Ranking Geographic and Sociodemographic Risk Factors for Youth Mental Health: Variable Importance Analysis by SHAP Values for Model Interpretation

by

Ke Wang

STAT 7320 Project Report

Department of Statistics

University of Manitoba

Winnipeg, Canada

Abstract

Analyzing risk factors is important in preventing mental health issues among young individuals. This project aims to prioritize features associated with youth mental health symptoms. Utilizing data from phases 3.5, 3.6, 3.7, 3.8, and 3.10 of the Household Pulse Survey (HPS) dataset, we trained models via the eXtreme Gradient Boosting (XGBoost). Following this, the SHapley Additive exPlanations (SHAP) were employed to determine the ranking of these features. Our findings revealed that the proportion of children in the household was the most significant variable, while tenure status and state of residence were the least significant variables.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The state of our mental health significantly influences our daily experiences, shaping our thoughts, emotions, and actions. Maintaining good mental health and preventing mental illness are essential for leading fulfilling lives and making sound decisions. Mental disorders in children manifest as significant deviations from typical learning, behavior, or emotional regulation, causing distress and hindering daily functioning. Common diagnoses among children include Attention Deficit Hyperactivity Disorder (ADHD), anxiety problems, behavioral disorders, and depression. According to the Centers for Disease Control and Prevention (2023), between 2016 and 2019, approximately 9.8% of U.S. children aged 3–17 were diagnosed with ADHD, 9.4% with anxiety disorders, 8.9% with behavior problems, and 4.4% with depression. Pew Research Center (2023) revealed that 40% of U.S. parents expressed serious concerns regarding their children's mental well-being.

Due to the growing concern surrounding children's mental health, many studies have been conducted to examine the risk factors associated with mental health issues, aiming to enhance diagnosis and treatment of mental disorders. Liu et al. (2023) conducted research estimating the prevalence of mental health symptoms in children with shared characteristics. Essex et al. (2006) employed the MacArthur moderator-mediator approach to identify various risk factors for children's mental health and to investigate their interrelations, categorizing them as proxy, overlapping, mediator, moderator, and independent risk factors. Panchal et al. (2023) conducted a systematic review, selecting papers based on specific inclusion criteria, and then applied meta-analysis to evaluate risk factors. However, assessments of how risk factors are ranked in terms of their impact on children's mental health issues are still in need of further exploration.

The aim of this study is to prioritize the features contributing to mental health symptoms among children through variable importance analysis. Determining the order of variables related to mental health symptoms in youth can assist professionals in gaining a clearer understanding of risk factors, thereby facilitating easier diagnosis and proper treatment.

The remaining part of this report is structured as follows: Chapter 2 outlines the methodology employed to identify the ranking of features. Chapter 3 provides an overview of the statistical analysis, encompassing data sources and findings. Chapter 4 details the conclusion, limitations and further work of this study.

Chapter 2

Methodology

2.1 XGBoost

We utilized eXtreme Gradient Boosting (XGBoost, Chen and Guestrin, 2016) to explore the relationship between various features and the occurrence of mental health symptoms in youth within U.S. households. Following its initial introduction in 2014, XGBoost quickly gained widespread attention from both academia and industry due to its outstanding performance in various machine learning tasks. The library was widely adopted by data scientists, machine learning practitioners, and data science competition participants for its efficacy and versatility. For example, among the 29 challenge-winning solutions highlighted in Kaggle's blog throughout 2015, 17 solutions used XGBoost, making it the most poular choice, followed by 11 solutions using deep neural networks (Chen and Guestrin, 2016).

In contrast to gradient boosting (Friedman, 2001), XGBoost employs the second-order Taylor's expansion of loss function for a more precise approximation. Moreover, it integrates a suite of advanced techniques to enhance its execution speed; see its official documentation at `https://xgboost.readthedocs.io/en/stable/index.html` for details. Although originally conceived to optimize tree-based models, its versatility has expanded to accommodate a broader range of models that are trained through loss function minimization. As of now, XGBoost stands as one of the leading algorithms for modeling large datasets effectively (Fu, 2024).

Consider a training set composed of paired observations $(\boldsymbol{x}_i, y_i)$, where $i = 1, \ldots, n$, with vectors $\boldsymbol{x}_i$ and scalar responses y_i. XGBoost develops a model in an additive manner, beginning with the minimizer of a given loss function $\sum_{i=1}^{n} L(y_i, f(\boldsymbol{x}_i))$ with respect to $f \in \mathcal{F}$, where $\mathcal{F}$ is a pre-specified space. In essence, at each iteration, the model construction progresses by adding a weak learner r which minimizes the sum of loss function and a regularization term Ω; see Algorithm 1 for a sketch of the algorithm.

The only difference between regression and classification by XGBoost is the assigned loss function. For regression, commonly used loss functions include squared error loss, squared log loss, and Pseudo Huber loss, which is an alternative to absolute loss that is twice differentiable. For binary classification, popular choices are logistic loss, hinge loss, and raw logistic loss, which outputs scores before logistic transformation. XGBoost also supports a range of other tasks such as survival analysis, ranking, and multi-class

Algorithm 1 A sketch of XGBoost.

1: Initialize model $f_0 \leftarrow \operatorname{argmin}_{f\in\mathcal{F}} \sum_{i=1}^{n} L(y_i, f(\boldsymbol{x}_i))$.

2: For $m = 1, \ldots, M$:

a: Train a model

$$r_m \leftarrow \underset{r\in\mathcal{F}}{\operatorname{argmin}} \sum_{i=1}^{n} L(y_i, f_{m-1}(\boldsymbol{x}_i) + \eta r(\boldsymbol{x}_{\boldsymbol{i}})) + \Omega(r),$$

where $0 < \eta < 1$ is the learning rate and $\Omega(\cdot)$ is a regularization/penalty term.

b: Update model $f_m(\cdot) \leftarrow f_{m-1}(\cdot) + \eta r_m(\cdot)$

3: Output $\hat{f} \leftarrow f_M$.

classification; the corresponding loss functions can be found in its official documentation. Since this project aims to perform a binary classification task, we selected the logistic loss

$$L(y_i, f(\boldsymbol{x}_i)) = -y_i \ln f(\boldsymbol{x}_i) - (1 - y_i) \ln\bigl(1 - f(\boldsymbol{x}_i)\bigr)$$

(with y_i encoded as 0 or 1) and the hinge loss

$$L(y_i, f(\boldsymbol{x}_i)) = \max(0, 1 - y_i f(\boldsymbol{x}_i))$$

(with y_i encoded as ± 1) for XGBoost model training.

For illustrative purposes, we demonstrate how tree-based XGBoost performs binary classification using logistic loss. XGBoost algorithm initializes the model f_0 by minimizing the logistic loss. For simplicity, we set learning rate $\eta \approx 1$, indicating a small number of iterations. At iteration m, XGBoost trains a tree r_m by minimizing the objective function

$$\mathrm{obj}^{(m)} = \sum_{i=1}^{n} L(y_i, f_{m-1}(\boldsymbol{x}_i) + r(\boldsymbol{x}_i)) + \Omega(r)$$

with respect to $r \in \mathcal{F}$. To obtain a nice form efficiently, XGBoost approximates $\mathrm{obj}^{(m)}$ via the second-order Taylor's expansion:

$$\mathrm{obj}^{(m)} \approx \sum_{i=1}^{n} [L(y_i, f_{m-1}(\boldsymbol{x}_i)) + g_{m,i} r(\boldsymbol{x}_i) + \frac{1}{2} h_{m,i} r(\boldsymbol{x}_i)] + \Omega(r) + \text{constants},$$

where $g_{m,i} = \frac{\partial L(y_i, f_{m-1}(\boldsymbol{x}))}{\partial f_{m-1}(\boldsymbol{x})}\Big|_{\boldsymbol{x}=\boldsymbol{x}_i}$ and $h_{m,i} = \frac{\partial^2 L(y_i, f_{m-1}(\boldsymbol{x}))}{\partial f_{m-1}(\boldsymbol{x})^2}\Big|_{\boldsymbol{x}=\boldsymbol{x}_i}$. After all constants are removed, the objective function is simplified to

$$\mathrm{obj}^{(m)} \approx \sum_{i=1}^{n} (g_{m,i} r(\boldsymbol{x}_i) + \frac{1}{2} h_{m,i} r^2(\boldsymbol{x}_i)) + \Omega(r).$$

To further compressing the objective function, we first need to introduce the definitions of the tree function r and the regularization term $\Omega(r)$. In XGBoost, the tree is refined as

$$r(\boldsymbol{x}) = w_{q(\boldsymbol{x})}$$

where $\boldsymbol{w} = [w_1, \ldots, w_T]^\top \in \mathbb{R}^T$ is a vector of scores on leaves with T as the number of leaves and $q : R^d \to \{1, 2, ..., T\}$ assigning each data point to a corresponding leaf. The regularization term is defined as

$$\Omega(r) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2,$$

where both γ and λ are tuning parameters. With the introduction of tree function r and regularization term $\Omega(r)$, the objective function can be further updated as

$$\begin{aligned}\mathrm{obj}^{(m)} &\approx \sum_{i=1}^{n}(g_{m,i}w_{q(\boldsymbol{x_i})} + \frac{1}{2}h_{m,i}w^2_{q(\boldsymbol{x_i})}) + \gamma T + \frac{1}{2}\sum_{j=1}^{T} w_j^2 \\ &= \sum_{j=1}^{T}[(\sum_{i\in I_j} g_{m,i})w_j + \frac{1}{2}(\sum_{i\in I_j} h_{m,i} + \lambda)w_j^2] + \gamma T,\end{aligned}$$

where $I_j = \{i \in \{1,\ldots,n\} \mid q(\boldsymbol{x_i}) = j\}$ is the set of indices of data points assigned to leaf j. Taking $G_{m,j} = \sum_{i\in I_j} g_{m,i}$ and $H_{m,j} = \sum_{i\in I_j} h_{m,i}$, the objective function can be eventually compressed as

$$\mathrm{obj}^{(m)} \approx \sum_{j=1}^{T}[G_{m,j}w_j + \frac{1}{2}(H_{m,j} + \lambda)w_j^2] + \gamma T, \tag{2.1}$$

where the best w_j is given by

$$w_j^* = -\frac{G_{m,j}}{H_{m,j} + \lambda}.$$

After obtaining the r_m by minimizing objective function (2.1) , the model at iteration m can be updated by

$$f_m(\cdot) = f_{m-1}(\cdot) + r_m(\cdot).$$

After M iterations, the final model $\hat{f}$ is set to f_M, which outputs probabilities. The prediction is made based on whether the probability is greater or less than 0.5.

The algorithm for XGBoost with hinge loss is similar with logistic loss. However, hinge loss is non-differentiable. To address this issue, at iteration m, XGBoost sets $g_{m,i} = -y_i$ and $h_{m,i} = 1$ when $y_i f_m(\boldsymbol{x}_i) < 1$, and sets $g_{m,i} = 0$ and $h_{m,i}$ as a small positive non-zero value otherwise.

2.2 Model Interpretation via SHAP

Shapley value is a solution concept from game theory. A solution concept is a formal rule that describes how players participate in the game. In transferable utility (TU) games, players collaborate to achieve an overall gain, with payoffs accruing to the coalition rather than individual players. In terms of distributing the coalition's worth, the most influential solution concept is the Shapley value, which was initially proposed by Shapley (1953) and extended to TU game theory by Owen (2013). In a game with n players and an additive set-function v determining the coalition's value, the Shapley value for player j is given by:

$$\phi_j(v) = \sum_{S \subseteq N} \gamma_n(s)[v(S) - v(S \backslash \{j\})], \tag{2.2}$$

where S is a subset of the full set N and $\gamma_n(s) = (s-1)!(n-s)!/n!$ with s denoting the cardinality of S. When player j is not included in S, $v(S\backslash\{j\}) -$

$v(S) = 0$ as $S\backslash\{j\} = S$. (2.2) quantifies the jth player's contribution to the final outcome of the coalition, with the sum of contributions equal to the outcome.

In the statistical context, the "coalition" represents the model prediction, where each player corresponds to a predictor, and the contribution measured by Shapley value corresponds to the importance of each predictor. The method of interpreting models based on Shapley values is called SHapley Additive exPlanations (SHAP, Lundberg and Lee, 2017).

Denote by $\hat{f}(\cdot \mid S)$ a predictive model trained by predictors in set S. Given a vector of covariates $\boldsymbol{x}$, accommodating for predictive models, substitute $\hat{f}(\boldsymbol{x} \mid S)$ and $\hat{f}(\boldsymbol{x} \mid S\backslash\{j\})$ for $v(S)$ and $v(S\backslash\{j\})$ in (2.2), respectively, where $\hat{f}(\cdot \mid S\backslash\{j\})$ is analogous to $\hat{f}(\cdot \mid S)$ but trained with the jth predictor excluded. (2.2) hence becomes a function of both $\boldsymbol{x}$ and $\hat{f}$:

$$\phi_j(\boldsymbol{x}, \hat{f}) = \sum_{S \subseteq N} \gamma_n(s)[\hat{f}(\boldsymbol{x} \mid S) - \hat{f}(\boldsymbol{x} \mid S\backslash\{j\})], \tag{2.3}$$

where $\phi_j(\boldsymbol{x}, \hat{f})$ is referred as SHAP values. The absolute value of (2.3) indicates the importance of covariate j for predicting $f(\boldsymbol{x})$, while its sign indicates whether the variables exert positive or negative effects on the prediction.

Chapter 3

Statistical Analysis

3.1 Data Sources

Household Pulse Survey (HPS) is one of the experimental data products from United States Census Bureau (2024). It was designed to collect data to measure how emergent issues impact U.S. households from social and economic perspectives. The survey contains multiple phases. Data collection for phase 1 started in April 2020 and phase 4.0 is currently ongoing until April 2024.

In this study, phases 3.5 through 3.8 and 3.10 were selected, as the other phases lacked information related to youth mental health. Phases 3.5 - 3.8 were conducted between June 1, 2022 and May 8, 2023, and phase 10 was conducted between August 23, 2023 and October 30, 2023. From a total of 951,878 households, 206,999 households were chosen, meeting the criteria of having at least one child under 18 and no missing values for the selected

13 variables. All data were assessed from HPS public use file on the U.S. Census Bureau's official website.

To assess the presence of mental health symptoms in households, feature extraction was performed on variables related to children's emotions. A household was considered to have youth with mental health symptoms if at least one child exhibited feelings of anxiety, clinginess, sadness, or depression in the past 4 weeks.

The original predictors include Hispanic origin, race/ethnicity, educational attainment, marriage status, total household population, total number of people under 18 years old in the household, recent household job loss, tenure status, total household income before taxes, and state of residence. Hispanic origin and race/ethnicity were combined as one race variable which contains 5 levels including White, Black, Asian, Hispanic and others. The children proportion in the household was defined as the number of people under 18 years old divided by the total household population.

3.2 Model Training and Evaluation

We employed XGBoost tree models using logistic loss and hinge loss on training data. Additionally, logistic regression and random forest models were fitted for comparative analysis.

XGBoost models need to be tuned properly to acheive optimal performance. For binary classification task, XGBoost tree algorithm offers multiple hyperparameters to manage model complexity and introduce randomness. These hyperparameters include boosting iterations, max tree depth, learning rate, minimum loss reduction, etc.. We performed 5-fold cross-validation using random search on the training data to identify the optimal hyperparameters, with F1 score serving as the optimization metric due to data imbalance. Unfortunately, visualizing the tuning process is challenging due to the involvement of multiple hyperparameters.

The evaluation results presented in Table 3.1 provide insights into the performance of each model. XGBoost with logistic loss outperforms other models since it has the highest F1 score, area under the curve (AUC) and overall accuracy (ACC). Logistic regression presents similar performance with XGBoost with logistic loss, suggesting a potential linear relationship between predictors and the response variable. Meanwhile, XGBoost with hinge loss reveals slightly worse performance compared to XGBoost with logistic loss and logistic regression, while random forest lags behind all other models. Note that in the comparison between ensemble methods, XGBoost consistently outperforms random forest.

Table 3.1: Model evaluation on test data using F1 score, AUC and ACC.

	F1	AUC	ACC
XGBoost: Logistic Loss	0.77314	0.60629	0.64473
XGBoost: Hinge Loss	0.77013	0.59524	0.64238
Logistic Regression	0.77072	0.59900	0.64356
Random Forest	0.74473	0.56641	0.62688

3.3 SHAP Results

Through SHAP analysis, we found that these models consistently rank most variables in similar orders. Figure 3.1 illustrates the mean absolute SHAP values for all predictors. These values are computed by averaging the absolute SHAP values across all observations. In the figure, we observe that the proportion of children in the household emerges as the most significant variable, followed by the race and marital status of respondents. Variables such as state of residence, tenure status, and recent work loss in the household appear to be insignificant. Notably, there is a discrepancy in the ranking of the household income variable between the two models. XGBoost with hinge loss identifies it as the second most important variable, while XGBoost with logistic loss places it third from the last.

The mean absolute SHAP values provide insight solely into the magnitude of SHAP values. To gain a comprehensive understanding that also incorporates the impact of directions, we can visualize the SHAP values through beeswarm plots, as depicted in Figure 3.2. These plots illustrate the SHAP value for each predictor across every observation, providing an information-dense summary of how variables influence response prediction. In Figure 3.2 , variables are arranged in descending order of significance from top to bottom, aligning with the orders revealed in the mean absolute SHAP values in Figure 3.1 . The color bar representing feature values on the right is applicable only to numerical predictors, such as the proportion of children

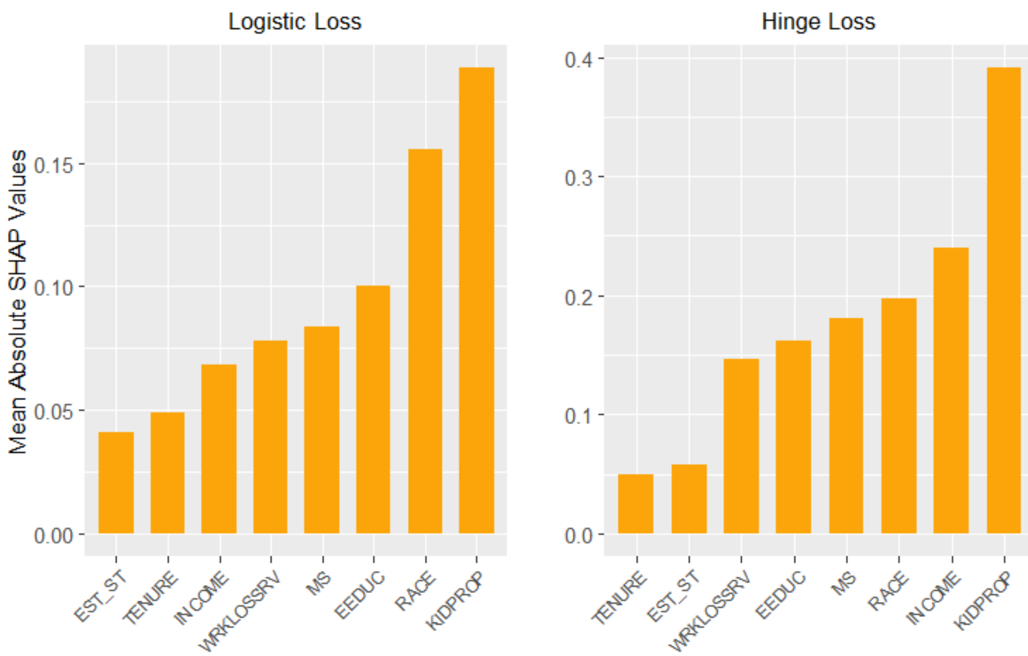

Figure 3.1: Left: Mean absolute SHAP values for variables on HPS data by XGBoost with logistic loss. Right: Mean absolute SHAP values for variables by XGBoost with hinge loss.

in our dataset, where we observe a positive correlation between children and SHAP values. For categorical predictors, colors represent different classes; for instance, there is a clear distinction between the two classes of work loss. However, for other predictors, we notice some overlap among their various levels.

To gain insight into the patterns within each variable, we can create dependence plots shown in Figure 3.3 and 3.4. These plots illustrate the SHAP values of individual variables plotted against their continuous values for numerical variables and against classes for categorical variables. The SHAP values depicted in the dependence plots indicate the extent to which the model prediction changes given specific variable values. Consistent patterns are observed across all variables in both models. For instance, households with respondents who have higher levels of education are more likely to have children revealing mental health symptoms. Households with higher incomes tend to have children who are less likely to show mental health symptoms. Furthermore, among all states, children residing in Utah are most likely to exhibit mental health symptoms as Utah shows the highest positive SHAP values, while children in Florida are least likely to exhibit such symptoms as Florida demonstrates approximately the lowest negative SHAP values.

SHAP values provide insights into individual observations as well. For instance, Figure 3.5 displays the SHAP values for the first observation in the test data. For this particular observation, both models highlight marital status and the proportion of children as influential variables, while variables

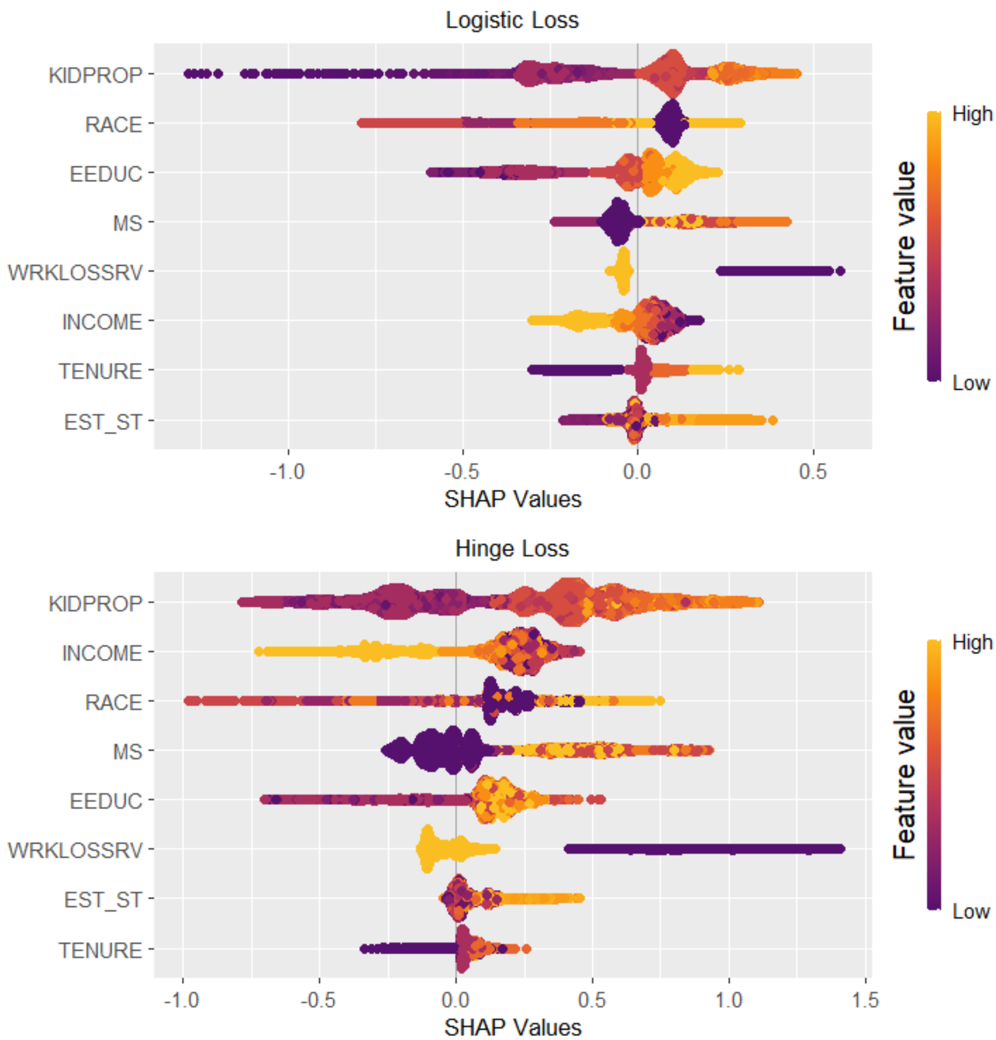

Figure 3.2: Top: SHAP values beeswarm plot for variables on HPS data by XGBoost with logistic loss. The color bar on the right is only valid for continuous predictors. For discrete predictors, different colors represent different levels. Bottom: SHAP values beeswarm plot for variables on HPS data by XGBoost with hinge loss.

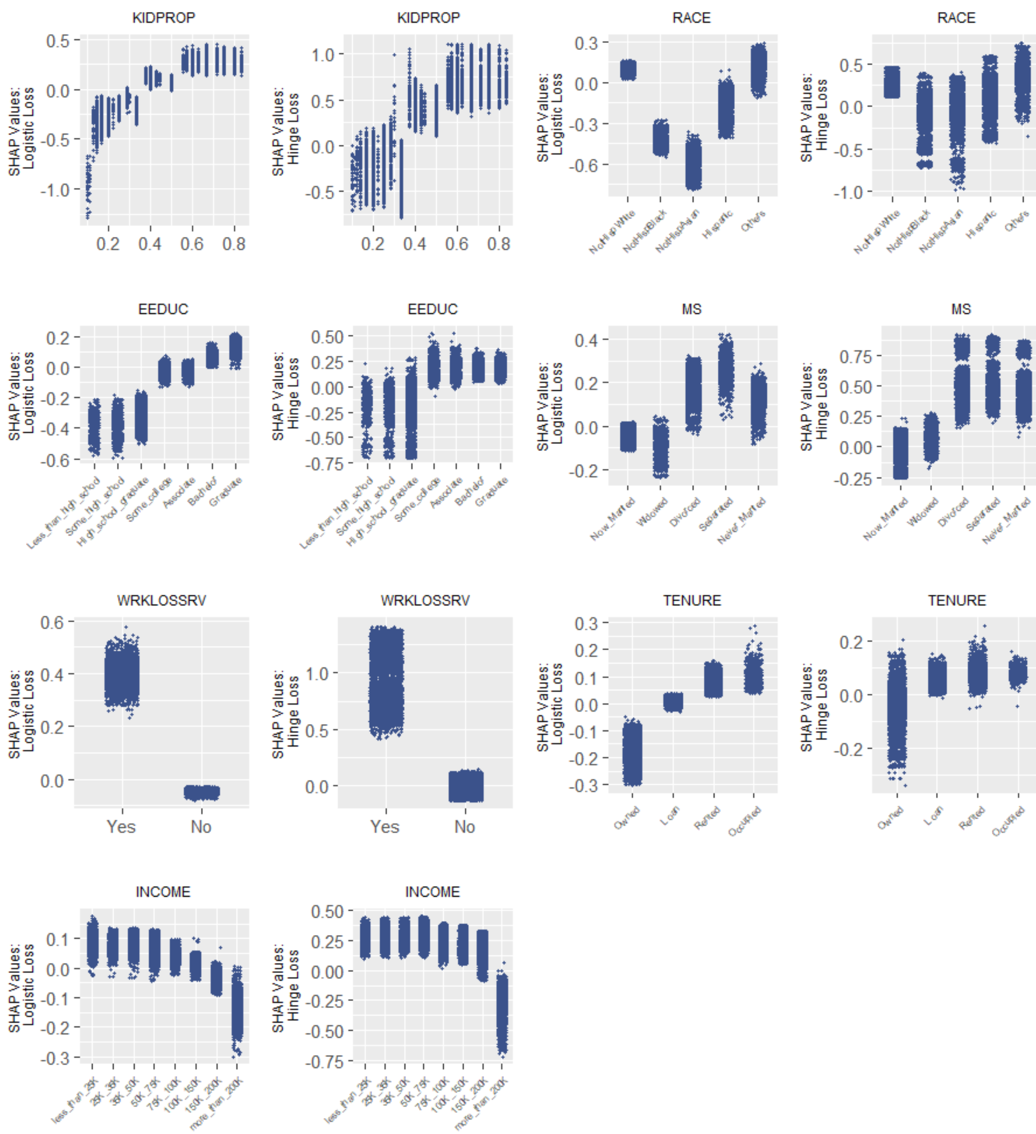

Figure 3.3: Dependence Plots for Variables Including Proportion of Children, Race, Education Attainment, Recent Work Loss, Tenure Status and Household Income

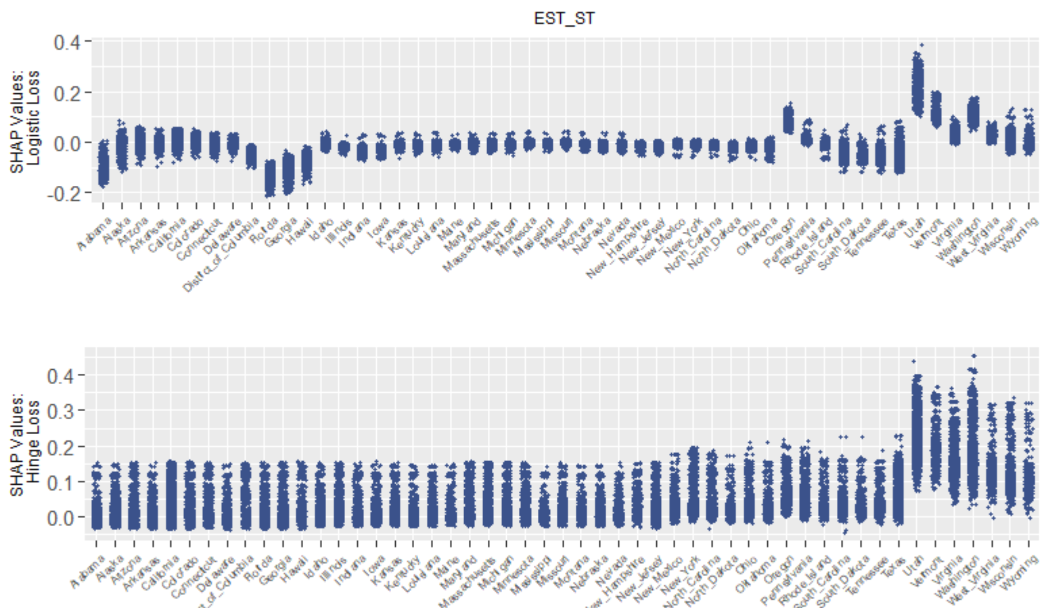

Figure 3.4: Dependence Plots for Variable State of Residence

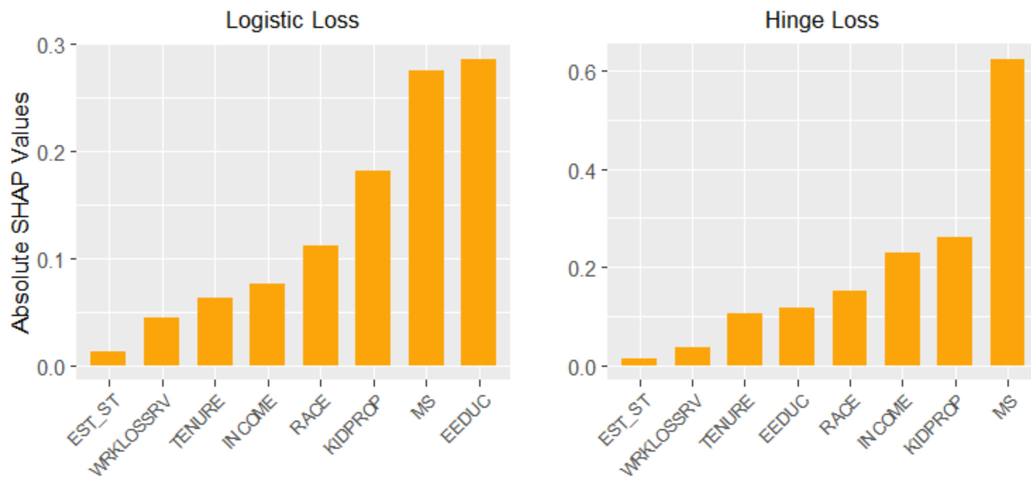

Figure 3.5: Left: Individual observation absolute SHAP values for variables on HPS data by XGBoost with logistic loss. The selected observation is the first observation from test data. Right: Individual observation absolute SHAP values for variables by XGBoost with hinge loss.

like state of residence, recent work loss, and tenure status seem to hold little significance. A notable disparity between the results from hinge loss and logistic loss lies in the ranking of education attainment. While XGBoost with logistic loss identifies it as the most significant variable, XGBoost with hinge loss ranks it fifth.

The sum of all SHAP values for an individual observation represents the difference between the baseline expected model prediction and a specific model prediction. We can visualize this difference by iteratively adding predictors one by one using a waterfall plot, which also demonstrates the additive atttribute of SHAP values. In Figure 3.6, we present the waterfall plot for the same first observation in the test data by XGBoost with logistic loss. The variable income has a SHAP value of 0.0762, and then we add this value to the baseline expectation of -0.572 yielding a sum of -0.4958. Variable education has a SHAP value of -0.286, and then we subtract 0.286 from the previous sum of -0.4958 yielding a new result of -0.7818. After repeating this process for all predictors, we obtain the final prediction of -0.569 for this observation. Note that SHAP values computed by the XGBoost algorithm are in log-odds format. To obtain the original probability p,, we can convert the SHAP value ϕ using (3.1):

$$p = \frac{\exp(\phi)}{1 + \exp(\phi)}. \tag{3.1}$$

After converting the sum of SHAP values to the original probability, we obtain a prediction response of 0.361, which is consistent with the prediction from XGBoost with logistic loss. Note that the waterfall plot may not always be suitable for extracting insights when using the entire test data. When drawing a waterfall plot for two or more observations, the algorithm computes the average SHAP values for each variable without first taking their absolute values. Consequently, the waterfall plot for variables with both positive and negative individual SHAP values can be misleading.

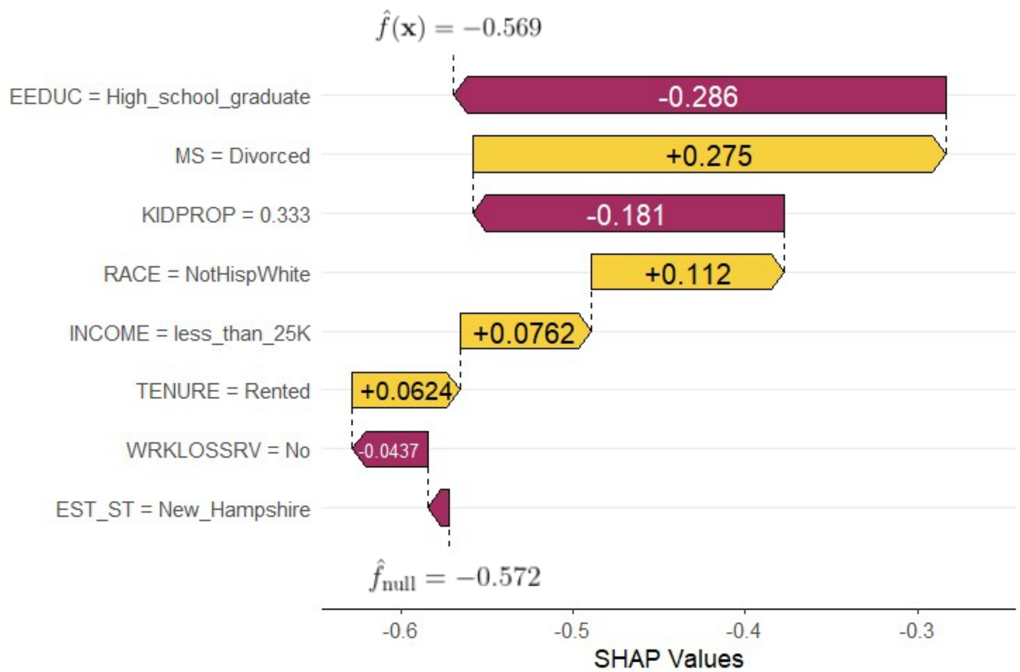

Figure 3.6: Individual observation SHAP Values waterfall plot on HPS data by XGBoost with logistic loss. $\hat{f}_{\text{null}}$ is the response prediction with null model and $\hat{f}(\mathbf{x})$ is the final response prediction for this observation.

Chapter 4

Discussions

By utilizing SHAP values with our XGBoost models, we effectively determined the ranking of selected variables. Our findings indicated that the proportion of children in the household was the most influential variable, whereas tenure status and state of residence were least significant. Furthermore, in the dependence plots analysis, we observed a positive correlation between education attainment and variable importance, and a negative correlation between household income and variable importance, consistent with findings from the prevalence estimation study of Liu et al. (2023). Professionals in the field of youth mental health can leverage the insights from this study to further explore these features, and enhance diagnostic and treatment strategies.

Several limitations in this study need to be noted. First, the assessment of all variables relied on the adults within the surveyed households. While household-related variables like state of residence and tenure status were likely to be accurately captured, individual-related factors such as race solely reflected the race of the respondent and not necessarily that of the children in the household. Moreover, the response variable relied on parental reports rather than medical diagnoses. Second, this study assumed independence among predictors, which is often not met in practice. Last, the study exclusively focused on sociodemographic and geographic variables in HPS data. This limitation in variable selection may have led to the missing of potential important variables that could impact the response.

Further analysis should be conducted from the following aspects. First, explore the correlation between predictors and add significant interaction terms. Second, expand the scope of variables considered. Last, as the number of predictors increases, variable selection can be considered for dimension reduction and model performance improvement. An option for this study is the SHAPley EXplaination Randomization Test (SHAP-XRT) proposed by Teneggi et al. (2023), which examines whether the distribution of responses changes with the introduction of a new predictor using SHAP values.

Bibliography

Centers for Disease Control and Prevention (2023). Data and statistics on children's mental health. http://web.archive.org/web/20080207010024/http://www.808multimedia.com/winnt/kernel.htm. Accessed: 2024-04-06. (Cited on page 1.)

Chen, T. and C. Guestrin (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. (Cited on page 3.)

dmlc XGBoost (2022). Random forests(tm) in xgboost. https://xgboost.readthedocs.io/en/stable/tutorials/rf.html. Accessed: 2024-03-26. (Not cited.)

Essex, M. J., H. C. Kraemer, J. M. Armstrong, W. T. Boyce, H. H. Goldsmith, M. H. Klein, H. Woodward, and D. J. Kupfer (2006). Exploring risk factors for the emergence of children's mental health problems. *Arch Gen Psychiatry 63*(11), 1246–56. (Cited on page 2.)

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics 29*, 1189–1232. (Cited on page 3.)

Fu, H. (2024). A story and potential actions. Presentation at *the Fireside Conversation on Empowering Statistics in the Era of AI*. Accessed: 2024-03-17. (Cited on page 4.)

Liu, J., Z. Zhou, X. Cheng, and N. Vangeepuram (2023). Geographic and sociodemographic variations in prevalence of mental health symptoms among us youths, 2022. *American Journal of Public Health 113*(10), 1116–1119. (Cited on pages 2 and 22.)

Lundberg, S. and S. I. Lee (2017). A unified approach to interpreting model predictions. (Cited on page 9.)

Meng, Y., N. Yang, Z. Qian, and G. Zhang (2021). What makes an online review more helpful: An interpretation framework using xgboost and shap values. *Journal of Theoretical and Applied Electronic Commerce Research 16*(3), 466–490. (Not cited.)

Owen, G. (2013). *Game Theory*. Emerald Group Publishing Limited. (Cited on page 8.)

Panchal, U., G. Salazar de Pablo, M. Franco, C. Moreno, M. Parellada, C. Arango, and P. Fusar-Poli (2023). The impact of COVID-19 lockdown on child and adolescent mental health: systematic review. *Eur. Child Adolesc. Psychiatry 32*(7), 1151–1177. (Cited on page 2.)

Pew Research Center (2023). Parenting in america today. https://www.pewresearch.org/social-trends/2023/01/24/parenting-in-america-today/. Accessed: 2024-02-26. (Cited on page 1.)

SHAP (2018a). An introduction to explainable ai with shapley values. https://shap.readthedocs.io/en/latest/index.html. Accessed: 2024-02-02. (Not cited.)

SHAP (2018b). Math behind linear explainer with correlation feature perturbation. https://shap.readthedocs.io/en/latest/index.html. Accessed: 2024-02-05. (Not cited.)

Shapley, L. S. (1953). A value for n-person games. In H. W. Kuhn and A. W. Tucker (Eds.), *Contributions to the Theory of Games (AM-28), Volume II*, pp. 307–318. Princeton: Princeton University Press. (Cited on page 8.)

Teneggi, J., B. Bharti, Y. Romano, and J. Sulam (2023). Shap-xrt: The shapley value meets conditional independence testing. (Cited on page 23.)

United States Census Bureau (2024). Household pulse survey: Measuring emergent social and economic matters facing u.s. households. https://www.census.gov/data/experimental-data-products/

household-pulse-survey.html. Accessed: 2024-01-30. (Cited on page 10.)