

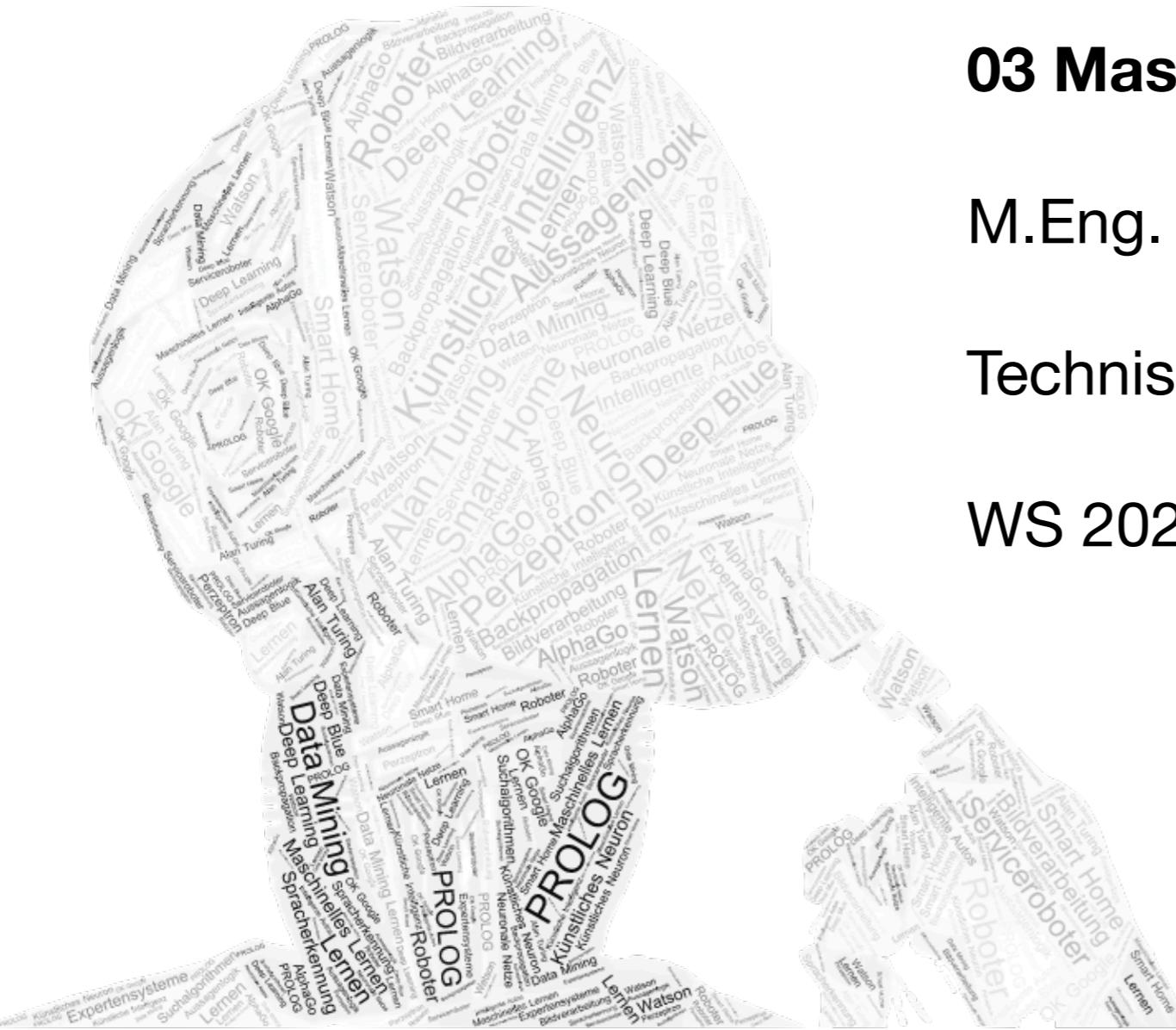
Künstliche Intelligenz

03 Maschinelles Lernen

M.Eng. Janine Breßler

Technische Hochschule Wildau

WS 2025/2026



Maschinelles Lernen

Maschinelles Lernen?

Menschliches Lernen:

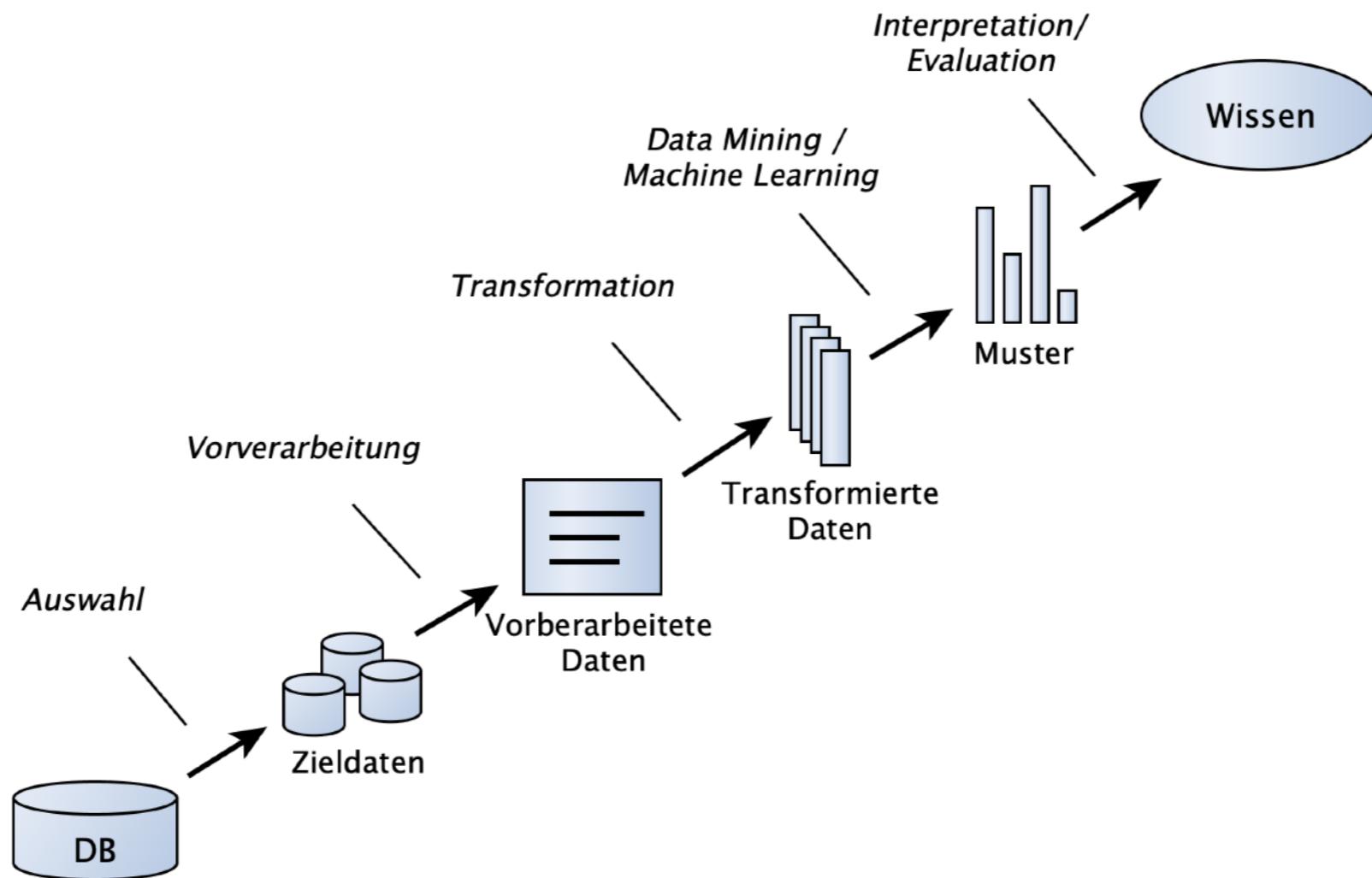
"Jede Verhaltensänderung, die sich auf Erfahrung, Übung oder Beobachtung zurückführen lässt."

Maschinelles Lernen, Frochte

Maschinelles Lernen:

- Die Maschine bzw. das Computerprogramm lernt aus Erfahrung und bewirkt damit eine Änderung des Verhaltens
- Statt einer statischen Programmierung erlernen die Computer anhand von Daten (Erfahrungen) ein Verhalten

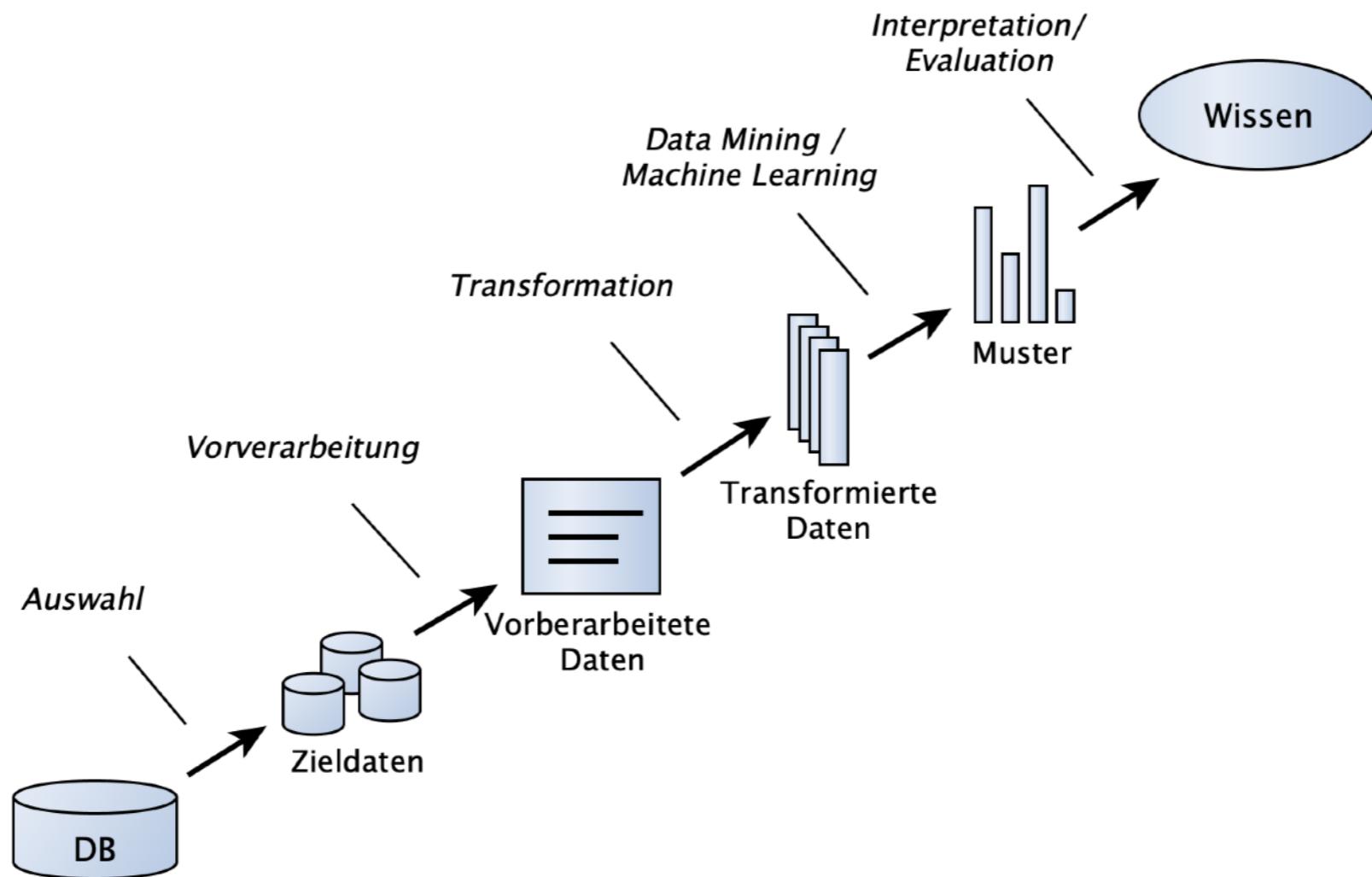
Knowledge Discovery in Databases



KDD als Prozess

- **Auswahl** von relevanten Daten für angestrebtes Ziel aus Datenbank
- **Vorverarbeitung** der Daten, indem Daten bereinigt werden, z.B. Werte wie NaN

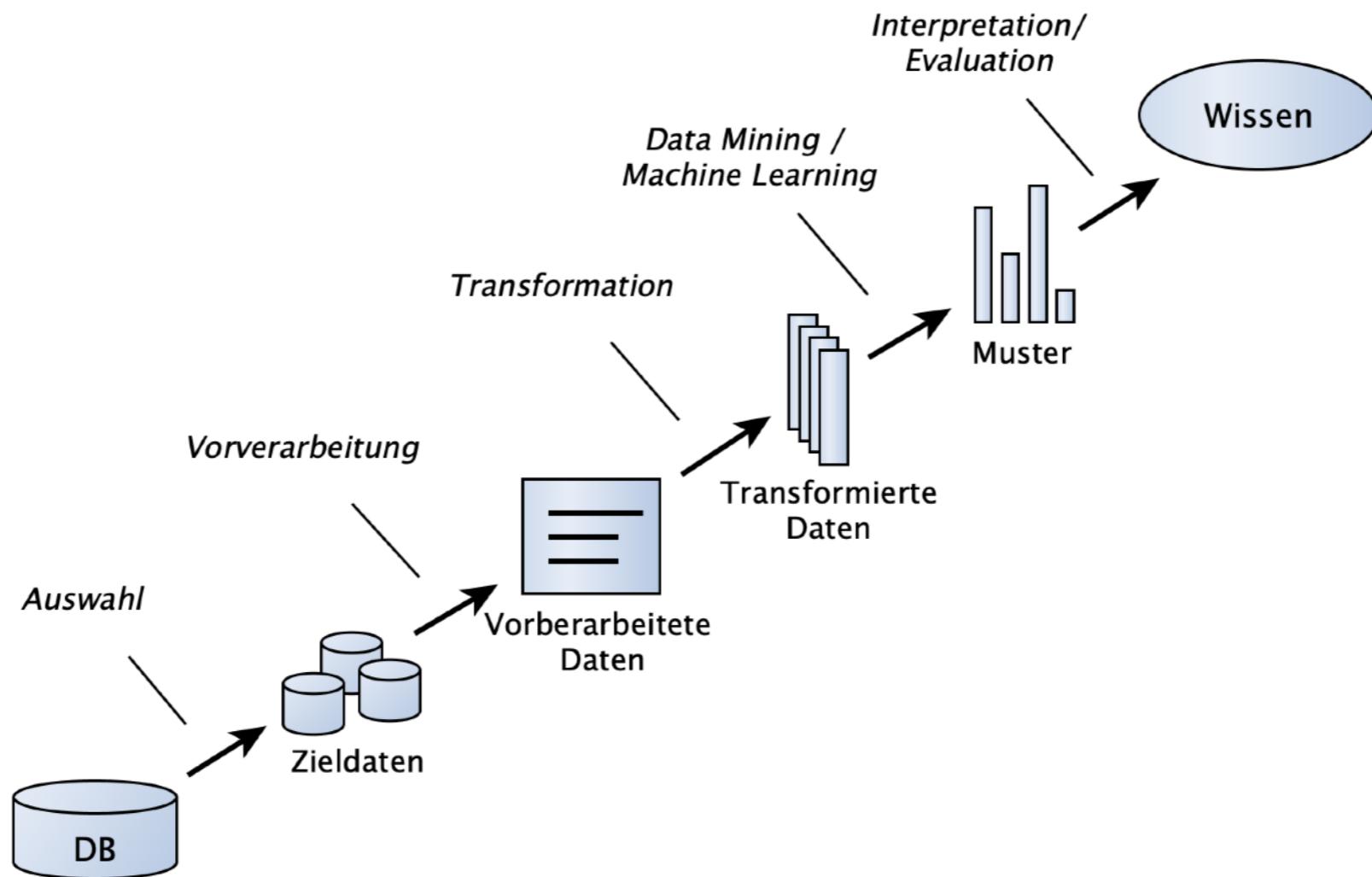
Knowledge Discovery in Databases



KDD als Prozess

- Transformation der Daten, um diese an den Algorithmus anzupassen
 - String-Daten werden in numerische Werte transformiert
 - Datenreduzierung:
 - Ausreißer-Datensätze werden entfernt
 - einzelne Merkmale werden zu einem Metamerkmal zusammengefasst
 - Normierung der Daten, bspw. auf Werte zwischen 0 und 1

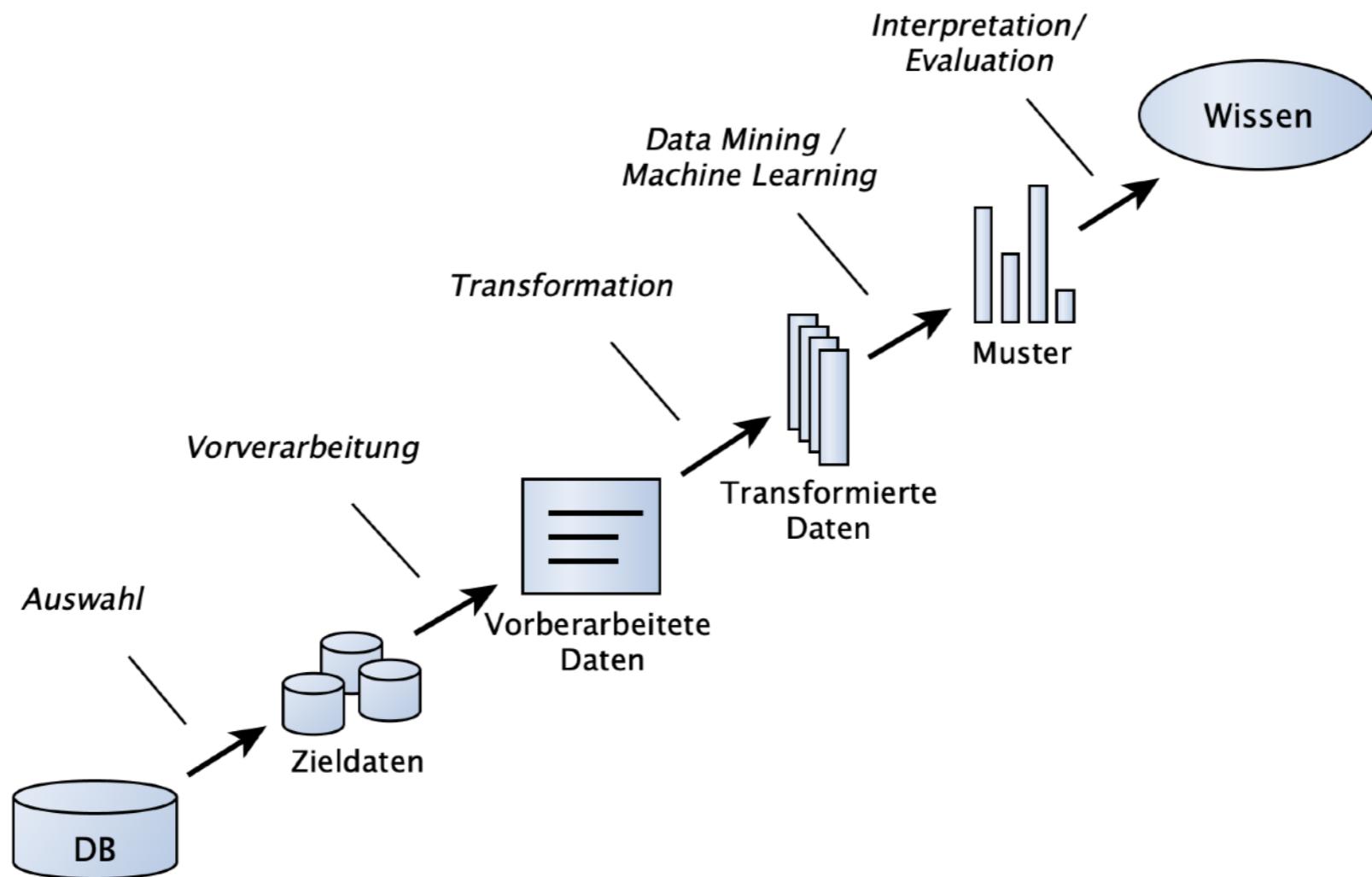
Knowledge Discovery in Databases



KDD als Prozess

- z.B. Machine Learning durch Auswahl eines geeigneten Werkzeugs durchführen
- Interpretation und Evaluation der Ergebnisse in den Daten, das Wissen
- Prozess kann iterativ durchgeführt werden, also Teilschritte wiederholt werden

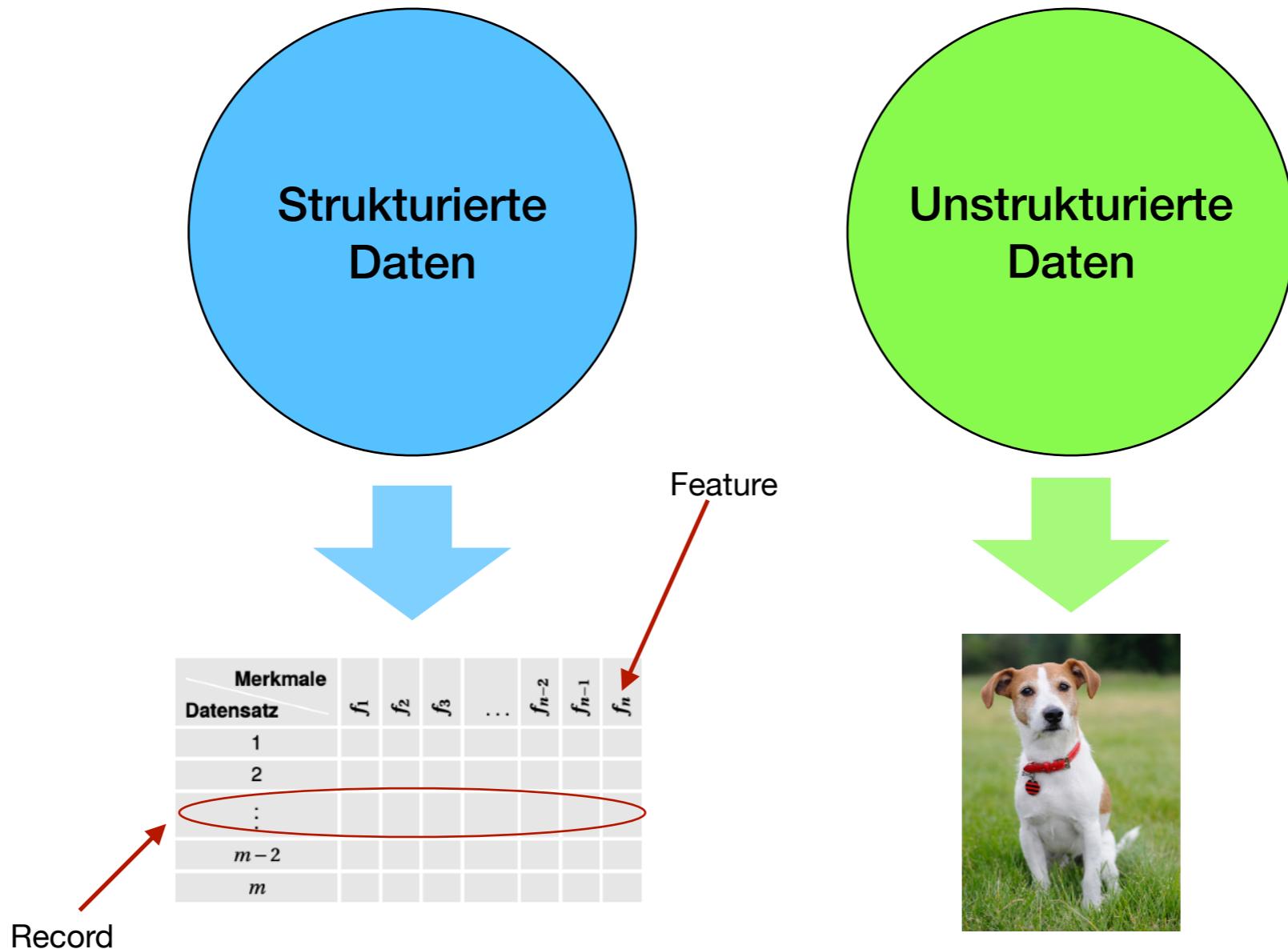
Knowledge Discovery in Databases



→ Maschinelles Lernen ist Teil des KDD-Prozesses

Daten, Daten und noch einmal Daten

Daten stellen Dreh- und Angelpunkt des Maschinellen Lernens dar



Strukturierte Daten sind in der Regel besser als Grundlage für das maschinelle Lernen geeignet!

Daten, Daten und noch einmal Daten

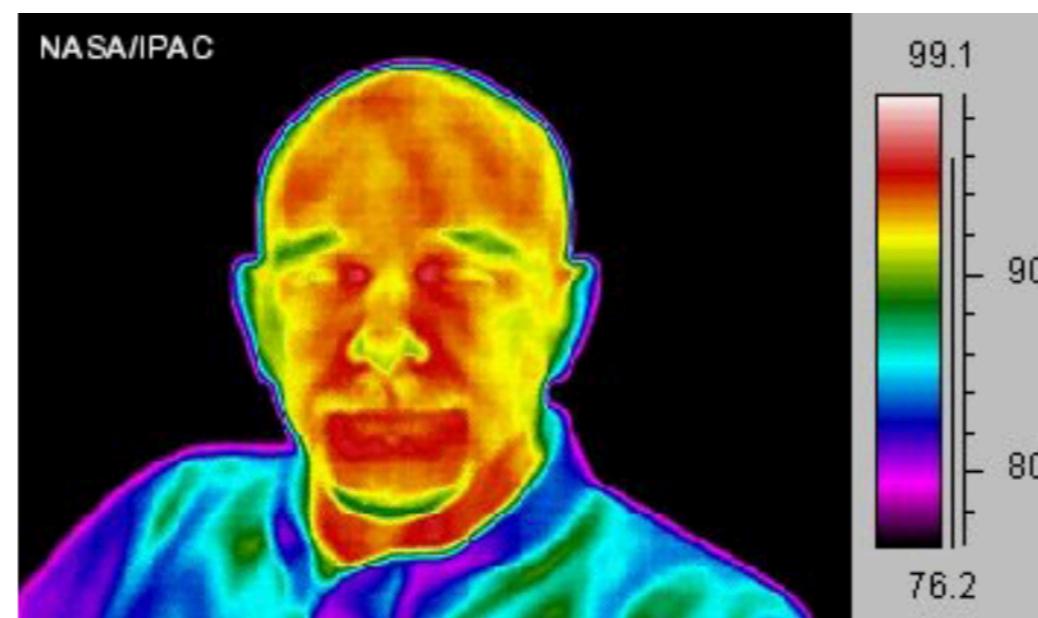
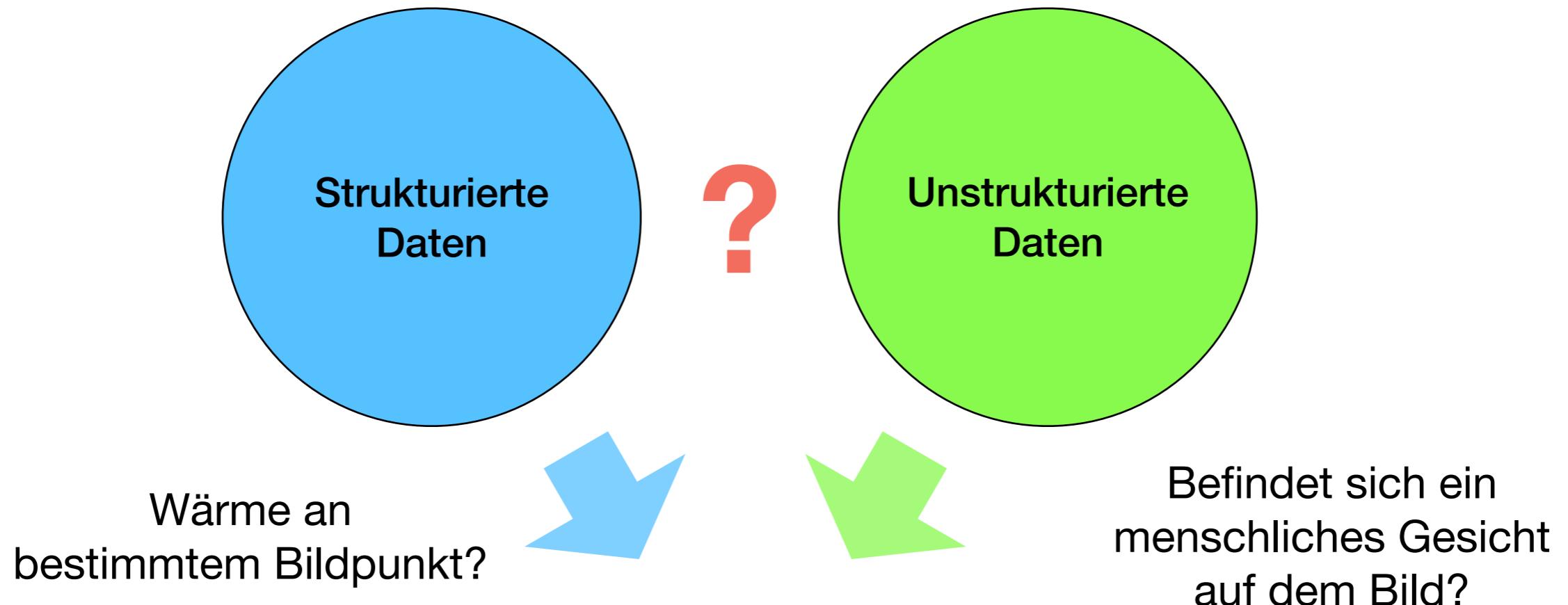


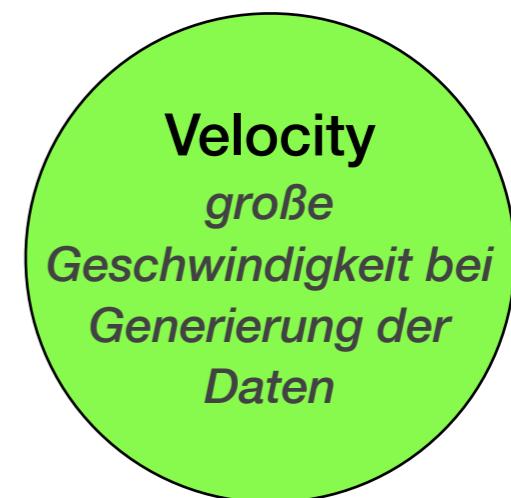
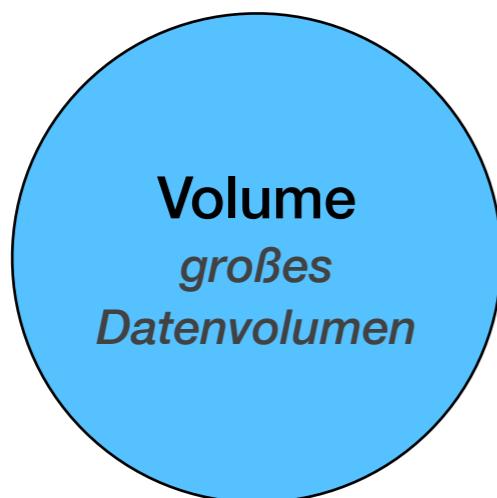
Bild von <https://spaceplace.nasa.gov/>

Daten, Daten und noch einmal Daten

Big Data

"[...] Datenbestände, die bzgl. ihrer Menge, Komplexität , schwachen Strukturierung und/oder Schnelllebigkeit ein Problem für die herkömmliche Datenverarbeitung bzw. Datenanalyse sind."

Maschinelles Lernen, Jörg Frochte

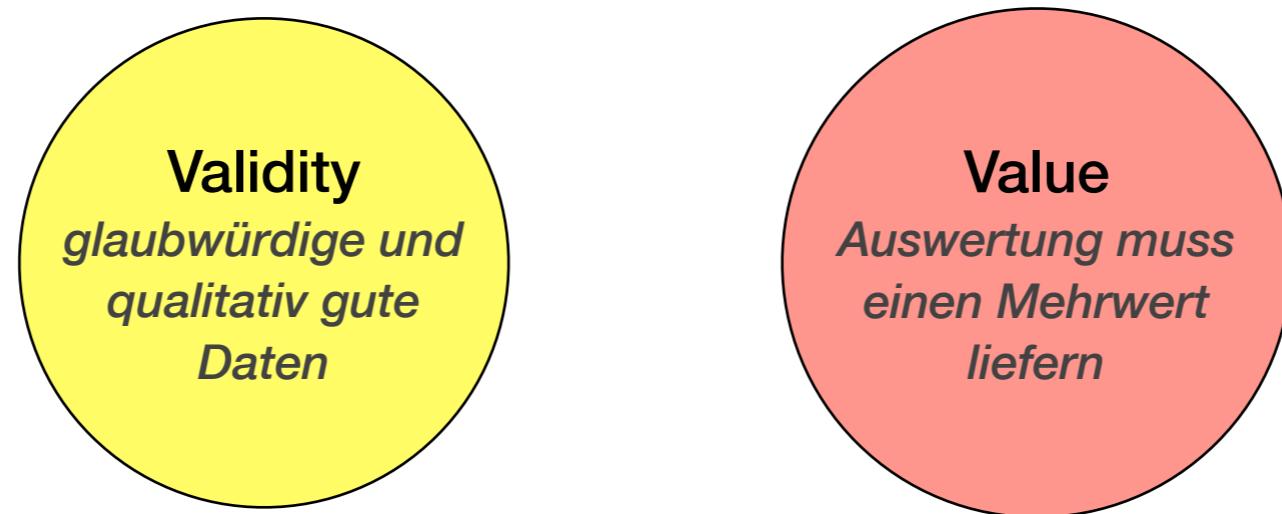


Ziel ist das Extrahieren von Informationen aus Daten,
um neues Wissen zu generieren.

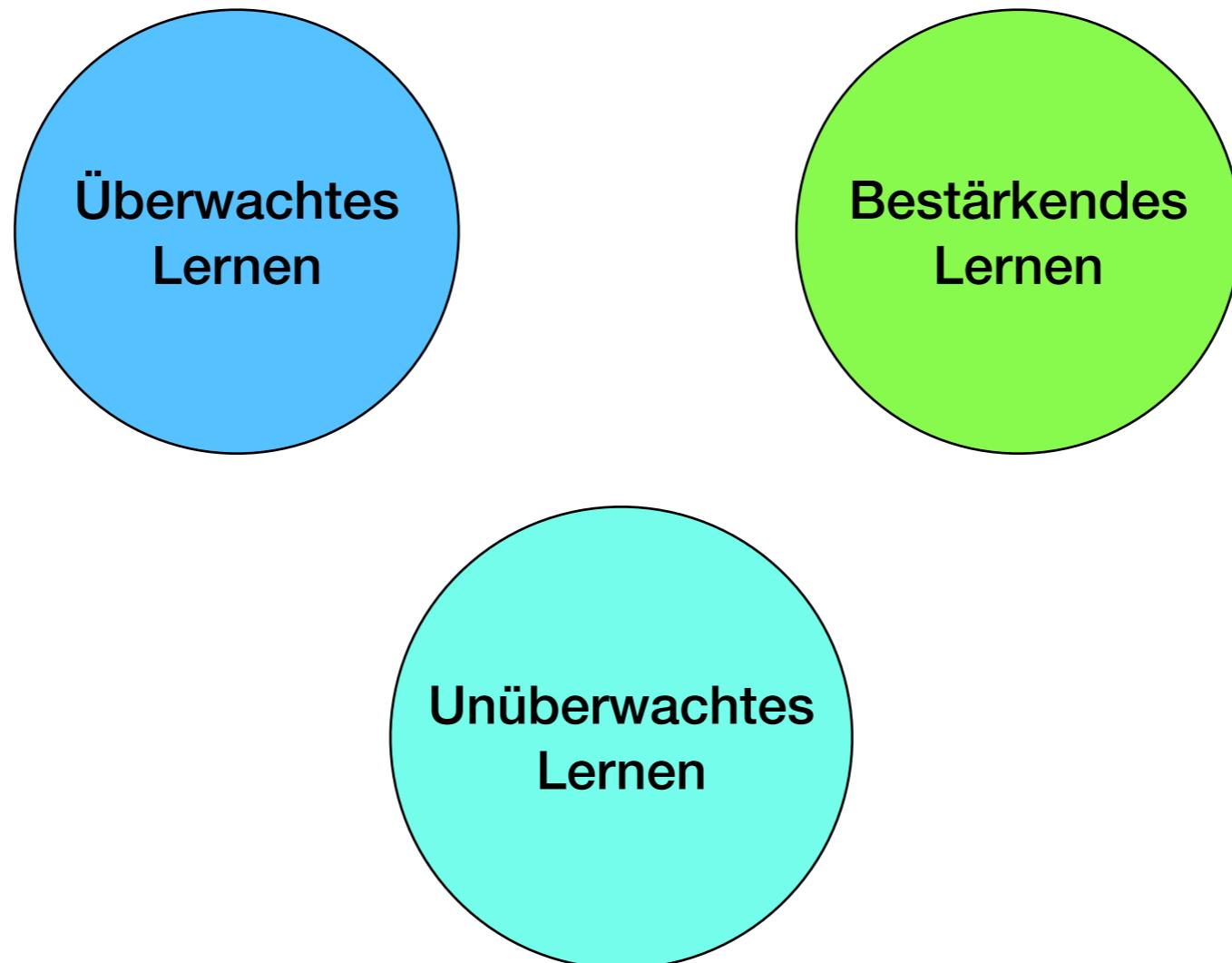
Daten, Daten und noch einmal Daten

Big Data

- Durch kombinierte und intelligente Analyse der Daten sollen Muster und Beziehungen erkannt werden
→ bessere Prognosen
- Intelligente Datenanalysen transformieren Big Data in Smart Data
- Anforderungen an die Daten:



Maschinelles Lernen



Alle Techniken haben gemein:

Konstruktion/Lernen einer mathematischen Funktion
 $f : X \rightarrow Y$

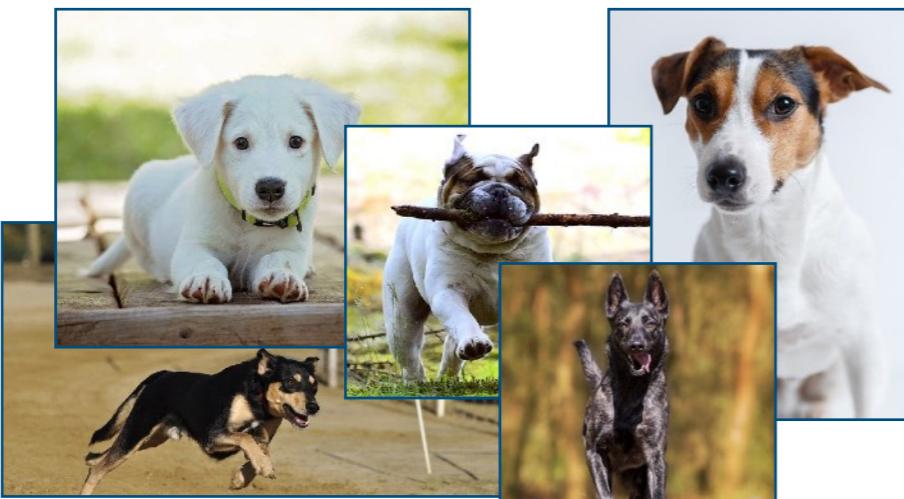


Quelle: PNGEgg

Überwachtes Lernen - Klassifikation



Große Menge an Ein- und Ausgabedaten (Datenpaare),
die über korrekten Funktionswert verfügen
→ gelabelte oder auch markierte Daten



Hund



Katze

Bilder von pixabay.com

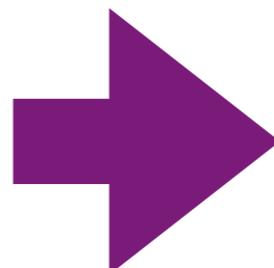
Überwachtes Lernen - Klassifikation

Computer wird mit Datensätzen trainiert

Anschließend soll Computer neue Bilder eigenständig den Labeln zuordnen (Klassifikation)

Die Zielmenge Y für das Ergebnis der Funktion ist diskret, die **Klassifikation** bildet auf Klassen (auch Label) mit nominalen Werten ab, hier: Hund oder Katze

Generalisierung von bekannten auf bisher unbekannte Datenpaare



Hund oder Katze?



Überwachtes Lernen - Regression

Regression erfolgt analog zur Klassifikation

Ziel ist das Erkennen von numerischen Zusammenhängen

Zielgröße ist im Gegensatz zur Klassifikation i.d.R. ein **kontinuierlicher Bereich**, mögliche Problemstellungen:

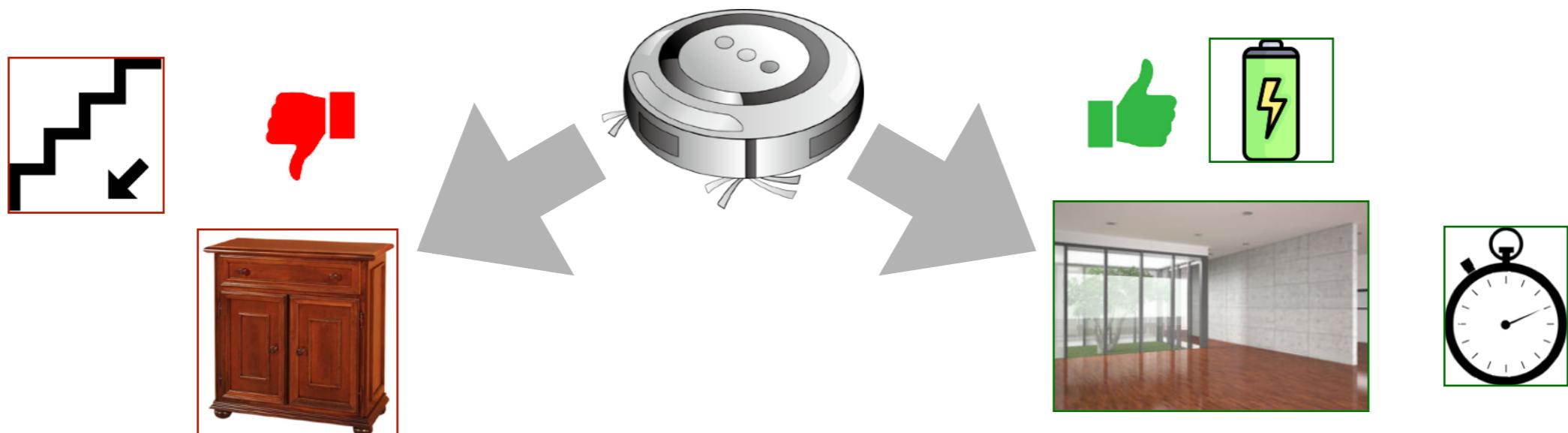
- Berechnung des optimalen Drehwinkels
- Berechnung des maximalen Kreditrahmens
- Prognose zur Anzahl der vermieteten Fahrräder



Bestärkendes Lernen

Ansatz interessant für Optimierungsprobleme

- optimale Strategie nicht bekannt
- Labeln von Daten nicht möglich, da nicht klar ist, welche **Abfolge von Teilschritten** genau richtig bzw. falsch ist
- bekannt ist jedoch, was einen wünschenswerten und nicht wünschenswerten Ausgang darstellt
- agentenbasiert, eine **optimale Strategie** wird anhand von Belohnung oder Bestrafung entwickelt



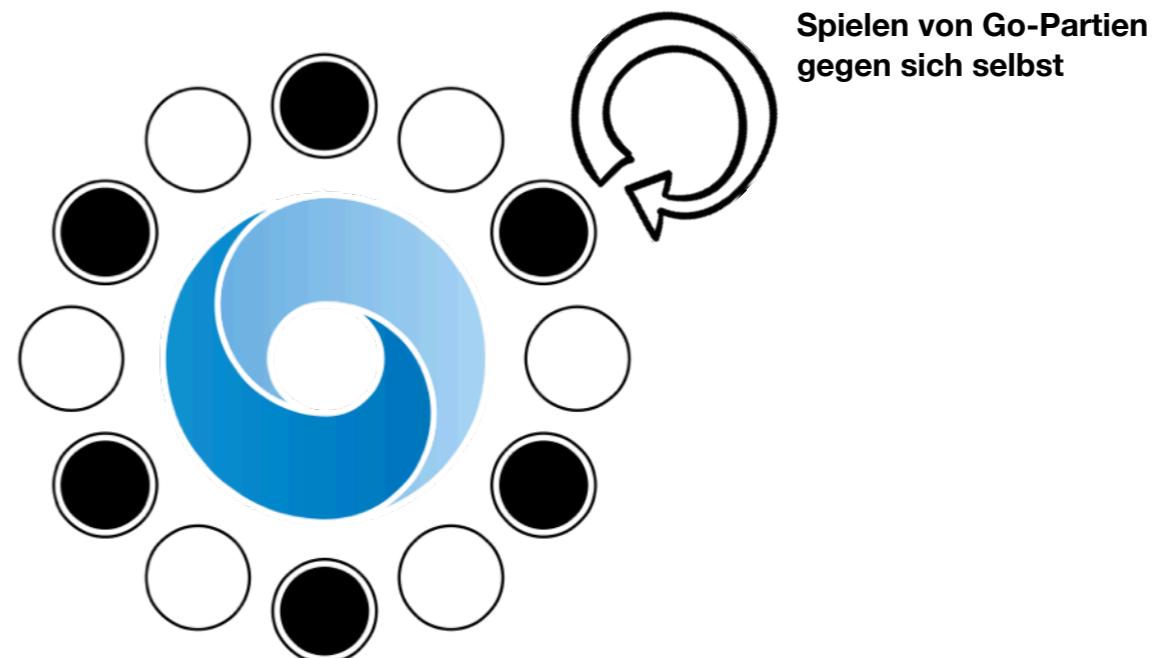
Bestärkendes Lernen

Voraussetzung für das bestärkende Lernen:

- Beschreibung der Aufgabe
 - Angabe, welche Aktionen möglich sind
 - verlässlicher Simulator, mit dem vorgenommene Aktionen bewertet werden können
- Den Rest findet die Maschine im Idealfall selbstständig heraus

Spielregeln des Go-Spiels:

- wann ist welcher Spielzug erlaubt
- was passiert bei dem jeweiligen Spielzug



Unüberbewachtes Lernen

Keine Zielwerte oder Zielergebnisse für Datenmenge vorhanden, da diese meist sehr groß und unstrukturiert

- keine Fehlerberechnung und Verteilen von Belohnungen/Strafen möglich
- Finden von Strukturen/Ähnlichkeiten in unmarkierten Datenmengen anhand von Merkmalen/Features
- Einteilung in Cluster/Gruppen (Clustering)

Unüberbewachtes Lernen

Bilden Sie zwei Gruppen für diese Pflanzen!



Einteilung in Gruppen:

- eine Gruppe á 3 Pflanzen und eine Gruppe á 1 Pflanze
- sowie zwei Gruppen á jeweils 2 Pflanzen

Unüberbewachtes Lernen

Anwendungsbeispiel:

"Kunden, die diesen Artikel gekauft haben, kauften auch ..."

"Das könnte Sie auch interessieren ..."

Ziel ist das Einsortieren von Kunden in Gruppen mit ähnlichen Merkmalsausprägungen, um einen Mehrwert zu erzeugen!

Clustering vs. Klassifikation

Beim **Clustering** ist die Intention, Gruppen von ähnlichen Daten zu finden. Zu Beginn des Lernalgorithmus steht noch nicht fest, durch welche Merkmale und Ähnlichkeiten/Unterschiede die Gruppen entstehen.

Beispiel E-Mail: Bei einer Menge von E-Mails werden zwei Cluster gebildet, die ein Experte anschließend als "Spam" bzw. "Wichtig" erkennt.

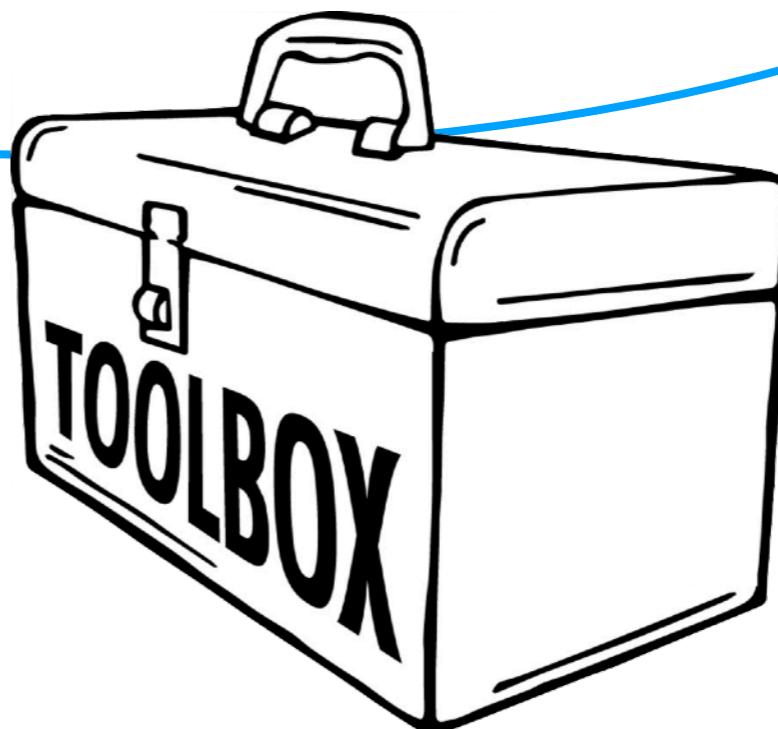
Bei der **Klassifikation** steht bereits vor Anwendung des Lernalgorithmus fest, in welche Gruppen ein Objekt eingeteilt werden kann. Das Ziel des Lernalgorithmus besteht darin, die Merkmale zu detektieren, die für die Zuordnung in die Gruppen signifikant sind.

Beispiel E-Mail: Spam und wichtige E-Mails unterscheiden sich z.B. in den Absendern und den verwendeten Wörtern.

Quelle: Maschinelles Lernen - Kompetenzen, Anwendungen und Forschungsbedarf
(Fraunhofer Gesellschaft)

Maschinelles Lernen

Welches Lernverfahren bietet sich für welche Problemstellung an?



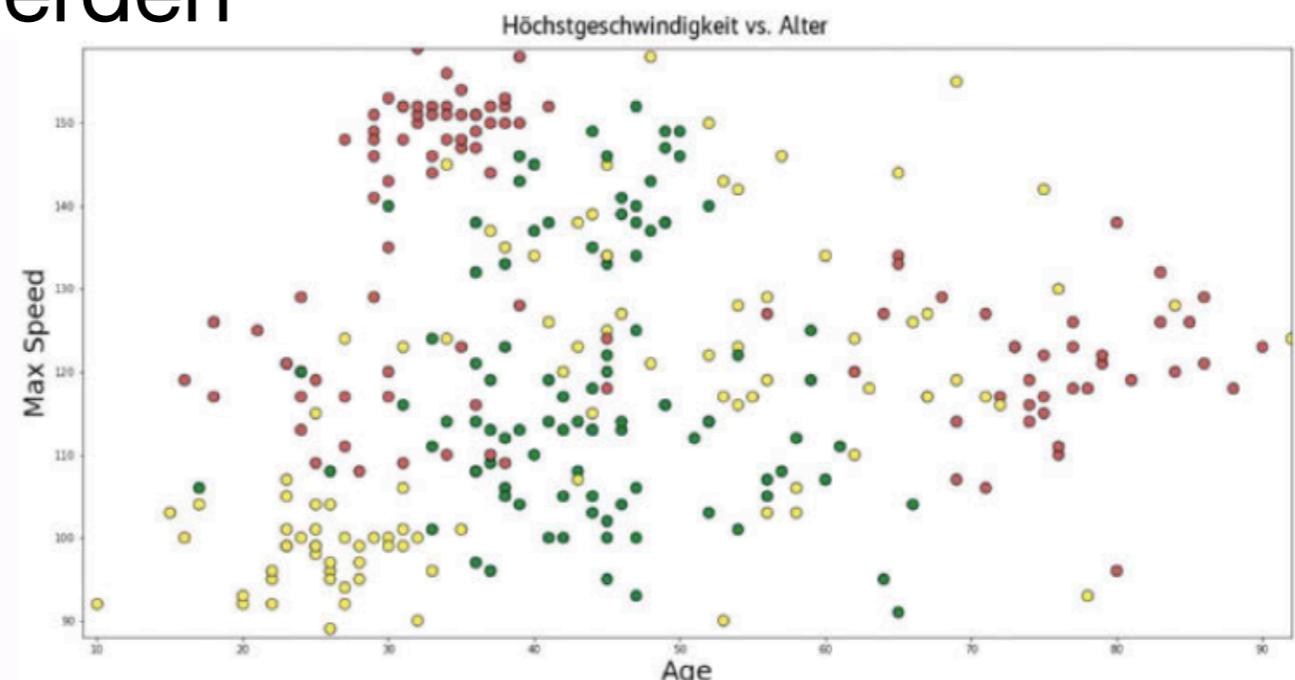
Maschinelles Lernen

Ein Beispiel:



Startup für KFZ-Versicherungen

- Beitragsklasse soll anhand von **tatsächlich aufgetretener Schadensfälle** bestimmt werden
- Folgende Daten liegen vor
 - Alter
 - Höchstgeschwindigkeit



Maschinelles Lernen

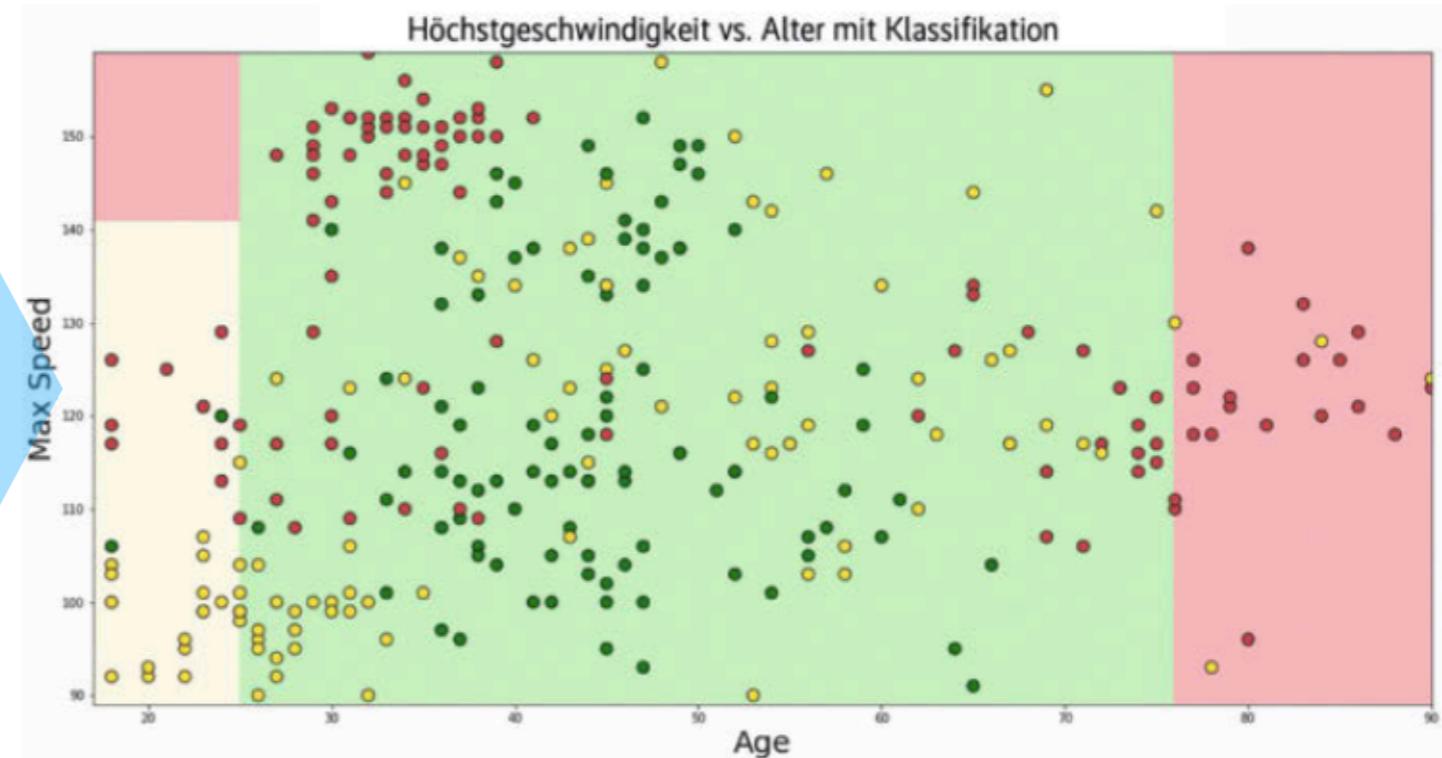
Herkömmlicher Ansatz:

Analytisches Herangehen an die Problemstellung:

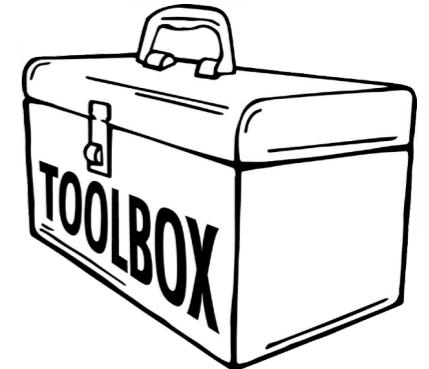
- Analysten finden anhand der Datensätze Regeln heraus
- Softwareentwickler setzen diese programmatisch um
- Software wird zur Klassifikation von Kunden genutzt

Regeln:

1. Jüngere Kunden haben eher ein höheres Risiko, mit schnellen Autos ein sehr hohes Risiko
2. ältere Kunden haben tendenziell ein hohes Risiko
3. andere Kunden haben ein geringes Risiko



Maschinelles Lernen



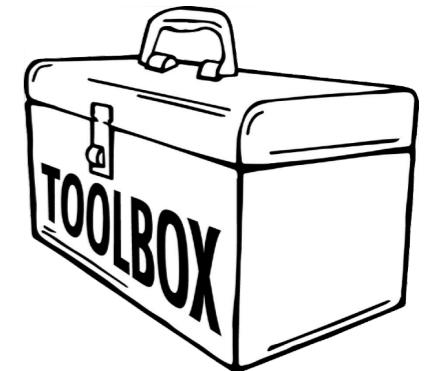
Bestärkendes Lernen:

Wichtig für das bestärkende Lernen:

- Beschreibung der Aufgabe
- Angabe, welche Aktionen möglich sind
- verlässlicher Simulator, mit dem vorgenommene Aktionen bewertet werden können

Für allgemeine Problemstellungen und auch für das Erstellen der Versicherungseinstufungen nicht geeignet.

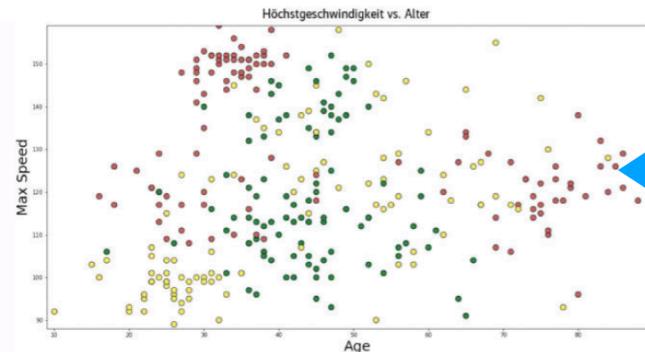
Maschinelles Lernen



Unüberwachtes Lernen:

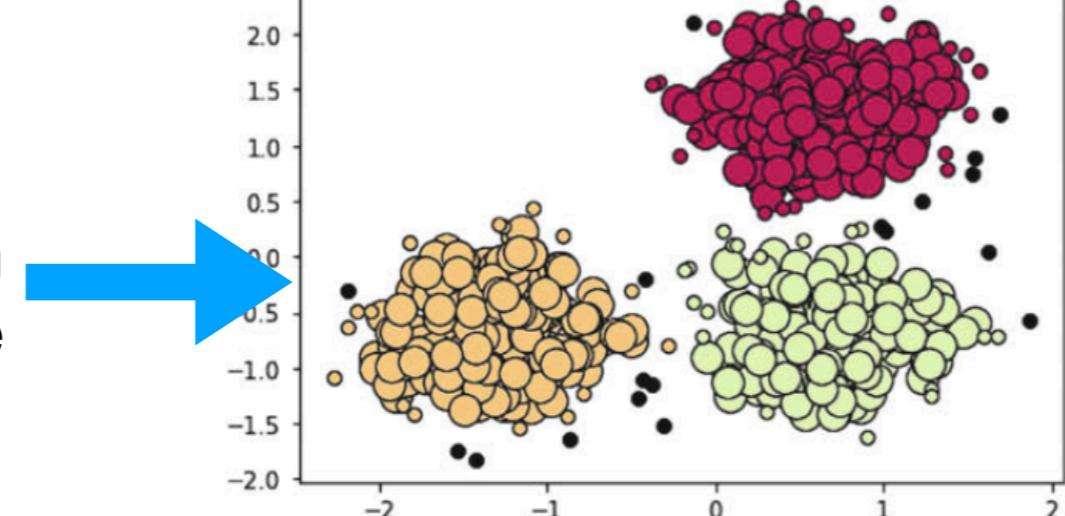
Wichtig für das Clustering:

- Daten sind nicht gelabelt, die Kundendaten werden also berücksichtigt, aber nicht die Einschätzung der Schadensklasse
- Anhand von Features versucht das Clustering-Verfahren die einzelnen Objekte Gruppen zuzuordnen

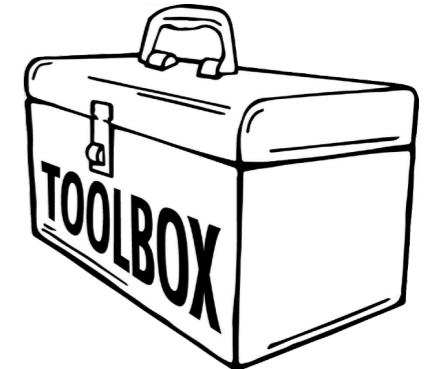


Färbung der Punkte (Label) für Clustering
nicht relevant

Ergebnis des Clusterings: Punkte mit gleicher Farbe gehören einem Cluster an, die schwarzen Punkte sind Ausreißer



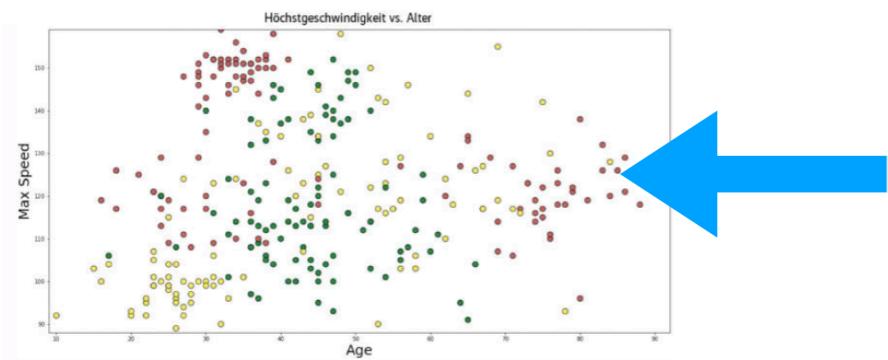
Maschinelles Lernen



Überwachtes Lernen:

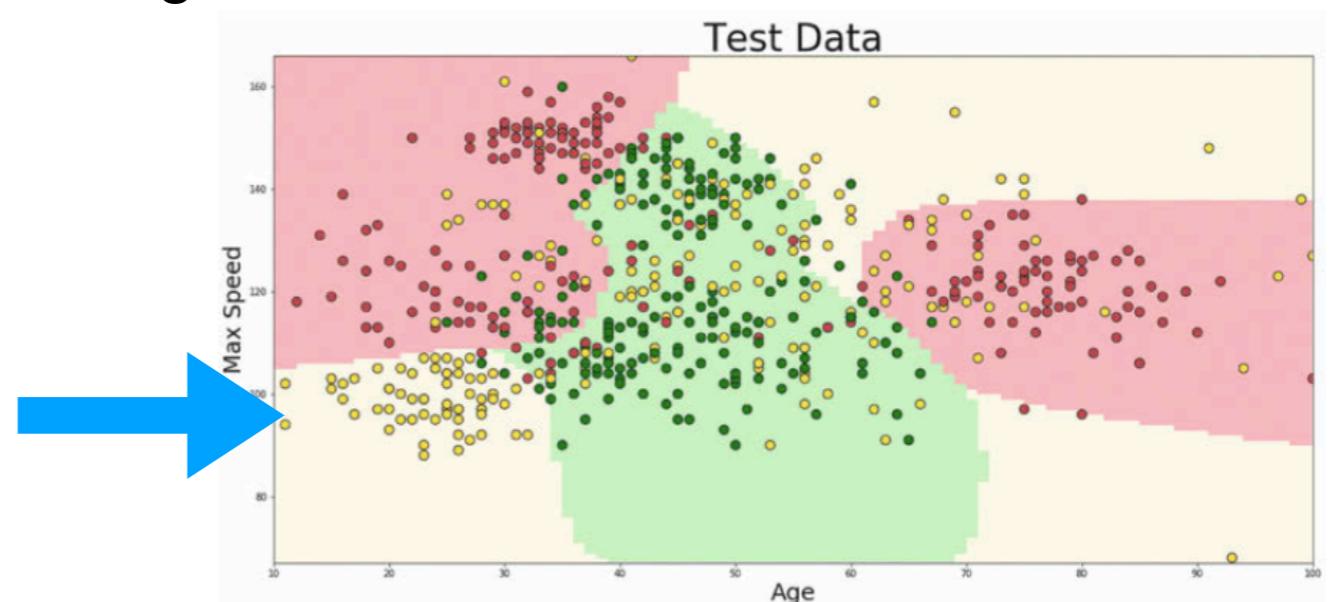
Wichtig für das überwachte Lernen:

- Daten müssen gelabelt sein: passende Paare von Ein- und Ausgabedaten
- Generalisierung von bekannten auf nicht bekannte Daten



Färbung der Punkte (Label) sind optimale Voraussetzungen für überwachtes Lernen

Ergebnis des Überwachten Lernens bzw. eine Vorhersage mit Hilfe von Testdaten. Auffällig sind hierbei die differenzierten und glatten Kanten zwischen den Klassen



Zusammenfassung

Beim Maschinellen Lernen (ML) erlernen die Computer anhand von Daten (Erfahrungen) ein Verhalten

Maschinelles Lernen ist Teil des Prozesses *Knowledge Discovery in Databases*

Aus Big Data soll mit Hilfe von Maschinellem Lernen neues Wissen generiert werden

Grundsätzlich wird bei ML zwischen überwachtem, unüberwachtem und bestärkendem Lernen unterschieden

Die Auswahl des Lernverfahrens ist abhängig von der Problemstellung und der Daten, die zur Verfügung stehen

Die Qualität der Daten ist ausschlaggebend für den Erfolg des Maschinellen Lernprozesses