Reacfin

IACA WEBINAR

Machine learning applications to non-life pricing and underwriting

JANUARY 2023

THIS PAGE IS LEFT BLANK INTENTIONALLY

SPEAKER'S INTRODUCTION

Xavier MARECHAL

CEO Reacfin and IA|BE qualified actuary

Expert in Non-Life and Health insurance (pricing, product development, reserving and risk management) and Data Science.

Reacfin Consulting

We offer consulting services in actuarial science & quantitative finance, including a.o. capital - portfolio - product - risk - and liquidity - management. We build our expertise on broad data science capacities.

Reaxii

By blending strong actuarial and financial business expertise with an in-depth understanding of cutting-edge IT technologies, Reaxii enables our clients to become more competitive and focus on their core business such as complex analysis, strategic decision-making and innovation.

Reacfin Academy

We share our knowledge with our clients. We offer a comprehensive learning platform, including on-site trainings, e-learning modules, webinars etc.

Challenges in insurance (with a focus on non-life insurance)

Increasing competition

- Commoditisation of insurance products
- Pricing comparison systems
- Sophistication in pricing
- Insurtechs simplifying products/processes

Availability of new data sources

- External data (IoT, open data,…)
- Use of unstructured data

New customers needs and behavior

- Digitalisation of underwriting process
  Direct vs Brokers
- New risks emerging
- Focus on price (made possible thanks to pricing comparison systems)

To adress these challenges, Insurers have to

- Innovate in product development and surrounding services
- **Capture and identify relevant features for pricing models**
- Adapt faster to market changes (identification of risks, building new models, faster product deployement)
- Improve processes (e.g. claims management) to increase added-value to clients.

AGENDA

Some useful ML techniques

Applications to pricing and underwriting

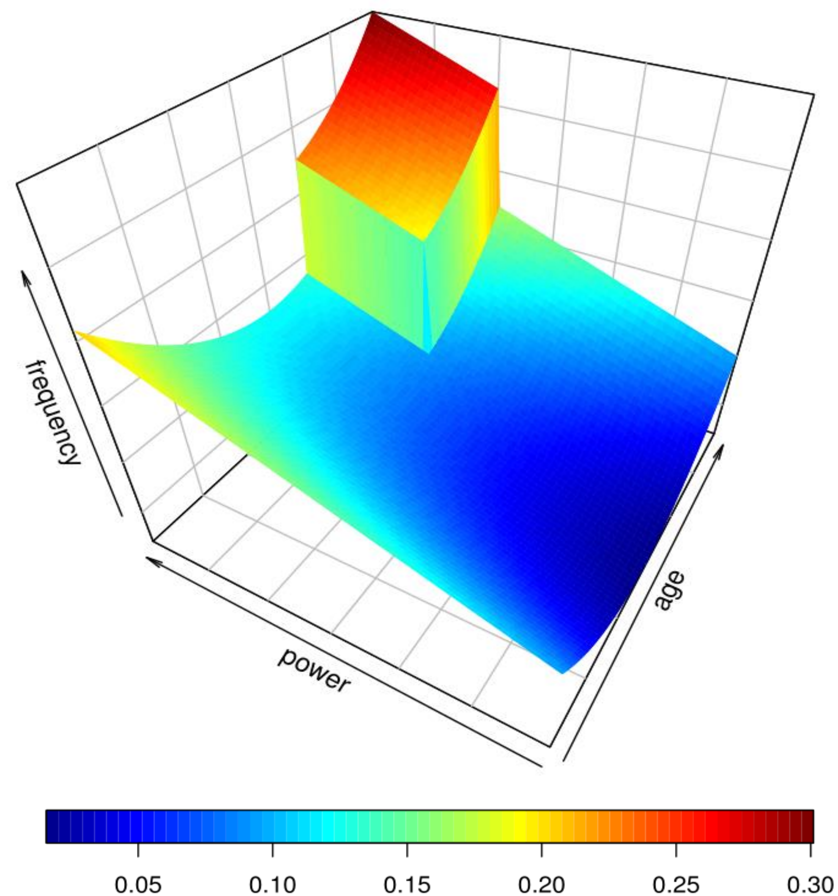
Challenges with Machine Learning techniques

Specificities of the Poisson frequency surface

- The Poisson frequency λ has the following properties:
    1. the first term is quadratic in the variable age,
    2. the second term is linear in the power,
    3. the third term is a nonlinear interaction between the two variables.

- It has been chosen to « fail » standard statistical methods (as GLM, see *infra*) and therefore show how some machine learning methods can « fix » these issues.

- We then divide our dataset in two subsets: a train dataset and a test dataset.

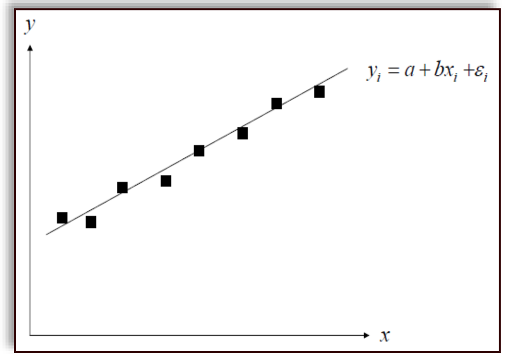


GENERALIZED LINEAR MODELS

GLM are still widely used by insurance companies for non-life pricing and other applications

Linear Model ("LM")

- $Y = \beta 0 + \beta_1 . X_1 + \cdots + \beta_n . X_n + \varepsilon$
- Y is a direct **linear combination** of explanatory variables
- The errors are assumed to be **Normally distributed**: $\varepsilon \sim N(0, \sigma^2)$
- And so, $Y \sim N(\mu, \sigma^2)$



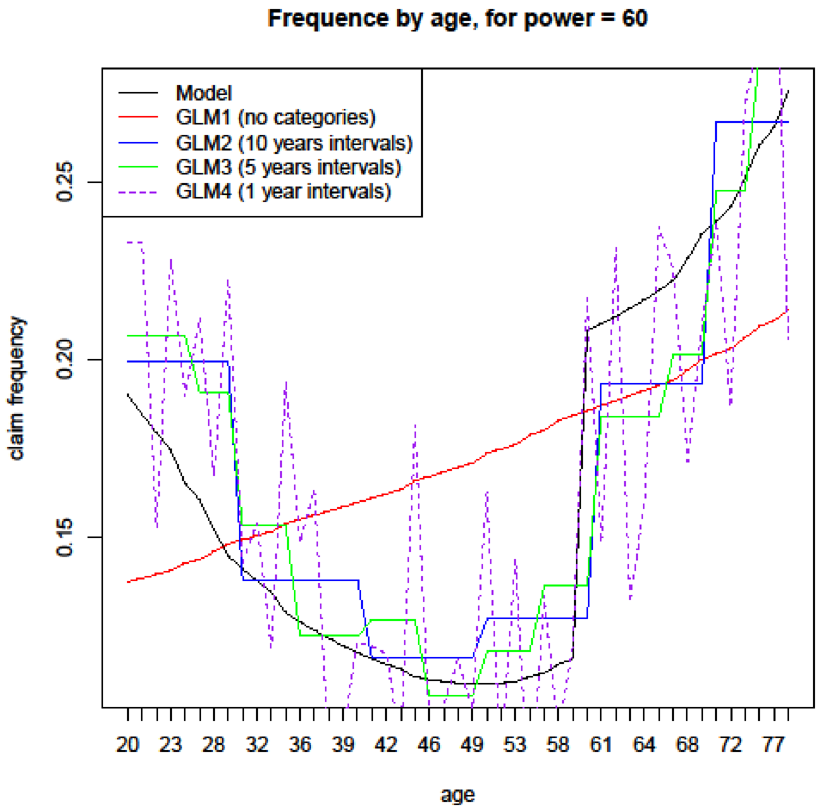
Generalized Linear Model ("GLM")

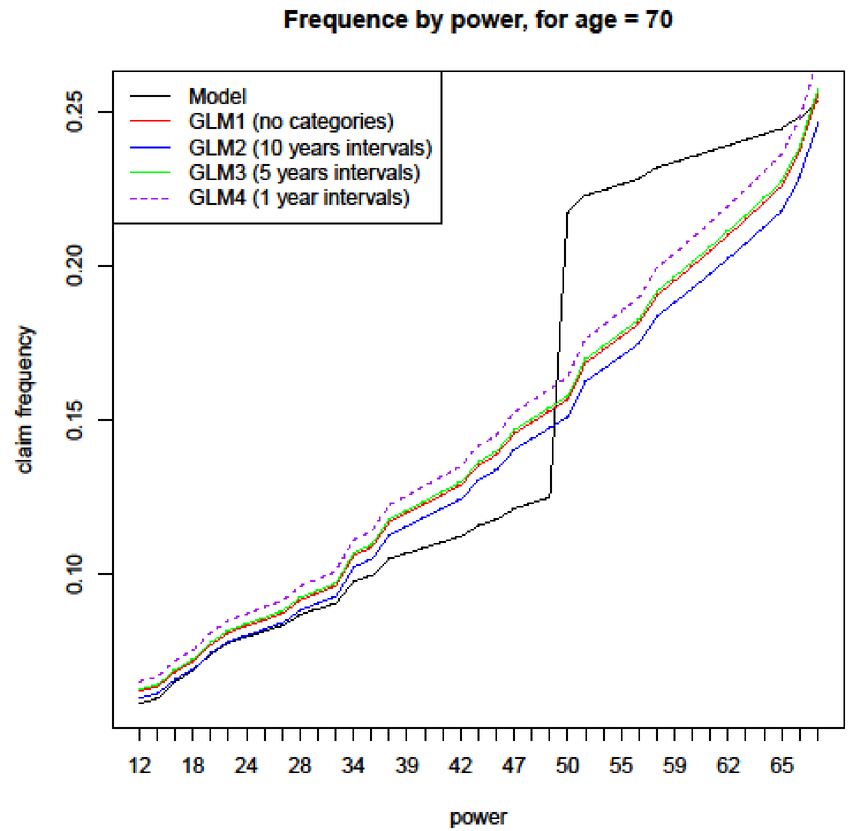
- $Y = g^{-1}(\beta 0 + \beta_1 . X_1 + \cdots + \beta_n . X_n) + \varepsilon$
- Y is now a **function (g^{-1}) of a linear combination** of the explanatory variables
- The distribution of the response variable **does not need to be Gaussian anymore**
- The features X_i are usually categorical as entering a continuous feature X_i^* in the linear predictor boils down to assume a linear effect of the X_i^* on the score scale: in log-linear models, this means that the mean is constrained to vary exponentially with X_i^*

Distributions
$Bin(1, \mu)$
$Poi(\mu)$
$Nor(\mu, \sigma^2)$
$Gam(\mu, \alpha)$
$IGau(\mu, \sigma^2)$

GENERALIZED LINEAR MODELS

GLM fail to adequately capture the interaction between age and power

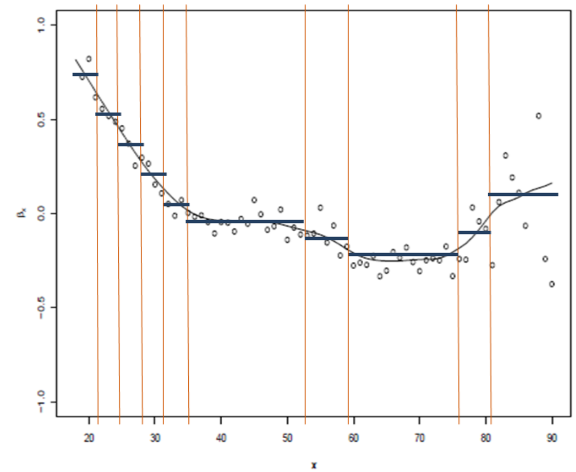




GENERALIZED ADDITIVE MODELS

Generalized Additive Models ("GAM") allow to model continuous variables

- A usually good solution to model continuous variables is to use a **semi-parametric approach**: if we are not sure about the type of influence of X we would prefer fitting a model with an additive score of the form



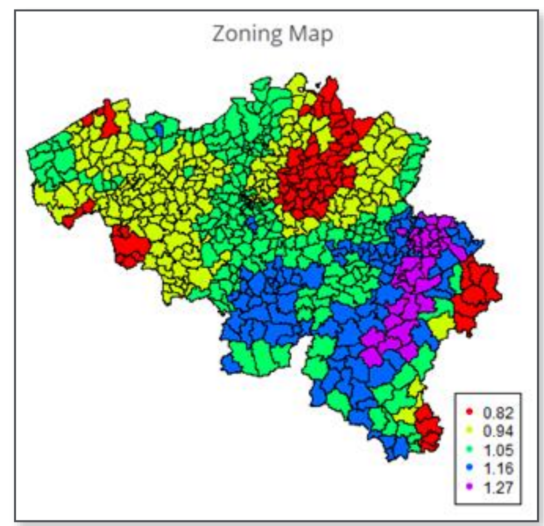
$$linear\ part\ +\ f\ (X)$$

where f is left unspecified and estimated from the data

- The mean μ_i of Y_i is linked to the nonlinear score via

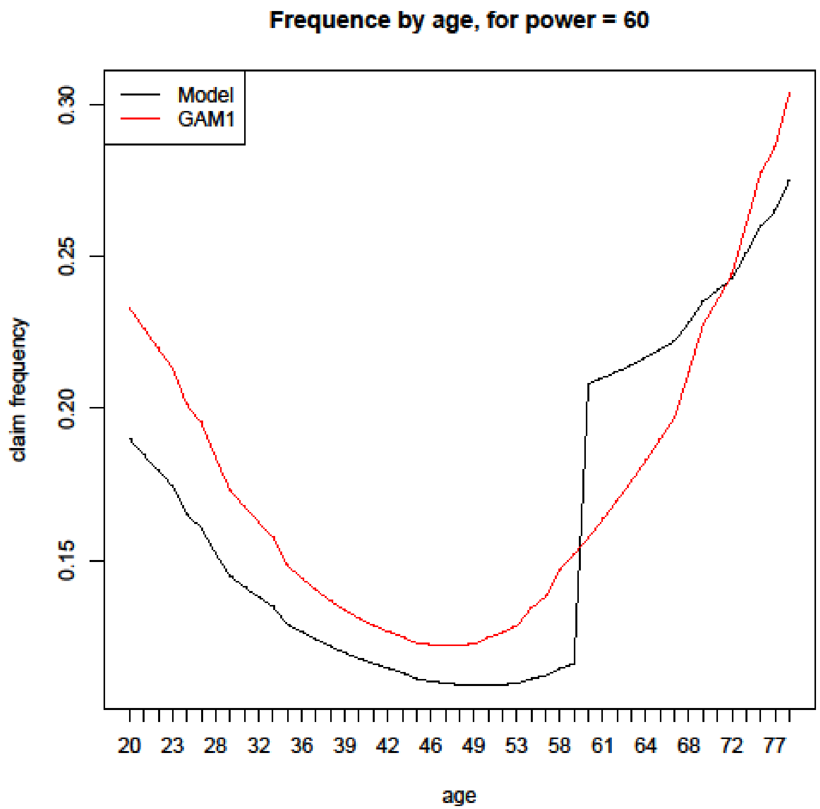
$$g(\mu_i) = \beta_0 + \sum_{j=1}^{p_{cat}} \beta_j x_{ij} + \sum_{j=p_{cat}+1}^{p} f_j(x_{ij}) = score_i$$

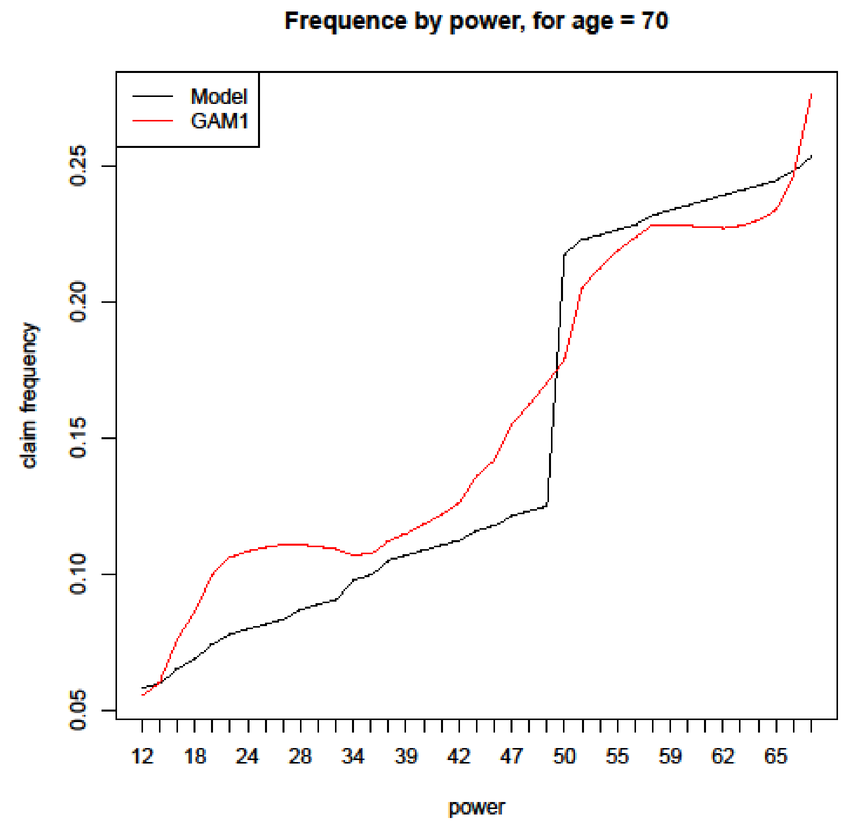
for some smooth unspecified functions f_j, where g is the link function



GENERALIZED ADDITIVE MODELS

GAM do not significantly improve GLM results





WHAT IS MACHINE LEARNING?

Objectives of Machine Learning ("ML")

ML algorithms aim at finding by themselves the method that best predicts the outcome of the studied phenomenon.

Supervised vs. Unsupervised learning

- **Supervised learning:**
  - Inputs and examples of their desired outputs are provided
  - The goal is to learn a **general rule that maps inputs to outputs**.

➔ *Given a set of training examples ($x_1, x_2,\ldots, x_n, y$), where y is the variable to be predicted, what is the most efficient algorithm to best approximate the realizations of y*

  - 2 main techniques
    - **Classification** : outputs are divided into two or more classes, and the learner must produce a model that assigns unseen inputs to one (or multi-label classification) or more of these classes.
    - **Regression**: the outputs are continuous rather than discrete.

- **Unsupervised learning:**
  - No labels are given to the learning algorithm
  - The goal is to **find structure in its input** (discovering hidden patterns in data).
  - Main technique
    - **Clustering**: a set of inputs is to be divided into groups. Unlike in classification, the groups may not be known beforehand.

Main use in non-life insurance

- Typically used to model **pricing or underwriting related variables**
  - Regression: frequency (#claims) or severity (claims cost)
  - Classification: lapse rates, conversion rates

- Typically used for **features engineering** (i.e. creating new variables)
  - E.g. vehicle classification, zoning,…

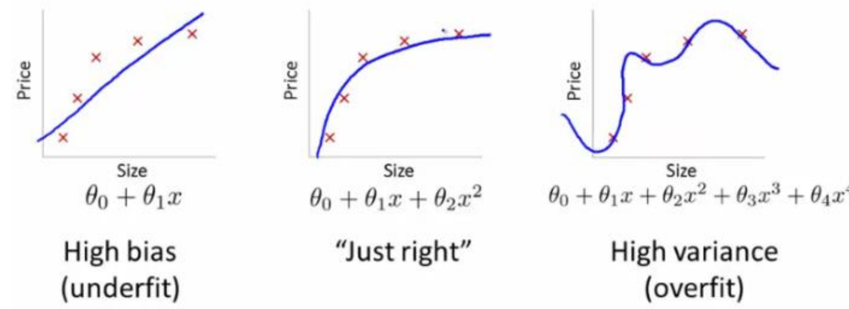
Focus on supervised models

OVERFITTING IS A CHALLENGE WITH MACHINE LEARNING MODELS

Overfitting deteriorates the predictive power of the model

The overfitting problem

- When modelling, we should be sensibilized with overfitting/lack of parcimony.
- It occurs when a statistical model **describes random error** or noise instead of the underlying relationship.
- The fact that the model fits our data well doesn't guarantee it will be a good fit to new data ➔ A good model is one that fits also well new data, i.e. that has a small predictive error



Bias-Variance Trade-off

- The **Prediction Error** can be decomposed as follows

$$E\left[\left(Y-\hat{Y}\right)^2\right] = \underbrace{\left(E[Y]-E[\hat{Y}]\right)^2}_{\text{Bias}} + \underbrace{Var\left(\hat{Y}\right)}_{\text{Estimation Variance}} + \underbrace{Var(Y)}_{\text{Pure randomness}}$$

- In general, we try to **minimize simultaneously the bias and the estimation variance** to get accurate predictions.
  - Usually, these two terms compete in the sense that a decrease in one of them typically leads to an increase in the other one.
  - This phenomenon is known as the **bias-variance trade-off** for which one needs to find a good balance (typically by controlling the complexity of the model).

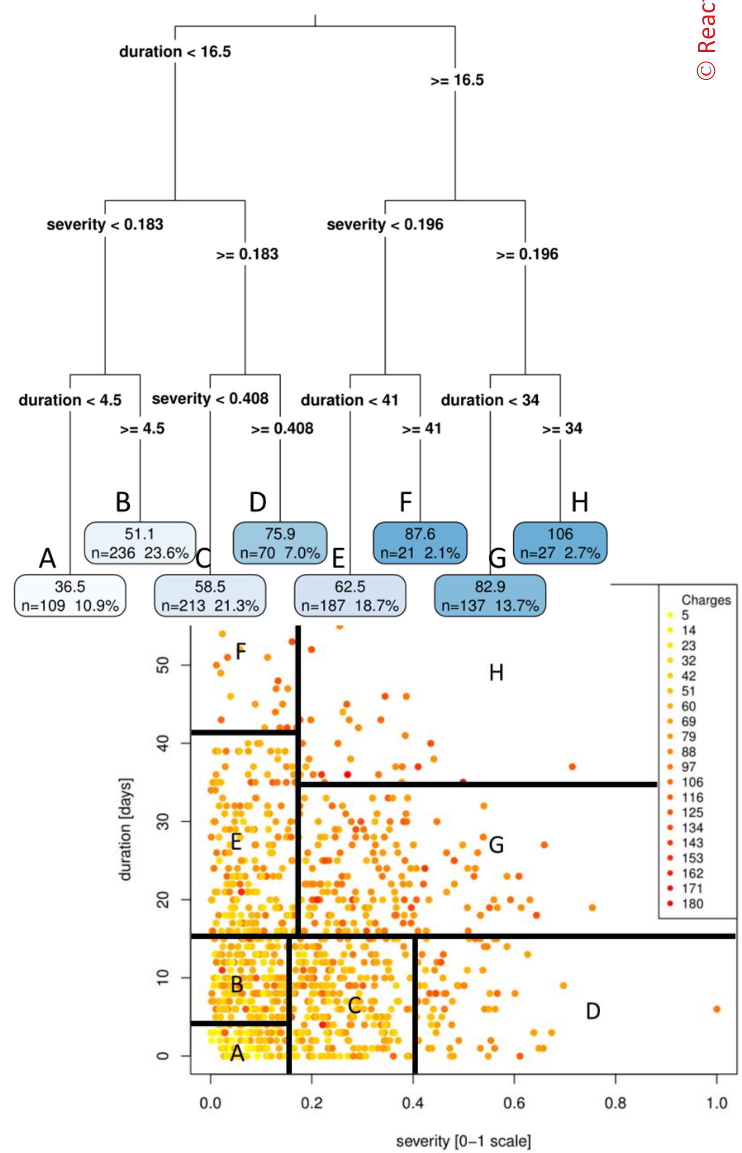
REGRESSION TREE ALGORITHM

Main idea

- Define a **loss/error** (or objective) **function** and
- Try to find regions $R_1, R_2, \dots, R_J$ that minimize (or maximize) the function retained
- All possible regions definitions can of course not be considered
- The tree algorithm therefore :
  - Starts with the global population and find the **optimal split of the predictor** at that level using the entire population
  - The same process is then applied on each sub-population
- In each sub-population, the estimation is obtained by **averaging** on the data points belonging to this sub-population

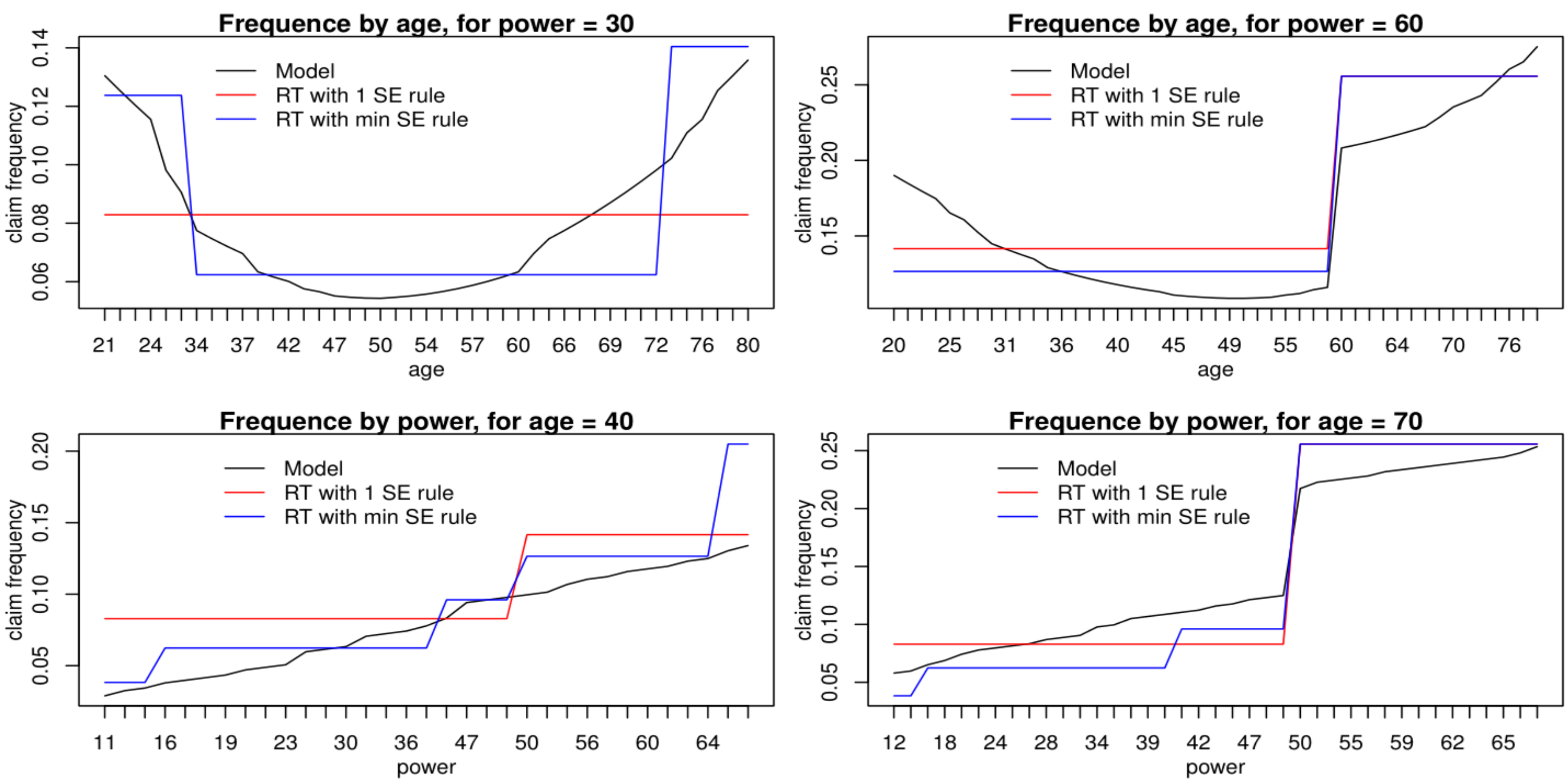
Important remarks

- The division decision is done in function of information available at moment before division execution
  - There is not warranty that the division decision taken is the best alternative insight to future divisions
- **Pruning** can be used to **reduce the size of the decision trees and its complexity**.
  - It is done by comparing its predictive power with trees having larger number of decision nodes.



REGRESSION TREE

Results of the simulated DB

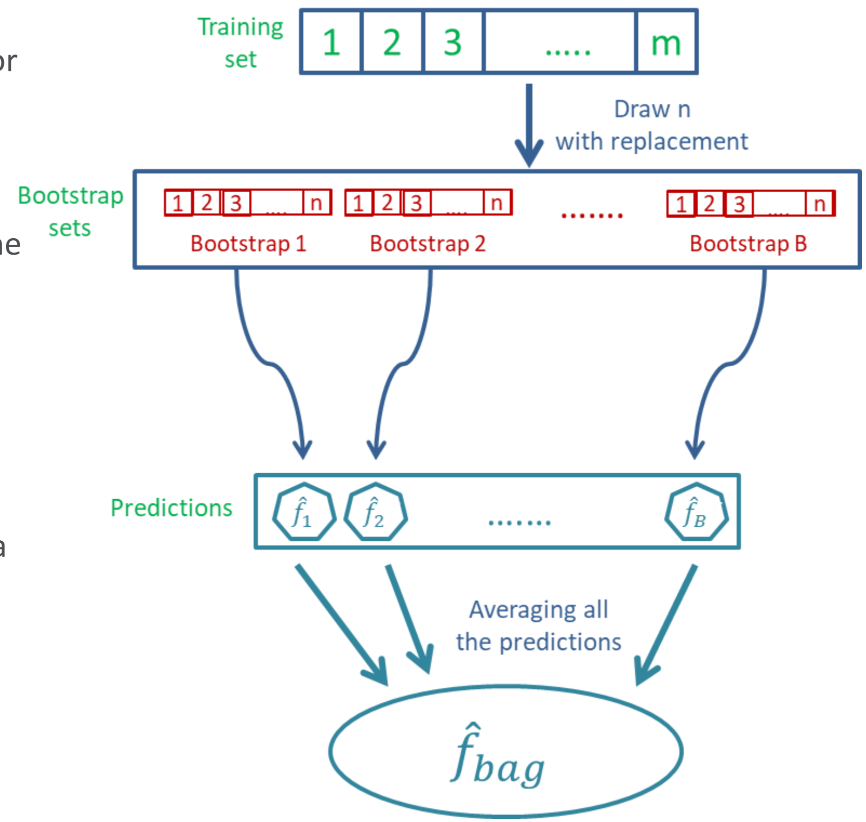


BOOTSTRAP AGGREGATION (BAGGING)

Bagging allows for variance reduction by averaging over several regression trees

Main idea

- **B**ootstrap **agg**regation, or **Bagg**ing, is a general-purpose procedure for reducing the variance of a statistical learning method
- Frequently used in the context of decision trees.
- Recall that given a set of n independent observations $Z_1, Z_2, \dots, Z_n$ each with variance σ^2, the variance of the mean $\bar{Z}$ of the observations is given by $\frac{\sigma^2}{n}$.
- **Averaging a set of observations reduces variance**. Usually multiple training sets are not at disposal



Algorithm

1. Bootstrap, by taking **repeated samples** from the (single) training data set.
2. Generate B different training data sets.
3. **Train our method** on the b^{th} bootstrapped training set to get $\hat{f}_b(x)$ the prediction at point x.
4. We then **average all the predictions** to obtain:

$$\hat{f}_{bag}(x) = \frac{1}{B}\sum_{b=1}^{B} \hat{f}_b(x)$$

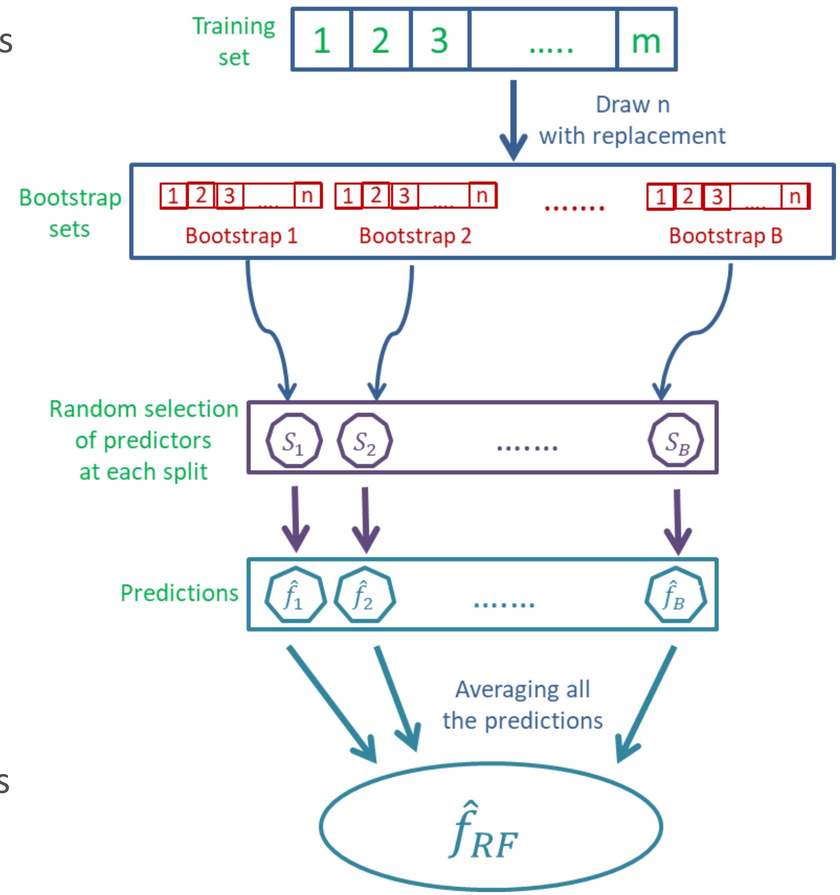
Random Forests improve bagging by decorrelating the trees

Main idea

- Random forests provide an improvement over bagging thanks to an additional step that **decorrelates the trees**. This **reduces the variance** when we average the trees.

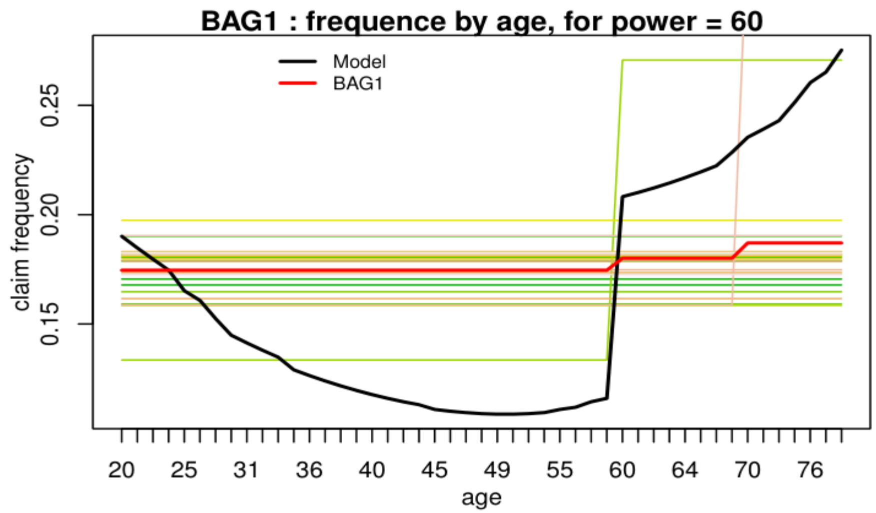
Algorithm

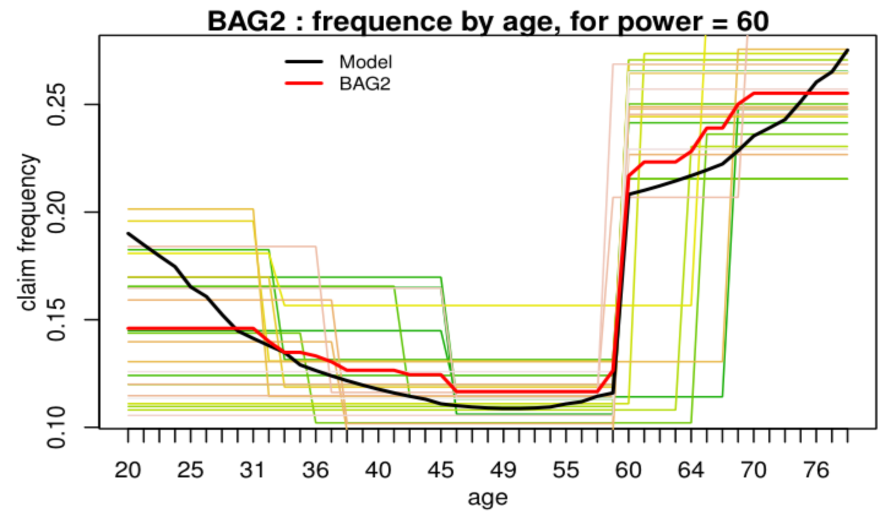
- As in bagging, we build several decision trees on **bootstrapped training samples**.
- But when building these decision trees, each time a split in a tree is considered, a **random selection of m predictors** is chosen as split candidates from the full set of p predictors. The split is allowed to use only one of those m predictors.
- A fresh selection of m predictors is taken at each split, and typically we choose $m \approx \sqrt{p}$ that is, the number of predictors considered at each split is approximately equal to the square root of the total number of predictors.

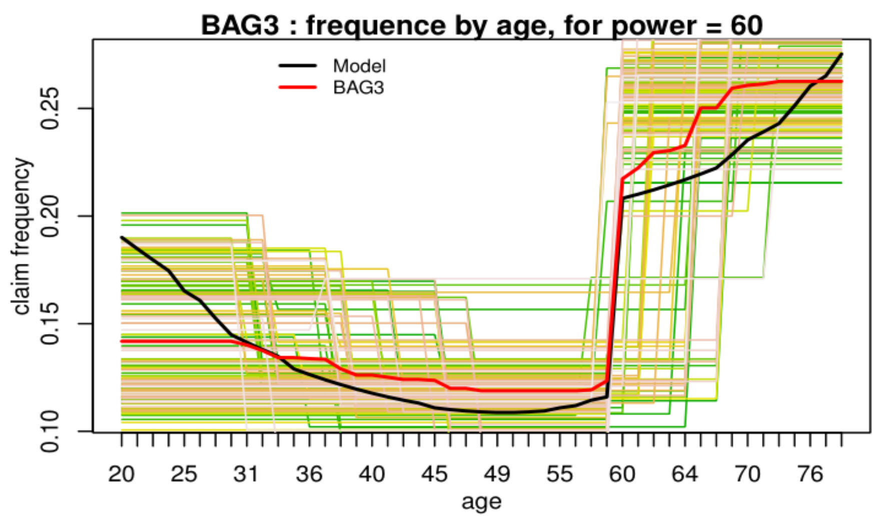


BAGGING

Results of the simulated DB







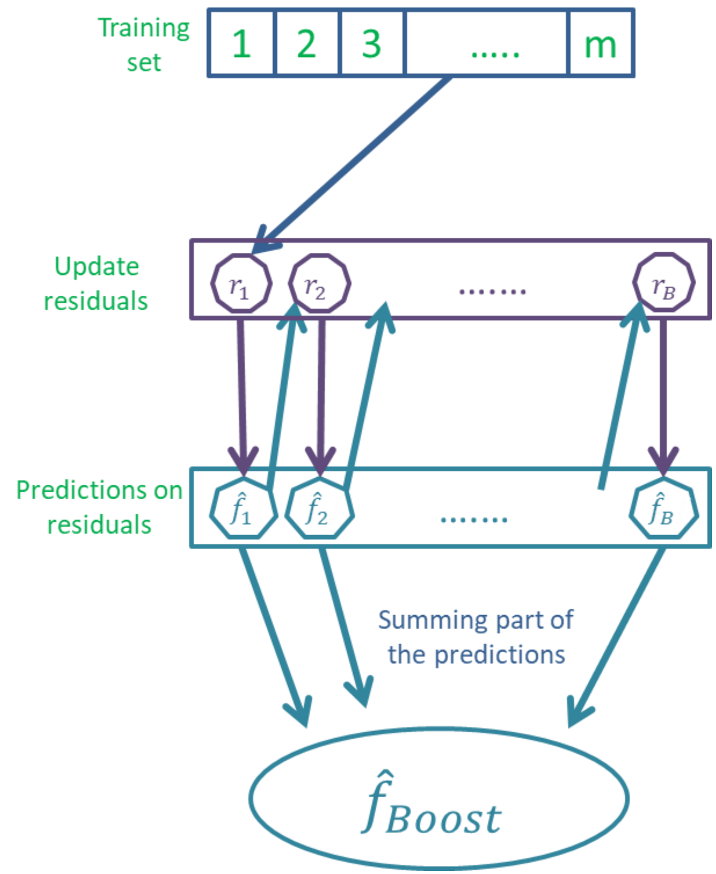
Algorithm

1. Set $\hat{f}(x) = 0$ and $r_i = y_i$ for all i in the training set

2. For $b = 1, 2, 3, \dots, B$, repeat :
   - Fit a tree $\hat{f}^b$ with d splits ($d + 1$ terminal nodes) to the training data (X, r)
   - Update $\hat{f}$ by adding in a reduced (shrunken) version of the new tree:
   $$\hat{f}(x) \leftarrow \hat{f}(x) + \lambda \hat{f}^b(x)$$
   - Update the residuals:
   $$r_i \leftarrow r_i - \lambda \hat{f}^b(x_i)$$

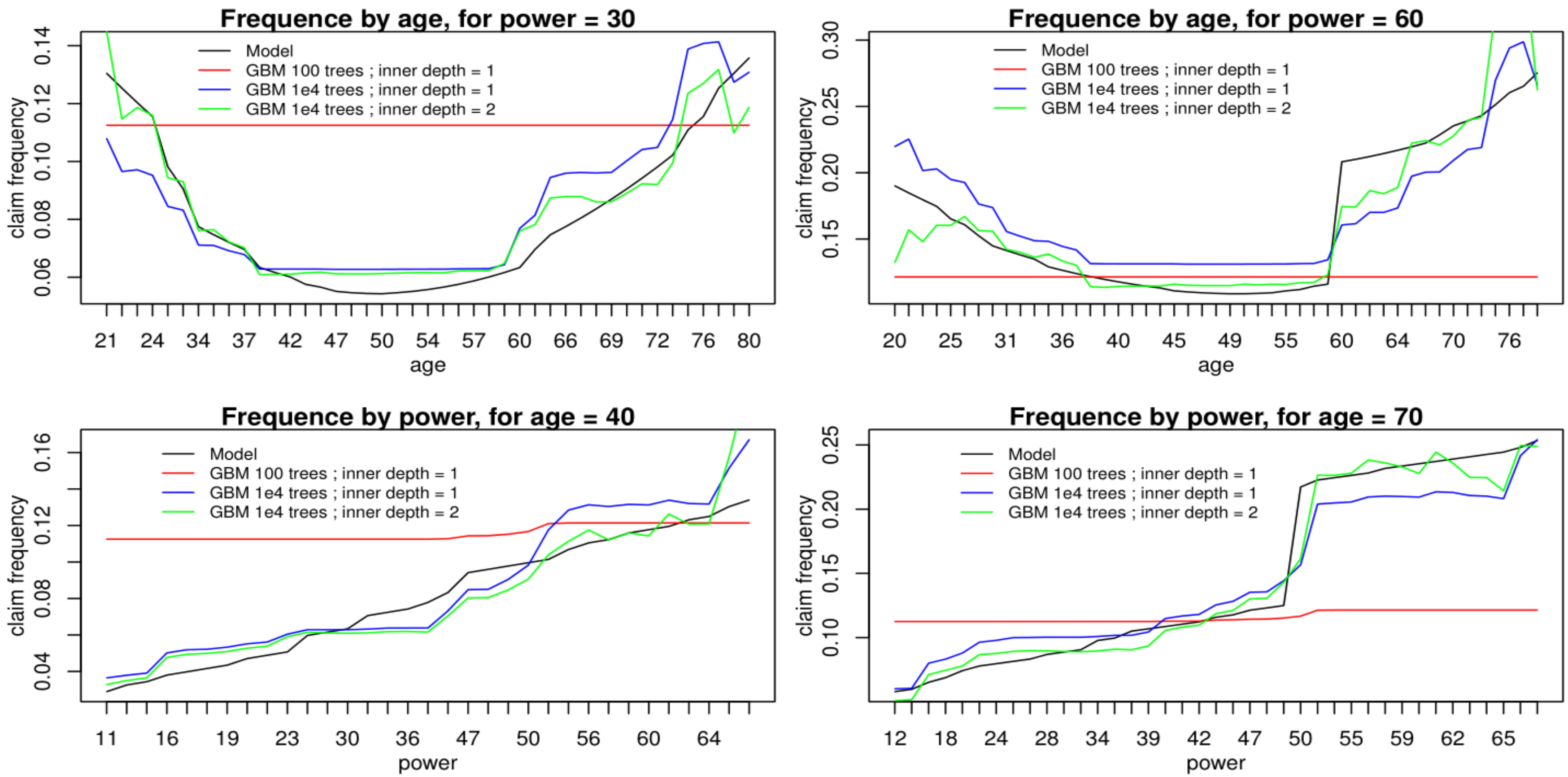
3. The final model is provided by

$$\hat{f}(x) = \sum_{b=1}^{B} \lambda \hat{f}^b(x)$$



GRADIENT BOOSTING MACHINE

Results of the simulated DB



AGENDA

Some useful ML techniques

Applications to pricing and underwriting

Challenges with Machine Learning techniques

TECHNICAL PRICING IS NOT THE ONLY APPLICATION OF ML TECHNIQUES: ML COULD ALSO HELP TO BOOST THE UNDERWRITING AND PORTFOLIO MANAGEMENT PROCESS

D SEGMENTATION

Segmentation and **pricing variables**
- Greater segmentation for greater risk selectivity and higher profitability
- Monitor concentrations of certain risk types

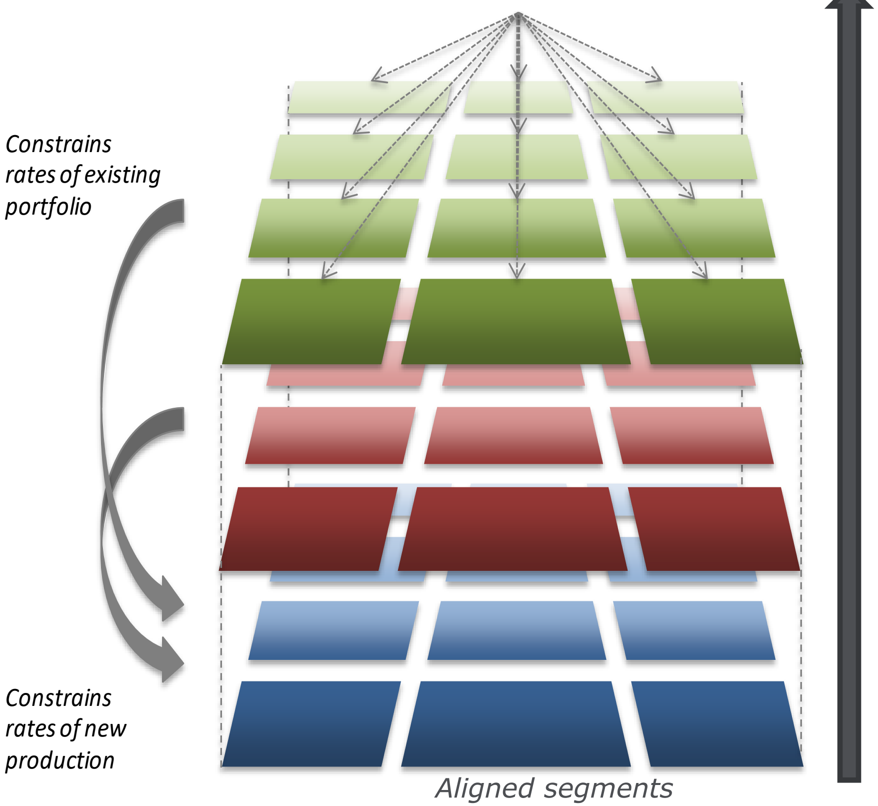
E **SCENARIO TESTING AND OPTIMISATION**

Impact of different scenarios on strategic indicators and optimization

C CLIENT BEHAVIOR

B COMPETITION

A TECHNICAL PRICING



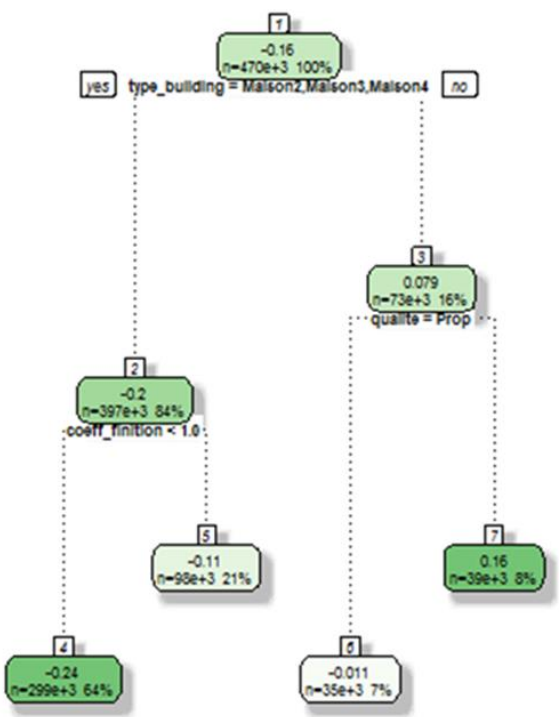
Customer behavior by segment
- Elasticity model help estimate pace at which rates can be increased by segment
- Focus Sales & Marketing to increase retention of better risks
- Building conversion rates model to better target clients

Competitor prices by segment and own current rates
- Position insurer to be competitive on certain segments

Portfolio profitability by segment, based on:
- Cost of risk (e.g. measured through GLMs or ML techniques)
- Portfolio composition (representativeness of each segment: total portfolio and recent new business)

Tree-based techniques can be used to compare Risk Premium and Commercial premium

- Thanks to tree-based methods (and variable importance) it is possible to identify the variables that are the most relevant to explain the differences between the risk premium and the current commercial premium
  - It helps in **defining the most relevant variables** that can, for example, then be included in a **profitability heatmap**

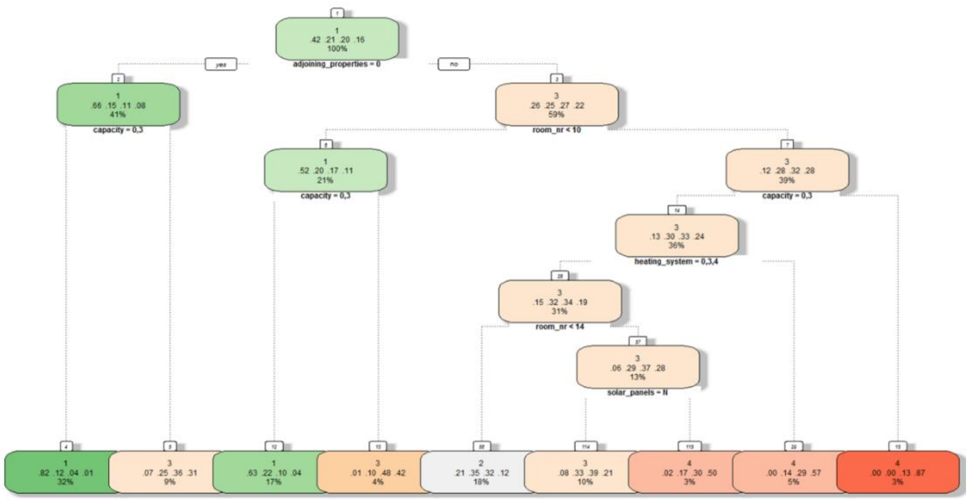


		coeff_finition				
type_building	qualite	A/0.8	B/1	C/1.1	D/1.15	E/1.2
Apparte	Loca	1.43	1.37	1.48	1.63	
	Prop	1.12	1.07	1.16	1.29	
Maison2	Loca	0.99	1.02	1.14	1.16	
	Prop	0.73	0.84	0.94	1.07	1.01
Maison3	Loca	0.92	0.93	1.05	1.14	
	Prop	0.72	0.80	0.90	0.98	0.96
Maison4	Loca	0.99	0.99	1.12	1.20	
	Prop	0.80	0.86	0.96	1.04	1.98

COMPETITION ANALYSIS TOOL

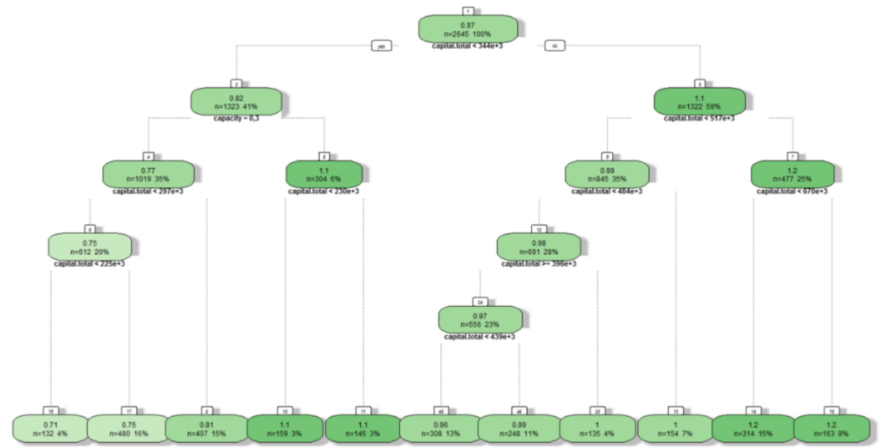
Tree-based techniques can be used to identify positioning on market segments and capture price differences

- **Identifying the segments** in which the insurance company is **well-positioned** with respect to its competitors is an important driver of a dynamic pricing process. E.g. **Classification of segments** in function of the ranking of the competitors with **regression trees**



- **Reverse engineering** of the pricing (structure) of competitors can be enhanced with ML techniques

- **Analyze the price dispersion** of the company with respect to its competitors or with respect to the average market price

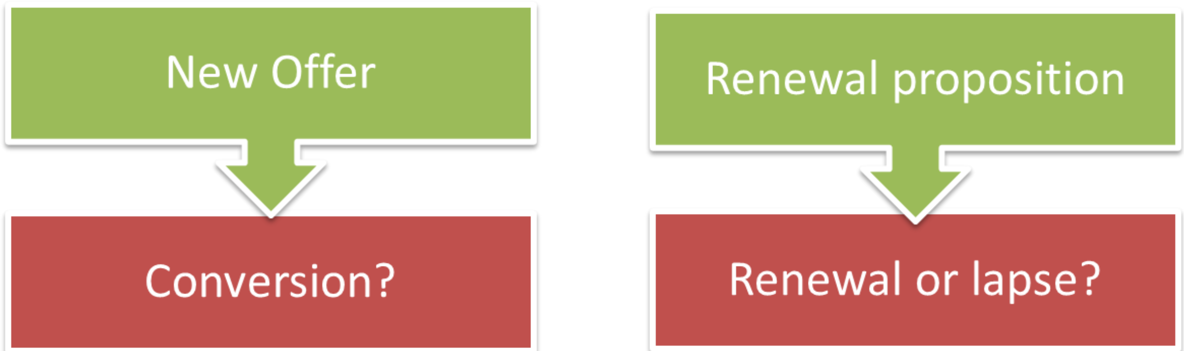


CLIENT BEHAVIOR

C

ML techniques can help improve the logistic regression

- The goal is to explain the conversion / lapse probabilities with some explanatory variables



- A dummy variable identifies the policies that were converted / renewed during the year
- Traditionaly Generalized Linear Models are used
  - E.g. A **logistic regression** can be performed on this dummy variable and potential explanatory variables

$$ln\left(\frac{\pi(x_1 \dots x_n)}{1-\pi(x_1 \dots x_n)}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$$

- Machine learning technique (e.g. **GBM**) are more and more often used as they usually **improve predictions** and allow to **find more complex patterns**

AGENDA

Some useful ML techniques

Applications to pricing and underwriting

Challenges with Machine Learning techniques

Comparing points of strengths

	Machine learning	Statistical modeling
Limits the number of **assumptions**	+	-
Inference: Assessing the reliability of modeling assumptions	-	+
Prediction: ability to extrapolate future or unobserved realizations of a variable given other explanatory observations	+	-/+
“Big Data”: ability to handle large sets of data both in terms of number of observations (“rows”) or variables (“columns”)	+	-
Human interactions: ability/need of incorporating material users ex-ante opinions (e.g. Expert Judgment)	-	+

- Results of Machine Learning algorithms will need careful attentions as they derive from automated procedures and could **induce conclusions which do not match a business logic ➔ Interpretability** is key for practical use as well as **ensuring fairness and avoiding discrimination**

- Another key challenge with Machine Learning is the risk of **overfitting**.
  - Overfitting relates to excessively complex models for which the large number of explanatory variables and parameters, is unreasonably important compared to the number of observations

AGENDA

Some useful ML techniques

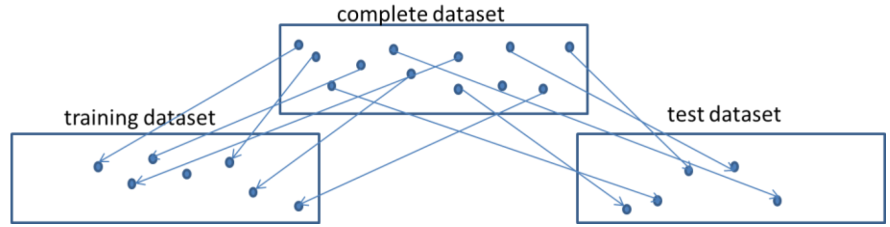
Applications to pricing and underwriting

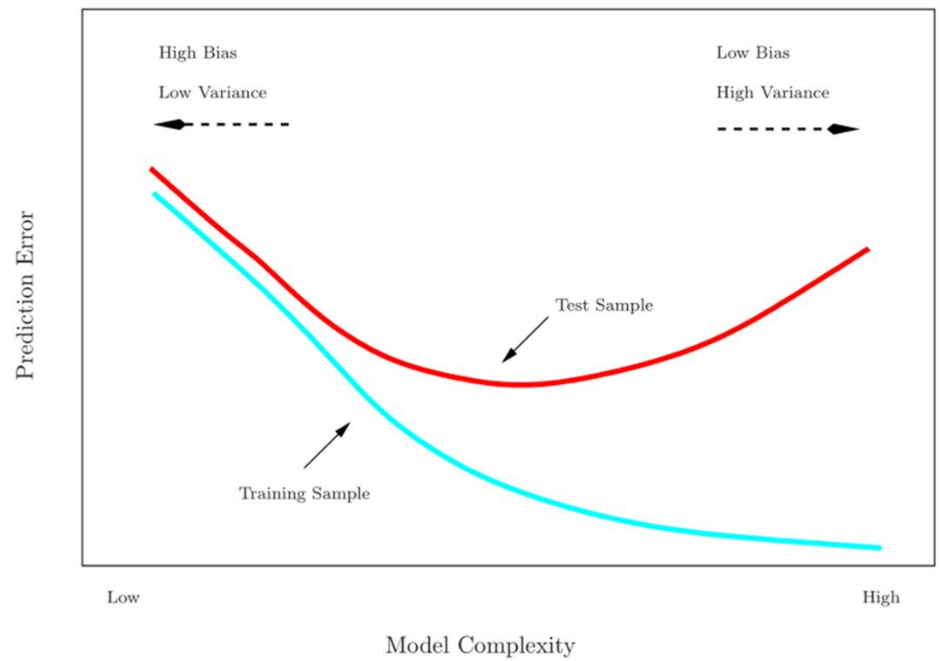
Challenges with Machine Learning techniques

- Overfitting
- Discrimination and fairness
- Interpretability

Overfitting can be reduced by separating the data into a training set and a test set

- Use two different datasets:
  - A **training set** to calibrate the model,
  - A **test set** to assess the model's predictive ability.
- Two different kinds of errors are defined:
  - The training error is calculated by applying the model to the observations used in its calibration
  - The test error is the average error that results from using the model to predict the response on a new observation, one that was not used in calibrating the model.
- The training error decreases with model complexity whereas the test error tends to increase when the level of model complexity creates overfitting
- The best solution is clearly to use a large test set. However, it is often not available!





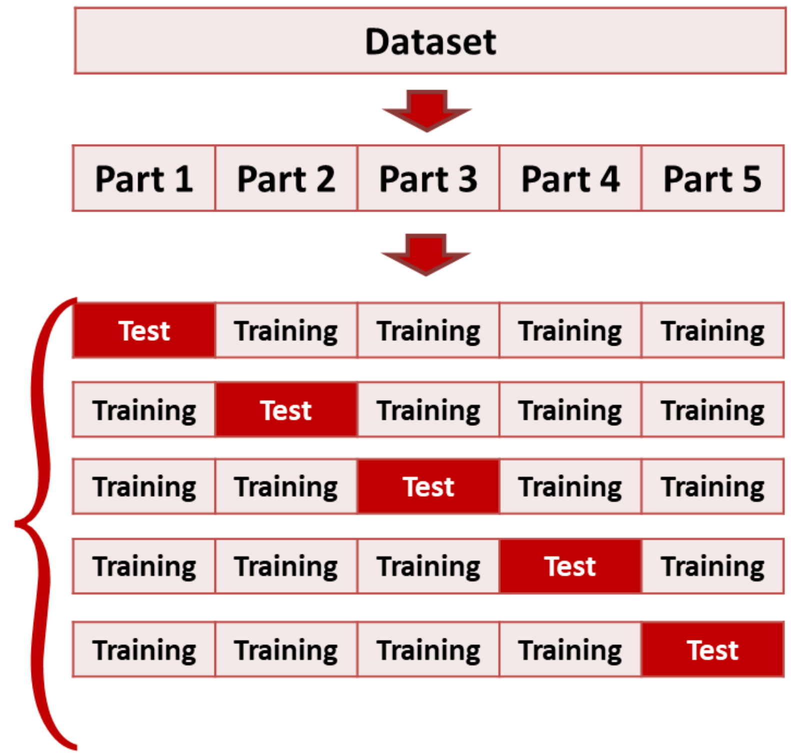
Drawbacks of training set / test set approach

- The method has some drawbacks:
  - The estimate of the test error can be highly variable, depending on precisely which observations are included in the training set and which observations are included in the test set.
  - In the test set approach, only a subset of the observations — those that are included in the training set rather than in the test set — are used to fit the model.

- This suggests that the test set error may tend to overestimate the test error for the model fit on the entire data set.

Cross-validation approach

- The idea of the method is to randomly divide the data into K equal-sized parts.
- We leave out part k, fit the model to the other $K - 1$ parts (combined), and then obtain predictions for the left-out k-th part.
- This is done in turn for each part $k = 1,2, \ldots K$, and then the results are combined.



AGENDA

Some useful ML techniques

Applications to pricing and underwriting

Challenges with Machine Learning techniques

- Overfitting
- Discrimination and fairness
- Interpretability

PRICING FAIRNESS CHALLENGE

Key challenge for insurance companies

Customer segmentation

- A fair premium, related to his/her risk profile, to minimize the potential for adverse selection.
- *i.e. the good risks could lapse and accept a lower premium elsewhere, leaving the insurer with an inadequately priced portfolio.*

Keeping pricing fairness :

Big data and ML models could lead to an increased segmentation among policyholders which has to be managed as well (to avoid non-insurability of some risks)

Risk pooling

- The use of machine learning for pricing should not lead to an extreme personalization of risk/premium
- *E.g. extremely high premiums for some risk profiles that imply no risk transfer.*
- The insurer has the social role of creating solidarity among the policyholders.

Best estimate price

- **Concepts**
  - **Non-protected variable** : discrimination based on these variable is permitted
  - **Protected variable** : discrimination based on these variables is not permitted
  - **Direct discrimination** : use of protected variables as a rating factor
  - **Indirect discrimination** : policyholders appear to be treated solely based on non-protected variables, but because of the correlation between protected and non-protected variables, model captures information on protected variables from non-protected variables.
- **Best-estimate price :** computed using the non-protected and protected variables

$$\mu(\boldsymbol{X_{NP}}, D) = E[Y|\ \boldsymbol{X_{NP}}, \mathrm{D}]$$

$\boldsymbol{X_{NP}}$ the non-protected variables, D the protected variables and Y the response variable

→ Direct discrimination

Unawareness price

- **Unawareness price :** computed using only the non-protected variables

$$\mu(\boldsymbol{X_{NP}}) = E[Y|\, \boldsymbol{X_{NP}}]$$

→ Indirect discrimination

- **Analytical unawareness price :** averaging the best-estimate prices with $P(D = d|\boldsymbol{X_{NP}})$

$$\mu(\boldsymbol{X_{NP}}) = E[Y|\, \boldsymbol{X_{NP}}] = \sum_{d} E[Y|\, \boldsymbol{X_{NP}}, \text{D=d}]\; P(D = d|\boldsymbol{X_{NP}})$$

→ Indirect discrimination

Non-discriminatory prices

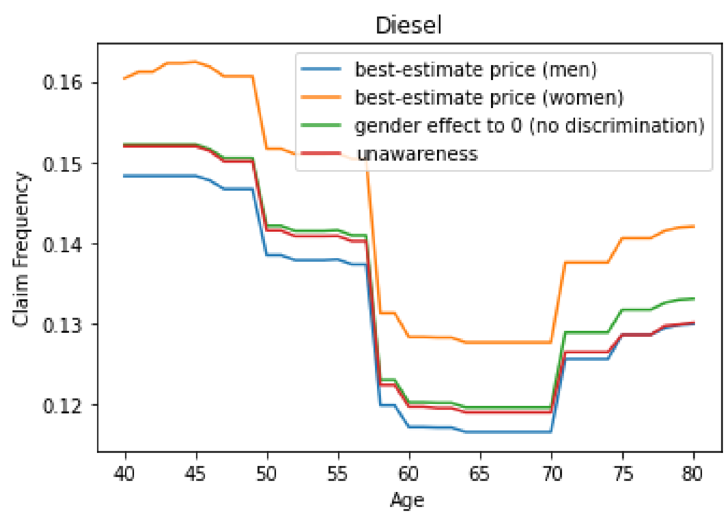
- **Discrimination-free price :** averaging the best-estimate prices with $P(D = d)$

$$h(\boldsymbol{X_{NP}}) = \sum_{d} E[Y|\, \boldsymbol{X_{NP}},\ \text{D=d}]\ P(D = d)$$

➡ No direct or indirect discrimination

- **Effect of the protected variable to 0 :** set the part of the score related to the protected variable to 0

➡ No direct or indirect discrimination



AGENDA

Some useful ML techniques

Applications to pricing and underwriting

Challenges with Machine Learning techniques

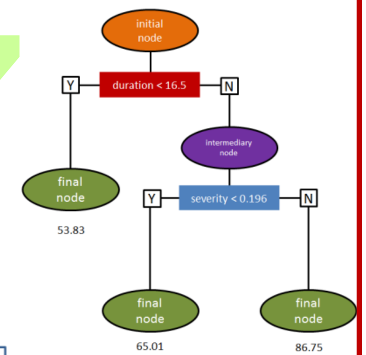
- Overfitting
- Discrimination and fairness
- Interpretability

SOME MACHINE LEARNING TECHNIQUES ARE BLACK BOXES AND INTERPRETATION OF THE RESULTS CAN BE QUITE DIFFICULT

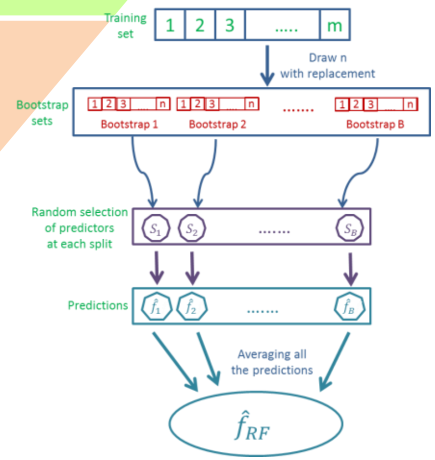
Understanding the results of ML techniques is not easy

Complexity

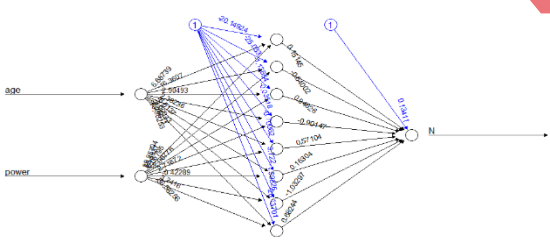
In the case of regression trees, understanding how the model predicts claims' cost or frequency values for new data points is not a problem, as it is very intuitive.



In the case of more complex methods such as Bagging and Random forests, even understanding how the model predicts values for new data points is rather difficult.



Things may be even worse for GBM and NN.



Interpretability

UNDERSTANDING THE RESULTS OF ML MODELS IS NEVERTHELESS KEY FOR SOUND BUSINESS DECISION-MAKING AS MANY STAKEHOLDERS USE THE RESULTS OF THE MODELS

Quant (Actuaries, data scientist,…)

- Able to understand the technical details
- Trust its outputs based on cross-validation, error measures and assesment plots

Other stakeholders

- Not necessarily « quantitative people »
- Should nevertheless understand and trust results to take decisions

Machine learning techniques usually improve predictive power but at the expense of a certain loss of interpretability ➔ Find trade-off between

- Predictive power
- Capacity to understand the results
- Ability to take sound decisions based on the results

High-end questions

Who will use the results? For what purpose? With which impact?

GLOBAL VS LOCAL INTERPRETABILITY OF ML TECHNIQUES

See previous IACA webinar on explainable ML

- Global Model Interpretability
  - How does the trained model make predictions?
    - Which features are important and what kind of interactions between them take place?
    - Global model interpretability helps to understand the distribution of your target outcome based on the features.
    - Global model interpretability is very difficult to achieve in practice → Any model that exceeds a handful of parameters or weights is difficult to understand
    - Some models are interpretable at a parameter level :
      - For linear models, the interpretable parts are the weights,
      - For trees interpretable parts are the splits (selected features plus cut-off points) and leaf node predictions.
  - Global Interpretable tools
    - Interpretable Models by nature (eg. Linear models, Regression Tree)
    - Feature Importance
    - Partial Dependant Plot (PDP), Individual Conditional Expectation (ICE) and Accumulated Local Effects (ALE)
    - Interaction Measures (H-statistic)
- Local Interpretability for a Single Prediction
  - Why did the model make a certain prediction for an instance?
    - If you look at an individual prediction, the behavior of the otherwise complex model might behave more pleasantly.
    - You can **zoom in on a single instance** and examine what the model predicts for this input and explain why.
      - Shapley Value
      - Breakdown

EXPLAINABLE BOOSTING MACHINE (EBM)

EBM is a special case of a GAM

$$g(E[y]) = \beta_0 + \sum f_j(x_j) + \sum f_{ij}(x_{ij})$$

- f_j is
  - a β coefficient if x_j is categorical
  - a function if x_j is continuous

- f_{ij} represents the interaction between x_i and x_j
  - Interactions automatically detected thanks to the FAST algorithm

- f_j and f_{ij} estimated thanks to **boosting and bagging** techniques

EXPLAINABLE BOOSTING MACHINE (EBM)

$$g(E[y]) = \beta_0 + \sum f_j(x_j) + \sum f_{ij}(x_{ij})$$

Algorithm with two explanatory variables

1. Fit a function F_1 with a tree using only $feature_1$
2. Compute $residual_1$ wrt F_1
3. Fit a function F_2 on $residual_1$ with a tree using only $feature_2$
4. Compute $residual_2$ wrt F_1 and F_2
5. Fit a function F_3 on $residual_2$ with a tree using only $feature_1$
6. …

- Run the algorithm to have n F_j for $feature_1$ and n F_j for $feature_2$
- Add them up to obtain f_1 for $feature_1$ and f_2 for $feature_2$
- We can add bagging : estimation of F_j with a forest instead of a tree

JOCO2024: REGISTRATION OPENS END FEBRUARY

- Stay tuned on https://www.joco2024.org/
- Currently selecting the speakers to finalize the program

CONTACT DETAILS

Xavier Maréchal
CEO – Managing Partner
M +32 497 48 98 48
xavier.marechal@reacfin.com

Reacfin
Place de l'Université 25
B-1348 Louvain-la-Neuve (Belgium)
T +32 (0) 10 68 86 07
www.reacfin.com

DISCLAIMER

The recipient of this document should treat all information as strictly confidential and only in the context stated below. Information may not be disclosed to any third party without the prior join-consent of Reacfin.

Estimates given in this presentation are based on our current knowledge, they can be based upon our previous experience within the Undertaking, as well as taking into account similar projects in the same context as the Undertaking, either locally, within majority of the EU countries as well as overseas.

This presentation is only the supporting document of a verbal presentation. Hence, it is not intended to be exhaustive. Quoting or using this document on its own might be misleading. As a result, these materials may not be used by anybody except their authors nor should they be relied upon in any way for any purpose other than as contemplated by joint written agreement with Reacfin.

Reacfin

Place de l'Université 25
B-1348 Louvain-la-Neuve
www.reacfin.com