

Data Science is **hot!**

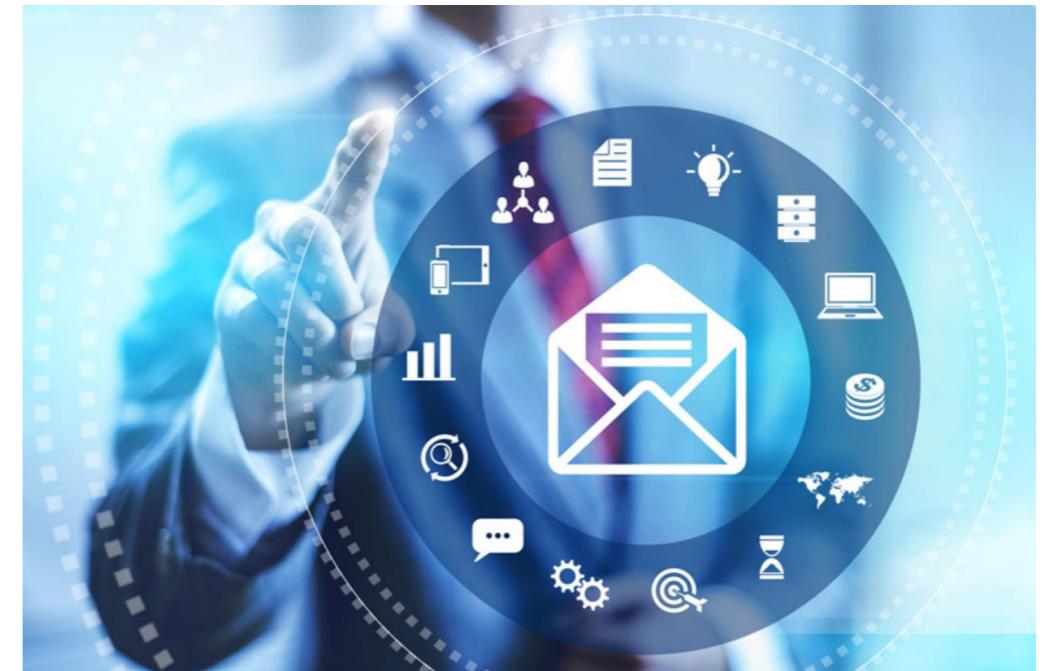
Grundverständnis Data Science
Studiengang Data Science



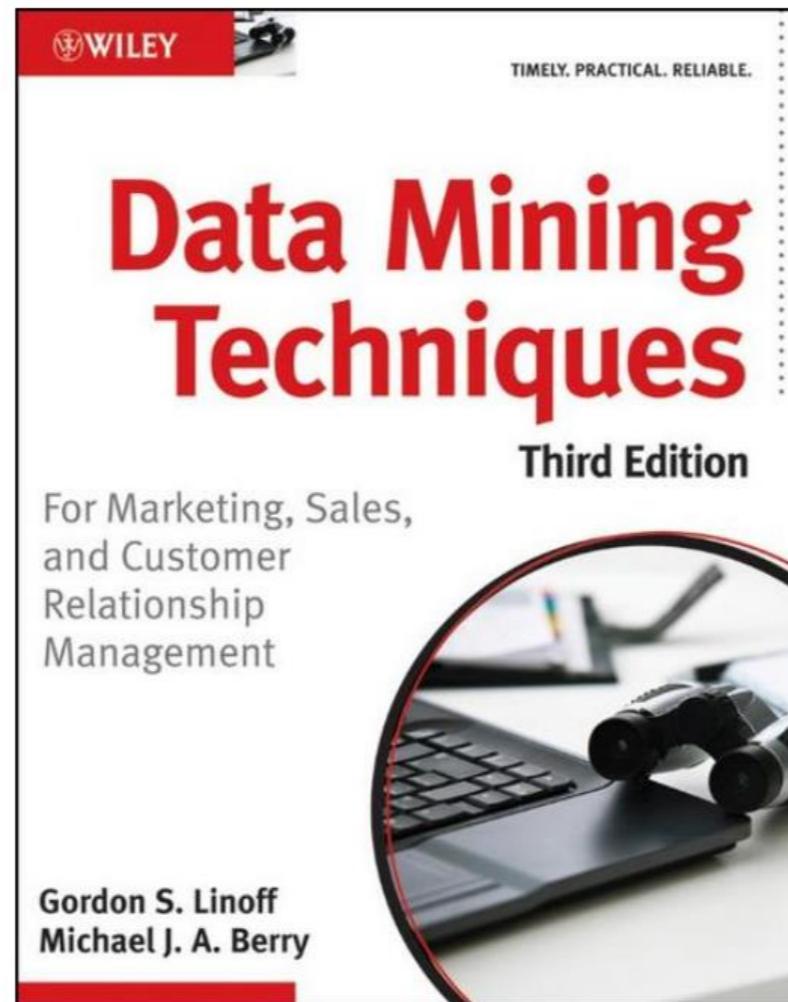
Warum gibt es “Grundverständnis Data Science”?

Data Scientist Interview...

- Die **Marketing & Sales** Abteilung der Bank bittet Sie um Unterstützung bei der **Kundenselektion** für eine geplante **Verkaufsaktion**
- Ziel ist es eine vor einen bestehenden **Anlagefonds zu promoten**
- 15% des Kundenstammes haben dieses Produkt in den letzten 12 Monaten bereits gekauft, der **Kundenanteil soll nun um 5-10% erhöht** werden.
- Das **Marketing-Budget** für die Promotion ist **beschränkt**, die Marketing-Abteilung möchte deshalb möglichst wenige Kunden **anschreiben**.
- Ihnen stehen für die erfolgreiche Ausgestaltung Verkaufs- und Kundendaten (Alter, Wohnort, Geschlecht, Umsatz, Vermögen, Nutzung anderer Produkte) der letzten 2 Jahre zu Verfügung.



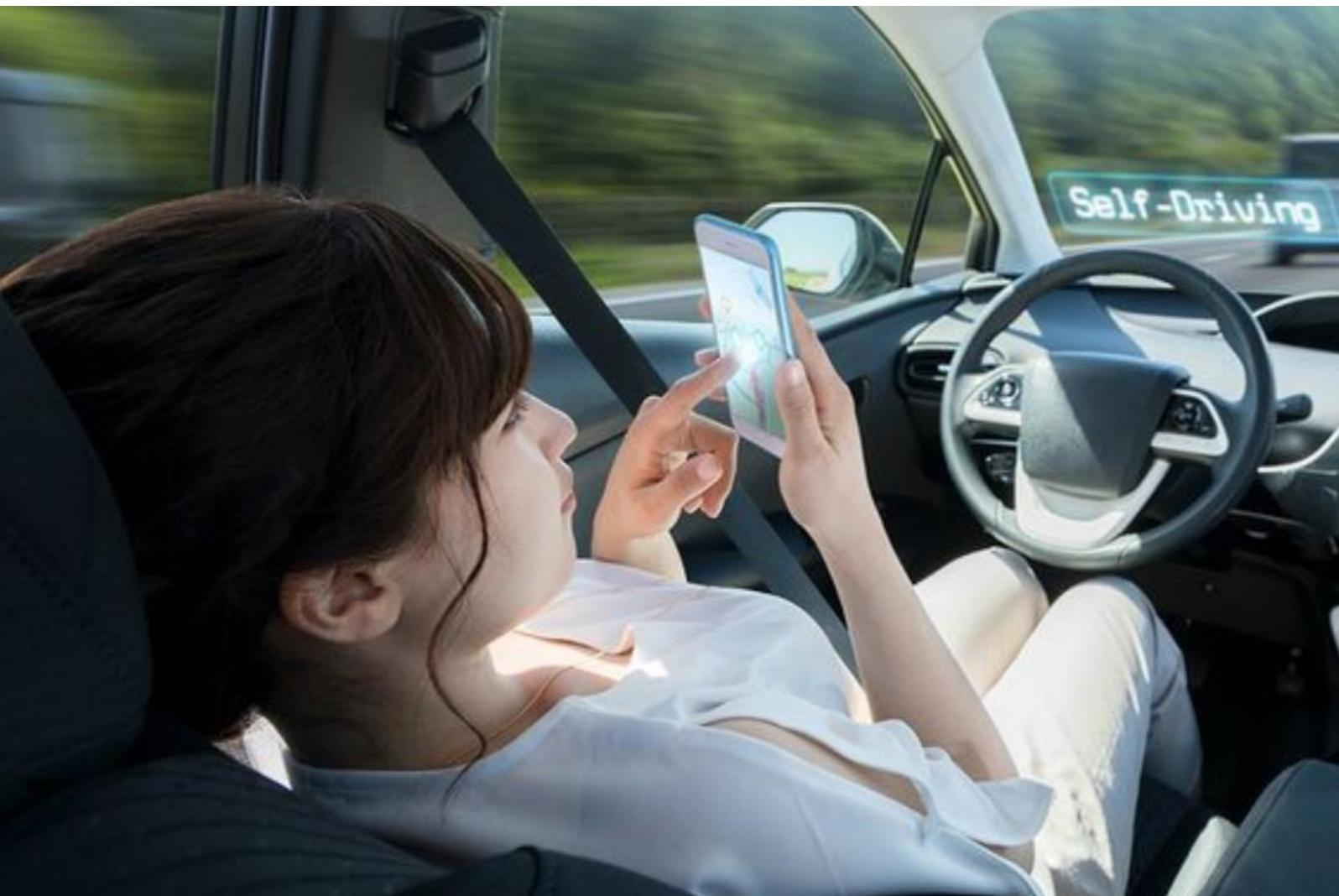
Wo erhaltet ihr ein konzeptionelles **Grundverständnis für Data Science?**



Kapitel 3 "The Data Mining Process" gibt einen ersten Überblick und erklärt wichtige Konzepte

Wo begegnen wir Data Science im Alltag?

Wo begegnen wir Data Science im Alltag?



<https://towardsdatascience.com/how-do-self-driving-cars-see-13054aee2503>

Wo begegnen wir Data Science im Alltag?



<https://www.post.ch/de/ueber-uns/innovation/innovationen-in-entwicklung/drohnen>

Wo begegnen wir Data Science im Alltag?



<https://www.technologyreview.com/s/613838/ai-climate-change-machine-learning/>

Wo begegnen wir Data Science im Alltag?



Wo begegnen wir Data Science im Alltag?



<https://emerj.com/ai-sector-overviews/machine-learning-managing-diabetes-5-current-use-cases/>

Wo begegnen wir Data Science im Alltag?



<https://towardsdatascience.com/hands-on-predict-customer-churn-5c2a42806266>

Wo begegnen wir Data Science im Alltag?



Wo begegnen wir Data Science im Alltag?



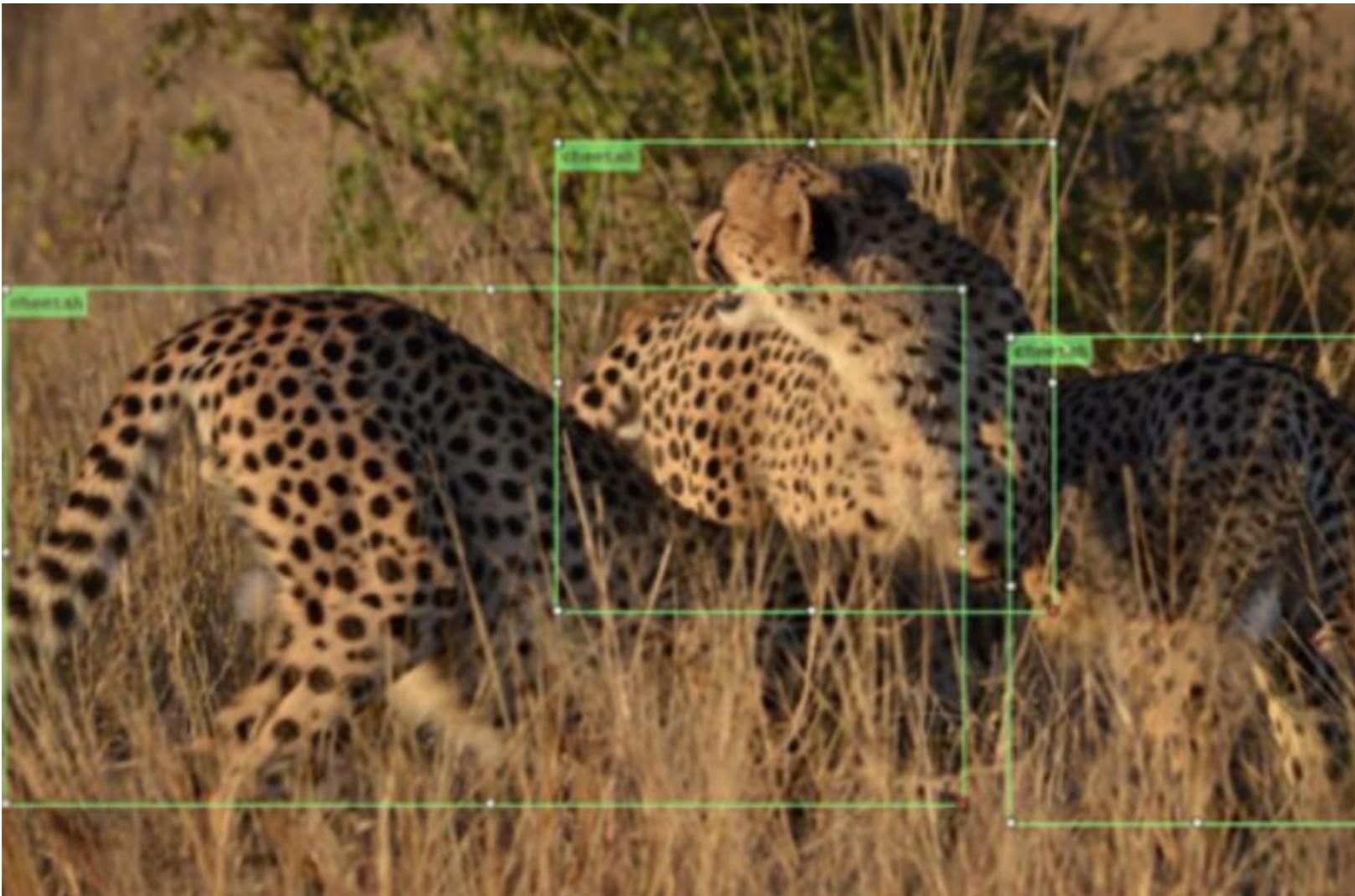
Wo begegnen wir Data Science im Alltag?



Wo begegnen wir Data Science im Alltag?



Wo begegnen wir Data Science im Alltag?

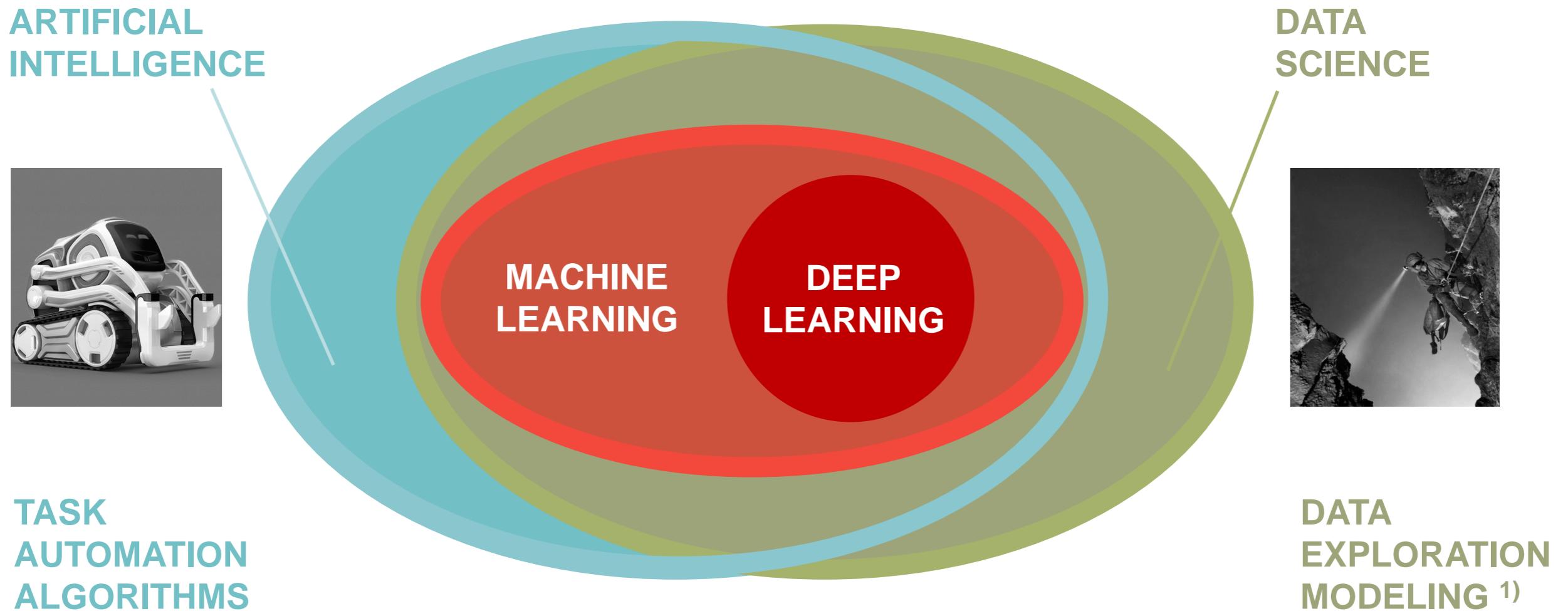


Was **ist** Data Science????

Was **ist** Data Science?



Was ist Data Science?

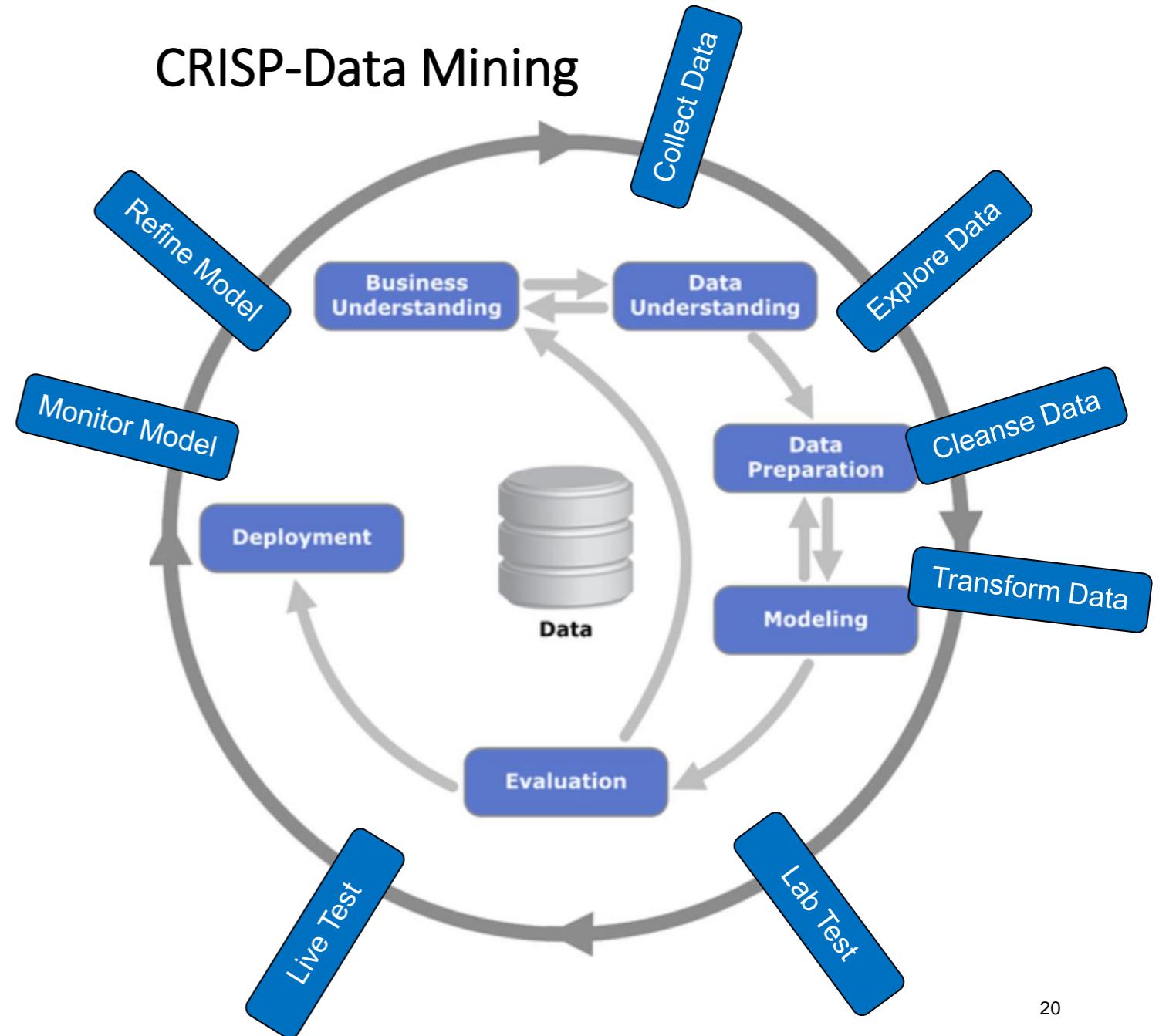


1) <https://www.amstat.org/asa/files/pdfs/EDU-DataScienceGuidelines.pdf>

Was ist Data Science?



CRISP-Data Mining



Wie sieht Data Science
konkret aus...????

Wie **ähnlich** sind sich Kunden?

PETER



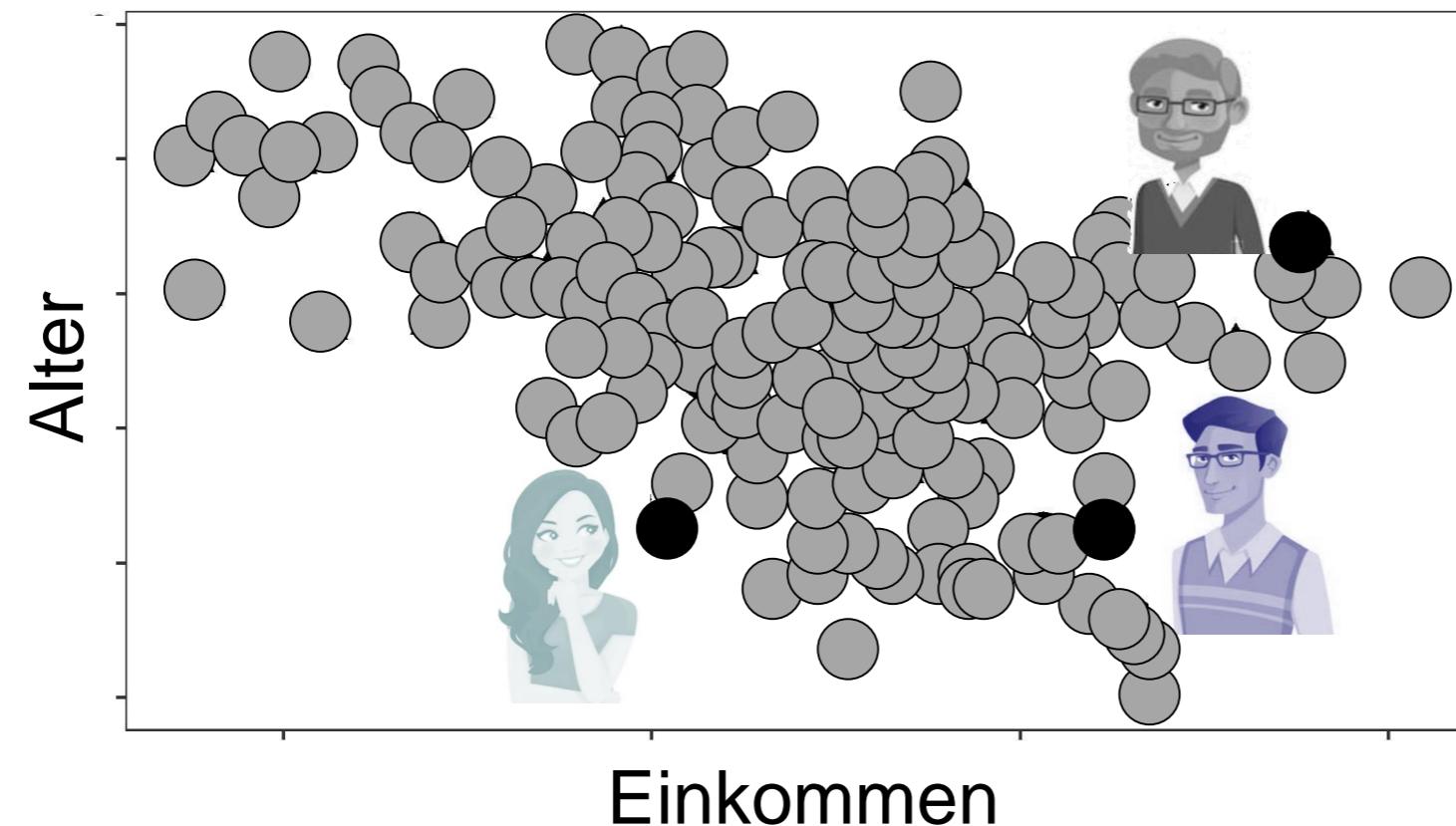
FRANK



ALICE



Wie **ähnlich** sind sich Kunden?



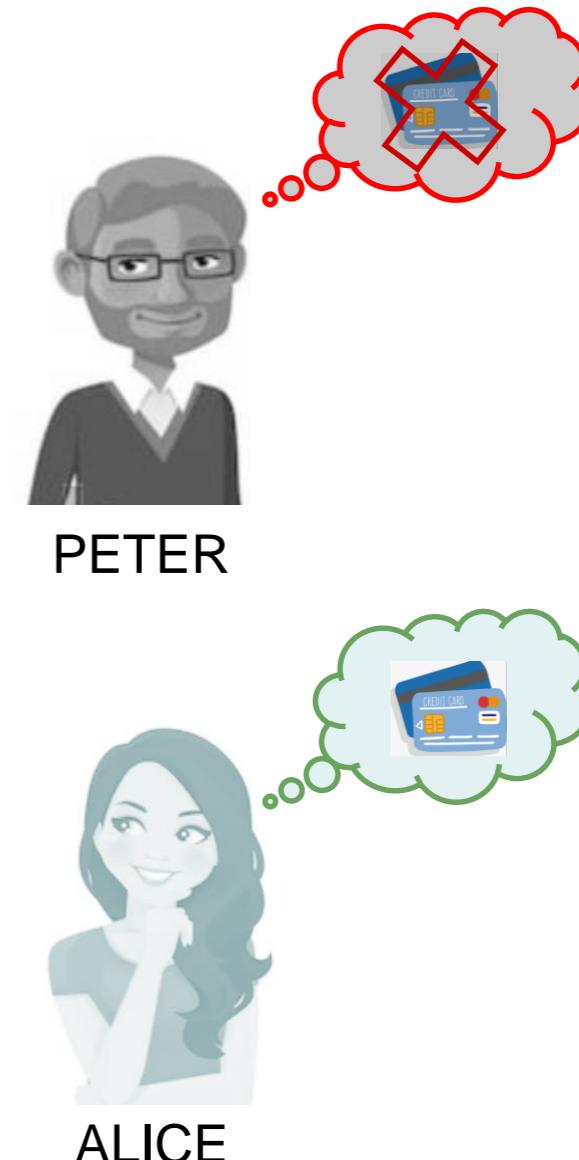
Wie ähnlich sind sich Kunden ...
...in Bezug auf ein spezifisches Produkt?



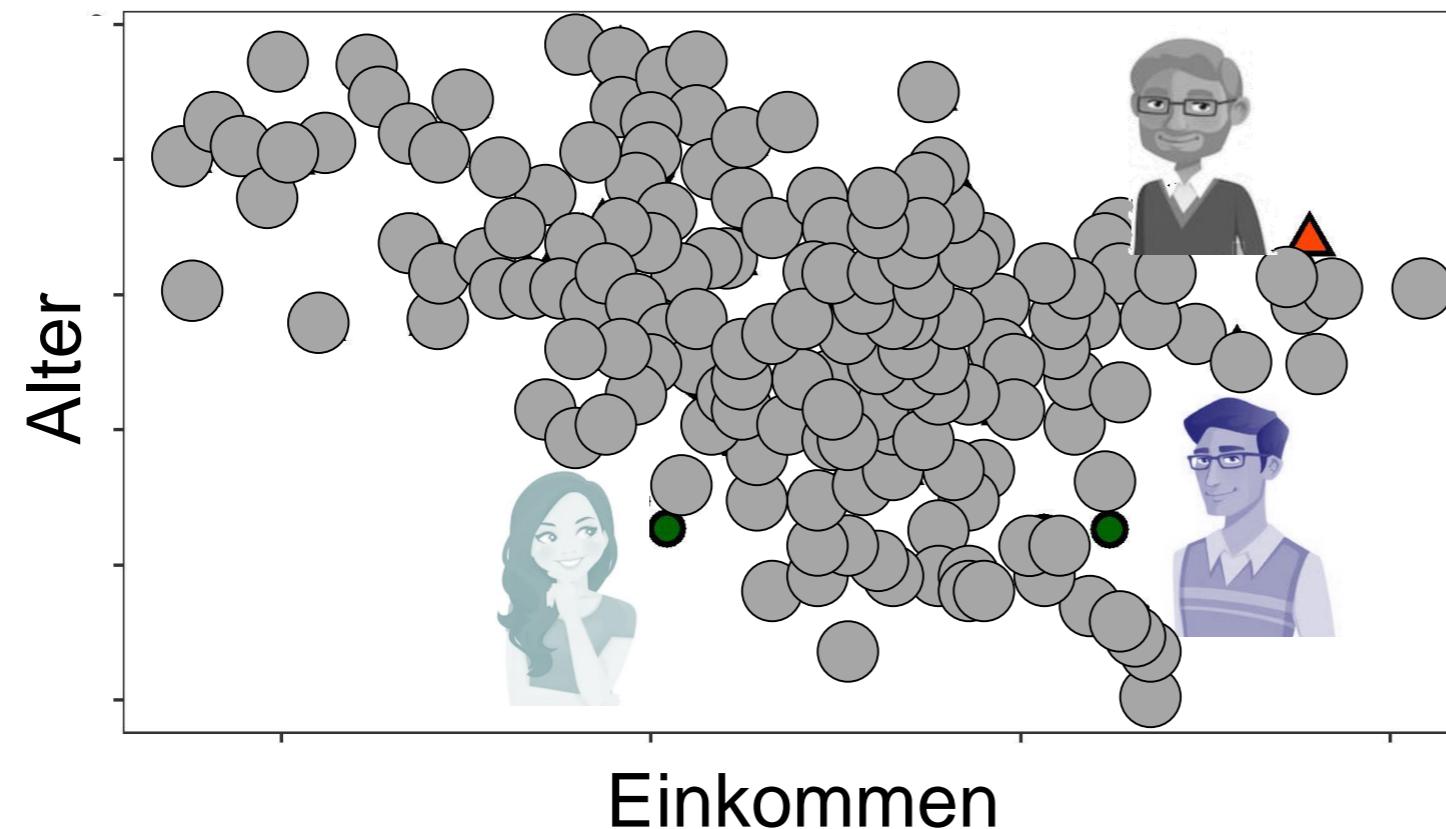
FRANK



Beispiel - Kreditkarte



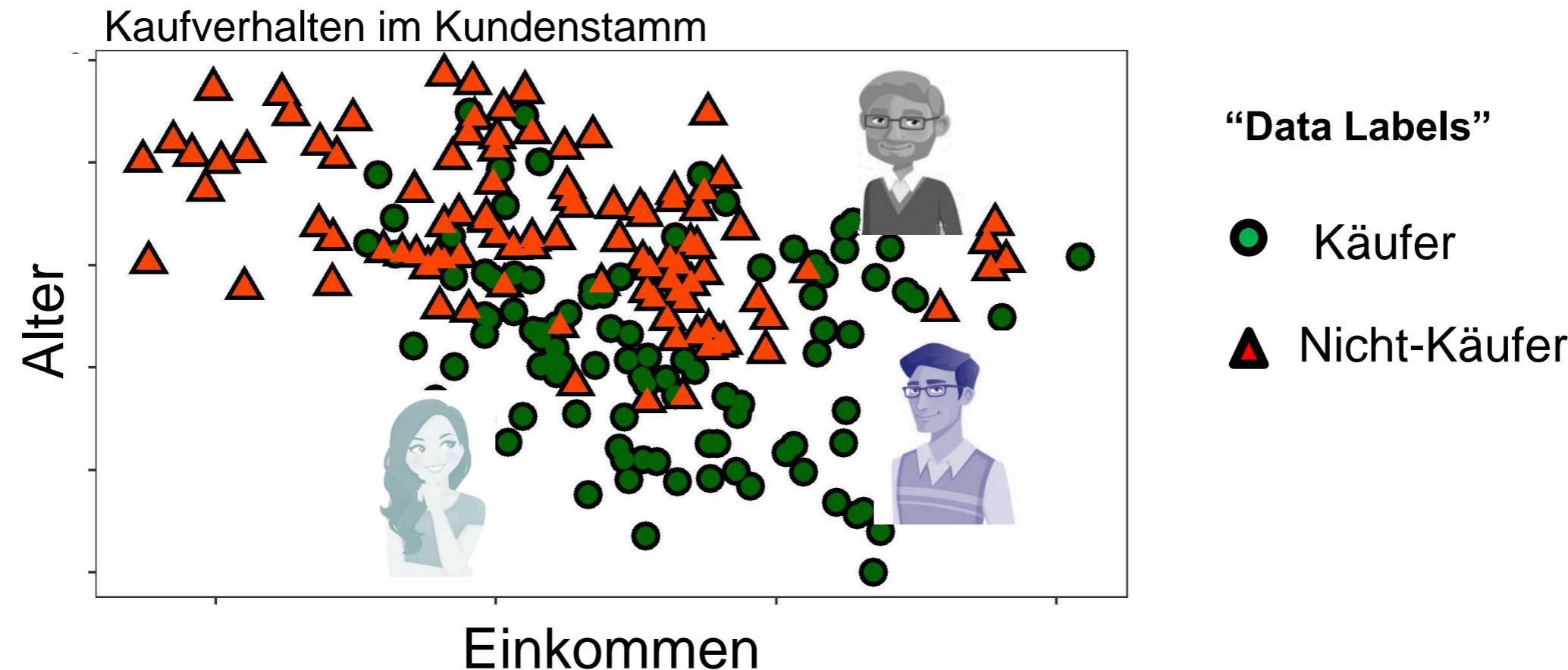
Wie ähnlich sind sich Kunden ...
...in Bezug auf ein spezifisches Produkt?



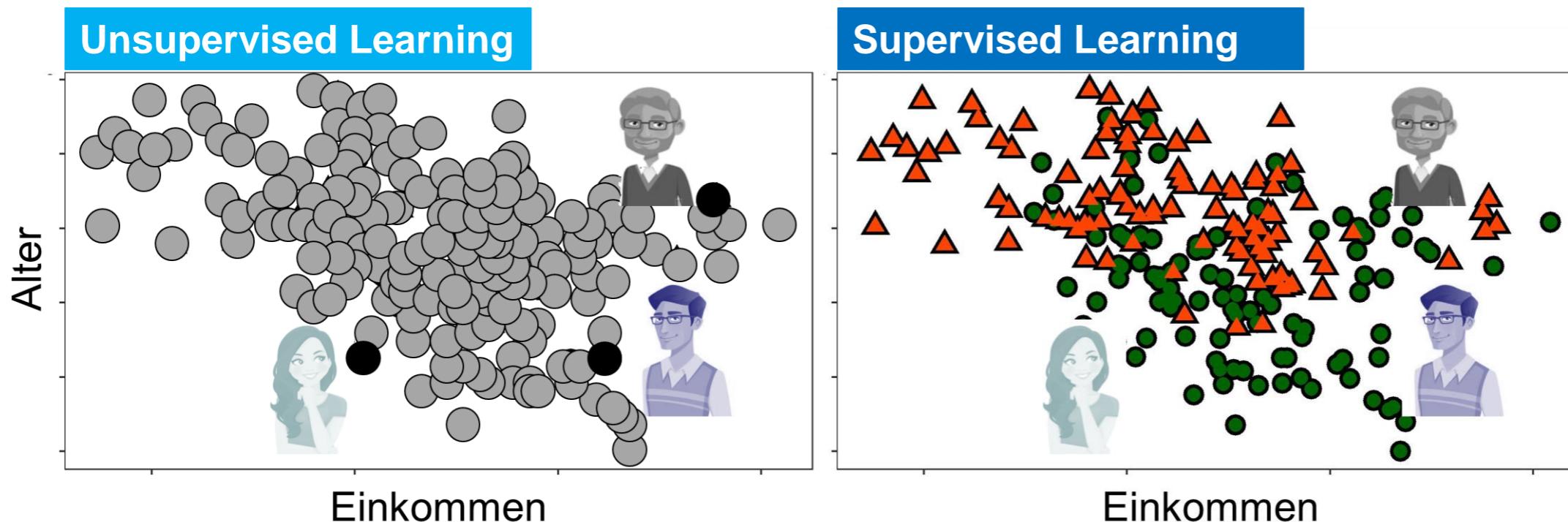
“Data Labels”

- Käufer
- ▲ Nicht-Käufer

Wie ähnlich sind sich Kunden ...
...in Bezug auf ein spezifisches Produkt?

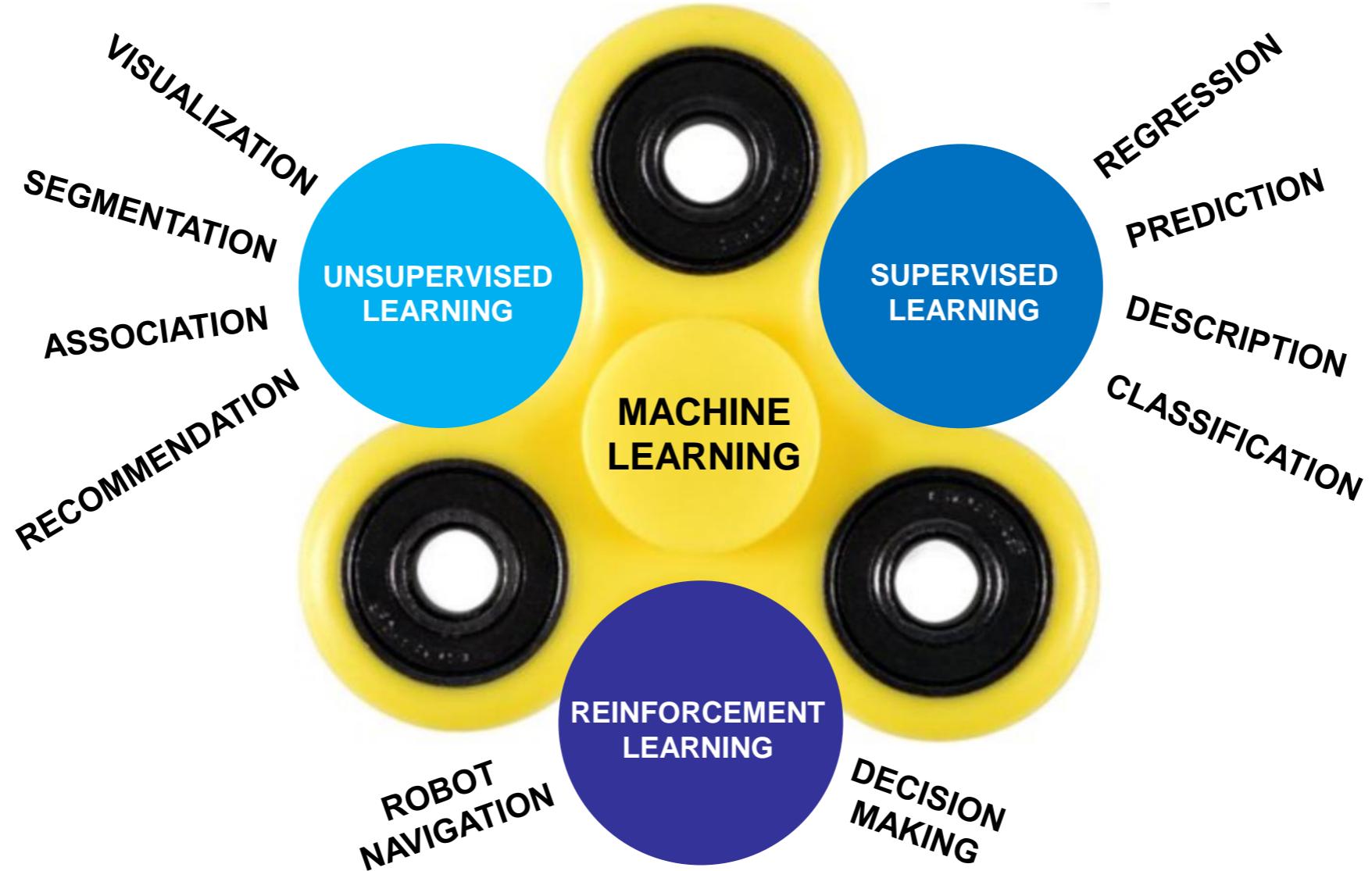


Wie ähnlich sind sich einzelne Kunden?



- “Data Labels”**
- Käufer
 - ▲ Nicht-Käufer

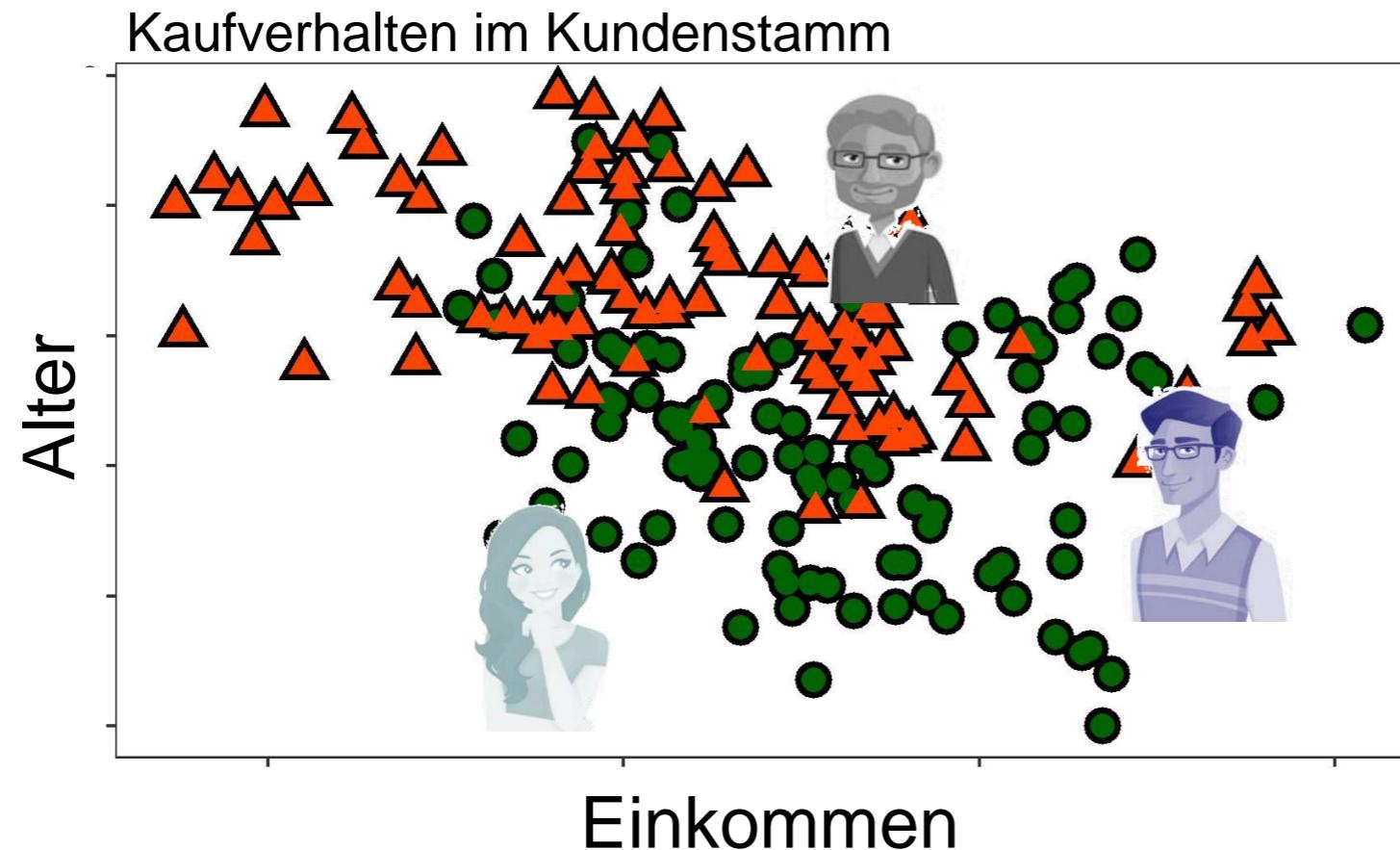
Was **ist** Data Science?



Was hat es mit dem Begriff
“... *Learning*” auf sich?

Was ist Supervised **Learning**?

LE 1

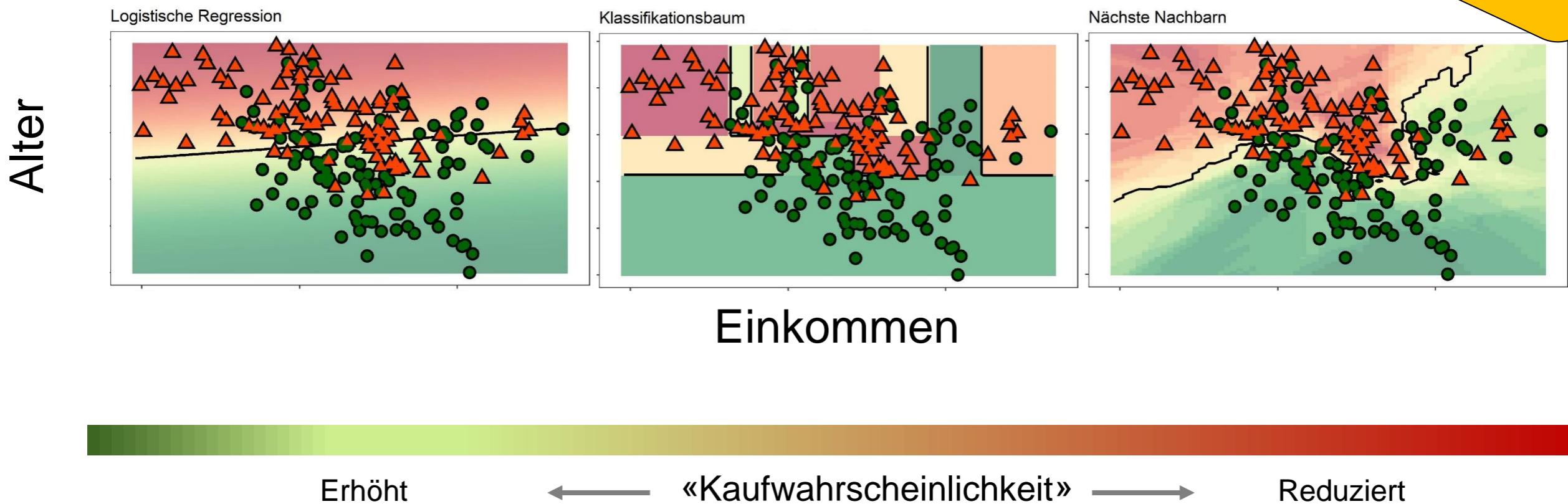


“Data Labels”

- Käufer
- ▲ Nicht-Käufer

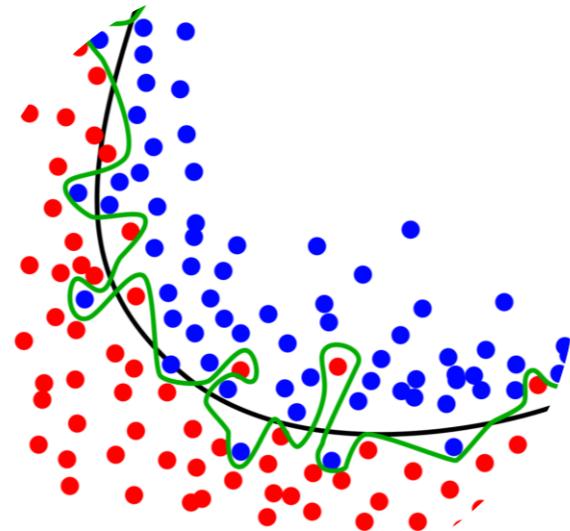
Was ist Supervised Learning?

GDS
Report



Was ist Supervised **Learning**?

Moderne
Algorithmen
können Daten
fast beliebig gut
fitten!



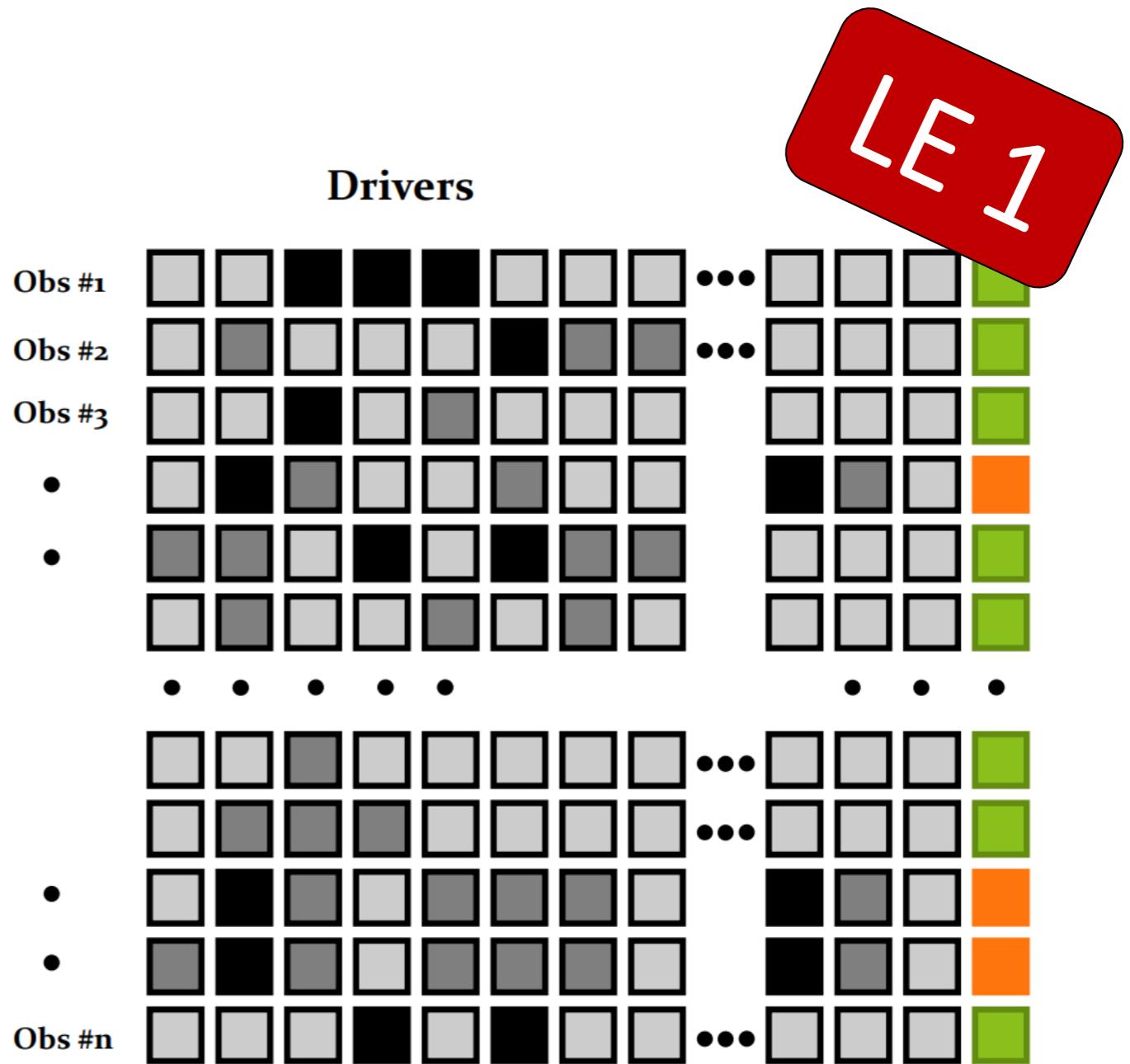
Aber:
There's
no free
lunch!



LE 1

Was ist **Supervised Learning**?

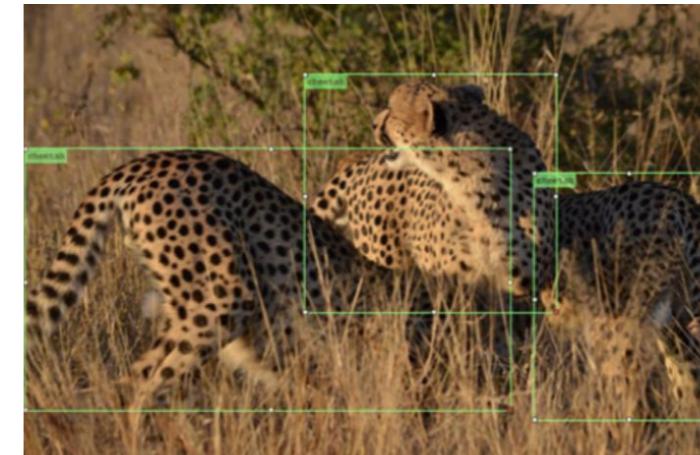
- **Aufgabe:** Identifiziere Einflussfaktoren (“Drivers”), um eine interessierende Zielgröße (“Target”) vorherzusagen
- **Voraussetzung:** Beobachtungen mit bekannten Werten der Zielgröße, von dem das “System” lernen kann (“labelled data”)



Was ist **Supervised Learning**?

- **Aufgabe:** Identifiziere Einflussfaktoren (“Drivers”), um eine interessierende Zielgröße (“Target”) vorherzusagen
- **Voraussetzung:** Beobachtungen mit bekannten Werten der Zielgröße, von dem das “System” lernen kann (“labelled data”)

Beispiel - Objekterkennung



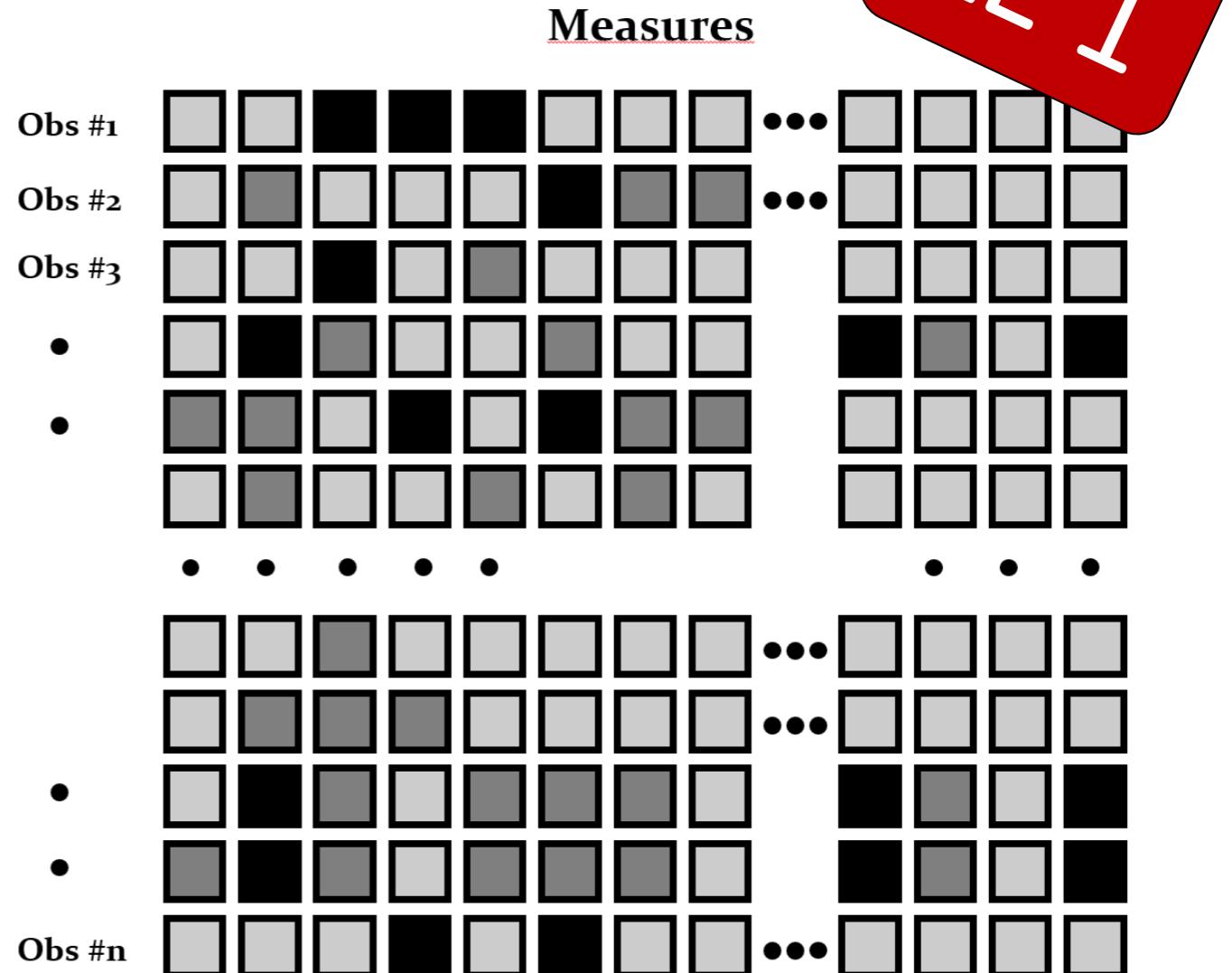
LE 1

Beispiel - Cross-Selling



Was ist **Unsupervised** Learning?

- **Aufgabe:** Identifiziere Zusammenhänge und Muster zwischen einer Vielzahl von gleichberechtigten Größen ("Measures")
- **Voraussetzung:** Definition des Ähnlichkeitsmasses muss der Fragestellung angepasst sein, Beobachtungen müssen normalisiert werden



Was ist **Unsupervised Learning**?

- **Aufgabe:** Identifiziere Zusammenhänge und Muster zwischen einer Vielzahl von gleichberechtigten Größen ("Measures")
- **Voraussetzung:** Definition des Ähnlichkeitsmaßes muss der Fragestellung angepasst sein, Beobachtungen müssen normalisiert werden

Beispiel - Warenkorbanalyse



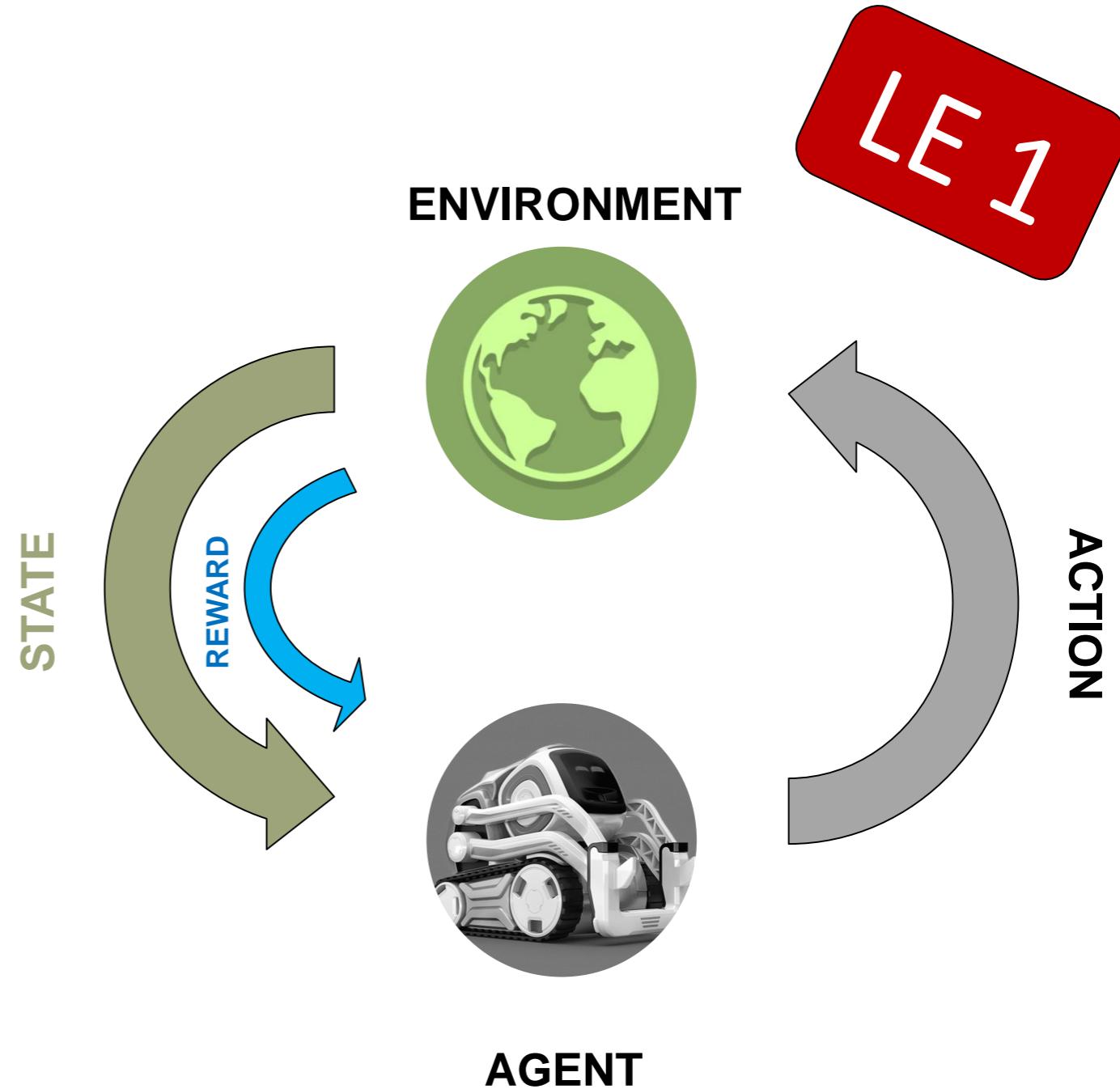
LE 1

Beispiel - Kundensegmentierung



Was ist Reinforcement Learning?

- **Aufgabe:** Finde die optimale Entscheidungsstrategie, um maximale Belohnung zu erhalten
- **Voraussetzung:** Unzählige Beobachtungen, um den Zusammenhang zwischen Aktionen des “Agents” und zu erwartendem “Reward” in jedem “Zustand” zu approximieren



Was ist Reinforcement Learning?

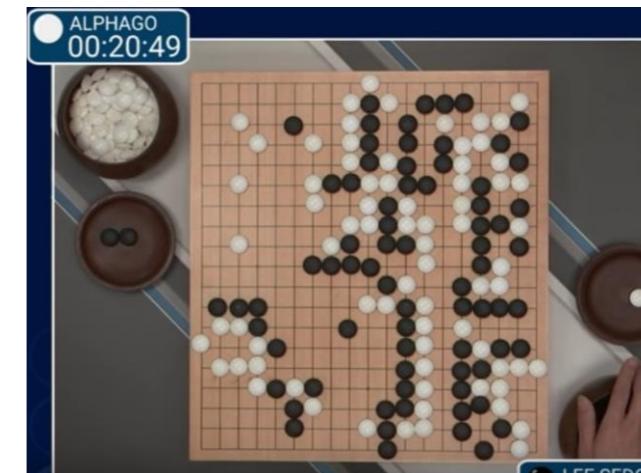
- **Aufgabe:** Finde die optimale Entscheidungsstrategie, um maximale Belohnung zu erhalten
- **Voraussetzung:** Unzählige Beobachtungen, um den Zusammenhang zwischen Aktionen des “Agents” und zu erwartendem “Reward” in jedem “Zustand” zu approximieren

Beispiel - Mondlandung



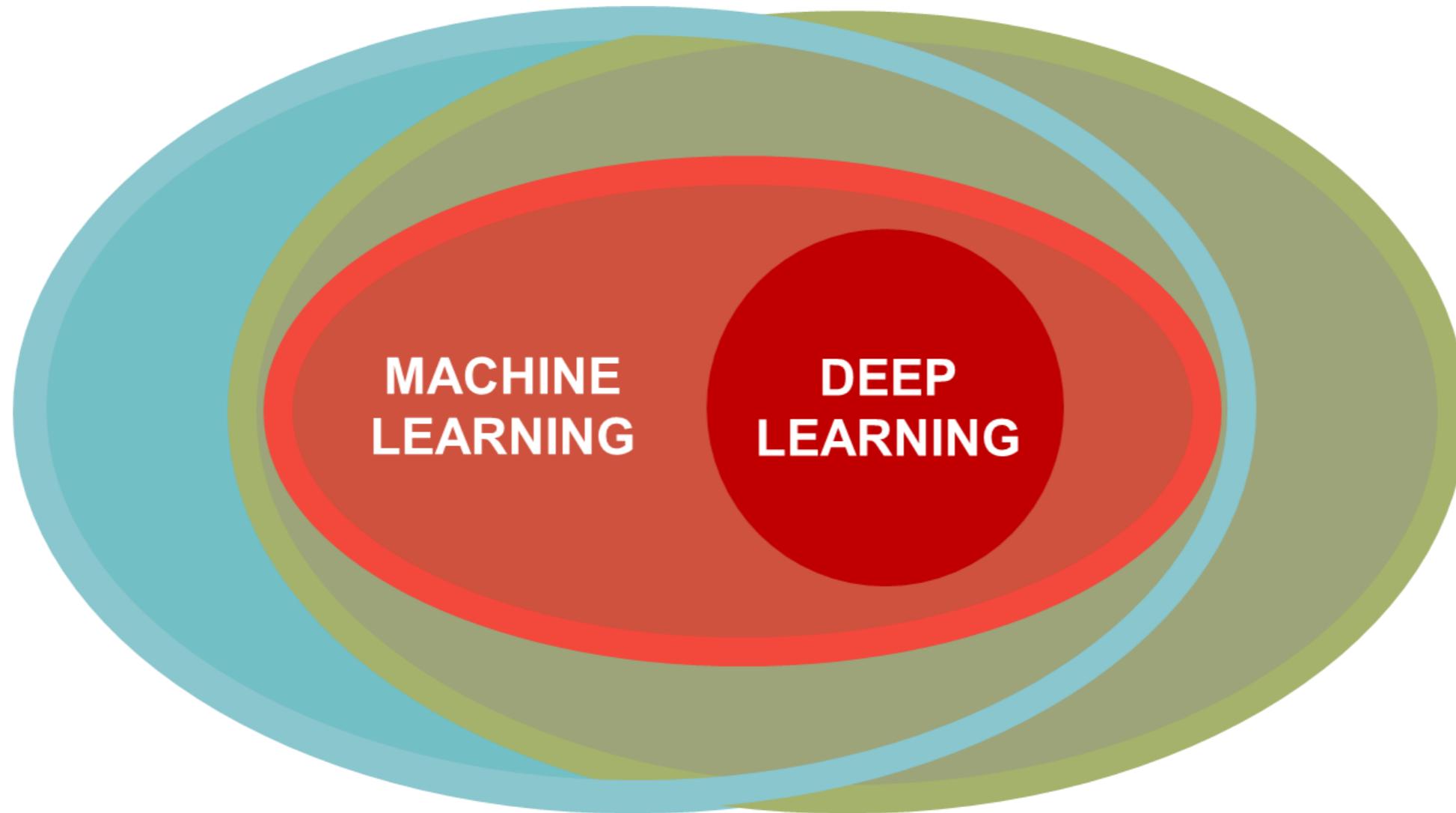
LE 1

Beispiel - Brettspiel GO

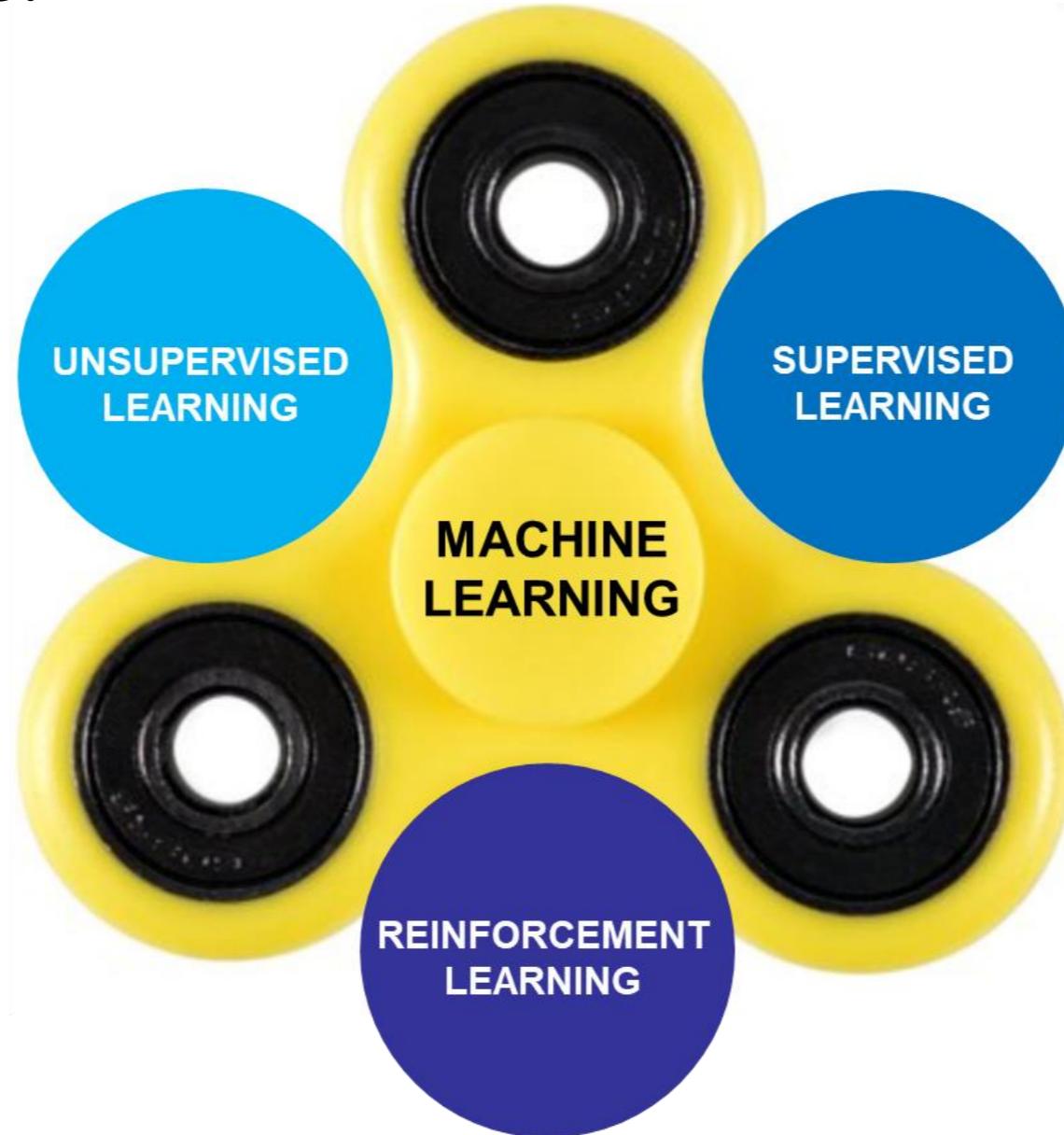


Was **ist** Data Science????

Was **ist** Data Science?



Was **ist** Data Science?



LE 1

Was sind unsere Anforderungen an Machine Learning Modelle...?

Was sind unsere Anforderungen an Machine Learning Modelle...?



- **Performance:** Präzise, verlässliche Vorhersagen
- **Explainability:** Erklärbare Vorhersagen (z.B. bezüglich Einflussfaktoren)
- **Robustness:** Robuste Vorhersagen (z.B. gegenüber Fehler/Extremwerten)
- **Convenience:** Einfachheit (z.B. kein Tuning)
- **Flexibility:** Breit anwendbar (z.B. für unterschiedliche Datentypen)
- **Scalability:** Breit anwendbar und schnell bei grossen Datenmengen
- ...

Wo steckt **Deep Learning** drin...?

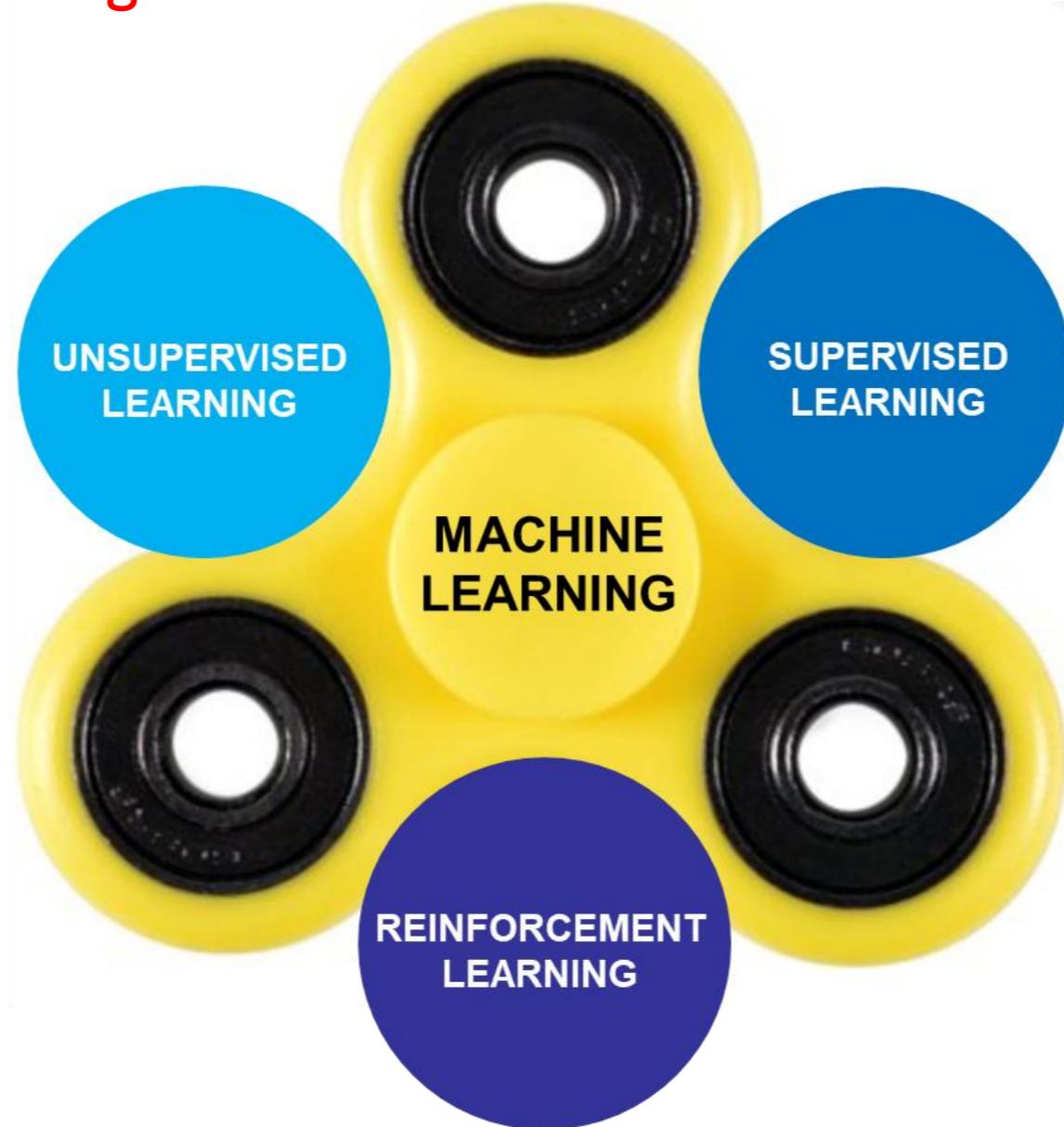


DEEP LEARNING

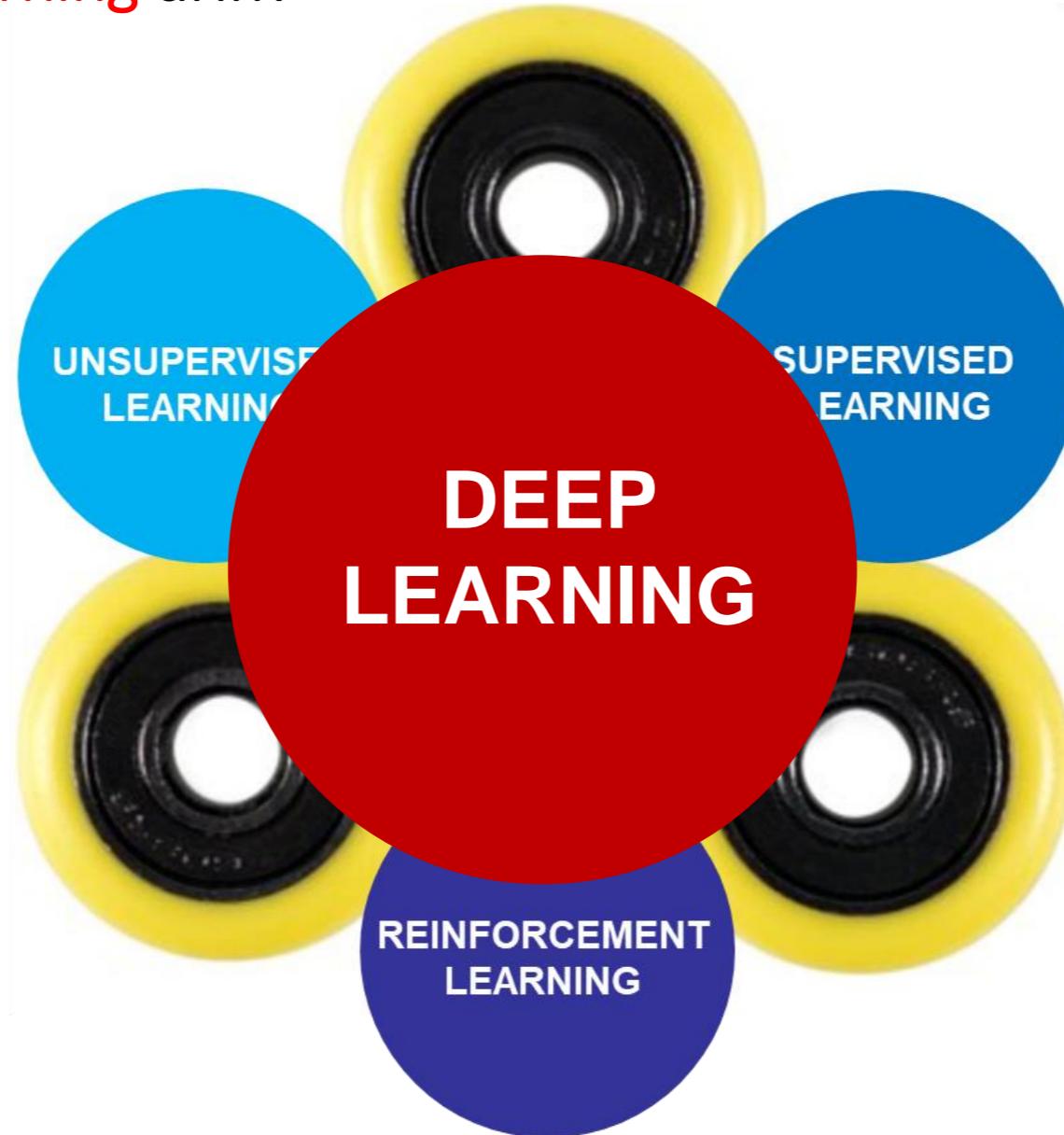


Wo steckt **Deep Learning** drin?

QUIZ



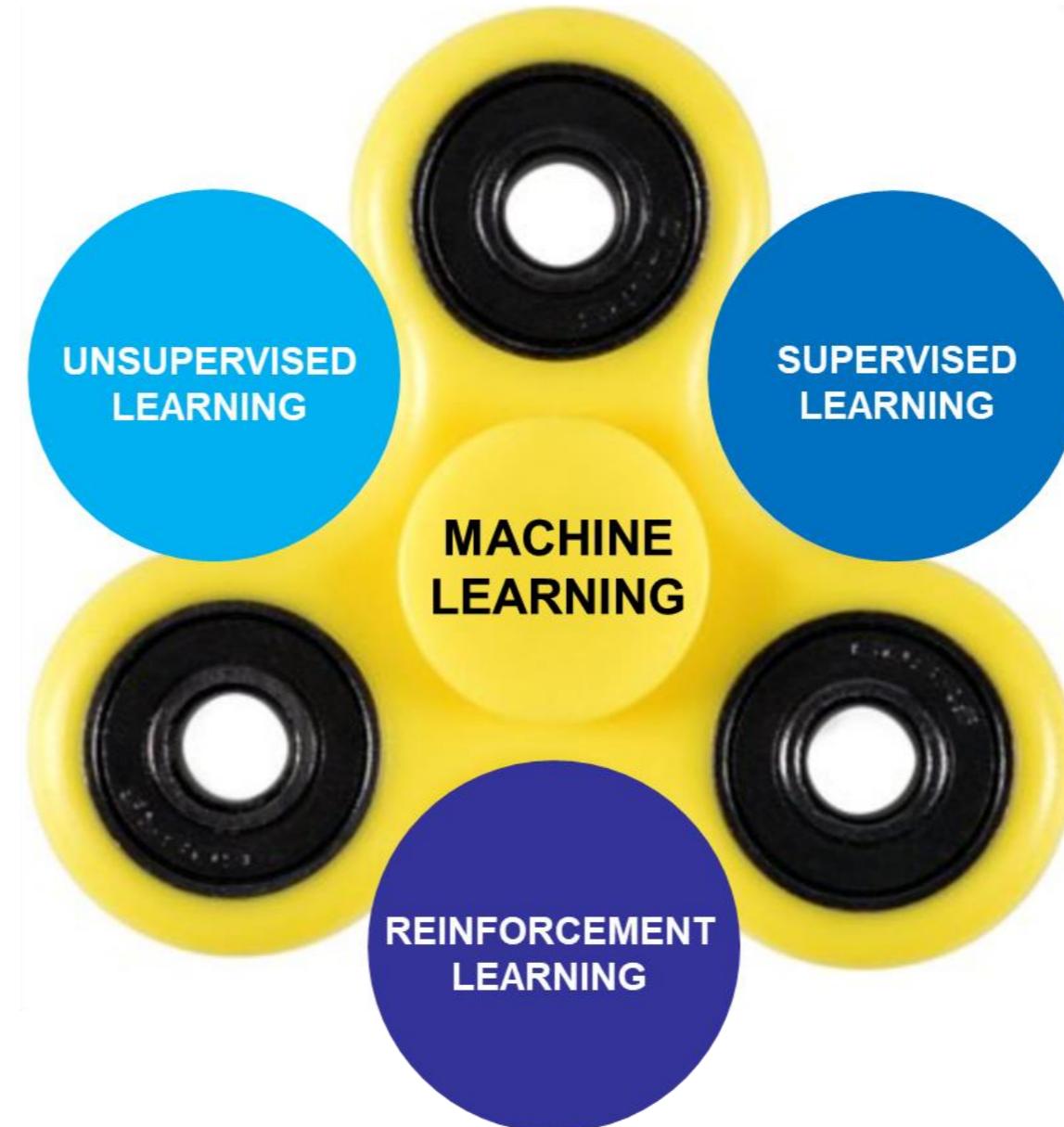
Wo steckt **Deep Learning** drin?



QUIZ

Was unterscheidet die
drei Arten des
Machine Learning?

QUIZ



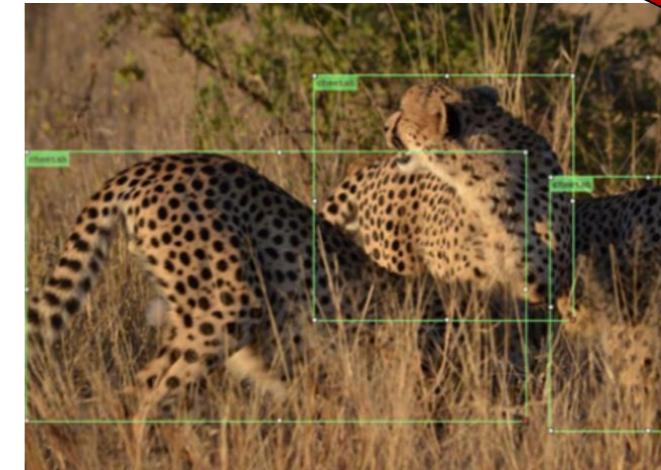
Mondlandung



Warenkorbanalyse



Objekterkennung



QUIZ

Kundensegmentierung



Cross-Selling



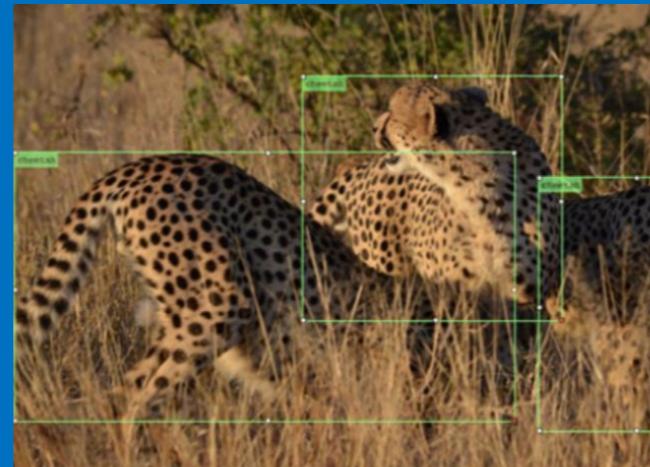
Brettspiel GO



Unsupervised Learning



Supervised Learning



Reinforcement Learning



QUIZ

Wer ist
“Lucy Charles”...?



Wer ist
“Lucy Charles”...?



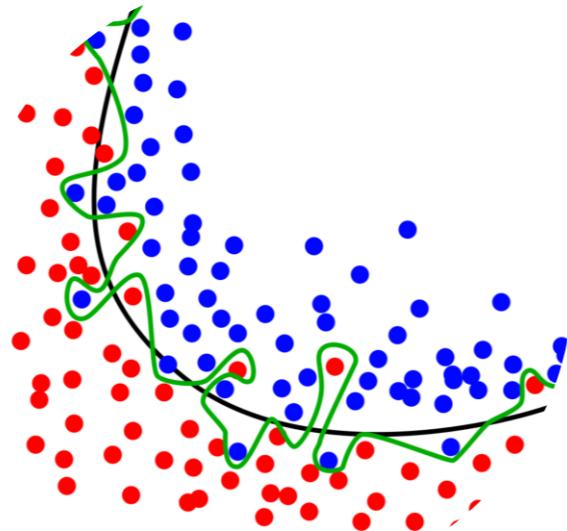
Wie trainiert **“Lucy Charles”** ...?

...um im
Wettkampf
erfolgreich zu
sein?



RECAP

Moderne
Algorithmen
können Daten
fast beliebig gut
fitten!



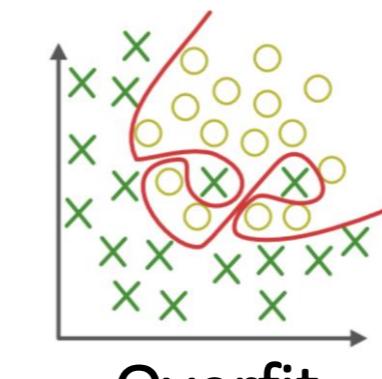
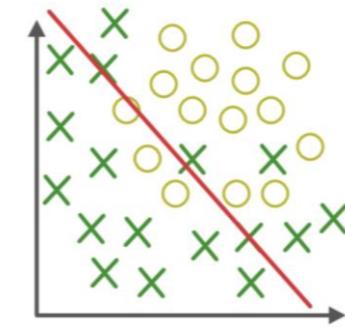
Aber:
There's
no free
lunch!



Was lernen wir von “*Lucy Charles*” ...?



Underfit



Fit



...trainiere unter verschiedenartigen Bedingungen, spezialisiere nicht zu stark!

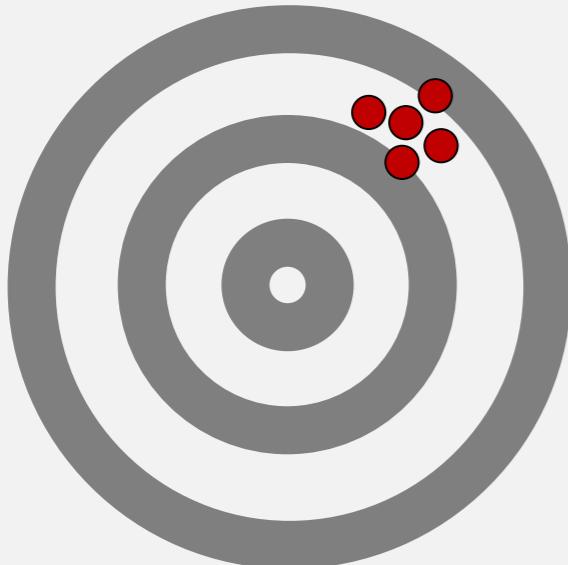
Was hat es mit dem Begriff “... *Learning*” auf sich?

... oder wie man sein Modell fit für die Anwendung macht!

Modellgenauigkeit im Supervised Learning

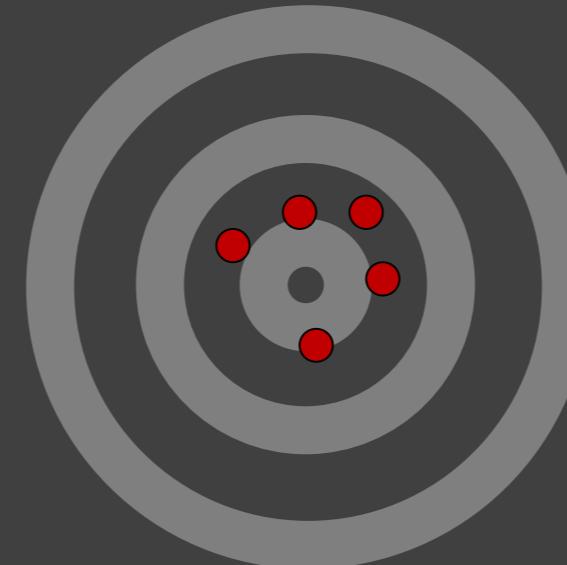
LE 2

High Bias
Low Variance

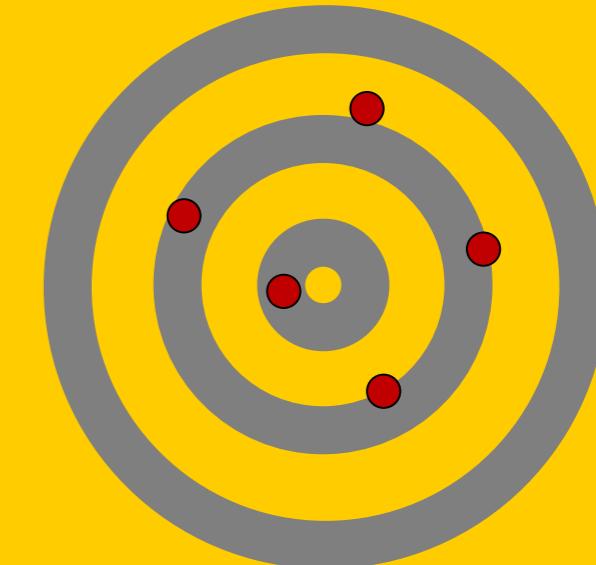


Konsistent daneben
geschossen

Bias-Variance
in balance



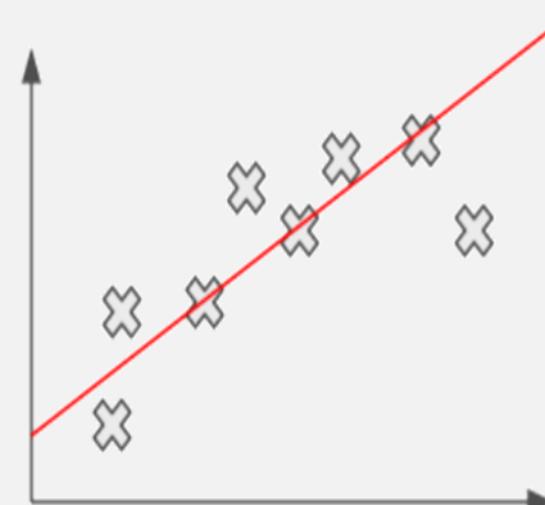
High Variance
Low Bias



Zufällig gezielt,
im Mittel so-so-la-la

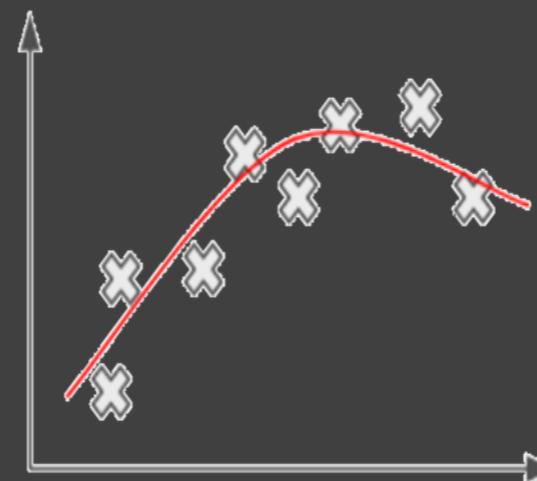
Modellgenauigkeit im Supervised Learning

High Bias
Low Variance



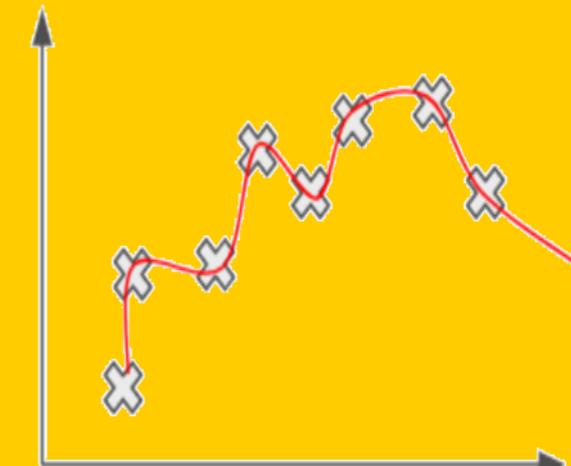
Modell konsistent, aber im
Mittel schlecht

Bias-Variance in
balance



Modell recht
konsistent und gut

High Variance
Low Bias



Modell im Mittel gut,
aber inkonsistent

Modellgenauigkeit im Supervised Learning

LE 2

Modell-Entwicklung

Drivers

	Drivers										Target
Obs #1	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]
Obs #2	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]
Obs #3	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]
•	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]
•	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]
•	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]
Obs #n	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]

Modell-Anwendung

Vorhersage für neue Daten...

Drivers

[]	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]	?
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	---

Beantwortet Fragen der Art...

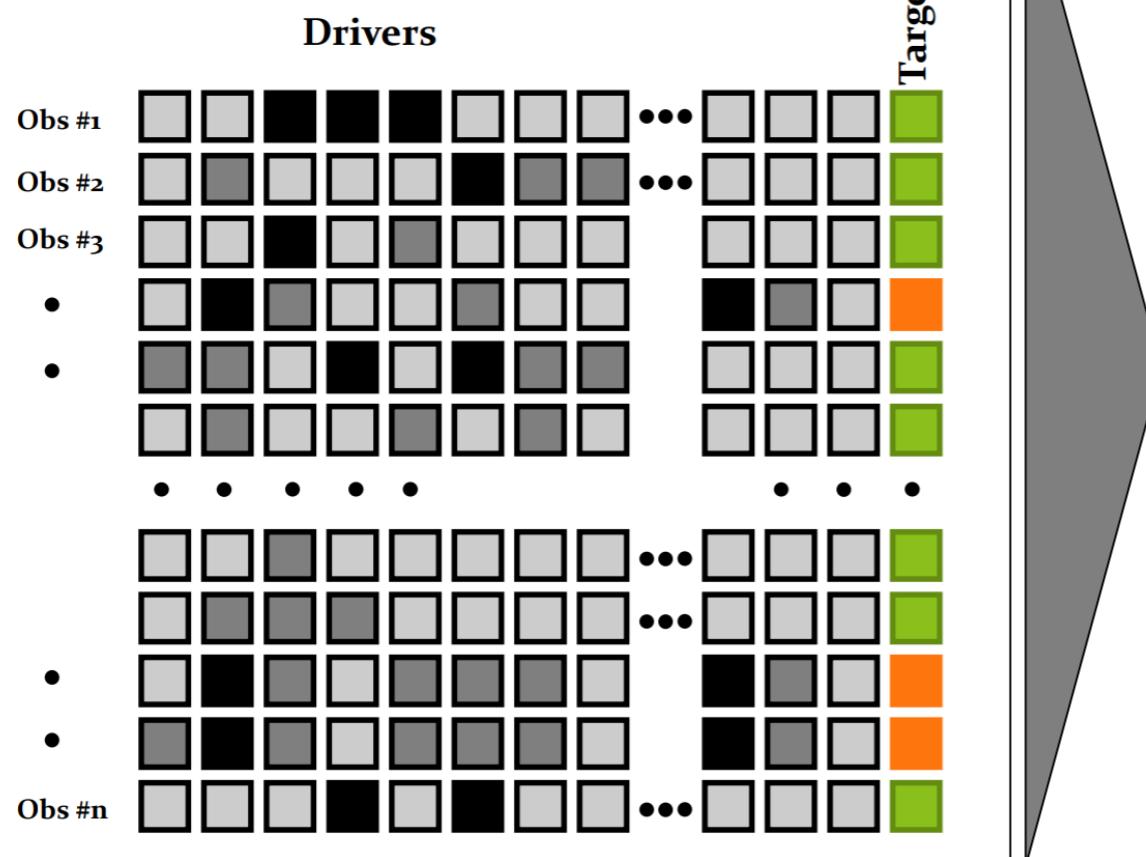
Wie teuer ist die Immobilie?

Kauft die Kund*In die Kreditkarte?

Geht Windturbine innerhalb von 2 Monaten kaputt?

Modellgenauigkeit im Supervised Learning

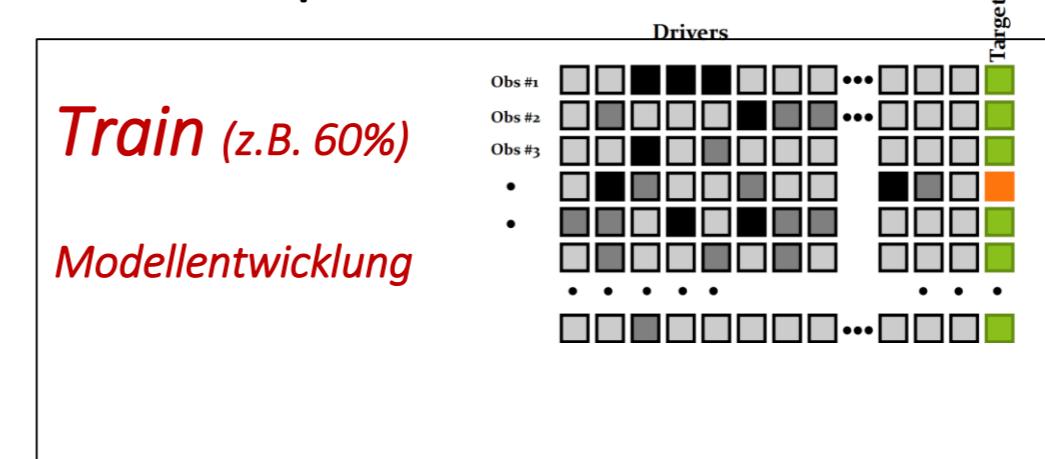
Modell-Entwicklung



Daten-Split

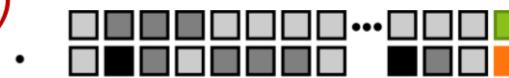
Train (z.B. 60%)

Modellentwicklung



Validate (z.B. 20%)

Modellselektion



Test (z.B. 20%)

Modelleinschätzung



Was hat es mit dem Begriff “... *Learning*” auf sich?

... oder was bei Data Science **schiefgehen** kann!

Was kann bei Data Science **schiefgehen?**

LE 3



Take a new look
at **Amazon jobs**

- Health care benefits day 1
- Performance-based bonuses



LE 3

Was kann bei Data Science **schiefgehen?**

BUSINESS NEWS

OCTOBER 10, 2018 / 5:12 AM / A YEAR AGO

Take a new look on jobs

Amazon scraps secret AI recruiting tool that showed bias against women

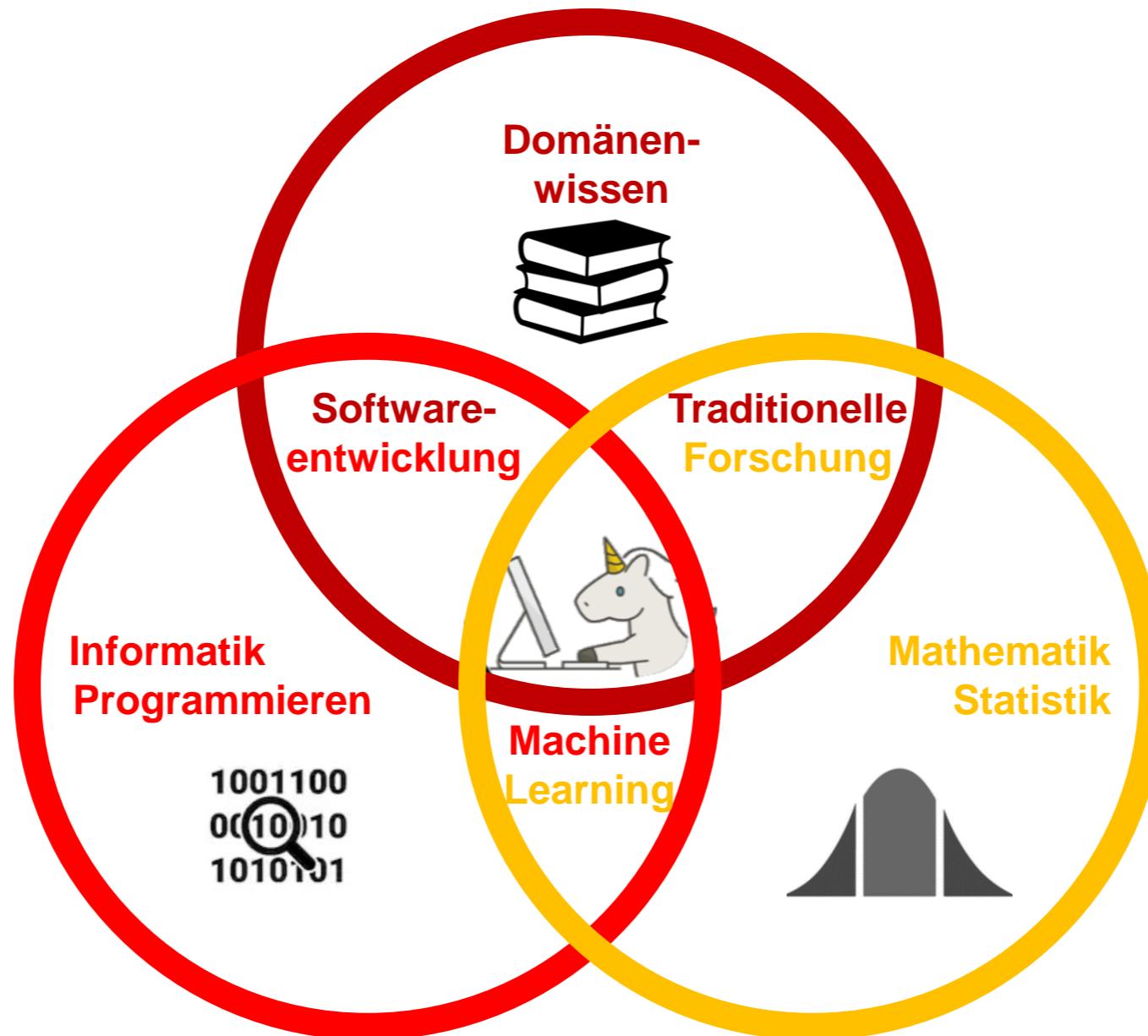
- Headline
- Benefits day
- Performance-based bonuses

The news article discusses how Amazon discontinued an AI tool used for job recruitment that was found to discriminate against women. The banner highlights the specific issue of gender bias in the AI tool.



Was zeichnet eine Data Scientist*in aus?





Deine ersten Schritte als Data Scientist*in!

Deine ersten Schritte als Data Scientist*in!

- Data Science Bericht: Tauche ein, recherchiere und schreibe einen Bericht zum Thema **“Entscheidungsbäume”**
- Umfang: PDF 4 Seiten (ohne Titelseite, Tabellen und Grafiken)
- Struktur: siehe nächste Seite
- Inhalt: siehe nächste Seite
- Abgabedatum: 1. Dezember 2020
- Abgabekanal: DS Spaces
- Namen: “GDS2020_Bericht_Vorname_Name.pdf”



GDS Report
Spezifikation

Stelle **Fragen** in der GDS Kontaktstunde **bis spätestens 17. November 2020!**

Deine ersten Schritte als Data Scientist*in!

Bericht «Entscheidungsbäume»	... beantwortet die Frage(n)...
Einsatzgebiet	Welche grundsätzlichen Problemstellungsklassen können behandelt werden?
Funktionsweise des Algorithmus	Wie sehen die theoretischen/konzeptionellen Konstruktionsschritte aus?
Entwicklung für die Anwendung	Welche Analyse-Schritte müssen für eine konkrete Anwendung/Fallstudie von der Entwicklung bis zur Ausrollen durchgeführt werden?
Vor- und Nachteile des Algorithmus	Welche grundsätzlichen Machine Learning Beurteilungskriterien kennst du für solche Algorithmen, wie schneiden Entscheidungsbäume ab und können sie verbessert werden?
Anwendungsbereiche	Welche Fallstudien sind prädestiniert für den Einsatz von Entscheidungsbäumen, und welche nicht?
Quellen	Welche Informationen hast du schlussendlich verwendet?



Deine ersten Schritte als Data Scientist*in!

- Zielsetzungen:
 - Bestehe dein “1. Data Science Interview”!
 - Erfahre die Faszination “Machine Learning”!
 - Strukturiere und bewerte neue Fakten!
 - Gestalte den Wissenstransport effektiv!

GDS Report
Zielsetzung

Deine ersten Schritte als Data Scientist*in!

- **Vorgehen:**

Bis 11. September 2020:

- erfasst die Aufgabenstellung individuell, recherchiert gezielt in 4-er Gruppen, rapportiert im Plenum (10 Minuten Präsentation)

Bis 17. November 2020

- sammelt und bewertet Informationen individuell, besprecht euch, schreibt Stichworte auf, kommt zu mir in die GDS-Kontaktstunde

Bis 1. Dezember 2020

- tauscht euch untereinander aus, lernt voneinander und schreibt euren Bericht



Wichtige Infos “Grundverständnis Data Science”

- Kürzel GDS
- Kontaktstunde DI, 11:15-12:00
- Ort LearningLab

- ECTS 1
- Leistungsnachweis 4-seitiger Bericht
Abgabe 1.12.2020
Benotete Einzelarbeit

