:

- Einfuehrungspräsentation Notizen.pdf
- Abschlusspräsentation Notizen.pdf

Lösungen für die Dozenten:

- AB 1_Daten filtern_Loesungen.pdf
- AB 2_Vorhersageregeln anwenden_Loesungen.pdf
- AB 3_Werbung schalten_Loesungen.pdf

Sonstige Materialien:

- Folienstifte
- OHP

Ablauf:

- | | | |
|----------|--------------------------|--|
| 8:15 | Vorbereitung | <input type="checkbox"/> Beginn der Vorbereitungen: Laptops in den Raum bringen, WLAN aktivieren, Material in den Raum bringen |
| | | <input type="checkbox"/> Matlab-Skript AB_1_Daten_filtern.mlx auf den Laptops öffnen. Command Window minimieren, da es nicht gebraucht wird. |
| | | <input type="checkbox"/> AB_1_Datei_filtern.pdf bereitlegen. |
| | | <input type="checkbox"/> opening_presentation.pdf auf Hiwi-Laptop öffnen |
| | | <input type="checkbox"/> Einführungsvortrag.pdf auf Hiwi-Laptop öffnen |
| | | <input type="checkbox"/> Modellierungsvortrag auf Hiwi-Laptop öffnen |
| 9:00 | Begrüßung | <input type="checkbox"/> Vortrag_OpeningPresentation.pdf halten |
| | Modellierungsvortrag | <input type="checkbox"/> wird von einem Doktoranden gehalten |
| 9:20 | Einstieg in des Projekt | <input type="checkbox"/> Einführungsvortrag.pdf halten, Notizen stehen zur Verfügung |
| | | <input type="checkbox"/> Vortrag bei Folie 13 pausieren und erste Arbeitsphase starten |
| | | <input type="checkbox"/> Matlab Einführung mit den SuS gemeinsam durchgehen und den SuS zeigen, wie sie die Hilfekarten und Aufgabenblätter in Matlab öffnen können |
| 9:30 | Partnerarbeit (Teil I) | <input type="checkbox"/> AB_1_Daten_filtern.pdf austeilen |
| 10:00 | | <input type="checkbox"/> Besprechung mit Folie AB_1_Daten_filtern_Besprechung.pdf |
| | | <input type="checkbox"/> Einführungsvortrag ab Folie 14 fortführen. |
| 10:15 | kurze Pause | |
| 10:30 | Partnerarbeit (Teil II) | <input type="checkbox"/> Matlab-Skript AB_2_Vorhersageregeln_testen.mlx von den SuS öffnen lassen. |
| | | <input type="checkbox"/> AB_2_Vorhersageregeln_testen.pdf austeilen |
| 11:15 | | <input type="checkbox"/> Besprechung der 2. Arbeitsphase mit Folie AB_2_Vorhersageregeln_anwenden_Besprechung.pdf ggf. nur bis zum Punkt "Neue Regeln", die restliche Besprechung erfolgt dann nach der Mittagspause |
| 12:00 | Mittagspause | |
| 13:00 | Partnerarbeit (Teil III) | <input type="checkbox"/> Matlab-Skript AB_3_Werbung_schalten.mlx von den SuS öffnen lassen |
| | | <input type="checkbox"/> AB_3_Werbung_schalten.pdf austeilen |
| parallel | Powerwall | <input type="checkbox"/> 3D-Simulation in der Gesundheitsforschung |
| 14:20 | | <input type="checkbox"/> Besprechung der 3. Arbeitsphase mit Folie AB_3_Werbung_schalten_Besprechung.pdf |
| 14:30 | Reflexion | <input type="checkbox"/> Abschlusspräsentation.pdf halten mit Präsentation Abschluss Notizen.pdf und die Diskussion mit den SuS durchführen |
| 14:45 | Werbeblock | <input type="checkbox"/> closing_presentation.pdf halten |
| 14:50 | Evaluation | <input type="checkbox"/> Mit der aktuellen CAMMP Evaluation |

C. Einführungsvortrag

C.1. Folien zum Einführungsvortrag

 <p>Sicherheit der Privatsphäre in sozialen Netzwerken - Was Facebook, Instagram und Co. über uns wissen</p> <p>Lehrstuhl für Mathematik Center for Computational Engineering Science</p> 	<h3>Kurzer Überblick</h3> <p>Wer von euch ist Mitglied in sozialen Netzwerken?</p>  <p><small>CAMMP day Sicherheit der Privatsphäre in sozialen Netzwerken - Was Facebook, Instagram und Co. über uns wissen 2/25</small></p>				
<h3>Nutzeranzahl bekannter sozialer Netzwerke (stand 2018)</h3> <table border="1"><tr><td>Facebook  2,2 Milliarden Nutzer</td><td>Instagram  800 Millionen Nutzer</td></tr><tr><td>WhatsApp  1,3 Milliarden Nutzer</td><td>Tik Tok  500 Millionen Nutzer</td></tr></table> <p><small>CAMMP day Sicherheit der Privatsphäre in sozialen Netzwerken - Was Facebook, Instagram und Co. über uns wissen 3/25</small></p>	Facebook  2,2 Milliarden Nutzer	Instagram  800 Millionen Nutzer	WhatsApp  1,3 Milliarden Nutzer	Tik Tok  500 Millionen Nutzer	<h3>Was denkt ihr?</h3> <p>Wie sicher ist eure Privatsphäre in sozialen Netzwerken?</p>  <p><small>CAMMP day Sicherheit der Privatsphäre in sozialen Netzwerken - Was Facebook, Instagram und Co. über uns wissen 4/25</small></p>
Facebook  2,2 Milliarden Nutzer	Instagram  800 Millionen Nutzer				
WhatsApp  1,3 Milliarden Nutzer	Tik Tok  500 Millionen Nutzer				
<h3>Was denkt ihr?</h3> <p>Wie sicher ist eure Privatsphäre in sozialen Netzwerken?</p>  <ul style="list-style-type: none">• Was bedeutet für euch Sicherheit in sozialen Netzwerken? <p><small>CAMMP day Sicherheit der Privatsphäre in sozialen Netzwerken - Was Facebook, Instagram und Co. über uns wissen 4/25</small></p>	<h3>Was denkt ihr?</h3> <p>Wie sicher ist eure Privatsphäre in sozialen Netzwerken?</p>  <ul style="list-style-type: none">• Was bedeutet für euch Sicherheit in sozialen Netzwerken?• Welche Informationen gebt ihr in einem sozialen Netzwerk über euch preis? Welche nicht? <p><small>CAMMP day Sicherheit der Privatsphäre in sozialen Netzwerken - Was Facebook, Instagram und Co. über uns wissen 4/25</small></p>				

<p>Wie sicher sind eure Daten?</p>  <p>Zeigt her eure Daten Sociale Netzwerke zeigen, was wir tun, was wir kaufen, alle Sicherheitsverletzungen. Wie können wir das Internet? Nicht alle. Angenommen, alle Sicherheitsverletzungen. Vor allem nicht in sozialen Netzwerken wie Facebook oder Twitter. Ein Malprogramm spürt den Benutzer aus und der Benutzer denkt, dass es nicht ist.</p> <p>Bis zu 87 Millionen Facebook-Nutzer betroffen Offiziell hat die Firma Databricks Analytics mehr Facebook-Nutzerinnen als bisher bekannt. Auch in Deutschland gibt es möglicherweise rund 200.000 Geschädigte.</p> <p>Diese Nutzerdaten sammeln Facebook und Twitter Sociale Netzwerke sind nur vorwiegend kostenlos. Tatsächlich zahlt der Nutzer mit seinen Daten. Große Anbieter wie Facebook oder Twitter ergötzen über die welche Informationen sie über einen sammeln.</p> <p>Kann man Facebook noch vertrauen? Viel Kommunikation, Millionen angelegter Profile und ein Internetauftritt, der nicht als Opfer steht. Facebook straschen, warum? Ein Blick in die Architektur des Netzwerks.</p> <p>Datenschutz ↑ Datensicherheit ↓</p> <p><small>CAMMP day Sicherheit der Privatsphäre in sozialen Netzwerken - Was Facebook, Instagram und Co. über uns wissen 5/25</small></p>	<p>Was denkt ihr?</p> <p>Welches Interesse hat eine Firma oder ein soziales Netzwerk an eure Daten?</p>  <p><small>CAMMP day Sicherheit der Privatsphäre in sozialen Netzwerken - Was Facebook, Instagram und Co. über uns wissen 6/25</small></p>
<p>Was denkt ihr?</p> <p>Welches Interesse hat eine Firma oder ein soziales Netzwerk an eure Daten?</p>  <ul style="list-style-type: none"> • Welche Informationen wären von besonderer Interesse? <p><small>CAMMP day Sicherheit der Privatsphäre in sozialen Netzwerken - Was Facebook, Instagram und Co. über uns wissen 6/25</small></p>	<p>Was denkt ihr?</p> <p>Welches Interesse hat eine Firma oder ein soziales Netzwerk an eure Daten?</p>  <ul style="list-style-type: none"> • Welche Informationen wären von besonderer Interesse? • Wie findet ihr die Tatsache, dass eure Daten analysiert werden? <p><small>CAMMP day Sicherheit der Privatsphäre in sozialen Netzwerken - Was Facebook, Instagram und Co. über uns wissen 6/25</small></p>
<p>Betreiber eines sozialen Netzwerks können Schattenprofile erstellen</p> <div style="border: 1px solid black; padding: 5px; margin: 10px 0;"> <p>Was ist ein Schattenprofil?</p> <p>Ein Profil für einen Nutzer eines Netzwerks, das Informationen über den Nutzer enthält, die dieser gar nicht angegeben hat.</p> </div> <p><small>CAMMP day Sicherheit der Privatsphäre in sozialen Netzwerken - Was Facebook, Instagram und Co. über uns wissen 7/25</small></p>	<p>Wie kommt man an unsere Daten?</p>  <p><small>CAMMP day Sicherheit der Privatsphäre in sozialen Netzwerken - Was Facebook, Instagram und Co. über uns wissen 8/25</small></p>

Wie kommt man an unsere Daten?



Wie kommt man an unsere Daten?



Wie kommt man an unsere Daten?



Wie kommt man an unsere Daten?

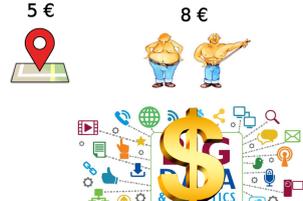
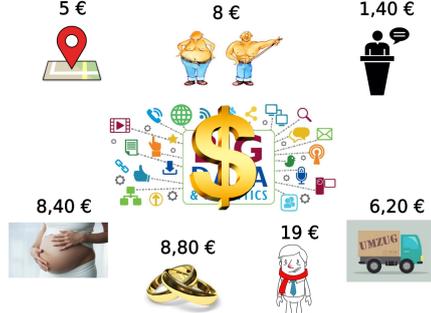


Wie kommt man an unsere Daten?

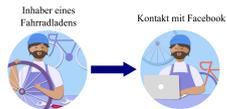


Wie kommt man an unsere Daten?



<p>Wie kommt man an unsere Daten?</p>  <p><small>CAMMP day Sicherheit der Privatsphäre in sozialen Netzwerken - Was Facebook, Instagram und Co. über uns wissen 8/25</small></p>	<p>Was sind unsere Daten wert?</p>  <p><small>CAMMP day Sicherheit der Privatsphäre in sozialen Netzwerken - Was Facebook, Instagram und Co. über uns wissen 8/25</small></p>
<p>Was sind unsere Daten wert?</p>  <p><small>CAMMP day Sicherheit der Privatsphäre in sozialen Netzwerken - Was Facebook, Instagram und Co. über uns wissen 8/25</small></p>	<p>Was sind unsere Daten wert?</p>  <p><small>CAMMP day Sicherheit der Privatsphäre in sozialen Netzwerken - Was Facebook, Instagram und Co. über uns wissen 8/25</small></p>
<p>Was sind unsere Daten wert?</p>  <p><small>CAMMP day Sicherheit der Privatsphäre in sozialen Netzwerken - Was Facebook, Instagram und Co. über uns wissen 8/25</small></p>	<p>Facebook Personalisierte Werbung</p>  <p><small>CAMMP day Sicherheit der Privatsphäre in sozialen Netzwerken - Was Facebook, Instagram und Co. über uns wissen 9/25</small></p>

Facebook | Personalisierte Werbung



Facebook | Personalisierte Werbung



Facebook | Personalisierte Werbung



Facebook | Personalisierte Werbung



Facebook | Personalisierte Werbung

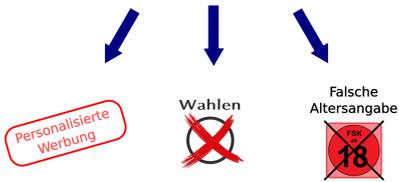


Das heutige Thema:

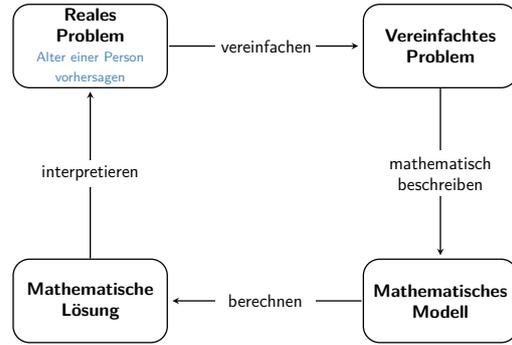
Wie gut kann man das Alter eines Nutzers in einem sozialen Netzwerk vorhersagen, obwohl dieses aus den Angaben in seinem Profil nicht ablesbar ist?

Das heutige Thema:

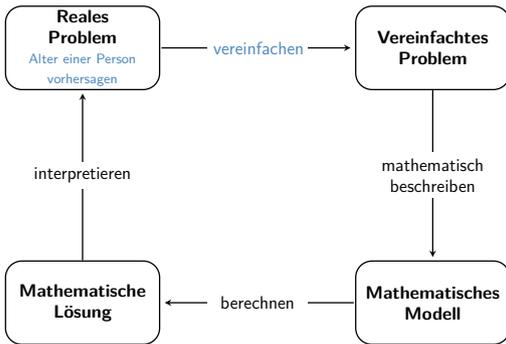
Wie gut kann man das Alter eines Nutzers in einem sozialen Netzwerk vorhersagen, obwohl dieses aus den Angaben in seinem Profil nicht ablesbar ist?



Modellierungskreislauf



Modellierungskreislauf



Vereinfachen | Profile von Nutzern des Netzwerks Friendster

1. Vereinfachung: Daten filtern
2. Vereinfachung: Angaben stimmen



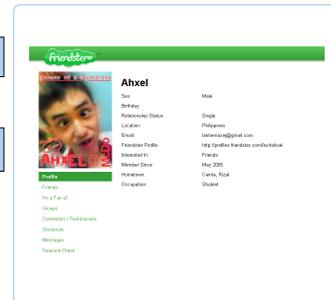
Vereinfachen | Daten aufarbeiten

1. Vereinfachung: Daten filtern
2. Vereinfachung: Angaben stimmen



Vereinfachen | Daten aufarbeiten

1. Vereinfachung: Daten filtern
2. Vereinfachung: Angaben stimmen



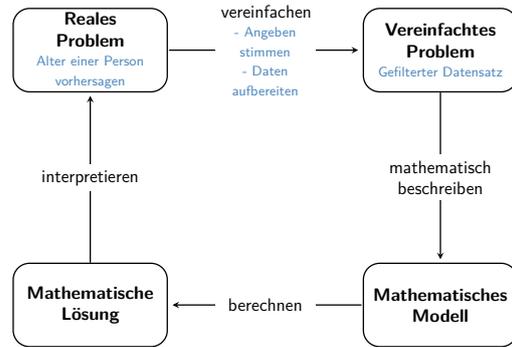
Vereinfachen | Jetzt seid ihr dran

1. Vereinfachung:
Daten filtern

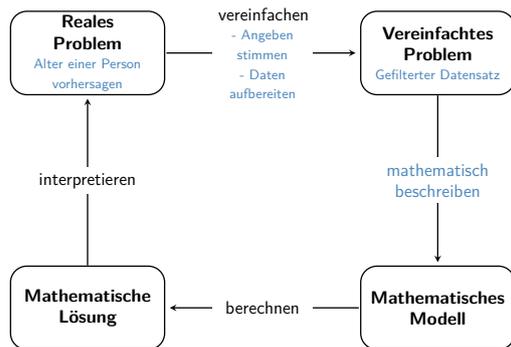
2. Vereinfachung:
Angaben stimmen



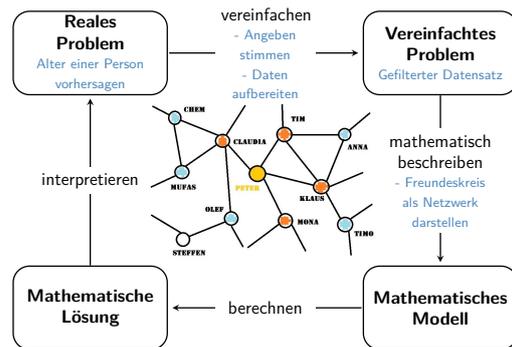
Modellierungskreislauf



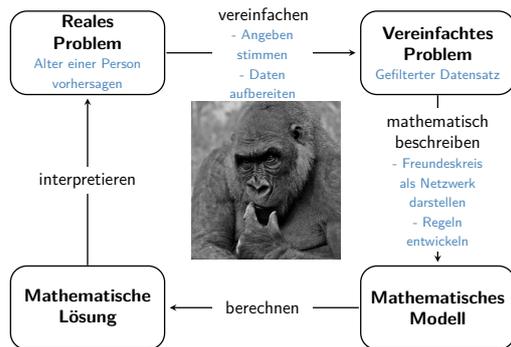
Modellierungskreislauf



Modellierungskreislauf



Modellierungskreislauf



Vorhersageregeln | Freunde

Regel 1

Das Alter des Nutzers ist das gerundete **arithmetische Mittel** des angegebenen Alters der *Freunde*.

Regel 2

Das Alter des Nutzers ist das gerundete **Median** des angegebenen Alters der *Freunde*.

Regel 3

Das Alter des Nutzers ist das gerundete **Modus** des angegebenen Alters der *Freunde*.



Vorhersageregeln | Freundes Freunde

Regel 4

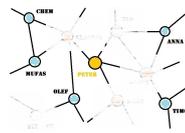
Das Alter des Nutzers ist das gerundete **arithmetische Mittel** des angegebenen Alters der *Freundes Freunde*.

Regel 5

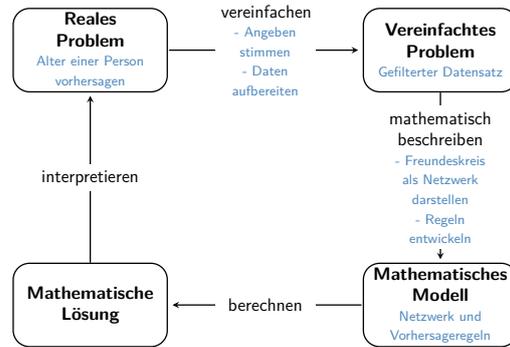
Das Alter des Nutzers ist der gerundete **Median** des angegebenen Alters der *Freundes Freunde*.

Regel 6

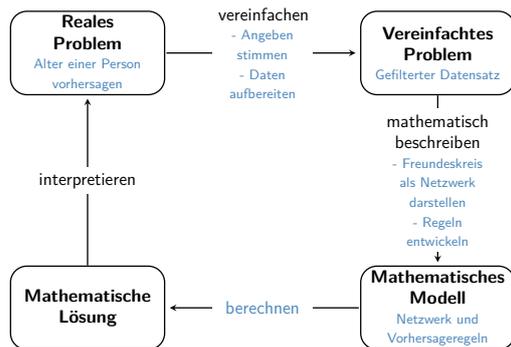
Das Alter des Nutzers ist der gerundete **Modus** des angegebenen Alters der *Freundes Freunde*.



Modellierungskreislauf



Modellierungskreislauf



Berechnen | Regeln anwenden

- 1 Wir kennen das Alter aller Nutzer, die nach der Filterung übrig geblieben sind.
- 2 Wähle eine bestimmte Person aus.
- 3 Blende ihr Alter aus.
- 4 Mache eine Vorhersage mit Hilfe der Regeln.
- 5 Blende das Alter wieder ein und vergleiche die Vorhersage mit dem tatsächlichen Alter.

Berechnen | Ergebnisse vergleichen

Wie können wir die Ergebnisse untereinander vergleichen, wenn jede Gruppe einen unterschiedlich großen Datensatz hat?



Berechnen | Ergebnisse vergleichen

Wie können wir die Ergebnisse untereinander vergleichen, wenn jede Gruppe einen unterschiedlich großen Datensatz hat?



Beispiel:

- Anzahl Nutzer von Gruppe 1: 78000
- Anzahl Nutzer von Gruppe 2: 73000

Berechnen | Ergebnisse vergleichen

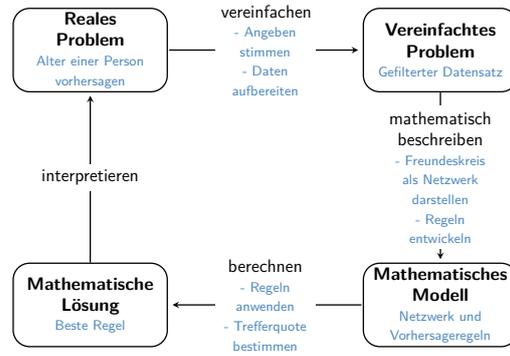
Wie können wir die Ergebnisse untereinander vergleichen, wenn jede Gruppe einen unterschiedlich großen Datensatz hat?



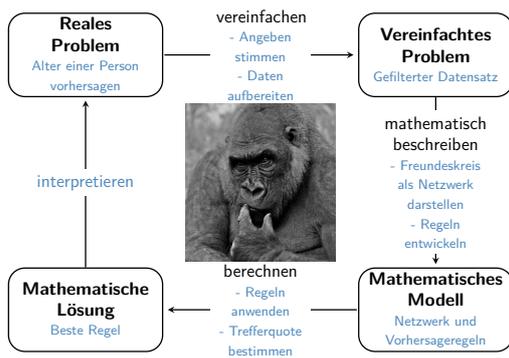
Lösung:

$$\text{Trefferquote} = \frac{\text{Anzahl der Nutzer, die richtig eingeschätzt wurden}}{\text{Gesamtanzahl der Nutzer, die überprüft wurden}}$$

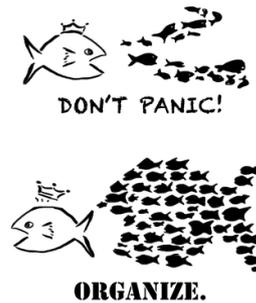
Modellierungskreislauf



Modellierungskreislauf



Jetzt seid ihr dran...



- Bearbeite die Arbeitsblätter!
- Nutze MATLAB!
- Arbeite im Team mit deinem Partner!
- Nutze die Hilfekarten!
- Frag' die Betreuer!
- Nutze das Internet!
- Präsentiere Ergebnisse deinen Mitschülern!
- Hilf deinen Mitschülern!

C.2. Notizen zum Einführungsvortrag

CAMMP day

Soziale Netzwerke Sek I

Hinweise zum Eröffnungsvortrag

Folie 1 | Zentrale Frage

Über dieses Thema wollen wir heute diskutieren.

Folie 2 | Kurzer Überblick

- Die SuS werden gefragt, wer Mitglied eines sozialen Netzwerks ist.
- Es wird erwartet, dass fast alle SuS sich bei der Umfrage melden werden. Sollte nur eine geringe Beteiligung zu erkennen sein, werden Netzwerke wie Instagram, WhatsApp oder Musik.ly aufgezählt.

Folie 3 | Nutzeranzahl bekannter sozialer Netzwerke (Stand 2018)

Den SuS wird mitgeteilt, dass das Ergebnis der Umfrage keine Überraschung ist, da die Nutzerzahlen von sozialen Netzwerken ständig wachsen. Dazu werden die Nutzerzahlen von vier sozialen Netzwerken präsentiert ¹:

- Facebook: 2,2 Milliarden
- Instagram: 800 Millionen
- WhatsApp: 1,3 Milliarden
- Tik Tok: Hier kann man sein eigenes kleines Musikvideo, Tutorial, Prank, etc. aufnehmen. Momentan hat es 500 Millionen Nutzer

Folie 4 | Was denkt ihr?

Die folgenden Fragen werden im Plenum kurz besprochen.

- Was bedeutet für euch Sicherheit in sozialen Netzwerken?
- Welche Informationen gebt ihr in einem sozialen Netzwerk über euch preis? Welche nicht?

Folie 5 | Wie sicher sind eure Daten?

Die Folie wird kurz auf die SuS wirken gelassen. Anschließend wird den SuS gesagt, dass die Themen Datenschutz und Datenmissbrauch seit der Gründung sozialer Netzwerke immer wieder Gesprächsstoff für Diskussionen liefern. Folgende Ausschnitte können vorgelesen werden ²:

- Kann man Facebook noch trauen? Wahlmanipulation, Millionen angezapfter Profile.
- Offenbar hat die Firma Cambridge Analytica mehr Facebook-Nutzer ausgespäht als bisher bekannt.
- Soziale Netzwerke sind nur vordergründig kostenlos. Tatsächlich zahlt der Nutzer mit seinen Daten
- Wie anonym bleibt man im Internet? Nicht sehr, sagen alle Sicherheitsexperten. Vor allem nicht in sozialen Netzwerken wie Facebook oder Twitter.

¹Quelle: <https://de.statista.com/statistik/daten/studie/181086/umfrage/die-weltweit-groessten-social-networks-nach-anzahl-der-user/> (Stand 01.06.2018)

²Quellen: <https://www.zeit.de/digital/internet/2018-04/datenmissbrauch-facebook-zuckerberg-cambridge-analytica> (Stand 21.05.2018)

<http://www.faz.net/aktuell/technik-motor/digital/soziale-netzwerke-zeigt-her-eure-daten-1938762.html> (Stand 21.05.2018)

<https://www.welt.de/wirtschaft/webwelt/article171429928/Diese-Nutzerdaten-sammeln-Facebook-und-Twitter.html> (Stand 21.05.2018)

<https://www.zeit.de/digital/datenschutz/2018-03/datenmissbrauch-cambridge-analytica-facebook-schutz-persoenele-daten> (Stand 21.05.2018)

Folie 6 | Was denkt ihr?

Die folgenden Fragen werden im Plenum kurz besprochen.

- Welches Interesse hat eine Firma oder ein Betreiber eines sozialen Netzwerks an eure Daten?
- Mögliche Antworten: Daten verkaufen, personalisierte Werbung, Informationen über die Person erhalten, etc.
- Welche Informationen wären von besonderen Interesse?
- Mögliche Antworten: Interessen wie Sport, Beruf; Alter, Geschlecht, Wohnort, ect.
- Wie findet ihr die Tatsache, dass eure Daten analysiert werden?
- Antwort offen

Folie 7 | Schattenprofil

Es wird nun der Begriff des Schattenprofils eingeführt:

Definition Schattenprofil:

- für Nutzer des Netzwerks: Profil, das Informationen über euch enthält, die ihr nicht angegeben habt
- (für Nicht-Nutzer: Profil mit Informationen wird erstellt.)

Folie 8 | Was sind unsere Daten wert?

Diese Folie besteht aus 12 Klicks. Die einzelne Punkte stehen für jeden Klick:

- Die Frage ist, wie kommen die Betreiber an unsere Daten?
- Das geht ganz einfach. Z.B. erlauben wir Apps auf unseren Standort zuzugreifen oder haben direkt GPS aktiviert.
- Und schon kann ganz einfach unser Wohnort herausgefunden werden.
- Auch durch Gefällt mir - Angaben geben wir einiges über uns preis. Mit diesen Angaben kann man ganz schnell die Vorlieben und Interessen eines Nutzers herausfinden.
- Z.B. das ein Nutzer sehr an Politik interessiert ist.
- Auch die Nutzung anderer Apps gibt viel über einen Preis. Als ein Beispiel kann hier die Nutzung einer Fitnessstracker-App (oft schon eine Standardapp bei neuen Handys) genannt werden, um seine Kalorien, Puls, gelaufene Strecke, usw. zu erfassen und zu speichern.
- Wie auch bei vielen anderen Apps, darf diese App auf die Fotos und Kontakte eines Nutzers zurückgreifen, die gesammelten Daten an soziale Netzwerke weitergeben und einem Werbung zusenden. Warum darf die App das? Weil man die genannten Punkte durch das Zustimmung der AGBs der App akzeptiert hat. Aber da die ja jeder kennt und liest, erzähle ich wahrscheinlich nichts neues ;)
- Kurzum durch das Nutzen einer Fitnessstracker-App kann man beispielsweise darauf schließen, dass jemand abnehmen möchte oder an Sport interessiert ist.
- Die nächste Frage lautet dann: Was sind unsere Daten und die daraus geschlossenen Informationen überhaupt wert?

2/5

- Dieser Frage ist Quarks und Co³ schon nachgegangen und hat Folgendes herausgefunden (die Preise gelten pro Tausender-Paket): Der Wohnort 5 €, das Interesse an Politik 1,40 € und das man abnehmen möchte 8 €.
- Große Veränderungen im Leben, wie eine Schwangerschaft ist 8,40 € wert oder eine Hochzeit ist 8,80 € wert.
- Die Information, dass man umzieht, ist 6,20 € wert. Und richtig lukrativ wird es, wenn man herausfindet, dass jemand krank ist. Für diese Information wird 19 € gezahlt.

Folie 9 | Facebook - personalisierte Werbung

Es wird ein Beispiele diskutiert, wie Facebook mit den Daten der Nutzer und die daraus resultierende personalisierte Werbung jährlich Milliarden an Gewinnen erzielt⁴:

- Ein Fahrradhändler möchte gerne Werbung über sein Angebot machen.
- Um möglichst viele Leute zu erreichen, wendet er sich an Facebook und beauftragt das Unternehmen gegen Bezahlung die Werbung an den Mann zu bringen.
- Auf der einen Seite kann er nun seine gewünschte Zielgruppe angeben: Z.B. Menschen, die Interesse an Fahrradfahren haben, zwischen 30 und 40 Jahren alt sind und in Sydney wohnen.
- Facebook hat über jeden Nutzer ein eigenes Profil erstellt. Viele Daten geben wir selber an. Interessen können über Gefällt mir-Angaben herausgefunden werden. Fehlende Informationen werden über unsere Freunde ermittelt. So erhalten die gewünschten Nutzer die Werbung des Händlers.
- Auf der anderen Seite kann der Händler Facebook die Telefonnummern und Emailadressen seiner bisherigen Kunden schicken. Facebook gleicht diese mit seiner Datenbank ab und schickt die Werbung an diese Kunden.

Folie 10 | Das heutige Thema

Das Thema des Tages wird nun vorgestellt:

Wie gut kann man das Alter eines Nutzers in einem sozialen Netzwerk vorhersagen, obwohl dieses aus den Angaben in seinem Profil nicht ablesbar ist?

Die SuS sollen zunächst eine Schätzung abgeben und anschließend diskutieren, wieso Betreiber eines Netzwerks an dieser Information interessiert sind.

Anschließend werden folgende drei Interessen vorgestellt:

- Die angesprochene personalisierte Werbung
- Wahlen: Wann ist man wahlberechtigt? Wahlthemen für bestimmte Altersgruppen
- Falsche Altersangaben herausfinden (z.B. im Playstation Netzwerk)

Folie 11-12 | Vereinfachen

Es folgt der Modellierungskreislauf. Das reale Problem, das Alter eines Nutzer zu bestimmen, muss zunächst vereinfacht werden.

³Quelle: Film *Quarks und Co: Verdatet und verkauft - wem gehören unsere Daten?*, Minuten 19:45 - 20:30

⁴Quelle: <https://de-de.facebook.com/ads/about> (Stand 18.05.2018)

Folie 13 | Vereinfachen

- echtes Netzwerk: Friendster
- diente als Vorbild für z.B. facebook.
- gegründet 2002
- bis zu 350 Millionen Nutzer (im Vergleich: Facebook 2,2 Milliarden Nutzer)
- die meisten Nutzer stammten aus den USA und aus Asien
- Netzwerk nicht mehr aktiv
- Daten der Nutzer sind gespeichert worden, sodass wir mit ihnen arbeiten können

1. Vereinfachung: Daten filtern

2. Vereinfachung: Angaben der Nutzer stimmen nach der Filterung

Hier wird die Präsentation pausiert und das erste Arbeitsblatt von den SuS bearbeitet. Nach der Besprechung geht die Präsentation hier weiter.

Folie 14-16 | Modellierungskreislauf

- Das vereinfachte Problem muss nun mathematisch beschrieben werden.
- Dazu stellen wir den Freundeskreis einer Person als ein Netzwerk dar, in dem die Freunde (orange) und Freundes-Freunde (blau) betrachtet werden.
- Anschließend müssen wir uns Vorhersageregeln überlegen, mit denen wir das Alter einer Person vorhersagen wollen und diese Regeln auf ihre Richtigkeit überprüfen.
- **Hier soll mit den SuS zunächst über mögliche Vorhersageregeln diskutiert werden, bevor die zu untersuchenden Regeln vorgestellt werden.**

Folie 17-18 | Vorhersageregeln

- Zunächst werden Regeln betrachtet, welche das genaue Alter eines Nutzers vorhersagen sollen.
- Dazu wird das Alter der Freunde und Freundes-Freunde betrachtet
- Von dem Alter der Freunde und Freundes-Freunde wird entweder das arithmetische Mittel, der Median oder der Modus berechnet. Der resultierende, gerundete Wert ist dann das vorhergesagte Alter des Nutzers (kurz mit den SuS die Grafik besprechen und für die gezeigten Freunde das arithmetische Mittel bestimmen lassen).

Folie 19-20 | Modellierungskreislauf

- Das Netzwerk und die Vorhersageregeln bilden unser mathematisches Modell.
- Als nächstes folgt die Berechnung, also das Anwenden unserer Regeln.

4/5

Folie 21 | Berechnen | Regeln anwenden

Für das Überprüfen der aufgestellten Regeln geht man wie folgt vor:

1. Das Alter aller Nutzer ist nach der Filterung bekannt.
2. Man wählt eine Person aus.
3. Man blendet das bekannte Alter aus.
4. Man trifft mit seiner aufgestellten Regel eine Vorhersage.
5. Man blendet das Alter wieder ein und überprüft, ob die Vorhersage richtig war.

Folie 22 | Berechnen | Ergebnisse vergleichen

Mit den SuS soll nun die Einführung der Trefferquote diskutiert werden:

- Durch die Filterung hat ggf. jede Gruppe einen Datensatz mit unterschiedlicher Stichprobengröße erzeugt.
- Mit den SuS wird die Frage diskutiert, wie man die Ergebnisse später trotzdem untereinander vergleichen kann (Die SuS erhalten kurz Zeit, um die Frage mit ihrem Partner zu besprechen und ihre Ideen zu notieren).
- Die Ideen der SuS werden besprochen. Zum Schluss wird die Formel für die Trefferquote vorgestellt.

Folie 23-24 | Modellierungskreislauf

Als mathematische Lösung erhält man dann die beste Regel, also die Regel mit der besten Trefferquote. Diese Lösung muss dann anschließend bezogen auf das reale Problem interpretiert werden.

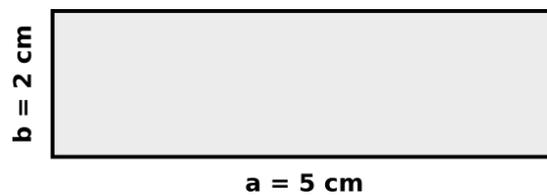
D. Einführung in MATLAB

Kurze Einführung in Matlab

Aufgabe 1

Ein Rechteck hat folgende Seitenlängen:

- Länge: $a = 5$ cm
- Breite: $b = 2$ cm



Bestimmt den Flächeninhalt des Rechtecks und überprüft eure Lösung mit Matlab.



```
a = 5;           % Länge des Rechtecks
b = 2;           % Breite des Rechtecks

Loesung = NaN;   % Ergebnis für den Flächeninhalt eintragen

Ergebnis(Loesung)
```

E. Arbeitsblatt 1

CAMMP day (Soziale Netzwerke) - Daten Filtern

Bevor wir uns über Regeln, die das Alter einer Person vorhersagen, Gedanken machen können, müssen wir uns zunächst einen Überblick über die vorhandenen Daten verschaffen. Wir müssen den vorliegenden Datensatz analysieren und uns entscheiden, welche Daten für unsere Zwecke genutzt werden können.



Aufgabe 1 - Daten laden



Zunächst muss der Datensatz, bestehend aus den ersten ca. 60.000 Nutzer des Netzwerks Friendster, in das Programm Matlab geladen werden. Klickt dazu auf den Button *Run Section* .



Daten_laden

Aufgabe 2 - Altersverteilung

Als nächstes müssen wir uns einen Überblick darüber verschaffen, wie die Altersverteilung in unserem Datensatz aussieht. Lasst euch dazu von Matlab ein Säulendiagramm anzeigen, welches euch die absoluten Häufigkeiten der einzelnen Altersangaben anzeigt. Drückt dazu wieder auf *Run Section* .



Tip: Geht ihr mit dem Mauszeiger über die Graphik, könnt ihr in der oberen rechten Ecke

das "Zoom-in" Symbol  auswählen. Klickt ihr dann mit der Maus in die Graphik, zoomt ihr in den ausgewählten Bereich hinein.



Beschreibt auf eurem Antwortblatt kurz die dargestellte Altersverteilung, in dem ihr euch an folgenden Leitfragen orientiert:

- Was beschreiben die beiden Achsen des Diagramms?
- In welchem Altersbereich liegt das Alter der meisten Nutzer?
- Wie entwickelt sich die absolute Häufigkeit der Nutzer mit wachsendem Alter?
- Gibt es auffällige Altersangaben?

Begründet, welche Nutzer eine falsche Altersangabe getätigt haben könnten.



Altersverteilung



Aufgabe 3 - Erste Filterung

Es wird Zeit eine erste Aufbereitung der Daten vorzunehmen und die Nutzer mit einer vermeintlich falschen Altersangabe aus den Datensatz herauszunehmen.



Wählt ausgehend von euren obigen Überlegungen für das Alter eine untere Grenze a und eine obere Grenze b , sodass die Nutzer mit einer Altersangabe unter- bzw. oberhalb dieser Grenze aus den Daten gefiltert werden. Notiert eure gewählten Grenzen auf dem Antwortblatt und klickt auf *Run Section* , um die Filterung zu starten.



```
a = NaN; % Alle Nutzer mit einer Altersangabe < a werden aus dem Datensatz gefiltert  
b = NaN; % Alle Nutzer mit einer Altersangabe > b werden aus dem Datensatz gefiltert
```

```
Filterung1(a,b)
```

Aufgabe 4 - Anzahl der Freunde und Freundes-Freunde

Um Informationen über euch herauszufinden, greifen Soziale Netzwerke unter anderem auf die Angaben eurer Freunde zurück. Mit einem Klick auf *Run Section*  erhaltet ihr einen Überblick über die Anzahl der Freunde und die Anzahl der Freundes-Freunde.



Tip: Geht ihr mit dem Mauszeiger über die Graphik, könnt ihr in der oberen rechten Ecke das "Zoom-in" Symbol  auswählen. Klickt ihr dann mit der Maus in die Graphik, zoomt ihr in den ausgewählten Bereich hinein.



Beschreibt auf eurem Antwortblatt die beiden dargestellten Grafiken. Begründet, welche Nutzer das Vorhersagen einer Information erschweren.



Daten
Angabe_Freunde(Users)



Aufgabe 5 - Zweite Filterung | Teil 1

Nun müssen wir eine zweite Aufbereitung der Daten vornehmen. Wir möchten später die Vorhersagen über das Alter eines Nutzers anhand der Angaben der Freunde oder Freundes-Freunde treffen. Wir wollen daher diejenigen Nutzer aus dem Datensatz aussortieren, welche 0 Freunde und 0 Freundes-Freunde besitzen.



Bringt dazu die folgenden Schritte zum Aussortieren der Nutzer in die richtige Reihenfolge und drückt auf *Run Section* .

1. Die Anzahl N der Nutzer im Datensatz bestimmen.
2. Für Nutzer mit 0 Freunden und 0 Freundes-Freunden in der bereits bestehenden Liste die 1 durch eine 0 ersetzen.
3. Den neuen Datensatz speichern.
4. Ab Schritt 2 den Code erneut durchlaufen lassen, um eventuell neu entstandene Nutzer mit 0 Freundes-Freunden auszusortieren.
5. Eine Liste mit N (Anzahl der Nutzer) Einsen erstellen.
6. Den Datensatz in das Programm laden.
7. Nutzer, an deren Position in der Liste eine 0 auftaucht, aus dem Datensatz entfernen. Gibt es in der Liste keine 0 als Eintrag, gehe zu Schritt 7.



```
Schritt_1 = NaN;           % Für NaN eine Nummer aus der Auflistung eintragen
Schritt_2 = NaN;           % Für NaN eine Nummer aus der Auflistung eintragen
Schritt_3 = NaN;           % Für NaN eine Nummer aus der Auflistung eintragen
Schritt_4 = NaN;           % Für NaN eine Nummer aus der Auflistung eintragen
Schritt_5 = NaN;           % Für NaN eine Nummer aus der Auflistung eintragen
Schritt_6 = NaN;           % Für NaN eine Nummer aus der Auflistung eintragen
Schritt_7 = NaN;           % Für NaN eine Nummer aus der Auflistung eintragen
```

```
Filterung2(Schritt_1,Schritt_2,Schritt_3,Schritt_4,Schritt_5,Schritt_6,Schritt_7)
```

Aufgabe 5 - Zweite Filterung | Teil 2



Führt die zweite Filterung für sechs Nutzer per Hand aus. Geht dabei wie folgt vor:

1. Drückt auf *Run Section* , um euch die Informationen über die sechs Nutzer anzeigen zu lassen.
2. Füllt die drei Listen entsprechend der Anleitung aus Teil 1 aus, indem ihr die Einträge für die einzelnen Listen zwischen den eckigen Klammern [] eingibt.
3. Überprüf euer Ergebnis mit *Run Section* .



Wichtig: Zwischen den einzelnen Einträgen in den Listen muss immer mindestens ein Leerzeichen gesetzt werden!

- In Liste_1 wird für jedem Nutzer zunächst eine 1 eingetragen.
- In Liste_2 wird für die Nutzer, die 0 Freunde oder Freundes-Freunde besitzen, die 1 durch eine 0 ersetzt.
- Nutzer ist eine Liste mit den Nutzerpositionen der verbleibenden Nutzer.



```
Liste_1 = [NaN NaN NaN NaN NaN NaN]; % Liste mit Einsen als Einträgen.  
% Denkt an die Leerzeichen!  
Liste_2 = [NaN NaN NaN NaN NaN NaN]; % Liste mit Nullen und Einsen als  
% Einträge. Denkt an die Leerzeichen!  
Nutzer = [NaN NaN NaN NaN]; % Liste mit den verbliebenden  
% Nutzerpositionen. Denkt an die Leerzeichen!  
  
Programmieren(Liste_1,Liste_2,Nutzer)
```

Aufgabe 6 - Übersicht zum neuen Datensatz verschaffen

Durch die beiden Filterungen habt ihr nun einen neuen Datensatz erzeugt. Drückt auf *Run Section* , um eine Übersicht über diesen Datensatz zu erhalten. Notiert diese Informationen auch auf eurem Antwortblatt.



```
Uebersicht_Datensatz
```



Zusatzaufgabe

Überlegt euch mögliche Vorhersageregeln, mit denen das Alter eines Nutzers vorhergesagt werden könnte. Notiert diese Regeln als "Wenn..., dann..." - Aussagen auf eurem Antwortblatt.



E.1. Zusatzmaterial von Arbeitsblatt 1

E.1.1. Antwortblatt

CAMMP day
Soziale Netzwerke Sek. I



RWTHAACHEN
UNIVERSITY

Antwortblatt 1: Daten filtern

Aufgabe 2 – Altersverteilung

- Beschreibt ausgehend von den Leitfragen die dargestellte Altersverteilung.

- Begründet, welche Nutzer eine falsche Altersangabe getätigt haben könnten.

Aufgabe 3 – Erste Filterung

Notiert eure gewählten Altersgrenzen a und b :

$a =$ _____ $b =$ _____

Aufgabe 4 – Anzahl der Freunde und Freundes-Freunde

- Beschreibt die beiden dargestellten Säulendiagramme.

- Begründet, welche Nutzer das Vorhersagen einer Information erschweren.

Aufgabe 6 – Übersicht zum neuen Datensatz

Gesamtanzahl N der Nutzer (Stichprobengröße):	
Minimales Alter im Datensatz:	
Maximales Alter im Datensatz:	
Durchschnittliche Anzahl an Freunde:	
Durchschnittliche Anzahl an Freundes-Freunden:	

Zusatzaufgabe

Notiert mögliche Regeln, mit denen das Alter eines Nutzers vorhergesagt werden könnte, als „Wenn..., dann...“ - Aussagen.

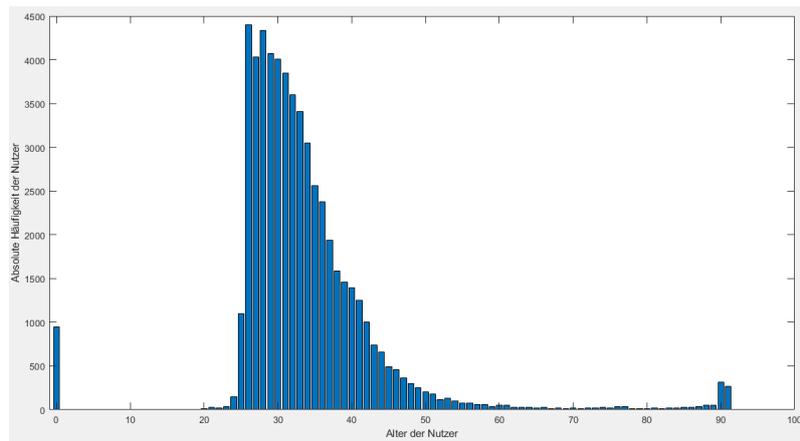
E.1.2. Besprechungsfolie

CAMMP day
Soziale Netzwerke Sek. I



Arbeitsblatt 1 | Daten filtern: Besprechung

Aufgabe 2 & 3 – Altersverteilung



- Begründet, welche Nutzer eine falsche Altersangabe getätigt haben könnten.

Beispiel:	1	2	3
Untere Schranke a:			
Obere Schranke b:			

Aufgabe 4 – Anzahl der Freunde und Freundes-Freunde

- Begründet, welche Nutzer das Vorhersagen einer Information erschweren.

Aufgabe 5 – Zweite Filterung | Teil II

Name	Nutzerposition	Alter der Freunde	Alter der Freundes-Freunde
Lisa	1	6	8
Tobias	2	3	0
Simon	3	8	16
Marius	4	6	5
Carolin	5	2	0
Miriam	6	7	12

Liste_1 = [___ ___ ___ ___ ___ ___]

Liste_2 = [___ ___ ___ ___ ___ ___]

Nutzer = [___ ___ ___ ___]

Aufgabe 6 – Übersicht zum neuen Datensatz

Beispiel:	1	2	3
Gesamtanzahl N der Nutzer (Stichprobengröße):			
Minimales Alter im Datensatz:			
Maximales Alter im Datensatz:			
Durchschnittliche Anzahl an Freunde:			
Durchschnittliche Anzahl an Freundes-Freunden:			

E.2. Musterlösung für Arbeitsblatt 1

Arbeitsblatt 1 | Daten filtern: Lösung

Aufgabe 2 & 3 – Altersverteilung

Begründet, welche Nutzer eine falsche Altersangabe getätigt haben könnten.

Hinweis zur Besprechung: Hier sind verschiedene Lösungen möglich. Wichtig ist, dass mit den SuS über die Wahl der Schranken a und b begründet diskutiert wird. Thematisiert und diskutiert werden sollten der Wahrheitsgehalt von Altersangaben von Null Jahren und von Angaben um die 80 und 90 Jahre. So kann beispielsweise diskutiert werden, dass Friendster eines der ersten sozialen Netzwerke war und gerade die älteren Personen an der damals gerade aufkommende Digitalisierung größtenteils nicht teilgenommen haben. So ist es zu dieser Zeit äußerst unüblich Personen um die 80 oder 90 Jahre in einem sozialen Netzwerk anzutreffen. Besonders auffällig ist zudem der starke Anstieg der Nutzer von 89 auf 90 sowie 91 Jahre. So gibt es fast sechs Mal so viele Nutzer mit einem Alter von 90 Jahre als 89-Jährige. Auch die Tatsache, dass die Anzahl der Personen in der Datenbank ab 28 Jahre mit zunehmendem Alter zurückgeht und plötzlich bei 90 Jahren wieder stark ansteigt, wirkt seltsam.

In der Musterlösung wurden Personen ab einem Alter von 90 Jahre aus dem Datensatz herausgenommen.

Aufgabe 4 – Anzahl der Freunde und Freundes-Freunde

Begründet, welche Nutzer das Vorhersagen einer Information erschweren.

Hinweis zur Besprechung: Wie bereits in der Aufgabe erwähnt, beziehen Betreiber von sozialen Netzwerken viele Informationen über uns mit Hilfe der Angaben unserer Freunde. Nutzer mit nur wenigen Freunden erschweren dabei eine Vorhersage. Vor allem aber sind die Nutzer mit Null Freunden oder Null Freundes-Freunde für eine Vorhersage des Alters ungeeignet.

Aufgabe 5 – Zweite Filterung | Teil I

Die richtige Reihenfolge lautet:

Schritt_1 = 6	Schritt_5 = 7
Schritt_2 = 1	Schritt_6 = 8
Schritt_3 = 5	Schritt_7 = 4
Schritt_4 = 2	Schritt_8 = 3

Aufgabe 5 – Zweite Filterung | Teil II

Da Tobias (Nutzer 2) und Carolin (Nutzer 5) keine Freundes-Freunde besitzen, müssen diese Nutzer aus dem Datensatz gefiltert werden. Aus diesem Grund tauchen die Nutzer - IDs nach der Filterung nicht mehr in der Liste der Nutzer auf.

Liste_1 = [1 1 1 1 1 1]

Liste_2 = [1 0 1 1 0 1]

Nutzer = [1 3 4 6]

Den SuS kann hier noch erklärt werden, dass es in der Informatik eine gängige Methode ist mit Listen aus Einsen und Nullen zu arbeiten. Wahre Aussagen werden dort ebenfalls mit einer Eins und falsche Aussagen mit einer Null codiert.

Auch kann den SuS gesagt werden, dass nach der Filterung die Nutzer - IDs aktualisiert werden. Das heißt, die Nutzer bekommen eine neue ID, sodass die IDs wieder durchgängig durchlaufen.

Aufgabe 6 – Übersicht zum neuen Datensatz

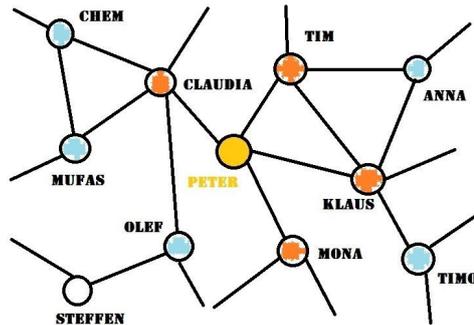
Beispielhafte Ausgabe bei einer Stichprobengröße von **N=44747**. Dieser Datensatz wurde auch bei der Erstellung der Musterlösung vom 2. und 3. Aufgabenblatt verwendet.

Gesamtanzahl N der Nutzer (Stichprobengröße):	44747
Minimales Alter im Datensatz:	18
Maximales Alter im Datensatz:	89
Durchschnittliche Anzahl an Freunde:	3.68
Durchschnittliche Anzahl an Freundes-Freunde:	23.06

F. Arbeitsblatt 2

CAMMP day (Soziale Netzwerke) - Vorhersageregeln anwenden

Das Alter der Personen soll nach bestimmten Regeln vorhergesagt werden. Ihr erhaltet dazu eine Liste mit Regeln, die wir untersuchen wollen. Im Folgenden sollt ihr überprüfen, wie gut die einzelnen Regeln das Alter einer Person vorhersagen.



Aufgabe 1 - Auffrischung des Wissens

Um in das Thema hineinzukommen, wollen wir kurz die Begriffe arithmetisches Mittel, Median und Modus wiederholen:

arithmetisches Mittel:	$\frac{\text{Summe aller Werte}}{\text{Anzahl der Werte}}$
Median:	Der Wert, welcher genau in der Mitte des Datensatzes liegt.
Modus:	Der Wert, welcher am häufigsten vorkommt.



Berechnet für einen Nutzer mit mehr als 5 Freunden für das Alter seiner Freunde jeweils das arithmetische Mittel, den Median und den Modus.

- Tragt dazu zunächst unten im Code die Nummer eines Nutzers für NaN ein und drückt auf *Run Section* .
- Berechnet anschließend die gesuchten Größen. Wenn beim Modus mehrere Werte zur Auswahl stehen, nehmt das arithmetische Mittel dieser Werte.
- Überprüft eure Ergebnisse, indem ihr sie für NaN eintragt und auf *Run Section*  drückt.



Hinweis: Zu dieser Aufgabe gibt es eine **Hilfekarte**. Öffnet dazu die Datei *Hilfekarte_1_Wiederholung.mlx*



```
Nutzer           = NaN;      % Nutzer, dessen Freunde angezeigt werden sollen
arithmetisches_Mittel = NaN;  % Euer berechnetes arith. Mittel für den Nutzer
Median           = NaN;      % Euer berechneter Median für den Nutzer
Modus            = NaN;      % Euer berechneter Modus für den Nutzer
```

```
Daten
Check_Loesung1(Users, Nutzer, arithmetisches_Mittel, Median, Modus)
```

Aufgabe 2 - Regeln durchlesen

1. Tragt zunächst die Gesamtzahl N an Nutzern in eurem Datensatz auf dem neuen Antwortblatt ein.
2. Lest euch die konkreten Regeln genau durch. Vermutet zunächst, welche der genannten Regeln die beste Vorhersage über das Alter treffen wird. Begründet eure Vermutung auf dem Antwortblatt.
3. Entscheidet euch, ob ihr im weiteren Verlauf die Regeln für die Freunde oder Freundes-Freunde der Nutzer testen wollt. Tragt eure Entscheidung auf dem Antwortblatt ein.



Aufgabe 3 - Regeln auf einzelne Nutzer anwenden



Zunächst wollen wir die Regeln an einzelnen Nutzern testen:

- Wählt dazu eine Regel aus, die ihr überprüfen wollt.
- Wählt einen Nutzer aus, auf dem die Regel angewendet werden soll.
- Drückt auf *Run Section* .
- Tragt euer Ergebnis auf dem Antwortblatt ein.



```
Regel = NaN;      % Regel, die überprüft werden soll  
Nutzer = NaN;    % Nummer des Nutzers, auf dem die Regel angewendet wird
```

```
Daten  
Check_Nutzer(Users, Regel, Nutzer)
```

Aufgabe 4 - Regeln auf viele Nutzer anwenden

Wir wollen die Regeln nun für viele Nutzer im Datensatz testen, um eine Aussage über die Trefferquote machen zu können.

$$\text{Trefferquote} = \frac{\text{Anzahl der Nutzer, die richtig eingeschätzt wurden}}{\text{Gesamtanzahl der Nutzer, die überprüft wurden}}$$



a) Tragt die Nummer der Regel, die ihr testen wollt, für NaN unter Regel ein. Drückt auf *Run*

Section  und die ausgewählte Regel wird auf alle Nutzer in eurem Datensatz angewendet und auf ihre Richtigkeit überprüft. Notiert eure Ergebnisse auf dem Antwortblatt und berechnet die Trefferquote.

b) Begründet auf eurem Antwortblatt, warum es sinnvoll ist die Vorhersageregeln auf alle Nutzer im Datensatz anzuwenden.



```
Regel = NaN;      % Regel, die überprüft werden soll
```

```
Daten  
Check_Regeln(Users, Regel)
```

Aufgabe 5 - Der Median und das arithmetische Mittel

Wir wollen uns kurz anschauen, warum der Median etwas bessere Ergebnisse als das arithmetische Mittel liefert. Dazu betrachten wir das Alter der Freunde von **Kim** im Jahr **2003** und **2004**. Im Jahr

2004 haben sich auch die Eltern von Kim in Friendster registriert und wurden von Kim als Freunde hinzugefügt. Das Alter von Kim beträgt 22.



a)

1. Gebt unter Nutzer für NaN zunächst Kim_2003 ein und drückt auf *Run Section* . Die '' müssen dabei stehen bleiben!
2. Berechnet das arithmetische Mittel und den Median für das Alter von Kims Freunden im Jahr 2003.
3. Überprüft eure Ergebnisse, indem ihr eure berechneten Werte für NaN einsetzt und auf *Run Section*  drückt. Tragt eurer Ergebnisse anschließend auf dem Antwortblatt ein.
4. Führt die gleichen Schritte für Kim_2004 aus.

b) Beschreibt auf eurem Antwortblatt bezogen auf Kims Alter den Unterschied zwischen dem arithmetischen Mittel und dem Median. Geht vor allem darauf ein, wie die Altersangaben von Kims Eltern im Jahr 2004 die beiden Werte beeinflussen.



```
Nutzer           = 'NaN';      % Nutzer, dessen Freunde angezeigt werden sollen.
                                     % Man braucht ''
arithmetisches_Mittel = NaN;    % Euer berechnetes arith. Mittel für den Nutzer
Median           = NaN;    % Euer berechneter Median für den Nutzer

Check_Loesung2(Nutzer, Median, arithmetisches_Mittel)
```

Neue Regeln

Wir haben jetzt schon ein paar Regeln auf den Datensatz angewendet, jedoch sind die bisherigen Trefferquoten noch nicht wirklich zufriedenstellend. Daher werden wir im weiteren Verlauf Regeln betrachten, die nicht das exakte Alter sondern den Bereich, in dem das Alter des Nutzers liegt, vorhersagen.

Aufgabe 6 a) - Boxplot zeichnen

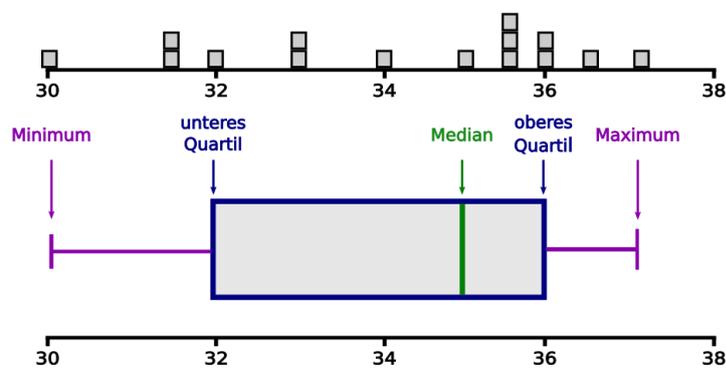
Mit Hilfe eines Boxplots wollen wir zunächst die Daten der Nutzer genauer analysieren, um neue Regeln für die Bestimmung des Alters herausfinden. Drückt auf *Run Section*  und ihr erhaltet die Altersangaben der Freunde von Julia.



Zeichnet für diese Daten einen Boxplot.



Hinweis: Für diese Aufgabe gibt es eine **Hilfekarten**. Öffnet dazu die Datei *Hilfekarte_2_Boxplot.mlx*



Boxplot

Aufgabe 6 b)

Die Box im Boxplot wird durch das untere und das obere Quartil begrenzt. Es ist festgelegt, dass höchstens 25 % der Werte unterhalb dem unteren Quartil und höchstens 25 % der Werte oberhalb des oberen Quartil liegen.



Bestimmt, wie viel Prozent der Werte mindestens innerhalb der Box liegen. Gebt den Wert als Bruch oder als Dezimalzahl an und tragt ihn für NaN bei Loesung ein. Überprüft eure Eingabe mit *Run*

Section .



```
Loesung = NaN;           % Prozentualer Anteil der Werte, die mindestens  
                          % innerhalb der Box liegen.
```

```
Check_LoesungBoxplot(Loesung)
```

Aufgabe 6 c) - Spannweite und Quartilsabstand



Bestimmt für die Daten von Julia die Spannweite und den Quartilsabstand. Überprüft eure Ergebnisse mit Matlab.

Spannweite = Maximum – Minimum

Quartilsabstand = oberes Quartil – unteres Quartil



```
Spannweite      = NaN;           % Tragt für NaN eure berechnete Spannweite ein.  
Quartilsabstand = NaN;           % Tragt für NaN euren berechneten Quartilsabstand ein.
```

```
Check_Loesung3(Spannweite, Quartilsabstand)
```

d) - Interpretation

Interpretiert auf eurem Antwortblatt die Bedeutung der Spannweite und des Quartilsabstands. Beschreibt dazu, wie viel Prozent der Daten innerhalb der Spannweite und des Quartilsabstand liegen. Begründet auch, welcher der beiden Werte sich am stärksten beim Auftreten eines Ausreißers verändern würde.



Aufgabe 7 a) - Neue Vorhersageregeln durchlesen

Meldet euch bei den Betreuern, um eine neue Liste mit Vorhersageregeln zu erhalten.



Lest euch die Regeln durch und vermutet zunächst, welche der Regeln die beste Vorhersage über das Alter macht. Tragt eure Vermutung auf das Antwortblatt ein und begründet diese.



Aufgabe 7 b) - Neue Regeln testen



Wir wollen nun die Regeln auf den Datensatz anwenden, um die Trefferquote der Regeln zu bestimmen.

1. Wählt zunächst wieder aus, ob ihr die Regeln für die Freunde oder Freundes-Freunde der Nutzer betrachten wollt.
2. Wendet für eure Wahl die Regel für die Spannweite und für den Quartilsabstand an.
3. Drückt auf *Run Section*  und tragt die Ergebnisse auf eurem Antwortblatt ein.
4. Bestimmt die Trefferquote für die beiden Regeln.



```
Regel = NaN;           % Regel, die überprüft werden soll  
  
Daten  
Check_Regeln2(Users, Regel)
```

Aufgabe 7 c) - Überblick über die Altersspanne verschaffen

Um eine Aussage darüber treffen zu können, wie groß die vorhergesagten Altersspannen sind, in denen das Alter des Nutzers liegen soll, müsst ihr euch die durchschnittliche Spannweite und den durchschnittlichen Quartilsabstand für die Freunde bzw. Freundes-Freunde anschauen.

- Drückt auf *Run Section*  und tragt die Informationen auf eurem Antwortblatt ein.
- Begründet auf eurem Antwortblatt, ob ihr die Spannweite oder den Quartilsabstand für eine Vorhersage des Alters wählen würdet.



Daten
Uebersicht_Streumass(Users)



Zusatzaufgabe - Bonusregel

Wir wollen noch eine weitere Vorhersageregeln betrachten, die sich mit der Standardabweichung beschäftigt.

Die Standardabweichung s gibt an, wie stark die einzelnen Werte um das arithmetische Mittel / den Mittelwert \bar{x} schwanken.

- Die Standardabweichung kann nur dann 0 sein, wenn alle Werte gleich sind.
- Die Standardabweichung ist klein, wenn die Werte nah am Mittelwert liegen.
- Je weiter die einzelnen Werte vom Mittelwert entfernt sind, desto größer wird die Standardabweichung.

Berechnung der Standardabweichung:

1. Zunächst bestimmt man die Anzahl n der Werte x_1, x_2, \dots, x_n .
2. Danach bestimmt man den Mittelwert \bar{x} der Werte x_1, x_2, \dots, x_n .
3. Anschließend erhält man die Standardabweichung mit folgender Formel:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}} \quad \text{Mit den Werten } x_1, x_2, \dots, x_n.$$

Das heißt, im Nenner steht die Größe der Stichprobe n . Im Zähler der Standardabweichung werden die quadratischen Abstände der einzelnen Werte x_1, x_2, \dots, x_n zum Mittelwert \bar{x} addiert. Damit steht unterhalb der Wurzel der mittlere Quadratische Abstand der einzelnen Werte zum Mittelwert.

Teil 1 - Standardabweichung kennen lernen



Berechnet die Standardabweichung für die Altersangaben von Pauls Freunden auf 2 Nachkommastellen genau:

[12 15 13 12 13 16 13 15]

Überprüft euer Ergebnis mit Matlab.



```
s = NaN;           % Berechnete Standardabweichung s
Check_Standardabweichung(s)
```

Teil 2 - Neue Regel überprüfen

Folgende Regeln zur Vorhersage des Alters sollen nun betrachtet werden:

Bonusregeln (Standardabweichung):

Für die Freunde:

Es wird für den Nutzer das arithmetische Mittel \bar{x} und die Standardabweichung s des angegebenen Alters der Freunde bestimmt. Das Alter des Nutzers liegt dann innerhalb der Werte $\bar{x} - s$ und $\bar{x} + s$.

Für die Freundes-Freunde:

Es wird für den Nutzer das arithmetische Mittel \bar{x} und die Standardabweichung s des angegebenen Alters der Freundes-Freunde bestimmt. Das Alter des Nutzers liegt dann innerhalb der Werte $\bar{x} - s$ und $\bar{x} + s$.



Wendet die neuen Regeln auf euren Datensatz an und bestimmt die Trefferquote. Geht dabei wie folgt vor:

- Legt zunächst unter Freundeskreis fest, ob das Alter der Freunde (1) oder der Freundes-Freunde (2) betrachtet wird.
- Klickt auf *Run Section* .



```
Freundeskreis = 1;           % Wählt 1 für die Freunde und 2 für die Freundes-Freunde
Daten
Test_Standardabweichung(Users, Freundeskreis)
```

Teil 3 - Vorhergesagte Altersspanne abschätzen

Drück auf *Run Section* , um eine Übersicht über die durchschnittliche Standardabweichung für die Freunde und Freundes-Freunde zu erhalten. So bekommt ihr einen Überblick, wie groß die vorhergesagten Altersspannen, in denen das Alter des Nutzers liegen soll, im Durchschnitt sind.



Uebersicht_Standardabweichung

F.1. Zusatzmaterial von Arbeitsblatt 2

F.1.1. Übersicht über die Vorhersageregeln 1-6

Vorhersageregeln

Für die Freunde:

Regel 1: (Arithmetisches Mittel der Freunde)

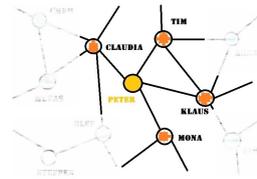
Das Alter des Nutzers ist das auf eine ganze Zahl gerundete arithmetische Mittel des angegebenen Alters der Freunde.

Regel 2: (Median der Freunde)

Das Alter des Nutzers ist der auf eine ganze Zahl gerundete Median des angegebenen Alters der Freunde.

Regel 3: (Modus der Freunde)

Das Alter des Nutzers entspricht dem Modus des angegebenen Alters der Freunde. Bei mehreren Modi ist das Alter des Nutzers das auf eine ganze Zahl gerundete arithmetische Mittel dieser Werte.



Peter und die Freunde von Peter.

Für die Freundes-Freunde:

Regel 4: (Arithmetisches Mittel der Freundes-Freunde)

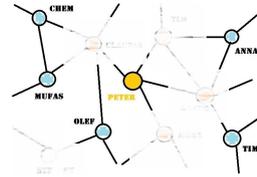
Das Alter des Nutzers ist das auf eine ganze Zahl gerundete arithmetische Mittel des angegebenen Alters der Freundes-Freunde.

Regel 5: (Median der Freundes-Freunde)

Das Alter des Nutzers ist der auf eine ganze Zahl gerundete Median des angegebenen Alters der Freundes-Freunde.

Regel 6: (Modus der Freundes-Freunde)

Das Alter des Nutzers entspricht dem Modus des angegebenen Alters der Freundes-Freunde. Bei mehreren Modi ist das Alter des Nutzers das auf eine ganze Zahl gerundete arithmetische Mittel dieser Werte.



Peter und die Freunde von Peters Freunden.

F.1.2. Übersicht über die Vorhersageregeln 7-10

Vorhersageregeln

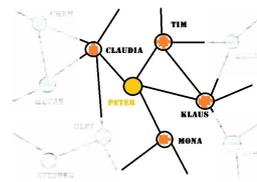
Für die Freunde:

Regel 7: (Spannweite der Freunde)

Das Alter des Nutzers liegt innerhalb der Spannweite des angegebenen Alters der Freunde.

Regel 8: (Quartilsabstand der Freunde)

Das Alter des Nutzers liegt innerhalb des Quartilsabstands des angegebenen Alters der Freunde.



Peter und die Freunde von Peter.

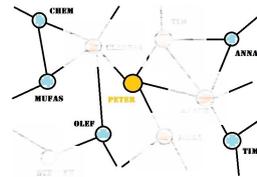
Für die Freundes-Freunde:

Regel 9: (Spannweite der Freundes-Freunde)

Das Alter des Nutzers liegt innerhalb der Spannweite des angegebenen Alters der Freundes-Freunde.

Regel 10: (Quartilsabstand der Freundes-Freunde)

Das Alter des Nutzers liegt innerhalb des Quartilsabstands des angegebenen Alters der Freundes-Freunde.



Peter und die Freunde von Peters Freunden.

F.1.3. Antwortblatt



Antwortblatt 2: Vorhersageregeln anwenden

Gesamtanzahl N der Nutzer (Stichprobengröße):	
Formel für die Trefferquote:	Trefferquote = $\frac{\text{Anzahl der Nutzer, die richtig eingeschätzt wurden}}{\text{Gesamtanzahl der Nutzer, die überprüft wurden}}$

Aufgabe 2 – Regeln durchlesen

Wir vermuten anhand der Betrachtung der konkreten Regeln, dass Regel _____ die beste Vorhersage über das Alter eines Nutzers macht, weil...

Im weiteren Verlauf betrachten wir die Vorhersageregeln für die _____ der Nutzer.

Aufgabe 3 – Regeln auf einzelne Nutzer anwenden

Notiert in der Tabelle, ob die Regeln das Alter des ausgewählten Nutzers richtig (✓) oder falsch (×) bestimmt haben.

Regel	Nutzer _____	Nutzer _____

Aufgabe 4 – Regeln auf viele Nutzer anwenden

Aufgabenteil a)

Regel:			
Anzahl richtiger Vorhersagen:			
Anzahl falscher Vorhersagen:			
Trefferquote (in %):			

Aufgabenteil b)

Warum ist es sinnvoll, die Vorhersageregeln auf alle Nutzer im Datensatz anzuwenden?

Aufgabe 5 – Der Median und das arithmetische Mittel

Aufgabenteil a)

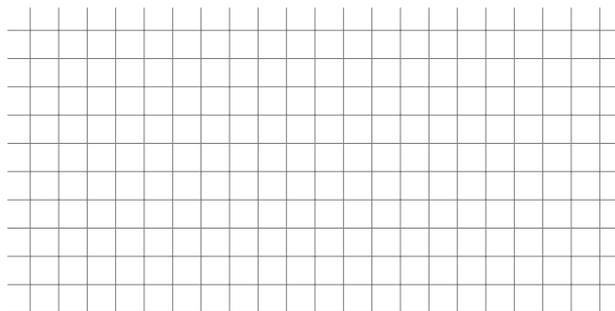
Vorhergesagtes Alter 2003 mit		Vorhergesagtes Alter 2004 mit	
Arithmetisches Mittel	Median	Arithmetisches Mittel	Median

Aufgabenteil b)

Welchen Unterschied könnt ihr bezogen auf Kims Alter zwischen dem arithmetischen Mittel und dem Median feststellen? Geht vor allem darauf ein, wie die Altersangaben von Kims Eltern im Jahr 2004 die beiden Werte beeinflussen.

Aufgabe 6 – Boxplot

Aufgabenteil a) Zeichnet einen Boxplot zu den angegebenen Daten.



Aufgabenteil d)

Interpretiert die Bedeutung der Spannweite und des Quartilsabstands. Beschreibt, wie viel Prozent der Daten innerhalb der beiden Werte liegen und welcher der beiden Werte sich am stärksten beim Auftreten eines Ausreißers verändern würde.

Aufgabe 7 – Regel zur Spannweite und zum Quartilsabstand

Aufgabenteil a) Wir vermuten, dass Regel _____ die beste Vorhersage macht, weil...

Aufgabenteil b)

Wir betrachten die Vorhersageregeln für die _____ der Nutzer.

Regel:		
Anzahl richtiger Vorhersagen:		
Anzahl falscher Vorhersagen:		
Trefferquote (in %):		

Aufgabenteil c) Wie groß sind für eure beiden Regeln im Durchschnitt die Altersspannen, in denen das vorhergesagte Alter des Nutzers liegt?

Für die Spannweite:

Für den Quartilsabstand:

Begründet, ob ihr die Spannweite oder den Quartilsabstand für eine Vorhersage des Alters wählen würdet. Geht dabei vor allem auf die Trefferquote und die Größe der durchschnittlichen Altersspannen ein.

F.1.4. Hilfekarte 1 - Wiederholung



Hilfekarte 1 - Wiederholung

Geht zu dem Abschnitt, zu dem ihr eine kleine Unterstützung benötigt:

1. Arithmetisches Mittel
2. Median
3. Modus

1) Arithmetisches Mittel \bar{x}



Zu der folgenden Datenreihe soll das arithmetische Mittel bestimmt werden:

[1 2 2 3 4 4 4 5]

Die **Summe der Werte** beträgt $1 + 2 * 2 + 3 + 3 * 4 + 5 = 22$.

Die **Anzahl der Werte** beträgt 8.

Dann ist das arithmetische Mittel der Datenreihe: $\bar{x} = \frac{\text{Summe aller Werte}}{\text{Anzahl der Werte}} = \frac{22}{8} = 2,75$

2) Median \tilde{x}

- Zunächst muss die Datenreihe aufsteigend vom kleinsten Wert sortiert werden. Das heißt, der kleinste Wert befindet sich am Anfang der Datenreihe und der größte Wert am Ende der Datenreihe.
- Der Median liegt genau in der Mitte einer sortierten Datenverteilung. Das heißt, die Hälfte der Daten ist immer kleiner, die andere Hälfte immer größer als der Median. Es werden folgende zwei Fälle unterschieden:



Ist die Anzahl der Daten **n ungerade**, befindet sich der Median bei einem sortierten Datensatz an der Position $\frac{n+1}{2}$.

11 12 12 13 13 14 15
|
Median

Ist die Anzahl der Daten **n gerade**, befindet sich der Median bei einem sortierten Datensatz zwischen den Positionen $\frac{n}{2}$ und $\frac{n}{2} + 1$. Man wählt dann das arithmetische Mittel dieser beiden Werte.

11 12 12 13,5 14 14 15 16
|
Median

3) Modus

Der Modus ist der Wert, welcher am häufigsten vorkommt. Kommen mehrere Werte gleich oft vor, wird der Mittelwert dieser Werte gebildet.



Zu der folgenden Datenreihe soll der Modus bestimmt werden:

[1 2 3 3 4 4 5 6]

In diesem Beispiel kommt sowohl der Wert 4 als auch der Wert 3 zwei Mal vor. Das heißt, der Modus dieser Datenreihe ist:

$$\frac{3+4}{2} = 3,5$$

F.1.5. Hilfekarte 2 - Boxplot



Hilfekarte 2 - Boxplot

Um einen Boxplot zu zeichnen, müssen folgende Schritte durchlaufen werden:

1. Die Daten sortieren
2. Maximum und Minimum bestimmen
3. Den Median bestimmen
4. Das untere Quartil und das obere Quartil bestimmen
5. Den Boxplot zeichnen

Je nachdem bei welchem Schritt ihr noch Schwierigkeiten habt, könnt ihr in den entsprechenden Abschnitt gehen, um eine kleine Unterstützung zu erhalten.

Hier nochmal die Altersangaben der Freunde von Julia, zu denen ihr den Boxplot erstellen sollt:

[42 40 41 38 42 42 40 45 41 43]

1) Die Daten sortieren

Als erstes müssen die Daten in die richtige Reihenfolge gebracht werden. Dazu werden die Daten aufsteigend vom kleinsten Wert aufgelistet. Das heißt, am Anfang steht der kleinste Wert und am Ende der größte Wert des vorliegenden Datensatzes.



Sortiert die Daten von Julia und überprüft euer Ergebnis mit *Run Section* .



```
Sortierte_Daten = [NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN];  
Boxplot_Schritt1(Sortierte_Daten)
```

2) Minimum und Maximum bestimmen

Das Minimum ist der kleinste Wert und das Maximum der größte Wert in eurem Datensatz. Wenn die Daten sortiert sind, ist das Minimum der erste und das Maximum der letzte Wert in eurer Liste.



Bestimmt das Minimum und das Maximum der Daten und überprüft euer Ergebnis mit *Run Section* .



```
Minimum = NaN;           % Tragt hier für NaN das Minimum ein
Maximum = NaN;           % Tragt hier für NaN das Maximum ein

Boxplot_Schritt2(Minimum, Maximum)
```

3) Den Median bestimmen

Der Median liegt genau in der Mitte einer sortierten Datenverteilung. Das heißt, die Hälfte der Daten ist immer kleiner, die andere Hälfte immer größer als der Median.

Ist die Anzahl der Daten **n ungerade**, befindet sich der Median bei einem sortierten Datensatz an der Position $\frac{n+1}{2}$.

11 12 12 13 13 14 15
 |
 Median

Ist die Anzahl der Daten **n gerade**, befindet sich der Median bei einem sortierten Datensatz zwischen den Positionen $\frac{n}{2}$ und $\frac{n}{2} + 1$. Man wählt dann das arithmetische Mittel dieser beiden Werte.

 13,5
11 12 12 13 | 14 14 15 16
 Median



Bestimmt den Median des Datensatzes und überprüft euer Ergebnis mit *Run Section* .



```
Median = NaN;           % Tragt hier für NaN den Median ein
```

```
Boxplot_Schritt3(Median)
```

4) Das untere und das obere Quartil bestimmen

Für das **untere Quartil** gilt: Höchstens ein Viertel (25 %) der Werte sind kleiner als das untere Quartil.

Für das **obere Quartil** gilt: Höchstens ein Viertel (25 %) der Werte sind größer als das obere Quartil.

Ist die Anzahl der Daten **gerade**, so teilt der Median die Datenreihe in zwei gleich große Datenhälften. Das untere Quartil ist dann der Median der unteren Datenhälfte und das obere Quartil der Median der oberen Datenhälfte

11	12		12	13		13	14		15	16
	12			13,5			14,5			
	unteres			Median			oberes			
	Quartil						Quartil			

Ist die Anzahl der Daten **ungerade**, so ist der Median der Wert in der Mitte. Auch hier sind die Quartile jeweils die Mediane der beiden Datenhälften.

11	12	12	13	13	14	15
	unteres		Median		oberes	
	Quartil				Quartil	



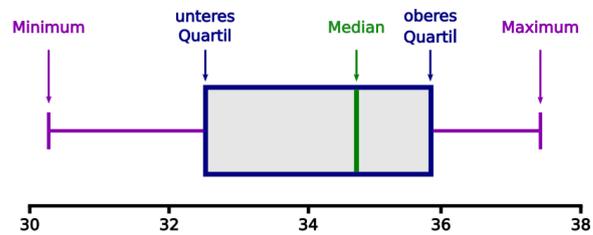
Bestimmt das untere und das obere Quartil der Datenverteilung und überprüft euer Ergebnis mit *Run Section* .



```
unteres_Quartil = NaN;           % Tragt für NaN das untere Quartil ein  
oberes_Quartil  = NaN;           % Tragt für NaN das obere Quartil ein
```

```
Boxplot_Schritt4(unteres_Quartil, oberes_Quartil)
```

5) Boxplot zeichnen



Zeichnet nun den Boxplot. Geht dabei wie folgt vor:

1. Zeichnet die Box ein, wobei das untere Quartil den linken Rand und das obere Quartil den rechten Rand der Box bilden.
2. Zeichnet den Median, wie in der obigen Abbildung gezeigt, an der entsprechenden Position ein.
3. Zeichnet das Minimum und das Maximum, wie in der obigen Abbildung gezeigt, ein.

Überprüft eure Zeichnung, indem ihr auf *Run Section*  drückt.



Boxplot_Schritt5

F.1.6. Besprechungsfolie

Arbeitsblatt 2 | Vorhersageregeln anwenden: Besprechung

Aufgabe 2 – Regeln durchlesen

Wie gut kann das Alter eines Nutzer mit den Regeln vorhergesagt werden?

Wir vermuten anhand der konkreten Regeln, dass _____ die beste Regel ist, weil...

Aufgabe 4 – Regeln auf viele Nutzer anwenden

	Regeln für die Freunde			Regeln für die Freundes-Freunde		
Regel:	1	2	3	4	5	6
Trefferquote:						

Warum ist es sinnvoll, die Vorhersageregeln auf alle Nutzer im Datensatz anzuwenden?

Aufgabe 5 – Der Median und das arithmetische Mittel

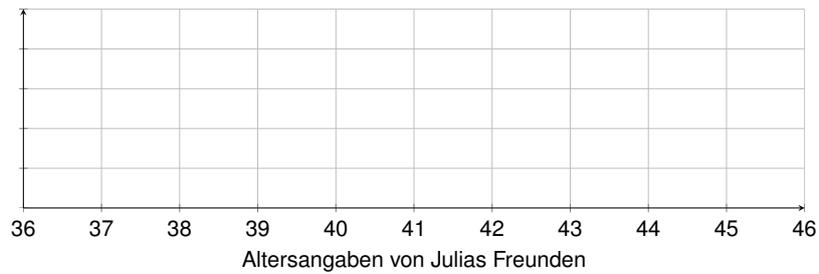
Kims wahres Alter: 22

Vorhergesagtes Alter 2003 mit		Vorhergesagtes Alter 2004 mit	
Arithmetisches Mittel	Median	Arithmetisches Mittel	Median

Welchen Unterschied könnt ihr bezogen auf Kims Alter zwischen dem arithmetischen Mittel und dem Median feststellen?

Aufgabe 6 – Boxplot

Zeichnet zu den gegebenen Daten einen Boxplot:



Spannweite =

Quartilsabstand =

Interpretation von Spannweite und Quartilsabstand:

Aufgabe 7 – Neue Regeln anwenden

Wir vermuten, dass Regel _____ die beste Vorhersage macht, weil...

Regel:	7	8	9	10
Trefferquote:				

Bei Betrachtung der Freunde:

Durchschnittliche Spannweite =

Durchschnittlicher Quartilsabstand =

Bei Betrachtung der Freundes Freunde:

Durchschnittliche Spannweite =

Durchschnittlicher Quartilsabstand =

F.2. Musterlösung für Arbeitsblatt 2

Arbeitsblatt 2 | Vorhersageregeln anwenden: Lösung

Aufgabe 1 – Auffrischung des Wissens

Das arithmetische Mittel, der Median und der Modus werden beispielhaft anhand folgendem Datensatz bestimmt:

[1 2 2 3 4 4 4 5]

Arithmetisches Mittel \bar{x} :

$$\bar{x} = \frac{1 + 2 * 2 + 3 + 3 * 4 + 5}{8} = \frac{22}{8} = 2,75$$

Median \tilde{x} :

Nach Cramer und Kamps (2017) ist der Median für metrische Daten wie folgt definiert ¹:

$x_{(1)} \leq \dots \leq x_{(n)}$ sei die Rangwertreihe eines metrischskalierten Datensatzes x_1, \dots, x_n . Der Median \tilde{x} ist definiert durch

$$\tilde{x} = \begin{cases} x_{(\frac{n+1}{2})}, & \text{falls } n \text{ ungerade,} \\ \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}), & \text{falls } n \text{ gerade.} \end{cases}$$

Für den obigen Datensatz liegt bereits die Rangwertreihe vor. Weiter ist $n = 8$ gerade. Daraus folgt:

$$\tilde{x} = \frac{1}{2}(x_{(4)} + x_{(5)}) = \frac{1}{2}(3 + 4) = 3,5$$

Modus:

Der Modus ist der Wert, welcher im Datensatz am häufigsten vorkommt. Kommen mehrere Werte gleich häufig vor, wird in diesem Lernmodul das arithmetische Mittel dieser Werte gebildet.

Für den obigen Datensatz gilt:

Modus: 4

Aufgabe 2 – Regeln durchlesen

Wie gut kann das Alter eines Nutzer mit den Regeln vorhergesagt werden?

Hinweis zur Besprechung: Hier sind verschiedene Lösungen möglich. Wichtig ist, dass die SuS ihre Antworten begründen. Hier soll lediglich ein erster Einschätzungsversuch der Regeln nach Gefühl vorgenommen werden.

Mögliche Antworten der SuS: Regel 1-3 sind genauer, da das Alter der engeren Freunde vermutlich mehr über das Alter des Nutzers aussagt. Auf der anderen Seite, haben die SuS bereits beim Filtern der Daten gesehen, dass die Nutzer im Schnitt nur etwa 3.5 Freunde besitzen. Dadurch könnten Regel 4-6 genauere Aussagen über das Alter eines Nutzers treffen, da bei den Freundes-Freunden mehr Daten berücksichtigt werden (im Durchschnitt etwa 23 Altersangaben).

Bei den Vor- und Nachteilen für das arithmetische Mittel, den Median und den Modus könnten die SuS auf mögliche Ausreißer bei den Altersangaben eingehen, welche vom arithmetischen Mittel wesentlich stärker berücksichtigt werden und dort eher zu einer verfälschten Vorhersage führen. Der Modus könnte dagegen zur besten Vorhersage führen, da sich ein Nutzer eventuell die meisten Freunde im gleichen Alter sucht.

¹vgl. E. Cramer & U. Kamps (2017): *Grundlagen der Wahrscheinlichkeitsrechnung und Statistik* (4. Auflage), Berlin: Springer Spektrum, S.27.

Aufgabe 4 – Regeln auf viele Nutzer anwenden

Aufgabenteil a)

Die angegebenen Trefferquoten beziehen sich auf eine Stichprobengröße von $N = 44747$ und sind lediglich als Richtwerte zu betrachten. Die Werte wurden auf zwei Nachkommastellen gerundet.

Regel:	Regeln für die Freunde			Regeln für die Freundes Freunde		
	1	2	3	4	5	6
Trefferquote:	16.61 %	18.75 %	18.39 %	11.06 %	14.23 %	14.41 %

Es ist zu erkennen, dass die Regeln für die Freunde durchweg zu höheren Trefferquoten führen als die der Freundes-Freunde. Außerdem kommen Regel 2 (Median) und Regel 3 (Modus) fast zum gleichen Ergebnis. Insgesamt ist aber festzuhalten, dass alle Trefferquoten zu gering sind, um das exakte Alter einer Person vorherzusagen.

Aufgabenteil b)

Warum ist es sinnvoll, die Vorhersageregeln auf alle Nutzer im Datensatz anzuwenden?

Je größer die Stichprobe N ist, desto aussagekräftiger und belastbarer sind die Ergebnisse. (Hinweis: Gesetz der großen Zahlen)

Aufgabe 5 – Der Median und das arithmetische Mittel

Kims wahres Alter: 22

Geordnete Altersangaben von Kims Freunden im Jahr 2003: 20 21 21 22 22 22 26

Geordnete Altersangaben von Kims Freunden im Jahr 2004: 20 21 21 22 22 22 26 56 60

Vorhergesagtes Alter 2003 mit		Vorhergesagtes Alter 2004 mit	
Arithmetisches Mittel	Median	Arithmetisches Mittel	Median
22	22	30	22

Im Jahr 2003 sind die Werte für das arithmetische Mittel sowie den Median gleich und entsprechen dem wahren Alter von Kim. Im Jahr 2004 sind in den Datensatz durch das Alter von Kims Eltern zwei Ausreißer hinzugekommen. Es ist zu erkennen, dass der Median von diesen extremen Werten nicht beeinflusst wurde, während das arithmetische Mittel für das Jahr 2004 um 8 Jahre größer ist.

Erklärung: Der Median ist gegenüber Ausreißern robust. Da er sich nicht aus den einzelnen Werten im Datensatz berechnet, sondern der Wert „im Zentrum“ ist, wird der Median von Ausreißern kaum beeinflusst. Anders ist das beim arithmetischen Mittel. Er berücksichtigt auch die Ausreißer, da er sich aus dem Durchschnitt aller Daten berechnet.

Aufgabe 6 – Boxplot

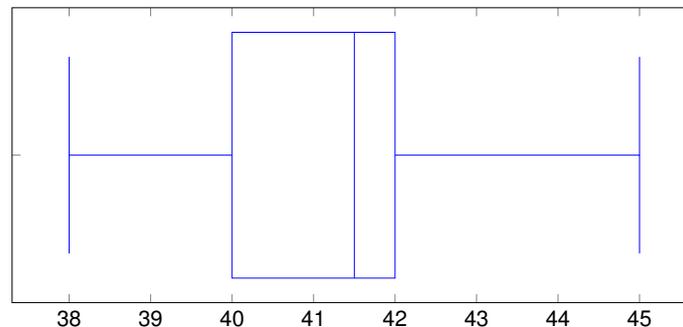
Aufgabenteil a)

Geordnete Altersangaben von Julias Freunden: 38 40 40 41 41 42 42 42 43 45

Hier die benötigten Werte für den Boxplot:

Minimum	Unteres Quartil	Median	Oberes Quartil	Maximum
38	40	41.5	42	45

Und der zugehörige Boxplot:



Aufgabenteil b)

Die Box wird durch das obere und das untere Quartil begrenzt, sodass die Länge der Box dem Quartilsabstand entspricht. Das untere Quartil $\tilde{x}_{0,25}$ ist so definiert, dass höchstens 25% aller Werte kleiner als $\tilde{x}_{0,25}$ sind. Ebenso gilt für das obere Quartil $\tilde{x}_{0,75}$, dass höchstens 25% aller Werte größer als $\tilde{x}_{0,75}$ sind.

⇒ Demnach befinden sich **mindestens 50 % der Werte** innerhalb der Box.

Aufgabenteil c)

Mit den in Aufgabenteil a) berechneten Werten ergibt sich:

$$\text{Spannweite} = x_{max} - x_{min} = 45 - 38 = 7$$

$$\text{Quartilsabstand} = \tilde{x}_{0,75} - \tilde{x}_{0,25} = 42 - 40 = 2$$

Aufgabenteil d) Interpretation von Spannweite und Quartilsabstand:

Die alleinige Angabe eines Lagemaßes, wie der Median, zur Beschreibung eines Datensatzes ist in der Regel unzureichend. So können beispielsweise die Werte in zwei unterschiedlichen Datensätzen mit dem selben Median von diesem Lagemaß unterschiedlich stark abweichen bzw. streuen. Das Abweichungsverhalten kann durch sogenannte Streuungsmaße, wie der Spannweite oder dem Quartilsabstand, angegeben werden. Allgemein kann die Aussage getroffen werden, je größer der Wert eines Streuungsmaßes ist, desto mehr streuen die Beobachtungen im Datensatz. Dabei muss jedoch immer beachtet werden, den konkreten Wert eines Streuungsmaßes in Abhängigkeit der Größenordnungen im Datensatz zu interpretieren.

Innerhalb der **Spannweite** liegen alle Daten eines Datensatzes. Sie basiert auf der Betrachtung der beiden extremsten Werte, also dem größten und dem kleinsten Wert. Damit ist die Spannweite jedoch sehr empfindlich auf Änderungen in diesen Werten. Vor allem Ausreißer haben einen direkten

3/5

Einfluss auf die Spannweite und können zu einem verfälschten Eindruck von der Streuung der Daten führen.

Der **Quartilsabstand** legt die Größen der Box in einem Boxplot fest und berechnet sich aus der Differenz von oberem und unterem Quartil der Daten. Wie bereits in Aufgabenteil b) festgestellt, liegen mindestens 50 % aller „zentralen“ Werte innerhalb des Quartilsabstands. Der Quartilsabstand wird durch Ausreißer in der Regel nicht verändert, da die größten und kleinsten Werte des Datensatzes zur Berechnung nicht herangezogen werden. Man sagt auch der Quartilsabstand ist robust gegenüber Ausreißern in der Stichprobe.

Aufgabe 7 – Neue Regeln anwenden

Aufgabenteil a)

Wie gut kann das Alter eines Nutzer mit den Regeln vorhergesagt werden?

Hinweis zur Besprechung: Wie in Aufgabe 2 sind auch hier verschiedene Lösungen möglich. Wichtig ist wieder, dass die SuS ihre Antworten begründen.

Mögliche Antwort der SuS: Die SuS könnten vermuten, dass die Spannweite eine bessere Vorhersage bzw. Trefferquote liefern wird. Denn in der Spannweite liegen alle Altersangaben der Freunde bzw. Freundes Freunde, wodurch das Intervall, in dem das Alter des Nutzers liegen soll, größer ist als beim Quartilsabstand.

Aufgabenteil b)

Die angegebenen Trefferquoten beziehen sich auf eine Stichprobengröße von **N = 44747** und sind lediglich als Richtwerte zu betrachten. Die Werte wurden auf zwei Nachkommastellen gerundet.

	Regeln für die Freunde		Regeln für die Freundes Freunde	
Regel:	7	8	9	10
Trefferquote:	53.21 %	47.34 %	76.07 %	56.11 %

Aufgabenteil c)

Bei Betrachtung der Freunde:

Durchschnittliche Spannweite = 6.47

Durchschnittlicher Quartilsabstand = 3.38

Bei Betrachtung der Freundes Freunde:

Durchschnittliche Spannweite = 19.97

Durchschnittlicher Quartilsabstand = 4.88

Begründete Wahl der Regel: Regel 9, also die Spannweite der Freundes Freunde, gibt zwar mit 76,07 % die beste Trefferquote ab, liefert jedoch mit einer durchschnittlichen Spannweite von etwa 20 ein zu ungenaues Intervall. Ein so großes, vorhergesagtes Altersintervall wäre für die Betreiber von sozialen Netzwerken nicht zufriedenstellend.

Werden die durchschnittlichen Intervallgrößen mit den Trefferquoten zusammen betrachtet, würde die Wahl auf Regel 10, der Quartilsabstand der Freundes Freunde, fallen. Diese Regel liefert nach Regel 9 mit 56,11 % die beste Trefferquote. Außerdem ist eine Intervallgröße von etwa 5 Jahren eine sehr gute Vorhersage. Auch ist die Größe des Intervalls im Vergleich zu Regel 8 nur um etwa 1 Jahr größer.

Insgesamt ergäbe sich auf Grundlage der Ergebnisse aus Aufgabenteil b) und c) für die Wahl der Regeln folgende Reihenfolge:

- 1) Regel 10,
- 2) Regel 8,
- 3) Regel 7,
- 4) Regel 9

Zusatzaufgabe | Teil I

Sortierte Altersangaben von Pauls Freunden: 12 12 13 13 13 15 15 16
Die Standardabweichung wird gerundet auf 2 Nachkommastellen angegeben.

Anzahl der Daten: $n = 8$

Arithmetisches Mittel: $\bar{x} = \frac{109}{8} = 13,625$

Standardabweichung: $s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_8 - \bar{x})^2}{n}} \approx 1,41$

Zusatzaufgabe | Teil II

Die angegebenen Trefferquoten beziehen sich auf eine Stichprobengröße von **N = 44747** und sind gerundet sowie lediglich als Richtwerte zu betrachten.

Regel:	für die Freunde	für die Freundes Freunde
Trefferquote:	53 %	72 %

Zusatzaufgabe | Teil III

Durchschnittliche Standardabweichung für die Freunde = 2,36

Durchschnittliche Standardabweichung für die Freundes Freunde = 4,81

Damit liegt die durchschnittliche Länge der vorhergesagten Altersintervallen bei der Betrachtung der Freunde bei etwa 5 Jahren und bei der Betrachtung der Freundes Freunde bei etwa 10 Jahren.

G. Arbeitsblatt 3 für Niveau 1

CAMMP day (Soziale Netzwerke) - Werbung schalten

Ihr seid nun Betreiber eines sozialen Netzwerkes und wollt mit dem Versand personalisierter Werbung Geld verdienen. Ihr habt eine Anfrage erhalten, bei der die Werbung Personen im Alter zwischen 25 und 30 Jahre erreichen soll. Um die Wünsche des Auftraggebers zu erfüllen, sollt ihr im Folgenden mit einer möglichst hohen Trefferquote Nutzer zwischen 25 und 30 Jahren in eurem sozialen Netzwerk identifizieren. Von euren Mitarbeitern wurden dazu Ideen vorgeschlagen, die ihr nun überprüfen sollt.



Erste Idee | Das arithmetische Mittel

Der erste Mitarbeiter schlägt folgende Vorhersageregeln vor:

Für die Freunde des Nutzers:

Man bestimmt für den Nutzer das gerundete arithmetische Mittel des angegebenen Alters der Freunde. Liegt das arithmetische Mittel zwischen 25 und 30 Jahre, dann befindet sich das Alter des Nutzers ebenfalls zwischen 25 und 30 Jahre.



a) Vorhersageregeln des 1. Mitarbeiters anwenden

Wendet die Vorhersageregeln des Mitarbeiters auf euren Datensatz an. Drückt dazu auf *Run*

Section . Füllt mit den Angaben die Spalte "Freunde des Nutzers" in der Tabelle auf eurem Antwortblatt aus.

b) Regel für die Freundes-Freunde anwenden

Die Regel des Mitarbeiters soll nun auf die Freundes-Freunde des Nutzers erweitert werden.

- Formuliert auf eurem Antwortblatt eine entsprechende Vorhersageregeln, welche sich auf die Freundes-Freunde eines Nutzers bezieht.
- Wendet die Vorhersageregeln für die Freundes-Freunde an. Tauscht dazu die Zahl 1 bei

Freundeskreis durch die Zahl 2 aus und drückt auf *Run Section* . Füllt auch hier mit den angezeigten Angaben die Spalte "Freundes-Freunde des Nutzers" in der Tabelle auf eurem Antwortblatt aus.



```
Freundeskreis = 1;      % Wählt 1 für die Freunde und 2 für die Freundes-Freunde
```

```
Daten
```

```
Check_meanIntervall(Users, Freundeskreis)
```

Zweite Idee | Entscheidungsregel mit Parametern

Ein anderer Mitarbeiter hat sich die folgende Vorhersageregeln ausgedacht:

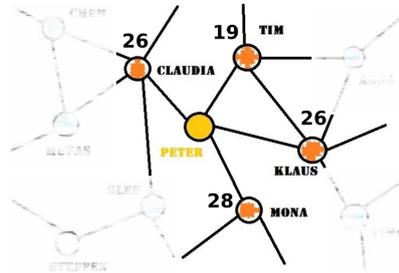
Für die Freunde des Nutzers:

Wenn der Nutzer mehr als **40% (=0.4)** Freunde mit einem Alter zwischen 25 und 30 Jahren besitzt, dann befindet sich sein eigenes Alter auch innerhalb dieser Altersspanne. Wenn der Nutzer weniger als **20% (=0.2)** Freunde mit einem Alter zwischen 25 und 30 Jahren besitzt, dann befindet sich sein eigenes Alter nicht in dieser Altersspanne. Sonst lasse den Zufall entscheiden.



a) Vorhersageregeln auf Peter anwenden

Wir wollen die Regel zunächst nur auf eine Person anwenden. Betrachtet dazu den Freundeskreis von Peter:



- Bestimmt, wie viel Prozent der Freunde von Peter zwischen 25 und 30 Jahre alt sind. Notiert den Wert auch auf eurem Antwortblatt.
- Begründet nun auf eurem Antwortblatt, wie die Vorhersage für Peter mit Hilfe der Regel des 2. Mitarbeiters lauten würde.



b) Vorhersageregeln des 2. Mitarbeiters anwenden

Wendet die Vorhersageregeln des Mitarbeiters auf euren Datensatz an. Drückt dazu auf *Run*

Section . Überträgt die Informationen auf euer Antwortblatt. Notiert auch, ob die Regel die Nutzer in der gesuchten Altersspanne (I) oder die Nutzer außerhalb der Altersspanne (\bar{I}) besser einschätzt.

c) Parameter ändern

Ändert die Parameter und versucht so eine noch bessere Trefferquote zu erzielen. Füllt für die gewählten Parameter die Tabelle auf eurem Antwortblatt aus.



Parameter = [0.4 0.2]; % Parameter für die Entscheidung

Daten
Check_Anteil(Users, Parameter)



Zusatzaufgabe Altersüberprüfung

Im ersten Arbeitsblatt habt ihr unter anderem die Nutzer aus eurem Datensatz herausgefiltert, die ein Alter von 0 Jahren angegeben haben. Wir wollen nun mit den bereits herausgefundenen Vorhersageregeln das tatsächliche Alter dieser Nutzer bestimmen.



a) Nutzer IDs mit einer Altersangabe von 0 Jahren herausfinden

Wir betrachten einen Datensatz, der auch noch die Nutzer mit einem Alter von 0 Jahren enthält. Zunächst müssen wir die Position dieser Nutzer in dem Datensatz (kurz Nutzer ID) herausfinden.



Klickt auf *Run Section*  und ihr bekommt die IDs der ersten 20 Nutzer mit einem Alter von 0 Jahren angezeigt.



```
Datensatz_laden
```

```
IDs_herausfinden(Users)
```

b) Alter vorhersagen



Sucht euch drei Nutzer heraus, deren wahres Alter ihr vorhersagen möchtet. Notiert die drei Nutzer IDs auf eurem Antwortblatt. Wählt anschließend für jeden der drei ausgesuchten Nutzer mindestens eine Vorhersageregeln aus, welche das exakte Alter vorhersagt und eine, die dem Nutzer eine Altersspanne

zuordnet. Tragt unten für NaN sowohl die ausgewählte Nutzer ID als auch die Regel ein und drückt auf *Run Section* , um eine Vorhersage über das wahre Alter zu treffen.



```
Nutzer_ID = NaN;      % Nutzer, dessen Alter vorhergesagt werden soll
Regel      = NaN;      % Regel (Zahl zwischen 1 und 10), mit der das
                       % Alter vorhergesagt werden soll.

Altersvorhersage(Users, Nutzer_ID, Regel)
```

G.1. Zusatzmaterial von Arbeitsblatt 3 für Niveau 1

G.1.1. Antwortblatt für Niveau 1

Antwortblatt 3: Werbung schalten

Gesamtanzahl N der Nutzer (Stichprobengröße):	
---	--

Erste Idee | Das arithmetische Mittel

Aufgabenteil a-b)

Untersucht:	Freunde des Nutzers	Freundes Freunde des Nutzers
Trefferquote einer richtigen Vorhersage innerhalb der gesuchten Altersspanne (in %):		
Trefferquote einer richtigen Vorhersage außerhalb der gesuchten Altersspanne (in %):		
Zusammenfassung:	•	•
Trefferquote für eine richtige Vorhersage (in %):		
Was wird besser vorhergesagt? I oder \bar{I} :		

Regel für die Freundes Freunde des Nutzers:

Zweite Idee | Entscheidungsregel mit Parametern

Aufgabenteil a)

Wie viel Prozent der Freunde von Peter sind zwischen 25 und 30 Jahre alt?: _____

Wie würde mit der Regel des 2. Mitarbeiters die Vorhersage des Alters für Peter lauten?

Aufgabenteil b-c) Ergebnisse nach Anwendung der Regel für die Freunde der Nutzer:

Parameter:	[0, 4 0, 2]			
Trefferquote einer richtigen Vorhersage innerhalb der gesuchten Altersspanne (in %):				
Trefferquote einer richtigen Vorhersage außerhalb der gesuchten Altersspanne (in %):				
Zusammenfassung:	•	•	•	•
Trefferquote für eine richtige Vorhersage (in %):				
Was wird besser vorhergesagt? I oder \bar{I} :				

Zusatzaufgabe Altersüberprüfung

Aufgabenteil b)

•	Vorhergesagtes Alter mit Regel				
•					
Für Nutzer mit ID _____					
Für Nutzer mit ID _____					
Für Nutzer mit ID _____					

G.1.2. Besprechungsfolie für Niveau 1

Arbeitsblatt 3 | Werbung schalten: Besprechung

Erste Idee | Das arithmetische Mittel

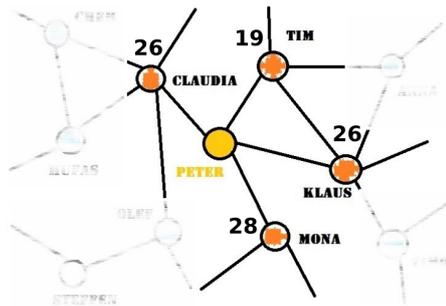
Aufgabenteil a-b)

Untersucht:	Freunde des Nutzers	Freundes Freunde des Nutzers
Trefferquote einer richtigen Vorhersage innerhalb der gesuchten Altersspanne (in %):		
Trefferquote einer richtigen Vorhersage außerhalb der gesuchten Altersspanne (in %):		
Zusammenfassung:	•	•
Trefferquote für eine richtige Vorhersage (in %):		
Was wird besser vorhergesagt? I oder \bar{I} :		

Regel für die Freundes Freunde des Nutzers:

Zweite Idee | Entscheidungsregel mit Parametern

Aufgabenteil a)



Wie viel Prozent der Freunde von Peter sind zwischen 25 und 30 Jahre alt?:

Aufgabenteil b-c) Ergebnisse nach Anwendung der Regel für die Freunde der Nutzer:

Parameter:	[0, 4 0, 2]			
Trefferquote einer richtigen Vorhersage innerhalb der gesuchten Altersspanne (in %):				
Trefferquote einer richtigen Vorhersage außerhalb der gesuchten Altersspanne (in %):				
Zusammenfassung:	•	•	•	•
Trefferquote für eine richtige Vorhersage (in %):				
Was wird besser vorhergesagt? I oder \bar{I} :				

G.2. Musterlösung für Arbeitsblatt 3 Niveau 1

AB 3 | Werbung schalten: Lösung Niveau 1

Hinweis: Die Lösung wurde mit einer Datensatzgröße von **44747** Nutzern erstellt.

Erste Idee | Das arithmetische Mittel

Aufgabenteil a-b)

Die Regel für die Freundes-Freunde lautet:

Man bestimmt für den Nutzer das gerundete arithmetische Mittel des angegebenen Alters der **Freundes-Freunde**. Liegt das arithmetische Mittel zwischen 25 und 30 Jahren, dann befindet sich das Alter des Nutzers ebenfalls zwischen 25 und 30 Jahren.

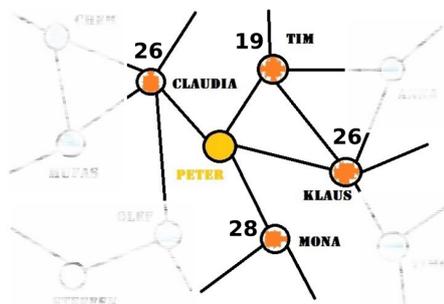
Die ermittelten Trefferquoten beziehen sich wieder auf einen Datensatz mit 44747 Nutzern.

Untersucht:	Freunde des Nutzers	Freundes-Freunde des Nutzers
Trefferquote einer richtigen Vorhersage innerhalb der gesuchten Altersspanne (in %):	71 %	56 %
Trefferquote einer richtigen Vorhersage außerhalb der gesuchten Altersspanne (in %):	87 %	88 %
Zusammenfassung:	•	•
Trefferquote für eine richtige Vorhersage (in %):	80 %	75 %
Was wird besser vorhergesagt? I oder \bar{I} :	\bar{I}	\bar{I}

Sowohl bei der Betrachtung der Freunde, als auch bei der Betrachtung der Freundes-Freunde sind die Trefferquoten für eine richtige Vorhersage innerhalb der gesuchten Altersspanne geringer als bei der zweiten Idee.

Zweite Idee | Entscheidungsregel mit Parametern

Aufgabenteil a)



Wie viel Prozent der Freunde von Peter sind zwischen 25 und 30 Jahre alt?:

75 %

⇒ **Antwort:** Da mehr als 40 % der Freunde von Peter zwischen 25 und 30 Jahren alt sind, wird Peter mit der Regel des 2. Mitarbeiters ebenfalls ein Alter zwischen 25 und 30 Jahren vorhergesagt.

Aufgabenteil b-c) Folgende Ergebnisse beziehen sich ebenfalls auf eine Stichprobengröße von **44747** Nutzern. Die angegebenen Parameter sind wieder nur als Beispiele zu betrachten.

Parameter:	[0,4 0,2]	[0,8 0,4]	[0,3 0,2]	[0,45 0,35]
Trefferquote einer richtigen Vorhersage innerhalb der gesuchten Altersspanne (in %):	85 %	69 %	87 %	83 %
Trefferquote einer richtigen Vorhersage außerhalb der gesuchten Altersspanne (in %):	76 %	86 %	73 %	80 %
Zusammenfassung:	•	•	•	•
Trefferquote für eine richtige Vorhersage (in %):	80 %	79 %	79 %	81 %
Was wird besser vorhergesagt? I oder \bar{I} :	I	\bar{I}	I	beide

Hinweis: Viele Parameterkombinationen führen zu Trefferquoten an die 80 %. Jedoch ist es für den Betreiber des sozialen Netzwerkes wichtig, dass besonders die Personen zwischen 25 und 30 Jahren gut vorhergesagt werden. Aus diesem Grund wäre beispielsweise von einer Wahl der Parameter [0,8 0,4] abzuraten.

Zusatzaufgabe Altersüberprüfung

Die SuS erhalten die ersten 20 Nutzer IDs, die ein Alter von 0 Jahre aufweisen. Sie wählen 3 dieser angezeigten Nutzer aus und versuchen mit den bereits bekannten Vorhersageregeln 1 bis 10 das wahre Alter der Nutzer herauszufinden. Mit den SuS kann hier nochmal auf die herausgefundenen Trefferquoten der einzelnen Regeln eingegangen werden, um die Zuverlässigkeit des vorhergesagten Alters zu diskutieren.

H. Arbeitsblatt 3 für Niveau 2

CAMMP day (Soziale Netzwerke) - Werbung schalten

Ihr seid nun Betreiber eines sozialen Netzwerkes und wollt mit dem Versand personalisierter Werbung Geld verdienen. Ihr habt eine Anfrage erhalten, bei der die Werbung Personen im Alter zwischen 25 und 30 Jahre erreichen soll. Um die Wünsche des Auftraggebers zu erfüllen, sollt ihr im Folgenden mit einer möglichst hohen Trefferquote Nutzer zwischen 25 und 30 Jahren in eurem sozialen Netzwerk identifizieren. Von euren Mitarbeitern wurden dazu Ideen vorgeschlagen, die ihr nun überprüfen sollt.



Erste Idee | Der Zufall soll entscheiden

Der erste Mitarbeiter schlägt folgende Vorhersageregeln vor:

Der Zufall:

Mit einer Wahrscheinlichkeit von p ist der Nutzer zwischen 25 und 30 Jahre alt. Mit der Wahrscheinlichkeit $1-p$ ist er es nicht.



Die Idee dahinter ist, für jeden Nutzer eine gezinkte Münze zu werfen. Zeigt die Münze Kopf, ordnen wir dem Nutzer ein Alter zwischen 25 und 30 Jahre zu. Zeigt die Münze Zahl, dann wird das Alter des Nutzers außerhalb der gesuchten Altersspanne vorhergesagt.



a) Relative Häufigkeit der Nutzer zwischen 25 und 30 Jahre bestimmen

Zunächst betrachten wir, wie viele Personen in eurem Datensatz zwischen 25 und 30 Jahre alt sind. Mit dieser Angabe legen wir die Wahrscheinlichkeit p fest, mit der die Münze Kopf zeigen soll.



- Klickt auf *Run Section* , um die absolute Häufigkeit der Nutzer zwischen 25 und 30 Jahre in eurem Datensatz zu erfahren. Tragt diesen Wert auf eurem Antwortblatt ein.
- Bestimmt nun die relative Häufigkeit der Nutzer zwischen 25 und 30 Jahre in eurem Datensatz. Gebt den Wert als eine Kommazahl zwischen 0 und 1 an und rundet sie auf 2 Nachkommastellen.
- Überprüft euren Wert mit Matlab und tragt ihn auf eurem Antwortblatt ein.



```
relative_Haeufigkeit = NaN;    % relative Häufigkeit als ein Wert zwischen 0 und 1  
  
Daten  
Uebersicht_Zufall(Users,relative_Haeufigkeit)
```

b) Zufallsregel anwenden

Ihr habt nun die relative Häufigkeit p bestimmt, mit der ein Nutzer zwischen 25 und 30 Jahre alt ist. Nun werft ihr für jeden Nutzer eure gezinkte Münze, die mit der Wahrscheinlichkeit p Kopf anzeigt. Zeigt die gezinkte Münze Kopf, wird die Vorhersage getroffen, dass der Nutzer zwischen 25 und 30 Jahre alt ist. Wendet nun die Zufallsregel auf euren Datensatz an.



- Gebt zunächst die Anzahl der Nutzer N ein, auf welche die Vorhersageregeln (Zufall) angewendet werden soll.
- Klickt auf *Run Section* .
- Tragt die Trefferquote auf eurem Antwortblatt ein.
- Variiert die Anzahl der Nutzer N .



```
N = NaN;    % Anzahl der Nutzer, auf welche die Regeln angewendet werden soll.
```

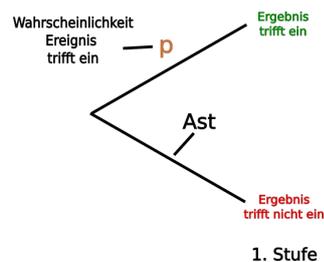
```
Check_Zufallsregel(Users, N)
```

c) Theoretische Trefferquote

Wir wollen nun die erhaltenen Trefferquoten von Aufgabenteil b) mit der erwarteten, theoretischen Trefferquote vergleichen. Denn es wird erwartet, dass sich die ermittelte Trefferquote für eine größer werdende Stichprobe N immer mehr an die theoretische Trefferquote annähert.



Zeichnet dazu auf eurem Antwortblatt ein zweistufiges Baumdiagramm, mit welchem ihr die theoretische Wahrscheinlichkeit (bzw. Trefferquote), dass das Alter eines beliebigen Nutzers mit Hilfe des Zufalls richtig vorhergesagt wird, bestimmen könnt. Nutzt dazu die ermittelte Wahrscheinlichkeit p , dass ein Nutzer zwischen 25 und 30 Jahre alt ist. Bestimmt anschließend mit Hilfe des Baumdiagramms die theoretische Trefferquote. Rundet euer Ergebnis auf 2 Nachkommastellen und gebt es wieder als einen Wert zwischen 0 und 1 an. Überprüft euer Ergebnis mit Matlab und tragt es anschließend auf eurem Antwortblatt ein.



Hinweis: Zu dieser Aufgabe gibt es eine **Hilfekarte**. Öffnet dazu die Datei `Hilfekarte_Baumdiagramm.mlx`



```
Loesung = NaN; % Wahrscheinlichkeit, dass das Alter mit dem Zufall richtig
               % vorhergesagt wird.

Check_Loesung_Wahrscheinlichkeit(Users, Loesung)
```

d) Trefferquoten vergleichen



Vergleiche die herausgefundenen Trefferquoten aus Aufgabenteil b) mit der theoretischen Trefferquote aus Aufgabenteil c). Geht dabei vor allem auf die Größe N der gewählten Stichproben in Aufgabenteil b) ein.



Zweite Idee | Entscheidungsregel mit Parametern

Ein anderer Mitarbeiter hat sich die folgende Vorhersageregeln ausgedacht:

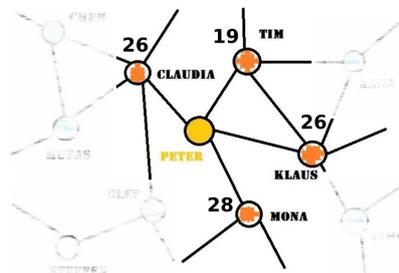
Für die Freunde des Nutzers:

Wenn der Nutzer mehr als **40% (=0.4)** Freunde mit einem Alter zwischen 25 und 30 Jahren besitzt, dann befindet sich sein eigenes Alter auch innerhalb dieser Altersspanne. Wenn der Nutzer weniger als **20% (=0.2)** Freunde mit einem Alter zwischen 25 und 30 Jahren besitzt, dann befindet sich sein eigenes Alter nicht in dieser Altersspanne. Sonst lasse den Zufall entscheiden.



a) Vorhersageregeln auf Peter anwenden

Wir wollen die Regeln zunächst nur auf eine Person anwenden. Betrachte dazu den Freundeskreis von Peter:





- Bestimmt, wie viel Prozent der Freunde von Peter zwischen 25 und 30 Jahre alt sind. Notiert den Wert auch auf eurem Antwortblatt.
- Begründet nun auf eurem Antwortblatt, wie die Vorhersage für Peter mit Hilfe der Regel des 2. Mitarbeiters lauten würde.



b) Vorhersageregeln des 2. Mitarbeiters anwenden

Wendet die Vorhersageregeln des Mitarbeiters auf euren Datensatz an. Drückt dazu auf *Run*

Section . Überträgt die Informationen auf euer Antwortblatt. Notiert auch, ob die Regel die Nutzer in der gesuchten Altersspanne (I) oder die Nutzer außerhalb der Altersspanne (\bar{I}) besser einschätzt.

c) Parameter ändern

Ändert die Parameter und versucht so eine noch bessere Trefferquote zu erzielen. Füllt für die gewählten Parameter die Tabelle auf eurem Antwortblatt aus.



```
Parameter = [0.4 0.2]; % Parameter für die Entscheidung
```

```
Daten  
Check_Anteil(Users, Parameter)
```



Zusatzaufgabe

Lasst euch von einem Betreuer das Antwortblatt für die Zusatzaufgaben geben und öffnet die Matlab-Datei *Zusatzaufgaben_AB3.mlx*.

H.1. Zusatzmaterial von Arbeitsblatt 3 für Niveau 2

H.1.1. Antwortblatt für Niveau 3

CAMMP day
Soziale Netzwerke Sek. I



RWTH AACHEN
UNIVERSITY

Antwortblatt 3: Werbung schalten

Gesamtanzahl N der Nutzer (Stichprobengröße):	
---	--

Erste Idee | Der Zufall entscheidet

Aufgabenteil a)

Absolute Häufigkeit der Nutzer zwischen 25 und 30 Jahren:	Relative Häufigkeit p der Nutzer zwischen 25 und 30 Jahren:

Aufgabenteil b) Ergebnisse nach Anwendung der Zufallsregel:

Größe N der gewählten Stichprobe:			
Trefferquote (in %):			

Aufgabenteil c) Zeichnet, wie in der Aufgabe beschrieben, das zweistufige Baumdiagramm:

Theoretische Trefferquote: _____ %

Aufgabenteil d) Vergleichen Sie die herausgefundenen Trefferquoten aus Aufgabenteil b) mit der theoretischen Trefferquote aus Aufgabenteil c). Gehen Sie dabei vor allem auf die Größe N der gewählten Stichproben ein.

Zweite Idee | Entscheidungsregel mit Parametern

Aufgabenteil a)

Wie viel Prozent der Freunde von Peter sind zwischen 25 und 30 Jahre alt?: _____

Wie würde mit der Regel des 2. Mitarbeiters die Vorhersage des Alters für Peter lauten?

Aufgabenteil b-c) Ergebnisse nach Anwendung der Regel für die Freunde der Nutzer:

Parameter:	[0,4 0,2]			
Trefferquote einer richtigen Vorhersage innerhalb der gesuchten Altersspanne (in %):				
Trefferquote einer richtigen Vorhersage außerhalb der gesuchten Altersspanne (in %):				
Zusammenfassung:	•	•	•	•
Trefferquote für eine richtige Vorhersage (in %):				
Was wird besser vorhergesagt? I oder \bar{I} :				

H.1.2. Zusatzarbeitsblatt für Niveau 2

Wichtig

Es gibt zwei Zusatzaufgaben und ihr könnt zunächst aussuchen, welche der beiden Aufgaben ihr bearbeiten wollt. In der ersten Zusatzaufgabe könnt ihr versuchen das tatsächliche Alter der Nutzer, die eine Altersangabe von 0 Jahre angegeben haben, mit euren Vorhersageregeln herauszufinden. Die zweite Zusatzaufgabe betrachtet nochmal die theoretische Trefferquote aus der Idee des ersten Mitarbeiters etwas genauer.

1) Zusatzaufgabe | Altersüberprüfung

Im ersten Arbeitsblatt habt ihr unter anderem die Nutzer aus eurem Datensatz herausgefiltert, die ein Alter von 0 Jahren angegeben haben. Wir wollen nun mit den bereits herausgefundenen Vorhersageregeln das tatsächliche Alter dieser Nutzer bestimmen.



a) Nutzer IDs mit einer Altersangabe von 0 Jahren herausfinden

Wir betrachten einen Datensatz, der auch noch die Nutzer mit einem Alter von 0 Jahren enthält. Zunächst müssen wir die Position dieser Nutzer in dem Datensatz (kurz Nutzer ID) herausfinden.



Klickt auf *Run Section*  und ihr bekommt die IDs der ersten 20 Nutzer mit einem Alter von 0 Jahren angezeigt.



```
Datensatz_laden
```

```
IDs_herausfinden(Users)
```

b) Alter vorhersagen



Sucht euch drei Nutzer heraus, deren wahres Alter ihr vorhersagen möchtet. Notiert die drei Nutzer IDs auf eurem Antwortblatt. Wählt anschließend für jeden der drei ausgesuchten Nutzer mindestens eine Vorhersageregeln aus, welche das exakte Alter vorhersagt und eine, die dem Nutzer eine Altersspanne zuordnet. Tragt unten für NaN sowohl die ausgewählte Nutzer ID als auch die Regel ein und drückt

auf `Run Section` , um eine Vorhersage über das wahre Alter zu treffen.

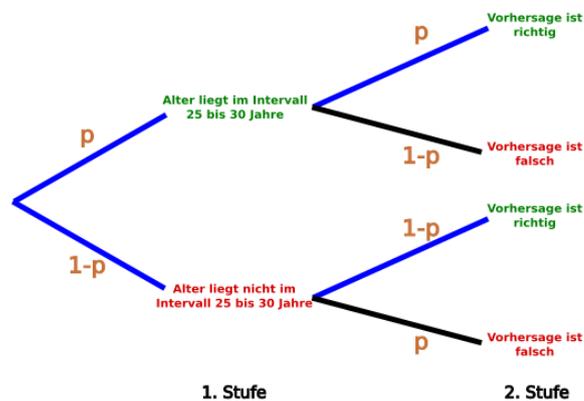


```
Nutzer_ID = NaN;      % Nutzer, dessen Alter vorhergesagt werden soll
Regel      = NaN;      % Regel (Zahl zwischen 1 und 10), mit der das
                       % Alter vorhergesagt werden soll.
```

```
Altersvorhersage(Users, Nutzer_ID, Regel)
```

2) Zusatzaufgabe | Theoretische Trefferquote

Für die Idee des ersten Mitarbeiters habt ihr folgendes Baumdiagramm aufgestellt, um die theoretische Trefferquote für eine richtigen Vorhersage mit Hilfe einer zufälligen Zuordnung zu bestimmen:



Dabei steht p für eure bestimmte relative Häufigkeit einen Nutzer mit einem Alter zwischen 25 und 30 Jahre in eurem Datensatz zu finden. Mit dieser Wahrscheinlichkeit habt ihr einem Nutzer ein Alter zwischen 25 und 30 Jahre zugeordnet.

a) relative Häufigkeit p



Klickt auf *Run Section* , um euch nochmal eure relative Häufigkeit p anzeigen zu lassen. Diesen Wert habt ihr bei der Idee des ersten Mitarbeiters für die Erstellung eures Baumdiagramms benutzt.



```
relative_Haeufigkeit
```

b) Wahrscheinlichkeit variieren

Bei der Idee des ersten Mitarbeiters habt ihr die relative Häufigkeit p als Wahrscheinlichkeit gewählt, einem Nutzer ein Alter zwischen 25 und 30 Jahre zuzuordnen. Als nächstes sollt ihr euch überlegen, wie sich die theoretische Trefferquote für eine richtige Vorhersage ändert, wenn ihr die Wahrscheinlichkeit einem Nutzer ein Alter innerhalb der Altersspanne 25 bis 30 Jahre zuzuordnen variiert.



Tragt dazu unten im Code unter Wahrscheinlichkeit die Wahrscheinlichkeit (als Zahl zwischen 0 und 1) ein, mit der ihr einem Nutzer ein Alter zwischen 25 und 30 Jahre zuordnen wollt. Berechnet anschließend die theoretische Trefferquote, die ihr für eure gewählte Wahrscheinlichkeit erwarten würdet. Gebt die theoretische Trefferquote auf zwei Nachkommastellen genau an. Drückt auf *Run*

Section , um eure berechnete theoretische Trefferquote zu überprüfen. Notiert eure Ergebnisse auch auf dem Antwortblatt. Überlegt euch zudem mit welcher Wahrscheinlichkeit ihr die beste theoretische Trefferquote erhaltet.



```
Wahrscheinlichkeit = NaN; % Wahrscheinlichkeit, mit der einem Nutzer ein Alter
                        % zwischen 25 und 30 Jahren zugeordnet wird.
Trefferquote        = NaN; % Die mit der gewählten Wahrscheinlichkeit zu
                        % erwartene theoretische Trefferquote

Loesung(Users, Wahrscheinlichkeit, Trefferquote)
```

H.1.3. Antwortblatt zur Zusatzaufgabe

Antwortblatt 3: Zusatzaufgaben

1) Zusatzaufgabe | Altersüberprüfung

Aufgabenteil b)

	Vorhergesagtes Alter mit Regel				
•					
•					
Für Nutzer mit ID _____					
Für Nutzer mit ID _____					
Für Nutzer mit ID _____					

2) Zusatzaufgabe | Theoretische Trefferquote

Aufgabenteil b)

Gewählte Wahrscheinlichkeit:				
Berechnete theoretische Trefferquote (in %):				

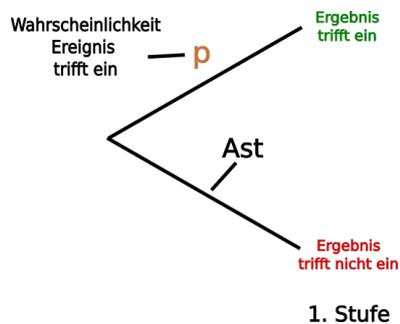
H.1.4. Hilfekarte 3 - Baumdiagramm



Hilfekarte 3 - Baumdiagramm

Je nachdem bei welchem Schritt ihr noch Schwierigkeiten habt, könnt ihr in den entsprechenden Abschnitt gehen, um eine kleine Unterstützung zu erhalten.

1. Die zwei Stufen im Baumdiagramm darstellen
2. Die Wahrscheinlichkeiten an den Ästen eintragen
3. Die richtigen Pfade aussuchen
4. Die Wahrscheinlichkeit der richtigen Pfade bestimmen
5. Die gesuchte Wahrscheinlichkeit berechnen



1) Die zwei Stufen im Baumdiagramm darstellen

In der ersten Stufe wird festgehalten, ob der Person ein Alter zwischen 25 und 30 Jahre zugeordnet wird oder nicht. In der zweiten Stufe wird dann notiert, ob die Vorhersage des Alters richtig oder falsch ist.



- Zeichnet die beiden Stufen in ein Baumdiagramm ein.
- Mit *Run Section*  könnt ihr euch die Lösung anschauen.



Baumdiagramm_Schritt1

2) Die Wahrscheinlichkeiten an den Ästen eintragen

An jedem Ast steht die Wahrscheinlichkeit, dass das jeweilige Ereignis eintritt. So steht an dem Ast, der zu dem Ereignis "Alter liegt zwischen 25 und 30 Jahre" gehört, eure berechnete Wahrscheinlichkeit p aus Aufgabenteil a). An dem anderen Ast steht dann die Gegenwahrscheinlichkeit $1-p$. Denn insgesamt müssen beide Äste addiert immer auf den Wert 1 bzw. 100% kommen.



- Tragt die Wahrscheinlichkeiten an jeden Ast in eurem Baumdiagramm ein.
- Mit *Run Section*  könnt ihr euch die Lösung anschauen.



Baumdiagramm_Schritt2

3) Die richtigen Pfade aussuchen

Ein Weg vom Anfang bis zu einem Ende des Baumdiagramms wird als Pfad bezeichnet.



- Markiert in eurem Baumdiagramm die Pfade, welche zu einer richtigen Vorhersage führen.
- Mit *Run Section*  könnt ihr euch die Lösung anschauen.



Baumdiagramm_Schritt3

4) Die Wahrscheinlichkeit der richtigen Pfade bestimmen

Die Wahrscheinlichkeit entlang eines Pfades wird mit der 1. Pfadregel, auch Produktregeln genannt, bestimmt:

Produktregel:

Die Wahrscheinlichkeit eines zusammengesetzten Ereignisses in einem mehrstufigen Vorgang ist gleich dem Produkt der Wahrscheinlichkeiten längs des Pfades, der zu diesem Ereignis führt.



- Bestimmt mit der Produktregel die Wahrscheinlichkeit der beiden gesuchten Pfade.
- Mit *Run Section*  könnt ihr euch die Lösung anschauen.



Baumdiagramm_Schritt4

5) Die gesuchte Wahrscheinlichkeit berechnen

Erfüllen mehrere Pfade die Fragestellung setzt sich die gesuchte Wahrscheinlichkeit mit Hilfe der 2. Pfadregel, auch Summenregel genannt, zusammen:

Summenregel:

Die Wahrscheinlichkeit eines Ereignisses in einem mehrstufigen Vorgang ist gleich der Summe der Wahrscheinlichkeiten der für dieses Ereignis günstigen Pfade.



- Bestimmt mit Hilfe der Summenregel die gesuchte Wahrscheinlichkeit.

H.1.5. Besprechungsfolie für Niveau 2

CAMMP day
Soziale Netzwerke Sek. I



RWTHAACHEN
UNIVERSITY

Arbeitsblatt 3 | Werbung schalten: Besprechung

Erste Idee | Der Zufall entscheidet

Aufgabenteil a)

Beispiele für die relative Häufigkeit p:				
--	--	--	--	--

Aufgabenteil b) Ergebnisse nach Anwendung der Zufallsregel:

Größe N der gewählten Stichprobe:				
Trefferquote (in %):				

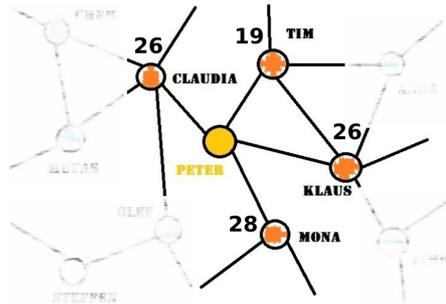
Aufgabenteil c) Zeichnet, wie in der Aufgabe beschrieben, das zweistufige Baumdiagramm:

Theoretische Trefferquote: _____

Aufgabenteil d) Vergleicht die herausgefundenen Trefferquoten aus Aufgabenteil b) mit der theoretischen Trefferquote aus Aufgabenteil c). Geht dabei auch auf die Größe N der gewählten Stichproben ein.

Zweite Idee | Entscheidungsregel mit Parametern

Aufgabenteil a)



Wie viel Prozent der Freunde von Peter sind zwischen 25 und 30 Jahre alt?:

Aufgabenteil b-c) Ergebnisse nach Anwendung der Regel für die Freunde der Nutzer:

Parameter:	[0, 4 0, 2]			
Trefferquote einer richtigen Vorhersage innerhalb der gesuchten Altersspanne (in %):				
Trefferquote einer richtigen Vorhersage außerhalb der gesuchten Altersspanne (in %):				
Zusammenfassung:	•	•	•	•
Trefferquote für eine richtige Vorhersage (in %):				
Was wird besser vorhergesagt? I oder \bar{I} :				

H.2. Musterlösung für Arbeitsblatt 3 Niveau 2

AB 3 | Werbung schalten: Lösung Niveau 2

Hinweis: Die Lösung wurde mit einer Datensatzgröße von **44747** Nutzern erstellt.

Erste Idee | Der Zufall entscheidet

Aufgabenteil a)

Absolute Häufigkeit der Nutzer zwischen 25 und 30 Jahren:	Relative Häufigkeit p der Nutzer zwischen 25 und 30 Jahren:
18701	0.42

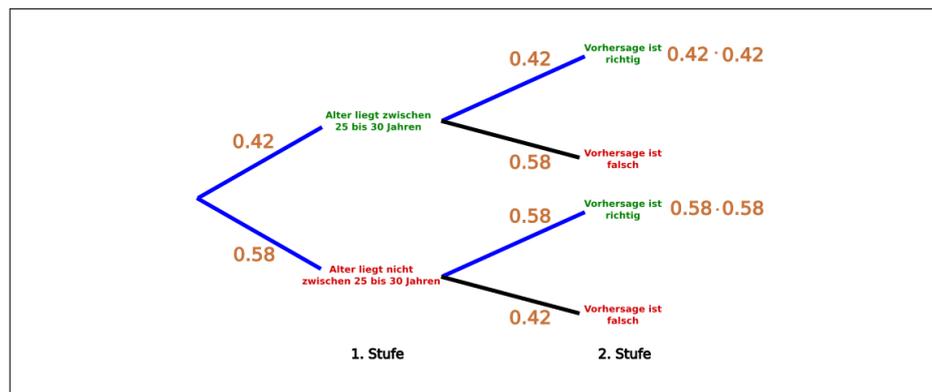
Hinweis: Die relative Häufigkeit wird als Wert zwischen 0 und 1 angegeben, welcher auf 2 Nachkommastellen gerundet wird.

Aufgabenteil b) Ergebnisse nach Anwendung der Zufallsregel:

Größe N der gewählten Stichprobe:	100	1000	10000	44747
Trefferquote (in %):	61 %	57 %	51 %	51 %

Hinweis zur Besprechung: Auch in der Besprechung sollten die Stichproben der Größe nach geordnet in der Tabelle notiert werden. Außerdem ist zu beachten, dass Matlab die Nutzer immer zufällig auswählt. Gerade bei kleinen Stichprobengrößen können so beim mehrmaligen Durchlaufen unterschiedliche Trefferquoten angezeigt werden.

Aufgabenteil c) Das zweistufige Baumdiagramm sieht wie folgt aus:



In blau sind die Pfade markiert, welche zu einer richtigen Vorhersage führen. Mit der Summenregel erhält man dann die gesuchte, theoretische Trefferquote:

$$\text{Theoretische Trefferquote: } 0,42^2 + 0,58^2 = 0,5128 \approx 0,51$$

Aufgabenteil d)

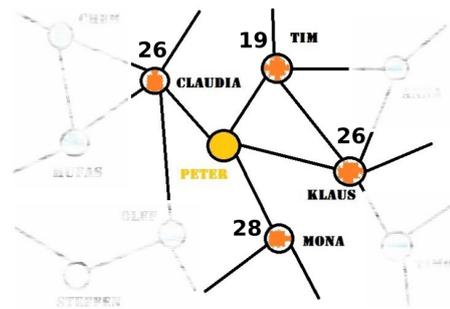
Vergleich der Trefferquoten aus Aufgabenteil b) mit der theoretischen Trefferquote aus Teil c):

Wird ein Zufallsexperiment immer unter den selben Bedingungen durchgeführt, so nähert sich die relative Häufigkeit immer weiter der erwarteten Wahrscheinlichkeit des Zufallsexperiments an. Dieses Phänomen wird als Gesetz der großen Zahlen beschrieben.

Eine Betrachtung der Trefferquoten in der Tabelle aus Aufgabenteil b) zeigt, dass sich die Werte mit wachsender Stichprobengröße N der theoretischen Trefferquote annähern. Bereits für $N = 10000$ schwankt die Trefferquote sehr nah um den erwarteten Wert. Das heißt, je größer die Anzahl an Nutzern gewählt wird, desto stärker nähert man sich der erwarteten, theoretischen Trefferquote an.

Zweite Idee | Entscheidungsregel mit Parametern

Aufgabenteil a)



Wie viel Prozent der Freunde von Peter sind zwischen 25 und 30 Jahre alt?:

75 %

⇒ **Antwort:** Da mehr als 40 % der Freunde von Peter zwischen 25 und 30 Jahren alt sind, wird Peter mit der Regel des 2. Mitarbeiters ebenfalls ein Alter zwischen 25 und 30 Jahren vorhergesagt.

Aufgabenteil b-c) Folgende Ergebnisse beziehen sich ebenfalls auf eine Stichprobengröße von 44747 Nutzern. Die angegebenen Parameter sind wieder nur als Beispiele zu betrachten.

Parameter:	[0,4 0,2]	[0,8 0,4]	[0,3 0,2]	[0,45 0,35]
Trefferquote einer richtigen Vorhersage innerhalb der gesuchten Altersspanne (in %):	85 %	69 %	87 %	83 %
Trefferquote einer richtigen Vorhersage außerhalb der gesuchten Altersspanne (in %):	76 %	86 %	73 %	80 %
Zusammenfassung:	•	•	•	•
Trefferquote für eine richtige Vorhersage (in %):	80 %	79 %	79 %	81 %
Was wird besser vorhergesagt? I oder \bar{I} :	I	\bar{I}	I	beide

Hinweis: Viele Parameterkombinationen führen zu Trefferquoten an die 80 %. Jedoch ist es für den Betreiber des sozialen Netzwerkes wichtig, dass besonders die Personen zwischen 25 und 30 Jahren gut vorhergesagt werden. Aus diesem Grund wäre beispielsweise von einer Wahl der Parameter [0,8 0,4] abzuraten.

1) Zusatzaufgabe | Altersüberprüfung

Die SuS erhalten die ersten 20 Nutzer IDs, die ein Alter von 0 Jahre aufweisen. Sie wählen 3 dieser angezeigten Nutzer aus und versuchen mit den bereits bekannten Vorhersageregeln 1 bis 10 das wahre Alter der Nutzer herauszufinden. Mit den SuS kann hier nochmal auf die herausgefundenen Trefferquoten der einzelnen Regeln eingegangen werden, um die Zuverlässigkeit des vorhergesagten Alters zu diskutieren.

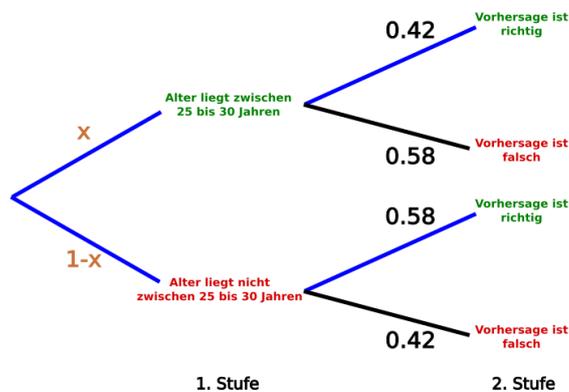
2) Zusatzaufgabe | Theoretische Trefferquote

Aufgabenteil a)

Für den betrachteten Datensatz aus den 44747 Nutzern liegt die relative Häufigkeit für die Nutzer zwischen 25 und 30 Jahre bei $p = 0.42$.

Aufgabenteil b)

Um die theoretische Trefferquote für eine richtige Vorhersage zu berechnen, wird das folgende zwei-stufige Baumdiagramm betrachtet. Dabei steht x für die von den SuS frei wählbare Wahrscheinlichkeit einem Nutzer das gesuchte Altersintervall $[25, 30]$ zuzuordnen. Die angegebenen Wahrscheinlichkeiten in der zweiten Stufe des Baumdiagramms sind für den jeweiligen Datensatz fest und beziehen sich hier wieder auf den Datensatz mit den 44747 Nutzern.

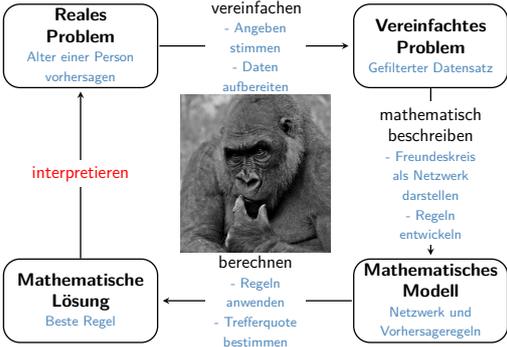
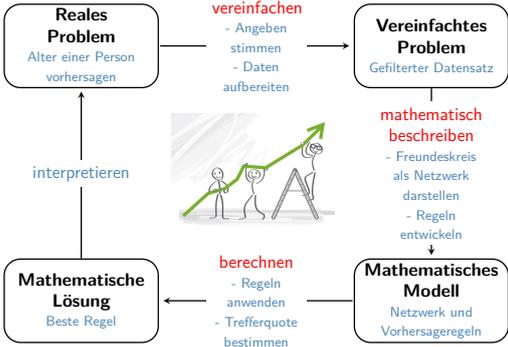


Es kann nun eine von x abhängige Funktion f aufgestellt werden, welche die theoretische Trefferquote bei vorgegebener Wahrscheinlichkeit x angibt:

$$\begin{aligned} f(x) &= (2 \cdot p - 1) \cdot x + 1 - p \\ &= (2 \cdot 0.42 - 1) \cdot x + 1 - 0.42 \\ &= -0.16 \cdot x + 0.58 \end{aligned}$$

I. Abschlussvortrag

I.1. Folien zum Abschlussvortrag

 <p>Sicherheit der Privatsphäre in sozialen Netzwerken - Was Facebook, Instagram und Co. über uns wissen</p> <p>Lehrstuhl für Mathematik Center for Computational Engineering Science</p> 	<p>Unsere heutige Frage:</p> <p>Wie gut kann man das Alter eines Nutzers in einem sozialen Netzwerk vorhersagen, obwohl dieses aus den Angaben in seinem Profil nicht ablesbar ist?</p> 
<p>Modellierungskreislauf</p>  <p>Reales Problem Alter einer Person vorhersagen</p> <p>vereinfachen - Angeben stimmen - Daten aufbereiten</p> <p>Vereinfachtes Problem Gefilterter Datensatz</p> <p>interpretieren</p> <p>Mathematische Lösung Beste Regel</p> <p>berechnen - Regeln anwenden - Trefferquote bestimmen</p> <p>Mathematisches Modell Netzwerk und Vorhersageregeln</p> <p>mathematisch beschreiben - Freundeskreis als Netzwerk darstellen - Regeln entwickeln</p>	<p>Modellierungskreislauf</p>  <p>Reales Problem Alter einer Person vorhersagen</p> <p>vereinfachen - Angeben stimmen - Daten aufbereiten</p> <p>Vereinfachtes Problem Gefilterter Datensatz</p> <p>interpretieren</p> <p>Mathematische Lösung Beste Regel</p> <p>berechnen - Regeln anwenden - Trefferquote bestimmen</p> <p>Mathematisches Modell Netzwerk und Vorhersageregeln</p> <p>mathematisch beschreiben - Freundeskreis als Netzwerk darstellen - Regeln entwickeln</p>
<p>Diskussion</p> <p>Ein letztes Mal die Gehirnzellen aktivieren!</p> 	<p>Diskussion Wer ist betroffen?</p> <p>Welche Personen sind vom Datensammeln und Vorhersagen persönlicher Informationen betroffen? Gibt es Personen die davor sicher sind?</p> 

Diskussion | Wer ist betroffen?

Welche Personen sind vom Datensammeln und Vorhersagen persönlicher Informationen betroffen? Gibt es Personen die davor sicher sind?



Diskussion | Wer ist betroffen?

Welche Personen sind vom Datensammeln und Vorhersagen persönlicher Informationen betroffen? Gibt es Personen die davor sicher sind?



Diskussion | Wer ist betroffen?

Welche Personen sind vom Datensammeln und Vorhersagen persönlicher Informationen betroffen? Gibt es Personen die davor sicher sind?



Diskussion | Was denkt ihr?

Wenn du nicht für einen Dienst bezahlst,...



... dann bist du nicht der Kunde, sondern das Produkt.

Diskussion | Vor- und Nachteile

Was sind die Vor- und Nachteile vom Sammeln und Benutzen persönlicher Daten?



Diskussion | Schutz der Daten

Wie kann man seine Daten und damit seine Privatsphäre schützen?



Diskussion | Konsequenzen

Welche Konsequenzen nehmt ihr aus dem heutigen Tag mit?



I.2. Notizen zum Abschlussvortrag

CAMMP day

Soziale Netzwerke Sek I

Hinweise zum Abschlussvortrag

Folie 1

Wir wollen nun nochmal die Ergebnisse des Tages sammeln und zusammen darüber diskutieren.

Folie 2 | Unsere heutige Frage

Mit dieser Fragestellung haben wir uns den gesamten Tag über beschäftigt.

- Die SuS werden gebeten, die Ergebnisse des Tages kurz zusammenzufassen.

Folie 3 | Modellierungskreislauf

Den SuS wird gesagt, dass der Modellierungskreislauf im Laufe des Tages mehrfach durchlaufen wurde:

- Gestartet sind wir mit der Vorhersage des exakten Alters.
- Dort haben wir jedoch erkannt, dass wir nicht zu befriedigenden Trefferquoten kommen.
- Wir haben deshalb das mathematische Modell angepasst und uns mit der Spannweite und dem Quartilsabstand Vorhersageregeln angeschaut, welche sich mit dem Altersintervall eines Nutzers beschäftigen. Ebenfalls wurde eine Regel zum Zufall und mit Parametern betrachtet.
- Nun müssen auch diese Ergebnisse wieder interpretiert werden.
- Falls von den SuS noch nicht erwähnt, wird nach der effektivsten Vorhersageregeln und deren Trefferquote gefragt.
- Falls von den SuS noch nicht erwähnt, wird gefragt, ob sie diese Trefferquote als hoch einschätzen.

Folie 4 | Modellierungskreislauf

Die SuS werden gefragt, welche weiteren Verbesserungen vorgenommen werden könnten, um noch eine höhere Trefferquote zu erzielen. Mögliche Beispiele wären:

- Beim Vereinfachen:
 - Die Vereinfachung „Alle Angaben sind nach der Filterung richtig“ anzweifeln.
- Beim mathematischen Modell:
 - Einen anderen (z.B. größeren) Datensatz benutzen.
 - Einen noch größeren Freundeskreis betrachten.
 - Weitere Angaben als das Alter der Freunde hinzunehmen (z.B. das Interesse der Nutzer).
 - Ein anderes Gütekriterium als die Trefferquote wählen.
- Für die mathematische Lösung:
 - Andere Vorhersageregeln (Vorschläge können gesammelt werden)

Folie 5 | Diskussion

Zum Schluss soll noch eine kleine Diskussion zu dem Thema, dass Betreiber von sozialen Netzwerken und auch andere Online-Giganten wie z.B. Google, unsere Daten sammeln und für ihre Zwecke (zum Geld verdienen) benutzen, geführt werden.

1/3

Folie 6 | Diskussion | Wer ist betroffen?

Die SuS sollen die Frage diskutieren, welche Personen von dem Sammeln von Daten und dem Vorhersagen von persönlichen Informationen betroffen sind. Auch wird die Frage gestellt, ob es ihrer Meinung nach Personen gibt, die davor sicher sind.

Nach der Diskussion können vom Betreuer noch folgende Punkte angesprochen werden:

- Soziale Netzwerke tauschen sich untereinander aus¹. Auch wenn ihr kein Facebook oder Instagram nutzt, so nutzt ihr vielleicht Whatsapp (Nach der JIM-Studie 2017 nutzten 94% der 12 bis 19-jährigen diese App mehrmals in der Woche².) So hat z.B. Facebook Instagram und auch WhatsApp gekauft. Für WhatsApp zahlte Facebook damals rund 20 Milliarden Dollar. Der teure Preis kam vor allem durch die 500 Millionen Nutzer zu stande, welche damals WhatsApp täglich nutzten und Facebook für Werbekunden noch attraktiver machte³.
- Aber auch von Personen, die kein Teil eines sozialen Netzwerkes wie WhatsApp und co sind, können Daten gesammelt und die bereits erwähnten Schattenprofile erstellt werden. Besitzt ein Freund z.B. WhatsApp und erlaubt der App auf das Kontaktbuch zugreifen zu dürfen, besitzt WhatsApp und damit auch Facebook die Namen, Telefonnummern und ggf. auch Emailadressen und den Wohnort von all seinen Kontakten.
Erinnerung:
Definition Schattenprofil:
 - für Nutzer des Netzwerks: Profil, das Informationen über euch enthält, die ihr nicht angegeben habt
 - (für Nicht-Nutzer: Profil mit Informationen wird erstellt.)
- Und selbst wenn man kein Handy besitzt, nutzen die meisten zu Hause Google zur Informationsbeschaffung (laut JIM-Studie 2017 85% der Jugendliche mehrmals die Woche⁴). Kaum jemand weiß so viel über die Vorlieben, Abneigungen oder Pläne der Menschen wie Google. Was Google mit den ganzen Suchanfragen macht, weiß keiner ganz genau. Aber eins ist sicher: Sucht man bei Google z.B. nach einem Auto, werden uns sofort die passenden Anzeigen vom freundlichen Autoverkäufer aus der Nachbarschaft angezeigt.
- Und shoppt man regelmäßig bei Amazon, werden die Daten ebenfalls fleißig gesammelt. Was man anklickt, was man kauft, was Kunden, die bestimmte Produkte gekauft haben, auch gekauft haben oder welche Interessen der Nutzer hat. Alles wird gesammelt und verwertet.
- Nun muss man sich aber auch die Fragen stellen, ob es vielleicht gerechtfertigt ist, dass diese Online-Giganten unsere persönlichen Daten benutzen. Schließlich zahlen wir keinen einzigen Cent für die ganzen Dienste, die uns quasi umsonst bereitgestellt werden. Dazu gibt es auch einen interessanten Satz, den man diesbezüglich immer mal wieder findet (nächste Folie öffnen).

Folie 7 | Diskussion | Was denkt ihr?

„Wenn du nicht für einen Dienst bezahlst, dann bist du nicht der Kunde, sondern das Produkt“ - Über diesen Satz sollen die SuS diskutieren.

- Den SuS kann vor Augen geführt werden, dass die Betreiber von sozialen Netzwerken viele Kosten aufbringen müssen, um ihre Dienste zur Verfügung zu stellen (z.B. Serverkosten, Mitarbeiterkosten, Kosten für die genutzten Räumlichkeiten, etc.) und jedes Unternehmen Gewinne erzielen möchte.

¹Quelle: H. Bleich (2016): Brüchiges Recht. Wie schwer es ist, Datenschutzverstöße von Facebook juristisch zu ahnden. In: c't 24/2016, S. 84

²Quelle: JIM Studie 2017, S. 35

³Quelle: <https://www.zeit.de/2016/37/whatsapp-facebook-daten-individualisierte-werbung-schutz>; Stand 26.06.2018

⁴Quelle: JIM-Studie 2017, S. 46

- Auch können die SuS gefragt werden, ob sie bereit wären für die Dienste Geld auszugeben. Z.B. gibt es für WhatsApp die Alternative Threema (Nutzerzahlen Stand 2018: 4,5 Millionen⁵), welches von anderen sozialen Netzwerken unabhängig ist, ohne Werbung funktioniert (WhatsApp bekommt ab 2019 Werbung) und die Nachrichten Ende-zu-Ende verschlüsselt verschickt. Die Kosten für diese App liegen bei 3 € (einmalig).

Folie 8 | Diskussion | Vor- und Nachteile

Die SuS sollen über die Vor- und Nachteile vom Sammeln und Benutzen persönlicher Daten diskutieren.

Als Anregung können folgende Beispiele genannt werden:

- Beispiele für Vorteile:
 - Unterstützung bei Ermittlungen der Polizei
 - Vorbeugen von Terroranschlägen
 - Wenn schon Werbung, dann auch Werbung die mich interessiert.
- Beispiele für Nachteile:
 - Die Firma Cambridge Analytica hat nach eigenen Angaben Trump bei der Wahl zum Präsidenten maßgeblich unterstützt. Durch personalisierte Wahlversprechen wurden die Menschen beeinflusst.
 - Fremde Personen wissen über unsere Interessen und Vorlieben Bescheid (der gläserner Mensch)
 - Fake-News und Manipulation
 - Filterblase bzw. Echokammer: Durch Algorithmen erhalten wir von Betreibern sozialer Netzwerke bevorzugt Nachrichten, die uns vermeintlich zusagen. Es besteht die Gefahr ein einseitiges Bild zu erhalten. Konträre Meinungen oder gesellschaftliche Widersprüche tauchen dort nur selten auf ⁶.

Folie 9 | Diskussion | Schutz der Daten

„Wie kann man seine Daten und damit seine Privatsphäre schützen?“ - Über diese Frage sollen die SuS diskutieren.

Hier ein paar Beispiele ⁷:

- Mehrere E-Mail-Adressen nutzen, um eine Zusammenführung der Daten zu erschweren.
- Mehrere Suchmaschinen über Google hinaus benutzen.
- Grundsätzlich überlegen, was man online mitteilen möchte. Was man im Internet einträgt, postet oder vermeintlich privat kommentiert, kann relativ schnell absichtlich oder unbeabsichtigt in der Öffentlichkeit landen. Und hat man erst einmal was gepostet, kann man es nur schwer oder gar nicht zurücknehmen.
- Lesen auf was die ganzen installierten Apps zugreifen und ggf. Zugriffsberechtigungen deaktivieren.
- Ein weiteres Beispiel ist der Live-Standort von WhatsApp. Wenn man diese Funktion einmal benutzt hat, erlaubt man WhatsApp jederzeit auf den aktuellen Standort zuzugreifen. Auch diese Erlaubnis kann nach der Benutzung wieder deaktiviert werden.

Folie 10 | Diskussion | Konsequenzen

Die SuS sollen mitteilen, welche Konsequenzen sie bezogen auf den Umgang mit sozialen Netzwerken aus dem heutigen Tag mitnehmen.

⁵Quelle: <https://de.statista.com/statistik/daten/studie/445619/umfrage/nutzer-des-schweizer-messaging-dienstes-threema/>; Stand 27.06.2018

⁶Quelle: <https://www.boell.de/de/2017/02/08/filter-bubble-echokammer-fake-news>; Stand 03.08.2018

⁷Quelle: <https://www.sz-online.de/nachrichten/mich-kriegt-ihr-nicht-2518740.html>; Stand 26.06.2018

J. Evaluation

J.1. Evaluationsbogen



CAMMP Day Evaluation SuS

Seite 1

Es besteht immer die Möglichkeit unsere Programme zu verbessern und wir würden gerne deine Meinung erfahren. Vielen Dank für deine Rückmeldung.

1. Persönliche Angaben

8 9 10/EF 11/Q1 12/Q2 13

Bitte gib dein Geschlecht an:

weiblich männlich

Seite 2

2. Bewertung des Workshops

	Trifft gar nicht zu (-)	Trifft eher nicht zu (-)	Trifft zum Teil zu (+)	Trifft voll zu (++)
Der CAMMP day hat mir Spaß gemacht.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Das Thema Sicherheit der Privatsphäre in sozialen Netzwerken war interessant.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich wusste bereits vor dem CAMMP day, dass Mathematik in meinem Alltag eine wichtige Rolle spielt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Ich habe verstanden, wie mit einfachen mathematischen Regeln Vorhersagen über das Alter getroffen werden können.

Ich wusste bereits, dass Informationen vorhergesagt werden können.

Der CAMMP day hat mich auf das Thema Sicherheit der Privatsphäre in sozialen Netzwerken aufmerksam gemacht.

Ich werde in Zukunft mehr über das Thema Sicherheit in sozialen Netzwerken nachdenken.

Ich habe auf Blatt 1 verstanden, warum die Daten gefiltert werden müssen.

Durch den Workshop habe ich mathematisches Modellieren besser verstanden.

Durch den CAMMP day habe ich erkannt, dass Mathematik eine wichtige Rolle in der Gesellschaft hat

Die Einführung in MATLAB war hilfreich.

Der Umgang mit MATLAB fiel mir schwer.

Die Aufgaben waren zu

einfach.

Die Aufgaben
waren zu
schwierig.

Die Hilfekarten
waren hilfreich.

Während des
CAMMP days
gab es Phasen,
in denen ich
nichts zu tun
hatte.

Folgende mathematische Begriffe kannte ich bereits vor dem CAMMP day:

Arithmetisches Mittel

Median

Modus

Spannweite

Quartilsabstand

Nach dem CAMMP day hat sich mein Umgang mit den folgenden mathematischen Begriffen verbessert:

Arithmetisches Mittel

Median

Modus

Spannweite

Quartilsabstand

3. Weiterführende Fragen

Wenn du über deine Erfahrungen mit diesem Kurs nachdenkst, wie sehr treffen die folgenden Aussagen auf dich zu?

	Trifft gar nicht zu (--)	Trifft eher nicht zu (-)	Trifft zum Teil zu (+)	Trifft voll zu (++)
Die Lern- und Arbeitsatmosphäre war angenehm.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Inhalte wurden klar vermittelt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Lern- und Arbeitszeiten waren angemessen (nicht zu lang oder kurz mit ausreichenden Pausen).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Der Kurs hat mein Interesse an Themen der Naturwissenschaften und Technik gesteigert.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Durch den Kurs habe ich interessante Berufs- und Studienmöglichkeiten kennen gelernt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich habe in diesem Kurs viel Neues gelernt, was mir für die Schule, für ein Studium, für einen Beruf weiterhelfen kann.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich kann mir vorstellen ein Studium oder eine Ausbildung im Bereich der Naturwissenschaften oder der Technik zu beginnen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich würde so einen Kurs gerne noch einmal besuchen und würde ihn auch anderen weiterempfehlen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Das Thema des Workshops war für mich neu.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Alles in allem hat mir die Veranstaltung gut gefallen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Ist dir etwas besonders leicht gefallen?

Nein.

Ja, und zwar

Ist dir etwas besonders schwer gefallen?

Nein.

Ja, und zwar

Hättest du dir stellenweise noch mehr Unterstützung gewünscht?

Nein.

Ja, und zwar bei

Hat dir etwas an dem Kurs absolut nicht gefallen?

Nein

Ja, und zwar:

Hat dir etwas an diesem Kurs besonders gut gefallen?

Nein

Ja, und zwar:

Hättest du gerne noch etwas anderes gesehen oder erfahren?

Nein

Ja, und zwar:

Seite 3

4. Lernzuwachs

Was hast du für dich persönlich durch die Teilnahme am Workshop gelernt?

5. Abschließende Bewertung:

Ich gebe dem CAMMP day die Schulnote:

- | | | |
|------------------------------------|--|--------------------------------------|
| <input type="radio"/> 1 (sehr gut) | <input type="radio"/> 3 (befriedigend) | <input type="radio"/> 5 (mangelhaft) |
| <input type="radio"/> 2 (gut) | <input type="radio"/> 4 (ausreichend) | <input type="radio"/> 6 (ungenügend) |

Ich gebe den Betreuern die Schulnote:

- | | | |
|------------------------------------|--|--------------------------------------|
| <input type="radio"/> 1 (sehr gut) | <input type="radio"/> 3 (befriedigend) | <input type="radio"/> 5 (mangelhaft) |
| <input type="radio"/> 2 (gut) | <input type="radio"/> 4 (ausreichend) | <input type="radio"/> 6 (ungenügend) |

Abschließender persönlicher Kommentar (z.B. Lob, Kritik, Verbesserungsvorschläge):

» [Umleitung auf Schlussseite von Umfrage Online](#)

J.2. Ergebnisse der Evaluation

CAMMP Day Evaluation SuS

1. 1. Persönliche Angaben

Anzahl Teilnehmer: 22

21 (95.5%): 8

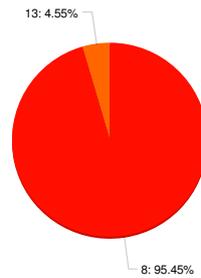
- (0.0%): 9

- (0.0%): 10/EF

- (0.0%): 11/Q1

- (0.0%): 12/Q2

1 (4.5%): 13

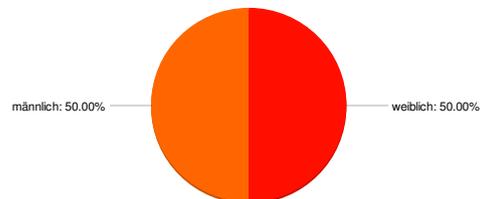


2. Bitte gib dein Geschlecht an:

Anzahl Teilnehmer: 22

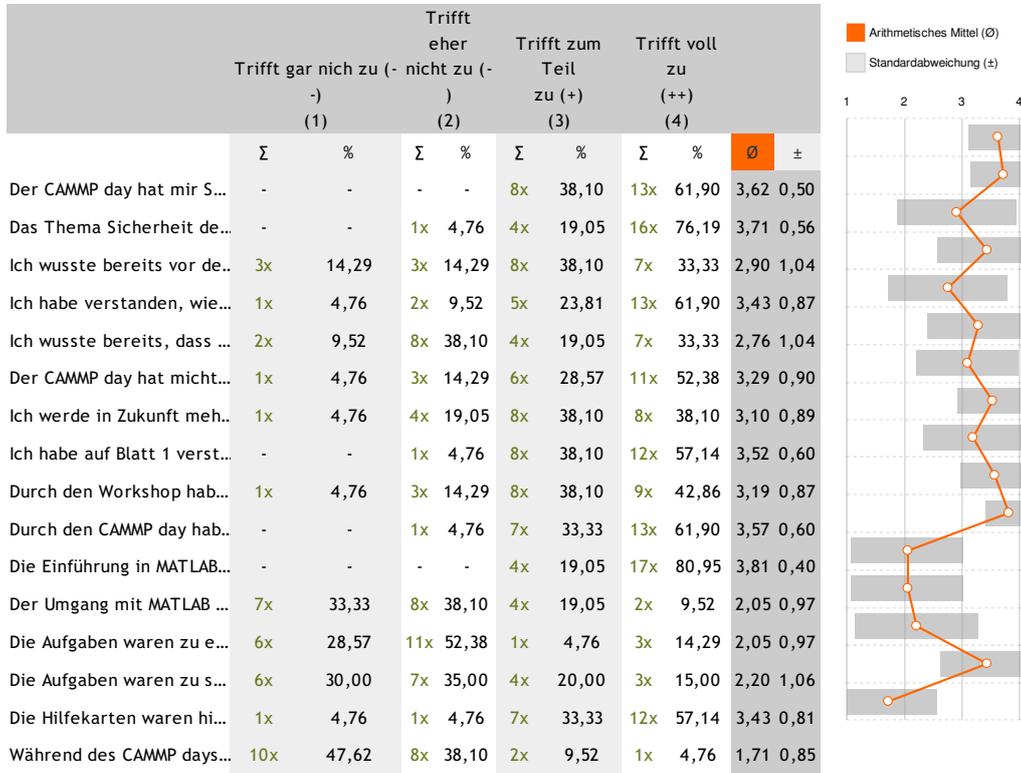
11 (50.0%): weiblich

11 (50.0%): männlich



3. 2. Bewertung des Workshops

Anzahl Teilnehmer: 21



4. Folgende mathematische Begriffe kannte ich bereits vor dem CAMMP day:

Anzahl Teilnehmer: 19

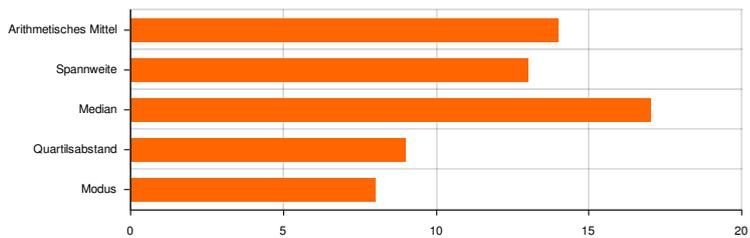
14 (73.7%): **Arithmetisches Mittel**

13 (68.4%): **Spannweite**

17 (89.5%): **Median**

9 (47.4%): **Quartilsabstand**

8 (42.1%): **Modus**



5. Nach dem CAMMP day hat sich mein Umgang mit den folgenden mathematischen Begriffen verbessert:

Anzahl Teilnehmer: 19

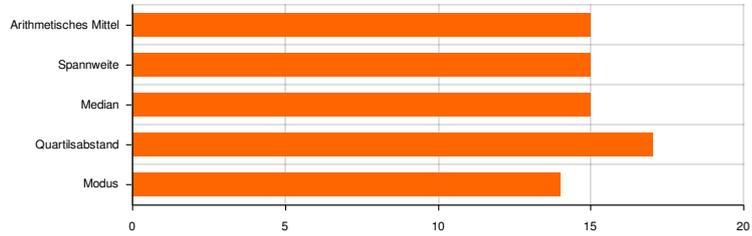
15 (78.9%): **Arithmetisches Mittel**

15 (78.9%): **Spannweite**

15 (78.9%): **Median**

17 (89.5%): **Quartilsabstand**

14 (73.7%): **Modus**



6. 3. Weiterführende Fragen

Wenn du über deine Erfahrungen mit diesem Kurs nachdenkst, wie sehr treffen die folgenden Aussagen auf dich zu?

Anzahl Teilnehmer: 21



7. Ist dir etwas besonders leicht gefallen?

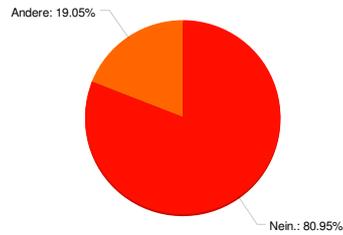
Anzahl Teilnehmer: 21

17 (81.0%): Nein.

4 (19.0%): Andere

Antwort(en) aus dem
Zusatzfeld:

- alles
- Den Median zu berchnen
- boxspot
- das filtern



8. Ist dir etwas besonders schwer gefallen?

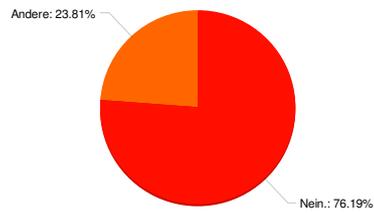
Anzahl Teilnehmer: 21

16 (76.2%): Nein.

5 (23.8%): Andere

Antwort(en) aus dem
Zusatzfeld:

- Die luft
- Ein paar aufgaben
- Das ausrechnen der
Treffquote
- ALLES
- Einige aufgaben

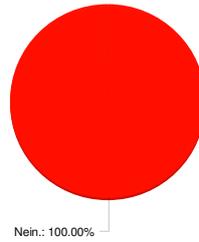


9. Hättest du dir stellenweise noch mehr Unterstützung gewünscht?

Anzahl Teilnehmer: 21

21 (100.0%): Nein.

- (0.0%): Andere



10. Hat dir etwas an dem Kurs absolut nicht gefallen?

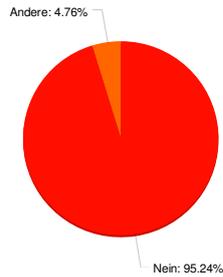
Anzahl Teilnehmer: 21

20 (95.2%): Nein

1 (4.8%): Andere

Antwort(en) aus dem
Zusatzfeld:

- Zu wenige pausen



11. Hat dir etwas an diesem Kurs besonders gut gefallen?

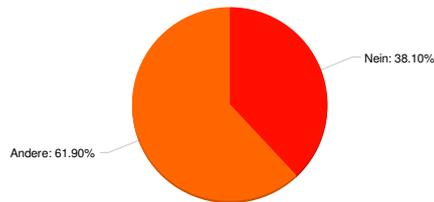
Anzahl Teilnehmer: 21

8 (38.1%): Nein

13 (61.9%): Andere

Antwort(en) aus dem
Zusatzfeld:

- Dass immer Hilfe da war wenn man sie brauchte
- Die Hilfe bei nicht verstandenen Aufgaben
- alles
- Thema
- alles
- Aufgaben waren nicht zu schwer
- Das wir mit pcs arbeiten
- Alles
- Das boxplot
- die freundlichkeit
- alles
- Die Hilfskarten haben einem sehr geholfen die Aufgabe zu lösen
- Das Thema war sehr aktuell und betrifft alle



12. Hättest du gerne noch etwas anderes gesehen oder erfahren?

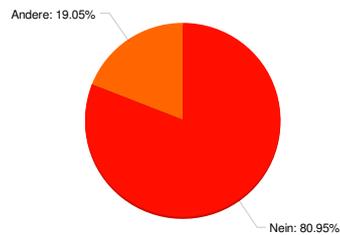
Anzahl Teilnehmer: 21

17 (81.0%): Nein

4 (19.0%): Andere

Antwort(en) aus dem
Zusatzfeld:

- Soziale Netzwerke
- Das wir mit pcs arbeiten
- Wie es besser geht und woran die meine standorte und E-mail so wie freunde herausfinden
- Ist Meerrettich ein Instrument



13. 4. Lernzuwachs

Was hast du für dich persönlich durch die Teilnahme am Workshop gelernt?

Anzahl Teilnehmer: 12

- Das Vorhersagen des Alters einer Person.
Ebenfalls mehr zum Thema Datenschutz.
- das man im internet vorsichtig sein muss
- Modellieren von realen Problemen
- Daten sind gefaerlich
- Zu viel um alles aufzuzählen
- Vieles
- Das ich vorsichtiger mit Medien sein muss
- wie wivhtig mathe ist
- gar nix
- daten besser schuetzen
- Das unsere Daten nicht sicher sind und man leicht Informationen vorhersagen kann
- Das man Informationen von Nutzern vorhersagen kann

14. 5. Abschließende Bewertung:

Ich gebe dem CAMMP day die Schulnote:

Anzahl Teilnehmer: 21

12 (57.1%): 1 (sehr gut)

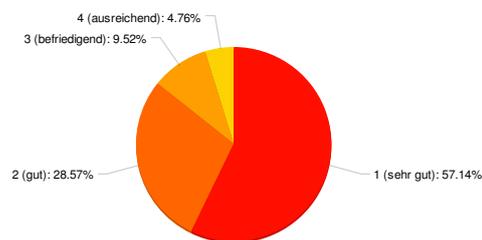
6 (28.6%): 2 (gut)

2 (9.5%): 3 (befriedigend)

1 (4.8%): 4 (ausreichend)

- (0.0%): 5 (mangelhaft)

- (0.0%): 6 (ungenügend)



15. Ich gebe den Betreuern die Schulnote:

Anzahl Teilnehmer: 21

18 (85.7%): 1 (sehr gut)

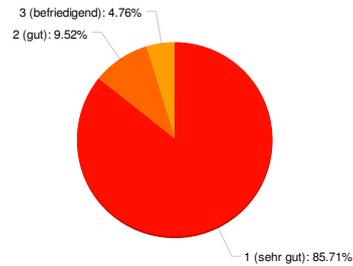
2 (9.5%): 2 (gut)

1 (4.8%): 3 (befriedigend)

- (0.0%): 4 (ausreichend)

- (0.0%): 5 (mangelhaft)

- (0.0%): 6 (ungenügend)



16. Abschließender persönlicher Kommentar (z.B. Lob, Kritik, Verbesserungsvorschläge):

Anzahl Teilnehmer: 12

- Die Aufgabenstellung war sehr gut und gut zu verstehen.
Wenn ich mal nicht weiterwusste, habt ihr mir geholfen und ich habe es sofort verstanden.
- nils und tobias sind nett :-)
- Bessere Luft und mehr Spaß
xD
- ;-)
- War sehr nice
- Ihr habt euch viel mpge gegeben
- Ich fand die Hilfe sehr Professionel und alles war in einer guten Stimmung.
- geeeeeiiiiilll
- Mir hat es Spass gemacht
- ich fands gut
- Der Tag hat mir Spaß gemacht.
- Das Material war sehr schön gestalltet.