Finance Research Letters 57 (2023) 104242

Contents lists available at ScienceDirect

Finance Research Letters

journal homepage: www.elsevier.com/locate/frl

Explainable AI for paid-up risk management in life insurance products

Lluís Bermúdez [a,*], David Anaya [b], Jaume Belles-Sampera [a,c]

[a] Riskcenter-IREA, Universitat de Barcelona, Spain
[b] Facultat d'Economia i Empresa, Universitat de Barcelona, Spain
[c] Grupo Catalana Occidente, S.A., Spain

ARTICLE INFO

JEL classification:
C10
C35
G22

MSC:
62H30
62P99
68T01

Keywords:
Machine learning
Shapley values
Kohonen networks
Risk analysis

ABSTRACT

Explainable artificial intelligence (xAI) provides a better understanding of the decision-making processes and results generated by *black-box* machine learning (ML) models. Here, we outline several xAI techniques in order to equip risk managers with more explainable ML methods. We illustrate this by describing an application for the more effective management of paid-up risk in insurance savings products. We draw on a database of real universal life policies to fit an initial logistic regression model and several tree-based models. We then use different xAI techniques, including a novel approach that leverages a Kohonen network of Shapley values, to offer valuable perspectives on tree-based models to the end-user. Based on these findings, we show how non-trivial ideas can emerge to improve paid-up risk management.

1. Introduction and motivation

Explainable artificial intelligence (xAI) techniques are receiving increasing attention in those fields in which artificial intelligence (AI) and machine learning (ML) models have been widely employed (Arrieta et al., 2020). Indeed, the adoption of xAI techniques appears to be critical to enhance understanding, trust and management of ML methods that are not directly interpretable (Adadi and Berrada, 2018).

One such field that has begun to use AI and ML methods to specifically analyze the probability of the occurrence of certain events is risk management. Here, the need to understand, trust and effectively manage ML results has become the challenge to which risk managers have had to commit themselves. In this paper, we discuss an example of how steps have been taken to face this particular challenge by adopting an application aimed at managing more effectively the paid-up risk in insurance savings products (an application, moreover, that is equally germane to financial savings products).

Life insurance products, including universal life insurance, are subject to various risks associated with policyholder behavior. One such risk, the paid-up risk, arises when policyholders exercise their option, at some point before maturity, to stop making the regular premium payments that they initially committed themselves to for the life of the policy. In a systematic review of xAI in insurance, Owens et al. (2022) show that only a small proportion of the articles reviewed report results related to the application of xAI techniques to risk management. Moreover, the articles that do primarily concern themselves with lapse or surrender risks (see Azzone et al., 2022), leaving paid-up risk largely overlooked.

* Correspondence to: Department of Economics, Financial and Actuarial Mathematics, Av.Diagonal 690, 08034 Barcelona, Spain.
E-mail address: lbermudez@ub.edu (L. Bermúdez).

https://doi.org/10.1016/j.frl.2023.104242
Received 19 May 2023; Received in revised form 29 June 2023; Accepted 21 July 2023
Available online 24 July 2023

Table 1
Definition of selected variables.

Variable	Definition	Mean
bin_resp	Premium payment status (1: Paid-up, 0: Active)	0.2111
cap	Additional sum insured in case of death (€)	3,502.8
res	Current value of the fund (€)	8,688.9
prem	Total annual premium (€)	1,303.5
age	Current age of the insured	47.88
loy	Number of years the policy has been in force	5.12
rem	Years remaining until final premium as per contract	19.30
gen	Gender of the insured (1: Female, 0: Male)	0.5563
unl	Unit-linked product (1: Yes, 0: No)	0.4068
tax	Product with tax advantages (1: Yes, 0: No)	0.3661
fee	Product with active surrender fee (1: Yes, 0: No)	0.1934
rate	Product with fixed guaranteed interest rate (1: Yes, 0: No)	0.0306

This paper seeks to address this deficiency by presenting risk managers with xAI techniques that can facilitate their decision-making process by increasing (1) their ability to distill knowledge and extract rules, and (2) their level of confidence as regards the predictive accuracy of *black-box* ML models. To do so, and as the main contribution of this paper, we propose a new combined use of Shapley values and Kohonen maps in an approach that takes its inspiration from Bussmann et al. (2020). It aims to group similar data points together based on their Shapley values and, hence, on their common characteristics. Risks within the same cluster are likely to share properties and may require similar risk response strategies. This allows risk managers to allocate resources effectively by focusing on the most critical clusters.

2. Methodology

In this paper, we focus our attention specifically on model-agnostic xAI methods, that is, methods that can be applied to any type of ML algorithm. Model-agnostic methods are usually post-hoc and may be either global (i.e. they explain the whole model) or local (i.e. they explain a single prediction) interpretable (Adadi and Berrada, 2018).

Global explanations provide insights into how the ML model works, the most common being: (1) feature importance (FI), which helps identify which features have the most impact on the model's output; and (2) visualizing techniques, such as partial dependence plots (PDP) or accumulated local effects (ALE), which help show how the model's predictions change as the value of a particular feature is varied while holding all other features constant. Local explanations help users understand the reasoning underpinning a model's decision on a particular input data point. The most frequently employed local xAI technique is SHapley Additive exPlanations (SHAP), which use game theory to assign a value to each input feature based on its contribution to the prediction (see Lundberg and Lee, 2017).

Here, briefly, we adopt the following approach. First, as is common, we employ global techniques – including FI and ALE – to determine the most relevant features used by the ML models to make their predictions. Second, as a novelty, we propose the transformation of the SHAP technique into a global xAI tool by (1) obtaining the Shapley values of the features for all data points belonging to each cell of the confusion matrix, and (2) using a Kohonen neural network (KNN) trained with these SHAP values to identify different policyholders profiles. The KNN, or self-organizing map (SOM), is a type of unsupervised learning algorithm commonly used for data visualization, clustering, and pattern recognition (see Huysmans et al., 2006). This second step provides us with a complementary global perspective on the different profiles used by ML models when making their predictions.

In what follows, we use *iml* (interpretable machine learning) and *kohonen* R packages to understand and visualize how ML models make predictions.

3. Data and modeling analysis

The data refer to policies sold between 2018 and 2019 by a life and non-life insurance company operating in the Spanish market. The policies are universal life products and all of them require policyholders to make constant, periodic (monthly) premium payments. We select only those active premium payment policies from 12/31/2018 that were in force as of 12/31/2019. The total number of policies is 7,886, while our subset of interest (i.e. policies with a paid-up premium as of 12/31/2019) represents 21.11% (or 1,665 policies). Table 1 contains a detailed description of policy features, including premium payment status and policyholder-product characteristics.

Note that the table includes a *fee* variable to indicate that some life insurance products include an initial deposit requirement which serves as a surrender charge in case of early termination and aims to mitigate the risk of lapsing. In practice, this deposit is a fee that policyholders agree to pay should they surrender or cancel the policy within a specified period of time, typically during the first few years of the contract.

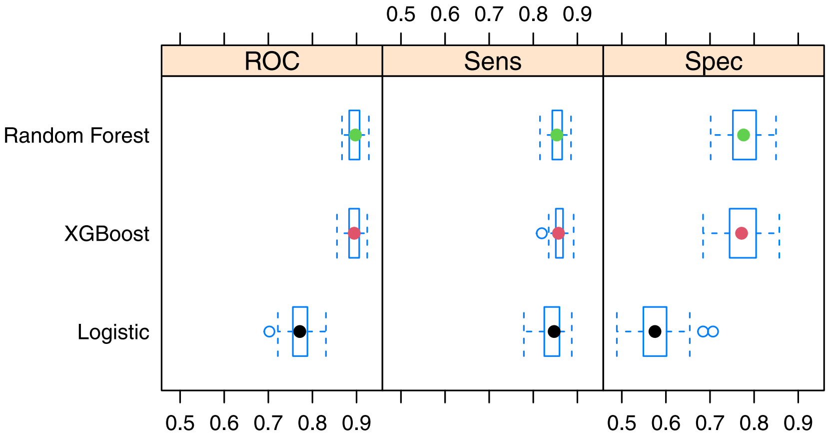


Fig. 1. Performance metrics for logistic regression, random forest and XGBoost models.

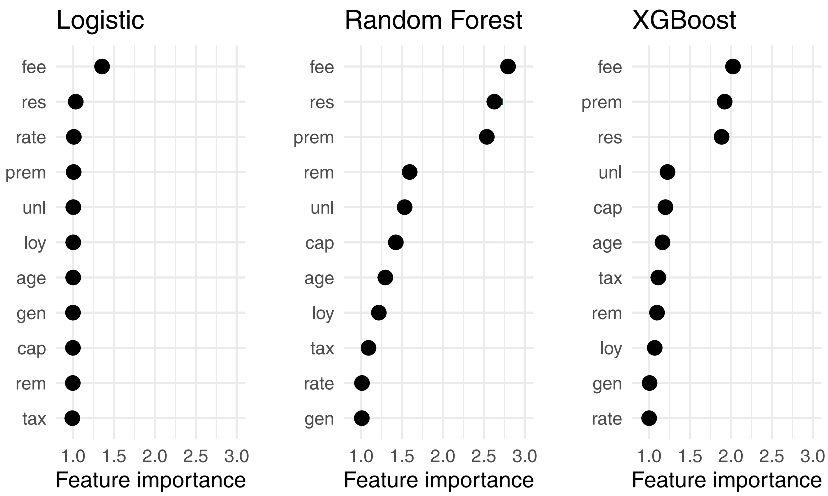


Fig. 2. Feature importance of the three models fitted, based on cross-entropy criteria.

The preparatory phase involves the analysis of alternatives aimed at balancing the imbalanced dataset (undersampling being the most appropriate technique), normalizing and scaling the continuous variables, and subsetting the data for cross-validation (80% were employed for training; the remaining 20% for testing).

After fitting logistic regression, random forest (RF), and XGBoost (XGB) models to these data, their respective predictive accuracies can be compared with the usual performance metrics (ROC, sensitivity and specificity) (see Fig. 1). In this case, *black-box* models (RF and XGB) obtained a higher level of predictive accuracy than that obtained by the logistic regression.

4. Empirical findings

In what follows, we show how xAI techniques can be specifically used to interpret, understand, and trust the RF model fitted to our paid-up data.

4.1. Interpreting the model

First, to boost managerial skills of knowledge distillation and rule extraction, we focus on the usability of global xAI techniques. Feature importance techniques seek to provide insights into the factors that have most influence when making predictions. FI scores are shown in Fig. 2. In general, the most influential features for the prediction of a paid-up event coincide across all three models: namely, *fee* and *unl* among the categorical features (i.e. related to a product's characteristics) and *res* and *prem* among the numerical features (i.e. measures of policyholder commitment to the product).

Have obtained scores for each feature, a manager might be interested in understanding the relationship between each input feature and the prediction output of a model. Fig. 3 shows the ALE plots for all features of the RF model. The ALE technique

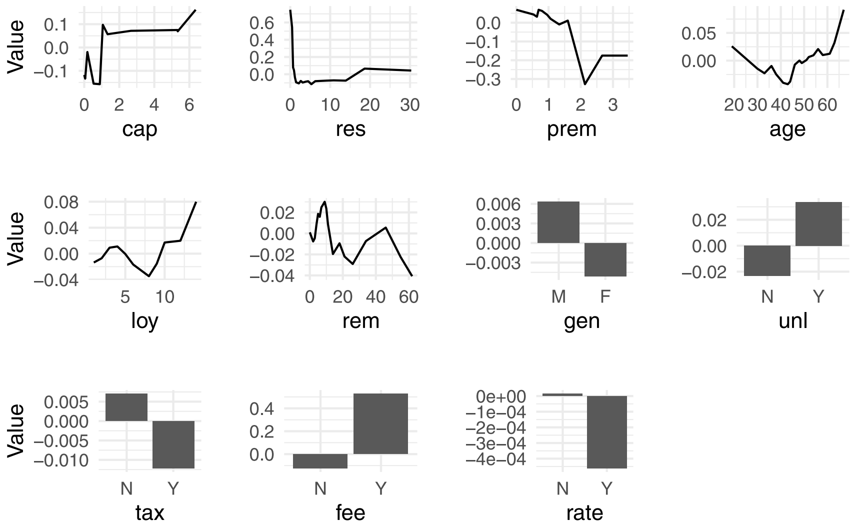


Fig. 3. Accumulated local effects for all features of the RF model. The first three features are expressed in thousands of €.

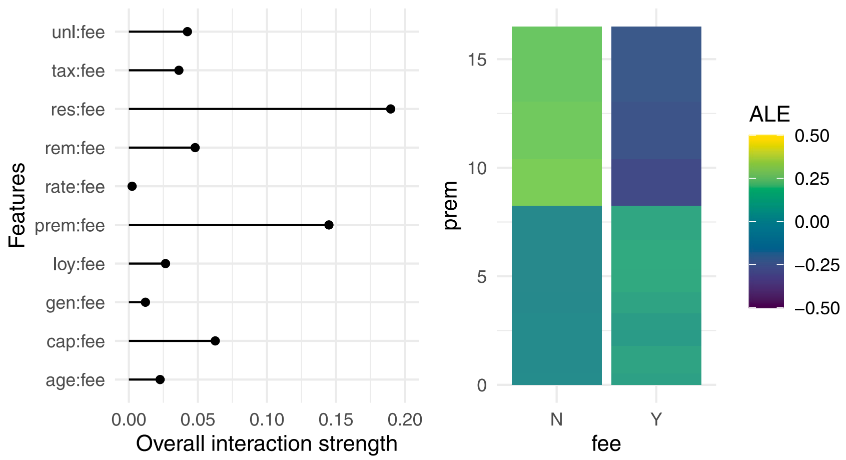


Fig. 4. H-statistics of two-way interactions between fee and the other features (left) and second-order ALE plot for *fee* and *prem* features of the RF model (right).

provides a variety of insights into the model's behavior, for example, identifying both non-linear effects of individual features and interactions between features that might otherwise be missed by simpler methods. Note that the most influential features identified in Fig. 2 also present the highest ALE values. Specifically, paid-up risk increases for premium payments up to 1,500€, while this risk falls for premiums higher than 1,500€, but especially for premiums around 2,000€. Interestingly, paid-up risk increases gradually as policyholders approach retirement age (after the age of 60 or with more than 15 years of policy duration). In general, in the case of the more influential categorical features, having a product with an active surrender charge (*fee*) or a unit-linked product (*unl*) increases paid-up risk, while for products offering tax benefits (*tax*) this risk is reduced.

Second-order ALE plots take this analysis one step further by estimating the effect of a pair of features on the predicted outcome and, hence, they enable managers to interpret interactions between pairs of features. Before using these plots, the features that interact with each other in a particular model have first to be identified and, then, the strength of those interactions must be quantified. A high H-statistic, as provided by Friedman and Popescu (2008), indicates a strong interaction effect. By way of illustration, Fig. 4 shows a significant interaction between two features of the RF model. Thus, the model's predictions change depending on the effect of the interaction between a product having (or not having) an active surrender fee and the value of annual premium payments (i.e. greater of less than 8,000€[1]). Premium payments above 8,000€increase the paid-up risk for non surrender charge products while for the same payments this risk is reduced markedly when the product has an active surrender fee. For premiums below 8,000€, the effect – albeit not so marked – is reversed.

[1] In Spain, at that time, individuals could deduct up to 8,000€from their taxable income for contributions made to certain pension products.

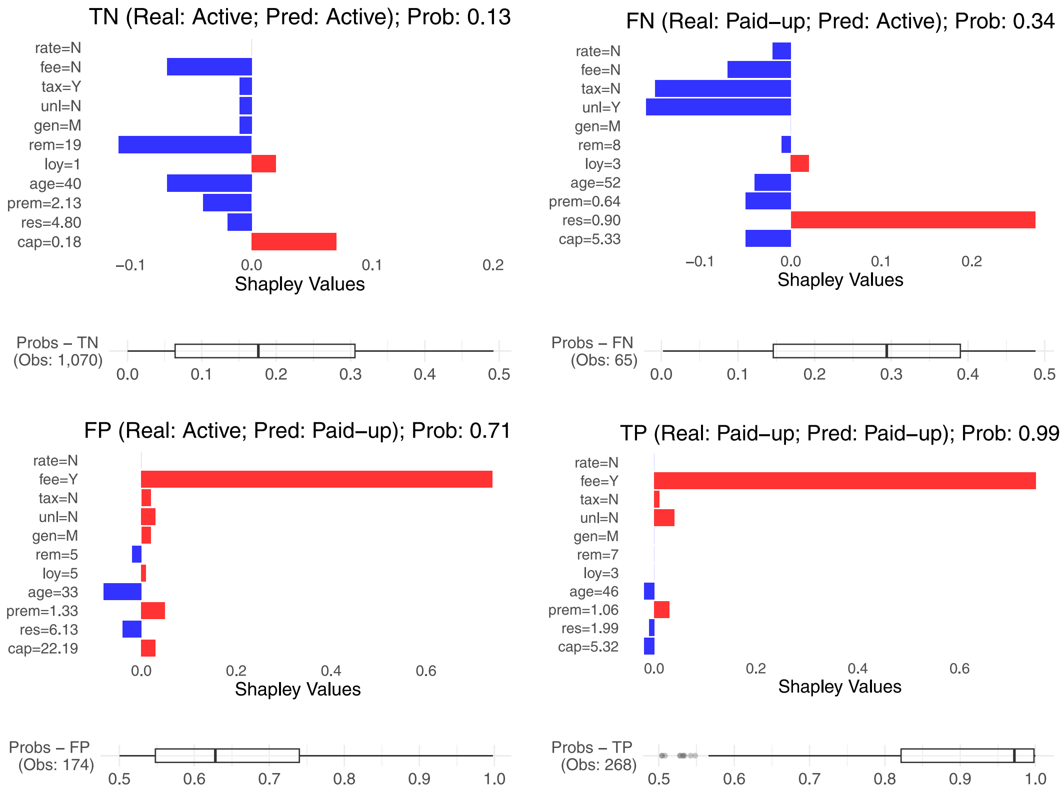


Fig. 5. Shapley values for random observations (from all the observations -1,577- in the testing dataset) of each cell of the confusion matrix, and the corresponding boxplots of the predicted paid-up risks.

4.2. Understanding and trusting the model

Second, to boost confidence in ML models, risk managers need to understand how the predictions are made. For a single observation, Shapley values are a frequently employed method to understand which features have the greatest weight in the model's prediction. A Shapley value represents the contribution made by one feature to the difference between the model's prediction for a single instance and the average prediction (c. 20% in our case). Fig. 5 shows, as expected, similar Shapley values (mostly positive) for the policyholders predicted by the model to be *Paid-up* – true positive (TP) and false positive (FP) observations – and the same degree of similarity (albeit mostly negatives) for policyholders predicted to be *Active* – true negative (TN) and false negative (FN) observations. In general, the most important features present higher Shapley values for *True* cases than they do for *False* cases. Our results also highlight the different scales of Shapley values according to whether the instance is predicted to be *Paid-up* or not.

Note that the boxplots of the predicted paid-up risks in Fig. 5 clearly show that the model mostly fails (FP and FN) when the prediction values are close to the threshold of 0.5 (between 0.3 and 0.7) and, in contrast, the model works (TP and TN) when these prediction values are more extreme (below 0.3 and above 0.7).

To avoid the limitations of the borderline cases, Fig. 6 shows the distribution of both features and Shapley values – focusing on the most important categorical and numerical features – for observations from the testing dataset with extreme prediction values (i.e. below 0.3 and above 0.7). The group with extreme predictions represents 70% of total observations (1,111), of which 92% are classified correctly. By so doing, managers can understand better how a model correctly classify policies as *Paid-up* or *Active*. First, marked differences are not evident in the feature distributions of the *True* and *False* cases, meaning that the errors (which represent 8% of the classifications) are due to other unobserved characteristics which, as such, are not detectable by the model. Second, the effect on the interaction between *fee* and premiums (and reserves) can be easily checked. For example, TN cases correspond to products both with and without a surrender fee, combining high premiums (and reserves) with a surrender fee and combining low premiums (and reserves) without a surrender fee, as illustrated by the different colors for the *prem* and *res* features in Fig. 6.

However, Fig. 6 does not provide a visualization of all possible combinations of features. It could be interesting, for example, to know how many different profiles, and their common feature combinations, can be identified from each cell of the confusion matrix. Here, as a novel approach of using Shapley values, we propose using a KNN (or SOMs) to identify clusters based on the Shapley values. In Figs. 7 and 8, the SOMs for observations with extreme and borderline predictions, respectively, are shown.

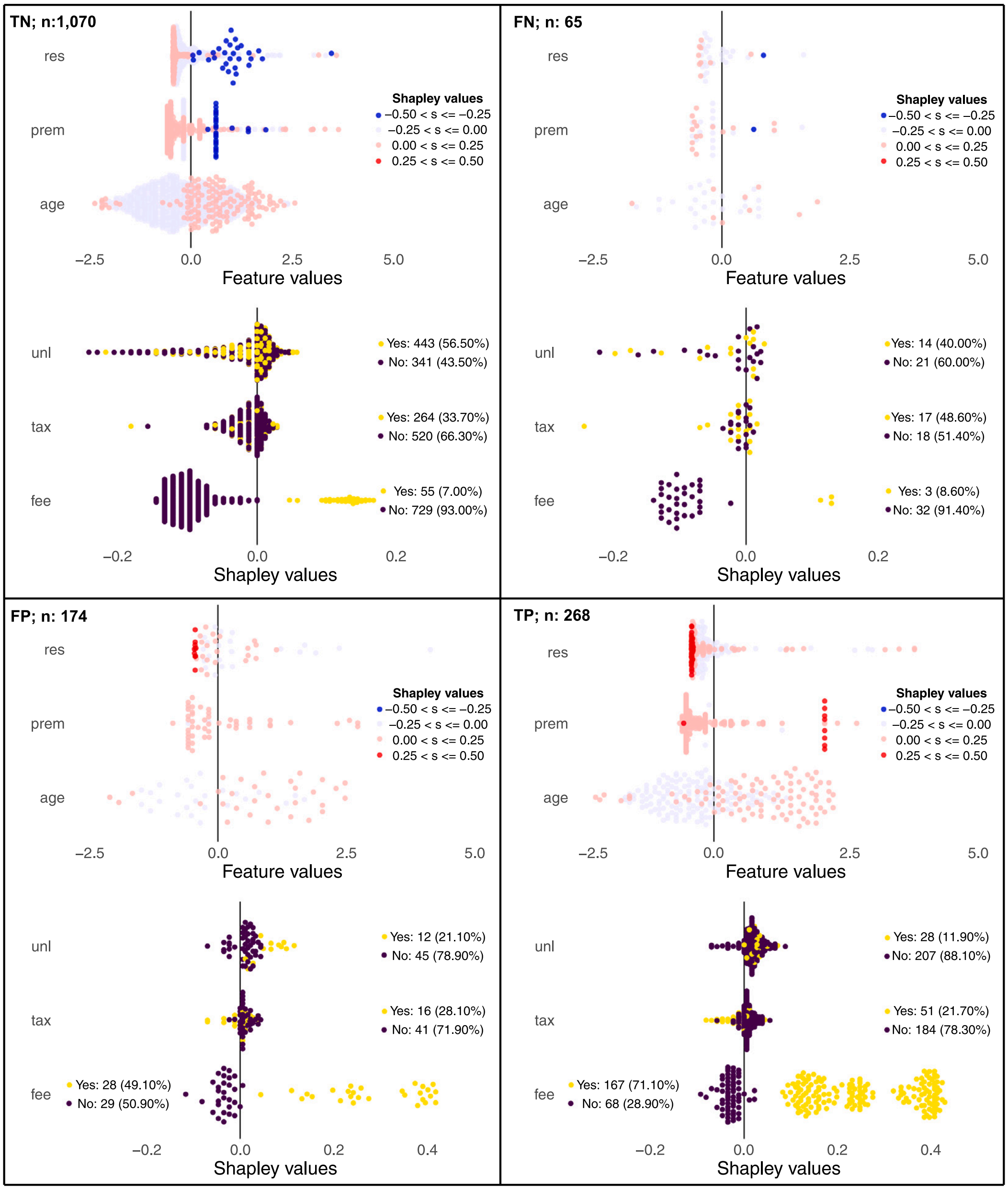
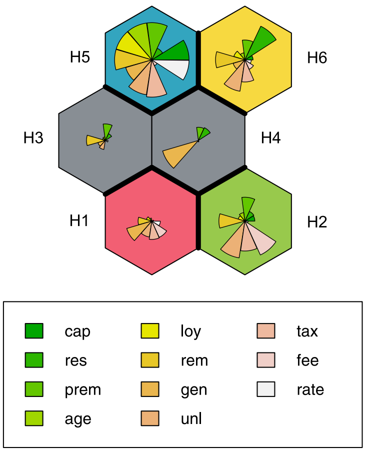


Fig. 6. Feature and Shapley values of each cell of the confusion matrix. Data with extreme prediction values (below 0.3 and above 0.7).

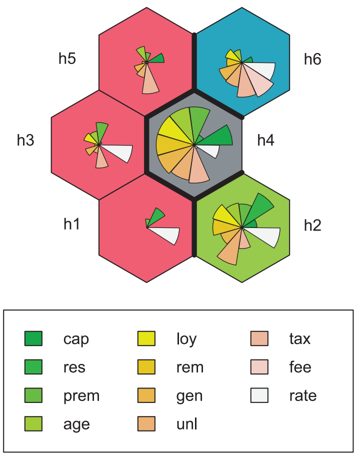
Fig. 7 shows that the Kohonen network consists of six clusters (or hexagons), three of them with paid-up probabilities below 0.3 (predicted as *Active*) and the rest with probabilities above 0.7 (predicted as *Paid-up*). The classification error indicates that the RF model produces more FP cases than FN. The hexagons with cases predicted as *Active* (H1, H3, and H4) can be divided according to whether or not the policy has a surrender fee and accounting for the interaction effect with premium and reserve values. H1 consists of non surrender fee products with high premiums and reserves, whereas H3 and H4 correspond to non surrender fee products with low premiums and reserves. Besides these features, H1 policies are neither unit-linked nor offer tax benefits, while H3 and H4 are mostly either unit-linked or offer tax benefits (with a larger proportion of unit-linked products). The main differences between H3 and H4 are due in the main to considerations of a time-related nature: thus, H4 policyholders are younger, have held the policy with the company for less time and have more years remaining on the contract than H3 policyholders. Similarly, the hexagons with cases predicted as *Paid-up* (H2, H5, and H6) can also be divided according to whether or not the policy has a surrender fee. Basically, H2 and H5 represent cases where the interaction effect is different: thus H2 corresponds to policies with a surrender fee and low premiums and reserves, whereas H5 corresponds, in the main, to policies without a surrender fee and high premiums and reserves. Uniquely, both policies with and without a surrender fee are equally represented in H6.



Hex	Class	Error	Prob(sd)	Prem	Res	Age	Fee	Unl	Tax
H1	N(100%)	FN(5%)	0.16(0.06)	2,664	21,541	49	Y(100%)	N(98%)	N(100%)
H2	P(100%)	FP(17%)	0.94(0.07)	720	4,244	47	Y(100%)	N(92%)	N(94%)
H3	N(100%)	FN(5%)	0.13(0.08)	1,065	4,131	49	N(100%)	Y(56%)	N(59%)
H4	N(100%)	FN(3%)	0.08(0.09)	639	1,050	41	N(100%)	Y(71%)	N(73%)
H5	P(100%)	FP(34%)	0.83(0.09)	1,555	9,908	59	N(80%)	N(86%)	N(61%)
H6	P(100%)	FP(13%)	0.93(0.08)	533	328	44	Y(51%)	N(78%)	N(63%)

Class (% of cases predicted as N: Negative or Active, or P: Positive or Paid-up);
Error (% of error cases, FN: False negative or FP: False positive);
Prob (Mean and standard error of Paid-up probabilities);
Prem, Res, Age (Median of the features);
Fee, Unl, Tax (% of majority class)

Fig. 7. SOM map of observations for extreme predictions (below 0.3 and above 0.7).



Hex	Class	Error	Prob(sd)	Prem	Res	Age	Fee	Unl	Tax
h1	N(100%)	FN(10%)	0.39(0.05)	639	718	44	N(100%)	N(73%)	Y(71%)
h2	P(100%)	FP(74%)	0.61(0.06)	888	663	44	N(100%)	N(71%)	Y(74%)
h3	N(100%)	FN(12%)	0.38(0.06)	746	6,576	52	N(98%)	N(59%)	N(53%)
h4	P(100%)	FP(78%)	0.58(0.05)	1,065	5,574	49	N(97%)	N(73%)	Y(54%)
h5	N(98%)	FP(2.35%)	0.39(0.06)	1,065	6,878	48	N(94%)	N(87%)	Y(59%)
h6	N(60%)	FN(7%) FP(81%)	0.46(0.10)	2,664	12,471	45	Y(100%)	N(82%)	N(98%)

Class (% of cases predicted as N: Negative or Active, or P: Positive or Paid-up);
Error (% of error cases, FN: False negative or FP: False positive);
Prob (Mean and standard error of Paid-up probabilities);
Prem, Res, Age (Median of the features);
Fee, Unl, Tax (% of majority class)

Fig. 8. SOM map of observations with borderline predictions (between 0.3 and 0.7).

Fig. 8 focuses on the observations that the model predicts lie closest to the borderline (466) and, as such, as the classification errors indicate, concentrate cases of model FN (33) and FP (286), providing a level of accuracy of 68%. The hexagons with cases predicted as *Active* (h1, h3, and h5) correspond to policies with similar characteristics (i.e. without a surrender fee and low premiums) to those in H3 and H4. However, the model mainly predicts larger paid-up probabilities for h1 than for H3 because the proportion of products offering tax benefits is higher than that of unit-linked products, while the same is true of h3 and H4 but in this case the higher predictions are due to the fact that h3 policies are associated with higher reserves than are H4 policies. Despite this, the model correctly predicts around 90% of *Active* cases. The hexagons with cases predicted as Paid-up (h2 and h4) present a larger proportion of errors (FP). Although both contain policies without a surrender fee, in common with H5, their premiums and reserves are lower than those in H5, reducing the paid-up probability greatly and failing in 75% of cases. Finally, hexagon h6 consists of policyholders (all without a surrender fee) with paid-up probabilities that are very close to 0.5 and predicted in almost equal proportions to be *Active* or *Paid-up*.

5. Conclusions

The xAI techniques proposed in this paper should help risk managers understand how ML models arrive at their predictions, thus, enabling them to identify patterns and gain insights that may not be immediately apparent in the raw data. These xAI techniques should serve to boost managerial confidence in the predictive accuracy of these models. By making the models more transparent and interpretable, risk managers can better understand the limitations and potential biases of the models, leading to more informed and more accurate decisions, and hence, better outcomes for their organizations and stakeholders.

Here, in illustrating the implications for risk management of using xAI techniques, and specifically the Kohonen network of Shapley values here proposed to interpret an ML model, we report a number of useful insights that should improve paid-up risk management. The most relevant conclusion to be drawn is that paid-up probability depends mainly on the interaction between the inclusion or otherwise of a surrender fee and the values of policy premiums and reserves. When a surrender fee is included in the product (to mitigate its lapse risk), the paid-up risk either increases (for policyholders with low premiums and reserves) or decreases (for policyholders with high premiums and reserve). This result may lead risk managers to rethink the rules governing surrender fees.

Indeed, this represents a breakthrough in the field of insurance risk management literature. The effectiveness of a managerial strategy aimed at mitigating one risk (such as setting surrender fees to reduce lapse risk) is called into question due to the discovery of an unintended consequence on another risk (the paid-up risk). Notably, this finding is primarily attributed to the skillful implementation of xAI techniques proposed in this manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Funding

This work has been partially supported by the Spanish Ministry of Science, Innovation and Universities (grant PID2019-105986GB-C21). Likewise, we want to acknowledge the support received by AGAUR, Government of Catalonia (grants 2020DI9 and 2021SGR00299).

References

Adadi, A., Berrada, M., 2018. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 6, 52138–52160.
Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F., 2020. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion 58, 82–115.
Azzone, M., Barucci, E., Moncayo, G., Marazzina, D., 2022. A machine learning model for lapse prediction in life insurance contracts. Expert Syst. Appl. 191, 116261.
Bussmann, N., Giudici, P., Marinelli, D., Papenbrock, J., 2020. Explainable AI in fintech risk management. Front. Artif. Intell. 3, 26.
Friedman, J., Popescu, B., 2008. Predictive learning via rule ensembles. Ann. Appl. Stat. 916–954.
Huysmans, J., Baesens, B., Vanthienen, J., Van Gestel, T., 2006. Failure prediction with self organizing maps. Expert Syst. Appl. 30 (3), 479–487.
Lundberg, S., Lee, S., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30.
Owens, E., Sheehan, B., Mullins, M., Cunneen, M., Ressel, J., Castignani, G., 2022. Explainable artificial intelligence (XAI) in insurance. Risks 10 (12), 230.