

DATA SCIENCE 1

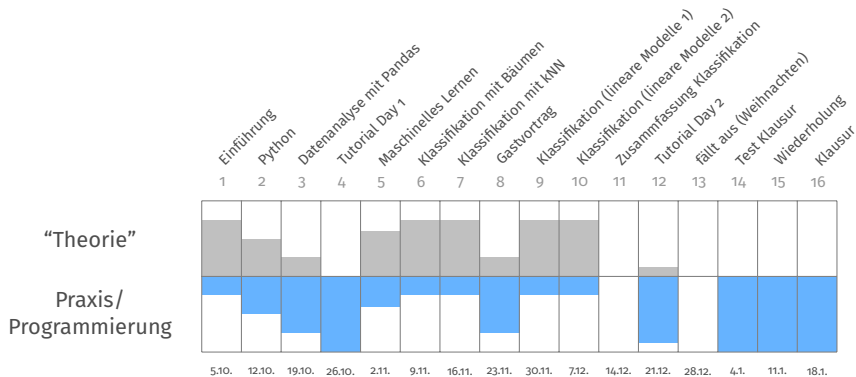
VORLESUNG 2 - WIEDERHOLUNG

PROF. DR. CHRISTIAN BOCKERMANN

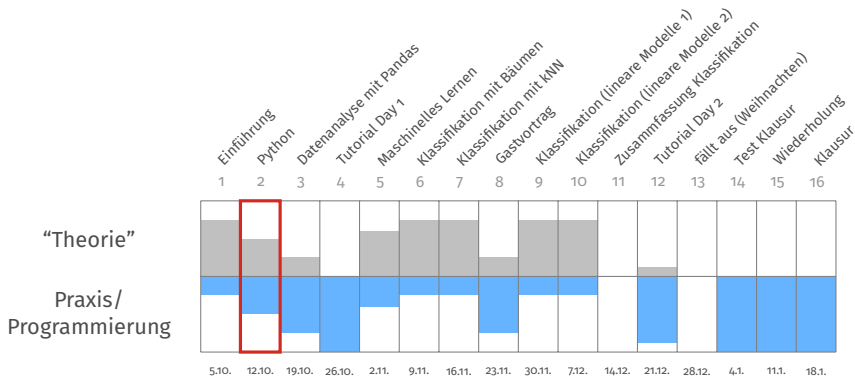
HOCHSCHULE BOCHUM

WINTERSEMESTER 2021/2022

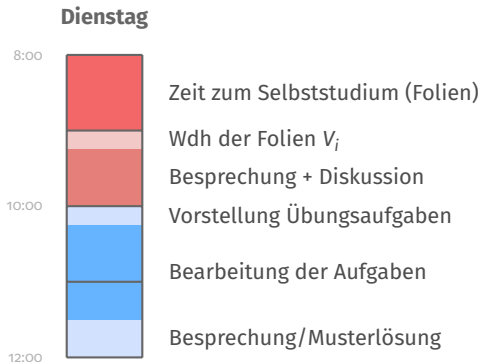
Aufbau der Vorlesung



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

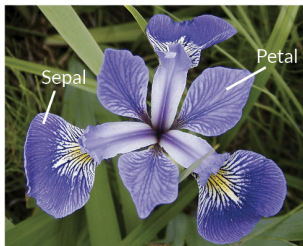
Beispiel: Klassifikation (Lernaufgabe)

- 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

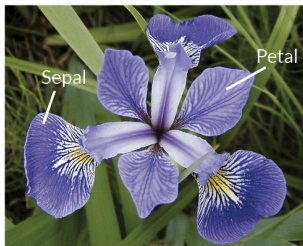


Iris Setosa



Iris Virginica

Ausflug in die Botanik: Schwertlilien



Iris Versicolor

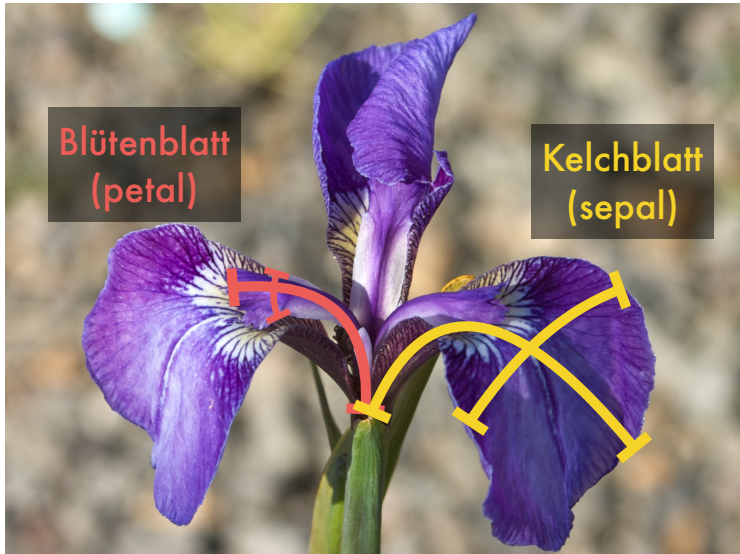


Iris Setosa



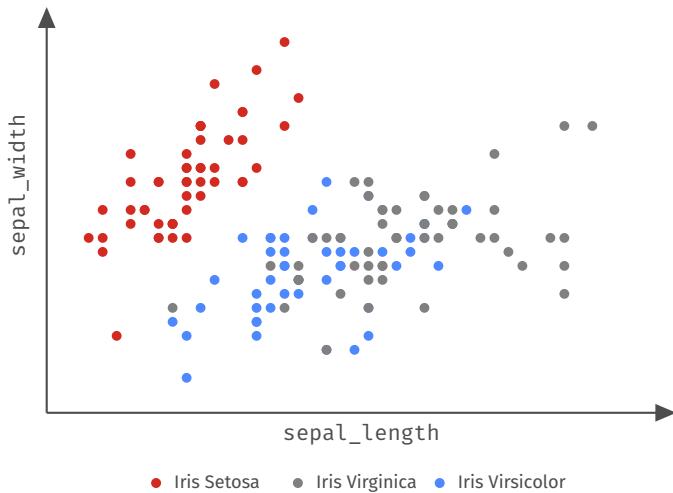
Iris Virginica

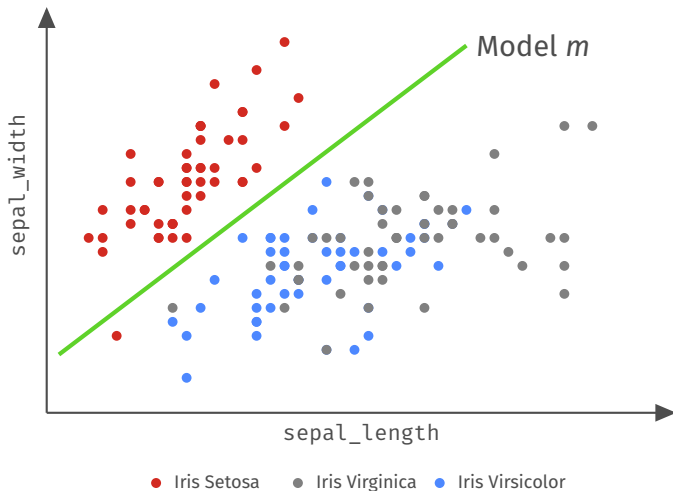
Wie soll man die auseinanderhalten?



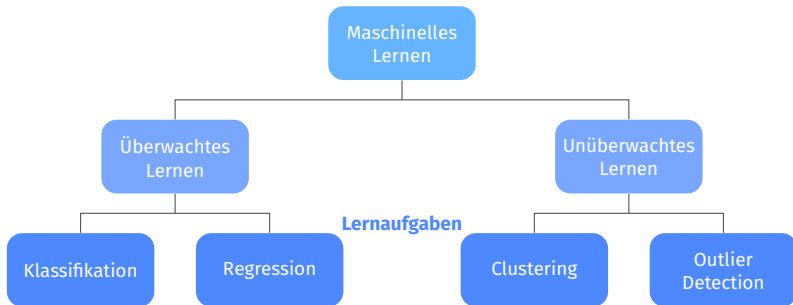
sepal_length	sepal_width	petal_length	petal_width	species
6.3	2.3	4.4	1.3	versicolor
6.4	2.7	5.3	1.9	virginica
5.4	3.7	1.5	0.2	setosa
6.1	3.0	4.6	1.4	versicolor
5.0	3.3	1.4	0.2	setosa
5.0	2.0	3.5	1.0	versicolor

Iris Datensatz, [Fisher, 1988]

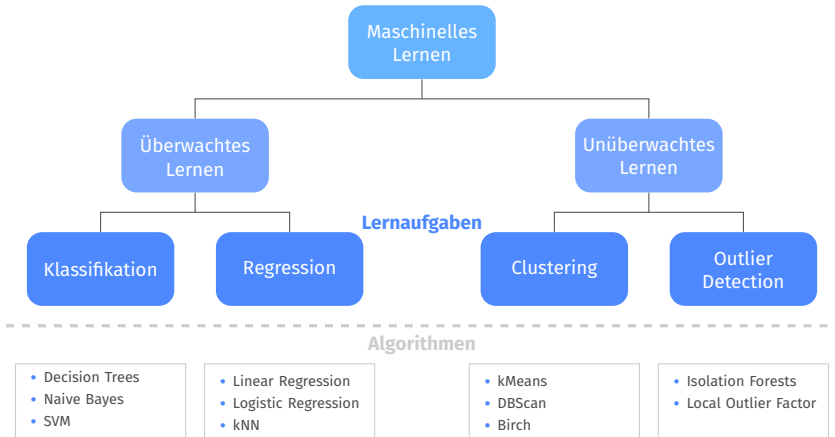




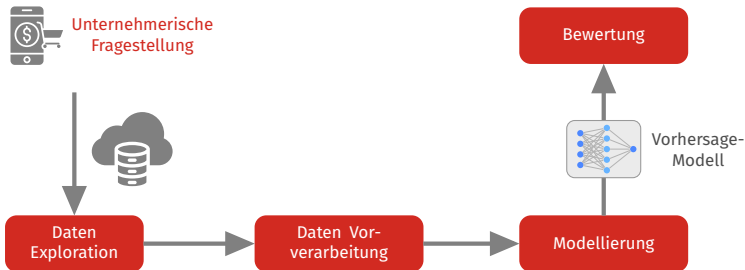
Lernaufgaben im Maschinellen Lernen



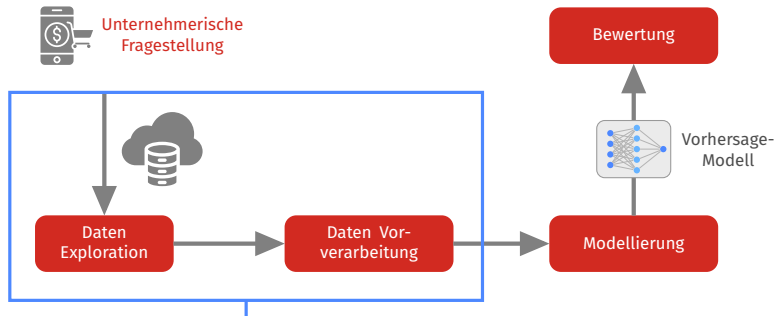
Lernaufgaben im Maschinellen Lernen



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 Business Model Canvas

Designed for: _____

Designed by: _____

Date: _____

Version: _____

<p>Key Partners</p>  <p>Who are the partners? What are they expected to do? What are the expected gains resulting from partners? What are the risks of not having partners? What are the risks of having partners?</p>	<p>Key Activities</p>  <p>What key activities do our Value Propositions require? Do they have a cost? Can they be automated? Can they be outsourced?</p>	<p>Value Propositions</p>  <p>What value does our offer bring to customers? What value does our customer segments value? What value does our customer segments expect? What value do we create for our business model? What value do we capture? What value do we lose?</p>	<p>Customer Relationships</p>  <p>What type of relationship do we want with our customers? What type of relationship do we have with our customers? What type of relationship do our competitors have with their customers? What type of relationship do we want to lose? What type of relationship do we want to gain?</p>	<p>Customer Segments</p>  <p>Who are our segments? What are our segments' needs? What are our segments' expectations? What are our segments' behaviors? What are our segments' characteristics?</p>
	<p>Key Resources</p>  <p>What key resources do our Value Propositions require? Do they have a cost? Can they be automated? Can they be outsourced?</p>		<p>Channels</p>  <p>Through which channels do we want to reach our customers? What are our channels? What are our channels' costs? What are our channels' reach? What are our channels' efficiency? What are our channels' flexibility? What are our channels' scalability?</p>	
<p>Cost Structure</p>  <p>What are the expected costs of our business model? What are the expected costs of our business model? What are the expected costs of our business model? What are the expected costs of our business model? What are the expected costs of our business model?</p>		<p>Revenue Streams</p>  <p>How do we generate revenue? What are our revenue streams? What are our revenue streams' costs? What are our revenue streams' reach? What are our revenue streams' efficiency? What are our revenue streams' flexibility? What are our revenue streams' scalability?</p>		

www.businessmodelgeneration.com

Icons: CC BY-NC-SA

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):

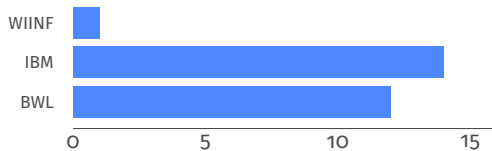
- Python “Crashkurs” – Grundlagen/Überblick
- Jupyter-Notebooks für Übungen

Vorlesung 2 (heute):

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

Vertiefung von Python wird die Vorlesung begleiten.

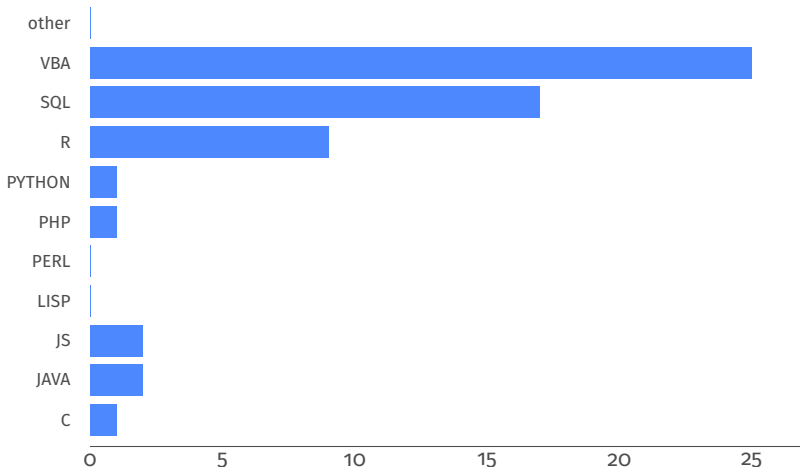
Vertretene Studiengänge



Berufsziele 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
```

Vorlesung 3 (nächste Woche):

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