



Institut für Finanz- und  
Aktuarwissenschaften

ifa



DAV/DGVFM  
Jahrestagung

2022

# Data-Analytics-Verfahren in der PKV

Use Case Stornoprävention



DGVFM  
DEUTSCHE GESELLSCHAFT  
FÜR VERSICHERUNGS- UND  
FINANZMATHEMATIK e.V.



DAV  
DEUTSCHE  
AKTUARVEREINIGUNG e.V.

Kinga Böhm (Hallesche Krankenversicherung a.G.) und Dr. Johannes Schupp (ifa)

DAV/DGVFM-Jahrestagung

Bonn, April 2022

# Agenda

**Einleitung**

**Daten**

**Data-Analytics-Verfahren zur Stornoprognose**

**Modelltuning und Optimierung**

**Ergebnis und Anwendung**

# Projektaufgabe

**Use Case:** Erstellung eines **Prognosemodells** für die **Stornowahrscheinlichkeit** jedes einzelnen **Versicherungsnehmers**, um mit gezielten Maßnahmen eine höhere Bestandsbindung in der Vollversicherung zu erreichen durch den Einsatz von **Data Analytics**.

## Data

- umfangreiches Domänenwissen
- Welche Daten sind **relevant und verfügbar**?
- Wo gibt es erwartbare **Unregelmäßigkeiten**?
- Datenverständnis hinsichtlich verlässlicher und sinnvoller Merkmale



## Analytics

- Erfahrung aus Projekten in allen Sparten
- **mathematisches** Methodenwissen
  - Wie muss man Modelle zur Fragestellung passend konfigurieren?
- **wirtschaftliches** Verständnis
  - Was bedeutet ein Modellergebnis ökonomisch?



- Projektablauf und -ziele:
  - Projektdauer etwa 6 Monate; vollständig virtuell über regelmäßige Videokonferenzen
  - zielgenauere Prognose der Stornowahrscheinlichkeit zur Vermeidung von Storno
  - Erhöhung der Akzeptanz und des Verständnisses von Data-Analytics-Verfahren
  - Erfahrungen sammeln mit Data-Analytics-Verfahren



## Benchmark: Bearbeitung der Projektaufgabe durch einen Fachexperten

Es ist eine Kundenbindungsaktion mit einem Budget für 1.000 Kunden geplant. Auftrag an Fachexperten: „Bitte selektieren Sie 1.000 Fälle aus 125.000, die stark stornogefährdet sein könnten!“

- „Sukzessives Filtern in einer Bestands-Excel Tabelle“

**200 Stornierer**  
**125.000 insgesamt**  
**= 0,16%**



Idee: Kunden mit hoher Leistungserstattung kündigen nicht.

**120 Stornierer**  
**57.500 insgesamt**  
**= 0,20%**



Idee: Junge und Kunden mit relevanter Alterungsrückstellung kündigen nicht.

**40 Stornierer**  
**5.450 insgesamt**  
**= 0,7%**



Idee: Kunden mit prozentual geringer Beitragsanpassung (BAP) kündigen nicht.

### Filter: Leistung > 500€

67.500 Fälle haben eine Leistung größer als 500€ erhalten. Davon haben 80 storniert.

→ es verbleiben:  
 $125.000 - 67.500 = 57.500$   
 $200 - 80 = 120$

### Filter: Alter < 28 oder Alter > 35

52.050 Fälle sind jünger als 28 oder älter als 35. Davon haben 80 storniert.

→ es verbleiben:  
 $67.500 - 52.050 = 15.450$   
 $120 - 80 = 40$

### Filter: BAP < k%

3.500 Fälle haben eine prozentual geringe BAP erhalten. Davon haben 25 storniert.

→ es verbleiben:  
 $15.450 - 3.500 = 11.950$   
 $40 - 25 = 15$

# Benchmark: Bearbeitung der Projektaufgabe durch einen Fachexperten

(Fortsetzung)

**15 Stornierer**  
**1.950 insgesamt**  
**= 0,8%**



Idee: Kunden mit hoher Bestandsdauer kündigen nicht.

▼ **Filter: Bestandsdauer**  
**> n Jahre**

600 Fälle haben eine Bestandsdauer größer als n Jahre. Davon haben 2 storniert.

→ es verbleiben:  
 $1.950 - 600 = 1.350$   
 $15 - 2 = 13$

**13 Stornierer**  
**1.350 insgesamt**  
**= 0,9%**



Idee: Kunden mit absolut geringer BAP kündigen nicht.

▼ **Filter: BAP < 12,50€**

350 Fälle haben eine BAP, die kleiner als 12,50€ ist. Davon haben 3 storniert.

→ es verbleiben:  
 $1.350 - 350 = 1.000$   
 $13 - 3 = 10$

**10 Stornierer**  
**1.000 insgesamt**  
**= 1,0%**  
 ursprünglich: 0,16%  
 (ca. 6-mal besser)

**1000 Fälle mit erhöhter Stornowahrscheinlichkeit identifiziert**


Vorhersage: **Storno**  
 (davon waren 990 Nichtstornierer)



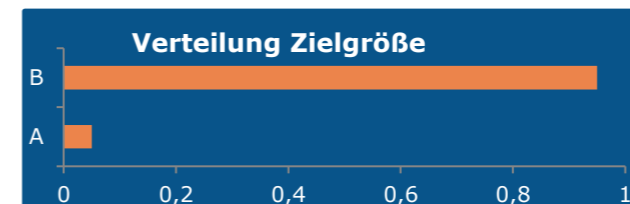
- Ziel des Use-Case: Die identifizierten Fälle sollen möglichst gut sein, z.B. ein Modell, welches 20 Stornofälle unter diesen 1.000 Fällen identifiziert, wäre doppelt so gut.
- **Use Case 2.0:** Erstellung eines Prognosemodells für die Stornowahrscheinlichkeit jedes einzelnen Versicherungsnehmers, um mit gezielten Maßnahmen eine höhere Bestandsbindung für **besonders gefährdete Kunden** in der Vollversicherung zu erreichen.

# Rahmenbedingungen des Use Case

- **Konzentration auf verhinderbares Storno:** Abgang in PKV
- Personen mit einer Krankheitskostenvollversicherung folgende Tarife / Personengruppen werden dabei nicht berücksichtigt:
  - Beihilfe-Tarife
  - Mediziner
  - Ausbildungstarife
  - große Anwartschaften
- Es werden nur Personen ab Alter 21 betrachtet.
- Für die Vorhersage der Abgänge in 2021 liegen uns Daten für die Jahre 2016-2020 vor.
- Die beobachtete Stornowahrscheinlichkeit im Bestand beträgt **0,16%**.

**Ampel** 

- systematische Anzeige eines Scorewertes
- aufgrund geringer Stornozahl insbesondere möglichst gute „rote Ampel“ wichtig



Die geringe Anzahl der Stornofälle ist sehr gut für die Hallesche, aber schlecht für Data Analytics.

# Agenda

**Einleitung**

**Daten**

**Data-Analytics-Verfahren zur Stornoprognose**

**Modelltuning und Optimierung**

**Ergebnis und Anwendung**

# Welche Daten spielen eine erklärende Rolle?

## versicherte Person

z.B. in der privaten  
Krankenvollversicherung



### Person

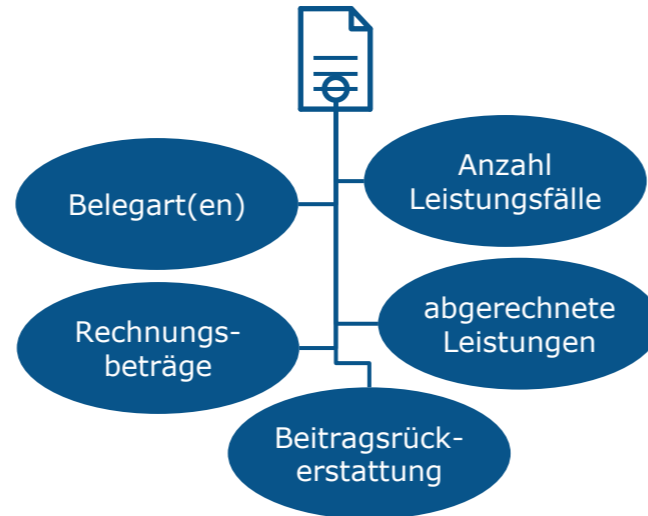
Alter  
Geschlecht  
Wohnort  
Anzahl VP  
...

### Tarif

Vertragsdauer  
Selbstbehalt  
Premium/Basic  
...

## eingereichte Belege

z.B. Leistungshistorie



## Sonstiges

z.B. Makler,  
Ausschließlichkeitsorganisation



### Vermittler

Historie  
Kundenfeedback  
Sozioökonomische  
Daten  
...

- Mögliche weitere Datenquellen: Stornostatistik des PKV-Verbands, dispositiver Bestand, operativer Bestand, Statistikdatenbank, Datenbank zur Beitragsrückerstattung, Statistisches Bundesamt



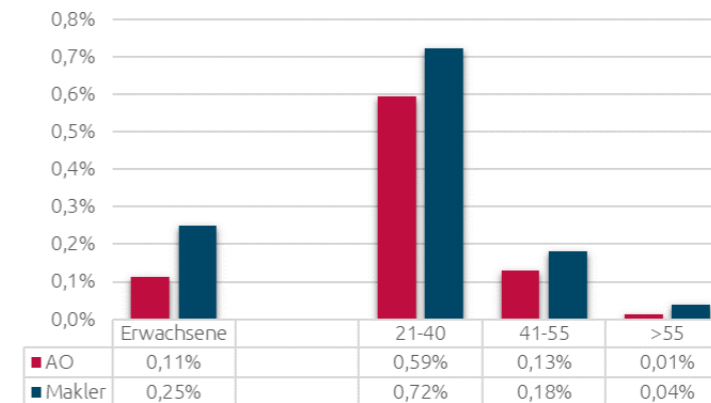
# Datengrundlage

Ein Data-Analytics-Verfahren kann nicht sämtliche Daten eines Krankenversicherers verarbeiten.

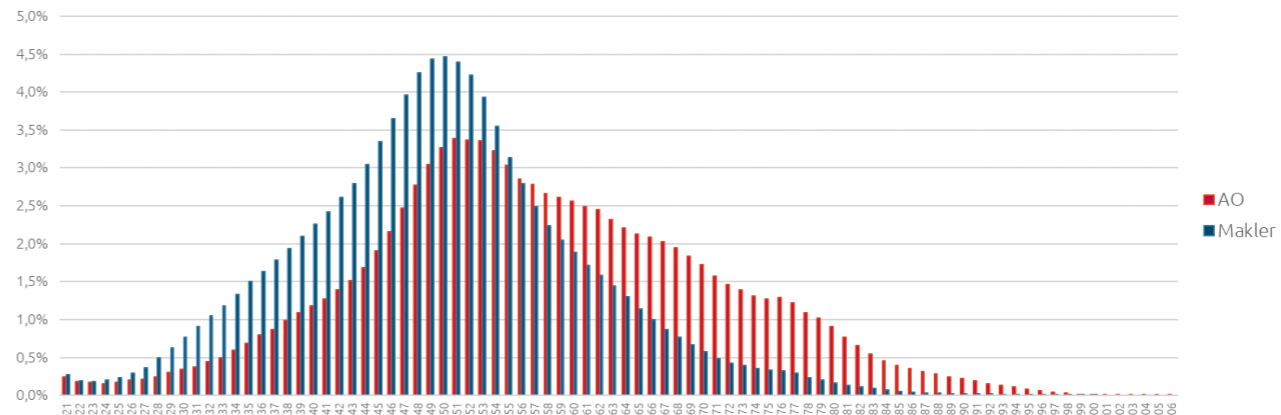
**Es gilt nicht: mehr Daten = bessere Prognose!**

- Probleme dabei:
  - Nicht alle Daten stehen zur Verfügung (Gründe: nicht erfasst, veraltet, technisch nicht möglich, sehr hoher Aufwand).
  - Merkmale müssen teilweise überarbeitet / korrigiert werden (teilw. sehr aufwändig).
  - unklar, ob jedes Merkmal wichtig
    - univariate und bivariate Analysen können helfen aber auch irreführen (s. Beispiel)
    - Methoden zur Merkmalsselektion benötigt
      - hinreichend große Vorauswahl
- Zwischenstand der Datengrundlage:
  - 814.000 Datensätze und 130 Spalten
  - **106 Mio. Datenpunkte**

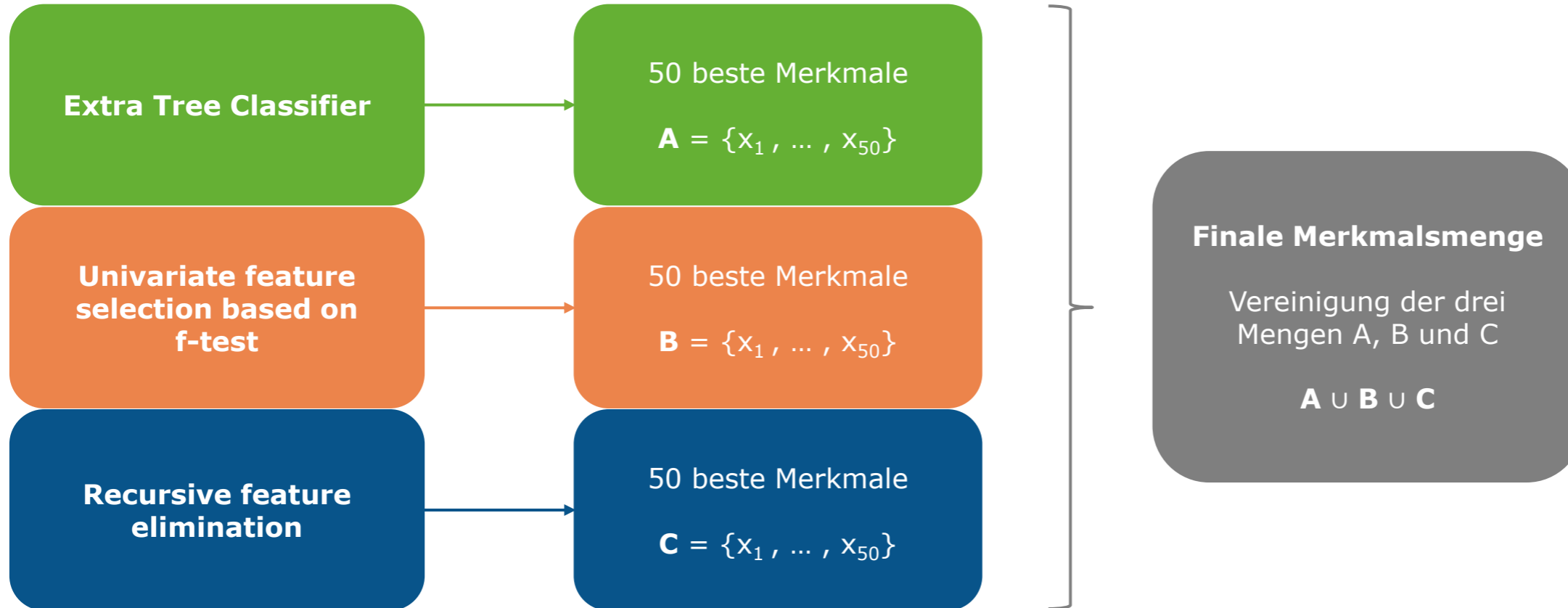
**Stornoquote nach Vertriebsweg**



**Bestandsverteilung für AO und Makler nach Alter**



## Durchführung der Feature Selection



- Endstand der Datengrundlage zur anschließenden Modellierung:
  - 814.000 Datensätze und 70 Spalten → **57 Mio. Datenpunkte**

# Agenda

**Einleitung**

**Daten**

**Data-Analytics-Verfahren zur Stornoprognose**

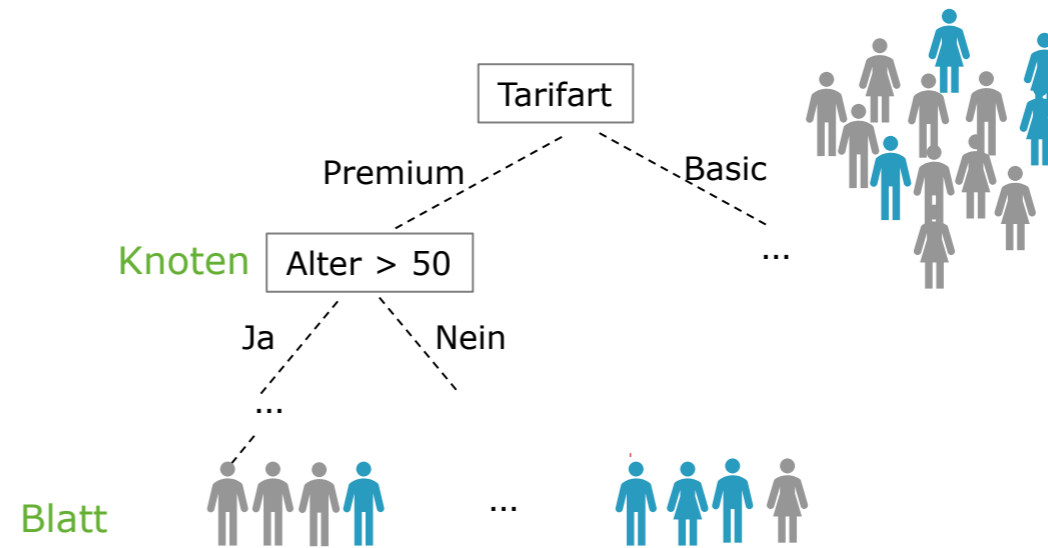
**Modelltuning und Optimierung**

**Ergebnis und Anwendung**

# Data-Analytics-Verfahren zur Stornoprognose

## Entscheidungsbäume

- Entscheidungsbäume unterteilen den Bestand gemäß der Merkmale in Gruppen.



- Der Algorithmus ermittelt, welche Merkmale in die Klassifizierung einfließen sollten, und zwar automatisiert und datengetrieben.**

- Ziel des Algorithmus (vereinfacht):
  - möglichst einheitliche Blätter erzeugen
  - d.h. die Stornierer von den Nichtstornierern zu separieren

# Data-Analytics-Verfahren zur Stornoprognose

## Entscheidungsbäume und Random Forest

- Entscheidungsbäume als Ausgangslage für das Modell

### Stärken von Entscheidungsbäumen

- Einfachheit und Verständlichkeit in der Modellstruktur
- intuitive Interpretation der Modellparameter
- robust gegenüber Ausreißern, Datenanomalien
- einfache und schnelle Kalibrierung

### Schwächen von Entscheidungsbäumen

- Restriktion bei der Mustererkennung **S1**
  - ausschließlich stückweise konstante Prognosen auf Blättern
- anfällig gegenüber Overfitting **S2**
  - zu viele Knoten & Blätter führen zu einer instabilen Performance für neue Daten

- **Random Forest** als geschickte Kombination vieler Bäume

- Ad **S1**: Random Forest als Kombination vieler Bäume ermöglicht bessere Prognosen, vergleichbar zu Neuronalen Netzen oder Boosting-Verfahren.
- Ad **S2**: Random Forest ist extrem robust gegenüber Overfitting.
- Dabei können weitgehend alle Stärken von Entscheidungsbäumen übernommen werden.

# Agenda

**Einleitung**

**Daten**

**Data-Analytics-Verfahren zur Stornoprognose**

**Modelltuning und Optimierung**

**Ergebnis und Anwendung**

# Entscheidungsbäume

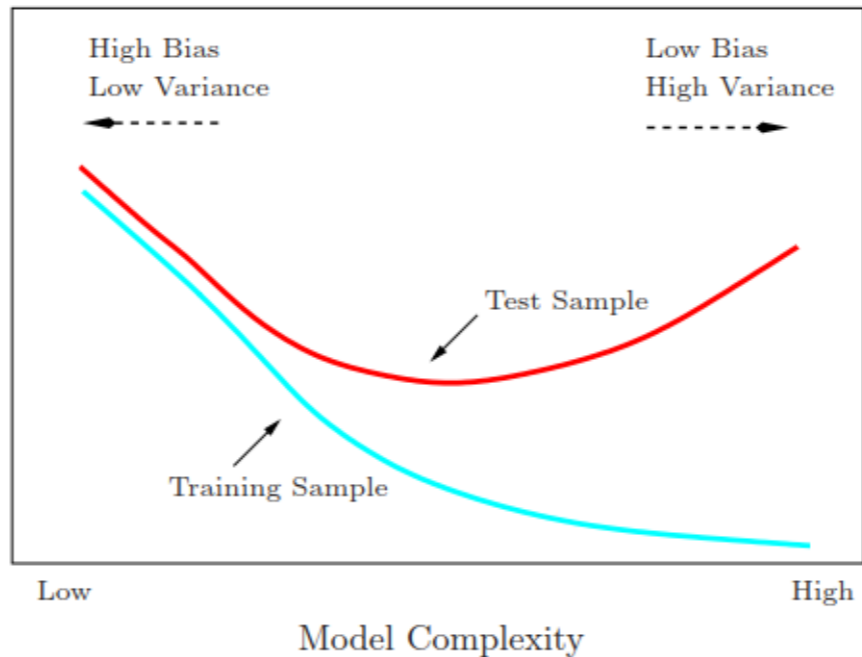
## der Ausgleich zwischen Bias und Varianz

- Der erwartete quadrierte Prognosefehler durch ein Modell  $\hat{f}$  lässt sich aufteilen:



$$E\left((y - \hat{f}(X))^2\right) = \text{Bias}(\hat{f})^2 + \text{Var}_{\text{Modell}}(\epsilon) + \text{Var}(\hat{f})$$

Abweichung der mittleren Modellvorhersagen von der Realität

$$\text{Bias } \hat{f} := E(\hat{f}(X)) - f(X)$$


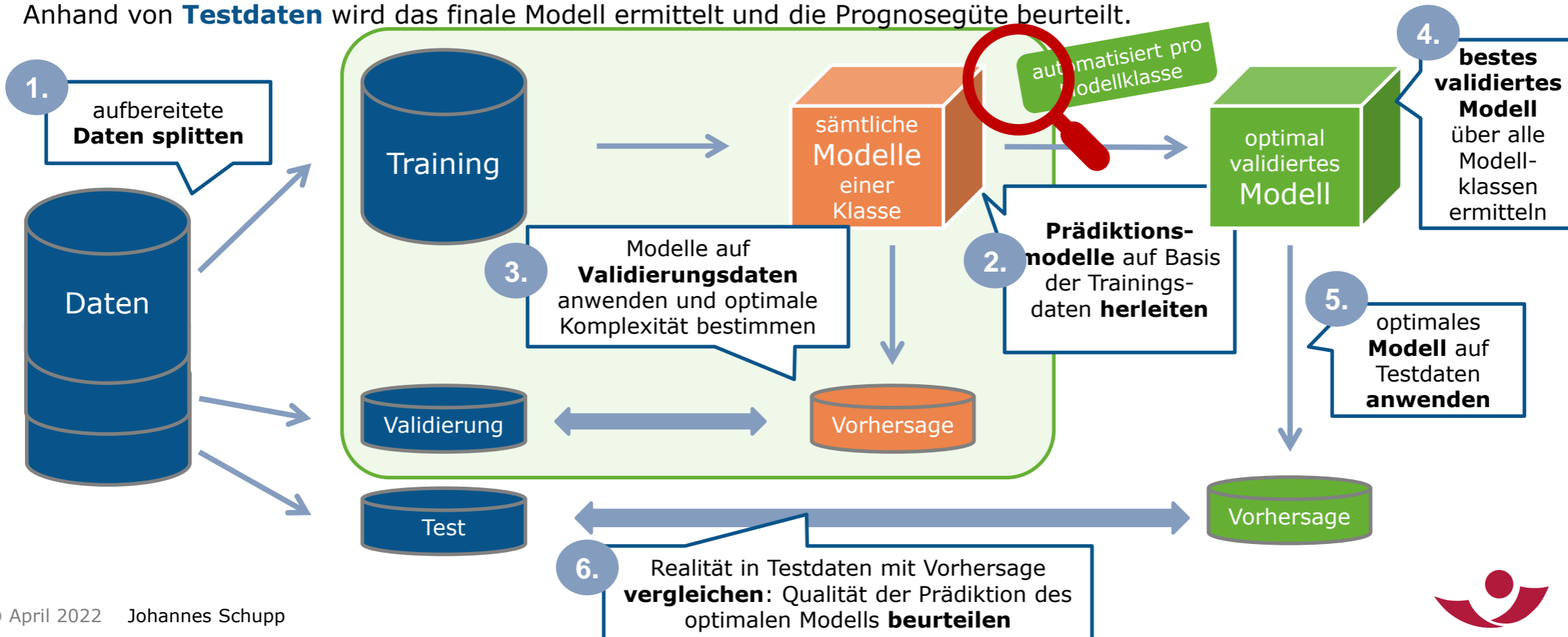
Variabilität in der Modellvorhersage durch die Daten

$$\text{Var } \hat{f} := E\left(\left(\hat{f}(X) - E(\hat{f}(X))\right)^2\right)$$

# Optimierungsprozess

## Training, Validierung und Test

- Die Optimierung des Lernprozess erfolgt mit **Aufteilung der Daten** für Training, Validierung und Test:
- Auf den Erfahrungen in den **Trainingsdaten** lernt jedes Modell (verschiedene Komplexitäten).
- Mittels Erfahrungen in den **Validierungsdaten** wird die optimale Komplexität pro Modell ermittelt.
- Anhand von **Testdaten** wird das finale Modell ermittelt und die Prognosegüte beurteilt.





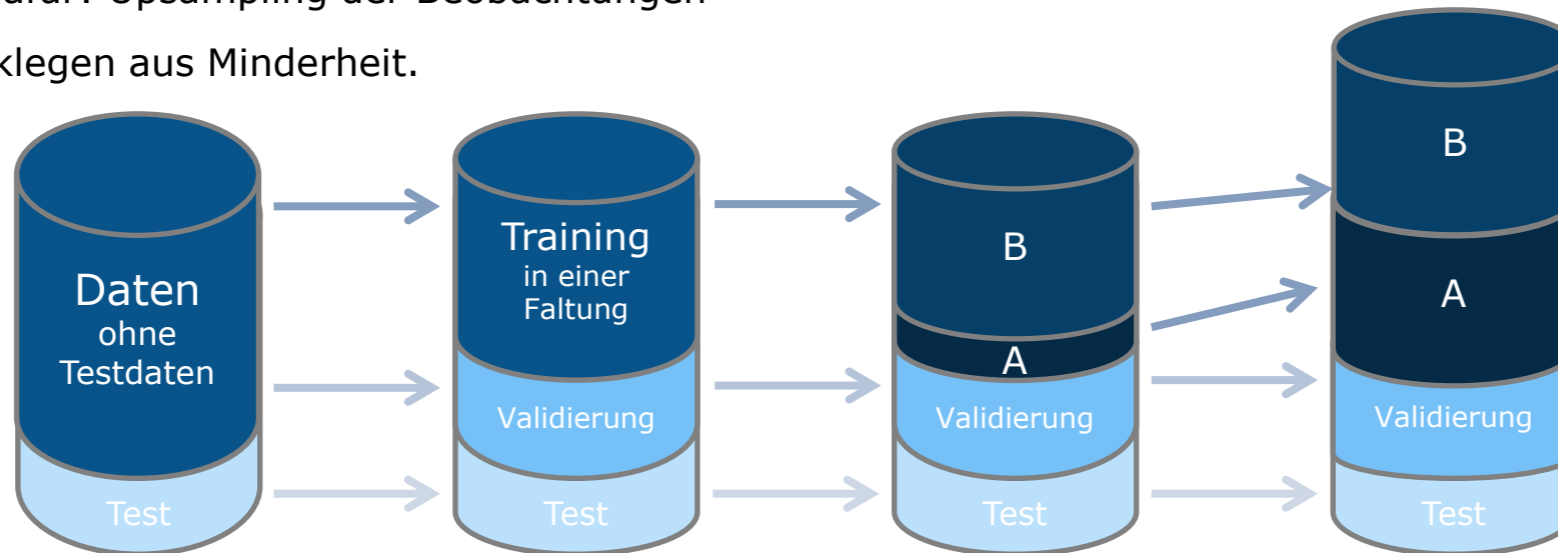
# Optimierungsprozess

## Vorgehen Imbalanced Data



## Umgang mit einem ungleichgewichteten (imbalanced) Datensatz

- Beispielsweise liegt in einem Datensatz die Zielgröße in 1% mit A und in 99% mit B vor.
  - Ein Modell, welches immer B vorhersagt, ist in 99% der Fälle richtig aber trotzdem unnützlich.
- Zum Umgang mit Ungleichgewicht in den Daten kann der Datensatz für das Training (ggf. in jeder Kreuzvalidierungsfaltung) **gewichtet** oder **verändert** werden (direkt oder indirekt).
- eine **Möglichkeit** dafür: Upsampling der Beobachtungen
  - Ziehe mit Zurücklegen aus Minderheit.



- weitere Möglichkeiten: Gewichtung, Downsampling, Ensemble Downsampling, spezielle Evaluationsmetriken

# Evaluationsmetriken: Konfusionsmatrix

Wie kann man Modelle miteinander vergleichen?

## Bewertung anhand der Konfusionsmatrix

- Für jeden Vertrag wird die Stornowahrscheinlichkeit  $P$  anhand **eines Schwellenwerts**  $s \in [0,1]$  in eine Prognose (Storno/Nichtstorno) umgewandelt:
  - Modell sagt Storno, wenn  $P > s$ ,
  - andernfalls: Modell sagt Nichtstorno

		Vorhergesagter Wert	
		Modell sagt: Kein Storno	Modell sagt: Storno
Tatsächlicher Wert	Tatsächlich: Kein Storno	a	b
	Tatsächlich: Storno	c	d

- Sensitivität /Recall  $\frac{d}{d + c}$
- Anteil der korrekt vorhergesagten Abgänge an allen Stornierern
  - zu niedrig: (zu) wenige Abgänge als Stornierer klassifiziert

- Spezifität  $\frac{a}{a + c}$
- Anteil tatsächlicher Nicht-Stornierer an Personen, für die Nicht-Storno vorhergesagt wird

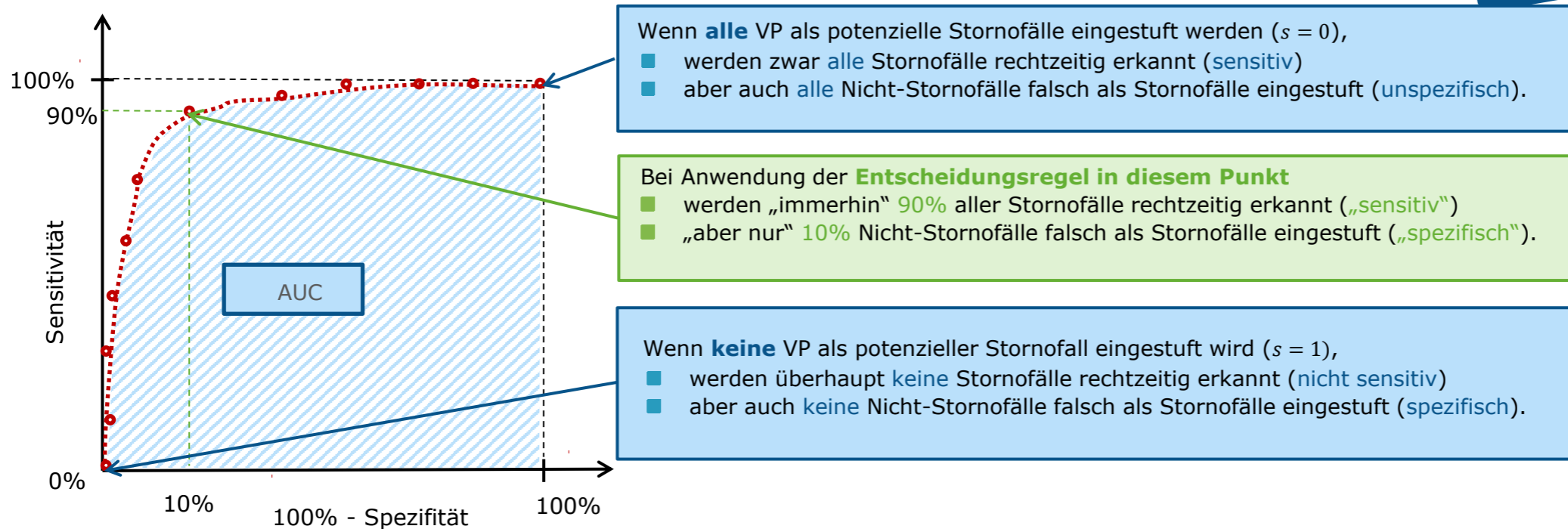
- weitere Metriken: Accuracy, F1-Score, F-beta-Score

# Modellvergleich, Modellgüte, Ergebnisformate

## ROC-Kurve – Stornomodell

- Die **Receiver Operating Characteristic (ROC)-Kurve** ergibt sich durch Anwendung **verschiedener Schwellenwerte**  $s$  auf die Wahrscheinlichkeiten  $P$  (jeder Punkt auf der Kurve steht für einen Schwellenwert).
  - Y-Achse: Anteil der richtigerweise erkannten Stornofälle (im Verhältnis zu allen tatsächlichen Stornofällen) versus
  - X-Achse: den Anteil der fälschlicherweise als Stornofälle identifizierten versicherten Personen VP (im Verhältnis zu allen Nicht-Stornofällen)
  - Die **Area Under the Curve (AUC)** kann zum Vergleich von Modellen verwendet werden.

Beispiel

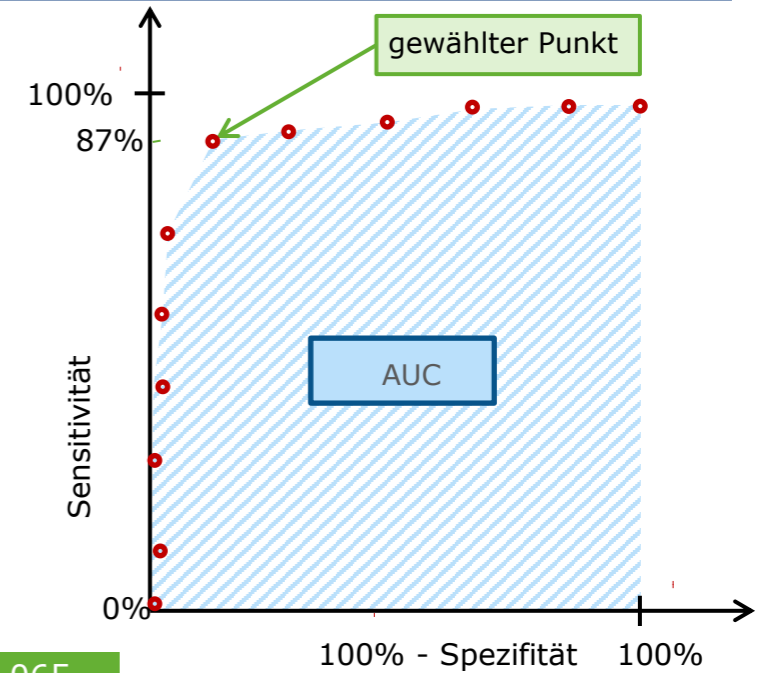


# Evaluationsmetriken: Konfusionsmatrix

Wie kann man Modelle miteinander vergleichen?

- Bewertung anhand der Konfusionsmatrix auf **Trainingsdaten**:

		Vorhergesagter Wert	
		Modell sagt: Kein Storno	Modell sagt: Storno
Tatsächlicher Wert	Tatsächlich: Kein Storno	94.065	30.581
	Tatsächlich: Storno	7.800	52.200



- Sensitivität / Recall**

$$\frac{52.200}{52.200 + 7.800} = 87\%$$

- Anteil der korrekt vorhergesagten Abgänge an allen Stornierern

- Spezifität**

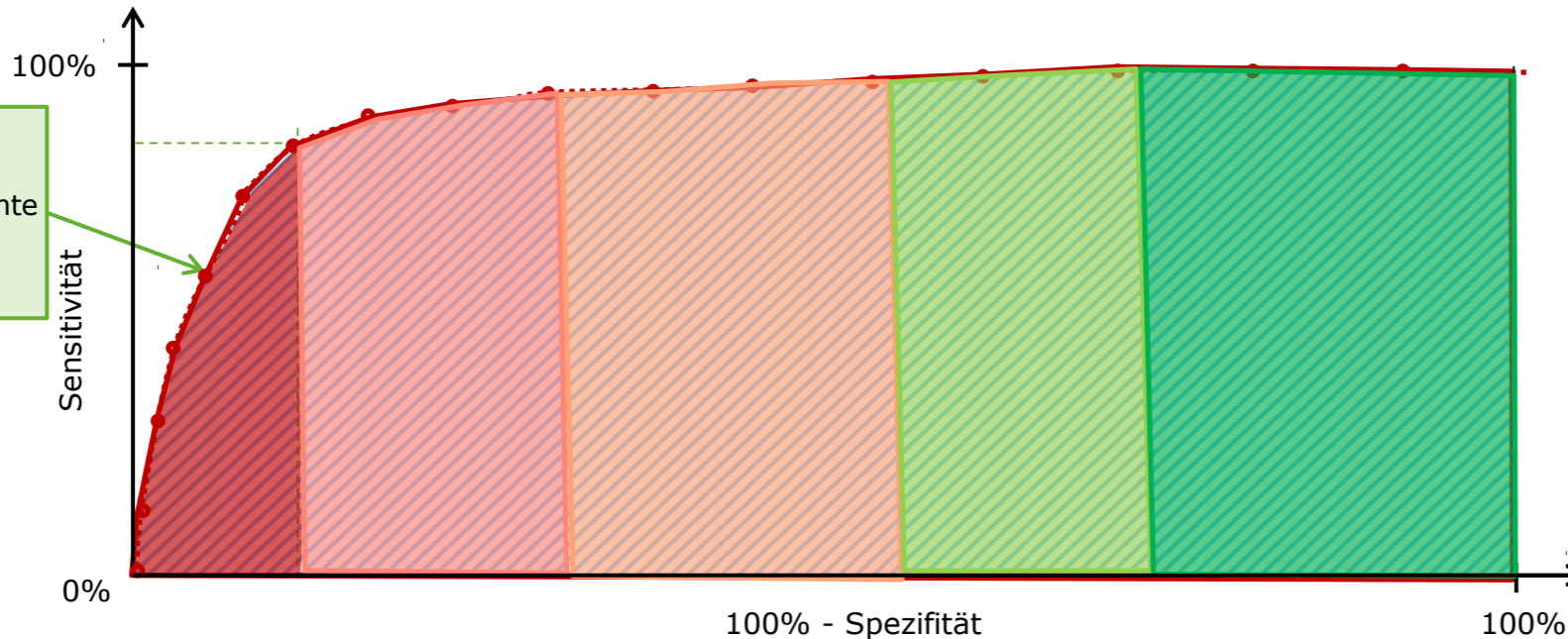
$$\frac{94.065}{94.065 + 7.800} = 92\%$$

- Anteil tatsächlicher Nicht-Stornierer an Personen, für die Nicht-Storno vorhergesagt wird

# Modellvergleich, Modellgüte, Ergebnisformate

## ROC-Kurve – Imbalanced Data

- Zahlenbeispiel: Bei einem Bestand von 500.000 Verträgen mit 1% Stornowahrscheinlichkeit und einem Schwellenwert passend zu 90% Spezifität & 90% Sensitivität:
  - 495.000 Verträge werden als Storno klassifiziert, wovon aber nur 4.500 stornieren, d.h. 490.500 sind falsch als Storno klassifiziert!
  - Der Fokus des Use Case liegt auf Bereichen mit **Spezifität > 90% und Sensitivität << 90%**.
- Anstatt des AUC kann auch nur ein Teil davon (Partial-AUC) optimiert werden.
  - der Höchstwert der False-Positive Rate (100% - Spezifität) wird eingeschränkt, z.B. auf 10%



Partial-AUC  
■ Fokus auf für Anwendung relevante Bereiche

AUC  
■ Bereiche, die irrelevant sind, werden in der Modelloptimierung berücksichtigt

## Finales Modell

### Erst ein maßgeschneidertes Hyperparametertuning liefert ein gutes Ergebnis!

- Sehr großer und unbalancierter Datensatz
  - vorgelagerte Feature Selection
  - stringenter Optimierungsprozess (Training, Validierung, Test, Kreuzvalidierung)
  - Upsampling
  - Wahl eines sehr robusten Algorithmus
  - **notwendig**: umfangreiches Methodenwissen und -Verständnis
- Tuning auf Ziele des Use Case abgestimmt:  
Modell muss nicht optimal gemäß Metrik  $xy$  sein, sondern das optimale Modell für den Use Case!
  - Partial-AUC
  - außerdem: Randomized Search zur Bewertung verschiedener Konfigurationen
  - **notwendig**: umfangreiches Verständnis der wirtschaftlichen Rahmenbedingungen des Use Case

# Agenda

**Einleitung**

**Daten**

**Data-Analytics-Verfahren zur Stornoprognose**

**Modelltuning und Optimierung**

**Ergebnis und Anwendung**

# Projektergebnisse

## Entwicklung eines Scorewertes



- Ziel der Modellkalibrierung: die „rote Ampel“ soll möglichst gut sein!
- Für neue Datensätze kann das Modell zur Bestimmung einer Stornowahrscheinlichkeit  $P$  für jeden Vertrag verwendet werden.
  - Der Schwellenwert  $s$  kann übernommen werden, siehe Stand 31.03.2021.
  - Der Schwellenwert  $s$  kann aber auch dynamisch angepasst werden, z.B. abhängig von Kundenbetreuerkapazitäten erhöht (reduziert) werden → weniger (mehr) prognostizierte Stornofälle.

Finale Vorhersage des Modells für Testdatensatz:

Scorewert	Stornogefahr	Anteil
0	keine	70,7%
1	kaum	13,1%
2	gering	5,5%
3	erhöht	5,0%
4	groß	4,9%
5	sehr groß	0,8%

gesamt

100%

Tatsächliche Beobachtung:

Scorewert-Stornoquote
0,01%
0,16%
0,23%
0,54%
1,30%
3,33%

0,16%

Finale Vorhersage des Modells für 2021:

Scorewert	Stornogefahr	Anteil
0	keine	71,1%
1	kaum	12,3%
2	gering	5,8%
3	erhöht	4,3%
4	groß	4,9%
5	sehr groß	1,5%

gesamt

100%

Tatsächliche Beobachtung:

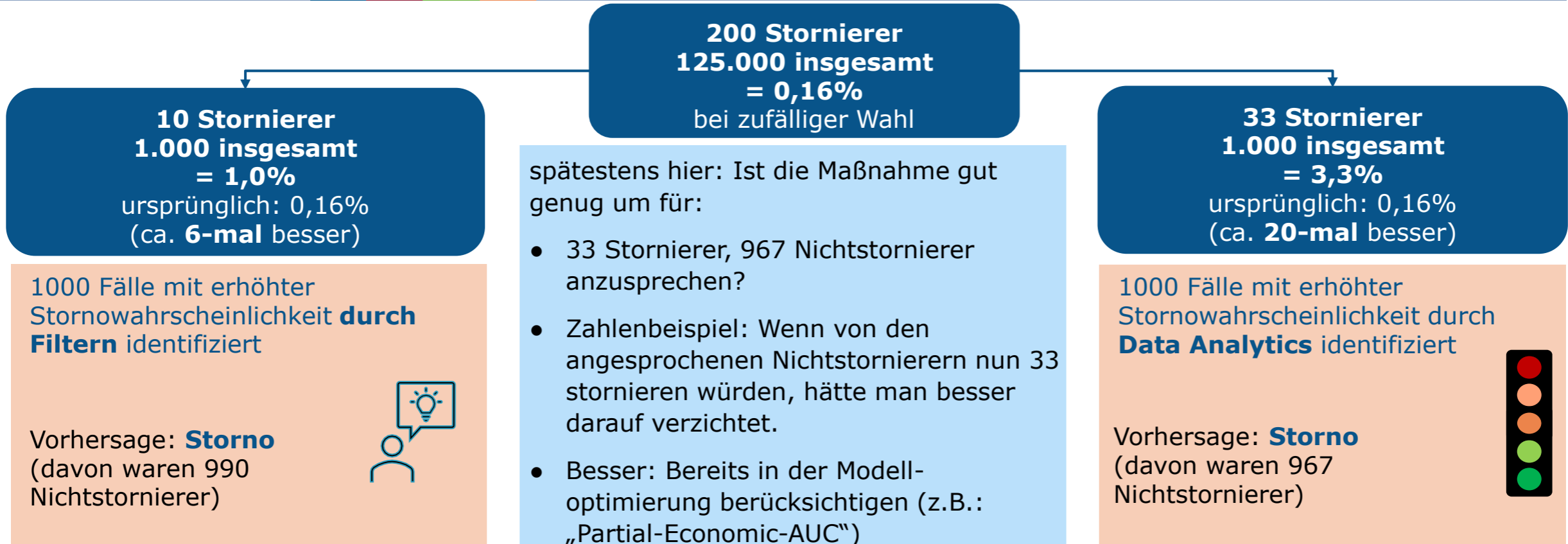
Scorewert-Stornoquote
0,03%
0,34%
0,78%
0,96%
1,81%
2,34%

0,28%

Stand  
31.03.2021



# Verbesserung durch Data-Analytics-Verfahren



Die ökonomische Anwendung (z.B. in Verbindung mit den angedachten Maßnahmen) steht im Zentrum jeder Data-Analytics-Anwendung.

## Fazit & Lessons Learned

- Data-Analytics-Verfahren können umfangreiche Datenmengen auf Muster untersuchen.
  - Für Storno besonders relevant: Alterungsrückstellung, BAP-Historie, Bestandsdauer
- Abhängig von der **zuvor festgelegten** Fragestellung in der Unternehmenssteuerung kann das Modell optimiert werden.

### Ampel



- Scorewerte
- zusätzlich: Ermittlung der Merkmale für hohen Scorewert

→ Das Modell liefert eine Stornowahrscheinlichkeit für jeden Vertrag.

→ Bei der Optimierung dieses Modells liegt der Fokus jedoch auf der "roten Ampel".

- entscheidend sind dabei
  - **versicherungstechnisches Wissen** → welche relevanten Daten kann man bekommen?
  - **Programmierkenntnisse** → Umsetzung in Python
  - **versicherungswirtschaftliches Wissen** → welche Maßnahme macht Sinn?
  - **statistisches Wissen** → wie werden Modelle eingestellt und optimiert?

# Vielen Dank für Ihre Aufmerksamkeit!

**Kinga Böhm**  
kinga.boehm@hallesche.de



**Dr. Johannes Schupp**  
+49 (731) 20 644-241  
j.schupp@ifa-ulm.de

