

## Datenbasierte Entscheidungsbäume mit unplugged Datenkarten als Einstieg in das maschinelle Lernen

### 1. Einleitung

Daten und maschinelles Lernen (ML) spielen eine zunehmend wichtige Rolle in verschiedenen Bereichen unseres Lebens und prägen durch resultierende Anwendungen künstlicher Intelligenz (KI) unseren Alltag. Dies hat Forderungen nach der Integration von Data Science und ML in Schulcurricula verstärkt (Engel, 2017; Ridgway, 2016). Engel (2017) betont die Notwendigkeit, Schülerinnen und Schüler zu kritischen Konsumenten datenbasierter Phänomene zu bilden. Ähnliche Ziele verfolgen Initiativen zur AI Literacy (Casal-Otero et al., 2023; Long & Magerko, 2020) sowie die deutsche Data Literacy Charta (Schüller et al., 2021), die fordert, Datenkompetenz in Bildungsstandards aufzunehmen und Lernende zu befähigen, datenbasierte Entscheidungen aktiv mitzugestalten.

Da ML oft als Expertenwissen gesehen wird (Sulmont et al., 2019), haben wir im ProDaBi-Projekt (prodabi.de) Ansätze entwickelt, ausgewählte Themen zu elementarisieren und altersgerecht zu kontextualisieren. Basierend auf Erfahrungen im Sekundarbereich II (Biehler & Fleischer, 2021; Fleischer et al., 2022) wurde eine Unterrichtseinheit für die 6. Klasse entwickelt. Mit unplugged Datenkarten zum Thema Lebensmittel und einer darauf basierenden Lernaktivität werden datenbasierte Entscheidungsbäume erstellt.

### 2. Datenbasierte Entscheidungsbäumen mit unplugged Datenkarten

Ein Entscheidungsbaum dient als hierarchisches Regelsystem, um Objekte datenbasiert zu klassifizieren. In Abb. 1 ist ein Entscheidungsbaum abgebildet zum Klassifizieren von Lebensmitteln als „eher empfehlenswert“ oder „eher nicht empfehlenswert“ anhand von Nährwertangaben pro 100 g.

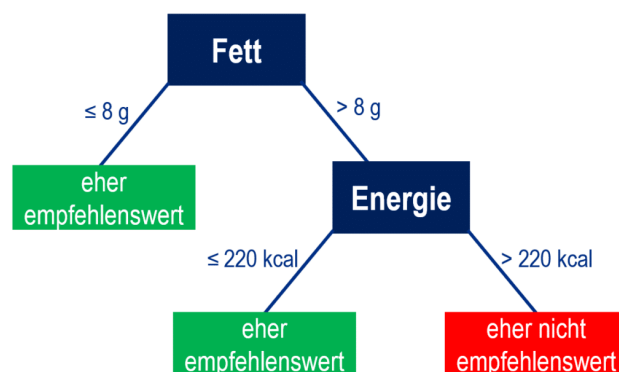


Abb. 1: Entscheidungsbaum zum Klassifizieren von Lebensmitteln

Im Unterricht werden 55 Datenkarten genutzt, die die typischen Nährwertangaben von Lebensmitteln enthalten (z. B. Apfel in Abb. 2). Ziel ist es, einen Baum zu entwickeln, der die vorliegenden Lebensmittel basierend auf ihren Nährwertangaben möglichst fehlerfrei klassifiziert. So ein Empfehlungssystem wird als Klassifikator bezeichnet, wobei die Nährwerte als Prädiktormerkmale dienen, um die Empfehlung als Zielmerkmal vorherzusagen. Lernende arbeiten mit 40 Karten zur Entwicklung des Baums (Training) und 15 Karten zum Testen. Ein einstufiger Entscheidungsbaum wird schrittweise basierend auf Trainingsdate entwickelt, indem ein Prozess mit den folgenden 4 Komponenten durchlaufen wird, der in Abb. 2 mit 11 Beispielkarten (Trainingsdaten) veranschaulicht ist:

- Datensplit durchführen: Ein Prädiktormerkmal (z. B. Fett) und ein Schwellenwert (z. B. 8 g) teilen die Daten in zwei Teildatensätze
- Entscheidungsregel erstellen: Innerhalb der Teildatensätze wird nach dem Mehrheitsprinzip eine Entscheidung getroffen. In Abb. 2 führt dies zu der Regel: "Wenn Fett kleiner oder gleich 8g pro 100g, dann eher empfehlenswert. Wenn Fett größer 8g pro 100g, dann eher nicht empfehlenswert"
- Evaluation der Entscheidungsregel: Die Regel wird evaluiert, indem die Anzahl falsch klassifizierter Objekte (Fehlklassifikationen) für beide Teildatensätzen ermittelt wird. In Abb. 2 sind es zwei (Avocado, Spiegelei).
- Darstellung des Entscheidungsbaums: Der resultierende Entscheidungsbaum wird verbal oder als Baumdiagramm dargestellt.

Ein einstufiger Entscheidungsbaum kann erweitert werden, indem der Teildatensatz mit Fehlklassifikationen weiter aufgeteilt wird, z. B. durch einen zweiten Datensplit mit dem Merkmal Energie (siehe Abb. 1). So entsteht ein mehrstufiger Baum, der den Datensatz immer fehlerfreier klassifiziert.

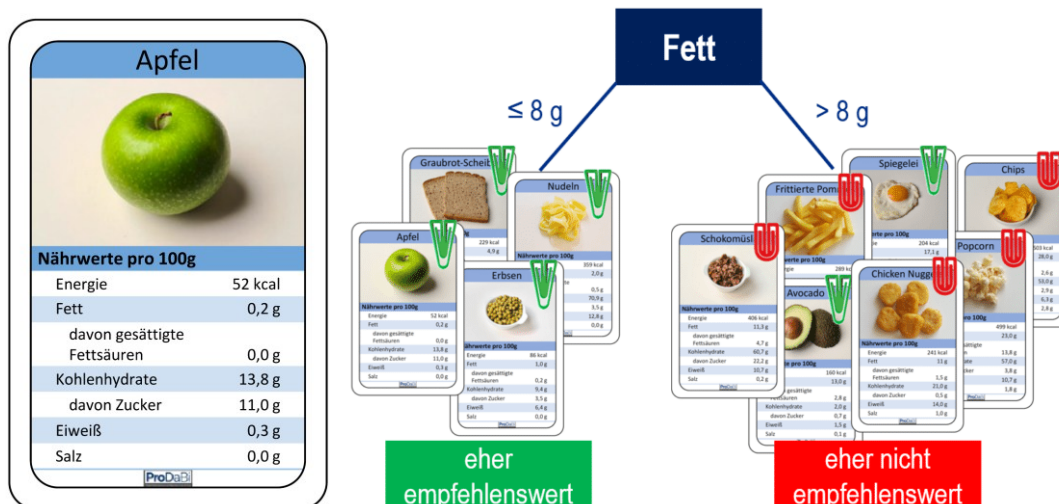


Abb. 2: Datensplit mit Datenkarten über Lebensmittel

Die Wahl eines optimalen Datensplits wird durchgeführt, indem für jedes Prädiktormerkmal der Datensplit ermittelt wird, der die geringste Anzahl an Fehlklassifikationen für Trainingsdaten liefert. Eine solch systematische Methode ist in professionellen Entscheidungsbaum Algorithmen implementiert. Anschließend wird ein so erstellter Baum mit Testdaten überprüft. Im Unterricht wird dies zunächst durch einfache Heuristiken vereinfacht dargestellt, um Lernenden die Grundlagen des maschinellen Lernens und der datenbasierten Entscheidungsfindung altersgerecht näherzubringen.

### **3. Studien zum Einsatz im Unterricht**

Im Rahmen von drei Studien wurde untersucht, ob das Unterrichtsdesign zugänglich ist für Klasse 6, welche inhaltlichen Hürde auftreten und wie Lernende eine konkrete Heuristik zur Erstellung datenbasierter Entscheidungsbäumen durchführen. Die Studien untersuchen Schülerinnen und Schüler, nachdem sie an einer achtstündigen Unterrichtsreihe teilgenommen haben.

233 Schülerinnen und Schüler nahmen anschließend an einer Umfrage teil. Eine Aufgabe darin war einen Entscheidungsbaum zur Klassifikation eines Objekts zu nutzen und die Entscheidungen entsprechend zu begründen. Die große Mehrheit der Lernenden war in der Lage dazu (Podworny et al., 2022).

In einer Interviewstudie im Anschluss an die Unterrichtsreihe untersuchten wir die Ansätze von vierzehn Lernenden zur Erstellung eines einstufigen datenbasierten Entscheidungsbaums mit Datenkarten (Fleischer et al., 2024). Alle Lernenden konnten einen Baum erstellen, indem sie Datensplits durchführten und Entscheidungsregeln per Mehrheitsentscheidung aufstellten; sechs von sieben Paaren gelang es, einen optimalen Entscheidungsbaum gemessen an der minimalen Fehlklassifikationsrate zu erstellen. Unterschiede waren in den Handlungen mit den Datenkarten festzustellen, sowie in der Argumentation zur Bewertung der Entscheidungsbäume und zur Auswahl von Entscheidungsregeln. Die meisten Schülerinnen und Schüler nutzten als Strategie die Datenkarten zu sortieren, während andere die Datenkarten wiederholt unterschiedlich gruppieren. Argumentationen zur Auswahl bestimmter Regeln beinhaltete zumeist die Minimierung der Fehlklassifikationen, während einige auch auf informelle Argumentationen zurückgriffen, um Datensplits basierend auf ihrem visuellen Eindruck zu bewerten.

Eine weitere Interviewstudie (Podworny et al., accepted) im Anschluss an die Unterrichtsreihe, die sich besonders mit dem Erstellen mehrstufiger Entscheidungsbäume auseinandersetzt ist aktuell im Veröffentlichungsprozess.

Die Ergebnisse der drei Studien zeigen zum einen, dass junge Lernende grundlegende Konzepte von datenbasierten Entscheidungsbäumen erfolgreich erlernen und anwenden können und geben zum anderen Hinweise auf

Schwierigkeiten, die für die Weiterentwicklung des Unterrichtsdesigns wichtig sind. Im Vortrag werden wir ausgewählte Ergebnisse der drei Studien vorstellen, sowie die Designideen unserer Unterrichtsreihe präsentieren.

## Literatur

- Biehler, R., & Fleischer, Y. (2021). Introducing students to machine learning with decision trees using CODAP and Jupyter Notebooks. *Teaching Statistics*, 43(S1), 133–142. <https://doi.org/10.1111/test.12279>
- Casal-Otero, L., Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K-12: A systematic literature review. *International Journal of STEM Education*, 10(1), 29. <https://doi.org/10.1186/s40594-023-00418-7>
- Engel, J. (2017). Statistical Literacy for Active Citizenship: A Call for Data Science Education. *Statistics Education Research Journal*, 16(1), 44–49. <https://doi.org/10.52041/serj.v16i1.213>
- Fleischer, Y., Biehler, R., & Schulte, C. (2022). Teaching and Learning Data-Driven Machine Learning with Educationally Designed Jupyter Notebooks. *Statistics Education Research Journal*, 21(2), 7. <https://doi.org/10/gqv59w>
- Fleischer, Y., Podworny, S., & Biehler, R. (2024). Teaching and learning to construct data-based decision trees using data cards as the first introduction to machine learning in middle school. *Statistics Education Research Journal*, 23(1), 3. <https://doi.org/10.52041/serj.v23i1.450>
- Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. <https://doi.org/10/ghbz2q>
- Podworny, S., Fleischer, Y., & Hüsing, S. (2022). Grade 6 Students' Perception and Use of Data-Based Decision Trees. *Bridging the Gap: Empowering and Educating Today's Learners in Statistics. Proceedings of the Eleventh International Conference on Teaching Statistics*. ICOTS 11, Rosario, Argentina. <https://doi.org/10.52041/iase.icots11.T2H3>
- Podworny, S., Fleischer, Y., & Biehler, R. (accepted). Young students' engagement with data to create decision trees
- Ridgway, J. (2016). Implications of the Data Revolution for Statistics Education. *International Statistical Review*, 84(3), 528–549. <https://doi.org/10/f3q6f6>
- Schüller, K., Koch, H., & Rampelt, F. (2021). Data Literacy Charta. Berlin: Stifterverband. [https://www.stifterverband.org/sites/default/files/data-literacy-charta\\_v1\\_2.pdf](https://www.stifterverband.org/sites/default/files/data-literacy-charta_v1_2.pdf)
- Sulmont, E., Patitsas, E., & Cooperstock, J. R. (2019). Can You Teach Me To Machine Learn? In E. K. Hawthorne, M. A. Pérez-Quñones, S. Heckman, & J. Zhang (Eds.), *Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (pp. 948–954). ACM. <https://doi.org/10/ghnpcs>