



Data Science

Aus Unternehmensdaten
Mehrwert schöpfen



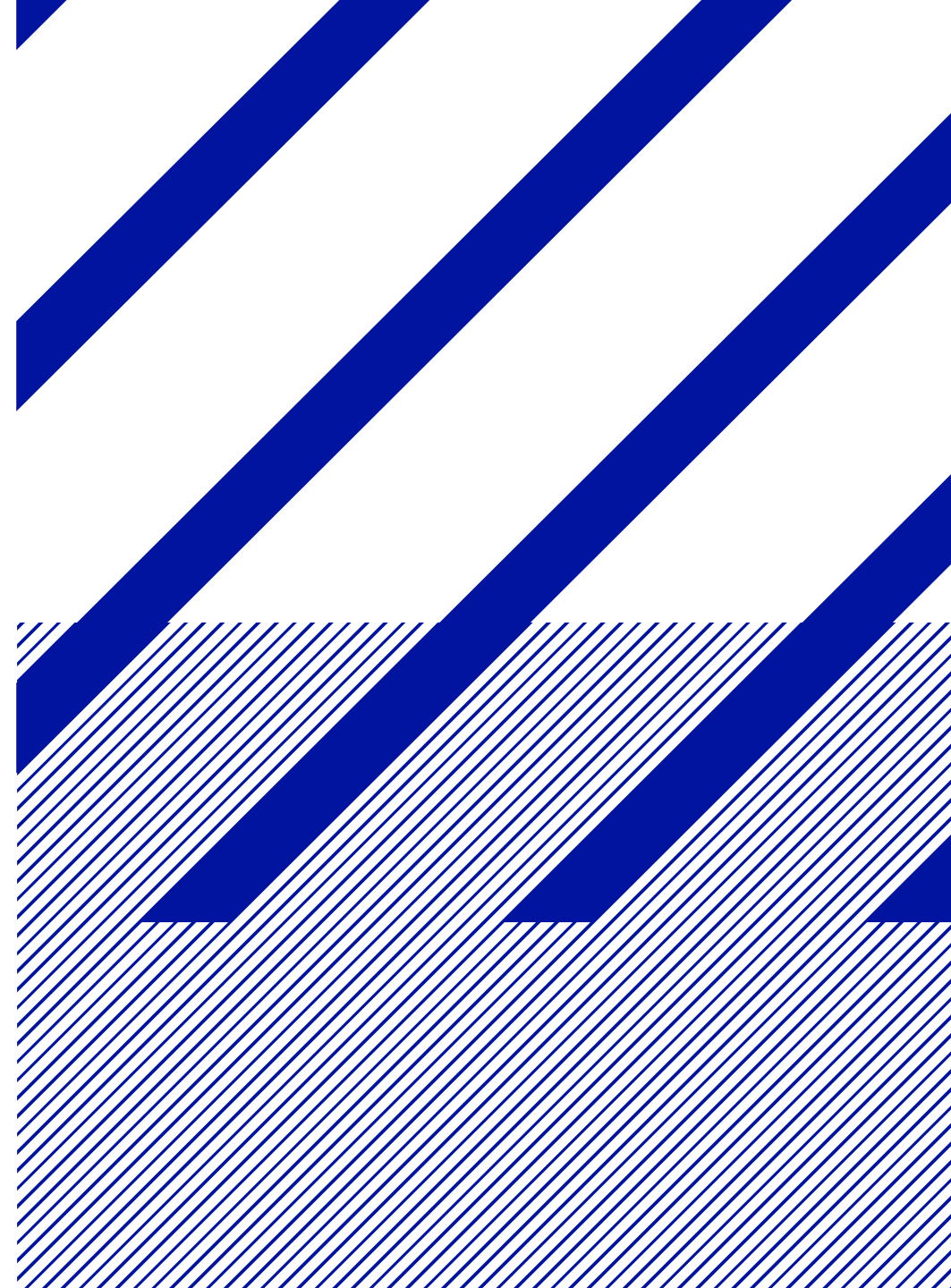
Prof. Dr. Michael Bücker

Professor für Data Science

Corrensstraße 25
D-48149 Münster

fon +49 (0)251.83 65-615
fax +49 (0)251.83 65-502

michael.buecker@fh-muenster.de
www.fh-muenster.de



Kurzvorstellung

Ihr Referent heute



FH MÜNSTER
University of Applied Sciences

Prof. Dr. Michael Bücker

*Professor für Data Science, Mathematik und
Wirtschaftsinformatik*

- Seit 03/2018: Professor an der FH Münster
- 05/2011 – 02/2018: Expert und Engagement Manager für Marketing Analytics und Data Science bei McKinsey & Company, Inc.
- 04/2011: Promotion in Statistik an der TU Dortmund
- 06/2008: Diplom in Statistik an der TU Dortmund



Datengestützte Entscheidungsprozesse

Daten als Grundlage – nicht als Ersatz für Expertise

In God we trust.

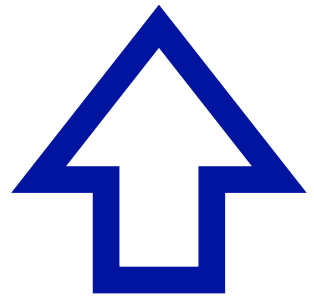
All others
must bring data.

William Edwards Deming



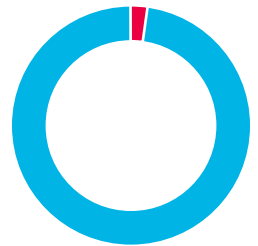
Datenmenge wächst stetig an

Bislang noch geringe Nutzung von Daten



2000%

Zuwachs gespeicherter Daten
zwischen 2013 und 2020



99,5%

Der Daten werden nie
analysiert

Sehr unterschiedliche Datenquellen

Kategorisierung von Datenquellen

- Erste Anlaufstelle: **interne strukturierte Daten** (z. B. CRM, Supply Chain, Produktion, Vertrieb, HR, ...)
- Etwas schwieriger zu verwerten: **Unstrukturierte Daten**, wie Texte/Dokumente, Videos, Bilder (z.B. Lieferscheine, Rechnungen, Emails, Sensoren, ...)
- Können beschafft werden: **Externe Daten** (z.B. demographische Daten, Reports, ...)
- Frei verfügbar: **Öffentliche Daten** (z.B. Social Media, Kartendienste, Medien, ...)



Verwendung von Daten in Unternehmen

Dashboards, Management Präsentationen



- Häufigste Verwendung: **Dashboards**, Präsentationen
- Erleichtert Interpretation und Verständnis der Daten; allerdings **schwierig komplexe Muster zu erkennen**
- Erkenntnisgewinn erfordert häufig gutes Verständnis von Daten und Zusammenhängen („**Data literacy**“)
- Keine direkten Handlungsempfehlungen („**Call to action**“)

Missbrauch von Daten

Sinnvoller Umgang mit Daten essentiell

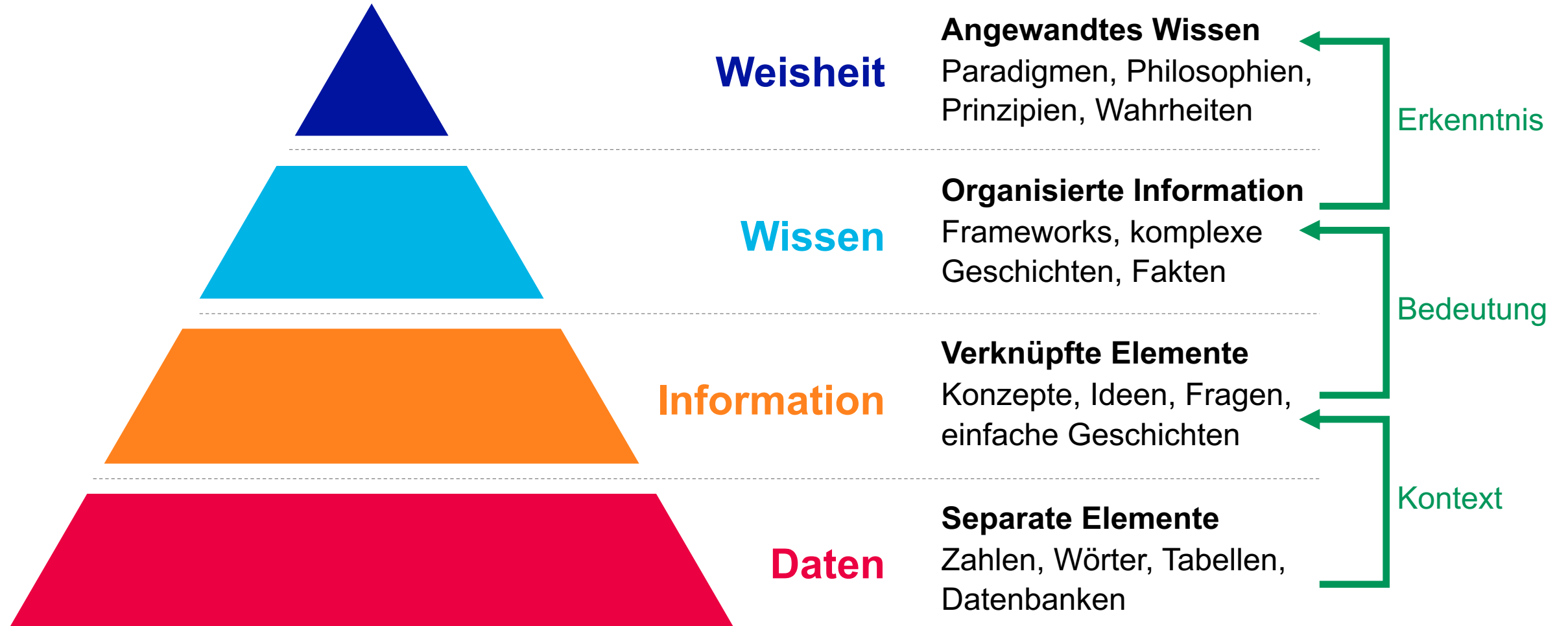


Torture the data
and it will confess
to anything.

Ronald Coase

Wertschöpfung aus Daten

Von Rohdaten zur Weisheit



Kontext, Bedeutung, Erkenntnis

Industrieexpertise unerlässlich

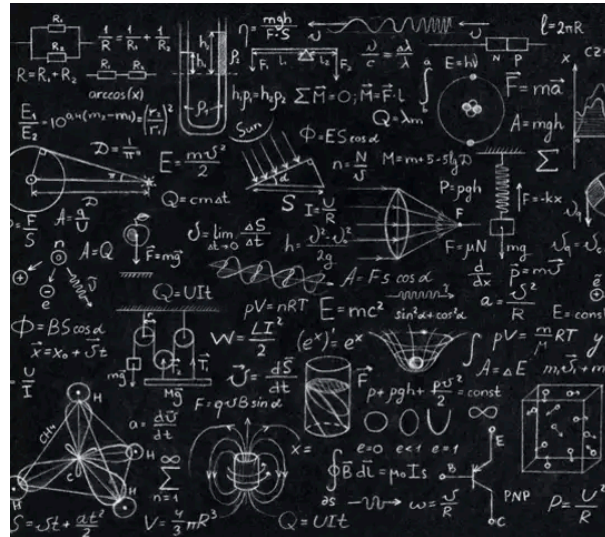
Daten



Analysen



Fachwissen

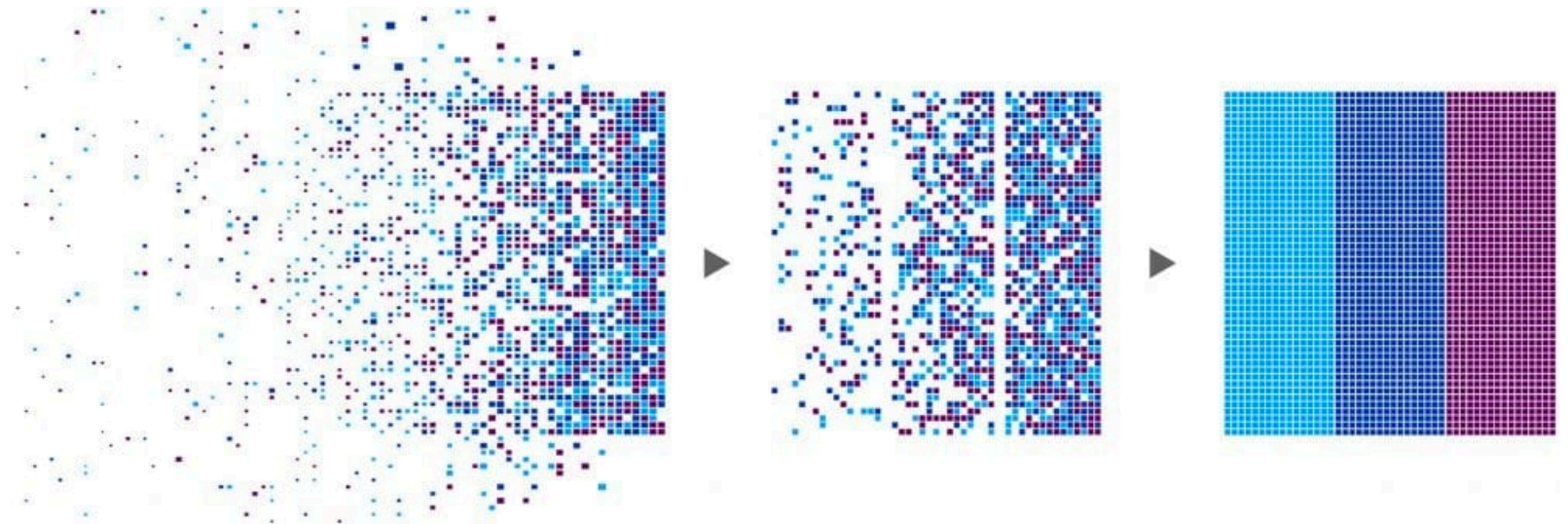


Strukturen in Daten entdecken

Statistik, Machine Learning, Algorithmen

- Mit Hilfe von **Statistik und mathematischen Algorithmen** lassen sich komplexe Strukturen in Daten entdecken, welche neue Erkenntnisse liefern können
- Fortschritte in Rechenleistung von Computern, neuartige Verfahren und die große Menge verfügbarer Daten eröffnen **immer neue Anwendungsfelder**
- Methoden der **künstlichen Intelligenz** lösen heute schon Probleme, die lange für unmöglich gehalten wurden

Wie aber lernen
Computer neue
Dinge mit Hilfe
von Daten?



Lernende Computer: Machine Learning

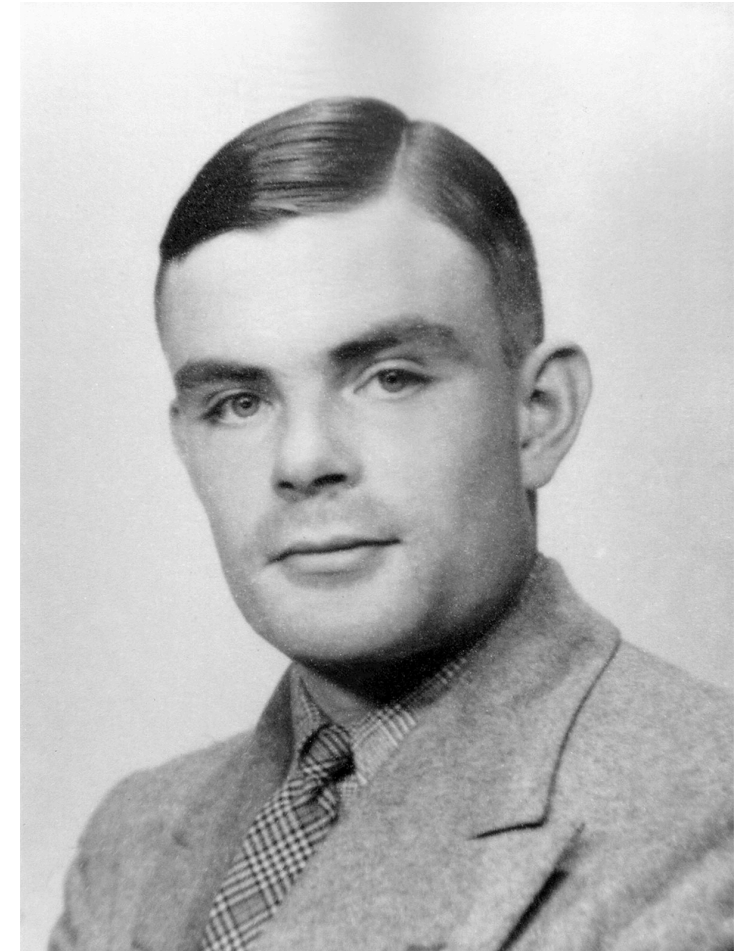
Lernen aus Erfahrung



FH MÜNSTER
University of Applied Sciences

What we want is
a machine that
can learn from
experience.

Alan Turing



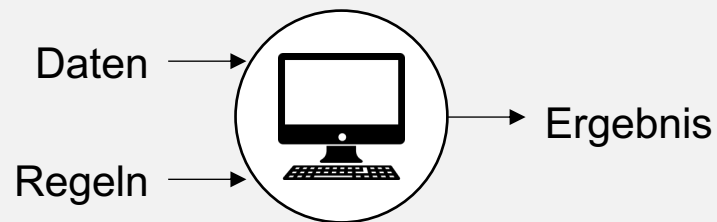
Lernende Computer: Machine Learning

Abgrenzung zu traditionellen Computerprozeduren

Traditionelles Computerprogramm

Definiere explizite Regeln, die der Computer befolgt und abarbeitet

```
if email contains "V!agra"  
  then mark is-spam;  
if email contains ...  
  then mark is-spam;
```

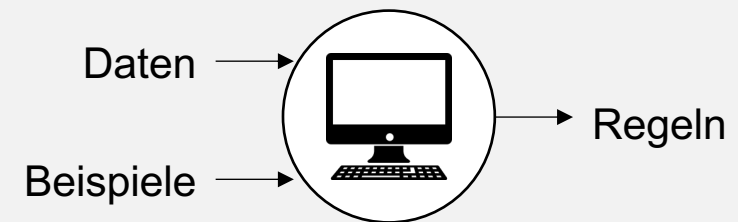


Erkläre mit Instruktionen

Machine Learning Programm

Computerprogramm, das anhand von Beispielen Regeln erlernt

```
try to classify some emails;  
change self to reduce errors;  
repeat;
```



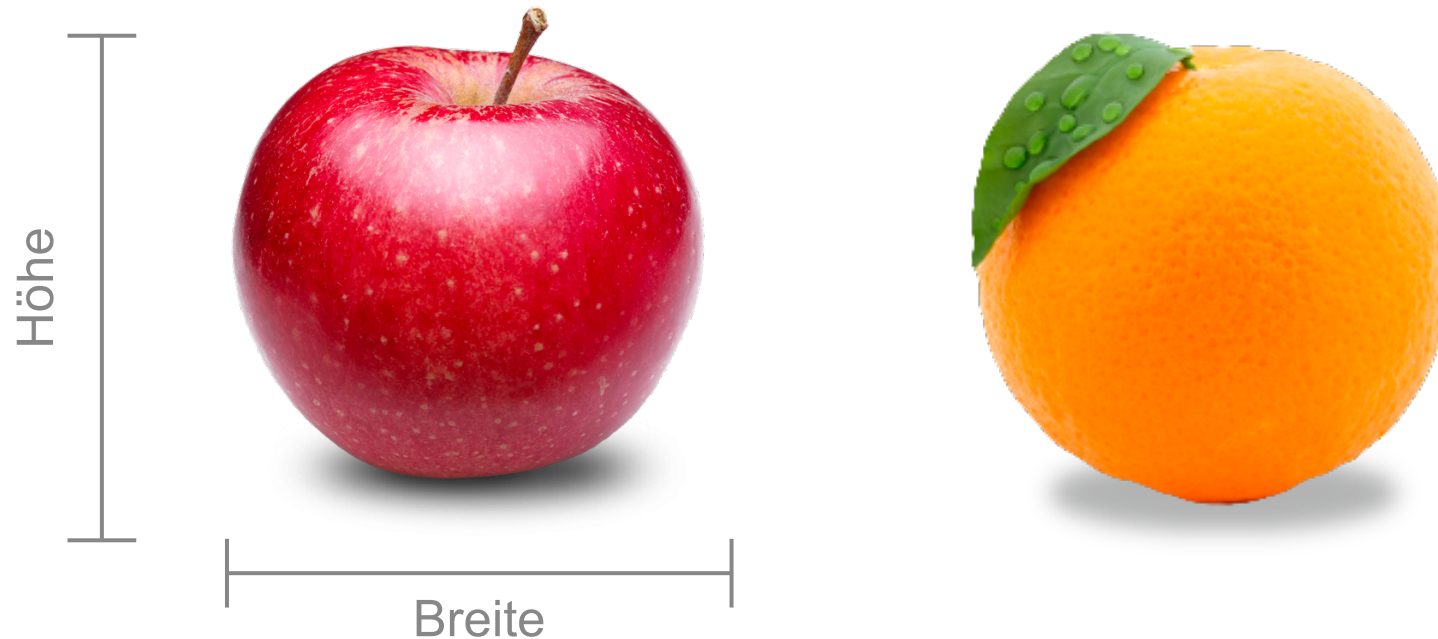
Erkläre mit Beispielen

Funktionsweise lernender Algorithmen

Entscheidungsbäume zur Klassifikation von Obst

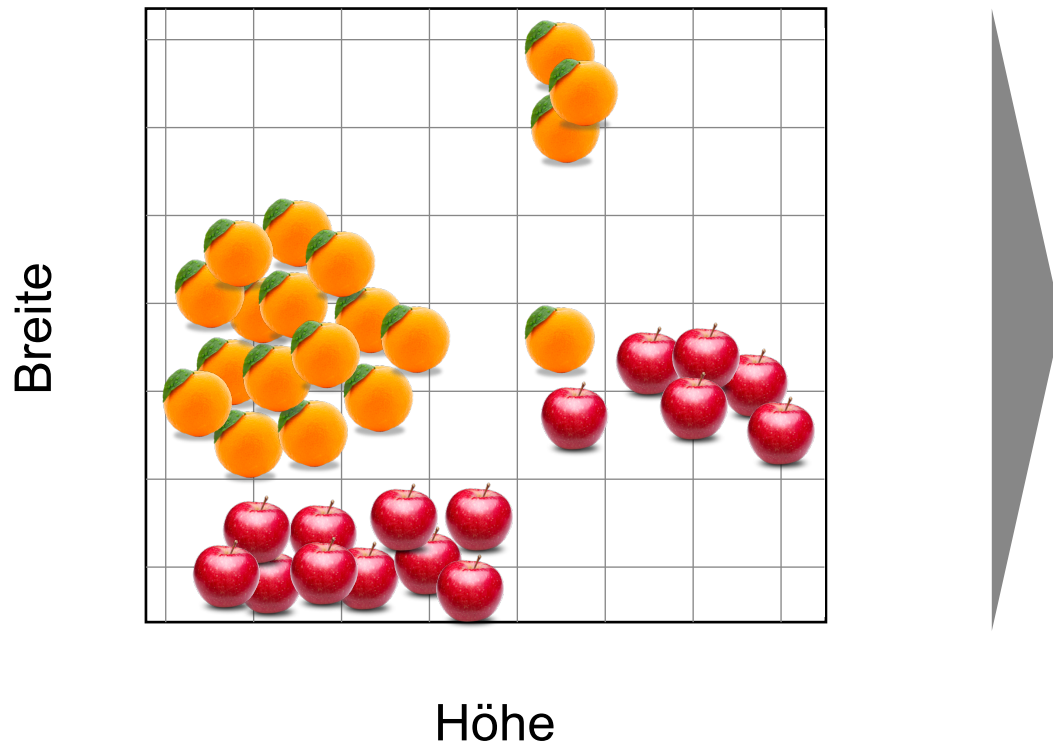


Jetzt: Wir wollen einem Computer beibringen, Äpfel von Orangen anhand der Breite und Höhe zu unterscheiden



Funktionsweise lernender Algorithmen

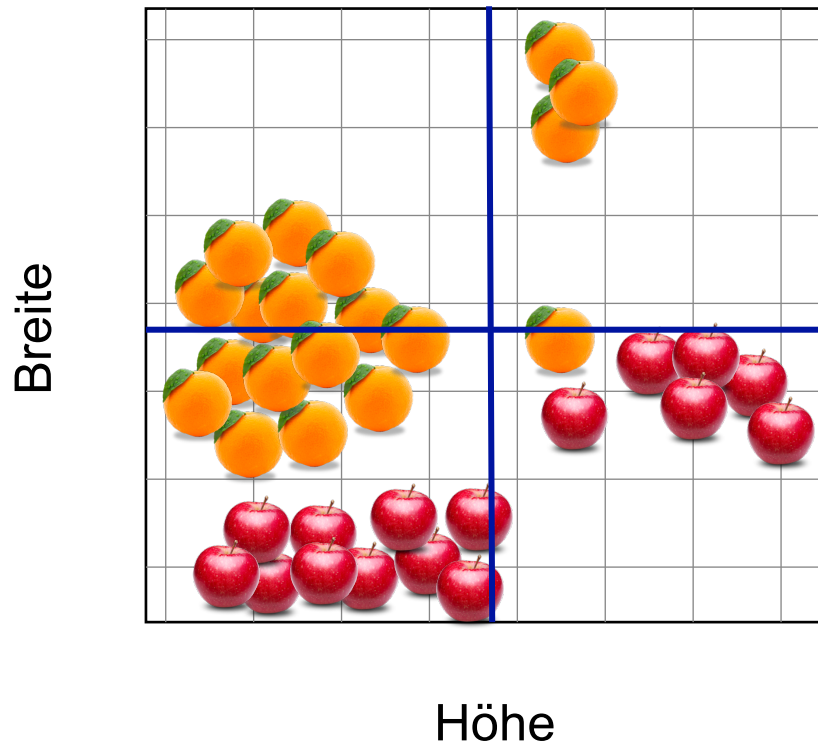
Entscheidungsbäume zur Klassifikation von Obst



- Wir sammeln **Beispiele**: eine Stichprobe von Orangen und Äpfeln
- Wir erfassen die **relevanten Daten**: Höhe und Breite jeder Frucht
- Wir kennen die **Zuordnung** der Früchte (Apfel, Orange)
- Wir nutzen die Beispiele um mit Hilfe eines Algorithmus **Regeln zu lernen**, mit denen sich die Früchte automatisch unterscheiden lassen
- Der Algorithmus wird so trainiert, dass möglichst **wenige Fehlzusordnungen** auftreten

Funktionsweise lernender Algorithmen

Entscheidungsbäume zur Klassifikation von Obst

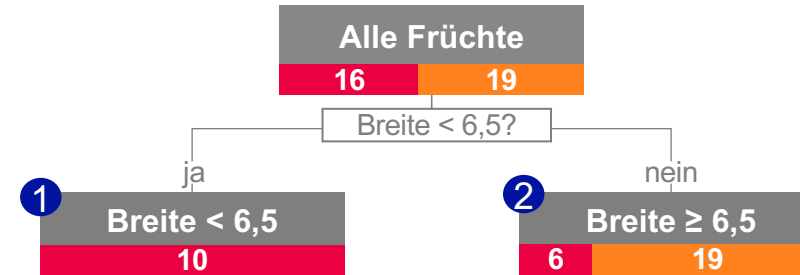
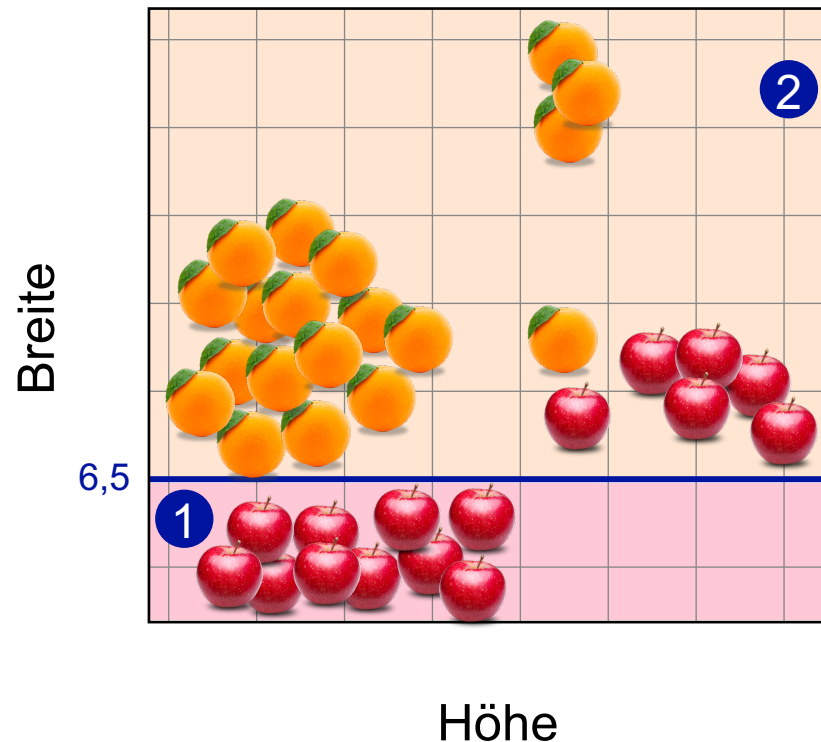


- Der Algorithmus prüft nun entlang jeder Dimension, wie sich **Äpfel und Orangen möglichst gut unterscheiden lassen**
- Eine gute Trennung gelingt,
 - wenn die aufgetrennten Gruppen **möglichst sortenrein** sind
 - wenn die aufgetrennten Gruppen **möglichst groß** sind
- Die beste Trennung wird als erste **Regel** festgehalten

Funktionsweise lernender Algorithmen

Entscheidungsbäume zur Klassifikation von Obst

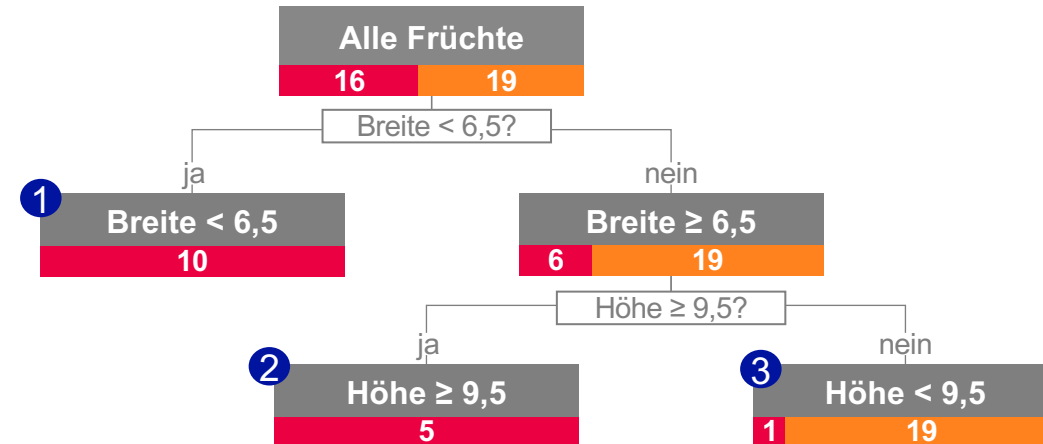
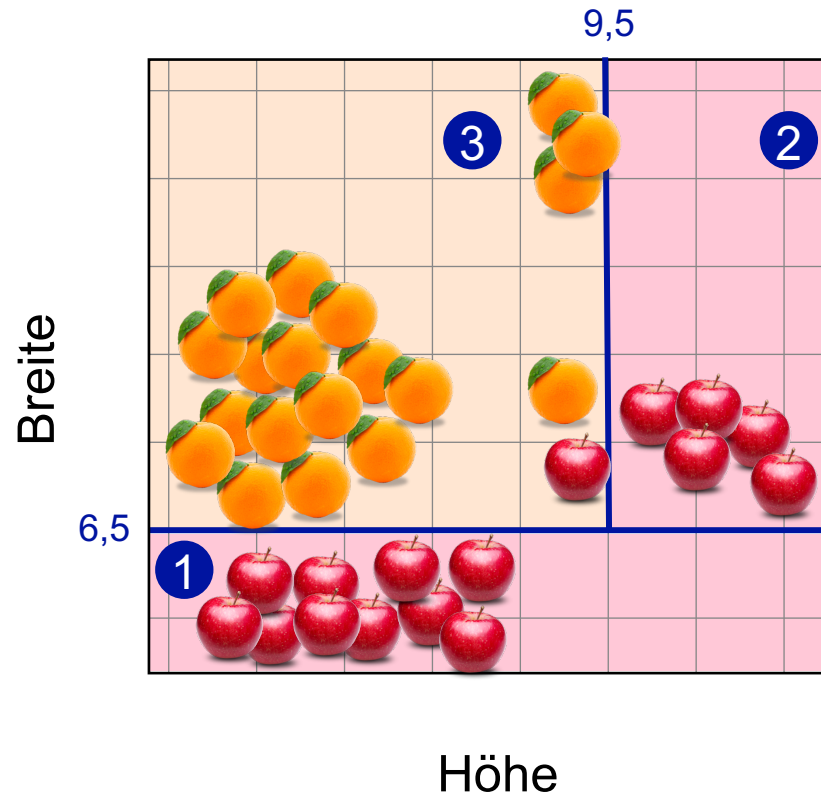
Ein Entscheidungsbaum lernt eine Folge von Entscheidung zur Vermeidung von Fehleinordnungen



Funktionsweise lernender Algorithmen

Entscheidungsbäume zur Klassifikation von Obst

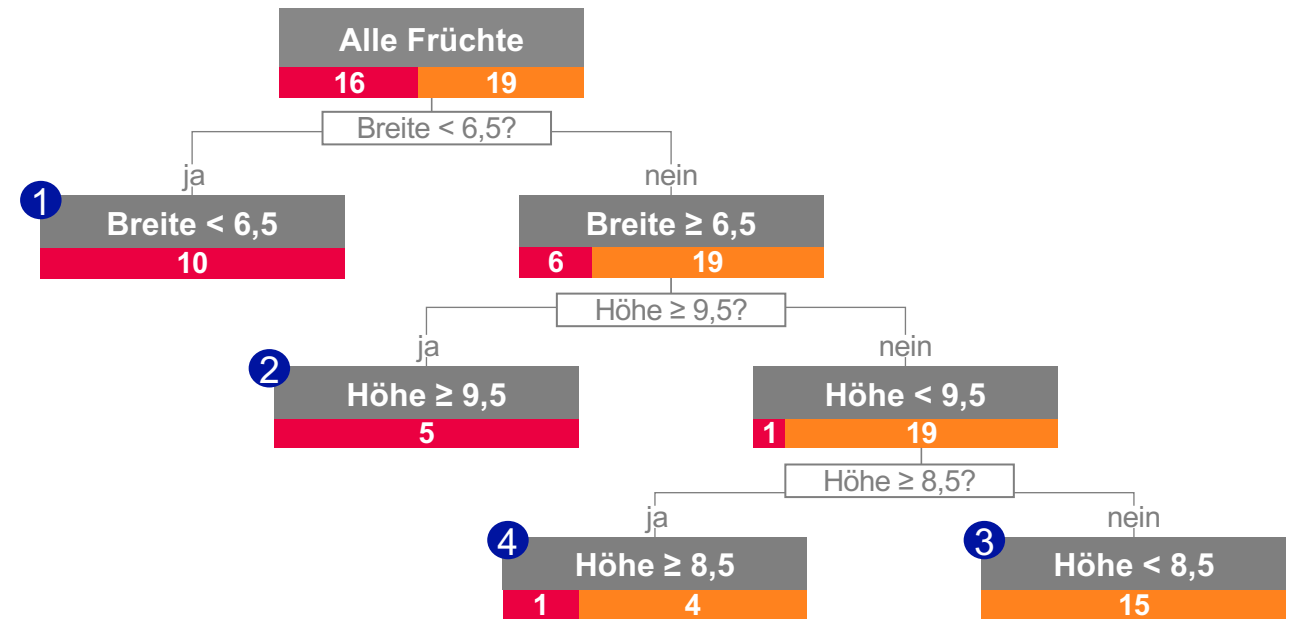
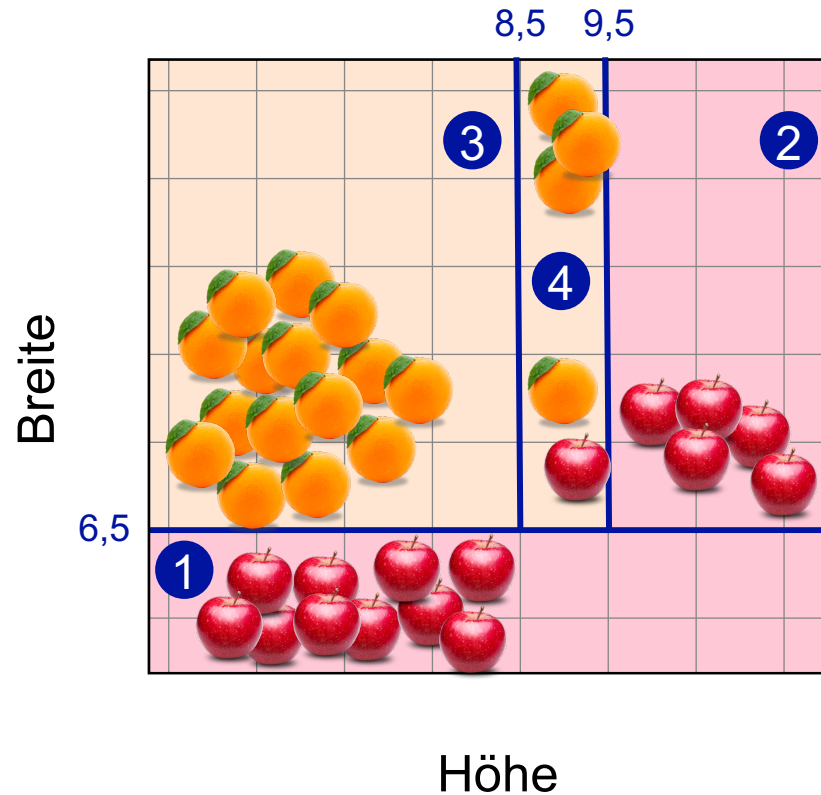
Ein Entscheidungsbaum lernt eine Folge von Entscheidung zur Vermeidung von Fehleinordnungen



Funktionsweise lernender Algorithmen

Entscheidungsbäume zur Klassifikation von Obst

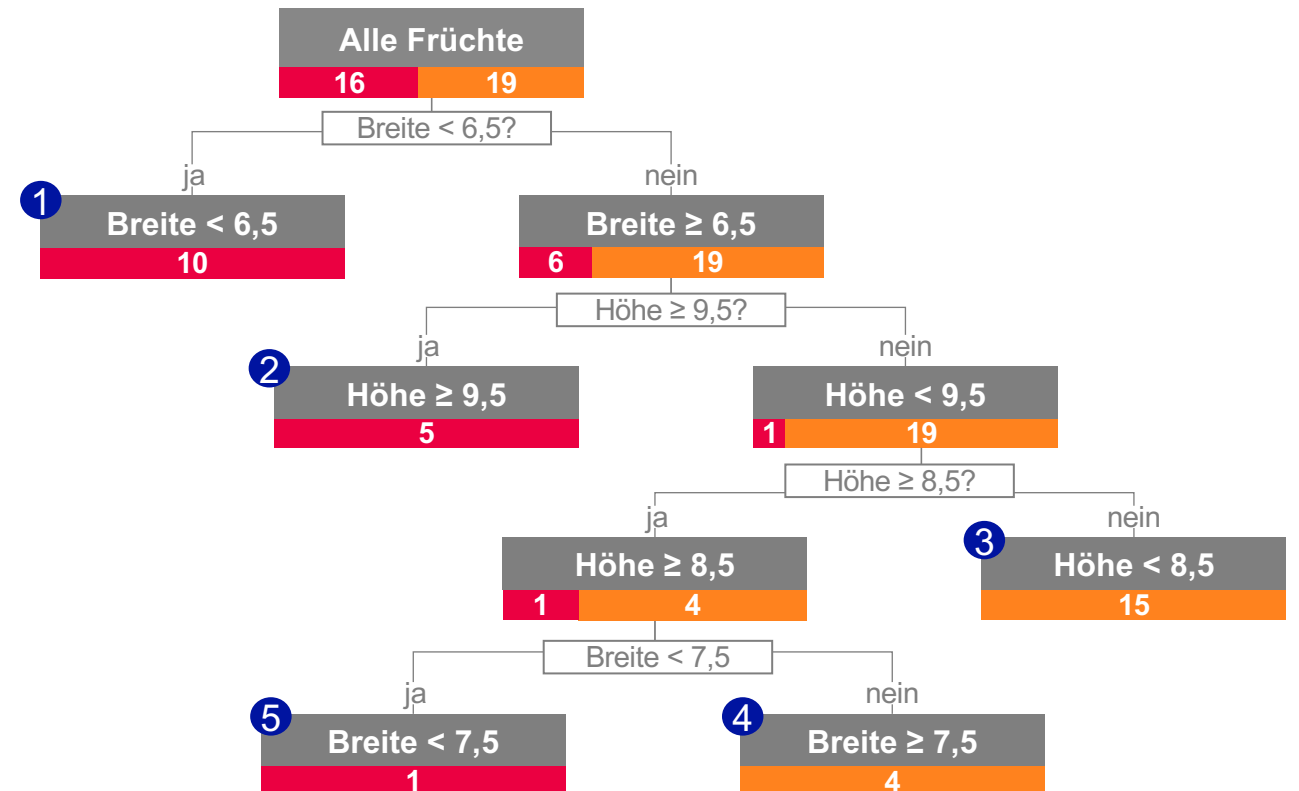
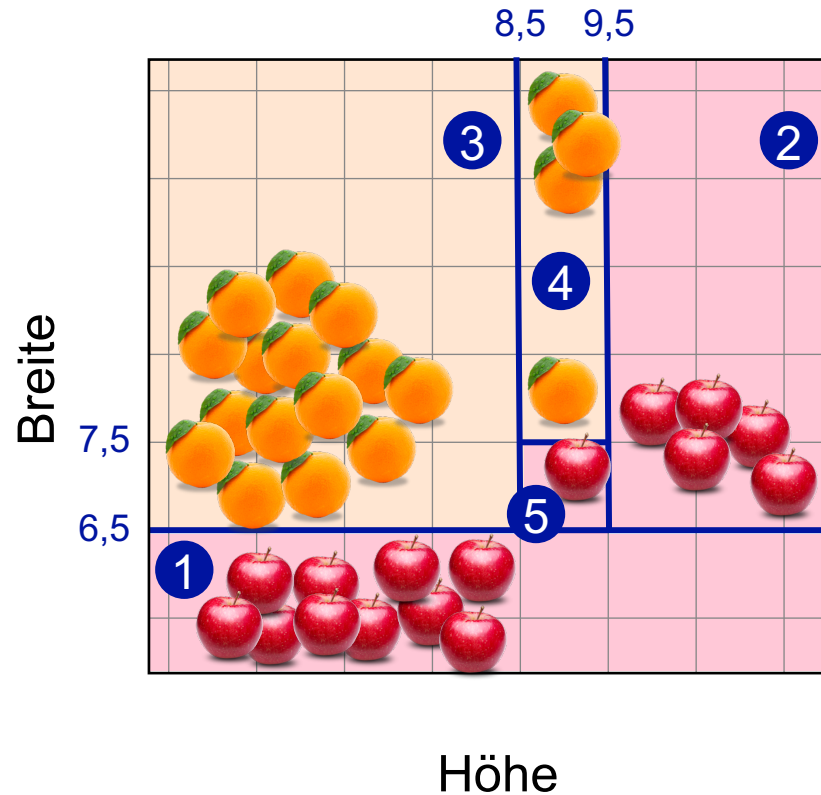
Ein Entscheidungsbaum lernt eine Folge von Entscheidung zur Vermeidung von Fehleinordnungen



Funktionsweise lernender Algorithmen

Entscheidungsbäume zur Klassifikation von Obst

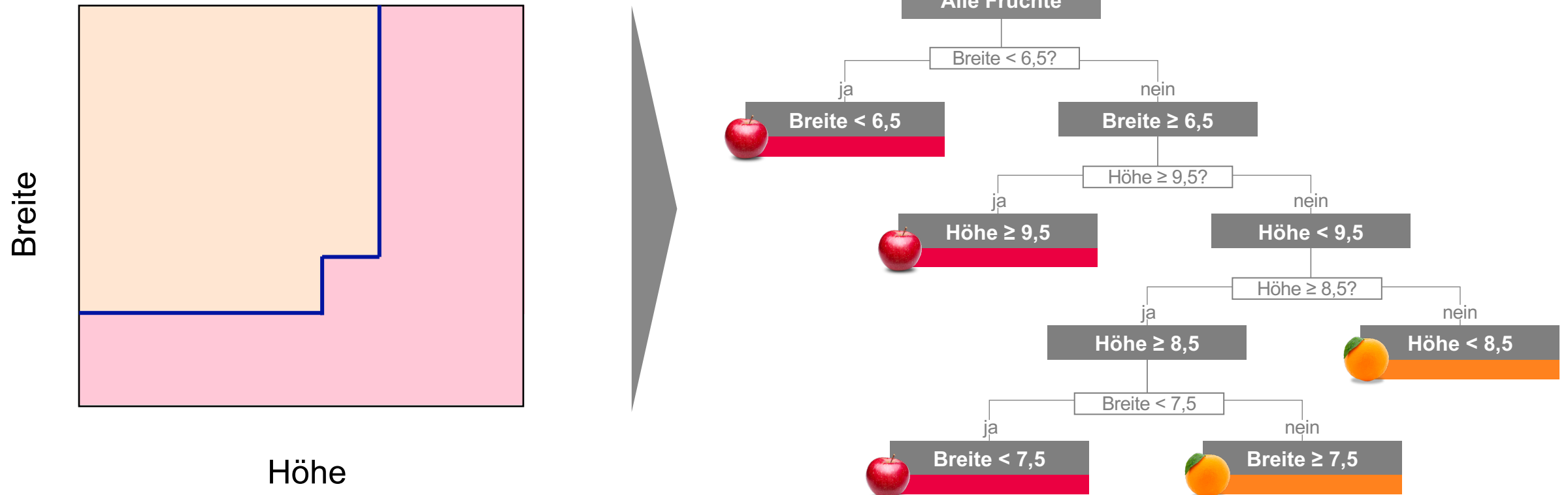
Ein Entscheidungsbaum lernt eine Folge von Entscheidung zur Vermeidung von Fehleinordnungen



Funktionsweise lernender Algorithmen

Entscheidungsbäume zur Klassifikation von Obst

Ein Entscheidungsbaum lernt eine Folge von Entscheidung zur Vermeidung von Fehleinordnungen



„Hot Topics“ für Unternehmen

Anwendungen im betriebswirtschaftlichem Kontext

Yield Optimization



Feineinstellung von **(Produktions-) Prozessen** zur Maximierung des Ertrags und der Qualität der Endprodukte

Personalisierung



Personalisierung von **Kundeninteraktionen** unter Berücksichtigung der individuellen Kundenbedürfnisse

Dynamic Pricing



Preisbildung in Echtzeit basierend auf Wettbewerb, Wertwahrnehmung, Kundenpräferenzen und -verhalten

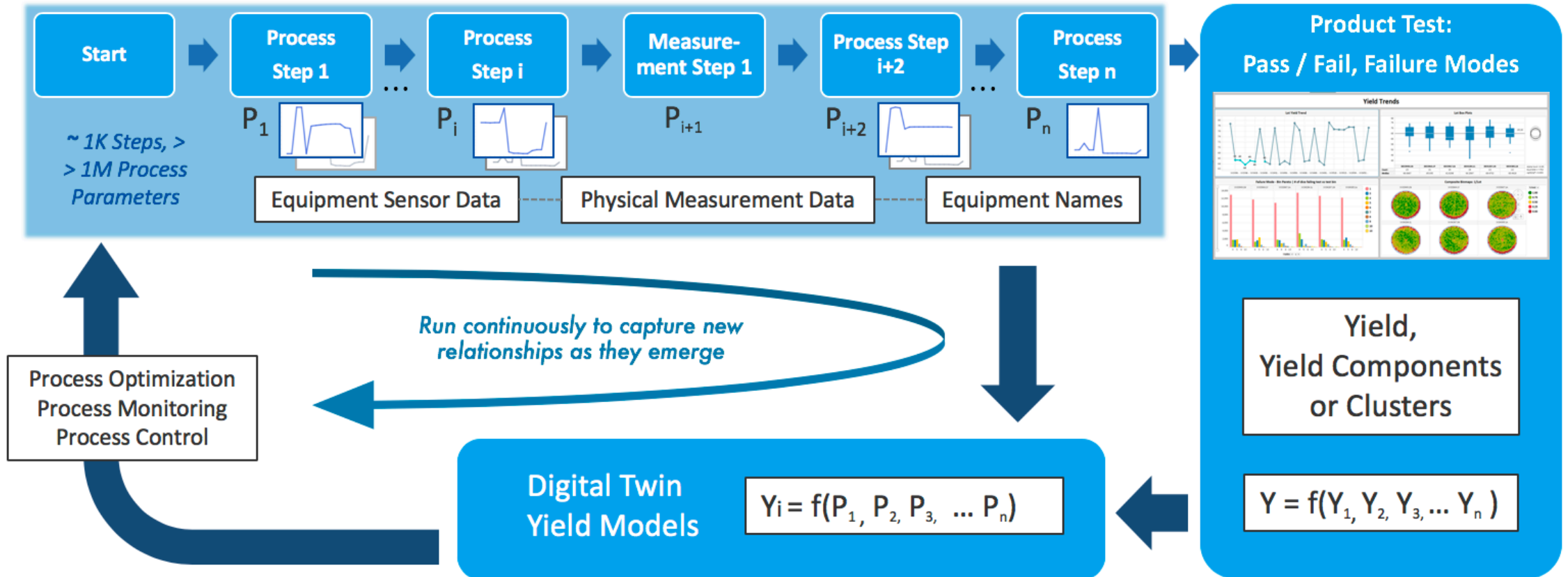
Predictive Maintenance



Vorhersage von **Störungen von Maschinen** bevor sie entstehen, um Kosten durch Ausfälle zu vermeiden

Optimierung von Produktionsprozessen

Beispiel aus der Halbleiter-Industrie



Missbrauch von Daten

Sinnvoller Umgang mit Daten essentiell



Ideas are easy.
Implementation is
hard.

Guy Kawasaki

Ressourcen für Data Science

Experten sind die knappe Ressource

Ressource	Beschreibung	
 Algorithmen	Algorithmen sind frei verfügbar; für die meisten existieren in Form von OpenSource Implementierungen	
 IT Infrastruktur	Infrastruktur zwar teuer, aber durch Cloud-Anbieter (AWS, GCP, Azure) schnell verfügbar und skalierbar	
 Software	Alle wesentlichen Softwaretools (Datenbanken, Datenverarbeitung, Machine Learning) sind OpenSource	
 Daten	Durch die digitale Transformation von Unternehmen sind zunehmend Daten von hoher Qualität verfügbar	
 Expertise	Experten sind sehr schwer zu finden; der Bedarf ist bei weitem nicht gedeckt und wird weiter wachsen ¹	

¹ The Quant Crunch: How the demand for data science skills is disrupting the job market; Research Paper von Burning Glass Technologies, BHEF, und IBM; <https://ibm.co/2LoV8li>

Identifizierung sinnvoller Anwendungen

Vorüberlegungen vor der Umsetzung

Problem	Anwendung	Nutzen	Machbarkeit
Beschreibung der Problemstellung	Wie lässt sich mit dem Ergebnis konkret umsetzen?	- Was ist der konkrete Nutzen? - Was sind die zu erwartenden Kosten?	- Sind die nötigen Daten verfügbar? - Gibt es die benötigte Infrastruktur?

Beispiel Kündigungsverhinderung

Welche Kunden werden kündigen?

Vorhersage der Kunden, die in x Wochen kündigen

Kunden mit hohem Kündigungsrisiko wird ein Angebot gemacht

Nutzen: Rückgang der Kündigungen

Zu berücksichtigen: Kosten der Angebote

Daten sind im ERP vorhanden, IT Infrastruktur besteht oder kann beschafft werden



Ich freue mich auf Ihre Fragen und eine gute Diskussion!

Prof. Dr. Michael Bücker

Professor für Data Science

Corrensstraße 25
D-48149 Münster

fon +49 (0)251.83 65-615
fax +49 (0)251.83 65-502

michael.buecker@fh-muenster.de
www.fh-muenster.de

