

Prof. Dr. Benedikt Funke (TH Köln)

Dr. Simon Hatzesberger (Allianz Private Krankenversicherungs-AG)

Dr. Lars Kunze (SIGNAL IDUNA Gruppe)

Analyse aktuarieller Fragestellungen an SOEP-Daten mittels ML- und XAI-Methoden

Fachgruppe Actuarial Data Science, 26.04.2024

Agenda

1 Motivation

2 SOEP-Daten

3 Anwendung und Evaluation von ML-Verfahren

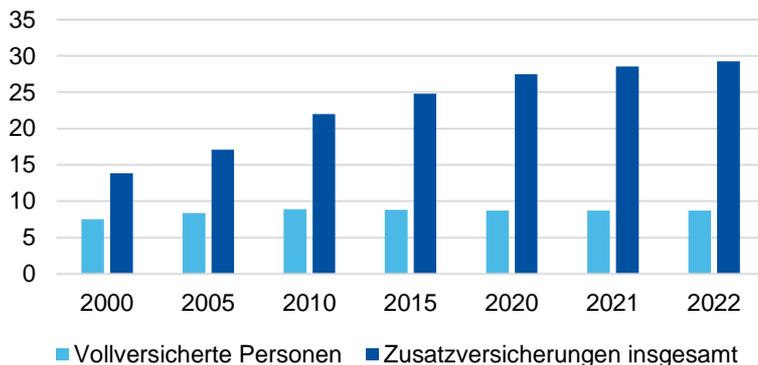
4 Erkenntnisgewinne mittels Explainable AI

5 Fazit & Ausblick

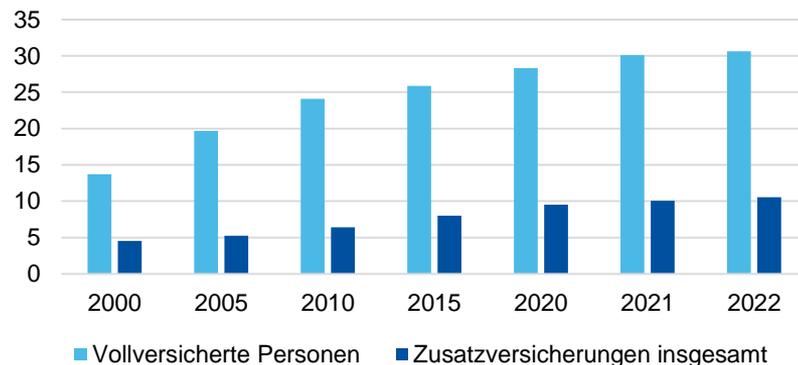
1 Motivation

Relevanz der privaten Krankenzusatzversicherung

Entwicklung der Krankheitsvoll- und Zusatzversicherungen in Mio.



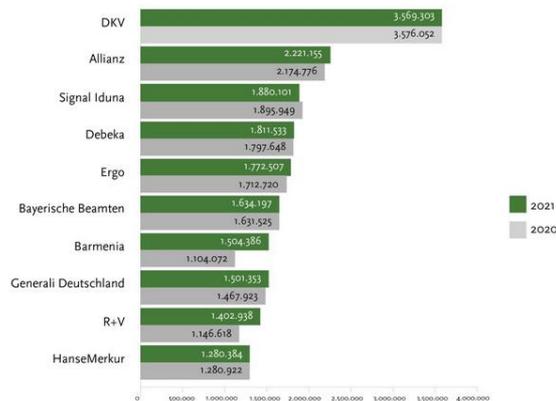
Entwicklung der Beitragseinnahmen in Mrd. EUR



- ▶ Starkes Bestandswachstum in der privaten Krankenzusatzversicherung (KZV) in den letzten Jahren
- ▶ Beitragsvolumen der privaten KZV ebenfalls deutlich gestiegen

Relevanz der privaten Krankenzusatzversicherung

Personen Krankenzusatzversicherung



Quelle: [Assekurata](#) (November 2022)

Boom bei betrieblicher
Krankenversicherung



Quelle: [PKV-Verband](#) (Januar 2024)

Relevanz der privaten Krankenzusatzversicherung

Pressemittteilung

AOK Bayern und Allianz kooperieren bei Zusatzversicherungen

16.10.2023 · AOK Bayern · 3 Min. Lesedauer

Abblättern



Unterszeichnung der neuen Kooperation in U.S. Daniel Böhr, Vorstand der Allianz Private Krankenversicherung, Nina Pfleger, Vorstandsvorsitzende der Allianz Private Krankenversicherung, Dr. Ingrid Stöper, Vorstandsvorsitzende der AOK Bayern, Stephan Abke, stellvertretender Vorstandsvorsitzender der AOK Bayern.

Quelle: [Versicherungsmonitor](#) (Oktober 2023)

TOP 10 InsurTechs Deutschland¹ (Gebuchte Beiträge in m€)



Confidential | Non Personal Data

Wachstum 2021/22 (in m€)

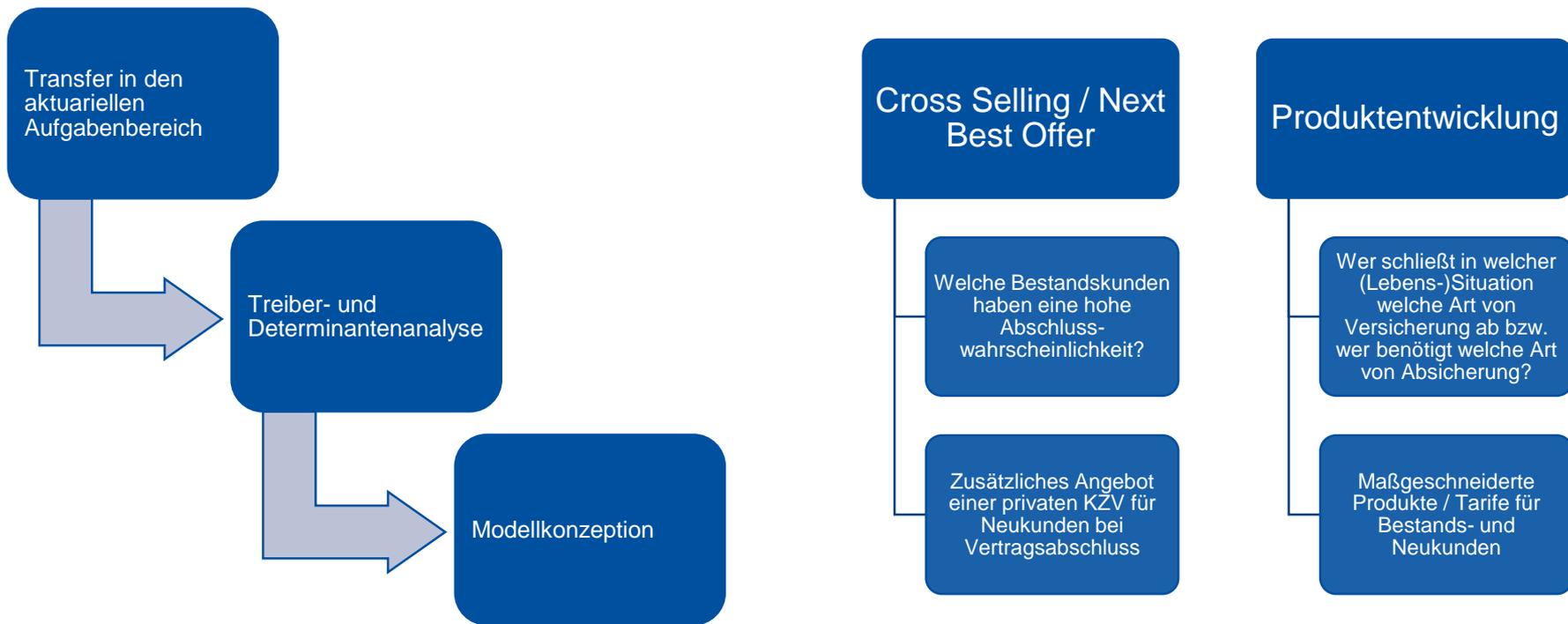


¹ IMWP Institute 2023. Quelle „Insurtechs in Deutschland“ ² nur KFZ (ohne Neckermann-Altbestand)

© Zurich

Quelle: [Versicherungsmonitor](#) (Juni 2023)

Relevanz der privaten Krankenzusatzversicherung



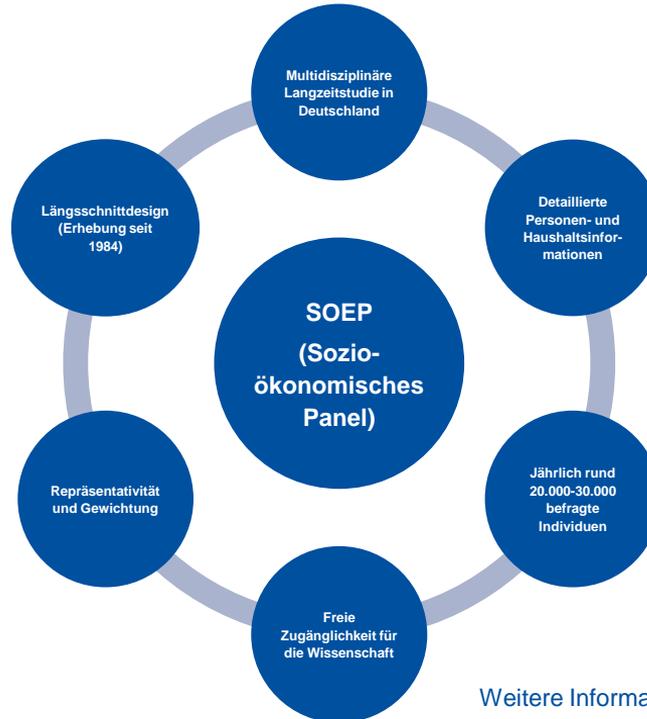
Relevanz der privaten Krankenzusatzversicherung

Ähnliche Fragestellungen wurden bereits in der wissenschaftlichen Literatur untersucht:

- ▶ **Lange et al. (2017):** Analyse des Erwerbs privater Krankenzusatzversicherungen in Deutschland.
Haupttreiber: Versicherungsneigung und Einkommen
- ▶ **Bonsang und Costa-Font (2022):** Analyse von Längsschnittdaten aus Deutschland.
Erkenntnisgewinn: Erwerb von privater Krankenzusatzversicherung ist von der Kontrollüberzeugung des eigenen Einflussbereichs (*internal locus of control*) abhängig
- ▶ **Eckert et al. (2021):** Methodisch ähnliche Vorgehensweise.
Thema: Identifikation von Risikofaktoren für die Berufsunfähigkeit mit Methoden des Maschinellen Lernens

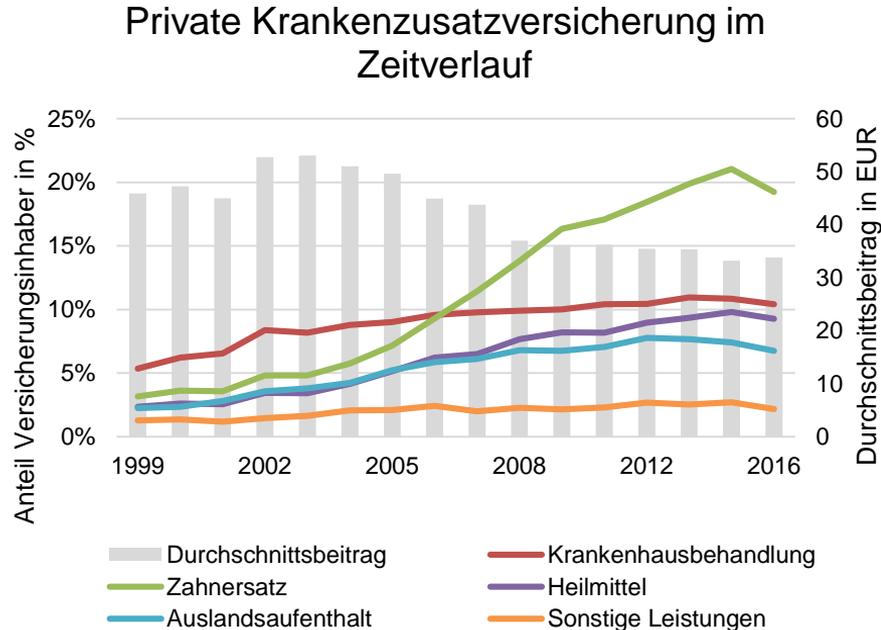
2 SOEP-Daten

SOEP-Studie: Grundlegendes



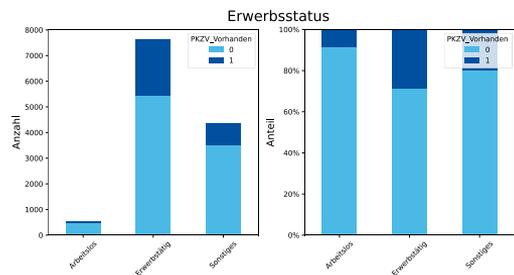
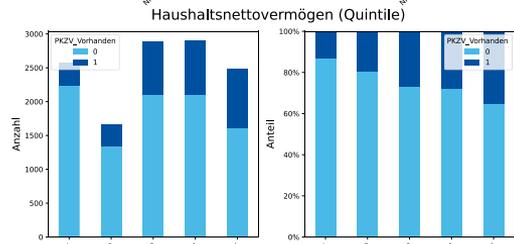
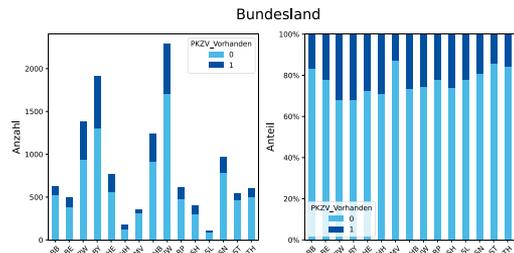
Weitere Informationen zur SOEP-Studie beim [DIW Berlin](#)

SOEP-Daten: Trend der Zielvariablen

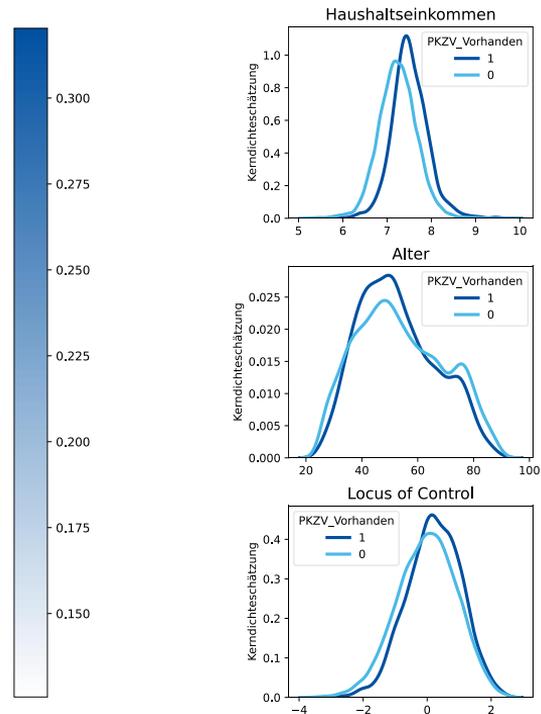
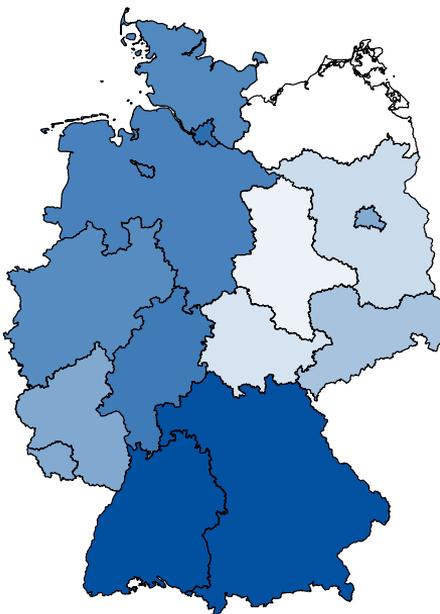


- ▶ Steigender Anteil des Vorhandenseins einer privaten KZV im Zeitverlauf; Haupttreiber: Zahnzusatzversicherungen
- ▶ Durchschnittsbeitrag rückläufig
- ▶ 45 % / 26 % / 18 % / 10 % / 2 % der VN haben eine / zwei / drei / vier / fünf Leistungsarten abgesichert
- ▶ Für die Analyse verwenden wir Daten für 2016 (Grund: Vergleichbarkeit zu Literatur, keine Kriseneffekte)

Verteilung der Zielvariablen bei ausgewählten Features



Anteil PKZV_Vorhanden je Bundesland



3

Anwendung und Evaluation von ML-Verfahren

Problemstellung

Binäres Klassifikationsproblem

Prognose des **Vorhandenseins einer privaten Krankenzusatzversicherung** auf Basis von Merkmalen wie z. B. Alter, Haushaltseinkommen, Locus of Control, Vorhandensein einer Lebensversicherung, ...

Bildungs- jahre	Alter	Haushalts- einkommen	Locus of Control	Bundes- land	Beschäf- tigung	LV vor- handen	...	PKZV vor- handen
10,5	41	7,485	1,205	BW	erwerbstätig	1	...	1
10,5	53	7,534	-0,042	MV	Sonstiges	0	...	0
12,0	72	7,268	-1,113	BY	Sonstiges	0	...	1
11,5	36	6,617	-2,023	BB	arbeitslos	1	...	0
12,0	28	7,483	1,209	NW	erwerbstätig	1	...	?
12,0	44	7,318	0,542	ST	erwerbstätig	0	...	?

Gütemaß

Primär: **AUC** (Area under Curve)

	AUC \in (0.5, 0.6] \triangleq unbrauchbar
	AUC \in (0.6, 0.7] \triangleq schlecht
	AUC \in (0.7, 0.8] \triangleq moderat
	AUC \in (0.8, 0.9] \triangleq gut
	AUC \in (0.9, 1.0] \triangleq exzellent

Sekundär: **Accuracy**

Klassischer Ansatz: Logistische Regression

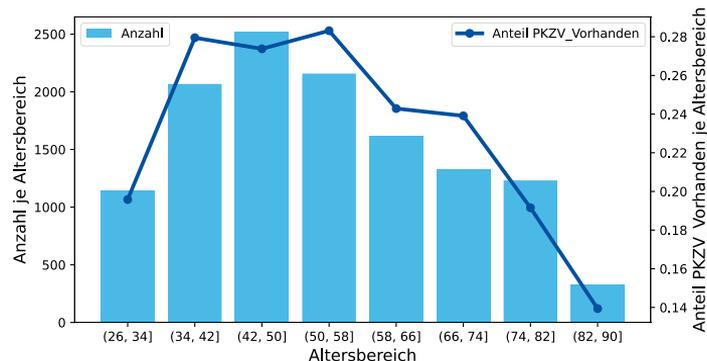
(Ausgewählte) Vorteile

- + etabliertes, wohlbekanntes Verfahren
- + einfach zu erklären und zu interpretieren

```
=====
              coef   std err          z      P>|z|
-----
const          -11.1070    0.651   -17.070    0.000
BildungAnzJahre    0.0728    0.011    6.568    0.000
Risikobereitschaft  0.0403    0.015    2.677    0.007
Alter          -0.0008    0.003   -0.300    0.764
HHEinkommenlrneq  0.09100   0.072   12.618    0.000
LV_Vorhanden     0.3825    0.053    7.171    0.000
...
=====
```

(Ausgewählte) Nachteile

- Prognosegüte oft lediglich akzeptabel
- Nichtlinearitäten nicht adäquat abbildbar



Evaluation der verwendeten ML-Verfahren

		Logistische Regression (LASSO)	Künstliches Neuronales Netz	Random Forest	LightGBM	XGBoost	CatBoost	AdaBoost
AUC	in-sample	0,725	0,727	0,758	0,794	0,820	0,805	0,765
	out-of-sample	0,720	0,715	0,720	0,725	0,723	0,732	0,720
Accuracy	in-sample	0,752	0,754	0,753	0,787	0,805	0,787	0,761
	out-of-sample	0,755	0,746	0,748	0,751	0,758	0,753	0,760



Erkenntnisse:

- ▶ Performance der Logistischen Regression lediglich moderat
- ▶ fortgeschrittenere Machine-Learning-Verfahren schneiden besser ab, wenngleich nur geringfügig
- ▶ AUC-Ergebnisse werden über weitere Fehlermaße (hier: Accuracy) bestätigt

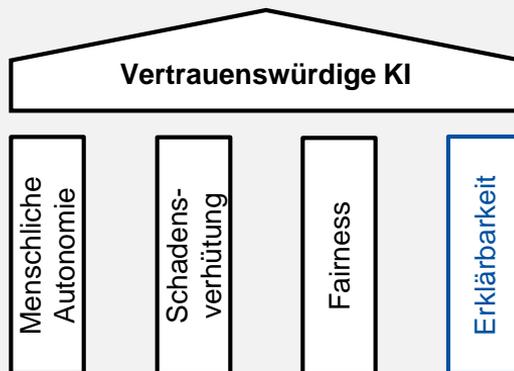
4 Erkenntnisgewinne mittels Explainable AI

Legitimation von Explainable AI (XAI)



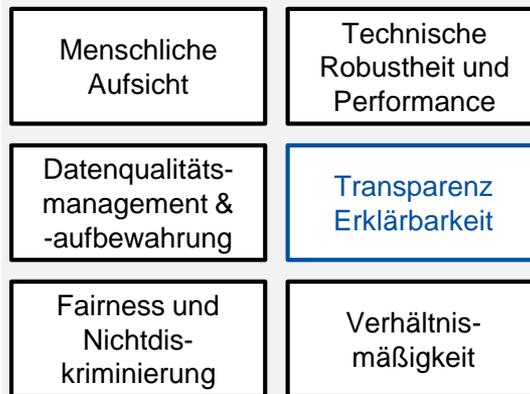
Branchenübergreifende Leitlinien

Ethik-Leitlinien der High-Level Expertengruppe der Europäischen Kommission



Versicherungsspezifische Leitlinien

AI Governance Principles



EU Artificial Intelligence Act

Inakzeptables Risiko

Hohes Risiko

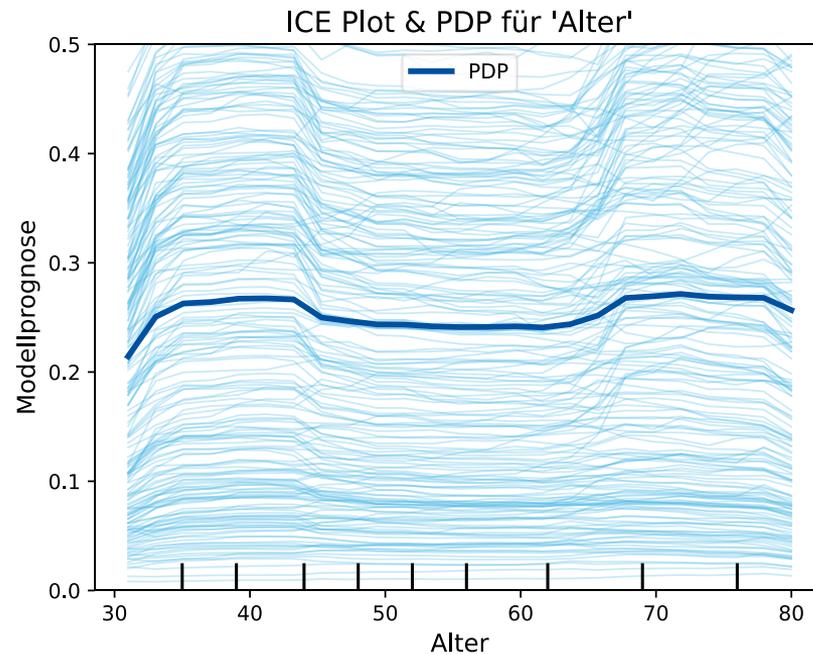
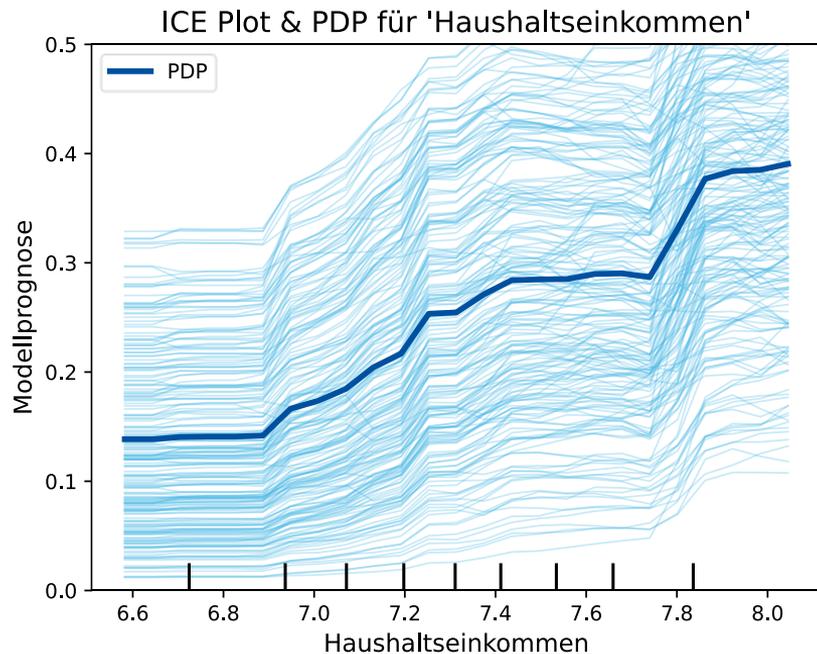
Begrenztes Risiko

Minimales Risiko

Article 13(1):

„High-risk AI systems shall be designed and developed in such a way to ensure that their operation is sufficiently transparent to enable deployers to *interpret the system's output* and use it appropriately. [...]“

XAI: ICE & Partial Dependence Plots



XAI: SHAP für lokale Interpretierbarkeit

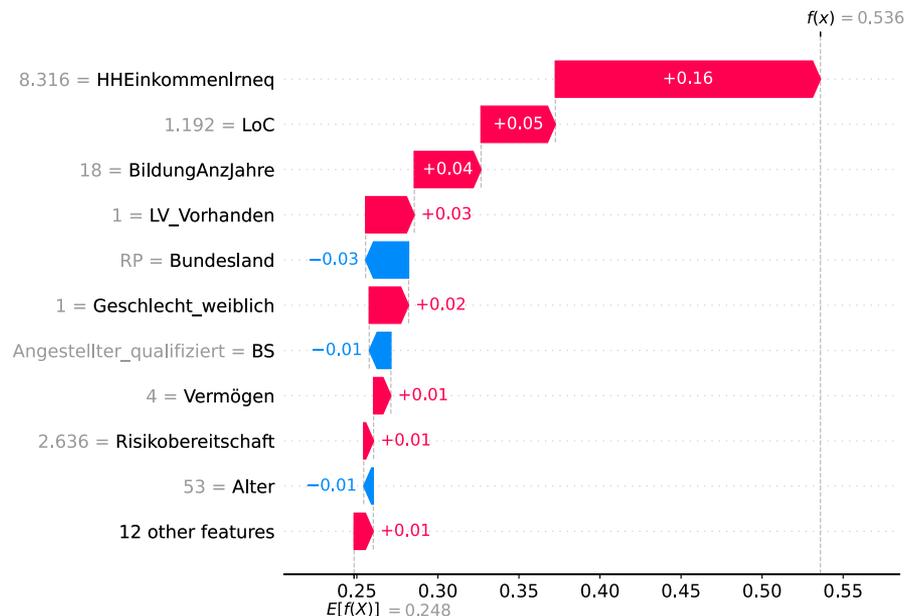
Grundgedanke zu SHAP:

Verwendung von Shapley-Werten zur additiven Zerlegung der Abweichung einer Vorhersage zum Mittel aller Vorhersagen; d. h.,

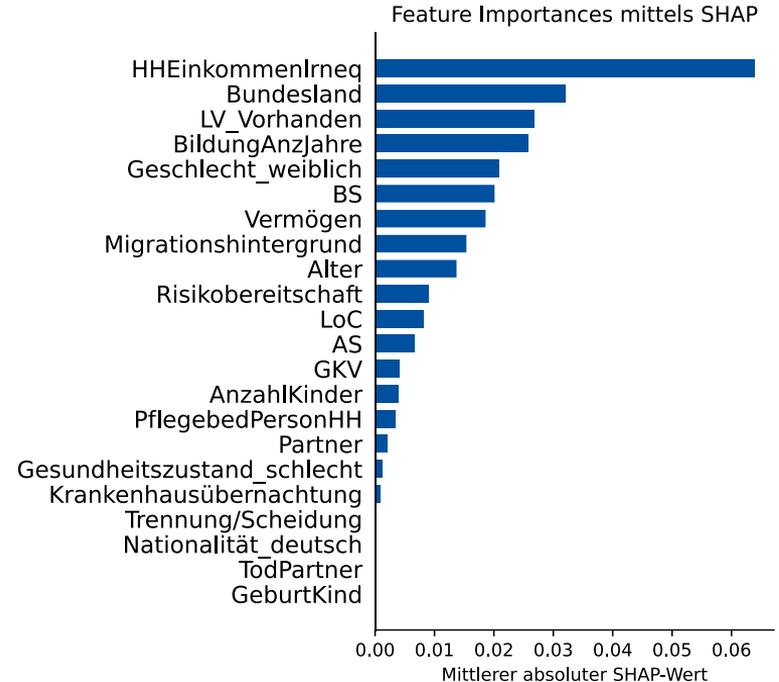
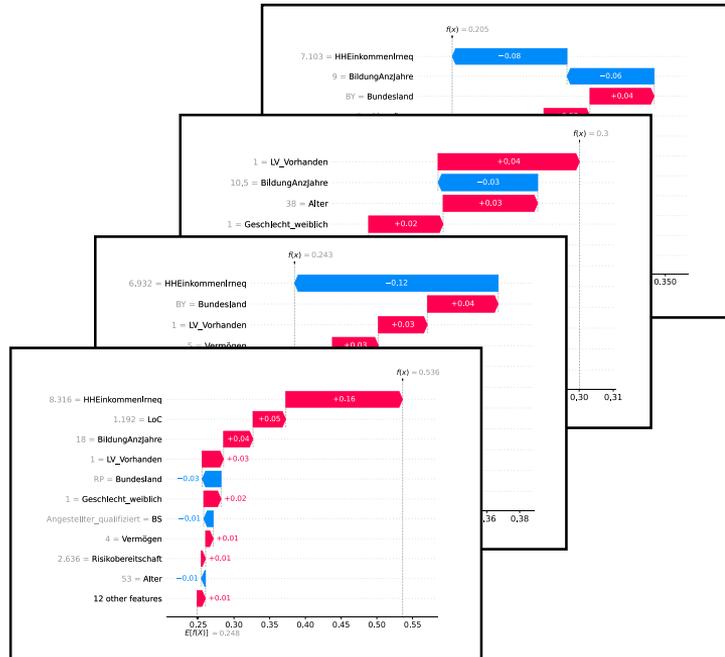
$$\hat{y}_i = \bar{y} + \sum_{j=1}^m \phi_{i,j}$$

wobei

- ▶ \hat{y}_i die Modellprognose zur i -ten Zeile ist,
- ▶ \bar{y} der Durchschnitt über alle Modellprognosen ist und
- ▶ $\phi_{i,j}$ der Shapley-Wert des Merkmals j der i -ten Zeile ist.



XAI: SHAP für globale Interpretierbarkeit



5 **Fazit & Ausblick**

Fazit & Ausblick

Fazit

- ▶ Prognose in vorliegender Problemstellung herausfordernd
- ▶ ML-Verfahren performen geringfügig besser als Logistische Regression
- ▶ identifizierte Determinanten (Haushaltseinkommen, Bundesland, Versicherungsneigung, Locus of Control) decken sich i. W. mit bisherigen Ergebnissen aus der Literatur

Ausblick

- ▶ Verwendung von zusätzlichem Feature Engineering und/oder zusätzlicher ML-Modelle
- ▶ Fokussierung auf spezielle Krankenzusatzversicherungsarten (z. B. Zahnzusatz)
- ▶ Fokussierung auf andere Versicherungsarten (z. B. Lebensversicherung)
- ▶ Prognose von “Erwerb” statt von “Vorhandensein”
- ▶ Prognose Zahlungsbereitschaft in EUR

Literaturverzeichnis

- ▶ Renate Lange, Jörg Schiller und Petra Steinorth, Demand and selection effects in supplemental health insurance in Germany, The Geneva Papers on Risk and Insurance – Issues and Practice (2017) 42: 5-30.
- ▶ Eric Bonsang und Joan Costa-Font, Buying control? ‘Locus of control’ and the uptake of supplementary health insurance, Journal of Economic Behavior and Organization (2022) 204: 476-489.
- ▶ Christian Eckert, Daniela Giesinger, Felix Müller und Antonia Schöning, Machine Learning in der Berufsunfähigkeitsversicherung? Eine Analyse von Risikofaktoren, Der Aktuar (2021), 2: 86-92.

**Vielen Dank für Ihre
Aufmerksamkeit.**

Prof. Dr. Benedikt Funke (TH Köln)

Dr. Simon Hatzesberger (Allianz Private Krankenversicherungs-AG)

Dr. Lars Kunze (SIGNAL IDUNA Gruppe)

*Analyse aktuarieller Fragestellungen an
SOEP-Daten mittels ML- und XAI-Methoden*

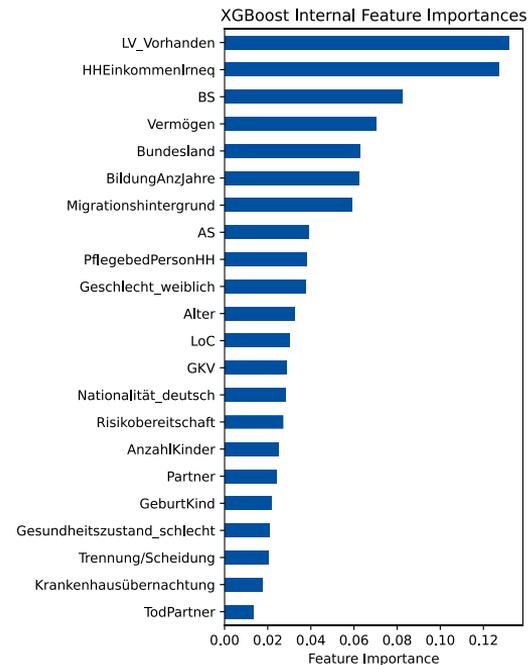
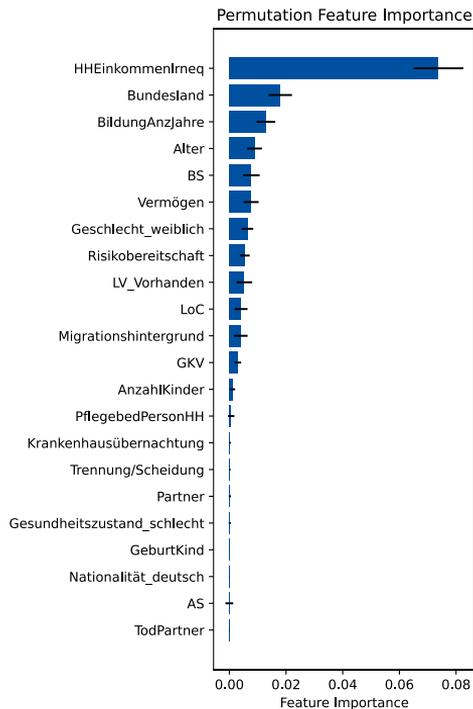
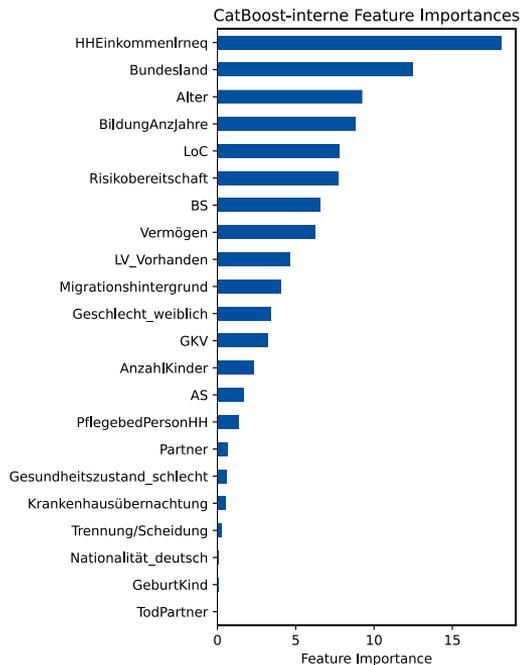
DAV/DGVFM
Jahrestagung
2024

Backup-Folien

Details zum Locus of Control

- ▶ Locus of Control (LoC) ist ein persönlichkeitspsychologisches Konstrukt, das sich auf die von einer Person wahrgenommene Verursachung von Situationen bezieht
- ▶ Personen mit einem sog. internalen LoC sind der Überzeugung, dass Ereignisse stark von ihren eigenen Fähigkeiten und ihren Anstrengungen gesteuert werden können
- ▶ Konstruktion: Hauptkomponentenanalyse auf Basis der nachfolgenden Items, welche jeweils auf einer Skala von 1 bis 7 gemessen werden (1: Stimme überhaupt nicht zu, 7: Stimme voll zu)
 - Habe nicht das erreicht, was ich verdiene
 - Mein Lebenslauf hängt von mir ab
 - Was man erreicht, hängt vom Glück ab
 - Andere bestimmten über mein Leben
 - Zweifle bei Schwierigkeiten an meinen Fähigkeiten
 - Möglichkeiten werden von sozialen Umständen bestimmt
 - Wenig Kontrolle über mein Leben
- ▶ Vorhersage des ersten Faktors liefert stetiges Maß für sog. internalen LoC (vgl. Bonsang und Costa-Font (2022) und Quellen darin)

XAI: Weitere Feature Importances (1/2)



XAI: Weitere Feature Importances (2/2)

	HHEinkommen	Bundesland	LV_Vorhanden	BildungAnzJahre	Geschlecht	Beschäftigung	Vermögen	Migrationshintergr.	Alter	Risikobereitschaft	Locus of Control	Anstellung	GKV	Anzahl Kinder	Pflegeb. Pers. HH	Partner	Gesundh.zustand	KH-Übernachtung	Trennung/Scheid.	Nationalität	Tod Partner	Geburt Kind
CatBoost SHAP	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
CatBoost PFI	1	2	9	3	7	5	6	11	4	8	10	21	12	13	14	17	18	15	16	20	22	19
CatBoost intern	1	2	9	4	11	7	8	10	3	6	5	14	12	13	15	16	17	18	19	20	22	21
XGBoost intern	2	5	1	6	10	3	4	7	11	15	12	8	13	16	9	17	19	21	20	14	22	18
∅	1,25	2,75	5,50	2,25	8,25	5,25	6,25	9,00	6,75	9,75	9,50	13,75	12,50	14,00	13,25	16,50	17,75	18,00	18,50	18,50	21,75	20,00