

Zielgruppe A: 10–12 Jahre (Mittelstufe)

Thema: Wie lernt ein Computer?

1. Datenerhebung

Sammle insgesamt 12 Datenpunkte (6 Erdbeeren, 6 Tomaten). Achte darauf, im Modus **Gewicht & Volumen** zu messen.

- **Beobachtung:** Vergleiche die Messwerte in der Tabelle. Was fällt dir bei den Zahlen auf, wenn du eine kleine Erdbeere mit einer großen Tomate vergleichst?

2. Modellbildung

Klicke auf „Modell trainieren“. Schau dir den Entscheidungsbaum (5) an.

- **Aufgabe:** Notiere den „Grenzwert“, den die KI für das Gewicht festgelegt hat.
- **Verständnis:** Warum hat die KI diesen Wert gewählt und nicht z. B. 5 oder 95?

3. Überprüfung (Testlauf)

Gib im Testlauf Werte ein, die „zwischen“ einer typischen Erdbeere und einer Tomate liegen (z. B. Gewicht 45g).

- **Frage:** Hat die KI recht? Diskutiere kurz: Kann eine KI auch Fehler machen, wenn die Daten nicht eindeutig sind?
-

Zielgruppe B: 15–18 Jahre (Oberstufe)

Thema: Klassifikationsalgorithmen und Feature-Engineering

1. Analyse der Trennschärfe

Führe Messreihen in allen drei Modi (A, B und C) durch.

- **Hypothese:** Welches Feature-Paar (z. B. Zucker/Gewicht vs. Gewicht/Volumen) erlaubt die präziseste Trennung der Klassen? Begründe deine Wahl anhand der Überlappungen im Diagramm.

2. Funktionsweise des Algorithmus

Das Programm nutzt eine Heuristik zur Bestimmung der *Decision Boundaries*.

- **Analyse:** Betrachte die gestrichelten Linien im Diagramm. Wie verändern sich diese Linien, wenn du gezielt „Ausreißer“ (sehr große Erdbeeren oder sehr kleine Tomaten) hinzufügst und neu trainierst?

3. Evaluation & Transfer

- **Kritik:** Der vorliegende Baum hat eine Tiefe von 2. Erkläre, warum ein tieferer Baum (mehr Verzweigungen) bei diesem Datensatz zu *Overfitting* führen könnte.
 - **Export:** Exportiere deine Daten als CSV und skizziere, wie die mathematische Bedingung für den ersten „Split“ (IF-Bedingung) lauten könnte, wenn man den Mittelwert beider Klassen als Grundlage nimmt.
-

Lehrer-Infoblatt: Die Mathematik hinter dem "KI-Labor"

1. Der Algorithmus: Geometrische Klassifikation

Die App verwendet einen vereinfachten **Entscheidungsbaum-Algorithmus (Decision Tree)**. Im Gegensatz zu komplexen Verfahren wie *Random Forests*, die auf Entropie-Berechnungen (Gini-Index) basieren, nutzt dieses Tool eine intuitive **Schwerpunkt-Heuristik**.

Der erste Split (Feature 1)

Das Modell berechnet die arithmetischen Mittelwerte des ersten Merkmals für beide Klassen:

$$\overline{x_{Erdbeere}} = \frac{1}{n} \cdot \sum x_{Erdbeere} \quad \text{und} \quad \overline{x_{Tomate}} = \frac{1}{m} \cdot \sum x_{Tomate}$$

Die Grenze (Decision Boundary) wird exakt in der Mitte zwischen diesen beiden Mittelwerten gezogen:

$$Split = \frac{\overline{x_{Erdbeere}} + \overline{x_{Tomate}}}{2}$$

2. Die Visualisierung im 2D-Raum

Da wir zwei Merkmale messen, arbeitet die KI in einem **zweidimensionalen Merkmalsraum**. Jedes Objekt ist ein Vektor $v = (f_1, f_2)$.

- **Split 1** teilt den Raum vertikal (entlang der x-Achse).
- **Split 2** teilt den verbleibenden Raum horizontal (entlang der y-Achse).

Dadurch entstehen drei rechteckige Regionen, die im Scatterplot farbig hinterlegt sind. Dies demonstriert das Konzept der **linearen Trennbarkeit**.

3. Didaktische Hinweise zum Modell-Verhalten

- **Sensitivität:** Da Mittelwerte verwendet werden, reagiert das Modell empfindlich auf Ausreißer. Ein einziger extrem schwerer Datenpunkt verschiebt die Trennlinie sofort.
- **Modell-Güte:** Schüler können sehen, dass das Modell "perfekt" trainiert sein kann (alle Trainingspunkte liegen in der richtigen Farbzone), aber beim "Testlauf" trotzdem scheitert, wenn die neuen Testdaten untypisch sind. Dies führt zum Thema **Generalisierung vs. Overfitting**.
- **Feature-Wahl:** Im Modus "Gewicht & Volumen" korrelieren die Daten stark (beide steigen linear an). In anderen Modi sind die Cluster eventuell klarer getrennt. Dies zeigt den Schülern, dass die Wahl der richtigen Sensoren (Features) entscheidender sein kann als der Algorithmus selbst.