

## Datenbewusstsein

# Wo, wie und wozu werden Daten gesammelt und verarbeitet? – Datenbewusstsein durch die Exploration von Empfehlungsdiensten

### Steckbriefinformationen des Unterrichtsmoduls:

<b>Titel:</b>	Wo, wie und wozu werden Daten gesammelt und verarbeitet? – Datenbewusstsein durch die Exploration von Empfehlungsdiensten im Kontext von Streamingdiensten
<b>Zielgruppe:</b>	Informatik in Klasse 8 bis 10 (alle Schulformen)
<b>Inhaltsfelder:</b>	„Informatik, Mensch und Gesellschaft“, „Information und Daten“, „Informatiksysteme“ und „Künstliche Intelligenz und maschinelles Lernen“
<b>Zeitlicher Umfang:</b>	6-8 Unterrichtsstunden je 45 Minuten

### 1 Überblick über den Verlauf des Unterrichtsmoduls:

Dieses Unterrichtsmodul vermittelt Datenbewusstsein: Die Kompetenz, sich die Rolle der erhobenen und verarbeiten Daten bei der Nutzung unterschiedlicher Anwendungen bewusst zu werden, um schließlich die Nutzung bewerten und eigene Handlungsoptionen identifizieren zu können. Damit wird das Ziel verfolgt, die Lernenden zu einer selbstbestimmten Interaktion mit datengetriebenen Anwendungen in ihrem Alltag zu befähigen. Das Modul setzt sich aus vier Teilen zusammen und thematisiert exemplarisch die Erhebung und Verarbeitung von persönlichen Daten bei der Nutzung von Streamingdiensten, bei denen Empfehlungsdienste eingesetzt werden. Fokussiert wird die Rolle der Daten für einen Empfehlungsdienst, wie etwa bei der Startseite bei einem Streamingdienst zu erkennen, wobei ebenfalls weiterführend auch andere Alltagskontexte der Lernenden aufgegriffen werden.

**Im ersten Teil** wird in Empfehlungsdienste als Kontext und in die Idee der personalisierten Filmempfehlungen eingeführt. Dabei wird erarbeitet, welche persönlichen Daten bei der Nutzung eines exemplarischen Streamingdienstes, insbesondere für den primären Zweck des Gebens von personalisierten Filmempfehlungen, erhoben werden. Zum Beispiel sind dafür Nutzungsdaten interessiert, wie etwa welche Filme ein Nutzender zuvor geschaut hat. Dabei entwickeln die Lernenden bereits erste Ideen, was es bedeutet, einem Nutzenden Empfehlungen anzuzeigen und wie solche Filmempfehlungen ermittelt werden könnten. **Im zweiten Teil** wird die Funktionsweise eines Filmempfehlungsdienstes rekonstruiert, wofür die Lernenden in einer vorbereiteten Lernumgebung mit einem funktionierenden Filmempfehlungsdienst (basierend auf realen Nutzungsdaten aus dem Streamingkontext) interagieren und schrittweise die Funktionsweise von der Erhebung von Daten bis hin zur automatisierten Ermittlung von Empfehlungen mit einem Verfahren des maschinellen Lernens erarbeiten (umfassende Erklärung des Empfehlungsdienstes sowie des ML-Verfahrens s. Abs. 9.4). **Im dritten Teil** wird exemplarisch eine Zweitverwertung der Nutzungsdaten durch einen Streamingdienst thematisiert, indem eine Diskussionsrunde zu dem sekundären Zweck einer personalisierten Bezahlschranke basierend auf der Idee des Empfehlungsdienstes betrachtet wird. Dabei wird der Interaktionskontext hinsichtlich der Erhebung und Verarbeitung persönlicher Daten reflektiert und Handlungsoptionen insbesondere auf einer individuellen Betrachtungsebene bedacht und bewertet. In diesem Teil werden verschiedene Aspekte der Wechselwirkung zwischen Nutzendem und dem Streamingdienst (mit Fokus auf Empfehlungsdienste) thematisiert, wie etwa Verstärkungen von Abhängigkeiten im Nutzungsverhalten oder Wirkungen im Sinne der Idee von Filterblasen. **Im vierten Teil** werden die gemachten Erfahrungen auf weitere mögliche Kontexte übertragen und so verallgemeinert, indem die Lernenden weitere datengetriebene Anwendungen aus ihrem Alltag untersuchen, in denen Empfehlungsdienste eingesetzt werden, wie zum Beispiel bei

bestimmten Apps auf ihrem Handy. Im Rahmen einer Evaluation und Bewertung der Datenerhebung und -verarbeitung in den verschiedenen Beispielen können Vor- und Nachteile der Erhebung und Verarbeitung persönlicher Daten (z.B. Nutzungsdaten) diskutiert werden, um so den Lernenden eine Grundlage für reflektierten Entscheidungen hinsichtlich der Interaktion mit datengetriebenen Anwendungen dieser Art zu vermitteln.

## **Inhalt des Dokuments:**

<b>1</b>	<b>Überblick über den Verlauf des Unterrichtsmoduls:</b>	<b>1</b>
<b>2</b>	<b>Was meint ‚Datenbewusstsein‘?</b>	<b>2</b>
<b>3</b>	<b>Didaktische Kernidee: Förderung von Datenbewusstsein in diesem Unterrichtsmodul:</b>	<b>3</b>
<b>4</b>	<b>Lernvoraussetzungen:</b>	<b>3</b>
<b>5</b>	<b>Ziele:</b>	<b>3</b>
<b>6</b>	<b>Zentrale Leitfragen:</b>	<b>4</b>
<b>7</b>	<b>Zusammenfassender Überblick über das Unterrichtsmodul:</b>	<b>5</b>
<b>8</b>	<b>Überblick über den Unterrichtsverlauf:</b>	<b>5</b>
<b>9</b>	<b>Beschreibungen ausgewählter Materialien und weiterführende Hintergrundinformationen:</b>	<b>15</b>

## **2 Was meint ‚Datenbewusstsein‘?**

Das Konzept Datenbewusstsein fasst für den Informatikunterricht das Ziel, ein Bewusstsein und Verständnis für die Erhebung, Verarbeitung und Verwendung persönlicher Daten während der Nutzung datengetriebener digitaler Artefakte<sup>i</sup> (s. Glossar in Abs. 10) bei Lernenden zu fördern. Die Erhebung persönlicher Daten während einer solchen Interaktion geschieht durch a) aktive Eingabe von Informationen seitens des Nutzenden, b) durch Beobachtung und Tracking des Verhaltens sowie c) durch Verarbeitung bereits zuvor erhobener Daten. Dies kann unterteilt werden in die explizite Datenerhebung, also die mit der Handlung intendierte aktive und direkte Bereitstellung von Informationen durch den Nutzenden, und die implizite Datenerhebung, also durch nebenher zur eigentlichen Handlung ablaufende Prozesse, wie etwa Beobachtung, Tracking und Generierung durch Datenverarbeitung<sup>ii</sup> (detaillierte Beschreibung der Begrifflichkeiten in Abs. 10). Nutzende von datengetriebenen digitalen Artefakten sind sich oft der explizit erhobenen Daten bewusst, der implizit erhobenen Daten jedoch oft nicht bewusst. Die so erhobenen persönlichen Daten im Rahmen einer Interaktion mit einem datengetriebenen digitalen Artefakt können durch verschiedene Methoden verarbeitet werden, von einfachen Data Moves (s. Glossar in Abs. 10) bis zu Methoden des Maschinellen Lernens. Mit der Verarbeitung und Verwendung der Daten werden verschiedene Zwecke verfolgt. Dabei können erhobene Daten zum („technischen“) Betreiben von Funktionen des datengetriebenen digitalen Artefakts (primäre Zwecke) und/oder für darüberhinausgehende Zwecke oder etwa zur Untersuchung von Weiterentwicklungen des datengetriebenen digitalen Artefakts (sekundäre Zwecke) verarbeitet und verwendet werden (detaillierte Beschreibung der Begrifflichkeiten in Abs. 10). Primäre Zwecke sind dabei aus Sicht der Nutzenden zu verstehen und sekundäre Zwecke eher aus Sicht der Anbietenden (Was kann ein Anbieter mit den Daten sonst noch machen?). Im Sinne der verschiedenen Zwecke kann etwa ein digitaler Doppelgänger als modellhafte Repräsentation der Nutzenden konstruiert werden. Mit dem Konzept Datenbewusstsein sollte zum einen eine Aufmerksamkeit für datengetriebene digitale Artefakte sowie zum anderen ein Verständnis der Prozesse der expliziten und impliziten Erhebung und Generierung von Daten und der automatisierten Datenverarbeitungsprozesse zu primären oder sekundären Zwecken vermittelt werden. Die Schülerinnen und Schüler sollen also dazu befähigt werden, in einer Interaktion mit einem datengetriebenen digitalen Artefakt die Erhebung und Verarbeitung persönlicher Daten erkennen und analysieren zu können sowie daraufhin selbstbestimmte Entscheidungen für Interaktionen dieser Art treffen zu können.

### **3 Didaktische Kernidee: Förderung von Datenbewusstsein in diesem Unterrichtsmodul**

Zur Umsetzung der Ziele und damit zum Fördern des Datenbewusstseins der Lernenden werden die Facetten von Datenbewusstsein in den vier Teilen des Unterrichtsmoduls umgesetzt. Das gewählte Beispiel im ersten Teil beschreibt ein Interaktionssystem bestehend aus einem Nutzenden und einem Streamingdienst bzw. dessen Empfehlungsdienst als datengetriebenes digitales Artefakt sowie der Interaktion zwischen diesen. Durch ein Spiel zu personalisierten Filmempfehlungen erarbeiten die Lernenden **im ersten Teil** die Bedeutung von personalisierten Filmempfehlungen und welche Rolle dabei Informationen bzw. Daten über die Person spielen. Dabei entwickeln sie Ideen für die explizite und implizite Erhebung von persönlichen Daten für den primären Zweck der Verwendung dieser Daten, des Ermittels von personalisierten Filmempfehlungen. Diesen primären Zweck im Sinne der automatisierten Ermittlung von personalisierten Filmempfehlungen durch einen Empfehlungsdienst erarbeiten die Lernenden **im zweiten Teil** detaillierter. Dabei werden insbesondere auch die Konstruktion und Bedeutung des digitalen Doppelgängers eines Nutzenden hervorgehoben. Für einen sekundären Zweck der Verwendung und Verarbeitung der erhobenen Daten oder auch der digitalen Doppelgänger von Nutzenden wird **im dritten Teil** eine exemplarische, fiktive personalisierte Bezahlschranke thematisiert, in der verschiedene Aspekte der Wechselwirkungen in dem Interaktionssystem aufgegriffen werden. Dies veranlasst die Lernenden die Rolle der Daten und des Selbst in diesem exemplarischen Interaktionssystem zu reflektieren und die Erhebung und Verarbeitung von persönlichen Daten in einem solchen Interaktionssystem zu bewerten. **Im vierten Teil** werden die erlernten Kenntnisse zum Datenbewusstsein auf weitere Beispiele aus ihrem eigenen Alltag angewandt: Interaktion mit einem datengetriebenen digitalen Artefakt; explizite und implizite Datenerhebung; primäre und sekundäre Zwecke der Verwendung und Verarbeitung sowie Konstruktion eines digitalen Doppelgängers. Diese Kontexte werden anschließend reflektiert und kriteriengeleitet bewertet.

### **4 Lernvoraussetzungen:**

Dieses Unterrichtsmodul setzt keine besonderen Vorkenntnisse der Lernenden voraus. Es sollte jedoch eine grundlegende Erfahrung im Umgang mit dem Computer vorhanden sein. Außerdem sind grundlegende Vorstellungen des Datenbegriffs wünschenswert, entsprechende Einführungen könnten aber auch in diesem Modul integriert werden. Ein Verständnis von Künstlicher Intelligenz oder Maschinellern ist nicht nötig, im Gegenteil werden in diesem Modul Aspekte dessen bereits aufgegriffen – diese könnten auch in einer Adaption des Unterrichtsmoduls erweitert werden.

### **5 Ziele:**

In den vier Teilen des Unterrichtsmoduls werden im Wesentlichen folgende Ziele verfolgt:

- **Teil 1: Filmempfehlungen und Datenerhebung durch einen Empfehlungsdienst**
  - Die Lernenden erkennen die Bedeutung von personalisierten Filmempfehlungen, indem sie exemplarisch anderen Lernenden in mehreren Schritten Filmempfehlungen geben und diesen Prozess der Verbesserung dieser Filmempfehlungen reflektieren.
  - Die Lernenden unterscheiden die Begriffe der explizit und implizit erhobenen Daten und entwickeln Ideen dafür, welche Daten für die automatisierte Ermittlung von Filmempfehlungen explizit und implizit erhoben werden.
  - Die Lernenden begründen exemplarisch die Notwendigkeit der expliziten und impliziten Erhebung von persönlichen Daten sowie deren Verarbeitung zum Ermitteln personalisierter Filmempfehlungen beispielhaft für die Erstellung einer Startseite bei einem Streamingdienst (primärer Zweck).
- **Teil 2: Aufbau und Funktionsweise von Filmempfehlungsdiensten**
  - Die Lernenden erklären wesentliche Schritte zur automatisierten Ermittlung von personalisierten Filmempfehlungen basierend auf explizit und implizit erhobenen Daten (z.B.

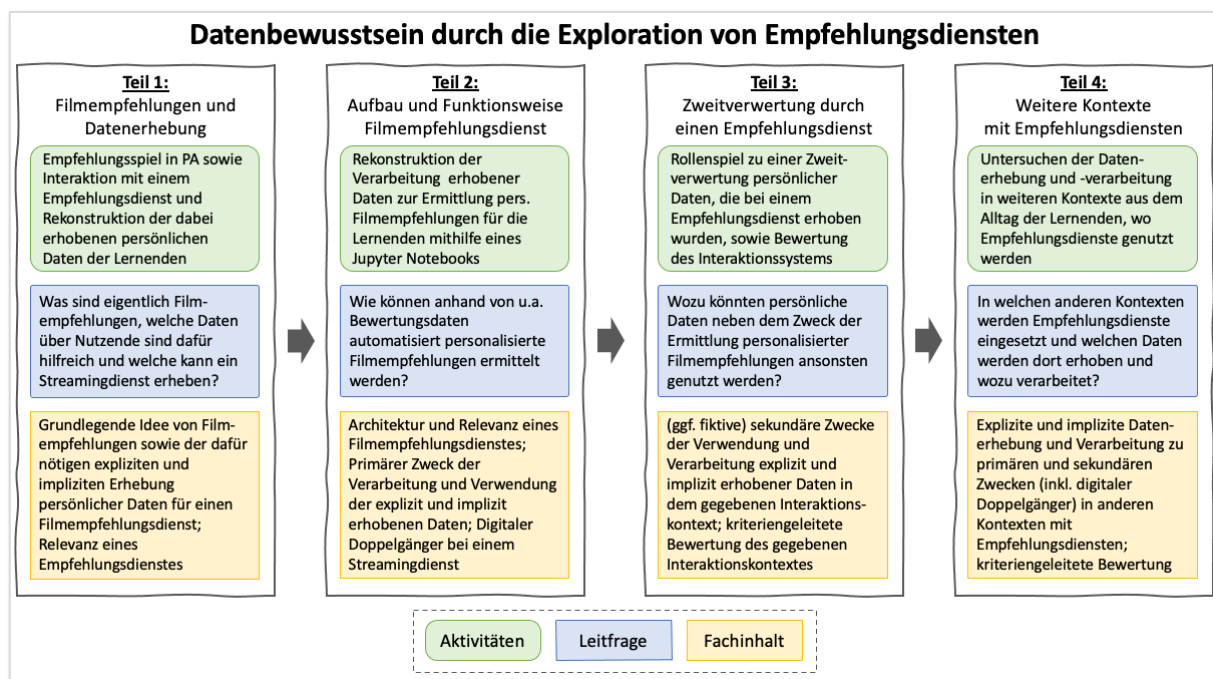
Nutzungsdaten), wobei sie auf die Grundidee des kollaborativen Filterns anhand des Verfahrens k-nearest-neighbors aus dem maschinellen Lernen eingehen.

- Die Lernenden beschreiben die Konstruktion eines digitalen Doppelgängers bei der Nutzung eines Streamingdienstes und begründen dessen Relevanz für einen Empfehlungsdienst.
- **Teil 3: Zweitverwertung durch einen Empfehlungsdienst**
  - Die Lernenden erkennen den Vorschlag der Zweitverwertung für eine personalisierte Bezahlschranke als Idee für einen sekundären Zweck der Verwendung und Verarbeitung der erhobenen Daten bzw. des digitalen Doppelgängers, indem sie diese Idee aus verschiedenen Perspektiven in einer Diskussionsrunde (bzw. Rollenspiel) beleuchten.
  - Die Lernenden erkennen mehrere Aspekte der Wechselwirkung zwischen Nutzendem und Streamingdienst, indem sie diese im Rahmen der Diskussionsrunde aufgreifen und bewerten.
  - Die Lernenden beschreiben aus ihrer individuellen Perspektive Handlungsoptionen bzgl. der Interaktion mit einem Streamingdienst mit einem Empfehlungsdienst, indem sie eine Bewertung bzgl. der Erhebung und Verarbeitung von Daten im Rahmen der Interaktion mit einem Streamingdienst vornehmen und ein Fazit dazu formulieren.
- **Teil 4: Weitere Kontexte mit Empfehlungsdiensten**
  - Die Lernenden wenden ihre gelernten Kenntnisse zum Datenbewusstsein auf weitere Beispiele eines datengetriebenen digitalen Artefakts aus ihrem Alltag an, indem sie an diesem Beispiel die explizite und implizite Datenerhebung, deren Verwendung und Verarbeitung zu primären und exemplarischen sekundären Zwecken sowie die Konstruktion von digitalen Doppelgängern identifizieren und beschreiben.
  - Die Lernenden nehmen eine begründete Bewertung der Erhebung und Verarbeitung von Standortdaten in den thematisierten Beispielen vor, indem sie zum Beispiel auf den Kompromiss zwischen einem datensparsamen Verhalten und das Nutzen von individuellen oder gesellschaftlichen Vorteilen eingehen.

## 6 Zentrale Leitfragen:

- **Teil 1: Filmempfehlungen und Datenerhebung durch einen Empfehlungsdienst**
  - Was sind personalisierte Filmempfehlungen, welche Daten über Nutzende sind dafür hilfreich und welche kann ein Streamingdienst explizit und implizit erheben?
- **Teil 2: Aufbau und Funktionsweise von Filmempfehlungsdiensten**
  - Wie können anhand von u.a. Bewertungsdaten (anhand explizit und implizit erhobener Daten) automatisiert personalisierte Filmempfehlungen ermittelt werden? (primärer Zweck)
  - Wie wird ein digitaler Doppelgänger von einem Nutzenden konstruiert und welche Rolle spielt dieser für die Funktionsweise im Empfehlungsdienst?
- **Teil 3: Zweitverwertung durch einen Empfehlungsdienst**
  - Wozu könnten persönliche Daten neben dem Zweck der Ermittlung personalisierter Filmempfehlungen ansonsten genutzt werden?
  - Welche Bedeutung hat die Rolle der Daten im Rahmen der Nutzung von Streamingdiensten mit Empfehlungsdiensten hinsichtlich Aspekte der Wechselwirkung zwischen Nutzendem und dem Streamingdienst?
- **Teil 4: Weitere Kontexte mit Empfehlungsdiensten**
  - In welchen anderen Kontexten werden Empfehlungsdienste eingesetzt, welche Daten werden dort erhoben und wozu werden sie verarbeitet?
  - Welche Handlungsoptionen hat ein Nutzender in diesen Kontexten?

## 7 Zusammenfassender Überblick über das Unterrichtsmodul



## 8 Überblick über den Unterrichtsverlauf:

In der folgenden Tabelle wird der Unterrichtsverlauf beschrieben. Dazu finden Sie in der Spalte „Inhalt“ Beschreibungen des Unterrichtsverlaufs mit jeweiligen didaktischen Bemerkungen, die sich in der Regel auf die Förderung von Datenbewusstsein konzentrieren. Zu den jeweiligen Phasen des Unterrichtsmoduls werden in der Tabelle die entsprechenden Materialien angegeben.

Phase	Inhalt	Ziele	Material
<b>Teil 1: Filmempfehlungen und Datenerhebung durch einen Empfehlungsdienst</b>			
1a	<p><b>Einführung in den Interaktionskontext und Problematisierung:</b></p> <p>Es wird mit einem Unterrichtsgespräch in den Kontext eingestiegen, wozu beispielsweise das beiliegende Beispielbild einer Startseite von der Streamingplattform Netflix gezeigt wird. Das Unterrichtsgespräch wird im Wesentlichen an drei Leitfragen orientiert:</p> <p>1. <i>Was ist ein Streamingdienst?</i> – Anhand des Beispielbildes erklären die Lernenden, was ein Streamingdienst ist, beispielsweise indem sie dabei auch auf ihre Erfahrungen mit dem Beispiel Netflix eingehen.</p> <p>2. <i>Wie ist eine solche Startseite aufgebaut?</i> – Die Lernenden erklären darauf aufbauend wie die Startseite bei dem Beispiel Netflix aufgebaut ist (lässt sich auch von ähnlichen Streamingplattformen übertragen). Insbesondere sollte dabei deutlich werden, dass eine solche Startseite personalisiert ist (Netflix beschreibt etwa, dass es für jeden Nutzenden eine komplett eigene Startseite gibt und sich keine doppeln würde). Die Personalisierung</p>	<ul style="list-style-type: none"> <li>Aktivieren von Vorkenntnissen zu Streamingdiensten und zur Personalisierung bzw. personalisierten Empfehlungen (i.d.R. aus Nutzerperspektive)</li> <li>Wecken von Neugierde an der Funktionsweise eines Empfehlungsdienstes</li> </ul>	<p>Beispielbild einer Netflix-Startseite;</p> <p>ggf. Aussage: „Für alle Nutzer:innen existieren eigene personalisierte Startseiten, keine doppeln sich.“;</p> <p>ggf. Bild von Netflix-Sortierung</p>

	<p>passiert bei Netflix etwa über die Reihenfolge der Kategorien (Zeilen) und innerhalb derer die Reihenfolge der Inhalte – zusätzlich werden auch die Bilder zu einem Inhalt abhängig vom Nutzenden gewählt. Es handelt sich also um Empfehlungen von Inhalten, die auf der Startseite angezeigt werden.</p> <p>3. <i>Wie kommen Filmempfehlungen zustande?</i> – Dies stellt die Leitfrage für die nächste Phase dar und wird noch nicht diskutiert.</p> <p><u>Didaktischer Kommentar:</u> Das Beispiel aufgrund seiner Bekanntheit – insbesondere auch bei den Lernenden – gewählt. Dennoch gibt es Lernende, die dazu weniger Erfahrungen haben als andere, sodass in dieser Phase alle Lernenden mit ihren mehr oder weniger vorhandenen Erfahrungen abgeholt werden sollen. Das Unterrichtsgespräch soll gerade auf die letzte Leitfrage hinarbeiten und diese motivieren; in anderen Worten soll problematisiert werden, wie solche personalisierten Filmempfehlungen ermittelt werden können.</p>		
1b	<p><b>Empfehlungsspiel: Bedeutung von Filmempfehlungen und zugehörige Datenerhebung</b></p> <p>In Partnerarbeit bearbeiten die Lernenden das AB1, das eine Art „Empfehlungsspiel“ darstellt. Bei der Partnerarbeit dürfen die Lernenden nur in bestimmten Aufgaben miteinander sprechen und nicht darüber hinaus. Durch das gegenseitige Geben von Filmempfehlungen werden die Lernenden darin eingeführt, was eine personalisierte Filmempfehlung ausmacht. Sie erarbeiten, welche Informationen dafür hilfreich sind, und entwickeln Ideen für entsprechende Daten, die von einem Streamingdienst explizit und implizit erhoben werden könnten (Aufg. 4 auf AB1).</p> <p>In der Auswertung dieser Erarbeitungsphase wird ausgewertet, welche Empfehlungen besser passten (s. Aufg. 3 auf AB1) und welche Fragen dafür besonders hilfreich waren (s. Aufg. 2 auf AB1). Außerdem werden die Ideen für die explizit und implizit erhobenen Daten (s. Aufg. 4 auf AB1) gesammelt.</p> <p><u>Didaktischer Kommentar:</u> Mit dieser Erarbeitungsphase sollte den Lernenden deutlich werden, worauf es ankommt, gut passende personalisierte Filmempfehlungen zu ermitteln. Dafür sollten in der PA möglichst Lernende zusammenarbeiten, die sich nicht gut kennen. In dieser Partnerarbeit wechseln die Lernenden zwischen der Anbieter- sowie Nutzerperspektive und reflektieren dabei den eigenen Prozess zum Ermitteln von personalisierten Filmempfehlungen. Die Begrifflichkeiten der <i>expliziten und impliziten</i></p>	<ul style="list-style-type: none"> <li>• Lernen/üben der Begrifflichkeiten zur expliziten und impliziten Datenerhebung sowie Identifizierung dessen in dem gegebenen Interaktionskontext</li> <li>• Erkennen des Konzepts der Personalisierung anhand personalisierter Filmempfehlungen als primärer Zweck durch die Reflektion des eigenständigen Gebens von personalisierten Filmempfehlungen</li> <li>• (Entwickeln erster Ideen zum Ermitteln personalisierter Filmempfehlungen)</li> <li>• Ggf. Wiederholung der Unterscheidung von Daten und Information</li> </ul>	AB1



	<p><i>Datenerhebung</i> werden in der Aufgabe 4 des AB1 eingeführt. Sofern diese für die Lernenden zuvor unbekannt waren, könnte vor Aufgabe 4 eine Besprechung zur konzeptionellen Einführung der Begriffe stattfinden (s. Begriffserklärung in Abs. 10). In der Auswertung der Erarbeitungsphase sollte dann einerseits auf die korrekte Einordnung von explizit und implizit erhobene Daten geachtet werden; andererseits sollte die Unterscheidung von Daten und Information berücksichtigt werden (z.B. sollte bei Daten nicht „Interesse“ stehen; die Frage ist doch viel mehr, welche Daten nötig sind, um Vermutungen zum Interesse einer Person aufzustellen).</p>		
<b>Teil 2: Aufbau und Funktionsweise von Filmempfehlungsdiensten</b>			
<b>2a</b>	<p><b>Überleitung zum Empfehlungsdienst im Jupyter Notebook:</b></p> <p>Mit Aufgreifen der vorherigen Leitfrage „Wie kommen Filmempfehlungen zustande?“ wird eine Überleitung zu diesem Teil 2 gemacht. Zuvor haben die Lernenden Ideen erarbeitet, welche Daten für das Finden von personalisierten Filmempfehlungen hilfreich sein könnten; offen ist jedoch noch, wie diese Daten denn verarbeitet werden, um solche Filmempfehlungen automatisiert zu ermitteln (also das, was die Lernenden bei AB1 im Kopf gemacht haben). Die Lehrkraft führt in das Jupyter Notebook und den Umgang damit ein. Dieses vorbereitete Jupyter Notebook hält einen „fertigen“ Filmempfehlungsdienst bereit (s. detaillierte Beschreibung in Abs. 9.4).</p> <p><u>Didaktischer Kommentar:</u></p> <p>Die noch offene Leitfrage aus Teil 1 wurde erst zur Hälfte beantwortet: Die Lernenden haben nun Ideen, welche Daten für einen Empfehlungsdienst explizit und implizit erhoben werden könnten; offen ist jedoch noch, wie damit dann automatisiert das gemacht werden soll, was die Lernenden intuitiv gemacht haben, um in der Partnerarbeit von AB1 personalisierte Filmempfehlungen zu geben. Je nach Vorerfahrungen der Lernenden kann hier mehr oder weniger auf viel auf die Bedienung eines Jupyter Notebooks eingegangen werden. Es sollte auf jeden Fall der Zellaufbau eines Jupyter Notebooks sowie das Ausführen von Zellen geklärt werden. In der Überleitung wird die Verbindung zwischen dem vorherigen eigenständigen Geben von personalisierten Filmempfehlungen zu einer digitalen Variante geschlagen.</p>	<ul style="list-style-type: none"> <li>Reflexion der vorherigen Leitfrage zum Ermitteln von personalisierten Filmempfehlungen, wofür zunächst nur Ideen für die Datenerhebung entwickelt wurden</li> </ul>	Jupyter Notebook
<b>2b</b>	<p><b>Interaktion und erste Erkundung eines Empfehlungsdienstes:</b></p> <p>Die Lernenden bearbeiten das gegebene Jupyter Notebook, in dem sie eigene Bewertungen von Filmen angeben und automatisiert personalisierte</p>	<ul style="list-style-type: none"> <li>Erfahrbar machen der expliziten und impliziten Datenerhebung während der Interaktion</li> </ul>	Jupyter Notebook

	<p>Filmempfehlungen angezeigt bekommen (Interaktion mit dem gegebenen Empfehlungsdienst).</p> <p>Zunächst rekonstruieren die Lernenden, welche Daten durch das gegebene Empfehlungsdienst explizit und implizit erhoben bzw. herangezogen wurden. Bei der Auswertung dessen sollte festgehalten werden, dass die Bewertungen sowohl eine Arte „explizite Bewertung“ explizit erhoben als auch Daten über das Anschauen von Filmen als eine Art „implizite Bewertung“ implizit erhoben werden. Bei dieser Auswertung wird ebenfalls das Konzept des digitalen Doppelgängers besprochen und gesichert, welches in dem Jupyter Notebook bereits dargestellt ist und somit transparent gemacht wird. Außerdem kann darauf eingegangen werden, welche weiteren Daten zur Verfügung stehen und somit zur Ermittlung von personalisierten Empfehlungen herangezogen werden können.</p> <p><u>Didaktischer Kommentar:</u></p> <p>Zunächst bekommen die Lernenden einen funktionierenden Empfehlungsdienst gegeben und rekonstruieren daran, welche Daten diesem Empfehlungsdienst vorliegen. Dabei sollte ein Bezug zu ihren Ideen aus AB1 hergestellt werden. Dabei sollten deutlich werden, welche explizit und implizit erhobenen Daten in diesem Jupyter Notebook tatsächlich herangezogen werden, damit in der nachfolgenden Rekonstruktion der Funktionsweise des Empfehlungsdienstes die Datengrundlage und dessen Bedeutung klar ist.</p> <p>Zu Beginn des Jupyter Notebooks werden bereits automatisiert Filmempfehlungen gegeben. Für die folgenden Phasen der Rekonstruktion steht die Leitfrage im Mittelpunkt, wie diese Empfehlungen zustande kommen.</p>	<ul style="list-style-type: none"> <li>• Identifizieren explizit sowie implizit erhobener Daten</li> <li>• Konzeptionelle Idee des digitalen Doppelgängers kennenlernen</li> <li>• Konstruktion eines digitalen Doppelgängers verstehen und auf den Kontext des Filmempfehlungsdienstes anwenden</li> </ul>	
2c	<p><b>Rekonstruktion der automatisierten Ermittlung von personalisierten Filmempfehlungen:</b></p> <p>Schrittweise erarbeiten die Lernenden anhand des Jupyter Notebooks, wie personalisierte Filmempfehlungen exemplarisch basierend auf den Bewertungs- bzw. Nutzungsdaten systematisch ermittelt werden können. Zunächst wird dies anhand eines reduzierten, händisch zu verarbeiteten Datenbeispiels rekonstruiert und anschließend mithilfe eines ML-Verfahrens umgesetzt.</p> <p><i>Reduziertes Beispiel von „Hand“:</i></p> <p>Anhand des reduzierten Beispiels mit fünf Nutzenden und drei Filmen rekonstruieren die Lernenden die systematische Idee der kollaborativen Filtermethodik basierend auf Ähnlichkeiten von Nutzenden (d.h. es werden ähnliche Nutzende gesucht und anhand derer Bewertungs- /</p>	<ul style="list-style-type: none"> <li>• Erkennen der Verknüpfung von digitalen Doppelgängern als Mittel der Konstruktion des digitalen Doppelgängers eines Nutzenden exemplarisch für das Ermitteln personalisierter Filmempfehlungen</li> <li>• Kennenlernen der automatisierten Ermittlung von personalisierten Filmempfehlungen auf der Basis des kollaborativen Filterns</li> </ul>	<p>Jupyter Notebook</p> <p>Ggf. AB-Sprinter-aufgabe</p>



<p>Nutzungsdaten wird entschieden, ob ein dritter Film einem Nutzenden empfohlen werden sollten (s. detaillierte Beschreibung in Abs. 9.1 und 9.4).</p> <p><i>Größeres Beispiel mit ML-Methode:</i> Die Lernenden erarbeiten in dem Jupyter Notebook anschließend stark geleitet die automatisierte Berechnung personalisierter Empfehlungen, wozu k-nearest-neighbor als ML-Methode genutzt wird, um automatisiert und basierend auf den Daten zu allen verfügbaren Filmen ähnliche Nutzende zu ermitteln (s. detaillierte Beschreibung in Abs. 9.4). Als Grundlage für diese Bearbeitung dient gerade die vorherige reduzierte Ermittlung personalisierter Filmempfehlungen.</p> <p>Als optionale Sprinteraufgabe können sich die Lernenden über das sogenannte Cold-Start Problem Gedanken machen (s. AB dazu). Die Lernenden befassen sich mit der Frage, inwiefern der primäre Zweck erfüllt werden kann, wenn ein Nutzender noch keinen digitalen Doppelgänger hat bzw. wenn ein neuer Inhalt hinzugefügt wird.</p> <p>Zwischen der händischen und der automatisierten Ermittlung von Filmempfehlungen kann bei Bedarf eine Zwischensicherung eingeschoben werden, um insbesondere die Idee der Ähnlichkeit und die darauf basierende Ermittlung der Vorhersage für ein Bewertungsmaß sowie Entscheidung für oder gegen die Empfehlung eines Films zu sichern.</p> <p>In der Auswertung werden Vorgehen zur Ermittlung der ähnlichen Nutzenden, das Aufstellen eines Modells (und dessen Bedeutung) sowie die Ermittlung der Vorhersagewerte (für Ratings) besprochen, sodass anschließend exemplarisch Wege zur Auswahl von Filmen für die personalisierte Empfehlung reflektiert werden können. Außerdem sollte insbesondere auch darauf eingegangen werden, dass das kollaborative Filtern im Endeffekt ein „Verbinden und Vergleichen“ von digitalen Doppelgängern ist, sodass der eigene digitale Doppelgänger auch bei den Filmempfehlungen anderer Personen eine Rolle spielt (und andersherum).</p> <p>Ggf. kann in der Auswertung die Idee des Cold-Start Problems aufgegriffen werden, beispielsweise indem dies von einzelnen Lernenden, die sich damit befasst haben, zusammen mit ihren Lösungsideen vorgestellt wird.</p> <p><u>Didaktischer Kommentar:</u> Zunächst wird sich auf zwei Filme beschränkt. Es werden also nur die Daten zu zwei Filmen betrachtet, um damit Empfehlungen „nicht-automatisiert“ zu</p>	<p>anhand persönlicher Bewertungs- und Nutzungsdaten</p> <ul style="list-style-type: none"> <li>• Kennenlernen der Nutzung eines ML-Verfahrens am Beispiel der Methodik k-nearest-neighbors</li> </ul>
--	--

	<p>ermitteln. Dies dient zunächst zur Vereinfachung, um die Idee der Ähnlichkeit einzuführen. Die hierbei verfolgte Idee ist vergleichbar mit dem Prinzip der ML-Methodik k-nearest-neighbors, die darauffolgend für die automatisierte Ermittlung personalisierter Filmempfehlungen genutzt wird. Diese Methode ist ein Beispiel, mit dem Zusammenhänge zwischen Nutzenden ermittelt werden können. Es werden also gerade zu einem festgelegten Nutzenden ein individuelles Cluster von ähnlichen Nutzenden ermittelt. Diese haben zu ihm/ihr scheinbar ein ähnliches Filminteresse. Dabei gilt die Grundannahme, dass die Filme, die die ähnlichen Nutzende interessant fanden, auch für die festgelegte Person vermutlich interessant sind. Für das ML-Verfahren könnten auch noch weitere Aspekte von ML integriert werden, wie etwa die Performance von ML-Modellen und dafür eine entsprechende Testphase, die in diesem Jupyter Notebook zugunsten des Ziels zum Verstehen des Prinzips der Ermittlung personalisierter Filmempfehlungen reduziert wurde. Sofern bereits Vorkenntnissen zu ML bestehen, könnte an dieser Stelle etwa auch ein Bezug zu diesen Aspekten hergestellt werden.</p>		
<b>Teil 3: Zweitverwertung durch einen Empfehlungsdienst</b>			
<b>3a</b>	<p><b>Überleitung zur exemplarischen Zweitverwertung:</b></p> <p>Im Unterrichtsgespräch wird die Rolle der Daten in dem exemplarischen Kontext der Interaktion mit einem Streamingdienst, der ein Empfehlungsdienst nutzt, reflektiert – begrenzt auf die bisherigen betrachteten Facetten: Die Lernenden fassen die explizite und implizite Erhebung von Daten in diesem Kontext zusammen, beschreiben die Rolle des digitalen Doppelgängers sowie die Verarbeitung der erhobenen Daten sowie die Verwendung des digitalen Doppelgängers für den primären Zweck des Ermitteln von personalisierten Filmempfehlungen. Die Legitimität der Datenerhebung und -verarbeitung (und damit einhergehenden Notwendigkeit) kann an dieser Stelle sinnstiftend diskutiert werden; wenn man das Feature der personalisierten Startseite schätzt, ist eine Erhebung und Verarbeitung persönlicher Daten nötig. Dabei können etwa zwei Aspekte hervorgehoben werden: (1) Reduzieren der Informationsüberflutung (Finden einer personalisierten Auswahl von Produkten, die dem Nutzer/der Nutzerin auf der Plattform präsentiert wird.) sowie (2) Verbesserung des Nutzungserlebnisses (Positive Erfahrungen mit der Plattform führen dazu, dass die Nutzenden z.B. mehr Filme schauen und mit dem Streamingdienst somit mehr Umsatz generiert wird.)</p> <p>Bei der Bewertung kann dann auf die Zweitverwertung übergeleitet werden, da diese für</p>	<ul style="list-style-type: none"> <li>• Reflektieren der Rolle von Daten in dem gegebenen Interaktionskontext hinsichtlich der zuvor thematisierten Facetten von Datenbewusstsein (explizite und implizite Datenerhebung, digitaler Doppelgänger, primärer Zweck)</li> <li>• Bewertung des primären Zwecks der Erhebung und Verarbeitung der Daten für die Anzeige einer personalisierten Startseite</li> </ul>	

	<p>die Bewertung ebenfalls relevant sein kann – kann zur Veranschaulichung der Betrachtung von sekundären Zwecken zusätzlich zu den primären Zwecken dienen.</p> <p><u>Didaktischer Kommentar:</u> Mit dieser Zwischenreflexion soll zum einen das Verständnis für die Rolle der Daten in diesem Kontext hinsichtlich der Facetten von Datenbewusstsein gefestigt werden. Zum anderen wird durch diese Reflexion ermöglicht eine größere Perspektive einzunehmen, mit der die jeweiligen Aspekte verknüpft und die Bedeutung der Daten erkannt werden kann, was für eine spätere Evaluation nötig ist.</p>		
<b>3b</b>	<p><b>Einführung in das Rollenspiel als Diskussionsrunde bzgl. einer Zweitverwertung:</b> Nun stellt sich die Frage, wofür ein Anbieter eines Streamingdienstes die Daten bzw. den digitalen Doppelgänger zweitverwerten könnte (sekundärer Zweck). Diese Frage wird im Unterrichtsgespräch aufgeworfen.</p> <p><u>Didaktischer Kommentar:</u> In den vorherigen Teilen wurde die explizite und implizite Datenerhebung sowie deren Verarbeitung für den primären Zweck der Konstruktion einer Startseite thematisiert, wobei die Rolle des digitalen Doppelgängers aufgegriffen wurde. In diesem Teil wird nun exemplarisch ein sekundärer Zweck aufgegriffen und zusammen mit Aspekten der Wechselwirkung zwischen Nutzendem und Streamingdienst diskutiert.</p>	<ul style="list-style-type: none"> <li>• Erkennen der Betrachtung eines sekundären Zwecks und wahrnehmen als durchaus relevanten Aspekt zusätzlich zu den aus Nutzungsperspektive nützlichen Aspekten der Personalisierung</li> <li>• Vornehmen eines Perspektivwechsels zur Anbieterperspektive</li> </ul>	
<b>3c</b>	<p><b>Rollenspiel zur Zweitverwertung:</b> <i>Ausgangssituation:</i> Die Lehrkraft stellt das Rollenspiel mit der Ausgangssituation vor, welche auf dem AB3 notiert ist. Die Lernenden können direkt Fragen dazu stellen. Anschließend wird das AB3 an die Lernenden ausgeteilt. Die Lernenden bilden Gruppen, in denen sie sich mit den Rollen für das Rollenspiel befassen.</p> <p><i>Vorbereitungs- &amp; Spielphase:</i> Die Lernenden bearbeiten in Gruppen die Rollen und bekommen dazu mit dem AB3 entsprechende Rollenkarten. [Folgende Rollen gibt es: Geschäftsführung, Leitung der technischen Abteilung, Leitung einer Forschungsabteilung, Leitung der Kundenbetreuung] Die Gruppen bearbeiten die zuvor beschriebene Ausgangssituation aus dem Blickwinkel der jeweiligen Rolle und u.a. eine Haltung zu diesem Vorschlag mit Argumenten festhalten (s. AB3). In der Spielphase vertreten die Lernenden ihre Rolle und diskutieren den Vorschlag der</p>	<ul style="list-style-type: none"> <li>• Vertiefen der Kenntnisse zur Erhebung und Verarbeitung der persönlichen Daten zu dem zuvor thematisierten primären Zweck</li> <li>• Vornehmen eines Perspektivwechsels zur Anbieterperspektive</li> <li>• Kennenlernen verschiedener Aspekte hinsichtlich der Wechselwirkung zwischen Nutzendem und Streamingdienst (bezogen auf den Einsatz eines Empfehlungsdienstes)</li> <li>• Abwägen verschiedener Perspektiven auf Zwecke der Verwendung und</li> </ul>	AB3 in 4 Varianten

	<p>Ausgangssituation. Diese Diskussion wird von einem Moderator/einer Moderatorin geleitet, die insbesondere die Beteiligung aller Rollen an der Diskussion sicherstellen soll. Die Lernenden sind angehalten ihre Argumente unterzubringen und ihre besprochene Haltung zu vertreten. Alle anderen Lernende sind Beobachter und machen sich zu den Argumenten und den Perspektiven Notizen.</p> <p><i>Auswertungsphase:</i> Zu Beginn der Auswertung findet eine Abstimmung zur Frage statt, ob der Vorschlag umgesetzt werden sollte oder nicht. (Ergebnis sollte notiert werden) Anschließend bewerten die Lernenden die Rollen, kommentieren diese und reflektieren, welche Positionen und Argumente besonders wichtig waren. Es wird nun final abgestimmt, ob der Vorschlag umgesetzt werden sollte oder nicht. Die Lernenden formulieren abschließend ein persönliches Fazit zur Erhebung sowie Verarbeitung und Verwendung der Daten durch einen Streamingdienst. In der Diskussion sollten insbesondere die Aspekte bzgl. der Wechselwirkung zwischen Nutzendem und Streamingdienst bezogen auf den Einsatz eines Empfehlungsdienstes aufgegriffen werden.</p> <p><u>Didaktischer Kommentar:</u> In dem Rollenspiel wird ein sekundärer Zweck, der im Kontext der Streamingdienste vielleicht eher als fiktiv angesehen werden kann. Dennoch bietet die Diskussionsrunde zu dieser Situation den Anlass über verschiedene Aspekte bzgl. der Wechselwirkung zwischen Nutzendem und Streamingdienst zu diskutieren, wie etwa Effekte der Filterblasen oder bzgl. verstärkter Abhängigkeiten der Nutzenden. Die Lernenden sollen die Situation aus den verschiedenen Perspektiven betrachten und reflektieren. Dabei sollen die Lernenden eine begründete Haltung dazu entwickeln und die mögliche Verwendung und Verarbeitung der erhobenen Daten zum sekundären Zweck bewerten.</p>	<p>Verarbeitung persönlicher Daten exemplarisch für einen fiktiven sekundären Zweck</p>	
3d	<p><b>Vertiefungs- oder Hausaufgabe (optional):</b> Geleitet durch Fragen reflektieren die Lernenden den Einsatz von Empfehlungsdiensten in ihrem Alltag und entwickeln dazu kontextabhängig eine Haltung.</p>	<ul style="list-style-type: none"> <li>Vorbereitende Reflexion der Rolle der Daten sowie Bewertung des Einsatzes von Empfehlungsdiensten</li> </ul>	AB-HA
<b>Teil 4: Weitere Kontexte mit Empfehlungsdiensten</b>			
4a	<p><b>Sammlung weiterer Beispiele mit Empfehlungsdiensten:</b> Im Plenum werden Beispiele für datengetriebenen digitale Artefakte aus dem Alltag der Lernenden gesammelt, in denen Empfehlungsdienste eingesetzt werden. Die können beispielsweise Feeds bei Social Media Plattformen, andere Plattformen aus dem</p>	<ul style="list-style-type: none"> <li>Anwendung des Gelernten über Empfehlungsdienste auf eigenen Alltagserfahrungen</li> <li>Identifizieren von datengetriebenen digitalen Artefakten, die</li> </ul>	Tafel, Beamer o.ä.

	<p>Streamingbereich oder auch eine Suchmaschine sein.</p> <p><u>Didaktischer Kommentar:</u> Den Lernenden soll es hiermit ermöglicht werden, dass sie ihre Vorstellungen über Empfehlungsdienste auf ihren Alltag insofern anwenden, dass sie diese Beispiele im Sinne datengetriebener digitale Artefakte identifizieren können, sodass erste Reflexionsprozesse angestoßen werden. Als Stütze kann dienen, dass Empfehlungsdienste im Sinne der individuellen Interaktion mit datengetriebenen digitalen Artefakten oft im Kontext der Personalisierung von Inhalten bzw. dessen Darstellung oder Auflistung wiederzufinden ist.</p>	Empfehlungsdienste einsetzen	
4b	<p><b>Anwendung der Facetten von Datenbewusstsein auf ausgewählte Beispiele aus ihrem Alltag:</b> Mithilfe des AB4 rekonstruieren die Lernenden nun für jeweils verschiedene Beispiele aus der vorherigen Sammlung (s. Phase 4a) die Rolle der Daten bei der Interaktion mit dem jeweiligen datengetriebenen digitalen Artefakt. Die Bearbeitung des AB4 findet wahlweise in Paaren oder Kleingruppen statt. Für die Bearbeitung sollten die Lernenden sich das jeweils gewählte datengetriebene digitale Artefakt anschauen können bzw. dazu recherchieren können. Die Wahl des zu untersuchenden Kontexts sollten die Lernenden selbst vornehmen können, es kann jedoch darauf geachtet werden, dass in der Lerngruppe insgesamt verschiedene Beispiele gewählt werden.</p> <p><u>Didaktischer Kommentar:</u> Die Lernenden sollen die Facetten von Datenbewusstsein (explizite und implizite Datenerhebung; primäre und sekundäre Zwecke sowie digitaler Doppelgänger) auf Beispiele aus ihrem Alltag anwenden. Damit wird der Transfer des Gelernten auf den eigenen Alltag sowie das Verknüpfen dessen mit eigenen Alltagserfahrungen gefördert. Außerdem soll dadurch angeregt werden, dass die Lernenden ihre alltäglichen Interaktionen mit datengetriebenen digitalen Artefakten reflektieren und erfahren, wie das Gelernte in alltäglichen Interaktionen angewendet werden könnte.</p>	<ul style="list-style-type: none"> <li>• Anwenden der Facetten von Datenbewusstsein auf weitere Kontexte aus dem Alltag der Lernenden</li> <li>• Vertiefung und Übung des Verständnisses bezüglich der Facetten von Datenbewusstsein</li> <li>• Kennenlernen der Anwendung dieser Perspektive für Alltagssituationen</li> </ul>	AB4, digitales Endgerät
4c	<p><b>Auswertung und Reflexion der weiteren Kontexte:</b> Ausgewählte Beispiele werden von den Lernenden im Plenum vorgestellt. Dabei sollen sie den Empfehlungsdienst vorstellen sowie ihre Entdeckungen/Vermutungen hinsichtlich der expliziten und impliziten Datenerhebung, der primären und sekundären Zwecke für die Verwendung und Verarbeitung dieser Daten sowie die Rolle des digitalen Doppelgängers erklären. Zu</p>	<ul style="list-style-type: none"> <li>• Kennenlernen verschiedener Kontexte aus Sicht der Facetten von Datenbewusstsein</li> <li>• Bewertung der Rolle von Daten in den verschiedenen Kontexten aus der Perspektive auf die</li> </ul>	Ggf. Beamer o.ä.

<p>den vorgestellten Ergebnissen sollten Rückfragen gestellt sowie Ergänzungen der anderen Lernenden vorgenommen werden können.</p> <p>Abschließend sollen die Lernenden auch eine Bewertung der Datenerhebung und -verarbeitung in den jeweiligen Kontexten vornehmen, die durchaus sowohl auf einer individuellen als auch auf einer gesellschaftlichen Betrachtungsebene verortet werden können. Dafür soll im Unterrichtsgespräch diskutiert werden, inwiefern ein Nutzender in den jeweiligen Kontexten Handlungsmöglichkeiten hat sowie sich für Handlungen entscheiden kann (z.B. bestimmte Datenerhebungen oder Zwecke einschränken). Bei dieser Sicherung sollte hervorgehoben werden, dass in den verschiedenen Kontexten hinsichtlich der verschiedenen Arten der Datenerhebung sowie der verschiedenen Zwecke unterschiedliche Bewertungen und Haltungen dazu möglich sind, ein Nutzender sich für verschiedene Handlungen entscheiden kann und ggf. Einfluss auf die Datenerhebung und -verarbeitung vornehmen kann. Wichtig dabei sollte sein, dass die Lernenden ihre eigene Haltung dazu entwickeln und sie verstehen, dass es durchaus auch wichtig ist, eine eigene Entscheidung zu treffen (gerade keine Resignation fördern!).</p> <p>Im Rahmen dieser Phase könnte ggf. auch eine Zusammenfassung des gesamten Unterrichtsmoduls vorgenommen werden.</p> <p><u>Didaktischer Kommentar:</u> Es sollte auf eine korrekte Anwendung der Facetten von Datenbewusstseins geachtet werden, um Fehlvorstellungen zu vermeiden. Durch diese Phase sollen die Lernenden verschiedene Kontexte aus der Perspektive des Konzepts Datenbewusstsein (d.h. mit den Facetten die Rolle der Daten beschreiben und damit evaluieren können) betrachten und dadurch das Gelernte mit Erfahrungen aus ihrem Alltag verknüpfen können. Die Bewertung dieser Kontexte zielt dann darauf ab, dass die Lernenden erkennen, dass es einerseits verschiedene Handlungsoptionen bzw. Handlungsentscheidungen gibt und andererseits es wichtig ist eine eigene Entscheidung diesbezüglich zu treffen. Dabei sollte dringend beachtet werden, dass Lernenden weder ein Verhalten noch eine Meinung vorgeschrieben wird. Gleichmaßen sollte vermieden werden, dass Lernende eine Art Resignation gegenüber der Datenerhebung und -verarbeitung durch datengetriebene digitale Artefakte in ihrem Alltag entwickeln und sie durchaus handlungsfähig sind.</p>	<p>Facetten von Datenbewusstsein sowie Entwicklung einer eigenen Haltung diesbezüglich</p> <ul style="list-style-type: none"> <li>• Reflektieren der Handlungsmöglichkeiten und -fähigkeit des Nutzenden in den jeweiligen Kontexten, insb. mit Bezug auf die Rolle der Daten</li> </ul>	
--	--	--



## 9 Beschreibungen ausgewählter Materialien und weiterführende Hintergrundinformationen:

### 9.1 Empfehlungsdienste im Allgemeinen (in engl.: Recommender System)

Ein Empfehlungsdienst verfolgt das Ziel die Menge aller vorhandenen Items (z.B. Filme, Musiktitel, Shopping-Produkte, ...) auf eine Vorauswahl (Empfehlungen) einzuschränken, um den Nutzer:innen bei der Entscheidungsfindung zu unterstützen. Dem Nutzer/der Nutzerin sollten also nicht alle Items angezeigt werden, sondern nur eine Auswahl an Items, für die sich der Nutzer potenziell interessieren könnte, um eine Informationsüberflutung zu umgehen. Die Anbietenden des Dienstes zielt damit auf eine Gewinnmaximierung ab, indem der Nutzer/die Nutzerin „neue und interessante“ Items „entdeckt“. Dadurch werden die Nutzer:innen zu längeren und häufigeren Zugriffen (Steigerung der Nutzungszeit) angeregt, wodurch sie mehr Daten hinterlassen und womöglich der Umsatz durch Käufe oder Werbungen gesteigert werden kann.

Im Wesentlichen gibt es inhaltsbasierte (content-based), kollaborative (collaborative) und hybride Methoden zum Filtern der Items. Beim kollaborativen Filtern werden ähnliche Nutzer:innen identifiziert, um dann Empfehlungen basierend auf deren Daten (bspw. Filmbewertungen) zu ermitteln (hier etwa: Mittelwerte der Bewertungen der ähnlichen Nutzer:innen). Beim inhaltsbasierten Filtern werden Daten herangezogen, welche inhaltliche Informationen über die Produkte enthalten bzw. zumindest operationalisieren (z.B. Tags, Genres, Wortvorkommen in Textbeschreibungen). Das hybride Filtern verbindet verschiedene Methoden des kollaborativen und inhaltsbasierten Filterns – i.d.R. nacheinander.

### 9.2 Empfehlungsdienste bei Streamingdiensten

Zwei Zitate zu Empfehlungsdiensten bei Netflix und Spotify:

*“A problem we face is that our catalog contains many more videos than can be displayed on a single page and each member comes with their own unique set of interests. Thus, a general algorithmic challenge becomes how to best tailor each member’s homepage to make it relevant, cover their interests and intents, and still allow for exploration of our catalog.”* - Netflix TechBlog (<https://netflixtechblog.com/learning-a-personalized-homepage-aa8ec670359a>)

*“Spotify has created engines to control and manage everything from your personal best home screen to carefully chosen and organized playlists like Discover Weekly, and continues to explore new ways to understand music, and why people listen to one song or genre over another. All this is achieved with a combination of different recommender systems.”* (<https://www.linkedin.com/pulse/how-spotify-recommender-system-works-daniel-roy-cfa/>)

In diesem Unterrichtsmodul liegt ein Fokus auf Streamingdienste, wodurch vor allem Plattformen wie Netflix und Spotify im Mittelpunkt stehen. Trotzdem ist dieser Markt stetig am Wachsen: Amazon Prime Video, Disney+, Apple Music oder Amazon Music sind nur einige der Wettbewerber. Schon seit dem Aufkommen von modernen Streamingdiensten arbeiten diese an folgender Frage: Wie kann einem Kunden/einer Kundin möglichst maßgeschneiderte (personalisierte) Produktempfehlungen angeboten werden und wie können somit möglichst viele Kunden zum Bezahlen im jeweiligen Preismodell motiviert werden? Für die Berechnung solcher Empfehlungen spielen verschiedene Faktoren eine entscheidende Rolle: Der grundsätzliche Aufbau (Layout) der Plattform, die Möglichkeiten zur Interaktion für den Kunden/die Kundin und beispielsweise auch das soziale Umfeld, welches sich über verschiedene Benutzer:innen hinweg bildet.

Netflix steht mit seinen über 15000 Filmen und Serien<sup>1</sup> und über 200 Millionen zahlenden Abonnenten weltweit<sup>2</sup> vor der Herausforderung, dass es seinen Nutzer:innen nur eine begrenzte Anzahl an Film- und Serientiteln vorschlagen und auf der Startseite anzeigen kann. Ein vergleichbares Bild ergibt sich bei Spotify: mit 70 Millionen Songs<sup>3</sup> und 365 Millionen monatlich aktive Nutzer:innen<sup>4</sup> findet sich dort zwar ein anderes Produkt wieder, der Aufbau Plattform-Seite und die Datenbasis sind jedoch ähnlich. Die folgenden Illustrationen verdeutlichen die Ähnlichkeit der beiden Plattformen. Auf der linken Seite befindet sich das Layout von Spotify's Oberfläche, das rechte Bild beschreibt selbigen Aufbau bei Netflix:



Bei diesen Startseiten werden sowohl Reihen als auch Spalten entsprechend an den Nutzer/die Nutzerin angepasst sortiert. Beide Übersichten verdeutlichen, wie wertvoll der Platz auf der Startseite ist und wie relevant damit der Einsatz eines Empfehlungsdienstes wird. Das Geschäftsmodell von Netflix als auch von Spotify ist abonnementbasiert. Dabei gibt es jeweils verschiedene Abo-Pläne, welche im Einzelnen jedoch nicht das Angebot erweitern, sondern z.B. die Anzahl der Nutzenden pro Account verändern. Das Ziel ist es also, die Anzahl der zahlenden Kunden zu maximieren und Abonnements immer wieder zu verlängern. Dies wird dadurch erreicht, möglichst passende Vorschläge für Filme/Musik zu realisieren. Neben Streamingdiensten mit vollem Zugriff auf das komplette Angebot nach Zahlung einer monatlichen Rate gibt es auch andere Geschäftsmodelle, wie etwa bei Amazon Prime Video. Diese Plattform stellt nach einem Abonnement ebenfalls einen Teil seines Film- und Serienangebots zur Verfügung (ca. 13000 Titel). Darüber hinaus gibt es jedoch Medien, welche nach wie vor durch die Zahlung eines einmaligen Betrages freigeschaltet werden müssen (ca. 25000 Titel). Dabei kann es sein, dass die ersten Staffeln einer Serie im Abonnementpreis enthalten sind, ab einer bestimmten Staffel jedoch eine zusätzliche Zahlung fällig wird.

### 9.3 Erhebung von Daten bei der Nutzung von Streamingdiensten

Das Nutzungsverhalten der Nutzer:innen von Streamdiensten ist essenziell, um das Geschäftsmodell und damit den Umsatz stetig zu optimieren. Dabei spielen grundsätzlich jegliche Arten von Interaktionen der Nutzer:innen eine Rolle. Dies fängt bei einfachen Feedbackmechanismen an, wie etwa der "Gefällt mir"-Button, und geht weiter zum Nutzungsverhalten über angeschauter Filme und angehörte Musik. Dies ist nur

<sup>1</sup> <https://www.comparitech.com/blog/vpn-privacy/netflix-statistics-facts-figures/>

<sup>2</sup> <https://de.statista.com/statistik/daten/studie/196642/umfrage/abonnenten-von-netflix-quartalszahlen/>

<sup>3</sup> <https://de.statista.com/statistik/daten/studie/378806/umfrage/anzahl-der-verfuegbaren-songs-auf-spotify/>

<sup>4</sup> <https://de.statista.com/infografik/13769/monatlich-aktive-nutzer-und-zahlende-abonnenten-von-spotify-weltweit/>

der Anfang von dem, wie Streamingdienste Daten erheben und verarbeiten. Ein Ziel ist es, einen digitalen Doppelgänger eines jeden Nutzenden zu konstruieren. Auf Basis dessen können dann Empfehlungen und weitere Verarbeitungen und Verwendungen dessen erfolgen.

Ein detailliertes Beispiel der Datenverwendung ist die sogenannte "Completion Rate" (deutsch: Abschlussrate)<sup>5</sup> einer Serie. Dabei kann folgende Frage beantwortet werden: "Wie viele Nutzer:innen, die Serie XY angefangen haben zu schauen, schauten diese bis zu Staffel 3?". Fällt die Antwort auf diese Frage z.B. auf 30% aus, öffnet sich ein weiterer Pool an Fragen: "An welchem Punkt haben die meisten Nutzer:innen die Serie abgebrochen?" und "Wie lange und mit wie großen Pausen zwischen den Folgen haben die 30% die Serie bis Staffel 3 angesehen?". Wenn nun ein gewisser Anteil eine Serie zu Ende geschaut hat, liegt es nahe, dass der Serienproduzent (wie inzwischen etwa Netflix) eine weitere Staffel produzieren wird – und im anderen Fall eventuell eine neue Serie produziert, die das Feedback der Nutzer:innen aufgreift, die die Serie nicht zu Ende gesehen haben.

Generell gilt: Bei allen Streamingdiensten werden zunächst alle verfügbaren (persönliche) Daten explizit und implizit erhoben und zu einem gegebenen Zeitpunkt verarbeitet, um das Geschäftsmodell weiter zu optimieren. Darüber hinaus können potenziell auch einzelne Daten an Dritte weitergegeben werden (z.B. Facebook oder andere Werbepartner).<sup>6</sup> Es ist jedoch wichtig anzumerken, dass gerade bei Netflix die monetäre Einnahmequelle auf den Abonnements und nicht auf der Weitergabe von Daten basiert. Diese Entscheidung wurde getroffen, um einen möglich Abgang von Nutzer:innen zu verhindern.<sup>7</sup>

Empfehlungsdienste finden sich heute in fast allen bekannten Online-Plattformen wieder. So nutzt Google solche Systeme beispielsweise bei der Google Suche zum Anzeigen der Suchergebnisse oder auf der Plattform YouTube zum Erzeugen einer personalisierten Startseite bzw. geben personalisierter Videoempfehlungen. Um jeweils zum Nutzer/zur Nutzerin passende Produkte vorzuschlagen, nutzt auch Amazon solchen Modellen. Auch die Bibliothek des KIT in Karlsruhe arbeitet mit ähnlichen Verfahren auf Basis dessen Literaturempfehlungen ausgegeben werden können.

Dabei spielen im Grunde immer ähnliche Daten eine Rolle. Auf der einen Seite die verschiedenen Produkte, welche auf der Plattform angeboten werden (Filme, Videos, Bücher, ...). Auf der anderen Seite stehen die Nutzer:innen und ermöglichen die Erhebung und Generierung wertvoller Daten durch die Interaktion mit den jeweiligen digitalen Artefakten (Schreiben von Rezensionen, Ansicht von Produkten, Verbindungen zu anderen Nutzer:innen, ...).

## 9.4 Empfehlungsdienst in diesem Unterrichtsmodul

### 9.4.1 Explizite und implizite Bewertungen

Bewertungen für Produkte, wie bspw. Filme, können explizit oder implizit vorliegen. Explizit sind Bewertungen dann, wenn der Nutzer/die Nutzerin das Produkt direkt beurteilt, bspw. über ein Gefällt-mir-Button oder eine Sternbewertung. Dadurch gibt der Nutzer/die Nutzerin i.d.R. seine Meinung von dem Produkt bzw. sein Interesse an dem Produkt aktiv zum Ausdruck. Implizite Bewertungen werden nicht von dem Nutzer/der Nutzerin direkt angegeben. Das bedeutet, dass bestimmte Daten erhoben, generiert und verarbeitet werden, welche bspw. als Operationalisierung für das Interesse an dem Produkt dienen können. Beispiele für implizite Bewertungen sind: Hat der Nutzer/die Nutzerin das Produkt gekauft? Hat er oder sie den Film vollständig geschaut oder früher beendet? Wurde der Film mehrmals geschaut? Wurde sich das Produkt gemerkt (Merklisten)?

---

<sup>5</sup> <https://neilpatel.com/blog/how-netflix-uses-analytics/>

<sup>6</sup> z.B. <https://www.spotify.com/de/legal/privacy-policy/>

<sup>7</sup> [https://geschaeftsmodell-workshop.de/geschaeftsmodell/beispiele/netflix-business-model#Das\\_Business\\_Model\\_Canvas\\_von\\_Netflix\\_als\\_Powerpoint\\_und\\_PDF](https://geschaeftsmodell-workshop.de/geschaeftsmodell/beispiele/netflix-business-model#Das_Business_Model_Canvas_von_Netflix_als_Powerpoint_und_PDF)

In dem Datensatz, welcher im Unterricht eingesetzt wird (Beschreiben siehe Abschnitt 9.4.2), wurden neben den expliziten Bewertungen zusätzlich implizite Beurteilungen generiert, um die beiden Konzepte zu veranschaulichen. Dabei sind implizite Filmbewertungen im Filmdatensatz die binäre Antwort auf die Frage, ob Nutzer:innen einen Film zu Ende geschaut hat oder nicht. Dieses neue Attribut wurde künstlich, jedoch auf Basis der vorhandenen Bewertungen angelegt. Mit einer Wahrscheinlichkeit von 85% wurden Bewertungen mit mehr als vier Sternen auf den Status "1" (Film zu Ende angeschaut) gesetzt. Liegt die Bewertung unter vier Sterne fand dies nur in 40% der Fälle statt. Alle übrigen Bewertungen erhielten den Status "0" (Film nicht zu Ende angeschaut). Im Anschluss wurden 25% der expliziten Bewertungen entfernt, um die Relevanz der impliziten Bewertungen darzustellen.

#### 9.4.2 Genutzte Ratingdaten und Aufbereitung dieser für das Unterrichtsmodul

In diesem Unterrichtsmodul ziehen wir reale Nutzungsdaten von Nutzer:innen der Plattform MovieLens (movielens.org) heran. Auf der Plattform angemeldete Nutzer:innen können dort u.a. Filme bewerten und Filmempfehlungen bekommen. Es ist also ein Empfehlungsdienst eingebettet. Die Betreiber haben Bewertungsdaten öffentlich zugänglich gemacht<sup>8</sup>. Für das Unterrichtsmodul haben wir diese Daten aus Performancegründen verkleinert, sodass wir lediglich ca. 50000 Bewertungen von ca. 5000 Usern zu insgesamt ca. 600 Filmen nutzen.

Die Filme, die bewertet werden können, wurden manuell nach einer subjektiven Einschätzung des Bekanntheitsgrades sowie unter Einbezug von IMDB-Hitlisten ausgewählt. Filme, welche unter den möglichen Empfehlungen erscheinen, haben eine Mindestanzahl an Bewertungen erhalten. Die Nutzer:innen in dem Datensatz wurden so ausgewählt, dass sie alle mindestens einen der Filme, welche über das Empfehlungsmodul bewertet werden können, selbst bewertet haben. Grundsätzlich lag das Hauptaugenmerk bei der Datengenerierung auf der Balance zwischen der Performance des Modells und den für die Berechnungen verfügbaren Hardware-Ressourcen.

Für das Unterrichtsmodul nutzen wir die Daten in Form von Datentabellen (DataFrames). Diese sind in den nachfolgenden Bildern dargestellt. Sie umfassen in der ersten Datentabelle Informationen über die Filme (Titel, Genre, Erscheinungsjahr) und in der zweiten Datentabelle gerade die explizit und implizit erhobenen *Bewertungen* der Nutzer:innen (Ids, Datum, Uhrzeit, Fertig\_Angeschaut, Rating). Sowohl die Nutzer:innen als auch die Filme bekommen eine ID zugewiesen, mit der sie eindeutig identifiziert werden.

---

<sup>8</sup> Referenz zum Projekt: <https://dl.acm.org/doi/10.1145/2827872;> Daten: <https://grouplens.org/datasets/movielens/>

	movielid	title	genres	Erscheinungsjahr
0	121364	Memories of My Melancholy Whores (2011)	Comedy Drama Romance	2011
1	177261	Jackals (2017)	Horror Thriller	2017
2	188769	Polish Legends: Twardowsky (2015)	Sci-Fi	2015
3	203831	Anabolic Life (2017)	Thriller	2017
4	190203	Kaleidoscope (2017)	Thriller	2017
...	...	...	...	...
5666	170993	Mini's First Time (2006)	Comedy Crime Drama	2006
5667	191137	La Désintégration (2012)	Drama	2012
5668	46850	Wordplay (2006)	Documentary	2006
5669	115680	Time Lapse (2014)	Crime Drama Sci-Fi Thriller	2014
5670	198503	The Isle (2018)	Horror	2018

5671 rows x 4 columns

	userid	movielid	Datum	Uhrzeit	Fertig_Angeschaut	rating
0	1	29	18-01-2017	16:56:07	1	4.5
2	1	77561	13-08-2015	15:40:40	1	4.0
3	1	72998	13-08-2015	16:03:43	1	4.0
4	1	70293	13-08-2015	15:58:02	1	4.0
5	1	68952	18-08-2019	03:18:04	1	2.5
...	...	...	...	...	...	...
1460217	9080	3979	27-05-2018	15:55:41	1	1.5
1460218	9080	3949	19-05-2018	09:02:01	1	4.5
1460219	9080	3418	28-05-2018	15:09:28	1	4.0
1460220	9080	3354	19-05-2018	09:38:53	0	2.0
1460222	9080	195163	21-01-2019	16:04:36	0	2.0

1210706 rows x 6 columns

### 9.4.3 Jupyter Notebooks im Allgemeinen

Jupyter Notebooks ermöglichen das Ausführen von Pythoncode in Echtzeit mit Hilfe sogenannter Codezellen. Die Ergebnisse werden jeweils unter der aktuellen Zelle angezeigt. Erläuterungen zu Aufgaben zwischen den einzelnen Codezellen können auf Basis der Auszeichnungssprache Markdown realisiert werden. Der gesamte Code kann dabei jederzeit manipuliert werden, was das spielerische Herantasten an Programmierung ermöglicht. Zu beachten ist, dass die Ausführung im Falle des Empfehlungsdienstes in diesem Modul nicht auf der lokalen Maschine, sondern auf dem zentralen Server der Universität Paderborn stattfindet.

### 9.4.4 Vorbereitetes Jupyter Notebook

Für dieses Unterrichtsmodul haben wir Bibliotheken entwickelt und ein Jupyter Notebook für den Unterricht vorbereitet. In diesem Jupyter Notebook (*Empfehlungsdienst für Filme.ipynb*) werden zunächst die Daten automatisiert eingelesen und ein Empfehlungsdienst am Beispiel von Netflix beschrieben. Anschließend ermitteln die Lernenden nach Eingabe von eigenen Bewertungen eigene Filmempfehlungen über einen bereits implementierten Empfehlungsdienst. Dieser basiert auf dem k-Nearest-Neighbor Algorithmus (Erklärung siehe unten) und nutzt als Basis seiner Vorschläge die vorgefilterten Bewertungsdaten. In der Standardeinstellung arbeitet der Dienst ausschließlich mit expliziten Bewertungen. Über einen Schalter im Code kann dieser jedoch die Empfehlungen auch basierend auf impliziten Bewertungen berechnen. Im nächsten Schritt wird die Frage behandelt, welche Daten erhoben wurden. Dafür können die Lernenden eine User-Movie-Tabelle (Erklärung siehe unten) aufrufen. Danach beschäftigt sich das Notebook mit der Frage, wie personalisierte Empfehlungen automatisiert berechnet werden können. Zur Visualisierung wird ein 2-dimensionales Koordinatensystem herangezogen. Darin können zwei Filme ausgewählt werden. Zu diesen Filmen werden dann alle vergebenen Bewertungen visualisiert. Somit sind einfache Analysen in Relation zur eigenen Bewertung möglich.

Alle nötigen Befehle werden in gelben Hinweisboxen erklärt. In blauen Boxen werden die Aufgaben detailliert formuliert und es werden grüne Einführungs- bzw. Erklärboxen eingeschoben.

Beim ersten Zugriff auf das Verzeichnis mit den Jupyter Notebooks muss man einen Login erstellen, mit dem zu einem späteren Zeitpunkt wieder an die letzte Bearbeitung angeschlossen werden kann. Andernfalls würden die Bearbeitungen nach schließen des Jupyter Notebooks gelöscht werden. Das Verzeichnis ist unter folgendem Link zu erreichen: <https://ddi-jupyter.cs.upb.de/empfehlungsdienst/>

#### 9.4.5 *k-Nearest-Neighbor Algorithmus zur Erstellung eines Modells:*

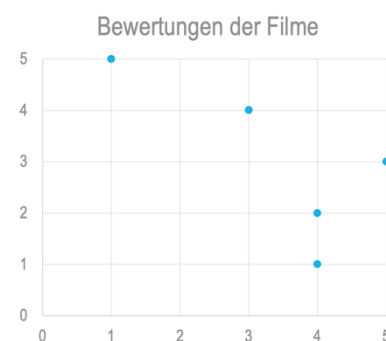
Der k-nearest-neighbor Algorithmus wird in dem vorbereiteten Jupyter Notebook mithilfe der Bibliothek sklearn zur Erstellung eines Modells verwendet. Dieses Modell kann anschließend zum Entscheiden von ähnlichen Nutzer:innen (eigentlich: nächsten Nachbarn) auf Basis von Daten aus z.B. einem Streamingdienst angewendet werden. Die konkrete Funktionsweise des Algorithmus wird im Unterricht nicht im Detail vermittelt, es soll lediglich die Idee der Vorgehensweise verstanden werden. An der Stelle der Modellerzeugung wird bewusst eine Black-Box gesetzt, um die im Rahmen dieser Unterrichtsreihe gesetzten Lernziele zu erreichen und keine Überforderung zu erzeugen.

##### Beispiel:

In der nebenstehenden Tabelle ist ein Minimalbeispiel gegeben. Es gibt Bewertungsdaten von fünf Nutzer:innen zu zwei Filmen. Anhand dieses Beispiels kann bereits das Suchen nach den k nächsten Nachbarn erklärt werden. Gesucht sind zum Beispiel zwei Nutzer:innen, die ähnlich zum markierten User 5 sind. Das sind dann etwa die User 1 und 4, da diese die kleinste Abweichung in ihren Bewertungen der beiden Filme zu User 5 haben. Konkret heißt das, dass die Abstände zwischen der Tabellenzeile von User 5 und denen von User 1 und 4 am kleinsten sind, die Differenz also möglichst klein ist. (Randnotiz: Mathematisch nutzen wir in unserer Umsetzung die euklidische Metrik für die Bestimmung von Abständen.) Dies kann auch in dem nebenstehenden Koordinatensystem visualisiert werden. Die Bewertungen zu Film A entsprechen den Werten auf der x-Achse und zu Film B denen auf der y-Achse. So stellt jeder Punkt im Koordinatensystem einen User da, der beide Filme bewertet hat. Mit dieser Vorgehensweise können zu einem gewählten User die k ähnlichsten Nutzer:innen einfach identifiziert werden. So können im Koordinatensystem etwa beliebig viele Nutzer:innen hinzugefügt werden. Um nun anhand dieser ähnlicher Nutzer:innen zu User 5 für einen dritten Film C herausfinden, ob dieser empfohlen werden sollte, wird eine Prediction ermittelt. Die Prediction wird etwa durch den Mittelwert der Bewertungen des Films C der ähnlichen Nutzer:innen ermittelt. In dem Beispiel der nebenstehenden Tabelle ist dies dann 4,5 (Mittelwert von 4 und 5). Das heißt, wenn User 5 den Film C schauen und bewerten würde, würde er wahrscheinlich eine Bewertung von 4,5 abgeben. Dem User 5 sollte der Film C also durchaus empfohlen werden. Dieses (hier stark reduzierte) Verfahren ist auf eine große Anzahl von Nutzer:innen und Filmen übertragbar.

	Film A	Film B	Film C
User 1	5	3	4
User 2	3	4	
User 3	1	5	4
User 4	4	2	5
User 5	4	1	

	Film A	Film B
User 1	5	3
User 2	3	4
User 3	1	5
User 4	4	2
User 5	4	1





#### 9.4.6 User-Movie-Tabelle als hilfreiche Tabelle zur Ermittlung von Empfehlungen

Die user-movie-Tabelle ist eine Datentabelle, die in diesem Kontext Filmbewertungen (Zellen) von Nutzer:innen (Zeilenweise userIds) zu den jeweiligen Filmen (Spaltenweise Filmtitel) aufführt. Diese Tabelle ist für den Empfehlungsdienst recht zentral, anhand dieser wird beispielsweise das vom k-Nearest-Neighbor Algorithmus ermittelte Modell mit einer aus der Tabelle erstellten sparse-Matrix berechnet. Ähnliche User werden also mithilfe der Abstände zwischen den jeweiligen Zeilen in dieser Tabelle bestimmt.

#### 9.5 Zusatzmaterialien und weitere Fakten

Bild einer Startseite von Netflix:

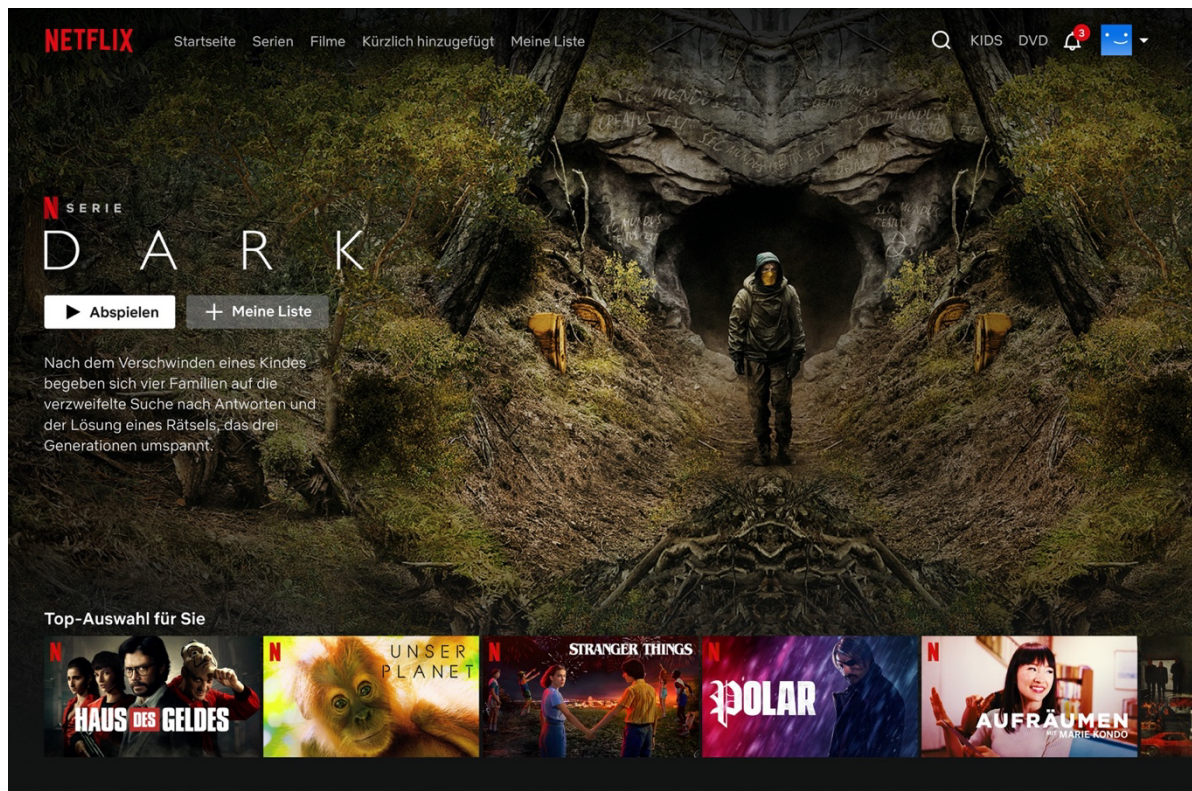


Abbildung 1: Bild von <https://about.netflix.com/de/company-assets>

Die Abbildung der Startseite wird als Motivation und Einstieg für die personalisierte Startseite verwendet. Weitere Informationen hierzu finden sich im Verlaufsplann.

### 10 Glossar relevanter Begriffe:

#### 10.1 Datenbegriff:

Wie im Kernlehrplan NRW für Informatik in Klasse 5 und 6 dokumentiert (s. Inhaltsfeld „Information und Daten“) kann zwischen Information und Daten unterschieden werden: Daten sind (digital) repräsentierte Informationen und können etwa in Informatiksystemen gespeichert und verarbeitet werden. Hinsichtlich des Konzepts Datenbewusstsein ist besonders hervorzuheben, dass Kontexte, Phänomene oder etwa auch Interessen, Emotionen oder Handlungen einer Person anhand ausgewählter Merkmale modelliert werden. Gerade die persönlichen Daten, welche für Datenbewusstsein eine große Rolle spielen, entstammen einem Kontext, wo die jeweilige Person involviert ist oder war. Bei der Thematisierung von Daten sollte der Modellierungsaspekt nicht vernachlässigt werden, da die Kontexte, Phänomene oder Personen nicht vollständig repräsentiert werden. Im Kontext des Datenbewusstseins bedeutet das gerade, dass die erhobenen, persönlichen Daten kein Abbild einer Person darstellen, sondern lediglich modellhaft anhand

ausgewählter Merkmale repräsentiert. Dadurch kann auch eine verzerrte Repräsentation einer Person entstehen. Außerdem sollte beachtet werden, dass gewisse Informationen, wie etwa Emotionen oder Interesse, im Sinne der Merkmale für die erhobenen Daten operationalisiert werden, was oft nicht trivial ist (Was heißt es, wenn ein datengetriebenes digitales Artefakt das „Interesse“ der Nutzenden speichert? – Was ist das? Woran wird das fest gemacht?)

### 10.2 Digitale Artefakte und datengetriebene digitale Artefakte:

Im Konzept Datenbewusstsein wurde der Begriff der *datengetriebenen digitalen Artefakte* (ddA) eingeführt. Dieser beschreibt eine spezielle Art von digitalen Artefakten. *Digitale Artefakte* sind ein Sammelbegriff für digitale Werkzeuge, Computersysteme aller Art, ihre Bestandteile, ihre Verbindung untereinander. Sie umfassen Sowohl Hardware, Software, Daten und Objekte sowie Algorithmen und Datenstrukturen. *Datengetriebene digitale Artefakte* sind dann digitale Artefakte, die sich selbst oder ihre Rückmeldung in der Interaktion mit diesem durch die Verarbeitung erhobener Daten verändert. Diese nutzen dann oft zum Beispiel auch Methoden des Maschinellen Lernens.

### 10.3 Architektur und Relevanz (Duale Natur digitaler Artefakte):

Die duale Natur digitaler Artefakte oder auch Dualität beschreibt eine analytische Trennung von Aspekten eines digitalen Artefakts (s.o.). Ein digitales Artefakt kann dieser Auffassung nach aus der Perspektive auf die Architektur und auf die Relevanz beschrieben werden<sup>iii</sup>. *Architektur* meint alle technologischen Strukturen und Mechanismen. *Relevanz* hingegen meint Intentionen, Funktionen, Meinungen, Interpretationen und der Kontext eines digitalen Artefakts.

### 10.4 Explizit und implizit erhobene Daten:

Im Konzept Datenbewusstsein wurden die Begrifflichkeiten der *explizit* und *implizit erhobenen Daten* eingeführt. Diese stehen in der Regel in der Verbindung zum Nutzenden - oft stellen sie personenbezogene Daten dar. Die explizit erhobenen Daten sind jene, die der Nutzende mit seiner Handlung intendiert zu erstellen, also direkt und aktiv eingegeben bzw. erzeugt hat. Darüber sind sich Nutzende in der Regel bewusst. Dies sind zum Beispiel bei Social Media Plattformen gepostete Texte und Bilder, bei einer Suchmaschine etwa der Suchbegriff oder beim Telefonieren über das Mobilfunknetz die Telefonnummer desjenigen, den man anrufen möchte. Im Gegensatz dazu, werden implizit erhobene Daten indirekt durch Beobachtung (Tracking) oder Verarbeitung bereits erhobener Daten nebenher zur eigentlichen Handlung des Nutzenden erhoben und generiert. Dieser Datenerhebung sind sich Nutzende oft nicht bewusst. Im Beispiel der Social Media Plattform sind dies etwa Likes und Klicks, bei der Suchmaschine etwa Klicks auf Suchergebnisse oder beim Telefonieren über das Mobilfunknetz etwa Standortdaten der verbundenen Basisstationen.

### 10.5 Primäre und sekundäre Zwecke der Verarbeitung und Verwendung:

Im Konzept Datenbewusstsein wurden die Begrifflichkeiten der *primären* und *sekundären Zwecke* der Verarbeitung und Verwendung erhobener Daten eingeführt. Diese beziehen sich auf die Verarbeitung und Verwendung von Daten über einen Nutzenden, die bei der Nutzung von datengetriebenen digitalen Artefakten erhoben werden. *Primäre* und *sekundäre Zwecke* beziehen sich auf die Intention, mit der diese zuvor erhobenen Daten verarbeitet und verwendet werden. Primäre Zwecke umfasst, dass die erhobenen Daten dazu verarbeitet und verwendet werden, um das datengetriebenen digitalen Artefakten mit den Features anbieten zu können. Diese beziehen sich auf einer Nutzerperspektive auf die Verarbeitung und Verwendung: Die Daten werden verarbeitet und verwendet, um Nutzenden Features anbieten zu können. Im Beispiel der Suchmaschine ist dies etwa das Anzeigen von Suchergebnissen. Auch inbegriffen wäre, wenn die Suchergebnisse personalisiert geordnet werden. Im Sinne des Features für den Nutzenden würde dies bedeuten, dass der Nutzende gerade die Ergebnisse angezeigt bekommt, die für ihn idealerweise relevant sind. Sekundäre Zwecke bedeutet, dass die Daten verarbeitet und verwendet werden, um andere/weitere Zwecke zu verfolgen – z.B. weitere wirtschaftliche oder wissenschaftliche Zwecke. Diese „Zweitverwertung“

der Daten bezieht sich auf einer Anbieterperspektive auf die Verarbeitung und Verwendung der erhobenen Daten: Wozu kann ein Anbieter eines datengetriebenen digitalen Artefakts die erhobenen Daten noch nutzen? Im Kontext von Streamingdiensten (z.B. Spotify) könnte dies etwa umfassen, dass Nutzungsdaten (z.B. gehörte Musik) zur Analyse der Emotionen der Nutzenden verwendet werden.

### 10.6 Data Moves:

Mit Data Moves werden Datenoperationen beschrieben. Diese umfassen etwa folgende<sup>iv</sup>:

- *Filtern*: Bilden einer Teilmenge der Daten
- *Gruppieren*: Daten in Teilgruppen unterteilen
- *Zusammenfassen*: Aggregieren von Daten
- *Berechnen*: Neue Attribute ausgehend von existierenden Daten erstellen (z.B. Ausgehend von zwei Spalten eine dritte Spalte erzeugen)
- *Merging/Joining*: Datensätze zusammenführen
- *Reorganisieren*: zum Beispiel ändern der Darstellung der Daten

---

<sup>i</sup> Höper, L. & Schulte, C., (2021). Datenbewusstsein: Aufmerksamkeit für die eigenen Daten. In: Humbert, L. (Hrsg.), *INFOS 2021 – 19. GI-Fachtagung Informatik und Schule*. Gesellschaft für Informatik, Bonn. (S. 73-82). DOI: 10.18420/infos2021\_f235

<sup>ii</sup> Höper, L. & Schulte, C., (2021). Datenbewusstsein im Kontext digitaler Kompetenzen für einen selbstbestimmten Umgang mit datengetriebenen digitalen Artefakten. In: Gesellschaft für Informatik (Hrsg.), *INFORMATIK 2021*. Gesellschaft für Informatik, Bonn. (S. 1623-1632). DOI: 10.18420/informatik2021-136

<sup>iii</sup> Schulte, C., & Budde, L. (2018). A Framework for Computing Education: Hybrid Interaction System: The need for a bigger picture in computing education. In *Proceedings of the 18th Koli Calling International Conference on Computing Education Research* (S. 1-10).

<sup>iv</sup> Erickson, T., Finzer, B., Reichsman, F., & Wilkerson, M. (2018). Data Moves: one key to data science at school level. In *Proceedings of the International Conference on Teaching Statistics (ICOTS-10)* (Vol. 6).