Responsible Prediction Making of COVID-19 Mortality (Student Abstract)

Hubert Baniecki[1], Przemyslaw Biecek[1, 2]

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

[2] Samsung Research and Development Institute, Poland

The Thirty-Fifth AAAI Conference on Artificial Intelligence, February 2-9, 2021

Responsible Prediction Making

In recent years we have seen a growing interest in the area of Responsible AI [Barredo Arrieta et al. 2019]. These concepts build upon research related to transparency, robustness and explainability of machine learning models; also an area of fairness, bias and accountability applied to the process of prediction making. Various effective methods were developed for model analysis. Unfortunately, we observed that they are not being used in such critical and sensitive domains as COVID-19 predictive modelling.

A broad overview of 145 predictive models for prognosis and diagnosis of COVID-19 is a starting point for our discussion [Wynants et al. 2020]. Many of the proposed models are so-called black-boxes, complex models like neural networks or tree ensembles, aiming for the best performance while overlooking interpretability and explanation of their reasoning.

Unfortunately, after reviewing these contributions, we concluded that little effort is being put into reassuring model robustness and transparency, especially for such a human-centred topic. Desired works exemplary could combine: using high-quality data moreover presenting its in-depth context, examining performance measures critically, providing complete documentation of the model, or at least model explanations.

Methods

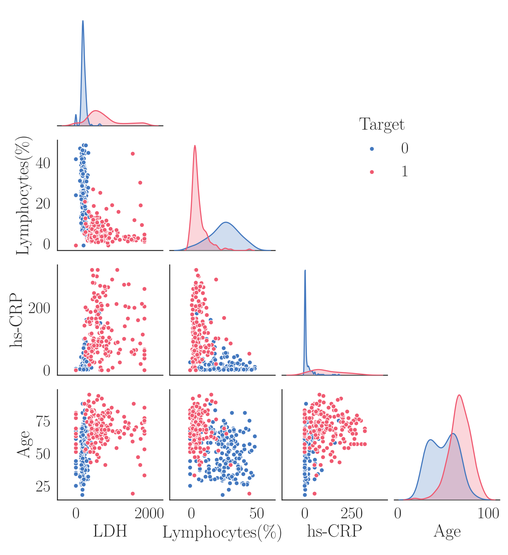
We show how to advance the current state-of-the-art black-box predictive models into the new responsible standards, by applying Interactive Explanatory Model Analysis (IEMA) implemented in *modelStudio* [Baniecki and Biecek 2019, 2020].

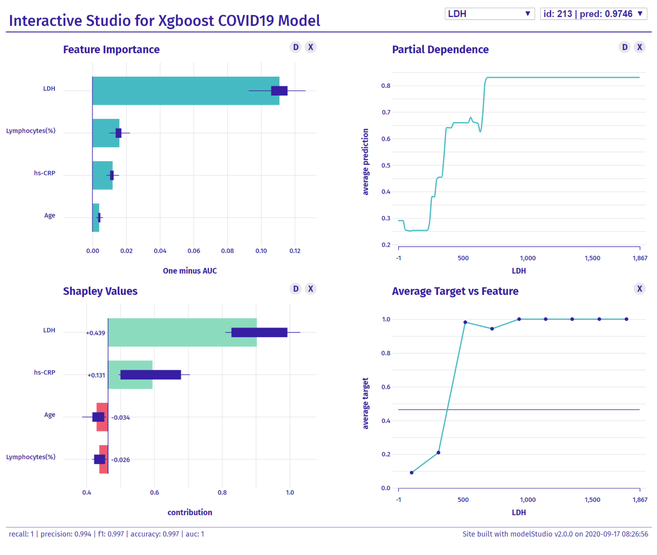
Use-case: An Interpretable Mortality Prediction Model for COVID-19 Patients

With context in mind, the work *"An interpretable mortality prediction model for COVID-19 patients"* of Yan et al. [2020] addresses an important issue of need, and with its recent popularity may be called state-of-the-art. Exploration of multiple XGBoost models that predict patients' mortality rates leads to the development of an interpretable decision tree. Analysing the results of the discussed work has led to multiple wh-questions about the data and models that we deem are necessary to address for an effective, and at the same time responsible, COVID-19 mortality prediction.

- **Why is LDH such a critical variable?** Yan and co-authors present performances of the Multi-tree XGBoost models constructed on three sets of variables. Nevertheless, we can observe that even a model with only one variable performs well (AUC over 0.90). The question arises due to its outstanding significance.
- **Why is age not used in the prediction making?** Multiple studies indicate that age has the potential to be a valuable mortality predictor. Yan and co-authors provide the data with age, but the described models are constructed only on blood test data. The question arises due to no comment on this significant factor.
- **What are the continuous relationships between variables and the target in the model?** The decision tree presented by Yan and co-authors rigorously dichotomizes continuous variables of high importance. What are the consequences? The question arises due to the potential of too significant model simplification.

We suggest answering sequences of the potential questions using the interactive and customisable dashboard.





Results

Explainable AI (XAI) techniques often provide a single-aspected view of the black-box model answering only the questions asked by the developers. We propose applying IEMA, which aims at the interactive juxtaposition of various XAI methods and data exploration techniques. This approach brings responsibility into prediction making of COVID-19 mortality as it allows answering all the potential questions about the process of models' reasoning. IEMA aims at an interactive multi-aspected view of the black-box and reassures full model transparency through providing a customisable dashboard for all stakeholders to review.

Left Figure: Data exploration should not be overlooked in AI prediction making. We see that the most important features (LDH and Lymphocytes) practically divide the data into target groups. Adding the third feature (hs-CRP) reassures almost complete separability; thus, age comes as not relevant for the model, which contradicts the knowledge. Such a simple visualisation adds significant insight into potential model training, evaluation, and explanation.

Right Figure: We present the *modelStudio* dashboard, which allows for performing Interactive Explanatory Model Analysis. It combines model explanations with data exploration visualisations for a broad view of the model's behaviour. For example, partial-dependence and average-target plots showcase the continuous relationships between variables and the target. We suggest using this framework to answer the potential questions since it is user-customisable and easy to share as a model's documentation.

Follow the URLs for further resources.

References

Baniecki, H.; and Biecek, P. 2019. modelStudio: Interactive Studio with Explanations for ML Predictive Models. *The Journal of Open Source Software* URL https://github.com/ModelOriented/modelStudio.

Baniecki, H.; and Biecek, P. 2020. The Grammar of Interactive Explanatory Model Analysis. *arXiv:2005.00497* .

Barredo Arrieta, A.; et al. 2019. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion* .

Wynants, L.; et al. 2020. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* .

Yan, L.; et al. 2020. An interpretable mortality prediction model for COVID-19 patients. *Nature Machine Intelligence* .

Dashboard & Code

rai-covid.drwhy.ai

https://github.com/hbaniecki/Pre-Surv-COVID-19

Contact

https://linkedin.com/in/hbaniecki

Acknowledgements

Work funded by the NCN Opus grant 2017/27/B/ST6/0130.