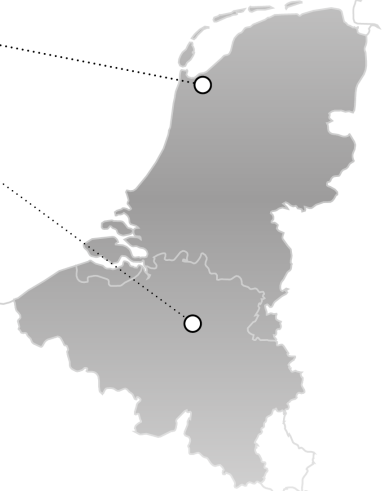
Boosting insights in insurance tariff plans with data science methods

Data Science Leuven, Meetup

Katrien Antonio | LRisk - KU Leuven | 2019-12-17

UNIVERSITEIT VAN AMSTERDAM

AMSTERDAM SCHOOL OF ECONOMICS Economics



KU LEUVEN

LRISK

Find me at...

@katrienantonio

@katrienantonio

katrienantonio.github.io

katrien.antonio@kuleuven.be

Four important principles

Inverted production cycle.

The contributions of the many to cover the misfortunes of the few.

Segmentation or risk classification.

Highly regulated business.

Inverted production cycle

Data driven business ~ estimation and prediction matter!

Pricing

- total cost of a contract only known ex post
- but, premium paid a start of the contract!

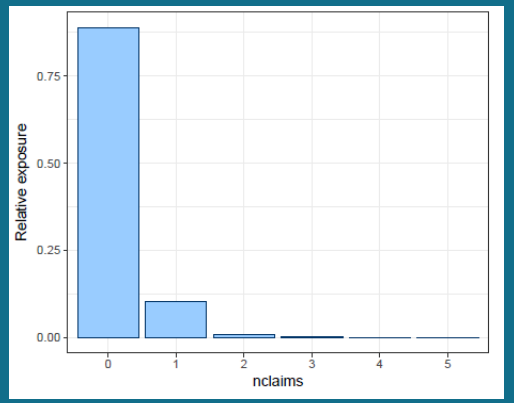
Reserving

- future development of claims?
- how much capital to hold?

Misfortunes of the few

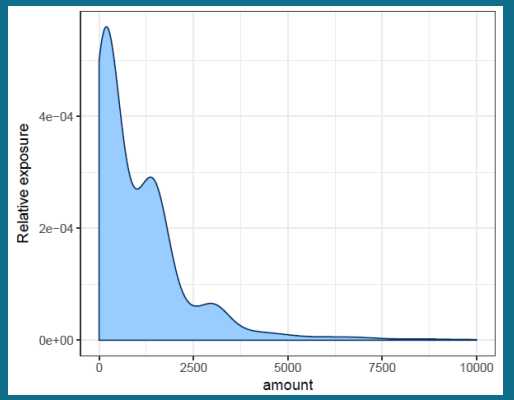
covered by contributions of the many

frequency



- number of (insured) events that occurred
- low frequency, very often zero!

severity



- only contracts with frequency > 0
- potentially high impact, (heavy) right tail

Figures from Henckaerts, Antonio et al. (2018). A data driven binning strategy for the construction of insurance tariff classes

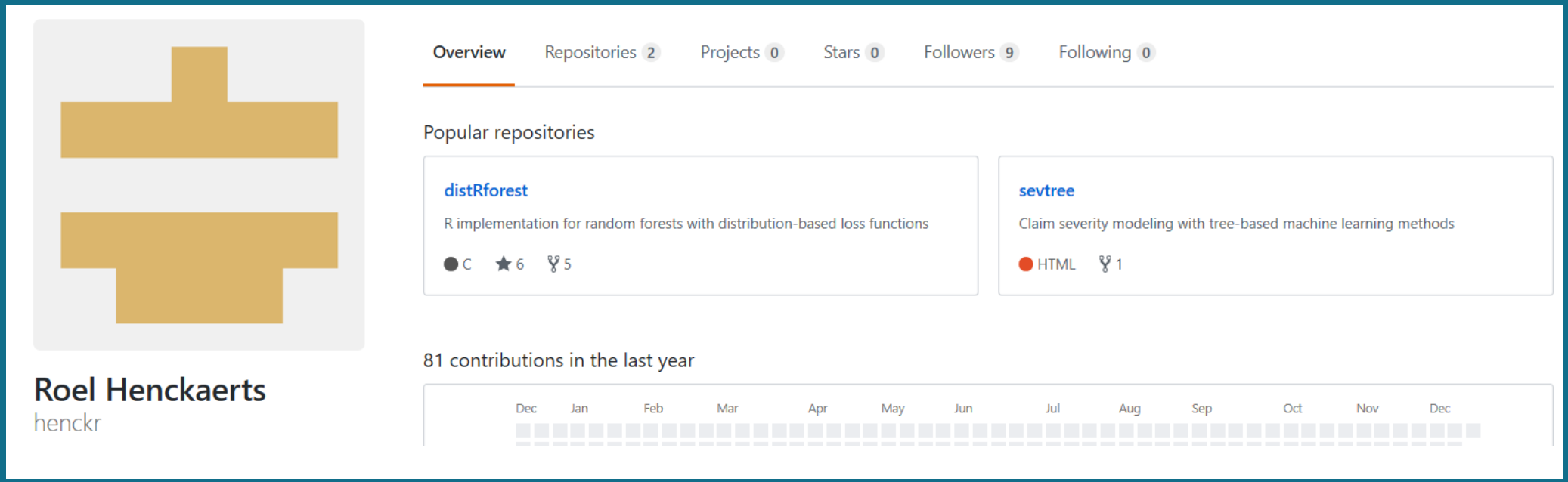
Contributions of the many

E(frequency) x E(severity)

Use suitable loss functions, e.g.

Poisson (for frequency) and gamma (for severity)

Henckaerts, Côté, Antonio et al. (2019). Boosting insights in insurance tariff plans with tree-based machine learning methods

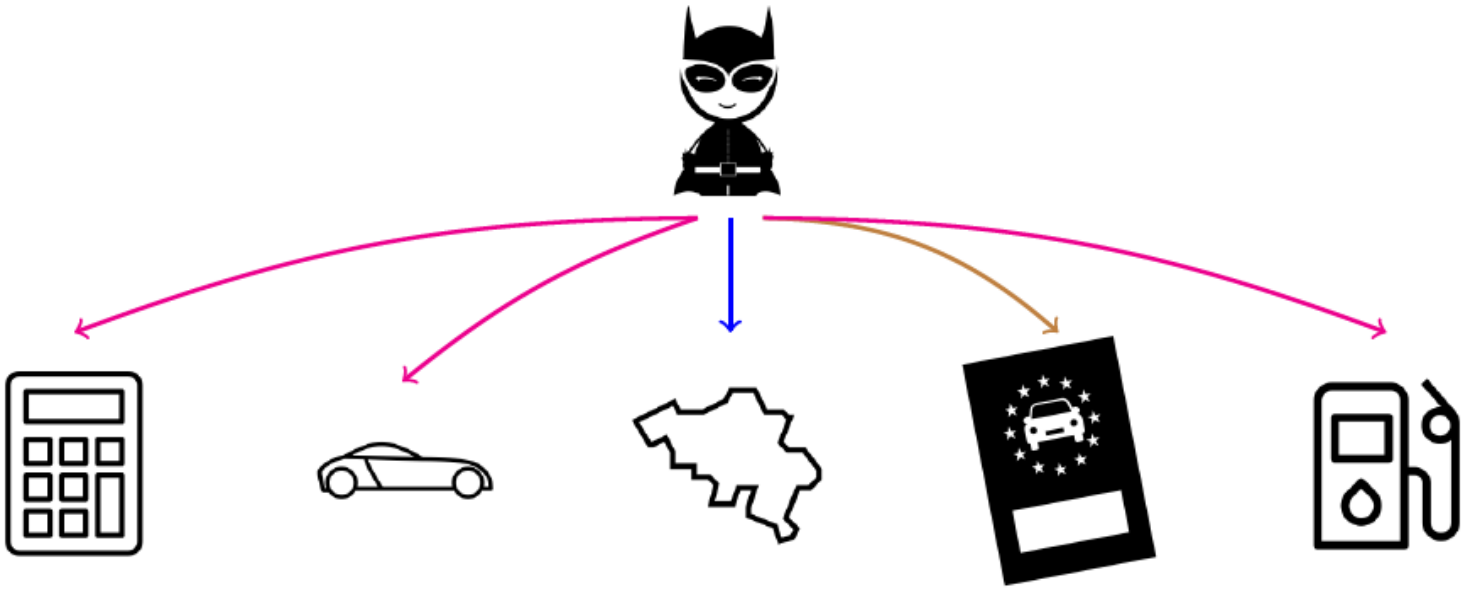


Suitable loss functions for frequency and severity tree-based modelling

Henckaerts, Côté, Antonio et al. (2019). Boosting insights in insurance tariff plans with tree-based machine learning methods

Risk classification

Multiple types of features ~ classics



Henckaerts, Antonio et al. (2018). A data driven binning strategy for the construction of insurance tariff classes

Devriendt, Antonio et al. (2018). Sparse regression with multi-type regularized feature modeling

Henckaerts, Côté, Antonio et al. (2019). Boosting insights in insurance tariff plans with tree-based machine learning methods

Multiple types of features ~ new stuff

James B.

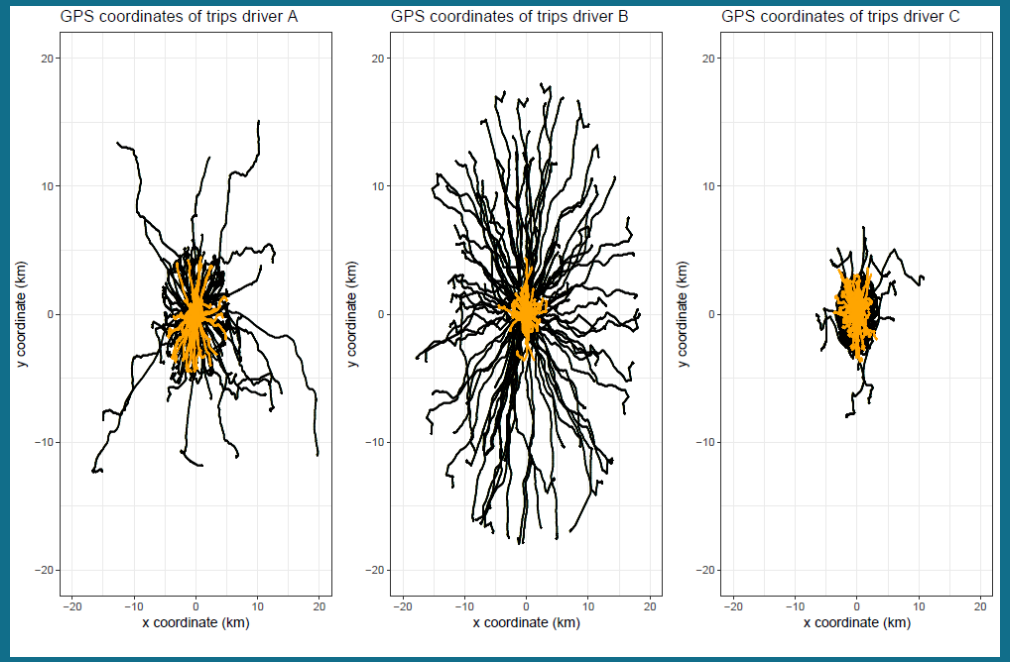
James B. drives 100 000 km with road type composition (0.15, 0.15, 0.5, 0.2).

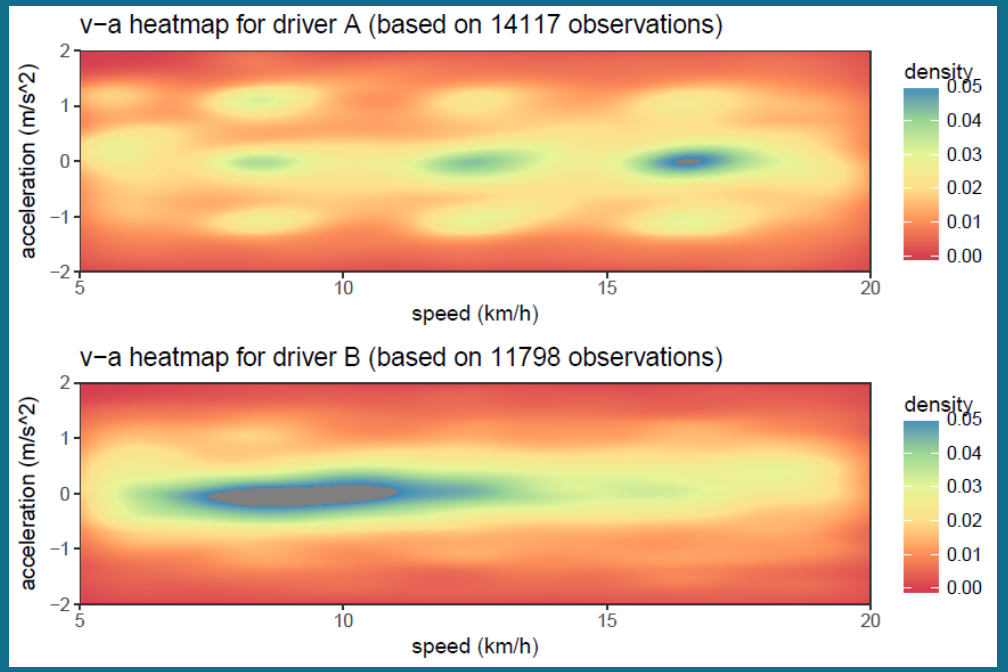
Eugene

Eugene drives 1 000 km with road type composition (0.5, 0.3, 0.2, 0).

Verbelen, Antonio et al. (2018). Unravelling the predictive power of telematics data in car insurance pricing

Multiple types of features ~ new stuff



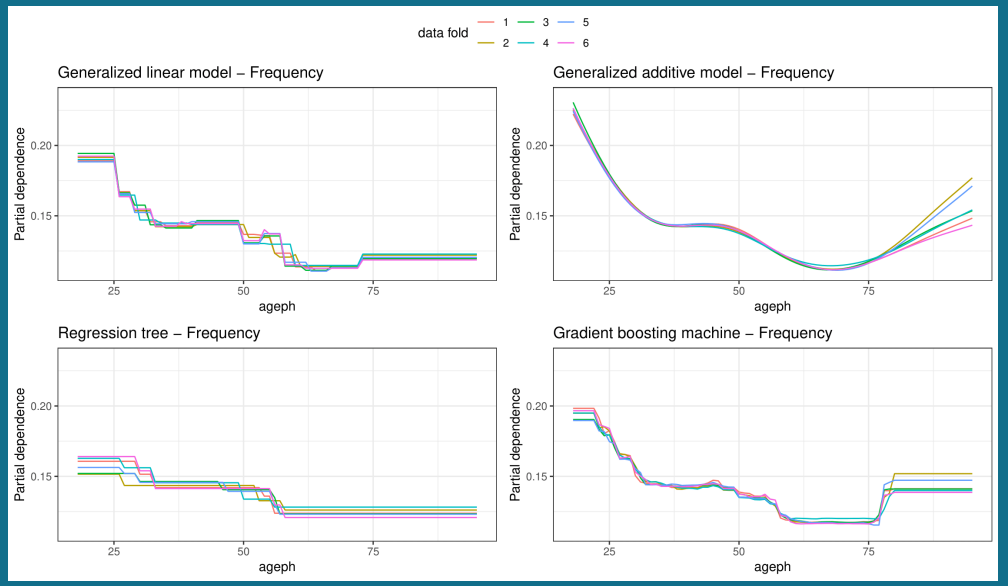


Feature engineering matters!

Pictures based on the work of prof. Mario Wüthrich (ETH) and data from the AXA Kaggle competition.

Highly regulated business

Interpretability matters



PDPs, ICEs, variable importance plots, ...

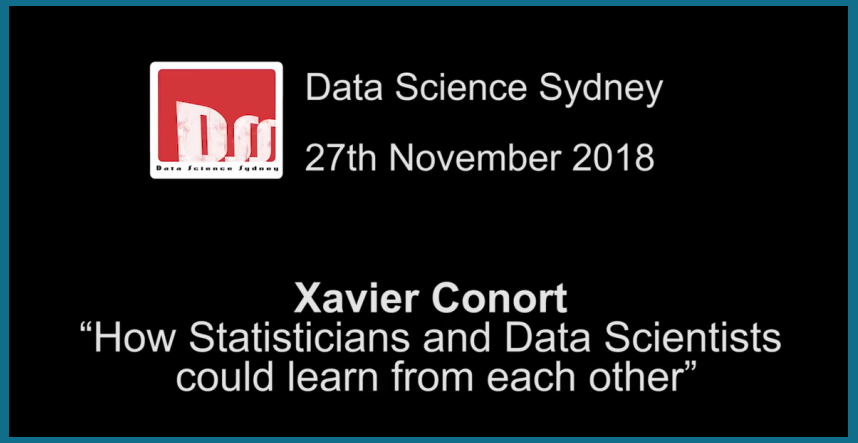
But, **GLMs** are still preferred tool, with

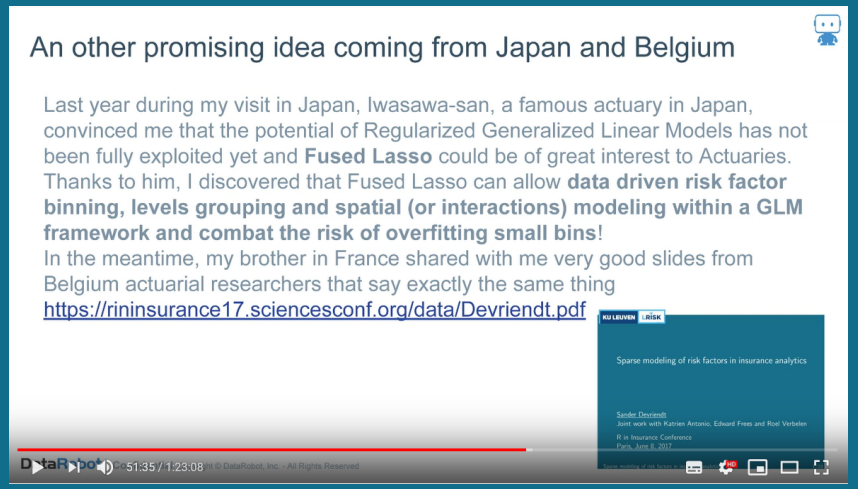
$$\mathrm{E}[Y] = \exp\left(\mathbf{x}^{'} \cdot \boldsymbol{\beta}\right),$$

and suitable loss distributions.

Figure from Henckaerts, Côté, Antonio et al. (2019). Boosting insights in insurance tariff plans with tree-based machine learning methods

Construction of the GLMs





GLMs with regularization

lasso-type penalties

$$-\log\ \mathcal{L}(\beta_0,\ \boldsymbol{\beta}) + \lambda \sum_{j=1}^{J} g_j(\beta_j)$$

`smurf` package for R

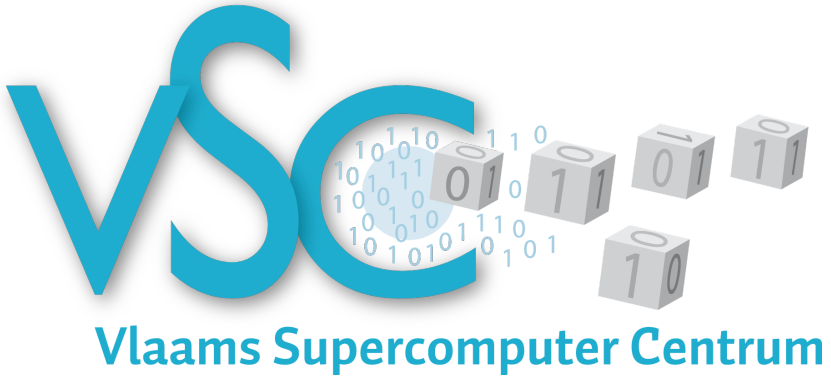
Devriendt, Antonio et al. (2018). Sparse regression with multi-type regularized feature modeling

Thanks to


ageas


fwo
Research Foundation
Flanders
Opening new horizons


ARGENTA


VSC
Vlaams Supercomputer Centrum

Thanks!

Slides created with the R package xaringan, inspired by the work of Alison Hill.