Data Science 1

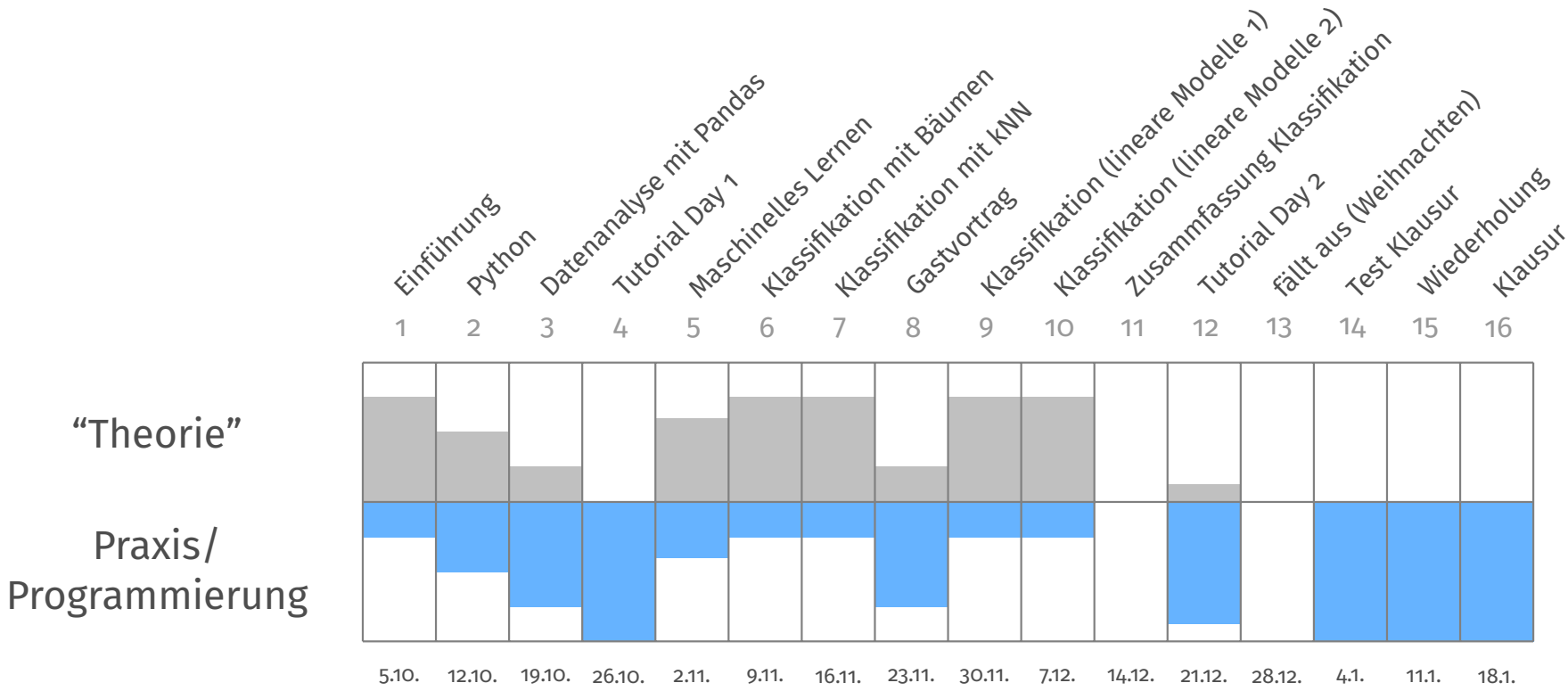
Vorlesung 2 - Wiederholung

Prof. Dr. Christian Bockermann

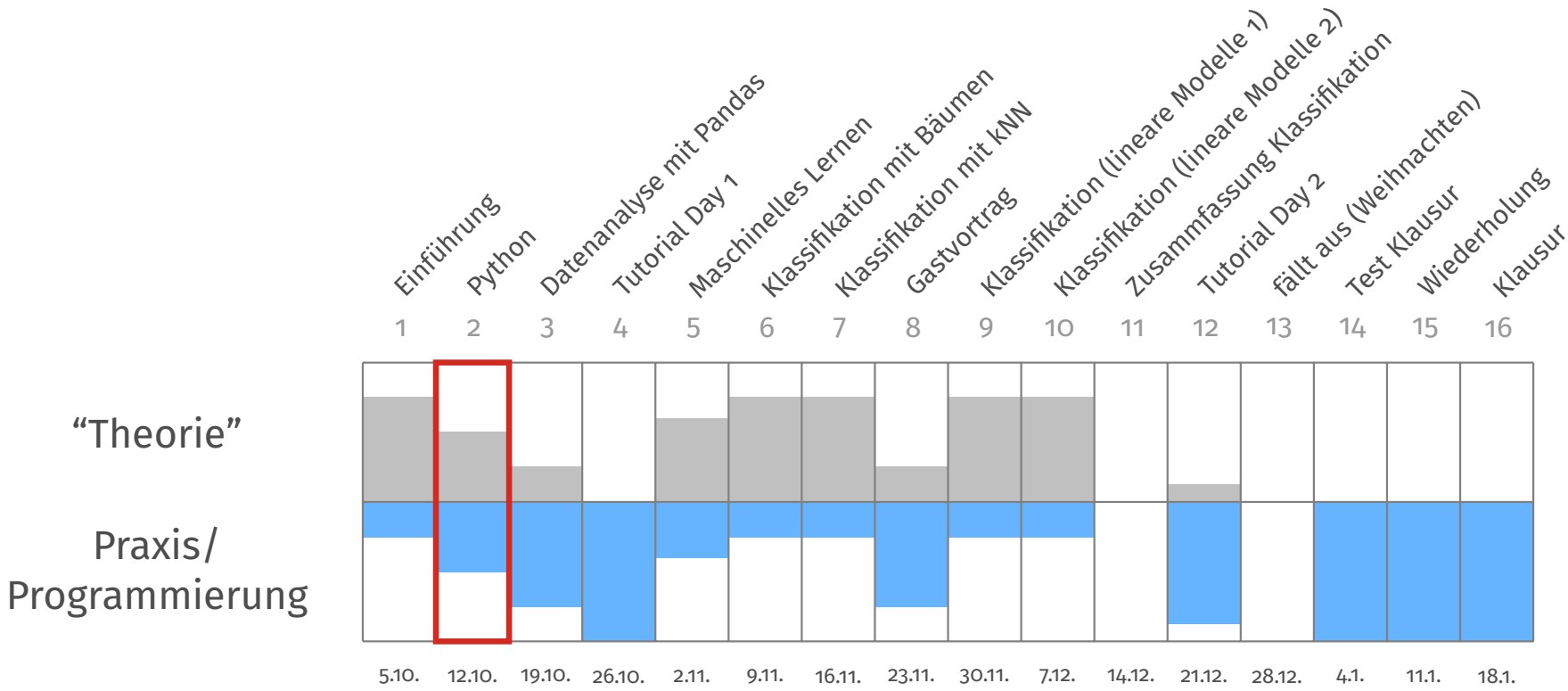
Hochschule Bochum

Wintersemester 2021/2022

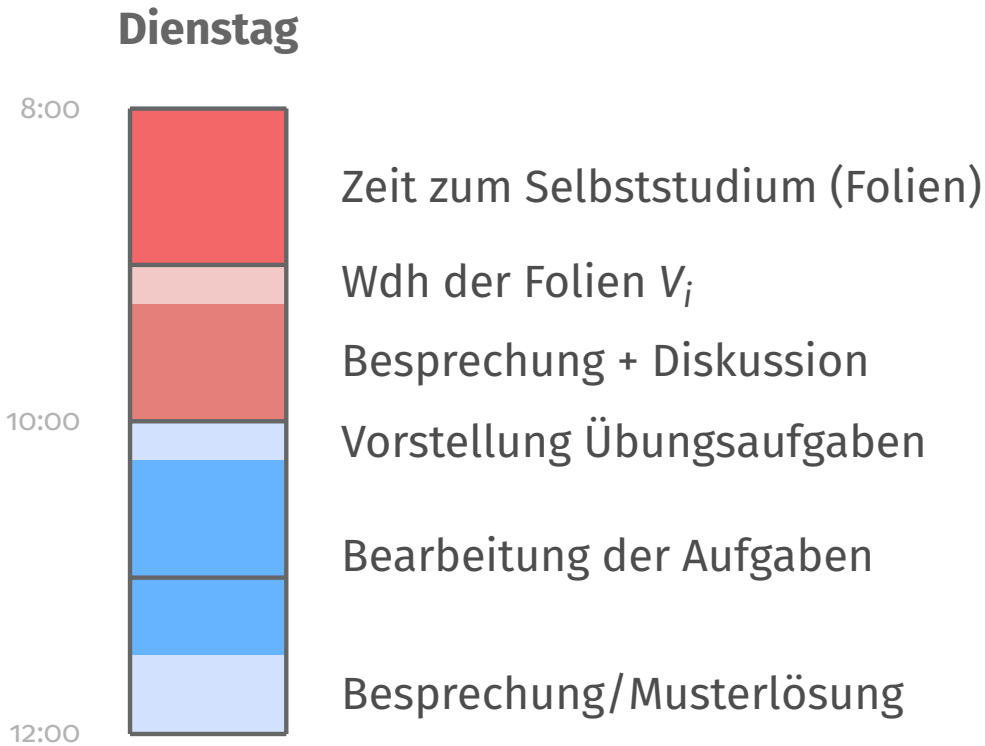
Aufbau der Vorlesung



Aufbau der Vorlesung



Ablauf der Vorlesung V_i



Wiederholung: Anwendungen für Data Science (*Use Cases*)

- Formel 1 - Überwachung, Vorhersage (Regression, Strategie)
- IP-TV - Marketing: Zielgruppenanalyse, Gruppen finden
- Handel - Kundenprofile, Gemeinsamkeiten erkennen

Wiederholung: Anwendungen für Data Science (*Use Cases*)

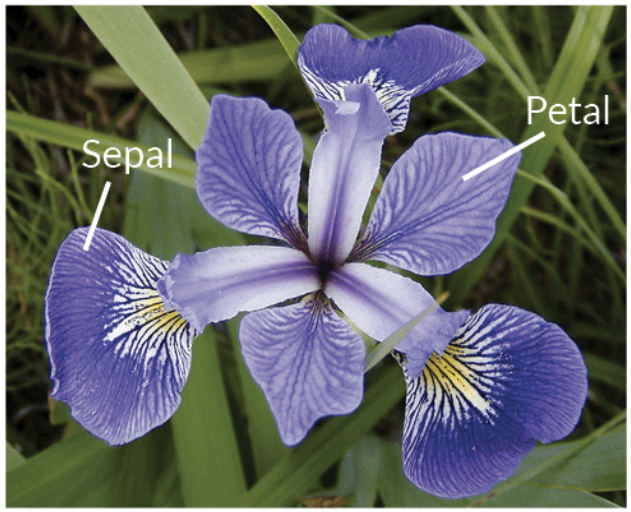
- Formel 1 - Überwachung, Vorhersage (Regression, Strategie)
- IP-TV - Marketing: Zielgruppenanalyse, Gruppen finden
- Handel - Kundenprofile, Gemeinsamkeiten erkennen

Beispiel: Klassifikation (Lernaufgabe)

- Spam-Erkennung (Text-Daten, Vorhersage: *Spam* / *No-Spam*)
- Weitere Use-Cases (aus Übungsblatt 1?)

Ausflug in die Botanik:

Ausflug in die Botanik: Schwertlilien

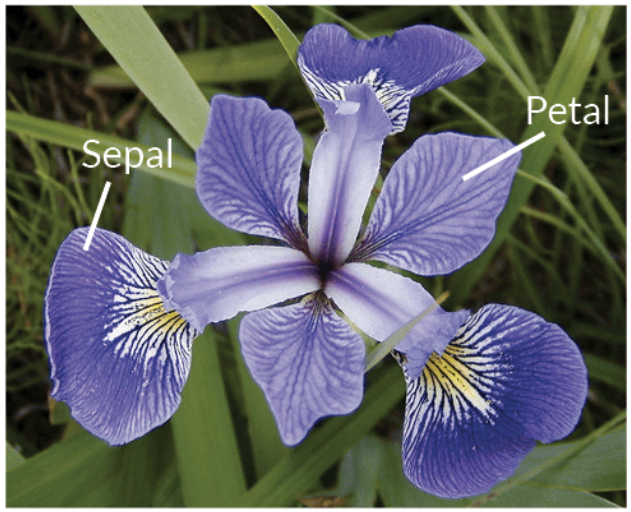


Iris Versicolor

Iris Setosa

Iris Virginica

Ausflug in die Botanik: Schwertlilien



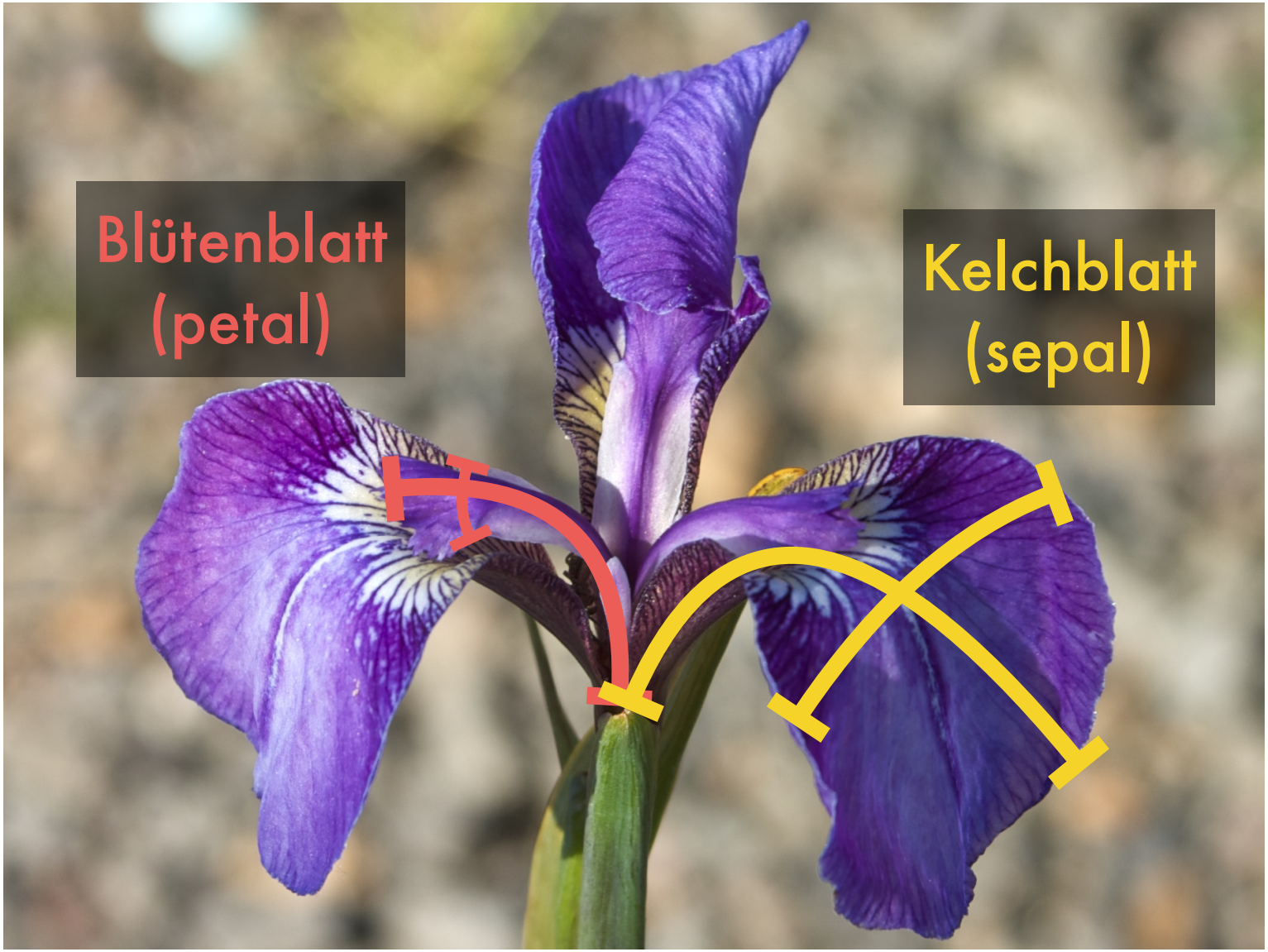
Iris Versicolor

Iris Setosa

Iris Virginica

Wie soll man die auseinanderhalten?

Beispiel: Klassifikation

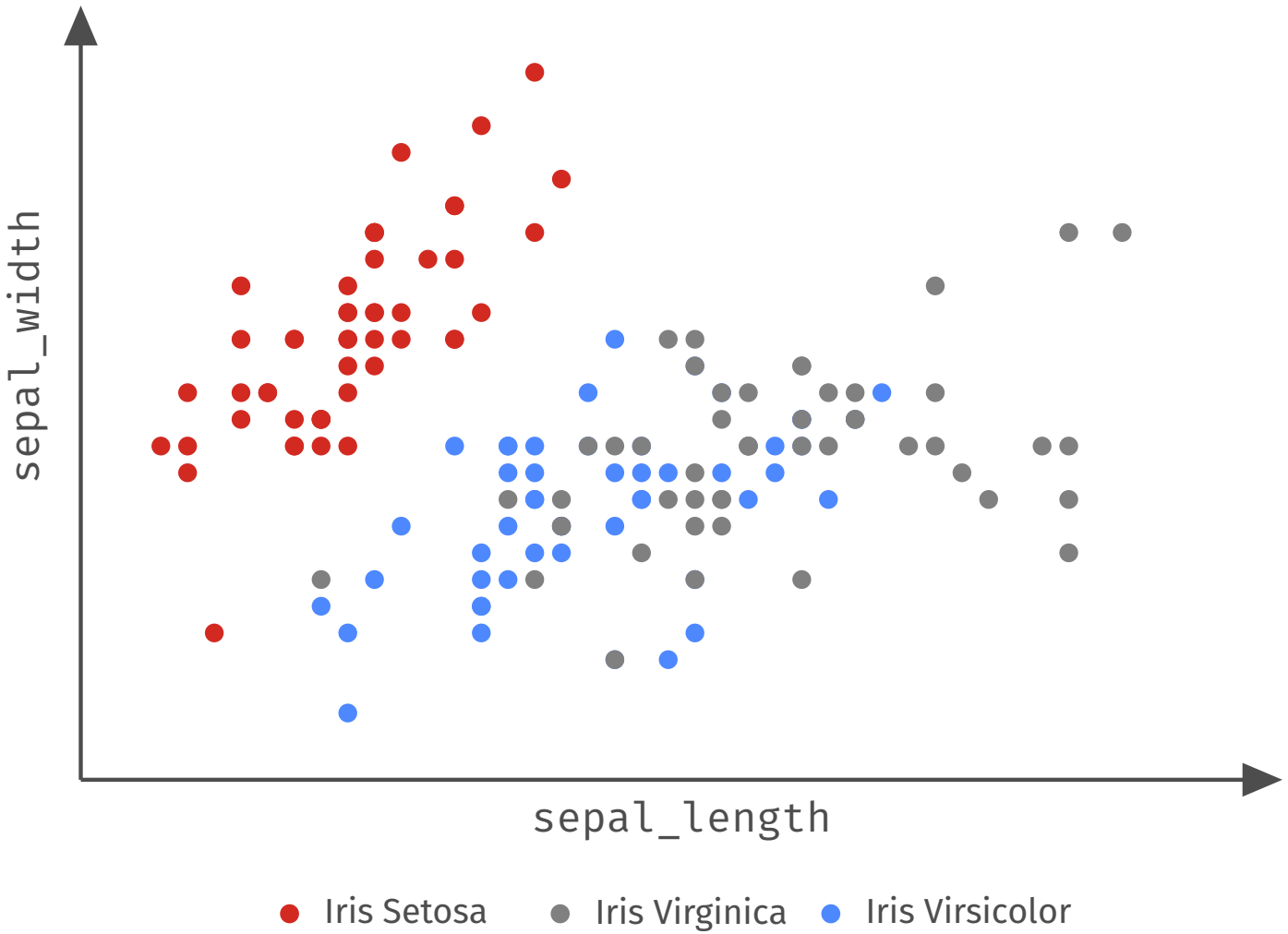


Beispiel: Klassifikation

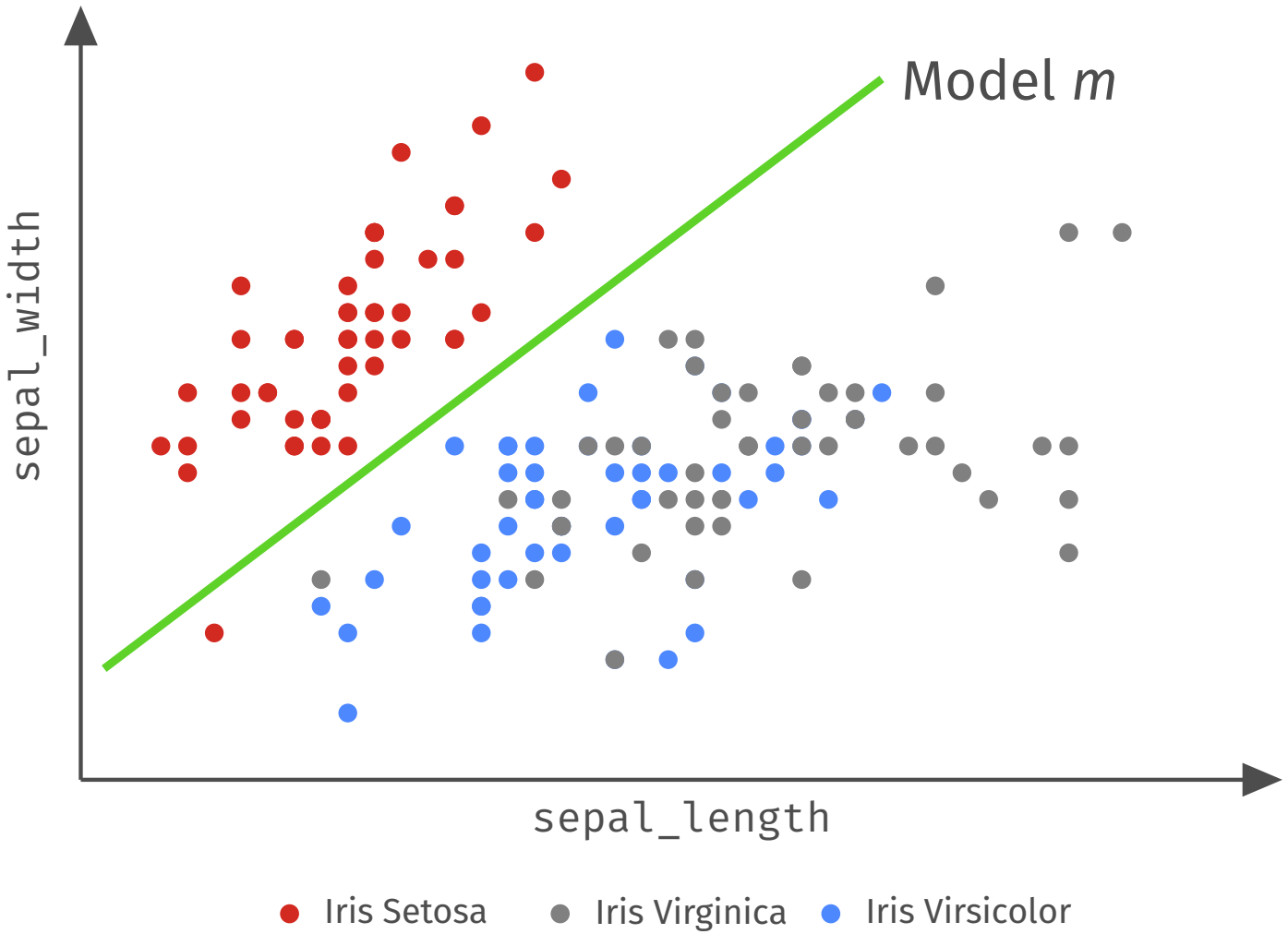
sepal_length	sepal_width	petal_length	petal_width	species
6.3	2.3	4.4	1.3	versicolor
6.4	2.7	5.3	1.9	virginica
5.4	3.7	1.5	0.2	setosa
6.1	3.0	4.6	1.4	versicolor
5.0	3.3	1.4	0.2	setosa
5.0	2.0	3.5	1.0	versicolor

Iris Datensatz, [Fisher, 1988]

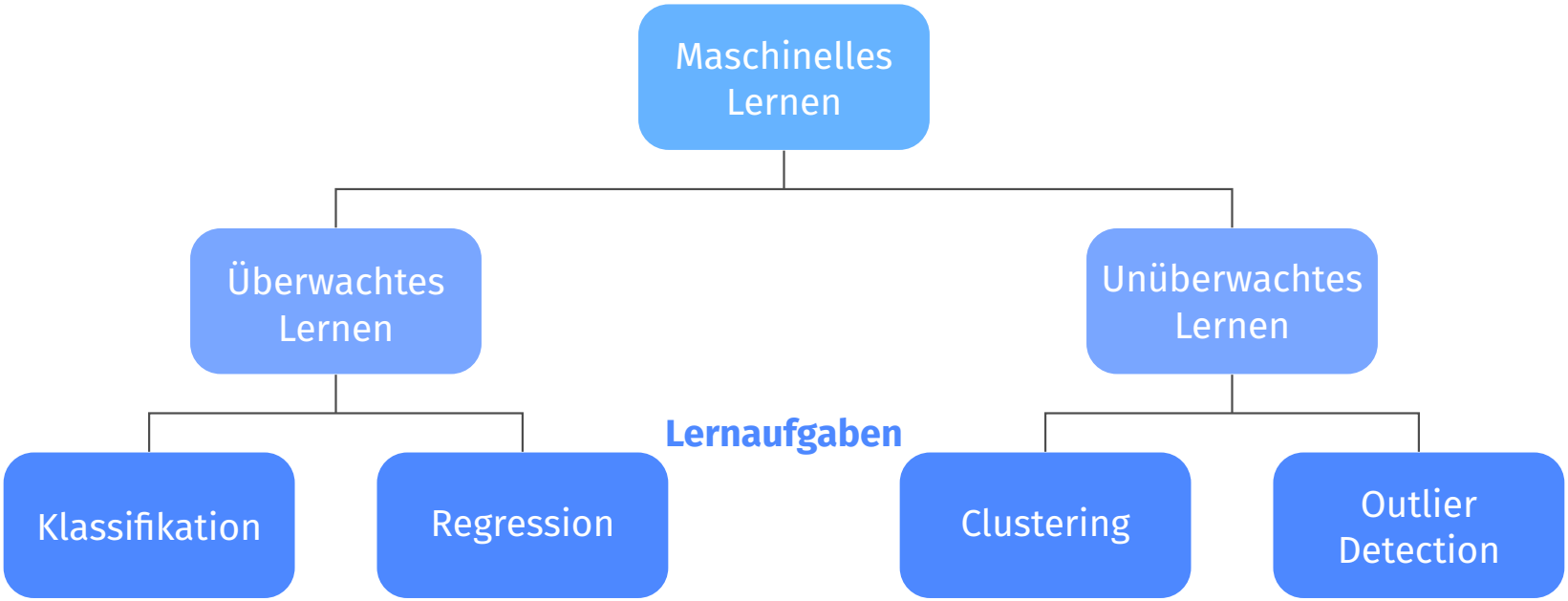
Beispiel: Klassifikation



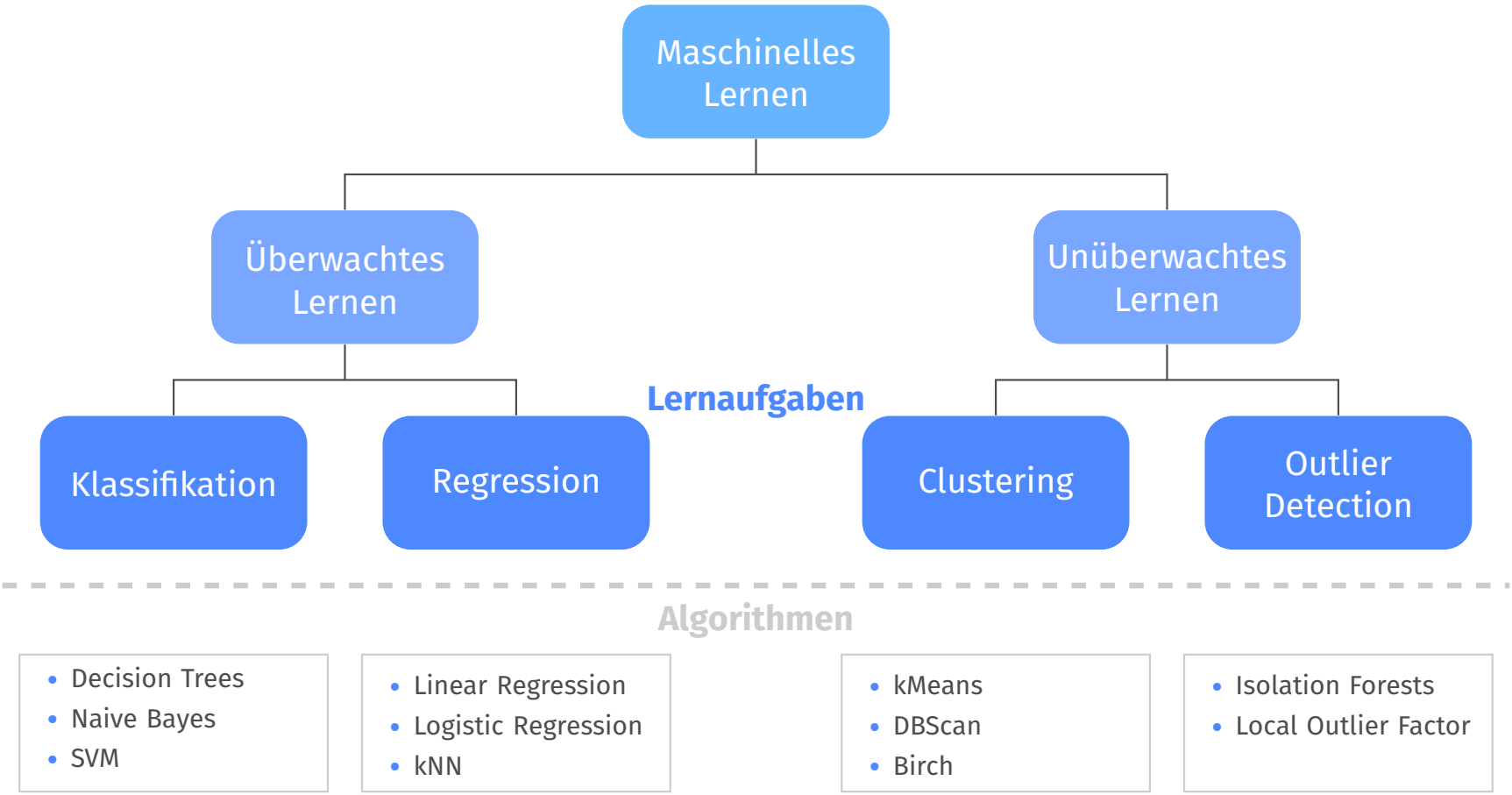
Beispiel: Klassifikation



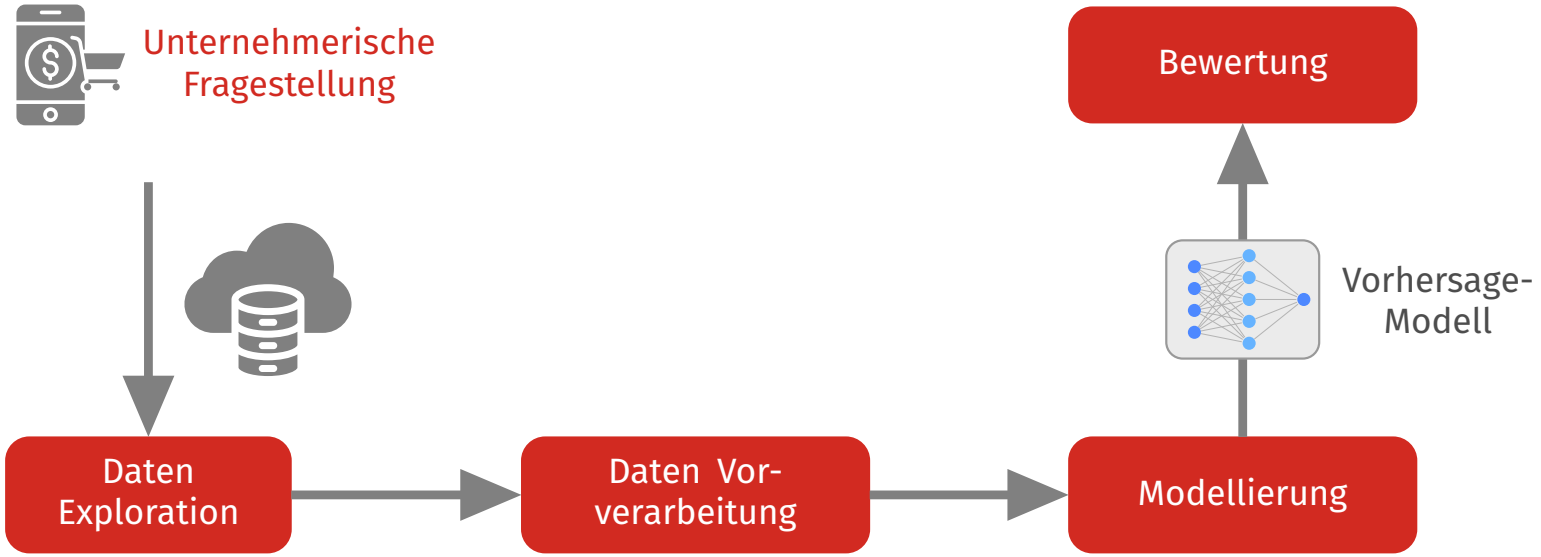
Lernaufgaben im Maschinellen Lernen



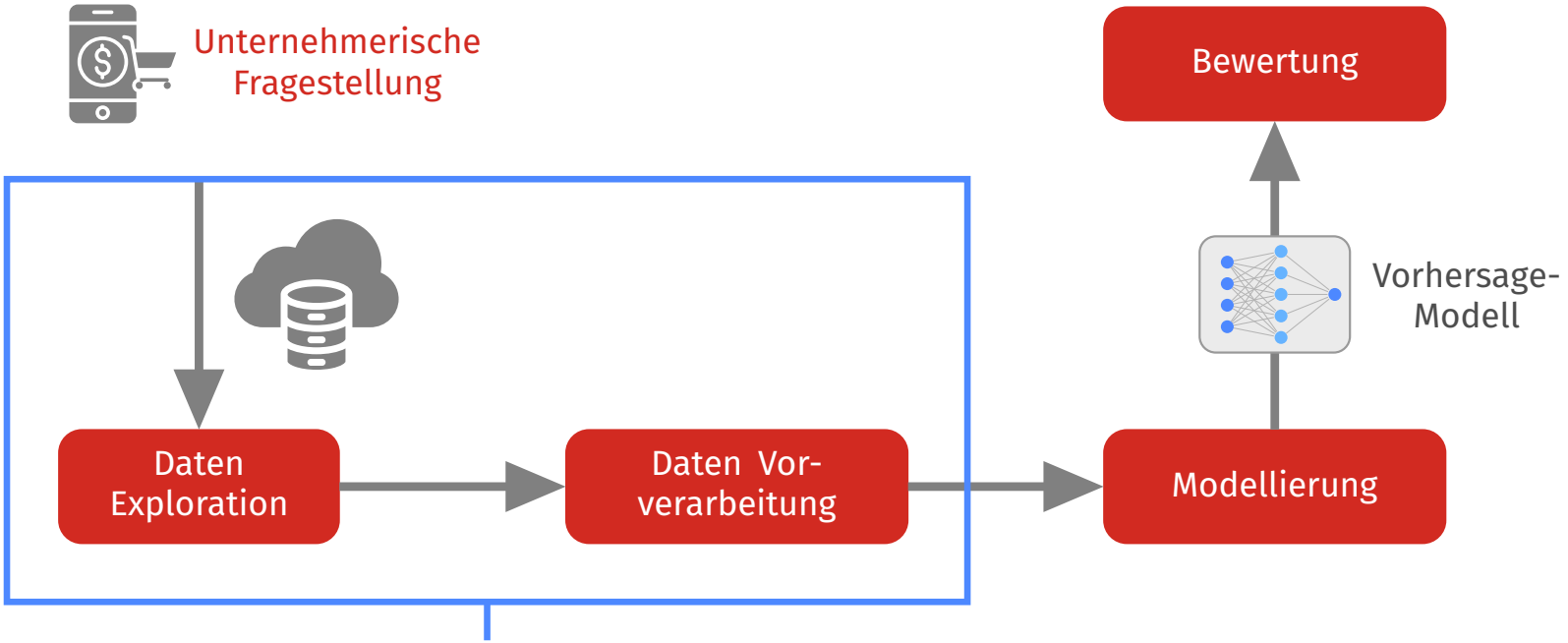
Lernaufgaben im Maschinellen Lernen



Wiederholung: Vorgehen bei der Datenanalyse (CRISP-DM)



Wiederholung: Vorgehen bei der Datenanalyse (CRISP-DM)



Datenvorverarbeitung
hier: Mit Python und Pandas

Data Science Anwendungsfälle (Aufgabe 3, Blatt 1)

- Welche Datenquellen/-arten gibt es?
- Welchen Nutzen hat ML/KI im gegebenen Anwendungsfall?
- Welche Lernaufgaben stecken dahinter?

Hochschule Bochum
Bochum University
of Applied Sciences

Fachbereich Wirtschaft
Prof. Dr. Christian Bockermann

Data Science

Wintersemester 2021/2022

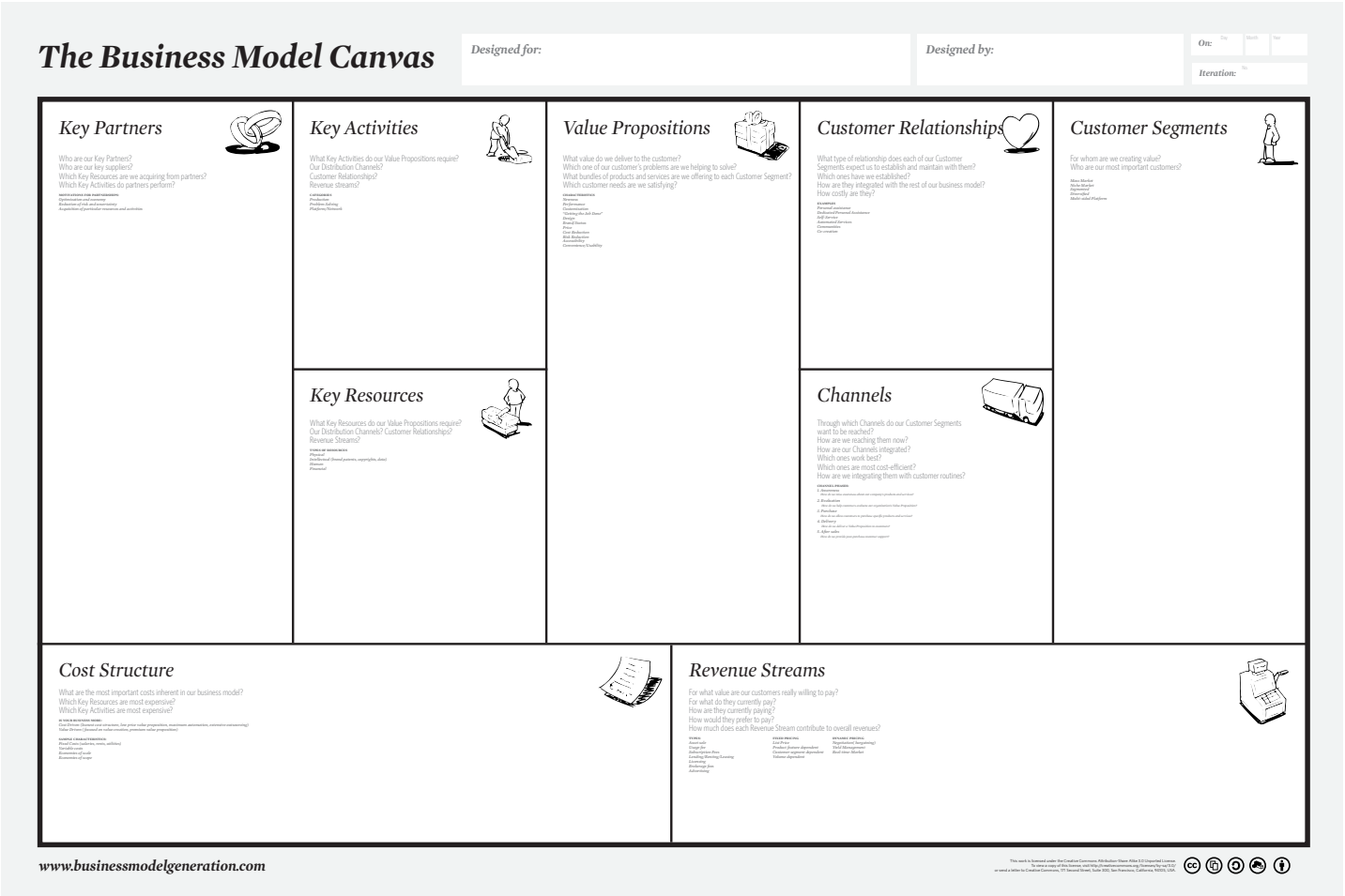
Übungsblatt 1

Aufgabe 1 (Fragebogen)

Unter der URL

`https://datascience.hs-bochum.de/umfrage/DS1`

Business Model Canvas



The Machine Learning Canvas (v0.4)

Designed for: Designed by: Date: Iteration:

Decisions

How are predictions used to make decisions that provide the proposed value to the end-user?

ML task

Input, output to predict, type of problem.

Value Propositions

What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?

Data Sources

Which raw data sources can we use (internal and external)?

Collecting Data

How do we get new data to learn from (inputs and outputs)?

Making Predictions

When do we make predictions on new inputs? How long do we have to featurize a new input and make a prediction?

Offline Evaluation

Methods and metrics to evaluate the system before deployment.

Features

Input representations extracted from raw data sources.

Building Models

When do we create/update models with new training data? How long do we have to featurize training inputs and create a model?

Live Evaluation and Monitoring

Methods and metrics to evaluate the system after deployment, and to quantify value creation.

Vorlesung 2 (heute):

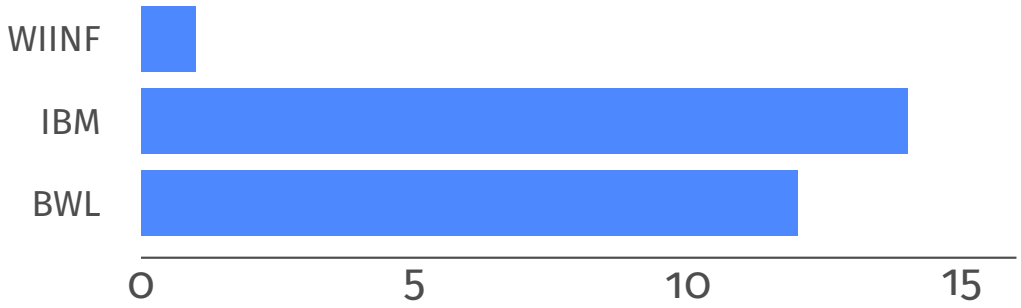
- Python “Crashkurs” – Grundlagen/Überblick
- Jupyter-Notebooks für Übungen

Vorlesung 2 (heute):

- Python “Crashkurs” – Grundlagen/Überblick
- Jupyter-Notebooks für Übungen

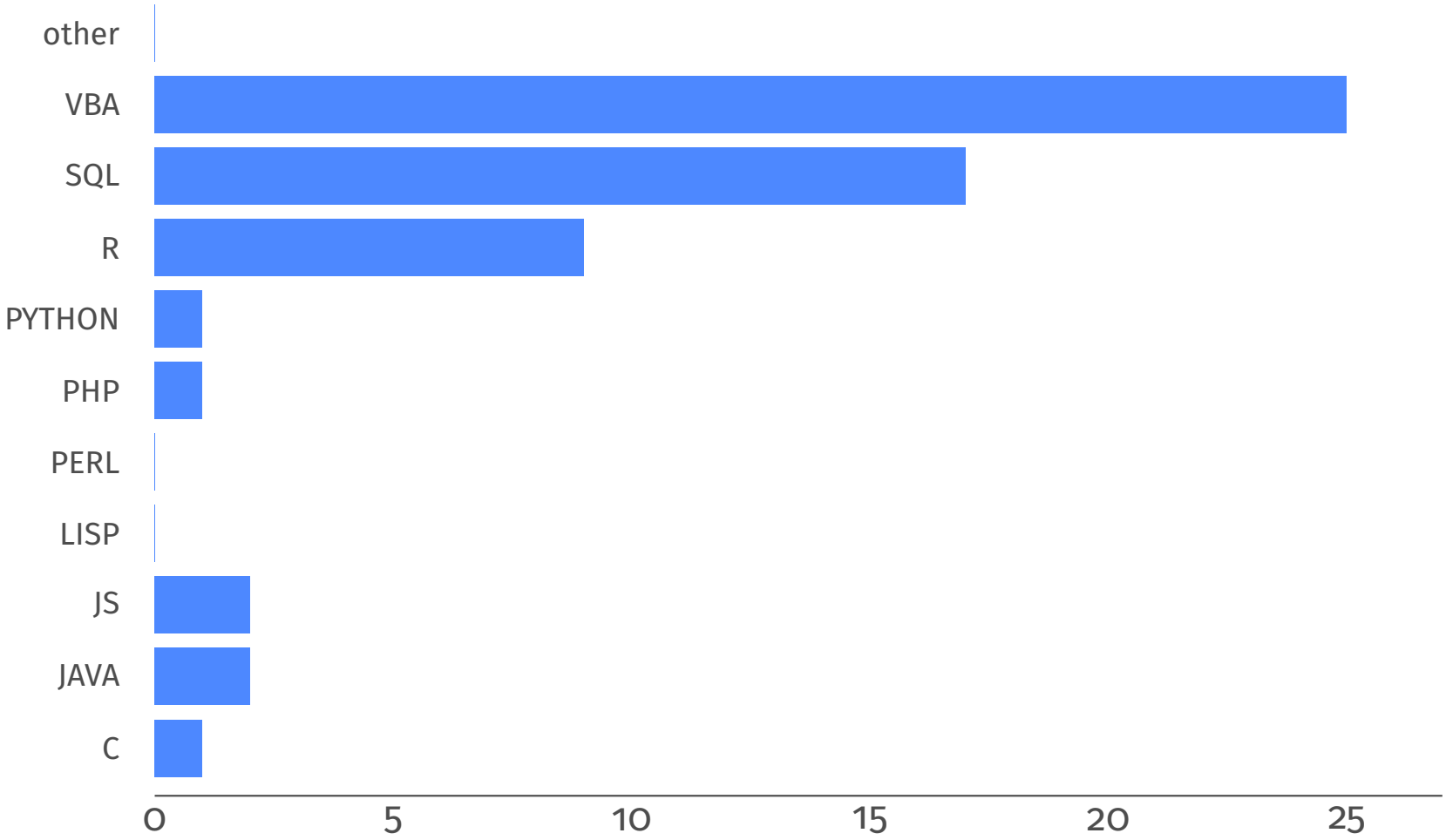
Vertiefung von Python wird die Vorlesung begleiten.

Vertretene Studiengänge



Berufziele aus dem Fragebogen

Vorkenntnisse: Programmiersprachen



```
Function berechneKosten(besucherAnzahl As Double,
                        groesse As String) As
                        Double
    Const pauschale As Double = 2500.0
    Const ticketPreis As Double = 25.0
    Const aufschlagStadion As Double = 0.10

    Dim kosten As Double
    kosten = 0

    kosten = pauschale + besucherAnzahl * ticketPreis

    If groesse = 'Stadion' Then
      kosten = kosten + kosten * aufschlagStadion
    End If

    berechneKosten = kosten
End Function
```

```
def berechneKosten(besucherAnzahl, groesse):
  pauschale = 2500.0
  ticketPreis = 25.0
  aufschlagStadion = 0.10

  kosten = pauschale + besucherAnzahl * ticketPreis

  if groesse == 'Stadion':
    kosten = kosten + kosten * aufschlagStadion

  return kosten
```

Vorschau auf nächste Woche

Vorlesung 3 (nächste Woche):

- Vorstellung des Pandas Moduls (Python)
- Lesen, Verarbeiten und Visualisieren von Daten