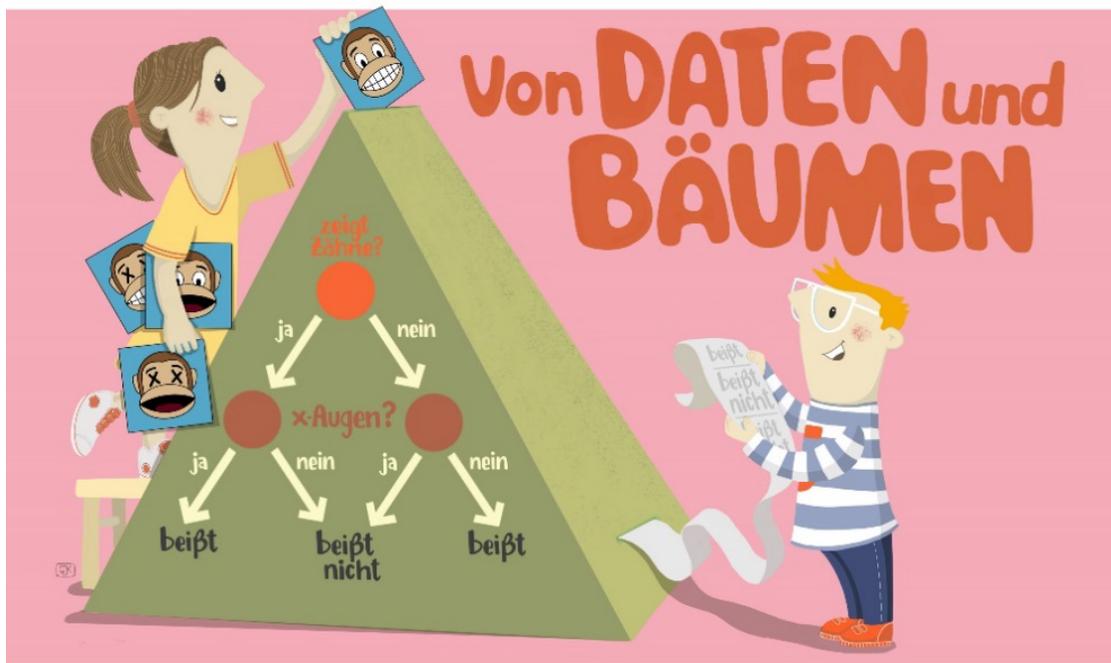


IT2School

Gemeinsam IT entdecken



KI-B4 Von Daten und Bäumen

Daten mit KI selbst auswerten

Eine Entwicklung in Kooperation von:

Inhalt

1	Von Daten und Bäumen	3
2	Warum gibt es das Modul?	4
3	Ziele des Moduls.....	4
4	Die Rolle der Unternehmensvertreterin/des Unternehmensvertreters.....	5
5	Inhalte des Moduls.....	5
1.	Klassifikation	5
2.	Over- und Underfitting	8
3.	Bewertung der Güte eines Modells	9
4.	Bias	11
5.	Orange3.....	12
6	Unterrichtliche Umsetzung.....	14
6.1	Grober Unterrichtsplan	15
6.2	Stundenverlaufsskizzen.....	16
6.2.1	Stunde 1: Datenanalyse und Entscheidungsbaum	16
6.2.2	Stunde 2-3: Datenanalyse mit Orange3	19
6.2.3	Stunde 4-6: Projektarbeit (nur Variante II).....	22
6.2.4	Stunde 7: KI und unsere Gesellschaft.....	24
7	Einbettung in verschiedene Fächer und Themen	25
8	Anschlussthemen.....	25
9	Literatur und Links	26
10	Arbeitsmaterialien	26
11	Glossar.....	27
12	Fragen, Feedback, Anregungen.....	28

1 Von Daten und Bäumen

In diesem Aufbaumodul wenden die Schülerinnen und Schüler KI-Methoden (vorwiegend das Lernen von Entscheidungsbäumen) aktiv auf **Daten** an und analysieren die gesellschaftliche Wirkung von KI-Systemen. Aufbauend auf den niederschweligen Beispielen, die auch im Modul "Schlag den Roboter" genutzt wurden, schlüpfen die Schülerinnen und Schüler in die Rolle eines Data Scientist und erzeugen mithilfe der freien Software Orange3 Modelle, um datenbasiert Entscheidungen abzuleiten oder Vorhersagen zu treffen. Die Schülerinnen und Schüler setzen damit nicht nur KI-Methoden aktiv ein, sondern lernen auch Berufsperspektiven wie die des Data Scientist kennen und diskutieren die gesellschaftliche Wirkung, die aus dem Einsatz ihrer Modelle resultieren würde.

Lernfeld/Cluster:	Mit KI gestalten
Zielgruppe/Klassenstufe:	4. bis 5. Klasse
	6. bis 7. Klasse
	X 8. bis 10. Klasse
	X 11. bis 12. Klasse
Geschätzter Zeitaufwand:	2 bis 8 Unterrichtsstunden, auch als Projekttag umsetzbar
Lernziele:	<p>Nach Abschluss des Teilmoduls können die Schülerinnen und Schüler...</p> <ul style="list-style-type: none"> • Probleme erläutern, in denen KI gewinnbringend eingesetzt werden kann (Welche Probleme werden bereits mit KI gelöst?). • Probleme unter Nutzung von KI-Methoden lösen (Wie kann ich KI nutzen, um Probleme zu lösen?). • Konsequenzen aus dem Einsatz von KI-Methoden diskutieren und bewerten. (Welche Konsequenzen hat der Einsatz von KI-Methoden auf uns und unsere Gesellschaft? Wo liegen die Gefahren von KI?) • Einsatzmöglichkeiten von KI-Methoden als "Werkzeug" bspw. bei der beruflichen Tätigkeit beschreiben (Wie arbeitet ein Data Scientist? Was kann ich mit KI später (beruflich) machen?).
Vorkenntnisse der Schülerinnen und Schüler:	<p>Empfohlen:</p> <ul style="list-style-type: none"> • Basismodul „Schlag den Roboter“
Vorkenntnisse der/des Lehrenden:	<p>Empfohlen:</p> <ul style="list-style-type: none"> • Basismodul „Schlag den Roboter“
Vorkenntnisse der Unternehmensvertreterin/des Unternehmensvertreters:	<p>Empfohlen:</p> <ul style="list-style-type: none"> • Beispiele für KI-Anwendungen aus dem eigenen Unternehmen benennen und erläutern können
Sonstige Voraussetzungen:	<ul style="list-style-type: none"> • Zugang zu Computern mit installierter Software Orange3 (kann ohne Administratorrechte installiert werden)

2 Warum gibt es das Modul?

Viele Anwendungen aus unserem Alltag basieren auf dem maschinellen Lernen, wie etwa Film- oder Musik-Streaming-Dienste. Diese Dienste sammeln Daten der Nutzerinnen und Nutzer, um ihre Vorlieben kennenzulernen, sie zu kategorisieren und um daraus passende Empfehlungen abzuleiten. Die Programme lernen durch die vielen Nutzerdaten, welche Inhalte sie welchen Nutzern vorschlagen können. Auch in anderen Bereichen, wie etwa im Gesundheitssektor zur besseren Erkennung von Krankheiten, wird das maschinelle Lernen eingesetzt. Zentral für maschinelles Lernen ist die Nutzung und Verwertung großer Mengen von Daten. Mit Hilfe von Beispieldaten lernt die Maschine und wird dann mit der Zeit und dem Zugriff auf noch mehr Daten, immer besser in ihren Berechnungen.

In diesem Modul wird der Fokus auf genau diese Daten gelegt, um das Phänomen des maschinellen Lernens zu erläutern. Diese Herangehensweise ermöglicht es auch Lernern ohne Programmiererfahrung, KI-Methoden zu verstehen und anzuwenden und die Hintergründe aktueller Fragestellungen gesellschaftlicher Tragweite selbst zu entdecken. Die Schülerinnen und Schüler haben darüber hinaus die Möglichkeit, realistische Szenarien aus dem Alltag zu diskutieren und einen Eindruck davon zu gewinnen, wie KI ihren eigenen Berufsweg beeinflussen kann. Durch das Verstehen der Funktionsweise, erhalten die Kinder und Jugendlichen einen kritischen Blick auf solche Systeme. Sie können ihr Verhalten und Nutzungsgewohnheiten reflektieren und werden in die Lage versetzt, selbstbestimmte Entscheidungen zu treffen, die nicht allein durch Maschinen gelenkt werden.

Durch den Fokus auf Daten ergeben sich im Rahmen dieses Moduls unterschiedliche Einsatz- bzw. Anknüpfungsmöglichkeiten für eine Vielzahl von Fächern, etwa in den Naturwissenschaften oder im Wirtschaftsunterricht.

3 Ziele des Moduls

Das Modul trägt auf vielfältige Weise zu den drei Dagstuhl-Perspektiven bei:

- Aus technologischer Perspektive lernen Schülerinnen und Schüler in diesem Modul typische Herangehensweisen für die Lösung von KI-Problemen mithilfe von maschinellem Lernen kennen.
- Aus anwendungsbezogener Perspektive erfahren die Schülerinnen und Schüler vielfältige Anwendungen von KI. Außerdem wird thematisiert, wie KI für die eigene berufliche Zukunft zur Auswertung von Daten (auch im Beruf) genutzt werden kann.
- Aus gesellschaftlich-kultureller Perspektive werden Konsequenzen des Einsatzes von KI-Methoden diskutiert und unter ethischen und moralischen Gesichtspunkten bewertet.



4 Die Rolle der Unternehmensvertreterin/des Unternehmensvertreters

In diesem Modul hat die Unternehmensvertreterin bzw. der Unternehmensvertreter mehrere Möglichkeiten aktiv mitzuwirken. Hier einige Anregungen:

- Als Special Guest in der Schule über die Bedeutung von maschinellem Lernen in der Wirtschaft und insbesondere im eigenen Unternehmen berichten
- Schülerinnen und Schülern eine Exkursion in das eigene Unternehmen ermöglichen und zeigen, wie Künstliche Intelligenz in der Praxis eingesetzt wird
- Fragen & Antworten-Runden mit „KI-ExpertInnen“ oder Data Scientists des eigenen Unternehmens organisieren, die berichten, warum sie sich für ein Studium im Bereich Informatik / Data Science entschieden haben
- „Kreativworkshop“ oder „Wettbewerb“ ausrichten, wie das Gelernte im Alltag der Schülerinnen und Schüler eingesetzt werden könnte und nach bestimmten Bewertungskriterien (Kreativität, Umsetzbarkeit, Innovationsgrad...) die entwickelten bzw. eingereichten Ideen bewerten
- Unterstützung von Jugend-forscht-Projekten im Bereich KI, die sich aus dem Unterricht ergeben

5 Inhalte des Moduls

Es lassen sich drei Arten unterscheiden, wie Maschinen „lernen“ können (siehe dafür das Basismodul KI-B3 „Schlag den Roboter“). In diesem Modul soll das überwachte Lernen vertieft werden. Hier werden dem Programm beschriftete Trainingsdaten zur Verfügung gestellt, auf deren Grundlage das Programm lernt, auch weiteren Daten eine Beschriftung zuzuordnen. Die wesentlichen Begriffe und Funktionen werden im Folgenden näher erläutert.

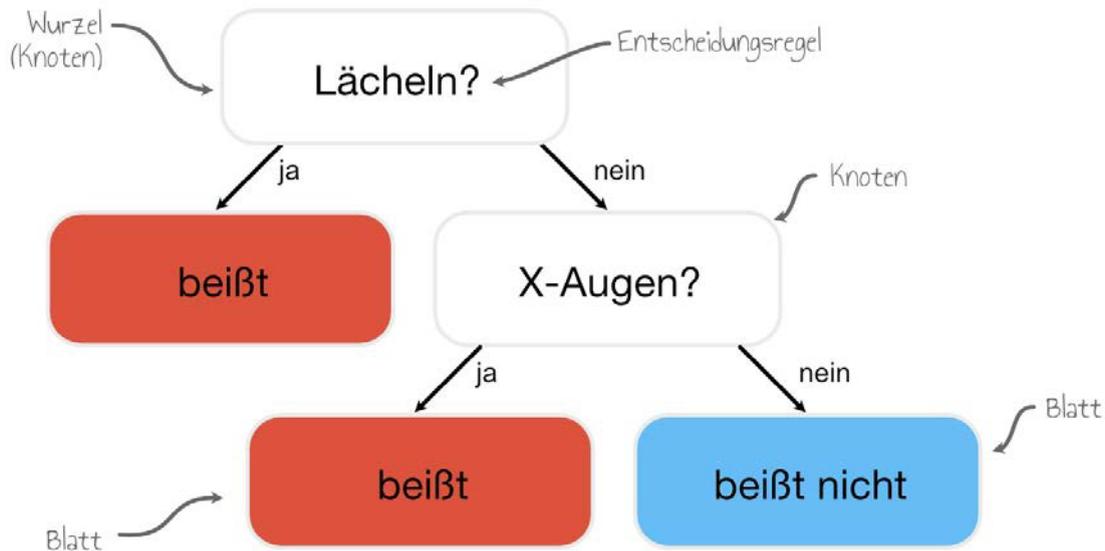
5.1 Klassifikation

Ein häufiger Einsatzzweck von überwachten Lernverfahren sind sogenannte Klassifikationsprobleme. Bei Klassifikationsproblemen lernt das Programm neuen Daten eine Kategorie zuzuordnen. Je nach Anwendungsbereich können dann beispielsweise medizinische Bilder in die Klassen „Tumor“ oder „kein Tumor“ sortiert werden oder bei der Bewertung der Bonität eines Bankkunden in die Klassen „kreditwürdig“ bzw. „nicht kreditwürdig“. Die Funktionsweise solcher Verfahren wird in diesem Modul anhand eines „Äffchen-Spiels“ verdeutlicht, wobei Äffchen aufgrund ihres Aussehens als „beißt“ oder „beißt nicht“ klassifiziert werden. In diesen Beispielen existieren jeweils nur zwei Klassen. In anderen Anwendungsfällen, z.B. bei der Erkennung des Inhalts auf Bildern können viele weitere Klassen zur Anwendung kommen, z.B. Hund, Katze, Maus, Straßenschild, Tisch, ...

Überwachte Lernverfahren erhalten eine Reihe von Trainingsdaten mit entsprechenden Beschriftungen. Sie lernen in der sogenannten Trainingsphase selbständig eine Verbindung zwischen Trainingsdaten und Beschriftung, ein sogenanntes „Modell“. In unserem Beispiel ist dieses Modell ein Entscheidungsbaum. Dieses Modell, wie z.B. der Entscheidungsbaum, kann dann auf beliebige Eingabedaten angewendet werden und die Klassen für Daten vorhersagen.

Klassifikation mittels Entscheidungsbaum

Ein häufig genutztes überwachtes Lernverfahren zur Klassifikation stellt das sogenannte Lernen von Entscheidungsbäumen (Decision tree learning) dar. Mit dem Begriff „Baum“ wird eine Datenstruktur aus der Informatik bezeichnet, mit der sich hierarchische Strukturen abbilden lassen. In einem sog. Entscheidungsbaum sind Entscheidungsregeln dargestellt.

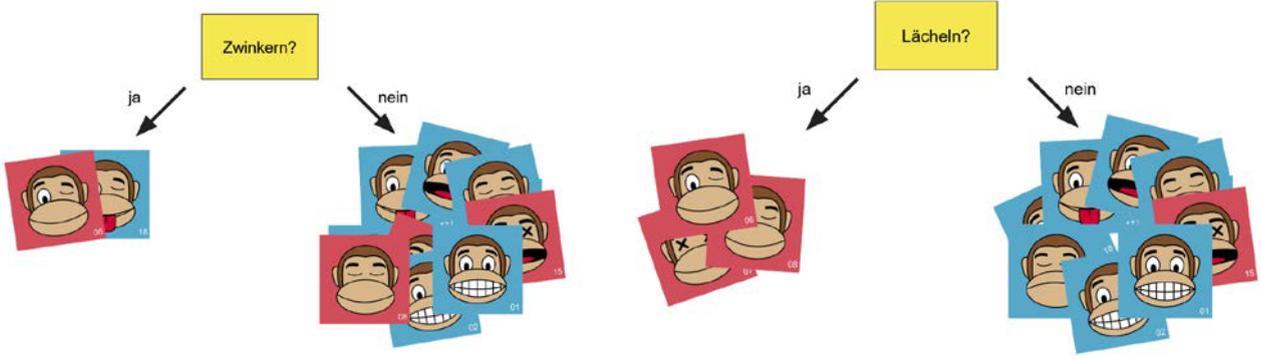
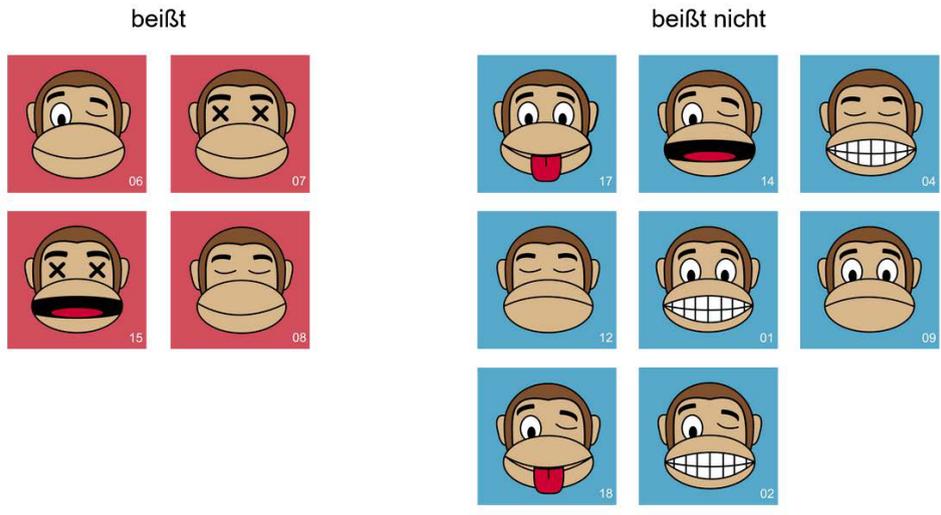


Ein **Knoten** (eine Abzweigung im Baum) repräsentiert eine zu treffende Entscheidung. Jeder Knoten kann je nach möglichen Entscheidungen mehrere Kinder (d.h. untergeordnete Knoten) haben. Der oberste Knoten wird als **Wurzel** bezeichnet. Hat ein Knoten keine Kinder, so sprechen wir von einem **Blatt**. In der Regel ergibt sich aus dem Blatt, an dem ein Datensatz nach Durchlaufen des Baumes endet, welcher Klasse er angehört. Mit Hilfe des Entscheidungsbaums können also Daten klassifiziert werden. Wenn bspw. ein Äffchen durch einen zugehörigen Entscheidungsbaum verarbeitet wird, entscheidet der letzte Knoten, nachdem alle Entscheidungsregeln ausgewertet wurden, über die Klasse, also entweder **beißt**, oder **beißt nicht**.

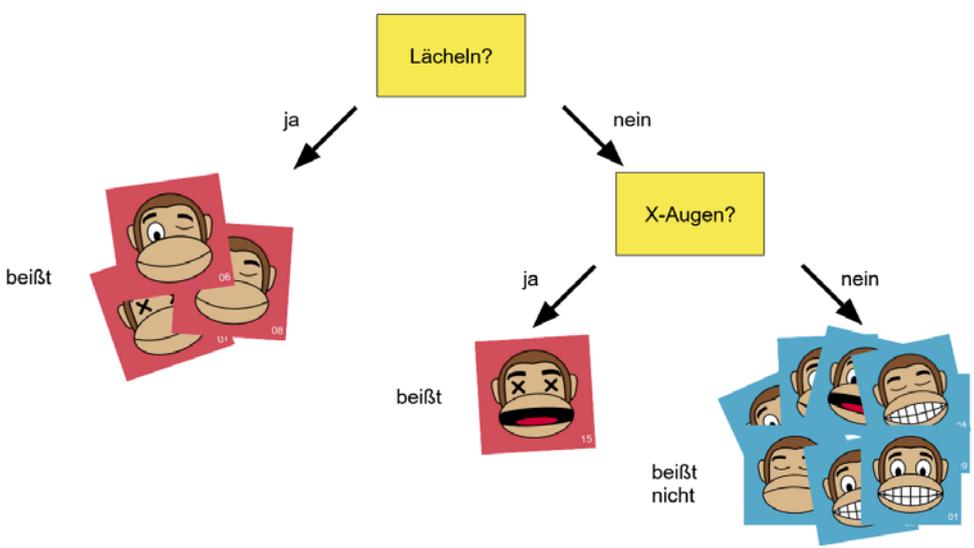
Beim Lernen von Entscheidungsbäumen werden in der Trainingsphase alle Trainingsdaten beginnend bei der Wurzel sukzessive mithilfe von Entscheidungsregeln aufgeteilt. Dabei wird stets die Regel gewählt, die am meisten Information über die zu vergebende Beschriftung liefert¹.

¹ Viele Algorithmen berechnen hierfür den sogenannten Informationsgewinn: https://de.wikipedia.org/wiki/Iterative_Dichotomiser_3

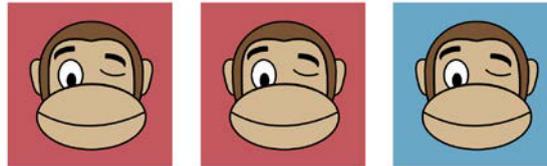




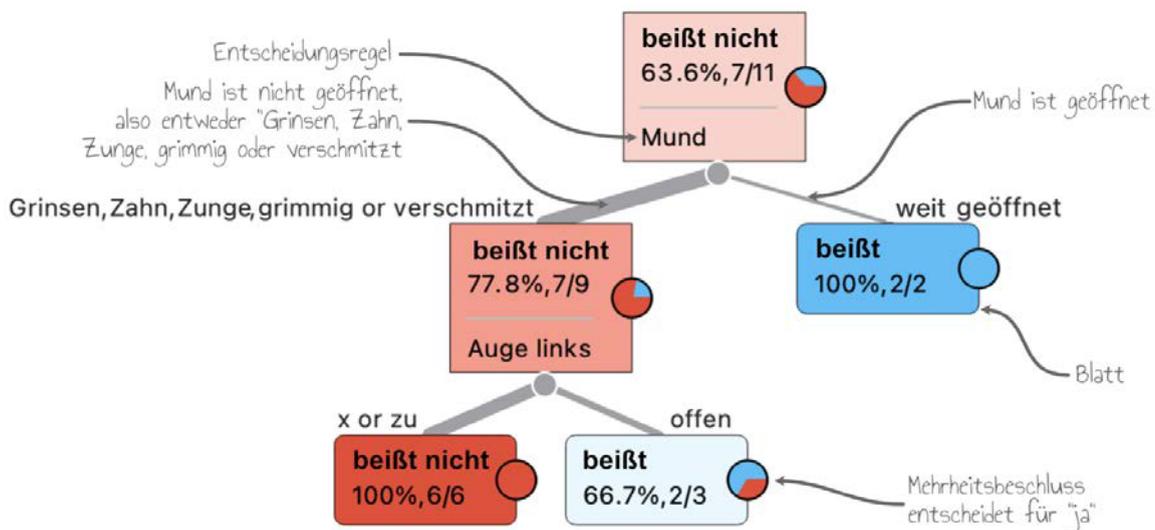
Im obigen Beispiel etwa liefert die Entscheidungsregel „Lächeln?“ deutlich mehr Information über die Beschriftung *beißt* bzw. *beißt nicht* als die Regel „Zwinkern?“. Entsprechend wählt das Lernverfahren selbstständig „Lächeln?“ als erste Entscheidungsregel aus. Dieses Vorgehen wird so lange wiederholt, bis entweder nur Äffchen mit der der gleichen Beschriftung übrig sind oder keine weiteren Merkmale zur Unterscheidung der Äffchen existieren. Der resultierende Baum für dieses Beispiel ist untenstehend abgebildet.



Gerade in realweltlichen Datensätzen und auch in diesem Beispiel, kann der Fall auftreten, dass sich die Daten in einem Knoten nicht weiter unterscheiden lassen. Betrachten wir dazu folgendes weitere Beispiel: Die Äffchen, deren Mund nicht geöffnet und deren linkes Auge geöffnet ist, unterscheiden sich nur noch bezüglich ihrer Beschriftung (beißt/beißt nicht), nicht aber in ihren Merkmalen.



In diesem Falls entscheidet der Mehrheitsbeschluss: Im Beispiel ist die Mehrheit der Äffchen beißend, daher wird die Klasse "beißt" vergeben. In der Software Orange, dem Werkzeug das in diesem Modul genutzt wird, ist das wie folgt visualisiert:



(Binärer) Entscheidungsbaum in Orange3

Von einem sogenannten binären Entscheidungsbaum sprechen wir, wenn jeder Knoten genau zwei oder keine Kinder hat.

5.2 Over- und Underfitting

Ziel des Lernprozesses ist es, ein Modell zu erhalten, das für den Anwendungsfall gute Ergebnisse liefert. Dabei kann es allerdings auch zu Problemen kommen. Beispielsweise kann es passieren, dass sich ein Modell zu stark an die Trainingsdaten anpasst und infolgedessen kaum auf neue Daten generalisieren kann. Dies könnte man auch mit „Auswendiglernen“ vergleichen. Das Programm wurde mit Hilfe der vorgegebenen Daten so trainiert, dass es nur mit den vorhandenen Daten gut funktioniert. Der Algorithmus ist aber nicht mehr in der Lage mit neuen Eingaben umzugehen. Man spricht in diesem Fall von einer Überanpassung (**Overfitting**).

Auf der anderen Seite kann es auch zu einer Unteranpassung (**Underfitting**) kommen, wenn das Modell aus den Trainingsdaten nicht genug über die zugrunde liegende Struktur der Daten lernen kann und deshalb zu stark generalisiert, sodass keine zufriedenstellenden Ergebnisse auf neuen Daten erzielt werden können. Es können aus den vorhandenen Daten einfach keine

sinnvollen Schlüsse gezogen werden, um auf Grundlage dessen neue Daten zu kategorisieren.

Um ein solches Over- bzw. Underfitting vor dem Einsatz in der Praxis aufzudecken, wird das Modell zunächst getestet.

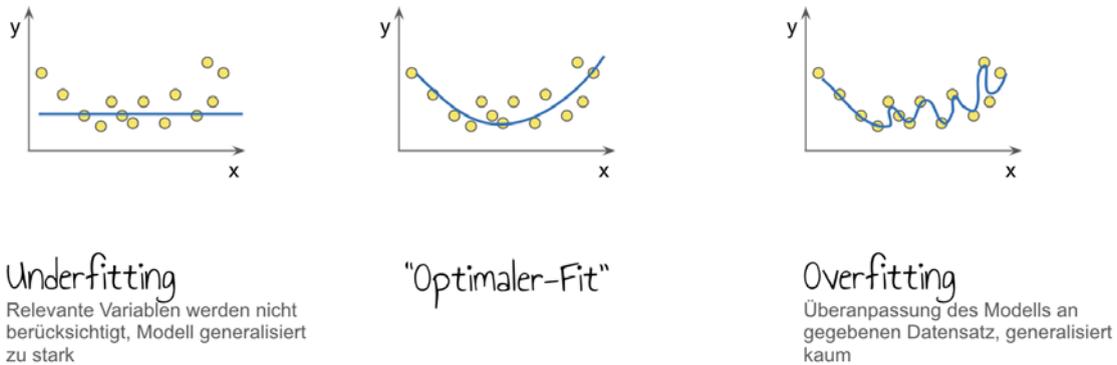


Abb.: Beispiel für Overfitting und Underfitting sowie dem „optimalen Fit“ bei einem einfachen Modell

5.3 Bewertung der Güte eines Modells

Um ein Modell zu testen werden Testdaten verwendet. Auch für diese Testdaten (wie für die Trainingsdaten) ist die Beschriftung, also tatsächliche Klasse der Daten, bekannt. Bei unseren Äffchen wissen wir also, ob sie beißen oder nicht. Um zu bewerten, wie „gut“ das Modell die Testdaten klassifiziert, wird die Güte des Modells bestimmt. Dazu wird das Modell auf die Testdaten angewendet. Anschließend vergleiche wir die vom Modell vorhergesagten mit den eigentlichen Beschriftungen.

Wenn bei einem Klassifikationsproblem nur zwei Klassen zu unterscheiden sind (wie „beißt“ und „beißt nicht“), bietet es sich an, die Ergebnisse dieses Vergleichs als Tabelle (sog. Konfusionsmatrix) darzustellen.

		Predicted		Σ
		ja	nein	
Actual	ja	4	0	4
	nein	1	8	9
Σ		5	8	13

Konfusionsmatrix: Vom Modell vorhergesagte Klassen (Predicted) werden spaltenweise, die tatsächlichen Klassen (Actual) zeilenweise eingetragen.

Anhand dieser Tabelle sind vier wichtige Fälle zu unterscheiden

- **True positive (TP)** (Richtig positiv): Fälle, in denen das Modell **ja** vorhergesagt hat und die tatsächliche Klasse ebenfalls **ja** war (Beispiel: in obiger Konfusionsmatrix: 4). Beispielsweise hat das Modell ein Äffchen als beißend klassifiziert und tatsächlich ist das Äffchen auch beißend.
- **True negative (TN)** (Richtig negativ): Fälle, in denen das Modell **nein** vorhergesagt hat und die tatsächliche Klasse ebenfalls **nein** war (Beispiel: in obiger Konfusionsmatrix: 8). Das Äffchen beißt nicht und wurde auch als „nicht beißend“ klassifiziert.
- **False positive (FP)** (Falsch positiv): Fälle, in denen das Modell **ja** vorhergesagt hat und die tatsächliche Klasse aber **nein** war (Beispiel: in obiger Konfusionsmatrix: 1). Das Modell hat das Äffchen als beißend klassifiziert, allerdings war das Äffchen tatsächlich als nicht beißend kategorisiert.
- **False negative (FN)** (Falsch negativ): Fälle, in denen das Modell **nein** vorhergesagt hat und die tatsächliche Klasse aber **ja** war (Beispiel: in obiger Konfusionsmatrix: 0). In diesem Fall hat das Modell das Äffchen als nicht beißend klassifiziert, allerdings beißt es doch.

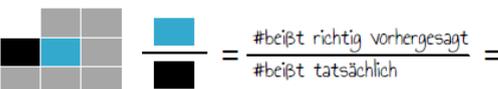
Mithilfe dieser Konfusionsmatrix lassen sich nun verschiedene Metriken zur Bewertung der Güte ganz einfach bestimmen. Zu den bekanntesten Metriken gehören dabei die **Genauigkeit** (accuracy), die **Trefferquote** (recall) und **Präzision** (precision).

Genauigkeit (accuracy): Die Genauigkeit gibt an, wie viele aller Fälle korrekt klassifiziert wurden, also:

$$\text{Genauigkeit} = \frac{TP+TN}{TP+TN+FP+FN}$$

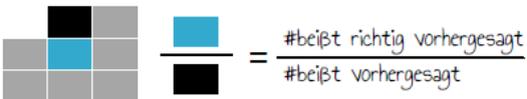

Genauigkeit stellt die allgemeinste Metrik dar und bietet sich beispielsweise dann an, wenn es darum geht, Hunde von Katzen zu unterscheiden, und beide Klassen ungefähr gleich oft in den Trainingsdaten vorkommen.

Trefferquote (recall): Die Trefferquote (oder Sensitivität) wiederum misst, wie viele der tatsächlich positiven Fälle richtig erkannt wurden:

$$\text{Trefferquote} = \frac{TP}{TP + FN}$$


Dies ist beispielsweise hilfreich, wenn hohe Kosten dadurch entstehen, dass kranke Personen (in der Medizinsprache *positiv* getestet) nicht erkannt werden. Vielleicht erinnern Sie sich noch an die Diskussionen bei der Einführung der Corona-Tests. Auch hier ging es um die Sensitivität der Tests.

Präzision (precision): Die Präzision gibt an, wie gut (oder wie präzise) das Modell tatsächlich positive Fälle erkennt:

$$\text{Präzision} = \frac{TP}{TP + FP}$$


$$= \frac{\text{\#beißt richtig vorhergesagt}}{\text{\#beißt vorhergesagt}}$$

Präzision eignet sich dann als Metrik, wenn die Kosten für falsch positive Fälle sehr hoch sind, beispielsweise wenn eine gesunde Person durch eine falsche Vorhersage des Modells eine teure und die Gesundheit beeinträchtigende Chemotherapie erhält.

Je nach Problem kann mal die eine und mal die andere Metrik besser geeignet sein, um die Güte des Modells zu bewerten.

Wenn, beispielsweise mithilfe dieser Metriken, beim Testen festgestellt wurde, dass das KI-Modell keine zufriedenstellenden Ergebnisse liefert, können sowohl die zum Training verwendeten Daten als auch Auswahl und Konfiguration des eingesetzten Verfahrens überarbeitet werden.

5.4 Bias

Von Algorithmen erhoffen wir uns im Allgemeinen faire und objektive Entscheidungen auf der Grundlage von Fakten, die im Gegensatz zu unseren menschlichen Entscheidungen, nicht von Emotionen, Stimmungen oder einseitigen Erfahrungen beeinflusst werden. Dennoch lesen wir, trotz Testphase, in den Medien von rassistischen Chatbots, ungerechtfertigten Verurteilungen durch KI und sexistischen Gesichtserkennungssystemen. Ursächlich dafür ist häufig ein Bias (Verzerrung) in den Daten, die zum Training und zum Testen des Modells eingesetzt wurden. Dabei können verschiedene Arten von Bias auftreten. Von einem historischen Bias sprechen wir beispielsweise dann, wenn sich KI-Systeme bei Einstellungsprozessen von Bewerberinnen und Bewerbern nur auf Einstellungsprozesse in der Vergangenheit beziehen. Von einer Stichprobenverzerrung (Selection Bias) ist beispielsweise die Rede, wenn ein Verfahren zur Gesichtserkennung vor allem mit hellhäutigen Männern trainiert wird und die Ergebnisse für dunkelhäutige Frauen deutlich schlechter ausfallen, diese Personengruppe also weniger gut erkannt wird. Nachdem die Test- und Trainingsdaten aus derselben Datengrundlage stammen und damit gegebenenfalls beide durch einen Bias verzerrt sind, zeigen sich daraus resultierende Probleme nicht in der Testphase sondern erst beim Einsatz in der Praxis und wir lesen in den Medien davon.

Maschinelle Lernverfahren leiten lediglich statistische Zusammenhänge aus Daten ab. So können dementsprechend immer nur **Korrelationen** (statistische Beziehungen) und nie **Kausalitäten** (Ursache-Wirkung-Beziehungen) identifiziert werden. Auch das ist bei der Bewertung des Einsatzes von entsprechenden KI-Systemen zu beachten.

Ein KI-System an sich ist also nicht „diskriminierend“ oder „sexistisch“. Aber wir Menschen sind eben oftmals unfair oder voreingenommen und das zeigt sich auch in den von uns produzierten Daten, die wir für maschinelle Lernverfahren heranziehen (etwa bei Bewerbungsprozessen). Damit übertragen wir unsere menschlichen Vorurteile auf die Technologie. Ein Bias kann aber auch durch Fehler in der Datenerhebung der verwendeten Datensätze (Stichprobenauswahl, soziale Erwünschtheit, ...) entstehen. Bias in den verwendeten Daten aufzudecken, ist in der Praxis nicht so einfach: In diesem Modul arbeiten wir zu Beginn mit sehr kleinen Datensätzen, die wir auch händisch auswerten und überblicken können. In der Praxis hingegen, werden

große Datensätze mit Millionen von Datenpunkten verwendet. Wie immer, wenn Informatiksysteme Entscheidungen mit direktem Einfluss auf unser Leben fällen, müssen sie insbesondere auch im Praxiseinsatz gründlich überprüft werden, um sicherzustellen, dass sie auch wie beabsichtigt funktioniert. Genauso wie bei anderen Werkzeugen, muss abgewogen werden, ob und wie wir sie als Gesellschaft einsetzen wollen.

5.5 Orange3

Datenanalyse mit Orange3

Orange3 ist ein Datenanalysetool, das an der Universität Ljubljana entwickelt wird und für alle gängigen Desktop-Betriebssysteme zum Download zur Verfügung steht (<http://orange.biolab.si/>). In Orange werden Datenanalysen nicht durch textbasierte Programmierung, sondern als Datenflussdiagramm beschrieben (siehe z.B. Arbeitsmaterial KI-B4.2.2). Entsprechend treten keine Syntaxfehler auf und der Fokus im Unterricht kann auf der Analyse und Interpretation der Daten liegen. Orange kann ohne Administratorrechte ausgeführt und via pip, via conda² oder als ausführbare Datei bezogen werden. Der Installationsprozess kann je nach System eine gewisse Zeit dauern. Deswegen sollte die Software vor der Unterrichtsstunde heruntergeladen und ausprobiert werden. Der Fokus in dieser Sequenz liegt auf dem überwachten Verfahren des Lernens von Entscheidungsbäumen, die im Folgenden genauer beschrieben werden. Für Tipps und Hilfestellungen zu den Funktionen von Orange3 sei an dieser Stelle auf die Hilfekarten KI-B4.2.5 verwiesen.

Mit Orange 3 können auch weitere Lernverfahren eingesetzt werden, z.B. Random Forests und k-Nearest-Neighbour. Diese Verfahren werden an dieser Stelle nicht vertieft, Sie finden dazu aber nähere Erläuterung im Zusatzmaterial KI-B4.2.6. Der Einsatz dieser Lernverfahren eignet sich vor allem für Fortgeschrittene im Rahmen tiefergehender Projektphasen. Die Schülerinnen und Schüler haben dann auch die Möglichkeit mit eigenen Daten zu arbeiten.

Datenerfassung und Modellierung

Um Daten in Orange3 verwenden zu können, müssen sie in einem für den Computer verarbeitbaren Format vorliegen. Für die Äffchen aus dem Unplugged-Spiel wählen wir dazu geeignete Merkmale aus den Bildern aus und übertragen die Informationen in eine Tabelle³. Mit diesem Schritt bereiten wir die Daten so auf, dass sie für den Computer verarbeitbar sind. Dabei fokussieren wir uns nur auf die Merkmale, die für den Anwendungsfall relevant sind. Dieses Vorgehen bezeichnet man als Modellierung. In der Tabelle gibt jede Spalte ein Merkmal an, jede Zeile beschreibt einen Datensatz (z.B. Bild von einem Äffchen). Bei der Auswahl der Merkmale der Äffchen (vgl. Abbildung) ergeben sich unterschiedliche Möglichkeiten: Die Augen lassen sich zusammen als Merkmal "Augen" mit Merkmalsausprägungen wie "beide offen" oder "nur ein Auge offen", einzeln wie in der Abbildung oder durch ja/nein-Merkmale wie "beide Augen offen", modellieren. Für überwachtetes Lernen muss außerdem ein Zielmerkmal, d.h. die vorherzusagende Beschriftung, festgelegt werden. In unserem Fall ist das „beißt“ oder „beißt nicht“. Für

² Bei pip und conda handelt es um Paketverwaltungsprogramme für die Programmiersprache Python. Sie müssen die Programme jedoch nicht kennen, um dieses Modul einsetzen zu können.

³ Der Computer könnte in unserem Beispiel auch direkt mit den Bildern arbeiten, wie das etwa bei Bilderkennung bzw. Klassifikation von Bildern der Fall ist. Allerdings liegen viele Daten in tabellarischer Form vor, wie auch in den späteren Beispielen in diesem Modul. Auch dort werden die für den Anwendungsfall relevanten Merkmale aus der „echten Welt“ erfasst und für den Computer verarbeitbar gemacht.

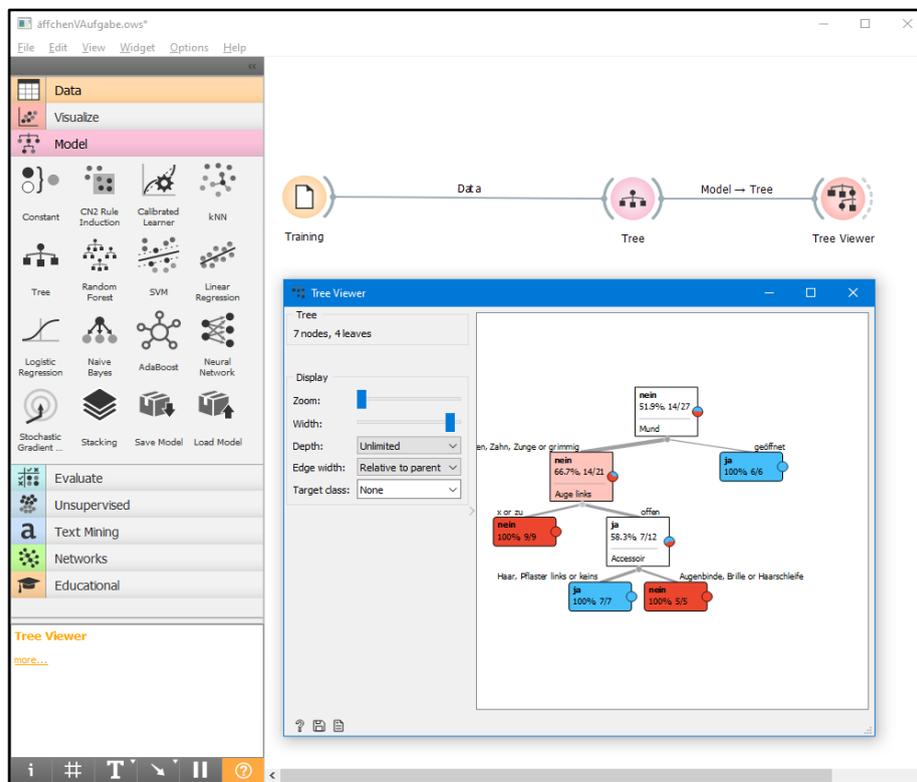
diese Unterrichtssequenz sind alle Daten bereits erfasst, liegen als csv-Datei vor und können direkt verwendet werden.

mS#Nummer	D#Auge links	D#Auge rechts	D#Mund	D#Accessoire	cD#Beschriftung
3	x	x	Zahn	keins	beißt nicht
6	offen	zu	Grinsen	keins	beißt
8	zu	zu	Grinsen	keins	beißt nicht
11	x	x	grimmig	keins	beißt nicht

Beispiel Kodierung der Äffchen: Jede Zeile stellt ein Äffchen dar, die verschiedenen Merkmale finden sich in den Spalten (die Zeichen vor # sind spezifisch für Orange3 und geben an, ob das Merkmal diskret (D) oder kontinuierlich (C) ist bzw. ob es sich lediglich um beschreibende Metadaten (m) oder das Zielmerkmal (c) handelt.

6 Unterrichtliche Umsetzung

Im Rahmen des Moduls KI-B3 „Schlag den Roboter“ haben die Schülerinnen und Schüler durch verschiedene Unplugged-Aktivitäten die grundlegenden Funktionsweisen verschiedener KI-Methoden untersucht sowie Beispiele kennengelernt und hinterfragt. Darauf aufbauend modellieren sie nun das bereits in der Unplugged-Variante erkundete Beispiel des Gute-Äffchen-Böse-Äffchen-Spiels mithilfe von Orange3 (siehe Abb.). Anschließend erfolgt eine Übertragung der gelernten Konzepte auf reale Datensätze, die die Schülerinnen und Schüler mithilfe von KI-Methoden auswerten. So erfahren sie wie Daten, die auch als Öl des 21. Jahrhunderts bezeichnet werden, erst die Grundlage für die jüngeren Entwicklungen künstlicher Intelligenz bilden⁴. Zum Einsatz kommen überwachte Lernverfahren. Aufbauend auf dieser Anwendung von KI-Methoden werden anschließend die gesellschaftliche Dimension betrachtet sowie drängende Fragen von KI-Verfahren diskutiert und bewertet. Das Modul ist dabei so gestaltet, dass es sowohl im Anschluss an das Modul „Schlag den Roboter“ als auch unabhängig davon im Unterricht eingesetzt werden kann.



Das Gute-Äffchen-Böse-Äffchen-Spiel in Orange3

⁴ In diesem Zusammenhang wird häufig der Begriff „Big Data“ verwendet.

6.1 Grober Unterrichtsplan

Variante I:

Diese Variante ist empfehlenswert, wenn weniger Unterrichtsstunden zur Verfügung stehen oder für die Lerngruppe eine einfachere Variante angemessen ist. Das Modul wird dann in einer verkürzten Version ohne Projektphase durchgeführt:

Unterrichtsszenarien	Kurze Zusammenfassung
Datenanalyse und Entscheidungsbaum (1 UZE)	Die Schülerinnen und Schüler ... erstellen händisch den Entscheidungsbaum und die Konfusionsmatrix für ein einfaches Beispiel.
Datenanalyse mit Orange3 (2 UZE)	Die Schülerinnen und Schüler ... nutzen das Werkzeug Orange3, um gegebene Daten zu analysieren. Hinweis: die Software Orange3 sollte vorher heruntergeladen werden.
KI und unsere Gesellschaft (1 UZE)	Die Schülerinnen und Schüler ... diskutieren und bewerten Konsequenzen aus dem Einsatz von KI-Methoden. ... erläutern den Unterschied von Korrelation und Kausalität. ... erläutern, warum Verzerrungen in Daten die Ergebnisse von KI-Systemen beeinflussen.

Variante II: In der ausführlichen Variante werden die erworbenen Kompetenzen zur Datenanalyse in Orange 3 vor der abschließenden Betrachtung gesellschaftlicher Auswirkungen in einer Projektphase vertieft. Diese schließt sich damit an die Doppelstunde zur Datenanalyse in Orange3 an.

Unterrichtsszenarien	Kurze Zusammenfassung
Projektarbeit (2-3 UZE)	Die Schülerinnen und Schüler ... lösen Probleme unter Nutzung von KI-Methoden. ... erläutern, warum bei Ergebnissen von ML-Verfahren KI-bedingt Fehler enthalten sein können. ... beschreiben Einsatzmöglichkeiten von KI-Methoden als „Werkzeug“ in beruflichen Tätigkeiten.

Vertiefungsmöglichkeiten: Die im Modul erlernten Kompetenzen im Umgang und zur Auswertung von Daten mithilfe von KI können auch auf weitere, insbesondere auch fachspezifische Fragestellungen angewandt werden. Dazu besteht auch die Möglichkeit, selbst Daten zu erheben. So könnten bspw. im naturwissenschaftlichen Unterricht Daten aus Experimenten gewonnen und anschließend, etwa in der Projektphase dieses Moduls, ausgewertet werden. Darüber hinaus bietet das Werkzeug Orange3 neben den im Modul eingeführten Verfahren weitere Möglichkeiten zur Auswertung und Visualisierung von Daten, die im Unterricht eingesetzt werden können.

6.2 Stundenverlaufsskizzen

Abkürzungen/Legende

AB = Arbeitsblatt/Arbeitsblätter; L = Lehrkraft; MuM = Mitschülerinnen und Mitschüler; SuS = Schülerinnen und Schüler;
UV = Unternehmensvertreterin/Unternehmensvertreter

6.2.1 Stunde 1: Datenanalyse und Entscheidungsbaum

In der ersten Einheit dieser Sequenz lernen die Schülerinnen und Schüler den Begriff des Entscheidungsbaums und der Konfusionsmatrix sowie Metriken zur Beurteilung der Güte eines Modells kennen, indem sie Äffchen nach ihrem Beißverhalten beim Füttern klassifizieren.

Zeit	Phase	Sozialform/ Lehrerimpuls	Inhalt/Unterrichtsgeschehen	Material
5	Einstieg	Lehrer-Schüler-Gespräch	<p>Zu Beginn der Stunde wird die Geschichte über die Firma Target auf dem Beamer aufgelegt und diese gemeinsam besprochen. Beispielsweise mit folgenden Diskussionsfragen:</p> <ul style="list-style-type: none"> - Welche Informationen hatte die Firma Target über ihre Kunden? - Welche Folgen können solche Vorhersagemodelle für die Betroffenen haben? - Kennt ihr auch andere Bereiche, in denen Daten über Euch gesammelt werden, um euch passende Angebote zu machen? Beispiel Spotify oder Netflix - Lebensweltbezug 	Foliensatz KI-B4.1
30	Eratbeitung	Einzel- oder Partnerarbeit	<ul style="list-style-type: none"> • Differenzierung I (leichter, Spiel noch nicht gespielt): Sie haben das Unplugged-Spiel mit ihrer Klasse noch nicht gespielt, dann nutzen Sie KI-B4.1.2a und im späteren Verlauf weiterhin die a-Varianten des Materials. • Differenzierung II (schwerer, Spiel noch nicht gespielt): Sie haben das Spiel noch nicht gespielt, wollen aber die komplexere Version durchführen. Nehmen Sie in diesem Fall die Äffchen aus KI-B4.1.2a als Beispiel her, anhand dessen sie die Spielregeln im Klassengespräch einführen, und führen sie anschließend KI-B4.1.2b durch. Im weiteren Verlauf des Moduls nutzen sie stets die b-Variante des Materials. 	<p>Differenzierung I: KI-B4.2.1 a) Differenzierung II und III: KI-B4.2.1 b) Kontrollblatt KI-B4.2.2</p>

			<ul style="list-style-type: none"> • Differenzierung III (schwerer, Spiel bereits gespielt): Sie haben das Modul „Schlag den Roboter“ bereits unterrichtet und das Gute-Äffchen-Böse-Äffchen-Spiel bereits gespielt. Nutzen Sie in diesem Fall KI-B4.1.2b und auch im späteren Verlauf weiterhin die b-Variante des Materials. 	
10	Diskussion	Lehrer-Schüler-Gespräch	Vergleich der Lösungen und Lehrer-Schüler-Gespräch zu den verschiedenen Metriken zur Beurteilung der Güte von Modellen.	

Stunde 1: Datenanalyse und Entscheidungsbaum

In der ersten Stunde lernen die Schülerinnen und Schüler den Begriff des Entscheidungsbaums und der Konfusionsmatrix sowie Metriken zur Beurteilung der Güte eines Modells kennen, indem sie die Aktivität 1 aus dem Modul KI-B3 „Schlag den Roboter“ erneut durchführen.

Präsentieren Sie zu Beginn der Stunde eine Zeitungsmeldung (siehe KI-B4.1.1) mit einer Geschichte über einen Einzelhändler, die sich in den USA zugetragen haben soll (z.B. über den Beamer).

Geben Sie den Schülerinnen und Schülern folgende Fragestellungen als Impuls mit:

- Welche Informationen hatte die Firma Target über ihre Kunden?
(Antwortmöglichkeiten: Welche Produkte wann gekauft wurden, Häufigkeit der Einkäufe, Tageszeiten, Menge gekaufter Produkte)
- Welche Folgen können solche Vorhersagemodelle für die Betroffenen haben?
(Antwortmöglichkeiten: Passgenaue Produktempfehlungen, Eindringen in die Privatsphäre, unangenehmes Gefühl beobachtet zu werden, Stigmatisierung, ...)
- Habt ihr Erfahrungen mit anderen Bereichen, in denen Daten gesammelt werden, um Euch passende Angebote zu liefern? (Antwortmöglichkeiten: Spotify, Netflix, Amazon Prime, Online-Shopping)

Diskutieren Sie mit Ihren Schülerinnen und Schülern über den Artikel. Leiten Sie dann dazu über, dass die Schülerinnen und Schüler in dieser Unterrichtssequenz selbst die Möglichkeit haben, solche Vorhersagemodelle zu erstellen.

Um solche Modelle nun auch selbst zu erstellen, wird zunächst die Unplugged-Aktivität zu überwachtem Lernen, das „Gute-Äffchen-böse-Äffchen“-Spiel aus Modul KI-B3 „Schlag den Roboter“ herangezogen. Bei der Erstellung der Regeln bzw. des Entscheidungsbaums auf dem zugehörigen Arbeitsblatt gibt es keine „eine Musterlösung“, auch wenn in den Begleitmaterialien eine mögliche „gute“ Lösung zur Orientierung angegeben ist. Auch diese Lösung ermöglicht aber keine vollständig korrekte Klassifikation der Testdaten, die beim maschinellen Lernen im Allgemeinen nicht erreicht werden kann. Die Schülerinnen und Schüler sollen bewusst mit ihren eigenen – und möglicherweise suboptimalen – Modellen das restliche Arbeitsblatt bearbeiten.

Vergleichen Sie anschließend die Lösungen mit den Schülerinnen und Schülern und sprechen Sie über die verschiedenen Metriken zur Beurteilung der Güte eines Modells im Lehrer-Schüler-Gespräch. Diese Metriken lassen sich zukünftig verwenden, um die Güte eines vom Computer automatisch erstellten Modells zu bestimmen.

6.2.2 Stunde 2-3: Datenanalyse mit Orange3

In der zweiten Einheit dieser Sequenz lernen die Schülerinnen und Schüler das Werkzeug Orange3 kennen und nutzen es zur Analyse von gegebenen Daten. Hinweis: Wenn die Software Orange 3 noch nicht auf den Schulrechnern heruntergeladen wurde, sollte dies vor der Stunde erledigt werden.

Zeit	Phase	Sozialform/ Lehrerimpuls	Inhalt/Unterrichtsgeschehen	Material
15	Einstieg	Einzel- oder Partnerarbeit	Die Schülerinnen und Schüler kodieren 3 gegebene Äffchen, sodass ein Computer diese verarbeiten könnte.	optional: Tabelle KI-B4.2.1
60	Erarbeitung & Vertiefung	Einzel- oder Partnerarbeit	Die Schülerinnen und Schüler nutzen Orange3, um Arbeitsblatt KI-B4.3.1 zu bearbeiten. Differenzierung I: Sie haben in Stunde 1 das Material a gewählt, dann wählen Sie auch hier KI-B4.2.2a. Differenzierung II+III: Sie haben in Stunde 1 das Material b gewählt, dann wählen Sie auch hier KI-B4.2.2b.	Differenzierung I: KI-B4.2.2a Differenzierung II und III: KI-B4.2.2b Datensatz KI-B4.2.3 optional: Hilfefkarten KI-B4.3.5
15	Besprechung	Lehrer-Schüler-Gespräch	Besprechung des Arbeitsblatts und ggf. Vergleich der Güte der verschiedenen Modelle. Thematisierung von Over- und Underfitting.	

Stunde 2 und 3: Datenanalyse mit Orange3

In der zweiten Einheit dieser Sequenz lernen die Schülerinnen und Schüler das Werkzeug Orange3 kennen und nutzen es zur Analyse von gegebenen Daten, um im Anschluss das Projekt durchführen zu können. Damit wenden die Schülerinnen und Schüler ein ähnliches Vorgehen wie die Data Scientists der Firma Target an.

Vermutlich ist Ihren Schülerinnen und Schülern klar, dass Computer mit strukturierten Daten und weniger mit Münzen oder Karten arbeiten. Dies dient auch als Fragestellung für den Beginn der Lerneinheit: Wie lässt sich dieses Verfahren nun durch einen Computer ausführen?

Dazu sollen die Schülerinnen und Schüler drei gegebene Äffchen geeignet modellieren, sodass ein Computer diese verarbeiten könnte. Dazu werden die Äffchen 1, 33 und 14 gezeigt. Optional kann bereits eine Tabelle vorgegeben werden (siehe KI-B4.2.1).

Es ergeben sich unterschiedliche Möglichkeiten zur Modellierung, etwa binär (Merkmal vorhanden/nicht vorhanden), oder aber über die verschiedenen Merkmalsausprägungen (bspw. Mund: offen/geschlossen/Zunge rausstrecken/ Zähne fletschen).

	Mund weit offen
	nein
	ja
	ja

binär

	Mund
	Zähne
	weit geöffnet
	weit geöffnet

nicht binär

Anschließend soll nun das Äffchenspiel in Orange3 gelöst werden. Findet die KI einen besseren Entscheidungsbaum? Dazu starten die Schülerinnen und Schüler Orange3 und bearbeiten Arbeitsblatt KI-B4.2.2 mit dem Datensatz KI-B4.2.3.

Wählen Sie dabei Variante KI-B4.2.2a, falls Sie

- ... in der erste Unterrichtsstunde Variante KI-B4.1.2a des Arbeitsblattes gewählt haben oder eine kürzere Variante bevorzugen.

Ansonsten wählen sie Sie KI-B4.2.2b.

Mögliche Differenzierung: Als mögliche Differenzierung können weitere Modelle (Random Forest und KNN) eingeführt werden. Dazu können die beigelegten Folien (KI-B4.2.4) verwendet werden. Anschließend sollten die Schülerinnen und Schüler ggf. noch einmal die Möglichkeit erhalten, die alternativen Modelle in Orange3 auszuprobieren.

Im Lehrer-Schüler-Gespräch wird nun das Arbeitsblatt besprochen und ggf. die Güte der verschiedenen Modelle verglichen. Außerdem können anhand von Aufgabe 3 die Begriffe Over- und Underfitting thematisiert werden.

6.2.3 Stunde 4-6: Projektarbeit (nur Variante II)

Während der (bis zu) dreistündigen Projektarbeitsphase arbeiten die Schülerinnen und Schüler an ihrem Arbeitsauftrag und erstellen ein Poster zur Präsentation ihrer Ergebnisse. Optional kann während des Projekts den Gruppen der Link zu einem Interview mit Data Scientist gegeben werden.

Zeit	Phase	Sozialform/ Lehrerimpuls	Inhalt/Unterrichtsgeschehen	Material
10	Wahl Projektthema	Lehrer-Schüler-Gespräch	Vorstellung und Wahl der Projektthemen	Themenübersicht KI-B4.3.1
50-95min	Durchführung des Projekts	Projektphase	Die Schülerinnen und Schüler bearbeiten das gewählte Projektthema in Gruppen und halten ihre Ergebnisse auf einem Plakat fest. Optional kann während des Projekts den Gruppen der Link zu einem Interview mit Data Scientist gegeben werden, die sich auf dem YouTube Kanal der Wissensfabrik finden: <ul style="list-style-type: none"> • Was macht eigentlich eine Data Scientist? Interview mit Elena Bruna https://youtu.be/EbURtR4NUj4 • Was macht eigentlich ein KI-Entwickler? Interview mit Dennis Hügele https://youtu.be/_sZnDZF7LgM 	KI-B4.3.3 Arbeitsanweisung Projekt KI-B4.2.5 Hilfefkarten für Orange3 ggf. KI-B4.3.4 Datensatz für das Projekt student data 1 x Plakat je Gruppe zur Darstellung der Ergebnisse (nicht in Klassensatz enthalten).
30	Präsentation	Schülerpräsentation und moderierte Fragegerunde	Präsentation und Diskussion der Ergebnisse (z.B. 5 x 4er Gruppe à 5 Minuten).	Von Schülerinnen und Schülern erstellte Präsentationen / Poster



Stunde 4 bis 6: Projektarbeit (nur Variante II)

Während der dreistündigen Projektarbeitsphase arbeiten die Schülerinnen und Schüler weitestgehend selbstständig. Stellen Sie der Klasse zunächst die verschiedenen Themen und damit verbundenen Fragestellungen für die Projektarbeit vor (vgl. KI-B4.3.1), bilden Sie Gruppen und lassen Sie die Gruppen dann jeweils ein Thema wählen.

Anschließend erhalten sie den Arbeitsauftrag.

Arbeitsauftrag (findet sich auch in den Begleitmaterialien)

Als Data Scientist ist es eure Aufgabe, Probleme durch Erstellen eines Klassifikationsmodells zu lösen. Dazu steht euch ein Datensatz zur Verfügung, der den Ausgangspunkt für euer KI-Modell darstellt.

Folgt dazu folgenden Schritten:

1. Erstellt ein neues Orange3-Projekt.
 - Wenn euer Datensatz bereits in Orange3 enthalten ist, zieht das Datasets-Widget auf die Leinwand und wählt den entsprechenden Datensatz mit einem Doppelklick aus.
 - Wenn euer Datensatz nicht in Orange3 enthalten ist, öffnet ihn wie den Äffchendatensatz mithilfe des File-Widgets.
2. Verschafft euch anschließend einen Überblick über den Datensatz mithilfe des Data Table-Widgets! Notiert, welche Informationen in den Spalten gespeichert sind. Was sind die möglichen Klassen? Welche Merkmale stehen euch zur Verfügung?
3. Teilt die Daten mithilfe des Data Sampler-Widget in Trainings- und Testdaten auf.
4. Trainiert mehrere Modelle, indem ihr unterschiedliche Merkmale für das Training auswählt und verschiedene Parameter des Trainingsprozesses festlegt.
5. Testet und bewertet eure Modelle mithilfe einer geeigneten Metrik.
6. Überlegt euch, welche Personengruppen wie von eurem Modell profitieren und welche Personengruppen wie von eurem Modell benachteiligt werden könnten.
7. Gestaltet ein Poster, das euren Datensatz, euer Ziel und eure Ergebnisse knapp und präzise zusammenfasst. Was findet ihr an eurer Auswertung überraschend?

Den Schülerinnen und Schülern stehen Hilfekarten zu den verschiedenen im Rahmen des Projekts relevanten Möglichkeiten innerhalb von Orange3. Die Schülerinnen und Schüler können zum Festhalten und Präsentieren ein Poster erstellen.

Zusätzlich besteht während des Projekts die Möglichkeit, den Gruppen einzeln ein aufgezeichnetes Interview mit zwei Data Scientists zu zeigen und so mögliche Berufs- und Karriereoptionen deutlich zu machen.

Abschließend präsentieren die Schülerinnen und Schüler ihre Ergebnisse. Bitten Sie sie dazu ihr Plakat an einer Stelle im Klassenzimmer aufzuhängen. Anschließend präsentieren die Gruppen der Reihe nach ihre Ergebnisse und die Klasse wandert mit. Dabei sollte jeweils dieselbe Zeit, die für die Präsentation verwendet wurde, auch für die anschließende Diskussion der Ergebnisse aufgewendet werden (z.B. 5 x 4er Gruppe à 5 Minuten mit 2,5 Präsentation und 2,5 Minuten Diskussion). Lassen Sie die Schülerinnen und Schüler dabei die Aufgaben des Menschen beim Erstellen von Modellen im Kontext des maschinellen Lernens (Konfigurieren des Verfahrens, Festlegen der Parameter, Güte der resultierenden Modelle kritisch einschätzen) reflektieren und benennen.



6.2.4 Stunde 7: KI und unsere Gesellschaft

In der letzten Einheit diskutieren die Schülerinnen und Schüler über die ethische Perspektive beim Einsatz von KI-Modellen.

Zeit	Phase	Sozialform/ Lehrerimpuls	Inhalt/Unterrichtsgeschehen	Material
25	Erarbeitung & Vertiefung	Partnerarbeit	Die Schülerinnen und Schüler bearbeiten in Partnerarbeit die vorgegebenen Szenarien und diskutieren über die Leitfragen. Alternativ können die Aufgaben mit Hilfe der Jigsaw-Methode bearbeitet werden.	KI-B4.4
10	Sicherung	Lehrer-Schüler-Gespräch	1-2 Gruppen stellen ihre Ergebnisse vor.	
10	Fazit und Ausblick	Lehrer-Schüler-Gespräch	Anschließend wird der Einsatz von KI im Plenum diskutiert. Leitfragen sind <ul style="list-style-type: none"> • Auf Basis des Gelernten: Wofür sollte KI eingesetzt werden -- und wofür nicht? • Welche Regeln sollten für einen fairen Einsatz von KI-Systemen gelten? 	

Stunde 7: KI und unsere Gesellschaft

In der letzten Einheit diskutieren Sie mit ihrer Klasse die ethische Perspektive beim Einsatz von KI-Modellen. Dazu erhalten die Schülerinnen und Schüler das Arbeitsmaterial KI-B4.4 mit drei Szenarien über unzuverlässige, diskriminierende bzw. fehlerhafte KI-Systeme. Diese werden zunächst in Partnerarbeit und anschließend im Plenum diskutiert. Durch den gesellschaftlich relevanten Bezug kann dieses Thema in unterschiedliche Fächer eingebettet werden.

7 Einbettung in verschiedene Fächer und Themen

Im Zuge der digitalen Transformation spielen Daten und Datenanalyse in allen Schulfächern und deren Bezugswissenschaften eine herausragende Rolle. So werden Datenanalysen und Künstliche Intelligenz oft als viertes Standbein der Wissenschaft bezeichnet. Jeder Wissenschaftler bzw. jede Wissenschaftlerin benötigt entsprechende Kompetenzen im Umgang und der Analyse von Daten. Und auch für den Schulunterricht werden entsprechende Vorgehensweisen zunehmend Inhalt und bieten weiterhin enormes Potential für einen handlungsorientierten und fachübergreifenden Unterricht. Ob bei der Durchführung von Experimenten im Physikunterricht, der Beantwortung von Fragen im Chemie- oder Biologieunterricht, der Auswertung von Geodaten im Geographieunterricht, der Analyse von Textdokumenten im Fach Deutsch oder der Diskussion ethischer, gesellschaftlicher und moralischer Konsequenzen im Ethik- oder Religionsunterricht: Überall finden die in diesem Modul eingeführten Verfahren zunehmende Bedeutung und können fachübergreifend zur Beantwortung fachspezifischer Fragestellungen eingesetzt werden.

8 Anschluss Themen

Als Anschluss Themen im Zusammenhang mit IT2School bieten sich folgende Bausteine an:

KI-A2: Mein persönlicher Assistent

Während in diesem Modul die datenbasierte Sicht eingenommen wurde, können die Schülerinnen und Schüler im Modul „Mein persönlicher Assistent“ selbst einen (einfachen) digitalen Sprachassistenten entwickeln. Dieses Modul ergänzt die hier eingenommene Sichtweise und zeigt, dass KI auch als kreatives Werkzeug für persönliche Projekte verwendet werden kann.



9 Literatur und Links

- **AI Unplugged:** <https://aiunplugged.org>
- **Werkzeug Orange3:** <https://orangedatamining.com> (kann ohne Administratorrechte genutzt werden)
- **Quellen für weitere Datensätze:** <https://datasetsearch.research.google.com/>, <https://archive.ics.uci.edu/ml/datasets/>
- **Unterrichtsreihe „Data Mining im Unterricht“:** <https://dataliteracy.education/unterrichtsmaterial/>

10 Arbeitsmaterialien

Das Modul besteht aus folgenden Materialien. Zu allen Arbeitsblättern gibt es auch Musterlösungen.

Nr.	Titel	Beschreibung
😊 KI-B4.1.1	Einstiegsimpuls	Einstiegsimpuls Zeitungsmeldung „Target“
😊 KI-B4.1.2	Entscheidungsbäume	Arbeitsblatt zur Einführung von Entscheidungsbäumen anhand der Äffchen in Variante a und b.
😊 KI-B4.1.3	Kontrollblatt	Kontrolle der Klassifizierung der Äffchen um anschließend die Metriken berechnen zu können
😊 KI-B4.2.1	Vorlage Tabelle	Tabellarische Vorlage zur Kodierung der Äffchen
😊 KI-B4.2.2	Orange3	Arbeitsblatt zur Einführung in die Nutzung von Orange3 zur Datenanalyse in Variante a und b.
😊 KI-B4.2.3	Datensatz Äffchen	Datensatz zur Bearbeitung des Arbeitsblattes
😊 KI-B4.2.4	Foliensatz Random Forest / KNN	Foliensatz zur optionalen Erläuterung weiterer Verfahren zur Datenanalyse
😊 KI-B4.2.5	Hilfen Orange3	Hilfekarten zu den verschiedenen für die Projektarbeit relevanten Funktionen von Orange3 (kann auch digital zur Verfügung gestellt werden)
😊 KI-B4.2.6	Erläuterung Random Forest / KNN	Fachlicher Hintergrund zu Random Forest bzw. KNN als optionalen weiteren Verfahren, die verwendet werden können.
😊 KI-B4.3.1	Übersicht Datensätze	Eine Übersicht über mögliche Datensätze für die Projektarbeitsphase
😊 KI-B4.3.2	Interviews Data Scientist	Interviews mit Unternehmensvertreter*innen: Was macht eigentlich ein Data Scientist? Interview 1: https://youtu.be/EbURtR4NUj4 Interview 2: https://youtu.be/_sZnDZf7LgM



😊 KI-B4.3.3	Arbeitsauftrag	Arbeitsauftrag für die Projektarbeit
😊 KI-B4.3.4	Datensatz student data	Datensatz für ein Projekt aus der Projektarbeit
😊 KI-B4.4	Arbeitsblatt Bias	Arbeitsblatt mit drei Szenarien zu Bias bei KI-Modellen

Legende

- 😊 Material für Schülerinnen und Schüler
- 👤 Material für Lehrkräfte sowie Unternehmensvertreterinnen und Unternehmensvertreter
- 🔵 Zusatzmaterial

11 Glossar

Begriff	Erläuterung
Maschinelles Lernen	Beim <u>maschinellen Lernen (ML)</u> leiten Computer Zusammenhänge aus Daten bzw. Erfahrungen ab. Das Gelernte wird in einem Modell gespeichert.
Überwachtes Lernen	Bei überwachtem Lernen wird aus beschrifteten Daten eine Zuordnung von Daten zu Beschriftung gelernt, die dann auf weitere, unbeschriftete Daten angewendet werden kann.
Trainingsdaten	Trainingsdaten sind Daten, die zur Erstellung eines mit überwachtem Lernen trainierten Modells herangezogen werden.
Testdaten	Testdaten werden genutzt, um zu beurteilen (zu „testen“), ob ein Modell zufriedenstellende Ergebnisse liefert. Sie sollten nicht Teil des Trainingsprozesses sein.
Entscheidungsbaum	Entscheidungsbäume stellen Entscheidungsregeln in einem geordneten und gerichteten Baum bestehend aus mehreren hierarchisch angeordneten Entscheidungen dar. Sie können im Fall von überwachtem Lernen auch automatisch aus Daten gelernt werden.
Konfusionsmatrix	Die Konfusionsmatrix ist eine Tabelle, in der die richtigen und falschen Vorhersagen eines Modells übersichtlich dargestellt werden.

12 Fragen, Feedback, Anregungen

Sie haben das Modul ausprobiert und nun Fragen, Anregungen oder Feedback für uns? Darüber freuen wir uns, denn mit Ihren Erfahrungen können wir Schritt für Schritt einen FAQ (Frequently Asked Questions) für die neuen KI-Module aufbauen oder die Module weiter entwickeln.

Bitte füllen Sie folgende Umfrage über SurveyMonkey aus: <https://bit.ly/3DQnqi8> über den folgenden QR-Code kommen Sie ebenfalls zur SurveyMonkey-Umfrage:



Sie können sich auch gerne unter bildung@wissensfabrik.de melden.

