auditor: visual validation of predictive models

Alicja Gosiewska[1], Agnieszka Ciepielewska[1,2,3], Przemysław Biecek[1,2]

[1]Faculty of Mathematics and Information Science, Warsaw University of Technology

[2]Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw

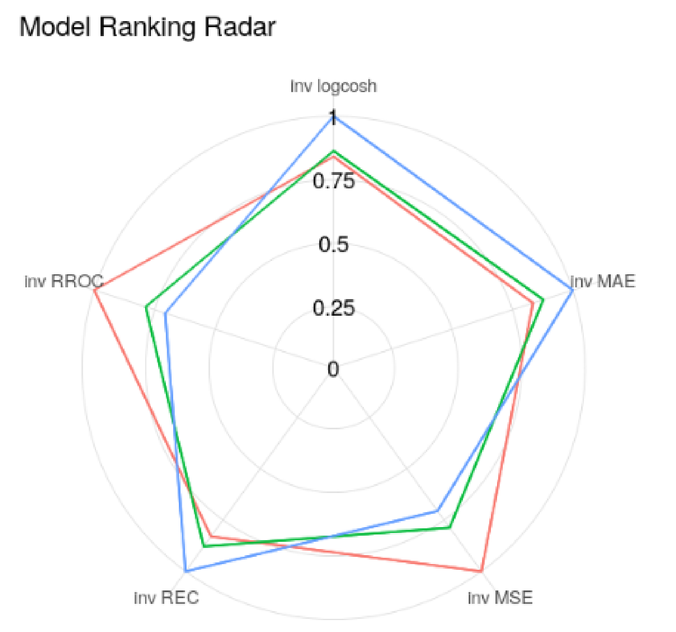
[3]Faculty of Physics, University of Warsaw

Introduction

The auditor is an R package for model-agnostic diagnostics. Implemented techniques facilitate assessing and comparing the goodness of fit and performance of models. In addition, they may be used for analysis of the similarity of residuals and for identification of outliers and influential observations. Due to flexible and consistent interface, it is simple to validate models of any class.

Comparison of models performance

The Model Ranking Plot compares models performance across multiple loss functions. This plot consists of two parts. On the left side there is a radar plot. Colors correspond to models, vertices to values of loss functions which are inverted and rescaled to [0, 1].
Models with better performance (larger inverted loss value) are closer to the edge. On the right side of the plot is a table with values of loss functions. The fourth column contains loss scaled to one of the models.



name	label	values	scaled
MAE	lm	3.33e+01	1.000
MAE	svm	3.17e+01	1.052
MAE	rf	2.78e+01	1.198
MSE	lm	1.66e+03	1.000
MSE	svm	2.10e+03	0.787
MSE	rf	2.34e+03	0.706
REC	lm	3.33e+01	1.000
REC	svm	3.15e+01	1.058
REC	rf	2.76e+01	1.204
RROC	lm	3.31e+09	1.000
RROC	svm	4.21e+09	0.787
RROC	rf	4.69e+09	0.706
logcosh	lm	2.29e+01	1.000
logcosh	svm	2.23e+01	1.027
logcosh	rf	1.93e+01	1.189

Figure 1: The linear model (red) has better performance in aspect of MSE and RROC, while random forest (blue) is better in aspect of logcosh, REC, and MAE.

Evaluation of error tolerance

Regression Error Characteristic (REC) curve is a generalization of Receiver Operating Characteristic (ROC) curve for regression. REC curve estimates the Cumulative Distribution Function of the error.
On the x axis of the plot there is an error tolerance. On the y axis there is an accuracy at the given tolerance level. The accuracy at tolerance ϵ is a percentage of observations predicted within the tolerance ϵ. In other words, residuals larger than ϵ are considered as errors.

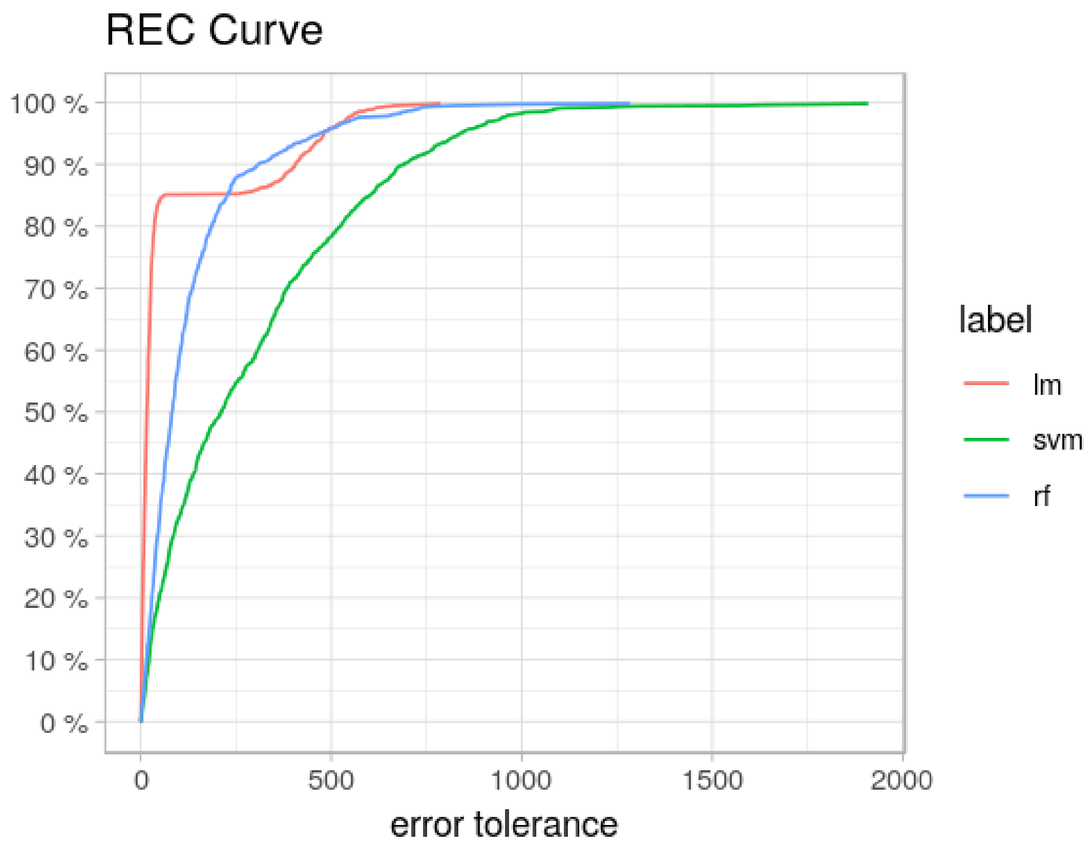


Figure 2: The linear model (red) has higher accuracy if we accept residuals lower than 200, but for tolerance level between 250 and 500 the random forest model (blue) has better performance.

Comparison of distributions of residuals

On the Residual Density plot, there are estimated densities of the models' residuals. This plot simplifies detecting the incorrect behavior of the residuals. If the residuals are not concentrated around zero, it is likely that the model predictions are biased.

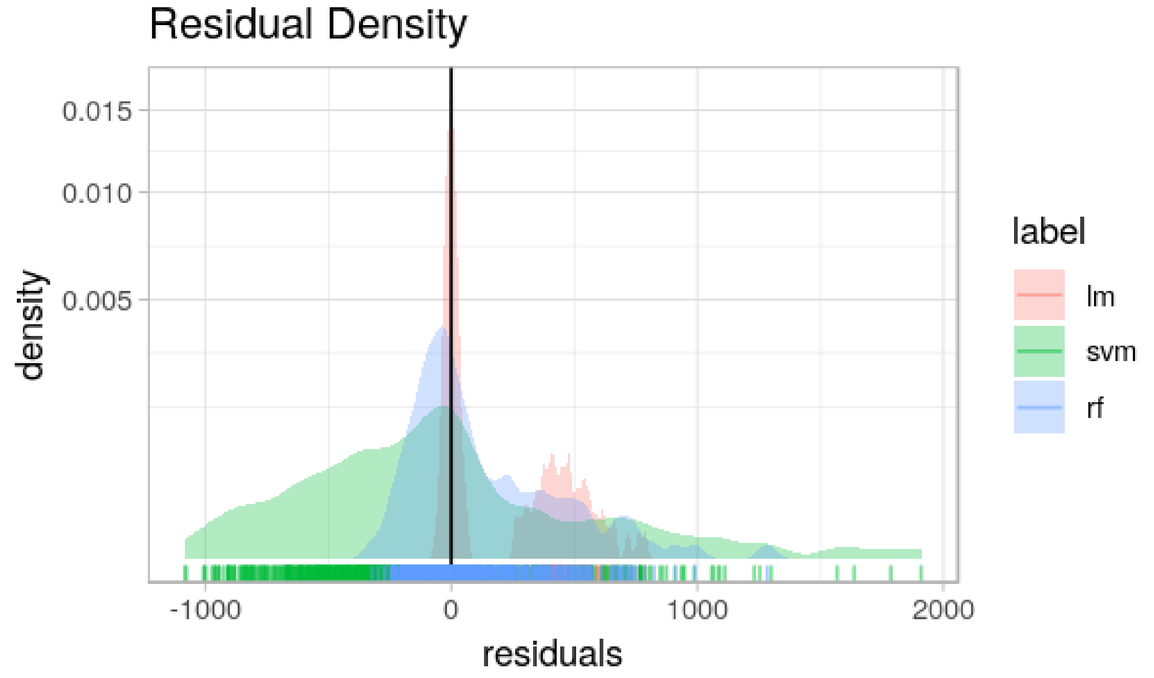


Figure 3: The variance of the svm and random forest residuals (green and blue) is much higher than for linear model (red). However the distribution for linear model is bimodal.

Comparison of predictions

Predicted Response Plot is a scatter plot of predicted values on the y axis against observed response on the x axis. Ideally, all points should be close to a black diagonal line.

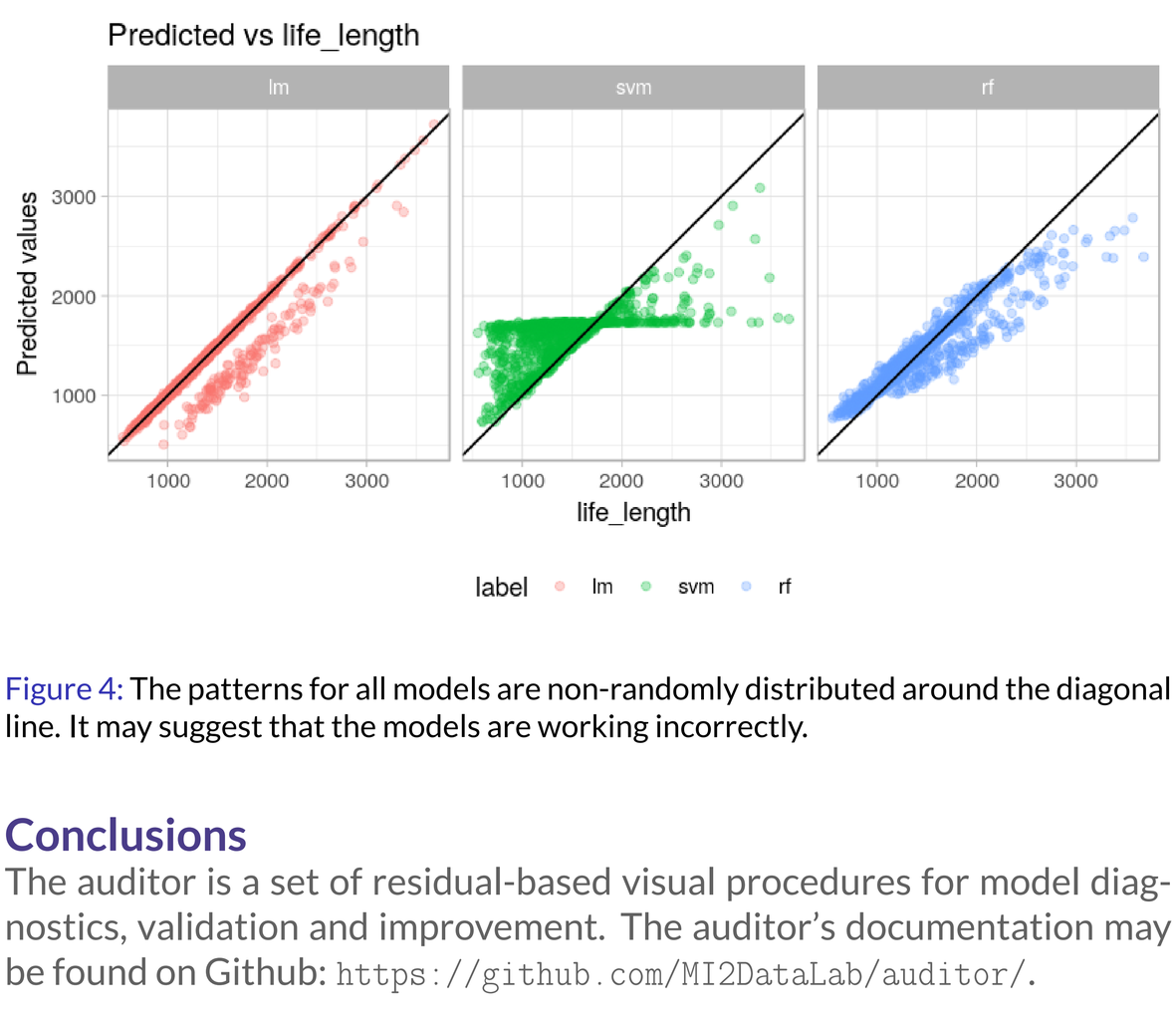


Figure 4: The patterns for all models are non-randomly distributed around the diagonal line. It may suggest that the models are working incorrectly.

Conclusions

The auditor is a set of residual-based visual procedures for model diagnostics, validation and improvement. The auditor's documentation may be found on Github: `https://github.com/MI2DataLab/auditor/`.

References

[1] H. Beyer. Tukey, john w.: Exploratory data analysis. addison-wesley publishing company reading, mass. — menlo park, cal., london, amsterdam, don mills, ontario, sydney 1977, XVI, 688 s. *Biometrical Journal*, 23(4):413–414, 1981. doi: 10.1002/bimj.4710230408. URL `https://doi.org/10.1002/bimj.4710230408`.

[2] Alicja Gosiewska and Przemyslaw Biecek. auditor: an R Package for Model-Agnostic Visual Validation and Diagnostic. *ArXiv e-prints*, art. arXiv:1809.07763, September 2018.

Acknowledgements

Work of Agnieszka Ciepielewska was financially supported by NCN Opus grant 2016/21/B/ST6/02176 while Alicja Gosiewska was supported by NCN Opus grant 2017/27/B/ST6/01307.

- ✉ `alicjagosiewskal@gmail.com`
- ✉ `agnieszkac27@gmail.com`