

Joachim ENGEL, Ludwigsburg

Data Science in der Lehrerbildung: Chancen, Herausforderungen und erste Einsichten

Data Science, die Wissenschaft des Lernens aus Daten, gilt als Wissenschaft des 21. Jahrhunderts. Die Data Literacy Charta (Schüller et al., 2020) fordert Data Literacy als unverzichtbaren, in allen Bildungsbereichen zu vermittelnden Teil der Allgemeinbildung. Dieser Beitrag plädiert für einen Kurs in Data Science für angehende Lehrkräfte, der über die Vermittlung von grundlegenden Kenntnissen in Statistik und Algorithmen hinausgehend gesellschaftliche Implikationen von automatisierten Entscheidungen thematisiert.

Kompetenter Umgang mit Daten als Teil der Allgemeinbildung

Data Science wurde als praktische Wissenschaft konzipiert, um konkrete Probleme in Wissenschaft, Technologie und Gesellschaft zu lösen. Als stark interdisziplinäres Gebiet bezieht sich Data Science auf eine Reihe von Fähigkeiten und Techniken, die Statistik, Informatik, Mathematik und Fachwissen in dem Bereich, aus dem die Daten stammen, umfassen. Innovation, sozialer Fortschritt und das Wohlergehen unserer Zivilgesellschaft erfordern, dass Menschen in Wissenschaft, Wirtschaft, Politik und Gesellschaft wissen, wie sie Daten auswerten und sinnvoll nutzen können, um ein fundiertes Verständnis unserer Welt zu entwickeln und drängende Herausforderungen mit empirischen Erkenntnissen und fundierten datengestützten Argumenten anzugehen. Mündige Bürger*innen brauchen im Zeitalter der Digitalisierung Fähigkeiten, um sich in der Fülle von Daten zurechtzufinden, um informierte Entscheidungen zu treffen - im Alltag ebenso wie auf verschiedenen gesellschaftlichen und politischen Ebenen. Gleichzeitig wirft Big Data mit seinen Möglichkeiten der Überwachung, Manipulation und Kontrolle ernsthafte Probleme für Demokratie und Freiheit auf (siehe z.B. Helbing et al., 2017). Algorithmen, die auf Daten zurückgreifen, werden verwendet, um Profile von Mitgliedern der Gesellschaft zu erstellen und wichtige Entscheidungen zu treffen, die sich unverhältnismäßig stark auf diejenigen auswirken, die weniger Privilegien und Ressourcen zur Verfügung haben (O'Neill, 2016). In einer zunehmend datenbasierten Gesellschaft erhalten Daten oft den Status einer objektiven Tatsache, obwohl sie oft partiell und verzerrt sind. Als Konsequenz kann die Vermittlung von Data Literacy und Data Science nicht auf das Erlernen von technischen Kenntnissen über Algorithmen, Big-Data-Management und Computertechnik reduziert werden. Im Mittelpunkt von Data Literacy stehen Fähigkeiten, Daten zu sammeln, zu verwalten, zu bewerten und auf kritische Weise neues Wissen aus Daten zu gewinnen.

nen. Dies schließt ausdrücklich die Fähigkeit ein, Daten und ihre Auswirkungen auf soziale und politische Interaktionen kritisch zu bewerten. Die Ausbildung in Data Science muss über eine Kombination von statistischen und technischen Fähigkeiten hinausgehen, kritisches Denken und Bürgersinn einschließen und die Fähigkeiten zur Bewertung, Analyse und Interpretation von Daten fördern. Ein solches Konzept weist viele Überschneidungen mit Zivild Statistik auf (Engel et. al., 2019; Ridgway, 2022).

Herausforderungen für Bildung und Didaktik

Data Literacy als Teil des allgemeinen Bildungsauftrags der Schulen im 21. Jahrhundert betrifft neben dem Fach Informatik und dem sozialwissenschaftlichen Unterricht in besonderer Weise den Mathematikunterricht. Diese Entwicklungen stellen eine fünffache Herausforderung dar, die über die Anwendung bestimmter statistischer Rechenverfahren und Argumentationsweisen des bisherigen Unterrichts weit hinausgeht:

- *Inhaltlich, kontextuell:* Die verfügbaren Daten müssen auf ihre Relevanz, Zuverlässigkeit und Glaubwürdigkeit für die ursprünglichen Fragen geprüft werden; es muss untersucht werden, inwieweit die gezogenen Schlussfolgerungen durch die Daten gestützt werden und welche Grenzen den Schlussfolgerungen gesetzt sind
- *Didaktisch:* Eine interdisziplinäre und fächerübergreifende Zusammenarbeit, z.B. mit den Sozialwissenschaften und der Informatik, ist notwendig, was neben möglichem Teamteaching auch die Vertrautheit mit der Fachdidaktik und der Unterrichtsmethodik unterschiedlicher Fachtraditionen erfordert.
- *Statistisch:* Geeignete Visualisierungen und statistische Zusammenfassungen müssen ausgewählt werden; multivariate Phänomene müssen auf erklärende Drittvariablen getestet werden; kausale Schlussfolgerungen müssen auf ihre Gültigkeit hin überprüft werden
- *Technologisch:* innovative Technologien und Software (CODAP, Gapminder, Python, R usw.) werden zur Visualisierung und Erforschung multivariater Datensätze eingesetzt; schlecht aufbereitete multivariate Datensätze, die aus dem Internet heruntergeladen werden, müssen möglicherweise zunächst für die Analyse und Visualisierung kuratiert (bereinigt und aufbereitet) werden.
- *Systemisch:* Fähigkeiten im Bereich Data Literacy können nicht an ein einzelnes Schulfach, sei es Mathematik, Informatik, Sozialkunde, delegiert werden, sondern gehen über die traditionelle Aufteilung des schulischen Fächerkanons hinaus. Dies stellt die Organisation von Lehrplänen und die Ausbildung von Lehrenden vor große Herausforderungen.

Erste Umsetzungen und Schlussfolgerungen

Inhalte und Ziele: Anknüpfend an frühere Seminarveranstaltungen für Lehramtsstudierende zu Zivild Statistik (Engel et al., 2019) verfolgte das Seminar „Data Science“ für Lehramtsstudierende mit Fach Mathematik an der PH

Ludwigsburg folgende globalen Ziele: Studierende sollen (1.) einige grundlegende Kenntnisse in Statistik und Datenvisualisierung erwerben (2.) die Herkunft und Qualität von Daten kritisch bewerten lernen (3.) an ausgewählten Verfahren des maschinellen Lernens (k-NN, CART, RandomForests) die Qualität von maschinellen Entscheidungen bewerten (4.) die Rolle menschlicher Kontrolle (bzw. Mensch-Maschine Interaktion) bei Entscheidungen verstehen, die die Persönlichkeitsrechte von Menschen betreffen.

Technische Ziele: Allgemeines technisches Ziel ist die Entwicklung eines basalen Verständnisses und eine kritische Würdigung einiger ML-Methoden des überwachten Lernen. Klassifizierungen mittels Entscheidungsbäume erweisen sich als ein sehr geeigneter Einstieg, der zunächst händisch durchgeführt werden kann bevor technische Hilfsmittel wie das Arbor-Plug-In der Software-Umgebung CODAP (Biehler et al., 2021) und später auch automatisierte Algorithmen wie z.B. das R-Paket rpart eingesetzt werden. Eine Herausforderung ist die Vermittlung des Problems des Overfitting.

Implementierung: Ein Seminar im Umfang von 2 SWS für Lehramtsstudierende im 2. Studienjahr wurde mehrfach durchgeführt. Zunächst standen Grundkonzepte der explorativen Datenanalyse und Statistik (Lage- und Streuungsparameter, Verteilung und Visualisierung von Daten, Indexzahlen) im Mittelpunkt. sowie Einführungen in CODAP, Gapminder und R als Werkzeuge zur Datenvisualisierung. Grundfragen zu Datenerhebung und Datenqualität, Verteilungen und Visualisierung von Daten, Lage- und Streuungsparameter, Indexzahlen, Korrelation, Regression & Multivariate Phänomene wurden thematisiert ebenso wie gesellschaftliche und ethische Herausforderungen von Big Data und Machine Learning.

Didaktisches Konzept: Eine idealtypische Seminarsitzung bestand aus 30-40 Minuten Instruktion durch den Dozenten, 30-40 Minuten Gruppenarbeit der Studierenden am PC und 20 Minuten Diskussion. Die Prüfungsleistung bestand aus regelmäßigen Hausaufgaben sowie eines im Team von 2 – 3 Personen zu bearbeitenden Datenanalyseprojektes, einzureichen in Form eines ca. 10 minütigen Videos. Letzteres erlaubt innere Differenzierung nach Interessen und technischen Kompetenzen: Während einige Teams inhaltliche Analysen mit Hilfe von Gapminder oder CODAP erstellten, vertiefte sich ein Team (Teilnehmer mit Zweitfach Informatik) in die Anwendung von Random Forests während wiederum andere eine Buchbesprechung (O’Neill, 2017) zu gesellschaftlichen Risiken von Big Data vorbereiteten.

Fazit

Insgesamt wurden die vielfältigen Lerngegenstände in der Regel mit hohem studentischem Engagement angenommen. Wichtig für die Motivation ist ein

reichhaltiges Angebot an inhaltlichen Kontexten (siehe <https://ase.web.org/islp/pcs>) sowie ein gestreutes Angebot bezüglich der Anforderungen technologischer Anwendungen, von Gapminder bis R. Die im Team produzierten Videos waren für viele Teilnehmende ein – mit viel Einsatz erreichtes - Highlight.

Eine gute Datenanalyse muss durch ein Ziel motiviert sein: Sie trägt zur Lösung eines klar umrissenen Problems bei oder beantwortet eine spezifische Frage. Die beste aller Analysen ist aber wertlos, wenn sie auf schwachen Daten beruht. Studierende müssen lernen, ihre Datenquellen und deren Herkunft zu dokumentieren. Algorithmen sind mächtige Werkzeuge zur Problemlösung, aber es ist letztlich der Mensch, der das Werkzeug auswählt und anwendet, der die Grenzen des Werkzeugs kennt, die Beschränkungen des Datensatzes versteht und die möglichen unbeabsichtigten Folgen eines Algorithmus verstehen kann, der ein Kriterium optimiert. An ausgewählten Beispielen können Studierende auch in Einführungsveranstaltungen die Grenzen der Algorithmen erfahren. Schließlich sollen Studierende lernen zu fragen, warum eine Analyse durchgeführt wird und die ethischen Konsequenzen der Antwort bedenken.

Literatur

- Biehler, R. & Fleischer, Y. (2021). Introducing students to machine learning with decision trees using CODAP and Jupyter Notebooks, *Teaching Statistics*, 43, 133–142.
- Engel, J., Biehler, R., Frischemeier, D., Podworny, S., Schiller, A. & Martignon, L. (2019). Zivilstatistik: Konzept einer neuen Perspektive auf Data Literacy und Statistical Literacy. *AStA Wirtschafts- und Sozialstatistisches Archiv* 13(3/4), 213–244.
- Helbing, D., Frey, B., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., van den Hoven, J., Zicari, R. & Zwitter, A. (2017). Digitale Demokratie statt Datendiktatur. In: C. Könneker (Hrsg.), *Unsere digitale Zukunft* (S. 3–21). Springer. https://doi.org/10.1007/978-3-662-53836-4_1
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality & Threatens Democracy*. New York, NY: Crown Publishing Group.
- Ridgway, J. (2022). *Statistics for empowerment and social engagement: teaching Civic Statistics to develop informed citizens*. Springer.
- Schüller, K., Koch, H. & Rampelt, F. (2020). *Data-Literacy Charta*. https://www.stifterverband.org/sites/default/files/data-literacy-charta_v1_2.pdf