Karlsruhe Institute of Technology

SDSC BW
Smart Data Solution Center Baden-Württemberg

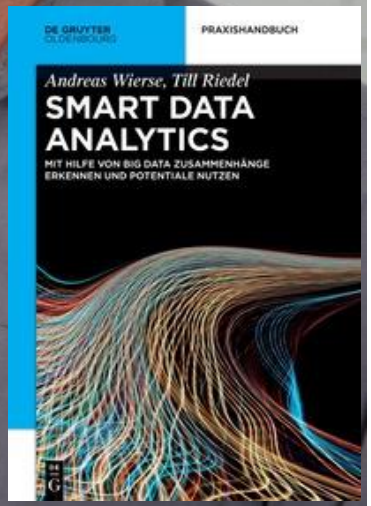
# Welche Daten sollte man sammeln??

## Weg zur erfolgsreiche Datenanalyse

Dr. Till Riedel



# Zu meiner Person




www.teco.edu/~riedel

Informatiker (Studium mit Vertiefung Compiler-Bau und Systemarchitektur)

Karlsruhe Institute of Technology

Lab Leader TECO (www.teco.kit.edu)

(seit 2010, seit 2005 am TECO)

Leiter des Analytics Teams

Smart Data Solution Center BW (sdsc-bw.de)

Koordination

Smart Data Innovation Lab (sdil.de)

Akademischer Rat am KIT

Vorlesung „Kontext Sensitive Systeme“

Forschung

Data Analytics für Sensor Netzwerke / Internet der Dinge (Promotionsthema: Modellgetriebe Middleware)

SDSC BW

Smart Data Solution Center Baden-Württemberg

# Unser gefördertes Angebot

- Was ist das SDSC-BW?
  - Passendes DataScience Transferangebot für KMUs in BW → kostenfrei bis Ende 2021
  - Unkompliziert, flexibel und schnell

Karlsruher Institut für Technologie
SICOS
Wir vermitteln Höchstleistung.

Baden-Württemberg
MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST

SDSC BW
Smart Data Solution Center Baden-Württemberg

SDSC BW
Smart Data Solution Center Baden-Württemberg

# Warum ein Solution Center für den Mittelstand?

- 99,4 Prozent aller Betriebe in BW haben weniger als 250 Mitarbeiter
- Besondere Herausforderungen bei Mittelständischen Unternehmen
- Anschluss an die Digitalisierung darf nicht verpasst werden

# Unser gefördertes Angebot

- Was leistet das SDSC-BW?
  - Kostenfreier Vor Ort Termin zur Vorstellung / als Ideengeber 
  - Kostenfreie Erstanalyse ihrer vorhandenen Bestandsdaten 
  - Empfehlungen für weiteres Vorgehen und ggf. Vorschlag für ein detailliertes Folgeprojekt 

# Vorgehen und Prozess im Projekt

- Wie sieht der Prozess im Projekt aus?
  - Einfacher und übersichtlicher Prozess mit direkten Ansprechpartnern


- Erstgespräch
- Voraussetzungen
- Zieldefinition
- Vorbereitungen

- Datenübergabe
- Analyse
- Auswertungen

- Ergebnisse
- Handlungsempfehlungen
- Ggf. nächste Schritte

Betreuung

# Vorgehen und Prozess im Projekt

- Was sind unsere Ziele?
  - Mit einem schnellen und einfachen Pilotprojekt den Wert der Daten zeigen
  - Gemeinsam mit Ihnen die ~~erste~~ Hürde nehmen
  - Ein Solution Center aufbauen, das kleine und mittelständische Unternehmen bei der Analyse ihrer Daten unterstützt

Herausforderung: Wir brauchen geeignete Daten von Ihnen!!!

## Welche Daten soll ich sammeln?

- Einleitung prädiktive Analysen
- Mehr Daten= Bessere Daten ?
- „Bessere“ Daten ≠ Bessere Daten !
- Bessere Daten !

Erfassung
Bereinigung
Merkmals-extraktion
Klassifikation
Schlussfolgerung
Sensor
Signal
Messwerte
Merkmalsvektor
Bewertung
Klasse

# KI heißt meist Prädiktive Analyse

| Var 1 | Var 2 | … | Var n | Zielvariable |
|---|---|---|---|---|
| | | | | z.B. Zahl, Kategorien |
| | | | | |
| | | | | |

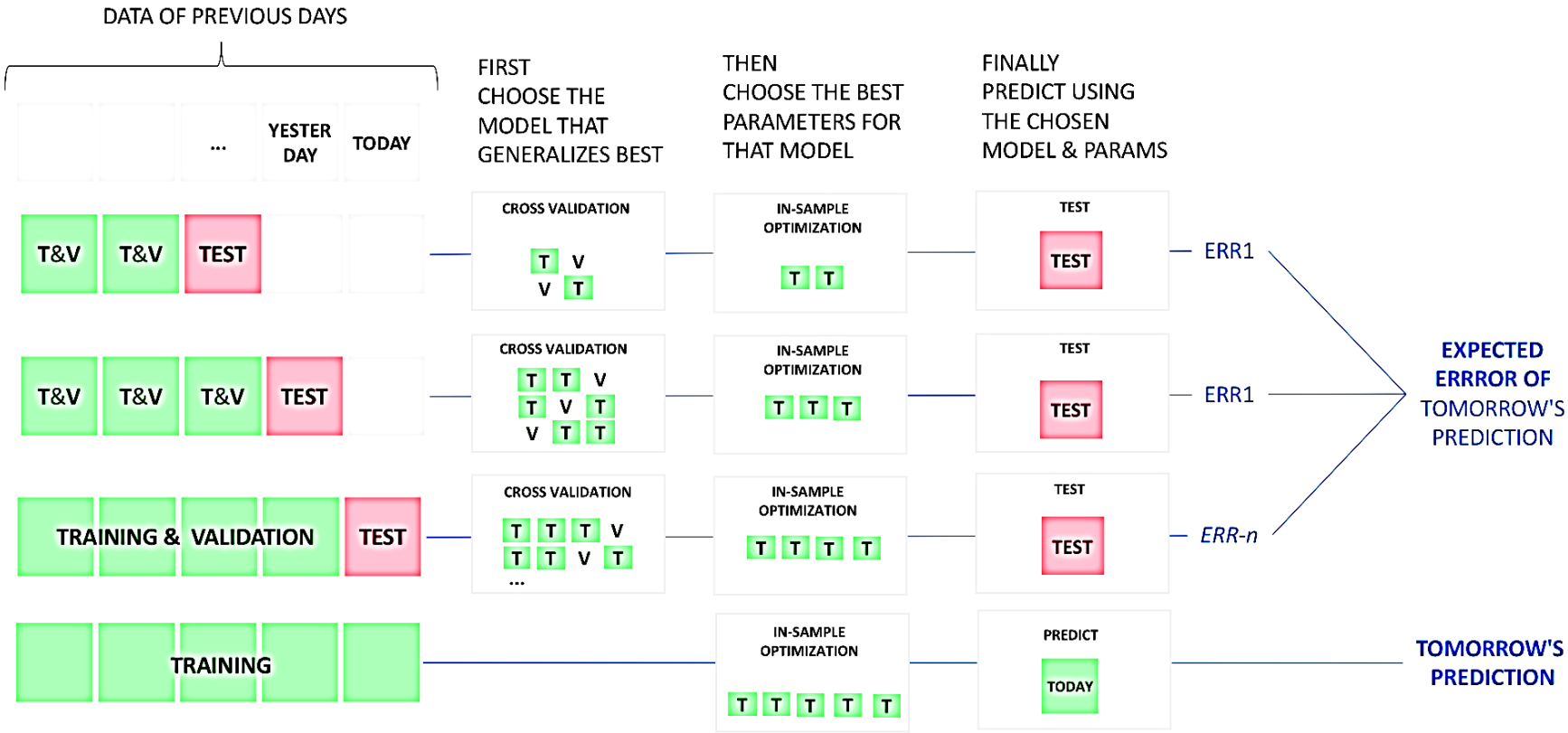
Relevant Information

Reflects your goal

- **Wichtig: Zielvariable muss Ihr Ziel widerspiegeln.**
- **Auflösung der Zielvariable definiert die effektive Datenmenge.**
- Mögliche Datentypen des Zielvariablen:
  - Kontinuierliche Zahl z.B. Temperatur, Preis, Energieverbrauch
  - Kategorie / Ordinale z.B. Qualitätsklasse (gut, schlecht), Fehler-Kennzeichen

- **NICHT** direkt geeignet: Fließtext über Ergebnisse der Qualitätskontrolle

# **PREDICT THE NEXT DAY** USING PREVIOUS DAYS DATA




Source: https://stats.stackexchange.com/questions/346907/splitting-time-series-data-into-train-test-validation-sets/346918

# Use Cases: Prädiktiv

## Predictive Maintenance

Wann sollte eine Maschine gewartet werden ?

| Zeitpunkt | Sensorwert1 | Sensorwert2 | ... | ... | Maschinenzustand |
|---|---|---|---|---|---|
| 15-02-2018 12:00 | 17.3 | 1032 | ... | ... | **5** |
| 20-02-2018 12:00 | 24.9 | 840 | ... | ... | **4.7** |

## Predictive Delay

Tritt eine Verspätung ein ?

| Zeitpunkt | Disponent | Status | Details | ... | Verspätung |
|---|---|---|---|---|---|
| 15-06-2018 16:36 | A | 7 | ... | ... | **Nein** |
| 16-06-2018 13:37 | B | 3 | ... | ... | **Ja** |

## Predictive Quality Control

Was ist die voraussichtliche Qualität der Produkte ?

| Produkt ID | Input 1 | Input 2 | Temperatur | Dauer | Qualitätsklasse |
|---|---|---|---|---|---|
| 0123 | 25 | Z1 | 34 | 10 min | **A** |
| 0124 | 25 | Z3 | 30 | 15 min | **B** |

# Typen der Potenzialanalyse

## Prädiktiv

- Entwicklung eines Modells mithilfe von Beispieldaten
- Beispiel:
  - Vorhersagen von Wartungsbedarf
  - Vorhersagen von Kundenabwanderung
  - Betrugserkennung
  - Mögliche Auswirkungen der Marketing Maßnahmen
- **Voraussetzung**: **Zielvariable**
- Evaluierung möglich
  - → **Geeignet für SDSC Projekte**
- *Supervised learning techniques*

## Explorativ

- Finden von Zusammenhängen zwischen Variablen
- Beispiel:
  - Korrelation von Variablengruppen
  - Kundensegmentierung
  - Clustering
- **Achtung**: Keine Referenzwerte
  - → Evaluierung nicht möglich
- *(Unsupervised learning techniques)*

Bevorzugt

# Use Cases: Explorativ

## Anomaly Detection

Visualisierung / Erkennung außergewöhntlicher Werte

| Zeit | Druck |
|---|---|
| 15-02-2018 00:20 | 4 bar |
| 15-02-2018 00:21 | 5 bar |

## Clustering

Kundensegmentierung

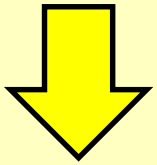
| Kunden ID | Geschlecht | Alter | Zeit | Umsatz |
|---|---|---|---|---|
| A111 | F | 50 | 14-02-2018 08:20 | 50,45 € |
| A112 | M | 25 | 14-02-2018 14:21 | 10,69 € |

## Rule Mining

Ableitung von Zusammenhängen in Form von Regeln

| Zeit | Auslastung 1 | Auslastung 2 | Auslastung 3 | Auslastung 4 | Spezifische Leistung |
|---|---|---|---|---|---|
| 15-02-2018 00:20 | 100% | 20% | 35% | 28% | 6 kW / $m^3$ |
| 15-02-2018 00:21 | 100% | 50% | 33% | 28% | 8 kW / $m^3$ |

# Prozess einer Potentialanalyse



## Technisches Gespräch

- Diskussion Hintergrund / Problem
- Zielsetzung

## Datenübergabe

- nach Datenanforderung übergeben
- Inkl. Datenbeschreibung
- Beginn Potentialanalyse

## Zwischen Präsentation

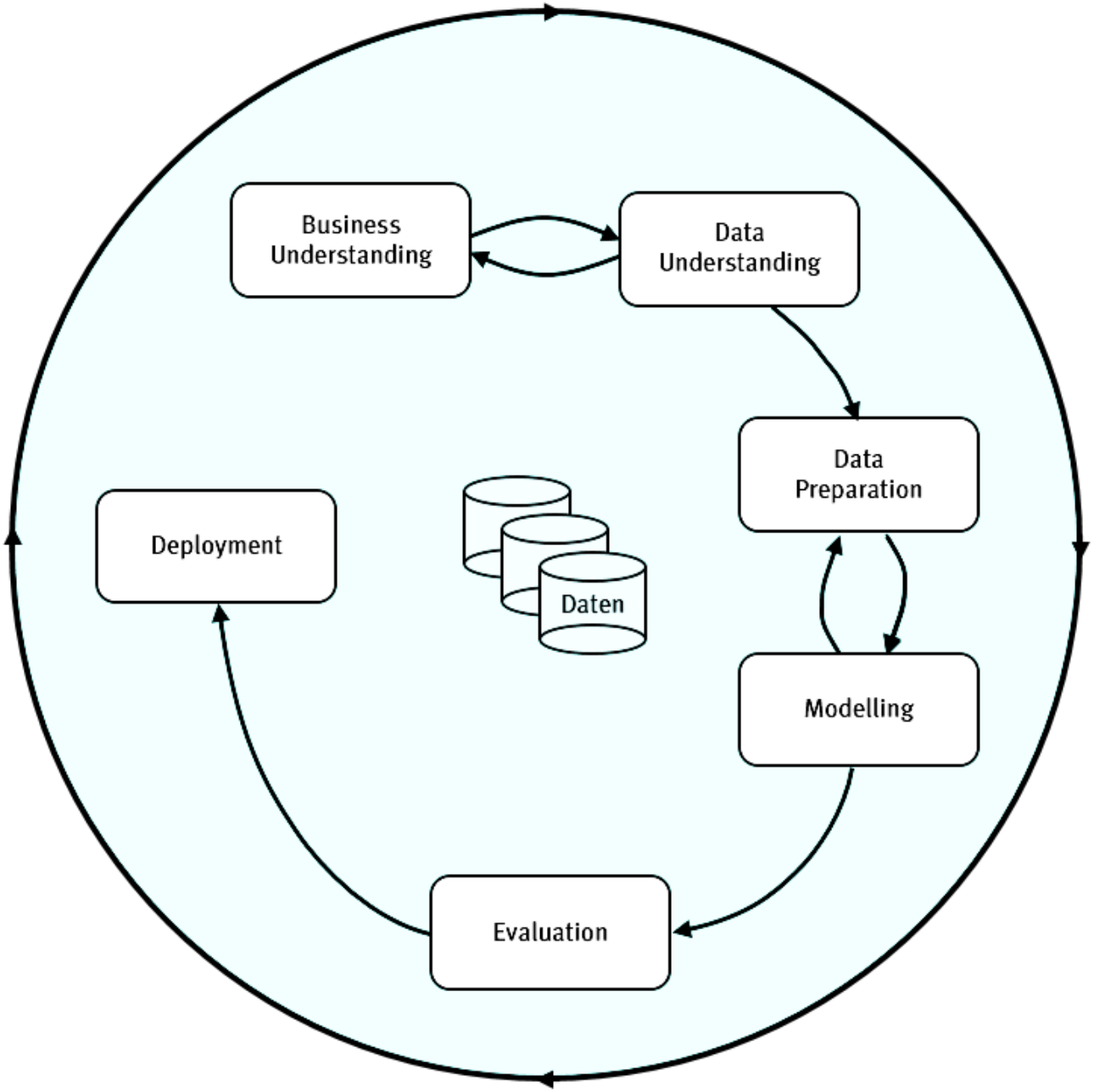
- Über Telekonferenz
- Diskussion Datenqualität
- Vorstellung Ansätze
- Auswahl geeigneter Ansätze
- Ggf. Überarbeitung des Ziels

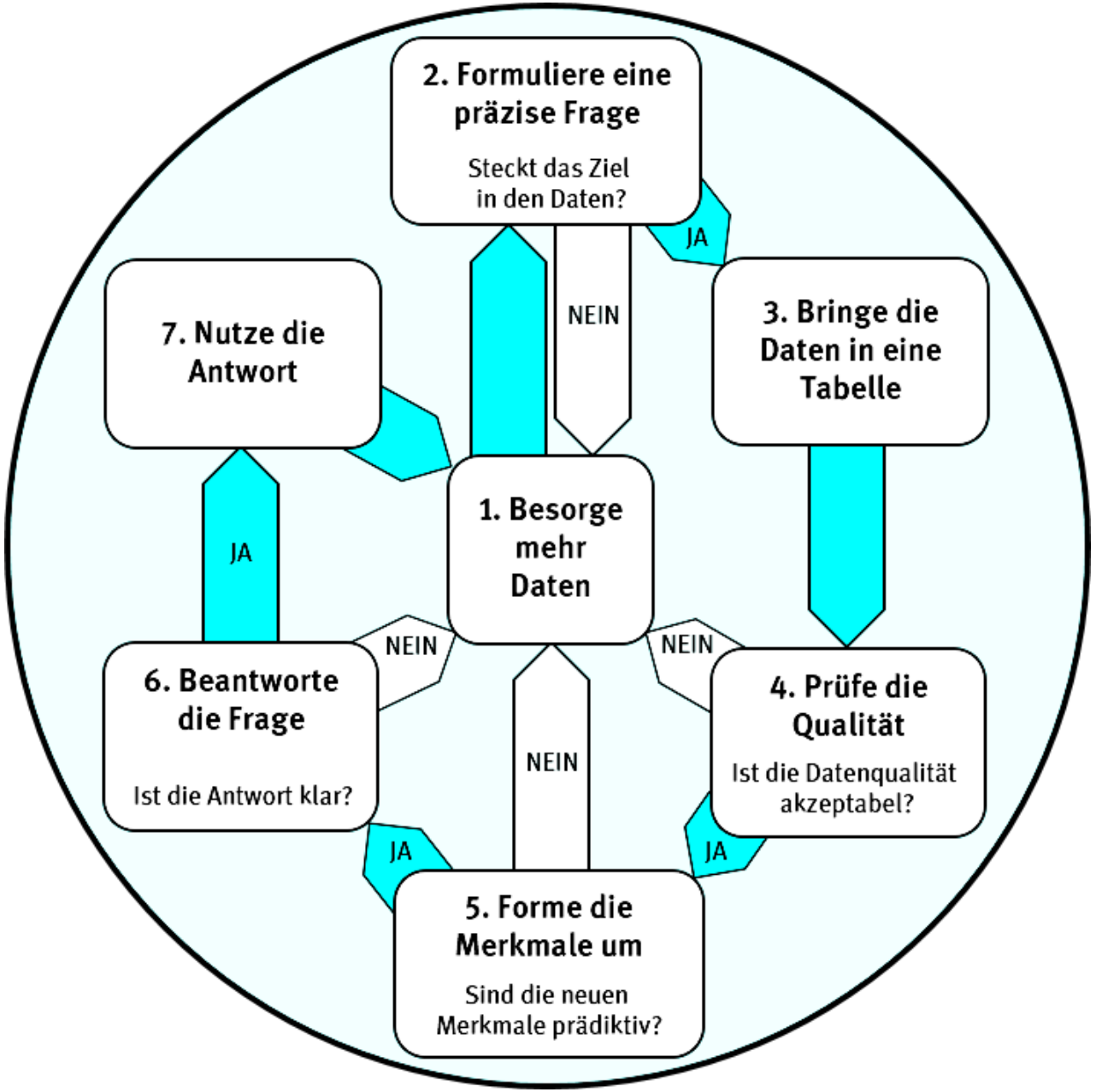
## Finale Präsentation

- vor Ort oder bei SDSC
- Vorstellung detaillierter Ansätze und Ergebnisse
- Diskussion weiterer Potenziale

# Gute Daten sind exportierbare Daten

- CSV noch immer defacto-Standard für den Datenaustausch
  - UTF-8
  - Beschreibung
    - Kommastelle: Komma oder Punkt?
    - Datumformat: YY-MM-DD, MM/DD/YYYY, Excel time stamp?
    - Interne Codierung
    - Spaltenbeschreibung
- Speicherung von Daten kann effizienter geschehen
  - aber Daten sollten exportiert und überprüft werden!!
  - (auch triviale) Datenanalysen selber können die Datenqualität am besten beschreiben
  - Data analytics arbeitet auf historischen Daten: Fehler können nur sehr langsam bzw. gar nicht korrigiert werden


Business Understanding
Data Understanding
Data Preparation
Modelling
Evaluation
Deployment
Daten


2. Formuliere eine präzise Frage
Steckt das Ziel in den Daten?
JA
NEIN
3. Bringe die Daten in eine Tabelle
7. Nutze die Antwort
JA
1. Besorge mehr Daten
NEIN
NEIN
6. Beantworte die Frage
Ist die Antwort klar?
4. Prüfe die Qualität
Ist die Datenqualität akzeptabel?
NEIN
JA
JA
5. Forme die Merkmale um
Sind die neuen Merkmale prädiktiv?

# WARUM MEHR DATEN

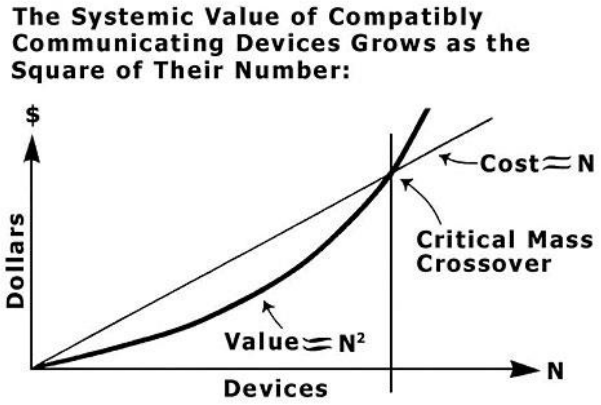
# Das Potential: Metcalf's Law und Vernetzung

*Vernetzte Systeme haben einen potentiellen Wert der sich quadratisch zu den Komponenten entwickelt ($N^2$)*

Robert Metcalfe
(source: U.S. Department of Commerce)

**Das ist dann der Fall wenn alle Teile miteinander interagieren können:**

**Smart Data heißt Vernetzung durch Daten!!**

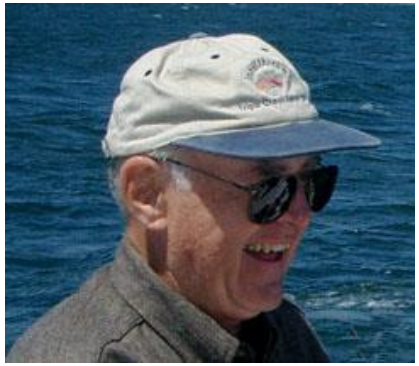



Source: IEEE Spectrum

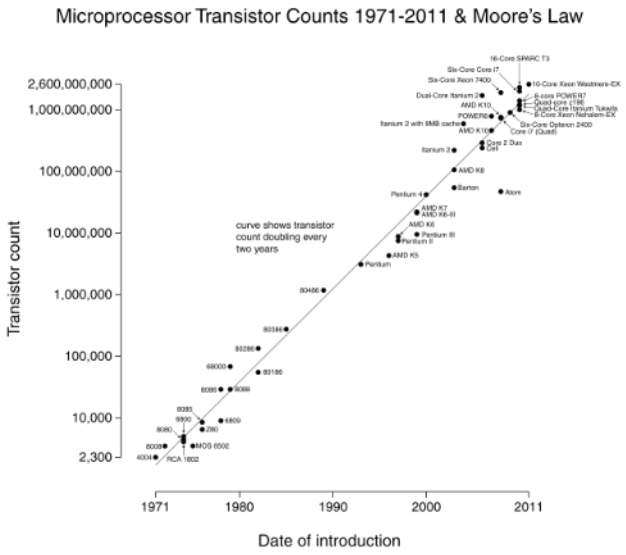
# Die Herausforderung: Moore's Law

*Leistungsdichte digitale Schaltungen verdoppelt sich alle 18 monate*

- → Konsumenten verlangen alle 1-2 Jahre eine neue Gerätegeneration
- → Siehe bereits PC HW, Photo Kameras, TVs, Telefone, Netzwerk Equipment…

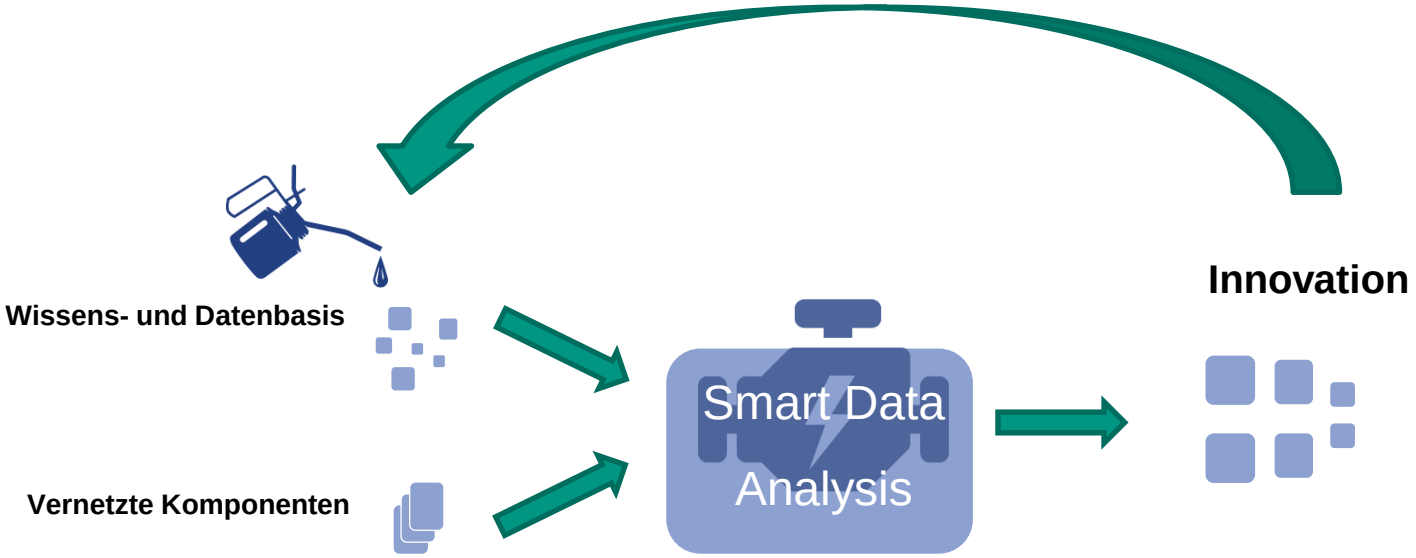


Gordon Moore
Photographer: Steve Jurvetson




**Hypothese: Nur agile, datengetriebene Innovation schafft 18 monatige Innovationszyklen…**

# Wie können Daten schnell zu Wachstum führen?




**“Wenn Daten das neue Öl sind,
dann ist Analytics ihr Verbrennungsmotor”
(Quelle: Gartner)**

# Mehr Daten: Wachsen nach innen!

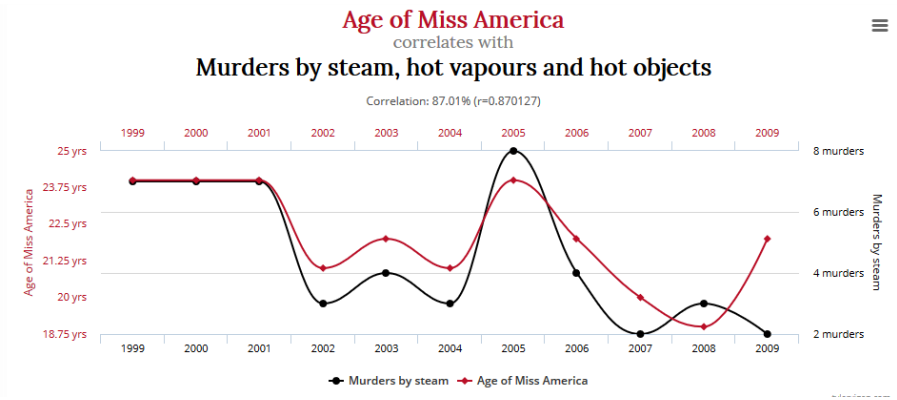
Maschinenhersteller sagen vorraus das 2025 32x mehr Rechnerknoten auf dem Hallenboden sind wie heute-> alle produzieren Daten

# Mehr Daten bedeutet nicht ein Meer von Daten

- Es zählt nicht die Anzahl der Einzeldaten sondern die Anzahl der Annotationen y
  - Qualitätsmaß
  - Kennzahl
  - Zukünftiger Wert
  - → die Einzeldaten X n:1 zugeordnet werden kann/abhängig davon ist

- Es zählt die relevanten und unhabhängige Daten
  - Extrem viele Daten erhöhen nur das Risiko von spurious correlations
  - Ableitbare Daten helfen nur bedingt

Source: https://www.tylervigen.com/spurious-correlations



# Rückblick: Prädiktive Analyse

## Predictive Maintenance

Wann sollte eine Maschine gewartet werden ?

| Zeitpunkt | Sensorwert1 | Sensorwert2 | ... | ... | Maschinenzustand |
|---|---|---|---|---|---|
| 15-02-2018 12:00 | 17.3 | 1032 | ... | ... | **5** |
| 20-02-2018 12:00 | 24.9 | 840 | ... | ... | **4.7** |

## Predictive Delay

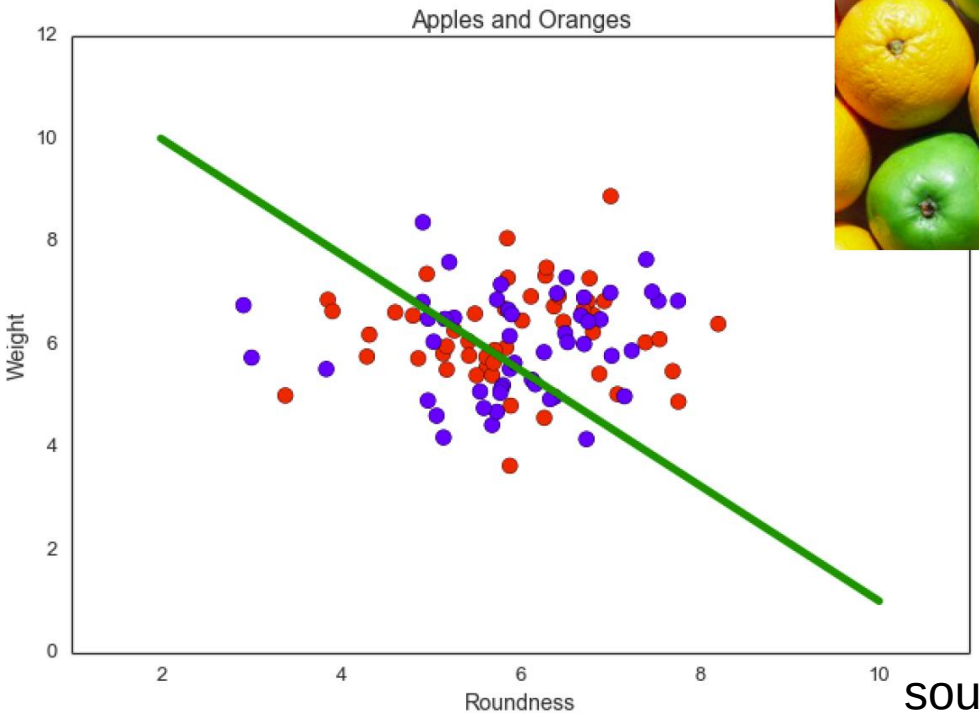
Tritt eine Verspätung ein ?

| Zeitpunkt | Disponent | Status | Details | ... | Verspätung |
|---|---|---|---|---|---|
| 15-06-2018 16:36 | A | 7 | ... | ... | **Nein** |
| 16-06-2018 13:37 | B | 3 | ... | ... | **Ja** |

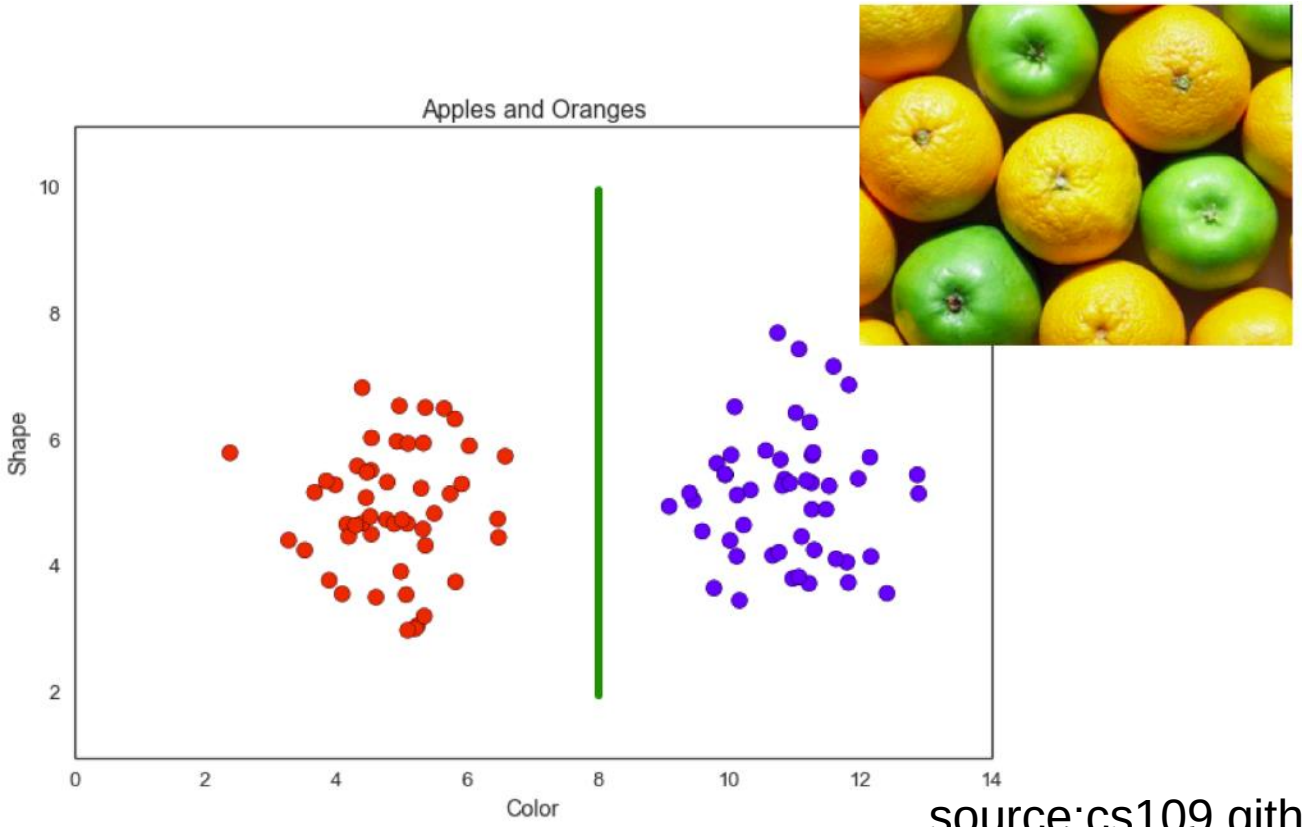
## Predictive Quality Control

Was ist die voraussichtliche Qualität der Produkte ?

| Produkt ID | Input 1 | Input 2 | Temperatur | Dauer | Qualitätsklasse |
|---|---|---|---|---|---|
| 0123 | 25 | Z1 | 34 | 10 min | **A** |
| 0124 | 25 | Z3 | 30 | 15 min | **B** |




source:cs109.github.io



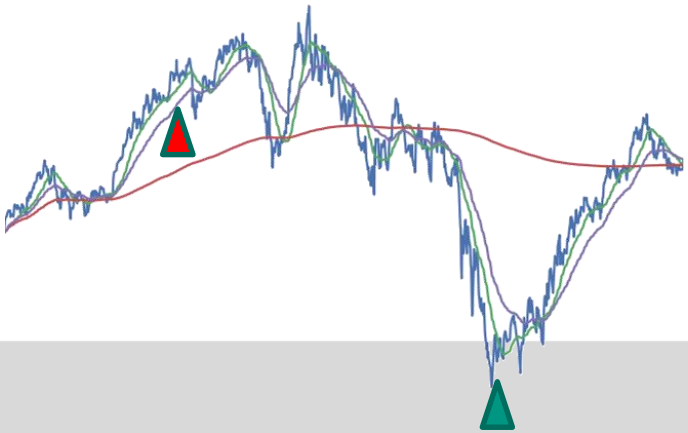

source:cs109.github.io

# „BESSERE“ DATEN

# Typische Probleme der Vorverarbeitung

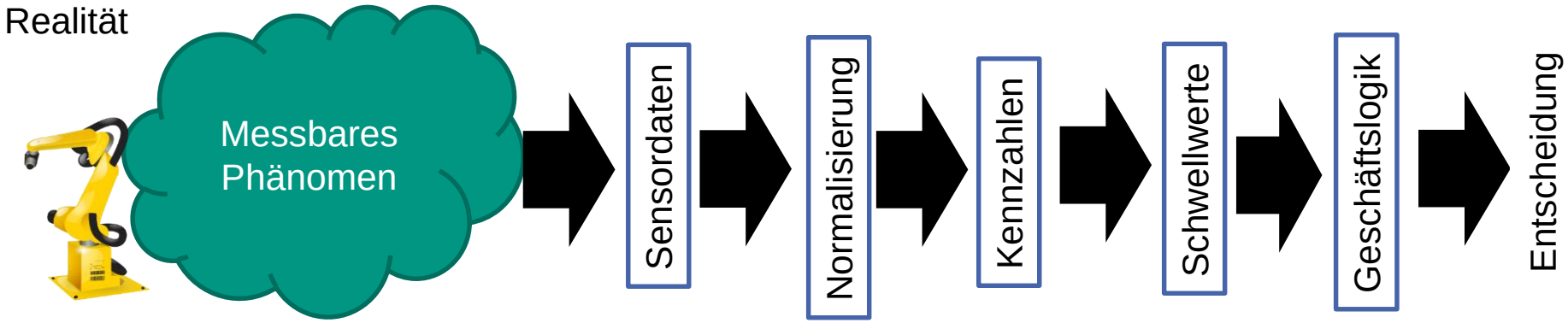
- Bessere Daten heißt nicht nur die „guten Daten“?
  - Nur Ausschuss ohne gute Beispiele  (Einklassen-Klassifikation schwierig)
  - Daten sollten auch das „Rauschen“ beschreiben (Overfitting)

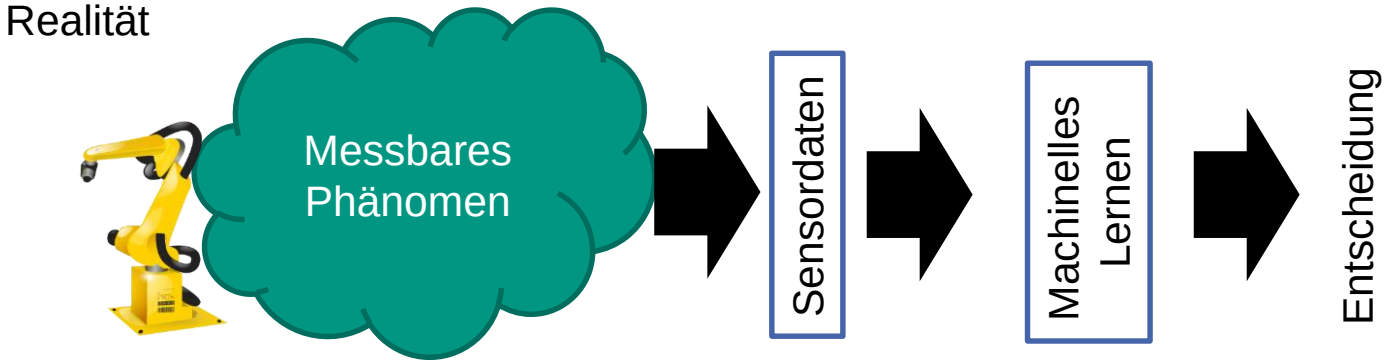
- Systembedingt keine Varianz in den Eingangsdaten
  - Randomisierung Stellgrößen
  - Systematische Exploration des Eingangsraumes
  - Ergebnis vor „Gut“/“Schlecht“ Vergleich aufzeichnen (Regression → Wie „gut“)
  - Daten unverarbeitet/ungemittelt aufzeichnen!!



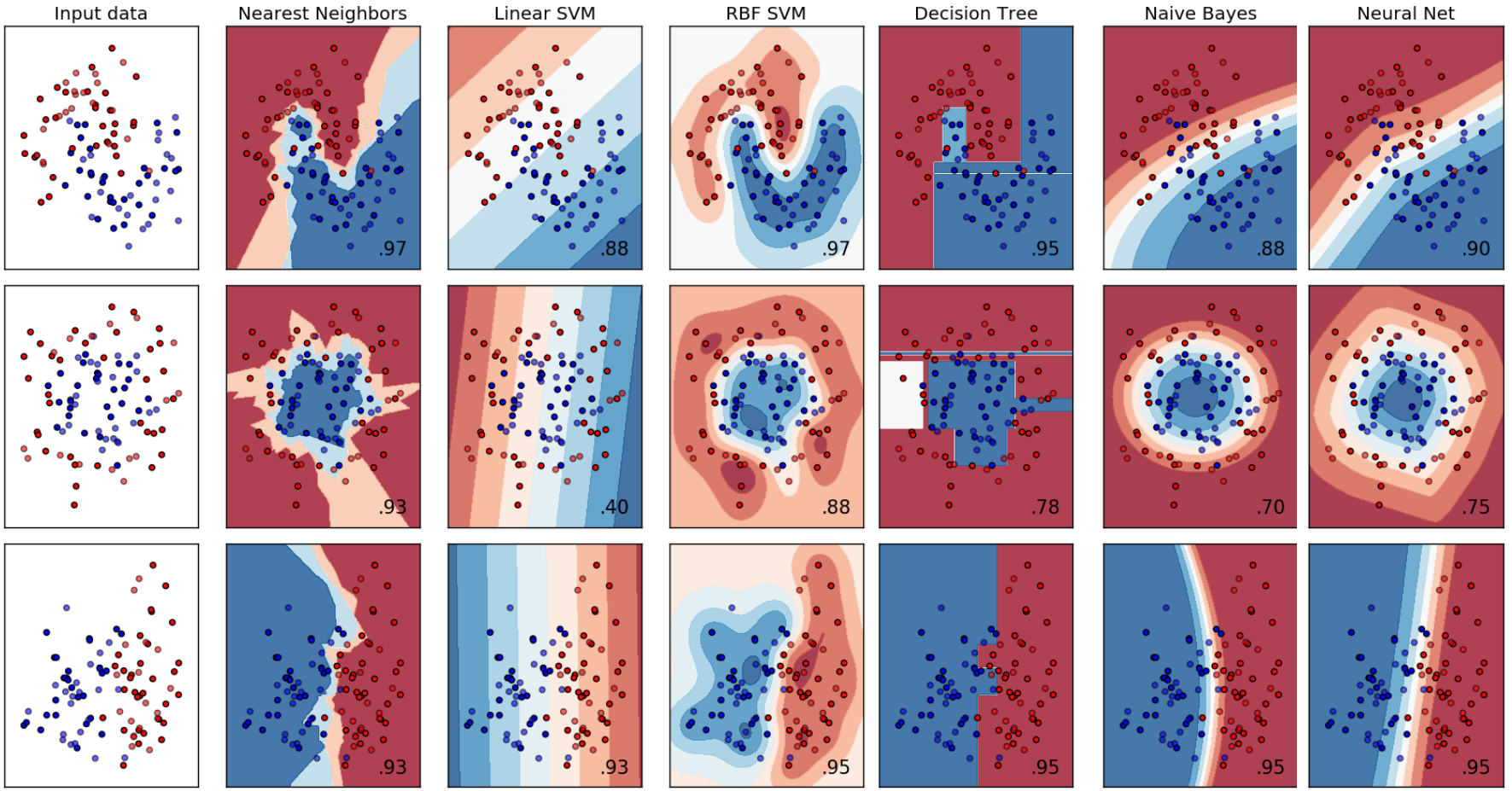
# Datenqualität ist nicht gleich dem Verabeitungsgrad

**Verkürzung der Entwicklungszeit durch Nutzung von Roh-Daten!**



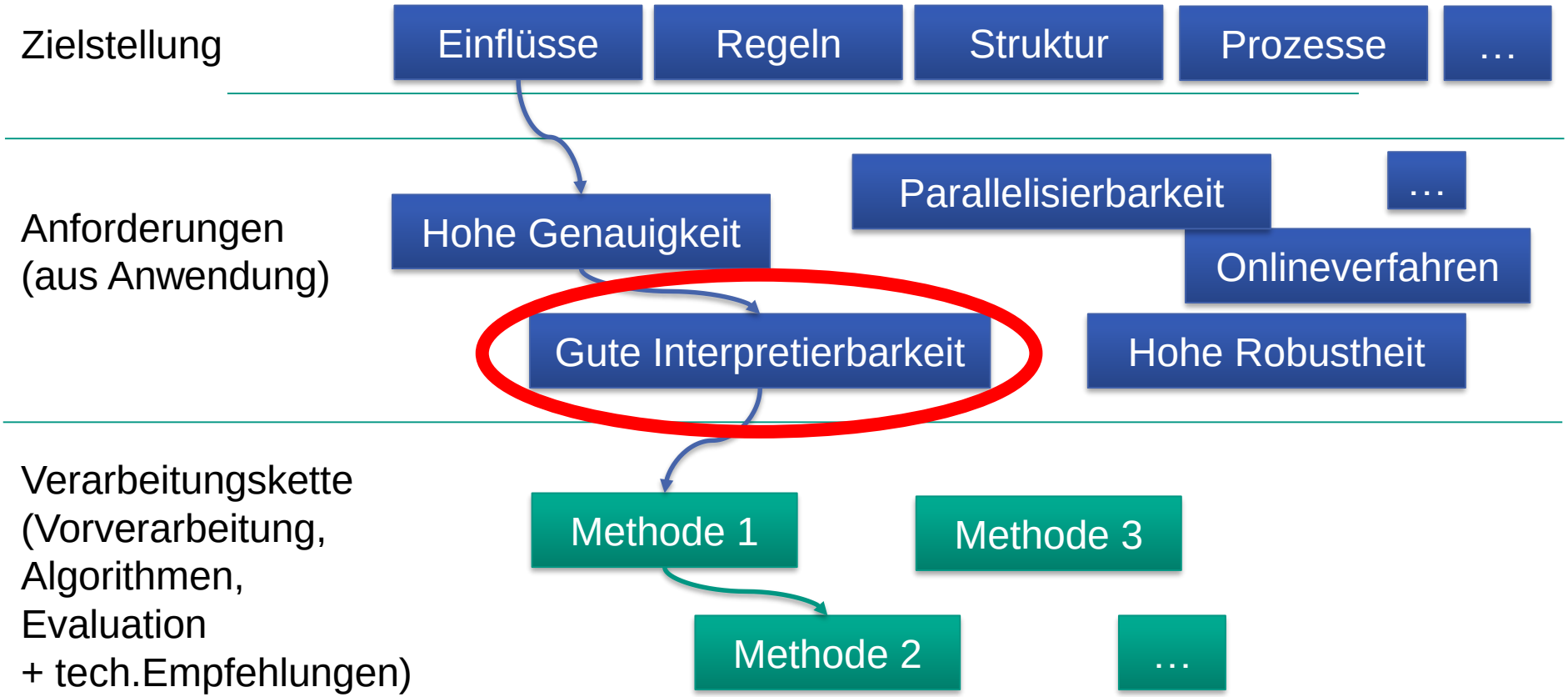



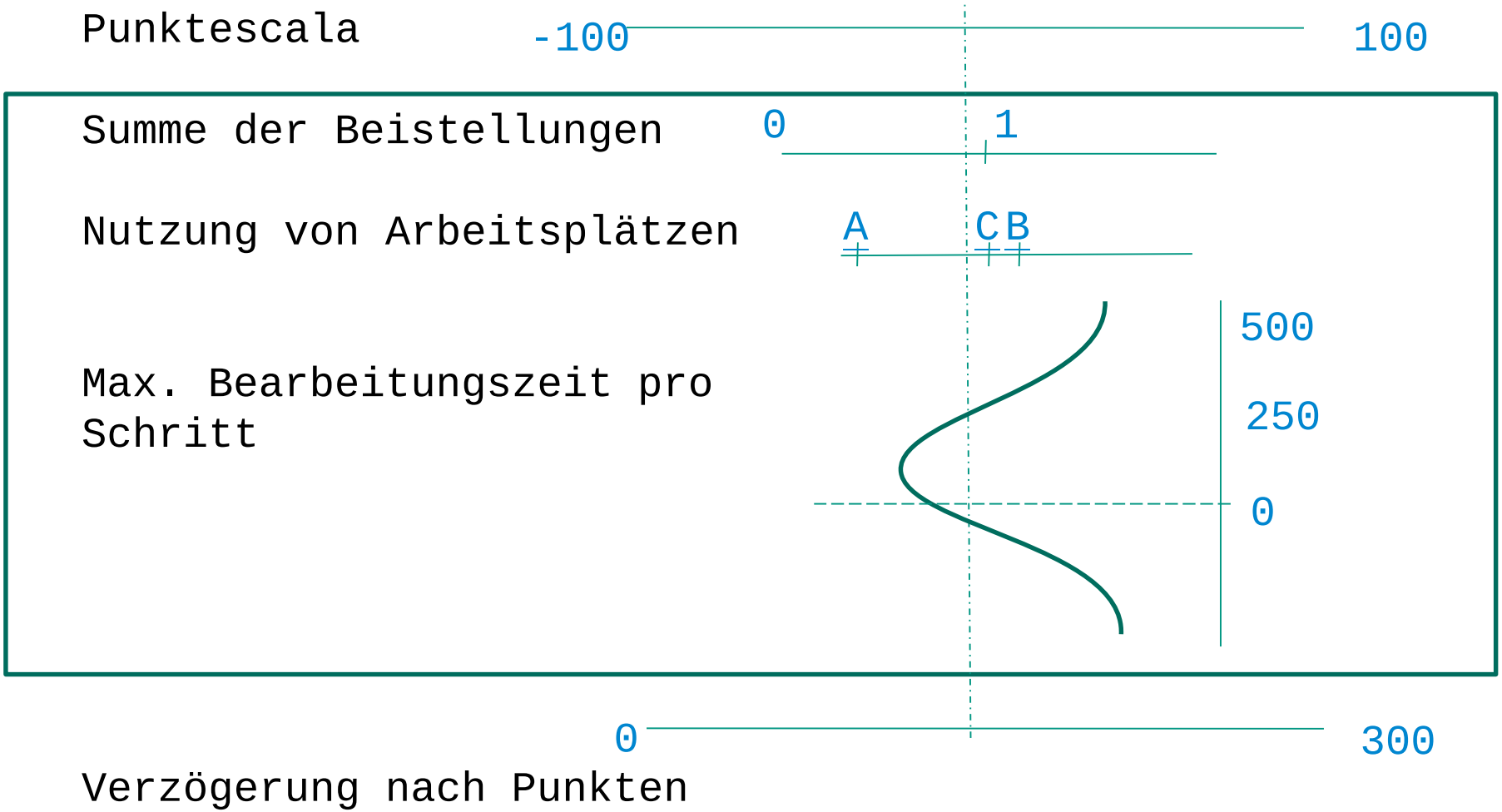
# Eignung der Daten ist leider modellspezifisch...
# ... auch die Anzahl der notwendigen Daten!

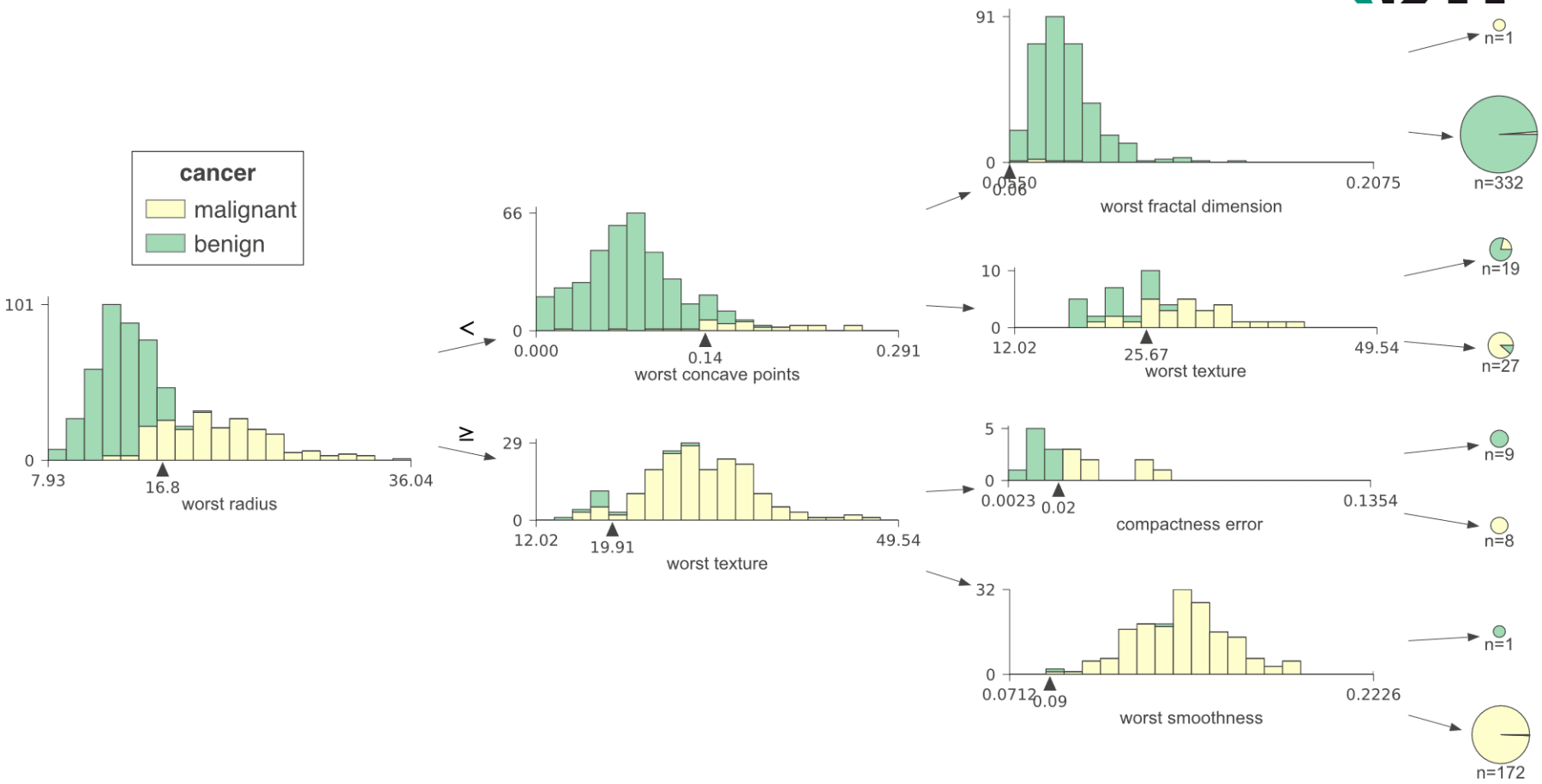



Source:https://scikit-learn.org

# Welche Auswirkung hat die Fragestellung bzw die Domäne auf die Datenauswahl, Vorverarbeitung und die Modellwahl ??

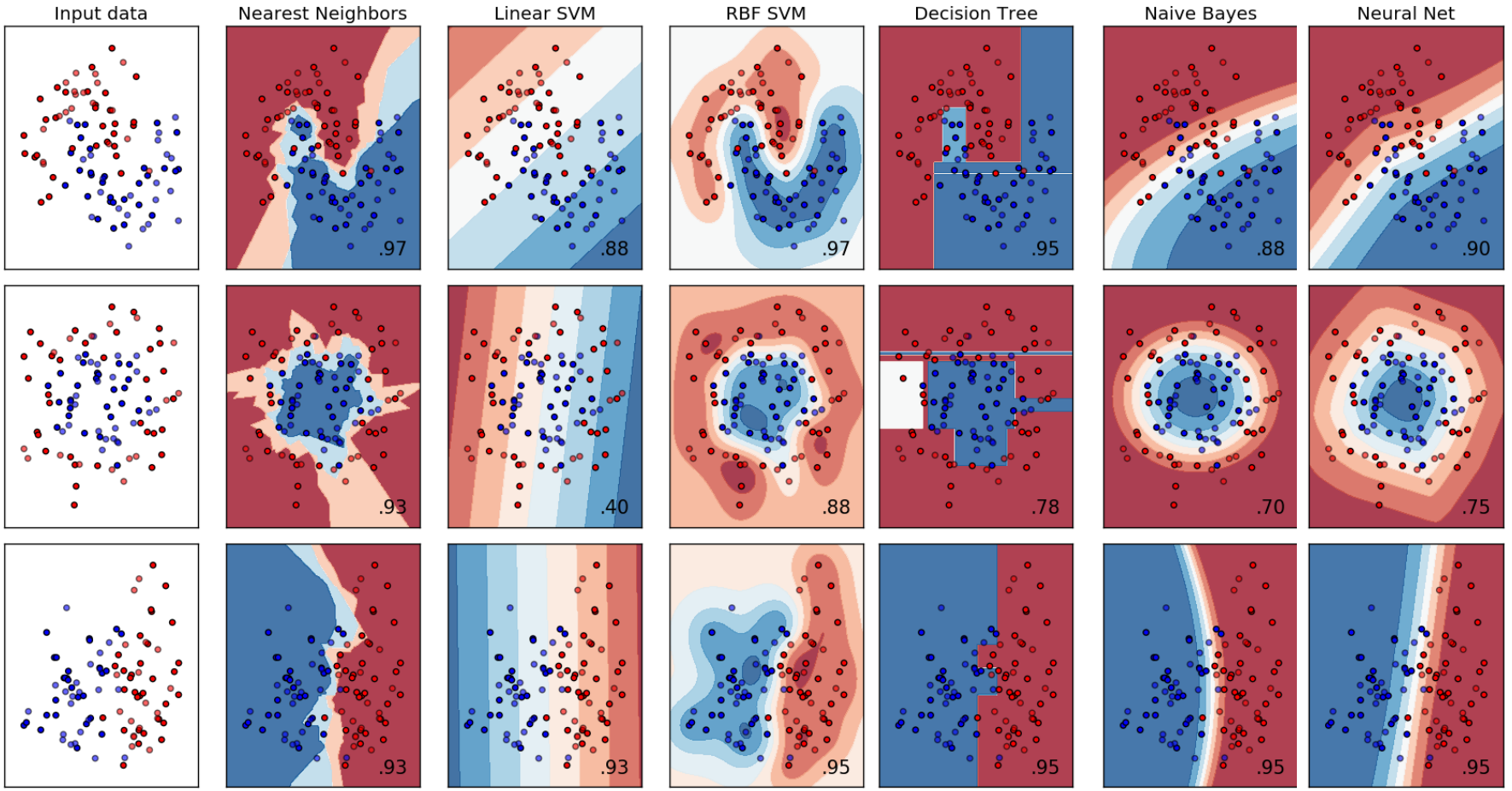



Punktescala
-100
100
Summe der Beistellungen
0
1
Nutzung von Arbeitsplätzen
A
CB
Max. Bearbeitungszeit pro Schritt
500
250
0
0
300
Verzögerung nach Punkten




Source: https://github.com/parrt/dtreeviz

# Eignung der Daten ist leider modellspezifisch...
# ... auch die Anzahl der notwendigen Daten!

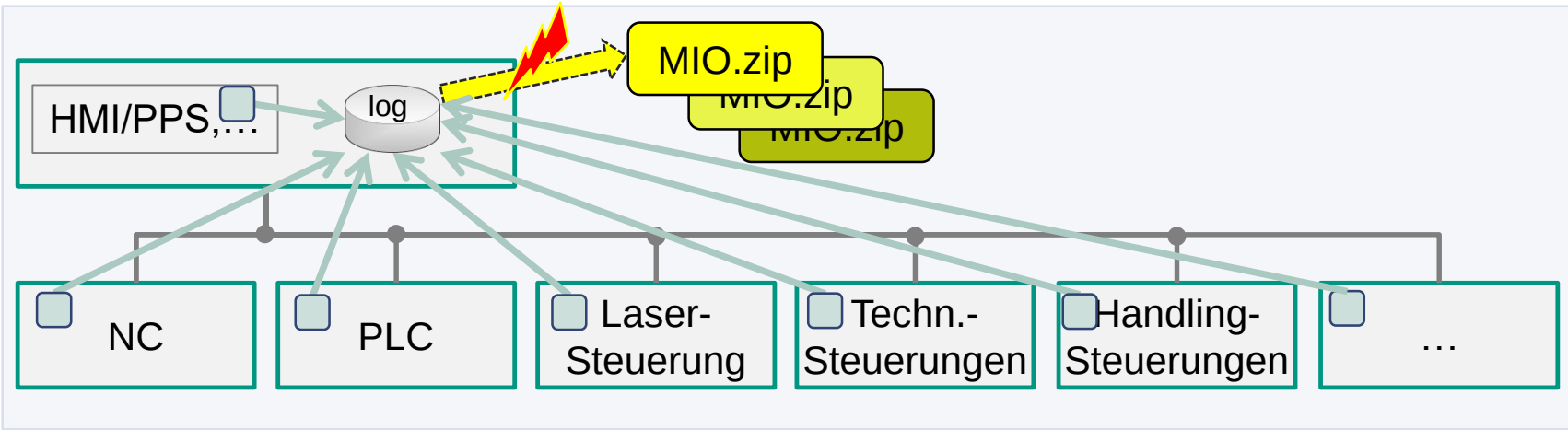



Source:https://scikit-learn.org

# BESSERE DATEN

# MIO (Machine Information Object)

- Snapshot des Maschinenzustandes




MIO.zip

- Loggings
- Konfigurationen
- Software Versionen
- Prozess-/Technologiedaten
- …

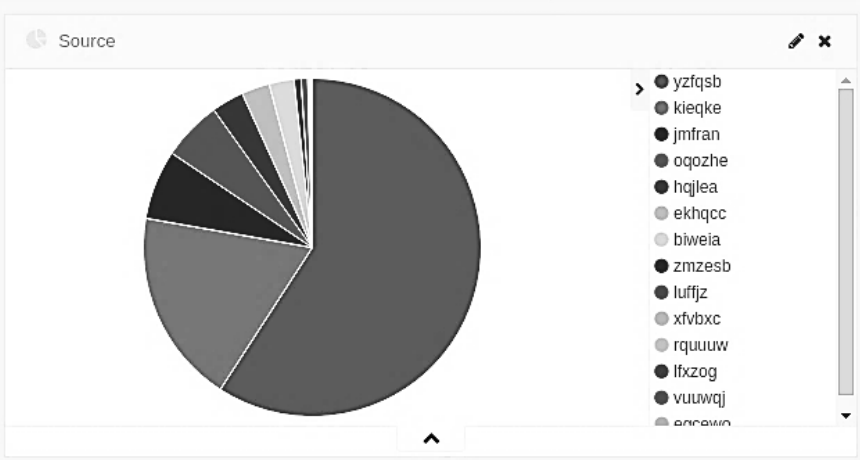
ca. 100 Mbyte Rohdaten
1.000 – 2.000 Einzeldateien
ca. 1 – 2 Mio‘s / Jahr und Maschine
Log-Zeitraum: 1 – 2 Wochen

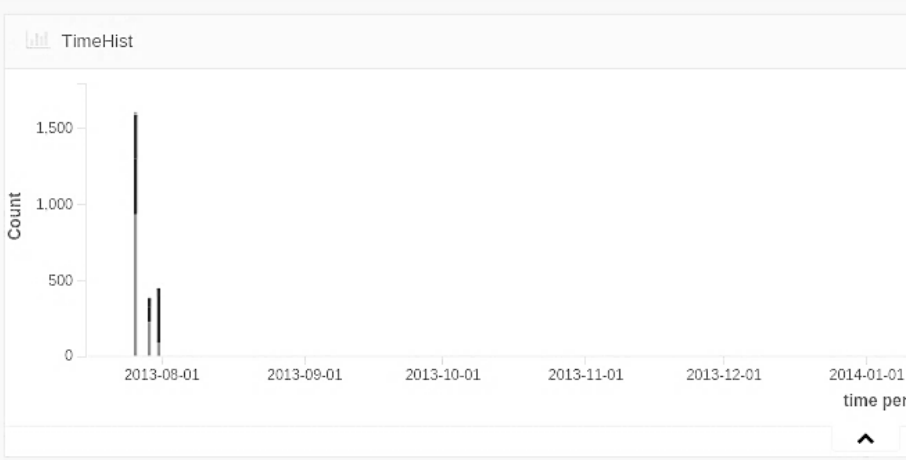
Nicht enthalten sind sensitive Daten, wie z.B.

- Konstruktionsdaten
- Bearbeitungsaufträge
- …

# Sequenzielle Regeln

| | Konfidenz | Support | Vorhersage-horizont |
|---|---|---|---|
| + <{8125885}, {0570745}, {1979269} > → 0456439 | 100% | 10.3% | 77.16s |



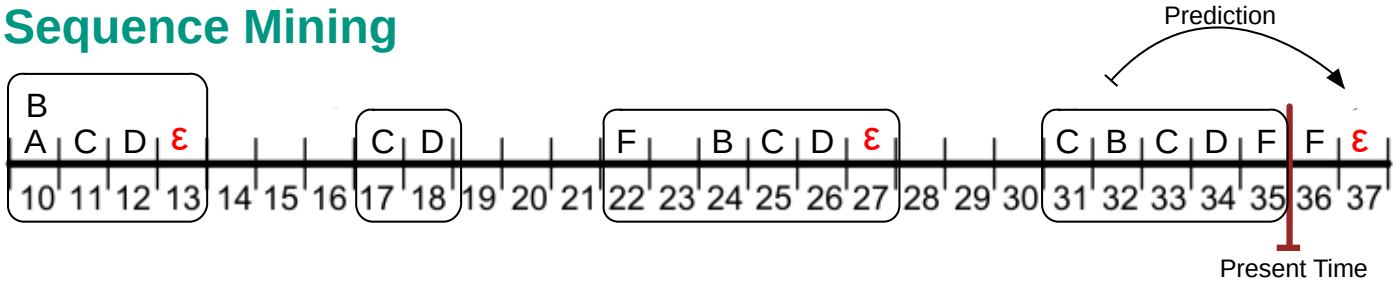




CodeTable

| code: Descending | message: Descending |
|---|---|
| 0570745 | Data transfer SHT-1.SHT |
| 0456439 | Nitrogen input pressure is lower than 27 Bar! |
| 8513211 | Gas flushing is active - flushing time: 3 |

# Praktische Herausforderung bei der Loganalyse

## Sequence Mining




→ Ist extrem komplex, wenn die Daten nicht in eine Reihenfolge gebracht werden können

A B C, A C B, B C A, C A B oder C B A ??

Probleme:

- echte Parallelität
- schlechte Zeitstempel

## Process Mining

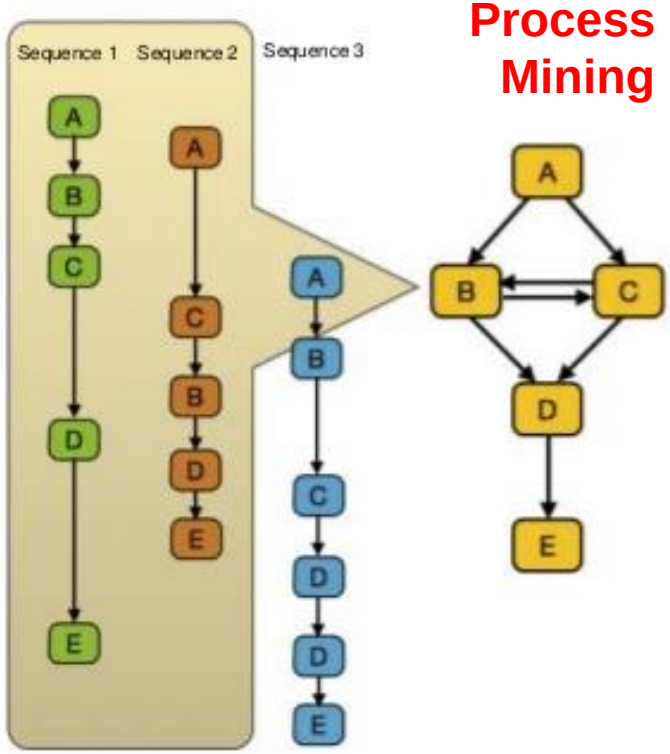


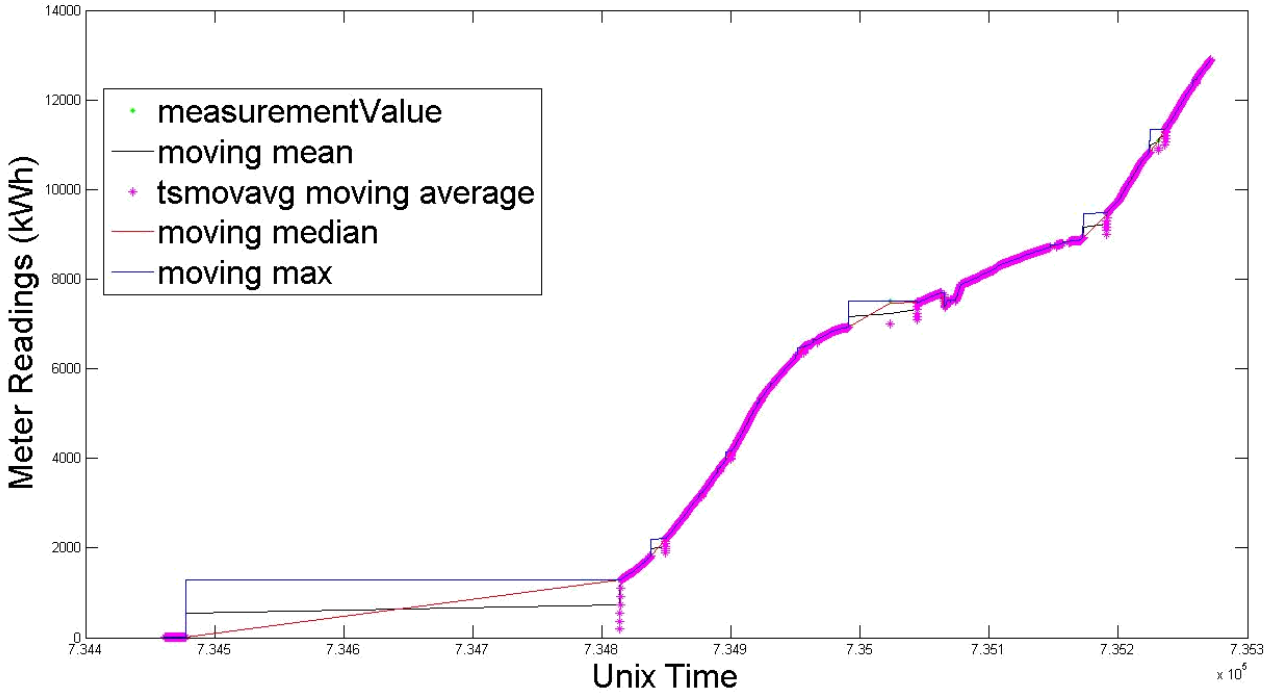

Bild Quelle: IT Service Management Forum - itSMF Österreich

# Ähnliches Problem: Welche Messdatum gehört zu welchem getesteten Produkt


Fügen sie Marker/ Sync ein!

Bei gleichen Durchlaufzeiten:  z.B. Unterbrechungen aufzeichnen!
Testen sie selbst ob sie die Daten zusammenbringen können!!!

# Beispiel fehlende/falsche Daten:



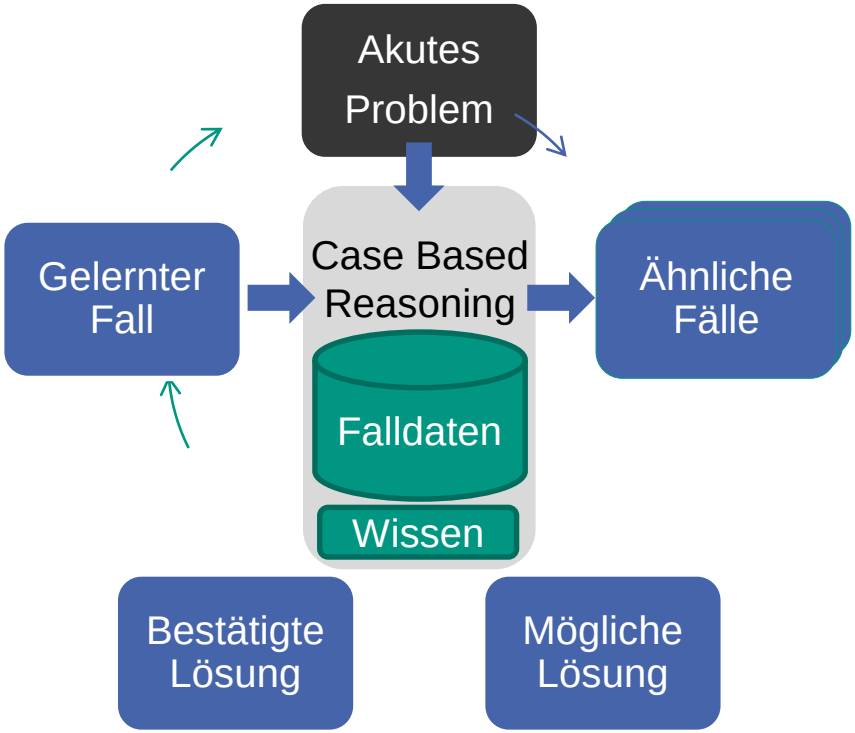

- NOBEL Database (smart meter messungen)

**Fehlende Daten müssen für viele Algorithmen künstlich eingeführt werden → aufwendig und führt zu Fehlern**

# Daten geben erst den möglichen Anwendungspotential vor: Nutzung für Diagnose?

- Daten müssen auch in der Anwendung effizient gesammelt werden

➔opportunistischer Ansatz

# BERATUNG SOWIE AUS- UND WEITERBILDUNGSANGEBOTE FÜR DEN MITTELSTAND

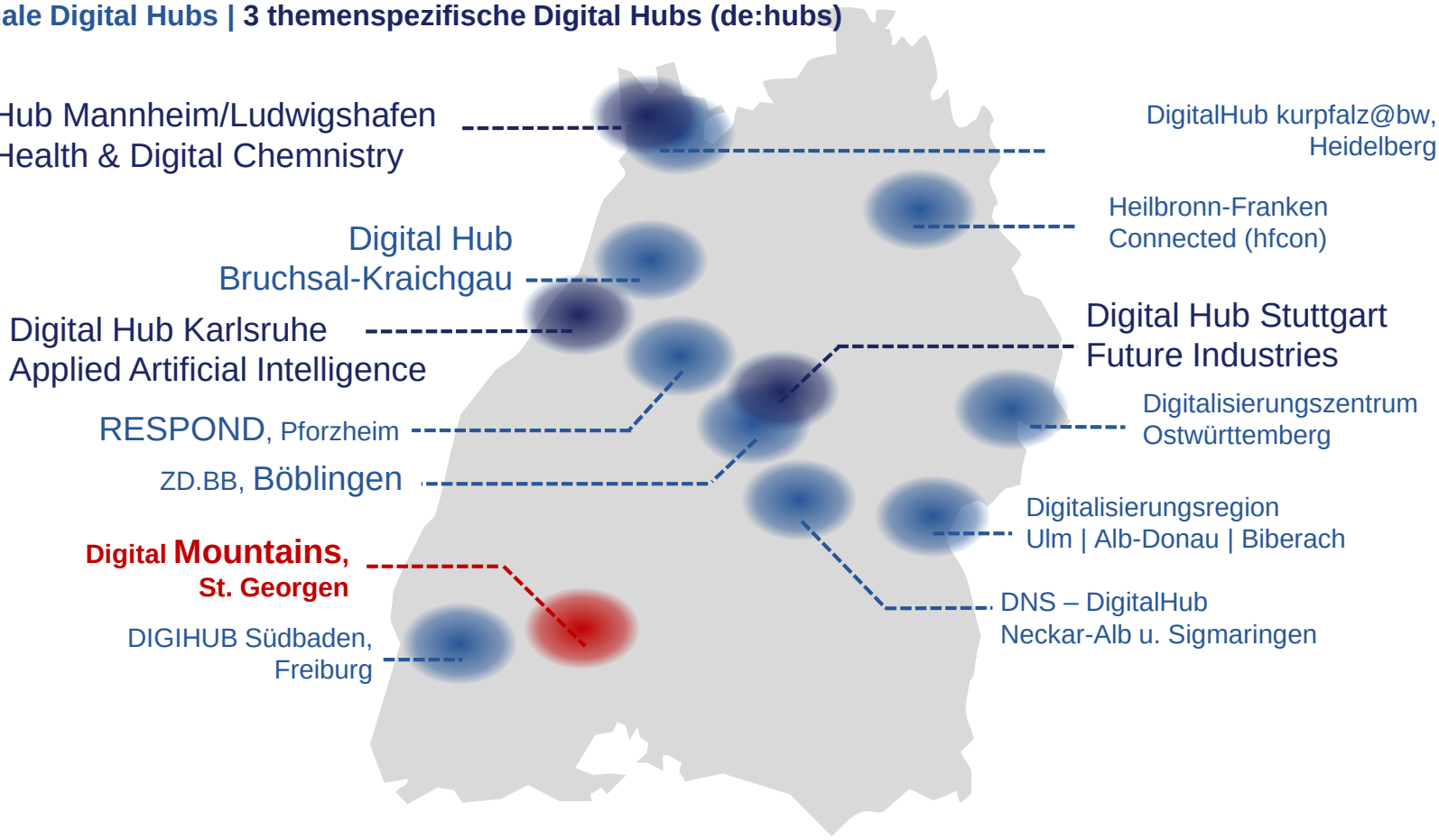
IN 3 MONATEN ZUM/ZUR DATENANALYT*IN

SMART DATA SOLUTION CENTER BADEN-WÜRTTEMBERG | **JETZT AUCH AM STANDORT ST. GEORGEN IM SCHWARZWALD**

DIGITAL Mountains

# ~~Digital Hubs~~ in Baden-Württemberg

**10 regionale Digital Hubs | 3 themenspezifische Digital Hubs (de:hubs)**




IW4.0 Initiative Wirtschaft 4.0 BW

Baden-Württemberg
MINISTERIUM FÜR WIRTSCHAFT, ARBEIT UND WOHNUNGSBAU

digital@bw

# Ihr Digi Hub in der Region: Digital Mountains

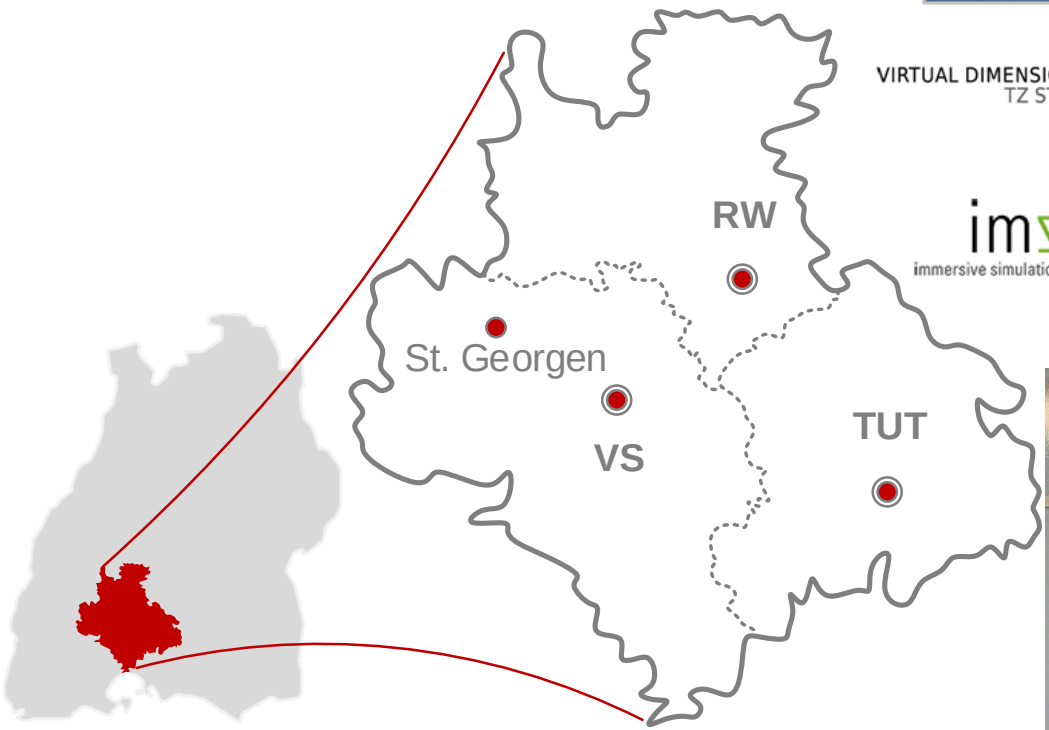


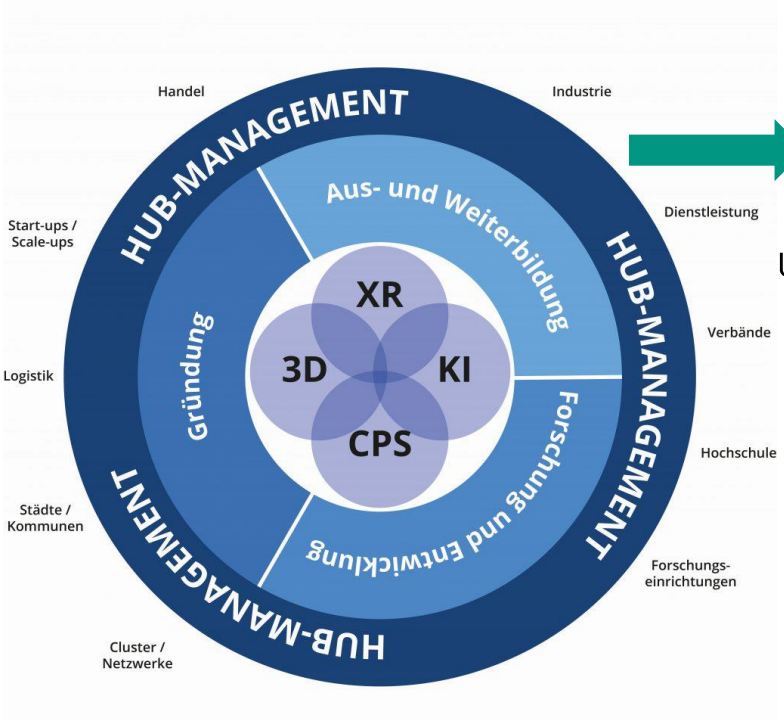


Foto: Schwarzwälder Bote

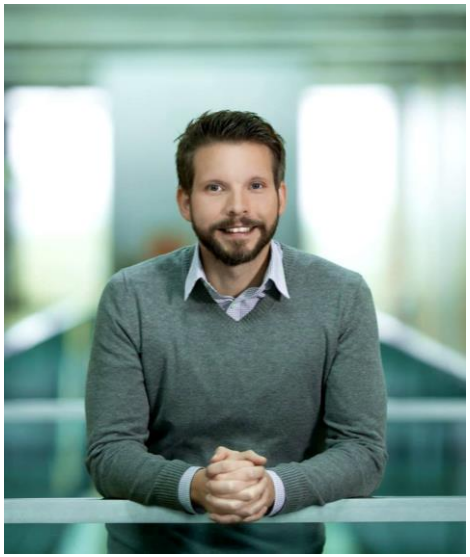
SDSC BW
Smart Data Solution Center Baden-Württemberg

KIT
Karlsruhe Institute of Technology

DIGITAL Mountains




HANDLUNGSFELD
**AUS- UND WEITERBILDUNG**
UNTERSTÜTZTUNG DURCH DAS KIT



IHK
Industrie- und Handelskammer
Schwarzwald-Baar-Heuberg

# DATENANALYST*IN
(ZERTIFIKATSLEHRGANG) IHK

1

## DATENBASIERTE GESCHÄFTSMODELLE UND ENTWICKLUNG VON DATENSTRATEGIEN.

DAUER 4 WOCHEN.*

2

## MIT SMART DATA ZUM BIG BUSINESS.

DAUER 5 WOCHEN.*

3

## ERLÖSORIENTIERTE BEWERTUNG VON MODELLEN.

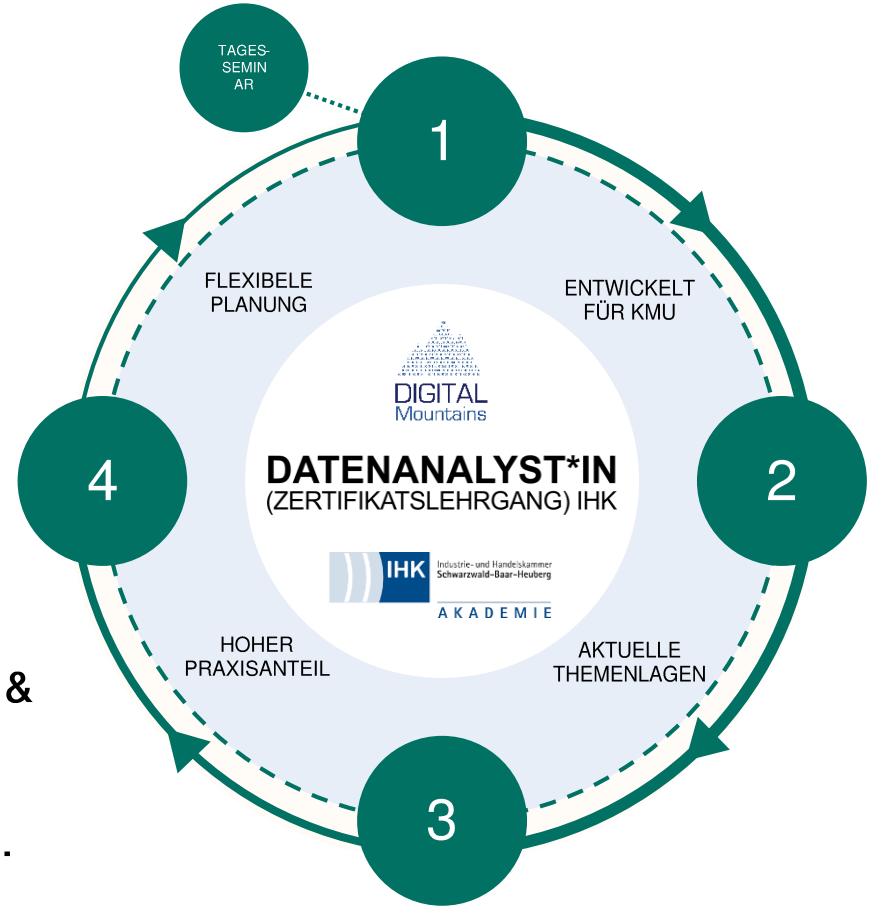
DAUER 4 WOCHEN.*

4

## DATENSICHERHEIT, DATENSCHUTZ & URHEBERRECHT ALS HERAUSFORDERUNGEN FÜR WIRKUNGSVOLLE DATENANALYSEN.

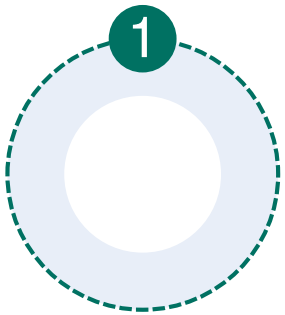
DAUER 1 WOCHE.*

*MODULE SIND EINZELN BUCHBAR, BESUCH VON SUBMODULEN AUF ANFRAGE.




**KERNMODULE** | AUS- UND WEITERBILDUNG ZUM/ZUR DATENANALYST*IN

# Seminarangebot Modul 1 (Start Oktober 2020)




## DATENBASIERTE GESCHÄFTSMODELLE UND ENTWICKLUNG VON DATENSTRATEGIEN.

- WELCHE MÖGLICHKEITEN GIBT ES, UM DATENANALYTISCHE DENKWEISEN ZU AKTIVIEREN / KULTIVIEREN?

  *(WIE SEHEN DIE ZUGRUNDELIEGENDEN DATENORIENTIERTEN GESCHÄFTSPROZESSE AUS?)*

- WIE LÄSST SICH SMART DATA ANALYTICS ALS GESCHÄFTSPROZESS ABBILDEN?

  (WIE IST EIN DATA-MINING STANDARD MODELL AUFGEBAUT?)

- WIE LASSEN SICH MIT ANGEWANDTEN DATENANALYTISCHEN DENKWEISEN WETTBEWERBSVORTEILE ERKENNEN UND VORSPRUNG BEWAHREN?

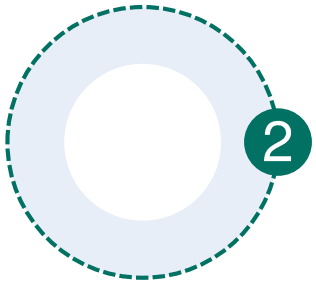
  *(VOR ALLEM UNTER DER BERÜCKSICHTIGUNG ETHISCH-MORALISCHER ASPEKTE?)*

- MIT WELCHEN MITTELN KÖNNEN KMU DAS MANAGEMENT UND DIE TALENTFÖRDERUNG VON DATENANALYSTEN FÖRDERN?

  *(+ WIE KÖNNEN ENTSCHEIDUNGSTRÄGER EINEN GUTEN DATENANALYSTEN ERKENNEN?)*

**SUBMODULE** | AUS- UND WEITERBILDUNG ZUM/ZUR DATENANALYST*IN

# Seminarangebot Modul 2




## MIT SMART DATA ZUM BIG BUSINESS.

- **WIE LÄSST SICH MASCHINELLES LERNEN AUF REALE PROBLEMSTELLUNGEN AUS DER WIRTSCHAFT ANWENDEN?**

  *(WIE LÄSST SICH MASCHINELLES LERNEN IM UNTERNEHMEN AUF BASIS VON OPEN SOURCE TOOLS ANWENDEN?)*

- **WIE WERDEN SINNVOLLE DATEN AGGREGIERT, UM WIRTSCHAFTLICHE AUSSAGEN ZU TREFFEN?**

  (WIE LÄSST SICH SINN AUS DATENMENGEN EXTRAHIEREN?)

- **WIE LÄSST SICH DIE UNTERNEHMERISCHE ENTSCHEIDUNGSFINDUNG DURCH MODERNE ML-MODELLE VERBESSERN?**

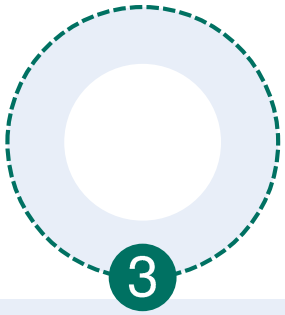
  *(WIE LÄSST SICH GEGENAUIGKEIT BEI VORHERSAGEMODELLEN VERBESSERN?)*

- **WELCHE VISUELLE STRATEGIE IST DIE PASSENDE, UM UNTERNEHMENSFRAGEN AUF BASIS VON ANALYSEERGEBNISSEN ZU BEANTWORTEN?**

  *(WIE LASSEN SICH ANALYSEERGEBNISSE FÜR NICHT-DATENANALYSTEN PRÄSENTIEREN - OHNE ÜBERSIMPLIFIZIERUNG/KOMPLEXITÄT?)*

**SUBMODULE** | AUS- UND WEITERBILDUNG ZUM/ZUR DATENANALYST*IN

# Seminarangebot Modul 3




## ERLÖSORIENTIERTE BEWERTUNG VON MODELLEN.

- **WELCHE MITTEL HAT EIN DATA ANALYST, UM DAS GEEIGNETESTE MODELLE FÜR EINE PROBLEMLÖSUNG ZU ERMITTELN?**

  *(WELCHES MODELL BESITZT DIE GERINGSTEN KOSTEN MIT DEM HÖCHSTEN NUTZEN?)*

- **WIE STRATEGIEN EINER MODELLBEURTEILUNG GIBT ES BEI APPLIKATIONEN IM BEREICH PREDICTIVE MAINTENANCE?**

  (WIE SIEHT DIE ANATOMIE EINER DATENANALYSE AUS BEI DER ERMITTELUNG DER REMAINING USEFUL LIFETIME - RUL?)

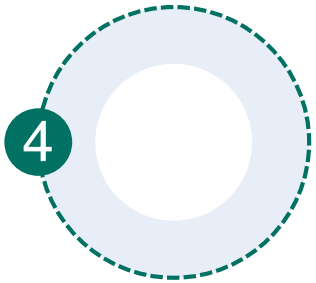
- **WIE LÄSST SICH DIE LEISTUNG EINES MODELL VISUELL VERDEUTLICHEN?**

  *(WELCHE HILFSMITTEL GIBT ES, UM DIE GETROFFENEN ENTSCHEIDUNGEN DEN ENTSCHEIDUNGSTRÄGERN PLAUSIBEL ZU VERMITTELN?)*

- **WIE LASSEN SICH DIE KOSTEN BEI DATENANALYSEN DURCH SMARTE VORGEHENSWEISEN VERRINGERN?**

  *(WELCHE MÖGLICHKEITEN HAT EIN DATENANALYST, UM SELBST MÖGLICHST EFFIZIENT ZU ARBEITEN?)*

# Seminarangebot Modul 4




## DATENSICHERHEIT, DATENSCHUTZ & URHEBERRECHT ALS HERAUSFORDERUNGEN FÜR WIRKUNGSVOLLE DATENANALYSEN.

- WANN IST DIE ERHEBUNG, VERARBEITUNG UND NUTZUNG VON PERSONENBEZOGENEN DATEN ÜBERHAUPT ZULÄSSIG?

  *(REICHEN GÄNGIGE ANONYMISIERUNGSVERFAHREN AUS, WIE STARK MUSS ANONYMISIERT WERDEN?)*

- DÜRFEN DATEN VORBEHALTLOS GESAMMELT WERDEN ZUR MODELLBILDUNG?

  *(WELCHE VORKEHRUNGEN MÜSSEN GETROFFEN WERDEN BEI VORRATSDATENSPEICHERUNGEN VON FREI VERFÜGBAREN DATEN?)*

- WELCHE TECHNISCHEN UND ORGANISATORISCHEN GESTALTUNGSMÖGLICHKEITEN GIBT ES, UM DATENPANNEN/DATENLEAKS ZU VERHINDERN?

  *(WIRKEN SICH DATENSPARSAMKEIT UND DATENVERMEIDUNG KONTRAPRODUKTIV AUF ANALYSEN AUS?)*

- UNTER WELCHEN RECHTLICHEN BEDINGUNGEN DÜRFEN MASSENDATEN GENUTZT WERDEN?

  *(+ WIE LASSEN SICH REGELUNGEN SEITENS DER DATENHOHEIT GEZIELT IMPLEMENTIEREN?)*

**SUBMODULE** | AUS- UND WEITERBILDUNG ZUM/ZUR DATENANALYST*IN

SDSC BW
Smart Data Solution Center Baden-Württemberg

KIT
Karlsruhe Institute of Technology

DIGITAL Mountains

IHK
Industrie- und Handelskammer Schwarzwald-Baar-Heuberg

**Schreiben Sie uns!!**
**Und besuchen Sie uns später im Jahr in St. Georgen!!**

# DATENANALYST*IN
(ZERTIFIKATSLEHRGANG) IHK


Ansprechpartner SDSC-BW:
Murat Malyemez info@sdsc-bw.de

Organisatorischer Ansprechpartner Weiterbildung:
Simone Mader mader@vs.ihk.de

Weitergehende Fragen/ Forschung
Till Riedel riedel@kit.edu

Inhaltlicher Ansprechpartner Weiterbildung / SDSC Büro St. Georgen:
Rainer Duda duda@kit.edu


**ES GIBT NOCH FREIE PLÄTZE!!**

Karlsruhe Institute of Technology

# Danke für Ihre Aufmerksamkeit

SDSC BW
Smart Data Solution Center Baden-Württemberg