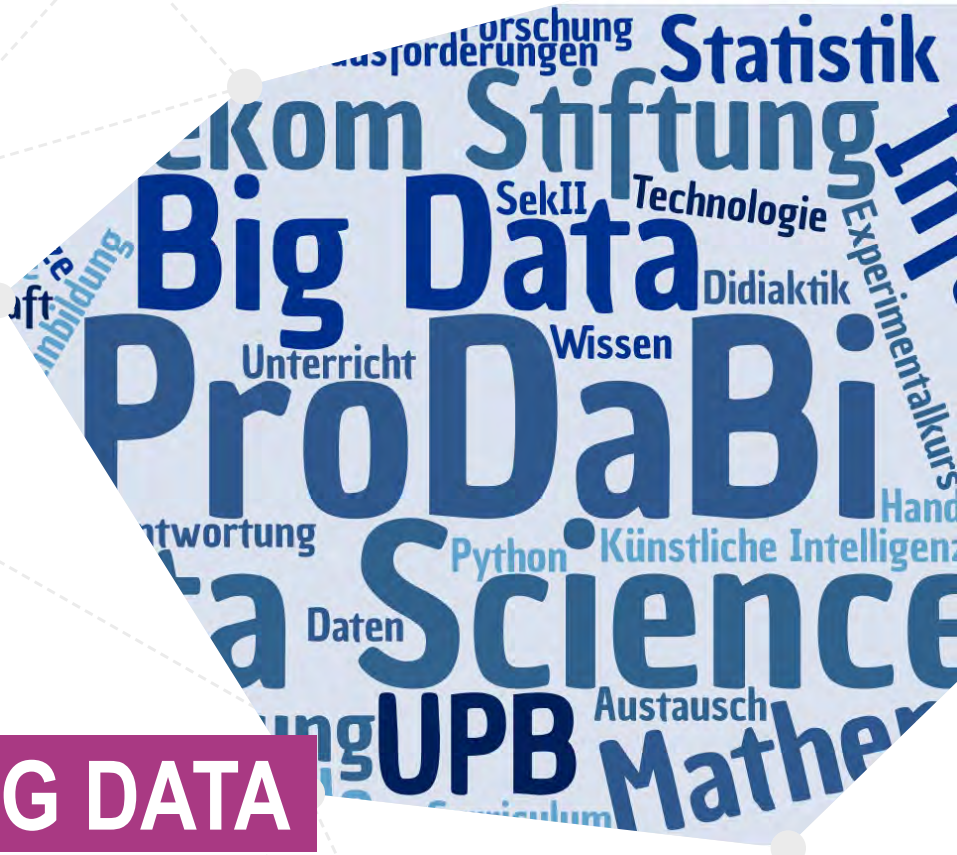
UNIVERSITÄT PADERBORN

ProDaBi
Projekt Data Science und Big Data in der Schule



DIDAKTIK DER INFORMATIK

DATA SCIENCE UND BIG DATA IN DER BERUFLICHEN BILDUNG – KONZEPTION UND ERPROBUNG EINES PROJEKTKURSES FÜR DIE SEKUNDARSTUFE II

27. BAG-FACHTAGUNG 11. – 13.03.2019 SIEGEN

Simone Opel & Michael Schlichtig

Deutsche Telekom Stiftung

Agenda

- *Data Science und Big Data und berufliche Informatikbildung*
- *Das Projekt ProDaBi*
- *Der Projektkurs zu ProDaBi*
- *Weiterentwicklung*
- *Fazit*

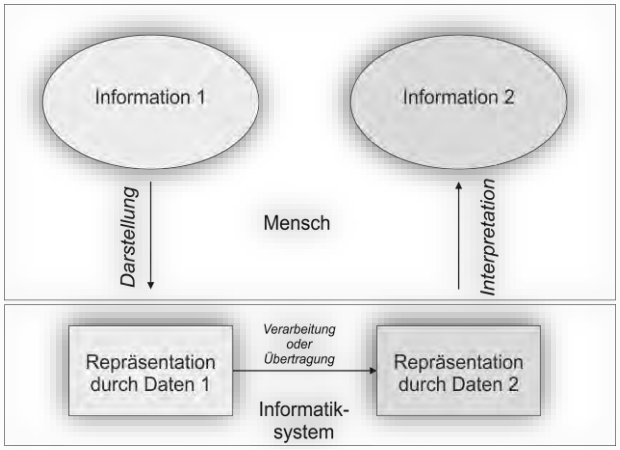
Data Science und Big Data in der beruflichen Informatikbildung

- Data Science, Big Data und Künstliche Intelligenz sind keine integralen Bestandteile verschiedener Lehrpläne
- **Aber:**
  - Bedeutung nimmt im Alltag immer mehr zu
  - Beruflicher Einsatz wird immer wichtiger
- **KMK – Handreichung 2011:**
  - „Die Förderung und der Erwerb einer umfassenden Handlungskompetenz stehen damit im Mittelpunkt des pädagogischen Wirkens.“ (S. 11)
  - „Damit befähigt die Berufsschule die Auszubildenden […] zur nachhaltigen Mitgestaltung der Arbeitswelt und der Gesellschaft in sozialer, ökonomischer, ökologischer und individueller Verantwortung.“ (S. 10)

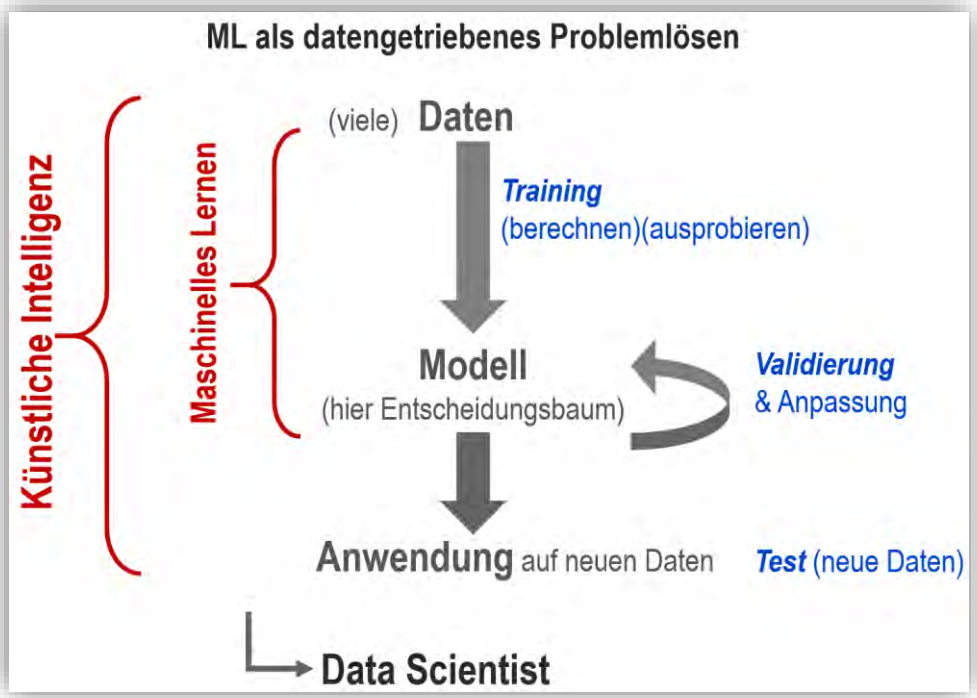
Bildquellen: vdi-nachrichten.de, amazon.de

Das Projekt ProDaBi - Data Science und Big Data in der Schule

Daten als Basis



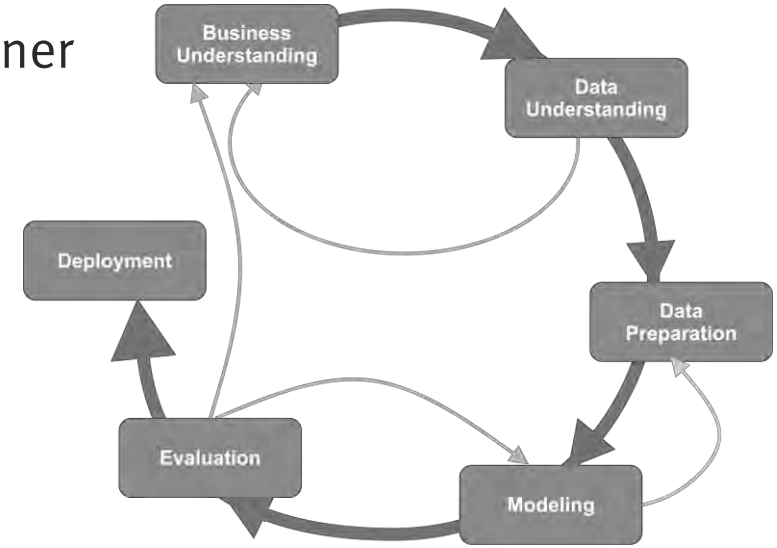
Information und Daten nach GI



Ergebnis einer Arbeitsphase während des Projektkurses

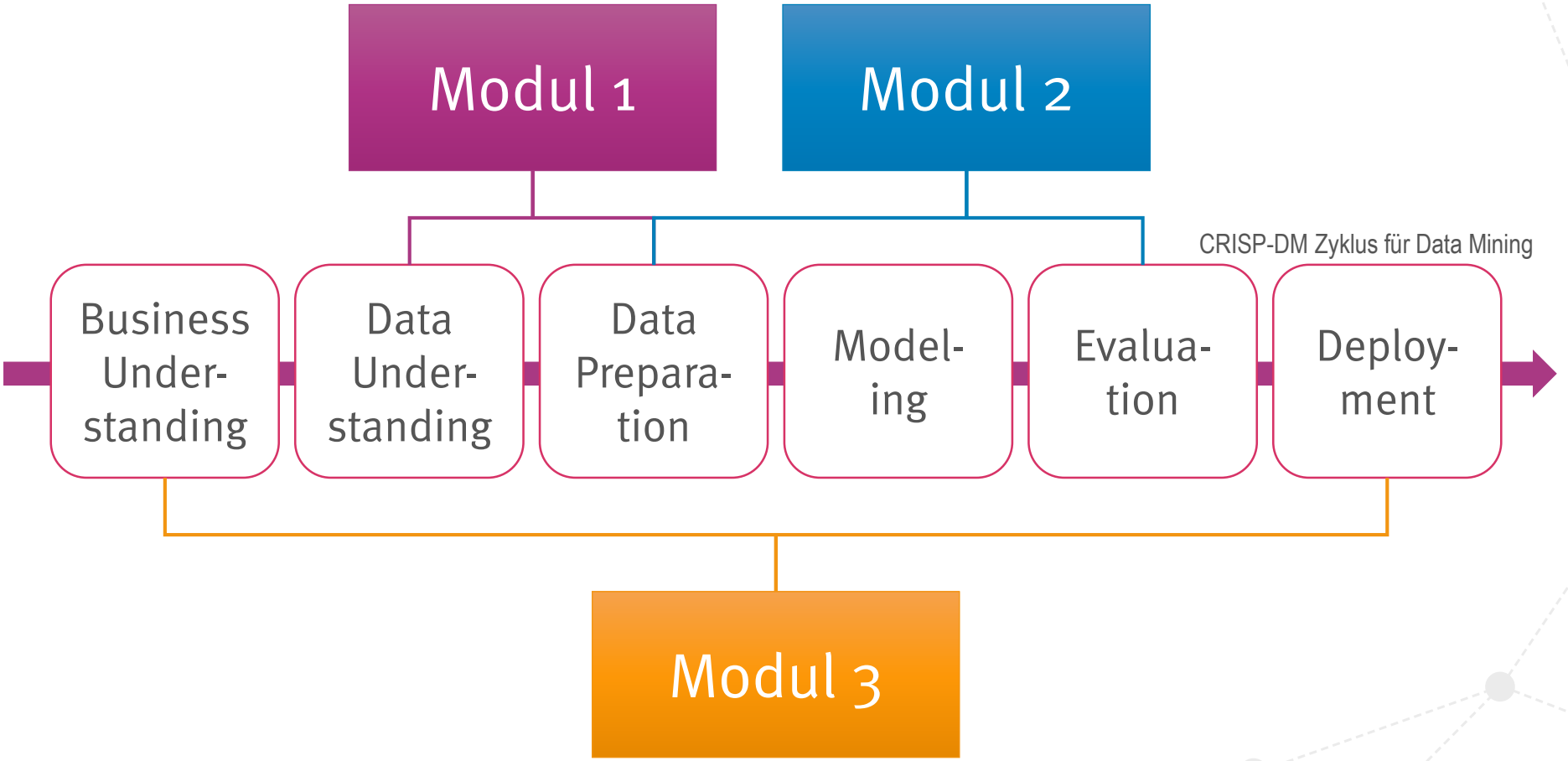
Der Projektkurs „Data Science“ – Planung

- Basis des Kurses ist der CRISP-DM Zyklus
- Entwicklung von einer algorithmen- zu einer datengetriebenen Sicht
- **Modularer Aufbau:**
  - **Modul 1:**
    Daten verstehen, aufbereiten und präsentieren – Datendetektive
  - **Modul 2:**
    Modelle erstellen und evaluieren – Maschinelles Lernen
  - **Modul 3:**
    Anwenden des Wissens auf eine konkrete Fragestellung - Projektmodul



CRISP-DM Zyklus für Data Mining nach Berthold et al. (2010)

Der Projektkurs „Data Science“ im Datenzyklus



Rahmen der Erprobung

- Kooperation mit 2 paderborner Gymnasien
- Projektkurs – Ersatz zu Facharbeit
- 2 Schülerinnen, 17 Schüler der Q2

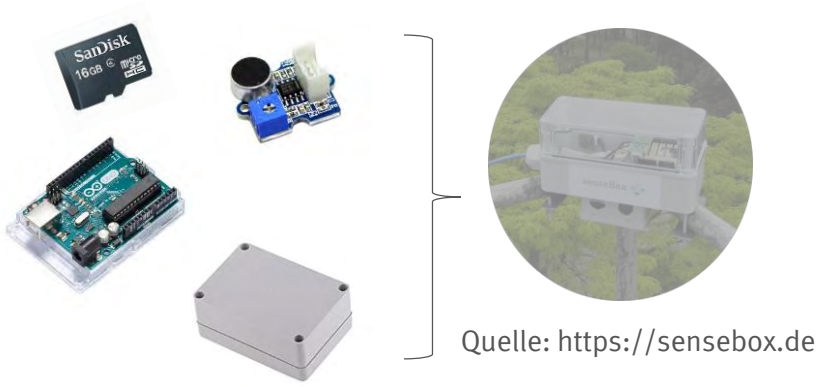
- Wöchentlich 3 Unterrichtsstunden (Montag 16:00 – 18:30)
- 25 Wochen (September 2018 – April 2019)

- Lehr-Lern-Labor *PIN-Lab* der Informatikdidaktik der Uni-Paderborn
  - Aufzeichnung durch Raumkameras und Bildschirmaufzeichnung
  - Beobachtung des Unterrichts durch 2 Personen

Modul 1: Datendetektive

- **Ziel:** Entwicklung von Datenkompetenz

- **Baustein 1:** Einführung in Data Science
  - Analyse von Akustikdaten
  - Erfassung mittels „Sensebox“
  - Verwendung von Jupyter Notebook und Python



Quelle: https://sensebox.de

Jupyter Notebook

File Edit View Insert Cell Kernel Widgets Help Not Trusted

Run Code

Sucht euch eine der folgenden CSV-Datein aus und erstellt ein entsprechendes DataFrame:

1. DATALOG_1.csv
2. DATALOG_2.csv
3. DATALOG_3.csv
4. DATALOG_4.csv
5. DATALOG_5.csv

In [1]:
```
#Hier euer Code

import csv
import pandas as pd


with open('DATALOG.CSV', 'r') as file, open('newFile.csv','w+') as newFile:
    newFile.write("Zeit,Wert\n")
    csv_reader = csv.reader(file, delimiter='\n')
    for row in csv_reader:
        s = row[0]
        if ';' in s:
            a = s.split(';')
            #print(a[0]+','+a[1]+'\n')
            newFile.write(a[0]+','+a[1]+'\n')

df = pd.read_csv('newFile.csv')
```

Daten ausgeben:

Wenn ihr diesen Befehl eingebt

```
print(df)
```

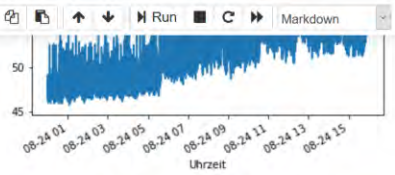
werden die Daten eures DataFrames ausgegeben.

print( ) ist eine spezielle Python-Funktion die den Inhalt der Klammer auf der Konsole ausgibt. Oft finden sich diese Funktion auch in den importierten Methode wieder.
Also muss *print()* nicht immer angegeben werden.

In [2]:
```
#Hier euer Code
print(df)
```

	Zeit	Wert
0	0	53.27
1	0	55.69
2	0	57.00
3	0	57.91
4	0	58.51
5	1	58.95
6	1	59.29
7	1	59.54
8	1	59.76
9	1	59.94

Run Markdown



Zusammenfassen der Akustikdaten:

Bis jetzt sind in unseren Graphen meist mehrere Werte pro Sekunde. Dadurch ist der Graph sehr überfüllt und sc zu vergleichen. Um die Daten noch übersichtlicher zu machen, können wir die Daten zusammenfassen. So gewi hinsichtlich des Lautstärkeprofils dieses Standortes.

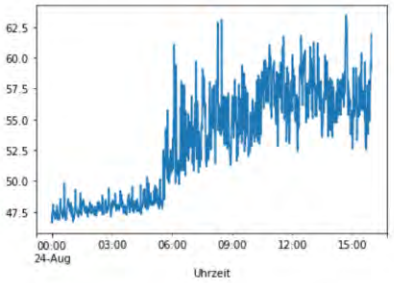
Hier können zwei Befehle nützlich sein: *resample()* und *mean()*:

```
df_Tag_zusammengefasst = df_Tag.resample('1Min').mean()
```

Das *'1Min'* steht hier für einen Zeitraum von 1 Minuten, für den die Daten zusammengefasst werden. Das Zusam Mittelwertberechnung. Ihr könnt damit mal ein bisschen rumprobieren und andere Zeiten wählen, und gucken, wa Stunden, oder Tagen! Z. B. *'30Min'*, *3H* oder *'2D'*

In [15]:
```
#Hier euer Code zum Zusammenfassen und anschließenden Zeichnen
df_Freitag_zsm = df_Freitag.resample('1Min').mean()
print(df_Freitag_zsm.Wert.plot())
```

```
AxesSubplot(0.125,0.125;0.775x0.755)
```



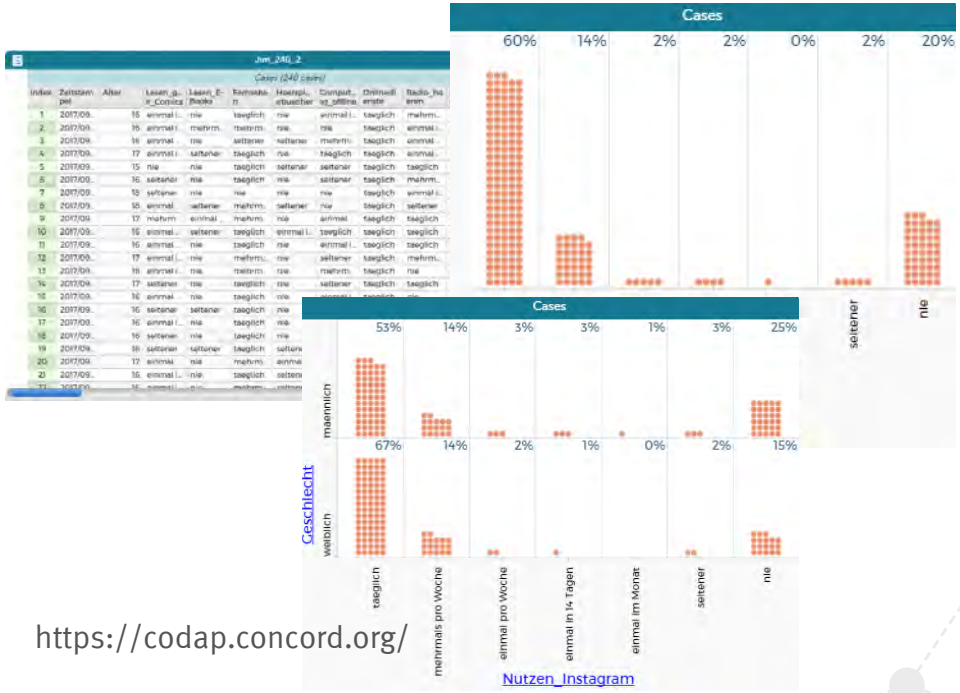
Überblick

Ihr habt nun eine Reihe an Funktionen kennengelernt, wie ihr mit den Akutsikdaten arbeiten könnt.

Hier noch einmal ein Überblick über die benutzen Methoden:

Modul 1: Datendetektive

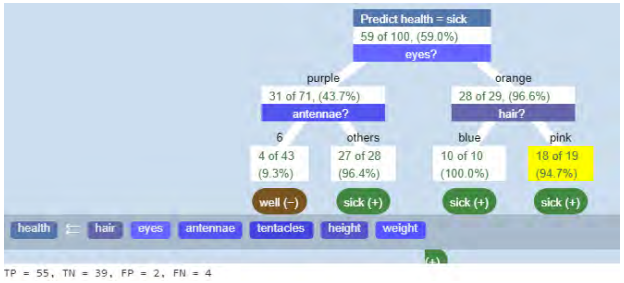
- **Baustein 2:** Einführung in die explorative Datenanalyse
  - Analyse des JIM-Datensatzes
  - Verwendung von CODAP
  - Multivariate Datenanalysen
  - Anwenden statistischer Konzepte
  - Analyse von Datenplakaten



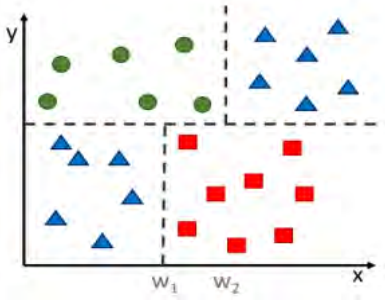
https://codap.concord.org/

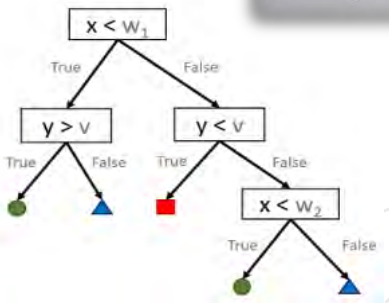
Modul 2: Einstieg in Maschinelles Lernen

- **Ziel:** Kennen von Grundkonzepten Künstlicher Intelligenz und Maschinellen Lernens
- **Baustein 1:** Einführung in Maschinelles Lernen mit Entscheidungsbäumen
  - **LAC-H** (**L**ernender **A**naloger **C**omputer für **H**exapawn)
  - Decision Trees (Entscheidungsbäume) mit CODAP (Treetool) und Jupyter Notbook u.a. mit JIM-Datensatz

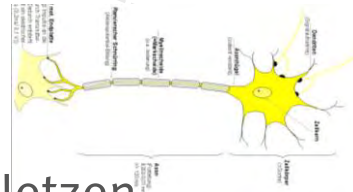


https://tinyurl.com/ydhn7fqm

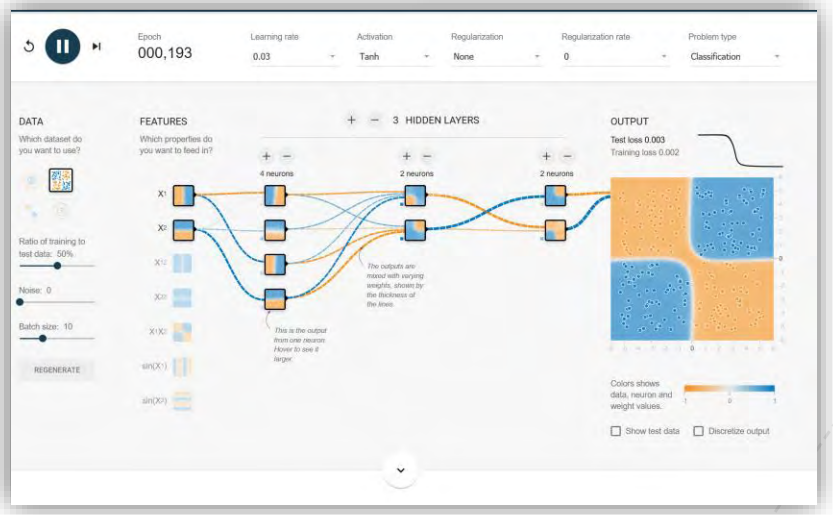


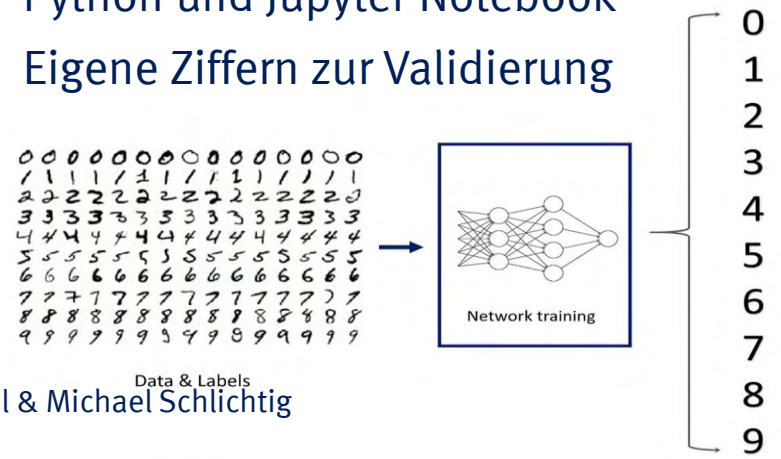


Modul 2: Einstieg in Maschinelles Lernen



- **Baustein 2:** Maschinelles Lernen mit Künstlichen Neuronalen Netzen
  - Einstieg durch Unplugged-Aktivität **„Brain in the Bag“**
  - Erkennen der Parameter und Kenngrößen durch Training von Netzen mit **„Playground Tensorflow“**
  - Modellierung eigener Netze
    - **MNIST-Datenbank** zur Erkennung handschriftlicher Ziffern
    - Python und Jupyter Notebook
    - Eigene Ziffern zur Validierung



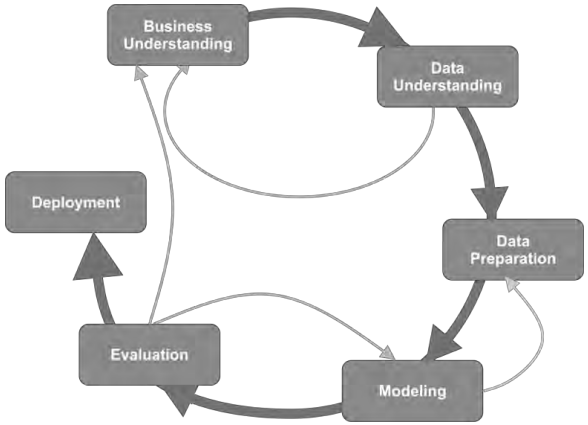


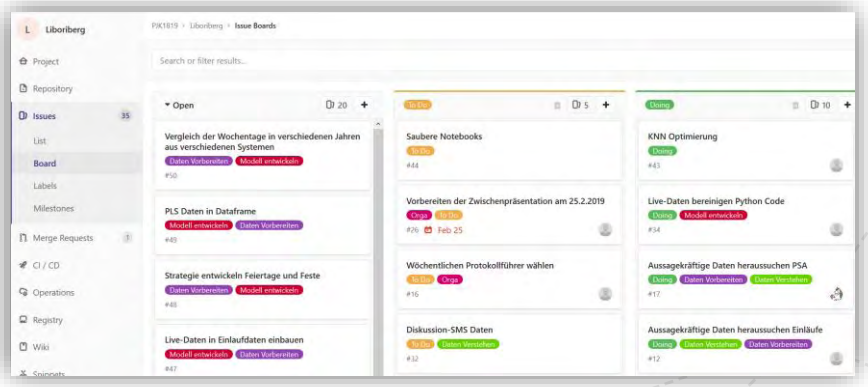
Modul 3: Wissen anwenden – Projektmodul

- Kooperation mit: Abfall- und Entsorgungsbetriebe Paderborn (ASP, Parkraumbewirtschaftung) und Fa. RTB
- **Aufgabenstellung:**
  - *Erstellung einer **Prognose** über die Belegung verschiedener bewirtschafteter Parkmöglichkeiten zu einem zukünftigen Zeitpunkt*
  - *Können die jeweiligen Belegungen – abhängig von verschiedenen Parametern – für einen zukünftigen Zeitpunkt mit einer hinreichend guten Genauigkeit **vorhergesagt** werden?*
- **Verfügbare Daten:**
  - Daten des Parkleitsystems (Induktionsschleifen)
  - Daten aus Parkautomaten und Kassensystemen

Modul 3: Wissen anwenden - Projektmodul

- **Projektgruppe 1:** Parkhausbelegung – Erfassung Parkdauer bei Bezahlung
- **Projektgruppe 2:** Parkplatzbelegung – Erfassung der geplanten Parkdauer
- Vorgehen:
  - Orientierung am **CRISP-DM Zyklus**
  - Arbeitsorganisation angelehnt an agilen Methoden
- Werkzeuge:
  - Jupyter Notebook mit Python
  - Git als Repository und Issue-Tool
- Abschlusspräsentation 1. April 2019





Weiterentwicklung des Projekts

Curriculum

- Entwicklung für Sek II allgemein
- Design Based Research-Ansatz wird vertieft
- Konzentration auf gesellschaftliche Aspekte

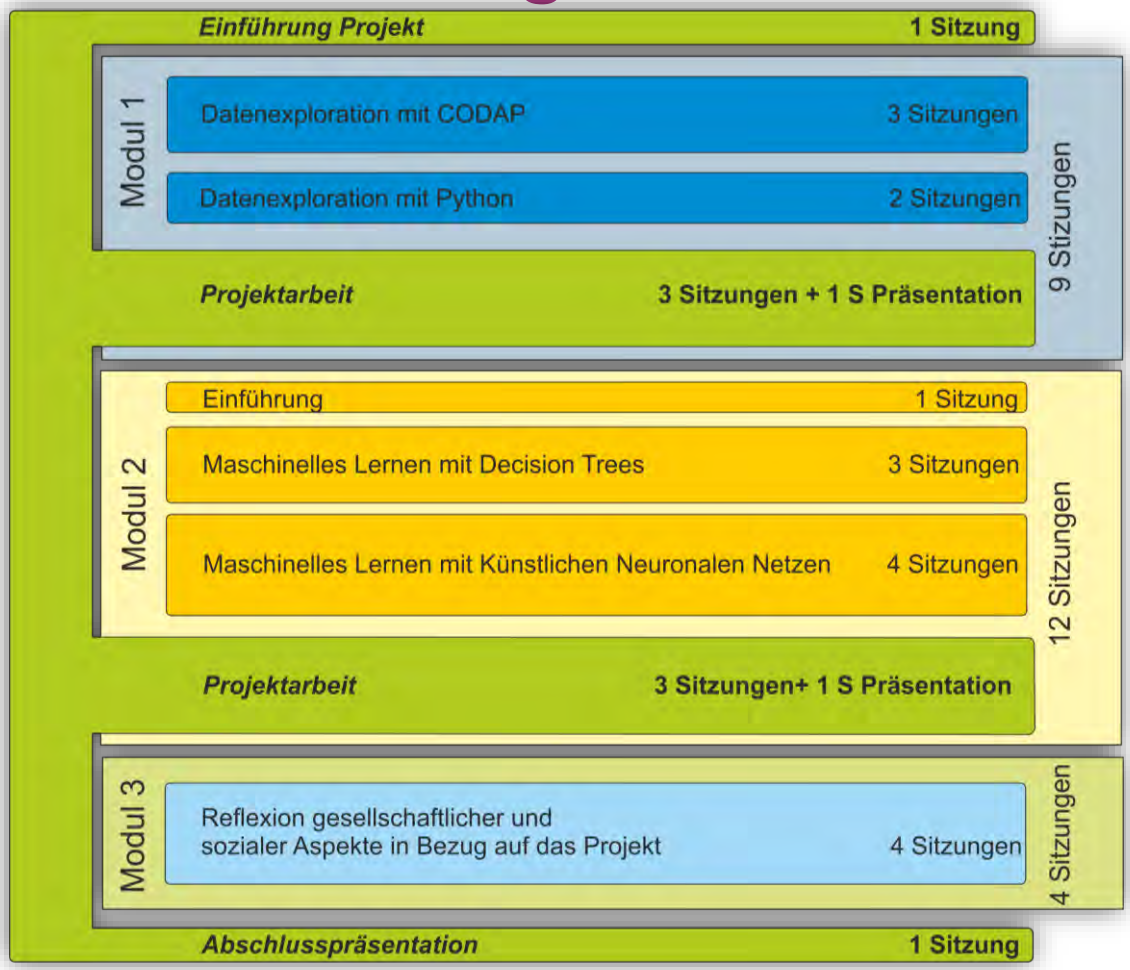
Projektkurs

- Stärkere Verzahnung des Projekts mit Theoriemodulen
- Neues Modul zu Gesellschaftlichen Auswirkungen von Big Data und Künstlicher Intelligenz

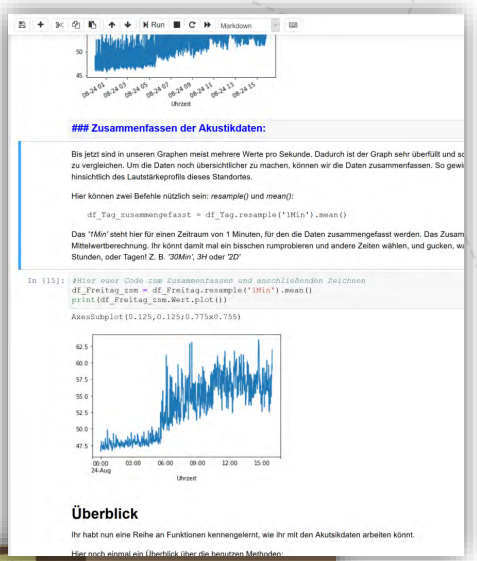
Unterrichtsmaterial

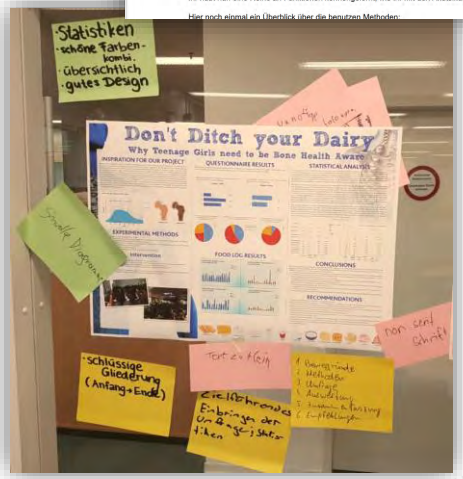
- … Bausteine als gesamter Projektkurs
- … Stand-Alone Bausteine mit minimalem Programmieranteil
- Evaluation und Weiterentwicklung in enger Verzahnung mit Lehrkräften

Geplanter Ablauf im Folgekurs



Fazit





- … es gibt noch viel zu tun
- … aber:
  - Notwendigkeit der Module erkennbar
  - Neue Kursstruktur und Modulstruktur gerade für Berufsfachschulen leichter umsetzbar
  - Einzelbausteine ermöglichen auch Einschübe oder dienen als Einstieg

- Wir suchen noch **Lehrkräfte**, die an Kooperationen interessiert sind!

https://www.prodabi.de