

CAS Machine Learning Working Party

Context and Key Issues in Ratemaking

Presenters

Marco De Virgilis - Marco is a Senior Actuarial Data Scientist at Allstate in Chicago. He is currently working on developing ML solutions to actuarial problems in both ratemaking and reserving.

Navarun Jain - Navarun is an analyst at Lux Actuaries & Consultants based in Dubai. As a Vice-Chair of the Working Party, Navarun leads the workstream on practical use cases/applications of machine learning in the actuarial field.

Nathaniel Loughin - Nate is a Director with KPMG's P&C Actuarial Practice, specializing in predictive analytics, large account pricing, and E&S Reserving and Operations.

Get the Code

Slides, code and data used in this presentation are available at:

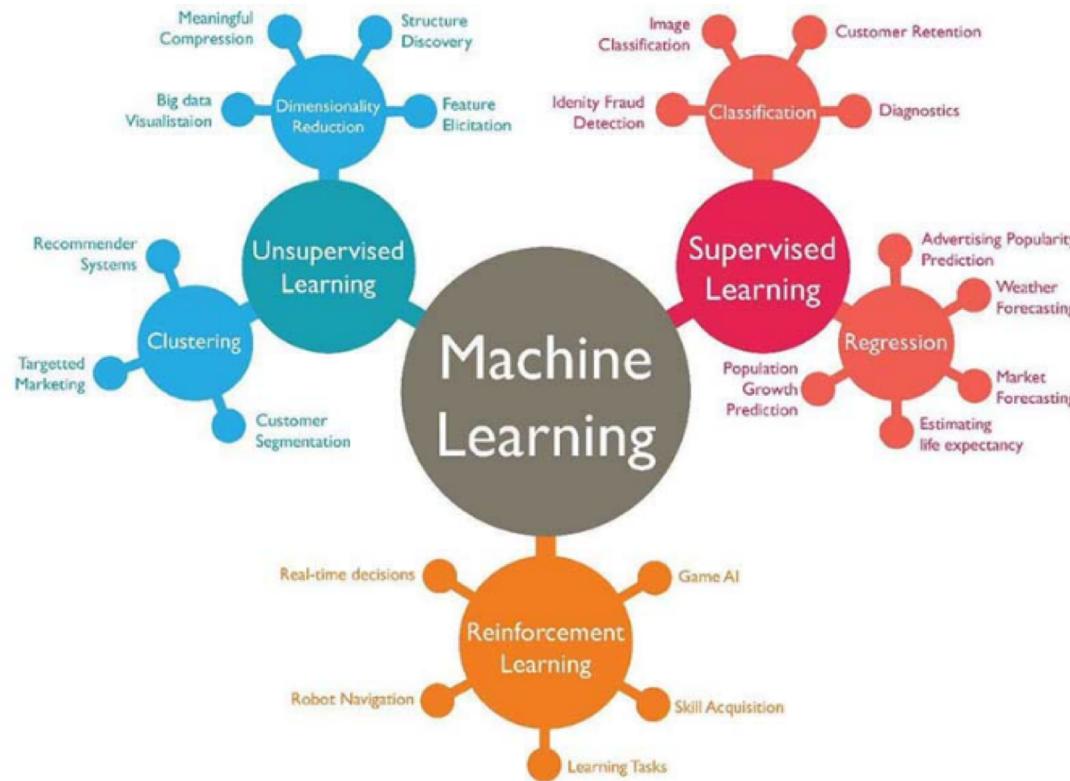
<https://github.com/mlwp3/CAMAR>

Introduction

What is Machine Learning?

- Catch-all term for a lot of concepts
- *Usually* involves a flexible algorithm that is *iteratively* adjusted based on optimizing some function of the data
 - E.g., take all the data, apply some transformations, and calculate how far you are from the answer you wanted, make adjustments, repeat
- Usually no closed-form solution to optimization problem, which necessitates iterative solutions
- Examples:
 - Computer vision
 - E-mail spam filtering
 - Netflix recommendations

What is Machine Learning?



Machine Learning Pros

- Good for open-ended problems (like computer vision) where it would be hard to manually engineer a model
 - Good for finding “hidden” relationships in data or selecting optimal subsets of predictors
- “On-line” learning and predicting possible
- Can fit highly non-linear functions that may be challenging for traditional approaches like GLMs
- Open-source software makes it easy!

Machine Learning Cons

- Not as transparent as statistical methods
- Not all statistical tools are available for evaluating model performance
- Can over-fit to data and create highly non-linear functions where you don't expect
- Computational cost - many of these models take a long time and a lot of computing power!

Why Should We Care About Machine Learning?

- It's cool, and it will make you cool
- It's going to be everywhere
- It can get much better results than more traditional models
- It can help explain results and identify patterns you might otherwise miss

Potential Applications to Ratemaking

- ML algorithms can enhance conventional models
- ML can enhance other insurance company functions
- ML can provide additional monitoring tools
- ML can enhance customer segmentation
- ML can expand profitability
- ...

Practical Applications

ML in action

The Data

freMPTL2 from R's *CASdatasets* package.

The data contains motor third-party liability policies from a French Insurer. Claim numbers and claim amounts, alongside a selection of risk features are available for analysis.

Variables

DRIVER

- Age
- Region
- Density

VEHICLE

- Age
- Brand
- Power
- Fuel Type

POLICY

- Exposure
- Bonus/Malus
- Claim Count
- Claim Amount

The Models

Models Considered

GLMs - The Classic Generalized Linear Model

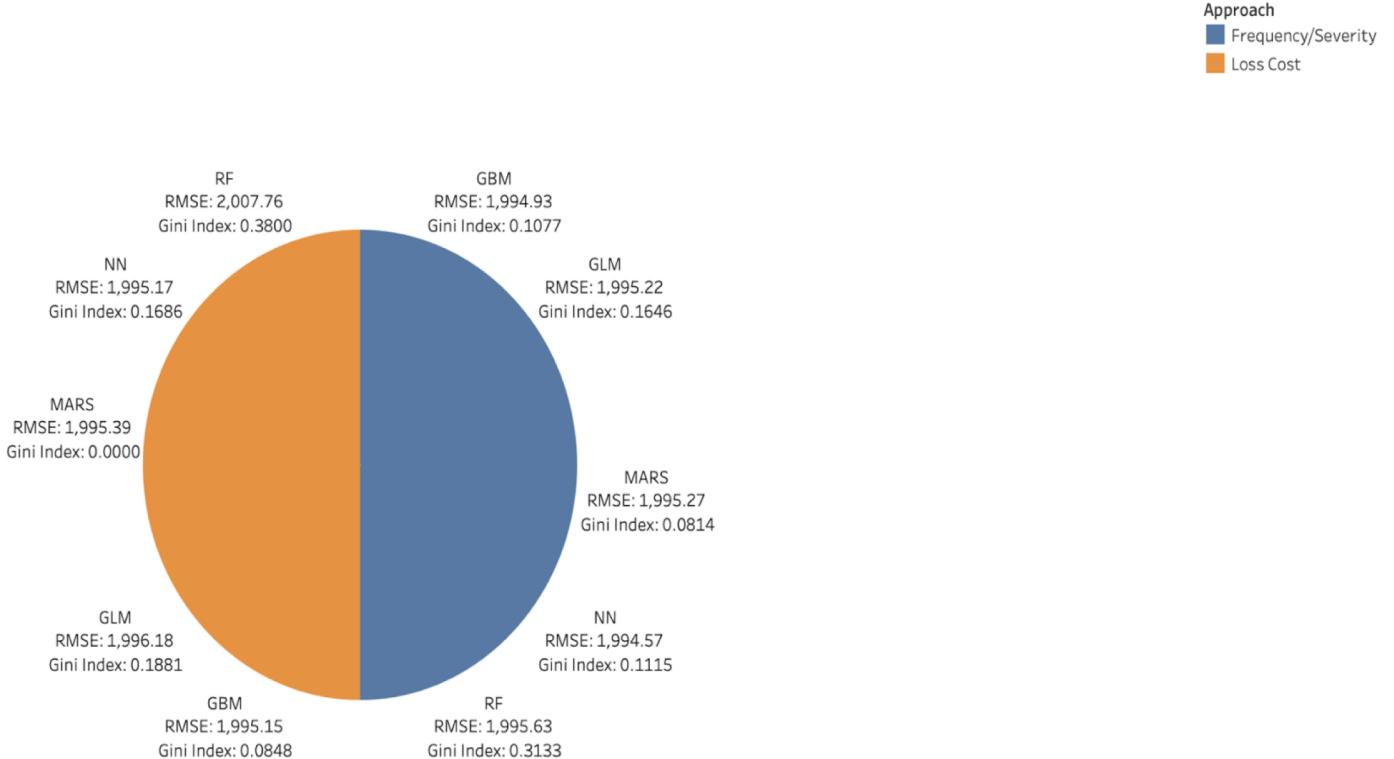
GBM - An approach that uses many weak predictors to generate robust estimates

NN - Layers of “neurons” that “learn” to reproduce desired output based on input

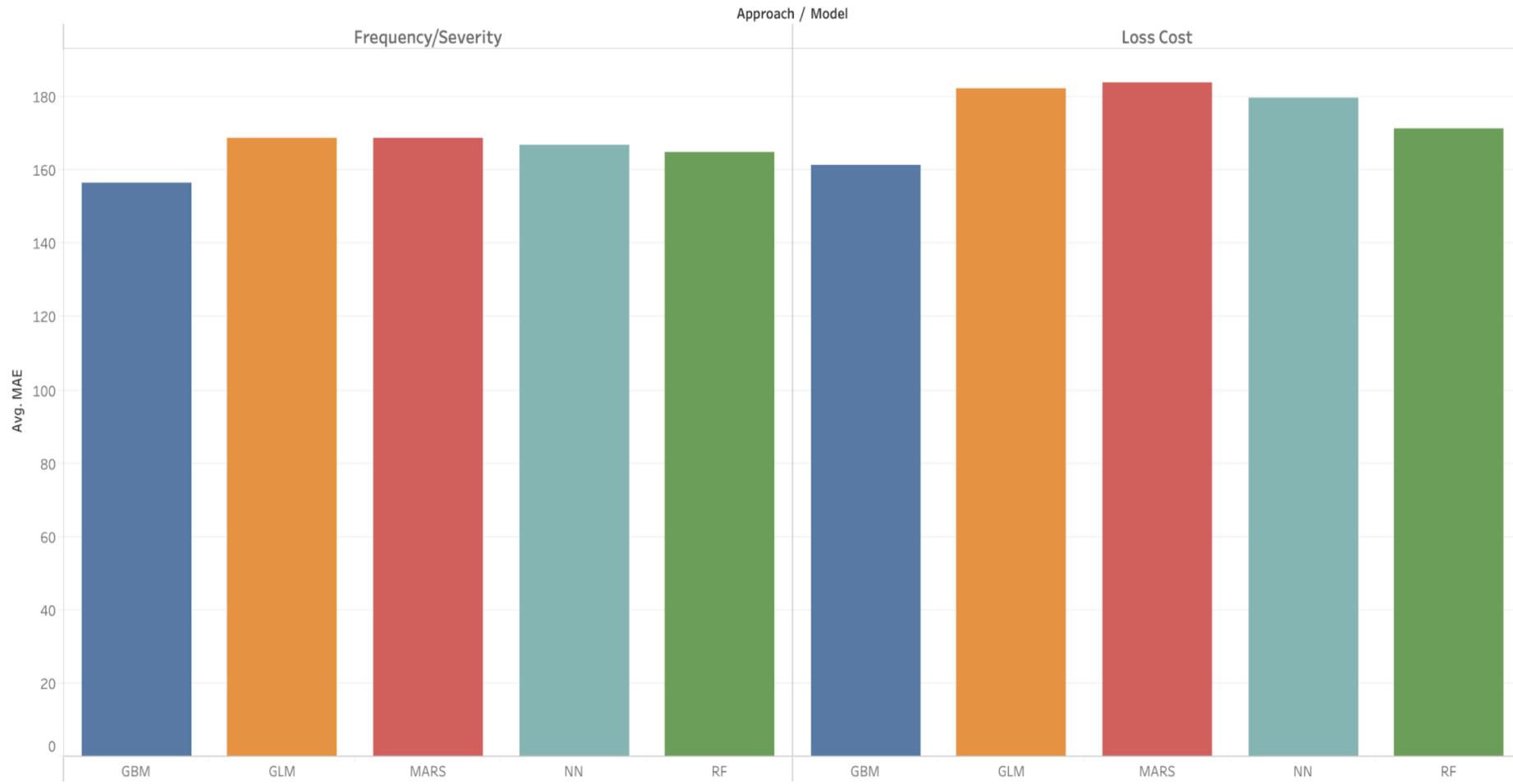
MARS - An automatic GLM that only uses linear splines

RF - A large number of big trees (vs GBMs which use small trees)

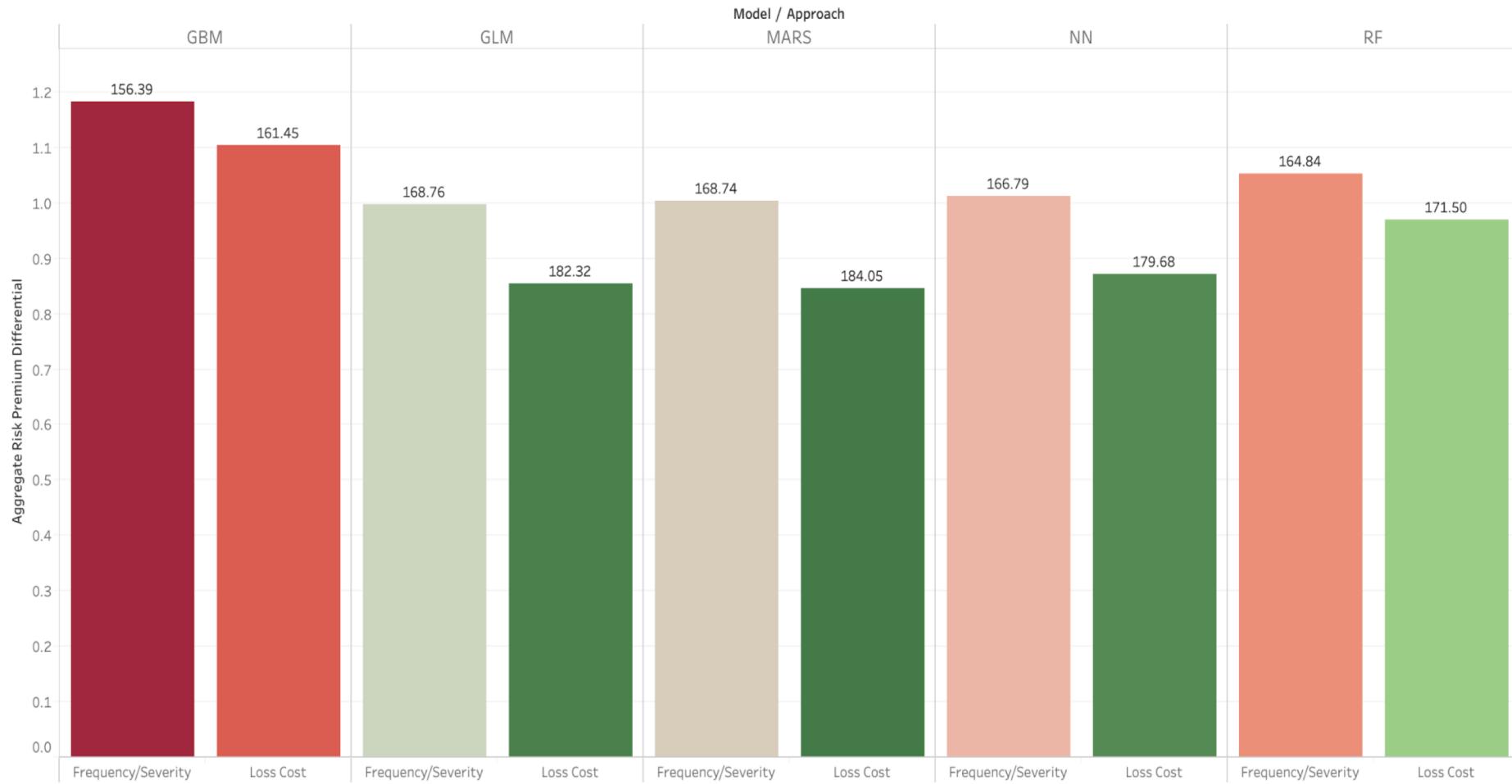
Models Considered



Comparison of Approaches across Models



Comparison of Models across Approaches



Communications Issues in ML

Towards Explainable AI (XAI)

Occam's Razor

The simplest explanation is usually the best

“...accuracy and simplicity (interpretability) are in conflict. For instance, linear regression gives a fairly interpretable picture of the x , y relation. But its accuracy is usually less than that of the less interpretable neural nets.”

L. Breiman

Start by Considering the Audience

- Technical Stakeholders
 - Other Actuaries
- Non-Technical Stakeholders
 - External
 - Regulators
 - Auditors
 - Internal
 - Profit Center Executives
 - Sales & Marketing
 - Agents & Insureds

ASOP 41 - Actuarial Communications

“...another actuary qualified in the same practice area could make an objective appraisal of the reasonableness...”

ASOP 41 - ML Issues

- The model includes the algorithm, data, hyperparameters, fitting methods
- ML is often “ad hoc” - many models are unique for their application
- ML algorithms and their underlying data are often proprietary

Regulators May Lack ML Capabilities

NAIC survey from 2017 indicates that:

- Not all states have personnel qualified to review GLMs
- Plurality of respondents note that filing complexity and/or lack of resources or expertise impeded their department's ability to review GLMs
- Not all states have an effective mechanism to protect confidentiality of models or other information submitted with a rate filing

Regulatory Issues

- Need to demonstrate that rates are not inadequate, excessive, or unfairly discriminatory
 - “Unfairly discriminatory” may be a challenge unless we can explain why a model produces a particular outcome.
- Need to file a rating plan
 - Does a black box meet the legal definition of a “filed rate”?
 - Is it necessary to convert the ML model to relativities for implementation?

Internal Communications

- Is the price change consistent with the corporate strategy and messaging?
- How do we explain the change to our management?
- Will our agents be able explain the change to their insureds?
- What do you say to insured whose premium changes because the model changed?
- Who will be impacted the most?

Bridging the Communication Gap

Basic Idea

ML can be a black box - let there be light!

MODEL INTERPRETATION

GLOBAL

Trying to understand the predictions on an *overall* level – ***In general, why does a model behave the way it does?***

LOCAL

Trying to understand predictions for *specific records* – ***For a given record, what led the model to predict what it did?***

Global Interpretation Strategies

TECHNICAL

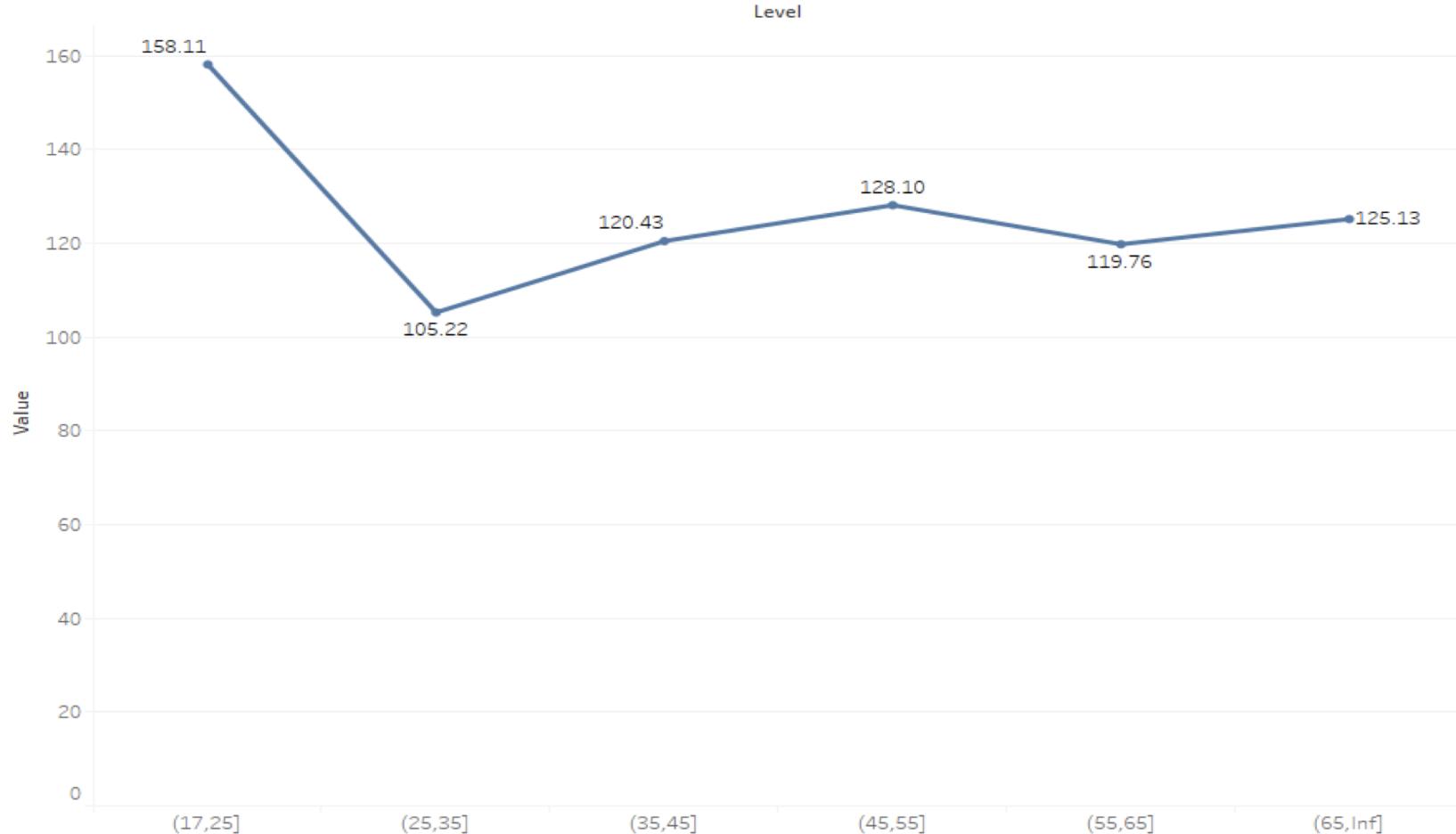
- Variable Importance
- Interaction Effect Analysis
- Feature Effect Analysis
 - Model Lift
- Gini Index/Gini Plot

NON-TECHNICAL

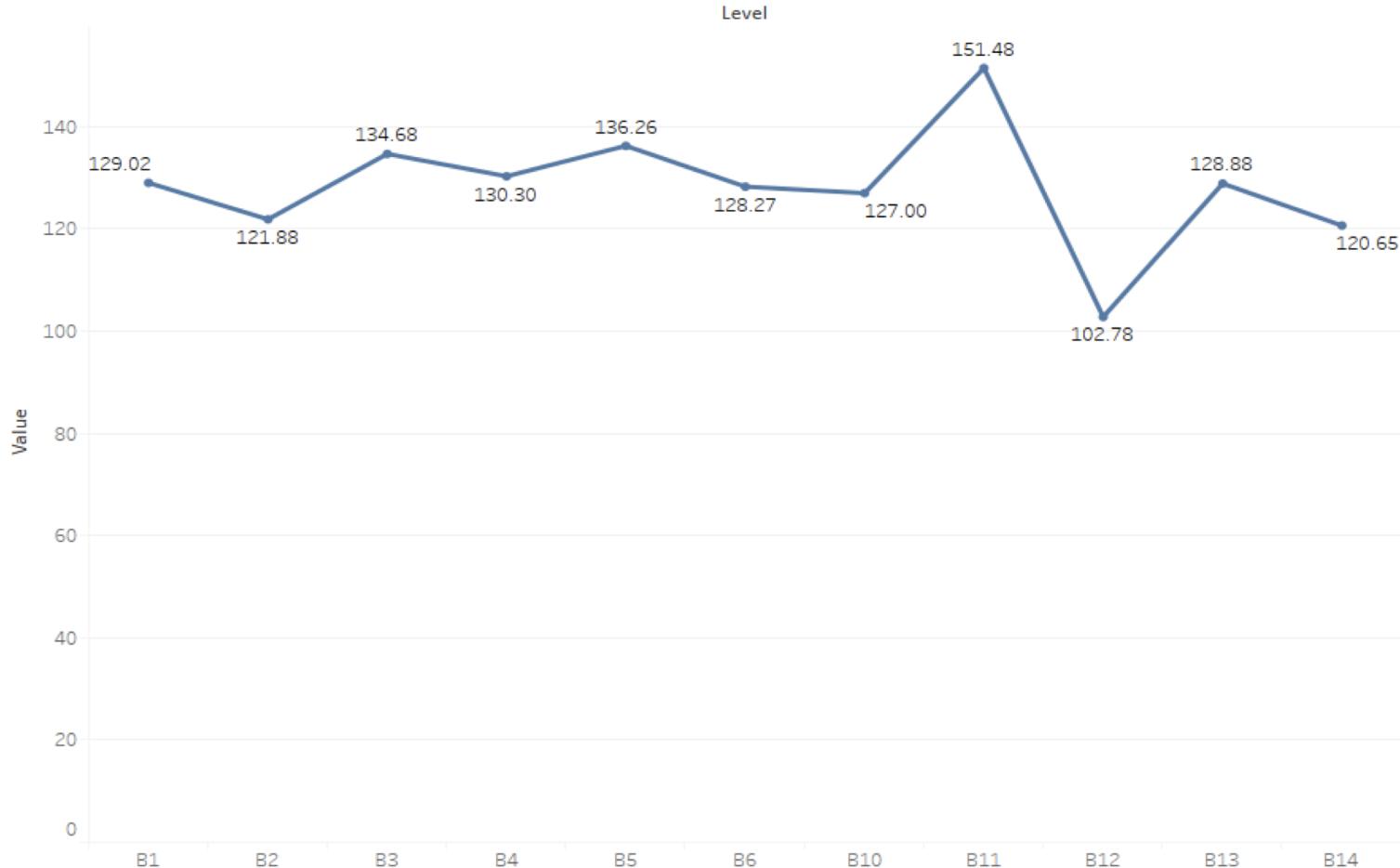
- Partial Dependence Plots

Partial Dependence Plots

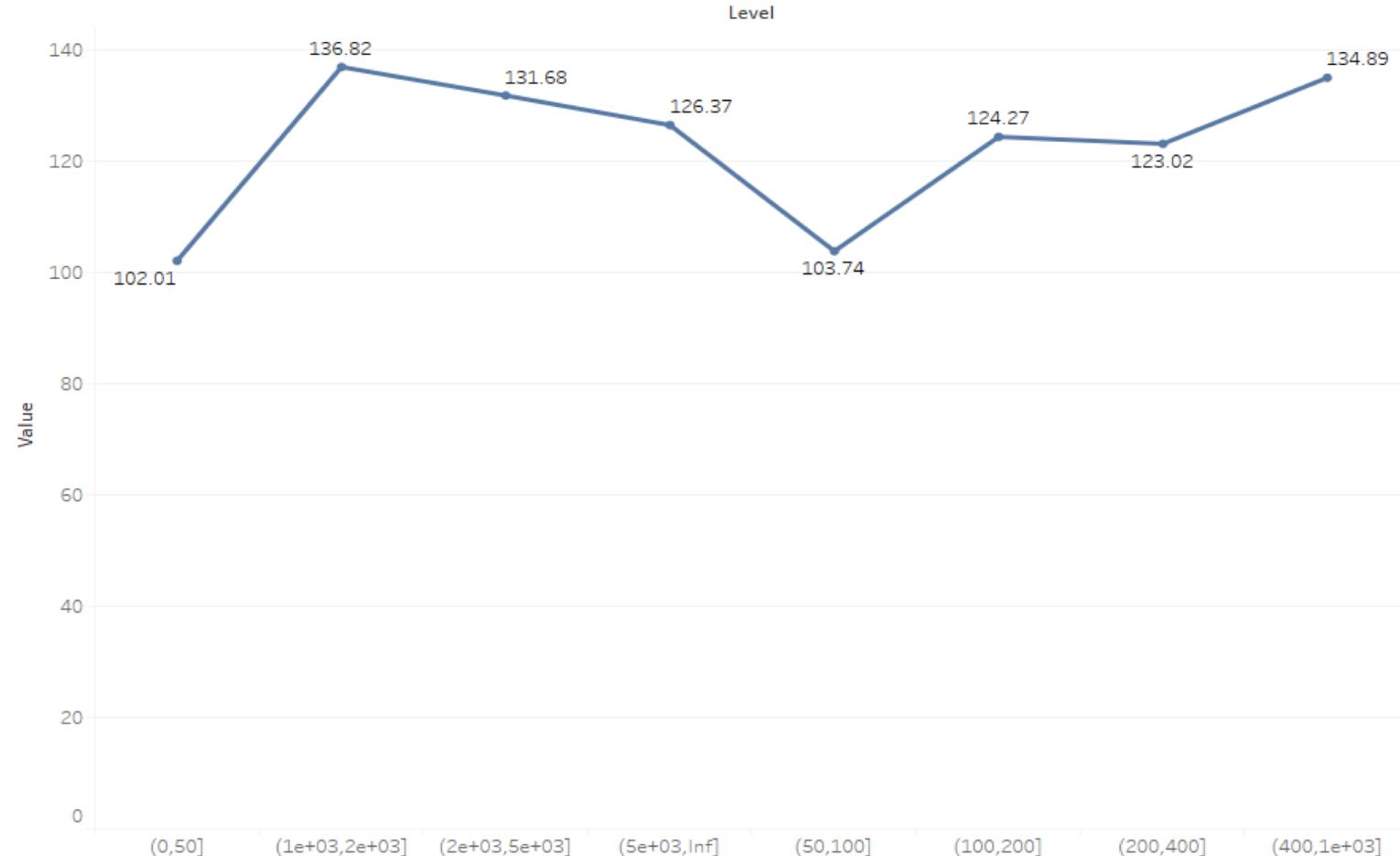
Partial Dependence - DrivAgeBand



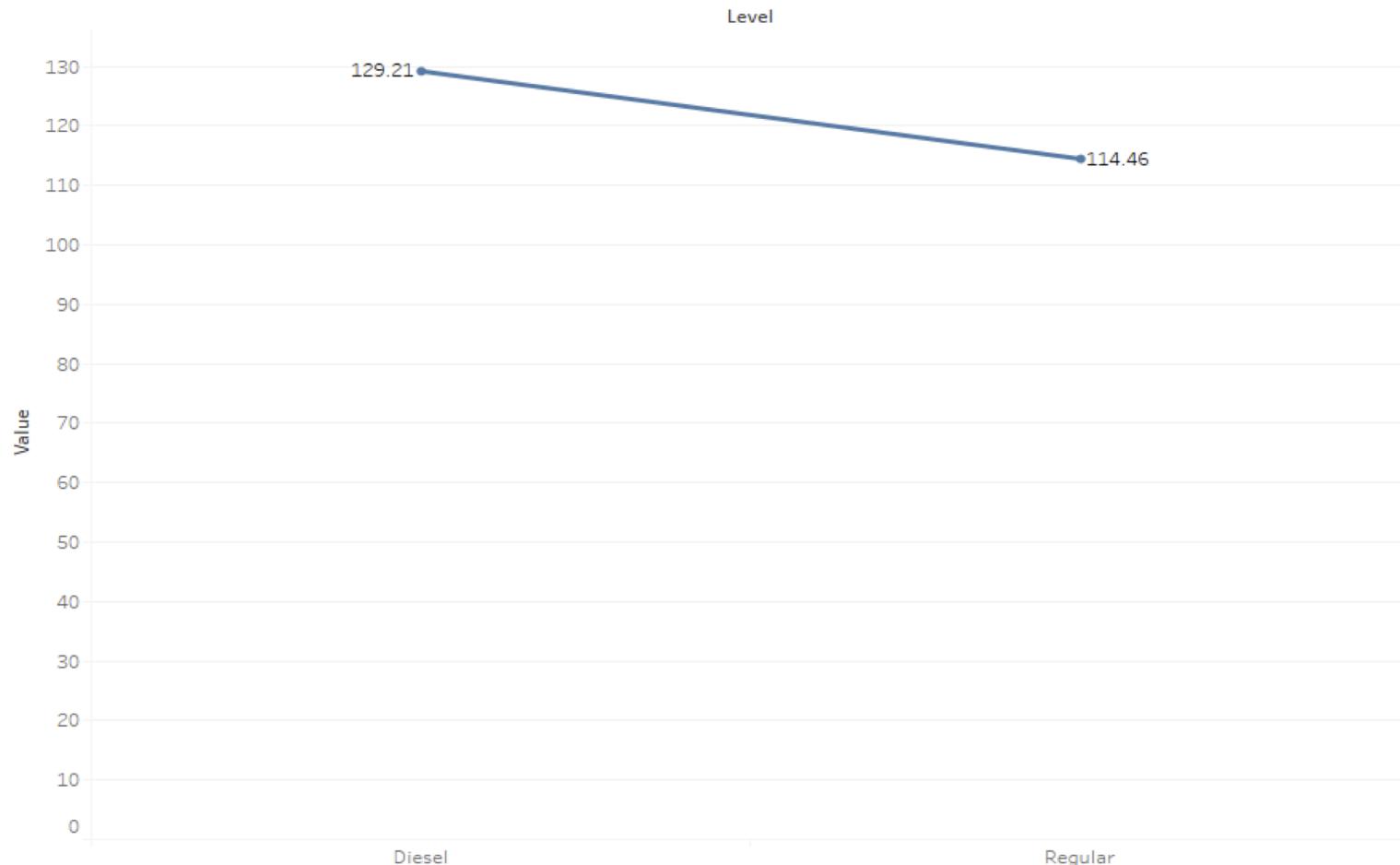
Partial Dependence - VehBrand



Partial Dependence - DensityBand



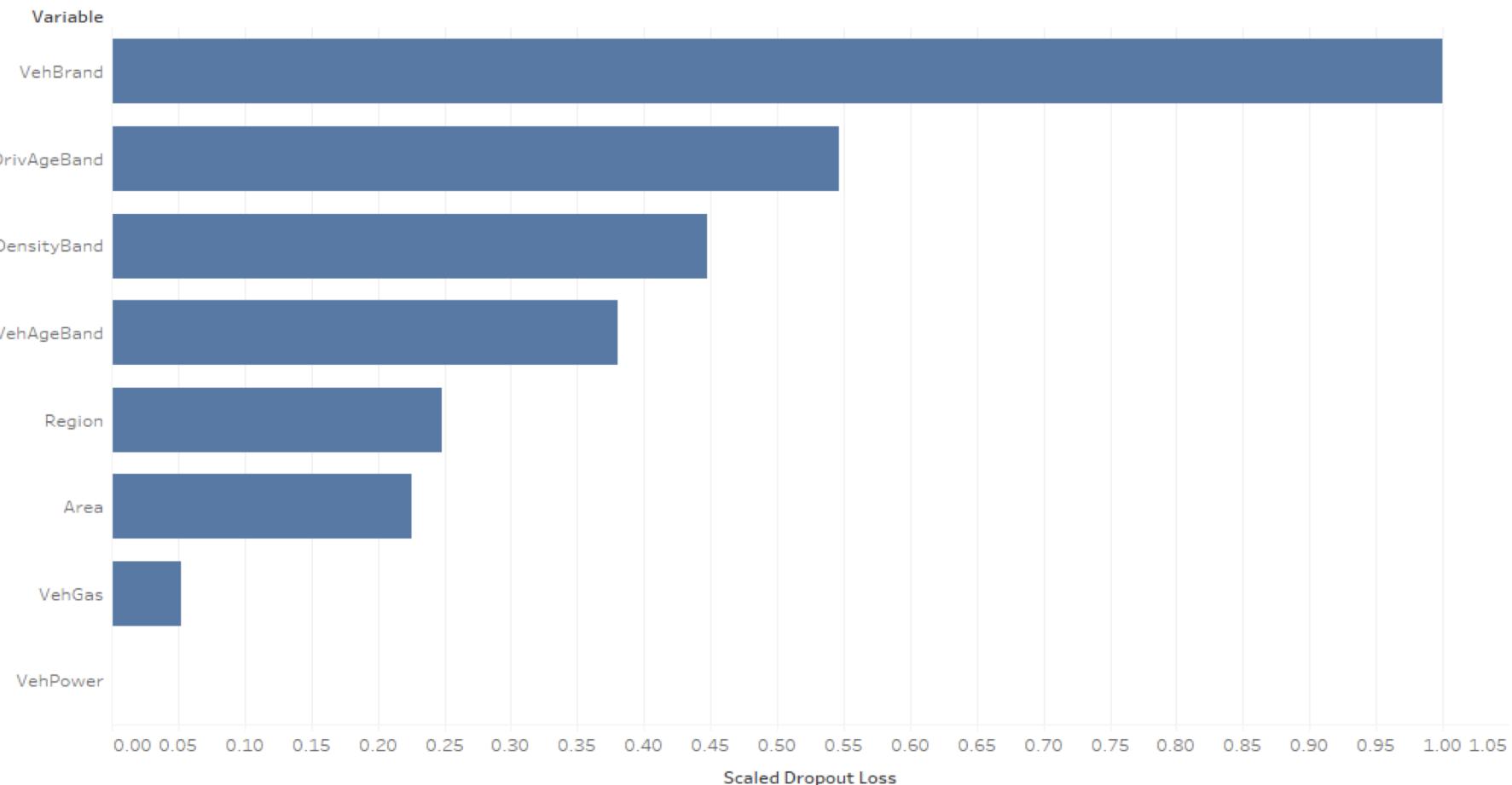
Partial Dependence - VehGas



Variable Importance

- Based on Permutation-based Loss Dropout
- Each rating variable is shuffled and model recomputed
- Degree of difference in RMSE w.r.t. original model indicates variable importance

Variable Importance

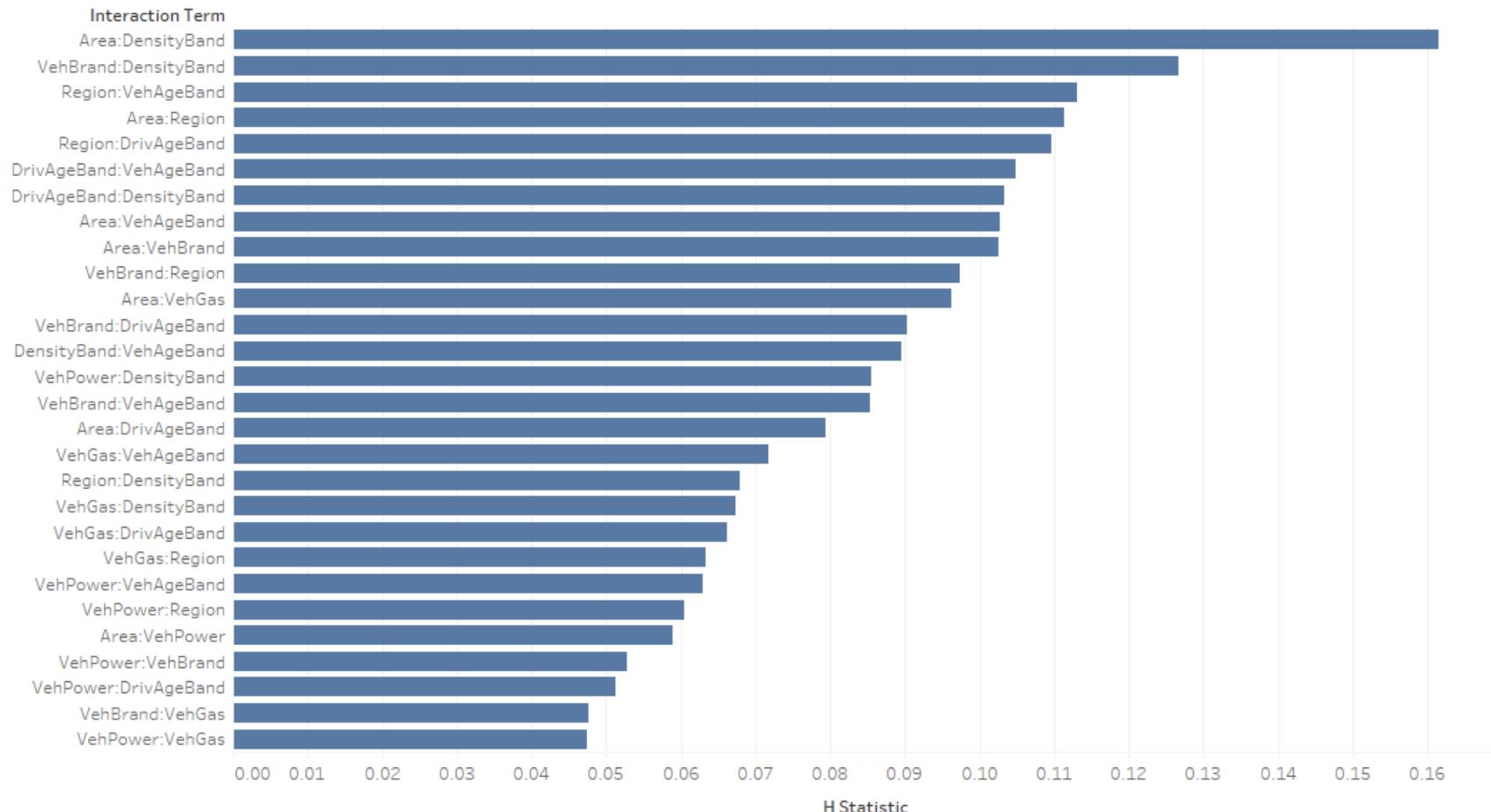


Interaction Effects

- Based on Partial Dependence (PD) - studies how model predictions depend on individual predictors
- Uses the Friedman H-Statistic
- Measures the degree of impact the joint PD of 2 variables has on the overall PD of the combination, intuitively,

$$PD(X, Y) = PD(X) + PD(Y) + PD(X \& Y)$$

Interaction Effects



Non-Technical Communication Strategies

- For the rating plan, the model must be converted to relativities.
 - Tools such as Lime may be needed to generate the relativities.
- ML can replace “judgement” in some rating plan components. For example:
 - Clustering used in a classification analysis
 - AI used to generate a brush-fire hazard map
- Rating examples help stakeholders can get a “feel” for what the model does.

References

- Henckaerts, Roel and Antonio, Katrien and Clijsters, Maxime and Roel, Verbelen, A Data Driven Binning Strategy for the Construction of Insurance Tariff Classes (May 12, 2017). Available at SSRN: <https://ssrn.com/abstract=3052174> or <http://dx.doi.org/10.2139/ssrn.3052174>
- Beuerlein, B. et al. Big Data and the Role of the Actuary. American Academy of Actuaries Big Data Task Force, June 2018
- Kuo, Kevin. DeepTriangle: A Deep Learning Approach to Loss Reserving. Risks 7.3 (2019): 97.
- Dai, Jie. Enhancing the Generalized Linear Modeling Approach with Machine Learning Technique. E-Forum, Casualty Actuarial Society, Spring 2018, vol. 2.
- Lally, Nathan & Hartman, Brian, 2018. Estimating loss reserves using hierarchical Bayesian Gaussian process regression with input warping, Insurance: Mathematics and Economics, Elsevier, vol. 82(C), pages 124-140.
- Jamal, S., et al. Machine Learning & Traditional Methods Synergy in Non-Life Reserving. Report of the ASTIN Working Party of the International Actuarial Association, 2018
- Wuthrich, Mario V., Machine Learning in Individual Claims Reserving (November 11, 2016). Swiss Finance Institute Research Paper No. 16-67. Available at SSRN: <https://ssrn.com/abstract=2867897> or <http://dx.doi.org/10.2139/ssrn.2867897>
- Spedicato, Giorgio and Dutang, Christophe and Petrini, Leonardo, Machine Learning Methods to Perform Price Optimization: A Comparison with Standard Generalized Linear Models. Variance, Casualty Actuarial Society, 2018, 12 (1), pp. 69-90. hal-01942038
- Gabrielli, Andrea and Richman, Ronald and Wuthrich, Mario V., Neural Network Embedding of the Over-Dispersed Poisson Reserving Model (November 21, 2018). Available at SSRN: <https://ssrn.com/abstract=3288454> or <http://dx.doi.org/10.2139/ssrn.3288454>

References, Continued

Wuthrich, Mario V., Neural Networks Applied to Chain-Ladder Reserving (July 6, 2018). Available at SSRN: <https://ssrn.com/abstract=2966126> or <http://dx.doi.org/10.2139/ssrn.2966126>

Guelman L., Guillén M., Pérez-Marín A.M. (2012) Random Forests for Uplift Modeling: An Insurance Customer Retention Case. In: Engemann K.J., Gil-Lafuente A.M., Merigó J.M. (eds) Modeling and Simulation in Engineering, Economics and Management. MS 2012. Lecture Notes in Business Information Processing, vol 115. Springer, Berlin, Heidelberg

Yeo, A. C., Smith, K. A., Willis, R. J., & Brooks, M. (2001). Modelling the effect of premium changes on motor insurance customer retention rates using neural networks. *Lecture Notes in Computer Science*, 2074, 390 - 399. https://doi.org/10.1007/3-540-45718-6_43

Salcedo-Sanz, S., DePrado-Cumplido, M., Segovia-Vargas, M.J., Pérez-Cruz, F. and Bousoño-Calzón, C. (2004), Feature selection methods involving support vector machines for prediction of insolvency in non-life insurance companies. *Intelligent Systems in Accounting, Finance and Management*, 12: 261-281. doi:10.1002/isaf.255

Ye Tian, Wei Yang, Gene Lai, Menghan Zhao. Predicting non-life insurer's insolvency using non-kernel fuzzy quadratic surface support vector machines. *Journal of Industrial & Management Optimization*, 2019, 15 (2) : 985-999. doi: 10.3934/jimo.2018081

Kartasheva, Anastasia V. and Traskin, Mikhail, Insurers' Insolvency Prediction Using Random Forest Classification (December 7, 2013). Available at SSRN: <https://ssrn.com/abstract=2364736>

Andreas Behr & Jurij Weinblat (2017) Default Patterns in Seven EU Countries: A Random Forest Approach, *International Journal of the Economics of Business*, 24:2, 181-222, DOI: 10.1080/13571516.2016.1252532

Nilsson, M., & Sandberg, E. (2018). Application and Evaluation of Artificial Neural Networks in Solvency Capital Requirement Estimations for Insurance Products.

References, Continued

De Virgilis, M., Pierluigi C., Estimation of Individual Claim Liabilities. Casualty Actuarial Society, 2020. Available at
<https://www.casact.org/research/wp/papers/working-paper-Virgilis-Cerqueti-2020-01.pdf>

ASTIN, Individual Claim Development with Machine Learning, 2017 Report. Available at
http://www.actuaries.org/ASTIN/Documents/ASTIN_ICDML_WP_Report_final.pdf

Jain, N., Machine Learning: “Pricing” the Way Forward. Institute & Faculty of Actuaries (IFoA) TIGI 2019; available at
https://www.actuaries.org.uk/system/files/field/document/Pricing%20Plenary%202_Navarun%20Jain.pdf

Jain, N., Fraud Detection: How can Machine Learning Help? IFoA GIRO Conference 2019; available at
https://www.actuaries.org.uk/system/files/field/document/A5_Navarun%20Jain.pdf

ML Working Party Members

Daniel Lupton (Chair)

Navarun Jain (Vice Chair)

Peter Dyson

Harsh Karthik

Liam McGrath

Marjan Qazvini

Les Vernon

Nathaniel Loughlin (Vice Chair)

Ralph Dweck

James Ely

Hao Li

Sharon Mott

Seth Roby

Marco De Virgilis