

# Enaktive Bestimmung der Hyperparameter beim Entscheidungsbaum- und k-nächste-Nachbarn-Algorithmus

## Vorstellung von Materialien eines KI-Experimentierkastens

Silvia Joachim <sup>1</sup> und Martin Hennecke <sup>2</sup>

**Abstract:** Für den Bereich der Künstlichen Intelligenz werden momentan zahlreiche Unterrichtsmaterialien veröffentlicht. Viele davon konzentrieren sich dabei auf den Teilbereich des Maschinellen Lernens. Die hier vorgestellten Unterrichtsmaterialien gehören ebenfalls in diese Kategorie – sie sprechen allerdings gezielt verschiedene Repräsentationsebenen an und sollen durch die haptische Zugänglichkeit auch Menschen mit einer Sehbeeinträchtigung den Zugang erleichtern. Bei den in diesem Beitrag vorgestellten Materialien handelt es sich um aus Holz gefertigte Pilze, die sich u.a. für die Erarbeitung des k-nächste-Nachbarn- und Entscheidungsbaum-Algorithmus eignen. Die Pilze verfügen dazu über binäre und metrische Merkmale. Der zugrunde liegende Datensatz ist bewusst klein und wurde so gewählt, dass der Einfluss von Hyperparametern gezeigt werden kann. Er enthält Validierungsdaten, die die Optimierung der Baumtiefe und der Anzahl der Nachbarn ermöglichen. Der Datensatz kann für weitere Algorithmen verwendet werden und ermöglicht somit eine effiziente Unterrichtsgestaltung.

**Keywords:** Unterrichtsmaterial, Experimentierkasten, Künstliche Intelligenz, k-nächste-Nachbarn-Algorithmus, Entscheidungsbaum-Algorithmus, Validierungsdaten, Hyperparameter, enaktiv


## 1 Einleitung

In der Öffentlichkeit wird das Thema Künstliche Intelligenz immer präsenter. Dies spiegelt sich auch in ersten Lehrplänen an allgemeinbildenden Schulen wider. So sieht beispielsweise der bayerische LehrplanPLUS für die Gymnasien in der Jahrgangsstufe 11 den Einsatz des k-nächste-Nachbarn- oder des Entscheidungsbaum-Algorithmus vor [ISB23a] [ISB23b]. Da das Fach Informatik in Bayern Pflichtfach ist muss daraus der Anspruch für eine inklusive Medienbildung abgeleitet werden, d.h. Unterrichtsmaterialien müssen z.B. auch für Schülerinnen und Schüler mit Sehbeeinträchtigung zugänglich sein.

Die Arbeit mit haptischen Unterrichtsmaterialien hat sich als Einstieg in neue Unterrichtsthemen in der Schulpraxis bewährt. Dieser Idee folgend stellt dieser Beitrag, real erfahr-, begreif- und fühlbare Unterrichtsmaterialien in Form aus Holz gefertigter Pilze vor. Sie

---

<sup>1</sup> Universität Würzburg, Didaktik der Informatik, Emil-Fischer-Str. 30, 97074 Würzburg, silvia.joachim@uni-wuerzburg.de,  <https://orcid.org/0009-0004-8095-0133>

<sup>2</sup> Universität Würzburg, Didaktik d. Informatik, Emil-Fischer-Str. 30, 97074 Würzburg, martin.hennecke@uni-wuerzburg.de,  <https://orcid.org/0000-0001-5586-0647>

verfügen über binäre und metrische Merkmale und sind daher für verschiedene Algorithmen des Maschinellen Lernens nutzbar, insbesondere für den k-nächste-Nachbarn-Algorithmus oder das Maschinelle Lernen von Entscheidungsbäumen. Die Arbeit mit den Pilzen konfrontiert die Schülerinnen und Schüler mit dem Problem, Pilze in „giftig“ (X) oder „essbar“ (O) zu klassifizieren. Die Bedeutung der Fragestellung wird dabei durch die Bedeutung der Entscheidung betont und eröffnet für eine anschließende kritische Reflexion der Vorgehensweise zahlreiche Ansätze für die Diskussion.



Abb. 1: Pilze mit sicht- und fühlbaren Merkmalen (Bildquelle: MEKRUPHY GMBH)

Als binäre Merkmale der Pilze sind Muster (Punkte, Streifen), Farbe (blau, rot), Randlinie (gerade, gewellt) und Kragen (ja, nein) ausgeprägt. Die metrischen Merkmale  $x$  (0 bis 12) und  $y$  (0 bis 8) geben den Standort der Pilze auf einem Koordinatensystem an. 14 dieser Datensätze sind als „giftig“ oder „essbar“ gelabelt und zwei stehen zur Klassifizierung zur Verfügung. Die Symbole oben im Pilzkopf geben zudem eine Unterteilung in Trainings-, Validierungs- und Testdaten vor. Der durch die Pilze kodierte Datensatz ermöglicht eine Optimierung der jeweiligen Modelle, ist aber dennoch klein genug um die Algorithmen auch enaktiv durchführen und manuell berechnen zu können.

## 2 Entscheidungsbaum-Algorithmus

Für den Entscheidungsbaum-Algorithmus [Er21] gibt es bereits ansprechende Materialien. So können etwa die „beißen“ oder „nicht beißen“ Äffchen von Seegerer, Lindner und Romeike [SLR19] genutzt werden. Die Arbeit mit den Pilzen ist grundsätzlich ähnlich gestaltbar. Die binären Merkmale Muster, Farbe (bzw. Form), Randlinie und Kragen erleichtern jedoch die direkte Nutzbarkeit der Daten beim Erlernen von binären Entscheidungsbäumen. Die gegebene Unterteilung in Trainings-, Validierungs- und Testdaten ist so gewählt, dass im Anschluss eine Optimierung des Hyperparameters Baumtiefe sinnvoll möglich ist. Die Pilze sind dabei nicht (nur) alternativ zu den Äffchen nutzbar, sondern können nach einer stärker spielerischen Begegnung mit den Äffchen auch für eine weitere Formalisierung der algorithmischen Idee des Lernalgorithmus eingesetzt werden.

Die fühlbaren binären Merkmale der Pilze sollen auch sehbeeinträchtigten Menschen einen Zugang zum Entscheidungsbaum-Algorithmus ermöglichen. Eine Platte mit Aussparungen für Wurzel, Knoten und Kantenbeschriftungen des Entscheidungsbaums ermöglicht eine graphische Repräsentation des Entscheidungsbaums, die für Schülerinnen und Schüler mit Sehbeeinträchtigung zugänglich ist. Die Pilze können immer wieder neu nach den Merkmalen Muster, Farbe (bzw. Form), Randlinie und Kragen sortiert werden, um das Merkmal mit dem größten Informationsgewinn zu ermitteln. Das anschauliche Ziel dabei ist, möglichst gut zwischen „essbaren“ und „giftigen“ Pilzen unterscheiden zu können.

Durch das im Folgenden beschriebene Zählen von Pilzen findet eine didaktische Reduktion beim Maschinellen Lernen eines Entscheidungsbaums statt. Die Schülerinnen und Schüler können anhand der acht Trainingsdaten (Pilze mit Dreieck) zunächst feststellen, dass es unter ihnen genauso viele „giftige“ wie „essbare“ Pilze gibt. Werden die Pilze nach dem Merkmal Muster sortiert, so sind sowohl unter den Pilzen mit Punkten als auch unter denen mit Streifen wieder genauso viele „essbare“ wie „giftige“ Pilze. Gleiches gilt für eine Sortierung nach den Merkmalen Farbe und Kragen. Bei einer Sortierung der Trainingsdaten nach dem Merkmal Randlinie können die Schülerinnen und Schüler feststellen, dass in der Gruppe mit gewellter Randlinie drei Pilze „essbar“ und zwei „giftig“ und in der Gruppe mit gerader Randlinie einer „essbar“ und zwei „giftig“ sind. Als Wurzel des Entscheidungsbaumes erhalten sie so das Merkmal Randlinie. Pilze mit gewellter Randlinie würden per Mehrheitsentscheidung als „essbar“ und Pilze mit gerader Randlinie als „giftig“ klassifiziert werden und ergäben so einen Entscheidungsbaum der Tiefe 1. Nach dem Divide-and-Conquer-Ansatz werden die Pilze danach entsprechend der Randlinie getrennt betrachtet (usw.).

Ein beim Entscheidungsbaum interessanter Hyperparameter ist dessen Tiefe. Bei großen Datensätzen kann Pruning vorteilhaft sein. Mit Hilfe der Validierungsdaten kann, wie bei Russel/Norvig [RN22] für einen großen Datensatz beschrieben, die optimale Baumtiefe gewählt werden, indem die minimale Fehlerrate für die Validierungsmenge bestimmt wird. Um dies nachvollziehbar zu machen, wurde der zugrunde liegende Datensatz so gestaltet, dass die Vorteile von Pruning nachvollzogen werden können.

### **3 k-nächste-Nachbarn-Algorithmus**

Der k-nächste-Nachbarn-Algorithmus lässt sich mit den metrischen Merkmalen umsetzen. Diese geben die Koordinaten der Fundorte der Pilze an und können in ein Koordinatensystem eingetragen werden. Die Klassifikation anhand der Koordinaten folgt der Idee, dass Pilze einer Art häufig in kleinen Gruppen beieinanderstehen. Beim Eintragen der Pilze lassen sich die giftigen und die essbaren Pilze z. B. durch die Symbole X und O visualisieren. Schülerinnen und Schüler mit einer Sehbeeinträchtigung können die Fundorte der Pilze räumlich auf einem Koordinatensystem ertasten. Entfernungen sind mit der Manhattan-Metrik erfassbar.

Anschließend kann nacheinander der Fundort der drei Validierungspilze betrachtet und der k-nächste-Nachbarn-Algorithmus (k ungerade) mit euklidischer Norm angewendet werden. D.h. es werden die Abstände von jeweils einem Validierungspunkt zu allen Trainingsdatenpunkten mit dem Lineal gemessen und die k-nächsten Nachbarn bestimmt. Dabei wird der Parameter k bewusst ungerade gewählt, so dass hier immer eine Mehrheitsentscheidung möglich ist. Da es 8 Trainingsdaten gibt, wird k=1, k=3, k=5 und k=7 betrachtet. Jeder Validierungspilz wird dann, entsprechend der Mehrheit der k-nächsten Nachbarn, klassifiziert und das Ergebnis mit seinem Label verglichen. Für k=5 werden alle Validierungs- und Testdaten richtig klassifiziert. Von den beiden unbekannt Pilzen wird einer als „essbar“ und einer als „giftig“ klassifiziert. Hier liefert der k-nächste-Nachbar-Algorithmus didaktisch sinnvoll ein anderes Ergebnis als der Entscheidungsbaum-Algorithmus mit binären Merkmalen; schließlich wurden ja sogar andere Merkmale betrachtet. Die Trefferquote ist maximal und bietet eine Möglichkeit zur Diskussion über das Vertrauen in KI-Systeme.

## 4 Fazit

Der durch Speise- bzw. Giftpilze gegebene Realitätsbezug macht unmittelbar klar, wie wichtig eine korrekte Klassifizierung ist. Die sich durch die unterschiedlichen Algorithmen ergebenden unterschiedlichen Klassifizierungen der unbekannt Pilzsarten bieten eine Möglichkeit zur Diskussion über die Qualität von Daten und die Vertrauenswürdigkeit von KI-Systemen, wie etwa einer Smartphone-App zur Pilzbestimmung. Die hier vorgestellten Pilze sind Teil eines Experimentierkastens, der in zahlreichen Lehrerfortbildungen getestet wurde und dort auf sehr positive Resonanz gestoßen ist.

## Literaturverzeichnis

- [Er21] Ertel, W.: Grundkurs Künstliche Intelligenz. Eine praxisorientierte Einführung, 5. Auflage, Springer Fachmedien Wiesbaden, 2021.
- [ISB23a] Staatsministerium für Schulqualität und Bildungsforschung München, LehrplanPLUS Gymnasium Bayern, Informatik 11 (NTG), <https://www.lehrplanplus.bayern.de/fachlehrplan/gymnasium/11/informatik/ntg>, Stand: 14.05.2023.
- [ISB23b] Staatsministerium für Schulqualität und Bildungsforschung München, LehrplanPLUS Gymnasium Bayern, spätbeginnende Informatik 11, [https://www.lehrplanplus.bayern.de/fachlehrplan/gymnasium/11/informatik/mug\\_swg\\_sg](https://www.lehrplanplus.bayern.de/fachlehrplan/gymnasium/11/informatik/mug_swg_sg), Stand: 14.05.2023.
- [RN22] Russel, S.; Norvig, P.: Artificial Intelligence. A Modern Approach, 4. Auflage, Pearson, United Kingdom, 2022.
- [SLR19] Seegerer, S.; Lindner, A; Romeike, R.: AI Unplugged – Wir ziehen Künstlicher Intelligenz den Stecker. In (Pasternak, A., Hrsg.): Informatik für alle. Bonn: Gesellschaft für Informatik. S. 325-334. DOI: 10.18420/infos2019-c18