

Inspection of Blackbox Models for Evaluating Vulnerability in Maternal, Newborn, and Child Health

William Ogallo¹, Skyler Speakman¹, Victor Akinwande¹, Kush R. Varshney^{1,2}, Aisha Walcott-Bryant¹, Charity Wayua¹ and Komminist Weldemariam¹

¹IBM Research – Africa, Nairobi, Kenya

²IBM Research – T. J. Watson Research Center, Yorktown Heights, NY, USA

Abstract

Improving maternal, newborn, and child health (MNCH) outcomes is a critical target for global sustainable development. Our research is centered on building predictive models, evaluating their interpretability, and generating actionable insights about the markers (features) and triggers (events) associated with vulnerability in MNCH. In this work, we demonstrate how a tool for inspecting “black box” machine learning models can be used to generate actionable insights from models trained on demographic health survey data to predict neonatal mortality.

1 Evaluating Vulnerability in MNCH

Improving maternal, newborn, and child health (MNCH) outcomes is an important health priority for sustainable global development. However, the progress towards MNCH-specific global targets such as reducing the maternal mortality ratio to less than 70 per 100,000 live births and lowering preventable deaths among newborns and children by the year 2030 remains uneven across countries¹.

Some of the key barriers to understanding and addressing MNCH challenges are the complicated interactions of various factors and interventions captured in data, including those stemming from existing MNCH-related efforts. Furthermore, due to the limitations of traditional statistical analysis approaches, it is difficult to generate new insights from data without the innovative use of machine learning algorithms. However, although machine learning algorithms are good at prediction, they are often considered “black box” models for their lack of interpretability. Consequently, MNCH stakeholders and policymakers find it difficult to adopt machine learning models for decision-making and intervention planning.

It is plausible that certain subpopulations of mothers and children have an increased susceptibility to shocks in their environment (biological or socio-economic vulnerabilities) that predispose them to disproportionately worse health outcomes such as neonatal and maternal mortality. Our goal is to help MNCH domain experts and stakeholders study, understand,

and identify these subpopulations. Towards this goal, we are developing predictive machine learning models to identify factors that contribute to poor MNCH health outcomes, and importantly, are using an interactive visual analytics tool, *Prospector*, to inspect these “black box” models and generate actionable insights. These insights could be used by MNCH stakeholders for planning targeted interventions for vulnerable subpopulations.

2 Prospector for Inspecting Black Box Models

Prospector is a visual analytics and explainability tool to inspect “black box” models [Krause *et al.*, 2016]. A data scientist can upload any predictive model into Prospector and then visualize feature importance rankings and partial dependence plots, as well as conduct localized inspection interactively. *Feature importance* is a measure of how important a feature is (relative to other markers in the model) in predicting the outcome. *Partial dependence plots* [Friedman, 2001] illustrate how, on average, varying the value of a given feature affects the predicted outcome while holding all other features constant. *Localized inspection* generates counterfactual data to illustrate how changing the feature profile of an individual subject affects the relationship between a single feature and the outcome for that individual.

One of our goals for using Prospector is to enable insight generation by MNCH domain experts and stakeholders. For example, using the enhanced capabilities of Prospector, MNCH experts can visualize, interact, and reason about what features are important for targeted intervention planning. We believe this to be the first practical real-world use of Prospector.

3 Inspection of Neonatal Mortality Models Trained on Demographic Health Survey Data

To demonstrate how Prospector can be used to generate MNCH insights, we analyzed Demographic and Health Surveys (DHS)² data from 10 Sub-Saharan countries. DHS data are generated from a series of household surveys conducted in over 90 countries every 3 to 5 years. These surveys are nationally representative and are primarily used for collecting data

¹<https://unstats.un.org/sdgs/report/2019/>

²<https://www.dhsprogram.com/>

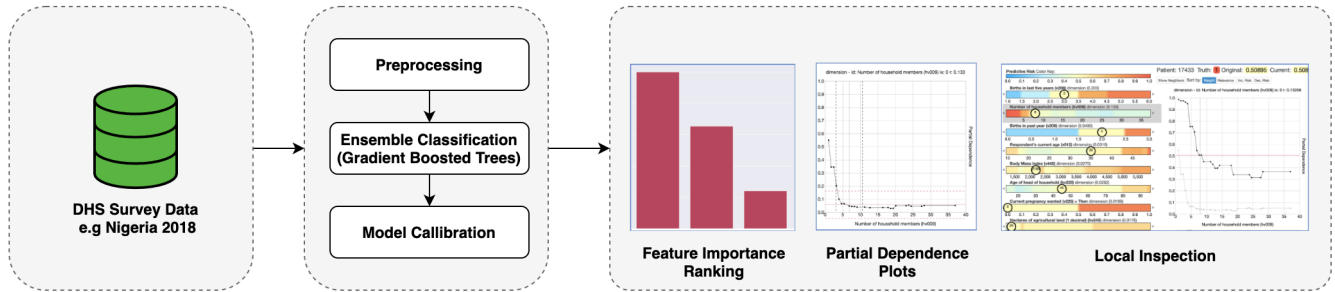


Figure 1: Pipeline for Modelling and Inspecting MNCH Models from Demographic Health Survey Data.

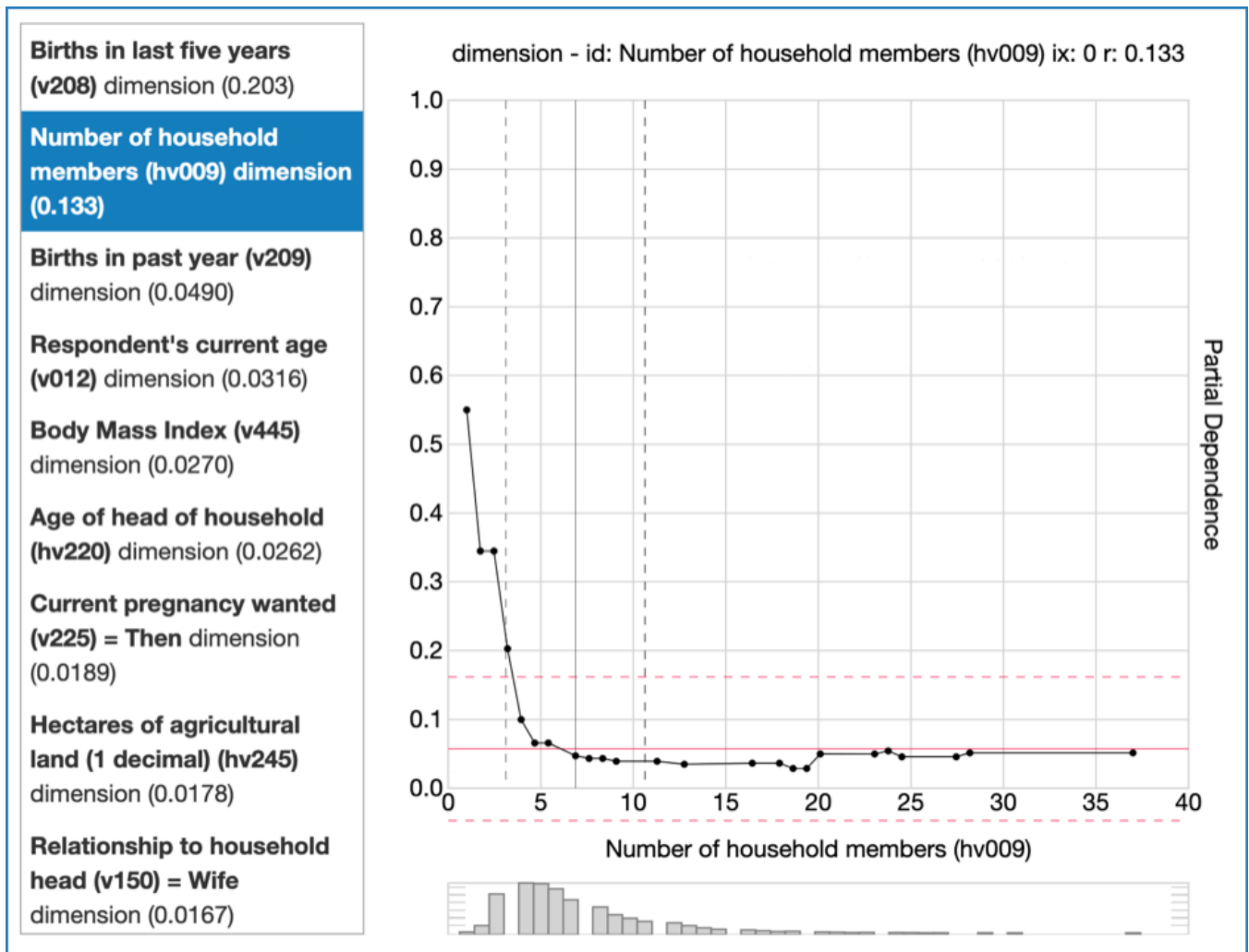


Figure 2: An example plot showing the feature importance rankings of the prediction of neonatal mortality (left) and the partial dependence of neonatal mortality on the number of household members (right) in the Nigeria 2018 Neonatal Mortality Model. Note that the user may interact with the plot in the tool.

for monitoring and impact evaluation of population, health, and nutrition indicators for individual countries, as well as for cross-country comparative analyses.

Figure 1 illustrates the pipeline used for generating predictive models and subsequently inspecting the models using Prospector. For each of the 10 countries, we analyzed the 2 most recent years of DHS data. For each survey dataset, we only considered women who reported giving birth at least once in the 5 years before the survey. We defined the unit of analysis as the mother. We extracted 43 household features (e.g. household size, wealth index, source of water, etc.) and 77 individual features (e.g. age, education level, ethnicity, etc.) giving a total of 120 features. All surveys have the same feature space. We trained an ensemble classifier using gradient boosted trees to identify women who had experienced the death of a neonate (i.e. child younger than 28 days) within 5 years before the date of the survey. For each model, hyperparameter tuning was done using grid search cross-validation.

We uploaded all the models into Prospector and examined feature importance ranking, partial dependence plots, and localized inspection. We observed that the reported number of births in the 5 years before a survey, as well as the number of household members (household size), are important markers of neonatal mortality that persist across years and countries. Furthermore, using partial dependence plots, we confirmed the previously known positive correlation between the number of births and neonatal mortality. Importantly, we identified a potentially novel finding about the negative correlation between household size and neonatal mortality. Figure 2 shows an example of how a domain expert would use the tool to visualize the partial dependence of neonatal mortality on the number of household members. It is worth noting that these findings are not causal relationships and that additional research is still needed to characterize the identified determinants of neonatal mortality.

4 Conclusions and Next Steps

Our work demonstrates a practical application of machine learning for generating novel insights on MNCH, an important global health problem. Tools such as Prospector are critical for improving explainability and transparency of “black box” machine learning models.

As avenues for future work, we are currently incorporating individual conditional expectation (ICE) plots into Prospector. This will help visualize the dependence of predictions on a given feature for individual subjects separately and in comparison to the overall partial dependence plots. Furthermore, we will develop functionalities that enable stakeholders to determine accurate representations of vulnerable subpopulations by identifying and visualizing the ICE plots of subgroups of individuals that show the highest change in prediction (risk of outcome) as the values of a given feature increases or decreases. This will involve developing anomalous pattern detection techniques for efficiently scanning the exponentially many subsets in a dataset to generate insights about specific subpopulations that are anomalous [McFowland *et al.*, 2013].

Acknowledgements

This work is funded by Bill & Melinda Gates Foundation, investment ID 52720. We thank the MNCH team from Bill & Melinda Gates Foundation, especially Nosa Orobato and Claire-Helene Mershon for their valuable domain expertise, insights, and interpretation of the findings.

References

- [Friedman, 2001] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, pages 1189–1232, 2001.
- [Krause *et al.*, 2016] Josua Krause, Adam Perer, and Kenney Ng. Interacting with predictions: Visual inspection of black-box machine learning models. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5686–5697, 2016.
- [McFowland *et al.*, 2013] Edward McFowland, Skyler Speakman, and Daniel B Neill. Fast generalized subset scan for anomalous pattern detection. *The Journal of Machine Learning Research*, 14(1):1533–1561, 2013.