



Data Science im Mittelstand

Kevin Yam, Leiter des Stabes für Mathematische Entwicklungen, Seghorn AG, Bremen

Ablauf

- 1 von der Datenbank zu Data Analytics – Wie kann der Schritt gelingen?
- 2 einige Verfahren aus der Praxis – Welche Lösung für welches Problem?
- 3 Kooperationen finden – Von welchem Partner kann ich was erwarten?
- 4 Adaption für das produzierende Gewerbe
- 5 Data Science in Deutschland – Wo stehen wir im internationalen Vergleich?

Von der Datenbank zu Data Analytics

„Data Scientist:

The Sexiest Job of the 21st Century“ –

Harvard Business Review im Oktober 2012

Anforderungen an den interne Aufbau



- ① Schaffung personeller Strukturen, um mit dem Aufbau einer Data Science-Unit zu beginnen - *Welche Lösung ist die beste für mein Unternehmen?*
- ② Informationstechnischer Aufbau für die gewählte Struktur – *Welche Lösung passt zu welchem Problem?*
- ③ die gesammelten Daten nutzbar machen - *Welche Software ist die Richtige?*
- ④ Data Science im Unternehmen richtig kommunizieren und positionieren - *Wie stelle ich mich intern auf? Wie beuge ich Widerständen vor?*

personelle Anforderungen

Was muss ein Data Scientist mitbringen?



analytische
Fähigkeiten

Datenbank-
management

Programmier-
fähigkeiten

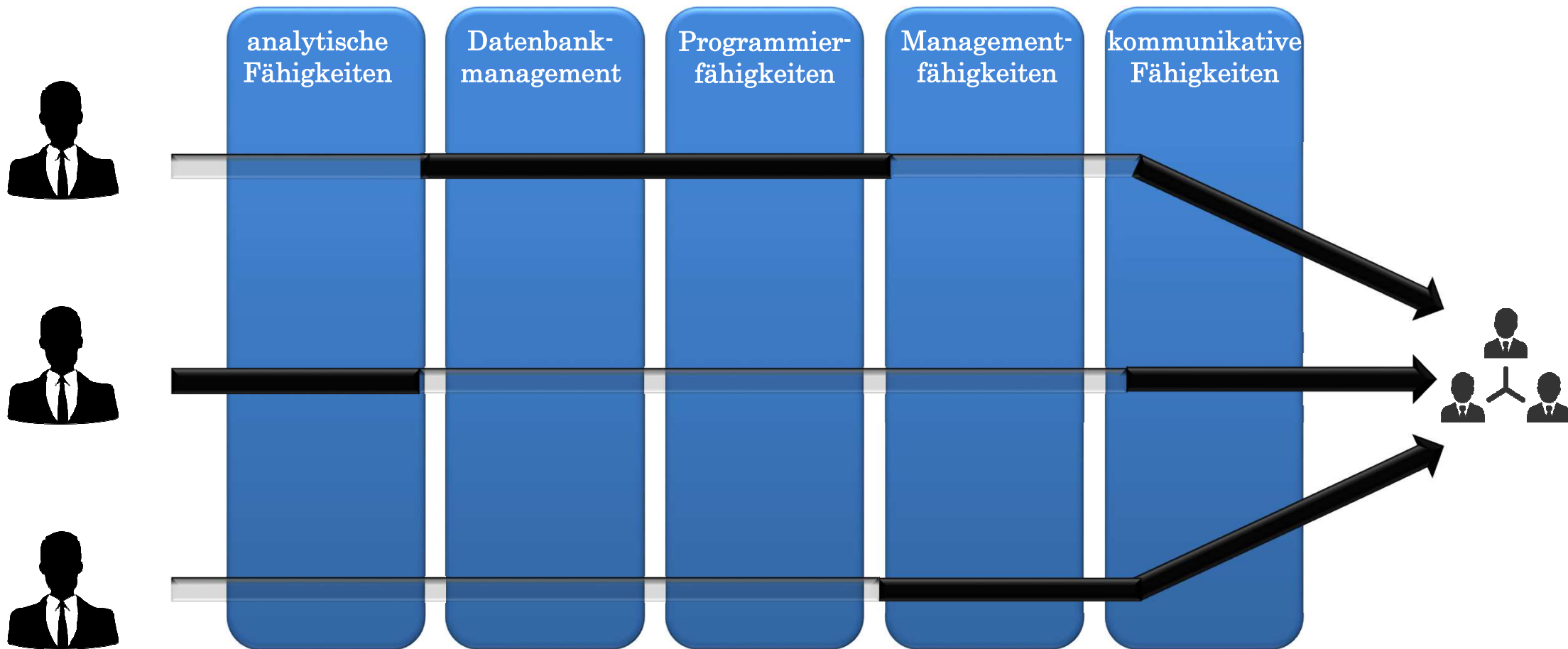
Management-
fähigkeiten

kommunikative
Fähigkeiten

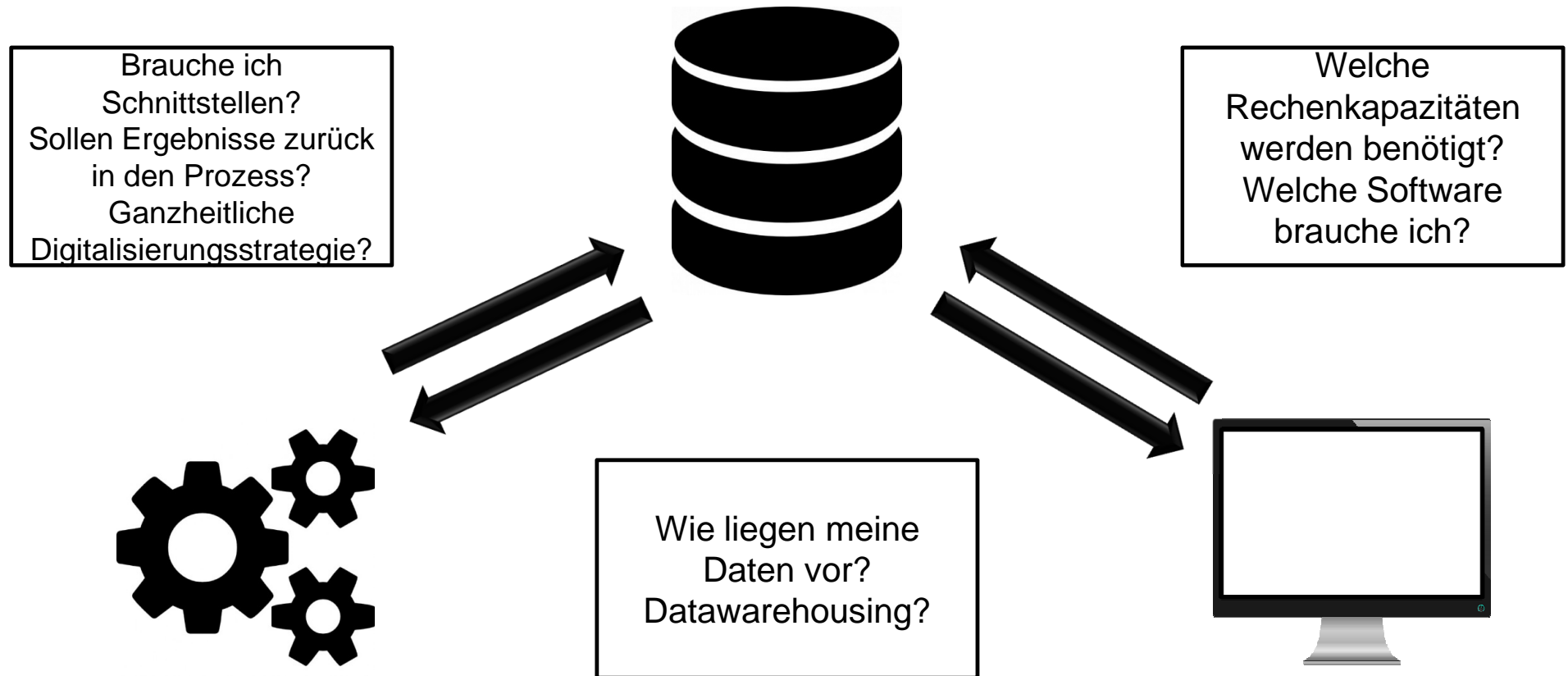


personelle Anforderungen

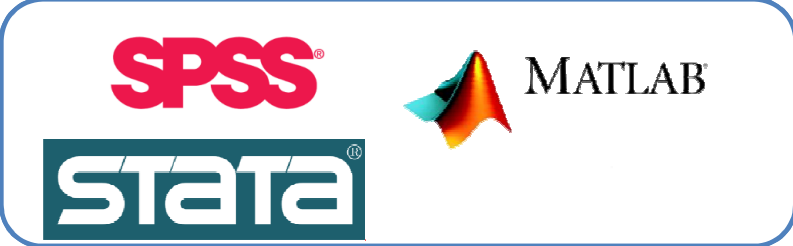

anstatt Individuallösung – auf Teamlösung zurückgreifen



technische Anforderungen



bestehende Softwarelösungen

	Commercial	Open Source
Usability	einfache, intuitive Bedienung	komplexere Bedienung
Kosten	kostenpflichtige Lizenzen	Freeware
sonstiges	Tooldatenbanken, regelmäßige Updates, Vorhandensein von Schnittstellen	gemeinsame Plattform mit der Wissenschaft, Quellcode einsichtig
Beispiele	z.B. SPSS, Stata oder Matlab	z.B. python, Octave oder R
		

1

Welche Software für welches Problem?

„make or buy“ Entscheidung? – zwei Beispiele

The screenshot shows the IBM SPSS Statistics Data Editor interface. The 'Analyze' menu is open, and the 'Regression' submenu is selected. The 'Linear...' option is highlighted. The data table below shows the following variables and their values for the first few rows:

gender	active	obesity	diabetes	bp
54 Female	Yes	No	No	Hypotension
-74 Male	Yes	Yes	No	Hypertension
-74 Female	Yes	Yes	Yes	Hypertension
-74 Male	Yes	No	No	Normal

The screenshot shows the Microsoft Excel interface. A chart is displayed with a trendline. The 'Trend and Progress' dialog box is open, showing the 'Trend' tab. The 'Trendline' section is selected, and the 'Linear' option is chosen. The 'Display Equation on Chart' checkbox is checked. The 'Display R-squared Value on Chart' checkbox is also checked. The 'Options' section is expanded, showing the 'Intercept' set to 0. The 'Display Equation on Chart' checkbox is checked. The 'Display R-squared Value on Chart' checkbox is also checked. The 'Options' section is expanded, showing the 'Intercept' set to 0. The 'Display Equation on Chart' checkbox is checked. The 'Display R-squared Value on Chart' checkbox is also checked.

Kommunikation & Positionierung

Zentralisierung

- disruptive Kompetenzen an verantwortlicher Stelle verorten
- gelebte Open Door Policy
- wichtig: gemeinsames Alignment mit der IT-Abteilung und Geschäftsführung

Dezentralisierung

- Fokussierung auf das Thema ermöglichen
- Personalressourcen klar regeln
- Sensibilisierung für das Thema schaffen und innerhalb des Unternehmens Data Science als gemeinsames Projekt angehen
- Kommunikationshürden abbauen



Wahl des Zentralisierungsgrades auch eine Ressourcenfrage

2

einige Verfahren aus der Praxis

„data rich but knowledge poor“

Wozu Data Science im Mittelstand?

branchenfremder Wettbewerb nimmt zu:

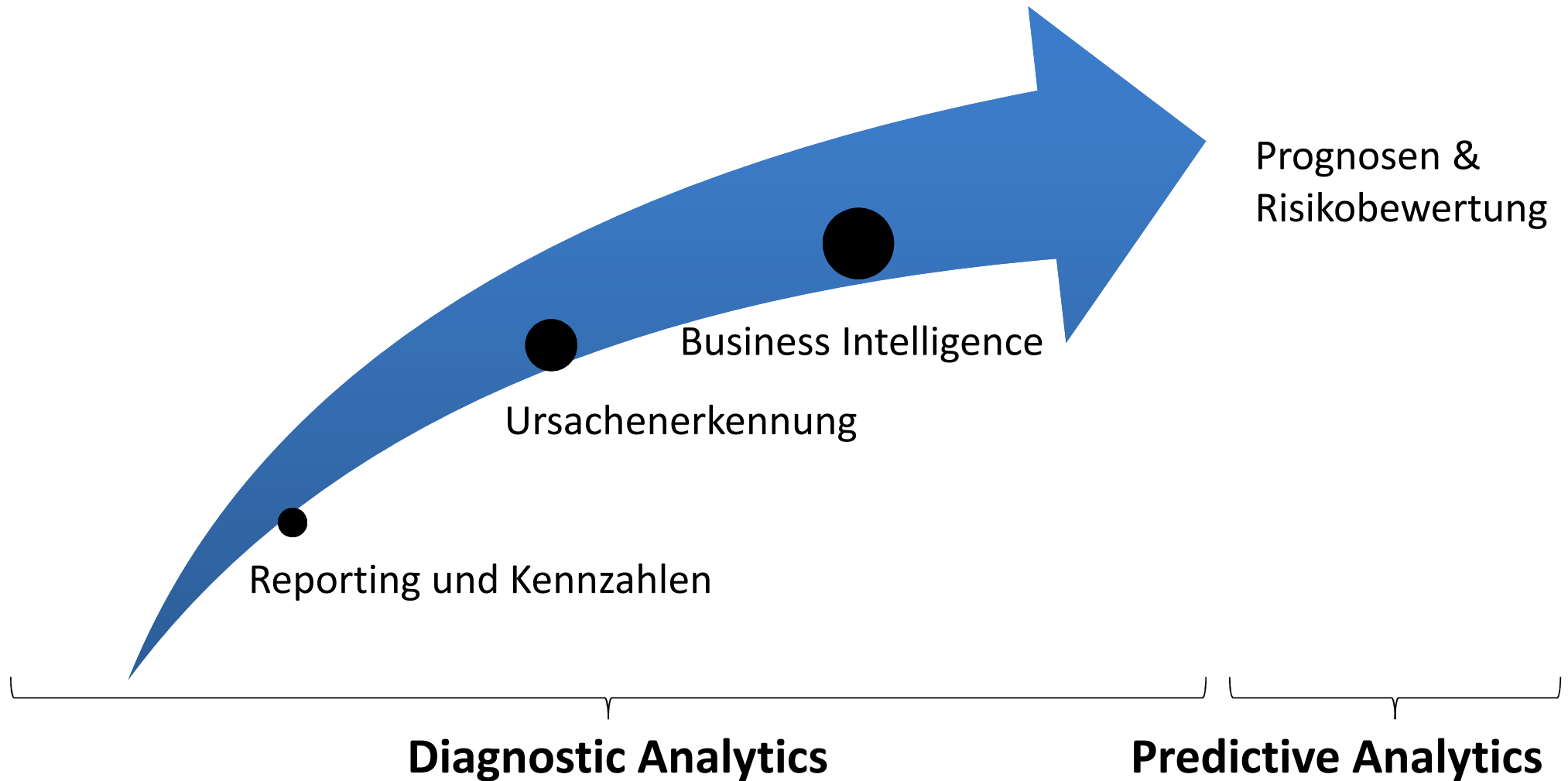
- datengetriebene Unternehmen zerstückeln die Wertschöpfungskette und dringen in konservative Märkte (z.B. Fintech-Unternehmen)
- Digitalisierung macht Data Science notwendig
- Data Science als Vorgabe angelsächsischer Investoren

aktueller Status im Risikomanagement/ Controlling:

- Mittelständler betreiben Risikomanagement mit Standard-Office Produkten
- wenig automatisiertes Risikomanagement
- Entscheidungen werden aufgrund von Bauchgefühl getroffen

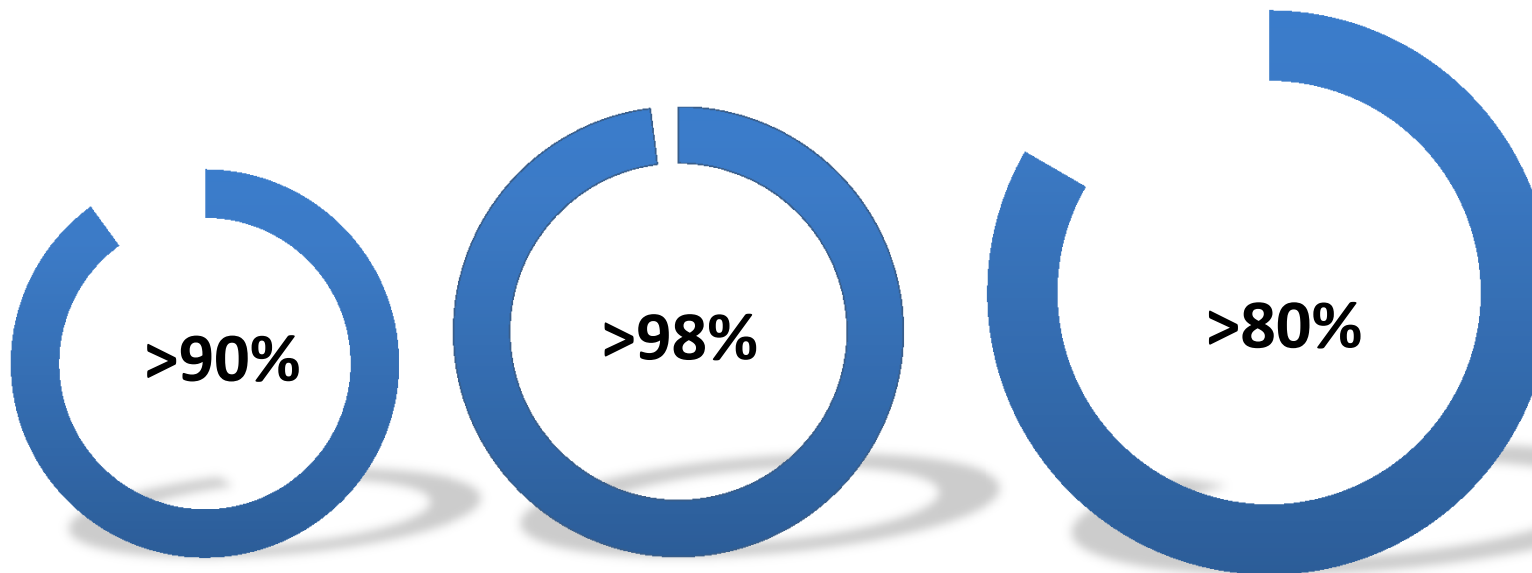


Diagnostic & Predictive Analytics



Business Case Forderungsmanagement

mithilfe von Data Science konnten



Scoringverfahren

entwickelt werden mit einer Genauigkeit von über 90%

laufende Portfolios

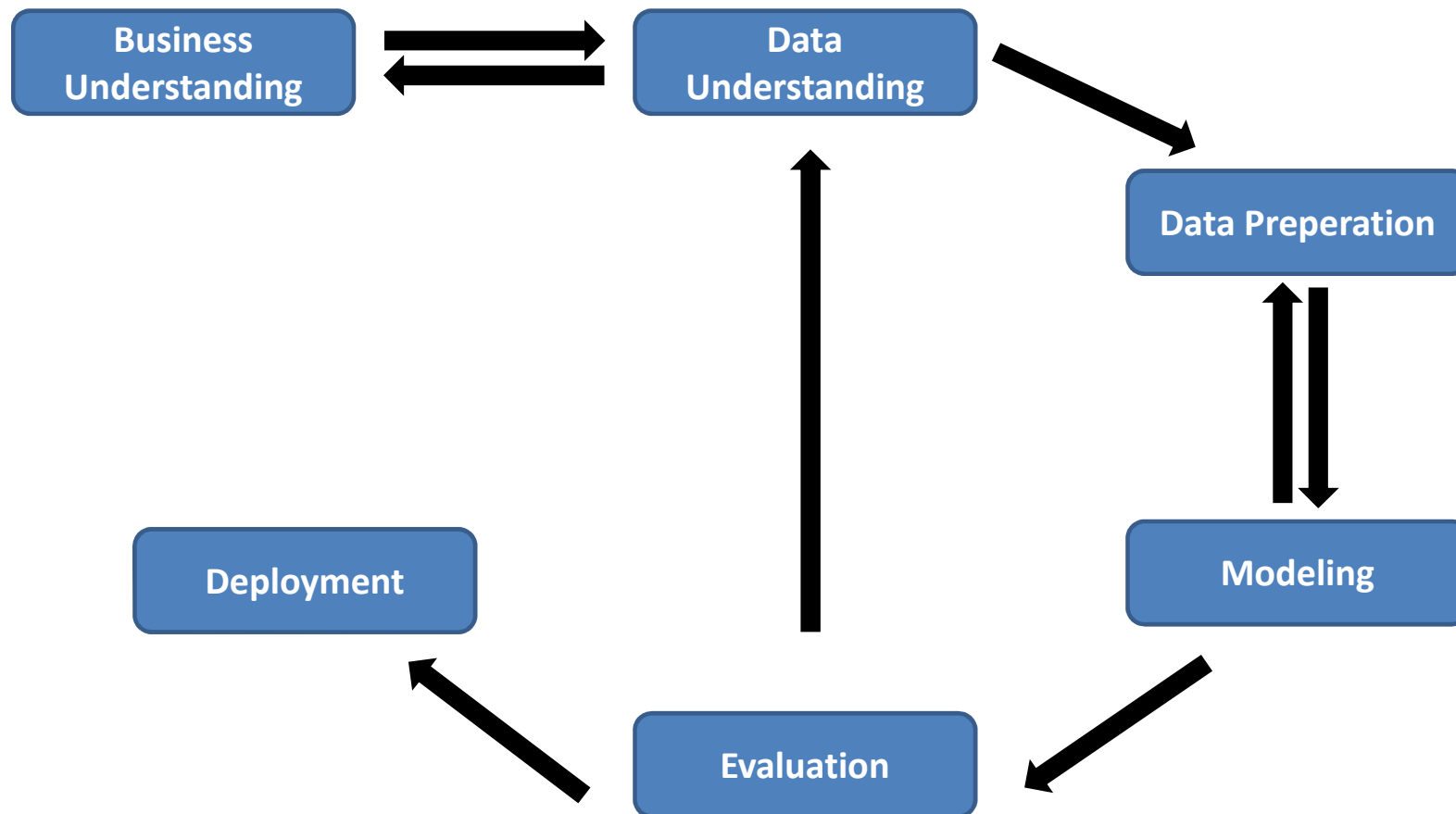
mit einer Genauigkeit von über 98% prognostiziert werden

sinnvolle Sekundärdaten

zu 80% der bestehenden Daten angefügt werden

- verlustbringende Prozessschritte identifiziert und eingespart werden
- weitreichende Automatisierungen der Bearbeitungsschritte durchgeführt werden
- quantitative Risikomodelle implementiert werden

Standardprozess zur Datenanalyse



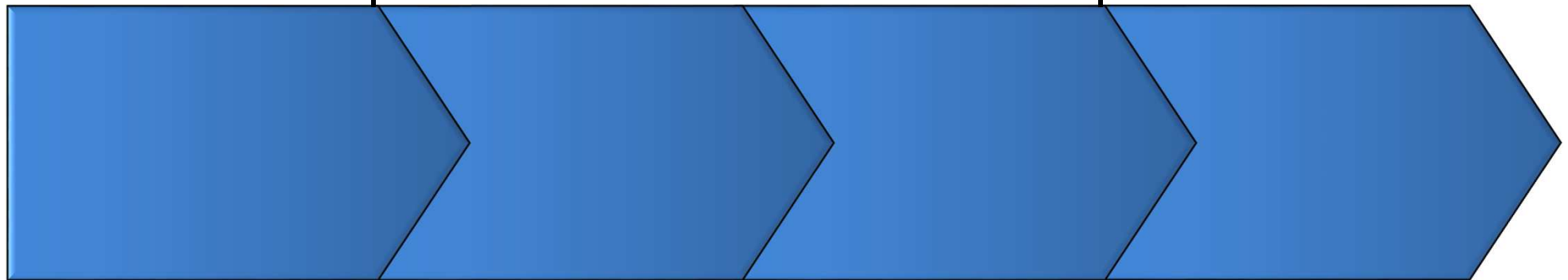
Prozess – und Kosteneinsparungen

Prozesse verstehen

- interne Wertschöpfungskette analysieren
- Handlungsspielräume finden
- rechtliche Parameter berücksichtigen

Einflussfaktoren finden

- Welche sind die entscheidenden Treiber?
- Ist es möglich Einflussfaktoren zu bewerten?



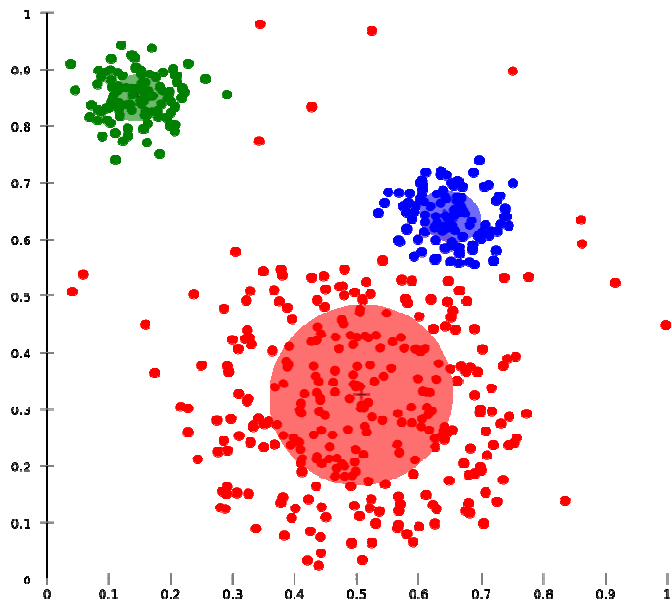
Clusteranalyse

- unterschiedliche Mandanten anhand der Einflussfaktoren in logisch sinnvolle Gruppen klassifizieren

Modellierung

- Suche nach einem Modell, das beobachtetes Verhalten bestmöglich abbildet

Klassifikation am Beispiel Clusteranalyse



hierarchischer Algorithmus

maximale interne
Homogenität bei maximaler
externen Heterogenität

Klassifizieren von Objekten
und Anomalien finden

Ziele:

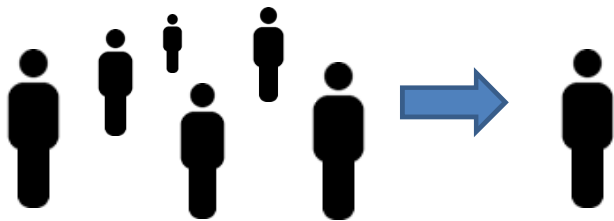
- Transparente Sicht auf den Geschäftsverlauf – Wo ist das Problem?
- effiziente Prozesse – Was lässt sich vermeiden?

Welche Verfahren eignen sich?

robuste Verfahren

z.B. Portfoliobewertung

- stabile Mittelung
- makroskopische Betrachtung
- spezifisches Anwendungsfeld
- kein Overfitting
- höhere Nachvollziehbarkeit der Ergebnisse



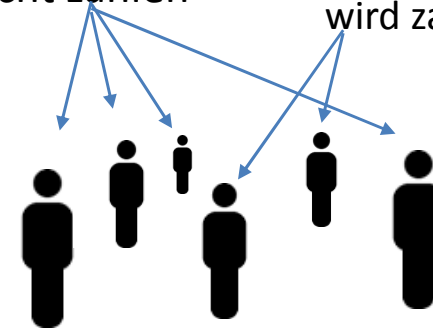
spezifische Verfahren

z.B. Scoring

- sehr sensitiv
- detaillierte Betrachtung
- fehleranfällig
- implementierungsaufwändig

wird nicht zahlen

wird zahlen



Risikomanagement mit Data Science

Datenanalyse ermöglicht:

- Erkennung von Einflussfaktoren auf das Ausfallrisiko
- Portfoliotaxierung
- Prozesskostenrisiken

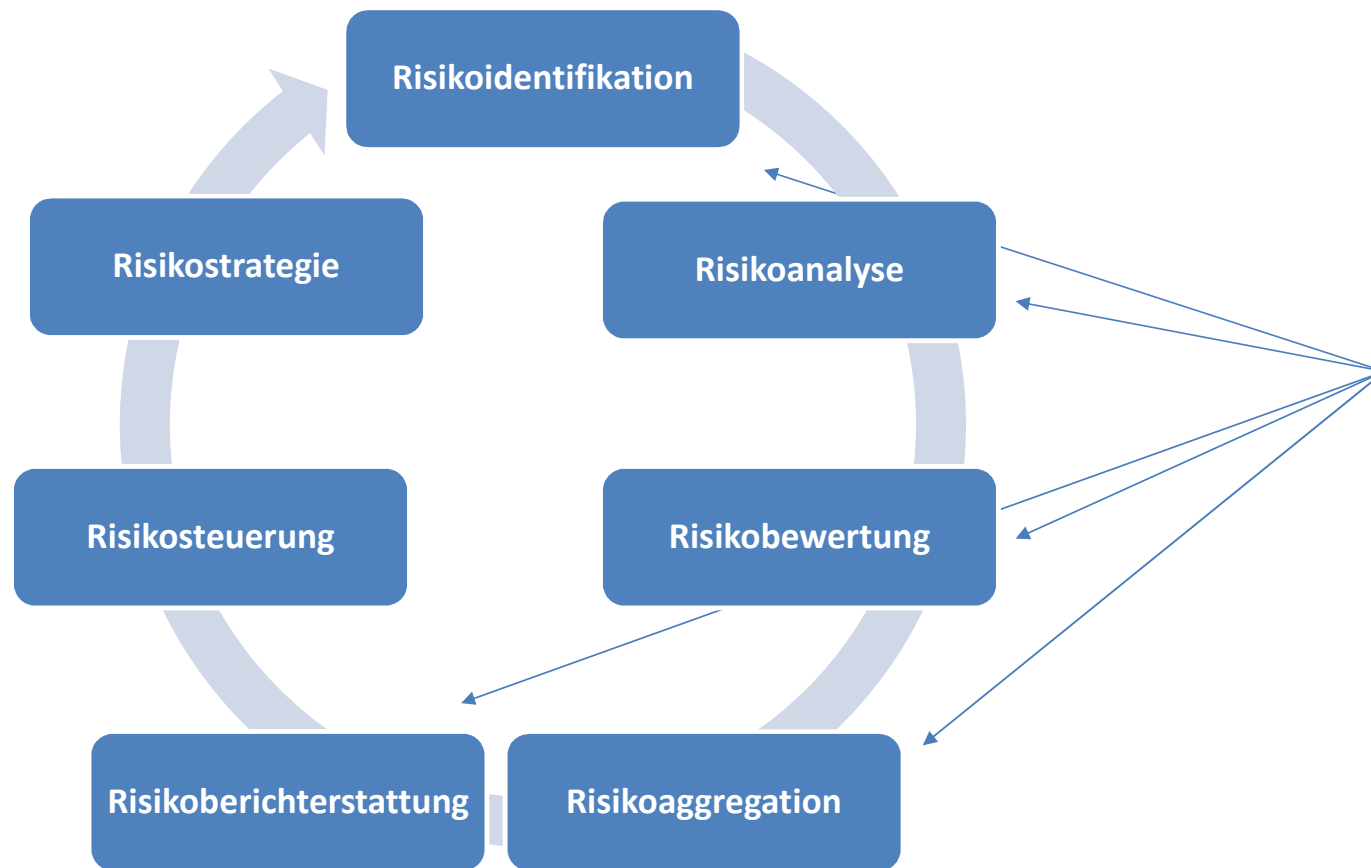
erforderlich aufgrund:

- gesetzliche Regulierungen (z.B. Basel, Solvency und Handelsgesetz)
- Compliance
- Marktanforderungen

Optimierung durch:

- Erkennen von wichtigen Prädiktoren
- Verbesserung der Datenqualität
- Methodenoptimierung

Risikokreislauf mit Data Science

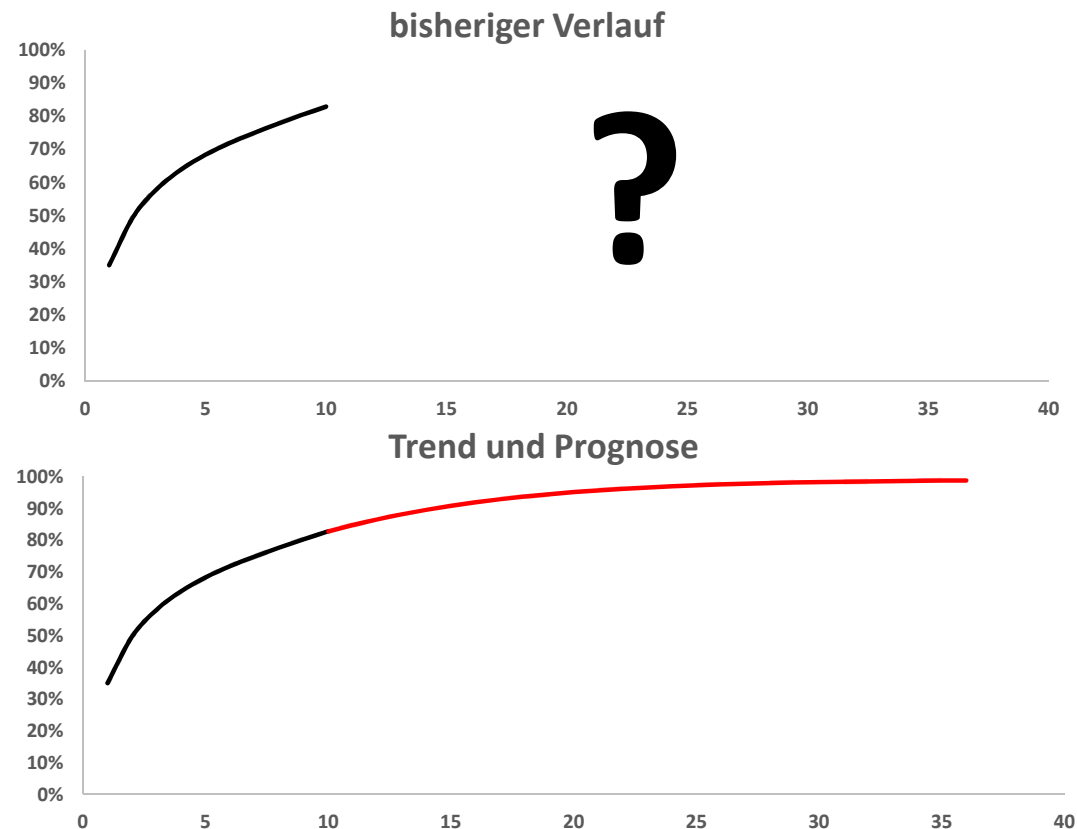


mithilfe von Data
Science nahezu
automatisierbar

Praxisbeispiel: Prognoseverfahren

von der Bleistiftmethode zur automatisierten Vorhersage mithilfe von Data Science

- Grundlage: Verlauf der ersten Monate
- Oszillation
- Modell für valide Prognosen erstellt und installiert werden

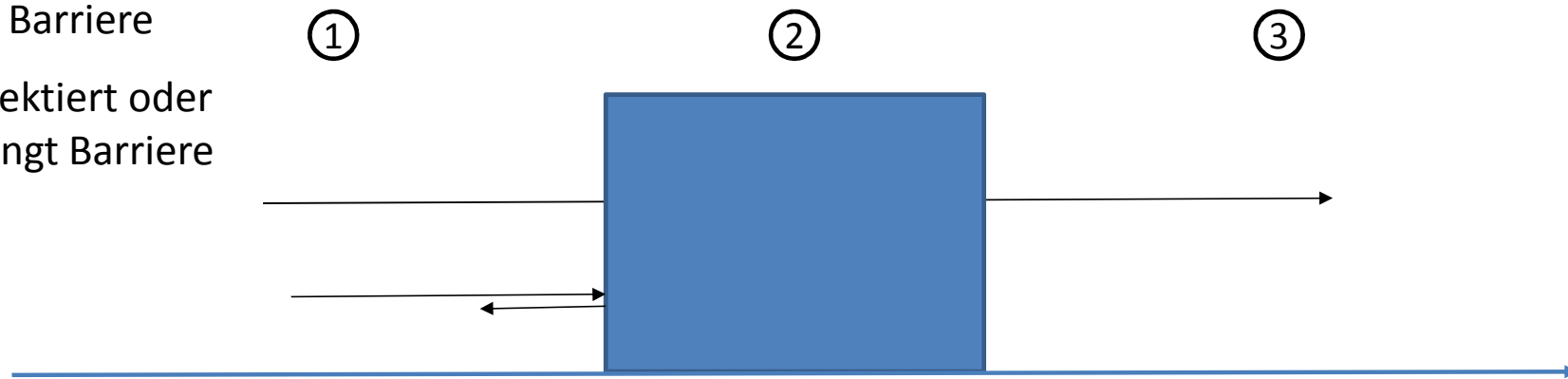


 Risikobewertung

Anpassung naturwissenschaftlicher Modelle

die Potentialbarriere für freies Teilchen aus der Physik

- ① freies Teilchen
- ② trifft auf Barriere
- ③ wird reflektiert oder durchdringt Barriere



Charakteristika des Individuellen

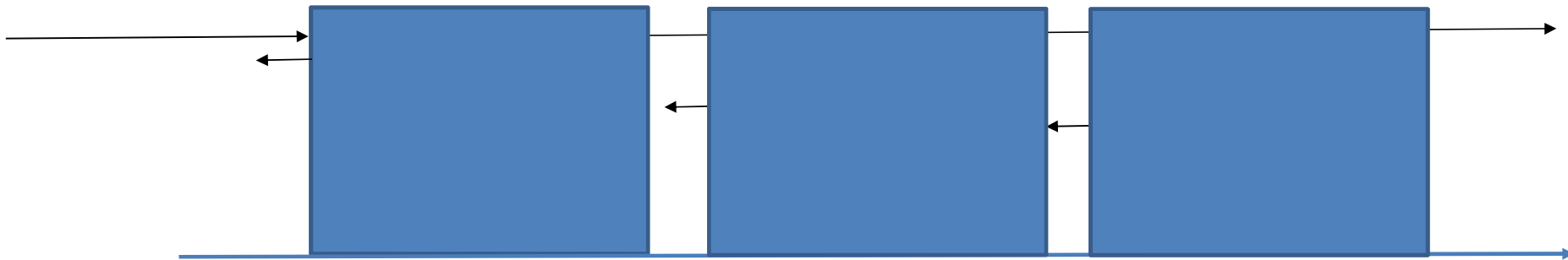


Modell



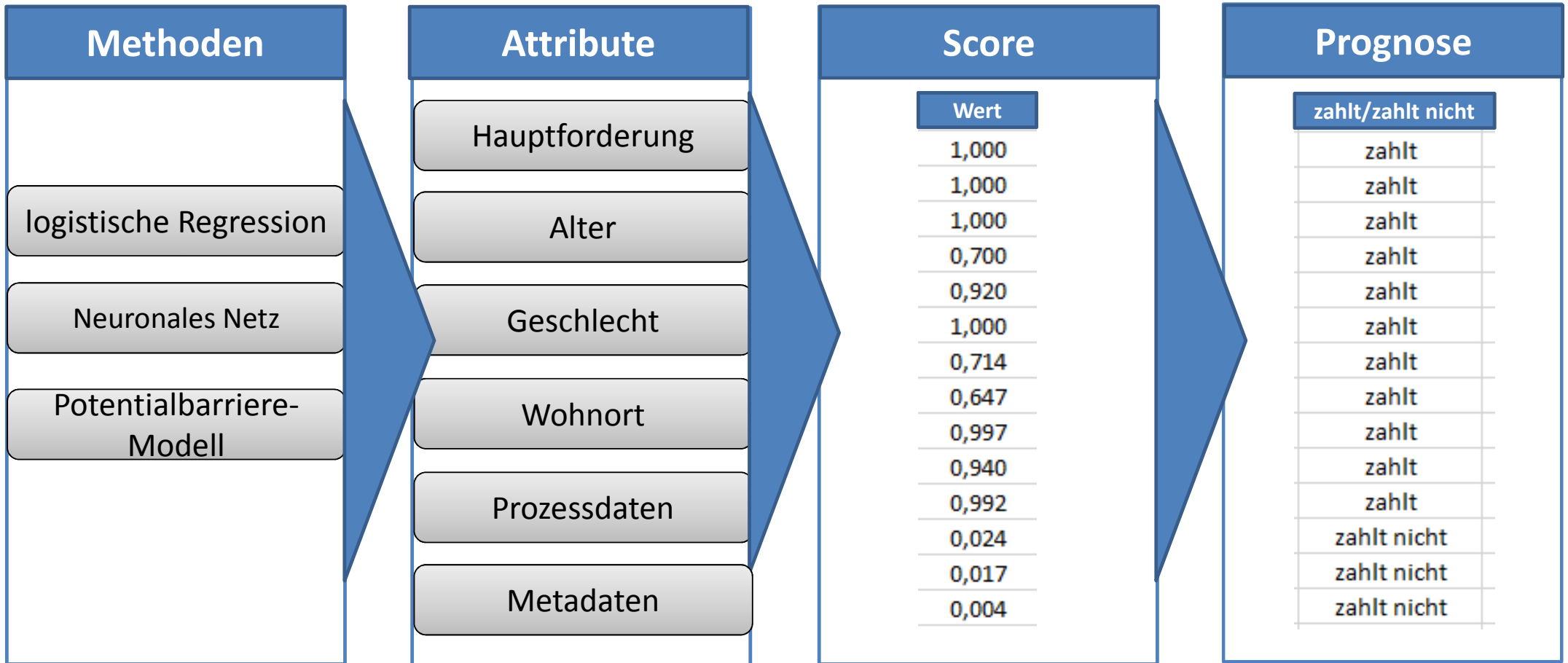
Score

mehrstufiges Potentialbarriere-Modell



- einzelne Barrieren symbolisieren die unterschiedlichen Mahnstufen
- nach jedem Prozessschritt kann für den Kunden eine neue Zahlungswahrscheinlichkeit berechnet werden
- einzelne Schritte im Mahnprozess können individuell bewertet werden
- Automatisierung im operativen Bereich

Praxisbeispiel: Scoring



statista 

schufa

Sekundärdaten



DLSTATIS
wissen.nutzen.

3

Kooperationen finden

wissenschaftliche Kooperationen

als Human-Resources Pool

als Entwickler & Sparringspartner



- Universitäten an Realdaten aus 1. Hand interessiert
 - über Kooperationen Risiken teilen
 - kostengünstig Projekte auslagern
 - Reputationssteigerung durch Publikationen in Journalen
- „war for talents“ im Bereich Data Science besonders ausgeprägt
 - mit interessanten Themen für Abschlussarbeiten Absolventen für das eigene Unternehmen begeistern
 - weitere Möglichkeiten des HR-Marketing

3 Forschungsverbund

Vorteile

- attraktive Fördergelder
- Synergieeffekte mit anderen Unternehmen auf thematischer Ebene
- Anknüpfungspunkte über die thematische Kooperation hinaus
- Reputationssteigerung
- Überwindung der eigenen Betriebsblindheit



Gefahren

- Knowhow an Konkurrenz
- zu viele Interna und Schwachstellen offenbaren
- Betriebsgeheimnisse erschweren Projektarbeit
- hoher Planungsaufwand

Zulieferer

Unternehmen

Kunde

die eigene Wertschöpfungskette als Kooperationspool nutzen

Adaption für das produzierende Gewerbe

4 Adaptionpotential

mögliche Anknüpfungspunkte mit Data Science



Beschaffung

z.B.

- Lagerbindung von Produktionsmitteln reduzieren
- bedarfsgerechte Produktionsmittelplanung



Fertigung

z.B.

- Kostentreiberanalyse
- Prozessoptimierung



Lagerhaltung

z.B.

- Just-in Time Produktion
- Auswertung bestehender Scanner-Systeme
- Optimierung von Arbeitsabläufen und des Lageraufbaus



Marketing und Vertrieb

z.B.

- individuelle Kundenansprache
- exakte Budgetplanung

Data Science als Marketingunterstützung

individuelle und abgestimmte Kundenansprache
Streuverluste im Marketing können verringert werden



abhängige Variable im Modell einfach Messbar: Umsatz

Klassifikationsanalysen

- effiziente Vertriebsplanung
- Beschwerdemanagement (automatische Bearbeitung von User-Feedback)
- Früherkennung neuer Markttrends (Inhalte von Social-Media Plattformen können ausgelesen werden und komprimiert zur Auswertung bereitgestellt werden)

Regressionsverfahren

- Pricing-Modelle (Erstellen eigener Preis-Absatz-Funktionen mithilfe historischer Daten)
- Optimierung von Marketingkampagnen
- Optimierung von Angeboten

Adaptionpotential im produzierende Gewerbe

Industrie 4.0 und „smart-factory“



auf der Managementebene

- Entscheidungsgrundlagen
- Handlungsempfehlungen
- Risikofrüherkennung
- Kapitalbindung reduzieren
- Produktqualität verbessern



- Wie kann ich meinen cashconversion cycle optimieren?

im Prozess

- Fehlererkennung
- Wartezeiten verkürzen
- Prozessoptimierung zur Reduktion von Gemeinkosten



- substituieren sich Faktoren im Prozess, sodass diese letztendlich nicht sichtbar werden?
- Welche sind entscheidende Kostentreiber und wie kann ich sie reduzieren?

Data Science in Deutschland

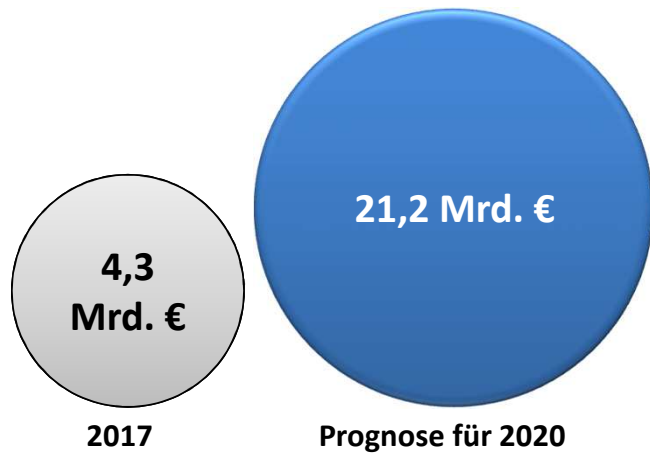
über

2 500 000 000 000 000 000

Bytes an Daten täglich

Marktentwicklung rund um das Thema KI

Umsatzvolumen im KI-Geschäft



Anstieg der Arbeitsproduktivität durch KI bis 2035 in Prozent

USA	+35%
Japan	+34 %
Deutschland	+29%
Großbritannien	+25%
Frankreich	+20%

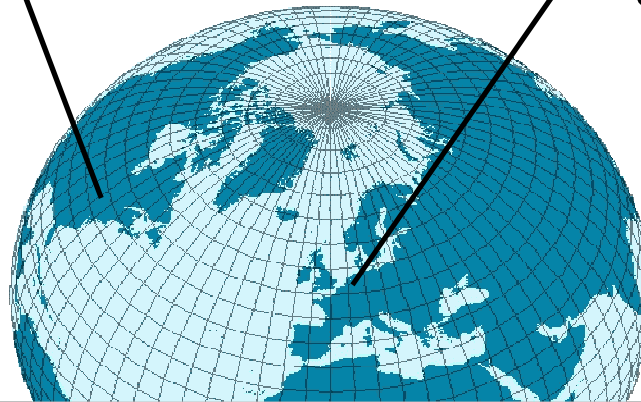
Automatisierungspotenzial

Gastgewerbe	66%
Produktion	64%
Logistik	60%
Bergbau	54%
Einzelhandel	54%
Land- und Forstwirtschaft	50%
Großhandel	50%
Versorger	45%
Finanzbranche	44%
Bauwesen	44%
Immobilien	44%
Unterhaltungsbranche	42%
Informationsdienstl.	41%
Verwaltung	41%
Techn. Dienstl.	39%

Wo stehen wir im internationalen Vergleich?

Amerika

- weltweite thematische Dominanz der Big Player wie Google, Amazon, IBM und Wal-Mart
- Google als Miterfinder von Big Data Technologien (MapReduce Algorithmus, Big Table etc.)
- universitär: seit einigen Jahren explizite Studiengänge
- nur 1/3 der US-Data Scientists stammen aus den USA



Europa

- aufgrund der Datenschutzes wurde das Thema lange Zeit in Deutschland kritisch betrachtet
- Start-Ups haben eine innovationstreibende Rolle (als Beispiel: Finleap, Pair Finance, Insur- und Legaltechfirmen)
- universitär: erst kürzlich Einzug in die deutschen Universitäten erhalten

Wo geht's hin im Mittelstand?



Chancen

- neue Marktfelder entstehen durch Erkenntnisse aus dem Bereich Data Science
- Prozesskosten können reduziert werden
- Controlling wird effizienter
- genauere Prognosen geben Planungssicherheit

Widerstände

- Datenschutzfragen
- Investitionshemmnisse
- Unwissenheit zu gewissen Fragestellungen
- kein externes Personal
- Personal ist schwer zu finden und für den Mittelstand zu begeistern

Lösungen

- Kooperationen nutzen
- Beratungsleistungen in Anspruch nehmen
- Kompetenzteams bilden
- Mut haben

Vielen Dank für Ihre Aufmerksamkeit



yam@seghorn.de



Kevin Yam

