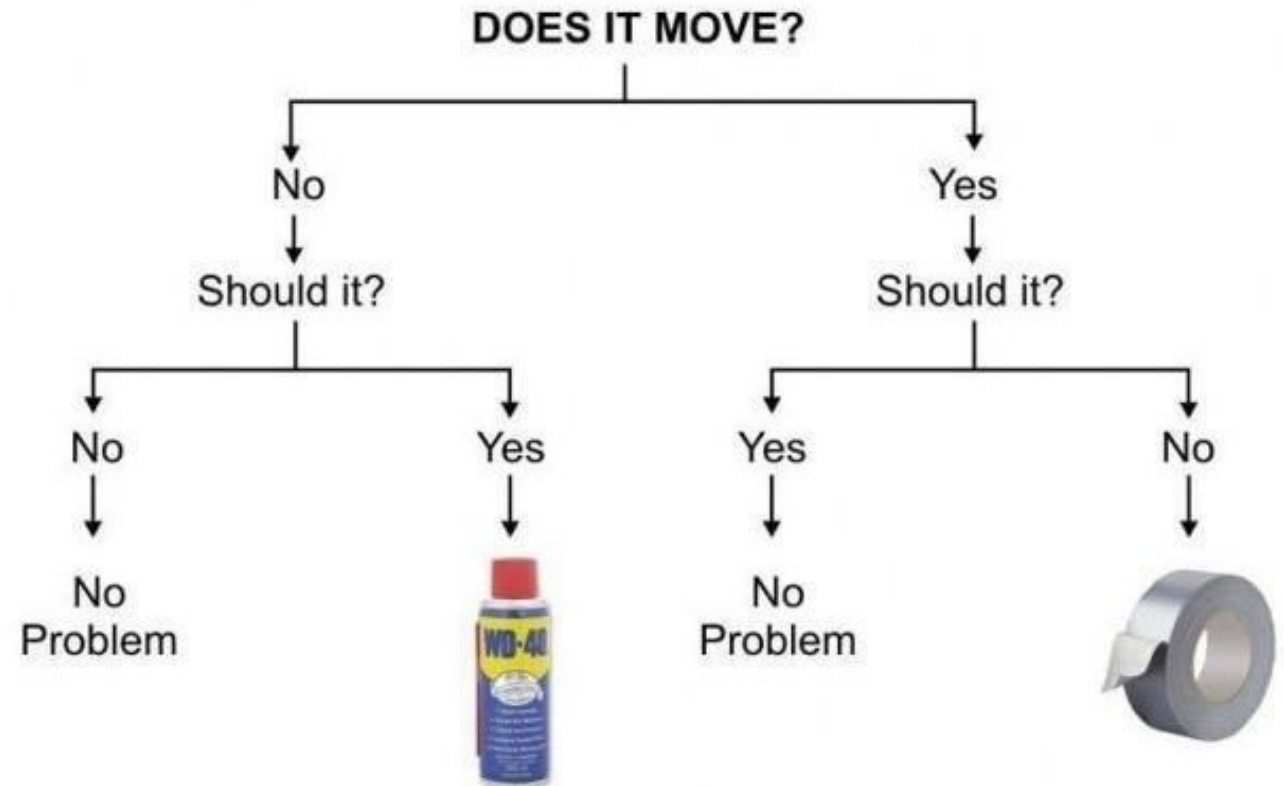


Maschinelles Lernen - Einführung

– Hans Friedrich Witschel, Andreas Martin



Die Dozenten...



Frieder

hansfriedrich.witschel@fhnw.ch



Manuel

Manuel.renold@fhnw.ch

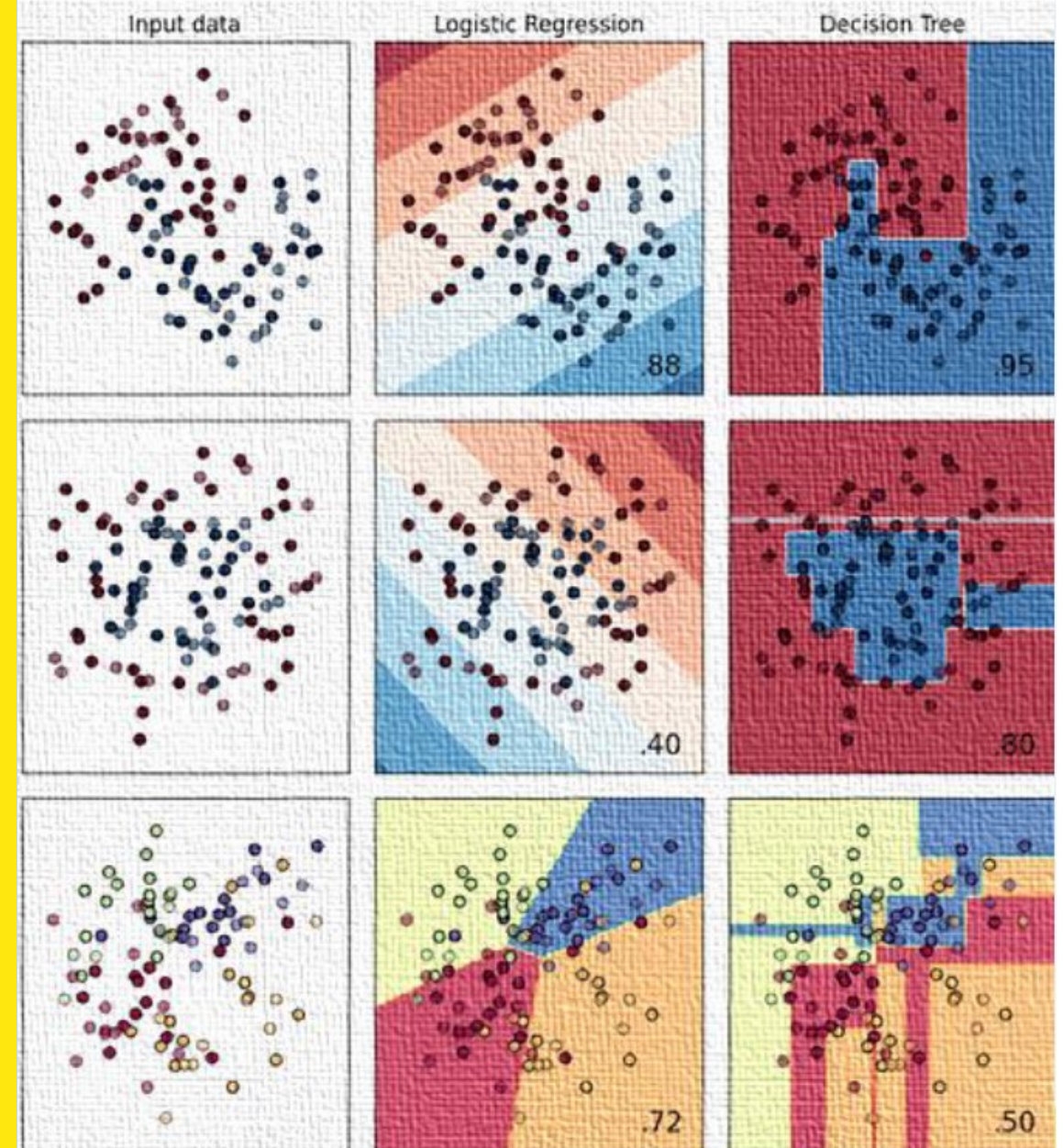
Semesterprogramm: Maschinelles Lernen

Semester		25 FS	Dozierende	Hans Friedrich Witschel und Manuel Renold	Durchführungsort	Olten
Nr.	Datum	Dozent	Themen			
1	19./21.02.	Frieder	Wiederholung ML, CRISP-DM, Formalisierung von Aufgaben des überwachten ML			
2	26./28.02.	Frieder	Datentransformation (für überwachtes Lernen)			
3	05./07.03.	Frieder	Modellierung (Klassifikation)			
4	12./14.03.	Frieder	Modellierung (Regression)			
5	19./21.03.	Frieder	Evaluationsprozeduren und Evaluationsmetriken (insbesondere Kostenmatrix)			
6	26./28.03.	Frieder	Typische Probleme			
7	02./04.04.	Frieder / Manuel	Coaching			
8	09./11.04.	Frieder	Typische Probleme 2			
9	16./18.04.	Kein Unterricht				
10	23./25.04.	Frieder	Unüberwachtes Lernen: Clustering			
11	30.04./02.05.	Kein Unterricht				
12	07./09.05.	Manuel	Reinforcement Learning			
13	14./16.05.	Frieder / Manuel	Coaching 2			
14	21./23.05.	Alle	Assignment-Präsentationen			
15	28./30.05.	Frieder	Explainable Machine Learning			
16	04./06.06.	Frieder	Ethische Fragestellungen, Wrap-up, Klausurvorbereitung			

Assessment

- 60% Klausur
- 40% semesterbegleitende Projektarbeit
 - a. Bitte macht Vierergruppen!
 - b. Datensätze und Aufgabenstellung folgen bald!

Was ist maschinelles Lernen?



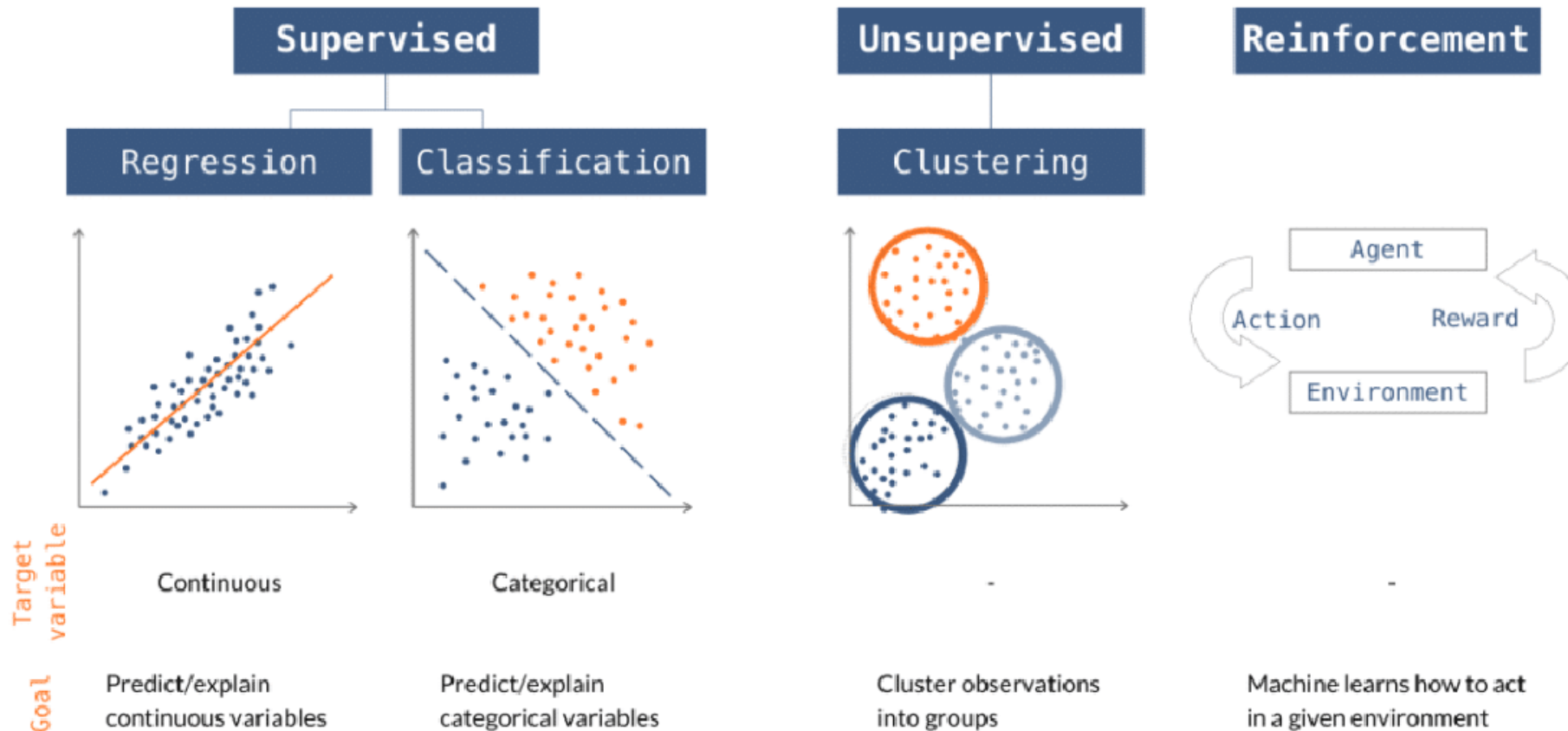
Muster



Interessante Muster

- **valide**: das Muster ist allgemein anwendbar, d.h. kann bei neuen Daten beobachtet werden
- **neu**: das Muster ist nicht bekannt und/oder trivial; es erweitert vorhandenes Wissen oder widerspricht ihm sogar
- **nützlich**: Es ist möglich, einen wirtschaftlichen Nutzen aus dem Modell zu ziehen
- **nachvollziehbar**: das Muster ist plausibel

Formen des maschinellen Lernens



Überwachtes Lernen – Business-Anwendungen

Klassifikation

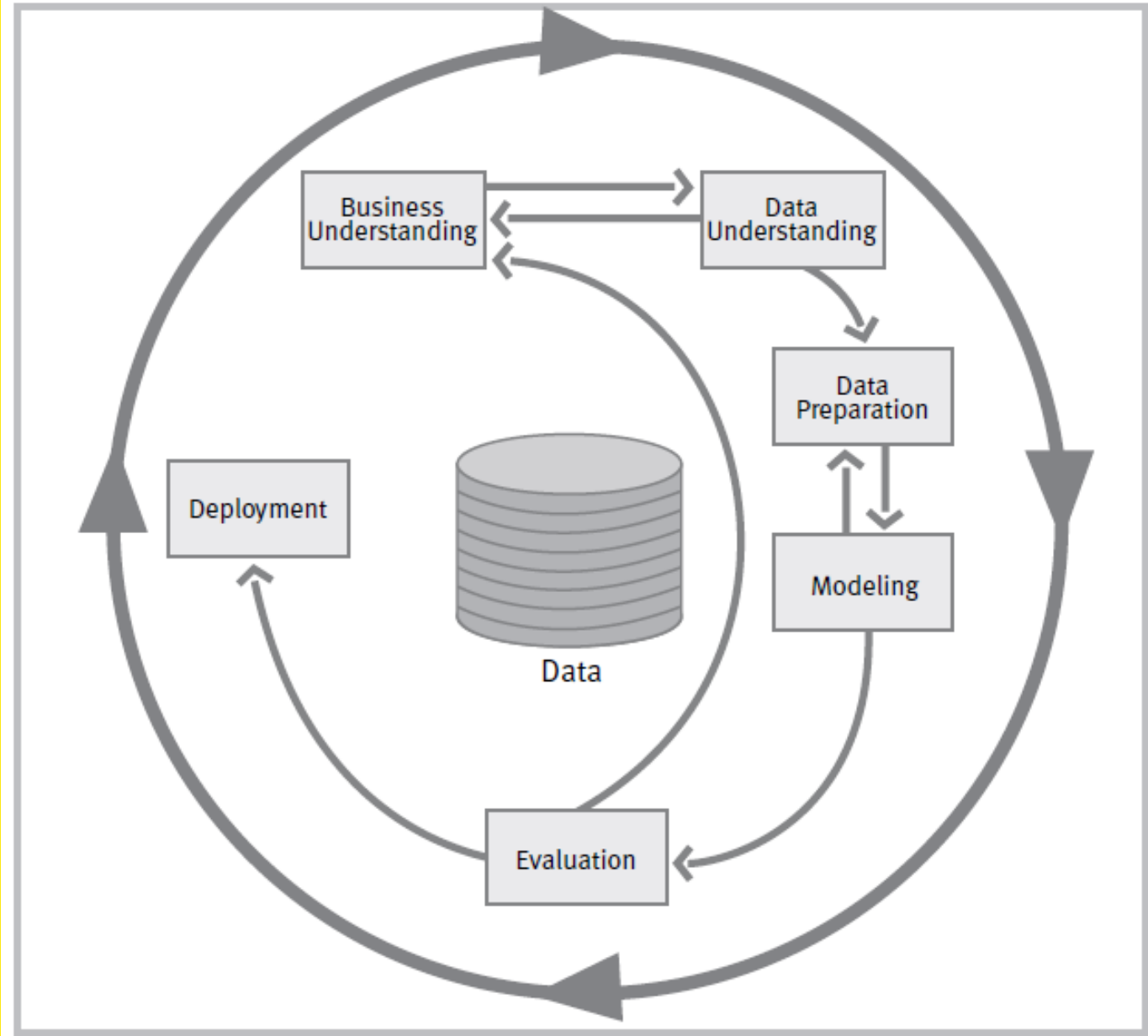
- Kreditvergabe
- Betrugserkennung
- Targeted Marketing
- Vorhersage
Kundenabwanderung («Churn»)
- Filtern von Inhalten (z.B. Emails)



Regression

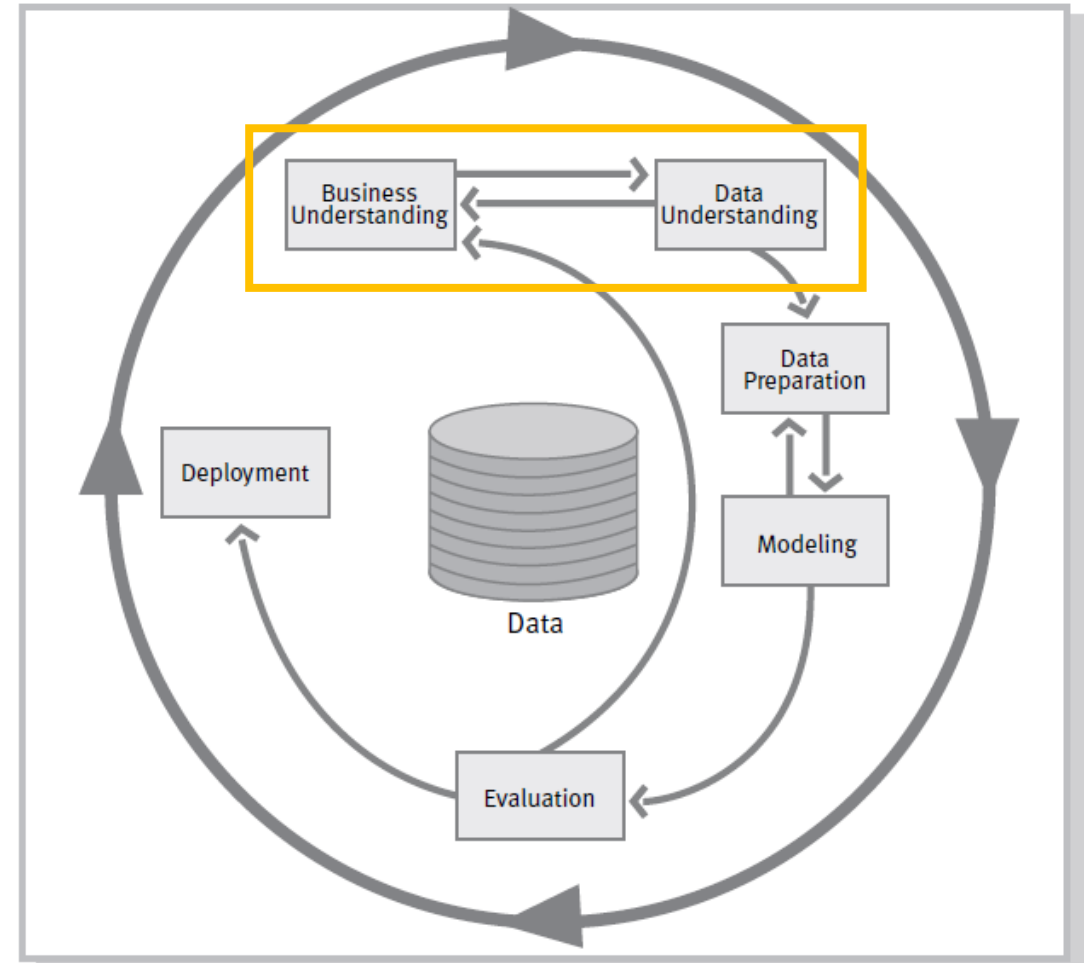
- Vorhersage der Nachfrage
- Preissensitivität
- Vorhersage von Kosten / Risiken
/ Erträgen
- Erfolgsvorhersage (z.B. für
Marketingkampagnen)

Der CRISP-DM-Zyklus



CRISP-DM: Problemformalisierung

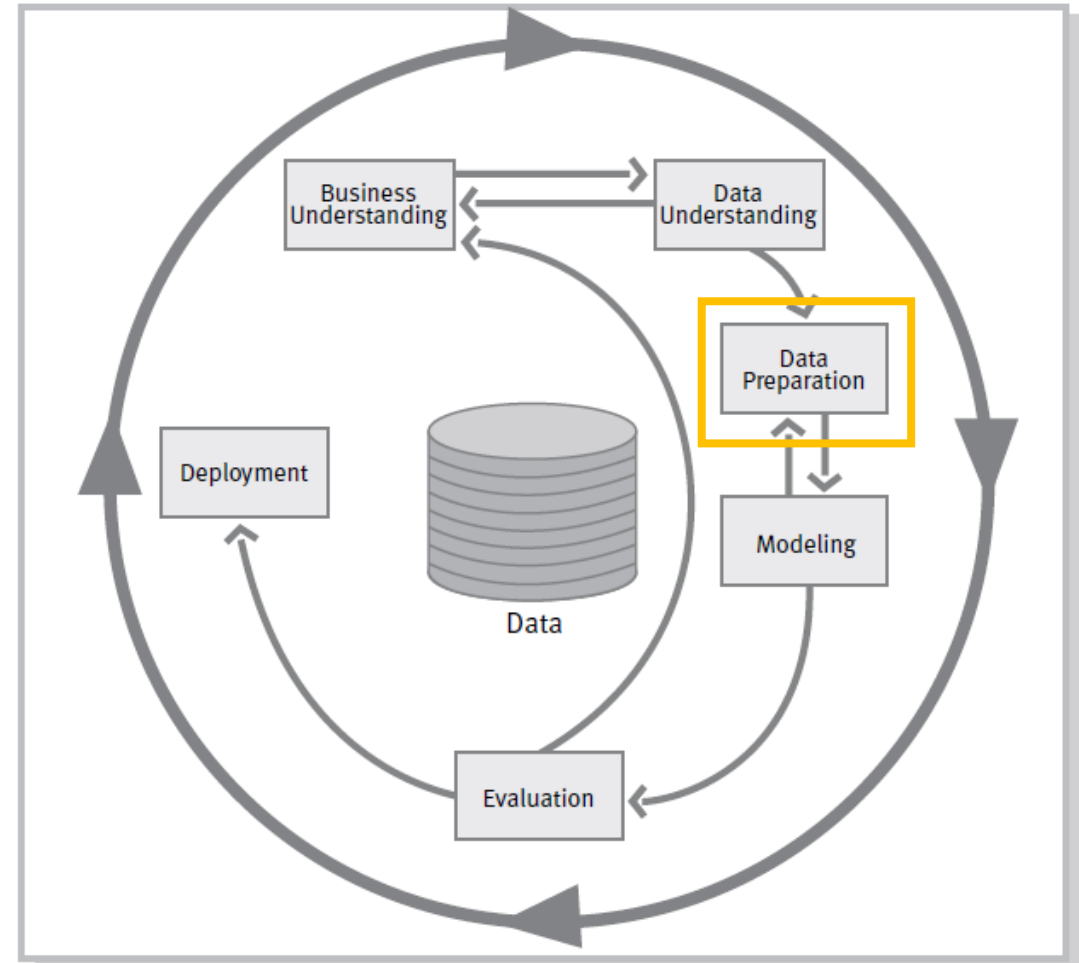
Lernziel: eine formalisierte
Problembeschreibung ableiten
(aus Geschäfts- und
Datenverständnis)



CRISP-DM: Daten vorbereiten

Lernziele:

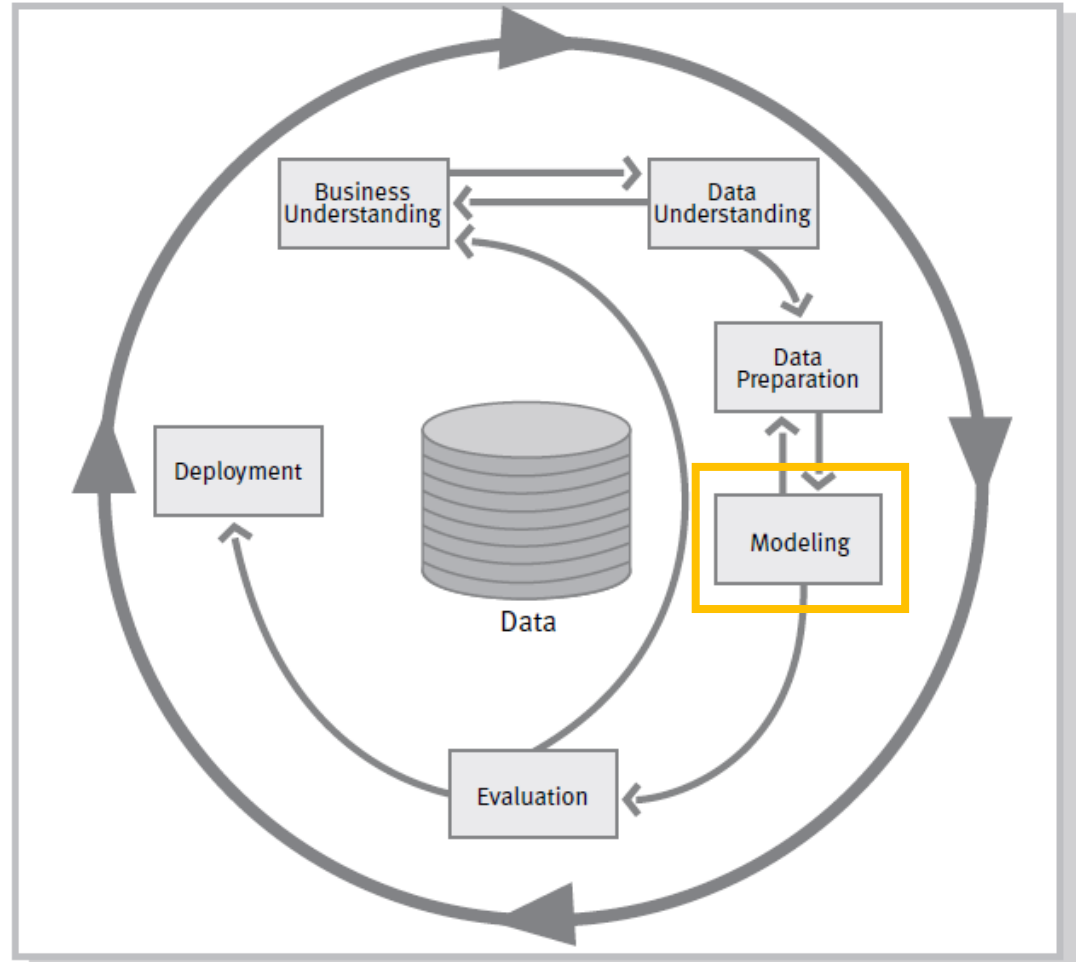
- Daten für ML anhand der Formalisierung aufbereiten
- spezielle Schritte verstehen: Handling fehlender Werte, Skalierung, Diskretisierung, ...



CRISP-DM: Modell bauen

Lernziele:

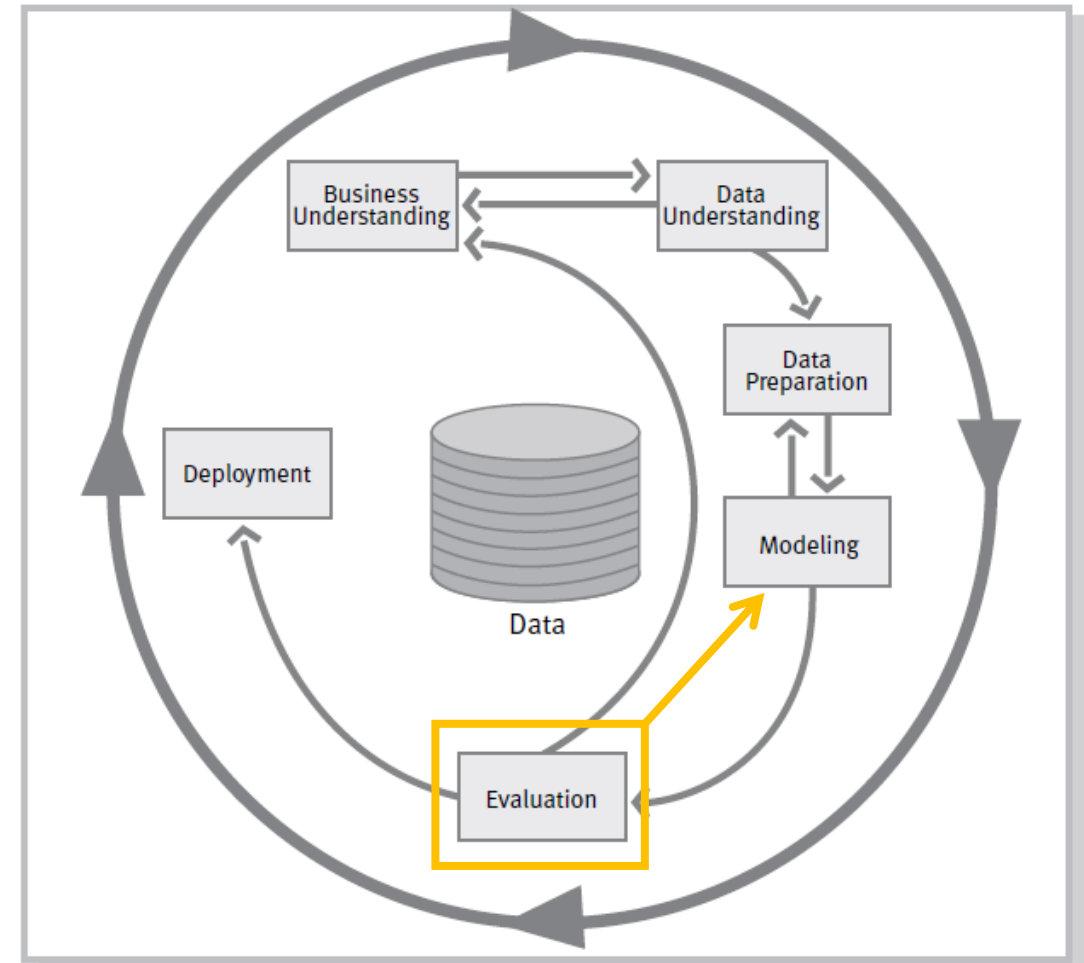
- Vor- und Nachteile und Besonderheiten von DM-Verfahren verstehen
- die richtigen Verfahren auswählen können



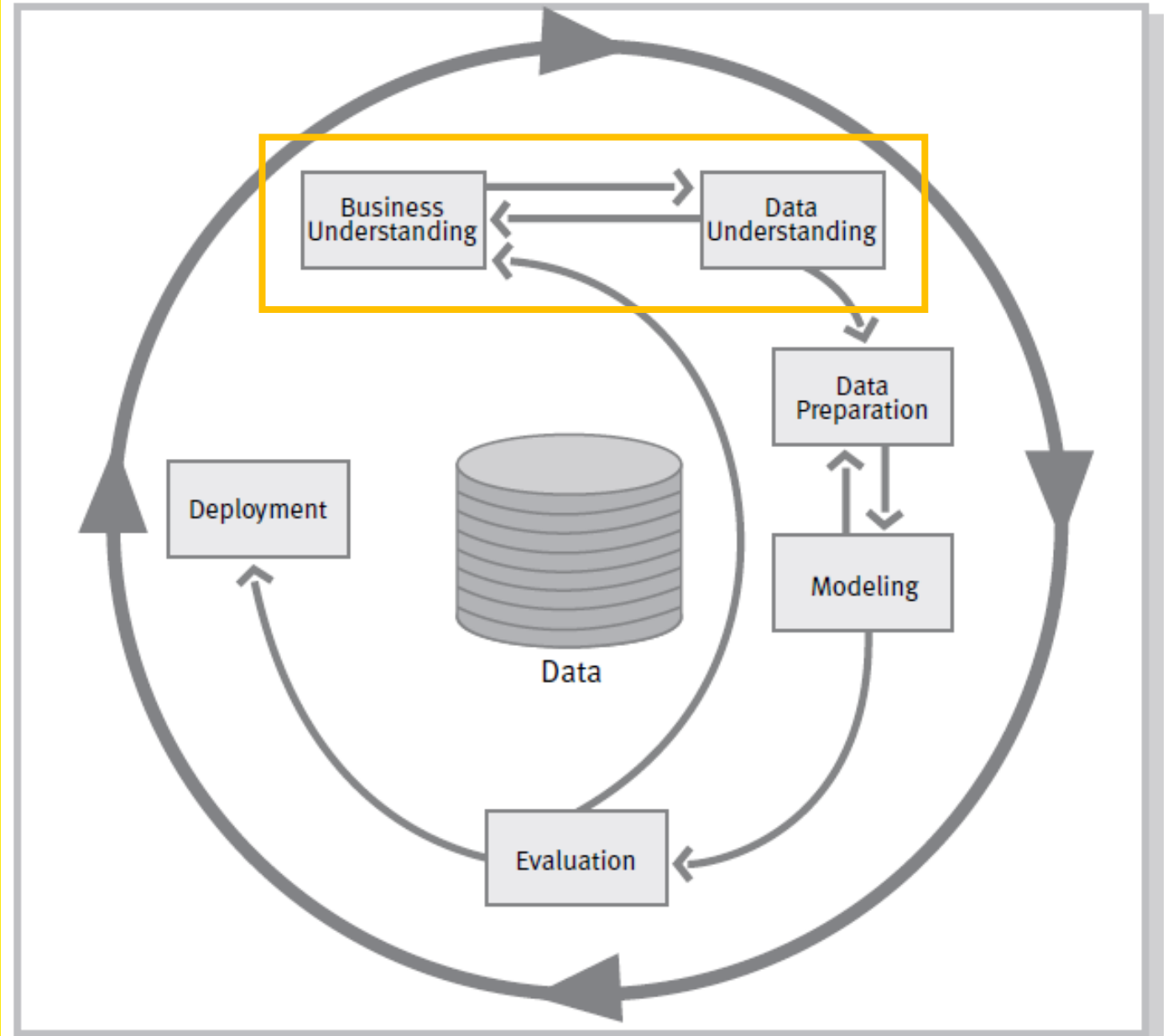
CRISP-DM: Modell evaluieren

Lernziele:

- Ergebnisse beurteilen
- Mögliche Probleme verstehen und vermeiden



Problem formalisieren



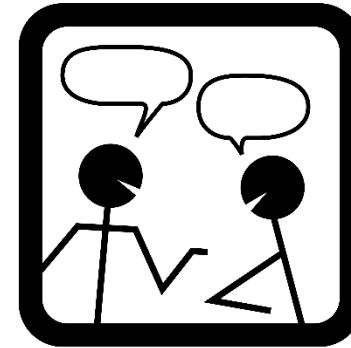
Zielformat

x_1	x_2	x_3	x_4	x_5	Y
0.1	5.2	2.1	0.2	2.9	A
0.2	4.9	1.9	0.1	3.1	A
5.1	4.8	1.9	1.2	6.1	B

← Header

Formalisierungsaufgabe

- Ausgangslage: «wir würden gern... vorhersagen»



- Was wir festlegen müssen:
 - a. Was sind die **Instanzen**, für die etwas vorhergesagt wird? → Zeilen
 - b. Was genau soll vorhergesagt werden (**Klassenattribut**)? → spezielle Spalte
 - c. Welche Merkmale (Features / **Attribute**) können dabei helfen? → andere Spalten

Formalisierungsangabe: Beispiel

- Jedes Jahr im Herbst bietet Swiss Bikes einen besonderen Service: «Machen Sie Ihr Bike fit für den Winter» (Bremsen, Reifen etc.)
- Im November verschickt Swiss Bikes ein Mailing zu diesem Angebot an alle Bestandskunden, die ihr Fahrrad mindestens einmal zur Reparatur gebracht haben. Das Mailing beinhaltet einen Gutschein für 10% Rabatt auf den Winter-Check.
- Swiss Bikes weiss, welche Kunden ihre Gutscheine in den letzten Jahren eingelöst haben



«Wir würden gern... dieses Jahr gezielt nur diejenigen Kunden anschreiben, die auch wirklich Interesse am Winter-Check haben!»

Formalisierungsaufgabe: Beispiel (2)

- **Instanzen:** Kunden
- **Klassenattribut:** Gutschein eingelöst? («Response»)
- **Attribute:** Datum letzte Reparatur, Anzahl Reparaturen (letzte 3 Jahre), Region, zuletzt gekauftes Fahrrad (Kategorie), ...

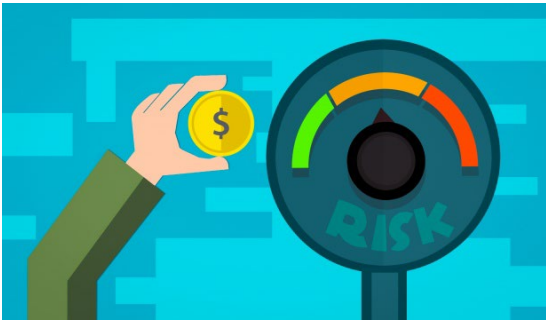
Formalisierungsaufgabe: Beispiel (3)

Kunden-ID	Datum letzte Reparatur	Anzahl Reparaturen (letzte 3 Jahre)	Region	Letztgekauftes Fahrrad (Kategorie)	Response
213232	May 24	1	city	Trekking	No
123244	Feb 25	5	city	Racing	Yes
546657	Nov 23	0	city	Single Speed	No
764566	Oct 21	0	rural	Single Speed	No
453232	Dec 22	0	rural	Mountain	No
785423	Apr 24	3	rural	Trekking	Yes
132567	Apr 24	2	city	Mountain	Yes
456467	Apr 23	1	rural	Mountain	No
342256	Nov 21	0	city	Single Speed	No
798888	Nov 24	4	rural	Trekking	Yes

Feature Engineering

- Woher weiss ich, was sich als Attribut eignet?
 - Kandidaten finden: Intuition («Business Understanding»)!
 - wichtig: alles, was helfen könnte, ist erlaubt!
 - dann: z.B. Korrelation mit Klassenattribut prüfen
 - oder einfach ausprobieren...
- Was eignet sich (sicher) nicht?
 - Attribute, die für jede Instanz einen anderen Wert haben (z.B. IDs)
 - Attribute, die immer den gleichen Wert haben
 - Attribute, deren Wert zum Zeitpunkt der Vorhersage nicht bekannt ist

Übung: Formalisierung Klassifikation



Bank: «könnte nicht die AI die Entscheidungen über Kreditvergaben übernehmen?»»



Mobilfunkanbieter: «wenn wir vorher wüssten, welche Kunden ihren Vertrag nicht verlängern, könnten wir sie mit bestimmten Angeboten halten...!»»



Leiterin Digitales Marketing:
«ich schaue immer, wie gut meine Kampagnen laufen. Wenn es z.B. nach der Hälfte der Laufzeit noch hakt, schiesse ich Budget nach. Aber manchmal ist das vielleicht gar nicht nötig? Ich wüsste gern schon zu dem Zeitpunkt, ob wir am Ende das gesteckte Ziel erreichen werden...!»»

Übung: Formalisierung Regression



Leiter Skigebiet: «ich plane jeden Tag die Ressourcen (Servicekräfte), die ich am nächsten Tag brauche. Das hängt natürlich davon ab, wieviele Leute am nächsten Tag zum Skifahren kommen...»



Private Krankenversicherung: «kann nicht die KI die Festlegung der Versicherungsbeiträge übernehmen?»