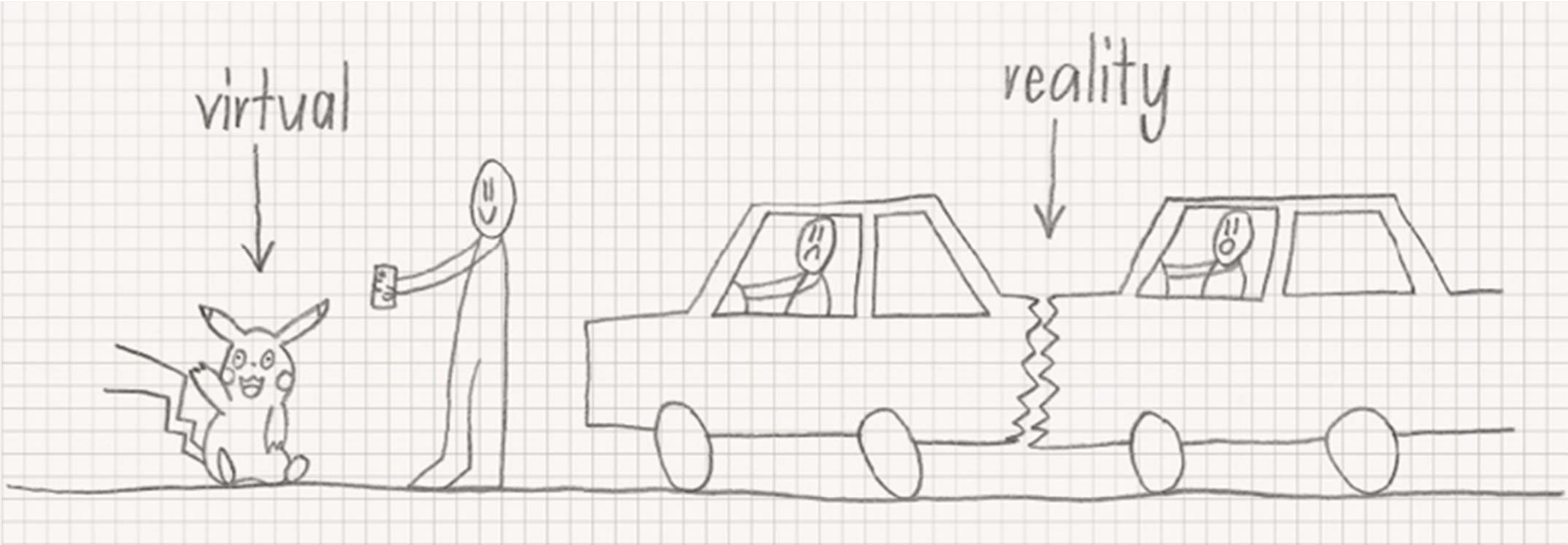


Non-Life Insurance Pricing goes XAI

Michael Mayer, October 6, 2021

Outline

1. Non-Life Insurance Pricing
2. Illustration with Real Data
3. The Role of ML and XAI

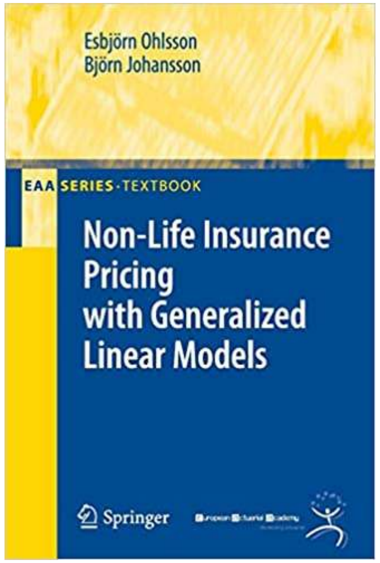
1. Non-Life Insurance Pricing

What is Non-Life Insurance Pricing?

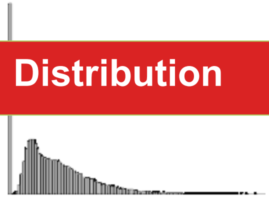
"Non-life insurance pricing is the art of setting the price of an insurance policy, taking into consideration various properties of the insured object and the policy holder.
The main source on which to base the decision is the insurance company's own historical data on policies and claims [...].
In a *tariff analysis*, the actuary uses this data to find a model which describes how the claim cost of an insurance policy depends on a number of explanatory variables."

Text color added retrospectively

- Preface, Ohlsson and Johansson (2010)

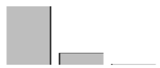
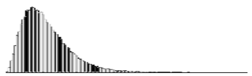


What does it mean from Statistical Perspective?

Response	Meaning	Distribution	Classic Approach
Claim cost	Losses ($$$) / Exposure		???

!

$$\boxed{Claim\ Cost = \frac{Losses}{Exposure} = \underbrace{\frac{Claim\ Count}{Exposure}}_{\text{Frequency}} \cdot \underbrace{\frac{Losses}{Claim\ Count}}_{\text{Severity}}}$$

Response	Meaning	Distribution	Classic Approach
Frequency	Claim count / Exposure		Poisson GLM with log-link
Severity	Losses / Claim count		Gamma GLM with log-link

Challenges?

2. Illustration with Real Data

Car Collisions with Large Animals

Annual figures for Switzerland

- Many large animals die on Swiss roads, among them ~8'000 deers.
- Total vehicle damage: ~25 Mio CHF
- Covered by partial damage coverage of motor insurance.

Source: https://de.wikipedia.org/wiki/Wildunfall

Annual figures for Swiss Mobiliar

- Ca. 2'000 animal collision claims
- Claim frequency around 0.3%-0.5%

Model for **claim frequency**, taking into account individual **risk factors**?

A simple GLM for Claim Frequency

Data

- Car policies with partial/full coverage
- Data over multiple years
- Millions of data rows
- Train/test split grouped by policy

Features

- **Driver:** place of living, age, gender, …
- **Car:** price, age, weight-to-power ratio, leased, …
- **Policy:** bonus protection, fully or partially comprehensive, year, …

Model*

- Poisson-GLM with log-link
  $$\log E(y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m$$
- Estimates of $\boldsymbol{\beta}$ minimize deviance
  $$2\sum w_i (y_i \log(y_i \,/\, \hat{y}_i) - (y_i - \hat{y}_i))$$
- Some estimates*:
  - Driver_age: -0.03
  - Town: -0.80
  - max_7000_km: -0.66

*Toy model for illustration only

3. The Role of ML and XAI

General View: Insurance Pricing Today and in Future

Today → Future?

Classic modelling

Manually build GLM

Difficult to select features, interactions, non-linearities
→ slow model building

Trade-off
Use insights from modern ML & XAI to build GLM

Efficient model building

Modern ML & XAI?

- Automatic model building
- Scales (more models, …)

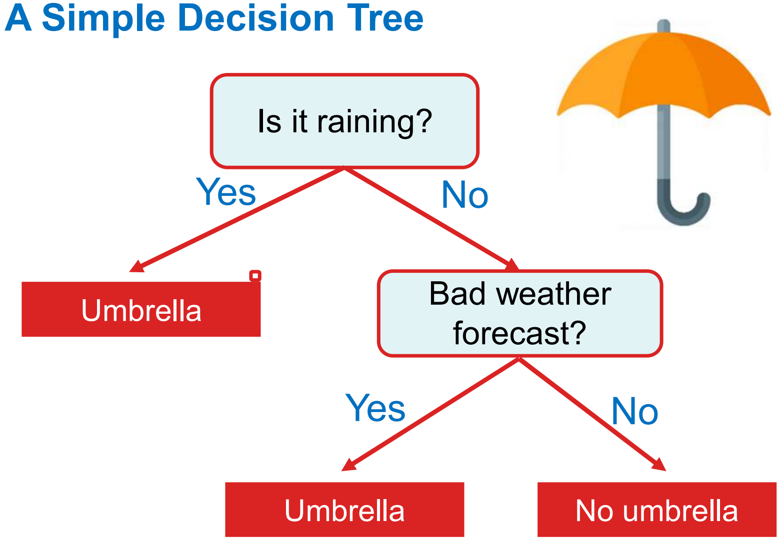
Advantages of GLMs?

1. Ease of interpretation and info gain
2. Transparency
3. Industry standard
4. Well-supported by standard software

Illustrated for toy model with the help of **Gradient Boosting** and **SHAP**

What is Gradient Boosting?

A Simple Decision Tree



Questions can be derived from data

Gradient Boosting

1. Fit small decision tree
2. Correct errors by another small tree
3. Repeat Step 2 multiple times

Implementations

- XGBoost
- **LightGBM**
- CatBoost

Click for more comparisons…

Aspect	XGBoost	LightGBM	CatBoost
Speed: CPU		❤	
Speed: GPU	❤		❤
Standard losses	✓	✓	✓
Special loss: Poisson/Gamma/Tweedie	✓	✓	✓
Special loss: Survival	✓		
Special loss: Robust		✓	✓
Special loss: Quantile		✓	✓
Tree size regularization	✓	✓	✓
Categorical input handling		❤	❤
Constraints: monotonic	✓	✓	✓
Constraints: interaction	✓	✓	

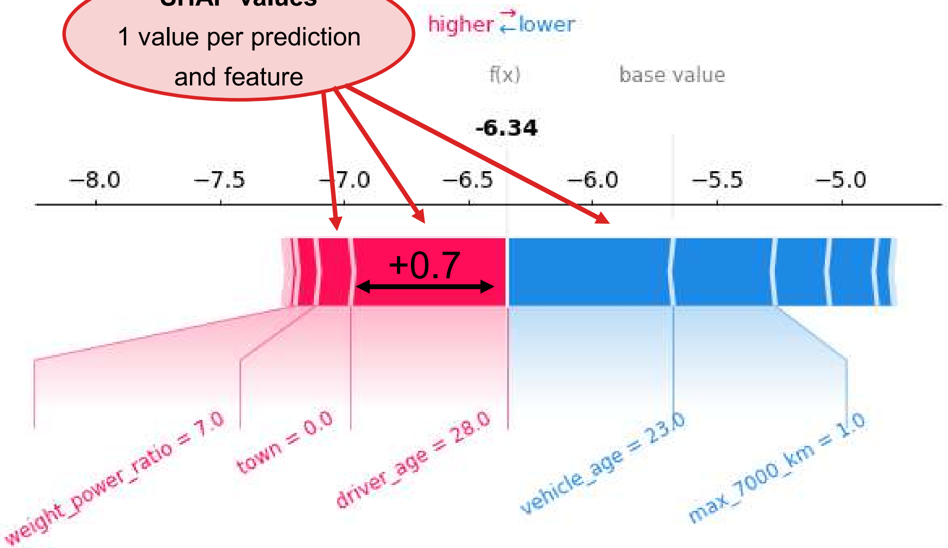
What is SHAP?

Repeat for many predictions

Fair, additive decomposition of single prediction

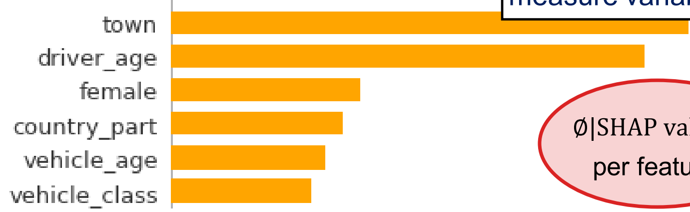
- Game-theoretic approach based on Shapley values
- Lundberg & Lee (2017)

SHAP values
1 value per prediction and feature



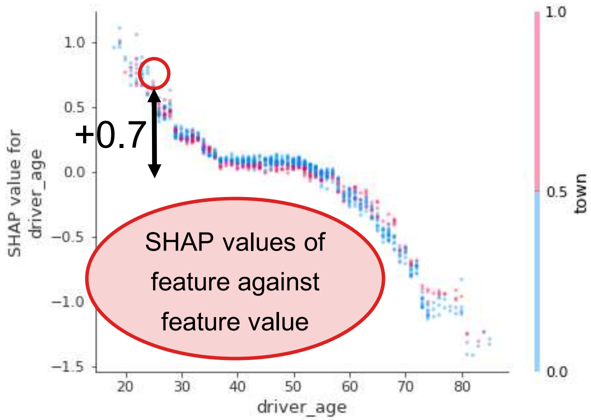
Variable Importance

Note: this is just one way to measure variable importance



Ø|SHAP value| per feature

Dependence Plots

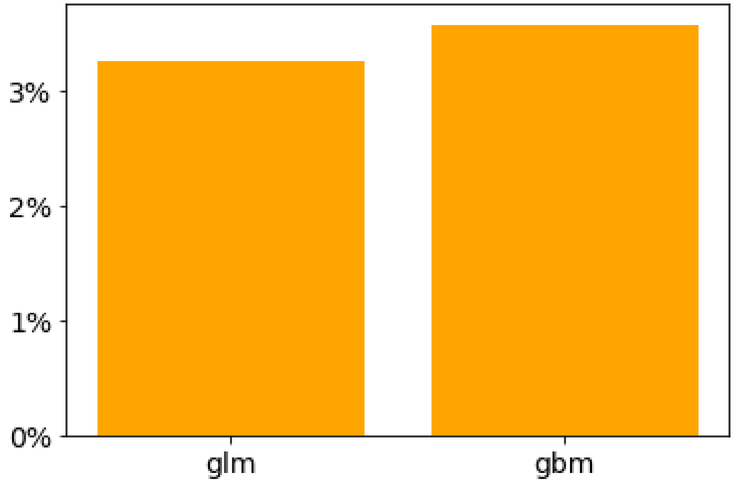


SHAP values of feature against feature value

Vertical scatter indicates interaction effects

Performance on 20% Test Data?

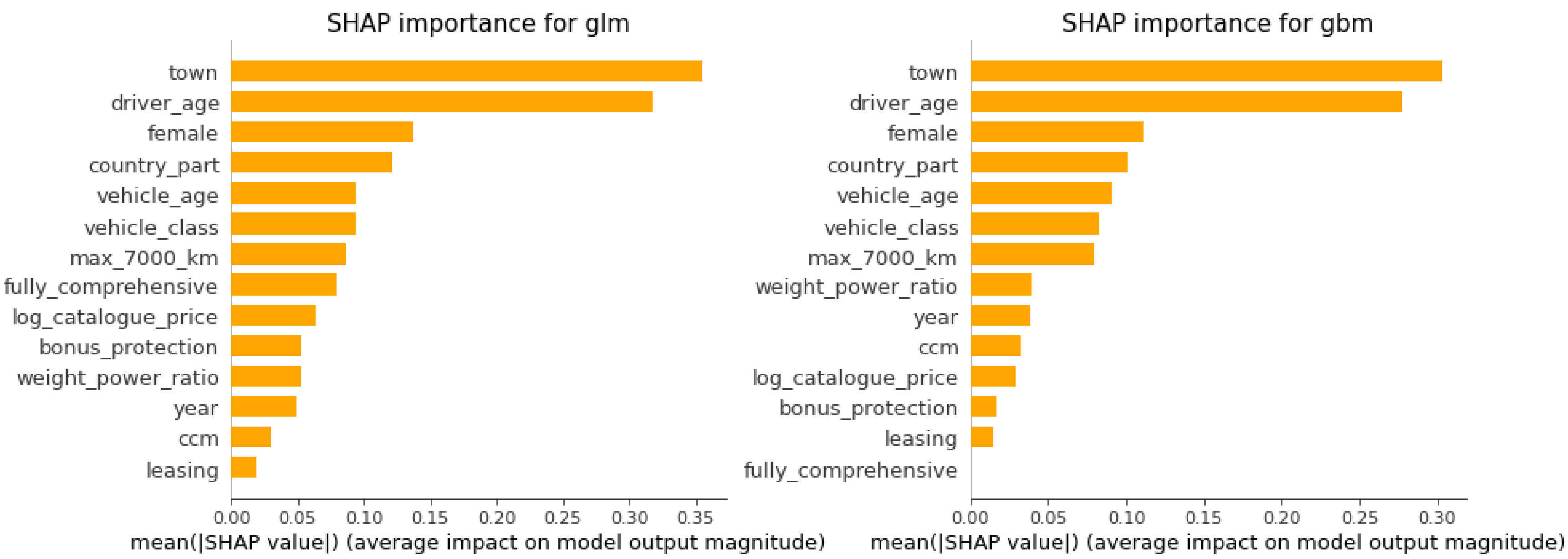
Relative reduction in Poisson deviance loss



Insights

- Makes sense: animal collisions are quite random and cannot be well predicted.
- There is room for improvement for GLM, but not too much.

Important Features?

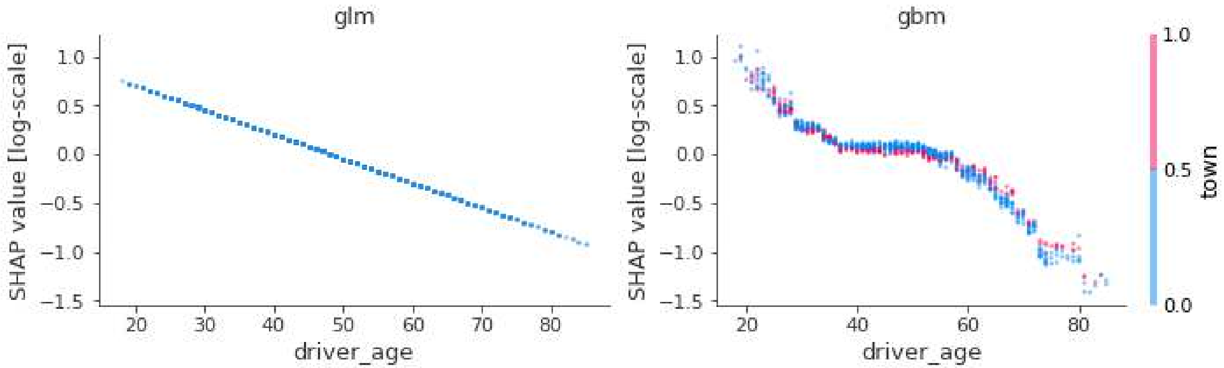


Insights

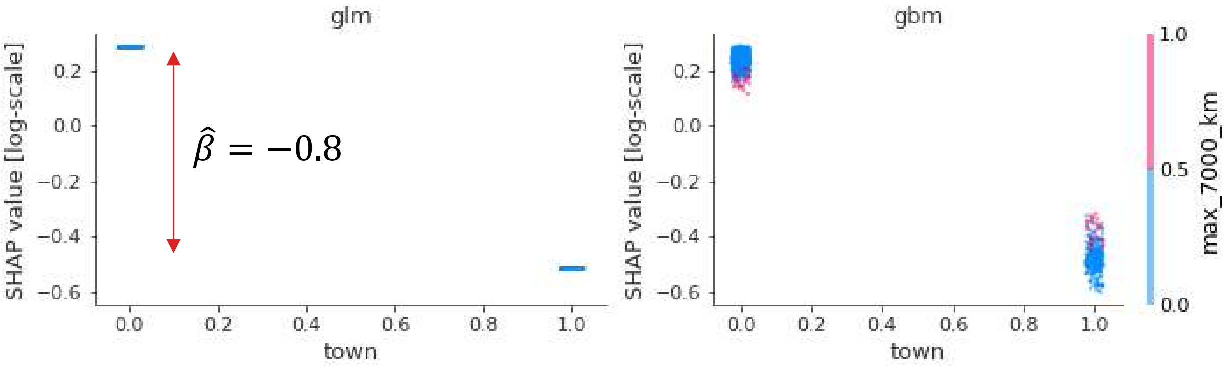
- Results make sense → trust in models grow
- Same variables important across models

Feature Effects?

Example: Driver's age



Example: Town (yes / no)



Insights

- Effects make sense
- Similar across models
- Use additional parameters for "driver's age" in GLM
- Add interaction of "town" and "max_7000_km" to GLM

Action summary

- Add parameters to GLM in guided way by XAI to reduce performance gap
- Go for GLM

Key Takeaways

Key Takeaways

- GLMs stay important in non-life insurance pricing.
- ML + XAI is a great way to improve them.

Resources

1. Ohlsson & Johansson (2010). Non-Life Insurance Pricing with Generalized Linear Models, Springer.
2. Mayer & Lorentzen (2020). Peeking into the Black Box: An Actuarial Case Study for Interpretable Machine Learning, SSRN.
3. SHAP Lundberg & Lee (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems 30.

Code examples:

– github.com/slundberg/shap
– https://github.com/mayer79/python_notebooks
– https://github.com/JSchelldorfer/ActuarialDataScience (Nb 8)