A Novel Approach to xAI Techniques: Improving Accuracy and Interpretability of Tree-Based Models

Jaume Belles-Sampera, PhD. Riskcenter IREA (UB), Grupo Catalana Occidente, S.A.

5th European
Congress of Actuaries
www.eca2024.org

About the speaker

- **Jaume Belles Sampera –** *Director of the Actuarial Function for Life, Health and Burial Businesses. GCO*
  Jaume holds a MsC in Mathematics and a PhD in Business from the University of Barcelona. He has a broad experience in the Spanish insurance and financial sectors. In 2015 he joined GCO to take part in the Actuarial Function, where he is responsible for the Life, Health and Burial Businesses since late 2020. He is a part-time researcher of the Riskcenter IREA (UB).

- GCO (Grupo Catalana Occidente) is one of the leaders in the Spanish insurance sector and credit insurance in the world, as well as the leader in the funeral sector in the Iberian Peninsula. With constant growth and extensive implementation, it has more than 8,500 employees and it is present in more than 50 countries.

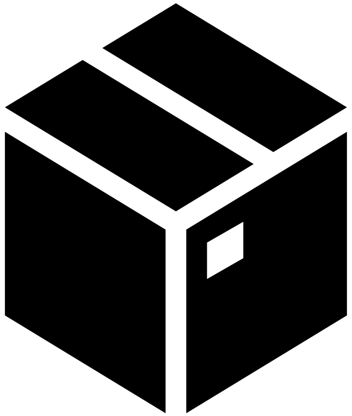
Starting point: our research goals

- Obtaining insights on **risks linked to the behaviour of policyholders** in universal life policies
- Exploring **how to explain** the classifying process behind a non-classical predictive model
- Can we contribute to **enhance xAI?**

Explainable AI – Model Agnostic

Tools for ***global*** explanations

- Feature Importance (FI)
- Partial Dependence Plots (PDP) with H-Statistic
- Accumulated Local Effects (ALE)

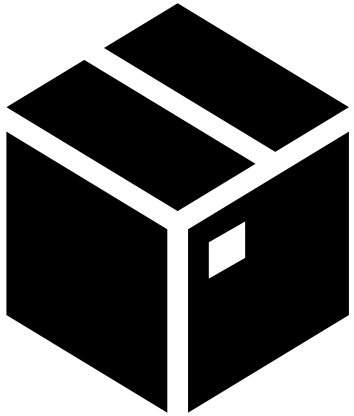


Tools for ***local*** explanations

- Shapley Additive Explanations (SHAP)

Following the classification proposed in **Adadi and Berrada (2018)**

Our main contribution



Tools for ***global*** explanations

- Feature Importance (FI)
- Partial Dependence Plots (PDP) with H-Statistic
- Accumulated Local Effects (ALE)

- KNN based on Shapley values

Tools for ***local*** explanations

- Shapley Additive Explanations (SHAP)

Inspired by **Bussmann et al. (2020)** and using Kohonen Neural Networks (KNN)

The Application

- Behavioral risk: *paid-up* risk

- Data:

  The database contains 7,886 UL policies that were actively paying premiums in 2018 and remained within the portfolio throughout the entirety of 2019.

  Paid-up: 1,665 policies (21.11%)
  Active: 6,221 policies (78.89%)

- Response variable: one-year ahead *paid-up* probability

- Predictive models: logistic regression | random forest | XGBoost

Explanatory variables

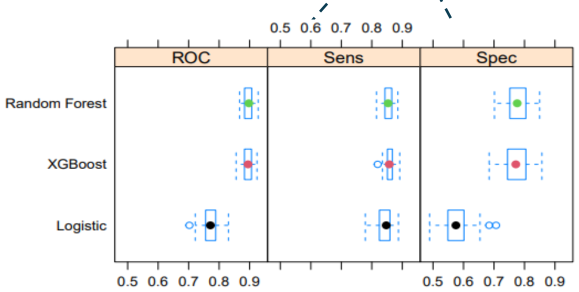
Categorical variables		
gen	factor	Gender of the insured (1: Female, 0: Male)
unl	factor	Unit-linked product (1: Yes, 0: No)
tax	factor	Product with tax advantages (1: Yes, 0: No)
fee	factor	Product with active surrender fee (1: Yes, 0: No)
rate	factor	Product with fixed guaranteed interest rate (1: Yes, 0: No)

Continuous variables		
cap	numeric	Additional sum insured in case of death
res	numeric	Current value of the fund
prem	numeric	Total annual premium
loy	numeric	Number of years the policy has been in force
age	numeric	Current age of the insured
rem	numeric	Years remaining until final premium as per contract

Steps

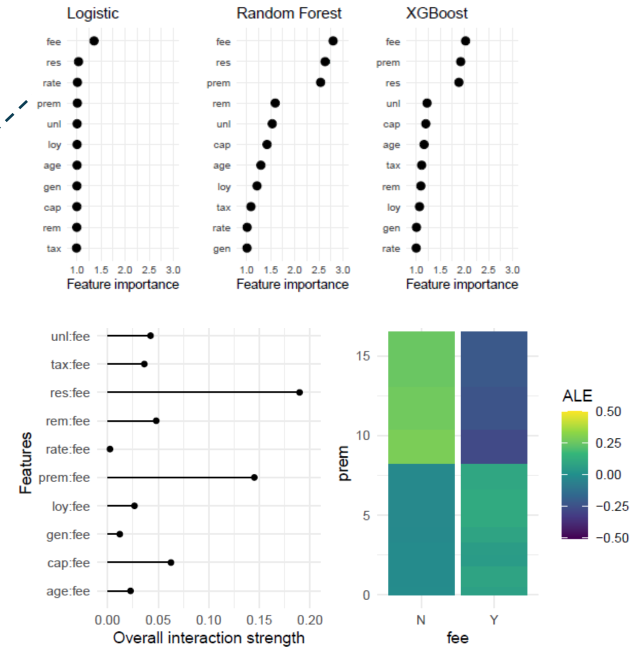
Preparatory phase

Accuracy analysis



Fitting models

Interpreting the selected model

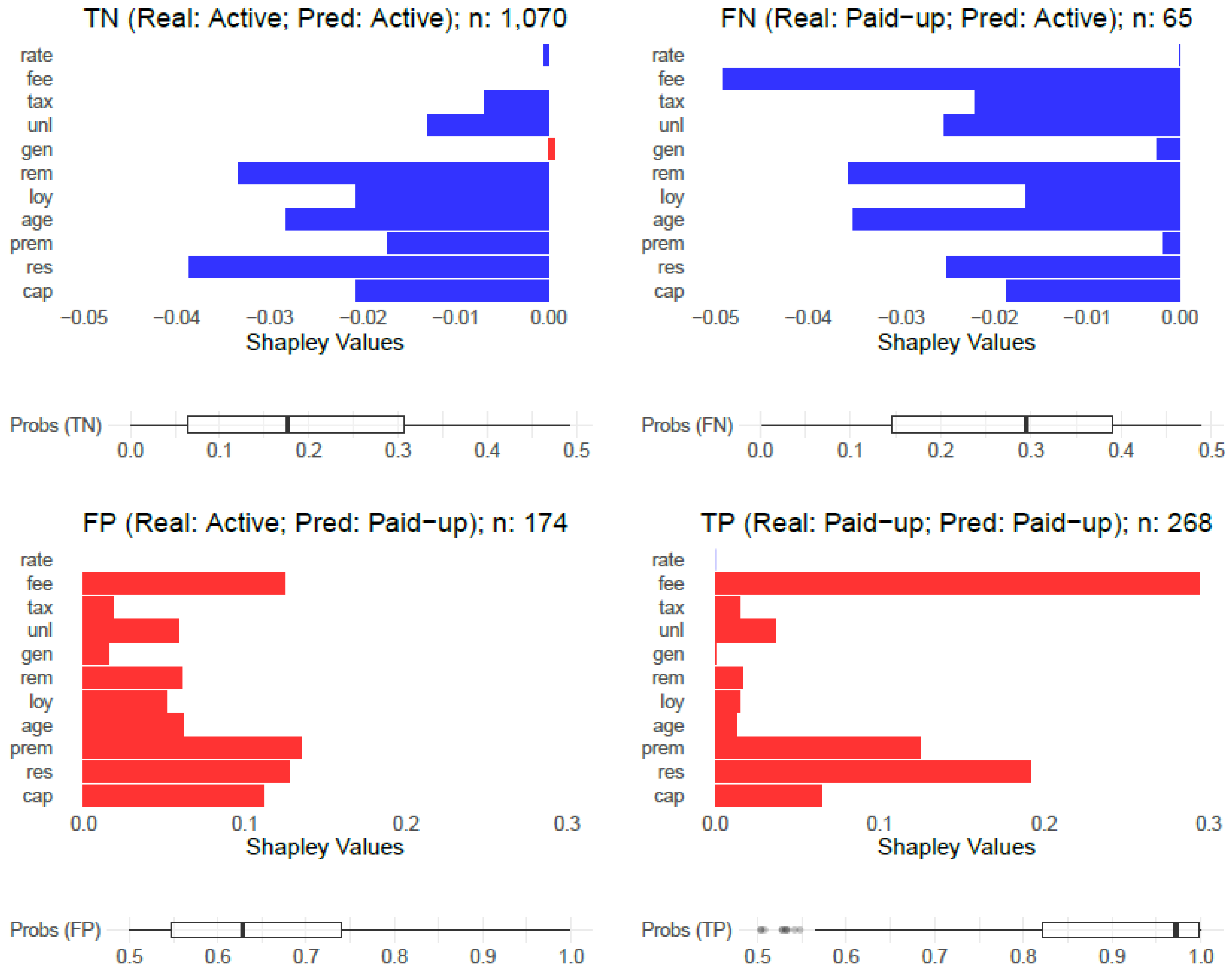


Improving explainability

Conclusions

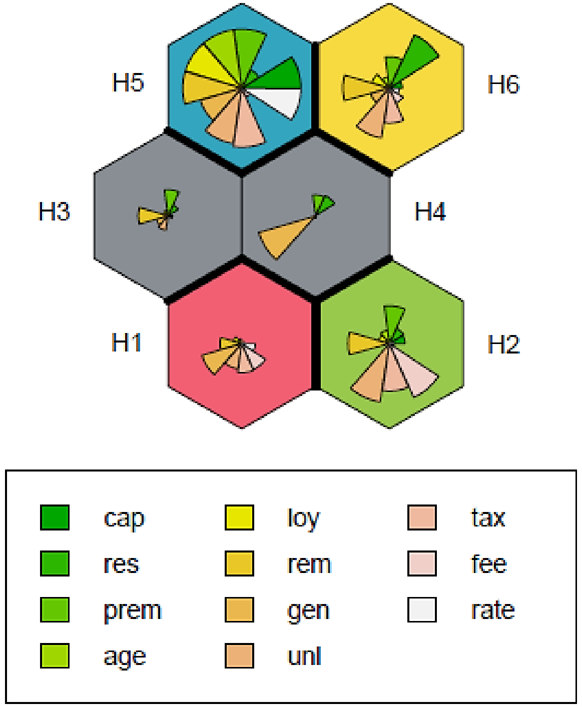
Improving explainability

First approach to transform a *local* explanatory tool (SHAP) into a *global* one



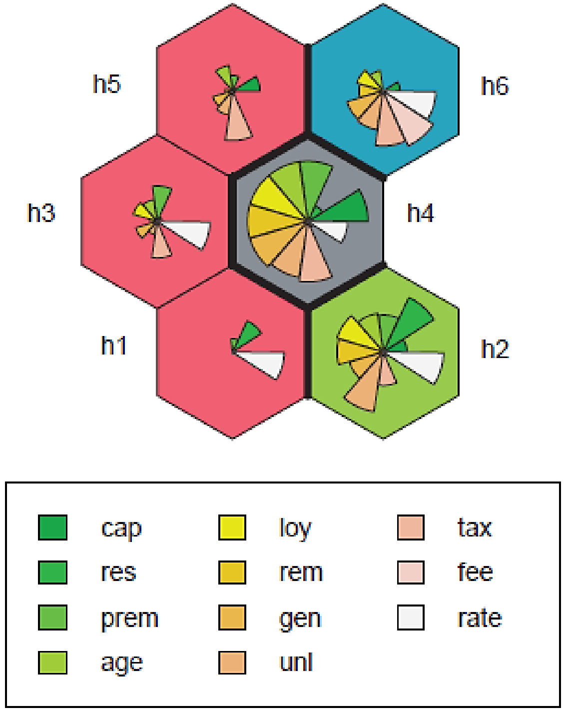
Improving explainability

KNN based on Shapley values



Extreme prediction values
(below 0.3 and above 0.7)

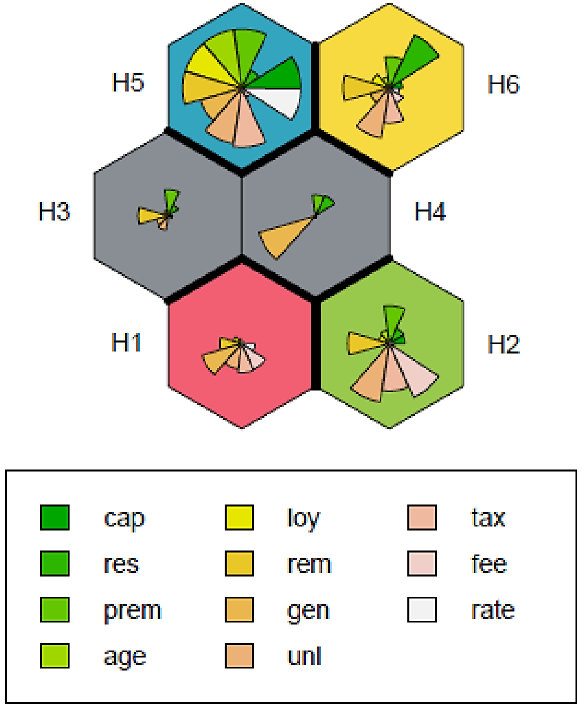
70% of total observations
92% are classified correctly



Prediction values close to the threshold of 0.5
(between 0.3 and 0.7)

30% of total observations
68% are classified correctly

Improving explainability



Extreme prediction values
(below 0.3 and above 0.7)

70% of total observations
92% are classified correctly

Predicted as Active cases:

- **H1**: surrender fee products with high premiums and reserves.
- **H3** and **H4**: non surrender fee products with low premiums and reserves.

Predicted as Paid-up cases:

- **H2**: surrender fee products with low premiums and reserves.
- **H5**: non surrender fee products with high premiums and reserves.
- **H6**: both surrender and non surrender fee products with lowest premiums and lowest reserves.

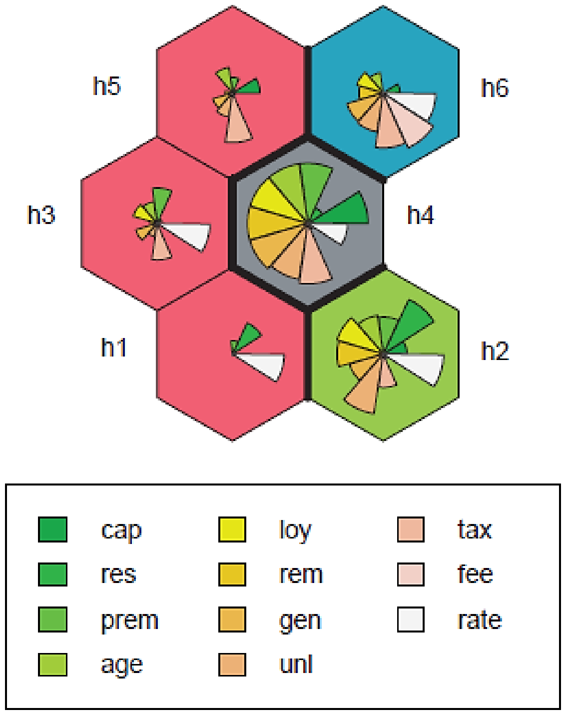
Improving explainability

Predicted as Active cases:

- **h1 (H3)**: non surrender fee products with low premiums and reserves, but the proportion of products offering tax benefits is higher than that of unit-linked products.
- **h3 (H4)**: non surrender fee products with low premiums and reserves, but with higher reserves than in H4 policies.
- **h5 (H5)**: non surrender fee products with high premiums and reserves, mainly unit-linked products.

Predicted as Paid-up cases:

- **h2** and **h4 (H5)**: non surrender fee products with high premiums and reserves, but lower than those of H5.
- **h6 (H1)**: surrender fee products with probabilities that are very close to 0.5 and predicted in almost equal proportions to be Active or Paid-up.



Prediction values close to the threshold of 0.5 (between 0.3 and 0.7)

30% of total observations
68% are classified correctly

Conclusions

- In the analysed portfolio, the paid-up probability depends mainly on the interaction between:
  - the inclusion (or not) 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).

- May risk managers rethink the rules governing surrender fees?

Link

- This work is part of the PhD dissertation (in progress) of Mr. David Anaya. It is a joint work with Professor Lluís Bermúdez and myself (his PhD supervisors).

- The opinions expressed in this work are those of the authors and do not necessarily represent those of GCO.

- You can find the full paper clicking on the link below (open access)

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

References

- Adadi, A. and M. Berrada (2018). “Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)”. *IEEE Access* 6, 52138-52160.
- Arrieta, A.B. et al. (2020). “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI”. *Information fusion* 58, 82-115.
- Azzone, M. and Barucci, E. and Moncayo, G.G. and Marazzina, D. (2022). “A machine learning model for lapse prediction in life insurance contracts”. *Expert Systems with Applications* 191, 116261.
- Bussmann, N. and Giudici, P. and Marinelli, D. and Papenbrock, J. (2020). “Explainable AI in fintech risk management”. *Frontiers in Artificial Intelligence* 3, 26.
- Costa e Silva, E. and Lopes, I.C. and Correia, A. and Faria, S. (2020). “A logistic regression model for consumer default risk”. *Journal of Applied Statistics* 47(13-15), 2879–2894.
- Eling, M. and Kiesenbauer, D. (2013). “What policy features determine life insurance lapse? An analysis of the German market”. *The Journal of Risk and Insurance* 81(2), 241–269.
- Friedman, J.H. and Popescu, B.E. (2008). “Predictive learning via rule ensembles”. *The Annals of Applied Statistics*, 916–954.
- Huysmans, J. and Baesens, B. and Vanthienen, J. and Van Gestel, T. (2006). “Failure prediction with self-organizing maps”. *Expert Systems with Applications* 30(3), 479-487.
- Lundberg, S.M. and Lee, S.I. (2017). “A unified approach to interpreting model predictions”. *Advances in Neural Information Processing Systems* 30, 4768-4777.

Thank you

Contact Details:

belles.sampera@ub.edu
jaume.belles@gco.com