Journal for the entire insurance science Duncker
& Humblot, 12165 Berlin

Explainable artificial intelligence using the example of ratings German life insurance company

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

Summary

Artificial intelligence (AI) is used in practice for decision-making (Lossos, Geschwill and Morelli 2021), increasingly also in the insurance industry. However, this requires trust in the various AI methods, especially when evaluating companies ("ratings"). Such trust is formed when decision makers and users can form mental models of the system and understand the output of the system. AI must therefore be explainable, a pure black box is insufficient even if the system is of high quality. "Explainable AI" (XAI) deals with the development of AI models that can be understood by humans (Adadi and Berrada 2018; European Commission 2020). In this article, desirable properties of industrial AI systems are examined - especially with regard to explainability - and presented and visualized using the application example of ratings of (German) life insurance companies. In addition to XAI as an aspect of technical acceptance, the interaction between the business model and customer acceptance in ratings of German life insurance companies is examined. Business figures of German life insurance companies are often considered to be non-transparent. This also applies if the HGB accounting is supplemented by the Solvency II reports on the solvency and financial position (SFCR). In this respect, the discussion of explainable AI methods in this context is a useful contribution to evaluation practice.

Prof. Dr. Mirko Kraft,
Coburg University of Applied Sciences, Faculty of Economics, Friedrich-Streib-Strasse 2, 96450 Coburg, Germany
Email: mirko.kraft@hs-coburg.de

dr Holger Bartel, RealRate GmbH
Email: holger.bartel@realrate.ai

Prof. Dr. Jochen L. Leidner,
Coburg University, Faculty of Economics and University of Sheffield, Department of Computer Science
Email: leidner@acm.org

DOI https://doi.org/10.3790/zverswiss.2023.04.Bartel.etal | Generated on 2023-02-18 13:06:43

Holger Bartel, Mirko Kraft and Jochen L. Leidner

Abstracts

Artificial intelligence (AI) is already used for decision-making in practice (Lossos/ Geschwill/ Morelli 2021), also increasingly 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. 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. 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.

*tags:*explainable artificial intelligence (AI), ratings, life insurance companies

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

1 Introduction

1.1 Motivation

In many areas and companies, the application of artificial intelligence (AI) is no longer just the future, but already a reality - including in the insurance industry (see e.g. Oletzky/Reinhardt 2022, Kurmann 2023). In some cases, great expectations are attached to the use of AI methods. B. Expected benefits from the automation of processes. With evaluation and decision-making processes, which were otherwise reserved for people, there is concern that the results are no longer comprehensible to the extent. This applies in particular to ratings of companies ("company ratings"), which are intended to make the financial solidity of different companies comparable.

Ratings automatically generated by artificial intelligence (AI) are used by market players to make their decisions. Therefore, they want to be able to understand AI systems (Samek/Müller 2019, p. 8). Explainability as an ethical principle of AI (HEG-KI 2018, p. 16) is therefore a prerequisite for transparency, and in particular for company ratings based on AI methods. In general, high demands are placed on company ratings

DOI https://doi.org/10.3790/zverswiss.2023.04.Bartel.etal | Generated on 2023-02-18 13:06:43

Changes have been made, including regulatory requirements to justify the use of external ratings (for rating regulation, see e.g. European Commission 2021).

However, trust in the underlying processes is particularly important when using AI methods, so that a wide range of efforts are made in practice and in research with AI systems[1] to create sufficient transparency and to ensure the explainability ("explainable AI", XAI) of the results. In addition to mathematical methods that, for example, analyze the sensitivities of the results to changes in the input data and other post hoc methods ("ex post explainability"), one approach is to focus on explainability as early as the design of the algorithms or the selection of the AI method to be observed ("explainability by design").

In order to achieve explainability for AI-based company ratings, such an approach is chosen in a use case - in contrast to the subsequent interpretation of any AI systems. In contrast to typical AI applications with big data and large neural networks (so-called deep learning), relatively small, structural networks are used here, which are given in the form of equations and represent expert knowledge. This guarantees the ability to explain using a directed graph that illustrates the causes and effects of the relevant variables (Bartel 2019). Users therefore need less professional and technical expertise to understand it.

Specifically, the example of the rating of German life insurers shows how a complex annual report, in particular the accounting figures, is analyzed (Sellhorn 2020). The strengths and weaknesses of the companies can thus be explained directly in a market comparison using the graph. The extended supervisory risk-based reports to the public, the so-called Solvency and Financial Condition Reports (SFCR), which have been mandatory for insurers since 2016 (on Pillar 3 of Solvency II cf. Gründl/ Kraft, for example), are also used as data for machine learning 2019, Van Hulle 2019).

It is also examined how the business model and the independence of the rater are related by using the scenarios "commissioning", the traditional model, afflicted with conflicts of interest (Crumley 2012; Stuwe et al. 2012), and "public rating", i.e. without being commissioned by the rated companies

[1] Under an AI system can be understood (*Holland/ Kavuri* 2021, p. 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, contrastingly compared and also differentiated from alternative approaches from the literature.

In the following article, the core of the transparency of ratings of German life insurance companies is presented through the use of explainable AI. Transparency and explainability are generally desirable criteria for AI systems (cf. also e.g. Oletzky/Reinhardt 2022, p. 505 f.). Expert knowledge and AI methods are combined here and also used in such a "hybrid" model to visualize the results, which also operationalizes the explainability for the users of the company ratings.

1.2 Overview

The article is structured as follows: After an introduction to the explainability of artificial intelligence (2.1) and an introductory link between ratings and artificial intelligence (2.2), a case study follows: The application of AI methods for company ratings, especially for German life insurance companies (3.) . The use case is then analyzed in more detail with regard to transparency and the business model (4.). The article concludes with a summary and an outlook on possible follow-up research projects (5.).

2. Explainable Artificial Intelligence

2.1 Approaches to Explainability

Explainable model types of machine learning can be distinguished on the one hand in *"White Box models*(also "ex ante" models, since the explainability is given from the outset), which are inherently explainable and therefore do not require any special discussion. Examples of this are linear models (e.g. linear or logistic regression (James et al. 2017)), decision trees (CART, Breiman et al. 1984), ID3 (Quinlan 1986) and rule systems (e.g. repeated incremental pruning to Produce Error Reduction (RIPPER, Cohen 1995)). On the other hand there is *"Black Box models*. Their ability to be explained can only be achieved “ex post” by “grafting on” additional mechanisms, if at all. Examples are “deep”, i.e. neural networks equipped with several non-input/output layers (so-called deep learning).

Explainable machine learning is now particularly concerned with generating explanations for such black box models (Adadi/Berrada 2018). This is particularly desirable because black box models currently provide the best predictions (see Figure 1). The *local*has data explainability

DOI https://doi.org/10.3790/zverswiss.2023.04.Bartel.etal | Generated on 2023-02-18 13:06:43

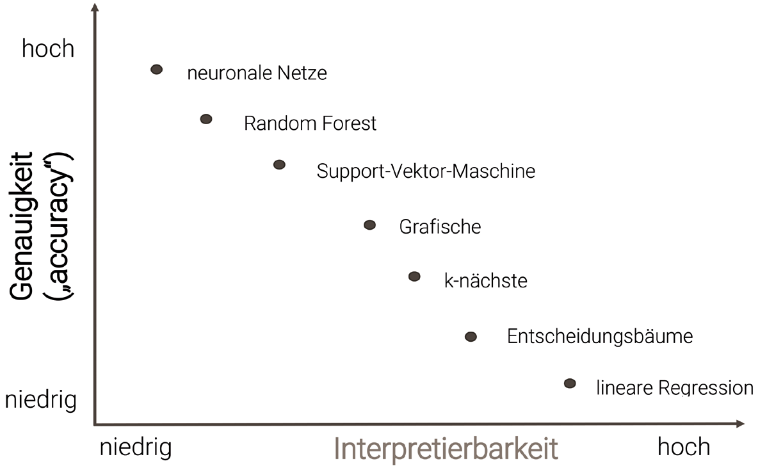

Fig. 1: Trade-off between accuracy and interpretability[2]

to explain why a given input x leads to a given output y while the *global* Model explainability aims to explain how a given model works as a whole. explainability *by design* tries to construct only such procedures that *ex ante* are explainable. Hybrid artificial intelligence (“hybrid AI”) is the combination of symbolic and connectionist methods such as neural networks (Wermter/Sun 2000), which increase the explainability by retaining symbolic aspects in a model.

An approach of *ex post* Explainability consists in inducing secondary proxy or surrogate models in addition to a given primary black box model; the surrogate model aims solely at generating explanations after the fact, while the primary model determines the actual prediction. Fidelity is relevant here: how closely do the predictions of the black box model and the surrogate model match? A simple approach is to induce a decision tree on neural network outputs, which are used as the gold standard. Unfortunately, such a simple method only results in low meaningfulness, fidelity and accuracy, but can be improved by applying regularization (Burkart/Huber 2021).

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

One of the earliest key works in the XAI environment is Ribeiro, Singh and Guestin's work Why Should I Trust You? Explaining the Predictions of Any Classifier, which introduced the LIME technique (short for Local Interpretable Model-agnostic Explanations, Ribeiro/Singh/Guestin 2016). The authors induce a white box classifier on the predictions of a given black box classifier as follows: LIME optimizes the dual criterion

$$\xi(x) = \text{argmin} -(f, G, \pi_x) + \omega(G)$$

ie simultaneously become the sum of the squared loss L and a complexity measure minimized in order to get explanations that work well (“locally faithful”), but which are also interpretable (“low complexity”). G is a "model" that characterizes whether there is an explanation for a dimension. To the local behavior of f to learn while the interpretable inputs vary, can $-(f, G, \pi_x)$ be approximated by using a random number of random samples weighted π_x is selected. Sample instances are procured at x_0, by dividing the elements that are not null, but are very much in x included are sampled uniformly at random. After that, labels with the existing classifier are obtained for these classes and the data classified in this way serve as training data for the new surrogate classifier. The points close to the point to be explained are weighted higher in our new, weighted linear model:

$$--(f, G, \pi_x) = \text{uh}_{e.g, e.g¢ \hat{I} Z} \, \pi_x(e.g)(f(e.g) - G(e.g¢))^2 \quad ,$$

whereby $\pi(e.g) = \exp\left(\frac{-D(x, e.g)^2}{\sigma^2}\right)$ and D en distance measure (e.g. similar to a cosine ability for text); this also represents values within Î[0;1] sure.

LIME generates locally trusted, linear explanations. A complexity measure Omega contains a bound K (e.g. B. the number of explanatory words in a text classification task) to ensure human interpretability.

Shapley values (Shapley 1951), a discovery originally from game theory (named after Lloyd Shapley, Nobel Prize in Economics 2012), offer ways to ex post explain nonlinear models. For a model trained with a set of features in the form of a function over a coalition of players, they offer an additive way to calculate which features contribute how much to a decision.

DOI https://doi.org/10.3790/zverswiss.2023.04.Bartel.etal | Generated on 2023-02-18 13:06:43

SHAP (short for "SHapley Additive exPlanations", Lundberg/Lee 2017) is a method that applies Shapley values to explainability. Starting from the a priori probability for a class, the features and their additive or subtractive effect on the overall decision are considered individually and sequentially. Put simply, the SHAP value of a feature is the difference between the mean of a feature and the partial dependency graph that results when we change a feature, but where the order matters.[3]

E

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

Fig. 2: Post hoc methods (model-agnostic)[4]

When considering any type of automation and in particular AI processes, which also automate cognitive, i.e. specifically human, abilities, it is essential to illuminate the ethical side. Ethical problems in this context include questions of the morality of automation in general, of justice and transparency (Leidner (previous year)). Is it generally moral to automate an activity? For example, there is the loss of jobs when the work of human analysts is replaced by a computer program that costs less, never sleeps, takes vacations, gets sick, etc. and can also calculate rating models faster. The question

[3] Please refer *Lundberg's* Lecture “The Science behind SHAP”: https://www.youtube.com/watch?v=-taOhqkiuIo [02/01/2023].

[4] Source: *Baniecki/ bicec* 2021

of transparency is: "Can humans understand why a specific decision was made?" The transparency of ratings by "human hands", without AI methods, has traditionally been problematic due to the likelihood of conflicts of interest, since the ratings are id R. were commissioned and paid for.

Automation offers protection against statistical distortion of individuals (bias), i.e. more objectivity. Traditional, particularly deep neural networks (i.e. those with more than one intermediate layer) are primarily black box models that classify very well on the one hand, but elude analysis on the other. The model proposed here provides ex ante or white box transparency “by design” (*by design*). In the following, first criteria for ethical AI are proposed, which do not claim to be exhaustive, but hopefully can be helpful when examining the moral side of models.[5]

As already mentioned, transparency is important. Structure and assumptions should be disclosed for models (which variables can be influenced and which cannot). A model should be legal, i.e. in accordance with laws and regulations. A model should also be free from discrimination (e.g. based on gender or religion). In addition, a model should also deliver the same output again with the same input data (reproducibility). If possible, it is desirable to make the program code of models available as open source software for review. Such a publication enables not only the reproducibility but also the understanding of the functionality and the adaptability to new questions.

When explaining a model, there should be clarity about causality or correlation: In the subsequent explanation of AI decisions, key figures such as correlation or Shapley values   are used to uncover the main causal relationships. However, causality cannot be read from data alone, this requires input from a human's technical expertise - a strong argument for explainability by design.

2.2 XAI and Ratings

Ratings are the evaluation of companies, (capital market) products or people, typically in terms of financial strength. A special branch relates to the assessment of creditworthiness using

---

[5]For a catalog of criteria for trustworthy AI, see HEG-KI 2018, for example. For explainability in the insurance sector, see in-depth *owens* et al. 2022.

DOI https://doi.org/10.3790/zverswiss.2023.04.Bartel.etal | Generated on 2023-02-18 13:06:43

Categorization, such as an AAA rating as the top rating from Standard & Poor's. Such Credit Rating Agencies (CRAs) are regulated in the EU by the Credit Rating Agencies Regulation (CRA Regulation) and are overseen by financial regulators.

A credit rating is an assessment of creditworthiness expressed in terms of rating categories. They are issued on a professional basis, are tied to a specific financial instrument, obligation or issuer, require analytical input from rating analysts, and are publicly announced or distributed by subscription.

If a credit rating is derived solely from the aggregation and presentation of data based on a pre-created statistical model and no significant rating-specific analytical inputs are included in the rating, the product is considered a "credit score" and is not subject to regulatory regulation. Therefore, the fully automatic ratings shown later do not fall under the credit rating regime, especially as long as they are not linked to a rating category.

Credit ratings help investors and lenders understand the risks associated with a particular investment or financial instrument. In the period leading up to the 2008 financial crisis, rating agencies failed to properly assess the risks of some of the more complex financial instruments. In response, the European Commission has strengthened the regulatory and supervisory framework for credit rating agencies in the EU to restore market confidence and increase investor protection. Since the end of 2009, rating agencies have had to be registered and are supervised by the competent national authorities.[6] In addition, rating agencies must avoid conflicts of interest and have sound rating methods and transparent rating activities.

But even independently of legal regulation, there is a need to ensure that such valuations, which can have a major economic impact on the assessed units, are fair, reliable and explainable. For example, there are considerations to grant the (private) customer the legal right to have the rating result explained. This is particularly important if the loan request is rejected. Regulation and customer protection will be key drivers for explainable rating models in the future.

---

[6] According to Art. 2 Para. 3 of Regulation EC/1060/2009, a rating agency must apply for registration in order to be recognized as an external rating agency in accordance with Directive 2006/48/EC.

DOI https://doi.org/10.3790/zverswiss.2023.04.Bartel.etal | Generated on 2023-02-18 13:06:43

AI models typically offer very good performance with a broad data basis, which has contributed to their great popularity. However, they generally cannot explain why the decision was made. Worse, they could be based on discrimination, e.g. in relation to gender, age, ethnicity, etc. It is therefore crucial to provide a valid, clear and legally compliant explanation of the rating decision that is understandable for companies and private users. Hybrid models are a promising way to achieve this.

In order to make rating results explainable, there is either the option of explaining a black box model afterwards or making the model explainable by design (2.1). Explainability by design is typically achieved by translating general technical and expert knowledge into graphic structures (knowledge graphs) and integrating them into the model right from the start. One then also speaks of hybrid models, since domain knowledge is linked with machine learning based on data.

On the one hand, external rating knowledge ensures later explainability and, on the other hand, the estimation is simplified because known knowledge no longer has to be re-learned from data. And especially in the case of smaller amounts of data, i.e. not big data, domain knowledge brings a structure that reduces the number of free parameters, which can then be estimated more precisely.

Causal structures must be implemented by experts. It is well known that correlations can be derived from data, but not causal relationships (Simpson 1951). Causal relationships of a specific use case must therefore be defined by an expert. For example, if banks want to assess the creditworthiness of private customers, financial wealth is a major factor in creditworthiness and should be allowed as such. Gender, on the other hand, should not affect creditworthiness, at least not directly. However, to ensure an overall high percentage of correct classifications, there may be an indirect effect, e.g. B. when gender has an impact on income, which in turn affects financial strength.

Predefined expertise and business rules can be represented in a causal graph showing what causes affect specific variables. The exact value of the effect can then be determined using machine learning. This is where the hybrid approach comes into play. This leads to highly structured models. So the neural network is not completely free to adapt arbitrary relationships to the given variables. Instead, many relations that represent the given structure are explicitly excluded.

There is basically a trade-off between performance and explainability: [7]Deep AI models typically offer a very high forecast quality or reproduction of observed data (see Fig. 1). This applies at least as long as large amounts of data are available. This is also the main reason why they are currently being used so successfully. A disadvantage in the application is the lack of interpretability. This can be increased by specifying structural relationships, but this may be associated with a certain decrease in the quality of the forecast. In the case of little data, as in the case of company ratings, a “frugal” model is required anyway so that the small degrees of freedom can be estimated or learned more precisely and to avoid the problem of overfitting.

Company rating analyzes are essentially based on accounting data. In the context of AI and accounting, Sellhorn (2020) points out that a lack of transparency costs trust. In the context of corporate reporting and auditing, Kokina and Davenport (2017) state: "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.” Dierkes and Sümpelmann (2019, p. 190) also point out that “ must not become a black box as a result of digitization", which "lacks a sufficient theoretical foundation". The following use case shows how this criticism can be countered in corporate ratings.

3. Use Case: Corporate Ratings

3.1 A hybrid rating model with expert knowledge

A rating of all German life insurers is used to demonstrate how the explainability of an AI rating is ensured. A hybrid AI model is used, which uses ex ante expert knowledge and estimates the resulting structural neural network using classic machine learning methods.

Input data are the published balance sheet and the associated profit and loss account from the annual reports. In addition to HGB and IFRS requirements, the data is standardized by the requirements of the Accounting Ordinance for Insurance Companies (RechVersV). In addition, the data

[7]See also *Oletzky/ Reinhardt*2022, p. 505.

DOI https://doi.org/10.3790/zverswiss.2023.04.Bartel.etal | Generated on 2023-02-18 13:06:43

Quality is high, since the annual financial statements must be audited by an auditor and the companies are supervised.

The company rating presented as an example is based on the approach of the rating agency RealRate. The rating includes all (active) German life insurers. These are a particularly challenging example, as the complexity of the business model is quite high and the balance sheet data has numerous industry-specific features that are related to the products (e.g. the guaranteed discount rate) or the business model (e.g. profit participation). The need is all the greater for both analysts and customers to be able to quickly grasp the essential relationships, strengths and weaknesses.

The model should only use externally available data in order to carry out a revaluation from the book value balance sheet to a market value balance sheet, similar to what is required by the Solvency II supervisory regime.[8th] For this reason, assumptions about the interest rate sensitivity of the liabilities side (liability duration) must be included in the model. However, the original approach by Bartel (2014a) still partly requires internal company information in order to be able to determine the insurer's risk. For the application considered here, the determination of the economic equity, i.e. the numerator of the solvency ratio, is therefore limited and as a reference value, i.e. the actual denominator of the solvency ratio, only the balance sheet total is used as a yardstick for simplification.

Bartel's (2014a, 2014b) approach proposes a highly simplified economic solvency model for life insurers, which reflects the main regulatory solvency rules. It takes into account the special features of the contract, in particular the customer's surplus participation in past profits (via the provision for premium refunds) and in future profits.

The model should only be as complex as necessary to map the business model of German life insurers, but also as simple as possible to ensure that it can be explained by design. The aim is to avoid complex cash flow projections or stochastic simulations and thus improve the transparency of the mode of operation. Only a few input parameters and variables are required in the RealRate model. However, due to the asymmetrical business model, these few variables are linked to one another in a strongly non-linear manner. A closed formula is also proposed to determine the value of the so-called guarantees and options. This arises when the company is valued, since the customer only participates in positive profits due to the guarantee granted to him

[8th] For the essential principles of the rating approach, see *barbel* 2014a.

DOI https://doi.org/10.3790/zverswiss.2023.04.Bartel.etal | Generated on 2023-02-18 13:06:43

but not in losses. This effect is evaluated using a closed option price formula, the so-called “buffer put” (see Bartel 2014a).

The model used is a holistic company model and not just a weighting of interesting key figures. The initial values for the effects between the variables result first from the expert system given in the form of equations. These effects are then selected in machine learning in such a way that the published solvency ratios (without supervisory transitional measures) are explained as well as possible in accordance with Solvency II. Instead of estimating an unrestricted neural network, in our case an optimization is carried out under (causal) constraints. The modeler determines which coefficients can be freely chosen. Other weights, on the other hand, remain unchanged, for example if they come from a pure definition equation.

The weights of the network correspond to the effects sought, which measure the quantitative effects between the variables. They serve to explain. While only overall effects in the sense of a total derivation are determined ex post in deep neural networks, direct effects in the sense of partial derivations can also be determined in causal models.

The graph resulting from the RealRate rating model resembles a neural network (see Fig. 6). The nodes or neurons correspond to the variables used in the model and the edges or weights to the effects sought. The individual neurons of the structured network are actually interpreted, which is not possible with deep networks. The difference to a classic neural network is, on the one hand, that not all variables of a layer are linked to each other, but many variables are straight due to the given causal structure *not* are linked together. On the other hand, this structural neural network is much smaller than typical deep neural networks. This is exactly what makes explainability possible.

The methodological peculiarities of the explainable AI approach consist in the fact that a causal model is specified. This structured model ensures later explainability by design and can also explicitly include regulatory requirements. In contrast to typical deep learning approaches, each node, i.e. each neuron, can be interpreted because it corresponds to a model variable. These are typically latent, i.e. not observable. Algebraic formulas are automatically derived for the individual sensitivities, i.e. the effects of each individual variable on the variable of final interest. This also improves the stability of the backpropagation and the optimization algorithm. For supervised learning, i.e. the estimation of the model weights represented by the directed edges, some

DOI https://doi.org/10.3790/zverswiss.2023.04.Bartel.etal | Generated on 2023-02-18 13:06:43

variables to be observable. This approach also enables significance testing of each individual effect. Finally, above all, the simple graphical representation of the relationships is possible.

The entire company model comprises a total of 32 equations (Bartel 2020d). Some equations are pure definitions, such as:

Equity capital ← subordinated liabilities
+ HGB equity
+ Participation rights

To express the direction (causality), a directional arrow ("←") is used instead of the equals sign ("="), as is usual in a mathematical equation. This results in the HGB equity as an output and the three inputs are:

1. the subordinated liabilities,

2. HGB equity without profit participation rights and subordinated liabilities and

3. the participation rights.

They are simply added up and together result in HGB equity. In this case, the three direct effects are simply one: they correspond to the partial derivatives of the output variables (equity) with respect to the three input variables. The corresponding partial graph thus looks like in Fig. 3:

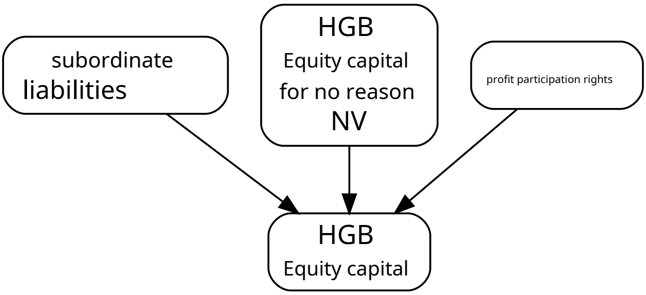

Fig. 3: Partial graph for HGB equity

Equity is thus an endogenous and latent (unobservable) model variable, while the other three variables are exogenous and manifest (observable). Each model equation thus causally determines a model variable.

Other equations are used for revaluation from book value balance sheet to market value balance sheet as shown in Fig. 4:

market value of investments ← active valuation reserves + book value investments

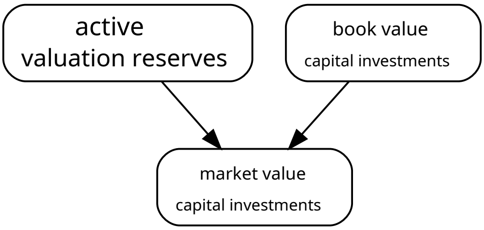

Fig. 4: Partial graph for the market value of investments

And still other equations are used to model the content of financial strength, as shown in Fig. 5:

economic equity ratio ← economic equity/HGB balance sheet total

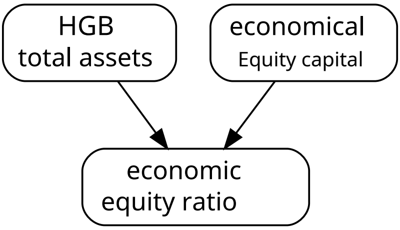

Fig. 5: Subgraph for the economic equity ratio

The economic equity ratio is the final variable of interest and is therefore the last model equation and thus last (bottom) variable in the explanatory graph. The input variables from the balance sheet and profit and loss account are shown in the graph in Fig. 6. Input quantities are also known as exogenous quantities and do not have incoming arrows in the graph. The output variables, on the other hand, are determined by the model and are also referred to as endogenous variables. In addition, most endogenous variables are usually latent, i.e. unobservable. However, there must be at least one observable endogenous variable so that the model can be checked using machine learning and the coefficients can be adjusted in such a way that reality is explained as well as possible. These coefficients used result from the optimized model by simply taking the appropriate derivatives. From all equations together results

DOI https://doi.org/10.3790/zverswiss.2023.04.Bartel.etal | Generated on 2023-02-18 13:06:43

A general graphic structure then arises, which is used directly for the calculation. In our approach, there are two special features compared to the commonly used networks:

1. Structured neural network: The network is highly restricted: relationships between the variables are only allowed if they are connected. However, most of the variables are not linked to one another. The structure is thus determined by these "non-connections" (mathematically zero restrictions). In addition, the direction of the arrow determines in which direction the causal relationship flows. In the case of A ← B, B can influence A, but not vice versa.

2. Shallow Learning/Small Data: In contrast to the commonly used deep learning with big data, only a very small network is used here. On the one hand, this has the advantage that only a very small amount of data is required for the estimation. For example, an entire industry with 100 companies, each with 100 balance sheet ratios, can be rated if only these 10,000 data are available. The less data you have, the smaller the models have to be so that you can estimate these few degrees of freedom in a statistically valid way. Other tasks such as image recognition by AI typically require millions of data points.

Parameters such as the profit participation level or the tax rate are later estimated by machine learning in such a way that the observed Solvency II solvency rates are explained as well as possible. The calibration of the individual effects in this structure is optimally determined by machine learning.[9]

3.2 The explainable causal graph

This structure is filled with different values and colors for each individual rated company. Table 1 shows the RealRate financial strength ranking of German life insurance in 2022 (based on the 2021 annual report data).[10] HUK-COBURG takes first place in the ranking with an economic equity ratio of 20.07%. Allianz Leben ranked 38th out of a total of 60 life insurers examined and has an economic equity ratio of 8.23%. This may seem surprising at first, since Allianz is the clear market leader in the German life insurance sector. the corporate

[9] For further details on the application of explainable artificial intelligence in the rating area see Bartel 2020a, 2020b, 2020c.

[10] This and other ratings are freely available online: https://realrate.ai/rankings [02/01/2023].

DOI https://doi.org/10.3790/zverswiss.2023.04.Bartel.etal | Generated on 2023-02-18 13:06:43

Table 1

RealRate financial strength ranking of German life insurers 2022

rank	*life insurer*	*economic equity ratio*
1.	HUK COBURG	20.07%
2.	PRC	18.20%
3.	Bavarian officials	17.20%
4.	BL the Bavarian	15.38%
5.	Heidelberger	14.63%
...		
36	HDI	8.42%
37	DEVK	8.38%
38	*alliance*	*8.23%*
39	NUREMBERG	8.10%
40	R + V	7.34%
...		

However, size in itself is not a quality feature in the AI financial strength ranking. On the contrary, it is clear that the top 5 are occupied by rather smaller life insurers with very good financial strength. Financial strength, measured as economic equity in relation to total assets, ranges from around –11% to around 20% in the market.

The relative strengths of the alliance are shown in green in Fig. 6, relative weaknesses are shown in red. The values in the nodes quantify the effect of each variable on financial strength compared to the market mean. It should be noted that this graph specifically represents the Allianz company; For the other companies, the structure of the graph is identical, but the individual values, i.e. the effects on financial strength, are of course different. In addition, the causal graph for each insurer is presented in such a way that the significant effects are highlighted, while insignificant effects are not shown at all. All quantitative effects shown are those on the final variable of interest, namely the economic equity ratio shown at the bottom of the graph.

The economic equity ratio, as indicated in the ranking table, is 8.23% and is thus just 0.652 percentage points below the market average of 8.9%. The greatest strength of Allianz Leben is the low promised interest rate (average standard interest rate), which is estimated indirectly from the available balance sheet data. In the phase of low interest rates, Allianz started earlier than other insurers to reduce the burden of guaranteed interest rates,

DOI https://doi.org/10.3790/zverswiss.2023.04.Bartel.etal | Generated on 2023-02-18 13:06:43

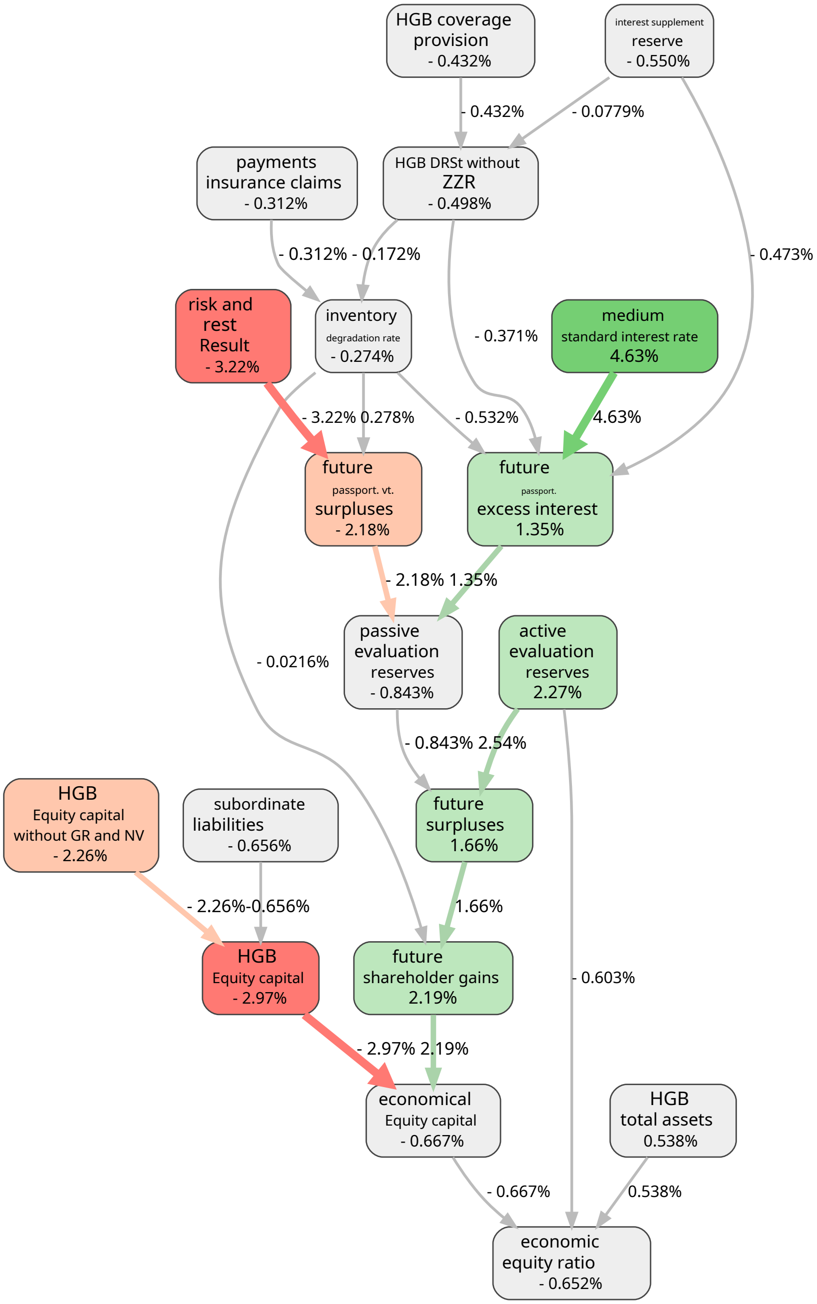

Fig. 6: RealRate financial strength analysis Allianz Lebensversicherung

DOI https://doi.org/10.3790/zverswiss.2023.04.Bartel.etal | Generated on 2023-02-18 13:06:43

by introducing additional modern products with lower interest rate guarantees. This strategy has strengthened the company's financial strength in the long term: the economic equity ratio has risen by 4.63 percentage points compared to the market average (see Fig. 6). Allianz is also successful in capital investments. It has extensive active valuation reserves. This means that the market values of the investments are higher than the conservative book values in the HGB balance sheet. This increases their economic equity ratio by 2.27 percentage points compared to the average of all life insurers. The actuarial interest rate and valuation reserves will be realized together in the form of future surpluses, which are recognized in the RealRate valuation model. This also corresponds to the logic of the Solvency II supervisory regime.

A relative weakness of Allianz, on the other hand, is the below-market risk and other result. This corresponds to below-average underwriting product profitability. This reduces financial strength by 3.22 percentage points. Although Allianz Leben's equity under commercial law (excluding profit participation rights and subordinated liabilities) is high at around EUR 3 billion in absolute terms, it is below the market average compared to the balance sheet total of around EUR 284 billion. This reduces the financial strength indicator by 2.26 percentage points. There is no absolute size bonus in the RealRate rating model. Instead, the relative balance sheet structure is compared, where size does not matter.

Overall, the essential economic relationships can be explained using the causal graph. In comparison to other purely indicator-based ratings, the multi-stage causal relationships become clear via the various mediator variables, which can be interpreted causally. The graph is easy to read and shows complex business models and relationships in just one image. In contrast to business or rating reports that are hundreds of pages long, this display is of great help in practice. Since the graph always shows the individual strengths and weaknesses of the company in relation to the overall market, it is directly suitable as a benchmark or peer group analysis.

In addition to actual financial strength, the causal graph is in fact the central rating result and explains this variable. The graph is provided and is the basis for the company's strategic analysis and future rating upgrades. In particular, it can serve as a basis for decision-making for the Executive Board, risk management and corporate planning. The effects shown can be interpreted as sensitivities. In the example, a green node, i.e. an individual strength, shows by how many percentage points the financial strength is increased by the fact that the corresponding size of the company deviates in a positive sense from the market mean.

The causal structure is the same for all companies, but the quantification of the individual effects is different. This is a direct consequence of the fact that the graph is fed with company-specific data, in this example with Allianz data. This gives you a comprehensive interpretation and explanation framework, but at the same time a very individual company analysis. The causal structure defined in advance ensures that the results can be explained later by design. The most important positive and negative effects can be easily read from the colour-coded graphic and thus meet the explainability requirement.

As Fig. 7 shows, the strength of the low guaranteed interest rate has steadily increased over time, up to the current positive effect of +4.63 percentage points on the economic equity ratio. At the same time, however, the weakness of below-average underwriting profitability has developed unfavorably.

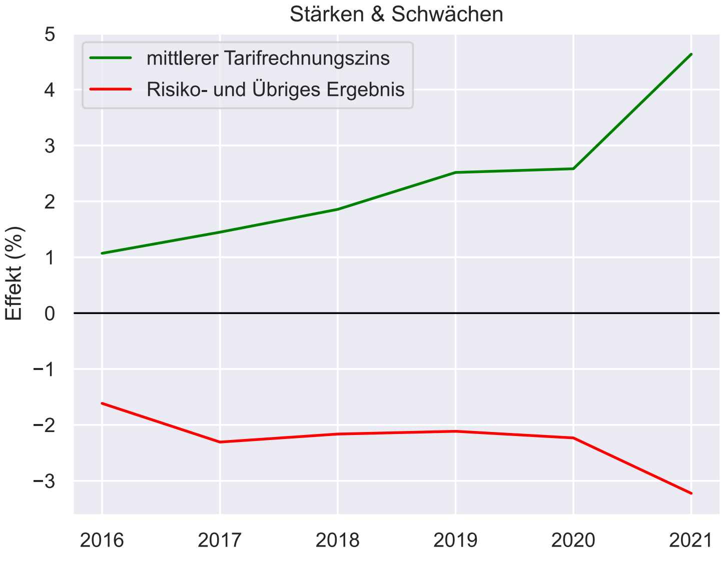

Fig. 7: The greatest strength and weakness of the life alliance over time, Effect on the economic equity ratio in percentage points

DOI https://doi.org/10.3790/zverswiss.2023.04.Bartel.etal | Generated on 2023-02-18 13:06:43

Finally, the question should be answered as to how well the explainable AI model can explain the reality of German life insurers in the 2021 financial year. For this purpose, the economic equity ratio is plotted on the x-axis (Fig. 8). The Solvency II own funds, without transitional measures, are plotted on the y-axis in relation to the balance sheet total. The orange line shows the theoretically desirable relationship, i.e. a 1:1 relationship between the two variables. The blue line indicates the actual regression fit. A connection between these two variables becomes clear; the rank correlation is 0.3. Thus, the model is able to explain a certain part of the published Solvency II values,

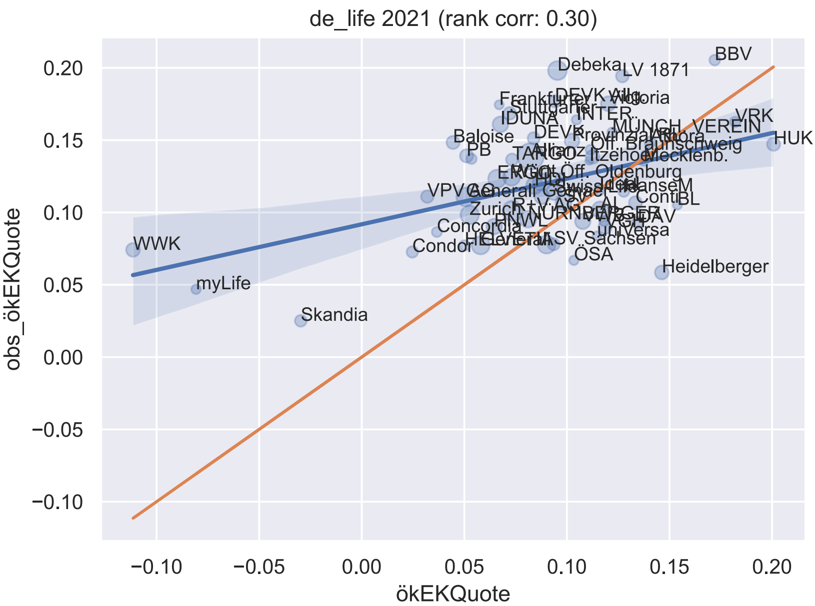

Fig. 8: Correlation between the economic capital ratio (x-axis) and the regulatory own funds according to Solvency II, without transitional measures, in relation to total assets (y-axis)

DOI https://doi.org/10.3790/zverswiss.2023.04.Bartel.etal | Generated on 2023-02-18 13:06:43

3.3 Estimation, Modeling Cycle and Software

The hybrid model is estimated in a similar way to the general, unconstrained case, but taking into account the causal restrictions. In fact, the given causal graph itself already looks like a neural network, but it is not fully connected. A target function to be minimized must be defined for the estimation of the model. In our example, this is the sum of squares between the modeled financial strength on the one hand and the published solvency ratios on the other hand (the objective function is formulated in Bartel 2019).

The optimization is a classic, non-linear optimization problem, but here with the additional constraint that certain nodes/neurons/variables are not causally linked to each other. This was technically implemented as a structural neural network (SNN). A similar procedure is already taking place behind the scenes at the technical level for general unrestricted neural networks: For the well-known backpropagation algorithm (see Rumelhart et al. 1986), the partial derivatives of the network must be determined, which is not possible for reasons of performance and numerical stability numerically - but not algebraically either, but the derivations are formed at the level of the program code. This is the so-called Automatic Differentiation (see Rall 1981). For example, the ADAM optimizer, which is particularly popular in machine learning, is used as an optimization algorithm (see Kingma/Ba 2014). Once the weights have been estimated, the model is validated. It is therefore necessary to check whether the assumed model fits the data. The computer code used is freely available as open source software.[11] The name Causing of this software stands for CAUSal INterpretation using Graphs. It is generally applicable to any topic. For example, there is an application that explains the wage level of young American workers based on their education and family background.[12]Causing is a multivariate graphical analysis tool that can be used to interpret the causal implications of a given system of equations. All you have to do as input is provide a data record and enter a system of equations. The endogenous variable on the left is assumed to be caused by the variables on the right side of the equation. Thus they provide the causal structure in the form of a directed acyclic graph. The output is an easy-to-understand colored diagram,

---

[11]Retrieved from https://github.com/realrate/Causing [02/01/2023].

[12]See https://github.com/realrate/Causing/blob/develop/docs/education.md [01.02. 2023].

DOI https://doi.org/10.3790/zverswiss.2023.04.Bartel.etal | Generated on 2023-02-18 13:06:43

which clearly shows the causal relationships between the variables. Entire chains of effects can be interpreted in this way.

After the model has been defined and estimated, modifications that result in different rules and goodness of fit can be checked. In the hybrid model approach, the following modeling cycle must be run through:

1. Define inputs

Determination of observed exogenous model variables that explain financial strength.

2. Define causal structure

Defining the causal structure by determining which variables have direct substantive relationships between them. In this way, the financial strength is ultimately determined.

3. Machine Learning/Estimation

Representation of the causal graph as a structured neural network and estimation of the free parameters using machine learning. The parameters are chosen in such a way that the published solvency ratios are reproduced as far as possible.

4. Evaluation

Measurement of the model performance and examination of the explainability in practice.

5. Model change

Modification of the model by changing the variables used or their causal relationships. Restart from step 1.

Overall, this hybrid modeling approach results in the following advantages from the user's point of view: explainability, transparency, scalability, small data instead of big data, speed. The methodology also facilitates model validation (measurement of model performance, explainability, identification of key modes of action, comparison of alternative explanations and quantification of performance losses). However, a transparent model alone is not sufficient. This only makes sense if the problem of conflicts of interest, which currently affects the relationship between companies and rating agencies, is also addressed. The next section addresses transparency and the design of business models.

DOI https://doi.org/10.3790/zverswiss.2023.04.Bartel.etal | Generated on 2023-02-18 13:06:43

4. Transparency and business model

Some of the market players consider the business figures of insurance companies to be non-transparent. In this respect, ratings of insurance companies have a more illuminating function, at least presumably. For the insurance industry, analysts are faced with the challenge that the various lines of business cause different interpretation difficulties. While the business results of property/casualty insurance companies and reinsurance companies can be subject to a large (random) range of fluctuation, the long-term nature and greater dependence on the capital market play a role for life insurance companies. The swings in property/casualty insurance companies are e.g. B. determined by natural disasters. An example would be the flooding in 2021 in the Ahr valley. However, these large and accumulated losses can only be normalized to a very limited extent on the basis of external balance sheet figures. Conversely, the interest rate sensitivities of the (German) life insurance portfolios, for example, are presented in the annual reports in only a slightly enlightening and comparable manner. In the case of insurance groups with several lines of business, the effects from different legal entities can also overlap. So-called conglomerate discounts are usual for this opacity. that the effects from different legal entities can overlap. So-called conglomerate discounts are usual for this opacity. that the effects from different legal entities can overlap. So-called conglomerate discounts are usual for this opacity.

The MCEV reporting on the initiative of international, capital market-oriented insurance groups was an attempt to respond to this (for the MCEV principles, see CFO Forum 2009). The dependency on ratings from international rating agencies could only be reduced to a limited extent. In particular, changes in the methodology of MCEV calculation were presented to the companies as an example of a lack of transparency. Attempts to reduce the market power of some rating agencies such as Standard & Poor's, Moody's and Fitch (particularly at European level) have also had little success.

The interaction of the business model of corporate ratings with independence is examined below. Scenario 1 "Commissioning" shows the current state of the rating sector: rating agencies have the rated companies as direct clients, which creates a strong psychological bond in the sense of an underlying obligation; sometimes the rated companies use this financial power expressly to put pressure on the rating agencies. The model is therefore sometimes viewed very critically by supervisors and also in politics and was the reason for calls for legal changes, especially during the financial crisis and other scandals. The regulation of such rating agencies has also reached a certain level in the EU, with some rating agencies, particularly in the US, appeal to freedom of expression. According to European

DOI https://doi.org/10.3790/zverswiss.2023.04.Bartel.etal | Generated on 2023-02-18 13:06:43

cal understanding of a commercial business model with only a few big players only conditionally the (sole) point of view.

The conflict of interest presented is resolved or at least alleviated by an alternative business model that decouples the relationship between the rating agency and the rated company. Scenario 2, "Public Rating", is per se less susceptible to unjustified influence.[13]The automation enabled by AI can also be used to scale (enlarge the addressable universe of valued companies) that reduce the opportunities for individual influence. The rating agencies issue ratings from many institutions, and rating agencies are subsequently paid by the rated institutions if the institutions want to republish an already existing rating for advertising purposes. The institutions evaluated do not affect the grade published. On the aspects of public data, transparency, conflicts of interest and the 2008 financial crisis, see Bartel 2023.

These public ratings are very common in the area of product ratings, such as at Stiftung Warentest. However, they have not yet established themselves in the area of company ratings, for various reasons. The requirement to use internal ratings instead of external ratings does not appear to be efficient from a macroeconomic point of view: Small and medium-sized companies in particular cannot afford a complex internal rating process. It therefore remains with external ratings and the associated conflicts of interest (on the conflicts of interest of the rating agencies from a US perspective, see Crumley 2012).

In addition to prohibiting conflicts, removing references to ratings, increasing liability, organizational barriers ("firewalls"), performance disclosure, disclosure of due diligence, increasing competition, "staleness" reforms, internal Management, administrative registration of alternative business models is required (Crumley 2012, p. iv). The corporate rating of German life insurance companies presented above could be considered such an alternative business model and be extended to other sectors and markets.

The business model of the RealRate agency differs from that of the traditional rating agencies in a number of key respects. The business model is completely data-based and no longer requires human analysts for the individual ratings. Once the model has been specified by an expert, it can be applied to all companies in the modeled industry. This automation leads to high scalability

---

[13]“Statistical” discrimination cannot be ruled out. See also *Bartlett*et al. 2022.

DOI https://doi.org/10.3790/zverswiss.2023.04.Bartel.etal | Generated on 2023-02-18 13:06:43

and favorable ratings. In this way, companies can also be covered that would otherwise not commission ratings. The business model is not based on the assignment, but all companies in an industry are always rated on the basis of publicly available data (public information rating). The rating model can be designed ex ante in such a way that it complies with legal requirements and social norms. In concrete terms, this means, for example, that distortions can be avoided, discrimination ruled out and fairness ensured. However, the main benefit of the non-commissioning approach by the individual company is that it avoids the inherent conflict of interest that arises from the fact that the company, as a client, commissions its rating for a fee.

In contrast to this, the top rated companies are offered the rating seal for advertising purposes (again comparable to the seal of the Stiftung Warentest for product tests). For example, only the top 25% of companies within an industry are awarded a seal. This eliminates the distortion that could result from the fact that only companies that can realistically expect a desired rating to commission ratings.

This business model is then based on annually recurring payments ("subscription") for the right to advertise with the rating result. The advantage of companies advertising with it is that they can advertise with a non-commissioned rating seal from an independent institution. The modified business model, together with modern AI technology, can help to avoid conflicts of interest. Since this is an external rating, it is particularly easy to compare within an industry due to the transparency.

5. Summary

In this article, a method for evaluating (German) insurance companies based on AI methods was presented. On the basis of a provisional catalog of criteria for industrial models of explainable artificial intelligence, it was then positioned in terms of its transparency properties. An explainable model for evaluating companies helps market players to better understand the decisions made. The model was applied to German insurance companies and guarantees that it can be explained using a directed graph that illustrates the causes and effects of the relevant variables. An advantage of the method is that strengths and weaknesses of the company directly above the

DOI https://doi.org/10.3790/zverswiss.2023.04.Bartel.etal | Generated on 2023-02-18 13:06:43

graphs are explainable and for which hardly any technical expertise is required. Subsequently, the application of such models was also examined from the point of view of transparency; In particular, it was also examined how the business model is related to the independence of the rater and how conflicts of interest from existing rating business models are reduced. After all, transparency is also a property of the system in which a method is embedded. The AI-based rating model is already being applied to German life insurance and health insurance companies, as well as risk insurers, and to assess future BU premium stability. In addition, approximately 2,000 listed US companies from 20 different industries are rated.

Extensions are conceivable in various directions: So far, only the quantitative data from the balance sheet and profit and loss account of the annual report are used. In future, disclosures in the notes, verbal descriptions of the accounting methods and the risk and management report should also be taken into account. ESG data (Environmental, Social, and Governance) will also play an important role in the future and expand the rating from purely financial data to a holistic analysis and evaluation. AI-based text summaries and ratings can be used here. Conversely, an extension with a text generation module is also conceivable, which generates a report on the rating achieved by each insurer from the network. Such a language generation component "Natural Language Generation") could further improve the explainability of the method. There has recently been enormous progress here, for example with the public discussion about ChatGPT from OpenAI[14]becomes clear. On the business model side, scenario analyzes would be desirable that show conflicts of interest and examine who generates ratings for whom and who pays for them. Quantifying the bias based on not publishing bad ratings would also provide valuable insights.

Acknowledgments: We would like to thank Niklas Häusle for valuable feedback at the DVfVW annual conference 2021.

literature

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

barbel, H. (2014a): Simple Solvency – A solvency model for German life insurers, DOI: 10.13140/2.1.2939.9041, https://www.researchgate.net/publication/267337608_Simple_Solvency_-_Ein_Solvenzmodell_fur_deutsche_Lebensverversicherunger [26.01.2023].

[14]https://openai.com/blog/chatgpt/ [02/01/2023].

barbel, H. (2014b): Simple Solvency – A solvency model for German life insurers, Lecture, qx Club, Regional Group of the German Actuary Association (DAV) for Berlin, DOI: 10.13140/2.1.1760.2560, https://www.researchgate.net/publication/267337670_Simple_Solvency_-_Ein_Solvenzmodell_fur_deutsche_Lebensververversicherunger [01/26/2023].

barbel, H. (2019): Causal analysis of equation systems with structural neuronal nets. Technical report. DOI: 10.13140/RG.2.2.16841.26729, https://www.researchgate.net/publication/335099531_Kausale_Analyse_von_Equationsystems_with_structural_neuronal_networks [01/26/2023].

barbel, H. (2020a): Causal Analysis – With an Application to Insurance Ratings, https://www.researchgate.net/publication/339091133_Causal_Analysis_-_With_an_Application_to_Insurance_Ratings [2023-01-26].

barbel, H. (2020b): Explainable Artificial Intelligence (XAI) in Ratings, https://www.researchgate.net/publication/344992217_Explainable_Artificial_Intelligence_XAI_in_Ratings [2023-01-26].

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

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

barbel, H. (2023): Financial strength ratings of German insurers using artificial intelligence enough In: Journal of Insurance, 74(2), pp. 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), pp. 30–56, DOI: 10.1016/j.jfineco.2021.05.047.

Breiman, L/ *Friedman*, JH/ *Olsen*, RA/ *Stone*, CJ (1984): Classification and regression trees, Monterey, CA, USA: Wadsworth & Brooks/Cole, DOI: 10.1201/9781315139470.

Burkart, N/ *Huber*, MF (2021): A Survey on the Explainability of Supervised Machines Learning, Journal of Artificial Intelligence Research 70, pp. 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, WW (1995): Fast Effective Rule Induction, Proc. 12th international conf machine learn ing (ICML).

Crumley, DG (2012): Credit Rating Agencies and Conflicts of Interest, Inaugural dissertation tation (JD thesis), University of Texas at Austin, Austin, TX, USA.

Dierkes, p./ *boorman*, J. (2019): Digitized peer group determination and beta fit. In: Ballwieser, W./Hachmeister, D. (ed.) (2019): Digitization and company valuation, pp. 173-192. Stuttgart: Schäffer-Poeschel.

Dziugaite, GK/ *Ben-David*, p./ *Roy*, DM (2020): Enforcing interpretability and its statistical tical impacts: trade-offs between accuracy and interpretability, https://arxiv.org/pdf/2010.13764.pdf [01/26/2023].

European Commission(2020): White Paper on Artificial Intelligence – A European sches Concept for Excellence and Trust, Brussels, https://ec.europa.eu/info/sites/ info/ files/commission-white-paper-artificial-intelligence-feb2020_de.pdf [26.01.2023].

European Commission(2021): Regulating credit rating agencies, Brussels, https://ec.eu ropa.eu/info/business-economy-euro/banking-and-finance/financial-supervision-andrisk-management/managing-risks-banks-and-financial-institutions/regulating-creditrating-agencies_en [26.01.2023].

High Level Expert Group on Artificial Intelligence(HEG-KI) (2018): Ethics guidelines for a trustworthy AI, Brussels, https://op.europa.eu/s/oVfc [01/26/2023].

GitHub(2021a): A Real World Example: Education and Wages for Young Workers, https://github.com/realrate/Causing/blob/develop/docs/education.md [2023-01-26].

GitHub (2021b): Causing: CAUSal INTERPRETATION using Graphs, https://github.com/real rate/causing.

reason, H./*Power*, M. (ed.) (2019): Solvency II – An Introduction. Basics of the new insurance supervision. 3rd edition Karlsruhe: VVW.

heskes, T./*Sijben*, E /*book*, IG/*class*, T (2020): Causal Shapley Values: Exploiting Causal Knowledge to Explain Individual Predictions of Complex Models, NeurIPS 2020, https://arxiv.org/abs/2011.01625 [01/26/2023].

Holland, CP/*Kavuri*, A. (2021): Artificial intelligence and digital transformation of domestic Surance Markets. In: The Capco Institute – Journal of Financial Transformation H. 54 (11/2021), pp. 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 ing: with Applications in R, New York, NY, USA: Springer.

Kingma, DP/*ba*, J. (2014): Adam: A Method for Stochastic Optimization, ICLR 2015 conference paper, https://arxiv.org/abs/1412.6980 [2023-01-26].

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

Kurmann, S. (2023): AI in the insurance industry: when science fiction meets reality meets, https://www.handelszeitung.ch/insurance/kunstliche-intelligenz-fur-versicherungs ungen/ki-in-der-versicherungsbranche-wenn-science-fiction-auf-reality-treffen-564992 [02/01/2023].

sufferer, JL (in preparation): A Survey of Ethical Problems of Artificial Intelligence.

lossos, C /*swollen*, p./*Morelli*, F. (2021): Openness through XAI in ML-enabled Decisions: A building block for optimizing decisions in the company? HMD practice of business informatics 58, pp. 303-320.

Lundberg, S. (2020): Lecture “The Science behind SHAP”, https://www.youtube.com/ watch?v=-taOhqkiuIo [2023-01-26].

Lundberg,SM/*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, pp. 4768-4777.

Oletzky, T./*Reinhardt*, A. (2022): Challenges of the regulation of and the view of the use of artificial intelligence in the insurance industry. In: Journal for the entire insurance science 111 (4), pp. 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), pp. 1-50, DOI: 10.3390/risks10120230.

Quinlan, JR (1986): Induction of Decision Trees, Machine Learning 1(1): pp. 81–106.

ral, LB (1981): Automatic Differentiation: Techniques and Applications. Lecture notes in computer science. 120. Springer.

Ribeiro, MT/*singh*, p./*guestrin*, C. (2016): "Why Should I Trust You?": Explaining the Predictions of Any Classifier, Proceedings of the 22ndACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Francisco, California, USA, pp. 1135-1144, DOI: 10.1145/2939672.2939778.

rumelhart, DE/*Hinton*, GE/*Williams*, RJ (1986): Learning representations by back-propagating errors, Nature 323 (6088): pp. 533-536, DOI 10.1038/323533a0.

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

Sellhorn, T. (2020): Machine Learning and Empirical Accounting Research: Some insights and open questions, Schmalenbach's Z. betriebswirtsch. Forsch 72, pp. 49-69, DOI: 10.1007/s41471-020-00086-1.

Shapley, LS (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.

Simpsons, EH (1951). The Interpretation of Interaction in Contingency Tables, Journal of the Royal Statistical Society, Series B. 13: pp. 238-241.

Stuwe, A./*White*, M./*Philippians*, J. (2012): Rating agencies: Are they necessary, superfluous, necessary evil or harmful?, Bonn: Friedrich-Ebert-Foundation, 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, p./*Sun*, R. (ed.) (2000): Hybrid Neural Systems. Springer Verlag, Heidelberg.