Zeitschrift für die gesamte Versicherungswissenschaft, 112 (2023), 3 – 32
Duncker & Humblot, 12165 Berlin

Erklärbare Künstliche Intelligenz am Beispiel von Ratings deutscher Lebensversicherungsunternehmen

Holger Bartel, *Mirko Kraft* und *Jochen L. Leidner*

Zusammenfassung

Künstliche Intelligenz (KI) wird in der Praxis zur Entscheidungsfindung eingesetzt (Lossos, Geschwill und Morelli 2021), zunehmend auch in der Versicherungsbranche. Dies erfordert jedoch Vertrauen in die verschiedenen KI-Methoden, speziell auch bei der Bewertung von Unternehmen („Ratings"). Solches Vertrauen bildet sich, wenn Entscheidungsträger und Nutzer mentale Modelle des Systems bilden können und sie die Ausgabe des Systems verstehen. KI muss also erklärbar sein, eine reine Black Box ist selbst bei hoher Qualität eines Systems unzureichend. Die „erklärbare KI" (engl. „eXplainable Artificial Intelligence", XAI) befasst sich mit der Entwicklung von KI-Modellen, die durch Menschen nachvollziehbar sind (Adadi und Berrada 2018; Europäische Kommission 2020). In diesem Beitrag werden wünschenswerte Eigenschaften industrieller KI-Systeme untersucht – speziell hinsichtlich der Erklärbarkeit – und am Anwendungsbeispiel von Ratings (deutscher) Lebensversicherungsunternehmen vorgestellt und auch visualisiert. Neben XAI als Aspekt der technischen Akzeptanz wird die Interaktion zwischen Geschäftsmodell und der kundenseitigen Akzeptanz bei Ratings von deutschen Lebensversicherungsunternehmen beleuchtet. Geschäftszahlen deutscher Lebensversicherungsunternehmen werden oft als intransparent erachtet. Dies gilt auch, wenn die HGB-Rechnungslegung um die Solvency II-Berichte zur Solvenz- und Finanzlage (SFCR) ergänzt betrachtet werden. Insofern ist die Auseinandersetzung mit erklärbaren KI-Methoden in diesem Kontext ein sinnvoller Beitrag für die Bewertungspraxis.

Prof. Dr. Mirko Kraft,
Hochschule Coburg, Fakultät Wirtschaftswissenschaften, Friedrich-Streib-Straße 2,
96450 Coburg, Deutschland
E-Mail: mirko.kraft@hs-coburg.de

Dr. Holger Bartel, RealRate GmbH
E-Mail: holger.bartel@realrate.ai

Prof. Dr. Jochen L. Leidner,
Hochschule Coburg, Fakultät Wirtschaftswissenschaften und University of Sheffield,
Department of Computer Science
E-Mail: leidner@acm.org

Abstract

Artificial intelligence (AI) is already used for decision-making in practice (Lossos/Geschwill/Morelli 2021), increasingly also in the insurance sector. However, it requires trust in the various AI methods, especially in the evaluation of companies („ratings"). Trust is formed when decision makers and users can form mental models of a system and they understand its output. AI must therefore be explainable; a pure black box is insufficient even if a system is of high quality. „Explainable AI" (eXplainable Artificial Intelligence, XAI) is concerned with the development of AI models that are comprehensible by humans (Adadi/Berrada 2018; European Commission 2020). In this paper, desirable properties of industrial AI systems are investigated – specifically with respect to explainability – and presented and visualized using the application example of ratings of German life insurance companies. In addition to XAI as one prerequisite for technical acceptance, the interaction between the business model and customer acceptance of ratings of German life insurance companies is examined. Financial key performance indicators for German life insurance companies are often said to lack transparency; this is still the case when HGB accounting is supplemented by the Solvency and Financial Condition Reports (SFCR) according to Solvency II. We argue that the examination of explainable AI methods is a useful contribution to the practice of valuation.

Stichworte: erklärbare Künstliche Intelligenz (KI), Ratings, Lebensversicherungsunternehmen

Key words: explainable artificial intelligence (XAI), ratings, life insurance companies

1. Einleitung

1.1 Motivation

Die Anwendung von Künstlicher Intelligenz (KI) ist in vielen Bereichen und Unternehmen nicht mehr nur Zukunft, sondern bereits Realität – so auch in der Versicherungsbranche (siehe z. B. Oletzky/Reinhardt 2022, Kurmann 2023). An den Einsatz von KI-Methoden werden teilweise große Erwartungen geknüpft, insbesondere werden z. B. Vorteile durch die Automatisierung von Prozessen erhofft. Bei Bewertungs- und Entscheidungsprozessen, die sonst Menschen vorbehalten waren, besteht aber die Sorge, dass die Ergebnisse nicht mehr in dem Maße nachvollziehbar sind. Das gilt insbesondere für Ratings von Unternehmen („Unternehmensratings"), die die finanzielle Solidität zwischen verschiedenen Unternehmen vergleichbar machen sollen.

Durch Künstliche Intelligenz (KI) automatisch generierte Bewertungen dienen Marktakteuren dazu, ihre Entscheidungen zu treffen. Daher wollen sie KI-Systeme nachvollziehen können (Samek/Müller 2019, S. 8). Erklärbarkeit als ein ethischer Grundsatz von KI (HEG-KI 2018, S. 16) ist demnach Voraussetzung für Transparenz, und insbesondere auch für auf Basis von KI-Methoden erstellte Unternehmensratings. Allgemein werden an Unternehmensratings hohe Anfor-

derungen gestellt, u. a. auch regulatorische Anforderungen, um den Einsatz von externen Ratings zu rechtfertigen (zur Rating-Regulierung siehe z. B. Europäische Kommission 2021).

Speziell beim Einsatz von KI-Methoden ist aber das Vertrauen in die unterliegenden Prozesse wichtig, sodass in der Praxis wie auch in der Forschung vielfältige Anstrengungen unternommen werden, bei KI-Systemen[1] ausreichend Transparenz zu schaffen und die Erklärbarkeit („explainable AI", XAI) der Ergebnisse sicherzustellen. Neben mathematischen Verfahren, die beispielsweise Sensitivitäten der Ergebnisse gegenüber Veränderungen der Eingangsdaten analysieren und anderen post hoc-Verfahren („ex post-Erklärbarkeit"), ist ein Ansatz, schon beim Design der Algorithmen bzw. bei der Auswahl der KI-Methode auf die Erklärbarkeit zu achten („Erklärbarkeit per Design").

Um Erklärbarkeit bei KI-basierten Unternehmensratings zu erreichen, wird in einem Use Case ein solcher Ansatz gewählt – im Gegensatz zur nachträglichen Interpretation beliebiger KI-Systeme. Anders als in typischen KI-Anwendungen mit Big Data und großen neuronalen Netzen (sog. Deep Learning) werden hier relativ kleine, strukturelle Netze verwendet, die in Gleichungsform vorgegeben sind, und Expertenwissen repräsentieren. Dies garantiert die Erklärbarkeit anhand eines gerichteten Graphen, der die Ursachen und Wirkungen der relevanten Größen veranschaulicht (Bartel 2019). Die Nutzer benötigen zum Verständnis daher weniger fachliche und technische Expertise.

Konkret wird anhand des Beispiels des Ratings deutscher Lebensversicherer gezeigt, wie die Analyse eines komplexen Geschäftsberichts, insbesondere der Rechnungslegungsgrößen, erfolgt (Sellhorn 2020). Stärken und Schwächen der Unternehmen sind damit direkt im Marktvergleich über den Graphen erklärbar. Als Daten für das maschinelle Lernen werden auch die erweiterten aufsichtlichen risikobasierten Berichte gegenüber der Öffentlichkeit, die sog. Solvency and Financial Condition Reports (SFCR) verwendet, die für Versicherer seit 2016 verpflichtend sind (zu Säule 3 von Solvency II vgl. beispielsweise Gründl/Kraft 2019, Van Hulle 2019).

Auch untersucht wird, wie das Geschäftsmodell und die Unabhängigkeit des Raters zusammenhängen, indem die Szenarien „Beauftragung", das traditionelle Modell, behaftet mit Interessenkonflikten (Crumley 2012; Stuwe et al. 2012), und „Public Rating", d. h. ohne Beauftragung durch das bewerteten Unterneh-

[1] Unter einem KI-System kann verstanden werden (*Holland*/*Kavuri* 2021, S. 106): „A set of inter-related elements of AI algorithms, big data, digital infrastructure and Management Information Systems (MIS), and the business context that encompasses business processes, products, and the business model of the firm, within an ethical, regulatory, and legal environment."

mens, kontrastierend verglichen und auch gegenüber alternativen Ansätzen aus der Literatur abgegrenzt werden.

In dem folgenden Beitrag wird im Kern die Transparenz von Ratings deutscher Lebensversicherungsunternehmen durch die Anwendung erklärbarer KI vorgestellt. Transparenz und Erklärbarkeit sind generell wünschenswerte Kriterien für KI-Systeme (vgl. auch z. B. Oletzky/Reinhardt 2022, S. 505 f.). Expertenwissen und KI-Methoden werden hier kombiniert und in einem solchen „hybriden" Modell auch für die Visualisierung der Ergebnisse genutzt, die die Erklärbarkeit auch für die Anwender der Unternehmensratings operationalisiert.

1.2 Überblick

Der Beitrag ist wie folgt aufgebaut: Nach Einführung in die Erklärbarkeit von Künstlicher Intelligenz (2.1) und einer einführenden Verknüpfung von Ratings und Künstlicher Intelligenz (2.2) folgt eine Fallstudie: Die Anwendung von KI-Methoden für Unternehmensratings speziell für deutsche Lebensversicherungsunternehmen (3.). Der Anwendungsfall wird dann im Hinblick auf Transparenz und Geschäftsmodell näher analysiert (4.). Der Beitrag schließt mit einer Zusammenfassung und einem Ausblick auf mögliche Folgeforschungsprojekte (5.).

2. Erklärbare Künstliche Intelligenz

2.1 Ansätze der Erklärbarkeit

Erklärbare Modelltypen des maschinellen Lernens lassen sich unterscheiden zum einen in *„White Box"-Modelle* (auch „ex ante"-Modelle, da die Erklärbarkeit von vornherein gegeben ist), die von Natur aus erklärbar sind und daher keiner speziellen Diskussion bedürfen. Beispiele davon sind lineare Modelle (z. B. lineare oder logistische Regression (James et al. 2017)), Entscheidungsbäume (CART, Breiman et al. 1984), ID3 (Quinlan 1986) sowie Regelsysteme (z. B. Repeated Incremental Pruning to Produce Error Reduction (RIPPER, Cohen 1995)). Zum anderen gibt es *„Black Box"-Modelle*. Deren Erklärbarkeit kann erst „ex post" durch „Aufpfropfen" zusätzlicher Mechanismen, wenn überhaupt, ansatzweise erreicht werden. Beispiele sind „tiefe", also mit mehreren Nicht-Ein-/Ausgabe-Schichten ausgestattete, neuronale Netze (sog. Deep Learning).

Erklärbares maschinelles Lernen ist nun insbesondere damit befasst, auch für solche Black Box-Modelle Erklärungen zu generieren (Adadi/Berrada 2018). Dies ist insbesondere wünschenswert, weil Black Box-Modelle derzeit die besten Vorhersagen liefern (siehe Abb. 1). Die *lokale* Daten-Erklärbarkeit hat

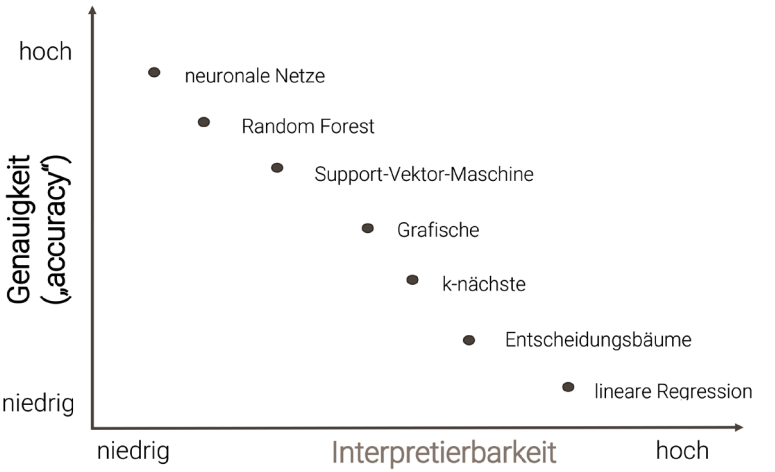

Abb. 1: Trade-off zwischen Genauigkeit („accuracy") und Interpretierbarkeit[2]

zum Ziel zu erklären, weshalb eine bestimmte Eingabe x zu einer bestimmten Ausgabe y führt, während die *globale* Modell-Erklärbarkeit zum Ziel hat zu erklären, wie ein bestimmtes Modell als Ganzes funktioniert. Erklärbarkeit *per Design* versucht, nur solche Verfahren zu konstruieren, die *ex ante* erklärbar sind. Hybride künstliche Intelligenz (engl. „hybrid AI") ist die Kombination symbolischer und konnektionistischer Methoden wie neuronaler Netze (Wermter/Sun 2000), die durch Beibehaltung symbolischer Aspekte in einem Modell die Erklärbarkeit steigern.

Ein Ansatz der *ex post*-Erklärbarkeit besteht im Induzieren von sekundären Stellvertreter- oder Surrogat-Modellen zusätzlich zu einem gegebenen primären Black Box-Modell; das Surrogat-Modell zielt ausschließlich auf die Generierung von Erklärungen im Nachhinein ab, während das primäre Modell die eigentliche Vorhersage bestimmt. Relevant ist hier die Wiedergabetreue (engl. „fidelity"): Wie stark stimmen die Vorhersagen von Black Box-Modell und Surrogat-Modell überein? Ein einfacher Ansatz besteht in der Induktion eines Entscheidungsbaums auf Ausgaben eines neuronalen Netzes, die als Gold-Standard verwendet werden. Ein solches einfaches Verfahren resultiert leider nur in geringer Aussagekraft, Wiedergabetreue und Genauigkeit, kann aber durch Anwendung von Regularisierung verbessert werden (Burkart/Huber 2021). Während Regularisierung üblicherweise der Verbesserung der Generalisierung dient, dient sie hierbei stattdessen der Verbesserung der Erklärbarkeit.

[2] Quelle: *Dziugaite* et al. 2020.

Eine der frühesten Schlüssel-Arbeiten im XAI-Umfeld ist Ribeiro, Singh und Guestins Arbeit „Why Should I Trust You? Explaining the Predictions of Any Classifier", das die LIME-Technik einführte (kurz für „Local Interpretable Model-agnostic Explanations", Ribeiro/Singh/Guestin 2016). Die Autoren induzieren einen White Box-Klassifikator zu den Vorhersagen eines gegebenen Black Box-Klassifikators wie folgt: LIME optimiert das duale Kriterium

$$\xi(x) = \arg\min \mathcal{L}(f, g, \pi_x) + \omega(g)$$

d. h. gleichzeitig werden die Summe des quadrierten Verlusts L sowie ein Komplexitätsmaß minimiert, um Erklärungen zu erhalten, die gut funktionieren („locally faithful"), die aber auch interpretierbar sind („low complexity"). g ist ein „Modell", das charakterisiert, ob für eine Dimension eine Erklärung vorliegt. Um das lokale Verhalten von f zu erlernen, während die interpretierbaren Eingaben variieren, kann $\mathcal{L}(f, g, \pi_x)$ angenähert werden, indem eine zufällige Anzahl von zufälligen Stichproben gewichtet mit π_x ausgewählt wird. Beispiel-Instanzen werden beschafft um x_0, indem die Elemente, die nicht Null sind, aber sehr wohl in x enthalten sind, uniform zufällig gesampelt werden. Danach erhält man für diese Klassen Etiketten mit dem existierenden Klassifikator und die so klassifizierten Daten dienen als Trainingsdaten für den neuen Surrogat-Klassifikator. Dabei werden diejenigen Punkte in der Nähe des zu erklärenden Punktes in unserem neuen, gewichteten linearen Modell höher gewichtet:

$$\mathcal{L}(f, g, \pi_x) = \sum_{z, z' \in Z} \pi_x(z)\left(f(z) - g(z')\right)^2,$$

wobei $\pi(z) = exp\left(\frac{-D(x,z)^2}{\sigma^2}\right)$ und D ein Distanzmaß ist (z. B. Kosinusähnlichkeit für Text); dies stellt auch Werte innerhalb $\in [0;1]$ sicher.

LIME erzeugt lokal vertrauenswürdige, lineare Erklärungen. Ein Komplexitätsmaß Omega enthält eine Schranke K (z. B. der Anzahl der Erklärungswörter in einer Textklassifikation-Aufgabe) zur Sicherstellung der menschlichen Interpretierbarkeit.

Shapley-Werte (Shapley 1951), eine Entdeckung, die ursprünglich aus der Spieltheorie stammt (benannt nach Lloyd Shapley, Nobelpreis für Ökonomie 2012), bieten Wege zur ex post-Erklärung nichtlinearer Modelle. Für ein Modell, welches mit einer Menge von Merkmalen in Form einer Funktion über einer Koalition von Spielern trainiert wurde, bieten sie eine additive Möglichkeit an zu berechnen, welche Merkmale wie viel zu einer Entscheidung beitragen.

SHAP (kurz für „SHapley Additive exPlanations“, Lundberg/Lee 2017) ist eine Methode, die Shapley-Werte auf die Erklärbarkeit anwendet. Dabei werden ausgehend von der A-priori-Wahrscheinlichkeit für eine Klasse die Merkmale und ihr additiver bzw. subtraktiver Effekt zur Gesamtentscheidung einzeln und sequenziell betrachtet. Vereinfacht gesprochen gilt: Der SHAP-Wert eines Merkmals ist die Differenz zwischen dem Mittelwert eines Merkmals und dem partiellen Abhängigkeitsgraph, der sich ergibt, wenn wir ein Merkmal ändern, wobei aber die Reihenfolge relevant ist.[3]

Diese und weitere post hoc-Methoden sind in der folgenden Übersicht nach Erklärungsbereichen und -ansätzen dargestellt:

What to explain?

How to explain?

	global	local	data
Profile	Partial Dependence Plot (PDP) Individual Conditional Expectation (ICE)	Ceretis Paribus Plot	
Parts	Global Feature Importance Leave-One-Covariate-Out (LOCO)	SHAP Attribution Break-Down Attribution	Graphical Networks
Distribution			Histogramm Boxplot Barplot

Abb. 2: Post hoc-Methoden (modellagnostisch)[4]

Bei der Betrachtung jeglicher Automatisierung und im speziellen bei Verfahren der KI, die ja auch kognitive, also spezifisch menschliche Fähigkeiten, automatisiert, ist eine Beleuchtung der ethischen Seite unerlässlich. Ethische Probleme in diesem Zusammenhang beinhalten Fragen der Moralität von Automatisierung generell, von Gerechtigkeit (engl. „fairness“/„justice“) und der Transparenz (Leidner (i. V.)). Ist die Automatisierung einer Aktivität generell moralisch zu vertreten? Hier ist beispielsweise der Verlust von Arbeitsplätzen zu nennen, wenn die Tätigkeit menschlicher Analysten durch ein Computerprogramm ersetzt wird, welches weniger kostet, nie schläft, Urlaub macht, krank wird usw. und noch dazu schneller Ratingmodelle berechnen kann. Die Frage

[3] Siehe *Lundbergs* Vortrag „The Science behind SHAP“: https://www.youtube.com/watch?v=-taOhqkiuIo [01.02.2023].

[4] Quelle: *Baniecki*/*Bizec* 2021.

der Transparenz lautet: „Kann der Mensch nachvollziehen, warum eine spezifische Entscheidung getroffen wurde?" Die Transparenz von Ratings von „Menschenhand", ohne KI-Methoden, war bereits traditionell problematisch aufgrund der Wahrscheinlichkeit von Interessenkonflikten, da die Ratings durch die bewerteten Unternehmen i. d. R. in Auftrag gegeben und bezahlt wurden.

Eine Automatisierung bietet Schutz vor statistischer Verzerrung einzelner (Bias), also mehr Objektivität. Traditionelle, besonders tiefe neuronale Netze (also solche mit mehr als einer Zwischenschicht), sind primär Black Box-Modelle, die einerseits sehr gut klassifizieren, sich andererseits aber einer Analyse entziehen. Das hier vorgeschlagene Modell bietet ex ante oder White Box-Transparenz „von Entwurfs wegen" (*per Design*). Nachfolgend werden erste Kriterien für ethische KI vorgeschlagen, die keinen Anspruch auf Vollständigkeit erheben, aber hoffentlich bei der Untersuchung der moralischen Seite von Modellen hilfreich sein können.[5]

Wie bereits erwähnt, kommt der Transparenz eine wichtige Stellung zu. Zu Modellen sollten Struktur und Annahmen offengelegt werden (welche Größen lassen sich beeinflussen und welche nicht). Ein Modell sollte legal sein, also in Einklang mit Gesetzen und Regulierungen stehen. Ein Modell sollte auch frei von Diskriminierung (z. B. aufgrund des Geschlechts oder Religion) sein. Zudem sollte ein Modell bei gleichen Eingabedaten auch die gleiche Ausgabe erneut liefern (Reproduzierbarkeit). Falls möglich, ist es wünschenswert, den Programmcode von Modellen als Open Source-Software (quelloffen) zur Überprüfung zur Verfügung zu stellen. Eine solche Veröffentlichung ermöglicht über die Reproduzierbarkeit hinaus das Verständnis der Funktionsweise, sowie die Anpassbarkeit auf neue Fragestellungen.

Bei den Erklärungen eines Modells sollte Klarheit über Kausalität oder Korrelation vorliegen: In der nachträglichen Erklärung von KI-Entscheidungen werden Kennzahlen wie Korrelation oder Shapley Values verwendet, um die wesentlichen Wirkungszusammenhänge aufzudecken. Kausalität kann jedoch nicht aus Daten alleine abgelesen werden, dies erfordert Input durch die fachliche Expertise eines Menschen – ein starkes Argument für die Erklärbarkeit per Design.

2.2 XAI und Ratings

Bei Ratings handelt es sich um die Bewertung von Unternehmen, (Kapitalmarkt-)Produkten oder Personen, typischerweise in Bezug auf die Finanzstärke. Ein spezieller Zweig bezieht sich auf die Bewertung der Kreditwürdigkeit mittels

[5] Für einen Kriterienkatalog für vertrauenswürdige KI vgl. beispielsweise HEG-KI 2018. Zu Erklärbarkeit im Versicherungsbereich vgl. vertiefend *Owens* et al. 2022.

Kategorisierung, wie beispielsweise ein AAA-Rating als beste Ratingeinstufung durch Standard & Poor's. Solche Kreditratingagenturen („Credit Rating Agencies", CRA) sind in der EU durch die Verordnung für Kreditratingagenturen (CRA-Verordnung) reguliert und werden durch die Finanzaufsichtsbehörden überwacht.

Ein Kreditrating ist eine Einschätzung über die Kreditwürdigkeit, die anhand von Ratingkategorien ausgedrückt wird. Sie sind auf professioneller Basis ausgestellt, an ein bestimmtes Finanzinstrument, eine Verpflichtung oder einen Emittenten gebunden, erfordern analytischen Input von Ratinganalysten und werden öffentlich bekannt gegeben oder im Abonnement verteilt.

Wenn eine Bonitätsbewertung ausschließlich aus der Zusammenfassung und Darstellung von Daten auf der Grundlage eines vorab erstellten statistischen Modells abgeleitet wird und keine wesentlichen Rating-spezifischen analytischen Eingaben in die Bewertung einfließen, gilt das Produkt als „Credit Score" und ist nicht aufsichtsrechtlich reguliert. Daher fallen die später gezeigten vollautomatischen Ratings nicht unter das Kreditratingregime, insbesondere solange sie nicht mit einer Ratingkategorie verknüpft werden.

Kreditratings helfen Anlegern und Kreditgebern, die mit einer bestimmten Anlage oder einem bestimmten Finanzinstrument verbundenen Risiken zu verstehen. In der Zeit vor der Finanzkrise im Jahr 2008 haben Ratingagenturen die Risiken einiger komplexerer Finanzinstrumente nicht richtig eingeschätzt. Als Reaktion darauf hat die Europäische Kommission den Regulierungs- und Aufsichtsrahmen für Ratingagenturen in der EU gestärkt, um das Marktvertrauen wiederherzustellen und den Anlegerschutz zu erhöhen. Seit Ende 2009 müssen Ratingagenturen registriert werden und werden von den zuständigen nationalen Behörden beaufsichtigt.[6] Darüber hinaus müssen Ratingagenturen Interessenkonflikte vermeiden und über solide Ratingmethoden und transparente Ratingaktivitäten verfügen.

Aber auch unabhängig von gesetzlicher Regulierung besteht der Bedarf, sicherzustellen, dass solche Bewertungen, die großen ökonomischen Einfluss auf die bewerteten Einheiten haben können, fair, verlässlich und erklärbar sind. Beispielsweise gibt es Überlegungen, dem (Privat-)Kunden das gesetzliche Recht einzuräumen, das Ratingergebnis erklärt zu bekommen. Gerade im Fall einer Ablehnung der Kreditanfrage ist dies besonders wichtig. Die Regulierung und der Kundenschutz werden künftig wesentliche Treiber für erklärbare Ratingmodelle sein.

[6] Eine Ratingagentur muss nach Art. 2 Abs. 3 der VO EG/1060/2009 eine Registrierung beantragen, um als externe Ratingagentur gemäß der Richtlinie 2006/48/EG anerkannt zu werden.

KI-Modelle bieten bei breiter Datengrundlage typischerweise eine sehr gute Performance, was zu deren starker Popularität beigetragen hat. Sie können jedoch im Allgemeinen nicht erklären, warum die Entscheidung getroffen wurde. Schlimmer noch, sie könnten auf Diskriminierung beruhen, z. B. in Bezug auf Geschlecht, Alter, ethnischer Zugehörigkeit usw. Es ist daher von entscheidender Bedeutung, eine valide, klare und gesetzeskonforme Erklärung der Ratingentscheidung zu liefern, die für Unternehmen und private Nutzer verständlich ist. Hybridmodelle sind ein vielversprechender Weg, dies zu erreichen.

Um Ratingergebnisse erklärbar zu machen, besteht entweder die Möglichkeit, ein Black Box-Modell im Nachhinein zu erläutern, oder das Modell schon per Design erklärbar zu machen (2.1). Erklärbarkeit per Design erfolgt typischerweise, indem allgemeines Fach- und Expertenwissen in graphische Strukturen übersetzt wird (Wissensgraphen, eng. „Knowledge Graph“) und schon von Beginn an in das Modell integriert wird. Man spricht dann auch von hybriden Modellen, da Domänenwissen mit Machine Learning auf Datenbasis verknüpft wird.

Externes Rating-Wissen sichert zum einen die spätere Erklärbarkeit und zum anderen wird auch die Schätzung vereinfacht, denn bekanntes Wissen muss nicht mehr neu aus Daten gelernt werden. Und gerade im Fall kleinerer Datenmengen, also nicht Big Data, bringt Domänenwissen eine Struktur, die die Anzahl freier Parameter reduziert, welche dann präziser geschätzt werden können.

Kausale Strukturen müssen von Experten implementiert werden. Es ist bekannt, dass aus Daten Korrelationen ableitbar sind, nicht jedoch kausale Beziehungen (Simpson 1951). Kausale Zusammenhänge eines konkreten Anwendungsfalls müssen also durch einen Experten definiert werden. Wenn Banken zum Beispiel die Bonität von Privatkunden bewerten wollen, so ist das finanzielle Vermögen ein Hauptfaktor der Kreditwürdigkeit und sollte als solcher zugelassen werden. Das Geschlecht hingegen sollte keinen Einfluss auf die Kreditwürdigkeit haben, zumindest nicht direkt. Um einen insgesamt hohen Prozentsatz korrekter Klassifikationen zu gewährleisten, kann es jedoch einen indirekten Effekt geben, z. B. wenn das Geschlecht einen Einfluss auf das Einkommen hat, welches wiederum die Finanzkraft beeinflusst.

Vordefiniertes Expertenwissen und Geschäftsregeln können in einem kausalen Graphen dargestellt werden, der zeigt, welche Ursachen Auswirkungen auf bestimmte Variablen haben. Der genaue Wert des Effekts kann dann mittels maschinellen Lernens bestimmt werden. Hier kommt der hybride Ansatz ins Spiel. Dies führt zu stark strukturierten Modellen. Das neuronale Netz ist also nicht völlig frei, beliebige Beziehungen zu den gegebenen Variablen anzupassen. Stattdessen sind viele Relationen explizit ausgeschlossen, die die vorgegebene Struktur repräsentieren.

Es gibt grundsätzlich einen Trade-Off zwischen Performance und Erklärbarkeit:[7] Tiefe KI-Modelle bieten typischerweise eine sehr hohe Prognosequalität bzw. Reproduktion beobachteter Daten (vgl. Abb. 1). Dies gilt zumindest solange große Datenmengen verfügbar sind. Dies ist auch der Hauptgrund, warum sie aktuell so erfolgreich eingesetzt werden. Als Nachteil zeigt sich in der Anwendung die mangelnde Interpretierbarkeit. Durch Vorgabe von strukturellen Zusammenhängen kann diese erhöht werden, jedoch ist dies ggfs. mit einem gewissen Rückgang der Prognosegüte verbunden. Gerade bei wenig Daten, wie im Fall der Unternehmensratings, ist aber ohnehin ein „sparsames" Modell erforderlich, damit die geringen Freiheitsgrade präziser geschätzt bzw. gelernt werden können und um das Problem des Overfittings zu vermeiden.

Ratinganalysen von Unternehmen basieren wesentlich auf Rechnungslegungsdaten. Im Zusammenhang von KI und Rechnungslegung weist Sellhorn (2020) darauf hin, dass mangelnde Transparenz Vertrauen kostet. Im Kontext der Unternehmensberichterstattung und Abschlussprüfung stellen Kokina und Davenport (2017) fest: „However, machine learning and deep learning neural networks, for example, are often ‚black boxes' that are difficult or impossible to understand and interpret, even for technical experts. Until such technologies are made more transparent, it may be difficult for regulatory bodies, accounting firms, and audited organizations to turn over decisions and judgments to them." Auch weisen Dierkes und Sümpelmann (2019, S. 190) darauf hin, dass es „durch die Digitalisierung nicht zu einer Black Box werden" darf, der es „an einer ausreichenden theoretischen Fundierung mangelt". Im nachfolgenden Use Case wird gezeigt, wie dieser Kritik bei Unternehmensratings begegnet werden kann.

3. Use Case: Unternehmensratings

3.1 Ein hybrides Ratingmodell mit Expertenwissen

Anhand eines Ratings aller deutschen Lebensversicherer wird demonstriert, wie die Erklärbarkeit eines KI-Ratings sichergestellt wird. Dabei wird ein hybrides Modell der KI eingesetzt, welches ex ante-Expertenwissen verwendet und das resultierende strukturelle neuronale Netz mit klassischen Methoden des maschinellen Lernens schätzt.

Inputdaten sind die veröffentlichte Bilanz sowie die zugehörige Gewinn- und Verlustrechnung aus den Geschäftsberichten. Die Daten sind – neben HGB- und IFRS-Vorgaben – durch die Vorgaben der Rechnungslegungsverordnung für Versicherungsunternehmen (RechVersV) normiert. Zudem ist die Daten-

[7] Vgl. dazu auch *Oletzky*/*Reinhard* 2022, S. 505.

qualität hoch, da die Jahresabschlüsse verpflichtend durch einen Wirtschaftsprüfer zu testieren und die Unternehmen beaufsichtigt sind.

Das beispielhaft vorgestellte Unternehmensrating basiert auf dem Ansatz der Ratingagentur RealRate. Das Rating umfasst alle (aktiven) deutschen Lebensversicherer. Diese sind ein besonders anspruchsvolles Beispiel, da die Komplexität des Geschäftsmodells recht hoch ist und die Bilanzdaten zahlreiche branchenspezifische Besonderheiten aufweisen, die mit den Produkten (z. B. der garantierte Rechnungszins) oder mit dem Geschäftsmodell (z. B. Überschussbeteiligung) zusammenhängen. Umso größer ist der Bedarf, sowohl für Analysten als auch für die Kunden, die wesentlichen Zusammenhänge, Stärken und Schwächen schnell erfassen zu können.

Das Modell soll ausschließlich mit extern verfügbaren Daten auskommen, um darauf basierend eine Umbewertung von der Buchwertbilanz in eine Marktwertbilanz durchzuführen, ähnlich wie es das Aufsichtsregime Solvency II fordert.[8] Daher müssen zum Beispiel Annahmen zur Zinssensitivität der Passivseite (Passivduration) in das Modell aufgenommen werden. Der ursprüngliche Ansatz von Bartel (2014a) benötigt jedoch teilweise noch interne Unternehmensangaben, um das Risiko des Versicherers bestimmen zu können. Für den hier betrachteten Anwendungsfall wird sich daher auf die Bestimmung des ökonomischen Eigenkapitals, also den Zähler der Solvabilitätsquote, beschränkt und als Bezugsgröße, also den eigentlichen Nenner der Solvabilitätsquote, vereinfachend nur die Bilanzsumme als Größenmaßstab verwendet.

Der Ansatz von Bartel (2014a, 2014b) schlägt ein stark vereinfachtes ökonomisches Solvenzmodell für Lebensversicherer vor, welches die wesentlichen aufsichtsrechtlichen Solvabilitätsregeln abbildet. Es berücksichtigt die vertraglichen Besonderheiten, insbesondere die Überschussbeteiligung der Kunden an den bisherigen Gewinnen (über die Rückstellung für Beitragsrückerstattung) sowie an den zukünftigen Gewinnen.

Das Modell soll nur so komplex wie nötig sein, um das Geschäftsmodell der deutschen Lebensversicherer abzubilden, aber auch so einfach wie möglich, um die Erklärbarkeit per Design zu gewährleisten. Ziel ist die Vermeidung aufwändiger Cashflow-Projektionen oder stochastischer Simulationen und somit die Verbesserung der Transparenz der Wirkungsweise. Im RealRate-Modell sind nur wenige Eingabeparameter und -variablen erforderlich. Diese wenigen Größen sind auf Grund des asymmetrischen Geschäftsmodells allerdings stark nichtlinear miteinander verknüpft. Es wird zudem eine geschlossene Formel vorgeschlagen, um den Wert der sogenannten Garantien und Optionen zu bestimmen. Dieser entsteht bei der Unternehmensbewertung, da der Kunde nur an positiven Gewinnen beteiligt wird, aufgrund der ihm gewährten Garantie

[8] Zu wesentlichen Grundlagen des Ratingansatzes siehe *Bartel* 2014a.

aber nicht an Verlusten. Die Bewertung dieses Effekts erfolgt mit einer geschlossenen Optionspreisformel, dem sogenannten „Puffer Put“ (siehe Bartel 2014a).

Bei dem verwendeten Modell handelt sich um ein ganzheitliches Unternehmensmodell und nicht nur um eine Gewichtung interessanter Kennzahlen. Die Ausgangswerte für die Effekte zwischen den Variablen ergeben sich zuerst aus dem in Gleichungsform vorgegebenen Expertensystem. Diese Effekte werden dann beim maschinellen Lernen so gewählt, dass die veröffentlichen Solvenzquoten (ohne aufsichtsrechtliche Übergangsmaßnahmen) gemäß Solvency II möglichst gut erklärt werden. Anstatt ein unrestringiertes neuronales Netz zu schätzen, erfolgt in unserem Fall eine Optimierung unter (kausalen) Nebenbedingungen. Dabei bestimmt der Modellierer, welche Koeffizienten frei gewählt werden können. Andere Gewichte bleiben hingegen unverändert, beispielsweise wenn sie aus einer reinen Definitionsgleichung stammen.

Die Gewichte des Netzwerkes entsprechen den gesuchten Effekten, welche die quantitativen Wirkungen zwischen den Variablen messen. Sie dienen der Erklärbarkeit. Während in tiefen neuronalen Netzen nur gesamthafte Effekte im Sinne einer totalen Ableitung ex post bestimmt werden, können in kausalen Modellen auch direkte Effekte im Sinne partieller Ableitungen bestimmt werden.

Der sich aus dem RealRate-Bewertungsmodell ergebende Graph ähnelt einem neuronalen Netz (siehe Abb. 6). Die Knoten bzw. Neuronen entsprechen dabei den im Modell verwendeten Variablen und die Kanten bzw. die Gewichte den gesuchten Effekten. Es werden also tatsächlich die einzelnen Neuronen des strukturierten Netzwerkes interpretiert, was bei tiefen Netzen nicht möglich ist. Der Unterschied zu einem klassischen neuronalen Netz besteht zum einen darin, dass nicht alle Variablen einer Schicht miteinander verknüpft sind, sondern viele Variablen aufgrund der gegebenen kausalen Struktur gerade *nicht* miteinander verknüpft sind. Zum anderen ist dieses strukturelle neuronale Netz wesentlich kleiner als typische tiefe neuronale Netze. Genau dies ermöglicht schließlich die Erklärbarkeit.

Die methodischen Besonderheiten des erklärbaren KI-Ansatzes bestehen darin, dass ein kausales Modell vorgegeben wird. Dieses strukturierte Modell stellt per Design die spätere Erklärbarkeit sicher und kann auch regulatorische Vorgaben explizit einbeziehen. Im Gegensatz zu typischen Deep Learning-Ansätzen wird jeder Knoten, also jedes Neuron, interpretierbar, denn es entspricht einer Modellvariablen. Diese sind typischerweise latent, also nicht beobachtbar. Für die einzelnen Sensitivitäten, also die Effekte jeder einzelnen Variablen auf die final interessierende Variable, werden automatisch algebraische Formeln abgeleitet. Dies verbessert zudem die Stabilität der Backpropagation und des Optimierungsalgorithmus. Für das Supervised Learning, also die Schätzung der durch die gerichteten Kanten repräsentierten Modellgewichte, müssen einige

Variablen beobachtbar sein. Dieser Ansatz ermöglicht auch Signifikanztest jedes einzelnen Effekts. Schließlich ist vor allem die einfache graphische Darstellung der Zusammenhänge möglich.

Das gesamte Unternehmensmodell umfasst insgesamt 32 Gleichungen (Bartel 2020d). Einige Gleichungen sind reine Definitionen, wie zum Beispiel:

Eigenkapital ← nachrangige Verbindlichkeiten
+ HGB-Eigenkapital
+ Genussrechte

Zum Ausdruck der Richtung (Kausalität) wird anstatt des Gleichheitszeichens („="), wie es in einer mathematischen Gleichung üblich ist, ein gerichteter Pfeil („←") verwendet. So ergibt sich das HGB-Eigenkapital als Output und die drei Inputs sind:

1. die nachrangigen Verbindlichkeiten,
2. das HGB-Eigenkapital ohne Genussrechte und nachrangige Verbindlichkeiten und
3. die Genussrechte.

Sie werden einfach aufsummiert und ergeben zusammen das HGB-Eigenkapital. In diesem Fall sind die drei direkten Effekte einfach gleich eins: Sie entsprechen nämlich den partiellen Ableitungen der Outputvariablen (Eigenkapital) in Bezug auf die drei Inputvariablen. Der entsprechende Teilgraph sieht damit aus wie in Abb. 3:

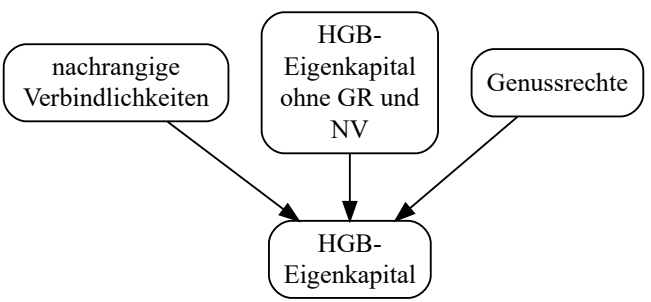

Abb. 3: Teilgraph für das HGB-Eigenkapital

Das Eigenkapital ist damit eine endogene und latente (nicht beobachtbare) Modellvariable, während die drei anderen Größen exogen und manifest (beobachtbar sind). Jede Modellgleichung determiniert damit kausal eine Modellvariable.

Andere Gleichungen dienen der Umbewertung von der Buchwertbilanz zur Marktwertbilanz, wie in Abb. 4 dargestellt:

Marktwert Kapitalanlagen ← aktivische Bewertungsreserven
+ Buchwert Kapitalanlagen

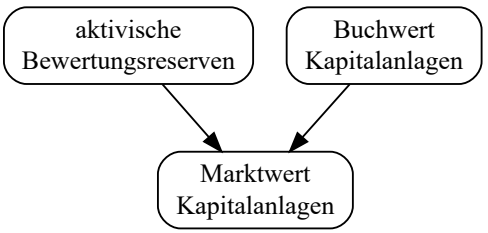

Abb. 4: Teilgraph für den Marktwert Kapitalanlagen

Und wieder andere Gleichungen dienen der inhaltlichen Modellierung der Finanzstärke, wie in Abb. 5 dargestellt:

ökonomische Eigenkapitalquote ← ökonomisches Eigenkapital/HGB-Bilanzsumme

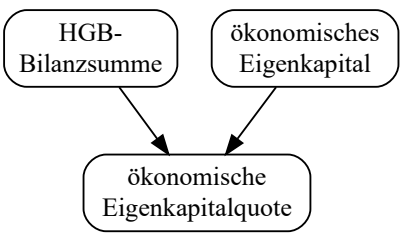

Abb. 5: Teilgraph für die ökonomische Eigenkapitalquote

Die ökonomische Eigenkapitalquote ist die finale interessierende Variable und ist daher die letzte Modellgleichung und damit letzte (unterste) Variable im erklärenden Graphen. Die Eingabegrößen aus Bilanz und Gewinn- und Verlustrechnung werden im Graphen Abb. 6 dargestellt. Eingabegrößen werden auch als exogene Größen bezeichnet und haben im Graphen keine eingehenden Pfeile. Die Ausgabegrößen werden hingegen durch das Modell bestimmt und sie werden auch als endogene Variablen bezeichnet. Die meisten endogenen Variablen sind zudem in der Regel latent, also nicht beobachtbar. Zumindest eine beobachtbare endogene Variable muss es aber geben, damit das Modell mittels maschinellen Lernens überprüft und die Koeffizienten so angepasst werden können, dass die Realität möglichst gut erklärt wird. Diese verwendeten Koeffizienten ergeben sich aus dem optimierten Modell, indem einfach die entsprechenden Ableitungen gebildet werden. Aus allen Gleichungen zusammen ergibt

sich dann eine allgemeine graphische Struktur, welche direkt für die Berechnung verwendet wird. Dabei gibt es in unserem Ansatz zwei Besonderheiten zu den üblicherweise verwendeten Netzen:

1. Strukturiertes neuronales Netz: Das Netz ist stark restringiert: Beziehungen zwischen den Variablen sind nur zulässig, wenn diese verbunden sind. Die allermeisten Variablen sind aber nicht miteinander verknüpft. Die Struktur wird also durch diese „Nicht-Verbindungen" (mathematisch Null-Restriktionen) bestimmt. Zudem bestimmt die Richtung des Pfeils in welche Richtung der kausale Zusammenhang fließt. Bei A ← B kann also B die Größe A beeinflussen, nicht aber umgekehrt.
2. Shallow Learning/Small Data: Im Gegensatz zum üblicherweise verwendeten Deep Learning mit Big Data wird hier nur mit einem sehr kleinen Netz gearbeitet. Dies hat zum einen den Vorteil, dass auch nur recht wenige Daten zu Schätzung benötigt werden. Zum Beispiel kann eine ganze Branche mit 100 Unternehmen mit jeweils 100 Bilanzkennzahlen geratet werden, wenn nur diese 10.000 Daten vorliegen. Je weniger Daten man hat, desto kleiner müssen die Modelle sein, damit man diese wenigen Freiheitsgrade statistisch valide schätzen kann. Bei anderen Aufgaben wie der Bilderkennung durch KI werden typischerweise Millionen Datenpunkte benötigt.

Parameter wie zum Beispiel das Überschussbeteiligungsniveau oder die Steuerquote werden später durch maschinelles Lernen so geschätzt, dass die beobachteten Solvency II-Solvenzquoten möglichst gut erklärt werden. Die Kalibrierung der einzelnen Effekte in dieser Struktur wird durch das maschinelle Lernen optimal bestimmt.[9]

3.2 Der erklärbare kausale Graph

Für jedes einzelne geratete Unternehmen ist diese Struktur mit unterschiedlichen Werten und Farben gefüllt. In Tabelle 1 ist das RealRate-Finanzstärke-Ranking der deutschen Lebensversicherung 2022 (basierend auf den Geschäftsberichtsdaten 2021) dargestellt.[10] Die HUK-COBURG belegt im Ranking den ersten Platz mit einer ökonomischen Eigenkapitalquote von 20,07 %. Die Allianz Leben hat im Ranking den Platz 38 von insgesamt 60 untersuchten Lebensversicherern belegt und verfügt über eine ökonomischen Eigenkapitalquote von 8,23 %. Dies mag auf den ersten Blick verwundern, da die Allianz der klare Marktführer im deutschen Lebensversicherungsbereich ist. Die Unternehmens-

[9] Für weitere Details zur Anwendung der erklärbaren künstlichen Intelligenz im Ratingbereich vgl. Bartel 2020a, 2020b, 2020c.

[10] Dieses und weitere Ratings sind frei online verfügbar: https://realrate.ai/rankings [01.02.2023].

Tabelle 1

RealRate-Finanzstärke-Ranking deutscher Lebensversicherer 2022

Rang	*Lebensversicherer*	*ökonomische Eigenkapitalquote*
1.	HUK-COBURG	20,07 %
2.	VRK	18,20 %
3.	Bayerische Beamten	17,20 %
4.	BL die Bayerische	15,38 %
5.	Heidelberger	14,63 %
…		
36.	HDI	8,42 %
37.	DEVK	8,38 %
38.	*Allianz*	*8,23 %*
39.	NÜRNBERGER	8,10 %
40.	R + V	7,34 %
…		

größe an sich ist jedoch kein Qualitätsmerkmal im KI-Finanzstärke-Ranking. Ganz im Gegenteil wird deutlich, dass die Top 5 mit eher kleineren Lebensversicherern mit sehr guter Finanzstärke belegt sind. Die Finanzstärke, gemessen als ökonomisches Eigenkapital im Verhältnis zur Bilanzsumme, hat im Markt eine Spannbreite von rund –11 % bis rund 20 %.

Die relativen Stärken der Allianz sind in Abb. 6 grün dargestellt, relative Schwächen sind rot dargestellt. Die Werte in den Knoten quantifizieren den Effekt der jeweiligen Variablen im Vergleich zum Marktmittel auf die Finanzstärke. Es sei betont, dass dieser Graph speziell das Unternehmen Allianz darstellt; für die anderen Unternehmen ist die Struktur des Graphen zwar identisch, aber die individuellen Werte, also die Effekte auf die Finanzstärke sind natürlich verschieden. Zudem wird der kausale Graph für jeden Versicherer so dargestellt, dass die wesentlichen Effekte hervorgehoben werden, während unwesentliche Effekte gar nicht dargestellt werden. Alle dargestellten quantitativen Effekte sind die auf die interessierende finale Variable, nämlich die ganz unten im Graphen dargestellte ökonomische Eigenkapitalquote.

Die ökonomische Eigenkapitalquote, wie in der Rankingtabelle angegeben, beträgt 8,23 % und liegt damit um 0,652 Prozentpunkte knapp unter dem Marktmittel von 8,9 %. Die größte Stärke der Allianz Leben ist der geringe zugesagte Garantiezins (mittlerer Tarifrechnungszins), welcher indirekt aus den verfügbaren Bilanzdaten geschätzt wird. Die Allianz ist in der Niedrigzinsphase früher als anderer Versicherer damit gestartet, die Last der Garantiezinsen zu senken,

Abb. 6: RealRate-Finanzstärkeanalyse Allianz Lebensversicherung

indem ergänzend moderne Produkte mit geringeren Zinsgarantien eingeführt wurden. Diese Strategie hat die Finanzstärke nachhaltig gestärkt: Die ökonomische Eigenkapitalquote steigt hierdurch im Vergleich zum Marktdurchschnitt um 4,63 Prozentpunkte (siehe Abb. 6). Auch in der Kapitalanlage ist die Allianz erfolgreich. Sie verfügt über umfangreiche aktivische Bewertungsreserven. Dies bedeutet, dass die Marktwerte der Kapitalanlagen höher sind als die konservativ angesetzten Buchwerte der HGB-Bilanz. Dadurch steigt ihre ökonomische Eigenkapitalquote um 2,27 Prozentpunkte im Vergleich zum Durchschnitt aller Lebensversicherer. Rechnungszins und Bewertungsreserven werden sich zusammen in Form von künftigen Überschüssen realisieren, welche im RealRate-Bewertungsmodell angesetzt werden. Dies entspricht auch der Logik des Solvency II-Aufsichtsregimes.

Eine relative Schwäche der Allianz ist hingegen das marktunterdurchschnittliche Risiko- und übrige Ergebnis. Dies entspricht einer unterdurchschnittlichen versicherungstechnischen Produktprofitabilität. Dies reduziert die Finanzstärke um 3,22 Prozentpunkte. Obwohl das handelsrechtliche Eigenkapital (ohne Genussrechte und nachrangige Verbindlichkeiten) der Allianz Leben in Höhe von rund 3 Mrd. Euro absolut gesehen hoch ist, so ist es im Vergleich zur Bilanzsumme von rund 284 Mrd. Euro marktunterdurchschnittlich. Dies reduziert die Finanzstärke-Kennziffer um 2,26 Prozentpunkte. Im RealRate-Ratingmodell gibt es keinen Bonus für absolute Größe. Stattdessen wird die relative Bilanzstruktur verglichen, bei der die Größe keine Rolle spielt.

Insgesamt werden die wesentlichen ökonomischen Zusammenhänge anhand des kausalen Graphen erklärbar. Im Vergleich zu anderen, rein kennzahlenbasierten Ratings, werden die mehrstufigen Wirkungszusammenhänge über die verschiedenen Mediatorvariablen deutlich, welche kausal interpretiert werden können. Der Graph ist leicht lesbar und stellt komplexe Geschäftsmodelle und Zusammenhänge in nur einem Bild dar. Im Gegensatz zu hunderten Seiten starken Geschäfts- oder Ratingberichten ist diese Darstellung in der Praxis eine große Hilfe. Da der Graph die individuellen Stärken und Schwächen des Unternehmens stets im Verhältnis zum Gesamtmarkt darstellt, ist er direkt als Benchmark- oder Peergroupanalyse geeignet.

Der kausale Graph ist in der Tat neben der eigentlichen Finanzstärke das zentrale Ratingergebnis und erklärt diese Größe. Der Graph wird zur Verfügung gestellt und ist die Grundlage für die strategische Analyse und künftige Ratingverbesserungen durch das Unternehmen. Er kann insbesondere dem Vorstand, dem Risikomanagement und der Unternehmensplanung als Entscheidungsgrundlage dienen. Die dargestellten Effekte sind als Sensitivitäten interpretierbar. In dem Beispiel zeigt ein grüner Knoten, also eine individuelle Stärke, um wie viel Prozentpunkte die Finanzstärke dadurch erhöht ist, dass die entsprechende Größe des Unternehmens im positiven Sinne vom Marktmittel abweicht.

Die kausale Struktur ist für alle Unternehmen gleich, allerdings fällt die Quantifizierung der individuellen Effekte unterschiedlich aus. Dies ist eine direkte Konsequenz aus der Tatsache, dass der Graph mit den unternehmensindividuellen Daten gespeist wird, in dem vorliegenden Beispiel hier mit den Daten der Allianz. So erhält man einen übergreifend gültigen Interpretations- und Erklärungsrahmen, aber gleichzeitig eine ganz individuelle Unternehmensanalyse. Die im Voraus definierte kausale Struktur sichert dabei per Design die spätere Erklärbarkeit der Ergebnisse. Die wichtigsten positiven und negativen Effekte lassen sich einfach aus der farbig abgestuften Grafik ablesen und erfüllen damit die Erklärbarkeitsanforderung.

Wie die Abb. 7 zeigt, hat sich die Stärke des geringen Garantiezinses über die Zeit kontinuierlich immer weiter verstärkt, bis zum aktuellen positiven Effekt in Höhe von +4,63 Prozentpunkte auf die ökonomische Eigenkapitalquote. Gleichzeitig hat sich jedoch auch die Schwäche der unterdurchschnittlichen versicherungstechnischen Profitabilität ungünstig entwickelt.

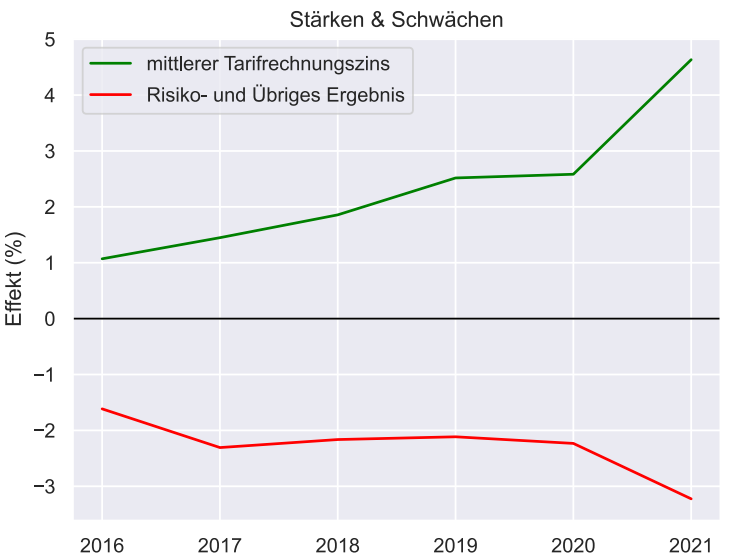

Abb. 7: Die größte Stärke und Schwäche der Allianz Leben im Zeitverlauf, Effekt auf die ökonomische Eigenkapitalquote in Prozentpunkten

Abschließend soll die Frage beantwortet werden, wie gut das erklärbare KI-Modell die Realität der deutschen Lebensversicherer im Bilanzjahr 2021 erklären kann. Hierfür ist auf der x-Achse die ökonomische Eigenkapitalquote abgetragen (Abb. 8). Auf der y-Achse werden die Solvency II-Eigenmittel, ohne Übergangsmaßnahmen, in Relation zur Bilanzsumme abgetragen. Die orange Linie zeigt den theoretisch wünschenswerten Zusammenhang an, also eine 1:1 Beziehung beider Größen. Die blaue Linie zeigt den tatsächlichen Regressions-Fit an. Es wird ein Zusammenhang zwischen diesen beiden Größen deutlich; die Rangkorrelation beträgt 0.3. Somit ist das Modell in der Lage, einen gewissen Teil der veröffentlichten Solvency II-Werte zu erklären, obwohl diese auf wesentlich mehr und nur intern verfügbaren Daten beruhen und zum Teil mittels komplexen stochastischen Simulationsmodellen erstellt wurden.

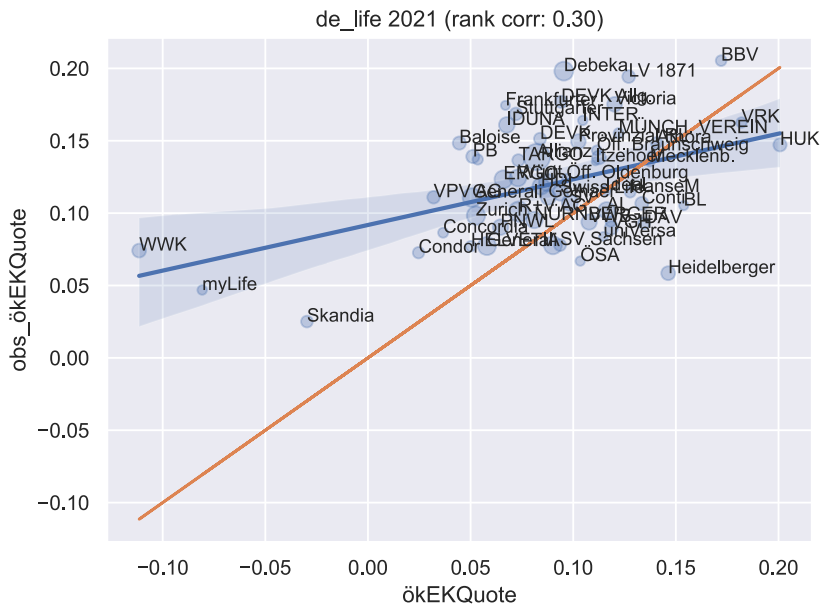

Abb. 8: Korrelation zwischen der ökonomische Eigenkapitalquote (x-Achse) und den aufsichtsrechtlichen Eigenmitteln gemäß Solvency II, ohne Übergangsmaßnahmen, in Relation zur Bilanzsumme (y-Achse)

3.3 Schätzung, Modellierungszyklus und Software

Die Schätzung des Hybridmodells erfolgt ähnlich wie im allgemeinen, uneingeschränkten Fall, jedoch unter Berücksichtigung der kausalen Restriktionen. Tatsächlich sieht der gegebene kausale Graph selbst schon wie ein neuronales Netz aus, ist aber nicht komplett vernetzt. Für die Schätzung des Modells ist eine zu minimierende Zielfunktion zu definieren. In unserem Beispiel ist dies die Fehlerquadratsumme zwischen der modellierten Finanzstärke auf der einen und den veröffentlichten Solvenzquoten auf der anderen Seite (die Zielfunktion ist ausformuliert in Bartel 2019).

Die Optimierung ist ein klassisches, nichtlineares Optimierungsproblem, hier jedoch unter der zusätzlichen Nebenbedingung, dass gewisse Knoten/Neuronen/Variablen kausal gerade nicht miteinander verknüpft sind. Dies wurde technisch als Strukturelles Neuronales Netz (SNN) umgesetzt. Ein ähnliches Vorgehen findet hinter den Kulissen auf technischer Ebene bereits heute für allgemeine unrestringierte Neuronale Netze statt: Für den bekannten Backpropagation Algorithmus (siehe Rumelhart et al. 1986) müssen die partiellen Ableitungen des Netzes bestimmt werden, was aus Gründen der Performance und numerischen Stabilität nicht numerisch erfolgt – allerdings auch nicht algebraisch, sondern die Ableitungen werden auf Ebene des Programmcodes gebildet. Dies ist die sogenannte Automatische Differentiation (siehe Rall 1981). Als Optimierungsalgorithmus wird zum Beispiel der im maschinellen Lernen besonders beliebte ADAM Optimierer verwendet (siehe Kingma/Ba 2014). Nach erfolgter Schätzung der Gewichte folgt die Modellvalidierung. So ist zu prüfen, ob das angenommene Modell zu den Daten passt. Der verwendete Computercode ist als Open Source Software frei verfügbar.[11] Der Name Causing dieser Software steht dabei für CAUSal INterpretation using Graphs. Sie ist allgemein für beliebige Themengebiete anwendbar. So ist dort beispielhaft eine Anwendung dargestellt, mit der die Lohnhöhe junger amerikanischer Arbeiter durch ihre Ausbildung und ihren familiären Hintergrund erklärt wird.[12] Causing ist ein multivariates grafisches Analysewerkzeug, mit dem die kausalen Auswirkungen eines gegebenen Gleichungssystems interpretiert werden können. Als Input muss lediglich ein Datensatz bereitgestellt und ein Gleichungssystem eingegeben werden. Es wird angenommen, dass die endogene Variable auf der linken Seite durch die Variablen auf der rechten Seite der Gleichung verursacht wird. Somit liefern sie die Kausalstruktur in Form eines gerichteten azyklischen Graphen. Als Output ergibt sich ein leicht verständliches farbiges Diagramm,

[11] Unter https://github.com/realrate/Causing [01.02.2023].

[12] Siehe https://github.com/realrate/Causing/blob/develop/docs/education.md [01.02.2023].

das die kausalen Wirkungszusammenhänge zwischen den Variablen anschaulich darstellt. Damit können ganze Wirkungsketten interpretiert werden.

Nachdem das Modell definiert und geschätzt wurde, können Modifikationen überprüft werden, die zu anderen Regeln und einer anderen Anpassungsgüte führen. Im hybriden Modellansatz ist folgender Modellierungszyklus zu durchlaufen:

1. Input definieren

Bestimmung der beobachteten exogenen Modellvariablen, die die Finanzstärke erklären.

2. Kausalstruktur definieren

Definieren der Kausalstruktur, indem festgelegt wird, zwischen welchen Variablen direkte inhaltliche Beziehungen bestehen. Auf diesem Weg wird letztlich die Finanzstärke bestimmt.

3. Maschinelles Lernen/Schätzung

Darstellung des kausalen Graphen als strukturiertes neuronales Netz und Schätzung der freien Parameter mittels maschinellen Lernens. Dabei werden die Parameter so gewählt, dass die veröffentlichten Solvenzquoten möglichst reproduziert werden.

4. Auswertung

Messung der Modellperformance und Prüfung der Erklärbarkeit in der Praxis.

5. Modelländerung

Modifikation des Modells durch Änderung der verwendeten Variablen oder deren kausaler Zusammenhänge. Neustart ab Schritt 1.

Insgesamt ergeben sich mit diesem hybriden Modellierungsansatz folgende Vorteile aus Anwendersicht: Erklärbarkeit, Transparenz, Skalierbarkeit, Small Data statt Big Data, Geschwindigkeit. Die Methodik erleichtert auch die Modellvalidierung (Messen der Modellperformance, Erklärbarkeit, Identifikation der wichtigsten Wirkungsweisen, Vergleich alternativer Erklärungen und Quantifizierung von Performanceverlusten). Ein transparentes Modell alleine ist jedoch nicht ausreichend. Ein solches ist nur dann sinnvoll, wenn auch das Problem des Interessenkonflikts adressiert wird, welcher aktuell das Verhältnis zwischen Unternehmen und Ratingagenturen betrifft. Im nächsten Abschnitt werden Transparenz und die Ausgestaltung von Geschäftsmodellen adressiert.

4. Transparenz und Geschäftsmodell

Die Geschäftszahlen von Versicherungsunternehmen werden von Marktakteuren zum Teil als intransparent erachtet. Insofern haben Ratings von Versicherungsunternehmen eine stärker erhellende Funktion, zumindest vermutlich. Für die Versicherungsbranche stehen Analysten vor der Herausforderung, dass die verschiedenen Sparten unterschiedliche Interpretationsschwierigkeiten hervorrufen. Während bei Schaden-/Unfallversicherungsunternehmen und bei Rückversicherungsunternehmen die Geschäftsergebnisse einer großen (zufälligen) Schwankungsbreite unterliegen können, spielt bei Lebensversicherungsunternehmen die Langfristigkeit und die stärkere Kapitalmarktabhängigkeit eine Rolle. Die Ausschläge bei Schaden-/Unfallversicherungsunternehmen sind z. B. durch Naturkatastrophen bestimmt. Als Beispiel zu nennen wären die Überflutungen 2021 im Ahrtal. Diese Groß- und Kumulschäden sind aber ohne weiteres auf Basis externer Bilanzzahlen nur sehr bedingt normalisierbar. Umgekehrt sind beispielsweise die Zinssensitivitäten der (deutschen) Lebensversicherungsbestände nur wenig erhellend und vergleichbar in den Geschäftsberichten dargestellt. Bei Versicherungsgruppen mit mehreren Sparten kommt hinzu, dass sich die Effekte aus verschiedenen rechtlichen Einheiten überlagern können. Üblich sind für diese Opazität sog. Konglomeratsabschläge.

Die MCEV-Berichterstattung auf Eigeninitiative internationaler, kapitalmarktorientierte Versicherungskonzerne war ein Versuch einer Antwort darauf (zu den MCEV-Prinzipien siehe CFO-Forum 2009). Die Abhängigkeit von Ratings internationaler Ratingagenturen konnte dadurch aber nur bedingt gemildert werden. Speziell wurden den Unternehmen Änderungen in der Methodik der MCEV-Berechnung wiederum als Beispiel für Intransparenz ausgelegt. Versuche, die hervorgehobene Marktmacht einiger Ratingagenturen wie beispielsweise Standard & Poor's, Moody's und Fitch (speziell auf europäischer Ebene) zu reduzieren, waren ebenfalls wenig erfolgreich.

Nachfolgend wird die Interaktion des Geschäftsmodells von Unternehmensratings mit der Unabhängigkeit untersucht. Szenario 1 „Beauftragung" zeigt den Ist-Zustand des Ratingsektors: Ratingagenturen haben die bewerteten Unternehmen als direkte Auftragskunden, was eine starke psychologische Bindung im Sinne einer unterschwelligen Verpflichtung erzeugt; manchmal nutzen die bewerteten Unternehmen diese finanzielle Macht auch expressis verbis, um Ratingagenturen Druck zu machen. Das Modell wird daher von Aufsehern und auch in der Politik teilweise sehr kritisch gesehen und war insbesondere auch in der Finanzkrise und bei anderen Skandalen Anlass für den Ruf nach gesetzlichen Veränderungen. Die Regulierung von solchen Ratingagenturen hat in der EU auch ein gewisses Maß erreicht, wobei die Ratingagenturen sich teilweise, insbesondere in den USA, auf die Meinungsfreiheit berufen. Dies ist nach europäi-

schem Verständnis für ein kommerzielles Geschäftsmodell mit nur wenigen großen Playern nur bedingt die (alleinige) Sichtweise.

Durch ein alternatives Geschäftsmodell, welches die Beziehung zwischen Ratingagentur und dem bewerteten Unternehmen entkoppelt, wird der dargestellte Interessenkonflikt gelöst bzw. zumindest gemildert. Das Szenario 2, „Public Rating", ist per se weniger anfällig für ungerechtfertigte Einflussnahmen.[13] Die durch KI ermöglichte Automatisierung kann weiterhin zu einer Skalierung (Vergrößerung des adressierbaren Universums bewerteter Unternehmen) genutzt werden, welche die Möglichkeiten individueller Einflussnahme verringern. Die Ratingagenturen erstellen Ratings von vielen Institutionen und Ratingagenturen werden von den bewerteten Institutionen im Nachhinein bezahlt, wenn die Institutionen ein bereits existierendes Rating für Werbezwecke republizieren möchten. Die bewerteten Institutionen haben keinen Einfluss auf die veröffentlichte Note. Zu den Aspekten öffentliche Daten, Transparenz, Interessenkonflikte und Finanzkrise 2008 siehe Bartel 2023.

Diese Public Ratings sind im Bereich der Produktratings sehr üblich, wie zum Beispiel bei der Stiftung Warentest. Im Bereich der Unternehmensratings haben sie sich allerdings noch nicht durchgesetzt, wofür verschiedene Gründe ursächlich sind. Die Forderung, interne Ratings statt externer Ratings zu verwenden, erscheint makroökonomisch nicht effizient: Einen aufwendigen internen Bewertungsprozess können gerade kleine und mittlere Unternehmen nicht leisten. Es bleibt somit bei externen Ratings und den damit verbundenen Interessenskonflikten (zu dem Interessenkonflikt der Ratingagenturen aus US-Perspektive siehe Crumley 2012).

Als mögliche Werkzeuge werden neben dem Verbot von Konflikten, dem Entfernen von Bezügen auf Ratings, der Erhöhung der Haftbarkeit, organisatorische Schranken („firewalls"), Leistungs-Offenlegung, Offenlegung der Due Diligence, Erhöhung des Wettbewerbs, „staleness"-Reformen, interne Verwaltung, administrative Registrierung eben auch alternativer Geschäftsmodelle gefordert (Crumley 2012, S. iv). Das oben dargestellte Unternehmensrating von deutschen Lebensversicherungsunternehmen könnte als solches, alternatives Geschäftsmodell gelten und auf andere Branchen und Märkte erweitert werden.

Das Geschäftsmodell der Agentur RealRate unterscheidet sich in wesentlichen Punkten von demjenigen der klassischen Ratingagenturen. Das Geschäftsmodell ist komplett datenbasiert und benötigt keinen menschlichen Analysten mehr für die einzelnen Ratings. Sobald das Modell einmal durch einen Experten spezifiziert wurde, kann es auf alle Unternehmen der modellierten Industrie angewendet werden. Diese Automatisierung führt zu einer hohen Skalierbarkeit

[13] Eine „statistische" Diskriminierung kann nicht ausgeschlossen werden. Vgl. auch *Bartlett* et al. 2022.

und günstigen Ratings. Auch können so Unternehmen abgedeckt werden, die sonst keine Ratings in Auftrag geben würden. Das Geschäftsmodell basiert auch gerade nicht auf der Beauftragung, sondern es werden stets alle Unternehmen einer Branche auf Basis öffentlich verfügbarer Daten bewertet (Public Information Rating). Das Ratingmodell kann ex ante per Design so konstruiert werden, dass es den gesetzlichen Vorgaben und den sozialen Normen entspricht. Konkret heißt das zum Beispiel, dass Verzerrungen vermieden, Diskriminierung ausgeschlossen und Fairness sichergestellt werden kann. Der größte Vorteil des Ansatzes der Nicht-Beauftragung durch die einzelnen Unternehmen besteht jedoch darin, den immanenten Interessenkonflikt zu vermeiden, der dadurch entsteht, dass das Unternehmen als Kunde sein Rating kostenpflichtig beauftragt. Möchte eine Ratingagentur erneut beauftragt werden, so könnte sie verleitet sein, ein zu gutes Rating auszusprechen.

Im Gegensatz dazu wird den am besten bewerteten Unternehmen das Ratingsiegel zu Werbezwecken angeboten (wiederum vergleichbar mit dem Siegel der Stiftung Warentest für Produkttests). Nur beispielsweise den besten 25 % der Unternehmen innerhalb einer Branche wird ein Siegel ausgestellt. Es entfällt dadurch die Verzerrung, die sich daraus ergeben könnte, dass nur Unternehmen Ratings in Auftrag geben, die ein gewünschtes Rating realistischerweise erwarten können.

Dieses Geschäftsmodell basiert dann auf jährlich wiederkehrenden Zahlungen („Abonnement“) für das Recht, mit dem Ratingergebnis zu werben. Der Vorteil der damit werbenden Unternehmen besteht darin, dass sie mit einem nicht beauftragten Ratingsiegel einer unabhängigen Institution werben können. So kann das modifizierte Geschäftsmodell zusammen mit moderner KI-Technologie helfen, Interessenkonflikte zu vermeiden. Da es sich um ein externes Rating handelt, ist zusätzlich wegen der Transparenz die Vergleichbarkeit innerhalb einer Branche besonders leicht möglich.

5. Zusammenfassung

In diesem Beitrag wurde ein Verfahren zur Bewertung von (deutschen) Versicherungsunternehmen auf Basis von KI-Methoden vorgestellt. Anhand eines vorläufigen Kriterienkatalogs für industrielle Modelle der erklärbaren Künstlichen Intelligenz wurde es dann bezüglich seiner Transparenz-Eigenschaften positioniert. Ein erklärbares Modell zur Bewertung von Unternehmen dient Marktakteuren dazu, gefällte Entscheidungen, besser nachvollziehen zu können. Das Modell wurde auf deutsche Versicherungsunternehmen angewendet und garantiert die Erklärbarkeit anhand eines gerichteten Graphen, der die Ursachen und Wirkungen der relevanten Größen veranschaulicht. Ein Vorteil der Methode ist, dass Stärken und Schwächen der Unternehmen direkt über den

Graphen erklärbar sind und zu dessen Verständnis kaum fachlich-technische Expertise erforderlich ist. Anschließend wurde die Anwendung solcher Modelle auch aus Sicht der Transparenz beleuchtet; es wurde insbesondere auch untersucht, wie das Geschäftsmodell mit der Unabhängigkeit des Raters zusammenhängt und Interessenkonflikte von bestehenden Rating-Geschäftsmodellen verringert werden. Denn Transparenz ist letztlich auch eine Eigenschaft des Systems, in das eine Methode eingebettet ist. Das KI-basierte Ratingmodell wird bereits angewendet auf deutsche Lebensversicherungs- und Krankenversicherungsunternehmen, sowie Risikoversicherer und zur Bewertung der künftigen BU-Beitragsstabilität. Außerdem werden ca. 2.000 börsengelistete US-amerikanische Unternehmen aus 20 verschiedenen Branchen geratet.

Erweiterungen sind in verschiedene Richtungen denkbar: Bisher werden lediglich die quantitativen Daten aus Bilanz und Gewinn-und-Verlustrechnung des Geschäftsberichts verwendet. Künftig sollten auch Anhangangaben, verbale Darstellungen der Bilanzierungsmethodik sowie des Risiko- und Lageberichts berücksichtigt werden. Auch ESG-Daten (Environmental, Social, and Governance) werden künftig eine wesentliche Rolle spielen und das Rating erweitern, von reinen Finanzdaten hin zu einer ganzheitlichen Analyse und Bewertung. Hierbei können KI-basierte Textzusammenfassungen und Bewertungen verwendet werden. Umgekehrt ist auch eine Erweiterung um ein Texterzeugungsmodul denkbar, welches aus dem Netzwerk einen Bericht zu dem erzielten Rating jedes Versicherers erzeugt. Eine solche Sprachgenerierungs-Komponente (engl. „Natural Language Generation") könnte die Erklärbarkeit der Methode weiter verbessern. Hier gab es zuletzt enorme Fortschritte, wie zum Beispiel anhand der öffentlichen Diskussion über ChatGPT von OpenAI[14] deutlich wird. Auf der Geschäftsmodell-Seite wären Szenarioanalysen wünschenswert, die Interessenkonflikte aufzeigen und untersuchen, wer Ratings für wen erzeugt und wer dafür bezahlt. Auch die Quantifizierung des Bias, der auf der Nicht-Veröffentlichung schlechter Ratings basiert, würde wertvolle Erkenntnisse liefern.

Danksagung: Wir danken für wertvolle Rückmeldungen auf der DVfVW-Jahrestagung 2021, insbesondere dem Diskutanten Niklas Häusle.

Literatur

Adadi, A./*Berrada*, M. (2018): Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI), IEEE Access 6, S. 52138–52160, DOI: 10.1109/ACCESS.2018.2870052.

Bartel, H. (2014a): Simple Solvency – Ein Solvenzmodell für deutsche Lebensversicherer, DOI: 10.13140/2.1.2939.9041, https://www.researchgate.net/publication/267337608_Simple_Solvency_-_Ein_Solvenzmodell_fur_deutsche_Lebensversicherer [26.01.2023].

[14] https://openai.com/blog/chatgpt/ [01.02.2023].

Bartel, H. (2014b): Simple Solvency – Ein Solvenzmodell für deutsche Lebensversicherer, Vortrag, qx Club, Regionale Gruppe der Deutschen Aktuarvereinigung e.V. (DAV) für Berlin, DOI: 10.13140/2.1.1760.2560, https://www.researchgate.net/publication/267337670_Simple_Solvency_-_Ein_Solvenzmodell_fur_deutsche_Lebensversicherer [26.01.2023].

Bartel, H. (2019): Kausale Analyse von Gleichungssystemen mit strukturellen neuronalen Netzen. Technischer Bericht. DOI: 10.13140/RG.2.2.16841.26729, https://www.researchgate.net/publication/335099531_Kausale_Analyse_von_Gleichungssystemen_mit_strukturellen_neuronalen_Netzen [26.01.2023].

Bartel, H. (2020a): Causal Analysis – With an Application to Insurance Ratings, https://www.researchgate.net/publication/339091133_Causal_Analysis_-_With_an_Application_to_Insurance_Ratings [26.01.2023].

Bartel, H. (2020b): Explainable Artificial Intelligence (XAI) in Ratings, https://www.researchgate.net/publication/344992217_Explainable_Artificial_Intelligence_XAI_in_Ratings [26.01.2023].

Bartel, H. (2020c): Causing: Causal Interpretation using Graphs, https://www.researchgate.net/publication/341878489_Causing_CAUSal_INterpretation_using_Graphs [26.01.2023].

Bartel, H. (2020d): RealRate Expert System Life Insurance, https://realrate.ai/download/publications/RealRate%20Expert%20System%20Life%20Insurance.pdf [26.01.2023]

Bartel, H. (2023): Finanzstärke-Ratings deutscher Versicherer mittels künstlicher Intelligenz. In: Zeitschrift für Versicherungswesen, 74(2), S. 42–51.

Bartlett, R./*Morse*, A./*Stanton*, R./*Wallace*, N. (2022): Consumer-lending discrimination in the FinTech Era. In: Journal of Financial Economics 143 (1), S. 30–56, DOI: 10.1016/j.jfineco.2021.05.047.

Breiman, L./*Friedman*, J. H./*Olshen*, R. A./*Stone*, C. J. (1984): Classification and regression trees, Monterey, CA, USA: Wadsworth & Brooks/Cole, DOI: 10.1201/9781315139470.

Burkart, N./*Huber*, M. F. (2021): A Survey on the Explainability of Supervised Machine Learning, Journal of Artificial Intelligence Research 70, S. 245–317, DOI: 10.1613/jair.1.12228.

CFO-Forum (2009): Market Consistent Embedded Value (MCEV)-Principles, http://www.cfoforum.eu/downloads/MCEV_Principles_and_Guidance_October_2009.pdf [26.01.2023].

Cohen, W. W. (1995): Fast Effective Rule Induction, Proc. 12th Int. Conf. Machine Learning (ICML).

Crumley, D. G. (2012): Credit Rating Agencies and Conflicts of Interest, Inauguraldissertation (J.D. thesis), University of Texas at Austin, Austin, TX, USA.

Dierkes, S./*Sümpelmann*, J. (2019): Digitalisierte Peer-Group-Bestimmung und Beta-Anpassung. In: Ballwieser, W./Hachmeister, D. (Hg.) (2019): Digitalisierung und Unternehmensbewertung, S. 173–192. Stuttgart: Schäffer-Poeschel.

Dziugaite, G. K./*Ben-David*, S./*Roy*, D. M. (2020): Enforcing interpretability and its statistical impacts: trade-offs between accuracy and interpretability, https://arxiv.org/pdf/2010.13764.pdf [26.01.2023].

Europäische Kommission (2020): Weißbuch zur Künstlichen Intelligenz – Ein europäisches Konzept für Exzellenz und Vertrauen, Brüssel, https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_de.pdf [26.01.2023].

Europäische Kommission (2021): Regulating credit rating agencies, Brüssel, https://ec.europa.eu/info/business-economy-euro/banking-and-finance/financial-supervision-and-risk-management/managing-risks-banks-and-financial-institutions/regulating-credit-rating-agencies_en [26.01.2023].

Hochrangige Expertengruppe für Künstliche Intelligenz (HEG-KI) (2018): Ethik-Leitlinien für eine vertrauenswürdige KI, Brüssel, https://op.europa.eu/s/oVfc [26.01.2023].

GitHub (2021a): A Real World Example: Education and Wages for Young Workers, https://github.com/realrate/Causing/blob/develop/docs/education.md [26.01.2023].

GitHub (2021b): Causing: CAUSal INterpretation using Graphs, https://github.com/realrate/Causing.

Gründl, H./*Kraft*, M. (Hg.) (2019): Solvency II – Eine Einführung. Grundlagen der neuen Versicherungsaufsicht. 3. Aufl. Karlsruhe: VVW.

Heskes, T./*Sijben*, E./*Bucur*, I. G./*Claassen*, T. (2020): Causal Shapley Values: Exploiting Causal Knowledge to Explain Individual Predictions of Complex Models, NeurIPS 2020, https://arxiv.org/abs/2011.01625 [26.01.2023].

Holland, C. P./*Kavuri*, A. (2021): Artificial intelligence and digital transformation of insurance markets. In: The Capco Institute – Journal of Financial Transformation H. 54 (11/2021), S. 104–115, https://www.capco.com/-/media/CapcoMedia/Capco-2/PDFs/Capco-Journal-54AI-and-Digital-Transformation-of-Insurance-Markets.ashx [26.01.2023].

James, G./*Witten*, D./*Hastie*, T./*Tibshirani*, R. (2017): An Introduction to Statistical Learning: with Applications in R, New York, NY, USA: Springer.

Kingma, D. P./*Ba*, J. (2014): Adam: A Method for Stochastic Optimization, ICLR 2015 conference paper, https://arxiv.org/abs/1412.6980 [26.01.2023].

Kokina, J./*Davenport*, T. H. (2017): The emergence of artificial intelligence: How automation is changing auditing. Journal of Emerging Technologies in Accounting 14(1), S. 115–122.

Kurmann, S. (2023): KI in der Versicherungsbranche: Wenn Science-Fiction auf Realität trifft, https://www.handelszeitung.ch/insurance/kunstliche-intelligenz-fur-versicherungen/ki-in-der-versicherungsbranche-wenn-science-fiction-auf-realitat-trifft-564992 [01.02.2023].

Leidner, J. L. (in Vorbereitung): A Survey of Ethical Problems of Artificial Intelligence.

Lossos, C./*Geschwill*, S./*Morelli*, F. (2021): Offenheit durch XAI bei ML-unterstützten Entscheidungen: Ein Baustein zur Optimierung von Entscheidungen im Unternehmen? HMD Praxis der Wirtschaftsinformatik 58, S. 303–320.

Lundberg, S. (2020): Vortrag „The Science behind SHAP“, https://www.youtube.com/watch?v=-taOhqkiuIo [26.01.2023].

Lundberg, S. M./*Lee*, S.-I. (2017): A unified approach to interpreting model predictions, Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS). Red Hook, NY, USA, S. 4768–4777.

Oletzky, T./*Reinhardt*, A. (2022): Herausforderungen der Regulierung von und der Aufsicht über den Einsatz Künstlicher Intelligenz in der Versicherungswirtschaft. In: Zeitschrift für die gesamte Versicherungswissenschaft 111 (4), S. 495–513. DOI: 10.1007/s12297-022-00541-4.

Owens, E./*Sheehan*, B./*Mullins*, M./*Cunneen*, M./*Ressel*, J./*Castignani*, G. (2022): Explainable Artificial Intelligence (XAI) in Insurance. In: Risks 10 (230), S. 1–50, DOI: 10.3390/risks10120230.

Quinlan, J. R. (1986): Induction of Decision Trees, Machine Learning 1(1): S. 81–106.

Rall, L. B. (1981): Automatic Differentiation: Techniques and Applications. Lecture Notes in Computer Science. 120. Springer.

Ribeiro, M. T./*Singh*, S./*Guestrin*, C. (2016): „Why Should I Trust You?": Explaining the Predictions of Any Classifier, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Francisco, California, USA, S. 1135–1144, DOI: 10.1145/2939672.2939778.

Rumelhart, D. E./*Hinton*, G. E./*Williams*, R. J. (1986): Learning representations by back-propagating errors, Nature 323 (6088): S. 533–536, DOI 10.1038/323533a0.

Samek, W./*Müller*, K.-R. (2019): Towards Explainable Artificial Intelligence. In: Samek, W./Montavon, G./Vedaldi, A. (Hg.): Explainable AI. Interpreting, explaining and visualizing deep learning (Lecture Notes in Computer Science 11700 /Lecture Notes in Artificial Intelligence), S. 5–22, DOI: 10.1007/978-3-030-28954-6.

Sellhorn, T. (2020): Machine Learning und empirische Rechnungslegungsforschung: Einige Erkenntnisse und offene Fragen, Schmalenbachs Z. betriebswirtsch. Forsch. 72, S. 49–69, DOI: 10.1007/s41471-020-00086-1.

Shapley, L. S. (1951): Notes on the n-Person Game – II: The Value of an n-Person Game, Technical Report RM-670, Santa Monica, CA, USA: RAND Corporation.

Simpson, E. H. (1951). The Interpretation of Interaction in Contingency Tables, Journal of the Royal Statistical Society, Series B. 13: S. 238–241.

Stuwe, A./*Weiß*, M./*Philipper*, J. (2012): Ratingagenturen: Sind sie notwendig, überflüssig, notwendiges Übel oder schädlich?, Bonn: Friedrich-Ebert-Stiftung, https://library.fes.de/pdf-files/managerkreis/09647.pdf [26.01.2023].

Van Hulle, K. (2019): Solvency requirements for EU insurers. Solvency II is good for you. Cambridge: Intersentia.

Wermter, S./*Sun*, R. (Hg.) (2000): Hybrid Neural Systems. Springer-Verlag, Heidelberg.