DAV/DGVFM
Jahrestagung
2024

Dr. Lukas Hahn, SV SparkassenVersicherung

XAI@SV

ML-Modelle im GLM-Gewand

DAV-Jahrestagung 2024 in Berlin – Fachgruppe ADS
26.04.2024

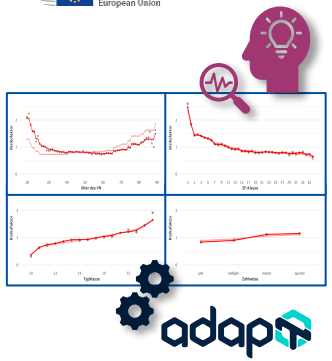
Das Fazit vorab

- explizite (neue?) Anforderung an Erklärbarkeit von KI-Anwendungen
- Abbildung über Post-Hoc-Erklärung i.A. nur approximativ
- Alternative: Surrogat-Modelle mit modell-intrinsischer Erklärbarkeit
- SV: Eigenentwicklung für automatisierte datengetriebene GLM-Erstellung
- Fallstudie: Kombination beider Welten funktioniert

➔ u.a. höhere Anforderungen an Erklärbarkeit

➔ Erfüllung von Anforderungen unklar

➔ historisch hohe Akzeptanz von GLMs in Versicherungen (inkl. Regulator)



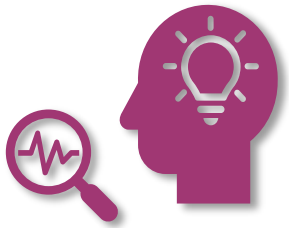
➔ neuer Anwendungsfall zur Erstellung erklärbarer Surrogat-Modelle für ML-Modelle

➔ *Teilung:* Mustererkennung + Erklärbarkeit
Güte: kein signifikanter Verlust
Prozess: zu großen Teilen automatisiert

Explainable AI – Begriffe und Definitionen

- Interpretierbarkeit / Erklärbarkeit
  = Verständnis über Zusammensetzung bzw. Wirkung eines **gesamten Regelwerks**

Interpretability
is the degree to which a
human…



“

…can **understand** the
cause of a decision.

Miller, Tim. “Explanation in artificial intelligence: Insights from the social sciences.” arXiv Preprint arXiv:1706.07269. (2017).

…can consistently **predict**
the model's result.

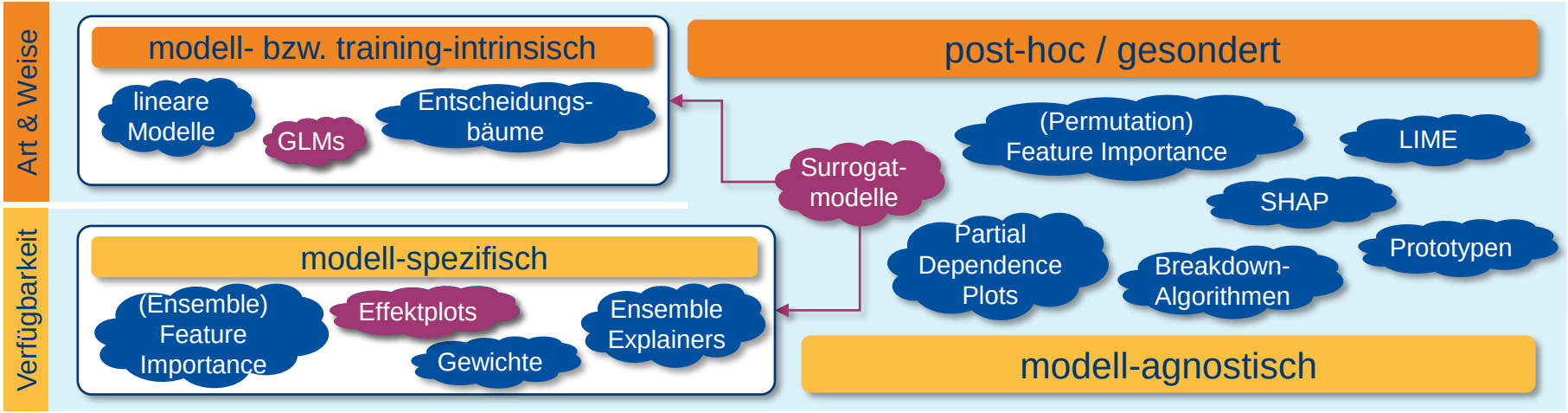
Kim, Been, Rajiv Khanna, and Oluwasanmi O. Koyejo. “Examples are not enough, learn to criticize! Criticism for interpretability.” Advances in Neural Information Processing Systems (2016).

➔ **Explainable AI** = Interpretierbarkeit eines gesamten Regelwerks auf Basis eines KI-Systems.

Das umfasst insb. die Extrahierung von Wissen aus einem maschinell erlernten KI-Modell über Zusammenhänge in den Daten und den daraus erkannten Mustern.

Quelle: Molnar, C. (2023). Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. https://christophm.github.io/interpretable ml book/

Verfahren für Explainable AI



- post-hoc Erklärverfahren, um ML-Modelle nachträglich zu interpretieren:
  - z.B. Feature Importances, Partial Dependence Plots
- eigenständige Anwendung **SV ADAPT**
  - automatisierte, datengetriebene GLM-Erstellung zur Risikomodellierung

DAV Deutsche Aktuarvereinigung e.V.

DGVFM Deutsche Gesellschaft für Versicherungs- und Finanzmathematik e.V.

1 Datengrundlage

2 Datenbasis

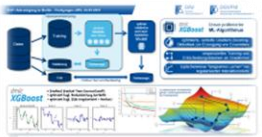
ML-Welt

3 Merkmalsauswahl

automatisierte ML-basierte Reduzierung des hochdimensionalen Raums

dmlc XGBoost

4 ML-Modellierung



Benchmark für Güte

5 Erklärungsverfahren

ML-basiertes GLM-Gewand

6 Surrogatmodell

adaptR

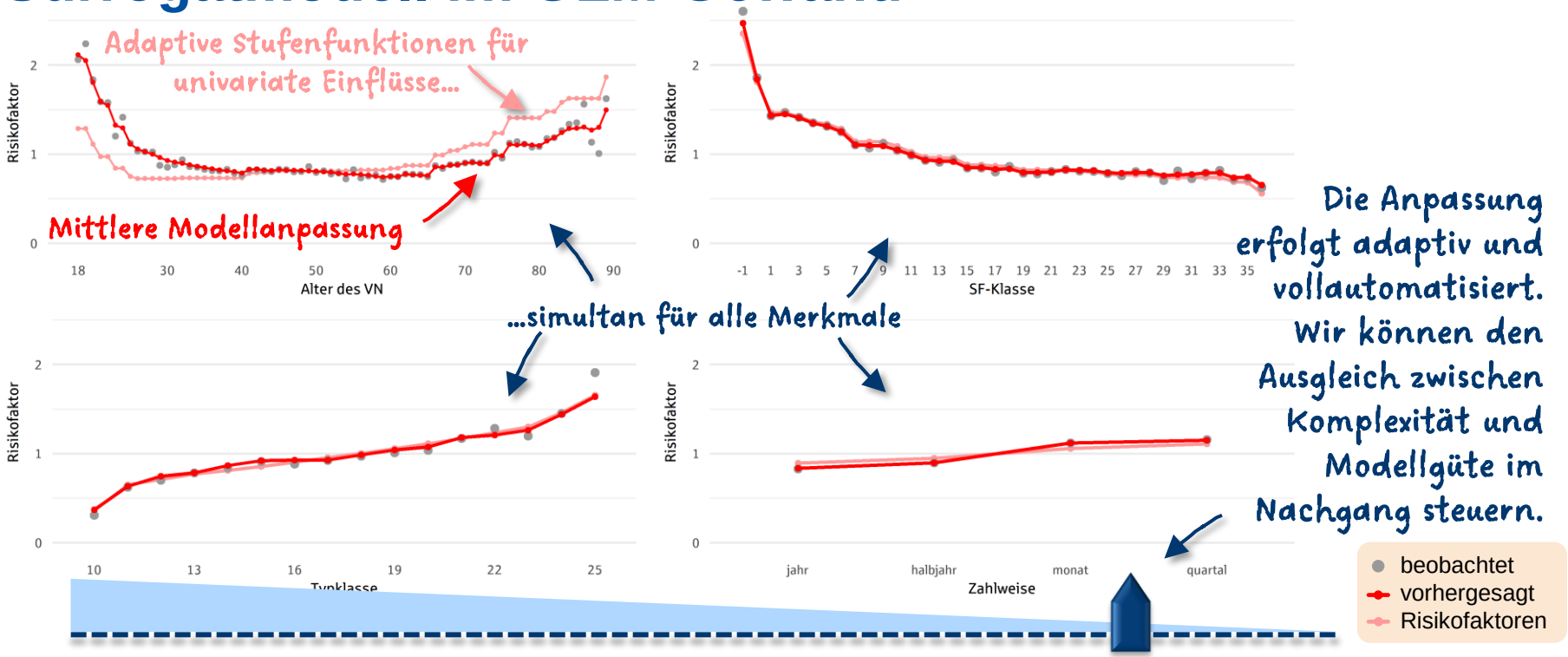
GLM-Welt

7 Bewertung/Vergleich

8 Verwendung/Erklärung

Dokumentation und Erfüllung notwendiger/gewünschter Anforderungen zu Transparenz

Surrogatmodell im GLM-Gewand



Regularisierte verallgemeinerte lineare Modelle

Verteilungsannahme $f(Y;\theta,\phi) = \exp\left(\frac{Y\theta - b(\theta)}{\phi} + c(Y;\phi)\right)$

Linkfunktion $g\colon \mathbb{R} \to \mathbb{R}$ sodass $\eta = g(\mu)$ mit $\mu = \mathbb{E}(Y)$

linearer Prädiktor $\eta = X\beta = \beta_0 + \beta_1 X_1 + \ldots + \beta_m X_m$

klassisch

Modellanpassung: Schätzung von β mit Maximum-Likelihood-Methode

$$\ell(\beta; y_i, x_i) = -\frac{1}{n}\sum_{i=1}^{n} \log f(y_i \mid x_i, \beta) \to \min$$

zusätzlich

Bestrafungsterm für zu hohe Komplexität: Vermeidung von Overfitting

$$\ell_{\text{pen}}(\beta; y_i, x_i, \lambda) = \ell(\beta; y_i, x_i) + \lambda\,\Omega(\beta_1, \ldots, \beta_m) \to \min$$

Shrinkage-Faktor $\lambda \geq 0$ zum Ausgleich zwischen
- Bestrafung $\Omega(\beta_1, \ldots, \beta_m)$ der Komplexität und
- Likelihood-Anpassung $\ell(\beta; y_i, x_i)$ an die Daten

Bestrafungsfunktion $\Omega\colon \mathbb{R}^m \to \mathbb{R}_+$ für die Parameter $\beta_1, \ldots, \beta_m$ (ohne Intercept)

technische Voraussetzung: Standardisierung der Merkmale x_i zur Vergleichbarkeit der β notwendig

DAV Deutsche Aktuarvereinigung e.V.

DGVFM Deutsche Gesellschaft für Versicherungs- und Finanzmathematik e.V.

Regularisierung für adaptive Mustererkennung

Idee

Regularisierung einer extrem großen Design-Matrix

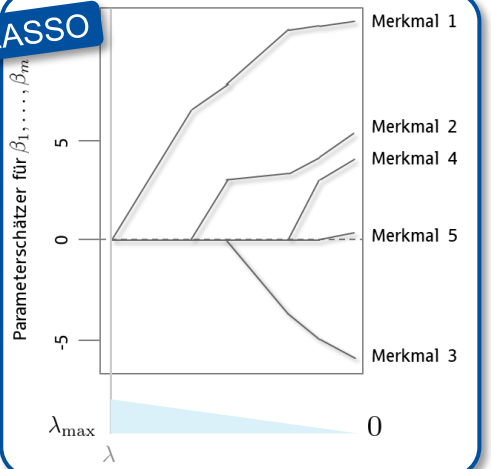
- geschickte Dummy-Codierung zum adaptiven Erlernen nicht-linearer Strukturen
- Haupteffekte und Interaktionen berücksichtigt
- Zu komplex?
  ➔ **Lass das mal das LASSO* machen!**

**Least Absolute Shrinkage and Selection Operator*

Alter (VN)	…	41	42	43	44	45	…
43-Jähriger	…	1	1	1	0	0	…
44-Jähriger	…	1	1	1	1	0	…

Beispiel: „Kontrast-Codierung“ für eine Stufen-Modellierung

LASSO



Vorteile

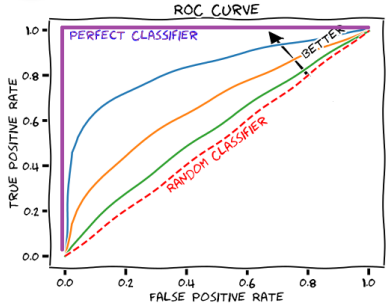
- LASSO-Verfahren ermöglicht **automatisierte Merkmalsauswahl**.
- Dummy-Codierung ermöglicht **automatisierte Effektauswahl** auch für **nicht-linearer Muster**.
- Am Ende dient **ein Hyperparameter** zum Fine-Tunen des Modells.
- Die Regularisierung sorgt für **hohe Robustheit** auch bei $p > n$.
- Über die Kontrast-Codierung ergibt sich **organisches Verhalten**.
  ➔ Einsatz z.B. in der Tarifierung
- Ergebnistyp ist ein **klassisches (erklärbares) GLM**.

Vortrag

https://www.ifa-ulm.de/fileadmin/user_upload/download/vortraege/2019_ifa_Hahn_Tarifierung-in-der-Schadenunfallversicherung-Ergaenzung-klassischer-Tarifierung-durch-moderne-Data-Analytics-Methoden_DAV-Jahrestagung.pdf

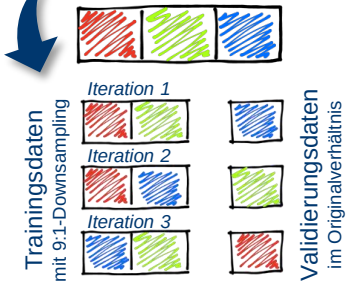
Zielgerichtetes Tuning des ML-Modells

AUC-Gütemaß

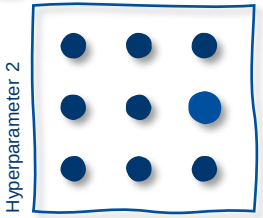


5-fache Kreuzvalidierung

Ausgangsdaten für das Tuning
im Originalverhältnis



Gittersuche über 4 Hyperparameter

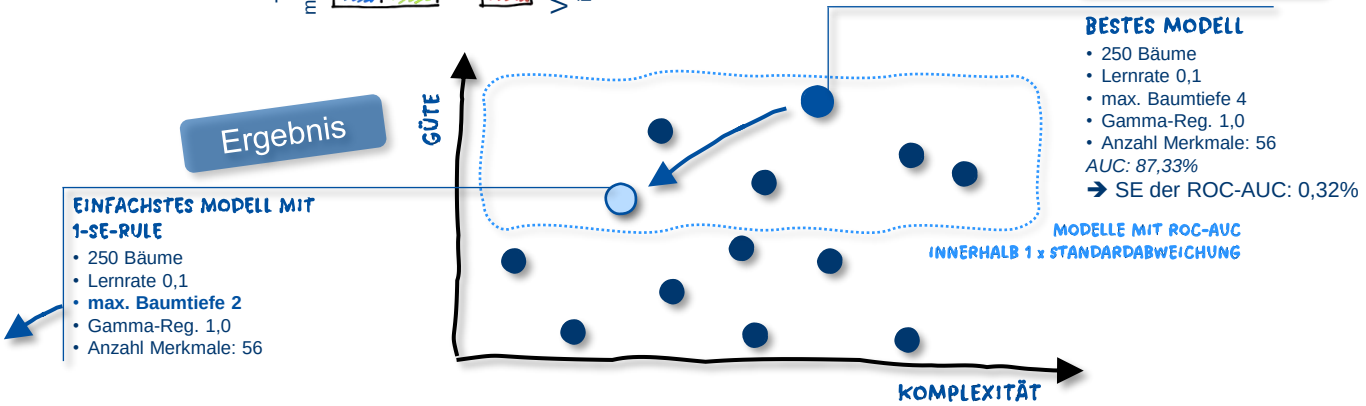


dmlc **XGBoost**

Intensive Gittersuche für
- Anzahl der Bäume
- Lernrate
- maximale Baumtiefe
- Gamma-Regularisierung

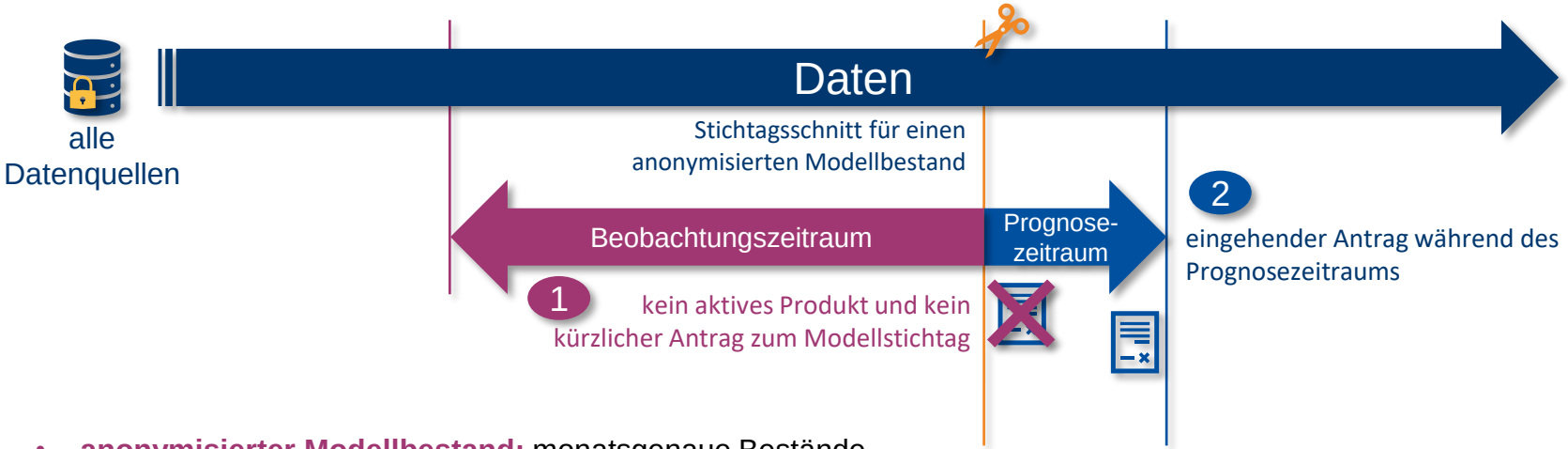
Weitere Parameter geprüft und bei Default-Einstellungen belassen.

Training



Erkenntnis:
Welche **Interaktionseffekte** sind **ausreichend?**
Zur Surrogatmodellierung genügen hier **paarweise Interaktionen.**

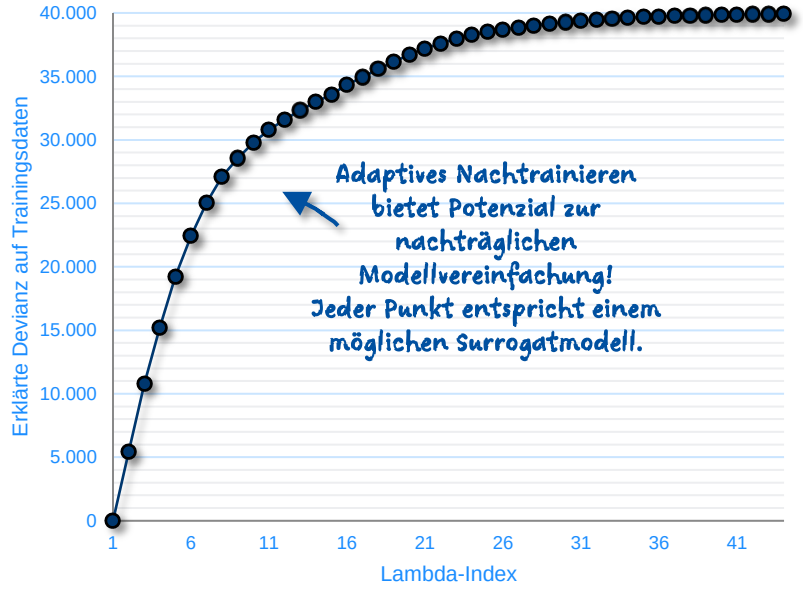
Fallstudie: Abschluss von VGV im Bestand



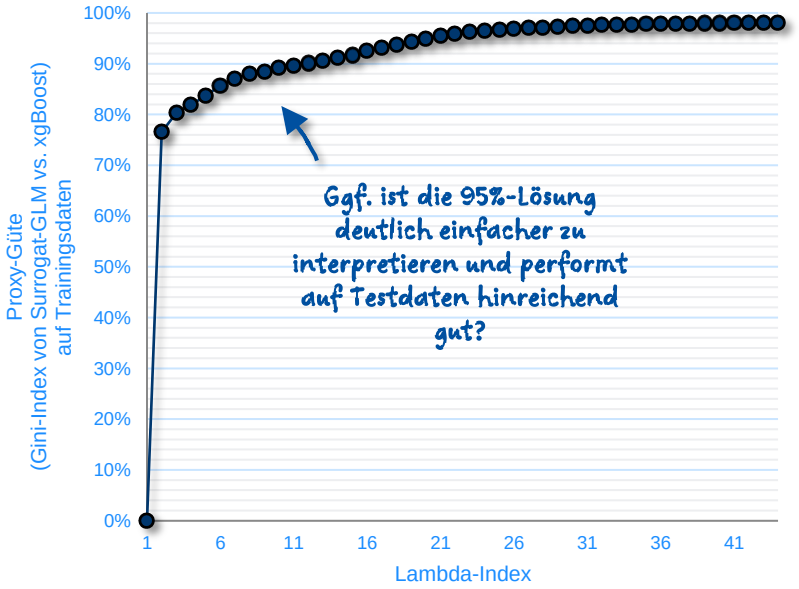
- **anonymisierter Modellbestand:** monatsgenaue Bestände
  - erklärende Merkmale zum Stichtag (stichtagsgenaue Informationen und Vorvergangenheit)
  - 2 Jahre zum Training (70%) und beiseitegelegte Testdaten (30%), 1 Jahr zum Backtesting
- **binäres Klassifikationsproblem:** Zielgröße als Abschluss im jeweiligen Folgemonat
  - Downsampling (9:1) auf Trainingsdaten

Fallstudie: Anpassung des Surrogatmodells

Erklärte Devianz ggü. ML-Vorhersage



Gini-Index ggü. ML-Vorhersage



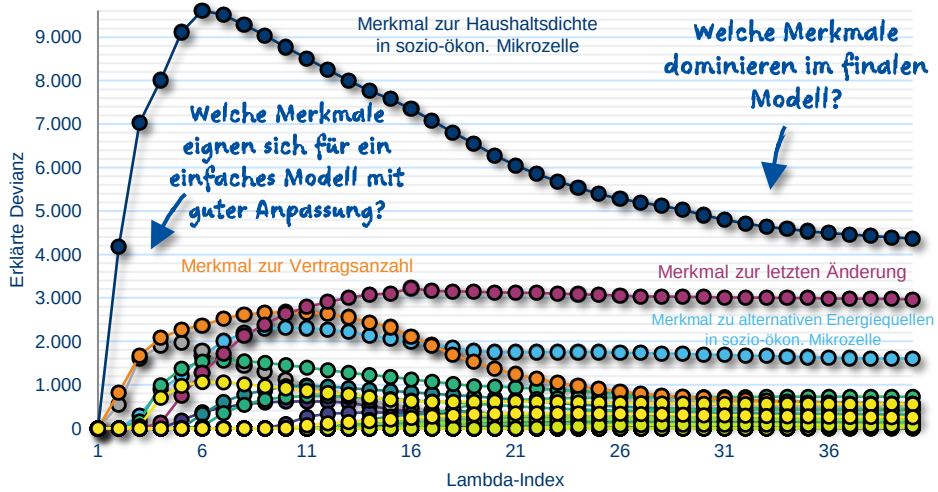
Fallstudie: Bewertung und Vergleich

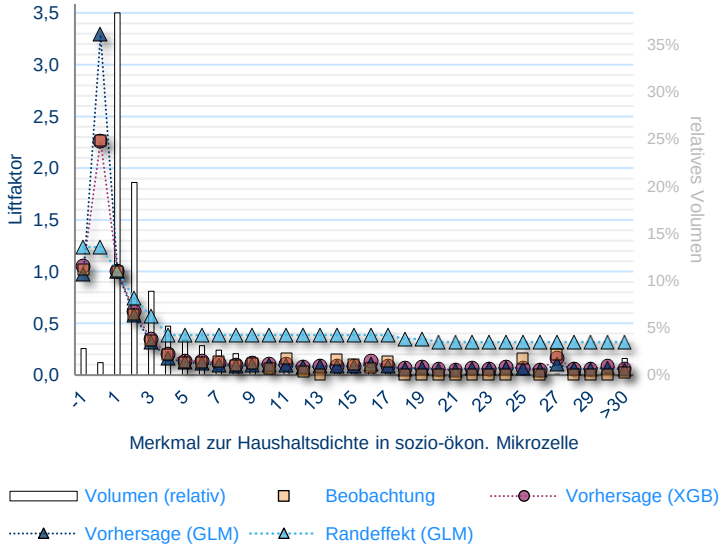
Anzahl Merkmale durch ML-Auswahlroutine	Abweichung ROC-AUC zwischen xgBoost & Surrogat-GLM*			Proxy-Güte (Gini-Index von Surrogat-GLM vs. xgBoost) auf Trainingsdaten
	Trainingsdaten	Testdaten	Backtesting	
56	-2,5%P	-0,6%P	-0,4%P	98,1%

Beispiel: nicht real, aber realistisch

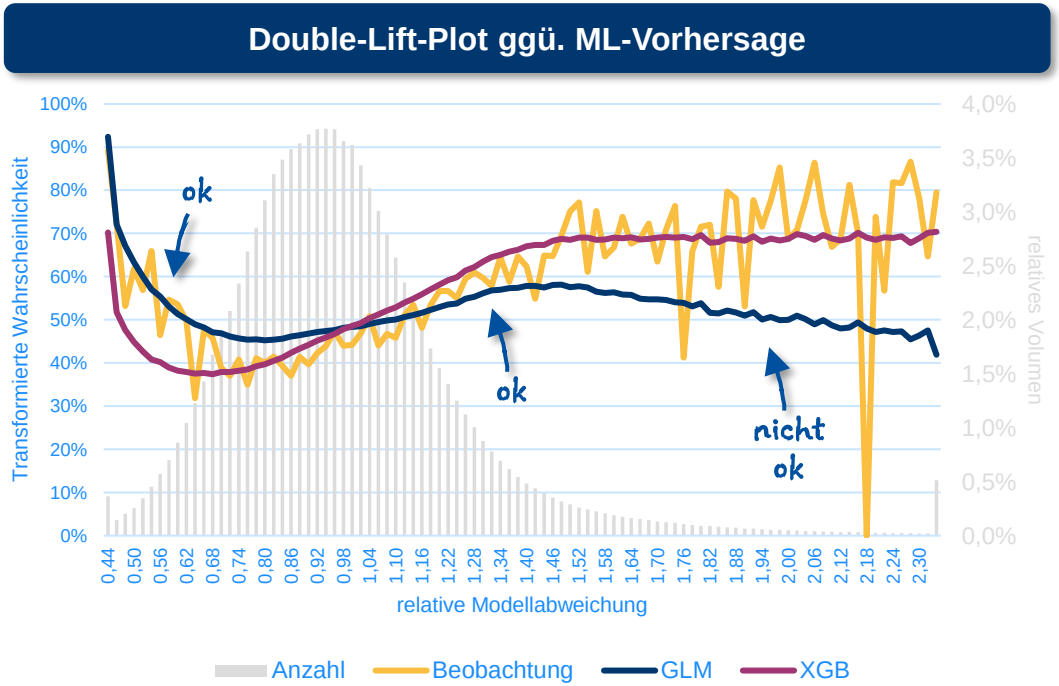
*Vergleiche Standardabweichung der ROC-AUC des xgBoost-Modells von 0,3%P (gemäß Tuning der zugehörigen Hyperparametrisierung). Für ein Surrogatmodell bestehend aus dem GLM und Anwendung eines nachgelagerten xgBoost-Modells mit 10 Bäumen der Tiefe 2 (= 10 Interaktionsterme) trainiert auf den vorläufigen Residuen kann das Backtesting um 0,1%P verbessert werden.

Merkmalsauswahl während der adaptiven Anpassung in SV ADAPT





Fallstudie: Analyse der Abweichungen



Fallstudie: Verwendung des Surrogatmodells

Das trainierte Surrogat-Modell…

…ist ein GLM.

➔ direkte Anwendung zum effizienten Scoring anstelle des ML-Modells:

Scoring: einfaches Ausrechnen der Regressionsgleichung

Produktive Anwendung

…ist ein GLM.

➔ unmittelbare, einfache und eindeutige Interpretation der Haupt- und Interaktionseffekte:

Interpretation: Wir sehen *exakt*, was das Modell tut.

Modellinterpretation

…ist ein GLM.

➔ unmittelbare Zerlegung eines Scores in seine additive/multiplikative Faktoren je Merkmal:

Erklärung: konsistent zu global interpretierbaren Effekten

Erklärung eines Scores

Dokumentation und Erfüllung notwendiger/gewünschter Anforderungen zu Transparenz

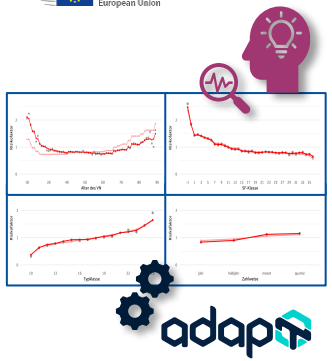
Fazit reloaded

- explizite (neue?) Anforderung an Erklärbarkeit von KI-Anwendungen
- Abbildung über Post-Hoc-Erklärung i.A. nur approximativ
- Alternative: Surrogat-Modelle mit modell-intrinsischer Erklärbarkeit
- SV: Eigenentwicklung für automatisierte datengetriebene GLM-Erstellung
- Fallstudie: Kombination beider Welten funktioniert

➔ u.a. höhere Anforderungen an Erklärbarkeit

➔ Erfüllung von Anforderungen unklar

➔ historisch hohe Akzeptanz von GLMs in Versicherungen (inkl. Regulator)



➔ neuer Anwendungsfall zur Erstellung erklärbarer Surrogat-Modelle für ML-Modelle

➔ *Teilung:* Mustererkennung + Erklärbarkeit
Güte: kein signifikanter Verlust
Prozess: zu großen Teilen automatisiert

DAV/DGVFM
Jahrestagung
2024

Dr. Lukas Hahn
Lead Data Scientist, SV SparkassenVersicherung Holding AG
Tel: 0711 898-43765
Email: lukas.hahn@sparkassenversicherung.de
https://de.linkedin.com/in/lukas-hahn-73a3659a