

DATA SCIENCE

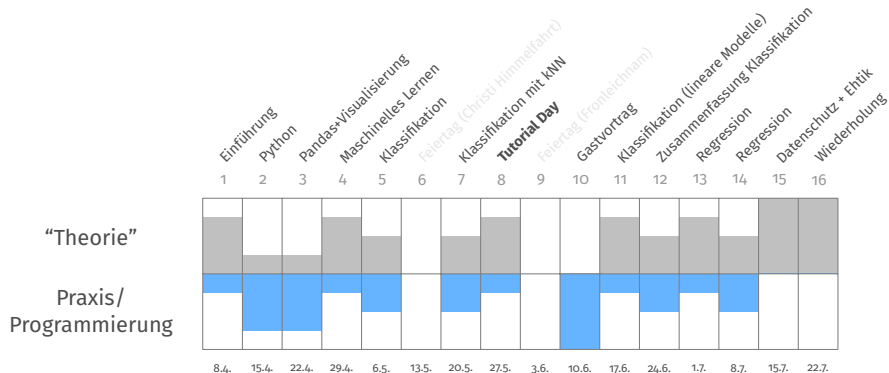
VORLESUNG 2 - WIEDERHOLUNG

PROF. DR. CHRISTIAN BOCKERMANN

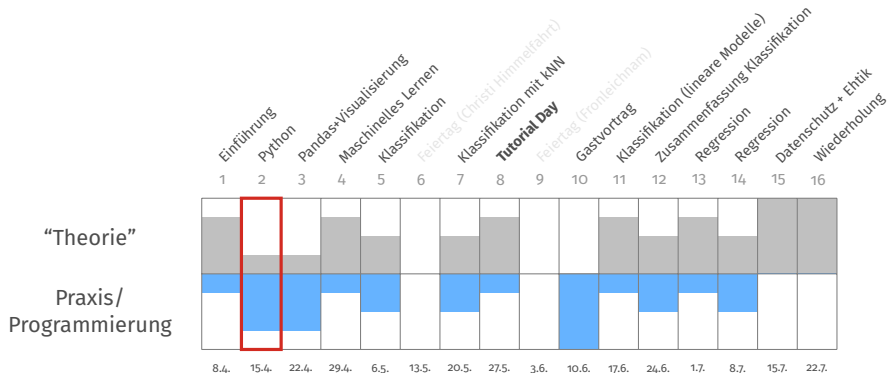
HOCHSCHULE BOCHUM

SOMMERSEMESTER 2021

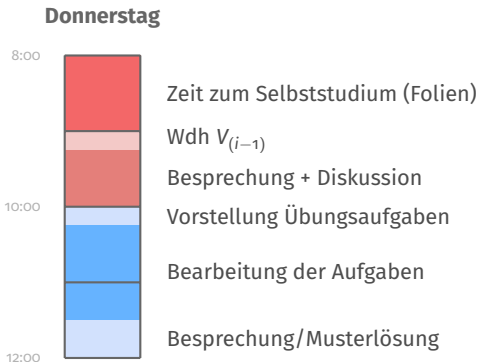
Aufbau der Vorlesung



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

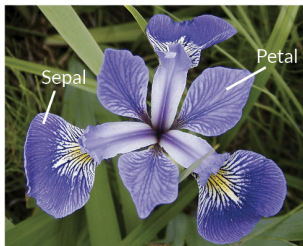
Beispiel: Klassifikation (Lernaufgabe)

- 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

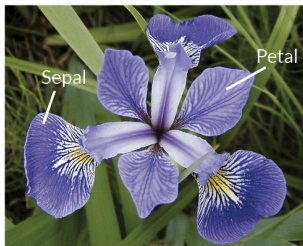


Iris Setosa



Iris Virginica

Ausflug in die Botanik: Schwertlilien



Iris Versicolor

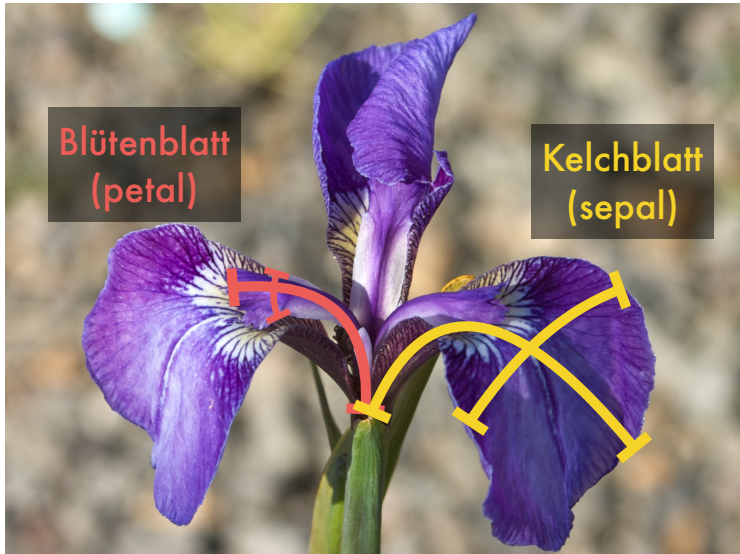


Iris Setosa



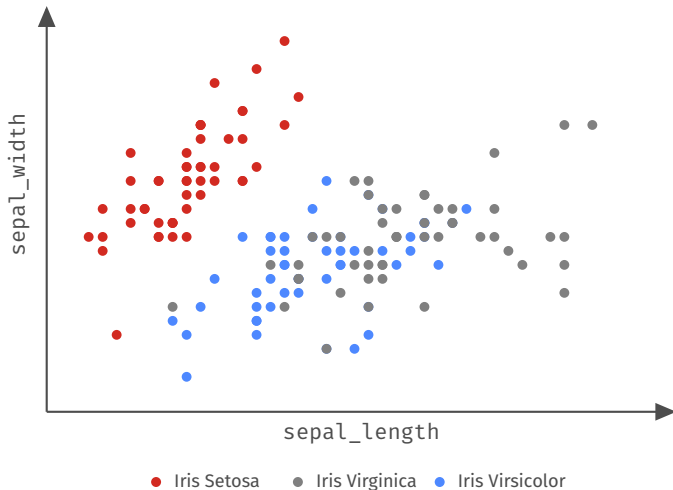
Iris Virginica

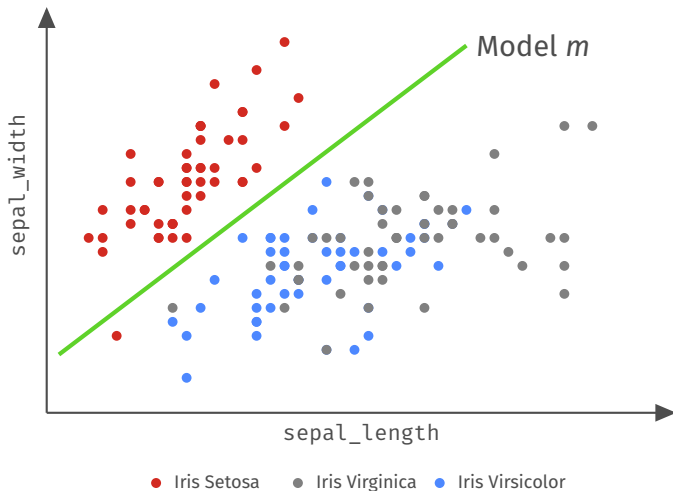
Wie soll man die auseinanderhalten?



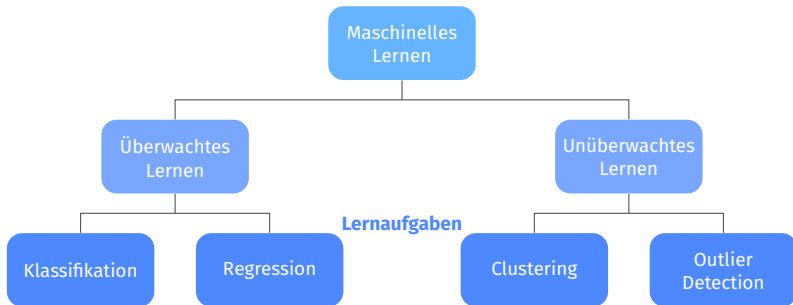
sepal_length	sepal_width	petal_length	petal_width	species
6.3	2.3	4.4	1.3	versicolor
6.4	2.7	5.3	1.9	virginica
5.4	3.7	1.5	0.2	setosa
6.1	3.0	4.6	1.4	versicolor
5.0	3.3	1.4	0.2	setosa
5.0	2.0	3.5	1.0	versicolor

Iris Datensatz, [Fisher, 1988]

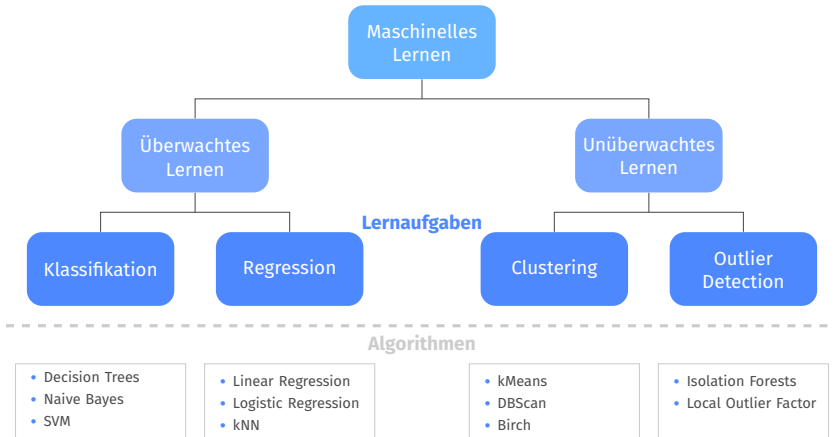




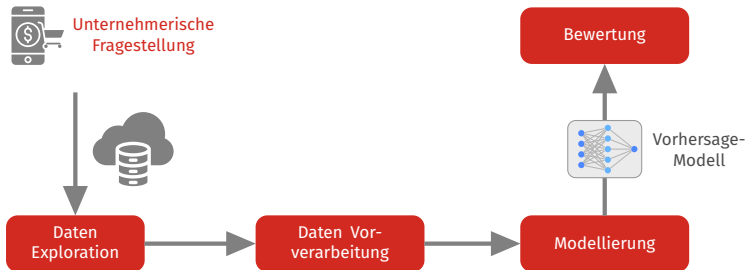
Lernaufgaben im Maschinellen Lernen



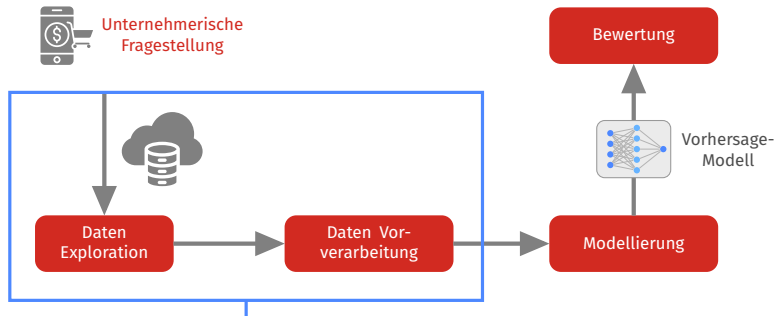
Lernaufgaben im Maschinellen Lernen



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

Sommersemester 2021

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? What are the expected costs? What are the expected benefits?</p>	<p>Key Activities</p>  <p>What key activities do our Value Propositions require? Do they have a Cost Advantage? Do they have a Differentiation Advantage? Do they have a Channel Advantage?</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 does our customer segments need? What value does our customer segments want? What value does our customer segments expect to receive? What value does our customer segments expect to pay for?</p>	<p>Customer Relationships</p>  <p>What type of relationship does our offer bring to customers? What type of relationship does our customer segments value? What type of relationship does our customer segments expect? What type of relationship does our customer segments need? What type of relationship does our customer segments want? What type of relationship does our customer segments expect to receive? What type of relationship does our customer segments expect to pay for?</p>	<p>Customer Segments</p>  <p>For whom are we creating value? What are the segments? What are the segments' needs? What are the segments' expectations? What are the segments' behaviors? What are the segments' characteristics?</p>
<p>Key Resources</p>  <p>What key resources do our Value Propositions require? Do they have a Cost Advantage? Do they have a Differentiation Advantage? Do they have a Channel Advantage?</p>			<p>Channels</p>  <p>Through which channels do our Customer Segments want to be reached? What are the channels? What are the channels' characteristics? What are the channels' costs? What are the channels' benefits? What are the channels' risks? What are the channels' opportunities?</p>	
<p>Cost Structure</p>  <p>What are the expected costs? What are the expected benefits? What are the expected risks? What are the expected opportunities?</p>		<p>Revenue Streams</p>  <p>For what value are our customers willing to pay? What are the revenue streams? What are the revenue streams' characteristics? What are the revenue streams' costs? What are the revenue streams' benefits? What are the revenue streams' risks? What are the revenue streams' opportunities?</p>		

www.businessmodelgeneration.com

Icons: CC BY-NC-SA











The Machine Learning Canvas (v0.4)

Designed for:

Designed by:

Date:

Iteration:

<p>Decisions </p> <p>How are predictions used to make decisions that provide the proposed value to the end-user?</p>	<p>ML task </p> <p>Input, output to predict, type of problem.</p>	<p>Value Propositions </p> <p>What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?</p>	<p>Data Sources </p> <p>Which raw data sources can we use (internal and external)?</p>	<p>Collecting Data </p> <p>How do we get new data to learn from (inputs and outputs)?</p>
<p>Making Predictions </p> <p>When do we make predictions on new inputs? How long do we have to featurize a new input and make a prediction?</p>	<p>Offline Evaluation </p> <p>Methods and metrics to evaluate the system before deployment.</p>	<p>Features </p> <p>Input representations extracted from raw data sources.</p>		<p>Building Models </p> <p>When do we create/update models with new training data? How long do we have to featurize training inputs and create a model?</p>
<p>Live Evaluation and Monitoring </p> <p>Methods and metrics to evaluate the system after deployment, and to quantify value creation.</p>				

Vorlesung 2 (heute):

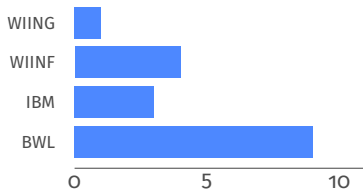
- Python “Crashkurs” – Grundlagen/Überblick
- Jupyter-Notebooks für Übungen

Vorlesung 2 (heute):

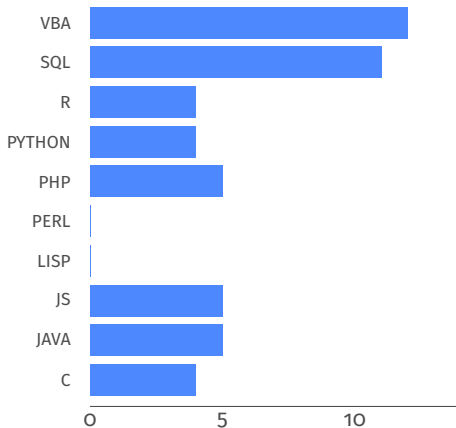
- Python “Crashkurs” – Grundlagen/Überblick
- Jupyter-Notebooks für Übungen

Vertiefung von Python wird die Vorlesung begleiten.

Vertretene Studiengänge



Vorkenntnisse: Programmiersprachen



Berufziele aus dem Fragebogen



Vorlesung 3 (nächste Woche):

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