Responsible actuarial learning

Katrien Antonio
LRisk - KU Leuven and ASE - University of Amsterdam

September 25, 2024

1. The expansion of the actuary's toolbox: from statistical and machine learning to **actuarial learning**.

2. On **smart feature engineering and feature selection** for actuarial work, and how data science methods can help.

3. **Responsible actuarial learning**: high-stakes decisions, interpretability tools, (local and global) surrogate models.

4. Don't overdo - the right tool, in the right place, at the right time.

Some (recent) initiatives:

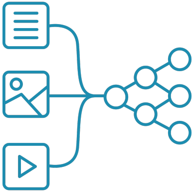
- working group on data science with the Swiss Association of Actuaries; see www.actuarialdatascience.org
- Actuarial Association of EU (AAE) on **CPD in Data Science**, Sept. 19 2024
- (since October 2021) the IA|BE Data Science Certificate in Belgium.

Picture taken from the strategy paper by the data science working group of the Swiss Association of Actuaries, August 2018.

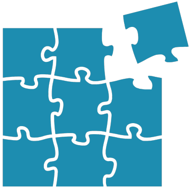
Actuarial learning

An opinionated guide to actuarial learning

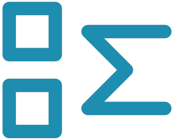
Target variables:
time-to-event, (low)
frequency, (high impact) severity



Multi-type input features:
continuous, spatial,
(high cardinality) factor,
compositional data



(In)complete:
due to reporting or settlement delay
policy modifications



Fine-grained or aggregate

- Predictive models for frequency, severity, churn or time-to-event data, with a (large) set of risk factors as input.

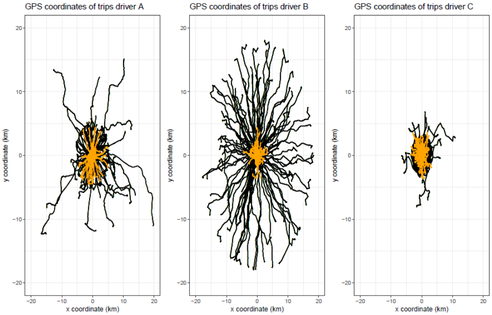
- Do mind the (actuarial) **distributional assumptions**:

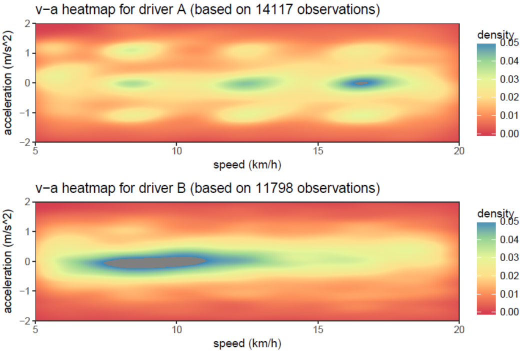
	Distribution	Prediction $f(\boldsymbol{x})$	Loss function $D(y, f(\boldsymbol{x}))$
Claim frequency	$N \sim$ Poisson	$\mathbb{E}(N \mid \boldsymbol{x}, e)$	$\frac{2}{n}\sum_{i=1}^{n}\left[y_i \ln\left\{\frac{y_i}{f^f(\boldsymbol{x}_i)}\right\} - \{y_i - f^f(\boldsymbol{x}_i)\}\right]$
Claim severity	$L/N \sim$ gamma	$\mathbb{E}(L/N \mid \boldsymbol{x})$	$\frac{2}{\sum_i N_i}\sum_{i=1}^{n} N_i\left[\frac{y_i - f^s(\boldsymbol{x}_i)}{f^s(\boldsymbol{x}_i)} - \ln\left\{\frac{y_i}{f^s(\boldsymbol{x}_i)}\right\}\right]$
Customer churn	$C \sim$ Bernoulli	$\mathbb{E}(C \mid \boldsymbol{x})$	$-\frac{1}{n}\sum_{i=1}^{n}\left[y_i \ln\{f^c(\boldsymbol{x}_i)\} + (1-y_i)\ln\{f^c(\boldsymbol{x}_i)\}\right]$

- **Advantages of ML**: smart engineering of features and selection.

Smart engineering of features

Example in telematics - IoT data collection leads to new or unexplored types of features





Pictures reproducing Wüthrich (2017, European Actuarial Journal) on Covariate selection from telematics car driving data , using data from AXA Kaggle competition.

Smart engineering of features

Example in telematics - IoT data collection leads to new or unexplored types of features

James B.

He drives **100 000 km** during one year.

His **road type composition** is $(15\ 000, 15\ 000, 50\ 000, 20\ 000)$ or $(0.15, 0.15, 0.5, 0.2)$.

Eugène from Man Bijt Hond

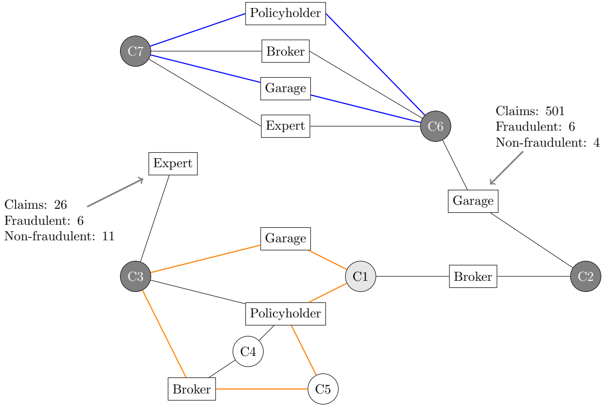
He drives **1 000 km** during one year.

His **road type composition** is $(500, 300, 200, 0)$ or $(0.5, 0.3, 0.2, 0)$.

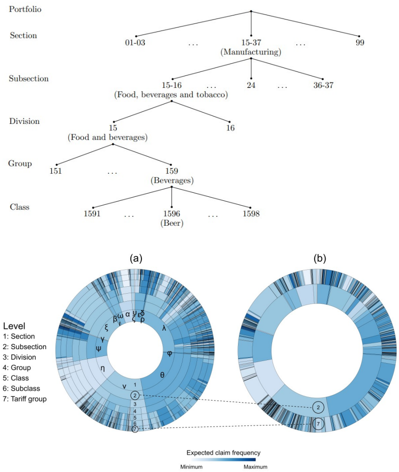
Verbelen, Antonio & Claeskens (2018, JRSS: Series C). Unravelling the predictive power of telematics data in car insurance pricing.

Smart engineering of features

Example in insurance fraud detection



Picture from Óskarsdóttir, Antonio et al. (2022, Risk Analysis). Social network analytics for supervised fraud detection in insurance.



Picture from Campo & Antonio (2024, Annals of Actuarial Science).

On clustering levels of a hierarchical risk factor.

arXiv:2403.03613v2 [stat.ME] 19 Aug 2024

Reducing the dimensionality and granularity in hierarchical categorical variables

Paul Wilsens[1,3,*], Katrien Antonio[1,2,3,4], and Gerda Claeskens[1,4,5]

[1] *Faculty of Economics and Business, KU Leuven, Belgium.*
[2] *Faculty of Economics and Business, University of Amsterdam, The Netherlands.*
[3] *Leuven Research Center on Insurance and Financial Risk Analysis, KU Leuven, Belgium.*
[4] *Leuven Statistics Research Center, KU Leuven, Belgium.*
[5] *Research Centre for Operations Research and Statistics, KU Leuven, Belgium.*
[*] *Corresponding author:* paul.wilsens@kuleuven.be

Abstract

Hierarchical categorical variables often exhibit many levels (high granularity) and many classes within each level (high dimensionality). This may cause overfitting and estimation issues when including such covariates in a predictive model. In current literature, a hierarchical covariate is often incorporated via nested random effects. However, this does not facilitate the assumption of classes having the same effect on the response variable. In this paper, we propose a methodology to obtain a reduced representation of a hierarchical categorical variable. We show how entity embedding can be applied in a hierarchical setting. Subsequently, we propose a top-down clustering algorithm which leverages the information encoded in the embeddings to reduce both the within-level dimensionality as well as the overall granularity of the hierarchical categorical variable. In simulation experiments, we show that our methodology can effectively approximate the true underlying structure of a hierarchical covariate in terms of the effect on a response variable, and find that incorporating the reduced hierarchy improves the balance between model fit and complexity. We apply our methodology to a real dataset and find that the reduced hierarchy is an improvement over the original hierarchical structure and reduced structures proposed in the literature.

MSC classification: 62H30, 68T07

Keywords: hierarchical categorical variable, entity embedding, clustering, predictive modelling, machine learning

Statements and declarations

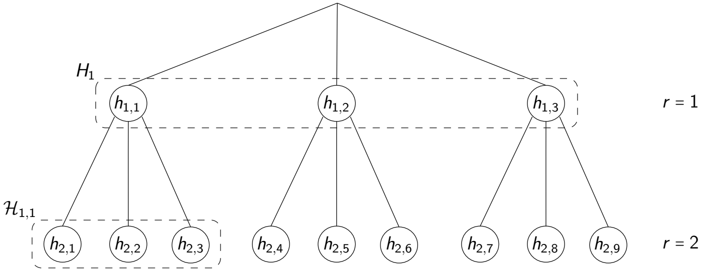
Data and code availability statement: Data and code are available on https://github.com/PaulWilsens/reducing-hierarchical-cat.

Funding statement and acknowledgements: The authors gratefully acknowledge funding from the FWO and Fonds De La Recherche Scientifique - FNRS (F.R.S.-FNRS) under the Excellence of Science (EOS) program, project ASTeRISK Research Foundation Flanders [grant number 40007517]. Katrien Antonio gratefully acknowledges support from the Chaire DIALog sponsored by CNP Assurances and the FWO network W001021N. We thank the editor and referees for their comments which helped improve the paper substantially.

Conflict of interest disclosure: The authors declare no conflict of interest.

Hierarchical categorical variable $\boldsymbol{h} = (h_1, \ldots, h_R)$ with R levels, where $h_r \in H_r = \{h_{r,1}, \ldots, h_{r,n_r}\}$ is a (non-hierarchical) categorical variable.

Aim to learn $\widetilde{\boldsymbol{h}} = (\widetilde{h}_1, \ldots, \widetilde{h}_{\widetilde{R}})$ with $\widetilde{R} \leq R$ levels, where we have that $\widetilde{n}_r \leq n_r \; \forall \; r = 1, \ldots, \widetilde{R}$.



Smart engineering of features

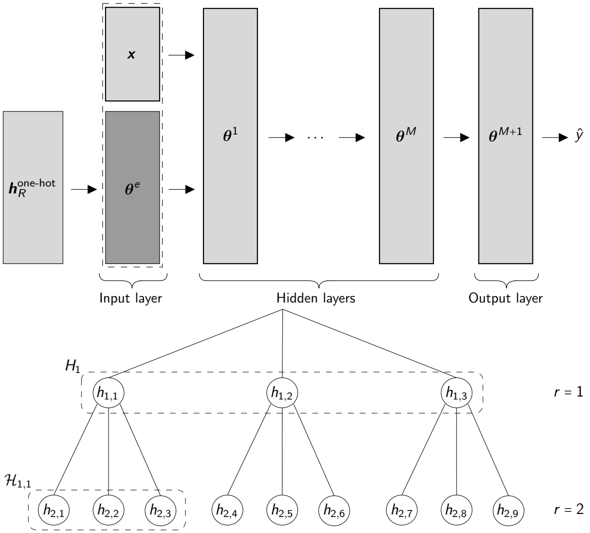
Via entity embedding

To embed the hierarchy for

1. $r = R$: we learn a feedforward neural network and apply **entity embedding** to h_R.

2. $r = 1..,R-1$: we **average** the embeddings over the hierarchical structure:

$$\boldsymbol{e}_{r,s} = \frac{1}{\dim(\mathcal{H}_{r,s})} \sum_{l|h_{r+1,l} \in \mathcal{H}_{r,s}} \boldsymbol{e}_{r+1,l}$$



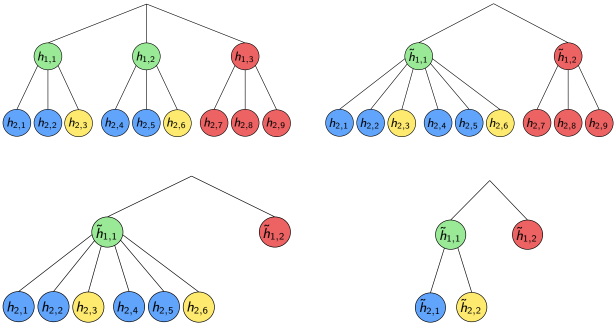
Smart engineering of features

Via entity embedding and clustering

We propose a **top-down** clustering algorithm that for a given level r

1. **merges** similar classes within level r, and
2. **collapses** descendant classes on level $r+1$ that are sufficiently close in the embedding space with their parent class on level r.

Both steps are **repeated** for every level in the hierarchy, starting from $r = 1$.



Smart selection of features

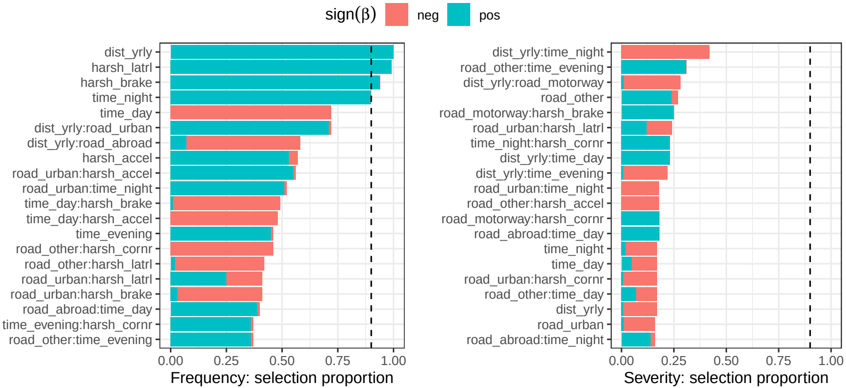
Example in telematics, from Henckaerts & Antonio (IME, 2022)

	Claim frequency		Claim severity		Customer churn	
Rank	Feature	%	Feature	%	Feature	%
1	geo_postcode	34.72	veh_weight	23.21	paym_split	43.48
2	driv_experience	14.08	veh_make	21.37	geo_postcode	11.67
3	driv_seniority	8.52	geo_postcode	10.54	veh_age	9.85
4	veh_make	6.25	veh_segment	10.48	paym_sepa	9.44
5	geo_mosaic	5.85	geo_mosaic	6.59	driv_seniority	6.90
6	veh_fuel	5.09	driv_seniority	5.83	veh_make	3.43
7	veh_segment	4.66	veh_value	3.50	driv_experience	2.85
8	paym_split	3.91	veh_age	3.44	geo_mosaic	2.45
9	driv_add_younger26	3.29	driv_experience	2.98	driv_age	2.43
10	driv_age	2.75	driv_add_younger26	2.91	veh_use	1.99
$\sum$		89.12		90.86		94.48

Henckaerts & Antonio (2022, IME). The added value of dynamically updating motor insurance prices with telematics collected driving behavior data.

Smart selection of features

Example in telematics, from Henckaerts & Antonio (IME, 2022)



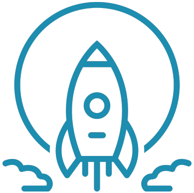
Equip **GLMs with a regularization** penalty to automatically select **most relevant telematics features**.

Responsible actuarial learning

An opinionated guide to (research on) responsible actuarial learning

High-stakes decisions
in a highly regulated industry

Explainable and transparent



Reproducible with
impact on other domains

Look under the hood of ML methods

Two roads or pathways to explainable AI (XAI):

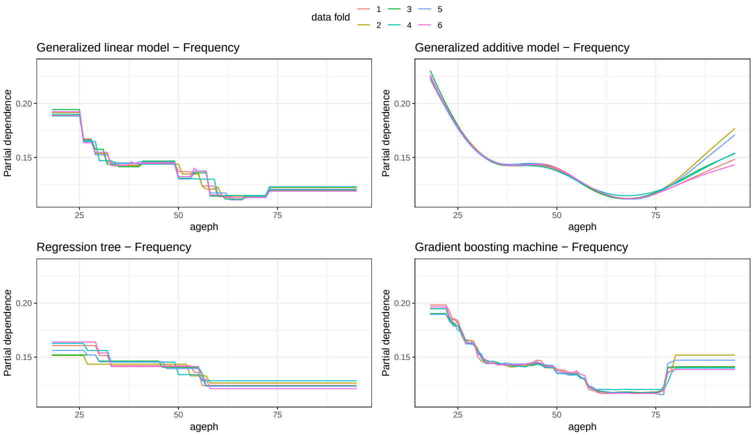
- **after the event**: use interpretation tools to (better) understand decision process in black box model

  Henckaerts, Côté, Antonio & Verbelen (2021, North American Actuarial Journal). Boosting insights in insurance tariff plans with tree-based machine learning methods.

  Holvoet, Antonio & Henckaerts (2024, on arxiv). Neural networks for insurance pricing with frequency and severity data: a benchmark study from data preprocessing to technical tariff.

- **by design**: develop and use transparent white box model, e.g. as a global surrogate for a more complex black box model

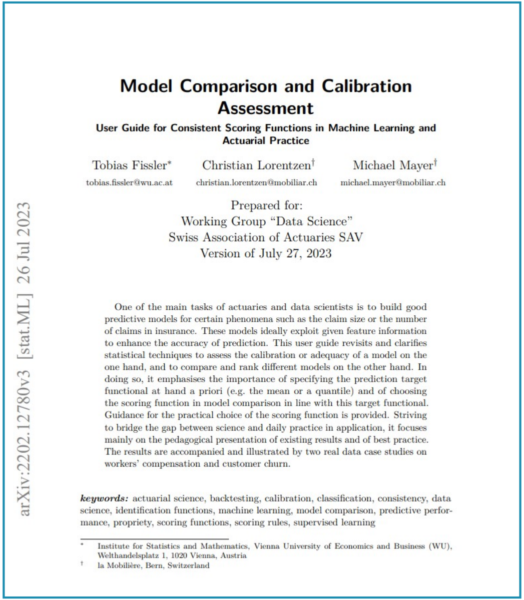
  Henckaerts, Antonio & Côté (2022, Expert Systems with Applications). When stakes are high: balancing accuracy and transparency with Model-Agnostic Interpretable Data-driven suRRogates.



Henckaerts, Côté, Antonio & Verbelen (2021, North American Actuarial Journal). Boosting insights in insurance tariff plans with tree-based machine learning methods. Society of Actuaries, annual prize for the best paper published in 2021.

Tools used in our paper on **neural networks benchmark study** (Holvoet et al., 2024):

- out-of-sample deviance and Diebold-Mariano test for predictive accuracy
- examine structure of predictions ~ well calibrated models?
- variable importance, partial dependence effects, Shapley values
- (ordered) Lorenz curves, Gini indices to compare technical tariffs.



arXiv:2202.12780v3 [stat.ML] 26 Jul 2023

Model Comparison and Calibration Assessment

User Guide for Consistent Scoring Functions in Machine Learning and Actuarial Practice

Tobias Fissler* tobias.fissler@wu.ac.at
Christian Lorentzen† christian.lorentzen@mobiliar.ch
Michael Mayer† michael.mayer@mobiliar.ch

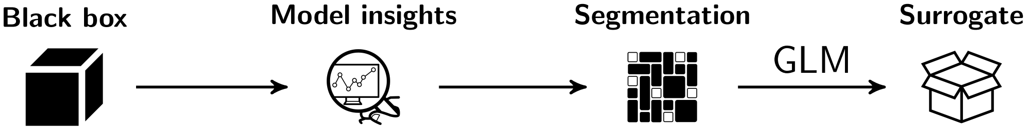
Prepared for:
Working Group "Data Science"
Swiss Association of Actuaries SAV
Version of July 27, 2023

One of the main tasks of actuaries and data scientists is to build good predictive models for certain phenomena such as the claim size or the number of claims in insurance. These models ideally exploit given feature information to enhance the accuracy of prediction. This user guide revisits and clarifies statistical techniques to assess the calibration or adequacy of a model on the one hand, and to compare and rank different models on the other hand. In doing so, it emphasises the importance of specifying the prediction target functional at hand a priori (e.g. the mean or a quantile) and of choosing the scoring function in model comparison in line with this target functional. Guidance for the practical choice of the scoring function is provided. Striving to bridge the gap between science and daily practice in application, it focuses mainly on the pedagogical presentation of existing results and of best practice. The results are accompanied and illustrated by two real data case studies on workers' compensation and customer churn.

keywords: actuarial science, backtesting, calibration, classification, consistency, data science, identification functions, machine learning, model comparison, predictive performance, propriety, scoring functions, scoring rules, supervised learning

* Institute for Statistics and Mathematics, Vienna University of Economics and Business (WU), Welthandelsplatz 1, 1020 Vienna, Austria
† la Mobilière, Bern, Switzerland

GLM as a global surrogate for a black box model



End product is a GLM, which features are **(1) smartly selected** and **(2) smartly engineered**, by looking under the hood of a more sophisticated black box model.

Picture from Henckaerts, Antonio & Côté (2022, Expert Systems with Applications). When stakes are high: balancing accuracy and transparency with Model-Agnostic Interpretable Data-driven suRRogates.

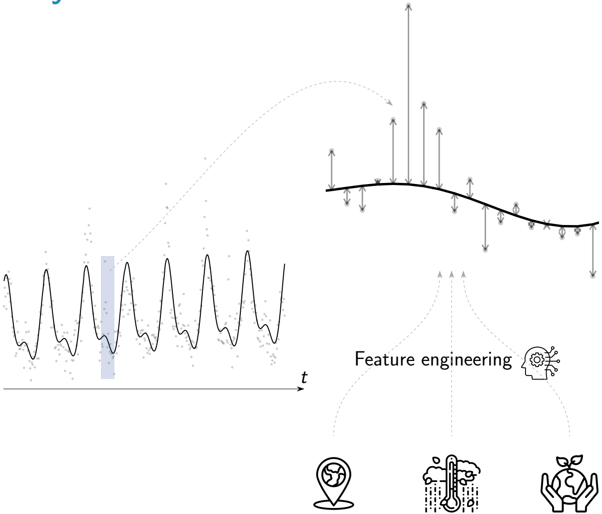
Learning from fine-grained open mortality data

Robben, Antonio & Kleinow (2024) on short-term association between **environmental variables** and mortality.

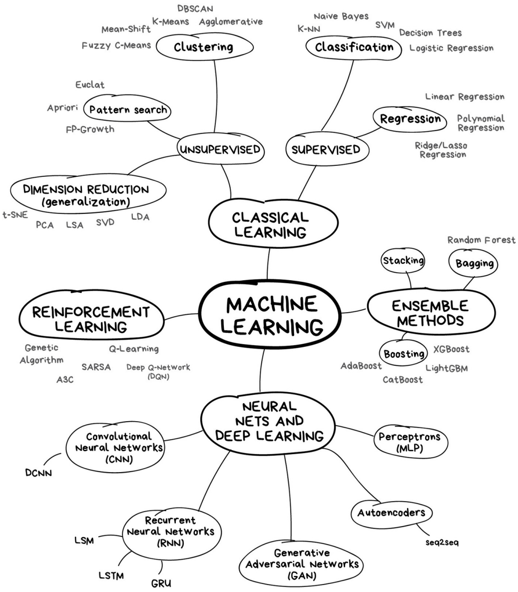
Approach:

1. a weekly, region-specific **baseline** mortality model to capture overall **seasonal** trends.

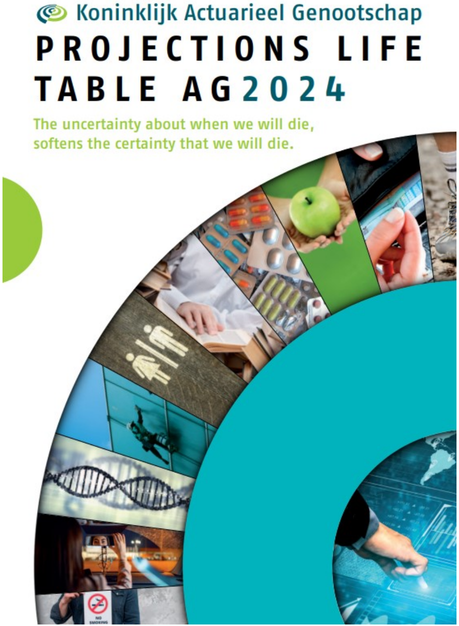
2. a machine learning model to analyze **mortality deviations** from the baseline model using region-specific environmental factors.



The right tool, in the right place, at the right time.



Picture taken from https://vas3k.com/blog/machine_learning/.



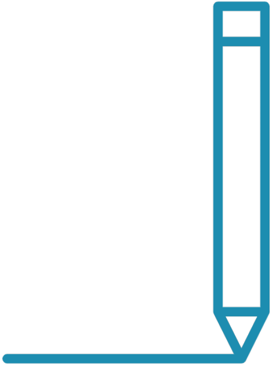
> The narrative must be that **actuaries are entering the data science world** not entirely to compete but also to bring the element of the **actuarial profession** where we build **integrity and transparency** into any work that we do, **and how documentation of that is possible.**

Quote from What data science means for the future of the actuarial profession, British Actuarial Journal, June 2018.

Responsible actuarial learning

Learning responsible actuaries

Responsible actuaries learn



Thanks!

For references, acknowledgements, and more see https://katrienantonio.github.io.