

Data Science Challenge 2022

Auto-ML und XAI

Dr. Felix Spangenberg (Deloitte), Andreas Zimmer (msg),
Yangyang Zhou (msg), Dirk Wegmann (msg),
Christian Brunner (Allianz)



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.



DGVFM

DEUTSCHE GESELLSCHAFT
FÜR VERSICHERUNGS- UND
FINANZMATHEMATIK e.V.

DAV/DGVFM-Jahrestagung, 26.–28. April 2023



Das Team

- Dr. Felix Spangenberg, Aktuar DAV und CADS
- Andreas Zimmer, Student der TH Rosenheim
- Yangyang Zhou, Aktuarin in Ausbildung
- Dirk Wegmann, Aktuar in Ausbildung
- Christian Brunner, Aktuar DAV

Während der Challenge waren alle Teilnehmer bei msg systems AG beschäftigt. Wir danken msg für die Unterstützung!





Das Team





Aufgabenstellung DAV Challenge 2022

- Bearbeitung einer aktuariellen Fragestellung mit Data-Science-Methoden
- Einsatz von Auto-ML
- Einsatz von XAI

Auswahl Datensatz und Fragestellung

- Datensatz: French MTPL
- Aufgabe: Vorhersage der Schadenfrequenz

- Datensatz stammt aus „Computational Actuarial Science with R“
- 677.991 Zeilen, 12 Spalten
- Anzahl Schäden und übliche Merkmale wie Alter, Region etc.
- Herkunft ist laut Beschreibung „unknown private insurer“
- Mehrere Tutorials zu dem Datensatz verfügbar, u.a.
 - <https://www.actuarialdatascience.org/ADS-Tutorials/> (SAV)
 - Friedrich Loser:
https://github.com/DeutscheAktuarvereinigung/claim_frequency



Datensatz (schon transformiert)

	ClaimNb	Area	VehPower	VehAge	DrivAge	Bonus-Malus	VehBrand	VehGas	Density	Region
0	1	4	5.0	1	6	50	B12	Regular	7.10	R82
1	1	4	5.0	1	6	50	B12	Regular	7.10	R82
2	1	2	6.0	2	6	50	B12	Diesel	3.99	R22
3	1	2	7.0	1	5	50	B12	Diesel	4.33	R72
4	1	2	7.0	1	5	50	B12	Diesel	4.33	R72
...
678012	0	2	7.0	2	3	54	B12	Diesel	4.17	R72

Poisson-Devianz

Für das Training der Modelle ist die Poisson-Devianz die richtige Metrik:

$$L(y, \hat{y}) = 2 \cdot \sum_{i=1}^n (y_i \cdot \log\left(\frac{y_i}{\hat{y}_i}\right) - (y_i - \hat{y}_i))$$

mit $0 := 0 \cdot \log 0$.

Vergleiche mit der quadratischen Metrik:

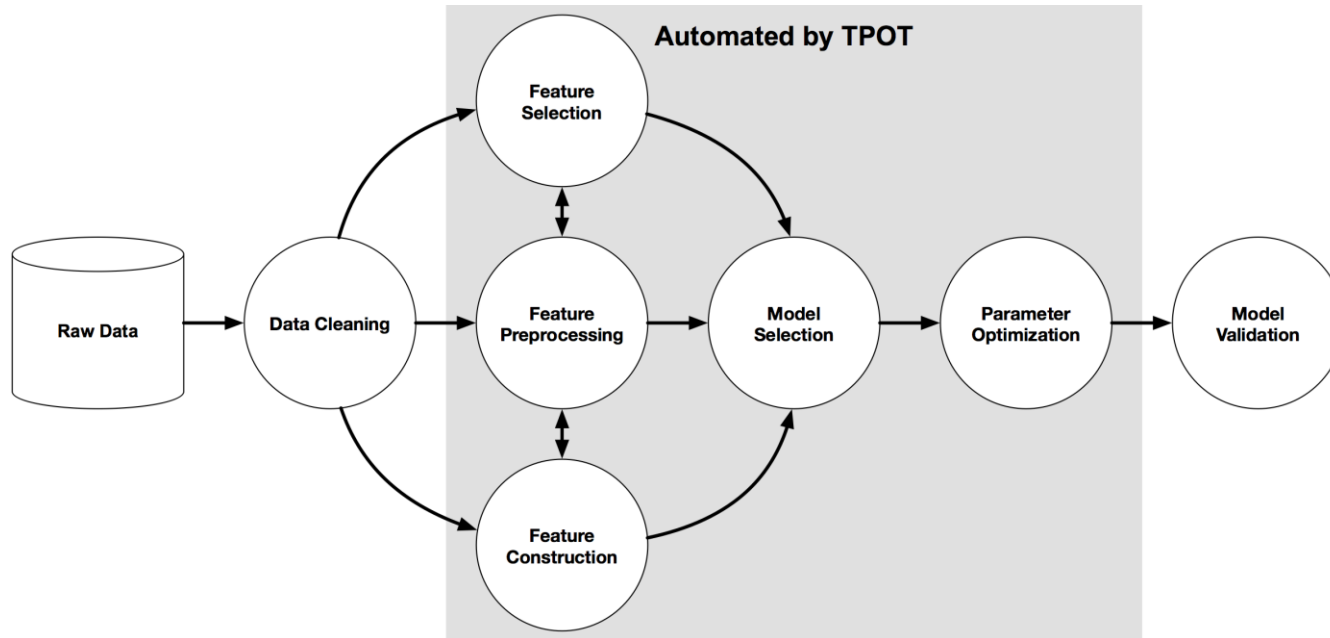
$$L(y, \hat{y}) = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Training der Modelle

Wir haben mit dem Datensatz verschiedene ML-Modelle mit Scikit-Learn und Tensorflow in Anlehnung an das Tutorial der SAV trainiert.

Modell	Train	Test
Dummy	0.6235	0.6304
GLM	0.6005	0.6094
TREE1	0.5967	0.6053
TREE2	0.5901	0.5980
Random Forest	0.5971	0.6067
Neuronales Netz	0.5931	0.6039
GBM	0.5752	0.5917

Auto-ML: Automatisiertes Machine-Learning



Grafik aus <https://github.com/EpistasisLab/tpot>



Auswahl Auto-ML-Tools

- Auswahl unterschiedlicher Verfahren für das Parametertuning
- CFO, Genetic Programming, Bayes Optimierung, Random Grid Search
- Scikit-Learn-API

Die Auswahl fiel auf:

- TPOT
- Auto-Sklearn
- H2O
- FLAML

Ergebnisse Auto-ML

Die Ergebnisse waren für die verschiedenen Tools:

Modell	Train	Test
TPOT	0.6369	0.7103
Auto-Sklearn	0.7085	0.7093
H2O	0.4918	0.5971
FLAML	0.5457	0.6072

Zum Vergleich:

Modell	Train	Test
Dummy	0.6235	0.6304
GBM	0.5752	0.5917

Probleme beim Auto-ML

- keine Unterstützung der Poisson-Devianz
- keine Unterstützung einer Gewichtung
- „Null-Vorhersagen“
- Trainierte Modelle zeigen Overfitting.
- Modelle schlechter als manuell trainierte Modelle
- Modelle teilweise sogar schlechter als die Baseline

Fazit:

- Auto-ML ist aktuell nur für Standard-ML-Fälle geeignet.
- Für spezielle aktuarielle Anwendungen sind die Tools nicht geeignet.
- Auch Auto-ML erfordert viel Arbeit.

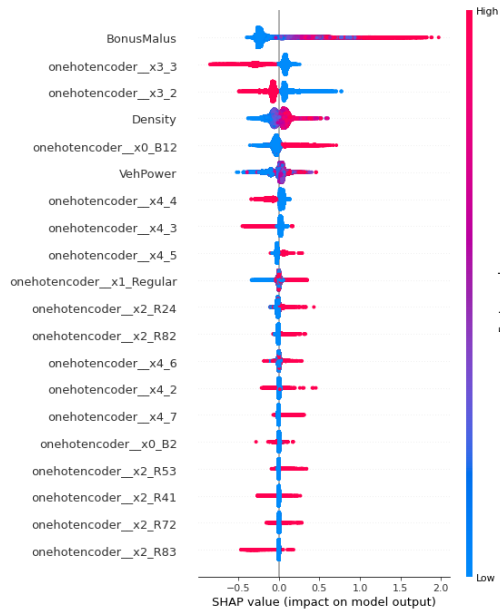


XAI (Explainable Artificial Intelligence)

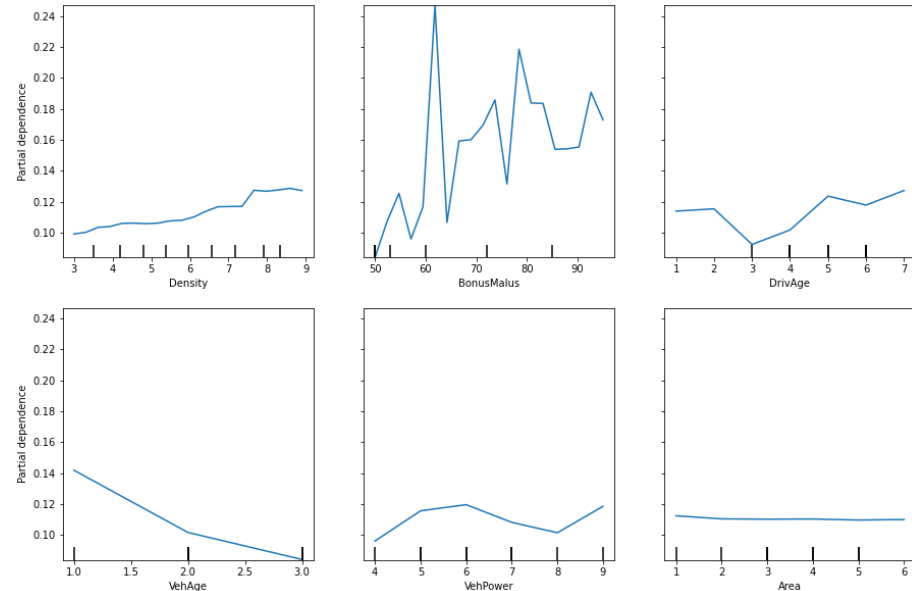
- ML-Modelle wie GBM und Neuronale Netze nicht verständlich
- Idee: ML-Modelle erklärbar machen
- Christoph Molnar: <https://christophm.github.io/interpretable-ml-book/>
- AG der DAV: Explainable Artificial Intelligence

XAI (Explainable Artificial Intelligence)

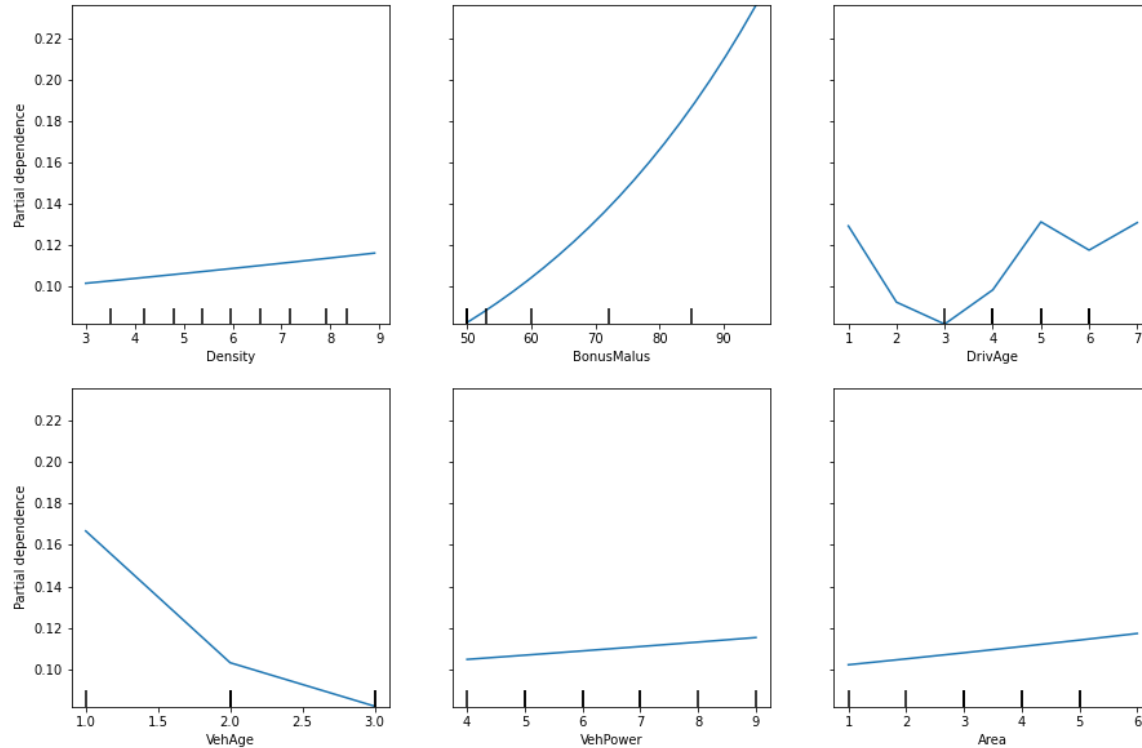
lokale Methoden (z.B. SHAP)



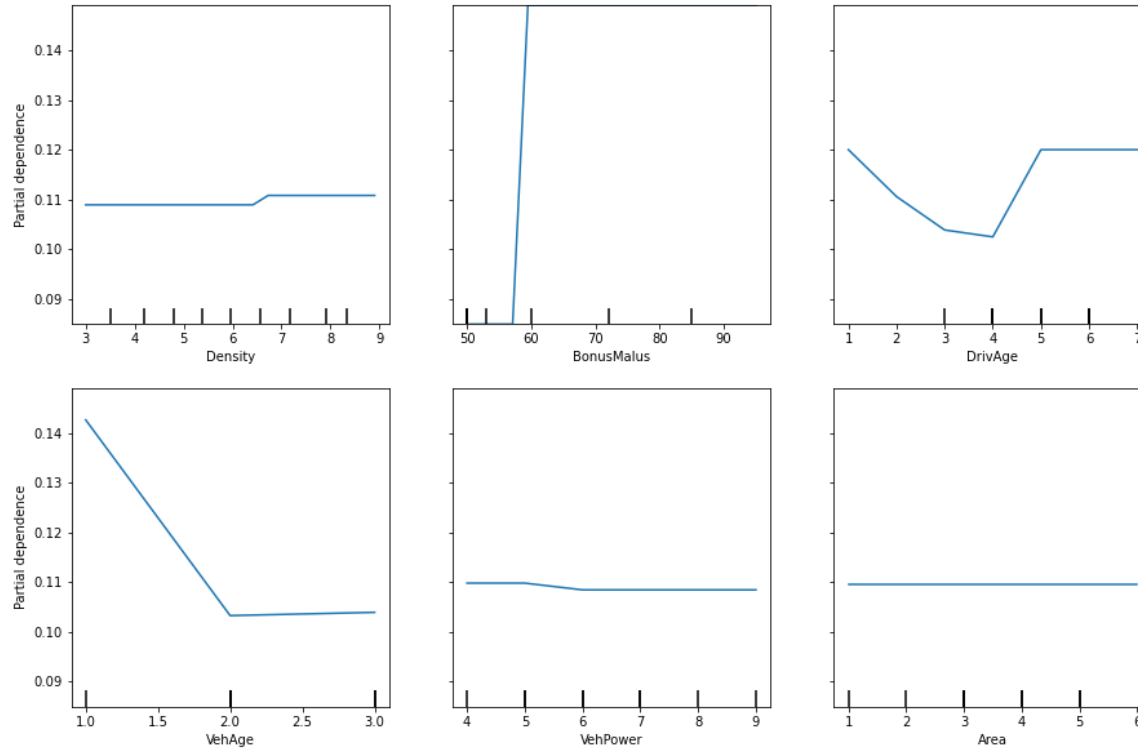
globale Methoden (z.B. PDP)



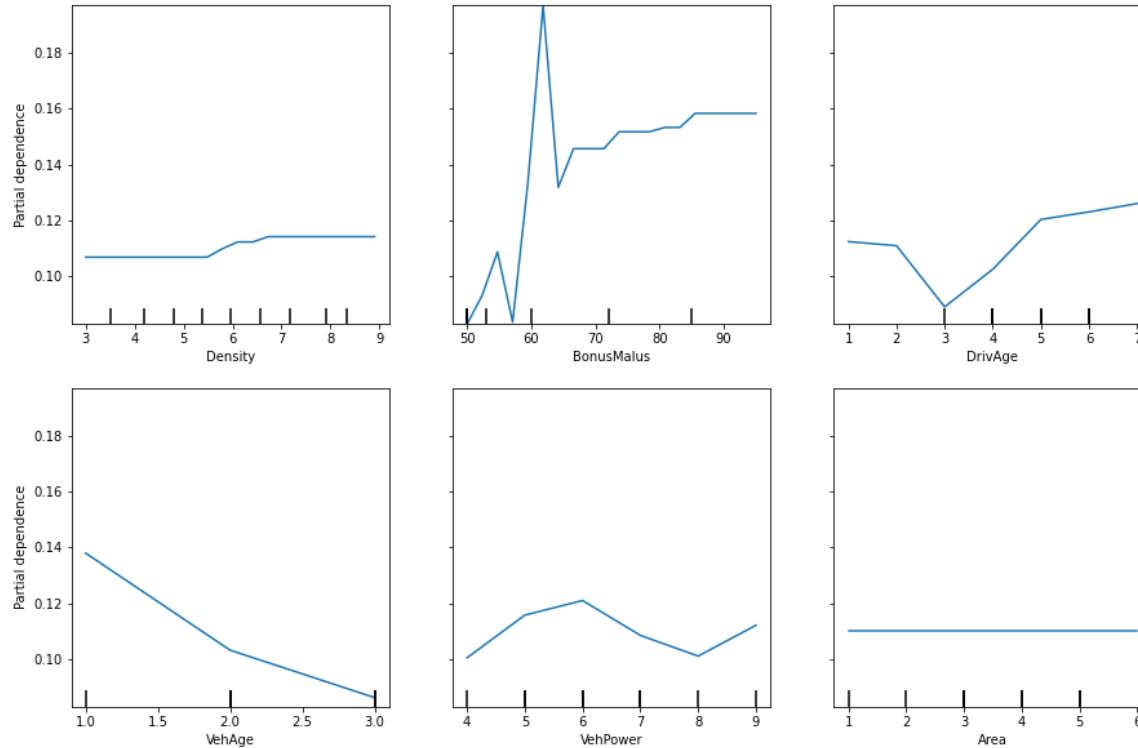
XAI: Partial Dependence Plots GLM



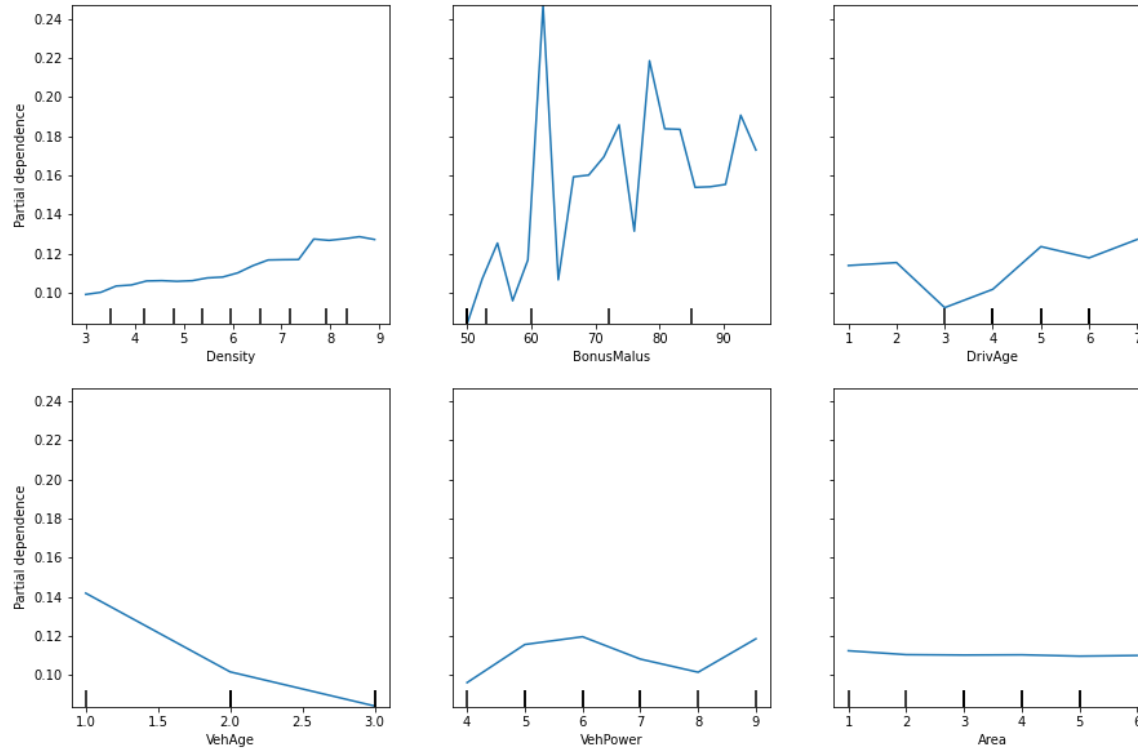
XAI: Partial Dependence Plots Tree1



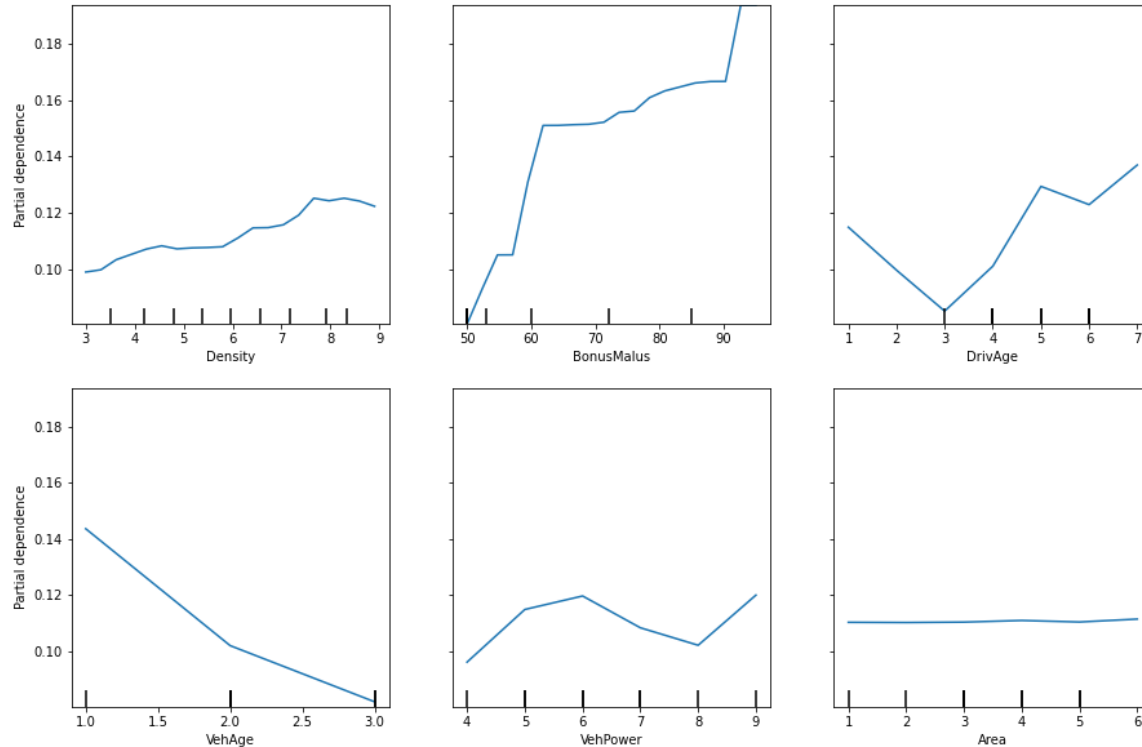
XAI: Partial Dependence Plots Tree2



XAI: Partial Dependence Plots GBM



XAI: Partial Dependence Plots GBM monotonisiert





Erneute Bewertung der Modelle nach XAI

- Die beiden Bäume und die GBM zeigen ein unplausibles Verhalten beim Bonus-Malus.
- Das GBM schien das beste Modell zu sein.
- Durch Einsatz von XAI ist klar, dass das GBM doch nicht geeignet ist!
- Bonus-Malus muss genauer untersucht werden!
- Aktuarielles Wissen ist bislang gar nicht in die Modelle eingeflossen!



Bonus-malus: comment ça marche

- In Frankreich wird eine Prämie ohne Bonus-Malus berechnet.
- Prämie wird anschließend mit dem Bonus-Malus multipliziert.
- Der Bonus-Malus darf daher nicht in die Modelle einfließen.
- Gewisse niedrige Bonus-Malus-Klassen werden paradoxerweise nur durch Schäden erreicht. Diese Zusammenhängen lernen die Modelle.

Zur Mechanik des französischen Bonus-Malus-Systems siehe:

<https://www.allianz.fr/assurance-particulier/vehicules/assurance-auto/conseils-pratiques/bonus-malus.html>



Zusammenfassung

- Manuelles Training weiterhin notwendig
- Auto-ML ist aktuell für spezielle aktuarielle Zwecke ungeeignet.
- Einsatz von XAI ist sehr sinnvoll.
- Aktuarielles Wissen sollte früh in den ML-Prozess einfließen.

Herzlichen Dank!

Dr. Felix Spangenberg
Andreas Zimmer
Yangyang Zhou
Dirk Wegmann
Christian Brunner

fspangenberg@deloitte.de
andreas.zimmer@msg.group
yangyang.zhou@msg.group
dirk.wegmann@msg.group
christian.brunner@allianz.de



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.



DGVFM

DEUTSCHE GESELLSCHAFT
FÜR VERSICHERUNGS- UND
FINANZMATHEMATIK e.V.

DAV/DGVFM-Jahrestagung, 26.–28. April 2023

ML-Modelle

Modell	Bechreibung	Güte	Rob.	Ex.		Weitere Maßnahme
Dummy	gewichtetes arithmetisches Mittel	0	+++	+++	nein	
GLM	GLM ohne Interaktionen	+	++	++	ja	Interaktionen, statistische Analyse
TREE1	20 Blätter, 10.000 Blattgröße	++	+	+	ja	Baumstruktur betrachten
TREE2	40 Blätter, 1.000 Blattgröße	++	0	+	ja	Baumstruktur betrachten
RF	Random Forest analog TREE1	++	-	-	nein	
NN	Neuronales Netz mit einer Hidden Layer	++	-	-	ja	XAI, Abhängigkeit vom Seed untersuchen
GBM	Gradient Boosting Machine	+++	-	-	ja	XAI, Robustheit erhöhen, Overfitting untersuchen



Auto-ML-Tools

Framework	Optimierung	Poisson-Devianz	Gewichte	Automatische Datenaufbereitung	Vorteile	Nachteile	Brauchbares Modell	Scikit-Learn-API
TPOT	Evolution	nein	mit etwas tricksen	ja	Export des Codes	lange Laufzeit	nein ohne Poisson-Devianz	ja
H2O	Random Grid Search	ja	ja	ja	XAI-Tools integriert	in unserem Fall starkes Overfitting, Datenformat	nein da Overfitting	nicht wirklich
AUTO-SKLEARN	Bayes	nein	nein	ja	wenig Transformationen notwendig	benötigt Linux oder Docker	nein ohne Gewichte	ja
FLAML	CFO	nein	ja	nein	schnelle Laufzeit	nur Baummodelle	nein ohne Poisson-Devianz	ja



Bewertung der Ergebnisse der Auto-ML-Tools

Framework	Modell	Beurteilung	sinnvolles Modell?
TPOT	Lassoregression + RF mit wenig sinnvollen Parametern	Underfitting, Training ohne Poisson-Devianz	nein
AUTO-SKLEARN	HistGradientBoostingRegressor	Underfitting, Training ohne Gewichte und ohne Poisson-Devianz	ja
H2O	Random Forest mit Tiefe 20 und 5.000-10.000 Blättern	Overfitting	ja, aber viel zu tiefe Bäume
FLAML	ExtraTreesRegressor oder XGBRegressor	Training ohne Poisson-Devianz, Overfitting	ja, aber merkwürdige Parameter