

Unsere Data Science Journey

Integration von Machine Learning Modellen in den Leistungsprozess



- 1. Einführung**
- 1. E2E-Prozess**
- 2. Fachliche Anforderungen und Datenhierarchie**
- 3. Modellentwicklung**
- 4. Testing und Monitoring**
- 5. Take Aways**

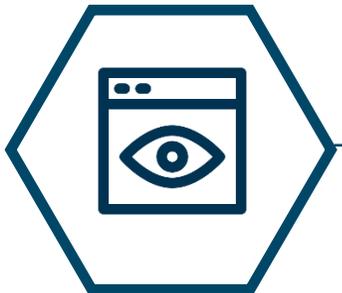


Hallesche Krankenversicherung



Platz 12 der PKV-Unternehmen

1.661 Mio. Euro Beitragseinnahmen (2024)
1.233 Mitarbeiterinnen und Mitarbeiter (2024)



Kerngeschäftsfelder

Vollversicherung für Arbeitnehmer
und Selbständige
bKV – betriebl. Krankenversicherung



Versichertenbestand

916.310 Gesamtversichertenbestand (2024)
225.190 Vollversicherte (2024)



Ihre Referenten



Jürgen Hoff

Gruppenleiter
Leistungscontrolling
juergen.hoff@hallesche.de

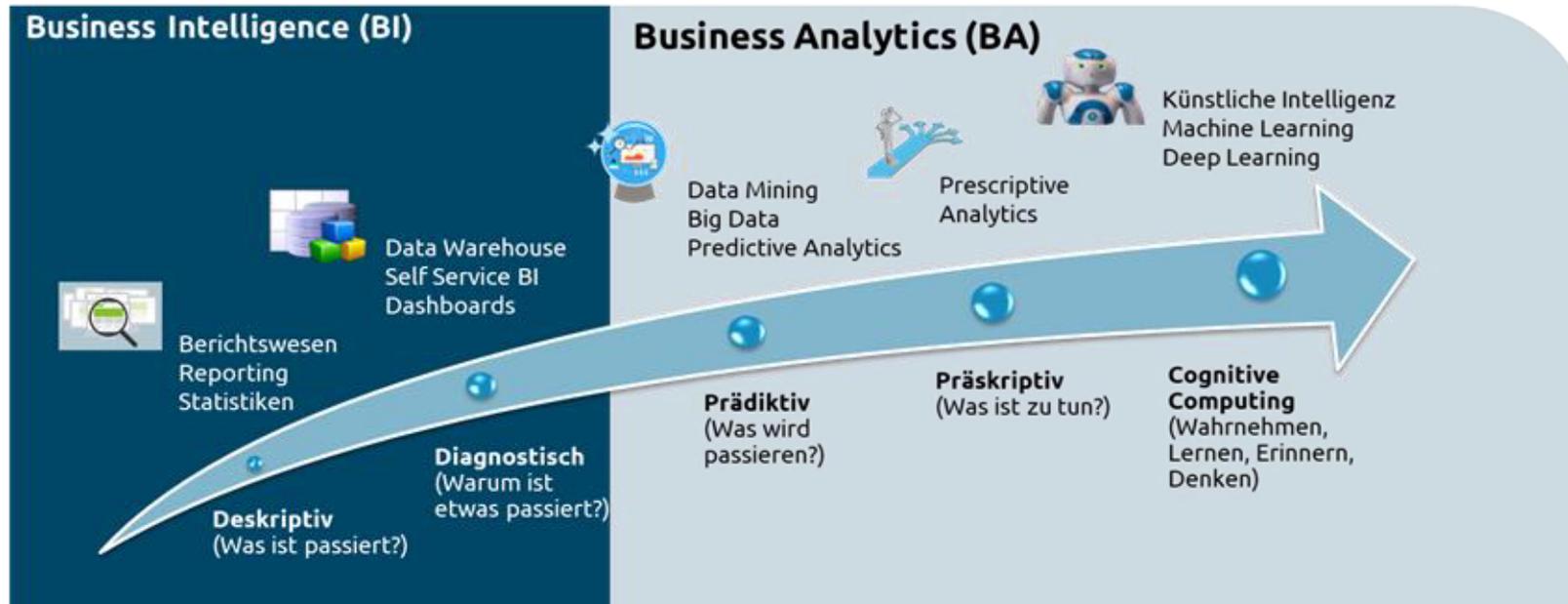


Harald Papp

Senior Data Scientist
Leistungscontrolling
harald-mircea.papp@hallesche.de



Evolution von Business Intelligence zu Business Analytics



BI

- 2015-2018: BI-Projekt – Aufbau DWH
- seit 2015: Erweiterung mit jedem Release

BA

- 2016-2020: Erste Piloten + Große Vorstudie
- 2022-2024: Projekt zur Einführung BA



Einführung von Data Science im Leistungsprozess der Hallesche

Ziele und Inhalte

Projekt aus Fachbereich initiiert und umgesetzt - Domänenwissen

- Großes interdisziplinäres Projektteam aus 6 Bereichen inkl. externer Support

Strategische Ziele

- Aufbau Infrastruktur und Personalressourcen
- Knowhow-Entwicklung Data Scientists und Data Analysts

Operative Ziele

- Umsetzung von 4 Use-Cases im Arzneimittel- und Arztbehandlungs-Umfeld
- Konzentration der Sachbearbeiter auf die prüfenswerten Fälle

Inhalte

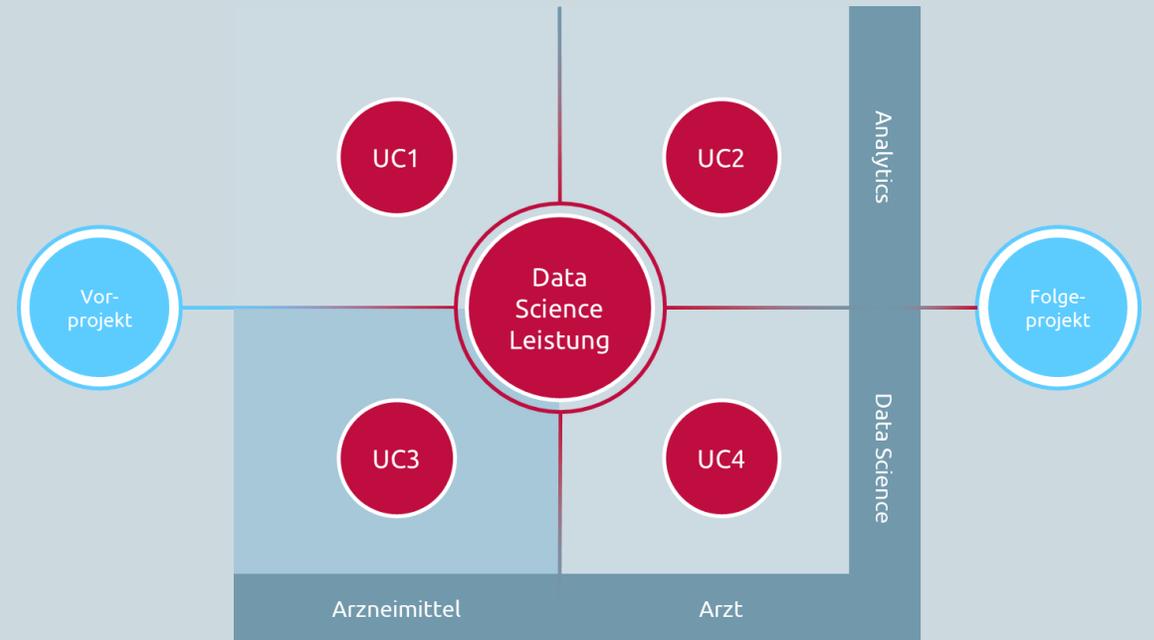
- **Use Case I & II:**
Optimierung der bestehenden deterministischer Arzneimittel- und Arztbehandlungsprüfsysteme mittel analytischer KI-Methoden
- **Use Case III & IV:**
Erweiterung des operativen Leistungssystems durch Arzneimittel- und Arztbehandlungs-KI

Gesamtziele

- Optimierung Prozesse und Reduzierung Sachbearbeitungsaufwand
- Erhöhung Automatisierung und Steigerung Einsparquoten

Nutzen der Use-Cases übersteigt Investitions- & Betriebskosten

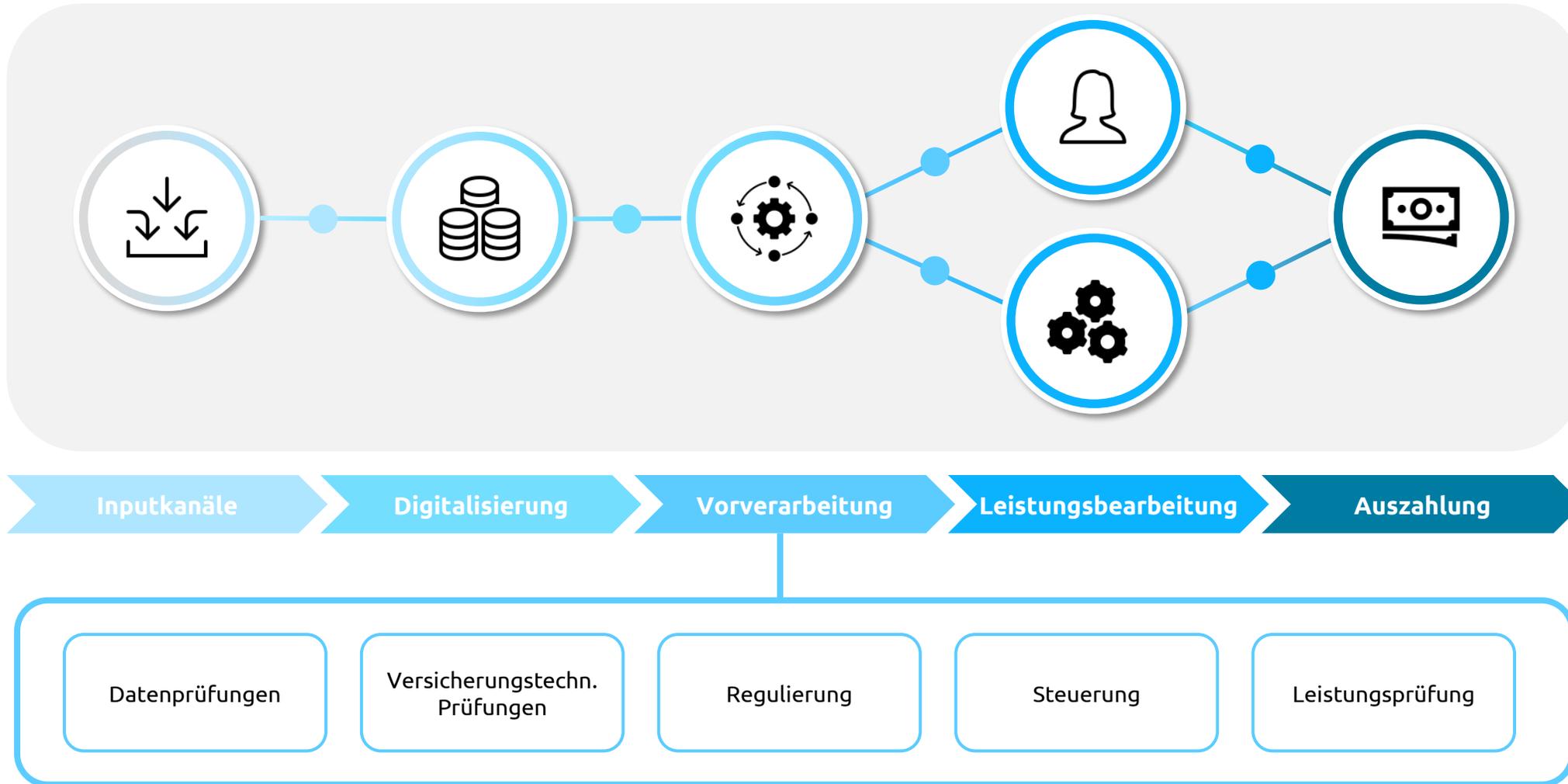
Der Beginn der Data Science Journey



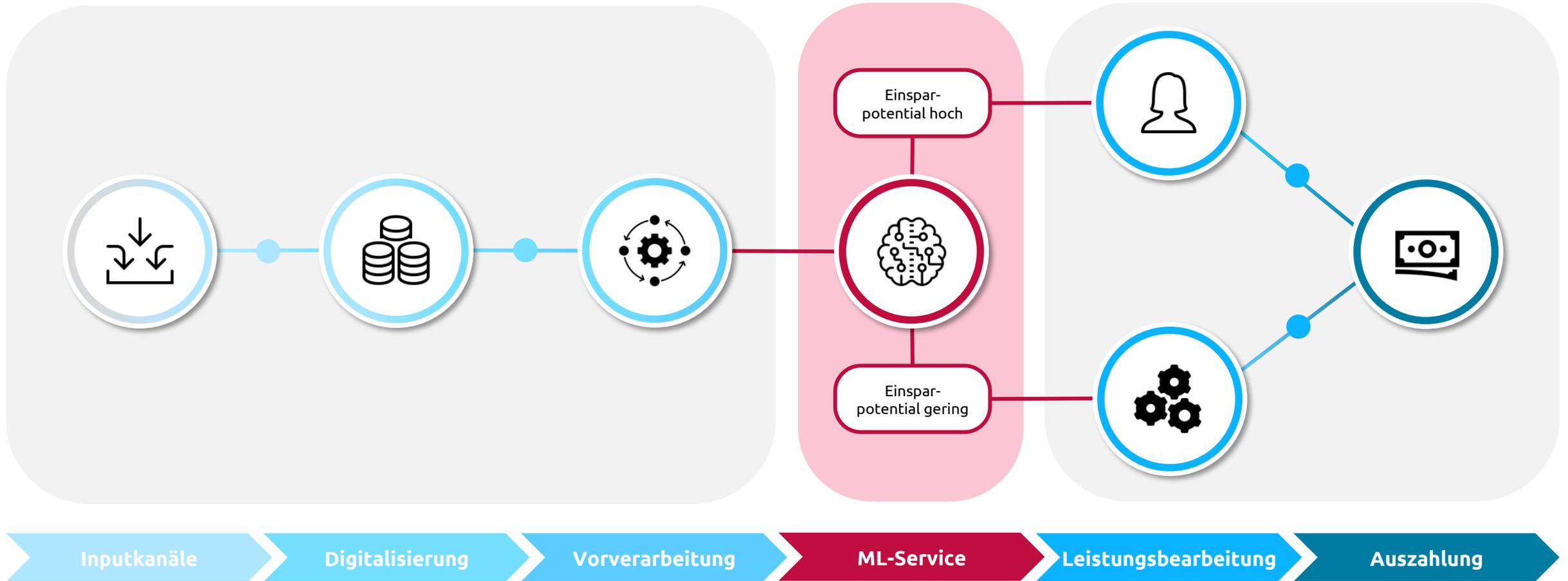
E2E-Prozess



Vom Eingang bis zu Auszahlung wandert ein Beleg durch viele Prozesse ...



... und zukünftig kommen durch den Einsatz von ML weitere Prozessschritte hinzu



Fachliche Anforderungen und Datenhierarchie



Fachanforderungen müssen erhoben und umgesetzt werden



Die Leistungsprüfung wird über Prüfregele gesteuert

Prüfregele unterscheiden sich in **Korrekturregele** und **Hinweisregle**

Klassifizierung von Prüfregele für ML-Service

Insgesamt werden 21 Regle als **ML-fähig** eingestuft

Zusammen mit den fachlichen Anforderungen reduzieren die nicht ML-fähigen Prüfregele die **Gesamtmenge der betrachteten Belege** stark

Prüfregele unterliegen **stetigen Veränderungen**, die beachtet werden müssen

Beleg

Übersicht | Diagnosen | Rechnungssicht | Leistungssicht

Positionen

1.

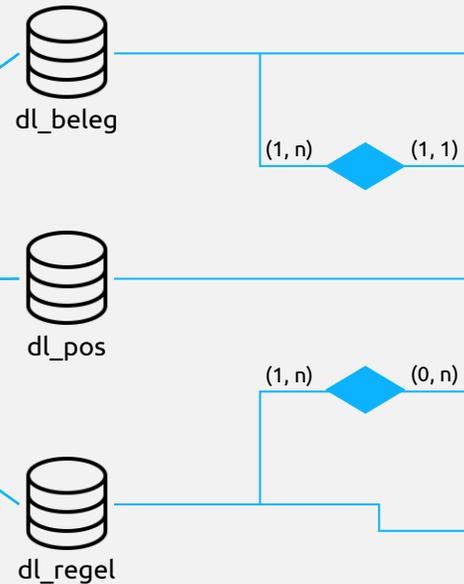
Ausgabe	PZN	Bez.	Anzahl	RBT	Korrektur
25.04.24	070145	Foltran	1	7,99	7,99

Korrektur:

100429 Präparat darf ohne Nachweis nicht erstattet werden.



Fachliche Sicht muss in eine Datensicht überführt werden



Beleg

Übersicht | Diagnosen | **Rechnungssicht** | Leistungssicht

Positionen

1.

Ausgabe	PZN	Bez.	Anzahl	RBT	Korrektur
25.04.24	070145	Foltran	1	7,99	7,99

Korrektur:
100429 Präparat darf ohne Nachweis nicht erstattet werden.



Der finale Datensatz weist eine hierarchische Struktur auf

Beleg	Position	Posten	Regel	VNR	Timestamp	Betrag	PZN	[...]
97124640	1	1	1000365	278680	2023-01-02 11:22:50	7.54	1234556	[...]
97124640	3	1	100037	278680	2023-01-02 11:22:50	4.25	7730256	[...]
97446851	1	1	100101	285941	2023-01-11 07:59:30	20.50	9999175	[...]
97176436	1	1	100084	147204	2023-01-03 11:25:38	73.95	14352754	[...]
97176436	1	1	100989	147204	2023-01-03 11:25:38	73.95	14352754	[...]
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]

Es werden knapp 100 Features betrachtet!



Modellentwicklung



Herausforderungen in der ML-Modellierung

Auf welcher hierarchischen Daten-Ebene schätze ich?

Wie aggregiere ich die Ergebnisse?

Custom Metrics vs. Standard Metrics?

Wie und wann wird retrained?

Data Drift?

Wie vermeide ich Data Leakage?

Concept Drift?

Wie gehe ich mit fehlenden Daten um?

Welche Modelle eignen sich für meine Daten?

Welchen Datenstand nehme ich?



Herausforderungen in der ML-Modellierung - Input

1

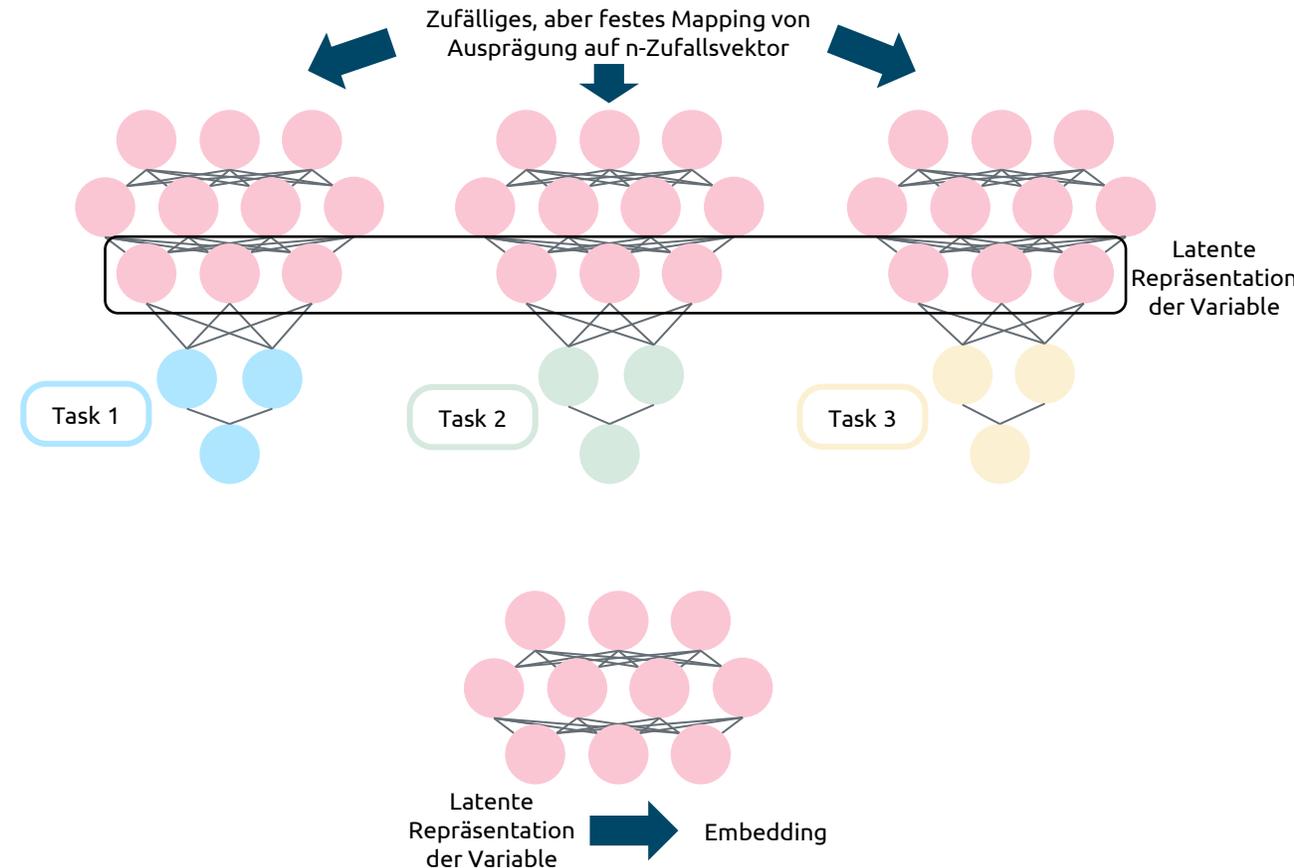
Kategorielle Variablen haben viele Ausprägungen

Problem

- One Hot Encoding führt zu vielen Zusatzdimensionen
- Curse of Dimensionality

Lösung

- Inspiriert von LLMs (aber eigentlich bewährte Methode): Embeddings
- Kategorielle Variable wird mittels Neuronalem Netzwerk in einen n-Vektorraum eingebettet, sodass semantische Zusammenhänge erhalten bleiben



2

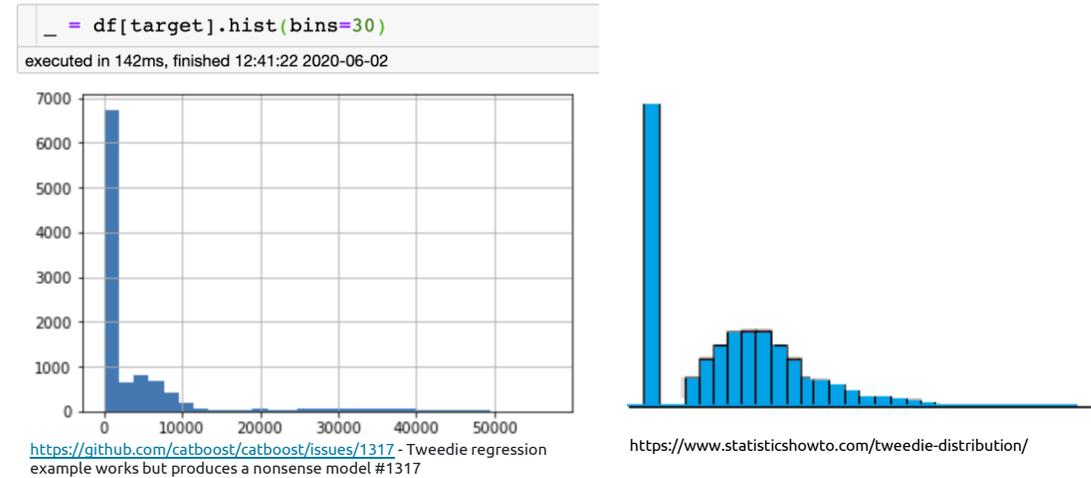
(bimodale) Tweedie-Verteilung

Problem

- Zielvariable weist (für Versicherungsdaten übliche) Tweedie Verteilung auf
- Verhalten für Datenpunkte bei $y = 0$ inhärent anders, als für Datenpunkte bei $y > 0$

Lösung

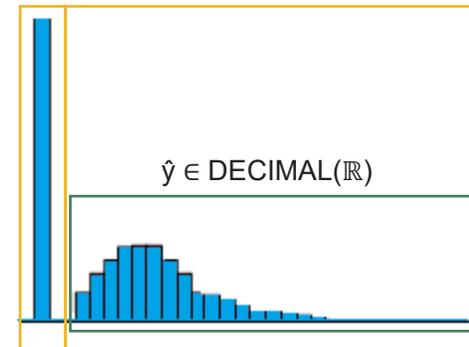
- Tweedie Regression? -> Nein!
- Zwei Modelle „stacken“: Binärer Klassifikator entscheidet, ob 0 oder nicht-0. Für nicht-0 Vorkommnisse anschließend Regression.



Klassifikation



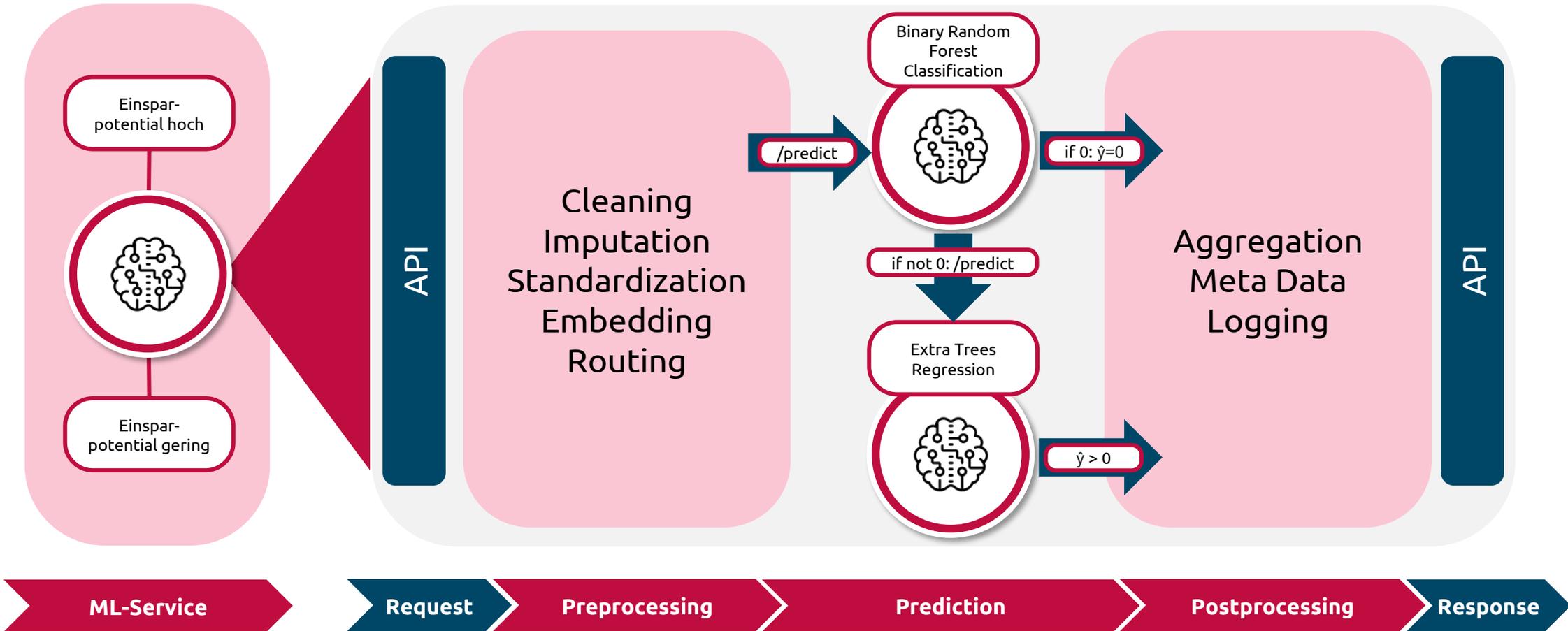
0 1



Regression



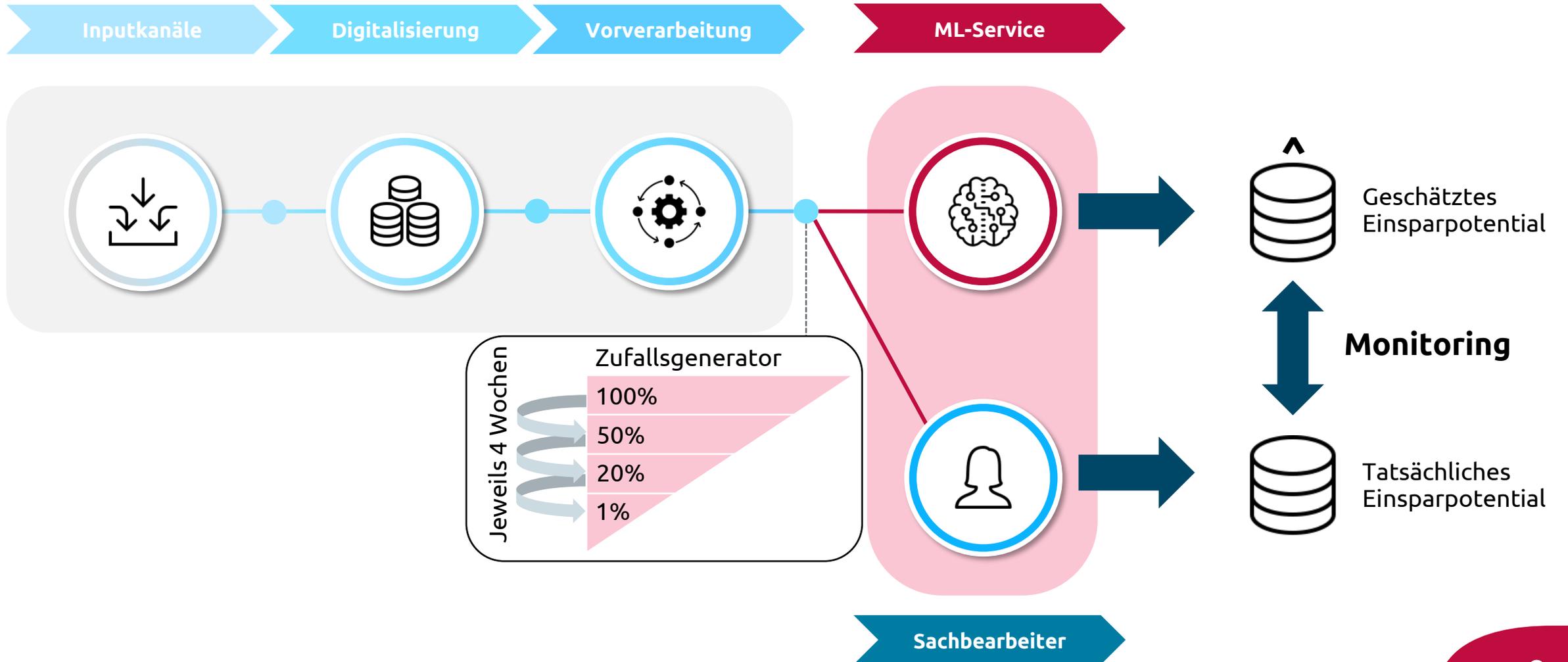
ML-Prozess mit Stacked Models



Testing und Monitoring



Monitoring als Qualitätssicherung/ Inspiration



Der Weg nach Prod ist lang, aber notwendig

1

Schätzung auf „unterster“ Ebene

Binary Random Forest Classification

Extra Trees Regression



Binäre Metriken, besonders **Recall**

Regressions-Metriken -> **RMSE**

Generelle Modellgüte + Modellauswahl

2

Schätzung Beleg-Ebene

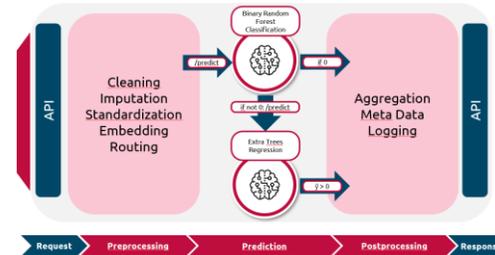


Binäre Metriken anhand Entscheidungsschwellwert + Benchmark vs. Dummy-Modell

Belegübergreifende Über-/Unterschätzung

3

Schätzung mit Echtweltdaten

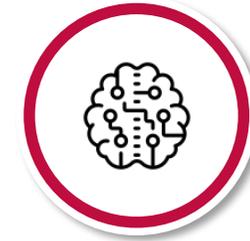


Ausgewiesene Kennzahlen: **Monetärer Fehler, Mittlerer Fehler, Anteil falscher Entscheidungen**

Abweichungen Train-Daten (Vorverarbeitung) und Livedaten

4

Benchmark mit Sachbearbeitern

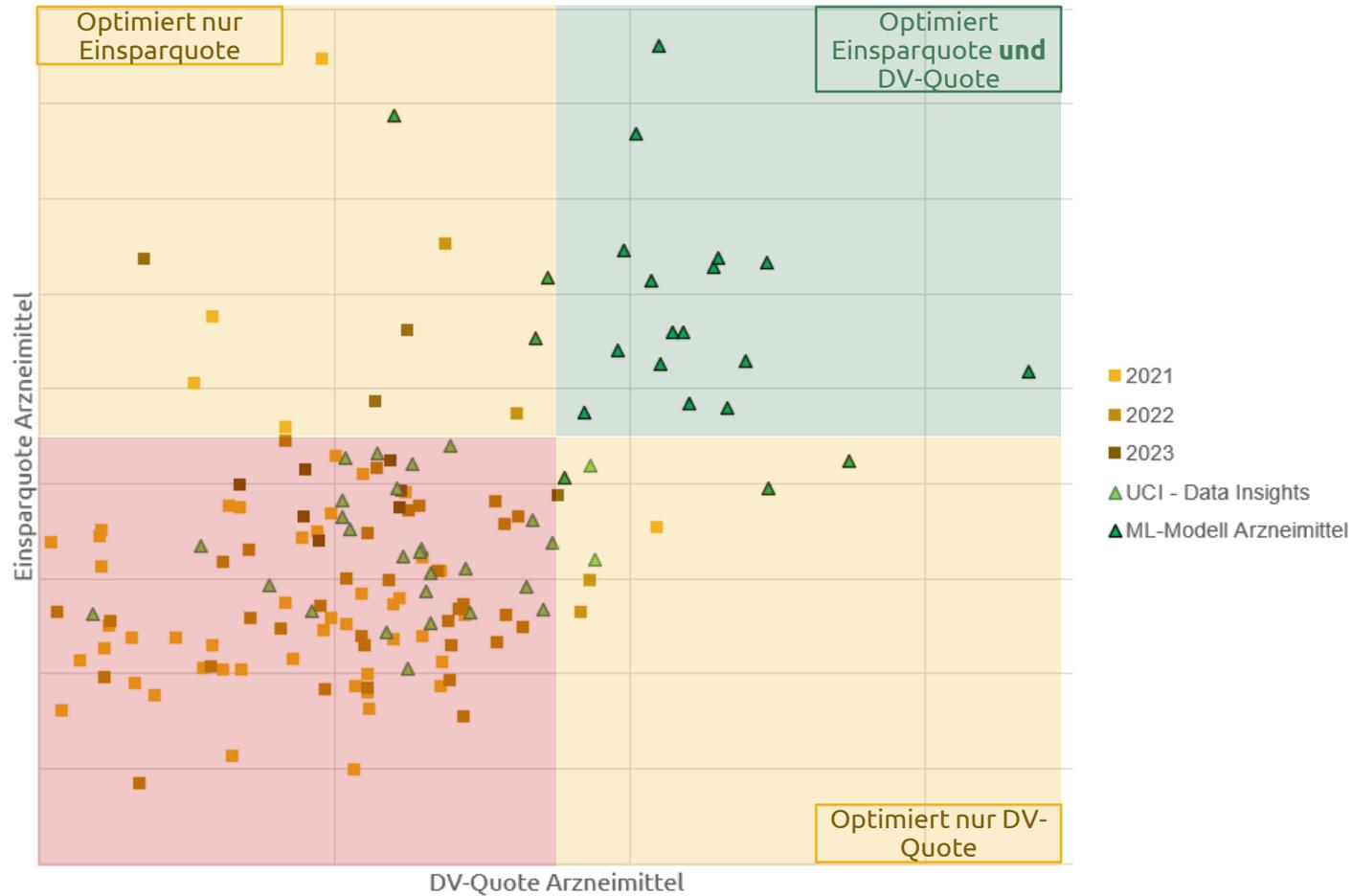


Ausgewiesene Kennzahlen: **Monetärer Fehler, Mittlerer Fehler, Anteil falscher Entscheidungen**

Data- und Concept-Drift + Unknown unknowns



Einsparquote und DV-Quote pro KW Arzneimittel



Take Aways



Take Aways

Gute Daten sind Daten, die schonmal benutzt wurden

Feature Engineering bietet die Möglichkeit Fachlichkeit miteinfließen zu lassen

Model Selection spielt bei Echtwelt-Daten nur eine untergeordnete Rolle

Der ML-Service steht nicht alleine da – Angrenzende Prozesse müssen mitgedacht werden

Akribisches Monitoring und Testing retten euch vor Stress in Prod



Rechtliche Hinweise

Gerne überlassen wir Ihnen diese Präsentation zu Informationszwecken. Bitte beachten Sie aber, dass die darin enthaltenen Informationen allgemeiner Natur sind und eine Beratung im konkreten Einzelfall nicht ersetzen können.

Diese Unterlage haben wir nach bestem Wissen erstellt und die Inhalte sorgfältig erarbeitet. Gleichwohl kann man Fehler nie ganz ausschließen. Bitte haben Sie deshalb Verständnis dafür, dass wir keine Garantie und Haftung für die Aktualität, Richtigkeit und Vollständigkeit übernehmen. Infolgedessen haften wir nicht für direkte, indirekte, zufällige oder besondere Schäden, die Ihnen oder Dritten entstehen. Der Haftungsausschluss gilt nicht für vorsätzliches oder grob fahrlässiges Handeln oder bei Nichtvorhandensein zugesicherter Eigenschaften.

In die Zukunft gerichtete Aussagen sind naturgemäß mit Ungewissheiten verbunden. Deshalb können die tatsächlichen Ergebnisse von diesen abweichen. Eine Verpflichtung zur Aktualisierung von Zukunftsaussagen wird nicht übernommen.

Bei Kapitalanlage-Produkten gilt zusätzlich: Die Präsentation stellt keine Anlageberatung dar und sollte auch nicht als Grundlage für eine Anlageentscheidung dienen. Aus den gegebenenfalls dargestellten Wertentwicklungen der Vergangenheit können keine Rückschlüsse auf zukünftige Wertsteigerungen gezogen werden.

Unsere Marken und Logos sind international markenrechtlich geschützt. Es ist nicht gestattet, diese Marken und Logos ohne unsere vorherige schriftliche Zustimmung zu nutzen.

Inhalt, Darstellung und Struktur dieser Unterlage sind urheberrechtlich geschützt und eine Nutzung, Verwendung, Reproduktion oder Weitergabe an Dritte – ganz oder teilweise – ist nur mit unserer ausdrücklichen vorherigen schriftlichen Zustimmung zulässig. Alle Rechte sind vorbehalten.

© ALH Gruppe

