



# Data Mining Projekte im Unternehmen

Prof. Dr. Robert Hable







# Industrie 4.0

Maschinen und Sensoren produzieren einen riesigen Datenstrom

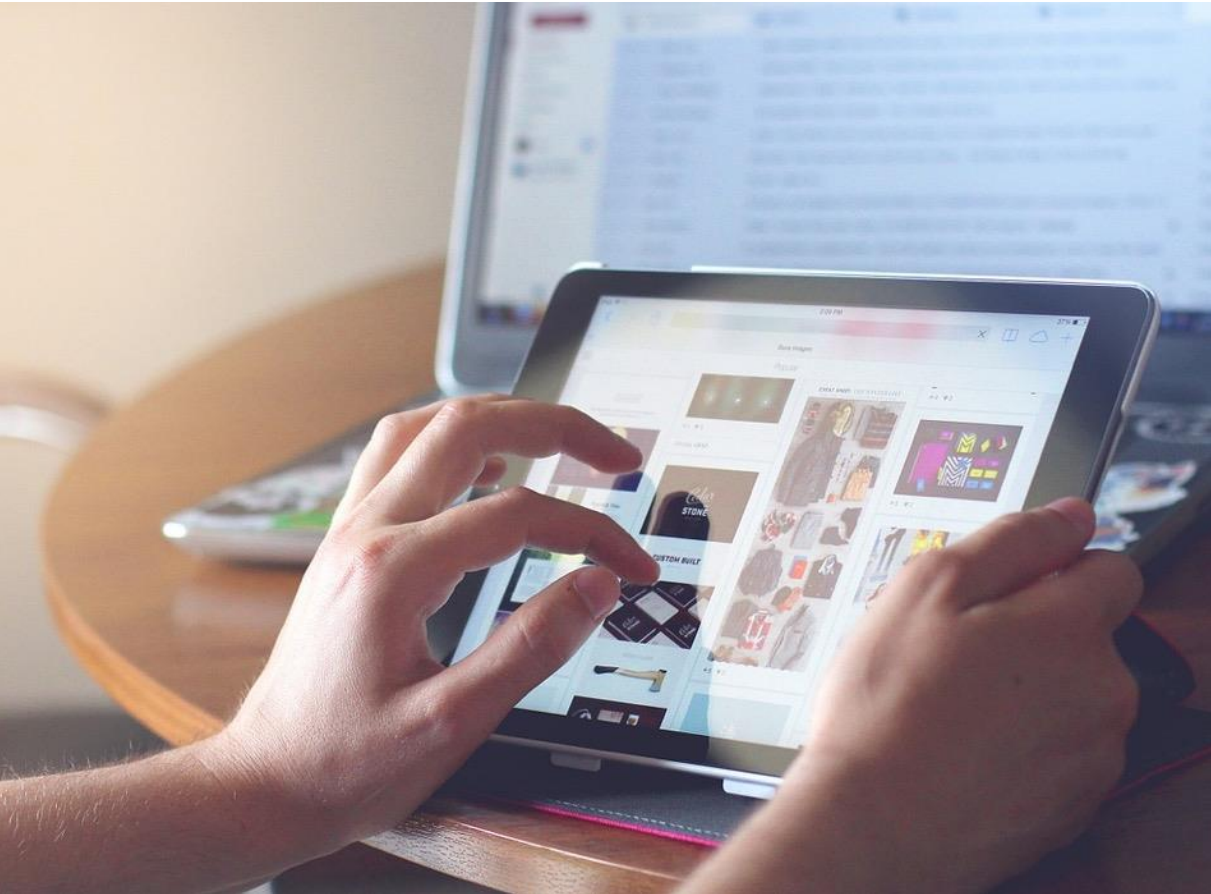
---



# Web 4.0

Internetnutzer produzieren massenhaft komplexe Daten (Texte, Klick-Verlauf).

---





# Finanzdaten

Automatische Datenerhebung durch elektronischen Zahlungsverkehr


---



# Unternehmensdaten

Unternehmen jeder Größe und Branche verfügen über Datenbestände

---



**Datenbestände  
lassen sich systematisch  
analysieren und verwerten.  
Z.B. Excel-Tabellen  
zu Lagerhaltung, Einkauf,  
Auftragseingängen,  
etc.**

---



# Daten sind das Öl des 21. Jahrhunderts ...

---



# Daten sind das Öl des 21. Jahrhunderts ...

... aber brauchen wir nicht eigentlich Benzin?

---





# Big Data Grundbausteine



# Big Data

5 V's: Volume, Variety, Velocity, Value, Validity

---



## Datenmanagement

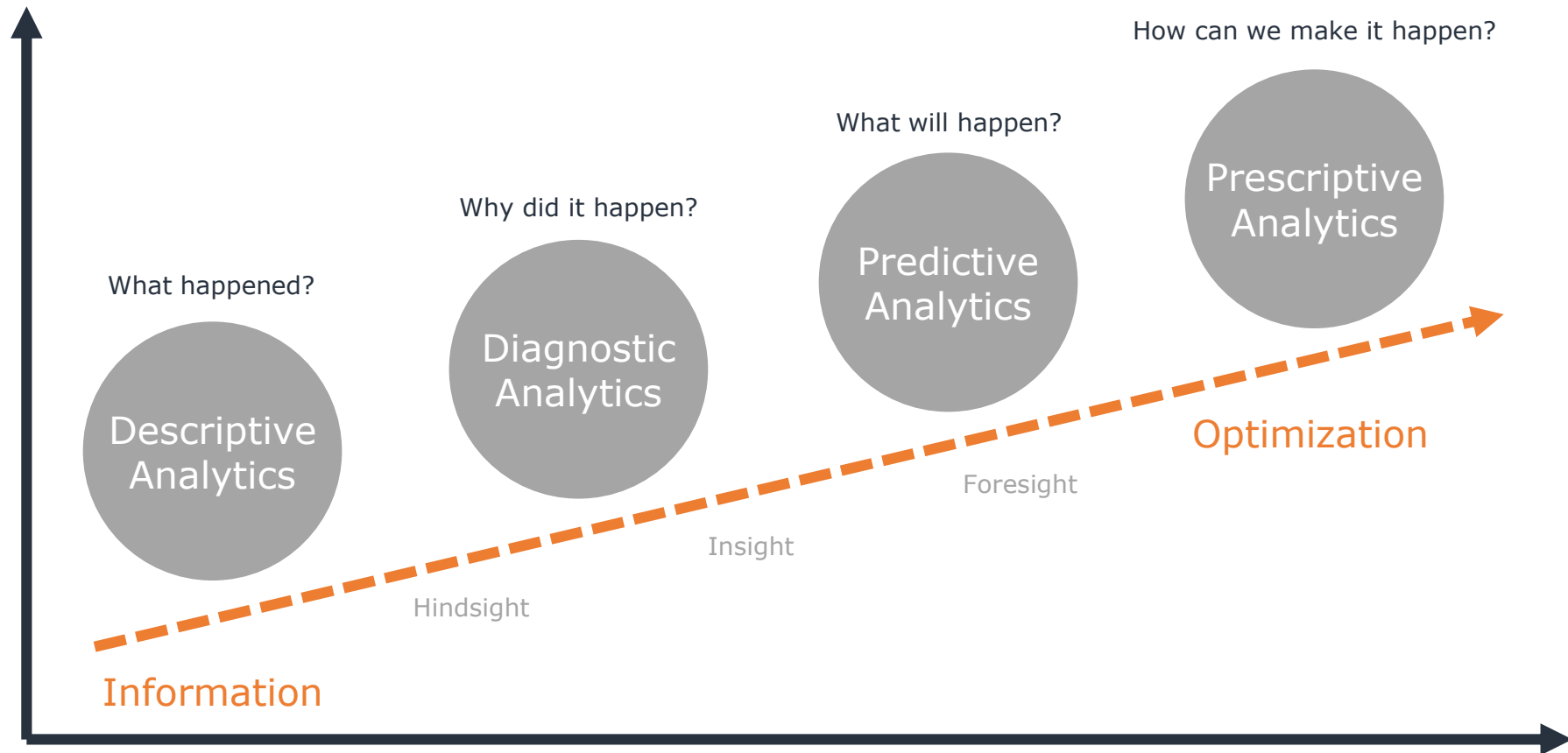
Daten sammeln,  
speichern, verfügbar  
machen

## Datenanalyse

Daten visualisieren,  
auswerten, produktiv  
nutzen



# Verschiedene Arten von Datenanalysen



# Verschiedene Arten von Datenanalysen?



## Elektronische Datenerfassung

- Wieviel von Produkt A wurde in Filiale X am 21.6.2016 verkauft?



## Datenbanken

- Wieviel von Produkt A wurde jeweils in den Filialen X, Y und Z in Woche 25 verkauft?



## Komplexe Datenanalysen

- Warum wird von Produkt A in Filiale X mehr verkauft als in Filiale Y ?
- Wieviel von Produkt A wird nächste Woche in Filiale X verkauft werden?
- Wieviel von Produkt A soll Filiale X auf Lager haben?



# Data Analytics – vielfältige Einsatzgebiete

## Marketing

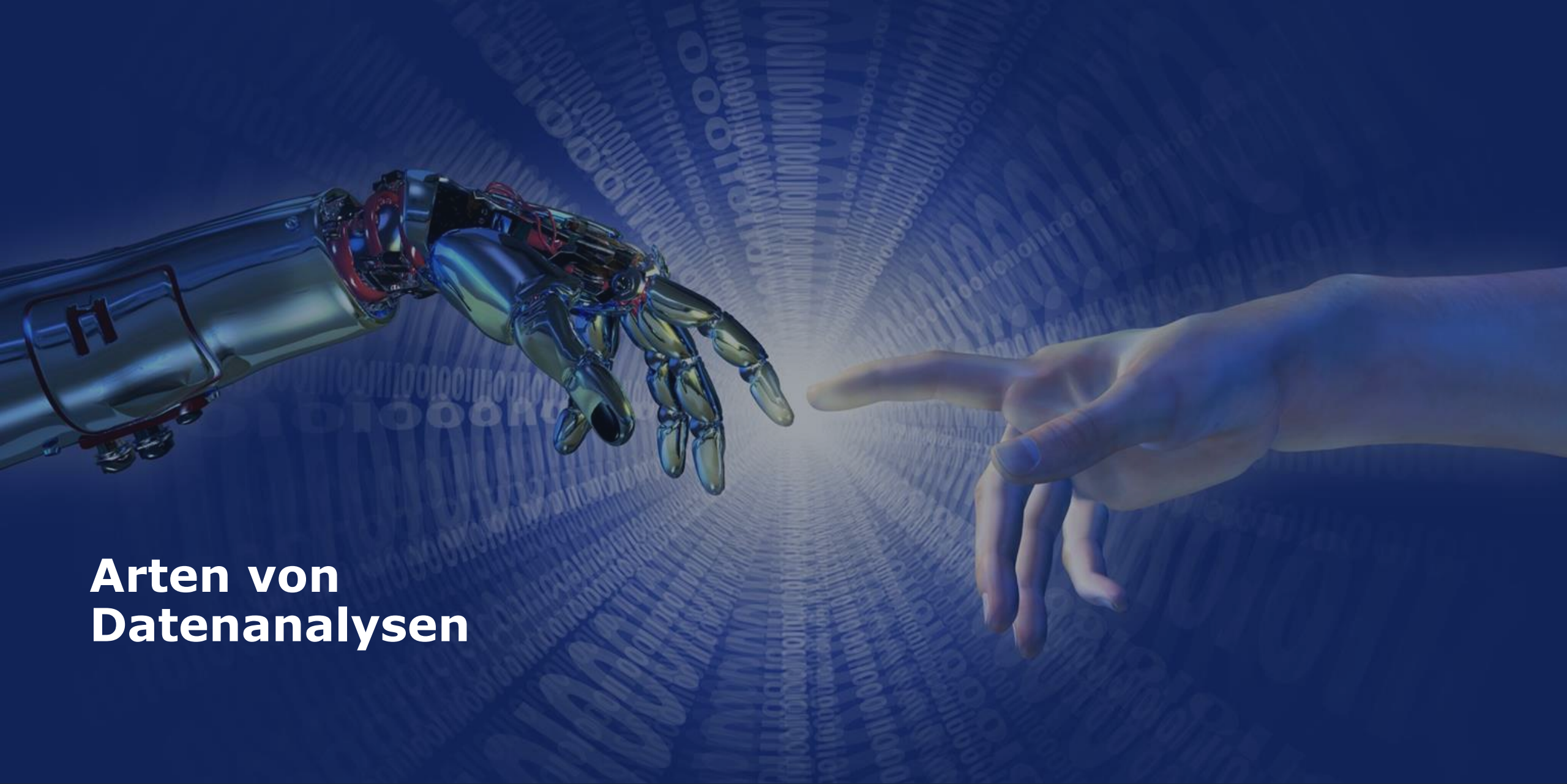
- Prognosen von Absatz und Kundenabrufen zur verbesserten Planung und Steuerung
- Effizienzsteigerung in Marketing und Vertrieb durch individualisiertes Marketing und objektive Bewertung von Maßnahmen

## Produktion

- Vorausschauende Instandhaltung für bedarfsgerechte und dynamische Wartung von Maschinenteilen
- Prozessüberwachung zur Früherkennung von Maschinenausfällen
- Erhöhung der Ressourceneffizienz durch Ermittlung von Ursachen hoher Verbräuche

## Finance

- Betrugserkennung durch Analyse von Buchungsdaten



# Arten von Datenanalysen



# Arten von Datenanalysen /I

## Überwachte Lernprobleme

---

**Einflussgröße**



**Zielgröße**

- Modellierung des Zusammenhangs zwischen ein oder mehreren Einflussgrößen und einer Zielgröße
- Prognose von Werten der Zielgröße basierend auf den Werten der Einflussgrößen

### Beispiele

- Welcher Artikel verkauft sich wie oft?
- Welche Sensorwerte deuten auf eine bevorstehende Maschinenstörung hin?

# Arten von Datenanalysen /I

## Überwachte Lernprobleme

---

**Einflussgröße**



**Zielgröße**

- Modellierung des Zusammenhangs zwischen ein oder mehreren Einflussgrößen und einer Zielgröße
- Prognose von Werten der Zielgröße basierend auf den Werten der Einflussgrößen

### Beispiele

- Welcher Artikel verkauft sich wie oft?
- Welche Sensorwerte deuten auf eine bevorstehende Maschinenstörung hin?

### Häufiges Problem

- Daten nur für Einflussgrößen (Artikelbeschreibung, Sensorwerte, Buchungsdaten) vorhanden, nicht aber für die Zielgröße

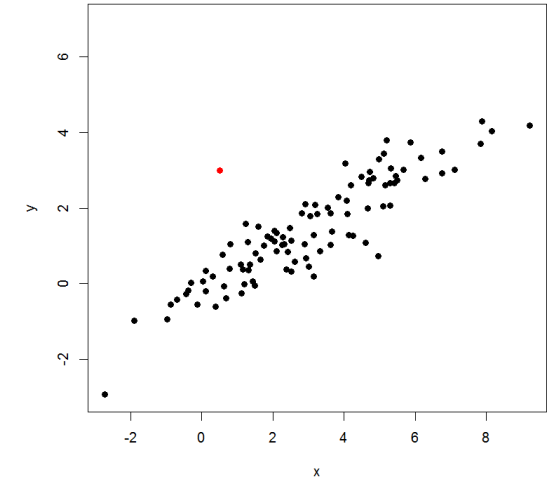


# Arten von Datenanalysen /II

## Überwachte Lernprobleme

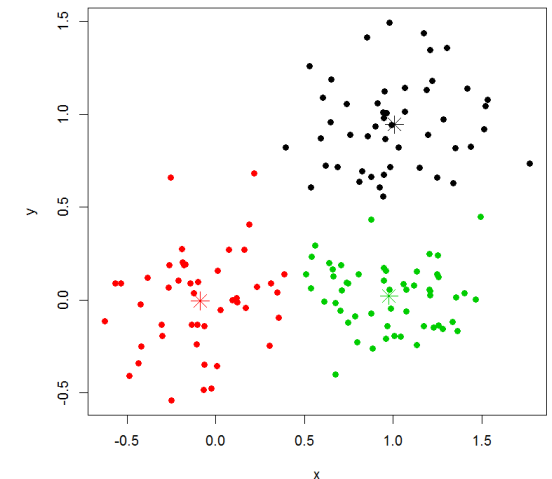
### Anomalie-Erkennung

- Betrugserkennung
- Warnsysteme



### Clusteranalyse

- Bsp. Marketing:  
Bilden von Kundengruppen
- Datensegmentierung  
(Reduktion der Datenkomplexität)

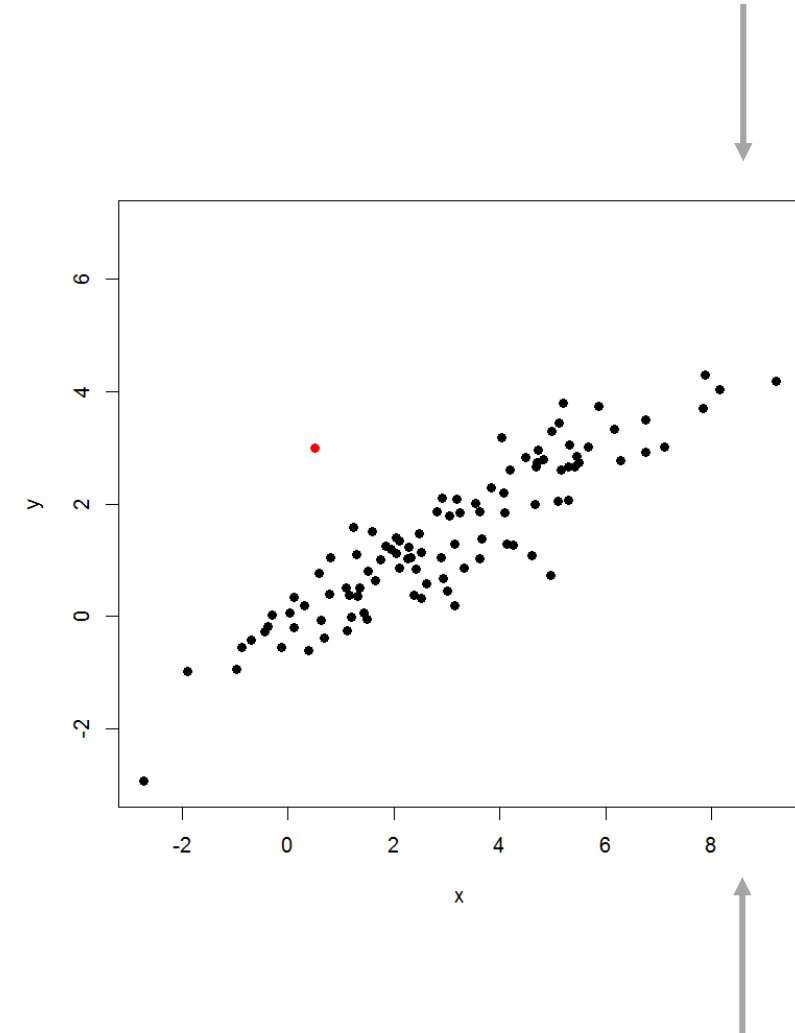


# Arten von Datenanalysen /III

## Warnsysteme durch Anomalie-Erkennung

Bei den Anschlägen auf das World Trade Center am 11. September 2001 hatten unter 80 Flugpassagieren 5 eine **ungewöhnliche Kombination von Merkmalen**:

- eine ausländische Staatsbürgerschaft (alle dieselbe)
- mit One-Way-Ticket unterwegs
- die Tickets am Gate gekauft und in bar bezahlt
- kein Gepäck aufgegeben





# Arten von Datenanalysen /IV

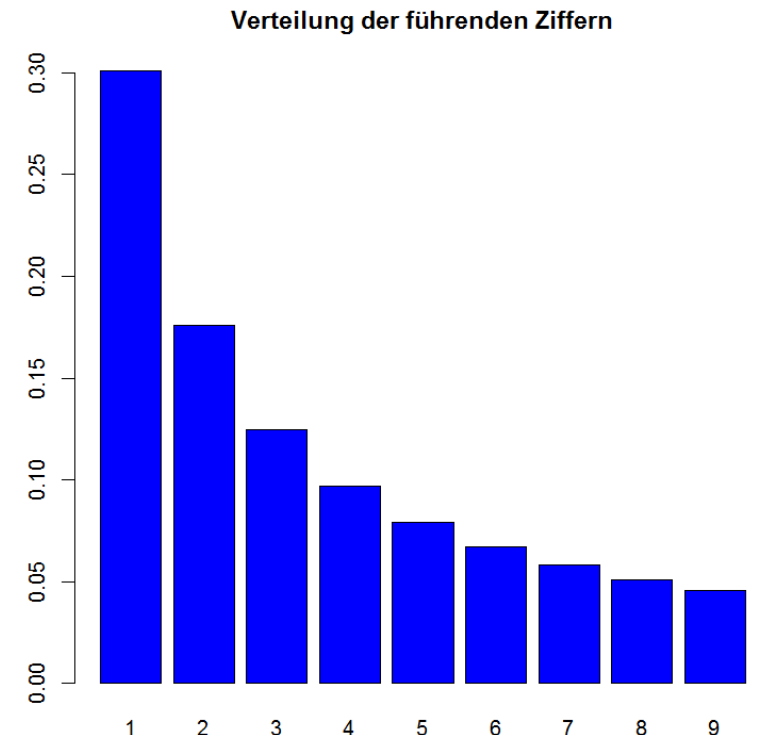
## Betrugserkennung durch Benford's Law

### Hintergrund

- Bei vielen Datensätzen folgt die Häufigkeit der führenden Ziffer bei Zahlen der Benford-Verteilung.
- Grund: Viele Datensätze sind log-normalverteilt (kleinere Zahlen kommen dann häufiger vor als größere).
- Abweichungen von der Benford-Verteilung können ein Indiz für Betrug sein.

### Prominente Beispiele

- Aufdeckung des Enron-Skandals
- Manipulation der Wirtschaftsdaten Griechenlands bei Euroeinführung







# Praxis- beispiele

---





# Praxisbeispiel Modeversandhandel/I



## Aufgabe



## Herausforderungen

Die Modeversandhandelsfirma

- produziert Artikel nicht selbst,
- sondern bestellt Artikel in Asien,
- und zwar ca. 6 Monate vor Katalogversand

- große Anzahl an unterschiedlichen Produkten,
- nahezu vollständiger Sortimentswechsel in jeder Saison,
- starke zeitliche Veränderungen (Modetrends),
- lange Produktions- und Lieferzeiten

# Praxisbeispiel Modeversandhandel/II

## Erfolgsfaktor Absatzprognosen

---

### Zu hohe Einkaufsmengen



- Preisabschritten
- gebundenes Kapital
- Ressourcenvernichtung

### Zu niedrige Einkaufsmengen



- Umsatzverlust
- Ergebnisverlust
- Kundenverärgerung

# Praxisbeispiel Modeversandhandel/III

Absatzprognosen: Mensch vs. Maschine

---

## Bisher

Absatzprognosen durch

- einfache Hochrechnungen
- Experteneinschätzungen



**UNSER ANSATZ**

---

## Absatzprognosen durch

- Datenanalyse
- künstliche Intelligenz: maschinelles Lernen



# Praxisbeispiel Modeversandhandel/IV

Absatzprognosen: Mensch vs. Maschine

---



**ERGEBNIS**

---



**Algorithmus um 18% besser als Modeexperten**

## Beispiel

- Wareneinkauf von 20 Mio. Euro pro Jahr
- Wert der hiervon nicht verkauften Waren: 6 Mio. Euro
- Datenanalyse vermeidet jedes Jahr Fehleinkauf im Wert von ca. 1 Mio. Euro

# Praxisbeispiel Elektromobilität/I



## Projekt e-Wald



## Herausforderungen

### Elektromobilität Bayerischer Wald

- Verbundforschungsprojekt unter Leitung der TH Deggendorf
- größtes Demonstrationsprojekt für Elektromobilität in Deutschland
- Entwicklung intelligenter Ladeinfrastruktur und innovativer Steuerungskonzepte

- Elektroautos besitzen noch immer eine deutlich geringere Reichweite als herkömmliche Autos



Zuverlässige Abschätzung der Reichweite notwendig



# Praxisbeispiel Elektromobilität/II

## Reichweitenschätzung durch Datenanalyse

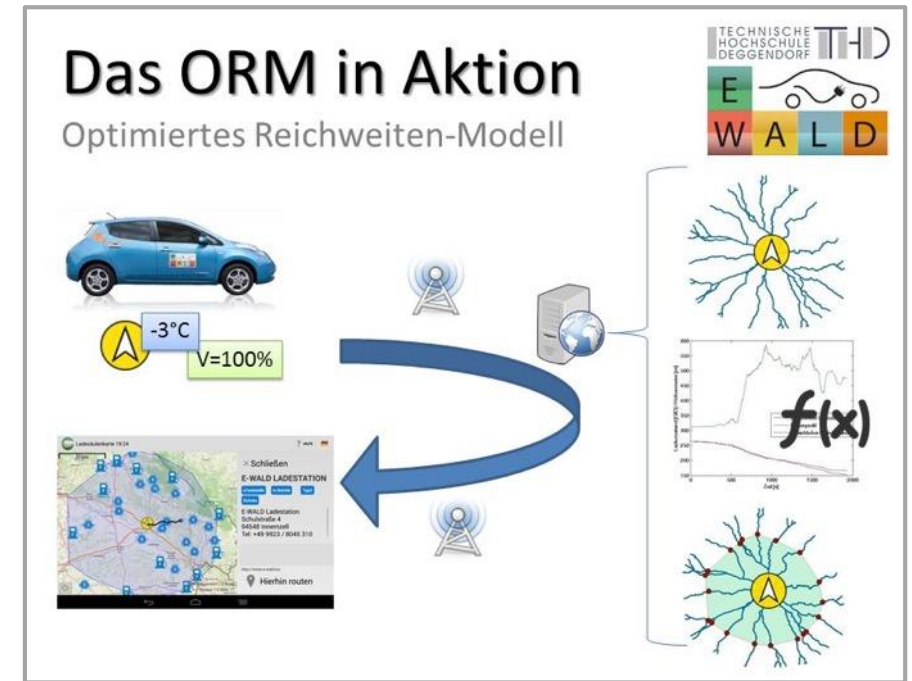
### Datengrundlage

echte Fahrdaten statt Testfahrten:

- Nutzerdaten aus E-Wald
- Datenerhebung im Sekundentakt bei jeder Fahrt

### Datenanalysen

Analyse der Fahrdaten mit Statistik  
und Methoden der künstlichen Intelligenz



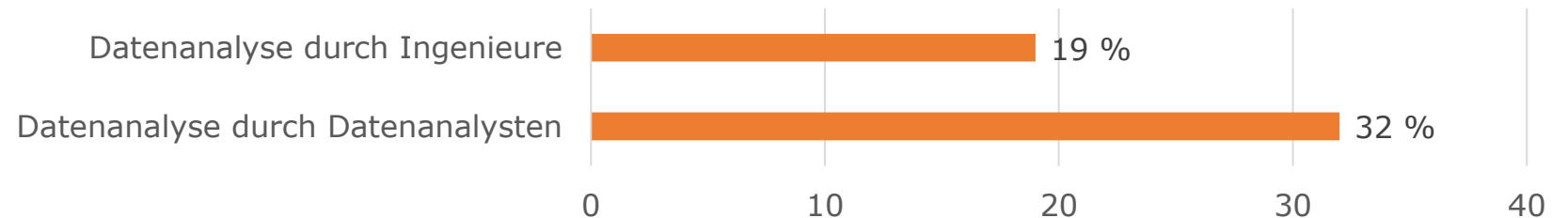


# Praxisbeispiel Elektromobilität/III

Reichweitenschätzung durch Datenanalyse



durch Datenerhebung und -analyse  
erzielte Verbesserung



# Praxisbeispiel Stapler/I



## Vorhandene Daten



## Erste Schritte

Fahraufträge für Stapler im Werk, etwa

- Zeitpunkt des Auftragseingangs
- Zeitpunkt der Auftragsannahme und der Beendigung des Auftrags
- Abhol- und Zielort

- Ermittlung von Hauptrouten
- Schwankungen der Auslastung (tages- und uhrzeitabhängig)
- Berechnung einer Fahrzeitmatrix



# Praxisbeispiel Stapler/II

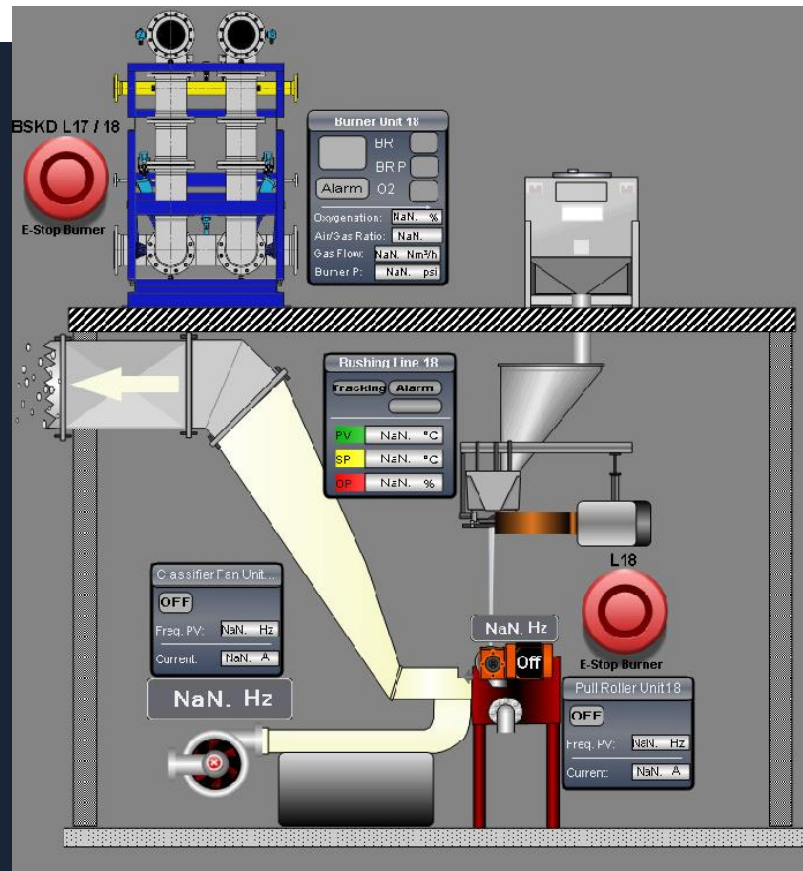


- **Analyse der Fahrten mit langer Erfüllungsdauer: Identifikation von Risikostrecken und -zeitpunkten**
- **Berechnung von Echtzeit-Kennzahlen zur Kapazitätsauslastung (prädiktiv)**



# Praxisbeispiel Schmelzen von Glas

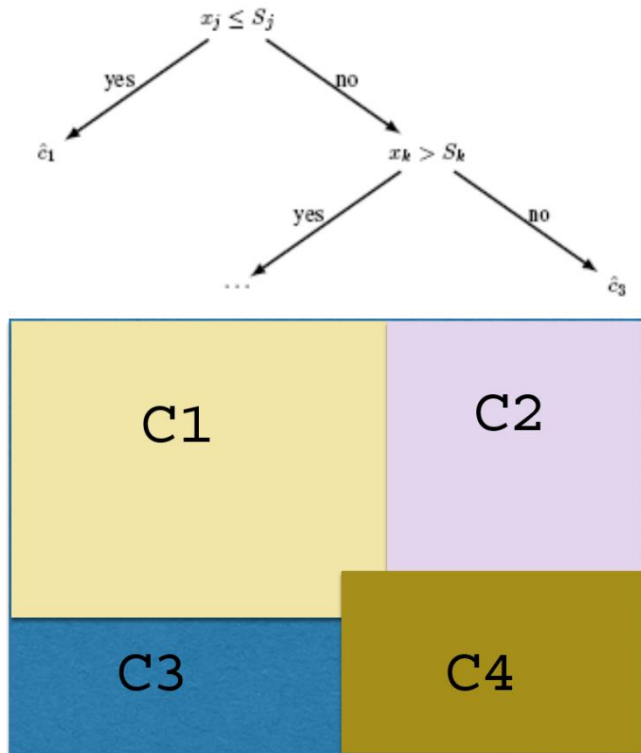
## Datenanalyse-Projekt



- Einstellung von Maschinenparametern bei der Glasbearbeitung
- Störungsfrüherkennung
- Predictive Maintenance

# Praxisbeispiel Schmelzen von Glas

## Datenanalyse-Projekt



### Entscheidungsbaum

- zur Mustererkennung (welche Einstellung der ANlagenparameter führt zu welcher Qualität)
- liefert einfach verständliche und nachvollziehbare Handlungsempfehlungen
- einfache Implementierung

IF  $x_j > S_j$  THEN output =  $\hat{c}_1$  .

IF  $x_j \leq S_j$  AND  $x_k > S_k$  THEN output =  $\hat{c}_3$  .

IF  $x_j \leq S_j$  AND  $x_k \leq S_k$  THEN output = ... .

# Praxisbeispiel Getränkeabsatz-Prognosen/I

## Datenanalyse-Projekt



### Absatzprognosen für Filialen von Getränkehändlern



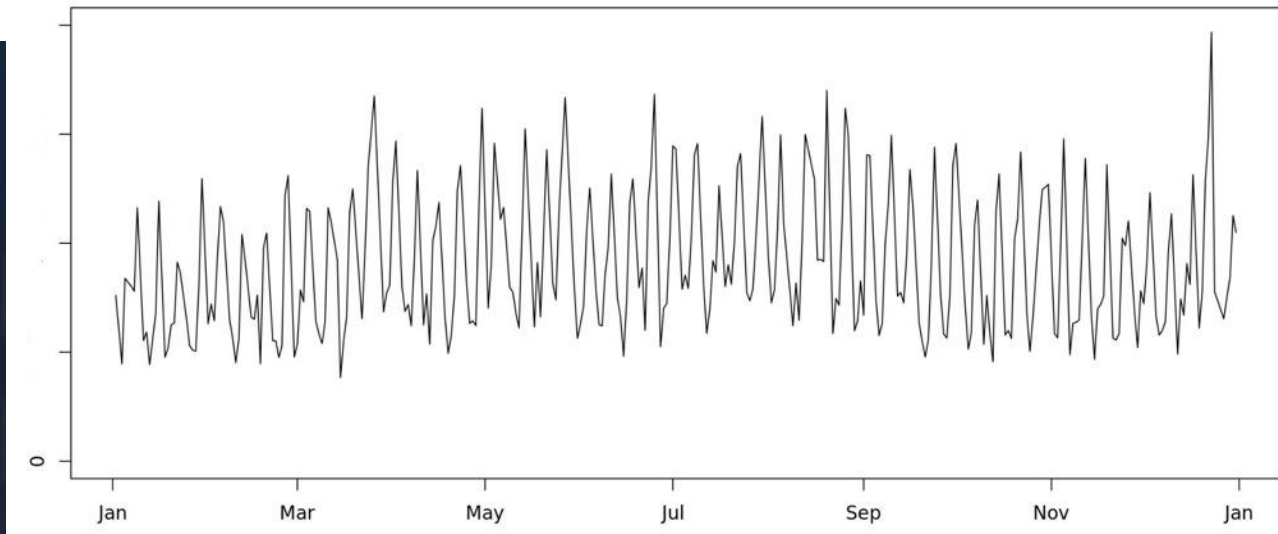
- Vermeidung von Stock-Outs
- Verringerung von Kosten für Lagerhaltung





# Praxisbeispiel Getränkeabsatz-Prognosen/II

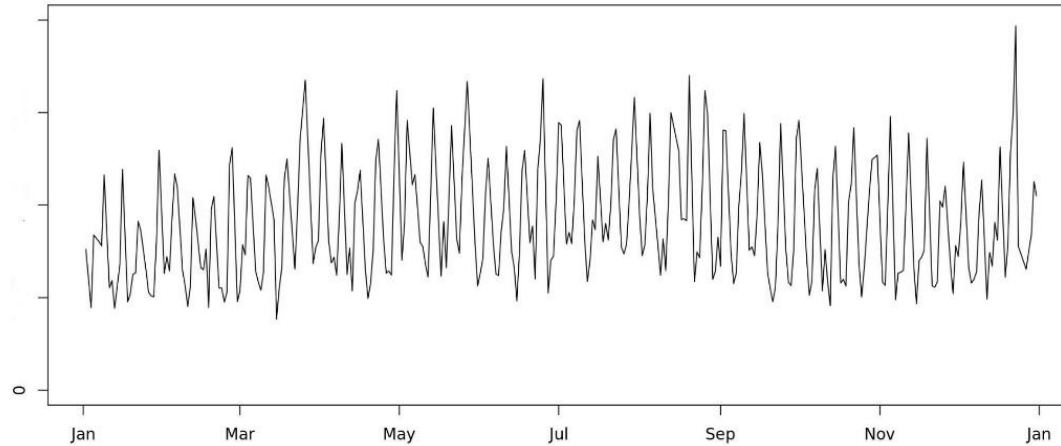
## Datenanalyse-Projekt



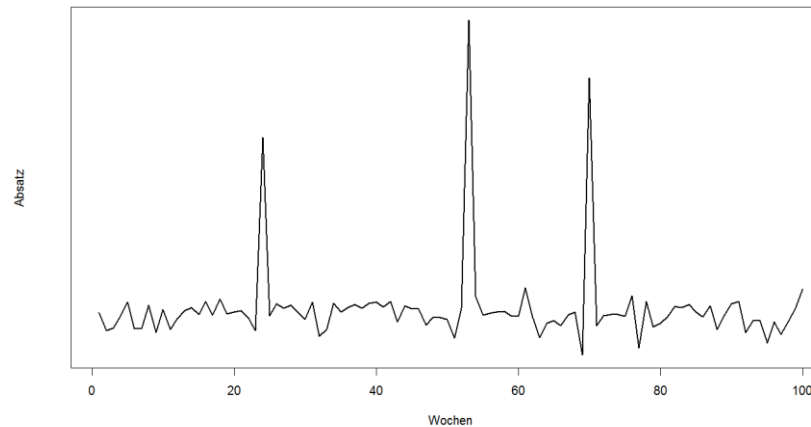
➤ **Täglicher Umsatz einer Filiale innerhalb eines Jahres**

# Absatzprognosen

Regelmäßige Absätze lassen sich gut prognostizieren:



Unregelmäßige Absätze lassen sich kaum prognostizieren:



# Handlungsempfehlungen

Erste Schritte wagen

---



große Investitionen in Datensammeln, Hard- und Software meiden

Auswahl eines geeigneten Use-Case ist der wichtigste erste Schritt



- Definition von konkreten Zielen/Herausforderungen
- Sichtung vorhandener Daten
- Agile Pilotprojekte
- Priorisierung nach Kosten-Nutzen bzw. Erfolgswahrscheinlichkeit




Wahl der Technologie, Software, Verfahren ist zweitrangig





# Trend

Vom  
Hersteller/Händler  
zum Dienstleister



**“People don’t want to buy a quarter-inch drill. They want a quarter-inch hole!”  
[Theodore Levitt]**

# Licht als Dienstleistung

Philips Pay-per-Lux

---

## Licht als Dienstleistung anstelle des Verkaufs von Beleuchtungssystemen

- Metro Washington
- Flughafen Amsterdam Schiphol



Der Kunde zahlt für das Licht. Philips bleibt der Eigentümer der Lichtanlage und ist zuständig für die Installation, Wartung und die Bezahlung der Stromrechnung.



**Damit wird Phillips vom Hersteller zum Dienstleister!**





# Vom Autohersteller zum Mobilitätsdienstleister/I

Daimler

---

Vom Autohersteller zum Anbieter von Mobilität durch Übernahme und Anteile an Mobilitäts-Service-Unternehmen:

- Carsharing-Dienst Car2go
- App MyTaxi
- Mitfahrzentrale Carpooling.com
- Busunternehmen Flixbus
- Chauffeurdienst Blacklane
- Mobilität-Service-Plattform Moovel



# Vom Autohersteller zum Mobilitätsdienstleister/II

BMW

---

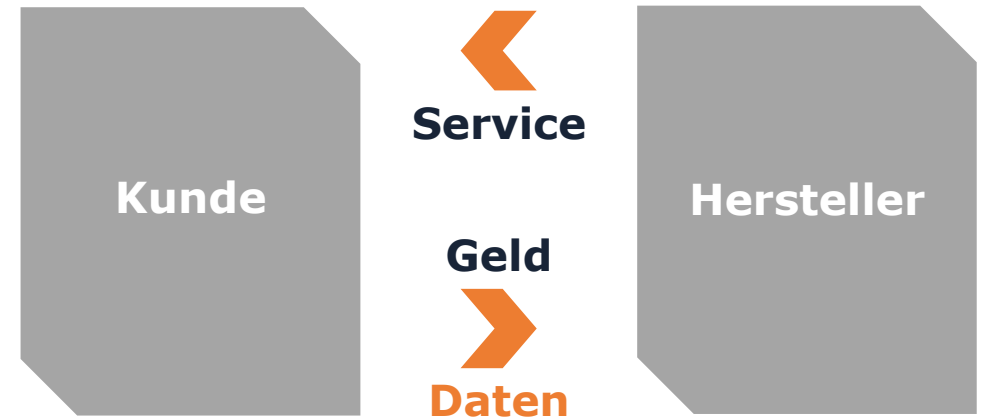
- BMW DriveNow
- Corporate CarSharing
- AlphaCity
- BMW On Demand



# Vom Maschinenhersteller zum Dienstleister

## Service für den Kunden

- Überwachung der Prozessdaten in Echtzeit
- Vorausschauende Instandhaltung
- Automatische Einstellung der Prozessparameter
- Datenprotokollierung zu Prozessverläufen





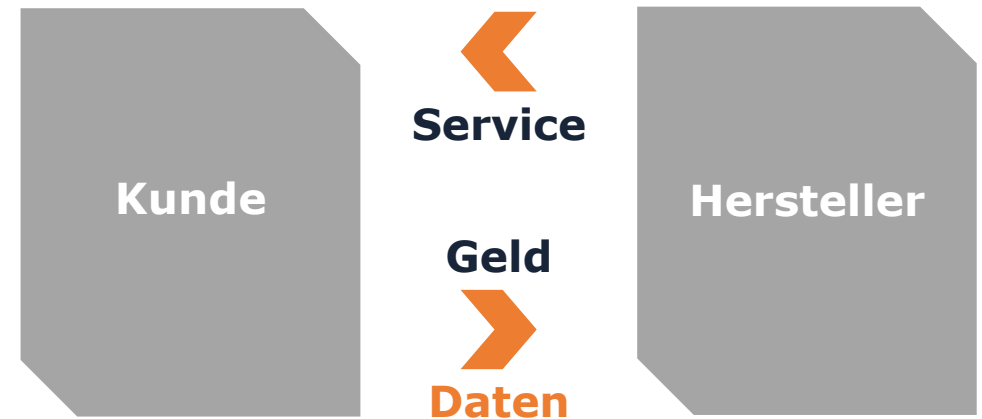
# Vom Maschinenhersteller zum Dienstleister

## Service für den Kunden

- Überwachung der Prozessdaten in Echtzeit
- Vorausschauende Instandhaltung
- Automatische Einstellung der Prozessparameter
- Datenprotokollierung zu Prozessverläufen

## Nutzen für den Hersteller

- Erschließung neuer Geschäftsfelder (After-Sales)
- Informationen über die Kunden
- Informationen über das eigene Produkt
- Informationen, wie Kunden das Produkt nutzen (jobs to be done)



# Digitale

Geschäftsmodelle  
Industrie 4.0

---

# Digitale Geschäftsmodelle Industrie 4.0

## Kernfragen

---

- Welche Kundenbedürfnisse kann ich bedienen?
- Welche Dienste schaffen einen Mehrwert für die Kunden?
- Wie erhalte ich Zugang zu Kunden-Nutzungsdaten?
- Wie kann ich einen Mehrwert aus den Daten ziehen?

# Handlungsempfehlungen

Erste Schritte wagen

---



große Investitionen in Datensammeln, Hard- und Software meiden

Auswahl eines geeigneten Use-Case ist der wichtigste erste Schritt



- Definition von konkreten Zielen/Herausforderungen
- Sichtung vorhandener Daten
- Agile Pilotprojekte
- Priorisierung nach Kosten-Nutzen bzw. Erfolgswahrscheinlichkeit



Wahl der Technologie, Software, Verfahren ist zweitrangig