

Rolf BIEHLER, Daniel FRISCHEMEIER, Susanne PODWORNÝ,  
Thomas WASSONG, Carsten SCHULTE, Simone OPEL &  
Michael SCHLICHTIG, Paderborn

## **Substantielle Digitale Bildung statt nur Anwendung digitaler Werkzeuge – Impulse aus einem Pilotprojekt zu Data Science in der Sekundarstufe**

### **Einleitung**

Die Diskussion um digitale Bildung wird in der Mathematikdidaktik manchmal verengt auf den Aspekt des Einsatzes digitaler Werkzeuge, in der Mediendidaktik auf die Aspekte der Nutzung des Internets, von Tablets und von interaktiven Whiteboards. Der Begriff des Algorithmus wird schon länger de facto nicht mehr als fundamental für den Mathematikunterricht angesehen, im Unterschied zu etwa Engel (1977). Demgegenüber werden in der Informatikdidaktik stärker fundamentale Ideen des Programmierens, Konzepte wie „algorithm literacy“ sowie die gesellschaftlichen Bedingungen und Auswirkungen der Digitalisierung zum Thema des Unterrichts gemacht werden sollen, wenn auch bisher im Unterricht Themen wie datengetriebene Algorithmen und maschinelles Lernen praktisch keine Rolle spielt. Seit einigen Jahren hat sich unter der Bezeichnung „Data Science“ eine neue Disziplin im Schnittbereich von Mathematik bzw. Statistik und Informatik etabliert. Bachelor- und Masterstudiengänge und universitäre Forschungszentren mit dieser Bezeichnung sprießen aus dem Boden. Als Komponente universitärer Schlüsselqualifikationen bzw. eines *studium generale* wird „data literacy“ diskutiert (vgl. Ridsdale et al., 2015). Es gibt eine Fülle von populärwissenschaftlichen Veröffentlichungen, die sich mit der neuen Rolle von Daten in der Gesellschaft und in allen privaten Bereichen beschäftigen und Konsequenzen für die Bildung artikulieren (Weigend, 2017; O, Neill, 2016; spitz, 2016). Fast täglich berichten Medien über diese Aspekte, z.B. über so genannte „Frankenstein Algorithmen“, die man nurmehr beobachten kann, ohne zu verstehen, warum sie sich so verhalten und wer dafür verantwortlich sei (sueddeutsche.de, 5.9.18). Bei Amazon wurden diskriminierende „sexistische Algorithmen“ entdeckt, die Frauen in Bewerbungsverfahren systematisch benachteiligen, weil die Algorithmen sich das selbst aus den vorliegenden Daten beigebracht hatten (meedia.de, 17.10.18).

Auf diesem Hintergrund ist an der Universität Paderborn eine Kooperation zwischen Mathematik- und Informatikdidaktik entstanden, die das Ziel verfolgt fundamentale Ideen der Data Science im Hinblick auf die schulische Bildung zu identifizieren, Unterrichts- und Fortbildungsmaterialien zu entwickeln, die zuvor im Unterricht getestet wurden.

## **ProDaBi: Projekt zu Data Science und Big Data in der Schule**

Das Projekt ProDaBi ([www.prodabi.de](http://www.prodabi.de)) wird seit 2017 von der Deutsche Telekom Stiftung gefördert, zunächst in Form der Unterstützung eines internationalen Symposiums, dann als Pilotprojekt, das die Universität Paderborn im Kooperation mit zwei Paderborner Gymnasien im Schuljahr 2018/2019 durchführt. Im interdisziplinären Symposium wurden die unterschiedlichen Perspektiven von Statistik und Statistikdidaktik, Informatik und Informatikdidaktik eingebracht und mit Vertretern industrieller und kommerzieller Data Science Anwendungen, Vertretern der Digital Humanities und der Sozioinformatik diskutiert (Biehler et. al, 2018). Auf dieser Basis konnte eine erste theoretische Basis für curriculare Ziele und Inhalte entwickelt werden (Biehler und Schulte, 2018).

### **Neue Akzente für die Stochastikdidaktik**

Die Stochastikdidaktik kann sich an Data Science sehr gut mit Konzepten der Explorativen Datenanalyse andocken (Biehler, 1991), die aber trotz der Aufnahme der Leitidee „Daten“ in die Bildungsstandards ein eher randständiges Thema geblieben ist, während „data detective“ mittlerweile eine Berufsbezeichnung geworden ist. Data Science bringt verschiedene neue Akzente die Stochastikdidaktik ein. Der Datenbegriff ist entscheidend zu erweitern. Multivariate Daten erheblichen Umfangs (Big Data), die z.B. auch durch Sensoren und Smartphones gewonnen wurden, die als Bewegungsdaten anfallen und Bilder und Texte im Internet und den sozialen Medien miteinschließen. Während in der Statistikdidaktik ein Datenanalysezyklus (Wild und Pfannkuch, 1999) mit den „conclusions“ endet, steht am Ende eines Data Science Zyklus („CRISP-DM“) oft das „deployment“ (Berthold ...), d.h. ein System wird in der Praxis implementiert und soziale Bedingungen und Auswirkungen sind mit zu bedenken. Die Aufbereitung und Bereinigung von Daten ist ein zentraler Bestandteil der Data Science und nicht etwas, was man besser den Schüler\*innen vorenthält. Bei der Modellierung spielt die Vorhersagekraft eine wesentliche Rolle, bei der Modellierung unterscheidet man Trainingsdaten, mit denen das Modell erstellt wird, von Testdaten, an denen sich das Modell bewähren muss und bei denen das häufig anzutreffende „Overfitting“ zu korrigieren ist. Eine weitere Umakzentuierung betrifft die erforderlichen Werkzeuge. In der datenbezogenen Stochastik werden für die schulische Bildung Werkzeuge wie Fathom, Tinkerplots ([www.stochastik-interaktiv.de](http://www.stochastik-interaktiv.de)) oder die damit verwandte, wenn auch nicht so leistungsstarke Webapplikation CODAP ([codap.concord.org](http://codap.concord.org)) favorisiert. In der Praxis der Data Science werden Werkzeuge wie *R* oder *Python* verwendet, weil darin Bibliotheken mit Data Science Algorithmen zur Verfügung stehen und diese Umgebungen das Erstellen, Codieren und

Testen eigener Algorithmen ermöglichen. Das sind einige der Umakzentuierungen aus Sicht der Stochastikdidaktik, für die Informatikdidaktik gilt Analoges (Biehler und Schulte, 2018).

### **Der Projektkurs Data Science im Schuljahr 2018/19**

In NRW gibt es für Schulen die Möglichkeit so genannte „Projektkurse“ einzurichten, die von den Schüler\*innen im Wahlpflichtbereich in der Oberstufe gebucht werden können und die ein Schuljahr lang gehen und bei denen die Noten für das Abitur zählen. Solche Projektkurse unterliegen keinen curricularen Vorgaben, sind aber einem oder mehreren schulischen Leitfächern zugeordnet. Zwei Paderborner Gymnasien (Theodorianum und Reismann) haben einen solchen Projektkurs zum Thema Data Science eingerichtet, der vom September 2018 bis April 2019 jede Woche im Umfang von 3 Schulstunden im Computerlabor der Informatikdidaktik an der Universität Paderborn durchgeführt wird. Als digitale Werkzeuge nutzen wir CODAP als Einstiegssoftware in interaktive Datenanalyse und in das Thema „Decision Trees“. Anschließend nutzen die Schüler\*innen die Programmiersprache in der Umgebung von Jupyter-Notebooks (Toomey, 2017), in denen in Text eingebettete Algorithmen und Verfahren in vorbereiteter Weise zur Verfügung gestellt werden, so dass die Lernenden erstmal in bereits fertigen Code eingeführt werden, den sie nachvollziehen und gezielt verändern können, als eine Möglichkeit für einen sanften Einstieg in die neue Programmierumgebung.

Der Projektkurs umfasst 4 Module: Modul 1: Daten und Datendetektive, Modul 2: Künstliche Intelligenz, maschinelles Lernen, datengetriebene Algorithmen, Modul 3: Data Science Projekt mit Partnern aus Industrie und Verwaltung, Modul 4: Data Science im gesellschaftlichen Kontext: Bedingungen und Implikationen. In Modul 1 stehen zunächst „klassische“ Umfrage Daten im Vordergrund: An Hand eines Teils des Fragebogens der JIM-Studie (Behrends und Rathgeb, 2017) wurden Daten an den beteiligten Schulen erhoben und von den Schüler\*innen nach unterschiedlichen Fragestellungen ausgewertet und die Ergebnisse präsentiert. Ferner wurden von Sensoren an ausgewählten Standorten in Paderborn (jeweils ca. 500.000 Datenpunkte) gesammelte Lärmdaten aufbereitet und analysiert. In Modul 2 wurden Entscheidungsbäume und künstliche neuronale Netze als zwei Methoden datengetriebenen maschinellen Lernens vorgestellt, an Daten erprobt und kontrastiert, wobei besonders die auf symbolischer KI beruhende Transparenz und Interpretierbarkeit Eigenschaften der Entscheidungsbäume zu den intransparenten neuronalen Netzen abgehoben wurde. In Modul 3 arbeiten die Schüler\*innen mit der Firma RTB ([www.rtb-bl.de/RTB/](http://www.rtb-bl.de/RTB/)) und dem Abfallentsorgungs- und Stadtreinigungsbetrieb Paderborn (ASP) zusammen, der für die

Parkraumbewirtschaftung in Paderborn zuständig ist ([www.paderborn.de/microsite/asp/](http://www.paderborn.de/microsite/asp/)), zusammen, um umfangreiche Daten aus Parkscheinautomaten in Paderborn so zu analysieren, dass eine bessere Vorhersage freier Parkplätze möglich wird. Modul 4 wird in diesem Durchgang in die anderen Module integriert.

## Literatur

- Behrens, P., & Rathgeb, T. (2017). JIM-Studie 2017 - Jugend, Information, (Multi-)Media, Basisstudie zum Medienumgang 12- bis 19-jähriger in Deutschland. Stuttgart: Medienpädagogischer Forschungsverbund Südwest.
- Berthold, M. R., Borgelt, C., Höppner, F., & Klawonn, F. (2010). *Guide to Intelligent Data Analysis - How to Intelligently Make Sense of Real Data*. London: Springer.
- Biehler, R. (Ed.) (1991). *Explorative Datenanalyse im Mathematikunterricht: Der Mathematikunterricht* 37, Heft 6.
- Biehler, R., Budde, L., Frischemeier, D., Heinemann, B., Podworny, S., Schulte, C., & Wassong, T. (Eds.). (2018). *Paderborn Symposium on Data Science Education at School Level 2017: The Collected Extended Abstracts*. Paderborn: Universitätsbibliothek Paderborn.
- Biehler, R., & Schulte, C. (2018). Perspectives for an interdisciplinary data science curriculum at German secondary schools. In R. Biehler, L. Budde, D. Frischemeier, B. Heinemann, S. Podworny, C. Schulte, & T. Wassong (Eds.), *Paderborn Symposium on Data Science Education at School Level 2017: The Collected Extended Abstracts* (pp. 2-14). Paderborn: Universitätsbibliothek Paderborn.
- Engel, A. (1977). *Elementarmathematik vom algorithmischen Standpunkt*. Stuttgart: Klett.
- O'Neil, C. (2016). *Weapons of Math Destruction - How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.
- Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., & Wuetherick, B. (2015). *Strategies and best practices for data literacy education: knowledge synthesis report*. Quelle: <http://hdl.handle.net/10222/64578>.
- Spitz, M. (2017). *Daten - das Öl des 21. Jahrhunderts? Nachhaltigkeit im digitalen Zeitalter*. Hamburg: Hoffmann and Campe.
- Toomey, D. (2017). *Jupyter for Data Science - Exploratory analysis, statistical modeling, machine learning, and data visualization with Jupyter*. Birmingham: Packt Publishing.
- Weigand, A. (2017). *Data for the people. Wie wir die Macht über unsere Daten zurückerobern*. Hamburg: Murmann Publ.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical Review*, 67(3), 223-265.