

ENGEL, Joachim  
Ludwigsburg

## **Data Science für informierte Bürgerinnen und Bürger**

### **Warum Data Science für alle?**

Die Informationslandschaft verändert sich im digitalen Zeitalter dramatisch aufgrund der zunehmenden Verfügbarkeit von Informationen über das Internet, der weit verbreiteten Nutzung digitaler Technologien, der Fülle an Daten und des einfachen Zugangs zu Datenanalysetools. 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 gesellschaftliche Herausforderungen mit empirischen Erkenntnissen und fundierten datengestützten Argumenten anzugehen.

### **Hoffnungen und Drohkulissen: Die Big Data Revolution**

Big Data birgt Hoffnungen, eine nachhaltige und lebenswerte Zukunft für unseren Planeten zu sichern. Es ermöglicht wichtige Erkenntnisse und Lösungen in essenziellen Bereichen wie Umweltschutz, effizientem Ressourcenmanagement und Krankheitsprävention und vieles mehr. Gleichzeitig bringt die Technologie aber auch erhebliche Risiken mit sich: Die weitreichenden Möglichkeiten zur Überwachung, Manipulation und Kontrolle stellen eine ernsthafte Bedrohung für Demokratie und Freiheit dar (vgl. Helbing et al., 2017). Die Fähigkeit, die Glaubwürdigkeit von Informationen und deren Quellen zu beurteilen, war noch nie so wichtig wie heute. Algorithmen, die sich auf Daten stützen, werden verwendet, um Profile von Mitgliedern der Gesellschaft zu erstellen und wichtige Entscheidungen zu beeinflussen, die sich unverhältnismäßig stark auf diejenigen auswirken, die weniger Privilegien und Ressourcen zur Verfügung haben. (O'Neil, 2016). Folglich darf die Vermittlung von Datenkompetenz und Data Science nicht nur auf das Erlernen von technischen Kenntnissen über Algorithmen, Big-Data-Management und Computertechnik reduziert werden.

### **Data Science für alle: Übergreifende Zielsetzungen**

Auf der Grundlage relevanter Literatur, einer Analyse von Notwendigkeiten zur Stärkung demokratischer Werte im digitalen Zeitalter und der Reflexion der eigenen Unterrichtspraxis will dieser Beitrag Anregungen und Hinweise für die Gestaltung von Data Science Education für informierte Bürger bieten. Es wird kein spezifischer mathematischer oder technologischer Hintergrund

In: L. Schick, M. Platz & A. Lambert (Hrsg.),  
Beiträge zum Mathematikunterricht 2025.

58. Jahrestagung der Gesellschaft für Didaktik der Mathematik. WTM.  
<https://doi.org/10.37626/GA9783959873307.0>

vorausgesetzt, so dass das Konzept und die Ziele für jede Gruppe informierter Bürgerinnen und Bürgern gelten können.

- *Bewusstsein dafür schärfen, wo wir im täglichen Leben auf Produkte von Data Science treffen.* Einige Beispiele: Smartphones, E-Commerce und Online-Shopping, Betrugserkennung; Banken und Finanzen; Bildung und Apps.
- *Bewusstsein für Datenqualität und Datentauglichkeit schulen:* Datenbasierte Evidenz (in Verbindung mit kritischem Denken) ist der Weg zu neuem Wissen. Daten sind jedoch nicht die „objektive Wahrheit“, sondern verzerrt und von Interessen beeinflusst. Wichtige kritische Fragen zu den in Erhebungen verwendeten Methoden sind u.a.:
  - Sind die Messgrößen (z. B. ein Fragebogen) gut definiert? Sind die Messgrößen robust und für die Zwecke, für die sie verwendet werden, geeignet?
  - Sind Metadaten (d. h. detaillierte Erläuterungen zur Definition der Variablen, den Merkmalen der Stichprobe usw.) verfügbar?
  - Waren die Stichprobenverfahren angemessen? Wer fehlt in den erhobenen Daten?

Über das technische Wissen hinaus ist es wichtig, dass der Einzelne in der Lage ist, kritische Fragen zu stellen, um die Glaubwürdigkeit und Gültigkeit aller Daten, Ergebnisse oder Schlussfolgerungen, auf die er stößt, sowohl aus technischen als auch aus logischen Gründen zu beurteilen. Es ist wichtig, Erzählungen und Interpretationen von Daten sowie die daraus gezogenen Schlussfolgerungen aus einer kritischen Perspektive zu betrachten: Sind die Schlussfolgerungen mit anderen verfügbaren Erkenntnissen vereinbar?

- *Bewusstsein für die Verzerrungen bei maschinellen Entscheidungen schärfen:* Jede Verzerrung in der Trainingsmenge wird in der Testmenge verstärkt und führt zu verzerrten Entscheidungen. Auch das Design des Algorithmus selbst kann zu Verzerrungen führen.
- *Vermittlung eines Bewusstseins für die Auswirkungen von Data-Science-Produkten auf die Gesellschaft:* Hierzu gehören Datenschutzfragen, Überwachung, Voreingenommenheit und Fairness, Autonomie und die Zukunft der Arbeit. Ein besonderes Problem ist die Intransparenz (oft als „Blackbox“ bezeichnet) der meisten Algorithmen für maschinelles Lernen. Die Übertragung der Entscheidungsfindung an eine Maschine, insbesondere in kritischen Bereichen wie der Strafjustiz oder bei medizinischen Entscheidungen über Leben und Tod, wirft schwerwiegende ethische und moralische Fragen auf

- *Grundlegende technologische Grundlagen des maschinellen Lernens*: Natürlich geht es bei der Data Science Education auch darum, etwas über Technologie zu lernen. Meiner Erfahrung nach (Erickson & Engel, 2023) sind Entscheidungsbäume ein guter Einstiegspunkt, um die Prinzipien des maschinellen Lernens zu verstehen. Entscheidungsbäume sind intuitiv, einfach anzuwenden und transparent. Je nach Lerngruppe kann das Thema Bäume auf anspruchsvollere ML-Themen wie Random Forests, Bagging und Boosting erweitert werden.

## **Schlussfolgerungen**

**Motivation, Problemdefinition und Kontext:** Datenanalyse muss durch ein Ziel motiviert sein. Und sie muss in einen klaren Kontext eingebettet sein, in dem sie angewendet und informiert werden soll. Eine gute Data Science-Anwendung löst ein klar definiertes Problem oder beantwortet eine spezifische Frage. Dies ist die harte Arbeit, die vor der Anwendung der automatisierten Tools geleistet werden muss. Und es ist eines der schwierigsten Dinge, die Studenten in der Schule und an der Universität lernen und verinnerlichen müssen.

**Datenherkunft und Metadaten:** Die ausgefeiltste Analyse ist wertlos, wenn sie auf schwachen oder fragwürdigen Daten beruht. Die Datenanalyse darf nicht blindlings durchgeführt werden, d. h. auf Daten angewendet werden, die ungeeignet oder voller Fehler und Lücken sind. Die Studierenden müssen lernen, ihre Datenquellen und deren Herkunft zu dokumentieren. Und, was noch wichtiger ist, skeptisch zu sein, was die Zuverlässigkeit ihrer Daten angeht.

**Mensch-Maschine-Interaktion und Entscheidungen:** Die Analyse muss eine Zusammenarbeit zwischen menschlichen Analytikern und Computeralgorithmen sein, wobei die Algorithmen als Werkzeuge dienen, die von Menschen bedient werden. Es ist der menschliche Analytiker, der sich an veränderte Umstände anpassen, die Grenzen des Modells erkennen, die Beschränkungen des Datensatzes verstehen, bewerten und korrigieren, außergewöhnliche und abweichende Werte ausschließen oder berücksichtigen und die möglichen unbeabsichtigten Folgen eines Modells verstehen kann.

**Ethik:** Zunehmend werden ethische Konsequenzen der Data-Science-Analyse aufgedeckt. Wir dürfen uns nicht auf Algorithmen verlassen und müssen unsere Studierenden darin schulen, ethisch zu denken und zu handeln und diese Grundsätze auf ihre Arbeit anzuwenden. Die Studierenden sollten lernen zu fragen, warum eine Analyse durchgeführt wird, und die ethischen Konsequenzen der Antwort in Betracht ziehen. Während viele im Bereich

der Datenwissenschaft Modelle als objektiv und unvoreingenommen betrachten, definiert O'Neil (S. 21) Modelle als „in Mathematik eingebettete Meinungen“. Während die Mathematik dem Modell den Anschein von Objektivität verleiht, hat in Wirklichkeit jemand das Modell erstellt und entschieden, welche Daten verwendet werden sollen, welche Variablen einbezogen werden sollen, welche Modellform verwendet werden soll und so weiter. Ein Modell ist in Wirklichkeit eine Meinung, die sowohl die Voreingenommenheit des Modellierers als auch die Voreingenommenheit der Daten selbst widerspiegelt. Diejenigen, die Data Science studieren, müssen für diese ethischen Fragen sensibilisiert und darin geschult werden, wie sie Voreingenommenheit und Diskriminierung in Modellen vermeiden können.

Problemlösung: Sicher, wir müssen auch technische Fähigkeiten wie Programmieren, Algorithmen für maschinelles Lernen und andere Big-Data-Themen vermitteln. Aber das sollte nicht der Schwerpunkt eines Data-Science-Lehrplans sein, genauso wenig wie Calculus (Differenzial- und Integralrechnung) der Schwerpunkt eines Physik-Lehrplans sein sollte. Es handelt sich dabei um Werkzeuge, und die Schüler sollten sie gut beherrschen - aber zuerst müssen sie lernen, warum und wie sie sie nutzen können. Der ultimative Maßstab für den Erfolg ist die Lösung des jeweiligen Problems durch die Bereitstellung nachhaltiger Lösungen, die greifbare Auswirkungen haben.

Studierende sollten schon früh in ihrer Ausbildung lernen, dass es bei Data Science NICHT um die Tools geht! Data-Science-Tools, wie leistungsfähig sie auch sein mögen, sind ein „Wie“, nicht das „Was“. Letztlich geht es nicht darum, die Tools gut zu kennen und zu nutzen, sondern darum, nachhaltige Lösungen für schwierige Probleme zu finden und zu nutzen. Andernfalls sollten wir uns nicht wundern, wenn die klügsten Köpfe, die wir ausbilden, ihre Intelligenz in erster Linie dazu nutzen, andere Menschen dazu zu bewegen, auf bestimmte Verbraucheranzeigen zu klicken, anstatt ihr Wissen zur Lösung dringender sozialer und gesellschaftlicher Probleme einzusetzen.

## Literatur

- Erickson, T., & Engel, J. (2023). What goes before the CART. Introducing classification trees with ARBOR and CODAP. *Teaching Statistics*, 45, S104–S113.
- Helbing, D., Frey, B., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., van den Hoven, J., Zicari, R. & Zwitter, A. (2017). Digitale Demokratie oder Datendiktatur. In: C. Könneker (Ed.), *Unsere digitale Zukunft*. [https://doi.org/10.1007/978-3-662-53836-4\\_1](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*. Crown Publishing Group.