

„Millionenbetrug beim Semmering-Tunnelbau“

- noe.ORF.at



Betrugsfall Semmering-Tunnelbau

- Chefeinkäufer der Baustelle & Komplizen unterschlugen Baumaterialien
- Auch gegen Zulieferfirmen wird wegen ausgestellter Scheinrechnungen ermittelt
- Schadensumme wird hier auf weit über 2 Millionen Euro geschätzt
- Die mutmaßlich illegalen Vorgänge blieben lange unentdeckt
- Fielen erst auf, als im Jahr 2018 rund 300.000 Liter Diesel von der Baustelle verschwanden
- Ermittlung gegen 33 Personen, rechtskräftige Anklage gegen 6 Beteiligte, Prozesstermin steht noch nicht fest (Stand Januar 2021)

→ Die meisten Betrugsfälle gelangen jedoch nicht an die Öffentlichkeit

Kurzvorstellung Applied Statistics



Unser Team

Unser Team besteht primär aus promovierten Mathematikern, Physikern und Informatikern mit starkem Statistik und Data Science Hintergrund.



Unsere Projektpartner

Wir bieten unseren Kunden seit 2014 individuelle Lösungen im Bereich datengetriebene Entscheidungsfindung, Statistik und Data Science.



Individuelle Lösungen

- Wir verstehen uns als Dienstleister
- All unsere Lösungen werden maßgeschneidert entwickelt
 - richten sich somit nach den individuellen Problemstellungen der Projektpartner
- Lösungen werden als Open-Source Software implementiert
 - Software kann so von Projektpartnern mit- und weiterentwickelt werden
 - Somit sollen nachhaltige Lösungen erzielt werden
- Es fallen aus Prinzip *keine Lizenzgebühren* an
- Kunden entscheiden über Umfang und Komponenten

Wir helfen Ihnen Ihre Industrie 4.0 Business Cases umzusetzen

Datengetriebene Entscheidungsfindung

- ✓ Data Mining
- ✓ Empfehlungsdienste
- ✓ Risiko Modellierung



- ✓ Empfehlungen für gezielte Investitionen in **OEBB** Bahnstationen
- ✓ Treffsichere Steuerungs- und Marketingmaßnahmen für **wikifolio** auf Basis von Datenanalysen

Statistische Optimierung

- ✓ Betriebsoptimierung
- ✓ Verbrauchsprognosen
- ✓ Automatisierte Prozesssteuerung



- ✓ Kosteneinsparungen durch automatische Routenplanungen bei der **Deutschen Bahn**
- ✓ Hohe Kosteneinsparungen durch verbesserte Produktionsressourcenplanung bei **BASF**

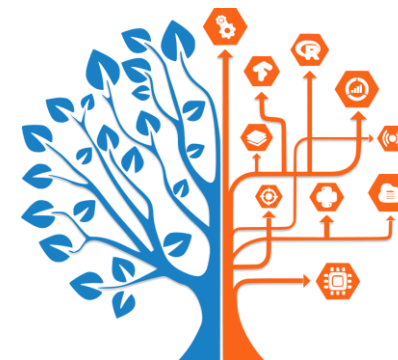
Zustandsüberwachung

- ✓ Soft Sensorik
- ✓ Predictive Maintenance
- ✓ Ausfallsvermeidung



- ✓ Hohe Reduktion von kostenintensiven Produktionsausfällen bei **BASF**

FRAUD DETECTION



Inhalt

- Betrugsfälle in der Geschichte, Gegenwart & Zukunft
- Use Cases aus der Praxis
- Strategie
- Methodik
- Erfolgsprüfung von KI-basierter Fraud Detection

Betrug oder Fehler?

- ✓ Groß angelegter Betrug, eigene Mitarbeiter beteiligt
- ✓ Betrug nur durch Lieferanten, unwissentliche Abnahme durch internen Mitarbeiter
- ✓ Unabsichtlicher Fehler des Lieferanten

Betrug hat viele Gesichter

Michelangelo fälschte antike Statue von Amor



Financial Fraud im antiken Griechenland



<https://divitaeblog.wordpress.com/2017/08/13/pioneers-of-fraud/>

https://www.online.uni-marburg.de/sds/secure/marburg/31-antike-mythologie/31_11_herrschaft/11_herrschaft_abb.html

Ein typisches Unternehmen verliert jedes Jahr 5 % seiner Einnahmen durch Betrugsfälle.

(globale Studie der Association of Certified Fraud Examiners - ACFE, 2018)

- Über Zeit und Regionen hinweg weisen die **Betrugsmuster** eine erstaunliche **Konsistenz** auf.
- Betrugsfälle, die in der Studie analysiert wurden, konnten im Durchschnitt (Median) erst nach 18 Monaten aufgedeckt werden.
- Trotz der konsistenten Betrugsmuster wurden die untersuchten Fälle am häufigsten anhand von konkreten Hinweisen entdeckt.
- Es fehlt an **systematischer Überprüfung** und durch **künstliche Intelligenz** gestützte Analysen.

Weltweit entstehen somit durch Betrug jedes Jahr Kosten in Höhe von rund 3,58 Billionen Euro.

Diese Summe entspricht dem BIP von Großbritannien und Italien zusammen

Was kann man dagegen tun?

1. Überprüfe nach dem Zufallsprinzip
2. Sachbearbeiter entscheidet nach Bauchgefühl was überprüft wird
3. Definiere Regeln zur Überprüfung
4. Nutze KI um verdächtige Machenschaften zu identifizieren

Betrug im Onlinehandel



Online Transaktionen

Kreditkarten Betrug

Identitätsdiebstahl

Friendly Fraud

Händlerbetrug

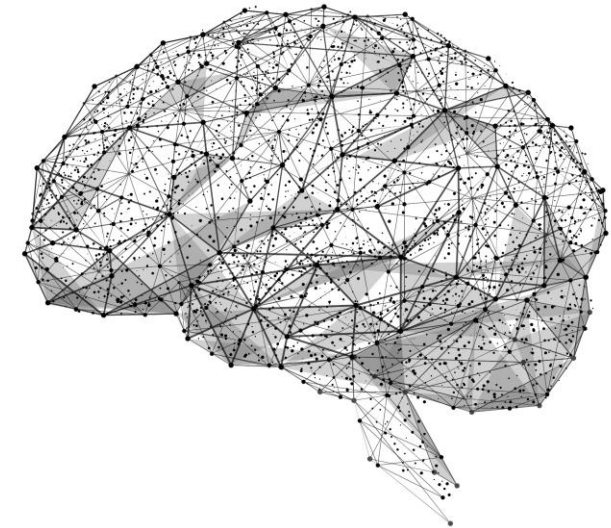
Clean Fraud

Betrug entwickelt sich ständig weiter...

... und so auch die Betrugsbekämpfung

Moderne Methoden zur Betrugserkennung nutzt:

- ✓ Big Data
- ✓ Netzwerkanalysen
- ✓ Künstliche Intelligenz



Unterscheidung Anforderung Fraud Detection

Credit Card Fraud Detection im **Bankwesen**

- Sofortreaktion notwendig (Entscheidung innerhalb von 8 Sekunden)
- Extrem hohes Datenvolumen
- Betrugsfälle fast zu 100% bekannt

Fraud Detection bei **Abrechnungsüberprüfungen**

- Längere Zeitspannen zur Überprüfung
- Datenvolumen und Datenqualität stark unternehmensabhängig
- Betrugsfälle meistens unbekannt
- Annahme: Es kann nicht jede Rechnung im Detail überprüft werden

Betrugsskandale bei der BASF

- Gerüstmontagen und Leitungsbau wurden falsch abgerechnet
- Diebstahl von Materialien
- Leistungen abgerechnet, die nie erbracht wurden
- Nach jedem Betrugsskandal neue Sicherheitsmechanismen



<https://www.n-tv.de/wirtschaft/Abrechnungsbetrug-schaedigt-BASF-article20280351.html>

<https://www.basf.com/>

<https://www.handelsblatt.com/unternehmen/industrie/chemieriese-basf-ermittelt-gegen-mitarbeiter-wegen-betrugsverdacht/20952692.html?ticket=ST-792255-QxtwcvXA7POdrEXIzgj-ap2>

Bayer entwickelt datengetriebene Betrugserkennung

Effizienz der Abrechnungsüberprüfungen

- Auf herkömmliche Weise: 5%
- Mit künstlicher Intelligenz: 25%



<https://www.bayer.com/de/>

<https://www.shutterstock.com/de/>

<https://theeriumpodcast.de/>
(16.02.2021, Walter Denk)

Business Case bei einem Kunden

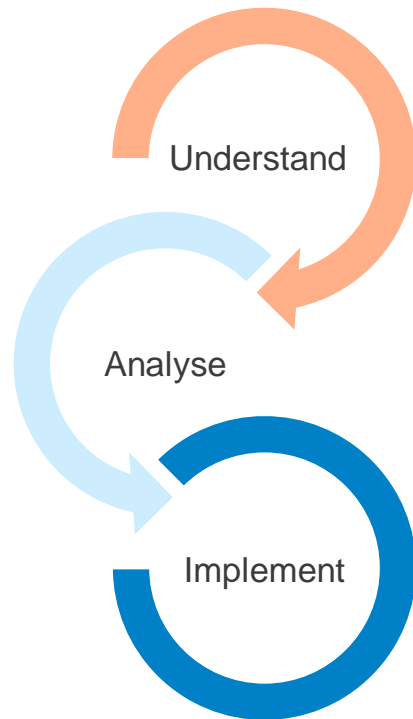
Effektivere Überprüfung von verdächtigen Rechnungen durch Fachpersonal

- Vor dem Einsatz unserer Software: 5-10%
- Durch die Unterstützung unserer Software: 30%

Betrugsprävention

- Abschreckungseffekt bei den Kontraktoren
- Abschreckungseffekt bei Mitarbeitern

Für uns steht die Umsetzung Ihres Gewinns im Zentrum



Verstehen der Abrechnungsprozesse

- ✓ Experten Know-how als wesentliche Komponente für erfolgreiche Modellierung
Welche Betrugsprävention gibt es bereits?
- ✓ Definition des Projekterfolgs auf Basis der wirtschaftlichen und strategischen Ziele des Kunden

Datenanalyse

- ✓ Explorative Datenanalyse und Anomaliedetektion
- ✓ Implementierung und Validierung von Expertenregeln

Integration ins Operativsystem

- ✓ Dashboards für Visualisierungen
- ✓ Integration von Ergebnissen in existierende Datenbanken
- ✓ Dockerisierung von Modellen für produktive Umgebungen

Welche Schritte wollen Sie setzen?

Wir begleiten den gesamten Prozess um aus Ihren Daten zu lernen und Betrug zu bekämpfen

Wo steht unser Unternehmen derzeit im Bereich der Betrugsbekämpfung?

Gibt es ungenutzte Daten, die bei der Betrugsbekämpfung helfen können?

Wo kann unternehmensspezifisches Wissen bei der Betrugsbekämpfung besser eingesetzt werden?

Nutzen wir KI um Betrugsfälle besser zu erkennen?

Unsere Strategie

- I. Datenquellen zusammenführen – Konsistenzprüfung
- II. Feature Engineering
- III. Anomaliedetektion
- IV. Implementieren von Expertenregeln und Firmenstrategien
- V. Ergebnis: Ranking nach Verdächtigkeit / Wichtigkeit der Überprüfung
- VI. Umsetzung der Ergebnisse und Arbeitsprozesse in BI Tools

I. Datenquellen zusammenführen

Tohuwabohu – *Am Anfang war das Chaos*

- Verfügbare Unternehmensdaten sind fast immer *unstrukturiert* oder *semi-strukturiert*
- unterschiedliche Datei- und Speicherformate je nach Unternehmensbereich, z.B.:
 - **Produktion** und **Logistik**: SAP, PDF
 - **IT-Abteilung**: relationale Datenbanken
 - **Buchhaltung**: SAP, Excel
 - **Management** und **Sales**: Word, Excel, PowerPoint, PDF
- “Grad der Strukturiertheit” hat großen Einfluss auf den Erfolg bzw. die Geschwindigkeit mit der Probleme datenbasiert gelöst werden können

Daten verschiedener Quellen werden zusammengeführt und auf Inkonsistenzen geprüft

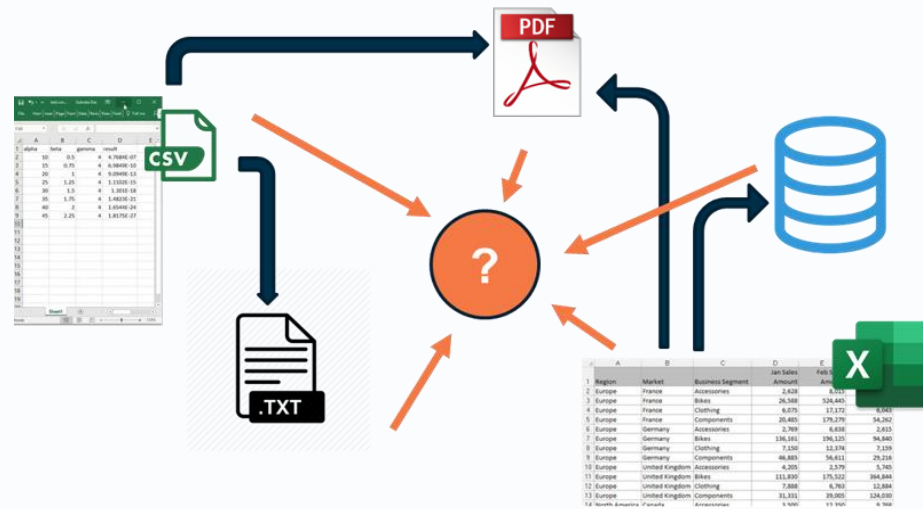
I. Datenquellen zusammenführen

Data Awareness

- Wie werden die gesammelten Daten gespeichert, und wie lang?
 - **Kurzfristig:** lokale Speicherung, einmalige Analysen
 - **Langfristig:** Datenbank/Data Warehouse, Data Lake
 - Speicherkosten
- Gibt es allgemein zugängliche Dokumentation zu Speicherorten oder Variablen- und File-Benennung?
 - Finden der benötigten Daten
- Kursieren mehrere Versionen der Daten im Unternehmen?
 - Datenkonsistenz
- Wie werden Zugriffsrechte vergeben?
 - Datenbanken erlauben viel besseres Management der Rechte verschiedener User
- Datenbanken sind erheblich *sicherer* als z.B. Netzwerklaufwerke oder lose Files auf öffentlich zugängliche PCs



I. Datenquellen zusammenführen Datenmanagement



- **Wie** gespeichert wird beeinflusst möglichen Mehrwert der gesammelten Daten

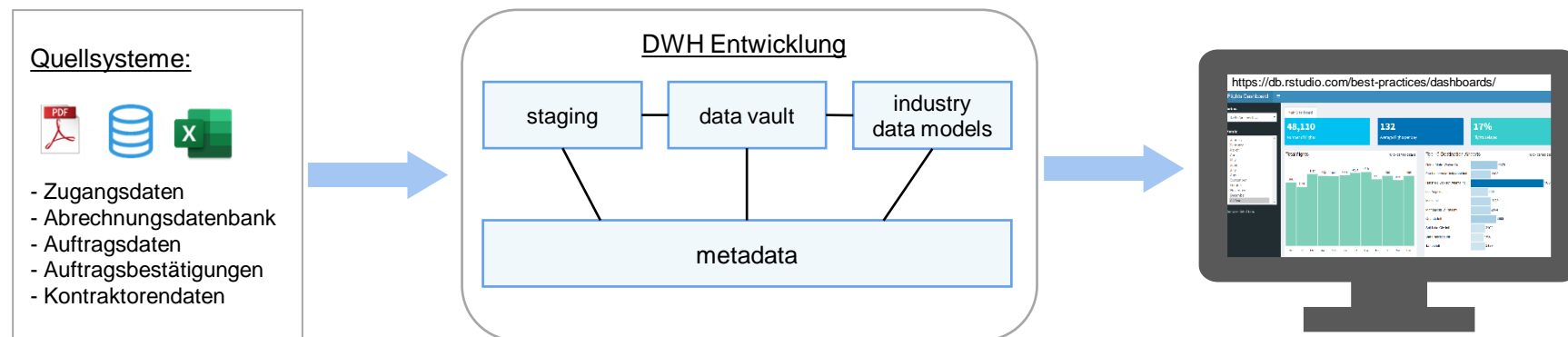
- Wissen **wo** welche Daten abgespeichert werden ist oft auf kleinen Personenkreis beschränkt/verteilt

- *Data Literacy* =
Die vorhandenen Daten und ihre Grenzen kennen.

I. Datenquellen zusammenführen Data Warehouse (DWH)

= eine Plattform, um...

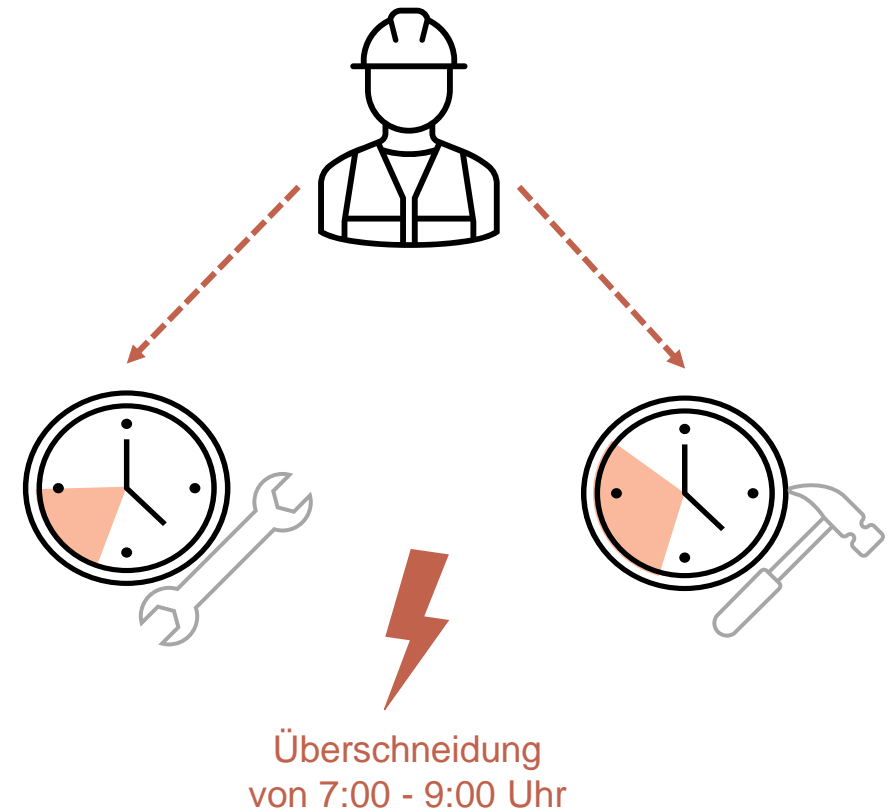
- Unterschiedliche Datenquellen zusammenzuführen
- Daten strukturiert abzulegen
- Langfristige Verfügbarkeit zu sichern
- Versorgung für datengetriebene Entscheidungsprozesse / Überwachungen sicherzustellen



I. Datenquellen zusammenführen

Auffälligkeiten durch Dateninkonsistenz

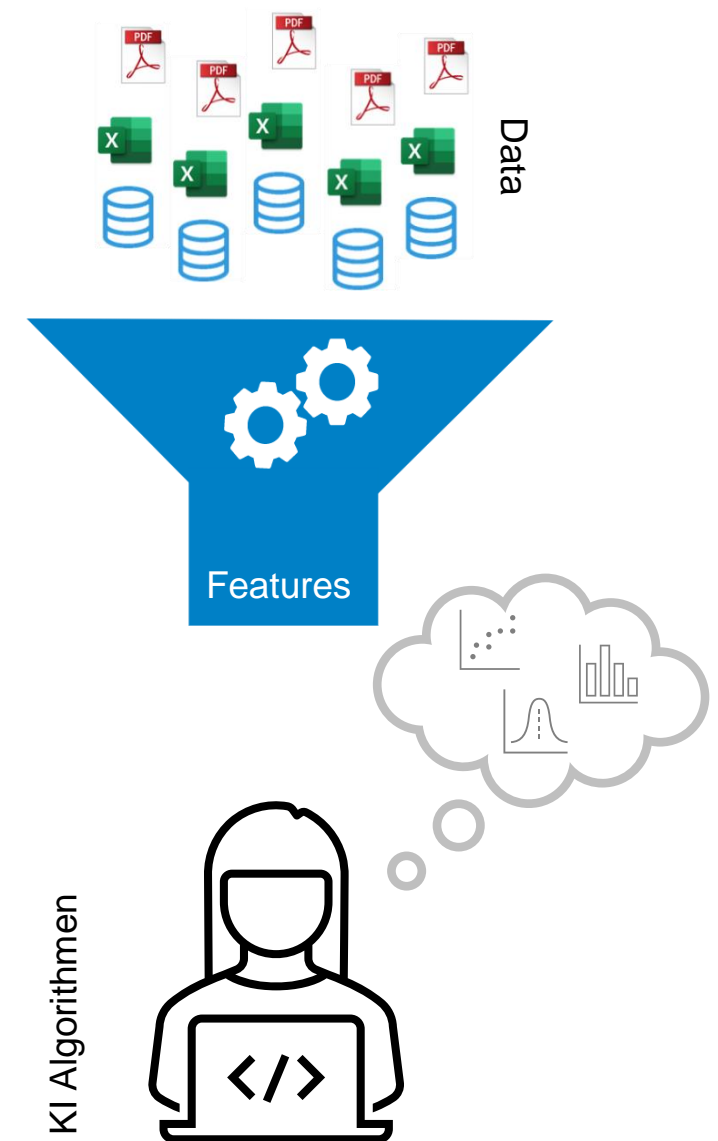
- Wenn Daten strukturiert und einheitlich abgelegt werden können manche Auffälligkeiten bereits **vor** der Verwendung eines KI Algorithmus erkannt werden
- Beispiel Lohnstundenverrechnung:
 - Lehrlinge mit hohen Stundenlöhnen
 - Lohnstunden werden gestückelt eingereicht
→ Pausenzeiten nicht eingehalten
→ Überschreitung der max. 12h/Tag
 - Kontraktorenmitarbeiter arbeiten für 2 Firmen gleichzeitig



II. Feature Engineering

Von Rohdaten zu verwertbaren Features

- Rohdaten sind für Machine Learning Modelle oft nicht geeignet
- Sinnvolle Features die relevante Informationen beinhalten müssen generiert werden
- Beispiele:
 - Verbrauch an Ressourcen/Arbeitsstunden eines Kontraktors für eine Tätigkeit
 - Abstand zwischen Wartungen
 - Zeitabstände zwischen Rechnungslegungen
 - Gesamtvolumina von Rechnungen
 - Textfeatures aus Freitextfeldern (Datumsangaben, Schlagwörter)



III. Anomaliedetektion

Erkennung von Ausreißern (1D)

Ausreißer = “Wert der aus der Reihe tanzt”

VARIABLENSTRUKTUR:

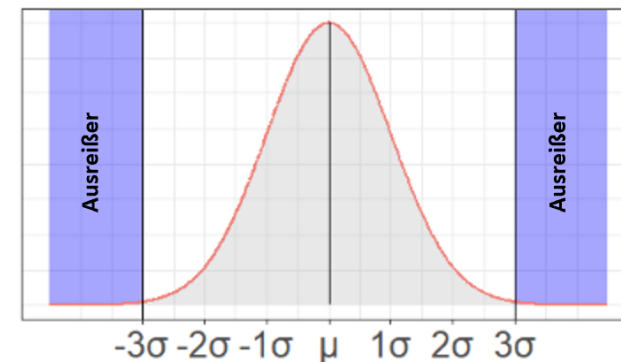
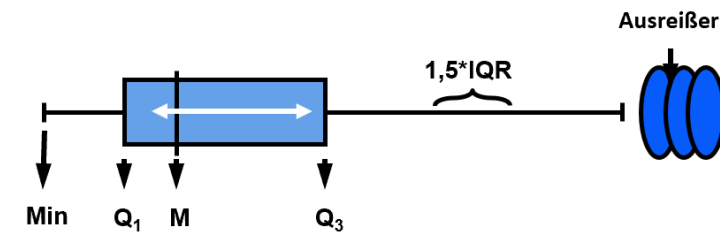
- z.B.: In einer Spalte mit dem Alter von Kunden sind Werte < 18 oder > 100 sehr sicher Ausreißer

AUS BOXPLOT:

- *Ausreißer:*
Beobachtungen außerhalb der 1.5-fachen Interquartils-Spannweite

“X-SIGMA”-REGEL:

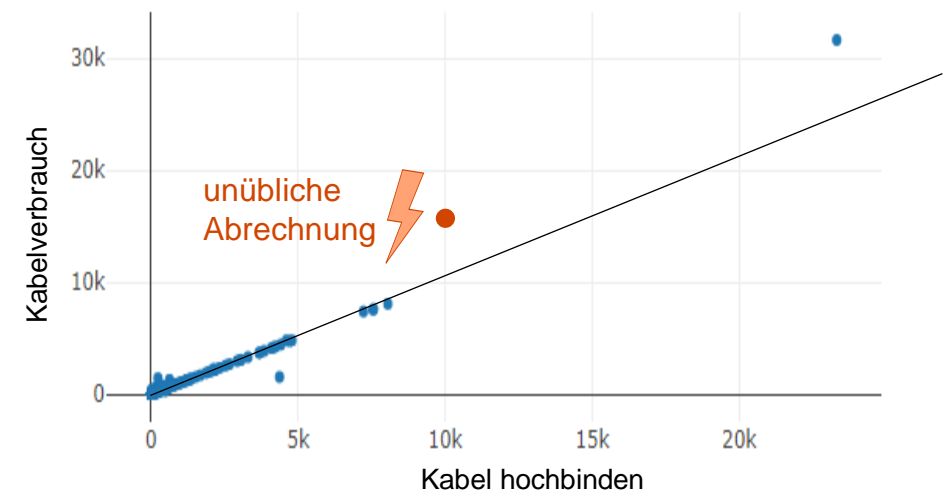
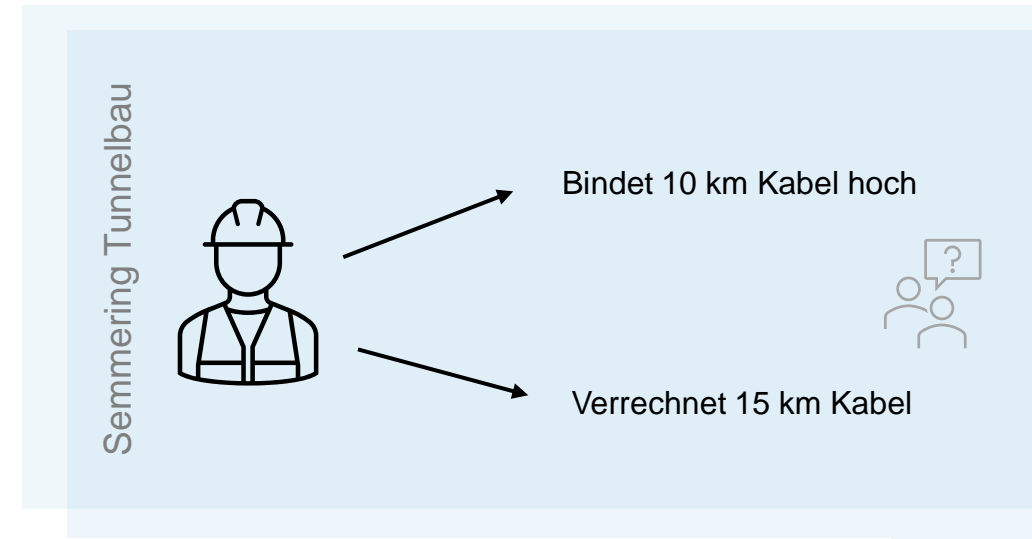
- *Ausreißer:*
Mehr als X Std-Abweichungen vom Mittelwert entfernte Beobachtungen



III. Anomaliedetektion

Ausreißer in statistischen Modellen

- Übliche Mengenkombinationen durch robuste Modelle Zusammenhang bringen
- **Achtung:** Ausreißer können auch schon in den zugrunde liegenden Daten vorkommen
 - Viele Modelle sind anfällig für Ausreißer
 - Ergebnisse werden verzerrt
 - Anwendung robuster Modellierungsmethoden



III. Anomaliedetektion

Achtung: nicht jede Anomalie muss ein Betrug sein

Beispiel I:

Ein Techniker wird regelmäßig mit IT Wartung beauftragt (zum Teil remote). Üblicherweise verbringt er 75% der verrechneten Zeit vor Ort. Plötzlich steigt der Anteil an Remote Arbeit stark an und es wird deutlich mehr abgerechnet.

→ viele mögliche Erklärungen:

- Spezielles Problem
- Coronabedingte Remotearbeit
- ...

Beispiel II:

Die üblichen Einnahmen beim Verkauf von Schreibtischlampen schwanken um die 90 T€/Monat, im Juli liegen sie allerdings nur bei 88,13 €

→ Ursache ist rasch gefunden:

- Dezimaltrenner “” wurde beim Erstellen als Komma “,” interpretiert (Encoding Hell)

Zusätzliche Prüfung durch Experten ist unerlässlich

IV. Expertenregeln und Firmenstrategien

Betrugsbekämpfung und Betrugsprävention

- Häufig werden zu kontrollierende Rechnungen zufällig ausgewählt
- Firmenstrategien in die Auswahl inkludieren:
 - Überprüfe alle Rechnungen ab einer bestimmten Höhe
 - Überprüfe jeden Kontraktor mindestens einmal im Jahr
 - Überprüfe jeden Mitarbeiter auf Korrektheit seiner Freigaben (heikel!)
- Erwünschte Resultate:
 - Maximiere die zurückgeforderten Beträge
 - Maximiere die Anzahl an Rechnungskorrekturen
 - Minimiere mutwillige Betrugsfälle

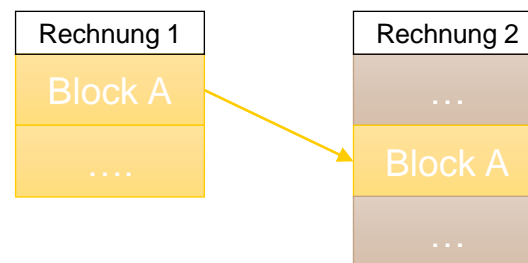
IV. Expertenregeln und Firmenstrategien

Integration des unternehmensspezifischen Know-Hows

Was könnten sinnvolle Regelwerke sein, um Betrug aufzudecken?

Allgemeine Expertenregeln:

- Rechnungsprüfer haben Erfahrung damit wie Betrugsfälle häufig zu erkennen sind
- Beispiel einer doppelten Verrechnung:



Themenspezifische Expertenregeln:

- Regelwerke für spezifische Themen
- Beispiel Semmering Basistunnel:

Wenn mehr als xx Liter Benzin pro Meter Tunnelbau verbraucht werden, ist das auffällig.

IV. Expertenregeln und Firmenstrategien

Kombiniere Expertenwissen und Machine Learning

Expertenregel:

Wenn mehr als xx Liter Benzin pro Meter Tunnelbau verbraucht werden, ist das auffällig.

- Auch bei schlechter Datengrundlage möglich
- Subjektive Sicht
- Aus Erfahrung und Prozesswissen entstanden

Anomaliedetektion:

Wieviel Liter Benzin werden pro Meter Tunnelbau verbraucht?

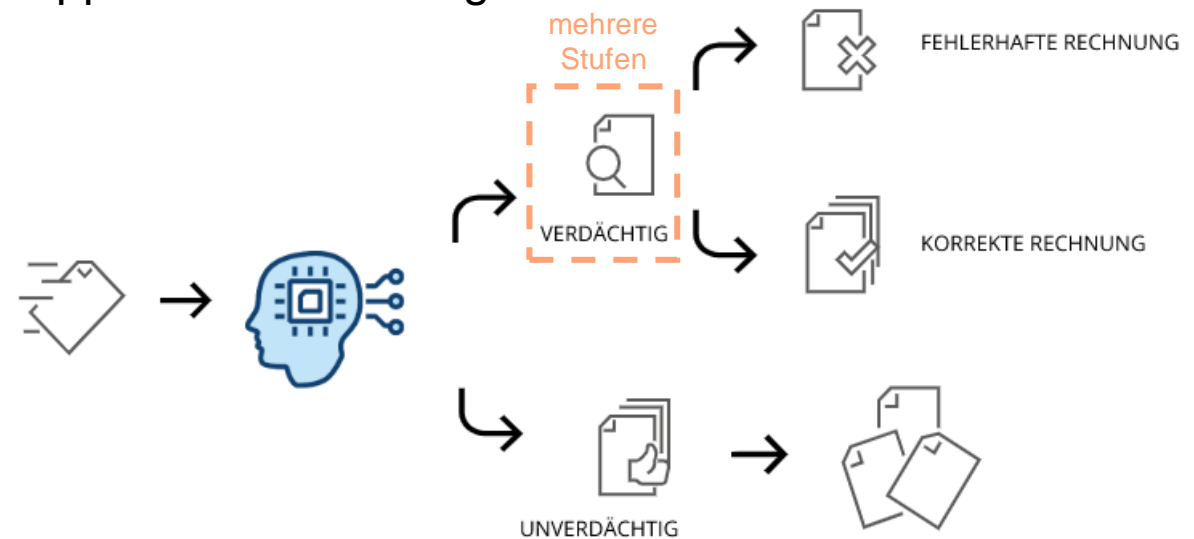
- Gute Datengrundlage notwendig
- Objektiv
- Anhand historischer Daten nachvollziehbar



V. Ranking

Sortierung nach Stufen der Verdächtigkeit

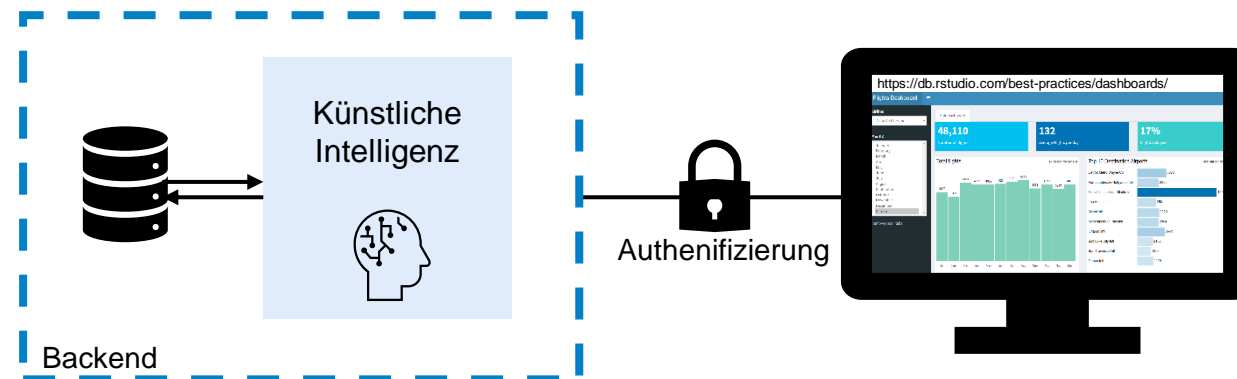
- Verdächtige Rechnungen können anhand von verschiedenen Kriterien nach Ihrer „Wichtigkeit“ geordnet werden
 - Gesamtbetrag der Rechnung
 - Betrag der Auffälligkeit
 - Ausmaß der Abweichung vom Üblichen
 - Grad der Überschneidung bei doppelter Verrechnung
 - Bestimmte Kontraktoren



VI. Umsetzung in BI Tool „From data to business value“

Implementieren einer Benutzeroberfläche macht Algorithmen für den User greifbar

- Bereitstellen der Ergebnisse in einem Dashboard
- Auffälligkeiten der KI Algorithmen werden angezeigt – **Benutzer entscheidet**
- Regelmäßig neu berechnete Auffälligkeiten werden direkt ins Dashboard übernommen
- Sowohl Python als auch R bieten Dashboardlösungen in Form von ‚Shiny‘ und ‚Dash‘ an



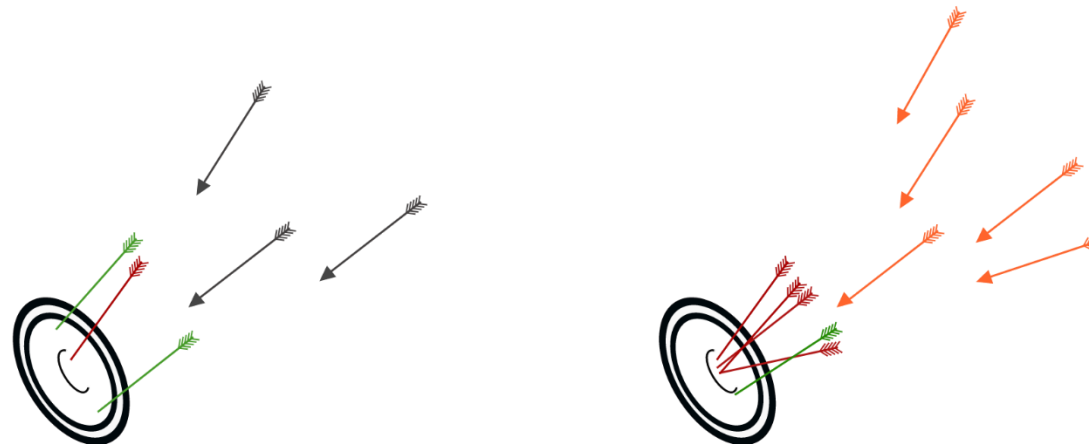
VI. Umsetzung in BI Tool

Vorteile einer Applikation

- ✓ Hosting auf internem Server oder Cloud
→ an verfügbare Infrastruktur anpassbar
- ✓ Zugriff über Web Browser
→ Keine Installation von Software auf Userseite notwendig
- ✓ Darstellung an Wünsche und Bedürfnisse der Benutzer anpassbar
→ Daten für den User optimal aufbereitet
- ✓ Strategische Prüfungen können leicht umgesetzt werden
→ Prüfungsprozess kann innerhalb der Applikation stattfinden
- ✓ Arbeitsaufwand für Prüfer pro Rechnung wird geringer

Wie prüfen wir unseren Erfolg?

- Algorithmen werden im Vorfeld an historischen Daten/Betrugsfällen getestet
- Durch Verwendung der KI erhöht sich die Wahrscheinlichkeit fehlerhafte Rechnungen zu erkennen was einer Steigerung der Effizienz bei der Überprüfung der Rechnungen gleichkommt
- KPIs einer erfolgreichen Umsetzung:
 - ✓ Zahl der geprüften Rechnungen steigt
 - ✓ Prozentsatz an Rückforderungen steigt
 - ✓ Vorallem beim langfristigen Kontraktoren tritt Abschreckungseffekt für mutwilligen Betrug ein



Betrug entwickelt sich ständig weiter...

... und so sollte es auch die Betrugsbekämpfung tun

Tools und Algorithmen müssen **anpassbar** und **erweiterbar** sein

- ✓ Neue Expertenregeln können schnell und einfach eingepflegt werden
- ✓ Regelmäßiges erheben und einbinden neuer Daten
- ✓ Modell Retraining

Quellen

https://www.parlament.gv.at/PAKT/VHG/XXVII/J/J_04087/fnameorig_847120.html

<https://steiermark.orf.at/stories/3083317/>

<https://www.process.vogel.de/betrug-in-millionenhoehe-falsche-fremdfirmen-rechnungen-erschuettern-basf-a-685991/>

<https://www.n-tv.de/wirtschaft/Abrechnungsbetrug-schaedigt-BASF-article20280351.html>

<https://www.handelsblatt.com/unternehmen/industrie/chemieriese-basf-ermittelt-gegen-mitarbeiter-wegen-betrugsverdacht/20952692.html?ticket=ST-792255-QxtwcwXA7POdrEXIzgjj-ap2>

Podcast <https://theeriumpodcast.de/> mit Walter Denk

Data Literacy Webtalks, DI Stefan Salzl, DI Andreas Blatt

Bildquellen - Logos:

<https://go.kount.com/>

<https://www.forter.com/>

<https://www.riskified.com/>

<https://www.signifyd.com/>

<https://www.clear.sale/>

<https://www.ravelin.com/>

<https://feedzai.com/>

<https://www.shutterstock.com/de/>

Vielen Dank für die
Aufmerksamkeit!



appliedstatistics
data driven solutions

Kontakt



Dr. Irene Ortner

E-Mail: irene.ortner@applied-statistics.at



Dr. Natalie Vollert

E-Mail: natalie.vollert@applied-statistics.at

WIEN, 25.03.2021

