

Methodik für Datenprojekte im Informatikunterricht

von Lukas Höper, Sven Hüsing, Hülya Malatyali, Carsten Schulte und Lea Budde

Das Projekt ProDaBi

Daten und somit auch Big Data spielen in der heutigen und zukünftigen Welt ohne Frage eine zunehmend wichtige Rolle. Künstliche Intelligenz und maschinelles Lernen erlauben die Verarbeitung von immer größer vorkommenden Datenmengen – von Big Data. Der Bereich Data Science beeinflusst somit sowohl den privaten Alltag jedes Einzelnen als auch die Wirtschaft und die Gesellschaft. Die Integration von Data Science Education in den Informatikunterricht gewinnt aktuell eine besondere Relevanz. Das **Projekt Data Science und Big Data** (kurz: ProDaBi) der Universität Paderborn (vgl. ProDaBi, 2018ff.), das von der Deutschen Telekom Stiftung initiiert und gefördert wurde, widmet sich dem Thema *Big Data und Data Science in der Schule*.

Für die Entwicklung und Durchführung von Unterrichtsprojekten mit Daten ist eine Methodik notwendig, die die Schritte in einem Datenprojekt strukturiert und verständlich macht. In diesem Beitrag schlagen die Autorinnen und Autoren eine Methodik für Datenprojekte im Schulunterricht vor und beschreiben dazu drei erprobte Unterrichtsprojekte.

Unterrichtsmethodik für Datenprojekte

Es gibt bereits mehrere Modelle, die die Phasen von Datenprojekten verdeutlichen. Im Folgenden wird eine Auswahl an Modellen vorgestellt und eine eigene Methodik vorgeschlagen, die die Kritikpunkte an den bestehenden Modellen für den Schulunterricht um eine erkenntnisorientierte Sichtweise ergänzen soll.

Ein Modell vor allem für das Bearbeiten von statistischen Problemen erläutern Chris Wild und Maxine Pfannkuch (vgl. Wild/Pfannkuch, 1999, S.225) unter

dem Namen PPDAC, das aus den zyklisch durchlaufenden Phasen *Problem, Plan, Data, Analysis* und *Conclusions* besteht.

Ein ähnliches Modell bezogen auf das Vorgehen bei einem Data-Mining-Projekt beschreibt der CRISP-DM (*Cross Industry Standard Process for Data Mining*; vgl. Shearer, 2000). Mit diesem werden die einzelnen Schritte bei einem Data-Mining-Prozess dargestellt. Die Phasen sind unterteilt in *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation* und *Deployment*, wobei zwischen einigen Phasen zyklische Bezüge existieren, sodass manche Phasen mehrmals durchlaufen werden können (vgl. Shearer, 2000).

Eine Adaption des CRISP-DM für Datenprojekte ist der *Data Analysis Cycle*, orientiert an Berthold u.a. (2010, S.9).

Für eine ausführliche Darstellung der drei Modelle verweisen wir deshalb auf Wild/Pfannkuch (1999), Shearer (2000) bzw. Berthold u.a. (2010). Relevante Aspekte werden in den unteren Abschnitten jedoch aufgegriffen.

Ziel und Notwendigkeit einer eigenen Methodik

Aufbauend auf den aufgeführten Modellen wird in diesem Abschnitt eine neue Methodik vorgeschlagen. Diese kann zur Erkenntnisgewinnung genutzt werden, die unseres Erachtens nach für die Unterrichtspraxis im Kontext von Datenprojekten vordergründig ist.

Dies wird durch die obigen Modelle nicht ausreichend aufgegriffen. Die Methodik wurde durch Extraktion und Weiterentwicklung der aufgeführten Modelle entwickelt und soll einer besseren Umsetzbarkeit für Datenprojekte im Informatikunterricht dienen und einen Rahmen für eine Reihen- und Unterrichtsplanung liefern.

Außerdem sollen die Prozesse, die in einem Datenprojekt existieren und für jeden Lernenden unterschiedlich sein können, nachvollziehbar und erklärbar werden.

Mithilfe der vorherigen Überlegungen schlagen wir somit die in Bild 1 (nächste Seite) visualisierte Methodik für die Erkenntnisgewinnung durch Daten vor.

Im Folgenden werden zunächst die vier Handlungsbereiche erläutert, die in Bild 1 als außenstehende Felder dargestellt sind, und mit möglichen Handlungen ausgefüllt, bevor die zentrale Position der Erkenntnisgewinnung begründet wird. Danach wird aufgezeigt, wie nach dieser Methodik schulische Datenprojekte geplant und durchgeführt werden können.

Fragen stellen und Ideen entwickeln

In diesem Handlungsbereich werden das Problem und die Situation, die dem Datenprojekt zugrunde liegen, erschlossen oder auch definiert (vgl. „Problem“ bei Wild/Pfannkuch, 1999, und Shearer, 2000; „project understanding“ bei Berthold u. a., 2010). Dazu werden Fragen gestellt bzw. formuliert, die mithilfe der Erkenntnisgewinnung durch Daten beantwortet werden sollen. Dabei findet eine Verknüpfung zum Vorwissen und Hintergrundwissen statt (vgl. Berthold u. a., 2010). Im Rahmen einer Datenanalyse bzw. eines Data-Mining-Prozesses soll ein Plan für das Datenprojekt bzw. sollen Ideen dafür und für den Lösungsprozess gesammelt werden (vgl. Wild/Pfannkuch, 1999; Shearer, 2000). Hinsichtlich des Plans können bereits – wie es u. a. bei Wild/Pfannkuch (1999) explizit vorgeschlagen wird – erste Pilotierungen mit gegebenen Daten oder auch Beispieldaten vorgenommen werden.

Die Formulierung „Ideen entwickeln“ soll unterstreichen, dass nicht in jedem Datenprojekt von vornherein ein fester Plan bestehen muss. Es können stattdessen verschiedene Ansätze zunächst ausprobiert oder ein exploratives und datengetriebenes Vorgehen ermöglicht werden.

Daten sammeln und aufbereiten

Für schulische Datenprojekte können die benötigten Daten entweder seitens der Lehrperson zur Verfügung gestellt werden, oder die Lernenden sammeln die Daten selbstständig.

Anschließend wird untersucht, welche Daten vorhanden und relevant sind (vgl. Berthold u. a., 2010). Gegebenenfalls kann durch eine Erkenntnis in diesem Schritt eine Frage formuliert werden (vgl. Berthold u. a., 2010). Außerdem sollte identifiziert werden, welche Daten beispielsweise zur Beantwortung einer Frage noch fehlen. In diesem Handlungsbereich sollte ebenfalls das erstmalige Anschauen der Daten und ggf. die Verifizierung der Datenqualität stattfinden, wie es im Schritt *Data Understanding* bei Shearer (2000) und Berthold u. a. (2010) beschrieben wird. Die Formulierung *Daten sammeln* beschreibt also nicht bloß das Heranziehen einer Datenquelle oder das eigene Erheben von Daten.

Neben dem Sammeln von Daten ist i. d. R. eine Aufbereitung der Daten nötig. Dies kann vielfältige Tätigkeiten umfassen, wie beispielsweise das Säubern, Zusammenfassen, Formatieren oder auch Filtern von Daten (vgl. Shearer, 2000; Wild/Pfannkuch, 1999).

In diesem Kontext definieren Erickson u. a. (2018) *Data Moves*, die spezifische Handlungen für einen *Data*

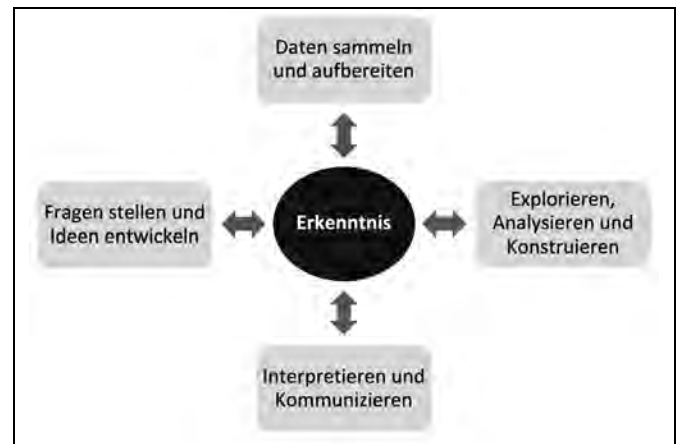


Bild 1: Methodik zur Erkenntnisgewinnung in Datenprojekten.

Scientist umfassen. Sie beschreiben folgende Handlungen als „Core Data Moves“: *Filtering, Grouping, Summarizing, Calculating, Merging* und *Reorganizing* (Erickson u. a., 2018). Einige, aber nicht alle dieser Data Moves, wie beispielsweise das Filtern, spielen im Rahmen der Aufbereitung der Daten eine Rolle. Die tiefere Verwendung der Data Moves, wie beispielsweise beim *Calculating*, wird erst im Handlungsbereich *Explorieren, Analysieren und Konstruieren* herangezogen.

Explorieren, Analysieren und Konstruieren

Wir postulieren für diesen Handlungsbereich und für die generelle Erkenntnisgewinnung in einem Datenprojekt die Wichtigkeit der Exploration von Daten, wie sie auch im PPDAC-Zyklus bei Wild/Pfannkuch (1999) im Schritt der Analyse vorgeschlagen wird. Durch das *Explorieren* bekommen die Lernenden die Möglichkeit, eigene Entdeckungen und somit schließlich Erkenntnisse hinsichtlich der Daten, darin enthaltener Informationen und ggf. Ideen für die Antwort der leitenden Frage zu erlangen. Sofern noch keine Frage formuliert wurde, könnte eine freie Exploration der Daten beim Stellen einer geeigneten Frage unterstützen. Somit können durch die Exploration tiefe Einblicke in die Struktur und den Aufbau der Daten gewonnen werden.

Als weiterer Schritt sind die Analyse und die Konstruktion der gesammelten oder gegebenen Daten zu nennen, mit denen schließlich eine Hypothese zur Beantwortung einer Frage oder zum Lösen eines Problems aufgestellt werden kann (vgl. Wild/Pfannkuch, 1999).

Im Rahmen dieses Handlungsbereichs sind ebenfalls die oben beschriebenen Data Moves von Erickson u. a. (2018) relevant und können für das *Explorieren, Analysieren und Konstruieren* verwendet werden.

Hinsichtlich der Modellierung, auf die beim CRISP-DM (vgl. Shearer, 2000) und im Data Analysis Cycle (vgl. Berthold u. a., 2010) großer Wert gelegt wird, sollte offengelegt werden, dass es sich höchstens um Mo-

delle der Realität handelt, die die Welt nicht eins zu eins widerspiegeln können.

Interpretieren und Kommunizieren

Nachdem Daten exploriert, analysiert oder konstruiert wurden, können diese interpretiert werden (vgl. Wild/Pfannkuch, 1999). Die Interpretationen können sich vor allem auf die zu untersuchende Frage oder das zu lösende Problem beziehen, sofern diese zuvor bestehen. Andernfalls können sie auch die Schritte im Bereich *Fragen stellen und Ideen entwickeln* unterstützen. Dabei können jeweils weitere Erkenntnisse gewonnen werden. Falls vor der Formulierung einer geeigneten Frage bereits eine Exploration der Daten stattfindet, können mithilfe einer Interpretation ebenfalls Ideen entwickelt werden (vgl. Wild/Pfannkuch, 1999). Das Entwickeln von Ideen kann sich sowohl auf eine erste Formulierung als auch auf eine neue Frage beziehen, die während der Untersuchung einer bestehenden Frage auftritt. So können Erkenntnisse beispielsweise während der Exploration, Analyse und Konstruktion dazu führen, dass neue Probleme oder neue Fragen entdeckt werden, die möglicherweise weitergehend untersucht werden können. Das adäquate Kommunizieren beispielsweise von Prozessschritten, vom aktuellen Vorgehen, von einer Protokollierung von Ergebnissen oder von jeglichen Erklärungen vor allem seitens der Lernenden ist für den Unterrichtsprozess essenziell, um die prozessbezogenen Kompetenzen der Lernenden zu fördern. Wie im CRISP-DM oder im Data Analysis Cycle könnten beispielsweise die Ergebnisse oder mitzunehmenden Erkenntnis(se) aus dem gesamten Projekt evaluiert und reflektiert werden (vgl. Shearer, 2000; Berthold u. a., 2010).

Der variable Start eines Datenprojekts

Wie bei den oben vorgestellten Modellen für die Schritte der Datenanalyse oder des Data Minings ist ein Start des Datenprojekts im Handlungsbereich *Fragen stellen und Ideen entwickeln* möglich (vgl. Wild/Pfannkuch, 1999; Shearer, 2000; Berthold u. a., 2010). So könnte zunächst eine Frage oder auch ein Problem, das i. d. R. der realen Welt entstammt, gefunden, beschrieben, formuliert oder gegeben werden. Darauf aufbauend werden in diesem Handlungsbereich Ideen entwickelt, wie die Frage untersucht oder das Problem gelöst werden kann. Entgegen der Ansätze einiger oben aufgeführter Modelle ist ebenfalls explizit ein Start in den anderen drei Handlungsbereichen sinnvoll. Ein Start beim *Explorieren, Analysieren und Konstruieren* wäre so denkbar, dass die Lernenden gegebene Daten explorieren und nach Informationen suchen, die sie aus den Daten entnehmen können. Für einen solchen Start ist Vorwissen über den Hintergrund des Themas bzw. des Kontextes nicht zwangsläufig nötig; durch das explorative Vorgehen können auch thematische Hintergründe oder Zusammenhänge erschlossen werden. Des Weiteren ist ein Start im Bereich *Daten sammeln und aufbereiten* möglich, beispielsweise indem Daten nach

einem vorgegebenen Verfahren erhoben oder gegebene Daten zunächst aufbereitet werden. Im Handlungsbereich *Interpretieren und Kommunizieren* könnte ein Start beispielsweise mit einer eigenen bzw. gegebenen Kommunikation oder möglicherweise auch durch eine Interpretation von bereits vorliegenden Ergebnissen vorgenommen werden.

Erkenntnis als zentrales Bindeglied

Zwischen jedem der vier zuvor beschriebenen Handlungsbereiche und dem Feld der Erkenntnis besteht in der Methodik eine Wechselbeziehung. Es ist möglich, mithilfe einer Erkenntnis, die aus einem Schritt eines Handlungsbereichs gewonnen werden konnte, einen weiteren und i. d. R. anderen Schritt eines anderen oder auch desselben Handlungsbereichs einzuleiten. Die Erkenntnisgewinnung in dem Prozess des Datenprojekts steht somit im Mittelpunkt der Methodik, da ohne diese keine Problemlösung und keine Beantwortung einer Forschungsfrage möglich ist. Noch feingranularer kann dies betrachtet werden, wenn beispielsweise ein Schritt in dem Bereich *Explorieren, Analysieren und Konstruieren* angeschaut wird: Die Daten werden mit einer bestimmten Handlung untersucht – z. B. ein Data Move –, und bevor daraus folgend eine Interpretation vorgenommen werden kann, geht eine Erkenntnis voraus. Als weiteres Beispiel können Schritte aus dem Handlungsbereich *Fragen stellen und Ideen entwickeln* betrachtet werden, die durch Erkenntnisse aus dem Bereich *Explorieren, Analysieren und Konstruieren* motiviert werden.

Während eines Datenprojekts sind durchaus flexible Wechsel zwischen den vier Handlungsbereichen erwünscht und sollten ermöglicht werden. So können beispielsweise die Ideen für den Prozess oder die zu untersuchende Frage während der Datenanalyse überarbeitet werden. Des Weiteren steht der Handlungsbereich *Interpretieren und Kommunizieren* i. d. R. in starken Wechselbeziehungen zu den anderen Handlungsbereichen, da zum Beispiel Schritte beim *Explorieren, Analysieren und Konstruieren* zu Ergebnissen führen, die beispielsweise mithilfe der *Interpretation und Kommunikation* immer wieder zu (neuen) Erkenntnissen führen.

Datenbezogene Unterrichtsprojekte

In dem vorherigen Abschnitt haben wir eine Methodik für die Erkenntnisgewinnung durch Daten vorgeschlagen und erläutert. Nun werden drei Unterrichtsprojekte vorgestellt, an denen beispielhaft das Vorgehen in der Methodik und ein möglicher Einsatz dieser beschrieben wird. Die Unterrichtsprojekte sind im Rahmen des Projekts ProDaBi entwickelt und mehrmals durchgeführt worden.

Tätersuche mit Jupyter Notebooks

Vorstellung

Das bereits mehrfach durchgeführte und evaluierte Unterrichtsprojekt *Tätersuche mit Jupyter Notebooks* befasst sich mit Aspekten von Datensammlung, Datenverarbeitung und Datenschutz im Kontext eines fiktiven Mordfalls, der mithilfe von gespeicherten Standortdaten aufgelöst werden soll. Die zur Verfügung gestellten Standortdaten werden von den Lernenden mithilfe von Jupyter Notebooks exploriert und analysiert. Aufgrund der vielen Differenzierungsmöglichkeiten werden für dieses Unterrichtsprojekt keine fachlichen Vorkenntnisse und keine Programmiererfahrungen vorausgesetzt. Somit besteht eine besondere Eignung für die Sekundarstufe I, wobei eine abgewandelte Umsetzung in der Sekundarstufe II durchaus denkbar ist. Die Jupyter Notebooks erlauben eine derartige Differenzierung (vgl. Jupyter Notebook, 2014 ff., und Wikipedia – Stichwort „Jupyter Notebook“).

Zielsetzung

In diesem Unterrichtsprojekt soll ein Vorgehen zur Analyse und Auswertung von Daten am Beispiel von Standortdaten erlernt und eigenständig in Jupyter Notebooks erprobt werden.

Dabei erfahren die Lernenden, welche Informationen über eine Person aus scheinbar anonymen Standortdaten gewonnen werden können. Es werden Grundlagen der Datenspeicherung und des Datenschutzes am Beispiel von Standortdaten vermittelt, indem die Lernenden Beispiele des eigenen Alltags untersuchen, in denen persönliche Daten gesammelt werden. Dabei erwerben die Lernenden einen kritischen und reflektierten Blick auf diese Themen und lernen, eine eigene Meinung fundiert zu formulieren.

Ablauf des Unterrichtsprojekts

Das Unterrichtsprojekt ist in vier Lerneinheiten gegliedert. Als wiederkehrende Begleitung dessen stellt der in Bild 2 aufgeführte Verlaufsplan, die sogenannte Roadmap, dar, mit der die vier Lerneinheiten für die Lernenden transparent dargestellt werden.

- 1. Lerneinheit – Einführung in das Unterrichtsprojekt:**
Zum Einstieg in das Unterrichtsprojekt wird ein Video zu dem Mordfall gezeigt. In diesem werden der Mordfall und alle relevanten Personen mit kurzen Steckbriefen vorgestellt. Anschließend werden die nötigen Schritte zum Lösen des Mordfalls besprochen. Gegebenenfalls sollte hier eine kurze Einführung in Jupyter Notebooks stattfinden.
- 2. Lerneinheit – Erarbeitung und Lösen des Mordfalls mit Jupyter Notebooks:**
Mithilfe von Analysen der gegebenen „anonymen“ Datensätze mit Standortdaten von Verdächtigen des Mordfalls müssen zunächst verlorengegangene Zuordnungen der verdächtigen Personen zu den Datensätzen ermittelt werden. Anschließend lösen die Lernenden mithilfe von Hinweisen den Mordfall und

stimmen innerhalb der Lerngruppe für einen der beiden zum Schluss möglichen Täter ab. Die Analysen finden in zwei vorbereiteten Jupyter Notebooks statt. Die Lerngruppe verurteilt durch Mehrheitsentscheidung einen Täter, wobei durch die Analyse keine Sicherheit gegeben ist. Das Urteil soll zum Abschluss aus ethischer Sicht und hinsichtlich der Fairness kritisch diskutiert werden.

3. Lerneinheit – Thematisierung von Datenschutz und Datenspeicherung:

Zum Einstieg der Lerneinheit wird der Begriff *Datenschutz* mithilfe eines Erklärvideos eingeführt. Die Themen *Datenschutz* und *Datensammlung* werden anhand einiger Beispiele (Apps, Unternehmen, Institutionen, ...) in Gruppen-/Partnerarbeit erarbeitet, im Plenum vorgestellt und diskutiert. Dabei liegt der Fokus darauf, welche persönlichen Daten gesammelt und wofür diese genutzt werden können. Nach einer kritischen Diskussion von allgemeinen Aspekten, Gemeinsamkeiten und Auffälligkeiten werden zum Abschluss Gefahren und Chancen der Sammlung von Daten (auch Standortdaten) beleuchtet.

4. Lerneinheit – Abschlussreflexion:

Diese Lerneinheit wird insbesondere durch Leitfragen strukturiert, wobei die Ergebnisse, Erkenntnisse und Meinungen zu Datenschutz und Datensammlung diskutiert werden. Dabei sollen die Erfahrungen des gesamten Unterrichtsprojekts und vor allem die Meinungen der Lernenden einbezogen werden.



Bild 2: Roadmap zum Unterrichtsprojekt.

Einordnung in die Methodik

Durch die Einführung in den Mordfall in der ersten Lerneinheit wird der Kontext des Mordfalls beschrieben. Dies ist in den Handlungsbereich *Fragen stellen und Ideen entwickeln* einzuordnen. Das Besprechen der nötigen Schritte zum Lösen des Mordfalls gleicht dem Entwickeln von Ideen zur Beantwortung der Täterfrage. Die resultierende Erkenntnis ist der Plan beziehungsweise die Idee für das weitere Vorgehen. Darauf folgend findet mit dem Übergang zu der zweiten Lerneinheit ein Wechsel zum Handlungsbereich *Daten sammeln und aufbereiten* statt. Die Daten sind bereits gegeben und müssen nicht mehr gesammelt werden. Dadurch soll der Schwierigkeitsgrad an dieser Stelle bewusst gesenkt werden, um den Einstieg in das Datenprojekt und das Verarbeiten von Daten mithilfe der Jupyter Notebooks zu vereinfachen. Die Aufbereitung der Daten findet in Form des Importierens der Datensätze in die Jupyter Notebooks statt. In diesem Handlungsbereich ist ebenfalls das Verstehen der Struktur der Daten sowie der Daten selbst inbegriffen. Mithilfe der Erkenntnisse aus diesem Handlungsbereich kann anschließend das *Explorieren, Analysieren und Konstruieren* folgen, um die Datensätze den Verdächtigen zuzuordnen und einen Täter zu ermitteln. Dabei finden wiederholt Wechsel zu Tätigkeiten des Handlungsbereichs *Interpretieren und Kommunizieren* statt, wobei u. a. der Kontext des Mordfalls in Form der Steckbriefe herangezogen wird. Infolgedessen werden Erkenntnisse gesammelt, die für die Zuordnung und die Wahl des Täters relevant sind. Bei diesen Erarbeitungen sind durchaus individuelle Wechsel zwischen den vier Handlungsbereichen möglich, werden jedoch an dieser Stelle aufgrund des begrenzten Rahmens dieses Beitrags nicht weiter ausgeführt.

Die dritte und vierte Lerneinheiten sind als Ergänzung des Datenprojekts zu verstehen, die dazu dienen, die Erkenntnisse der ersten beiden Lerneinheiten in einen gesellschaftlichen und persönlichen Rahmen einzuordnen.

Umweltanalyse mit der Sensebox und Jupyter Notebooks

Vorstellung

Im Folgenden wird ein Unterrichtsprojekt beschrieben, bei dem Lernende mithilfe einer Sensebox – ein Arduino, an den verschiedene Sensoren angeschlossen werden können – Umweltdaten wie Temperatur, Luftfeuchtigkeit oder Feinstaub sammeln und diese im Anschluss mithilfe eines Jupyter Notebooks auswerten. Die Auswertung soll dabei auf eine von den Lernenden formulierte Forschungsfrage abzielen, die im Rahmen einer Projektpräsentation zum Ende des Unterrichtsprojekts beantwortet werden soll.

Zielsetzung

In diesem Unterrichtsprojekt sollen die Lernenden Methoden zur Datenanalyse kennenlernen und diese auf verschiedene Datensätze anwenden. Dabei spielen

besonders Auswertungs- und Darstellungsmöglichkeiten großer Datensätze eine entscheidende Rolle, jedoch auch das Formulieren und Beantworten von Forschungsfragen, die sich teilweise – im Sinne eines Projektstartpunkts im Handlungsbereich *Explorieren, Analysieren und Konstruieren* – erst durch Erkenntnisse aus der Exploration der Datensätze ergeben. Auf der anderen Seite stellt auch die Reflexion der durchgeführten Datensammlung und -analyse ein wesentliches Ziel des Unterrichtsprojekts dar: Die Lernenden entdecken, wie gesammelte Daten und Datenanalysen den (datengetriebenen) Blick auf die Welt und das eigene Verhalten beeinflussen können.

Ablauf des Unterrichtsprojekts

Das vorliegende Unterrichtsprojekt ist in fünf Lerneinheiten gegliedert, die die Lernenden in Kleingruppen durchlaufen. Zuvor wird ihnen ein kurzer Überblick über das Unterrichtsprojekt sowie über das Messinstrument der Sensebox (siehe Bild 3, nächste Seite) und das Auswertungsinstrument der Jupyter Notebooks gegeben.

1. Lerneinheit – Einführung in das Projekt und Generieren einer Fragestellung für die geplante datengetriebene Analyse:

Zunächst entwickeln die Lernenden erste Fragestellungen, denen sie in der späteren Datenanalyse nachgehen wollen. Ein Beispiel für eine solche Fragestellung könnte sein: „Wann und wie oft überschreitet die Feinstaubbelastung an unserer Schule EU-Grenzwerte?“ Im weiteren Verlauf des Unterrichtsprojekts können diese Fragen jedoch noch aufbauend auf zwischenzeitlich erhaltenen Erkenntnissen angepasst oder ergänzt werden.

2. Lerneinheit – Planen und Erstellen des Instruments zur Datensammlung:

In der zweiten Phase überlegen die Lernenden, welche Daten sie zur Beantwortung ihrer Forschungsfragen benötigen und wie bzw. wo sie diese mit der Sensebox erheben können. Im Anschluss daran programmieren sie die Sensebox so, dass die gewünschten Daten gesammelt werden und stellen sie an der geplanten Stelle auf. Gegebenenfalls muss der Aufstellort im Rahmen der Datensammlung nach einer bestimmten Zeit gewechselt werden. Alternativ lässt sich das Unterrichtsprojekt auch mit dieser Lerneinheit starten, indem zunächst alle möglichen Sensoren bei der Messung verwendet werden und dann nach einer ersten Sichtung der erhobenen Daten eine Forschungsfrage formuliert wird, die durch weitere Datenexploration beantwortet werden soll.

3. Lerneinheit – Programmiersprache PYTHON kennenlernen:

Während die Daten mit der Sensebox über mehrere Wochen gesammelt werden, machen sich die Lernenden mit der Programmiersprache PYTHON vertraut. Hierzu durchlaufen sie einen PYTHON-Online-Kurs der University of Waterloo sowie zwei Jupyter Notebooks, in denen die Arbeit mit Jupyter Notebooks und spezifische Datentypen für die Datenauswertung erläutert werden.

4. Lerneinheit – Datenbereinigung und -analyse:

In der vierten Phase pflegen die Lernenden die gesammelten Daten der Sensebox in die Jupyter Notebooks ein, bereinigen diese und werten sie unter Verwendung verschiedener Visualisierungen und statistischer Kenngrößen hinsichtlich ihrer eingangs formulierten Forschungsfrage aus. Dabei benutzen und adaptieren sie bereits vorgefertigte Jupyter Notebooks, die sie bei ihrer Datenauswertung unterstützen. Weiterhin erstellen sie in dieser Phase eine Präsentation, die die wesentlichen Ergebnisse ihrer Datenanalyse sowie daraus resultierende gesellschaftliche Implikationen beinhaltet.

5. Lerneinheit – Auswertungsphase:

Im letzten Schritt stellen die Kleingruppen ihre Präsentationen vor und diskutieren die herausgearbeiteten Ergebnisse und gesellschaftlichen Implikationen mit den anderen Lernenden. Dabei reflektieren sie auch, inwiefern sie die Datensammlung und die Datenauswertung dabei unterstützt hat, Erkenntnisse und darauf aufbauende gesellschaftliche Implikationen in Bezug auf ihre Umwelt zu gewinnen.

Einordnung in die Methodik

Zu Beginn dieses Unterrichtsprojekts können die Lernenden Forschungsfragen erstellen, auf denen ihre anschließende Datenanalyse aufbaut. Dabei entwickeln sie ebenso erste Ideen hinsichtlich erwartbarer Ergebnisse oder auch bezüglich eines Plans zur Erhebung und Auswertung geeigneter Daten. Folglich lässt sich diese Projektphase dem Handlungsbereich *Fragen stellen und Ideen entwickeln* zuordnen. Alternativ können die Forschungsfragen auch aufbauend auf einer ersten Exploration von zuvor erhobenen Umweltdaten erstellt werden.

Anknüpfend an die Fragestellungen und ersten Ideen zur Datenerhebung und -auswertung entwickeln die Lernenden durch eine selbstständige Programmierung der Sensebox ein eigenes Erhebungswerkzeug. Dabei stellen sie sich Fragen, die sich auf die konkrete Datenerhebung beziehen. Während die Beantwortung dieser Fragen noch zum Handlungsbereich *Fragen stellen und Ideen entwickeln* gehört, beziehen sich die Programmierung und das Aufstellen der Sensebox bereits auf den Handlungsbereich *Daten sammeln und aufbereiten*, da hiermit der Grundstein für die Datensammlung gelegt wird.

Die Bearbeitung des PYTHON-Online-Kurses und der zwei einführenden Jupyter Notebooks dient in diesem Unterrichtsprojekt der Aneignung grundlegenden Wissens für die Datenanalyse. Aufgrund der gleichzeitig stattfindenden Datensammlung durch die Sensebox ist diese Phase als Schritt im Handlungsbereich *Daten sammeln und aufbereiten* zu sehen.

In der vierten Phase des Unterrichtsprojekts bereiten die Lernenden die gesammelten Daten zunächst auf, indem sie Messfehler löschen und die zur Beantwortung ihrer Forschungsfrage benötigten Daten herausfiltern. Während diese Aktivitäten zum Handlungsbereich *Daten sammeln und aufbereiten* gehören, lässt sich die anschließende Exploration und Analyse der Daten, was insbesondere das Erstellen von Grafiken und die Berechnung statistischer Kennwerte beinhaltet,

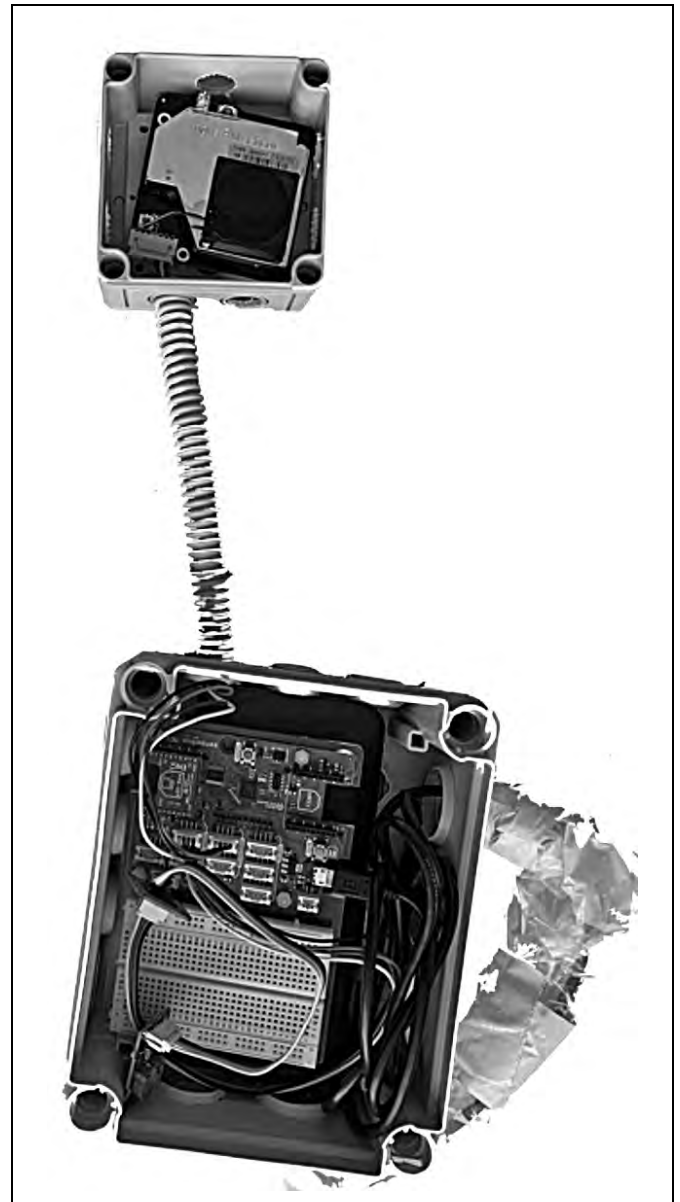


Bild 3: Sensebox mit Feinstaubsensor und Temperatur-/Luftfeuchtigkeitssensor.

klar dem Handlungsbereich *Explorieren, Analysieren und Konstruieren* zuordnen. Dabei sei jedoch auch erwähnt, dass durch die Exploration der Daten in dieser Unterrichtsphase ebenfalls neue Forschungsfragen oder Ideen entwickelt werden können.

Die Erstellung der Präsentationen in dieser Phase bezieht sich auf den Handlungsbereich *Interpretieren und Kommunizieren*, da die Lernenden ihre Ergebnisse aus der Datenanalyse in den Präsentationen darstellen und interpretieren sollen, gerade auch im Hinblick auf gesellschaftliche Implikationen. Daran knüpft auch die letzte Phase dieses Unterrichtsprojekts an, in der diese Ergebnisse und Implikationen kommuniziert und im Anschluss von allen Lernenden diskutiert werden sollen.

Datenanalyse mit CODAP am Beispiel eigener JIM-Daten

Vorstellung

In diesem Unterrichtsprojekt explorieren die Lernenden mithilfe des Datenanalysetools CODAP einen multivariaten, mittels einer Onlinebefragung erhobenen Datensatz zum Medienumgang von Jugendlichen. Die Onlinebefragung orientiert sich dabei am Fragebogen der JIM-Studie 2016 (vgl. mpfs, 2016). Die Daten werden den Lernenden zur Verfügung gestellt, die in diesem Kontext eigene statistische Fragestellungen entwickeln, analysieren und beantworten. Die Analyse der Daten findet hauptsächlich mit CODAP statt. Für tiefere Untersuchungen und eine mögliche Differenzierung des Unterrichtsprojekts steht auch ein entsprechendes Jupyter Notebook zur Datenanalyse der JIM-Daten zur Verfügung.

Zielsetzung

In diesem Unterrichtsprojekt steht insbesondere die Datenanalyse bis hin zum Erkenntnisinteresse bezüglich bestimmter Aspekte der Ergebnisse der JIM-Umfrage im Fokus. Die Lernenden entdecken, wie man multivariate Daten analysiert, auswertet und so aus Daten im Kontext Erkenntnisse gewinnen kann.

Konkret steht hierbei die Aneignung erster Data Moves wie beispielsweise *Datenbereinigung*, *Datenzusammenführung* und *Datenvisualisierung* im Vordergrund, indem die Lernenden diese auf einen vorgegebenen Datensatz anwenden.

Ablauf des Unterrichtsprojekts

Das vorliegende Unterrichtsprojekt ist in folgende vier Lerneinheiten gegliedert.

1. Lerneinheit – Einführung in den JIM-PB-Datensatz und in CODAP:

Nach einer kurzen Vorstellung des Erhebungsverfahrens des JIM-Datensatzes werden die wesentlichen Elemente des Datenanalysetools CODAP erklärt. Die Lernenden steigen dann selbst in die Exploration der Daten mit CODAP ein und lernen dabei den Datensatz mit seinen verschiedenen Merkmalen und Ausprägungen kennen. Diese Phase endet mit dem Zusammentragen der ersten Ergebnisse der Datenexploration und dem intuitiven Generieren statistischer Fragestellungen.

2. Lerneinheit – Einige statistische Konzepte:

Basierend auf den bereits entwickelten Fragestellungen erwerben die Lernenden Daten-Kompetenzen zu statistischen Konzepten im Kontext der JIM-Daten. Hierbei stehen die Exploration von Zusammenhängen numerischer und kategorialer Variablen, die Entwicklung und Bewertung von statistischen Diagrammen sowie die adäquate Präsentation von statistischen Ergebnissen durch Datenplakate im Vordergrund. Die Ergebnisse werden schließlich mithilfe von Normen zusammengefasst, sodass diese für die nächste Phase zur Verfügung stehen.

3. Lerneinheit – Projektphase:

In dieser Phase analysieren die Lernenden in Kleingruppen die JIM-PB-Daten hinsichtlich ihrer eigenen statistischen Fragestellungen. Während des Prozesses der Datenanalyse steht die Lehrperson als Lernbegleiter und die zuvor entwickelten Normen für statistische Diagramme und Präsentationen als Orientierung zur Verfügung, während die Gruppen eigenständig ein Datenplakat zur Dokumentation ihrer Datenanalyse und Beantwortung ihrer Fragestellung erstellen.

4. Lerneinheit – Vorbereitung und Präsentation der Datenplakate:

Diese Phase dient dem Abschluss des Unterrichtsprojekts. Die Präsentationen werden durch die Gruppen finalisiert und anschließend im Plenum kommuniziert. Die anschließende Diskussion und Bewertung findet erneut mithilfe der Normen aus Lerneinheit 2 statt. Die Lernenden können sich so gegenseitig differenzierte Rückmeldungen geben und die eigene Leistung begründet reflektieren.

Einordnung in die Methodik

Die erste Phase dieses Unterrichtsprojekts setzt sich aus mehreren Schritten der oben beschriebenen Methodik zusammen. Während die Erläuterung des Datenerhebungsverfahrens Erkenntnisse im Sinne des Handlungsbereiches *Daten sammeln und aufbereiten* liefert, bezieht sich die Einführung der Lehrperson zu CODAP auf Schritte im Handlungsbereich *Explorieren, Analysieren und Konstruieren*. Anschließend werden diese Schritte bei der ersten Exploration der Daten von den Lernenden ausgeführt. Dabei erhalten sie Erkenntnisse, auf deren Grundlage sich erste intuitive Fragen stellen lassen, denen die Lernenden in der nachfolgenden Exploration weiter nachgehen können (*Fragen stellen und Ideen entwickeln*). Zudem können sie in dieser Phase bereits ihre ersten gesammelten Ergebnisse interpretieren und hinterfragen, wie diese zustande kommen (*Interpretieren und Kommunizieren*).

In der zweiten Lerneinheit steht die Vorbereitung der weiteren Schritte im Sinne der Handlungsbereiche *Fragen stellen und Ideen entwickeln, Explorieren, Analysieren und Konstruieren* sowie *Interpretieren und Kommunizieren* im Fokus, indem die Lernenden entsprechende Informationen über statistische Konzepte im Kontext der JIM-Daten erhalten und diese als Normen festhalten.

Darauf aufbauend durchlaufen die Lernenden in der Projektphase Schritte aus allen vier oben genannten Handlungsbereichen, indem sie ihre erste intuitive Fragestellung überarbeiten (*Fragen stellen und Ideen entwickeln*), geeignete Daten und Datensätze zur Beantwortung dieser Fragestellung auswählen (*Daten sammeln und aufbereiten*), diese Daten dann explorieren und auswerten (*Explorieren, Analysieren und Konstruieren*) und die aus dieser Datenanalyse erhaltenen Ergebnisse (wie beispielsweise Grafiken und statistische Kennwerte) gemeinsam mit Erkenntnissen hieraus in einem Datenplakat zusammenführen (*Interpretieren und Kommunizieren*). Wichtig ist hierbei anzumerken, dass es sich nicht um einen linearen Ablauf dieser vier

Bereiche handelt, sondern die Lernenden vielmehr dauerhaft erkenntnisgeleitet zwischen diesen wechseln. In der letzten Phase werden die entstandenen Datenplakate präsentiert und diskutiert, sodass die Lernenden sich in dieser Phase in dem Handlungsbereich *Interpretieren und Kommunizieren* bewegen, auch weil Rückschlüsse aus den vorgestellten Ergebnissen gezogen und besprochen werden.

Fazit und Ausblick auf weitere Arbeit in ProDaBi

In diesem Beitrag wurde eine Methodik für Datenprojekte im Schulunterricht, vor allem für den Informatikunterricht, vorgeschlagen und an drei beispielhaften Unterrichtsprojekten aufgezeigt, wie eine Umsetzung dieser Methodik in der Praxis aussehen kann. Um die Schwächen und Stärken dieser Methodik zu untersuchen, werden die Autoren diese weitergehend einsetzen und evaluieren sowie in weitere Diskurse einsteigen.

Lukas Höper
Sven Hüsing
Hülya Malatyali
Prof. Dr. Carsten Schulte
Lea Budde
Universität Paderborn
Fürstenallee 11
33102 Paderborn

E-Mails: lukas.hoeper | sven.huesing | huelya.malatyali | carsten.schulte | [@upb.de](mailto:lea.budde)

Danksagung:
Dieser Beitrag wurde von Mitarbeitern des Projekts ProDaBi verfasst, wobei Ideen und Gedanken des gesamten ProDaBi-Teams eingeflossen sind. Bei weiterem Interesse an dem Projekt

oder an den Materialien für die aufgeführten Unterrichtsprojekte wenden Sie sich gerne an das ProDaBi-Team:
prodabi@campus.upb.de

Literatur und Internetquellen

Berthold, M.R.; Borgelt, Chr.; Höppner, F.; Klawonn, F.: Guide to Intelligent Data Analysis – How to Intelligently Make Sense of Real Data. Reihe „Texts in Computer Science“. London: Springer, 2010.

Erickson, T.; Finzer, B.; Reichsman, F.; Wilkerson, M.: Data Moves – One key to data science at the school level. In: M.A. Sorto, A. White, L. Guyot (Hrsg.): Looking back, looking forward – Proceedings of the Tenth International Conference on Teaching Statistics – ICOTS10, 8–13 July, 2018, Kyoto, Japan. Voorburg (NL): ISI/IASE, 2018 (Electronic Edition).
<https://t1p.de/g7zv>

Jupyter Notebook, 2014 ff.:
<https://jupyter.org/>

mpfs – Medienpädagogischer Forschungsverbund Südwest (Hrsg.): JIM-Studie 2016 – Jugend, Information, (Multi-) Media. Basisuntersuchung zum Medienumgang 12- bis 19-Jähriger. Stuttgart: mpfs, 2018.
<https://t1p.de/n43i>

ProDaBi – Projekt Data Science und Big Data. 2018 ff.
<https://www.prodabi.de/>

Shearer, C.: The CRISP-DM Model – The New Blueprint for Data Mining. In: Journal of Data Warehousing, Vol. 5 (2000), No. 4, S.13–22.
<https://t1p.de/rkz0>

Wild, C.J.; Pfannkuch, M.: Statistical Thinking in Empirical Enquiry. In: International Statistical Review, Vol. 67 (1999), No. 3, S.223–248.
<https://t1p.de/k10v>

Wikipedia – Stichwort „Jupyter Notebook“:
https://de.wikipedia.org/wiki/Project_Jupyter#Jupyter_Notebook

Alle Internetquellen wurden zuletzt am 18. Februar 2021 geprüft und können auch aus dem Service-Bereich des LOG IN Verlags (<https://www.log-in-verlag.de/>) heruntergeladen werden.