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
hubert.baniecki.stud@pw.edu.pl, przemyslaw.biecek@pw.edu.pl

Abstract

For high-stakes prediction making, the Responsible Artificial Intelligence (RAI) is more important than ever. It builds upon Explainable Artificial Intelligence (XAI) to advance the efforts in providing fairness, model explainability, and accountability to the AI systems. During the literature review of COVID-19 related prognosis and diagnosis, we found out that most of the predictive models are not faithful to the RAI principles, which can lead to biassed results and wrong reasoning. To solve this problem, we show how novel XAI techniques boost transparency, reproducibility and quality of models.

Introduction: Responsible Prediction Making

In recent years we have seen a growing interest in the area of Responsible Artificial Intelligence (Barredo Arrieta et al. 2019; Gill et al. 2020; Biecek and Burzykowski 2021). 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.

An overview of 145 predictive models for prognosis and diagnosis of COVID-19 presented by Wynants et al. (2020) is a starting point for our discussion. 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. Tayarani-N (2020) also points out this significant research gap. 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 explanations of its reasoning. In this paper, we show how to advance the current state-of-the-art predictive models into the new responsible standards, by applying Interactive Explanatory Model Analysis (IEMA) implemented in modelStudio (Baniecki and Biecek 2019, 2020).

Explainable AI techniques reviewed by Barredo Arrieta et al. often provide a single-aspected view of the black-box model answering only the questions asked by the developers. While the Model Cards framework (Mitchell et al. 2019) focuses on providing short static model documentation, 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. We showcase its application on the real use-case.

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. Authors explore multiple XGBoost models that predict the mortality rates of patients, which leads to the development of an interpretable decision tree. Analyzing the results of Yan et al. and co-authors 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 making.

Why is LDH such a critical variable? Yan, L. 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 (Wynants et al. 2020). Yan, L. 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, L. and co-authors rigorously dichotomizes continuous

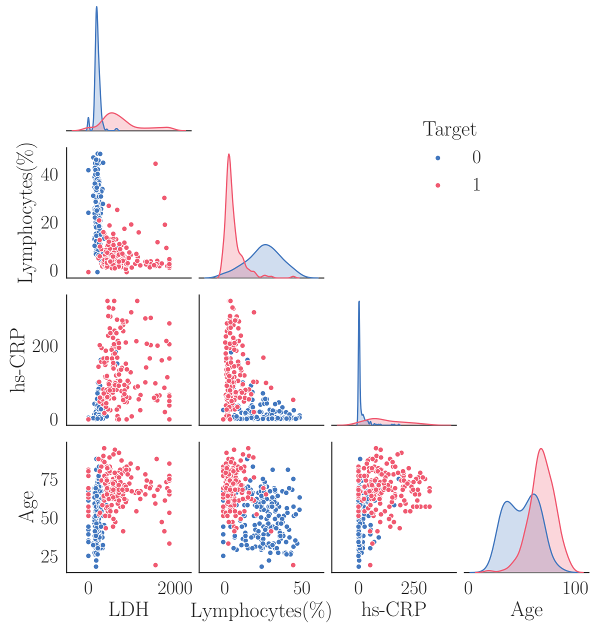

Figure 1: Data exploration should not be overlooked in AI prediction making. We see that the most important features (LDH and Lymphocytes) practically divide the data into target groups. Adding the third feature (hs-CRP) reassures almost complete separability; thus, age comes as not relevant for the model, which contradicts the knowledge. Such a simple visualization adds significant insight into potential model training, evaluation, and explanation.

variables of high importance. Such treatment, especially in medicine, has well documented detrimental consequences. The question arises due to the potential of too significant model simplification.

We provide visualizations using the supplementary resources to answer the above questions[1]. Figure 1 presents simple data exploration, while Figure 2 presents model explanation, accompanied by data exploration in the modelStudio dashboard. Additionally, we find that missing values, in continuous variables of positive values and monotonous correlation with the target, were imputed with an almost minimum value of `-1`. More queries arise since such treatment may be harmful in the final shallow tree model. Novel XAI techniques like IEMA facilitate finding answers to the sequences of potential questions.

Conclusion

While the prediction making of COVID-19 mortality is an important issue of need, current literature lacks explainability of models and reproducibility (Wynants et al. 2020; Tayarani-N 2020). With this contribution, we showcase how to embrace the responsible principles by providing a framework to cope with these shortcomings. In future work, we plan to calibrate the dashboard for COVID-19 use-case.

[1] dashboard: https://rai-covid.drwhy.ai/
code: https://github.com/hbaniecki/Pre-Surv-COVID-19/

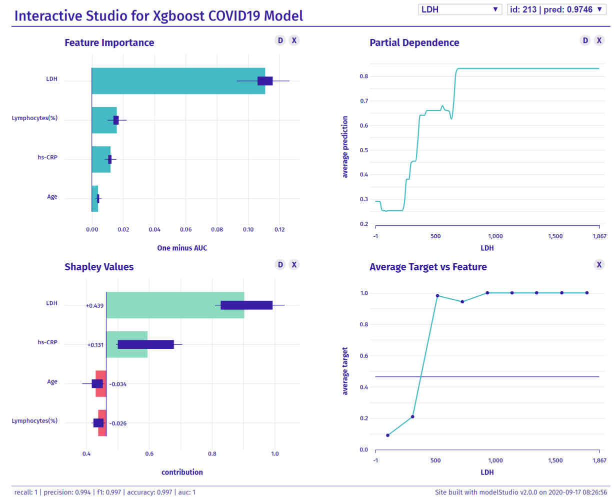

Figure 2: We present the modelStudio dashboard[1], which allows for performing Interactive Explanatory Model Analysis. It combines model explanations with data exploration visualizations for a broad view of the model's behaviour. For example, partial-dependence and average-target plots showcase the continuous 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. These traits are crucial for responsible prediction making of COVID-19 mortality.

Acknowledgments

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

References

Baniecki, H.; and Biecek, P. 2019. modelStudio: Interactive Studio with Explanations for ML Predictive Models. *Journal of Open Source Software* .

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

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

Biecek, P.; and Burzykowski, T. 2021. *Explanatory Model Analysis*. Chapman and Hall/CRC.

Gill, N.; Hall, P.; Montgomery, K.; et al. 2020. A Responsible Machine Learning Workflow with Focus on Interpretable Models, Post-hoc Explanation, and Discrimination Testing. *Information* .

Mitchell, M.; Wu, S.; Zaldivar, A.; et al. 2019. Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19* .

Tayarani-N, M.-H. 2020. Applications of Artificial Intelligence in Battling Against Covid-19: A Literature Review. *Chaos, Solitons & Fractals* .

Wynants, L.; Van Calster, B.; Collins, G. S.; 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* .