

Datenanalyse und Modellierung als Instrumente der Wirtschaft



250125-1 Konversatorium
„Berufsbild Mathematik
(Finanzmathematik in der Praxis)“

27. Juni 2018

Mag. Christoph Krischanitz

Erfahrungen in der Datenanalyse

Seit 1982

- Personal(kosten)daten
- Trendanalysen, Mustererkennung, Forecasting, Simulation, Hochrechnung
- Sterblichkeiten, Invaliditäten, Krankheitskosten, Fluktuationen, ...

Seit 2002

- Versicherungsdaten
- Nichtlineare Regressionen, Clustering, Häufigkeitsanalysen, Principal Component Analysis, Distribution-Fitting, Simulation
- Schadendaten, Naturkatastrophen, Kundendaten

Seit 2002

- Finanzdaten
- Zeitreihen, Korrelationsanalysen, Optimierung, Optionspreise, Zinsmodelle
- Asset-Liability-Management, Portfoliomanagement, Pricing

Seit 2007

- Allg. Datenanalysen
- Ausfallrisiken, Kundenrating, Segmentierung, Mustererkennung, Abweichungsanalysen, Regressionen
- Controlling, Risikomanagement, HR

Trendlinie

Warum Datenanalyse?

- Motivation
- Kompetenz
- Strategie und Entscheidung

Branchen/Sektoren

- Regulierte Märkte
- Unregulierte Märkte

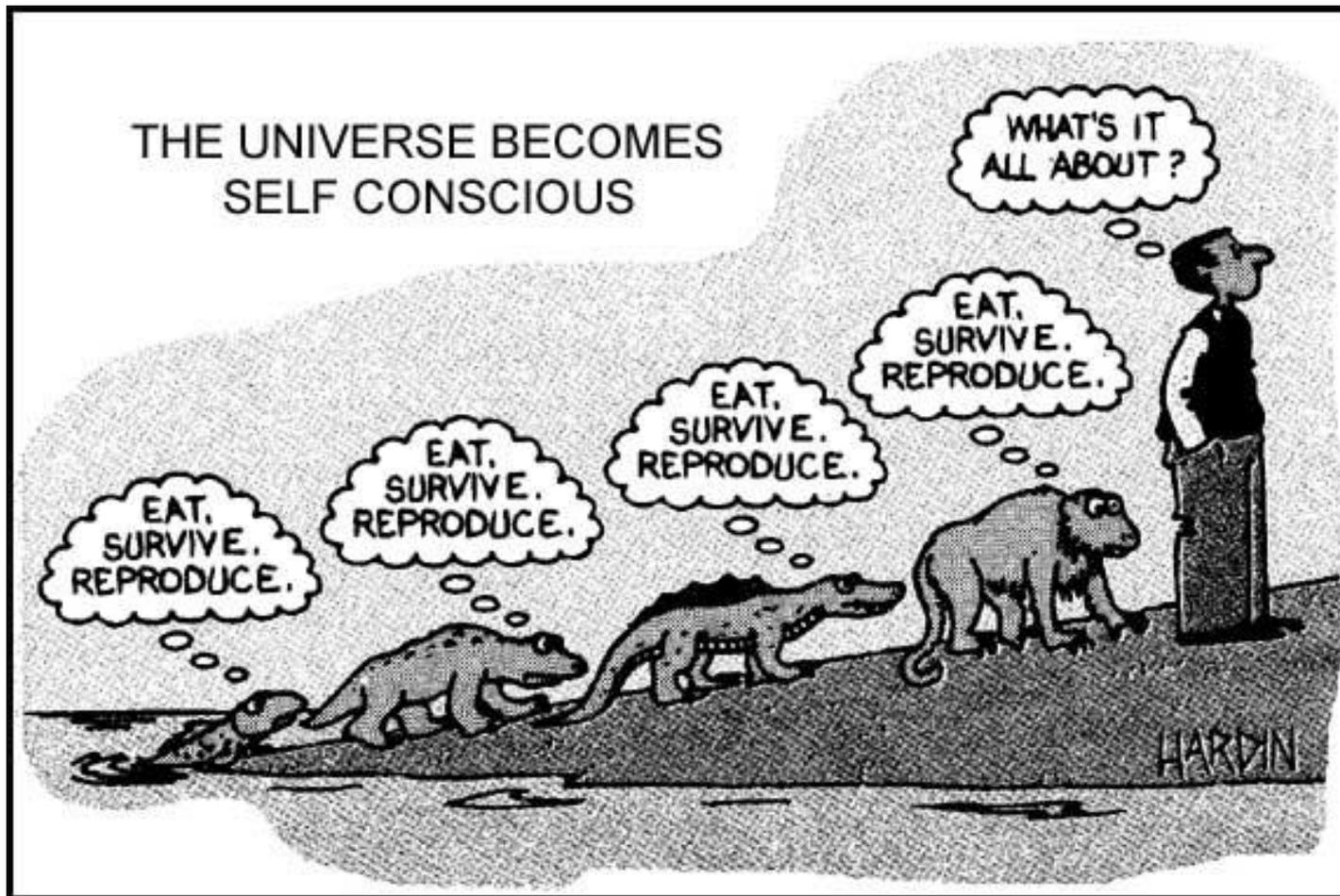
Datenanalyse im Unternehmen

- Unternehmensbereiche
- Prozesse
- Daten
- Governance
- Ziele
- Methoden
- Tools
- Case Studies

Von der Analyse zur Simulation



Warum Datenanalyse?



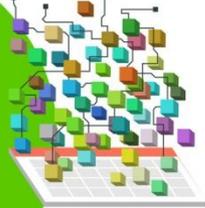
40 ZETTABYTES
[43 TRILLION GIGABYTES]
of data will be created by 2020, an increase of 300 times from 2005



Volume SCALE OF DATA



It's estimated that **2.5 QUINTILLION BYTES** [2.3 TRILLION GIGABYTES] of data are created each day



Most companies in the U.S. have at least **100 TERABYTES** [100,000 GIGABYTES] of data stored



The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015 **4.4 MILLION IT JOBS** will be created globally to support big data, with 1.9 million in the United States



As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES
[161 BILLION GIGABYTES]



30 BILLION PIECES OF CONTENT are shared on Facebook every month



By 2014, it's anticipated there will be **420 MILLION WEARABLE, WIRELESS HEALTH MONITORS**

4 BILLION+ HOURS OF VIDEO are watched on YouTube each month



400 MILLION TWEETS are sent per day by about 200 million monthly active users

Variety DIFFERENT FORMS OF DATA



The New York Stock Exchange captures **1 TB OF TRADE INFORMATION** during each trading session



Modern cars have close to **100 SENSORS** that monitor items such as fuel level and tire pressure

Velocity ANALYSIS OF STREAMING DATA

By 2016, it is projected there will be **18.9 BILLION NETWORK CONNECTIONS** – almost 2.5 connections per person on earth



1 IN 3 BUSINESS LEADERS don't trust the information they use to make decisions



27% OF RESPONDENTS

in one survey were unsure of how much of their data was inaccurate

Veracity UNCERTAINTY OF DATA

Poor data quality costs the US economy around **\$3.1 TRILLION A YEAR**



Was ist Data Analytics?

Multivariate
Statistics

Advanced
Analytics

Predictive
Analytics

Cognitive
Analytics

Data
Mining

Knowledge
Discovery

Machine
Learning

Artificial
Intelligence

Data
Science

arithmetica

Beispiele für Vorteile der Datenanalyse

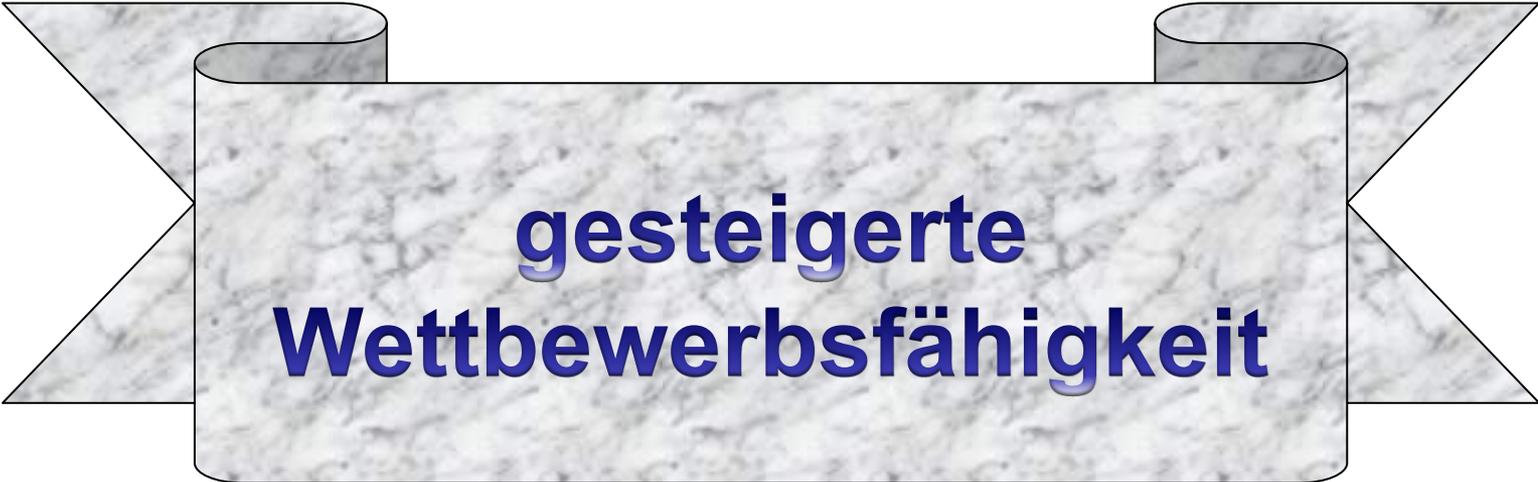
PwC

- Klare Entscheidungsgrundlagen schaffen
- Geschäftsprozesse optimieren
- Risiken kalkulieren
- Profitabilität steigern
- Preisgestaltung dynamisch anpassen
- Optimale Kundenorientierung
- Ausschöpfen unentdeckter Marktpotenziale

KPMG

- Wachstum optimieren
- Risiken managen
- Kosten sparen
- Aus Daten lernen
- Vertrauen schaffen

Letztendlich geht es um ...



**gesteigerte
Wettbewerbsfähigkeit**

Ineffizienzen
identifizieren

Potenziale
erkennen

Bessere
Differenzierung

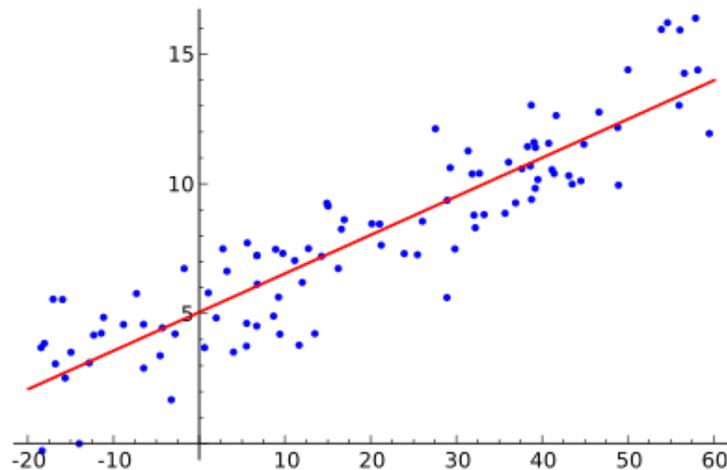
Genauere
Rechnung

Schnellere
Reaktion

arithmetica

Ziel der Datenanalyse

- Zufällige Effekte von systemischen Effekten zu trennen
- Typische Fragestellungen:
 - > Welche systemischen Zusammenhänge („Muster“, „Trends“, ...) sind in den Daten bemerkbar?
 - > Wie stark beeinflusst der Zufall die Ergebnisse?



Zufälligkeit

- Welche Reihe ist zufällig?

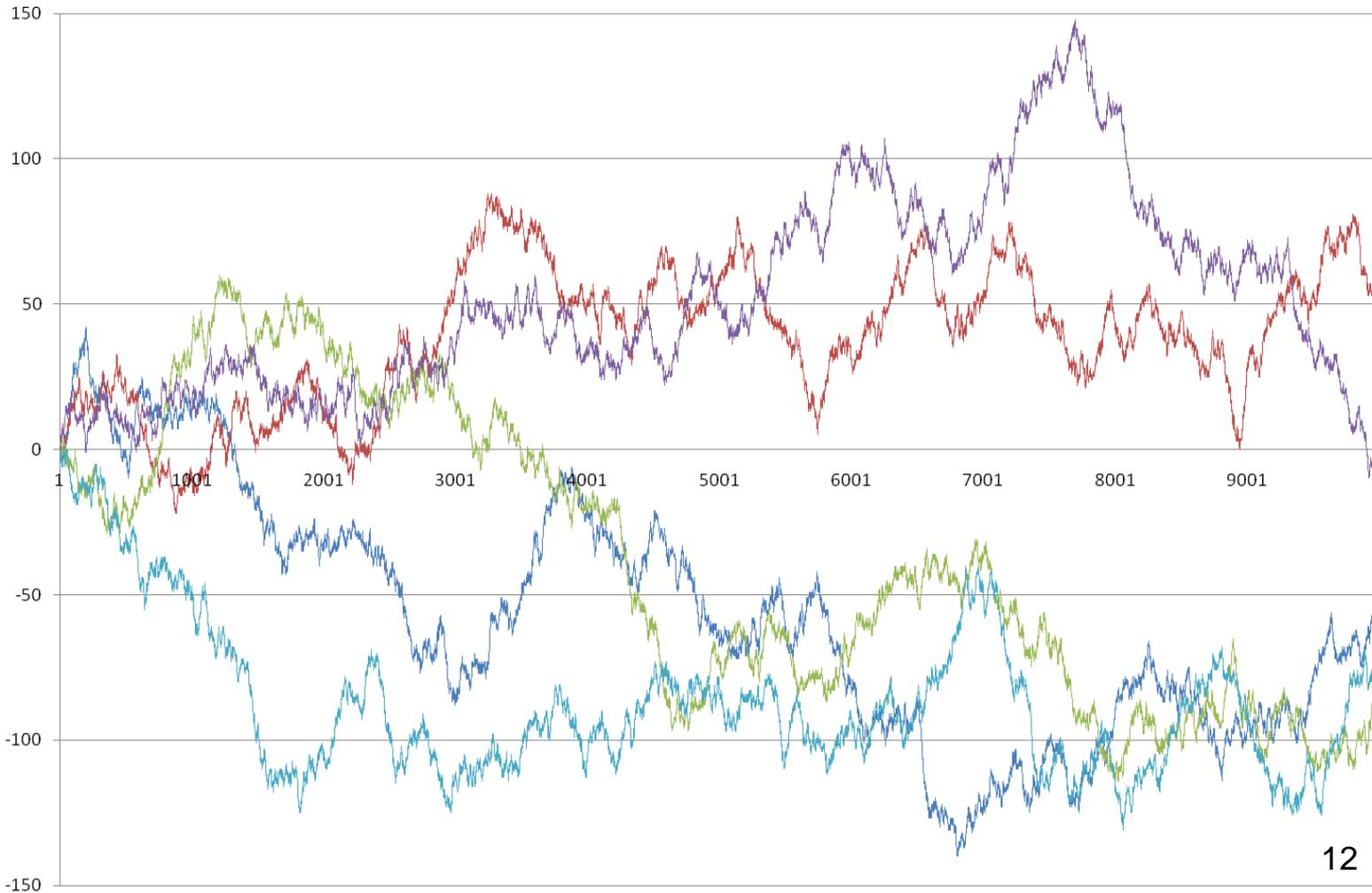
A: 10

B: 11001100110011001100110011001100110011001100

C: 1011001110001100101100111000110010110011

D: 0010000101011101110010000100100000111110

Random Walk



arithmetica

Der Zufall ist nicht immer als solcher erkennbar

Gefahr der Überinterpretation von gewissen Datenentwicklungen

Daraus entstehen falsche Modellansätze bzw. falsche Annahmen

Die Analyseergebnisse sind irreführend

Falsche Entscheidungen werden getroffen

Daher:

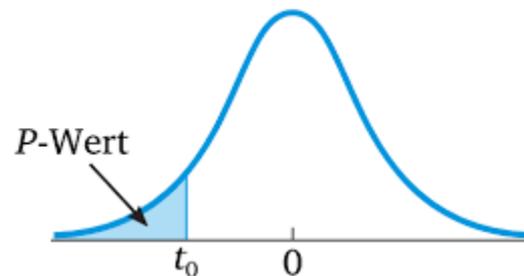
- 1. Ausreichende Ressourcen für Analyse zur Verfügung stellen
- 2. Regelmäßige Validierung der Ansätze, Annahmen und Ergebnisse



arithmetica

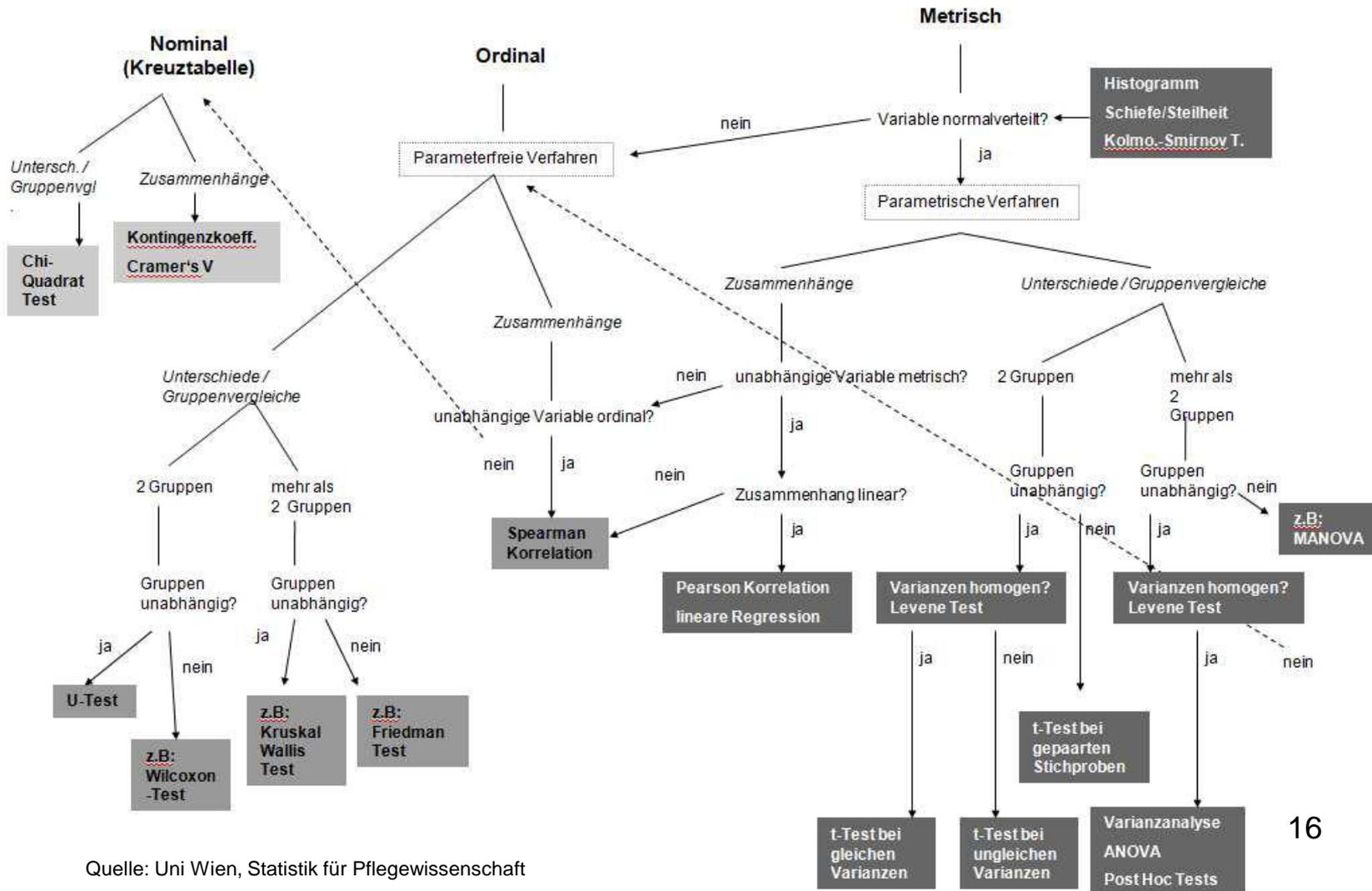
Der p-Wert

- Wahrscheinlichkeit, dass ein beobachteter Effekt rein zufällig auftritt („Nullhypothese“)
- Je kleiner, desto unwahrscheinlicher, dass reiner Zufall im Spiel war
- „Signifikanzniveau“: Schwelle für den p-Wert
 - > häufig: p-Wert < 5% gilt als signifikant
 - > keine universelle Größe, p-Wert individuell festzulegen



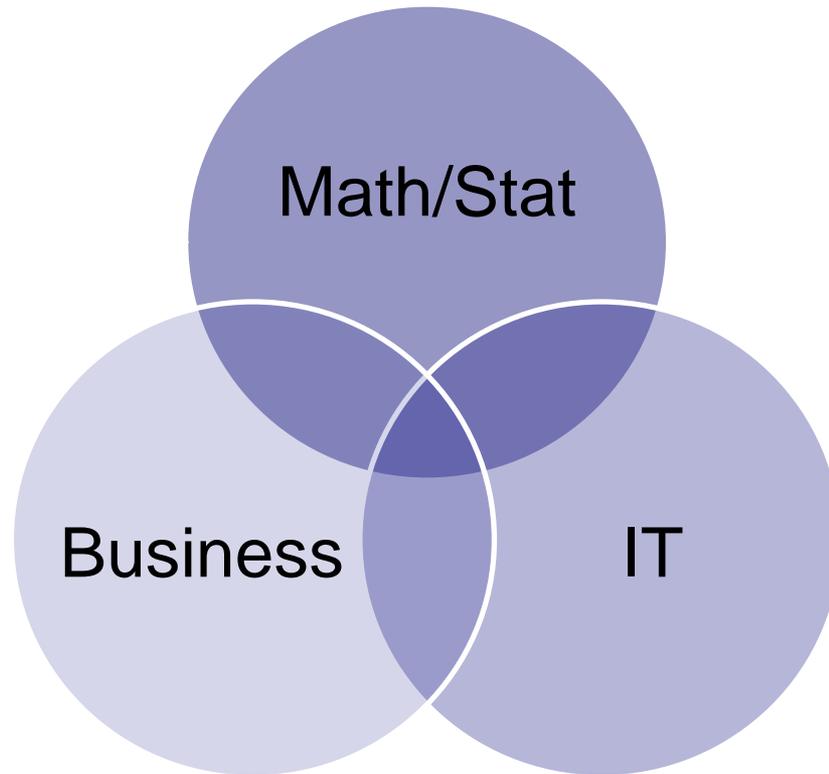
Signifikanztests

Datenniveau der abhängigen Variable



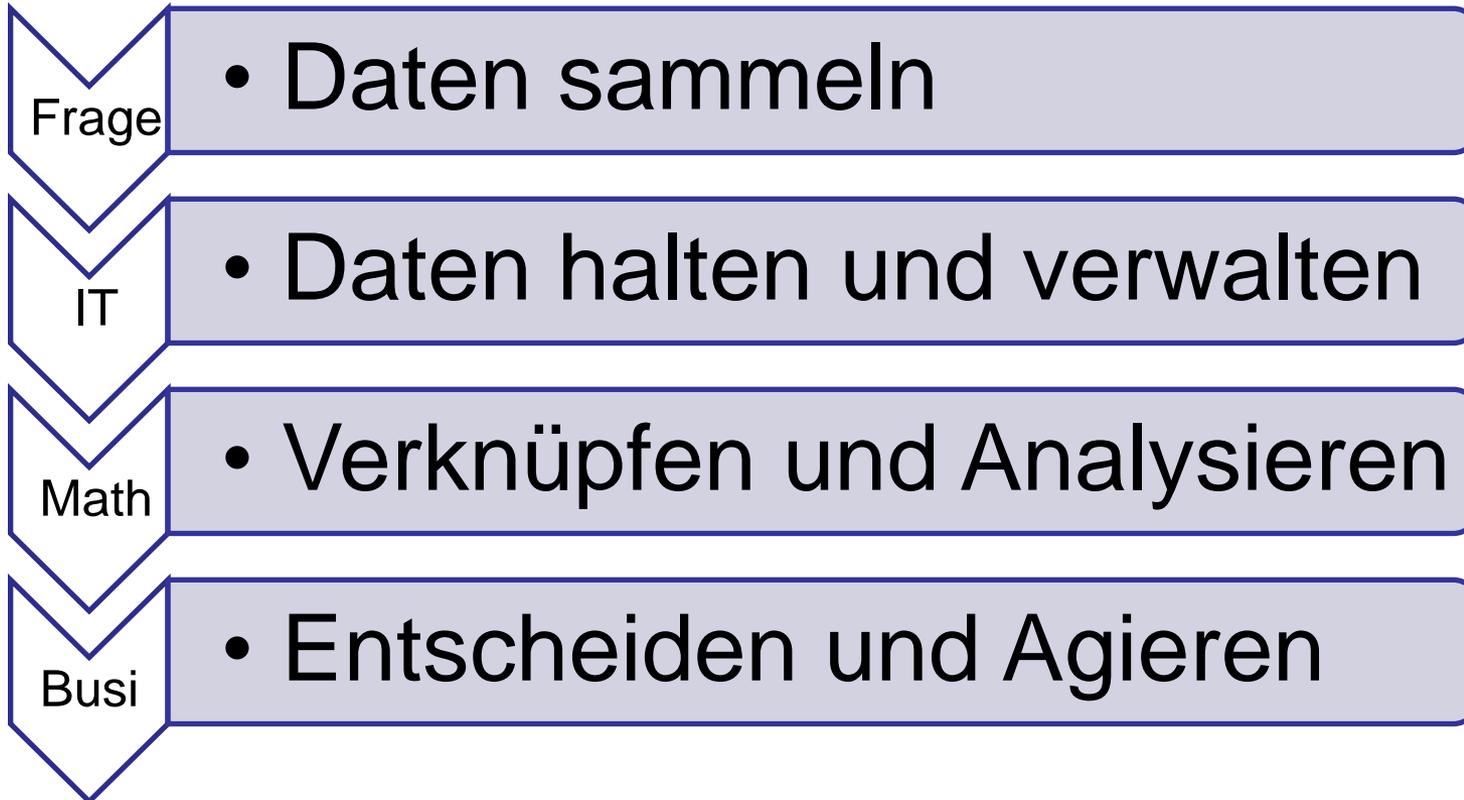
arithmetica

Data Science = Kombination von Kompetenzen



arithmetica

Verkettung der Kompetenzen

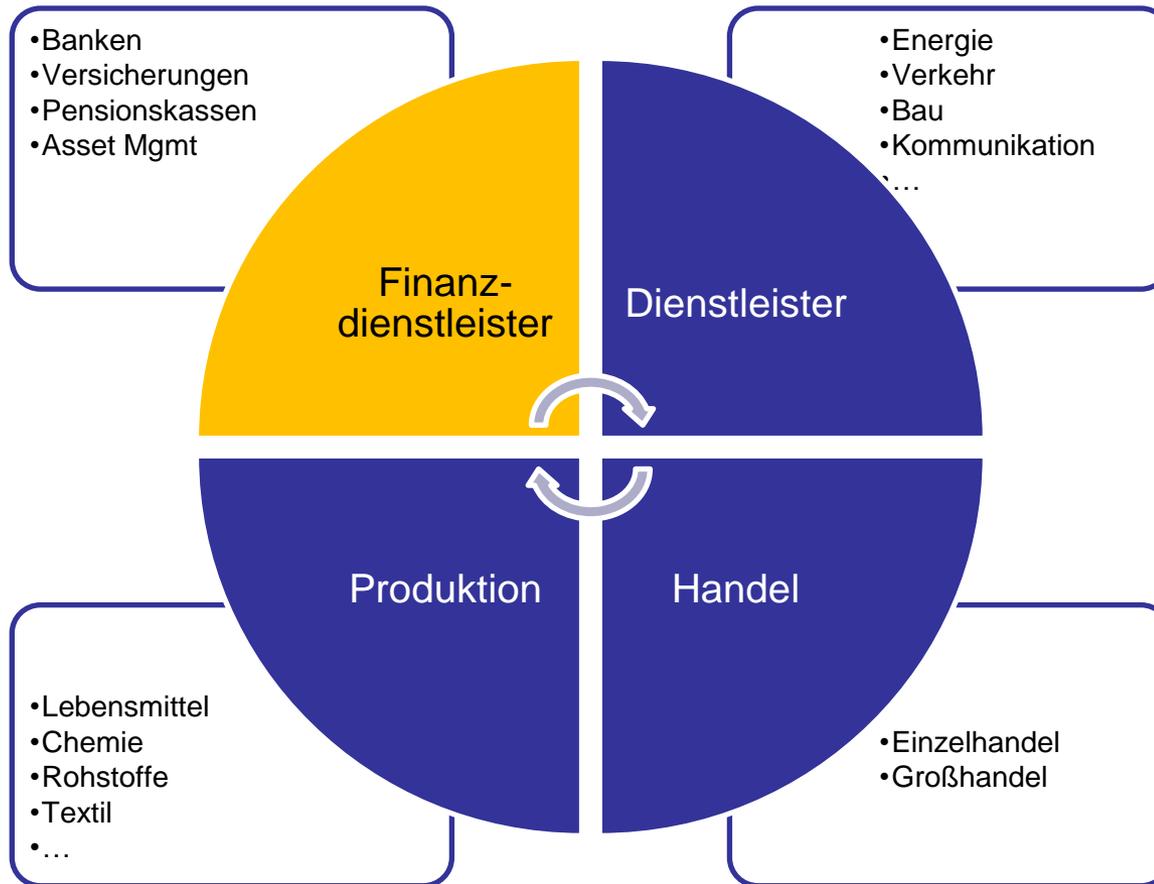


Branchen/Sektoren

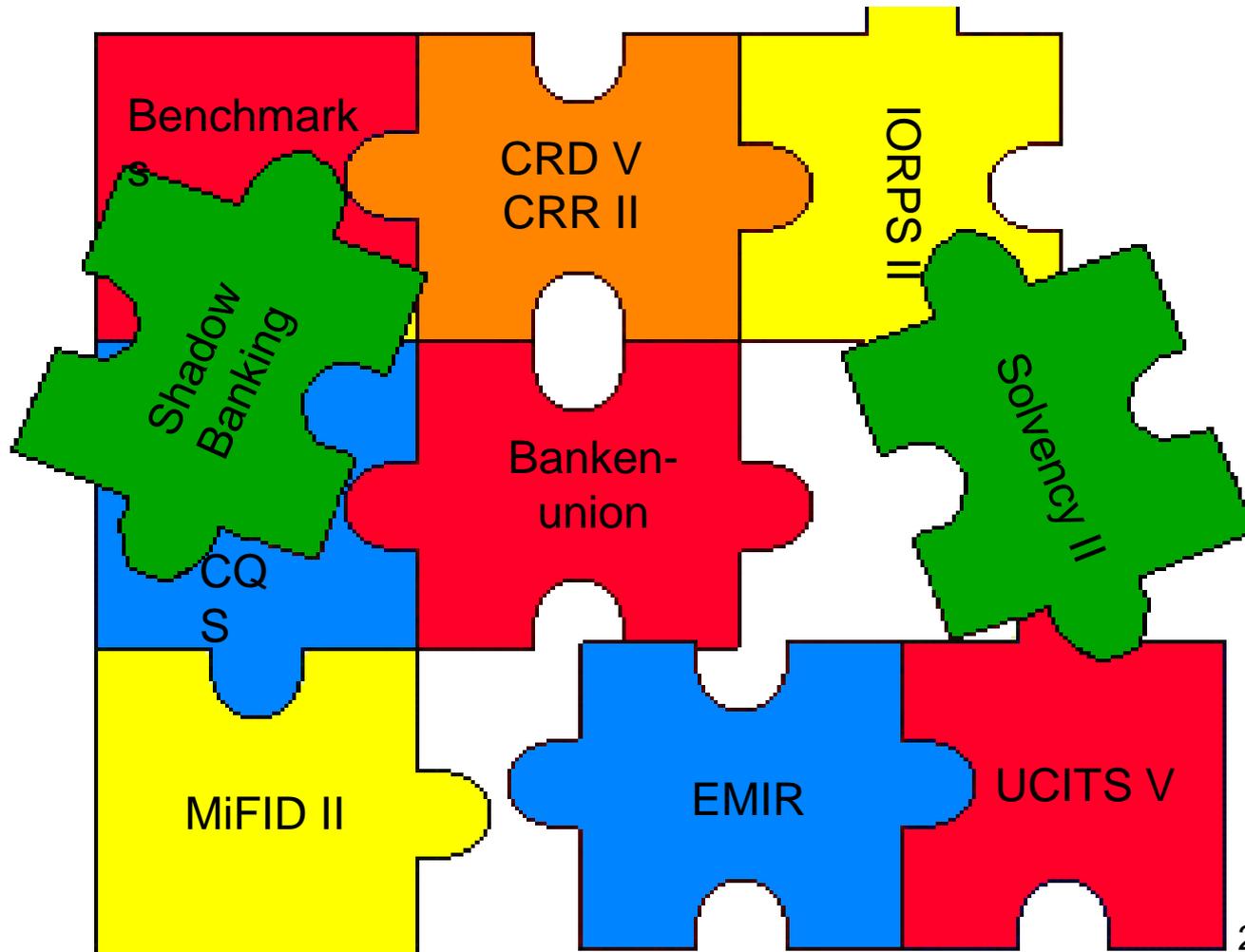


arithmetica

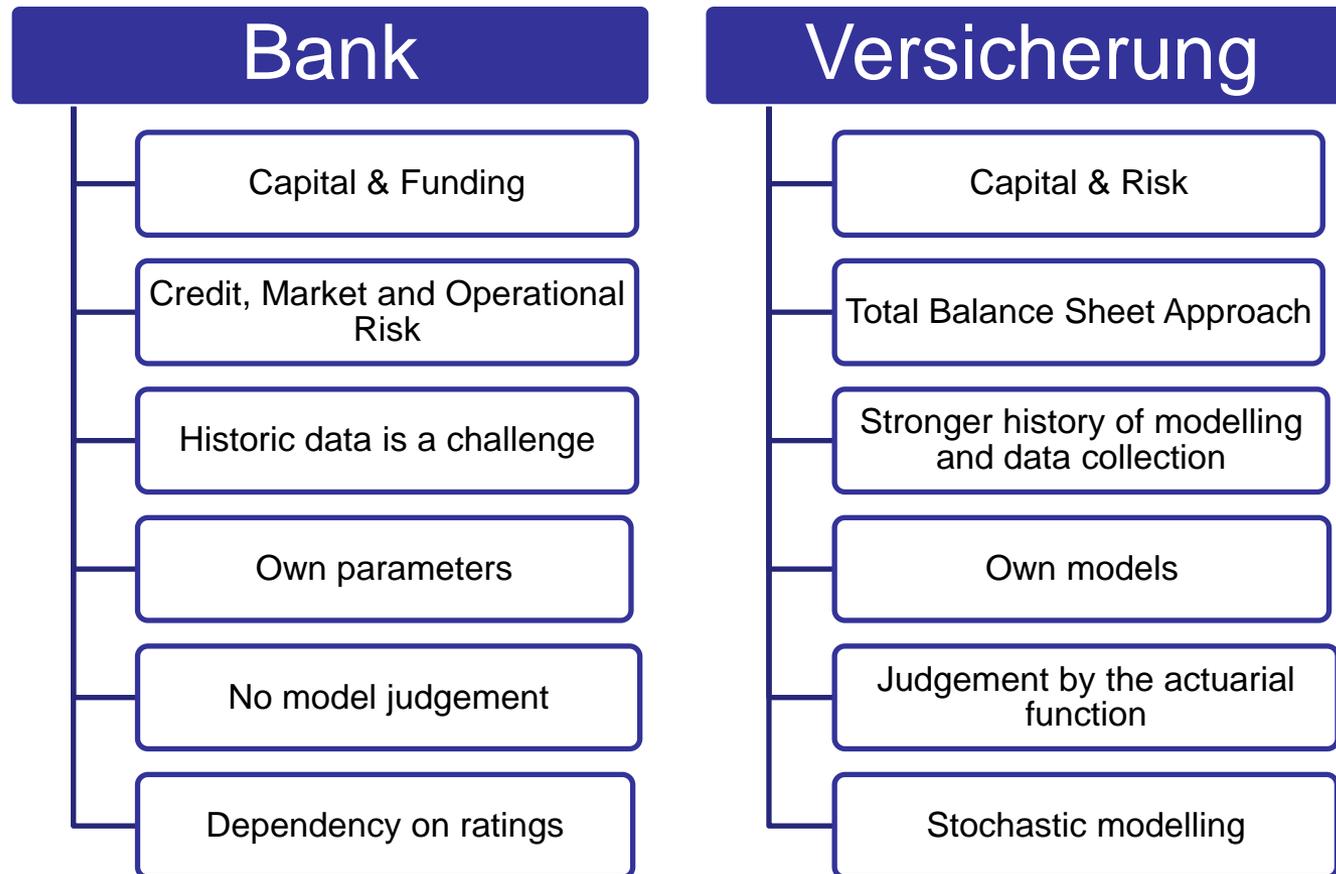
Wichtige Branchen



Das europäische Aufsichtspuzzle



Unterschiede Banken/Versicherungen



How Data Makes Insurance Work Better for You

- 1. THE DIGITAL REVOLUTION AND BIG DATA** pg6
The world is changing at an unprecedented pace – so just what is so different about data these days, and why does that affect the insurance industry?
- 2. UNDERSTANDING YOU** pg10
Insurers need to understand you in order to provide you with products and services that really work for you.
- 3. GETTING THE RIGHT PRICE** pg14
Getting the right insurance isn't simply about getting the right product, it's also about getting it at the right price, which fairly reflects your circumstances.
- 4. ENABLING YOU TO MAKE THE RIGHT DECISIONS TO MINIMISE YOUR RISK** pg18
Insurers take on risk on behalf of their customers, and are keen to use the changing digital world to help customers understand and control the risks they face.
- 5. IMPROVING THE CLAIMS PROCESS** pg22
Insurers want to use data and technology to make the claims process as easy and flexible as possible, and to combat fraud more effectively.
- 6. HOW WE TREAT YOUR DATA** pg26
Using data innovatively means that insurers have to continue to build a trusting relationship with their customers, which involves being transparent about how we use your data.

Anwendungen von Advanced Analytics bei Versicherungen - Beispiele

Finanzdaten

- Schadenreservierung
- Schadenanalyse
- Asset-Management
- Forderungsausfälle
- Kapitalanforderung

Scoring

- Risikoprüfung
- Tarifierung
- Rabattierung

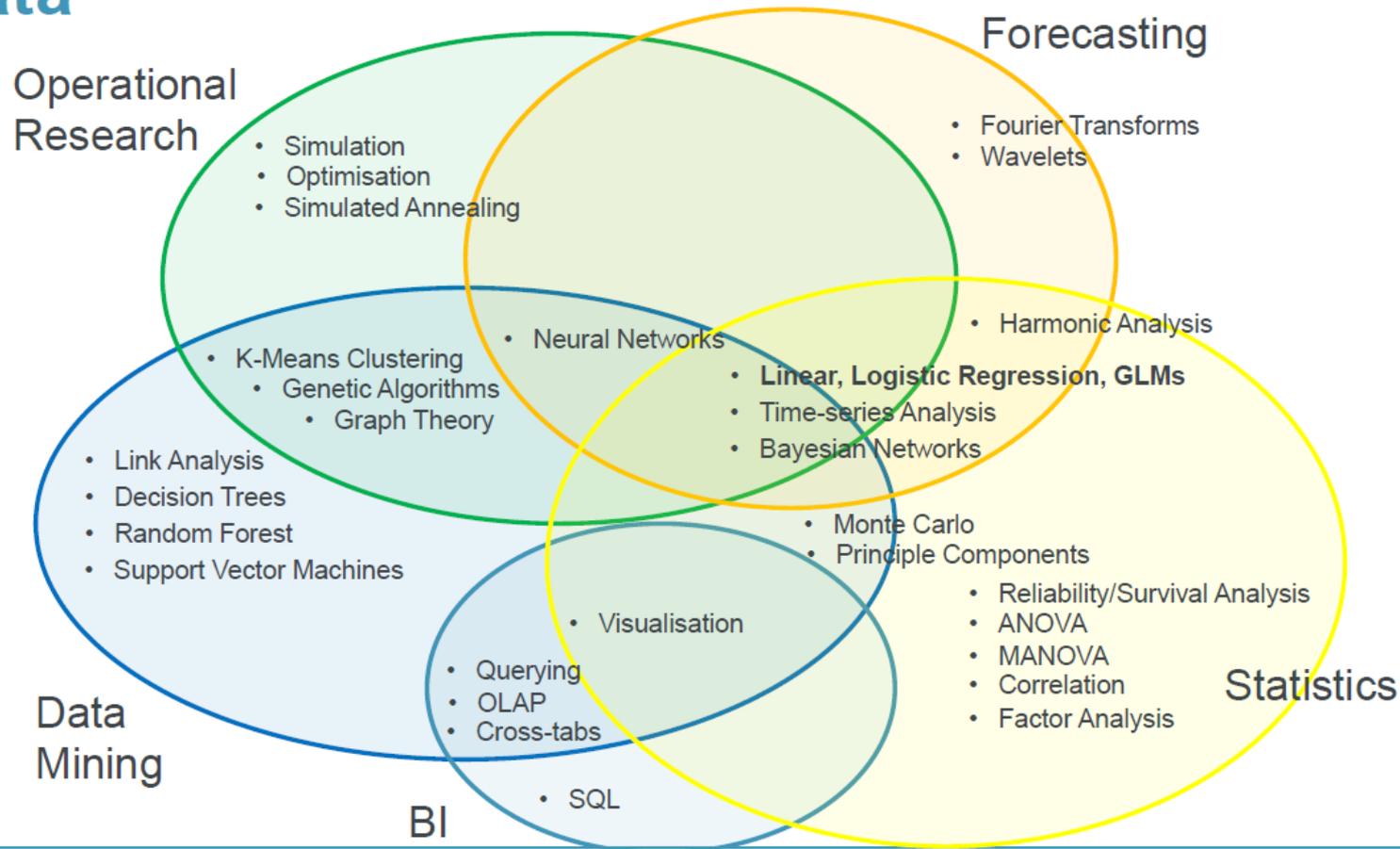
Kundenverhalten

- Abschlussaffinität
- Cross- und Up-Selling-Potenzial
- Fraud
- Fahrverhalten / Telematik (PAYD -> PHYD)
- IoT (zB wearables)

Stimmungen

- Fotos
- Social Media
- Kommunikation

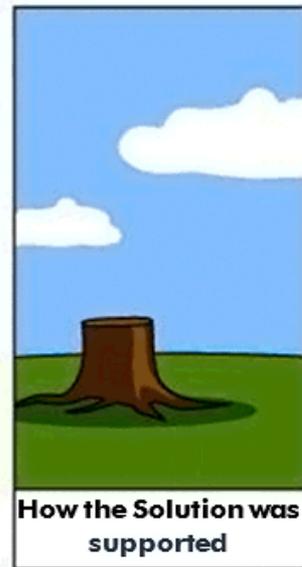
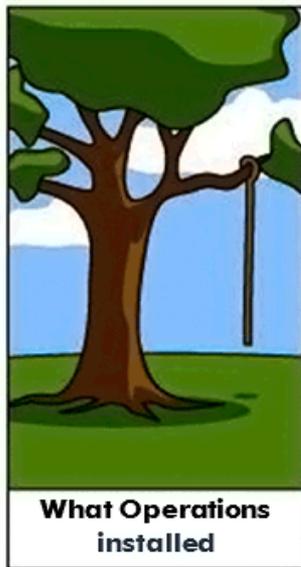
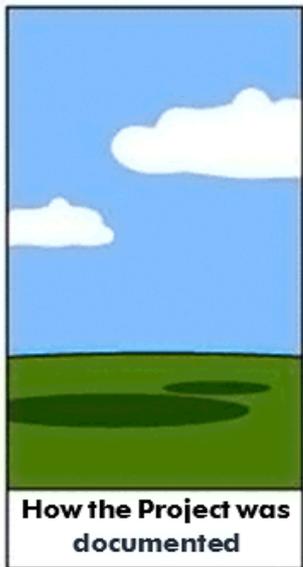
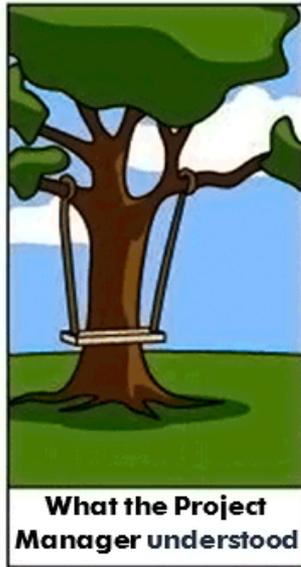
Analytics are needed to realise the value of data



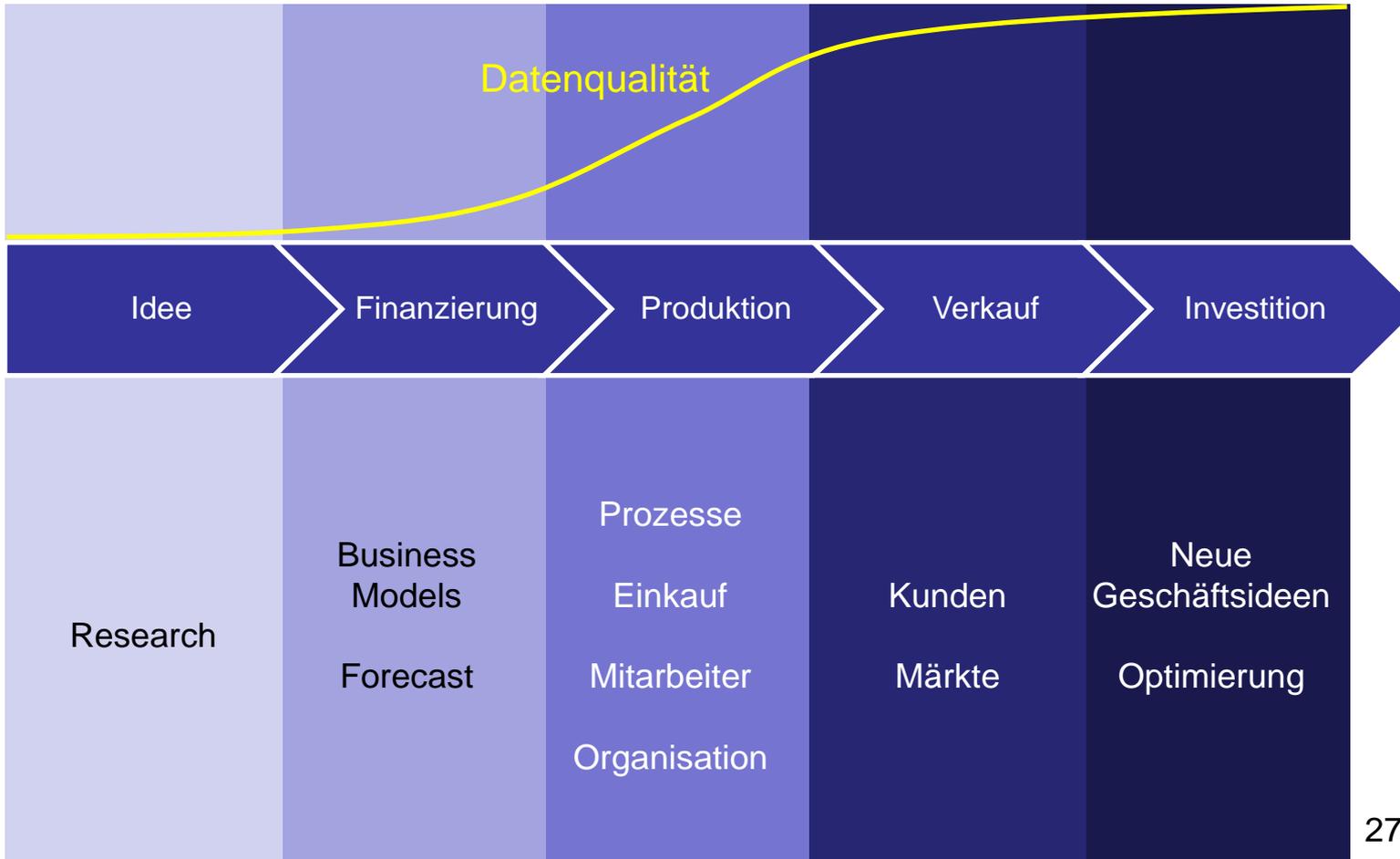
arithmetica

Modified from a version presented by John Elder, www.datamininglab.com, 2012.

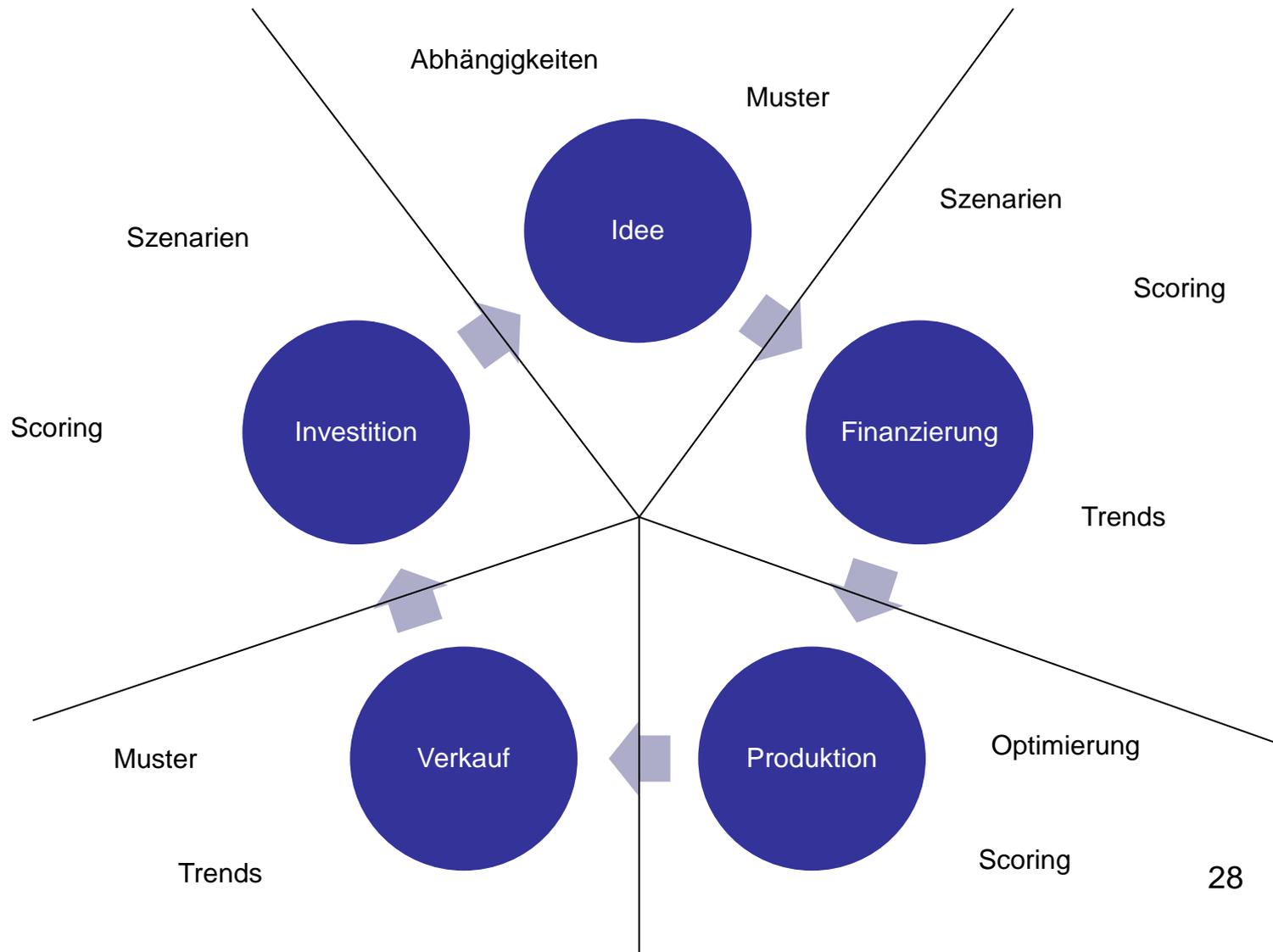
Datenanalyse im Unternehmen



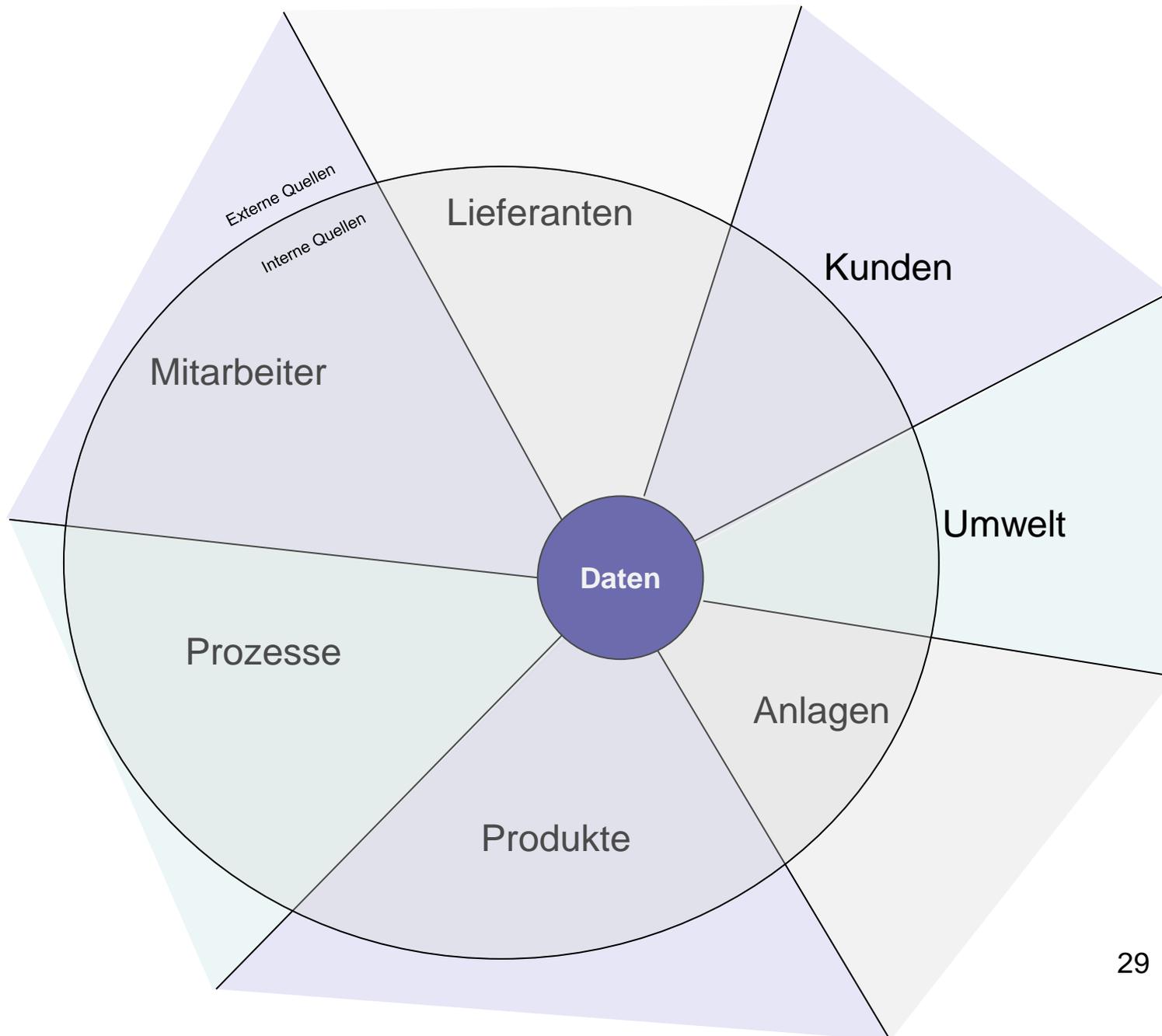
Typische Unternehmensphasen



Analysetechniken pro Phasen (Versuch einer Zuordnung)



Typische Datenquellen



Analysethemen nach strategischer Ausrichtung

Kostenführerschaft

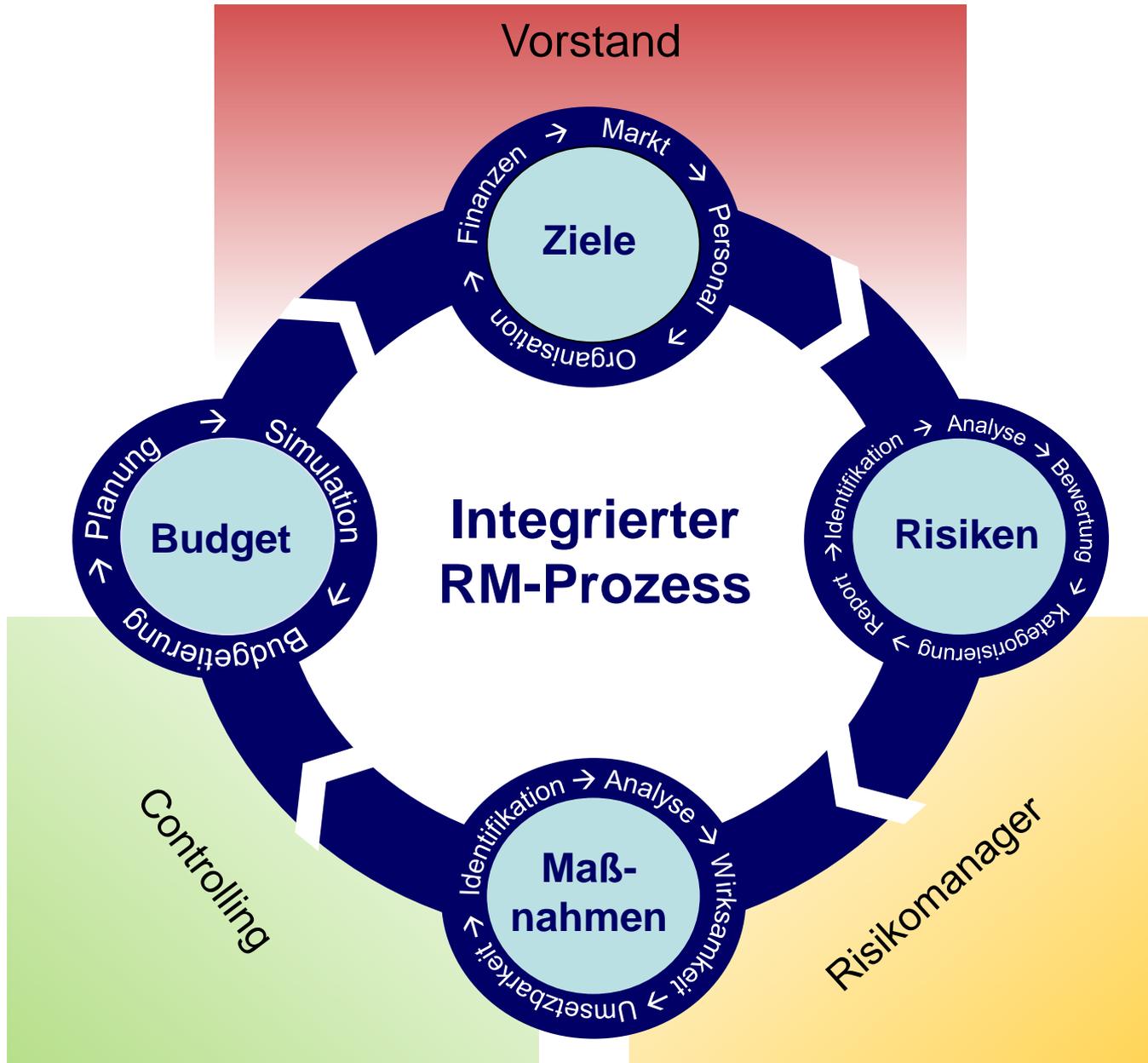
- Lieferanten
- Prozesse
- Auslastung
- Fehlzeiten
- Mengen

Produktdifferenzierung

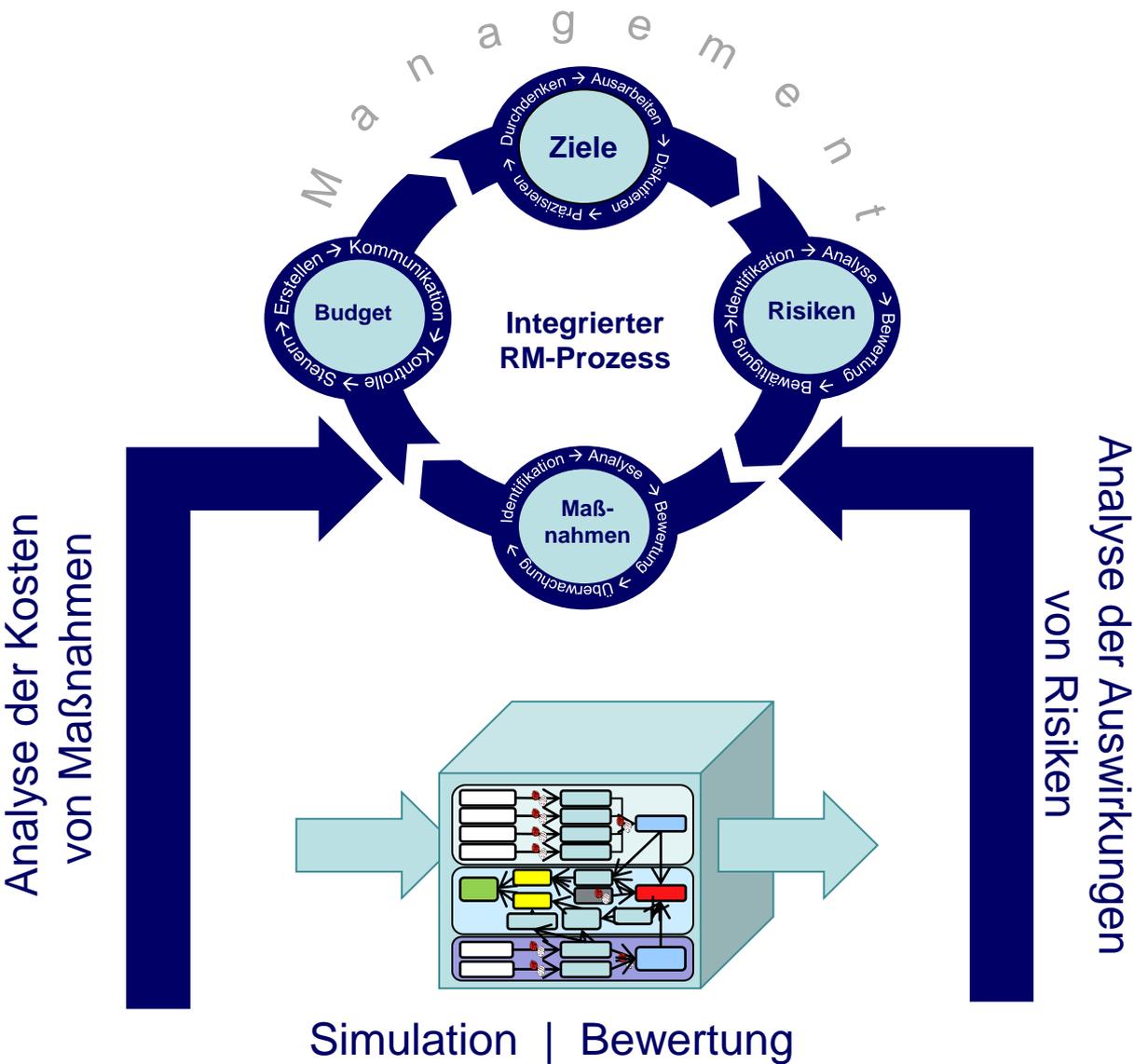
- Kunden
- Märkte
- Talente
- Qualitätsmanagement
- Innovation

Typische Unternehmensziele

- Finanzielle Stabilität: $\text{NetDebt/EBITDA} < x$
- Kapitalrentabilität: $\text{ROCE} > y\%$
- Mitarbeiterfluktuationsrate $< z\%$
- Positionierung als ...
- Umsetzung des ...-Programms
- Ausbau des Geschäftsbereichs XY



Einsatz von Analyse- und Simulationsmodellen



Kundenbedürfnisse (aus persönlichen Gesprächen)

Controlling/Finanz

- Predictive Maintenance
- Zahlungsausfälle
- Selbstversicherung
- Bandbreitenplanung
- Projektkosten
- Immobilienbewertung

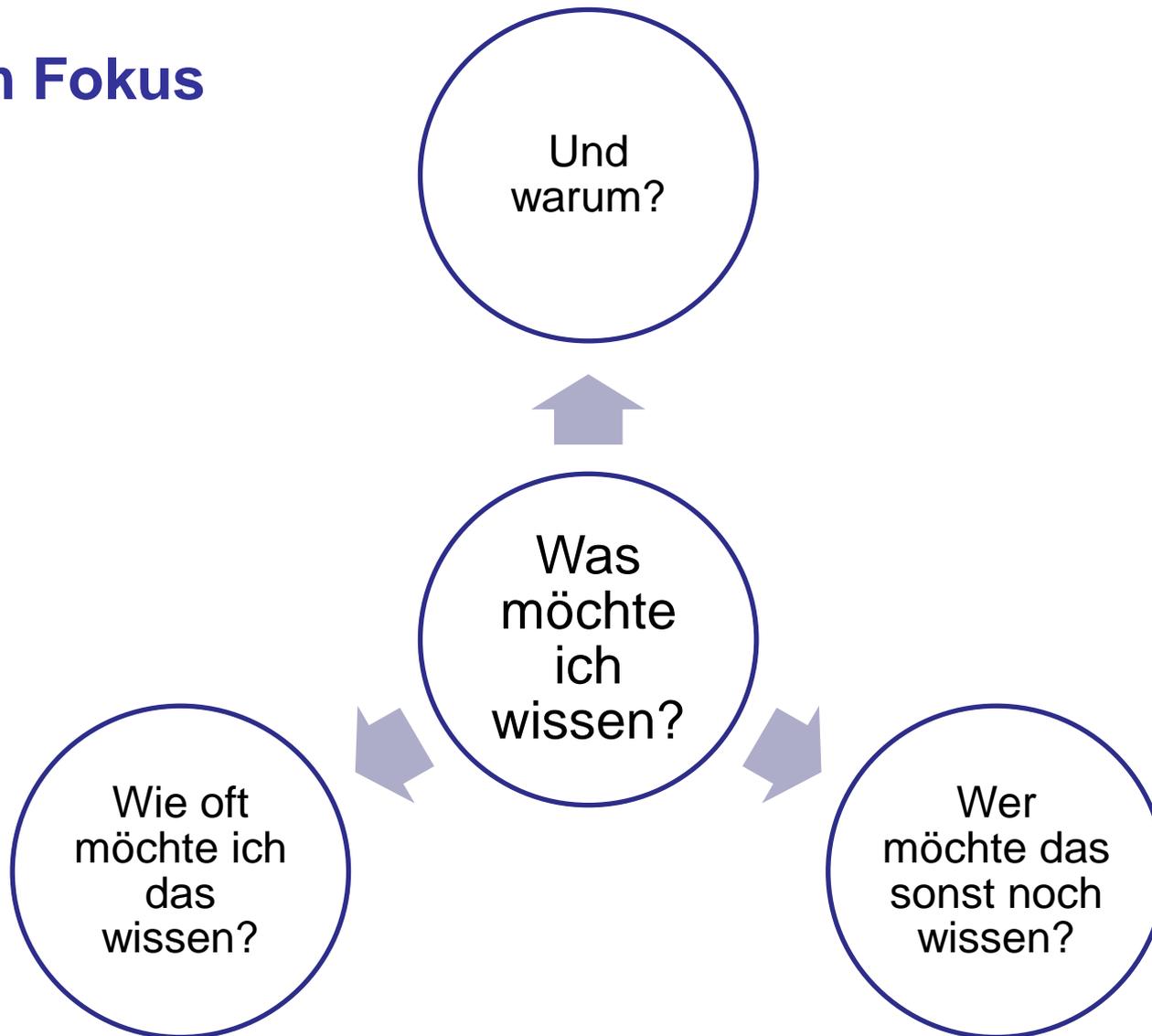
Personal

- Personalkostenanalysen
- Personalplanung

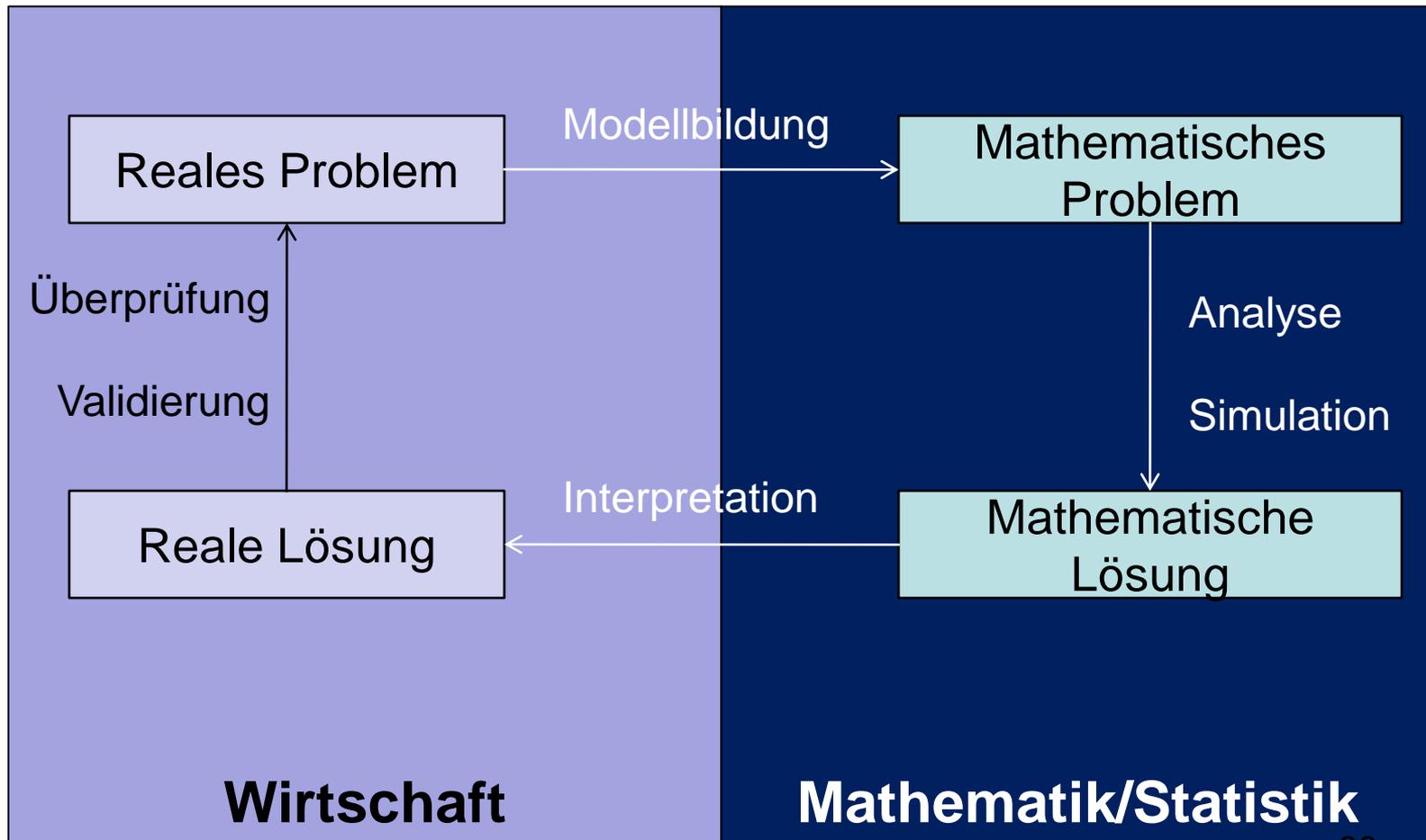
Marketing/Vertrieb

- Kundenprofile
- Abschlussaffinität
- Kundenausfall

Im Fokus

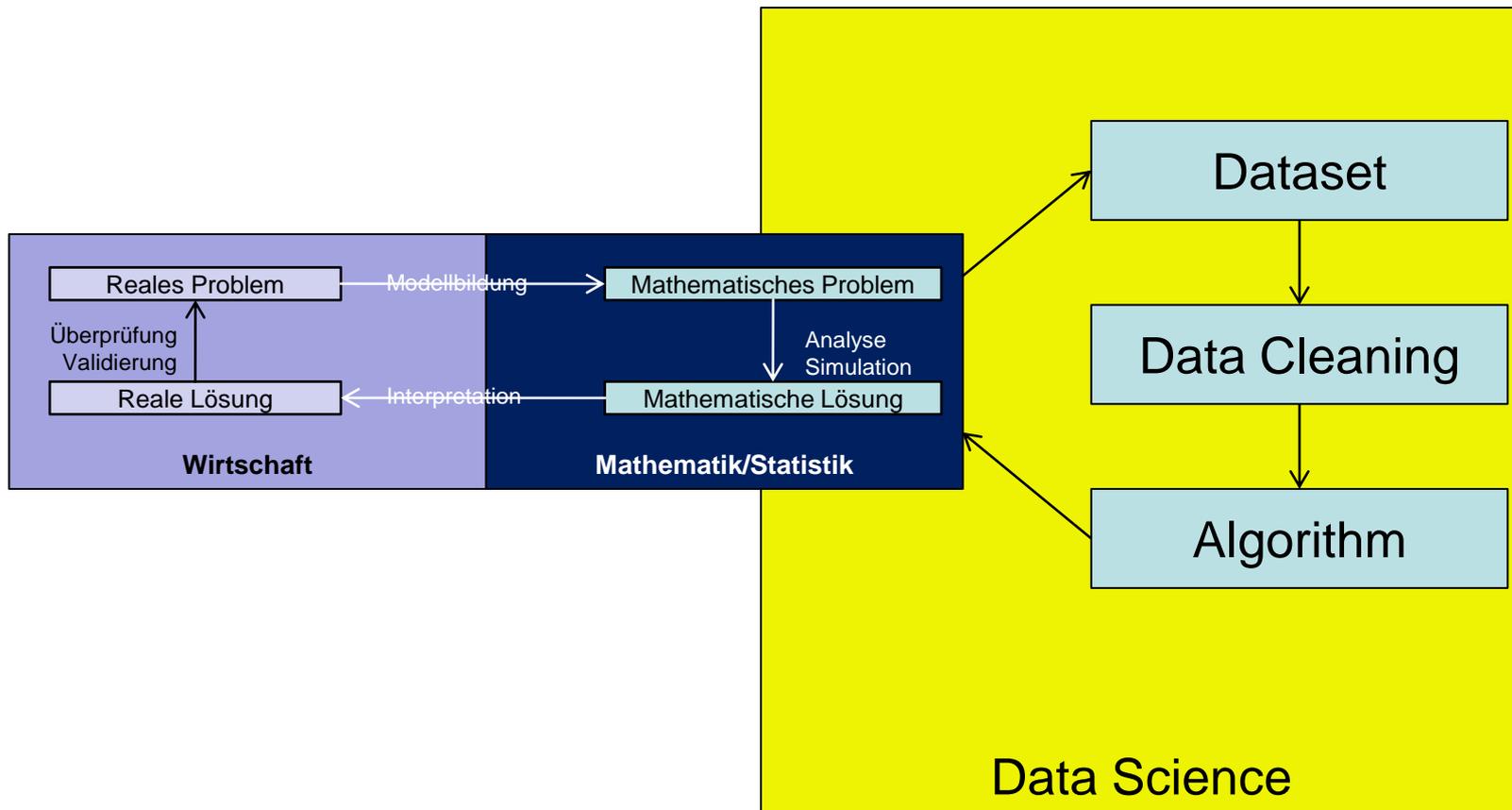


Wie werden Probleme gelöst?



arithmetica

Data Science in der Problemlösung



arithmetica

Kriterien zur Modellauswahl

Richtigkeit

Ein Modell ist richtig, solange es bekannten Tatsachen nicht widerspricht

Prüfung der Richtigkeit eines Modells hängt vom Betrachtungsgegenstand ab

Zulässigkeit

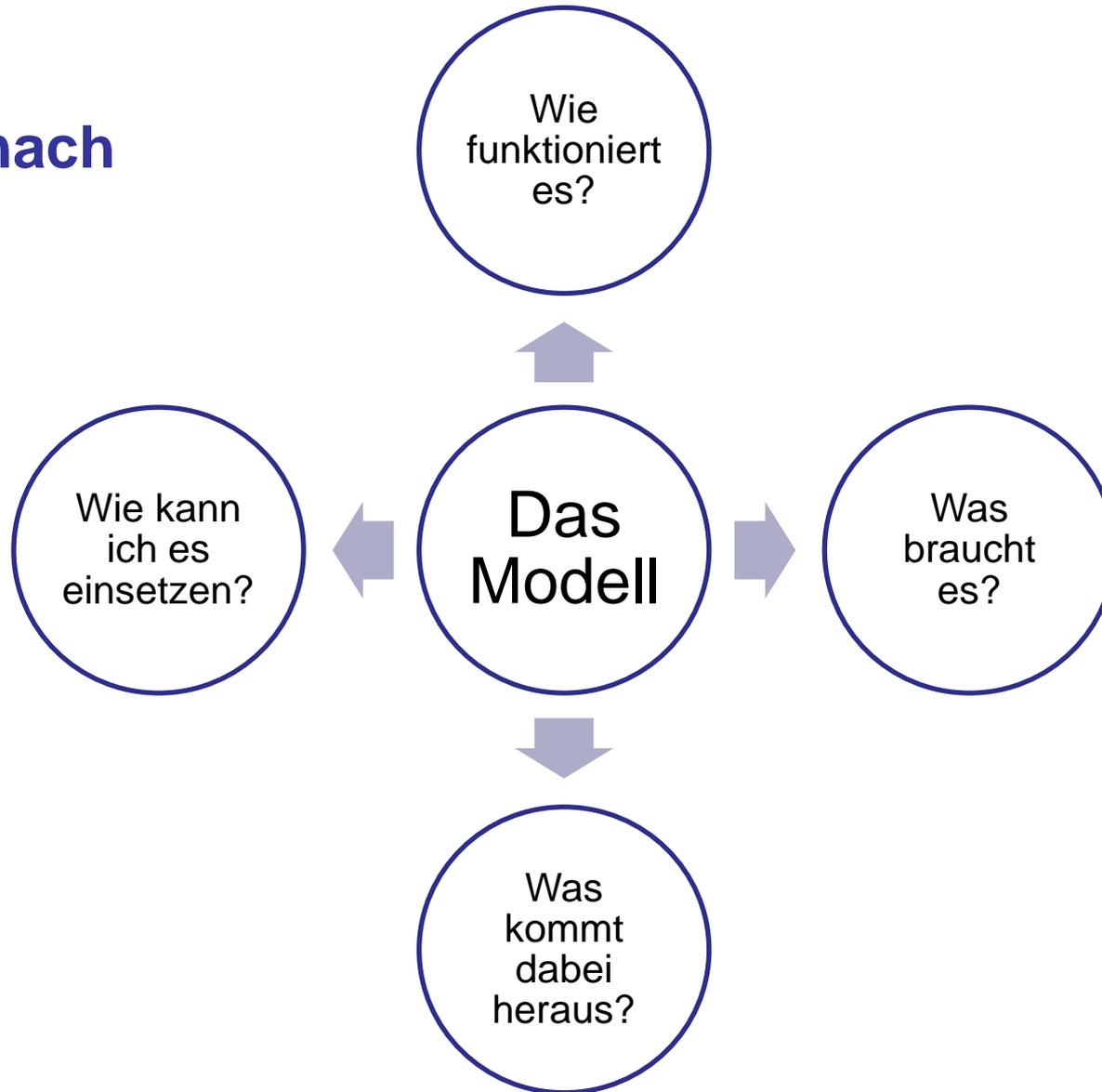
Ein Modell ist (logisch) zulässig, wenn es auf eindeutige Weise formuliert ist und keine Widersprüche enthält

Zweckmäßigkeit

Ein Modell ist zweckmäßig, wenn es keine für das behandelte Problem überflüssigen Anteile enthält

Ein Modell sollte so einfach wie möglich und so kompliziert wie nötig sein. (Minimum Description Length Principle)

Danach



Datenqualität

- Es gibt keine perfekten Daten!!!
 - > Vollständigkeit
 - > Korrektheit
 - > Aktualität
 - > Konsistenz
 - > Verständlichkeit
- Datenpunkte sind Beobachtungen aus der Vergangenheit
 - > Was wurde (nicht) beobachtet
 - > Wann wurde (nicht) beobachtet
 - > Mit welcher Methode wurde gemessen
 - > Wie sorgfältig/genau wurde gemessen
 - > Wie und wann wurden die Daten in das System gebracht
 - > Welche Kontrollen/Prozesse bestehen

Advanced Analytics Techniken

Classification

Decision trees

Rule based classification

Bayesian Classification

Neural Networks

Clustering

K-means clustering

Fuzzy C-means clustering

Self-organizing maps

Hierarchical clustering

Regression

Linear regression

Non-linear regression

Logistic regression

Projection pursuit regression

Kernel based regression

Advanced Analytics Techniken

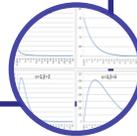
- Anzahl Ereignisse
- Eintrittswahrscheinlichkeiten
- Trends

Häufigkeiten



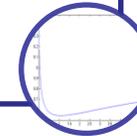
- Wahrscheinlichkeitsverteilungen
- Szenarien
- Benchmarks

Auswirkungen



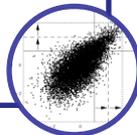
- Katastrophen/ Krisen
- Risikotoleranzen
- Kumulereignisse

Extremwerte



- Korrelationen
- Diversifikation
- Kausalitäten

Abhängigkeiten



- Ratings
- Pricing
- Früherkennung

Scores



- Cluster
- Typologien
- Auffälligkeiten

Ähnlichkeiten



Tools?

- Alle herkömmlichen Statistikpakete sind geeignet
 - > R
 - > Matlab
 - > SPSS
 - > SAS
 - > ...
- Excel ist nur bedingt geeignet (siehe Bsp.)
- Kriterien:
 - > Aufgabenstellung
 - > Methodenvielfalt
 - > Numerische Präzision
 - > Performance
 - > Integrierbarkeit, Automatisierbarkeit

Frisch gekocht oder Dosengericht?

- Üblicherweise erfolgt Integration der Datenanalyse in drei Prozessschritten:
 1. Entwicklung der Analysemethodik anhand konkreter Fragestellung
 2. Automatisierung und Integration
 3. Regelmäßige Validierung/Überprüfung
- Fertige „Tools“ können meist nur 2.
- Datenanalyse „von der Stange“ ohne eigenen Aufwand und Ressourcen liefert wenig werthaltige Ergebnisse

CASE STUDIES

Fragestellung 1: Einkaufsgenossenschaft

A

- Woran kann man erkennen, ob ein Händler in Zahlungsschwierigkeiten ist?

B

- Lässt sich die Anzahl/Höhe der Zahlungsausfälle prognostizieren?

Lösungsansatz zu Fragestellung 1

Statistische Analyse des historischen Einkaufsverhaltens der Händler

Unterschiede im Einkaufsverhalten zwischen ausfallgefährdeten und gesunden Händlern

Verknüpfung des Einkaufsverhaltens mit Bilanzinformationen

A

Frühwarnsystem (Rating) zur frühzeitigen Erkennung auftretender Zahlungsschwierigkeiten

B

Prognosemodell für die statistische Häufigkeit von Zahlungsausfällen/Einzelwertberichtigungen

Fragestellung 2: Liquidität Projektgeschäft

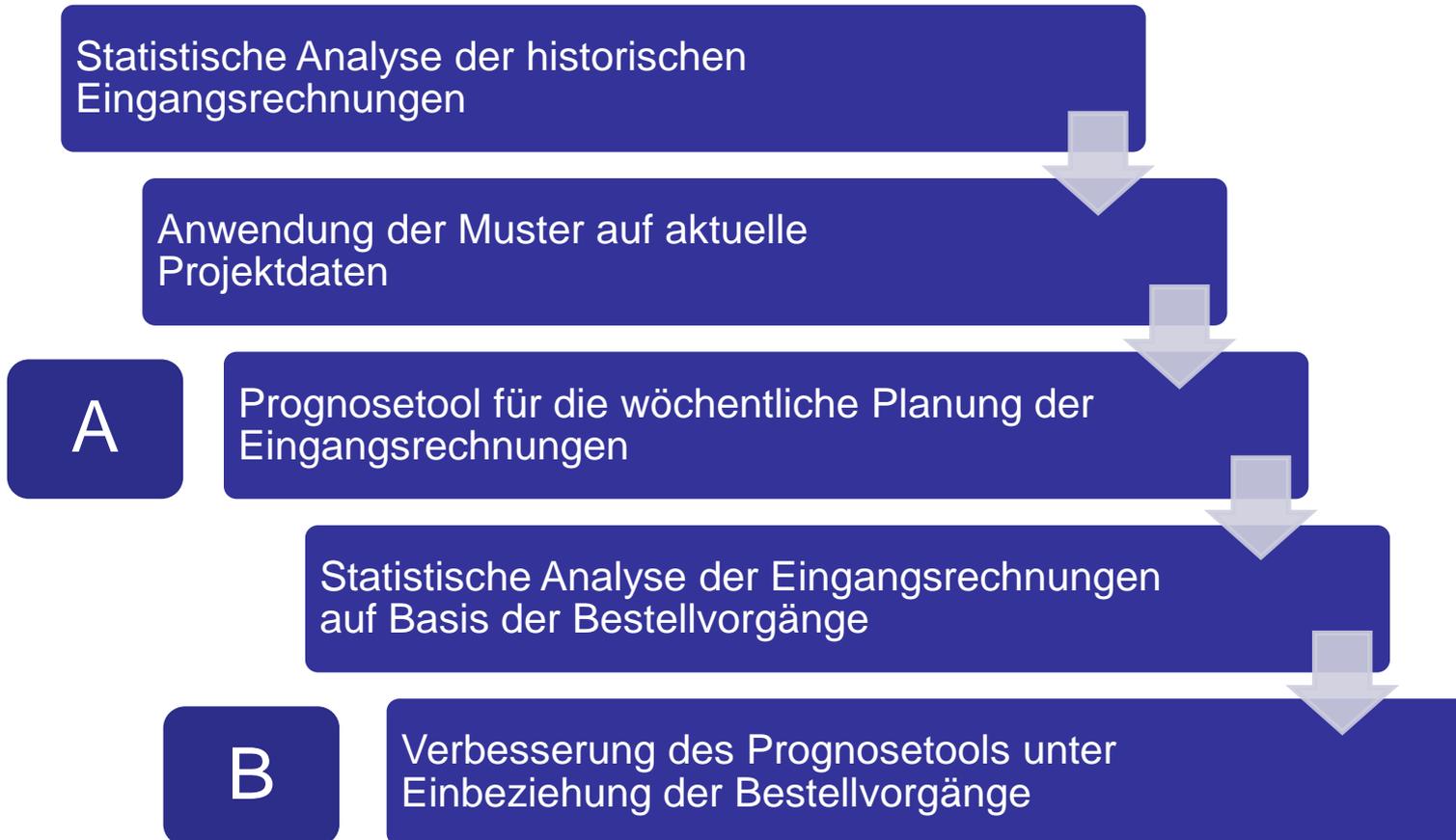
A

- Gibt es erkennbare Muster bei den Eingangsrechnungen, die die Qualität der wöchentlichen Planung verbessern können?

B

- Lässt sich die Cash-Flow-Prognose durch Einbeziehung der Bestellvorgänge verbessern?

Lösungsansatz zu Fragestellung 2



Fragestellung 3: Prognosetool Personalkosten

A

- Das bisherige Tool erfordert viel manuellen Aufwand, insbesondere bei Änderungen in den Annahmen

B

- Das Tool beinhaltet keinen systematischen Soll-/Ist-Vergleich und keine Abweichungsanalyse

C

- Das Prognosemodell berücksichtigt Arbeitsmarktentwicklungen nicht adäquat.

D

- Das Reporting an die Fachbereiche ist nicht ausgereift und passiert zu einem großen Teil manuell.

Lösungsansatz zu Fragestellung 3

Weiterentwicklung des bestehenden Prognose-
tools in MS Excel (ev. mit VBA und Addins)

A

Freistellung der Annahmen zur besseren
Übersicht und für Szenarienberechnungen

B

Aufbau einer IST-Datenbasis und standardisierte
Soll-/Ist-Vergleiche

C

Weiterentwicklung des Modells um zusätzliche
Einflussfaktoren zu berücksichtigen

D

Weitgehend automatisierte Reporting-Funktion

Weitere Projekte

Auffälligkeiten bei Kassabons

- Kundenkarten, Retouren/Stornos, ...

Einflussfaktoren Produktumschlag

- Wetter, Tageszeit, Wochentage, ...

Simulation Preismodelle

- Ergebnissteuerung

Mustererkennung Kaufverhalten

- Externe Einflussfaktoren auf Bestellvorgänge (zB Ölpreis)

Früherkennung Kundenausfall

- Quantitative (Umsatz, Preise) und qualitative (Zufriedenheit, Besuche) Faktoren

...

Schlussfolgerungen aus einem CFO-Round Table mit 10 großen österreichischen Unternehmen (Juni 2017)

- Know-how erforderlich
 - > Eigene Ressourcen oder externer Ressourcenpool?
- Datenqualität ist essentiell
 - > ABER: es gibt keine perfekten Daten!
- Evolutionärer Prozess -> (klein) anfangen!
 - > Der erste Schritt ist immer der schwierigste
- B2C weiter vorangeschritten als B2B
 - > Daten über Kundenverhalten sind mannigfaltiger, haben aber auch schlechtere Qualität
- Zukunft bleibt unsicher, aber bessere Vorbereitung auf zukünftige mögliche Ereignisse
 - > Ziel ist nicht die Glaskugel, sondern fitter zu werden
- Wenig Tradition mit Modell-Governance
 - > Erfahrungen aus dem Finanzsektor nutzen

Future skills

- Critical thinking and problem solving
- Collaboration across networks and leading by influence
- Agility and adaptability
- Initiative and entrepreneurship
- Effective oral and written communication
- Assessing and analysing information
- Curiosity and imagination

Fragen?



Geschäftsführer

Mag. Christoph Krischanitz

Tel.: +43 1 310 59 01-19

Email: christoph.krischanitz@arithmetica.at

www.arithmetica.at