Institute and Faculty of Actuaries

Machine Learning

Duncan Anderson

GIRO 2017

19 October 2017

GIRO 2016 - Response to machine learning

Don’t panic!

We’re doomed!

This is not all new

Actuaries adopt GLMs

First computer

Neural nets

Trees

GLMs

CART

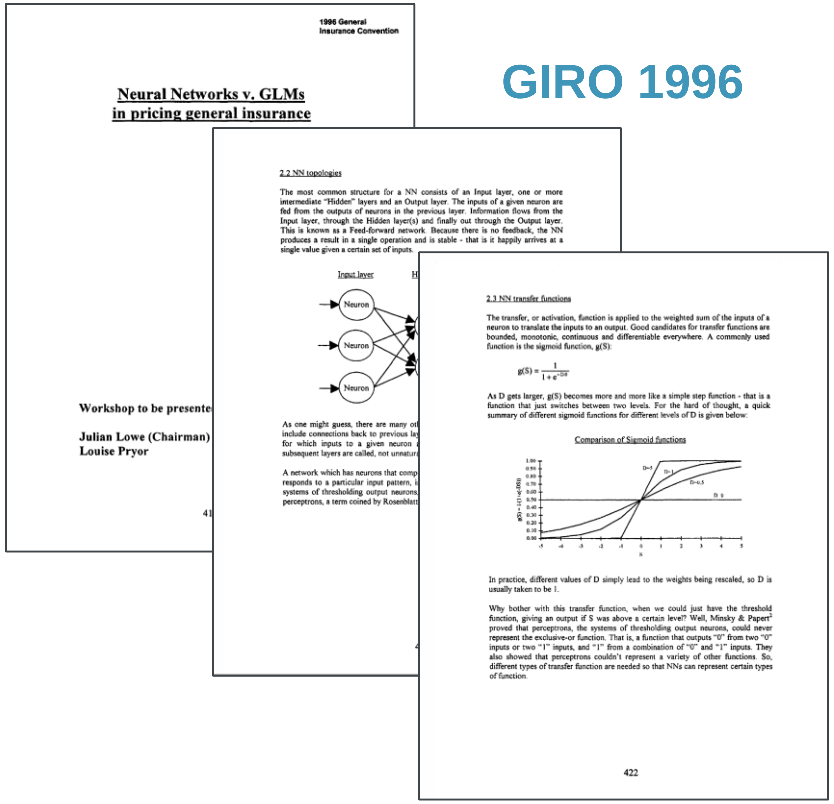
Random forests

GBMs

1940s | 1950s | 1960s | 1970s | 1980s | 1990s | 2000s | 2010s | Today

Deep Blue vs Kasparov

AlphaGo vs Lee Sedol


1996 General Insurance Convention

Neural Networks v. GLMs in pricing general insurance

Workshop to be presente

Julian Lowe (Chairman)
Louise Pryor

41

2.2 NN topologies

The most common structure for a NN consists of an Input layer, one or more intermediate "Hidden" layers and an Output layer. The inputs of a given neuron are fed from the outputs of neurons in the previous layer. Information flows from the Input layer, through the Hidden layer(s) and finally out through the Output layer. This is known as a Feed-forward network. Because there is no feedback, the NN produces a result in a single operation and is stable - that is it happily arrives at a single value given a certain set of inputs.

As one might guess, there are many ot include connections back to previous la for which inputs to a given neuron subsequent layers are called, not unnatur

A network which has neurons that comp responds to a particular input pattern, systems of thresholding output neurons perceptrons, a term coined by Rosenblatt

2.3 NN transfer functions

The transfer, or activation, function is applied to the weighted sum of the inputs of a neuron to translate the inputs to an output. Good candidates for transfer functions are bounded, monotonic, continuous and differentiable everywhere. A commonly used function is the sigmoid function, g(S):

$$g(S) = \frac{1}{1 + e^{-DS}}$$

As D gets larger, g(S) becomes more and more like a simple step function - that is a function that just switches between two levels. For the hard of thought, a quick summary of different sigmoid functions for different levels of D is given below:

Comparison of Sigmoid functions

In practice, different values of D simply lead to the weights being rescaled, so D is usually taken to be 1.

Why bother with this transfer function, when we could just have the threshold function, giving an output if S was above a certain level? Well, Minsky & Papert[2] proved that perceptrons, the systems of thresholding output neurons, could never represent the exclusive-or function. That is, a function that outputs "0" from two "0" inputs or two "1" inputs, and "1" from a combination of "0" and "1" inputs. They also showed that perceptrons couldn't represent a variety of other functions. So, different types of transfer function are needed so that NNs can represent certain types of function.

422

GIRO 1996

- Hyper scale parallel computing
- NoSQL databases
- Free software environments, languages, analytics libraries
- Data stream and real-time processing supporting IoT
- Distributed Big Data storage/ processing with Hadoop
- Data visualisation tools
- Integrated environments and services

There is a spectrum of complexity

“Vastly more risky than North Korea”

GLM stepwise macro

AI comprehension | Bespoke image recognition | Speech analytics | Machine learning predive modelling

Full autonomous driving | Object recognition | Topic modelling | Automated GLMs

Hard
Evolving
Requires significant expertise

Not at all hard
Already in use
Actuaries can do this stuff

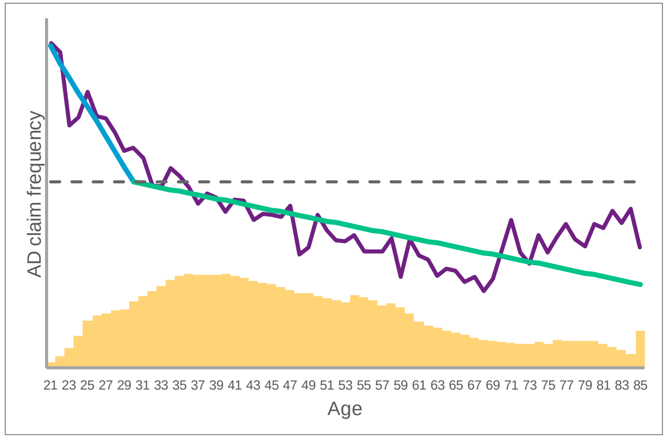
Example machine learning methods

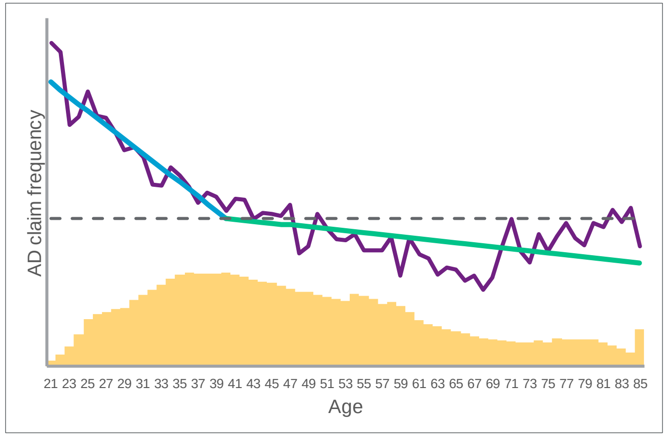
- Ensemble Methods
- Classifications Trees
- "Earth"
- Gradient Boosting Machines
- K-Means Clustering
- Support Vector Machines
- Elastic Net
- Neural Networks
- Naïve Bayes
- Random Forests
- Regression Trees
- Principal Components Analysis
- Lasso
- K-nearest Neighbours
- Ridge Regression

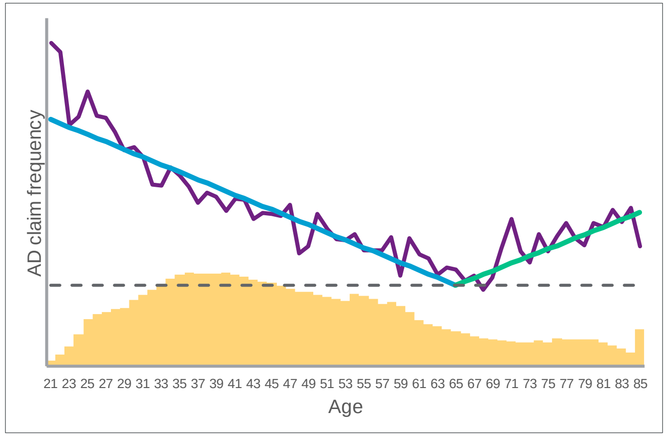
Example machine learning methods

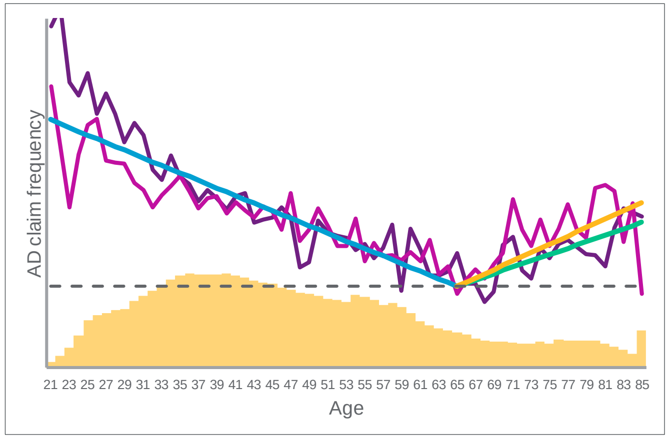
- Ensemble Methods
- Classifications Trees
- "Earth"
- Gradient Boosting Machines
- K-Means Clustering
- Support Vector Machines
- Elastic Net
- Neural Networks
- Naïve Bayes
- Random Forests
- Regression Trees
- Principal Components Analysis
- Lasso
- K-nearest Neighbours
- Ridge Regression

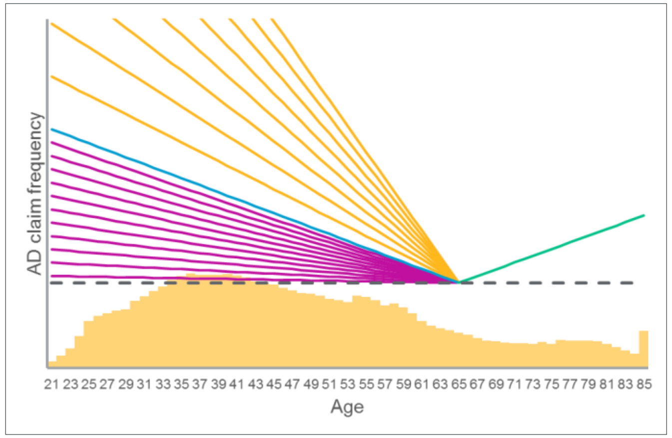
Multivariate adaptive regression splines ("Earth")











Penalised regression (Lasso, Ridge, Elastic Net)

GLM Lasso Ridge

f($\underline{x}$) = $g^{-1}(\mathbf{X}.\underline{\beta})$ where $\underline{\beta}$ estimated by minimising $L(\beta|X,y) + \boxed{\boxed{\lambda_1 \sum_i |\beta_i|} + \boxed{\lambda_2 \sum_i \beta_i^2}}$

Elastic Net

1	2	3	4	5
Fit	Fit	Fit	Fit	Test
Fit	Fit	Fit	Test	Fit
Fit	Fit	Test	Fit	Fit
Fit	Test	Fit	Fit	Fit
Test	Fit	Fit	Fit	Fit
Holdout	Holdout	Holdout	Holdout	Holdout

Example machine learning methods

Ensemble Methods

Classifications Trees

"Earth"

Gradient Boosting Machines

K-Means Clustering

Support Vector Machines

Elastic Net

Neural Networks

Naïve Bayes

Random Forests

Regression Trees

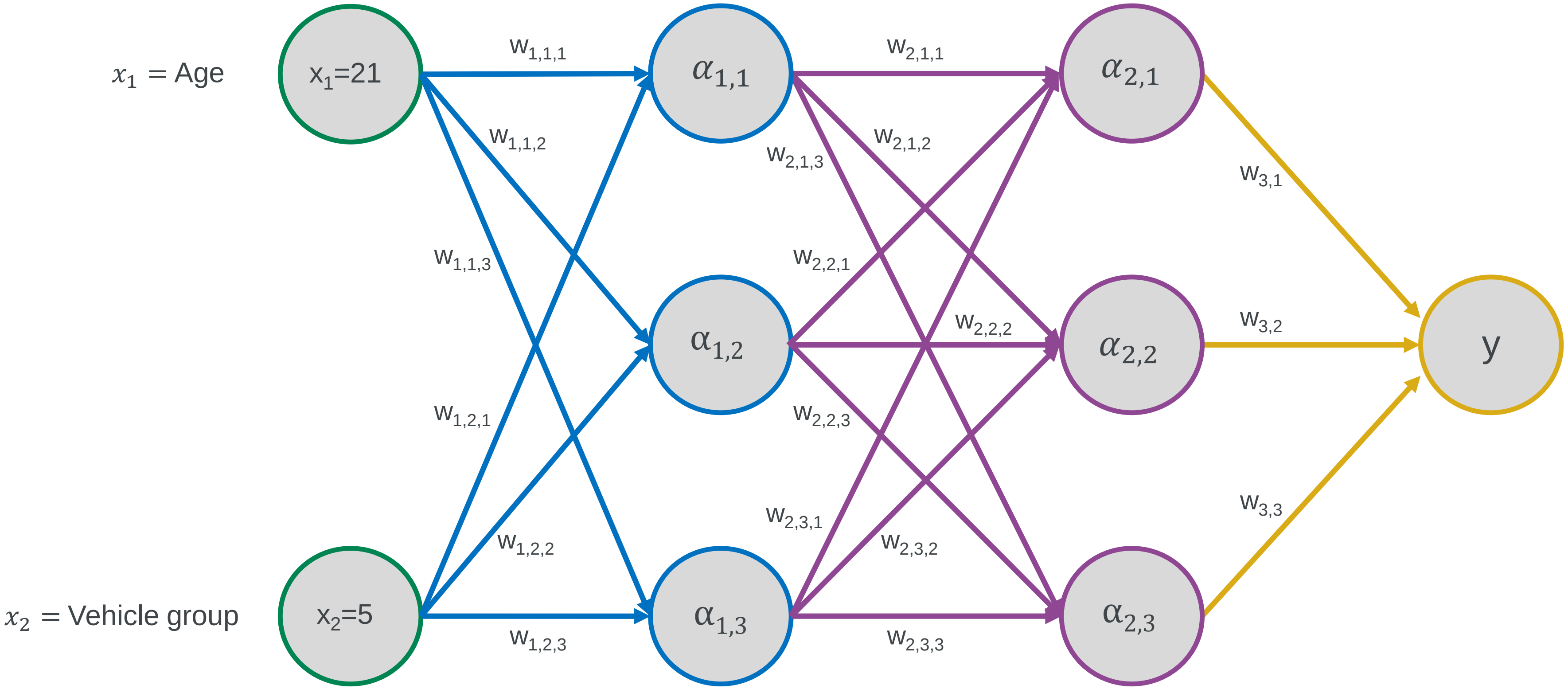
Principal Components Analysis

Lasso

K-nearest Neighbours

Ridge Regression

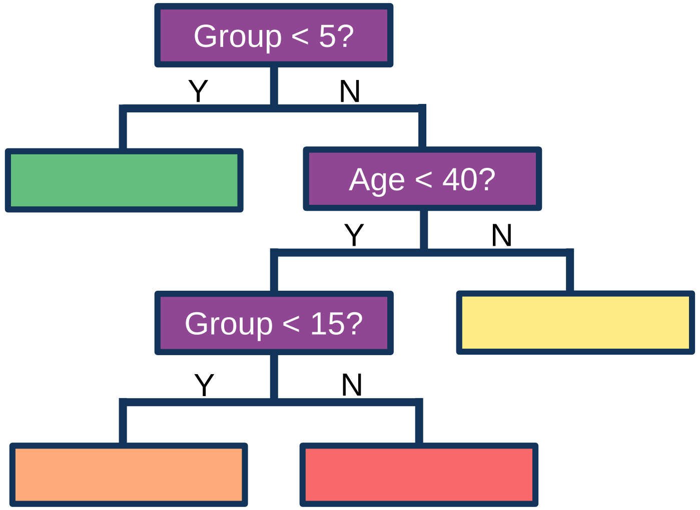
Neural networks

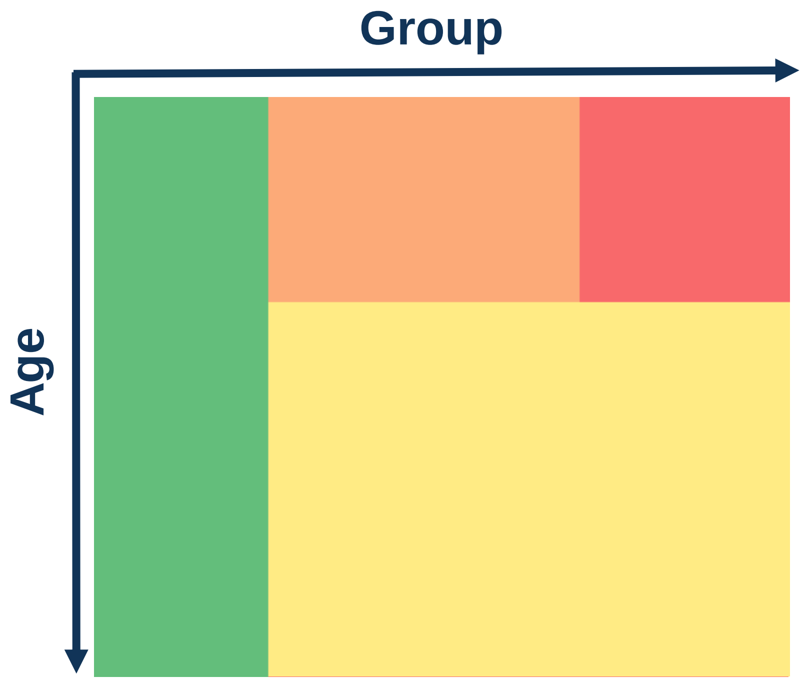


Example machine learning methods

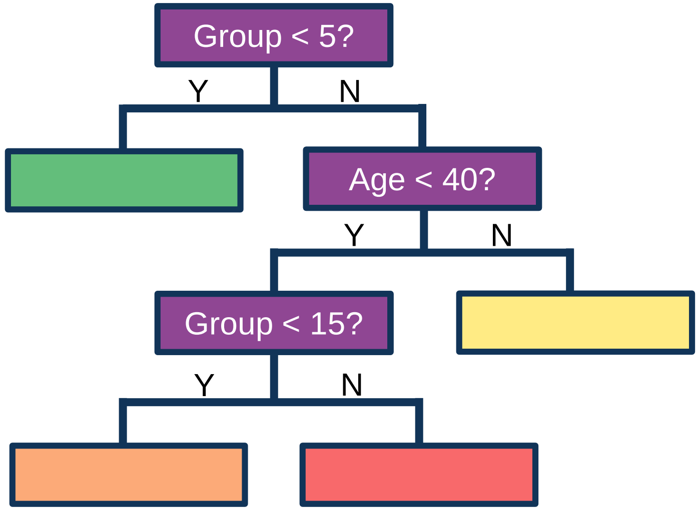
- Ensemble Methods
- Classifications Trees
- "Earth"
- Gradient Boosting Machines
- K-Means Clustering
- Support Vector Machines
- Elastic Net
- Neural Networks
- Naïve Bayes
- Random Forests
- Regression Trees
- Principal Components Analysis
- Lasso
- K-nearest Neighbours
- Ridge Regression

Decision trees



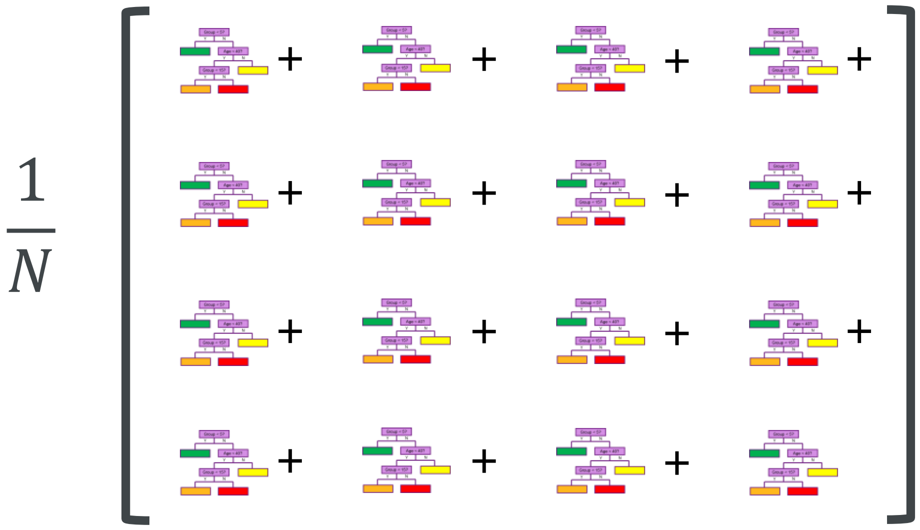


Random Forests

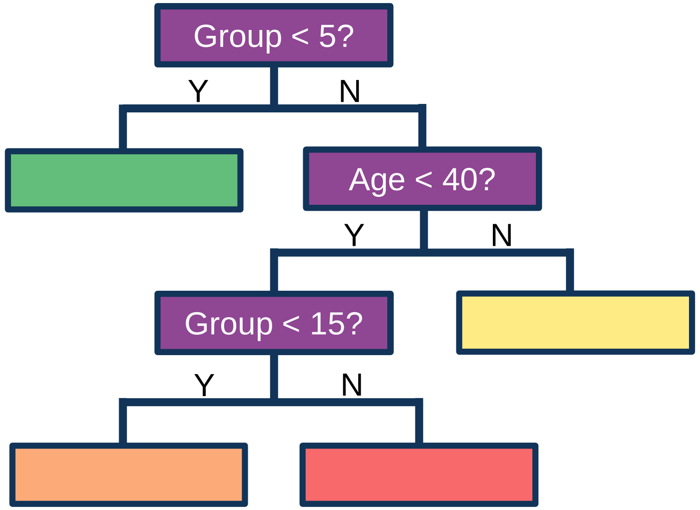


A random forest

$$f(x) = \frac{1}{N}\sum_{n=1}^{N} f_n(x)$$

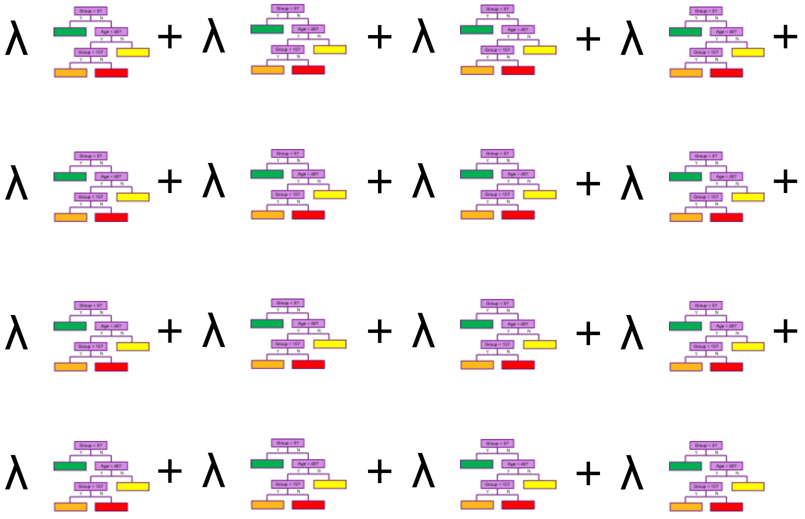


Gradient Boosted Machine or “GBM”



A GBM

$$f(x) = \lambda \sum_{n=1}^{N} f_n(x)$$



Example machine learning methods

- Ensemble Methods
- Classifications Trees
- "Earth"
- Gradient Boosting Machines
- K-Means Clustering
- Support Vector Machines
- Elastic Net
- Neural Networks
- Naïve Bayes
- Random Forests
- Regression Trees
- Principal Components Analysis
- Lasso
- K-nearest Neighbours
- Ridge Regression

Do they add value?

Ensemble Methods

Classifications Trees

"Earth"

Gradient Boosting Machines

K-Means Clustering

Support Vector Machines

Elastic Net

Neural Networks

Naïve Bayes

Random Forests

Regression Trees

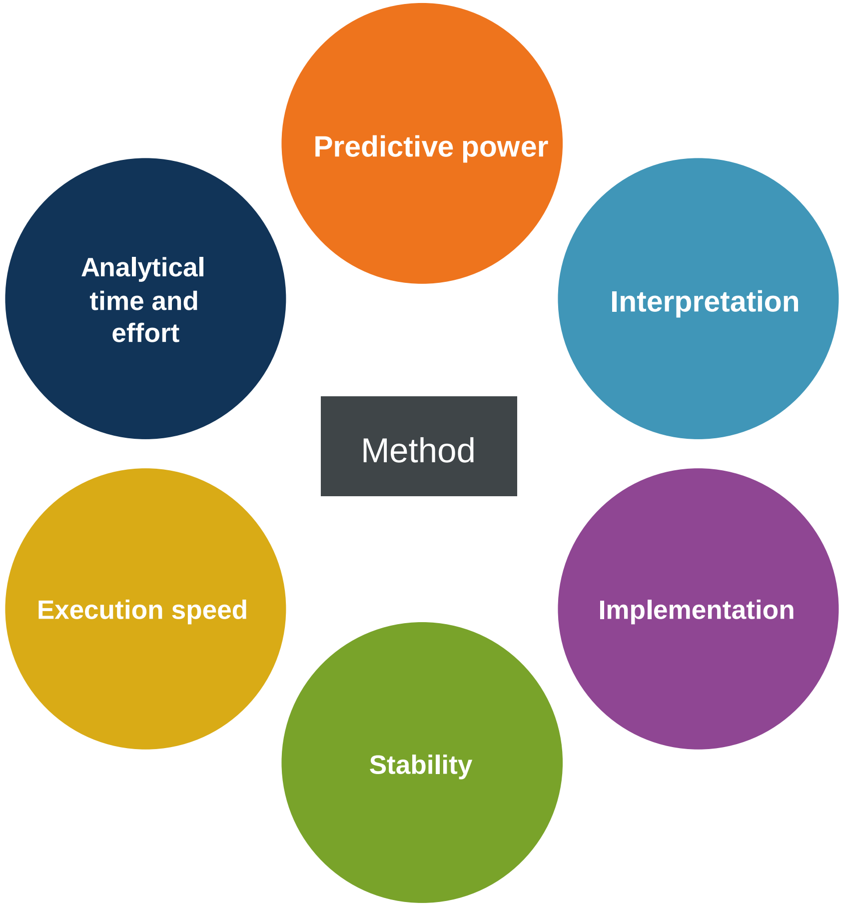
Principal Components Analysis

Lasso

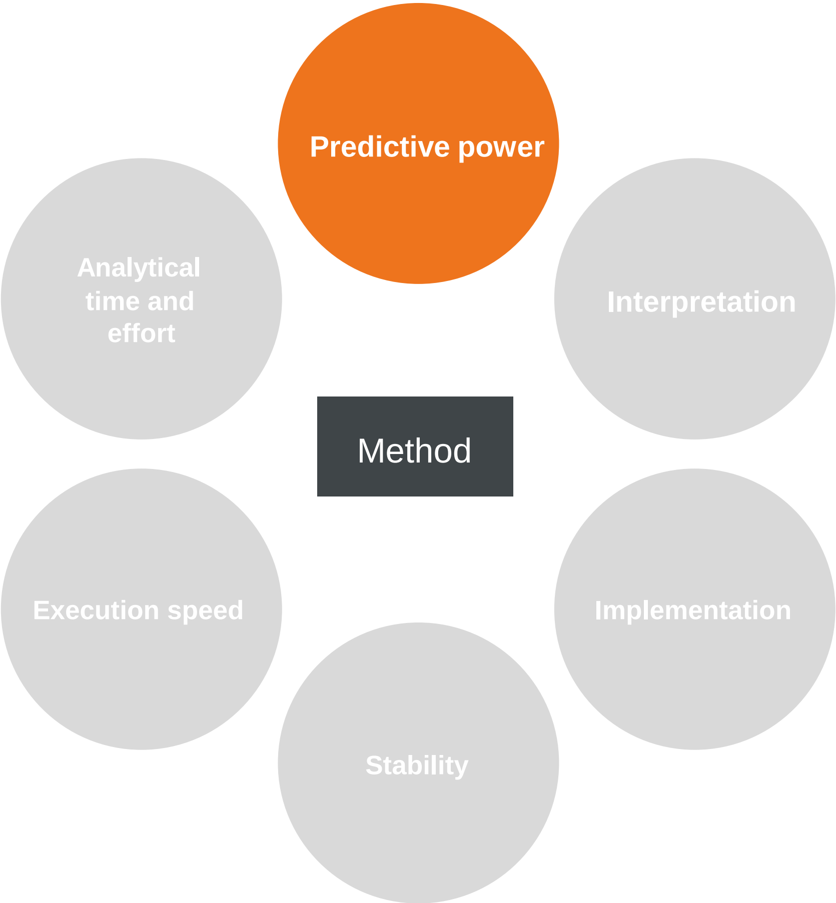
K-nearest Neighbours

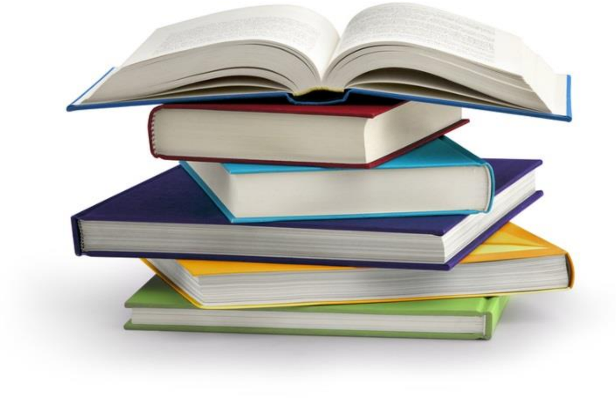
Ridge Regression

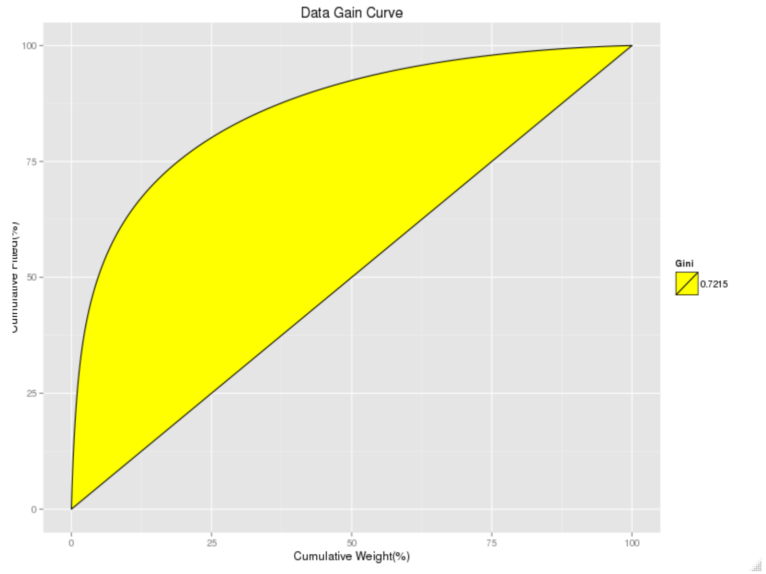
Dimensions of utility

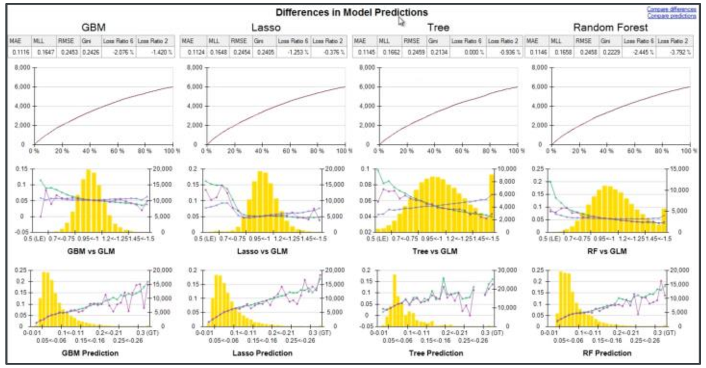


Dimensions of utility









Loss ratio improvement 3.1%!

Dimensions of utility

Predictive power

Analytical time and effort

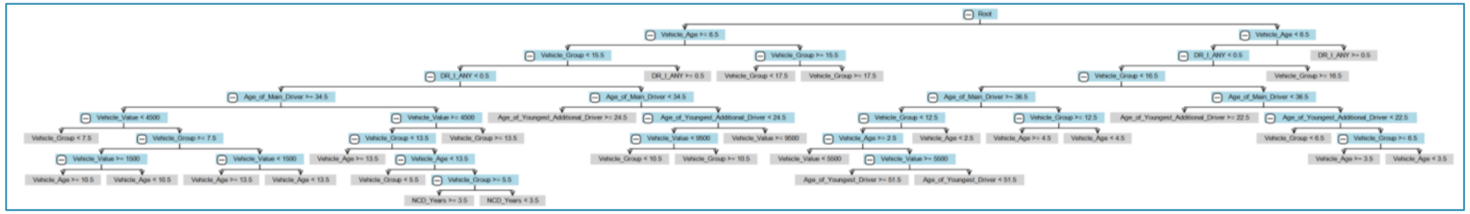
Interpretation

Method

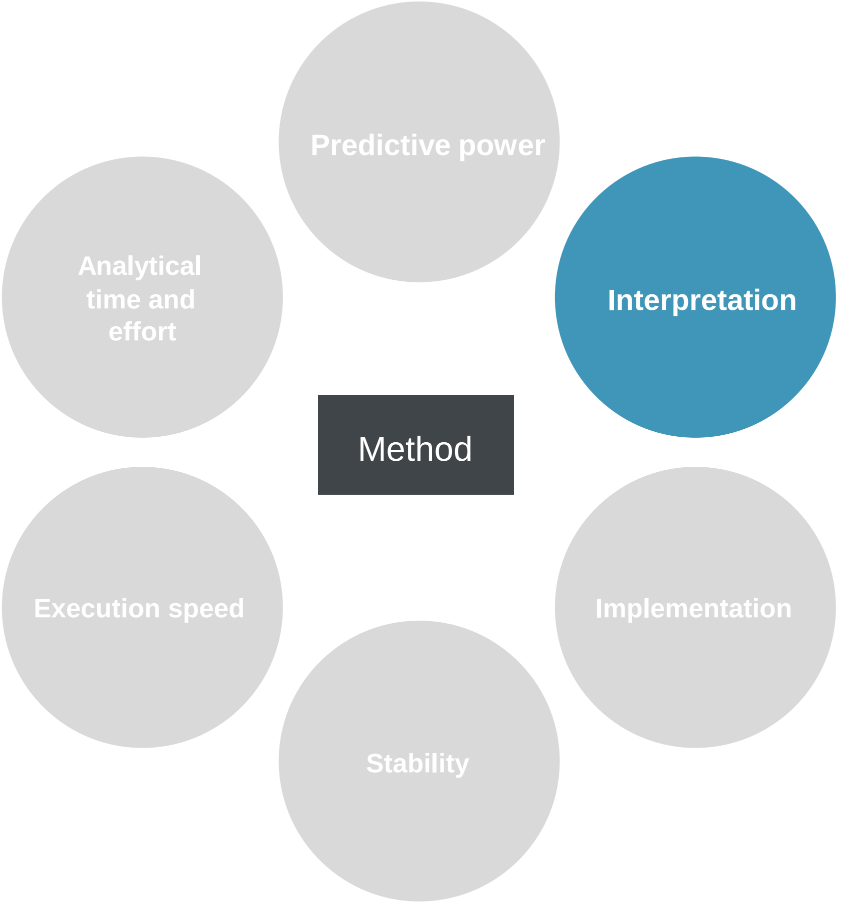
Execution speed

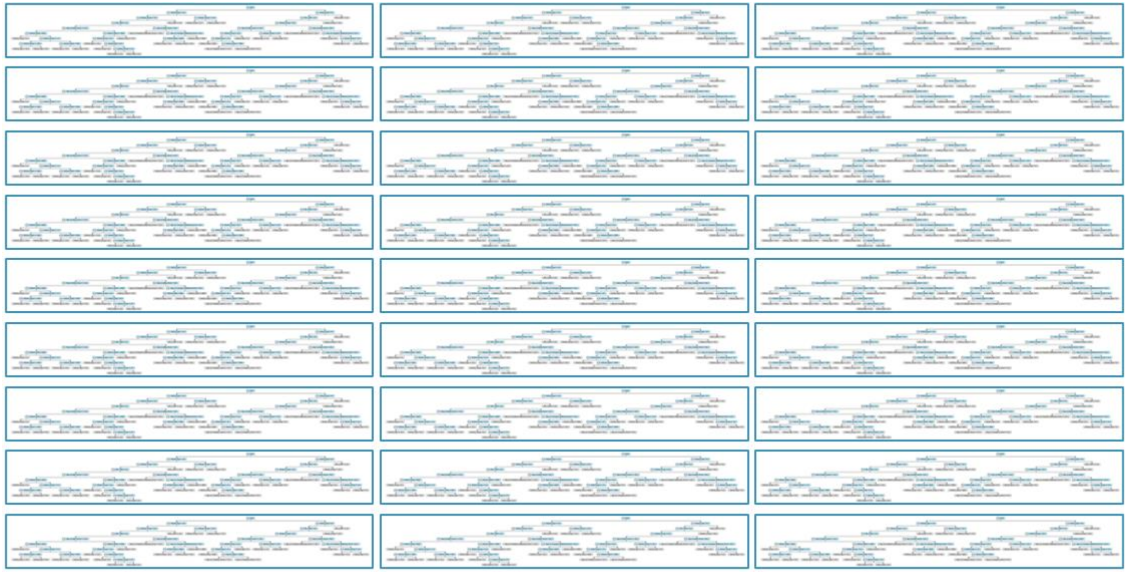
Implementation

Stability



Dimensions of utility





Dimensions of utility

Predictive power

Analytical time and effort

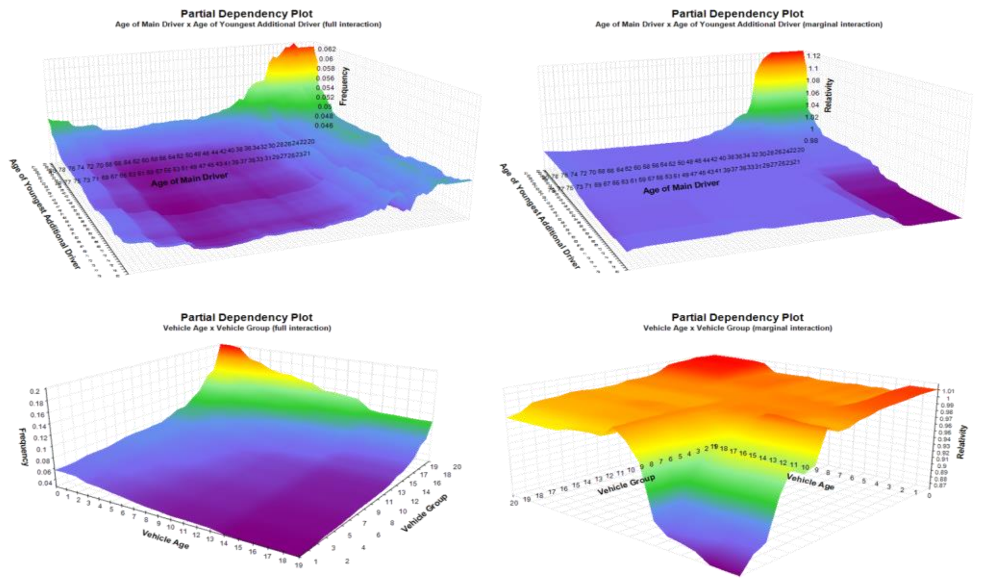
Interpretation

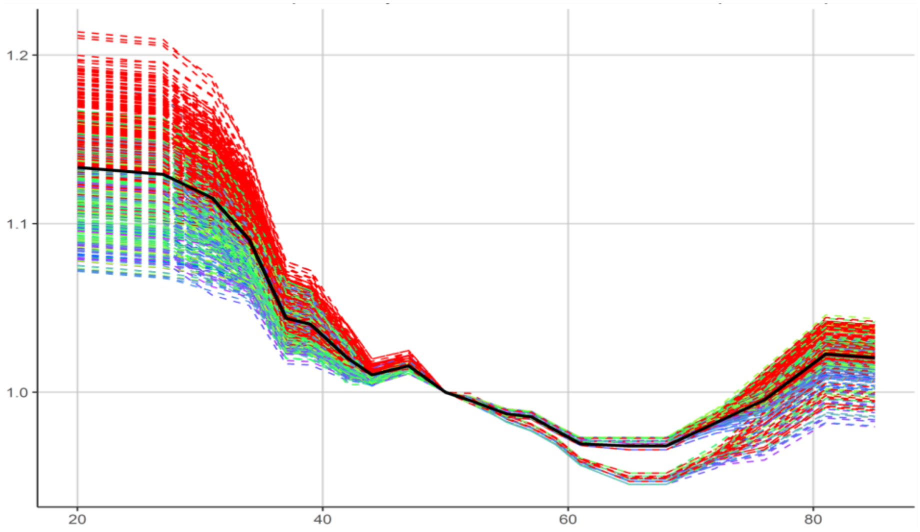
Method

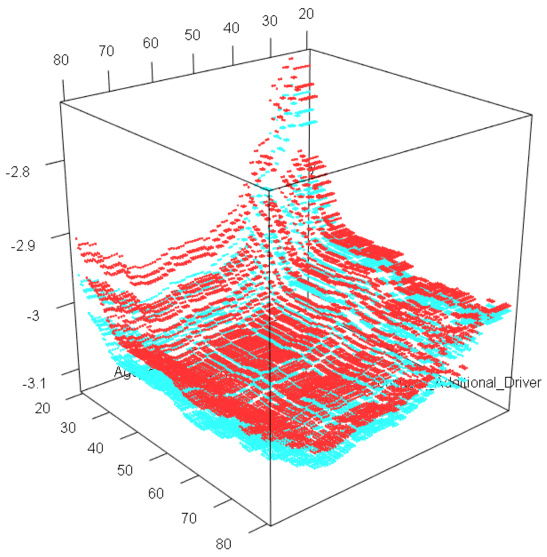
Execution speed

Implementation

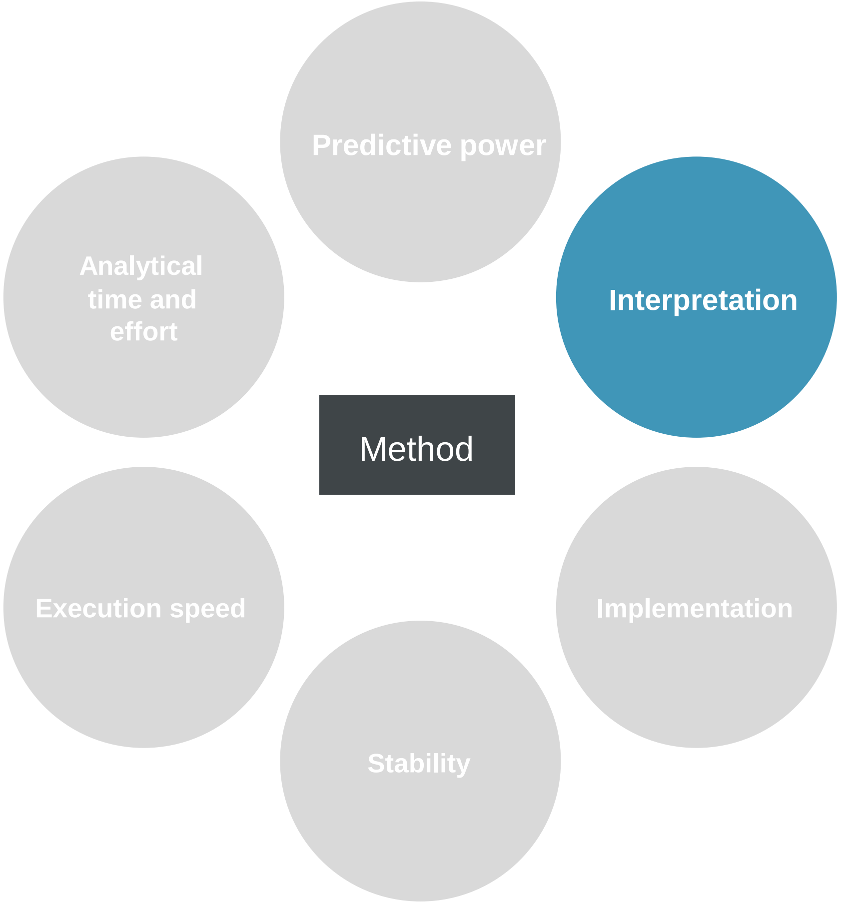
Stability



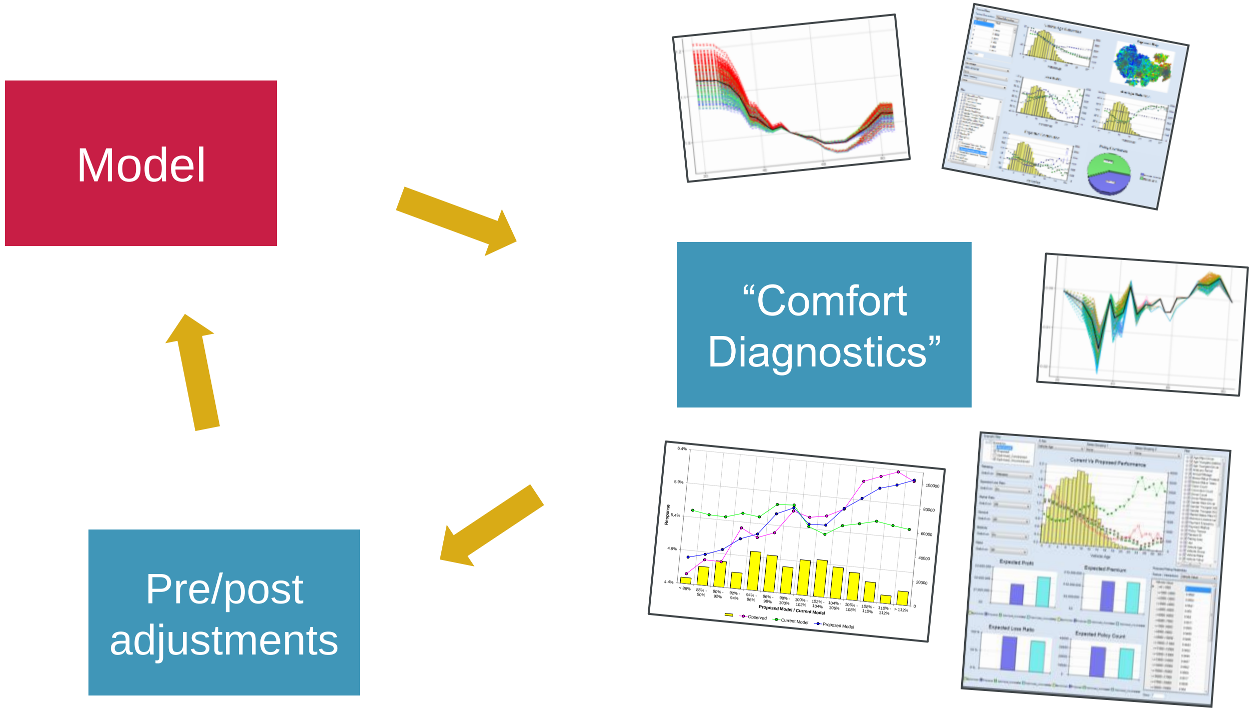




Dimensions of utility



- How much do you need to understand?
  - How much would you normally understand? (eg vehicle classification)
  - Cost of error? (eg marketing)
  - Regulatory requirements
  - Professional standards
- “Comfort diagnostics”



Dimensions of utility

Predictive power

Analytical time and effort

Interpretation

Method

Execution speed

Implementation

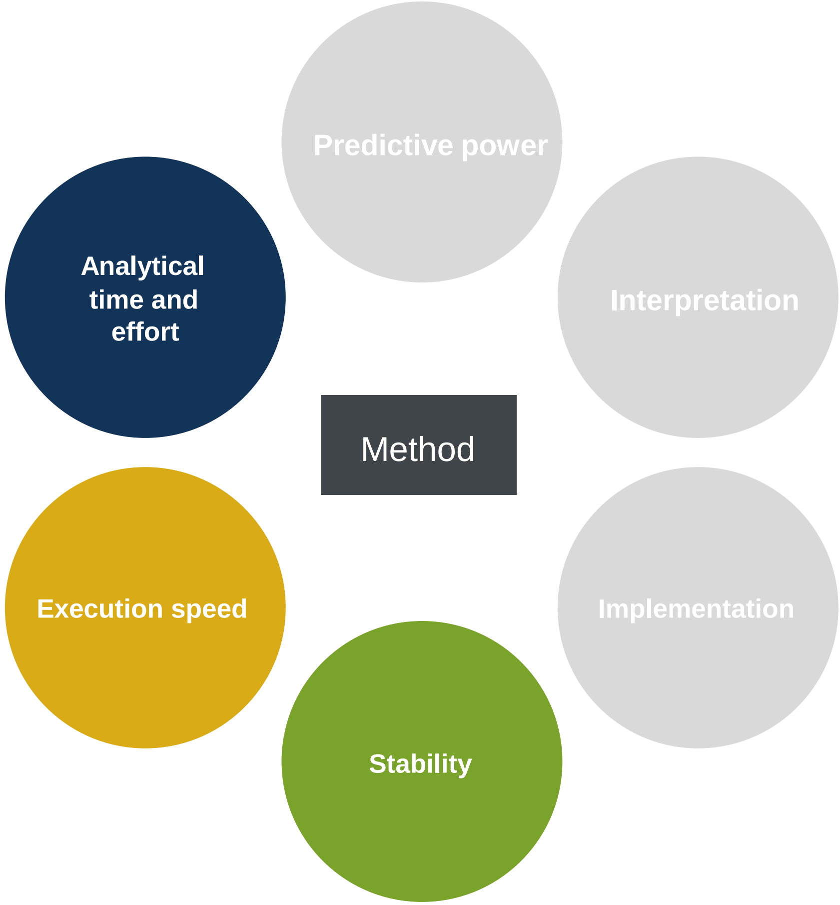
Stability

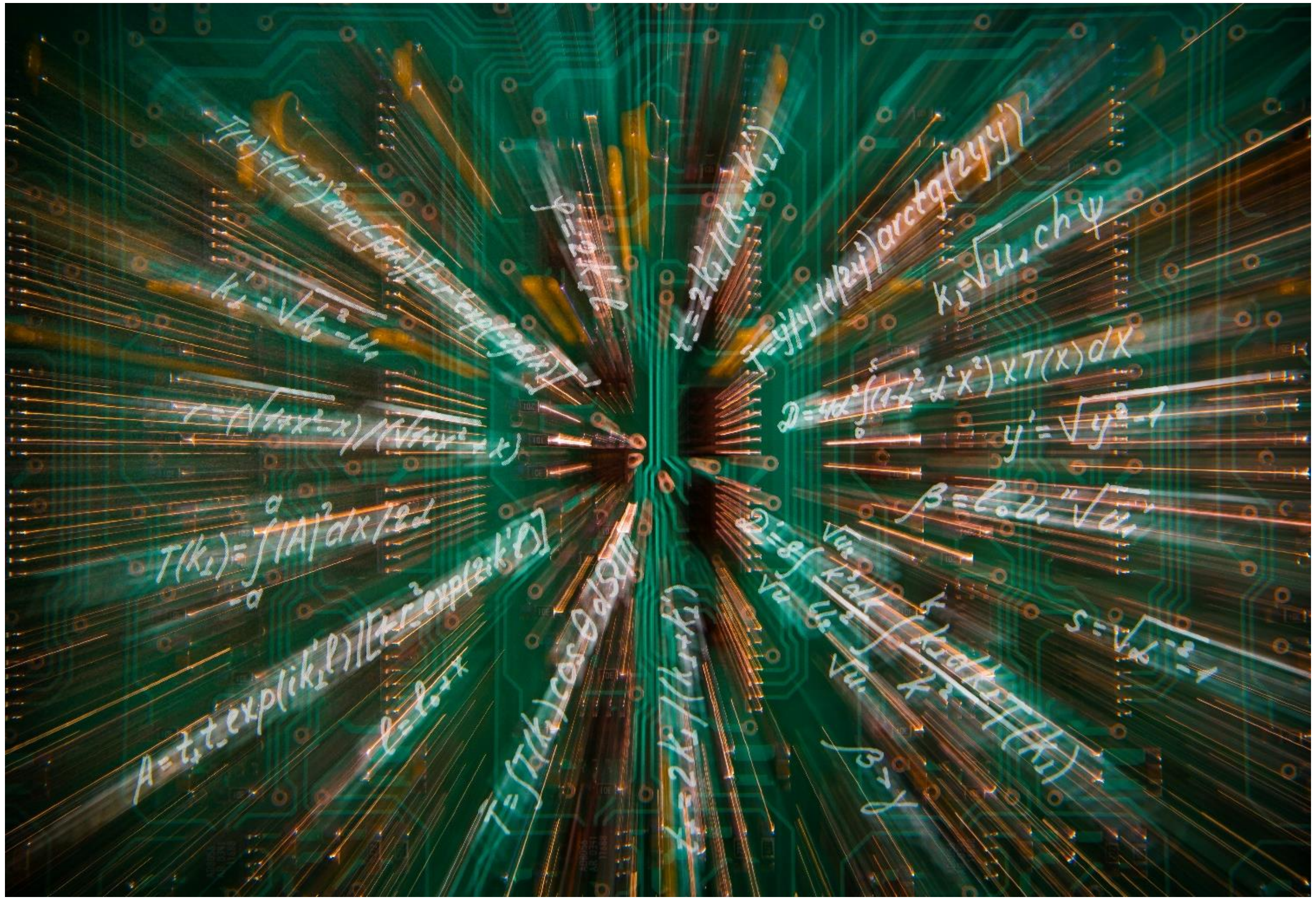
Analytical environment

Next generation rating engine

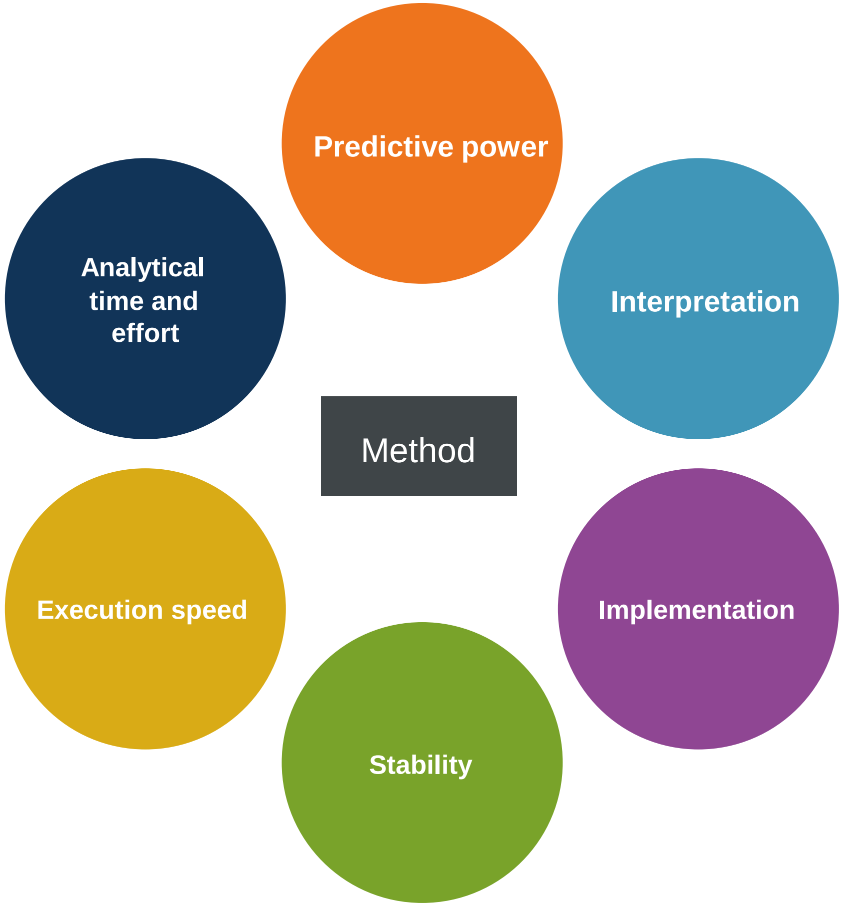
Policy administration system

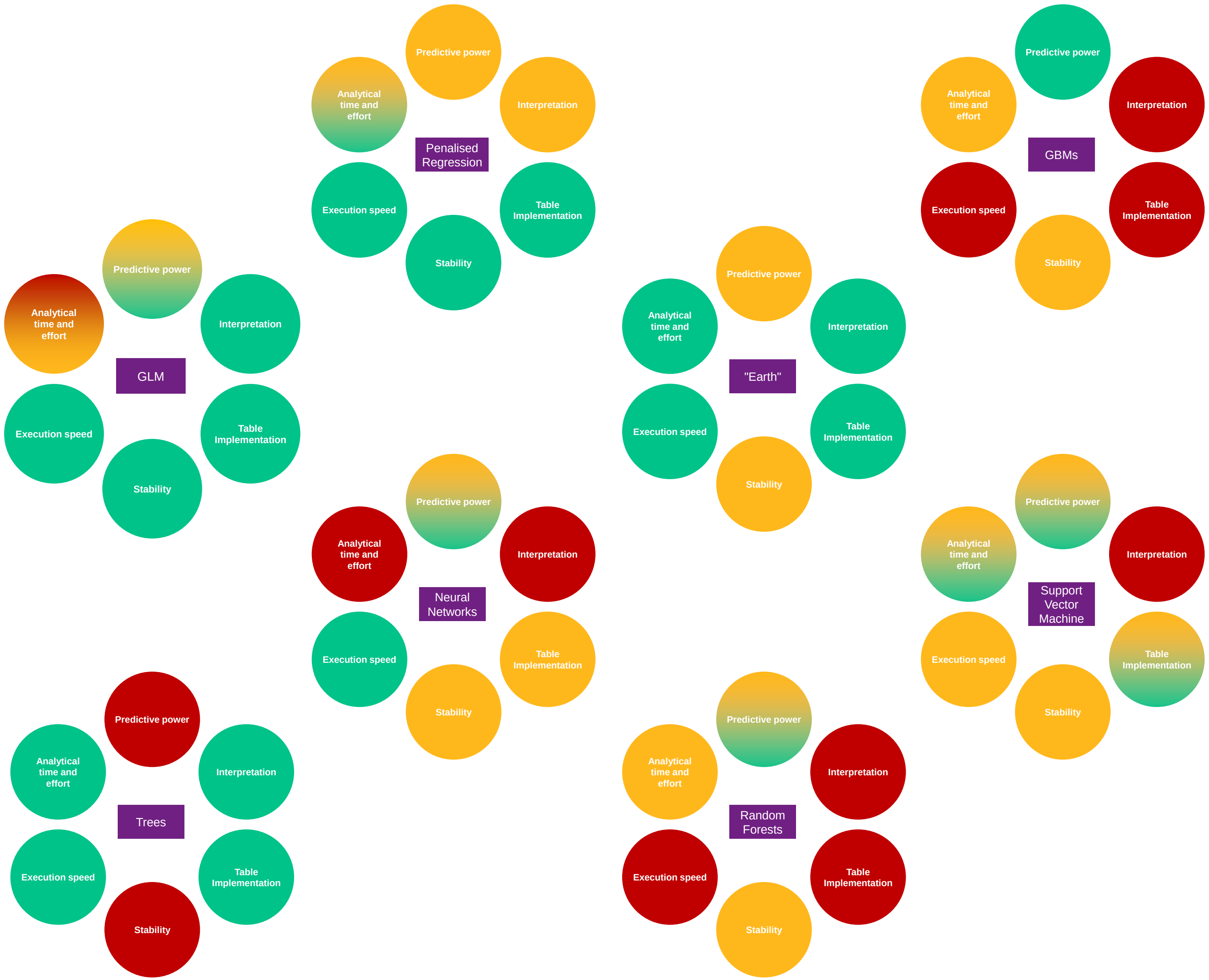
Dimensions of utility



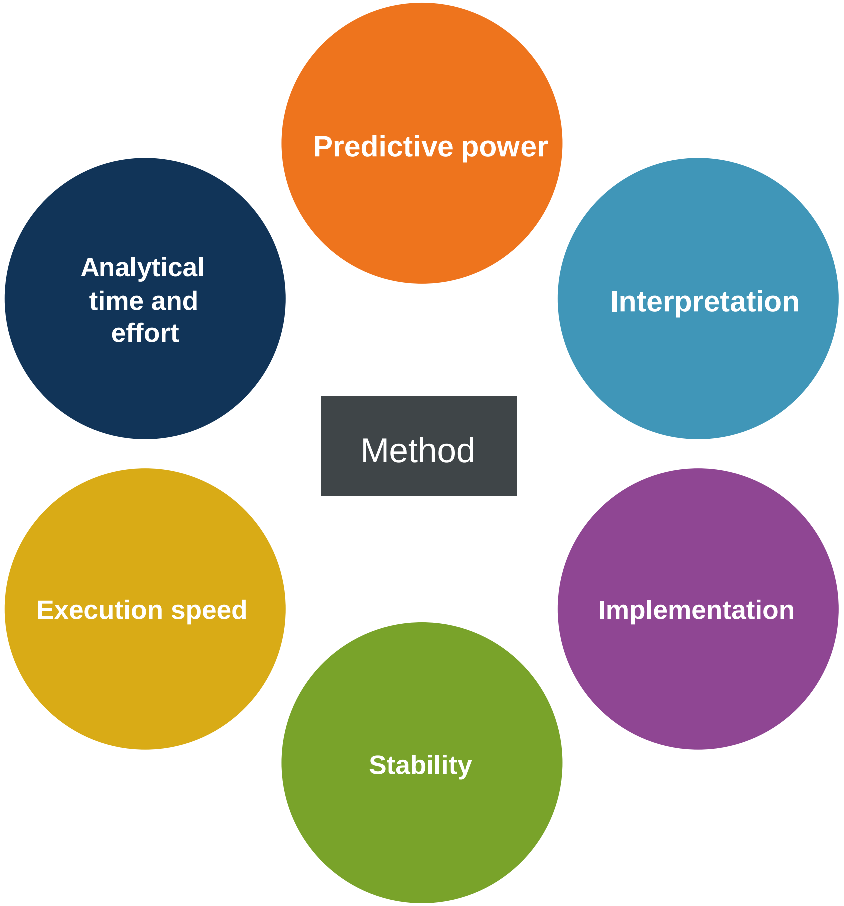


A toolkit…


Predictive power
Analytical time and effort
Interpretation
Method
Execution speed
Implementation
Stability

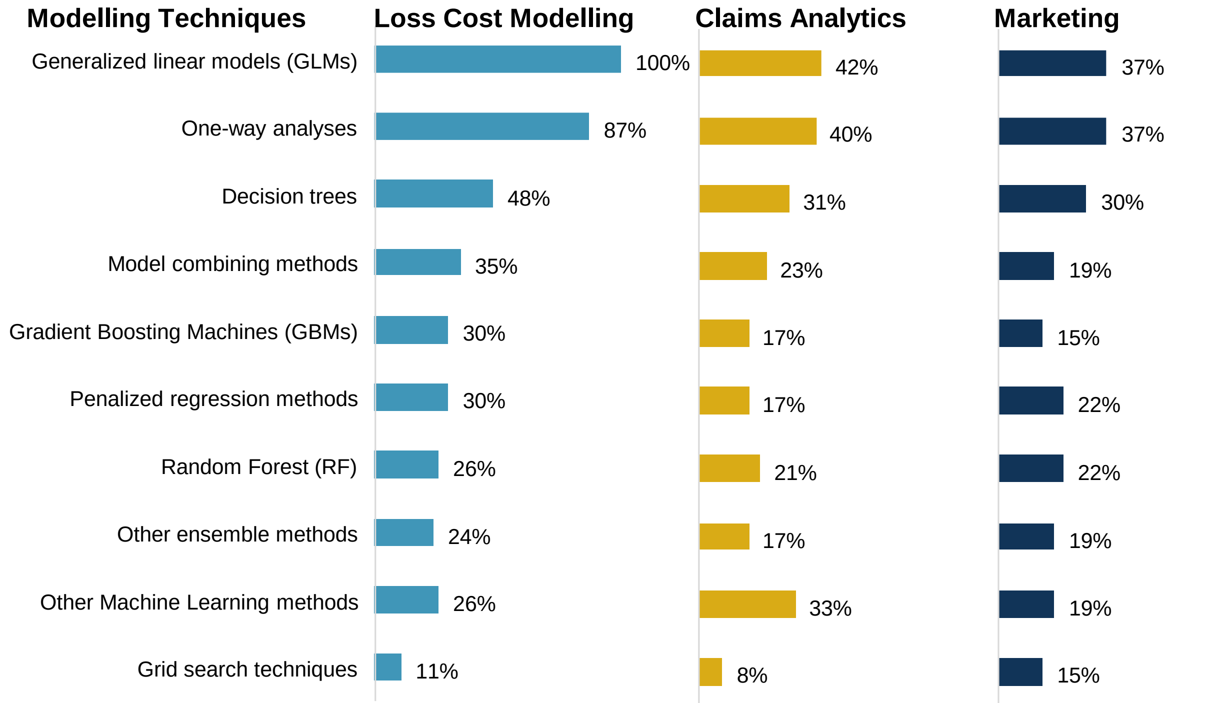

GLM
Predictive power
Analytical time and effort
Interpretation
Execution speed
Table Implementation
Stability
Penalised Regression
Predictive power
Analytical time and effort
Interpretation
Execution speed
Table Implementation
Stability
GBMs
Predictive power
Analytical time and effort
Interpretation
Execution speed
Table Implementation
Stability
"Earth"
Predictive power
Analytical time and effort
Interpretation
Execution speed
Table Implementation
Stability
Neural Networks
Predictive power
Analytical time and effort
Interpretation
Execution speed
Table Implementation
Stability
Support Vector Machine
Predictive power
Analytical time and effort
Interpretation
Execution speed
Table Implementation
Stability
Trees
Predictive power
Analytical time and effort
Interpretation
Execution speed
Table Implementation
Stability
Random Forests
Predictive power
Analytical time and effort
Interpretation
Execution speed
Table Implementation
Stability

…that is already in use



2016 US market survey

For which business applications do you use or plan to use these methods?



Willis Towers Watson Predictive Modeling Survey 2016

So…

Machine learning is already in use

Actuaries are already involved

It's not just about methods

Data beats models

It's not just about methods

Working out what to model matters

It's not just about predictiveness

A broader set of problems can be analysed - rapid basic insight adds value

Evolution not revolution

Models are complementary to existing methods

Issues for the Profession

Training

- A generation less familiar with stats?
- CAS, SOA ahead? (eg CSPA)
- GIRO too big now to help?
- IFoA on the case, but fast enough?

Role of the actuary

- Domain expertise matters (at least currently)
- Easier for an actuary to pick up machine learning than for a data scientist to understand insurance?
- Siloed teams don't work
- Familiarity and the right vernacular can help
- Scope of involvement?
  Pricing ✓ Reserving ✓ Claims analytics ✓
  Customer management ? Marketing ???

Regulatory issues

- TAS: Judgement - what judgement?
- GDPR
- FCA
- Government Select Committee (Science and Technology)

That spectrum of complexity

We can do this stuff

AI comprehension
Bespoke image recognition
Speech analytics
Machine learning predive modelling
Full autonomous driving
Object recognition
Topic modelling
Automated GLMs

This end could be interesting…

1_Ferrari.jpg

ferrari 24.15 %
nissan 13.55 %
jaguar 10.12 %
mclaren 7.32 %
ford 6.14 %
Elapsed: 2968

2_Ford.jpg

ford 79.46 %
dodge 6.30 %
chevrolet 5.08 %
am 2.15 %
gmc 1.81 %
Elapsed: 1422

3_LandRover.jpg

land 64.69 %
jeep 21.45 %
mazda 4.73 %
am 1.89 %
toyota 0.84 %
Elapsed: 1391

4_Mini.jpg

mini 68.86 %
chrysler 7.02 %
spyker 5.95 %
bmw 5.16 %
aston 3.39 %
Elapsed: 1390

09_Volkswagen.jpg

volkswagen 21.89 %
suzuki 12.35 %
acura 9.27 %
gmc 7.50 %
cadillac 7.09 %
Elapsed: 1406

5_Tesla.jpg

tesla 20.04 %
porsche 18.24 %
jaguar 8.88 %
lamborghini 6.47 %
honda 5.27 %
Elapsed: 1359

6_Ford.jpg

ford 22.99 %
honda 21.39 %
suzuki 10.98 %
hyundai 9.37 %
cadillac 8.37 %
Elapsed: 1390

7_Mini.jpg

mini 64.58 %
chrysler 15.28 %
buick 2.64 %
infiniti 2.01 %
ford 1.94 %
Elapsed: 1500

8_LandRover.jpg

land 63.36 %
audi 6.88 %
volvo 6.02 %
bmw 6.01 %
bentley 4.64 %
Elapsed: 1375

10_Mitsubishi.jpg

mitsubishi 33.07 %
chrysler 13.51 %
toyota 12.34 %
acura 7.06 %
buick 5.92 %
Elapsed: 1406

Institute and Faculty of Actuaries

Machine Learning

Duncan Anderson

GIRO 2017

19 October 2017