

Entwicklung und Reflexion einer Unterrichtssequenz zum Maschinellen Lernen als Aspekt von Data Science in der Sekundarstufe II

Abstract: Die Bereiche „Data Science“ und „Big Data“ sowie ihre technischen, ethischen und gesellschaftlichen Auswirkungen werden zunehmend nicht nur in der Wissenschaft, sondern auch in diversen Medien diskutiert und somit verstärkt auch zu einem wichtigen Thema für alle. Um den Schülerinnen und Schülern der Sekundarstufe II einen theoretisch und fachwissenschaftlich fundierten Einstieg in diesen Themenbereich zu ermöglichen, wurde ein erster Entwurf eines interdisziplinären Curriculums entwickelt, das neben fachlichen Aspekten von Data Science einen Fokus auf gesellschaftliche Fragestellungen, die sich in diesem Zusammenhang ergeben, legt. Im Folgenden werden neben der Konzeption des Kurses die bisherigen Erfahrungen aus der Durchführung in einem Oberstufenkurs- insbesondere in Hinsicht der darin enthaltenen Unterrichtseinheit zum Maschinellen Lernen - berichtet, sowie die sich hieraus ergebenden Implikationen für die Weiterentwicklung diskutiert und dargestellt.

Keywords: Data Science, Maschinelles Lernen, KI, Künstliche Intelligenz, Künstliche Neuronale Netze, Entscheidungsbäume, Big Data, Curriculum

1 Einleitung

Data Science sowie Fragen zu Maschinellern Lernen (ML) und Künstlicher Intelligenz (KI) werden im geschäftlichen Bereich intensiv diskutiert und dort auch in großem Maß in verschiedene Systeme implementiert. So nehmen sie nicht nur dort, sondern auch in der gesellschaftlichen Diskussion über die Funktion, den Nutzen und die Gefahren dieser Systeme viel Raum ein. Es erscheint uns wichtig, dass all diese Fragestellungen in den Unterricht integriert werden. Daher entwickelten wir auf Basis curricularer Ideen aus der Informatik und Mathematik einen ersten Entwurf eines Curriculums für die Sekundarstufe II und führten dies in ein konkretes Unterrichtskonzept über. Im Rahmen dieses Artikels stellen wir neben diesem Kurs auch erste Erfahrungen der Umsetzung – insbesondere der Einheiten aus KI und ML – sowie die hinter dem Entwicklungsprozess liegenden Ideen vor.

2 Data Science und Maschinelles Lernen – Aspekte für die Bildung

Data Science, ML und der Umgang mit Big Data ist ein sehr umfassendes Feld, das über technische und wissenschaftliche Aspekte unterschiedlicher Disziplinen weit hinaus geht

und auch ethische, gesellschaftliche und soziale Auswirkungen in sich trägt – und daher nicht umsonst immer häufiger auch in sehr verschiedenen Kontexten diskutiert wird. Um möglichst viele Aspekte zu verstehen und ein gemeinsames Verständnis von *Data Science* zu entwickeln, wurden im Rahmen eines interdisziplinären, internationalen Symposiums diese Aspekte diskutiert und zwei für uns curricular relevante Bereiche wurden identifiziert: Für die *Informatikdidaktik* sind das die Entwicklung von Computational Thinking [TD16], verschiedene Bereiche der Mensch-Maschine-Interaktion und ihre Auswirkungen auf die Gesellschaft – und damit auch unser Umgang mit Big Data sowie den Methoden und Auswirkungen von Data Science [SBS18]. Im Bereich der *Statistikdidaktik* sollten die fundamentalen Ideen der Statistik um verschiedene Aspekte der *statistischen Kompetenz* [Ri16] erweitert werden.

2.1 Daten und Datenprozesse als strukturgebende Komponenten

Im Gegensatz zur praktischen Informatik stehen im Bereich der *Data Science* nicht Algorithmen und Abläufe sowie deren Modellierung und Implementierung im Vordergrund, sondern der Umgang mit *Daten*. Um jedoch eine gemeinsame Sicht auf Daten zu entwickeln, muss zunächst ein gemeinsames begriffliches Verständnis entwickelt werden. Nach der Definition der Empfehlungen GI zu den Bildungsstandards in der Sekundarstufe sind „Daten eine Darstellung von Information in formalisierter Art [...]. Daten werden wieder zu Information, wenn sie in einem Bedeutungskontext interpretiert werden“ ([Rö16], S. 9). Somit kann ein Informatiksystem nur Daten verarbeiten und darstellen, die darin enthaltene Information werden durch Interpretation durch den Menschen gewonnen, Daten sind also „nicht nur Zahlen, sie sind Zahlen mit einem Kontext“ [CM97]. Es stellt sich die Frage, wie Schülerinnen und Schüler *Datenkompetenz* erwerben können, die Voraussetzung für einen kompetenten Umgang mit fehlerbehafteten oder unterschiedlich strukturierten Daten innerhalb eines Kontextes ist. Ridsdale et al. definieren *Datenkompetenz* als die prozessorientierte „Fähigkeit, Daten kritisch zu sammeln, zu analysieren, zu bewerten und anzuwenden“ ([Ri15], S. 3). Damit liegt der Schwerpunkt der Handlungen der Lernenden im Umgang mit verschiedenen Daten, so dass „Prozessmodelle zur Datenanalyse“ einen sinnvollen Ausgangspunkt zur Planung von unterrichtlichem Handeln bieten. Der Definition von Ridsdale et al. ähnelt das CRISP-DM-Modell³ [Be10], das ein vollständiges Verfahren zum Umgang mit Daten beschreibt (vgl. Abb. 1).

Dieses Prozessmodell erscheint uns einen sinnvollen Rahmen zur Erarbeitung eines Curriculums und entsprechenden Unterrichtsmaterials zu bieten, so dass entschieden wurde, das CRISP-DM-Modell als Basis weiterer Entwicklungen zu verwenden.

³ CRISP-DM = **C**ross-**I**ndustry Standard Process for **D**ata **M**ining, Phasen: Verstehen der Fragestellung – Verstehen der Daten – Aufbereitung der Daten – Modellbildung – Bewertung des Modells – Einsatz

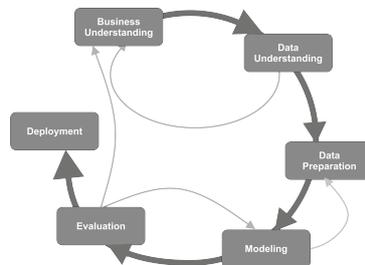


Abb. 1: Das CRISP-DM-Modell als Standardprozessmodell für Data Mining

2.2 Relevanz sozialer und gesellschaftlicher Aspekte

Eine ebenfalls große Relevanz im Bereich *Data Science* besitzen die hieraus folgenden gesellschaftlichen und sozialen Implikationen. Diskussionen über gesellschaftliche Aspekte im Informatikunterricht werden – wenn überhaupt – nur entkoppelt von technologischen Fragen der Unterrichtsinhalte geführt und sind nicht fest in die Arbeit in Softwareprojekten und Lernaufgaben integriert. Auch eines der wenigen Data Science Curricula für das schulische Umfeld [GR17] stellt die verschiedenen Aspekte von Data Science zwar umfassend, aber nur aus fachwissenschaftlicher Sicht dar. Dies stützt die Annahme, dass Informatiklehrkräfte gesellschaftliche Fragen wohl als wenig wichtig ansehen. Daher stellt die Frage, wie diese Aspekte integraler Bestandteil des Unterrichtsmaterials und des Curriculums werden können, eine wichtige Herausforderung dar.

2.3 Auf dem Weg zum Data Science Kurs - Didaktische Ansätze

Eine wichtige Erkenntnis der Vorarbeiten war die Notwendigkeit eines gemeinsamen Grundverständnisses aller Begriffe und Verfahren (vgl. [Tv09]) – ein Schritt, der auch während der ersten Durchführung des Projektkurses noch nicht abgeschlossen ist. Weiterhin wurde evaluiert, ob weitere schon existierende Materialien oder Curricula verwendet werden können, allerdings zeigte sich, dass sämtliche, meist hochschulische Curricula gut strukturiert sind, aber nur auf technologische Aspekte fokussieren. Daher ergibt sich die Notwendigkeit, diese Materialien selbst zu entwickeln und die für die Umsetzung relevanten Inhalte aus den Beiträgen des in Kapitel 1 erwähnten Symposiums zu generieren. Das Auffinden relevanter Information aus Data Mining und Statistik auf der einen Seite, sowie die notwendigen Kompetenzen aus dem Bereich des ML sowie dem Design Künstlicher Neuronaler Netzwerke (KNN) auf der anderen Seite, ist dank der breiten Basis aller analysierten Unterlagen relativ einfach. Schwieriger gestaltet sich das Einbinden der gesellschaftlichen, sozialen und interdisziplinären Fragestellungen (vgl. Abschnitt 2.2). Daher entschieden wir uns, in diesem Projekt den Ansatz des „*Design-Based Research*“ (DBR) [Co03] zu verfolgen: Ausgehend von einem ersten Entwurf eines Data Science Curriculums wird ein entsprechender Kurs entwickelt, durchgeführt und evaluiert. Aus den Erkenntnissen dieser ersten Durchführung

wird der Kurs und damit auch das Curriculum in mehreren Zyklen weiterentwickelt. Die im folgenden beschriebene Version des Data Science Kurses wurde unter diesem Ansatz entwickelt und wird als sog. „Projektkurs“ in Kooperation mit einem Gymnasium vor Ort während des SJ 2018/19 erprobt und evaluiert.

3 Der Data Science Kurs

Der so entstandene Kurs ist modular angelegt, wobei die beiden Bereiche „Data Mining und Statistik“ sowie „Künstliche Intelligenz und Maschinelles Lernen“ klar abgegrenzt sind und dem CRISP-DM-Modell (vgl. Abschnitt 2.1) folgend aufeinander aufbauend gestaltet werden. Zum Erwerben von Kompetenz, zumindest einfache Data Science Projekte selbst durchzuführen, wird als drittes Modul ein Projektmodul entwickelt und durchgeführt, so dass der Kurs in seiner ersten Version aus drei Modulen besteht:

1. *Von Daten zu Informationen:* Dieses Modul ist eine Einführung in Data Science und den Umgang mit Big Data und zielt darauf ab, das statistische Denken zu verbessern und Datenkompetenz zu entwickeln, um so statistische Methoden auf Daten zur Informationsgewinnung anwenden zu können.

2. *Künstliche Intelligenz und Maschinelles Lernen:* Im Rahmen dieses Moduls lernen die Schülerinnen und Schüler zwei unterschiedliche Methoden kennen, Erkenntnisse aus Daten zu gewinnen, indem sie exemplarisch Entscheidungsbäume als Vertreter einer Symbolischen KI und KNN (hier Back-Propagation-Netze) als typischen Vertreter überwachter Lernens kennenlernen, analysieren und auf eigene Beispiele anwenden.

3. *Datenprojekte:* Im Rahmen der Durchführung von Datenprojekten können die Schülerinnen und Schüler ihr bisheriges Wissen einsetzen, um reale Fragestellungen zu bearbeiten. Dabei werden sie motiviert, ihr Vorgehen im Sinne des CRISP-DM-Modells zu planen, um so zu für sie optimalen Ergebnissen kommen zu können und diese auch am Ende zu präsentieren und deren gesellschaftlichen Implikationen diskutieren zu können.

Es wurde auf eine enge Verflechtung der informatischen und mathematischen Inhalte mit gesellschaftlichen und sozialen Aspekten geachtet, um eine mehrdimensionale und interdisziplinäre Sicht auf alle Aspekte der Themen zu erhalten.

3.1 Modul 1: Von Daten zu Informationen – Datendetektive

Folgend dem in Abschnitt 2.1 beschriebenen CRISP-DM-Modell ist für ein Vorhaben aus dem Bereich Data Science zunächst wichtig, die Fragestellung zu verstehen („Business Understanding“), entweder sinnvolle Daten zu erheben oder schon gesammelte Daten zu verstehen, („Data Understanding“) und sie zu analysieren und aufzubereiten („Data Preparation“). Inspiriert von diesem Zyklus wird in dem Modul „Datendetektive“ im

ersten Baustein zunächst die Verwendung von großen und offen verfügbaren Datenmengen diskutiert und anschließend an Hand von „Lärmdaten“ erprobt, derartige Daten selbst zu analysieren, unter Verwendung von Jupyter Notebooks⁴ aufzubereiten und zu visualisieren. Im zweiten Baustein führen die Schülerinnen und Schüler eine explorative Datenanalyse unter Verwendung des multivariaten JIM-Datensatzes mit Hilfe des Online-Tools CODAP⁵

3.2 Modul 2: Künstliche Intelligenz und Maschinelles Lernen

Betrachtet man den CRISP-DM-Zyklus weiter, so folgt als nächster Schritt das Bilden eines Modells („Modeling“). Hier erkennen die Lernenden die Unterschiede zwischen klassischen algorithmischen Problemlösungsverfahren und datengetriebenen Prozessabläufen am Beispiel des ML (vgl. Abb. 2) und können so auch die damit verbundene Rolle des Menschen innerhalb dieser Mensch-Maschine-Interaktion diskutieren und reflektieren. Da der Bereich des ML zu groß und komplex ist, um umfassend mit Schülerinnen und Schülern bearbeitet zu werden, wurden im Vorfeld wichtige Verfahren exemplarisch ausgewählt und in zwei Bausteinen mit den Schülerinnen und Schülern bearbeitet.

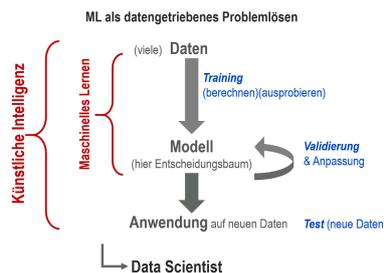


Abb. 2: Mit den Schülerinnen und Schülern erarbeitete Darstellung zum Vorgehen zur Erstellung von KI-Modellen mittels Maschinellern (ML)

Im ersten Baustein erwerben die Schülerinnen und Schüler unter Verwendung des „Sweet Learning Computers“ [Cu16b], einer Unplugged-Aktivität, ein grundsätzliches Verständnis über ML und diskutieren auf Basis dieses so erarbeiteten Wissens über aktuelle und zukünftige Chancen und Risiken dieser Technologien sowie ihren vielfältigen Einsatz. Dies wird vertieft durch den Einsatz von *Entscheidungsbäumen*, die in relativ kurzer Zeit für die Lernenden verständlich sind und auch die wesentlichen Verfahren zumindest im Grundsatz für die Lernenden transparent und erkennbar sind. Als Werkzeug wird hier CODAP mit einem zusätzlichen Plug-In zum Darstellen von Entscheidungsbäumen verwendet. Allerdings stößt dieses Tool sehr bald an seine Grenzen. Insbesondere durch das manuelle Aufbauen des Entscheidungsbaumes kann kein tieferes Verständnis darüber erworben werden, wie Entscheidungsbäume algorithmisch erzeugt und zum automatischen

⁴ <https://jupyter.org/>

⁵ <https://codap.concord.org>

Klassifizieren eingesetzt werden können. Daher werden im Anschluss mit Jupyter Notebook eigene Bäume erzeugt, so dass die Möglichkeit besteht, die Passung des entwickelten Baumes durch die Hinzunahme von Validierungsdaten zu überprüfen und anschließend zu optimieren. Im darauf folgenden zweiten Baustein werden zunächst in Analogie zur Funktionsweise des Gehirns durch die Unplugged-Aktivität „Brain in a Bag“ [Cu16a] die Grundbegriffe von KNN zusammen mit den Schülerinnen und Schülern erarbeitet. Durch dieses Spiel, das durch weitere Erläuterungen und theoretische Inhalte ergänzt wird, erkennen die Schülerinnen und Schüler so die grundlegenden Eigenschaften und Parameter von KNN ohne die Hürde eigener Programmierung. Um diese Erkenntnisse praktisch zu erproben und ein Gefühl für den Einfluss der verschiedenen Parameter zu erhalten, trainieren sie unter Verwendung des Online-Tools „Playground Tensorflow“⁶ Netze für verschiedene Aufgaben und beobachten dabei direkt grafisch aufbereitet die Auswirkungen ihrer Änderungen. Für ein tieferes Verständnis modellieren, trainieren und validieren sie selbst mit von Jupyter Notebook eigene KNN unter Verwendung der handschriftlichen Ziffern aus der MNIST-Datenbank [LCB98]. Da die Ziffern dieser Datenbank der amerikanischen Schreibweise entsprechen, führt eine Validierung durch handgeschriebene Zahlen der Schülerinnen und Schüler zu schlechten Ergebnissen, so dass hier ein Anlass geschaffen wird, die Grenzen und Möglichkeiten von verschiedenen Ansätzen von ML zu diskutieren.

3.3 Modul 3: Datenprojekte

Während in den ersten beiden Modulen die Erarbeitung neuen Wissens im Mittelpunkt stand, wird im dritten Modul ein gemeinsames Datenprojekt mit „realen“ Daten durchgeführt und dabei werden die in den vorherigen Modulen erworbenen Kompetenzen vertieft. Zur Organisation und zur Unterstützung der arbeitsteiligen Arbeit an den Daten steht den Schülerinnen und Schülern ein Gitlab zur Verfügung, das sowohl zur Daten- als auch zur Aufgabenorganisation genutzt werden kann. Die betreuenden Personen fungieren im Rahmen dieses Projekts als Lernbegleiter und kümmern sich um den Kontakt mit den Projektpartnern sowie die interne Kommunikation und Organisation.

Zum Zeitpunkt des Verfassens dieses Artikels befindet sich das Datenprojekt in der ersten Hälfte der Durchführung, so dass hier noch keine Erfahrungen der Durchführung vorliegen.

4 Erfahrungen der ersten Kursdurchführung

Der Projektkurs findet dreistündig mit 2 Schülerinnen und 17 Schülern der Jgst. 12, die alle Informatik belegt haben und daher über grundlegende Kenntnisse von Java verfügen, in einer Laborumgebung statt. Die in diesem Kurs verwendete Programmiersprache Python sowie Jupyter Notebook waren für sie neu – der Einstieg war für sie aber mit etwas Unterstützung gut zu bewältigen.

⁶ <https://playground.tensorflow.org>

4.1 Erkenntnisse aus Modul 1: Daten und Informationen – Datendetektive

Das erste Modul dient zur Vermittlung von Kompetenzen zu statistischen Exploration, Verarbeitung und Darstellung von Daten (vgl. Abschnitt 3.1). Die Modulabschlusspräsentationen der Schülerinnen und Schüler zur statistischen Untersuchung von eigenen Fragestellungen anhand des JIM-Datensatzes zeigen ebenso wie verschiedene Diskussionsrunden, dass es den Schülerinnen und Schülern gelungen ist, ein kritisches Verständnis zu Daten, Information und ihrer Visualisierung zu entwickeln. So wurde beispielsweise in den Präsentationen häufig die Größe des verwendeten JIM-Datensatzes kritisch gewürdigt und die damit verbundene geringe Aussagekraft der sich ergebenden kleinen Teilmengen eingeordnet. Am Ende der Bausteine wurde von den Schülerinnen und Schülern jeweils ein Feedback eingeholt. Die verwendeten Items erhoben die Motivation und Vorstellungen der Lernenden nach Knogler & Lewalter [KL14]. Ein Teil der Rückmeldungen ist in Abb. 3 dargestellt. Die Aussagen wurden jeweils zwischen „Trifft zu“ (gewichtet mit Schulnote 1) bis „Trifft nicht zu“ (gewichtet mit 5) bewertet. Es zeigt sich, dass die Schülerinnen und Schüler den Umgang mit den „Lärmdaten“ als spannender fanden als die Auswertung der JIM-Daten, die persönliche Relevanz jedoch in beiden Bausteinen als nicht sehr hoch erachtet wurde. Bei der weiteren Überarbeitung des Materials und der Fragestellungen ist zu überlegen, wie mit dieser hier berichteten fehlenden Relevanz umgegangen werden kann.

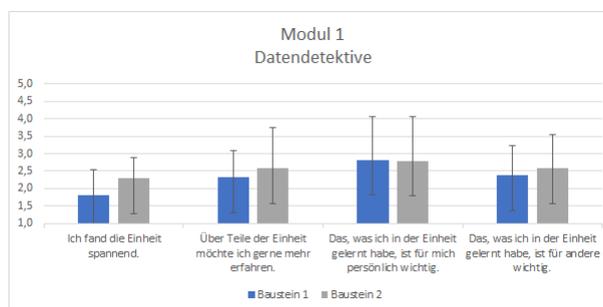


Abb. 3: Schülerrückmeldungen zum Modul 1 - mit gewichtetem Mittelwert und Standardabweichung. ($N = 16$ für Baustein 1; $N = 14$ für Baustein 2)

4.2 Erkenntnisse aus Modul 2: Künstliche Intelligenz und Maschinelles Lernen

In Modul 2 liegt der Schwerpunkt auf der Vermittlung der Grundideen von ML und KI sowie dem Erwerb der Kompetenz zum eigenständigen Programmieren von KI-Modellen (vgl. Abschnitt 3.2). Die beiden Unplugged-Einheiten haben sich als guter thematischer Einstieg in den jeweiligen Baustein erwiesen und waren für die Lernenden hilfreich zur Bildung eines ersten Verständnisses des Themenbereichs. Das Feedback zeigt, dass das angemessene schrittweise Anheben des Schwierigkeitsniveaus bei dem komplexen Themenfeld von ML und KI eine Herausforderung darstellt: Während der „Sweet Learning Computer“-durchweg positiv bewertet wurde, wurden die darauf folgenden einfachen

Beispiele zum manuellen Erstellen von Entscheidungsbäumen mittels CODAP teils als monoton und wenig motivierend empfunden. Für die Erarbeitung der algorithmischen Darstellung von Entscheidungsbäumen wurde beispielhaft der *CART*-Algorithmus in der Variante des Python-Packages *scikit-learn*⁷ mit der Entropie nach Shannon verwendet. Der Versuch, den Begriff der Entropie sowie zugehörige Aspekte der Informationstheorie in Jupyter Notebook als Selbstlerneinheit zu gestalten, hat sich als zu schwierig für die Schülerinnen und Schüler erwiesen, so dass dies stattdessen gemeinsam im Plenum erarbeitet wurde. Auffallend ist in den Rückmeldungen zu diesem Baustein, dass nur wenige Lernende angaben, die Anforderungen bewältigen zu können ($n = 6$ bei $N = 17$ Rückmeldungen), aber die meisten meinten, zugleich viel gelernt zu haben ($n = 12$). Auch ist anzunehmen, dass diese Hürde ein zentraler Punkt sein könnte, warum die Schülerinnen und Schüler den Baustein 1 des Moduls schlechter als den zweiten Baustein bewertet haben (vgl. Abb. 4). Ein weiterer Einflussfaktor ergab sich im Gespräch mit den Schülerinnen und Schülern, die bemängelten, dass auch in diesem Baustein aus didaktischen Gründen nur kleine Datensätze mit nur begrenzter Aussagekraft zum Einsatz kamen. Die Unplugged-Einheit „Brain in a Bag“ sowie die im ersten Baustein erarbeitete Darstellung (vgl. Abb. 2) zum datengetriebenen Problemlösen ermöglichte den Lernenden bereits eine sachlich fundierte Diskussion zur gesellschaftlichen Bedeutung der Verwendung von KI-Systemen, in der Probleme wie Vorurteile in Trainingsdaten („Diskriminierende Algorithmen“) divers betrachtet und diskutiert wurden. Im weiteren Verlauf des Moduls erarbeiteten sich die Schülerinnen und Schüler das eigene Erstellen, Trainieren und Validieren von KNN unter Verwendung von Python und Jupyter Notebook mit dem MNIST-Datensatz zur Ziffernerkennung. Dass es damit gelang, ein grundlegendes Verständnis zu ML und KI sowie der Programmierung eines passenden Netzes zu entwickeln, zeigte sich auch in der Modulabschlusspräsentation. Die Schülerinnen und Schüler programmierten und optimierten in Gruppen für den schon bekannten JIM-Datensatz für die beiden gelernten Verfahren zum ML je ein Modell zur Vorhersage des Geschlechts des Studienteilnehmers und präsentierten anschließend das Ergebnis. Es zeigt sich, dass die Schülerinnen und Schüler ein Grundverständnis für Konzepte von ML entwickeln konnten und Fertigkeiten, eigene KI-Modelle zu programmieren, entwickelt haben. Dies sind wichtige Kompetenzen, die im anschließenden Projektmodul (vgl. Abschnitt 3.3) gefestigt und weiterentwickelt werden sollen.

5 Ausblick und Fazit

Im Rahmens dieses Artikels beschreiben wir die Entwicklung eines Data Science Kurses, der in der Sekundarstufe II als Projektkurs unterrichtet werden kann. Uns ist bewusst, dass nicht jede Lehrkraft, die gerne dieses Thema in den Unterricht einbinden möchte, diese Zeitressource hat. Aus diesem Grund planen wir, die Module und Bausteine auf Basis unserer Erfahrungen der ersten Durchführung weiter zu entwickeln, so dass es auch möglich sein wird,

⁷ <https://scikit-learn.org/stable/modules/tree.html#tree>

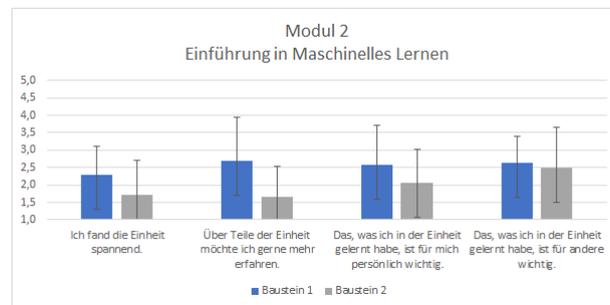


Abb. 4: Schülerurückmeldungen zum Modul 2 - mit gewichtetem Mittelwert und Standardabweichung. ($N = 17$ für Baustein 1; $N = 18$ für Baustein 2)

nur ausgewählte Teile durchzuführen. Sobald die Erkenntnisse der aktuellen Durchführung in unsere Materialien eingearbeitet sein werden, werden diese mit interessierten Lehrkräften unterschiedlicher Schulen erprobt, dabei sukzessive optimiert und einer breiteren Anzahl von Lehrkräften zur Verfügung gestellt. Zur Weiterentwicklung des Gesamtkurses im Sinne des DBR wird dieser im nächsten Schuljahr mit der bisherigen Partnerschule in modifizierter Form durchgeführt. Da die Schülerinnen und Schüler einerseits das durch die Modulstruktur bedingte „Vorratslernen“ bemängelten und andererseits an einigen Stellen Schwierigkeiten haben, die erlernten Inhalte zielgerichtet auf die Fragestellungen des Projekts anzuwenden, ist geplant, die Theoriephasen so mit der Arbeit an einem Datenprojekt zu verzahnen, dass die theoretischen Inhalte gezielt während des Projektablaufs eingebaut werden. Zudem sollen die in den verschiedenen Modulen enthaltenen gesellschaftlichen, sozialen und ethischen Fragestellungen nochmals gebündelt in einem zusätzlichen Modul diskutiert und bearbeitet werden, um die Schülerinnen und Schüler noch mehr mit den darin enthaltenen Fragestellungen vertraut zu machen. Dies alles erfordert jedoch noch einiges an konzeptioneller Arbeit, die sich jedoch durch einen höheren Kompetenzerwerb seitens der Schülerinnen und Schüler bemerkbar machen wird. Zusammenfassend ist zu sagen, dass wir trotz aller noch zu leistenden Entwicklungsarbeit, überzeugt sind, durch den vorgestellten *Data Science Kurs* einen großen Schritt hin zu einem umfassenden Curriculum für diesen sehr komplexen Bereich getan zu haben, so dass wir gespannt auf die nächsten Durchführungen des Kurses und seiner Module sind.

Literatur

- [Be10] Berthold, M. R.; Borgelt, C.; Höppner, F.; Klawonn, F.: Guide to intelligent data analysis: how to intelligently make sense of real data. Springer Science & Business Media, London, 2010.
- [CM97] Cobb, G. W.; Moore, D. S.: Mathematics, statistics, and teaching. *The American Mathematical Monthly* 104/9, S. 801–823, 1997.

-
- [Co03] Cobb, P.; Confrey, J.; diSessa, A.; Lehrer, R.; Schauble, L.: Design Experiments in Educational Research. *Educational Researcher* 32/1, S. 9–13, 2003.
- [Cu16a] Curzon, P.: Brain in a Bag, A CS4FN Computing Activity, https://www.youtube.com/watch?v=lux_ybamCIU, Accessed: 2018-06-15, 2016.
- [Cu16b] Curzon, P.: The Sweet Learning Computer, A CS4FN Computing Activity, www.cs4fn.org/machinelearning, Accessed: 2019-02-10, 2016.
- [GR17] Grillenberger, A.; Romeike, R.: Key concepts of data management: an empirical approach. In: *Proceedings of the 17th Koli Calling Conference on Computing Education Research*, Koli. ACM, New York, S. 30–39, 2017.
- [KL14] Knogler, M.; Lewalter, D.: Design-Based Research im naturwissenschaftlichen Unterricht. *Psychologie in Erziehung und Unterricht* 61/, S. 2–14, 2014.
- [LCB98] LeCun, Y.; Cortes, C.; Burges, C. J.: MNIST handwritten digit database, <http://yann.lecun.com/exdb/mnist/>, Accessed: 2018-06-15, 1998.
- [Ri15] Ridsdale, C.; Rothwell, J.; Smit, M.; Ali-Hassan, H.; Bliemel, M.; Irvine, D.; Kelley, D.; Matwin, S.; Wuetherick, B.: Strategies and best practices for data literacy education: knowledge synthesis report. 2015.
- [Ri16] Ridgway, J.: Implications of the Data Revolution for Statistics Education. *International Statistical Review* 84/3, S. 528–549, 2016.
- [Rö16] Röhner, G.; Brinda, T.; Denke, V.; Hellmig, L.; Heußner, T.; Pasternak, A.; Schwill, A.; Seiffert, M.: Bildungsstandards Informatik für die Sekundarstufe II. Beilage zu LOG IN, Heft/183/184, 2016.
- [SBS18] Sentance, S.; Barendsen, E.; Schulte, C.: *Computer Science Education: Perspectives on Teaching and learning in school*. Bloomsbury Academic, London, 2018.
- [TD16] Tedre, M.; Denning, P. J.: The Long Quest for Computational Thinking. In: *Proceedings of the 16th Koli Calling International Conference on Computing Education Research*. Koli Calling '16, ACM, Koli, Finland, S. 120–129, 2016.
- [Tv09] Thijs, A.; van den Akker, J.: *Curriculum in development*. SLO - Netherlands Institute for curriculum development, Enschede, 2009.