Institute and Faculty of Actuaries

Why are ML Predictions so Leaky

Martin Cairns & Benjamin Gaby

15 April 2025

There's a perception our prices aren't fair…

UK car insurance premiums rise by 34% – and far higher than in rest of Europe

ABI reveals industry payouts for claims have risen at far lower rate, raising questions about value of policies

The Association of British Insurers (ABI) put the average rise for car cover in the UK at 34% between Q4 2022 and Q4 2023. Photograph: Piotr Adamowicz/Alamy

House of Commons Library

UK Parliament > House of Commons Library > Insight > The rising cost of UK car insurance

Insight

The rising cost of UK car insurance

Published Tuesday, 13 August, 2024

express.co.uk • 21/08/2024

Drivers urged to 'make note of this date' before they buy car insurance to save money

The company noted that a number of factors determine the overall cost of car insurance(Image: Getty)

Expert says one thing can 'sway' your car insurance price

Everyone needs car insurance, but costs have shot up over the past couple of years – however drivers can find a better deal in most circumstances

Drivers warned as these 10 job titles will make car insurance prices soar

Drivers have been warned that their car insurance premiums could increase over the next year if they work under any of these common job titles

Some drivers 'unable' to get car insurance due to new 'strict new rule'

Some car insurance providers have an age restriction in place meaning those above a certain age are unable to get cover, a finance expert has claimed while speaking about the issue

Insurance costs are soaring – so here's how my family cut them

Tips and tricks for getting the best deals when renewing home insurance and car cover

Average cost of UK car insurance rises by one-third in a year, analysis finds

ABI reports annual jump of £157 in first quarter of 2024 but says 1% increase on previous quarter indicates rises are easing

PRICES & INTEREST RATES • 7 MAR 2024 MARK FARRELL, JOHN NICHOLAS

Why have insurance premiums gone up so much?

'Martin Lewis tool helped me save £618 on my car insurance – anyone can do it'

The MSE newsletter said the rising costs of car insurance was the "final nail in the coffin of many budgets" with it rising by 70% since 2021

Wasn't Machine Learning supposed to make our prices better?

Institute and Faculty of Actuaries

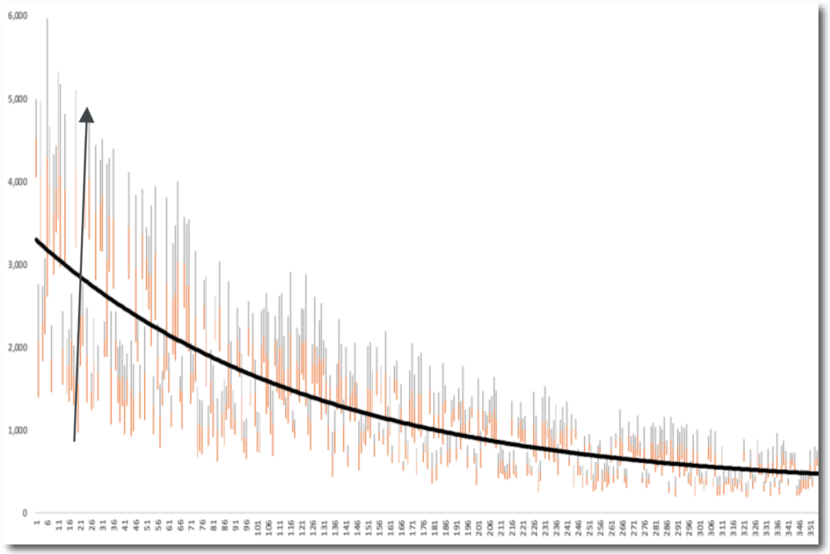
How we’ve got here

Where We Were

- GLMs
- Limited numbers of factors
- Good quality of fit
- Predictable response
- Lots of human understanding
- Multi-disciplined oversight and controls

Where We Are

- GBMs
- High numbers of factors
- Great fit/able to represent micro adjustments
- Lack of transparency & ability to predict
- Limited ability for businesses to interact with the modelling output
- How do we correct models without destroying their value?



Institute and Faculty of Actuaries

On sink holes and the winner's curse

The models have the ability to find **small pockets of business** which perform differentially, and set a more appropriate price for it

But what if there was limited exposure? And an unusually lucky performance?

You get a sink hole in your model!

You won't find it - but it's only a very small amount of the book so it doesn't matter?

We live in a price competitive market

So the group of potential policyholders **will** find your underpriced segment

So it will grow into an outsized part of your portfolio

And produce losses, becoming a big share of your loss experience

Are we just throwing sabots at the machine?

Institute and Faculty of Actuaries

How is Machine Learning used in practice?

Some firms continue to shy away from GBMs as not sufficiently understood and trusted, particularly for risk modelling

- *Pros*: stable modelling performance
- *Cons*: Falling behind market standards of predictive power

Some firms use a GBM as a test framework to hunt through features finding which ones add value. These are then brought into the main model (a GLM)

- *Pros*: Relatively efficient extensive search for additional features
- *Cons*: GBM ranking can be flawed where features are highly correlated. Only GLM predictive performance; falling behind market standards

Build initial GLM. Fit a GBM to the residuals. Import GBM score as an additional factor for the GLM

- *Pros*: GBM derives extra insights from the initial GLM noise
- *Cons*: Using the score in a GLM on the same data it was derived from imports noise; overfits leading to poor generalisation. Lacks interpretation

Build a GBM and call directly in the pricing stack (possibly as part of an ensemble)

- *Pros*: Maximum predictive performance
- *Cons*: Results not fully understood; poor performance in the market with high volatility in prices for similar risks between refreshes

There's no free lunch. If the GBM model isn't stable, importing the 'good bits' doesn't work

Why are current approaches failing?

Unique Features of Insurance Data

Applying ML to insurance pricing presents unique challenges:

- Sparse, skewed & noisy responses and high dimensionality means data is "**insufficient**"
- "**Biased**" training data (consequence of footprint & prices) with highly correlated features; need for extrapolation is inevitable
- **Winners' curse** means models need to perform well on a different mix
- Cost of error is high
- Observing error takes **time**

Problems with Current Practice

Feature Engineering

- Inadequate cleaning & unnecessarily complex feature spaces make extracting the underlying "signals" difficult

Model Fitting

- Insufficient control of ML models leads to overfitting & capture of excessive noise
- No direct injection of domain knowledge

Implementation

- Lack of transparency and instability prevent direct use of ML models
- Inability to monitor mix changes and understand the impact of rate changes

Underperformance

The bottom line:

Anticipated predictive lift is not realised and significant variations in segmental prices appear between analyses

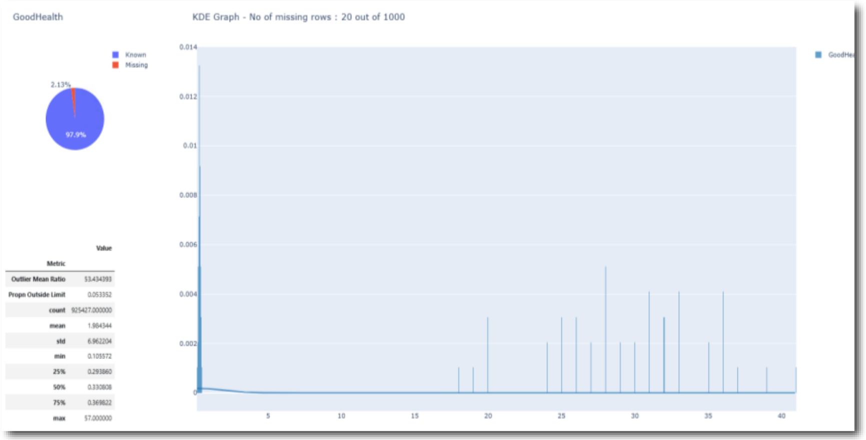
Institute and Faculty of Actuaries

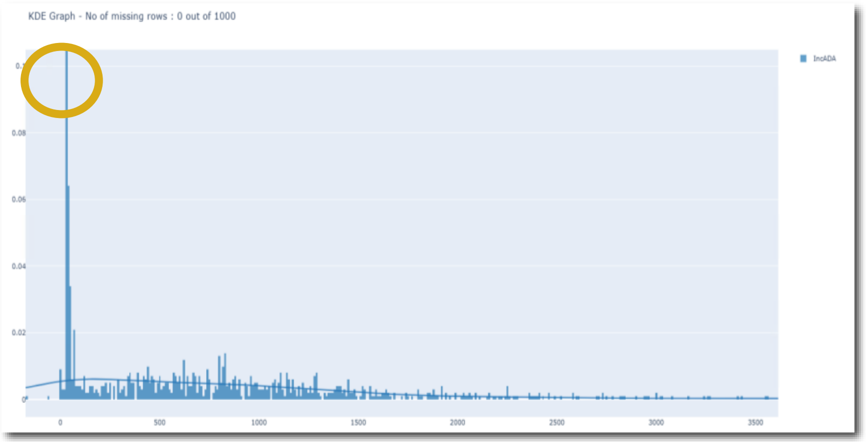
Insurance data – “Dirty Data”

Insurance companies often have multiple complex databases and extraction processes, resulting in “Dirty Data”

The top chart shows that the Census factor “Good Health” has extreme values indicating <1% of the UK are in good health. Using such a factor will lead to spurious results.

The bottom chart identifies a spike of fee-only claims which have not been removed from the AD incurred, resulting in the response we are modelling being incorrect





Conclusion: Data must be investigated before being used

Institute and Faculty of Actuaries

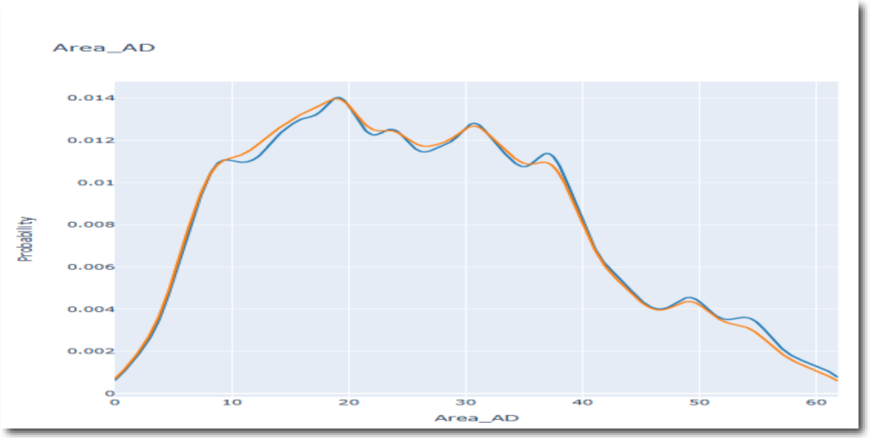
Insurance data – missing data

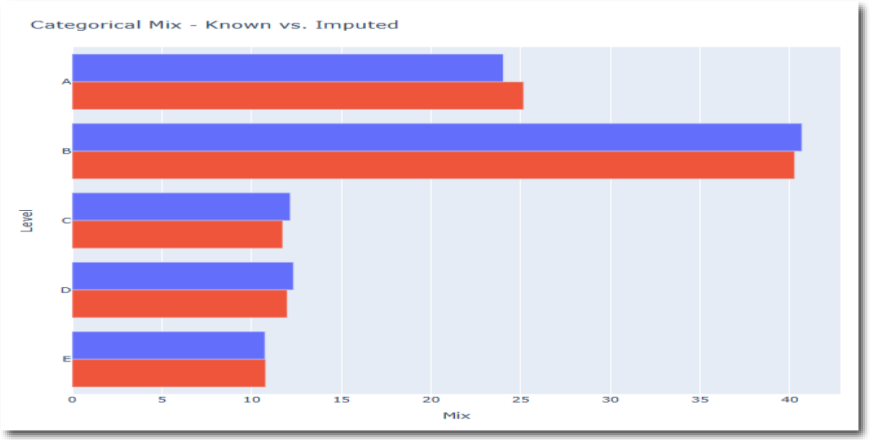
Allowing missing data to enter the ML model without understanding it can lead to:

- A spurious model
- Unwanted model bias

It is also not appropriate to use the traditional GLM approach of “Default Values”. The ML model just blindly free fits to whatever the data says.

We expect the factor distributions of the unknown data to follow similar distributions of the known data. Ideally, we should use some imputing algorithm which reflects this.





Conclusion: Methods aren’t clever enough to deal with missing data, so we need to be cleverer in ‘healing’ it before modelling

Institute and Faculty of Actuaries

Insurance data – categorical factors

Blindly one-hot encoding our categorical factors is very dangerous

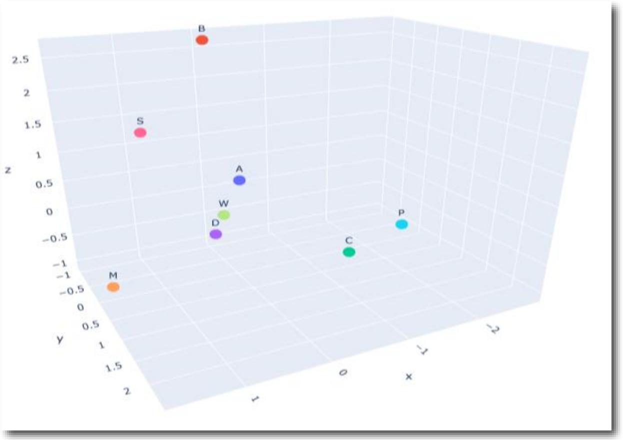
- Sparse and noisy levels of the factor are free-fitted exactly to the data
- Native methods in GBMs will also fit to noise

In the GLM world, we would fit variates to categorical factors

- Natural ordering
- Group similar levels together using custom factors

Before building an ML model we must **transform** our categorical factors

Be careful of hidden model bias: we need to group Civil Partnership with Married before we transform. A visualisation can help!

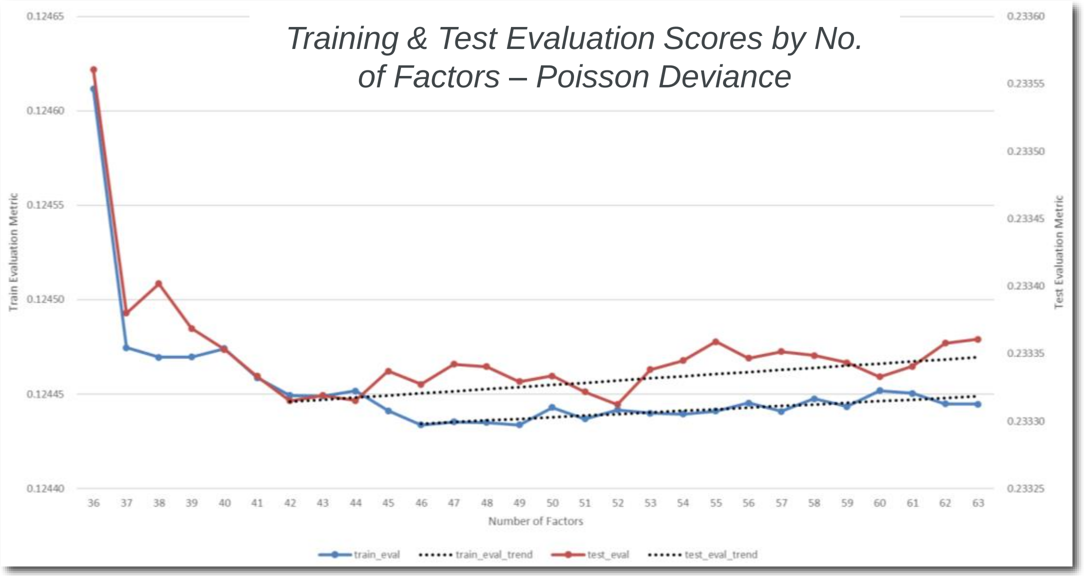


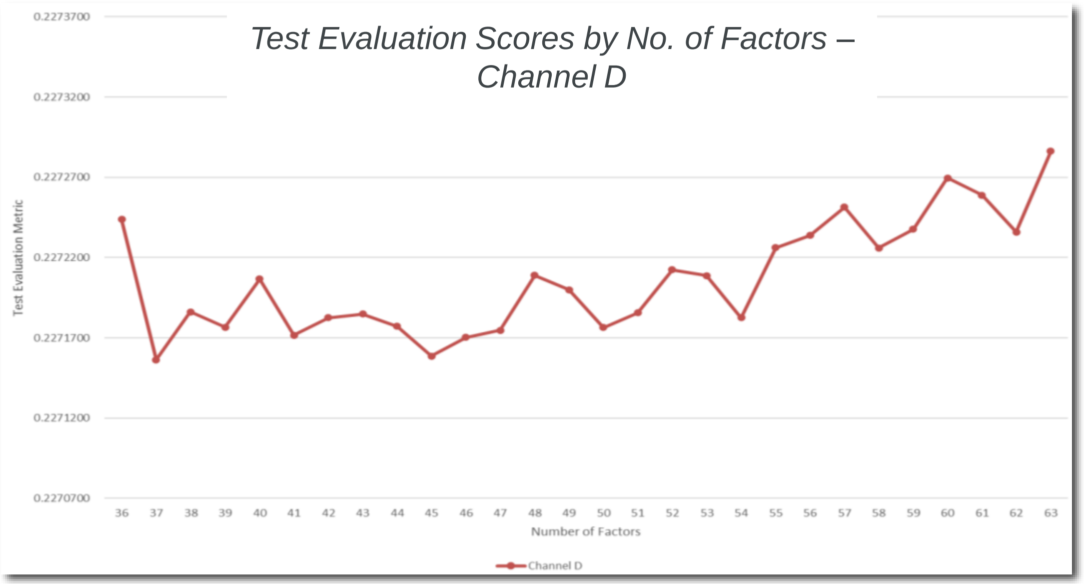
	Marital Status Level	Exposure Proportion	0	1	2
A	Separated	1%	-0.424	-0.817	0.168
B	Civil Partnership	0.01%	0.003	-0.771	2.668
C	Living with Partner/ Common Law	16%	-0.223	2.382	0.457
D	Divorced	3%	-0.108	-1.025	-0.938
M	Married	42%	1.706	0.263	-0.741
P	Partnered	1%	-2.812	-0.254	-1.044
S	Single	36%	0.895	-0.751	1.294
W	Widowed	1%	-0.167	-0.818	-0.457

Conclusion: ML models need to be told that different levels are ‘close’

Institute and Faculty of Actuaries

The danger of “over-featuring”





- Adding extra features is not improving the performance of the model on training or test data on the overall mix
- There is a slight deterioration in performance which is greater for test data

- For one particular channel (which has a different risk mix with reasonable exposure) there is significant deterioration as the factor count increases

Institute and Faculty of Actuaries

The danger of “over-featuring” – what went wrong?

Due to the correlation in insurance data, one needs to sensibly pick a subset of factors that explains the data with the least number of factors. Use the following criteria:

1. Their absolute strength
2. Stability i.e. the factors are time- and random factor- consistent
3. Minimise correlations with any factors already selected
4. Using domain knowledge, as you may expect there to be lots of strong interactions with a weaker factor

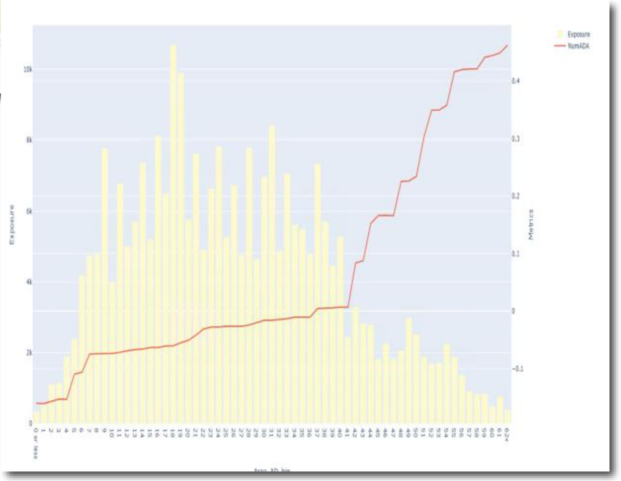
Our research below, performed on real client data, proves models can achieve higher and much more stable predictive performance on $\sim\frac{1}{3}$ of the factors in use today

Conclusion: Too many features = more noise. Factors should earn their way in.

Institute and Faculty of Actuaries

Understanding what your model is giving you

- Industry practice is to validate a GBM with Partial Dependency Plots (“PDPs”). **This is very dangerous.** PDPs are based on averaging scenarios varying only one factor.
  - However, this creates unreasonable combinations (e.g. a 21-year-old who has held a license for 30 years) and includes these in the average.
  - When you have a model which is already over-reacting to noise and interactions across correlated variables, these averages can become uninformative.
- It is better to perform a Shap analysis, which:
  - Retains meaningful combinations of factors, and reliable outputs
  - But can still struggle with correlated factors



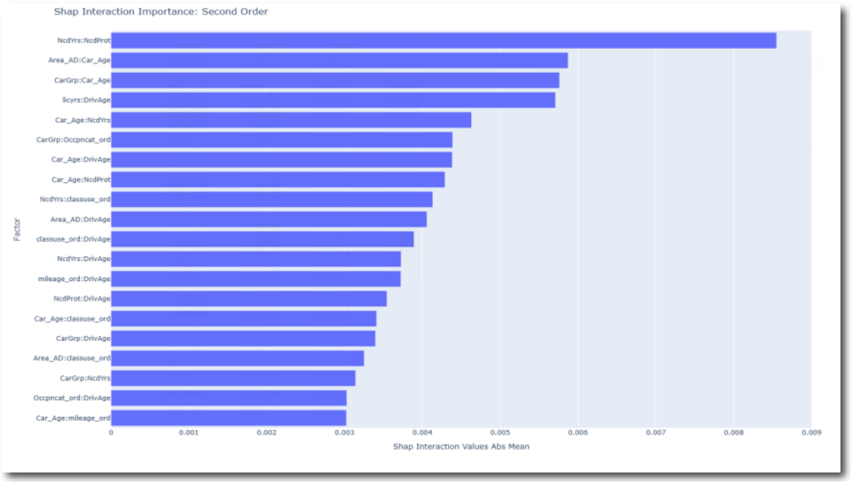
- The implied trends for the one-way factors are not smooth ⇒ an unsmooth customer journey of prices
- The trends potentially fit to the extremes. This may give unfair prices to vulnerable customers

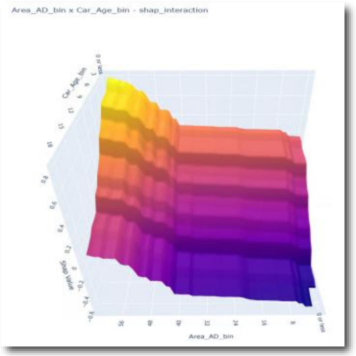
Institute and Faculty of Actuaries

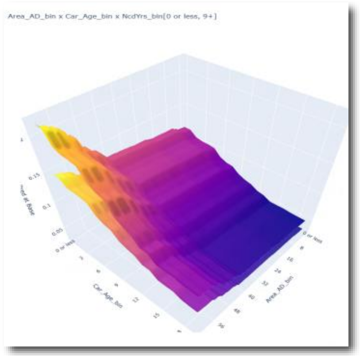
Typical issues seen with interactions in GBMs

The outputs above from the Shap-Interaction Analysis present the following problems:

1. We will not want to implement all interactions
2. Some of them will not make sense when they are validated on surfaces
3. A lot of the weaker interactions further down the list are just fitting to noise
4. The surfaces are very noisy leading an un-smooth customer journey
5. We (still) cannot validate on surfaces any interaction of higher order than three







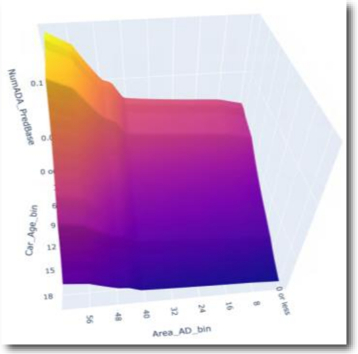
Knowing our model has an interaction we don't agree with is half the battle – but what can we do about it?

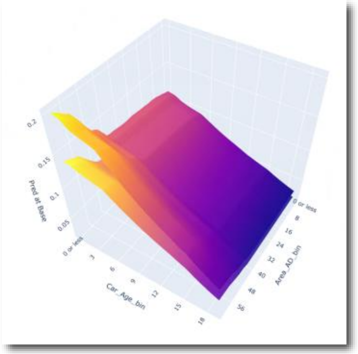
Institute and Faculty of Actuaries

Smoothing of predictions to avoid cliff edges

Our research shows the following if we linearise the GBM to a (automatically backfitted) GLM we get:

1. Retention of over 80% of the predictive uplift from the GBM over a traditional best-practice hand-crafted GLM
2. To choose what interactions are fitted in the GLM backfit, avoiding implementing counter-intuitive and higher than three-way interactions
3. A model that is less sensitive to mix change and generalises better
4. A model that is more transparent, understood and can be challenged by all stakeholders





Is this the way to validate GBMs?
Is this the (controllable, governed) model which should be deployed?

Institute and Faculty of Actuaries

Final thoughts for you

Are we being professional in allowing models to be deployed that the business does not understand?

Are actuaries (and the business) being side-lined from the modelling process and not able to inject domain knowledge? How does business and marketing strategies align with prices that no-one understands?

Has actuarial and business governance been lost as we move into a ML world? Is a compliance and/or Consumer Duty disaster inevitable if the process does not change?

Institute and Faculty of Actuaries