

Original Paper

Identifying Psychosocial and Ecological Determinants of Enthusiasm In Youth: Integrative Cross-Sectional Analysis Using Machine Learning

Roberta M Dolling-Boreham^{1,2}, BAsC, MD; Akshay Mohan^{2,3}, BTech, SM; Mohamed Abdelhack², MSc, PhD; Tara Elton-Marshall^{4,5}, PhD; Hayley A Hamilton^{5,6,7}, PhD; Angela Boak⁵, MA; Daniel Felsky^{2,6,8,9}, BSc, PhD

¹Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada

²Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health, Toronto, ON, Canada

³Center for Industrial Relations and Human Resources, University of Toronto, Toronto, ON, Canada

⁴School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

⁵Institute for Mental Health Policy Research, Centre for Addiction and Mental Health, Toronto, ON, Canada

⁶Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

⁷Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada

⁸Department of Psychiatry, University of Toronto, Toronto, ON, Canada

⁹Institute of Medical Science, University of Toronto, Toronto, ON, Canada

Corresponding Author:

Daniel Felsky, BSc, PhD

Krembil Centre for Neuroinformatics

Centre for Addiction and Mental Health

250 College Street

Toronto, ON, M5T 1R8

Canada

Phone: 1 416 535 8501 ext 33587

Email: daniel.felsky@camh.ca

Abstract

Background: Understanding the factors contributing to mental well-being in youth is a public health priority. Self-reported enthusiasm for the future may be a useful indicator of well-being and has been shown to forecast social and educational success. Typically, cross-domain measures of ecological and health-related factors with relevance to public policy and programming are analyzed either in isolation or in targeted models assessing bivariate interactions. Here, we capitalize on a large provincial data set and machine learning to identify the sociodemographic, experiential, behavioral, and other health-related factors most strongly associated with levels of subjective enthusiasm for the future in a large sample of elementary and secondary school students.

Objective: The aim of this study was to identify the sociodemographic, experiential, behavioral, and other health-related factors associated with enthusiasm for the future in elementary and secondary school students using machine learning.

Methods: We analyzed data from 13,661 participants in the 2019 Ontario Student Drug Use and Health Survey (OSDUHS) (grades 7-12) with complete data for our primary outcome: self-reported levels of enthusiasm for the future. We used 50 variables as model predictors, including demographics, perception of school experience (i.e., school connectedness and academic performance), physical activity and quantity of sleep, substance use, and physical and mental health indicators. Models were built using a nonlinear decision tree-based machine learning algorithm called extreme gradient boosting to classify students as indicating either high or low levels of enthusiasm. Shapley additive explanations (SHAP) values were used to interpret the generated models, providing a ranking of feature importance and revealing any nonlinear or interactive effects of the input variables.

Results: The top 3 contributors to higher self-rated enthusiasm for the future were higher self-rated physical health (SHAP value=0.62), feeling that one is able to discuss problems or feelings with their parents (SHAP value=0.49), and school belonging (SHAP value=0.32). Additionally, subjective social status at school was a top feature and showed nonlinear effects, with benefits to predicted enthusiasm present in the mid-to-high range of values.

Conclusions: Using machine learning, we identified key factors related to self-reported enthusiasm for the future in a large sample of young students: perceived physical health, subjective school social status and connectedness, and quality of relationship

with parents. A focus on perceptions of physical health and school connectedness should be considered central to improving the well-being of youth at the population level.

(*JMIR Public Health Surveill* 2024;10:e48705) doi: [10.2196/48705](https://doi.org/10.2196/48705)

KEYWORDS

subjective well-being; Ontario Student Drug Use and Health Survey (OSDUHS); machine learning; Shapley additive explanations (SHAP); extreme gradient boosting (XGBoost); psychosocial; ecological; determinants; enthusiasm; mental health; well-being; youth; public health; student; self-reported

Introduction

Mental illnesses among youth account for up to 70% of all disability-adjusted life years [1] in high-income countries, and 20% of Canadian youth experience symptoms of mental illness, with many needing medical intervention [2]. Although there is no universally agreed upon definition of youth, the traditional definition refers to those aged 10-19 years [3] but can vary across cultures, often reflecting social responsibilities, and encompass those up to the age of 25 years [4]. Over 75% of mental disorders occur before the age of 25 years [5], with this period being crucial for developing the skills and habits necessary for lifelong mental well-being. Mental illness in youth can lead to long-lasting negative physical, psychological, and social impacts on individuals and their families [6,7]. Therefore, much work has been dedicated to understanding the behavioral, experiential, environmental, and demographic correlates of mental illness in youth. However, mental health is not defined as merely the absence of illness [8], and much less effort has been spent on identifying the multifactorial correlates of wellness [9].

Understanding adolescent well-being at a population level rather than evaluating mental illness in clinical settings is critical for implementing universal mental health promotion measures. Population-wide studies on well-being thus far highlight the influence of social connectedness [10,11] and competence in social, occupational, and academic domains [11]. Other studies suggest that depression and anxiety symptoms [12,13] and chronic disease [12] have important influence on well-being.

Enthusiasm for the future (herein “enthusiasm”) is an established component of mental wellness [14] and is among the attitudes gauged with the Positive and Negative Affect Schedule scale [15] to measure positive affect. Enthusiasm is strongly correlated with mental wellness [16,17] and is able to independently predict other aspects of well-being, including life satisfaction, positivity, personal growth, improved social connections, and enhanced self-purpose [18]. However, the multifactorial and complex nature of youth well-being [1] makes the determination of policy and programming with the most potential for impact extremely challenging. Thus, understanding the major components of youth well-being, and therefore, potential targets of public health intervention call for methodologies that inherently manage such complexities.

Machine learning is an excellent tool for analyzing complex, multifactorial data across several domains, including public health, diagnostics, treatment, and prognosis [19]. Interpretable methods such as extreme gradient boosting (XGBoost) [20] and

Shapley additive explanations (SHAP) [21] gaining popularity, are nonlinear interactive models capable of identifying important features. This approach has been successfully applied to population-based studies to better understand multifactorial contributors to mental illness [22]. However, few studies have applied machine learning to population-based data to investigate mental well-being [23]. In this study, we used a gradient-boosted tree-based machine learning algorithm to better understand multifactorial contributors to self-reported enthusiasm in 13,661 participants from the 2019 Ontario Student Drug Use and Health Survey (OSDUHS). We then applied SHAP analysis to explain the resulting models.

Methods

Survey Data Collection

Data from the 2019 cycle of OSDUHS were used for analyses. OSDUHS is a biennial cross-sectional voluntary survey conducted among students in grades 7-12 attending publicly funded schools within Ontario, with the goal of collecting data on their physical, mental, and social well-being and the prevalence of self-reported risk behaviors such as gambling and drug use. OSDUHS data have been widely used to help direct policy decision-making, public programming aimed at supporting youth, and setting priorities for enhancing the health of youth [24]. Administered since 1977, OSDUHS uses a 2-stage cluster sample design based on a random selection of schools stratified by region and school level (elementary and secondary) and a random selection of classes within each school [24].

The 2019 cycle of OSDUHS, administered in 263 schools across 47 school boards within Ontario, had a sample of 14,142 students. The cycle administered 4 versions of the survey: 2 versions for elementary students (grades 7-8) and 2 for secondary school students (grades 9-12). Each version consisted of a set of core questions that were common across all 4 surveys regarding demographics, perception of school experience, family life, substance use, physical activity, hours of sleep, and other physical and mental health indicators. Responses to these core questions constituted the set of features used in our analysis. Survey weights were not used in our analysis. Further details about the 2019 OSDUHS are publicly available [24].

Definition of the Primary Outcome

Our primary outcome of interest was student self-reported level of agreement to the question “I feel enthusiastic about my future” [25], answered on a 4-point Likert scale (0: strongly disagree to 3: strongly agree). This single Likert-type item of self-reported enthusiasm was used as an outcome given its simplicity and interpretability when building a machine learning

model, while also being cognizant of its limitations in being able to capture the multidimensional nature of the well-being construct. This question was posed to all students, maximizing the included sample, and enthusiasm is a known contributor to well-being and positive affect [14,15]. Furthermore, the Likert scale is a fundamental psychometric tool used to quantify qualitative attributes of the human experience [26]. A Likert scale style survey is easier to administer and minimizes the participant burden, increasing the likelihood of survey completion. Of the total sample, 13,661 student participants provided a response to self-reported enthusiasm and were included in our modelling (481/14,142, 3.5% missing). Raw values and 2 transformations of this variable were used to fit 3 different models, as described in the Statistical Modelling section below. Likert-style outcomes have been used in previous machine learning classification studies [27,28].

Processing of Model Inputs

A complete set of 50 variables was selected from the aforementioned common core questions to maximize the number of student responses that could be included for model training and testing. A range of data types were used in the analysis, including numeric, ordinal, and unordered categorical. Variables such as household composition (family members present in the home environment), racial background, and geographic region within Ontario were one-hot encoded (the process of splitting a categorical variable into multiple binary dummy variables with either yes or no membership), a method often used for categorical variables in tree-based learning models. Ordinal variables were treated as numeric (the machine learning algorithm selected for analysis was XGBoost, later described, which is a nonlinear model and therefore makes no scalar assumptions about levels of ordinal variables). Some variables such as the 10-item responses to “How often did you drink alcohol in the last 12 months?” were collapsed into binary responses (yes vs no) corresponding to whether any alcohol was consumed in the last 12 months. These collapsed variables were created by the OSDUHS team and were provided to us in the original OSDUHS data set [24]. This was the case for all questions pertaining to substance use and was done based on the very low response rates—largely in categories of substance use—for extreme values of some variables.

Additional processing involved re-encoding and collapsing of categories for ease of interpretation. First, any response option that could be interpreted as a “no” was collapsed into a common category. For example, the 2 response options “use internet, but not social media,” and “don’t use the internet” were combined to “don’t use social media” for the question “How many hours do you typically spend on social media?” Second, the weight of the student was considered to be the average of the 2-kg range options provided to students. Third, the many categories for language spoken were condensed to categories of “English only,” “French only,” “English and French only,” and “other multiple languages.” All encoded ordered categorical variables were shifted to start with a value of zero. Finally, the remaining categorical variables were one-hot encoded (0=no, 1=yes for category membership), with “don’t know” or “unsure” responses being treated as missing. Crucially, XGBoost was designed to manage missing values by learning appropriate tree branches

during the training process and implementing a default mechanism for evaluating new data, which negates the need for a priori imputation. A complete descriptive list of the input variables used in our analysis can be found in Table S1 in [Multimedia Appendix 1](#). Variable transformations are detailed in Table S2 in [Multimedia Appendix 1](#).

Statistical Modelling With XGBoost

Python (v 3.8.13; Python Software Foundation) was used for all analyses. Three classification models were generated using the XGBoost algorithm [20], with the outcome being class labels derived from self-reported enthusiasm, modelled by our set of 50 input variables. XGBoost is a decision tree-based algorithm, whereby a set of trees are initialized and iterated during model training to improve fit. The weighted average of the output of each tree within the set is considered as the prediction of the trained model. XGBoost was selected due to its ability to handle linear, nonlinear, and interactive effects between predictor variables.

A total of 3 classification models were built, each for a different grouping of responses to our primary outcome of enthusiasm. These 3 models were specified as follows:

1. Multiclass classification: Outcome being the 4 classes of the original survey responses, that is, strongly disagree (470/13,661, 3.44%), somewhat disagree (1388/13,661, 10.16%), somewhat agree (6331/13,661, 46.34%), and strongly agree (5472/13,661, 40.06%) (N=13,661).
2. Binary classification: Outcome being binary with the 2 classes being students who chose strongly agree (5472/13,661, 40.06%) and students who chose any other response (8189/13,661, 59.94%) (N=13,661).
3. Binary classification: Outcome being binary with the first class being “enthusiastic” (5472/7330, 74.7%) comprising students who selected “strongly agree” and the second class being “not enthusiastic” (1858/7330, 25.3%) comprising students who chose either “strongly disagree” or “somewhat disagree.” Students who chose “somewhat agree” were removed from this analysis to improve class discrimination, as it was considered more likely to represent ambivalence when compared to a “somewhat disagree” response due to acquiescence bias [29]. Those who responded “somewhat disagree” were included to compensate for the low percentage of students in the “strongly disagree” class (n=7330).

In the binary classification of model 3, a sensitivity analysis to demonstrate ambivalence is reported in [Multimedia Appendix 1](#). The data were divided into 2 nonoverlapping subsets through random sampling, stratified by outcome categories (to maintain balance in outcome groups), with 80% as training data and 20% as withheld test data to evaluate unbiased model performance. Random oversampling of the outcome group with fewer observations—within only the training data—was used to help mitigate distributional imbalances (using the imbalanced-learn Python library). The area under the receiver operating characteristic curve (AUROC) was used to evaluate the model’s accuracy on the validation data during the fitting process due to its scale and classification-threshold invariance. XGBoost automatically handled missing data by assignment of a default

direction at each decision tree node such that loss of AUROC was minimized. The optimal hyperparameters for the XGBoost algorithm, which are values that control the XGBoost process and are selected before model training, were selected through Bayesian optimization using the Hyperopt Python library [30]. Hyperparameters that were optimized can be seen in Table S3 in [Multimedia Appendix 1](#). Accuracy, precision, and recall were also calculated for each model within test data, and confusion matrices were used to visualize model performance.

Determination of Variable Importance With SHAP

Given that XGBoost is not inherently interpretable, the importance of each input feature in model classification was determined by calculating the absolute mean SHAP values [21] based on the test data. A positive SHAP value indicated that the variable had a positive influence on the outcome (pushing the model classification toward enthusiastic), with a greater magnitude indicative of a greater impact on the model output. In addition, following our previously published approach [22], SHAP values for the interactions among the top 15 input features were calculated to determine the importance of interaction among these variables on classification in comparison to the importance of the individual variables. SHAP analysis was performed on all models; however, the best model based on the aforementioned metrics was selected to present detailed results and structure discussion.

Ethics Approval and Recruitment

The 2019 OSDUHS was approved by the institutional review boards at the Center for Addiction and Mental Health, York University, and 34 Ontario school boards. Schools were randomly selected. After seeking approval from relevant school boards, the randomly selected schools were invited to participate in the survey. Schools that could not participate were replaced by schools within the same stratum. Once a school was approved, 1 or 2 classes in the relevant grades were randomly selected from a master list of all classes. Students in the selected classes were given a parental consent–student assent form to take home to parents, which explained the survey’s purpose and method. Only students with parental consent could participate. Students completed the survey in the classroom during regular school hours. The survey data were anonymous [24]. Generative artificial intelligence was not used in the creation of this manuscript. Analyses of OSDUHS data were approved by the CAMH Research Ethics Board (CAMH 099/2019).

Results

Participant Characteristics

[Table 1](#) summarizes the demographic characteristics of the survey participants included in our study. The sample had a mean age of 14.9 (SD 1.8) years, was well balanced for biological sex assigned at birth (7612/13,661, 55.72% female), and included a majority of individuals who self-identified as White (8617/13,661, 63.08%).

Table 1. Baseline characteristics of the respondents used in the modelling process (N=13,661).

Characteristics	Values ^a
Age (years)	
Mean (SD)	14.86 (1.77)
Range (min-max)	9 (11-20)
Missing responses (n)	5
Grade	
Mean (SD)	9.61 (1.65)
Range (min-max)	5 (7-12)
Missing responses	10
Sex at birth, n (%)	
Female	7612 (55.72)
Male	6049 (44.28)
Region, n (%)	
Greater Toronto Area (GTA)	5257 (38.48)
Northern Ontario	918 (6.72)
Western Ontario	4382 (32.08)
Eastern Ontario	3104 (22.72)
Race/ethnicity (respondents were allowed to select more than one), n (%)	
White	8617 (63.08)
Chinese	766 (5.61)
South Asian	1265 (9.26)
Black	1293 (9.46)
Indigenous	378 (2.77)
Filipino	736 (5.39)
Latin/Central/South American	584 (4.27)
Southeast Asian	259 (1.90)
West Asian/Arab	731 (5.35)
Korean	124 (0.91)
Japanese	64 (0.47)
Missing responses	67 (0.49)

^aThe percentages shown are not weighted.

XGBoost Modelling of Youth Enthusiasm

Table 2 summarizes the key performance metrics for the 3 models built using XGBoost: the multiclass classification of all enthusiasm responses (model 1), the binary classification of enthusiastic respondents against all others (model 2), and the binary classification of enthusiastic versus not enthusiastic respondents, excluding ambivalent classes (model 3). For each of the 3 models, the accuracy, precision, and recall metrics were highly similar, being within 1% of each other, indicating that the model classified a similar number of false positives as false negatives [31]. The AUROC metrics were 0.68 for model 1, 0.75 for model 2, and 0.86 for model 3. The AUROC metrics are shown in **Table 2**, which highlights the calculated performance metrics for each model on the withheld test data.

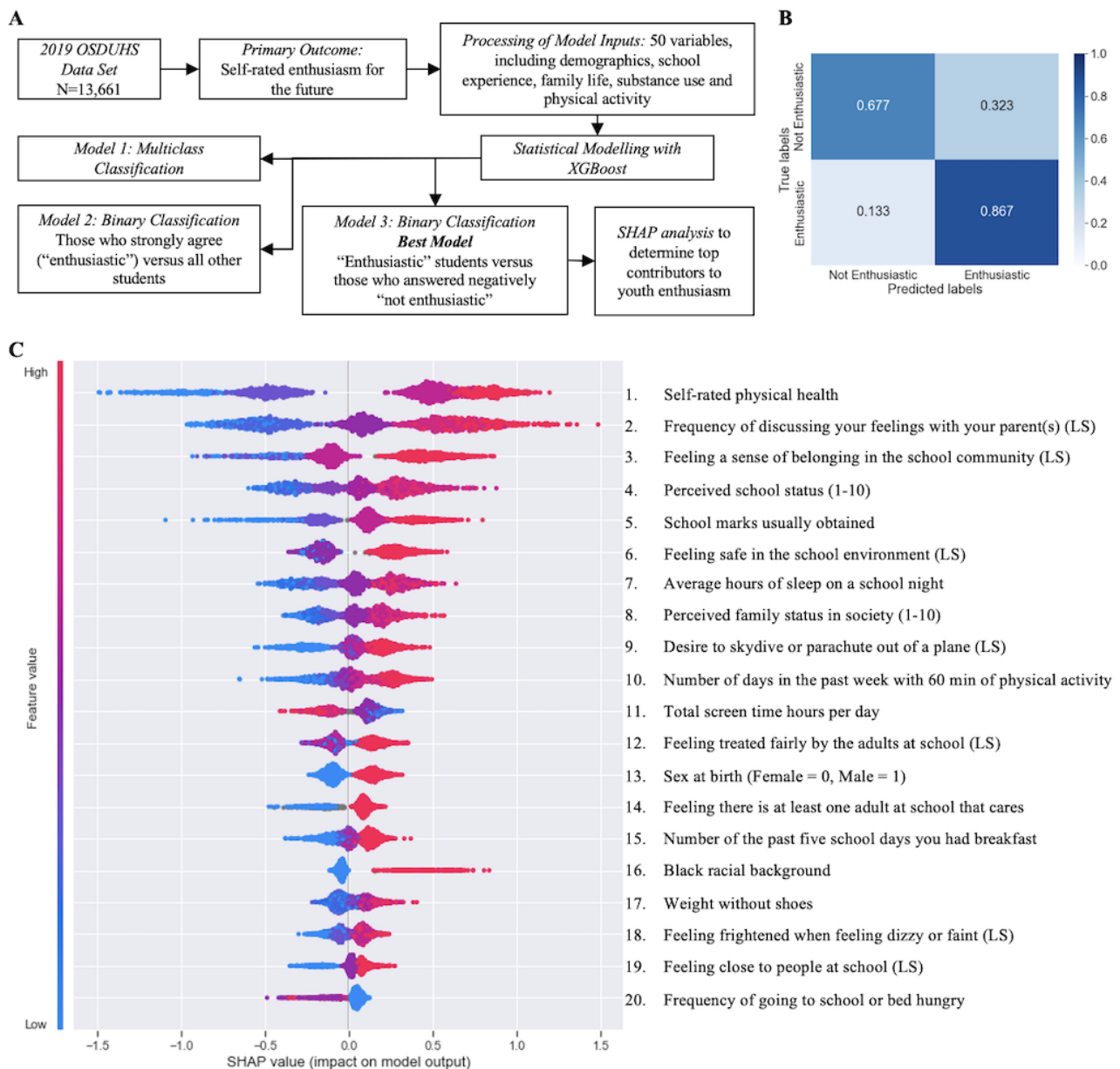
Model 3 was selected as the best classification model for self-reported enthusiasm using the set of 50 input features, as it had the best performance across accuracy (0.81, 95% CI 0.79-0.83), precision (0.82, 95% CI 0.80-0.84), and recall (0.81, 95% CI 0.79-0.83) in held-out test data. The confusion matrix depicting model 3 performance is shown in **Figure 1B**. Confusion matrices for models 1 and 2 are available in **Figures S1-S2 in Multimedia Appendix 1**. Optimal hyperparameters for our trained model are provided in **Table S3 in Multimedia Appendix 1**. The superior performance of this model when compared to those of model 1 (accuracy 0.53, 95% CI 0.51-0.55) and model 2 (accuracy 0.68, 95% CI 0.66-0.70) is likely due to the exclusion of potentially ambivalent respondents, polarizing the sample and offering more contrast between classes for training.

Table 2. Performance metrics calculated for each model on withheld test data.

	Training data (80% of the data), n	Test data (20% of the data), n	AUROC ^a	Accuracy (95% CI)	Precision (95% CI)	Recall (95% CI)
Model 1: multiclass classification	10,888	2773	0.68	0.53 (0.51-0.55)	0.52 (0.50-0.54)	0.53 (0.51-0.55)
Model 2: binary classification (enthusiastic vs all others)	10,888	2773	0.75	0.68 (0.66-0.70)	0.69 (0.67-0.71)	0.68 (0.66-0.70)
Model 3: binary classification (enthusiastic vs not enthusiastic)	5864	1466	0.86	0.81 (0.79-0.83)	0.82 (0.80-0.84)	0.81 (0.79-0.83)

^aAUROC: area under the receiver operating characteristic curve.

Figure 1. Results of the extreme gradient boosting model of binarized self-reported enthusiasm (model 3, n=7330). A: Simplified flow diagram of the study design. B: Confusion matrix for model 3, with elements normalized to the true prediction class population sizes. The main diagonal cells indicate predictions that match the true labels. C: Shapley additive explanations summary plot for model 3 showing the top 20 variables ranked by mean absolute Shapley additive explanations values on test data. Each point represents an individual student’s response to the question listed, the color of that point represents the actual value of the response, and the horizontal position of the point on the figure represents the impact that answer had on the predicted outcome. The further right on the figure the point is (indicating a higher SHAP value), the more positive impact it had toward predicting “enthusiastic,” as opposed to “not enthusiastic.” LS: Likert scale; OSDUHS: Ontario Student Drug Use and Health Survey; SHAP: Shapley additive explanations; XGBoost: extreme gradient boosting.



Identifying the Most Important Variables With SHAP

SHAP analysis ranked the most important features used by the model for classification. Figure 1B shows the top 20 input variables, as determined by the mean of the absolute values of their SHAP values in the test data. Figure 2 illustrates the impact of the response of each individual student on model outputs for the top 5 input features within the test data. The most important

contributor to youth enthusiasm based on SHAP analysis was self-rated physical health. In general, themes surrounding physical health, family relationships, and school experience appeared among the top input features across all 3 models. Details for the SHAP values of the top 5 input features in each model are listed in Table 3. A complete list of SHAP values for model 3 is available in Table S4 in Multimedia Appendix 1.

Figure 2. Shapley additive explanations values plotted for the top 5 variables predictive of binarized enthusiasm (model 3, n=7330). Each point represents a single student's response to the question plotted against its corresponding Shapley additive explanations value (impact it had on predicting enthusiastic over not enthusiastic). Positive Shapley additive explanations values indicate more impact toward "enthusiastic," whereas negative Shapley additive explanations values indicate more impact toward "not enthusiastic." A line of best fit (locally weighted scatterplot smoothing smooth curve) was added to each plot only to demonstrate the overall trend; however, they were not fit using the underlying model and are not meant to represent statistical significance. A: Self-rated physical health (most impactful feature). B: "Talk about your problems or feelings with parent(s) (Likert scale)" (second most impactful feature). C: "I feel like I am part of this school (Likert scale)" (third most impactful feature on model prediction). D: School status on a scale of 1-10 (fourth most impactful feature). E: School marks usually obtained (fifth most impactful feature). Blue dashes on the y-axis indicate observations for which the variable indicated on the x-axis were missing. LS: Likert scale; SHAP: Shapley additive explanations.

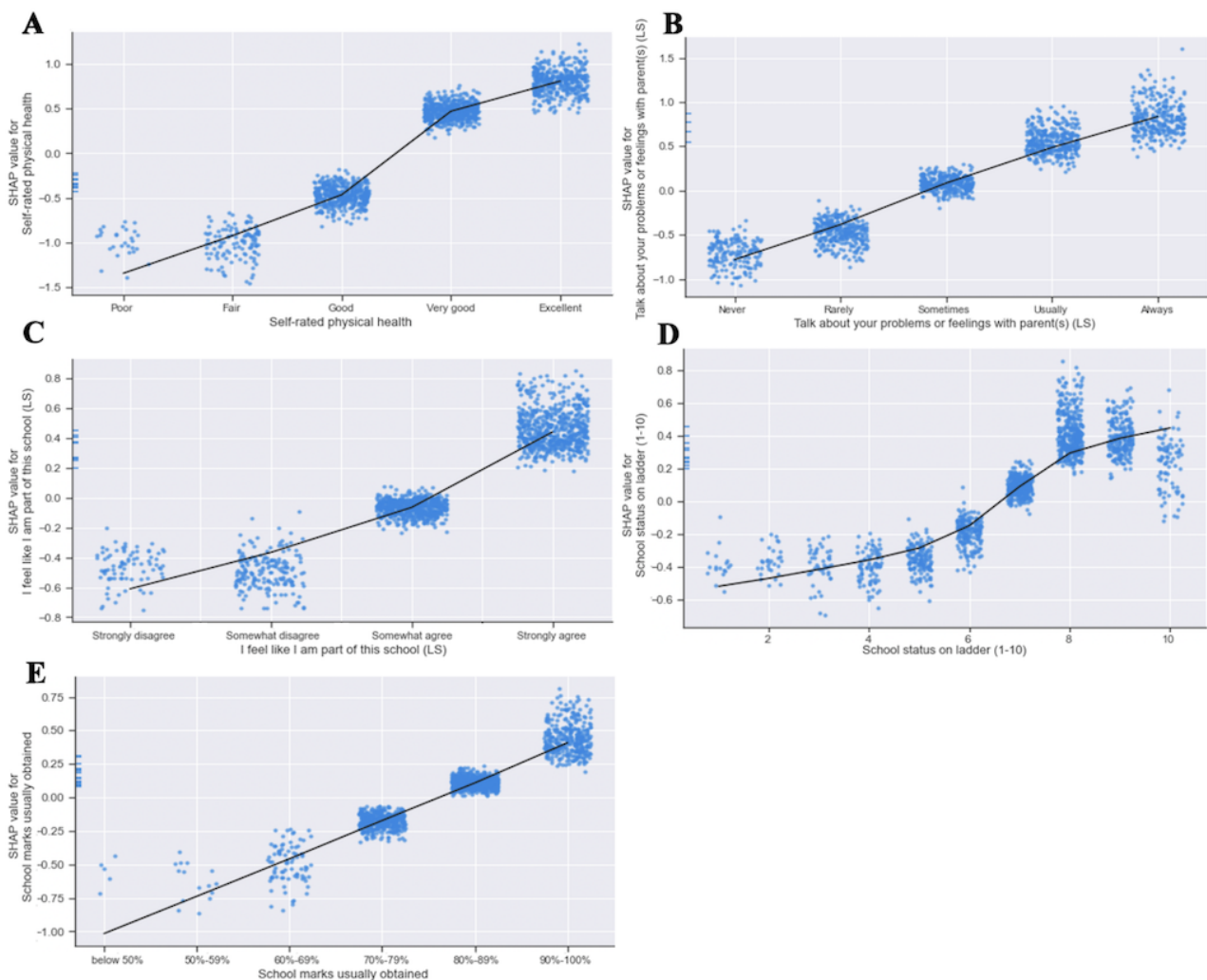


Table 3. Top 5 variables identified by magnitude of importance (rank) for all 3 models.

Rank	Model 1: multiclass classification		Model 2: binary classification (enthusiastic vs all others)		Model 3: binary classification (enthusiastic vs not enthusiastic)	
	Input feature	Mean absolute SHAP ^a value	Input feature	Mean absolute SHAP value	Input feature	Mean absolute SHAP value
1	Self-rated physical health	0.299	Feeling comfortable sharing one's thoughts or feelings with their parents	0.266	Self-rated physical health	0.621
2	Feeling comfortable sharing one's thoughts or feelings with their parents	0.222	Self-rated physical health	0.245	Feeling comfortable sharing one's thoughts or feelings with their parents	0.494
3	Feeling a sense of belonging in the school community	0.179	School marks usually obtained	0.203	Feeling a sense of belonging in the school community	0.322
4	Perceived school status on a scale of 1-10	0.173	Feeling a sense of belonging in the school community	0.164	Perceived school status on a scale of 1-10	0.260
5	Perceived family status in society on a scale of 1-10	0.126	Perceived school status on a scale of 1-10	0.137	School marks usually obtained	0.231

^aSHAP: Shapley additive explanations.

Assessment of Variable Interactions

Examination of the impact of pairwise variable interactions on model classification indicated that the interactions between input features did not substantially affect the model output, with mean absolute SHAP values for these interactions being lower than the mean absolute SHAP values of any of the top 15 most important explanatory variables in each of the 3 tested models. For our top model (model 3), the most important interaction with a SHAP value of 0.054 was between “talk about your problems or feelings with your parents” and “self-rated physical health.” Variable interaction data are presented in Figures S3-S8 in [Multimedia Appendix 1](#).

Discussion

We used a gradient-boosted tree-based machine learning algorithm to classify self-reported youth enthusiasm (as an indicator of well-being [14,15]) and identify the most important contributing input features by using the population-level OSDUHS, conducted among elementary and secondary school students attending publicly funded Ontario schools. A wide range of variables were used to model enthusiasm, including sociodemographic factors, physical activity and quantity of sleep, other physical and mental health indicators, perception of school experience, and substance use. The XGBoost algorithm was used to generate 3 models to classify youth enthusiasm, with SHAP values being used to explain the importance of each input feature across the sample. The top explanatory variables in classifying enthusiasm were related to physical health, relationship with parents, and school experience. A crucial aspect of this insight is that these rankings are derived from an approach that accounts for the context of all variables in the model (ie, the coalition of variables for each participant)

in a nonlinear interactive way rather than in mass bivariate testing or specific hypothesis-driven tests.

Our ranking of top features supports a body of evidence describing the close interconnection between a person's physical and mental health—both actual and perceived [32-34]. For example, our top feature contributing to enthusiasm and by proxy well-being was self-rated physical health, which has many contributors, including self-esteem, self-awareness, and physical activity levels. Physical activity itself is a well-known protective factor against mental illness [35-37], with physical activity in youth consistently improving well-being [34,38]. Conversely, physical inactivity is a potential risk factor for mental illness, with increasing physical activity contributing to more effective treatment [37]. Actual physical activity and self-perceived physical health status are related, and both could be used to identify children at risk for associated physical and mental health problems [32]. The number of days in the past week with 60 minutes of physical activity ranked tenth in variable importance. In addition to perceived and actual physical activity, number of hours slept per night was the seventh most important feature for classifying enthusiasm, with students who slept more hours reporting greater enthusiasm and therefore potentially greater sense of well-being. Previous studies have demonstrated a significant association between poor sleep quality and poor mental health. It is also possible that sleep is a mediator of other factors that contribute to poor mental health in youth, including social media use [39]. Given our findings suggesting a connection between physical health and enthusiasm, as well as the body of evidence demonstrating the interconnection of mental and physical health, it is likely that the promotion of physical health can enhance youth well-being [36]. Thus, public health measures directed at improving youth well-being should

include the promotion of physical health and importantly, a healthy self-perception of physical fitness.

Relationships with parents and school, specifically as they relate to feeling socially supported in these environments, were also important predictors of enthusiasm, and therefore probably well-being (feeling comfortable sharing thoughts and feelings with parents, feeling a sense of belonging in the school community, feeling safe in school, and high perceived school status). The connection of family and school relationships with youth well-being has also been highlighted in several previous studies: family and teacher relationships have been demonstrated to be significantly associated with reduced substance use [40]. Peer connectedness in the school environment has also played a role with a greater sense of well-being [40]. Support from teachers and family has also been associated with significant improvements in mental well-being [41]. Additionally, it has been suggested that school environments that are structured to maximize connection between teachers and students can enhance student engagement and lead to improved student well-being [42,43]. The importance of social connectedness as it relates to youth enthusiasm may be even more relevant post COVID-19 pandemic. Although our study was conducted before the pandemic, a recent study of Canadian youth subjective well-being (an all-encompassing term for happiness, satisfaction, morale, and positive affect [44]) during the COVID-19 pandemic showed that having access to friends and areas to play was correlated with improved subjective well-being [10]. Given the connection of support within school and family environments with enthusiasm and youth well-being, increased social support should be another consideration in public health measures aimed at improving well-being in youth.

Our study has several limitations. First, our analyses were limited to the set of 50 core questions used in OSDUHS and

excluded potentially important questions and topics unique to particular split ballot versions of the questionnaire. These topics included bullying, gambling, and antisocial behaviors that could also be related to enthusiasm. Second, the possible response categories for the self-reported enthusiasm for the future question in OSDUHS did not include a neutral option. Although we believe modelling the outcome in multiple ways helped provide a more fulsome picture of enthusiasm based on different groupings, we cannot assume that all respondents in the “somewhat agree” or “somewhat disagree” categories were truly ambivalent. Third, students who did not respond to the question of enthusiasm were excluded from the analysis. It is possible that this missingness is nonrandom, with students who felt less enthusiastic about their future deciding to not participate in the survey or to respond to the question. However, missingness for this variable was low (481/14,142, 3.5% missing of the total sample). Additionally, the OSDUHS responses are self-reported, meaning that they could be affected by recall and memory. Finally, it is to be noted that the conducted analysis is cross-sectional and not causal.

In summary, we used XGBoost to identify the set of behavioral, environmental, and psychosocial factors related to self-reported enthusiasm for the future in a large sample of young students. The most important factors were perceived physical health, school social status and connectedness, and quality of parental relationships. These factors were found to have a stronger association with enthusiasm than many common intervention targets, including social media, drug, and alcohol use. With the close interconnection of enthusiasm and well-being, our findings suggest that a focus on physical health and school connectedness should be central to impactful public health programming aimed at improving the mental well-being of youth, particularly when it comes to improving enthusiasm for the future.

Acknowledgments

This study was supported by The Koerner Family Foundation New Scientist Award, The Krembil Foundation, The Centre for Addiction and Mental Health Discovery Fund, and The Canadian Institutes of Health Research.

Data Availability

The data set analyzed during this study is available upon reasonable request to the Center for Addiction and Mental Health at osduhs@camh.ca.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary data.

[\[DOCX File, 3668 KB-Multimedia Appendix 1\]](#)

References

1. Malla A, Shah J, Iyer S, Boksa P, Joober R, Andersson N, et al. Youth mental health should be a top priority for health care in Canada. *Can J Psychiatry*. Apr 2018;63(4):216-222. [FREE Full text] [doi: [10.1177/0706743718758968](https://doi.org/10.1177/0706743718758968)] [Medline: [29528719](https://pubmed.ncbi.nlm.nih.gov/29528719/)]
2. Canadian community health survey: mental health, 2012. Statistics Canada. 2013. URL: <https://www150.statcan.gc.ca/n1/daily-quotidien/130918/dq130918a-eng.htm> [accessed 2024-08-21]

3. World Health Organization. Child and adolescent mental health policies and plans. In: World Health Organization. Geneva. World Health Organization; 2005.
4. Patel V, Flisher AJ, Hetrick S, McGorry P. Mental health of young people: a global public-health challenge. *Lancet*. Apr 14, 2007;369(9569):1302-1313. [doi: [10.1016/S0140-6736\(07\)60368-7](https://doi.org/10.1016/S0140-6736(07)60368-7)] [Medline: [17434406](https://pubmed.ncbi.nlm.nih.gov/17434406/)]
5. Kessler RC, Angermeyer M, Anthony JC, DE Graaf R, Demyttenaere K, Gasquet I, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry*. Oct 2007;6(3):168-176. [FREE Full text] [Medline: [18188442](https://pubmed.ncbi.nlm.nih.gov/18188442/)]
6. Kessler RC, Heeringa S, Lakoma MD, Petukhova M, Rupp AE, Schoenbaum M, et al. Individual and societal effects of mental disorders on earnings in the United States: results from the national comorbidity survey replication. *Am J Psychiatry*. Jun 2008;165(6):703-711. [FREE Full text] [doi: [10.1176/appi.ajp.2008.08010126](https://doi.org/10.1176/appi.ajp.2008.08010126)] [Medline: [18463104](https://pubmed.ncbi.nlm.nih.gov/18463104/)]
7. Wittchen HU, Nelson CB, Lachner G. Prevalence of mental disorders and psychosocial impairments in adolescents and young adults. *Psychol Med*. Jan 1998;28(1):109-126. [doi: [10.1017/s0033291797005928](https://doi.org/10.1017/s0033291797005928)] [Medline: [9483687](https://pubmed.ncbi.nlm.nih.gov/9483687/)]
8. Fusar-Poli P, Salazar de Pablo G, De Micheli A, Nieman DH, Correll CU, Kessing LV, et al. What is good mental health? A scoping review. *Eur Neuropsychopharmacol*. Feb 2020;31:33-46. [FREE Full text] [doi: [10.1016/j.euroneuro.2019.12.105](https://doi.org/10.1016/j.euroneuro.2019.12.105)] [Medline: [31901337](https://pubmed.ncbi.nlm.nih.gov/31901337/)]
9. Gentzler AL, Root AE. Positive affect regulation in youth: taking stock and moving forward. *Social Development*. Jan 29, 2019;28(2):323-332. [doi: [10.1111/sode.12362](https://doi.org/10.1111/sode.12362)]
10. Mitra R, Waygood EOD, Fullan J. Subjective well-being of Canadian children and youth during the COVID-19 pandemic: The role of the social and physical environment and healthy movement behaviours. *Prev Med Rep*. Sep 2021;23:101404. [FREE Full text] [doi: [10.1016/j.pmedr.2021.101404](https://doi.org/10.1016/j.pmedr.2021.101404)] [Medline: [34189017](https://pubmed.ncbi.nlm.nih.gov/34189017/)]
11. Urke HB, Holsen I, Larsen T. Positive youth development and mental well-being in late adolescence: the role of body appreciation. Findings from a prospective study in Norway. *Front Psychol*. 2021;12:696198. [FREE Full text] [doi: [10.3389/fpsyg.2021.696198](https://doi.org/10.3389/fpsyg.2021.696198)] [Medline: [34497559](https://pubmed.ncbi.nlm.nih.gov/34497559/)]
12. Wang X, Jia X, Zhu M, Chen J. Linking health states to subjective well-being: an empirical study of 5854 rural residents in China. *Public Health*. Jun 2015;129(6):655-666. [doi: [10.1016/j.puhe.2015.03.014](https://doi.org/10.1016/j.puhe.2015.03.014)] [Medline: [25931439](https://pubmed.ncbi.nlm.nih.gov/25931439/)]
13. Zhang N, Liu C, Chen Z, An L, Ren D, Yuan F, et al. Prediction of adolescent subjective well-being: A machine learning approach. *Gen Psychiatr*. 2019;32(5):e100096. [FREE Full text] [doi: [10.1136/gpsych-2019-100096](https://doi.org/10.1136/gpsych-2019-100096)] [Medline: [31552391](https://pubmed.ncbi.nlm.nih.gov/31552391/)]
14. Lopez S, Snyder C. *The Oxford Handbook of Positive Psychology*. Oxford. Oxford University Press; 2009.
15. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*. 1988;54(6):1063-1070. [doi: [10.1037/0022-3514.54.6.1063](https://doi.org/10.1037/0022-3514.54.6.1063)] [Medline: [3397865](https://pubmed.ncbi.nlm.nih.gov/3397865/)]
16. Jain R, Jain S. *The Science and Practice of Wellness: Interventions for Happiness, Enthusiasm, Resilience, and Optimism (HERO)*. New York. WW Norton & Company; 2020.
17. Yaklin S, Jain R, Cole SP, Raison C, Rolin D, Jain S. HERO wellness scale: Examining a new mental wellness scale. *Ann Clin Psychiatry*. Feb 2020;32(1):33-40. [Medline: [31990967](https://pubmed.ncbi.nlm.nih.gov/31990967/)]
18. Sun J, Kaufman SB, Smillie LD. Unique associations between big five personality aspects and multiple dimensions of well-being. *J Pers*. Apr 2018;86(2):158-172. [doi: [10.1111/jopy.12301](https://doi.org/10.1111/jopy.12301)] [Medline: [28035671](https://pubmed.ncbi.nlm.nih.gov/28035671/)]
19. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. *Psychol Med*. Jul 2019;49(9):1426-1448. [doi: [10.1017/S0033291719000151](https://doi.org/10.1017/S0033291719000151)] [Medline: [30744717](https://pubmed.ncbi.nlm.nih.gov/30744717/)]
20. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 2016:785-794. [doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785)]
21. Lundberg S, Lee S. A unified approach to interpreting model predictions. In: Luxburg UV, Guyon I, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al, editors. *Advances in Neural Information Processing Systems 30*. Red Hook, NY. Curran Associates, Inc; 2018:4765-4774.
22. Hueniken K, Somé NH, Abdelhack M, Taylor G, Elton Marshall T, Wickens CM, et al. Machine learning-based predictive modeling of anxiety and depressive symptoms during 8 months of the COVID-19 global pandemic: Repeated cross-sectional survey study. *JMIR Ment Health*. Nov 17, 2021;8(11):e32876. [FREE Full text] [doi: [10.2196/32876](https://doi.org/10.2196/32876)] [Medline: [34705663](https://pubmed.ncbi.nlm.nih.gov/34705663/)]
23. Thieme A, Belgrave D, Doherty G. A systematic review of the HCI literature to support the development of effective and implementable ML systems. *ACM Trans Comput Hum Interact*. Aug 17, 2020;27(5):34:1-34:53. [doi: [10.1145/3398069](https://doi.org/10.1145/3398069)]
24. Boak A, Elton-Marshall T, Mann R, Henderson J, Hamilton H. The mental health and well-being of Ontario students, 1991-2019: Detailed findings from the Ontario Student Drug Use and Health Survey (OSDUHS). Centre for Addiction and Mental Health. 2020. URL: <https://www.camh.ca/-/media/files/pdf---osduhs/osduhs-mh-report2019-pdf.pdf> [accessed 2024-08-28]
25. Woicik PA, Stewart SH, Pihl RO, Conrod PJ. The substance use risk profile scale: a scale measuring traits linked to reinforcement-specific substance use profiles. *Addict Behav*. Dec 01, 2009;34(12):1042-1055. [doi: [10.1016/j.addbeh.2009.07.001](https://doi.org/10.1016/j.addbeh.2009.07.001)] [Medline: [19683400](https://pubmed.ncbi.nlm.nih.gov/19683400/)]
26. Joshi A, Kale S, Chandel S, Pal D. Likert scale: explored and explained. *BJAST*. Jan 10, 2015;7(4):396-403. [doi: [10.9734/bjast/2015/14975](https://doi.org/10.9734/bjast/2015/14975)]

27. Tokmic F, Hadzikadic M, Cook JR, Tcheremissine OV. Development of a behavioral health stigma measure and application of machine learning for classification. *Innov Clin Neurosci*. Jun 06, 2018;15(5-6):34-42. [[FREE Full text](#)] [Medline: [30013818](#)]
28. Rahman A, Siraji M, Khalid L, Faisal F, Nishat M, Ahmed A. Perceived stress analysis of undergraduate students during COVID-19: a machine learning approach. 2022. Presented at: 2022 IEEE 21st Mediterranean Electrotechnical Conference (MELECON); June 14-16; Palermo, Italy. [doi: [10.1109/melecon53508.2022.9843081](#)]
29. Soto CJ, John OP, Gosling SD, Potter J. The developmental psychometrics of big five self-reports: acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *J Pers Soc Psychol*. Apr 2008;94(4):718-737. [[FREE Full text](#)] [doi: [10.1037/0022-3514.94.4.718](#)] [Medline: [18361680](#)]
30. Bergstra J, Yamins D, Cox D. Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures. *Proceedings of Machine Learning Research*. 2013. URL: <https://proceedings.mlr.press/v28/bergstra13.html> [accessed 2024-08-28]
31. Juba B, Le HS. Precision-recall versus accuracy and the role of large data sets. *Proceedings of the AAAI conference on artificial intelligence*. 2019;33(01):4039-4048. [doi: [10.1609/aaai.v33i01.33014039](#)]
32. Sollerhed AC, Aпитzsch E, Råstam L, Ejlerstsson G. Factors associated with young children's self-perceived physical competence and self-reported physical activity. *Health Educ Res*. Feb 01, 2008;23(1):125-136. [doi: [10.1093/her/cym010](#)] [Medline: [17347524](#)]
33. Timo J, Sami YP, Anthony W, Jarmo L. Perceived physical competence towards physical activity, and motivation and enjoyment in physical education as longitudinal predictors of adolescents' self-reported physical activity. *J Sci Med Sport*. Sep 01, 2016;19(9):750-754. [doi: [10.1016/j.jsams.2015.11.003](#)] [Medline: [26671711](#)]
34. Kemel PN, Porter JE, Coombs N. Improving youth physical, mental and social health through physical activity: a systematic literature review. *Health Promot J Austr*. Jul 2022;33(3):590-601. [doi: [10.1002/hpja.553](#)] [Medline: [34735738](#)]
35. Harvey SB, Hotopf M, Overland S, Mykletun A. Physical activity and common mental disorders. *Br J Psychiatry*. Nov 2010;197(5):357-364. [doi: [10.1192/bjp.bp.109.075176](#)] [Medline: [21037212](#)]
36. Jacka FN, Mykletun A, Berk M. Moving towards a population health approach to the primary prevention of common mental disorders. *BMC Med*. Nov 27, 2012;10:1-6. [[FREE Full text](#)] [doi: [10.1186/1741-7015-10-149](#)] [Medline: [23186355](#)]
37. Stathopoulou G, Powers MB, Berry AC, Smits JAJ, Otto MW. Exercise interventions for mental health: a quantitative and qualitative review. *Clinical psychology: Science and Practice*. 2006;13(2):179-193. [doi: [10.1111/j.1468-2850.2006.00021.x](#)]
38. Ahn S, Fedewa AL. A meta-analysis of the relationship between children's physical activity and mental health. *J Pediatr Psychol*. May 01, 2011;36(4):385-397. [doi: [10.1093/jpepsy/jsq107](#)]
39. Alonzo R, Hussain J, Stranges S, Anderson KK. Interplay between social media use, sleep quality, and mental health in youth: A systematic review. *Sleep Med Rev*. Apr 2021;56:101414-101414. [doi: [10.1016/j.smrv.2020.101414](#)] [Medline: [33385767](#)]
40. Moore GF, Cox R, Evans RE, Hallingberg B, Hawkins J, Littlecott HJ, et al. School, peer and family relationships and adolescent substance use, subjective wellbeing and mental health symptoms in Wales: a cross sectional study. *Child Ind Res*. Dec 2018;11(6):1951-1965. [doi: [10.1007/s12187-017-9524-1](#)]
41. Bonell C, Shackleton N, Fletcher A, Jamal F, Allen E, Mathiot A, et al. Student- and school-level belonging and commitment and student smoking, drinking and misbehaviour. *Health Education Journal*. Mar 2017;76(2):206-220. [doi: [10.1177/0017896916657843](#)]
42. Markham WA, Aveyard P. A new theory of health promoting schools based on human functioning, school organisation and pedagogic practice. *Soc Sci Med*. Mar 2003;56(6):1209-1220. [doi: [10.1016/s0277-9536\(02\)00120-x](#)] [Medline: [12600359](#)]
43. Rowe F, Stewart D, Patterson C. Promoting school connectedness through whole school approaches. *Health Education*. 2007;107(6):524-542. [doi: [10.1108/09654280710827920](#)]
44. Diener E. Subjective well-being. In: *The Science of Well-Being*. Berlin. Springer Nature; 2009:11-58.

Abbreviations

AUROC: area under the receiver operating characteristic curve

OSDUHS: Ontario Student Drug Use and Health Survey

SHAP: Shapley additive explanations

XGBoost: extreme gradient boosting

Edited by A Mavragani; submitted 03.05.23; peer-reviewed by Y Yang, M Marques da Cruz; comments to author 07.12.23; revised version received 27.01.24; accepted 16.05.24; published 12.09.24

Please cite as:

Dolling-Boreham RM, Mohan A, Abdelhack M, Elton-Marshall T, Hamilton HA, Boak A, Felsky D

Identifying Psychosocial and Ecological Determinants of Enthusiasm In Youth: Integrative Cross-Sectional Analysis Using Machine Learning

JMIR Public Health Surveill 2024;10:e48705

URL: <https://publichealth.jmir.org/2024/1/e48705>

doi: [10.2196/48705](https://doi.org/10.2196/48705)

PMID: [39264706](https://pubmed.ncbi.nlm.nih.gov/39264706/)

©Roberta M Dolling-Boreham, Akshay Mohan, Mohamed Abdelhack, Tara Elton-Marshall, Hayley A Hamilton, Angela Boak, Daniel Felsky. Originally published in JMIR Public Health and Surveillance (<https://publichealth.jmir.org>), 12.09.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <https://publichealth.jmir.org>, as well as this copyright and license information must be included.