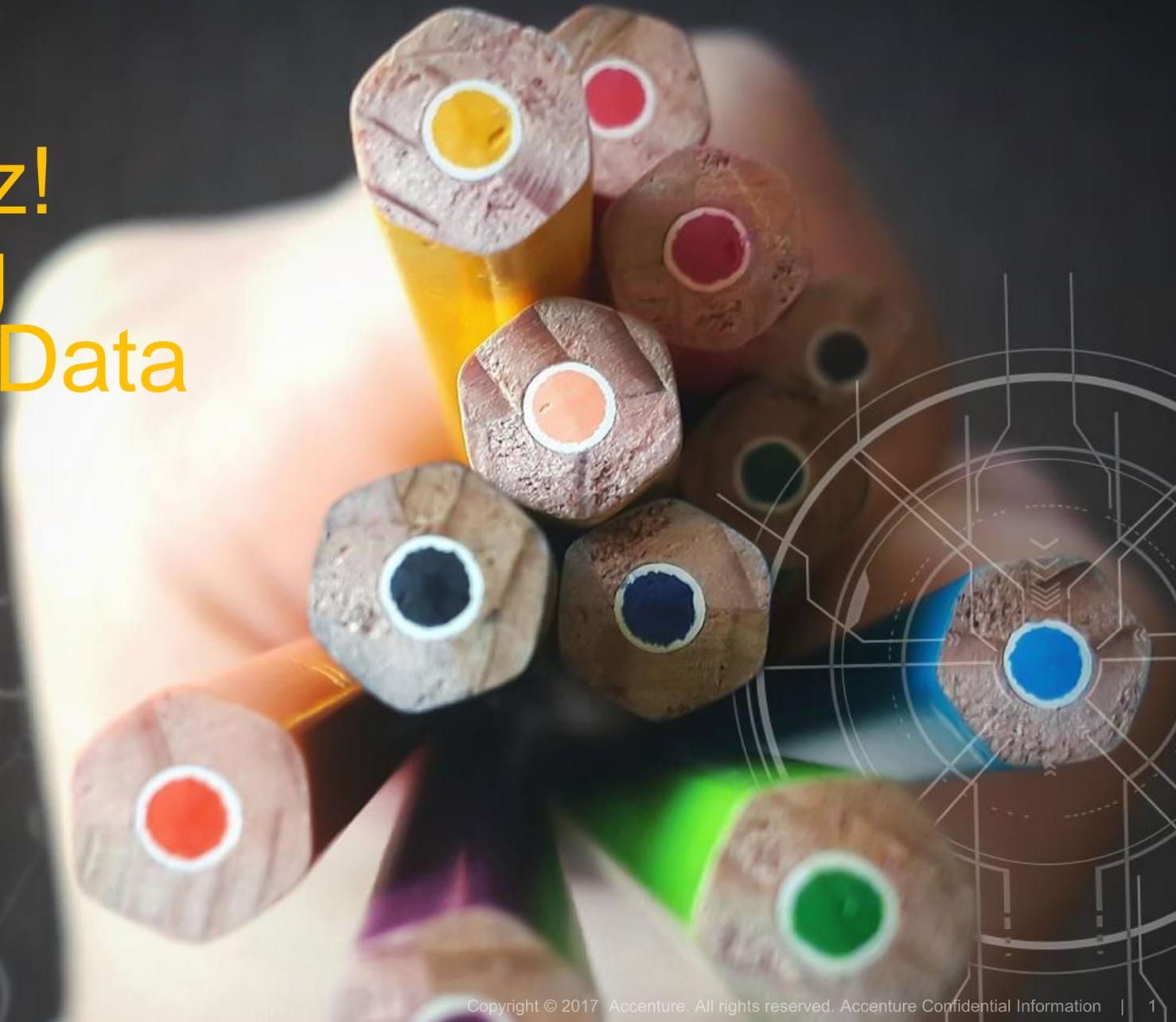


accenture[>]digital

Analytic im Einsatz! Betrugserkennung auf Basis von Big Data

Karol Sobiech

High performance. Delivered.



accenture
High performance. Delivered.

accentureconsulting

accenturestrategy

accenturedigital

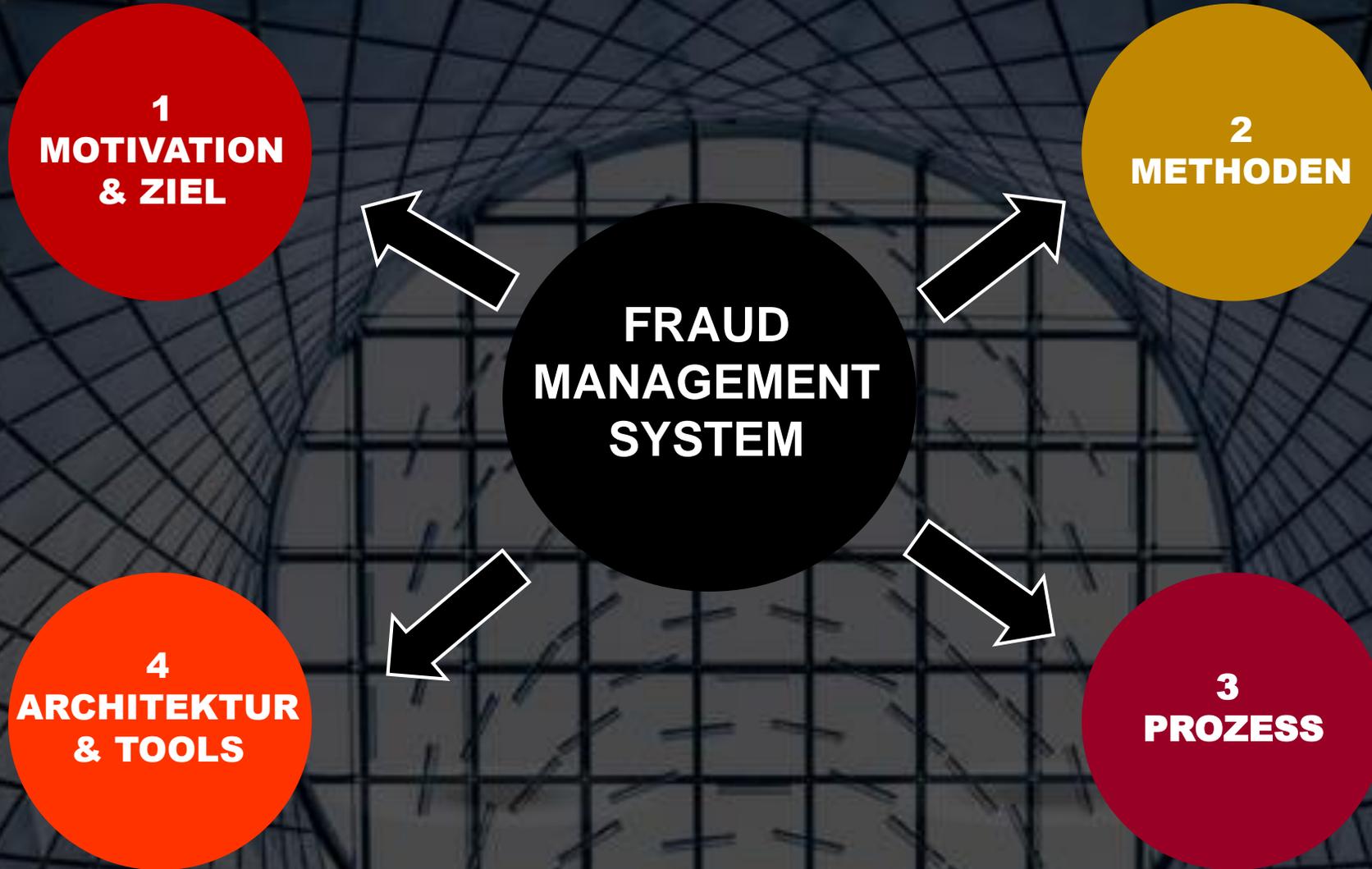
accenturetechnology

accentureoperations

ACCENTURE GLOBAL DELIVERY NETWORK



AGENDA



MOTIVATION UND ZIEL

Wie soll das Unternehmen mit potenziellen Betrugsfällen umgehen?



✓ **Ziel:** Optimierung - Verluste möglichst minimieren bei geringeren Kosten

✓ **Weg zum Ziel:** Identifikation der Fälle mit einer hohen Betrugswahrscheinlichkeit

 **BUSINESS CASE**
Lohnt sich das überhaupt?
Wie betroffen ist mein Unternehmen durch den Betrug?
PROOF-OF-VALUE FÜR FMS

METHODEN FÜR BETRUGSERKENNUNG

HERAUSFORDERUNGEN

- Evolution der Betrugsmuster
- Betrüger versuchen unauffällig zu bleiben
- Betrüger arbeiten zusammen
- Es gibt nicht viele bekannte Betrugsfälle



EXPERTENREGELN

Industrieexperten definieren, in welchen Fällen eine Transaktion verdächtig ist.

- (+) einfach zu implementieren
- (-) müssen vorher definiert werden
- (-) basieren auf bekannten Mustern



ANOMALIE-ERKENNUNG

Untypische, auffällige Fälle (Anomalien / Outliers) werden als verdächtig gekennzeichnet.

- (+) unbekannte Muster können entdeckt werden
- (+) verlangt keine historischen Fälle
- (-) "Anomalie" muss definiert werden

FOKUS DIESER PRÄSENTATION



PREDICTIVE ANALYTICS

Auf Basis der historischen Fälle, werden Merkmale identifiziert, die die Wahrscheinlichkeit des Betrugs erhöhen.

- (+) sehr effizient
- (-) historische Beobachtungen sind erforderlich
- (-) neue Muster können nicht entdeckt werden



SOZIALE NETZWERKANALYSE

Durch die Analyse der Beziehungen zwischen der Parteien werden Betrugsmuster identifiziert.

- (+) erkennen Kooperationsschemen
- (-) sind eher eine Ergänzung für andere Methoden

FOKUS DIESER PRÄSENTATION

ANOMALIE-ERKENNUNG

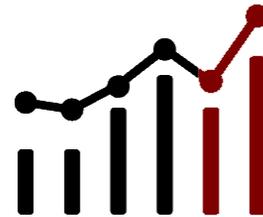
Anhand eines Beispiels aus Krankenversicherung

Use Case:

Wie können die Versicherten identifiziert werden, deren Krankenversicherung (bewusst oder unbewusst) mitbenutzt wird?

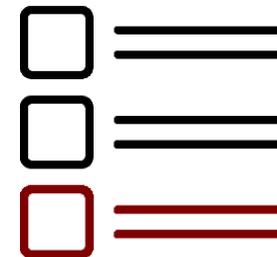
PEER-GROUP-ANALYSE

„Der Patient verhält sich anders als ähnliche Patienten“



ASSOZIATIONSANALYSE

„Die Kombinationen der Leistungen des Patienten sind untypisch“



BREAK-POINT-ANALYSE

„Das Verhalten des Patienten hat sich mit der Zeit geändert“

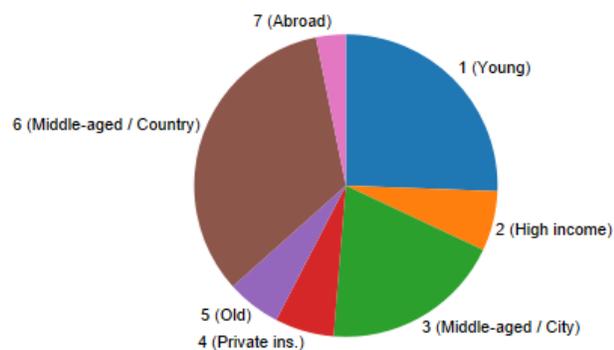


PEER-GROUP-ANALYSE

„Der Patient verhält sich anders als ähnliche Patienten“

1) CLUSTERING

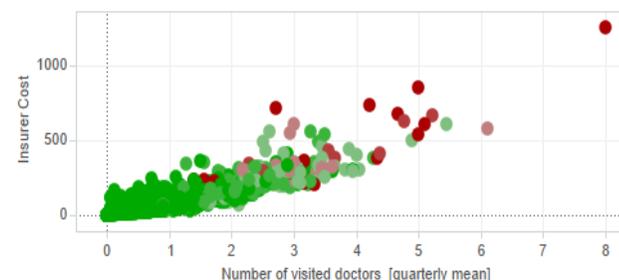
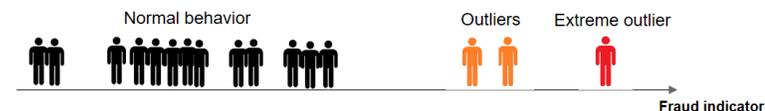
Teile Patienten
in homogene Segmente



Input: demografische Daten

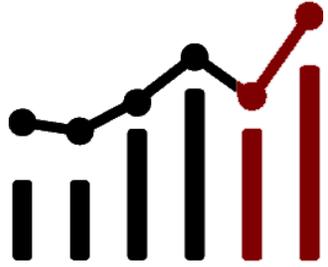
2) OUTLIER-DETECTION

Vergleiche Patienten
Innerhalb der Segmente



Beispiele der analysierten KPIs:

- Anzahl der gefahrenen km zum Arzt
- Anzahl der besuchten Fachgebiete

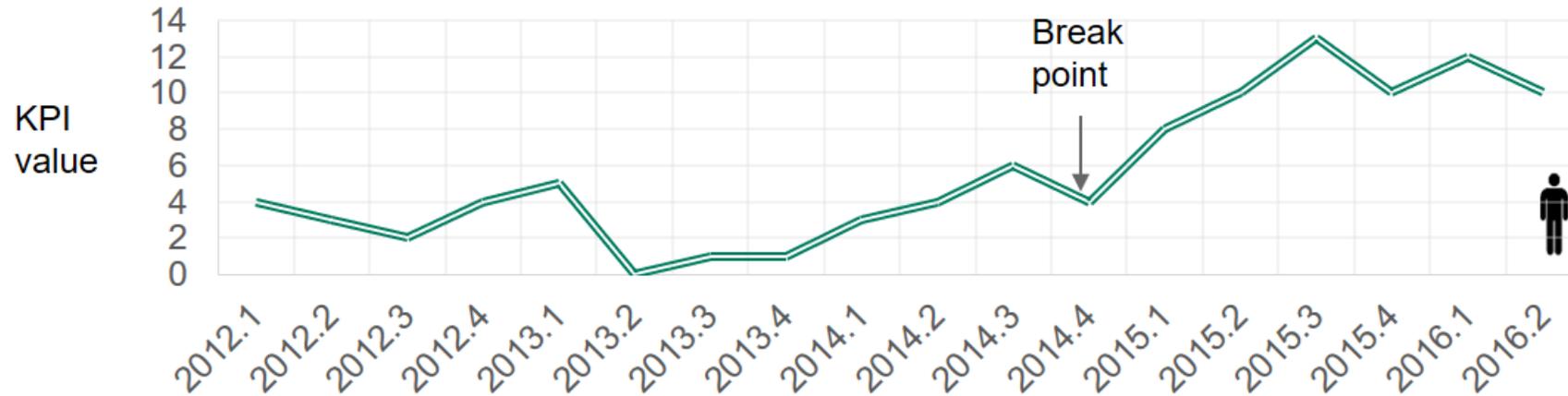


BREAK-POINT-ANALYSE

„Das Verhalten des Patienten hat sich mit der Zeit geändert“

ZEITREIHENANALYSE

In der Historie der Patienten werden Wendepunkte gesucht

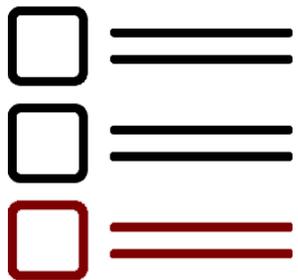


Beispiele der analysierten KPIs:

- Anzahl der besuchten Ärzte (pro Monat)
- Die Kosten der Leistungen (pro Monat)



*Hat diese Änderung eine Erklärung?
z.B. eine Krankheit (s. Diagnose)
in diesem Zeitraum?*



ASSOZIATIONSANALYSE

„Die Kombinationen der Leistungen des Patienten sind untypisch“



Wie definiert man, welche Kombinationen von Diagnosen und Leistungen / Medikamente „normal“ und erlaubt sind

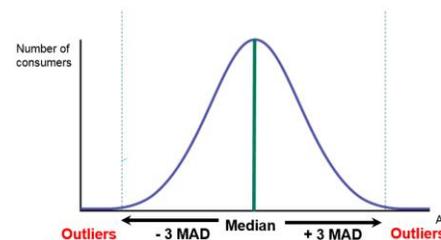
Expertenregeln

Statistische Analyse

Validierung und Korrektur

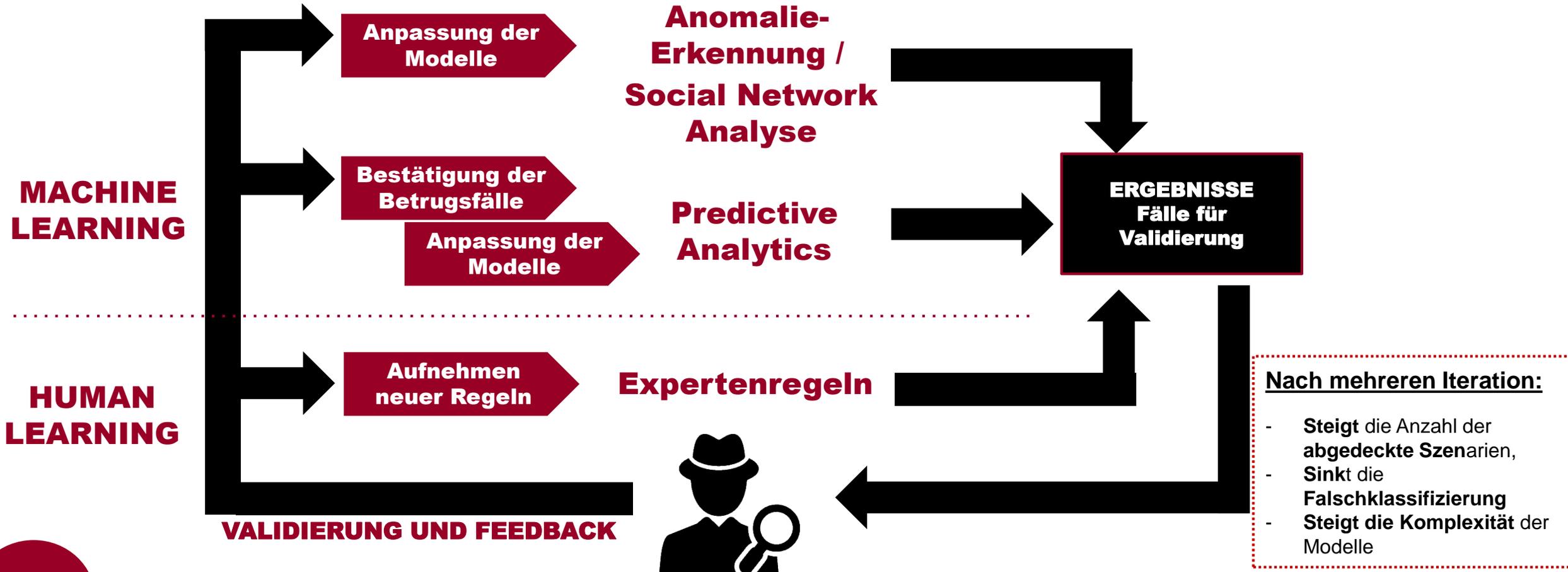
1) Anhand der Historie definiere die häufigen Kombinationen von Diagnosen und Leistungen / Medikamenten
Lerne aus den Daten – häufige Kombinationen im System werden als normal und konform betrachtet.

2) Identifiziere die Ausreißer in der Verteilung
Was ist die Verteilung der Leistungen im Bezug auf Patientenmerkmale? Was ist der typischer Alter der Leistungsbezieher?



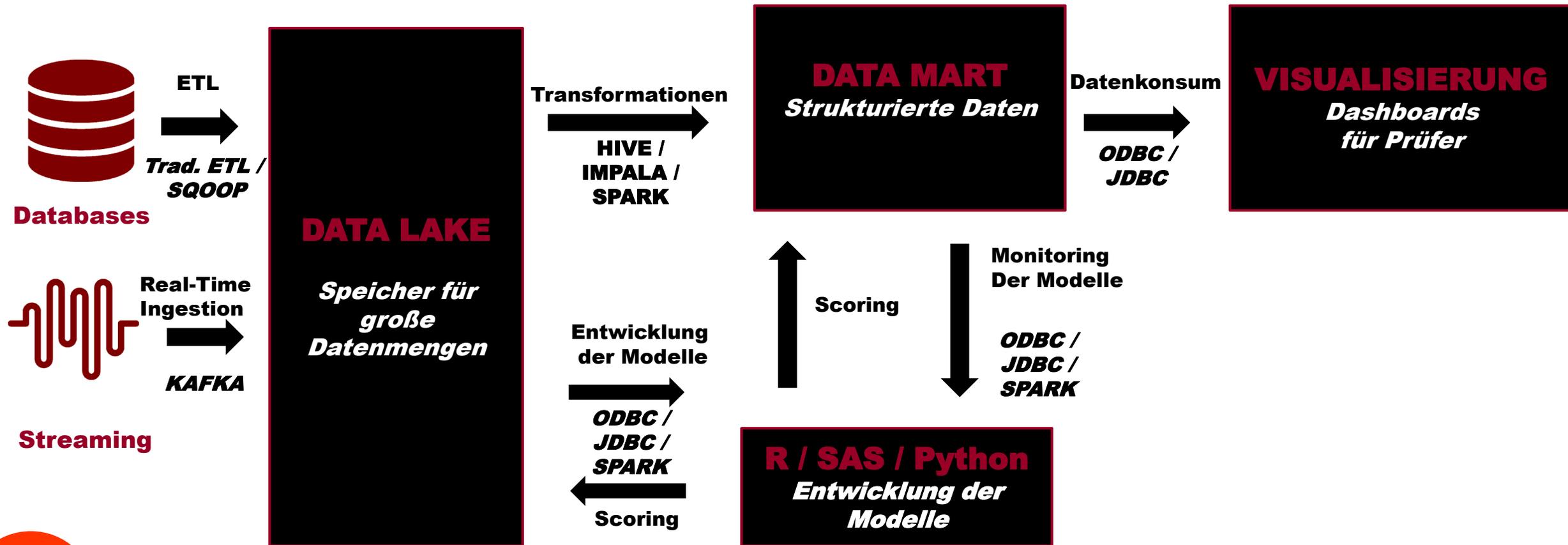
PROZESS

Evolution des Betrugserkennungssystems



ARCHITEKTUR

Big Data Architektur für Fraud Management System



SPARKLYR

Synergie von R und Spark

R

- die (wahrscheinlich) populärste Programmiersprache für Data Science
- große Community
- große Datenmengen müssen in die Umgebung kopiert werden
- abhängig vom Memory der Umgebung

Spark

- schnelle Datenverarbeitung in Memory auf der Hadoop Plattform
- bietet Machine Learning Library MLlib
- verfügbar über API für Scala, Java & Python

SparklyR

- Interface für Spark von R
- Produkt von Rstudio (<http://spark.rstudio.com/>)
- Verfügbar auf CRAN
- Einfacher Einstieg in Spark für R Entwickler



dplyr

Datenverarbeitung



mllib

Spark Machine Learning



h2o

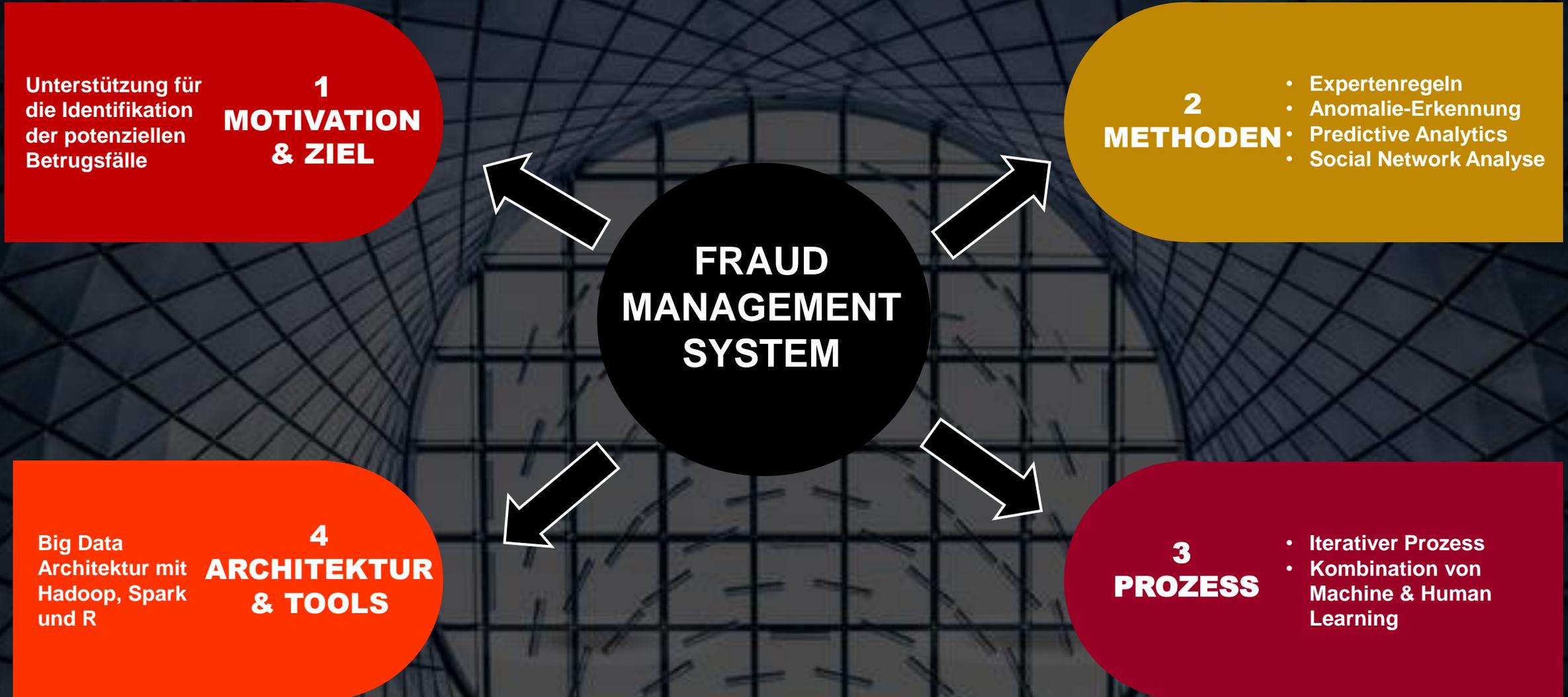
zusätzliche Bibliothek für
Machine Learning



Extensions

mit Spark API

ZUSAMMENFASSUNG



FRAGEN?



**KAROL
SOBIECH**

.....
Consultant,
Accenture Analytics
Kaistr. 20
40221 Düsseldorf
karol.sobiech@accenture.com
+49 175 5763628