Responsible Machine Learning
Assignment 2
10 points

November 5, 2026

In Assignment 2, analyze and explain machine learning (ML) models developed for Assignment 1. Keep your work in a public GitHub repository, or in a private repository shared with the instructor (`jphall663`). You may use AI tools, but you are responsible for validating and understanding your results.
For each section, consider whether the explanatory results make sense given domain knowledge and whether differences between the models are logical and informative.

1 Calculate and Plot Global Feature Importance.

Use regression coefficients, Shapley values, or another technically sound method to calculate global feature importance for your models. Plot the results as bar charts and compare how the models rank and weight input features.

2 Calculate and Plot Local Feature Importance.

Calculate local feature importance for individuals at the 10^{th}, 50^{th}, and 90^{th} percentiles of predicted probability for each model. Plot the local feature importance values and compare how the models explain individuals at similar points in their predicted probability distributions.

3 Calculate and Plot Feature Behavior.

Plot partial dependence for each main effect feature under each model. You may also plot interaction contours or surfaces, such as interactions identified by EBM. Alternative techniques include accumulated local effects (ALE), individual conditional expectation (ICE), and EBM feature scores.

4 Submit Code Results.

Submit your code link and prediction file to the class evaluation server at a TBD URL by the deadline below. The class evaluation server will rank submissions using the cross-validated ranking method described in the model evaluation notebook. It will also use an LLM-as-judge to assess whether the submitted code follows assignment instructions. The four points for the prediction deliverable will be based on the ranking. You will have opportunities to improve your rank each week of the class.

Grading rubric:

Global feature importance plots: 2 points.
Local feature importance plots: 2 points.
Feature behavior plots: 2 points.
Ranking on holdout data: 4 points.

Due: TBD.