

ReComMeND

Prof. Dr. Robert Knappe, HWR Berlin

Nicki Kämpf, BHT

66. Tagung des AK Berlin-Brandenburg / 11. November 2021



Lufthansa Systems



**Lufthansa
Industry Solutions**

Projektpartner:

Gefördert durch:



Ziele von ReComMeND

Revenue **C**ontrolling mit **m**athematischer
Modellierung auf Basis **e**mpirischer Daten
für **N**ahverkehrs-**D**ienstleistungen

1. Prognosemethodik unter Einsatz mathematischer Modelle optimieren
2. Treibereinflüsse besser verstehen
3. Predictive Analytics in die Controllingpraxis integrieren
4. Verallgemeinerbare Erkenntnisse für (Nah-)Verkehrsdienstleistungen ableiten

Projektpartner:



Lufthansa Systems



**Lufthansa
Industry Solutions**

Laufzeit: 1. 10. 2019 – 31. 12. 2022, Budget ca. 300 T€

<https://www.ifaf-berlin.de/projekte/recommend/>

Gliederung

1. Einblicke in die Herausforderungen des Transformationsprozesses zum datengetriebenen Unternehmen
2. Überblick zu mathematisch-statistischen Verfahren & Technologien der Künstlichen Intelligenz
3. Modernisierung des Rollenverständnisses für Controlling und Management
4. Diskussion

1. Einblicke in die Herausforderungen des Transformationsprozesses zum datengetriebenen Unternehmen

Quelle: Die Herausforderung "Digitalisierung" - Was den digitalen Wandel ausmacht und wie man ihn meistert
Link: <https://www.youtube.com/watch?v=sexwoHDhkrI> (0:00-1:15)

„Daten sind das Öl des 21. Jahrhunderts, und Datenanalyse der Verbrennungsmotor.“

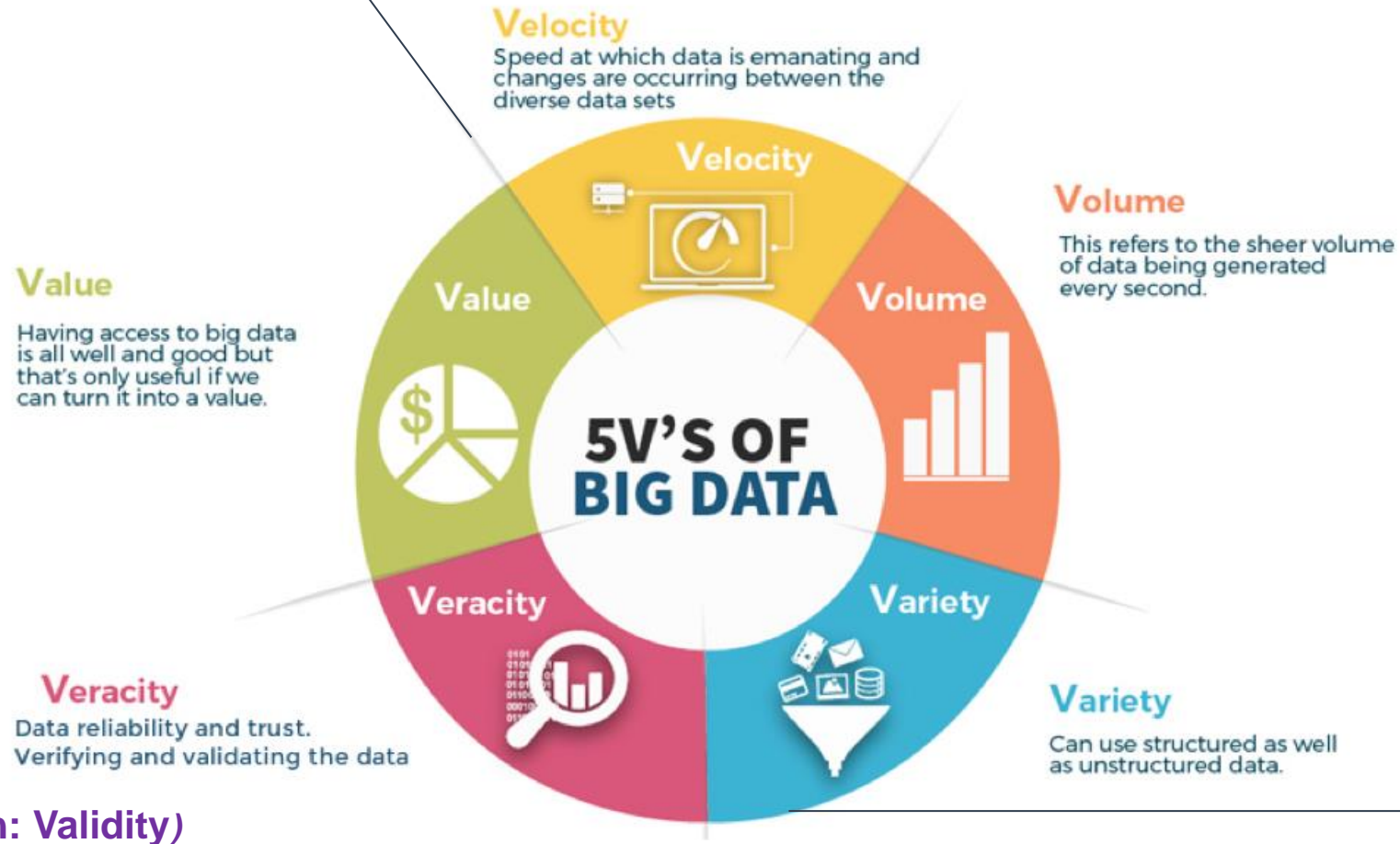
Peter Sondergaard, Head of Research & Advisory, Gartner Inc.

Definition: Datengetriebenes Unternehmen

- Datengetriebenes Unternehmen zu sein, meint die stetige Suche nach verbesserten Optionen der Entscheidungsunterstützung durch Datennutzung (Beispiel: Forecast)
- Dies impliziert auch strategische Überlegungen:
 - ändert sich das Geschäftsmodell im Zuge der Digitalisierung?
 - Muss das Datenmanagement entsprechend umgestellt werden?

*Vgl. BI oder DIE Data Talk – Was ist eine Data Driven Company?
Link: <https://www.youtube.com/watch?v=v3-dtnbWqKM> (1:55-2:30)*

Definition von Grundbegriffen: Big Data



**Traditionelle
Definition: 3 V's**

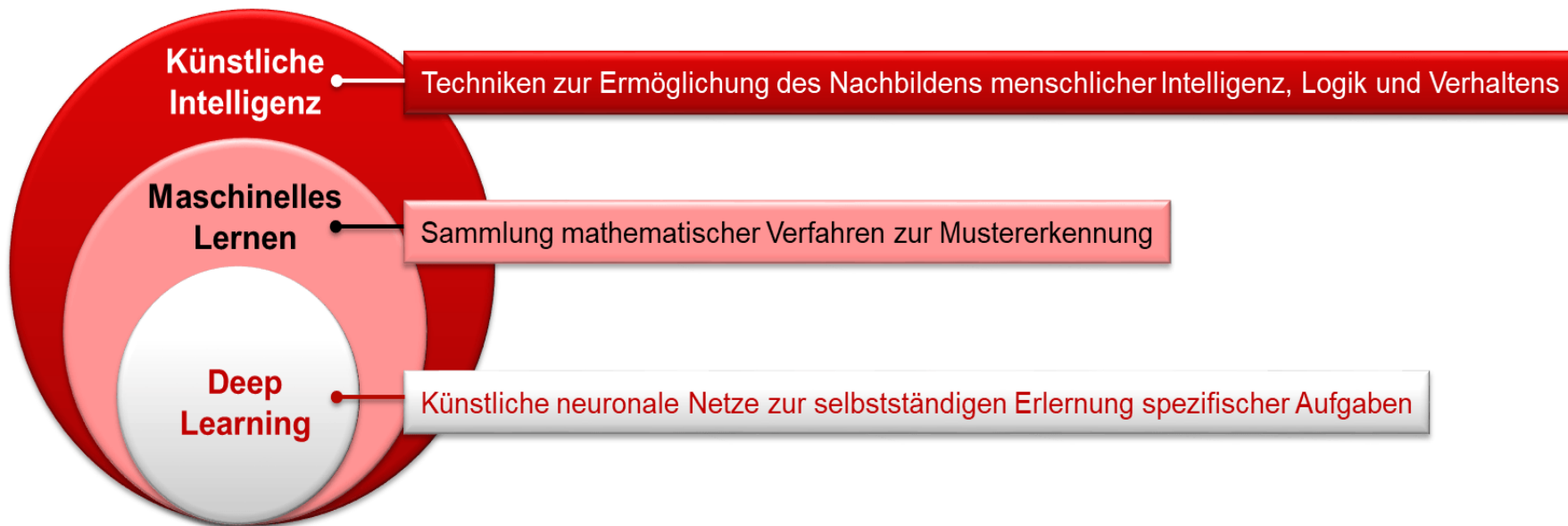
**Jüngere Definitionen
erweitert um 2-3
weitere V's**

(auch: Validity)

Quelle: <https://www.techentice.com/the-data-veracity-big-data/> Vgl. Bauer et al. (2018), S.131f.

Definition von Grundbegriffen: Maschinelles Lernen

„Machine learning is the field of study that gives computers the ability to learn without being explicit programmed.“ – *Arthur L. Samuel, 1959*



→ Automatisierte Erstellung eines Modells auf Basis von Erfahrungen (= Daten)
→ Computer schreibt eigenes Programm

Quelle: <https://kompetenzzentrum-saarbruecken.digital/technologieradar-maschinelles-lernen/>

Vgl. Lanquillon (2019), S. 91

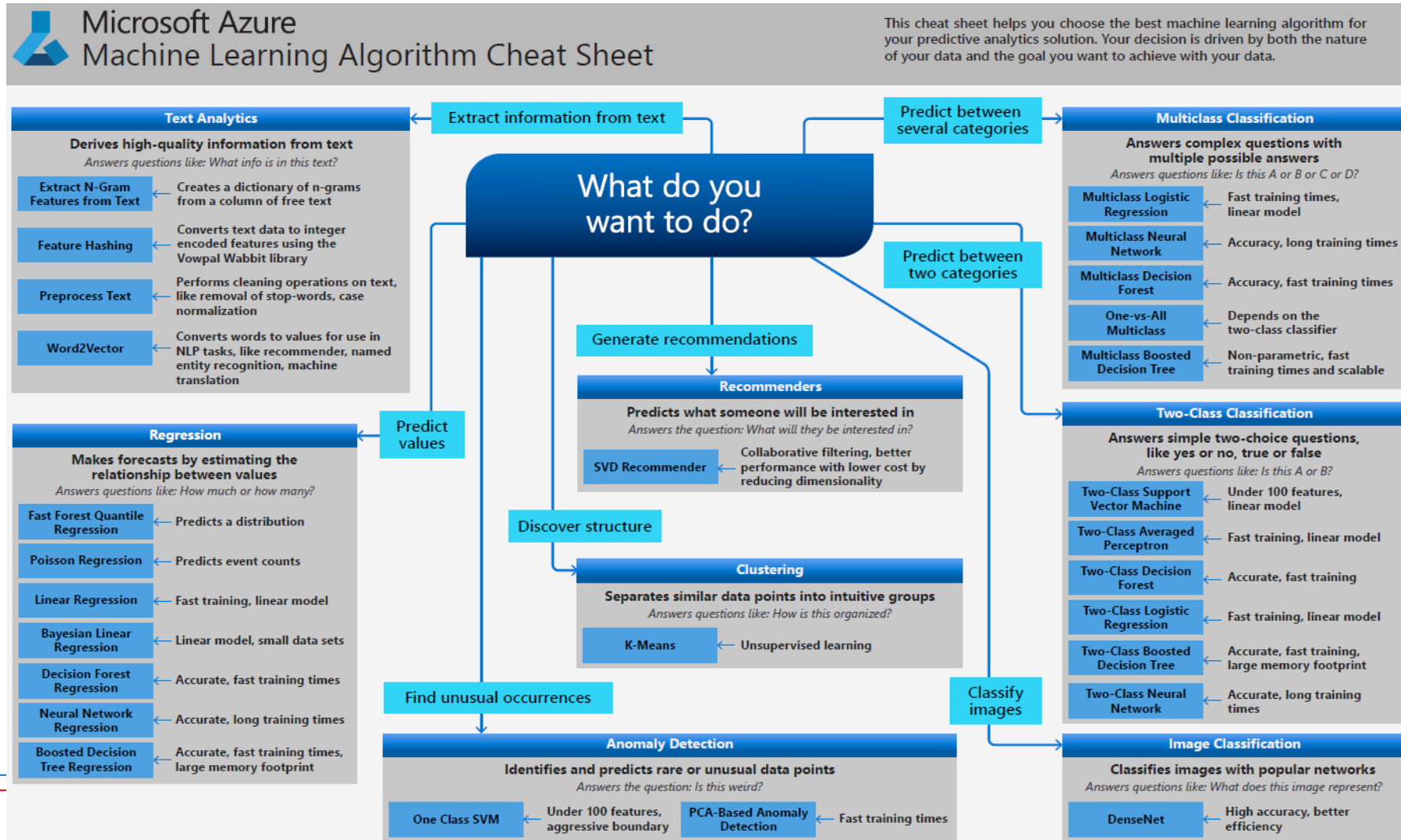
Definition von Grundbegriffen der KI und Anwendungsmöglichkeiten

*Quelle: Vortrag von
Prof. Sabina
Jeschke auf dem
45. Congress der
Controller,
26./27.04.2021*

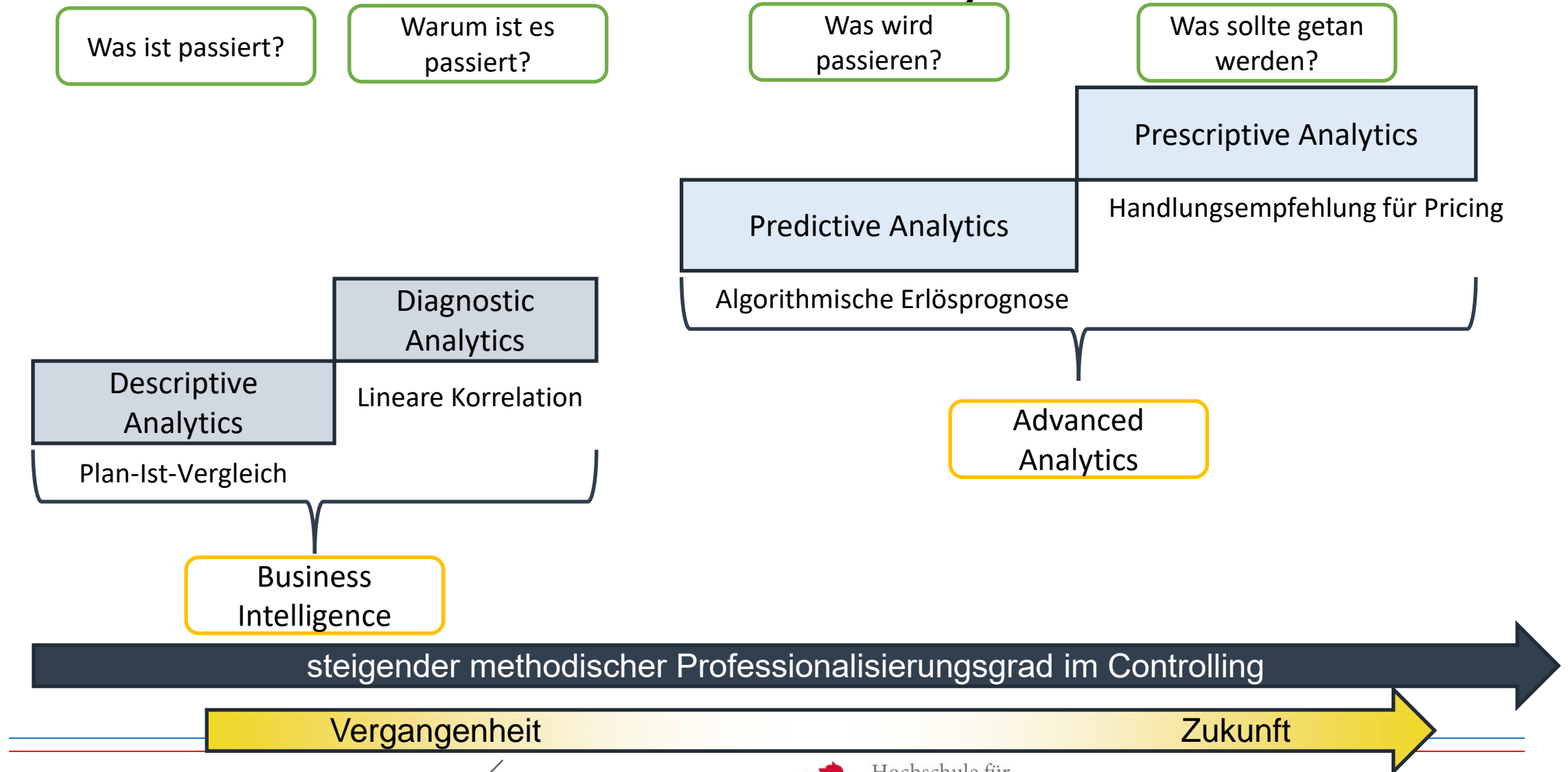
Anwendungsfelder von Data Science und Künstlicher Intelligenz

- *Clustering*: Gruppieren ähnlicher Strukturen, z. B. Kunden, Produkte
- *Classification*: Vorhersage von Gruppenzugehörigkeiten u. Verhalten, z. B. Wahrscheinlichkeiten des Wiederholungskaufs / der Kündigung bestimmen, Bilderkennung
- *Regression*: Vorhersage numerischer Werte, z. B. Absatzmenge, Preisschätzung
- *Time Series Forecast*: Periodische Muster erkennen und fortschreiben, z. B. Umsatzplanung, Mitarbeiterereinsatzplanung, Mitarbeiterfluktuation
- *Anomaly Detection*: Fehler und Manipulationen erkennen, z. B. Qualitätsmanagement, menschliche Fehler, Tippfehler und Datenfehler in Formularen erkennen
- *Feature Importance*: Einflüsse, Zusammenhänge (Signifikanzen) und deren Stärke erkennen, z. B. Zusammenhang Werbung und Umsatz (Marketingcontrolling), Temperatur und Umsatz etc.
- *Survival Analysis*: Schätzen von Überlebenswahrscheinlichkeiten, z. B. im Vertrieb Kündigungs- bzw. Kaufabschlusswahrscheinlichkeit, im Bereich Finanzen: Forderungsausfall
- *Association Rule Learning*: auffällige Zusammenhänge erkennen, z. B. zusammen gekaufte Produkte (Onlinemarketing / Empfehlungsmarketing), Lager-/Versandoptimierung
- *Natural Language Processing (NLP)*: Erkennung von gesprochener und geschriebener Sprache, z. B. Sprachsteuerung, automatische Interpretation von Texten, Gesetzen, Verträgen, Fallbearbeitung, Stimmungserkennung (Tonfall bzw. Wortlaut) im Reklamationswesen

Beispiel für Algorithmen-Bibliothek aus Microsoft Azure zur Verwendung im Controlling



Von der BI zu Advanced Analytics

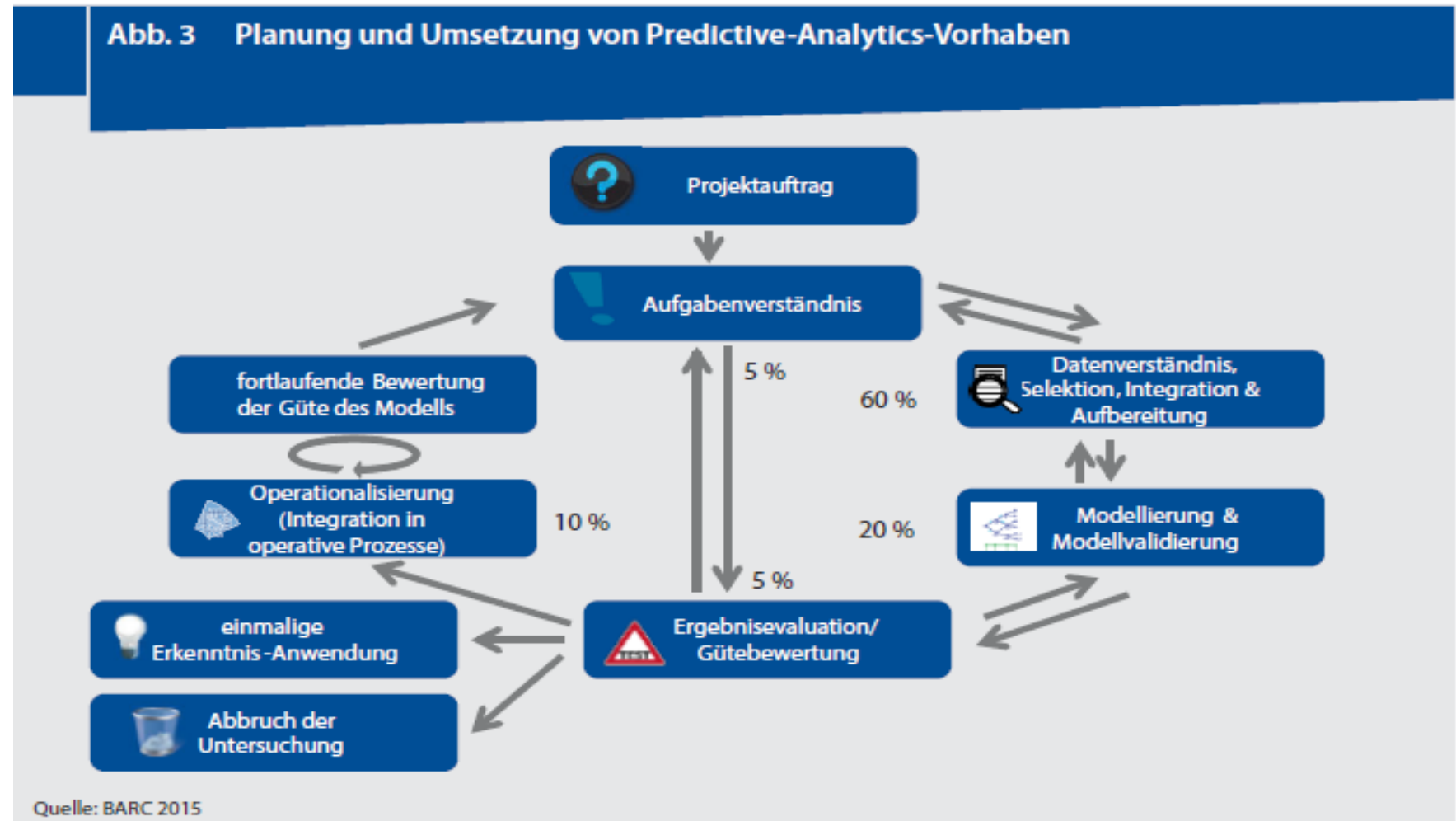


Typische Projektstruktur Implementierung von Predictive Analytics

Mischung aus:

- KDD-Methodik (Knowledge Discovery in Databases)
- CRISP DM (Cross Industry Standard Process for Data Mining) und
- Erfahrungswissen

Quelle: Iffert, Lars: Predictive Analytics richtig einsetzen. In: Controlling & Management Review, Sonderheft 1 / 2016, S. 20

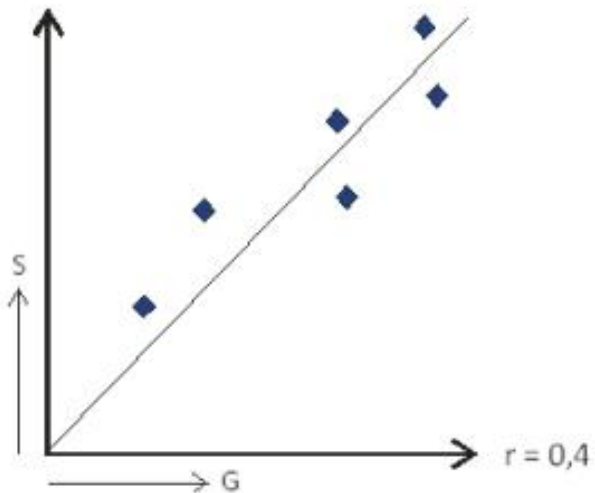


2. Überblick zu mathematisch-statistischen Verfahren & Technologien der künstlichen Intelligenz

Maschinelles Lernen – Welche Aufgaben können gelernt werden?

Regression

Geburtenrate / Storchendichte



S = Storchendichte (V_{a1})

G = Geburtenrate (V_{a2})

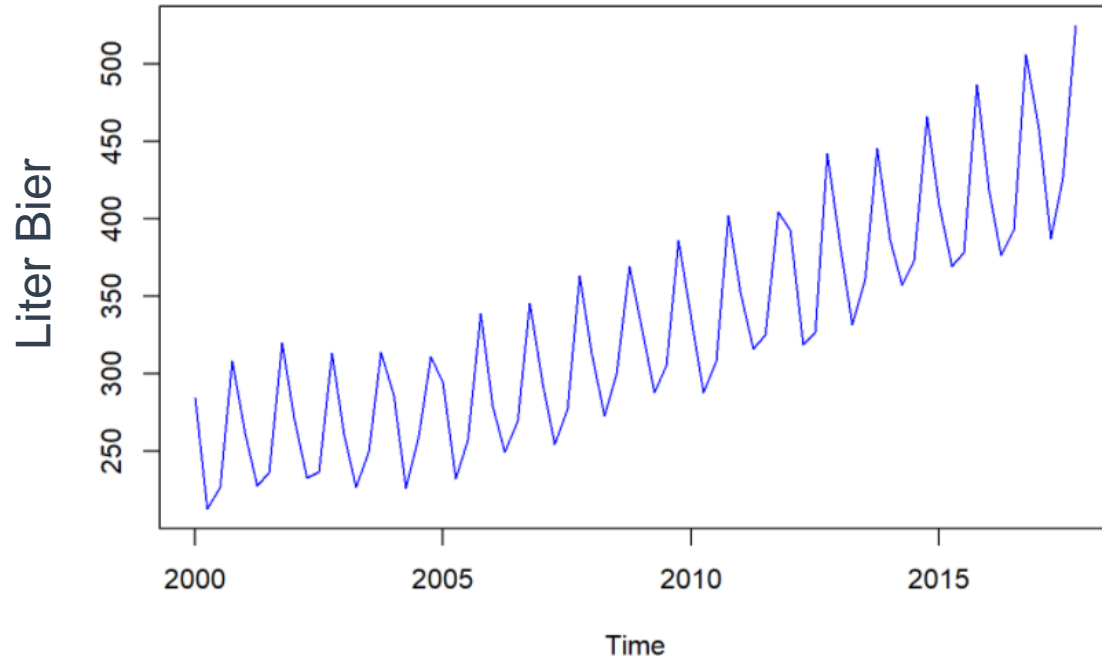
- Prognose von kontinuierlichen Zahlen
- Zusammenhang zwischen:
 - Bildungsjahre und Einkommen
 - Geburtenrate und Storchendichte
- Korrelation vs. Kausalität:
 - Expertenmeinung!

Quelle: <https://www.univie.ac.at/sowi-online/esowi/cp/methodologiesowi/methodologiesowi-38.html>

Maschinelles Lernen – Welche Aufgaben können gelernt werden?

Regression

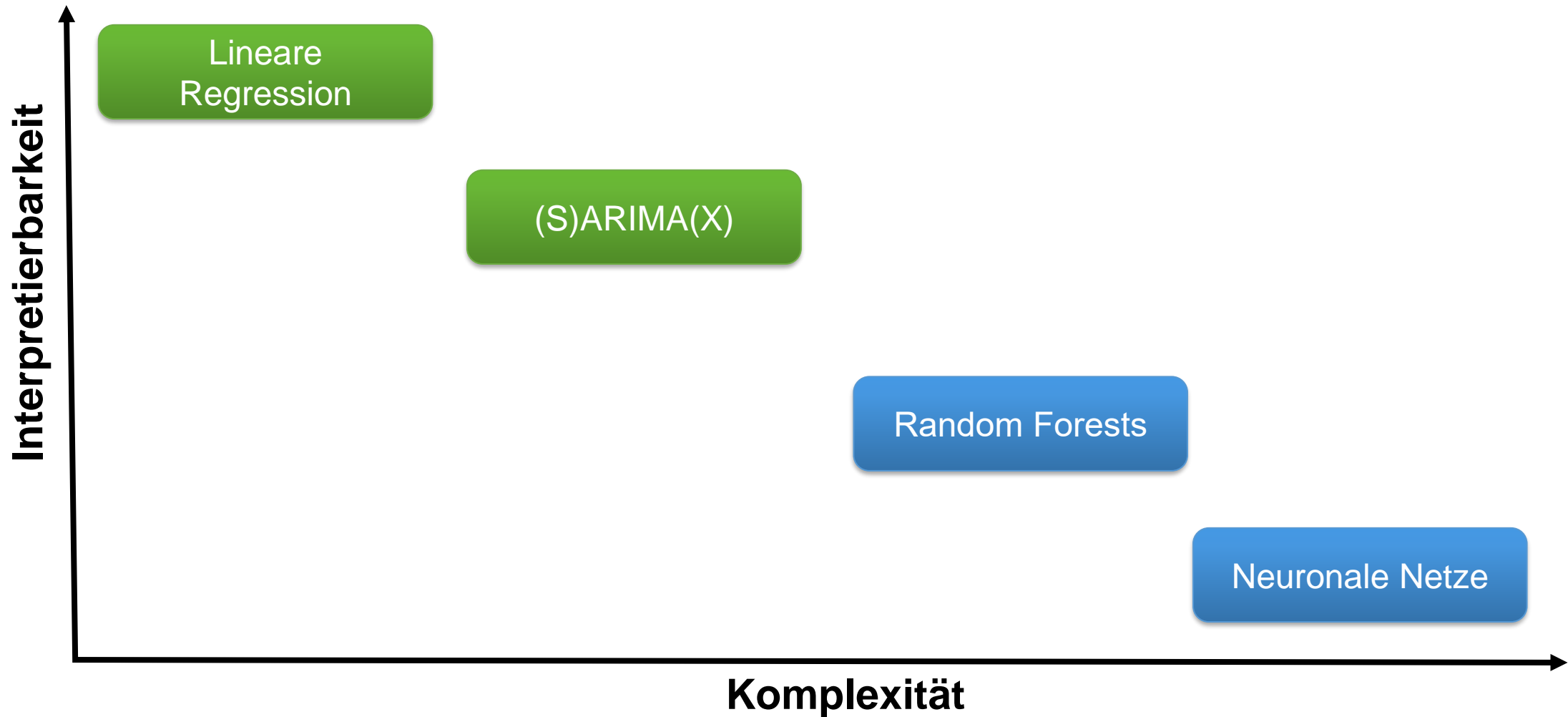
Beer sales time series data between 2000 and 2017



Quelle: <https://rpubs.com/amitkayal/369361>

- Untergebiet: Zeitreihenanalyse
 - Abhängig von historischen Werten und Saisonalitäten
 - Beispiele: Umsätze, Verkaufszahlen, Besucher, BIP usw.

Methodenübersicht



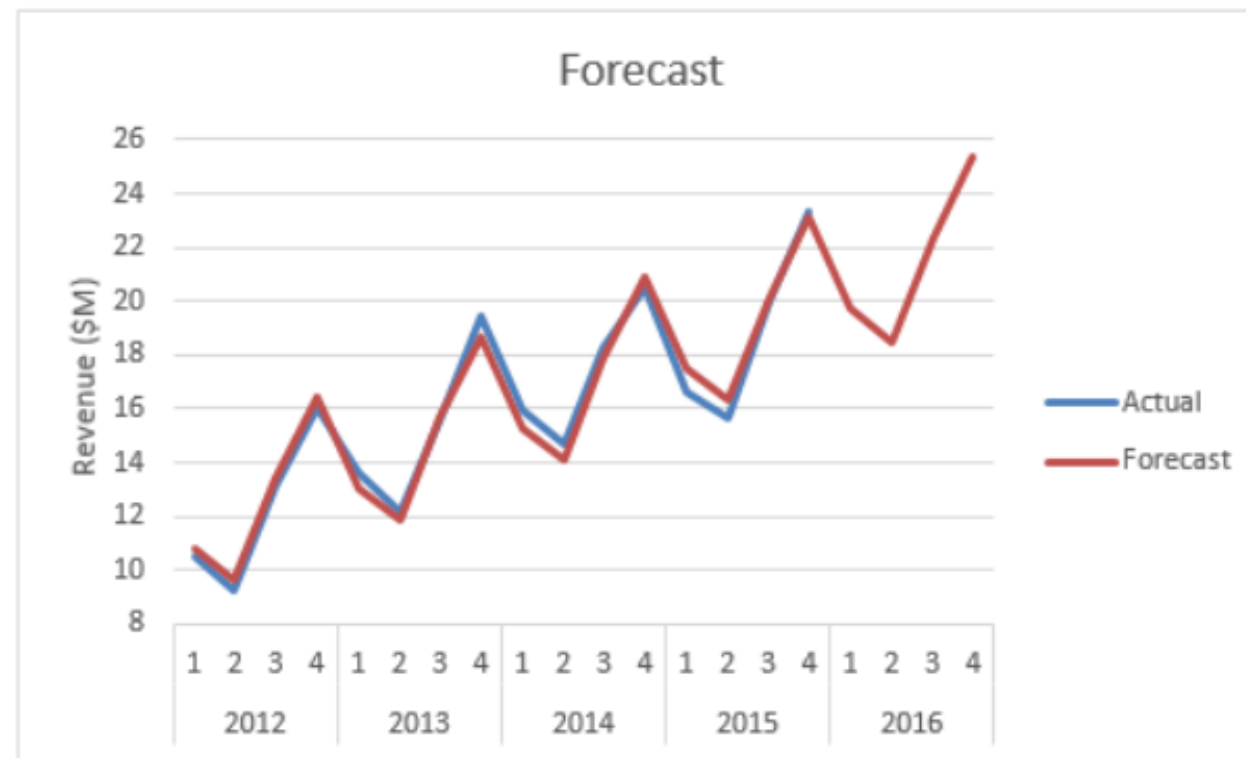
Vergleich der Methoden - Gemeinsamkeiten

- Voraussetzungen für alle Methoden:
 - Aufspaltung in Trainings- und Testdaten
 - je mehr Daten, desto besser
 - Treiber müssen vorher identifiziert werden
 - Trainings- und Testdaten (inkl. Treiber) müssen im selben Format vorliegen
- Trainieren der Methoden:
 - Parameter der Methoden werden anhand von Trainingsdaten gefunden
↔ Hyperparameter müssen vorab festgelegt werden
 - Testdaten werden nicht für das Training benutzt
 - Testdaten dienen nur dazu, Methoden auf neuen Daten zu testen

Vergleich der Methoden – Lineare Regression

Beispiel: Prognose von Umsatz eines Unternehmens in Millionen US Dollar

$$\text{Umsatz} = \text{Konstante} + \beta_1 t + \beta_2 Q1 + \beta_3 Q2 + \beta_4 Q3$$



Quelle: Charles Zaiontz

Vergleich der Methoden – Lineare Regression

	A	B	C	D	E	F	G
2							
3	Year	Quarter	Rev (\$M)	t	Q1	Q2	Q3
4	2012	1	10.5	1	1	0	0
5		2	9.2	2	0	1	0
6		3	13.1	3	0	0	1
7		4	16.0	4	0	0	0
8	2013	1	13.6	5	1	0	0
9		2	12.2	6	0	1	0
10		3	15.6	7	0	0	1
11		4	19.4	8	0	0	0
12	2014	1	15.9	9	1	0	0
13		2	14.7	10	0	1	0
14		3	18.3	11	0	0	1
15		4	20.5	12	0	0	0
16	2015	1	16.6	13	1	0	0
17		2	15.7	14	0	1	0
18		3	20.0	15	0	0	1
19		4	23.3	16	0	0	0
20	2016	1		17	1	0	0
21		2		18	0	1	0
22		3		19	0	0	1
23		4		20	0	0	0

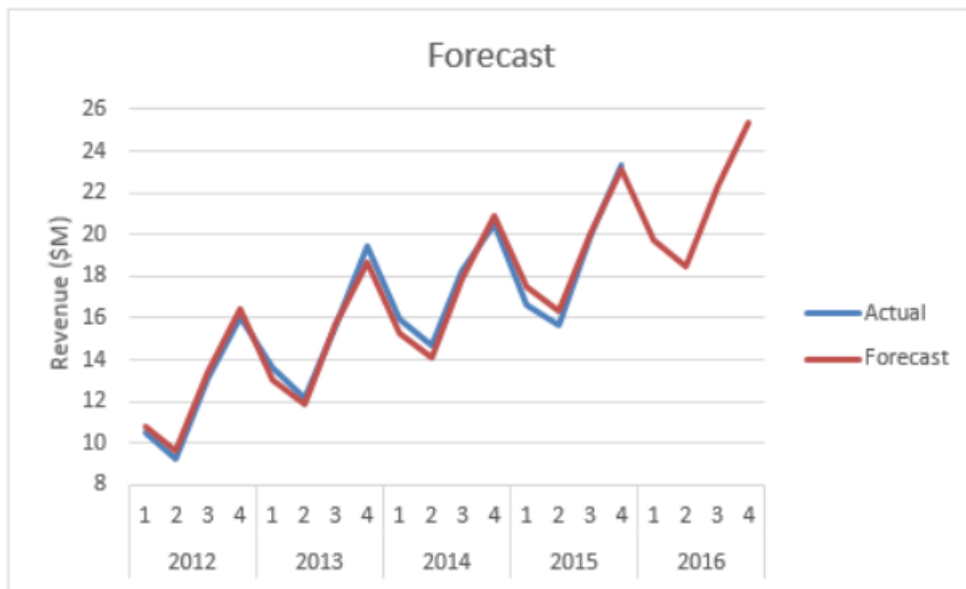
- Erstellung von Dummyvariablen erfolgt automatisch in Streamlit
- lediglich Auswahl der Treiber im Menü notwendig

- Vorheriges Erkennen und Erstellen des Treibers t notwendig

Quelle: Charles Zaiontz

Vergleich der Methoden – Lineare Regression

$$\text{Umsatz} = \text{Konstante} + \beta_1 t + \beta_2 Q1 + \beta_3 Q2 + \beta_4 Q3$$



Quelle: Charles Zaiontz

Interpretation der Koeffizienten möglich:

Treiber	Koeffizient
Konstante	14,273
t	0,55
Q1	- 3,98
Q2	- 5,73
Q3	- 2,49

Wenn **t um eine 1 Einheit steigt** (1 Quartal mehr) und die anderen Faktoren konstant sind, so steigt der Umsatz um 0,55 Millionen US Dollar.

Wenn die anderen Faktoren konstant sind, so wird im **Q1** 3,98 Millionen US Dollar weniger umgesetzt als im Q4.

Vergleich der Methoden – Lineare Regression

Vorteile:

- Interpretierbar
- einfach zu implementieren und zu maintainen
- kein Hyperparameter Tuning notwendig

Nachteile

- nur einfache (lineare) Zusammenhänge abbildbar
- schlechte Performance, wenn Treiber korreliert sind (Multikollinearität) oder vergessen werden
- saisonale Effekte müssen explizit mit aufgenommen werden

Vergleich der Methoden – (S)ARIMA(X)

(Seasonal) Auto Regressive Integrated Moving Average Process (with Exogenous Variables)

Lineare Regression, die:

- Vergangene Werte der zu prognostizierenden Variable (Lags) berücksichtigt (AR)
- Eigene Prognosefehler der Vergangenheit berücksichtigt (MA)
- Saisonelle Komponenten berücksichtigt (ARS, MAS)
- Treiber berücksichtigt

Vergleich der Methoden – (S)ARIMA(X)

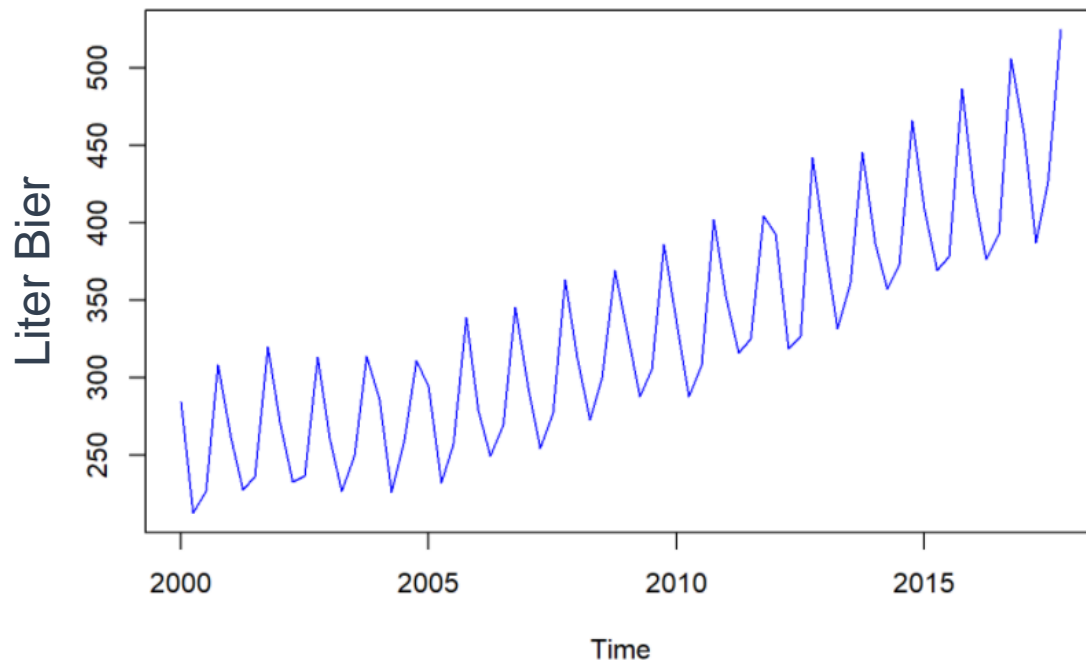
Umsatz = Konstante + $\beta \times \text{Lags}$ + $\gamma \times \text{Prognosefehler}$ + $\alpha \times \text{saisonale Lags}$ + $\delta \times \text{saisonale Prognosefehler}$ + $\sigma \times \text{Treiber}$

Welche Hyperparameter sind festzulegen?

- Wie viele und welche Lags sollen berücksichtigt werden?
 - Wie viele und welche Prognosefehler sollen berücksichtigt werden?
- automatisches Finden der besten Hyperparameter in Streamlit**

Vergleich der Methoden – (S)ARIMA(X)

Beer sales time series data between 2000 and 2017



Quelle: <https://rpubs.com/amitkayal/369361>

Interpretation der Koeffizienten möglich:

Treiber	Koeffizient
Konstante	10,05
Marketingausgaben	0,055
ARL1	0,24
MAL1	- 0,45
ARSL12	0,55

Wenn die **Marketingausgaben** um eine 1 Einheit steigen und die anderen Faktoren konstant sind, so steigt der Verkauf um 0,055 Liter Bier.

Vergleich der Methoden – (S)ARIMA(X)

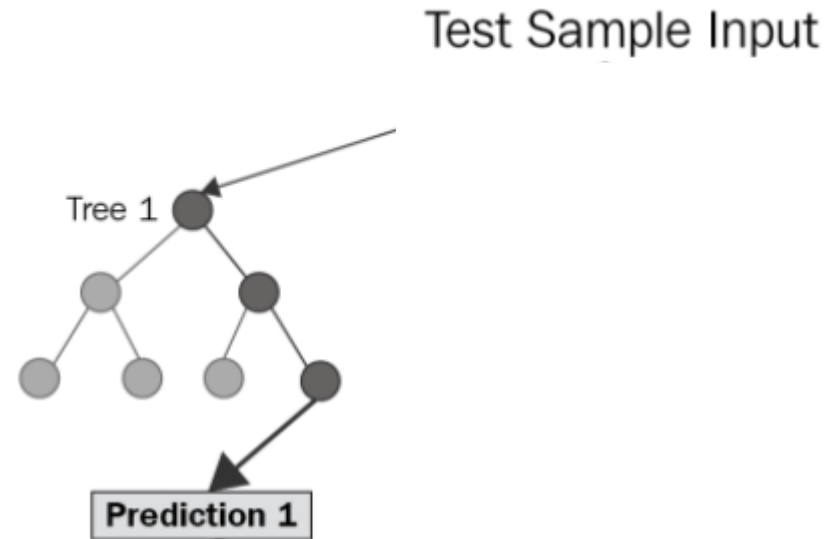
Vorteile:

- Interpretierbar
- Abhängigkeiten zu historischen Werten abbildbar
- einfach zu implementieren und zu maintainen

Nachteile

- nur einfache Zusammenhänge abbildbar
- schlechte Performance, wenn Treiber korreliert sind (Multikollinearität)

Vergleich der Methoden – Random Forest

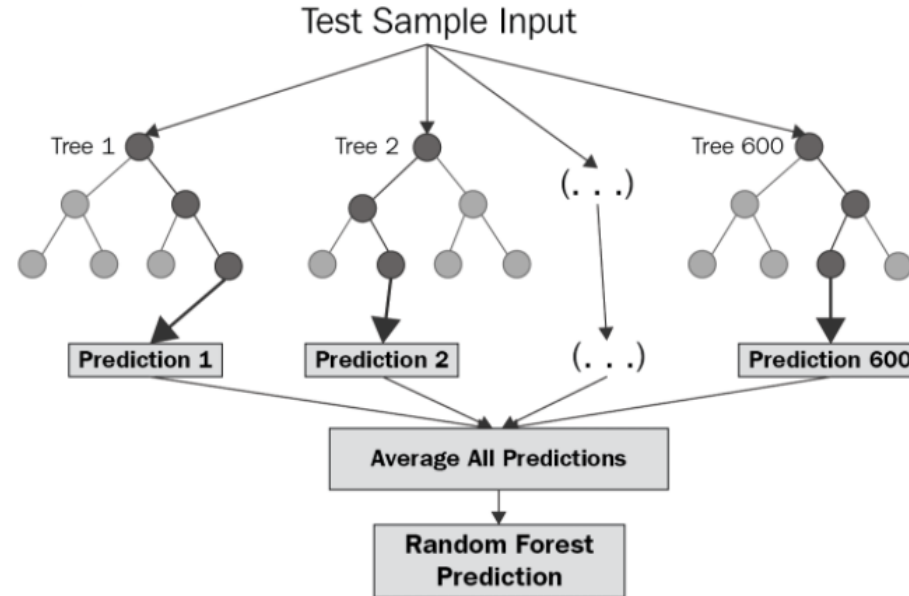


Quelle: <https://ichi.pro/de/random-forest-und-seine-implementierung-124804450305359>

Vergleich der Methoden – Random Forest

Feature Importance:

- erlaubt eine qualitative Aussage über die Treiber
- je höher der Wert, desto mehr trägt der Treiber zur Verbesserung der Prognose bei



Quelle: <https://ichi.pro/de/random-forest-und-seine-implementierung-124804450305359>

Vergleich der Methoden – Random Forest

Welche Hyperparameter sind festzulegen?

- Anzahl der zu schätzenden Bäume
- max. Anzahl der Tiefe
- ...

→ Streamlit soll diese Hyperparameter selbst optimieren

Vergleich der Methoden – Random Forest

Vorteile:

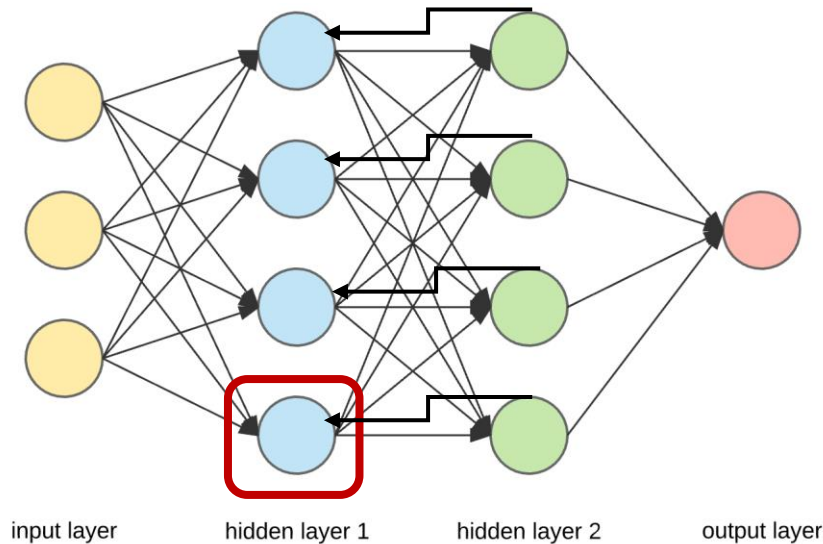
- nicht-lineare Zusammenhänge abbildbar
- Mittelwertbildung ermöglicht robuste Ergebnisse
- Wichtigkeit der Treiber interpretierbar (Feature Importance)

Nachteile

- kann nur Werte prognostizieren, die in den Trainingsdaten vorkommen
- saisonale Effekte und Lags müssen explizit mit aufgenommen werden
- erlaubt keine quantitativen Aussagen

Vergleich der Methoden – Neuronale Netze

Neuronale Netzwerke sind nichts anderes als die Optimierung komplexer Funktionen!



$$Output = f(\dots f(\dots f(1 - \sigma(w_z \times [h_{t-1}, x_t]) \times h_{t-1} + \sigma(w_z \times [h_{t-1}, x_t]) \times \tanh(W \times [r_t \times h_{t-1}, x_t])))$$

Quelle: <https://gongster.medium.com/how-does-a-neural-network-work-intuitively-in-code-f51f7b2c1e3f>

Vergleich der Methoden – Neuronale Netze

Welche Hyperparameter sind festzulegen?

- Anzahl der Layer
- Anzahl der Units
- Aktivierungsfunktionen
- Optimierungsverfahren zum Trainieren der Daten
- ...

→ **Streamlit soll diese Hyperparameter selbst optimieren**

Vergleich der Methoden – Neuronale Netze

Vorteile:

- können jeden Zusammenhang abbilden
- Abhängigkeiten zu historischen Werten abbildbar

Nachteile

- nicht interpretierbar
- schwierig zu implementieren und zu maintainen

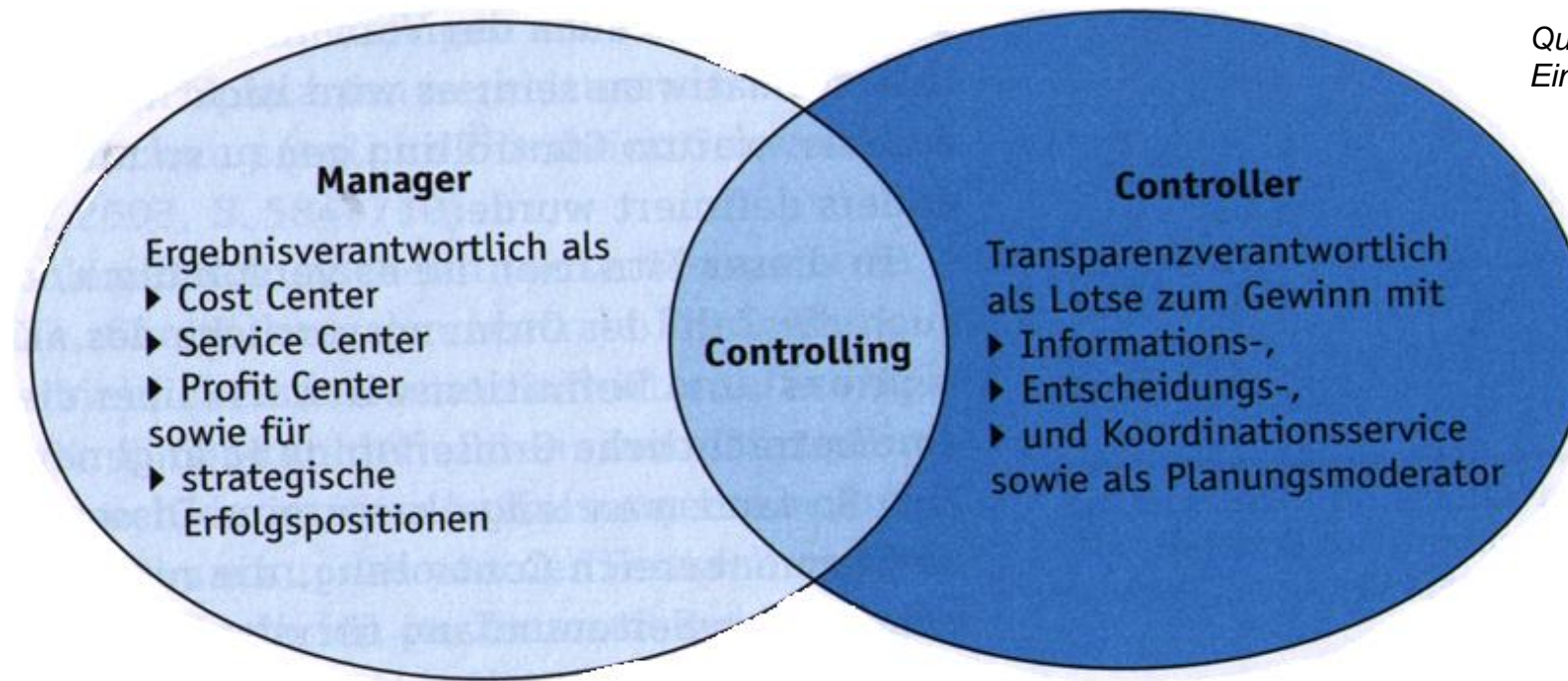
Vergleich der Methoden – Zusammenfassung

- Trade-Off: Interpretierbarkeit vs. Komplexität
- welches Modell sinnvoll ist, hängt stark von den Daten und der Aufgabe ab
- je mehr Daten, desto besser → vor allem bei komplexere Methoden
- Streamlit soll selbstständig die besten Hyperparameter bestimmen
- Expertenmeinung → Controller*in
- Maintenance der Methoden / Datenmanagement → Data Scientist?

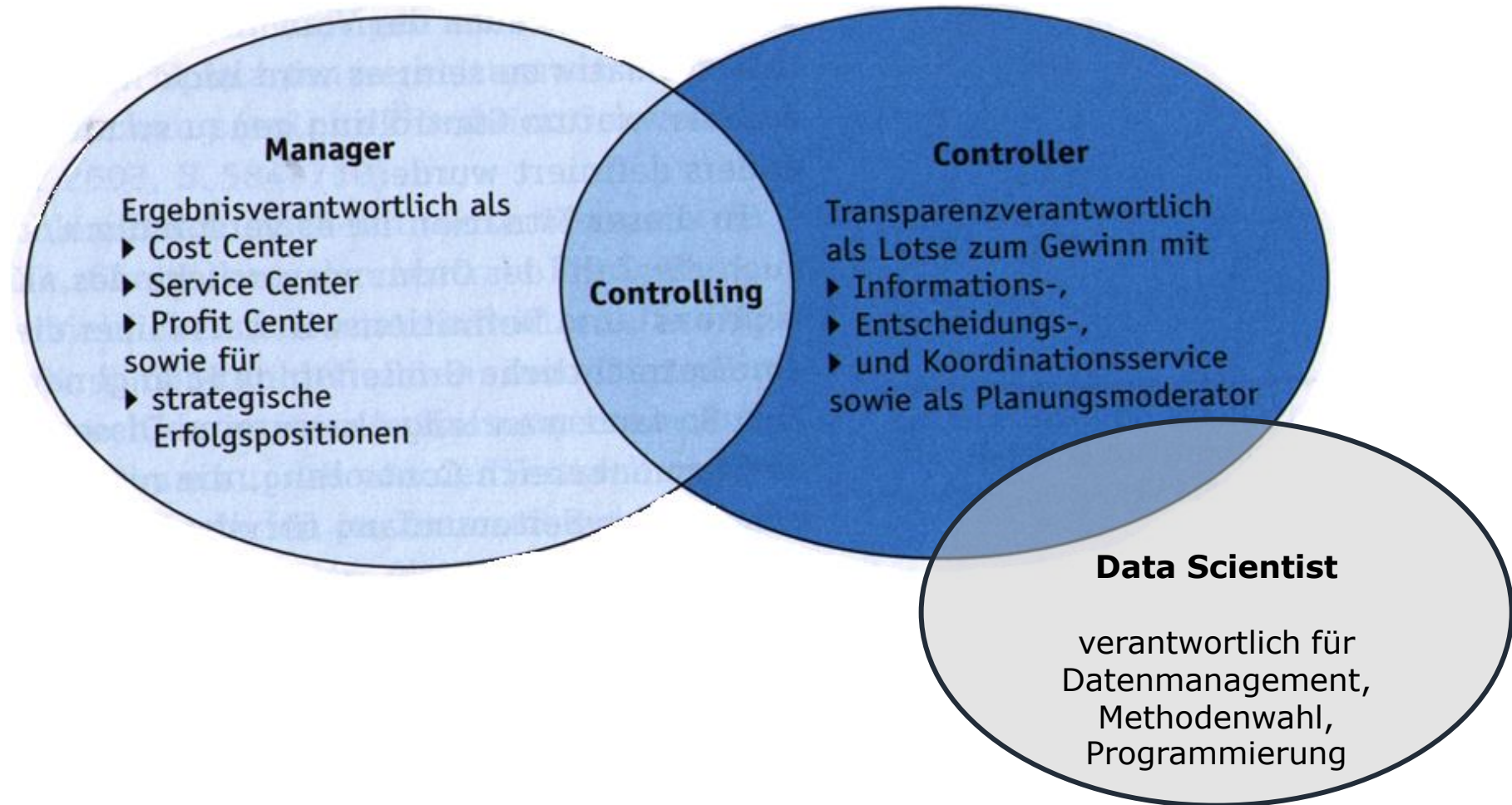
3. Modernisierung des Rollenverständnisses für Controlling und Management

Erweiterung des Akteur- und Aufgabenspektrums

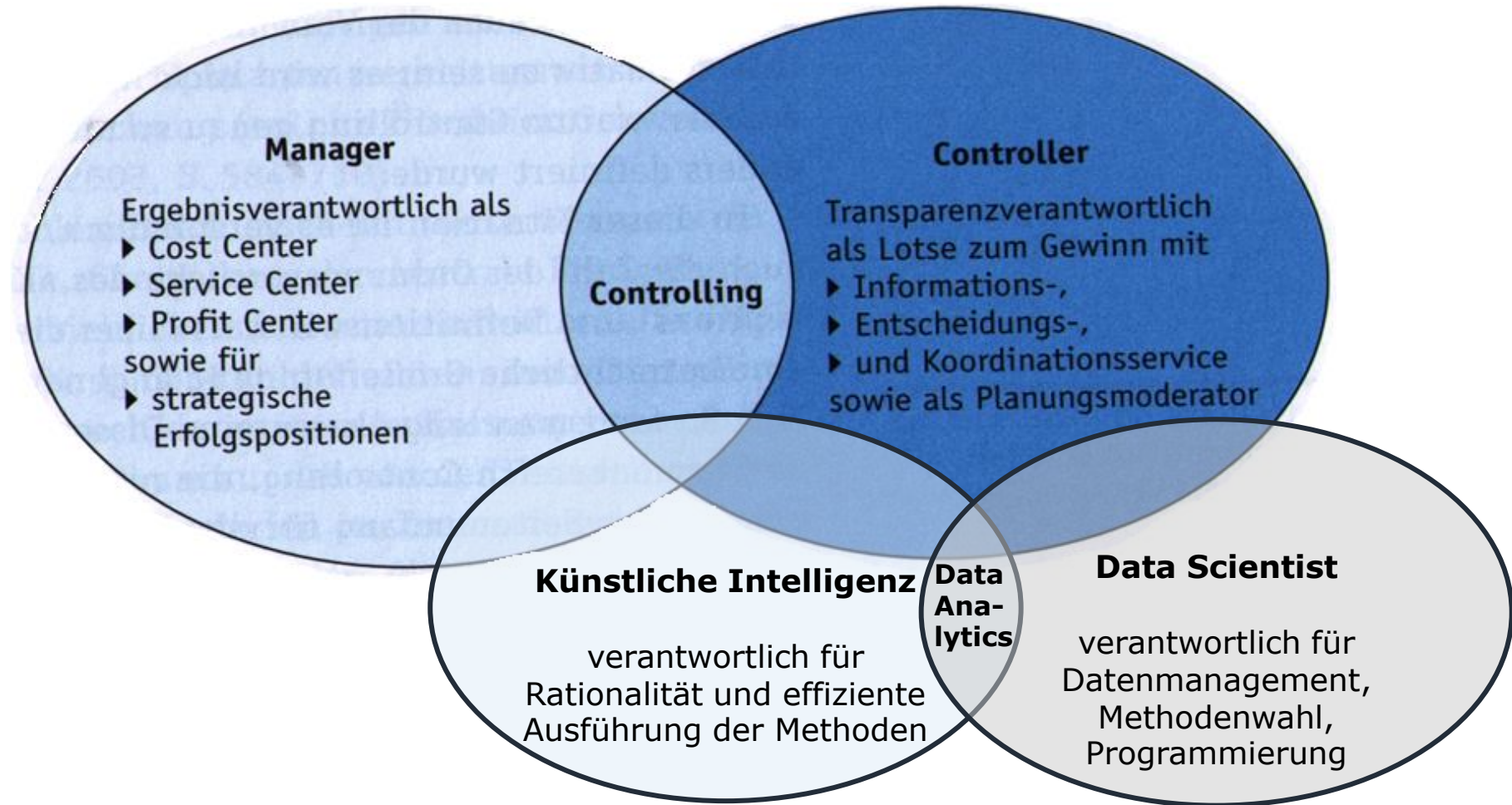
Quelle: Weber / Schäffer,
Einführung in das Controlling



Erweiterung des Akteur- und Aufgabenspektrums



Erweiterung des Akteur- und Aufgabenspektrums



Auswirkung der Digitalisierung auf die (Teil-)Prozesse im Controlling

Hauptprozess	Teilprozesse							
Strategische Planung	Strategische Analyse	Prüfung/Anp. Vision, Mission, Werte	Prüfung/Anp. Geschäftsmodell und strategische Stoßrichtung	Definition Ziele, Maßnahmen & Messgrößen	Finanzielle Bewertung d. Strategie	Abstimmung der Strategie mit Stakeholdern	Kommunikation der Strategie	Monitoring der Strategieumsetzung
Operative Planung, Budgetierung	Festlegen/Kommunizieren von Prämissen & top-down Ziele		Erstellung von Einzelplänen & Budgets	Zusammenfassung & Konsolidierung von Einzelplänen		Prüfung/Anp. der Planungsergebnisse	Präsentation & Verabschiedung der Planung	
Forecast	Ermittlung einer Datenbasis für den Forecast sowie Erstellung Forecast		Datenanalyse & Abweichungsanalyse (Forecast bzw. Plan/Budget)		Erarbeitung von Gegensteuerungsmaßnahmen		Verabschiedung des Forecasts	
Kosten-, Leistungs-, Ergebnisrechnung	Definition & Pflege Stammdaten	Kostenartenrechnung und Kostenstellenrechnung (inkl. Leistungsverrechnung)		Angebots-/ Auftragsplankalkulation	Mitfld.- & Nachkalkulation	Periodenerfolgsrechnung	Periodenabschluss der Kostenrechnung	Abweichungsanalyse
Management Reporting	Management des Reportingsystem- & Datenprozesses		Berichtserstellung (Zahlenteil)		Berichtserstellung (Abweichungsanalyse und Kommentar)		Bewertung durch Management & Einleitung von Maßnahmen	
Projekt- und Investitionscontrolling	Planung des Projektes/ Investitionen	Unterstützung des Genehmigungsverfahrens		Erstellung von Investitions-/ Projektberichten	Erstellung von Entscheidungsvorlagen		Nachkalkulation und Abschlussbericht	
Risiko- management	Identifikation & Klassifikation von Risiken	Analyse & Bewertung von Risiken	Aggregation der Einzelrisiken Gesamtrisikopositionen		Ableiten & Verfolgen von Risikomaßnahmen		Erstellung eines Risikoberichts	
Betriebswirtschaftliche Beratung & Führung	Begleitung Entscheidungsprozess	Begleitung/Einleitung von Maßnahmen zum Ergebnis-/ Kostenmanagement		Mitarbeit/Initiation von Prozessanalyse und -optimierung		Projektmitarbeit	Förderung von betriebswirtschaftlichem Know-how im Unternehmen	

= stark betroffen
 = mittelstark betroffen
 = wenig/leicht betroffen

Quelle: Kirchberg; Müller (2016): Digitalisierung im Controlling, in: Gleich; Grönke; et al. (Hrsg.): Konzerncontrolling 2020, S. 91ff.

Herausforderungen

- Top-1-Herausforderung ist Datenmanagement
- Zu viele Daten aus unterschiedlichsten Quellen, welche harmonisiert werden müssen
- steigende Anzahl Anspruchsgruppen außerhalb des Controllings
- Müssen in einer Quelle zusammengeführt und harmonisiert werden sowie an Interessensgruppen verteilt werden

*Vgl. the factlights news - Die Top-Herausforderungen der Digitalisierung
Link: <https://the-factlights.de/top-herausforderungen/>*

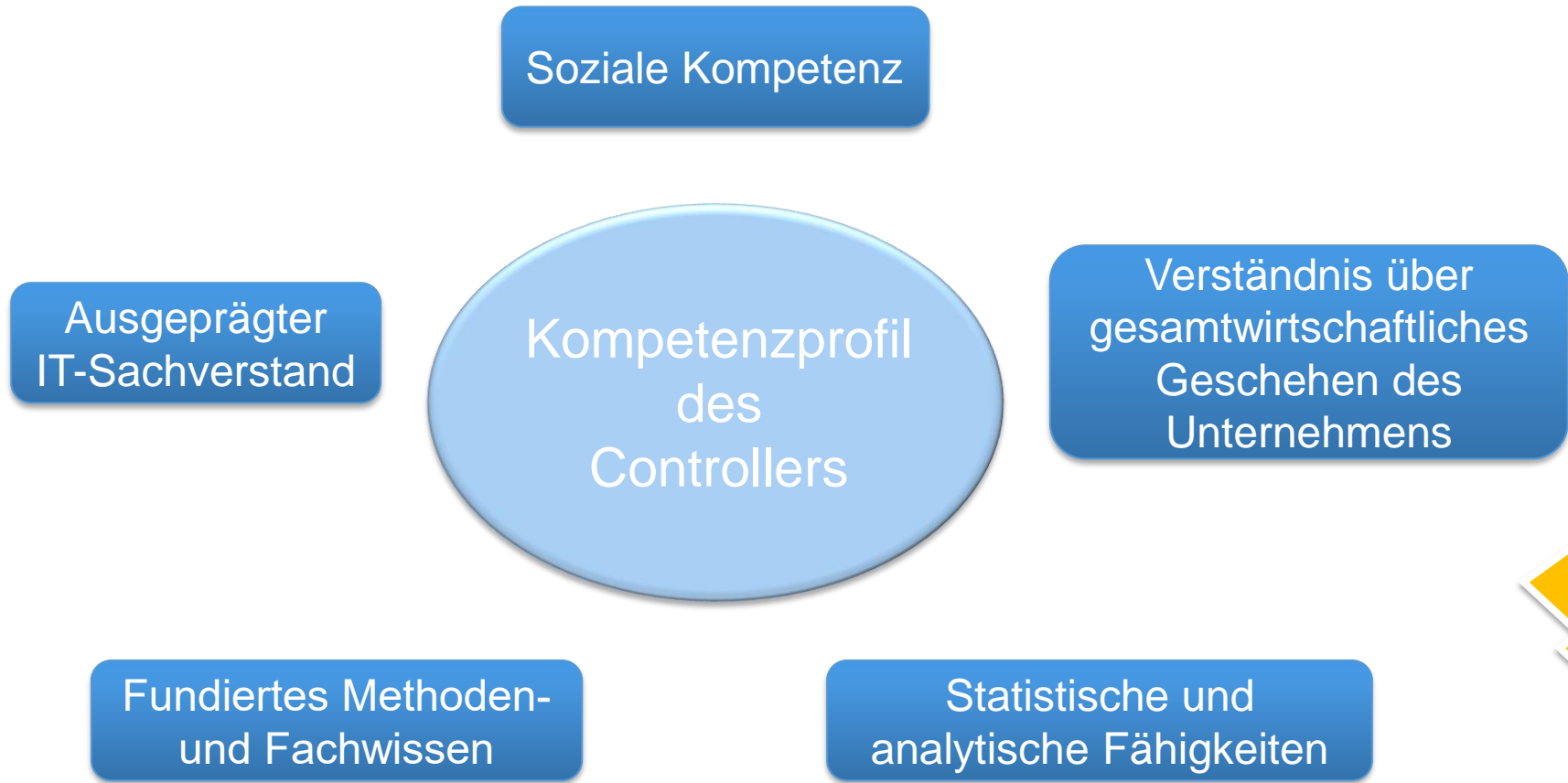
Leitbild Controlling 4.0

- **Standardisierung** als Voraussetzung für Integration, Flexibilität und Sicherheit von Daten und Analyseergebnissen
 - Sicherstellung einer “single source of truth“
- **Automatisierung** manueller/repetitiver Controllingprozesse
 - Herausforderung: Nutzung der freiwerdenden Kapazitäten durch Weiterentwicklung von Controller-Aufgaben
- **Spezialisierung** in Form veränderter Kompetenzanforderungen

→ Controller in der Verantwortung, **Kulturwandel** hin zum datengetriebenen Unternehmen voranzutreiben

Vgl. Möller; Pieper (2015), S. 44 Vgl. HeimeI, Müller (2019), S. 421ff.

Leitbild Controlling 4.0



- Data Scientist (große Unternehmen) & Business Analyst
- Gelingt Anpassung nicht, werden Funktionen langfristig von Maschinen und Managern übernommen
- Klassischer Business Partner wird überflüssig



Eigene Darstellung, vgl. Heimel, Müller (2019), S.424f.

Buchtipp zum Einstieg in Themenkomplex



Link: <https://link.springer.com/book/10.1007/978-3-658-26763-6>

Linksammlung

- <https://www.kaggle.com/> Große internationale Data Science-Community mit Lernvideos zu verschiedenen Methoden, Datensets, Fallbeispielen etc.
- <https://azure.microsoft.com/de-de/> Sehr verbreitete Plattform von Microsoft zur Anwendung und Übung von Maschinellem Lernen, zugleich Cloud; teilweise kostenfrei
- <https://www.icv-controlling.com/de/arbeitskreise/bi-big-data-und-controlling.html> Arbeitskreis Big Data und Controlling im Internationalen Controllerverein, vgl. hier auch Unterpunkt Publikationen
- <https://machinelearningmastery.com/start-here/> Kostenfreie Online-Einführungskurse und vertiefende Informationen zum Machine Learning
- <https://r4ds.had.co.nz/> und <https://adv-r.hadley.nz/> Populäre kostenfreie Literatur zur verbreiteten Programmiersprache R für Maschinelles Lernen; unter <https://rstudio.com/products/rstudio/> gibt es die notwendige Software R
- <https://otexts.com/fpp2/> Kostenfreie Literatur zum Thema Prognosen
- <https://big-data-factory.de/frei-verfuegbare-testdaten-fuer-big-data-projekte/>

Quellen

Bauer, Thomas et al. (2018): Big Data in der wirtschaftswissenschaftlichen Forschung, in: König, Christian et al. (Hrsg.) (2018): Big Data. Chancen, Risiken, Entwicklungstendenzen, Wiesbaden: Springer Fachmedien Wiesbaden, S. 129-148

Heimel, Jana; Müller, Michael (2019): Controlling 4.0 – Wie veränderte Datenverfügbarkeit und Analysemöglichkeiten das Controlling erneuern, in: Erner, Michael (Hrsg.) (2019): Management 4.0 – Unternehmensführung im digitalen Zeitalter, Berlin, Heidelberg: Springer Berlin Heidelberg, S. 389-430

The factlights Studie 2020 zu Digitalisierung und Datenarbeit. Link: <https://the-factlights.de/studie2020/#a>

Hyndman, Rob J.; Athanasopoulos, George (2018). Forecasting: principles and practice, OTexts.

Niklas Donges (2021): A Guide to RNN: Understanding Recurrent Neural Networks and LSTM Networks, Link: <https://builtin.com/data-science/recurrent-neural-networks-and-lstm>

Herzlichen Dank für Ihre Aufmerksamkeit!

Kontakte

Projekt-Exposé mit allen Beteiligten: <https://www.ifaf-berlin.de/projekte/recommend/>

Prof. Dr. Robert Knappe

Professur für Betriebswirtschaftslehre der
öffentlichen Verwaltung

Hochschule für Wirtschaft und Recht Berlin
Alt Friedrichsfelde 60
10315 Berlin
Tel. (+49) (0)30 - 30877 – 2664
Email: robert.knappe@hwr-berlin.de

Nicki Lena Kämpf

Wissenschaftliche Mitarbeiterin

Berliner Hochschule für Technik
Email: NickiLena.Kaempf@bht-berlin.de

Mitarbeit bei der Vorbereitung der Präsentation: Robert Brust, B.A.

4. Diskussion