8
Explainable and Unbiased Predictive Analytics for Institutional Researchers

Rishika Samala[1], Tradara McLaurine[2], **Dilip Francies**[3], Jesse Talley[2], Gina Deom[1], Stefano Fiorini[1], Christina Downey[2]

[1]Indiana University. [2]Indiana University Indianapolis. [3]Indiana University Bloomington

Brief Summary of Topic and Session Outline

With the growth of analytical methodologies, the ethical and transparent use of algorithmic outputs in university operations is of paramount importance. This session covers participatory and analytical approaches to ensure actionable analytics, support explainable AI and ML, and address bias in advanced analytics. It concludes with a discussion on effective, accessible, and unbiased application of advanced analytics in institutional research.

Proposal Abstract

With the rapid growth of analytical methodologies from both statistics and data science, the ethical and transparent application of algorithmic outputs in support of university operations and activities has become increasingly important and has generated concern in the higher education community. This session discusses effective participatory approaches that ensure actionability of analytics, support the production of explainable Artificial Intelligence and Machine Learning, and minimize and effectively address the risk of bias in advanced analytics. The session will close with an interactive discussion on what it means to provide effective, accessible, and unbiased application of advanced analytics in institutional research.

Proposal Narrative

The application of analytical algorithms to student record data significantly enhances the quality of support that a university can provide to its students. However, it can also have unintended adverse effects. This is why the Association for Institutional Research (AIR) acknowledges, in its Statement of Ethical Principles, the profound impact of Institutional Research on the people we serve (AIR Statement of Ethical Principles, 2019).

With the growth of analytical methods from both statistics and data science (e.g., The AIR Professional File Fall 2023 Volume), ethical and transparent use of algorithms in university operations and activities has become critical, raising concerns in the higher education community (Mowreader, 2024).

Our proposed session builds on over a decade of experience in developing machine learning and advanced analytics with stakeholders to support student success. Our approach is grounded in three key pillars:

1. Inclusion and participation of those who will use the analytics for actions from the outset, both to provide input on model development and to evaluate the quality and effectiveness of the models (Fiorini et al. 2018).

2. The development of explainable AI, particularly through the application of Shapley variable importance plots, to unpack the elements affecting an outcome and enhance the actionability of model predictions (Gopinath 2021, Murphy 2022, SHAPforxboost 2023).

3. The adoption of industry standards and tools to evaluate the presence and magnitude of bias in predictions and implement corrections, if necessary (Bellamy et al. 2018, Prince 2019, Gándara et al. 2024).

The session will focus on a process and models we implemented to support advising actions for priority populations on one of our university campuses. This process allows us to harness resources to create a supportive educational experience for students who might experience adverse outcomes in their first semester. After providing an overview and a rapid introduction to the participatory work implemented in the project (see point 1 above), we will focus our presentation on pillars two and three.

A key challenge is translating complex insights from advanced analytics into actionable knowledge for non-technical stakeholders. Current methods for analyzing variable importance, such as Shapley variable importance plots, permutation importance, tree-based feature importance, and partial dependence plots, need careful introduction to allow for proper interpretability by non-experts. Simplifying these insights for non-technical audiences has proven valuable, and we will present methods for doing so, focusing on how to balance the benefits and limitations of these models.

Bias in machine learning, stemming from data collection or algorithms or historical practices, affects reliability and fairness. This is particularly crucial in educational applications, where biased models can perpetuate inequalities. Ensuring fairness in machine learning is essential for providing equitable treatment across demographic groups. In our session, we will discuss our use of the Equal Opportunity Difference (EOD) metric to evaluate bias. EOD measures true positive rate differences between privileged and unprivileged groups, with a goal of achieving a value of 0 (indicating fairness).

We will present our results using this fairness metric and discuss the factors contributing to the positive outcomes we observed. This will open the floor for broader audience engagement – concluding with an interactive discussion on the effective, accessible, and unbiased application of advanced analytics in institutional research, which will allow attendees to share their experiences and implementation plans.

Learning Outcomes

Our goal for this session is to contribute and expand knowledge sharing and conversations in the areas of interpretable AI and bias in advanced modeling. Both topics are critical for the ethical practice of Institutional Research.

Session outcomes will be:

1. Understanding of the concepts and current methodologies applied in the field of data science,
2. Overview of practical applications of these two approaches and insights in how they contribute to practitioner's use of the data and its contribution to inform programs,
3. An understanding of current work and initiatives implemented and under development at various other institutions.

Keywords

Explainable AI and ML, bias in advance modeling, actionable analytics, participatory action research.

References

1. Bellamy, R. K., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., ... & Zhang, Y. (2019). AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development*, *63*(4/5), 4-1.

2. Chu, W., Hosseinalipour, S., Tenorio, E., Cruz, L., Douglas, K., Lan, A., Brinton, C. (2023). Multi-Layer Personalized Federated Learning for Mitigating Biases in Student Predictive Analytics. *Journal of LaTeX Class Files*, 14(8).

3. Gándara, D., Anahideh, H., Ison, M. P., & Picchiarini, L. (2024). Inside the Black Box: Detecting and Mitigating Algorithmic Bias Across Racialized Groups in College Student-Success Prediction.

AERA Open, *10*. https://doi.org/10.1177/23328584241258741

4. Mowreader, Ashley. (2024). Report: Predictive Models May Have Bias Against Black and Hispanic Learners. *Inside Higher Ed.*

5. Prince, S. (2019). Tutorial #1: bias and fairness in AI. Borealis AI.

6. Gopinath, D. (2021). The Shapley Value for ML Models. What is a Shapley value, and why is it crucial to so many explainability techniques? *TowardsDataScience.*

7. Murphy, A. Shapley Values – A Gentle Introduction. (2022). *H20.ai Blog.*

8. Package 'SHAPforxgboost'. (2023). https://cran.rproject.org/web/packages/SHAPforxgboost/SHAPforxgboost.

9. Kwon, Yongchan, and James Y. Zou. "Weightedshap: analyzing and improving shapley based feature attributions." Advances in Neural Information Processing Systems 35 (2022): 34363-34376.