Persistent Sadness and Hopelessness in Youth: A Supervised Machine Learning Classification Approach to Identify Risk Factors

Catalina Cañizares[1], Accacia Russell[1], Mark J. Macgowan[1], Gabriel Odom[1]

[1] Robert Stempel College of Public Health and Social Work, Florida International University

Introduction

The Centers for Disease Control and Prevention (CDC) recently released the Youth Risk Behavior Survey Data Summary and Trends Report: 2011-2021 (CDC, 2023). This is the first YRBS data to be collected since the beginning of the pandemic and the findings are particularly worrying. Three out of five teenage girls reported persistent feelings of sadness or hopelessness, which is double the rate for boys and the highest level reported in a decade. This highlights the urgent need to better understand hopelessness and reverse this negative trend, ensuring that youth have the support they need to be healthy and thrive.

Numerous studies have consistently demonstrated the strong association between feelings of hopelessness and mental health issues, such as depression and suicidality (Park & Kim, 2018). Various factors have been identified as contributing to feelings of hopelessness, including school-related stress (Lane & Miranda, 2018), adapting to academic expectations (Lum et al., 2019), socioeconomic status (Kaniuka et al. (2020), generational trauma (Jelsma et al. (2022), childhood trauma (Aafjes-van Doorn et al. (2020), personality (Kelliher-Rabon et al. (2022), and negative life events (Hirsch et al. (2021).

It is crucial to understand these risk factors to prevent persistent sadness and hopelessness in youth and its associated mental health consequences. However, traditional research has focused on individual risk factors in isolation, limiting the understanding of how multiple variables interact and contribute to feelings of hopelessness. As a result, programs may not consider the effects of interacting risk factors. Therefore, there is a pressing need to develop more complex models that account for the presence of multiple risk factors to better understand the complex interplay between these variables.

Objective
This study aims to analyze multiple risk factors using supervised machine learning methods to accurately predict sadness and hopelessness and identify the principal risk factors that contribute to these negative emotions.

Methods

- Data from the 2019 YRBSS was used for this study.
- The sample consisted of 13,491 adolescents.
- The outcome variable, feelings of hopelessness, was measured using a binary response to the question: "In the past 12 months, have you experienced such intense sadness or hopelessness for two or more weeks in a row that you stopped engaging in your usual activities?"
- The model incorporated **58** variables related to behaviors that contribute to unintentional injuries and violence, sexual behaviors, alcohol and other drug use, tobacco use, unhealthy diet, and physical activity. Age, sex, race/ethnicity, and sexual orientation were included as covariates during analysis.

Analysis plan

- To identify the principal risk factors related to feelings of hopelessness, we used the *Least Absolute Shrinkage and Selection Operator (Lasso)*, *Random Forest*, and *Xgboost* algorithms.
- The algorithm was trained and tested using 75% and 25% of the sample, respectively.
- To prevent overfitting, we employed 5-fold cross-validation during the tuning process.

Results

In Table 1, we present the demographic characteristics of the study sample stratified by their hopelessness status. The adolescents analyzed in this study had a mean age of 16.1 years, with a standard deviation of 1.23, a minimum age of 12, and a maximum age of 18. Of the total sample, 37% of adolescents reported persistent sadness and hopelessness. Our findings indicate that teenage girls and individuals who identify as LGBTQ+ are significantly more vulnerable to experiencing feelings of hopelessness, with teenage girls reporting twice the amount of hopelessness compared to their male counterparts and more than half of the adolescents who identified as LGBTQ+ reporting persistent sadness. Additionally, multiple Hispanic and multiple non-Hispanic adolescents reported more feelings of hopelessness.

Table 2 displays the performance of the three machine learning models employed in our study, all of which showed high predictive accuracy. XGboost and Random Forest exhibited the best overall performance with an AUC of 0.77 and accuracy of 0.73, followed closely by Lasso with an AUC of 0.76 and accuracy of 0.73. However, our models showed an imbalance between sensitivity and specificity, indicating high sensitivity but low specificity. Our findings suggest that XGboost outperforms the other models, demonstrating the highest Kappa specificity and Log-loss scores.

Furthermore, we extracted the most important variables to predict hopelessness. As shown in Figure 1, bullying and cyberbullying, hours of sleep, being assigned male at birth, and identifying as LGBTQ+ were consistently ranked among the most important predictors for the outcome classification.

Finally, the LASSO model was also effective in identifying the key variables that contribute to hopelessness, as well as protective factors against it. Figure 2 illustrates that being bullied in school, identifying as LGBTQ+, and experiencing cyberbullying were significant variables used to classify students in the hopelessness group. In contrast, being assigned male at birth, obtaining more sleep, and having breakfast were the most crucial predictors for classifying an adolescent as non-hopeless.

Table 1

Demographic Characteristics of Adolescents by Hopelessness Status

	Persistent Sadness and Hopelessness			
	No %	Ratio No/Yes	Yes %	Overall
Sex				
Female	53.36	1.14	46.64	6,616
Male	73.21	2.73	26.79	6,759
Missing for Sex	51.72	1.07	48.28	116
Race				
White	63.99	1.78	36.01	6,726
Multiple-Hispanic	58.68	1.42	41.32	2,217
Black or African American	68.48	2.17	31.52	1,577
Hispanic/Latino	62.47	1.66	37.53	1,191
Multiple-Non-Hispanic	54.78	1.21	45.22	586
Other	66.58	1.99	33.42	793
Missing Race/Ethnicity	64.84	1.84	35.16	401
Sexual Orientation				
Heterosexual	67.77	2.10	32.23	10,725
LGBQT	39.10	0.64	60.90	1,967
Missing for Sexual Orientation	62.70	1.68	37.30	799
Feeling Hopeless	63.29	1.72	36.71	13,491

Age: Mean = 16.1, Median = 16, Min = 12, Max = 18, SD = 1,23
Results from the Youth Risk Behavior Survey (YRBS) 2019

Table 2

Performance Metrics for the LASSO, Random Forest and Xgboost Models
Predicting Sadness and Hopelesseness for 13,491 Adolescents Using the YRBS 2019 Data

Method	Kappa	Accuracy	Sensitivity	Specificity	AUC	Log loss
Lasso	0.371	0.731	**0.895**	0.447	0.769	0.548
Random Forest	0.388	**0.733**	0.875	0.488	**0.772**	0.552
XGboost	**0.389**	0.731	0.861	**0.508**	0.771	**0.543**

Kappa: value < 0 no agreement, 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial Landis and Koch (1977)
Accuracy, Sensitivity, Specificity, and AUC: Have a range of [0, 1]. The greater the value, the better is the performance of the model
Log loss is a measure of the performance of a classification model. A perfect model has a log loss of 0.

Figure 1

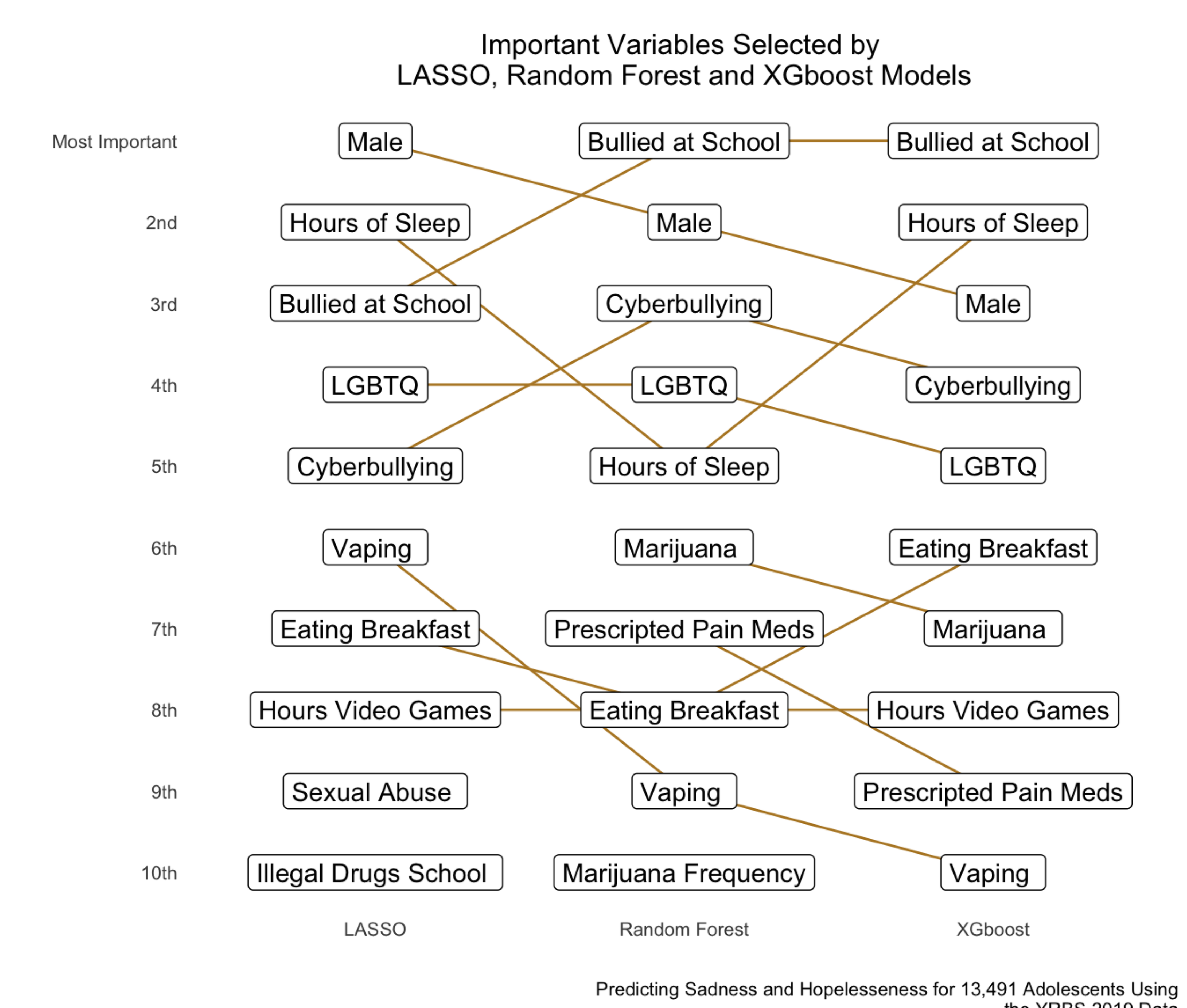
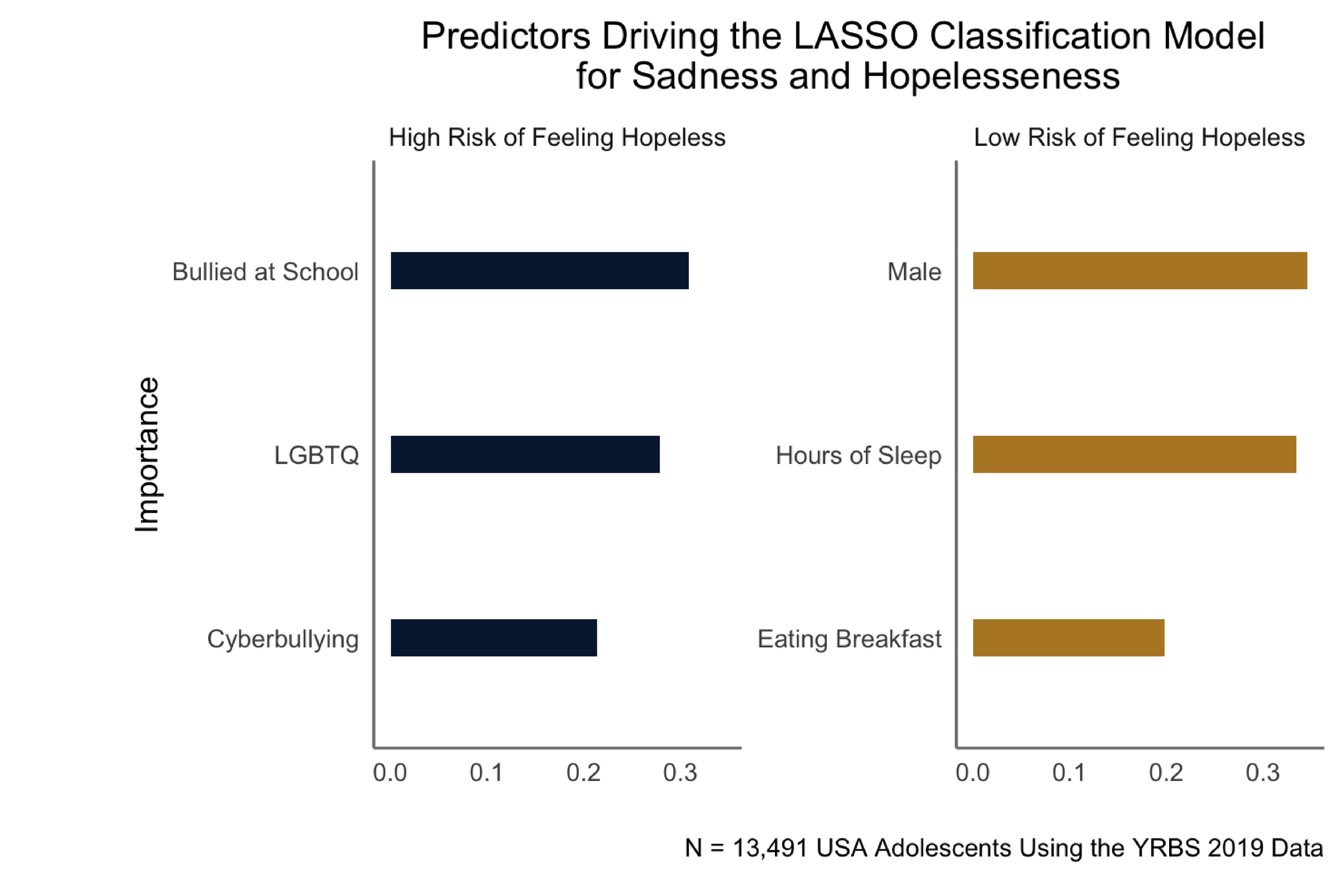


Figure 2



Discussion

The present study successfully developed three distinct supervised machine learning algorithms that identified the primary risk factors for persistent sadness and hopelessness among adolescents from a set of multiple risk factors. The models consistently identified five main variables as risk factors, which aligns with past research indicating the correlation between sleep disturbances, trauma, and stressful life events with an increased likelihood of hopelessness (Li et al., 2020; Lane & Miranda, 2018). The study also found that adolescent girls are more affected than their male counterparts and that LGBQT+ youth are substantially more likely to experience persistent sadness and hopelessness, which highlights the need for selective interventions to be developed specifically for this group. By targeting these factors and incorporating them into school-based programs, struggling youth could potentially be provided with a lifeline.

These results are essential in calling for more time and attention to be committed to addressing the issue of hopelessness during intervention given its increasing prevalence and significant relationship with suicide ideation and attempts (Park & Kim, 2018). Understanding correlated risk factors is crucial to help clinicians better identify and target at-risk adolescents.

In conclusion, this study represents a significant step forward in acknowledging hopelessness as a crucial research concern and identifying factors that must be considered in prevention and intervention efforts. This innovative methodological approach underscored multiple risk factors that increase the likelihood of persistent sadness and hopelessness, confirmed the high presence of hopelessness among teenage girls and LGBQT+ youth, and highlighted consistent risk factors that should be included in prevention efforts.

Limitations

It is important to note, that this study has certain limitations. The construction of the algorithm in this study did not consider the imbalance of the outcome, leading to better sensitivity than specificity. Having better sensitivity than specificity means that our model is more likely to correctly identify adolescents who will feel hopeless, but it may also generate a higher number of false positives. Future research should employ techniques such as downsampling to improve model performance.

References